

Правительство Российской Федерации  
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ  
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ  
«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ  
«ВЫСШАЯ ШКОЛА ЭКОНОМИКИ»  
(НИУ ВШЭ)

УДК 004.93"1 004.932  
Рег. № НИОКТР 123022000053-6  
Рег. № ИКРБС

УТВЕРЖДАЮ  
И.о. проректора НИУ ВШЭ,  
канд. биол. наук  
\_\_\_\_\_ А.В. Балышев  
«\_\_\_» \_\_\_\_\_ 2023 г.

ОТЧЕТ  
О НАУЧНО-ИССЛЕДОВАТЕЛЬСКОЙ РАБОТЕ

АНАЛИЗ МОДЕЛЕЙ МНОГОКРИТЕРИАЛЬНОГО, КОЛЛЕКТИВНОГО,  
ОПТИМИЗАЦИОННОГО И СТРУКТУРНОГО ВЫБОРА РЕШЕНИЙ  
(заключительный)

Руководитель НИР,  
директор, д-р техн. наук, проф.

\_\_\_\_\_ Ф.Т. Алескеров

Москва 2023

## СПИСОК ИСПОЛНИТЕЛЕЙ

Руководитель НИР, директор, д-р техн. наук, проф. \_\_\_\_\_  
подпись, дата

Ф. Т. Алескеров  
(введение, заключение)

Отв. исполнитель, гл. науч. сотр., PhD \_\_\_\_\_  
подпись, дата

Э. Маскин  
(пункты 2.2.1, 2.2.2)

**Исполнители:**

Зам. директора центра, д-р физ.-мат. наук, проф. \_\_\_\_\_  
подпись, дата

А. Е. Лепский  
(пункты 1.4.1–1.4.3)

Ведущий науч. сотр., д-р техн. наук, проф. \_\_\_\_\_  
подпись, дата

А. С. Беленький  
(пункты 2.2.3, 2.2.4)

Ведущий науч. сотр., д-р техн. наук, проф. \_\_\_\_\_  
подпись, дата

Б. Г. Миркин  
(подраздел 3.1)

Ведущий науч. сотр., д-р техн. наук \_\_\_\_\_  
подпись, дата

В. В. Подиновский  
(подраздел 2.1)

Ст. науч. сотр., канд. физ.-мат. наук \_\_\_\_\_  
подпись, дата

Ю. А. Веселова  
(подпункты 1.3.1.3, 1.3.1.4)

Ст. науч. сотр., канд. экон. наук \_\_\_\_\_  
подпись, дата

Д. С. Карабекян  
(подпункты 1.3.1.1, 1.3.1.2)

Ст. науч. сотр., канд. экон. наук \_\_\_\_\_  
подпись, дата

А. В. Карпов  
(пункт 1.1.1)

Ст. науч. сотр., канд. наук по прикл. мат. \_\_\_\_\_  
подпись, дата

Т. А. Ломоносов  
(подраздел 4.1)

Ст. науч. сотр., канд. физ.-мат. наук	_____	В. С. Молостцов
	подпись, дата	(пункты 2.3.3–2.3.8)
Ст. науч. сотр., канд. техн. наук, доц.	_____	А. Л. Мячин
	подпись, дата	(пункты 3.3.1–3.3.3)
Ст. науч. сотр., д-р наук по прикл. мат.	_____	И. В. Романов
	подпись, дата	(подраздел 4.2)
Ст. науч. сотр., канд. физ.-мат. наук	_____	А. Н. Субочев
	подпись, дата	(пункты 1.2.1–1.2.3)
Ст. науч. сотр., канд. техн. наук	_____	В. И. Якуба
	подпись, дата	(пункт 1.3.5)
Науч. сотр., PhD	_____	Э. Доган
	подпись, дата	(пункт 1.1.2)
Науч. сотр., канд. физ.-мат. наук	_____	Л. Г. Егорова
	подпись, дата	(пункты 2.3.1, 2.3.2)
Науч. сотр.	_____	Н. Г. Мещерякова
	подпись, дата	(пункты 3.1.1, 3.1.2)
Мл. науч. сотр.	_____	С. С. Демин
	подпись, дата	(пункты 1.4.4, 1.4.5)
Мл. науч. сотр.	_____	Ю. В. Зонтов
	подпись, дата	(пункты 1.3.3, 1.3.4)
Мл. науч. сотр	_____	А. А. Иванов
	подпись, дата	(пункт 1.3.4)

Мл. науч. сотр

\_\_\_\_\_  
подпись, дата

Д. С. Ткачев  
(пункты 3.4.1, 3.4.2)

Стажёр-исследователь

\_\_\_\_\_  
подпись, дата

М. С. Богоявленский  
(пункты 3.4.3, 3.4.4)

Стажёр-исследователь

\_\_\_\_\_  
подпись, дата

К. Е. Зиновьева  
(пункты 3.1.3, 3.1.4)

Стажёр-исследователь

\_\_\_\_\_  
подпись, дата

А. А. Микаелян  
(пункты 3.4.5, 3.4.6)

Стажёр-исследователь

\_\_\_\_\_  
подпись, дата

К. В. Мячина  
(пункты 3.1.5, 3.1.6)

Стажёр-исследователь

\_\_\_\_\_  
подпись, дата

А. К. Степочкина  
(пункты 3.1.7–3.1.9)

Стажёр-исследователь

\_\_\_\_\_  
подпись, дата

К. В. Сурова  
(пункты 3.1.10, 3.1.11)

Стажёр-исследователь

\_\_\_\_\_  
подпись, дата

Д. А. Чубарова  
(пункты 3.3.4–3.3.8)

Стажёр-исследователь

\_\_\_\_\_  
подпись, дата

А. В. Юдина  
(пункты 1.2.4, 1.2.5)

Нормоконтроль

\_\_\_\_\_  
подпись, дата

Е. А. Афанасьева

## **РЕФЕРАТ**

Отчет 196 с., 1 кн., 53 рис., 20 табл., 212 источн.

**ТЕОРИЯ ПРИНЯТИЯ РЕШЕНИЙ, ТЕОРИЯ КОЛЛЕКТИВНОГО ВЫБОРА, ТЕОРИЯ ИГР,  
КЛАСТЕРНЫЙ АНАЛИЗ, АНАЛИЗ ПАТТЕРНОВ, СЕТЕВОЙ АНАЛИЗ, МЕТОДЫ  
ОПТИМИЗАЦИИ, ТЕОРИЯ СВИДЕТЕЛЬСТВ**

Объектами исследования были задачи и методы теории принятия решений, методы кластеризации и паттерн-анализа данных, методы теории свидетельств, базы данных торговли электроэнергией и редкоземельными металлами, база данных российских банков, рейтинги научных журналов, библиометрические данные, исторические данные курсов валют и пр.

Целью исследования были разработка и анализ моделей и методов многокритериального, коллективного, оптимизационного и структурного выбора решений с применением этих моделей к решению ряда прикладных задач моделирования социально-экономических систем.

Методологической базой проекта были: теория принятия решений, теория коллективного выбора, теория игр, методы кластеризации и паттерн-анализа данных, методы сетевого анализа данных, теории свидетельств, методы оптимизации и др.

Область применения полученных результатов: системы выборов на всех уровнях, принятие решений в финансовой и банковской сферах, принятие решений во внешней торговле, менеджмент образования, управление наукой, библиометрический анализ и др.

## СОДЕРЖАНИЕ

ВВЕДЕНИЕ .....	10
1 Исследования по теории выбора решений и её применению .....	15
1.1 Исследование доменов Кондорсе .....	15
1.1.1 Симметричные домены Кондорсе .....	15
1.1.2 Обобщенные предпочтения Фишберна .....	17
1.2 Исследование аксиоматических свойств и разработка эффективных алгоритмов вычисления новых обобщений устойчивых турнирных решений .....	19
1.2.1 Общие теоретические сведения .....	21
1.2.2 Свойства версий минимального покрывающего множества .....	24
1.2.3 Проектирование и разработка MVP .....	26
1.2.4 Эмпирическое сравнение методов ранжирования .....	28
1.2.5 Заключение .....	34
1.3 Исследования по манипулированию процедурам агрегирования в теории коллективного выбора .....	34
1.3.1 Манипулирование при неполной информации и неопределенности относительно действий других избирателей .....	35
1.3.2 Разработка структуры базы данных и сайта для хранения и визуализации данных по индивидуальной и коллективной манипулируемости процедур агрегирования ....	37
1.3.3 Разработка алгоритмов распараллеливания задачи оценки манипулируемости процедур агрегирования .....	42
1.3.4 Оценка соотношений между манипулируемостью различных процедур агрегирования для различных методов построения расширенных предпочтений .....	46
1.3.5 Оценка соотношений между манипулируемостью q-Паретовских правил коллективного выбора; .....	53
1.4 Доверительное агрегирование и ранжирование в порядковой шкале .....	57
1.4.1 Введение .....	57
1.4.2 Необходимые сведения из теории свидетельств .....	58
1.4.3 Постановка задачи и общая схема доверительного ранжирования .....	59
1.4.4 Численный пример .....	65

1.4.5 Заключение .....	68
<b>2 Разработка и исследование оптимизационных моделей.....</b>	<b>69</b>
2.1 Аппроксимация функций и регрессионный анализ: многокритериальный подход ..	69
2.1.1 Введение.....	69
2.1.2 Многокритериальная задача аппроксимации.....	70
2.1.3 О построении множеств $U^\Gamma$ и $Y^\Gamma(x)$ .....	72
2.1.4 Аппроксимация при отсутствии информации $\Gamma$ .....	74
2.1.5 Аппроксимация при равноважных критериях .....	77
2.1.6 Аппроксимация при равноважных критериях со шкалой первой порядковой метрики.....	78
2.1.7 О взаимосвязи недоминируемых и классических аппроксимирующих функций.....	82
2.1.8 Заключение .....	83
2.2 Разработка инструментария решения задачи о выборе оптимальной структуры курсов для смешанного обучения .....	83
2.2.1 Введение.....	83
2.2.2 Краткий обзор научных публикаций по использованию смешанных курсов в образовательном процессе .....	88
2.2.3 Постановки задач и их математические формулировки .....	93
2.2.4 Проблемы, связанные с проведением смешанных курсов .....	103
2.3 Оптимизация гарантированного риска с учетом дополнительных связей между неопределенными факторами .....	109
2.3.1 Введение.....	109
2.3.2 Принцип минимаксного сожаления Сэвиджа .....	110
2.3.3 Постановка задачи о выборе структуры вклада .....	111
2.3.4 Явный вид функции гарантированного риска и возможные оптимальные решения .....	112
2.3.5 Алгоритм построения гарантированного по риску решения.....	115
2.3.6 Экономическая и игровая интерпретация.....	116

2.3.7 Результаты модельных экспериментальных расчетов .....	119
2.3.8 Некоторые выводы.....	122
3 Разработка и применение методов сетевого и кластерного анализа данных .....	123
3.1 Сетевые модели цитирований авторов и организаций в области болезни Паркинсона.....	123
3.1.1 Введение.....	123
3.1.2 Описание данных .....	124
3.1.3 Построение сети цитирований организаций .....	125
3.1.4 Построение сети цитирований авторов.....	125
3.1.5 Результаты анализа сети цитирования организаций .....	126
3.1.6 Результаты паттерн-анализа цитируемости организаций .....	127
3.1.7 Результаты анализа устойчивости сети цитирования организаций .....	127
3.1.8 Результаты анализа сети цитирования авторов.....	128
3.1.9 Результаты паттерн-анализа цитируемости журналов .....	130
3.1.10 Результаты анализа устойчивости сети цитирования журналов .....	131
3.1.11 Заключение .....	131
3.2 Применение метода аномальных кластеров к анализу температурных карт .....	131
3.2.1 Введение.....	131
3.2.2 Особенности применения алгоритма выделения аномальной группы на океанских данных.....	132
3.2.3 Аномальные кластеры .....	134
3.3 Разработка человеко-машинной процедуры паттерн-кластеризации данных .....	137
3.3.1 Введение.....	137
3.3.2 Аналитический обзор литературы.....	138
3.3.3 Математическая модель построения $\delta$ -трубки .....	138
3.3.4 Описание используемых программных продуктов .....	141
3.3.5 Спецификации входных и выходных данных программы.....	142
3.3.6 Описание интерфейса программы .....	143
3.3.7 Результаты кластеризации данных .....	144

3.3.8 Заключение .....	149
3.4 Паттерн-анализ экспорта-импорта электроэнергии .....	149
3.4.1 Введение.....	149
3.4.2 Обзор литературы .....	150
3.4.3 Описание и предобработка данных .....	151
3.4.4 Описание метода и алгоритма .....	154
3.4.5 Результаты и дискуссия .....	156
3.4.6 Заключение .....	166
4 Исследование непрерывных моделей.....	168
4.1 Разработка алгебраического метода решения линейных разностных стационарных уравнений высшего порядка.....	168
4.1.1 Введение.....	168
4.1.2 Алгебраическая модификация метода неопределенных коэффициентов .....	170
4.1.3 Практическая демонстрация метода .....	174
4.1.4 Принцип суперпозиции .....	177
4.2 Локальная гладкая управляемость для некоторой распределенной системы четвёртого порядка .....	178
ЗАКЛЮЧЕНИЕ.....	180
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ .....	182

## ВВЕДЕНИЕ

В отчете представлены результаты исследований по проекту «Анализ моделей многокритериального, коллективного, оптимизационного и структурного выбора решений», который выполнялся в 2023 году сотрудниками Международного центра анализа и выбора решений (МЦАВР) НИУ ВШЭ.

Проект выполнялся по следующим основным научным направлениям:

- исследования по теории выбора решений и её применением;
- разработка и исследование оптимизационных моделей;
- разработка и применение методов сетевого и кластерного анализа данных;
- исследование непрерывных моделей.

Данный проект является продолжением проекта 2022 года «Теоретические и прикладные задачи анализа данных и принятия решений в социально-экономической и финансовой сферах» [1]. В частности, в 2022 г. были получены следующие результаты:

- поставлена и исследована задача принятия решения при наличии информации о важности групп критериев; найдены способы проверки непротиворечивости информации о важности и способы построения введенных отношений предпочтения; выявлена взаимосвязь качественной значимости и качественной вероятности;
- исследованы вычислительные свойства задач коллективного выбора для случая, когда предпочтения избирателей являются сепарабельными; показано, что правило Янга и правило Чемберлина – Куранта являются NP-трудными; разработан алгоритм генерирования сепарабельных предпочтений;
- исследованы турнирные способы выбора оптимальных вариантов на основе обобщений минимального покрывающего множества; рассмотрена задача агрегирования международных рейтингов вузов на основе указанных обобщений и других методов ранжирования на основе турнирных решений;
- в рамках исследования манипулируемости правил коллективного выбора разработаны алгоритмы, позволяющие получить точные значения индексов манипулируемости для случая трех альтернатив; вычислены значения индексов индивидуального манипулирования для числа участников от 3 до 100, для различных процедур голосования и для 4 типов расширенных предпочтений;
- исследована стабильность паросочетаний между государственными адвокатами и судебными делами, в которых истцы и/или ответчики не могут позволить себе правовых услуг; введена новая концепция стабильности, которая устраняет все возможные блокирующие пары в практических сценариях;

- рассмотрены алгоритмы паросочетаний, максимизирующие доход для обобщенных паросочетаний с платежами участников (традиционно, эта модель носит название «рынка бракосочетаний»); проанализированы условия, при которых существуют равновесия с нулевым доходом;
- исследована аксиоматика линейных функций интенсивности связи для кардинальных индексов влияния;
- разработан программный комплекс в среде Excel решения задачи о нахождении структур трехвалютного депозита, оптимальных по гарантированному риску, в условиях неизвестных будущих обменных курсов валют; проведено численное моделирование зависимости минимального гарантированного риска от максимальных возможных будущих значений курсов валют;
- предложена методика формирования гибких бизнес-правил для улучшения процесса закупки компонентов для систем управления на основе критерия минимума максимального риска;
- исследована задача нахождения набора учебных курсов по выбору, который был бы оптимальным в некотором смысле, как для студента с точки зрения администрации, так и для самих студентов;
- разработан метод максимального правдоподобия для решения задачи обобщения нечетких кластеров в таксономиях предметных областей; метод апробирован на материалах публикаций издательства Springer в области Науки о данных;
- в рамках теории функций доверия разработаны два метода кластеризации тел свидетельств на основе вычисления функции плотности конфликта между фокальными элементами и на основе минимизации среднего внутрикластерного конфликта;
- проанализированы аксиоматические свойства ряда мер центральности сетевых структур;
- выполнен сетевой анализ научных публикаций в области исследований болезни Паркинсона за период 2015–2021 гг. с помощью классических и новых индексов центральности, которые учитывают групповое влияние и позволяют выделить группы ссылающихся друг на друга исследователей;
- разработана имитационная модель разлива нефти для специфических условий Арктики (низкие температуры, лед на поверхности моря и т. д.); модель апробирована для разных частей арктического региона и различных сценариев возможных разливов;
- исследованы библиометрические и наукометрические показатели международных журналов по бизнесу, менеджменту и бухгалтерскому учету из базы данных Scopus;

- выполнен анализ регионального развития методом динамической паттерн-кластеризации; выявлены схожие траектории развития регионов, найдены группы регионов со схожим развитием, проанализирована устойчивость регионального развития;
- выявлены ключевые факторы, влияющие на принятие решений относительно вакцинации против COVID-19 молодыми людьми, с помощью обработки результатов опроса, проведенного методом эксперимента дискретного выбора;
- разработана методология повышения точности прогноза прироста числа больных COVID-19 по странам мира с использованием методов динамического кластерного анализа на признаковом пространстве (уровень образования, законопослушность граждан, строгость ограничительных мер);
- предложена и реализована на языке Python процедура унификации и оптимизации мониторинга текущей ситуации с COVID-19, прогнозирование роста заражений с использованием моделей кластеризации и линейной регрессии, анализа эффективности с помощью метода оболочечного анализа данных.

Ряд задач исследования проекта 2023 года являются логическим продолжением задач и результатов проекта 2022 года. В частности, были продолжены исследования по манипулированию правил коллективного выбора, по формированию оптимального набора курсов (только теперь речь идет о смешенном обучении), разработка новых методов паттерн-анализа, продолжены исследования по применению сетевого анализа данных и т. д. Другие исследования 2023 года являются новыми. Например, разработка метода аппроксимации функций на основе многокритериального подхода, разработка метода доверительного агрегирования и ранжирования в порядковой шкале и др. Некоторые исследования этого года предполагают дальнейшее развитие в следующем году.

Методологической базой исследования были теория коллективного выбора, методы оптимизации, многокритериальное принятие решений, теоретико-игровые модели, сетевой и кластерный анализ данных, теория свидетельств и др.

Получены как теоретические, так и прикладные задачи. Некоторые задачи потребовали разработки комплексов программ и использования открытых баз данных.

Отчет состоит из 4 разделов.

В первом разделе приведены результаты исследований по теории выбора решений и ее применением.

Подраздел 1.1 посвящен результатам исследования доменов Кондорсе. В частности, в рамках данного исследования были получены характеристики важного в теории предпочтений класса максимальных доменов Кондорсе.

В подразделе 1.2 приведены результаты продолжающегося исследования аксиоматических свойств новых обобщений устойчивых турнирных решений. Кроме того, в этом году было разработано веб-приложение для агрегирования рейтингов научных журналов, основанных на различных библиометрических показателях.

Подраздел 1.3 посвящен результатам продолжающегося исследования по манипулированию процедурам агрегирования. В частности, найдены условия, при которых вероятность существования избирателя, имеющего стимул к манипулированию равна нулю. Разработана структура базы данных и сайт для хранения и визуализации данных по индивидуальной и коллективной манипулируемости процедур агрегирования. Разработаны алгоритмы распараллеливания задачи оценки манипулируемости процедур агрегирования. Проведено сравнение манипулируемости различных процедур агрегирования в случае Impartial Culture, множественного выбора и 4-х методов построения расширенных предпочтений. Кроме того, проведено сравнение между манипулируемостью q-Паретовских правил коллективного выбора.

Наконец, в подразделе 1.4 приведены результаты по разработке в рамках теории свидетельств процедуры агрегирования информации от экспертов и последующего ранжирования этих агрегированных оценок в том случае, когда оценки представлены в порядковой шкале с указанием их достоверности.

Во втором разделе отчета приведены результаты по разработке и исследованию некоторых оптимизационных моделей.

Так, подраздел 2.1 посвящен исследованию свойства введенных ранее аппроксимирующих функций на основе многокритериального подхода. Исследованы взаимосвязи новых аппроксимирующих функций с классическими. Предложены вычислительные методы построения множеств аппроксимирующих функций.

В подразделе 2.2 описан инструментарий принятия решений по оптимальному формированию набора учебных курсов, преподавательского состава и бюджета при смешанном формате обучения студентов.

В подразделе 2.3 приведены результаты модельных вычислений и анализа задачи о структуре трехвалютного вклада, оптимального по Сэвиджу, с учетом динамики валютных курсов в 2023 году.

Третий раздел отчета посвящен результатам, связанным с разработкой и применением методов сетевого и кластерного анализа данных.

Так в подразделе 3.1 приведены результаты исследований методами сетевого анализа публикаций в области болезни Паркинсона.

В подразделе 3.2 описано применение метода выделения аномальных кластеров, который был разработан ранее участниками проекта, к анализу карт температур морской поверхности. В частности, с помощью этого метода было описано такое явление как апвеллинг у берегов Португалии.

В подразделе 3.3 описан программный комплекс паттерн-кластеризации данных с различными типами кластерных трубок. Кроме того, описана разработанная человеко-машинная процедура, позволяющая объединять и обрабатывать в полуавтоматическом режиме результаты разных кластеризаций.

Наконец, в подразделе 3.4 методами паттерн-кластеризации проведен анализ на страновом уровне экспорта-импорта электроэнергии.

Последний четвёртый раздел посвящен исследованиям непрерывных моделей.

Так в подразделе 4.1 приведены результаты по разработке алгебраической модификации метода неопределенных коэффициентов для решения линейных разностных стационарных уравнений высшего порядка.

А в подразделе 4.2 описана постановка и решение задача приведения в состояние покоя некоторой распределенной системы четвёртого порядка на двумерном торе посредством локального гладкого управляющего воздействия.

Результаты, отраженные в отчете, были апробированы на ряде авторитетных семинаров и конференций, были своевременно опубликованы или находятся в печати.

## **1 Исследования по теории выбора решений и её применением**

В этом разделе отчета отражены результаты исследований в области теории выбора, которые проходили в следующих основных научных направлениях:

- исследование доменов Кондорсе;
- исследование аксиоматических свойств и разработка эффективных алгоритмов вычисления новых обобщений устойчивых турнирных решений;
- исследования по манипулированию процедурами агрегирования в теории коллективного выбора;
- доверительное агрегирование и ранжирование в порядковой шкале.

### **1.1 Исследование доменов Кондорсе**

Результаты этого раздела основаны на статьях [2, 3], являющихся продолжением работ в области исследования структурированных предпочтений и доменов Кондорсе [4, 5].

#### **1.1.1 Симметричные домены Кондорсе**

Знаменитый парадокс Кондорсе показывает, что если множество допустимых предпочтений избирателей неограниченно, то результат голосования методом простого большинства может быть нетранзитивным. Домены Кондорсе — это подмножества линейных порядков, обладающие тем свойством, что всякий раз, когда предпочтения всех избирателей принадлежат этому набору, голосование методом простого большинством никогда не приводит к циклам в общественных предпочтениях. Домены Кондорсе являются довольно интересными объектами с комбинаторной точки зрения.

Давно известно [6], что домен Кондорсе должен удовлетворять полному набору так называемых запрещающих условий, которые могут быть трех различных видов: не-первый, не-второй и не-третий. Каждое не-условие для каждой тройки кандидатов запрещает одному из кандидатов занимать определенную позицию по отношению к двум другим кандидатам.

Обычно домен Кондорсе удовлетворяет смеси запрещающих условий, т. е. для разных троек он может удовлетворять разным запрещающим условиям, например в перемежающей схеме Фишберна [7] используются условия не-первый и не-третий. Однако даже домены, удовлетворяющие одному и тому же не-условию для каждой тройки, далеки от полной классификации.

Основная цель этого исследования — охарактеризовать класс максимальных доменов Кондорсе, которые имеют запрещающее условие никогда не второй в каждой тройке

альтернатив. Целесообразно классифицировать те домены, которые не являются композициями (ни в каком смысле) более мелких доменов. Для этого мы определим концепцию декомпозиции доменов Кондорсе.

Домен Кондорсе является *симметричным*, если для каждого линейного порядка из домена верно, что обратный порядок также принадлежит домену. Домен Кондорсе является *максимальным*, если не существует другого домена Кондорсе, который содержал бы данный домен в качестве подмножества. Первым исследованием симметричных доменов Кондорсе является работа [8], где поставлен ряд нерешенных проблем, ответ на которые пытается найти работа [2].

Множество альтернатив обозначается через  $A = \{1, \dots, m\}$ , а множество агентов через  $\mathcal{N} = \{1, \dots, n\}$ . Каждый агент  $i \in \mathcal{N}$  имеет линейный порядок предпочтения  $P_i$  на множестве  $A$  (далее, если не оговорено иное, то под предпочтением подразумевается линейный порядок). Максимальный элемент в этом порядке является наилучшим, минимальный – наихудшим. Для  $aP_i c$  и  $bP_i c$  вводится обозначение  $\{a, b\}P_i c$ . Пусть  $\mathcal{L}(A)$  будет множеством всех возможных линейных порядков на множестве  $X$ . Кортеж из  $n$  линейных порядков является профилем предпочтений:  $\mathcal{P} = (P_1, \dots, P_n) \in \mathcal{L}(A)^n$ . Доменом предпочтений является подмножество линейных порядков  $D \subseteq \mathcal{L}(A)^n$ .

Показано, что симметричность домена Кондорсе определяется ограничениями на тройки альтернатив, т. е. локальными условиями симметричности.

**Теорема 1.1.** [2] Максимальный домен Кондорсе является симметричным тогда и только тогда, когда для каждой тройки альтернатив верно, что существует альтернатива, никогда не стоящая на втором месте в сужении домена на данную тройку.

Пусть  $D$  является доменом Кондорсе на множестве  $A = \{a_1, \dots, a_m\}$ ,  $D_1, \dots, D_m$  являются доменами Кондорсе на непересекающихся множествах  $C_1, \dots, C_m$ . Тогда определим композицию доменов как домен на множестве  $C_1 \cup \dots \cup C_m$ ,  $D(a_1 \rightarrow D_1, \dots, a_m \rightarrow D_m) := \{u_1, \dots, u_m \mid u_i \in D_{i,j}, a_{i_1}, \dots, a_{i_m} \in D\}$ . Если домены  $D, D_1, \dots, D_m$  являются симметричными, то и их композиция является симметричной. Если композиция является симметричной, то и все участвующие в ней домены являются симметричными. Например, имея домены  $D = \{ab, ba\}$ ,  $D_1 = \{12, 21\}$ ,  $D_2 = \{34, 43\}$ , получим домен  $\{1234, 1243, 2134, 2143, 3412, 3421, 4312, 4321\}$ .

Домен является недекомпозируемым, если его невозможно представить в виде композиции доменов, в которой  $|D| > 1$  и хотя бы для одного домена нижнего уровня выполнено  $|D_i| > 1$ .

Домен  $K = \{1234, 4321, 2413, 3142\}$  (здесь и далее линейный порядок описывается в виде последовательности альтернатив) является примером недекомпозиуемого максимального симметричного домена Кондорсе. Рассмотрим домены вида  $D = \{1234, 4321, x, \bar{x}\}$ , где  $\bar{x}$  – последовательность альтернатив, имеющая обратный порядок относительно последовательности  $x$ . Такие домены получили название домены Рейно [9].

Последовательность является простой, если не содержит интервала длиной более 1 и менее  $m$  элементов, чьи значения образуют интервал натуральных чисел. К примеру, последовательность 3124 не является простой, так как содержит интервал 312.

**Теорема 1.2.** [2] Домен  $D = \{1234, 4321, x, \bar{x}\}$  является максимальным симметричным недекомпозиуемым доменом Кондорсе, если последовательность  $x$  является простой.

В [2] показано, что верно и обратное утверждение для числа альтернатив, не превышающем шести. Домены, описанные в теорема 1.2, являются основой для постройки разнообразных симметричных доменов Кондорсе, используя операцию композиции доменов. Содержательным примером таких доменов является домены сепарабельных предпочтений.

Симметричные максимальные области Кондорсе образуют богатый класс областей Кондорсе, описание которых сводится к характеристике неразложимых симметричных максимальных доменов. Помимо тривиальных доменов на множестве единственной альтернативы, обнаружен богатый класс доменов мощностью 4, которые называются доменами Рейно. Он охарактеризован в терминах простых последовательностей, которые являются хорошо изученным комбинаторным объектом.

Авторы исследования предполагают, что других неразложимых симметричных максимальных областей Кондорсе не существует. Это проверено для случая пяти и шести альтернатив, но оставили вопрос в полной общности открытым. Другой интересный связанный с этим открытый вопрос заключается в характеристике возможных размеров симметричных максимальных областей Кондорсе. Все известные конструкции дают нам размер, равный степени двойки.

### 1.1.2 Обобщенные предпочтения Фишберна

Однопиковые предпочтения моделируют предпочтения на множестве желаемых альтернатив. Основополагающим свойством этих предпочтений является минимальное разнообразие, каждая альтернатива является наилучшей хотя бы в одном предпочтении из домена. Обратные к ним, однолунковые предпочтения, моделируют предпочтения на нежелательных альтернативах, например на возможных местах размещения загрязняющего окружающую среду предприятия. В данном домене каждая альтернатива является наихудшей хотя бы в одном предпочтении.

Модель структурированных предпочтений, описывающих предпочтения как на желаемых, так и на нежелаемых предпочтениях построена в [3]. Такой моделью является обобщенные предпочтения Фишберна [7]. Домен определяется подмножеством альтернатив  $X \subseteq A$  и последовательностью альтернатив  $a_1, \dots, a_m$ . Домен определяется следующими условиями на тройки альтернатив. Для каждой тройки альтернатив  $a_i, a_j, a_k$  верно, что если медиана чисел  $i \leq j \leq k$  принадлежит  $X$ , то альтернатива  $a_j$  никогда не стоит на первом месте в сужении домена на эту тройку, и если медиана чисел  $i \leq j \leq k$  не принадлежит  $X$ , то альтернатива  $a_j$  никогда не стоит на последнем месте в сужении домена на эту тройку.

Домен однопиковых предпочтений и домен Фишберна являются частными случаями обобщенного домена Фишберна.

Домен Кондорсе является слабо минимально разнообразным, если для каждой альтернативы найдется линейный порядок, в котором данная альтернатива находится на первом или последнем месте. Домен является связным, если один линейный порядок домена можно получить из другого с помощью последовательного обращения пар соседних альтернатив линейных порядков, при этом все промежуточные линейные порядки тоже принадлежат тому же домену.

**Теорема 1.3** [3]. Максимальный домен Кондорсе является связным и слабо минимально разнообразным тогда и только тогда, когда он является обобщенным доменом Фишберна.

Данный результат является обобщением аксиоматической характеристацией однопиковых предпочтений К. Пуппе [10]. Однопиковые на окружности предпочтения определяются аналогично однопиковым. Альтернативы размещены на окружности. Каждый агент имеет свою наилучшую и наихудшую альтернативу. Альтернативы между наилучшей и наихудшей каждой агент упорядочивает согласно их размещению на окружности. Альтернативы с разных сторон от наилучшей могут быть упорядочены произвольным образом. Домен однопиковых на окружности предпочтений не является доменом Кондорсе.

**Теорема 1.4** [3]. Обобщенный домен Фишберна является подмножеством однопикового на окружности домена.

Обобщенный домен Фишберна является максимальным доменом Кондорсе. Таким образом, этот домен совмещает алгоритмические преимущества однопиковых на окружности предпочтений [11] и нормативные свойства доменов Кондорсе.

Обобщенный домен Фишберна имеет несколько содержательных интерпретаций, совмещающих расположенные на окружности альтернативы и разбиение альтернатив на приемлемые и неприемлемые. Например, окружностью может быть 24-часовой временной

интервал. Коллектив ищет возможность согласовать общее время встречи онлайн. Альтернативы могут лежать как внутри интервала рабочих часов, так и вне его, но только альтернативы, лежащие в интервале рабочих часов, могут быть наилучшими, остальные не могут быть наилучшими, но могут быть наихудшими.

Эмпирические исследования [12] показывают, что однопиковые предпочтения являются теоретической концепцией без примеров из реальных выборов. Предпочтения близкие к однопиковым [13] направлены на то, чтобы приблизить чисто однозначные предпочтения и расширить набор рассматриваемых профилей предпочтений. Существует много способов определить предпочтения, близкие к однопиковым: профили предпочтений, которые становятся однопиковыми после удаления k агентов, или после удаления k альтернатив, или после разделения на k частей, или после выполнения k перестановок последовательных альтернатив в порядке предпочтения и т.д. Все эти обобщения являются эвристическими и не обладают четкими теоретическими свойствами, например, все такие области не являются областями Кондорсе.

Несмотря на хорошие математические свойства домена Фишберна, он не находит применения в теории общественного выбора. В этом разделе аксиоматически обосновываются домены, являющиеся почти однопиковыми, (GF-домены) и дается интерпретация этих предпочтений. Введение GF-доменов — это шаг к применению нового класса структурированных предпочтений, включая домен Фишберна. Совмещение моделей предпочтений на множестве желаемых и нежелаемых альтернатив (смешанная манна) является современным развитием экономических моделей выбора [14].

## **1.2 Исследование аксиоматических свойств и разработка эффективных алгоритмов вычисления новых обобщений устойчивых турнирных решений**

Государство ежегодно тратит сотни миллиардов рублей на развитие науки. Для получения наибольшей отдачи подобные инвестиции проводятся с учетом оценки результативности ученых. Данная оценка делается на основе анализа значимости научных журналов, в которых публикуется ученый. Для сравнения научных журналов в библиометрии предложено множество библиометрических показателей. Очевидно, что для большей объективности обсуждаемая оценка результативности должна проводиться по совокупности этих показателей, однако на практике оказывается, что несовпадение относительных оценок одних и тех же научных журналов по различным библиометрическим показателям приводит к отсутствию (в общем случае) транзитивности. Возникает проблема оптимального выбора в ситуации противоречивых предпочтений, то есть проблема выбора в некотором

(не наивном) смысле лучших альтернатив на основании результатов их попарного сравнения, репрезентированных нетранзитивным отношением. Подобную проблему можно представить как задачу выбора победителя (или победителей) в спортивном турнире, и потому концепции её решения были названы *турнирными решениями*.

Очевидная сложность как в построении теоретического обоснования, так и в практической реализации нетривиальных методов разрешения описанных выше противоречий привела к отсутствию в открытом доступе информационных систем, позволяющих строить агрегированные ранжирования научных журналов на основе их библиометрических показателей.

В данном исследовании для решения поставленной проблемы были использованы обобщения такого турнирного решения, как минимальное покрывающее множество (*minimal covering set*), введенное в [15].

Основными задачами исследования были, во-первых, проверка сохранения рассматриваемыми версиями свойства устойчивости минимального покрывающего множества, во-вторых, разработка прототипа (*minimum viable product, MVP*) информационной системы, позволяющей производить оценку научных журналов на основе значений их библиометрических показателей, и в-третьих, проверка на реальных данных качества ранжирований, полученных с помощью устойчивых версий обсуждаемого решения.

В результате проведённого исследования было получено описание свойств обобщений минимального покрывающего множества и сделан вывод о целесообразности их использования в прикладных задачах. Кроме того, был разработан (на Python) MVP библиометрической информационной системы, аналогов которой на рынке не представлено. На примере реальных и актуальных данных о научных журналах по менеджменту показано, что турнирные решения репрезентируют совокупность исходных ранжирований, построенных на основе значений шести наиболее популярных библиометрических показателей, лучше, чем каждое из них.

В пункте 1.2.1 описывается теоретическая модель выбора на основании попарного сравнения вариантов, приводятся определения основных понятий. Формулировки определений заимствованы в основном из [16].

В пункте 1.2.2 содержатся формулировки теорем, показывающих наличие или отсутствие у рассматриваемых обобщений минимального покрывающего множества важных свойств турнирных решений, определённых в разделе 1.1.

В пункте 1.2.3 приводится описание разработанного MVP, также в ней указаны инструменты, использованные в процессе разработки. Принцип работы веб-приложения описывается на методологию, описанную в [17].

В пункте 1.2.4 приводится результат сравнения сортировок с помощью рассматриваемых обобщений минимального покрывающего множества с некоторыми другими методами построения ранжирований, основанными на турнирных решениях. Приведенные в данном разделе результаты были получены с помощью разработанного MVP.

В Заключении кратко резюмируются основные результаты исследования.

### 1.2.1 Общие теоретические сведения

#### 1.2.1.1 Выбор и предпочтения

*Универсальным множеством*  $A$  называют множество всех потенциально доступным для выбора альтернатив,  $|A| > 2$ . *Предъявление*  $X$  представляет из себя непустое конечное подмножество универсального множества  $A$ ,  $X \subseteq A$ . В рассматриваемой задаче выбора предъявление является переменной величиной. Выбор есть разделение предъявления на два подмножества – подмножество выбранных и подмножество отвергнутых альтернатив. *Функция выбора* или *решение задачи выбора*  $S(X)$  – это функция, которая каждому предъявлению ставит в соответствие подмножество выбранных альтернатив  $S \subseteq X$ .

Выбор из пары альтернатив однозначно определяет на  $A$  полное, рефлексивное, и антисимметричное бинарное отношение  $\geq$ , представляющее нестрогие предпочтения выбирающего. Асимметричная часть  $>$  этого отношения является полным, антирефлексивным и асимметричным бинарным отношением, репрезентирующим строгие предпочтения субъекта выбора. Симметричную часть  $\sim$  отношения  $\geq$  составляют упорядоченные пары равноценных альтернатив. Без потери общности можно считать, что в этом случае выбираются обе альтернативы.

Естественным решением задачи выбора является множество *максимальных элементов* сужения бинарного отношения  $\geq_X$  на предъявление  $X$ . Однако хорошо известно, что в рассматриваемой задаче коллективного выбора максимальных элементов отношения  $\geq_X$  может не существовать (парадокс Кондорсе).

#### 1.2.1.2 Турнир и турнирные решения

*Турниром*  $T$  называют пару предъявления  $X$  и сужения отношения  $>$  на это предъявление,  $(X, >_X)$ . *Слабым турниром*  $W$  называют пару  $(X, \geq_X)$ .

Функция выбора  $S$ , которая каждому турниру ставит в соответствие некоторое множество выбора  $S(T) \subseteq X$  (использующиеся далее обозначения  $S(T)$ ,  $S(X, >_X)$ ,  $S(X, >)$ ,  $S(>_X)$ ,  $S(X)$  являются эквивалентными), называется *турнирным решением*, если она удовлетворяет следующим трём свойствам:

- 1)  $\forall T, S(T) \neq \emptyset$  (непустота);

2)  $\forall T, MAX(\gtrsim_X) \subseteq S(T)$ , где  $MAX(\gtrsim_X)$  – множество максимальных элементов  $\gtrsim_X$  (согласованность по Кондорсе);

3)  $S(\pi(X)) = \pi(S(X))$ , где  $\pi(X)$  – произвольная перестановка альтернатив из  $X$  (Нейтральность).

### 1.2.1.3 Свойства турнирных решений

Турнирные решения оценивают и сравнивают по наличию у них важных свойств, отсутствие которых заставляет сомневаться в разумности и оптимальности такого способа выбора.

#### 1. Монотонность.

Пусть турниры  $\gtrsim_X$  и  $\gtrsim'_X$  совпадают для всех пар, не содержащих альтернативу  $a$ :  $\gtrsim_{X \setminus \{a\}} = \gtrsim'_{X \setminus \{a\}}$ . Пусть  $\forall b \in X \setminus \{a\}, (a \gtrsim b \Rightarrow a \gtrsim' b) \wedge (a > b \Rightarrow a > b)$ . Тогда решение  $S$  является *монотонным*, если из  $a \in S(\gtrsim_X)$  следует, что  $a \in S(\gtrsim'_X)$ .

#### 2. Устойчивость.

Турнирное решение  $S$  является *устойчивым*, если для любых двух предъявлений  $Y, Z \subseteq A$  таких, что  $Y \cap Z \neq \emptyset$ , имеет место эквивалентность:

$$S(Y, \gtrsim_Y) = S(Z, \gtrsim_Z) = X \Leftrightarrow X = S(Y \cup Z, \gtrsim_{Y \cup Z}). \quad (1.1)$$

Это утверждение можно рассматривать как конъюнкцию прямой и обратной импликаций. Прямая импликация в работе [18] называется свойством  $\hat{\gamma}$ , а обратная – свойством  $\hat{\alpha}$ .

Помимо самого свойства устойчивости далее нам потребуются определения и некоторых связанных с ним свойств. Турнирное решение  $S$  является *идемпотентным*, если  $S(S(X)) = S(X)$ . Турнирное решение  $S$  удовлетворяет *свойству Айзермана – Александрова*, если  $S(X) \subseteq Y \subseteq X \Rightarrow S(Y) \subseteq S(X)$ .

#### 3. Композиционная согласованность.

Орграф, репрезентирующий турнир, можно разделить на компоненты. Альтернативы  $x, y \in X$  принадлежат одной *компоненте* орграфа тогда и только тогда, когда они одинаково относятся ко всем остальным альтернативам из других компонент. Турнирное решение называется *композиционно-согласованным*, если для любых турниров  $T, T_1, \dots, T_n$ , и турнира компонент  $\tilde{T}$  таких, что  $T = \prod(\tilde{T}, T_1, \dots, T_n)$ , верно

$$S(T) = \bigcup_{i \in S(\tilde{T})} S(T_i). \quad (1.2)$$

#### 1.2.1.4 Непокрытое множество ( $UC$ ) и его обобщения

Непокрытое множество основано на отношении покрытия, для определения которого требуется ввести понятия верхнего и нижнего срезов и горизонта альтернативы. *Верхний срез*  $D(x) = \{y \in X | y > x\}$ , *нижний срез*  $L(x) = \{y \in X | x > y\}$ , *горизонт*  $H(x) = \{y \in X \setminus \{x\} | x \sim y\}$ .

Альтернатива  $x$  покрывает альтернативу  $y$  (обозначим это отношение как  $xCy$ ), если  $(x > y) \wedge (L(y) \subset L(x))$  [19]. Альтернатива называется *непокрытой* тогда и только тогда, когда в предъявлении нет ни одной покрывающей ее альтернативы. *Непокрытое множество*  $UC$  есть множество всех непокрытых альтернатив предъявления,  $UC(X, >) = \{x \in A | \forall y \in X, \neg yCx\}$ .

Существует 15 различных версий определения отношения покрытия [20], а следовательно, и непокрытого множества.

#### 1.2.1.5 Минимальное покрывающее множество ( $MC$ ) и его обобщения

В случае, если рассматриваемый турнир является полным, *покрывающим множеством*  $Y \subseteq X$  называют множество, удовлетворяющее следующим условиям

- 1)  $UC(Y) = Y$ ;
- 2)  $\forall x \in X \setminus Y, x \notin UC(Y \cup \{x\})$ .

*Минимальным покрывающим множеством*  $MC$  [15] называют единственное минимальное по включению покрывающее множество. Известно, что минимальное покрывающее множество является монотонным, устойчивым, композиционно-согласованным и вычислительно простым [21].

Согласно оригинальному определению [15; 18], в слабом турнире *покрывающим по версии V множеством*  $Y \subseteq X$  называют множество, удовлетворяющее следующим двум условиям:

- 1)  $UC_V(Y) = Y$ ;
- 2)  $\forall x \in X \setminus Y, x \notin UC_V(Y \cup \{x\})$ , где  $V \in \{S, M, IM, F, IF, McK, IMcK, D\}$ .

Минимальное покрывающее по некоторой версии множество определяется как единственное минимальное по вложению покрывающее по соответствующей версии отношения покрытия множество.

Существенным минусом этого определения является его излишняя требовательность. В [21], а также на рисунке 1.1 показано, что из-за данного недостатка обобщения минимального покрывающего множества, определенные через некоторые версии непокрытого множества, могут быть пустыми.

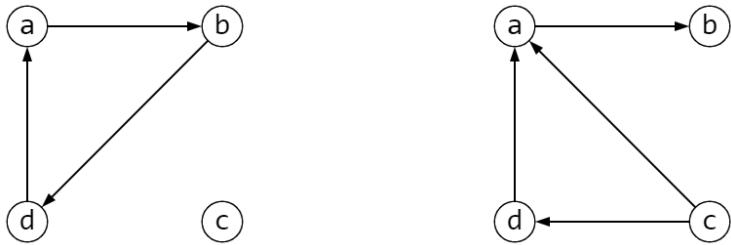


Рисунок 1.1 –  $MC_{IF}$  и  $MC_{IMcK}$  могут быть пустыми

Помимо оригинального определения обобщений минимального покрывающего множества существует определение, основанное на понятии внешней устойчивости [22; 23]. Согласно ему, в слабом турнире *покрывающим по версии V множеством*  $Y \subseteq X$  называют такое множество, что  $\forall x \in X \setminus Y, x \notin UC_V(Y \cup \{x\})$ , где  $V \in \{S, M, IM, F, IF, McK, IMcK, D\}$ .

Минимальное покрывающее по некоторой версии множество определяется как объединение всех минимальных по вложению покрывающих по соответствующей версии отношения покрытия множеств.

Новое определение позволяет обобщить минимальное покрывающее множество даже с помощью версий отношения покрытия, которые ранее могли давать пустое решение.

#### 1.2.1.6 Другие турнирные решения и их обобщения

Помимо определенных ранее турнирных решений в работе также использовались обобщения других популярных турнирных решений на случай слабых турниров, а именно: *объединение минимальных внешне устойчивых множеств*  $ES$  [24; 25; 26; 27] и обобщение *двухпартийного множества*  $BP$  [28], называемое *существенным множеством*  $E$  [29]. Также в исследовании использовались вторая и третья версии правила Коупланда, согласно которым альтернативы расставляются в порядке убывания значений оценок Коупланда  $s_2$  и  $s_3$  соответственно, где  $s_2(x) = |L(x)|$ ,  $s_3(x) = |X| - |D(x)|$ .

#### 1.2.2 Свойства версий минимального покрывающего множества

Относительно свойств введенных ранее обобщений минимального покрывающего множества известно, что  $MC_{McK}$  и  $MC_D$  являются устойчивыми, а  $MC_{McK}$  – монотонным [30] и вычислительно простым решением [31]. По аналогии с  $MC_{McK}$  можно показать, что  $MC_D$  также удовлетворяет свойству монотонности и является вычислительно простым. Другим версиям минимального покрывающего множества в литературе уделено не так много внимания. В частности, не было доказано, удовлетворяют ли они интересующему нас свойству устойчивости. Результат проверки выполнения связанных с устойчивостью свойств, описанных в прошлом разделе, приведен в теореме 1.5.

Теорема 1.5.

- 1)  $MC_S$  не удовлетворяет свойству идемпотентности, свойству Айзermana – Алексерова, а также свойству  $\hat{\alpha}$  и  $\hat{\gamma}$ .  $MC_S$  не является устойчивым турнирным решением.
- 2)  $MC_M$  не удовлетворяет свойству идемпотентности, свойству Айзermana – Алексерова, а также свойству  $\hat{\alpha}$  и  $\hat{\gamma}$ .  $MC_M$  не является устойчивым турнирным решением.
- 3)  $MC_{IM}$  не удовлетворяет свойству идемпотентности, свойству Айзermana – Алексерова, а также свойству  $\hat{\alpha}$  и  $\hat{\gamma}$ .  $MC_{IM}$  не является устойчивым турнирным решением.
- 4)  $MC_F$  удовлетворяет свойству идемпотентности.  $MC_F$  не удовлетворяет свойству Айзermana – Алексерова, а также свойству  $\hat{\alpha}$  и  $\hat{\gamma}$ .  $MC_{IM}$  не является устойчивым турнирным решением.
- 5)  $MC_{IF}$  не удовлетворяет свойству идемпотентности, свойству Айзermana – Алексерова, а также свойству  $\hat{\alpha}$  и  $\hat{\gamma}$ .  $MC_{IF}$  не является устойчивым турнирным решением.
- 6)  $MC_{McK}$  удовлетворяет свойству идемпотентности, свойству Айзermana – Алексерова, а также свойству  $\hat{\alpha}$  и  $\hat{\gamma}$ .  $MC_{McK}$  является устойчивым турнирным решением.
- 7)  $MC_{IMcK}$  не удовлетворяет свойству идемпотентности, свойству Айзermana – Алексерова, а также свойству  $\hat{\alpha}$  и  $\hat{\gamma}$ .  $MC_{IMcK}$  не является устойчивым турнирным решением.
- 8)  $MC_D$  удовлетворяет свойству идемпотентности, свойству Айзermana – Алексерова, а также свойству  $\hat{\alpha}$  и  $\hat{\gamma}$ .  $MC_D$  является устойчивым турнирным решением.

В Таблице 1.1 символ “+” означает, что турнирное решение, расположенное в столбце, удовлетворяет свойству,енному в соответствующей строке. Знак “–” означает, что турнирное решение, расположенное в столбце, не удовлетворяет свойству,енному в соответствующей строке.

Таблица 1.1 – Свойства обобщений МС

	$MC_S$	$MC_M$	$MC_{IM}$	$MC_F$	$MC_{IF}$	$MC_{McK}$	$MC_{IMcK}$	$MC_D$
Идемпотентность	–	–	–	+	–	+	–	+
Свойство Айзermana-Алексерова	–	–	–	–	–	+	–	+
Свойство $\hat{\alpha}$	–	–	–	–	–	+	–	+
Свойство $\hat{\gamma}$	–	–	–	–	–	+	–	+
Устойчивость	–	–	–	–	–	+	–	+

Заметим, что  $MC_{IM}$  и  $MC_{IMcK}$  не являются турнирными решениями. Было выяснено, что они могут не содержать в себе множество максимальных элементов отношения коллективного предпочтения, называемое ядром.

Устойчивость, монотонность и вычислительная простота минимального покрывающего по МакКельви и минимального покрывающего по Дуггану множества позволили сделать вывод о целесообразности использования данных обобщений для решения практических задач. Ввиду отсутствия устойчивости другие версии минимального покрывающего множества далее не использовались.

### 1.2.3 Проектирование и разработка MVP

#### 1.2.3.1 Описание предметной области

Одной из задач в области управления в академической сфере является оценка работы ученых и сравнение их вкладов в науку. Данная оценка результативности обычно производится на основе уровня (престижа) научных журналов, в которых публикуется ученый. Решение обозначенной задачи крайне важно, ведь именно на уровне журналов обычно завязана система стимулирования научных сотрудников, а следовательно, и развития науки.

Сравнение научных журналов затрудняется существованием большого количества баз данных научного цитирования (Scopus, Web of Science, РИНЦ и т. д.) и наличием десятков несогласующихся и несравнимых библиометрических показателей, по-разному учитывающих имеющуюся информацию о научных журналах и особенностях областей науки, в которых они представлены. Очевидно, что для улучшения качества результирующей сортировки научных журналов важно одновременно использовать сразу несколько библиометрических показателей. Однако производить оценку уже на основе трех показателей довольно сложно, ввиду наличия противоречий в оценках.

Очевидная сложность как в построении теоретического обоснования, так и в практической реализации нетривиальных методов разрешения этих противоречий привела к отсутствию в открытом доступе информационных систем, позволяющих строить агрегированные ранжирования научных журналов на основе их библиометрических показателей.

#### 1.2.3.2 Описание MVP

В рамках настоящего исследования был разработан MVP библиометрической информационной системы (в виде функционирующего сайта), позволяющий производить оценку научных журналов на основе их библиометрических показателей [32]. В качестве способов решения обсуждаемой задачи были выбраны такие турнирные решения, так непокрытое по версии Миллера множество, объединение минимальных внешне устойчивых множеств, существенное множество, минимальное покрывающее по МакКельви множество

и минимальное покрывающее по Дуггану множество, а также вторая и третья версия правила Коупланда. Причина выбора минимального покрывающего по МакКельви и минимального покрывающего по Дуггану множества заключалась в их хороших аксиоматических свойствах, упомянутых в прошлом разделе. Другие же перечисленные турнирные решения были взяты из-за того, что ранее они уже использовались для решения подобных задач и демонстрировали хорошие результаты [17]. В обсуждаемом веб-приложении были реализованы эффективные алгоритмы вычисления всех указанных выше турнирных решений. Разработанное программное решение позволило автоматизировать процесс построения агрегированных ранжирований научных журналов. Стоит отметить, что принцип работы MVP опирается на методологию, описанную в статье [17]. В качестве меры корреляции в MVP используется коэффициент Кендалла  $\tau_b$ .

С исходным кодом созданного MVP можно ознакомиться по ссылке [33]. Каждая функция веб-приложения сопровождается строками документации (docstrings), описывающими ее предназначение, входные и выходные параметры.

Разработанное программное решение может быть использовано в качестве вспомогательного инструмента, облегчающего принятие решений. В частности, конечными пользователями продукта могут выступать ученые, сотрудники высших учебных заведений, члены комиссий по грантам и т. п. Ученые могут использовать MVP как для выбора научных журналов для публикации своих статей, так и для анализа трендов развития интересующих их предметных областей, высшие учебные заведения же могут внедрять разработанную систему с целью повышения эффективности распределения денежных средств.

Основными преимуществами MVP являются его теоретическая обоснованность, простота и интуитивность интерфейса, а также высокий уровень автоматизации.

Помимо достоинств веб-приложения стоит выделить и его ограничения:

- 1) MVP может работать только с файлами весом до 200 Мб.
- 2) Для корректной работы MVP необходимо постоянное соединение с интернетом.
- 3) Веб-приложение совместимо с такими браузерами, как Яндекс Браузер, Google Chrome, Mozilla Firefox, Microsoft Edge и Safari. Использование MVP в других браузерах и/или на мобильный устройствах возможно, но не рекомендуется.
- 4) На данный момент интерфейс MVP прописан на русском языке. В дальнейшем планируется увеличение количества поддерживаемых языков.

### 1.2.3.3 Обоснование выбора инструментов

В качестве языка программирования для создания MVP был выбран Python. Данный язык программирования часто используется для научных вычислений. В качестве его основных преимуществ можно выделить простоту и интуитивность синтаксиса, разнообразие библиотек и фреймворков и высокую скорость разработки. Стоит отметить, что разрабатываемое веб-приложение было написано на версии Python 3.10 и не может быть скомпилировано с помощью более ранних версий интерпретатора.

В коде разработанной программы были использованы такие библиотеки, как pandas (2.0.1), numpy (1.24.3), scipy (1.10.1), XlsxWriter (3.1.0), openpyxl (3.1.2), pulp (2.7.0) и streamlit-ext (0.1.7).

Также для создания MVP был использован фреймворк Streamlit (1.22.0), позволяющий развертывать веб-приложения. Одним из основных преимуществ Streamlit является отсутствие необходимости самостоятельного создания и компоновки веб-элементов сайта, что значительно ускоряет разработку.

### 1.2.4 Эмпирическое сравнение методов ранжирования

Для оценки эффективности минимального покрывающего по МакКельви множества и минимального покрывающего по Дуггану множества как инструментов выбора проведено сравнение результатов их применения в прикладной задаче с результатами, которые дают другие турнирные решения. Для решения поставленной задачи использовалось веб-приложение, описанное в прошлом разделе.

Как и в любой области, связанной с рейтингами, в библиометрии существует огромное количество различных способов оценки журналов. Каждый год исследователи придумывают все новые и новые показатели, постоянно усложняя их в попытка нивелировать недостатки классических индексов. Отсутствие согласованности этих рейтингов заметно затрудняет сравнение научных журналов.

Целесообразность использования турнирных решений в задаче построения агрегированных ранжирований применительно к библиометрическим данным уже была показана в статье [17]. Данное исследование использует схожий подход, но несколько расширяет его благодаря добавлению новых турнирных решений.

#### 1.2.4.1 Библиометрические показатели

Данный параграф посвящен описанию шести наиболее популярных библиометрических показателей, а именно: (двухлетнего) импакт-фактора, пятилетнего импакт-фактора, индекса влияния статьи, SNIP, SJR и CiteScore. Полученные с их помощью рейтинги являлись входными данными в рассматриваемом примере.

#### 1.2.4.1.1 Двухлетний импакт-фактор (2YIF) и пятилетний импакт-фактор (5YIF)

Импакт-фактор, введенный в [34] является одним из самых известных библиометрических показателей. Импакт-фактор журнала рассчитывается как среднее количество цитирований  $CIT$  в расчетный год  $T$  на статью  $PUB$ , опубликованную в этом журнале в промежутке от  $T - n$  до  $T - 1$ ,

$$IF = \frac{\sum_{t=1}^n CIT(T, T-t)}{\sum_{t=1}^n PUB(T-t)}. \quad (1.4)$$

Задействованный в определении параметр  $n$  принято называть публикационным окном. На практике публикационное окно обычно берут равным двум или пяти годам, тем самым получая двухлетний импакт-фактор и пятилетний импакт-фактор соответственно.

Несмотря на простоту как понимания, так и подсчета, у импакт-фактора также есть серьезные недостатки. Например, импакт-фактор не учитывает, что в разных областях науки статьи цитируются по-разному. В частности, задержка в цитировании может возникать в связи с особенностями способов получения новых результатов. Описанный недостаток подтверждает важность одновременного использования импакт-факторов с разными значениями публикационного окна.

Оба рассматриваемых в работе импакт-фактора рассчитываются на основе базы данных научного цитирования Web of Science (WoS) и ежегодно публикуются в Journal Citation Report (JCR).

#### 1.2.4.1.2 CiteScore (CS)

Введенный в 2016 году показатель CiteScore представляет из себя разновидность импакт-фактора. В случае CiteScore публикационное окно берется равным 4 годам, причем окно цитирования тоже равно 4. Таким образом CiteScore находится как среднее количество цитирований  $CIT$  за расчетный период от  $T - 3$  до  $T$  на статью  $PUB$ , опубликованную в этом журнале в этом же промежутке,

$$CS = \frac{\sum_{m=0}^3 \sum_{t=m}^3 CIT(T-m, T-t)}{\sum_{t=0}^3 PUB(T-t)}. \quad (1.5)$$

Стоит отметить, что в работе описан обновленный метод нахождения CiteScore. До 2019 года данный показатель рассчитывался абсолютно аналогично импакт-фактору с публикационным окном в 4 года.

В отличие от показателей, рассмотренных в прошлом пункте, CiteScore рассчитывается на основе данных из Scopus.

#### 1.2.4.1.3 SNIP (source normalized impact per paper)

Введенный в 2009 году индекс SNIP [35] призван исправить некоторые несовершенства импакт-фактора.

Вычисление SNIP начинается с нахождения импакт-фактора с публикационным окном в 3 года, после чего полученный результат нормируются на специфический показатель, отражающий значение цитирования предметной области журнала. В качестве такого показателя принято использовать среднее количество источников в списке литературы всех цитирующих определенный журнал публикаций. Таким образом, цитирования работ в областях с небольшим средним списком источников имеют больший вес и наоборот. Подобный подход интересен тем, что он производит некоторую классификацию областей науки без знания количества и примеров представителей исходных классов.

В отличие от импакт-фактора SNIP рассчитывается на основе базы данных научного цитирования Scopus, а результаты расчетов публикуются издательством Elsevier.

