

Институт проблем управления
им. В.А. Трапезникова РАН

УПРАВЛЕНИЕ БОЛЬШИМИ СИСТЕМАМИ

Выпуск 112
Ноябрь 2024

**СБОРНИК
ТРУДОВ**

ISSN 1819-2467

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

Москва – 2024

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

**УПРАВЛЕНИЕ
БОЛЬШИМИ
СИСТЕМАМИ**

СБОРНИК ТРУДОВ

Выпуск 112

Москва – 2024

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. Щербakov 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

Горбунова А.В.

Оценки копулы и квантилей распределения времени отклика системы с разделением и параллельным обслуживанием заявок и распределением Парето времени обслуживания 7

Моисеева С.П., Панкратова Е.В.

Асимптотический анализ многопоточной гетерогенной СМО в условии предельно редких изменений состояний управляющей входящими потоками цепи Маркова..... 30

Пауль С.В., Назаров А.А., Лапатин И.Л.

Исследование двумерного маркированного ММРР в предельном условии высокой интенсивности 45

Математическая теория управления

Мухин А.В.

Об исключительном случае в задаче о размещении полюсов..... 64

Анализ и синтез систем управления

Юрченков А.В.

Синтез управления для стационарных систем при дополнительных ограничениях на внешнее возмущение в рамках анизотропной теории 74

Сетевые модели в управлении

Гасанов И.И.

Проект системы распределения заказа внутри ассоциации производителей посредством множественных аукционов..... 95

Губанов Д.А., Чхартишвили А.Г.

Влияние алгоритмов ранжирования, ботов и модерации контента на формирование мнений в социальной сети 109

Управление в социально-экономических системах

Авдеева З.К., Коврига С.В.

Интеграционное прогнозирование нестационарных процессов, представленных временными рядами. Обзор..... 129

Базилевский М.П.

Применение аппарата математического программирования для построения неэлементарных квазилинейных регрессий..... 168

Чернов И.В.

Методология сценарного планирования и управления в системе обеспечения национальной безопасности .. 187

***Управление техническими системами
и технологическими процессами***

**Галин Р.Р., Мамченко М.В., Галина С.Б.,
Зорин В.А.**

Подход к распределению задач технологического процесса среди неоднородных участников КРТС с учетом их состояния 233

Управление подвижными объектами и навигация

Трефилов П.М.

Алгоритм интеграции сенсорных данных для оценки состояния автономных роботов в интеллектуальной транспортной среде..... 257

***Надежность и диагностика средств и систем
управления***

Острер Л.А., Русев В.Н., Скориков А.В.

Асимптотика моментов и их производных для избыточных распределений..... 274

*Программы и системы моделирования объектов,
средств и систем управления*

**Граничин О.Н., Граничина О.А., Трофимов С.А.,
Щербаков П.С.**

*Калибровка большого массива ультразвуковых
сенсоров.....* 294

**Лемперт А.А., Жарков М.Л., Казаков А.Л.,
Ву Х.З.**

*Моделирование морского контейнерного терминала
с использованием сети массового обслуживания* 310

Технические и программные средства управления

Роганов В.Р., Короп Г.В., Михеев М.Ю.

*Имитатор визуальной обстановки, синтезирующий
в зрительном диапазоне электромагнитных волн
внешний вид пространства вокруг кабины
авиационного тренажёра* 338

ОЦЕНКИ КОПУЛЫ И КВАНТИЛЕЙ РАСПРЕДЕЛЕНИЯ ВРЕМЕНИ ОТКЛИКА СИСТЕМЫ С РАЗДЕЛЕНИЕМ И ПАРАЛЛЕЛЬНЫМ ОБСЛУЖИВАНИЕМ ЗАЯВОК И РАСПРЕДЕЛЕНИЕМ ПАРЕТО ВРЕМЕНИ ОБСЛУЖИВАНИЯ

Горбунова А. В.¹

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

Рассматривается система с разделением и параллельным обслуживанием заявок. Предполагается, что распределение времени обслуживания на всех приборах имеет распределение Парето. Изучается зависимость между временами пребывания подзаявок в подсистемах, являющаяся основной причиной сложности анализа подобных систем. Время пребывания заявки в системе (или среднее время отклика) является максимумом из зависимых случайных величин пребывания подзаявок в системе. Получены приближения совместного распределения времен пребывания подзаявок с помощью теории копул. Также предложен подход для определения квантилей распределения времени отклика системы с помощью диагонального сечения копул. Данный подход ранее применялся для случая анализа аналогичной системы, но с экспоненциальным распределением времени обслуживания. Однако основное отличие исследуемой системы от экспоненциального случая заключается в том, что вид функции распределения времени пребывания подзаявки в подсистеме неизвестен. Поэтому используется аналитическое приближение для квантилей распределения времени отклика в подсистеме в предположении полученной ранее аппроксимации распределения времени пребывания подзаявки в подсистеме распределением Фреше. Оценки, полученные для квантилей и копулы распределения времени отклика, показывают хорошее соответствие с данными имитационного моделирования.

Ключевые слова: система с параллельным обслуживанием заявок, система массового обслуживания, время отклика, квантили распределения, копула.

1. Введение

В статье исследуется частный случай fork-join-системы, или, что то же самое, системы с разделением и параллельным обслуживанием заявок. Рассматривается система, состоящая из двух

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

подсистем. Предполагается, что входящий поток является пуассоновским, а время обслуживания на каждом из приборов имеет распределение Парето. Более детальное описание системы будет представлено в следующем разделе.

Fork-join-системы массового обслуживания (СМО) используются для моделирования процессов функционирования различных структур, предполагающих разделение исходной задачи на более мелкие составляющие с последующей их параллельной обработкой. Распараллеливание решаемой задачи является одним из способов повышения эффективности производительности физических систем и применяется в различных сферах: в области информационных технологий (параллельные или распределенные вычисления), в сфере материального производства (изготовление продукции, состоящей из множества деталей), логистике (сборка и доставка в пункт назначения многокомпонентного заказа) и т.п. [2, 3, 5, 16, 18].

Анализ подобных систем является довольно сложным, точные характеристики (среднее время отклика системы) получены только для случая двух подсистем с экспоненциальным распределением времени обслуживания [13]. Для систем с неэкспоненциальными временами обслуживания уже даже в случае двух подсистем известны только аппроксимации различной точности в определенных диапазонах параметров [14, 15, 19]. С оценками для моментов времени пребывания заявки в системе с распределением Парето времени обслуживания можно ознакомиться, например, в [6].

В настоящей работе исследуется fork-join-система с точки зрения оценки не менее важных характеристик случайной величины времени отклика, чем его моменты. А именно, предлагается подход к оценке квантилей распределения времени отклика системы, основанный на элементах теории копул. Более подробно с элементами теории можно ознакомиться, например, в [8, 9, 10, 12]. Копулы, наряду с коэффициентами корреляции, в полной мере позволяют охарактеризовать зависимость между случайными величинами. В данном случае речь идет о наличии

зависимости между временами пребывания подзаявок, на которые разделяется заявка, в соответствующих подсистемах. Среди работ, посвященных оценке квантилей распределения времени пребывания заявки в fork-join СМО с различными распределениями можно отметить [1, 7, 14, 15, 17]. Здесь предлагается подход, успешно примененный к оценке квантилей в экспоненциальном случае. Однако в случае распределения Парето закон (функция) распределения времени пребывания подзаявки в подсистеме неизвестен, поэтому за основу взято приближение для квантилей времени отклика в базовом случае из [1]. Оценки, полученные для квантилей и копулы распределения времени отклика, показывают хорошее соответствие с эмпирическими данными.

Статья организована следующим образом: во втором разделе подробно описана математическая модель исследуемой fork-join системы, в третьем разделе описывается подход к построению оценок квантилей времени отклика с помощью диагонального сечения копулы, в следующем разделе приводится описание метода оценки самой копулы, в заключении подводятся некоторые итоги.

2. Математическая модель системы с разделением и параллельным обслуживанием

Рассматривается система с разделением и параллельным обслуживанием заявок, при поступлении в которую заявка мгновенно разделяется на $K = 2$ подзаявки. После деления каждая из подзаявок в свою очередь поступает на обслуживание с единственным прибором (рис. 1). Входящий поток заявок является пуассоновским с интенсивностью $\lambda > 0$. Время обслуживания имеет распределение Парето с функцией распределения следующего вида:

$$(1) \quad B(x) = 1 - \left(\frac{\alpha - 1}{\alpha} \cdot \frac{1}{x} \right)^\alpha, \quad x \geq \frac{\alpha - 1}{\alpha}, \quad \alpha > 3.$$

Тогда среднее время обслуживания b составляет одну условную временную единицу, а второй момент времени обслуживания $b^{(2)}$ определяется соотношением, зависящим от параметра распределе-

ния α , т.е.

$$b_{Pa} = 1, \quad b_{Pa}^{(2)} = \frac{(\alpha - 1)^2}{\alpha(\alpha - 2)}.$$

Фактически каждая из подзаявок поступает на обслуживание в подсистему типа $M|Pa|1$. Коэффициент загрузки системы определяется как $\rho = \lambda b_{Pa} = \lambda < 1$.

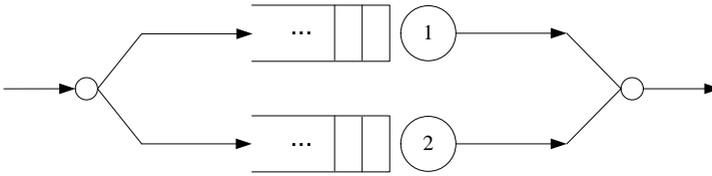


Рис. 1. Модель fork-join СМО с двумя подсистемами типа $M|Pa|1$

Целая заявка может покинуть систему только после окончания обслуживания обеих составляющих ее подзаявок. При этом считается, что сборка заявки (из обслуженных подзаявок) происходит мгновенно. Поэтому если обозначить времена пребывания каждой из двух подзаявок в соответствующих подсистемах через случайные величины ξ_1 и ξ_2 , то время пребывания заявки будет являться максимумом из этих случайных величин:

$$R_2 = \max\{\xi_1, \xi_2\}.$$

Стоит отметить, что случайные величины ξ_1 и ξ_2 коррелируют между собой, причем их зависимость является положительной. Наличие корреляции заметно усложняет анализ fork-join систем. По этой причине представляет интерес исследование систем с разделением и параллельным обслуживанием с точки зрения изучения имеющейся зависимости, которую характеризуют не только коэффициенты корреляции, но в большей степени копулы. Копулы позволяют перейти от частных распределений случайных величин к их совместному распределению, тем самым

предполагая полноценное описание зависимости двух величин (в данном случае двух).

Далее будет описан подход для оценки копулы и квантилей распределения времени отклика на примере, когда параметр $\alpha = 4$. При этом предложенный подход может распространяться и на другие значения α .

3. Подход к оценке квантилей распределения времени отклика с помощью диагонального сечения копулы

В статье [1] был предложен подход для оценки квантилей распределения времени отклика fork-join-системы с распределением Парето времени обслуживания и различными вариантами распределений для входящего потока, а именно, распределением Эрланга, гиперэкспоненциальным распределением и пуассоновским входящим потоком. В рамках предложенного подхода была обоснована допустимость использования распределения Фреше для аппроксимации распределения времени отклика R_K fork-join-системы с распределением Парето времени обслуживания и числом $K \geq 2$ подсистем.

Далее применялся метод моментов для нахождения оценок параметров \hat{a} и \hat{b} , участвующих в выражении для приближения квантилей времени отклика \hat{x}_p уровня p , в результате было получено следующее аналитическое выражение:

$$(2) \quad \hat{x}_{p,R_K} = \hat{a}_K + \hat{b}_K (-\ln p)^{-1/\alpha}, \quad 0 < p < 1, \quad 2 \leq K \leq 20,$$

где \hat{a}_K и \hat{b}_K определялись с помощью оценок математического ожидания $E[R_K]$ и дисперсии времени отклика $Var[R_K]$ fork-join СМО.

Поскольку рассматривались различные варианты распределений для времен между соседними поступлениями заявок, а формул для оценок математического ожидания и дисперсии времени отклика fork-join-систем известно не так много, то в качестве оценок $E[R_K]$ и $Var[R_K]$ использовались значения, полученные с помощью имитационного моделирования.

В данной работе предлагается еще один подход к оцениванию квантилей времени отклика fork-join СМО для частного случая системы с разделением и параллельным обслуживанием, когда число подсистем $K = 2$. Этот подход тесно связан с элементами теории копул и аналогичен подходу, предложенному в [7] для fork-join-системы с двумя подсистемами типа $M|M|1$. Однако основное отличие случая подсистем типа $M|Pa|1$ от случая экспоненциального обслуживания заключается в том, что вид функции распределения времени пребывания подзаявки в данной подсистеме неизвестен. В то время как для подсистемы типа $M|M|1$ время пребывания подзаявки имеет также экспоненциальное распределение. Этим фактором и обуславливается использование аппроксимации вида (2).

Выражение (2) применимо не только для диапазона значений $2 \leq K \leq 20$, который рассматривался в более ранних работах при анализе fork-join СМО с распределением Парето времени обслуживания вида (1), но в том числе и для базового случая, а именно, когда $K = 1$. Поскольку подход основывается на оценке двумерных копул, то предлагается использовать результаты оценки для базового случая $K = 1$ при оценке квантилей в случае $K = 2$. При этом в выражении (2) вероятность p необходимо будет заменить на его оценку, которая будет получена далее при анализе случая $K = 2$ с помощью диагонального сечения копул.

Итак, в дальнейшем будем использовать следующее выражение для оценки квантилей распределения времени отклика при $K = 1$:

$$(3) \quad \hat{x}_{p,R_1} = \hat{a}_1 + \hat{b}_1(-\ln p)^{-1/\alpha}, \quad 0 < p < 1,$$

где для параметров \hat{a}_1 и \hat{b}_1 предположим следующие оценки:

$$(4) \quad \hat{a}_1 = \frac{A_0 + A_1\rho + A_2\rho^2}{1 - \rho}, \quad \hat{b}_1 = \frac{B_0 + B_1\rho + B_2\rho^2}{1 - \rho}.$$

Здесь исходим из того, что времена отклика растут асимптотически пропорционально $1/(1 - \rho)$ при $\rho \rightarrow 1$, а значит, так же ведут себя коэффициенты a и b , так что если мы умножим их на $(1 - \rho)$, то получим какие-то нелинейные ограниченные функции от ρ на отрезке $[0, 1]$, которые попробуем приблизить квадратичными.

Для определения коэффициентов A_i и B_i , $i = 0, 1, 2$, из (4) воспользуемся методом оптимизации Нелдера – Мида [11]. А именно с помощью симуляции системы $M|Pa|1$ получим множество реализаций случайной величины времени пребывания заявки в СМО, после чего статистически на основе эмпирических данных определим для нее квантили распределения x_{p,R_1} . Затем минимизируем с помощью метода Нелдера – Мида модуль относительной погрешности приближения оценки квантилей, рассчитанных по формулам (3) и (4), относительно данных, полученных с помощью имитационного моделирования, что позволит определить искомые коэффициенты:

$$\left| \frac{x_{p,R_1} - \left(\hat{a}_1 + \hat{b}_1 (-\ln p)^{-1/\alpha} \right)}{x_{p,R_1}} \right| \xrightarrow{A_0, A_1, A_2, B_0, B_1, B_2} \min.$$

Для коэффициента загрузки системы рассматривается диапазон значений $\rho = \{0,1; 0,2; \dots; 0,9\}$. А для вероятностей p , т.е. уровней квантилей, выбраны значения $\{0,30; 0,35; \dots; 0,85; 0,90\}$, поскольку, как правило, на практике в большей степени интерес представляют квантили именно более высокого порядка.

В результате получим

$$(5) \quad \begin{aligned} A_0 &\approx 0,129674, & A_1 &\approx -1,650335, & A_2 &\approx 0,316858, \\ B_0 &\approx 0,702442, & B_1 &\approx 0,917551, & B_2 &\approx -0,149744, \end{aligned}$$

Оценки погрешностей приближений для формул (3)–(5) представлены в таблице 1 и для наглядности изображены на рис. 5а.

Таблица 1. Погрешности приближений значений квантилей распределения времени отклика системы x_{p,R_1} ($K = 1$), рассчитанных с помощью аналитических формул (3), (4) и (5) в сравнении с результатами имитационного моделирования

| Оцениваемая характеристика | Типы ошибок | | |
|----------------------------|-------------|------------|----------|
| | Max APE, % | Min APE, % | MAPE, % |
| Квантиль x_{p,R_1} | 5,731694 | 0,004086 | 2,240293 |

В таблице приведены абсолютные значения относительных погрешностей приближений для 117 рассчитанных значений \hat{x}_{p,R_1} : максимальная погрешность приближения (Max APE, %), минимальная погрешность приближения (MinAPE, %) и средняя погрешность (MAPE, %).

Далее допустим, что частная функция распределения случайных величин ξ_1 и ξ_2 времен пребывания подзаявок в подсистемах $M|Pa|1$ имеет вид $G(x)$ и является строго возрастающей. Тогда их совместная функция распределения согласно теореме Склера представима в виде

$$G_{\xi_1, \xi_2}(x_1, x_2) = P(\xi_1 < x_1, \xi_2 < x_2) = C(G(x_1), G(x_2)),$$

где $C(u_1, u_2)$ — это копула-функция совместного распределения случайных величин ξ_1 и ξ_2 . Тогда для случайной величины их максимума $R_2 = \max(\xi_1, \xi_2)$ функция распределения примет вид

$$\begin{aligned} G_{R_2}(x) &= P(\max(\xi_1, \xi_2) < x) = \\ &= P(\xi_1 < x, \xi_2 < x) = C(G(x), G(x)) = C(u, u). \end{aligned}$$

При этом величина

$$\delta(u) = C(u, u), \quad 0 \leq u \leq 1,$$

называется диагональным сечением копула-функции. Соответственно, имеем, что

$$G_{R_2}(x) = C(G(x), G(x)) = \delta(G(x)).$$

Поэтому уравнение для определения квантили времени отклика принимает вид

$$G_{R_2}(x_{p,R_2}) = \delta(G(x_{p,R_2})) = p,$$

откуда

$$(6) \quad x_{p,R_2} = x_p = G^{-1}(\delta^{-1}(p)) = x_{\delta^{-1}(p), R_1}.$$

Стоит отметить, что диагональное сечение характеризуется (необходимыми и достаточными) свойствами

$$(7) \quad \begin{aligned} & \max\{2u - 1, 0\} \leq \delta(u) \leq u; \\ & 0 \leq \delta(u_2) - \delta(u_1) \leq 2(u_2 - u_1), \quad 0 \leq u_1 \leq u_2 \leq 1, \end{aligned}$$

которыми обладает, в частности, степенная функция

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

При этом значение параметра $\beta = 1$ определяет абсолютную положительную зависимость случайных величин, а значение $\beta = 2$ — их полную независимость. Поэтому при дальнейших исследованиях обратимся к копуле Гумбеля

$$(8) \quad \begin{aligned} C(u_1, u_2) &= \exp\{-((-\ln u_1)^\theta + (-\ln u_2)^\theta)^{1/\theta}\}, \\ &\theta \geq 1, \quad u_1, u_2 \in [0, 1], \end{aligned}$$

так как она характеризуется степенным диагональным сечением следующего вида:

$$(9) \quad \delta(u) = u^{2^{1/\theta}}.$$

Введем случайные величины $U_1 = G(\xi_1)$ и $U_2 = G(\xi_2)$, согласно методу обратного преобразования Н.В. Смирнова случайные величины U_1 и U_2 будут иметь равномерное распределение на отрезке $[0, 1]$, т.е.

$$U_i = G(\xi_i) \sim R[0, 1], \quad i = 1, 2.$$

Затем рассмотрим случайную величину $V = \max(U_1, U_2)$, для которой в силу предположения о том, что функция распределения $G(x)$ является строго возрастающей, будет справедливо следующее:

$$(10) \quad \begin{aligned} V &= \max(U_1, U_2) = \max(G(\xi_1), G(\xi_2)) = \\ &= G(\max(U_1, U_2)) = P(V < u) = p, \end{aligned}$$

где p — это вероятность, $0 \leq p \leq 1$.

С другой стороны, для диагонального сечения справедливо

$$\begin{aligned} \delta(u) &= C(u, u) = P(\xi_1 < x, \xi_2 < x) = \\ &= P(G(\xi_1) < G(x), G(\xi_2) < G(x)) = P(U_1 < u, U_2 < u) = \end{aligned}$$

$$= P(\max(U_1, U_2) < u) = P(V < u) = p,$$

т.е.

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

где u_p – это квантиль уровня p распределения случайной величины V . В условиях предположения о степенном виде диагонального сечения получаем выражение для определения вероятности p , которое впоследствии будет использоваться для оценки квантилей времени отклика R_2 :

$$(11) \quad p = \delta(u_p) = u_p^\beta.$$

Далее с помощью имитационного моделирования системы с разделением и параллельным обслуживанием заявок определяется множество пар значений величин (ξ_1, ξ_2) . Речь идет о порядке 5–10 миллионов таких пар, т.е. получаем $N = 5$ миллионов (для низких уровней загрузки системы) или $N = 10$ миллионов (для высоких уровней загрузки) реализаций случайных величин (ξ_1^i, ξ_2^i) , где $i = 1, \dots, N$.

Затем для каждой пары (ξ_1^i, ξ_2^i) необходимо определить величину V_i , которая согласно (10) определяется выражением $G(\max[G(\xi_1^i), G(\xi_2^i)]) = G(\max(U_1^i, U_2^i))$. Однако, как уже упоминалось ранее, в силу того, что вид функции распределения $G(x)$ неизвестен, напрямую определить V_i по значениям (ξ_1^i, ξ_2^i) не получится. Поэтому воспользуемся универсальным методом нормированных рангов (описанным, например, в [10, §5.5.2]), согласно которому можно использовать асимптотическую оценку

$$\begin{aligned} \hat{V}_i &= \max \left(\hat{G}(\xi_1^i), \hat{G}(\xi_2^i) \right) = \max \left(\hat{U}_1^i, \hat{U}_2^i \right) = \\ &= \max \left(\frac{\text{rang}(\xi_1^i)}{N+1}, \frac{\text{rang}(\xi_2^i)}{N+1} \right), \end{aligned}$$

где $\text{rang}(\cdot)$ – это ранг, т.е. порядковый номер реализации аргумента после сортировки в порядке возрастания, $i = 1, \dots, N$.

Далее для полученных значений реализаций случайной величины \hat{V} оцениваем квантили ее распределения, т.е. находим ста-

статистические значения (u_p, p) , упорядочивая значения \hat{V}_i по возрастанию, в результате получаем

$$(u_p, p) = \left(\hat{V}_{(k)}, \frac{k}{N+1} \right),$$

где $\hat{V}_{(k)}$ – это k -я порядковая статистика, $k = 1, \dots, N$. Пары (u_p, p) оцениваются, как и ранее, для различных значений коэффициента загрузки системы $\rho = \{0,10; 0,15; \dots; 0,90\}$ и вероятностей $p = \{0,30; 0,25; \dots; 0,85; 0,90\}$.

Далее на имеющихся данных (u_p, p) для каждого выбранного значения коэффициента загрузки ρ проводится графический анализ для определения конкретного вида функциональной зависимости p от u_p из (11)

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

На рис. 2 изображена зависимость $\ln p$ от значений $\ln u_p$.

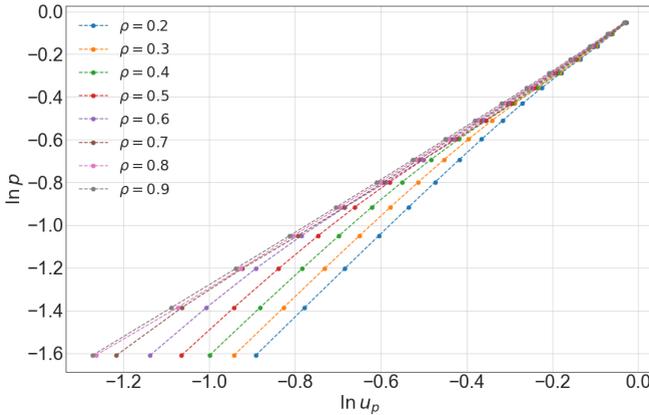


Рис. 2. Зависимость $\ln p$ от $\ln u_p$

Как следует из вида графика, между $\ln p$ и $\ln u_p$ можно допустить линейную зависимость, причем пучок прямых проходит через начало координат, соответственно, свободный коэффициент отсутствует, т.е. получаем соотношение

$$\ln p \approx \widehat{\beta}(\rho) \cdot \ln u_p,$$

что соответствует предположению о степенной зависимости p от u_p .

Далее подберем функциональную зависимость для $\widehat{\beta}(\rho)$. Для этого построим на имеющихся данных график зависимости отношения $\ln p / \ln u_p$. Из вида рис. 3 можно предложить квадратичную зависимость.

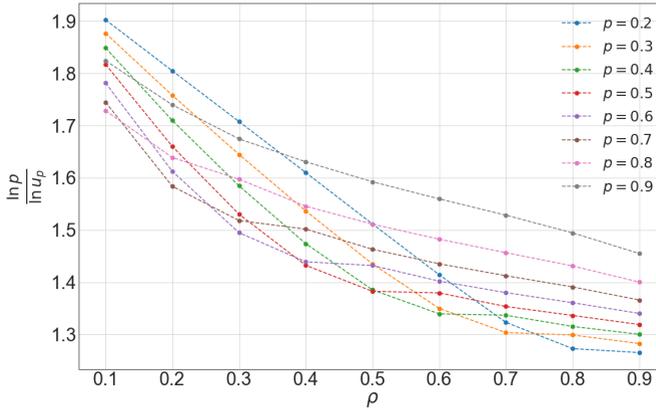


Рис. 3. Зависимость $(\ln p / \ln u_p)$ от ρ

При $\rho \rightarrow 0$ времена пребывания подзаявок асимптотически независимы, что в соответствии с теорией копул дает $\beta \rightarrow 2$. Поэтому логично допустить, что кривые проходят через точку с координатами $(0, 2)$, при этом вершина параболы сдвинута вправо и обращена вниз, что говорит об отрицательном значении числового коэффициента перед ρ и положительном при ρ^2 . Таким образом, имеем

$$\frac{\ln p}{\ln u_p} = \beta(\rho) \approx 2 - C_1\rho + C_2\rho^2.$$

В результате получаем следующее выражение:

$$(12) \quad p = \delta(u_p, \rho) \approx u_p^{2 - C_1\rho + C_2\rho^2}.$$

Теперь необходимо найти значения коэффициентов C_1 и C_2 . Для этого воспользуемся методом оптимизации Нелдера – Мида

и минимизируем модуль относительной погрешности полученного аналитического (степенного) приближения \hat{p} по сравнению со значениями p , полученными с помощью имитационного моделирования:

$$\left| \frac{p - u_p^{2-C_1\rho+C_2\rho^2}}{p} \right| \xrightarrow{C_1, C_2} \min .$$

Таким образом, значения коэффициентов равны
 (13) $C_1 \approx 1,334476, \quad C_2 \approx 0,550919,$

соответственно, искомая оценка имеет вид

$$(14) \quad p = \delta(u_p, \rho) \approx u_p^{2-1,334476\rho+0,550919\rho^2}.$$

На рис. 4 представлены результаты имитационного моделирования вероятностей или уровней p квантилей u_p случайной величины $V = G(\max(U_1, U_2))$ в сравнении с результатами вычислений по аналитической формуле (14) в диапазоне значений $[0,30; 0,90]$ с шагом 0,05.

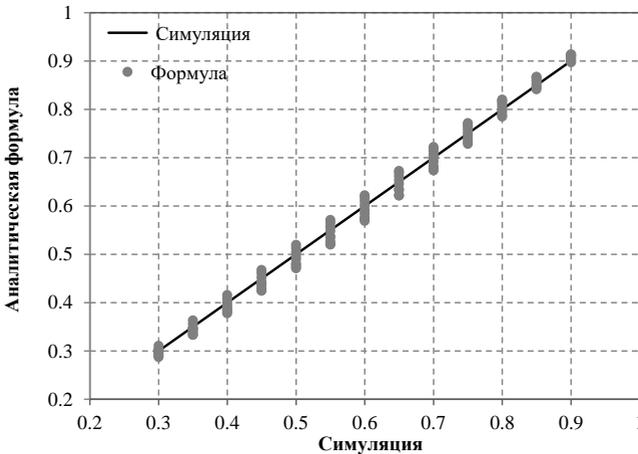


Рис. 4. Сравнение аналитических результатов формулы (14) с имитационным моделированием значений p квантилей u_p случайной величины $V = G(\max(U_1, U_2))$ для значений $\rho \in \{0,10; 0,20; \dots; 0,90\}$

Каждая точка, изображенная на графике, представляет собой множество из 9 точек по числу значений коэффициента загрузки $\rho \in \{0,10; 0,20; \dots, 0,90\}$, которые накладываются друг на друга. Для ясности в таблице 2 приведены абсолютные значения относительных погрешностей приближений для 117 рассчитанных значений p .

Таблица 2. Погрешности приближений значений вероятностей p , рассчитанных с помощью аналитической формулы (14) в сравнении с результатами имитационного моделирования

| Оцениваемая характеристика | Типы ошибок | | |
|----------------------------|-------------|------------|----------|
| | Max APE, % | Min APE, % | MAPE, % |
| p из (14) | 5,731694 | 0,061445 | 2,311958 |

Сравнительная анализ результатов имитационного моделирования уровня p квантилей u_p с.в. V с результатами аналитической формулы (14) показал, что средняя погрешность приближения составляет около 2,2%.

Далее, поскольку вид функции распределения $G(x)$ случайных величин ξ_1 и ξ_2 времен пребывания подзаявок в подсистемах $M|Pa|1$ неизвестен, то воспользуемся соотношением (3) для квантилей. А именно, при $K = 1$ имеем

$$G(x_{p,R_1}) = p, \quad x_{p,R_1} = G^{-1}(p) = \hat{a}_1 + \hat{b}_1(-\ln p)^{-1/\alpha}.$$

Теперь с учетом уравнения (6) для определения квантилей времени отклика R_2 в случае $K = 2$ выразим $\delta^{-1}(p)$:

$$\begin{aligned} \delta(u_p) &= u_p^{2-C_1\rho+C_2\rho^2} = p, \\ (2 - C_1\rho + C_2\rho^2) \ln u_p &= \ln p, \\ u_p = \delta^{-1}(p) &= p^{\frac{1}{2-C_1\rho+C_2\rho^2}}. \end{aligned}$$

Поэтому можем записать следующее

$$x_{p,R_2} = x_p = G^{-1}(\delta^{-1}(p)) = G^{-1}\left(p^{\frac{1}{2-C_1\rho+C_2\rho^2}}\right) =$$

$$= \hat{x}_p = \hat{a}_1 + \hat{b}_1 \left(-\ln p^{\frac{1}{2-C_1\rho+C_2\rho^2}} \right)^{-1/\alpha}.$$

В результате окончательно получаем следующее выражение для оценки квантилей времени отклика R_2 уровня p , $0,30 \leq p \leq 0,90$:

$$(15) \quad \hat{x}_p = \hat{a}_1 + \hat{b}_1 \left(-\frac{\ln p}{2 - 1,334476\rho + 0,550919\rho^2} \right)^{-1/\alpha},$$

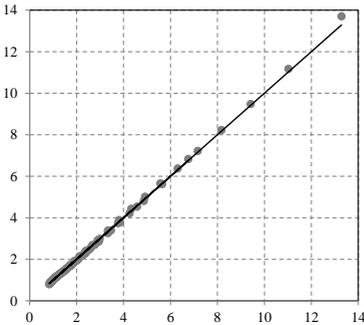
$$0 < p < 1,$$

где

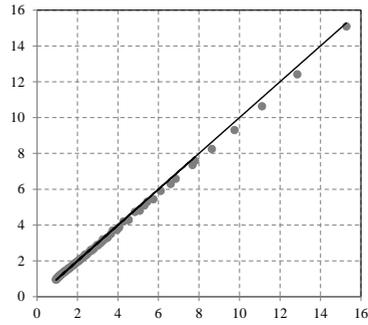
$$(16) \quad \hat{a}_1 \approx \frac{0,129674 - 1,650335\rho + 0,316858\rho^2}{1 - \rho},$$

$$\hat{b}_1 \approx \frac{0,702442 + 0,917551\rho - 0,149744\rho^2}{1 - \rho}.$$

Далее аналогично оценим качество аппроксимации полученного выражения (15) для 117 рассчитанных значений квантилей при $\rho \in \{0,10; 0,20; \dots; 0,90\}$ и $p \in \{0,20; 0,25; \dots; 0,90\}$. Результаты представлены в таблице 3 и на рис. 5б. Получаем, что максимум модуля относительной ошибки составляет около 5,7%, а среднее значение этого модуля равно примерно 2,2%.



а)



б)

Рис. 5. Сравнение эмпирических и аналитических квантилей распределения случайной величины времени отклика R_K системы с разделением и параллельным обслуживанием: а) рассчитанных по формулам (3)–(5), $K = 1$; б) рассчитанных по формулам (15)–(16), $K = 2$

Таблица 3. Погрешности приближений значений квантилей распределения времени отклика системы x_p ($K = 2$), рассчитанных с помощью аналитических формул (15) и (16) в сравнении с результатами имитационного моделирования

| Оцениваемая характеристика | Типы ошибок | | |
|----------------------------|-------------|------------|----------|
| | Max APE, % | Min APE, % | MAPE, % |
| Квантиль x_{p,R_2} | 5,476451 | 0,001325 | 2,949008 |

4. Приближение копулы времен пребывания подзаявок в подсистемах копулой Гумбеля

В предыдущем разделе была получена оценка диагонального сечения копулы $\delta(u)$. В данном разделе будет представлено аналитическое выражение, оценивающее саму копулу $C(u_1, u_2)$. Для этого потребуются эмпирические данные, проанализировав которые можно будет сделать вывод о близости исследуемой копулы к одному из известных семейств, в частности, копулам Гумбеля.

Алгоритм построения эмпирической копулы будет следующим:

1) имитационное моделирование множества пар (ξ_1^k, ξ_2^k) случайных величин времен пребывания в подсистемах $M|Pa|1$ fork-join СМО, где k – порядковый номер смоделированной пары значений, $k = 1, \dots, N$, N – объем выборки (общее число пар случайных величин);

2) преобразование случайных величин (ξ_1^k, ξ_2^k) методом нормированных рангов (см. [10, §5.5.2]) в случайные величины с асимптотически равномерным распределением на отрезке $[0, 1]$, $U_i \sim R[0, 1]$, $i = 1, 2$:

$$(U_1^k, U_2^k) = \left(\frac{\text{rang}(\xi_1^k)}{N + 1}, \frac{\text{rang}(\xi_2^k)}{N + 1} \right);$$

3) разбиение единичного квадрата на более мелкие квадраты (сетку) со сторонами длиной $h = 1/m$, где, например, $m = 20$ и определение числа точек (U_1^k, U_2^k) , попадающих в каждый из квадратов, вершинами которого являются точки $(0, 0)$, $(ih, 0)$,

$(0, jh), (ih, jh), i, j = 1, \dots, m$, и нормирование полученного значения, т.е.

$$C_{ij} = C(ih, jh) \approx \hat{C}_{ij} = \frac{1}{N} \sum_{k=1}^N \mathbf{1}\{U_1^k < ih, U_2^k < jh\},$$

где $\mathbf{1}\{\cdot\}$ – функция-индикатор события $\{\cdot\}$.

На рис. 6 представлен график эмпирической копулы или, что то же самое, совместной функции распределения случайного вектора (U_1, U_2) , построенной в соответствии с представленным выше алгоритмом.

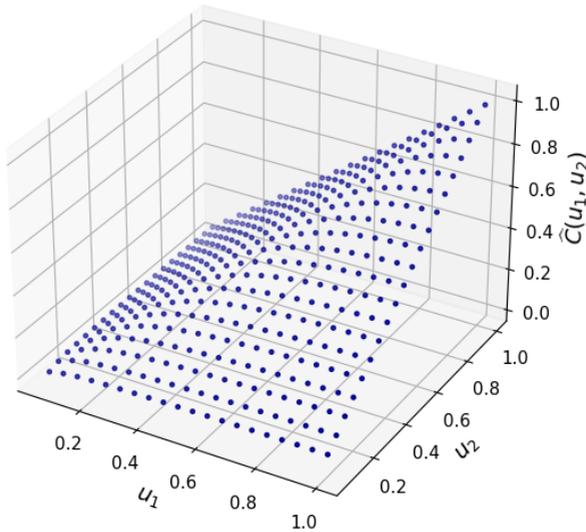


Рис. 6. Эмпирическая копула $\hat{C}(u_1, u_2)$

Исходя из внешнего вида полученной эмпирической функции на рис. 6, а также учитывая, что диагональное сечение рассматриваемой копулы было приближено в предыдущем разделе выражением вида

$$(17) \quad \delta(u) \approx u^\beta, \quad \beta = 2 - C_1\rho + C_2\rho^2,$$

будем приближать искомую копулу $C(u_1, u_2)$ копулой Гумбеля, которая имеет вид

$$(18) \quad C_g(u_1, u_2) = \exp\{-[(-\ln u_1)^\theta + (-\ln u_2)^\theta]^{\frac{1}{\theta}}\},$$

где $\theta \in [1, +\infty)$ – параметр копулы, который предстоит оценить.

Поскольку для копулы Гумбеля диагональное сечение имеет вид

$$\delta_g(u) = C_g(u, u) = u^{2^{1/\theta}},$$

то с учетом (17) получаем, что

$$(19) \quad \theta \approx \frac{\ln 2}{\ln \beta} = \frac{\ln 2}{\ln(2 - C_1\rho + C_2\rho^2)}.$$

Далее снова воспользуемся методом оптимизации Нелдера – Мида для минимизации модуля относительной ошибки приближения функции копулы Гумбеля (18) с учетом того, что параметр θ определяется выражением (19), при сравнении с «истинными» значениями функции копулы Гумбеля, полученными с помощью имитационного моделирования для различных коэффициентов загрузки $\rho \in \{0,10; 0,20; \dots; 0,90\}$. Как и раньше, не будем рассматривать квантили низкого уровня, т.е. пусть $u_1, u_2 \in \{0,30; 0,35; \dots; 0,90\}$. В результате получаем следующие значения искоемых коэффициентов:

$$(20) \quad C_1 \approx 1,068768, \quad C_2 \approx 0,202125,$$

поэтому

$$(21) \quad C(u_1, u_2) \approx \exp \left\{ - \left((-\ln u_1)^{\frac{\ln 2}{\ln(2-1,069\rho+0,202\rho^2)}} + \right. \right. \\ \left. \left. + (-\ln u_2)^{\frac{\ln 2}{\ln(2-1,069\rho+0,202\rho^2)}} \right)^{\frac{\ln(2-1,069\rho+0,202\rho^2)}{\ln 2}} \right\}.$$

Как видно из (13) и (20), полученные оценки коэффициентов различны, однако соответствующие им оценки функции $\beta(\rho)$ мало различаются между собой на рассматриваемом промежутке загрузки ρ , что позволяет говорить о значительной согласованности между ними.

Что касается погрешности аппроксимации формулы (21), то в таблице 4 представлены значения максимальной (Max APE),

минимальной (Min APE) и средней относительной ошибки аппроксимации (MAPE), первая из которых не превышает 10%, на наборе данных из 1521 троек (ρ, u_1, u_2) . В той же таблице приведены результаты в случае оценки по коэффициентам из (13). На рис. 7 также представлены графики эмпирической функции копулы и копулы, определяемой выражением (21) на заданном диапазоне значений $0,3 \leq u_1, u_2 \leq 0,9$.

Таблица 4. Погрешности приближений функции копулы Гумбеля $C(u_1, u_2)$ формулой (21)

| Оцениваемая характеристика | Типы ошибок | | |
|---------------------------------------|-------------|------------|----------|
| | Max APE, % | Min APE, % | MAPE, % |
| $C(u_1, u_2)$, C_1, C_2 из (20) | 8,248810 | 0,000686 | 2,735456 |
| $C(u_1, u_2)$, C_1, C_2 из (13) | 14,733185 | 0,000385 | 1,999350 |

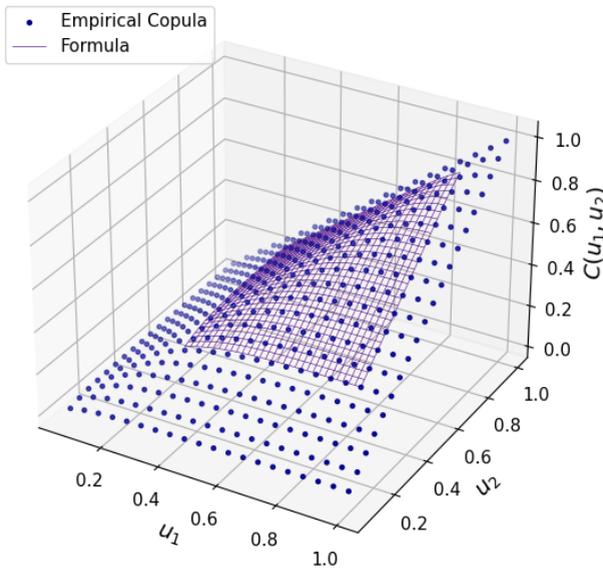


Рис. 7. Сравнение эмпирической копулы $\hat{C}(u_1, u_2)$ и аналитической формулы (21)

Заметим также, что если проводить оценку параметра θ классическим методом максимального правдоподобия (соответствующей функцией Python для копулы Гумбеля), то полученные значения, количество которых в данном случае будет соответствовать количеству значений коэффициента корреляции $\rho \in [0,1; 0,9]$ с шагом 0,10 (т.е. их будет всего 9), на тех же 1521 тройках значений (ρ, u_1, u_2) приближение копулой Гумбеля показывает большие погрешности. В этом случае Max APE $\approx 15,7\%$, Min APE $\approx 0,0009\%$ и MAPE $\approx 2,34\%$.

5. Заключение

В настоящей работе изучены приближения совместного распределения времен пребывания подзаявок с помощью теории копул. Получено хорошее соответствие с данными для степенных диагональных сечений. На основе оценок диагональных сечений выведены оценки квантилей времени отклика в широком диапазоне значений вероятностей или уровней квантилей, а также значений коэффициента загрузки системы. Несмотря на то, что в рамках данной работы рассматривалось определенное значение параметра распределения Парето времени обслуживания, представленный в статье подход, основанный на элементах теории копул, можно применить аналогичным образом и для других значений параметра, а также обобщить на системы с большим количеством подсистем.

Литература

1. ГОРБУНОВА А.В., ЛЕБЕДЕВ А.В. *Квантили распределения времени отклика в fork-join системах с распределением Парето времени обслуживания* // Вестник ВГУ. Серия: Системный анализ и информационные технологии. – 2024. – №3. – С. 5–16.
2. ARMONY M., ISRAELIT S., MANDELBAUM A. et al. *Patient flow in hospitals: a data-based queueing-science*

- perspective* // Stochastic Systems. – 2015. – Vol. 5, No. 1. – P. 146–194.
3. BACCELLI F., MAKOWSKI A.M. *Queueing models for systems with synchronization constraints* // Proc. of the IEEE. – 1989. – Vol. 77, No. 1. – P. 138-161.
 4. ENGANTI P., ROSENKRANTZ T., SUN L. et al. *ForkMV: Mean-and-Variance Estimation of Fork-Join Queueing Networks for Datacenter Applications* // IEEE Int. Conf. on Networking, Architecture and Storage (NAS). – 2022. – P. 1–8.
 5. GALLIEN J., WEIN L.M. *A simple and effective component procurement policy for stochastic assembly Systems* // Queueing Systems. – 2001. – Vol. 38. – P. 221–248.
 6. GORBUNOVA A.V., LEBEDEV A.V. *Nonlinear approximation of characteristics of a fork-join queueing system with Pareto service as a model of parallel structure of data processing* // Mathematics and Computers in Simulation. – 2023. – Vol. 214. – P. 409–428. – DOI: <https://doi.org/10.1016/j.matcom.2023.07.029>.
 7. GORBUNOVA A.V., LEBEDEV A.V. *Copulas and quantiles in fork-join queueing Systems* // Advances in Systems Science and Applications. – 2024. – Vol. 24, No. 1. – P. 1–19.
 8. GUDENDORF G., SEGERS J. *Extreme-Value Copulas* // In: Copula theory and Its Application. – Springer, 2010. – P. 127–145.
 9. LEBEDEV A.V. *On the Interrelation between Dependence Coefficients of Bivariate Extreme Value Copulas* // Markov Proc. Relat. Fields. – 2019. – Vol. 25, No. 4. – P. 639–648.
 10. MCNEIL A.J., FREY R., EMBRECHTS P. *Quantitative risk management*. – Princeton: Princeton University Press, 2005. – 538 p.
 11. NELDER J.A., MEAD R. *A simplex method for function minimization* // The Computer Journal. – 1965. – Vol. 7. – P. 308–313.
 12. NELSEN R.B. *An introduction to copulas*. – Springer Science & Business Media, 2007. – 272 p.

13. NELSON R., TANTAWI A.N. *Approximate analysis of fork/join synchronization in parallel queues* // IEEE Trans. Comput. – 1988. – Vol. 37, No. 6. – P. 739–743.
14. NGUYEN M., ALESAWI S., LI N. et al. *ForkTail: A black-box fork-join tail latency prediction model for user-facing datacenter workloads* // Proc. of the 27th Int. Symposium on High-Perform. Parallel Distrib. Comput.. – 2018. – P. 206–217.
15. NGUYEN M., ALESAWI S., LI S. et al. *A black-box fork-join latency prediction model for data-intensive applications* // IEEE Trans. on Parallel and Distributed Systems. – 2020. – Vol. 31, No. 9. – P. 1983–2000.
16. OLIVEIRA D.C.M., LIU J., PACITTI E. *Data-intensive workflow management: for clouds and data-intensive and scalable computing environments* // Synthesis Lectures on Data Management. – 2019. – Vol. 14, No. 4. – P. 1–179.
17. QIU ZH., PEREZ J.F., HARRISON P.G. *Beyond the mean in fork-join queues: Efficient approximation for response-time tails* // Performance Evaluation. – 2015. – Vol. 91. – P. 99–116.
18. SCHOL D., VLASIOU M., ZWART B. *Large fork-join queues with nearly deterministic arrival and service times* // Mathematics of Operations Research. – 2022. – Vol. 47, No. 2. – P. 1335–1364. – DOI: <https://doi.org/10.1287/moor.2021.1171>.
19. THOMASIAN A. *Analysis of fork/join and related queueing systems* // ACM Computing Surveys (CSUR). – 2014. – Vol. 47, No. 2. – P. 17:1–17:71.
20. VARKI E., MERCHANT A., CHEN H. *The M/M/1 fork-join queue with variable subtasks* // Unpublished. – URL: <https://www.cs.unh.edu/~varki/publication/2002-nov-open.pdf> (дата обращения: 02.09.2024).
21. VARMA S., MAKOWSKI A.M. *Interpolation approximations for symmetric fork-join queues* // Performance Evaluation. – 1994. – Vol. 20. – P. 245–265.

**ESTIMATES OF THE COPULA AND QUANTILES
OF THE RESPONSE TIME DISTRIBUTION
FOR A FORK-JOIN QUEUEING SYSTEM
WITH THE PARETO DISTRIBUTION OF SERVICE
TIME**

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

Abstract: A fork-join queueing system is considered. It is assumed that the service time distribution on all servers has a Pareto distribution. The dependence between the sojourn times of subtasks in subsystems is studied, which is the main reason for the complexity of analyzing such systems. The sojourn time of a task in the system (or the average response time) is the maximum of the dependent random variables of the sojourn time of subtasks in the system. Approximations of the joint distribution of the sojourn times of subtasks are obtained using copula theory. An approach is also proposed for determining the quantiles of the system response time distribution using a diagonal section of copulas. This approach was previously used to analyze a similar system, but with an exponential distribution of service time. However, the main difference between the system under study and the exponential case is that the type of the distribution function of the sojourn time of a subtask in the subsystem is unknown. Therefore, an analytical approximation is used for the quantiles of the response time distribution in the subsystem under the assumption that the distribution of the time of stay of a subtask in the subsystem is approximated by the Frechet distribution obtained earlier. The estimates obtained for the quantiles and copula of the response time distribution show good agreement with the simulation data.

Keywords: fork-join queueing system, queueing system, distribution quantiles, copula.

УДК 519.2

ББК 22.17

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

Поступила в редакцию 25.10.2024.

Дата опубликования 30.11.2024.

АСИМПТОТИЧЕСКИЙ АНАЛИЗ МНОГОПОТОЧНОЙ ГЕТЕРОГЕННОЙ СМО В УСЛОВИИ ПРЕДЕЛЬНО РЕДКИХ ИЗМЕНЕНИЙ СОСТОЯНИЙ УПРАВЛЯЮЩЕЙ ВХОДЯЩИМИ ПОТОКАМИ ЦЕПИ МАРКОВА¹

Моисеева С. П.²

(ФГАОУ Национальный исследовательский
Томский государственный университет, Томск)

Панкратова Е. В.³

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

В настоящее время многомодальные системы набирают популярность с развитием многомодальных интерфейсов. Многомодальные потоки представляют собой интегрированные разнотипные потоки, включающие передачу голоса, текстовых данных и видео, поэтому для их описания логично применять непуассоновские модели. В качестве математической модели многомодальной обслуживающей системы рассматривается многопоточная система массового обслуживания с потоками, меняющими свою интенсивность в зависимости от состояний марковской случайной среды. Поступающие требования различных потоков обслуживаются в течении экспоненциально распределенного случайного времени с параметрами, определяемыми типом потока. Ставится задача исследования многомерного марковского процесса числа занятых приборов в системе в стационарном режиме. Используя свойства характеристических функций, получены выражения для нахождения допределительных значений основных вероятностных характеристик числа занятых приборов каждого типа. Асимптотическое исследование проводится в условии предельно редких изменений состояний среды. Получен вид многомерной асимптотической характеристической функции. Полученное асимптотическое распределение является многомодальным, так как имеет несколько локальных максимумов, что имеет принципиальное значение для применения результатов на практике. Доказано, что одномерные (маргинальные) стационарные распределения вероятностей числа занятых приборов каждого типа являются взвешенными суммами пуассоновских распределений. Проведен численный анализ области применимости полученной аппроксимации.

¹ Исследование выполнено за счет гранта Российского научного фонда № 24-21-00454 <https://rscf.ru/project/24-21-00454>.

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

³ Екатерина Владимировна Панкратова, к.ф.-м.н., с.н.с. (pankate@gmail.com).

Ключевые слова: марковски модулированные пуассоновские потоки, асимптотический анализ, предельно редкие изменения состояний цепи Маркова.

1. Введение

С повсеместным распространением беспроводных пользовательских устройств стала актуальна беспроводная передача данных различных приложений пользователей. Несмотря на то, что по прогнозам пропускная способность беспроводных сетей значительно вырастет в ближайшие годы, также вырастет и объем передаваемого трафика. Это обусловлено, с одной стороны, ростом технологии Интернета вещей (Internet of Things, IoT), с другой стороны, существенно увеличивается роль телекоммуникаций в профессиональной деятельности, обучении и обеспечении общественных отношений. Таким образом, в структуре информационных потоков увеличивается доля услуг, потребляющих значительные объемы ресурса передачи данных. Ввиду этого обеспечение эффективной передачи данных в сети остается актуальной задачей.

Многомодальные системы также в настоящее время набирают популярность с развитием многомодальных интерфейсов. Под модальностью при этом понимаются физически регистрируемые элементы коммуникации (человеко-машинной и/или межличностной), включающие как собственно передаваемую информацию (сообщение), так и информацию о самом индивиде (его состоянии; отношении к сообщению, к собеседнику, к коммуникации и пр.) [4].

Тенденция передачи данных по беспроводным каналам связи актуальна и для многомодальных информационных систем. Набор многомодальных данных и их размер могут быть различными в зависимости от задачи. Так, например, для речевой модальности в системе распознавания речи по аудиозаписи может быть достаточно 70–80 кбайт, тогда как для жестовой модальности в системе распознавания русского жестового языка объем одной записи модальности может составлять примерно 125 Мбайт [1].

Учитывая, что многомодальные потоки представляют собой интегрированные разнотипные потоки, включающие передачу голоса, текстовых данных и видео, для их описания логично применять непуассоновские модели. Так как на обслуживание различных информационных единиц затрачивается различное время в зависимости от формата, применяемых протоколов и других параметров, то в качестве моделей процессов в информационных системах используют системы массового обслуживания с разнотипными заявками, требующими разного времени обслуживания.

В данной работе предлагается модель многопоточной гетерогенной системы передачи данных по каналам различной интенсивности обслуживания в виде гетерогенной системы массового обслуживания.

Решение задач анализа немарковских многомерных моделей массового обслуживания с непуассоновскими входящими потоками, к сожалению, представлено лишь отдельными работами, в которых, как правило, рассматриваются модели узкого класса или специфической конфигурации [2, 3, 5, 10, 11]. Выработка же общих подходов и методов исследования немарковских моделей массового обслуживания в настоящее время является актуальной научной проблемой.

2. Математическая модель

2.1. ОПИСАНИЕ МОДЕЛИ

На вход системы поступают N пуассоновских потоков с интенсивностями, зависящими от состояния случайной среды. Случайная среда является цепью Маркова с непрерывным временем с числом состояний $k(t) = 1, 2, \dots, K$ и определяется матрицей инфинитезимальных характеристик $\mathbf{Q} = [q_{ij}]$, $i, j = 1, \dots, K$.

Таким образом, можно говорить о том, что входящие потоки являются марковски модулированными пуассоновскими потоками (Markov Modulated Poisson Process) [8, 9], которые имеют общий управляющий процесс $k(t)$ и определяются диагональными матрицами условных интенсивностей $\Lambda^{(l)}$, $l = 1, \dots, N$, с элементами $\lambda_k^{(l)}$ на главной диагонали.

Поступающее требование всегда застает свободный прибор и занимает его для обслуживания в течение случайного времени, распределенного по экспоненциальному закону с параметром μ_l , $l = 1, \dots, N$, соответственно. Таким образом, в системе находятся разнотипные требования, отличающиеся параметрами обслуживания, которые мы условно определим как разные блоки.

То есть рассматривается бесконечнолинейная система массового обслуживания с поступающими на вход ММРР-потоками разнотипных заявок и, соответственно, их разнотипным обслуживанием, отличающимся скоростью обслуживания. Используя символику Кендалла, будем обозначать такую систему массового обслуживания $\text{MMPP}^{(n)}|M_n|\infty$.

Ставится задача исследования N -мерного немарковского случайного процесса $\{i_1(t), \dots, i_N(t)\}$, описывающего число заявок в соответствующем блоке обслуживания.

2.2. СИСТЕМА УРАВНЕНИЙ КОЛМОГОРОВА

Для построения марковского случайного процесса воспользуемся методом дополнительной компоненты [5]. Введем случайный процесс $k(t)$ – состояние случайной среды, управляемое цепью Маркова, и будем рассматривать $(N + 1)$ -мерный марковский процесс $\{k(t), i_1(t), i_2(t), \dots, i_N(t)\}$. Определим совместное распределение вероятностей

$$(1) \quad P(k, i_1, i_2, \dots, i_N, t) = P\{k(t) = k, i_1(t) = i_1, i_2(t) = i_2, \dots, i_N(t) = i_N\}.$$

Для более компактной записи в дальнейшем будем использовать векторную запись. Для этого введем следующие обозначения: $\mathbf{i}(t) = [i_1(t), i_2(t), \dots, i_N(t)]$, $\mathbf{e}_1 = [1, 0, \dots, 0]$, $\mathbf{e}_2 = [0, 1, \dots, 0]$, ..., $\mathbf{e}_N = [0, 0, \dots, 1]$ – вектор-строки размерности $1 \times N$.

Используя формулу полной вероятности, для распределения вероятностей рассматриваемого случайного процесса составим систему равенств

$$(2) \quad P(k, \mathbf{i}, t + \Delta t) = P(k, \mathbf{i}, t) \left\{ 1 + \left[q_{kk} \sum_{l=1}^N (\lambda_k^{(l)} + i_l \mu_l) \Delta t \right] \right\} + \\ + \sum_{l=1}^N \lambda_k^{(l)} \Delta t P(k, \mathbf{i} - \mathbf{e}_l, t) + \sum_{l=1}^N (i_l + 1) \mu_l \Delta t P(k, \mathbf{i} + \mathbf{e}_l, t) + \\ + \sum_{v=1, v \neq k}^K q_{vk} \Delta t P(k, \mathbf{i}, t) + o(\Delta t).$$

Откуда получаем прямую систему дифференциальных уравнений Колмогорова:

$$(3) \quad \frac{\partial P(k, \mathbf{i}, t)}{\partial t} = -P(k, \mathbf{i}, t) \left[\sum_{l=1}^N (\lambda_k^{(l)} + i_l \mu_l) \right] + \sum_{l=1}^N \lambda_k^{(l)} P(k, \mathbf{i} - \mathbf{e}_l, t) + \\ + \sum_{l=1}^N (i_l + 1) \mu_l P(k, \mathbf{i} + \mathbf{e}_l, t) + \sum_{v=1}^K q_{vk} P(k, \mathbf{i}, t), \quad k = 1, \dots, K.$$

В стационарном режиме функционирования система уравнений (3) примет вид

$$(4) \quad 0 = -\Pi(k, \mathbf{i}) \left[\sum_{l=1}^N (\lambda_k^{(l)} + i_l \mu_l) \right] + \sum_{l=1}^N \lambda_k^{(l)} \Pi(k, \mathbf{i} - \mathbf{e}_l) + \\ + \sum_{l=1}^N (i_l + 1) \mu_l \Pi(k, \mathbf{i} + \mathbf{e}_l) + \sum_{v=1}^K q_{vk} \Pi(k, \mathbf{i}), \quad k = 1, \dots, K.$$

3. Вероятностные характеристики числа занятых приборов каждого типа

Введем в рассмотрение частичные характеристические функции вида

$$(5) \quad h(k, u_1, u_2, \dots, u_N) = H(k, \mathbf{u}) = \sum_{i_1} e^{ju_1 i_1} \sum_{i_2} e^{ju_2 i_2} \dots \sum_{i_N} e^{ju_N i_N} \Pi(k, \mathbf{i}),$$

$$\mathbf{u} = (u_1, u_2, \dots, u_N),$$

для которых система уравнений (4) переписется в виде

$$(6) \quad j \sum_{l=1}^N \mu_l (e^{-ju_l} - 1) \frac{\partial h(k, \mathbf{u})}{\partial u_l} = \sum_{l=1}^N \lambda_k^{(l)} (e^{ju_l} - 1) h(k, \mathbf{u}) + \\ + \sum_{v=1}^K q_{vk} h(v, \mathbf{u}), \quad k = 1, \dots, K,$$

с начальным условием

$$(7) \quad h(k, u_1, u_2, \dots, u_N) \Big|_{u_1 = \dots = u_N = 0} = H(k, \mathbf{0}) = r(k),$$

$r(k)$ – стационарное распределение вероятностей состояний управляющей цепи Маркова $k(t)$.

Здесь и далее принимаем, что $j = \sqrt{-1}$ – мнимая единица.

Запишем систему (6) в матричном виде:

$$(8) \quad j \sum_{l=1}^N \mu_l (e^{-j\mu_l} - 1) \frac{\partial \mathbf{h}(\mathbf{u})}{\partial u_l} = \mathbf{h}(\mathbf{u}) \left[\sum_{l=1}^N \Lambda^{(l)} (e^{j\mu_l} - 1) + \mathbf{Q} \right],$$

где введены вектор строки

$$(9) \quad \mathbf{h}(\mathbf{u}) = [h(1, \mathbf{u}), h(2, \mathbf{u}), \dots, h(K, \mathbf{u})],$$

$$\frac{\partial \mathbf{h}(\mathbf{u})}{\partial u_l} = \left[\frac{\partial h(1, \mathbf{u})}{\partial u_l}, \frac{\partial h(2, \mathbf{u})}{\partial u_l}, \dots, \frac{\partial h(K, \mathbf{u})}{\partial u_l} \right],$$

$$\frac{\partial \mathbf{h}(\mathbf{u})}{\partial u_l} = \left[\frac{\partial h(1, \mathbf{u})}{\partial u_l}, \frac{\partial h(2, \mathbf{u})}{\partial u_l}, \dots, \frac{\partial h(K, \mathbf{u})}{\partial u_l} \right].$$

Используя свойства характеристических функций, можно показать, что среднее число занятых приборов каждого типа определяется по формуле

$$(10) \quad m_1^{(l)} = M \{i_l\} = \frac{\mathbf{r} \Lambda^{(l)} \mathbf{e}}{\mu_l}, \quad l = 1, \dots, N,$$

где $\mathbf{r} = [r(1), r(2), \dots, r(K)]$ – вектор-строка стационарного распределения вероятностей состояний случайной среды, определяемая системой линейных уравнений

$$(11) \quad \begin{cases} \mathbf{r} \mathbf{Q} = \mathbf{0}, \\ \mathbf{r} \mathbf{e} = 1. \end{cases}$$

Здесь и далее \mathbf{e} – единичный вектор-столбец.

Вторые начальные моменты:

$$(12) \quad m_2^{(l)} = \left\{ \mathbf{r} \Lambda^{(l)} [\mu_l \mathbf{I} - \mathbf{Q}]^{-1} [2\Lambda^{(l)} - \mu_l \mathbf{I}] + \mathbf{r} \Lambda^{(l)} \right\} [2\mu_l \mathbf{I} - \mathbf{Q}]^{-1} \mathbf{e},$$

$$l = 1, \dots, N,$$

\mathbf{I} – матрица размерности $K \times K$, у которой главная диагональ состоит из единиц, а все остальные элементы 0.

Корреляционный момент:

$$(13) m^{(ls)} = M \{i_l \cdot i_s\} = \frac{1}{\mu_l + \mu_s} \left[\mathbf{m}_1^{(s)} \Lambda^{(l)} + \mathbf{m}_1^{(l)} \Lambda^{(s)} \right] \mathbf{e},$$

$$\mathbf{m}_1^{(l)} = \mathbf{r} \Lambda^{(l)} [\mu_l \mathbf{I} - \mathbf{Q}]^{-1}, \quad l, s = 1, \dots, N, \quad l \neq s.$$

4. Асимптотический анализ при условии предельно редких изменениях случайной среды

4.1. УСЛОВИЕ ПРЕДЕЛЬНО РЕДКИХ ИЗМЕНЕНИЙ СОСТОЯНИЙ СЛУЧАЙНОЙ СРЕДЫ

Далее будем искать решение системы (8) при условии предельно редких изменений состояний случайной среды.

Известно [5], что значения диагональных элементов матрицы инфинитезимальных характеристик определяют продолжительность времени пребывания потока в соответствующем состоянии.

Обозначив T_k – длительность пребывания цепи Маркова в k -м состоянии, имеем:

$$(14) T_k = -\frac{1}{q_{kk}}, \quad k = 1, \dots, K.$$

Если q_{kk} принимает достаточно малые значения, то продолжительность времени пребывания потока в k -м состоянии неограниченно возрастает, что оправдывает название рассматриваемого асимптотического условия.

Таким образом, условием предельно редких изменений состояний цепи Маркова будем называть равенства

$$(15) q_{vk} = \varepsilon q_{vk}^*, \quad \varepsilon \rightarrow 0.$$

Здесь ε – некоторый малый положительный параметр.

4.2. АСИМПТОТИЧЕСКИЙ АНАЛИЗ СИСТЕМЫ

Принимая во внимание (15), систему уравнений (8) перепишем в виде

$$(16) j \sum_{l=1}^N \mu_l (e^{-j\mu_l} - 1) \frac{\partial \mathbf{h}(\mathbf{u})}{\partial u_l} = \mathbf{h}(\mathbf{u}) \left[\sum_{l=1}^N \Lambda^{(l)} (e^{j\mu_l} - 1) + \varepsilon \mathbf{Q}^* \right].$$

Решение $\mathbf{h}(\mathbf{u})$ этой системы, зависящее от параметра ε , обозначим

$$(17) \mathbf{h}(\mathbf{u}) = \mathbf{f}(\mathbf{u}, \varepsilon) = \mathbf{f}(u_1, \dots, u_N, \varepsilon)$$

с начальными условиями

$$(18) \mathbf{f}(\mathbf{0}, \varepsilon) = \mathbf{h}(\mathbf{0}) = \mathbf{r}.$$

В силу теоремы о непрерывной зависимости решения системы дифференциальных уравнений от параметра существует предел

$$(19) \lim_{\varepsilon \rightarrow 0} \mathbf{f}(\mathbf{u}, \varepsilon) = \mathbf{f}_1(\mathbf{u}).$$

Тогда, выполнив системе (16) предельный переход при $\varepsilon \rightarrow 0$, получим матричное уравнение

$$(20) j \sum_{l=1}^N \mu_l (e^{-j\mu_l} - 1) \frac{\partial \mathbf{f}_1(\mathbf{u})}{\partial u_l} = \mathbf{f}_1(\mathbf{u}) \sum_{l=1}^N \Lambda^{(l)} (e^{j\mu_l} - 1).$$

Очевидно, что система (20) представляет собой систему уравнений в частных производных вида

$$(21) \sum_{l=1}^N j \mu_l (e^{-j\mu_l} - 1) \frac{\partial f_1(k, u_1, \dots, u_N)}{\partial u_l} = \\ = f_1(k, u_1, \dots, u_N) \sum_{l=1}^N \lambda_k^{(l)} (e^{j\mu_l} - 1), \quad k = 1, \dots, K.$$

Запишем соответствующую систему обыкновенных дифференциальных уравнений [7]:

$$(22) \frac{du_1}{j\mu_1(e^{-j\mu_1} - 1)} = \dots = \frac{du_N}{j\mu_N(e^{-j\mu_N} - 1)} = \frac{df_1(k, \mathbf{u})}{f_1(k, \mathbf{u}) \sum_{l=1}^N \lambda_k^{(l)} (e^{j\mu_l} - 1)},$$

$$k = 1, \dots, K.$$

Рассмотрим первые N интегралов вида

$$(23) \frac{du_1}{j\mu_1(e^{-j\mu_1} - 1)} = \frac{du_s}{j\mu_s(e^{-j\mu_s} - 1)}, \quad s = 2, \dots, N.$$

Нетрудно убедиться, что решение имеет вид

$$(24) \left(\frac{e^{j\mu_1} - 1}{C_s} \right)^{\frac{1}{\mu_1}} = (e^{j\mu_s} - 1)^{\frac{1}{\mu_s}}, \\ (e^{j\mu_1} - 1)^{\frac{\mu_s}{\mu_1}} = C_s (e^{j\mu_s} - 1), \quad s = 2, \dots, N.$$

Далее рассмотрим дифференциальное уравнение

$$(25) \quad \frac{du_1}{j\mu_1(e^{-ju_1} - 1)} = \frac{df_1(k, \mathbf{u})}{f_1(k, \mathbf{u}) \sum_{l=1}^N \lambda_k^{(l)} (e^{ju_l} - 1)}, \quad k = 1, \dots, K.$$

которое перепишем в виде

$$(26) \quad \frac{\sum_{l=1}^N \lambda_k^{(l)} (e^{ju_l} - 1) du_1}{j\mu_1(e^{-ju_1} - 1)} = \frac{df_1(k, \mathbf{u})}{f_1(k, \mathbf{u})}.$$

Далее, учитывая (24), получим

$$(27) \quad \frac{\lambda_k^{(1)} (e^{ju_1} - 1) + \sum_{l=2}^N \lambda_k^{(l)} \frac{(e^{ju_l} - 1)^{\mu_l}}{C_l} du_1}{j\mu_1 e^{-ju_1} (e^{ju_1} - 1)} = \frac{df_1(k, \mathbf{u})}{f_1(k, \mathbf{u})}.$$

Откуда получаем

$$(28) \quad \left(\frac{\lambda_k^{(1)}}{\mu_1} + \sum_{l=2}^N \lambda_k^{(l)} \frac{(e^{ju_l} - 1)^{\mu_l}}{\mu_l C_l} \right) d(e^{ju_1} - 1) = \frac{df_1(k, \mathbf{u})}{f_1(k, \mathbf{u})}.$$

Общее решение принимает вид

$$(29) \quad f_1(k, \mathbf{u}) = C \exp \left\{ \frac{\lambda_k^{(1)}}{\mu_1} (e^{ju_1} - 1) + \sum_{l=2}^N \lambda_k^{(l)} \frac{(e^{ju_l} - 1)^{\mu_l}}{\mu_l C_l \mu_1} \right\}.$$

Далее, подставляя (24), получаем

$$(30) \quad f_1(k, \mathbf{u}) = C(k) \exp \left\{ \sum_{l=1}^N \frac{\lambda_k^{(l)}}{\mu_l} (e^{ju_l} - 1) \right\}, \quad k = 1, \dots, K.$$

Константа $C(k)$ определяется начальными условиями (18):

$$(31) \quad f_1(k, \mathbf{0}) = C(k) = r(k), \quad k = 1, \dots, K.$$

Окончательно получаем

$$(32) \quad h(k, \mathbf{u}) = \mathbf{f}(k, \mathbf{u}, \varepsilon) \approx \mathbf{f}_1(k, \mathbf{u}) = r(k) \exp \left\{ \sum_{l=1}^N \frac{\lambda_k^{(l)}}{\mu_l} (e^{ju_l} - 1) \right\}.$$

Тогда, просуммировав (32) по всем $k = 1, \dots, K$, находим, что характеристическая функция совместного распределения вероятностей числа занятых приборов каждого типа имеет вид

$$(33) \quad h(\mathbf{u}) \approx \sum_{k=0}^N r(k) \exp \left\{ \sum_{l=1}^N \frac{\lambda_k^{(l)}}{\mu_l} (e^{ju_l} - 1) \right\}.$$

Нетрудно убедиться, что одномерные характеристические функции стационарного распределения вероятностей числа занятых приборов каждого типа для всех $s = 1, \dots, N$ имеют вид

$$(34) \quad h(u_s) = M \left\{ e^{ju_s i_s} \right\} \approx \sum_{k=0}^N r(k) \exp \left\{ \frac{\lambda_k^{(s)}}{\mu_s} (e^{ju_s} - 1) \right\} = \\ = \sum_{k=1}^K r(k) \sum_{i=0}^{\infty} \frac{\left(\frac{\lambda_k^{(s)}}{\mu_s} e^{ju_s} \right)^i}{i!} e^{-\frac{\lambda_k^{(s)}}{\mu_s}} = \sum_{i=0}^{\infty} (e^{ju_s})^i \sum_{k=1}^K r(k) \frac{\left(\frac{\lambda_k^{(s)}}{\mu_s} \right)^i}{i!} e^{-\frac{\lambda_k^{(s)}}{\mu_s}}.$$

Следовательно, маргинальные распределения вероятностей числа занятых приборов каждого типа являются взвешенной суммой пуассоновских распределений:

$$(35) \quad \lim_{\varepsilon \rightarrow 0} \Pi(i_s, \varepsilon) = \Pi(i_s) = \sum_{k=1}^K r(k) \frac{(\rho_k^{(s)})^{i_s}}{i_s!} e^{-\rho_k^{(s)}},$$

где

$$(36) \quad \rho_k^{(s)} = \lambda_k^{(s)} / \mu_s.$$

Аналогично можно показать, что двумерное распределение вероятностей определяется формулой

$$(37) \quad \Pi(i_s, i_l) = \sum_{k=1}^K r(k) \frac{(\rho_k^{(s)})^{i_s}}{i_s!} \frac{(\rho_k^{(l)})^{i_l}}{i_l!} e^{-\rho_k^{(s)}} e^{-\rho_k^{(l)}}, \quad l, s = 1, \dots, N.$$

На рис. 1–2 представлены графики одномерных распределений вероятностей при следующих параметрах системы: $N = 2$, $\mu_1 = 0,1$, $\mu_2 = 0,2$, а входящие потоки заданы матрицами условных интенсивностей вида

$$(38) \Lambda^{(1)} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 10 & 0 & 0 \\ 0 & 0 & 5 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad \Lambda^{(2)} = \begin{bmatrix} 20 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 6 & 0 \\ 0 & 0 & 0 & 20 \end{bmatrix},$$

и матрицей инфинитезимальных характеристик

$$(39) \mathbf{Q} = \varepsilon \cdot \begin{bmatrix} -6 & 3 & 2 & 1 \\ 0,4 & -1,5 & 0,1 & 1 \\ 1 & 2 & -6 & 3 \\ 0 & 1 & 6 & -7 \end{bmatrix}.$$



Рис. 1. Распределение вероятностей числа заявок первого типа при $\varepsilon = 0,1$

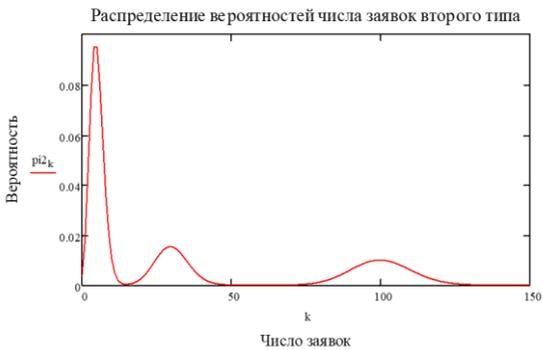


Рис. 2. Распределение вероятностей числа заявок второго типа при $\varepsilon = 0,1$

Следует заметить, что полученное асимптотическое распределение является многомодальным, так как имеет несколько локальных максимумов, что имеет принципиальное значение для применения результатов на практике.

5. Численный анализ

Поскольку результат (33) получен в условии предельно редких изменений состояний среды, необходимо установить, достаточно ли точна полученная аппроксимация, чтобы ее можно было применять на практике.

Из (33) согласно свойствам характеристических функций нетрудно получить асимптотические вероятностные характеристики.

Математическое ожидание числа заявок каждого типа:

$$(40) M_1 as^{(l)} = j \sum_{k=1}^K r(k) \cdot \frac{\lambda_k^{(l)}}{\mu_l}, \quad l = 1, \dots, N.$$

Вторые начальные моменты числа заявок каждого типа:

$$(41) M_2 as^{(l)} = \sum_{k=1}^K r(k) \left[\left(\frac{\lambda_k^{(l)}}{\mu_l} \right)^2 + \frac{\lambda_k^{(l)}}{\mu_l} \right], \quad l = 1, \dots, N.$$

Корреляционный момент:

$$(42) Mas\{i, i_s\} = \sum_{k=1}^K r(k) \frac{\lambda_k^{(1)}}{\mu_1} \frac{\lambda_k^{(2)}}{\mu_2}, \quad l, s = 1, \dots, N.$$

Для определения точности аппроксимации сравним ее асимптотические вероятностные значения с точными значениями (10), (12), (13).

Для анализа рассмотрим систему с двумя входящими потоками с параметрами (38)–(39) при различных значениях ε .

Так как асимптотические и точные средние определяются одинаковыми выражениями, то оценим относительную погрешность, сравнивая моменты второго порядка:

$$(43) \Delta_l = \frac{|m_2^{(l)} - M_2 as^{(l)}|}{m_2^{(l)}}.$$

В таблице 1 приведены значения относительных погрешностей между значениями точных и аппроксимационных начальных моментов второго порядка.

Таблица 1. Относительная погрешность аппроксимации

| Δ | $\varepsilon = 0,1$ | $\varepsilon = 0,01$ | $\varepsilon = 0,005$ |
|------------|---------------------|----------------------|-----------------------|
| $\Delta 1$ | 0,24 | 0,07 | 0,04 |
| $\Delta 2$ | 0,69 | 0,15 | 0,08 |

Очевидно, что при уменьшении параметра ε уменьшаются значения относительных погрешностей. При $\varepsilon \leq 0,005$ относительная погрешность составляет менее 10%.

6. Заключение

В настоящей статье представлена математическая модель мультипоточковой системы передачи данных в виде гетерогенной бесконечнолинейной системы массового обслуживания с несколькими входящими потоками разной модальности.

На практике обычно предоставляемые для использования резервы ресурсов ограничены, поэтому допущение, что количество каналов неограниченное, достаточно сильное, так как в этом случае в рассматриваемой системе нет потерь. Но полученные результаты, а именно асимптотическое распределение вероятностей и выражения для моментов, позволяют оценить оптимально необходимое количество ресурсов каждого блока для системы с ограниченным ресурсом, обеспечивающее заданную вероятность потерь.

В дальнейшем планируется реализовать имитационную модель, чтобы сравнить асимптотическое и эмпирическое распределения вероятностей исследуемых процессов, а также для решения задач оптимизации.

Литература

1. БАСОВ О.О., ПАКУЛОВА Е.А., САИТОВ И.А. *Методологические основы построения интеллектуальных инфоком-*

- муникационных систем. – Орёл: Академия ФСО России, 2020. – 272 с.
2. ВИШНЕВСКИЙ В.М., ДУДИН А.Н., КЛИМЕНОК В.И. *Стохастические системы с коррелированными потоками. Теория и применение в телекоммуникационных сетях.* – М.: Рекламно-издательский центр "ТЕХНОСФЕРА", 2018. – 564 с.
 3. ГОРБАТЕНКО А.Е. *Асимптотики произвольного порядка для системы $MAR|GI|_{\infty}$ в условии растущей интенсивности входящего потока* // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. – 2010. – №2(11). – С. 35–43.
 4. МАТВЕЕВ Ю.Н. *Технологии биометрической идентификации личности по голосу и другим модальностям* // Вестник Московского государственного технического университета им. Н.Э. Баумана. – 2012. – №3(3). – С. 5.
 5. НАЗАРОВ А.А., МОИСЕЕВА С.П. *Методы асимптотического анализа в теории массового обслуживания.* – Томск: Изд-во НТЛ, 2006. – 112 с.
 6. НАУМОВ В.А., САМУЙЛОВ К.Е. *О моделировании систем массового обслуживания с множественными ресурсами* // Вестник РУДН. Серия: Математика, информатика, физика. – 2014. – №3. – С. 60–64.
 7. ЭЛЬСГОЛЬЦ Л.Э. *Дифференциальные уравнения и вариационное исчисление.* – М.: Наука, 1969. – 424 с.
 8. LUCANTONI D.M. *New results on the single server queue with a batch Markovian arrival process* // Stochastic Models. – 1991. – Vol. 7. – P. 1–46.
 9. NEUTS M.F., He Q.-M. *Markov arrival process with marked transitions* // Stochastic Processes and Applications. – 1998. – Vol. 74. – P. 37–52.
 10. SINGH V.P. *Markovian queues with three heterogeneous servers* // AIEE Transactions. – 1971. – Vol. 3 (1). – P.45–48.
 11. SINGH V.P. *Two-server Markovian queues with balking: Heterogeneous vs. homogeneous servers* // Operations Research. – 1970. – No. 18(1). – P. 145–159.

ASYMPTOTIC ANALYSIS OF A MULTI-FLOW HETEROGENEOUS QS UNDER CONDITIONS OF EXTREMELY RARE STATE CHANGES MANAGER OF INPUT FLOWS MARKOV CHAINS

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

Ekaterina Pankratova, V.A. Trapeznikov Institute of Control Sciences of Russian Academy of Sciences, Moscow, PhD (pankatya86@gmail.com).

Abstract: Currently, multimodal systems are gaining popularity with the development of multimodal interfaces. Multimodal streams are integrated streams of different types, including the transmission of voice, text data and video, so it is logical to use non-Poisson models to describe them. As a mathematical model of a multimodal servicing system, a multi-threaded heterogeneous queuing system with flows changing their intensity depending on the states of the Markov random environment is considered. Incoming requests from various flows are serviced during an exponentially distributed random time with parameters determined by the type of flow. Expressions are obtained for finding the maximum values of the main probabilistic characteristics of the number of occupied devices of each type. Asymptotic research is carried out under the condition of extremely rare changes in the states of the environment. The form of the multidimensional asymptotic characteristic function is obtained. It is proven that one-dimensional (marginal) stationary probability distributions of the number of occupied devices of each type are weighted sums of Poisson distributions. A numerical analysis of the range of applicability of the obtained approximation was carried out.

Keywords: Markov modulated Poisson flows, asymptotic analysis, extremely rare changes in the states of a Markov chain.

УДК 519.872

ББК 22.18

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

Поступила в редакцию 17.06.2024.

Опубликована 30.11.2024.

ИССЛЕДОВАНИЕ ДВУМЕРНОГО МАРКИРОВАННОГО ММРР В ПРЕДЕЛЬНОМ УСЛОВИИ ВЫСОКОЙ ИНТЕНСИВНОСТИ¹

Пауль С. В.², Назаров А. А.³, Лапатин И. Л.⁴
(Национальный исследовательский Томский
государственный университет, Томск)

Рассматривается математическая модель потока разнородных данных в виде двумерного маркированного ММРР. Исследование таких моделей необходимо для анализа нагрузки на многомодальные системы. Многомодальные интерфейсы способны обрабатывать несколько естественных для человека способов ввода информации, каждый из которых требует определенных ресурсов для распознавания, обработки и передачи. Для проектирования таких систем необходимо строить оценки требуемых ресурсов. Эти оценки могут строиться на основании совместного распределения вероятностей количества событий каждого типа за определенный промежуток времени. В работе предлагается асимптотический подход оценки двумерного распределения вероятностей числа событий, наступивших в высокоинтенсивном маркированном марковски модулированном потоке за некоторое время. Предельное условие высокой интенсивности определяется ростом параметра интенсивности наступления событий в исследуемом потоке. Метод асимптотического анализа проводится в два этапа. На первом этапе находятся параметры, которые определяют асимптотические средние числа событий первого и второго типа, наступивших в высокоинтенсивном потоке. На втором этапе находятся параметры, определяющие асимптотические дисперсии и ковариацию числа событий первого и второго типов. Показано, что предельное распределение числа событий, наступивших в высокоинтенсивном маркированном ММРР, является двумерным гауссовским. Полученные формулы для нахождения распределения и его характеристик имеют достаточно простые выражения, неизвестные в которых находятся решением систем линейных уравнений.

Ключевые слова: маркированный ММРР, метод асимптотического анализа, многомодальные системы.

¹ Исследование выполнено за счет гранта Российского научного фонда № 24-21-00454, <https://rscf.ru/project/24-21-00454/>.

² Светлана Владимировна Пауль, д. ф.-м. н., доцент (paulsv82@mail.ru).

³ Анатолий Андреевич Назаров, д.т.н., профессор (nazarov.tsu@gmail.com).

⁴ Иван Леонидович Лапатин, к. ф.-м. н. (ilapatin@mail.ru).

1. Введение

Многомодальные системы [1, 4, 8], такие как картографические, медицинские, системы виртуальной реальности, робототехнические и управляемые системы, web-приложения, «умный дом», технологии интернета вещей, в настоящее время набирают стремительную популярность и становятся более сложными и многофункциональными. При проектировании многомодальных систем возникают проблемы эффективного взаимодействия человека с компьютером, решение которых позволяет разработать современные многомодальные интерфейсы, обеспечивающие привычные и естественные для пользователя способы взаимодействия. Они должны непрерывно интерпретировать поступающую информацию от различных визуальных, слуховых и тактильных каналов, реализовывать максимальную функциональность для достижения надежности в работе инфотелекоммуникационных систем.

Основным отличием многомодальных интерфейсов от профессиональных средств взаимодействия является применение способов коммуникации, характерных для общения между людьми. Задача разработки многомодальных интерфейсов возникла при появлении научной парадигмы окружающего интеллектуального пространства [9].

Для проектирования многомодальных интерфейсов необходимо прогнозировать нагрузку, которая на него поступает, чтобы обеспечить необходимое качество обработки запросов. Поскольку одновременно поступают запросы разных типов, которые требуют различных алгоритмов обработки и выделения разных ресурсов, необходимо исследовать характеристики их совместного поступления.

Современные потоки информации в многомодальных системах могут быть суперпозицией потоков речевых сообщений или звуковых сигналов, аудио-, видео-, мультимедиа информации, потоков интерактивных данных. Моделируя такие телекоммуникационные потоки с пакетным поступлением заявок или несколькими типами сообщений, иногда важно выделять (маркировать) необходимые входящие сообщения. Для модели-

рования таких неоднородных потоков используются маркированные потоки, самым общим из которых принято считать ВММАР (Batch Marked Markov Arrival Process) [2]. В случае моделирования коррелированных неоднородных потоков часто используют ММАР (Marked Markov Arrival Process) [11, 13]. ММАР являются обобщением МАР [6, 7, 12] на случай гетерогенных сообщений (разнородных по типу передаваемых данных). Такие потоки часто используются как входящие в системах массового обслуживания [10, 14, 15, 20, 21].

В работе [5] представлены два эквивалентных способа задания МАР: классический способ задания МАР, согласно определению Д. М. Лукантони и М. Ф. Ньютса, и эквивалентный способ задания МАР-потока, который предложил Д. Кокс, которым и воспользуемся в данной работе. Одним из наиболее распространенных частных случаев МАР является марковски модулированный пуассоновский поток событий (ММРР) [3, 16, 17, 18, 19].

2. Математическая модель и постановка задачи

Рассмотрим один из классов маркированных МАР – маркированный ММРР, заданный генератором $\mathbf{Q}_{dop} = [q_{vm}]$ инфинитезимальных характеристик цепи Маркова $m(t) = 0, 1, \dots, M$ с непрерывным временем, управляющей потоком; диагональными матрицами $\Lambda_{dop}^{(1)}$ и $\Lambda_{dop}^{(2)}$ условных интенсивностей $\lambda_m^{(1)}$ и $\lambda_m^{(2)}$ наступления события в m -м состоянии двумерного маркированного ММРР, когда его управляющая цепь $m(t)$ принимает значение $m = 0, 1, \dots, M$ для первого и второго типа событий потока соответственно. Значения управляющего процесса $m(t)$ также будем называть состояниями исследуемого двумерного маркированного ММРР.

Пусть в некоторый момент времени t_m управляющая цепь Маркова перейдет в состояние v или, что то же самое, двумерный маркированный ММРР сменит свое состояние на v . В этом состоянии поток будет находиться до момента t_{m+1} . Длина $(t_{m+1} - t_m)$ интервала постоянства распределена по экспоненци-

альному закону с параметром $-q_{vm}$. В течение времени пребывания потока в состоянии v наступают события первого и второго типов с интенсивностями $\lambda_m^{(1)}$ и $\lambda_m^{(2)}$ соответственно. В момент времени t_{m+1} цепь Маркова (двумерный маркированный ММРР) перейдет в некоторое состояние m . Далее процедура повторяется для состояния m потока.

Обозначим распределение вероятностей

(1) $P_m(n_1, n_2, t) = P\{n_1(t) = n_1, n_2(t) = n_2, m(t) = m\}$, $m = 0, 1, \dots, M$, где $n_1(t)$ и $n_2(t)$ – число событий первого и второго типов соответственно, наступивших в двумерном маркированном ММРР за время t .