#### 1.2.4.1.4 SJR (SCImago Journal Rank)

Индекс SJR был предложен исследовательской группой SCImago. Особенностью обсуждаемого показателя является то, что в рамках его вычисления цитированиям присваивается вес, зависящий от влиятельности цитирующего журнала. Влиятельность журнала, в свою очередь, изменяется его цитируемостью. Данный подход является крайне популярным и, например, лежит в основе такого известного алгоритма, как PageRank. Публикационное окно при этом берется равным 3 годам, а цитирования берутся только за расчетный год.

Сам алгоритм является итеративным и хорошо описан в [36]. На первой итерации всем журналам присваивается одинаковое значение авторитетности (влиятельности). Далее на каждой итерации эти значения обновляются в зависимости от авторитетности ссылающихся на него журналов. Процесс повторяется до тех пор, пока изменения не станут незначительными.

Очевидно, что в графе цитирований могут быть петли. Они соответствуют так называемым самоцитированиям, то есть ситуациям, когда научный журнал ссылается на самого себя. Самоцитированием иногда злоупотребляют неблагонадежные журналы, поэтому при нахождении SJR слишком высокие показатели самоцитирования искусственно снижаются до 33% от всех цитирований, сделанных в этом научном журнале.

Важно отметить, что SJR журнала нормируется на количество статей, опубликованных в нем, что делает этот показатель независимым от объема.

Как и в случае SNIP, значения SJR рассчитываются на основе базы данных научного цитирования Scopus и публикуются издательством Elsevier.

#### 1.2.4.1.5 Индекс влияния статьи (AI)

Публикуемый с 2007 года индекс влияния статьи определяется как нормированная на объем научного журнала линейная свертка количества полученных цитирований с векторами из собственного вектора матрицы цитирований, соответствующего наибольшему собственному значению.

Так как на практике собственный вектор находится итеративно, то процедура нахождения индекса влияния статьи оказывается очень похожей на уже описанный выше способ вычисления SJR. Как и при расчете SJR цитирования берутся только за расчетный год, но в случае AI пубикационное окно принимается равным 5 годам. Отличия также есть и в отношении индексов к самоцитированиям журналов. Если при нахождении SJR их количество может быть занижено, но все равно принимается во внимание, то в AI самоцитирования не учитываются вовсе. Различаются и базы данных, на основе которых формируются показатели. Так AI строится на основе данных о научных журналах, рецензируемых в Web of Science (WoS).

#### 1.2.4.2 Описание данных

В качестве входных данных были взяты шесть рейтингов научных журналов по менеджменту, построенных по значениям наиболее популярных библиометрических показателей (двуухлетнему импакт-фактору, пятилетнему импакт-фактору, индексу влияния статьи, SNIP, SJR и CiteScore), рассчитанным для 2022 года. Данные были собраны из открытых источников. Так значения показателей, формирующихся для научных журналов, рецензируемых в Web of Science, были взяты из отчета JCR, оставшиеся же собирались из Scopus. Собранные данные были объединены в одну таблицу, после чего из нее были удалены научные журналы, у которых отсутствовало значение хотя бы по одному из показателей. Важно отметить, что названия некоторых научных журналов в двух используемых базах данных отличались. Например, в названии одного и того же журнала в Web of Science мог использоваться амперсанд “&”, в то время как в Scopus он был заменен на “and”. Подобные ситуации были разрешены вручную.

Итоговая таблица содержит в себе 177 научных журналов по менеджменту [Репозиторий проекта на GitHub]. В рассматриваемой совокупности журналов имеется 816 равнозначенные пары.

#### 1.2.4.3 Корреляционный анализ

С файлом, содержащим в себе результаты работы веб-приложения на данных из [33], можно ознакомиться по ссылке [37]. На рисунке 1.2 приведены значения коэффициента  $\tau_b$ , рассчитанные для отобранных научных журналов.

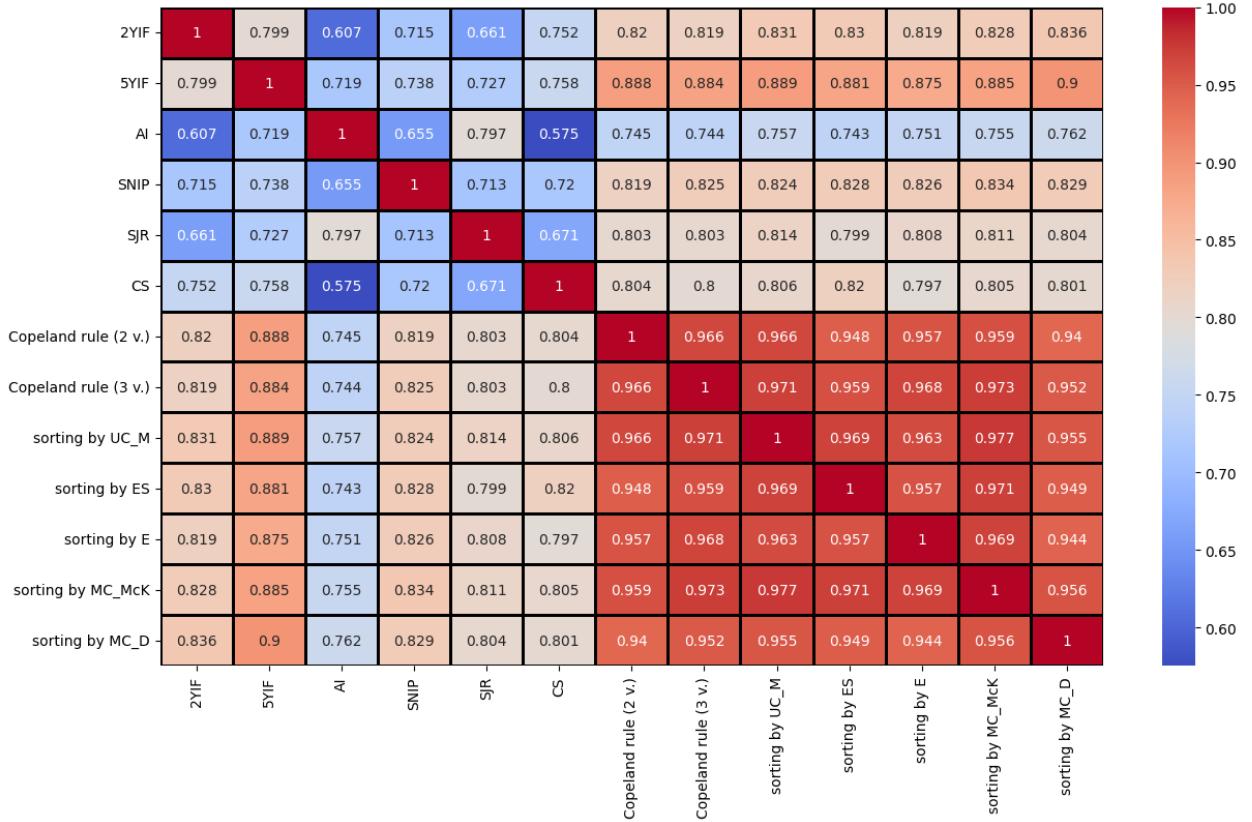


Рисунок 1.2 – Значения коэффициента  $\tau_b$

#### 1.2.4.4 Сравнение ранжирований

Ранжирования, исходные и агрегированные, сравнивались попарно и упорядочивались следующим образом. Вслед за авторами [17] предполагалось, что ранжирование  $I$ reprезентирует множество агрегируемых ранжирований лучше ранжирования  $II$ , если ранжирование  $I$  коррелирует с большинством из агрегируемых ранжирований сильнее, чем ранжирование  $II$ . Ранжирование  $I$  коррелирует с ранжированием  $III$  сильнее согласно коэффициенту  $\tau_b$ , чем ранжирование  $II$ , если значение  $\tau_b$  для пары  $\{I, III\}$  больше значения  $\tau_b$  для пары  $\{II, III\}$ . С помощью попарного сравнения было определено мажоритарное отношение на множестве ранжирований. Матрица строгого мажоритарного отношения, полученная на основе найденных значений коэффициента  $\tau_b$ , приведена в Таблице 1.2.

Таблица 1.2 – Матрица строгого мажоритарного отношения ( $\tau_b$ )

	2YIF	5YIF	AI	SNIP	SJR	CS	Copeland rule (2 v.)	Copeland rule (3 v.)	sorting by $UC_M$	sorting by $ES$	sorting by $E$	sorting by $MC_{McK}$	sorting by $MC_D$	Copeland score s2
2YIF	0	0	1	0	1	0	0	0	0	0	0	0	0	2
5YIF	1	0	1	1	1	1	0	0	0	0	0	0	0	5
AI	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SNIP	0	0	1	0	1	0	0	0	0	0	0	0	0	2
SJR	0	0	1	0	0	0	0	0	0	0	0	0	0	1
CS	0	0	1	0	1	0	0	0	0	0	0	0	0	2
Copeland rule (2 v.)	1	1	1	1	1	1	0	1	0	0	0	0	0	7
Copeland rule (3 v.)	1	1	1	1	1	1	0	0	0	0	0	0	0	6
sorting by $UC_M$	1	1	1	1	1	1	1	1	0	1	1	1	0	11
sorting by $ES$	1	1	1	1	1	1	0	0	0	0	1	0	0	7
sorting by $E$	1	1	1	1	1	1	0	1	0	0	0	0	0	7
sorting by $MC_{McK}$	1	1	1	1	1	1	1	1	0	1	1	0	0	10
sorting by $MC_D$	1	1	1	1	1	1	1	1	1	1	1	0	0	11

Для ранжирования имеющихся ранжирований к полученному мажоритарному отношению была применена вторая версия правила Коупланда. Таким образом альтернативы были расставлены в порядке убывания значений оценок Коупланда  $s_2$ . Ранжирование ранжирований, полученное на основе этих оценок, приведено в Таблице 1.3.

Таблица 1.3 – Ранжирование ранжирований ( $\tau_b$ )

Ранг	Метод
1	Сортировка с помощью $MC_D$ /
2	Сортировка с помощью $UC_M$
3	Сортировка с помощью $MC_{McK}$
4	Сортировка с помощью $E$ /
5	Сортировка с помощью $ES$ /
6	Правило Коупланда (2 версия)
7	Правило Коупланда (3 версия)
8	Пятилетний импакт-фактор (5YIF)
9	Двухлетний импакт-фактор (2YIF) /
10	SNIP /
11	CiteScore (CS)
12	SJR
13	Индекс влияния статьи (AI)

Проведенный корреляционный анализ показал, что обсуждаемые обобщения минимального покрывающего множества ( $MC_{McK}$  и  $MC_D$ ) представляют исходную совокупность агрегируемых рейтингов лучше, чем каждый из них, а в случае минимального покрывающего по Дуггану множества, еще и лучше, чем агрегированные ранжирования, построенные с помощью иных используемых в работе процедур выбора.

### 1.2.5 Заключение

В ходе проведенного исследования были рассмотрены различные способы обобщения минимального покрывающего множества на случай слабого турнира. Анализ наличия у них крайне важного свойства устойчивости показал непригодность обобщений минимального покрывающего множества, отличных от полученных через версии МакКельви и Дуггана отношения покрытия, для работы с реальными данными.

В рамках настоящего исследования было разработано и описано веб-приложение на языке программирования Python [32]. Данное приложение позволяет агрегировать противоречивые рейтинги научных журналов, основанные на различных библиометрических показателях, тем самым решая актуальную прикладную задачу из области управлением в академической сфере.

Также был приведен пример работы приложения на актуальных данных о значениях шести популярных библиометрических показателей статей, опубликованных в 2022 году для 177 журналов по менеджменту. Корреляционный анализ показал, что турнирные решения представляют исходную совокупность агрегируемых рейтингов лучше, чем каждый из них, а в случае минимального покрывающего по Дуггану множества, еще и лучше, чем агрегированные ранжирования, построенные с помощью иных процедур выбора.

## 1.3 Исследования по манипулированию процедур агрегирования в теории кол- лективного выбора

Продолжающееся исследование по анализу процедур агрегирования с точки зрения возможности манипулирования ими в проекте 2023 года осуществлялось в следующих направлениях:

- манипулирование при неполной информации и неопределенности относительно действий других избирателей;
- разработка структуры базы данных и сайта для хранения и визуализации данных по индивидуальной и коллективной манипулируемости процедур агрегирования;
- разработка алгоритмов распараллеливания задачи оценки манипулируемости процедур агрегирования;

- оценка соотношений между манипулируемостью различных процедур агрегирования для различных методов построения расширенных предпочтений;
- оценка соотношений между манипулируемостью q-Паретовских правил коллективного выбора.

### 1.3.1 Манипулирование при неполной информации и неопределенности относительно действий других избирателей

#### 1.3.1.1 Введение

Для модели манипулирования со стороны избирателей при неполной информации и неопределенности относительно действий других участников голосования, предложенной в 2021 году, ранее были получены значения вероятности манипулирования в случае трех альтернатив. Результаты показали, что для многих правил коллективного выбора значение этой вероятности становится нулевым, начиная с некоторого числа избирателей при публичной информации о победителях голосования. Можно ли теоретически показать, что это верно не только для трех альтернатив? В рамках проекта МЦАВР в 2023 году такой теоретический результат был получен для класса правил подсчета очков. В наивной модели манипулирования, Модели 1, стимул к манипулированию у избирателя не зависит от наличия стимула к манипулированию у других, в то время как в Модели 2 предполагается, что стимул к манипулированию пропадает в случае, если существует опасность получить менее предпочтительный исход при одновременном искажении предпочтений всеми остальными избирателями, имеющими стимул к манипулированию.

#### 1.3.1.2 Определения и обозначения

Пусть  $N = \{1, 2, \dots, n\}$  – множество избирателей, множество альтернатив –  $X$ ,  $|X| = m$ . Предпочтения избирателя  $i$ ,  $P_i$  – бинарное отношение, являющиеся линейным порядком,  $L(X)$  – множество всех линейных порядков. Профиль предпочтений всех избирателей обозначается за  $\mathbf{P} = (P_1, P_2, \dots, P_n)$ , а профиль предпочтений всех избирателей, кроме  $i$  – за  $\mathbf{P}_{-i}$ . Множество всех профилей предпочтений есть  $L(X)^N$ , включающее  $(m!)^n$  элементов. Отображение  $F : L(X)^N \rightarrow 2^X \setminus \emptyset$  называется правилом коллективного выбора. Если результат правила включает в себя более одной альтернативы, то применяется алфавитное правило устранения множественности выбора,  $T : 2^X \setminus \emptyset \rightarrow X$ . Согласно ему, на множестве  $X$  задается линейный порядок  $P_T$ , например,  $aP_TbP_Tc\dots$ , и если по правилу коллективного выбора было выбрано несколько альтернатив, то из них выбирается альтернатива, недоминируемая по  $P_T$ , т.е.  $T(A) = \{a \in A \mid \forall x \in A, x \neq a \ (a, x) \in P_T\}$ . Композиция функций  $F$  и  $T$ ,

т. е.  $T(F(\mathbf{P}))$ , обозначается как  $V(\mathbf{P})$ . Пусть  $v(a, \mathbf{P})$  – вектор распределения рангов альтернативы  $a$ ,  $j$ -ый элемент которого,  $v_j(a, \mathbf{P})$ , равен количеству избирателей, в предпочтениях которых альтернатива  $a$  стоит на  $j$ -ом месте. Пусть  $r_i(a, P_i) = |\{b \in A \mid aP_i b\}|$  – ранг альтернативы  $a$  в предпочтениях  $P_i$ .

Перед голосованием избиратели получают информацию о профиле предпочтений  $\mathbf{P}$ . Если информация полная, то они знают весь профиль. В случае неполной информации то, что известно избирателям о  $\vec{P}$  представлено в модели манипулирования с помощью функции публичной информации (ФПИ)  $\pi$ . При ФПИ 1Winner,  $\pi(\vec{P}) = V(\vec{P})$ .

Правило подсчета очков определяется вектором очков  $s = (s_1, \dots, s_m)$ , где  $s_j$  – количество очков, которое получает альтернатива за  $j$ -ю позицию в предпочтениях избирателя,  $s_1 \geq \dots \geq s_m$  и  $s_1 > s_m$ . Выбор по данному правилу определяется как множество альтернатив, для которых сумма очков по всем избирателям является максимальной, т. е.  $c \in F(\mathbf{P}) \Leftrightarrow c \in \arg \max_{a \in X} (\langle s, v(a, \mathbf{P}) \rangle)$ .

Пусть  $W_i^{\pi(\vec{P})}$  – информационное множество избирателя  $i$ , множество всех профилей предпочтений, согласующихся с имеющейся у избирателя  $i$  информацией  $\pi(\vec{P})$ . Формально,  $W_i^{\pi(\mathbf{P})} = \{\mathbf{P}'_{-i} \in L(X)^{N \setminus \{i\}} : \pi(P_i, \mathbf{P}'_{-i}) = \pi(\mathbf{P})\}$ .

### 1.3.1.3 Модели манипулирования

#### Модель 1

**Определение 1.1.** Избиратель  $i$  имеет стимул манипулировать в *Модели 1* при ФПИ  $\pi$  в профиле  $\vec{P}$  при правиле коллективного выбора  $F$ , если существует  $\tilde{P}_i$  такое, что 1) либо  $V(\tilde{P}_i, \mathbf{P}'_{-i}) = V(\mathbf{P})$ , либо  $V(\tilde{P}_i, \mathbf{P}'_{-i}) P_i V(\mathbf{P})$  для всех  $\mathbf{P}'_{-i} \in W_i^{\pi(\mathbf{P})}$ ; 2) существует  $\mathbf{P}''_{-i} \in W_i^{\pi(\mathbf{P})}$  такой, что  $V(\tilde{P}_i, \mathbf{P}''_{-i}) P_i V(\mathbf{P})$ .

Если в профиле  $\mathbf{P}$  существует хотя бы один избиратель, имеющий стимул манипулировать при информации  $\pi$  и правиле  $F$ , то профиль называется манипулируемым при ФПИ  $\pi$ . Множество всех избирателей, которые имеют стимул манипулировать в  $\mathbf{P}$  при  $\pi$ , обозначается за  $\Pi(\mathbf{P})$ .

В общем случае у избирателя, имеющего стимул манипулировать, может быть несколько стратегий манипулирования. Для простоты мы фиксируем одну стратегию для каждого избирателя и выбираем её по принципу наилучшего возможного результата.

Обозначим за  $\tilde{\mathbf{P}}$  профиль предпочтений, полученный из  $\mathbf{P}$ , в котором все избиратели из  $\Pi(\mathbf{P}) \setminus \{i\}$  манипулируют.

#### Модель 2

**Определение 1.2.** Избиратель  $i$  имеет стимул манипулировать в Модели 2 при ФПИ  $\pi$  в профиле  $\Pi(\mathbf{P})$  при правиле коллективного выбора  $F$ , если существует  $\tilde{P}_i$  такое, что

- 1) избиратель  $i$  имеет стимул манипулировать в Модели 1 при  $\pi$  в  $\mathbf{P}$  и при правиле  $F$ ;
- 2) либо  $V(\tilde{P}_i, \tilde{\mathbf{P}}'_{-i}) = V(P_i, \tilde{\mathbf{P}}'_{-i})$ , либо  $V(\tilde{P}_i, \tilde{\mathbf{P}}'_{-i}) > V(P_i, \tilde{\mathbf{P}}'_{-i})$  для всех  $\mathbf{P}'_{-i} \in W_i^{\pi(\mathbf{P})}$ .

Таким образом, стимул к манипулированию пропадает в случае, если существует опасность получить менее предпочтительный исход при одновременном искажении предпочтений всеми избирателями из  $\Pi(\mathbf{P}) \setminus \{i\}$ .

#### 1.3.1.4 Результаты

Основной результат исследования этого пункта содержится в следующей теореме. Пусть  $I^{M^2}(m, n, F, \pi)$  – вероятность, что в случайно выбиравом профиле предпочтений существует по крайней мере один избиратель, имеющий стимул для манипулирования в Модели 2 при  $\pi$  в  $\mathbf{P}$  и при правиле  $F$ .

**Теорема 1.6.** Для любого правила подсчета очков  $F$  и любого числа альтернатив  $m$  существует конечное число избирателей  $n^*$  такое, что для всех  $n > n^*$  имеет место  $I^{M^2}(m, n, F, 1\text{Winner}) = 0$ .

#### 1.3.1.5 Некоторые выводы

Мы показали, что если избирателям известна информация о победителях голосования после устранения множественности выбора, то для правил подсчета очков при любом количестве альтернатив более 3 существует такое число избирателей, при котором вероятность существования избирателя, имеющего стимул к манипулированию с помощью наивной стратегии равна нулю, если другие избиратели тоже манипулируют. Вычисления, сделанные в предыдущих исследованиях, показывают, что, скорее всего, такая ситуация имеет место и для других правил. Остается открытым вопрос, а существует ли вообще такая стратегия, которая позволяет манипулировать в условиях неопределенности относительно предпочтений и стимулов других избирателей, представляющий большой интерес и открывающий возможности для дальнейшего изучения стратегического поведения при голосовании.

Результаты проведенных исследований нашли отражение в [39] и [38].

**1.3.2 Разработка структуры базы данных и сайта для хранения и визуализации данных по индивидуальной и коллективной манипулируемости процедур агрегирования**

#### 1.3.2.1 Введение

В Международном центре анализа и выбора решений накоплен большой объем расчетных данных по индивидуальной и коллективной манипулируемости процедур агрегирования. В течение более 30 лет проведены расчеты более чем для 40 процедур агрегирования

для 10 различных индексов. Данные представляют собой многомерные массивы, визуализация и анализ которых представляет сложную задачу, для решения которой была разработана информационная система, представленная в этой работе.

### 1.3.2.2 Основная часть

В ходе данной работы разработаны структура реляционной базы данных (рисунок 1.3), позволяющая представить многомерные массивы данных по индивидуальной и коллективной манипулируемости процедур агрегирования в формате, оптимизированном для задачи визуализации.

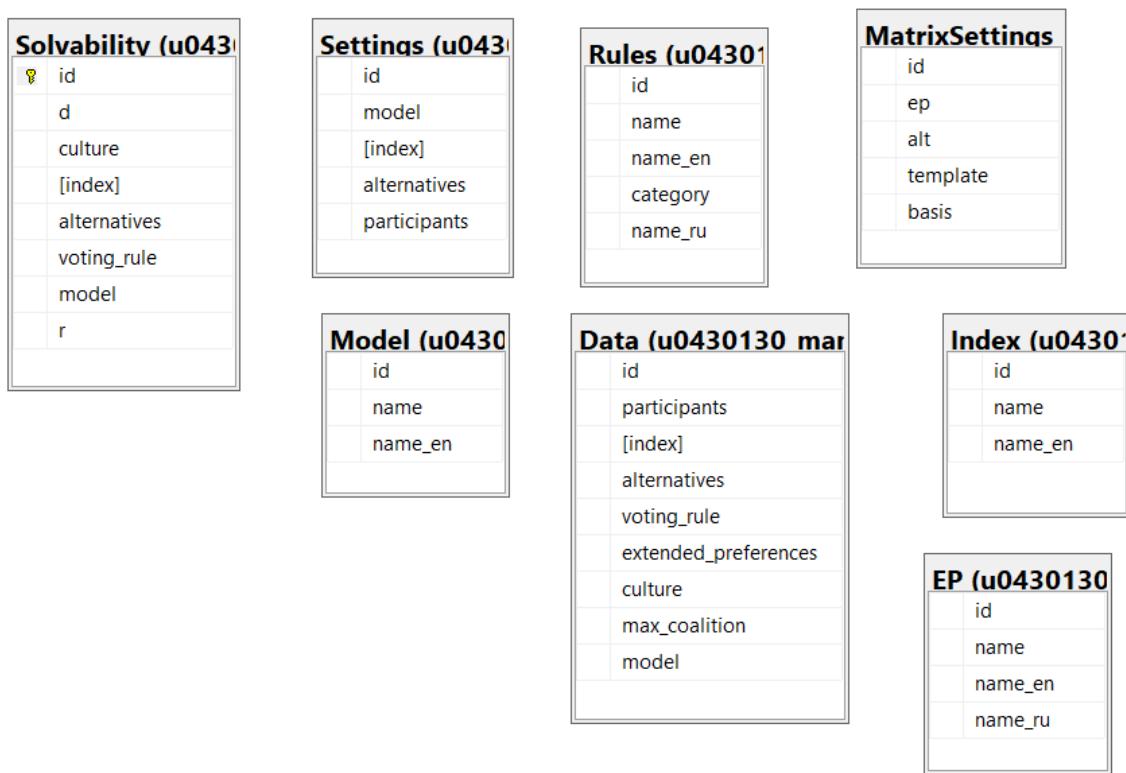


Рисунок 1.3 – Структура БД

Разработанная база данных состоит из 8 таблиц. Основные таблицы Data и Solvability содержат «плоское» представление исходных многомерных массивов. Поля Data.participants, Solvability.R и Solvability.D содержат массивы значений соответствующих индексов для всех количеств участников (как правило 98, от 3 до 100). Остальные таблица содержат справочники видов моделей, процедур агрегирования, расширенных предпочтений. Таблица MatrixSettings предназначена для хранения параметров для работы с матрицами переходов.

Основная страница портала (рисунок 1.4) содержит информацию о проекте и его участниках.

← → ⌂ demo.moscowlab.info

**HSE UNIVERSITY** Project: Manipulability of voting rules Ru

- Home page
- Main Notions
- Probabilistic Models
- Voting procedures
- Indices of manipulability
- Multi-valued Choice and Extended Preferences
- Results of voting procedures
- References
- Properties of voting rules
- Degree of Manipulability
- Decisiveness
- Resoluteness
- Freedom of Manipulability  
( $I_1; I_2^+, I_2^0, I_2^-$ )
- Efficiency of Manipulability
- Compare properties
- Manipulability Minimum
- Index Extremum

**Disclaimer:** Please note that this web site is still under development and some data may be not fully validated or incomplete!

## Manipulability of Social Choice Rules

The problem of manipulation in voting is that a voter can achieve a better social decision for herself by misrepresenting her sincere preference. Gibbard (1973) and Satterthwaite (1975) showed that for at least three alternatives and single-valued choice every non-dictatorial voting rule is individually manipulable. Later Duggan and Schwartz (2000) had generalized this result for the case of multiple choice (when there can be more than one alternative as the result of voting).

Then interesting question arises: if we know that every social choice rule is manipulable, how can we find the least manipulable one? We address this problem using several models of manipulation in single-valued and multiple-valued choice.

This web-portal is the result of the ongoing work conducted at the National Research University Higher School of Economics since 2006. Please navigate the menu on your left to explore our results.

### Submit your research

If you would like to make the results of your studies of the manipulability of voting rules available on this portal, please contact us.

### The Team

 **Fuad T. Aleskerov**  
Department Head, Professor: Faculty of Economic Sciences / Department of Mathematics  
Director: International Centre of Decision Choice and Analysis

Tenured Professor (2005)  
Distinguished Professor (2019)  
Member of the HSE Academic Council

Рисунок 1.4 – Главная страница портала

Портал содержит список публикаций и справочную информацию по существующим процедурам агрегирования (рисунок 1.5), индексам, вероятностным моделям формирования профилей предпочтений и расширенным предпочтениям. Используются средства визуализации математических формул.

← → ⌂ demo.moscowlab.info/MainNotions/VotingProcedures

- Probabilistic Models
- Voting procedures
- Indices of manipulability
- Multi-valued Choice and Extended Preferences
- Results of voting procedures
- References
- Properties of voting rules
- Degree of Manipulability
- Decisiveness
- Resoluteness
- Freedom of Manipulability  
( $I_1; I_2^+, I_2^0, I_2^-$ )
- Efficiency of Manipulability
- Compare properties
- Manipulability Minimum
- Index Extremum

## Voting procedures

We consider a finite set  $A$  consisting of  $m$  alternatives,  $m = 3, 4$ . Let  $A = 2^A \setminus \{\emptyset\}$  denote the set of all non-empty subsets of  $A$ . Each agent from a finite set  $N = \{1, \dots, n\}$ ,  $n > 1$ , is assumed to have a preference  $P_i \in L$  over alternatives where  $L$  is the set of linear orders on  $A$ . An ordered  $n$ -tuple of preferences  $P_i$  is called a (preference) profile,  $\vec{P}$ . A group decision is made by a social choice rule based on  $\vec{P}$  and is considered to be an element of  $A$ . Thus we define a social choice rule as a mapping  $C : L^n \rightarrow A$ . We consider 27 social choice rules. We divided them into 5 groups.

### Scoring Rules

#### 1. Plurality Rule.

Choose alternatives that are ranked first by the maximum number of agents, i.e.

$$a \in C(\vec{P}) \Leftrightarrow [\forall x \in A \quad n^+(a, \vec{P}) \geq n^+(x, \vec{P})]$$

where  $n^+(a, \vec{P}) = \text{card}\{i \in N \mid \forall y \in A \quad a P_i y\}$

#### 2. q-Approval rule

Let us define

$$n^+(a, \vec{P}, q) = \text{card}\{i \in N \mid \text{card}\{D_i(a)\} \leq q - 1\},$$

where  $D_i(a) = \{y \in A : y P_i a\}$  is the upper contour set of  $a \in A$  in  $P_i \in L$ . Let  $n^+(a, \vec{P}, q)$  be the number of agents for which  $a$  is ranked among the first  $q$  alternatives in their preference ordering. The integer  $q$  can be called as the degree of the procedure. We define q-Approval as follows

$$a \in C(\vec{P}) \Leftrightarrow [\forall x \in A \quad n^+(a, \vec{P}, q) \geq n^+(x, \vec{P}, q)],$$

i.e., the alternatives which are admitted to be among the  $q$  best by the highest number of agents are chosen. It can be easily seen that Plurality Rule is a special case of q-Approval where  $q = 1$ .

#### 3. Borda's Rule

Let  $r_i(x, \vec{P})$  be the cardinality of the lower contour set of  $x \in A$  in  $P_i \in L$ , i.e.  $r_i(x, \vec{P}) = |L_i(x)| = |\{b \in A : x P_i b\}|$ . The sum of  $r_i(x, \vec{P})$  over all  $i \in N$  is called the Borda score of alternative  $a$ .

$$r(a, \vec{P}) = \sum_{i=1}^n r_i(a, P_i).$$

Рисунок 1.5 – Раздел с описанием процедур агрегирования

Портал содержит многофункциональные средства визуализации (рисунки 1.6, 1.7), позволяющие гибко настроить параметры отображения данных, произвести экспорт данных и графика.

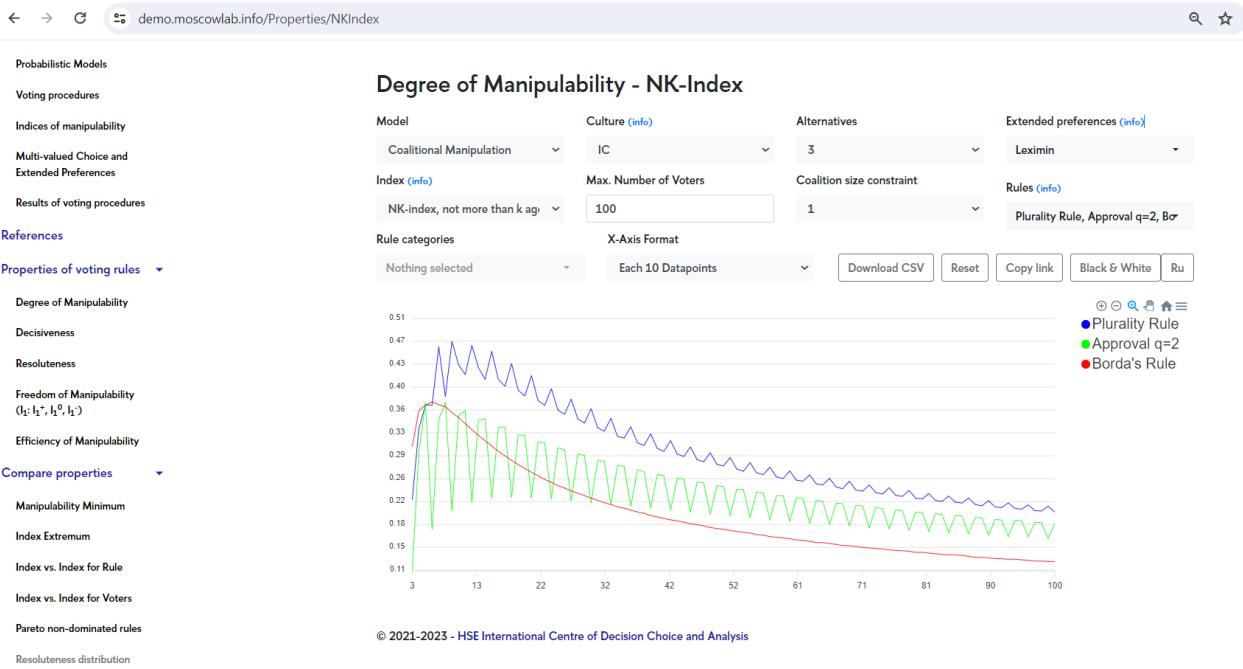


Рисунок 1.6 – Раздел визуализации данных манипулирования (NK-индекс)

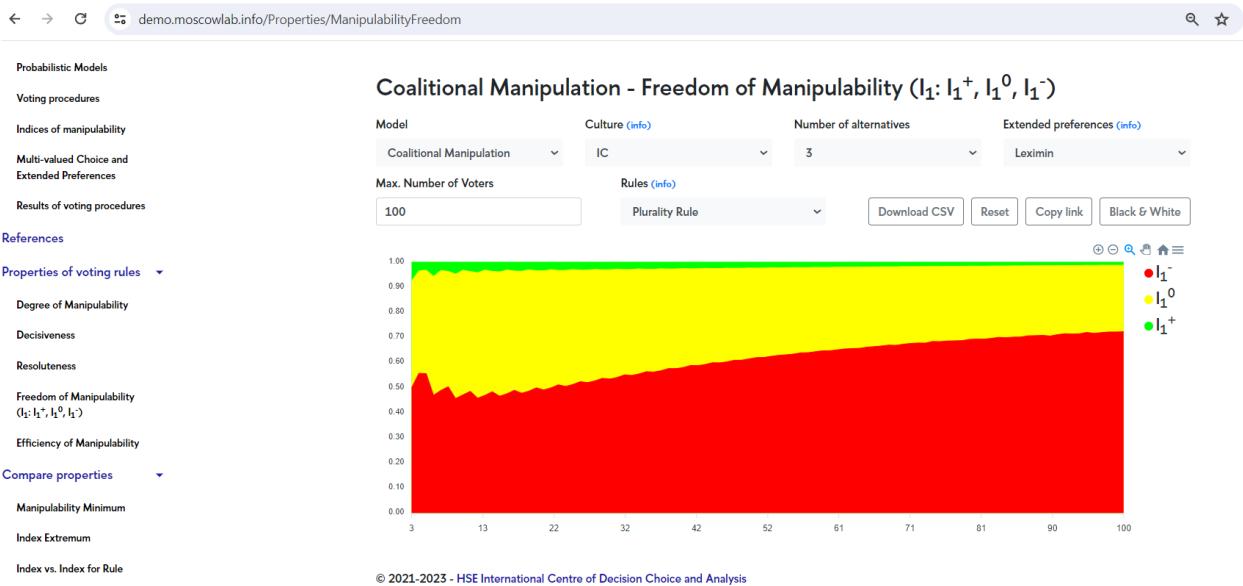


Рисунок 1.7 – Раздел визуализации данных индекса I<sub>1</sub>

Представлены разнообразные виды графиков, предназначенные для наиболее наглядной визуализации определенных данных.

Имеются средства табличного представления (рисунок 1.8), позволяющие отобрать процедуры агрегирования с наибольшим или наименьшим значением выбранного индекса для заданного набора параметров.

demo.moscowlab.info/Compare/Extremum

**Coalitional Manipulation - Index extremum**

Model	Culture	Extremum	Number of alternatives	Extended preferences
Coalitional Manipulati	IC	Minimum	3	Leximin
Coalition size constraint	Index	Rule categories	Rules	Maximum deviation
1	NK index, not n	Scoring Rules	Plurality Rule, Apr	0
<input type="button" value="Reset"/> <input type="button" value="Copy link"/>				
Agents/Voters	Rules	Extremum		
3	Black's Procedure, Nanson's Procedure, Inverse Borda's Procedure	0.0553		
4	Black's Procedure, Nanson's Procedure, Inverse Borda's Procedure	0.2365		
5	Nanson's Procedure, Inverse Borda's Procedure	0.0696		
6	Hare's Procedure	0.1698		
7	Nanson's Procedure, Inverse Borda's Procedure	0.0721		
8	Hare's Procedure	0.1778		
9	Nanson's Procedure, Inverse Borda's Procedure	0.0736		
10	Nanson's Procedure	0.2257		
11	Nanson's Procedure	0.0750		
12	Hare's Procedure	0.1798		
13	Nanson's Procedure	0.0756		

Рисунок 1.8 – Раздел табличного представления данных

Представлены средства сравнительного анализа индексов (рисунок 1.9) и процедур агрегирования (рисунок 1.10).

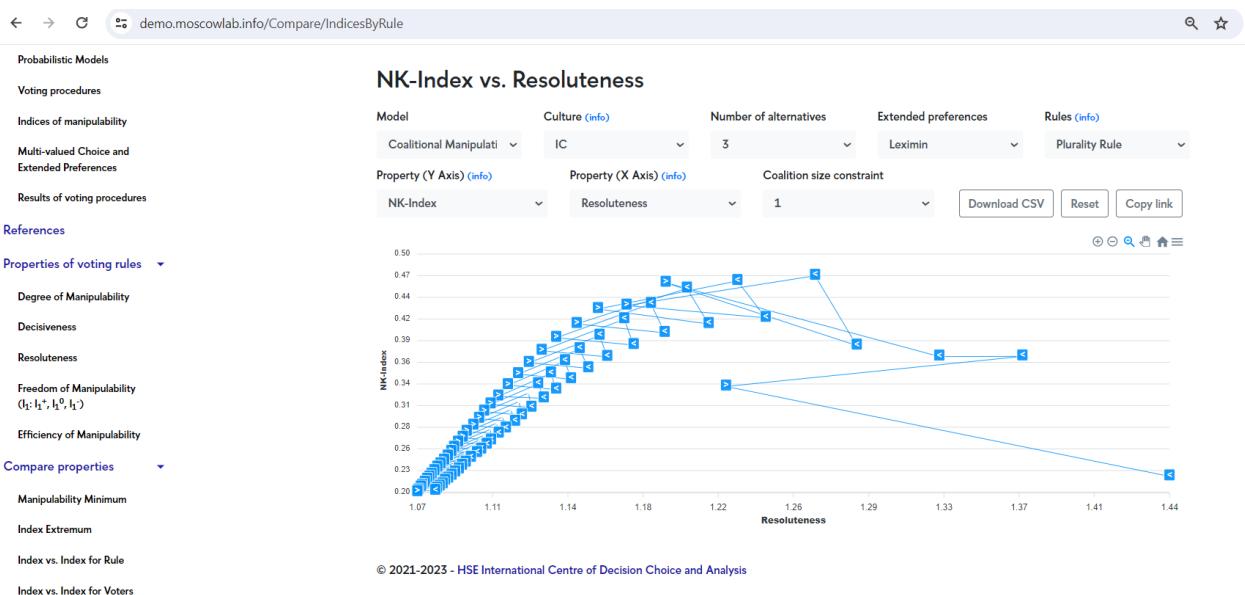


Рисунок 1.9 – Раздел сравнительного анализа индексов

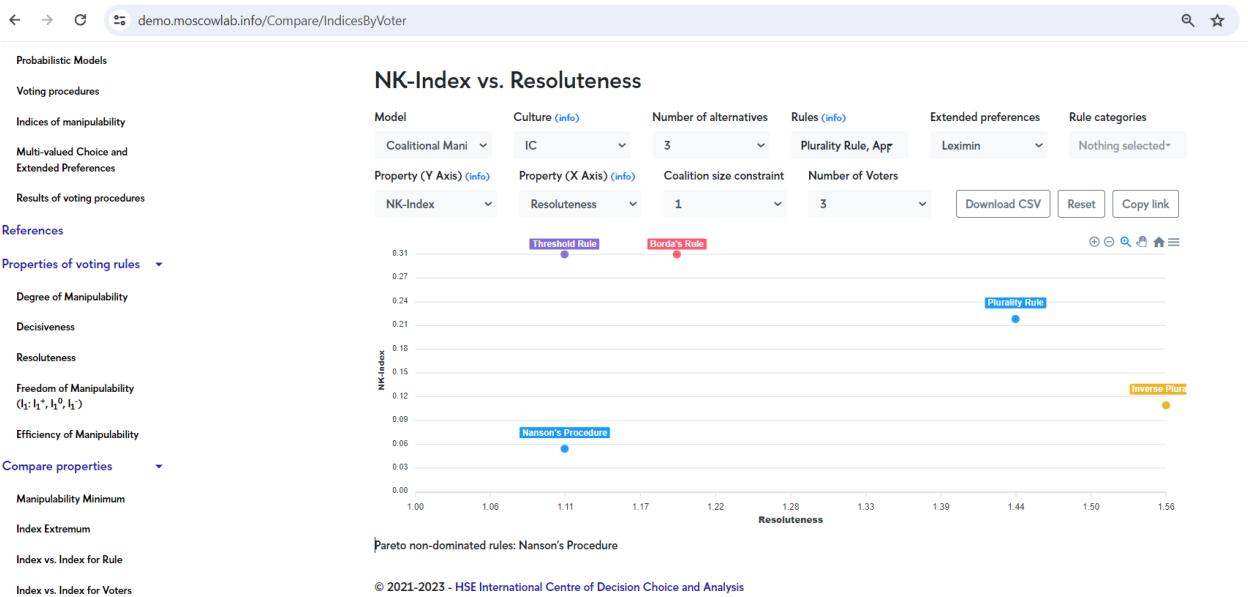


Рисунок 1.10 – Раздел сравнительного анализа индексов (относительно числа участников)

### 1.3.2.3 Заключение

В ходе исследования разработано полнофункциональное веб-приложение, содержащее как теоретические сведения по задачам исследования манипулируемости процедур агрегирования, так и интерактивные средства визуализации данных.

Версия приложения, содержащая функционал для работы с матрицами переходов, развернута на внутреннем сервере в сети ВШЭ. Версия, не содержащая эту функциональность, доступна в сети Интернет.

Разработанные средства визуализации применяются при проведении исследований и подготовке презентационных материалов.

### 1.3.3 Разработка алгоритмов распараллеливания задачи оценки манипулируемости процедур агрегирования

#### 1.3.3.1 Введение

Оценка манипулируемости процедур агрегирования представляет из себя задачу, предполагающую проверку профилей предпочтений, количество которых быстро возрастает с ростом числа участников и альтернатив.

Общее число профилей предпочтений для вероятностной модели Impartial Culture (IC) при расчете индекса манипулируемости рассчитывается как  $(m!)^n$ , где  $m$  – количество альтернатив, а  $n$  – количество участников. Очевидно, что полный перебор профилей для количества альтернатив более 5 и количества участников до 100 и более является чрезвычайно ресурсоемкой задачей, т.к. для проверки манипулируемости необходимо вычислить

значение процедуры агрегирования для всех вариантов манипулирования по каждому профилю. Для решения используются приближенные методы, когда расчет производится для 1 миллиона случайных профилей, однако даже при данном ограничении размерность задачи слишком высока для персональных компьютеров.

Представляется целесообразным проведения расчетов на суперкомпьютерном кластере НИУ ВШЭ.

### 1.3.3.2 Основная часть

В ходе исследования проанализированы существующие на данный момент последовательные реализации процедур агрегирования, разработанные ранее, а также правила формирования расширенных предпочтений, подхода к определению манипулируемости процедур агрегирования и расчета *индекса манипулируемости* (индекс Нитцана – Келли).

Установлено, что процедура расчета индекса манипулируемости обладает ресурсом распараллеливания. Для запуска процедуры на мощностях суперкомпьютерного кластера НИУ ВШЭ исходные программы на языке C# были переписаны на C++. Для компиляции, которая также происходит на узле кластера, используется предустановленный пакет Intel oneAPI Toolkit с установленными параметрами оптимизации. Данный подход, в соответствии с методическими материалами по работе с кластером, является рекомендуемым.

Так при компиляции рекомендуется указывать особые флаги оптимизаций, позволяющие ускорить выполнение программы. Флаг оптимизации -O2 активирует векторизацию и базовую оптимизацию циклов. Данный флаг в большинстве случаев обеспечивает наилучшее быстродействие работы программы.

При использовании программ с большим количеством вычислений с плавающей точкой, как в случае данного исследования, или обрабатывающих объёмные наборы данных рекомендуется использовать флаг -O3, активирующий более агрессивное преобразование циклов и условных выражений.

При компиляции с использованием компиляторов Intel также рекомендуется указывать флаг -xHOST. Данный флаг позволяет компилятору использовать максимальный набор инструкций, доступных для процессоров на суперкомпьютере. Данный флаг ещё больше оптимизирует приложение, но снижает его переносимость на другие платформы (на всех узлах вычислительного кластера НИУ ВШЭ используются процессоры одного поколения с одинаковым набором инструкций), что не является критичным в данном исследовании.

При распараллеливании задачи определения манипулируемости профиля предпочтений для группы участников возможно разделение набора (1 миллион) профилей на равное количество долей (по количеству используемых вычислительных ядер, эксперимен-

тальным путем установлено, что число ядер = 32 является оптимальным с точки зрения баланса между быстродействием и эффективностью обработки задачи планировщиком очереди кластера). Используются средства распараллеливания библиотеки STL C++ для создания потоков с помощью объектов класса std::thread.

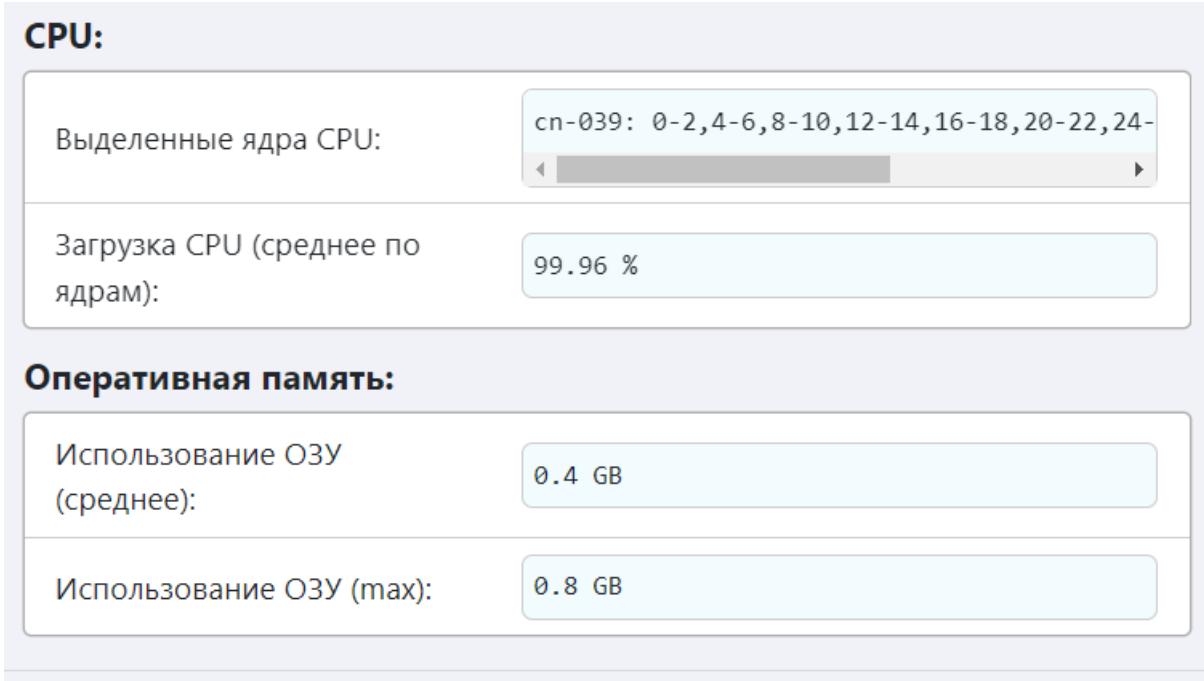


Рисунок 1.11 – Загрузка вычислительного узла

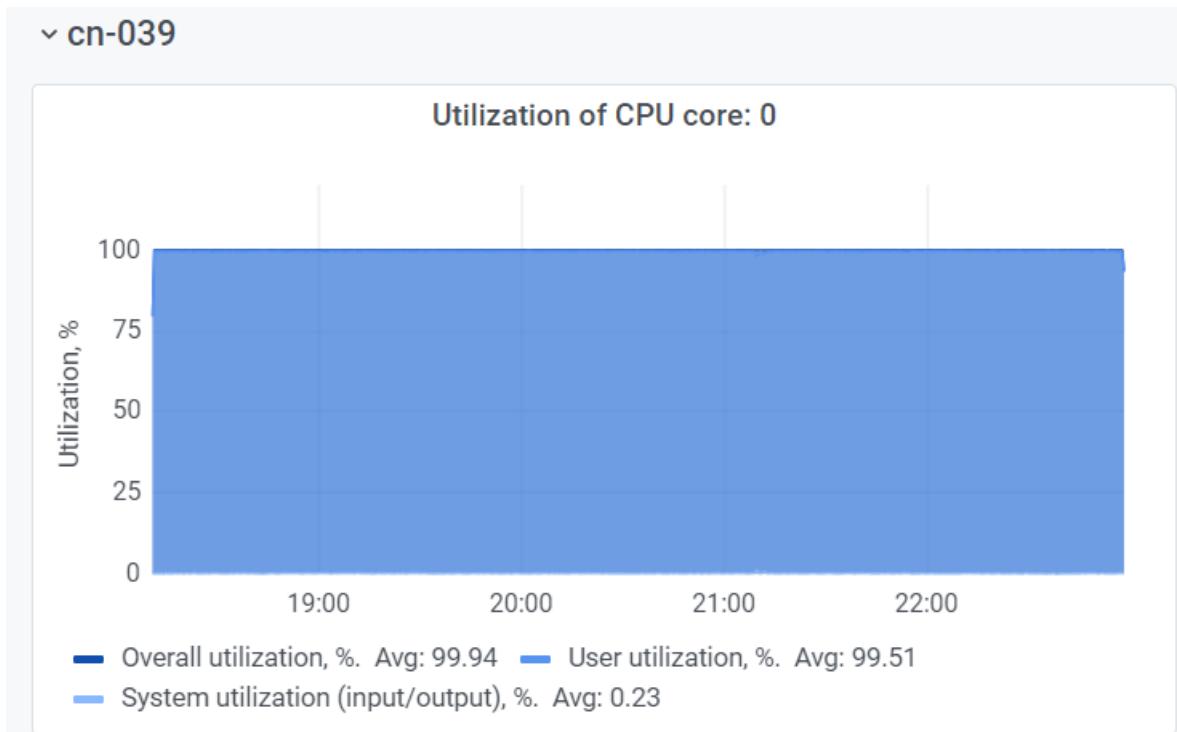


Рисунок 1.12 – Загрузка ядра

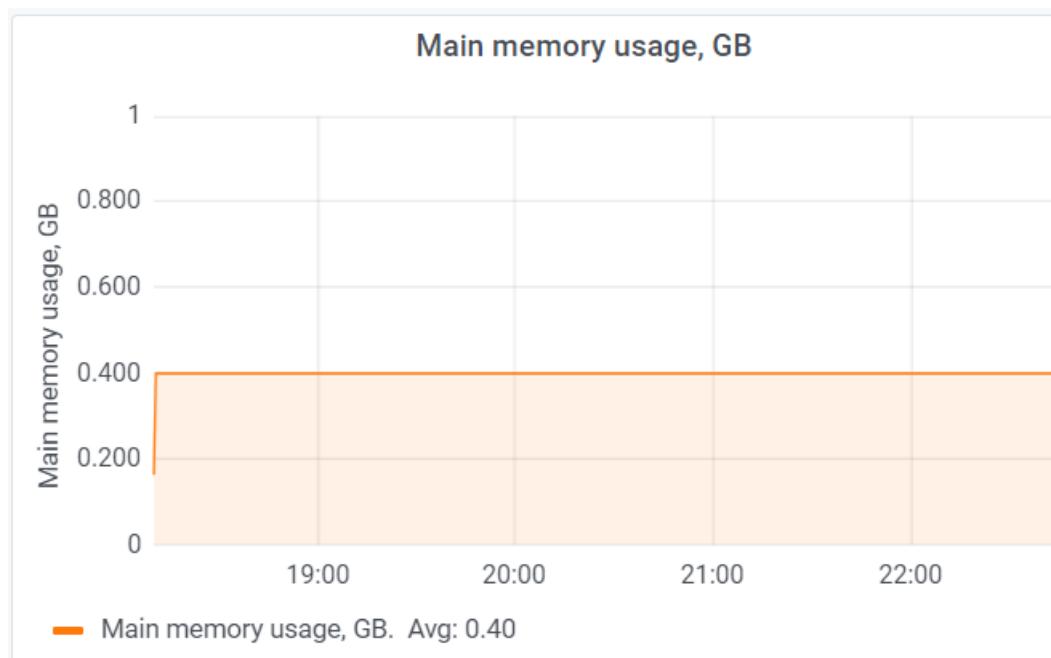


Рисунок 1.13 – Использование памяти

Особенность алгоритма расчета индекс Нитцана – Келли состоит в следующем: при проверке всех вариантов манипуляций для заданного профиля в случае, если хотя бы один дает выигрыш, то дальнейшая проверка не производится, профиль считается манипулируемым. Данных подход в случае последовательной программной реализации позволяет ускорить решений задачи.

Однако, в случае параллельной программной реализации, это приводит к тому, что некоторые ядра (потоки) завершают проверку своей части профилей быстрее остальных, что в среднем приводит к показателю загруженности 80% по всем ядрам.

Использование другого подхода (основанного на паттерне проектирования “Объектный пул”), при котором каждое вычислительное ядро забирает по одному профилю из набора, и, после завершения проверки манипулируемости, переходит к следующему не проверенному профилю, позволило повысить среднюю загрузку по всем ядрам до 99.8%.

Также в ходе исследования, в рамках решения задачи генерации множеств профилей и их хранения, была обнаружена зависимость между порядковым номером и составом профиля предпочтений группы участников для вероятностной модели Impartial Culture (IC). Реализован симметричный алгоритм расчета соответствия, позволяющий получить профиль по его порядковому номеру и наоборот.

Рассмотрим указанный алгоритм подробнее. Пусть  $r$  – число перестановок без повторений из  $m$  альтернатив, где каждая перестановка упорядочена в соответствии с предпочтениями  $n$ -го участника. Тогда профиль коллективных предпочтений представляет собой

целочисленный вектор, где в  $k$ -й позиции указан порядковый номер индивидуального предпочтения участника с номером  $k$ . Было установлено, что профиль предпочтений, сформированный таким образом, есть представление своего порядкового номера в смешанной (факториальной) системе счисления, по основанию, зависящему от числа альтернатив  $m$ .

### 1.3.3.3 Заключение

Благодаря использованию мощностей суперкомпьютерного кластера НИУ ВШЭ произведены расчеты для скоринговых (scoring) правил для количества альтернатив вплоть до 9 при количестве участников до 100.

Установленная зависимость между порядковым номером и составом профиля предпочтений группы участников позволяет осуществить хранение цифровых представлений профилей предпочтений значительно компактнее, чем использовавшийся ранее текстовый формат.

Программа для расчета степени манипулируемости для 10 скоринговых правил перенесена на язык программирования C++, оптимизирована и распараллелена с использованием возможностей библиотеки STL и Intel oneAPI Toolkit, что сделало возможными расчеты для случая более 5 альтернатив.

Использование данного стека технологий продиктовано ориентацией на выполнение расчетов на суперкомпьютерном кластере НИУ ВШЭ, оптимизированного для использования технологий компании Intel.

Исходные коды выложены в открытый доступ на общедоступном ресурсе GitHub.

Результаты этой части отчета отражены в [40].

### 1.3.4 Оценка соотношений между манипулируемостью различных процедур агрегирования для различных методов построения расширенных предпочтений

#### 1.3.4.1 Введение

Манипулирование – это ситуация, которая возникает при голосовании, когда участник или группа участников записывают в бюллетень неискренние предпочтения, чтобы достичь для себя более хорошего результата голосования.

Было доказано в работах [41, 42], что не существует недиктаторской неманипулируемой процедуры голосования. Позднее такой же вывод был показан [43] и для случаев множественного выбора, то есть для ситуаций, когда в процедуре голосования возможны ничьи между двумя и более альтернативами.

После этого возник вопрос: если не существует неманипулируемых процедур голосования, можно ли найти наименее манипулируемую процедуру голосования?

За последние десятилетия был опубликован целый ряд работ в области исследования процедур голосования на предмет манипулируемости. Такие работы затрагивали исследование как индивидуальной [44, 45], так и коалиционной манипулируемости [46]. В работах по исследованию манипулируемости чаще всего используются два подхода.

Первый – аналитический вывод формул. Для определенного правила голосования выводится формула для определенного индекса манипулируемости и определенного количества альтернатив. Ключевая проблема такого подхода в том, что для каждого правила, каждой модели и каждого индекса придётся заново выводить новую формулу. Кроме того, если для скоринговых правил (Правило относительного большинства, Правило Борда и т. д.) вывод формул представляется возможным, то для более сложных правил, например, основанных на мажоритарной матрице, или  $q$ -Паретовских правил, вывод формул выглядит как очень сложная аналитическая задача.

Второй подход – генерация случайного набора профилей и оценка на нём индексов манипулируемости [47]. Такой подход не позволяет получить точную оценку индексов манипулируемости, однако позволяет произвести оценку индексов манипулируемости с определённой погрешностью, которая при моделировании на большом количестве профилей (1 миллион) будет относительно маленькой: 0.001 по индексу манипулируемости Нитцана – Келли [48].

В этой работе мы используем второй подход: мы генерируем 1 миллион случайных профилей для каждого набора параметров, чтобы получить значения индекса манипулируемости Нитцана – Келли для различных процедур агрегирования. Далее мы строим соотношения индексов для различных комбинаций правил, чтобы найти закономерности.

#### 1.3.4.2 Модель манипулирования и основные обозначения

Мы рассматриваем модель, в которой  $n$  участников голосования выбирают среди  $m$  альтернатив. У каждого участника есть предпочтение, выраженное линейным порядком, на множестве альтернатив, то есть одно из  $m!$  предпочтений.

Профиль  $P$  – это совокупность участников голосования и их предпочтений. Процедура голосования  $C(P)$  ставит в соответствие заданному профилю результат процедуры голосования.

Если обозначить за  $P$  исходный профиль, а за  $P'$  профиль, где манипулирующий участник или группа участников предъявили неискренние предпочтения, то манипулирование считается успешным, если  $C(P') > C(P)$  для манипулирующего участника или группы участников, где знак  $>$  означает, что результат  $C(P')$  является более предпочтительным, чем  $C(P)$ .

Существует два подхода к тому, каким может быть результат голосования  $C(P)$ : это одиночный и множественный выбор. В рамках концепции одиночного выбора предполагается, что если между двумя или более альтернативами возникла ничья, то используется некоторый дополнительный механизм для определения победителя. Один из самых популярных вариантов – использование алфавитного (лексикографического) принципа устранения несравнимости. Как результат, при одиночном выборе процедура голосования даёт на выходе одну альтернативу – победителя. Например, если альтернативы  $\{a\}$  и  $\{b\}$  набрали по 4 голоса, и возникла ничья, то выбор будет  $\{a\}$  по алфавитному принципу.

В модели множественного выбора предполагается, что в случае возникновения ничьи между двумя или более альтернативами, все такие альтернативы будут входить в результат коллективного выбора. Например, если альтернативы  $\{a\}$  и  $\{b\}$  набрали по 4 голоса, то выбор будет  $\{a, b\}$ . То есть результатом процедуры голосования при множественном выборе может являться непустое подмножество множества альтернатив.

Таким образом, исходных предпочтений участника (например,  $a > b > c$ ) может оказаться недостаточно для сравнения результатов процедуры голосования до и после манипулирования, например, если до манипулирования результат был  $\{b\}$ , а после он стал  $\{a, c\}$ , то стал ли этот результат для участника лучше или хуже? Что лучше: получить гарантированно вторую лучшую альтернативу или лучше получить ничью между лучшей и худшой альтернативами?

Для ответа на этот вопрос в литературе используются расширенные предпочтения [49, 47], которые являются линейным порядком на множестве всех возможных результатов процедуры агрегирования. Для случая 3 альтернатив это 4 метода построения расширенных предпочтений (при предположении, что исходные предпочтения участника  $a > b > c$ ):

### 1) Лексимин

Результаты процедур агрегирования сравниваются лексикографически по наихудшей альтернативе. Выбор, у которого худшая альтернатива лучше, является более предпочтительным. Расширенные предпочтения Лексимин:  $\{a\} > \{a, b\} > \{b\} > \{a, c\} > \{a, b, c\} > \{b, c\} > \{c\}$ .

### 2) Лексимакс

Результаты процедур агрегирования сравниваются лексикографически по наилучшей альтернативе. Выбор, у которого лучшая альтернатива лучше, является более предпочтительным. Расширенные предпочтения Лексимакс:  $\{a\} > \{a, b\} > \{a, b, c\} > \{a, c\} > \{b\} > \{b, c\} > \{c\}$ .

### 3) Рискофил

Результаты процедур агрегирования сравниваются по вероятности наилучшей альтернативы в выборе. Выбор, у которого вероятность наилучшей альтернативы в множестве выше, является более предпочтительным. Расширенные предпочтения Рискофил:  $\{a\} > \{a, b\} > \{a, c\} > \{a, b, c\} > \{b\} > \{b, c\} > \{c\}$ .

### 4) Рискофоб

Результаты процедур агрегирования сравниваются по вероятности наихудшей альтернативы в выборе. Выбор, у которого вероятность наихудшей альтернативы в множестве ниже, является более предпочтительным. Расширенные предпочтения Рискофоб:  $\{a\} > \{a, b\} > \{b\} > \{a, b, c\} > \{a, c\} > \{b, c\} > \{c\}$ .

Таким образом, расширенные предпочтения позволяют сравнить все возможные результаты коллективного выбора. Например, если до манипулирования выбор был  $\{b\}$ , а после стал  $\{a, c\}$ , то при расширенных предпочтениях Лексимакс и Рискофил это будет успешное манипулирование, а при расширенных предпочтениях Лексимин и Рискофоб – неуспешное.

Манипулирование может осуществляться как одним участником, так и группой участников. Если манипулирует ровно один участник, то это индивидуальное манипулирование. Если манипулировать может группа участников сообща, то это коалиционное манипулирование.

Профиль считается манипулируемым, если существует хотя бы один участник или коалиция, которые могут провести успешное манипулирование. Профиль считается неманипулируемым, если не существует ни одной успешной попытки манипулирования.

#### 1.3.4.3 Индекс манипулируемости и схема расчётов

В литературе известно несколько индексов манипулируемости. Мы используем самый популярный из них – индекс Нитцана – Келли (NK-index). Он был введён в работах [50, 51]. Индекс Нитцана – Келли – это доля манипулируемых профилей в общем количестве профилей.

$$NK = \frac{\text{количество манипулируемых профилей}}{\text{общее количество профилей}}. \quad (1.6)$$

Алгоритм вычисления индекса NK:

- 1) Для каждого профиля сгенерировать все возможные попытки манипулирования

- 2) Если хотя бы одна попытка манипулирования является успешной, то есть  $C(P') > C(P)$  для манипулирующего участника или группы участников согласно их расширенных предпочтений, то профиль помечается как манипулируемый
- 3) Если для профиля не существует ни одной успешной попытки манипулирования, то он помечается как неманипулируемый

Стандартный механизм расчёта манипулируемости состоит из генерации профилей и проверки их на манипулируемость. Алгоритм состоит из нескольких шагов:

- 1) Генерация 1 миллиона случайных профилей
- 2) Для каждого профиля – генерация всех возможных попыток манипулирования, то есть всех возможных изменений предпочтений каждым участником
- 3) Подсчёт нового результата коллективного выбора для каждой попытки манипулирования каждого участника
- 4) Расчёт индекса  $NK$ : если хотя бы одна попытка манипулирования привела к лучшему результату, то профиль считается манипулируемым. Если для заданного профиля ни одна попытка манипулирования не привела к улучшению результата для манипулирующего участника, то профиль считается неманипулируемым

После проведения расчётов индекса  $NK$  для различных процедур агрегирования мыстроим графики с соотношениями индексов  $NK$  для различных правил.

#### 1.3.4.4 Результаты

Мы провели компьютерное моделирование и расчёт индекса  $NK$  для случаев от 3 до 100 участников для различных процедур голосования для 4 типов расширенных предпочтений (РП): Лексимин, Лексимакс, Рискофил, Рискофоб для случая IAC.

Ниже на рисунке 1.14 приведён пример графика индекса  $NK$  для Правила Относительного большинства и Процедуры Нансона для РП Лексимин:

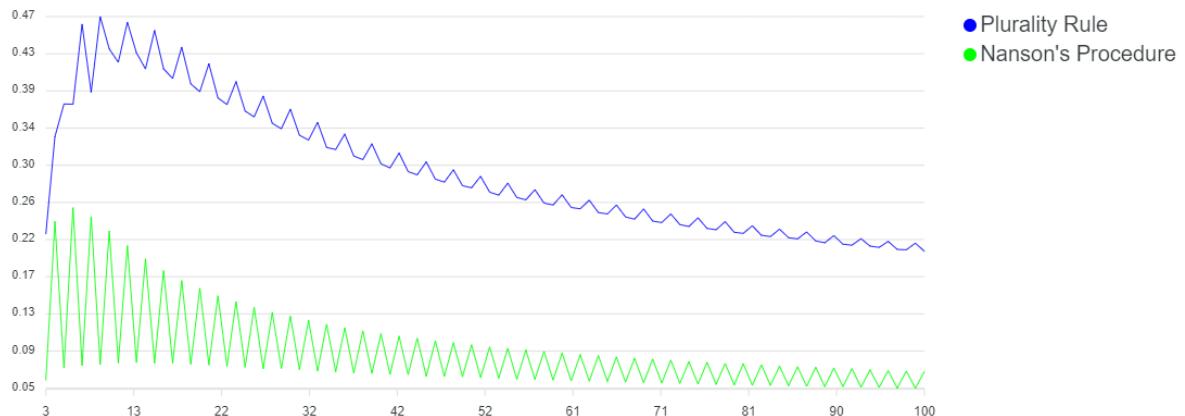


Рисунок 1.14 – Графики индекса  $NK$  для Правила Относительного большинства и Процедуры Нансона для РП Лексимин

На этом и других подобных графиках для других правил мы можем увидеть наличие колебаний в значениях индексов NK. Для Правила Относительного большинства цикл равен 3 участникам, для Процедуры Нансона цикл равен 2 участникам. При этом известно, что с ростом количества участников манипулируемость правил снижается.

Теперь приведём график сравнение относительных значений индексов NK для разных пар процедур агрегирования. На рисунке 1.15 приведены 5 процедур агрегирования: 3 наиболее известных (Правило Относительного большинства, Одобряющее голосование с  $q=2$  и Правило Борда), а также 2 наименее манипулируемых процедуры согласно предыдущим исследованиям из литературы, например, из [47] – Процедура Нансона и Процедура Хара. По горизонтальной оси представлено количество участников, по вертикальной – отношение индекса NK данной процедуры к индексу NK «эталонной» процедуры. В данном случае в качестве «эталонной» взята процедура Нансона как наименее манипулируемая в большинстве случаев.

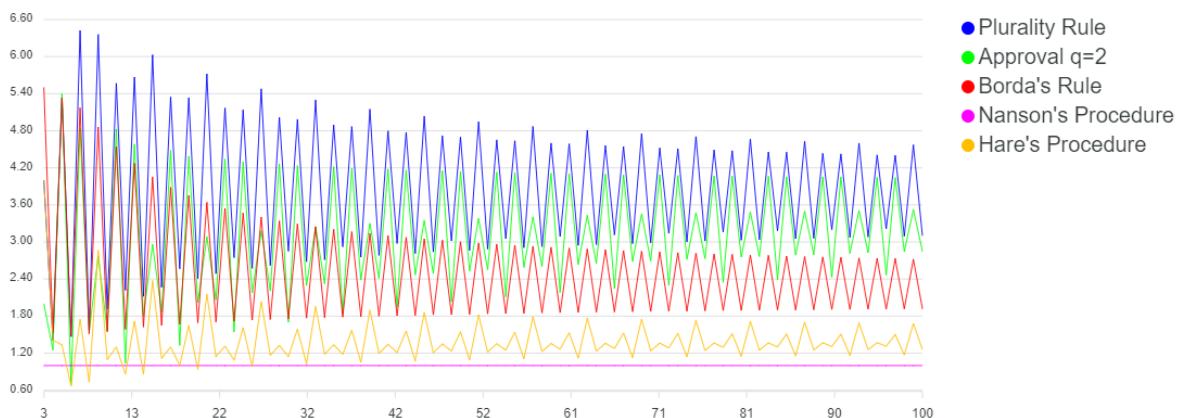


Рисунок 1.15 – Графики сравнение относительных значений индексов NK для 5 процедур агрегирования для РП Лексимин

Можно видеть, что соотношения индексов NK для небольшого количества участников сильно колеблются: например, для случая 5 участников Правило Относительного большинства и Одобряющее голосование с  $q=2$  более чем в 5 раз более манипулируемы, чем процедура Нансона, а для случая 6 альтернатив Правило Относительного большинства – лишь в 1.5 раза более манипулируемо, а Одобряющее голосование и вовсе в 1.5 раза менее манипулируемо.

Однако с ростом количества участников колебания заметно уменьшаются и наблюдается некоторая сходимость. Так, например, процедура Хара для большого количества участников в 1.17–1.68 раз более манипулируема, чем процедура Нансона, правило Борда – в 1.92–2.72 раз, Одобряющее голосование – в 2.85–3.52 раз, Правило Относительного большинства – в 3.1–4.57 раз.

Другая ситуация наблюдается для РП Лексимакс (те же 5 процедур представлены на рисунке 1.16):

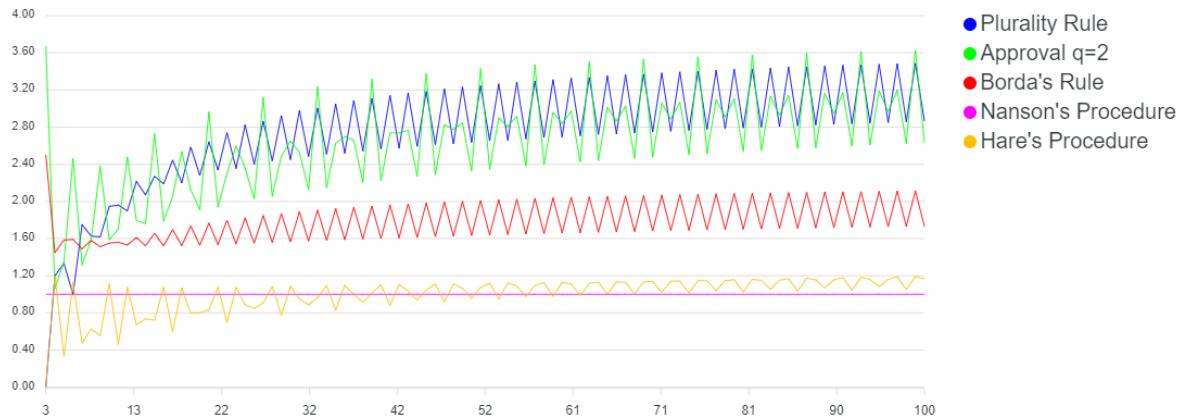


Рисунок 1.16 – Графики сравнение относительных значений индексов NK для 5 процедур агрегирования для РП Лексимакс

Здесь мы видим постепенное снижение манипулируемости «эталонной» процедуры (вновь процедуры Нансона) относительно других процедур. Процедура Хара для небольших значений количества участников была менее манипулируема, чем процедура Нансона, однако начиная с 63 участников становится постоянно более манипулируемой. Для больших количеств участников процедура Хара в 1.05–1.2 раза более манипулируема, чем процедура Нансона, Правило Относительного большинства – в 2.86–3.48 раз, Правило Борда – в 1.73–2.11 раз, Одобряющее голосование с  $q=2$  – в 2.63–3.62 раза.

Для РП Рискофоб график (рисунок 1.17), а как следствие, и выводы, очень похож на ситуацию с РП Лексимин:



Рисунок 1.17 – Графики сравнение относительных значений индексов NK для 5 процедур агрегирования для РП Рискофоб

А для РП Рискофил график (рисунок 1.18) похож на график для РП Лексимакс:

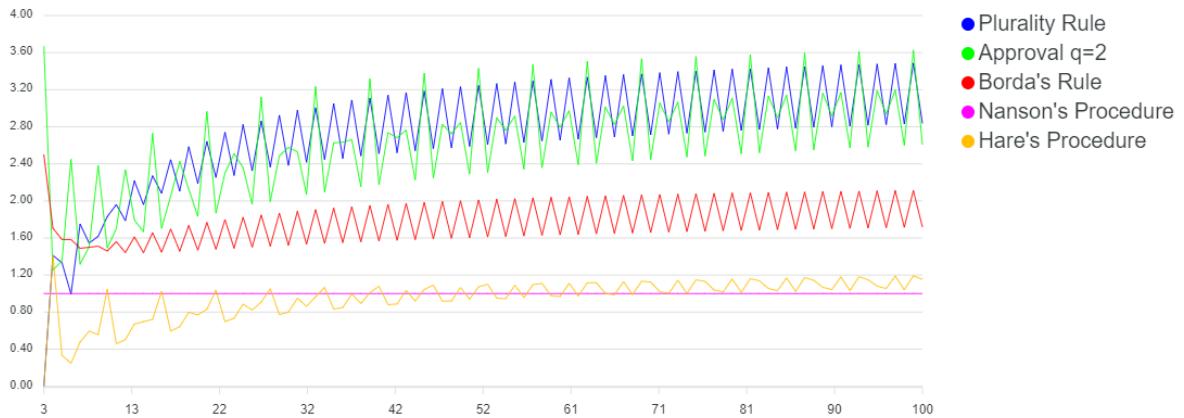


Рисунок 1.18 – Графики сравнение относительных значений индексов NK для 5 процедур агрегирования для РП Рискофил

Подобные графики и сравнения позволяют увидеть поведение индексов манипулируемости для различных процедур агрегирования при росте количества участников. Например, наглядно видно, что процедура Нансона может не быть наименее манипулируемой на небольшом количестве участников, однако с ростом количества участников она становится наименее манипулируемой. Для случая РП Лексимин и Рискофоб – с 27 участников, для РП Лексимакс и Рискофил – с 63 участников.

Кроме того, для РП Лексимин и Рискофоб при росте количества участников наблюдается сходимость соотношения между индексами NK между процедурами. Такая гипотеза заслуживает внимания и рассмотрения на большем количестве участников в качестве направления для будущих исследований.

### 1.3.5 Оценка соотношений между манипулируемостью q-Паретовских правил коллективного выбора;

В этом пункте отчета приведены результаты расчета степени индивидуальной и коалиционной манипулируемости для процедур агрегирования: Strong q-Pareto simple majority rule, Strong q-Pareto plurality rule, Strongest q-Pareto simple majority rule, Condorcet Practical rule. В рассмотренных правилах параметр  $q$ , характеризующий нечувствительность правила, формируется эндогенно, начиная с наименьшего значения, и увеличивается до тех пор, пока правила не начнут возвращать непустой выбор.

Приведем определения рассмотренных правил.

#### 1. Strong q-Pareto simple majority rule.

Пусть  $f(\vec{P}; i, q) = \left\{ x \in A \mid \text{card}(D_i(x)) \leq q \right\}$  и  $\mathfrak{I} = \left\{ I \subset N \mid \text{card}(I) = \left\lceil \frac{n}{2} \right\rceil \right\}$  – семейство коалиций простого большинства. Определим функцию:  $C(A) = \bigcup_{I \in \mathfrak{I}} \bigcap_{i \in I} f(\vec{P}; i, q)$ .

Определим выбор, как состоящий из наилучших ( $q=1$ , первое место в упорядочении) альтернатив для каждого участника в, по крайней мере, одной коалиции простого большинства. Если таких альтернатив не существует, то определим выбор, как состоящий из альтернатив, стоящих на первом или втором местах ( $q=2$ ). Значение  $q$  увеличивается до тех пор, пока выбор не станет непустым.

## 2. Strong q-Paretian plurality rule.

Правило совпадает с правилом Strong q-Paretian simple majority rule, если выбор одиночный. Если же в выборе содержатся несколько альтернатив, то подсчитывается число коалиций простого большинства, содержащих каждую из альтернатив. В выборе содержатся альтернативы, входящие в максимальное число таких коалиций.

## 3. Strongest q-Paretian simple majority rule.

Определим функцию  $C(X) = \bigcap_{I \in \mathfrak{I}} f(\vec{P}; I, q)$ , где  $f(\vec{P}; I, q) = \{x \in A \mid \text{card} \left( \bigcap_{i \in I} D_i(x) \right) \leq q\}$

$$\text{card}(I) = \left\lceil \frac{n}{2} \right\rceil.$$

Выбор состоит из альтернатив, Парето-оптимальных в каждой коалиции простого большинства при  $q=0$ . Если таких альтернатив не существует, то  $q$  увеличивается на 1, пока выбор не станет непустым.

## 4. Condorcet Practical rule

Пусть  $f(\vec{P}; i, q) = \{x \in A \mid \text{card}(D_i(x)) \leq q\}$ , где  $D_i(x) = \{y \in A \mid y P_i x\}$  – верхний контур альтернативы  $x$  для участника  $i$ . Определим выбор, как функцию  $C(A) = \bigcap_{i \in N} f(\vec{P}; i, q)$  при  $q = 0$ . Если таких альтернатив нет, то  $q$  увеличивается на 1 до тех пор, выбор не станет непустым.

В рамках исследования был проведен расчет для следующих моделей манипулирования:

- индивидуальное манипулирование; В случае индивидуального манипулирования предполагается, что только один участник искажает свои предпочтения, чтобы получить лучший для себя результат коллективного выбора.

- коалиционное манипулирование, в котором в коалицию входят участники с одинаковыми предпочтениями и в качестве неискренних предъявляют также одинаковые предпочтения;

– коалиционное манипулирование, в котором коалиция продвигает одну альтернативу, ставя ее на первое место в своих предпочтениях (модель same-first). В этом варианте коалиционного манипулирования предполагается, что коалиция может состоять не только из участников с одинаковыми предпочтениями, но и из участников с разными предпочтениями, если им выгодно продвигать одну и ту же альтернативу.

Для оценки степени манипулируемости использовались следующие индексы.

Индекс Ницана – Келли (NK), определяемый, как доля манипулируемых профилей в общем числе рассмотренных профилей.

Кроме того, рассчитываются три индекса свободы манипулирования. Такие индексы рассчитываются, как нормированное число попыток манипулирования, которые приводят к, соответственно, улучшению, равенству или ухудшению результата коллективного выбора.

Также рассчитываются индексы эффективности манипулирования в двух вариантах, как средний прирост и как максимальный прирост в упорядочении альтернатив коллективного выбора. Среди всех успешных попыток манипулирования рассчитывается выгода от манипулирования с точки зрения количества мест в упорядочении, на которое поднялась альтернатива (или множество альтернатив) в результате такого манипулирования. Затем вычисляется средний, и, соответственно, максимальный прирост в упорядочении альтернатив.

И, наконец, рассчитывается индекс разрешимости (D), равный среднему числу альтернатив в коллективном выборе. Для каждого профиля рассчитывается мощность коллективного выбора, затем это значение усредняется. Индекс D может варьироваться от 1 до числа альтернатив  $m$ . Значения, близкие к 1, показывают, что правило обладает высокой разрешимостью и чаще возвращает однозначный выбор. Если значение D близко к числу альтернатив  $m$ , то правило имеет плохую разрешающую способность и чаще всего возвращает полный набор альтернатив. Идея индекса состоит в том, что если определенное правило часто возвращает полный набор альтернатив в качестве коллективного выбора, то оно будет иметь низкие показатели манипулируемости не из-за того, что оно само по себе не подвержено манипулированию, и из-за того, что оно почти всегда возвращает один и тот же набор альтернатив, что почти бесполезно с практической точки зрения.

Процесс оценки степени манипулируемости выглядит следующим образом.

Генерируются случайные 1 млн профилей. Каждое правило рассматривается отдельно, для каждого профиля рассчитывается множественный коллективный выбор по каждому правилу. Мощность коллективного выбора используется для расчета индекса разре-

шимости D. Каждая из 3-х моделей манипулирования рассматривается отдельно. Для каждого профиля для каждой модели манипулирования генерируются все возможные попытки манипулирования, т. е. для случая индивидуального манипулирования, генерируются все попытки каждого участника, для случая коалиционного манипулирования, генерируются все попытки всех рассматриваемых коалиций. Результаты манипулирования сравниваются, учитывая расширенные предпочтения.

Каждый способ построения расширенных предпочтений рассматривается отдельно (4 способа для  $m=3$  альтернатив, 10 способов для  $m=4$  альтернатив, и 12 способов для  $m=5$  альтернатив).

Для каждой попытки манипулирования подсчитывается результат для профиля с неискренним предпочтением. Результат попытки манипулирования – лучше искреннего результата, ситуация равенства, или ситуация ухудшения – учитывается при расчете индекса свободы манипулирования.

Если для профиля существует хотя бы одна успешная попытка манипулирования, он помечается как манипулируемый для данной модели и для данного правила. В противном случае он помечается как не подверженный манипулированию. На основании этого рассчитывается индекс Ницана – Келли (как доля таких манипулируемых профилей). Для расчета индекса свободы манипулирования учитывается число способов, которыми манипулирование может быть достигнуто. Для расчета индексов эффективности манипулирования учитывается то, на сколько мест в упорядочении в коллективном выборе поднялось то или иное множество альтернатив.

В результате расчетов получены следующие результаты.

Для случая 3-х альтернатив, для количества участников меньше 8, наименее манипулируемыми становятся разные правила. Начиная же с числа участников  $n=8$ , рассмотренные правила отличаются друг от друга по значениям индекса NK, и возможно расположить их от наименее до наиболее манипулируемого:

- 1) Condorcet Practical rule (самые низкие значения NK),
- 2) . Strong q-Paretian simple majority rule,
- 3) Strongest q-Paretian simple majority rule,
- 4) . Strong q-Paretian plurality rule.

Для числа альтернатив  $m=4$  и  $m=5$ , для большинства случаев, манипулируемость правил выглядит следующим образом:

- 1) Condorcet Practical rule (самые низкие значения NK),
- 2) Strongest q-Paretian simple majority,
- 3) . Strong q-Paretian plurality,

#### 4) Strong q-Paretian simple majority.

Следует также отметить, что для Condorcet Practical rule наблюдаются высокие значения индекса D=m. Это означает, что в большинстве случаев Condorcet Practical rule возвращает полный выбор, т. е. разрешимость правила крайне низка.

Кроме того, следует отметить, что причина различий в индексе NK для правила Strong q-Paretian plurality для различного количества альтернатив также в низкой разрешимости. Для 3 альтернатив это правило показывает высокие значения индекса D (низкая разрешающая способность), и это второе наименее манипулируемое правило. Но для 4 и 5 альтернатив, при сопоставимой с другими правилами разрешимости, его индекс NK является наибольшим.

По результатам исследования опубликована работа [52].

### **1.4 Доверительное агрегирование и ранжирование в порядковой шкале**

#### **1.4.1 Введение**

Одной из важных задач теории принятия решений является задача ранжирования альтернатив, представленных оценками от ряда агентов (экспертов) в той или иной шкале измерений и (возможно) с учетом некоторой дополнительной информации (надежности оценок, статистических данных о прошлых ранжирований и т. д.). При решении этой задачи часто используется подход построения агрегированных индексов относительно каждой альтернативы с последующим ранжированием таких индексов. В этой работе будем рассматривать задачу агрегирования информации и ранжирования альтернатив в случае ординальной (порядковой) шкалы измерений. Существуют ряд методов агрегирования в таких шкалах. Так для агрегирования индивидуальных предпочтений используются методы теории коллективного выбора [53].

При построении правил агрегирования желательно учитывать степень противоречивости информации, поступающей от разных источников (например, экспертов), степень ее неопределенности, надежность этих источников.

Все эти особенности (противоречивость, неопределенность, надежность) хорошо моделируются в теории свидетельств (функций доверия) [54, 55]. Но в теории свидетельств альтернативы выбираются из некоторого множества, порядок элементов в которых не определен. В последнее время структуры доверия стали рассматриваться и на упорядоченных базовых множествах, на структурах предпочтений [56, 57, 58]. В частности, в [57] вводятся и рассматриваются функции доверия на парных сравнениях альтернатив из двух источников и вводится мера различия между парными сравнениями альтернатив из разных источников.

В [58] вводятся и рассматриваются основные конструкции теории свидетельств (функции доверия, правила комбинирования, метрики на множестве функций доверия и т. д.) на множестве всех подмножеств упорядоченных последовательных элементов некоторого базового множества.

В нашем исследовании была рассмотрена задача агрегирования информации от экспертов относительно альтернатив и последующее ранжирование этих альтернатив в том случае, когда информация представлена в некоторой порядковой шкале. Предполагалось, что помимо самих оценок известны их надежности. Основная идея состоит в представлении информации от экспертов в виде тела свидетельств. Задача агрегирования информации от экспертов решается с помощью применения правил комбинирования свидетельств. И наконец задача ранжирования альтернатив решается с помощью ранжирования соответствующих агрегированных тел свидетельств.

Решение этой задачи проиллюстрировано на примере ранжирования статей по результатам оценок рецензентов и с учетом степеней уверенности рецензентов в своих оценках. Такие процедуры ранжирования могут быть реализованы, например, в системах управления конференциями типа *EasyChair*.

#### 1.4.2 Необходимые сведения из теории свидетельств

Напомним некоторые необходимые сведения из теории свидетельств [55]. Пусть:

$T = \{t_1, \dots, t_s\}$  — некоторое конечное множество,

$2^T$  — множество всех подмножеств из  $T$ ,

$m: 2^T \rightarrow [0,1]$ ,  $\sum_{A \in 2^T} m(A) = 1$ ,  $m(\emptyset) = 0$  — базовое доверительное назначение (функция масс),

$A = \{A\}$  — множество всех фокальных элементов, т. е.  $A \in A$ , если  $m(A) > 0$ ,

$F = (A, m)$  — тело свидетельств,

$F(T)$  — множество всех тел свидетельств на  $T$ .

Тело свидетельств вида  $F_A = (\{A\}, 1)$  называют категоричным. В частности, категоричное тело свидетельств  $F_T = (\{T\}, 1)$  называют бессодержательным, поскольку оно не несет никакой информации о принадлежности истинной альтернативы подмножеству из  $T$ . Произвольное тело свидетельств  $F = (A, m)$  можно представить в виде:  $F = \sum_{A \in A} m(A)F_A$ . Тело свидетельств вида  $F_A^\alpha = \alpha F_A + (1-\alpha)F_T$ ,  $\alpha \in (0,1)$  называют простым. Тело свидетельств называют недогматичным, если  $m(T) > 0$ .

Ниже будем рассматривать тела свидетельств на упорядоченном множестве  $T = \{t_1, \dots, t_s\}$ ,  $t_1 < \dots < t_s$ . Причем будем считать, что фокальными элементами могут быть только упорядоченные последовательные наборы элементов из  $T = \{t_1, \dots, t_s\}$ , т. е. множества вида  $A = \{t_i, t_{i+1}, \dots, t_{i+r}\} \subseteq T$ . Множество всех таких подмножеств из  $T$  будем обозначать через  $\binom{2^T}{od} \subseteq 2^T$ ,  $\left| \binom{2^T}{od} \right| = \frac{1}{2} |T|(|T|+1)$ . Такие конструкции рассматривались в [58].

Телу свидетельств  $F = (A, m)$  можно поставить в соответствие так называемую пигнистическую вероятность (pignistic probability)  $Bet_F$  [59], значения которой на элементах  $t_i \in X$ ,  $i = 1, \dots, s$  равны вероятности этих событий при условии, что в пределах фокальных элементов случайные величины распределены равномерно:

$$Bet_F(t_i) = \sum_{A \in \binom{2^T}{od}, t_i \in A} \frac{m(A)}{|A|}, \quad i = 1, \dots, s. \quad (1.7)$$

Пигнистическое преобразование используют для принятия решения о выборе наилучшей альтернативы из множества тел свидетельств.

Количество незнания в информации, содержащейся в теле свидетельств  $F = (A, m)$  оценивают с помощью мер неопределенности [60]. В настоящее время существует несколько десятков таких мер (см. обзор [61]). В нашем исследовании использовалась энтропия Денга (Deng entropy) [62]

$$DE(F) = - \sum_{A \in \binom{2^T}{od}} m(A) \log_2 \left( \frac{m(A)}{2^{|A|}-1} \right), \quad (1.8)$$

которая приближенно равна сумме энтропии Шеннона для функции масс  $-\sum_{A \in \binom{2^T}{od}} m(A) \log_2 m(A)$  и меры неточности  $\sum_{A \in \binom{2^T}{od}} m(A)|A|$ . Ниже введем и будем использовать одну модификацию энтропии Денга, которая учитывает специфику тел свидетельств на упорядоченном множестве.

#### 1.4.3 Постановка задачи и общая схема доверительного ранжирования

Пусть  $X$  — множество альтернатив, которые представлены  $n$ -мерными векторами  $\mathbf{x} = (x_1, \dots, x_n)$ , где  $x_i$  — оценки порядковой шкалы  $T = \{t_1, \dots, t_s\}$ ,  $t_1 < \dots < t_s$ , которые соответствуют термам:  $t_1$  — «очень низкий»,  $t_2$  — «низкий» и т. д. Кроме самих оценок известны

элементы  $h_i$ , заданные в порядковой или числовой шкале и характеризующие степень надежности (уверенности ЛПР в правильности своего решения) относительно оценки  $x_i \in T$ . Необходимо агрегировать информацию от экспертов и ранжировать альтернативы. В данном исследовании предлагается агрегировать информацию от экспертов используя методы теории функций доверия.

Общая схема агрегирования информации от экспертов и ранжирования альтернатив будет следующей:

- 1) построение тел свидетельств отдельных экспертных оценок и с учетом информации об их надежности;
- 2) агрегирование сформированных тел свидетельств по каждой альтернативе используя методы комбинирования свидетельств;
- 3) ранжирование агрегированных оценок.

#### 1.4.3.1 Построение тел свидетельств отдельных экспертных оценок

Каждой оценке  $x_i \in T = \{t_1, \dots, t_s\}$  и ее степени надежности (уверенности ЛПР в правильности своего решения)  $h_i$  поставим в соответствие простое тело свидетельств

$$F_{x_i} = \mu(h_i)F_{A(x_i, h_i)} + (1 - \mu(h_i))F_T, \quad (1.9)$$

где  $A(x_i, h_i)$  — «размытие» точечного значения  $x_i \in T = \{t_1, \dots, t_s\}$  в зависимости от надежности оценки  $h_i$ ,  $\mu(h_i)$  — функция дисконтирования массы в зависимости от надежности оценки  $h_i$ . Ниже в примере будем считать, что  $\mu(h) = h$ . Процедуры размытие точечных и интервальных данных на числовой оси для построения фокальных элементов тел свидетельств рассматривались в работах [64, 63, 65].

Информация в виде коэффициента  $h \in [0, 1]$  о степени уверенности ЛПР в правильности своего решения может быть использована для размытия точечной оценки  $x \in T$ . Будем считать, что  $h = 0$  соответствует абсолютной неуверенности,  $h = 1$  — абсолютной уверенности. Чем меньше степень уверенности, тем больше должна быть степень размытия.

В общем случае функция  $A(x, h) : T \times [0, 1] \rightarrow \binom{2^T}{od}$  должна удовлетворять условиям:

- 1)  $A(x, 1) = \{x\} \quad \forall x \in T;$
- 2)  $A(x, 0) = T \quad \forall x \in T;$
- 3)  $A(x, h_2) \subseteq A(x, h_1) \quad \forall x \in T, \text{ если } h_1 \leq h_2.$

Из 1) и 3), в частности, следует, что  $x \in A(x, h) \quad \forall x \in T$  и  $\forall h \in [0, 1]$ . Кроме того, если в ранговой шкале  $T$  есть нейтральная оценка  $t_k \in T$ , то потребуем, чтобы при размытии элемента  $x \in T$  множество  $A(x, h)$  имело большее смещение к нейтральной оценке, чем к «крайним» элементам. Точнее, чтобы выполнялось условие:

$$4) \quad A(t_j, h) = \{t_{j-p}, \dots, t_j, \dots, t_{j+r}\}, \text{ где } p \leq r, \text{ если } t_j < t_k \text{ и } p \geq r, \text{ если } t_j > t_k.$$

Это можно интерпретировать следующим образом: смещение оценки с низкой степенью уверенности к нейтральной оценке делает ее «в среднем» менее конфликтной с другими оценками, что соответствует парадигме «с незнанием нет конфликта». Конкретный вид функции  $A(x, h)$  зависит от приложения и определяется экспертыным образом методами машинного обучения.

#### 1.4.3.2 Энтропия тела свидетельств на упорядоченном множестве

На упорядоченном множестве энтропия тела свидетельств должна зависеть не только от масс и мощностей фокальных множеств, но и от «разброса» элементов фокальных множеств. Такая энтропия рассматривалась, например, в [66]. В ней учитывался «разброс» элементов равномощных фокальных множеств. В случае упорядоченных последовательных наборов элементов (а в нашей работе рассматриваются только такие множества) все равномощные фокальные элементы будут иметь один и тот же «разброс». С другой стороны, если в ранговой шкале  $T$  есть нейтральная оценка  $t_k \in T$ , то два равномощных фокальных множества  $A$  и  $B$  с одинаковыми массами будут вносить разный вклад в энтропию в зависимости от расположения их элементов относительно нейтральной оценки. Если в  $A$  элементы симметрично расположены относительно нейтральной оценки, а в  $B$  — нет, то вклад  $A$  в итоговую энтропию должен быть больше, чем вклад  $B$ . Для оценки этих вкладов введем следующий индекс  $I(A) \in [I_{\min}, I_{\max}]$ ,  $A \in \binom{2^T}{od}$ , который характеризует несимметричность расположения элементов относительно нейтральной оценки.

Пусть  $A_- = \{t \in A : t < t_k\}$ ,  $A_+ = \{t \in A : t > t_k\}$ . Индекс  $I$  должен удовлетворять условиям:

$$1) \quad I(A) = I_{\min}, \text{ если } A \in \binom{2^T}{od} : A_- = \emptyset \vee A_+ = \emptyset;$$

$$2) \quad I(A') \leq I(A''), \text{ если } A', A'' \in \binom{2^T}{od} : |A'| = |A''| \text{ и } \|A'_+ - A'_-\| \geq \|A''_+ - A''_-\|.$$

Например, этим условиям удовлетворяет индекс  $I(A) = \frac{2 \min\{|A_-|, |A_+|\}}{\|A_- - A_+\| + 1} \in [0, |A| - 1]$ ,

который и будем использовать ниже в модификации энтропии Денга

$$DEM(F) = - \sum_{A \in A} m(A) \log_2 \left( \frac{m(A)}{2^{|A|+I(A)} - 1} \right). \quad (1.10)$$

В этом случае к энтропии Денга приближенно добавляется мера средней несимметричности  $\sum_{A \in A} m(A)I(A)$ . Нетрудно видеть, что:

- a)  $I(A) = I_{\min} = 0 \Leftrightarrow A_- = \emptyset \vee A_+ = \emptyset$  (в частности,  $I(A) = 0$ , если  $A$  — синглетон);
- б) если  $A \in (2^T)_{od} \setminus \{\emptyset\}$  и  $|A_-| = |A_+|$ , то  $I(A) = I_{\max} = |A| - 1$ ;
- в)  $DEM(F) \geq DE(F)$ .

#### 1.4.3.3 Агрегирование тел свидетельств экспертических оценок

Далее для вектора оценок  $\mathbf{x} = (x_1, \dots, x_n)$  все тела свидетельств  $F_{x_i}$  агрегируются в соответствии с некоторым правилом  $\otimes_R : \underbrace{F(T) \times \dots \times F(T)}_n \rightarrow F(T)$ . В результате получим тело свидетельств

$$F_{\mathbf{x}} = \bigotimes_{i=1}^n F_{x_i}. \quad (1.11)$$

В качестве правил агрегирования будем рассматривать конъюнктивные правила (Демпстера, Ягера и др.), в которых новые фокальные элементы строятся на пересечении фокальных элементов агрегируемых тел свидетельств. Это важно, поскольку  $A \cap B \in (2^T)_{od}$  для любых  $A, B \in (2^T)_{od}$ . Ненормализованное конъюнктивное правило агрегирования тел свидетельств  $F_i = (A_i, m_i)$ ,  $i = 1, \dots, n$  на  $T$  имеет вид  $F_{\cap} = \bigotimes_{i=1}^n F_i = (A \cup \{\emptyset\}, m_{\cap})$ , где

$$m_{\cap}(A) = \sum_{B_1 \cap \dots \cap B_n = A} m_1(B_1) \dots m_n(B_n). \quad (1.12)$$

Но в этом случае может оказаться, что  $m_{\cap}(\emptyset) = \sum_{B_1 \cap \dots \cap B_n = \emptyset} m_1(B_1) \dots m_n(B_n) > 0$ . Величина  $Con(F_1, \dots, F_n) = m_{\cap}(\emptyset) \in [0, 1]$  является мерой несогласованности источников информации, представленных телами свидетельств или мерой конфликта. Чтобы получить тело

свидетельств  $F = (A, m)$  необходимо перераспределить массу  $m_{\cap}(\emptyset)$  по другим фокальным элементам. В классическом правиле Демпстера [54]  $\otimes_D$  осуществляется равномерное перераспределение массы  $m_{\cap}(\emptyset)$  по всем фокальным элементам:  $m_D(A) = \frac{1}{1 - m_{\cap}(\emptyset)} m_{\cap}(A)$ , если  $A \neq \emptyset$  и  $m_D(\emptyset) = 0$ . Если  $Con(F_1, \dots, F_n) = 1$  (абсолютный конфликт), то правило Демпстера неприменимо. Заметим, что если рассматриваются только недогматичные тела свидетельств, то  $Con(F_1, \dots, F_n) < 1$ .

В правиле Ягера [67]  $\otimes_Y$  вся масса  $m_{\cap}(\emptyset)$  добавляется к  $m_{\cap}(T)$ , увеличивая вес «незнания»:  $m_Y(A) = m_{\cap}(A)$ , если  $A \neq \emptyset, T$ ,  $m_Y(\emptyset) = 0$  и  $m_Y(T) = m_{\cap}(T) + m_{\cap}(\emptyset)$ .

Рассмотрим ещё одну процедуру перераспределения массы  $m_{\cap}(\emptyset)$ . Будем перераспределять её по фокальным элементам пропорционально их «величине». Такая схема является модификацией правила Инагаки [68]. Поэтому будем называть ее модифицированным правилом Инагаки  $\otimes_{MI}$ . Точнее, пусть  $m_{MI}(A) = m_{\cap}(A)(1 + \alpha_A \varphi(A))$ , если  $A \neq \emptyset$ , где  $\varphi: 2^T \rightarrow [0, +\infty)$  — монотонная функция (т. е.  $\varphi(A) \leq \varphi(B)$ , если  $A \subseteq B$ ),  $\alpha_A, A \in A$  — неотрицательные коэффициенты, которые должны удовлетворять условию нормировки:  $1 = \sum_{A \in A} m_{MI}(A) = \sum_{A \in A} m_{\cap}(A) + \sum_{A \in A} m_{\cap}(A)\alpha_A \varphi(A) = 1 - m_{\cap}(\emptyset) + \sum_{A \in A} m_{\cap}(A)\alpha_A \varphi(A)$ . Откуда следует, что  $\alpha_A, A \in A$  должны удовлетворять равенству  $m_{\cap}(\emptyset) = \sum_{A \in A} m_{\cap}(A)\alpha_A \varphi(A)$ .

Рассмотрим частный случай, когда  $\alpha_A = \alpha = \text{const}$ . Тогда  $\alpha = \frac{m_{\cap}(\emptyset)}{H(F_{\cap})}$ , где  $H(F_{\cap}) = \sum_{A \in A} m_{\cap}(A)\varphi(A)$  — мера неопределенности тела свидетельств  $F_{\cap}$ . В этом случае правило перераспределения масс примет вид:  $m_{MI}(A) = m_{\cap}(A) + \frac{m_{\cap}(A)\varphi(A)}{H(F_{\cap})} m_{\cap}(\emptyset)$ ,  $A \in A$ .

Величина  $\frac{m_{\cap}(A)\varphi(A)}{H(F_{\cap})}$  — это часть меры неопределенности, приходящейся на фокальный элемент  $A \in A$ . Поэтому смысл этого правила заключается в следующем: величина конфликта  $m_{\cap}(\emptyset)$  перераспределяется по фокальным элементам пропорционально приходящейся на них неопределенности. Ниже в примере в качестве функции  $\varphi$  будем рассматривать мощность множества  $\varphi(A) = |A|$ .

Заметим, что во всех рассмотренных конъюнктивных правилах агрегирования в результате свидетельств  $F = (A, m)$  множество фокальных элементов  $A$  не меняется. А если агрегируемые свидетельства имеют нулевой конфликт, то все правила дадут один и тот же результат.

#### 1.4.3.4 Ранжирование агрегированных оценок

Пусть  $\{F_x\}_{x \in X}$  — множество тел свидетельств, соответствующих векторным оценкам  $\mathbf{x} = (x_1, \dots, x_n) \in X$ ;  $Bet_{F_x}$  — пигнистическая вероятность, соответствующая телу свидетельств  $F_x$ . Для ранжирование векторных оценок  $\{\mathbf{x}\}$  на основе найденных тел свидетельств будем использовать понятие интервальной медианы  $[\underline{me}_x, \overline{me}_x] \subseteq T$ ,  $x \in X$ , границы которой  $\underline{me}_x, \overline{me}_x$  определяются следующим образом.

Рассмотрим два множества  $\underline{I}_x = \left\{ t_k : \sum_{i=1}^k Bet_{F_x}(t_i) \leq \frac{1}{2}, Bet_{F_x}(t_k) \neq 0 \right\}$  и  $\overline{I}_x = \left\{ t_k : \sum_{i=k}^s Bet_{F_x}(t_i) \leq \frac{1}{2}, Bet_{F_x}(t_k) \neq 0 \right\}$  (хотя бы одно из них непустое). Пусть  $\Delta_x = X \setminus (\underline{I}_x \cup \overline{I}_x)$ .

Возможны три случая:

- 1) если  $|\Delta_x| = 0$  (т. е.  $\underline{I}_x \cup \overline{I}_x = X$ ), то положим  $\underline{me}_x = \sup \underline{I}_x, \overline{me}_x = \inf \overline{I}_x$ ;
- 2) если  $|\Delta_x| = 1$ , то  $\underline{me}_x = \overline{me}_x = t \in \Delta_x$ ;
- 3) если  $|\Delta_x| > 1$ , то нетрудно показать, что в этом случае либо  $Bet_{F_x}(t) = 0 \quad \forall t \in \Delta_x$  и тогда положим  $\underline{me}_x = \sup \underline{I}_x, \overline{me}_x = \inf \overline{I}_x$ , либо существует единственный элемент  $t_0 \in \Delta_x$ :  $Bet_{F_x}(t_0) > 0$  и тогда пусть  $\underline{me}_x = \overline{me}_x = t_0 \in \Delta_x$ .

Например, для тела свидетельств  $F = \frac{1}{4} F_{\{t_1, t_2\}} + \frac{1}{3} F_{\{t_1, t_2, t_3\}} + \frac{1}{3} F_{\{t_2, t_3, t_4\}} + \frac{1}{12} F_{\{t_1, t_2, t_3, t_4\}}$ , заданном на упорядоченном множестве  $T = \{t_1, t_2, t_3, t_4\}$  пигнистическое вероятностное распределение равно:  $Bet_F(t_1) = \frac{37}{144}, Bet_F(t_2) = \frac{53}{144}, Bet_F(t_3) = \frac{35}{144}, Bet_F(t_4) = \frac{19}{144}$ . Тогда (случай 2)  $\underline{me}_x = \overline{me}_x = t_2$ .

Далее рассмотрим множество векторов «точечных» медиан  $M_x = \left\{ (me_x)_{x \in X} : \underline{me}_x \leq me_x \leq \overline{me}_x, x \in X \right\}$ . Для каждого фиксированного набора медиан

$\mathbf{me} = (me_x)_{x \in X} \in M_X$  ранжируем альтернативы  $X$  в соответствии с ранжированием этих медиан:  $\mathbf{a} \succ_{\mathbf{me}} \mathbf{b}$ , если  $me_a > me_b$  и  $\mathbf{a} \sqsupseteq_{\mathbf{me}} \mathbf{b}$ , если  $me_a = me_b$ . В результате получим множество частных ранжирований (предпочтений) на  $X$ . Итоговое ранжирование альтернатив осуществляется путем применения некоторого правила агрегирования предпочтений. Широкий выбор таких правил предоставляет теория коллективного выбора [53]. Например, может быть использовано правило Борда. В этом случае вычисляем турнирную матрицу

$$S = \{s(\mathbf{a}, \mathbf{b})\}_{\mathbf{a}, \mathbf{b} \in X}, \quad s(\mathbf{a}, \mathbf{b}) = \sum_{m \in M_X} Ind_m(\mathbf{a}, \mathbf{b}), \quad Ind_m(\mathbf{a}, \mathbf{b}) = \begin{cases} 1, & \text{если } \mathbf{a} \succ_m \mathbf{b}, \\ 0.5, & \text{если } \mathbf{a} \sqsupseteq_m \mathbf{b}, \\ 0, & \text{если } \mathbf{b} \succ_m \mathbf{a}. \end{cases}$$

суммарный ранг каждой альтернативы  $r_a = \sum_{b \in X} s(\mathbf{a}, \mathbf{b})$ . Агрегированное ранжирование получается в результате упорядочивания по убыванию суммарного ранга  $r_a$ ,  $a \in X$ .

#### 1.4.4 Численный пример

Проиллюстрируем предложенную схему агрегирования экспертных оценок и последующего ранжирования на данных о рецензировании статей/докладов конференций в системе управления конференциями *EasyChair* (<https://easychair.org>). В этой системе используется семиранговая порядковая шкала оценивания  $T = \{t_1, \dots, t_7\}$ . Ранги  $t_i$ ,  $i = 1, \dots, 7$  соответствуют рекомендациям (в порядке возрастания) “strong reject”, “reject”, “weak reject”, “borderline paper”, “weak accept”, “accept”, “strong accept”. Кроме того, рецензент дает оценку в пятиранговой шкале ( $h_1$  – “none”,  $h_2$  – “low”,  $h_3$  – “medium”,  $h_4$  – “high”,  $h_5$  – “expert”) о степени уверенности (confidence) в правильности своего решения:  $h_1 < \dots < h_5$ .

Будем считать, что степени уверенности заданы в числовой шкале формулой  $h_j = 0.2j$ ,  $j = 1, \dots, 5$ .

Точечные данные  $\left\{ \left( x_r^{(k)}, h_r^{(k)} \right) \right\}_{r=1}^3$   $n = 3$  рецензентов относительно 4-х статей  $\{a_1, \dots, a_4\}$  представлены в Таблице 1.4 ( $r$  – индекс рецензента,  $k$  – индекс статьи,  $k = 1, \dots, 4$ ), где  $x_r^{(k)} \in T$ .

Таблица 1.4 – Исходные данные об оценках рецензентов

	статья $a_1$	статья $a_2$	статья $a_3$	статья $a_4$
рецензент 1	$(t_6, 0.8)$	$(t_6, 1)$	$(t_5, 0.8)$	$(t_4, 0.6)$
рецензент 2	$(t_5, 1)$	$(t_5, 0.4)$	$(t_6, 0.6)$	$(t_5, 0.4)$
рецензент 3	$(t_4, 0.8)$	$(t_5, 0.6)$	$(t_3, 0.6)$	$(t_5, 1)$

Найдем простые тела свидетельств вида (1.1) для каждой оценки всех рецензентов. Будем использовать следующие функции размытия  $A(x, h)$ , которые удовлетворяют всем

рассмотренным выше условиям:  $A(t_i, 1) = \{t_i\}$ ,  $i = 1, \dots, 7$ ;  $A(t_i, 0.8) = \begin{cases} \{t_i, t_{i+1}\}, i = 1, 2, 3, \\ \{t_i\}, i = 4, \\ \{t_{i-1}, t_i\}, i = 5, 6, 7; \end{cases}$

$$A(t_i, 0.6) = \begin{cases} \{t_i, t_{i+1}, t_{i+2}\}, i = 1, \\ \{t_{i-1}, t_i, t_{i+1}\}, i = 2, \dots, 6, \\ \{t_{i-2}, t_{i-1}, t_i\}, i = 7; \end{cases} \quad A(t_i, 0.4) = \begin{cases} \{t_1, t_2, t_3, t_4\}, i = 1, 2, \\ \{t_2, t_3, t_4, t_5\}, i = 3, \\ \{t_3, t_4, t_5\}, i = 4, \\ \{t_3, t_4, t_5, t_6\}, i = 5, \\ \{t_4, t_5, t_6, t_7\}, i = 6, 7; \end{cases}$$

$$A(t_i, 0.2) = T, \quad i = 1, \dots, 7.$$

В результате получим тела свидетельств  $F_{rk} = h_r^{(k)} F_{A(x_r^{(k)}, h_r^{(k)})} + (1 - h_r^{(k)}) F_T$ ,  $r = 1, 2, 3$ ,

$k = 1, \dots, 4$ . Чтобы все тела свидетельств были недогматичными (и, как следствие, правило Демпстера было заведомо применимо), заменим значения  $h_r^{(k)} = 1$  на близкие значения  $h_r^{(k)} = 0.95$ . В Таблице 1.5 тела свидетельств представлены в виде пар  $(A(x_r^{(k)}, h_r^{(k)}), h_r^{(k)})$ ,  $r = 1, 2, 3$ ,  $k = 1, \dots, 4$ . В последней строке приведены значения меры конфликта оценок всех рецензентов по каждой статье  $Con(F_{1k}, F_{2k}, F_{3k})$ ,  $k = 1, \dots, 4$ .

Таблица 1.5 – Тела свидетельств оценок рецензентов

	статья $a_1$	статья $a_2$	статья $a_3$	статья $a_4$
рецензент 1	$(\{t_5, t_6\}, 0.8)$	$(\{t_6\}, 0.95)$	$(\{t_4, t_5\}, 0.8)$	$(\{t_3, t_4, t_5\}, 0.6)$
рецензент 2	$(\{t_5\}, 0.95)$	$(\{t_3, t_4, t_5, t_6\}, 0.4)$	$(\{t_5, t_6, t_7\}, 0.6)$	$(\{t_3, t_4, t_5, t_6\}, 0.4)$
рецензент 3	$(\{t_4\}, 0.8)$	$(\{t_4, t_5, t_6\}, 0.6)$	$(\{t_2, t_3, t_4\}, 0.6)$	$(\{t_5\}, 0.95)$
$Con$	0.79	0	0.36	0

Применим конъюнктивные правила для агрегирования тел свидетельств всех рецензентов по каждой статье. Так как для статей  $a_2$  и  $a_4$  конфликт равен нулю, то агрегирование по этим статьям не будет зависеть от вида конъюнктивного правила. Например, для правила

Демпстера  $F_k = \bigotimes_{r=1}^3 F_{rk}$  получим следующие результаты:

$$F_1 = 0.038F_{\{t_4\}} + 0.914F_{\{t_5\}} + 0.038F_{\{t_5, t_6\}} + 0.01F_T,$$

$$F_2 = 0.95F_{\{t_6\}} + 0.03F_{\{t_4, t_5, t_6\}} + 0.008F_{\{t_3, t_4, t_5, t_6\}} + 0.012F_T,$$

$$F_3 = 0.3F_{\{t_4\}} + 0.3F_{\{t_5\}} + 0.2F_{\{t_4, t_5\}} + 0.075F_{\{t_2, t_3, t_4\}} + 0.075F_{\{t_5, t_6, t_7\}} + 0.05F_T,$$

$$F_4 = 0.95F_{\{t_5\}} + 0.03F_{\{t_3, t_4, t_5\}} + 0.008F_{\{t_3, t_4, t_5, t_6\}} + 0.012F_T.$$

Пигнистические вероятности  $Bet_{F_k}(t_i)$ ,  $k=1,\dots,4$ ,  $i=1,\dots,7$  для свидетельств агрегированных по правилу Демпстера приведены в Таблице 1.6 (не показаны вероятности меньшие, чем 0.002).

Таблица 1.6 – Правило Демпстера: пигнистические вероятности

	$t_1$	$t_2$	$t_3$	$t_4$	$t_5$	$t_6$	$t_7$	$DEM$
$F_1$	0	0	0.004	0.04	0.934	0.022	0	0.73
$F_2$	0	0	0.007	0.014	0.014	0.965	0	0.63
$F_3$	0.008	0.032	0.032	0.432	0.432	0.032	0.032	3.67
$F_4$	0	0	0.016	0.014	0.965	0.005	0	0.7

Кроме того, в последнем столбце Таблице 1.6 приведены значения модифицированной энтропии Денга, характеризующей степень неопределенности тел свидетельств, полученных в результате агрегирования оценок рецензентов.

Все медианы этих распределений будут точечными и равны соответственно:  $me_1 = \{t_5\}$ ,  $me_2 = \{t_6\}$ ,  $me_3 = \{t_4\}$ ,  $me_4 = \{t_5\}$ . Поэтому ранжирование статей будет следующим:  $a_2 \succ_{me} a_1 \sqcap_{me} a_4 \succ_{me} a_3$ .

При использовании правила Ягера или модифицированного правила Иногаки получим другие тела свидетельств и пигнистические вероятности по статьям  $a_1$  и  $a_3$ , но их медианы и соответственно итоговое ранжирование статей не изменится.

Сравним этот результат с линейной сверткой критериев. В этом случае будем считать, что оценки рецензентов даны в числовой шкале  $t_i = i$ ,  $i=1,\dots,7$ . Тогда линейная свертка критериев (рецензентов) по каждой статье с весами  $h_r^{(k)}$ ,  $r=1,2,3$  примет вид  $C(a_k) = \sum_{r=1}^3 h_r^{(k)} x_r^{(k)}$ . Для наших данных мы получим:  $C(a_1) = 12.75$ ,  $C(a_2) = 10.7$ ,  $C(a_3) = 9.4$ ,  $C(a_4) = 9.15$ . Таким образом, ранжирование статей в этом случае будет следующим:  $a_1 \succ a_2 \succ a_3 \succ a_4$ . Отличие в ранжированиях можно объяснить тем, что рецензии по статьям  $a_1$  и  $a_3$  сильно конфликтны, что учитывается при применении конъюнктивных правил. Кроме того, при линейной свертке критериев оценки с малым весом значительно снижают значение линейной свертки. В то время как при доверительном агрегировании малый вес только «размывает» оценку рецензента.

#### 1.4.5 Заключение

В этой части отчета приведены результаты по разработке доверительной процедуры агрегирования информации от экспертов и последующего ранжирования этих агрегированных оценок в том случае, когда оценки представлены в порядковой шкале с указанием их достоверности. Процедура основана на размытии точечных оценок в соответствии со степенями их достоверности. В результате оценки экспертов преобразуются в тела свидетельств. Рассмотрено агрегирование тел свидетельств экспертных оценок на основе конъюнктивных правил комбинирования теории свидетельств. Ранжирование осуществляется на основе вычисления (возможно интервальных) медиан соответствующих пигнистическим вероятностям тел свидетельств. Метод проиллюстрирован на примере ранжирования оценок рецензентов в системе управления конференциями типа *EasyChair*. Метод продемонстрировал определенную устойчивость результата к выбору вида конъюнктивного правила агрегирования. В отличие от метода линейной свертки критериев (если считать, что данные представлены в числовой шкале) предложенный метод не занижает оценки экспертов с низкой уверенностью, а только сильнее размывает их. Кроме того, использование правил агрегирования теории свидетельств позволяет учитывать конфликтность экспертных оценок. Метод адаптируется под конкретные приложения и требования за счет выбора функций размытия, правил агрегирования тел свидетельств, правил агрегирования ранжирований.

Результаты исследований нашли отражение в [69].

## **2 Разработка и исследование оптимизационных моделей**

В этом разделе отчета приведены результаты разработки и исследования некоторых оптимизационных моделей. Работа осуществлялась в следующих направлениях:

- аппроксимация функций и регрессионный анализ: многокритериальный подход;
- разработка инструментария решения задачи о выборе оптимальной структуры курсов для смешанного обучения;
- оптимизация гарантированного риска с учетом дополнительных связей между неопределенными факторами.

### **2.1 Аппроксимация функций и регрессионный анализ: многокритериальный подход**

#### **2.1.1 Введение**

Аппроксимация (приближение), т. е. замена одних математических объектов другими, в том или ином смысле близкими к исходным, широко применяется в математике и ее приложениях. В частности, она является основой для аналитического представления табличных данных функцией заданного вида с неизвестными параметрами, значения которых определяются так, чтобы график функций был по возможности максимально близок к заданным точкам согласно выбранной числовой мере близости. Задачи такого оценивания являются базовыми в теории аппроксимации функций и регрессионном анализе [70, 71].

Меры близости в регрессионном анализе достаточно обоснованно (в частности, с привлечением принципа максимума правдоподобия) выбираются в тех случаях, когда распределение вероятностей ошибок измерения подчинены некоторому известному закону, чаще всего нормальному закону, или закону Гаусса (при этом мерой служит сумма квадратов уклонений рассчитанных значений аппроксимирующей функции от табличных), реже – закону Лапласа (и тогда мерой служит сумма модулей таких уклонений).

В общем же случае при решении задач аппроксимации проблема обоснованного выбора меры близости остается нерешенной, на практике ей обычно уделяется недостаточно внимания (нередко о ней вообще не вспоминают) и в качестве меры берут сумму квадратов уклонений, так как это – «общепринятый подход», для него имеются давно и хорошо разработанный математический аппарат аппроксимации и созданы компьютерные программы. А ведь выбор меры близости самым существенным образом влияет на результат аппроксимации.