Трехмерный случайный процесс $\{n_1(t), n_2(t), m(t)\}$ является марковским, что позволяет для распределения вероятностей (1) получить систему дифференциальных уравнений Колмогорова:

$$(2) \quad \frac{\partial P_m(n_1, n_2, t)}{\partial t} = -(\lambda_m^{(1)} + \lambda_m^{(2)})P_m(n_1, n_2, t) + \lambda_m^{(1)}P_m(n_1 - 1, n_2, t) + \lambda_m^{(2)}P_m(n_1, n_2 - 1, t) + \sum_{v=0}^M P_v(n_1, n_2, t)q_{vm}.$$

Введем частичные характеристические функции

$$H_m(u_1, u_2, t) = \sum_{n_1=0}^{\infty} e^{ju_1 n_1} \sum_{n_2=0}^{\infty} e^{ju_2 n_2} P_m(n_1, n_2, t), \quad m = 0, 1, \dots, M.$$

Систему (2) перепишем в следующем виде:

$$(3) \quad \frac{\partial H_m(u_1, u_2, t)}{\partial t} = \left[\lambda_m^{(1)}(e^{ju_1} - 1) + \lambda_m^{(2)}(e^{ju_2} - 1) \right] H_m(u_1, u_2, t) + \sum_{v=0}^M H_v(u_1, u_2, t)q_{vm}.$$

Обозначим вектор-строку

$$\mathbf{H}(u_1, u_2, t) = \{H_1(u_1, u_2, t), H_2(u_1, u_2, t), \dots, H_M(u_1, u_2, t)\},$$

тогда систему уравнений (3) перепишем в матричном виде

$$(4) \quad \frac{\partial \mathbf{H}(u_1, u_2, t)}{\partial t} = \mathbf{H}(u_1, u_2, t) \left\{ \mathbf{Q}_{dop} + (e^{ju_1} - 1)\mathbf{\Lambda}_{dop}^{(1)} + (e^{ju_2} - 1)\mathbf{\Lambda}_{dop}^{(2)} \right\}.$$

Обозначим

$$(5) \quad \mathbf{H}(0, 0, t) = \mathbf{r},$$

где \mathbf{r} – вектор-строка стационарного распределения вероятностей значений марковского процесса $m(t)$. Из (4) при $u_1 = 0$ и $u_2 = 0$ получим систему уравнений, которая совместно с условием нормировки однозначно определяет вектор \mathbf{r} :

$$(6) \quad \mathbf{r}\mathbf{Q} = 0, \\ \mathbf{r}\mathbf{e} = 1.$$

Здесь \mathbf{e} – единичный вектор-столбец. Получим задачу Коши в виде

$$(7) \quad \frac{\partial \mathbf{H}(u_1, u_2, t)}{\partial t} = \mathbf{H}(u_1, u_2, t) \left\{ \mathbf{Q}_{dop} + (e^{ju_1} - 1) \mathbf{\Lambda}_{dop}^{(1)} + (e^{ju_2} - 1) \mathbf{\Lambda}_{dop}^{(2)} \right\}, \\ \mathbf{H}(0, 0, t) = \mathbf{r},$$

решение которой, применяя метод матричной экспоненты, запишем в виде

$$(8) \quad \mathbf{H}(u_1, u_2, t) = \mathbf{r} \exp \left\{ \left[\mathbf{Q}_{dop} + (e^{ju_1} - 1) \mathbf{\Lambda}_{dop}^{(1)} + (e^{ju_2} - 1) \mathbf{\Lambda}_{dop}^{(2)} \right] t \right\}.$$

Нахождение значений двумерной характеристической функции $\mathbf{H}(u_1, u_2, t)$ при больших значениях интенсивностей входящего потока требует значительных затрат машинного времени, которое не всегда приводит к определенным результатам. Поэтому получим предельное при высокой интенсивности потока выражение для распределения вероятностей значений числа событий, наступивших в двумерном маркированном ММРР.

Умножив матричное уравнение (7) на единичный вектор-столбец \mathbf{e} , получим систему

$$(9) \quad \frac{\partial \mathbf{H}(u_1, u_2, t)}{\partial t} = \mathbf{H}(u_1, u_2, t) \left\{ \mathbf{Q}_{dop} + (e^{ju_1} - 1) \mathbf{\Lambda}_{dop}^{(1)} + (e^{ju_2} - 1) \mathbf{\Lambda}_{dop}^{(2)} \right\}, \\ \frac{\partial \mathbf{H}(u_1, u_2, t)}{\partial t} \mathbf{e} = \mathbf{H}(u_1, u_2, t) \left\{ (e^{ju_1} - 1) \mathbf{\Lambda}_{dop}^{(1)} + (e^{ju_2} - 1) \mathbf{\Lambda}_{dop}^{(2)} \right\} \mathbf{e},$$

которая будет основной в дальнейших исследованиях.

Ставится задача нахождения распределения вероятностей числа событий первого и второго типов, наступивших в двумерном маркированном ММРР за время t .

3. Асимптотический анализ маркированного ММРР в условии большой интенсивности

Высокая интенсивность исследуемого потока обеспечивается за счет представления матриц интенсивностей $\Lambda_{dop}^{(l)} = \Lambda^{(l)}N$, $l = 1, 2$, $Q_{dop} = QN$, где $\Lambda^{(l)}$ Q – фиксированные матрицы, N – неограниченно возрастающий параметр. Величину N будем называть параметром высокой интенсивности двумерного маркированного ММРР, а предельное условие $N \rightarrow \infty$ будем называть условием высокой интенсивности потока [5]. Заданный в таком предельном условии двумерный ММРР будем называть высокоинтенсивным. Проведем анализ высокоинтенсивного двумерного маркированного ММРР, а именно, получим предельное распределение вероятностей числа событий, наступивших в исследуемом потоке за время t .

Реализация метода асимптотического анализа проводится в два этапа. На первом этапе находятся параметры $\kappa_1^{(1)}$ и $\kappa_1^{(2)}$, которые определяют асимптотические средние $\kappa_1^{(1)}Nt$ и $\kappa_1^{(2)}Nt$ числа событий первого и второго типа, наступивших в двумерном маркированном ММРР за время t .

При реализации второго этапа асимптотического анализа находятся параметры $\kappa_2^{(1)}$, $\kappa_2^{(2)}$ и κ_{12} определяющих асимптотические дисперсии числа событий $\kappa_2^{(1)}Nt$ и $\kappa_2^{(2)}Nt$ первого и второго типа, наступивших в двумерном маркированном ММРР за время t , и ковариацию $\kappa_{12}Nt$ числа событий первого и второго типов. Показано, что предельное распределение числа событий, наступивших в высокоинтенсивном исследуемом потоке, является двумерным гауссовским.

Решим систему (9) методом асимптотического анализа в предельном условии высокой интенсивности потока.

3.1. АСИМПТОТИКА ПЕРВОГО ПОРЯДКА

Введем обозначение $\frac{1}{N} = \varepsilon$, тогда систему (9) запишем в виде

$$(10) \quad \frac{1}{N} \frac{\partial \mathbf{H}(u_1, u_2, t)}{\partial t} = \mathbf{H}(u_1, u_2, t) \left\{ \mathbf{Q} + (e^{ju_1} - 1) \mathbf{\Lambda}^{(1)} + (e^{ju_2} - 1) \mathbf{\Lambda}^{(2)} \right\},$$

$$\frac{1}{N} \frac{\partial \mathbf{H}(u_1, u_2, t)}{\partial t} \mathbf{e} = \mathbf{H}(u_1, u_2, t) \left\{ (e^{ju_1} - 1) \mathbf{\Lambda}^{(1)} + (e^{ju_2} - 1) \mathbf{\Lambda}^{(2)} \right\} \mathbf{e},$$

и в системе (10) сделаем следующие замены

$$(11) \quad u_1 = \frac{w_1}{N} = \varepsilon w_1, \quad u_2 = \frac{w_2}{N} = \varepsilon w_2, \quad \mathbf{H}(u_1, u_2, t) = \mathbf{F}_1(w_1, w_2, t, \varepsilon).$$

Получим систему

$$(12) \quad \varepsilon \frac{\partial \mathbf{F}_1(w_1, w_2, t, \varepsilon)}{\partial t} = \mathbf{F}_1(w_1, w_2, t, \varepsilon) \left\{ \mathbf{Q} + (e^{j\varepsilon w_1} - 1) \mathbf{\Lambda}^{(1)} + (e^{j\varepsilon w_2} - 1) \mathbf{\Lambda}^{(2)} \right\},$$

$$\varepsilon \frac{\partial \mathbf{F}_1(w_1, w_2, t, \varepsilon)}{\partial t} \mathbf{e} = \mathbf{F}_1(w_1, w_2, t, \varepsilon) \left\{ (e^{j\varepsilon w_1} - 1) \mathbf{\Lambda}^{(1)} + (e^{j\varepsilon w_2} - 1) \mathbf{\Lambda}^{(2)} \right\} \mathbf{e}.$$

Решение $\mathbf{F}_1(w_1, w_2, t, \varepsilon)$ будем искать в классе вектор-функций, для которых существует конечный предел при $\varepsilon \rightarrow 0$ $\lim_{\varepsilon \rightarrow 0} \mathbf{F}_1(w_1, w_2, t, \varepsilon) = \mathbf{F}_1(w_1, w_2, t)$ и предел производной по t

$$\lim_{\varepsilon \rightarrow 0} \frac{\partial \mathbf{F}_1(w_1, w_2, t, \varepsilon)}{\partial t} = \frac{\partial \mathbf{F}_1(w_1, w_2, t)}{\partial t}.$$

Теорема 1. *Асимптотическое решение системы (12) $\lim_{\varepsilon \rightarrow 0} \mathbf{F}_1(w_1, w_2, t, \varepsilon) = \mathbf{F}_1(w_1, w_2, t)$ в предельном условии высокой интенсивности потока имеет вид*

$$(13) \quad \mathbf{F}_1(w_1, w_2, t) = \mathbf{r} \exp \left\{ \left(jw_1 \kappa_1^{(1)} + jw_2 \kappa_1^{(2)} \right) t \right\},$$

где вектор \mathbf{r} является решением системы уравнений (6), а величины $\kappa_1^{(1)}$ и $\kappa_1^{(2)}$ определяются равенствами

$$(14) \quad \kappa_1^{(1)} = \mathbf{r} \mathbf{\Lambda}_1^{(1)} \mathbf{e}, \quad \kappa_1^{(2)} = \mathbf{r} \mathbf{\Lambda}_1^{(2)} \mathbf{e}.$$

Доказательство. В (12) разложим экспоненты в ряд Тейлора, сгруппируем слагаемые порядка малости не выше ε , разделим на ε и устремим $\varepsilon \rightarrow 0$. Получим систему для предельных характеристических функций

$$(15) \quad \mathbf{F}_1(w_1, w_2, t) \mathbf{Q} = 0,$$

$$\frac{\partial \mathbf{F}_1(w_1, w_2, t)}{\partial t} \mathbf{e} = j \mathbf{F}_1(w_1, w_2, t) \left\{ w_1 \mathbf{\Lambda}^{(1)} + w_2 \mathbf{\Lambda}^{(2)} \right\} \mathbf{e}.$$

Решение системы (15) будем искать в виде

$$(16) \mathbf{F}_1(w_1, w_2, t) = \mathbf{r}\Phi(w_1, w_2, t).$$

Получим систему

$$(17) \mathbf{r}\mathbf{Q} = 0,$$

$$\frac{\partial \Phi_1(w_1, w_2, t) / \partial t}{\Phi_1(w_1, w_2, t)} = j \left\{ w_1 \mathbf{r}\Lambda^{(1)} + w_2 \mathbf{r}\Lambda^{(2)} \right\} \mathbf{e}.$$

Так как отношение $\frac{\partial \Phi_1(w_1, w_2, t) / \partial t}{\Phi_1(w_1, w_2, t)}$ не зависит от t , то

функция $\Phi(w_1, w_2, t)$ имеет вид

$$(18) \Phi(w_1, w_2, t) = \exp \left\{ \left(jw_1 \kappa_1^{(1)} + jw_2 \kappa_1^{(2)} \right) t \right\},$$

где ввели обозначения

$$(19) \kappa_1^{(1)} = \mathbf{r}\Lambda_1^{(1)} \mathbf{e}, \quad \kappa_1^{(2)} = \mathbf{r}\Lambda_1^{(2)} \mathbf{e}.$$

Подставляя полученное решение (18) в (16), доказано, что выполняется равенство (13). \square

В силу замены (11) $\mathbf{H}(u_1, u_2, t) = \mathbf{F}_1(w_1, w_2, t, \varepsilon)$ и найденного в (16) и (18) предельного значения $\mathbf{F}_1(w_1, w_2, t)$ можно записать приближенное равенство для $\mathbf{H}(u_1, u_2, t)$:

$$\mathbf{H}(u_1, u_2, t) \approx \mathbf{r} \exp \left\{ ju_1 \kappa_1^{(1)} Nt + ju_2 \kappa_1^{(2)} Nt \right\}.$$

Обозначим предельную характеристическую функцию $h(u_1, u_2, t) = \mathbf{H}(u_1, u_2, t) \mathbf{e}$ числа событий первого и второго типа, наступивших в двумерном маркированном ММРР за время t

$$h(u_1, u_2, t) = \exp \left\{ ju_1 \kappa_1^{(1)} Nt + ju_2 \kappa_1^{(2)} Nt \right\}.$$

Доказанная теорема 1 определяет лишь асимптотические средние значения $\kappa_1^{(1)} Nt$ и $\kappa_1^{(2)} Nt$ первого и второго типа событий, наступивших в двумерном маркированном ММРР за время t в условии высокой интенсивности наступления событий.

Для построения распределения вероятностей числа событий, наступивших в исследуемом потоке, построим асимптотику второго порядка.

3.2. АСИМПТОТИКА ВТОРОГО ПОРЯДКА

На втором этапе реализации метода асимптотического анализа учитывая полученный вид характеристической функции $h(u_1, u_2, t)$, в системе (10) выполним замену

$$(20) \mathbf{H}(u_1, u_2, t) = e^{(ju_1 N \kappa_1^{(1)} + ju_2 N \kappa_1^{(2)})t} \mathbf{H}_2(u_1, u_2, t).$$

Получим систему уравнений

$$(21) \begin{aligned} & \frac{1}{N} \frac{\partial \mathbf{H}_2(u_1, u_2, t)}{\partial t} + (ju_1 \kappa_1^{(1)} + ju_2 \kappa_1^{(2)}) \mathbf{H}_2(u_1, u_2, t) = \\ & = \mathbf{H}_2(u_1, u_2, t) \left\{ \mathbf{Q} + (e^{ju_1} - 1) \mathbf{\Lambda}^{(1)} + (e^{ju_2} - 1) \mathbf{\Lambda}^{(2)} \right\}, \\ & \frac{1}{N} \frac{\partial \mathbf{H}_2(u_1, u_2, t)}{\partial t} \mathbf{e} + (ju_1 \kappa_1^{(1)} + ju_2 \kappa_1^{(2)}) \mathbf{H}_2(u_1, u_2, t) \mathbf{e} = \\ & = \mathbf{H}_2(u_1, u_2, t) \left\{ (e^{ju_1} - 1) \mathbf{\Lambda}^{(1)} + (e^{ju_2} - 1) \mathbf{\Lambda}^{(2)} \right\} \mathbf{e}. \end{aligned}$$

Обозначим $\frac{1}{N} = \varepsilon^2$, в системе (21) выполним замены

$$(22) u_1 = \frac{w_1}{\sqrt{N}} = \varepsilon w_1, \quad u_2 = \frac{w_2}{\sqrt{N}} = \varepsilon w_2,$$

$$\mathbf{H}_2(u_1, u_2, t) = \mathbf{F}_2(w_1, w_2, t, \varepsilon).$$

Получим систему для характеристических функций

$$(23) \begin{aligned} & \varepsilon^2 \frac{\partial \mathbf{F}_2(w_1, w_2, t, \varepsilon)}{\partial t} + (j\varepsilon w_1 \kappa_1^{(1)} + j\varepsilon w_2 \kappa_1^{(2)}) \mathbf{F}_2(w_1, w_2, t, \varepsilon) = \\ & = \mathbf{F}_2(w_1, w_2, t, \varepsilon) \left\{ \mathbf{Q} + (e^{j\varepsilon w_1} - 1) \mathbf{\Lambda}^{(1)} + (e^{j\varepsilon w_2} - 1) \mathbf{\Lambda}^{(2)} \right\}, \\ & \varepsilon^2 \frac{\partial \mathbf{F}_2(w_1, w_2, t, \varepsilon)}{\partial t} \mathbf{e} + (j\varepsilon w_1 \kappa_1^{(1)} + j\varepsilon w_2 \kappa_1^{(2)}) \mathbf{F}_2(w_1, w_2, t, \varepsilon) \mathbf{e} = \\ & = \mathbf{F}_2(w_1, w_2, t, \varepsilon) \left\{ (e^{j\varepsilon w_1} - 1) \mathbf{\Lambda}^{(1)} + (e^{j\varepsilon w_2} - 1) \mathbf{\Lambda}^{(2)} \right\} \mathbf{e}. \end{aligned}$$

Решение $\mathbf{F}_2(w_1, w_2, t, \varepsilon)$ будем искать в классе вектор-функций, для которых существует конечный предел при $\varepsilon \rightarrow 0$ $\lim_{\varepsilon \rightarrow 0} \mathbf{F}_2(w_1, w_2, t, \varepsilon) = \mathbf{F}_2(w_1, w_2, t)$ и предел производной по t

$$\lim_{\varepsilon \rightarrow 0} \frac{\partial \mathbf{F}_2(w_1, w_2, t, \varepsilon)}{\partial t} = \frac{\partial \mathbf{F}_2(w_1, w_2, t)}{\partial t}.$$

Теорема 2. Асимптотическое решение системы (23) $\lim_{\varepsilon \rightarrow 0} \mathbf{F}_2(w_1, w_2, t, \varepsilon) = \mathbf{F}_2(w_1, w_2, t)$ в предельном условии высокой интенсивности потока имеет вид

$$(24) \mathbf{F}_1(w_1, w_2, t) = \mathbf{r} \exp \left\{ \frac{(jw_1)^2}{2} \kappa_2^{(1)} t + \frac{(jw_2)^2}{2} \kappa_2^{(2)} t + jw_1 \cdot jw_2 \cdot \kappa_{12} t \right\},$$

где вектор \mathbf{r} является решением системы уравнений (6), а величины $\kappa_2^{(1)}$, $\kappa_2^{(2)}$ и κ_{12} определяются равенствами

$$(25) \begin{aligned} \kappa_2^{(1)} &= 2\mathbf{g}_1 \mathbf{\Lambda}^{(1)} \mathbf{e} + \kappa_1^{(1)}, \\ \kappa_2^{(2)} &= 2\mathbf{g}_2 \mathbf{\Lambda}^{(2)} \mathbf{e} + \kappa_1^{(2)}, \\ \kappa_{12} &= \mathbf{g}_1 \mathbf{\Lambda}^{(2)} \mathbf{e} + \mathbf{g}_2 \mathbf{\Lambda}^{(1)} \mathbf{e}. \end{aligned}$$

Здесь вектор-строки \mathbf{g}_1 и \mathbf{g}_2 являются решениями неоднородных систем уравнений

$$(26) \begin{aligned} \mathbf{g}_1 \mathbf{Q} &= \mathbf{r} \left(\kappa_1^{(1)} \mathbf{I} - \mathbf{\Lambda}^{(1)} \right), \\ \mathbf{g}_1 \mathbf{e} &= 0. \end{aligned}$$

$$(27) \begin{aligned} \mathbf{g}_2 \mathbf{Q} &= \mathbf{r} \left(\kappa_1^{(2)} \mathbf{I} - \mathbf{\Lambda}^{(2)} \right), \\ \mathbf{g}_2 \mathbf{e} &= 0. \end{aligned}$$

Доказательство. В систему (23) подставим следующее разложение

$$(28) \mathbf{F}_2(w_1, w_2, t, \varepsilon) = \Phi_2(w_1, w_2, t) \{ \mathbf{r} + j\varepsilon w_1 \mathbf{f}_1 + j\varepsilon w_2 \mathbf{f}_2 \} + O(\varepsilon^2),$$

и, разложив экспоненту в ряд Тейлора, учитывая (19), получим

$$(29) \left(w_1 \kappa_1^{(1)} + w_2 \kappa_1^{(2)} \right) \mathbf{r} = w_1 \mathbf{r} \mathbf{\Lambda}^{(1)} + w_2 \mathbf{r} \mathbf{\Lambda}^{(2)} + w_1 \mathbf{f}_1 \mathbf{Q} + w_2 \mathbf{f}_2 \mathbf{Q} + \frac{O(\varepsilon^2)}{\varepsilon},$$

$$\begin{aligned} \frac{\partial \Phi_2(w_1, w_2, t)}{\partial t} &= \frac{(jw_1)^2}{2} \left(2\mathbf{f}_1 \mathbf{\Lambda}^{(1)} \mathbf{e} - 2\kappa_1^{(1)} \mathbf{f}_1 \mathbf{e} + \kappa_1^{(1)} \right) + \\ &+ \frac{(jw_2)^2}{2} \left(2\mathbf{f}_2 \mathbf{\Lambda}^{(2)} \mathbf{e} - 2\kappa_1^{(2)} \mathbf{f}_2 \mathbf{e} + \kappa_1^{(2)} \right) + \\ &+ jw_1 jw_2 \left(\mathbf{f}_1 \mathbf{\Lambda}^{(2)} \mathbf{e} + \mathbf{f}_2 \mathbf{\Lambda}^{(1)} \mathbf{e} - \kappa_1^{(1)} \mathbf{f}_2 \mathbf{e} - \kappa_1^{(2)} \mathbf{f}_1 \mathbf{e} \right) + \frac{O(\varepsilon^3)}{\varepsilon^2}. \end{aligned}$$

Перейдем к пределу при $\varepsilon \rightarrow 0$ в последней системе (29), получим

$$(30) jw_1 \left(\mathbf{r} \left(\mathbf{\Lambda}^{(1)} - \kappa_1^{(1)} \right) + \mathbf{f}_1 \mathbf{Q} \right) + jw_2 \left(\mathbf{r} \left(\mathbf{\Lambda}^{(2)} - \kappa_1^{(2)} \right) + \mathbf{f}_2 \mathbf{Q} \right) = 0,$$

$$\begin{aligned} \frac{\partial \Phi_2(w_1, w_2, t) / \partial t}{\Phi_2(w_1, w_2, t)} &= \frac{(jw_1)^2}{2} (2\mathbf{f}_1 \Lambda^{(1)} \mathbf{e} - 2\kappa_1^{(1)} \mathbf{f}_1 \mathbf{e} + \kappa_1^{(1)}) + \\ &+ \frac{(jw_2)^2}{2} (2\mathbf{f}_2 \Lambda^{(2)} \mathbf{e} - 2\kappa_1^{(2)} \mathbf{f}_2 \mathbf{e} + \kappa_1^{(2)}) + \\ &+ jw_1 jw_2 (\mathbf{f}_1 \Lambda^{(2)} \mathbf{e} + \mathbf{f}_2 \Lambda^{(1)} \mathbf{e} - \kappa_1^{(1)} \mathbf{f}_2 \mathbf{e} - \kappa_1^{(2)} \mathbf{f}_1 \mathbf{e}). \end{aligned}$$

Из первого уравнения системы (30) получим два неоднородных матричных уравнения, определяющих векторы \mathbf{f}_1 и \mathbf{f}_2

$$(31) \quad \mathbf{f}_1 \mathbf{Q} = \mathbf{r} (\kappa_1^{(1)} \mathbf{I} - \Lambda^{(1)}),$$

$$(32) \quad \mathbf{f}_2 \mathbf{Q} = \mathbf{r} (\kappa_1^{(2)} \mathbf{I} - \Lambda^{(2)}).$$

Системы (31) и (32) являются неоднородными системами линейных алгебраических уравнений для \mathbf{f}_1 и \mathbf{f}_2 . Эти системы совместны и имеют множество решений, поскольку определитель матрицы коэффициентов системы равен 0, а ранг расширенной матрицы равен рангу матрицы коэффициентов.

Система $\mathbf{rQ} = 0$ является однородной системой для неоднородных систем (31) и (32). Тогда решения систем (31) и (32) можно записать в виде

$$(33) \quad \mathbf{f}_1 = C\mathbf{r} + \mathbf{g}_1,$$

$$(34) \quad \mathbf{f}_2 = C\mathbf{r} + \mathbf{g}_2.$$

Здесь C – константа, вектор вероятностей \mathbf{r} определен выше, а векторы \mathbf{g}_1 и \mathbf{g}_2 являются частными решениями неоднородных систем (31) и (32), удовлетворяющих дополнительным условиям $\mathbf{g}_1 \mathbf{e} = 0$, $\mathbf{g}_2 \mathbf{e} = 0$. Подставим решения (33) и (34) в систему (30):

$$(35) \quad \mathbf{g}_1 \mathbf{Q} = \mathbf{r} (\kappa_1^{(1)} \mathbf{I} - \Lambda^{(1)}),$$

$$\mathbf{g}_2 \mathbf{Q} = \mathbf{r} (\kappa_1^{(2)} \mathbf{I} - \Lambda^{(2)}),$$

$$\begin{aligned} \frac{\partial \Phi_2(w_1, w_2, t) / \partial t}{\Phi_2(w_1, w_2, t)} &= \frac{(jw_1)^2}{2} (2\mathbf{g}_1 \Lambda^{(1)} \mathbf{e} + \kappa_1^{(1)}) + \\ &+ \frac{(jw_2)^2}{2} (2\mathbf{g}_2 \Lambda^{(2)} \mathbf{e} + \kappa_1^{(2)}) + jw_1 jw_2 (\mathbf{g}_1 \Lambda^{(2)} \mathbf{e} + \mathbf{g}_2 \Lambda^{(1)} \mathbf{e}). \end{aligned}$$

Так как отношение $\frac{\partial \Phi_2(w_1, w_2, t) / \partial t}{\Phi_2(w_1, w_2, t)}$ не зависит от t , то скалярную функцию $\Phi_2(w_1, w_2, t)$ определим в виде

$$(36) \quad \Phi_2(w_1, w_2, t) = \exp \left\{ \frac{(jw_1)^2}{2} \kappa_2^{(1)} t + \frac{(jw_2)^2}{2} \kappa_2^{(2)} t + jw_1 \cdot jw_2 \cdot \kappa_{12} t \right\},$$

Тогда

$$\frac{\partial \Phi_2(w_1, w_2, t) / \partial t}{\Phi_2(w_1, w_2, t)} = \frac{(jw_1)^2}{2} \kappa_2^{(1)} + \frac{(jw_2)^2}{2} \kappa_2^{(2)} + jw_1 \cdot jw_2 \cdot \kappa_{12}.$$

Тогда систему (35) можно переписать в виде совокупности систем

$$(37) \quad \mathbf{g}_1 \mathbf{Q} = \mathbf{r} \left(\kappa_1^{(1)} \mathbf{I} - \mathbf{\Lambda}^{(1)} \right),$$

$$\mathbf{g}_1 \mathbf{e} = 0.$$

$$(38) \quad \mathbf{g}_2 \mathbf{Q} = \mathbf{r} \left(\kappa_1^{(2)} \mathbf{I} - \mathbf{\Lambda}^{(2)} \right),$$

$$\mathbf{g}_2 \mathbf{e} = 0,$$

где

$$(39) \quad \kappa_2^{(1)} = 2\mathbf{g}_1 \mathbf{\Lambda}^{(1)} \mathbf{e} + \kappa_1^{(1)},$$

$$\kappa_2^{(2)} = 2\mathbf{g}_2 \mathbf{\Lambda}^{(2)} \mathbf{e} + \kappa_1^{(2)},$$

$$\kappa_{12} = \mathbf{g}_1 \mathbf{\Lambda}^{(2)} \mathbf{e} + \mathbf{g}_2 \mathbf{\Lambda}^{(1)} \mathbf{e}.$$

Решив системы (37) и (38), найдем векторы \mathbf{g}_1 и \mathbf{g}_2 , которые подставим в выражения (39) для $\kappa_2^{(1)}$, $\kappa_2^{(2)}$ и κ_{12} . Полученные выражения полностью определяют предельную функцию $\Phi(w_1, w_2, t)$ (36), которая полностью определяет предельную функцию (24), учитывая разложение (28). \square

Возвращаясь к вектор-функции $\mathbf{H}_2(u_1, u_2, t)$, в силу замен (22) можно записать

$$\mathbf{H}_2(u_1, u_2, t) \approx \mathbf{r} \exp \left\{ \frac{(ju_1)^2}{2} \kappa_2^{(1)} Nt + \frac{(ju_2)^2}{2} \kappa_2^{(2)} Nt + ju_1 \cdot ju_2 \cdot \kappa_{12} Nt \right\},$$

поэтому для вектора-функции $\mathbf{H}(u_1, u_2, t)$ из (20) получим приближенное равенство

$$\mathbf{H}(u_1, u_2, t) \approx \mathbf{r} \exp \left\{ ju_1 \kappa_1^{(1)} Nt + ju_2 \kappa_1^{(2)} Nt + \frac{(ju_1)^2}{2} \kappa_2^{(1)} Nt + \frac{(ju_2)^2}{2} \kappa_2^{(2)} Nt + ju_1 \cdot ju_2 \cdot \kappa_{12} Nt \right\}.$$

Учитывая ранее введенное обозначение характеристической функции $h(u_1, u_2, t) = \mathbf{H}(u_1, u_2, t)\mathbf{e}$ числа событий, наступивших в двумерном маркированном ММРР, запишем предельную при достаточно больших N характеристическую функцию

$$(40) \quad h(u_1, u_2, t) \approx \exp \left\{ ju_1 \kappa_1^{(1)} Nt + ju_2 \kappa_1^{(2)} Nt + \frac{(ju_1)^2}{2} \kappa_2^{(1)} Nt + \frac{(ju_2)^2}{2} \kappa_2^{(2)} Nt + ju_1 \cdot ju_2 \cdot \kappa_{12} Nt \right\}.$$

Таким образом получили, что распределение вероятностей числа заявок, наступивших в двумерном маркированном ММРР, можно аппроксимировать двумерным нормальным распределением, которое определяется асимптотическими средними $\kappa_1^{(1)} Nt$ и $\kappa_1^{(2)} Nt$, дисперсиями $\kappa_2^{(1)} Nt$ и $\kappa_2^{(2)} Nt$ и ковариацией $\kappa_{12} Nt$ числа событий первого и второго типов, наступивших в двумерном маркированном ММРР за время t .

4. Численный пример

Опишем алгоритм построения аппроксимации распределения вероятностей числа заявок, наступивших в высокоинтенсивном двумерном маркированном ММРР.

Алгоритм построения гауссовской аппроксимации распределения вероятностей числа заявок, наступивших в высокоинтенсивном двумерном маркированном ММРР.

1. Задание значений параметров потока: элементов матриц \mathbf{Q} , $\mathbf{\Lambda}^{(1)}$, $\mathbf{\Lambda}^{(2)}$, значение параметра N .

2. Нахождение стационарного распределения вероятностей \mathbf{r} управляющего ММРР-потокком процесса $m(t)$, которое определяется системой уравнений $\mathbf{r}\mathbf{Q} = 0$, $\mathbf{r}\mathbf{e} = 1$.

3. Нахождение параметров $\kappa_1^{(1)}$, $\kappa_1^{(2)}$ по формулам (14).

4. Нахождение векторов \mathbf{g}_1 , \mathbf{g}_2 как решение матричных систем уравнений (26) и (27).

5. Нахождение параметров $\kappa_2^{(1)}$, $\kappa_2^{(2)}$, κ_{12} по формулам (25).

6. Построение двумерной гауссовской плотности $p(x_1, x_2)$ распределения вероятностей числа заявок, наступивших в высокоинтенсивном двумерном маркированном ММРР с математическими ожиданиями $\kappa_1^{(1)} Nt$ и $\kappa_1^{(2)} Nt$, дисперсиями $\kappa_2^{(1)} Nt$ и $\kappa_2^{(2)} Nt$, ковариацией $\kappa_{12} Nt$ числа событий первого и второго типов.

$$(41) p(x_1, x_2) = \frac{1}{2\pi \cdot \sqrt{\kappa_2^{(1)} Nt \cdot \kappa_2^{(2)} Nt \cdot (1 - (\kappa_{12} Nt)^2)}} \times \\ \times \exp \left\{ -\frac{1}{2(1 - (\kappa_{12} Nt)^2)} \times \left[\frac{(x_1 - \kappa_1^{(1)} Nt)^2}{\kappa_2^{(1)} Nt} - \right. \right. \\ \left. \left. - 2\kappa_{12} Nt \frac{(x_1 - \kappa_1^{(2)} Nt)(x_2 - \kappa_2^{(2)} Nt)}{\sqrt{\kappa_2^{(1)} Nt \cdot \kappa_2^{(2)} Nt}} + \frac{(x_2 - \kappa_1^{(2)} Nt)^2}{\kappa_2^{(2)} Nt} \right] \right\}.$$

7. Построение аппроксимации двумерного дискретного распределения вероятностей $P(n_1, n_2)$ числа событий, наступивших в высокоинтенсивном двумерном маркированном ММРР, по формуле

$$(42) P(n_1, n_2) = \frac{p(n_1, n_2)}{\sum_{n_1=0}^{\infty} \sum_{n_2=0}^{\infty} p(n_1, n_2)}.$$

Для предложенной модели двумерного маркированного ММРР положим

$$\mathbf{Q} = \begin{bmatrix} -1 & 0,4 & 0,6 \\ 0,2 & -0,5 & 0,3 \\ 0,3 & 0,4 & -0,7 \end{bmatrix}, \quad \mathbf{\Lambda}^{(1)} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}, \quad \mathbf{\Lambda}^{(2)} = \begin{bmatrix} 0,3 & 0 & 0 \\ 0 & 0,1 & 0 \\ 0 & 0 & 0,2 \end{bmatrix},$$

$$N = 30, t = 1$$

Решая систему $\mathbf{rQ} = 0$, $\mathbf{re} = 1$, получим
 $\mathbf{r} = \{0,197; 0,444; 0,359\}$.

Для рассматриваемых исходных данных \mathbf{Q} , $\Lambda^{(1)}$, $\Lambda^{(2)}$, N , t значения $\kappa_1^{(1)}$, $\kappa_1^{(2)}$, $\kappa_2^{(1)}$, $\kappa_2^{(2)}$, κ_{12} составляют

$$\kappa_1^{(1)} = 2,162, \quad \kappa_1^{(2)} = 0,175, \quad \kappa_2^{(1)} = 3,001, \quad \kappa_2^{(2)} = 0,272, \quad \kappa_{12} = -0,226.$$

График аппроксимаций дискретного распределения вероятностей $P(n_1, n_2)$ числа событий, наступивших в высокоинтенсивном двумерном маркированном ММРР, представлен на рис. 1.

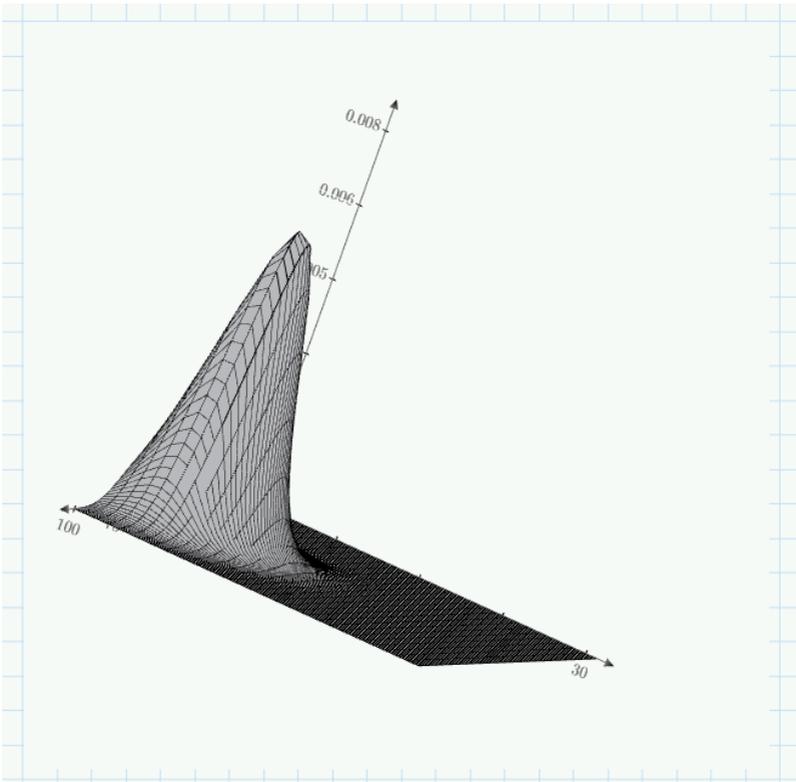


Рис. 1. Двумерная гауссовская аппроксимация распределения вероятностей $P(n_1, n_2)$ числа заявок, наступивших в высокоинтенсивном двумерном маркированном ММРР

5. Заключение

На основе реализации метода асимптотического анализа в предельном условии высокой интенсивности наступления событий и введения неограниченно возрастающего параметра N построена гауссовская двумерная аппроксимация распределения вероятностей числа событий, наступивших в высокоинтенсивном двумерном маркированном ММРР за время t . При этом параметры гауссовской аппроксимации находятся с использованием достаточно простых математических процедур, без применения сложных преобразований, требующих многократного двумерного интегрирования. Для целей исследования нагрузки на многомодальные системы важно, что мы получаем именно оценку двумерного распределения, т.е. можем учитывать взаимосвязь нагрузки одного и второго типа. А относительная простота полученных формул позволяет рассчитывать эту зависимость для разных соотношений параметров потока и оценивать чувствительность общей нагрузки от них.

Литература

1. БАСОВ О.О., ПАКУЛОВА Е.А., САИТОВ И.А. *Методологические основы построения интеллектуальных инфокоммуникационных систем: монография.* – Орёл : Академия ФСО России, 2020. – 272 с.
2. ВИШНЕВСКИЙ В.М., ДУДИН А.Н., КЛИМЕНОК В.И. *Стохастические системы с коррелированными потоками. Теория и применение в телекоммуникационных сетях.* – М.: ТЕХНОСФЕРА, 2018. – 564 с.
3. ГНЕДЕНКО Б.В., КОВАЛЕНКО И.Н. *Введение в теорию массового обслуживания* : учебное пособие. 4-е изд. – М.: изд-во ЛКИ, 2007. – 400 с.
4. КАГИРОВ И.А., РЮМИН Д.А., АКСЁНОВ А.А. и др. *Мультимедийная база данных жестов русского жестового языка в трехмерном формате // Вопросы языкознания.* – 2020. – №1 – С. 104–123.

5. МОИСЕЕВ А.Н., НАЗАРОВ А.А. *Бесконечнолинейные системы и сети массового обслуживания*. – Томск: Изд-во НТЛ, 2015. – 240 с.
6. НАУМОВ В.А., САМУЙЛОВ К.Е. *О марковских и рациональных потоках случайных событий. I* // Информатика и ее применение. – 2020. – Т. 14, вып. 3. – С. 13–19.
7. НАУМОВ В.А., САМУЙЛОВ К.Е. *О марковских и рациональных потоках случайных событий. II* // Информатика и ее применение. – 2020. – Т. 14, вып. 4. – С. 37–46.
8. РОНЖИН А.Л., КАРПОВ А.А. *Проектирование интерактивных приложений с многомодальным интерфейсом* // Доклады ТУСУРа. – 2010 – №1(21), часть 1 – С. 124–127.
9. ЮСУПОВ Р.М., РОНЖИН А.Л. *От умных приборов к интеллектуальному пространству* // Вестник Российской академии наук. – 2010. – Т. 80, №1. – С. 45–51.
10. BAI X., JIN S. *Performance analysis of an energy-saving strategy in cloud data centres based on a MMAP[K]/M[K]/N₁ + N₂ non-preemptive priority queue* // Future Gener. Comput. Syst. – 2022. – No. 136. – P. 205–220.
11. DUDIN A.N., DUDIN S.A., DUDINA O.S. *Randomized Threshold Strategy for Providing Flexible Priority in Multi-Server Queueing System with a Marked Markov Arrival Process and Phase-Type Distribution of Service Time* // Mathematics. – 2023. – No. 11. – P. 2669.
12. DUDIN A.N., KLIMENOK V.I., VISHNEVSKY V.M. *The Theory of Queueing Systems with Correlated Flows*. – Springer Nature: Cham, Switzerland, 2020. – 410 p.
13. HE Q.M. *Queues with marked customers* // Adv. Appl. Probab. – 1996. – No. 28. – P. 567–587.
14. HE Q.M. *Fundamentals of Matrix-Analytic Methods*. – Springer: New York, NY, USA, 2014. – 349 p.
15. KLIMENOK V., DUDIN A., VISHNEVSKY V. *Priority multi-server queueing system with heterogeneous customers* // Mathematics. – 2020. – No. 8. – P. 1501.
16. NAUMOV V., GAIDAMAKA Y., YARKINA N. et al. *Matrix and Analytical Methods for Performance Analysis of Telecom-*

- munication Systems.* – Springer: Berlin/Heidelberg, Germany, 2021. – 305 p.
17. NAZAROV A., MOISEEV A., LAPATIN I. et al. *Real Waiting Time in Single-Server Resource Queue with Markovian Arrival Process* // *Communications in Computer and Information Science.* – 2023. – Vol. 1803. – P. 116–125.
 18. NAZAROV A., PAUL S., PHUNG-DUC T. et al. *Analysis of Tandem Retrial Queue with Common Orbit and MMPP Incoming Flow* // *Lecture Notes in Computer Science.* – 2022. – Vol. 13766. – P 270–283.
 19. NAZAROV A., PHUNG-DUC T., PAUL S. et al. *Two-Way Communication Retrial Queue with Markov Modulated Poisson Input and Multiple Types of Outgoing Calls* // *Communications in Computer and Information Science.* – 2022. – Vol. 1748. – P. 370–381.
 20. RAJ R., JAIN V. *Optimization of traffic control in MMAP[2]/PH[2]/S priority queueing model with PH retrial times and the preemptive repeat policy* // *J. Ind. Manag. Optim.* – 2023. – No. 19. – P 2333–2353.
 21. SAMOUYLOV K., DUDINA O., DUDIN A. *Analysis of Multi-Server Queueing System with Flexible Priorities* // *Mathematics.* – 2023. – No. 11. – P. 1040.

STUDY OF TWO-DIMENSIONAL MARKED MMPP UNDER THE HIGH RATE LIMIT CONDITION

Svetlana Paul, Tomsk State University, Tomsk, Doctor of Science, professor (paulsv82@mail.ru)

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

Ivan Lapatin, Tomsk State University, Tomsk, Cand.Sc., associate professor (ilapatin@mail.ru).

Abstract: This paper considers a mathematical model of a heterogeneous flow in the form of a two-dimensional marked MMPP. The study of such models is necessary to analyze the load on multimodal systems. Multimodal interfaces are capable of processing multiple natural human input methods, each of which requires specific resources for recognition, processing and transmission. To design such systems, it is

necessary to estimate the required resources. These estimates can be based on the joint probability distribution of the number of calls of each type over a certain period of time. The paper proposes an asymptotic method to estimating the two-dimensional probability distribution of the number of arrivals in a high-intensity marked Markov Modulated Process. The limiting condition of high intensity is determined by the parameter of the rate of arrivals in the process over a certain time. The asymptotic analysis method is carried out in two stages. At the first stage, the parameters are obtained that determine the asymptotic mean numbers of arrivals of the first and second types that occurred in the high-intensity flow. At the second stage, the parameters are found that determine the asymptotic variances and the covariance of the number of events of the first and second types. It is shown that the limiting distribution of the number of events that occurred in a high-intensity marked MMPP is a two-dimensional Gaussian. The resulting formulas for finding the distribution and its characteristics have fairly simple expressions, the unknowns in which are found by solving systems of linear equations.

Keywords: Marked MMPP, asymptotic analysis method, multimodal system.

УДК 519.872

ББК 22.171

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

Поступила в редакцию 03.06.2024.

Опубликована 30.11.2024.

ОБ ИСКЛЮЧИТЕЛЬНОМ СЛУЧАЕ В ЗАДАЧЕ О РАЗМЕЩЕНИИ ПОЛЮСОВ

Мухин А. В.¹

(Нижегородский государственный университет
им. Н.И. Лобачевского, Нижний Новгород)

Рассматривается задача о размещении полюсов с помощью статического регулятора по выходу. Если задача разрешима, то спектр матрицы замкнутой системы можно расположить в любых заданных, симметричных относительно действительной оси точках комплексной полуплоскости. Это дает возможность не просто стабилизировать систему, но и задавать требуемые характеристики, такие, как запас устойчивости, время переходных процессов и другие. Известно, что если произведение числа входов и выходов превышает размерность системы, то задача о размещении полюсов для системы, заданной в виде передаточной матрицы, разрешима. В статье показано, что данное соотношение не является достаточным условием для системы, заданной в пространстве состояний. Существует исключительный случай, при котором задача о размещении полюсов принципиально неразрешима. Этот случай легко обнаруживается с помощью перемножения матриц выхода и входа. Если это произведение дает нулевую матрицу, то в силу неизменности следа матрицы замкнутой системы, задача неразрешима как в действительной, так и в комплексной области. Причем произведение матриц выхода и входа инвариантно относительно базиса. Сформулировано необходимое условие разрешимости.

Ключевые слова: задача о размещении полюсов, статический регулятор по выходу, след матрицы.

1. Введение

Задача о размещении полюсов с помощью статического регулятора по выходу представляет большой практический интерес. Если задача разрешима, то спектр матрицы замкнутой системы можно расположить в любых заданных, симметричных относительно действительной оси точках комплексной полуплоскости. Проблема размещения собственных значений входит в число трудных задач теории управления [3]. Необходимые условия, достаточные условия разрешимости, выражаемые в виде соотношения между размерностью системы, количеством

¹ Алексей Валерьевич Мухин, аспирант (muhin-aleksei@yandex.ru).

входов и выходов получены в [7, 8, 11, 16, 17]. Различные подходы к решению задачи изложены в [5, 9, 12–14, 18–20]. Подробный обзор основных результатов можно найти в [4]. Задача затрагивалась также и в более ранних обзорах [6, 15].

В статье получен новый результат, дополняющий достаточные условия [17] для систем, заданных в пространстве состояний. Показано, что если произведение матриц выхода и входа дает нулевую матрицу, то задача о размещении полюсов с помощью статического регулятора по выходу принципиально неразрешима. Сформулировано необходимое условие разрешимости.

Структура статьи стандартная: во втором разделе приведены предварительные сведения; третий раздел содержит основной результат; в четвертом разделе рассмотрен практический пример; последний раздел – заключение.

2. Предварительные сведения

Дана линейная непрерывная стационарная управляемая и наблюдаемая система:

$$(1) \quad x = Ax + Bu, \quad x(0) = x_0,$$

$$y = Cx,$$

где $x \in R^n$ – состояние; $u \in R^m$ – вход; $y \in R^p$ – измеряемый выход; $A \in R^{n \times n}$ – обратимая матрица системы; $B \in R^{n \times m}$ – матрица входа; $C \in R^{p \times n}$ – матрица выхода.

Без потери общности полагаем, что размерности матриц B и C не избыточны:

$$\text{rank}(B) = m,$$

$$\text{rank}(C) = p.$$

Применим к (1) закон управления по измеряемому выходу:

$$u = Ky,$$

где $K \in R^{m \times p}$.

Основной результат статьи связан с задачей о размещении полюсов. Соответствующая задача формулируется следующим образом: *существует ли матрица K , которая обеспечивает совпадение спектра матрицы замкнутой системы*

$$\widetilde{A}_c = A + BKC$$

с любым наперед заданным множеством точек в комплексной плоскости, расположенных симметрично относительно действительной оси:

$$\{\alpha_1, \alpha_2, \dots, \alpha_n\} \subset \mathbb{C}.$$

Задачу можно рассматривать как частный случай задачи о статическом регуляторе по выходу. Из разрешимости задачи о размещении полюсов следует разрешимость задачи о статическом регуляторе по выходу. Если матрица выхода является обратимой, то управляемость системы является необходимым и достаточным условием для разрешимости задачи (см., например, [4]). Для практических целей такая постановка менее интересна, поскольку осуществляется при полностью измеряемом состоянии. Наибольший интерес – когда матрицы B и C одновременно необратимы. Необходимым условием разрешимости задачи является соотношение $mp \geq n$ [16]. В случае комплексных матриц K это условие является также достаточным.

3. След матрицы замкнутой системы

Покажем условие, выражаемое в виде произведения матриц C и B , при котором след матрицы замкнутой системы является постоянным. Для этого нам понадобится новый базис, в котором это условие легко обнаруживается. С этой целью выполним линейное преобразование базиса системы (1) посредством обратной матрицы $S \in \mathbb{R}^{n \times n}$ так, чтобы матрица выхода приняла следующий вид:

$$C = (I_p \quad 0_{p \times (n-p)}),$$

где I_p – единичная матрица ранга p .

В новом базисе система (1) примет вид

$$(2) \quad \begin{aligned} x &= SAS^{-1}x + SBu = Ax + Bu, \quad x(0) = x_0, \\ u &= CS^{-1}x = Cx = (I_p \quad 0)x, \end{aligned}$$

где x – вектор состояния системы в новом базисе.

Отметим, что никаких требований на структуру матрицы A не накладывается. Существует, вообще говоря, множество мат-

риц S , обеспечивающих требуемый вид матрицы выхода. Действительно, запишем матрицу преобразования в виде

$$S = \begin{pmatrix} S_p \\ S_{n-p} \end{pmatrix}.$$

Из уравнения

$$CS^{-1} = \begin{pmatrix} I_p & 0 \end{pmatrix}$$

следует равенство $C = S_p$. Нижние $(n - p)$ строк являются свободными и должны лишь обеспечивать обратимость матрицы S .

Тогда в качестве матрицы S можно взять, например,

$$S = \begin{pmatrix} C \\ \mathcal{N}_C^T \end{pmatrix},$$

где \mathcal{N}_C – любое решение уравнения $C\mathcal{N}_C = 0$.

Обратной матрицей будет:

$$S^{-1} = \begin{pmatrix} C^T(CC^T)^{-1} & \mathcal{N}_C(\mathcal{N}_C^T\mathcal{N}_C)^{-1} \end{pmatrix}.$$

Запишем также уравнение замкнутой системы в новом базисе:

$$(3) \quad x_c = A_c x = (A + BKC)x = \begin{pmatrix} A_p + BK & A_{n-p} \end{pmatrix} x,$$

где $A_p \in R^{n \times p}$.

Равенство спектров \widetilde{A}_c и A_c гарантируется подобием матриц. Далее будем рассматривать матрицу A_c . Докажем лемму:

Лемма. Если произведение $CB = 0$, то след матрицы A_c равен следу матрицы A .

Доказательство. Разобьем матрицы A и B на блоки следующим образом:

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix},$$

$$B = \begin{pmatrix} B_{11} \\ B_{21} \end{pmatrix},$$

где $A_{11} \in R^{p \times p}$; $B_{11} \in R^{p \times m}$.

Произведение матриц BKC примет вид

$$BKC = \begin{pmatrix} B_{11}K & 0_{p \times (n-p)} \\ B_{21}K & 0_{(n-p) \times (n-p)} \end{pmatrix}.$$

Тогда матрицу замкнутой системы (3), можно записать следующим образом:

$$(4) \quad A_c = \begin{pmatrix} A_{11} + B_{11}K & A_{12} \\ A_{21} + B_{21}K & A_{22} \end{pmatrix}.$$

Если $B_{11} = 0$, то (4) примет вид

$$A_c = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} + B_{21}K & A_{22} \end{pmatrix}.$$

След матрицы A_c в таком случае равен следу матрицы A :

$$(5) \quad \text{tr}(A_c) = \text{tr}(A) = \sum_{i=1}^n \lambda_i.$$

Но так как в выбранном базисе имеет место равенство

$$CB = (I_p \quad 0) \begin{pmatrix} B_{11} \\ B_{21} \end{pmatrix} = B_{11},$$

то если $B_{11} = 0$, то и $CB = 0$. Лемма доказана.

Характеристический многочлен A_c можно записать в виде:

$$\lambda^n - (\text{tr}(A_c))\lambda^{n-1} + \dots + \det(A_c) = 0.$$

Из спектрального ограничения (5) следует, что расположить произвольным образом n корней такого многочлена невозможно, так как $\text{tr}(A_c) = \text{const}$. Таким образом, справедливо следующее утверждение:

Утверждение 1. Необходимым условием разрешимости задачи о размещении полюсов для управляемой и наблюдаемой системы является

$$(6) \quad CB \neq 0.$$

В общем случае относительно ранга произведения матриц C и B имеет место соотношение

$$0 \leq \text{rank}(CB) \leq \min\{m, p\}.$$

Обратимость хотя бы одной из матриц гарантированно обеспечивает выполнение (6). Следует принять во внимание, что произведение матриц выхода и входа не зависит от базиса. Поэтому из (6) следует $CB \neq 0$ и наоборот. Следовательно, проверке условия (6) можно выполнить и в исходном базисе. Отметим, что произведение матриц выхода и входа связаны с понятием относительного порядка системы [10]. Обратимся теперь к результату [17], согласно которому, для системы, заданной в виде передаточной матрицы, соотношение

$$(7) \quad mp > n$$

является достаточным условием разрешимости задачи о размещении полюсов. Докажем, что соотношение (7) не является достаточным условием разрешимости задачи о размещении полюсов для системы, заданной в пространстве состояний. Для этого достаточно доказать, что из (7) не следует (6). Рассмотрим следующий пример. Пусть матрицы (2) имеют вид

$$(8) \quad A = \begin{pmatrix} \mathbf{a}_{11} & a_{12} & a_{13} & a_{14} & a_{15} \\ a_{21} & \mathbf{a}_{22} & a_{23} & a_{24} & a_{25} \\ a_{31} & a_{32} & \mathbf{a}_{33} & a_{34} & a_{35} \\ a_{41} & a_{42} & a_{43} & \mathbf{a}_{44} & a_{45} \\ a_{51} & a_{52} & a_{53} & a_{54} & \mathbf{a}_{55} \end{pmatrix} \in R^{5 \times 5},$$

$$B = \begin{pmatrix} 0_{3 \times 2} \\ I_2 \end{pmatrix},$$

$$C = (I_3 \quad 0_{3 \times 2}).$$

Проверяем, что $CB = 0_{3 \times 2}$. Таким образом, условие (7) выполняется, а условие (6) – нет. Пусть $K = \{k_{ij}\} \in R^{2 \times 3}$. Запишем матрицу A_c :

$$A_c = \begin{pmatrix} \mathbf{a}_{11} & a_{12} & a_{13} & a_{14} & a_{15} \\ a_{21} & \mathbf{a}_{22} & a_{23} & a_{24} & a_{25} \\ a_{31} & a_{32} & \mathbf{a}_{33} & a_{34} & a_{35} \\ a_{41} + k_{11} & a_{42} + k_{12} & a_{43} + k_{13} & \mathbf{a}_{44} & a_{45} \\ a_{51} + k_{21} & a_{52} + k_{22} & a_{53} + k_{23} & a_{54} & \mathbf{a}_{55} \end{pmatrix}.$$

Видим, что $\text{tr}(A_c) = \text{tr}(A)$. Убеждаемся, что из (7) не следует (6). Если в приведенном выше примере увеличить число входов до $m = 3$, то условие (6) будет выполняться:

$$CB = (I_3 \quad 0_{3 \times 2}) \begin{pmatrix} 0_{2 \times 3} \\ I_3 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}.$$

Соответствующая матрица A_c будет равна:

$$A_c = \begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{25} \\ a_{31} + k_{11} & a_{32} + k_{12} & \mathbf{a}_{33} + k_{13} & a_{34} & a_{35} \\ a_{41} + k_{21} & a_{42} + k_{22} & a_{43} + k_{23} & a_{44} & a_{45} \\ a_{51} + k_{31} & a_{52} + k_{32} & a_{53} + k_{33} & a_{54} & a_{55} \end{pmatrix}.$$

Теперь $\text{tr}(A_c) \neq \text{tr}(A)$. Справедливо утверждение:

Утверждение 2. Соотношение (7) не является достаточным условием разрешимости задачи о размещении полюсов для системы, заданной в пространстве состояний.

Покажем, что условия (6) и (7) можно обобщить. Введем соотношение:

$$(9) \quad m + p > n.$$

Так как $\text{rank}(C) = p$ и $\text{rank}(B) = m$, то для выполнения (7) достаточно, чтобы $m > n - p$. Поэтому из неравенства (9) следует (6). Обратное неверно. Из (7) следует, что $\min\{m, p\} > 1$. Тогда из (9) следует также условие (7). Если $\min\{m, p\} > 1$, то соотношение (9) выполняется при полной обратной связи. Задача

о размещении полюсов в таком случае имеет простое решение. Следовательно, из соотношения (9) следуют условия (6) и (7). Необходимо отметить, соотношение (9) хорошо известно. В частности, в [11] получено достаточное условие разрешимости задачи в виде неравенства:

$$(10) \quad n - m - p + 1 \leq q,$$

где q – порядок регулятора.

В случае статического регулятора ($q = 0$) неравенство (10) эквивалентно (9). Схожие соотношения получены также и в других работах (см. обзор [4]).

4. Практический пример

В практических задачах, когда в измерении и управлении задействуются отдельные переменные, а не их комбинации, равенство $CB = 0$ вполне возможно. Покажем это на примере системы, представляющей вертикальный жесткий ротор, вращающийся в электромагнитных подшипниках [1]. Электромагнитные подшипники представляют большой практический интерес для целого ряда промышленных применений. Преимущества таких систем – отсутствие физического контакта и как следствие, механического трения. Линеаризованная модель, описывающая динамику такого тела, имеет вид [2]:

$$\begin{aligned} x_1 &= x_5, \\ x_2 &= x_6, \\ x_3 &= x_7, \\ x_4 &= x_8, \\ x_5 &= \lambda(x_9 - x_{10}) + 2\lambda x_1 - \rho x_6, \\ x_6 &= \lambda(x_{12} - x_{11}) + 2\lambda x_2 + \rho x_5, \\ x_7 &= -(x_{11} + x_{12}) + 2x_3, \\ x_8 &= -(x_9 + x_{10}) + 2x_4, \\ x_9 &= (x_8 - x_5) - \mu x_9 + u_1, \\ x_{10} &= -(x_8 + x_5) - \mu x_{10} + u_2, \\ x_{11} &= (x_7 + x_6) - \mu x_{11} + u_3, \\ x_{12} &= -(x_7 - x_6) - \mu x_{12} + u_4, \end{aligned}$$

где λ, ρ, μ – конструктивные параметры.

Первые две переменные описывают углы наклона центра масс ротора, вторые две – смещения в двух плоскостях, следующие четыре – скорости изменения первых четырех перемен-

ных, последние четыре – токи в цепях электромагнитов. Матрица входа задана следующим образом:

$$B = \begin{pmatrix} 0_{8 \times 4} \\ I_4 \end{pmatrix}.$$

Предположим, что измеряются первые восемь переменных. Матрица выхода имеет вид:

$$C = (I_8 \quad 0_{8 \times 4}).$$

Видим, что $CB = 0$. Решить задачу о размещении полюсов с помощью статического регулятора по выходу при такой организации измерений принципиально невозможно. В данном случае можно лишь стабилизировать систему.

5. Заключение

В статье показано, что при решении задачи о размещении полюсов с помощью статического регулятора по выходу для системы, заданной в пространстве состояний, необходимо учитывать структуру матриц входа и выхода. Если произведение матриц выхода и входа образует нулевую матрицу, то задача о размещении полюсов неразрешима как в действительной, так и в комплексной области. Достаточное условие разрешимости, выражаемое в виде произведения числа входов и выходов, в таком случае неприменимо для систем, заданных в пространстве состояний.

Литература

1. БАЛАНДИН Д.В., КОГАН М.М. *Управление движением вертикального жесткого ротора, вращающегося в электромагнитных подшипниках* // Известия РАН. ТИСУ. – 2011. – №5. – С. 13–17.
2. МУХИН А.В. *Математическое моделирование процесса стабилизации жесткого ротора, вращающегося в электромагнитных подшипниках* // Труды НГТУ им. Р.Е. Алексеева. – 2021. – №2. – С. 36–48.
3. ПОЛЯК Б.Т., ЩЕРБАКОВ П.С. *Трудные задачи линейной теории управления. Некоторые подходы к решению* // Автоматика и телемеханика. – 2005. – №5. – С. 4–46.
4. ШУМАФОВ М.М. *Стабилизация линейных систем управления. Проблема назначения полюсов. Обзор* // Вестник СПбГУ.

- Математика. Механика. Астрономия. – 2019. – Т. 6(64), вып. 4. – С. 564–591.
5. BELOZYOROV V.Ye. *New solution method of linear static output feedback design problem for linear control systems* // Linear Algebra and its Applications. – 2016. – Vol. 54. – P. 204–227.
 6. BERNSTEIN D.S. *Some Open Problems in Matrix Theory Arising in Linear Systems and Control* // Linear Algebra and its Applications. – 1992. – Vol. 162–164. – P. 409–432.
 7. BROCKETT R., BYRNES C. *Multivariable Nyquist criteria, loci root and pole placement: A geometric viewpoint* // IEEE Trans. Automat. Control. – 1981. – Vol. 26. – P. 271–284.
 8. EREMENKO A., GABRIELOV A. *Pole placement by static output feedback for generic linear system* // SIAM J. Control Optim. – 2002. – Vol. 41. – P. 303–312.
 9. HAMID M. *The resolution of the equation and the pole assignment problem: A general approach* // Automatica. – 2017. – Vol. 79. – P. 162–166.
 10. ISIDORI A. *Nonlinear Control System*. – Berlin; New York: Springer-Verlag, 1985.
 11. KIMURA H. *On pole placement by gain output feedback* // IEEE Trans. Automat. Control. – 1975. – Vol. 20. – P. 509–519.
 12. MOSTAFA EL-SAYED M.E., TAWHID M.A., ELWAN E.R. *Nonlinear conjugate gradient methods for the output feedback pole assignment problem* // Pacific Journal Optimization. – 2016. – Vol. 12(1). – P. 55–85.
 13. SCHMID R., NTOGRAMATZIDIS L., NGYEN T. et al. *A unified method for optimal arbitrary placement* // Automatica. – 2014. – Vol. 50. – P. 2150–2154.
 14. SHANA M., PRASANNA C., MADHU N.B. *Approximating constrained minimum cost input–output selection for generic arbitrary pole placement in structured systems* // Automatica. – 2019. – Vol. 107. – P. 200–210.
 15. SYRMOS V.L., ABDALLAH C.T., DORATO P et al. *Static Output Feedback. A Survey* // Automatica. – 1997. – Vol. 33(2). – P. 125–137.
 16. WILLEMS J.C., HESSELINK W.H. *Generic properties of the pole placement problem* // Proc. of the 7th IFAC Congress. – 1978. – P. 1725–1729.

17. WANG X. *Pole placement by static output feedback* // J. Math. System, Estimation and Control. – 1992. – Vol. 2. – P. 205–218.
18. YANG K., ORSI R. *Generalized pole placement via based on projections* // Automatica. – 2006. – Vol. 42. – P. 2143–2150.
19. YANNAKOUDACIS A.G. *The static output feedback from the invariant point of view* // IMA Journal of Mathematical Control and Information. – 2016. – Vol. 33. –P. 639–669.
20. ZUBOV N.E., MIKRIN E.A., MISRIKHANOV M.Sh. et al. *Controlling the finite eigenvalues of the descriptor system* // Doklady Akademy Nauk. – 2015. – No. 460(4). – P. 381–384.

ABOUT AN UNUSUAL CASE IN THE POLE PLACEMENT PROBLEM

Aleksey Mukhin, Lobachevsky State University, Nizhny Novgorod, postgraduate student (myhin-aleksey@yandex.ru).

Abstract: The pole placement problem using a static output feedback is considered. If the problem is solvable, then the spectrum of the closed-loop system matrix can be located at any given points of the complex half-plane, symmetric with respect to the real axis. This makes it possible not only to stabilize the system, but also to set the required characteristics, such as stability margin, transition time, and others. It is known that if the multiplication of the number of inputs and outputs is greater than the dimension of the system, then the pole placement problem for a system in the form of a transfer matrix is solvable. The article shows that this ratio is not a sufficient condition for a system defined in the state space. There is an exceptional case in which the pole placement problem is fundamentally unsolvable. This case is simple discovered by means of multiplying of matrixes output and input. If this product gives a zero matrix, then due to the matrix trace consistency of the closed system matrix, the problem is unsolvable both in the real and in the complex domain. Moreover, the product of the output and input matrices is invariant with respect to the basis. A necessary condition for solvability is formulated.

Keywords: the pole placement problem, static output feedback, trace of matrix.

УДК 517.977

ББК 22.18

*Статья представлена к публикации
членом редакционной коллегии И.Б. Фуртатом.*

Поступила в редакцию 23.07.2024.

Опубликована 30.11.2024.

СИНТЕЗ УПРАВЛЕНИЯ ДЛЯ СТАЦИОНАРНЫХ СИСТЕМ ПРИ ДОПОЛНИТЕЛЬНЫХ ОГРАНИЧЕНИЯХ НА ВНЕШНЕЕ ВОЗМУЩЕНИЕ В РАМКАХ АНИЗОТРОПИЙНОЙ ТЕОРИИ¹

Юрченков А. В.²

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

Рассматривается линейная дискретная стационарная система с управлением под влиянием окрашенного возмущения. Внешнее возмущение выбирается из класса нецентрированных стационарных гауссовских последовательностей случайных векторов с известным ограничением на уровень средней анизотропии. Для указанного класса объектов управления вводится динамический регулятор, с помощью которого необходимо обеспечить ограничение анизотропийной нормы от внешнего возмущения к управляемому выходу замкнутой системы. Задача синтеза анизотропийного динамического регулятора заключается в нахождении пространственной реализации регулятора из условия ограниченности анизотропийной нормы замкнутой системы. Используя линеаризующую обратимую замену переменных, поставленную задачу можно свести к численному решению задачи выпуклой оптимизации с ограничениями специального вида, характерными для анизотропийной теории. В постановке задачи считается, что среднее внешнего возмущения неизвестно, но известно ограничение на него в виде неравенства. Этот параметр обуславливает появление дополнительного ограничения в задаче выпуклой оптимизации. Результирующая система неравенств представляет собой линейные матричные неравенства в совокупности с неравенством специального вида, которое является нелинейным относительно неизвестных параметров, но одновременно является выпуклым по этим параметрам. Задача поиска матриц регулятора может быть решена стандартными методами.

Ключевые слова: анизотропийная теория, выпуклая оптимизация, нецентрированные возмущения.

1. Введение

Рассматривая самые различные динамические объекты, практически невозможно найти такие, которые в процессе функ-

¹ Автор признателен к.ф.-м.н. А.Ю. Кустову за ценное обсуждение содержания статьи.

² Александр Викторович Юрченков, к.ф.-м.н., с.н.с.
(alexander.yurchenkov@yandex.ru)

ционирования не были бы подвержены влиянию внешних факторов в самом широком смысле: окружающая среда, неизмеряемые силы, помехи в каналах связи, – список можно продолжать, и каждый инженер может добавить свои пункты. Поэтому и задачи теории управления можно разделить на два больших класса: с внешними возмущениями и без них. Каждая из таких задач имеет свои допущения и методы решения, причем для одного и того же объекта возможно применение суперпозиции управлений. В данной работе будет сделан упор на подавлении влияния внешнего возмущения из специального класса. Хорошо известные в теории управления \mathcal{H}_2 - и \mathcal{H}_∞ -оптимальные подходы используют соответствующие нормы передаточной функции замкнутой системы в пространствах Харди. Однако выбор конкретного критерия качества зависит от гипотезы о принадлежности внешнего возмущения к конкретному классу. Например, \mathcal{H}_∞ -оптимальная теория управления подразумевает, что внешнее возмущение (входной сигнал) квадратично интегрируем или квадратично суммируем (в зависимости от выбранного способа описания времени), что подробно обсуждается в [14]. Однако законы управления, синтезируемые в рамках этой теории, достаточно консервативны [16] (т.е. требуют значительных ресурсов на формирование управления, поскольку рассчитаны на «наихудший случай»). С другой стороны, \mathcal{H}_2 -оптимальная теория управления обладает наилучшими показателями качества в случае гауссовского возмущения для управляемой системы. Каждая из перечисленных теорий накладывает ограничения на входной сигнал при минимизации соответствующей нормы для систем с замкнутым контуром. Стоит упомянуть, что как и \mathcal{H}_2 -, так и \mathcal{H}_∞ -оптимальный синтез управления может быть реализован численно с помощью решения уравнений Риккати. Но несмотря на простоту реализации, регуляторы на базе \mathcal{H}_2 - и \mathcal{H}_∞ -теорий эффективны только в том случае, если верны основные гипотезы о внешних возмущениях. Многокритериальные задачи в терминах смешанных формулировок $\mathcal{H}_2/\mathcal{H}_\infty$ -оптимизации могут компенсировать некоторые недостатки обеих теорий [1, 17, 18, 33].

Но более интересный и физически обоснованный случай возникает, когда свойства экзогенного возмущения точно не известны. Для измерения неопределенности стохастических свойств возмущений можно применить некоторые концепции теории информации. Таким образом была создана теория, основанная на анизотропии. В этой части теории управления в качестве критерия эффективности анализируется анизотропийная норма системы.

Теория, предложенная И.Г. Владимировым в последнем десятилетии 20-го века, обобщает \mathcal{H}_2 - и \mathcal{H}_∞ -подходы с точки зрения описания множества внешних возмущений. Основы этой теории были представлены в работах [3, 25, 28]. Введенное понятие анизотропии случайного вектора позволило максимально расширить класс внешних возмущений, действующих на динамические системы, эталоном при этом служили вектора с равномерным распределением на единичной сфере. В зависимости от введенной меры отличия от эталонного вектора внешнего возмущения, анизотропийная норма находится между масштабированной \mathcal{H}_2 - и \mathcal{H}_∞ -нормами той же системы, включая и эти предельные случаи. Существует несколько подходов к вычислению анизотропийной нормы системы, но стоит выделить алгоритм на основе выпуклой оптимизации [29]. Методами выпуклой оптимизации можно найти минимальное пороговое значение анизотропийной нормы, которое гарантировано ограничит анизотропийную норму сверху. В основе такого подхода лежит утверждение, называемое леммой о вещественной ограниченности в рамках анизотропийной теории [22].

Изначально в анизотропийной теории были рассмотрены линейные дискретные стационарные системы с центрированным возмущением [13]. За тридцать лет были рассмотрены различные задачи теории управления в анизотропийной постановке: от синтеза ПИД-регулятора [8] и предельных случаев [2, 11] до параметризации оптимальных регуляторов [4] и постановки задачи в непрерывном случае [12]. Также были рассмотрены случаи ненулевого среднего внешнего возмущения [20, 23]. В настоящее время класс объектов не ограничивается детерминирован-

ным случаем и включает в себя системы с мультипликативными шумами [9, 10, 19, 32] и стохастические системы [21].

В данной работе рассматривается задача построения анизотропийного субоптимального управления для стационарной системы, на вход которой поступает возмущение с ограничениями. Решение задачи сводится к решению системы линейных матричных неравенств с выпуклым ограничением специального вида, характерного для анизотропийной теории. Численно задача может быть решена стандартными прикладными пакетами [26]. Статья организована следующим образом: во второй части приводится необходимый теоретический минимум для дальнейшего изложения, третья содержит постановку задачи, четвертая часть посвящена основному результату, в пятой части продемонстрированы результаты численного моделирования, последняя часть – заключение.

2. Предварительные сведения

В этом разделе приводятся краткие теоретические сведения, касающиеся дискретных линейных стационарных систем в рамках анизотропийной теории, а также сведения относительно нецентрированного внешнего возмущения.

2.1. СРЕДНЯЯ АНИЗОТРОПИЯ И АНИЗОТРОПИЙНАЯ НОРМА

Первое определение анизотропии случайного m -мерного вектора было дано в статье [3]. В более поздних работах, например, в статье [13], под анизотропией случайного вектора w с плотностью распределения $f(x)$ понимается следующее.

Определение 1 [3]. Анизотропией случайного m -мерного вектора с плотностью распределения $f(x)$ называют величину

$$A(w) = \min_{\lambda > 0} D(f || p_\lambda),$$

где

$$D(f || p_\lambda) = \mathbf{E} \left[\ln \frac{f}{p_\lambda} \right],$$

что представляет собой относительную энтропию или информационное уклонение Кульбака – Лейблера функции f по отношению к плотности p_λ центрированного симметричного гауссовского распределения

$$p_\lambda(x) = (2\pi\lambda)^{-m/2} \exp\left(-\frac{\|x\|^2}{2\lambda}\right).$$

В качестве внешнего возмущения для дискретных стационарных систем выбирают последовательности векторов, в анизотропной теории в качестве характеристики последовательности $\{w_k\}$ рассматривают среднюю анизотропию.

Определение 2 [3]. Средней анизотропией последовательности $\overline{W} = \{w_k\}$ называют величину

$$(1) \quad \overline{\mathbf{A}}(W) = \lim_{N \rightarrow \infty} \frac{\mathbf{A}(W_{0:N-1})}{N},$$

где $W_{0:N-1} = [W_0^\top, \dots, W_{N-1}^\top]^\top$ – расширенный вектор.

Обозначим в качестве входа и выхода линейной системы F последовательности $W \in \mathbb{L}_2^m$ и $Z \in \mathbb{L}_2^p$ соответственно, где \mathbb{L}_2^* обозначает пространство Лебега векторов с конечной нормой. Причем возмущающая последовательность W получена с помощью линейного формирующего фильтра G из стандартного гауссовского белого шума V :

$$w_j = \sum_{k=0}^{\infty} g_k v_{j-k}, \quad j \in \mathbb{Z},$$

где $g_k \in \mathbb{R}^{m \times m}$. Формирующий фильтр G имеет передаточную матричнозначную функцию $G(z)$:

$$G(z) = \sum_{k=0}^{\infty} g_k z^k,$$

которая является аналитичной внутри единичного диска $|z| < 1$, $z \in \mathbb{Z}$. Передаточная функция $G(z)$ имеет конечную \mathcal{H}_2 -норму,

равную

$$\|G\|_2 = \left(\sum_{k=0}^{\infty} \text{tr}(g_k g_k^\top) \right)^{1/2}.$$

Линейная система F имеет передаточную функцию $F(z)$ в соответствующем пространстве Харди $\mathcal{H}_\infty^{p \times m}$ с конечной \mathcal{H}_∞ -нормой:

$$\|F\|_\infty = \sup_{|z| < 1} \bar{\sigma}(F(z)) = \text{ess sup}_{-\pi \leq \omega \leq \pi} \bar{\sigma}(\hat{F}(\omega)),$$

где $\bar{\sigma}(\cdot)$ обозначает максимальное сингулярное значение матрицы.

Рассмотрим множество формирующих фильтров

$$\mathcal{G}_a = \{G \in \mathcal{H}_2^{m \times m} : W = GV, \bar{A}(W) \leq a\},$$

где $V = \{v_k\}_{k \in \mathbb{Z}}$ – стандартный гауссовский белый шум с нулевым средним и единичной ковариационной матрицей, как это постулируется в работе [25].

Определение 3 [28]. Анизотропийной нормой $\|F\|_a$ линейной системы F называют величину

$$(2) \quad \|F\|_a = \sup_G \left\{ \frac{\|FG\|_2}{\|G\|_2} : G \in \mathcal{G}_a \right\}.$$

Для любой несферичной системы $F \in \mathcal{H}_\infty^{p \times m}$ выполнено следующее ограничение на анизотропийную норму:

$$\frac{1}{\sqrt{m}} \|F\|_2 = \|F\|_0 \leq \|F\|_a \leq \lim_{a \rightarrow \infty} \|F\|_a = \|F\|_\infty,$$

что в определенном смысле и служит основанием для утверждения, что анизотропийная теория обобщает результаты \mathcal{H}_2 - и \mathcal{H}_∞ -теорий, поскольку в предельных случаях получает результаты, характерные для отмеченных теорий.

2.2. ЧАСТОТНАЯ ОБЛАСТЬ

Выражения для средней анизотропии и анизотропийной нормы линейной дискретной системы можно сформулировать на

языке передаточных функций. Для этого рассмотрим набор специальных функций:

$$(3) \quad \mathcal{A}_0(q) = \frac{m}{2} (\ln(\Phi(q)) - \Psi(q)),$$

$$(4) \quad \mathcal{N}_0(q) = \left(\frac{\Phi(q) - 1}{q\Phi(q)} \right)^{1/2},$$

где

$$(5) \quad \Phi(q) = \frac{1}{2\pi m} \int_{-\pi}^{\pi} \text{tr}((I_m - q\Lambda)^{-1}) d\omega$$

$$(6) \quad \Psi = \frac{1}{2\pi m} \int_{-\pi}^{\pi} \ln \det(I_m - q\Lambda) d\omega$$

и $\Lambda = (\widehat{F}(\omega))^* \widehat{F}(\omega)$, $\omega \in [-\pi; \pi)$. Здесь под $\widehat{F}(\omega)$ понимается значение передаточной функции на границе единичного диска

$$\widehat{F}(\omega) = \lim_{r \rightarrow 1_-} F(r \exp(i\omega)), \quad \omega \in [-\pi; \pi).$$

Функции (3)–(6) являются неубывающими по параметру $q \in (0; \|F\|_{\infty}^{-2})$.

Теорема 1 [27]. Для любой линейной системы с передаточной функцией $F(z)$, удовлетворяющей выражению

$$\frac{1}{\sqrt{m}} \|F\|_2 < \|F\|_{\infty},$$

анизотропная норма будет равна

$$(7) \quad \|F\|_a = \mathcal{N}_0(\mathcal{A}_0^{-1}(a)),$$

где $q = \mathcal{A}_0^{-1}(a)$ и любой формирующий фильтр, удовлетворяющий представлению

$$\widehat{G}\widehat{G}^* = (I_m - q\Lambda)^{-1},$$

генерирует последовательность случайных векторов W с ограниченным уровнем средней анизотропии $\overline{\mathbf{A}}(W) \leq a$.

Также такое множество формирующих фильтров обозначается следующим образом:

$$\mathcal{G}_a^{\diamond} = \arg \max_{G \in \mathcal{G}_a} \frac{\|FG\|_2}{\|G\|_2}.$$

2.3. ВРЕМЕННАЯ ОБЛАСТЬ

Рассмотрим стандартное представление динамического дискретного линейного объекта в пространстве состояний:

$$(8) \quad F : \begin{cases} x_{k+1} = Ax_k + Bw_k, \\ z_k = Cx_k + Dw_k, \end{cases}$$

где $x_k \in \mathbb{R}^n$ – состояние, $w_k \in \mathbb{R}^m$ – вход, $z_k \in \mathbb{R}^p$ – выход системы. Матрицы A, B, C, D с вещественными элементами известны и имеют соответствующую размерность. На матрицу A накладывается дополнительное условие устойчивости по Шуру. Полагаем, что формирующий фильтр G генерирует внешнее возмущение $W = GV$ из гауссовской белозумной последовательности V случайных векторов. Тогда внутренняя динамика формирующего фильтра может повторять динамику объекта (8) в совокупности со следующим уравнением:

$$w_k = Lx_k + \Sigma^{1/2}v_k,$$

где $\Sigma \in \mathbb{R}^{m \times m}$ – симметричная положительно определенная матрица, а матрица $L \in \mathbb{R}^{m \times n}$ выбрана из соображений, что выражение $A + BL$ обеспечивает асимптотическую устойчивость. Пользуясь перечисленными обозначениями, анизотропийную норму стационарной системы можно вычислить согласно следующему утверждению.

Теорема 2 [27]. *Для системы (8) с устойчивой матрицей A и внешним возмущением с ограниченным уровнем a средней анизотропии анизотропийная норма может быть вычислена согласно выражению*

$$\|F\|_a = \left(\frac{1}{q} \left(1 - \frac{m}{\text{tr}(LPL^\top + \Sigma)} \right) \right)^{1/2},$$

где $P \in \mathbb{R}^{n \times n}$ – решение уравнения Ляпунова

$$(9) \quad P = (A + BL)P(A + BL)^\top + B\Sigma B^\top,$$

и скалярный параметр q совместно с матрицами L, Σ удовлетворяют уравнению Риккати:

$$(10) \quad \begin{aligned} R &= A^\top RA + qC^\top C + L^\top \Sigma^{-1} L, \\ \Sigma &= (I_m - qD^\top D - B^\top RB)^{-1}, \\ L &= \Sigma(B^\top RA + qD^\top C) \end{aligned}$$

в совокупности с уравнением специального вида

$$(11) \quad a = -\frac{1}{2} \ln \det \left(\frac{m\Sigma}{\text{tr}(LPL^T + \Sigma)} \right).$$

Таким образом, независимо от способа описания стационарной системы – на языке передаточных функций (частотная область) или пространства состояний (временная область) – существуют точные формулы для вычисления анизотропийной нормы.

2.4. СУБООПТИМАЛЬНАЯ ЗАДАЧА

Несмотря на наличие точных формул для вычисления анизотропийной нормы, поиск единственного решения системы (10)–(11) из нелинейных матричных уравнений представляет собой достаточно трудную задачу. Для того чтобы получить эффективный инструмент численного решения задач в анизотропийной теории, была предложена идея заместить оптимальную постановку (поиск точного значения анизотропийной нормы (2)) субоптимальной (т.е. поиск верхней границы этой нормы). Методика перехода от системы равенств к системе неравенств описана в работе [5]. Позднее в статьях [29, 30] авторы развивают идею перехода к выпуклой задаче оптимизации для численного решения задач в анизотропийной постановке. Приведем утверждение, основанное на лемме о вещественной ограниченности для анизотропийной нормы системы с представлением (8).

Теорема 3 [7]. *Анизотропийная норма $\|F\|_a$ системы (8) строго ограничена пороговым значением γ , если существует число $\eta > \gamma^2$, такое что система неравенств*

$$(12) \quad \eta - (\exp(-2a) \det \Xi)^{1/m} < \gamma^2,$$

$$(13) \quad \begin{bmatrix} \Xi - \eta I_m & * & * \\ B & -\Theta^{-1} & * \\ D & 0 & -I_p \end{bmatrix} \prec 0,$$

$$(14) \quad \begin{bmatrix} -\Theta & * & * & * \\ 0 & -\eta I_m & * & * \\ A & B & -\Theta^{-1} & * \\ C & D & 0 & -I_p \end{bmatrix} \prec 0$$

имеет некоторое положительно определенное решение $\Xi \in \mathbb{R}^{m \times m}$ и $\Theta \in \mathbb{R}^{n \times n}$.

Относительно неизвестных матриц система (12)–(14) не является ни линейной, ни выпуклой, однако существуют приемы, включающие замену переменных и конгруэнтные преобразования, позволяющие в каждом конкретном случае успешно свести эти неравенства к виду линейных с выпуклым ограничением.

2.5. СЛУЧАЙ НЕЦЕНТРИРОВАННОГО ВОЗМУЩЕНИЯ

Преыдушие разделы посвящены базовым понятиям анизотропийной теории для стационарных систем при центрированном возмущении. В случае, когда внешнее возмущение имеет ненулевое матожидание, впервые рассмотрен в работах [20, 23], где определение и вычисление анизотропийной нормы адаптированы для указанного условия. В текущей работе накладываются дополнительные условия на первые два момента возмущающей последовательности $\{w_k\}$ в системе (8) следующего вида:

$$|\mathbf{E}W| \geq \tau, \quad \mathbf{E}(|W - \mathbf{E}W|^2) \leq \sigma, \quad (\sigma = 1 - \tau^2),$$

где τ и σ являются некоторыми положительными константами. В таком случае средняя анизотропия возмущения W может быть определена как

$$\begin{aligned} (15) \quad \bar{\mathbf{A}}_\tau(W) &= \lim_{N \rightarrow \infty} \frac{\mathbf{A}_\tau(W_{0:N-1})}{N} \\ &= \lim_{N \rightarrow \infty} \frac{\mathbf{A}(W_{0:N-1}) - \frac{mN}{2} \ln(1 - \tau^2)}{N} \\ &= \bar{\mathbf{A}}(W) - \frac{m}{2} \ln(1 - \tau^2), \end{aligned}$$

где $\bar{\mathbf{A}}(W)$ определяется согласно (1). Ограниченность среднего уровня анизотропии (15) следует из ограниченности $\bar{\mathbf{A}}(W) < b$. Рассмотрим следующее выражение:

$$\|F\|_{a,\tau} = \sqrt{\|F\|_{b,0}^2 (1 - \tau^2) + \|F\|_\infty^2 \tau^2},$$

где $\|F\|_{b,0} = \|F\|_a$ при условии $b = a$. Выражение $\|F\|_{a,\tau}$ будем называть (a, τ) -анизотропийной нормой системы F . Задача построения управления для линейной нестационарной системы

с нецентрированным возмущением разобрана в работе [20], откуда взяты некоторые используемые обозначения.

Специальные функции (3) и (4) изменяются следующим образом:

$$(16) \quad \mathcal{A}_\tau(q) = \mathcal{A}_0(q) - \frac{m}{2} \ln(1 - \tau^2) = \\ = \frac{m}{2} \left(\ln \left(\frac{\Phi(q)}{1 - \tau^2} \right) - \Psi(q) \right),$$

$$(17) \quad \mathcal{N}_\tau(q) = \left((\mathcal{N}_0(q))^2 (1 - \tau^2) + \|F\|_\infty^2 \tau^2 \right)^{1/2} = \\ = \left(\frac{\Phi(q) - 1}{q\Phi(q)} (1 - \tau^2) + \|F\|_\infty^2 \tau^2 \right)^{1/2}.$$

В формулах (16)–(17) вспомогательные функции $\Phi(q)$ и $\Psi(q)$ определяются согласно (5) и (6).

При использовании выражений (16) и (17) анизотропийная норма системы F , выраженная через специальные функции (7), изменяется следующим образом:

$$\|F\|_a = \mathcal{N}(\mathcal{A}^{-1}(b)),$$

где $q = \mathcal{A}^{-1}(b)$, $b = a + \frac{m}{2} \ln(1 - \tau^2)$, что детально описано в работе [20].

3. Постановка задачи

В этом разделе рассмотрим задачу построения управления для линейной дискретной стационарной системы с нецентрированным возмущением в рамках анизотропийной теории.

Задача 1. Для объекта управления, динамика которого описывается в пространстве состояний следующим образом:

$$(18) \quad \mathcal{F} : \begin{cases} x_{k+1} = Ax_k + B_w w_k + B_u u_k, \\ z_k = C_z x_k + D_{zw} w_k + D_{zu} u_k, \\ y_k = C_y x_k + D_{yw} w_k, \end{cases}$$

где $x_k \in \mathbb{R}^n$ – внутренне состояние, $w_k \in \mathbb{R}^m$ – внешнее возмущение с ограничением на первый момент $|\mathbf{E}w_k| < \tau$ и средний

уровень анизотропии $a \geq 0$, $u_k \in \mathbb{R}^q$ – управляющее воздействие, $z_k \in \mathbb{R}^p$ – управляемый выход, $y_k \in \mathbb{R}^r$ – наблюдаемый выход, требуется найти представление динамического регулятора в пространстве состояний

$$(19) \quad \mathcal{F}_c : \begin{cases} \xi_{k+1} = A_c \xi_k + B_c y_k, \\ u_k = C_c \xi_k + D_c y_k, \end{cases}$$

такого, чтобы анизотропийная норма $\|\mathcal{F}_c\|_a$ замкнутой управлением системы (18) была бы ограничена минимально возможным пороговым значением $\gamma = \gamma_{min}$. В системе (19) вектор $\xi_k \in \mathbb{R}^n$ соответствует внутреннему состоянию регулятора.

4. Решение задачи

Сформулируем в виде утверждения решение поставленной задачи 1.

Теорема 4. Рассмотрим линейную дискретную стационарную систему $\bar{\mathcal{F}}$ в пространстве состояний (18), на которую действует внешнее возмущение в виде гауссовской последовательности случайных векторов с ограниченным уровнем средней анизотропии числом a при дополнительном условии $|\mathbf{E}w_k| < \tau$, параметры a и τ считаются фиксированными. Тогда если выпуклая задача оптимизации

$$\gamma^2 \rightarrow \min$$

при ограничениях

$$(20) \quad \left[\begin{array}{cccc} -\Pi_{11} & * & & \\ -I_n & -\Phi_{11} & & \\ 0_{m \times n} & 0_{m \times n} & & \\ \text{АП}_{11} + B_u C & A + B_u D C_y & \dots & \\ \mathbf{A} & \Phi_{11} A + B C_y & & \\ C_z \Pi_{11} + D_{zu} C & C_z + D_{zu} D C_y & & \\ * & * & * & * \\ * & * & * & * \\ \dots & -\eta_i I_m & * & * \\ B_w + B_u D D_{yw} & -\Pi_{11} & * & * \\ \Phi_{11} B_w + B D_{yw} & -I_n & -\Phi_{11} & * \\ D_{zw} + D_{zu} D D_{yw} & 0_{p \times n} & 0_{p \times n} & -I_p \end{array} \right] \prec 0,$$

$$(21) \quad \begin{bmatrix} -\Psi_i - \eta_i I_m & * & * & * \\ B_w + B_u \mathbf{D} D_{yw} & -\Pi_{11} & * & * \\ \Phi_{11} B_w + \mathbf{B} D_{yw} & -I_n & -\Phi_{11} & * \\ D_{zw} + D_{zu} \mathbf{D} D_{yw} & 0_{p \times n} & 0_{p \times n} & -I_p \end{bmatrix} \prec 0, \quad i = \overline{1, 2},$$

$$(22) \quad \begin{bmatrix} \Pi_{11} & I_n \\ I_n & \Phi_{11} \end{bmatrix} \succ 0, \quad \eta > \gamma_1^2,$$

$$(23) \quad \eta - \exp(-2b/m) (\det \Psi_1)^{1/m} < \gamma_1^2,$$

$$(24) \quad \gamma_1^2 (1 - \tau^2) + \gamma_2^2 \tau^2 < \gamma_2^2,$$

где $b = a + \frac{m}{2} \ln(1 - \tau^2)$, $\eta_1 = \eta$, $\eta_2 = \gamma_2^2$, имеет решение $\mathbf{A} \in \mathbb{R}^{n \times n}$, $\mathbf{B} \in \mathbb{R}^{n \times r}$, $\mathbf{C} \in \mathbb{R}^{q \times n}$, $\mathbf{D} \in \mathbb{R}^{q \times r}$, $\Pi_{11} \succ 0$, $\Phi_{11} \succ 0$, $\Psi_1 \succ 0$, $\Psi_2 \succ 0$, $\eta > 0$, $\gamma_1 > 0$, $\gamma_2 > 0$, то для системы (18) матрицы динамического регулятора (19) имеют вид:

$$\begin{aligned} A_c &= \Phi_{12}^{-1} (\mathbf{A} - \Phi_{12} \mathbf{B} C_y \Pi_{11} - \Phi_{11} B_u \mathbf{C} \Pi_{12}^\top - \\ &\quad - \Phi_{12} (\mathbf{A} + B_u \mathbf{D} C_y) \Pi_{11}) \Pi_{12}^{-\top}, \\ B_c &= \Phi_{12}^{-1} (\mathbf{B} - \Phi_{11} B_u \mathbf{D}), \\ C_c &= (\mathbf{C} - \mathbf{D} C_y \Pi_{11}) \Pi_{12}^{-\top}, \\ D_c &= \mathbf{D}, \end{aligned}$$

где $\Pi_{12} \Phi_{12}^\top = I_n - \Pi_{11} \Phi_{11}$.