В главе 1 отчета за 2021 год [72] был предложен новый подход: проблема выбора меры близости решается за счет того, что близость оценивается при помощи векторного

критерия, состоящего из модулей уклонений значений аппроксимирующей функции от заданных для всех табличных значений аргумента. Была сформулирована общая многокритериальная постановка задачи аппроксимации (регрессионного анализа) и намечен вычислительный путь построения аппроксимирующих функций. В настоящем отчете изучены свойства введенных аппроксимирующих функций и их связь с классическими (получаемыми методами наименьших квадратов, наименьших модулей и наименьшего максимального модуля уклонений). Предложены вычислительные методы построения множеств аппроксимирующих функций.

### 2.1.2 Многокритериальная задача аппроксимации

Исходными данными в задаче аппроксимации являются пары значений независимой переменной  $x$  и зависимой переменной  $y$ :

$$(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n). \quad (2.1)$$

В регрессионном анализе независимая переменная  $x$  именуется фактором, а зависимая переменная  $y$  – откликом (наблюденным значением).

Далее будем полагать, что аппроксимирующая функция (функция регрессии) является линейной:

$$y = ax + b, \quad (2.2)$$

где  $a$  и  $b$  – коэффициенты аппроксимации, которые нужно оценить с использованием исходных данных (2.1). Для такого оценивания вводятся в рассмотрение модули невязок  $\delta_i$ , или уклонений теоретических значений зависимой переменной – рассчитанных по формуле (2.3) – от заданных:

$$\delta_1 = ax_1 + b - y_1, \delta_2 = ax_2 + b - y_2, \dots, \delta_n = ax_n + b - y_n. \quad (2.3)$$

Общепринятым при аппроксимации является путь, основанный на введении числовой меры близости аппроксимирующей функции к исходным данным, такой, как сумма квадратов невязок, сумма модулей невязок или максимальный модуль невязок. Наиболее распространенным методом оценивания является метод наименьших квадратов – МНК (the least squares method – LS), согласно которому значения коэффициентов регрессии определяются из условия:

$$\sum_{i=1}^n \delta_i^2 \rightarrow \min_{a, b}. \quad (2.4)$$

Для расчета значений коэффициентов согласно (2.4) имеются простые формулы [70, 71].

Достаточно распространенным, особенно в последние годы, стал метод наименьших модулей – МНМ (the least modulus method – LM), согласно которому значения коэффициентов аппроксимации определяются из условия:

$$\sum_{i=1}^n |\delta_i| \rightarrow \min_{a, b}. \quad (2.5)$$

Разработаны специальные алгоритмы нахождение значений коэффициентов согласно (2.5) [73].

Метод, основанный на использовании наибольшего модуля невязок, называемый далее методом наименьшего максимального модуля – МНММ (the least maximum modulus method – LMM), по сути, использует меру Чебышёва и предполагает получение значений коэффициентов регрессии из условия:

$$\max_{i \in N} |\delta_i| \rightarrow \min_{a, b}, \quad (2.6)$$

где  $N = \{1, 2, \dots, n\}$ . Отметим, что нелинейные оптимизационные задачи (2.5) и (2.6) за счет введения дополнительных переменных можно свести к эквивалентным задачам линейного программирования [74].

Далее рассматривается альтернативный метод оценивания, учитывающий, что по своей сути задача оценивания параметров регрессии изначально является многокритериальной: критериями  $f_i$ , подлежащими минимизации ( $i \in N$ ), служат модули невязок  $\delta_i$  из (2.3), составляющие векторный критерий  $f(a, b) = (f_1(a, b), f_2(a, b), \dots, f_n(a, b))$ , где  $f_i(a, b) = z_i = |\delta_i|$ . Значение векторного критерия для пары  $(a, b)$  называется ее векторной оценкой:  $z = f(a, b)$ .

Пусть на множестве векторных оценок  $Z = \mathbb{R}_+^n = [0, +\infty)^n$  задан строгий частичный порядок  $P^\Gamma$ , где  $\Gamma$  – информация, касающаяся оценивания удаленности: если верно  $z' P^\Gamma z''$ , то при модулях невязок  $z'_1, z'_2, \dots, z'_n$  аппроксимирующая функция считается ближе к заданным точкам (2.1), чем при модулях невязок  $z''_1, z''_2, \dots, z''_n$ . Отношение  $P^\Gamma$  индуцирует на плоскости  $\mathbb{R}^2$  строгий частичный порядок  $P_\Gamma$ :

$$(a', b')P_\Gamma(a'', b'') \Leftrightarrow f(a', b')P^\Gamma f(a'', b''), \quad (2.7)$$

имеющий аналогичный смысл: если верно  $(a', b')P_\Gamma(a'', b'')$ , то при значениях параметров  $a', b'$  аппроксимирующая функция считается ближе к заданным точкам (2.1), чем при значениях параметров  $a'', b''$ . В качестве решения задачи оценивания параметров аппроксимации рассматривается множество точек (векторов)  $(a, b) \in \mathbb{R}^2$ , которые недоминируются по отношению  $P_\Gamma$ . (Точка  $(a^*, b^*)$  называется недоминируемой по  $P_\Gamma$ , если не существует точки  $(a, b)$  такой, что верно  $(a, b)P_\Gamma(a^*, b^*)$ ).

При указанном многокритериальном подходе в качестве решения задачи аппроксимации выделяется не одно оптимальное значение коэффициентов  $(a^*, b^*)$ , а множество недоминируемых значений  $U^\Gamma \subset \mathbb{R}^2$ . Каждой паре  $(a, b) \in U^\Gamma$  соответствует «своя» функция  $y = f(x|a, b) = ax + b$  и ее график – прямая (2.2), а множеству  $U^\Gamma$  – множество функций  $F^\Gamma = \{f(x|a, b) | (a, b) \in U^\Gamma\}$  и множество их графиков – связка (совокупность) прямых. При этом фиксированному значению аргумента  $x$  ставится в соответствие множество чисел  $Y^\Gamma(x) = \{y \in \mathbb{R} | y = ax + b, (a, b) \in U^\Gamma\}$ . Иными словами, полученная указанным путем аппроксимирующая функция  $F^\Gamma$  оказывается многозначной. Такие аппроксимирующие функции будем называть недоминируемыми, или, более конкретно,  $\Gamma$ -недоминируемыми.

Многозначность аппроксимации (множественность выделяемых значений параметров) – плата за то, что при многокритериальной постановке задачи их оценивания не делается никаких дополнительных ограничительных допущений о характере (свойствах) исходных данных (2.1) типа фиксирования вида закона распределения вероятностей ошибок измерений и т. п.

Для снятия неопределенности из-за многозначности аппроксимации (за счет снижения информативности) можно множество  $Y^\Gamma(x)$  представлять его центром тяжести (центроидом)  $c^\Gamma(x)$ . В частности, если  $Y^\Gamma(x)$  – отрезок  $[\alpha(x), \beta(x)]$ , то  $c^\Gamma(x)$  – его середина:  $c^\Gamma(x) = \frac{1}{2}(\alpha(x) + \beta(x))$ . Если функция  $y = c^\Gamma(x)$  оказывается линейной (на интересующем интервале оси  $Ox$ ), то ее можно принять за однозначное упрощенное решение задачи линейной аппроксимации.

### 2.1.3 О построении множеств $U^\Gamma$ и $Y^\Gamma(x)$

Для приближенного построения (конечной аппроксимации) множества  $U^\Gamma$  можно построить «достаточно мелкую» сетку, покрывающую «подходящую» область в  $\mathbb{R}^2$  (она определяется с учетом специфики задачи, т. е. расположения точек из (2.1) на плоскости),

рассматривать эту сетку вместо множества  $\mathbb{R}^2$  и выделять те из ее узлов, которые являются недоминируемыми по  $P_\Gamma$ . Подобный подход широко используется в многокритериальной оптимизации, когда множество допустимых вариантов решения является областью в  $\mathbb{R}^n$  (см., например, [75]). Он применялся также при построении множества ПН-средних, определяемого в рамках многокритериального подхода как совокупности точек прямой, каждая из которых недоминируется по отношению предпочтения на множестве векторов, компонентами которых служат расстояния от неё до заданных точек [76, 77].

В качестве ориентира при первоначальном определении области поиска в  $\mathbb{R}^2$  можно использовать точки  $(a^\bullet, b^\bullet)$ , соответствующие прямым, проходящим через 2 или более точек множества (2.1). Тырсин и Азарян [78] называют такие точки «узловыми» и доказывают, что линию, определяемую по МНМ, следует искать среди  $y = a^\bullet x + b^\bullet$ . Также полезными могут оказаться точки  $(a^\blacksquare, b^\blacksquare)$ , соответствующие прямым, проходящим через 2 или более середин отрезков, соединяющих пары точек множества (2.1). В последующих примерах линии, определяемые по МНММ, оказываются среди  $y = a^\blacksquare x + b^\blacksquare$ .

Для сокращения области поиска множества  $U^\Gamma$  можно воспользоваться следующим замечанием. Зафиксируем значение параметра  $a = a^0$  и рассмотрим локальную задачу поиска точек  $(a^0, b) \in U^\Gamma(a^0)$ , недоминируемых по  $P_\Gamma$  на множестве  $\{(a, b) \mid a = a^0, b \in \mathbb{R}\}$ . Обозначим  $b_i = y_i - a^0 x_i$  для всех  $i \in N$  и преобразуем выражение для уклонений (2.3):

$$\delta_i = a^0 x_i + b - y_i = b - b_i. \quad (2.8)$$

В результате получаем задачу определения множества ПН-средних  $G^\Gamma(B(a^0))$  среди множества вещественных чисел  $B(a^0) = \{b_1, b_2, \dots, b_n\}$  (см. [76, 77]). Если точка  $b$  не входит в множество средних  $G^\Gamma(B(a^0))$ , то это означает, что точка  $(a^0, b)$  оказывается доминируемой по  $P_\Gamma$  на рассматриваемом множестве  $\{(a, b) \mid a = a^0, b \in \mathbb{R}\}$ , а значит и на всём множестве  $\mathbb{R}^2$ . Следовательно, должно выполняться включение

$$\{(a, b) \in U^\Gamma \mid a = a^0\} \subseteq \{(a, b) \mid a = a^0, b \in G^\Gamma(B(a^0))\}. \quad (2.9)$$

В ходе перебора точек методом построения сетки, это замечание можно использовать следующим образом. Для каждого перебираемого значения  $a$  сначала можно определить ПН-средние  $G^\Gamma(B(a))$ , а затем перебирать только значения  $b$  из этого множества. Это может быть удобно, так как для некоторых типов информации  $\Gamma$  существуют быстрые аналитические методы определения ПН-средних.

Для произвольного фиксированного значения  $x^0$  множество  $Y^\Gamma(x^0)$  графически представляется сечением вертикальной прямой  $x = x^0$  связки прямых  $F^\Gamma$ . Для построения множеств  $Y^\Gamma(x)$  можно использовать также графоаналитический метод, не требующий построения связки прямых и основанный на следующих соображениях. Произвольной точке  $(x^0, y^0)$  пространства  $\mathbb{R}^2$  данных  $(x, y)$  соответствует прямая  $b = -x^0a + y^0$  в пространстве  $\mathbb{R}^2$  параметров  $(a, b)$ . Если эта прямая содержит хотя бы одну точку  $(a^*, b^*) \in U^\Gamma$ , то точке  $(x^0, y^0)$  соответствует по крайней мере одна прямая из связки аппроксимирующих прямых  $F^\Gamma$ . Фиксируем  $x = x^0$  и будем менять  $y$ . Тогда соответствующая прямая  $b = -x^0a + y^0$  будет сдвигаться параллельно самой себе:  $b = -x^0a + y$ . То есть, вертикальной прямой  $x = x^0$  в пространстве  $\mathbb{R}^2$  данных  $(x, y)$  соответствует множество прямых  $b = -x^0a + y$  в пространстве  $\mathbb{R}^2$  параметров  $(a, b)$  с одинаковым угловым коэффициентом  $-x^0$ .

Следовательно, для построения  $Y^\Gamma(x^0)$  можно действовать так. Вначале следует построить множество  $U^\Gamma$  или его конечную аппроксимацию в системе координат параметров  $aOb$ . Далее нужно взять (начертить) прямую с наклоном  $-x^0$ . Параллельно сдвигая эту прямую (т. е. сохраняя угол наклона), нужно определить все те положения, в которых прямая  $b = -x^0a + y$  будет пересекать множество  $U^\Gamma$ . Соответствующие этим положениям значения  $b$  на оси ординат  $Ob$  составят множество  $Y^\Gamma(x^0)$ . В частности, если  $U^\Gamma$  – компактное (т. е. замкнутое и ограниченное) связное множество, то  $Y^\Gamma(x^0)$  – отрезок с концами  $y_{\min}$  и  $y_{\max}$ , соответствующими крайним (верхнему и нижнему) положениям передвигаемой прямой:  $y_{\min} = b_{\min}$  и  $y_{\max} = b_{\max}$ , ибо линейное отображение  $b = -x^0a + y^0$  является непрерывным и поэтому, как известно из курса анализа, переводит компактное связное множество в  $\mathbb{R}^2$  в компактное и связное множество в  $\mathbb{R}$ .

Примеры построения множеств  $U^\Gamma$  и  $Y^\Gamma(x)$  приводятся ниже. В них для конечной аппроксимации множеств  $U^\Gamma$  и  $Y^\Gamma(x)$  использовалась специально разработанная компьютерная программа.

#### 2.1.4 Аппроксимация при отсутствии информации $\Gamma$

Рассмотрим вначале случай, когда никакой информации  $\Gamma$  нет:  $\Gamma = \emptyset$ . Здесь в роли частичного порядка  $P^\Gamma$  выступает отношение Парето  $P^\emptyset$ , которое определяется (с учетом желательности минимизации критериев) так [79]:

$$z' P^\emptyset z'' \Leftrightarrow z'_i \leq z''_i, i = 1, 2, \dots, n, z' \neq z''. \quad (2.10)$$

Структура множества  $U^\varnothing$  исследовалась в [80]. Было установлено, что оно компактно (замкнуто и ограничено), связно, возможно, не выпукло и является объединением конечного числа политопов (выпуклых оболочек конечных наборов двумерных векторов). Из этого свойства, как указывалось выше, следует, что множества  $Y^\varnothing(x)$  – отрезки.

Изложенные теоретические положения иллюстрирует следующий числовой пример.

Пример 2.1. Пусть исходные данные (2.1) заданы в Таблице 2.1.

Таблица 2.1 – Исходные данные для задач примеров 2.1–2.3

$x_i$	1	2	3	4	5	7
$y_i$	3	2	5	5	4	7

Здесь уравнения аппроксимирующих функций, определяемые МНК, МНМ и МНММ, соответственно, таковы:

$$f^{LS}(x) = \frac{47}{70}x + 1\frac{61}{70} \approx 0.671x + 1.871; \quad (2.11)$$

$$f^{LM}(x) = \frac{2}{3}x + 2\frac{1}{3} \approx 0.667x + 2.333; \quad (2.12)$$

$$f^{LMM}(x) = \frac{2}{3}x + 1\frac{5}{6} \approx 0.667x + 1.833. \quad (2.13)$$

В этом примере при расчетах для аппроксимации множества  $U^\varnothing$  использовалась сетка с шагом 0.005 по параметру  $a$  и с шагом 0.05 по параметру  $b$ , покрывающая область  $[-1, 3] \times [-4, 9]$  для параметров  $(a, b)$ . Чтобы избежать вычислительных ошибок при сравнении чисел с плавающей точкой, при вычислениях использовались только рациональные числа, представленные целыми числителем и знаменателем. Например,  $a = \frac{\hat{a}}{\check{a}}$ , где числитель  $\hat{a}$  и знаменатель  $\check{a} > 0$  – целые числа. В таком представлении числитель и знаменатель уклонений (2.3) рассчитываются по следующим формулам:

$$\hat{\delta}_i = \hat{a}\check{x}_i\check{b}\check{y}_i + \hat{b}\check{a}\check{x}_i\check{y}_i - \hat{y}_i\check{a}\check{x}_i\check{b}, \quad \check{\delta}_i = \check{a}\check{b}\check{x}_i\check{y}_i. \quad (2.14)$$

Связка аппроксимирующих функций представлена в области  $[0, 8] \times [1, 8]$  для  $(x, y)$  на рис. 1 сверху в виде затемненной области, а множество точек  $U^\varnothing$  представлено в области  $[-1, 3] \times [-4, 9]$  для  $(a, b)$  на рисунке 2.1 снизу.

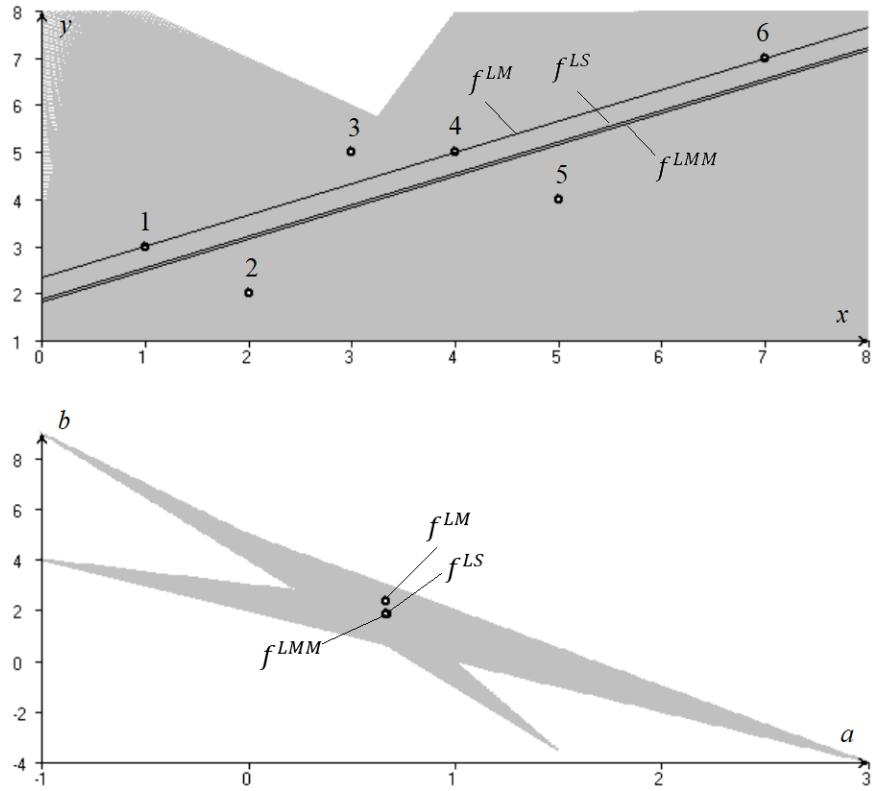


Рисунок 2.1 – Связка прямых и множество  $U^\varnothing$

Для информации  $\Gamma = \emptyset$  множество ПН-средних  $G^\varnothing(B(a))$  (множество значений функций, реализующих среднее по Коши) определяется легко: это отрезок  $[\min_{i \in N} b_i, \max_{i \in N} b_i]$ . Графики нижней и верхней границы этого отрезка в зависимости от параметра  $a$  представлены на рисунке 2.2.

Также на рисунке 2.2 изображены «узловые» точки  $(a^*, b^*)$ , соответствующие прямым, проходящим через 2 или более точек множества (2.1) (номера  $i \in N$  этих точек указаны рядом через дефис, например, точке «1-4-6» соответствует линия, определяемая по МНМ). Очевидно, что эти точки  $(a^*, b^*)$  недоминируемые по Парето, так как для каждой такой точки (и только для нее) все соответствующие уклонации  $\delta_i = 0$ .

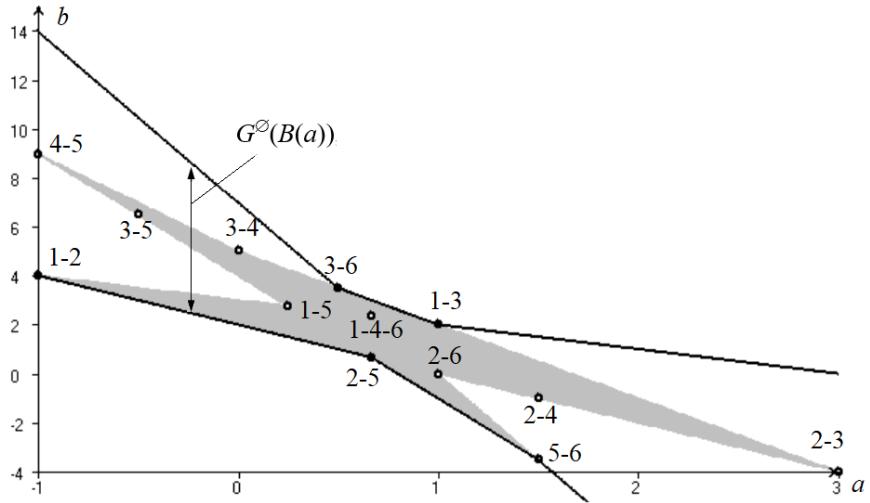


Рисунок 2.2 – Множество  $U\emptyset$ , графики границ отрезков  $G\emptyset(B(a))$ , а также узловые точки

### 2.1.5 Аппроксимация при равноважных критериях

Пусть теперь все критерии имеют равную важность (см. [81, 82]): информация  $\Gamma = E$ . Содержательно равноважность критериев означает, что после произвольной перестановки компонент векторной оценки  $z$  получится векторная оценка, эквивалентная (равная по предпочтительности) исходной. В рассматриваемой постановке задачи аппроксимации все точки (2.1) одинаково существенны («равноправны») и поэтому критерии  $f_i$  равноважны.

Обозначим через  $z_{\downarrow} = (z_{[1]}, z_{[2]}, \dots, z_{[n]})$  вектор, полученный упорядочением по невозрастанию компонент вектора  $z$ :  $z_{[1]} \geq z_{[2]} \geq \dots \geq z_{[n]}$ . Отношение  $P^E$  определяется следующим аналитическим решающим правилом [81]:

$$z' P^E z'' \Leftrightarrow z'_{[1]} \leq z''_{[1]}, z'_{[2]} \leq z''_{[2]}, \dots, z'_{[n]} \leq z''_{[n]}, z'_{\downarrow} \neq z''_{\downarrow}. \quad (2.15)$$

Поскольку  $P^E \supset P^\emptyset$ , то имеет место включение  $U^E \subseteq U^\emptyset$ . Оказывается, что здесь множество  $U^E$  может быть несвязным, и тогда связка аппроксимирующих прямых расщепляется на несколько частей. (Несвязность  $U^E$  – аналог несвязности средних по  $P_E$  [76]). Эту возможность иллюстрирует следующий пример.

**Пример 2.2.** Пусть исходные данные (2.1) заданы в Таблице 1.1. При расчетах для аппроксимации множества  $U^E$  использовалась сетка с шагом 0.001 по параметру  $a$  и с шагом 0.01 по параметру  $b$ , покрывающая область  $[0, 1.1] \times [0, 5]$  для параметров  $(a, b)$ . На рисунке 2.3 представлены связка графиков аппроксимирующих функций и множество  $U^E$ , а

также графики функций (2.8) – (2.10) и исходные точки. Рисунок 2.3 показывает, что множество  $U^E$  не связно, и из-за этого связка линий аппроксимации расщепилась на несколько частей, а множества  $Y^E(x)$  не являются числовыми промежутками.

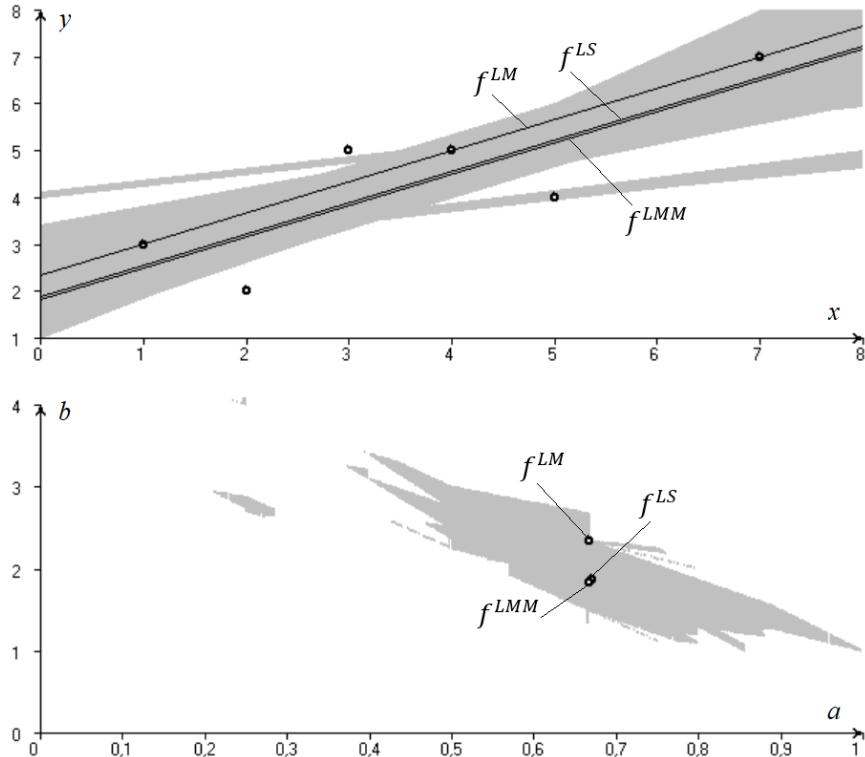


Рисунок 2.3 – Связка прямых и множество  $U^E$

#### 2.1.6 Аппроксимация при равноважных критериях со шкалой первой порядковой метрики

Пусть, наконец, все критерии имеют равную важность, и шкала критериев является шкалой первой порядковой метрики (см. [83]): информация  $\Gamma = E\Delta$ . Такая шкала обладает следующим свойством: если в произвольной векторной оценке  $z$ , в которой  $z_i > z_j$ , заменить  $z_i$  на  $z_i - q$ , а  $z_j$  – на  $z_j + q$ , но при этом обеспечить выполнение неравенства  $z_i - q \geq z_j + q$ , где  $q$  – положительное число, то полученная таким образом векторная оценка будет более предпочтительна, чем исходная. Это означает, что при указанном переходе большее отклонение от одной из точек выборки уменьшается за счет равного по величине увеличения меньшего отклонения от другой точки выборки, так что набор отклонений становится более близким к идеальному набору из минимально возможных равных уклонений. Такой подход является аналогом известного в теории общественного благосостояния принципа Пигу – Дальтона: «Общество станет лучше, если единица дохода будет передана от более богатого индивида более бедному» [84].

В рассматриваемом случае имеем частичный порядок  $P^{E\Delta}$ , задаваемый следующим решающим правилом [85]:

$$z'P^{E\Delta}z'' \Leftrightarrow z'_{[1]} \leq z''_{[1]}, z'_{[1]} + z'_{[2]} \leq z''_{[1]} + z''_{[2]}, \dots, \quad (2.16)$$

$$z'_{[1]} + z'_{[2]} + \dots + z'_{[n]} \leq z''_{[1]} + z''_{[2]} + \dots + z''_{[n]}, z'_\Delta \neq z''_\Delta. \quad (2.17)$$

Поскольку  $P^{E\Delta} \supset P^E$ , имеет место включение  $U^{E\Delta} \subseteq U^E$ .

Пример 2.3. Пусть исходные данные (2.1) заданы в Таблице 2.1. На рисунке 2.4 представлены связки линий аппроксимации и множество  $U^{E\Delta}$ , а также линии регрессии (2.8) – (2.10) и исходные точки.

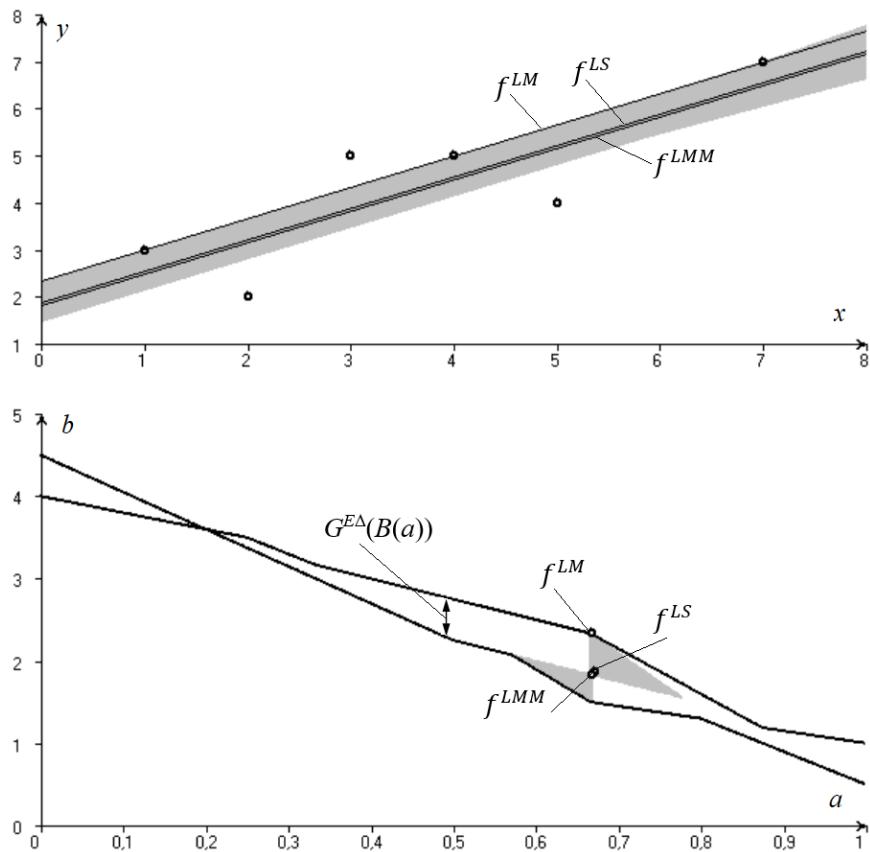


Рисунок 2.4 – Связка прямых, множество  $U^{E\Delta}$  и графики границ отрезков  $G^{E\Delta}(B(a))$

ПН-средние  $G^{E\Delta}(B(a))$  представляют собой отрезок  $[\alpha, \beta]$ , границы которого легко вычисляются по формулам [77]:

$$\alpha = \frac{1}{2} \min_{p \in H} (b_{(p)} + b_{(n+1-p)}), \quad (2.18)$$

$$\beta = \frac{1}{2} \max_{p \in H} (b_{(p)} + b_{(n+1-p)}), \quad (2.19)$$

где  $H = \{1, 2, \dots, h\}$ , а  $h$  – целая часть числа  $(n+1)/2$ . Еще одно полезное свойство ПН-средних при информации  $E\Delta$  – устойчивость к изменению положения точек множества  $B(a)$ . Поскольку при небольшом изменении параметра  $a$  на  $a + \varepsilon$ , положение точек  $b_i$  множества  $B(a + \varepsilon)$  меняется относительно точек  $B(a)$  на величину  $\varepsilon x_i$ , то согласно [77], максимальное изменение границ множества  $G^{E\Delta}(B(a + \varepsilon))$  по сравнению с  $G^{E\Delta}(B(a))$  не превышает  $\varepsilon \cdot \max_{i \in H} |x_i|$ . Из этого следует, что функции границ  $\alpha(a)$  и  $\beta(a)$  должны быть непрерывными. Графики этих функций приведены на рисунке 2.4. Здесь множество  $U^{E\Delta}$  связно, и поэтому множества  $Y^{E\Delta}(x)$  являются числовыми промежутками.

Рассмотрим графоаналитический метод определения  $Y^{E\Delta}(x)$ . На рисунке 2.5 снизу представлено множество  $U^{E\Delta}$  и графики прямой  $b = -x^0 a + y$  для точки  $x^0 = 2$  в двух крайних положениях, которые определяют приближенные значения  $b_{\min} = 2.85$  и  $b_{\max} = 3.7$ , которым соответствует  $Y^{E\Delta}(2) = [2.85, 3.7]$ . Это согласуется с результатом, который можно получить при помощи рисунка 2.5 сверху, если построить вертикальную прямую  $x = 2$  и рассечь ею связку прямых.

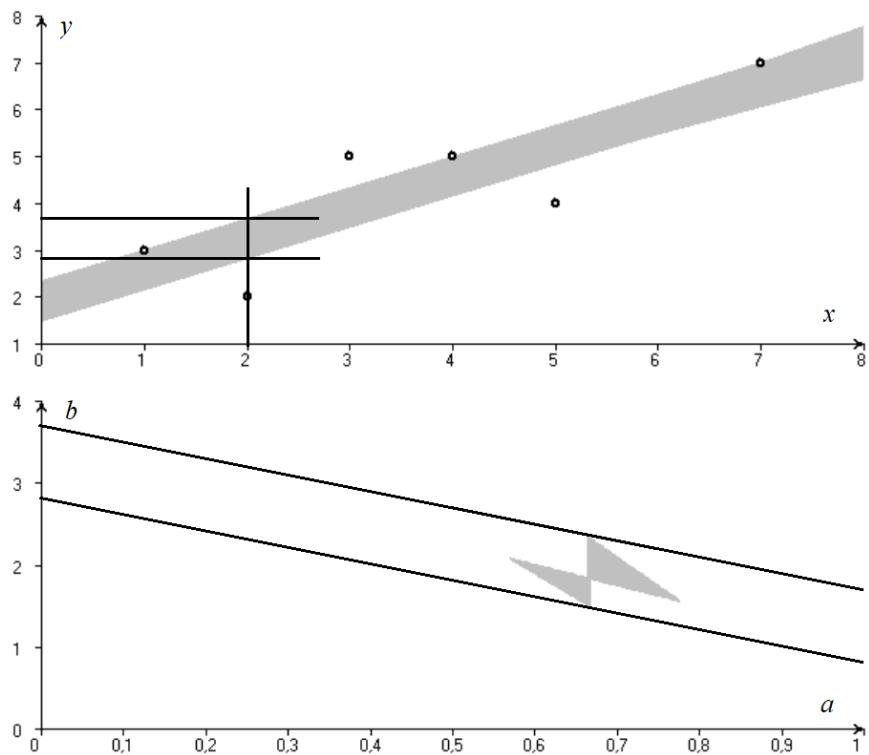


Рисунок 2.5 – Графоаналитический метод определения  $Y^{E\Delta}(x)$  при  $x = 2$

Рассмотрим более сложный пример с реальными данными.

Пример 2.4. В Таблице 2.2 представлены исходные данные (2.1) по ВВП на душу населения в странах Евросоюза за 2000–2021 гг. [86]. Нумерация годов  $x_i$  для удобства

начинается с 0, т. е.  $x_1 = 0$  соответствует 2000 году. Величины  $y_i$  заданы в тысячах евро на человека.

Таблица 2.2 – Исходные данные по ВВП на душу населения в Евросоюзе

$x_i$	0	1	2	3	4	5	6	7	8	9	10
$y_i$	22.45	22.90	23.09	23.21	23.71	24.06	24.82	25.51	25.58	24.41	24.90
$x_i$	11	12	13	14	15	16	17	18	19	20	21
$y_i$	25.32	25.10	25.06	25.42	25.95	26.41	27.11	27.62	28.07	26.39	27.83

Здесь уравнения аппроксимирующих функций, определяемые МНК, МНМ и МНММ, соответственно, таковы:

$$f^{LS}(x) \approx 0.2313 x + 22.7953; \quad (2.20)$$

$$f^{LM}(x) \approx 0.2478 x + 22.5944; \quad (2.21)$$

$$f^{LMM}(x) \approx 0.197 x + 23.3885. \quad (2.22)$$

При расчетах для аппроксимации множества  $U^{E\Delta}$  использовалась сетка с шагом 0.0005 по параметру  $a$  и с шагом 0.005 по параметру  $b$ , покрывающая область  $[0.15, 0.3] \times [22, 24]$  для параметров  $(a, b)$ . Связка аппроксимирующих функций представлена в области  $[-1, 22] \times [22, 28]$  для  $(x, y)$  на рисунке 2.6 сверху, а множество точек  $U^{E\Delta}$  представлено в области  $[0.15, 0.3] \times [22, 24]$  для  $(a, b)$  на рисунке 2.6 снизу. По разбросу значений параметра  $a$  примерно от 0.2 до 0.25 можно судить о ежегодном среднем приращении показателя ВВП на данном периоде времени.

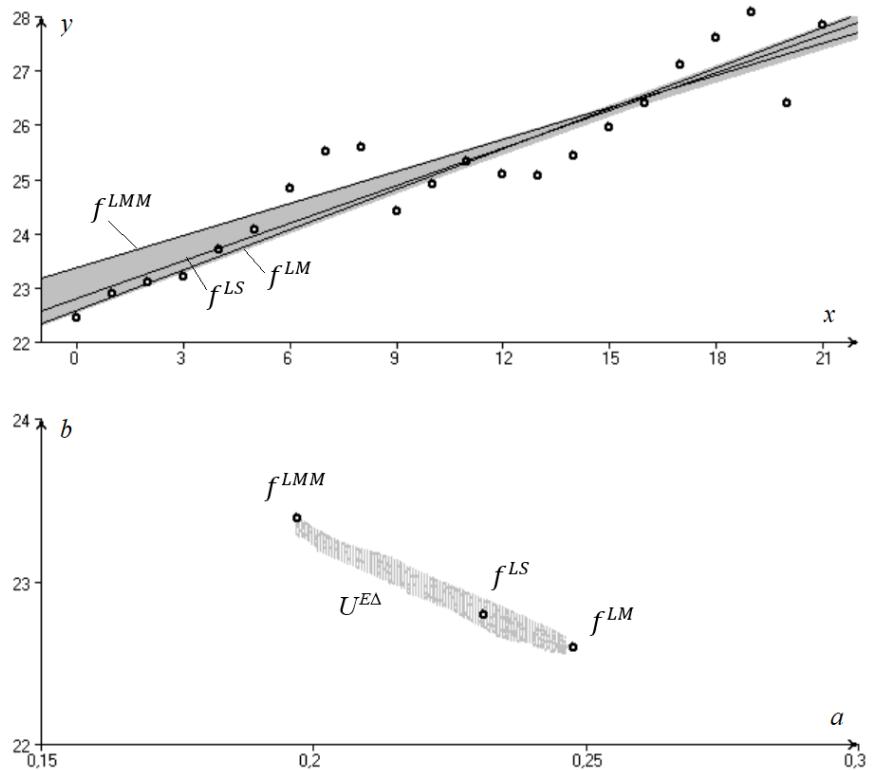


Рисунок 2.6 – Связка прямых и множество  $U^{E\Delta}$

### 2.1.7 О взаимосвязи недоминируемых и классических аппроксимирующих функций

Поскольку информация  $E\Delta$  уточняет информацию  $E$ , то, как уже указывалось выше, верно  $P^\emptyset \subset P^E \subset P^{E\Delta}$ , и поэтому имеют место включения  $U^\emptyset \supseteq U^E \supseteq U^{E\Delta}$ , а соответствующие связки аппроксимирующих функций последовательно входят друг в друга. Это хорошо видно при сравнении рисунков 2.1, 2.3 и 2.4.

Интересен также вопрос о том, всегда ли графики функций  $f^{LS}$ ,  $f^{LM}$  и  $f^{LMM}$ , полученные классическими методами – МНК, МНМ и МНММ – входят в связки прямых для  $\Gamma = \emptyset$ ,  $\Gamma = E$  и  $\Gamma = E\Delta$ , т. е. когда справедливы включения  $f^{LS} \in F^\emptyset$ ,  $f^{LM} \in F^E$  и  $f^{LMM} \in F^{E\Delta}$ .

Ответ на этот вопрос дает следующая теорема.

**Теорема 2.1.** Справедливы соотношения  $f^{LS} \in F^{E\Delta} \subseteq F^E \subseteq F^\emptyset$ ,  $f^{LM} \in F^E \subseteq F^\emptyset$  и  $f^{LMM} \in F^\emptyset$ . Если функция  $f^{LM}$  (соответственно функция  $f^{LMM}$ ) единственна, то  $f^{LM} \in F^{E\Delta}$  (соответственно  $f^{LMM} \in F^E$ ).

Доказательство теоремы опускается (оно есть в статьях, указанных в Заключении)

### 2.1.8 Заключение

Множественный подход к определению параметров аппроксимации выделяет не одну прямую линию аппроксимации, а связку таких линий. Множественность выделяемых значений параметров аппроксимации – плата за то, что при многокритериальной постановке задачи их оценивания не делается никаких дополнительных ограничительных допущений о характере (свойствах) исходных данных (например, о законе распределения их вероятностей). Получающуюся при предлагаемом подходе многозначную аппроксимирующую функцию можно интерпретировать двояко: либо как совокупность возможных однозначных функций аппроксимации, либо как множество функций, характеризующие аппроксимацию в целом.

За единственную линию аппроксимации, особенно в случае равноважных критериев со шкалой первой порядковой метрики, можно взять геометрическое место точек – центров тяжести (центроидов) множеств  $Y^\Gamma(x)$ .

Актуальной остается проблема разработки эффективных вычислительных методов построения множеств недоминируемых значений параметров аппроксимации  $U^\Gamma$ , а также множеств  $Y^\Gamma(x)$ . Такие методы должны учитывать специфические особенности отношений  $P_\Gamma$  для различных видов информации  $\Gamma$ .

Предложенный подход очевидным образом обобщается на случай аппроксимации функций нескольких переменных, но вычислительные проблемы из-за повышения размерности при этом существенно усложняются.

Поставленные перед началом исследований задачи выполнены полностью. Исследования проводились совместно с к.ф.-м.н. Нелюбиным А.П. Полученные результаты опубликованы в статьях [87, 88].

## 2.2 Разработка инструментария решения задачи о выборе оптимальной структуры курсов для смешанного обучения

### 2.2.1 Введение

Дистанционное обучение является одним из быстро развивающихся направлений высшего образования и, согласно многочисленным отчетам только в США, в настоящее время оно охватывает миллионы студентов [89]. Специалисты в области высшего образования продолжают обсуждать различные философские и педагогические проблемы дистанционного обучения, в том числе связанные с качеством образования, которое оно может обеспечить (по сравнению с традиционным высшим образованием) [90]. В то же время экономика дистанционного обучения становится одной из областей, которая начинает волновать администрации учебных заведений, так как это касается колледжей и университетов,

в которых курсы, доступные через Интернет, или уже предлагаются или планируются к использованию [91].

Одна из тем, активно обсуждаемых в научных публикациях в рамках экономики дистанционного обучения, касается конкретных программных комплексов, называемых смешанными курсами [92, 93, 94, 95]. Данные имеющихся исследований, анализирующих эффективность различных форм дистанционного обучения, позволяют предположить, что с точки зрения качества образования, смешанные курсы могут быть не менее эффективными, чем курсы, записываемые студентами при стандартной форме обучения [96, 97]. Более того, на это качество, по-видимому, положительно влияет так называемый «эффект взаимодействия сверстников», который неизбежно присутствует во всех смешанных курсах [98, 99]. Однако, проведение смешанного курса по конкретному предмету может потребовать от администрации колледжа/университета предоставить больше профессоров и ассистентов преподавателя, чем для онлайнового курса, в то время как в большинстве случаев, такое расширение преподавательского состава может быть осуществлено только в финансовых рамках бюджета колледжа/университета.

Удовлетворение желаемых требований к качеству при определенных финансовых ограничениях является типичной проблемой принятия решений, которая решается в промышленных, транспортных, сельскохозяйственных, деловых, финансовых и других системах (см., например, [100, 101, 102, 103, 104, 105, 106]). Эти системы уже давно используют инструменты поддержки принятия решений, позволяющие своим пользователям находить наилучшее размещение финансовых ресурсов, которые они могут позволить себе потратить, чтобы быть конкурентоспособными на соответствующих рынках. Думается, что администрация каждого колледжа/университета выиграла бы от наличия в своем распоряжении аналогичных инструментов, позволяющих решать проблемы, которые рынок образования потенциальных новых студентов ставит сейчас и будет ставить в будущем. Это происходит потому, что все колледжи/университеты конкурируют, особенно за новых студентов, заинтересованных в хорошем высшем образовании.

Администрациям колледжей/университетов следует обратить внимание на тот факт, что при должной рекламе обязательство предлагать смешанные курсы, основанные на записанных лекциях и семинарах профессоров из самых известных университетов мира, является фактором, положительно влияющим на набор студентов. Более того, принятие такого стратегического решения о предоставлении разработанных и рекламируемых таким образом смешанных курсов, скорее всего, положительно повлияет на положение колледжа/университета на рынке образования в целом, а также на его бюджет, по крайней мере, по двум причинам.

Во-первых, почти все потенциальные студенты-новички колледжей/университетов, которые заинтересованы в получении знаний (а не только диплома), мечтают учиться в Гарварде, МИТ, Кембридже, Оксфорде, Принстоне, Стэнфорде и других всемирно известных университетах. Тем не менее, многие из этих студентов не могут учиться там либо из-за низких оценок в школе, либо по финансовым причинам (или по обеим причинам). Если конкретный колледж/университет мог бы убедить потенциальных новых студентов вышеупомянутого типа в том, что большую часть курсов они будут слушать у того же самого лектора что и студенты из этих знаменитых университетов, но при этом им: (а) будут предлагаться эффективные подготовительные программы перед началом каждого такого курса, и (б) объясняться все нюансы курса в простой и понятной для них форме, это будет иметь значение в принятии этими потенциальными новыми студентами решений о поступлении на учебу, благоприятных для этого колледжа/университета.

Во-вторых, если таким потенциальным студентам-новичкам колледжа/университета будет предложено изучить и эти курсы, и подготовительные программы по доступной цене (которая в итоге может оказаться ниже, чем стоимость изучения тех же предметов онлайн в других колледжах/университетах), их желание поступить в этот колледж/университет только усилится.

Таким образом, основной вопрос, на который администрация колледжа/университета (заинтересованная в проведении смешанных курсов в принципе, должна ответить) выглядит следующим образом: как сделать описанную выше образовательную стратегию реальностью и как убедить потенциальных новых студентов поступить в этот колледж/университет?

Чтобы ответить на этот вопрос, заинтересованному колледжу/университету необходимо решить две группы проблем.

К первой группе относятся проблемы, связанные с поиском такой структуры каждого конкретного смешанного курса (основанного на использовании упомянутых выше записях лекций и семинаров профессоров из самых известных университетов мира и подготовительных программ), которая помогла бы администрации колледжа/университета убедить потенциальных новых студентов, что с первых же лекций они обязательно добьются успеха в изучении курса.

Ко второй группе относятся проблемы, связанные с финансовыми аспектами организации и проведения смешанных курсов таким образом, чтобы сделать их финансово доступными как для колледжа/университета, так и для студентов.

Решение задач первой группы (связанных с выбором структуры каждого смешанного курса) предполагает проведение комплексных исследований того, насколько разные

категории студентов фокусируют внимание как на предмете преподавания в целом, так и на конкретной информации, предлагаемой в курсах, так и на методиках, используемых при проведении этих курсов. Кроме того, исследования должны определить лучший способ проведения лекции путем выбора оптимальной последовательности фрагментов из записанных курсов и объяснений объема этих фрагментов, которые должны либо предшествовать конкретному фрагменту, либо сразу же следовать за ним. Эти исследования должны привести к разработке контрольных вопросов, которые будет задаваться преподавателем, преподающим курс, прежде чем перейти к следующему фрагменту. Однако все это должно осуществляться таким образом, чтобы не превратить лекцию в дискуссию с аудиторией, которая может занять большую часть лекционного времени. Наконец, структура подготовительных программ, предлагаемых колледжем/университетом, позволяющая студентам добиться успеха в изучении каждого смешанного курса, разработанного в выбранном порядке, должна быть определена и объявлена заранее.

Опубликованные рекомендации преподавателей, а также, возможно, даже специальные консультации этих преподавателей, могут существенно повлиять на эффективность результатов такого обучения. Разумеется, опыт и педагогические навыки тех, кто будет выбран для ведения смешанных курсов, имеют большое значение. Особенно, если каждому преподавателю, выбранному для преподавания смешанного курса, удается протестировать конкретную версию (или даже разные конкретные версии) этого смешанного курса, которых он предполагает придерживаться при работе с группами студентов, похожих на тех (или, по крайней мере, близких к тем), кто, как ожидается, будет изучать этот курс в колледже/университете.

Решение задач из второй группы предполагает определение:

(а) минимального бюджета для организации и преподавания набора из выбранных смешанных курсов, с тем чтобы обеспечить процент студентов, которые будут изучать курсы из этого набора и которые, как ожидается, добьются такого успеха в изучении каждого конкретного курса из набора — не ниже определенного желательного уровня;

(б) максимального процента студентов (которые будут изучать смешанные курсы из этого набора выбранных курсов), которые, как ожидается, преуспеют в изучении всех курсов из набора в рамках определенного бюджета.

Хотя для решения проблем из первой группы есть место (если нет необходимости) для использования математических методов, в настоящее время эти проблемы обсуждаются в научных публикациях по этим проблемам преимущественно на уровне гипотез [107, 108, 109]. Авторы этих публикаций либо только формулируют, либо формулируют и проверяют предложенные гипотезы, проводя опросы как среди студентов, так и среди преподавателей.

Публикации по проблемам из второй группы вообще редко считаются научными публикациями, и никакого полезного количественного анализа этих проблем пока в них не было предложено. Такое положение дел оставляет администрацию колледжей/университетов практически «безоружной» перед экономическими вопросами организации и проведения смешанных курсов или включения других форм дистанционного обучения в учебные программы.

В настоящем исследовании (и отчете) основное внимание уделено проблемам из второй группы упомянутых выше задач, которые связаны с поиском и анализом финансовых вариантов, доступных для реализации администрации колледжа/ университета, заинтересованной в организации и проведении смешанных курсов рассматриваемого типа. Представленные в отчете материалы позволяют администрации колледжа/ университета количественно оценить размер бюджета, необходимого для разработки и преподавания набора смешанных курсов любой выбранной структуры. Исходя из этого бюджета, администрация заинтересованного колледжа/ университета может построить схемы для расчета платы за обучение для своих потенциальных студентов в режиме онлайн, с учетом действующей финансовой политики колледжа/ университета.

Предлагаемый инструментарий может помочь администрациям заинтересованных колледжей/ университетов разработать оптимальную стратегию организации и проведения смешанных курсов с использованием доступных записанных материалов (курсов) профессоров ведущих университетов мира (как и, разумеется, с использованием материалов других вузов). Он реализуется с помощью стандартных пакетов прикладных программ, которые существуют, в частности, для персональных компьютеров.

Предлагаемый инструмент принятия решений, который предложен в [110], является первым средством, способным помочь решать конкретные экономические проблемы, с которыми сталкивается администрация колледжа/ университета, заинтересованная в организации и проведении смешанных курсов в своей попытке сделать дистанционное обучение частью учебного процесса обучения в нем. В то же время в [110] представлен краткий обзор научных публикаций по пяти важным (с точки зрения автора) направлениям, связанным с организацией использования конкретных записанных (видео) курсов при дистанционном обучении в целом и проведении смешанных курсов в колледжах/ университетах; фрагменты этого обзора представлены в настоящем отчете.

Помимо Введения, отчет содержит три пункта. В пункте 2.2.2 представлены упомянутые выше фрагменты обзора, опубликованного автором в [110], позволяющие получить представление об основных направлениях исследований в области экономики онлайн образования. В пункте 2.2.3 описана новая математическая модель, предложенная автором для

расчета оптимальной стратегии найма преподавателей для набора смешанных курсов с использованием записанного материала (видео) лекций профессоров крупнейших колледжей/университетов мира. На основе этой модели сформулированы две задачи целочисленного программирования, формализующие задачи, указанные во Введении. В пункте 2.2.4 сформулирован ряд проблем, связанных с проведением смешанных курсов, которые могут представлять интерес для администрации колледжей/университетов и могут стать предметом дальнейших научных исследований в области экономики онлайн образования.

## 2.2.2 Краткий обзор научных публикаций по использованию смешанных курсов в образовательном процессе

Несмотря на (а) важность использования онлайн курсов в рамках смешанного обучения как способа радикального улучшения качества образования во всем мире, и (б) очевидную необходимость решения экономических проблем, связанных с наймом новых университетских преподавателей, для проведения конкретных смешанных курсов, как из-за границы, так и из других университетов в стране, (по крайней мере, с точки зрения автора) количественный анализ этих проблем не проводился с достаточной глубиной. Насколько известно автору, только качественные соображения из двух перечисленных аспектов онлайн образования, как в целом, так и в отношении смешанных курсов, до сих пор были представлены либо в форме тематических исследований, либо в форме открытых дискуссий. Как можно видеть из обзора, представленного автором в [110] (фрагменты из которого приводятся в настоящем разделе отчета), а также из публикаций, указанных во «Введении», имеющиеся научные исследования в области смешанного обучения, охватывающие пять важных аспектов образования, связанных со смешанным обучением (а) описание общих преимуществ дистанционного обучения онлайн, (б) анализ особенностей смешанных курсов, (в) построение подходов к оценке эффективности смешанного обучения, (г) обсуждение “эффекта взаимодействия сверстников” в смешанном обучении и (д) выбор структуры смешанного обучения охватывают лишь качественный аспект соответствующих тем.

Цель представленного ниже обзора — дать представление о:

- сути указанных выше (и рассмотренных в [110]) пяти тем, затронутых современными публикациями по смешанному обучению;
- состоянии дел в этой области, что послужило мотивацией предлагаемого в настоящем исследовании (как в публикации [110], так и в настоящем отчете) подхода к количественному анализу проблемы использования (доступных) онлайн (видео) курсов профессоров ведущих университетов мира при разработке курсов по смешанному обучению студентов колледжей/университетов.

Проблемы онлайн дистанционного обучения. Важная миссия онлайн образования, близкая к теме настоящего исследования, состоит в предоставлении возможностей для доказательства высокого качества образования, предлагаемого самыми престижными факультетами/университетами мира, до колледжей/университетов более скромного (в настоящее время) уровня качества образования. Такие колледжи/университеты могут, тем не менее, обеспечить более высокий (чем в настоящее время) уровень качества образования, финансово доступный как для этих учебных заведений, так и для их студентов, широко используя онлайн (видео) курсы, разработанные ведущими колледжами/ университетами мира [91], в частности, путем включения записанных фрагментов этих онлайн курсов в смешанные курсы, которые они могут предложить своим студентам. Помимо социальных задач онлайн образования, администрациям указанных колледжей/университетов необходимо изучить две группы проблем, связанных с дистанционным обучением.

К первой группе этих проблем относятся те, которые связаны с а) «дистанционным обучением», такие как общение среди онлайн обучающихся [90], б) подходами к структурированию оценок изучаемых частей онлайн курса, влияющими на стратегии обучения студентов [110], в) созданием соответствующей социальной среды, влияющей на мотивацию к обучению [111] и поощрением присутствия феномена так называемого «массового обучения» [112], г) выбором структуры инструкций, предоставляемых во всех формах онлайн обучения [113], а также взаимоотношениями учитель-ученик, влияющими на достижение учащимися успехов [114], д) эффективностью онлайн обучения в форме дискуссий и так называемого «тихого обучения» [115], е) педагогическими аспектами восприятия знаний обучающимися студентами и поведения участников онлайн образования, а также микро- и макро-уровнями среды, окружающей онлайн образование [116].

Вторая группа указанных проблем интересует прежде всего преподавателей и администрации колледжей/университетов, участвующих в проведении дистанционных курсов обучения, и включает такие проблемы как подготовка учителей к проведению онлайн курсов [117] и поощрение преподавателей к переходу на онлайн обучение [118] или/и (по крайней мере) к использованию онлайн материалов в смешанных курсах. Среди рассматриваемых проблем второй группы, в частности, изучаются такие проблемы как а) получение информации об эффективности онлайн обучения [119], б) сравнение этой эффективности с эффективностью традиционного образования [120], в) анализ примеров успешной замены онлайн занятий дистанционным обучением [121], г) описание основных преимуществ онлайн обучения [122], д) создание обучающих сообществ [123], е) обеспечение доступности онлайн образования людям с ограниченными возможностями [124].

Обе группы вышеуказанных проблем изучаются в рамках ответов на некоторые анкеты [125, 126, 127, 128] или тематических исследований [111, 124, 129].

Особенности смешанных курсов. Сравнение успеваемости студентов в рамках традиционного обучения и смешанного обучения представлено в [92], где авторы, в частности, приходят к (довольно естественному) выводу о том, что студенты с высокими средними баллами в изучении предыдущих традиционных курсов достигают лучших результатов и при изучении смешанных курсов. Авторы [93] приходят к аналогичному (естественному) выводу, о том, что фактическое участие в работе курса (посещение) является одним из ключевых факторов для прогнозирования итоговых оценок за онлайн курс. Анализ влияния смешанного обучения на изучение как STEM, так и not-STEM дисциплин, представленный в [130], показывает, что изучение STEM дисциплин в рамках смешанного обучения более эффективно, чем изучение их в рамках традиционного обучения. В то же время сравнение результатов экзаменов и викторин в смешанном и традиционном курсах, представленные в [95] для проведения вводных экономических курсов, не выявили существенных преимуществ смешанного обучения.

Восемь проблем, с которыми сталкиваются преподаватели, проводящие смешанные курсы, описаны в [131] как части индуктивных категорий—учебных процессов, проблем сообщества и технических проблем—выявленных в результате обработки ответов 117 преподавателей из четырех университетов Турции на вопросы, которые им задавали в ходе интервью. В исследованиях, представленных в [132], утверждается, что лишь меньшинство студентов используют определенные инструменты при изучении смешанных курсов, поддерживаемые системами управления контентом (CMS), и студенты регулируют использование этих инструментов по мере развития смешанного курса. В [133] изучаются подходы студентов к исследованию и сотрудничеству в ходе работы над дополнительным курсом, там же сообщается о разделении студентов на подгруппы внутри группы численностью более 200 студентов со схожими подходами к исследованиям и технологиями обучения внутри каждой подгруппы применительно к конкретному односеместровому курсу. Авторы исследования полагают, что такое разделение помогает объяснить, почему студенты из некоторых подгрупп более успешны в изучении курса, чем студенты из других подгрупп.

В обзоре 640 источников в рамках 20 исследований, представленных в [134], изложены четыре ключевые проблемы при разработке смешанных курсов: 1) обеспечение гибкости курса; 2) стимулирование взаимодействия учитель-ученик; 3) облегчение учебных процессов для студентов; 4) содействие созданию эмоционально мотивированного учебного климата и подходы к решению четырех указанных проблем. В [135] сравниваются оценки студентами онлайновых, смешанных и офлайн курсами и опыт их работы с ними.

Эффективность смешанного обучения. Исследование, связанное с внедрением подхода смешанного обучения к разработке курса, в котором обсуждаются как реакция студентов на среду смешанного обучения, так и мысли автора курса о требованиях, которым должен соответствовать успешный смешанный курс, описано в [136]. Совокупность узких мест современного высшего образования (какими они представляются авторам [137]) рассматривается, в частности, с точки зрения способности смешанных курсов способствовать эффективному решению соответствующих задач управления высшим образованием.

Двадцать исследований, проанализированных в рамках обзора, представленного в [138], позволяют его авторам предположить, что существуют две группы основных факторов, каждая из которых влияет на решение творческих задач в учебной среде, которые следует учитывать при построении смешанного курса. В первую группу входят четыре частных фактора, влияющих на процесс решения, а остальные пять факторов, определяющих смешанную среду обучения, образуют вторую группу.

Результаты эксперимента с адаптивной, интегрированной в учебный процесс онлайн системой средой, разработанной для поддержки процессов обучения, в которых участвуют несколько областей знаний, курсов и преподавателей обсуждаются в [96], где авторы публикации утверждают, что использование этой онлайн системы помогает успешно сочетать традиционные альтернативные методы обучения с различными средами обучения.