Доказательство. Сформируем замкнутую систему F_{cl} на основе объекта (18) и регулятора (19):

$$(25) \quad \begin{aligned} \zeta(k+1) &= \mathcal{A} \zeta(k) + \mathcal{B} w(k), \\ z(k) &= \mathcal{C} \zeta(k) + \mathcal{D} w(k), \end{aligned}$$

где $\zeta(k) \in \mathbb{R}^{2n}$ обозначает объединенный вектор состояния:

$$\zeta(k) = \begin{bmatrix} x(k) \\ x_c(k) \end{bmatrix},$$

матрицы системы $\mathcal{A} \in \mathbb{R}^{2n \times 2n}$, $\mathcal{B} \in \mathbb{R}^{2n \times m}$, $\mathcal{C} \in \mathbb{R}^{p \times 2n}$, $\mathcal{D} \in \mathbb{R}^{p \times m}$ имеют следующий вид:

$$(26) \quad \begin{aligned} \mathcal{A} &= \begin{bmatrix} \mathbf{A} + B_u D_c C_y & B_u C_c \\ B_c C_y & A_c \end{bmatrix}, \quad \mathcal{B} = \begin{bmatrix} B_u D_c D_{yw} \\ B_c D_{yw} \end{bmatrix}, \\ \mathcal{C} &= [C_1 + D_{zu} D_c C_y \quad D_{zu} C_c], \quad \mathcal{D} = D_{zu} D_c D_{yw}. \end{aligned}$$

Для замкнутой системы (25) можно сформулировать утверждение об ограниченности анизотропийной нормы замкнутой системы значением γ_1 в виде задачи выпуклой оптимизации, в основе которой лежит линеаризующая замена, предложенная в работах [15, 24]. Эта задача будет состоять в виде неравенств (20)–(23).

Так как внешнее возмущение нецентрировано, (a, τ) -анизотропийная норма системы \mathcal{F} удовлетворяет условию $\|\mathcal{F}\|_{a,\tau} \leq \gamma$, если существуют параметры γ_1 и γ_2 , такие что

$$\gamma_1^2(1 - \tau^2) + \gamma_2^2\tau^2 \leq \gamma^2,$$

$$\|\mathcal{F}\|_{b,0} \leq \gamma_1, \quad \|\mathcal{F}\|_{\infty} \leq \gamma_2,$$

где $b = a + \frac{m}{2} \ln(1 - \tau^2)$. Поскольку функция (17) ограничена величиной γ для всех q , при которых $\mathcal{A}_{\tau}(q) \leq a$, всегда найдутся такие γ_1 и γ_2 , при которых

$$\mathcal{N}_0(q) \leq \gamma_1, \quad \|F\|_{\infty} \leq \gamma_2,$$

что как раз эквивалентно условию ограниченности анизотропийной нормы, что указано в работе [20]. Условие ограниченности анизотропийной нормы $\|\mathcal{F}\|_{b,0} \leq \gamma_1$ в терминах матричных неравенств сформулировано в [29]. Неравенство (12), учитывающее связь среднего уровня анизотропийной нормы с ограничением на анизотропийную норму в случае ненулевого среднего примет, соответственно, следующий вид:

$$\eta - \exp(-2b/m)(\det \Psi_1)^{1/m} < \gamma_1^2.$$

5. Численное моделирование

Данный раздел содержит результаты численного моделирования вычисления анизотропийной нормы и сравнение полученной нормы с \mathcal{H}_{∞} -нормой той же системы.

Пример 1. Данные взяты из работы [31]. Объект управления представляет собой линеаризованную модель биомеханической руки. Матрицы объекта в пространстве состояний (18) имеют вид:

$$A = \begin{bmatrix} 1 & \Delta & 0 & 0 \\ 0 & 1 & \Delta/m & 0 \\ 0 & 0 & 1 - \Delta/\delta_2 & \Delta/\delta_2 \\ 0 & 0 & 0 & 1 - \Delta/\delta_1 \end{bmatrix}, B_w = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix},$$

$$B_u = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \Delta/\delta_1 \end{bmatrix}, C_z = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, D_{zw} = I, D_{zu} = 0,$$

$$C_y = C_z, D_{yw} = 0, D_{yu} = 0,$$

где $m = 1$, $\Delta = 0,01$, $\delta_1 = \delta_2 = 0,04$. Цель моделирования состоит в поиске пространственной реализации γ -субоптимального динамического регулятора (19) в зависимости от различных значений уровня средней анизотропии a и параметра τ , участвующего в ограничении матожидания возмущения. Численно задача была решена с помощью пакетов Yalmip Matlab toolbox и SeDuMi optimization package [26]. Сравнительные результаты сведены в таблицу 1, где в каждой клетке дано отношение анизотропийной нормы $\|F_{cl}\|_a$ и \mathcal{H}_∞ -нормы $\|F_{cl}\|_\infty$.

В первом столбце таблицы 1 приведены разные уровни средней анизотропии a , используемые при моделировании, в первой строке – параметр τ нецентрированности внешнего возмущения. В каждой клетке таблицы приведено отношение анизотропийной нормы $\|F_{cl}\|_a$ системы к соответствующему значению $\|F_{cl}\|_\infty$ -нормы этой же системы. Из приведенных соотношений можно заметить, что при слабо окрашенных возмущениях (т.е. небольшим значением параметра a , ограничивающего уровень средней анизотропии) преимущество системы управления на основе анизотропийного подхода с точки зрения выбранного показателя качества может достигать 25%. В то же время при увеличении параметров a и τ анизотропийная норма стремится к \mathcal{H}_∞ -норме,

что при наличии априорной информации о внешнем возмущении может служить критерием выбора закона управления, поскольку реализация \mathcal{H}_∞ -подхода при формировании управления требует меньше вычислительных ресурсов.

Таблица 1. Отношение норм $\|F_{cl}\|_a / \|F_{cl}\|_\infty$

| $a \setminus \tau$ | 0,01 | 0,05 | 0,1 | 0,15 |
|--------------------|-------|-------|-------|-------|
| 0,01 | 0,730 | 0,730 | 0,834 | 0,926 |
| 0,1 | 0,755 | 0,844 | 0,943 | 0,942 |
| 1 | 0,761 | 0,872 | 0,967 | 0,956 |
| 10 | 0,863 | 0,952 | 0,986 | 0,986 |

6. Выводы

Применение анизотропного подхода к решению задач подавления влияния внешних возмущений имеет как свои плюсы в виде более гибкого механизма учета информации о внешнем возмущении, так и минусы в виде более сложного численного решения по сравнению со стандартными задачами \mathcal{H}_2 - и \mathcal{H}_∞ -оптимальных теорий.

Литература

1. БАЛАНДИН Д.В., БИРЮКОВ Р.С., КОГАН М.М. *Многокритериальная оптимизация индуцированных норм линейных операторов: прямая и двойственная задачи управления и фильтрации* // Известия РАН. Теория и системы управления. – 2022. – №2. – С. 43–57.
2. БЕЛОВ И.Р., КУСТОВ А.Ю. *О применении фильтра Калмана в задаче оценивания при слабо окрашенных входных шумах* // Управление большими системами. – 2023. – Вып. 103. – С. 94–120.

3. ВЛАДИМИРОВ И.Г., КУРДЮКОВ А.П., СЕМЕНОВ А.В. *Анизотропия сигналов и энтропия линейных стационарных систем* // Доклады РАН. – 1995. – Т. 342, №5. – С. 583–585.
4. КУСТОВ А.Ю. *Параметризация оптимальных анизотропийных регуляторов* // Автоматика и телемеханика. – 2023. – №10. – С. 59–71.
5. ЧАЙКОВСКИЙ М.М. *Нахождение сильно минимизирующего ранг решения линейного матричного неравенства* // Автоматика и телемеханика. – 2007. – №9. – С. 96–105.
6. ЧАЙКОВСКИЙ М.М., КУРДЮКОВ А.П. *Критерий строгой ограниченности анизотропийной нормы заданным значением в терминах матричных неравенств* // Доклады Академии наук. – 2011. – Т. – 441, №3. – С. 318–321.
7. ЧАЙКОВСКИЙ М.М. *Синтез субоптимальных регуляторов методами выпуклой оптимизации и полуопределенного программирования* // Управление большими системами. – 2015. – Вып. 42. – С. 100–152.
8. ЧАЙКОВСКИЙ М.М., ТИМИН В.Н., КУРДЮКОВ А.П. *Синтез анизотропийного субоптимального пид регулятора для дискретной линейной стационарной системы: одномерный случай* // Автоматика и телемеханика. – 2019. – Вып. 9. – С. 156–172.
9. ЮРЧЕНКОВ А.В., КУСТОВ А.Ю., КУРДЮКОВ А.П. *Условия ограниченности анизотропийной нормы системы с мультипликативными шумами* // Доклады Академии наук. – 2016. – Т. 467, №4. – С. 396–399.
10. ЮРЧЕНКОВ А.В., БЕЛОВ И.Р. *Лемма об ограниченности анизотропийной нормы стационарной системы с мультипликативными шумами* // Дифференциальные уравнения. – 2023. – Т. 59, №11. – С. 1550–1560.
11. BELOV I.R. *On the Approximation of Anisotropic Controller by \mathcal{H}_2 -Optimal Controller* // Proc. of the 32th Mediterranean Conference on Control and Automation. – 2024. – Vol. 9. – P. 891–895.

12. BOICHENKOV V.A., BELOV A.A., ANDRIANOVA O.G. *Axiomatic Foundations of Anisotropy-Based and Spectral Entropy Analysis: A Comparative Study* // Mathematics. – 2023. – Vol. 11, No. 12. – P. 2751 (1–10).
13. DIAMOND P., VLADIMIROV I., KURDYUKOV A. et al. *Anisotropy-based performance analysis of linear discrete time invariant control systems* // Int. Journal of Control. – 2001. – Vol. 74, No. 1. – P. 28–42.
14. DOYLE J.C., GLOVER K., KHARGONEKAR P.P. et al. *State-space solution to standard \mathcal{H}_2 and \mathcal{H}_∞ control problems* // IEEE Trans. Automat. Contr. – 1989. – Vol. 34. – P. 831–846.
15. GAHINET P. *Explicit controller formulas for LMI-based \mathcal{H}_∞ synthesis* // Automatica. – 1996. – Vol. 32. – P. 1007–1014.
16. GU D.-W., TSAI M.C., O'YOUNG S.D. et al. *State-space formulae for discrete-time \mathcal{H}_∞ optimization* // Int. J. Contr. – 1989. – Vol. 49. – P. 1683–1723.
17. HADDAD W.M., BERNSTEIN D.S., MUSTAFA D. *Mixed $\mathcal{H}_2/\mathcal{H}_\infty$ regulation and estimation: The discrete time case* // Syst. Control Lett. – 1991. – Vol. 16. – P. 235–247.
18. KHARGONEKAR P.P., ROTEA M.A. *Mixed $\mathcal{H}_2/\mathcal{H}_\infty$ control: a convex optimization approach* // IEEE Trans. Automat. Contr. – 2011. – Vol. 36. – P. 824–837.
19. KUSTOV A.YU., KURDYUKOV A.P., YURCHENKOV A.V. *On the Anisotropy-Based Bounded Real Lemma Formulation for the Systems with Disturbance-Term Multiplicative Noise* // Proc. of the 12th IFAC Int. Workshop on Adaptation and Learning in Control and Signal Processing. – 2016. – P. 1–5.
20. KUSTOV A.YU., TIMIN V.N. *Suboptimal Anisotropy-based Control for Linear Discrete Time Varying Systems with Noncentered Disturbances* // IFAC-PapersOnLine. – 2017. – Vol. 50, Iss. 1. – P. 6122–6127.
21. KUSTOV A.YU. *State-Space Formulas for Anisotropic Norm of Linear Discrete Time Varying Stochastic System* // Proc. of the 15th Int. Conf. on Electrical Engineering, Computing Science and Automatic Control. – 2018. – P. 1–6.

22. KURDYUKOV A.P., MAXIMOV E.A. TCHAIKOVSKY M.M. *Anisotropy-Based Bounded Real Lemma* // Proc. of the 19th Int. Symposium on Mathematical Theory of Networks and Systems. – 2010. – P. 2391–2397.
23. KURDYUKOV A.P., YURCHENKOV A.V., KUSTOV A.YU. *Robust Stability in Anisotropy-Based Theory with Non-Zero Mean of Input Sequence* // Proc. of the 21st Int. Symposium on Mathematical Theory of Networks and Systems. – 2014. – P. 208–214.
24. SCHERER C.W., GAHINET P., CHILALI M. *Multiobjective output-feedback control via LMI optimization* // IEEE Trans. on Automatic Control. – 1997. – Vol 42. – P. 896–911.
25. SEMYONOV A.V., VLADIMIROV I.G., KURDJUKOV A.P. *Stochastic approach to \mathcal{H}_∞ -optimization* // Proc. of the 33rd Conf. on Decision and Control. – 1994. – Vol. 3. – P. 2249–2250.
26. STURM J.F. *Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones* // Optimization Methods and Software. – 1999. – Vol. 11–12. – P. 625–653.
27. VLADIMIROV I.G., KURDYUKOV A.P., SEMENOV A.V. *On computing the anisotropic norm of linear discrete-time-invariant systems* // Proc. of the 13 IFAC World Congr. – 1996. – Paper IFAC–2d–01.6. – H. – P. 179–184.
28. VLADIMIROV I.G., KURDYUKOV A.P., SEMENOV A.V. *State-space solution to anisotropy-based stochastic \mathcal{H}_∞ -optimization problem* // Proc. of the 13 IFAC World Congr. – 1996. – Paper IFAC–3d–01.6. – H. – P. 427–432.
29. TCHAIKOVSKY M.M., KURDYUKOV A.P., TIMIN V.N. *Strict anisotropic norm bounded real lemma in terms of inequalities* // Proc. of the 18th IFAC World Congr. – 2011. – P. 2332–2337.
30. TCHAIKOVSKY M.M. *Static Output Feedback Anisotropic Controller Design by LMI-based Approach: General and Special Cases* // Proc. of the American Control Conf. ACC. – 2012. – P. 5208–5213.

31. WEIWEI L., TODOROV E., SKELTON R.E. *Estimation and Control Systems with Multiplicative Noise via Linear Matrix Inequalities* // Am. Contr. Conf. – 2005. – P. 1811–1816.
32. YURCHENKOV A.V., KUSTOV A.YU., TIMIN V.N. *The sensor network estimation with dropouts: Anisotropy-based approach* // Automatica. – 2023. – Vol. 151. – P. 110924 (1–8).
33. ZHOU K., GLOVER K., BODENHEIMER B.A. et al. *Mixed \mathcal{H}_2 and \mathcal{H}_∞ performance objectives I: Robust performance analysis, II: Optimal control* // IEEE Trans. Automat. Contr. – 1994. – Vol. –39. – P. 1564–1587.

ANISOTROPY-BASED CONTROL DESIGN FOR LINEAR TIME-INVARIANT SYSTEMS WITH MOMENTS CONSTRAINTS OF DISTURBANCES

Alexander Yurchenkov, V.A. Trapeznikov Institute of Control Sciences of RAS, Cand.Sc., Senior Researcher (alexander.yurchenkov@yandex.ru).

Abstract: In this paper, a linear discrete time-invariant system with control under the influence of a colored disturbance is considered. The external disturbance is selected from the class of non-centered stationary Gaussian sequences of random vectors with a known restriction on the level of mean anisotropy. For the specified class of control objects, a dynamic regulator is introduced, with the help of which it is necessary to ensure the boundedness of the anisotropic norm from an external disturbance to the controlled output of a closed-loop system. The control design problem is to construct an anisotropy-based dynamic regulator in terms of state-space representation. The boundedness of the closed-loop system is provided by anisotropy-based small gain theorem. Using linearizing reversible variable change, the problem can be reduced to a numerical solution of the convex optimization problem with special constraints characteristic of anisotropy-based theory. In the formulation of the problem, it is assumed that the expectation of the external disturbance is unknown, but a condition on it in the form of inequality is known. This parameter causes an additional constraint to appear in the convex optimization problem. The resulting system of inequalities is linear matrix inequalities in combination with an inequality of a special type, which is nonlinear with respect to unknown parameters, but at the same time convex in these parameters. The problem of finding the regulator matrices can be solved by standard methods.

Keywords: anisotropy-based theory, convex optimization, noncentered disturbance.

УДК 62-5

ББК 30в6

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

Поступила в редакцию 22.08.2024.

Дата опубликования 30.11.2024.

ПРОЕКТ СИСТЕМЫ РАСПРЕДЕЛЕНИЯ ЗАКАЗА ВНУТРИ АССОЦИАЦИИ ПРОИЗВОДИТЕЛЕЙ ПОСРЕДСТВОМ МНОЖЕСТВЕННЫХ АУКЦИОНОВ

Гасанов И. И.¹
(ФИЦ ИУ РАН, Москва)

Описывается проект системы, заказанной компанией, которая занимается координацией совместного сборочного производства изделий из комплектующих, в изготовлении которых занято большое количество независимых предприятий. Всё производство конечных изделий может быть представлено в виде направленного ациклического графа, в котором для сборки деталей более низкого ранга используются детали, имеющие более высокий ранг. Координация такого производства среди большого множества независимых производителей требует от головного предприятия значительных усилий по согласованию планов, учёта интересов участников, постоянного мониторинга их возможностей. При этом добиться оптимального решения, минимизации затрат не представляется возможным. Была сформулирована задача разработать альтернативный подход к формированию производственного плана, использующий современные идеи распределённых решений на основе сетевых принципов, сходные с организацией взаимодействия внутри блокчейна, когда функционирование системы в целом определяется индивидуальными интересами и действиями её участников. Автором была предложена система взаимодействия внутри ассоциации производителей на основе множественных локальных аукционов, на которых торгуются предложения участников по производству комплектующих, необходимых для выполнения общего заказа. Такой подход имеет свои преимущества, но несёт в себе и определённые риски. В статье даётся общий обзор аукционной схемы и обсуждаются трудности, связанные с её реализацией.

Ключевые слова: аукцион, сетевое взаимодействие, планирование производства.

1. Введение

В настоящее время наблюдается большой интерес к созданию систем с сетевым принципом управления. Это структуры, которые полностью или частично состоят из элементов, обладающих определённой автономией и модифицирующих своё состояние в результате локального взаимодействия с другими

¹ Игорь Искендерович Гасанов, с.н.с. (gasanov48@yandex.ru).

элементами системы. Материальный и информационный обмен между элементами системы интегрируется в отклик системы на внешние воздействия. В результате такого взаимодействия система в целом эволюционирует в полезном её создателям – природе или человеку – направлении.

Сама подобная архитектура сложных систем не является новым изобретением. За ней стоит некоторый фундаментальный принцип организации живой природы, начиная с колоний простейших до сообщества клеток, образующих организмы высших животных. В настоящее время тот же принцип активно внедряется в сферу информационных технологий. Два наиболее ярких примера – это нейросети [1, 4, 6] и блокчейны [5, 8]. В основу данной работы положена переработанная и дополненная статья автора [1], которая в дальнейшем была процитирована авторами статьи [3]. В настоящей работе описывается проект организации взаимодействия между предприятиями, образующими ассоциацию, которая выполняет заказы по производству и сборке некоторого фиксированного набора изделий. Сетевой принцип взаимодействия предлагается использовать на этапе выработки плана исполнения очередного заказа и его распределения внутри ассоциации. Предполагается, что такая форма взаимодействия между участниками ассоциации может быть реализована на базе ещё одной конструкции, активно исследуемой в экономической науке и используемой на практике, аукциона [7, 9, 10].

2. Модель ассоциации

Рассматривается ассоциация предприятий, занимающихся производством и сборкой некоторого фиксированного набора изделий $\widehat{M} = \{m\}_{m=1}^M$, которые будем обозначать просто цифрами. Предприятия, входящие в ассоциацию, будем называть участниками. Обозначим их множество через $\widehat{L} = \{l\}_{l=0}^L$.

Участники – самостоятельные предприятия, производственная деятельность которых не ограничена участием в ассоциации. Участие в ассоциации предоставляет предприятиям дополнительные возможности расширить и упорядочить рынок сбыта продукции. Ассоциация заключает контракты с внешни-

ми заказчиками и выполняет сборку изделий из набора \widehat{M} . Члены ассоциации предлагают свои услуги и принимают на себя обязательства по совместному исполнению этих контрактов.

Будем придерживаться следующих обозначений.

Векторы будем обозначать большими буквами с верхней крышкой, а их компоненты – строчными буквами с индексами, т.е. по шаблону $\overline{R} = (r_1, r_2, \dots, r_N)$.

В процессе выполнения заказа участники осуществляют взаимопоставки своих изделий. Одни изделия используются как комплектующие, необходимые для сборки других изделий. Всё множество поставок по заказу описывается матрицей $Y = (y_n^{i,j})$. У её элементов $y_n^{i,j}$ нижний индекс будет обозначать изделие, первый верхний индекс – поставщика, а второй – получателя изделия.

Обозначим через $\Omega \subset \mathbb{R}^M$ множество векторов продукции $\overline{W} = (w_1, w_2, \dots, w_M)$, в которых компоненты w_m соответствуют количеству (объёму) изделий $m \in \widehat{M}$. Через \overline{w}_m будем записывать вектор $\overline{W} \in \Omega$ с единственной ненулевой компонентой m , равной w_m . Вектор \overline{w}_m , у которого компонента $w_m = 1$, будем обозначать \overline{e}_m . Если $\overline{W} \in \Omega$, то набор изделий, соответствующий его ненулевым компонентам, будем обозначать W .

Часть изделий $m \in \widehat{M}$ собирается из других изделий, производимых внутри или вне ассоциации. Изделию m сопоставим множество $G(m) \subset \widehat{M}$ тех деталей, из которых оно монтируется, и которые производятся внутри ассоциации. Через $\overline{G}(\overline{w}_m) \in \Omega$ обозначим вектор объёмов изделий $G(m)$, необходимых для сборки объёма w_m . В $\overline{G}(\overline{w}_m)$ входят только детали финальной сборки изделия m : если деталь $k \in G(m)$ сама является сборной, то объёмы комплектующих $\overline{G}(\overline{w}_k)$ не входят в компоненты вектора $\overline{G}(\overline{w}_m)$.

Для множества изделий $W \subset \widehat{M}$ через $G(W) \subset \widehat{M}$ обозначим множество тех деталей, которые необходимы для монтажа

изделий из W : $G(W) = \bigcup_{m \in W} G(m)$. Отметим, что в множество $G(W)$ может входить и часть изделий из W , так как одни изделия из набора W могут служить комплектующими для других. Через $\bar{G}(\bar{W}) \subset \Omega$ будем обозначать вектор объёмов комплектующих $G(W)$, необходимых для сборки вектора \bar{W} , т.е. $\bar{G}(\bar{W}) = \sum_{m \in W} \bar{G}(\bar{w}_m)$.

Определение 1. Назовём множество $G(W)$ проекцией множества W , а вектор $\bar{G}(\bar{W})$ – проекцией вектора \bar{W} .

Опишем структуру производства, для которой предназначена, рассматриваемая в статье схема.

В множестве \hat{M} выделяем подмножество изделий \hat{M}^1 . Это те изделия, которые ассоциация производит по внешним заказам или на продажу и которые не используются как комплектующие для производства других изделий из \hat{M} . Этим изделиям припишем ранг 1, $\hat{M}^1 \cap G(\hat{M}^1) = \emptyset$.

Изделия первого ранга монтируется из деталей, составляющих множество $G(\hat{M}^1)$. Припишем ранг 2, таким изделиям из множества $G(\hat{M}^1)$, которые не имеют ранг 1 и используются как комплектующие только для изделий ранга 1. Множество таких изделий обозначим \hat{M}^2 , т.е. $t \in \hat{M}^2$, если $t \in G(\hat{M}^1) \& t \notin G(\hat{M} \setminus \hat{M}^1)$.

Для сборки изделий из множества \hat{M}^2 , в свою очередь, могут понадобиться комплектующие. Припишем ранг 3 таким изделиям из множества $G(\hat{M}^2)$, которые используются в качестве комплектующих только для изделий рангов 1 и 2. Их множество обозначим \hat{M}^3 . В силу определения рангов 1 и 2, изделия из \hat{M}^3 не входят в множества \hat{M}^1 и \hat{M}^2 . Таким образом, $t \in \hat{M}^3$, если $t \in G(\hat{M}^2) \& t \notin G(\hat{M} \setminus (\hat{M}^1 \cup \hat{M}^2))$.

И так далее, пока очередное множество $G(\hat{M}^r)$ не окажется пустым. Понятно, что всего таким образом будет выделено некоторое ограниченное число рангов.

2.1. ГРАФИЧЕСКОЕ ПРЕДСТАВЛЕНИЕ ПРОИЗВОДСТВА

Описанную выше структуру производства можно изобразить в виде направленного ациклического графа. Назовём его технологическим графом и обозначим Γ . Узлы графа Γ соответствуют изделиям $t \in \widehat{M}$, обозначим их теми же буквами. Дуги, исходящие из узла $t \in \widehat{M}$ графа Γ , направлены к узлам, сопоставленным изделиям, для сборки которых используются комплектующие t . На нижнем уровне графа Γ находятся узлы, соответствующие изделиям уровня \widehat{M}^1 .

Пример технологического графа приводится на рис. 1. Узлы 1 и 2 – это корень графа, т.е. множество \widehat{M}^1 . К множеству \widehat{M}^2 относятся узлы 3–7 и 13; к множеству \widehat{M}^3 – узлы 8–10 и 11; узлы 12, 13 – к множеству \widehat{M}^4 ; узел 14 – к множеству \widehat{M}^5 . Концевые узлы – это 6, 9, 13, 14.

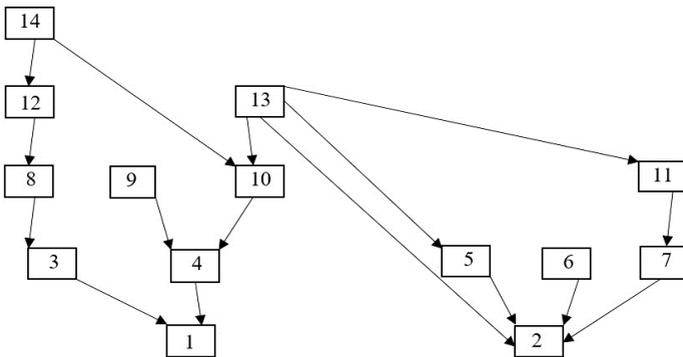


Рис. 1.

Если в графе Γ идёт дуга от узла α к узлу β , то α будем называть прямым предшественником β , а β – прямым потомком α .

Совокупность прямых предшественников множества узлов $W \subset \Gamma$ соответствует множеству $G(W)$, т.е. проекции W . Двигаясь по дугам от узла потомка к узлу предшественнику, мы неизбежно упрёмся в концевые узлы, которые разбросаны по разным множествам $G^n(W)$. Материалы, необходимые для изготовления

изделий, соответствующих этим узлам, закупаются участниками вне ассоциации.

Определение 2. Рангом изделия $m \in \widehat{M}$ назовём длину максимального пути (числа дуг) на технологическом графе от корня графа до узла m плюс 1.

Определение 3. Вес дуги $m' \mapsto m$ – это число деталей m' , которое используется при сборке одного изделия m .

Пусть $\bar{W} \in \Omega$. Компоненты вектора $\bar{G}(\bar{W})$ рассчитываются как сумма произведений компонент вектора \bar{W} на веса дуг:

$$g_k(\bar{W}) = \sum_{m \in M} w_m \cdot h_{m,k},$$

где $h_{m,k}$ – вес дуги $m \mapsto k$

Определение 4. Обозначим через $\bar{H}(\bar{W})$ и назовём полной проекцией вектора $\bar{W} \in \Omega$ вектор деталей $m \in \widehat{M}$ всех уровней, необходимых для сборки \bar{W} . Вектор $\bar{H}(\bar{W})$ – это сумма последовательных проекций $\bar{G}(\bar{W}) + \bar{G}(\bar{G}(\bar{W})) + \bar{G}(\bar{G}(\bar{G}(\bar{W}))) + \dots$

Полная проекция $\bar{H}(\bar{W}) = \sum_{n \in \widehat{M}} \bar{H}(\bar{w}_n)$. Компоненты вектора $\bar{H}(\bar{w}_n)$ нетрудно рассчитать, используя веса дуг на поддереве технологического графа с корнем в узле n .

2.2. УЧАСТНИКИ

Каждый из участников ассоциации производит своё подмножество изделий из набора \widehat{M} . Не вдаваясь в подробности, модель производства участника $l \in \widehat{L}$ можно описать так.

Текущие объёмы ресурсов, которые имеются в распоряжении предприятия l будем записывать в виде вектора $\bar{S}_l = (s'_1, s'_2, \dots, s'_p)$. Задана вектор-функция $\bar{\Psi}'(\bar{W}) : \mathbb{R}^m \rightarrow \mathbb{R}^p$, показывающая, сколько ресурсов каждого вида p расходует участник l при производстве вектора продукции \bar{W} .

Пусть компонента ψ'_1 обозначает финансы. Расход ресурса $\psi'_1(\bar{W})$ отражает финансовые издержки предприятия на производство вектора продукции \bar{W} . В $\psi'_1(\bar{W})$ не входит стоимость комплектующих, а только собственные траты участника на

сборку продукции. Предполагается, что стоимость деталей, поставляемых участнику l , войдёт в стоимость его изделий и вернётся поставщикам после выполнения заказа.

Осуществление поставок между предприятиями внутри ассоциации приводит к транспортным расходам. Предполагается, что у каждого предприятия l имеются таблицы

$$T^l = (t_m^{l,j}), m \in \widehat{M},$$

в которых $t_m^{l,j}$ – это цена доставки единицы продукции m от l к j . Эти цены участники добавляют к стоимости своей продукции.

У ассоциации есть выделенный участник – Центр, которому сопоставим номер $l=0$. В рамках рассматриваемой схемы функции Центра – это приём заказов от внешних клиентов, организация и координация аукционов, а также распределение заказов и выручки от их исполнения между изготовителями изделий ранга 1.

3. Задача распределения производства внутри ассоциации

Допустим, Центр принимает заказ на производство изделия $m \in \widehat{M}^1$ в объёме w_m . Производство заказанных изделий, включая изготовление комплектующих, распределяется между участниками ассоциации, и после выполнения заказа готовые изделия m поступают в Центр.

Для реализации заказа должны быть обеспечены поставки комплектующих между участниками, занятыми в его исполнении. Должны выполняться балансовые соотношения:

$$(1) \quad \forall l \in \widehat{L}: \bar{G}(\bar{W}^l) = \sum_{j \in \widehat{L}} \bar{Y}^{j,l},$$

и ресурсные ограничения

$$(2) \quad \forall l \in \widehat{L}: \bar{\Psi}^l(\bar{W}^l) \leq \bar{S}^l,$$

где \bar{W}^l – вектор изделий участника l , а $\bar{Y}^{j,l}$ – вектор поставок от предприятий j к l .

В данной работе не моделируются создание и использование запасов комплектующих. Предполагается, что для производства используются текущие ресурсы предприятий.

Так как ассоциация стремится опередить конкурентов в борьбе за заказы, то она заинтересована в снижении себестоимости своей продукции. Ставится задача минимизировать себестоимость:

$$(3) \sum_{l \in L} \Psi_l \left(\sum_{j \in L} \bar{Y}^{l,j} \right) + \sum_{m \in M} \sum_{l \in L} \sum_{j \in L} t_m^{l,j} \cdot y_m^{l,j} \rightarrow \min.$$

4. Обзор аукционной схемы

В настоящее время вычислительные мощности позволяют решать задачи линейного программирования очень большой размерности. Если модель из раздела 3 записана в линейных соотношениях, то можно ожидать, что трудности распределения заказа внутри ассоциации будут связаны не собственно с расчётами, а с созданием очень объёмной модели, объединяющей производство всех участников, и постоянным мониторингом состояния производства на каждом из составляющих ассоциацию предприятий. Для достаточно крупной ассоциации такой централизованный подход может оказаться слишком затратным, если вообще осуществимым. Поэтому возникает потребность в декомпозиции задачи, с тем чтобы большую часть функций по оценке производственных мощностей предприятий и разработке общего плана производства передать самим предприятиям. Идея состоит в том, чтобы план производства вырабатывался как результат множественных аукционов внутри коалиции, на которых предприятия самостоятельно выбирают поставщиков и потребителей продукции. Представляется, что у такой схемы есть ещё одно важное преимущество перед централизованным планированием. Она снимает возможные претензии участников к Центру при распределении заказов между предприятиями, подозрения, что это распределение осуществляется не вполне по экономическим соображениям. При аукционной схеме распределение заданий между участниками может проводиться согласно простым принципам и публично.

Сделаем краткое описание предлагаемой схемы, цель которого дать о ней общее представление и указать на некоторые проблемы, связанные с её реализацией.

Предположим, что Центру поступил заказ на производство некоторого объёма w^* изделия $m \in \tilde{M}^1$. Центру известны технологические графы для всех изделий, производимых внутри ассоциации. Он публикует полученный заказ вместе с его полной проекцией – вектором $\bar{H}(\bar{w}_m^*)$.

Предприятия $l \in \tilde{L}$, получив от Центра общий запрос на производство в виде полной проекции $\bar{H}(\bar{w}_m^*)$ и опираясь на текущие объёмы своих ресурсов \bar{S}^l , формируют собственные планы производства векторов изделий $\bar{W}^l \leq \bar{H}(\bar{w}_m^*)$. После этого они публикуют запросы на поставки комплектующих $\bar{G}(\bar{W}^l)$.

В ответ на запрос предприятия l другие члены ассоциации, исходя из собственных планов на участие в исполнении заказа, публикуют предложения по объёмам и ценам поставок запрошенных деталей. После этого участник l , с учётом объёмов предложенных ему поставок, корректирует исходный вектор \bar{W}^l . Именно эту форму взаимодействия между производителями будем называть аукционом. Это один из вариантов аукциона закрытого типа. Публикации предприятиями запросов на поставки комплектующих – суть объявление аукционов на эти поставки.

Аукцион на поставки комплектующей k для изделия n , производимый участником l , будем обозначать $A_{k,n}^l$.

Обозначим множество предприятий, производящих изделия из набора $W \subset \tilde{M}$, как $L(W)$.

Пусть n – некоторая ненулевая компонента вектора $\bar{H}(\bar{w}_m^*)$, такая что $G(n) \neq \emptyset$. На заказы изделий из множества $G(n)$ должны откликнуться участники $j \in L(G(n))$. Часть своих ресурсов они выделяют для производства комплектующих $G(n)$. Для планирования производства изделий n участнику $l \in L(n)$ необходимо знать, какими объёмами комплектующих $G(n)$ он будет располагать. Для этого необходимо, чтобы завершились аукционы по всем комплектующим $G(n)$. Поэтому аукционы следует проводить в определённом порядке. Этот порядок регулирует Центр, следуя структуре технологического графа.

Определение 5. Положим ранг любого аукциона $A_{k,n}^l$, $k \in G(n)$, $l \in L(n)$, равным рангу изделия n .

Условие 1. Аукционы, у которых ранг выше, всегда предпочитают аукционам с более низким рангом.

В соответствии с определением рангов, комплектующие $G(n)$ всегда имеют более высокий ранг, чем изделие n . Поэтому в силу условия 1 участник l , производящий изделия n , завершив аукционы $A_{k,n}^l$ по изделиям $k \in G(n)$, знает объёмы всех комплектующих, которыми он располагает для производства изделий n . Он может рассчитать доступный ему объём их производства и распределить его по аукционам следующих уровней $A_{n,k}^j$, $n \in G(k)$, $j \in L(k)$.

Замечание. Материалы, не входящие в набор изделий \widehat{M} , могут быть необходимы и закупаться участниками для производства планируемых изделий вне ассоциации. Вообще говоря, вне ассоциации участники могут закупать и какие-то комплектующие, производимые внутри ассоциации. Это схемы не меняет.

Когда нисходящая волна аукционов по комплектующим всех рангов завершается, Центр получает предложения по выпуску продукции t от предприятий, которые являются его конечными сборщиками. Если суммарный объём этих предложений не меньше заказа w^* , то Центр распределяет заказ между этими производителями, т.е. производителями изделия ранга 1. В свою очередь те, получив задание от Центра, распределяют задание на комплектующие ранга 2 между их производителями, согласно их предложениям на прошедших аукционах и т.д.

Проиллюстрируем предлагаемую схему на примере, используя технологический граф на рис. 1.

Пример 1. Предположим, что Центру поступил заказ на производство изделия $1 \in \widehat{M}^1$ в объёме w^* . После того как Центр рассылает участникам вектор комплектующих $\bar{H}(\bar{w}_1^*)$, производители изделий из множеств $L(10)$ и $L(12)$ проводят аукционы на поставки изделий 14. (Отметим, что в этом и других случаях множества L могут пересекаться.) Предложения поставщиков на аукционах – это объёмы и цены поставок с учётом транспорт-

ных расходов. Каждое из предложений ограничено объёмом, объявленным инициатором аукциона до его проведения. Если на каком-то из таких аукционов объёмы предложений в сумме превосходят тот объём, который необходим его организатору, то он делает выбор между поступившими предложениями, отдавая предпочтение более дешёвым вариантам. Собрав пул из предложений по поставкам комплектующих 14, участник из множества $L(12)$ определяет, сколько изделий 12 и по каким ценам он может произвести, не превышая объёма, объявленного им на аукционе. После этого он готов сформировать предложения о поставках комплектующих 12 на аукционах по изделиям 8.

Затем аналогичным образом проводят аукционы участники из множества $L(8)$. Они делают выбор между предложениями по изделиям 12 и формируют собственные предложения для участников из множества $L(3)$.

Одновременно с участниками из $L(8)$ аукционы по комплектующим 13 проводят предприятия из множества $L(10)$. К этому времени у них уже сформированы пулы комплектующих 14. После аукционов по комплектующим 13 они определяют объёмы и цены поставок изделий 10 предприятиям из $L(4)$.

После этого по той же схеме аукционы проводят предприятия из множества $L(3)$ по изделиям 8 и из множества $L(4)$ по изделиям 9 и 10.

Наконец, участниками из $L(1)$ проводятся аукционы по комплектующим 3 и 4. По результатам этих аукционов они рассчитывают, какой объём изделий 1 и по каким ценам они могут произвести, и объявляют результаты Центру. На этом заканчивается нисходящая волна аукционов.

Начинается заключительный этап вёрстки плана по выполнению заказа. Центр распределяет производство изделий 1 между участниками из множества $L(1)$. Назначенные объёмы не должны превосходить предложения участников. Участники из $L(1)$, в соответствии с назначенными объёмами производства, распределяют задание на поставки комплектующих 3 и 4 и т.д. Восходящая волна заданий доходит до концевых узлов, и на этом формирование плана производства завершается.

Если число участников велико и их ресурсы производства значительно превосходят те, что необходимы для исполнения

заказа, то схема представляется эффективной. Однако, как нетрудно убедиться, даже если допустимый план существует, его формирование посредством аукционной схемы не гарантировано.

Пример 2. Предположим, что в ассоциацию помимо Центра входят 4 предприятия (рис. 2). Участники l_1 и l_2 производят изделие 1. Для сборки одного изделия 1 требуются комплектующие 2 и 3, каждое в единичном объёме.

Изделия 2 и 3 производятся участниками l_3 и l_4 , причём из-за ресурсных ограничений каждый из них может изготовить либо одно изделие 2, либо одно изделие 3.

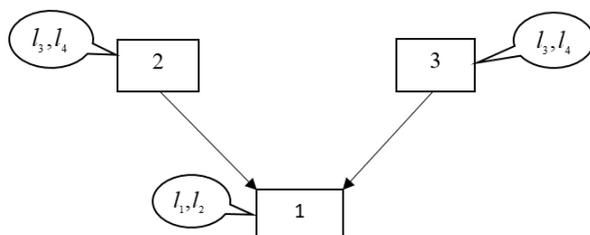


Рис. 2.

Ассоциации заказано изделие 1 в объёме 1. Участники l_1 и l_2 объявляют аукционы на поставку изделий 2 и 3. Участник l_3 решает изготовить комплектующую 2 в объёме 1 и предложить её как на аукционе $A_{2,1}^1$, так и на аукционе $A_{2,1}^2$. При этом он не рискует тем, что не сможет выполнить обязательства, так как согласно объёму заказа на сборку изделия 1 требуемые поставки изделия 2 не могут превысить объём 1. Если участник l_4 решает заняться производством изделия 3 и предложить его на аукционах $A_{3,1}^1$ и $A_{3,1}^2$, то план исполнения заказа может быть сформирован и изделие 1 может быть собрано или участником l_1 , или участником l_2 по выбору Центра.

Однако ничто в условиях примера не препятствует участнику l_4 , как и участнику l_3 , запланировать производство комплектующей 2 и предложить её на аукционы $A_{2,1}^1$ и $A_{2,1}^2$. Тогда

предприятия l_1, l_2 останутся без комплектующей 3 и формирование плана окажется невозможным.

5. Заключение.

Описанная схема представляется достаточно наглядной, но чтобы аукционные предложения участников по поставкам не оказались в противоречии с их исполнимостью, объёмы заказов и предложений на аукционах должны быть определённым образом регламентированы. От того, каких правил будут придерживаться участники при формировании аукционных заявок, зависят шансы на то, что производственный план будет сформирован. Этот и другие вопросы реализации аукционной схемы рассматриваются в следующей статье, в которой будет проведён подробный анализ предлагаемой схемы.

Литература

1. ГАСАНОВ И.И. *Организация аукционов в сетевых моделях* // Материалы конференции «Управление развитием крупномасштабных систем» (MLSD'2021). – М. ИПУ РАН, 2021. – С. 414 – 422.
2. ГУДФЕЛЛОУ Я., БЕНДЖИО И., КУРВИЛЛЬ А. *Глубокое обучение*. – М.: ДМК-пресс, 2018. – 652 с.
3. ЕРЕШКО Ф.И., МУШКОВ А.Ю., ТУРКО Н.И. и др. *Управление в крупномасштабных проектах многоукладной экономики*. // Автоматика и телемеханика. – 2022. – №5. – С. 102–132.
4. ЛЮГЕР ДЖ.Ф. *Искусственный интеллект: стратегии и методы решения сложных проблем*. – М.: Издательский дом «Вильямс», 2003. –864 с.
5. ПОППЕР Н. *Цифровое золото. Невероятная история Биткойна или как идеалисты и бизнесмены изобретают деньги заново*. – М., СПб, Киев: Диалектика, 2016. – 368 с.
6. ШОЛЛЕ Ф. *Глубокое обучение на Python*. – СПб.: Питер, 2018. – 400 с.
7. СОНИН Л.И. *Основы теории аукционов* // Вопросы экономики. – 2021. – №1. – С. 5-32.

8. АНТОНОПУЛОС А.М. *Осваиваем биткойн*. – М.: ДМК Пресс 201. – 428 с.
9. KLEMPERER P. *Auction Theory: A Guide to the Literature* // Journal of Economic Surveys. – 1999. – Vol. 13(3). – P. 227–286.
10. KRISHNA V. *Auction theory*. – New York: Elsevier, 2002. – 303 p.

THE PROJECT OF THE ORDER DISTRIBUTION SYSTEM WITHIN THE ASSOCIATION OF MANUFACTURERS THROUGH MULTIPLE AUCTIONS

Igor Gasanov, FRC CSC RAS, Moscow, Senior Researcher (gasanov48@yandex.ru).

Abstract: The paper describes the design of a system ordered by a company that coordinates the joint assembly production of products from components, the manufacture of which employs a large number of independent enterprises. The entire production of end products can be represented as a directed acyclic graph, in which parts of a lower rank are used to assemble parts of a higher rank. Coordination of such production among a large number of independent manufacturers requires significant efforts from the parent company to coordinate plans, take into account the interests of participants, and constantly monitor their capabilities. At the same time, it is not possible to achieve an optimal solution and minimize costs. The task was formulated to develop an alternative approach to the formation of a production plan using modern ideas of distributed solutions based on network principles, similar to the organization of interaction within the blockchain, when the functioning of the system as a whole is determined by the individual interests and actions of its participants. The author proposed a system of interaction within the association of manufacturers based on multiple local auctions, where participants' offers for the production of components necessary to fulfill a common order are traded. This approach has its advantages, but it also carries certain risks. The article provides an overview of the auction scheme and discusses the difficulties associated with its implementation.

Keywords: auction, networking, production planning.

УДК 519.83

ББК 22.18

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

Поступила в редакцию 17.06.2024.

Опубликована 30.11.2024.

ВЛИЯНИЕ АЛГОРИТМОВ РАНЖИРОВАНИЯ, БОТОВ И МОДЕРАЦИИ КОНТЕНТА НА ФОРМИРОВАНИЕ МНЕНИЙ В СОЦИАЛЬНОЙ СЕТИ¹

Губанов Д. А.², Чхартишвили А. Г.³
(ФГБУН Институт проблем управления
им. В.А. Трапезникова РАН, Москва)

Рассматривается модель формирования информационных каскадов в онлайн-овых социальных сетях, учитывающая влияние алгоритмов ранжирования контента, действий ботов и модерации контента. Особое внимание уделено динамике мнений, которая критически важна для прогнозирования и управления социальными процессами. В отличие от традиционных моделей, здесь мнения агентов (пользователей) не наблюдаемы напрямую: их действия, такие как публикация комментариев, служат косвенными индикаторами взглядов. Эти действия влияют на мнения других пользователей, приводя к формированию информационного каскада в сети. Модель дополнена такими факторами, как алгоритмы показа комментариев, поведение ботов и модерация контента администратором информационного ресурса. Вычислительные эксперименты показывают, что алгоритмы ранжирования существенно влияют на динамику мнений и действий, особенно при ограниченной глубине просмотра пользователей. Кроме того, введение ботов и модерации может существенно изменить ход обсуждений. В работе исследуется взаимодействие стратегических игроков, включая модератора и ботов с противоположными позициями, и прогнозируется результат их взаимодействия на основе равновесий Нэша. Наконец, формализована и решена для частного случая задача управляющего органа, который, стремясь продвинуть нужную ему точку зрения в сети, осуществляет влияние на количество ботов.

Ключевые слова: онлайн-овые социальные сети, алгоритмы социальной сети, формирование мнений пользователей, боты, модерация контента, информационное противоборство, имитационное моделирование.

¹ Работа выполнена при финансовой поддержке Российского научного фонда в рамках проекта №23-21-00408.

² Дмитрий Алексеевич Губанов, д.т.н., в.н.с. (dmitry.a.g@gmail.com).

³ Александр Гедеванович Чхартишвили, д.ф.-м.н., г.н.с. (sandro_ch@mail.ru).

1. Введение

В последние десятилетия большой интерес теоретиков и практиков привлекают социальные сети. Выявление и прогнозирование динамики предпочтений пользователей онлайн-социальных сетей имеет огромную важность при моделировании информационного управления и информационного противоборства [10]. Эти предпочтения могут иметь социально-политическую, экономическую, психологическую или какую-либо другую природу.

Одним из методов исследования динамики мнений в социальных сетях является имитационное моделирование. Оно основано на задании на микроуровне правила изменения мнения пользователей (*агентов*) в зависимости от наблюдаемых ими действий тех соседних узлов сети – агентов, с которыми имеется та или иная связь (см., например, [16]). Отметим, что альтернативным подходом является моделирование динамики на макроуровне, например при помощи системы дифференциальных уравнений [17].

При моделировании динамики мнений в социальных сетях традиционно считается, что мнение и действие агента (индивида, являющегося узлом сети) отождествляются – см., например, [11, 12]. Применительно к онлайн-социальным сетям это означает, что агент (в данном случае – пользователь сети) без искажения транслирует свое внутреннее состояние и другие агенты имеют возможность это состояние наблюдать. В последнее десятилетие ситуация меняется в сторону разработки более сложных и реалистичных моделей [1, 4, 6, 14, 15]. Кроме того, предлагаются модели, в которых рассматривается влияние средств коммуникации на информационные процессы в сетях [13, 18].

В данной работе развивается ранее предложенная авторами модель [2, 3, 5, 7, 8] именно такого класса, где мнения (или предпочтения) агентов не наблюдаемы, а наблюдаемые действия не полностью отражают их мнения. Модель дополнена факторами алгоритма показа комментариев пользователям социальной сети, действий виртуальных агентов (ботов)

и модерации контента со стороны администратора страницы информационного источника.

В первом разделе статьи описана модель формирования информационного каскада и стратегические игроки, влияющие на него. Во втором разделе исследовано влияние ранжирования комментариев со стороны сети. В третьем разделе исследовано влияние ботов и модерации контента со стороны администратора. Наконец, в четвертом разделе рассмотрена возможность влияния управляющего органа, который может влиять на количество ботов, являющееся важным параметром ситуации информационного противоборства.

2. Формирование информационного каскада и стратегические игроки

В данном разделе мы опишем, следуя [9], модель формирования последовательности комментариев к сообщению (посту) в социальной сети.

Сначала приведем краткое вербальное описание. Будем считать, что имеется фиксированное множество пользователей социальной сети, являющихся подписчиками информационного источника. В информационном источнике публикуется пост, который рано или поздно увидят все пользователи-подписчики. У каждого пользователя есть мнение, которое он корректирует, прочитав часть ранее оставленных комментариев (отметим, что сам пост мы рассматриваем лишь как точку для сбора комментариев, т.е. он может как выразить какое-либо мнение, так и носить чисто информационный характер). После этого пользователь выбирает действие (в соответствии со своим сформированным мнением), ставит лайк соответствующим комментариям (из числа просмотренных), затем с некоторой вероятностью сам пишет комментарий (мы считаем, что каждый пользователь оставляет не более одного комментария).

Приведем теперь формальное описание модели. В начальный момент имеется множество $N = \{1, \dots, n\}$ пользователей, которые подписаны в онлайн-сети на данный информационный источник. Считаем, что у каждого пользователя $i \in N$ в начальный момент времени имеются

следующие параметры: мнение $x_i \in [0; 1]$, вероятность написать комментарий p_i , а также n_i – максимальное количество комментариев, которые пользователь просмотрит перед выбором своего действия (число комментариев / глубина просмотра – это параметр, отражающий характеристики как самого агента, так и характеристики поста). Также заданы неотрицательные числа $b_{ij}, i, j \in N$, характеризующие степень доверия пользователя i пользователю j .

Формирование последовательности комментариев после появления в информационном источнике сообщения (поста) осуществляется посредством выполнения n однотипных шагов.

На каждом шаге i с равной вероятностью выбирается любой из еще не видевших сообщение пользователей, не ограничивая общности будем считать его i -м. Он просматривает сообщение и либо первые n_i комментариев, либо, если количество всех имеющихся к данному шагу комментариев меньше n_i , все комментарии. Обозначим множество авторов просмотренных i -м пользователем комментариев через N_i . Каждый комментарий j -го пользователя является отражением его действия $y_j \in \{0,1\}$ – выбора позиции «за» (действие $y_j = 1$) или «против» (действие $y_j = 0$).

Будем считать, что под влиянием просмотренных комментариев i -й пользователь корректирует свое мнение x_i следующим образом:

$$(1) \quad x_i' = \frac{b_{ii}x_i + \sum_{j \in N_i} b_{ij}y_j}{b_{ii} + \sum_{j \in N_i} b_{ij}}.$$

После этого i -й пользователь выбирает свое действие $y_i \in \{0,1\}$ в соответствии с параметром x_i' : действие 1 («за») с вероятностью x_i' и действие 0 («против») с вероятностью $(1 - x_i')$. Далее i -й пользователь ставит лайк тем из просмотренных комментариев, которые соответствуют выбранному им действию (т.е. ставит лайк комментарию j -го пользователя при условии $y_i = y_j$). Наконец, в завершение шага i пользователь с вероятностью p_i сам пишет комментарий «за» или «против» в соответствии с выбранным действием (соответственно, с вероятностью $(1 - p_i)$ i -й пользователь не оставляет комментарий под сообщением). Введем параметр

$z_i \in \{0,1\}$, который равен 1, если i -й пользователь оставил комментарий, и равен 0 в противоположном случае.

В результате n шагов алгоритма формируется последовательность комментариев. Обозначим через N_z множество оставивших комментарий пользователей. Будем считать, что наиболее важной характеристикой последовательности является доля комментариев «за», т.е. $\delta = \sum_{i \in N_z} y_i / N_z$.

Пользователи социальной сети не являются стратегическими игроками, однако на них нацелено воздействие стратегических игроков. Будем рассматривать стратегических игроков трех типов, целью каждого из которых является либо максимизация, либо минимизация величины δ .

Первый тип – боты. Будем считать, что боты составляют множества $M^0 = \{n + 1, \dots, n + m^0\}$ и $M^1 = \{n + m^0 + 1, \dots, n + m^0 + m^1\}$, как бы дополняющие множество пользователей N . Бот отличаются от обычного пользователя тем, что всегда (с вероятностью 1) пишет комментарий, и его действие предопределено заранее: для j -го бота, $j \in M^0$, это действие $y_j = 0$ (и проставление лайков комментариям, соответствующим этому действию); для k -го бота, $k \in M^1$, это действие $y_k = 1$ (и, аналогично, проставление лайков комментариям, соответствующим этому действию). Таким образом, боты из множества M^1 (далее для краткости будем называть их 1-ботами) стремятся увеличить долю комментариев «за», а боты из множества M^0 (их будем называть 0-ботами) – уменьшить.

Второй тип стратегического игрока – администратор страницы информационного источника в сети (далее – администратор). Будем считать, что администратор источника может удалять нежелательные для него (по какой-либо причине) комментарии и лайки пользователей и ботов.

Наконец, третий тип стратегического игрока – сама онлайн-вая социальная сеть (далее – онлайн-вая сеть). Будем считать, что онлайн-вая сеть может управлять параметрами алгоритма ранжирования комментариев к постам.

Опишем теперь стратегии игроков, которые мы будем рассматривать.

Для ботов (т.е., по сути, для команд однотипных ботов) будем рассматривать два варианта:

(б1) боты с равной вероятностью оказываются на любом месте в последовательности комментаторов (как обычные пользователи);

(б2) боты являются первыми комментаторами.

Администратор может осуществлять модерацию, стирая определенное количество нежелательных для него комментариев (нежелательными являются либо комментарии «за», либо комментарии «против»). Будем считать, что администратор стирает каждый нежелательный комментарий, как только тот появляется с фиксированной вероятностью $q \in [0; 1]$, особо будем рассматривать два случая:

(а1) $q=0,2$;

(а2) $q=0,8$.

Онлайновая сеть может применять один из трех вариантов показа комментариев:

(с1) в обратном хронологическом порядке – сначала новые, потом старые;

(с2) в порядке убывания количества лайков (при одинаковом количестве лайков – в обратном хронологическом порядке, как в п. (с1));

(с3) сначала комментарии «за», затем комментарии «против» (внутри обоих множеств – в обратном хронологическом порядке, как в п. (с1)).

3. Влияние ранжирования комментариев

Для введенной выше модели будем оценивать характеристики информационных каскадов при помощи имитационного моделирования, позволяющего рассчитать усредненную долю комментариев «за» в зависимости от варианта показа (ранжирования) комментариев (действия ботов и администрации информационного источника рассмотрены в следующем разделе). Будем считать, что в сети $n = 100$ участников (например, подписчиков данной информационного ресурса) и она представляет собой полный граф, в котором каждый участник одинаково доверяет всем агентам (в том числе

самому себе). Мнения агентов в начальный момент времени равномерно распределены на отрезке $[0; 1]$, кроме того, для всех агентов одинакова как вероятность написать комментарий $p_i = 0,5$, так и «глубина» просмотра $n_i = 7$.

Сначала в качестве иллюстрации приведем результаты одиночных экспериментов (см. рис. 1, вертикальными линиями обозначены моменты высказывания комментариев, начальные мнения агентов в последовательности обозначены звездочками).

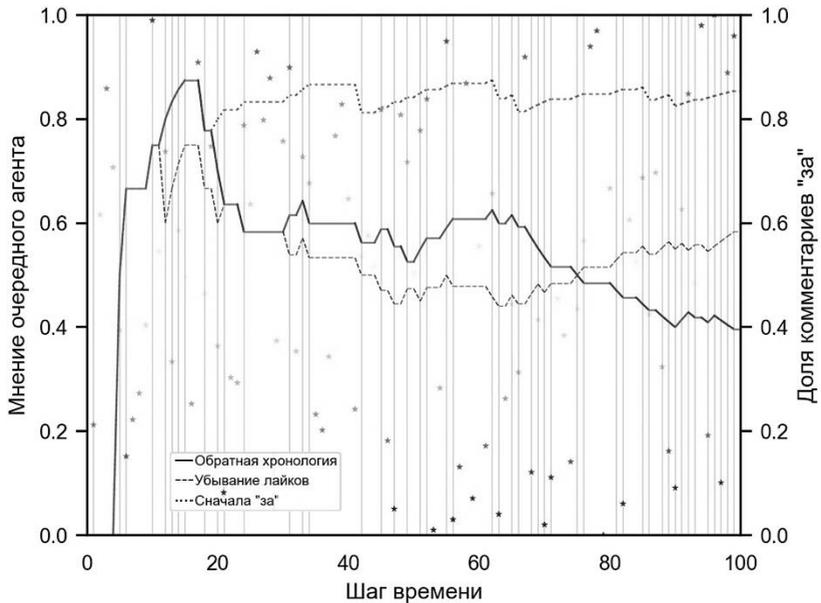


Рис. 1. Динамика доли комментариев «за»

Для случая показа в обратном хронологическом порядке (см. п.1) доля комментариев «за» резко растет, а затем спадает до уровня 0,40. Для случая показа в порядке убывания лайков (см. п.2) после резкого роста происходит падение, а затем рост восстанавливается до уровня 0,58. Показ сначала комментариев «за» приводит к удержанию уровня 0,85.

Для каждого варианта показа выполним 1000 запусков, а затем усредним результаты. Для варианта показа в обратном хронологическом порядке доля комментариев «за» составила 0,5,

для варианта показа в порядке убывания лайков – 0,5, а для показа сначала комментариев «за» – 0,9.

Введем теперь еще одну характеристику информационного каскада – долю комментариев «за» для первых 10 комментариев, которую дискретизируем следующим образом:

$$(2) \quad d = \begin{cases} 0, & \text{доля} \leq 1/3, \\ 1, & 1/3 < \text{доля} \leq 2/3, \\ 2 & \text{иначе.} \end{cases}$$

Увеличим количество агентов до $n = 500$ и проведем анализ доли «за» в зависимости от алгоритма показа, вероятности написать комментарий (p_i), глубины просмотра (n_i) и доли «за» (d), см. рис. 2.

| | n_i | 1 | | | 5 | | | 10 | | | 50 | | | Вся история | | | |
|-----------------------|-----------------------|------|------|------|------|------|------|------|------|------|------|------|------|-------------|------|------|---|
| | | d | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 |
| Вариант показа | p | | | | | | | | | | | | | | | | |
| Обратная хронология | 0,1 | 0,45 | 0,50 | 0,57 | 0,32 | 0,46 | 0,66 | 0,21 | 0,53 | 0,77 | 0,13 | 0,50 | 0,86 | 0,13 | 0,50 | 0,86 | |
| | 0,5 | 0,49 | 0,50 | 0,51 | 0,47 | 0,50 | 0,53 | 0,39 | 0,49 | 0,59 | 0,17 | 0,52 | 0,85 | 0,15 | 0,51 | 0,87 | |
| | 0,9 | 0,49 | 0,50 | 0,50 | 0,48 | 0,51 | 0,52 | 0,44 | 0,49 | 0,55 | 0,18 | 0,50 | 0,82 | 0,14 | 0,49 | 0,86 | |
| Сначала "за" | 0,1 | 0,63 | 0,71 | 0,76 | 0,60 | 0,83 | 0,91 | 0,43 | 0,82 | 0,93 | 0,13 | 0,50 | 0,86 | 0,13 | 0,50 | 0,86 | |
| | 0,5 | 0,73 | 0,74 | 0,75 | 0,85 | 0,90 | 0,92 | 0,83 | 0,93 | 0,95 | 0,48 | 0,84 | 0,96 | 0,15 | 0,51 | 0,87 | |
| | 0,9 | 0,74 | 0,74 | 0,75 | 0,88 | 0,91 | 0,92 | 0,89 | 0,94 | 0,95 | 0,66 | 0,90 | 0,97 | 0,14 | 0,49 | 0,86 | |
| Убывание лайков | 0,1 | 0,24 | 0,48 | 0,76 | 0,15 | 0,51 | 0,85 | 0,13 | 0,51 | 0,86 | 0,13 | 0,50 | 0,86 | 0,13 | 0,50 | 0,86 | |
| | 0,5 | 0,25 | 0,50 | 0,75 | 0,14 | 0,49 | 0,87 | 0,13 | 0,50 | 0,89 | 0,14 | 0,51 | 0,88 | 0,15 | 0,51 | 0,87 | |
| | 0,9 | 0,25 | 0,52 | 0,75 | 0,12 | 0,48 | 0,87 | 0,10 | 0,47 | 0,88 | 0,13 | 0,49 | 0,86 | 0,14 | 0,49 | 0,86 | |

Рис. 2. Доли комментариев «за» в зависимости от параметров

Сильно влияет на динамику вариант показа «сначала за», причем увеличение глубины просмотра до определенного момента позволяет усилить воздействие алгоритма (поскольку влияние окружения усиливается), но затем – по мере приближения n_i к n – воздействие ослабляется (поскольку агент начинает видеть все разнообразие позиций в сети). Кроме того, сложившееся в начале каскада «усредненное общественное мнение» (характеристика d) оказывает существенное влияние на итоговую долю комментариев «за» для всех алгоритмов показа. Особенно это верно для варианта показа «убывание лайков», что объясняется подкреплением влияния начального состояния лайками и возникновением положительной обратной связи.

Содержательно, ранний «захват» обсуждений, например, ботами приведет к достижению цели их владельца.

Рассмотрим теперь случай, когда распределение начальных мнений подчиняется бета-распределению (с параметрами $\alpha = 1$, $\beta = 5$). Проведем анализ доли «за» в зависимости от алгоритма показа, вероятности написать комментарий (p_i), глубины просмотра (n_i) и доли «за» (d), см. рис. 3.

| | n_i | 1 | | | 5 | | | 10 | | | 50 | | | Вся история | | | |
|-----------------------|-----------------------|------|------|------|------|------|------|------|------|------|------|------|------|-------------|------|------|---|
| | | d | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 |
| Вариант показа | p | | | | | | | | | | | | | | | | |
| Обратная хронология | 0,1 | 0,16 | 0,25 | 0,28 | 0,11 | 0,35 | 0,52 | 0,08 | 0,45 | 0,69 | 0,06 | 0,46 | 0,80 | 0,06 | 0,46 | 0,80 | |
| | 0,5 | 0,17 | 0,18 | 0,18 | 0,16 | 0,19 | 0,23 | 0,14 | 0,26 | 0,34 | 0,06 | 0,46 | 0,73 | 0,05 | 0,48 | 0,78 | |
| | 0,9 | 0,17 | 0,17 | 0,18 | 0,16 | 0,17 | 0,19 | 0,15 | 0,23 | 0,22 | 0,08 | 0,42 | 0,70 | 0,06 | 0,46 | 0,79 | |
| Сначала "за" | 0,1 | 0,42 | 0,57 | 0,63 | 0,35 | 0,79 | 0,86 | 0,21 | 0,78 | 0,90 | 0,06 | 0,46 | 0,80 | 0,06 | 0,46 | 0,80 | |
| | 0,5 | 0,55 | 0,58 | 0,59 | 0,72 | 0,85 | 0,86 | 0,65 | 0,90 | 0,92 | 0,21 | 0,83 | 0,93 | 0,05 | 0,48 | 0,78 | |
| | 0,9 | 0,57 | 0,58 | 0,59 | 0,78 | 0,86 | 0,86 | 0,77 | 0,91 | 0,92 | 0,40 | 0,89 | 0,96 | 0,06 | 0,46 | 0,79 | |
| Убывание лайков | 0,1 | 0,09 | 0,56 | 0,63 | 0,07 | 0,44 | 0,79 | 0,06 | 0,48 | 0,82 | 0,06 | 0,46 | 0,80 | 0,06 | 0,46 | 0,80 | |
| | 0,5 | 0,08 | 0,54 | 0,59 | 0,05 | 0,45 | 0,78 | 0,04 | 0,47 | 0,80 | 0,05 | 0,48 | 0,78 | 0,05 | 0,48 | 0,78 | |
| | 0,9 | 0,09 | 0,54 | 0,59 | 0,05 | 0,45 | 0,80 | 0,04 | 0,46 | 0,81 | 0,05 | 0,45 | 0,80 | 0,06 | 0,46 | 0,79 | |

Рис. 3. Доли комментариев «за» в зависимости от параметров

Предыдущие выводы сохраняются, однако мы видим общее для всех случаев естественное уменьшение доли «за».

Дадим содержательную интерпретацию полученным результатам.

Вариант показа по убыванию лайков. Нужно отметить, что агенты могут не публиковать комментарии, но ставят лайки. На начальном этапе появляется больше агентов с мнением «против», поэтому в показе будут доминировать комментарии с соответствующей позицией. Доминирование усугубляется в случае увеличения вероятности написать комментарий (p_i), поскольку агенты с мнением «против» не будут молчать и тем самым будут влиять на развитие информационного каскада (особенно при увеличении глубины просмотра агентов).

Вариант с обратной хронологией ведет к не столь быстрому доминированию доли комментариев «против». Агенты в последовательности поддаются влиянию уже высказавшихся участников сети. И здесь увеличение глубины наблюдения комментариев

«усугубляет» ситуацию, поскольку «консенсус» сложился, агенты поддаются влиянию большинства.

Вариант показа «сначала за» ведет в итоге к доминированию доли комментариев «за», но в случае агентов с небольшой глубиной наблюдений; в противном случае алгоритм может и не найти достаточное число комментариев «за» для перелома ситуации.

4. Влияние воздействия ботов и модерации комментариев

В данном разделе учтем действия ботов и действия администратора, который может осуществлять модерацию, стирая определенное количество нежелательных для него комментариев.

Рассмотрим следующие параметры имитационного моделирования:

- доля m_0 от n : $\{0; 0,1; 0,3\}$;
- доля m_1 от n : $\{0; 0,1; 0,3\}$;
- «глубина» просмотра всех агентов n_i : 5 или $(n + m_0 + m_1)$;
- вероятность написать комментарий $p_i = 1$;
- позиция администратора: {«за», «против»};
- алгоритм сети: $\{c1, c2, c3\}$.

Таким образом, имеется следующий набор изменяемых в ходе имитационного моделирования параметров: количество 0-ботов, количество 1-ботов, глубина просмотра, алгоритм сети c , позиция администратора ({«за», «против»}). В этой ситуации имеется три стратегических игрока: администратор (его стратегии $(a1), (a2)$), 0-боты (стратегии $(b1), (b2)$) и 1-боты (стратегии $(b1), (b2)$). Выигрыш каждого игрока определяется итоговой долей δ комментариев «за» пользователей (не ботов!) – со знаком плюс для администратора с позицией «за» и 1-ботов, со знаком минус для администратора с позицией «против» и 0-ботов. Будем считать, что имеется игра в нормальной форме – игроки принимают решение одновременно и независимо, стремясь максимизировать свой выигрыш.

Решением игры будем считать, как обычно, равновесие Нэша – ситуацию (набор стратегий игроков), в которой ни один игрок не может увеличить свой выигрыш, изменив свою стратегию при фиксированных стратегиях остальных игроков.

Зафиксируем каждый набор значений параметров и оценим выигрыши игроков при тех или иных стратегиях (проведем 100 запусков каждой конфигурации). Построив матрицу выигрышей, найдем решение игры для каждой ситуации – равновесия Нэша в чистых стратегиях (напомним, что равновесием Нэша называется ситуация, в которой ни один игрок не может увеличить свой выигрыш, изменив стратегию при неизменных стратегиях других игроков).

Нетрудно видеть, что всего возможно $2^23^3 = 108$ ситуаций. Оказывается, что в 105 из них существует равновесие Нэша в чистых стратегиях. Упорядочим ситуации с равновесиями по убыванию итоговой доли «за». Максимум достигается при максимальной глубине просмотра, максимальном числе ботов «за», отсутствии ботов «против» и позиции администратора «за». Минимум достигается при максимальной глубине просмотра, максимальном числе ботов «против», отсутствии ботов «за» и позиции администратора «против». Глубина просмотра уменьшает влияние алгоритма ранжирования, поскольку пользователи видят всю историю сообщений.

Во всех ситуациях доминирующая стратегия администратора – удалять как можно чаще «нежелательные» сообщения. Для ботов, как правило, лучше стратегия (б2), при которой боты стремятся быть первыми комментаторами, задавая начальный тон обсуждения. Начальное преобладание мнений может существенно повлиять на последующих агентов, склоняя их к тому же мнению. Однако есть случаи, когда равновесной является стратегия (б1), при которой боты с равной вероятностью оказываются на любом месте в последовательности комментаторов (как обычные пользователи). Во всех этих случаях глубина просмотра пользователей маленькая, при этом: а) либо сетью используется алгоритм «сначала новые»; б) либо алгоритм «по убыванию лайков» и как ботов с противоположной позицией больше, так и модератор имеет противоположную позицию; в) либо алгоритм «сначала за» и ботов с противоположной позицией («за») максимальное число (здесь равновесие (б1, б2)).

Теперь рассмотрим ситуацию, когда администратор по каким-то причинам фиксирует стратегию (а1) – это означает,

напомним, что он удаляет 20% нежелательных для него комментариев и это известно ботам. И снова лучшей стратегией является (б2).

Равновесия со стратегиями ботов «как обычные пользователи» (б1) возникают в случае, когда глубина просмотра пользователей маленькая ($n_i = 5$) и применяется алгоритм «сначала новые» (с1). Вариант равновесия «как обычные пользователи» (стратегия 0-ботов), «первые комментаторы» (стратегия 1-ботов) возникает в том случае, когда:

- позиция администратора «за»;
- доля 0-ботов – 10% или 30 %, доля 1-ботов – 10 %;
- глубина просмотра пользователей $n_i = 5$;
- алгоритм «сначала за» (с3).

Отметим случаи, когда изменение стратегии администратора приводит (см. рис. 4а и 4б, соответствующие стратегии линии $q = 0,2$ и $0,8$):

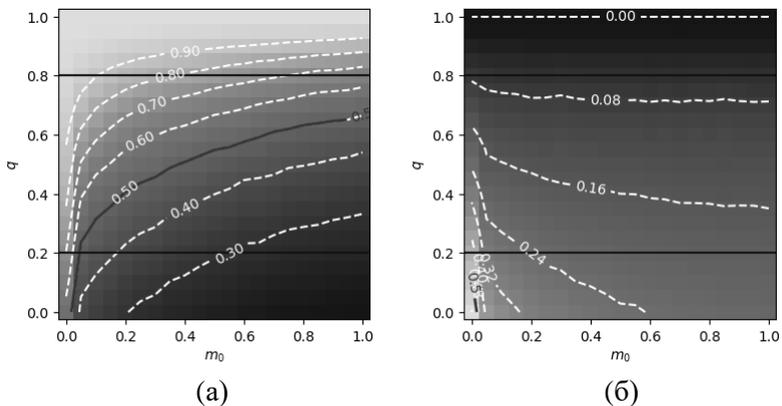


Рис. 4. Тепловые карты δ : а) позиции администратора ($y = 1$) и ботов противоположны; б) позиции администратора ($y = 0$) и ботов совпадают

а) к значительному изменению δ ($>0,5$) – это случай противостояния администратора с ботами с противоположной позицией, когда поддерживающие администратора боты отсутствуют, а глубина просмотра пользователей максимальна;

б) к незначительному изменению δ ($<0,007$) – это случай совпадения интересов администратора и ботов, когда боты с противоположной позицией отсутствуют, а глубина просмотра пользователей максимальна.

На рис. 4а видно, что ситуация условного баланса выигрышей противоборствующих сил ($\delta = 0,5$) описывается вогнутой кривой: при малом количестве ботов его увеличение «требует» от администратора значительного повышения усилий по удалению комментариев.

5. Воздействие управляющего органа

Выше были рассмотрены возможности стратегических игроков в зависимости от значений параметров ситуации. Эти игроки действуют оптимально в своих интересах, и равновесным (в теоретико-игровом смысле) результатом информационного процесса является средняя доля δ голосов «за», которая при каждом наборе параметров может быть оценена на основе многократных экспериментов. Таким образом, величина δ является функцией параметров ситуации. При этом важнейшими параметрами являются количества 0-ботов и 1-ботов – m_0 и m_1 .

Предположим теперь, что в ситуации участвует некий управляющий орган (далее будем называть его *центром*), который может влиять на параметры ситуации. Целевой функцией центра является максимизация доли голосов «за». Центр не управляет ботами (т.е. не определяет их стратегию), но может влиять на их количество, неся при этом определенные затраты.

Формализуем задачу центра. Пусть, понеся затраты $c_0 \geq 0$, он может добиться количества 0-ботов $m_0(c_0)$, а понеся затраты $c_1 \geq 0$ – количества 1-ботов $m_1(c_1)$; если соответствующего результата добиться невозможно, то затраты на его достижение условно можно считать бесконечными. Суммарные затраты ограничены величиной c , т.е. бюджетное ограничение имеет вид $c_0 + c_1 \leq c$. При этом выгода центра от доли «за» δ составляет $f(\delta)$. При этих условиях задача максимизации целевой функции центра F имеет следующий вид:

$$(1) \quad F(P, c_0, c_1) = f(\delta(P, m_0(c_0), m_1(c_1))) - c_0 - c_1 \xrightarrow{c_0, c_1} \max,$$

при ограничениях $c_0 \geq 0, c_1 \geq 0, c_0 + c_1 \leq c$, где через P обозначены остальные параметры ситуации. Она может быть решена при помощи имитационного моделирования.

В конкретных частных ситуациях принятие решений выбор центра может формулироваться более просто на основе той же задачи (1). Пусть, например, при данном наборе значений параметров P имеется 30% от n 0-ботов и отсутствуют 1-боты, а у центра имеются только три возможности, удовлетворяющие бюджетному ограничению:

- не влиять на ситуацию, тогда затраты отсутствуют и $F = f(\delta(P; 0, 3n; 0))$;
- уменьшить количество 0-ботов до нуля, тогда затраты c'_0 и $F = f(\delta(P; 0; 0)) - c'_0$;
- уменьшить количество 0-ботов до 10% от n (затраты c''_0), при этом одновременно увеличив количество 1-ботов до тех же 10% от n (затраты c''_1), тогда $F = f(\delta(P; 0, 1n; 0, 1n)) - c''_0 - c''_1$.

Тогда оптимальное действие центра соответствует максимальному значению целевой функции.

Пусть теперь $c_0 = c'_0 = \frac{3}{2}c''_0$, $c = 1$ (т.е. $c_0 + c_1 \leq 1$), $f(\delta(\cdot)) = k\delta(\cdot)$, где $k \in \{0,5; 1; 2; 20\}$.

Тогда целевая функция:

- в случае первой стратегии (A1): $k\delta_1$;
- в случае второй стратегии (A2): $k\delta_2 - c_0$;
- в случае третьей стратегии (A3): $k\delta_3 - \frac{2}{3}c_0 - c_1$.

Область доминирования стратегии A1:

$$k\delta_1 > k\delta_2 - c_0, \quad k\delta_1 > k\delta_3 - \frac{2}{3}c_0 - c_1, \text{ т.е.}$$

$$c_0 > k(\delta_2 - \delta_1), \quad c_1 > k(\delta_3 - \delta_1) - \frac{2}{3}c_0.$$

Область доминирования стратегии A2:

$$k\delta_2 - c_0 > k\delta_1, \quad k\delta_2 - c_0 > k\delta_3 - \frac{2}{3}c_0 - c_1, \text{ т.е.}$$

$$c_0 < k(\delta_2 - \delta_1), \quad c_1 > k(\delta_3 - \delta_2) + \frac{1}{3}c_0.$$

Область доминирования стратегии A3:

$$k\delta_3 - \frac{2}{3}c_0 - c_1 > k\delta_1, \quad k\delta_3 - \frac{2}{3}c_0 - c_1 > k\delta_2 - c_0, \text{ т.е.}$$

$$c_1 < k(\delta_3 - \delta_1) - \frac{2}{3}c_0, \quad c_1 < k(\delta_3 - \delta_2) + \frac{1}{3}c_0.$$

Численные расчеты показывают, что $\delta_1 = 0,115$, $\delta_2 = 0,509$, $\delta_3 = 0,498$.

Рассмотрим области доминирования для различных k .

Пусть $k = 0,5$ (см. рис. 5). Тогда область доминирования A1 определяется

- $c_0 > 0,197$, $c_1 > 0,192 - 0,666c_0$.

Область доминирования A2:

- $c_0 < 0,197$, $c_1 > -0,006 + 0,333c_0$.

Область доминирования A3:

- $c_1 < 0,192 - 0,666c_0$, $c_1 < -0,006 + 0,333c_0$.

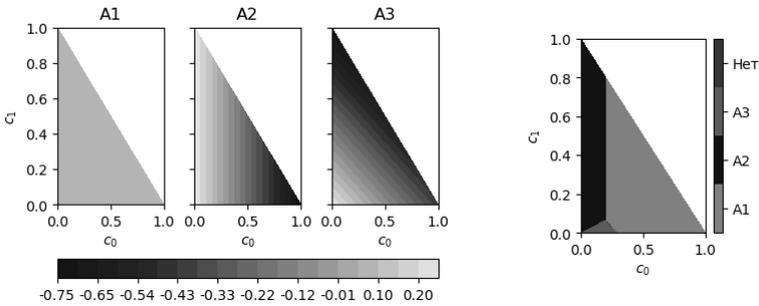


Рис. 5. $k = 0,5$. Значения целевой функции при разных действиях центра: A1, A2, A3; области доминирования представлены справа

Пусть $k = 1,0$ (см. рис. 6). Тогда области доминирования определены справа на рис. 6.

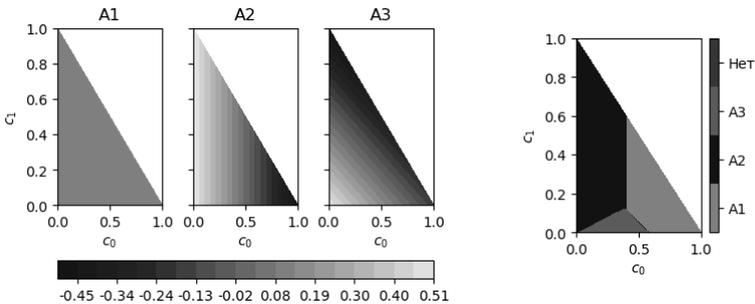


Рис. 6. $k = 1,0$. Значения целевой функции при разных действиях центра: A1, A2, A3; области доминирования представлены справа

Пусть $k = 2,0$ (см. рис. 7). Тогда области доминирования определены справа на рис. 7.

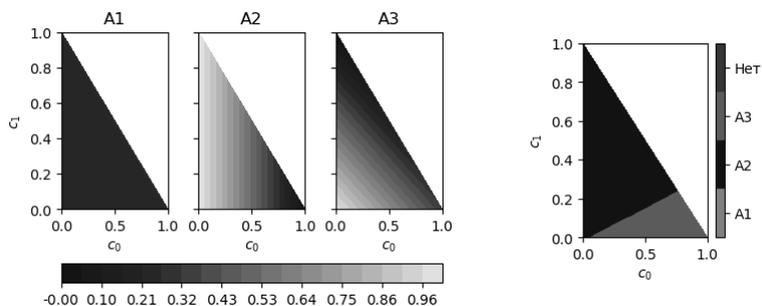


Рис. 7. $k = 2$. Значения целевой функции при разных действиях центра: A1, A2, A3; области доминирования представлены справа

Пусть $k = 2,0$ (см. рис. 8). Тогда области доминирования определены справа на рис. 8.

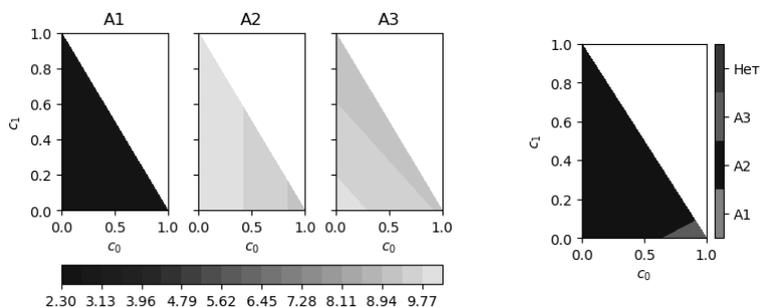


Рис. 8. $k = 20$. Значения целевой функции при разных действиях центра: A1, A2, A3; области доминирования представлены справа

По мере роста коэффициента k (определяющего выигрыш от числа голосов «за») увеличивается область допустимых затрат, при которых выгоднее действовать (т.е. выполнять A2 или A3). Выбор A2 или A3 определяется соотношением между c_0 и c_1 (см. рис. 5–8).

6. Заключение

В работе рассмотрена модель формирования информационных каскадов, в которой мнения (относительно некоторого вопроса) агентов не наблюдаемы, а наблюдаемые действия полностью отражают их мнения. Совершаемые агентами действия (написание комментариев) влияют на мнения действующих впоследствии агентов, тем самым формируя информационный каскад мнений и действий. Как показали вычислительные эксперименты, существенное влияние на такой каскад оказывает алгоритм показа предшествующих действий агенту сети: в обратном хронологическом порядке, по убыванию лайков, или сначала комментарии с заданной позицией. Особенно это верно в том случае, когда агенты просматривают небольшое число комментариев (возможно в силу когнитивных ограничений). Следовательно, относительно простые изменения в алгоритмах онлайн-социальной сети могут оказать косвенное, но решающее воздействие на мнения и предпочтения пользователей в сети, и в итоге – на их действия.

Существенное влияние на мнения агентов также оказывают действия ботов и действия администратора, который может осуществлять модерацию, стирая определенное количество нежелательных для него комментариев. В этих случаях при различных параметрах модели рассматривается протывоборство, для которого рассчитываются равновесия Нэша в чистых стратегиях. Во всех ситуациях доминирующая стратегия администратора – удалять как можно чаще «нежелательные» сообщения. Для ботов, как правило, лучше стратегия, при которой боты стремятся быть первыми комментаторами, задавая начальный тон обсуждения.

Начальное преобладание мнений может существенно повлиять на последующих агентов, склоняя их к тому же мнению, – это подтверждает важность первоначальной реакции пользователей для итогового результата обсуждения. Однако есть случаи, когда равновесной является стратегия «распределения», при которой боты с равной вероятностью оказываются на любом месте в последовательности комментаторов (как обычные пользователи). Это случаи характеризуются малой глубиной просмотра пользо-

вателей, когда администратор занимает противоположную позицию и (или) когда преобладают боты с противоположной позицией, – тогда ботам эффективнее предпринимать действия в случайные моменты времени, а не пытаться переломить ситуацию в самом начале (условно говоря, это «стратегия слабых»).

Формализована задача управляющего органа (центра), который может влиять на параметры ситуации, при этом неся определенные затраты и стремясь к максимизации доли голосов «за». Показаны области значений доступных центру затрат, где ему выгоднее отказаться от влияния либо предпринять то или иное действие.

Литература

1. ГУБАНОВ Д.А., НОВИКОВ Д.А. *Модели совместной динамики мнений и действий в онлайн-овых социальных сетях. Ч. 2. Линейные модели* // Проблемы управления. – 2023. – №3. – С. 40–64.
2. ГУБАНОВ Д.А., ПЕТРОВ И.В. *О модели поляризации мнений в социальных сетях* // Материалы 12-й Междунар. конф. «Управление развитием крупномасштабных систем» (MLSD'2019), Москва, ИПУ РАН. – М., 2019. – С. 1200–1202.
3. ГУБАНОВ Д.А., ПЕТРОВ И.В., ЧХАРТИШВИЛИ А.Г. *Многомерная модель динамики мнений в социальных сетях: индексы поляризации* // Проблемы управления. – 2020. – №3. – С. 26–33.
4. НОВИКОВ Д.А. *Модели динамики психических и поведенческих компонент деятельности в коллективном принятии решений* // Управление большими системами. – 2020. – Вып. 85. – С. 206–237.
5. ЧХАРТИШВИЛИ А.Г. *Задача нахождения медианного предпочтения индивидов в стохастической модели* // Автоматика и телемеханика. – 2021. – №5. – С. 139–150.
6. BANISCH S., OLBRICH E. *An argument communication model of polarization and ideological alignment* // arXiv:1809.06134. – 2018.
7. CHKHARTISHVILI A.G., GUBANOV D.A. *A Study on the Control of the Dynamics of Multidimensional Opinions in Social Networks* // Proc. of the 14th Int. Conf. "Management of Large-

- Scale System Development" (MLSD-2021). – Moscow: IEEE. – 2021. – DOI: 10.1109/MLSD52249.2021.9600250.
8. CHKHARTISHVILI A.G., GUBANOV D.A. *Forming Opinions in Social Networks: The Confrontation of Several Information Sources* // Proc. of the 15th Int. Conf. Management of Large-Scale System Development (MLSD-2022). – Moscow: IEEE. – 2022. – DOI: 10.1109/MLSD55143.2022.9934221.
 9. CHKHARTISHVILI A.G., GUBANOV D.A. *The Impact of Online Social Network Algorithms on User Opinion Formation* // Proc. of the 16th Int. Conf. Management of Large-Scale System Development (MLSD-2023). – Moscow: IEEE. – 2023. – DOI: 10.1109/MLSD58227.2023.10303932.
 10. CHKHARTISHVILI A.G., GUBANOV D.A., NOVIKOV D.A. *Social Networks: Models of information influence, control and confrontation*. – Cham, Switzerland: Springer International Publishing, 2019. – 158 p.
 11. DEGROOT M.H. *Reaching a Consensus* // Journal of American Statistical Association. – 1974. – No. 69. – P. 118–121.
 12. FLACHE A., MĂS M. et al. *Models of Social Influence: Towards the Next Frontiers* // The Journal of Artificial Societies and Social Simulation. – 2017. – Vol. 20, No. 4. – URL: <https://jasss.soc.surrey.ac.uk/20/4/2.html>.
 13. HUSZÁR F. et al. *Algorithmic amplification of politics on Twitter* // Proc. of the National Academy of Sciences. – 2022. – Vol. 119, No. 1. – e2025334119.
 14. KOZITSIN I.V. *A general framework to link theory and empirics in opinion formation models* // Scientific Reports. – 2022. – Vol. 12, No. 1. – URL: <https://www.nature.com/articles/s41598-022-09468-3>.
 15. MĂS M., FLACHE A. *Differentiation without distancing. Explaining bi-polarization of opinions without negative influence* // PloS one. – 2013. – Vol. 8, No. 11. – e74516.
 16. PERRA N., ROCHA L.E. C. *Modelling opinion dynamics in the age of algorithmic personalization* // Scientific reports. – 2019. – Vol. 9, No. 1. – P. 1–11.
 17. PETROV A.P., LEBEDEV S.A. *Online Political Flashmob: The Case of 632305222316434* // Computational Mathematics and Information Technologies. – 2019. – No. 1. – P. 17–28.

18. ROSSI W.S., POLDERMAN J.W., FRASCA P. *The Closed Loop Between Opinion Formation and Personalized Recommendations* // IEEE Trans. on Control of Network Systems. – 2021. – Vol. 9, No. 3. – P. 1092–1103.