Эффект взаимодействия сверстников в обучении. Эффекты взаимодействия сверстников как явление, связанное с влиянием, которое взаимодействие со сверстниками может оказывать на поведение человека при групповом обучении, имеют некоторые общие черты. Это (в какой-то мере) объясняет почему эти эффекты в настоящее время изучаются, в основном, в общем виде (хотя они могут иметь определенные особенности в смешанном обучении и влиять на состав учащихся, которые, в частности, изучают смешанные курсы, и на достижение ими лучших академических результатов).

Основы влияния сверстников в высшем образовании, включая определение этого явления, обсуждаются в [98], где авторы утверждают, что эффекты взаимодействия сверстников в обучении влияют на экономику высшего образования, главным образом потому, что (по мнению этих авторов) они «устраняют неуклюжие аномалии в институциональном поведении колледжей и университетов, а также в экономической структуре высшего образования, если они существуют». В [139] обсуждаются три схемы взаимодействия учащихся в беседах со сверстниками: накопительно-исследовательская схема, схема вопросов и ответов под руководством преподавателя, и схема спорного исследования. Авторы [139] полагают, что использование указанных схем может повысить эффективность бесед со сверстниками при обучении по всем предметам.

Теоретическая модель, пытающаяся объяснить (и сравнить) различные эффекты взаимодействия сверстников в разных областях обучения основанная на исследовании поведения студентов первого курса государственного университета среднего размера в Италии представлена в [140]. Влияние эффекта взаимодействия сверстников на академическую успеваемость студентов изучалось в [141] на основе набора публикаций в этой области, и один из результатов, представленных в [141], состоит в утверждении о том, что студенты с более высокими способностями имеют более высокий положительный эффект взаимодействия сверстников в отношении других студентов с высокими способностями как на уровне школы, так и на уровне колледжа/университета.

Выбор структуры смешанного курса. В [109] обсуждается подход к структурированию бакалаврской программы по электротехнике, рассчитанной на группу студентов, изучающих этот предмет одновременно с работой (т. е. трудоустроенных). При изучении курсов программы используются несколько вариантов обмена информацией, такие как установление связей с курсом математики, получение справочной информации, необходимой для изучения курсов, электронные модульные анкеты, а также канал обратной связи, функционирующий на протяжении всего курса.

В [107] рассматривается методика проектирования (и структурирования) смешанных образовательных курсов для выявления перспектив обучения зачисленных студентов, а также особенностей восприятия ими курса, что (по мнению автора) помогает оценить их возможные достижения в результате изучения курса. Автор [107] выделяет четыре типа такого восприятия и предлагает свои наблюдения за академическими достижениями студентов с некоторыми из этих типов восприятия.

Подход к математическому мышлению, лежащий в основе структуры конечного курса в части курса, связанного с исчислением функций многих переменных, предложен в [142], где авторы утверждают, что создание способной и целенаправленной учебной среды адекватно развитию математического мышления студентов. В [143] предложен набор рекомендаций, основанных на теории учебного дизайна для преподавателей письма, который направлен на поддержку эффективного обучения студентов. Авторы этой публикации предлагают преподавателю сосредоточиться на пяти вопросах при разработке курса: 1) обосновании необходимости курса; 2) структуре аудитории для изучения курса; 3) преимуществах курса в онлайн форме; 4) основных педагогических принципах, используемых при его проведении; 5) доступных ресурсах.

Концепция смешанного обучения, а также базовые элементы модели обучения, обсуждается в [144]. Набор руководств, документов и публикаций, связанных с практичес-

скими аспектами разработки смешанных курсов, а также общие принципы разработки таких курсов, соответствующая терминология и стратегии, необходимые для удовлетворения аккредитационных требований, рассматриваются в [145]. Авторы этой публикации обсуждают подходы к использованию онлайн технологий и реализации курсов, а также проблемы и трудности, которые следует учитывать при разработке структуры смешанных курсов, и намечают направления дальнейших исследований в этой области. Подробные инструкции по разработке смешанных курсов для достижения целей обучения, а также некоторые инструменты такой разработки и шаблоны представлены в [108] и в [146], где содержатся как теоретические, так и практические соображения, связанные с разработкой структуры смешанных курсов, и отражается опыт авторов этих книг в разработке таких курсов и в преобразовании традиционных курсов в курсы смешанного обучения.

Как видно из представленного обзора, (и как было отмечено во Введении к настоящему отчету), дискуссии и рассуждения по всем вышеперечисленным пяти направлениям на сегодняшний день (по мнению автора) проводятся на уровне лишь философских и педагогических наблюдений и предложений. В то же время, очевидно, что конкурентоспособность колледжа/университета на рынке потенциальных новых студентов не может основываться только на таких наблюдениях и дискуссиях. Администрация колледжа/университета, обладающая инструментами для принятия стратегических решений на основе формализованного финансового анализа, из которых экономически доступные решения могут быть приняты, несомненно, будет иметь существенные конкурентные преимущества на указанном рынке.

### 2.2.3 Постановки задач и их математические формулировки

Университет/колледж (для определенности Университет далее в настоящем отчете) намеревается начать преподавание смешанных курсов на одном из иностранных языков, например, на английском языке, и ему необходимо решить, как организовать это обучение. Число смешанных курсов, преподаваемых на английском языке, равно  $K$ , а количество преподавателей-носителей английского языка, способных (а) охватить все эти  $K$  курсов, и (б) быть нанятыми Университетом из-за границы, равно (или не превышает)  $L$ . Помимо привлечения преподавателей из-за границы для проведения указанных  $K$  смешанных курсов, Университет рассматривает возможность предложить преподавать некоторые курсы из этого набора  $K$  смешанных курсов (или все  $K$  смешанных курсов) (не более чем  $R$ ) преподавателям-носителям языка страны Университета, хорошо владеющих английским языком. С этой целью Университет рассматривает потенциальных кандидатов на преподавание смешанных курсов из числа тех, кто в настоящее время работает либо в Университете, либо

в других университетах страны. Университет планирует покупать (или брать из открытых источников) записанные онлайн (видео) курсы, которые преподают выдающиеся профессора из ведущих университетов мира, и использовать эти записи при разработке всех вышеупомянутых  $K$  смешанных курсов, независимо от того, кто, в конечном итоге, будет выбран для преподавания конкретного курса. Каждый из преподавателей: (а) приглашенных из-за границы; (б) приглашенных из других университетов страны; (в) работающих в настоящее время в Университете, которые приглашены преподавать конкретный курс из набора  $K$  смешанных курсов, должны преподавать его “одним и тем же способом”. Именно, этот преподаватель должен преподавать такой курс в форме лекций и семинаров, а избранные для этих лекций и семинаров материалы должны быть основаны на (или, по крайней мере, использовать) вышеупомянутые материалы, полученные из записанных онлайн (видео) курсов.

Каждый из преподавателей, которых Университет считает приглашенными для преподавания курсов из набора  $K$  смешанных курсов (из числа тех, кто может быть выбран из профессоров, работающих в настоящее время в Университете, нанят из других вузов страны или приглашен из-за границы), может преподавать не более определенного числа таких смешанных курсов. Если преподавателю из Университета поручено преподавать некоторые курсы из набора  $K$  смешанных курсов, ее/его существующие обязанности по преподаванию в Университете курсов на языке страны должны быть выполнены другими преподавателями, либо работающими в настоящее время в Университете, либо теми, кто будет нанят из других университетов страны (но не кем-либо из тех, кто приглашен для преподавания курсов из набора  $K$ ) на почасовой основе. Каждый потенциальный кандидат на преподавание любого конкретного курса из набора  $K$  смешанных курсов проходит проверку экспертами, признанными Университетом. Такие эксперты должны оценить процент студентов, которые с вероятностью, имеющей приемлемое для Университета значение, преуспеют в изучении этого курса, если указанный кандидат будет выбран для преподавания курса.

Университет заинтересован в оценке двух (указанных во Введении) чисел, связанных с организацией и проведением рассматриваемых  $K$  смешанных курсов:

– минимального бюджета для организации преподавания этих  $K$  смешанных курсов на английском языке, чтобы гарантировать, что процент студентов Университета (из числа тех, кто будет изучать курсы из этого набора), которые, как ожидается, преуспеют в изучении каждого конкретного курса (из этих  $K$  смешанных курсов), будет не ниже определенного желательного для Университета процента,

– максимального процента студентов (из числа тех, кто будет изучать курсы из набора  $K$  смешанных курсов), которые, как ожидается, преуспеют в изучении каждого конкретного курса (из этих  $K$  смешанных курсов) при конкретном бюджете, который Университет может позволить себе потратить на организацию и проведение этих курсов.

Все преподаватели-носители языка страны (как из Университета, так и из других вузов страны), которые потенциально способны преподавать курсы на английском языке из набора  $K$  смешанных курсов, далее называются разработчиками курсов. Если курс из смешанного курса  $K$  должен преподаваться разработчиком-носителем английского языка, приглашенным из-за границы, студенты, включенные в группы, назначенные для изучения этого курса, должны до начала курса (а) пройти углубленный курс английского языка, (б) пройти подготовительные программы и иметь соответствующие учебные пособия, чтобы быть готовыми к пониманию фрагментов из соответствующих записанных (видео) курсов.

Пусть:

$K$  — число смешанных курсов, которые Университет планирует преподавать на английском языке в ближайшие  $T$  лет, ежегодно, начиная с некоторого момента  $T^*$ ,

$B$  — годовой бюджет, выделяемый Университетом для покрытия расходов, связанных с разработкой и проведением набора  $K$  смешанных курсов в течение  $T$  лет,

$B_0$  — стоимость записанных онлайн (видео) курсов, которые Университет планирует приобрести для использования в разработке набора  $K$  смешанных курсов,

$M$  — количество потенциальных разработчиков смешанных курсов, которые в настоящее время работают в Университете,

$R$  — число потенциальных разработчиков смешанных курсов, которые в настоящее время работают в других университетах страны, заинтересованных в работе в Университете,

$L$  — число потенциальных разработчиков-носителей английского языка из-за границы, которые могут удовлетворить потребности Университета преподавании курсов из набора  $K$  смешанных курсов и которых Университет финансово может позволить себе пригласить преподавать эти курсы,

$\nabla_{ik}$  — годовая надбавка к базовой зарплате потенциального разработчика  $i \in \overline{1, M}$ , который в настоящее время работает в Университете и который может быть выбран (назначен) для преподавания курса  $k$  из набора  $K$  смешанных курсов,  $k \in \overline{1, K}$ ,

$b_r$  — базовая годовая зарплата потенциального разработчика  $r \in \overline{1, R}$  из другого университета в стране, который может быть выбран для преподавания курсов из набора  $K$  смешанных курсов,

$\delta_{rk}$  — годовая надбавка к базовой зарплате потенциального разработчика  $r \in \overline{1, R}$  из другого университета в стране, который может быть выбран для преподавания курса  $k$  из набора  $K$  смешанных курсов,  $k \in \overline{1, K}$ ,

$g_r$  — затраты на размещение, связанные с наймом потенциального разработчика  $r \in \overline{1, R}$  из другого университета в стране, который может быть выбран для преподавания курсов из набора  $K$  смешанных курсов (принимающие некоторое конечное число значений),

$a_l$  — базовая годовая зарплата потенциального разработчика  $l \in \overline{1, L}$ , который может быть приглашен из-за границы для преподавания курсов из набора  $K$  смешанных курсов,

$\Delta_{lk}$  — годовая надбавка к базовой зарплате потенциального разработчика курса  $l \in \overline{1, L}$ , который может быть приглашен из-за границы для преподавания курса  $k$  из набора  $K$  смешанных курсов,  $k \in \overline{1, K}$ ,

$h_l$  — затраты на размещение, связанные с наймом потенциального разработчика  $l \in \overline{1, L}$ , который может быть приглашен из-за границы для преподавания курсов из набора  $K$  смешанных курсов (принимающие некоторое конечное число значений),

$d_i$  — почасовая зарплата преподавателя (из числа работающих в настоящее время в Университете), который может заменить разработчика  $i \in \overline{1, M}$  (которому поручено преподавать курс из набора  $K$  смешанных курсов) во всех видах деятельности, связанных с преподаванием курса, “освобождаемых” разработчиком  $i \in \overline{1, M}$ , которого она/он заменяет (если есть такие преподаватели),

$t_{ik}$  — количество часов в год, “освобождаемых” разработчиком  $i \in \overline{1, M}$ , который в настоящее время работает в Университете и который может быть выбран (назначен) для преподавания курса  $k \in \overline{1, K}$  из набора  $K$  смешанных курсов — в результате ее/его “переключения” на преподавание курса  $k \in \overline{1, K}$ , которые должны быть охвачены другими преподавателями (либо из числа преподавателей, кто в настоящее время работает в Университете, либо из числа тех, кто может быть приглашен из других университетов страны),

$\alpha_{ik}$  — экспертная оценка процента студентов, которые, как ожидается, преуспеют в изучении смешанного курса  $k$ , если этот курс будет преподаваться разработчиком  $i$ , работающим в настоящее время в Университете,  $i \in \overline{1, M}$ ,  $k \in \overline{1, K}$ ,

$\beta_{rk}$  — экспертная оценка процента студентов, которые, как ожидается, преуспеют в изучении смешанного курса  $k$ , если этот курс будет преподаваться разработчиком  $r$ , приглашенным из другого университета в стране  $r \in \overline{1, R}$ ,  $k \in \overline{1, K}$

$\gamma_{lk}$  — экспертная оценка процента студентов, которые, как ожидается, преуспеют в изучении смешанного курса  $k$ , если этот курс будет преподаваться разработчиком  $l$ , приглашаемым из-за границы  $l \in \overline{1, L}$ ,  $k \in \overline{1, K}$ ,

$\omega_k$  — желательный для Университета процент студентов, ожидаемых успешно изучить смешанный курс  $k$ ,  $k \in \overline{1, K}$ ,

$u_k$  — булева переменная, равная 1, если предполагается, что курс  $k$  из набора  $K$  смешанных курсов будет преподаваться разработчиком, приглашаемым из-за границы (так что специальный курс английского языка и подготовительные программы должны проводиться и учебные пособия должны выдаваться Университетом для студентов, которые выберут смешанный курс  $k$  для изучения), и равная 0 в противном случае,

$f_k$  — годовая стоимость проведения углубленного курса английского языка, проведения подготовительных программ и подготовки соответствующих учебных пособий для студентов, выбравших изучение смешанного курса  $k$ ,  $k \in \overline{1, K}$ ,

$x_{ik}$  — булева переменная, равная 1, если предполагается, что разработчику  $i \in \overline{1, M}$  из Университета будет поручено преподавать смешанный курс  $k \in \overline{1, K}$  и равна 0, в противном случае,

$s_r$  — булева переменная, равная 1, если предполагается, что разработчик  $r \in \overline{1, R}$  из другого университета в стране, будет приглашен для преподавания курсов из набора  $K$  смешанных курсов, и равна 0, в противном случае,

$y_{rk}$  — булева переменная, равная 1, если предполагается, что разработчик  $r \in \overline{1, R}$  из другого университета в стране будет приглашен для преподавания смешанного курса  $k \in \overline{1, K}$  и равна 0, в противном случае,

$w_l$  — булева переменная, равная 1, если предполагается, что разработчик  $l \in \overline{1, L}$  из-за границы будет приглашен для преподавания курсов из набора  $K$  смешанных курсов и равна 0, в противном случае,

$z_{lk}$  — булева переменная, равная 1, если предполагается, что разработчик  $l \in \overline{1, L}$  из-за границы будет приглашен для преподавания смешанного курса  $k \in \overline{1, K}$ , и равен 0, в противном случае,

$M\theta_i \subset \overline{1, K}$  — подмножество курсов из набора  $K$  смешанных курсов, которые разработчик  $i \in \overline{1, M}$  из Университета не может преподавать,

$R\theta_r \subset \overline{1, K}$  — подмножество курсов из набора  $K$  смешанных курсов, которые разработчик  $r \in \overline{1, R}$  из другого университета в стране не может преподавать,

$L\theta_l \subset \overline{1, K}$  — подмножество курсов из набора  $K$  смешанных курсов, которые разработчик  $l \in \overline{1, L}$  из-за границы, не может преподавать,

$v_i$  — максимальное число курсов из набора  $K$  смешанных курсов, которые разработчик из Университета может преподавать одновременно,  $i \in \overline{1, M}$ ,

$\mu_r$  — максимальное число курсов из набора  $K$  смешанных курсов, которые разработчик  $r \in \overline{1, R}$  из другого университета в стране может преподавать одновременно,

$\pi_l$  — число курсов из набора  $K$  смешанных курсов, которые преподаватель  $l \in \overline{1, L}$  из-за границы может преподавать одновременно,

$q_k$  — годовой бюджет, который Университет может потратить на дополнительную заработную плату (надбавку), выплачиваемую тому из разработчиков, который (как предполагается) будет преподавать курс  $k \in \overline{1, K}$  из набора смешанных курсов, включая (если необходимо) проведение курса английского языка и подготовительной программы, а также подготовку соответствующих учебных пособий и тому преподавателю, работающему в Университете, который заменяет указанного разработчика (если необходимость в такой замене возникнет).

### Предположения.

1. Число потенциальных разработчиков из других университетов страны, из которых следует выбирать, чтобы (а) пригласить их стать разработчиками смешанных курсов и (б) заменить часы, “освобождаемые” этими разработчиками из Университета, является извест-

ным целым числом. Разработчикам смешанных курсов из множеств  $\overline{1, M}$  и  $\overline{1, R}$  выплачивается дополнительная зарплата за разработку и проведение курсов из набора  $K$  смешанных курсов. Эта разработка осуществляется в соответствии со структурой каждого смешанного курса, которые эти разработчики приглашаются преподавать, и оплачивается Университетом. При этом базовая зарплата разработчика  $i \in \overline{1, M}$  остается такой же, какой она была до назначения этого разработчика на преподавание смешанных курсов (и она не покрывается бюджетом  $B$ , выделяемым на покрытие расходов, связанных с наймом разработчиков для преподавания курсов из набора  $K$  смешанных курсов).

2. Разработчики, приглашенные из-за границы для преподавания курсов из набора  $K$  смешанных курсов, прибывают с уже разработанными такими курсами (в соответствии со структурой этих курсов, которая должна быть заранее предоставлена администрацией Университета). Тем не менее, этим разработчикам выплачиваются дополнительные зарплаты (надбавки)  $\nabla_{ik}$ ,  $l \in \overline{1, L}$ ,  $k \in \overline{1, K}$  за преподавание на этих курсах, в то время как их основные зарплаты определяются соответствующими контрактами между каждым из них и Университетом. Все заработные платы, выплачиваемые каждому приглашенному Университетом разработчику, как каждому приглашенному из других университетов страны, так и каждому приглашенному из-за границы, покрываются бюджетом  $B$ , выделяемым на покрытие расходов, связанных с наймом разработчиков и проведения ими курсов из набора  $K$  смешанных курсов.

3. Подготовительные программы и углубленные курсы английского языка проводятся и учебные пособия предоставляются для смешанного курса только в том случае, если этот курс преподается разработчиком, приглашенным из-за границы (хотя, как правило, по крайней мере, учебные пособия могут быть использованы преподавателями любого из двух других типов при проведении этого курса). Введение булевых переменных, подобных  $u_k$  может отразить этот случай в системе ограничений (2.11), (2.12).

4. Все контракты с каждым разработчиком, приглашенным Университетом из-за границы для преподавания курсов из набора  $K$  смешанных курсов, также как и все контракты с каждым разработчиком из других университетов страны, нанятым Университетом для преподавания смешанных курсов из этого набора, заключаются сроком на  $T$  лет, начиная с момента  $T^*$ .

5. Учитывая использование записанных фрагментов (видео) лекций при разработке смешанных курсов, а также приглашение разработчиков для преподавания смешанных курсов.

сов не только из Университета (включая разработчиков, приглашенных из-за границы), администрация Университета выбирает общий набор смешанных курсов, которые она планирует начать предлагать в Университете, начиная с момента  $T^*$ , ежегодно, в течение  $T$  лет, так что ни один новый смешанный курс не будет добавлен Университетом в этот общий набор в течение указанных  $T$  лет.

6. Вся информация, необходимая для оценки значений параметров  $\alpha_{ik}, i \in \overline{1, M}$ ,  $k \in \overline{1, K}$ ,  $\beta_{rk}$ ,  $r \in \overline{1, R}$ ,  $k \in \overline{1, K}$ , и  $\gamma_{lk}$ ,  $l \in \overline{1, L}$ ,  $k \in \overline{1, K}$ , считается доступной в отношении всех разработчиков, которых предполагается пригласить в Университет для преподавания смешанных курсов из множества  $K$ . Потенциальным разработчикам из-за рубежа, которые рассматриваются Университетом на предмет их приглашения преподавать курсы из множества  $K$ , предлагается предоставить часть своих записанных (видео) лекций по соответствующим предметам заранее, чтобы позволить администрации Университета и экспертам сделать соответствующие оценки. Получив такие оценки, Университет организует тестирование представленных материалов на выбранной группе своих студентов, чтобы выяснить, какие виды подготовительных программ (а, возможно, и углубленных курсов английского языка) целесообразно провести и какие виды учебных пособий целесообразно обеспечить с тем, чтобы помочь студентам лучше понять содержание смешанных курсов и язык, который будет использоваться для преподавания этих курсов.

В приводимой ниже системе ограничений, предложенной (в [110] и в настоящем отчете), предполагается, что параметры  $\alpha_{ik}, \beta_{rk}$  и  $\gamma_{lk}$ , представляющие собой экспертные оценки процента студентов, которые (как предполагается) выберут для изучения курс  $k \in \overline{1, K}$ , добьются успеха в изучении этого курса, если курс преподается разработчиком из университета, или разработчиком из другого университета в стране, или разработчиком, приглашенным из-за границы, соответственно — это известные числа, предоставленные администрации Университета лицами, признанными Университетом как эксперты. Эта система ограничений

$$\begin{aligned}
& \sum_{i=1}^M x_{ik} + \sum_{r=1}^R y_{rk} + \sum_{l=1}^L z_{lk} = 1, \quad k \in \overline{1, K}, \\
& \sum_{i=1}^M x_{ik} \leq 1, \quad k \in \overline{1, K}, \quad \sum_{k=1}^K x_{ik} \leq v_i, \quad i \in \overline{1, M}, \\
& x_{ik} = 0, \quad k \in M \theta(i) \subset \overline{1, K}, \quad i \in \overline{1, M}, \\
& y_{rk} \leq s_r \leq \sum_{k=1}^K y_{rk}, \quad r \in \overline{1, R}, \quad k \in \overline{1, K}, \\
& \sum_{r=1}^R y_{rk} \leq 1, \quad k \in \overline{1, K}, \quad \sum_{k=1}^K y_{rk} \leq \mu_r, \quad r \in \overline{1, R}, \\
& y_{rk} = 0, \quad k \in R \theta(r) \subset \overline{1, K}, \\
& \sum_{l=1}^L z_{lk} - u_k = 0, \quad k \in \overline{1, K}, \\
& z_{lk} \leq w_l \leq \sum_{k=1}^K z_{lk}, \quad l \in \overline{1, L}, \quad k \in \overline{1, K}, \\
& \sum_{l=1}^L z_{lk} \leq 1, \quad k \in \overline{1, K}, \quad \sum_{k=1}^K z_{lk} \leq \pi_l, \quad l \in \overline{1, L}, \\
& z_{lk} = 0, \quad k \in L \theta(L) \subset \overline{1, K}, \\
& \sum_{l=1}^L \Delta_{lk} z_{lk} + \sum_{r=1}^R \delta_{rk} y_{rk} + \sum_{i=1}^M \nabla_{ik} x_{ik} + \sum_{i=1}^M d_i t_{ik} x_{ik} + f_k u_k \leq q_k, \quad k \in \overline{1, K},
\end{aligned} \tag{2.23}$$

и

$$\begin{aligned}
& \sum_{l=1}^L w_l (a_l + h_l) + \sum_{k=1}^K \sum_{l=1}^L \Delta_{lk} z_{lk} + \sum_{r=1}^R s_r (b_r + g_r) + \sum_{k=1}^K \sum_{r=1}^R \delta_{rk} y_{rk} + \\
& \sum_{k=1}^K \sum_{i=1}^M \nabla_{ik} x_{ik} + \sum_{k=1}^K \sum_{i=1}^M d_i t_{ik} x_{ik} + \sum_{k=1}^K f_k u_k \leq B - B_0.
\end{aligned} \tag{2.24}$$

должна выполняться начиная с первого года в множестве из  $T$  лет.

Естественно также предполагать, что а) в Университете, намеревающемся предлагать своим студентам смешанные курсы из множества  $K$ , работают преподаватели, способные разрабатывать и вести хотя бы некоторые курсы из множества  $K$  (т. е., преподаватели, которых Университет рассматривает как потенциальных разработчиков смешанных курсов из множества  $\overline{1, K}$ ), что отражается отсутствием в системе ограничений (2.11) некоторых ограничений для переменных  $x_{ik}$ , которые аналогичны соответствующим ограничениям для переменных  $y_{rk}$  и  $z_{lk}$ , б) базовая заработка разработчика из Университета не

является частью бюджета  $B$  и бюджетов  $q_k, k \in \overline{1, K}$  (что имеет место независимо от числа смешанных курсов, которые этот разработчик может быть приглашен преподавать).

Исходя из системы ограничений (2.23), (2.24) и с учетом сделанных выше предположений, задачи, поставленные ранее в этом разделе отчета, можно математически сформулировать следующим образом:

Задача 1. Эта задача состоит в максимизации минимального процента от общего числа студентов, которые, как ожидается, преуспеют в изучении курса из набора  $K$  смешанных курсов, когда эти курсы считаются одинаково важными с точки зрения Университета, и эта задача формулируется исходя из фиксированного годового бюджета  $B$  на  $T$  последовательных лет и является задачей булева программирования

$$\min_{k \in \overline{1, K}} \left( \sum_{i=1}^M \alpha_{ik} x_{ik} + \sum_{r=1}^R \beta_{rk} y_{rk} + \sum_{l=1}^L \gamma_{lk} z_{ik} \right) \rightarrow \max_{(x_{ik}, y_{rk}, z_{ik}, u_k, s_r, w_l)} \quad (2.25)$$

с системой ограничений (2.23), (2.24).

Задача 2. Эта задача состоит в минимизации значения годового бюджета на проведение курсов из набора  $K$  смешанных курсов в первые  $T$  последовательных лет (начиная с некоторого момента  $T^*$ ) при условии достижения целевого значения процента студентов, которые, как ожидается, преуспеют в изучении этих курсов, определяемого числами  $\omega_k$ ,  $k \in \overline{1, K}$ , и является задачей булева программирования

$$\begin{aligned} & \sum_{l=1}^L w_l (a_l + h_l) + \sum_{k=1}^K \sum_{l=1}^L \Delta_{lk} z_{ik} + \sum_{r=1}^R s_r (b_r + g_r) + \sum_{k=1}^K \sum_{r=1}^R \delta_{rk} y_{rk} + \\ & \sum_{k=1}^K \sum_{i=1}^M \nabla_{ik} x_{ik} + \sum_{k=1}^K \sum_{i=1}^M d_i t_{ik} x_{ik} + \sum_{k=1}^K f_k u_k \rightarrow \min_{(x_{ik}, y_{rk}, z_{ik}, u_k, s_r, w_l)} \end{aligned} \quad (2.26)$$

с системой ограничений (2.23) и дополнительной системой ограничений

$$\sum_{i=1}^M \alpha_{ik} x_{ik} + \sum_{r=1}^R \beta_{rk} y_{rk} + \sum_{l=1}^L \gamma_{lk} z_{ik} \geq \omega_k, \quad k \in \overline{1, K}. \quad (2.27)$$

В обеих задачах предполагается, что системы ограничений (2.23), (2.24) (в задаче 1) и (2.23), (2.24), (2.27) (в задаче 2) совместны, что можно проверить с использованием методики, указанной, в частности, в [147], и (если потребуется), внести соответствующие поправки в эти системы ограничений, а задача 2 решается для конечного набора значений пар (известных) чисел ( $g_r, h_l$ ).

#### 2.2.4 Проблемы, связанные с проведением смешанных курсов

1. Ясно, что, в принципе, задачи булевого программирования (2.23)–(2.24) и (2.23), (2.14), (2.15) могут служить математическими формулировками задач, связанных с отысканием оптимального состава преподавателей также для проведения любого набора онлайн курсов в колледже/университете, например, по физике, математике, биологии, без использования (видео) лекций преподавателей других учебных заведений по этим предметам. В этом случае подготовительные программы по соответствующим предметам могут проводиться (а могут и не проводиться), учебные пособия по ним могут предлагаться (а могут и не предлагаться), а преподаватели из-за границы могут приглашаться (а могут и не приглашаться) для преподавания курсов по этим предметам; при этом в формулировках задач 1 и 2 некоторые переменные и параметры, присутствующие в системах ограничений (2.23), (2.24), следует опустить.

Такая возможность использовать предложенные задачи булева программирования вытекает из очевидного замечания: с экономической точки зрения, схема формирования набора смешанных курсов и набора стандартных онлайн курсов, в принципе, одна и та же, что должна отражать математическая формализация этих схем. Действительно, в обоих случаях необходимо определить оптимальное сочетание преподавателей из некоторого доступного их набора для преподавания курсов в определенных финансовых пределах. Однако, проведение стандартных онлайн курсов в колледже/университете, имеющем скромную репутацию в какой-либо конкретный период времени, существенно отличается от формирования в нем смешанных курсов, в частности, по набору проблем, требующих решения. Среди таких проблем, подлежащих решению для того, чтобы рассчитывать на то, что предлагаемые колледжем/университетом смешанные курсы будут привлекательны для его потенциальных новых студентов, будучи в пределах финансовых возможностей колледжа/университета, администрации этого колледжа/университета необходимо:

а) должным образом рекламировать эти курсы как курсы, которые будут преподаваться с использованием фрагментов (видео) лекций профессоров известных университетов мира,

б) убедить потенциальных новых студентов в том, что к каждому предлагаемому смешанному курсу прилагаются учебные пособия, предлагаются подготовительные программы, а, если необходимо, то и углубленные курсы английского языка, которые помогут каждому заинтересованному студенту добиться успеха в изучении этого смешанного курса,

в) выплачивать выбранным для ведения курсов разработчикам надбавку за разработку каждого смешанного курса (в соответствии со структурой такого курса, предоставляемой администрацией колледжа/университета), и

г) применять надежную процедуру оценки способности каждого потенциального разработчика эффективно преподавать смешанный курс с тем, чтобы экспертные оценки процента ожидаемых успехов потенциальных студентов в изучении рекламируемых смешанных курсов были не ниже определенных желательных их значений (которые известны для стандартных онлайн курсов по тем же предметам).

Наконец, смешанные курсы, включенные в программу колледжа/университета должны быть разработаны и реализованы так, чтобы студенты, добившиеся наибольших успехов в их изучении, имели бы реальные шансы на подготовку к продолжению образования в тех самых всемирно известных университетах, (видео) фрагменты из лекций профессоров которых являлись частью этих смешанных курсов. Например, к моменту окончания ими бакалавриата, указанные студенты могли бы успешно подготовиться к поступлению в эти университеты на магистерские программы, тогда как после окончания магистратуры, они имели бы реальные шансы на поступление на обучение там в докторантуре для получения степени доктора философии. Ясно, что реальность таких шансов была бы существенным вкладом в престиж колледжа/университета и резко противопоставила бы описанную стратегию разработки и проведения смешанных курсов традиционным методикам разработки и проведения онлайн курсов. Колледж/университет может даже организовать группы из наиболее продвинутых студентов, заинтересованных в том, чтобы сделать дополнительные шаги в направлении подготовки к обучению в указанных ведущих университетах мира, и помочь им добиться успеха в осуществлении своего желания. Это может быть сделано, в частности, путем предложения таким студентам специально разработанных смешанных курсов с более интенсивным использованием фрагментов онлайн (видео) лекций, прочитанных выдающимися профессорами из этих университетов.

2. В то время как разработка смешанных курсов с использованием фрагментов (видео) лекций профессоров ведущих университетов мира представляется перспективным вариантом разработки для продвинутых обучающихся и потенциальных новых студентов колледжей/университетов, фрагменты (видео) лекций обычных преподавателей (включая

преподавателей, преподающих в этих учебных заведениях) могут использоваться при подготовке смешанных курсов для менее подготовленных потенциальных и обучающихся в этих колледжах/университетах студентов. В этих случаях задачи (2.23)–(2.24) и (2.23), (2.24), (2.17) (с изменениями, отражающими особенности стратегии разработки и проведения смешанных курсов, могут быть использованы для получения соответствующих количественных финансовых оценок.

3. Постановки задач 1 и 2 позволяют предположить, что предлагаемый инструмент поддержки принятия решений может помочь администрации колледжа/университета обосновать потребности бюджета колледжа/университета в переговорах с юридическими лицами, которые заботятся о качестве образования в нем и о его престиже (в частности, в ее переговорах с местными и федеральными администрациями и с частными спонсорами). Например, этот инструмент может помочь обосновать размер  $B - B_0$  той части бюджета колледжа/университета, которую этот колледж/университет планирует потратить на организацию и проведение смешанных  $K$  курсов в переговорах с (указанными выше) юридическими лицами о его конкурентоспособности на рынке новых потенциальных студентов, в частности, обосновать минимально необходимый размер такой части.

4. В настоящем отчете конкурентоспособность колледжа/университета на рынке потенциальных студентов рассматривается с точки зрения только новых потенциальных студентов. Однако, очевидно, что, конкурентоспособность колледжа/университета также зависит от того, насколько хорошо учебный процесс организован для студентов, уже обучающихся в этом колледже/университете хотя бы какое-то время. Соответствующие мероприятия по организации учебного процесса должны мотивировать таких студентов завершить свое образование там, где они в настоящее время обучаются, и не вызывать у них желание перевестись в другой колледж/университет и получить соответствующую степень там. Что касается смешанных курсов, которые могут быть как основными, так и факультативами, те же особенности организации учебного процесса, которые могут привлечь внимание потенциальных новых студентов, должны присутствовать и в смешанных курсах, предлагаемых уже зачисленным студентам колледжа/университета. Такое положение дел естественно, так как качество преподавания этих курсов, влияющее на намерения новых потенциальных студентов поступить учиться в колледж/университет, отчасти мотивированы информацией об этом качестве, полученной от обучающихся там студентов; однако, его реализация представляет собой проблему, требующую решения.

5. Предлагаемая схема использования фрагментов записанных (видео) курсов — это лишь один из вариантов использования таких записей. Так как большинство этих онлайн

курсов доступны бесплатно, администрация колледжа/университета может решить использовать онлайн курс как замену офлайн курса, без проектирования и запуска нового (смешанного) курса. Однако, имеющийся у автора педагогический опыт свидетельствует о том, что записанные онлайн лекции вряд ли могут заменить присутствие лектора в аудитории для большинства студентов, изучающих эти курсы, в то время как замена такого рода не способствует увеличению процента студентов, которые будут изучать соответствующие смешанные курсы успешно. В любом случае, представляется разумным «проверить» оба подхода к использованию онлайн (видео) курсов, прежде чем принять окончательное решение по этому вопросу.

6. Организация подготовительных программ для отдельных групп студентов колледжей/ университетов для успешного изучения ими новых смешанных курсов всегда является испытанием для их администраций. Основная проблема здесь связана со своевременной подготовкой студентов: а) к пониманию фактов, которые будут обсуждаться в конкретном смешанном курсе; б) к тому, чтобы чувствовать себя комфортно со стилем подачи материала разработчиком, выбранным для проведения этого смешанного курса. Время, требуемое на такую подготовку, во многом зависит от того, когда отбор разработчиков для проведения конкретного смешанного курса будет завершен, а результаты проведения подготовительных программ существенно зависят от того, кто проводит эти занятия для каждого нового курса из набора  $K$  смешанных курсов. Анализ таких результатов может стать сложной проблемой, если разработчик, выбранный для проведения конкретного смешанного курса, в настоящее время не преподает в колледже/университете.

Также необходимо подчеркнуть, что, как указано выше во Введении, разработчик, выбранный для преподавания конкретного смешанного курса в колледже/университете, может разрабатывать этот курс самостоятельно, при условии, что он/она придерживается структуры курса, определенной администрацией колледжа/университета. Однако, выбор подходов к разработке конкретных смешанных курсов в соответствии с определенной (администрацией колледжа/университета) структурой этих курсов является проблемой, не обсуждаемой в рамках тех исследований, результаты которых отражены в [110] и настоящем отчете.

Наконец, в модели (2.23), (2.24) предполагалось, что учебные пособия и курсы английского языка организуются и проводятся только для тех смешанных курсов, которые будут преподавать разработчики, приглашаемые из-за границы, что не всегда может иметь место. Разработчики каждого из двух других типов разработчиков курсов (из колледжа/ университета и из других университетов страны) могут потребовать организовать подготовительные программы (а также углубленные курсы английского языка) для смешанных

курсов (которые они могут быть приглашены преподавать), и эти программы и курсы должны будут предлагаться соответствующими кафедрами колледжа/ университета.

Пусть  $\tilde{M} \subset M$  и  $\tilde{R} \subset R$  – множества потенциальных разработчиков (из колледжа/ университета и разработчиков из других университетов), которые считают необходимым провести подготовительные программы колледжем/университетом, если они будут приглашены Университетом для преподавания смешанных курсов. Пусть далее  $\tilde{f}_{ik}^U$  и  $\tilde{f}_{rk}^{OU}$  – расходы на проведение указанных подготовительных программ,  $\tilde{\psi}_{ik}^U$  и  $\tilde{\psi}_{rk}^{OU}$  – расходы на проведение углубленных курсов английского языка, запрошенных этими разработчиками из колледжа/университета и из других университетов страны (т. е., соответственно,  $k \in \overline{1, K}$ ,  $i \in \overline{1, \tilde{M}}$ ,  $r \in \overline{1, \tilde{R}}$ ). Тогда, например, неравенство (2.12) примет вид

$$\begin{aligned} & \sum_{l=1}^L w_l(a_l + h_l) + \sum_{k=1}^K \sum_{l=1}^L \Delta_{lk} z_{lk} + \sum_{r=1}^R s_r(b_r + g_r) + \sum_{k=1}^K \sum_{r=1}^R \delta_{rk} y_{rk} + \\ & \sum_{k=1}^K \sum_{i=1}^M \nabla_{ik} x_{ik} + \sum_{k=1}^K \sum_{i=1}^M d_i t_{ik} x_{ik} + \sum_{k=1}^K f_k u_k + \sum_{i \in \tilde{M}} \sum_{k=1}^K \tilde{f}_{ik}^U x_{ik} + \\ & \sum_{r \in \tilde{R}} \sum_{k=1}^K \tilde{f}_{rk}^{OU} y_{rk} + \sum_{i \in \tilde{M}} \sum_{k=1}^K \tilde{\psi}_{ik}^U x_{rk} + \sum_{r \in \tilde{R}} \sum_{k=1}^K \tilde{\psi}_{rk}^{OU} y_{rk} \leq B - B_0 \end{aligned} \quad (2.28)$$

Здесь предполагается, что если разработчик  $i^*$  из множества  $\tilde{M}$  не требует организации и проведения подготовительных программ для проведения смешанного курса  $k^* \in \overline{1, K}$ , то коэффициент  $\tilde{f}_{i^* k^*}^U$  равен нулю. То же предположение имеет место для коэффициента  $\tilde{f}_{r^{**} k^{**}}^{OU}$  (если разработчик  $r^{**}$  из набора  $\tilde{R}$  не требует организации и проведения подготовительных программ для смешанного курса  $k^{**} \in \overline{1, K}$ ) и для коэффициентов  $\tilde{\psi}_{i^{**} k^{**}}^U$  и  $\tilde{\psi}_{r^{**} k^{**}}^{OU}$ .

7. Как правило, администрация колледжа/университета может осуществлять два подхода к использованию фрагментов онлайн (видео) лекций в смешанных курсах, отличающихся бюджетом на проведение этих курсов. В рамках первого подхода, она может предлагать один и тот же набор смешанных курсов ежегодно в течение, скажем,  $T$  лет, чтобы не было новых смешанных курсов (которые бы добавлялись в течение указанных  $T$  лет; см. предположение 5 из раздела 3). Второй подход предполагает расширение спектра этих смешанных курсов ежегодно (или даже чаще) в зависимости от результатов проведения

смешанных курсов конце конкретного года (или даже каждого года), включая финансовые результаты, связанные с их проведением.

Можно легко убедиться в том, что предложенную в [110] и выше в настоящем отчете математическую модель можно использовать для формализации соответствующих задач в рамках второго подхода, хотя структура системы ограничений в соответствующих задачах будут разными. Именно, в рамках второго подхода указанная система будет иметь блочно-диагональную структуру ограничений, относящиеся к каждому конкретному году (в течение  $T$  лет) внутри отдельного блока, вместе с отдельным блоком, связывающим все переменные системы ограничений [148].

8. По мнению автора, предложенный (в [110] и выше в настоящем отчете) инструментарий демонстрирует потенциал системного анализа, (адекватного) математического моделирования, исследования операций, методов оптимизации и программного обеспечения для решения практических проблем, возникающих в экономике образования.

Еще одно из дальнейших направлений исследования проблем, рассмотренных в рамках проведенного исследования, может касаться развития подходов к решению задач 1 и 2 в условиях неопределенности. Эта неопределенность связана с невозможностью определить точные значения экспертиных оценок параметров  $\alpha_{ik}, \beta_{rk}, \gamma_{lk}$  — процентов студентов, ожидаемых преуспеть в изучении каждого из  $K$  смешанных курсов при преподавании его каждым из преподавателей, рассматриваемых администрацией колледжа/университета в качестве потенциального разработчика этого курса в принципе. В простейшем, однако, практически важном случае, следует считать, что каждое из значений  $\alpha_{ik}, \beta_{rk}, \gamma_{lk}$  меняется в некоторых известных пределах, так что каждый из трех векторов, образуемых числами  $\alpha_{ik}, \beta_{rk}, \gamma_{lk}$ , можно считать принадлежащими некоторому выпуклому многограннику в соответствующем конечномерном пространстве. Аналогичные предположения имеют место и для ряда других параметров, присутствующих в системе ограничений (2.23), (2.24).

Хотя предлагаемый инструментарий позволяет администрации колледжа/университета проводить многократные расчеты для любых наборов экспертов, оценки которых ей может быть интересно изучить, следует подчеркнуть, что информация о значениях этих оценок имеет вероятностный характер. Более того, установление каких-либо конкретных закономерностей и параметров законов распределения, описывающих динамику изменения этих параметров, представляет значительные трудности. В то же время теоретико-игровые подходы к рассмотрению подобной информации (о значениях параметров в соответствующих математических моделях), позволяют формулировать задачи в условиях неопределен-

ности, аналогичные рассмотренным в настоящем отчете, решение которых может проводиться на основе использования комплексов стандартных программ даже на персональных компьютерах [149]. Однако применимость таких подходов к рассматриваемым задачам зависит от того, справедливы ли для них утверждения, подобные утверждениям, сформулированным в [149].

## **2.3 Оптимизация гарантированного риска с учетом дополнительных связей между неопределенными факторами**

### **2.3.1 Введение**

По отношению к управлению рисками и доходами лица, принимающие решение (ЛПР) в задачах оптимизации при неопределенности, можно условно разделить на три категории. «Рискофобы» стремятся исключить какой-либо риск и предпочитают максимизировать гарантированный доход. «Рискофилы» в своих решениях учитывают только риск и стремятся его минимизировать. Поскольку понятие риска неоднозначно, заметим, что в данной работе используется концепция риска по Сэвиджу как потери от незнания неопределенных факторов. «Нейтральные» ЛПР стремятся учитывать оба показателя - как доход, так и риск, что приводит к задаче многокритериального принятия решений при неопределенности. Возможны различные подходы к ее решению. Далее будем рассматривать однокритериальную задачу определения структуры трехвалютного вклада, оптимальной по гарантированному риску (по Сэвиджу). Неопределенными факторами в ней являются будущие значения обменных курсов валют - предполагаются известными только лишь границы их возможных значений. Это приводит к возникновению риска, понимаемого как отклонение получаемого результата от желательного или ожидаемого результата. Применением к функции риска принципа наилучшего гарантированного результата [150] вычисляется наименьший гарантированный риск и соответствующая оптимальная структура. Именно такой подход используется ниже в задаче оптимальной диверсификации депозита по трем видам валют – риск понимается в соответствии с [151] (риск по Сэвиджу или сожаление). Используется точка зрения ЛПР, допускающего риск и стремящегося минимизировать его.

Случаи ЛПР, не допускающих риска и нейтральных [152] изучались ранее в [153]. В частности, в явном виде было получено решение для наилучшего гарантированного дохода в рассматриваемой задаче с  $n$  типами валют. Нейтральное ЛПР принимает во внимание как доход, так и риски, стремясь и повысить доход, и уменьшить риск. Соответствующая двухкритериальная задача в условиях неопределенности изучалась в [153].

Концепция оптимальности по Сэвиджу [151] и другие подходы к задачам с неопределенными параметрами рассматривались в публикациях [154, 155, 156]. В работе [153] исследовалось решение минимаксного сожаления Сэвиджа в задаче диверсификации депозита по двум видам валют. Статья [157] посвящена поиску решения с оптимальным гарантированным риском в аналогичной задаче с тремя типами валют. В ней получен явный вид оптимального решения и предложен конструктивный алгоритм его вычисления, реализованный в Excel. Проведены отладочные и экспериментальные расчеты. Случай равной гарантированной доходности всех трех валют (при минимальных будущих курсах) исследован полностью аналитически. Проведено численное моделирование зависимости минимального гарантированного риска и структуры оптимального депозита от параметров задачи, таких как верхние границы будущих значений курсов валют и процентные ставки.

В данном исследовании проведен анализ зависимости оптимального гарантированного риска и структуры вклада от параметров задачи с учетом дополнительных связей между границами неопределенности. Полученные результаты могут быть использованы при анализе проблем управления финансами в условиях неполной информации.

### 2.3.2 Принцип минимаксного сожаления Сэвиджа

Пусть  $f(x, y)$  – доход,  $X$  – множество стратегий ЛПР,  $Y$  – множество неопределенностей. Тогда  $\max_{z \in X} f(z, y)$  будет наилучшим возможным доходом, если имеет место неопределенность  $y$ . Однако ЛПР не знает заранее, каким будет значение неопределенного параметра  $y$ . Разница

$$\Phi(x, y) = \max_{z \in X} f(z, y) - f(x, y) \quad (2.29)$$

называется функцией риска (сожаления) по Сэвиджу. Она представляет собой «потерю из-за незнания» – разницу между наилучшим результатом, который мог бы быть получен при известной неопределенности  $y$ , и реальным результатом с некоторой стратегией  $x$ . Этот риск зависит как от стратегии  $x$ , так и от неопределенности  $y$ .

Для минимизации риска ЛПР может использовать концепцию наилучшего гарантированного результата (принцип Вальда) [158].

Эти соображения приводят к следующему определению (принцип минимаксного сожаления Сэвиджа [151]).

**Определение.** Решение  $x^*$  называется оптимальным по гарантированному риску решением (ОГРР), если

$$\Phi^r = \max_{y \in Y} \Phi(x^r, y) = \min_{x \in X} \max_{y \in Y} \Phi(x, y), \quad (2.30)$$

где функция риска (потерь)  $\Phi(x, y)$  определена в (2.29). Для краткости будем такое решение называть также оптимальным (по Сэвиджу), а значение  $\Phi^r$  – оптимальным (минимальным) гарантированным риском.

В соответствии с Определением (2.18) построение ОГРР сводится к следующему:

1) Вычислить  $f[y] = \max_{z \in X} f(z, y)$  для каждой неопределенности  $y \in Y$ .

2) Вычислить функцию риска  $\Phi(x, y) = f[y] - f(x, y)$ .

3) Для всех  $x \in X$  вычислить гарантированный риск

$$\max_{y \in Y} \Phi(x, y) = \max_{y \in Y} (f[y] - f(x, y)) = \Phi[x] \geq \Phi(x, y) .$$

4) Вычислить (минимальный) гарантированный риск  $\Phi^r = \min_{x \in X} \Phi[x] = \Phi[x^r]$ .

Точка минимума  $x^r$  и будет оптимальным по гарантированному риску решением, а значение  $\Phi^r = \Phi[x^r]$  – минимальным гарантированным риском.

### 2.3.3 Постановка задачи о выборе структуры вклада

Депозитор (ЛПР) распределяет один рубль по трем видам вкладов, покупая доллары в объеме  $x_d$  рублей и евро в объеме  $x_e$  рублей по начальным курсам  $K_d$  и  $K_e$  соответственно, при этом  $1 - x_d - x_e$  – остаток на рублевом вкладе. В конце периода вклада наращенные по известным процентным ставкам  $d_d, d_e$  суммы долларов и евро конвертируются в рубли по неизвестным на начало периода курсам  $y_d \in [a_d, b_d], y_e \in [a_e, b_e]$ , где  $a_d$  и  $b_d$  – нижняя и верхняя границы возможных изменений финального курса доллара,  $a_e, b_e$  – то же для евро. В результате итоговый доход в рублевом выражении можно представить в следующем виде:

$$f(x, y) = (1 + r)(1 - x_d - x_e) + x_d \frac{1+d_d}{K_d} y_d + x_e \frac{1+d_e}{K_e} y_e. \quad (2.31)$$

Депозитор (ЛПР) стремится максимизировать итоговый показатель (2.31) путем выбора допустимой стратегии  $x = (x_d, x_e) \in X = \{x_d + x_e \leq 1, x_i \geq 0 (i = d, e)\}$ , причем он должен учитывать возможность реализации любых значений неопределенности  $y = (y_d, y_e) \in Y = [a_d, b_d] \times [a_e, b_e]$ . Математическая модель задачи диверсификации связана с антагонистической игрой  $\Gamma = \langle X, Y, f(x, y) \rangle$ , где  $f(x, y)$  – функция дохода ЛПР (2.31), множество стратегий ЛПР

$$X = \{x = (x_d, x_e) | x_d + x_e \leq 1, x_i \geq 0 (i = d, e)\}, \quad (2.32)$$

множество неопределенностей  $y$  – прямоугольник

$$Y = \{y = (y_d, y_e) | y_i \in [a_i, b_i] (i = d, e)\}. \quad (2.33)$$

### 2.3.4 Явный вид функции гарантированного риска и возможные оптимальные решения

Приведенное выше определение оптимальности по Сэвиджу справедливо для любых  $f, X$  и  $Y$ . Для задачи (2.19)–(2.21) об оптимальной структуре вклада план диверсификации  $(1 - x_d^r - x_e^r, x_d^r, x_e^r)$  будет оптимальным (стратегия  $x^r = (x_d^r, x_e^r)$  – гарантированным по риску решением), соответствующее  $\Phi^r$  – минимальным гарантированным риском.

В статье [157] был получен явный вид функции риска для задачи диверсификации трехвалютного депозита:

$$\begin{aligned} \Phi(x, y) &= f[y] - f(x, y) = \\ &= \begin{cases} \Phi_1(x, y) = (1 + r) - f(x, y) = \\ [(1 + r) - \xi_d]x_d + [(1 + r) - \xi_e]x_e, \quad y \in Y_1, \\ \Phi_2(x, y) = \xi_d y_d - f(x, y) = \\ (1 - x_d)[\xi_d y_d - (1 + r)] + [1 + r - \xi_e]x_e, \quad y \in Y_2, \\ \Phi_3(x, y) = \xi_e y_e - f(x, y) = \\ [1 + r - \xi_d]x_d + [\xi_e y_e - (1 + r)](1 - x_e), \quad y \in Y_3, \end{cases} \end{aligned} \quad (2.34)$$

где для краткости были использованы обозначения  $\xi_d = \frac{1+d_d}{K_d}$ ,  $\xi_e = \frac{1+d_e}{K_e}$ .

Возвращаясь к исходным параметрам задачи, получим следующий вид функции риска, более удобный для экономических интерпретаций получаемых результатов:

$$\Phi(x, y) = f[y] - f(x, y) =$$

$$\begin{aligned}
& \Phi_1(x, y) = (1 + r) - f(x, y) = \\
& = \left[ 1 + r - \frac{1+d_d}{K_d} y_d \right] x_d + \left[ 1 + r - \frac{1+d_e}{K_e} y_e \right] x_e, \quad y \in Y_1, \\
& \Phi_2(x, y) = \frac{1+d_d}{K_d} y_d - f(x, y) = \\
= & \left( 1 - x_d \right) \left[ \frac{1+d_d}{K_d} y_d - (1 + r) \right] + \left[ 1 + r - \frac{1+d_e}{K_e} y_e \right] x_e, \quad y \in Y_2, \\
& \Phi_3(x, y) = \frac{1+d_e}{K_e} y_e - f(x, y) = \\
& = \left[ 1 + r - \frac{1+d_d}{K_d} y_d \right] x_d + \left[ \frac{1+d_e}{K_e} y_e - (1 + r) \right] (1 - x_e), \quad y \in Y_3.
\end{aligned} \tag{2.35}$$

Множества  $Y_1, Y_2, Y_3$  являются областями линейности по  $y$  функции риска, то есть функция  $\Phi(x, y)$  совпадает с линейной (по  $x$ ) функцией  $\Phi_i(x, y)$  при любом фиксированном  $y \in Y_i$ . В этом случае говорят, что функция риска  $\Phi(x, y)$  определяется в точке  $(x, y)$  функцией  $\Phi_i(x, y)$ . Функции  $\Phi_i(x, y)$  ( $i = 1, 2, 3$ ) являются билinearными функциями переменных  $x$  и  $y$ . Поэтому функция риска  $\Phi(x, y)$  при любой фиксированной стратегии  $x$  является кусочно-линейной функцией неопределенности  $y$ .

Там же для каждой стратегии  $x \in X$  вычислена функция гарантированного риска  $\Phi[x]$ :

$$\Phi[x] = \max\{\alpha_d x_d + \alpha_e x_e, \beta_d (1 - x_d) + \alpha_e x_e, \alpha_d x_d + \beta_e (1 - x_e)\}, \quad x \in X, \tag{2.36}$$

где использованы следующие обозначения:

$$\alpha_d = \left[ (1 + r) - \frac{1+d_d}{K_d} a_d \right], \quad \alpha_e = \left[ (1 + r) - \frac{1+d_e}{K_e} a_e \right], \tag{2.37}$$

$$\beta_d = \left[ \frac{1+d_d}{K_d} b_d - (1 + r) \right], \quad \beta_e = \left[ \frac{1+d_e}{K_e} b_e - (1 + r) \right]. \tag{2.38}$$

Экономический смысл параметров  $\alpha_d, \alpha_e, \beta_d, \beta_e$  изложен ниже.

Наилучший (наименьший) гарантированный риск  $\Phi^r = \min_{x \in X} \Phi[x] = \Phi[x^r]$  может быть вычислен анализом функции гарантированного риска  $\Phi[x]$  на границе и на внутренности множества стратегий  $X$  (треугольника АOB) с последующим выбором итогового решения:

$$\min_{x \in X} \Phi[x] = \min\{\min_{x \in [0, A]} \Phi[x], \min_{x \in [0, B]} \Phi[x], \min_{x \in [A, B]} \Phi[x], \min_{x \in \text{int}X} \Phi[x]\}. \tag{2.39}$$

Полученные таким образом в [157] результаты можно переформулировать в экономических терминах в следующей форме.

Утверждение 2.1. Если оптимальный гарантированный риск достигается комбинацией долларового и рублевого компонента, то оптимальным является решение вида

$$x^{OA} = (x_d^{OA}, 0), \quad 0 < x_d^{OA} < 1, \quad (2.40)$$

$$\text{где } x_d^{OA} = \begin{cases} \frac{\beta_d - \beta_e}{\alpha_d + \beta_d}, & \text{если } \beta_e \geq 0, \\ \frac{\beta_d}{\alpha_d + \beta_d}, & \text{если } \beta_e \leq 0. \end{cases} \quad (2.41)$$

Две последние формулы можно объединить в одно выражение:

$$x_d^{OA} = \frac{\beta_d - \beta_e (\operatorname{sign} \beta_e + 1)/2}{\alpha_d + \beta_d}. \quad (2.42)$$

Утверждение 2.2. Если оптимальный гарантированный риск достигается комбинацией евро и рублей, то оптимальным является решение вида

$$x^{OB} = (0, x_e^{OB}), \quad 0 < x_e^{OB} < 1, \quad (2.43)$$

$$\text{где } x_e^{OB} = \begin{cases} \frac{\beta_e - \beta_d}{\alpha_e + \beta_e}, & \text{если } \beta_d \geq 0, \\ \frac{\beta_e}{\alpha_e + \beta_e}, & \text{если } \beta_d \leq 0. \end{cases} \quad (2.44)$$

Здесь также обе формулы могут быть объединены одним выражением:

$$x_e^{OB} = \frac{\beta_e - \beta_d (\operatorname{sign}(\beta_d) + 1)/2}{\alpha_e + \beta_e}. \quad (2.45)$$

Утверждение 2.3. Если оптимальный гарантированный риск достигается комбинацией долларов и евро, то оптимальным является одно из следующих трех решений

$$\{x^{12} = (x_e^{12}, 1 - x_e^{12}), x^{13} = (x_e^{13}, 1 - x_e^{13}), x^{23} = (x_e^{23}, 1 - x_e^{23})\} \cap \operatorname{int} X, \quad (2.46)$$

$$\text{где } x_e^{12} = \frac{\alpha_d}{\alpha_d + \beta_d}, \quad x_e^{13} = \frac{\beta_e}{\alpha_e + \beta_e}, \quad x_e^{23} = \frac{\alpha_d + \beta_e}{\alpha_e + \alpha_d + \beta_e + \beta_d}. \quad (2.47)$$

Остается рассмотреть, случай, когда функция гарантированного риска  $\Phi[x]$  достигает минимального значения  $\Phi^*$  на множестве  $X$  в некоторой внутренней точке  $X$  и не достигает его на границе  $frX$ .

**Утверждение 2.4.** Если оптимальный гарантированный риск достигается только при использовании всех трех валют, то оптимальным является решение

$$x^* = (x_d^*, x_e^*) = \left( \frac{\beta_d}{\alpha_d + \beta_d}, \frac{\beta_e}{\alpha_e + \beta_e} \right) \quad (2.48)$$

при условии, что  $0 < x_d, x_e < 1$  и  $x_d + x_e < 1$ .

Любая точка локального минимума выпуклой функции на выпуклом множестве одновременно является точкой глобального (на этом множестве) минимума. Следовательно, если точка внутреннего минимума (2.44) существует и  $x^* = (x_d^*, x_e^*) \in intX$ , это и будет ОГРР. Существует она или нет, зависит от взаимосвязи параметров рассматриваемой задачи. Очень часто ОГРР оказывается на границе множества допустимых решений. Это означает, что оптимальный план диверсификации содержит только две (или даже одну) из трех валют.

К списку возможных кандидатов на роль оптимального решения, указанных в утверждениях 2.1–2.4, необходимо добавить также одновалютные варианты:  $(0,0)$  – только рубли,  $(1,0)$  – только доллары,  $(0,1)$  – только евро. В результате получим исчерпывающее множество из девяти решений, хотя бы одно из которых будет оптимальным.

### 2.3.5 Алгоритм построения гарантированного по риску решения

Объединение утверждений 2.1–2.4 приводит к следующей схеме [157] вычисления гарантированной по риску стратегии диверсификации вклада и наименьшего гарантированного риска.