THE INFLUENCE OF RANKING ALGORITHMS, BOTS AND CONTENT MODERATION ON OPINION FORMATION IN SOCIAL NETWORKS

Dmitry Gubanov, V.A. Trapeznikov Institute of Control Sciences of RAS, Moscow, Doctor.Sc. (dmitry.a.g@gmail.com).

Alexander Chkhartishvili, V.A. Trapeznikov Institute of Control Sciences of RAS, Moscow, Doctor.Sc. (sandro_ch@mail.ru).

Abstract: This paper considers a model of information cascade formation in online social networks, accounting for the influence of content ranking algorithms, bot actions, and content moderation. Special attention is given to opinion dynamics, which is critically important for predicting and managing social processes. Unlike traditional models, in this case, the opinions of agents (users) are not directly observable: their actions, such as posting comments, act as indirect indicators of their opinions. These actions influence the opinions of other users, leading to the formation of an information cascade within the network. The model incorporates additional factors such as comment ranking algorithms, bot behavior, and content moderation. Computational experiments demonstrate that ranking algorithms significantly affect the dynamics of opinions and actions, particularly when users have a limited view depth. Moreover, the introduction of bots and moderation can substantially alter the course of discussions. The study explores the interaction of strategic players, including the moderator and bots with opposing views, and predicts the outcome of their interactions based on Nash equilibria. Finally, the problem of a control subject (principal) is formalized and solved for a specific case, where it seeks to advance a specific viewpoint in the network by influencing bots within the network.

Keywords: online social networks, social network algorithms, user opinion generation, bots, content moderation, information warfare, simulation modeling.

УДК 519.8+51-77
ББК 22.18

Статья представлена к публикации членом редакционной коллегии Ф.Т. Алескеровым.

*Поступила в редакцию 28.09.2024.
Опубликована 30.11.2024.*

ИНТЕГРАЦИОННОЕ ПРОГНОЗИРОВАНИЕ НЕСТАЦИОНАРНЫХ ПРОЦЕССОВ, ПРЕДСТАВЛЕННЫХ ВРЕМЕННЫМИ РЯДАМИ. ОБЗОР¹

Авдеева З. К.², Коврига С. В.³
(ФГБУН Институт проблем управления
им. В.А. Трапезникова РАН, Москва)

Обзор охватывает основные направления и подходы к интеграционному прогнозированию нестационарных процессов, представленных временными рядами. Ключевым источником порождения нестационарности являются быстрые и плохо предсказуемые изменения внешней среды, под влиянием которых происходят структурные сдвиги в процессах, протекающих в сложных экономических и социально-политических системах. Решение задач прогнозирования динамики таких объектов в контексте повышения точности прогноза усложняется по мере увеличения горизонта прогнозирования, что обуславливает потребность в моделях и методах, способных обрабатывать разнородную информацию. Интеграционные методы – это методы, позволяющие наряду с количественными данными учитывать суждения (прогнозистов, экспертов, аналитиков) и информацию из разнородных информационных источников на разных этапах решения задачи и тем самым прямым или косвенным способом включать их в формируемый прогноз. Развитие таких методов направлено на повышение точности прогноза через использование всей доступной информации об объекте прогнозирования, включая данные об эндогенных и экзогенных факторах влияния на него. В обзоре внимание было сконцентрировано на современном состоянии в области интеграционного прогнозирования, на существующих проблемах и путях их решения.

Ключевые слова: нестационарные процессы, временные ряды, прогнозирование, суждения.

1. Введение

Анализ и прогнозирование временных рядов является популярным прогностическим инструментом для работы с большими объемами данных, отражающих закономерности поведе-

¹ Работа выполнена при частичной финансовой поддержке Российского научного фонда (грант № 23-21-00455).

² Зинаида Константиновна Авдеева, к.т.н., с.н.с. (avdeeva@ipu.ru).

³ Светлана Вадимовна Коврига, н.с. (kovriga@ipu.ru).

ния исследуемых процессов в экономических (планирование на разных уровнях управления, товарные и финансовые рынки, управление цепочками поставок и др.), социально-политических (протестная активность, избирательные компании и др.), социотехнических системах (социальные сети) [4, 6].

Несмотря на то, что количественные методы анализа и прогнозирования временных рядов постоянно развиваются, их потенциала недостаточно для эффективного прогнозирования нестационарных процессов, описывающих динамику таких систем. Одним из ключевых источников порождения нестационарности являются быстрые, плохо предсказуемые изменения внешней среды, под влиянием которых происходят структурные сдвиги и структурные изменения прогнозируемых процессов.

При этом проблема точности прогноза нарастает по мере увеличения горизонта прогнозирования. Прогнозирование временного ряда на несколько шагов вперед на длительный период времени является сложной задачей, поскольку уровень неопределенности прогноза увеличивается с ростом горизонта [34].

Многочисленные исследования показывают, что количественные (статистические) и качественные (основанные на экспертных знаниях) методы имеют ограничения и проблемы с точностью. Оба класса методов зависят от качества исходных данных и знаний, от типа неопределенности внешних факторов, от волатильности значимых параметров и пр., и, следовательно, от способности улавливать в информационном пространстве сигналы, указывающие на начало значимых изменений (см., например, [39, 46]).

Такие сигналы невозможно идентифицировать исключительно на основе анализа количественных данных, поскольку они отсутствуют в данных временных рядов. Необходимо привлечение экспертных знаний и выбор подходящих методов обработки этих знаний в контексте решения задачи прогнозирования. Поэтому в условиях высокой изменчивости факторов воздействия и самого прогнозируемого процесса, когда статистические методы теряют свою эффективность, повышение качества прогноза по мере роста горизонта прогнозирования воз-

можно путем сочетания качественных и количественных методов.

Основная цель данной статьи – представить обзор основных направлений и подходов к прогнозированию нестационарных процессов с прямым и косвенным включением в формируемые прогнозы суждений (прогнозистов, экспертов и аналитиков¹) и информации из разнородных информационных источников. Рассматривается прогнозируемый объект, процесс изменения которого описывается временным рядом или группой временных рядов, динамика которых может изменяться не только вследствие эндогенных изменений, но и под воздействием экзогенных изменений в окружающей среде объекта. Под влиянием экзогенных факторов в процессе могут возникать структурные сдвиги: скачки уровня или волатильности, изменения трендов временных рядов, изменения связей между компонентами рядов. Факторы влияния могут быть описаны 1) временными рядами; 2) данными, полученными в результате структурирования и обработки различных источников информации о состоянии и направлении влияния факторов на текущее состояние объекта прогнозирования.

Проведенный в рамках обзора анализ подходов к прогнозированию такого рода объектов позволяет выделить общую тенденцию развития методов прогнозирования временных рядов. За период начиная с основополагающей работы [20] и вплоть до наших дней широкий круг специалистов подтверждает, что статистические комбинации прогнозов повышают точность прогнозов. Наряду с этим эмпирические данные свидетельствуют об усилении тенденций, связанных с развитием комбинированных методов по пути 1) сочетания статистических прогнозов и прогнозов, полученных с помощью машинного обучения [45], и 2) включения суждения в методы формирования прогнозов [43, 45]. При этом также отмечается необходимость повышения

¹ *Прогнозист – специалист по прогнозированию, носитель знаний о моделях и методах прогнозирования; эксперт, аналитик – носитель знаний предметной области (допускается, что специалист может быть носителем и тех и других знаний).*

интерпретируемости результатов прогнозов, чтобы они были пригодными для принятия решений.

С учетом ориентации на объекты прогнозирования – нестационарные процессы, порождаемые изменениями внешней среды, а также с учетом современных тенденций построения прогнозов на основе комбинирования методов, в том числе развития интеграционных методов, позволяющих включать суждения в формируемый прогноз, в обзоре внимание было сконцентрировано на современном состоянии в этой области, в большей степени отражающих возможности и ограничения интеграционного прогнозирования, использующего разнотипную информацию, а также на нерешенных проблемам. При этом в качестве основных критериев применимости моделей и методов к прогнозированию нестационарных процессов выделены точность формируемого прогноза, интерпретируемость его результатов для принятия решений, своевременность получения прогноза.

Насколько известно авторам, в отечественной публикационной среде обзор, систематизирующий основные направления и подходы к развитию интеграционных методов прогнозирования временных рядов, не представлен. Существующие обзоры, известные авторам, посвящены количественному прогнозированию временных рядов (например, один из последних обзоров представлен в [1]). Зарубежные обзоры, посвященные интеграционному прогнозированию, как правило, затрагивают отдельные аспекты проблемы включения суждений в формируемые прогнозы, в частности, рассматривают вопросы понимания и оценки основных когнитивных особенностей и поведенческих проблем прогнозистов, которые необходимо учитывать при разработке интеграционных методов, применимость таких методов в некоторой прикладной области (например, прогнозирования спроса на продукцию/продаж в контексте цепочки поставок и т.д.), сравнение эффективности оценочных и статистических методов (см. например, недавние обзоры [15, 32, 43]) и др.

Раздел 2 включает вступительную часть к обзору. В разделе 3 представлено общее описание особенностей и подходов к формированию прогнозов с использованием количественных

методов как необходимой базы для решения задач прогнозирования нестационарных процессов в условиях структурных сдвигов. Раздел 4 содержит обзор основных направлений и подходов к интеграционному прогнозированию, использующему суждения и количественные данные на разных этапах прогнозирования, выделены проблемы разработки и применения таких методов и пути их решения. В заключении представлены обобщающие выводы.

2. Общие положения

При выделении основных направлений и подходов к прогнозированию нестационарных процессов со структурными сдвигами ориентация была на следующие основные классификационные признаки: степень формализации, общий принцип действия, способ получения прогнозной информации, горизонт прогноза, тип информации и данных о процессе.

Следует отметить, что рассмотренные направления и подходы к прогнозированию процессов, представленных временными рядами, в большей степени являются общими, независимо от их стационарности или нестационарности. Специфика, как правило, связана с применением конкретных моделей и методов прогнозирования нестационарных процессов. Выбор метода прогнозирования зависит от ряда факторов: свойств прогнозируемого ряда (случайное блуждание или фрактальный процесс); регулярности решения задачи (одноразовый или повторяющийся); формы представления прогноза (конкретное значение, конкретное значение и/или интервал, направление тренда – восходящий, нисходящий – и оценка его продолжительности), длины горизонта прогноза; типа информации и данных о прогнозируемом процессе, их доступности и полноте; требований к точности прогноза [1].

По типу используемых данных и информации и степени формализации все методы прогнозирования делятся на 1) интуитивные (экспертные), 2) формализованные (количественные) и 3) интеграционные (смешанные), использующие 1) и 2).

На рис. 1 отражена применимость различных классов методов на горизонте прогнозирования в зависимости от обеспеченности информацией об объекте прогнозирования.

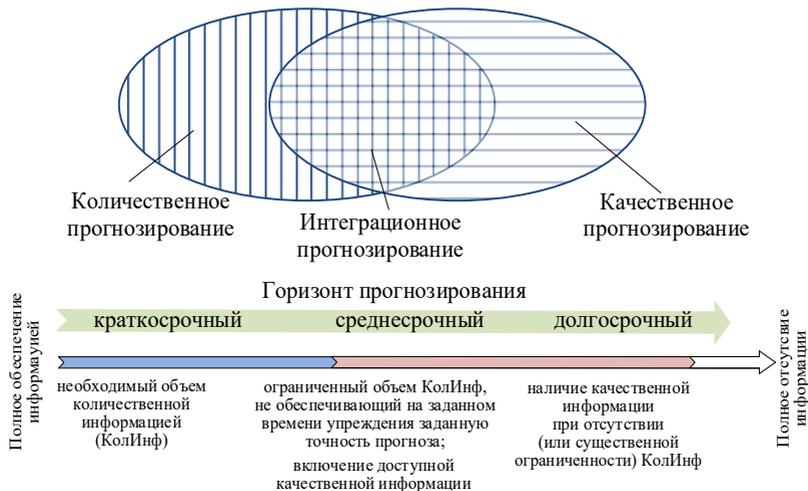


Рис. 1. Применимость методов прогнозирования в зависимости от обеспеченности информацией на горизонте прогнозирования

Класс интуитивных (экспертных) методов применяется в случаях, когда объект прогнозирования либо слишком прост, либо настолько сложен, что аналитически учесть влияние многих факторов практически невозможно [4], либо прогноз формируется на слишком большом горизонте прогнозирования (как правило, это долгосрочное прогнозирование).

В результате опроса экспертов полученные экспертные оценки используют как конечные прогнозы. В зависимости от общих принципов действия эти методы делятся на две группы: индивидуальные и коллективные экспертные оценки. К типовым экспертным методам, применяемых для прогнозирования развития таких объектов, относятся метод интервью, метод сценариев, метод Дельфи [45].

В обзоре подробно не представлен класс чисто экспертных методов, так как рассматривается объект прогнозирования, ди-

намика которого представлена временным рядом. В основе анализа и прогнозирования временных рядов, как правило, лежат методы формальной обработки. Включение в обзор общего описания особенностей и подходов к формированию прогнозов на основе количественных методов обусловлено тем, что они рассматриваются как необходимая методологическая база для решения задач прогнозирования нестационарных процессов со структурными сдвигами на горизонте прогнозирования.

Далее более подробно рассмотрим общие подходы к прогнозированию нестационарных процессов на основе количественных и интеграционных методов.

3. Прогнозирование на основе количественной информации

Количественные методы по общему принципу действия можно разделить на единичные (когда прогноз формируется на основе единственного метода) и комбинированные (на основе интеграции нескольких методов).

Комбинирование прогнозов зарекомендовало себя в качестве эффективного подхода к прогнозированию, приводящего в основном к повышению точности прогнозов [5, 8, 14, 45], в том числе в задачах обнаружения структурных сдвигов временных рядов, обусловленных изменяющимися во времени внешними условиями. При этом изменения динамики могут быть вызваны не только постепенным изменением воздействия внешних факторов, но и внезапными событиями [18]. Как отмечалось во введении, наблюдается усиление тенденции развития комбинированных методов.

Актуальность совместной применимости методов для комбинированных прогнозов (независимо от используемых методов: только количественных или интеграционных) следует из следующих условий [14]:

– если нет уверенности, какой метод прогнозирования является наиболее точным. Это может быть вызвано разными причинами: новой ситуацией, наличием разнородного набора временных рядов или ожиданием того, что будущее будет особенно

нестабильным, изменчивым. Даже если можно определить лучший метод, объединение может быть полезным, если другие методы дают некоторую дополнительную информацию;

– когда нет уверенности в ситуации прогнозирования. В ситуациях неопределенности объединение может уменьшить ошибку. Неопределенность ситуации может быть оценена на основании необъяснимых значительных отклонений в прошлом или ожидаемых неустойчивых изменений в будущем. Поскольку неопределенность увеличивается с увеличением горизонта прогноза, комбинирование должно быть особенно полезным для среднесрочных и долгосрочных прогнозов;

– когда важно избежать больших ошибок.

Базовыми принципами объединения прогнозов является использование [7, 8, 14]

– нескольких источников данных, позволяющих добавить дополнительную полезную информацию, способствующую более полному и точному описанию прогнозируемого процесса, а также скорректировать предвзятость (при привлечении экспертов);

– разных методов или данных, или сочетаний различных методов и данных;

– формальных процедур для объединения прогнозов;

– знаний предметной области, чтобы варьировать весами методов;

– равных весовых коэффициентов, если нет веских доказательств в поддержку неравного веса прогнозов;

– большего веса для метода, который априори известен как лучший по сравнению с остальными.

Линейная форма является наиболее распространенной формой при объединении прогнозов в силу своей простоты и удобства в использовании [8]:

$$(1) \hat{F}_{fin} = \sum_{i=1}^n w_i \hat{F}_i,$$

где \hat{F}_{fin} – объединенный прогноз; \hat{F}_i – i -й частный прогноз; w_i – весовой коэффициент, с которым i -й частный прогноз был включен в объединенный прогноз.

Одной из важнейших проблем объединения прогнозов является нахождение весовых коэффициентов, с которыми частные прогнозные значения объединяются в общем прогнозе. От выбора метода построения весовых коэффициентов зависит уровень повышения точности прогнозирования. На сами весовые коэффициенты накладываются только два ограничения: сумма всех весовых коэффициентов должна быть равно единице и весовые коэффициенты должны находиться в интервале $[0, 1]$ [8].

В [8] рассмотрены основные требования, предъявляемые к выбору весов, преимущества и недостатки различных подходов к их определению. В недавнем обзоре [56] можно ознакомиться с современным состоянием в области комбинирования прогнозов (независимо от того, объединяются только количественные прогнозы или, в том числе, с включением суждений), в части определения потенциала и ограничений различных стратегий комбинирования (выбора весов объединяемых прогнозов).

Также поднимается проблема выбора из пула прогнозов до комбинирования и отмечается необходимость разработки эмпирических рекомендаций и количественных метрик, которые помогут прогнозистам в выборе [55, 56].

По способу обработки и получения прогнозной информации можно выделить три основных подхода [43]: 1) экстраполяция, 2) многомерные методы, 3) компьютероёмкие методы.

Экстраполяция использует исторические данные для прогнозирования. Из экстраполяционных методов экспоненциальное сглаживание, которое усредняет (сглаживает) данные временных рядов, является одним из широко практикуемых классических методов. Авторегрессионное интегрированное скользящее среднее (ARIMA) и его многочисленные варианты (например, ARIMAX, ARMA-GARCH, ARFIMA) также являются популярными экстраполяционными методами, поскольку они могут учитывать тенденции, сезонность, ошибки и нестационарные аспекты временного ряда.

Многомерные методы являются формами регрессионного анализа и предполагают наличие причинной связи между зависимой переменной и одной или несколькими независимыми (объясняющими) переменными. Такие методы способны решить

проблему среднесрочного прогнозирования нестационарных временных рядов, рассматривая различные экзогенные переменные, которые могут повлиять на объект прогнозирования.

Компьютероемкие (computer intensive) методы¹ прогнозирования являются более поздними разработками. Недавний рост их популярности стимулирован ростом интеллектуального анализа данных и прогнозной аналитики, а также ростом накопленных количественных массивов нужной длины и качества. Еще одной причиной их развития обусловлено тем, что традиционные методы многомерного прогнозирования ограничены задачами низкой размерности, линейными зависимостями и короткими горизонтами. Технологические достижения вместо этого смещают акцент на проблемы, характеризующиеся большим количеством переменных, нелинейными зависимостями и длительными горизонтами прогнозирования. За последние несколько лет большинство наиболее эффективных методов многомерного прогнозирования были основаны на моделях глубокого обучения. Однако такие модели характеризуются высокими требованиями к доступности данных и вычислительным ресурсам и страдают от недостаточной интерпретируемости [25].

Также по способу обработки информации с учетом горизонта прогнозирования можно выделить два стандартных подхода: прогнозирование на один шаг и на несколько шагов вперед.

Последовательное применение одношаговой прогнозной модели, учитывающей взаимосвязи между последовательными значениями процесса, может быть полезен только на ограниченном интервале, поскольку не учитывает изменения, возникающие на горизонте прогнозирования. Если состояние объекта зависит не только от значений процесса, но и от значений упреждающих факторов, описываемых временными рядами других процессов, то использование информации о состоянии упреждающих факторов позволяет повысить качество прогноза.

Многошаговое прогнозирование ориентировано на формирование средне- и долгосрочных прогнозов.

¹ Это методы, требующие значительных вычислительных ресурсов и времени.

В общем виде задача многошагового прогнозирования

– для одномерного исторического временного ряда, состоящего из T наблюдений, состоит в прогнозировании следующих H значений, где $H > 1$ – горизонт прогноза:

$$(2) \{y_1, y_2, \dots, y_T\} \Rightarrow \{y_{T+1}, y_{T+2}, \dots, y_{T+H}\}.$$

– для многомерного ряда, состоящего из N рядов исторических наблюдений вида $\{y_1^k, y_2^k, \dots, y_T^k\}$, –

$$(3) \{Y_T^1, \dots, Y_T^k, \dots, Y_T^N\} \Rightarrow \{Y_{T+H}^1, \dots, Y_{T+H}^k, \dots, Y_{T+H}^N\},$$

где $Y_{T+H}^k = \{y_{T+1}^k, y_{T+2}^k, \dots, y_{T+H}^k\}$.

В отличие от прогнозирования на один шаг вперед, прогнозирование состояния объекта (например, цен на товарных рынках) на заданный горизонт H с пошаговым разбиением представляет трудную задачу; качество и методы решения зависят от свойств прогнозируемого объекта. Также сложность обусловлена тем, что приходится иметь дело с различными проблемами: накоплением ошибок, снижением точности и увеличением неопределенности [38, 52]. Однако задачи прогнозирования на несколько шагов все больше находят распространение в различных приложениях.

Существует три основные стратегии решения задачи многошагового прогнозирования: рекурсивная, прямая (независимая) и совместная стратегии [52, 53].

В рекурсивной стратегии (Recursive, S^{Rec}) формирование прогноза \hat{y}_{T+h} на h -шагов вперед происходит путем итеративного вычисления значений в $t+1, t+2, \dots$ по сформированной одношаговой модели прогноза $M^{one-step}$:

$$(4) M^{one-step}: \hat{y}_{t+1} = f(y_t, y_{t-1}, \dots, y_{t-d+1}) + w,$$

где $t \in d, \dots, T-1$; d – количество прошлых наблюдений; w – слагаемое, включающее ошибку модели и внешние возмущения/шум;

$$(5) S^{Rec}: \hat{y}_{T+h} = \begin{cases} \hat{f}(y_T, \dots, y_{T-d+1}), & \text{если } h = 1; \\ \hat{f}(\hat{y}_{T+h-1}, \dots, \hat{y}_{T+1}, y_T, \dots, y_{T-d+1}), & \text{если } h \in \{2, \dots, d\}; \\ \hat{f}(\hat{y}_{T+h-1}, \dots, \hat{y}_{T+h-d}), & \text{если } h \in \{d+1, \dots, H\}. \end{cases}$$

В зависимости от шума, присутствующего во временных рядах и горизонте прогнозирования, рекурсивная стратегия может быть недостаточно эффективной в задачах многошагового прогнозирования. Эта стратегия чувствительна к накоплению ошибок на горизонте прогнозирования.

Прямая стратегия (*Direct*, S^{Dir}) заключается в формировании многомасштабного прогноза \hat{y}_{T+h} на горизонте H , когда исходный временной ряд преобразуется в ряды в h -временных масштабах, например, для годового прогноза в четырёх масштабах: месяц, квартал, полгода, год, а далее формируется h прогнозных моделей:

$$(6) \quad M^h: y_{t+h} = f_h(y_t, \dots, y_{t-d+1}) + w,$$

где $t \in \{d, \dots, T-H\}$, $h \in \{1, \dots, H\}$.

В этой стратегии прогноз есть h условно независимых прогнозов

$$(7) \quad S^{Dir}: \hat{y}_{T+h} = \hat{f}_h(y_T, \dots, y_{T-d+1}).$$

S^{Dir} не использует приближительных значений для расчета прогнозов, что делает ее невосприимчивой к накоплению ошибок. Такая стратегия применяется и для улучшения качества краткосрочных прогнозов, и для формирования среднесрочных прогнозов. При этом эта стратегия требует большого вычислительного времени, поскольку необходимо построить h моделей. Сочетание многомасштабного прогноза с комбинированием результатов прогноза при поступлении новых данных обеспечивает возможность коррекции за счет выбора лучших из имеющихся наборов. Однако предположение о независимости h прогнозов между собой не всегда верно, так как есть методы, которые показывают сложные зависимости между поведением ряда в разных масштабах.

Рекурсивную (*Recursive*), прямую стратегию (*Direct*) и их объединение *RecDir* относятся к стратегиям с одним выходом, поскольку они моделируют данные как функцию с одним выходом (с несколькими входами).

Совместная стратегия, также называемая «множество входов и много выходов» (*MIMO*, S^{MIMO}) (доступна только для моделей нелинейного прогнозирования) предполагает обучение

одной многовыходной модели на истории наблюдений $[y_1, y_2, \dots, y_T]$:

$$(8) M^{MIMO}: [y_{T+H}, \dots, y_{t+1}] = F(y_t, \dots, y_{t-d+1}) + \mathbf{w},$$

где $t \in \{d, \dots, T-H\}$, $F: R^d \rightarrow R^H$ является векторной функцией, \mathbf{w} – вектор шума с необязательно диагональной ковариационной матрицей.

Прогноз формируется за один шаг:

$$(9) S^{MIMO}: [\hat{y}_{T+H}, \dots, \hat{y}_{t+1}] = \hat{F}(y_T, \dots, y_{T-d+1}).$$

Введение стратегии MIMO мотивировано необходимостью избегать моделирования по стратегии с одним выходом, которое игнорирует существование стохастических зависимостей между будущими значениями, что влияет на точность прогноза. S^{MIMO} формируется на основе стохастической зависимости между прогнозируемыми значениями, характеризующей временной ряд. Она позволяет избежать условной независимости h прогнозов прямой стратегии, а также накопления ошибок, свойственного рекурсивной стратегии. Однако необходимость сохранения стохастических зависимостей за счет использования одной модели имеет недостаток, поскольку ограничивает все прогнозируемые горизонты одной и той же структурой модели. Это ограничение может снизить гибкость подхода к прогнозированию. Для решения этой проблемы предложена стратегия объединения MIMO и стратегии Direct – DIRMO. Также применение стратегии MIMO требует большего объема данных при обучении и подвержено проблеме переобучения.

Прогнозирование обнаружения структурных сдвигов в процессе и динамики развития процесса после того как в нем произошёл структурный сдвиг является важным элементом формирования многошагового прогноза. Решение этой задачи позволяет проанализировать динамику ряда и своевременно обнаружить наличие изменений (структурных сдвигов).

Поэтому важнейшей составной частью задачи прогнозирования процессов со структурными сдвигами является задача мониторинга – обнаружения разладки временного ряда, характеризующей момент времени, связанный с существенными изменениями свойств ряда. Еще больше возрастает роль мони-

торинга, если изменение динамики наблюдаемого процесса на горизонте прогнозирования обусловлено влиянием событий внешней среды.

В частности, механизмы формирования цен на товарных рынках, состав, сила влияния и направление других факторов, включенных в модель прогноза, могут изменяться на горизонте прогноза. В этих условиях модель перестает описывать динамику прогнозируемой цены, становится непригодной для прогноза и требует замены. Причинами изменений могут служить изменения внешней среды (политические или экономические события, изменения в планах развития масштабных проектов, пандемия, военно-политические конфликты и пр.) и внутренние механизмы функционирования рынка (конкуренция, ожидания относительно будущего участников рынка, используемые ими стратегии). Поэтому необходимо отслеживать происходящие изменения, своевременно изменять и корректировать модели посредством мониторинга, целями которого являются: обнаружение изменений временных рядов (структурных сдвигов), выявление причин, которые вызвали или могут вызвать возникновение структурных сдвигов.

Для обнаружения и идентификации типов изменений нестационарных процессов используются два широко распространенных подхода [33]:

- 1) флуктуационные тесты, модифицированные для применения в режиме мониторинга;
- 2) алгоритмы последовательного обнаружения и идентификации изменений, в основе которых лежит подход А. Вальда.

Алгоритмы мониторинга, основанные на флуктуационных тестах, представляют собой модификацию ретроспективного теста рекурсивных оценок момента изменения свойств, границы которого расширяются с ростом объема выборки. Эти тесты просты и удобны в применении.

В основе алгоритмов последовательного обнаружения лежит схема обработки поступающих наблюдений с использованием статистики отношения правдоподобия. В качестве критерия качества обнаружения используется минимизация запаздывания при ограничениях на частоту ложных обнаруже-

ний и вероятность ошибочной идентификации. Последовательные алгоритмы ранее использовались преимущественно в технических приложениях. Работы, исследующие эффективность их применения для экономических данных, появились сравнительно недавно.

Для прогнозирования нестационарных процессов используются различные методы: построение параметрических моделей, сингулярный спектральный анализ, декомпозиция по эмпирическим модам, машины опорных векторов, нейронные сети и др. [21, 23]. Однако высокий уровень неопределенности в динамике процессов (например для таких объектов, как товарные рынки) создает проблемы для применения только количественных моделей прогнозирования.

Чтобы уменьшить эту неопределенность, некоторые авторы [48, 52] исследовали различные стратегии многошагового прогнозирования (обзор стратегий приведен выше) и показали, что наиболее эффективными для многошагового прогнозирования являются стратегии с несколькими выходными данными. Применение этих стратегий предполагает сохранение стохастических зависимостей между членами прогнозируемой последовательности. Этим условиям удовлетворяют регулярные процессы, когда динамика ряда может быть с достаточной точностью описана некоторой динамической системой. Проблема, однако, в том, что динамика временных рядов часто меняется со временем, а причины изменения связаны с влиянием внешних событий. Поэтому в таких ситуациях более эффективен подход, основанный на извлечении и учете в прогнозе дополнительной информации о событиях прогнозируемого периода, влияющих на прогнозируемый процесс [58].

Однако эта информация не всегда может быть своевременно отражена в данных. Для нестационарных процессов, изменение динамики которых часто связано с внешними событиями, вызывающими структурные сдвиги, все больше внимания привлекают подходы, в основе которых лежит применение для формирования прогнозов информации о событиях, влияющих на значение прогнозируемого показателя на горизонте прогноза. Эта информация, извлекаемая из дополнительных источников,

далеко не всегда представляет собой временные ряды. Она может быть получена в результате:

- анализа текущей ситуации экспертами и аналитиками, прогнозистами и лицами, принимающими решения,
- поиска (по ключевым словам), анализа и обработки информации из разнородных информационных источников.

Поэтому возникает проблема как получить разнотипную информацию и как включать ее в прогноз.

Обобщая результаты анализа основных подходов и направлений количественного прогнозирования применительно к нестационарным процессам со структурными сдвигами на горизонте прогноза, можно сделать вывод, что существующие на сегодня методы количественного прогнозирования не обеспечивают требуемое качество прогноза в силу возрастающей неопределенности в динамике обозначенного объекта на горизонте среднесрочного прогнозирования. На краткосрочном горизонте все три группы методов, обозначенные в данном разделе, способны формировать прогнозы заданной точности; по мере расширения горизонта такая способность снижается (рис. 2).

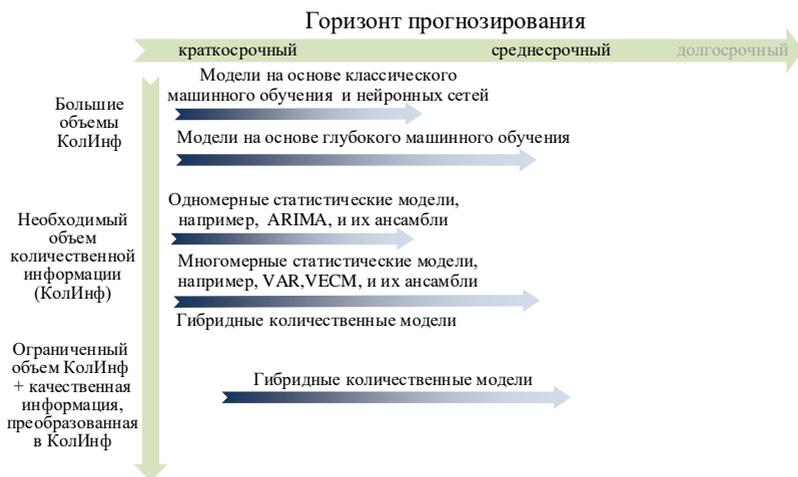


Рис. 2. Применимость количественных методов на горизонте прогнозирования

При этом многомерные статистические модели и их ансамбли, модели на основе глубокого обучения и количественные гибридные количественные модели можно рассматривать как основу для развития интеграционного прогнозирования.

Также важной характеристикой методов количественного прогнозирования является интерпретируемость (объяснимость) полученные результаты их применения для принятий решений (рис. 3).

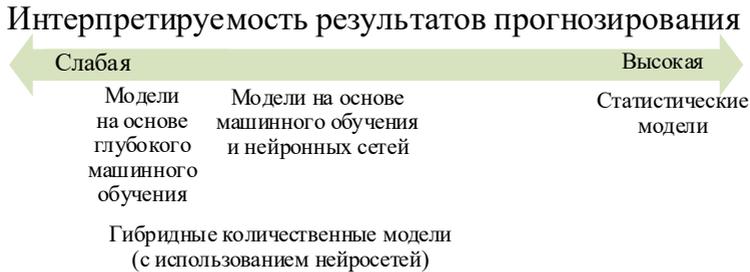


Рис. 3. Интерпретируемость (объяснимость) результатов применения методов для принятия решений

Еще одним важным показателем применимости методов является точность прогноза. Для ее оценки могут быть использованы различные меры оценки ошибки прогнозирования (MAE, MAPE, MSE, RNSE, MAP, R2, MAD и др.). Выбор зависит от специфики решаемой задачи, от используемых методов (в частности, когда прогноз формируется разными методами) и т.п. Статистика использования различных мер точности для разных областей здесь не рассматривается, это тема отдельного обзора.

4. Прогнозирование на основе интеграционных методов, использующих количественные данные, суждения и информацию из разнородных источников

Одно из важных преимуществ интеграции количественных методов и методов извлечения и обработки суждений и инфор-

мации из разнородных источников связано с их способностью обрабатывать различные типы информации.

Интеграционные методы – это методы, позволяющие учитывать разнотипную информацию (структурированную и неструктурированную) на разных этапах решения задачи, и тем самым прямым или косвенным способом включать ее в формируемый прогноз.

4.1. ТИПЫ ИНФОРМАЦИИ И ПОДХОДЫ К ЕЕ ИСПОЛЬЗОВАНИЮ НА РАЗЛИЧНЫХ ЭТАПАХ ПРОЦЕССА ПРОГНОЗИРОВАНИЯ

С увеличением горизонта прогноза возрастает потребность в учете различных типов информации для решения задач прогнозирования нестационарных процессов в условиях структурных сдвигов и повышения точности формируемых прогнозов.

При применении интеграционных методов можно выделить три типа информации (знаний), включаемых в прогноз [40, 43, 50].

Контекстная информация (contextual information) – конкретная информация, доступная в среде прогноза. Это информация, отличная от временных рядов и общего опыта, которая помогает в объяснении, интерпретации и прогнозировании поведения временных рядов. Такая информация (включающая как структурированные данные, так и неструктурированные текстовые данные) может собираться из различных информационных источников с использованием поисковых систем или специализированных программных средств интеллектуального анализа текстов.

Знания предметной области (domain knowledge) – знания, которые практикующие специалисты, эксперты получают благодаря опыту в рамках своей работы. Знание предметной области является результатом применения человеческой интерпретации к контекстной информации (например, информации о значимых событиях внешней среды, которые не отражены в данных текущих наблюдений процесса, но могут проявиться в них только с задержкой).

Технические знания (technical knowledge) – знания формальных методов и процедур анализа данных и прогнозирования.

Суждения, представленные контекстной и/или экспертной информацией, могут включаться на разных этапах процесса прогнозирования [40, 44]:

1) использование суждений при выборе и разработке количественного прогноза (ввод информации в статистический процесс путем выбора соответствующих моделей и параметров);

2) декомпозиция проблемы (задачи) – разбиение ее на части для формирования прогнозов по этим частям с последующим объединением полученных отдельных прогнозов, вместе определяющих прогноз целевого значения прогнозируемого объекта. Формы декомпозиции (мультипликативная или аддитивная) определяются в соответствии с характером проблемы и знаниями о взаимосвязи между ее компонентами [15, 32, 40];

3) комбинирование (совместное использование) качественных прогнозов (на основе суждений) и количественных (цифровых) прогнозов (распространенная линейная форма объединения представлена уравнением (1));

4) субъективная корректировка (judgmental adjustment) количественных прогнозов:

$$(10) \hat{F}_{fin} = \hat{F}_0 + \Delta\hat{F},$$

где \hat{F}_0 – статистический прогноз, полученный формальным методом; $\Delta\hat{F}$ – корректировка прогноза \hat{F}_0 , полученная на основе экспертной информации.

Несмотря на то, что наиболее эффективный способ использования суждений – это ввод данных в статистический процесс путем выбора соответствующих моделей и параметров (подход 1), сделать это часто невозможно в условиях, когда прогнозист получает информацию несвоевременно [50]. К тому же роль суждения в этом подходе недостаточно изучена. Основное внимание уделялось автоматическим (статистическим) методам выбора моделей (в [45] данный подход получил название прогнозирование на основе признаков (feature)), несмотря на то что

специалисты по прогнозированию часто склонны игнорировать рекомендации системы и применять суждения при выборе модели прогнозирования. Поэтому вопрос о выборе подходящих моделей прогнозирования по-прежнему остается актуальным. Кроме того отмечается, что по мере увеличения объема информации специалисты не в состоянии обрабатывать ее эффективно и одновременно, поэтому декомпозицию (подход 2) целесообразно включать на этапе выбора модели [44].

В сообществе прогнозистов применяются два подхода с использованием признаков временных рядов: выбор модели на основе признаков и сочетание моделей на основе признаков. Обоснованием применения указанных подходов является то, что единственная модель не всегда работает наилучшим образом для всех временных рядов. Поэтому вместо выбора одной модели для всех данных предлагается использовать признаки для получения наиболее подходящей модели или оптимальной комбинации моделей-кандидатов. Признак временного ряда может быть любым статистическим представлением характеристик временного ряда. Эти признаки включают сводные данные временных рядов с точки зрения их корреляционной структуры, распределения, стационарности, свойств масштабирования и соответствия набору моделей временных рядов [29, 45].

Для выбора моделей (или их комбинаций) применяют автоматические (статистические) алгоритмы на основе информационных критериев [44] (таких как информационный критерий Акайке (AIC), байесовский информационный критерий (BIC) и т.д.) [22], на основе оценки эффективности различных моделей на наборе валидационных данных (перекрестная валидация) [42]. Наконец в рамках автоматического выбора модели оцениваются различные характеристики временных рядов (таких как тренд, сезонность, случайность, асимметрия, прерывистость, изменчивость, количество доступных наблюдений), а также учитываются переменные для принятия решений (таких как горизонт прогноза). С учетом этих характеристик (признаков) предлагается подход к выбору подходящих моделей на основе экспертных правил ([12, 13, 15, 24]) или с развитием машинного обучения – процедур мета-обучения процесса автоматического

получения знаний для выбора моделей (например, в исследовании [47] применяются деревья решений). Однако такие подходы очень чувствительны к выбранным правилам или функциям мета-обучения. Общепринятого набора таких правил не существует. Также предложен автоматизированный подход к извлечению признаков временных рядов на основе их визуализации, в котором признаки, извлеченные из изображений, используются для оценки оптимальных весов комбинации прогнозов [37].

Все перечисленные подходы для автоматического выбора моделей основаны на статистике или могут быть реализованы с помощью алгоритмического процесса. Общим для них является то, что выбор основан на исторических данных. Ни один из этих алгоритмов не может оценивать прогнозы, когда соответствующие фактические данные еще недоступны [44].

Подходы 3 («комбинирование») и 4 («субъективная корректировка количественных прогнозов») относятся к основным подходам к интеграции суждений со статистическими прогнозами [15, 43] и наиболее представлены в публикациях.

Комбинирование происходит, когда используется некоторая форма процедуры объединения статистических прогнозов и прогнозов на основе суждений [31, 43].

Подход 4 представляет собой апостериорное включение субъективной корректировки количественного прогноза с учетом дополнительной информации.

Субъективная корректировка полученных прогнозов может быть полезной при определенных условиях, когда высокая изменчивость временных рядов ухудшает эффективность статистического моделирования. Эта изменчивость вызвана структурными сдвигами, присущей им случайностью и неопределенностью, которые не могут быть объяснены причинными факторами, но могут быть отражены в контекстной информации (например, о событиях, которые не отражены в модели) [15]. Такая корректировка прогноза может быть в «ручном» режиме, проводимая прогнозистом на основе экспертных суждений, или с использованием формализованных процедур.

Например, в [40] представлен подход к структурированию и формализации экспертных знаний для коррекции статистиче-

ского прогноза из-за предстоящих особых событий. Структурирование достигается с помощью четырех формализованных факторов (transient factors, quantum jump factor, transferred impact factor, trend change factor), определяющих типы динамики, для идентификации событий и соотнесения их с этими факторами. Эксперт-прогнозист оценивает силу (веса) воздействия выявленных корректирующих факторов, ассоциированных с событиями, эти эффекты суммируются и добавляются к исходному статистическому прогнозу. Развитие данного подхода представлено в недавнем исследовании [19], где прогноз спроса на новую продукцию формируется на основе структурирования информации из социальных сетей и субъективной экспертной корректировки прогноза (с использованием корректирующих факторов) с учетом изменения настроений под воздействием событий, ассоциируемых с этими факторами.

Подход 4) относится к типичной практике в компаниях и представляет собой основную конкурирующую альтернативу комбинированному процессу интеграции суждений в прогнозы [15].

Как отмечалось в разделе 3, объединение (подход 3)) наиболее эффективно, когда объединенные прогнозы не коррелированы и привносят в процесс прогнозирования различные виды информации [14, 40]. При этом в [49] показано, что 1) контекстная информация более важна по сравнению с техническими знаниями при объединении прогнозов; 2) чем более нестабильны временные ряды, тем больше контекстной информации требуется для хорошей комбинации.

Интеграционное прогнозирование при объединении прогнозов часто использует ту же идею нахождения весов при комбинации количественных прогнозов (раздел 3), при этом допускается использование суждений. В [7] проведен сравнительный анализ методов формирования весов, использующих экспертную информацию, для объединения частных прогнозов в общий; субъективная корректировка (подход 4) также рассматривается в контексте объединения количественной и экспертной информации при формировании итогового прогноза.

Комбинированный подход (подход 3) к прогнозированию опирается на использование различных типов информации

и/или методов (алгоритмов). В последнее время все большее распространение получают гибридные модели, применение которых подразумевает обработку разных типов информации несколькими моделями для формирования конечного прогноза.

Одним из таких подходов является улучшение прогнозной модели целевого показателя за счет включения в нее неструктурированных данных, обработанных алгоритмами поисковых систем и используемых в качестве объясняющих переменных изменения динамики целевого показателя [30].

В частности, для учета экологических событий в прогнозе авторы [54] предложили методику многомасштабного прогнозирования цен на нефть, используя многофакторные данные поисковых систем в разных временных масштабах.

В [30] представлен обзор применения традиционных статистических моделей и методов машинного обучения для прогнозирования туристического спроса с использованием неструктурированной информации. Отнесение данного подхода к интеграционным методам обусловлено тем, что несмотря на совершенствование поисковых алгоритмов и интеллектуального анализа текстов, остается значимой проблема качественного отбора ключевых слов, являющегося наиболее важным этапом для формирования поисковых запросов в массивах неструктурированной информации. Общепринятого метода, позволяющего систематическим образом осуществлять такой отбор, не существует [30]; как правило, полагаются на экспертное знание предметной области и использование различных терминов, связанных с прогнозируемым процессом.

Перспективным направлением решения задач прогнозирования временных рядов является когнитивный анализ и моделирование ситуаций, о чем свидетельствуют недавние обзоры [26, 41].

В основе этого направления лежит применение нечетких когнитивных карт¹ (НКК) – формализованных моделей пред-

¹ В зарубежных публикациях общепринят термин «нечеткая когнитивная карта» (НКК) (предложен Б. Коско [36]) при исследовании динамических систем в контексте мягких вычислений. НКК представляют собой графовую

ставления знаний о причинно-следственных влияниях между значимыми факторами (концептами) ситуации.

Можно выделить два направления развития методов на основе НКК в этой области.

Одно из них ориентировано на преобразование временного ряда конкретного параметра (или нескольких параметров) в НКК для последующего формирования прогноза динамики этого параметра (параметров) (см., например, [51]) с использованием различных методов обучения.

В контексте целей настоящего обзора представляет интерес второе направление по развитию интеграционных методов – гибридный подход, где НКК используются совместно со статистическими моделями или нейронными сетями. В этом направлении можно выделить ряд исследований, где формирование и/или коррекция прогнозной модели опирается на данные о событиях (инфоповодах) внешней среды, извлеченных из разнородных источников информации и обработанных с использованием НКК (см. [9, 16, 35]).

В [35] представлена гибридная модель – знаковая НКК и нейронная сеть – для прогнозирования рядов финансовых данных. В знаковой НКК отражаются только направления положительного (усиливающего) или отрицательного (ослабляющего) причинно-следственного влияния между факторами (концептами). Гибридная модель составляет основу автоматизированной системы KBNMiner, включающей базу экспертных знаний, формируемую по НКК различных предметных областей; информационно-поисковую подсистему для автоматического сбора новостной информации о событиях из интернета, где направленный поиск организован с использованием концеп-

структуру представления причинно-следственных рассуждений, где нечеткость обусловлена размытой (неоднозначной) степенью причинности между размытыми причинными концептами (факторами) [36] (как правило, степень причинности (сила влияния) между концептами лежит в диапазоне $[-1, 1]$ (или $[0, 1]$)). В отечественных публикациях большее распространение получил термин «когнитивная карта» при том же смысле. В данном обзоре используется НКК независимо от употребляется термина в цитируемых публикациях.

тов НКК; подсистему обработки выделенных положительных и отрицательных событий на основе анализа структуры причинно-следственных влияний НКК для последующего ввода результатов в прогнозную модель нейронной сети в виде параметра, характеризующего относительную силу (положительного или отрицательного) воздействия на прогнозируемый показатель, вместе с другими связанными с ним финансовыми показателями. Эмпирическая проверка на задаче прогнозирования процентных ставок подтвердила, что учет качественной информации оказывает значительное влияние на эффективность нейронной сети при прогнозировании.

В [2, 16] рассматривается задача прогнозирования цен на сырьевые товары на год вперед с разбивкой по месяцам. Сложность решения задачи обусловлена ростом уровня неопределенности прогноза по мере увеличения горизонта прогнозирования. Для уменьшения неопределенности предложена гибридную модель формирования и корректировки ежемесячного прогноза цен на год вперед, которая основана на использовании ансамблей многомерных цифровых моделей временных рядов, построенных в различных временных масштабах. Помимо прогнозируемых показателей и оказывающих влияние на прогноз временных рядов цен и макропоказателей модель использует экспертно-аналитическую информацию, обработанную с помощью НКК ситуации. Для коррекции прогнозной модели на горизонте прогнозирования применяется совместный мониторинг цен на товарных рынках [17], включающий 1) цифровой мониторинг для выявления структурных сдвигов во временном ряду прогнозируемого процесса (и рядах связанных с ним процессов); 2) ситуационный мониторинг внешней среды (по разнородным информационным источникам) и формирование сигналов по результатам сценарного моделирования на НКК возможных последствий влияния значимых внешних событий (инфоповодов) на изменение прогнозируемого процесса и связанных с ним процессов; 3) управление обменом информацией между 1) и 2) и формирование выходных агрегированных сигналов. Работоспособность предложенного подхода подтверждена на примере прогнозирования цен на черный лом на 2019 г.

В [3, 9] предложена модель модулярной системы прогнозирования временных рядов, включающая три основных модуля. Гибридная нейро-нечеткая сеть формирует прогноз временного ряда – количественный прогноз, результаты которого верифицируются по критерию заданной точности. НКК, учитывающая факторы причинно-следственного влияния на прогнозируемый показатель на основе собранных данных о событийном влиянии на временной ряд, формирует прогноз с оценкой влияния события на прогнозируемый показатель. Итоговый прогноз формируется моделью нейронной сети, агрегирующей результаты работы первых двух моделей.

В контексте повышения точности прогнозов также рассматривается возможности сочетания подходов 3) и 4). В [28] представлен подход к комбинированию частных прогнозов, скорректированных экспертами; при этом показано, что комбинированные прогнозы, основанные на модели, могут быть превзойдены комбинированными прогнозами, полученными на основе частных прогнозов, скорректированных экспертами, даже когда сами скорректированные частные прогнозы менее точны, чем прогнозы, лежащие в основе модели.

Таким образом, для решения задачи прогнозирования нестационарных процессов в условиях структурных сдвигов на длительном горизонте прогнозирования H необходимо учитывать различные типы информации:

$$(5) \quad \hat{F}_{fin} = f(\{Y_T^1, \dots, Y_T^k, \dots, Y_T^N\}, H, I_{cont}, I_{exp}, I_{tech}),$$

где $\{Y_T^1, \dots, Y_T^k, \dots, Y_T^N\}$ – количественные данные в виде многомерных рядов; I_{cont} – контекстная информация; I_{exp} – экспертная информация; I_{tech} – техническая информация.

Кроме того, специфика задачи требует применения в качестве методологической базы ее решения – стратегии многошагового (многомасштабного) прогнозирования многомерных рядов.

Обобщая проведенный анализ основных направлений и подходов к интеграционному прогнозированию нестационарных процессов со структурными сдвигами (раздел 3 и раздел 4.1), на рис. 4 систематизированы модели и методы, приме-

нимые для решения задачи среднесрочного прогнозирования таких процессов. Слева приведена стандартная схема решения указанной задачи, справа – модели и методы, которые могут комбинироваться с учетом длины горизонта, типа информации и данных о прогнозируемом процессе, их доступности и полноте; требований к точности прогноза.



Рис. 4. Основные модели и методы, применимые для решения задачи среднесрочного прогнозирования нестационарного процесса со структурными сдвигами

4.2. ОСНОВНЫЕ ПРОБЛЕМЫ РАЗРАБОТКИ И ПРИМЕНЕНИЯ ИНТЕГРАЦИОННЫХ МЕТОДОВ ПРОГНОЗИРОВАНИЯ И ПУТИ ИХ РЕШЕНИЯ

Несмотря на признанную значимость интеграционных методов прогнозирования с включением суждений и неструктурированной информации, основными проблемами их разработки и применения в контексте повышения точности формируемых прогнозов являются предвзятость специалистов по прогнозированию, экспертов и аналитиков, недостаточная технологичность методов структурирования неоднородной информации, необхо-

димых для выделения и формализации полезной информации, формирования и обработки сигналов на ее основе для включения в прогноз [14, 40, 50].

Сложность своевременного получения полезной и достоверной информации о характере и тенденциях изменения процессов воздействия внешней среды на объект прогнозирования обусловлена, с одной стороны, недостатком информации и понимания существенных разнородных факторов – движущих сил, определяющих изменение этих процессов, а с другой стороны – проблемой большого «шума» в данных, характеризующих исследуемые процессы, из-за их избыточной доступности. Такая доступность обусловлена большим количеством источников и порождаемыми ими большими потоками информации в условиях повсеместного внедрения цифровых и информационно-телекоммуникационных технологий, что может привести к информационной перегрузке, влияющей на способность специалистов определять достоверность, актуальность и надежность имеющейся информации [57]. К тому же обработка в таких объемах и не требуется прогнозисту для решения поставленных задач. Кроме того, контекстная информация часто представляется в неколичественной форме, и ее количественная оценка может быть сложной задачей. Поэтому такую информацию сложно интегрировать с данными временных рядов.

Путь решения обозначенных проблем интеграционного прогнозирования специалисты и исследователи в области анализа и прогнозирования временных рядов в контексте повышения точности прогнозов видят в совершенствовании систем поддержки прогнозирования (СПП), предназначенных не для замены прогнозистов, а для структурированной и систематической поддержки деятельности прогнозистов, экспертов и аналитиков в жизненном цикле решения практических задач прогнозирования [10, 11, 15, 27].

Согласно [15] СПП определяется как программное обеспечение, фреймворк (готовый набор инструментальных средств) и/или структурированная процедура, которые могут использовать экспертные суждения, разнотипную информацию и количественные методы для интеграции исторических данных

и контекстной информации с целью оказания значимой поддержки прогнозисту для формирования более точных прогнозов и анализа результатов.

На практике СПП не получили широкого распространения. Большинство компаний используют коммерческие пакеты для прогнозирования спроса на продукцию. Прогнозы, создаваемые этими пакетами, часто корректируются на основе управленческие суждений. Причинами медленного развития и внедрения практических СПП являются: высокая стоимость разработки и внедрения СПП; сложность обслуживания и модернизации СПП в условиях быстрых изменений в типе и объеме данных и информации. Еще одна проблема связана с тем, что существующие СПП систематически не собирают и не включают экспертную информацию в прогнозы, что приводит к системному недоверию и ручной корректировке результатов прогнозистами [15].

Так как суждения играют важную роль в жизненном цикле решения практических задач прогнозирования, начиная от выбора моделей и до коррекции прогноза, их невозможно исключить из этого процесса даже при наличии хорошо продуманной СПП [43, 44]. На рис. 5 представлена схема расширенного информационного обеспечения основных этапов прогнозирования нестационарных процессов в условиях структурных сдвигов (в пунктирном прямоугольнике выделен необязательный этап – декомпозиция задачи).

Как видно на рис. 5, экспертные знания являются основой информационного обеспечения, так как используются на все этапах жизненного цикла прогнозирования, но следует отметить, что периодичность их актуализации на разных этапах различна.

СПП может рассматриваться как способ интеграции моделей и методов прогнозирования с учетом разнотипной информации, включаемой в прогноз.

Вопросы разработки принципов и требований к формированию функционала такой СПП в контексте решения задачи прогнозирования нестационарных процессов в условиях структурных сдвигов не рассматриваются в данном обзоре, так как

это требует серьезной проработки и не является предметом данного обзора.

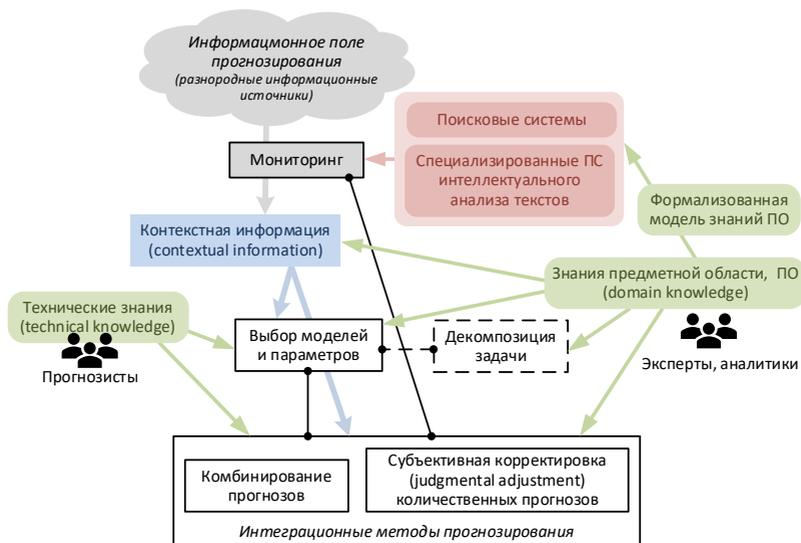


Рис. 5. Схема расширенного информационного обеспечения основных этапов решения задачи прогнозирования нестационарных процессов в условиях структурных сдвигов

Однако можно выделить ряд общих взаимосвязанных принципов ее построения для решения указанных задач в контексте повышения точности прогнозов на заданном горизонте прогнозирования.

Соответствие функционала сложности решаемых задач, адаптивность СПП. Понимание характеристик задач и соответствующих потребностей в поддержке позволяет создавать СПП, которые лучше соответствуют потребностям пользователей и адаптируются к ним.

Технологичность процесса решения задачи прогнозирования, обеспечивающая систематическую и структурированную процедуру решения задачи. К тому же подгонка технологической поддержки к характеристикам и особенностям задач спо-

способствует выявлению пробелов между функциональностью системы и потребностями пользователей.

Поддержка многомодельности (гибридный подход) в СПП, дающая возможность работы со структурированной и неструктурированной информацией (рис. 5) в жизненном цикле решения задачи прогнозирования. Для облегчения выбора моделей – поддержка теоретическими знаниями по свойствам и поведению моделей; накопление базы знаний по положительному и отрицательному опыту применения моделей.

Включение модели представления знаний предметной области с целью повышения качества направленного поиска и фильтрации неструктурированной информации при выделении полезной контекстной информации для формируемого прогноза. Например, в качестве моделей представления знаний можно использовать НКК, графы знаний.

Снижение когнитивной нагрузки и ресурсоемкости обработки информации человеком. СПП, которые дополняют возможности человека по обработке информации, могут повысить эффективность решений, поскольку они снижают когнитивную нагрузку на человека при решении сложных задач среднесрочного и долгосрочного прогнозирования.

Информативность, позволяющая обеспечить постоянный анализ результатов в жизненном цикле решения задачи.

5. Заключение

Несмотря на многолетнюю историю исследований в области прогнозирования с использованием суждений (Forecasting with Judgment), публикуемых зарубежных обзорах, затрагивающих различные аспекты этого направления, актуальность его развития сохраняется, особенно для решения задач среднесрочного и долгосрочного прогнозирования процессов в сложных динамических системах. В этих системах протекают нестационарные процессы, динамика которых может изменяться не только вследствие эндогенных изменений, но и под воздействием экзогенных факторов в окружающей среде объекта на длительном горизонте прогнозирования. В таких условиях повыше-

ние точности прогноза зависит не только от используемых количественных методов прогнозирования, но и от методов, способных обнаруживать и учитывать информацию о будущих событиях, изменяющих динамику целевого показателя в интервале между прогнозами, отсутствующей в данных на момент составления прогноза.

Обобщая результаты проведенного обзора, можно сделать вывод, что для задач прогнозирования нестационарных процессов в условиях структурных сдвигов на длительном горизонте прогнозирования является перспективным применение интеграционных методов. Такие методы должны сочетать в себе возможности гибридных моделей, комбинирования современных количественных методов и методов обработки всей доступной полезной контекстной и экспертной информации об объекте прогнозирования, учитываемой на разных этапах решения задачи прогнозирования.

При этом, учитывая сложность решения задач среднесрочного и долгосрочного прогнозирования для таких объектов, необходимость решения проблемы полноты и достаточности включаемой разнородной информации в формируемый прогноз, развитие таких методов перспективно по пути их интеграции с СПП.

Литература

1. АВДЕЕВА З.К., КОВРИГА С.В., ГРЕБЕНЮК Е.А. *Анализ методов среднесрочного прогнозирования процессов со структурными сдвигами на финансовых и товарных рынках* // Проблемы управления. – 2024. – №5 (в печати).
2. АВДЕЕВА З.К., КОВРИГА С.В., ГРЕБЕНЮК Е.А. *Формирование среднесрочных месячных прогнозов цен на сырье на основе экспертной и количественной информации* // Автоматизация в промышленности. – 2022. – №5. – С. 38–45.
3. АВЕРКИН А.Н., ЯРУШЕВ С.А., ПАВЛОВ В.Ю. *Когнитивные гибридные системы поддержки принятия решений и прогнозирования* // Программные продукты и системы. – 2017. – №4. – С. 632–642.

4. АНДРИАНОВА Е.Г., ГОЛОВИН С.А., ЗЫКОВ С.В. и др. *Обзор современных моделей и методов анализа временных рядов динамики процессов в социальных, экономических и социотехнических системах* // Russian Technological Journal. – 2020. – №8(4). – С. 7–45. – DOI: <https://doi.org/10.32362/2500-316X-2020-8-4-7-45>.
5. АСТАФЬЕВА Е.В., ТУРУНЦЕВА М.Ю. *Анализ возможностей улучшения качества прогнозов цен на природные ресурсы методами комбинирования на основе регрессионных оценок весов* // Экономическое развитие России. – 2023. – Т. 30, №12. – С. 24–33.
6. СЕМЕНОВ А.В. *Анализ временных рядов в политической науке: возможности и ограничения* // Политическая наука. – 2021. – №1. – С. 76–97.
7. СУРКОВ А.А. *Объединение экономических прогнозов с использованием экспертной информации* // Статистика и экономика. – 2019. – №16(5). – С. 4–14.
8. ФРЕНКЕЛЬ А.А., СУРКОВ А.А. *Методологические подходы к улучшению точности прогнозирования путем объединения прогнозов* // Вопросы статистики. – 2015. – №8. – С. 17–36.
9. ЯРУШЕВ С.А., АВЕРКИН А.Н., ФЕДОТОВА А.В. *Модулярная модель прогнозирования временных рядов на основе нейро-нечетких сетей и когнитивного моделирования* // Нечеткие системы и мягкие вычисления. – 2017. – Т. 12, вып. 2. – С. 159–168.
10. ABOLGHASEMI M., HURLEY J., ESHRAGH A. et al. *Demand forecasting in the presence of systematic events: Cases in capturing sales promotions* // Int. Journal of Production Economics. – 2020. – Vol. 230. – 107892.
11. ADYA M., LUSK E.J. *Designing Effective Forecasting Decision Support Systems: Aligning Task Complexity and Technology Support* / C. Jao (ed.) Decision Support Systems. – IntechOpen Limited, United Kingdom, 2012.
12. ADYA M., LUSK E.J. *Development and validation of a rule-based time series complexity scoring technique to support design of adaptive forecasting DSS* // Decision Support Systems. – 2016. – No. 83(3). – P. 70–82

13. ARINZE B. *Selecting appropriate forecasting models using rule induction* // Omega. – 1994. – No. 22(6). – P. 647–658.
14. ARMSTRONG J.S. *Combining forecasts* // In: Principles of Forecasting: A Handbook for Researchers and Practitioners / J.S. Armstrong (ed.). – Springer Nature, 2001. – P. 417–440.
15. ARVAN M., FAHIMNIA B., REISI M. et al. *Integrating human judgement into quantitative forecasting methods: A review* // Omega. – 2019. – Vol. 86. – P. 237–252. – DOI: <https://doi.org/10.1016/j.omega.2018.07.012>.
16. AVDEEVA Z.K., GREBENYUK E.A., KOVRIGA S.V. *Raw Material Price Forecasting on Commodity Markets: Application of Expert and Quantitative Information* // Advances in Systems Science and Applications. – 2022. – Vol 22, No. 4. – P. 126–143.
17. AVDEEVA Z.K., GREBENYUK E.A., KOVRIGA S.V. *Detection of structural shifts in commodity markets in the mode of situation and digital monitoring* // IFAC-PapersOnLine. – 2023. – Vol. 56, Iss. 2. – P. 7778–7783.
18. AZMY W.M., ATIYA A.F., EL-SHISHINY H. *Forecast Combination Strategies for Handling Structural Breaks for Time Series Forecasting* // In: Multiple Classifier Systems. MCS 2010. Lecture Notes in Computer Science / Eds.: N.E. Gayar, J. Kittler, F. Roli. – Springer, Berlin, Heidelberg. – 2010. – Vol. 5997. – DOI: https://doi.org/10.1007/978-3-642-12127-2_25.
19. BADULESCU Y., CAÑAS F., CHEIKHROUHOU N. *Judgmental adjustment of demand forecasting models using social media data and sentiment analysis within industry 5.0 ecosystems* // Int. Journal of Information Management Data Insights. – 2024. – Vol. 4, Iss. 2. – 100272.
20. BATES J.M., GRANGER C.W.J. *The combination of forecasts* // Journal of the Operational Research Society. – 1969. – No. 20(4). – P. 451–468.
21. BAYÜKŞAHINA Ü., ERTEKIN S. *Improving forecasting accuracy of time series data using a new ARIMA-ANN hybrid method and empirical mode decomposition* // Neurocomputing. – 2019. – Vol. 361. – P. 151–163.

22. BRUNHAM K.P., ANDERSON D.R. *Model selection and multi-model inference: A practical information-theoretic approach*. – Springer-Verlag, Heidelberg, 2002. – 490 p.
23. CHENG C., SA-NGASOONGSONG A., BEYCA O. et al. *Time series forecasting for nonlinear and non-stationary processes: a review and comparative study* // IISE Trans. – 2015. – Vol. 47(10). – P. 1053–1071.
24. COLLOPY F., ARMSTRONG J.S. *Rule-Based Forecasting: Development and Validation of an Expert Systems Approach to Combining Time Series Extrapolations* // Management Science. – 1992. – No. 38(10). – P. 1394–1414.
25. DE STEFANI J., BONTEMPI G. *Factor-Based Framework for Multivariate and Multi-step-ahead Forecasting of Large-Scale Time Series* // Front. Big Data. – 2021. – Vol. 4. – 690267.
26. FELIX G., NÁPOLES G., FALCON R. et al. *A review on methods and software for fuzzy cognitive maps* // Artif Intell Rev. – 2019. – Vol. 52. – P. 1707–1737.
27. FILDES R., GOODWIN P., LAWRENCE M. *The design features of forecasting support systems and their effectiveness* // Decision Support Systems. – 2006. – Vol. 42, Iss. 1. – P. 351–361.
28. FRANCES F.H., VAN DIJK D. *Combining expert adjusted forecasts* // Journal of Forecasting. – 2019. – Vol. 38, Iss. 5. – P. 415–421.
29. FULCHER B.D., JONES N.S. *Highly comparative feature-based time-series classification* // IEEE Trans. on Knowledge and Data Engineering. – 2014. – No. 26(12). – P. 3026–3037.
30. GEORGAKIS A., PROFILIDIS V., BOTZORIS G.N. *Forecasting tourism demand using search engine data* // Proc. of the 10th Int. Congress on Transportation Research (ICTR2021). – Rhodes, Greece. – 2021.
31. GOODWIN P. *Integrating management judgment and statistical methods to improve short-term forecasts* // Omega. – 2002. – Vol. 30, Iss. 2. – P. 127–135.
32. GOODWIN P., FILDES R. *Forecasting with Judgment* // In: The Palgrave Handbook of Operations Research / Eds.: S. Salhi, J. Boylan. – Palgrave Macmillan, Cham, 2022. – DOI: https://doi.org/10.1007/978-3-030-96935-6_16.

33. GREBENYUK E.A. *Algorithms for Detecting Changes in the Properties of Non-Stationary Processes in the Monitoring Mode* // Proc. of the 14th Int. Conf. "Management of Large-Scale System Development" (MLSD). – Moscow: IEEE. – 2021. – URL: <https://ieeexplore.ieee.org/document/9600109>.
34. HOGARTH R.M., MAKRIDAKIS S. *Forecasting and planning: an evaluation* // Management science. – 1981. – No. 27(2). – P. 115–138.
35. HONG T., HAN I. *Integrated approach of cognitive maps and neural networks using qualitative information on the World Wide Web: the KBNMiner* // Expert Systems. – 2004. – No. 21(5). – P. 243–252.
36. KOSKO B. *Fuzzy cognitive maps* // Int J Man Mach Stud. – 1986. – No. 24(1). – P. 65–75.
37. LI X., KANG Y., LI F. *Forecasting with time series imaging* // Expert System with Applications. – 2020. – Vol. 160. – 113680.
38. LIVIERIS I.E., PINTELAS P. *A novel multi-step forecasting strategy for enhancing deep learning models' performance* // Neural Computing and Applications. – 2022. – Vol. 34. – P. 19453–19470.
39. MAKRIDAKIS S., HOGARTH R., GAB A. *Forecasting and uncertainty in the economic and business world* // Int. Journal of Forecasting. – 2009. – Vol. 25. – P. 794–812.
40. MARMIER F., CHEIKHROUHOU N. *Structuring and integrating human knowledge in demand forecasting: a judgmental adjustment approach* // Production Planning & Control. – 2010. – No. 21(4). – P. 399–412.
41. ORANG O., DE LIMA E SILVA P.C., GUIMARÃES F.G. *Time series forecasting using fuzzy cognitive maps: A survey* // Artif. Intell. Rev. – 2022. – DOI: <https://doi.org/10.1007/s10462-022-10319-w>.
42. ORD J.K., FILDES R., KOURENTZES N. *Principles of Business Forecasting*. – Wessex Press Publishing Co, 2017.
43. PERERA H.N., HURLEY J., FAHIMNIA B. et al. *The human factor in supply chain forecasting: A systematic review* // European Journal of Operational Research. – 2019. – Vol. 274(2). – P. 574–600.

44. PETROPOULOS F., KOURENTZES N., NIKOLOPOULOS K. et al. *Judgmental selection of forecasting models* // Journal of Operations Management. – 2018. – No. 60. – P. 34–46.
45. PETROPOULOS F., APILETTI D. et al. *Forecasting: theory and practice* // Int. Journal of Forecasting. – 2022. – Vol. 38(3). – P. 705–871.
46. POLAT C. *Forecasting as a strategic decision-making tool: A review and discussion with emphasis on marketing management* // European Journal of Scientific Research. – 2008. – Vol. 20(2). – P. 419–442.
47. PRUDÊNCIO R.B., LUDERMIR T.B. *Meta-learning approaches to selecting time series models* // Neurocomputing. – 2004. – No. 61. – P.121–137.
48. SAHOO D., SOOD N., RANI U. et al. *Comparative Analysis of Multi-Step Time-Series Forecasting for Network Load Dataset* // Proc. of the 11th Int. Conf. on Computing, Communication and Networking Technologies. – Kharagpur, India. – 2020. – P. 1–7. – DOI: <https://doi.org/10.1109/ICCCNT49239.2020.9225449>.
49. SANDERS N.R., RITZMAN L.P. *Bringing judgment into combination forecasts* // J. Oper. Manag. – 1995. – No. 13. – P. 311–321.
50. SANDERS N.R., RITZMAN L.P. *Judgmental Adjustment of Statistical Forecasts* // In: Principles of Forecasting: A Handbook for Researchers and Practitioners / J.S. Armstrong (ed.). – Springer Nature, 2001. – P. 405–416.
51. SHAN D., LU W., YANG J. *The data-driven fuzzy cognitive map model and its application to prediction of time series* // Int. Journal of Innovative Computing, Information and Control. – 2018. – No. 14(5). – P. 1583–1602.
52. TAIEB S., BONTEMPI G., ATIYA A. et al. *A review and comparison of strategies for multi-step ahead time series forecasting based on the NN5 forecasting competition* // Expert System with Applications. – 2012. – Vol. 39(8). – P. 7067–7083. – DOI: <https://doi.org/10.1016/j.eswa.2012.01.039>.
53. TAIEB S.B., ATIYA A.F. *A bias and variance analysis for multistep-ahead time series forecasting* // IEEE Trans. Neural Netw. Learn. Syst. – 2016. – No. 27. – P. 62–76.

54. TANG L., ZHANG C., LI L. et al. *A multi-scale method for forecasting oil price with multi-factor search engine data* // Applied Energy. – 2020. – Vol. 257. – 114033. – DOI: <https://doi.org/10.1016/j.apenergy.2019.114033>.
55. THOMSON M.E., POLLOCK A.C., ÖNKAL D. et al. *Combining forecasts: Performance and coherence* // Int. Journal of Forecasting. – 2019. – Vol. 35, Iss. 2. – P. 474–484.
56. WANG X., HYNDMAN R.J., LI F. et al. *Forecast combinations: An over 50-year review* // Int. Journal of Forecasting. – 2023. – Vol. 39, Iss. 4. – P. 1518–1547. – DOI: <https://doi.org/10.1016/j.ijforecast.2022.11.005>.
57. WEBBY R., O'CONNOR M., LAWRENCE M. *Time-series forecasting using domain knowledge* // In: Principles of Forecasting: A Handbook for Researchers and Practitioners / J.S. Armstrong (ed.). – Springer Nature, 2001. – P. 389–403.
58. WEBBY R., O'CONNOR M., EDMUNDSON B. *Forecasting support systems for the incorporation of event information: an empirical investigation* // Int. J. Forecast. – 2005. – Vol. 21. – P. 411–423.

INTEGRATING FORECASTING OF NON-STATIONARY PROCESSES REPRESENTED BY TIME SERIES. OVERVIEW

Zinaida Avdeeva, V.A. Trapeznikov Institute of Control Sciences of RAS, Moscow, Cand. Sc., Senior Researcher (avdeeva@ipu.ru).

Svetlana Kovriga, V.A. Trapeznikov Institute of Control Sciences of RAS, Moscow, Researcher (kovriga@ipu.ru).

Abstract: The review covers the main directions and approaches to integration forecasting of non-stationary processes represented by time series. The key source of non-stationarity generation is rapid and poorly predictable changes in the external environment, under the influence of which structural shifts occur in the complex economic and socio-political processes. The solution of problems of forecasting the dynamics of such objects in the context of improving the accuracy of the forecast becomes more complex as the forecast horizon increases. It determines the need for models and methods capable of processing heterogeneous information. Integration methods are methods that allow, along with quantitative data, to take into account judgments (of forecasters, experts, analysts) and information from heterogeneous

information sources at different stages of problem solving, and thus directly or indirectly include them in the forecast being formed. The development of such methods is aimed at increasing the accuracy of the forecast through the use of all available information about the forecasting object, including data on endogenic and exogenic factors of influence on it. The review focused on the current state of integration forecasting, on the existing problems and ways to solve them.

Keywords: non-stationary processes, time series, forecasting, judgments.

УДК 338.27

ББК 65.054

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

Поступила в редакцию 01.10.2024.

Опубликована 30.11.2024.

ПРИМЕНЕНИЕ АППАРАТА МАТЕМАТИЧЕСКОГО ПРОГРАММИРОВАНИЯ ДЛЯ ПОСТРОЕНИЯ НЕЭЛЕМЕНТАРНЫХ КВАЗИЛИНЕЙНЫХ РЕГРЕССИЙ

Базилевский М. П.¹

(ФГБОУ ВО Иркутский государственный университет
путей сообщения, Иркутск)

В неэлементарных квазилинейных регрессиях объясняющие переменные сначала преобразуются с помощью элементарных функций, после чего пары полученных факторов снова преобразуются с помощью неэлементарных функций \tan и \tanh . Такие модели нелинейны как по факторам, так и по параметрам, поэтому даже их оценивание представляется сложной вычислительной задачей. А если неизвестен состав входящих в модель переменных, а также их элементарные и неэлементарные преобразования, то сложность задачи существенно возрастает. На решение этой проблемы направлено данное исследование. Вместо трудоёмких переборных процедур использован хорошо развитый за последнее время аппарат математического программирования. Метод построения неэлементарных квазилинейных регрессий формализован в виде задачи частично-булевого линейного программирования. Предложенный метод реализован в специальной компьютерной программе. Её достоинство в том, что пользователь может регулировать в процессе построения число преобразованных переменных, поэтому программой можно пользоваться как для решения простых задач управления на обычных персональных компьютерах, так и для обработки массивов больших данных с помощью облачных сервисов. Неэлементарные квазилинейные регрессии могут быть использованы для решения задач управления в технических, социально-экономических, медицинских и других системах.

Ключевые слова: регрессионный анализ, нелинейная регрессия, неэлементарная квазилинейная регрессия, метод наименьших квадратов, задача частично-булевого линейного программирования.

1. Введение

С помощью машинного обучения [15, 18] в настоящее время решается множество различных задач управления в технике, экономике, медицине, строительстве и других областях человеческой деятельности. Обученные модели используются, в частности, для прогнозирования будущего состояния функциониру-

¹ Михаил Павлович Базилевский, к.т.н., доцент (mik2178@yandex.ru).

вания исследуемого объекта или процесса. При этом для обучения с целью прогнозирования существует множество типов моделей: деревья решений, алгоритмы кластеризации и пр., но исследователи зачастую выбирают искусственные нейронные сети (ИНС) [22], либо регрессионные модели [16]. Нельзя сказать однозначно, прогнозы по какой из этих двух разновидностей моделей точнее на практике, поскольку все зависит от конкретной ситуации. Данная статья посвящена вопросам автоматизации процесса построения регрессионных моделей.

Давно миновали те времена, когда оценивание линейной регрессии с помощью метода наименьших квадратов (МНК), например, по выборке из ста наблюдений для пяти объясняющих переменных, считалось весьма солидной вычислительной задачей. На современных среднестатистических компьютерах такая задача решается практически мгновенно. Однако линейные регрессии, благодаря своей простой содержательной интерпретации, сегодня всё же находят применение (см., например, [23, 24]). Но большинство реальных объектов и процессов в мире подчиняется нелинейным законам и закономерностям, для описания которых линейных регрессий недостаточно, поэтому приходится прибегать к оцениванию нелинейных зависимостей.

Проведен анализ следующих прикладных научных работ, посвященных нелинейному регрессионному моделированию. В [17] решается проблема поддержки оптимальной температуры внутри помещения, для чего строится нелинейная регрессия зависимости между энергопотреблением здания и температурой наружного воздуха. В [19] разработана трехслойная ИНС и модели нелинейной регрессии для прогнозирования скорости производства биогаза из анаэробного гибридного реактора. При этом оба вида моделей хорошо справились с прогнозированием. В [20] прогнозируются ключевые показатели (осадка гребня, внутренняя осадка и прогиб лицевой плиты) деформаций при строительстве каменно-насыпных плотин с бетонным замком. В [25] представлен анализ и сравнение процесса отслеживания дронов на основе линейных фильтров Калмана по сравнению с нелинейной полиномиальной регрессией. При этом сделан вывод, что оба метода целесообразно использовать в разных усло-

виях шумовых измерений. Работы [8, 11] посвящены прогнозированию прочности бетона на сжатие. В [11] обычный заполнитель в бетонной смеси предлагается заменить переработанными отходам резиновых шин, а в [8] – частично заменить цемент порошком стеклянных отходов. В обоих случаях ИНС оказались несколько лучше по качеству, чем нелинейные регрессии. Хорошие результаты при прогнозировании прочности бетона на сжатие с помощью квазилинейной регрессии продемонстрированы в [5]. Помимо этого учеными ведутся фундаментальные исследования в области нелинейного регрессионного моделирования. Так, в [10] решается проблема усреднения моделей регрессии и исследуется критерий нелинейной информации. А в [9] разработана так называемая оптимальная остаточная регрессия.

Проведенный анализ показывает, что в качестве нелинейных регрессий исследователи зачастую применяют традиционные формы связи между переменными – степенные, полиномиальные, логарифмические и пр. Однако в [2, 3] на основе функции Леонтьева [6] были разработаны новые, показывающие хорошие результаты на практике модели – неэлементарные линейные регрессии (НЛинР), которые линейны по объясняющим переменным, но нелинейны по параметрам. В зависимости от количества объясняющих переменных НЛинР могут иметь большое количество регрессоров, поэтому при моделировании целесообразно решать задачу отбора информативных регрессоров (ОИР) [13, 14]. Для этого в [1] был использован хорошо развитый в последнее время аппарат частично-булевого линейного программирования (ЧБЛП) [21]. Тем самым задача ОИР в НЛинР, оцениваемой с помощью МНК, была сведена к задаче ЧБЛП. Исходя из этого, в работе [4] были предложены неэлементарные квазилинейные регрессии (НКЛинР), в которых объясняющие переменные преобразуются с помощью элементарных математических функций. Там же подробно описана первая версия программы ВИнтер-2 для построения НКЛинР. Однако формализация задачи ОИР в НКЛинР в виде задачи ЧБЛП до сегодняшнего дня нигде не была описана. К тому же ещё никогда, за исключением небольшого примера в [4], не решалась

задача построения НКЛинР по реальным данным. На решение указанных проблем направлено данное исследование.

2. Оптимизационная задача построения НКЛинР

Предположим, что имеется выборочная совокупность объема n , соответствующая измерениям зависимой (объясняемой) переменной y , и l независимых (объясняющих) переменных x_1, \dots, x_l . Тогда предложенная в [1] НЛинР с бинарными операциями \min и \max и $(1 + l + 4C_l^2)$ неизвестными параметрами $\alpha_0, \alpha_1, \dots, \alpha_l, \alpha_1^{\min}, \dots, \alpha_{C_l^2}^{\min}, \alpha_1^{\max}, \dots, \alpha_{C_l^2}^{\max}, k_1^{\min}, \dots, k_{C_l^2}^{\min}, k_1^{\max}, \dots, k_{C_l^2}^{\max}$, имеет вид

$$(1) \quad y_i = \alpha_0 + \sum_{j=1}^l \alpha_j x_{ij} + \sum_{j=1}^{C_l^2} \alpha_j^{\min} \min \{ x_{i,\mu_{j,1}}, k_j^{\min} x_{i,\mu_{j,2}} \} + \\ + \sum_{j=1}^{C_l^2} \alpha_j^{\max} \max \{ x_{i,\mu_{j,1}}, k_j^{\max} x_{i,\mu_{j,2}} \} + \varepsilon_i, \quad i = 1, 2, \dots, n,$$

где C_l^2 – число сочетаний без повторений из l элементов по 2; $\varepsilon_1, \dots, \varepsilon_n$ – ошибки аппроксимации регрессии; $\mu_{j,1}, \mu_{j,2}, j = 1, 2, \dots, C_l^2$, – элементы индексной матрицы \mathbf{M} размера $C_l^2 \times 2$, содержащей в строках все возможные комбинации (сочетания) пар индексов объясняющих переменных.

Введем множество F , состоящее из числа $elem$ элементарных математических функций, т.е. $F = \{f_1(x), f_2(x), \dots, f_{elem}(x)\}$. Преобразуя каждую объясняющую переменную с помощью этого множества, получим расширенный набор из $l \times elem$ переменных. Это известный приём, описанный в монографии [7]. В качестве элементарных функций можно использовать $x^2, \sqrt{x}, \ln(x), e^x$ и т.д.

Предложенная в [4] для введенного расширенного набора переменных НКЛинР с $(1 + l \cdot elem + 4C_{l \cdot elem}^2)$ неизвестными параметрами $\alpha_0, \alpha_{jk}, j = 1, 2, \dots, l, k = 1, 2, \dots, elem, \alpha_j^{\min}, \alpha_j^{\max}, k_j^{\min}, k_j^{\max}, j = 1, 2, \dots, C_{l \cdot elem}^2$, имеет вид

$$\begin{aligned}
 (2) \quad y_i &= \alpha_0 + \sum_{j=1}^l \sum_{k=1}^{elem} \alpha_{jk} f_k(x_{ij}) + \\
 &+ \sum_{j=1}^{C_{l,elem}^2} \alpha_j^{min} \min \left\{ f_{\mu_{j,1,2}}(x_{i,\mu_{j,1,1}}), k_j^{min} f_{\mu_{j,2,2}}(x_{i,\mu_{j,2,1}}) \right\} + \\
 &+ \sum_{j=1}^{C_{l,elem}^2} \alpha_j^{max} \max \left\{ f_{\mu_{j,1,2}}(x_{i,\mu_{j,1,1}}), k_j^{max} f_{\mu_{j,2,2}}(x_{i,\mu_{j,2,1}}) \right\} + \varepsilon_i, \\
 i &= 1, 2, \dots, n,
 \end{aligned}$$

где $\mu_{j,1,1}, \mu_{j,1,2}, \mu_{j,2,1}, \mu_{j,2,2}, j = 1, 2, \dots, C_{l,elem}^2$ – элементы трехмерного массива (куба) \mathbf{M}^* размера $C_{l,elem}^2 \times 2 \times 2$, в котором первое измерение – «Номер пары преобразованных переменных», второе – «Индексы переменных», третье – «Индексы преобразований». Иными словами, горизонтальные срезы куба представляет собой двухмерные массивы (матрицы) двух измерений – «Индексы переменных» и «Индексы преобразований».

НКЛинР (2) также можно называть неэлементарной регрессией с элементарными преобразованиями переменных.

Например, если имеется три переменных x_1, x_2, x_3 , а множество $F = \{x^3, 2^x\}$, то без использования операции \max НКЛинР (2) примет вид

$$\begin{aligned}
 y_i &= \alpha_0 + \alpha_{11} x_{i1}^3 + \alpha_{12} 2^{x_{i1}} + \alpha_{21} x_{i2}^3 + \alpha_{22} 2^{x_{i2}} + \alpha_{31} x_{i3}^3 + \alpha_{32} 2^{x_{i3}} + \\
 &+ \alpha_1^{min} \min \{x_{i1}^3, k_1^{min} x_{i2}^3\} + \alpha_2^{min} \min \{x_{i1}^3, k_2^{min} x_{i3}^3\} + \alpha_3^{min} \min \{x_{i1}^3, k_3^{min} 2^{x_{i1}}\} + \\
 &+ \alpha_4^{min} \min \{x_{i1}^3, k_4^{min} 2^{x_{i2}}\} + \alpha_5^{min} \min \{x_{i1}^3, k_5^{min} 2^{x_{i3}}\} + \alpha_6^{min} \min \{x_{i2}^3, k_6^{min} x_{i3}^3\} + \\
 &+ \alpha_7^{min} \min \{x_{i2}^3, k_7^{min} 2^{x_{i1}}\} + \alpha_8^{min} \min \{x_{i2}^3, k_8^{min} 2^{x_{i2}}\} + \alpha_9^{min} \min \{x_{i2}^3, k_9^{min} 2^{x_{i3}}\} + \\
 &+ \alpha_{10}^{min} \min \{x_{i3}^3, k_{10}^{min} 2^{x_{i1}}\} + \alpha_{11}^{min} \min \{x_{i3}^3, k_{11}^{min} 2^{x_{i2}}\} + \alpha_{12}^{min} \min \{x_{i3}^3, k_{12}^{min} 2^{x_{i3}}\} + \\
 &+ \alpha_{13}^{min} \min \{2^{x_{i1}}, k_{13}^{min} 2^{x_{i2}}\} + \alpha_{14}^{min} \min \{2^{x_{i1}}, k_{14}^{min} 2^{x_{i3}}\} + \\
 &+ \alpha_{15}^{min} \min \{2^{x_{i2}}, k_{15}^{min} 2^{x_{i3}}\} + \varepsilon_i, \quad i = 1, 2, \dots, n,
 \end{aligned}$$

для которой вертикальный срез куба \mathbf{M}^* измерений «Номер пары преобразованных переменных» и «Индексы переменных» –

$$\begin{pmatrix}
 1 & 1 & 1 & 1 & 1 & 2 & 2 & 2 & 2 & 3 & 3 & 3 & 1 & 1 & 2 \\
 2 & 3 & 1 & 2 & 3 & 3 & 1 & 2 & 3 & 1 & 2 & 3 & 2 & 3 & 3
 \end{pmatrix}^T,$$

а вертикальный срез измерений «Номер пары преобразованных переменных» и «Индексы преобразований» –

$$\begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 2 & 2 & 2 \\ 1 & 1 & 2 & 2 & 2 & 1 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 \end{pmatrix}^T.$$

Как видно, даже при $l = 3$ и $elem = 2$ в НКЛинР (2) уже содержится слишком много регрессоров (21 штука). А при использовании бинарной операции \max их количество ещё существенно возрастет (до 36 штук), что серьезно осложнит процесс оценивания модели. Поэтому возникает необходимость выбора из расширенного набора преобразованных переменных только наиболее информативных из них в некотором смысле. Формализуем далее такую задачу ОИР в виде задачи ЧБЛП так, как это сделано для НЛинР (1) в [1].

Для удобства запишем НКЛинР (2) в виде

$$(3) \quad y_i = \alpha_0 + \sum_{j=1}^l \sum_{k=1}^{elem} \alpha_{jk} \cdot w_{ijk} + \\ + \sum_{j=1}^{C_{l \cdot elem}^2} \alpha_j^{min} \min \left\{ w_{i, \mu_{j,1,1}, \mu_{j,1,2}}, k_j^{min} \cdot w_{i, \mu_{j,2,1}, \mu_{j,2,2}} \right\} + \\ + \sum_{j=1}^{C_{l \cdot elem}^2} \alpha_j^{max} \max \left\{ w_{i, \mu_{j,1,1}, \mu_{j,1,2}}, k_j^{max} \cdot w_{i, \mu_{j,2,1}, \mu_{j,2,2}} \right\} + \varepsilon_i, \quad i = 1, 2, \dots, n,$$

где $w_{ijk} = f_k(x_{ij})$, $i = 1, 2, \dots, n$, $j = 1, 2, \dots, l$, $k = 1, 2, \dots, elem$.

Из [1] следует, что области возможных значений параметров k_j^{min} , k_j^{max} , $j = 1, 2, \dots, C_{l \cdot elem}^2$, в регрессии (3) при $w_{ijk} \neq 0$ можно записать в виде следующих промежутков:

$$(4) \quad k_j^{нижн} < k_j^{min} < k_j^{верхн}, \quad k_j^{нижн} < k_j^{max} < k_j^{верхн}, \quad j = 1, 2, \dots, C_{l \cdot elem}^2,$$

$$\text{где } k_j^{нижн} = \min \left\{ \frac{w_{1, \mu_{j,1,1}, \mu_{j,1,2}}}{w_{1, \mu_{j,2,1}, \mu_{j,2,2}}}, \frac{w_{2, \mu_{j,1,1}, \mu_{j,1,2}}}{w_{2, \mu_{j,2,1}, \mu_{j,2,2}}}, \dots, \frac{w_{n, \mu_{j,1,1}, \mu_{j,1,2}}}{w_{n, \mu_{j,2,1}, \mu_{j,2,2}}} \right\},$$

$$k_j^{верхн} = \max \left\{ \frac{w_{1, \mu_{j,1,1}, \mu_{j,1,2}}}{w_{1, \mu_{j,2,1}, \mu_{j,2,2}}}, \frac{w_{2, \mu_{j,1,1}, \mu_{j,1,2}}}{w_{2, \mu_{j,2,1}, \mu_{j,2,2}}}, \dots, \frac{w_{n, \mu_{j,1,1}, \mu_{j,1,2}}}{w_{n, \mu_{j,2,1}, \mu_{j,2,2}}} \right\}.$$

Затем равномерно разобьем каждый из промежутков (4) p точками и перепишем регрессию (3) в виде

$$(5) \quad y_i = \alpha_0 + \sum_{j=1}^l \sum_{k=1}^{elem} \alpha_{jk} \cdot w_{ijk} + \\ + \sum_{j=1}^{C_{l \cdot elem}^2} \sum_{k=1}^p \alpha_{jk}^- \min \left\{ w_{i, \mu_{j,1,1}, \mu_{j,1,2}}, \lambda_{jk} \cdot w_{i, \mu_{j,2,1}, \mu_{j,2,2}} \right\} + \\ + \sum_{j=1}^{C_{l \cdot elem}^2} \sum_{k=1}^p \alpha_{jk}^+ \max \left\{ w_{i, \mu_{j,1,1}, \mu_{j,1,2}}, \lambda_{jk} \cdot w_{i, \mu_{j,2,1}, \mu_{j,2,2}} \right\} + \varepsilon_i, \quad i = 1, 2, \dots, n,$$

где λ_{jk} , $j = 1, 2, \dots, C_{l \cdot elem}^2$, $k = 1, 2, \dots, p$ – элементы матрицы Λ размера $C_{l \cdot elem}^2 \times p$. Элемент λ_{jk} равен значению k -й точки j -го промежутка (4). Поскольку коэффициенты λ_{jk} , $j = 1, 2, \dots, C_{l \cdot elem}^2$, $k = 1, 2, \dots, p$, известны, то регрессия (5) относится к квазилинейным, т.е. становится линейной по неизвестным параметрам α_{jk} , $j = 1, 2, \dots, l$, $k = 1, 2, \dots, elem$; α_{jk}^- , α_{jk}^+ $j = 1, 2, \dots, C_{l \cdot elem}^2$, $k = 1, 2, \dots, p$. Поэтому удобнее записать её в виде

$$(6) \quad y_i = \alpha_0 + \sum_{j=1}^l \sum_{k=1}^{elem} \alpha_{jk} \cdot w_{ijk} + \sum_{j=1}^{C_{l \cdot elem}^2} \sum_{k=1}^p \alpha_{jk}^- \cdot z_{ijk}^- + \\ + \sum_{j=1}^{C_{l \cdot elem}^2} \sum_{k=1}^p \alpha_{jk}^+ \cdot z_{ijk}^+ + \varepsilon_i, \quad i = 1, 2, \dots, n,$$

где $z_{ijk}^- = \min \left\{ w_{i, \mu_{j,1,1}, \mu_{j,1,2}}, \lambda_{jk} \cdot w_{i, \mu_{j,2,1}, \mu_{j,2,2}} \right\}$,

$z_{ijk}^+ = \max \left\{ w_{i, \mu_{j,1,1}, \mu_{j,1,2}}, \lambda_{jk} \cdot w_{i, \mu_{j,2,1}, \mu_{j,2,2}} \right\}$, $i = 1, 2, \dots, n$, $j = 1, 2, \dots,$

$C_{l \cdot elem}^2$, $k = 1, 2, \dots, p$.

Сформулируем задачу ОИР для регрессии (6) следующим образом. Требуется выбрать оптимальное число регрессоров так, чтобы коэффициент детерминации R^2 модели был максимален, каждая объясняющая переменная входила в модель не более одного раза, а знаки всех МНК-оценок параметров α_{jk} , α_{jk}^- , α_{jk}^+ удовлетворяли содержательному смыслу факторов. Последнее условие сначала проверяется априори, для чего нужно обратиться к экспертам из данной предметной области, задача которых проанализировать коэффициенты корреляции зависимой переменной со всеми объясняющими переменными w_{jk} , z_{jk}^- , z_{jk}^+ .

Не удовлетворяющие смыслу переменные следует исключить, либо, например, дополнить выборку новыми наблюдениями. Затем согласованность содержательному смыслу проверяется уже после оценивания с использованием следующих неравенств:

$$(7) \alpha_{jk} \cdot r_{y, w_{jk}} > 0, j = 1, 2, \dots, l, k = 1, 2, \dots, elem;$$

$$\alpha_{jk}^- \cdot r_{y, z_{jk}^-} > 0, \alpha_{jk}^+ \cdot r_{y, z_{jk}^+} > 0, j = 1, 2, \dots, C_{l \cdot elem}^2, k = 1, 2, \dots, p,$$

где символом r обозначены коэффициенты корреляции между переменными.

Введем принимающие два значения «0» и «1» бинарные переменные

$$(8) \delta_{jk} \in \{0,1\}, j = 1, 2, \dots, l, k = 1, 2, \dots, elem,$$

$$(9) \delta_{jk}^-, \delta_{jk}^+ \in \{0,1\}, j = 1, 2, \dots, C_{l \cdot elem}^2, k = 1, 2, \dots, p,$$

которые отвечают за вхождение регрессоров w_{jk} , z_{jk}^- , z_{jk}^+ в модель. Например, если $\delta_{jk} = 1$, то в модель входит регрессор без бинарной операции (j -я переменная с k -м преобразованием), а если $\delta_{jk} = 0$, то нет. Если $\delta_{jk}^- = 1$, то в модель входит регрессор с бинарной операцией \min (j -я пара преобразованных переменных для k -й точки разбитого промежутка), а если $\delta_{jk}^- = 0$, то нет. Аналогично можно описать переменную δ_{jk}^+ .

С помощью переменных, преобразованных по правилам

$$y_i^* = \frac{y_i - \bar{y}}{\sigma_y}, i = 1, 2, \dots, n; w_{ijk}^* = \frac{w_{ijk} - \bar{w}_{jk}}{\sigma_{w_{jk}}},$$

$$i = 1, 2, \dots, n, j = 1, 2, \dots, l, k = 1, 2, \dots, elem;$$

$$z_{ijk}^* = \frac{z_{ijk}^- - \bar{z}_{jk}^-}{\sigma_{z_{jk}^-}}, z_{ijk}^{**} = \frac{z_{ijk}^+ - \bar{z}_{jk}^+}{\sigma_{z_{jk}^+}}, i = 1, 2, \dots, n,$$

$$j = 1, 2, \dots, C_{l \cdot elem}^2, k = 1, 2, \dots, p,$$

перейдем от модели (6) к стандартизованной регрессии с неизвестными коэффициентами β_{jk} , $j = 1, 2, \dots, l$, $k = 1, 2, \dots, elem$, β_{jk}^- , β_{jk}^+ , $j = 1, 2, \dots, C_{l \cdot elem}^2$, $k = 1, 2, \dots, p$, вида

$$(10) \quad y_i^* = \sum_{j=1}^l \sum_{k=1}^{elem} \beta_{jk}^* \cdot w_{ijk}^* + \sum_{j=1}^{C_{elem}^2} \sum_{k=1}^p \beta_{jk}^- \cdot z_{ijk}^* + \\ + \sum_{j=1}^{C_{elem}^2} \sum_{k=1}^p \beta_{jk}^+ \cdot z_{ijk}^{**} + \varepsilon_i^*, \quad i = 1, 2, \dots, n,$$

где $\varepsilon_1^*, \dots, \varepsilon_n^*$ – ошибки аппроксимации модели.

Коэффициент детерминации R^2 регрессий (6) и (10) находится [1] как сумма произведений стандартизованных коэффициентов на соответствующие корреляции регрессоров с переменной y , поэтому введем следующую целевую функцию:

$$(11) \quad R^2 = \sum_{j=1}^l \sum_{k=1}^{elem} \beta_{jk} \cdot r_{y, w_{jk}} + \sum_{j=1}^{C_{elem}^2} \sum_{k=1}^p \beta_{jk}^- \cdot r_{y, z_{jk}} + \\ + \sum_{j=1}^{C_{elem}^2} \sum_{k=1}^p \beta_{jk}^+ \cdot r_{y, z_{jk}^{**}} \rightarrow \max.$$

МНК-оценивание модели (10) состоит в решении системы линейных алгебраических уравнений. Однако при комбинировании регрессоров нужно учесть изменение структуры такой системы: обнуление оценок и исключение из системы лишних уравнений. Обнуление оценок, с учётом введенных бинарных переменных и условий (7), осуществляется с помощью следующих ограничений:

$$(12) \quad 0 \leq \beta_{jk} \leq M \cdot \delta_{jk}, \quad (j, k) \in \left\{ (s_1, s_2) \mid r_{y, w_{s_1 s_2}} > 0 \right\},$$

$$(13) \quad -M \cdot \delta_{jk} \leq \beta_{jk} \leq 0, \quad (j, k) \in \left\{ (s_1, s_2) \mid r_{y, w_{s_1 s_2}} < 0 \right\},$$

$$(14) \quad 0 \leq \beta_{jk}^- \leq M \cdot \delta_{jk}^-, \quad (j, k) \in \left\{ (s_1, s_2) \mid r_{y, z_{s_1 s_2}^-} > 0 \right\},$$

$$(15) \quad -M \cdot \delta_{jk}^- \leq \beta_{jk}^- \leq 0, \quad (j, k) \in \left\{ (s_1, s_2) \mid r_{y, z_{s_1 s_2}^-} < 0 \right\},$$

$$(16) \quad 0 \leq \beta_{jk}^+ \leq M \cdot \delta_{jk}^+, \quad (j, k) \in \left\{ (s_1, s_2) \mid r_{y, z_{s_1 s_2}^+} > 0 \right\},$$

$$(17) \quad -M \cdot \delta_{jk}^+ \leq \beta_{jk}^+ \leq 0, \quad (j, k) \in \left\{ (s_1, s_2) \mid r_{y, z_{s_1 s_2}^+} < 0 \right\},$$

где M – большое положительное число, способ выбора которого обсуждается в [1]. Заметим, что использовать ограничения

(12)–(17) на МНК-оценки не обязательно. Однако, как утверждается в [12], их наличие существенно повышает скорость решения задачи.

Исключение из системы лишних уравнений для обнуленных бинарных переменных организуется с помощью ограничений

$$(18) \quad -\left(1 - \delta_{jk}^-\right) M \leq \sum_{s_1=1}^l \sum_{s_2=1}^{elem} r_{w_{jk} w_{s_1 s_2}} \cdot \beta_{jk} + \sum_{s_1=1}^{C_{l,elem}^2} \sum_{s_2=1}^p r_{w_{jk} z_{s_1 s_2}^-} \cdot \beta_{s_1 s_2}^- + \\ + \sum_{s_1=1}^{C_{l,elem}^2} \sum_{s_2=1}^p r_{w_{jk} z_{s_1 s_2}^+} \cdot \beta_{s_1 s_2}^+ - r_{y, w_{jk}} \leq (1 - \delta_{jk}^-) M, \\ j = 1, 2, \dots, l, k = 1, 2, \dots, elem,$$

$$(19) \quad -\left(1 - \delta_{jk}^-\right) M \leq \sum_{s_1=1}^l \sum_{s_2=1}^{elem} r_{w_{s_1 s_2} z_{jk}^-} \cdot \beta_{s_1 s_2} + \sum_{s_1=1}^{C_{l,elem}^2} \sum_{s_2=1}^p r_{z_{s_1 s_2} z_{jk}^-} \cdot \beta_{s_1 s_2}^- + \\ + \sum_{s_1=1}^{C_{l,elem}^2} \sum_{s_2=1}^p r_{z_{s_1 s_2} z_{jk}^+} \cdot \beta_{s_1 s_2}^+ - r_{y z_{jk}^-} \leq (1 - \delta_{jk}^-) M, \\ j = 1, 2, \dots, C_{l,elem}^2, k = 1, 2, \dots, p,$$

$$(20) \quad -\left(1 - \delta_{jk}^+\right) M \leq \sum_{s_1=1}^l \sum_{s_2=1}^{elem} r_{w_{s_1 s_2} z_{jk}^+} \cdot \beta_{s_1 s_2} + \sum_{s_1=1}^{C_{l,elem}^2} \sum_{s_2=1}^p r_{z_{s_1 s_2} z_{jk}^+} \cdot \beta_{s_1 s_2}^- + \\ + \sum_{s_1=1}^{C_{l,elem}^2} \sum_{s_2=1}^p r_{z_{s_1 s_2} z_{jk}^-} \cdot \beta_{s_1 s_2}^+ - r_{y z_{jk}^+} \leq (1 - \delta_{jk}^+) M, \\ j = 1, 2, \dots, C_{l,elem}^2, k = 1, 2, \dots, p.$$

Как видно, если некоторая бинарная переменная «срабатывает», т.е. принимает значение 1, то соответствующее ограничение из множества (18)–(20) трансформируется в строгое равенство. Иначе это ограничение игнорируется.

Ограничения на единственность вхождения каждой объясняющей переменной в модель можно записать в виде

$$(21) \quad \sum_{k=1}^{elem} \delta_{jk} + \sum_{i \in Y_j} \sum_{k=1}^p \delta_{ik}^- + \sum_{i \in Y_j} \sum_{k=1}^p \delta_{ik}^+ \leq 1, \quad j = 1, 2, \dots, l,$$

где Y_j – множество номеров горизонтальных срезов трехмерной матрицы \mathbf{M}^* , содержащих среди своих элементов число j .

Таким образом, решение задачи ЧБЛП с целевой функцией (11) и с линейными ограничениями (8), (9), (12)–(21) гарантирует построение НКЛинР (2) с требуемыми свойствами.

3. Пример

Для построения НКЛинР была использована выборочная совокупность объема $n = 21$, по которой в [1] идентифицировалась НЛинР. Из исходного набора были задействованы следующие переменные: y – отправление грузов железнодорожным транспортом общего пользования в Иркутской области (млн тонн); x_2 – процент трудоспособного населения от общей численности; x_3 – численность рабочей силы (тыс. чел.); x_{18} – число предприятий и организаций; x_{22} – производство электроэнергии (млрд кВт*ч).

Выбор именно этих четырех объясняющих переменных продиктован следующими соображениями.

1. На объемы отправленных грузов железнодорожным транспортом, несомненно, влияют объемы производства в регионе. А повышение объемов производства стимулирует, в частности, увеличение мощностей генерации электроэнергии, числа предприятий и численности рабочей силы.

2. Все коэффициенты корреляций объясняющих переменных с y значимы для уровня значимости 0,05.

3. Коэффициенты корреляции всех этих четырех объясняющих переменных с y положительны, что соответствует экономическому смыслу решаемой задачи.

Заметим, что единственный значимый коэффициент корреляции равен 0,82 между переменными x_2 и x_3 . Однако, как будет показано далее, это не повлияет на интерпретацию НКЛинР.

Рассмотрим этапы построения НКЛинР в программе Винтер-2.

Поскольку все четыре объясняющих переменных по знаку корреляции с y согласованы, то можно сразу переходить к загрузке выборки в программу. Загрузив статистические данные из текстового файла, на панели «Квазилинейная» сначала нужно выбрать элементарные преобразования объясняющих перемен-

ных. На данный момент в ВИнтер-2 встроено 9 функций: x^2 , $x^{-1.5}$, x^{-1} , $x^{-0.5}$, \sqrt{x} , x , $x^{1.5}$, x^2 , $\ln(x)$. Выбор слишком большого числа преобразований негативно влияет на скорость решения задачи, поэтому было принято решения сформировать множество F из трех функций: x^{-1} , \sqrt{x} и $\ln(x)$. Выбрав преобразования, следует нажать на кнопку «Исключить». В результате нажатия формируются все преобразованные объясняющие переменные и каждая из них автоматически проходит проверку на: 1) согласованность знака корреляции с y смыслу задачи; 2) слишком малое значение коэффициента корреляции с y ; 3) слишком высокие значения критерия нелинейности [5]. Для последних двух условий пороговые значения в ВИнтер-2 были назначены 0 и 1 соответственно. Для первого условия ничего задавать не нужно. В итоге ни одна из 12 преобразованных переменных не была исключена.

Затем нужно выбрать неэлементарные преобразования для всех двенадцати элементарно преобразованных переменных. Для этого на панели «Неэлементарные преобразования» нужно указать два параметра: 1) число разбиений; 2) корреляция с y . Первый параметр означает число точек разбиения p промежутков (4). Второй параметр – пороговое значение для абсолютной величины коэффициента корреляции преобразованной переменной с y . Было принято решение задать эти параметры равными 4 и 0,7 соответственно. После чего, нажав кнопку «Исключить», автоматически сформируются все возможные неэлементарные преобразования переменных и каждая из них пройдет проверку на: 1) согласованность знака корреляции с y смыслу задачи; 2) слишком малое значение коэффициента корреляции с y . Всего было сформировано $2C_{12}^{elem} \cdot p = 2C_{12}^4 \cdot 4 = 528$ неэлементарных преобразований. В результате исключения из них осталось только 80. Тем самым, к построению НКЛинР мы подошли, имея в распоряжении 12 элементарно и 80 не элементарно преобразованных переменных.

Для формирования в ВИнтер-2 задачи ЧБЛП (8), (9), (11)–(21) на панели «Задача ЧБЛП» были выбраны следующие параметры: 1) регрессоры – 0; 2) точность – «0,000000000000»;

3) вхождение – 1. Первый параметр означает, что нет ограничений на число регрессоров в модели, третий – что каждая объясняющая переменная входит в модель не более одного раза. Второй параметр означает, что все величины в задаче должны округляться до 12 знаков после запятой. При нажатии на кнопку «Создать» автоматически сформировалась задача ЧБЛП для пакета LPSolve.

Решение сформированной задачи в LPSolve на обычном персональном компьютере с процессором AMD Ryzen 3 4300 (2,70 ГГц) было получено за 2,724 с. С использованием информации, расположенной в текстовом поле на панели «Неэлементарные преобразования», была произведена расшифровка спецификации модели. Таким образом, построена следующая НКЛинР:

$$(22) \tilde{y} = -618,246 + \overset{(0,5363)}{0,43} \min\{\sqrt{x_{18}}; 32,4774\sqrt{x_{22}}\} + \\ + \overset{(0,4244)}{\underset{(12,11)}{72,594}} \max\{\sqrt{x_2}; 1,1024 \ln x_3\}.$$

Коэффициент детерминации R^2 модели (22) составляет 0,960711, что на 0,014528 выше, чем у построенной в [1] НЛинР. К тому же число степеней свободы НКЛинР (22) $df = n - m - 1 = 21 - 2 - 1 = 18$, что на 2 единицы больше, чем у приведенной в [1] модели. Как видно, каждая объясняющая переменная входит в НКЛинР ровно 1 раз, а знаки всех МНК-оценок в этой регрессии согласуются с содержательным смыслом задачи.

В уравнении (22) в скобках под коэффициентами приведены наблюдаемые значения t -критерия Стьюдента, а над оценками – значения абсолютных вкладов переменных в общую детерминацию, которые в сумме дают 0,9607. Оба регрессора в регрессии (22) значимы по t -критерию Стьюдента для уровня значимости $\alpha = 0,01$. Причем значение коэффициента корреляции между переменными $\min\{\sqrt{x_{18}}; 32,4774\sqrt{x_{22}}\}$ и $\max\{\sqrt{x_2}; 1,1024 \ln x_3\}$ составляет 0,225, что говорит об отсутствии в модели мультиколлинеарности.

Представим модель (22) в виде кусочно-заданной функции:

$$(23) \quad \tilde{y} = \begin{cases} -618,246 + 13,97\sqrt{x_{22}} + 72,59\sqrt{x_2} & \text{при } \frac{x_{18}}{x_{22}} \geq 1054,78, \quad \frac{\sqrt{x_2}}{\ln x_3} \geq 1,102, \\ -618,246 + 13,97\sqrt{x_{22}} + 80,03\ln x_3 & \text{при } \frac{x_{18}}{x_{22}} \geq 1054,78, \quad \frac{\sqrt{x_2}}{\ln x_3} < 1,102, \\ -618,246 + 0,43\sqrt{x_{18}} + 72,59\sqrt{x_2} & \text{при } \frac{x_{18}}{x_{22}} < 1054,78, \quad \frac{\sqrt{x_2}}{\ln x_3} \geq 1,102, \\ -618,246 + 0,43\sqrt{x_{18}} + 80,03\ln x_3 & \text{при } \frac{x_{18}}{x_{22}} < 1054,78, \quad \frac{\sqrt{x_2}}{\ln x_3} < 1,102. \end{cases}$$

Видно, что функция (23) меняет своё аналитическое выражение в зависимости от значений соотношений x_{18} / x_{22} и $\sqrt{x_2} / \ln x_3$. Если $x_{18} / x_{22} \geq 1054,78$, то на y влияет регрессор $\sqrt{x_{22}}$, а если $x_{18} / x_{22} < 1054,78$, то $\sqrt{x_{18}}$. Если $\sqrt{x_2} / \ln x_3 \geq 1,102$, то на y влияет регрессор $\sqrt{x_2}$, а если $\sqrt{x_2} / \ln x_3 < 1,102$, то $\ln(x_3)$. Условие $x_{18} / x_{22} \geq 1054,78$ сработало для 2003–2009, 2013–2018 гг., а $x_{18} / x_{22} < 1054,78$ – в остальных случаях. Условие $\sqrt{x_2} / \ln x_3 \geq 1,102$ сработало для 2003–2008 гг., а условие $\sqrt{x_2} / \ln x_3 < 1,102$ – в остальных случаях. Для интерпретации коэффициентов при преобразованных переменных $\sqrt{x_2}$, $\sqrt{x_{18}}$, $\sqrt{x_{22}}$ и $\ln(x_3)$ можно использовать приём, предложенный в [5]. Из-за переключения преобразованных переменных x_2 и x_3 высокая степень их корреляции не влияет на интерпретацию оценок.

Далее было принято решение с использованием других известных спецификаций попытаться получить модель с двумя регрессорами, т.е. с числом степеней свободы $df = 18$, превосходящую НКЛинР (22) по значению R^2 . В результате были оценены следующие регрессии.

1. Линейная регрессия со всеми четырьмя переменными:

$$\tilde{y} = -76,682 + 1,849x_2 - 0,0162x_3 + 0,0005x_{18} + 0,22x_{22},$$

для которой $R^2 = 0,867334$. Во-первых, в этой регрессии, в отличие от модели (22), знак коэффициента при переменной x_3 не удовлетворяет содержательному смыслу задачи. Во-вторых, если даже линейная регрессия с четырьмя переменными оказалась

хуже модели (22), то двухфакторные линейные регрессии, для которых $df = 18$, тем более окажутся хуже.

2. Квазилинейная модель:

$$\tilde{y} = -126,157 - 1,77 \cdot 10^6 x_{18}^{-1} + 27,434 \sqrt{x_2},$$

для которой $R^2 = 0,852327$. Эта регрессия получена с использованием той же технологии, что и (22), поэтому знаки коэффициентов в ней корректны, но качество её ниже.

3. Параболический тренд:

$$\tilde{y} = 49,2728 + 2,9061t - 0,1551t^2,$$

для которого $R^2 = 0,698581$, $t = 1, 2, \dots, 21$, – переменная времени.

4. Тренд с фиктивной переменной:

$$\tilde{y} = 44,756 + 3,379t - 5,0275d,$$

для которого $R^2 = 0,911306$, а фиктивная переменная задана по

правилу $d = \begin{cases} 0, & \text{если } t < 7, \\ t - 7, & \text{если } t \geq 7. \end{cases}$

5. Авторегрессия второго порядка:

$$\tilde{y}_t = 5,944 + 0,936y_{t-1} - 0,0443y_{t-2},$$

для которой $R^2 = 0,732612$.

Таким образом, все пять типов построенных моделей проиграли по величине R^2 НКЛинР (22).

4. Заключение

В статье впервые предложен метод построения неэлементарных квазилинейных регрессий, которые могут быть использованы для выработки эффективных управленческих решений в различных системах. Сформулированная задача ЧБЛП содержит линейные ограничения на знаки МНК-оценок и количество вхождений объясняющих переменных в модель. При снятии этих ограничений результаты моделирования могут существенно улучшиться, но при этом увеличится и время решения задачи. Предложенный метод реализован в программе ВИнтер-2, которая с помощью решателя LPsolve по заданным пользователем начальным параметрам автоматически строит оптимальную

по коэффициенту детерминации неэлементарную квазилинейную регрессию, т.е. автоматически определяет состав входящих в неё объясняющих переменных, а также их элементарные и неэлементарные преобразования. В программе имеется возможность управлять количеством преобразованных переменных и тем самым контролировать сложность задачи и время её решения. Построенная с помощью ВИнтер-2 модель по данным о железнодорожных перевозках оказалась лучше, чем построенная ранее неэлементарная линейная регрессия. Однако в построенных нашим методом регрессиях некоторые регрессоры могут быть незначимы по тем или иным критериям, а также может присутствовать мультиколлинеарность. Решению этих проблем будут посвящены дальнейшие исследования.

Литература

1. БАЗИЛЕВСКИЙ М.П. *Метод построения неэлементарных линейных регрессий на основе аппарата математического программирования* // Проблемы управления. – 2022. – №4. – С. 3–14.
2. БАЗИЛЕВСКИЙ М.П. *Отбор информативных операций при построении линейно-неэлементарных регрессионных моделей* // Int. Journal of Open Information Technologies. – 2021. – Т. 9, №5. – С. 30–35.
3. БАЗИЛЕВСКИЙ М.П. *Оценивание линейно-неэлементарных регрессионных моделей с помощью метода наименьших квадратов* // Моделирование, оптимизация и информационные технологии. – 2020. – Т. 8, №4(31).
4. БАЗИЛЕВСКИЙ М.П. *Программа построения вполне интерпретируемых элементарных и неэлементарных квазилинейных регрессионных моделей* // Труды Института системного программирования РАН. – 2023. – Т. 35, №4. – С. 129–144.
5. БАЗИЛЕВСКИЙ М.П. *Технология построения вполне интерпретируемых квазилинейных регрессионных моделей* // Прикладная математика и вопросы управления. – 2024. – №1. – С. 123–138.

6. КЛЕЙНЕР Г.Б. *Производственные функции: Теория, методы, применение.* – М.: Финансы и статистика, 1986. – 239 с.
7. НОСКОВ С.И. *Технология моделирования объектов с нестабильным функционированием и неопределенностью в данных.* – Иркутск: РИЦ ГП «Облформпечать», 1996. – 320 с.
8. AHMAD S.A., RAFIQ S.K., AHMED H.U. et al. *Innovative soft computing techniques including artificial neural network and nonlinear regression models to predict the compressive strength of environmentally friendly concrete incorporating waste glass powder // Innovative Infrastructure Solutions.* – 2023. – Vol. 8, No. 4. – P. 119.
9. CHEN D., HU F., NIAN G. et al. *Deep residual learning for nonlinear regression // Entropy.* – 2020. – Vol. 22, No. 2. – P. 193.
10. FENG Y., LIU Q., YAO Q. et al. *Model averaging for nonlinear regression models // Journal of Business & Economic Statistics.* – 2022. – Vol. 40, No. 2. – P. 785–798.
11. JAF D.K.I., ABDALLA A., MOHAMMED A.S. et al. *Hybrid nonlinear regression model versus MARS, MEP, and ANN to evaluate the effect of the size and content of waste tire rubber on the compressive strength of concrete // Heliyon.* – 2024. – Vol. 10, No. 4.
12. KONNO H., YAMAMOTO R. *Choosing the best set of variables in regression analysis using integer programming // Journal of Global Optimization.* – 2009. – Vol. 44. – P. 273–282.
13. MAZUMDER R., RADCHENKO P., DEDIEU A. *Subset selection with shrinkage: Sparse linear modeling when the SNR is low // Operations Research.* – 2023. – Vol. 71, No. 1. – P. 129–147.
14. MILLER A. *Subset selection in regression.* – Chapman and hall/CRC, 2002.
15. MOLNAR C. *Interpretable machine learning.* – Lulu.com, 2020.
16. MONTGOMERY D.C., PECK E.A., VINING G.G. *Introduction to linear regression analysis.* – John Wiley & Sons, 2021.
17. OMOGOROYE O.O., OLANIYI O.O., ADEBIYI O.O. et al. *Electricity consumption (kW) forecast for a building of interest*

- based on a time series nonlinear regression model // Asian Journal of Economics, Business and Accounting. – 2023. – Vol. 23, No. 21. – P. 197–207.*
18. SHARIFANI K., AMINI M. *Machine learning and deep learning: A review of methods and applications // World Information Technology and Engineering Journal. – 2023. – Vol. 10, No. 7. – P. 3897–3904.*
 19. TUFANER F., DEMIRCI Y. *Prediction of biogas production rate from anaerobic hybrid reactor by artificial neural network and nonlinear regressions models // Clean Technologies and Environmental Policy. – 2020. – Vol. 22. – P. 713–724.*
 20. WEN L., LI Y., CHAI J. *Multiple nonlinear regression models for predicting deformation behavior of concrete-face rockfill dams // Int. Journal of Geomechanics. – 2021. – Vol. 21, No. 2. – P. 04020253.*
 21. WOLSEY L.A. *Integer programming. – John Wiley & Sons, 2020.*
 22. YANG G.R., WANG X.J. *Artificial neural networks for neuroscientists: a primer // Neuron. – 2020. – Vol. 107, No. 7. – P. 1048–1070.*
 23. YANG Y., DU L., LI Q. et al. *Vibration prediction and analysis of the main beam of the TBM based on a multiple linear regression model // Scientific Reports. – 2024. – Vol. 14, No. 1. – P. 3498.*
 24. ZHANG Z., YIN Z., CHEN Y. et al. *Evaluation and prediction of water resources carrying capacity using a multiple linear regression model in Taizhou City, China // Human and Ecological Risk Assessment: An International Journal. – 2023. – Vol. 29, No. 2.– P. 553–570.*
 25. ZITAR R.A., MOHSEN A., SEGHTROUCHNI A.E. et al. *Intensive review of drones detection and tracking: linear kalman filter versus nonlinear regression, an analysis case // Archives of Computational Methods in Engineering. – 2023. – Vol. 30, No. 5. – P. 2811–2830.*

CONSTRUCTING NON-ELEMENTARY QUASILINEAR REGRESSIONS USING MATHEMATICAL PROGRAMMING APPARATUS

Mikhail Bazilevskiy, Irkutsk State Transport University, Irkutsk, Candidate of Technical Sciences, Assistant Professor (mik2178@yandex.ru).

Abstract: In non-elementary quasilinear regressions, the explanatory variables are first transformed using elementary functions, after which the pairs of resulting factors are again transformed using the non-elementary functions min and max. Such models are nonlinear in both factors and parameters, so even their estimation seems to be a complex computational task. And if the composition of the variables included in the model, as well as their elementary and non-elementary transformations, is unknown, then the complexity of the problem increases significantly. This study aims to solve this problem. Instead of labor-intensive exhaustive search procedures, a well-developed mathematical programming apparatus has been used recently. The method for constructing non-elementary quasilinear regressions is formalized as a mixed 0-1 integer linear programming problem. The proposed method is implemented in a special computer program. Its advantage is that the user can regulate the number of transformed variables during the construction process, so the program can be used both for solving simple control problems on ordinary personal computers and for processing large data arrays using cloud services. Non-elementary quasilinear regressions can be used to solve control problems in technical, socio-economic, medical and other systems.

Keywords: regression analysis, nonlinear regression, non-elementary quasilinear regression, ordinary least squares method, mixed 0-1 linear programming problem.

УДК 519.862.6

ББК 22.18

*Статья представлена к публикации
членом редакционной коллегии Р.М. Нижегородцевым.*

Поступила в редакцию 24.07.2024.

Опубликована 30.11.2024.

МЕТОДОЛОГИЯ СЦЕНАРНОГО ПЛАНИРОВАНИЯ И УПРАВЛЕНИЯ В СИСТЕМЕ ОБЕСПЕЧЕНИЯ НАЦИОНАЛЬНОЙ БЕЗОПАСНОСТИ

Чернов И. В.¹

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

Разрабатывается методология сценарного планирования и управления обеспечением национальной безопасности Российской Федерации. Проведен анализ основных направлений развития методологии сценарного анализа и моделирования как инструмента поддержки принятия решений. Приведены результаты разработки понятийного аппарата сценарного исследования, на основе которого разработана общая схема процедуры формирования сценариев. В основу экспертного описания модели развития рассматриваемых процессов положено понятие экспертно-значимого разбиения расширенного фазового пространства, а также экспертно-значимых событий, которые происходят в последовательности, указанной в системе упорядочения в рамках базисной модели. В результате исследования данной модели формируются сценарии поведения объекта управления или развития ситуации в сфере безопасности. Разработана многоуровневая модель жизненного цикла планирования и управления мероприятиями безопасности, обеспечивающая возможность упреждающей верификации результатов реализации ключевых этапов управленческого цикла, что позволяет перейти от пассивного учета свершившихся событий к проактивному управлению. Полученные результаты могут послужить основой для автоматизации сценарного исследования, а также аналитической прогнозной поддержки процессов подготовки и принятия управленческих решений в области обеспечения безопасности сложных организационных систем.

Ключевые слова: безопасность, сценарий, жизненный цикл, управление, моделирование.

1. Введение

Сложившаяся в настоящее время ситуация привела к нарастанию существующих и появлению принципиально новых и крайне опасных стратегических угроз национальной безопасности (НБ) России. Многофакторный и многоцелевой характер внешних угроз (а по ряду направлений скрытный, а также ориен-

¹ Игорь Викторович Чернов, к.т.н., в.н.с. (chernov@ipi.ru).

тированный на имеющиеся уязвимости социально-экономической системы нашей страны) характер выбранных геополитическими противниками способов и механизмов военного, политического и экономического давления на Российскую Федерацию диктуют острую необходимость повышения эффективности методов и механизмов планирования и управления в сфере обеспечения НБ России.

Обеспечение национальной безопасности государства является одной из наиболее сложных проблем теории и методологии организационного управления и целого ряда смежных научных дисциплин.

Стратегия национальной безопасности Российской Федерации определяет систему обеспечения национальной безопасности как «совокупность осуществляющих реализацию государственной политики в сфере НБ органов публичной власти и находящихся в их распоряжении инструментов» [13]. Таким образом, национальная безопасность и как предметная область, и как объект управления представляет собой сложную распределенную иерархическую мультиструктурную систему, включающую множество различных взаимосвязанных элементов на федеральном, региональном, отраслевом и т.д. уровнях. При этом можно выделить внешний и внутренний контуры национальной безопасности (рис. 1). Первый уровень отражает внешнюю (военно-политическую, внешнеэкономическую, дипломатическую и т.д.) составляющую комплекса проблем обеспечения национальной безопасности и тесно связан с уровнем международной безопасности (состоянием и тенденциями развития международных отношений). Внутренний контур соответственно отражает проблемы противодействия внутренним угрозам устойчивому и поступательному развитию общества и государства (социально-экономическому, производственно-технологическому, научно-техническому и т.д.) [6, 20].

Необходимым условием повышения эффективности процессов управления обеспечением НБ является смещение акцента с задач реактивного характера на опережающие, основной целью которых является недопущение кризисных ситуаций на основе упреждающего выявления негативных тенденций и принятия содействующих воздействий.

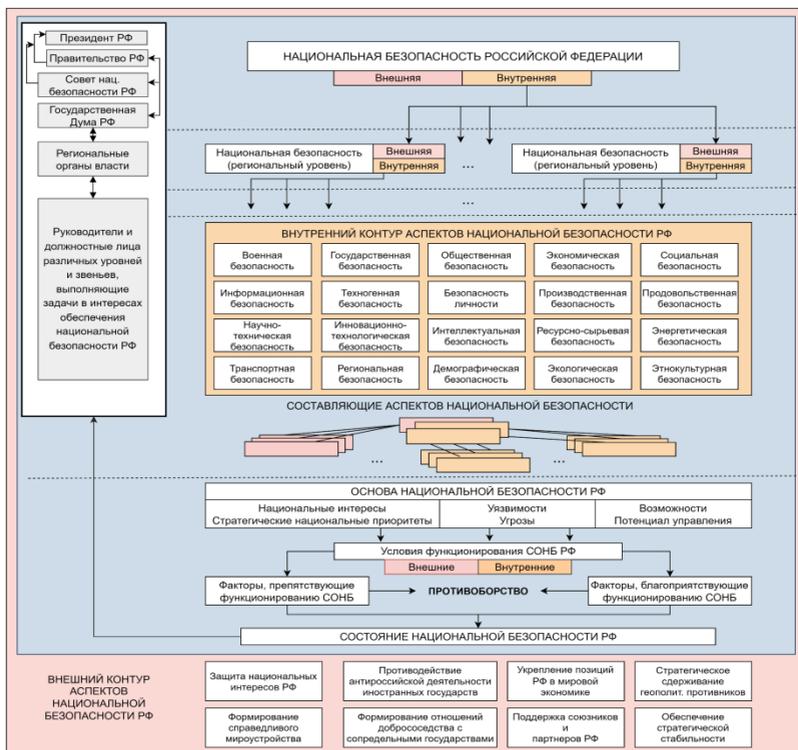


Рис. 1. Система обеспечения национальной безопасности (СОНБ)

Методология формирования альтернативных сценариев развития ситуации и поведения сложных объектов организационного управления направлена на выделение и последующий опережающий анализ (упреждающий мониторинг) совокупности ключевых факторов, характеризующих уязвимости и риски (как внешние, так и внутренние), идентификацию текущих и потенциальных угроз реализации целей управления обеспечением как НБ в целом, так и ее важнейших функционально-структурных компонент, а также опережающую оценку качества и последствий реализации управленческих решений в рассматриваемой предметной области.

Сценарная технология принципиально позволяет обеспечить возможность решения сложных проблем в условиях неполной наблюдаемости происходящих внешних и внутренних процессов; неполноты и недостаточной достоверности исходной количественной и качественной информации; отсутствия точных значений большинства описывающих складывающуюся обстановку факторов; труднопредсказуемости тенденций возможного развития проблемных ситуаций; скрытного характера многих существующих и потенциальных угроз; отсутствия практической возможности заблаговременно предвидеть и оценить как позитивные, так и возможные нежелательные результаты реализации разрабатываемых планово-управленческих решений [20]. Кроме того, методология формирования и анализа сценариев развития проблемных ситуаций является в достаточной степени универсальным инструментом поддержки принятия решений, что обеспечивает возможность его использования при решении достаточно широкого класса планово-управленческих задач в рассматриваемой предметной области.

В настоящее время сценарные технологии достаточно широко используются при решении самых разнообразных задач организационного управления в различных областях человеческой деятельности, причем сфера их успешного применения непрерывно расширяется, о чем свидетельствует значительное количество научных публикаций как в отечественных, так и в зарубежных источниках. Одновременно с этим большинство применяемых в настоящее время методов решения прикладных задач опирается в основном на экспертные методы формирования сценариев, которые разрабатываются и исследуются «вручную». При этом формализованные методы решения рассматриваемых задач если и применяются, то их функционал ориентируется преимущественно на специфические особенности конкретной и строго ограниченной области исследования, т.е. являются в значительной мере узкоспециализированными, что объективно ограничивает область их практического использования. При этом существенно возрастает актуальность исследований комплексных проблем формализации задач сценарного подхода, направлений алгоритмической и программной реализации процессов разработки сценарных моделей, генерации и анализа сценариев,

а также формирования сценарно-прогнозных оценок реализуемости и качества альтернативных управленческих решений с целью информационно-аналитической поддержки процессов их подготовки.

В целом проблемы оценивания и выбора эффективных стратегических решений при управлении гео- и военно-политическими, социальными, макроэкономическими, научно-техническими, производственно-технологическими, информационно-психологическими и др. процессами, оказывающими непосредственное влияние на уровень НБ государства, являются исключительно сложными. В процессе решения задач такого типа необходимо использование адекватных их сложности и разнообразию формализованных моделей, позволяющих трансформировать экспертные оценки возможных путей развития проблемных ситуаций или характера деструктивного воздействия внешних и внутренних угроз в математические конструкции, обеспечивающие возможность разработки соответствующих средств автоматизации сценарного анализа и поддержки принятия решений.

2. Формализованные методы формирования и анализа сценариев развития обстановки

2.1. ПОНЯТИЙНЫЙ АППАРАТ СЦЕНАРНОГО ИССЛЕДОВАНИЯ

Основопологающим принципом в решении проблем управления безопасностью сложных систем является моделирование полного цикла процесса выработки, принятия и реализации решений, при этом следует учитывать следующие особенности:

- каждый объект или ситуацию в сфере безопасности необходимо рассматривать в качестве структурной части (системного элемента) более сложной системы [4];
- необходимы определения роли, интересов, ресурсов и др. свойств каждой исследуемой ситуации или объекта в сфере безопасности в общем процессе, направленном на обеспечение безопасного и устойчивого развития страны;
- обязательным элементом системы сценарных моделей являются субъекты действия, способные для реализации своих целей и интересов оказывать активное влияние на характер развития

политических, экономических, социальных и иных процессов как на объекте управления, так и во внешней среде;

– оценка результативности внешних или управленческих воздействий на исследуемый объект в сфере безопасности должна формулироваться только после завершения исследования всех фаз процесса управленческого воздействия и включать не только конечный итог, но и важные с точки зрения безопасности страны промежуточные результаты.

Далее введем ряд определений, необходимых для формирования общей методологической схемы сценарного исследования.

Определение 1. Системный элемент – формальное описание рассматриваемого подмножества составляющих системы, которая в рамках данного исследования представляет собой заданный набор объектов, объединенных общими отношениями (взаимодействиями) и обладающих определенными свойствами, и реализующих определенные функции и задачи.

Определение 2. Экспертно-значимая декомпозиция исследуемой системы на системные элементы – объединение системных элементов по заданному набору параметров. Естественная группировка системных элементов осуществляется по традиционным стратам: экономическая, информационная, политическая, военная и др.

Определение 3. Расширенное фазовое пространство – объединение (прямое произведение) фазового пространства исследуемой системы и внешнего пространства (внешней среды). В качестве элементов фазового пространства могут выступать аналитические данные моделирования, например показатели, отражающие характер динамики изменения факторов во времени, определяющих состояние исследуемой системы.

Определение 4. Разбиение расширенного фазового пространства – подмножество данного пространства, переменные которого выделяют по определенному критерию [9]. Разбиения фазового пространства позволяют формировать качественно различные сценарии изменения обстановки. В сценарном исследовании набор таких критериев (правила выбора) и их значимость для реализации целей управления можно задавать экспертным путем, в этом случае говорят об экспертно-значимых разбиениях.

Определение 5. Правила выбора экспертно-значимых разбиений – критерий, по которому происходит выделение экспертно-значимых разбиений.

Определение 5. Состояние объекта управления – совокупность значений эндогенных и экзогенных переменных (точка в расширенном фазовом пространстве) объекта управления (ОУ).

Определение 7. Модель измерения состояния ОУ – совокупность правил фиксации (определения) состояний ОУ.

Определение 8. Событие – фиксация текущего состояния ОУ посредством анализа модели измерения состояний (точка в расширенном фазовом пространстве).

Определение 9. Динамическая модель поведения ОУ – совокупность состояний ОУ, фиксированных на основе модели измерения в заданные моменты времени. Множество таких состояний представляет собой траекторию поведения ОУ.

Определение 10. Шкала траекторий – шкала модельного времени (множество моментов времени), определяющая моменты фиксации траектории поведения объекта в соответствии с правилами выбора.

Определение 11. Шкала событий – дискретная шкала, определяющая последовательность экспертно-значимых событий.

Определение 12. Квазиинформационная гипотеза (КИГ) – формальное описание неопределенности в заданный момент времени, учитываемой при формировании сценариев. В КИГ входит возможная реакция ОУ, а также изменение структуры, свойств отношений взаимодействия и т.п.

Определение 13. Экспертно-значимое событие (ЭЗС) – выделенное в соответствии с правилами выбора состояние исследуемой системы (точка в экспертно-значимом разбиении фазового пространства).

2.2. ОБЩАЯ СХЕМА ФОРМИРОВАНИЯ СЦЕНАРИЯ

На основе использования приведенного выше понятийного аппарата сценарного исследования разработана общая схема формирования сценария (рис. 2). Основная задача автоматизации процессов генерации сценариев изменения обстановки – описать экспертные неформализованные знания о предметной области исследования формальными математическими конструкциями.

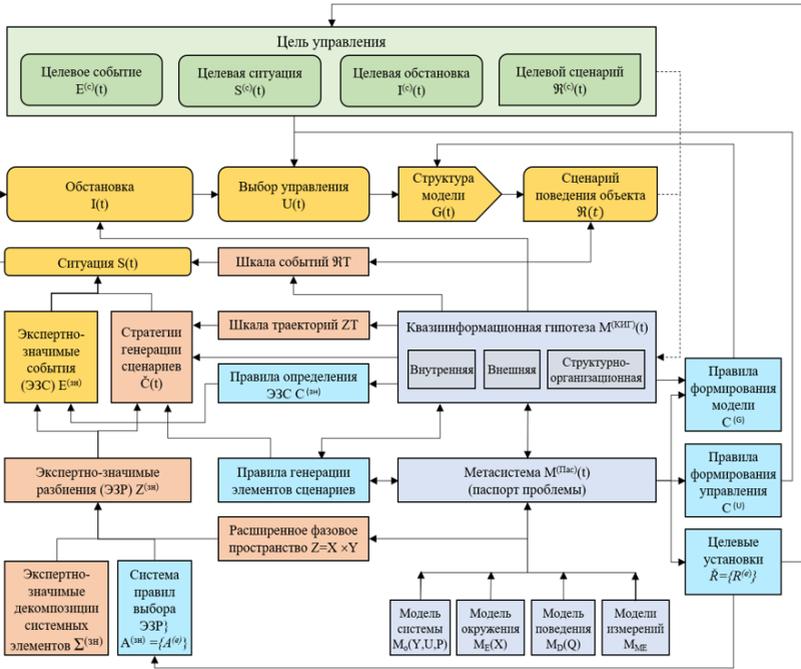


Рис.2. Общая схема формирования сценария

Цель – создать спектр сценариев поведения исследуемой системы как модели развития обстановки при реализации альтернативных вариантов внешних и внутренних условий и управленческих воздействий.

Предлагаемая схема описывает разветвленный многоуровневый механизм формирования сценария. С одной стороны, он отражает основные этапы процесса моделирования изменения обстановки при функционировании сложной системы. С другой – при развернутом сценарном исследовании условий функционирования ОУ позволяет осуществлять полный цикл стадий и этапов сценарного анализа и синтеза.

Результатом применения разработанной схемы является сгенерированный в автоматизированном режиме спектр сценариев (модель развития обстановки).

С этой целью в схеме формирования сценария на начальном этапе формализуются три основные компоненты проблемной ситуации.

1. Основные формальные модели, описывающие совместное поведение ОУ и его окружения (системные метаэлементы, мета-система) и базирующиеся на экспертном описании проблемной ситуации (паспорт проблемы) [9].

2. Формальная модель неопределенности – квазиинформационная гипотеза.

3. Формализованное описание предметной области на основе выделения экспертно-значимых декомпозиций, экспертно-значимых разбиений (ЭЗР) расширенного фазового пространства и экспертно-значимых событий (ЭЗС) [4, 9].

4. В основе сценарного исследования в широком смысле и сценарного моделирования – в узком лежит анализ данных как об исследуемом объекте (системе), так и об обстановке, влияющей на его характеристики, функционирование и развитие.

Среди параметров системы можно выделить [4]: вектор экзогенных переменных $x \in X$; вектор эндогенных переменных $y \in Y$; вектор управляемых переменных $u \in U$, вектор ресурсов $p \in P$ и ограничения Q , которые накладываются на поведение объекта или развитие ситуации.

Для сбора и представления исходных данных используется ряд системных элементов сценарной системы, представляющих собой следующие модели [9, 4]:

- идентифицированная модель системы – $M_O(Y; U; P)$;
- модель окружения – $M_E(X)$;
- модель поведения – $M_D(Q)$;
- модель измерения состояний системы – M_{MO} ;
- модель измерения состояния окружения – M_{ME} .

Совокупность выделенных моделей является метанабором сценарной системы и служит основой создания и поддержания в актуальном состоянии паспорта проблемной ситуации:

(1) $M = (M_O(Y; U; P); M_E(X); M_D(Q), M_{MEO})$.

Прежде всего необходимо идентифицировать исследуемую систему как объект управления, выделив и описав параметры

и связи между ними в модели $M_O(Y; U; P)$. Аналогичное описание необходимо для экзогенных переменных в модели $M_E(X)$, которые служат основой для получения спектра альтернативных сценариев, поскольку именно с помощью данной модели проводится анализ параметров окружения и формируются альтернативные гипотезы об изменчивости внешней среды, в том числе с учетом неопределенности.

При формировании, а затем и исследовании сценарной модели развития ситуации в сфере безопасности необходимо сочетание обеих моделей $M_O(Y; U; P)$ и $M_E(X)$ в рамках составного системного элемента сценарной системы \tilde{S}_{OE}^{SC} с набором внутренних элементов и элементов окружения (y, u, p, x) . Конкретный состав элемента сценарной системы \tilde{S}_{OE}^{SC} , как и состав моделей $M_O(Y; U; P)$ и $M_E(X)$, определяется в соответствии с целью сценарного исследования. В качестве примера здесь и далее используется математический аппарат функциональных графов. В основе сценарных графовых моделей лежат следующие базовые понятия [4–7, 19–20]:

- Фактор – любой, являющийся значимым в модели для представления и анализа систем или ситуаций показатель с учетом заданной степени детализации. Множество факторов (вершин графа): $F = X \cup Y \cup U \cup P$.

- Взаимодействие факторов (отношения между факторами) – количественное или качественное описание воздействия изменения одного фактора на изменение других факторов. В простейшем случае это веса дуг (множество E), связывающие эти факторы.

- Состояние – это текущие значения параметров факторов, значимое для оценки ситуации. Анализ состояния факторов служит основой для формирования событий сценария.

- Событие – это смена состояния одного или нескольких факторов.

- Сценарий поведения системы – последовательность событий, показывающая, как из существующего или какого-либо заданного состояния управляемая система этап за этапом переходит в другое состояние.

Модель $M_D(Q)$ задает преобразование параметров системы и определяет характер их взаимодействия с параметрами окружения (внешней среды), что в итоге характеризует динамику изменения фазовых состояний. Представленный в качестве примера математический аппарат является расширением классической графовой модели: кроме орграфа $G(F, E)$, где F – конечное множество вершин, а E – множество дуг графа, в модель включаются дополнительные компоненты. В частности, вводится множество параметров вершин [15]:

$$V = \{ v_i, i \leq N = \|X\| \}.$$

В соответствие каждой вершине x_i ставится ее параметр $v_i \in V$. Вводится также функционал преобразования дуг $F(V, E)$, т.е. в соответствие каждой дуге ставится функция, которая в частных случаях является знаком $(+1, -1)$ либо весом ($W_{ij} \in \mathbb{R}$):

$$F(v_i, v_j, e_{ij}) = f_{ij}(v_i, v_j).$$

На расширенных таким образом орграфах вводится понятие импульса и импульсного процесса в дискретном временном пространстве. Импульсом $P_i(n)$ в вершине x_i в момент времени $n \in N$ называется изменение параметра в этой вершине в момент времени n :

$$P_i(n) = v_i(n) - v_i(n - 1).$$

Значение параметра в вершине x определяется соотношением

$$v_i(n) = v_i(n - 1) + \sum_{j=1, j \neq i}^N F(v_i, v_j, e_{ij}) P_j(n - 1) + P_i^0(n).$$

Здесь $P_i^0(n)$ — внешний импульс, вносимый в вершину e_i в момент времени n . Из двух последних конечно-разностных уравнений легко получить уравнение для импульса в исследуемом процессе:

$$P_i(n) = \sum_{j=1, j \neq i}^N F(v_i, v_j, e_{ij}) (P_j(n - 1) + P_i^0(n)).$$

Сегодня уже накоплен определенный опыт использования аппарата функциональных графов в качестве методологической основы автоматизации процессов сценарного исследования широкого круга проблем НБ. На рис. 3 представлены начальные

уровни иерархии сценарных моделей, структура которых в виде множеств факторов и взаимосвязей между ними, а также полученные при их исследовании сценарии представлены в [4, 15, 19–20]. Классификация сценарных моделей при этом соответствует аспектам НБ РФ [13].



Рис. 3. Национальные интересы, стратегические национальные приоритеты и сценарные модели

Особое место в общей схеме формирования сценариев в части представления результатов, на основе которых принимаются управленческие решения, занимает модель измерения состояний $M_{МЕО}$. Она включает две модели: модель измерения состояния системы $M_{МО}$ (объекта управления или проблемной ситуации) и модель измерения состояния внешней среды $M_{МЕ}$.

В основе любого сценария лежит определение событий $E^{(3н)}$ (ЭЗС), которые формируются на основе изменяемых фазовых со-

стояний системы. Всякий раз определение $E^{(зн)}$ должно осуществляться формализованно на основе модели измерений. От выбора модели измерений зависит оценка обстановки и, соответственно, выбор конкретного управленческого решения. Следовательно, модель измерений также может являться объектом косвенного управления (например, реализованного в форме экономического, информационно-психологического и т.д. воздействия) [4]. Поскольку технологически сценарий представляет собой упорядоченную во времени последовательность событий, на основе оценивания которых, как и сценария в целом, происходит выработка управленческих решений, то критически важным является возможность идентификации и «отсечки» данных событий в полученной общей картине, образуемой совокупностью динамических характеристик изменения значений определенного подмножества факторов модели на значительном временном интервале. С этой целью используются процедуры анализа динамики факторов, полученной в результате сценарного моделирования. Пусть для i -го фактора в момент времени t определен тип динамики $FSc(t)$. Для определения события используется пять типов поведения факторов, алгоритм расчета которых приведен в [15], например: рост ($FSc(t) = 1$); падение ($FSc(t) = 2$), неизменность ($FSc(t) = 3$) и т.д. Тогда события сценария являются функцией от поведения факторов:

$$E_j^{(зн)}(t) = f(FSc(v_1(t)), FSc(v_2(t)), \dots, FSc(v_m(t))).$$

В качестве иллюстрации можно привести следующие характерные примеры поведения факторов, определяющие угрозы (формирующие нежелательные события) в сфере НБ: рост инфляции, рост дефицита федерального бюджета, рост риска нарушения режима нераспространения ядерного оружия, падение рождаемости, снижение уровня жизни граждан, рост киберпреступности и кибертерроризма, падение объемов геологоразведки полезных ископаемых, рост объемов «утечки» капиталов, рост дефицита квалифицированной рабочей силы, рост уровня зависимости от импорта современных технологий и т.д.

Результаты сценарного исследования, например, полученные с помощью специализированного программного комплекса, основанного на использовании рассмотренного математического

аппарата, по сути, являются «протосценариями», поскольку для получения пригодного для дальнейшего анализа сценария необходимо обеспечить смысловое соответствие между полученными расчетными результатами и терминологией конкретной предметной области [4, 9, 15]. Использование модели измерения состояний M_{MO} позволяет проводить качественный анализ важнейших параметров поведения исследуемого объекта или процесса развития ситуации, что способствует их адекватной идентификации при выработке управленческих решений. В частности, можно отметить, что различные модели измерений позволяют оценивать и интерпретировать события в пользу того или иного субъекта действия, что является одним из направлений подготовки управленческих решений. В ряде практических случаев, например, при решении задач информационного управления, модель измерения может выступать самостоятельно, не опираясь на реальные события [20].

Модель измерения M_{MO} регламентирует масштаб и варианты измерения объекта, так же как и модель измерения состояния внешней среды M_{ME} для внешнего окружения объекта. Степень детализации может зависеть от цели управления, сложившейся обстановки и возможностей субъекта управления. Следовательно, имеет смысл рассматривать объединённую модель измерения $M_{MEO} = (M_{MO}, M_{ME})$.

Таким образом, в качестве входной информации для системного элемента выступают значения внешних и внутренних параметров. Далее с помощью объединённой модели $M_{MEO} = (M_{MO}, M_{ME})$ пособытийно формируется сценарно-субъективное представление о состоянии фазового пространства. Следует отметить, что подобная субъективность построения модели измерения может являться как сильной, так и слабой стороной в управлении сложными организационными системами.

Системные метаэлементы представляют собой совокупность моделей, описывающих совместное функционирование ОУ и окружения: модель ОУ, модель окружения, модель их совместного взаимодействия, а также применяемые инструментальные средства измерения (фиксации) событий.

Объединение системных метаэлементов (метасистемы) составляет паспорт проблемной ситуации, который, в свою очередь, является основой формирования сценарной модели. Таким образом, паспорт представляет собой базу данных накапливаемой и изменяемой информации о существующих и новых ситуациях [9].

На основе паспорта проблемной ситуации формируется многомерное расширенное фазовое пространство Z – множество возможных состояний ОУ и его окружения.

Квазиинформационная гипотеза $M^{(КИГ)}(t)$ представляет собой формальное описание неопределенности (в том числе состояний (обстановок) других субъектов управления (СУ)) в момент времени t , учитываемой при формировании сценариев. Для принятия решения о возможном очередном ожидаемом событии в формируемом сценарии следует провести процедуру учета неопределенности. В качестве основы для этой процедуры могут, например, выступать:

- вероятностная модель, предполагающая принятие решения на основе построения математического ожидания значений неопределенных величин;
- модель игры с Природой, в которой проводится расчет рисков, а критерием принятия решений является их минимизация;
- модель согласования знаний экспертов;
- модель влияния внешней среды на процесс сценарного исследования;
- модель игры с активным противником, в которой принятие решения проводится на основе определения тех или иных равновесных ситуаций;
- модель рефлексивной неопределенности, когда принятие решения проводится на основе расчета возможных ответов противника на реализацию предпринимаемых действий.

Традиционные проблемы формирования моделей – это обеспечение их полноты и адекватности. Любая модель сложной системы строится на основе некоторых допущений, которые отражаются в КИГ, разрешая таким образом проблему структурной

неопределенности модели, а также выделяя и структурируя область множества условных решений, в которой будут разворачиваться сценарные события [4].

В зависимости от решаемых задач управления могут быть изучены различные компоненты КИГ: внешняя, внутренняя, структурная, функциональная, ресурсная и другие ее составляющие.

Формализованное описание предметной области представляет собой структурированное представление знаний экспертов о возможных направлениях развития исследуемой ситуации и последствия предпринимаемых действий.

С точки зрения системного анализа в основу формализованного описания предметной области положено понятие системного элемента. Перечень свойств (системных параметров), присутствующих рассматриваемым ОУ, определяется в квазиинформационной гипотезе. Примерами системных параметров являются как общеуправленческие свойства (активности, детерминированности, изменяемости, замкнутости и открытости, управляемости, устойчивости, уязвимости и т.п.), так и свойства, относящиеся к различным аспектам предметной области (форма собственности, отраслевая или региональная принадлежность, используемые ресурсы, методы и инструменты управления и т.д.). Экспертно-значимые декомпозиции системных элементов представляют собой их объединение по заданному набору системных параметров.

Пусть $\mathfrak{R}^{(c)} \in \check{R}$ – желаемый сценарий – одна из целей сценарного исследования, где \check{R} – множество целей. В качестве описания цели могут использоваться текстовые конструкции, что вполне соответствует экспертной значимости сценарной концепции для сложных систем, объектов и ситуаций.

Следующим шагом формализации предметной области является декомпозиция построенного расширенного фазового пространства Z в соответствии с целью сценарного исследования на экспертно-значимые его разбиения (ЭЗР) $Z^{(c)}$, в которых будут построены качественно различные сценарии изменения обстановки.

В основе формализованного экспертного описания расширенного фазового пространства лежит теоретико-множественный подход, предполагающий разбиение данного пространства на подмножества, характеризующие качественные, экспертно-значимые свойства изучаемого объекта (системы или проблемы) [4, 9]. Процесс моделирования осуществляется в компакте $Z^{(o)} \subseteq Z$, выделение которого происходит на основе экспертных оценок.

В случае, если множество целей велико, для упрощения процедуры сценарного исследования необходимо на первых этапах ограничить сбор информации для формирования расширенного фазового пространства и паспорта проблемы. Таким образом, сценарное исследование реализации цели $R^{(e)}$ будет основано на ограниченном метанаборе, соответствующем конкретной страте [9]:

$$(2) \quad M^{(e)} = (M_O^{(e)}(Y; U; P); M_E^{(e)}(X); M_D^{(e)}(Q), M_{MEO}^{(e)}),$$

где $M_O^{(e)}(Y; U; P)$, $M_E^{(e)}(X)$ – модели системы и окружения в e -й целевой страте; $M_D^{(e)}(Q)$ – модель поведения в e -й целевой страте; $M_{MEO}^{(e)} = \{M_{MO}^{(e)}, M_{ME}^{(e)}\}$ – модели измерения параметров системы и окружения в e -й целевой страте.

Те же самые соображения имеют смысл при разработке иерархии сценарных моделей. В исследуемой предметной области иерархия сценарных моделей соответствует иерархии организационных систем и зависит от уровня и специфики решаемых ими задач. Как правило, подчиненные организационные системы отвечают за реализацию целей верхнего уровня в одной или нескольких стратегиях. Следовательно, возможна организационная стратификация модели верхнего уровня.

В основе экспертного описания модели поведения объектов или развития ситуации, как уже было упомянуто выше, лежат понятия экспертно-значимого разбиения расширенного фазового пространства и экспертно-значимых событий, которые происходят в последовательности, указанной в системе упорядочения (шкала событий) [9]. В основе формирования шкал траекторий и событий лежит следующие параметры [17]: период оценки тен-

денций и динамики изменения значений факторов в шагах моделирования; количество шагов моделирования, учитываемое при анализе динамик факторов (имитируется глубина мониторинга моделируемой системы); количество шагов моделирования, используемое для имитации временной задержки процесса анализа исследуемой ситуации и принятия решения (период запаздывания управленческих решений в шагах моделирования).

Шкала событий вводится с целью дифференциации восприятия событий, а также их оценки при управлении в системе НБ РФ. Сценарная модель безопасности ОУ содержит условия безопасности, классификацию ситуаций опасности, а также возможности обеспечения безопасности функционирования и развития [17]. Например, в самом общем виде классификацию складывающихся в какой-либо из сфер НБ события и ситуации можно представить с помощью следующей шкалы:

- 1) штатная (находится в заданных пределах показателей и индикаторов безопасности);
- 2) целевая (полностью управляемая, находится в заданных пределах целевого режима функционирования);
- 3) ожидаемой опасности (прогнозируется возможный выход значений показателей безопасности за допустимые пределы);
- 4) кризисная (значения параметров безопасности вышли за допустимые пределы, но ситуация ограниченно управляема, причем существует несколько альтернативных путей ее нормализации);
- 5) критическая (существует единственный управленческий вариант разрешения возникших проблем);
- 6) неконтролируемая (развитие ситуации неуправляемо).

Различные варианты разрешения неопределенностей на основе КИГ формируют соответствующие связанные множества ЭЗР – ЭЗС – сценариев $\mathfrak{R}(t)$, определяемые также спектром управляющих воздействий $U(t)$ (аттрактивные сценарии) или отсутствием управления (синергические сценарии).

Этапы ограничения фазового пространства Z связаны с выделением системных элементов на основе экспертно-значимой их декомпозиции по различным критериям или правилам. Например, по критерию применения политических, военных, социальных, информационных, экономических и т.д. технологий.

Пусть $A^{(3H)}$ – множество правил (модель) выбора ЭЗР, соответствующих целевым установкам \check{R} .

Определение 14. Экспертно-значимым разбиением $Z^{(3H)} \in Z(A^{(3H)})$ расширенного фазового пространства Z в соответствии с целевыми установками является пара $Z^{(e)} = \{\{\Sigma^{(3H)}\}, A^{(e)}\}$ [9], где:

- $A^{(e)} \in A^{(3H)}$ – правило (модель) выбора экспертно-значимого разбиения (области пространства Z), соответствующего реализации поставленных целей $R^{(e)} \in \check{R}$;
- $\{\Sigma^{(3H)}\}$ – экспертно-значимые декомпозиции системных элементов.

В качестве инструмента определения ЭЗР может служить паспорт проблемной ситуации или дерево целей, составленное из измеряемых параметров $y(t)$ и $x(t)$.

В каждом из сформированных разбиений следует выделить экспертно-значимые события $E^{(3H)}$, которые определяются условиями их осуществления $C^{(3H)}$. Критериями (условиями) выделения событий могут быть:

- достижение заданных состояний;
- достижение нормативных состояний;
- достижение пограничных состояний (например, кризисное, критическое и т.п.);
- смена методов управления, в том числе цели, стратегии, механизмов реализации;
- смена динамики изменения значений факторов модели и т.д.

Критерии задаются экспертным путем исходя из состояния защищенности текущих национальных интересов РФ [15] и возможностей управления. Нормативы характеризуют состояние безопасности на каждом уровне организационной иерархии, например нормативы успешности (качества) функционирования конкретной организационной системы для реализации целей НБ на своем уровне иерархии.

Характеристики экспертно-значимых разбиений являются основой для формирования экспертно-значимых событий [19], которые являются наиболее существенными для лица, принима-

ющего решения (ЛПР), а также опорными событиями, характеризующими динамику изменения системы и являющиеся основой для формирования ситуации $S(t)$.

В процессе моделирования происходит формирование текущих событий (точек $\mathbf{z}^{(i)}$ в расширенном фазовом пространстве Z), последовательность которых в моменты времени t_i формирует траекторию в Z . Состояние системы в каждый момент времени может характеризовать экспертно-значимое событие, если оно находится в экспертно-значимом разбиении. Также в зависимости от целей сценарного исследования в качестве ЭЗС может выступать сам факт выхода состояния системы в определенный момент времени из ЭЗР.

Для того чтобы полностью задать модель сценарного исследования, необходимо задать сценарно-событийное пространство $(E, E^{(3н)}, C^{(e)})$, а именно:

- описать множество E возможных событий;
- ввести множество $E^{(3н)}$ подмножеств множества событий E , которые являются экспертно-значимыми, т.е. по сути задать область определения функции сценарной значимости $C^{(3н)}(E)$;
- задать способ определения сценарной значимости события $C^{(3н)}(E)$.

Текущее событие $\mathbf{z}(t)$ в момент времени t представляет собой точку в расширенном фазовом пространстве Z .

Определение 15. Экспертно-значимым событием $E^{(3н)}$ назовем тройку $(Z^{(3н)}, \mathbf{z}(t), t, C^{(3н)})$, где $C^{(3н)}$ – правила определения ЭЗС; $Z^{(3н)}$ – подмножество пространства Z ; $\mathbf{z}(t)$ – точка входа/выхода расширенной фазовой траектории движения объекта управления на множество $Z^{(3н)}$ в момент времени, задаваемый шкалой траекторий $t \in ZT$ [9].

Совокупность текущих событий (ЭЗС), наложенная на шкалу событий RT , до момента t фиксирует ситуацию $S(t)$ в момент времени t . Определение ситуации дает возможность получить исходные данные о реализациях предыдущих КИГ (накопленный опыт). В КИГ входит возможная реакция ОУ, изменение структуры и т.п.

Обстановка $I(t)$ в момент времени t объединяет ситуацию и квазиинформационную гипотезу, т.е. совокупность предыду-

щих событий и возможных изменений ситуации (возможных событий). Если моменты времени дискретны (используется дискретная шкала событий), то в качестве аргумента времени уместно употреблять обозначение t_i , где i – номер временного интервала. В таком случае переменная $\tau_i = t_{i+1} - t_i$ – шаг генерации сценария, который может быть постоянным или меняться в процессе исследования. Обстановка оценивается исходя из целей обеспечения НБ, которые заключаются в реализации национальных интересов и национальных приоритетов [13].

На разных этапах управления критериями могут быть достижение события, ситуации, обстановки или целевого сценария.

Обозначим ожидаемое событие как $E^{(p)}(t) = (\mathbf{z}(t), t)$.

Определение 16. Ситуация $S(t)$ в момент времени t – последовательность экспертно-значимых событий, происшедших к моменту времени t [9].

Определение 17.

$$(3) S(t) = \{E^{(3H)}(t_i), (t - s) \leq t_i \leq t\},$$

где s глубина ситуации (количество временных интервалов в дискретной шкале событий), t_s – горизонт событий.

Определение 18. Обстановка $I(t)$ в момент времени t определяется как [9]

$$(4) I(t) = (S(t), M^{(КИГ)}(t)),$$

где $S(t)$ – ситуация в момент времени t ; $M^{(КИГ)}(t)$ – квазиинформационная гипотеза в момент времени t .

Очередное событие $E^{(3H)}(t_i + 1)$ выбирается на основе стратегии формирования сценария $\check{C}(t)$ и в общем случае определяется ситуацией непосредственно до этого момента времени $S(t_i)$, а также на основе предполагаемых условий развития системы (функционирования объекта управления) $M^{(КИГ)}(t_i)$:

$$(5) E^{(3H)}(t_i + 1) = E(S(t_i), M^{(КИГ)}(t_i), \check{C}(t_i)).$$

Таким образом, очередное событие формируется в зависимости от обстановки и применяемой стратегии, в том числе с учетом правил учета неопределенности в рассматриваемом разбиении предметной области, механизма управления, наличия ресурсов и т.д.

Определение 19. Сценарием \mathfrak{R} поведения системного объекта является последовательность обстановок

(6) $\mathfrak{R} = \{(I(t_i), t_i), i = 0, 1, \dots, N\}$,

где $N \in \mathfrak{RT}$ – глубина сценария (количество временных интервалов в дискретной событийной шкале); $t_N = T \in ZT$ – горизонт сценария. Обстановкой $I(t_i)$ при $t_i \in ZT$ является пара: ситуация $S(t_i)$ и квазиинформационная гипотеза $M^{(КИГ)}(t_i)$.

Таким образом, очередное событие $E^{(зн)}(t_{i+1})$ в момент времени t_{i+1} по шкале ZT или в момент времени $i + 1$ по шкале \mathfrak{RT} определяются:

- ситуацией $S(t_i)$ в момент времени $t_i \in ZT$;
- квазиинформационной гипотезой $M^{(КИГ)}(t_i)$ в момент времени t_i ;
- стратегией формирования сценария $\check{C}(t_i)$ в момент времени t_i .

Далее приведены укрупненные этапы пошагового формирования сценария.

1. Формирование текущей ситуации $S(t)$ на основе ограниченного ряда предыдущих ЭЗС.

2. Формулирование КИГ на текущий момент времени $M^{(КИГ)}(t)$ и определение на ее основе возможных событий $ME(M^{(КИГ)}(t))$.

3. Выбор из множества возможных очередных событий $ME(M^{(КИГ)}(t))$ на основе стратегии формирования сценария $\check{C}(t)$ очередного ЭЗС $E^{(зн)}(t + 1)$.

4. Отображение текущей обстановки $I(t)$ на основе текущей ситуации и включение в нее результатов текущей КИГ.

5. Формирование сценария \mathfrak{R} как последовательности обстановок и переход к продолжению формирования цепочки событие-ситуация-обстановка-сценарий-событие. Таким образом, если выбранное на шаге 3 очередное событие в соответствии с заданным критерием является экспертно-значимым, то его следует включить в формируемый сценарий и продолжить формирование рассматриваемой цепочки.

Эффективно построенная сценарная система позволяет перейти к созданию системы обеспечения безопасности объекта, структуры или процесса функционирования сложной системы.

Объединение системно-логического, структурно-организационного и сценарного подходов дает возможность формализовать систему моделей иерархического децентрализованного группового управления [9]. Это позволяет проводить разработку эффективных сценарных механизмов управления комплексом взаимосвязанных мероприятий в сфере безопасности, осуществляемых различными субъектами управления.

Совокупность управляющих воздействий, удерживающих развитие системы в допустимой окрестности выбранной траектории, является оптимальным управлением в широком смысле.

В случае существования нескольких возможных путей достижения целей развития системы может быть поставлен вопрос о построении наилучшей траектории развития с точки зрения некоторого критерия эффективности управления. Допустимые управляющие воздействия, обеспечивающие экстремальные значения показателя эффективности, выраженного в виде целевой функции, являются оптимальным управлением в смысле заданного показателя эффективности.

3. Разработка информационно-аналитических моделей сценарного планирования процессов обеспечения безопасности на базе процессного подхода

Разработанная методология использования сценарного подхода в процессе решения задач планирования и управления в условиях возрастания неопределенности базируется на комплексе механизмов прогнозирования, а также диагностирования и анализа не только уязвимостей и угроз, но и оценки возможностей и потенциала управления с учетом альтернативных вариантов возможного развития обстановки [7, 8, 18, 20]. Под потенциалом управления здесь понимается прямая или косвенная возможность достижения целей управления. Необходимость введения этого понятия обусловлено тем, что в Стратегии национальной безопасности отсутствует понятие обратное угрозе, тем не менее Стратегия направлена не только на обеспечение состояния защищенности, но и на реализацию национальных интересов и стратегических национальных приоритетов страны.

Сценарии при этом служат инструментом снижения большей части неопределенностей, ограничивая их несколькими альтернативными направлениями развития ситуации. Возможные пути и способы учета неопределенности в процессе сценарного моделирования представлены в таблице 1.

Таблица 1. Способы учета различных видов неопределенности в процессе сценарного имитационного моделирования

| Виды неопределенности | Сценарный подход |
|--|--|
| Неопределенность условий и исходных данных как об объекте управления, так и о состоянии факторов внешней среды | <p><u>Вариант 1:</u> Формирование и исследование качественных моделей на основе анализа причинно-следственных связей, которые малочувствительны к точности и достоверности исходных данных.</p> <p><u>Вариант 2:</u> Дополнение имеющейся неполной информации об обстановке прогнозными данными качественного характера о возможных событиях и последствиях реализации управленческих решений.</p> |
| Неопределенность целеполагания (в том числе в условиях многостороннего конфликта). Неоднозначность, а иногда и невозможность выбора единой цели. | Формирование и анализ сценариев, оценивающих уровень противоречивости и достижимости целей управления обеспечением НБ при различных условиях. |
| Неопределенность в оценке достижимости целей. | Анализ сценариев, оценивающих возможность достижения целей управления путем воздействия на ограниченное множество управляемых факторов сценарной модели, в том числе со стороны как взаимодействующих, так и конфликтующих (конкурирующих) субъектов. |
| Неопределенность действий, связанная с отсутствием однозначности критериев выбора решений. | Формирование и анализ альтернативных прогнозных сценариев с целью оценки последствий реализации принимаемых управленческих решений. |

Таблица 1 (продолжение)

| | |
|---|---|
| <p>Перспективная неопределенность вследствие возможности появления факторов, которые при подготовке решения были неизвестны, либо недооценены в процессах анализа предметной области.</p> | <p>Оперативное изменение структуры сценарной модели и управление процессом моделирования с целью определения необходимых действий, направленных на корректировку целей или путей их достижения.</p> |
| <p>Ретроспективная неопределенность, связанная с отсутствием информации о поведении изучаемого объекта или процесса в прошлом, или при принципиальной невозможности использования такой информации для решения задач управления из-за асимметрии событий в прошлом и будущем.</p> | <p>Формирование сценарных моделей на основе исследования различных трендов развития ситуации и экспертных оценок.</p> |
| <p>Неопределенность, связанная с невозможности предсказания точных результатов принимаемых решений из-за лавинообразного нарастания объема исходной и оперативной информации об объекте управления, внешней среде или обстановке.</p> | <p>Формирование и исследование сценарных моделей с использованием устойчивых закономерностей на основе знаний, полученных в результате предварительного анализа больших объемов информации.</p> |

Прогнозные сценарии, полученные в результате моделирования, целесообразно использовать в качестве исходной информации для целеполагания и в дальнейшем – для планирования. С другой стороны, сценарное моделирование помогает строить прогнозы развития ситуации на различных уровнях управления и на основе результатов их анализа корректировать стратегическое видение ситуации. Таким образом, прогноз можно считать особым и востребованным в рассматриваемых условиях методом управления в организационных системах [12].

3.1. СЦЕНАРНОЕ ПЛАНИРОВАНИЕ В УПРАВЛЕНИИ ОБЕСПЕЧЕНИЕМ НБ

При использовании сценарного подхода в процессах разработки долгосрочных и среднесрочных планов мероприятий по обеспечению безопасности можно выделить две базовые укрупненные стратегии [4, 20]:

- сфокусированное сценарное планирование (СфП), основанное на принятом в качестве базового единственного сценария;
- сбалансированное сценарное планирование (СбП), основанное на использовании нескольких, описывающих возможные альтернативы развития событий, различных сценариев.

В рамках выделенных стратегий для решения практических задач возможно использовать следующие подходы к преобразованию сценариев в конкретные планы.

I. *Реакция на наиболее вероятные события.* В качестве основного преимущества такого подхода можно отметить использование единственного сценария в процессе разработки стратегических, тактических и оперативных планов как единого и достаточно компактного описания последовательности и взаимозависимости будущих событий, что в определенной мере упрощает процессы и процедуры разработки планов. К недостаткам можно отнести довольно значительную зависимость результатов от точности и достоверности прогнозных оценок характера и тенденций изменения ситуации и, соответственно, деструктивного влияния внешней среды, что порождает остаточную неопределенность. Ее снижение возможно за счет расширения области сценарного анализа или проведения последующих уточняющих сценарных исследований по мере поступления новой информации или приобретения новых знаний (уже в рамках стратегии СбП).

II. *Минимизация возможных ущербов.* При таком подходе целесообразно рассматривать в качестве его основы пессимистические сценарии нанесения максимально возможного ущерба от угроз, как очевидных, так и прогнозируемых. Подход применим в том случае, когда небольшая вероятность угроз связана с критичными крупномасштабными негативными последствиями. Недостатками данного подхода является возможная пассивность

управленческих решений, а также нерациональное использование резервируемых ресурсов, которые, как правило, только способствуют наихудшему варианту возможного развития обстановки.

III. *Гибкость планирования* позволяет за счет использования альтернативных сценариев соотносить плановые мероприятия не только с исполнителями, сроками и желаемыми результатами, но и с происходящими или прогнозируемыми событиями как позитивного, так и негативного толка. Это делает возможным упреждающий учет возможных изменений социальных, политических, организационных, экономических и других условий не только на начальной стадии планирования, но и на последующих этапах, что позволяет корректировать приоритеты, цели, набор задач и критерии эффективности.

IV. *Максимизация результативности.* При данном подходе к планированию при использовании стратегии СфП в качестве основы берутся оптимистические сценарии достижения поставленной цели с заданным уровнем эффективности. При этом уже на начальных этапах управленческого цикла имеется возможность оценивать имеющийся потенциал и определять направления его эффективного использования для успешной реализации сформулированных задач. Одновременно с этим появляется возможность прогнозировать рост потенциала управления при решении заданного множества функциональных задач особенно на длительном временном горизонте. Однако в рамках стратегии СфП негативным последствием выбора оптимистического сценария в качестве базового является отсутствие возможности полноценного учета высоких рисков при прочих альтернативах развития ситуации. Решение данной проблемы лежит в использовании стратегии СБП, предполагающей более сбалансированный подход к управлению безопасностью. В рамках данной стратегии имеется возможность анализа расширенного множества альтернативных сценариев развития ситуации и учета неблагоприятного влияния процессов и факторов внешней и внутренней среды.

V. *Формирование благоприятных внешних условий.* Основой стратегического целеполагания и планирования при этом под-

ходе также является позитивный сценарий, назначением которого является выявление или формирование благоприятных условий управления, направленного на реализацию поставленных целей. Реализация этой задачи осуществляется на основе планирования управленческих воздействий со стороны организационных систем на внешние и внутренние источники угроз или на их компенсацию. При этом необходимо рассматривать ряд локальных сценариев оценки результативности возможного противодействия при ликвидации источников угроз, особенно критически важных. Рассматриваемый подход как инструмент прямого или косвенного влияния на происходящие или прогнозируемые события, а также условия их осуществления, носит ярко выраженный проактивный и «наступательный характер».

При традиционном подходе, когда разработка и анализ сценариев развития ситуации осуществляется экспертами «вручную» и при этом требуется значительный объем временных затрат, в основном развивалась стратегия СфП [4]. Применение средств автоматизации процессов генерации сценариев позволяет перейти от планирования, основанного на использовании единственного базового сценария, к задействованию множества необходимых для принятия плановых решений альтернативных сценариев. При этом появляется возможность проведения сценарных исследований с целью опережающей оценки эффективности принимаемых решений практически на любом этапе подготовки и реализации разрабатываемых плановых документов на основе как формируемой прогнозной, так и поступающей оперативной информации о развитии или изменении ситуации.

Любой перспективный план или программа в процессе разработки и практической реализации проходит несколько стадий своего жизненного цикла: от выявления требующих решения проблем и до финальной оценки полученных результатов. Принятые на любом этапе неэффективные и ошибочные управленческие решения влекут за собой целый комплекс проблем вплоть до необходимости возврата на начальные этапы целеполагания и планирования с целью внесения существенных корректировок, которые, в свою очередь, приведут к необходимости внесения изменений по всему управленческому циклу [15].

Повысить результативность планирования и управления, снизить влияние имеющейся неопределенности, а также сократить риски позволяет разработка и применение моделей жизненного цикла (ЖЦ), обеспечивающих возможность эффективного управления взаимосвязанными и иерархическими по структуре взаимосвязей этапами рассматриваемых процессов.

3.2. РАЗРАБОТКА КОМПЛЕКСНОЙ W-ОБРАЗНОЙ МОДЕЛИ ЖИЗНЕННОГО ЦИКЛА ПРОЦЕССА УПРАВЛЕНИЯ НАЦИОНАЛЬНОЙ БЕЗОПАСНОСТЬЮ НА ОСНОВЕ КОНЦЕПЦИИ СЦЕНАРНОГО АНАЛИЗА И ПРОГНОЗИРОВАНИЯ

В качестве исходной для формирования модели жизненного цикла взята V-образная модель, поскольку с точки зрения системы обеспечения НБ в ней важное место занимает планирование, с повышенным постоянным вниманием к стадиям контроля и аттестации на ранних этапах разработки планов мероприятий [1, 3, 6]. Здесь жизненным циклом является процесс последовательного изменения состояния объекта, обусловленный видом производимых на него воздействий. Планирование и управление обеспечением НБ также является процессом последовательного изменения состояния сложной системы, который должен находиться под управленческим воздействием.

Основным недостатком классической V-образной модели жизненного цикла является аттестация результатов по факту их достижения [16]. Кроме того, следует учесть современные тенденции развития методов и организационных механизмов обеспечения НБ в реалиях сегодняшнего дня. С учетом данных тенденций для реализации целей обеспечения НБ предлагается модифицировать V-образную модель, добавив в нее промежуточную ветвь, отражающую использование сценарного подхода для решения рассматриваемых задач, осуществив, таким образом, переход к W-образной модели сценарного планирования и управления. Целесообразность данного перехода определяется тем, что сценарный подход основан на способах и технологиях управления неопределенностями будущего и направлен на выявление изменений в окружающем мире, новых потребностей и вызовов,

а также выработку ответных действий [11, 14]. Если в V-образной модели подчеркивается важность верификации и аттестации результатов планирования и управления, то W-образная модель наряду с традиционным анализом результатов предоставляет возможность упреждающей или прогнозной верификации, что позволяет перейти от пассивного учета уже свершившихся событий к упреждающему управлению обеспечением НБ [15]. Возможности разработанного сценарного подхода и его ориентированность на упреждение позволяет продлить ветви планирования до более ранних этапов управленческого цикла, оказывающих влияние не только на процессы целеполагания, связанные с развитием собственно объекта управления, но и на оценку целевых установок остальных активных акторов, оказывающих непосредственное влияние на уровень НБ РФ.

Предлагаемая модель жизненного цикла должна объединять два принципа и, соответственно, две взаимосвязанные части. Верхняя часть модели ориентирована на контроль обеспечиваемого текущего уровня НБ и анализ тенденций его изменения, что является необходимым условием для эффективного целеполагания. Нижняя часть модели должна быть ориентирована на управление целями и конкретными механизмами их достижения. Разработанная модель жизненного цикла за счет применения сценарной методологии должна объединять два процесса: во-первых, трансформацию национальных интересов и стратегических приоритетов в конкретные цели и планы, обладающие различной степенью детализации и взаимоувязанные по иерархии управленческого процесса, и во-вторых, – их реализацию конкретными организационными системами.

Процесс сценарного планирования и управления должен включать фазы стратегического, тактического и оперативного управления. Внутри каждой фазы должны осуществляться нескольких этапов сценарного исследования: создание сценария развития обстановки, выявление возможностей и уязвимостей; стратегическое планирование, связанное с задействованием и созданием новых возможностей, а также снижением уязвимостей и угроз достижению целей управления; разработку управляющих решений в форме направлений и моментов воздействий и форми-

рование целей для субъектов управления нижних уровней иерархии. При этом всегда должна иметься возможность вернуться на предыдущую фазу [20].

Применение методологии сценарного планирования и управления безопасностью в сложных иерархических системах с группами объектов, обладающих собственными системами управления, а также собственными интересами и целями, позволяет в меняющемся сверхконкурентном мире в условиях нестабильности, неопределенности и деструктивных воздействий перейти от интеграции объектов и процессов в рамках отдельного мероприятия при реализации относительно постоянных целей к стратегической и основанной на сценарном анализе альтернативных вариантов будущего интеграции, осуществляемой в прогнозируемых временных и событийных периодах. Долгосрочный прогноз развития в сфере безопасности, используемый в качестве основы планирования конкретных мероприятий, должен являться результатом серии взаимосвязанных сценарных прогнозов, полученных на предыдущих верхних уровнях иерархии [10].

3.3. СТРУКТУРА И ОСНОВНЫЕ ЭЛЕМЕНТЫ W-ОБРАЗНОЙ МОДЕЛИ ЖИЗНЕННОГО ЦИКЛА РЕАЛИЗАЦИИ ПЛАНИРОВАНИЯ И УПРАВЛЕНИЯ

Целевые параметры процессов управления обеспечением НБ на стратегическом уровне определяются основополагающими доктринальными документами федерального уровня, устанавливающими долгосрочные цели, основные направления и механизмы реализации государственной политики в области обеспечения национальной, международной, военной, государственной и т.д. безопасности, а также устойчивого социально-экономического, инновационно-технологического и т.д. развития государства и общества. На тактическом и оперативном уровнях рассматриваемые параметры определяются уполномоченными органами законодательной и исполнительной власти в рамках предоставленных им действующим законодательством полномочий и возложенных обязанностей.

В качестве примера можно привести целевые показатели оценки состояния национальной безопасности, содержащиеся

в предпоследней редакции «Стратегии национальной безопасности Российской Федерации» [13].

- удовлетворенность граждан степенью защищенности своих конституционных прав и свобод, личных и имущественных интересов, в том числе от преступных посягательств;
- доля современных образцов вооружения, военной и специальной техники в Вооруженных Силах Российской Федерации, других войсках, воинских формированиях и органах;
- ожидаемая продолжительность жизни;
- валовой внутренний продукт на душу населения;
- децильный коэффициент (соотношение доходов 10 процентов наиболее обеспеченного населения и 10 процентов наименее обеспеченного населения);
- уровень инфляции;
- уровень безработицы;
- доля расходов в валовом внутреннем продукте на развитие науки, технологий и образования;
- доля расходов в валовом внутреннем продукте на культуру;
- доля территории Российской Федерации, не соответствующая экологическим нормативам.

В последней редакции (2021 г.) Стратегии основное внимание уделено путям достижения целей обеспечения национальной безопасности и решения задач, предусмотренных в рамках стратегических национальных приоритетов [13].

Используемый в процессе разработки сценариев набор факторов и параметров играет исключительно важную роль в управлении обеспечением национальной безопасности, поскольку во многом определяет качество подготовки и принятия управленческих решений не только на федеральном, но и на отраслевом и региональном уровнях и, по сути, является основой их информационно-аналитического обеспечения.

В самом общем виде состав показателей национальной безопасности крайне широк и включает внешнеполитические, военно-политические, макроэкономические, социальные, демографические, внутренние общественно-политические, финансово-экономические, организационно-управленческие, институцио-

нальные, юридические (нормативно-правовые), социокультурные, межконфессиональные (межэтнические), информационно-коммуникативные, экологические, производственные, техногенные, природно-климатические и др. группы факторов [14, 20].

При этом задача выделения подмножества наиболее значимых из них является достаточно сложной, поскольку состав выделенных факторов индивидуален не только для различных составляющих национальной безопасности, но и в значительной степени зависит от целевых установок и конкретного состава решаемых функциональных задач.

Процесс принятия решения в сфере НБ должен основываться на результатах анализа текущих и сценарно-прогнозных оценках тенденций развития международной обстановки (МО), политической обстановки (ПО), а также социально-экономических систем (СЭС). Предлагаемая W-образная модель (рис. 4) состоит из трех ветвей и двух частей.