- 1) Задать значения процентных ставок  $r, d_d, d_e$  и начальных обменных курсов  $K_d, K_e$ , установить значения верхних и нижних границ возможных изменений курсов доллара  $a_d, b_d$  и евро  $a_e, b_e$ .
- 2) Вычислить вторичные параметры  $\alpha_d, \alpha_e, \beta_d, \beta_e$  по формулам (2.37), (2.38).
- 3) Вычислить девять точек-кандидатов на оптимальное решение. Соответствующие формулы для этих точек представлены в приведенной ниже Таблице 2.3.
- 4) Удалить точки, не принадлежащие множеству допустимых решений  $X$ .
- 5) Вычислить значения гарантированного риска в оставшихся точках, выбрать решение с наилучшим (минимальным) гарантированным риском.

Таблица 2.3 – Кандидаты в решения и проверка условий их допустимости

N	Кандидаты в решения	Формулы	Проверка условий
1	План O=(0,0) (рубли)	(0,0)	Не требуется
2	План A=(1,0) (доллары)	(1,0)	Не требуется
3	План B=(0,1) (евро)	(0,1)	Не требуется
4	План $(x_d^{OA}, 0)$ (рубли и доллары)	$x_d^{OA} = \frac{\beta_d - \beta_e(\text{sign}\beta_e + 1)/2}{\alpha_d + \beta_d}$	$0 < x_d^{OA} < 1$
5	План $(0, x_e^{OB})$ (рубли и евро)	$x_e^{OB} = \frac{\beta_e - \beta_d(\text{sign}\beta_d + 1)/2}{\alpha_d + \beta_d}$	$0 < x_e^{OB} < 1$
6	План $(1 - x_e^{12}, x_e^{12})$ (доллары и евро)	$x_e^{12} = \frac{\alpha_d}{\alpha_d + \beta_d}$	$0 < x_e^{12} < 1$
7	План $(1 - x_e^{13}, x_e^{13})$ (доллары и евро)	$x_e^{13} = \frac{\beta_e}{\alpha_e + \beta_e}$	$0 < x_e^{13} < 1$
8	План $(1 - x_e^{23}, x_e^{23})$ (доллары и евро)	$x_e^{23} = \frac{\alpha_d + \beta_e}{\alpha_e + \alpha_d + \beta_e + \beta_d}$	$0 < x_e^{23} < 1$
9	Внутренняя точка в X (рубли, доллары и евро)	$(x_d^*, x_e^*) = \left( \frac{\beta_d}{\alpha_d + \beta_d}, \frac{\beta_e}{\alpha_e + \beta_e} \right)$	$0 < x_d, x_e < 1$ $x_d + x_e < 1$

### 2.3.6 Экономическая и игровая интерпретация

Напомним, что  $r, d_d, d_e$  – процентные ставки по соответствующим видам вкладов (рубль, доллар, евро);  $K_d$  и  $K_e$  – курсы доллара и евро по отношению к рублю, по которым закупается валюта в начале года,  $a_d, a_e$  – нижние, а  $b_d, b_e$  – верхние пределы значений обменных курсов доллара  $y_d$  и евро  $y_e$  в конце периода депозита. Эти параметры полагаются известными ЛПР в начале периода вклада. Однако значения неопределенных параметров – финальных значений курсов доллара и евро известны только с точностью до множеств:  $y_d \in [a_d, b_d]$ ,  $y_e \in [a_e, b_e]$ . Рублевый доход при распределении рубля  $(1 - x_d - x_e, x_d, x_e)$  определяется линейной по  $x = (x_d, x_e)$  функцией (2.19) с неопределенными коэффициентами.

Входящие в функцию гарантированного риска (2.24) производные коэффициенты  $\alpha_d, \alpha_e, \beta_d, \beta_e$  зависят от вышеперечисленных исходных параметров и имеют следующую содержательную интерпретацию. Выражение  $\alpha_d = \left[ (1 + r) - \frac{1+d_d}{K_d} a_d \right]$  задает разницу между доходностью  $(1 + r)$  одного рубля в рублевом вкладе и доходностью при использовании его через покупку доллара в начале периода по курсу  $K_d$  руб./долл. и продажу итоговой суммы  $(1 + d_d)$  по минимальному курсу  $a_d$  в конце периода. Аналогичный смысл имеет и параметр  $\alpha_e = \left[ (1 + r) - \frac{1+d_e}{K_e} a_e \right]$ , задающий разность доходностей рубля и евро в случае, если неопределенный курс евро принимает минимальное возможное значение.

Коэффициент  $\beta_d$  ( $\beta_e$ ) задает разность между доходностью одного рубля, вложенного в депозит через доллар (евро), и доходностью рубля в рублевом вкладе при самом высоком финальном курсе доллара (евро).

Задача о наилучшем гарантированном доходе (2.31)–(2.33) может рассматриваться как антагонистическая игра ЛПР с Природой [159]. Она имеет при данной структуре неопределенности довольно простое минимаксное решение в случае любого конечного числа валют [153]. Принцип минимаксного сожаления Сэвиджа связан с рассмотрением более сложной конструкции – целевой функции риска (2.17), к которой затем применяется минимаксный принцип Вальда. Таким образом, задачу о вычислении оптимального риска также можно интерпретировать как антагонистическую игру ЛПР против Природы – неопределенности с функцией риска в качестве функции выигрыша. Однако, в отличие от вышеупомянутой игры «ЛПР – Природа» с функцией дохода (2.31) в качестве функции выигрыша, эта игра не имеет, как правило, седловой точки [160] в чистых стратегиях. Действительно, при сделанных предположениях максимин и минимакс в этой игре существуют, но не равны. Максимин равен нулю – это оптимальное значение функции риска при известной ЛПР наперед неопределенности, то есть фактически без неопределенности. Минимакс же для неотрицательной функции риска неотрицателен и равен нулю только в вырожденных случаях, когда существует стратегия, оптимальная при любой неопределенности. Отсюда следует [161] отсутствие седловой точки.

В данном исследовании гарантирующая стратегия ЛПР была найдена прямым вычислением («по определению») минимакса функции риска.

По изначальной постановке рассматриваемая задача близка к задаче Марковица об оптимальной структуре портфеля инвестиций [162], впервые сформулированной еще в 1952 году и с тех пор получившей развитие в многочисленных публикациях. Несмотря на столь долгую историю вопроса, до сих пор «многие показатели эффективности инвестиций основываются на портфельной теории Марковица» [163]. В этой теории доходности отдельных активов портфеля – случайные величины, относительно которых предполагаются известными математические ожидания и ковариационная матрица [164]. Ожидаемая доходность всего портфеля и ее дисперсия рассматриваются как подлежащие оптимизации два показателя качества – доход и риск. С формальной точки зрения это задача двухкритериальной оптимизации [165] с линейным (доход) и квадратичным (риск как дисперсия доходности) критериями и простейшими линейными ограничениями. Разработаны различные методы нахождения Парето-оптимальных решений, способы построения траекторий эффективных портфелей [166]. Большинство из них использует для построения Паретовской границы приемы ее параметризации, например, оптимизацию по одному критерию при ограничении на значения другого.

Предмет данного исследования близок к теме оптимизации портфеля с тремя активами, два из которых – рисковые. Однако имеются два существенных отличия – в доступной информации и в оценке риска. По ряду причин не всегда можно получить удовлетворительные оценки статистических характеристик доходности активов и приходится довольствоваться интервальными прогнозами. Использование дисперсии как меры риска удобно с вычислительной точки зрения, но имеет свои известные недостатки. В их числе – необходимость в информации о коррелированности активов. Применение альтернативной меры риска – риска по Сэвиджу развивает дополнительные направления разработки портфельной теории и практики.

В некоторых специальных случаях возможно аналитическое получение решения задачи о наилучшем гарантирующем решении. Так, для случая  $\alpha_d = \alpha_e = 0$  был получен следующий результат:

**Утверждение 2.5.** Если  $\alpha_d = \alpha_e = 0$ , т. е. все три валюты при минимальных финальных значениях курсов доллара и евро дают равный доход, то функция гарантированного риска  $\Phi[x]$  достигает минимального на множестве  $X$  значения в некоторой точке из множества

$$\{(1,0), (0,1), x_e^{23} = (\beta_d/(\beta_e + \beta_d), \beta_e/(\beta_e + \beta_d))\}. \quad (2.49)$$

Сравнение значений функции гарантированного риска в трех точках-кандидатах выделяет единственную точку оптимума  $x^r = (\beta_d/(\beta_e + \beta_d), \beta_e/(\beta_e + \beta_d))$ , которая и дает минимальный гарантированный риск  $\Phi^r = \Phi[x^r] = \beta_d\beta_e/(\beta_e + \beta_d)$ .

Таким образом, в случае равной рублевой доходности трех валют (при минимальных значениях финальных курсов) оптимальное по гарантированному риску решение состоит в полном распределении рубля между долларом и евро пропорционально верхним возможным значениям их финальных курсов. Если  $\beta_d = \beta_e = \beta$ , оптимальное решение  $x^r = (0.5; 0.5)$ .

В общем же случае поиск оптимального решения задачи диверсификации требует использования численных алгоритмов. Предложенный метод был реализован в среде Excel 2016 [167] в двух вариантах – пользовательском и исследовательском. Первый рассчитан на пользователя, имеющего минимальные навыки работы в готовой таблице Excel. Работа в нем сводится к вводу упомянутых выше исходных данных и содержательному анализу полученных результатов. Возможности исследовательского варианта частично иллюстрируются (но не ограничиваются) приводимыми далее результатами расчетов.

### 2.3.7 Результаты модельных экспериментальных расчетов

Были проведены экспериментальные вычисления с учетом динамики валютных курсов в 2023 году. Кроме того, выполнены дополнительные экспериментальные расчеты, учитывающие некоторые взаимозависимости параметров задачи.

В Таблице 2.4 приведены результаты расчетов минимального гарантированного риска в зависимости от значений двух параметров - верхних границ курсов  $b_d$ ,  $b_e$ , варьируемых от 100 до 130 для доллара и от 105 до 135 для евро с шагом 1,5 руб. Значения нижних границ курсов предполагаются фиксированными на уровне 95 руб./дол. и 100 руб./евро. Значения использованных при моделировании параметров приведены в Таблице 2.5. При их назначении учитывалась бурная динамика курса рубля осенью 2023 года.

На рисунке 2.7 полученные результаты представлены в графическом виде, где i-я снизу кривая соответствует зависимости оптимального риска от верхней границы курса евро  $b_e$  при фиксированном i-ом значении курса доллара  $b_d$ .

Таблица 2.4 – Минимальный гарантированный риск как функция верхних границ неопределенности обменных курсов. Вторая строка содержит значения курса доллара, левый столбец – евро

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
	100,0	101,5	103,0	104,5	106,0	107,5	109,0	110,5	112,0	113,5	115,0	116,5	118,0	119,5	121,0	122,5	124,0	125,5	127,0	128,5	130,0
105,0	0,000	0,000	0,000	0,000	0,000	0,007	0,019	0,029	0,037	0,044	0,050	0,055	0,058	0,062	0,066	0,068	0,072	0,075	0,077	0,079	0,081
106,5	0,000	0,000	0,000	0,000	0,000	0,007	0,019	0,029	0,037	0,044	0,050	0,055	0,058	0,062	0,066	0,068	0,072	0,075	0,077	0,079	0,081
108,0	0,000	0,000	0,000	0,000	0,000	0,007	0,019	0,029	0,037	0,044	0,050	0,055	0,058	0,062	0,066	0,068	0,072	0,075	0,077	0,079	0,081
109,5	0,000	0,000	0,000	0,000	0,000	0,007	0,019	0,029	0,037	0,044	0,050	0,055	0,058	0,062	0,066	0,068	0,072	0,075	0,077	0,079	0,081
111,0	0,000	0,000	0,000	0,000	0,000	0,007	0,019	0,029	0,037	0,044	0,050	0,055	0,058	0,062	0,066	0,068	0,072	0,075	0,077	0,079	0,081
112,5	0,000	0,000	0,000	0,000	0,000	0,007	0,019	0,029	0,037	0,044	0,050	0,055	0,058	0,062	0,066	0,068	0,072	0,075	0,077	0,079	0,081
114,0	0,012	0,012	0,012	0,012	0,012	0,021	0,032	0,041	0,048	0,055	0,061	0,065	0,068	0,070	0,073	0,077	0,080	0,083	0,085	0,087	0,089
115,5	0,023	0,023	0,023	0,023	0,023	0,034	0,047	0,056	0,064	0,071	0,079	0,087	0,091	0,095	0,098	0,101	0,104	0,107	0,110	0,113	0,116
117,0	0,031	0,031	0,031	0,031	0,033	0,047	0,061	0,074	0,080	0,085	0,091	0,098	0,103	0,108	0,112	0,117	0,121	0,125	0,129	0,133	0,136
118,5	0,039	0,039	0,039	0,039	0,039	0,041	0,046	0,051	0,056	0,063	0,073	0,080	0,087	0,093	0,097	0,102	0,106	0,109	0,113	0,116	0,118
120,0	0,045	0,045	0,045	0,045	0,048	0,054	0,059	0,065	0,071	0,079	0,087	0,094	0,100	0,105	0,110	0,114	0,118	0,122	0,124	0,126	0,128
121,5	0,051	0,051	0,051	0,051	0,054	0,061	0,066	0,073	0,079	0,085	0,094	0,101	0,108	0,113	0,119	0,123	0,128	0,130	0,132	0,135	0,137
123,0	0,056	0,056	0,056	0,056	0,059	0,064	0,073	0,080	0,087	0,093	0,100	0,108	0,115	0,121	0,125	0,128	0,133	0,136	0,139	0,142	0,144
124,5	0,061	0,061	0,061	0,061	0,063	0,071	0,078	0,086	0,093	0,100	0,108	0,115	0,122	0,128	0,133	0,138	0,141	0,144	0,147	0,149	0,151
126,0	0,063	0,063	0,063	0,063	0,067	0,075	0,083	0,091	0,099	0,107	0,115	0,123	0,127	0,130	0,134	0,137	0,140	0,143	0,146	0,149	0,152
127,5	0,067	0,067	0,067	0,067	0,071	0,079	0,087	0,096	0,104	0,112	0,121	0,126	0,130	0,134	0,138	0,141	0,145	0,148	0,151	0,154	0,157
129,0	0,070	0,070	0,070	0,070	0,074	0,083	0,091	0,100	0,108	0,117	0,125	0,129	0,134	0,138	0,142	0,146	0,149	0,153	0,155	0,158	0,161
130,5	0,073	0,073	0,072	0,072	0,077	0,086	0,095	0,104	0,113	0,122	0,127	0,132	0,137	0,141	0,145	0,148	0,153	0,156	0,160	0,163	0,166
131,0	0,075	0,075	0,075	0,075	0,075	0,079	0,086	0,098	0,107	0,116	0,125	0,130	0,135	0,140	0,144	0,149	0,153	0,157	0,160	0,164	0,167
133,5	0,077	0,077	0,077	0,077	0,077	0,081	0,091	0,101	0,110	0,120	0,127	0,133	0,138	0,143	0,147	0,152	0,156	0,160	0,164	0,168	0,171
135,0	0,079	0,079	0,079	0,079	0,079	0,084	0,093	0,103	0,113	0,123	0,129	0,135	0,140	0,145	0,150	0,155	0,160	0,164	0,168	0,172	0,175

Таблица 2.5 – Значения параметров, использованные при вычислении Таблицы 2.4

Рубль, %	Доллар, %	Евро, %	Дол. мин.	Дол. макс.	Евро мин.	Евро макс.	Долл. текущ.	Евро текущ.
10,00%	3,00%	2,50%	95	100 – 130	100	105 – 135	100	105

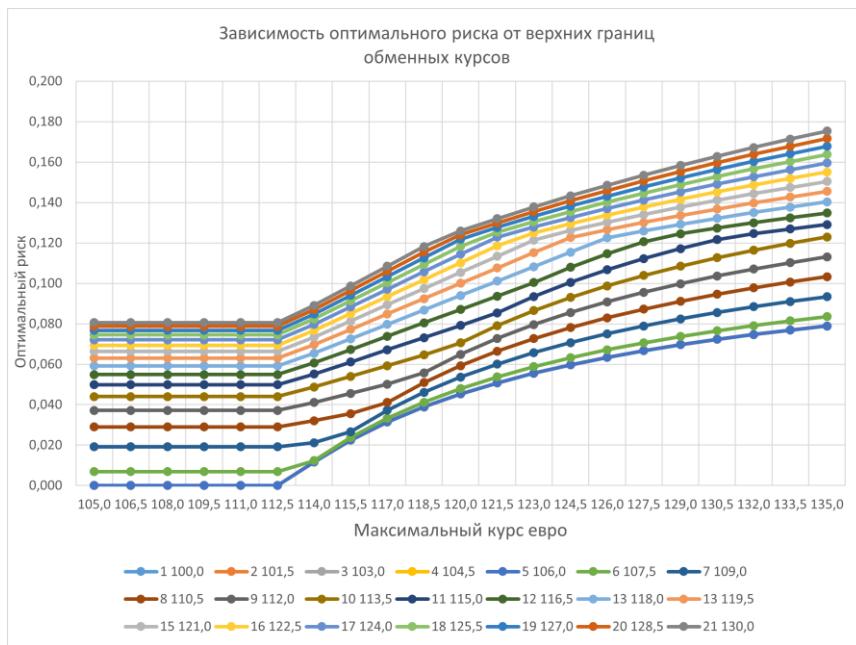


Рисунок 2.7 – Зависимости минимального риска от  $b_e$  при фиксированных значениях  $b_d$  от 100 до 130 руб./долл.

Ввиду тесной связи между курсами доллара и евро представляет особый интерес поведение оптимального риска при синхронном, связанном изменении верхних границ. Рассмотрим варьирование  $b_d, b_e$  при условии  $105 \leq b_d + 5 = b_e \leq 135$  и при тех же значениях постоянных параметров, что и выше (см. Таблицу 2.5). Результат представлен в Таблицу 2.6, в графической форме – на рисунке 2.8.

Таблица 2.6 – Минимальный гарантированный риск при связанном изменении верхних границ неопределенности обменных курсов.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
\$ макс.	100	101,5	103	104,5	106	107,5	109	110,5	112	113,5	115	116,5	118	119,5	121	122,5	124	125,5	127	128,5	130
€ макс.	105	106,5	108	109,5	111	112,5	114	115,5	117	118,5	120	121,5	123	124,5	126	127,5	129	130,5	132	133,5	135
Оптр.риск	0,000	0,000	0,000	0,000	0,000	0,007	0,021	0,036	0,050	0,065	0,079	0,094	0,108	0,123	0,130	0,138	0,145	0,153	0,160	0,168	0,175



Рисунок 2.8 – Минимальный гарантированный риск при связанном изменении верхних границ курсов  $b_d, b_e$  ( $b_e = b_d + 5$ )

В проведении расчетов по сценарию убывания нижних границ курсов при фиксированных значениях верхних границ на начальном уровне или ниже необходимости нет. Действительно, при этом доллар и евро менее доходны, чем рубль, и оптимальным решением с нулевым риском будет чисто рублевый депозит.

В Таблице 2.7 приведены результаты расчетов зависимостей производных параметров – коэффициентов функции гарантированного риска  $\alpha_d, \alpha_e, \beta_d, \beta_e$  от процентной ставки по рублю. Содержательный смысл этих величин был пояснен выше. Значения прочих показателей фиксированы в соответствии с Таблицей 2.8.

Приведенные в верхних строках данные относятся к оптимальной структуре вклада и оптимальному гарантированному риску. Минимальный гарантированный риск снижается с ростом ставки рубля.

Рисунок 2.9 показывает линейную зависимость параметров  $\alpha_d, \alpha_e, \beta_d, \beta_e$ , а попарная параллельность графиков иллюстрирует их взаимосвязанность. Действительно, в силу определений (2.37) и (2.38) коэффициенты  $\alpha_d$  и  $\alpha_e$  отличаются на константу  $\alpha_e - \alpha_d = \frac{1+d_e}{K_d} a_d - \frac{1+d_e}{K_e} a_e$ , не зависящую от процентной ставки  $r$ . Аналогично разность  $\beta_e - \beta_d = \frac{1+d_e}{K_e} b_e - \frac{1+d_d}{K_d} b_d$  не зависит от  $r$ .

Таблица 2.7 – Структура решения и зависимость параметров  $\alpha_d, \alpha_e, \beta_d, \beta_e$  от процентной ставки по рублю

Ставка	6,0%	6,5%	7,0%	7,5%	8,0%	8,5%	9,0%	9,5%	10,0%	10,5%	11,0%	11,5%	12,0%	12,5%	13,0%	13,5%	14,0%	14,5%	15,0%
доллар	0,645	0,645	0,645	0,645	0,645	0,645	0,645	0,645	0,645	0,645	0,325	0,325	0,325	0,325	0,325	0,325	0,325	0,325	0,325
евро	0,355	0,355	0,355	0,355	0,355	0,355	0,355	0,355	0,355	0,355	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
рубль	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,675	0,675	0,675	0,675	0,675	0,675	0,675	0,675	0,675
риск	0,099	0,099	0,099	0,099	0,099	0,099	0,099	0,099	0,099	0,099	0,096	0,093	0,090	0,088	0,083	0,079	0,076	0,073	0,069
ALFA <sub>d</sub>	0,035	0,040	0,045	0,050	0,055	0,060	0,065	0,070	0,075	0,080	0,085	0,090	0,095	0,100	0,105	0,110	0,115	0,120	0,125
BETA <sub>d</sub>	0,193	0,188	0,183	0,178	0,173	0,168	0,163	0,158	0,153	0,148	0,143	0,138	0,133	0,128	0,123	0,118	0,113	0,108	0,103
ALFA <sub>e</sub>	0,086	0,091	0,096	0,101	0,106	0,111	0,116	0,121	0,126	0,131	0,136	0,141	0,146	0,151	0,156	0,161	0,166	0,171	0,176
BETA <sub>e</sub>	0,119	0,114	0,109	0,104	0,099	0,094	0,089	0,084	0,079	0,074	0,069	0,064	0,059	0,054	0,049	0,044	0,039	0,034	0,029

Таблица 2.8 – Значения параметров, использованные при вычислении Таблицы 2.7

Рубль, %	Доллар, %	Евро, %	Дол. мин	Дол. макс	Евро мин.	Евро макс	Долл начальн.	Евро начальн.
10,00%	2,50%	2,50%	90	110	95	115	90	100

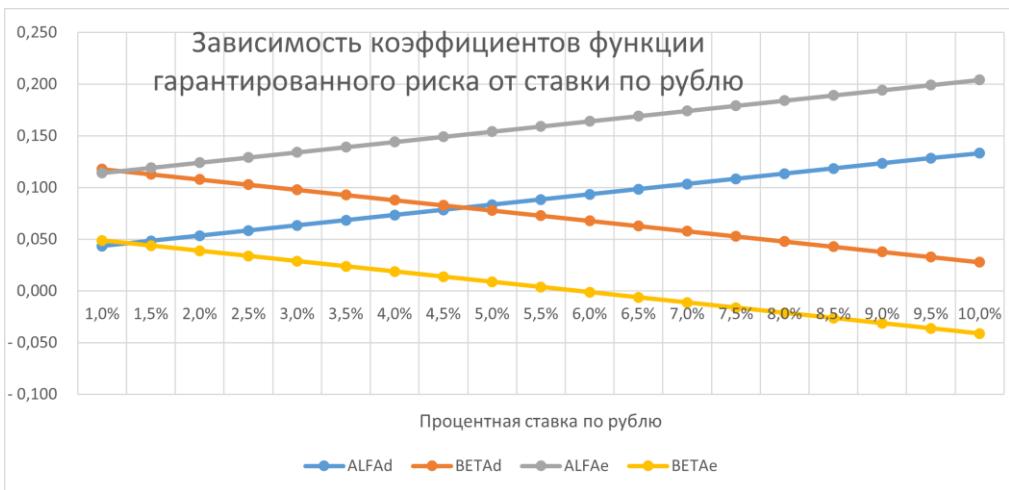


Рисунок 2.9 – Зависимость параметров  $\alpha_d, \alpha_e, \beta_d, \beta_e$  от процентной ставки по рублю

### 2.3.8 Некоторые выводы

Один из основных подходов к принятию решений в условиях существенной неопределенности состоит в минимизации гарантированного риска. Для оценки последнего обычно используется концепция минимаксного сожаления Сэвиджа (риск по Сэвиджу), дающая наилучший гарантированный результат. С этих позиций был проведен анализ задачи об оптимальной по риску структуре трехвалютного вклада (условно в рублях, долларах и евро). В работе [157] были найдены функции риска и гарантированного риска, предложен алгоритм расчета в Excel оптимального решения. В данном исследовании изучалась зависимость оптимального гарантированного риска и структуры вклада от параметров задачи с учетом дополнительных связей между границами неопределенности. При проведении модельных вычислений принималась во внимание динамика валютных курсов в 2023 г.

Задачу о вычислении оптимального риска можно интерпретировать как антагонистическую игру ЛПР против Природы-неопределенности с функцией риска в качестве функции выигрыша. Множеством стратегий противника ЛПР – Природы является множество всех возможных значений неопределенных переменных, в данном случае – будущих обменных курсов валют. Проведено численное моделирование влияния формы и размеров этого множества неопределенностей на решение рассматриваемой задачи. Выполнен численный анализ зависимости оптимального по Сэвиджу решения при различных вариантах изменения границ неопределенности. Результаты позволяют оценить степень влияния масштаба неопределенности на оптимальное решение.

Проведенные исследования нашли отражение в [157, 167]. Они дают представление о возможностях предлагаемого метода и его программной реализации и могут быть использованы при анализе проблем управления финансами в условиях неполной информации.

### **3 Разработка и применение методов сетевого и кластерного анализа данных**

В этом разделе отчета приведены результаты разработки и применения методов сетевого и кластерного анализа данных. Работа над этой тематикой осуществлялась в 2023 году в следующих направлениях:

- сетевые модели научных исследований в области болезни Паркинсона;
- применение метода аномальных кластеров к анализу температурных карт;
- разработка человека-машинной процедуры паттерн-кластеризации данных;
- исследование основных свойств критических групп показателей, выявленных методом анализа паттернов;
- паттерн-анализ экспорта-импорта электроэнергии.

#### **3.1 Сетевые модели цитирований авторов и организаций в области болезни Паркинсона**

##### **3.1.1 Введение**

Болезнь Паркинсона (БП) представляет собой нейродегенеративное заболевание, которое повреждает нейроны в конкретной области мозга, известной как черная субстанция. Главные клинические проявления БП включают в себя дрожь, гипокинезию (снижение и замедление двигательной активности), повышенный мышечный тонус, нарушения в походке и балансе, а также различные немоторные и когнитивные расстройства. Первое описание Болезни Паркинсона было сделано Джеймсом Паркинсоном в 1817 году.

Следует отметить, что Болезнь Паркинсона занимает второе место по распространенности среди нейродегенеративных заболеваний после болезни Альцгеймера. В настоящее время более 10 миллионов людей страдают от этого недуга. По мере старения населения этот показатель продолжает расти, что оказывает непосредственное и косвенное воздействие на затраты государств, например, в США эти расходы составляют 52 миллиарда долларов в год.

Научное сообщество регулярно публикует тысячи исследовательских статей, посвященных этой болезни. Для качественного поиска потенциальных методов лечения БП крайне важно изучить предметную область: анализировать научные статьи, изучать журналы, выявлять авторов и организации, занимающиеся исследованиями в этой области.

Эффективное исследование предметной области можно провести с помощью построения различных сетевых моделей, основанных на цитированиях, и анализа полученных сетей при помощи соответствующих методов. Этот вид анализа позволит выявить ключевых участников, перспективы развития и тенденции в данной области исследований.

Данное исследование продолжает работу прошлого года [1], в котором были построены сети цитирований публикаций и журналов в области исследований болезни Паркинсона, а также проанализированы классические и новые индексы центральности. Целями исследования этого 2023 года были построение сетей цитирований авторов и их аффилиаций, анализ центральности полученных сетей, паттерн-анализ цитируемости и анализ стабильности сетей.

В данной исследовательской работе использовались разнообразные методы сетевого анализа для изучения паттернов цитирования в области исследований по болезни Паркинсона. Мы построили сети цитирования, охватывающие статьи, журналы, авторов и их аффилиации. Сеть цитирования представляет собой взвешенный направленный граф, где вершинами являются различные объекты (статьи, журналы, авторы, организации), а ребра между ними указывают на наличие и количество цитирований между этими объектами.

Для оценки влиятельности каждого из этих объектов мы рассчитали и проанализировали различные индексы центральности, включая 4 классических и 2 новых, учитывавших различные параметры вершин. Эти индексы были рассчитаны как за весь рассматриваемый период, так и по годам исследования. Также, для журналов и аффилиаций авторов мы провели кластеризацию паттернов изменения их цитируемости в течение разных лет.

Для выявления самых значимых вершин в сети, могут быть использованы различные методы, в первую очередь это анализ центральности. Расчет индексов центральности позволяет оценить и ранжировать влиятельность той или иной вершины относительно остальных. Существует большое количество различных индексов, которые могут быть рассчитаны для сети в зависимости от задачи. Подробное описание данных индексов приведено в книге Марка Ньюмана.

### 3.1.2 Описание данных

Из базы данных научных публикаций Microsoft Academic были загружены статьи, опубликованные в период с 2015 по 2021 год, содержащие одновременно слова "parkinson" и "disease" в заголовке или аннотации. Этот временной интервал был выбран для определения основных тенденций в исследованиях за последние годы. Всего было найдено 70119 статей. Анализ данных показал, что 10681 публикация не имеет цифрового идентификатора DOI, 59438 содержат его. Из них 7315 не имеют аннотации, 52132 имеют ее. Около 45940 статей содержат информацию о журнале, в котором они были опубликованы. Для дальнейшего анализа использовались только статьи с DOI, аннотацией и информацией о журнале, чтобы можно было интерпретировать результаты.

### 3.1.3 Построение сети цитирований организаций

Для построения сети цитирования аффилиаций мы использовали аффилиацию первого автора, указанного среди всех авторов, так как это обычно главный автор, который руководит исследовательской группой, и часто именно это заведение финансирует исследование. Однако не все статьи содержали правильную аффилиацию авторов: некоторые не имели названия аффилиации, другие были написаны с ошибками или указаны в неправильной форме (например, только город или страна), по которой невозможно определить аффилиацию. В результате было найдено 3045 уникальных идентификаторов аффилиаций. Из этих статей удалили 5631, которые не имели идентификатора аффилиации. В итоге в сети аффилиаций получилось 3029 вершин и 170340 ребер, учитывая параллельные за разные годы и 128602 с группировкой по годам.

Сеть имеет следующую структуру: ребро между аффилиацией AfId1 и AfId2 означает, что статья с аффилиацией AfId1 ссылается на статью с аффилиацией AfId2. Вес ребра — это количество цитирований за определенный год.

### 3.1.4 Построение сети цитирований авторов

Таким же образом была построена сеть цитирований для авторов. Для каждой статьи было решено выбрать первого автора из списка, так как, как правило, в медицинских статьях первым указывается главный автор, либо начальник лаборатории. Ребро отвечает за цитирование статьей, опубликованной автором с AuId1, статьи автора с AuId2. Вес ребра — это количество ссылок за определённый год. В такой сети возможны самоцитирования. В Таблице 3.1 представлены некоторые показатели сети цитирований для авторов.

Таблица 3.1 – Показатели сети цитирования авторов

Название показателя	Значение
Число вершин	27551
Число ребер	271623
Число компонент связности	82
Размер наибольшей компоненты	27402
Плотность	0,0004
Минимальное число цитирований	1
Максимальное число цитирований	2416
Среднее число цитирований	14,75

В описанной сети насчитывается 82 компоненты связности. Самая большая из них содержит 27402 автора. Также в сети 58 компонент с парами авторов, 5 компонент с тройками авторов и 18 компонент, состоящих только из одного автора, который цитирует только самого себя.

Во всех тройках есть разные авторы из разных организаций, но их работы посвящены схожим специальным темам.

Те же паттерны можно найти в компонентах с размером 2. В основном эти авторы узко специализируются на конкретных вопросах, связанных с БП. Для анализа центральности сети цитирования авторов были выбраны аналогичные значения квот и добавлена квота 0,1% от общего числа цитирований журнала.

### 3.1.5 Результаты анализа сети цитирования организаций

Индекс In-degree отражает количество цитирований или входящих ребер. В Таблице 3.2 представлен топ-10 самых цитируемых аффилиаций по In-degree. National Institutes of Health – это 27 национальных институтов и исследовательских центров здравоохранения США, является основным центром в США, ответственным за исследования в области медицины. UCL Institute of Neurology – институт факультета наук о мозге Университетского колледжа Лондона. Эти заведения являются профильными, поэтому в них занимаются исследованиями нейродегенеративных заболеваний, в том числе и болезнью Паркинсона. Кроме того, в таблице представлены университеты из разных стран, такие как Великобритания, США, Бельгия, Швеция, которые являются крупными и старейшими в мире. Они известны своими качественными исследованиями в различных научных областях и имеют множество нобелевских лауреатов. Каролинский институт в Швеции является одним из крупнейших медицинских университетов в Европе.

Таблица 3.2 – Топ-10 аффилиаций по In-degree

№	Название	In-degree	Ранг по Betweenness	Ранг по Eigenvector	Ранг по Pagerank
1	National Institutes of Health	0,0156	5	1	1
2	UCL Institute of Neurology	0,0134	3	4	2
3	University of Cambridge	0,0131	1	3	4
4	University of Cambridge	0,0122	2	2	3

### Продолжение таблицы 3.2

5	University of Oxford	0,01206	6	5	5
6	University College London	0,0114	14	6	6
7	Harvard University	0,0100	8	8	7
8	University of Pennsylvania	0,0085	24	7	8
9	Katholieke Universiteit Leuven	0,0080	26	10	12
10	Karolinska Institutet	0,0079	10	9	9

#### 3.1.6 Результаты паттерн-анализа цитируемости организаций

Паттерн-анализ проводился для кластеризации изменений количества цитирований и индекса PageRank по годам. Для этого были построены сети цитирований по годам. Были применены методы с использованием фиксированной трубы для абсолютных значений и тангенсов угла наклона, ширина трубы принята равной 0.1. Данные были нормализованы с помощью минимаксного метода по годам. При таких параметрах было получено 28 паттернов, наибольший из которых содержит 2833 организации, а 18 паттернов содержат только одну организацию. В ходе анализа полученных паттернов удалось выявить интересные особенности. Например, найден паттерн, в котором содержатся Кембриджский университет и Университетский колледж Лондона, расположенные в одной стране и имеющие похожую динамику цитирований. Более подробные результаты доступны на странице <https://www.hse.ru/DeCAn/parkinson>.

#### 3.1.7 Результаты анализа устойчивости сети цитирования организаций

Для оценки степени сходства сетей за все рассматриваемые годы были рассчитаны матрицы интервального порядка и влияния для последовательных годов и расстояние между ними (Таблица 3.3). Значение  $\epsilon$  было установлено равным 10, что означает, что изменение ранга аффилиации менее чем на 10 единиц не считается существенным и ранги считаются равными.

Близкое расстояние между матрицами интервальных порядков указывает на незначительные изменения в ранжировании аффилиаций со временем. Это означает, что наиболее важные аффилиации сохраняют свои позиции. Аналогично, значения матриц влияний очень близки к 0, что означает, что организации цитируют друг друга в примерно одинако-

вых пропорциях. В результате метрика стабильности также имеет близкие значения и остается почти неизменной со временем. Следовательно, цитирования аффилиаций остаются постоянными, и нет заметных тенденций, которые могли бы изменить структуру цитирований в будущем. При выборе организаций для инвестирования можно ориентироваться на их популярность и значимость в предыдущие годы.

Таблица 3.3 – Устойчивость сети цитирований аффилиаций

	2015-2016	2016-2017	2017-2018	2018-2019	2019-2020	2020-2021
Расстояние между матрицами интервальных порядков	0.0555	0.0554	0.0422	0.0581	0.0541	0.0523
Расстояние между матрицами влияния	0.00016	0.00023	0.00029	0.00034	0.00039	0.00044
Устойчивость	0.03921	0.03916	0.02983	0.04108	0.03826	0.03697

### 3.1.8 Результаты анализа сети цитирования авторов

В Таблице 3.4 представлены топ-10 авторов по индексу Total Influence с минимальной квотой 0,1%. Результаты почти аналогичны топ-10 по индексу In-degree (самые цитируемые авторы в научной области). Интересно, что для индекса PI первый автор Ronald B. Postuma имеет значение почти 1, а другие имеют значения, близкие к нулю. Это означает, что у первого автора наибольшее количество ключевых вершин (pivotal nodes), на порядок большее, чем у других.

Таблица 3.4 – Топ-10 авторов по TI, q = 0,1%

№	Имя	Аффилиация	In-degree	BI, q = 0,1%	PI, q = 0,1%	TI, q = 0,1%
1	Ronald B. Postuma	Montreal General Hospital	0,0071	0,577	0,9999	0,528
2	E. Ray Dorsey	University of Rochester	0,0023	0,0251	0	0,0091
3	Anthony H.V. Schapira	University College London	0,0024	0,0245	0	0,009

Продолжение таблицы 3.4

4	Alicia M. Pickrell	National Institute of Neurological Disorders and Stroke (NINDS)	0,0018	0,0142	0	0,0053
5	Ole-Bjørn Tysnes	Haukeland University Hospital	0,00164	0,0122	0	0,0046
6	Timothy R. Sampson	California Institute of Technology	0,00168	0,012	0	0,00457
7	Alberto J. Espay	University of Cincinnati	0,0018	0,011	0	0,0043
8	Mike A Nalls	National Institutes of Health	0,00197	0,0104	0	0,0041
9	Alberto Ascherio	Harvard University	0,0016	0,0105	0	0,004
10	Daniel Weintraub	Veterans Health Administration	0,0017	0,0096	0	0,0038

В Таблице 3.5 представлены топ-10 авторов по TI с наибольшей квотой в 10%. Удивительно, но самый цитируемый автор Ronald B. Postuma снова имеет максимальное значение. По квоте в 5% он также на первом месте. Но он имеет нулевые значения новых индексов. Другие авторы в таблице имеют одинаковые значения индексов. Это может быть объяснено тем фактом, что в сети цитирования авторов веса ребер меньше, чем в сети цитирования журналов, поэтому для большой квоты количество критических множеств и ключевых вершин гораздо меньше. В таблице 3 автора из немецких университетов и 2 из китайских — они интенсивнее других ссылаются друг на друга.

Таблица 3.5 – Топ-10 авторов по TI, q = 10%

№	Имя	Аффилиация	In-degree	BI, q = 10%	PI, q = 10%	TI, q = 10%
1	Ronald B. Postuma	Montreal General Hospital	0,0071	0	0	0,0024
2	Thomas J. Hirschauer	Ohio State University	0,0001	0,0009	0,0015	0,0009

Продолжение таблицы 3.5

3	Yimeng Chen	Chinese Academy of Sciences	0,0001	0,0009	0,0015	0,0009
4	Tatsuya Sasaki	Okayama University	0,0001	0,0009	0,0015	0,0009
5	Michael Khalil	Medical University of Graz	0,0001	0,0009	0,0015	0,0009
6	Inês Caldeira Brás	University of Göttingen	0,0001	0,0009	0,0015	0,0009
7	Hsin Hsi Tsai	National Taiwan University	0,0001	0,0009	0,0015	0,0009
8	Jun Sung Lee	New Generation University College	0,0001	0,0009	0,0015	0,0009
9	Qiu-Jin Yu	Capital Medical University	0,0001	0,0009	0,0015	0,0009
10	Robert D. Salazar	Saarland University	0,0001	0,0009	0,0015	0,0009

### 3.1.9 Результаты паттерн-анализа цитируемости журналов

Для 3292 журналов был произведен анализ паттернов с различными параметрами.

Основная задача такого анализа — кластеризация журналов и выделение отличающихся паттернов. Это позволит определить журналы, цитируемость которых выросла за период с 2015 по 2021, а также те журналы, популярность которых угасает со временем. Было решено проводить паттерн-анализ по изменению индекса In-degree по годам.

Первый вариант кластеризации — кластеризация по абсолютному значению. Таким образом возможно определить журналы, цитируемость которых сильно отличается от цитируемости основной массы по значению. Параметры для кластеризации были выбраны следующие:  $\epsilon = 0,05$ . Было выделено 20 паттернов. Подробное описание каждого паттерна с анализом журналов приведено на электронном ресурсе <https://www.hse.ru/DeCAn/parkinson>. В самый большой паттерн попало 3228 журналов из 3292: значения их цитируемости было близким к нулю все 7 лет. Было выявлено несколько паттернов, цитируемость которых была больше остальных, но она не изменялась на протяжении 7 лет и не была самой высокой среди журналов.

### **3.1.10 Результаты анализа устойчивости сети цитирования журналов**

Для сети цитирования журналов, состоящей из 3292 вершин, был проведен анализ устойчивости. Было построено 7 сетей для каждого года с 2015 по 2021, оценивалось динамическое изменение сети цитирования.

Можно отметить, что для сети цитирований журналов большее значение имеет изменение в центральности, поскольку за 7 лет изменения в структуре такой сети практически незаметны. Максимальные изменения в центральности произошли с 2015 по 2016 год, а максимальные изменения в топологии сети - с 2020 по 2021. Измеряя стабильность сети таким способом, можно заключить, что максимальное изменение в сети произошло с 2015 по 2016 год и с каждым годом сеть становилась все стабильнее.

### **3.1.11 Заключение**

Новые модели центральности позволяют нам учитывать различные параметры вершин, такие как групповое взаимодействие и влияние ключевых узлов. Классические и новые индексы центральности были подсчитаны для более 3 тысяч журналов, 3 тысяч организаций и почти 30 тысяч авторов, из которых выявлены наиболее значимые. Используя данный метод, можно выявить группы интенсивно цитирующих друг друга авторов и журналов. Более того, этот метод может быть использован в различных научных областях для того, чтобы выделить журналы, организации и авторов, которые могут стать прибыльными объектами для инвестиций.

Также был произведен паттерн-анализ для более 3 тысяч журналов и организаций и оценена устойчивость сетей цитирования во временном промежутке с 2015 по 2021, которая учитывает изменения в топологии сети и центральности вершин.

Данный подход к анализу научной области является универсальным и может быть применен в исследовании различных областей науки для определения активно развивающихся групп журналов, авторов, организаций и, как следствие, направлений исследования.

Результаты данного исследования нашли отражение в [168].

## **3.2 Применение метода аномальных кластеров к анализу температурных карт**

### **3.2.1 Введение**

Работа по кластер-анализу шла в 2023 году, главным образом, по распространению ранее предложенного нами метода аномальных кластеров на картографические данные, меняющиеся во времени. Метод аномальных кластеров в применении к стандартным данным вида объект-признак описан ниже в пункте 3.2.3. Задача состоит в том, чтобы по карте тем-

пературы морской поверхности выделить «холодный участок», соответствующий так называемому апвеллингу – солитоно-подобному образованию типа горизонтальной трубы длиной в сотни километров и диаметром несколько метров, которая вращается вдоль своей оси, создавая условия, массово привлекающие к себе рыбу и тем самым крайне важная для рыбного промысла, играющего большую роль в экономике прибрежных стран таких, как Португалия. Наша работа относится к числу пионерских исследований этого феномена. В результате применения описываемой далее трехступенчатой методики кластерного анализа к картам температуры морской поверхности у берегов Португалии за многие годы нам удалось показать, что апвеллинг в этих водах является кусочно-постоянным образованием с 3–5 периодами стабильности, переходящими друг в друга через не слишком большие изменения.

### 3.2.2 Особенности применения алгоритма выделения аномальной группы на океанских данных

Речь идет о так называемых картах температуры морской поверхности (SST grids), т. е. прямоугольных решеток  $G_t$ , ( $t=1, 2, \dots, T$  – моменты времени; в нашем анализе – 27 средне-недельных данных с апреля по октябрь), в узлах  $(i,j)$  которых задана температура  $t_{ij}$ .

Методика применения нашего метода аномального кластер-анализа включает три ступени:

Ступень 1. Применение метода аномального кластера к температурным данным для формирования образа апвеллинга в каждый данный момент  $t=1, 2, \dots, T$ .

На этой ступени метод аномального кластера в его версии, описанной в [169], формирует образ апвеллинга как индикаторную функцию  $u=(u_{ij})$ , где  $u_{ij}=1$  для точек  $(i,j)$ , принадлежащих апвеллингу, и  $u_{ij}=0$  для точек, не принадлежащих ему (процесс проиллюстрирован ниже с помощью рисунка 3.1).

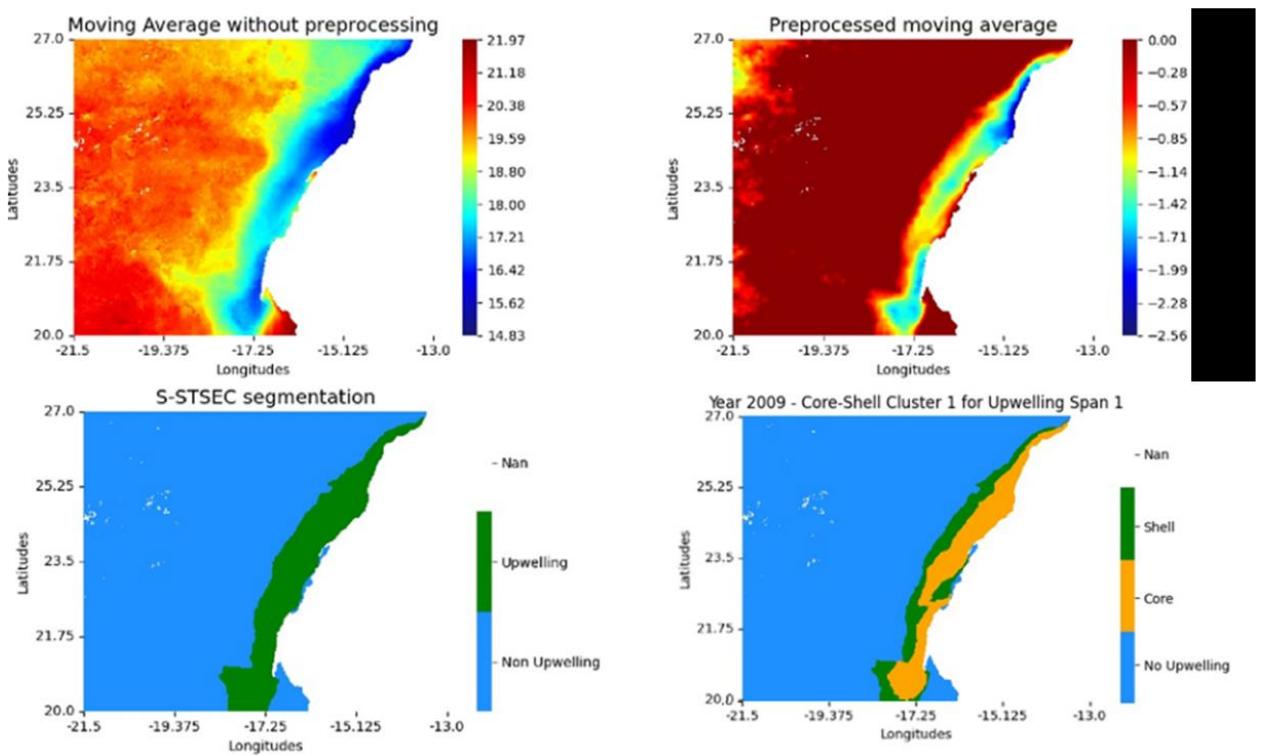


Рисунок 3.1 – Начальные и финальные шаги процесса применения методики аномального кластер-анализа для карт температуры морской поверхности. Исходные данные (слева вверху); пред-обработанные (путем перехода к скользящему среднему) данные (справа вверху); выделение сегмента апвеллинга (слева внизу), разделение сегмента на постоянную и «оболочечную» части (справа внизу). Фрагмент белого цвета – поверхность суши.

### Ступень 2. Выявление периодов стабильности.

Охарактеризуем каждый образ апвеллинга следующими четырьмя признаками

- площадь образа апвеллинга;
- средняя температура в образе апвеллинга;
- максимальная широта образа апвеллинга;
- минимальная широта образа апвеллинга.

Таким образом, на вход алгоритма аномального кластер-анализа поступает матрица размера  $T \times 4$ . Алгоритм последовательно разбивает последовательность недель  $t=1, 2, \dots, T$  на кластеры, характеризующие периоды стабильности, при этом автоматически определяя их количество [170].

### Ступень 3. Выявление постоянного ядра в периоде стабильности.

В каждом заданном периоде стабильности мы рассматриваем карты температуры морской поверхности и применяем к ним модификацию алгоритма аномального кластера, в которой динамика характеризуется наличием постоянного «ядра» и переменной «оболочки», как предложено нами в [171] и подробно расписано в [172].

Получаемое на этой ступени множество «ядер» кластеров и образует кусочно-постоянную модель явления апвеллинг, наблюдаемого ежегодно у берегов Португалии и некоторых других стран. Океанографические данные подтверждают адекватность наших построений, как показано в [172].

### 3.2.3 Аномальные кластеры

Данный метод в какой-то мере «разведывает» структуру данных. Он основан на последовательном выявлении и удалении так называемых «аномальных» групп. «Аномальность» понимается как удаленность от некой «реперной» точки.

Реперная точка выбирается как проявление «нормы», «среднего» или «нормального» объекта, не обязательно среди множества наблюденных объектов. Во многих случаях центр гравитации множества объектов, то есть его среднее, может быть выбран как реперная точка отсчета.

Использование реперной точки позволяет сравнивать объекты не друг с другом, а именно с ней, что существенно экономит объем вычислений: вместо перебора всех парных расстояний между объектами, есть возможность сконцентрировать усилия на вычислении только расстояний между объектами и реперной точкой, что даёт снижение порядка количества шагов с  $N^2$  до  $N$ .

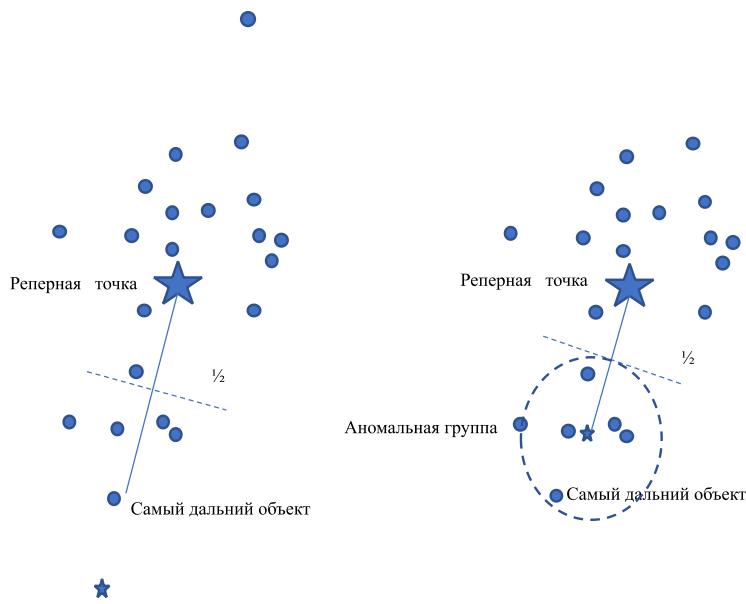


Рисунок 3.2 – Извлечение аномальной группы при реперной точке, расположенной в центре масс (большая звезда); малая звезда представляет центр аномальной группы. Первая итерация показана на левой части рисунка, а финальная – на правой

Аномальная группа конструируется как кластер, наиболее удаленный от реперной точки. Процесс начинается с того, что объект, наиболее удаленный от реперной точки, объявляется центром аномальной группы. Затем версия метода K-средних с  $K=2$  применяется к двум центрам: один – это реперная точка, которая не меняется всё время процесса вычислений, а второй – это центр аномальной группы, который обновляется согласно стандартной процедуре. А именно, при заданном аномальном центре, аномальная группа определяется как множество объектов, которые ближе к этому центру, чем реперной точке. При заданной аномальной группе её центр вычисляется как центр масс с помощью нахождения среднего всех входящих в группу объектов. Процедура повторяется до тех пор, пока не сойдется (см. рисунок 3.2).

Таким образом, метод аномальной группы – это версия метода K-средних, в которой:

- (A) количество кластеров  $K$  равно 2;
- (Б) центром одного из кластеров является 0, перенесённый в реперную точку и не меняющийся в процессе итераций;
- (В) начальный центр аномальной группы выбирается как максимально удалённый от точки 0.

Свойство (B) автоматизирует определение начального центра, исходя из того, что центр должен быть максимально удаленным от точки отсчета 0.

Прежде чем описать математическую модель аномальной группы, дадим формулировку алгоритма выделения аномальной группы.

Алгоритм выделения аномальной группы ВАГ

- 1) Инициализация аномального центра. Найдем объект, максимально удаленный от начала координат, 0, и поместим туда начальный аномальный центр,  $c$ .
- 2) Обновление аномальной группы. Определим аномальную группу  $S$  вокруг  $c$  правилом: объект  $y_i$  относится к кластеру  $S$ , если  $d(y_i, c) < d(y_i, 0)$ .
- 3) Обновление аномального центра. Вычислим внутригрупповое среднее  $c'$  для множества объектов  $S$  и проверим, отличается ли оно от предыдущего центра  $c$ . Если  $c'$  и  $c$  не равны друг другу, обновляем центр присвоением  $c \leftarrow c'$  и переходим к шагу 2. В противном случае переходим к шагу 4.
- 4) Выдача результатов. Список группы  $S$  и центр  $c$  выдаются как результат работы алгоритма.

Нетрудно доказать, что метод аномальной группы, как и метод K-средних, альтернативно минимизирует некий критерий

$$W(S, c) = \sum_{i \in S} d(y_i, c) + \sum_{i \notin S} d(y_i, 0), \quad (3.1)$$

где  $S$  – это искомое подмножество  $I$ , а  $c$  – центр  $S$  в пространстве признаков. При этом метод аномальной группы отличается от метода К-средних при  $K=2$  в следующем: здесь имеется только один центр,  $c$ , который и обновляется в алгоритме; другой центр, 0, не меняется никогда и используется только для приписывания ему объектов, не входящих в аномальную группу. Поэтому метод К-средних при  $K=2$  на выходе дает два кластера, а метод аномальной группы – только один, наиболее удаленный от реперной точки, 0.

Можно показать, что критерий (3.1) – не что иное, как критерий наименьших квадратов для модели

$$y_{iv} = \begin{cases} c_v + e_v, & i \in S \\ 0 + e_v, & i \notin S \end{cases} \quad (3.2)$$

где  $i$  – номер объекта,  $v$  – номер признака, а  $S$  – множество объектов искомой аномальной группы, тогда как  $c = (c_1, \dots, c_v, \dots, c_V)$  – её центр. Эта модель имеет очень простую структуру: объект либо принадлежит аномальной группе, либо нет, причем в первом случае он должен как можно точнее совпадать с  $c$ , а во втором – с 0! (Не забудем, что 0 здесь – это реперная точка, например точка средних значений всех признаков.) Подобная простая модель характерна и для метода К-средних: в ней каждый объект должен как можно точнее совпадать с центром кластера, которому он принадлежит. Подобные простые модели лежат в основе практических всех популярных методов анализа данных. В частности, путем раскрытия скобок в квадратичных расстояниях  $d$ , критерий (3.1) может быть преобразован к следующему виду:

$$W(S, c) = \sum_{i=1}^N \sum_{v=1}^V y_{iv}^2 - |S| \sum_{v=1}^V c_v^2. \quad (3.3)$$

Таким образом, приходим к разложению разброса данных на сумму двух слагаемых

$$T(Y) = |S| \|c\|^2 + W(S, c), \quad (3.4)$$

где  $T(Y)$  обозначает разброс данных. Первое слагаемое выражает ту часть разброса данных, которая объяснена аномальной группой, а второе – необъясненную часть. Поскольку метод

минимизирует необъясненную часть, а разброс данных – постоянная величина, то одновременно максимизируется объясненная часть. Ее относительная величина, результат деления на разброс данных, выражает вклад аномальной группы в разброс данных. Чем больше вклад, тем лучше эта группа отделена от остальных данных. Выдача вклада может производиться на шаге 4 Алгоритма ВАГ.

### **3.3 Разработка человеко-машинной процедуры паттерн-кластеризации данных**

#### **3.3.1 Введение**

В данной части отчета рассматривается задача кластеризации данных, основанная на структурном сходстве объектов. Понятие "паттерн" широко используется в задачах анализа данных и происходит от английского слова pattern, которое означает "образ". Но в задачах машинного обучения зачастую используется в значении «шаблон», то есть ряд схожих, общих характеристик объектов, объединенных в один класс или кластер. Задача стоит не только в том, чтобы объединить схожие объекты, но и в делении объектов на кластеры, максимально удаленные друг от друга по признакам.

В последнее время активное развитие получила сфера машинного обучения и кластерного анализа. Значительное увеличение количества данных приводит к необходимости их анализа, объединения схожих объектов в кластеры и описание их паттернов.

Применяя различные алгоритмы кластеризации к данным, может возникнуть ряд проблем, таких как проблема вычислительной сложности существующих алгоритмов кластеризации, а также отсутствие визуализации при использовании кластеризаций для анализа шаблонов. Ряд алгоритмов хорошо работает только на больших данных. В таком случае эти алгоритмы не применимы к анализу на региональном или страновом уровне за малый период времени.

Также существующие алгоритмы кластеризации не позволяют внедрять анализ результатирующих паттернов непосредственно во время кластеризации данных.

В нашей работе был реализован метод кластеризации с линейной сложностью. Основной целью была разработка человеко-машинной процедуры, которая позволяет анализировать объекты и их структуры непосредственно в процессе кластеризации на основе визуально отображаемых признаков.

### 3.3.2 Аналитический обзор литературы

В [173] рассматриваются существующие методы кластеризации, выделения схожих признаков объектов, предсказания поведения объектов, уточняет основные понятия, связанные с этой областью. Было уточнено понятие паттерна и его использование для динамики данных.

Работа [174] описывает метод параллельных координат и его применение к анализу возможных результатов голосования на выборах по трем округам и анализу банков России и Турции.

В [175] анализируются 100 крупнейших банков России по стереотипам их поведения. Выполнен анализ изменения структуры банковского сектора России. Цель данной работы состояла в выявлении паттернов поведения банков России и их тенденций развития. Паттерны поведения изучены методом динамического анализа, описанного в [173, 174].

В [175] кластерный анализ применен к анализу макроэкономических показателях 119 стран. Рассматривались такие показатели, как темпы роста реального ВВП, денег и инфляции для каждой страны. В результате все 119 стран были сгруппированы в 6 кластеров, состоящих из похожих стран.

### 3.3.3 Математическая модель построения δ-трубки

Каждый объект – вектор размера  $k$  со значениями от 0 до 1, который можно представить в виде графика в параллельных координатах

$$u: u_i \in [0, 1], \dim(u) = k. \quad (3.5)$$

Для каждого объекта его параметры отображаются в параллельных координатах, а затем проводятся отрезки, соединяющие точки, соответствующие значению  $u_i$ , на соседних координатах. В результате получается график объекта (рисунок 3.3).