Верхний уровень модели ориентирован на контроль состояния сферы НБ и процесса ее развития, а также видения того, что является необходимым для определения системы целей и состава мероприятий обеспечения НБ.

Нижний уровень модели ориентирован на управление целями, конкретные механизмы их достижения и результаты управления. Предлагаемая модель объединяет оба процесса: превращение идей, результатов анализа обстановки и видения складывающейся ситуации в конкретные цели и планы, а также управленческие решения.

Следует отметить, что подобная полная двухуровневая W-образная модель жизненного цикла применима только для верхнего уровня планирования и управления НБ, оценивающей основные тенденции развития ситуации в МО, ПО и СЭС. Назначение верхнего уровня W-образной модели является основой для инициирования конкретных мероприятий безопасности и выработки общей цели. Для объектов нижних уровней иерархии в качестве основы будет использоваться вторая (нижняя) часть W-образной модели жизненного цикла процессов планирования и управления.

Инструменты автоматической генерации сценариев должны быть задействованы на всех этапах планирования для выявления

угроз, уязвимостей и возможностей управления безопасностью сложных многоуровневых организационных систем.



Рис. 4. W-образная модель жизненного процесса планирования и управления национальной безопасностью

Таким образом, в процессе планирования – от анализа условий (наблюдение) до исполнения (действие) – методы создания сценариев представляют собой эффективные инструменты для выявления изменений, не только возникающих уязвимостей и угроз [7], но и возможностей и потенциала управления. Рассмотрим основные этапы использования W-образной модели для решения рассматриваемых задач.

I. Первый (верхний) уровень W-образной модели жизненного цикла планирования и управления безопасностью:

1-й этап: долгосрочный прогноз развития мировых тенденций, который лежит в основе оценки уровня НБ РФ, а также тенденций развития будущей международной обстановки в будущем. Этап реализуется в основном на базе использования экспертных процедур и оценок. Роль сценарного моделирования на этом этапе не является решающей, а заключается в его использовании только в качестве одного из инструментов поддержки экспертных процедур. На этом этапе также выделяются основные движущие силы в мире, которые могут повлиять на саму формулировку проблемы, критерии эффективности управления и направления развития сферы НБ;

2-й этап: долгосрочный прогноз развития МО, который существует в форме множества альтернативных сценариев, созданных на основе единой сценарной модели, включающей интересы сторон исследуемой ситуации;

3-й этап: долгосрочный прогноз альтернатив развития политической обстановки как следствия развития МО, обладает существенными политическими особенностями, отраженными в соответствующих сценарных моделях;

4-й этап: прогноз динамики развития СЭС, войн и конфликтов [12];

5-й этап: прогноз социально-экономического, научно-технического и военно-политического развития страны, который является основой для планов общественного и государственного развития, создания материальной, интеллектуальной и социальной базы для обеспечения НБ [12] (для реализации данного этапа необходимы создание и поддержка в актуальном состоянии интегральной сценарной модели развития РФ);

II. Второй (нижний) уровень W-образной модели жизненного цикла планирования и управления национальной безопасностью:

6-й этап: целевые параметры безопасности определяются на основе сценарного прогноза как внешних, так и внутренних особенностей развития МО и ПО и СЭС РФ. На основе результатов выполнения этапов 2–5 выбираются наиболее актуальные целевые установки на множестве критериев НБ, формулируются желаемые значения и тенденции изменения факторов, которые составляют эти критерии. Сценарная поддержка реализации рассматриваемого этапа состоит в получении сценариев верификации поставленных целей на достижимость, причем наиболее важной функцией является получение прогнозных сценариев долгосрочных последствий целеполагания. Таким образом, отсеиваются ложные цели, выявление которых затруднительно или даже невозможно без сценарно-прогнозной экспертизы. Ложными целями являются направления обеспечения безопасности, которые невозможно осуществить, например, по причинам их противоречивости, ресурсно-временных ограничений и т.д.

7-й этап: оценка обстановки, направленная на идентификацию последствий принимаемых решений и оценку их эффективности. На этом этапе происходит выявление возможностей, угроз, рисков и уязвимостей, которые влияют на реализацию целей планирования и управления. Мониторинг обстановки на этом этапе позволяет вносить в сценарную модель коррекции в виде управленческих воздействий конкурирующих или кооперирующихся объектов управления.

8-й этап: на этом этапе происходит выработка замысла, которая направлена на формирование главного направления, определяющего способы достижения цели, раскрытие основной содержательной части и сути решения на осуществление мероприятия по обеспечению НБ [13]. В замысле определяются:

- задействованные силы и средства;
- направления главных и вспомогательных действий, точки сосредоточения основных усилий;

- основные возможности и уязвимости, связанные с достижением поставленных целей в условиях вероятного внешнего и внутреннего противодействия;
- способы реализации поставленных целей;
- основные привлекаемые ресурсы, силы и средства.

Конкретная суть каждого замысла организационного мероприятия зависит от его вида, намеченных целей, задач и состава имеющихся сил и средств. Замысел становится основой для разработки плана мероприятия по обеспечению НБ РФ и всей последующей работы по определению задач силам и средствам, а также отработки взаимодействия, обеспечения и управления в предстоящих действиях. Сценарий при этом используется для прогнозной оценки эффективности выполнения замысла мероприятия, а также для оценки текущей, а самое важное – прогнозируемой обстановки, в которой действует объект управления.

9-й этап: разработка плана мероприятия. Идентификация и использование в процессе планирования сценарно-технологического резерва повышает возможность достижения цели управления при полном или частичном поражении функциональности элементов многоуровневой организационной структуры, выполняющей отдельные этапы и мероприятия в сфере НБ за счет недостатка ресурсов, а также при реализации возможных угроз или по другим причинам. Одним из результатов этапа 9 являются целевые значения для подчиненных структур – нижних объектов иерархии (ОИ), а также управляющие воздействия, необходимые для их реализации с учетом влияния внешней среды. Участие подчиненных ОИ задается не только технологией реализации целевых установок, но и сценарными условиями их выполнимости. Таким образом, реализуются два критерия привлечения к выполнению мероприятия подчиненных ОИ. Во-первых, очевидное подчинение обусловлено участием ОИ в технологических цепочках достижения цели. Во-вторых, подчинение или включение ОИ в иерархию может быть обусловлено ситуационно, благодаря их влиянию на достижение целевых параметров (рис. 5).

Во втором случае также подключаются технологии, но уже технологии безопасности. Для каждого объекта иерархии строится многоуровневая модель рисков. Затем с помощью элемента

автоматизированной системы сценарного прогнозирования и анализа, который предназначен для раннего распознавания угроз и вызовов НБ в прогнозируемый период [20], выделяются прогнозируемые риски и угрозы. После этого определяются те объекты иерархии (ОУ), которые способны создать условия выполнения целевых параметров с учетом противодействия.

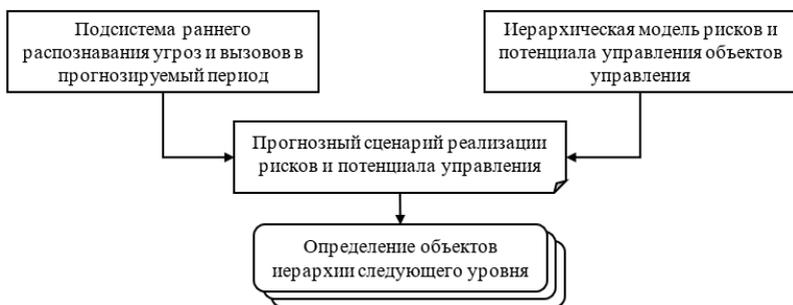


Рис. 5. Выбор подчиненных объектов иерархии по включению в ситуационный сценарий достижения целевых параметров

Разработка плана мероприятий для выбранного объекта управления предполагает одновременную разработку целевых установок для подчиненных объектов иерархической модели управления. Таким образом, реализуется принцип иерархического группового децентрализованного сценарного управления при реализации сложной системы мероприятий по управлению обеспечением НБ РФ (рис. 6):

- для получения значений целевых параметров объекты нижнего уровня (объекты-подчиненные) передают для включения в структуру сценарной модели верхнего уровня свои целевые факторы, которые определяются специализацией этих объектов иерархии (на рис. 6 зеленые стрелки);

- для каждого объекта в сложной иерархической структуре строится модель, которая служит основой для получения сценария достижения целевых параметров определенных на этапе 9 верхнего уровня (объекта-начальника) (на рис. 6 синие стрелки);

- результаты достижения цели для каждого подчиненного объекта управления (6-й этап) передаются наверх для контроля

выполнения плана мероприятия объекту иерархии верхнего уровня (этап 9) (на рис. 6 красные стрелки).

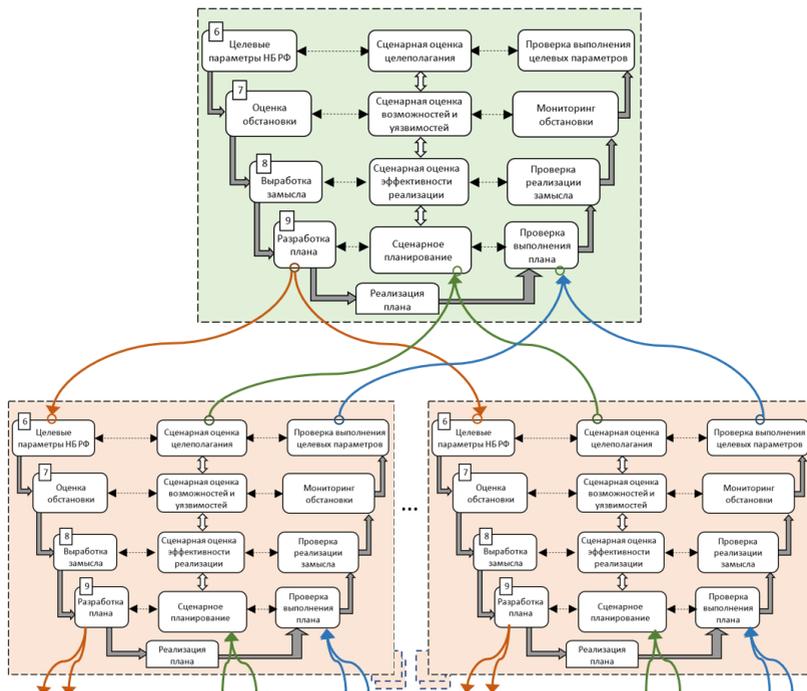


Рис. 6. Фрагмент иерархии W-образных моделей жизненного процесса планирования и управления безопасностью

Иерархия W-образных моделей жизненного цикла упреждающего сценарно-прогнозного мониторинга и анализа позволяет:

- ускорять реакцию всех уровней управления на возможные изменения, влияющие на достижение целей;
- переходить на всех уровнях иерархии к упреждающему управлению обеспечением НБ РФ;
- создавать виртуальные рабочие группы, объединенные по сходным целевым факторам, факторам управления или этапам прогнозных сценариев, переходя тем самым при необходимости к матричным подструктурам;

– включать в иерархическую модель негативно воздействующие на состояние НБ РФ субъекты, а также учитывать возможные сценарии их поведения, основанные на реализации интересов этих субъектов на всех этапах планирования.

Каждый объект на любом уровне иерархии имеет собственную структуру управления, которая, в свою очередь, строит собственный управляющий сценарий достижения целевых параметров, определенных для него верхним уровнем. Данный сценарий, таким образом, является одновременно и основой планирования для этого объекта, и координирующим сценарием для подчиненных объектов. Целевыми параметрами, полученными из координирующего сценария, задаются цели для подчиненных объектов и порядок (режим) их координации.

3.5. ИЕРАРХИЯ СЦЕНАРНОГО УПРАВЛЕНИЯ

Схема сценарного исследования в сочетании с разработанной W-образной моделью позволяет выстраивать иерархию децентрализованного группового управления сложными системами на основе координирующих сценариев (рис. 7).

Следует отметить, что если в иерархическую модель включены субъекты, негативно воздействующие на состояние НБ РФ, то в модель вышестоящего уровня иерархии включаются факторы субъектов противодействия, непосредственно влияющие на выполнение целей и задач управления по обеспечению НБ.

В разработанной модели жизненного цикла предусматриваются процедуры аттестации и верификации, которые осуществляются как на основе данных мониторинга, так и с использованием результатов применения сценарной методологии (внешних и внутренних данных, описывающих решаемую задачу планирования и содержащихся в паспорте проблемы $M^{(Пас)}(t)$).

Разработанная W-образная модель определяет результаты, которые должны быть получены на завершающем этапе управленческого цикла. При этом получаемые итоговые данные на каждой стадии или этапе должны подвергаться сценарному анализу и контролю.

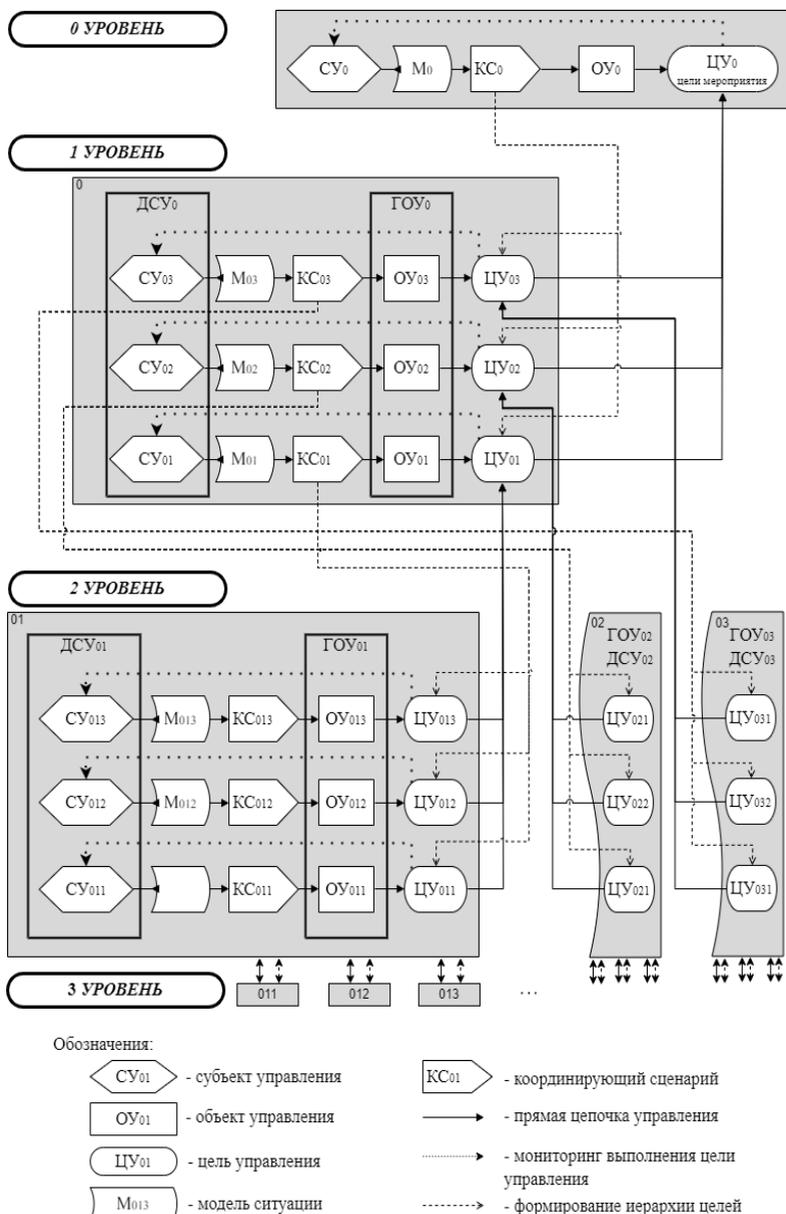


Рис. 7. Модель управления иерархическими системами НБ на основе координирующих сценариев

Применение рассматриваемой модели позволяет системе НБ РФ отслеживать и прогнозировать ход процесса управления обеспечением НБ $\mathfrak{R}(t)$, а также вносить упреждающие управленческие воздействия $U(t)$, пользуясь в основном событийной временной шкалой \mathfrak{RT} .

Фаза проверки хода и результатов выполнения мероприятий подтверждает правильность детализации плана. Фазы проверки исполнения отдельных этапов плановых мероприятий и сценарно-прогнозного анализа и мониторинга обстановки обеспечивают правильность ее оценки, а также замысла рассматриваемого мероприятия в сфере НБ. Фаза проверки выполнения поставленной цели мероприятия подтверждает правильность целевых установок, а также состава и содержания планируемых мероприятий по их реализации.

4. Заключение

Проведен анализ основных особенностей процессов целеполагания, планирования и управления обеспечением безопасности сложных систем в условиях неопределенности и объективных ограничений на противодействие угрозам.

Приведены результаты разработки понятийного аппарата сценарного исследования, на основе которого предложена общая схема формирования сценария, базирующаяся на трех основных компонентах исследуемой ситуации: 1) базовые формальные модели, описывающие совместное поведение объекта управления и внешней среды и основывающиеся на экспертном описании проблемной ситуации; 2) формальная модель неопределенности – квазиинформационная гипотеза; 3) формализованное описание предметной области на основе выделения экспертно-значимых декомпозиций, экспертно-значимых разбиений расширенного фазового пространства состояний исследуемой системы и экспертно-значимых событий, обеспечивающих многовариантность разрабатываемых сценариев. Полученные результаты являются основой для автоматизации сценарного исчисления, а также информационной поддержки процессов подготовки и принятия управленческих решений.

Для решения задач обеспечения НБ на различных уровнях проведен анализ возможности использования процессного подхода на базе моделей жизненного цикла. Сделан вывод о целесообразности использования в качестве основы для решения рассматриваемых задач V-образной модели с учетом ее основного недостатка – верификации запланированных результатов только по факту их достижения.

С целью исключения упомянутых выше ограничений и повышения адекватности отражения реальных особенностей рассматриваемых процессов разработана модифицированная W-образная модель жизненного цикла, обеспечивающая возможность наряду с традиционными методами анализа проводить упреждающую прогнозную оценку ожидаемых результатов в условиях неопределенности, что позволяет перейти от реактивного к проактивному управлению мероприятиями по обеспечению НБ.

Разработанная модификация заключается в дополнении структуры традиционной V-образной модели связанной со сценарным планированием промежуточной ветвью, отражающей предоставляемые использованием сценарного подхода возможности опережающей оценки эффективности планово-управленческих решений с учетом альтернатив возможного будущего развития событий на объектах управления и во внешней среде в условиях риска и неопределенности.

Разработана структура, состав основных элементов и стадий решения задач управления обеспечением НБ в рамках W-образной модели жизненного цикла, а также методология использования иерархии рассматриваемых моделей для решения сложно-подчиненных проблем планирования и управления в рассматриваемой предметной области.

Обоснована необходимость встраивания сценарной технологии в функциональную структуру системы подготовки стратегических и тактических плановых решений в рассматриваемой предметной области, что должно повысить эффективность процессов планирования и обоснованность планово-управленческих решений на основе использования результатов анализа альтернатив развития проблемных ситуаций, а также упреждающей оценки результативности и возможных негативных последствий принятия и реализации разрабатываемых решений.

Литература

1. БЕРЕГОВСКАЯ Е.О., ШАЛИНА А.И., КРАСНОВА А.С. *Военно-экономический анализ стадий жизненного цикла высокотехнологичной продукции* // Экономика высокотехнологичных производств. – 2022. – Т. 3, № 1. – С. 27–38.
2. БЕРГ Д.Б., УЛЬЯНОВА Е.А., ДОБРЯК П.В. *Модели жизненного цикла: учеб. пособие.* – Екатеринбург: Изд-во Урал. ун-та, 2014. – 74 с.
3. ГОЛУБЕВ С.С., КУКУШКИНА Г.Р. *Проблемы развития системы управления полным жизненным циклом вооружения, военной и специальной техники* // Экономика высокотехнологичных производств. – 2020. – №4. – С. 183–196.
4. ДРАНКО О.И., НОВИКОВ Д.А., РАЙКОВ А.Н. и др. *Управление развитием региона. Моделирование возможностей.* – М.: URSS, ООО «ЛЕНАНД», 2023. – 432 с.
5. КЛОЧКОВ В.В. *Взаимосвязь военных и социально-экономических угроз национальной безопасности России* // Экономическая безопасность. – 2021. – №1. – С. 43–56.
6. ЛАРЮХИН В.Б., ОВЧИННИКОВ С.А., СКОБЕЛЕВ П.О. и др. *Управление процессами и ресурсами в системе полного жизненного цикла вооружения и военной техники на основе цифровой экосистемы адаптивного менеджмента* // Вопросы инновационной экономики. – 2020. – №3. – С. 1259–1274.
7. ЛИНДГРЕН М., БАНДХОЛЬД Х. *Сценарное планирование: связь между будущим и стратегией* // Корпоративный менеджмент. – Москва. – URL: <https://www.cfin.ru/management/strategy/plan/scenario.shtml> (дата обращения: 10.03.24).
8. МАСЛИХИНА В.Ю. *Сценарное планирование развития пространственных социально-экономических систем: методические подходы* // Региональная экономика: теория и практика. – 2020. – Т. 18, №10(481). – С. 1839–1868.
9. *Модели и методы анализа и синтеза сценариев развития социально-экономических систем: в 2-х кн.* / Под ред. В.Л. Шульца и В.В. Кульбы. – М.: Наука, 2012. – Кн. 1: 304 с.; кн. 2: 358 с.

10. ПОДБЕРЕЗКИН А.И. *Методика планирования стратегического сдерживания*. – URL: <http://eurasian-defence.ru/?q=node/44776> (дата обращения: 21.08.2020).
11. РИНГЛАНД Д. *Сценарное планирование для разработки бизнес-стратегии*. – М.: Диалектика, 2008. – 559 с.
12. *Стратегическое прогнозирование международных отношений: кол. монография / Под ред. А.И. Подберезкина, М.В. Александрова*. – МГИМО МИД России, Центр военно-политических исследований. – М.: МГИМО-Университет, 2016. – 743 с.
13. *Указ Президента Российской Федерации от 02.07.2021 г. №400 «О Стратегии национальной безопасности Российской Федерации»*. – URL: <http://www.kremlin.ru/acts/bank/4704644776> (дата обращения: 20.05.2024).
14. ХИЖНЯК Е.А. *Сценарное планирование как эффективный инструмент стратегического управления // Российское предпринимательство*. – 2007. – Т. 8, №3. – С. 17–20.
15. ЧЕРНОВ И.В. *Сценарно-когнитивное моделирование сложных систем на основе событийной идентификации динамики факторов // Проблемы управления*. – 2023. – №3. – С. 65–76.
16. ЧЕРНОВ И.В. *Сценарные методы повышения эффективности реализации жизненного цикла программно-целевого управления (анализ концепции) // Проблемы управления*. – 2021. – №5. – С. 88–93.
17. ЧЕРНОВ И.В. *Сценарный анализ уязвимости при управлении сложными системами // Автоматика и телемеханика*. – 2022. – №5. – С. 133–147.
18. ШЕВЧЕНКО И.В., ЛУНЕВ В.Д. *Современные подходы к сценарному планированию в России // Ученые записки Крымского инженерно-педагогического университета*. – 2022. – №4(78). – С. 192–198.
19. ШУЛЬЦ В.Л., БОЧКАРЕВ С.А., КУЛЬБА В.В. и др. *Сценарное исследование проблем обеспечения общественной безопасности в условиях цифровизации*. – М.: Проспект, 2020. – 240 с.

20. ШУЛЬЦ В.Л., ЧЕРНОВ И.В., КУЛЬБА В.В. и др. *Сценарное планирование в управлении обеспечением НБ: методологические основы* // Национальная безопасность / Nota bene. – 2023. – №5. – С. 36–61.

ANALYSIS OF THE METHODOLOGY OF SCENARIO PLANNING AND NATIONAL SECURITY MANAGEMENT

Igor Chernov, V.A. Trapeznikov Institute of Control Sciences of RAS, Moscow, Candidate of Technical Sciences, Leading Researcher (chernov@ipu.ru).

Abstract: The article is devoted to the development of a methodology for scenario planning and management of ensuring the national security of the Russian Federation. The results of developing a conceptual framework for scenario research are presented, on the basis of which a general scheme for the procedure for generating scenarios has been developed. The expert description of the development model of the processes under consideration is based on the concept of an expert-significant partition of the extended phase space, as well as expert-significant events that occur in the sequence specified in the ordering system within the framework of the basic model. As a result of the study of this model, scenarios for the behavior of the control object or the development of the security situation are formed. A multi-level model of the life cycle of planning and managing security measures has been developed, providing the possibility of proactive verification of the results of the implementation of key stages of the management cycle, which allows us to move from passive accounting of accomplished events to proactive management. The results obtained can serve as the basis for automating scenario research, as well as analytical predictive support for the processes of preparing and making management decisions in the field of ensuring the security of complex organizational systems.

Keywords: security, scenario, life cycle, management, modeling.

УДК 005.7

ББК 13.42

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

Поступила в редакцию 17.06.2024.

Опубликована 30.11.2024.

ПОДХОД К РАСПРЕДЕЛЕНИЮ ЗАДАЧ ТЕХНОЛОГИЧЕСКОГО ПРОЦЕССА СРЕДИ НЕОДНОРОДНЫХ УЧАСТНИКОВ КРТС С УЧЕТОМ ИХ СОСТОЯНИЯ¹

Галин Р. Р.², Мамченко М. В.³, Галина С. Б.⁴,
Зорин В. А.⁵

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

Описываются результаты серии экспериментов для задач формирования оптимального состава исполнителей и минимизации времени выполнения задач технологического процесса (по три случая для каждой задачи, отражающие три реализованных механизма решения). Представлены показатели эффективности распределения задач между участниками и оценки эффективности экспериментов других авторов, посвященных распределению задач в много-агентных и коллаборативных робототехнических системах (КРТС). Представлен качественный сравнительный анализ полученных результатов экспериментов, проведена оценка условий и результатов экспериментов на основе сформированных показателей оценки эффективности взаимодействия людей и коботов и эффективности распределения задач между участниками. Анализ показывает сопоставимые результаты по минимизации времени выполнения задач с использованием авторских решений по сравнению с рассмотренными работами. Следует отметить, что авторская постановка задач распределения работ в КРТС по сравнению со другими работами обладает комплексностью. В частности, во всех моделируемых задачах (подбор оптимального состава исполнителей, все варианты минимизации времени выполнения задач технологического процесса и затрат) обязательно проверяется требование коллаборации людей и коботов (совместное выполнение любой задачи хотя бы одним коботом и хотя бы одним человеком), при этом отслеживается степень задействования каждого участника. Кроме того, в случае подбора оптимального количества людей и коботов из состава КРТС осуществляется проверка выполнимости технологического процесса расчетным составом участников.

¹ Исследование выполнено за счет гранта Российского научного фонда №23-29-00690, <https://rscf.ru/project/23-29-00690>.

² Ринат Романович Галин, к.т.н., с.н.с. (grr@ipu.ru).

³ Марк Владиславович Мамченко, н.с. (markmamcha@gmail.com).

⁴ Саня Болаткызы Галина, н.с. (kameshevasaniya@gmail.com).

⁵ Василий Александрович Зорин, соискатель, ИПУ РАН (v.a.zorin@mail.ru).

Ключевые слова: коллаборативная робототехническая система, взаимодействие человека и робота, распределение задач, технологический процесс, оптимизация.

1. Введение

В настоящее время на производстве для совместного выполнения технологических операций (задач) с человеком активно применяются коллаборативные роботы (коботы). Подобная совместная деятельность людей и коллаборативных роботов называется коллаборацией, а организационно-техническую систему, состоящую из людей и коботов, объединенных общей задачей (целью), выполняемой в едином рабочем пространстве, принято обозначать коллаборативной робототехнической системой (КРТС) [13].

Одной из ключевых задач повышения эффективности функционирования КРТС является создание системы принятия решений, обеспечивающей способность динамически распределять задачи среди участников. Настоящая работа является заключительной статьей цикла работ (ряд результатов представлен в [1, 2, 11–14, 24]), посвященных распределению технологических задач в смешанных неоднородных командах КРТС.

В представленной статье предлагаются показатели оценки эффективности взаимодействия человека и кобота и распределения работ среди участников КРТС, а также сравнение полученных результатов численных экспериментов в моделирующем комплексе с аналогичными данными, полученными другими авторами (в научных публикациях).

2. Предложенные подход, методика и моделирующий комплекс

В предыдущих работах предложены и описаны сформулированы функциональные требования, предъявляемые к разрабатываемому решению (методика и подход) и соответствующему математическому и алгоритмическому обеспечению, разработаны подход и реализующий его алгоритм распределения задачи

в КРТС, а также методика назначения исполнителей на задачи технологического процесса [1, 2, 11–14, 24].

В рамках цикла работ предполагается решение трех основных задач:

1. Минимизация количества задействуемых участников КРТС: формирование оптимального состава исполнителей для задаваемого технологического процесса (с учетом или без учета ограничений на превышение изначального времени технологического процесса).

2. Минимизация времени на выполнение всех задач технологического процесса. Представляет собой решение оптимизационной задачи (отдельно для каждой задачи технологического процесса). В качестве минимизируемой функции (значения) может выступать как общее время выполнения технологического процесса, так и набор значений времени технологического процесса и каждой задачи в его составе.

3. Минимизация затрат для выполнения задач технологического процесса. Предполагается минимизация энергетических или финансовых затрат, требуемых для выполнения технологического процесса (с учетом задаваемых значений эффективности участников КРТС).

Требования к разрабатываемому моделирующему комплексу, в том числе стек разработки, необходимый перечень функциональных возможностей, а также требования к его графическому интерфейсу представлены в авторской работе [24]. На основании выявленных требований авторами был разработан моделирующий комплекс для апробации метода и реализующего его алгоритма распределения задач среди участников в КРТС, который обладает функциональными возможностями, представленными в работе [24], основными из которых являются: задание исходных данных и проведение моделирования для всех основных сценариев (минимизация времени, состава исполнителей и затрат), задание и учет пространственной модели рабочих мест, возможность переназначения и доназначения исполнителей в ходе выполнения работ, учет использования материалов, деталей и других объектов, учет времени перемещения людей между рабочими местами, возможность внесения изме-

нений в параметры коботов, людей и задач непосредственно в ходе выполнения технологического процесса и др.

2.1. ЗАДАЧА 1. МИНИМИЗАЦИЯ КОЛИЧЕСТВА ЗАДЕЙСТВУЕМЫХ УЧАСТНИКОВ КРТС

Исходные данные. Пусть дана смешанная неоднородная команда, состоящая из двух множеств: n людей $H = \{h_i\}_{i=1}^n$ и m коботов $B = \{b_j\}_{j=1}^m$. В качестве смешанной неоднородной команды может рассматриваться как сама КРТС в целом, так и одна или несколько ее подмножеств (команд КРТС), отвечающих требованиям коллаборативности (наличие не менее одного человека и не менее одного кобота). В предельном случае – один человек и один кобот – подобная смешанная команда называется коллаборативной робототехнической ячейкой (или коллаборативной ячейкой).

У каждого участника КРТС задано значение эффективности, влияющее на время выполнения задачи. Это значение может быть одинаковым для всех задач или задаваться отдельно для каждой p задачи. Пусть эффективность k человека и s кобота в команде задаются как r_{h_k} и r_{b_s} соответственно, при этом данные значения будут одинаковы для всех задач и задаваться в диапазоне дискретных значений $[0, \dots, 1]$. Если участник команды имеет нулевую эффективность, он не может выполнять задачу (задачи); в случае единичной эффективности полагается, что эффективность такого участника максимальная для задачи (задач).

Пусть также дан технологический процесс, в котором в каждый дискретный момент времени $t = 0, \dots, t_{TP}$ начинается $p_t \in \mathbb{N}$, $p_t \geq 1$ новых задач (одновременно завершаются задачи, время завершения которых вычислено как t). Любая p задача имеет две временные характеристики:

– $t_{op.st_p}$ – время выполнения p задачи минимальным составом (коллаборативной ячейкой) с максимальной эффективностью человека и кобота; данное значение полагается изначально заданным (известным);

– $t_{op,real_p}$ – время выполнения p задачи вычисленным (расчетным) количеством из n_p людей и m_p роботов; расчет данного значения показан в формулах (5) и (8).

Кроме того, в момент времени t при принятии решении о назначении состава исполнителя для каждой p задачи из множества задач p_t учитываются также следующие величины:

– минимальное количество людей $n_{p_{min}}$ и роботов $m_{p_{min}}$, требуемых для выполнения задачи;

– доступное к моменту времени t количество людей $n_{p_{av}}$ и роботов $m_{p_{av}}$, которые не заняты выполнением каких-либо других задач в этот момент времени;

– максимальное (предельное) количество людей $n_{p_{max}}$ и роботов $m_{p_{max}}$, которых возможно назначить на выполнение p задачи.

$$\text{В общем случае } \begin{cases} n_{p_{min}} \leq n_p \leq \min(n_{p_{max}}, n_{p_{av}}), \\ m_{p_{min}} \leq m_p \leq \min(m_{p_{max}}, m_{p_{av}}). \end{cases}$$

Постановка задачи. Требуется найти минимальный состав исполнителей для всех задач технологического процесса. Решение задачи минимизации числа задействуемых участников КРТС рассматривается для следующих случаев [4–5, 7]:

Случай 1. Не учитывается время выполнения задачи. Для всех дискретных значений времени $t = 0, \dots, t_{TP}$ найти количество людей и роботов по следующей формуле:

$$(1) \quad \max \min \left(\sum_{p=1}^{p_t} n_p \right), \quad \max \min \left(\sum_{p=1}^{p_t} m_p \right),$$

где t_{TP} – время окончания технологического процесса, p_t – количество задач, начавшихся в момент времени t . При этом соблюдается ограничение

$$(2) \quad \begin{cases} n_{p_{min}} \leq n_p, \\ m_{p_{min}} \leq m_p. \end{cases}$$

Случай 2. Учитывается время выполнения задачи [9, 12, 14]. Для каждой p задачи найти такие значения n_p и m_p , удовлетворяющие следующему условию:

$$(3) \quad \begin{cases} \min(n_p + m_p + n_p \times m_p), \\ \min \left(t_{op.real_p} = \frac{t_{op.st_p} (n_{p_{min}} + m_{p_{min}})}{\sum_{k=1}^{n_p} r_{h_k} + \sum_{s=1}^{m_p} r_{b_s}} \right), \end{cases}$$

учитывая ограничения

$$(4) \quad \begin{cases} n_{p_{min}} \leq n_p, \\ m_{p_{min}} \leq m_p, \\ \max \min \left(\sum_{p=1}^{p_t} n_p \right), \max \min \left(\sum_{p=1}^{p_t} m_p \right). \end{cases}$$

2.2. ЗАДАЧА 2. МИНИМИЗАЦИЯ ВРЕМЕНИ НА ВЫПОЛНЕНИЕ ВСЕХ ЗАДАЧ ТЕХНОЛОГИЧЕСКОГО ПРОЦЕССА

Исходные данные соответствуют задаче 1.

Постановка задачи. Требуется найти минимальное время выполнения всех задач технологического процесса. Рассмотрим три случая (варианта) назначения исполнителей на работы.

Случай 1. Для каждой p задачи найти такие значения n_p и m_p , что выполняется условие [15, 17]

$$(5) \quad \min \left(t_{op.real_p} = \frac{t_{op.st_p} (n_{p_{min}} + m_{p_{min}})}{\sum_{k=1}^{n_p} r_{h_k} + \sum_{s=1}^{m_p} r_{b_s}} \right),$$

учитывая ограничения

$$(6) \quad \begin{cases} n_{p_{min}} \leq n_p \leq \min(n_{p_{av}}, n_{p_{max}}), \\ m_{p_{min}} \leq m_p \leq \min(m_{p_{av}}, m_{p_{max}}), \end{cases}$$

при этом значения n_p и m_p выбираются случайным образом с учетом ограничений (6).

Случай 2. Для каждой p задачи найти такие значения n_p и m_p , что выполняется условие

$$(7) \quad \begin{cases} \min \left(\left| \sum_{i=1}^{n_p} r_{h_i} - \sum_{i=n_p+1}^{\min(n_{p_{av}}, n_{p_{max}})} r_{h_i} \right| \right), \\ \min \left(\left| \sum_{j=1}^{m_p} r_{b_j} - \sum_{j=m_p+1}^{\min(m_{p_{av}}, m_{p_{max}})} r_{b_j} \right| \right), \end{cases}$$

учитывая ограничения (6).

Случай 3. Для каждой p задачи найти такие значения n_p и m_p , что выполняется условие [22]

$$(8) \quad \min \left(t_{op.real_p} = \frac{t_{op.st_p} (n_{p_{min}} + m_{p_{min}})}{\sum_{k=1}^{n_p} r_{h_k} + \sum_{s=1}^{m_p} r_{b_s}} \right),$$

учитывая ограничения

$$(9) \quad \begin{cases} n_p \leq n_{p_{av}}, \\ n_p \geq n_{p_{min}}, \\ n_p \leq n_{p_{max}}, \\ m_p \leq m_{p_{av}}, \\ m_p \geq m_{p_{min}}, \\ m_p \leq m_{p_{max}}, \\ t_{op.real_p} \geq 0. \end{cases}$$

2.3. ЗАДАЧА 3. МИНИМИЗАЦИЯ ЗАТРАТ ДЛЯ ВЫПОЛНЕНИЯ ЗАДАЧ ТЕХНОЛОГИЧЕСКОГО ПРОЦЕССА

Заключительной рассматриваемой задачей являлась минимизация энергетических или финансовых затрат, требуемых для выполнения задач технологического процесса. В работе [2] представлены постановка задачи (в трех случаях) для вычисления оптимальных значений объемов работ для участников сме-

шанных неоднородных команд (с разными видами функций затрат), а также результаты численных экспериментов для пяти сценариев.

Следует отметить, что при проведении имитационного моделирования в статье [2] не предполагалась возможность выполнения человеком большего функционала (объема работ), чем у робота. Это обусловлено тем, что функция производственных затрат для роботов задавалась в линейном виде, а для людей – в виде квадратичной производственной функции. Таким образом, назначение большего объема работ роботу является предпочтительным, так как эффективность его работы не снижается со временем. Несмотря на то, что случай назначения большего объема работ человеку является неоптимальным, имитационное моделирование данного случая и случая необходимости изменения объема работ человека в ходе выполнения задач может быть рассмотрено в последующих работах.

3. Результаты экспериментов

В проведенной серии экспериментов рассматривается смешанная неоднородная команда КРТС, членами которой являются множество людей $H = \{h_i\}_{i=1}^n$ (n – количество людей в команде) и множество роботов $B = \{b_j\}_{j=1}^m$ (m – количество роботов). Задание и описание переменных, используемых в моделирующем комплексе, изложено в работе [11].

В среде MATLAB было проведено моделирование субоптимального распределения задач между участниками КРТС. Изначальный технологический процесс представлен в виде взвешенного ориентированного графа (см. рис. 1), вершины которого обозначают моменты окончания предыдущих и начало новых задач, а ребра – длительность соответствующих задач.

Оптимизационная задача (8) с учетом ограничений (9) решается с помощью методов оптимизации (например, генетического алгоритма) отдельно для каждой задачи [25–31].

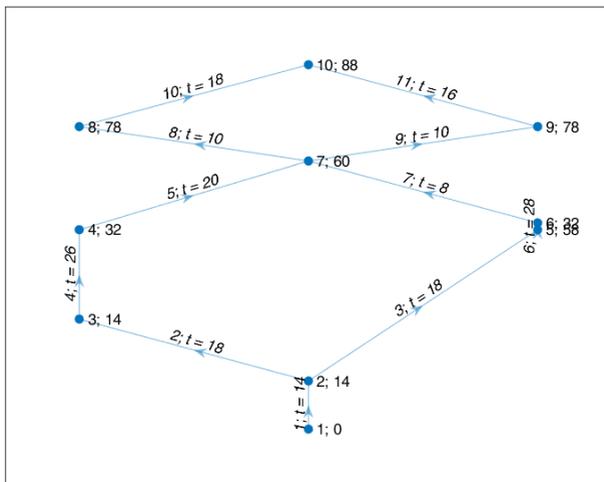


Рис. 1. Технологический процесс в виде ориентированного графа с исходными параметрами

Результаты численного моделирования случаев для Задач 1 и 2 представлены в таблице 1. Используемые в таблице обозначения для КРТС и ее участников и задач технологического процесса соответствуют обозначениям, представленным в исходных данных для задачи 1.

Таблица 1. Результаты численного моделирования случаев для Задач 1 и 2

| Значения | Задача 1. Случай 1 | | | Задача 1. Случай 2 | | | Задача 2. Случай 1 | | | Задача 2. Случай 2 | | | Задача 2. Случай 3 | | |
|---|-----------------------|------|------|-----------------------|------|------|-----------------------|------|------|-----------------------|------|------|-----------------------|------|------|
| | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 |
| <p><u>Исходные данные:</u> количество людей $n = 25$, количество роботов $m = 20$, минимальный состав на выполнение задачи $n_{p_{min}} = m_{p_{min}} = 1$, максимальный $n_{p_{max}} = m_{p_{max}} = 10$ для каждой задачи.</p> <p><u>Обозначения:</u> r_h – диапазон значений (значения) эффективности для людей, r_b – диапазон значений (значения) эффективности для роботов, t_{before} – время выполнения технологического процесса до оптимизации (в секундах), t_{after} – время выполнения технологического процесса после оптимизации (в секундах), $t_{elapsed}$ – общее время работы вычисления в моделирующем комплексе (в секундах).</p> | | | | | | | | | | | | | | | |
| r_h | 0–1 | 0,3 | 0,1 | 0–1 | 0,3 | 0,1 | 0–1 | 0,3 | 0,1 | 0–1 | 0,3 | 0,1 | 0–1 | 0,3 | 0,1 |
| r_b | 0–1 | 0,8 | 0,8 | 0–1 | 0,8 | 0,8 | 0–1 | 0,8 | 0,8 | 0–1 | 0,8 | 0,8 | 0–1 | 0,8 | 0,8 |
| t_{before} | 106 | 106 | 106 | 106 | 106 | 106 | 106 | 106 | 106 | 106 | 106 | 106 | 106 | 106 | 106 |
| t_{after} | 283 | 192 | 235 | 115 | 192 | 235 | 67 | 39 | 63 | 38 | 42 | 53 | 79 | 69 | 83 |
| $t_{elapsed}$ | 1,82 | 1,27 | 1,44 | 47,2 | 39,4 | 39,9 | 0,25 | 0,26 | 0,23 | 0,27 | 0,28 | 0,27 | 39,7 | 40,5 | 42,8 |

4. Оценка эффективности взаимодействия человека и робота и распределения задач

Анализируя постановку и решение задач 1–3, возможно сформировать следующие показатели оценки эффективности функционирования КРТС, учитывая взаимодействие людей и роботов в едином рабочем пространстве и эффективность распределения задач между участниками [1, 12]:

1. При минимизации времени выполнения задач:

- время выполнения отдельных задач и всего технологического процесса расчетным составом людей и роботов;
- разность исходного и расчетного (полученного) времени выполнения отдельных задач и всего технологического процесса (сэкономленное времени);
- выигрыш по времени (снижение времени выполнения задачи/технологического процесса в процентном соотношении).

2. При минимизации состава исполнителей:

- расчетный состав исполнителей;
- выигрыш по составу (снижение количества людей и роботов для выполнения задачи/технологического процесса, в процентном соотношении);
- наличие коллаборации людей и роботов (обязательно для КРТС как для отдельной задачи, так и для всего технологического процесса).
- выполнимость технологического процесса расчетным составом исполнителей (количеством людей и роботов), учитывая количество рабочих мест, задач технологического процесса и его модель в виде ориентированного взвешенного графа;
- степень задействования каждого участника КРТС – отношение количества операций, выполненных участником, к их общему количеству (не может быть нулевой, в идеальном случае примерно одинакова отдельно для людей и роботов).

3. При минимизации затрат:

- финансовые затраты на выполнение задачи/всего технологического процесса;

- энергетические затраты на выполнение задачи/всего технологического процесса;
- другие задаваемые затраты;
- выигрыш по затратам (снижение затрат, требуемых для выполнения задачи/технологического процесса, в процентном соотношении, отдельно для каждого типа затрат).

Кроме того, в предложенных подходе и методике подбор состава исполнителей на выполнение задачи осуществляется на основе одного из следующих критериев [1, 12]:

- доступность участников КРТС (количество незанятых людей и коботов к моменту начала задачи), с учетом минимального и максимального возможного состава исполнителей для конкретной задачи;
- эффективность участников КРТС, доступных к моменту начала конкретной задачи (с учетом минимального и максимального возможного состава исполнителей).

В разработанных подходе и методике предполагается, что по умолчанию для каждой задачи существует ее стандартное время выполнения задаваемым минимально возможным составом исполнителей (не менее одного человека и не менее одного кобота), при этом каждый исполнитель имеет максимальную эффективность. Способ расчета стандартного времени выполнения задачи отражает ее оценочное или априорно заданное время выполнения и может быть задан произвольным образом.

5. Сравнительный анализ полученных результатов моделирования

Для сравнения полученных результатов моделирования с аналогичными данными, полученными другими авторами, приведем краткий обзор некоторых численных результатов моделирования распределения задач в коллаборативных и многоагентных робототехнических системах (РТС).

В работе [21] предложен алгоритм планирования и распределения работ в КРТС с учетом ограничений ресурсов с минимизацией времени выполнения за счет минимизации функции затрат участников (решение оптимизационной задачи). Учиты-

вается время на передачу деталей (объектов воздействия) между человеком и коботом. Численный эксперимент проведен на базе коллаборативной ячейки (один кобот и один человек).

Эксперимент 1. Начальные условия. Задача: поиск оптимальной последовательности действий участников команды при разборке деталей (моделирование), 14 задач. Начальное общее время выполнения задач: 71 с. Результат: общее время выполнения задач после применения предложенного алгоритма – 51 с, 120 итераций. Время выполнения вычислений составило 383 с.

Эксперимент 2. Начальные условия соответствуют эксперименту 1. Начальное общее время выполнения задач: 69 с. Результат: общее время выполнения задач после применения предложенного алгоритма: менее 49 с, 127 итераций. Время выполнения вычислений составило 275 с.

Эксперимент 3. Реальная производственная задача по разборке жесткого диска. Начальное общее время выполнения задач: 203 с, после применения предложенного алгоритма – 151 с, 189 итераций. Время выполнения вычислений составило 1349 с.

В работе [23] предложен алгоритм распределения задач участникам КРТС на основе смешанно-целочисленного линейного программирования для минимизации функции затрат (оптимизационная задача).

Эксперимент проведен для технологического процесса по сборке изделия (14 сборочных задач и 4 задачи контроля выполнения работы коботом(ами) со стороны человека). Исходные данные: 1 человек и 2 кобота. Начальное общее время выполнения задач: 668,37 с, после применения предложенного алгоритма – 497,97 с.

В работе [19] предложен авторский метод динамического разбиения задач на последовательности подзадач и их назначения исполнителям многоагентной системы (в том числе в КРТС) с учетом приоритизации задач. Четыре эксперимента проведены на базе коллаборативной ячейки.

Эксперимент 1. Перепланировка распределения задач при поступлении новой задачи в процессе выполнения технологического процесса. Начальное общее время выполнения задач: 140 с, после применения предложенного метода – 135 с.

Эксперимент 2. Перераспределение задач при увеличении срока выполняемой задачи. Начальное общее время выполнения задач: 205 с, после применения предложенного метода – 190 с.

Эксперимент 3. Перераспределение задач при принудительном изменении порядка их выполнения. Начальное общее время выполнения задач: 125 с, после применения предложенного метода – 140 с.

Эксперимент 4. Параллельное выполнение нескольких задач. Начальное общее время выполнения задач: 140 с, после применения предложенного метода – 135 с.

В работе [6] предложен алгоритм распределения задач в многоагентных РТС на основе дискретно-событийной модели и минимизации функций затрат участников. Эксперимент проведен на базе малой КРТС (один человек и два кобота) для технологического процесса из 18 задач. Начальное общее время выполнения задач: 135 с, после применения предложенного алгоритма – 114 с. В постановке задачи отсутствует необходимость обязательной коллаборации человека и робота (задачи могут выполняться исполнителями независимо друг от друга).

В работе [8] предложена метод принятия решений при распределении задач в КРТС на основе модифицированной архитектуры FlexHRC (flexible human-robot-collaboration) и минимизации функции затрат участников для формирования последовательности задач. Эксперимент проведен на базе коллаборативной ячейки, технологический процесс включал в себя 32 задачи. Общее время выполнения задачи (с применением метода) составило 348,08 с, время выполнения вычислений – 6,29 с.

В работе [10] предложен метод распределения задач в РТС на основе деревьев поведения с временными и логическими ограничениями. Эксперимент проведен на базе малой РТС (четыре агента) для выполнения технологического процесса, состоящего из 12 задач. Начальное общее время выполнения задач: около 125 с, после применения предложенного метода – менее 120 с.

В статье [16] описан подход к формированию расписания работ в многоагентных РТС на основе методов смешанного целочисленного линейного программирования и программирова-

ния в ограничениях. Проведена серия численных экспериментов, в том числе на базе больших РТС. В частности, время выполнения вычислений при построении оптимального расписания для 100 агентов и технологического процесса, состоящего из 1080 задач, составило около 183 с.

В [18] предложен метод распределения задач в КРТС с использованием многоагентного подхода и алгоритм поиска в графе A^* . Вычислительный эксперимент проведен на базе коллаборативной ячейки для простейшего технологического процесса, состоящего из 6 задач. Общее время выполнения задач с применением предложенного метода составило 3,5 с, вычисления проведены за 40 мс.

В работе [20] предложен подход ConcHRC для распределения задач между агентами РТС с возможностью их распараллеливания, в том числе в сценариях коллаборативного взаимодействия. Проведено два эксперимента на базе коллаборативной ячейки для технологического процесса, состоящего из 12 задач (для двух разных роботов).

Эксперимент 1. Общее время выполнения задач с применением предложенного подхода составило 242 с, время вычислений – 4,23 с.

Эксперимент 2. Общее время выполнения задач с применением предложенного подхода составило 307 с, время вычислений – 3,19 с.

Наконец, в статье [3] предложен алгоритм распределения задач в децентрализованных роевых робототехнических системах в условиях неполноты информации (на основе методов системного анализа и аналитической геометрии). В работе представлена серия из 100 вычислительных экспериментов по распределению 100 задач, проведенных для РТС из 50 агентов. Среднее начальное общее время выполнения задач: 80,1 с, после применения алгоритма – 51,9 с. Следует отметить, что в постановке задачи коллаборация человека и робота присутствует только на этапе начального распределения задач, в дальнейших этапах – только в процессе управления группировкой роботов и контроля хода выполнения задач.

5.1. СРАВНИТЕЛЬНЫЙ АНАЛИЗ

В таблице 2 представлен сравнительный анализ результатов вычислительных экспериментов, полученных в рамках цикла работ [1, 2, 11–14, 24] по сравнению с аналогичными данными, полученные другими авторами [3, 6, 8, 10, 16, 18–21, 23]. Следует отметить, что при наличии больших серий экспериментов средние представлялись значения.

Из таблицы 2 видно, что авторская постановка задач распределения работ в КРТС по сравнению со сравниваемыми работами обладает комплексностью. В частности, во всех моделируемых задачах (подбор оптимального состава исполнителей, все варианты минимизации времени выполнения задач, технологического процесса и затрат) обязательно проверяется требование коллаборации людей и коботов (совместное выполнение любой задачи хотя бы одним коботом и человеком), ведется учет степени задействования каждого участника. Кроме того, осуществляется верификация выполнимости технологического процесса расчетным составом участников при подборе оптимального количества людей и коботов из состава КРТС.

Следует отметить, что минимизация состава исполнителей, финансовых затрат и времени выполнения каждой задачи возможна в любых конфигурациях задаваемого технологического процесса, предельной численности людей и коботов в КРТС и их параметров.

Таблица 2. Сравнительный анализ результатов вычислительных экспериментов

| Показатель, величина измерения (при наличии) | Значения | | | | | | | | | | |
|---|----------------|--------|--------------------|-----|--------|------|------------|------|------------|------------|-------------------------------------|
| | [21] | [23] | [19] | [6] | [8] | [10] | [16] | [18] | [20] | [3] | Авторские решения [1, 2, 11–14, 24] |
| Результаты вычислительного эксперимента (серии экспериментов) | | | | | | | | | | | |
| Количество экспериментов | 3 | 1 | 4 | 1 | 1 | 1 | 80 | 1 | 2 | 100 | 15; 5 |
| Количество людей в эксперименте | 1 | 1 | 1 | 1 | 1 | 4 | max 100 | 1 | 1 | 50 | 25; 2–4 |
| Количество коботов (роботов) в эксперименте | 1 | 2 | 1 | 2 | 1 | | | 1 | 1 | | 20; 2–4 |
| Количество задач (в том числе в рамках технологического процесса) | 14 | 18 | 4; 5; 5; 6 | 18 | 32 | 12 | 1080 | 6 | 12 | 100 | 11 |
| Общее время выполнения задач (без применения предложенных решений), с | 71; 69; 203 | 668,37 | 140; 205; 125; 140 | 135 | – | ~125 | – | – | – | 80,1 (ср.) | 106 |
| Общее время выполнения задач (с применением предложенных решений), с | 51; 49; 151 | 497,97 | 135; 190; 140; 135 | 114 | 348,08 | <120 | ~183 | 3,5 | 242; 307 | 51,9 (ср.) | 89,5 (ср.) |
| Время вычислений (при наличии), с | 383; 275; 1349 | – | – | – | 6,29 | – | – | 0,04 | 4,23; 3,19 | – | 17,04 (ср.) |

Таблица 2 (продолжение)

| Показатели оценки эффективности взаимодействия людей и роботов | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Учет выполнимости технологического процесса оптимальным (расчетным) составом участников (0 – не учитывается, 1 – учитывается) | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 |
| Учет степени задействования участников (0 – не учитывается, 1 – учитывается) | 1 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 |
| Степень коллаборации (0 – отсутствует, 1 – возможна для отдельных задач, 2 – обязательна для всего технологического процесса) | 1 | 1 | 2 | 1 | 1 | 1 | 2 | 1 | 1 | 1 | 2 |
| Показатели эффективности распределения задач между участниками | | | | | | | | | | | |
| Учет (минимизация) количества людей и роботов, назначаемых на выполнение задач(и) (+/-) | + | + | + | - | - | - | + | - | - | - | + |
| Учет (минимизация) финансовых затрат на выполнение отдельной задачи/всего технологического процесса (+/-) | - | - | - | + | - | - | - | + | + | - | + |
| Учет (минимизация) времени выполнения отдельной задачи (+/-) | + | + | + | + | - | + | + | - | + | - | + |

6. Заключение

В работе представлено краткое описание основных результатов, полученных в рамках цикла работ. Предложенные подход, методика и моделирующий комплекс позволяют решать три основные задачи при распределении работ в КРТС (формирование оптимального состава исполнителей, минимизация времени на выполнение всех задач и затрат для задаваемого технологического процесса), описание которых кратко представлено в настоящей работе.

Описаны результаты серии экспериментов для задач формирования оптимального состава исполнителей и минимизации времени выполнения задач технологического процесса (по три случая для каждой задачи, отражающие три реализованных механизма решения).

Представлены показатели эффективности распределения задач между участниками и оценки эффективности взаимодействия людей и коботов. Показан обзор вычислительных экспериментов других авторов, посвященных распределению задач в многоагентных и коллаборативных робототехнических системах. Представлен качественный сравнительный анализ результатов экспериментов (количество людей и роботов, задач, общее время выполнения задач), проведена оценка условий и результатов экспериментов на основе сформированных показателей эффективности взаимодействия людей и коботов и эффективности распределения задач между участниками.

Анализ показывает сопоставимые результаты по минимизации времени выполнения задач с использованием авторских решений по сравнению с рассмотренными работами. Следует отметить, что авторская постановка задач распределения работ в КРТС по сравнению с другими работами обладает комплексностью. В частности, во всех моделируемых задачах (подбор оптимального состава исполнителей, все варианты минимизации времени выполнения задач технологического процесса и затрат) обязательно проверяется требование коллаборации людей и коботов (совместное выполнение любой задачи хотя бы одним коботом и хотя бы одним человеком), ведется учет сте-

пени задействования каждого участника. Осуществляется проверка выполнимости технологического процесса расчетным (оптимальным) составом участников КРТС (в случае минимизации количества исполнителей). Кроме того, минимизация состава исполнителей, финансовых затрат и времени выполнения каждой задачи возможна в любых конфигурациях задаваемого технологического процесса, предельной численности людей и роботов в КРТС и их параметров.

Литература

1. ГАЛИН Р.Р., ГАЛИНА С.Б., МАМЧЕНКО М.В. *Подход к распределению работ в коллаборативной робототехнической системе с учетом модели рабочего пространства и динамического переназначения исполнителей* // Известия Кабардино-Балкарского научного центра РАН. – 2023. – Т. 6, №116. – С. 21–32.
2. ГАЛИН Р.Р., ШИРОКИЙ А.А., МАГИД Е.А. и др. *Эффективное функционирование смешанной неоднородной команды в коллаборативной робототехнической системе* // Информатика и автоматизация. – 2021. – Т. 20, №6. – С. 1224–1253.
3. ПЕТРЕНКО В.И., ТЕБУЕВА Ф.Б., ГУРЧИНСКИЙ М.М. и др. *Алгоритм планирования последовательности выполнения задач агентами роевых робототехнических систем в условиях неполноты информации* // Сборник материалов V Международной научно-практической конференции «Развитие современной науки и технологий в условиях трансформационных процессов». – 2022. – С. 160–169.
4. ALI A., AZEVEDO-SA H., TILBURY D.M. et al. *Heterogeneous human–robot task allocation based on artificial trust* // Scientific Reports. – 2022. – Vol. 12, No. 1. – P. 1–15.
5. ALI A., AZEVEDO-SA H., TILBURY D.M. et al. *Using Trust for Heterogeneous Human-Robot Team Task Allocation* // AAAI Fall Symposium on AI for HRI. – 2021. – P. 1–4.
6. CHEN F., SEKIYAMA K., CANNELLA F. et al. *Optimal sub-task allocation for human and robot collaboration within hybrid*

- assembly system* // IEEE Trans. on Automation Science and Engineering. – 2013. – Vol. 11, No. 4. – P. 1065–1075.
7. CHOWDHURY M., MAIER M. *Local and nonlocal human-to-robot task allocation in fiber-wireless multi-robot networks* // IEEE Systems Journal. – 2017. – Vol. 12, No. 3. – P. 2250–2260.
 8. DARVISH K., BRUNO B., SIMETTI E. et al. *Interleaved online task planning, simulation, task allocation and motion control for flexible human-robot cooperation* // 27th IEEE Int. Symposium on Robot and Human Interactive Communication (RO-MAN-2018). – 2018. – P. 58–65.
 9. DJEZAIRI S., AKLI I., ZAMOUM R.B. et al. *Mission allocation and execution for human and robot agents in industrial environment* // 27th IEEE Int. Symposium on Robot and Human Interactive Communication (RO-MAN-2018). – 2018. – P. 796–801.
 10. FUSARO F., LAMON E., DE MOMI E. et al. *An integrated dynamic method for allocating roles and planning tasks for mixed human-robot teams* // 30th IEEE Int. Conf. on Robot & Human Interactive Communication (RO-MAN-2021). – 2021. – P. 534–539.
 11. GALIN R.R., GALINA S.B. *Numerical Modeling Results of Work Allocation Algorithm in Collaborative Robotic System* // Int. Russian Automation Conference (RusAutoCon-2024). – 2024. – P. 843–848.
 12. GALIN R.R., MAMCHENKO M.V., GALINA S.B. *Task Allocation Methodology in Collaborative Robotic Systems* // Int. Russian Automation Conference (RusAutoCon-2023). – 2023. – P. 1004–1009.
 13. GALIN R., SHIROKY A., MAGID E. et al. *Effective Functioning of a Mixed Heterogeneous Team in a Collaborative Robotic System* // Informatics and Automation. – 2021. – Vol. 20, No. 6. – P. 1224–1253.
 14. GALINA S.B., GALIN R.R. *Approach to Efficient Task Allocation and Cost Minimization in Collaborative Robotic Systems* // International Russian Smart Industry Conference (SmartIndustryCon-2023). – 2023. – P. 574–579.
 15. HABIBIAN S., LOSEY D.P. *Encouraging human interaction with robot teams: Legible and fair subtask allocations* // IEEE

- Robotics and Automation Letters. – 2022. – Vol. 7, No. 3. – P. 6685–6692.
16. HAM A., PARK M.J. *Human-robot task allocation and scheduling: Boeing 777 case study* // IEEE Robotics and Automation Letters. – 2021. – Vol. 6, No. 2. – P. 1256–1263.
 17. HOWARD A.M. *Role allocation in human-robot interaction schemes for mission scenario execution* // Proc. of the IEEE Int. Conf. on Robotics and Automation (ICRA-2006). – 2006. – P. 3588–3594.
 18. JOHANNSMEIER L., HADDADIN S. *A hierarchical human-robot interaction-planning framework for task allocation in collaborative industrial assembly processes* // IEEE Robotics and Automation Letters. – 2016. – Vol. 2, No. 1. – P. 41–48.
 19. JUNG Y., KIM H., SUH K.D. et al. *Human-Centered Dynamic Service Scheduling Approach in Multi-Agent Environments* // Applied Sciences. – 2022. – Vol. 12, No. 21, 10850. – P. 1–18.
 20. KARAMI H., DARVISH K., MASTROGIOVANNI F. *A task allocation approach for human-robot collaboration in product defects inspection scenarios* // 29th IEEE Int. Conf. on Robot and Human Interactive Communication (RO-MAN-2020). – 2020. – P. 1127–1134.
 21. LEE M.L., BEHDAD S., LIANG X. et al. *Task allocation and planning for product disassembly with human–robot collaboration* // Robotics and Computer-Integrated Manufacturing. – 2022. – Vol. 76, No. 102306. – P. 1–15.
 22. LI X., ZHANG Z., WU D.D. et al. *A multi-robot allocation model for multi-object based on Global Optimal Evaluation of Revenue* // Int. Journal of Advanced Robotic Systems. – 2021. – Vol. 18, No. 6. – P. 1–18.
 23. LIPPI M., MARINO A. *A mixed-integer linear programming formulation for human multi-robot task allocation* // 30th IEEE Int. Conf. on Robot & Human Interactive Communication (RO-MAN-2021). – 2021. – P. 1017–1023.
 24. MAMCHENKO M.V., GALINA S.B. *Modified Algorithm and Simulation Tool for Task Allocation and Work Distribution in Collaborative Robotic Systems for Technological Processes* //

- Int. Russian Automation Conference (RusAutoCon-2024). – 2024. – P. 691–696.
25. MINA T., KANNAN S.S., JO W. et al. *Adaptive workload allocation for multi-human multi-robot teams for independent and homogeneous tasks* // IEEE Access. – 2020. – Vol. 8. – P. 152697–152712.
 26. NOORMOHAMMADI-ASL A., AYUB A., SMITH S.L., DAUTENHAHN K. *Task Selection and Planning in Human-Robot Collaborative Processes: To be a Leader or a Follower?* // 31st IEEE Int. Conf. on Robot and Human Interactive Communication (RO-MAN-2022). – 2022. – P. 1244–1251.
 27. RAHMAN S.M.M., WANG Y. *Mutual trust-based subtask allocation for human–robot collaboration in flexible lightweight assembly in manufacturing* // Mechatronics. – 2018. – Vol. 54. – P. 94–109.
 28. RIEDELBAUCH D., HENRICH D. *Coordinating flexible human-robot teams by local world state observation* // 26th IEEE Int. symposium on Robot and Human Interactive Communication (RO-MAN-2017). – 2017. – P. 1000–1005.
 29. RIEDELBAUCH D., HENRICH D. *Exploiting a human-aware world model for dynamic task allocation in flexible human-robot teams* // Int. Conf. on Robotics and Automation (ICRA-2019). – 2019. – P. 6511–6517.
 30. TRAM A.V.N., RAWEEWAN M. *Optimal Task Allocation in Human-Robotic Assembly Processes* // 5th Int. Conf. on Robotics and Automation Engineering (ICRAE-2020). – 2020. – P. 158–162.
 31. ZHANG F., ZHANG Y., XU S. *Collaboration effectiveness-based complex operations allocation strategy towards to human–robot interaction* // Autonomous Intelligent Systems. – 2022. – Vol. 2, No. 20. – P. 1–12.

**APPROACH TO ALLOCATION OF TASK
OF TECHNOLOGICAL PROCESSES AMONG CRS
HETEROGENEOUS PARTICIPANTS TAKING INTO
ACCOUNT THEIR STATE**

Rinat Galin, V.A. Trapeznikov Institute of Control Sciences of RAS, Moscow, Cand.Sc., Senior Researcher (grr@ipu.ru).

Mark Mamchenko, V.A. Trapeznikov Institute of Control Sciences of RAS, Moscow, Researcher (markmamcha@gmail.com).

Saniya Galina, V.A. Trapeznikov Institute of Control Sciences of RAS, Moscow, Researcher (kameshevasaniya@gmail.com).

Vasiliy Zorin, V.A. Trapeznikov Institute of Control Sciences of RAS, Moscow, Applicant (v.a.zorin@mail.ru)

Abstract: The paper describes the results of a series of experiments for the problems of forming an optimal composition of performers, and minimizing the execution time of the tasks in the technological process (three cases for each problem). Criteria of the efficiency of task allocation and work distribution among the participants and human-robot collaboration are presented. An overview of computational experiments of other authors on task allocation and work distribution in multi-agent and collaborative robotic systems (CRS) is shown. Qualitative comparative analysis of the results obtained from the experiments, and their evaluation based on the established criteria are given. The analysis shows comparable results of the authors' solutions compared to the reviewed works. The authors' statement of the problem of task allocation and work distribution in CRS compared to other sources has complexity. In particular, when selecting optimal composition of the performers, minimizing tasks/technological process execution time or costs, the results are verified to ensure that people and robots always collaborate. The degree of engagement of each participant is monitored. In addition, when optimal number of people and cobots from the CRS is chosen, the feasibility of the technological process for the calculated composition is also checked.

Keywords: collaborative robotic system, human-robot interaction, task allocation, technological process, optimization.

УДК 004.896+519.85

ББК 30.606

*Статья представлена к публикации
членом редакционной коллегии Л.Б. Рапопортом.*

Поступила в редакцию 07.11.2024.

Опубликована 30.11.2024.

АЛГОРИТМ ИНТЕГРАЦИИ СЕНСОРНЫХ ДАННЫХ ДЛЯ ОЦЕНКИ СОСТОЯНИЯ АВТОНОМНЫХ РОБОТОВ В ИНТЕЛЛЕКТУАЛЬНОЙ ТРАНСПОРТНОЙ СРЕДЕ¹

Трефилов П. М.²

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

Предложен алгоритм интеграции сенсорных данных для оценки состояния автономных объектов в интеллектуальной транспортной среде (ИТС). Основное внимание уделяется обеспечению точности навигации в сложных и динамически изменяющихся условиях городской среды, где традиционные методы навигации, такие как GPS, могут быть недостаточно эффективными. Предложенный алгоритм объединяет данные от различных сенсоров (LIDAR, камеры, инерциальные сенсоры, GPS) и элементов ИТС, обеспечивая точную оценку положения и траектории движения автономных систем. Экспериментальные результаты, полученные в условиях имитационного моделирования и натурных испытаний, подтвердили повышение точности, что делает его перспективным для применения в автономных транспортных средствах. В работе также обсуждаются возможности дальнейшего развития алгоритмов машинного обучения и методов защиты данных для улучшения эффективности и безопасности навигационных систем в ИТС.

Ключевые слова: интеграция сенсорных данных, интеллектуальные транспортные системы, автономные объекты, алгоритмы навигации, безопасность.

1. Введение

Разработка навигационных систем для робототехнических платформ в интеллектуальной транспортной среде (ИТС) – это активная область исследований, направленная на создание методов точного и надежного перемещения роботов в динамических условиях. Вектор состояния навигационной системы, включающий такие параметры, как положение, ориентация, скорость

¹ Исследование выполнено за счет гранта Российского научного фонда №23-19-00338, <https://rscf.ru/project/23-19-00338/>.

² Петр Михайлович Трефилов, н.с. (petertrfi@ipu.ru).

и траектория движения динамических объектов, требует постоянной оценки в реальном времени для обеспечения безопасного и эффективного функционирования автономных систем.

Особое внимание уделяется не только разработке методов точного позиционирования и навигации, но и обеспечению эффективного взаимодействия роботов с элементами интеллектуальной транспортной инфраструктуры, такими как адаптивные светофоры, интеллектуальные дорожные знаки и автоматизированные системы управления движением. Интеграция информации от различных источников является ключевым фактором для повышения надежности и устойчивости навигационных решений.

Одной из основных проблем является обеспечение высокой точности и надежности навигации в сложных условиях эксплуатации: при ограниченной видимости, неблагоприятных погодных условиях, в плотной городской застройке, где сигналы глобальных навигационных спутниковых систем (GNSS) могут быть искажены или недоступны. В таких ситуациях необходимо разрабатывать алгоритмы, способные эффективно использовать данные от бортовых сенсоров (лидаров, радаров, камер, инерциальных навигационных систем) и внешних источников, интегрированных в ИТС, для обеспечения точного определения положения и ориентации робота.

Кроме того, системы должны быть достаточно гибкими, чтобы адаптироваться к изменениям в транспортной среде, таким как временные перекрытия дорог, изменения в разметке или появление новых дорожных знаков. Это требует использования алгоритмов машинного обучения и искусственного интеллекта, которые способны быстро обучаться и адаптироваться к новым условиям, а также анализировать большие объемы данных для принятия оптимальных решений.

Развитие навигационных систем для робототехнических платформ в ИТС направлено на создание полностью автономных транспортных средств, способных безопасно и эффективно перемещаться в сложных и динамически изменяющихся условиях городской среды. В будущем ожидается дальнейшее совершенствование сенсорных технологий, методов обработки данных и алго-

ритмов искусственного интеллекта, что позволит повысить точность и надежность навигационных систем. С развитием технологий Интернета вещей (IoT) и увеличением количества подключенных устройств ИТС будут становиться все более комплексными, интегрируя в себя не только транспортные средства и дорожную инфраструктуру, но и различные устройства, которые могут предоставить дополнительную информацию для улучшения навигации и управления движением.

Таким образом, успешное развитие и внедрение навигационных систем для роботов в ИТС требует комплексного подхода, объединяющего достижения в области сенсорных технологий, машинного обучения, коммуникаций и кибербезопасности. Такие системы могут сыграть ключевую роль в создании умных городов будущего, где транспорт будет безопасным, эффективным [2].

2. Основные подходы к локализации объектов в ИТС

Одним из ключевых направлений в этой области является одновременная локализация и построения карты Simultaneous Localization and Mapping (SLAM). Этот метод позволяет робототехническим платформам одновременно создавать карту окружающей среды и определять свое собственное местоположение. Ведущие ученые в этой области, такие как Хью Дуррант-Уайт и Джон Леонард, заложили фундамент современных подходов к локализации и картографии в своих работах, включая статью [5]. Их исследования существенно повысили точность оценки вектора состояния, особенно в условиях неизвестной и динамически изменяющейся среды.

Современные усовершенствования метода SLAM, как показано в [4], позволяют интегрировать дополнительные сенсорные данные для повышения точности. Преимуществом этих методов является возможность одновременного построения карты окружающей среды и определения собственного местоположения без предварительно известной карты или зависимости от GPS-сигнала. Метод адаптируется к различным типам сенсоров, включая лидары, камеры и ультразвуковые датчики. Однако процесс

одновременной локализации и построения карты требует значительных вычислительных ресурсов, что может ограничивать его применение в реальном времени на платформах с ограниченными вычислительными возможностями. Кроме того, ошибки в данных сенсоров могут привести к накоплению погрешностей в построении карты и ухудшению общей локализации системы.

Еще одним важным направлением является применение методов машинного обучения и искусственного интеллекта для улучшения навигационных систем. Себастьян Трун, один из ведущих ученых в этой области, предложил использование байесовских сетей и других методов машинного обучения для повышения точности и надежности автономных систем [13]. Его подходы позволяют роботам адаптироваться к изменяющимся условиям и улучшать свои способности к самообучению. Современные модели глубокого обучения, такие как глубокие сверточные нейронные сети (Convolutional Neural Networks, CNN), позволяют роботам совершенствовать навыки навигации и взаимодействия с окружающей средой [8]. Преимущество применения методов машинного обучения заключается в способности роботов обучаться на основе накопленного опыта и улучшать модели поведения и локализации со временем. Эти методы эффективно анализируют большие объемы данных от множества сенсоров, учитывая сложные нелинейные зависимости. В отличие от численных методов, машинное обучение позволяет создавать предсказательные модели, которые учитывают поведение других динамических объектов и прогнозируют изменения в дорожной обстановке. Однако методы машинного обучения требуют значительных объемов данных для обучения моделей, что может быть затруднительно в условиях быстро меняющейся среды. Кроме того, результаты, полученные с помощью искусственного интеллекта, иногда трудно интерпретировать, что осложняет диагностику ошибок и их устранение.

Алгоритмы фильтрации и оптимизации также играют ключевую роль в оценке состояния динамических систем. Фильтр Калмана, предложенный Рудольфом Калманом [7], остается фундаментальным инструментом для оценки вектора состояния в реальном времени. Этот метод используется для сглаживания и предсказания параметров, что особенно важно в условиях шума

и неопределенности. Расширенный фильтр Калмана (ЕКФ) и другие современные методы продолжают быть эффективными для работы в реальном времени и широко используются для предсказания состояния систем [15]. Основным достоинством стоит выделить простоту реализации и высокую вычислительно эффективность (работа в режиме реального времени на многих платформах с относительно небольшими вычислительными способностями). ФК используется в широком диапазоне приложений: от локализации роботов до контроля движения транспортных средств, способен эффективно уменьшать влияние шума и неточностей в данных, предоставляемых сенсорами. Однако ФК имеет ряд ограничений на применение нелинейных моделей, что ограничивает его применимость в условиях, где присутствуют нелинейные зависимости. Ошибки в модели системы или неверные предположения о распределении ошибок могут привести к значительным отклонениям в оценке состояния.

Интеграция данных от различных сенсоров для повышения точности и надежности является еще одним важным аспектом разработки навигационных систем. Так, в статье [3] предложены методы объединения данных от лидаров, камер и инерциальных сенсоров, что позволяет повысить устойчивость систем к различным внешним воздействиям. Современные методы объединения данных от нескольких сенсоров, такие как алгоритмы на основе байесовских сетей, позволяют компенсировать недостатки одного сенсора за счет преимуществ другого, повышая общую точность системы [9]. В случае отказа одного сенсора система может продолжать функционировать, используя данные от других сенсоров. Система может адаптироваться к различным условиям, например, использовать оптические сенсоры при хорошей видимости и инерциальные данные при плохой видимости. Однако в случае значительной информационной избыточности фильтрация погрешностей, основанных на различных физических принципах, не предоставляется возможным.

В контексте интеллектуальных транспортных систем интеграция роботов в ИТС требует учета взаимодействия с инфраструктурой. В статье [12] описаны подходы к интеграции автономных транспортных средств в городскую среду, подчеркивая

важность взаимодействия с интеллектуальной транспортной инфраструктурой для обеспечения безопасности и эффективности движения. Использование данных от ИТС (например, о пробках и дорожных условиях) позволяет роботам оптимизировать свои маршруты и снижать время в пути. Современные алгоритмы взаимодействия позволяют автономным транспортным средствам эффективно координировать свои действия с элементами ИТС и другими участниками движения, что улучшает общую эффективность транспортной средств [16]. Взаимодействие с ИТС позволяет автономным транспортным средствам заранее получать предупреждения о потенциальных опасностях, таких как аварии или закрытые дороги. Однако эффективность таких систем сильно зависит от уровня развития и распространенности ИТС, что ограничивает их применение в менее развитых регионах. Передача данных между автономными транспортными средствами и ИТС может создавать риски утечки информации и кибератак, что требует дополнительных мер безопасности.

Развитие ИТС требует создания надежных и эффективных механизмов защиты данных, передаваемых между автономными транспортными средствами и элементами инфраструктуры. Для минимизации рисков, связанных с кибератаками и утечкой данных, разрабатываются современные криптографические методы и протоколы безопасной передачи данных. Важную роль в этом играют технологии распределенных реестров, которые позволяют обеспечить прозрачность и защиту передаваемой информации, а также улучшить доверие между участниками транспортной системы [1, 10].

Тем не менее внедрение этих технологий требует решения ряда технических и организационных проблем. Во-первых, необходимо обеспечить совместимость и стандартизацию протоколов передачи данных между различными элементами ИТС и автономными транспортными средствами. Это включает в себя разработку единых стандартов для связи и взаимодействия, а также создание централизованных и децентрализованных систем управления доступом к данным.

Во-вторых, интеграция криптографических методов защиты данных может привести к увеличению вычислительных затрат

и задержек в передаче информации. В условиях реального времени это может негативно сказаться на безопасности и эффективности работы ИТС. Для решения этой проблемы разрабатываются алгоритмы оптимизации вычислительных процессов, такие как методы параллельных вычислений и распределенной обработки данных, которые позволяют снизить нагрузку на системы и улучшить их производительность [14].

Кроме того, важно учитывать аспекты этики и конфиденциальности при разработке и внедрении ИТС. Сбор и обработка больших объемов данных о перемещении транспортных средств и их взаимодействии с инфраструктурой поднимают вопросы о защите персональных данных и соблюдении прав пользователей. В этой связи особое внимание уделяется разработке методов анонимизации и минимизации данных, которые позволяют сохранять приватность пользователей, не снижая при этом точности и эффективности работы систем [6].

Интеллектуальные транспортные системы (ИТС) играют важную роль в развитии и улучшении функциональности беспилотных летательных аппаратов (БПЛА). Внедрение ИТС позволяет значительно повысить точность навигации, безопасность полетов и эффективность выполнения задач в условиях сложной и динамически изменяющейся среды. ИТС обеспечивает беспилотные летательные аппараты критически важной информацией в реальном времени, что позволяет им корректировать маршрут, избегать столкновений, эффективно координировать действия с другими участниками воздушного движения и выполнять задачи даже в условиях ограниченной видимости или отсутствия GPS-сигнала. Так, в работе [11] используется метод оппортунистического слияния данных с различных сенсоров с применением байесовского фильтра Монте-Карло для повышения точности навигации. Особенностью подхода является возможность динамической интеграции наблюдений от внешних дружественных сенсорных устройств, что повышает устойчивость системы и делает ее пригодной для работы как в помещениях, так и на открытом воздухе. Управление процессом сбора информации осуществляется с помощью многоагентного слоя, что обеспечивает прозрачность для основного фильтрующего решения. В качестве

доказательства концепции представлены предварительные результаты, полученные на реальной платформе, специализирующейся на навигации на открытом воздухе.

Таким образом, разработка навигационных систем для робототехнических платформ в интеллектуальной транспортной среде является многогранной задачей, требующей комплексного междисциплинарного подхода. Это включает в себя не только совершенствование методов локализации, картографии и машинного обучения, но и интеграцию современных технологий безопасности и защиты данных. Решение этих задач позволит создать более эффективные, надежные и безопасные интеллектуальные транспортные системы, способные удовлетворить растущие потребности современного общества.

3. Алгоритм интеграции сенсорных данных для оценки состояния автономных объектов в ИТС

Для обеспечения безопасного и точного перемещения беспилотного летательного аппарата (БПЛА) в ИТС необходимо разработать алгоритм интеграции сенсорных данных. В ИТС используются многочисленные источники информации, включая данные от других транспортных средств, элементов дорожной инфраструктуры и различных сенсоров. Традиционные методы объединения данных могут сталкиваться с проблемой информационной избыточности, когда избыточные данные не уменьшают, а наоборот увеличивают погрешность, что снижает общую точность системы. Поэтому требуется особый подход, который учитывает и минимизирует такие риски, обеспечивая при этом высокую точность оценки состояния аппарата.

Таким образом алгоритм должен:

1. Собирать данные с различных сенсоров (LIDAR, камеры, IMU, GPS, и др.), а также агрегировать информацию, поступающую от элементов ИТС.
2. Оценивать качество получаемых данных для выявления возможных ошибок или недостатков.
3. Выбирать наиболее релевантные данные, исключая те, которые могут ухудшить общую оценку состояния.

4. Адаптивно объединять выбранные данные, используя взвешенные методы, чтобы минимизировать погрешность.

5. Предсказывать состояние БЛА на основе модели движения и предыдущих данных.

6. Корректировать предсказанное состояние с учетом собранной информации.

7. Проверять условия завершения миссии и повторять процесс, пока цель не будет достигнута.

Для математической формализации алгоритма примем начальное состояние вектора состояния x_0 и ковариационную матрицу ошибки P_0 , равную начальным значениям, в которой $x(t)$ включает параметры положения, ориентации, скорости и другие динамические параметры БПЛА.

В каждый момент времени t БПЛА получает данные с различных сенсоров: $z_{CAM}(t)$, $z_{IMU}(t)$, $z_{GPS}(t)$, $I_{ITS}(t)$, где: $z_{CAM}(t)$ – измерение с камеры; $z_{IMU}(t)$ – измерения от инерциальной навигационной системы; $z_{GPS}(t)$ – измерения спутниковой навигационной системы; $I_{ITS}(t)$ – измерения о местоположении и скорости объекта с использованием ИТС.

Ковариационная матрица измерений примет вид

$$(1) R_i(t) = Cov(z_i(t)).$$

Стандартные отклонения и уровни неопределенности для каждого сенсор имеют вид

$$(2) \sigma_i(t) = \sqrt{diag R_i(t)}.$$

Для выбора релевантных данных необходимо произвести оценку. На основе анализа качества данных выбираются наиболее информативные сенсоры. Пороговое значение (*threshold*) математически определяется в зависимости от задачи и используемого метода анализа данных. В контексте интеграции сенсорных данных *threshold* используется для отбора релевантных данных, минимизации погрешности и устранения ненадежных данных. Допустим, выбранные сенсоры обозначаются индексами j .

Выбираются сенсоры, удовлетворяющие условию

$$(3) Z(t) = \{z_j(t) : \sigma_j(t) < threshold\}.$$

Проверка на корреляцию между данными сенсоров будет выглядеть как:

$$(4) \quad p_{ij}(t) = \frac{\text{Cov}(z_i(t), z_j(t))}{\sigma_{z_i}(t)\sigma_{z_j}(t)},$$

где $\text{Cov}(z_i(t), z_j(t))$ – ковариация между измерениями $z_i(t)$ и $z_j(t)$; $\sigma_{z_i}(t)$ и $\sigma_{z_j}(t)$ – стандартные отклонения измерений $z_i(t)$ и $z_j(t)$ соответственно.

Если $|p_{ij}(t)| > \text{corr_threshold}$, то данные $z_j(t)$ исключаются, чтобы избежать разнородной ошибки, возникшей из-за разных физических принципов. corr_threshold используется для контроля качества и релевантности данных, участвующих в оценке состояния объекта, исключая те данные, которые могут привести к ухудшению точности оценки из-за их высокой взаимозависимости. Этот порог задается в зависимости от конкретной задачи и методов анализа данных

Для каждого оставшегося сенсора рассчитывается объединенное значение измерений $z^{comb}(t)$ полученных с различных сенсоров, которые прошли фильтрацию и были отобраны как релевантные:

$$(5) \quad z^{comb}(t) = \sum_j \omega_j(t)z_j(t),$$

где веса $\omega_j(t)$ определяются как

$$(6) \quad \omega_j(t) = \frac{1/\sigma_j^2(t)}{\sum_k 1/\sigma_k^2(t)}.$$

Таким образом, $z^{comb}(t)$ представляет собой среднее взвешенное значение измерений, учитывающее точность каждого сенсора, что позволяет минимизировать погрешность в оценке состояния системы. То есть сенсоры с меньшей неопределенностью имеют больший вес.

В случае отсутствия сигналов от сенсоров необходимо проведение экстраполяции измерений. Используя модель движения БЛА, предсказание состояния производится следующим образом:

$$\hat{x}(t|t-1) = f(\hat{x}(t-1), u(t-1)).$$

Обновление ковариационной матрицы:

$$(7) \quad P(t|t-1) = F(t-1)P(t-1)F^T(t-1) + Q(t-1),$$

где $F(t-1)$ – матрица частных производных функции перехода, а $Q(t-1)$ – ковариационная матрица процессного шума.

Для реализации коррекции вектора состояния с помощью обратной связи на основе объединения измерительной информации с различных сенсоров используется фильтр Калмана:

$$(8) \hat{x}(t) = \hat{x}(t | t - 1) + K(t)(z^{comb}(t) - h(\hat{x}(t | t - 1)))$$

где матрица Калмана $K(t)$:

$$K(t) = P(t | t - 1)H^T(t) (H(t)P(t | t - 1)H^T(t) + R_{comb}(t))^{-1},$$

где $H(t)$ – матрица частных производных функции наблюдения, $R_{comb}(t)$ – объединенная ковариационная матрица ошибок наблюдений.

Обновление ковариационной матрицы состояния происходит следующим образом:

$$(9) P(t) = (I - K(t)H(t))P(t | t - 1)$$

где I — единичная матрица, $H(t)$ — матрица частных производных функции наблюдения, которая связывает измерения с состоянием.

4. Проведение эксперимента

Для оценки эффективности предложенного алгоритма интеграции сенсорных данных были проведены две серии экспериментов: имитационное моделирование и натурные испытания.

Целью эксперимента является оценка эффективности предложенного алгоритма интеграции сенсорных данных для автономного объекта (БПЛА) в условиях имитационной и реальной среды.

Имитационная модель представляет собой виртуальную среду, где автономный объект оснащен несколькими сенсорами (камера, IMU, GPS), генерирующими данные о состоянии объекта (положение, ориентация, скорость и т.д.). В модели учитываются различные факторы, такие как шумы и погрешности в измерениях, задержки в передаче данных и внешние воздействия на сенсоры.

Для оценки эффективности предложенного алгоритма интеграции сенсорных данных, включая обработку данных, фильтрацию, коррекцию состояния с использованием фильтра Калмана и управление на основе объединенной информации, задаются

начальные параметры для симуляции, включая значения `corr_threshold`.

Для сравнения использовались траектории, полученные следующим образом:

1) БПЛА полагается только на собственные сенсоры (ИНС, камеры, GPS) и предварительно загруженные карты для обхода запрещенной зоны;

2) БПЛА получает поддержку от ИТС, включая обновления данных о статических и динамических препятствиях (например, других летательных аппаратах, строительных кранах), а также рекомендации по оптимальному маршруту через городской район.

Эксперимент проводился с использованием пяти различных сценариев:

1) БПЛА пролетает через имитацию застройки, избегая препятствий, таких как башни и другие высокие объекты.

2) БПЛА выполняет миссию и сталкивается с нештатной ситуацией, требующей экстренной посадки (например, потеря связи или отказ оборудования). Оценивается способность системы безопасно приземлить БПЛА с использованием информации от ИТС.

3) БПЛА должен выполнить задание в условиях ограниченной видимости, таких как туман, дождь или ночь. В качестве эксперимента камера была отключена и информации от нее не поступало.

4) БПЛА должен доставить груз в указанную точку назначения, при этом условия на маршруте могут динамически изменяться (например, изменение конечной точки доставки).

5) БПЛА должен выполнить задание вблизи зоны, в которой запрещены полеты. Необходимо безопасно обойти запрещенную зону, не нарушая ее границы (запретная зона сформирована искусственно).

В качестве платформы использовался БПЛА, собранный из общедоступных комплектующих и изображенный на рис. 1.

Во время полета местоположение БПЛА постоянно изменялось и сравнивалось с запланированным маршрутом. Запись данных осуществлялась с частотой 100 Гц для ИНС и 10 Гц для СНС и камеры. Отклонение от маршрута определялось как расстояние

между фактическим положением БПЛА и ближайшей точкой на запланированном маршруте. Процент точности удержания маршрута рассчитывался как отношение времени, когда БПЛА находился в пределах допустимого отклонения, к общему времени полета по маршруту.



Рис. 1. Модель используемого БПЛА

По результатам эксперимента рассчитывался процент точности удержания маршрута. Он рассчитывался, как отношение количества времени, когда БПЛА находился в пределах допустимого отклонения, к общему времени полета по маршруту.

По результатам серии экспериментов получены значения, изображенные на рис. 2.

Результаты продемонстрировали, что при слабой корреляции данных от сенсоров предложенный алгоритм интеграции сенсорных данных способен обеспечивать высокую точность и стабильность в оценке состояния автономных объектов. Это свидетельствует о том, что алгоритм может эффективно минимизировать погрешности, вызванные несовпадением данных от различных сенсоров, и способен работать в условиях информационной избыточности, сохраняя при этом более высокую точность.

Сравнение точности удержания маршрута БПЛА с различными конфигурациями сенсоров показало, что использование предложенного подхода позволило улучшить контроль за траекторией движения объекта, особенно в условиях, где сигналы GPS

нестабильны или отсутствуют. Таким образом, алгоритм доказал свою эффективность в условиях реальной эксплуатации, где интеграция данных от множества источников играет ключевую роль в обеспечении безопасности и точности навигации.

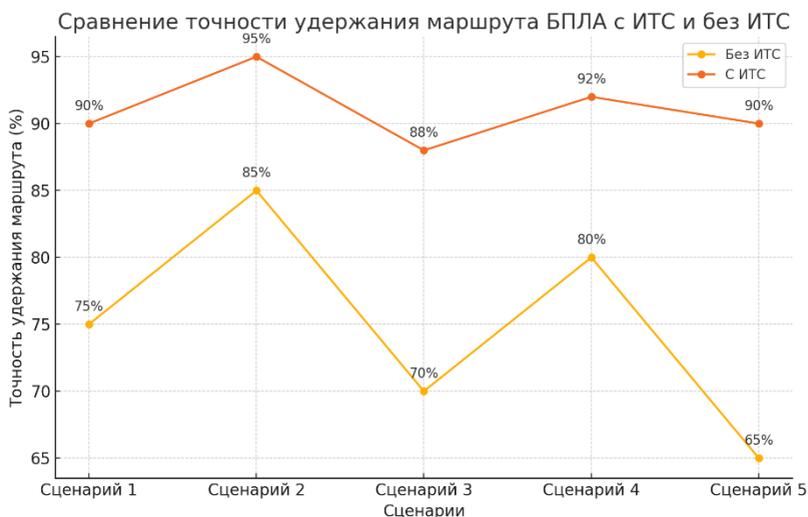


Рис. 2. Сравнение точности удержания маршрута БПЛА

5. Заключение

В данной работе предложен и экспериментально проверен алгоритм интеграции сенсорных данных для оценки состояния автономных объектов в интеллектуальных транспортных системах. Результаты проведенных экспериментов показали, что предложенный подход обеспечивает более высокую точность оценки состояния по сравнению с традиционными методами.

Перспективы дальнейших исследований включают совершенствование алгоритмов обработки и интеграции данных, что позволит повысить устойчивость и адаптивность систем автономного управления. В частности, планируется улучшение методов машинного обучения и фильтрации данных для уменьшения вычислительных затрат и повышения эффективности работы в реальном времени. Также в ближайших планах расширение

возможностей алгоритма для работы в условиях ограниченной видимости и сложных погодных условий, а также интеграция с новыми типами сенсоров и элементами ИТС.

С учетом быстрого развития технологий такие системы могут стать основой для создания более безопасных и эффективных умных городов будущего. Автономные транспортные средства смогут взаимодействовать с элементами инфраструктуры и друг с другом на высоком уровне, обеспечивая безопасное и эффективное управление транспортными потоками.

Литература

1. ЖАРКО Е.Ф., ПРОМЫСЛОВ В.Г., ИСХАКОВ А.Ю. и др. *Кибербезопасность беспилотных транспортных средств. Архитектура. Методы проектирования.* – М.: Радиотехника, 2021. – 160 с.
2. ROMANOVA M.A., GALIN R.R., TREFILOV P.M. *Application of IoT technologies for automation of management in agriculture* // IOP Conference Series: Earth and Environmental Science. – Krasnoyarsk: IOP Publishing Ltd, 2019. – Vol. 315. – DOI: <https://iopscience.iop.org/article/10.1088/1755-1315/315/3/032032/pdf>.
3. ABDI B., AL HAGE J., EL BADAoui EL NAJJAR M. et al. *Multi-Robot Autonomous Navigation System Using Informational Fault Tolerant Multi-Sensor Fusion with Robust Closed Loop Sliding Mode Control* // 21st International Conference on Information Fusion (FUSION) – 2018. – P. 2258–2265 – DOI: 10.23919/ICIF.2018.8455787.
4. BASHAR Md. S., ASHIKUZZAMAN A.K.M., RAFID M. et al. *Recent Advancements in Visual-Inertial Simultaneous Localization and Mapping (VIO-SLAM) for Autonomous Vehicles: A Review* // Int. Journal of Engineering Research & Technology (IJERT). – May, 2023. – Vol. 12, Iss. 05. – P. 280–284.
5. DURRANT-WHYTE H., BAILEY T. *Simultaneous Localization and Mapping: Part I* // IEEE Robotics & Automation Magazine. – 2006. – Vol. 13, No. 2. – P. 99–110. – DOI: 10.1109/MRA.2006.1638022.

6. GOSMAN C., DOBRE C., POP F. *Privacy-Preserving Data Aggregation in Intelligent Transportation Systems* // IFIP/IEEE Symposium on Integrated Network and Service Management (IM-2017), Lisbon, Portugal, 2017. – P. 1059–1064. – DOI: 10.23919/INM.2017.7987438.
7. KALMAN R. *A New Approach to Linear Filtering and Prediction Problems* // Journal of Basic Engineering. – 1960. – P. 35–45.
8. KENDALL A., GRIMES M. *End-to-End Learning of Geometry and Context for Deep Stereo Regression* // IEEE Trans. on Pattern Analysis and Machine Intelligence. – 2017. – P. 678–692.
9. LABBE M., MICHAUD F. *RTAB-Map as an Open-Source Lidar and Visual SLAM Library for Large-Scale and Long-Term Online Operation* // Journal of Field Robotics. – 2021. – Vol. 36, No. 3. – P. 545–564.
10. MADAAN G., KUMAR A., BHUSHAN B. *Blockchain Assisted Secure Data Sharing in Intelligent Transportation Systems* // Smart and Sustainable Approaches for Optimizing Performance of Wireless Networks – 2022. – P. 1834–1843.
11. MARTI E., GARCIA J., MOLINA J. *Opportunistic Multi-Sensor Fusion for Robust Navigation in Smart Environments* // User-Centric Technologies and Applications – 2011. – Vol. 94. – P. 1–10.
12. SHLADOVER S., RIZZETTO C. *Automated Vehicles and the Smart Transportation System* // IEEE Trans. on Intelligent Transportation Systems. – 2013. – P. 23–33.
13. THRUN S. *Probabilistic Robotics*. – MIT Press, 2005. – P. 45–78.
14. XIAO J., LIU Y., ZOU Y. et al. *An Efficient Elliptic Curve Cryptography-Based Secure Communication with Privacy Preserving for Autonomous Vehicle* // Journal of Advanced Transportation. – 2024. – Vol. 2024, No. 5808088. – P. 1–10. – DOI: 10.1155/2024/5808088.
15. XU M., CHEN S., HUANG S. et al. *Invariant EKF Based 3D Active SLAM with Exploration Task*. // Springer Nature – 2022. – P. 1–14.
16. BITHAS P. *Special Issue "Challenges in V2X Communications for Cooperative Autonomous Driving."* // Signal Processing – 2019. – P. 1–5.

SENSOR DATA INTEGRATION ALGORITHM FOR STATE ESTIMATION OF AUTONOMOUS ROBOTS IN AN INTELLIGENT TRANSPORT SYSTEM

Peter Trefilov, V.A. Trapeznikov Institute of Control Sciences of RAS, Moscow, (petertrfi@ipu.ru).

Abstract: This paper deals with the development and experimental testing of a sensor data integration algorithm for estimating the state of autonomous objects in intelligent transport systems (ITS). The main attention is paid to ensuring accuracy and reliability of navigation in complex and dynamically changing conditions of urban environment, where traditional navigation methods, such as GPS, may be insufficiently effective. The proposed algorithm combines data from different sensors (LiDAR, cameras, inertial sensors, GPS) and ITS elements to provide accurate position and trajectory estimation of autonomous systems. Experimental results obtained in simulation and field tests confirmed high accuracy and adaptability of the algorithm, which makes it promising for application in autonomous vehicles. The paper also discusses the possibilities of further development of machine learning algorithms and data protection methods to improve the efficiency and safety of navigation systems in ITS.

Keywords: sensor data integration, intelligent transport systems, autonomous objects, navigation algorithms, safety.

УДК 004.3+004.6
ББК 16.2

*Статья представлена к публикации
членом редакционной коллегии Н.А. Коргиным.*

*Поступила в редакцию 20.09.2024.
Опубликована 30.11.2024.*

АСИМПТОТИКА МОМЕНТОВ И ИХ ПРОИЗВОДНЫХ ДЛЯ ИЗБЫТОЧНЫХ РАСПРЕДЕЛЕНИЙ

Острер Л. А.¹, Русев В. Н.², Скориков А. В.³

(РГУ нефти и газа (НИУ) имени И.М. Губкина, Москва)

Функционирование современных сложных систем характеризуется различными видами рисков. Анализ данных таких систем показывает, что обычно наборы данных обладают характерным свойством: поведением распределения при больших значениях аргумента, которое называется тяжелым хвостом. Рассматриваются классы распределений с тяжелыми хвостами, которые имеют важные приложения в теории страховых случаев и теории надежности: распределения Гнеденко – Вейбулла; Бенктандера I, II; Бурра XII. Асимптотика момента для функции превышения среднего значения и функции превышения дисперсии были получены специально для рассматриваемых распределений с тяжелыми хвостами и могут быть использованы для получения аппроксимации при больших значениях временной переменной. В работе подробно изучается оценка погрешности для асимптотического разложения функции среднего избытка распределения Гнеденко – Вейбулла при любых значениях параметра формы. Отмечено существенное различие в поведении оценок погрешности при значениях параметра формы меньших единицы, соответствующих тяжелому хвосту распределения Гнеденко – Вейбулла. В частности, найдены значения параметра формы, при которых разложения точны, т.е. имеют конечное число слагаемых. Для распределений Гнеденко – Вейбулла; Бенктандера I, II; Бурра XII доказаны асимптотические разложения производных остаточных моментов. Рассмотрено также описание поведения системы как области притяжения предельного экстремального состояния. Результаты статьи служат инструментом для приложений к теории риска, надежности и экстремальным событиям.

Ключевые слова: асимптотические разложения, средние избыточные функции, избыточная дисперсия, распределения с тяжелыми хвостами.

1. Введение

Существование рисков для жизни, собственности, окружающей среды характерно для функционирования современных больших систем. В работе Микош [16] отмечено, что данные о страховании от пожаров в Дании и данные о промышленных

¹ Леонид Александрович Острер, ст. преподаватель (leonidostrer@gmail.com).

² Владимир Николаевич Русев, к.т.н., доцент (rusev.v@gubkin.ru).

³ Александр Васильевич Скориков, к.ф.-м.н, доцент (skorikov.a@gubkin.ru).

пожарах в США могут быть смоделированы с помощью распределений с тяжелыми хвостами, которые характеризуются более медленным убыванием при больших значениях аргумента, чем хвосты любого экспоненциального распределения. Другой характеристикой распределения с тяжелым хвостом является отсутствие конечного экспоненциального момента любого порядка. Для моделей распределений с тяжелыми хвостами характерен принцип единого большого скачка, который лежит в основе вероятностного поведения сумм независимых случайных величин: вероятность того, что сумма случайных величин может превышать некоторое большое значение x , совпадает с вероятностью того, что максимум одной из этих отдельных случайных величин также превышает x . Систематическое изложение теории распределений с тяжелыми хвостами дано в работах Румянцева, Морозова [3] и Фосс [10]. Избыточные распределения и избыточные моменты, введенные Эмбрехтс и др. [9], являются известным инструментом, используемым для характеристики пиков, превышающих пороговое значение. Пусть X – неотрицательная случайная величина с функцией распределения F с неограниченным носителем, т.е. $F(x) < 1$ для всех x . Рассмотрим случайную величину $X_t = (X - t | X > t)$ с функцией распределения F_t , которая называется функцией распределения превышения порогового значения t . Функция распределения F_t в теории надежности известна как функция распределения избыточного срока службы или остаточного ресурса. Функция $\bar{F}(x) = 1 - F(x) = P(X > x)$ есть распределение хвоста. Математическое ожидание случайной величины X_t называется функцией среднего избытка (ME – mean excess function) X , т.е. является средней ожидаемой продолжительностью жизни в зависимости от возраста t или средним остаточным сроком службы (MRL – mean residual life) в теории надежности. Функция среднего избытка (ME) определяется формулой

$$(1) \quad \mu(t) = E(X - t | X > t) = \int_t^{\infty} \bar{F}(x) dx / \bar{F}(t).$$

Заметим, что для распределений с тяжелыми хвостами обычно функции $\mu(t)$ стремятся к бесконечности при $t \rightarrow \infty$.

Важное значение при моделировании играет дисперсия. В работе Калашникова, Константиноидиса [2] отмечено, что если дисперсия размеров выплат мала, то даже при наличии тяжелого хвоста функция риска будет как у распределений с конечными экспоненциальными моментами. Функция избыточной дисперсии (функция остаточной дисперсии) определяется по формуле

$$(2) \quad \sigma^2(t) = E(X_t^2) - \mu^2(t) = \frac{2}{\bar{F}(t)} \int_t^{\infty} \bar{F}(x) \cdot \mu(x) dx - \mu^2(t).$$

Функции превышения пороговых значений, функции среднего превышения играют фундаментальную роль в управлении рисками, актуарной науке, проблемах экстремальных значений, анализе надежности. Подробное обсуждение функции превышения среднего значения в работах Гош и Резник [12] и обзор MRL см. в статье Банджевич [7]. В статье Русева, Скорикова [4] найдены асимптотики функции риска в случае выплат для известных стандартных распределений с тяжелыми хвостами. Приложение MRL к надежности системы «скважина – насос» дано в статье авторов [1]. Главные члены асимптотического разложения функций среднего избытка для основных стандартных распределений актуарной теории можно увидеть в книге Эмбрехтс и др. [9].

Целью настоящего исследования является уточнение асимптотических разложений функций среднего избытка и избыточной дисперсии для некоторых известных распределений. Также исследуются скорости увеличения или уменьшения средних избыточных функций и избыточных дисперсий этих распределений при увеличении порогового значения t , а также их выпуклость или вогнутость.

2. Распределение Гнеденко – Вейбулла

2.1. АСИМПТОТИЧЕСКИЕ РАЗЛОЖЕНИЯ МОМЕНТОВ

Двухпараметрическое распределение Гнеденко – Вейбулла с хвостом

$$(3) \quad \bar{F}(t) = e^{-(\alpha t)^\beta}, \quad t > 0, \quad \alpha > 0, \quad \beta > 0,$$

является одним из наиболее широко применяемых распределений для анализа времени до отказа оборудования в теории надежности. Наиболее полной книгой по распределению Гнеденко – Вейбулла является справочник Ринне [18], в котором дано описание истории развития распределения в статистической теории и прикладной статистике. Распределение Гнеденко – Вейбулла является распределением с тяжелым хвостом при значениях параметра формы в интервале $(0 < \beta < 1)$. Асимптотические разложения моментов, данные авторами ранее [20], имеют вид

$$(4) \quad \mu(t) = \frac{t}{\beta(\alpha t)^\beta} \times \left(1 + \frac{1-\beta}{\beta(\alpha t)^\beta} + \frac{(1-\beta)(1-2\beta)}{\beta^2(\alpha t)^{2\beta}} + o\left(\frac{1}{t^{2\beta}}\right) \right),$$

$t \rightarrow \infty$;

$$(5) \quad \sigma^2(t) = \frac{1}{(\beta\alpha^\beta t^{(\beta-1)})^2} \times \left(1 + \frac{4(1-\beta)}{\beta(\alpha t)^\beta} + \frac{(1-\beta)(11-17\beta)}{\beta^2(\alpha t)^{2\beta}} + o\left(\frac{1}{t^{2\beta}}\right) \right),$$

$t \rightarrow \infty$.

Получим асимптотическое разложение производных. Заметим, что асимптотическое разложение производных проясняет зависимость остаточных моментов от параметров распределения. В частности, из доказанной ниже формулы (8) следует, что значение параметра, равное 0,5, является критическим для скорости изменения распределения с тяжелым хвостом. Отметим, что избыточные функции распределений выражаются обычно через специальные функции. Для изучения таких функций используются асимптотические разложения. Возникает проблема дифференцирования асимптотического разложения поскольку условия дифференцирования асимптотического разложения не всегда выполняются (например, см. Эрдейи [5]). Результат Гупта [14] позволяет обойти эту трудность и заменить операцию дифференцирования асимптотики моментов алгебраическими операциями для обширного класса распределений:

$$(6) \quad \frac{d\mu(t)}{dt} = -\frac{\bar{F}'(t)}{\bar{F}(t)} \mu(t) - 1;$$

$$(7) \quad \frac{d\sigma^2(t)}{dt} = -\frac{\bar{F}'(t)}{\bar{F}(t)} [\sigma^2(t) - \mu^2(t)].$$

Теорема 1. Пусть X – случайная величина, которая имеет распределение Гнеденко – Вейбулла, тогда

$$(8) \quad \frac{d\mu}{dt} = \frac{(1-\beta)}{\beta(\alpha t)^\beta} \left[1 + \frac{(1-2\beta)}{\beta(\alpha t)^\beta} + o\left(\frac{1}{t^\beta}\right) \right], \quad t \rightarrow \infty;$$

$$(9) \quad \frac{d\sigma^2(t)}{dt} = \frac{2(1-\beta)}{\beta^2} \cdot \frac{t^{1-2\beta}}{\alpha^{2\beta}} \left[1 + \frac{2(2-3\beta)}{\beta(\alpha t)^\beta} + o\left(\frac{1}{t^\beta}\right) \right], \quad t \rightarrow \infty.$$

Доказательство. Подставив функции из (3), (4) в (6), получим

$$\frac{d\mu(t)}{dt} = \alpha\beta(\alpha t)^{\beta-1} \cdot \frac{1}{\alpha\beta(\alpha t)^{\beta-1}} \left[1 + \frac{(1-\beta)}{\beta(\alpha t)^\beta} + \frac{(1-\beta)(1-2\beta)}{\beta^2(\alpha t)^{2\beta}} + o\left(\frac{1}{t^{2\beta}}\right) \right] - 1,$$

что дает асимптотическое разложение (8). Производную дисперсии можно вычислить по формуле

$$\frac{d\sigma^2(t)}{dt} = (\alpha\beta)(\alpha t)^{\beta-1} \left[\sigma^2(t) - \mu^2(t) \right],$$

где

$$\mu^2(t) = \frac{t^2}{\beta^2(\alpha t)^{2\beta}} \left[1 + \frac{2(1-\beta)}{\beta(\alpha t)^\beta} + \frac{(1-\beta)(3-5\beta)}{\beta^2(\alpha t)^{2\beta}} + o\left(\frac{1}{t^{2\beta}}\right) \right], \quad t \rightarrow \infty.$$

После некоторых преобразований получаем:

$$\sigma^2 - \mu^2 = \frac{2(1-\beta) \cdot t^2}{\beta^3(\alpha t)^{3\beta}} \left[1 + \frac{2(2-3\beta)}{\beta(\alpha t)^\beta} + o\left(\frac{1}{t^\beta}\right) \right], \quad t \rightarrow \infty.$$

Из формулы (7) и функции распределения (3) имеем

$$\frac{d\sigma^2(t)}{dt} = (\alpha\beta) \cdot (\alpha t)^{\beta-1} \cdot \frac{2(1-\beta) \cdot t^2}{\beta^3(\alpha t)^{3\beta}} \left[1 + \frac{2(2-3\beta)}{\beta(\alpha t)^\beta} + o\left(\frac{1}{t^\beta}\right) \right], \quad t \rightarrow \infty;$$

$$\frac{d\sigma^2(t)}{dt} = \frac{2(1-\beta)}{\beta^2} \cdot \frac{t^{1-2\beta}}{\alpha^{2\beta}} \left[1 + \frac{2(2-3\beta)}{\beta(\alpha t)^\beta} + o\left(\frac{1}{t^\beta}\right) \right], \quad t \rightarrow \infty.$$

Следствие. Таким образом, при больших значениях t имеем для тяжелых хвостов:

$$(10) \quad \frac{d\mu(t)}{dt} > 0 \Rightarrow \mu(t) \uparrow; \quad \frac{d\mu(t)}{dt} \downarrow 0, \quad 0 < \beta < 1;$$

$$(11) \quad \frac{d\sigma^2(t)}{dt} > 0 \Rightarrow \sigma^2(t) \uparrow, \quad 0 < \beta < 1.$$

Однако поведение производной существенно зависит от порогового значения параметра формы: $\beta = 1/2$:

$$(12) \frac{d\sigma^2(t)}{dt} \sim c^2 t^\varepsilon, \quad \varepsilon > 0, \quad \frac{d\sigma^2(t)}{dt} \uparrow, \quad 0 < \beta < \frac{1}{2};$$

$$(13) \frac{d\sigma^2(t)}{dt} \sim c^2 t^{-\varepsilon}, \quad \varepsilon > 0, \quad \frac{d\sigma^2(t)}{dt} \downarrow, \quad \frac{1}{2} < \beta < 1;$$

$$(14) \frac{d\sigma^2(t)}{dt} = \frac{4}{\alpha} \left[1 + \frac{1.2}{\sqrt{\alpha t}} + o\left(\frac{1}{\sqrt{t}}\right) \right], \quad \beta = \frac{1}{2}.$$

2.2. ОЦЕНКА ПОГРЕШНОСТИ

Получим оценку погрешности асимптотического разложения функции среднего избытка распределения Гнеденко – Вейбулла для любого n . Представим хвост распределения в виде

$$(15) \quad \bar{F}(x) = -\frac{x^{1-\beta}}{\alpha^\beta \cdot \beta} \cdot \frac{d\bar{F}(x)}{dx}.$$

Преобразуем интеграл $\mu(t)$ по частям:

$$\begin{aligned} \mu(t) &= \frac{1}{\bar{F}(t)} \int_t^{+\infty} \bar{F}(x) dx = \frac{1}{\bar{F}(t)} \int_t^{+\infty} \left[-\frac{x^{1-\beta}}{\alpha^\beta \cdot \beta} \right] d\bar{F}(x) = \\ &= \frac{1}{\bar{F}(t)} \left\{ \frac{t^{1-\beta} \bar{F}(t)}{\alpha^\beta \cdot \beta} + \int_t^{+\infty} \frac{(1-\beta)x^{-\beta}}{\alpha^\beta \cdot \beta} \bar{F}(x) dx \right\}. \end{aligned}$$

Учитывая (15), имеем

$$(16) \quad \mu(t) = \frac{t^{1-\beta}}{(\alpha^\beta \cdot \beta)} - \frac{1}{\bar{F}(t)} \cdot \frac{(1-\beta)}{(\alpha^\beta \cdot \beta)^2} \int_t^{+\infty} x^{1-2\beta} d\bar{F}(x).$$

Обозначим

$$(17) \quad I_k(t) = \frac{1}{\bar{F}(t)} \int_t^{+\infty} \left[-\frac{x^{1-k\beta}}{(\alpha^\beta \cdot \beta)^k} \right] d\bar{F}(x).$$

Формула (16) принимает вид

$$(18) \quad \mu(t) = I_1(t) = \frac{t^{1-\beta}}{(\alpha^\beta \cdot \beta)} + (1-\beta)I_2(t).$$

Аналогичным образом получим из (17) рекуррентную формулу

$$(19) I_k(t) = \frac{t^{1-k\beta}}{(\alpha^\beta \cdot \beta)^k} + (1-k\beta)I_{k+1}(t).$$

Подстановка в (18) даёт

$$(20) \mu(t) = \frac{t^{1-\beta}}{(\alpha^\beta \cdot \beta)} + (1-\beta) \left[\frac{t^{1-2\beta}}{(\alpha^\beta \cdot \beta)^2} + (1-2\beta)I_3(t) \right].$$

Дальнейшие итерации $I_k(t)$ дают точное разложение $\mu(t)$ по степенным функциям для любых t :

$$(21) \mu(t) = \frac{t^{1-\beta}}{(\alpha^\beta \cdot \beta)} + (1-\beta) \frac{t^{1-2\beta}}{(\alpha^\beta \cdot \beta)^2} + (1-\beta)(1-2\beta) \frac{t^{1-3\beta}}{(\alpha^\beta \cdot \beta)^3} + \dots$$

$$+ (1-\beta)(1-2\beta)(1-3\beta) \dots (1-n\beta) \times$$

$$\times \left[\frac{t^{1-(n+1)\beta}}{(\alpha^\beta \cdot \beta)^{n+1}} + (1-(n+1)\beta)I_{n+2}(t) \right].$$

Или, обозначив степенную функцию через

$$z(t) = t^{-\beta} / (\alpha^\beta \cdot \beta),$$

равенство (21) перепишем в виде

$$(22) \mu(t) = t \cdot z(t) \cdot \left[1 + (1-\beta)z + 1-\beta(1-2\beta)z^2 + \right.$$

$$+ 1-\beta(1-2\beta)(1-3\beta)z^3 + \dots$$

$$\left. + 1-\beta(1-2\beta) \dots (1-n\beta)z^n \right] + R_n(t).$$

Остаточный член:

$$R_n(t) = \prod_{k=1}^{n+1} (1-k\beta) \cdot I_{n+2}(t).$$

Подставим выражение $I_{n+2}(t)$ в форме интеграла (17):

$$(23) R_n(t) = \prod_{k=1}^{n+1} (1-k\beta) \cdot \frac{1}{F(t)} \int_t^{+\infty} \left[-\frac{x^{1-(n+2)\beta}}{(\alpha^\beta \cdot \beta)^{n+2}} \right] d\bar{F}(x).$$

Если отбросить остаточный член в формуле (22), то получим приближённое представление $\mu(t)$:

$$(24) \mu(t) \approx t \cdot z(t) \cdot \left[1 + (1-\beta)z + (1-\beta)(1-2\beta)z^2 + \right.$$

$$\left. + (1-\beta)(1-2\beta)(1-2\beta)z^3 + \dots + (1-\beta)(1-2\beta) \dots (1-n\beta)z^n \right]$$

с погрешностью

$$\Delta_n = |R_n(t)|.$$

При значениях параметра формы $\beta > 1$ модуль подынтегральной степенной функции убывает, и оценка остаточного члена принимает простой вид:

$$(25) |R_n(t)| \leq \prod_{k=1}^{n+1} (k\beta - 1) \cdot \frac{t^{1-(n+2)\beta}}{(\alpha^\beta \cdot \beta)^{n+2}}.$$

Число членов в формуле (24) (при любом значении β) надо выбирать так, чтобы последующее слагаемое по модулю было меньше предыдущего. В частности,

$$\delta_n(t) = \left| \frac{R_n(t)}{u_n(t)} \right| \leq \frac{|1 - (n+1)\beta|}{\beta} (\alpha t)^{-\beta} = \left(n + 1 - \frac{1}{\beta} \right) (\alpha t)^{-\beta} < 1$$

при достаточно больших допустимых αt , т.е. погрешность приближения должна быть меньше последнего учтённого слагаемого.

Более сложным является случай $0 < \beta < 1$. Отметим, что в формуле (22) остаточный член исчезает при

$$(26) \beta = \frac{1}{n+1}, \quad n=1, 2, \dots$$

При таких β формула (22) даёт представление $\mu(t)$ в виде конечной суммы степенных функций.

Проиллюстрируем алгоритм оценки погрешности на примере $1/2 < \beta < 1$. Положим в формулах (21), (23) $\beta = 1/2 + \varepsilon$, где $0 < \varepsilon < 1/2$. Получим, ограничившись для определённости тремя явно выписанными членами,

$$(27) \mu(t) = \frac{t^{1/2-\varepsilon}}{(\alpha^\beta \cdot \beta)} + \left(\frac{1}{2} - \varepsilon \right) \frac{t^{-2\varepsilon}}{(\alpha^\beta \cdot \beta)^2} + \left(\frac{1}{2} - \varepsilon \right) (-2\varepsilon) \frac{t^{-\frac{1}{2}-3\varepsilon}}{(\alpha^\beta \cdot \beta)^3} + R_3(t).$$

Из монотонного убывания подынтегральной функции в оценке R_3 вытекает оценка погрешности:

$$(28) \quad |R_3(t)| \leq \left(\frac{1}{2} - \varepsilon\right) (2\varepsilon) \left(\frac{1}{2} + 3\varepsilon\right) \frac{1}{F(t)} \int_i^\infty \frac{x^{1-4\left(\frac{1}{2}+\varepsilon\right)}}{(\alpha^\beta \cdot \beta)} d\bar{F}(x) \leq \\ \leq \left(\frac{1}{2} - \varepsilon\right) (2\varepsilon) \left(\frac{1}{2} + 3\varepsilon\right) \frac{t^{-1-4\varepsilon}}{\alpha^\beta \cdot \beta^4}.$$

При достаточно больших (αt) отношение

$$\delta_3 = \left| \frac{R_3(t)}{u_3(t)} \right| \leq \frac{\frac{1}{2} + 3\varepsilon}{\frac{1}{2} + \varepsilon} (\alpha t)^{-\left(\frac{1}{2} + \varepsilon\right)} \text{ меньше 1.}$$

2.3. ОЦЕНКА ПОГРЕШНОСТИ В МОДЕЛИ ГНЕДЕНКО – ВЕЙБУЛЛА, ПОСТРОЕННОЙ ПО ДАННЫМ ОТКАЗОВ ПОГРУЖНОГО ОБОРУДОВАНИЯ

По данным отказов погружного оборудования в работе [1] было получено, что время работы до отказа системы «скважина – насос» имеет распределение Гнеденко – Вейбулла с параметрами $\alpha = 0,014$; $\beta = 1,49$. Для выбора необходимого числа членов асимптотического разложения остаточного времени жизни $\mu(t)$ воспользуемся условием

$$\delta_n(t) = \left| \frac{R_n(t)}{u_n(t)} \right| < 1$$

и построим графики отношений $\delta_n(t)$, $n = 1, 2, \dots$ (рис. 1).

Вероятность безотказной работы системы «скважина – насос» в течение времени более чем 200 суток при полученных значениях параметров распределения меньше 1%. Поэтому выбор числа слагаемых в асимптотическом разложении и соответствующего номера n остаточного члена $R_n(t)$ зависит от выбора допустимого времени t эксплуатации установки. Оценки погрешностей при трех и четырех слагаемых в разложении можно не рассматривать, так как условия $\delta_2(t) < 1$, $\delta_3(t) < 1$, которые соответствуют трем и четырем слагаемым, выполняются при $t > 125$ и $t > 160$ соответственно. Вероятности безотказной работы равны $P(125) = 0,009$ и $P(160) = 0,03$, т.е. число проработавших устано-

вок в первом случае не превосходит 9%, во втором – 3%. Рассмотрим оценки погрешности при выборе одного или двух слагаемых в асимптотическом разложении $\mu(t)$.

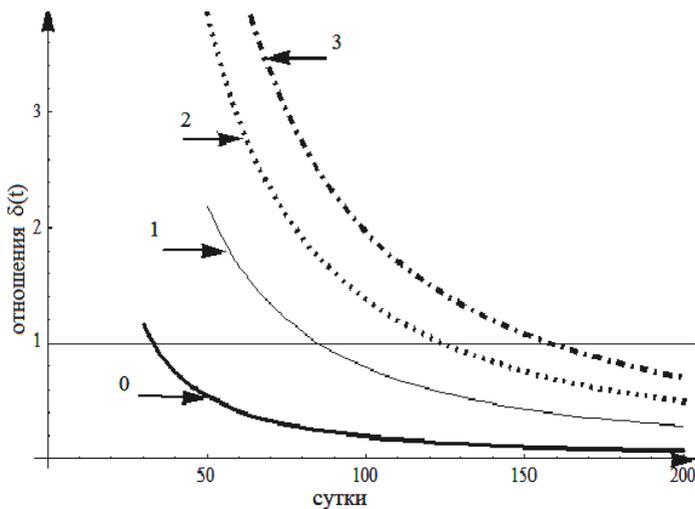


Рис. 1. Графики отношений $\delta_0(t)$, $\delta_1(t)$, $\delta_2(t)$, $\delta_3(t)$ (обозначены через 0, 1, 2, 3 соответственно)

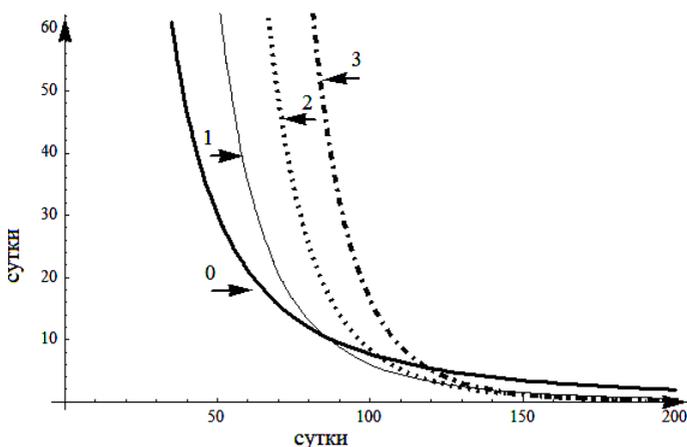


Рис. 2. Графики погрешностей для остаточных членов $R_0(t)$, $R_1(t)$, $R_2(t)$, $R_3(t)$ (обозначены через 0, 1, 2, 3 соответственно)

Минимальное допустимое время при выборе одного слагаемого в формуле (24) асимптотического разложения с остаточным членом $R_0(t)$, согласно графику на рис. 1, можно взять 34 суток. Этому времени соответствует погрешность 64 суток. При выборе двух слагаемых с остаточным членом $R_0(t)$, при минимальном допустимом времени 85 суток получаем погрешность в 10 суток. Отметим, что нулевой член разложения t дает такую же погрешность, так как графики пересекаются в этой точке. Однако при $t > 85$ разложение с двумя слагаемыми дает меньшую погрешность. Таким образом, асимптотическое разложение с двумя слагаемыми – рекомендуемый выбор в формуле (24) для модели Гнеденко – Вейбулла времени работы системы «скважина – насос».

3. Распределение Бенкандера типа I

Помимо распределения Гнеденко – Вейбулла, для моделирования в страховании используются важные распределения, такие как распределения Бенкандера I и II. Бенкандер определил эти распределения для моделирования крупных убытков в актуарной науке. Теоретически предсказанные размеры требований с помощью распределений Бенкандера I и II типов согласуются со статистическими данными о размере требований в области актуарных наук.

Хвост распределения Бенкандера типа I имеет вид

$$(29) \quad 1 - F(x) = \left(1 + 2 \frac{\beta}{\alpha} \ln x \right) e^{-\beta(\ln x)^2 - (\alpha+1)\ln x}, \quad \alpha > 0, \beta > 0, x > 1.$$

Распределение Бенкандера типа I аналогично логнормальному распределению, и известна функция среднего избытка (см. таблицу 3.4.7 в книге [9]):

$$(30) \quad \mu(t) = \frac{t}{2\beta \ln t + \alpha}.$$

Получим асимптотическое разложение дисперсии.

Теорема 2. Для распределения Бенкандера типа I справедливы следующие асимптотические разложения дисперсии и её производной:

$$(31) \quad \sigma^2(e) = \frac{t^2}{(2\beta \ln t + \alpha)(2\beta \ln t + \alpha - 1)} \times \\ \times \left[1 + \frac{1}{2\beta \ln t + \alpha} + o\left(\frac{1}{\ln t}\right) \right], t \rightarrow \infty;$$

$$(32) \quad \frac{d\sigma^2(e)}{dt} = \frac{2t}{(2\beta \ln t + \alpha)^2} [1 + o(1)]; \quad t \rightarrow \infty.$$

Доказательство. Подставим в интеграл формулы (2) функцию среднего избытка (30) и хвост распределения (29):

$$I = \int_t^{+\infty} \bar{F}(x) \mu(x) dx = \frac{1}{\alpha} \int_t^{+\infty} e^{-\alpha \left[\frac{\beta}{\alpha} \ln^2 x + \ln x \right]} dx.$$

После замены

$$\xi(t) = \sqrt{\beta} \left(\ln t + \frac{(\alpha - 1)}{2\beta} \right)$$

получаем представление в виде интеграла Лапласа:

$$I = \frac{e^{\frac{(\alpha-1)^2}{4\beta}}}{\alpha \sqrt{\beta}} \int_{\xi(t)}^{+\infty} e^{-\xi^2} d\xi.$$

Далее используем асимптотическое разложение этого интеграла

$$I(t) = \frac{e^{(\alpha-1)^2/4\beta}}{\alpha \sqrt{\beta}} e^{-\xi^2(t)} \cdot \frac{1}{2\xi(t)} \left[1 - \frac{1}{(2\xi^2)^2} + \frac{1 \cdot 3}{(2\xi^2)^2} - \frac{1 \cdot 3 \cdot 5}{(2\xi^2)^3} + \dots \right], \quad \xi \rightarrow \infty.$$

Таким образом,

$$(33) \quad \sigma^2(t) = Z(t) \left[1 - \frac{1}{(2\xi^2)^2} + \frac{1 \cdot 3}{(2\xi^2)^2} - \frac{1 \cdot 3 \cdot 5}{(2\xi^2)^3} + \dots \right] - \frac{t^2}{(2\beta \ln t + \alpha)^2},$$

где

$$Z(t) = \frac{2t^2}{(2\beta \ln t + \alpha)(2\beta \ln t + \alpha - 1)}, \quad \xi^2(t) = \frac{1}{4\beta} (2\beta \ln t + (\alpha - 1))^2.$$

Главный член асимптотического разложения (33) имеет вид

$$U(t) = Z(t) - \frac{t^2}{(2\beta \ln t + \alpha)^2} =$$

$$(34) = \frac{t^2}{(2\beta \ln t + \alpha)} \left[\frac{2}{(2\beta \ln t + (\alpha - 1))} - \frac{1}{(2\beta \ln t + \alpha)} \right] =$$

$$= \frac{t^2}{(2\beta \ln t + \alpha)(2\beta \ln t + \alpha - 1)} \left[1 + \frac{1}{2\beta \ln t + \alpha} \right].$$

Отсюда следует (31).

Для вывода асимптотического разложения (32) используется представление

$$(35) \frac{d\sigma^2(t)}{dt} = -\frac{d \ln \bar{F}(t)}{dt} [\sigma^2(t) - \mu^2(t)].$$

Имеем для хвоста распределения Бенкандера типа I и производной

$$\ln \bar{F}(t) = \ln \left(1 + 2 \frac{\beta}{\alpha} \ln t \right) - [\beta \ln^2 t + (\alpha + 1) \ln t];$$

$$\frac{d \ln \bar{F}(t)}{dt} = \left\{ \frac{2 \frac{\beta}{\alpha}}{1 + 2 \frac{\beta}{\alpha} \ln t} - [2\beta \ln t + (\alpha + 1)] \right\} \frac{1}{t};$$

$$\frac{d \ln \bar{F}(t)}{dt} = -\frac{1}{t} \left\{ 2\beta \ln t + (\alpha - 1) + 2 - \frac{1}{\ln t} + o \left(\frac{1}{\ln t} \right) \right\}, \quad t \rightarrow \infty.$$

Асимптотическое разложение второго множителя (35) можно получить, используя (33). Отсюда главный член асимптотического разложения (35) имеет вид

$$A(t) = Z(t) - 2\mu^2(t) = \frac{2t^2}{(2\beta \ln t + \alpha)^2 (2\beta \ln t + \alpha - 1)}.$$

Тогда (32) следует из (35) путем замены множителей асимптотическими разложениями, приведенными выше.

Следствие. Для распределения Бенкандера типа I ($0 < \alpha$; $0 < \beta$) с тяжелым хвостом при $t \rightarrow \infty$ избыточная дисперсия увеличивается и скорость роста увеличивается.

4. Распределение Бенктандера типа II

Рассмотрим распределение Бенктандера типа II с хвостом распределения

$$(36) \bar{F}(t) = e^{\frac{\alpha}{\beta} t^{\beta-1}} \cdot e^{-\frac{\alpha}{\beta} t^{\beta}} \quad \alpha > 0, 0 < \beta < 1, t > 1.$$

Это распределение близко к распределению Гнеденко – Вейбулла. Функция среднего избытка Бенктандера типа II имеет вид

$$(37) \mu(t) = \frac{1}{\alpha} t^{1-\beta},$$

с точностью до множителя $\frac{1}{\beta\alpha^{\beta-1}}$, совпадающий с главным членом разложения $\mu(t)$ распределения Гнеденко – Вейбулла.

Асимптотическое разложение избыточной дисперсии для Бенктандера типа II

$$(38) \sigma^2(t) = \frac{t^{2-2\beta}}{\alpha^2} \times \left(1 - 2 \frac{(\beta-1)}{\alpha} t^{-\beta} + 2 \frac{(\beta-1)(2\beta-1)}{\alpha^2} t^{-2\beta} + o(t^{-2\beta}) \right)$$

найдено авторами ранее и опубликовано в материалах конференции [21].

Получим асимптотическое разложение производной избыточной дисперсии.

Теорема 3. Имеет место следующее асимптотическое разложение производной избыточной дисперсии для распределения Бенктандера типа II:

$$(39) \frac{d\sigma^2(t)}{dt} = \frac{2(1-\beta)}{\alpha^2} t^{1-2\beta} \left[1 + \frac{(2-3\beta)}{\alpha} \frac{1}{t^{\beta}} + o\left(\frac{1}{t^{\beta}}\right) \right], \quad t \rightarrow \infty,$$

$$\alpha > 0, 0 < \beta < 1, t > 1.$$

Доказательство. Для доказательства (39) используем представление (35) производной дисперсии. Имеем

$$\frac{\bar{F}'(t)}{\bar{F}(t)} = -\alpha t^{\beta-1} \left[1 + \frac{(1-\beta)}{\alpha} \frac{1}{t^{\beta}} \right];$$

$$\sigma^2(t) - \mu^2(t) = \frac{2(1-\beta)}{\alpha^3} t^{2-3\beta} \left[1 + \frac{(1-2\beta)}{\alpha} \frac{1}{t^{\beta}} + o\left(\frac{1}{t^{\beta}}\right) \right], \quad t \rightarrow \infty.$$

Подставив в (35), получаем

$$\frac{d\sigma^2(t)}{dt} = \frac{2(1-\beta)}{\alpha^2} \cdot t^{1-2\beta} \left[1 + \frac{(1-\beta)}{\alpha} \cdot \frac{1}{t^\beta} \right] \left[1 + \frac{(1-2\beta)}{\alpha} \cdot \frac{1}{t^\beta} + o\left(\frac{1}{t^\beta}\right) \right], t \rightarrow \infty;$$

$$\frac{d\sigma^2(t)}{dt} = \frac{2(1-\beta)}{\alpha^2} \cdot t^{1-2\beta} \left[1 + \frac{(2-3\beta)}{\alpha} \cdot \frac{1}{t^\beta} + o\left(\frac{1}{t^\beta}\right) \right], t \rightarrow \infty.$$

Следствие. Для распределения Бенкандера II типа асимптотическое поведение избыточной дисперсии аналогично поведению избыточной дисперсии распределения Гнеденко – Вейбулла.

5. Распределение Бурра типа XII

Хвост распределения Бурра типа XII задается формулой

$$(40) \quad \bar{F}(t) = \frac{1}{(1+t^c)^k}, t \geq 0, c > 0, k > 0.$$

Бурр ввёл двенадцать различных видов распределений, см. [19]. В настоящей статье рассматривается одно из этих распределений, которое называется «Бурра типа XII». Многие авторы рассматривали расширение распределения Бурра XII, например Олападе [17], Гад и др. [11]. Распределение Бурра XII используется во многих приложениях, таких как страхование и финансовая экономика, анализ выживаемости (см., например, Бурнецки и др. [8], Махмуд и др. [15]). Асимптотические разложения средней избыточной функции и избыточной дисперсии для распределения Бурра типа XII были получены авторами ранее и имеют вид

$$(41) \quad \mu(t) = \frac{t}{(kc-1)} \left[1 + \frac{kc}{((k+1)c-1)} \cdot \frac{1}{t^c} + o\left(\frac{1}{t^c}\right) \right], t \rightarrow \infty;$$

$$(42) \quad \sigma^2(t) = \frac{t^2}{(kc-1)^2} \times \left[1 + \frac{2}{kc-2} + \frac{2c^2k(k+1)}{((k+1)c-1)((k+1)c-2)} \cdot \frac{1}{t^c} + o\left(\frac{1}{t^c}\right) \right], t \rightarrow \infty.$$

Докажем теорему об асимптотическом разложении производных моментов распределения Бурра типа XII.

Теорема 4.

$$(43) \quad \frac{d\mu(t)}{dt} = \frac{1}{(kc-1)} \left[1 + \frac{kc(1-c)}{(kc-1+c)} \cdot \frac{1}{t^c} + o\left(\frac{1}{t^c}\right) \right], \quad t \rightarrow \infty;$$

$$(44) \quad \frac{d\sigma^2(t)}{dt} = t \cdot \frac{2kc}{(kc-1)^2(kc-2)} [1 + o(1)], \quad t \rightarrow \infty.$$

Доказательство. На первом шаге доказательства вычислим логарифмическую производную функции распределения:

$$\left[\ln \bar{F}(t) \right]' = \frac{-kct^{c-1}}{1+t^c} = -\frac{kc}{t} \cdot \frac{1}{1+\frac{1}{t^c}} = -\frac{kc}{t} \left[1 - \frac{1}{t^c} + \frac{1}{t^{2c}} + o\left(\frac{1}{t^{2c}}\right) \right].$$

Отсюда

$$(45) \quad \left[\ln \bar{F}(t) \right]' = -\frac{kc}{t} \left[1 - \frac{1}{t^c} + o\left(\frac{1}{t^c}\right) \right], \quad t \rightarrow \infty.$$

Используя (41) и (45), получаем разложения

$$\left[\ln \bar{F}(t) \right]' \cdot \mu(t) = -\frac{kc}{kc-1} \left[1 + \frac{1}{t^c} \left(\frac{kc}{(k+1)c-1} - 1 \right) + o\left(\frac{1}{t^c}\right) \right];$$

$$\left[\ln \bar{F}(t) \right]' \cdot \mu(t) + 1 = \left(-\frac{1}{kc-1} \right) - \frac{kc(1-c)}{(kc-1)(kc-1+c)} \cdot \frac{1}{t^c} + o\left(\frac{1}{t^c}\right), \quad t \rightarrow \infty.$$

Отсюда по формуле (6) получаем асимптотическое разложение (43). Докажем (44), используя представление (7). Имеем

$$\sigma^2(t) - \mu^2(t) = \frac{2t^2}{(kc-1)^2(kc-2)} \left[1 + \frac{kc \cdot 2(kc-2)}{((k+1)c-1) \cdot ((k+1)c-2)} \cdot \frac{1}{t^c} + o\left(\frac{1}{t^c}\right) \right].$$

$$\begin{aligned} \frac{d\sigma^2(t)}{dt} &= -\frac{d \ln \bar{F}(t)}{dt} \left(\sigma^2(t) - \mu^2(t) \right) = \\ &= \frac{2tkc}{(kc-1)^2(kc-2)} \left[1 + \frac{1}{t^c} \left[\frac{2kc(kc-2)}{((k+1)c-1) \cdot ((k+1)c-2)} - 1 \right] + o\left(\frac{1}{t^c}\right) \right] = \\ &= t \cdot \frac{2kc}{(kc-1)^2(kc-2)} [1 + o(1)], \quad t \rightarrow \infty. \end{aligned}$$

6. Предельные распределения

Рассмотрим асимптотическое поведение рассмотренных функций распределения превышения порогового значения при $t \rightarrow \infty$.

В теории экстремальных значений широко используется понятие области притяжения предельных распределений, введенное Гнеденко [13]. Для каждого предельного распределения G область притяжения остаточного времени жизни состоит из всех функций распределения F , для которых существуют такие нормализующие функции $a(t)$, $b(t)$, что выполняется равенство

$$\lim_{t \rightarrow \infty} F_t(b(t) + xa(t)) = G(x).$$

Критерий области притяжения для предельного экспоненциального распределения был доказан Балкема, Де Хаан [6] в виде

$$(46) \quad c_v(t) = \frac{\sigma(t)}{\mu(t)} \rightarrow 1, \quad t \rightarrow \infty.$$

Используя теоремы 2,3,4, можно легко проверить, что условие (46) выполняется для распределений Гнеденко – Вейбулла, Бенкандера I, II. В силу разложений (41), (42) и критерия (46) область притяжения экспоненциального распределения не включает распределение Бурра XII.

Литература

1. ДЕНЬГАЕВ А.В., РУСЕВ В.Н., СКОРИКОВ А.В. *Исследование средней остаточной наработки в модели Гнеденко – Вейбулла распределения отказов. Оценки остаточного времени ресурса погружного насосного оборудования* // Сборник «Труды Российского государственного университета нефти и газа имени И.М. Губкина (НИУ)» – 2020. – №1(298). – С. 61–73.
2. КАЛАШНИКОВ В.В., КОНСТАНТИНИДИС Д.Г. *Вероятность разорения* // *Фундаментальная и прикладная математика* – 1996. – Т. 2, вып. 4. – С. 1055–1100.
3. РУМЯНЦЕВ А.С., МОРОЗОВ Е.В. *Распределения с тяжелыми хвостами и их приложения*. – Петрозаводск: Изд-во ПетрГУ, 2013. – 67 с.

4. РУСЕВ В.Н., СКОРИКОВ А.В. *Асимптотика риска страхования для требований, имеющих тяжёлые хвосты* // Автоматизация и информатизация. ТЭЖ. – 2023. – № 11(604). – С. 35–40.
5. ЭРДЕЙИ А. *Асимптотические разложения*. – М.: Физматгиз, 1962. – 127 с.
6. BALKEMA A.A., DE HAAN L. *Residual life time at great age* // The Annals of probability. – 1974. – Vol.2, No. 5. – P. 792–804.
7. BANJEVIC D. *Remaining useful life in theory and practice*. – 2009. – P. 337–349. – DOI: <https://doi.org/10.1007/s00184-008-0220-5>.
8. BURNECKI K., HÄRDLE W., WERON R. *An introduction to simulation of risk processes. Encyclopedia of Actuarial Science*. – Chichester: Wiley, 2004. – P. 1–7.
9. EMBRECHTS P., KLÜPPELBERG C., MIKOSCH T. *Modeling Extremal Events for Insurance and Finance*. – Berlin: Springer-Verlag, 1997.
10. FOSS S., KORSHUNOV D., ZACHARY S. *An Introduction to Heavy-Tailed and Subexponential Distributions*. – New York: Springer, 2011. – 123 p.
11. GAD A.M., HAMEDANI G.G., SALEHABADI S.M. et al. *The Burr XII-Burr XII Distribution: Mathematical Properties and Characterizations* // Pak. J. Statist. – 2019. – Vol. 35(3). – P. 229–248.
12. GHOSH S., RESNICK S. *Discussion on mean excess plots* // Stochastic Processes and their Applications. – 2010. – Vol. 120, Iss. 8. – P. 1492–1517.
13. GNEDENKO B. *On the limiting distribution of the maximum term in a random series* // In: S.J. Kotz, Breakthroughs in Statistics. Volume 1. Foundations and Basic Theory. – New York, NY, USA: Springer Science + Business Media, 1993. – P. 185–225.
14. GUPTA R.C. *On the monotonic properties of residual variance and their applications in reliability* // Journal of Statistical Planning and Inference. – 1987. – Vol. 16. – P. 329–335.
15. MAHMOUD K. OKASHA M., MATTER Y. *On the three-parameter Burr type XII distribution and its application to heavy tailed lifetime data* // Journal of advances in mathematics. – 2015. – Vol. 10, No.4. – P. 3429–3442.
16. MIKOSCH T. *Non-Life insurances mathematics. An introduction with stochastic processes*. – Berlin: Springer, 2004. – 235 p.

17. OLAPADE A. *On a six-parameter generalized Burr XII distribution* // arXiv:0806.1579v1 [math.ST]. – 2008. – DOI: <https://doi.org/10.48550/arXiv.0806.1579>.
18. RINNE H. *The Weibull Distribution: A handbook*. – London, New York: Chapman and Hall/CRC Press, 2009. – 784 p.
19. RODRIGES R.A. *Guide to the Burr type XII distributions* // *Biometrika*. – 1977. – Vol. 64. – P. 129–134.
20. RUSEV V., SKORIKOV A. *Residual Life Time of the Gnedenko Extreme – Value Distributions. Asymptotic Behavior and Applications* // In: *Recent Developments in Stochastic Methods and Applications* / Eds.: Shiryayev A.N., Samouylov K.E., Kozyrev D.V. – Springer Proceedings in Mathematics and Statistics. – 2021. – Vol. 371. – P. 292–305.
21. RUSEV V., SKORIKOV A. *Asymptotics of Moments for the Remaining Time of Heavy-Tail Distributions* // *Comput. Sci. Math. Forum*. – 2023. – Vol.7(1), No. 52. – DOI: <https://doi.org/10.3390/IOCMA2023-14435>.

THE ASYMPTOTICS OF MOMENTS AND MOMENT'S DERIVATIVES FOR EXCESS DISTRIBUTION

Leonid Ostrer, National University of Oil and Gas «Gubkin University», Moscow, Senior Lecturer (leonidostrer@gmail.com).

Vladimir Rusev, National University of Oil and Gas «Gubkin University», Moscow, Cand. Sc., Associate Professor (rusev.v@gubkin.ru).

Alexander Skorikov, National University of Oil and Gas «Gubkin University», Moscow, Cand. Sc., Associate Professor (skorikov.a@gubkin.ru).

Abstract: The functioning of modern complex systems is characterized by various types of risks. Data analysis of such systems shows that data sets have characteristic properties: heavy distribution tails. An important issue is the impact of individual extreme events on the global behavior of the entire system too. The proposed article discusses classes of distributions with heavy tails, which are important in the theory of insurance claims and reliability theory: Gnedenko-Weibull distribution; Benktander I, II; Burr XII. The moment's asymptotics have been derived for mean excess function and excess variance function especially for the heavy-tailed distributions. The paper studies in detail the error estimate for the asymptotic expansion of the mean excess function of the Gnedenko-Weibull distribution for any values of

the shape parameter. There is a significant difference in the behavior of error estimates for values of the shape parameter less than one corresponding to the heavy tail of the Gnedenko –Weibull distribution. The values of the shape parameter are found for which the decompositions are accurate in particular. That is, the expansion has finite quantity members. Asymptotic expansions of derivatives of residual moments are proved for Gnedenko-Weibull; Benktander I, II; Burr XII distributions. The description of the behavior of the system as a region of attraction of the ultimate extreme state is also considered. These results serve as a tool for the applications to risk theory, reliability and extremal event.

Keywords: distribution tail, asymptotic expansion, mean excess functions, excess variance, heavy-tail distributions.

УДК 519.218.4 + 517.956.8 + 517.968.22

ББК 22.161.6 + 30.14

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

Поступила в редакцию 04.08.2024.

Опубликована 30.11.2024.

КАЛИБРОВКА БОЛЬШОГО МАССИВА УЛЬТРАЗВУКОВЫХ СЕНСОРОВ¹

Граничин О. Н.²

(Санкт-Петербургский государственный университет,
Санкт-Петербург)

Граничина О. А.³

(Российский государственный педагогический университет
им. А.И. Герцена, Санкт-Петербург)

Трофимов С. А.⁴

(Санкт-Петербургский государственный университет,
Санкт-Петербург)

Щербаков П. С.⁵

(ФГБУН Институт проблем управления РАН, Москва;
Московский физико-технический институт, Долгопрудный)

Предлагается метод калибровки плоского кругового массива сенсоров для ультразвуковой медицинской томографии, позволяющий оценивать координаты сенсоров и задержки приема/передачи сигналов. Он основан на модели распространения ультразвуковых волн и использует триангуляцию для оптимизации параметров модели. На каждой итерации метода оценки координат и задержек пересчитываются отдельно друг от друга путем решения систем линейных уравнений малой размерности. Моделирование на искусственно сгенерированных данных показало высокую эффективность и точность предложенного подхода. Метод устойчив к шумам и способен работать в различных условиях, включая неидеальность модели распространения ультразвуковых волн. Предложенная процедура калибровки может быть легко масштабирована на большие размеры массива, что актуально при практической реализации систем ультразвуковой томографии. Представляется, что разработанный подход позволит повысить точность результатов ультразвуковой томографии при работе на реальных данных, что делает его ценным инструментом для широкого спектра применений, таких как медицинская диагностика и робототехника.

¹ Работа О.А. Граничиной в разделах 2 и 4 выполнена при частичной финансовой поддержке внутреннего гранта РГПУ им. А.И. Герцена (проект №28ВГ).

Результаты исследований П.С. Щербакова в разделе 3 получены за счет средств РНФ (проект №21-71-30005, <https://rscf.ru/project/21-71-30005>).

² Олег Николаевич Граничин, д.ф.-м.н., профессор (oleg_granichin@mail.ru).

³ Ольга Александровна Граничина, д.пед.н., проф. (olga_granichina@mail.ru).

⁴ Степан Артемович Трофимов, аспирант (steve.trofimov@gmail.com).

⁵ Павел Сергеевич Щербаков, д.ф.-м.н., г.н.с. (cavour118@mail.ru).

Ключевые слова: ультразвуковая томография, плоский круговой массив сенсоров, идентификация параметров модели, триангуляция, системы большой размерности.

1. Введение

Ультразвуковая томография (УЗТ) широко применяется в медицинской визуализации благодаря своей неинвазивности и экономичности. Разработан ряд систем УЗТ, предназначенных специально для скрининга рака молочной железы [4, 11, 16, 18, 19], и предложены алгоритмы для реконструкции структуры исследуемой области по результатам применения этих аппаратов [5, 6, 10, 15, 20]. Настоящая работа находится в русле предшествующих исследований и направлена на повышение точности реконструкции изображений в УЗТ. Цель – разработка более эффективных методов калибровки УЗИ-сенсоров для дальнейшей реконструкции изображений, которые позволят улучшить качество изображений и повысить точность диагностики. Рассматриваемый здесь подход в предварительной форме был представлен в [1, 8]. В настоящей работе изучается более сложная модель неопределенностей и помех в наблюдениях. Предложена новая процедура оценивания ключевых параметров и разработана программная платформа для генерирования тестовых данных и их обработки. Описаны результаты моделирования.

Учитывая рост количества сенсоров и усложнение индустриально производимых систем диагностики, их калибровка приобретает первостепенное значение для обеспечения достоверности измерений. Надежная и качественная калибровка является критически важным требованием для многих приложений обработки сигналов, включая ультразвуковую томографию. Она позволяет точно определять параметры сенсоров, такие как их координаты и задержки приема/передачи сигналов, что имеет решающее значение для получения достоверных результатов измерений и корректного восстановления УЗИ-картинки на их основе. В литературе [15, 17] продемонстрирована эффективность использования калиброванных датчиков для улучшения качества рекон-

струкции в ультразвуковой томографии. В статье [15] рассматривается аналогичная задача калибровки и предлагается итеративный алгоритм, устойчивый к шумам и использующий для оптимизации матрицу расстояний. Однако алгоритм, предложенный в [15], имеет высокую вычислительную сложность, что ограничивает его применимость для систем с большим количеством сенсоров. Это побуждает к разработке новых эффективных методов калибровки, которые могут достигать высокой точности без существенных вычислительных затрат.

В работе [12] предложен итеративный метод для совместной оценки положения датчиков и скорости звука в воде. Однако одновременная аппроксимация этих параметров возможна лишь с точностью до некоторого общего масштабирующего множителя, идентификация которого требует серьезных дополнительных усилий. Кроме того, невыпуклая формулировка задачи не гарантирует глобальной сходимости алгоритма. Более того, метод не учитывает задержки приема/передачи сигналов и другие шумы, что может повлиять на точность калибровки. Таким же недостатком страдает метод из [17], где вводимая целевая функция также не является выпуклой, что не гарантирует сходимость к глобальному минимуму при оптимизации.

В статье [13] предложен алгоритм глобальной оптимизации для определения местоположения сенсоров путем максимизации резкости реконструируемого изображения. Работа алгоритма была успешно продемонстрирована на примере массива из 28 сенсоров. Однако часто наблюдалась его сходимость к локальным минимумам и плохая масштабируемость с увеличением количества датчиков. В данной статье предлагается альтернативный подход, который оказывается эффективным при решении задач калибровки УЗТ-систем с числом сенсоров порядка двух тысяч.

В [14] предложен подход к исправлению ошибок в изображениях, которые связаны с неточностями определения координат сенсоров при отсутствии калибровки. Этот подход использует алгоритмы глубокого обучения для реконструкции изображения. Однако модели глубокого обучения часто являются сложными

и не вполне интерпретируемыми [9], что затрудняет понимание и настройку процесса реконструкции. Кроме того, использование глубокого обучения зачастую приводит к потере линейности процесса восстановления изображения, что может повлиять на точность и надежность получаемых результатов.

2. Постановка задачи. Математическая модель

Рассмотрим устройство, представляющее собой плоскую круговую раму с расположенными на ней n сенсорами, которые могут излучать и принимать ультразвуковой сигнал, см. рис. 1. Для точного определения времени полета сигнала (Time of Flight, ToF) от сенсора i к сенсору j , которое является основой реконструкции, требуется знать задержки r_i передачи и приема e_j сигнала, а также точные координаты расположения сенсоров на раме.

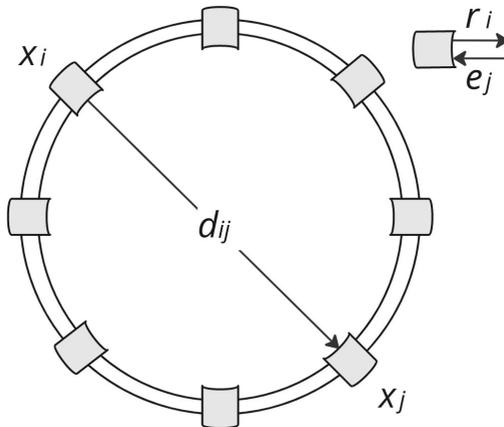


Рис. 1. Схема кругового ультразвукового аппарата с массивом сенсоров

В рамках рассматриваемой модели измеренное время m_{ij} прохождения сигнала от сенсора i к сенсору j задается следующим соотношением:

$$(1) \quad m_{ij} = t_{ij} + r_i + e_j + \delta_{ij},$$

где t_{ij} – истинное время полета сигнала; r_i и e_j – задержки излучения и приема этого сигнала; δ_{ij} – шум измерений, обусловленный дискретной природой процесса регистрации сигнала сенсором-приемником; его абсолютное значение предполагается ограниченным величиной, не превышающей времени прохождения двух длин волны сигнала. Координаты сенсоров $x_i = (x_{i1}, x_{i2})^T \in \mathbb{R}^2$ не известны и отличны от идеальных – равномерно распределенных на кольце – вследствие неизбежных неточностей производственного процесса; по той же причине неизвестны величины r_i и e_i , и ниже будет предложен алгоритм их оценивания, называемый калибровкой устройства.

Так как процесс происходит в однородной среде (воде), то первый обнаруженный на приемнике сигнал ультразвукового импульса соответствует лучу, следующему по кратчайшему пути между сенсорами, т.е. по прямой. Тогда

$$(2) \quad d_{ij} = c \cdot t_{ij},$$

где $d_{ij} = \|x_i - x_j\|_2$ – расстояние между сенсором-излучателем i и сенсором-приемником j в евклидовой норме; c – скорость распространения звука в однородной среде, предполагаемая известной точно.

Учитывая формулы (1)–(2), перепишем уравнение зависимости расстояния между сенсорами от времени прохождения сигнала:

$$(3) \quad (x_{i1} - x_{j1})^2 + (x_{i2} - x_{j2})^2 = (c \cdot t_{ij})^2 = c^2 \cdot (m_{ij} - e_i - r_j - \delta_{ij})^2,$$

где ошибка измерения входит в член, возводимый в квадрат, что затрудняет использование традиционных методов оптимизации.

3. Алгоритм калибровки

Каждая итерация алгоритма оценивания расположения x_i сенсоров и запаздываний r_i , e_i разбивается на два этапа, которые выполняются последовательно:

- 1) оценивание координат x_i ;
- 2) оценивание задержек r_i , e_i .

3.1. Оценивание расположения сенсоров

Пусть после шага s алгоритма имеем оценки x^s координат сенсоров и оценки e^s, r^s задержек приема и передачи сигнала.

Зафиксируем номер i сенсора; переход к оценке его координат на шаге $s + 1$ осуществляется следующим образом. Выберем случайно равномерно на множестве $\{1, \dots, n\} \setminus i$ два индекса $l \neq j$ и рассмотрим оценки координат и задержек «пробной» пары сенсоров l и j , полученные на шаге s .

Из уравнения (3) имеем:

$$(4) \quad \begin{cases} (x_{i1} - x_{l1}^s)^2 + (x_{i2} - x_{l2}^s)^2 = c^2(m_{il} - e_i^s - r_l^s - \delta_{il})^2, \\ (x_{i1} - x_{j1}^s)^2 + (x_{i2} - x_{j2}^s)^2 = c^2(m_{ij} - e_i^s - r_j^s - \delta_{ij})^2. \end{cases}$$

Раскроем скобки и вычтем второе уравнение из первого; получим линейное уравнение относительно координат x_{i1}, x_{i2} , которое выглядит следующим образом:

$$(5) \quad \begin{aligned} & a_1(x_l^s, x_j^s)x_{i1} + a_2(x_l^s, x_j^s)x_{i2} = \\ & = u_{lj}(m_{il}, m_{ij}, \delta_{il}, \delta_{ij}, e_i^s, r_j^s, x_l^s, x_j^s), \end{aligned}$$

а выражения для a_1, a_2, u_{lj} имеют следующий вид:

$$(6) \quad \begin{aligned} a_1 &= 2(x_{j1}^s - x_{l1}^s), \\ a_2 &= 2(x_{j2}^s - x_{l2}^s), \\ u_{lj} &= c^2((m_{il} - e_i^s - r_l^s - \delta_{il})^2, \\ & - (m_{ij} - e_i^s - r_j^s - \delta_{ij})^2) - ((x_{l1}^s)^2 + (x_{l2}^s)^2) + ((x_{j1}^s)^2 + (x_{j2}^s)^2). \end{aligned}$$

При реализации алгоритма для оценивания расположения датчика i используется $k > 1$ пробных пар сенсоров, см. ниже. Пропедев описанные выше действия k раз, получим k линейных уравнений относительно координат сенсора x_i .

Запишем эти уравнения в матричном виде:

$$(7) \quad A_i x_i = U_i,$$

где $A_i \in \mathbb{R}^{k \times 2}$ – матрица коэффициентов вида a_1, a_2 , а $U_i \in \mathbb{R}^k$ – вектор из элементов вида u_{lj} .

Заметим, что оценивание расположения двух разных сенсоров производится независимо, т.е. возможно ускорение алгоритма путем параллелизации вычислений.

Обозначим через $\tau_i = \{(j, l)\}$ подмножество множества сенсоров, состоящее из пробных пар (j, l) , выбранных для сенсора i

так, как описано выше, т.е. случайно равномерно на множестве $\{1, \dots, n\} \setminus i$. На этом множестве введем функционал среднего риска (качества) для сенсора i :

$$(8) \quad F_i(x_i) = \mathbb{E} \|A_i x_i - U_i\|^2,$$

где \mathbb{E} – значок математического ожидания, которое берется по всем парам и всем помехам.

Рассмотрим случайную выборку $\hat{\tau}_i \subset \tau_i$ и составим эмпирический функционал:

$$(9) \quad \hat{F}_i(x_i) = \frac{1}{|\hat{\tau}_i|} \sum_{(j,l) \in \hat{\tau}_i} \|A_i x_i - U_i\|^2,$$

где $|\hat{\tau}_i|$ – мощность множества $\hat{\tau}_i$, усреднение производится по выбранному набору пар и по реализовавшимся помехам. Результат минимизации функционала (9) по переменным x_{i1}, x_{i2} принимаем за оценку координат сенсора i на шаге $(s + 1)$:

$$x_i^{s+1} = \operatorname{argmin} \hat{F}_i(x_i).$$

Величина k – количество пар пробных сенсоров – является параметром алгоритма, и при численном тестировании будут использоваться ее различные значения, а описанную выше процедуру оценивания координат будем называть триангуляцией.

Суммирование в (9) происходит по элементам $\hat{\tau}_i$, которые по определению независимы, случайно и одинаково распределены. Покажем, что для любого набора $\{(j, l)\}$ выражения $\|A_i x_i - U_i\|^2$ ограничены.

Обозначим через $f_t(x_i) = \|A_i x_i - U_i\|^2$ невязку для какой-то одной пробной пары сенсоров. Пусть x_i^* минимизирует f_t и $f_t^* = f_t(x_i^*)$. Тогда $f_t(\vec{0}) \geq f_t^*$, где $\vec{0}$ – нулевой вектор. Так как $f_t(\vec{0}) = \|U_i\|^2 < \infty$. Таким образом, величина f_t может быть ограничена сверху квадратом нормы правой части уравнения (7).

Тогда в соответствии с законом больших чисел имеем

$$\lim_{|\hat{\tau}_i| \rightarrow \infty} \hat{F}_i \rightarrow F_i$$

с вероятностью 1 и в среднеквадратичном.

3.2. Оценивание задержек

Для оценивания задержек зафиксируем координаты сенсоров x и воспользуемся уравнением (3):

$$(10) \quad (x_{i1} - x_{j1})^2 + (x_{i2} - x_{j2})^2 = (ct_{ij})^2.$$

Зная оценки x^{s+1} координат сенсоров и используя выражение (1), получаем:

$$(11) \quad t_{ij}^{s+1} = \frac{1}{c} \sqrt{(x_{i1}^{s+1} - x_{j1}^{s+1})^2 + (x_{i2}^{s+1} - x_{j2}^{s+1})^2}.$$

Просуммируем уравнения (1) по i для каждого j , получим n линейных уравнений относительно $2n$ переменных e_i, r_i :

$$(12) \quad \sum_{i=1}^n e_i + nr_j = \sum_{i=1}^n (m_{ij} - t_{ij}^{s+1}); \quad j = 1, \dots, n.$$

Аналогично, просуммировав (1) по j для каждого i , получим еще n линейных уравнений относительно переменных e_i, r_i :

$$(13) \quad ne_i + \sum_{j=1}^n r_j = \sum_{j=1}^n (m_{ij} - t_{ij}^{s+1}); \quad i = 1, \dots, n.$$

Таким образом, получаем $2n$ линейных уравнений относительно $2n$ переменных e_i, r_i . Удобным свойством этой системы уравнений является то, что коэффициенты в левой части фиксированы при любой оценке правой части. Это означает, что она останется неизменной на всех итерациях алгоритма, что позволяет вычислить её единожды. Для решения системы линейных уравнений потребуется лишь вычисление правых частей и умножение на заранее полученную матрицу, что значительно ускоряет процесс.

Перепишем систему уравнений в следующем виде:

$$(14) \quad B \begin{pmatrix} e \\ r \end{pmatrix} = v^{s+1}.$$

Здесь

$$B = \begin{pmatrix} \mathbf{1}_{n \times n} & nI_{n \times n} \\ nI_{n \times n} & \mathbf{1}_{n \times n} \end{pmatrix} \in \mathbb{R}^{2n \times 2n}$$

– матрица коэффициентов в левых частях уравнений (12), (13), а через $\mathbf{1}_{n \times n} \in \mathbb{R}^{n \times n}$ и $I_{n \times n} \in \mathbb{R}^{n \times n}$ обозначены матрица из единиц и единичная матрица соответственно; $e = (e_1, \dots, e_n)^\top$

и $r = (r_1, \dots, r_n)^\top$ – переменные, а $v^{s+1} \in \mathbb{R}^{2n}$ – вектор правых частей уравнений (12, 13).

Значения переменных e, r , удовлетворяющие соотношению (14) (т.е. уравнениям (12), (13)), естественно принять за новые оценки e^{s+1} и r^{s+1} запаздываний. Нетрудно видеть, что матрица B вырождена: нулевому собственному значению отвечает нормализованный собственный вектор вида

$$p = t/\|t\|; \quad t = \underbrace{(1, \dots, 1)}_n, \underbrace{(-1, \dots, -1)}_n)^\top,$$

поэтому за оценку запаздываний, получаемую на шаге $(s + 1)$ принимаем

$$\begin{pmatrix} e^{s+1} \\ r^{s+1} \end{pmatrix} = B^+ v^{s+1},$$

где B^+ – матрица, псевдообратная к B .

4. Моделирование

4.1. Тестовые данные

Для численного моделирования была разработана система генерирования тестовых данных, которая позволяет моделировать неточные (неидеальные, не равномерные на окружности) координаты датчиков с любой заранее известной помехой. Также генерируются задержки передачи и приема ультразвукового сигнала. В типичном ультразвуковом томографе используется четырехсекторное кольцо сенсоров, расстояние между секторами заранее определено, что учитывается при моделировании.

Результатом работы системы моделирования являются: x_{ij} – «неидеальные» координаты сенсоров, r_i, e_i – дискретные задержки приема и передачи сенсоров с известным шумом.

Далее эти данные используются для апробации алгоритмов и проверки качества его работы по выбранным метрикам.

Ниже приведен пример генерирования координат $n = 32$ сенсоров, разбитых на четыре сектора на кольце. Такое количество сенсоров выбрано для наглядности рисунка, все чис-

ленные эксперименты были проведены для $n = 2048$ сенсоров. Расстояния между секторами полагались равными двум расстояниям между соседними датчиками одного сектора. Радиус кольца принят равным единице. Идеальные координаты сенсоров возмущены непрерывно равномерно распределенным шумом $\mathcal{U}[-0.01, 0.01]$ (см. рис. 2).

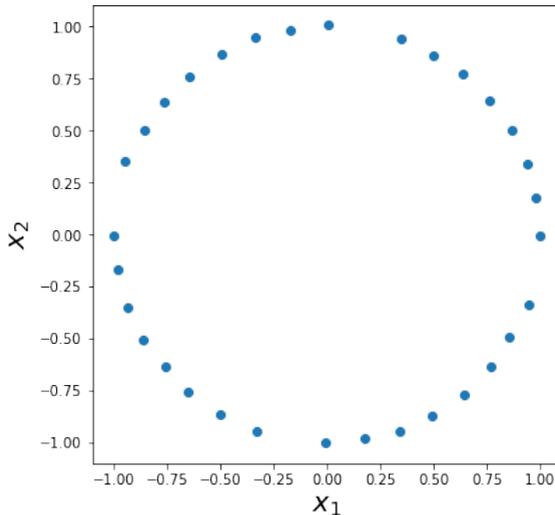


Рис. 2. Четыре сектора координат сенсоров с равномерно распределенной помехой

4.2. Апробация алгоритма

Для рассматриваемой задачи была проведена апробация алгоритма с учетом разных входных параметров: количества сенсоров, участвующих в алгоритме, а также разных типов шумов.

Для тестирования был сгенерирован набор из $n = 2048$ сенсоров. Для расположения сенсоров и для помех δ в экспериментах были рассмотрены два типа шума: равномерно распределенный на $[-0,01; 0,01]$ и усеченный гауссовский на $[-7, 7]$ с нулевым средним и дисперсией $\sigma = 2$. Для большей наглядности были приведены результаты в виде линейной комбинации оши-

бок для задержек и координат датчиков, которые были масштабированы для приведения к одним порядкам величин. Искусственно сгенерированные данные, использовавшиеся при тестировании, представляются близкими по природе к реальным данным. Так, равномерное распределение (ошибки в координатах сенсоров) с малой дисперсией и нулевым средним до определенной степени соответствует практике, характеризуя мелкие неточности производственного процесса изготовления прибора. Шум, добавляемый в задержку при передаче и приёме сигналов, был подобран на основе эмпирических данных, полученных с существующего оборудования.

В экспериментах рассматривались значения $k = 16, 32, 64, 128, 256, 512, 1024$; набор из $k = 1024$ пар отвечает использованию всех возможных пробных пар сенсоров.

5. Результаты

Для заданных параметров были получены результаты калибровки координат, представленные на графике (см. рис. 3) при равномерно распределенном шуме.

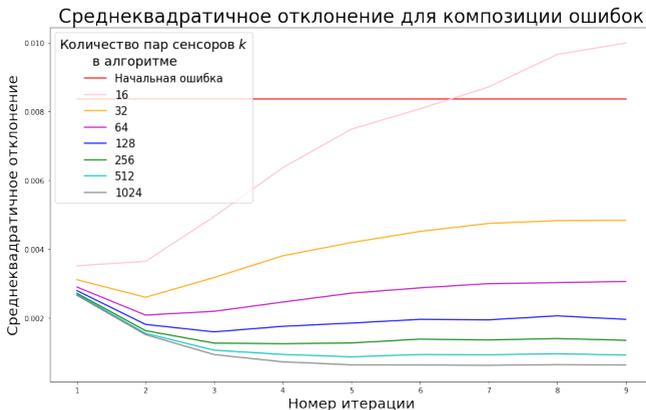


Рис. 3. Зависимость суммы ошибок от количества итераций алгоритма при равномерном распределении всех шумов

Как указано выше, для оценки ошибки было принято суммарное среднеквадратическое отклонение по задержкам e , r и координатам x . График отражает зависимость ошибки от количества итераций алгоритма.

Аналогичный график для нормального распределения представлен на графике (см. рис. 4).

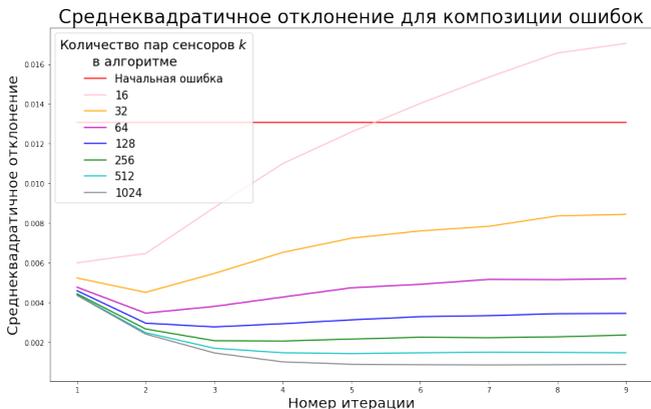


Рис. 4. Зависимость суммы ошибок от количества итераций алгоритма при нормальном распределении всех шумов

6. Выводы и направления дальнейших исследований

Алгоритм показал уменьшение ошибки местоположения датчиков и задержек относительно неоткалиброванных данных. Видно, что алгоритм зависит от выбора параметра k , при малых значениях которого оптимизационная задача расходится, – предположительно из-за накапливающейся ошибки при вычислениях с плавающей точкой.

Дальнейшее направление работы заключается в разработке гибридных алгоритмов калибровки, основанных на представленном в этой работе алгоритме. Планируется использование элементов подходов, описанных в работах [2, 3, 7], которые позволяют работать с накапливающейся ошибкой и, вероятно, при-

дадут новым синтезированным алгоритмам свойства робастности. Кроме того, упомянутые методы имеют ряд параметров, выбор которых для рассматриваемой задачи калибровки может быть критичен, и этот правильный выбор предполагается осуществить в дальнейших исследованиях. Применение теоретически обоснованных методов с доказанной сходимостью к конкретной задаче, имеющей практическое происхождение, представляет собой нетривиальную задачу. Наконец, предполагается апробация алгоритма на реальных данных, характер шумов в которых может быть отличным от принятых в настоящей работе.

Литература

1. ГРАНИЧИН О.Н., ЩЕРБАКОВ П.С. *Об одном подходе к решению задачи о калибровке большого массива ультразвуковых датчиков* // В Сб. материалов XIV Всероссийского совещания по проблемам управления (ВСПУ-2024), Россия, Москва, ИПУ РАН, 17–20 июня 2024. – 2024.
2. CAMPI M., GARATTI S. *Compression, generalization and learning* // Journal of Machine Learning Research. – 2023. – Vol. 24, No. 339. – P. 1–74.
3. CSÁJI B.C., CAMPI M.C., WEYER E. *Sign-perturbed sums: A new system identification approach for constructing exact non-asymptotic confidence regions in linear regression models* // IEEE Trans. on Signal Processing. – 2014. – Vol. 63, No. 1. – P. 169–181.
4. DURIC N., LITTRUP P., POULO L. et al. *Detection of breast cancer with ultrasound tomography: First results with the Computed Ultrasound Risk Evaluation (CURE) prototype* // Medical physics. – 2007. – Vol. 34, No. 2. – P. 773–785.
5. EROFEEVA V., GALYAMINA V., GRANICHIN O. et al. *Detection of specific areas and densities for ultrasound tomography* // Cybernetics and Physics. – 2019. – Vol. 8, No. 3. – P. 121–127.

6. FANG X., WU Y., SONG J. et al. *Zone-shrinking Fresnel zone travel-time tomography for sound speed reconstruction in breast USCT* // Sensors. – 2020. – Vol. 20, No. 19. – P. 5563.
7. GRANICHIN O., EROFEEVA V., IVANSKIY Y. et al. *Simultaneous perturbation stochastic approximation-based consensus for tracking under unknown-but-bounded disturbances* // IEEE Trans. on Automatic Control. – 2021. – Vol. 66, No. 8. – P. 3710–3717.
8. GRANICHIN O., SCHERBAKOV P., MING Y. *A Calibration algorithm for a large circular array of ultrasound transducers* // Proc. of the Int. Workshop on Medical Ultrasound Tomography (MUST), 10-12 June 2024, Amsterdam, Netherlands. – 2024.
9. GRÖHL J., SCHELLENBERG M., DREHER K. et al. *Deep learning for biomedical photoacoustic imaging: A review* // Photoacoustics. – 2021. – Vol. 22. – P. 100241.
10. JOVANOVIĆ I. *Inverse problems in acoustic tomography: theory and applications*. – Suisse: EPFL, 2008. – 139 c.
11. LENOX M.W., WISKIN J., LEWIS M. et al. *Imaging performance of quantitative transmission ultrasound* // Journal of Biomedical Imaging. – 2015. – Vol. 2015. – P. 773–785.
12. LIN L., ZHANG Y., HU P. et al. *High-speed three-dimensional photoacoustic computed tomography for preclinical research and clinical translation* // Nature communications. – 2021. – Vol. 12. – P. 882.
13. RAHAMAN J., PRAKASH R., RANJBARAN M. et al. *Transducer misplacement compensation for in-lab-made 3D photoacoustic tomography systems using nature-inspired algorithms* // Photons Plus Ultrasound: Imaging and Sensing. – 2023. – Vol. 12379. – P. 123790W.
14. RAJENDRAN P., PRAMANIK M. *Deep-learning-based multi-transducer photoacoustic tomography imaging without radius calibration* // Optics Letters. – 2023. – Vol. 46, No. 18. – P. 4510–4513.
15. ROY O., JOVANOVIĆ I., DURIC N. et al. *Robust array calibration using time delays with application to*

- ultrasound tomography* // Medical Imaging: Ultrasonic Imaging, Tomography, and Therapy. – 2011. – Vol. 7968. – P. 46–56.
16. RUITER N.V., ZAPF M., HOPP T. et al. *3D ultrasound computer tomography of the breast: A new era?* // European Journal of Radiology. – 2012. – Vol. 81. – P. S133–S134.
 17. SASTRY K., ZHANG Y., HU P. et al. *A method for the geometric calibration of ultrasound transducer arrays with arbitrary geometries* // Photoacoustics. – 2023. – Vol. 32. – P. 100520.
 18. SONG J., WANG S., ZHOU L. et al. *A prototype system for ultrasound computer tomography with ring array* // 2nd IET Int. Conf. on Biomedical Image and Signal Processing (ICBISP-2017). – 2017. – P. 1–4.
 19. SONG J., ZHANG Q., ZHOU L. et al. *Design and implementation of a modular and scalable research platform for ultrasound computed tomography* // IEEE Trans. on Ultrasonics, Ferroelectrics and Frequency Control. – 2021. – Vol. 69, No. 1. – P. 62–72.
 20. WANG S., ZENG L., SONG J. et al. *Variational mode decomposition for ultrasound computed tomography* // Medical Imaging: Ultrasonic Imaging and Tomography. – 2020. – Vol. 11319, No. 19. – P. 140–145.

CALIBRATION OF A LARGE ARRAY OF ULTRASOUND SENSORS

Oleg Granichin, Saint Petersburg State University, Saint Petersburg, Doctor of Science, professor (oleg_granichin@mail.ru).

Olga Granichina, Russian State Pedagogical University in the name of A. I. Herzen, Saint Petersburg, Doctor of Science, professor (olga_granichina@mail.ru).

Stepan Trofimov, Saint Petersburg State University, Saint Petersburg, Ph.D. student (steve.trofimov@gmail.com).

Pavel Shcherbakov, V.A. Trapeznikov Institute of Control Sciences,

RAS, Moscow; Moscow Institute of Physics and Technology,
Dolgoprudny, Doctor of Science, (cavour118@mail.ru).

Abstract: The paper presents a method for calibration of planar circular arrays of sensors used in ultrasound tomography, which enables the estimation of sensor coordinates and signal reception/transmission delays. The method uses an ultrasound wave propagation model and is based on the triangulation approach to identify the model parameters. At each iteration of the method, the estimates of coordinates and delays are recalculated separately from each other by solving systems of linear equations of low dimensionality. Simulation over synthetic data demonstrated high efficiency and accuracy of the approach: It is noise-resistant and is capable of operating in diverse conditions, including non-ideal model of ultrasound wave propagation. The proposed approach can be easily scaled to calibrate devices with large number of sensors, which makes it relevant to correct industrial practical implementation. This method can significantly improve the accuracy of ultrasound imaging results, making it a valuable tool for a variety of applications, such as medical diagnostics, robotics, etc.

Keywords: ultrasound tomography, planar circular array of sensors, identification of the model parameters, triangulation, large-scale systems.

УДК 519.7
ББК 22.18

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

*Поступила в редакцию 06.06.2024.
Дата опубликования 30.11.2024.*

МОДЕЛИРОВАНИЕ МОРСКОГО КОНТЕЙНЕРНОГО ТЕРМИНАЛА С ИСПОЛЬЗОВАНИЕМ СЕТИ МАССОВОГО ОБСЛУЖИВАНИЯ¹

Лемперт А. А.^{2а}, Жарков М. Л.^{3а},
Казаков А. Л.^{4аб}, Ву Х.З.^{5б}

(^а *Институт динамики систем и теории управления имени
В.М. Матросова СО РАН, Иркутск*)

(^б *Иркутский национальный исследовательский
технический университет, Иркутск*)

Статья посвящена проблеме моделирования и прогнозирования работы морского контейнерного терминала при учете неравномерности поступления грузов и влияния случайных факторов на продолжительность их обработки. Математическое описание системы строится в виде сети массового обслуживания (СеМО), узлы которой соответствуют структурным элементам терминала. Движение заявок между узлами происходит согласно заданной маршрутной матрице. Входящие транспортные потоки, прибывающие как с моря, так и с суши, моделируются с помощью пуассоновских потоков заявок, в том числе групповых. Построенная модель идентифицируется для морского контейнерного терминала, расположенного на севере Вьетнама и являющегося одним из крупнейших в Юго-Восточной Азии. Его математическая модель состоит из 14 узлов и включает два входящих потока, один из которых является неординарным. Маршрутная матрица строится на основе статистической обработки данных из открытых источников и натурных наблюдений, которые также используются для определения параметров входящих потоков. Предлагается и программно реализуется алгоритм имитации функционирования построенной СеМО. Выполняются сценарные расчеты для анализа текущей пропускной способности системы и прогноза ее работоспособности при увеличении объемов контейнерных потоков, в частности, при поступлении в терминал сверхбольших контейнеровозов.

Ключевые слова: математическое моделирование, теория массового обслуживания, имитационное моделирование, морской транспорт.

¹ Исследование выполнено за счет гранта Российского научного фонда, проект №24-21-00264.

² Анна Ананьевна Лемперт, к.ф.-м.н., в.н.с. (lempert@icc.ru).

³ Максим Леонидович Жарков, к.т.н., н.с. (zharkm@mail.ru).

⁴ Александр Леонидович Казаков, д.ф.-м.н., профессор, г.н.с. (kazakov@icc.ru).

⁵ Хоанг Занг Ву, аспирант (giangtrang14022016@gmail.com).

1. Введение

В настоящее время наблюдается быстрый рост контейнерных перевозок. Ожидается, что к 2029 году их рынок в стоимостном выражении достигнет 145,70 млрд долларов, что на 22% больше, чем в 2024 году [1]. До 80% подобных перевозок осуществляется морским транспортом, движение которого подвержено влиянию погодных и прочих труднопрогнозируемых условий. Вследствие увеличения грузопотоков и неравномерности их движения возникает потребность поиска эффективных стратегий управления портами, что подразумевает решение следующих основных задач: согласование подвода грузов и назначения причалов [2, 8], оценки допустимой загрузки системы [7], формирование парка перегрузочного оборудования [5], используемого при пиковых нагрузках. Последняя считается ключевой задачей при управлении портом [5, 7].