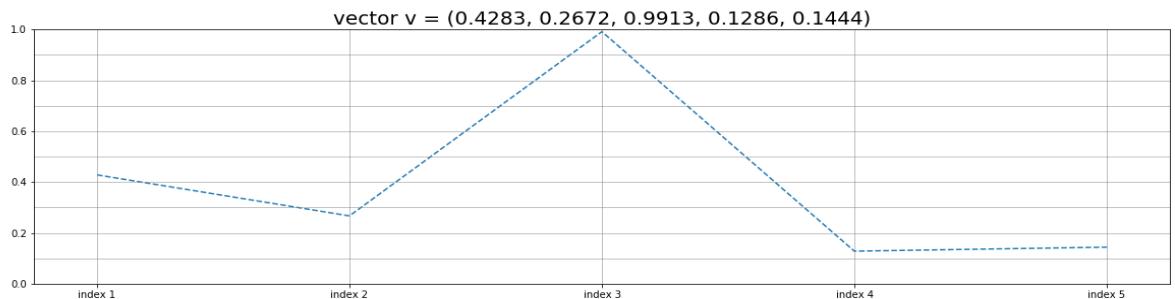


Рисунок 3.3 – Пример графика объекта

Далее каждому объекту сопоставляется вектор  $v$ , который зависит от типа кластеризации:

- кластеризация объектов по абсолютным значениям вектора  $u$ ;
- кластеризация объектов по тангенсам углов наклонов отрезков графика абсолютных значений вектора  $u$ .

Мы считаем, что соседние параллельные координаты находятся на расстоянии 1 друг от друга. То есть  $x_{i+1} - x_i = 1$ . Откуда получаем

$$\operatorname{tg} \alpha_i = \frac{y_{i+1} - y_i}{x_{i+1} - x_i} = \frac{u_{i+1} - u_i}{1} = u_{i+1} - u_i. \quad (3.6)$$

Откуда

$$v \in R^{k-1}: v_i = u_{i+1} - u_i. \quad (3.7)$$

Кластеризация строится по случайно выбранному вектору  $v$ . По вектору  $v$  считается  $\delta$ -трубка, зависящая от параметров  $\varepsilon$  и  $h$ . Это вектор, размерность которого совпадает с размерностью  $v$ : для каждого значения  $v_i$  считается максимально возможное отклонение  $\delta_i$  от этого значения  $v_i$  внутри кластера  $K(v)$ , образованного  $v$ . Вектора, попадающие в  $\delta$ -трубку вектора  $v$ , объединяются с ним в один кластер (рисунок 3.5)

$$v' \in K(v) \Leftrightarrow \forall i \quad |v'_i - v_i| < \delta_i. \quad (3.8)$$

Типы построения  $\delta$ -трубки:

- фиксированная трубка: ширина трубы не зависит от значений вектора  $v$  (см. рисунок 3.4):  $\delta_i = \varepsilon$ ;
- адаптивная трубка: ширина трубы пропорциональна модулюм значений вектора  $v$  (см. рисунок 3.5):  $\delta_i = \varepsilon |v_i| + h$ ;
- комбинированная трубка: трубка с фиксированной шириной, пропорциональная максимальному модулю значений вектора  $v$  (см. рисунок 3.6):  $\delta_i = \varepsilon \max_i(|v_i|) + h$ .

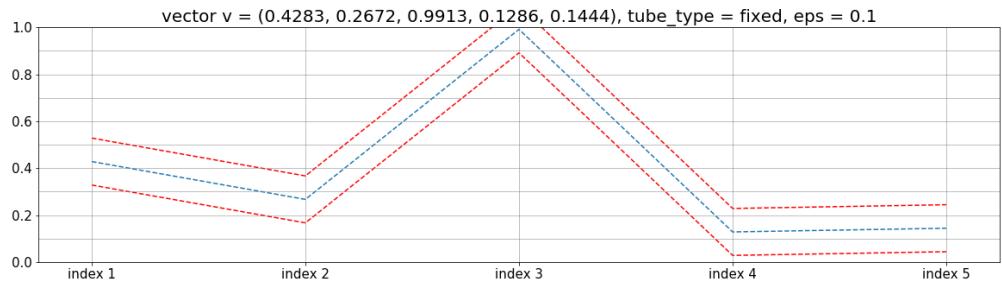


Рисунок 3.4 – Фиксированная трубка

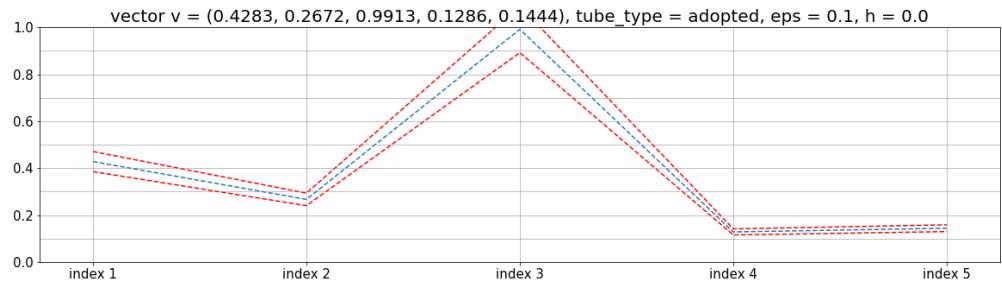


Рисунок 3.5 – Адаптивная трубка

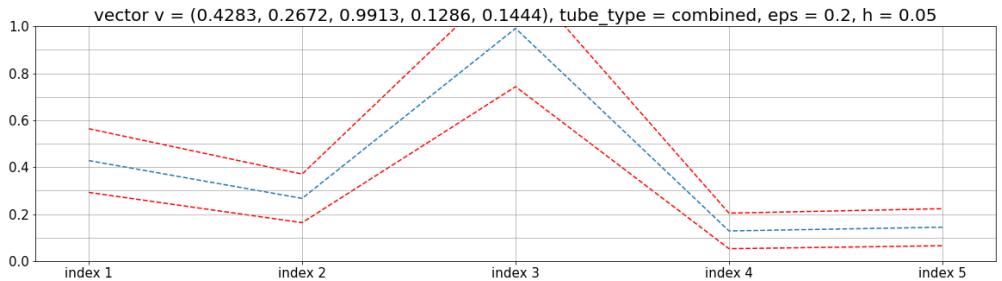


Рисунок 3.6 – Комбинированная трубка

Кластеризация запускается  $t$  раз, в результате чего получается набор разбиений объектов  $\Omega = \{\Omega_j\}_{1 \leq j \leq t}$ . В каждом наборе  $m_j$  кластеров:  $\Omega_j = \{K_k\}_{1 \leq k \leq m_j}$ . В каждом кластере  $n_k$  векторов  $K_k = \{v_l\}_{1 \leq l \leq n_k}$ .

Введем  $p_k$  – паттерн кластера  $K_k$  или его середина:

$$p_{k_i} = \frac{\sum_{v \in K_k} v_i}{n_k}. \quad (3.9)$$

Далее выбирается разбиение  $\Omega_j$ , максимизирующее разность функционалов  $f_1$  и  $f_2$ :

$$f_1(\Omega_j) = \frac{\sum_{1 \leq k_1, k_2 \leq m_j} (\rho(p_{k_1}, p_{k_2}))}{m_j^2},$$

$$f_2(\Omega_j) = \frac{\sum_{1 \leq k \leq m_j} \left( \frac{\sum_{v \in K_k} \rho(v, p_k)}{n_k} \right)}{m_j}, \quad (3.10)$$

где  $\rho(x, y)$  – квадрат евклидова расстояния

$$\rho(x, y) = (\sqrt{\sum_i (x_i - y_i)^2})^2 = \sum_i (x_i - y_i)^2. \quad (3.11)$$

Первый функционал – среднее квадратов попарных евклидовых расстояний между паттернами кластеров. При его максимизации паттерны максимально удалены друг от друга.

Второй функционал – среднее квадратов евклидовых расстояний от векторов кластера до середины этого кластера. При его минимизации векторы каждого кластера максимально плотно расположены друг к другу.

### 3.3.4 Описание используемых программных продуктов

Использовались следующие программные продукты:

При

- python 3.11;
- NumPy 1.23.4;
- Pandas 1.5.1;
- Dash 2.7.0;
- Dash-core-components 2.0.0;
- Dash-html-components 2.0.0;
- Dash-table 5.0.0;
- Flask 2.2.2;
- Plotly 5.11.0;
- Matplotlib 3.5.3.

Проект состоит из двух файлов:

- app.py;
- mathmodel.py

Файл mathmodel.py подключается к файлу app.py в качестве библиотеки. И содержит в себе реализованный алгоритм кластеризации.

Файл app.py содержит в себе объект класса Dash, к которому применяются методы layout для определения макета приложения.

Numpy и Pandas позволяют использовать векторные вычисления при работе с данными.

### 3.3.5 Спецификации входных и выходных данных программы

Входные данные программы:

- load-data – csv-file с сепаратором "запятая". Представляет из себя нормированные от 0 до 1 данные;
- pattern-data – csv-file с сепаратором "запятая" – результат найденной ранее кластеризации;
- input1 – список групп номеров паттернов, которые необходимо объединить. Группы паттернов вводятся через точку с запятой. Номера паттернов, которые необходимо объединить, вводятся через пробел;
- input2 - список номеров групп, которые необходимо пересчитать. Вводится через пробел;
- input3 – значение  $\varepsilon$ ;
- input4 – значение  $h$ ;
- input5 – тип кластеризации;
- input6 – тип трубы;
- calculate – кнопка, при нажатии которой применяется алгоритм кластеризации к выбранным паттернам или ко всей выборке данных;
- patternlist-button – кнопка, скрывающая и открываящая список кластеров с количеством объектов в них;
- draw – кнопка, при нажатии которой рисуются интерактивные графики;
- draw-save – кнопка, при нажатии скачиваются графики zip-file с графиками;
- Download numbers of patterns – ссылка, при нажатии на которую скачивается csv-file с результатами кластеризации.

Выходные данные программы

- csv-file с сепаратором "запятая" содержащий список кластеров для каждого объекта;
- zip-file, содержащий нарисованные в формате png графики.

### 3.3.6 Описание интерфейса программы

Изначально открывается страница загрузки данных (рисунок 3.7). В поле load data загружаются данные для кластеризации. В поле Load numbers of patterns загружаются готовые результаты кластеризации, если кластеризация для этих данных проводилась ранее.

После загрузки этих полей открывается панель с вводом параметров кластеризации и кнопками “Download charts as zip”, “Show list of patterns” и, если загружены данные, панель отрисовки и скачивания картинок (рисунок 3.8).

При нажатии кнопки “Show list of patterns” список паттернов открывается. При нажатии кнопки “Hide list of patterns” прячется.

При некорректном вводе данных в поля ввода поле обводится в красную рамку и нажатие кнопки enter игнорируется.

При нажатии “Download numbers of patterns”, скачивается результат кластеризации.

При нажатии “Download charts as zip”, скачивается zip-файл с графиками паттернов в параллельных координатах.

При нажатии кнопки “Draw patterns” рисуются графики паттернов в параллельных координатах (рисунок 3.9).

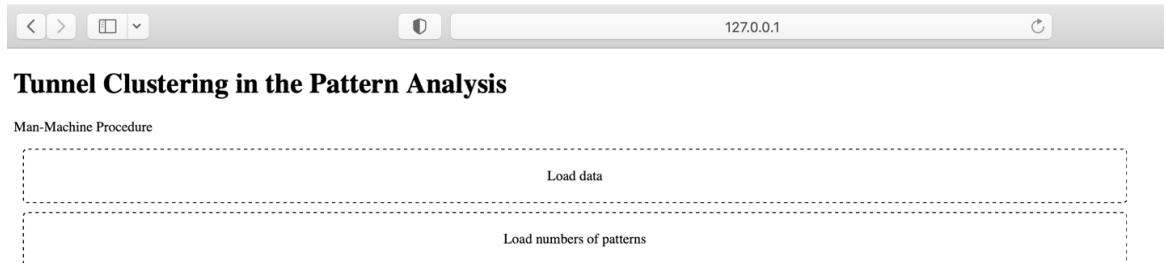


Рисунок 3.7 – Страница загрузки данных

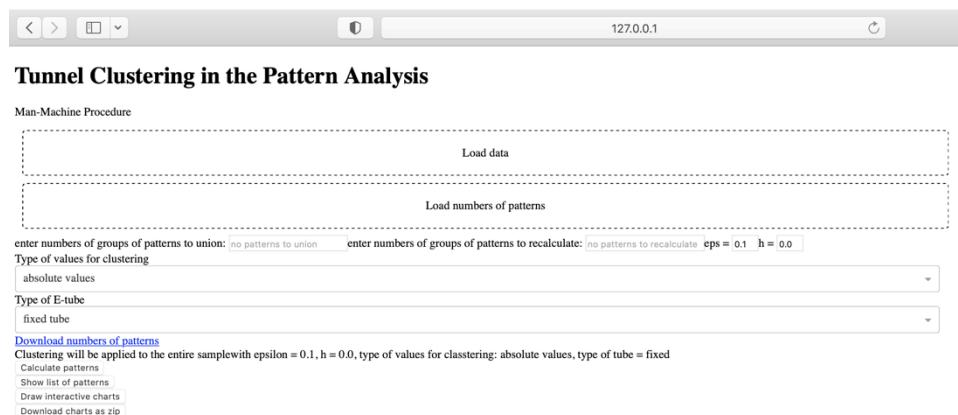


Рисунок 3.8 – Панель с вводом параметров кластеризации, отрисовки и скачивания картинок

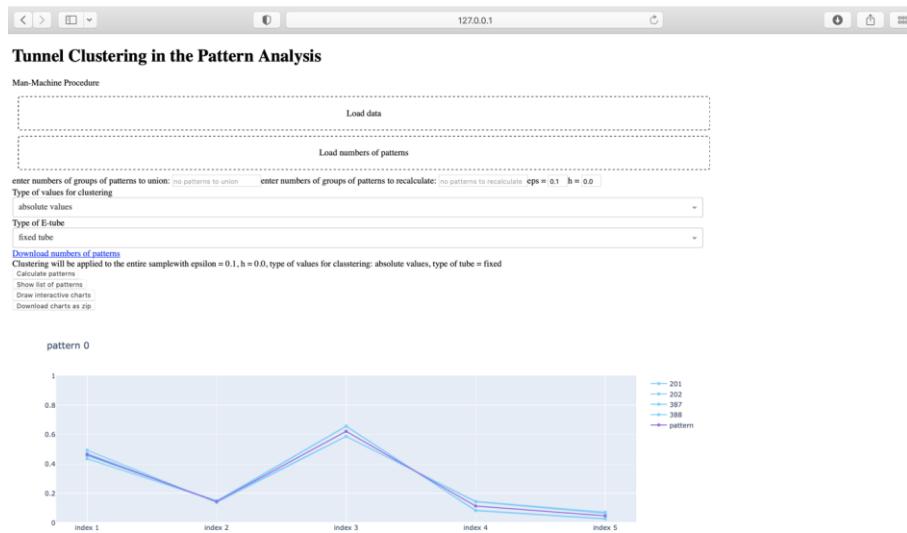


Рисунок 3.9 – Отрисовка графиков паттернов в параллельных координатах

### 3.3.7 Результаты кластеризации данных

Во всех нижеописанных примерах применения человека-машинной процедуры использовалась минимаксная нормировка

#### 3.3.7.1 Кластеризация аффилиаций авторов статей, посвященных болезни Паркинсона

Была проведена кластеризация по 3029 аффилиациям авторов статей, посвященных болезни Паркинсона. В качестве векторов объектов был рассмотрен индекс In-degree в каждый год в период с 2015 по 2021 года.

Параметры кластеризации:  $\epsilon = 0.1$ ,  $h=0$ .

Тип кластеризации:

- кластеризация с фиксированной трубкой;
- кластеризация объектов по тангенсам углов наклона.

В результате кластеризации были получены 29 паттернов, из которых одну аффилиацию содержат 18 паттернов. Так же был получен один крупный кластер, составляющий 93, 5% выборки.

Данная кластеризация позволила выделить схожие по количеству цитирований с течением времени аффилиации.

В выделенный крупный паттерн входят аффилиации, с незначительным изменением индекса цитирования с течением времени.

### 3.3.7.2 Первая кластеризация журналов со статьями, посвященными болезни Паркинсона

Было проведено кластеризация данных по 3292 журналам. В качестве векторов объектов был рассмотрен индекс In-degree в каждый год в период с 2015 по 2021 года.

Параметры кластеризации:  $\epsilon = 0.05$ ,  $h=0$ .

Тип кластеризации:

- кластеризация с фиксированной трубкой;
- кластеризация объектов по абсолютным значениям.

В результате кластеризации было выделено 19 паттернов. Самый крупный из них составляет 98% выборки. Индекс цитируемости этих паттернов меньше 0.05.

Индекс In-degree для остальных журналов значительно выше. Наиболее цитируемые журналы:

- Journal of Parkinson's Disease;
- Brain;
- Neurobiology of Disease;
- Scientific Reports;
- Lancet Neurology;
- PLOS ONE;
- Frontiers in Neurology;
- Movement Disorders – Parkinsonism & Related Disorders;
- Neurology.

### 3.3.7.3 Вторая кластеризация журналов со статьями, посвященными болезни Паркинсона

Данная кластеризация так же была проведена кластеризация по 3292 журналам. В качестве векторов объектов был рассмотрен индекс In-degree в каждый год в период с 2015 по 2021 года.

Параметры кластеризации:  $\epsilon = 0.1$ ,  $h=0$ .

Тип кластеризации:

- кластеризация с фиксированной трубкой;
- кластеризация объектов по тангенсам углов наклона.

Был выделен один крупный паттерн, включающий в себя журналы, индекс цитирования для которых не менялся с течением времени. В него входит 3281 из 3292 журналов.

Кластеры журналов, для которых наблюдается изменение количества цитирований с течением времени:

- Scientific reports;
- Journal of Neurochemistry и Parkinsonism & Related;
- Neurobiology of Aging и Neuroscience;
- Plos one;
- Train;
- Frontiers in neurology;
- Neurobiology of disease;
- Frontiers in neuroscience;
- Lancet neurology.

### 3.3.7.4 Кластеризация банков России по шести показателям

В данной работе была проведена кластеризация банков России за период с 2017 по 2021 года по месяцам по пяти показателям:

- соотношение объемов депозитного кредитного портфелей;
- коэффициент рентабельности активов;
- норматив текущей ликвидности;
- норматив достаточности капитала.

Соотношение резервов на возможные потери по ссудам к кредитам 270 банков были разделены на три группы:

- крупные банки (1140 банков);
- средние банки (3840 банков);
- малые банки (11220 банков).

В результате кластеризации крупных банков было выделено 19 паттернов, 5 из которых составляют 89.4%. Это такие банки, как Сбербанк, Газпромбанк, Альфа-Банк, Россельхозбанк, Московский кредитный банк, Райффайзенбанк, Росбанк, АБ Россия, ЮниКредит Банк, ВБРР, СМП банк и Новикомбанк. Сбербанк и Росбанк.

В результате кластеризации средних банков было выделено 18 паттернов. Самый крупный паттерн составляет 89.7%. На 97.8% приходится первые 5 паттернов. На оставшиеся паттерны приходятся банки, которые активно привлекают депозиты, банки, которые имеют высокую долю средств вкладчиков, а также банки, бизнес-стратегия которых опи-

сывается данным паттерном, характеризуются значительным переизбыtkом капитала, низким качеством активов, высокой долей депозитов, низкой прибыльностью и приемлемым запасом ликвидности.

В результате кластеризации малых банков России было выделено 16 паттернов. Самый крупный из них составляет 97% выборки, 5 самых крупных паттернов составляют 99%. В самый крупный паттерн входят все банки, кроме Ю Би Эс Банка.

Данная кластеризация позволила определить основные стратегии поведения банков среди крупных, средних и малых банков. А также выделить банки, стратегия которых отличается от стратегий остальных банков.

### 3.3.7.5 Кластеризация стран мира по экспорту-импорту электроэнергии в период с 2000 по 2020 годы

Кластеризация была проведена по 81 стране за 2000, 2005, 2010, 2015 и 2020 года. Каждая страна в каждый конкретный год воспринималась, как отдельный объект. Были рассмотрены следующие параметры этих стран:

- ВВП на душу населения;
- импорта на душу населения;
- ВВП на импорт;
- потребление на душу населения.

К этим данных была применена человеко-машинная процедура со следующими параметрами:  $\epsilon = 0.175$ ,  $h = 0.2$ .

Тип кластеризации:

- кластеризация с комбинированной трубкой;
- кластеризация объектов по тангенсам углов наклона.

Десять самых крупных кластеров составляют 57% выборки. Два самых крупных кластера с 45 наблюдениями, с 33 наблюдениями и с 23 наблюдениями соответственно. Включают в себя страны с низкими показателями. В первый кластер входят страны с развивающейся экономикой, во второй как страны с развивающейся экономикой, так и страны и наибеднейшие страны, в третий так же входят развивающиеся страны.

Отдельно был выделен кластер, состоящий из 33 наблюдений, включающий в себя наиболее развитые, европейские страны. Они обладают достаточно высоким ВВП и потреблением на душу населения, при этом низкими показателями импорта на душу населения и импорта на ВВП.

В отдельные паттерны, состоящие 3–4 объектов, вошли такие страны как, Швейцария, Европейский союз, Черногория, Люксембург, Норвегия. Из чего был сделан вывод о высокой изменчивости рассмотренных показателей, связанных с электричеством.

### 3.3.7.6 Паттерн-анализ в исследовании рынка соединений редкоземельных металлов

Был проведен кластерный анализ 110 стран мира за период с 2005 по 2020 года. Были рассмотрены показатели:

- импорт соединений редкоземельных металлов на душу населения в долларах;
- ВВП на душу населения в миллиардах долларов.

В ходе анализа были выделены 18 паттернов, 4 самых крупных из которых: 0, 7, 9, 12, – составляют 77% выборки.

В нулевой паттерн входят такие страны, как США, Австрия, Франция, Швейцария, Бельгия. Это развитые страны, с наибольшим средним значением ВВП на душу населения. У этих стран также наибольшее значение импорта редкоземельных металлов на душу населения, что свидетельствует о значительном потреблении редкоземельных металлов в этих странах.

Седьмой паттерн – самый крупный паттерн. Сюда входят в основном развивающиеся страны такие, как Китай, Россия, Белоруссия, Тунис. Однако в данном паттерне присутствуют и развитые страны: Словакия, Португалия, Израиль, Словения, Литва, Латвия, Чехия, Южная Корея. Здесь также есть неразвитые страны: ЮАР, Таиланд, Шри-Ланка, Барбадос, Намибия, Маврикий, Ливан. Эти страны в разные годы находятся или 7-м паттерне, или в 9-м.

Среднее значение импорта на душу населения в этих странах незначительное. Всего 0.6 долларов на душу населения. ВВП на душу населения в этих странах значительное 16461.38 млрд долларов на душу населения.

В девятый паттерн вошли в основном слаборазвитые страны, с низкими значениями показателей. Однако здесь также встречаются Китай, который меняет свой паттерн с девятого на седьмой в 2015 году, и Индия, которая меняет свой паттерн с 12 на 9 в 2010 году. Объем импорта соединений редкоземельных металлов в этих странах значительный, но из-за огромного населения в него импорт на душу населения незначительный.

Двенадцатый паттерн включает в себя страны преимущественно отсталые и слаборазвитые страны.

В отдельные кластеры попали Эстония и Катар. В Эстонии находится репарационная установка, производящая редкоземельные металлы высокой частоты. Учитывая низкое население этой страны, импорт на душу населения в Эстонии значителен. Катар, в свою

очередь, за счет богатых ресурсов входит в топ-5 стран по ВВП на душу населения. Импорт же в Катаре низкий, за счет чего он был выделен в отдельный кластер.

### 3.3.8 Заключение

Средствами языка Python реализована модель паттерн-кластеризации данных с различными типами кластерных трубок. Разработана человеко-машинная процедура, позволяющая объединять и обрабатывать вручную результаты разных кластеризаций.

Алгоритм кластеризации позволяет выделять объекты, отличающиеся от остальной выборки, исходя из их структурной схожести. Написанный веб-интерфейс осуществляет отрисовку графиков результатов кластеризации. Анализ этих графиков дает возможность выделять новые паттерны или объединять существующие в полуавтоматическом режиме.

Рассмотренные примеры кластеризации показывают, что данный алгоритм может выделить как уникальные объекты, чьи характеристики резко отличаются от характеристик большинства объектов, так и основные паттерны характеристик схожих объектов.

Алгоритм хорошо работает, как на малых выборках, так и на больших.

## 3.4 Паттерн-анализ экспорта-импорта электроэнергии

### 3.4.1 Введение

Электроэнергетические показатели находятся в тесной связи с индексами экономического развития страны: например, уровень потребления указывает как на масштаб промышленности в государстве (для технологичного производства требуются большие объемы электричества), так и на обеспеченность домохозяйств (чем крупнее экономика – тем более развита энергосеть и обширнее использование электроэнергии для повседневных нужд). Данное исследование посвящено анализу показателей электроэнергии и связанных с ними макроэкономических процессов, политических и экономических тенденций.

Метод паттерн-анализа ранее использовался в изучении социоэкономических [177] и финансовых [178] процессов и явлений на страновом и региональном уровнях, позволяя изучить совокупность всех релевантных признаков и дать комплексную оценку исследуемой теме, что даёт преимущество перед классическими статистическими методами.

В ходе данной работы был использован анализ паттернов по уровню ВВП и показателям, связанным с производством, потреблением и закупкой электроэнергии, изучено поведение стран в каждом из исследуемых временных периодов и сформированы кластеры, объединяющие похожие по паттернам объекты. Такая группировка даёт возможность охарактеризовать каждую из стран и описать с макроэкономической точки зрения причины, обуславливающие тот или иной набор значений показателей. Динамический анализ паттер-

нов в дальнейшем демонстрирует направленность изменений поведения стран в показателях электроэнергии, как для конкретных стран, так и глобально. Рассмотрение отдельных, наиболее значимых для исследования объектов даёт возможность проанализировать изменения энергетических политик крупнейших экономик и, с помощью изученной литературы, найти им объяснение.

Совокупность вышеописанных методов в этой работе ставит целью формирование обобщённого представления о глобальном рынке электроэнергии, выявление зависимостей отдельных стран от импорта данного блага, анализ макроэкономических и бизнес-процессов, приводящих к изменению тех или иных показателей, а также создание основы для дальнейших исследований в области торговли электроэнергией.

### 3.4.2 Обзор литературы

Большинство статей и отчётов обращают внимание на агрегированные показатели производства, потребления и глобального спроса на электричество. Например, в Ember's Global Electricity Review [179] отмечается, что в структуре спроса в последние годы доминируют Азия и Северная Америка с более чем 75% суммарно. Крупнейшими странами по объёму спроса на электричество являются Китай (31%) и США (15%), а Россия занимает лишь 4% от общего спроса. При этом большая часть спроса удовлетворяется за счёт собственных производственных мощностей.

Говоря о структуре предложения [180], обозреватели уделяют особое внимание альтернативным и возобновляемым источникам энергии (солнечная энергетика, ветряная, водородная и другие), хотя лишь небольшая часть спроса удовлетворяется за счёт них.

Международное энергетическое агентство (IEA) отмечает [181], что тенденция к росту спроса на электроэнергию сохранится, хотя и замедлит темпы за счёт шоков 2020–2022 годов, повлекших за собой существенный рост цен на генерацию электричества, и будет удовлетворяться, в основном, за счёт возобновляемых источников энергии. Такая ситуация может привести к большей энергетической независимости стран, которые способны наладить выработку электроэнергии из возобновляемых источников, а значит к уменьшению межстранового импорта.

Если же говорить о торговле электроэнергией на страновом уровне, то глобального рынка, в действительности, не существует, что обусловлено технологическими особенностями доставки электричества – значительными потерями при передаче на большие расстояния, и отсутствием торговых площадок и общемировых договорённостей.

В торговле электроэнергией значительную роль играют экономические союзы и международные соглашения. Например, в Северной Америке большое значение в межгосударственной торговле электроэнергией играет соглашение NAFTA – за счёт него обеспечивается взаимодействие между Мексикой, США и Канадой [182]. Подобные соглашения существуют на территории Южной Азии (SAFTA), Юго-Восточной Азии (ASEAN), Евросоюза (The European Commission's Target Electricity Model или TEM), СНГ (ЕАЭС), Скандинавии (Nordic council). В Южной Америке с 1994 года существует торговая организация MERCOSUR, хотя на 2003 год отмечается [182] отсутствие эффективности соглашения и незначительные возможности передачи электроэнергии.

Важной частью при организации рынка электричества является централизованная энергетическая сеть [183]. Например, предполагается, что TEM может увеличить объёмы торговли с 10 до 15% от общего спроса внутри ЕС и сэкономить до полутора миллиардов евро в год, что оправдывает затраты на перестройку сетей [184]. Подобным образом оценивается возможное объединение сетей США и Канады [185].

Во многих работах упоминается положительное влияние роста энергоэффективности сетей, как следствие их объединения, на долю используемой возобновляемой энергии (в частности, в Европе [184] и Азии [186]).

В странах с наименее развитой экономикой описана взаимосвязь между импортом и количеством домохозяйств, имеющих доступ к энергосети. Налаживание импорта электроэнергии в такой ситуации будет стимулировать подключение домохозяйств к общей энергосети, что окажет положительное влияние на экологическое состояние окружающей среды и стабильность обеспечения электроэнергией [187].

Еще одним фактором, оказывающим влияние на импорт и экспорт электроэнергии, являются особенности её производства и распределения внутри страны. Многие авторы называют олигополистические рынки электричества (в США, Европе, скандинавских странах) и отсутствие централизованной инфраструктуры главными причинами неспособности адаптироваться к продаже и покупке электричества из соседних стран, вследствие чего развитие международной торговли останавливается [184] или откладывается [185].

### 3.4.3 Описание и предобработка данных

#### 3.4.3.1 Источники данных

Для проводимого анализа были выбраны данные по объёмам годового взаимного странового импорта электроэнергии (код классификации по Гармонизированной системе – HS2716) за 2000-2020 гг. с интервалами в 5 лет, выраженные в долларах США; с портала

Всемирной торговой организации [188] по статистике Всемирной таможенной организацией; всего было представлено 512 наблюдений.

Для проведения более точного анализа были удалены данные об экспорте/импорте, не превышающие 50 тыс. долларов США, после чего сохранилось 417 наблюдений. Помимо этого, для расчётов показателей с портала Всемирного банка [189] были загружены данные по реальному ВВП исследуемых стран.

В силу особенностей полученных данных в разные периоды встречается пропуски в торговле между двумя странами. Это может быть обусловлено как нулевым импортом, так и изменением статуса государства, например, вступлением в экономический союз, с прекращением передачи данных о конкретной стране. Тем не менее, такая особенность данных не влияет на общие результаты исследования, что объясняется незначительностью доли стран с пропусками информации в общем объеме импорта/экспорта электроэнергии. При этом страны, вступившие в экономический союз, учитывались в составе этого союза.

Пропуски в данных были заполнены с помощью статистики, предоставленной ОЕС [190] (на портале размещена информация из базы данных СЕРП) – от общего количества наблюдений добавленные данные составили 24%.

Для расчёта показателей импорта электроэнергии на душу населения доля импорта в ВВП и потребления энергии на душу населения были взяты данные по потреблению [191] и производству [192] электроэнергии, а также статистика по ВВП на душу населения и численность населения [193].

Были рассчитаны показатели ВВП на душу населения, суммарного импорта на душу населения, суммарного импорта на реальный ВВП и потребления электроэнергии на душу населения. Анализ паттернов проводился для 81-го объекта за 5 временных периодов.

### 3.4.3.2 Проверка данных на достоверность

Недостоверными могли считаться наблюдения, которые значительно отклоняются от значений по производству и потреблению электричества в предшествующие и последующие периоды и не могут объясняться циклическими факторами. Такие данные выделялись по трём параметрам: падение импорта электроэнергии более чем на 30% за пятилетний период, рост ВВП за то же время и отсутствие увеличения производства электроэнергии за соответствующий временной промежуток.

Таким образом, было выделено 12 наблюдений, соответствующих вышеупомянутым критериям. Для каждого из них можно найти объяснение: Албания 2015, 2020 (рост ВВП может объясняться увеличением производства нефтепродуктов), Канада 2005 (можно объяснить общей незначительностью импорта для Канады на фоне больших производственных

мощностей электричества внутри страны), Европейский союз 2010 (может объясняться увеличением производства электричества внутри ЕС, и упадком экономики и производства после кризиса 2008 года), Молдова 2010, 2015 (упадок промышленности, случившийся в 2009 и 2015 гг., мог существенно снизить спрос на электроэнергию, хотя и не повлиять на общий тренд роста ВВП при измерении раз в 5 лет), Черногория 2015 (в период 2014-2015 наблюдалось существенное падение ВВП – при увеличении к 2010 году, чем и может объясняться падение импорта и производства электроэнергии), Нигерия 2020 (импорт за год равен нулю, что можно объяснить как отсутствием данных, так и отсутствием импорта на фоне экономического упадка после 2014 года или возможностью Нигерии вырабатывать достаточное количество энергии с помощью запасов нефти), Дания 2015, Финляндия 2015, Люксембург 2015, Австрия 2015 (у всех вышеупомянутых европейских стран наблюдается значительное падение показателей ВВП в период 2014-2015, причины которого могли также отрицательно повлиять на потребление и импорт электроэнергии, хотя за период 2010-2015 незначительный рост ВВП существует).

#### 3.4.3.3 Корреляционный анализ данных

Наблюдается сильная корреляционная связь (в среднем за рассматриваемые периоды – 0,93) между объёмами производства электроэнергии и ВВП (см. Таблицу 3.6), при этом значение коэффициента корреляции убывает с 2000 по 2020. Однако корреляция между ВВП на душу населения и производством электроэнергии значительно слабее (около 0,3), аналогично можно наблюдать снижение за рассматриваемый период.

Также существует существенная корреляция между показателями ВВП и потребления электричества на душу населения (в среднем – 0,81), что может объясняться связью уровня жизни и уровнем потребления электричества домохозяйствами.

Наблюдается значительная корреляционная связь производства и потребления электроэнергии – за все рассматриваемые периоды она варьируется в интервале от 0,99 до 1, что говорит о почти линейной прямой взаимосвязи. Исходя из этого, можно сделать вывод о незначительности импорта электроэнергии в общей структуре энергоснабжения стран. Такая связь может быть объяснена нецелесообразностью покупки электричества для регионов страны, удалённых от импортёра, нежеланием зависеть от других стран в области энергетики и отсутствием необходимости закупки в целом.

Таблица 3.6 – Корреляционные связи между данными

Показатели / коэффициент корреляции	2000	2005	2010	2015	2020
Производство электроэнергии и ВВП	0,981	0,96	0,93	0,906	0,885

Продолжение таблицы 3.6

ВВП и потребление на душу населения	0,829	0,835	0,794	0,808	0,784
Производство и потребление электроэнергии	0,998	0,99	0,99	0,99	0,99
ВВП на душу населения и производство электроэнергии	0,344	0,298	0,232	0,192	0,185

#### 3.4.4 Описание метода и алгоритма

Чтобы изучить изменения в поведении стран на протяжении рассматриваемого периода времени, а также выделить внутри них сообщества, характеризующиеся похожим поведением, были использованы методы паттерн-анализа [194]. Данный подход предполагает представление данных с использованием параллельных координат: по оси абсцисс откладываются исследуемые показатели, а по оси ординат их значения. Таким образом, каждая страна в определённом периоде представляет собой вектор из значений исследуемых показателей. Такой подход делает возможным не только сравнительный анализ конкретной страны в разные периоды времени, но и анализ её схожести с другими странами с последующим выделением сообществ-кластеров.

В данном исследовании были выбраны следующие показатели для анализа: ВВП на душу населения, суммарный импорт электроэнергии на душу населения, доля суммарного импорта электроэнергии в ВВП и потребление электроэнергии на душу населения.

Стоит отметить, что при изучении паттернов рассматривались не абсолютные значения показателей, а их соотношение, поскольку более важной целью является выделение тех стран, которые характеризуются схожим набором параметров в относительном смысле, а не в абсолютном, что в свою очередь позволяет не учитывать разницу в благосостоянии стран. Таким образом, все значения показателей были предварительно приведены к шкале от 0 до 1 при помощи процедуры минимаксной нормализации. Формула для расчета нормированных значений представлена ниже:

$$X_i^* = \frac{X_i - X_{min}}{X_{max} - X_{min}}, \quad (3.12)$$

где  $X_i^*$  – нормированное значение,  $X_i$  – изначальное значение,  $X_{min,max}$  – минимальное и максимальное значения выборки

Первым этапом паттерн-анализа является выделение кластеров, которые характеризуются похожим соотношением рассматриваемых параметров. На данном этапе каждая из

рассматриваемых стран в определённом периоде выступает как отдельный объект, например, «Российская Федерация 2000» и «Российская Федерация 2005» выступают двумя отдельными объектами анализа. Такой подход позволяет в дальнейшем оценить частоту, с которой тот или иной объект меняет свой паттерн.

Для решения задач подобного рода существует множество различных алгоритмов, такие как метод k-ближайших соседей или DBScan, однако для задач данного исследования они оказались недостаточно эффективными, поскольку приводили к возникновению слишком обширных кластеров, содержащих множество разнородных элементов или же, наоборот, слишком большого числа кластеров, содержащих малое количество элементов. Для решения этой задачи был использован следующий алгоритм туннельной кластеризации:

Для каждого объекта составим вектор его тангенсов  $v$ , считая расстояние между параллельными координатами равным 1:

$$v \in \mathbb{R}^{k-1}; v_i = y_{t+1} - y_i. \quad (3.13)$$

Отсюда

$$\operatorname{tg} \alpha_i = \frac{y_{i+1} - y_i}{x_{i+1} - x_i} = y_{t+1} - y_i. \quad (3.14)$$

Алгоритм:

1. Вектор  $v'$  входит в кластер  $K(v)$ , если для случайного вектора  $v$  верно

$$v' \in K(v) \Leftrightarrow \forall i |v'_i - v_i| < \epsilon, \quad (3.15)$$

где величина  $\epsilon$  рассчитывается как:

$$\epsilon_i = \max(Y_i) + 0,2. \quad (3.16)$$

2. Алгоритм запускается  $t$  раз, в результате чего получается набор разбиений  $\Omega = (\Omega_j)_{1 \leq j \leq t}$ . В каждом наборе  $m_j$  кластеров  $\Omega_j = (K_k)_{1 \leq k \leq m_j}$ . В каждом кластере  $n_j$  векторов  $K_k = (v_l)_{1 \leq l \leq n_j}$ . Пусть  $p_k$  — паттерн кластера  $K_k$ , который вычисляется по формуле

$$p_{ki} = \frac{\sum_{v \in K_k} v_i}{n_k}. \quad (3.17)$$

3. Выбирается разбиение  $\Omega_j$ , максимизирующее разницу между функционалами  $f_1$  и  $f_2$ :

$$f_1(\Omega_j) = \frac{\sum_{1 \leq k_1, k_2 \leq m_j} (\rho(p_{k_1}, p_{k_2}))}{m_j^2}, \quad f_2(\Omega_j) = \frac{\sum_{1 \leq k \leq m_j} \frac{1}{n_k} \sum_{v \in K_k} \rho(p_{k_1}, p_{k_2})}{m_j}, \quad (3.18)$$

где

$$\rho(x, y) = \sum_i (x_i - y_i)^2. \quad (3.19)$$

Второй этап данного анализа предполагает анализ паттернов в динамике, который исходит из идеи, что структура параметров некоторых стран будет неизменна во времени. Таким образом, все страны оказываются разделены на несколько групп в зависимости от уровня их устойчивости [195]. Будем выделять группы абсолютно устойчивых, устойчивых, частично устойчивых, неустойчивых и абсолютно неустойчивых, при количестве изменений паттерна 0, 1, 2, 3 и 4 раза соответственно.

### 3.4.5 Результаты и дискуссия

Анализ паттернов был проведён для 80-и стран и Европейского союза в период с 2000 по 2020 гг. с интервалами в 5 лет, что привело к формированию 67-и кластеров.

Количество наблюдений, входящих в один кластер, варьируется от 1 до 45, со средним показателем равным 5,9. При этом в 10 крупнейших кластерах находятся 57% наблюдений, в 12–61%, а в 18–72%.

Кластер 1: China 2000, Colombia 2000, Ecuador 2000, Egypt 2000, Jordan 2000, North Macedonia 2000, Peru 2000, Thailand 2000, Zambia 2000, Lithuania 2000, China 2005, Colombia 2005, Egypt 2005, Guatemala 2005, Kyrgyz Republic 2005, Peru 2005, Zambia 2005, Armenia 2010, Colombia 2010, Egypt 2010, Eswatini 2010, India 2010, Indonesia 2010, Kyrgyz Republic 2010, Moldova. Republic of 2010, Peru 2010, Samoa 2010, Zambia 2010, Colombia 2015, Egypt 2015, India 2015, Indonesia 2015, Moldova. Republic of 2015, Peru 2015, Samoa 2015, Ecuador 2020, Egypt 2020, India 2020, Indonesia 2020, Kyrgyz Republic 2020, Morocco 2020, Pakistan 2020, Peru 2020, Samoa 2020, Zambia 2020.

Результаты визуализации этого кластера представлены на рисунке 3.10.

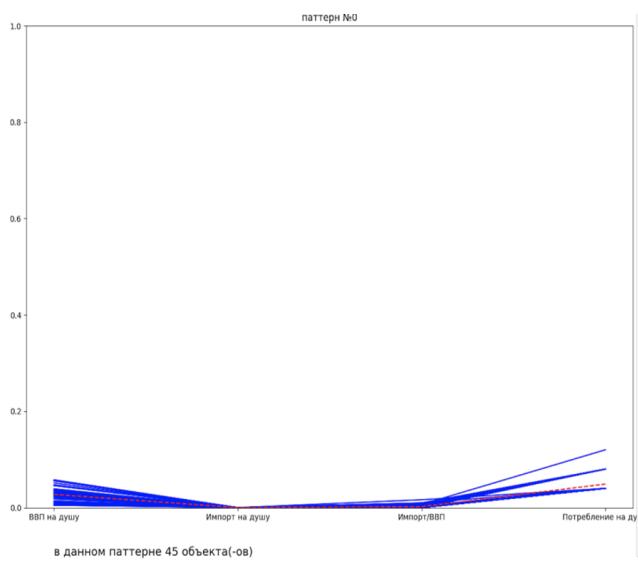


Рисунок 3.10 – Визуализация 1-го кластера (паттерн №0)

Это самый большой кластер, состоящий из 45 объектов, объединяет в себе страны с низкими значениями показателей (не более 0,1). Стоит заметить, что в данном кластере преобладают страны с развивающейся экономикой – Китай в 2000 и 2005, Индия в 2010–2020, Индонезия, Замбия, Уганда, Пакистан, Перу и другие.

Кластер 2: Burkina Faso 2000, Côte d'Ivoire 2000, Guatemala 2000, India 2000, Indonesia 2000, Lesotho 2000, Myanmar 2000, Nigeria 2000, Pakistan 2000, Samoa 2000, Uganda 2000, Burkina Faso 2005, Côte d'Ivoire 2005, India 2005, Indonesia 2005, Lesotho 2005, Myanmar 2005, Nigeria 2005, Pakistan 2005, Samoa 2005, Uganda 2005, Burkina Faso 2010, Côte d'Ivoire 2010, Myanmar 2010, Nigeria 2010, Pakistan 2010, Côte d'Ivoire 2015, Nigeria 2015, Pakistan 2015, Côte d'Ivoire 2020, Myanmar 2020, Nigeria 2020, Uganda 2020.

Результаты визуализации этого кластера представлены на рисунке 3.11.

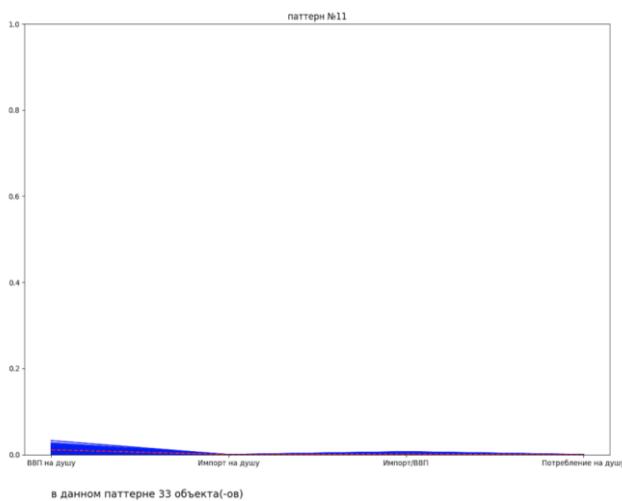


Рисунок 3.11 – Визуализация 2-го кластера (паттерн №11)

В данном кластере также можно наблюдать страны со всеми показателями меньше 0,1, хотя от рассмотренного выше его отличает ещё более низкий уровень потребления электричества на душу населения. В него входят бедные страны (Кот-д'Ивуар, Мьянма, Нигерия, Уганда и др.)

Кластер 3: Malaysia 2000, Mexico 2000, Poland 2000, Serbia 2000, South Africa 2000, Ukraine 2000, Malaysia 2005, Mexico 2005, Serbia 2005, Turkey 2005, Ukraine 2005, Brazil 2010, China 2010, Malaysia 2010, Mexico 2010, Turkey 2010, Ukraine 2010, Brazil 2015, China 2015, Malaysia 2015, Mexico 2015, Malaysia 2020, Mexico 2020.

Результаты визуализации этого кластера представлены на рисунке 3.12.

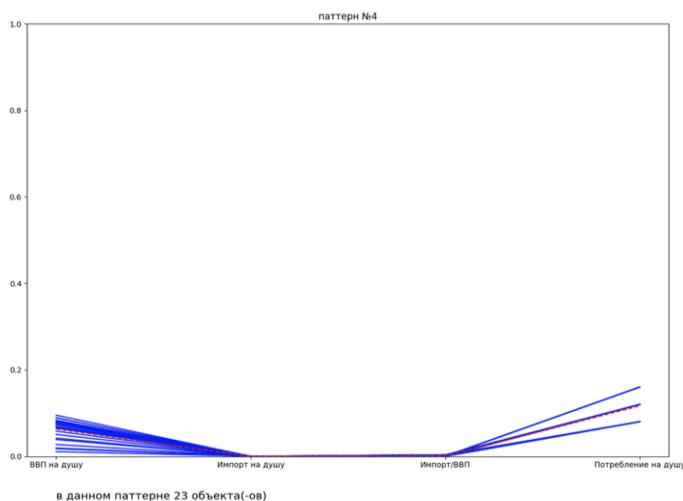


Рисунок 3.12 – Визуализация 3-го кластера (паттерн №4)

Третий по размеру кластер схож с вышеописанными, хотя и отличается относительно большим ВВП на душу и потреблением. В него входят развивающиеся страны, такие как Мексика, Турция, Украина и другие.

Вышеописанные кластеры объединяют в себе развивающиеся страны со средним уровнем дохода и относительно небольшим потреблением электричества на душу населения, способные при этом производить её в достаточном для внутреннего потребления объеме. Следовательно, данные страны незначительно зависят от импорта электроэнергии.

Кластер 4: Belgium 2000, Denmark 2000, Austria 2000, Hong Kong. China 2005, Macao. China 2005, Belgium 2005, Denmark 2005, France 2005, Germany 2005, Ireland 2005, Italy 2005, Netherlands 2005, Austria 2005, Hong Kong. China 2010, Belgium 2010, Denmark 2010, Germany 2010, Netherlands 2010, Austria 2010, Hong Kong. China 2015, United Kingdom 2015, Belgium 2015, Denmark 2015, Germany 2015, Netherlands 2015, Austria 2015, Hong Kong. China 2020, United Kingdom 2020, Belgium 2020, France 2020, Germany 2020, Netherlands 2020, Austria 2020.

Результаты визуализации этого кластера представлены на рисунке 3.13.

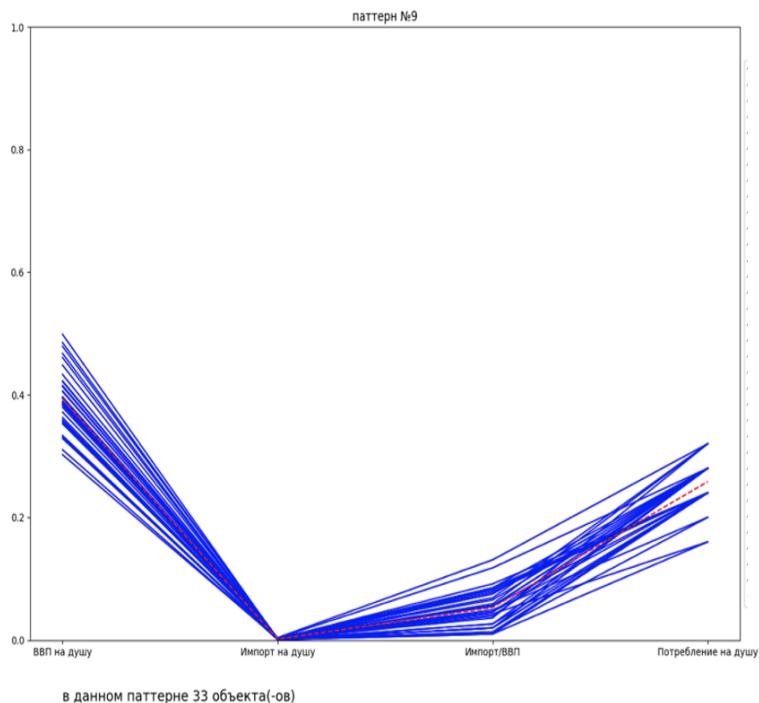


Рисунок 3.13 – Визуализация 4-го кластера (паттерн №9)

Отдельным паттерном поведения характеризуются преимущественно развитые европейские страны – они обладают достаточно высоким ВВП и потреблением на душу населения, но при этом низкими показателями импорта электроэнергии на душу населения и импорта электроэнергии на ВВП. В основном это крупные промышленные государства с развитым производственным сектором, и, в большей степени, самостоятельно удовлетворяющие спрос на электроэнергию. К ним относятся Бельгия, Австрия, Гонконг, Германия и другие.

Кластер 5. В этот кластер (паттерн №3) входит 21 страна: Estonia 2000, Slovak Republic 2000, Slovenia 2000, Portugal 2000, Chile 2005, Estonia 2005, Poland 2005, South Africa 2005, Lithuania 2005, Poland 2010, South Africa 2010, Thailand 2010, Argentina 2015, Kazakhstan 2015, Poland 2015, Turkey 2015, Bulgaria 2015, Brazil 2020, South Africa 2020, Ukraine 2020, Uruguay 2020.

Кластер 6. В этот кластер (паттерн №64) входит 22 страны: Albania 2000, Armenia 2000, Eswatini 2000, Mongolia 2000, Morocco 2000, Turkey 2000, Zimbabwe 2000, Georgia 2005, Mongolia 2005, Ecuador 2010, Jordan 2010, Mongolia 2010, Zimbabwe 2010, Ecuador 2015, Jordan 2015, Morocco 2015, Niger 2015, Thailand 2015, Zimbabwe 2015, Albania 2020, Armenia 2020, Colombia 2020.

Модель поведения стран из кластеров 5 и 6 схожа с описанными выше – низкий показатель ВВП на душу населения ( $<0,2$ ), относительно низкие показатели импорта на ВВП и потребления электроэнергии на душу населения ( $<0,4$ ) и крайне высокий показатель импорта электроэнергии на душу населения (в среднем  $>0,8$ ). К этим странам относятся развивающиеся страны с относительно высоким уровнем развития экономики, в том числе, некоторые европейские – Польша в 2005–2015, Болгария, Литва и другие. Можно предположить, что это страны без больших производственных энергетических мощностей. Разделение похожих объектов на разные кластеры может объясняться разной политикой закупки электроэнергии или разным уровнем избыточного спроса – в первом кластере показатель импорта электроэнергии на ВВП немного меньше, чем во втором, а также разным уровнем потребления (ко второму кластеру относятся страны с более жарким климатом и менее развитые).

Кластер 7. В этот кластер (паттерн №17) входит 14 стран: Brazil 2000, Kazakhstan 2000, Bulgaria 2000, Albania 2005, Argentina 2005, Brazil 2005, Jordan 2005, Morocco 2005, El Salvador 2010, Armenia 2015, Georgia 2015, Guatemala 2020, Jordan 2020, Moldova 2020. Он состоит из государств с показателями, схожими с вышеописанными, но с меньшим уровнем импорта электроэнергии на душу населения. Некоторые страны (Аргентина, Бразилия) оказываются в этих кластерах попеременно, в зависимости от года, что является дополнительным аргументом в пользу их связи.

Отдельный кластер из 12 стран (паттерн №1: Albania 2010, North Macedonia 2010, Lithuania 2010, Croatia 2015, Eswatini 2015, Latvia 2015, Mongolia 2015, Montenegro 2015, North Macedonia 2015, Croatia 2020, Mongolia 2020, Serbia 2020.) образуют маленькие по территории и экономике страны, в которых показатель импорта электроэнергии на ВВП крайне высок, при низких уровнях потребления, ВВП и импорта электроэнергии на душу населения. К нему относятся Сербия, Ботсвана, Северная Македония, Черногория и другие.

Помимо этого, выделяются страны с потреблением электроэнергии и ВВП на душу населения приблизительно равными 0,2, близким к нулю импортом электроэнергии на душу населения. При этом они подразделяются на два кластера: паттерн №18 – Cyprus 2000, Greece 2000, Malta 2000, Czech Republic 2005, Greece 2005, Malta 2005, Malta 2010, Portugal 2010, Chile 2015, Uruguay 2015, Greece 2015, Portugal 2015, Chile 2020, Türkiye 2020, Portugal 2020; паттерн №19 – Croatia 2005, Croatia 2010, Hungary 2010, Latvia 2010, Slovak Republic 2010, Hungary 2015, Slovak Republic 2015, Hungary 2020, Lithuania 2020, Slovenia 2020. Данное разделение обусловлено различными показателями импорта электроэнергии на ВВП – приблизительно 0,1 и 0,3, соответственно. В оба кластера входят небольшие, с точки зрения экономики, государства, при этом №19 образуется из европейских стран, а в №18 добавляются некоторые африканские и латиноамериканские страны.

Вышеописанные кластеры (паттерны № 3, 64, 17, 1, 18, 19) содержат в себе страны с развивающейся экономикой с относительно высокими показателями импорта электроэнергии. Можно предположить, что производственные мощности данных государств не позволяют им полностью обеспечивать внутренний спрос (в отличие от паттернов № 0, 11, 4, 9).

Выделяются в качестве отдельного кластера (паттерн №15: Canada 2000, Finland 2000, Sweden 2000, Finland 2005, Canada 2010, United Arab Emirates 2010, Finland 2015, United Arab Emirates 2020.) северные европейские страны, Канада за некоторые годы и ОАЭ, характеризующиеся высоким ВВП на душу (0,4–0,5) и потребления и низкими показателями импорта электроэнергии (<0,1). Такое выделение можно объяснить повышенным потреблением электроэнергии, обусловленное климатическими особенностями в ОАЭ и в Канаде.

Достаточно велико количество кластеров, состоящих из одной страны. В основном в них представлены развитые страны с высокими показателями ВВП и потребления электричества на душу населения. Это объясняется отличием этих стран в показателях электроэнергии от их соседей, а также высоким уровнем стабильности в вопросах закупки и потребления электричества. В такие кластеры были отнесены Швейцария, Европейский союз, Черногория, Люксембург, Норвегия.

Говоря в целом о паттернах поведения стран в потреблении, производстве и закупке электроэнергии, можно сделать вывод о существовании некоторой зависимости между принадлежностью к паттерну и уровнем экономического развития страны, её площадью, а также географическим положением.

Для лучшего понимания структуры изменений в паттернах поведения, был проведён динамический анализ паттернов [195], позволяющий проследить изменения показателей объектов на горизонте анализа.

В данном исследовании наибольший интерес представляет оценка стабильности объектов, характеризующая частоту изменения их паттерна, а также выделение определённых закономерностей в выборе паттернов среди специфичных групп стран.

Распределение количества изменений паттерна (см. рисунок 3.14) является весьма неравномерным со средним значением равным 3,83. Среди рассматриваемых стран нет ни одной абсолютно устойчивой, а большая часть из них, наоборот, относится к неустойчивым и абсолютно неустойчивым типам. Единственной страной, сменившей паттерн за весь рассматриваемый период лишь два раза, является Казахстан (17-й паттерн в 2000–2005 гг., 0-й паттерн в 2005–2010 гг. и 19-й в 2020 году). Если же рассматривать неустойчивые объекты, то к этой группе относятся страны со средним уровнем дохода, а именно Молдова, Сербия, Киргизия и другие. В группу абсолютно неустойчивых входят страны с различным уровнем экономического развития и импорта в сфере электроэнергетики, что затрудняет выделение специфической группы стран.

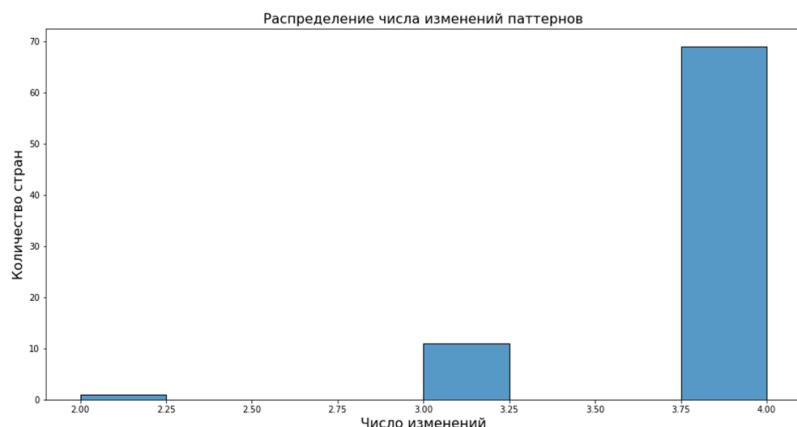


Рисунок 3.14 – Распределение числа изменений паттерна

Если же рассматривать динамику доли стран, изменивших свой паттерн (рисунок 3.15), то можно заметить, что она достигает максимума в 0,99 в 2010 году и снижается далее, достигая минимума (менее 0,93) в 2020 году.

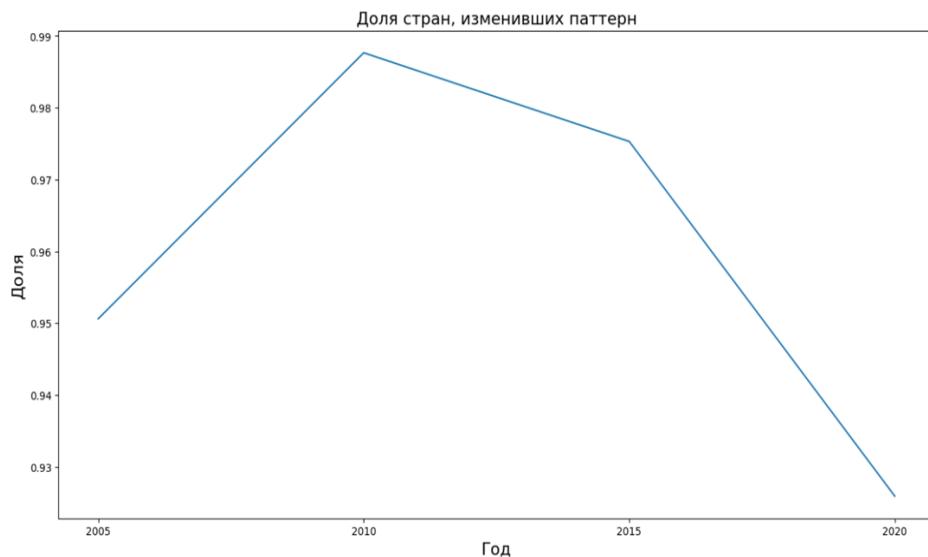


Рисунок 3.15 – Доля стран, изменивших паттерн

Такая динамика может быть объяснена некоторыми факторами: в 2010 году высокая доля изменений паттерна обусловлена снижением значений всех параметров, связанным с мировым экономическим кризисом и необходимостью адаптации стран к новым условиям. Дальнейшее снижение доли стран, изменивших паттерн к рассматриваемому году, можно объяснить как относительной макроэкономической стабильностью, так и развивающейся интеграцией стран в сфере электроэнергетики (например, появление и развитие концепции общей энергосистемы ЕС).