В настоящее время для описания логистических операций в морских контейнерных терминалах с учетом действия случайных факторов применяются сети Петри [5], вероятностно-статистические [7] и сетевые модели [27], а также теория массового обслуживания (ТМО) [16], которая еще с 80-х годов прошлого века считается одним из эффективных математических инструментов для решения проблем портовой логистики [17, 20].

Одной из первых значимых работ по применению ТМО для моделирования морских контейнерных терминалов, по-видимому, является [17]. В ней рассмотрена задача определения минимального необходимого количества причальных кранов, поэтому работа всего порта моделируется укрупненно, в виде одной системы массового обслуживания (СМО). При этом типовая структура морских терминалов включает три подсистемы и два маршрута движения контейнеров, поэтому, как показали дальнейшие исследования, более адекватным инструментом в данном случае являются сети массового обслуживания (СеМО), которые представляют собой совокупность конечного числа взаимосвязанных СМО [3, 10].

Пионерскими работами по применению СеМО в области портовой логистики являются [13] и [21], в которых моделиру-

ется перегрузка контейнеров со склада на судно и оптимизируется работа козловых кранов на складе. К настоящему времени опубликовано большое число работ по данной тематике, поэтому обсудим только наиболее релевантные из них. Так, в [24, 25] моделируется работа контейнерного склада, а в [18] – процесс перегрузки контейнеров с судна на железнодорожный транспорт. Статьи [15, 28, 29] посвящены применению нестационарных СеМО для учета суточных колебаний интенсивности поступления транспорта в порт. В работах [20, 22, 23] моделируются перевозки контейнеров на внутренних тягачах между причалом и складом, а в [26] – по всей территории морского терминала с учетом того, что число таких тягачей ограничено, при этом предполагается, что часть заявок (контейнеры) поступает из внешнего источника, а другая часть (тягачи) циркулирует внутри системы.

Отметим, что во всех рассмотренных выше работах моделируются отдельные подсистемы контейнерных терминалов или рассматривается только одно направление движения контейнеров. Однако для эффективного формирования парка перегрузочного оборудования в условиях роста контейнерных перевозок необходимо комплексное моделирование работы морских терминалов с учетом всех существующих подсистем, использования разных видов транспорта и наличия двух направлений движения материальных потоков внутри системы. Настоящее исследование посвящено решению этой актуальной проблемы.

У авторов есть опыт моделирования транспортных систем, имеющих в своем составе разные виды транспорта и подверженных влиянию случайных факторов. Так, в [6, 12] предложен подход к изучению работы наземных систем, базирующийся на использовании СеМО с несколькими неординарными потоками заявок, на основе которого построены модели мультимодального пересадочного узла [30], железнодорожной станции [11] и участков железнодорожной сети [4, 19]. В данном исследовании этот подход совершенствуется и применяется для сценарного анализа и прогнозирования работы принципиально нового объекта транспорта – морского контейнерного терминала.

Статья организована следующим образом. В разделе 2 дано описание объекта исследования. В разделе 3 представлены ис-

пользованный математический аппарат и модель морского терминала в виде СеМО. В разделе 4 описаны алгоритм имитационного моделирования ее функционирования и его программная реализация, с помощью которой в разделе 5 выполнен вычислительный эксперимент. На основе результатов расчетов дана оценка текущей пропускной способности системы и построен прогноз ее работы в будущем.

2. Описание объекта исследования

Морской контейнерный терминал – комплекс, специализирующийся на временном хранении, обработке и перевалке контейнеров. Соответственно, он имеет две основные задачи: обеспечение эффективной погрузки/разгрузки транспорта; сглаживание несоответствия между поступлением и отправлением контейнеров.

Крупные морские терминалы подразделяются на порты-хабы и узловые порты. Первые предназначены для перевалки грузов между магистральными и фидерными судами для дальнейшей перевозки. В узловых портах (гейтвейх) стыкуются магистральные морские и внутренние водные маршруты, а также сухопутный транспорт. Типовая их схема включает три основные подсистемы: грузовой фронт для наземного транспорта; контейнерную площадку; причал с береговыми кранами. Контейнеры движутся в двух направлениях: судно – контейнерная площадка – наземный транспорт и обратно.

В данной работе мы рассмотрим Международный контейнерный терминал Хайфон (далее Терминал), который является узловым глубоководным морским портом и имеет типовую структуру. Он расположен в северном экономическом регионе Вьетнама, обслуживает 17 регулярных международных маршрутов. Главные направления: США (7), Китай (5) и Индия (3). По числу перегружаемых контейнеров в год Терминал занимает первое место в северной части Вьетнама и четвертое место среди всех систем страны. Общий вид Терминала и карта-схема ее расположения представлены на рис. 1.

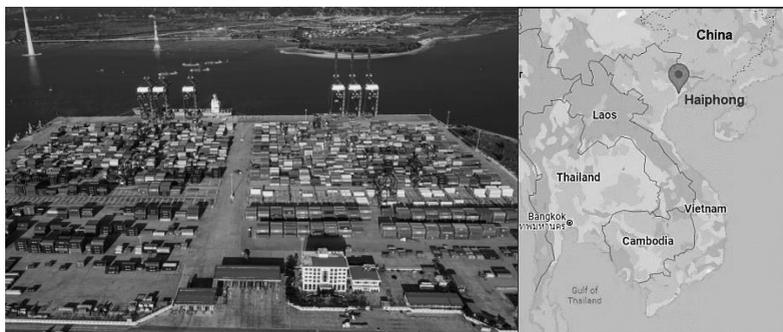


Рис. 1. Общий вид Терминала

2.1. КОНТЕЙНЕРНЫЕ ПОТОКИ

Терминал обрабатывает рефрижераторные, груженые и порожние контейнеры двух типов: 20-футовые и 40-футовые, которые прибывают на автомобильном и водном видах транспорта. Общепринятой единицей измерения их количества является TEU (twenty-foot equivalent unit), что соответствует одному 20-футовому контейнеру. В таблице 1 представлены статистические данные о количестве обработанных контейнеров в 2018–2023 гг., полученные из открытых источников (см. <https://hict.net.vn/thu-vien/Pages/thu-vien.aspx>).

Таблица 1. Количество обработанных контейнеров

| Год | 2018 | 2019 | 2020 | 2021 | 2022 | 2023 |
|----------|------|------|------|------|------|------|
| Тыс. TEU | 65 | 419 | 662 | 697 | 1181 | 1273 |
| Прирост | – | 545% | 58% | 5% | 70% | 7,8% |

В среднем в Терминал поступает 10,3 контейнеровозов в неделю, вмещающих от 950 до 13000 TEU, при этом средний размер партии контейнеров составляет 1280 TEU. Контейнеры также прибывают в Терминал на внешних тягачах с интенсивностью 872 тягача в сутки. Два указанных транспортных потока независимы и обрабатываются в отдельных подсистемах.

2.2. СТРУКТУРНЫЕ ЭЛЕМЕНТЫ ТЕРМИНАЛА

В Терминале имеются 9 основных структурных элементов, которые представлены на рис. 2.



Рис. 2. Функциональная схема

Целью моделирования является оценка пропускной способности объекта, поэтому опишем только те элементы, которые непосредственно участвуют в перевалке контейнеров.

Причалная зона (№ 1) включает два причала, на которых установлено по три береговых крана (№ 2). Средняя скорость погрузки/разгрузки каждого из них равна 30 контейнеров в час.

Контейнерная площадка (склад – № 3) вмещает 20160 TEU. На ней работают 24 козловых крана и два ричстакера, каждый из которых перегружает в среднем 50 контейнеров за час.

На пропускном пункте (ворота – № 8) действуют 9 въездных и 5 выездных полос, каждая из которых обслуживает в среднем 80 тягачей за час. Также перед воротами в терминал со стороны дороги имеется площадка, вмещающая до 9 тягачей.

Перевозка контейнеров между причалами и складом осуществляется с помощью 30 внутренних тягачей. Среднее время движения каждого составляет 5 минут, вместимость – 2 TEU. Перевозка контейнеров между воротами и складом выполняется внешними тягачами. На основе анализа спутниковых снимков установлено, что одновременно на территории может находиться

ся до 52 автомобилей. Их среднее время движения от ворот к складу составляет 4 минуты.

При описании Терминала необходимо также выделить якорную стоянку, которая находится в 34 км в Южно-Китайском море, так как именно там суда ожидают разрешения на швартовку. Время движения от стоянки до Терминала с учетом швартовки в среднем составляет 2,3 часа.

Таким образом, в Терминале имеется три крупные зоны: причальная зона, контейнерная площадка и ворота, а также якорная стоянка. Контейнеры доставляются двумя видами транспорта, которые независимы и обслуживаются в различных подсистемах.

3. Математическая модель

В данном разделе приводится структура модели морского контейнерного терминала и математический аппарат теории СеМО, необходимый для ее построения. Затем модель идентифицируется для выбранного транспортного объекта.

3.1. МАТЕМАТИЧЕСКИЙ АППАРАТ И СТРУКТУРНАЯ ИДЕНТИФИКАЦИЯ МОДЕЛИ

Математическая модель работы морского контейнерного терминала строится в виде открытой сети массового обслуживания в два этапа. На первом описываются входящие потоки контейнеров, на втором – процесс их обслуживания в подсистемах и маршруты движения отдельных контейнеров внутри системы.

Первый этап. При описании технических систем с помощью ТМО моделью прибывающего материального потока является поток заявок, поступающих из внешнего мира и требующих обслуживания [3].

Для описания поступления внешних тягачей и контейнеровозов предлагается использовать простейшие потоки [23, 26], поскольку данные транспортные средства движутся независимо друг от друга (ординарность и отсутствие последействия), и на больших промежутках времени неравномерностью потоков можно пренебречь (стационарность). Как правило, тягачи привозят один контейнер за раз, поэтому порождаемый ими поток

контейнеров также является простейшим. Суда перевозят контейнеры партиями, поэтому здесь поток поступающих контейнеров моделируется неординарным пуассоновским потоком [16, 25]. Для идентификации параметров входящих потоков заявок используются результаты анализа статистических данных.

Второй этап. СеМО состоит из конечного множества S , содержащего $s \in \mathbb{N}$, $s \geq 2$, узлов [3, 10]. Узел $y \in S$ – система массового обслуживания, включающая $1 \leq n_y < \infty$ однородных каналов и $0 \leq m_y < \infty$ мест в очереди [3]. В узел поступает входящий поток заявок A_y . Время их обслуживания в каналах является случайной величиной с законом распределения B_y . В открытой системе внешний источник, как правило, считается дополнительным (фиктивным) узлом.

Прибывшие в морской терминал контейнеры проходят несколько этапов обработки в структурных элементах системы. Они выполняют различные функции, в частности разгрузку, погрузку, перевозку и хранение контейнеров. Тем не менее эти элементы имеют общие признаки: наличие технических устройств, обрабатывающих поток контейнеров, регулярное повторение однотипных операций, недетерминированное время их выполнения, наличие очереди перед устройствами. Можно видеть, что эти признаки удовлетворяют требованиям однофазной СМО, а так как элементы терминала взаимосвязаны, то они соответствуют узлам СеМО. Контейнерные потоки с суши и моря прибывают в разные подсистемы морского контейнерного терминала, поэтому в структуре СеМО будут присутствовать два и более фиктивных узла. Число узлов СеМО и фиктивных узлов, а также их параметры определяются на основе натурального обследования конкретного объекта.

С помощью символики Кендалла – Башарина [3, 4] узел представляется в следующем виде:

$$(1) A_y / B_y / n_y / m_y.$$

В данной нотации для указания конкретных потоков используются следующие обозначения: M – стационарный пуассоновский поток заявок, G – произвольный стационарный поток; * – закон распределения промежутков времени между последовательным поступлением двух заявок неизвестен. Для неординар-

ных потоков применяются обозначения M^X и G^X , где X – закон распределения числа заявок в поступающей группе.

При групповом поступлении заявок используются различные дисциплины их принятия в узлы. Наиболее часто при описании транспортных систем [16, 19] применяется дисциплина «полный отказ»: если не хватает места хотя бы для одной заявки из группы, то теряется вся группа заявок. Для рассматриваемого объекта эта дисциплина означает, что суда и тягачи, получив отказ, уходят на внешнюю стоянку и ожидают там освобождения достаточного места в системе. Затем они вновь поступают в нее и рассматриваются как новая группа заявок на обслуживание.

Также необходимо определить дисциплину обслуживания очереди (при наличии) – алгоритма постановки в очередь при поступлении новых заявок и сдвигов при изменении числа заявок, а также предоставления обслуживания. Для большинства транспортных систем естественной является FIFO (First In, First Out), которая заключается в том, что из очереди выбирается заявка с наибольшим временем ожидания [10].

Маршрут движения заявок между узлами хранят в матрице вероятностей переходов (маршрутной матрице) P размера $(s + 1) \times (s + 1)$, элементы которой:

- $P_{y;u}$, $y, u \in S$ – вероятность того, что после завершения обслуживания в узле y заявка перейдет в узел u ;
- $P_{0;u}$ – вероятность поступления заявки из источника в узел u , $P_{0;0} = 0$, $P_{0;u} \geq 0$;
- $P_{y;0}$ – вероятность того, что заявка покинет СеМО сразу после завершения обслуживания в узле y .

Элементы маршрутной матрицы удовлетворяют условию

$$(2) \quad \sum_{u=0}^s P_{y;u} = 1 \quad \forall y \in S.$$

В ситуации, когда свободных мест в узле y не остается, а на вход продолжают поступать заявки из других узлов, применяются блокировки работы каналов. Это означает, что заявка остается в канале предыдущего узла, временно блокируя его работу, и ожидает освобождения места в узле y .

Целью моделирования является нахождение показателей эффективности СеМО, на основе которых можно сделать вывод о качестве функционирования моделируемой транспортной системы и построить прогноз ее работы при изменении объемов транспортных потоков в будущем.

Под состоянием СМО понимается число заявок, которые находятся в ней. Поскольку СеМО состоит из множества СМО, то будем рассматривать многомерное пространство состояний (k_1, k_2, \dots, k_s) , где $k_y = 0, 1, \dots, N_y$ – число заявок в узле y .

Определение показателей эффективности требует вычисления стационарных вероятностей состояний СеМО. Обозначим через $\pi(k_1, k_2, \dots, k_s)$ стационарную вероятность состояния (k_1, k_2, \dots, k_s) , тогда вероятности $\pi_y(k)$ того, что узел y содержит ровно $k_y = k$ заявок, вычисляются как

$$(3) \quad \pi_y(k) = \sum_{\substack{k_y=0, \\ i=1,2,\dots,s, i \neq y}}^{N_i} \{ \pi(k_1, k_2, \dots, k_s) | k_y = k \}$$

при условии нормировки $\sum_{\substack{k_y=0, \\ i=1,2,\dots,s}}^{N_i} \pi(k_1, k_2, \dots, k_s) = 1$.

На основе стационарных вероятностей (3) рассчитываются показатели эффективности [10]:

- вероятность P_L отказа в обслуживании заявки;
- среднее время T_S пребывания заявки в СеМО и отдельно \bar{t}_y в узле y ;
- суммарная продолжительность блокировки каналов \bar{b}_y ;
- среднее количество \bar{V}_y заявок в узле y :

$$\bar{V}_y = \sum_{k=1}^{N_y} k \pi_y(k);$$

- среднее число \bar{K}_y работающих каналов в узле y :

$$\bar{K}_y = \sum_{k=1}^{m_y-1} k \pi_y(k) + \sum_{k=m_y}^{N_y} m_y \pi_y(k);$$

- средняя длина \bar{L}_y очереди в узле y :

$$\bar{L}_y = \sum_{k=m_y+1}^{N_y} (k - m_y) \pi_y(k).$$

Найти стационарные вероятности аналитическими методами возможно лишь для наиболее простых типов СеМО (см. [3, 10]). В настоящее время общепринятой практикой для определения показателей эффективности СеМО является имитационное моделирование [14, 23, 25], которое дает возможность не только построить наглядные модели функционирования сложных систем с любой степенью детализации, но и реализовать различные виды вычислительных алгоритмов для обработки и анализа данных.

3.2. ПОСТРОЕНИЕ МОДЕЛИ КОНТЕЙНЕРНЫХ ПОТОКОВ

Тягачи перевозят два 20-футовых или один 40-футовый контейнер. При этом 75% контейнеров на судах являются 40-футовыми. Примем, что *одной заявкой* на обслуживание является 40-футовый контейнер. Их типы не учитываются. Таким образом, внешний тягач считается одной заявкой, а контейнеровоз – группой заявок (см. выше).

Контейнерный поток, прибывающий на морском транспорте, моделируется неординарным пуассоновским потоком M_1 . На основе результатов анализа статистических данных (см. <https://hict.net.vn/thu-vien/Pages/thu-vien.aspx>) установлено, что размеры групп контейнеров подчиняются дискретному равномерному закону $U(a_{min}, a_{max})$. Вероятность появления группы размера i определяется по формуле

$$p_i = \frac{1}{a_{max} - a_{min} + 1},$$

где a_{min} , a_{max} соответствуют наименьшей и наибольшей наблюдаемым партиям контейнеров.

Контейнерный поток, поступающий на наземном транспорте, моделируется простейшим потоком M_2 . Параметры входящих потоков заявок представлены в таблице 2. В ней: λ_i – интенсивность поступления групп заявок в час, V_i – закон распределение их размеров, $i = 1, 2$.

Таблица 2. Модели входящих контейнерных потоков

| Поток | Вид транспорта | λ_i | V_i |
|-------|----------------|-------------|-------------|
| M_1 | Контейнеровоз | 0,057 | U(320; 960) |
| M_2 | Внешний тягач | 36,32 | 1 |

3.3. МАТЕМАТИЧЕСКОЕ ОПИСАНИЕ РАБОТЫ СИСТЕМЫ

Работа элементов Терминала моделируется 14 узлами разных типов. Для их обозначения перед порядковым номером узла ставится один из следующих символов:

- O – источник входящего поток;
- a – многоканальная СМО с отказами и групповым обслуживанием заявок,
- b – многоканальная СМО с очередью,
- c – многоканальная СМО с отказами,
- d – одноканальная СМО с изменяющейся числом мест в очереди и групповым обслуживанием заявок.

Узлы O_0 и O_{13} – фиктивные узлы, источники входящих потоков M_1 и M_2 соответственно. Узел a_1 описывает работу якорной стоянки, в котором размер обслуживаемых групп соответствует максимальному числу прибывающих контейнеров на судах, а количество каналов равно среднему числу контейнеровозов, которые прибывают в систему за неделю.

При моделировании причальной зоны предполагается, что один (первый) причал предназначен для разгрузки контейнеровозов, а второй – для загрузки. Работа первого причала представляется узлом b_2 , в котором каналы описывают работу береговых кранов, а число мест в очереди соответствует максимальному размеру прибывающих партий контейнеров. Второй причал моделируется узлом d_3 . В нем канал – само судно, а число мест Y в очереди меняется динамически и становится равной размеру следующей обслуживаемой группе заявок Y в канале. СМО с такими характеристиками согласно [9] относится к классу систем, функционирующих в случайной среде.

Работа контейнерной площадки моделируется двумя узлами c_7 и b_8 , из которых первый отвечает за погрузку контейнеров на склад, а второй – за их хранение и выгрузку. Очередь узла b_8

соответствует вместимости склада. Аналогичным образом описывается работа ворот – узлы b_{11} и c_{12} .

Далее в таблице 3 представлено формальное описание этих узлов с помощью символики Кендалла – Башарина (1); T_y – закон распределения времени обслуживания, в часах; X_y – закон распределения размера обслуживаемой группы заявок в канале узла y .

Таблица 3. Описание структурных элементов Терминала

| Узел | Элемент | Модель | T_y | X_y |
|---------------|-----------------------------|----------------|----------------|---------------|
| a_1 | Якорная стоянка | $M_1/G^X/10/0$ | $N(2,3; 1,67)$ | 960 |
| b_2 | Причал для разгрузки судов | $*/M/3/960$ | $\exp(30)$ | 1 |
| d_3 | Причал для загрузки судов | $*/M^X/1/X$ | $\exp(0,057)$ | $U(420; 960)$ |
| c_4 | Береговые краны | $*/M/3/0$ | $\exp(30)$ | 1 |
| c_5, c_6 | Внутренние тягачи | $*/M/15/0$ | $\exp(12)$ | 1 |
| c_7 | Козловые краны / ричстакеры | $*/M/13/0$ | $\exp(50)$ | 1 |
| b_8 | Склад | $*/M/13/20160$ | $\exp(50)$ | 1 |
| c_9, c_{10} | Внешние тягачи | $*/M/26/0$ | $\exp(15)$ | 1 |
| b_{11} | Въездные ворота | $M_2/M/9/9$ | $\exp(80)$ | 1 |
| c_{12} | Выездные ворота | $*/M/5/0$ | $\exp(80)$ | 1 |

Здесь $N(\mu; \sigma)$ – нормальное распределение, функция плотности вероятностей которого имеет вид

$$f_{\mu,\sigma}(t) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(t-\mu)^2}{2\sigma^2}};$$

$\exp(\lambda)$ – экспоненциальный закон распределения с функцией плотности вероятностей

$$f_{\lambda}(t) = \begin{cases} \lambda e^{-\lambda t}, & t \geq 0, \\ 0, & t < 0. \end{cases}$$

Маршрутная матрица P имеет следующий вид:

| | O_0 | a_1 | b_2 | d_3 | c_4 | c_5 | c_6 | c_7 | b_8 | c_9 | c_{10} | b_{11} | c_{12} | O_{13} |
|----------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|----------|----------|----------|----------|
| O_0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| a_1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| b_2 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| d_3 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| c_4 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| c_5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| c_6 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| c_7 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| b_8 | 0 | 0 | 0 | 0 | 0 | 0 | 0,5 | 0 | 0 | 0 | 0,5 | 0 | 0 | 0 |
| c_9 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| c_{10} | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| b_{11} | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| c_{12} | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| O_{13} | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |

Переход заявки из узла y в узел x возможен только при $p_{y,x} > 0$. Сумма элементов одной строки удовлетворяет условию нормировки (2). На рис. 3 представлена схема СеМО в виде ориентированного графа.

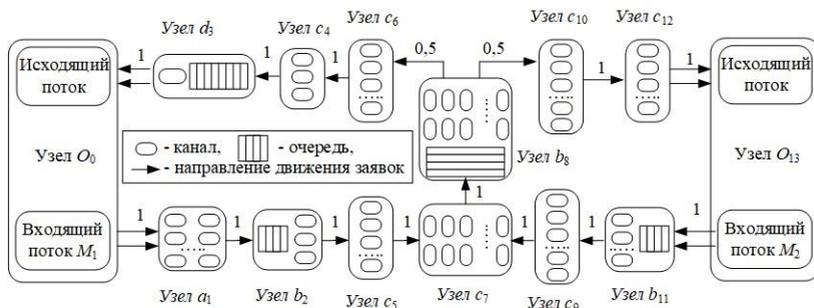


Рис. 3. Схема СеМО

Таким образом, математическая модель, описывающая работу Терминала, имеет вид СеМО с двумя входящими потоками и 14 узлами. Один входящий поток является простейшим и мо-

делирует поступление контейнеров на внешних тягачах, а второй – неординарным пуассоновским и моделирует прибытие контейнеров на судах. Узлы СеМО отличаются типом, числом каналов и вместимостью. Два из них являются фиктивными – источники потоков заявок, один – многоканальная СМО с отказами и групповым обслуживанием – отображает работу якорной стоянки, три – многоканальные с конечной очередью – функционирование причала для разгружаемых судов, склада и въездных ворот, семь – многоканальные СМО с отказами – работа причальных кранов, выездных ворот и движение внутренних и внешних тягачей внутри системы. Также имеется один узел – СМО с изменяющимся числом мест в очереди, который моделирует загрузку судна и его отправление из системы. Дисциплина принятия групп заявок – «полный отказ», дисциплина обслуживания очереди – FIFO.

4. Имитационная модель

Исследование построенной СеМО выполняется при помощи имитационной модели, которая основана на дискретно-событийном подходе к моделированию. Для ее программной реализации использовался язык программирования Object Pascal.

Пользователь задает в программе следующие параметры СеМО: количество входящих потоков и для каждого интенсивность поступления групп заявок и законы распределения их размеров; число узлов и для каждого длину очереди, число каналов, законы распределения размеров обслуживаемой группы заявок и времени их обслуживания, дисциплину принятия заявок и обслуживания очереди; маршрутную матрицу.

В программе отдельный входящий поток заявок имеет свой счетчик, который генерирует размеры групп заявок и промежутки времени между их поступлением по соответствующим параметрам. Единицей измерения модельного времени является один цикл T_c выполнения алгоритма имитации функционирования СеМО (см. рис. 4). Он обычно принимается равным одной минуте. Программа завершает работу после выполнения T_m циклов, число которых задает пользователь. Для получения ста-

ционных показателей эффективности с точностью $\pm 0,01$ необходимо, чтобы $T_m \geq 50000$. Результаты работы программы выводятся в табличном и графическом видах и могут быть экспортированы в таблицы Excel.



Рис. 4. Блок-схема алгоритма имитации функционирования СеМО

Алгоритм имитации функционирования СеМО имеет относительно низкую сложность, поэтому для его выполнения достаточно ресурсов персонального компьютера. Алгоритм выполняется параллельно в многопроцессорных системах. В этом случае реализующая его программа запускается отдельно на каждом процессоре. При этом для каждого процесса исполь-

зуется своя область оперативной памяти, в которую дублируются данные, внесенные пользователем.

Точность программы тестировалась на задачах трех типов с известными характеристиками: двух- и трехфазная СМО с блокировками и неординарным входящим потоком [30], открытая экспоненциальная СеМО с 6 узлами [14]. В результате сравнения расчетных показателей эффективности СМО и известных установлено, что максимальная относительная погрешность работы программы не превышает 4%. Поскольку в транспортных системах такая погрешность является вполне приемлемой, перейдем к вычислительным экспериментам.

5. Вычислительный эксперимент

С помощью программы выполнено три вычислительных эксперимента:

- Целью первого является анализ текущей загрузки Терминала.
- Во втором устанавливается максимальное число контейнеров, которое система способна обработать за год. Рассматриваются два наиболее показательных случая. В первом увеличиваются размеры партий контейнеров на судах, во втором – интенсивности поступления самих контейнеровозов.
- В третьем исследуется поведение Терминала при поступлении в него сверхбольших контейнеровозов.

Для каждого эксперимента проводилось по 10 пусков программы, виртуальное время каждого из которых – 365 дней. В начальный момент времени СеМО пуста. Усредненные результаты десяти пусков представлены далее в таблицах и на рисунках.

Дополнительно были проведены пуски в случае, когда в начальный момент времени в очереди Узла b_8 (склада) находилось $n_{b_8} / 2$ заявок. Оказалось, что полученные показатели эффективности незначительно отличаются от таковых для случая, когда в начальный момент времени система была пуста.

5.1. ЭКСПЕРИМЕНТ 1.

В таблице 4 приведены результаты численного исследования СеМО (см. п. 3). Далее V_y – общее число заявок, поступивших в узел y ; \bar{V}_y – среднее число заявок, находящиеся в узле в сутки; $K_y = \bar{K}_y/m_y$ – коэффициент занятости каналов; L_y , \bar{t}_y , \bar{b}_y и P_L означают то же, что и в п. 3.1.

Таблица 4. Результаты эксперимента 1

| | V_y | \bar{V}_y | K_y | \bar{L}_y | \bar{t}_y (ч) | \bar{b}_y (ч) |
|------------------|----------|-------------|--------|------------------|-----------------|-----------------|
| Узел a_1 | 308325,6 | 844,2 | 0,020 | – | 3,87 | 719,0 |
| Узел b_2 | 308148,6 | 836,2 | 0,406 | 198,73 | 5,67 | 61,6 |
| Узел d_3 | 306535,4 | 843,7 | 0,812 | 534,55 | 32,17 | – |
| Узел c_4 | 306532,4 | 839,8 | 0,841 | – | 0,07 | 11004,4 |
| Узел c_5 | 307933,0 | 843,6 | 0,201 | – | 0,08 | 226,3 |
| Узел c_6 | 306550,4 | 839,8 | 0,798 | – | 0,34 | 74696,6 |
| Узел c_7 | 620156,4 | 1699,1 | 0,113 | – | 0,02 | 252,0 |
| Узел b_8 | 620155,0 | 1678,8 | 0,768 | 3447,34 | 52,39 | 70659,7 |
| Узел c_9 | 312229,0 | 855,4 | 0,092 | – | 0,07 | 488,0 |
| Узел c_{10} | 306227,4 | 839,0 | 0,092 | – | 0,07 | 19,2 |
| Узел b_{11} | 312229,0 | 855,4 | 0,052 | 0 | 0,01 | 165,2 |
| Узел c_{12} | 306226,6 | 839,0 | 0,093 | – | 0,01 | – |
| Поступило | | M_1 | M_2 | Отклонено | M_1 | M_2 |
| Групп | | 487 | 312868 | Заявок | 0 | 639 |
| Заявок | | 308326 | 312868 | P_L | 0,00103 | |

Для проверки адекватности модели сравнивались следующие показатели: а) общее число прибывших $V_1 + V_{11}$ и отправленных $V_3 + V_{12}$ заявок из таблицы 4 с объемами импорта и экспорта контейнеров за 2023 год; б) соотношение прибывших V_1/V_{11} и отправленных V_3/V_{12} заявок с соответствующими соотношениями импортных и экспортных объемов контейнеров для разных видов транспорта. В результате наибольшая относительная погрешность составила 3,9%, а средняя – 2,3%.

Проанализируем представленные в таблице 5 результаты и интерпретируем их.

1. В среднем 639 (0,2%) заявок из потока M_2 получают отказ в Узле b_{11} (ворота). Однако он имеет низкую загрузку, так как $K_{11} = 0,052 < 1$. Поэтому заявки получают отказ из-за блокировки Узла c_9 , которая вызвана переполнением Узла b_8 (склад).

2. Средняя длина очереди в Узле b_8 равна $\bar{L}_8 = 3447,34$, что соответствует 17% от вместимости очереди. Следовательно, переполнение этого узла возникает только при пиковых нагрузках, которые появляются относительно редко.

3. Средняя продолжительность блокировки каналов Узла a_1 составляет $\bar{b}_1 / 487,2 = 1,48$ ч, что можно интерпретировать как среднее время ожидания на якорной стоянке.

Таким образом, Терминал эффективно справляется с нагрузкой, в частности, среднее время ожидания на якорной стоянке 1,48 ч и среднее время хранения контейнеров на складе $\bar{t}_8 = 52,39$ ч относительно малы. Наблюдаемое число непринятых контейнеров с внешних тягачей несущественно, так как на территории морского терминала возможно размещение до 1500 контейнеров вне склада.

5.2. ЭКСПЕРИМЕНТ 2

Случай 1. По сравнению с Экспериментом 1 увеличены на 10%, 20% и 30% следующие параметры модели:

- средний размер групп заявок из потока M_1 ;
- интенсивность поступления заявок из потока M_2 ;
- средний размер обслуживаемой группы заявок в Узле d_3 .

Изменение размера обслуживаемых групп в Узле d_3 происходит за счет того, что доставка и вывоз контейнеров в подавляющем большинстве случаев производится одними и теми же контейнеровозами. В таблице 5 и на рис. 5 представлены результаты моделирования.

Таблица 5. Результаты Эксперимента 2, случай 1

| № | $M_1 + M_2$ | V_L | P_L | T_S | \bar{L}_8 |
|---------|-------------|---------|--------|-------|-------------|
| 1. +10% | 692167,9 | 1358,7 | 0,0020 | 79,0 | 4554,6 |
| 2. +20% | 753467,9 | 6282,7 | 0,0083 | 108,1 | 7297,8 |
| 3. +30% | 831316,6 | 24959,0 | 0,0330 | 161,0 | 12220,6 |

Здесь и далее V_L – число отклоненных заявок.

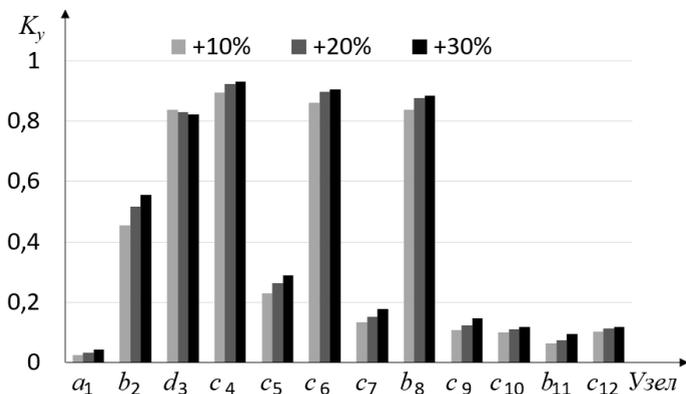


Рис. 5. Эксперимент 2, случай 1. Коэффициент использования каналов (K_y) в узлах

Случай 2. По сравнению с Экспериментом 1 интенсивности поступления заявок в обоих потоках и интенсивность обслуживания в Узле d_3 увеличены на 10%, 20% и 30%. Увеличение интенсивности обслуживания обусловлено ростом числа судов, прибывающих и, следовательно, отправляющихся из Терминала за год. Полученные результаты представлены в таблице 6 и на рис. 6.

Таблица 6. Результаты Эксперимента 2, случай 2

| № | $M_1 + M_2$ | V_L | P_L | T_S | \bar{L}_8 |
|---------|-------------|---------|--------|-------|-------------|
| 1. +10% | 695999,6 | 2224,9 | 0,0032 | 96,6 | 6035,1 |
| 2. +20% | 760171,6 | 7483,1 | 0,0099 | 107,1 | 7444,9 |
| 3. +30% | 826456,9 | 11567,1 | 0,0140 | 120,3 | 9361,8 |

Интерпретируем результаты Экспериментов 2. Из таблиц 5 и 6 видно, что при увеличении объемов потоков контейнеров на 30% (строки №3) система в обоих случаях перестает справляться с нагрузкой. Так, число непринятых контейнеров за год в среднем по двум экспериментам составляет 17552 (1,06%), что является критическим. При увеличении нагрузки на 20% в среднем 6883 (0,46%) контейнеров не будут обслужены. По мнению экспертов, в течение года такое число контейнеров возможно

разместить вне склада и обработать их в периоды низкой нагрузки на систему.

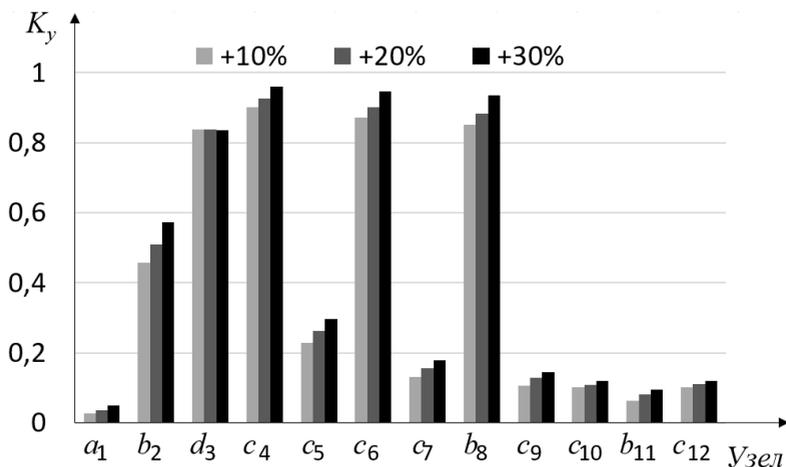


Рис. 6. Эксперимент 2, случай 2. Коэффициент использования каналов (K_y) в узлах

Таким образом, объемы перегружаемых контейнеров могут быть увеличены до 1,45–1,5 млн TEU в год, что на 20% больше показателя 2023 года. Однако это потребует четкой и бесперебойной работы Терминала, в частности, своевременного прибытия и отправления контейнеровозов.

Далее рассмотрим рис. 5 и 6. Наибольшее значение коэффициента занятости каналов (K_y) наблюдается в Узлах c_4 (береговые краны), c_6 и b_8 . При этом Узлы c_6 и b_8 имеют наибольшую продолжительность блокировки, которая учитывается при расчете K_y . Следовательно, наблюдаемые значения K_6 и K_8 обусловлены высокой загрузкой Узла c_4 . Поэтому ограничивающим элементом следует считать береговые краны и можно обоснованно предполагать, что их числа будет недостаточно при поступлении сверхбольших контейнеровозов в будущем.

Рассмотрим работу Терминала при увеличении числа береговых кранов до 8 и поступлении в него сверхбольших контейнеровозов, которые прибывают А) раз в полтора месяца (см. <https://hict.net.vn/en/about/Pages/history.aspx>) и Б) раз в месяц.

Примем, что такой контейнеровоз привозит в Терминал в среднем 5000 контейнеров [26].

5.3. ЭКСПЕРИМЕНТ 3

По сравнению с Экспериментом 1 были внесены следующие изменения в модель:

1. Добавлен новый поток заявок M_3 , который описывает поступление сверхбольших контейнеровозов. Размер групп заявок имеет распределение $U(4500; 5500)$, интенсивность поступления $\lambda_A = 0,0009$ и $\lambda_B = 0,0014$ групп заявок в час.

2. Число каналов в Узлах b_2 и c_4 увеличено до четырех. Также в Узле b_2 максимальная длина очереди составляет 5500 мест.

3. В Узле d_3 средний размер обслуживаемой группы заявок увеличен на 16% и соответствует среднему размеру прибывающих групп заявок из потоков M_1 и M_3 . Интенсивность обслуживания также увеличена на λ_A и λ_B . Эти изменения являются следствием того, что общее число отправляемых из Терминала судов возрастает из-за новых маршрутов сверхбольших контейнеровозов, которые при этом вмещают большее количество контейнеров, чем другие суда. Результаты моделирования представлены в таблице 7.

Таблица 7. Результаты Эксперимента 3

| № | $M_1 + M_2 + M_3$ | V_L | P_L | T_S | \bar{L}_8 |
|-----------------|-------------------|--------|--------|-------|-------------|
| 1. $+\lambda_A$ | 679155,7 | 1278,2 | 0,0019 | 68,8 | 3531,2 |
| 2. $+\lambda_B$ | 693532,2 | 3136,7 | 0,0045 | 73,7 | 3884,5 |

Можно видеть, что при поступлении одного сверхбольшого контейнеровоза в месяц (строка №2 таблицы 7) за год в среднем не будет принято 3136,7 (0,45%) контейнеров, которые поступают с внешних тягачей. Данное значение не оказывает существенного влияния на работу системы из-за возможности разместить контейнеры непосредственно на территории морского терминала и обработать их при освобождении ресурсов системы. Дальнейшее увеличение интенсивности прибытия таких контейнеровозов труднореализуемо, так как Терминал будет способен эффективно работать только при отсутствии пиковых нагрузок. В частности, при поступлении сверхбольших контей-

неровозов раз в три недели необходимо, чтобы склад был загружен не более чем на 40–45% непосредственно перед поступлением такого судна. Этого можно добиться за счет принудительного ожидания на якорной стоянке, что приведет и к увеличению финансовых потерь из-за простоя судов.

5.4. ВЫВОД О РАБОТЕ СИСТЕМЫ

Таким образом, Терминал в настоящее время имеет запас пропускной способности, который позволит перегрузить до 1,5 млн TEU в год. Однако для эффективной разгрузки и погрузки сверхбольших контейнеровозов необходима его модернизация, в первую очередь – увеличение числа береговых кранов с 6 до 8. Также потребуется увеличить и количество внутренних тягачей для обеспечения бесперебойной работы этих кранов, как минимум до 40 шт. [22, 23]. В результате система будет способна эффективно обрабатывать текущий транспортный поток, а также прибытие сверхбольших контейнеровозов в среднем раз в месяц.

6. Заключение

В статье предложена математическая модель морского контейнерного терминала, которая представляет собой сеть массового обслуживания с неординарными входящими потоками. За счет применения аппарата теории массового обслуживания достигается высокая адаптивность математического описания, что позволяет применить его для анализа текущего состояния и прогнозирования работы различных морских терминалов в относительно короткие сроки, а также оценить эффективность проектов по их модернизации.

Построенная модель идентифицирована для международного контейнерного терминала, который обладает типовой конфигурацией и является одним из крупнейших в Юго-Восточной Азии. Для определения показателей эффективности работы объекта предложен алгоритм имитации функционирования СеМО, реализованный в виде программы, которая не предполагает использования высокопроизводительного оборудования при вы-

полнении расчетов и может выполняться в параллельном режиме на многопроцессорных компьютерах.

На основе результатов численного моделирования удалось не только оценить текущую и максимальную допустимую загрузку системы, но и спрогнозировать работу морского терминала при поступлении в него сверхбольших контейнеровозов. В результате сформулированы обоснованные рекомендации по увеличению парка перегрузочного оборудования.

Дальнейшие исследования в данном направлении могут быть направлены на развитие предлагаемого подхода, в частности, за счет учета нестационарности поступления транспорта в систему. Также интересным видится адаптация и применение представленного инструментария для моделирования других типов морских терминалов: угольных, нефтеналивных и зерновых.

Литература

1. *Анализ размера и доли рынка контейнерных перевозок - тенденции роста и прогнозы (2024–2029 гг.)* [Электронный ресурс]. – Режим доступа: <https://www.mordorintelligence.com/ru/industry-reports/global-container-shipping-market>.
2. АЛЕКСАНДРОВ А.Э., ЯКУШЕВ Н.В. *Стохастическая постановка динамической транспортной задачи с задержками с учетом случайного разброса времени доставки и времени потребления* // Управление большими системами. – 2006. – №12–13. – С. 5–14.
3. БАШАРИН Г.П., БОЧАРОВ П.П., КОГАН Я.А. *Анализ очередей в вычислительных сетях. Теория и методы расчета*. – М.: Наука, 1989. – 336 с.
4. БЫЧКОВ И.В., КАЗАКОВ А.Л., ЖАРКОВ М.Л. *Формирование предложений по увеличению пропускной способности Улан-Баторской железной дороги с применением теории массового обслуживания* // Вычислительные технологии. – 2023. – Том 28, №6. – С. 17–36. – DOI: 10.25743/ICT.2023.28.6.003.
5. ЗУБ И.В., ЕЖОВ Ю.Е., СТЕНИН Н.Н. *Модель выбора портового перегрузочного оборудования на основе моделирова-*

- ния технологической линии порта // Вестник Государственного университета морского и речного флота имени адмирала С.О. Макарова. – 2020. – №12. – С. 1016–1028. – DOI: 10.21821/2309-5180-2020-12-6-1016-1028.
6. КАЗАКОВ А.Л., МАСЛОВ А.М. *Построение модели неравномерного транспортного потока на примере железнодорожной грузовой станции // Современные технологии. Системный анализ. Моделирование.* – 2009. – №3. – С. 27–32.
 7. КУПЦОВ Н.В., КУЗНЕЦОВ А.Л., ШАТИЛИН А.В. *Разработка модели вероятностной оценки пропускной способности морского грузового фронта экспортного угольного терминала // Вестник государственного университета морского и речного флота им. адмирала С.О. Макарова.* – 2020. – Т.12, №1. – С. 17–34. – DOI: 10.21821/2309-5180-2020-12-1-17-34.
 8. МОРОЗОВ Н.Ю., ГРИШИН Е.М., ПРАВДИВЕЦ Н.А. и др. *Оптимизация пунктов перевалки "морской порт – железная дорога" // Управление большими системами: сборник трудов.* – 2022. – №99. – С. 135–156. – DOI: 10.25728/ubs.2022.99.6.
 9. ПОЛИН Е.П. МОЙСЕЕВА С.П., МОЙСЕЕВ А.Н. *Применение отрицательного биномиального распределения для аппроксимации стационарного распределения числа заявок в СМО с входящим МАР-поток, интенсивность которого зависит от состояния системы // Управление большими системами: сборник трудов.* – 2024. – №108. – С. 40–56.
 10. BOLCH G., GREINER S., DE MEET H. et al. *Queueing Networks and Markov Chains: Modeling and Performance Evaluation with Computer Science Applications.* – NY.: John Wiley & Sons, 1998. – 726 p.
 11. BYCHKOV I., KAZAKOV A., LEMPERS A. et al. *Modeling of railway stations based on queuing networks // Applied Sciences.* – 2021. – Vol. 11(5). – P. 2425. – DOI: 10.3390/app11052425.
 12. BYCHKOV I.V., KAZAKOV A.L., LEMPERS A.A. et al. *An intelligent management system for the development of a regional transport logistics infrastructure // Automation and Remote*

- Control. – 2016. – Vol. 77. – P. 332–343. – DOI: 10.1134/S0005117916020090.
13. CANONACO P., LEGATO P., MAZZA R.M. et al. *A queuing network model for the management of berth crane operations* // Computers & Operations Research. – 2008. – Vol. 35. – P. 2432–2446. – DOI: 10.1016/j.cor.2006.12.001.
 14. CHEE-HOCK NG, BOON-HEE S. *Queueing Modelling Fundamentals With Applications in Communication Networks*. – Chichester: John Wiley & Sons, 2008. – 294 p.
 15. CHEN G., GOVINDAN K., GOLIAS M.M. *Reducing truck emissions at container terminals in a low carbon economy: proposal of a queueing-based bi-objective model for optimizing truck arrival pattern* // Transportation Research Part E: Logistics and Transportation Review. – 2013. – Vol. 55. – P. 3–22. – DOI: 10.1016/j.tre.2013.03.008.
 16. DRAGOVIĆ B., PARK N.-K., ZRNIĆ N.D. et al. *Mathematical Models of Multiserver Queueing System for Dynamic Performance Evaluation in Port* // Mathematical Problems in Engineering. – 2012. – Vol. 2012. – P. 19. – DOI: 10.1155/2012/710834.
 17. EDMOND E.D., MAGGS R.P. *How useful are queue models in port investment decisions for container berths?* // Journal of the Operational Research Society. – 1978. – Vol. 29, No. 8. – P. 741–750.
 18. BABELI K., HESS S., HESS M. *Capacity utilization of the container terminal as multiphase service system* // European Transport / Trasporti Europei. – 2022. – Vol. 86, No. 4. – P. 1–15.
 19. KAZAKOV A., LEMPERT A., ZHARKOV M. *An approach to railway network sections modeling based on queuing networks* // Journal of Rail Transport Planning & Management. – 2023. – Vol. 27. – P. 100404. – DOI: 10.1016/j.jrtpm.2023.100404.
 20. KOZAN E. *Comparison of analytical and simulation planning models of seaport container terminals* // Transportation Planning and Technology. – 1997. – Vol. 20. – P. 235–248. – DOI: 10.1080/03081069708717591.
 21. LEGATO P., CANONACO P., MAZZA R.M. *Yard Crane Management by Simulation and Optimisation* // Maritime Economics & Logistics. – 2009. – Vol. 11. – P. 36–57. – DOI: 10.1057/mel.2008.23.

22. LEGATO P., MAZZA R.M. *Queueing analysis for operations modeling in port logistics* // Maritime Business Review. – 2020. – Vol. 5. – P. 67–83. – DOI: 10.1108/MABR-09-2019-0035.
23. LEGATO P., MAZZA R.M. *Queueing networks for supporting container storage and retrieval* // Maritime Business Review. – 2023. – Vol. 8. – P. 301–317. – DOI: 10.1108/MABR-01-2023-0009.
24. RAJ G., ROY D., DE KOSTER R. et al. *Stochastic modeling of integrated order fulfillment processes with delivery time promise: Order picking, batching, and last-mile delivery* // European Journal of Operational Research. – 2024. – Vol. 316. – P. 1114–1128. – DOI: 10.1016/j.ejor.2024.03.003.
25. ROY D., DE KOSTER R. *Optimal Stack Layout Configurations at Automated Container Terminals Using Queueing Network Models* // In: Böse J.W. (eds) Handbook of Terminal Planning. Operations Research / Computer Science Interfaces Series. – Cham: Springer, 2020. – DOI: 10.1007/978-3-030-39990-0_19.
26. ROY D., VAN OMMEREN J.-K., DE KOSTER R. et al. *Modeling landside container terminal queues: Exact analysis and approximations* // Transportation Research Part B: Methodological. – 2022. – Vol. 162. – P. 73–102. – DOI: 10.1016/j.trb.2022.05.012.
27. ROŽIĆ T., IVANKOVIĆ B., BAJOR I. et al. *A Network-Based Model for Optimization of Container Location Assignment at Inland Terminals* // Appl. Sci. – 2022. – Vol. 12. – P. 5833. – DOI: 10.3390/app12125833.
28. ZHANG X., ZENG Q., CHEN W. *Optimization model for truck appointment in container terminals* // Procedia – Social and Behavioral Sciences. – 2013. – Vol. 96. – P. 1938–1947. – DOI: 10.1016/j.sbspro.2013.08.219.
29. ZHANG X., ZENG Q., YANG Z. *Optimization of truck appointments in container terminals* // Marit. Econ. Logist. – 2019. – Vol. 21. – P. 125–145. – DOI: 10.1057/s41278-018-0105-0.
30. ZHARKOV M.L., KAZAKOV A.L., LEMPERT A.A. *Transient process modeling in micrologistic transport systems* // IOP Conf. Series: Earth and Environmental Science. – 2021. – Vol. 629. – P. 012023. – DOI: 10.1088/1755-1315/629/1/012023.

MODELING OF A SEA CONTAINER TERMINAL USING A QUEUING NETWORK

Anna Lempert, Matrosov Institute for System Dynamics and Control Theory of Siberian Branch of RAS, Irkutsk, Cand.Sc. (lempert@icc.ru).

Maxim Zharkov, Matrosov Institute for System Dynamics and Control Theory of Siberian Branch of RAS, Irkutsk, Cand.Sc. (zharkm@mail.ru).

Alexander Kazakov, Matrosov Institute for System Dynamics and Control Theory of Siberian Branch of RAS, Irkutsk National Research Technical University, Irkutsk, Doctor of Science, professor (kazakov@icc.ru).

Giang Vu, Irkutsk National Research Technical University, Irkutsk, graduate student (giangtrang14022016@gmail.com).

Abstract: The article focuses on the problem of modeling and forecasting the operation of a sea container terminal under the unevenness of cargo arrival and the influence of random factors on the service duration. The mathematical model takes the form of a queuing network (QN), whose nodes simulate the operation of the terminal structural elements. The movement of requests between nodes occurs according to a given route matrix. To describe transport flows arriving from sea and land, we use both ordinary and batch Poisson flows of requests. The constructed model is identified for a sea container terminal in the north of Vietnam – one of the largest in Southeast Asia. The model consists of 14 nodes and includes two incoming Poisson flows, one of which is batch. The route matrix and the parameters of incoming flows are determined by statistical processing of data from open sources and field observations. The mathematical model is implemented, and scenario simulation is performed. Based on the results obtained, we conclude the current capacity of the system and forecast its performance with an increase in the volume of container flows, in particular, with the arrival of ultra-large container vessels in the future.

Keywords: mathematical modeling, queuing network, simulation model, sea transport.

УДК 519.872

ББК 22.18

Статья представлена к публикации членом редакционной коллегии М.И. Гераськиным.

Поступила в редакцию 11.07.2024.

Опубликована 30.11.2024.

ИМИТАТОР ВИЗУАЛЬНОЙ ОБСТАНОВКИ, СИНТЕЗИРУЮЩИЙ В ЗРИТЕЛЬНОМ ДИАПАЗОНЕ ЭЛЕКТРОМАГНИТНЫХ ВОЛН ВНЕШНИЙ ВИД ПРОСТРАНСТВА ВОКРУГ КАБИНЫ АВИАЦИОННОГО ТРЕНАЖЁРА¹

Роганов В. Р.², Короп Г. В.³, Михеев М.Ю.⁴
(ФГБУ ВО «Пензенский государственный
технологический университет», Пенза)

Авиационный тренажёр – это комплекс из нескольких имитаторов, объединённых с целью обучения лётчиков профессиональным навыкам пилотирования и самолётовождения. Так как моделировать все возможные ситуации при полёте на реальном летательном аппарате не представляется возможным, каждый авиационный тренажёр разрабатывается для формирования составляющих когнитивной модели лётчика с профессиональными навыками пилотирования летательного аппарата в определённых ситуациях, названных «учебными ситуациями». Из всех имитаторов авиационного тренажёра выделяется эргатический оптико-программно-технический имитатор, синтезирующий для лётчика визуально наблюдаемую 3D-модель внешней среды с качеством, достаточным для решения навигационных задач, включая использование глазомера для оценки расстояния до видимых 3D-моделей реперных объектов, из которых набирается 3D-модель заданного района полётов. Эта задача делится на ряд подзадач, включающих конструирование 3D-моделей реперных объектов, их описание и размещение в базе данных компьютерного генератора изображения, выбор 3D-индикатора, обучение лётчика видеть 3D-объекты в то время как он видит видеоряд из 3D-проекций 3D-моделей реперных объектов. В статье проведён анализ имитаторов визуальной обстановки и 3D-индикаторов с точки зрения формирования как во время реальных полётов на летательном аппарате, так и при полётах на авиационных тренажёрах у лётчика постоянно обновляющихся составляющих его когнитивной модели, позволяющих ему у спешно пилотировать летательный аппарат и решать навигационные задачи как «при полётах в сплошной облачности», так и при полёте с визуальным наблюдением реперных объектов с целью определения места нахождения модели летательного аппарата и направления её полёта.

¹ Часть статьи подготовлена при поддержке при поддержке гранта РФ №23-21-10046, <https://rscf.ru/project/23-21-10046/>.

² Владимир Робертович Роганов, к.т.н, доцент (vladimir_roganov@mail.ru).

³ Геннадий Викторович Короп, к.т.н, доцент (korop_g@mail.ru).

⁴ Михаил Юрьевич Михеев, д.т.н. профессор, зав. каф. (mix1959@gmail.com).

Ключевые слова: имитатор визуальной обстановки, составляющие когнитивной модели лётчика.

1. Введение

В настоящее время все лётчики совершенствуют свои профессиональные навыки, тренируясь на авиационных тренажёрах (АТ) [1]. Во время учебных полётов лётчик, сидя в кабине АТ, управляет моделью летательного аппарата, перемещаясь в 3D-модели внешней среды [2]. Эффект полёта как на реальном летательном аппарате (ЛА) возникает за счёт воздействия на m -го лётчика в каждый момент $t_j^{(S)}$ информационных потоков $I^{(S)}(t_j)$, поступающих к m -му лётчику от $K^{(S)}$ числа имитаторов. Совокупность имитаторов составляет комплекс технических средств конкретного АТ [3] и разрабатывается для обучения лётчика в учебных ситуациях из утверждённого Заказчиком списка u_n :

$$(1) \quad I^{(S)}(t_j^{(S)}, u_n) = \bigcup_{k=1}^{K^{(S)}} i_k^{(S)}(t_j^{(S)}, u_n).$$

Имитаторы АТ синтезируют информационные потоки $I^{(S)}(t_j, u_n)$, максимально идентичные информационным потокам $I^{(F)}(t_j, u_n)$, получаемым лётчиком во время полёта на реальном ЛА, в условиях, аналогичных условиям из списка u_n , что обеспечивается информацией поступающей к лётчику [4] от имитаторов, число которых $K^{(F)}$:

$$(2) \quad I^{(F)}(t_j^{(F)}, u_n) = \bigcup_{k=1}^{K^{(F)}} i_k^{(F)}(t_j^{(F)}, u_n),$$

$$(3) \quad K^{(F)} > K^{(S)}.$$

Так как создать полную модель невозможно [5], каждый k -й имитатор синтезирует в $t_j^{(S)}$ момент времени $i_k^{(S)}(t_j, u_n)$ с качеством, допускающим обучение лётчика в строго определённых n -х учебных ситуациях на интервале времени [6]:

$$(4) \quad t_j^{(S)} = t_j^{(F)} \pm \Delta t_k^{(S)},$$

когда

$$(5) \quad i_k^{(S)}(t_j^{(S)}, u_n) = F_k(t_j^{(S)}, u_n) \approx i_k^{(R)}(t_j^{(R)}, u_n).$$

Учебная ситуация определяется Заказчиком, и им же проверяется возможность обучения m -го лётчика на конкретном АТ во время полётов на нём лётчиков-испытателей [7]. Заказчик проверяет, что поступающая в $t_j^{(S)}$ момент времени информация $i_k^{(S)}(t_j)$, синтезированная k -м имитатором АТ успешно формирует у m -го лётчика заданные составляющие его когнитивной модели $C_m^{(S)}(i_k^{(S)})$, в которых откладывается необходимый для полётов на реальном ЛА опыт пилотирования летательного аппарата и решения задач самолётовождения [8].

2. Материалы и методы

Разработка и эксплуатация АТ показала, что при учебных полётах в заданных учебных ситуациях у m -го лётчика успешно формируются навыки пилотирования ЛА [9] и навыки решения задач самолётовождения [10], откладываемые в составляющих его когнитивной модели $C_m^{(S)}(i_k^{(S)})$, каждая из которых имеет два параметра [11]:

$$(6) \quad C_m^{(S)}(i_k^{(S)}) = C_m^{(S)'}(i_k^{(S)}) + C_m^{(S)''}(i_k^{(S)}),$$

где $C_m^{(S)}(i_k^{(S)})$ – параметр положительного опыта m -го лётчика, максимально приближенный к составляющей $C_m^{(F)}(i_k^{(F)})$, формирующейся у m -го лётчика во время реального полёта на ЛА, когда он получает $i_k^{(S)}(t_j^{(S)})$ [12]; $C_m^{(S)''}(i_k^{(S)})$ – параметр «ложного навыка» [13] у m -го лётчика, появление которого зависит от отклонений $i_k^{(S)}(t_j^{(S)})$ от $i_k^{(F)}(t_j^{(F)})$ и причиной появления которых являются ошибки при синтезе $I^{(S)}(t_j)$ программно-технической частью k -го имитатора АТ, синтезирующего $i_k^{(S)}$ [14], который описывает поведение конкретного узла ЛА, являющегося прототипом k -го имитатора АТ.

При разработке АТ (6) рассчитывается для каждого имитатора применительно для каждой ситуации, указанной Заказчиком как u_i -я «учебной ситуации» [7]. Если во время пусконаладочных работ фиксируется отклонение допустимого предела $i_k^{(S)}$ в любой момент $t_j^{(S)}$,

$$(7) \quad \left| i_k^{(S)}(t_j^{(S)}, u_i) - i_k^{(F)}(t_j^{(F)}, u_i) \right| \geq \text{const } k,$$

АТ не может применяться для обучения лётчика в данной ситуации.

Таким образом, задачей разработчиков любого k -го имитатора авиационного трена является формирование у m -го лётчика $C_m(i_m^{(S)})$ с учётом (8) при полётах в учебных ситуациях из списка $u_i \in u_n$. Общее число учебных ситуаций

$$(8) \quad U^{(S)} = \bigcup_{i=1}^{I_N^{(S)}} u_i,$$

где $I_N^{(S)}$ – максимально возможное число учебных ситуаций – является основной характеристикой конкретного АТ.

Большинство пилотажных и комплексных АТ разрабатывались с учётом первоочередного обучения m -го лётчика профессиональному пилотированию ЛА с формированием в u_i -й учебной ситуации составляющих его когнитивной модели $C_{m_k}^{(S)}(t_j^{(S)}, C_{m_k}(t_{j-1}), u_i)$ с учётом уже имеющихся у него в предыдущий момент времени $C_m^{(S)}(t_{j-1})$ [15], в которых откладывается его опыт решения поставленных задач как во время реальных полётов, так и при полётах на АТ [16]. Ориентация АТ на обучение лётчика пилотированию ЛА объясняется двумя причинами:

1) задачи обучения лётчика профессиональным навыкам пилотирования ЛА являются более актуальными с точки зрения повышения уровня безопасности полётов ЛА [17], и для их решения достаточно синтезировать в u_i -й учебной ситуации в режиме реального времени (РРВ) с циклом 120 мс на каждом k -м имитаторе приборов кабинного оборудования информацию

$$(9) \quad i_k^{(S)}(t_j^{(S)}) = F_k^{(S)}(D_k(t_j^{(S)}), u_i),$$

отражающую в момент времени $t_j^{(S)}$ изменение функции $F_k^{(S)}$, моделирующей поведение рассматриваемого узла ЛА $D_k(t_j^{(S)})$ в u_i -й учебной ситуации, что фиксируется в $C_m^{(S)}(i_k^{(S)})$ как навыки работы с конкретным прибором кабинного оборудования, зависящие от особенностей m -го лётчика и его общего времени налёта T_m в условиях, аналогичных u_i -й учебной ситуации:

$$(10) \quad T_m(u_i) = T_m^{(F)}(u_i) + T_m^{(S)}(u_i),$$

где $T_m^{(F)}(u_i)$ – время налёта m -го лётчика на реальном ЛА;
 $T_m^{(S)}(u_i)$ – время налёта m -го лётчика на реальном АТ.

$$(11) C_m(i_k(T_m(u_i))) \rightarrow \text{const } C_{m_k} \quad \text{при } T_m(u_i) \rightarrow \infty;$$

2) обучение m -го лётчика максимально возможным навыкам решения навигационных задач при полёте на АТ предполагает наличие профессиональных навыков ориентировки в пространстве, что составляет:

– умение $C_m(i_V(t_j))$ считывать информацию с панелей k -х приборов радиотехнических средств самолётовождения $i_k^{(S)}(t_j^{(S)})$ (авиационных радиокompасов (АРК), самолётной радиостанции ближней навигации (РСБНс) и других систем, ориентированных на работу с наземными радиомаяками), используя карту с нанесённым местом расположения наземных средств самолётовождения для определения места положения ЛА (при этом точность определения места нахождения ЛА зависит и от квалификации лётчика $C_m(i_V(t_j))$ и от погрешностей показаний радиотехнических средств самолётовождения [18]); это определяет формирование заданных составляющих его когнитивной модели с профессиональными навыками решения навигационных задач

$$(12) C_m i_k(T_m(u_i)) \rightarrow \text{const } C_{m_k} \quad \text{при } T_m(u_i) \rightarrow \infty;$$

что аналогично (11) и позволяет разрабатывать АТ, имитаторы которого синтезируют информационные потоки, позволяющие обучать лётчика и пилотированию ЛП и решению навигационных задач «при полётах в сплошной облачности», используя одинаковые подходы к разработке методов синтеза в РРВ информационных потоков от всех имитаторов о поведении основных узлов модели ЛА и о перемещении модели ЛА в модели пространства, что лётчик может наблюдать по показаниям имитаторов приборов кабинного оборудования (при этом допускается организация обратных связей между программно-техническими системами разных имитаторов для увеличения точности синтезируемой информации, и это не зависит особенностей подготовленности анализаторов m -го лётчика воспринимать необходимую информацию);

– умение $C_m(i_V(t_j))$ восстанавливать $t_j^{(S)}$ -й момент времени ориентировку в пространстве по визуально наблюдаемым $3D^{(R)}$ -объектам, которые лётчик во время подготовки к полёту определяет как «реперные» ($C_m(i_V(t_j))$), что также включает также профессиональное умение m -го лётчика определять расстояние до выбранного реперного объекта используя свой глазомер), в итоге для успешного формирования $C_m(i_V(t_j))$ необходимо сконструировать $3D$ -модель района полётов с достаточным числом визуально наблюдаемых $3D_o^{(R_M)}$ - моделей реперных o -объектов, каждый из которых при полётах на АТ должен распознаваться лётчиками как реальный реперный o -объект с вероятностью $P_o^{(3D_M)} \geq 98\%$ (98% – принятый в авиации предел, превышение которого считается обязательным для оценки полученных результатов методами математической статистики и вероятности), или условная вероятность визуального распознавания лётчиком наблюдаемой модели реперного o -объекта $3D_o^{(R)}$ как искомого o -объекта, выбранного для ориентирования на реальной местности $3D_o^{(R_N)}$, должна быть

$$(13) P_o^{(3D_M)} = P(3D_o^{(R_M)} \text{ модель} / 3D_o^{(R_N)} \text{ объект}) \geq 98\% .$$

Имитатор, синтезирующий в РРВ для лётчика визуально наблюдаемые им во время полёта $3D$ -модели реперных объектов с качеством, позволяющим лётчику распознавать из как реальные $3D$ -объекты, называется «Имитатор визуальной обстановки» (ИВО) (рис. 1). В его состав входит база данных с информацией о $3D$ -модели района полётов (размер которой может быть до 1500×1500 км), компьютерный генератор изображения (КГИ), синтезирующий в каждый цикл РРВ (80–120 мс) на экране $2D$ -проекции реперных объектов, попавших в камеру наблюдения, и $3D^{(V)}$ -индикатор, при просмотре через оптику которого лётчик считает, что он видит $3D$ -объект. Во время полёта на АТ на его вход поступает информация о положении в $3D$ -модели внешней среды подвижного наблюдателя и направления его взгляда (совпадает с продольной осью управляемой модели ЛА).



Рис. 1. Схема имитатора визуальной обстановки

Требования к основным узлам ИВО:

1) в базе данных должна храниться информация о всех 3D-моделях реперных объектов в виде специальных примитивов (тип примитивов зависит от программ компьютерной графики) с информацией, необходимой для синтеза на экране за каждый цикл РРВ 2D-проекции 3D-моделей реперных объектов, попавших в пирамиду видимости (пирамида видимости – условная пирамида без основания, с вершиной в точке расположения глаза подвижного наблюдателя и телесными углами, заданными как 40 градусов по горизонтали и 30 градусов по вертикали (рис. 2)); структура базы данных позволяет по запросу отсечь лишнюю информацию и за минимальное время извлечь все не-

обходимые примитивы, из которых набирается конструкции 3D-моделей реперных объектов, совокупность которых составляет 3D-модель района полётов;

2) многопроцессорный КГИ реализует один из возможных алгоритмов компьютерной графики, но всегда первый шаг – по полученным в начале цикла РРВ координатам подвижного наблюдателя и направления оси пирамиды видимости извлечь из базы данных КГИ всю информацию для синтеза на экране 2D-проекций 3D-моделей реперных объектов, попавших в пирамиду видимости (решение этой задачи предполагает конструирование 3D-модели района полётов из прилегающих друг к другу сегментов первого уровня с заданной дальностью видимости, что позволяет сразу отобразить все активные сегменты и уже учитывая их число распределять по ним 3D-модели реперных объектов так, чтобы не превысить заданный цикл РРВ, её частное и наиболее часто встречающееся решение – в центре модели района полётов иметь один сегмент (обычно размер 15×15 км) с 3D-моделью ВПП в центре и несколькими 3D-моделями реперных объектов, при вылете за пределы этого района полёт на АТ проводится в «сплошной облачности»: в этом случае можно обучать лётчиков навыкам пилотирования ЛА (11) и умению ориентироваться по приборам (12), но невозможно обучать большинству навыков $C_m(i\nu(t_j))$, определяющих умение решать навигационные задачи, предполагающие умение ориентироваться по визуально наблюдаемым реперным объектам);

3) эргатический $3D^{(V)}$ -индикатор, через который подготовленный человек рассматривает на экране видеоряд из 2D-проекций 3D-моделей реперных объектов в углах обзора 40 градусов по горизонтали и 30 градусов по вертикали, при этом оптика 3D-индикатора активизирует заданные составляющие зрительного аппарата человека, заставляя его видеть 3D-объект с качеством, позволяющим ему профессионально тренировать глазомер; его наличие позволяет улучшить формирование $C_m(i\nu(t_j))$.

ИВО, в отличие от остальных имитаторов АТ, формирует у лётчика составляющие его когнитивной модели, которые поз-

воляют лётчику визуально ориентироваться в визуально наблюдаемом пространстве и ориентированы на решение лётчиком навигационных задач $C_m(i_V(t_j))$, для этого необходима 3D-модель района полётов. ИВО является «тупиковой системой», не предполагающей наличие обратной связи с другими имитаторами АТ, поэтому борьба с возможными ошибками синтеза наблюдаемого изображения ведётся только на этапе разработки ИВО. Исследования показали, что ошибки синтеза визуально наблюдаемых 3D-моделей реперных объектов, из которых набирается 3D-модель района полётов, являются следствием шумов программно-технических систем КГИ и должны сводиться к минимуму с учётом включения в технологический процесс синтеза 3D-объектов особенностей зрительного аппарата конкретного человека.

Таким образом, качество 3D-объектов, визуально наблюдаемых лётчиком, синтезируемых ИВО и формирующих $C_m(i_V(t_j))$, зависит:

- от конструктивных особенностей $3D^{(V)}$ -индикатора;
- от выбранного алгоритма компьютерного синтеза на экране 2D-проекций 3D-моделей реперных объектов, реализованного программно-технической системой КГИ;
- от конструкции 3D-модели района полётов.

При исследовании ИВО с КГИ «Аксай» был проведён следующий опыт. Подвижный наблюдатель был расположен на расстоянии 200 км от центра мировой системы координат, при этом положение пирамиды видимости делило экран на верхнюю и нижнюю часть, что позволяло видеть срез 3D-модели поверхности района полётов. Ожидалось, что изображение будет статичным, однако из-за шумов, возникающих в процессе синтеза, изображение поверхности двигалось верх-вниз. Был проведён ещё один опыт: по траектории параллельно подстилающей поверхности по центру 3D-модели ВПП от её начала до конца должна был шагать подвижная 3D-модель робота. Однако после начала движения подвижный робот или поднимался вверх, или «проваливался» в 3D-модель ВПП, что было очень хорошо заметно при использовании $3D^{(V)}$ -индикатора «ОКУ». Таким образом, исследования ИВО показали, что необходим комплексный

подход к решению задачи синтеза визуально наблюдаемой 3D-модели внешней среды для лётчика и этот подход должен учитывать борьбу с возможными ошибками ИВО.

Создать полную визуально наблюдаемую 3D-модель пространства окружающую кабину АТ невозможно. Будем считать, что имеются ограничения, которые можно описать функционалом

$$(14) F^{(3D)}(x_1^{(3D)}, x_2^{(3D)}, x_3^{(3D)}, x_4^{(3D)}, x_5^{(3D)}, x_6^{(3D)}, x_7^{(3D)}),$$

где $x_1^{(3D)}$ – угол обзора 3D-модель пространства; $x_2^{(3D)}$ – минимальная дистанция между расположением наблюдателя и ближайшей видимой 3D-моделью; $x_3^{(3D)}$ – разрешающая способность изображения 3D-модели; $x_4^{(3D)}$ – внешний вид конструкции 3D-модели реперных объектов, включая заданный вариант их окраса, для выполнения (13); $x_5^{(3D)}$ – число одновременно наблюдаемых 3D-модели реперных объектов, обеспечивающих решение поставленных Заказчиком задач обучения лётчика; $x_6^{(3D)}$ – объём зрачка наблюдения (сколько человек могут одновременно наблюдать 3D-объекты: если для одного человека – это «узкий зрачок» (рис. 3); если могут наблюдать несколько человек – это «широкий зрачок» (рис. 4)); $x_7^{(3D)}$ – компактность конструкции 3D-индикатора.



Рис. 2. Одно из возможных положений пирамиды видимости

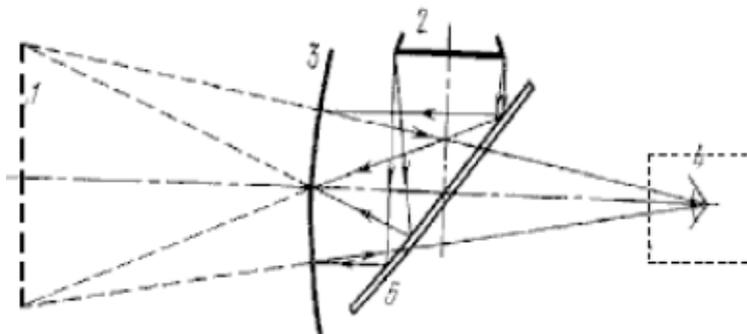


Рис. 3 Схема одноканального безочкового $3D^{(V)}$ -индикатора с «узким зрачком»: 1 – мнимый экран; 2 – телевизионный экран; 3 – вогнутое сферическое зеркало; 4 – зрачок; 5 – светоделительная пластина

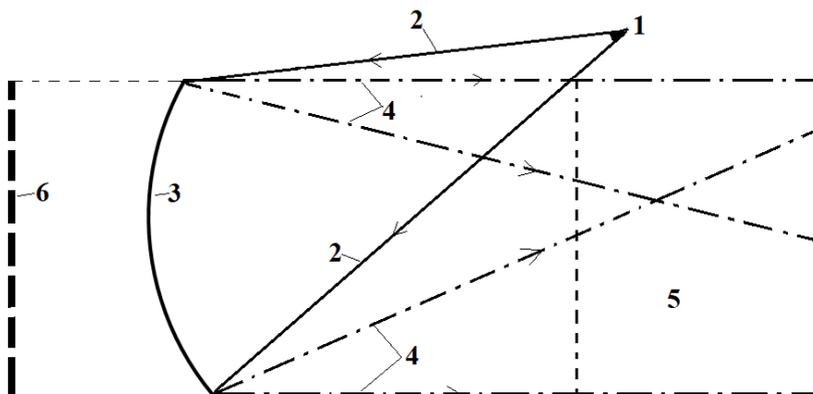


Рис. 4 Схема одноканального безочкового $3D^{(V)}$ -индикатора с «широким зрачком»: 1 – телепроектор; 2 – лучи от телепроектора; 3 – сферическое вогнутое зеркало большого радиуса; 4 – отражённые лучи; 5 – зрачок; 6 – мнимый экран

Параметры функционала (14) показывают, на что обращать внимание при выборе $3D^{(V)}$ -индикатора. При расчёте $x_1^{(3D)}$ для одноканального безочкового $3D^{(V)}$ -индикатора считается, что вершина пирамиды видимости находится между глаз лётчика (при разработке АТ одноместного ЛА) или между глаз левого

лётчика (при разработке АТ для ЛА управляемого двумя лётчиками, при этом правый лётчик видит 3D-модели окружающего пространства «глазами левого лётчика»). При расчёте $x_1^{(3D)}$ для двухканального $3D^{(V)}$ -индикатора с диспаратными очками считается, что вершина пирамиды видимости упирается в советующий глаз лётчика; $x_2^{(3D)}$ рассчитывается при разработке конструкции самого $3D^{(V)}$ -индикатора; $x_3^{(3D)}$ определяется техническими характеристиками выбранного экрана и $x_1^{(3D)}$; $x_4^{(3D)}$, $x_5^{(3D)}$, $x_6^{(3D)}$, $x_7^{(3D)}$ определяются конструктивными особенностями разработанного $3D^{(V)}$ -индикатора.

Исследования показали, что оптимальные углы обзора для одного канала – 40 градусов по горизонтали и 30 градусов по вертикали. В большинстве случаев используют составной многоканальный $3D^{(V)}$ -индикатор. Например, составной трёхканальный $3D^{(V)}$ -индикатор позволяет лётчику видеть 3D-модель окружающего пространства в углах обзора 120 градусов по горизонтали и 30 градусов по вертикали (рис. 5). Учитывая, что у большинства людей углы эффективного зрения – 30 градусов по горизонтали и 30 градусов по вертикали, углы обзора 3D-индикаторов ИВО позволяют лётчику достаточно эффективно решать задачи пилотирования летательного аппарата и самолётотовждения визуально рассматривая 3D-модель окружающего пространства.

В настоящее время в АТ используется три конструкции псевдообъёмных эргатических $3D^{(V)}$ -индикаторов (термин «псевдообъёмный» показывает, что во всех случаях лётчики наблюдают экране видеоряд из 2D-проекций 3D-моделей реперных объектов, но оптическая часть таких индикаторов активизирует заданные составляющие зрительного аппарата подготовленного человека, заставляя его считать, что он видит 3D-объекты, при этом, если необходимо, он может профессионально тренировать свой глазомер):

1) эргатические безочковые одноканальные $3D^{(V)}$ -индикаторы на базе зеркального коллиматора (рис. 3, 5), имеющие «узкий зрачок наблюдения» – область пространства, где человеку комфортно наблюдать 3D-объекты (при условии постоянного перемещения подвижного наблюдателя в трёхмерной модели

пространства), её размер позволяет разместить в ней глаза одного человека (имеется эффект «мёртвой зоны» – дистанции от глаза наблюдателя до мнимого экрана или места в виртуальном пространстве, где наблюдаем ближайший 3D-объект, начиная от которого и далее человек может профессионально тренировать свой глазомер): за счёт эффекта коллимации зрачок человека настраивается «на взор в бесконечность» (воздействует на аккомодацию зрительного аппарата человека-наблюдателя), а наблюдение постоянно изменяющегося изображения воздействует на конвергенцию зрительного аппарата человека-наблюдателя, что в итоге заставляет человека-наблюдателя считать, что он видит 3D-объекты (таким образом в технологическую цепочку процесса синтеза 3D-объекта включается зрительный аппарат человека);

2) эргатические безочковые одноканальные 3D^(V)-индикаторы, использующие эффект коллимации с «широким зрачком наблюдения» – область пространства (рис. 4), в которой помещается кабина широкофюзеляжного самолёта (рис. 6), где 3D-объекты могут видеть все члены экипажа: условия наблюдения 3D-объектов и технологическая цепочка процесса синтеза 3D-объектов такие же, как и в одноканальных безочковых 3D^(V)-индикаторах с «узким зрачком наблюдения»;

3) безочковые одноканальные 3D-индикаторы с экранами высокой четкости и высокого разрешения (не ниже 4К) (рис. 7) с заданной дистанцией до наблюдения ближайшего 3D-объекта с учётом подтверждения визуально получаемой информации от других имитаторов приборов кабинного оборудования: позволяют обучать лётчика посадке ЛА на ВПП (такие индикаторы применяются в стационарных АТ; за счёт удаления лётчика от экрана и высокого разрешения создаются условия для рассмотрения лётчиком на экране синтезированного видеоряда из 2D-проекций 3D-моделей реперных объектов с воздействием на аккомодацию и конвергенцию зрительного аппарата человека, что близко к технологии синтеза визуально наблюдаемых 3D-объектов, как и в первых двух случаях, но нет экспериментально подтверждённых данных о возможности в этом случае профессионально тренировать глазомер лётчика);

4) эргатические двухканальные $3D^{(V)}$ -индикаторы с диспаратными очками позволяющими каждому глазу лётчика с учётом его положения в $3D$ -модели пространства видеть синтезированный только для этого глаза видеоряд из $2D$ -проекции $3D$ -моделей реперных объектов (позволяет обучать лётчика профессиональным навыкам работы с визуально наблюдаемыми $3D$ -объектами, находящимися вблизи него, например при обучении дозаправки самолёта в воздухе, когда заправочный конус находится недалеко от кабины ЛА) (рис. 8).



Рис. 5. Кабина АТ с ИВО с составным $3D^{(V_2)}$ -индикатором из трёх безочковых одноканальных $3D^{(V_2)}$ -индикаторов с «узким зрачком»



Рис. 6. Кабина АТ с ИВО с составным четырёхканальным безочковым $3D^{(V_2)}$ -индикатором из четырёх безочковых одноканальных $3D^{(V)}$ -индикаторов с «широким зрачком»

Исследования показали, что в зависимости от z -й конструкции оптической части эргатических $3D^{(V_2)}$ -индикаторов для того, чтобы m -й лётчик видел $3D$ -объекты, необходим интервал времени $\Delta t 3D_m^{(V_2)}$ от нескольких секунд до полугода. За этот интервал зрительный аппарат m -го лётчика обучается видеть $3D$ -объекты, рассматривая через z -ю конструкцию оптической части на плоском экране изменяемый с частотой не менее 120 мс видеоряд из $2D$ -проекции $3D$ -моделей, и при этом профессио-

нально тренировать свой глазомер. Таким образом, при полётах на АТ для полномасштабной тренировки предварительно у лётчика должны быть сформированы составляющие его когнитивной модели $C_m^{(3D^{(Vz)})}(t)$, позволяющие ему в каждый момент времени t ориентироваться в синтезированной визуально наблюдаемой трёхмерной среде и после интервала $\Delta t 3D_m^{(Vz)}$, необходимого для обучения его зрительного аппарата, работать с конкретным индикатором $3D^{(Vz)}$:

$$(15) C_m^{(3D^{(Vz)})}(t) = F(C_m^{(3D^{(V)})}, \Delta t(3D_m^{(Vz)})).$$



Рис. 7. Пилотажный АТ для обучения лётчика управлению ЛА с тремя 3D-индикаторами позволяющему лётчику пилотировать модель ЛА наблюдая линию горизонта



Рис. 8. Специализированный АТ для обучения лётчика дозаправки в воздухе с ИВО с двухканальным 3D-индикатором с диспаратными очками

3. Выводы

1. При выборе ИВО для АТ для полномасштабной тренировки лётчика необходимо учитывать, как должны формироваться составляющие его когнитивной модели $C_m^{(3D^{(Vz)})}(t)$, позволяющие ему в каждый момент времени t ориентироваться в синтезированной визуально наблюдаемой трёхмерной среде.

2. В отличие от остальных имитаторов АТ, позволяющих обучать лётчика профессиональным навыкам пилотирования ЛА и решению навигационных задач «при полётах в сплошной

облачности», которые определяются временем его налёта на реальном ЛА и на АТ, ИВО позволяет обучать лётчика решению навигационных задач с визуальным ориентированием по выделенным на фоне модели местности 3D-моделям реперных объектов при условии, что он распознаёт эти 3D-модели как реальные 3D-объекты с вероятностью более 98%, с учётом конструктивных особенностей 3D-моделей реперных объектов и подготовленности зрительного аппарата m -го лётчика во время полёта на АТ видеть 3D-объекты и при этом профессионально тренировать свой глазомер, в то время как он через 3D (V_z) -индикатор наблюдает видеоряд из 3D-проекций, попавших в пирамиду видимости 3D-моделей объектов, с циклом не более 120 мс.

Литература

1. ВЕНИКОВ В.А. *Теория подобия и моделирования*. – М.: Высшая школа, 1976. – 479 с.
2. КЕМАЛОВ Б.К., КУАТОВ Б.Ж., ЮРКОВ Н.К. *Формирование моделирующей среды авиационного тренажера // Надёжность и качество сложных систем*. – 2015. – №1(9). – С. 9–16.
3. КРАСОВСКИЙ А.А., ЛОПАТИН В.И. и др. *Авиационные тренажеры*. – М.: Изд-во ВВИА им.: Жуковского, 1992. – 320 с.
4. КУАТОВ Б.Ж. *Системный подход к организации тренажерной подготовки летного состава в современных условиях // Надёжность и качество сложных систем*. – 2014. – №4(8). – С. 34–40.
5. КУАТОВ Б.Ж., КЕМАЛОВ Б.К., БИКЕЕВ Р.Р. *Влияние адекватности тренажеров на планирования тренажерной подготовки // Надёжность и качество сложных систем*. – 2016. – №4(16). – С. 76–81.
6. ЛАПШИН Э.В. *Исследование информационных процессов, протекающих в тренажерах // Надёжность и качество сложных систем*. – 2013. – №2(2). – С. 87–93.
7. ЛАПШИН Э.В., КУАТОВ Б.К. КЕМАЛОВ Б.Ж. *Влияние адекватности тренажеров на планирования тренажерной*

- подготовки // Надёжность и качество сложных систем. – 2016. – №4(16). – С. 128–144.
8. МАРЧУК В.И., ГЛЕБОВА Т.А., КУВШИНОВА О.А. и др. *Задачи моделирования района полетов для авиационных тренажеров и операторов беспилотных летательных аппаратов* // Надёжность и качество сложных систем. – 2021. – №3(35). – С. 80–86.
 9. МАМАЕВ В.Я., СИНЯКОВ А.Н., ПЕТРОВ К.К. и др. *Воздушная навигация и элементы самолётовождения : учебное пособие.* – СПб.: СПбГУАП, 2002. – 256 с.
 10. РЫБКИН П.Н. *Компьютерный тренажёр для пилотов вертолётов МИ-8МТВ для отработки навигационных задач* // Научные проблемы транспорта Сибири и дальнего востока. – 2012. – №2. – С. 157–160.
 11. РОГАНОВ В.Р., КУВШИНОВА О.А., ЕСИМОВА Н.С. и др. *Проектирование авиационных тренажеров с распараллеливанием вычислительных процессов* // Надёжность и качество сложных систем. – 2020. – №2(30). – С. 96–105
 12. ШУКШУНОВ В.Е. *Автоматизированные обучающие системы профессиональной подготовки операторов летательных аппаратов.* – М.: Машиностроение, 1986. – 240 с.
 13. ROGANOV V., MIKHEEV M., ESIMOVA N. et al. *On the assessment of the image model of 3d models synthesized by optical-software-technical systems* // Proc. of the Moscow Workshop on Electronic and Networking Technologies (MWENT-2020). – 2020. – P. 9067481.
 14. ROGANOV V., DOLGOVESOV B., GUDKOVAI C. *Simulators of Visual Environment Of Aviation And Space Simulators* // E3S WEB of Conferences. –2023. – P. 03014.
 15. *Jane's military training and simulation systems.* – USA. – 471 p.
 16. ROGANOV V.R., SAGYNDYK A.B., AKHTARIEVA R.F. et al. *Integrated Organization of the System for Forming the Information Support of Aeronautical Simulator* // Int. Journal of Applied Engineering Research. – 2017. – Vol. 12, No. 15. – P. 5207-5213.
 17. ROGANOV V., DOLGOVESOV B., ASMOLOVA E. et al. *Analysis of Directions for Improvement Of Flight Simulators* //

- Web of Conferences. – 2023. – P. 04037. – DOI: 10.1051/e3sconf/202346004037.
18. MIKHEEV M.Yu., ROGANOV V.R., ANDREEV P.G. et al. *Developing the structure of the quality control system of power supply units in mobile robots // Proc. of the Int. Siberian Conf. on Control and Communications (SIBCON). – Astana, Kazakhstan, Astana, June, 29-30, 2017. – P. 728–728.*
 19. http://yandex.ru/images/search?img_url=https%3A%2F%2Fwww.gctc.ru%2Fmedia%2Fimages%2Fnews%2F2012%2Fv.zpk%2Fkosmozentr%2FSHL_1928.jpg&lr=11101&pos=0&rpt=simage&source=serp&text=авиационные%20тренажёры%20картинки (дата обращения 01.07.2024).

IMITATOR OF VISUAL ENVIRONMENT SYNTHESIZING IN THE VISUAL RANGE OF ELECTROMAGNETIC WAVES THE EXTERNAL APPEARANCE OF THE SPACE AROUND THE CABIN OF AN AIR TRAINING SIMULATOR

Vladimir Roganov, Federal State Budgetary Institution of Higher Education “Penza State Technological University”, Penza, Cand of Techn.Sc., assistant professor (vladimir_roganov@mail.ru)

Gennady Korop, Federal State Budgetary Institution of Higher Education “Penza State Technological University”, Penza, Cand of Techn.Sc., Junior Research Fellow (korop_g@mail.ru)

Mikhail Mikheev, Federal State Budgetary Institution of Higher Education “Penza State Technological University”, Penza, Doctor of Techn.Sc., professor (mix1959@gmail.com)

Abstract: An aviation simulator is a complex of several imitators combined for the purpose of training pilots in professional piloting and aircraft navigation skills in training situations. The purpose of flights on an aviation simulator is to form in pilots the components of its cognitive model, which will allow them to successfully operate in normal and pre-emergency situations in the future. Of all the imitators of an aviation simulator, the ergatic optical-software-hardware imitator stands out, synthesizing for the pilot a visually observable 3D-model of the external environment with a quality sufficient for solving navigation problems, including the use of the eye to estimate the distance to visible 3D-models of reference objects, from which a 3D-model of a given flight area is collected. This task is divided into a number of subtasks including the construction of 3D-models of reference objects, their description and placement in the database of a computer image generator,

the selection of a 3D-indicator, and training the pilot to see 3D-objects while he sees a video sequence of 3D-projections of 3D-models of reference objects. The article analyzes visual environment imitators and 3D-indicators from the point of view of the formation of constantly updated components of the pilot's cognitive model during both real approaches to an aircraft and flights on flight simulators, allowing him to successfully pilot an aircraft and solve navigation problems both "during flights in continuous cloud cover" and during flights with visual observation of reference objects in order to determine the location of the aircraft model and the direction of its flight.

Keywords: visual environment imitator, components of the pilot's cognitive model.

УДК: 629.7

ББК: 39.52

*Статья представлена к публикации
членом редакционной коллегии Б.Р. Андриевским.*

Поступила в редакцию 29.07.2024.

Опубликована 30.11.2024.