В целом, на исследуемом интервале в 20 лет паттерны поведения стран в показателях электроэнергии обладают низкой устойчивостью. Любые изменения могут быть объяснены как сменой политики государств в отношении электроэнергии, качественным изменением экономики данной страны, так и меняющимися условиями рынка электроэнергии.

Тем не менее, неустойчивость в странах с развитой или стремительно развивающейся экономикой может быть обусловлена не только экзогенными факторами, сколько приверженностью государственных властей к политике, направленной на обеспечение энергетической безопасности, снижения уровня выбросов и других. В странах с высокими темпами роста смена паттернов часто обусловлена трансформацией экономики и растущим спросом на электроэнергию. Подробнее динамика смены паттернов рассмотрена на примере Европейского союза, Китая и США.

Из диаграммы на рисунке 3.16 видно, что Европейский союз является весьма устойчивым в таких показателях, как ВВП на душу населения, доля импорта электроэнергии в ВВП, потребление на душу населения, в отличие от импорта электроэнергии на душу

населения, который устойчиво растет на всем рассматриваемом периоде (исключением является 2005–2010 гг.). Объяснением этого может являться стратегия государств-членов ЕС, направленная на снижение уровня выбросов CO<sub>2</sub> и активно реализуемая в течение последнего десятилетия. Общая выработка электроэнергии в 27 странах-членах ЕС в период с 2011 по 2020 год снизилась на 300 ТВт/ч [196], что связано с постепенным отказом от газовой и атомной выработки электроэнергии и переходу к возобновляемым источникам энергии. Тем не менее, традиционные источники электричества временно не могут быть полностью заменены, что приводит к росту импорта электроэнергии ЕС из стран-соседей: Албания, Белоруссия, Украина, Сербия и др.

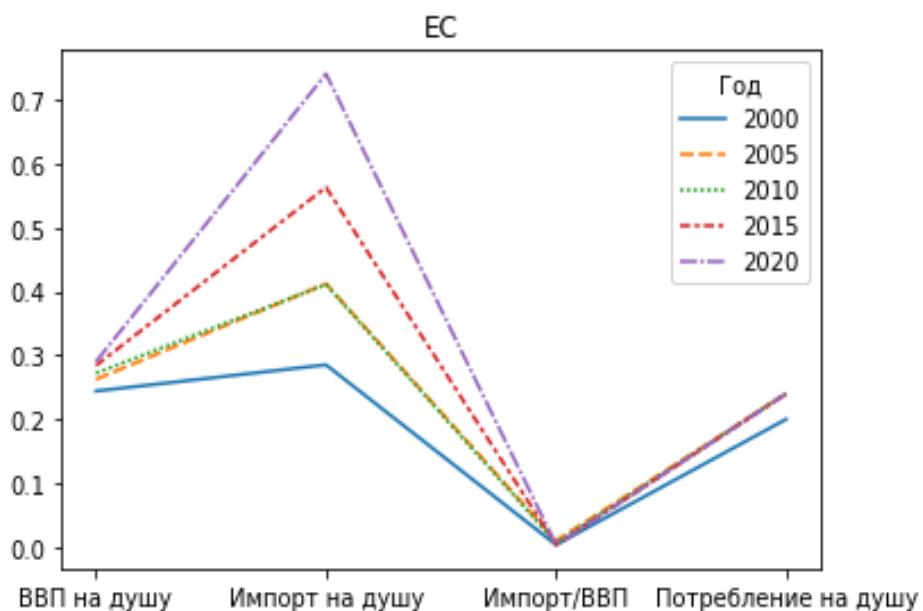


Рисунок 3.16 – Изменение паттерна ЕС

Учитывая сложившуюся неопределенность, которая выражается в снижении возможности импорта электроэнергии из некоторых стран-соседей, рост цен на газ, а также невозможностью достижения углеродной нейтральности к 2035 году, стоит ожидать, что в будущем рост импорта электроэнергии в ЕС может сохраниться, особенно в Германии, наиболее стремительно отказывающейся от выработки из ископаемых источников. Кроме того, доля угольной выработки может продолжить расти в случае сохранения высоких цен на газ (например, в Польше и Чехии).

В случае с Китаем (рисунок 3.17) основной причиной смены паттерна является устойчивый рост доли импорта электроэнергии в ВВП. Такая динамика связана со стремительным экономическим развитием КНР, создающим значительный спрос на электроэнергию для промышленного сектора, который не может быть удовлетворён локальными мощ-

ностями. Основным экспортером электроэнергии в КНР является Россия, некоторое значение для импорта также имеют Мьянма и Гонконг. Кроме того, рост импорта косвенно обусловлен политикой отказа от угольной выработки. При этом стоит отметить, что значения импорта и потребления электроэнергии на душу населения меняются незначительно на всём рассматриваемом временном интервале, что обусловлено высокими темпами роста населения, компенсирующими возрастающие объёмы импорта. Таким образом, Китай всё ещё относится к странам с достаточно низким уровнем энергообеспеченности домохозяйств.

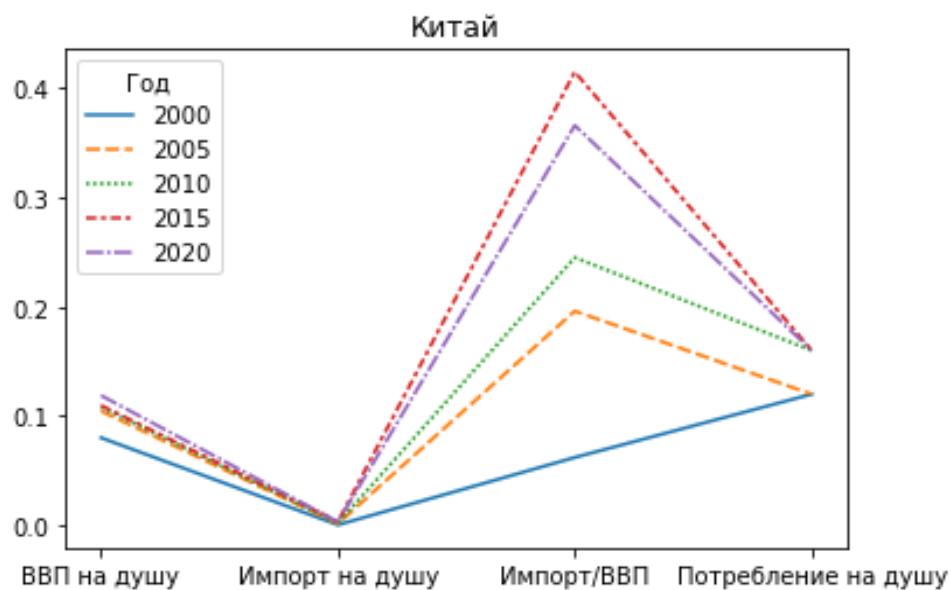


Рисунок 3.17 – Изменение паттерна Китая

Как и в ЕС, основной причиной смены паттерна в США (рисунок 3.18) является изменение объёмов импорта электроэнергии на душу населения, однако, это обусловлено несколько иными причинами. Несмотря на то, что США также придерживается политики по снижению уровня выбросов, основным причиной для наращивания импорта скорее является изменение стоимости выработки на локальном рынке и возможность её замещения более дешёвой электроэнергией из Мексики и Канады. Это предположение подтверждается данными: рост импорта на душу населения наблюдается в периоды высоких цен на электроэнергию.

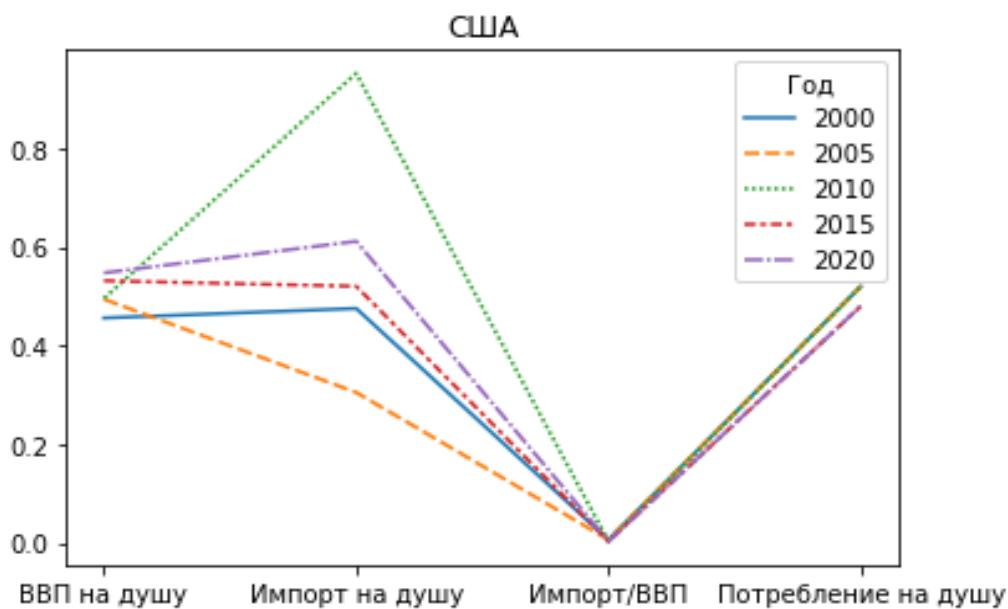


Рисунок 3.18 – Изменение паттерна США

Таким образом, можно сделать вывод, что в развитых странах основной причиной неустойчивости являются колебания импорта электроэнергии на душу населения, обусловленные проводимой политикой или рыночными условиями, в то время как в развивающихся странах больше подвержена колебаниям доля импорта в ВВП, что обусловлено необходимостью наращивания импорта в условиях избыточного спроса.

#### 3.4.6 Заключение

Подводя итог, стоит сказать, что вопросы, связанные с электроэнергетикой, являются весьма актуальными для большинства стран, что обусловлено не только возрастающей озабоченностью вопросами экологии, но и стремлением обеспечить высокую степень энергетического суверенитета. Несмотря на стремительное развитие технологий, позволяющее создавать инфраструктуру для транспортировки электроэнергии на более протяженные расстояния, процесс создания глобального рынка находится только лишь на ранних стадиях.

В данной работе были применены методы паттерн-анализа, позволяющие отметить наиболее перспективные направления развития отрасли, актуальные тенденции, а также регионы и страны, сталкивающиеся с определенными проблемами в области электроэнергии. Были сделаны выводы о группах стран, различающихся по уровню экономического развития и его связи с показателями экспорт-импорта электроэнергии, об изменениях их уровней в крупнейших странах.

Таким образом, полученные результаты способствуют углублению понимания глобального рынка электроэнергии и могут внести вклад в решение актуальных задач отраслей бизнеса и экономического развития.

## **4 Исследование непрерывных моделей**

В этом году исследование непрерывных моделей осуществлялось в двух направлениях:

- разработка алгебраического метода решения линейных разностных стационарных уравнений высшего порядка;
- решение задачи о локальной гладкой управляемости для некоторой распределенной системы четвёртого порядка.

### **4.1 Разработка алгебраического метода решения линейных разностных стационарных уравнений высшего порядка**

#### **4.1.1 Введение**

Идея представления различных математических процессов как операторов не нова. Представлению анализа, дифференцирования и интегрирования в таком виде посвящена обширная литература, см., к примеру, [197, 198, 199]. В них авторы впервые предложили матрично-дифференциально-операторный метод нахождения частного решения определённых дифференциальных уравнений, который, по своей сути, является модификацией метода неопределённых коэффициентов. Предложенный метод позволяет получать систему линейных алгебраических уравнений относительно коэффициентов разложения квазиполиномиальной правой части уравнения без фактического использования операции дифференцирования.

Конечно, можно возразить, что метод неопределённых коэффициентов достаточно хорошо описан во всех классических учебниках по дифференциальным уравнениям (см., к примеру, [200, 201, 202, 203, 204, 205]) и поднять вопрос о целесообразности его модификаций. На это возражение можно ответить, что серьёзным недостатком данного метода является его трудоёмкость и невозможность использования на контрольных мероприятиях самых сложных и интересных задач, решаемых этим методом, а именно тех, в которых возникает комплексный резонанс.

Однако на некоторых специальностях помимо дифференциальных уравнений также изучаются и разностные уравнения. Тем не менее разностные уравнения, по большей части, оставлены в стороне, но теория разностных уравнений применима во многих областях, которые требуют моделирования различных систем. Самой известной из таких моделей, конечно, является паутинообразная модель рынка. Паутинообразная модель со складом, например, описывается разностным уравнением второго порядка. Более того, применение

метода неопределённых коэффициентов к разностным уравнениям куда затратнее по времени, чем в дифференциальном случае, см., например, [206] потому, что в комплексном случае нам приходится рассматривать выражения вида  $\sin \alpha(n+k)$  и  $\cos \alpha(n+k)$ , что существенно сокращает объём материала, который можно преподнести студентам за выделенные часы. В связи с этим, даже в [206] среди задач отсутствуют примеры таких разностных уравнений, в которых комплексные собственные числа оператора в левой части уравнения имеют достаточно широкий набор аргументов.

В данной работе по аналогии с [197, 199] разработана алгебраическая модификация метода неопределенных коэффициентов для решения линейных разностных стационарных уравнений высшего порядка. Представлены фрагменты кода на языке Wolfram Mathematica, благодаря которым разностные уравнения этим методом решаются достаточно быстро.

Важнейшим и неоспоримым преимуществом данного подхода является компактность и читаемость найденного частного решения по сравнению с базовыми возможностями веб-версии Wolfram Alpha. На рисунках 4.1 и 4.2 приведены примеры плохо читаемого решения в данной среде, а также отсутствия такового.

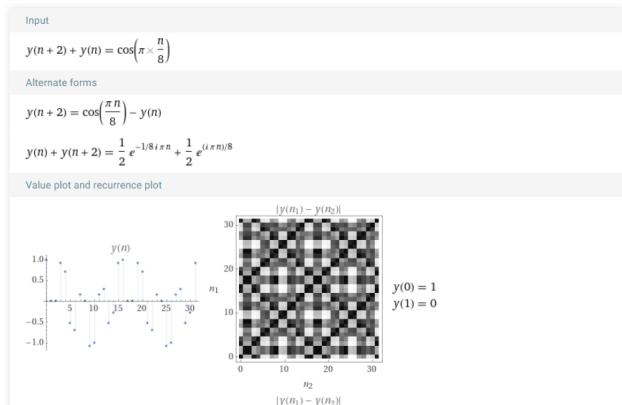


Рисунок 4.1 – Отсутствие точного решения в среде Wolfram Alpha

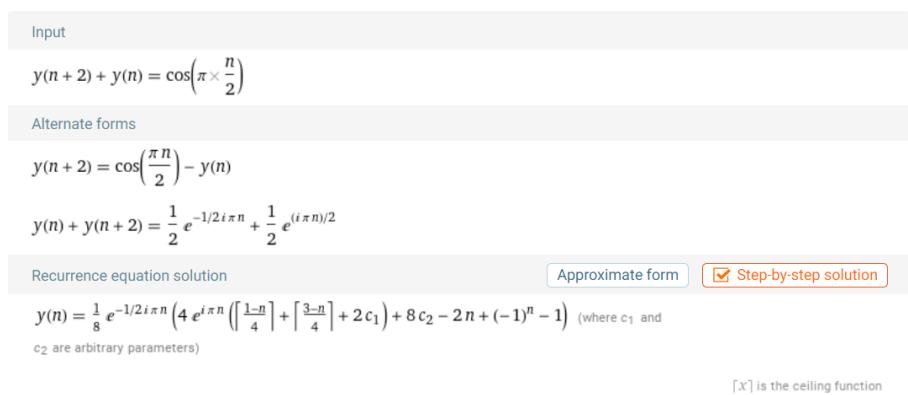


Рисунок 4.2 – Плохо читаемое решение простого разностного уравнения с комплексным резонансом

#### 4.1.2 Алгебраическая модификация метода неопределенных коэффициентов

##### 4.1.2.1 Геометрия простейшего оператора сдвига

Пусть  $N_0$  – расширенный ряд натуральных чисел. Рассмотрим пространство  $C^{N_0}$  всех комплекснозначных числовых последовательностей, индекс которых начинается с нуля. Пусть  $e_0 = (1, 0, \dots, 0)$ . Тогда  $L[e_0]$  является подпространством  $C^{N_0}$ . Теперь положим  $S$  факторпространством  $C^{N_0} / L[e_0]$ .

Рассмотрим теперь простейший оператор сдвига  $T : S \rightarrow S$ , действующий по правилу  $(Ty)_n = y_{n+1}$ . Очевидно, этот оператор является линейным.

Следующие утверждения достаточно легко доказать.

**Утверждение 4.1.** Всякое число  $\lambda \in C \setminus \{0\}$  является собственным числом оператора  $T$ .

**Утверждение 4.2.** Пусть  $\lambda \in C \setminus \{0\}$ . Тогда корневое подпространство оператора  $T$ , которое соответствует собственному числу  $\lambda$ , является множеством  $V_\lambda = L\{\lambda^n, n\lambda^n, n^2\lambda^n, \dots, n^k\lambda^n, \dots\}$ ,

Доказательство утверждения 2 вне теории разностных уравнений можно вывести, например, из [207].

##### 4.1.2.2 Матричное представление простейшего оператора сдвига

Рассмотрим пространство  $L_\mu^s = L\{\mu^n, n\mu^n, \dots, n^s\mu^n\}$ , где  $\mu \in C \setminus \{0\}$  и  $s \in N$ . Также для любого комплексного числа  $\mu = |\mu|e^{i\varphi}$  с ненулевой мнимой частью, рассмотрим пространство

$$L_{\mu, C}^s = L\{|\mu|^n \sin n\varphi, |\mu|^n \cos n\varphi, n|\mu|^n \sin n\varphi, n|\mu|^n \cos n\varphi, \dots, n^s|\mu|^n \sin n\varphi, n^s|\mu|^n \cos n\varphi\}. \quad (4.1)$$

Последнее получено из  $L_\mu^s$  ограничением скаляров на поле вещественных чисел, элементов на вещественные числовые последовательности и сменой базиса.

Доказаны следующие теоремы.

Теорема 4.1. 1. Сужение  $T_\mu^s = T|_{L_\mu^s}$  оператора Т на  $L_\mu^s$  является автоморфизмом пространства  $L_\mu^s$ , а матрица  $T_\mu^s$  оператора сужения в базисе  $\{\mu^n, n\mu^n, \dots, n^s\mu^n\}$  имеет вид

$$T_\mu^s = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & \dots & \begin{pmatrix} s \\ 1 \end{pmatrix} \\ 0 & 0 & 1 & \dots & \begin{pmatrix} s \\ 2 \end{pmatrix} \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}. \quad (4.2)$$

2. В  $L_\mu^s$  существует такой базис, что матрица  $T_\mu^s$  в этом базисе имеет вид

$$T_\mu^s = \mu I_{s+1} + U_{s+1}, \quad (4.3)$$

где  $I_k$  – единичная матрица  $k$ -го порядка и  $U_k$  – верхне-сдвиговая матрица  $k$ -го порядка (т.е.,  $U_{ij} = \delta_{i+1,j}$ , где  $\delta_{ij}$  – дельта-символ Кронекера). Иными словами, жорданова форма матрицы  $T_\mu^s$  состоит из одной клетки.

Теорема 4.2. 1. Сужение  $T_{\mu,C}^s = T|_{L_{\mu,C}^s}$  оператора Т на  $L_{\mu,C}^s$  является автоморфизмом пространства  $L_{\mu,C}^s$ , а матрица  $T_{\mu,C}^s$  оператора сужения в базисе

$$\{|\mu|^n \sin n\varphi, |\mu|^n \cos n\varphi, n|\mu|^n \sin n\varphi, n|\mu|^n \cos n\varphi, \dots, n^s|\mu|^n \sin n\varphi, n^s|\mu|^n \cos n\varphi\} \quad (4.4)$$

имеет вид

$$T_{\mu,C}^s = \begin{pmatrix} J & J & J & J & J \\ 0 & J & 2J & \dots & \binom{s}{1}J \\ 0 & 0 & J & \dots & \binom{s}{2}J \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & J \end{pmatrix}, \quad (4.5)$$

где  $J = \begin{pmatrix} \cos \varphi & -\sin \varphi \\ \sin \varphi & \cos \varphi \end{pmatrix}$ .

2. В  $L_{\mu,C}^s$  существует такой базис, что матрица  $T_{\mu,C}^s$  в этом базисе имеет вид

$$T_{\mu,C}^s = |\mu| I_{s+1} \otimes J + U_{s+1} \otimes I_2, \quad (4.6)$$

где  $\otimes$  – матричное произведение Кронекера.

#### 4.1.2.3 Алгебраическая модификация метода неопределённых коэффициентов

Пусть  $S_R$  – подпространство  $S$ , которое состоит только из вещественных числовых последовательностей и является линейным пространством над полем вещественных чисел. Рассмотрим неоднородное линейное стационарное разностное уравнение

$$y_{n+k} + a^{(1)} y_{n+k-1} + \dots + a^{(k)} y_n = f_n, \quad (4.7)$$

где  $y \in S_R$  – искомая последовательность,  $a^{(1)}, \dots, a^{(k)}$  – заданные числа, причём  $a^{(k)} \neq 0$ , и  $f \in S_R$  – заданная последовательность.

Теорема 4.3. Пусть  $\mu$  является вещественным числом и в (4.1)  $f \in L_\mu^s$ .

1. Если  $\mu$  не является характеристическим числом однородного уравнения (1) с  $f = 0$ , то частное решение неоднородного уравнения (1)  $y^{(0)} \in L_\mu^s$ , а его коэффициенты в базисе  $\{\mu^n, n\mu^n, \dots, n^s\mu^n\}$  можно найти из системы уравнений  $R_\mu^s \mathbf{y} = \mathbf{f}$ , где

$R_\mu^s = \left(T_\mu^s\right)^k + a^{(1)} \left(T_\mu^s\right)^{k-1} + \dots + a^{(k-1)} T_\mu^s + a^{(k)} I$  и  $\mathbf{f}$  – вектор-столбец, полученный разложением правой части  $f$  по заданному базису, причём система имеет единственное решение.

2. Если  $\mu$  является характеристическим числом однородного уравнения (4.1) кратности  $m$ , то частное решение  $y^{(0)} \in L_\mu^{m+s} = K_\mu^m \oplus L_\mu^{m,s}$ , где  $K_\mu^m = L\left\{\mu^n, n\mu^n, \dots, n^{m-1}\mu^n\right\} = \text{Ker } R_\mu^s|_{L_\mu^{m+s}}$  и  $L_\mu^{m,s} = L\left\{n^m\mu^n, n^{m+1}\mu^n, \dots, n^{m+s}\mu^n\right\}$ . В этом случае можно найти проекцию  $y_p$  частного решения  $y^{(0)}$  на  $L_\mu^s$  и коэффициенты этой проекции в базисе из п. 1 из системы уравнений  $R_\mu^{m,s}y_p = \mathbf{f}$ , где  $R_\mu^{m,s}$  – матрица, полученная вычёркиванием первых  $m$  столбцов и  $m$  строк из матрицы  $\left(T_\mu^s\right)^k + a^{(1)} \left(T_\mu^s\right)^{k-1} + \dots + a^{(k-1)} T_\mu^s + a^{(k)} I$ , и  $\mathbf{f}$  – вектор-столбец, полученный разложением правой части  $f \in L_\mu^s$  по базису  $\left\{\mu^n, n\mu^n, \dots, n^s\mu^n\right\}$ , причём система имеет единственное решение.

Теорема 4.4. Пусть  $\mu$  является комплексным числом с ненулевой мнимой частью и в (4.1)  $f \in L_{\mu,C}^s$ .

1. Если  $\mu$  не является характеристическим числом однородного уравнения (1) с  $f = 0$ , то частное решение неоднородного уравнения (1)  $y^{(0)} \in L_{\mu,C}^s$ , а его коэффициенты в базисе  $\left\{|\mu|^n \sin n\varphi, |\mu|^n \cos n\varphi, n|\mu|^n \sin n\varphi, n|\mu|^n \cos n\varphi, \dots, n^s|\mu|^n \sin n\varphi, n^s|\mu|^n \cos n\varphi\right\}$  можно найти из системы уравнений  $R_{\mu,C}^s \mathbf{y} = \mathbf{f}$ , где  $R_{\mu,C}^s = \left(T_{\mu,C}^s\right)^k + a^{(1)} \left(T_{\mu,C}^s\right)^{k-1} + \dots + a^{(k-1)} T_{\mu,C}^s + a^{(k)} I$  и  $\mathbf{f}$  – вектор-столбец, полученный разложением правой части  $f$  по заданному базису, причём система имеет единственное решение.

2. Если  $\mu$  является характеристическим числом однородного уравнения (1) кратности  $m$ , то частное решение  $y^{(0)} \in L_{\mu,C}^{m+s} = K_{\mu,C}^m \oplus L_{\mu,C}^{m,s}$ , где

$$\begin{aligned}
K_{\mu,C}^m &= \\
&= L \left\{ |\mu|^n \sin n\varphi, |\mu|^n \cos n\varphi, n|\mu|^n \sin n\varphi, n|\mu|^n \cos n\varphi, \dots, n^{m-1} |\mu|^n \sin n\varphi, n^{m-1} |\mu|^n \cos n\varphi \right\} = \\
&= \text{Ker } R_{\mu,C}^s |_{L_{\mu,C}^{m+s}}
\end{aligned} \tag{4.8}$$

и

$$\begin{aligned}
L_{\mu}^{m,s} &= L \{ n^m |\mu|^n \sin n\varphi, n^m |\mu|^n \cos n\varphi, n^{m+1} |\mu|^n \sin n\varphi, n^{m+1} |\mu|^n \cos n\varphi, \dots, \\
&\quad n^{m+s} |\mu|^n \sin n\varphi, n^{m+s} |\mu|^n \cos n\varphi \}.
\end{aligned} \tag{4.9}$$

В этом случае можно найти проекцию  $y_p$  частного решения  $y^{(0)}$  на  $L_{\mu,C}^s$  и коэффициенты этой проекции в базисе из п.1 из системы уравнений  $R_{\mu,C}^{m,s} y_p = \mathbf{f}$ , где  $R_{\mu,C}^{m,s}$  – матрица, полученная вычёркиванием первых  $2m$  столбцов и  $2m$  строк из матрицы  $(T_{\mu,C}^s)^k + a^{(1)}(T_{\mu,C}^s)^{k-1} + \dots + a^{(k-1)}T_{\mu,C}^s + a^{(k)}I$ , и  $\mathbf{f}$  – вектор-столбец, полученный разложением правой части  $f \in L_{\mu}^s$  по базису  $\left\{ |\mu|^n \sin n\varphi, |\mu|^n \cos n\varphi, n|\mu|^n \sin n\varphi, n|\mu|^n \cos n\varphi, \dots, n^s |\mu|^n \sin n\varphi, n^s |\mu|^n \cos n\varphi \right\}$ , причём система имеет единственное решение.

#### 4.1.3 Практическая демонстрация метода

В данном подразделе мы продемонстрируем практическую применимость этого метода и покажем, как использовать Wolfram Language для быстрого получения решений нелинейных разностных стационарных уравнений.

Сперва найдём частное решение уравнения  $y_{n+2} + y_n = \cos \frac{\pi n}{8}$ . Решим его, воспользовавшись предложенным методом. Заметим, что в этом уравнении нет резонанса, поскольку характеристические числа однородного уравнения равны  $\pm i$ , где  $i$  – мнимая единица, а правая часть представляет собой квазимногочлен, соответствующий паре комплексно-сопряжённых чисел  $e^{\pm i \frac{\pi}{8}}$ .

Матрица оператора  $T|_{L^1_{\frac{\pi}{8}i,C}}$  имеет вид  $T^1_{\frac{\pi}{8}i,C} = \begin{pmatrix} \cos \frac{\pi}{8} & -\sin \frac{\pi}{8} \\ \sin \frac{\pi}{8} & \cos \frac{\pi}{8} \end{pmatrix}$ .

Значения этой матрицы можно рассчитать при помощи компьютерных средств, как представлено на рисунке 4.3.

```
In[10]:= arg = π / 8;
T = FunctionExpand[{{Cos[arg], -Sin[arg]}, {Sin[arg], Cos[arg]}}];
T // MatrixForm
Out[12]//MatrixForm=

$$\begin{pmatrix} \frac{\sqrt{2+\sqrt{2}}}{2} & -\frac{1}{2}\sqrt{2-\sqrt{2}} \\ \frac{\sqrt{2-\sqrt{2}}}{2} & \frac{\sqrt{2+\sqrt{2}}}{2} \end{pmatrix}$$

```

Рисунок 4.3 – Матрица оператора в рассматриваемом пространстве средствами Wolfram Mathematica

Тогда матрицу оператора  $R|_{L^1_{\frac{\pi}{8}i,C}}$  можно вычислить следующим образом (рисунок 4.4).

```
In[23]:= R = T.T + IdentityMatrix[2] // FullSimplify;
R // MatrixForm
Out[24]//MatrixForm=

$$\begin{pmatrix} 1 + \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & 1 + \frac{1}{\sqrt{2}} \end{pmatrix}$$

```

Рисунок 4.4 – Вычисление матрицы разностного оператора из левой части уравнения

Теперь находим частное решение (рисунок 4.5).

```
In[28]:= f = {{0}, {1}};
y = LinearSolve[R, f] // FullSimplify;
y // MatrixForm
Out[30]//MatrixForm=

$$\begin{pmatrix} -\frac{1}{2} + \frac{1}{\sqrt{2}} \\ \frac{1}{2} \end{pmatrix}$$

```

Рисунок 4.5 – Частное решение неоднородного линейного разностного стационарного уравнения матричным методом

Таким образом, искомое частное решение есть  $y_n = \frac{\sqrt{2}-1}{2} \sin \frac{\pi n}{8} + \frac{1}{2} \cos \frac{\pi n}{8}$ .

Рассмотрим теперь уравнение, в котором есть резонанс:  $y_{n+2} + y_n = \sin \frac{\pi n}{2}$ . В пра-

вой части  $f \in L^1_{\frac{\pi}{2}\mathbf{i}, C}$  есть резонанс. Поэтому матрица оператора  $T|_{L^1_{\frac{\pi}{2}\mathbf{i}, C}}$  имеет вид

$$T^1_{\frac{\pi}{8}\mathbf{i}, C} = \begin{pmatrix} 0 & -1 & 0 & -1 \\ 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \end{pmatrix}. \quad (4.10)$$

Соответствующий код в Wolfram записывается следующим образом (рисунок 4.6).

```
In[49]:= arg = π/2;
J = FunctionExpand[{{Cos[arg], -Sin[arg]}, {Sin[arg], Cos[arg]}}];
J // MatrixForm;
Z = {{0, 0}, {0, 0}};
T = ArrayFlatten[{{J, J}, {Z, Z}}];
T // MatrixForm

Out[54]//MatrixForm=

$$\begin{pmatrix} 0 & -1 & 0 & -1 \\ 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

```

Рисунок 4.6 – Формирование блочной матрицы в резонансном случае в среде Wolfram Mathematica

Тогда матрицу оператора  $R|_{L^1_{\frac{\pi}{2}\mathbf{i}, C}}$  можно вычислить следующим образом (рисунок 4.7).

```
In[57]:= R = T.T + IdentityMatrix[Length[T]] // FullSimplify;
R // MatrixForm

Out[58]//MatrixForm=

$$\begin{pmatrix} 0 & 0 & -2 & 0 \\ 0 & 0 & 0 & -2 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

```

Рисунок 4.7 – Вычисление матрицы оператора в левой части разностного уравнения

Остаётся лишь найти частное решение (рисунок 4.8).

```

In[74]:= m = 1 (*кратность*);
L = Length[T];
f = {{1}, {0}};
Rf = R[[1;; 2*m, 2*m+1;; L]]
yp = LinearSolve[Rf, f] // FullSimplify;
yp // MatrixForm

Out[77]=  $\{\{-2, 0\}, \{0, -2\}\}$ 

Out[79]//MatrixForm=

$$\begin{pmatrix} -\frac{1}{2} \\ 0 \end{pmatrix}$$


```

Рисунок 4.8 – Получение частного решения неоднородного линейного разностного стационарного уравнения с комплексным резонансом в правой части

В этом примере, искомым частным решением является  $y_n = -\frac{1}{2}n \sin \frac{\pi n}{2}$ .

#### 4.1.4 Принцип суперпозиции

**Теорема 4.5 (Принцип суперпозиции).** Пусть в уравнении (4.1) правая часть  $f \in L \left\{ L_{\mu_1}^{s_1}, \dots, L_{\mu_k}^{s_k}, L_{\mu_{k+1},C}^{s_{k+1}}, \dots, L_{\mu_l,C}^{s_l} \right\}$ , где  $k$  и  $l$  – целые числа,  $\mu_1, \dots, \mu_k$  – вещественные числа, а  $\mu_{k+1}, \dots, \mu_l$  – комплексные числа.

Тогда для нахождения решения уравнения (4.1) нужно найти матрицу оператора  $T$  в базисе пространства  $L \left\{ L_{\mu_1}^{s_1}, \dots, L_{\mu_k}^{s_k}, L_{\mu_{k+1},C}^{s_{k+1}}, \dots, L_{\mu_l,C}^{s_l} \right\}$ , которая имеет вид

$$T = \begin{pmatrix} T_{\mu_1}^{s_1} & O & \dots & \dots & \dots & O \\ O & \ddots & \dots & \vdots & \dots & \vdots \\ \vdots & \dots & T_{\mu_k}^{s_k} & O & \dots & \vdots \\ \vdots & \dots & O & T_{\mu_{k+1},C}^{s_{k+1}} & \dots & \vdots \\ \vdots & \dots & \dots & \dots & \ddots & \dots \\ O & \dots & \dots & \dots & \dots & T_{\mu_l,C}^{s_l} \end{pmatrix}, \quad (4.11)$$

построить матрицу  $R = T^k + a^{(1)}T^{k-1} + \dots + a^{(k-1)}T + a^{(k)}I$ , вычёркиванием из матрицы нулевых столбцов и строк получить невырожденную матрицу  $R_{nd}$  и решением уравнения

$R_{nd}\mathbf{y} = \mathbf{f}$ , где  $\mathbf{f}$  – вектор-столбец, полученный разложением правой части по базису пространства  $L\left\{L_{\mu_1}^{s_1}, \dots, L_{\mu_k}^{s_k}, L_{\mu_{k+1}, C}^{s_{k+1}}, \dots, L_{\mu_l, C}^{s_l}\right\}$ , получить искомое частное решение.

Результаты исследований отражены в работе [206].

#### **4.2 Локальная гладкая управляемость для некоторой распределенной системы четвёртого порядка**

В данном подразделе опишем один результат исследования управляемости систем с распределенными параметрами. А именно, рассматривается уравнение по своему виду напоминающее уравнение колебания пластины, областью рассмотрения является двумерный тор. Ставится задача привести решение уравнения и первую производную по времени в нулевое состояние за конечное время посредством локального воздействия на поверхность тора. Заметим, что локальное воздействие может быть приложено к сколь угодно малой части данного тора.

Ранее в мировой научной литературе рассматривались подобные задачи. Например, различными авторами исследовались проблемы управляемости для механической системы, описываемой уравнением колебания мембранны. Ставилась задача привести мембрану в состояние покоя за конечное время с помощью локального силового воздействия, приложенного к части области, занимаемой мембранны. Заметим, что добиться поставленной цели в данном случае за сколь угодно малое время невозможно в силу того, что скорость распространения волн у мембранны конечна. Более того оказалось, что для доказательства управляемости приходится требовать выполнения дополнительных условий, накладываемых на геометрию исходной области и подобласти, к которой приложено управление. Условия, накладываемые на область, подобласть и время управления получили название Geometric Control Condition (GCC) и с успехом применялись многими авторами (см., например, [209]) для решения сходных задач.

Похожие проблемы управляемости ставились также для уравнения Шрёдингера в различных областях, в том числе и на торе такими авторами как, например, Розье и Чанг [210]. Для уравнения типа Шрёдингера на торе была установлена локальная управляемость, причем выполнение GCC не требовалось, т. е. подобласть была произвольной и время сколь угодно малым. Кроме того, этими авторами было доказано, что гладкость управляющего воздействия можно повысить, повышая гладкость начальных данных задачи. В работе Джффара [211] доказывалось, что колебания квадратной пластины можно привести в покой локальным воздействием без условий типа GCC.

В нашем исследовании решается задача управляемости также без наложения условий типа GCC и дополнительно, в отличие от работы Джффара, удается доказать возможность повышения гладкости управления за счет повышения гладкости начальных данных. Этого удается достичь применением метода расщепления оператора основного уравнения и использования в дальнейшем известного Hilbert Uniqueness Method, введенного впервые Лионсом.

По результатам исследования опубликована работа [212].

## ЗАКЛЮЧЕНИЕ

В проекте «Анализ моделей многокритериального, коллективного, оптимизационного и структурного выбора решений» в 2023 году были получены следующие основные результаты:

- получена характеристика важного в теории предпочтений класса максимальных доменов Кондорсе;
- исследованы аксиоматические свойства обобщений минимального покрывающего множества на случай слабых турниров;
- разработано веб-приложение для агрегирования рейтингов научных журналов, основанных на различных библиометрических показателях;
- проанализированы различные агрегирования шести популярных библиометрических показателей статей, опубликованных в 2022 году для 177 журналов по менеджменту;
- в рамках исследования по манипулированию процедур агрегирования найдены условия, при которых вероятность существования избирателя, имеющего стимул к манипулированию равна нулю;
- разработана структура базы данных и сайт для хранения и визуализации данных по индивидуальной и коллективной манипулируемости процедур агрегирования;
- разработаны алгоритмы распараллеливания задачи оценки манипулируемости процедур агрегирования;
- проведено сравнение манипулируемости различных процедур агрегирования в случае Impartial Culture, множественного выбора и 4-х методов построения расширенных предпочтений;
- проведено сравнение между манипулируемостью q-Паретовских правил коллективного выбора;
- в рамках теории свидетельств разработана процедура агрегирования информации от экспертов и последующего ранжирования этих агрегированных оценок в том случае, когда оценки представлены в порядковой шкале с указанием их достоверности;
- изучены свойства введенных ранее аппроксимирующих функций на основе многокритериального подхода; исследованы взаимосвязи новых аппроксимирующих функций с классическими; предложены вычислительные методы построения множеств аппроксимирующих функций;

- разработан инструментарий принятия решений по оптимальному формированию набора учебных курсов, преподавательского состава и бюджета при смешанном формате обучения студентов;
- выполнены модельные вычисления и анализ задачи о структуре трехвалютного вклада, оптимального по Сэвиджу, с учетом динамики валютных курсов в 2023 году;
- методами сетевого анализа исследованы публикации в области болезни Паркинсона;
- с помощью разработанного ранее метода выделения аномальных кластеров, примененных к анализу карт температур морской поверхности, описано такое явление как апвеллинг у берегов Португалии;
- средствами языка Python реализована модель паттерн-кластеризации данных с различными типами кластерных трубок; разработана человеко-машинная процедура, позволяющая объединять и обрабатывать в полуавтоматическом режиме результаты разных кластеризаций;
- методами паттерн-кластеризации проведен анализ на страновом уровне экспорта-импорта электроэнергии; были выявлены группы стран, различающиеся по уровню экономического развития и его связи с показателями экспорта-импорта электроэнергии, изменениях этих показателей за последние годы;
- разработана алгебраическая модификация метода неопределенных коэффициентов для решения линейных разностных стационарных уравнений высшего порядка;
- решена задача приведения в состояние покоя некоторой распределенной системы четвёртого порядка на двумерном торе посредством локального гладкого управляющего воздействия.

## СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Отчет МНУЛ анализа и выбора решений, Т3-53 за 2022г. – 2022. – URL: [https://www.hse.ru/data/2023/01/10/2039164567/%D0%98%D1%82%D0%BE%D0%B3%D0%BE%D0%B2%D1%8B%D0%B9%20%D0%BE%D1%82%D1%87%D0%B5%D1%82%20%D0%A2%D0%97-53\\_02.12.2022.pdf](https://www.hse.ru/data/2023/01/10/2039164567/%D0%98%D1%82%D0%BE%D0%B3%D0%BE%D0%B2%D1%8B%D0%B9%20%D0%BE%D1%82%D1%87%D0%B5%D1%82%20%D0%A2%D0%97-53_02.12.2022.pdf) (дата обращения 20.11.2023)
2. Karpov A., Slinko A. Symmetric maximal Condorcet domains// Order. – 2023. – N 40. – P. 289–309.
3. Karpov A. Preferences over mixed manna // In: Goldengorin, B., Kuznetsov, S. (Eds.) Data Analysis and Optimization: In Honor of Boris Mirkin’s 80th Birthday. – Cham: Springer, 2023. – P. 169–178.
4. Karpov A., Slinko A. Constructing large peak-pit Condorcet domains// Theory and Decision. – 2023. – N 94. – P. 97–120.
5. Karpov A. V. Structured preferences: A literature survey // Automation and Remote Control. – 2022. – N 83. – P. 1329–1354.
6. Sen A. K. A Possibility theorem on majority decisions // Econometrica. – 1966. – N 34(2). – P. 491–499.
7. Fishburn P. C. Acyclic sets of linear orders// Social Choice and Welfare. – 1996. – N 14. – P. 113–124.
8. Danilov V. I., Koshevoy G. A. Maximal Condorcet domains // Order. – 2013. – N 30(1). – P. 181–194.
9. Raynaud H.: Paradoxical results from Inada’s conditions for majority rule / Technical report, Stanford University Institute for Mathematical Studies in the Social Sciences. – 1981.
10. Puppe C. The single-peaked domain revisited: A simple global characterization // J. Econ. Theory. – 2018. – N 176. – P. 55–80.
11. Peters D., Lackner M. Preferences single-peaked on a circle// J. of Artif. Intell. Res. – 2020. – N 68 – P. 463–502.
12. Mattei N., Walsh T. Preflib: A library of preference data / Proceedings of Third International Conference on Algorithmic Decision Theory (ADT 2013). – Lecture Notes in Artificial Intelligence. – 2013. – P. 259–270.
13. Bredereck R., Chen J., Woeginger G. J. Are there any nicely structured preference profiles nearby?// Math. Soc. Sci. – 2016. – N 79. – P. 61–73.
14. Bogomolnaia A., Moulin H., Sandomirskiy F. Competitive division of a mixed manna// Econometrica – 2017. – N 85(1). – P. 1847–1871.

15. Dutta B. Covering sets and a new Condorcet choice correspondence // Journal of Economic Theory. – 1988. – N 44. – P. 63–80.
16. Brandt F., Brill M., Harrenstein P. Tournament solutions // Handbook of Computational Social Choice / Brandt, F., Conitzer, V., Endriss, U., Lang, J., Procaccia, A.D. (eds). – Cambridge: Cambridge University Press. – 2016. – P. 57–84.
17. Aleskerov F., Pislyakov V., Subochev A. Ranking Journals in Economics, Management and Political Science by Social Choice Theory Methods // NRU Higher School of Economics. Series WP BRP "Science, Technology and Innovation". – 2014. – N WP BRP 27/STI/2014.
18. Brandt F., Harrenstein P. Set-rationalizable choice and self-stability // Journal of Economic Theory. – 2011. – N 146(4). – P. 1721–1731.
19. Miller N. A new solution set for tournaments and majority voting: Further graph-theoretical approaches to the theory of voting // American Journal of Political Science. – 1980. – Vol. 24. – P. 68–96.
20. Aleskerov F., Subochev A. Matrix-vector representation of various solution concepts // NRU Higher School of Economics. Series WP7 "Математические методы анализа решений в экономике, бизнесе и политике". – 2009. – No. WP7/2009/03.
21. Brandt F., Brill M., Harrenstein P. Extending tournament solutions // Social Choice and Welfare. – 2018. – N 51(2). – P. 193–222.
22. Laslier J.-F. Tournament solutions and majority voting // Heidelberg, New-York: Springer-Verlag. – 1997. – P. 134.
23. Brandt F. Minimal stable sets in tournaments // Journal of Economic Theory. – 2011. – Vol. 146, N 4 – P. 1481–1499.
24. Subochev A. Dominant, Weakly Stable, Uncovered Sets: Properties and Extensions // Working paper WP7/2008/03 – Moscow: Higher School of Economics Publ. House. – 2008. – P. 28.
25. Subochev A. Dominating, Weakly Stable, Uncovered Sets: Properties and Extensions // Automation and Remote Control. – 2010. – N 1. – P. 130–143.
26. Wilhelm J. Generalized solution principles and out-ranking relations in multi-criteria decision-making // European Journal of Operational Research. – 1977. – N 1(6). – P. 376–385.
27. Wuffle A., Feld S. L., Owen G., Grofman B. Finagle's Law and the Finagle Point, a New Solution Concept for Two-Candidate Competition in Spatial Voting Games Without a Core // American Journal of Political Science. – 1989. – N 33(2). – P. 348–375.

28. Laffond G., Laslier J.-F., Le Breton M. The bipartisan set of a tournament game // Games Economic Behavior. – 1993. – V. 5. – P. 182–210.
29. Dutta B., Laslier J.-F. Comparison functions and choice correspondences // Social Choice and Welfare. – 1999. – V. 16, N 4. – P. 513–532.
30. Peris J. E., Subiza B. Condorcet choice correspondences for weak tournaments // Social Choice and Welfare. – 1999. – V. 16, N 2. – P. 217–231.
31. Brandt F., Fischer F. Computing the minimal covering set // Mathematical Social Sciences. – 2008. – V. 56, N 2. – P. 254–268.
32. MVP библиометрической информационной системы: Streamlit [электронный ресурс]. – 2023. – URL: <https://bibliometric-information-system.streamlit.app/> (дата обращения 19.05.2023).
33. Репозиторий проекта на GitHub: GitHub [электронный ресурс]. – 2023. – URL: <https://github.com/AngelinaYudina/Tournament-Solutions-web-application> (дата обращения 19.05.2023).
34. Garfield E., Sher I. H. New factors in the evaluation of scientific literature through citation indexing // American Documentation. – 1963. – V. 14, N 3. – P. 195–201.
35. Moed H. F. Measuring contextual citation impact of scientific journals // Journal of Informetrics. – 2010. – V. 4, N 3. – P. 265–277.
36. Gonzalez-Pereira B., Guerrero-Bote V., Moya-Anegon F. A new approach to the metric of journals scientific prestige: The SJR indicator // Journal of Informetrics. – 2010. – V. 4, N 3. – P. 379 – 391.
37. Файл с результатами работы MVP: Google Drive [электронный ресурс]. – 2023. – URL: <https://bit.ly/3InniM8> (дата обращения 19.05.2023).
38. Veselova Y. A. Manipulation by Coalitions in Voting with Incomplete Information// In: Data Analysis and Optimization. In Honor of Boris Mirkin's 80th Birthday / B. Goldengorin, S. Kuznetsov (eds.). – Springer, 2023. – P. 377–395.
39. Peters H., Veselova Y. A. On the safety of group manipulation// Social Choice and Welfare. – 2023. – N 61. – P. 713–732.
40. Zontov Y. V., Aleskerov F. T. On estimation of manipulability indices for large amounts of agents and high number of alternatives// (готовится к публикации)
41. Gibbard A. Manipulation of voting schemes // Econometrica. – 1973. – N 41. – P. 587–601.
42. Satterthwaite M. Strategy-proofness and Arrow's conditions: existence and correspondence theorems for voting procedures and social welfare functions // Journal of Economic Theory. – 1975. – N 10. – P. 187–217.

43. Duggan J., Schwartz T. Strategic manipulability without resoluteness or shared beliefs: Gibbard–Satterthwaite generalized // Social Choice and Welfare. – 2000. – N 17. – P. 85–93.
44. Chamberlin J. R. An investigation into the relative manipulability of four voting systems // Behavioral Science. – 1985. – N 30(4). – P. 195–203.
45. Aleskerov F., Kurbanov E. Degree of manipulability of social choice procedures // Alkan A. et al. (eds.) Current Trends in Economics. Studies in Economic Theory. Berlin Heidelberg, N.Y.: Springer, 1999. – N 8. – P. 13–27.
46. Xia L., Conitzer V, Procaccia A. A scheduling approach to coalitional manipulation// Proceedings of the ACM Conference on Electronic Commerce. – 2010. – P. 275–284. 10.1145/1807342.1807386.
47. Aleskerov F., Karabekyan D., Sanver R., Yakuba V. On manipulability of positional voting rules // SERIEs: Journal of the Spanish Economic Association. – 2011. – N 2(4). – P. 431–446.
48. Карабекян Д. С. Манипулирование в задаче коллективного принятия решений// диссертация на соискание ученой степени кандидата экономических наук, Москва. – 2012.
49. S. Barbera, W. Bossert, P. Pattanaik. Ranking Sets of Objects. Handbook of Utility Theory/ Barbera, S., Hammond, P.J. and C. Seidl (eds.). – Boston: Kluwer Academic Publishers. –2004.
50. Nitzan S. The vulnerability of point-voting schemes to preference variation and strategic manipulation // Public Choice. – 1985. – N 47. – P. 349–370.
51. Kelly J. Almost all social choice rules are highly manipulable, but few aren't // Social Choice and Welfare. – 1993. – N 10. – P. 161–175.
52. Aleskerov F. T., Ivanov A., Karabekyan D., Yakuba V. I. On the Individual and Coalitional Manipulability of q-Paretoian Social Choice Rules// in: Advances in Collective Decision Making: Interdisciplinary Perspectives for the 21st Century. Springer. – 2023. – P. 95–111.
53. Aleskerov F. Arrovian Aggregation Models. — Dordrecht: Kluwer, 1999.
54. Dempster A.P. Upper and lower probabilities induced by multivalued mapping// Ann. Math. Statist. – 1967. – N 38. – P. 325–339.
55. Shafer G. A mathematical theory of evidence. – Princeton University Press, Princeton, N.J., 1976.
56. Zhang Y., Bouadi T., Martin A. A Clustering Model for Uncertain Preferences Based on Belief Functions// In: Ordonez C., Bellatreche L. (eds) Big Data Analytics and

Knowledge Discovery. DaWaK 2018. Lecture Notes in Computer Science. – 2018. – N 11031. – Springer, Cham. – P. 111–125.

57. Zhang Y., Bouadi T., Wang Y., Martin A. A distance for evidential preferences with application to group decision making// Information Sciences. – 2021. – N 568. – P. 113–132.

58. Martin A. Belief Functions on Ordered Frames of Discernment// In: Le Hégarat-Mascle S., Bloch I., Aldea E. (eds) Belief Functions: Theory and Applications. BELIEF 2022. Lecture Notes in Computer Science. – 2022. – N 13506. – Springer, Cham. – P. 129–138.

59. Smets P. Decision making in TBM: the necessity of the pignistic transformation // International Journal of Approximate Reasoning. – 2005. – N 38. – P. 133–147.

60. Bronevich A., Lepskiy A. Imprecision indices: axiomatic, properties and applications // Int. J. of General Systems. – 2015. – N 44(7-8). – P. 812–832.

61. Dezert J., Tchamova A. On the Effectiveness of Measures of Uncertainty of Basic Belief Assignments // Information & Security: An International Journal. – 2022. – N 52. – P.9–36.

62. Deng Y. Deng entropy// Chaos Solitons Fract. – 2016. – N 91. – P. 549–553.

63. Lepskiy A. On Optimal Blurring of Point Expert Estimates and their Aggregation in the Framework of Evidence Theory// Procedia Computer Science. – 2022. – N 214. – P. 573–580.

64. Lepskiy A. Conflict Measure of Belief Functions with Blurred Focal Elements on the Real Line// In: Denœux T., Lefèvre E., Liu Z., Pichon F. (eds) Belief Functions: Theory and Applications. BELIEF 2021. Lecture Notes in Computer Science. – 2021. – N 12915. – P. 197–206. Springer, Cham.

65. Lepskiy A. Fuzzy Threshold Aggregation// In: Kahraman, C., Sari, I.U., Oztaysi, B., Cebi, S., Cevik Onar, S., Tolga, A.Ç. (eds) Intelligent and Fuzzy Systems. INFUS 2023. Lecture Notes in Networks and Systems. – 2023. – N 758. – P.69-76. Springer, Cham.

66. Zhang H., Deng Y. Entropy Measure for Orderable Sets// Information Sciences. – 2021. – N 561. – P.141–151.

67. Yager R.R. On the Dempster-Shafer framework and new combination rules // Information Sciences. – 1987. – N 41. – P. 93–138.

68. Inagaki T. Interdependence between Safety-Control Policy and Multiple-Sensor Schemes Via Dempster-Shafer Theory// IEEE Transactions on Reliability. – 1991. – N 40(2). – P. 182–188.

69. Lepskiy A. Evidence-Based Aggregation and Ranking in an Ordinal Scale// Procedia Computer Science. – 2023. – N 221C. – P. 1066–1073.

70. Демидович Б. П., Марон И. А., Шувалова Э. З. Численные методы анализа. Приближение функций, дифференциальные и интегральные уравнения. – М.: Наука, 1967. – 368 с.
71. Малов С. В. Регрессионный анализ: теоретические основы и практические рекомендации. – СПб.: Издательство Санкт-Петербургского университета, 2013. – 280 с.
72. Отчет о научно-исследовательской работе «Анализ данных и выбор решений в задачах исследования социально-экономических и политических систем». – М.: ВШЭ, 2021. – 323 с.
73. Мудров В. И., Кушко В. Л. Метод наименьших модулей. – М.: URSS, 2013. – 57 с.
74. Charnes A., Cooper W. W. Management models and industrial applications of linear programming. Vol. 1. – N.Y.: Wiley, 1961. – 859 p.
75. Статников Р. Б., Матусов И. Б. О недопустимых, допустимых и оптимальных решениях в задачах проектирования // Проблемы машиностроения и надежности машин. – 2012. – № 4. – С. 10–19.
76. Подиновский В. В., Нелюбин А. П. Средние величины: многокритериальный подход // Проблемы управления. – 2020. – № 5. – С. 3–16.
77. Подиновский В. В., Нелюбин А. П. Средние величины: многокритериальный подход. II // Проблемы управления. – 2021. – № 2. – С. 33–41.
78. Тырсин А. Н., Азарян А. А. Точные алгоритмы реализации метода наименьших модулей на основе спуска по узловым прямым // Вычислительная математика. – 2017. – № 4. – С. 21–32.
79. Подиновский В. В., Ногин В. Д. Парето-оптимальные решения многокритериальных задач. Изд. второе, испр. и доп. – М.: Физматлит 2007. – 256 с.
80. Зоркальцев В. И. Октаэдрические и евклидовы проекции точки на линейное многообразие // Труды инст. математики и механики УрО РАН. – 2012. – № 3. – С. 106–118.
81. Подиновский В. В. Многокритериальные задачи с однородными равноценными критериями // Журнал вычислительной математики и математической физики. – 1975. – № 2. – С. 330–344.
82. Подиновский В. В. Идеи и методы теории важности критериев в многокритериальных задачах принятия решений. – М.: Наука, 2019. – 103 с.
83. Нелюбин А. П., Подиновский В. В. Алгоритмическое решающее правило, использующее ординальные коэффициенты важности критериев со шкалой первой порядковой метрики // Журнал вычислительной математики и математической физики. – 2012. – № 1. – С. 43–59.

84. Fishburn P. C., Willig R. D. Transfer principles in income redistribution // Journal of Public Economics. – 1984. – V. 25. – P. 323–328.
85. Marshall A. W., Olkin I. Inequalities: Theory of majorization and its applications. – N.Y.: Academic press, 1979. – 569 p.
86. Eurostat. Real GDP per capita, 2022. – [https://ec.europa.eu/eurostat/databrowser/view/sdg\\_08\\_10/default/table](https://ec.europa.eu/eurostat/databrowser/view/sdg_08_10/default/table)
87. Нелюбин А. П., Подиновский В. В. Аппроксимация таблично заданных функций: многокритериальный подход // Журнал вычислительной математики и математической физики. – 2023. – 63(5). С. 717–730.
88. Nelyubin A. P., Podinovski V. V. Approximation of functions defined in tabular form: Multicriterial approach. Computational Mathematics and Mathematical Physics. – 2023. – N 63(5). – P. 730–742.
89. Kolowich S. Exactly how many students take online courses// The Chronicle of Higher Education. – 2014. – 16.
90. Joksimovic S., Gasevic D., Loughin T. M., Kovanovic V., Hatala M. Learning at distance: effects of interaction traces on academic achievement// Comput. Educ. – 2015. – N 87. – P. 204–217.
91. Carnoy M., Kuz'minov Y. Online learning: how it affects the university structure and economics. Panel discussion// Educ. Stud. Moscow. – 2015. – N 3. – P.8–43.
92. Asarta C. J., Schmidt J.R. Comparing student performance in blended and traditional courses: does prior academic achievement matter?// Internet High. Educ. – 2917. – N 32. – P.29–38.
93. Zhu Y., Au W., Yates G. University students' self-control and self-regulated learning in a blended course// Internet High. Educ. – 2016. – N 30. – P.54–62.
94. Zacharis N. Z. A multivariate approach to predicting student outcomes in web-enabled blended learning courses// Internet High. Educ. – 2015. – N 27. – P.44–53.
95. Olitsky N. H., Cosgrove S. B. The effect of blended courses on student learning: evidence from introductory economics courses// Int. Rev. Econ. Educ. – 2015. – N 15. – P.17–31.
96. Alvarez A., Martin M., Fernandez-Castro I., Urretavizcaya M. Blending traditional teaching methods with learning environments: Experience, cyclical evaluation process and impact with MAgAdI// Comput. Educ. – 2013. – N 68. – P.129–140.
97. Yigit T., Koyun A., Yuksel A. S., Cankaya I. A. Evaluation of blended learning approach in computer engineering education// Procedia Soc. Behav. Sci. – 2014. – N 141. – P.807–812.

98. Winston G. C., Zimmerman D. J. Peer effects in higher education// In: The Economics of Where to Go, When to go, and How to Pay for it. – 2014. – P.395–421. University of Chicago Press.
99. Griffith A. L., Rask K. N. Peer effects in higher education: a look at heterogeneous impacts// Econ. Educ. Rev. – 2014. – N 39. – P.65–77.
100. Bazerman M., Moore D. Judgment in Managerial Decision Making, – 8th edn. – Wiley. – 2012.
101. Gilboa I. Making Better Decisions: Decision Theory in Practice. – 1st edn. – Wiley/Blackwell. – 2010.
102. Saaty T. Decision Making for Leaders: The Analytic Hierarchy Process for Decisions in a Complex World. – 3rd Revised edn. – RWS Publications. – 2012.
103. Noyes J., Cook M. Decision Making in Complex Environments. – 1st edn. – CRC Press. – 2017.
104. de Haan A., de Heer P. Solving Complex Problems: Professional Group Decision-Making Support in Highly Complex Situations. – 2nd Revised edn. – Eleven International Publishing. – 2015.
105. Hammond J., Keeney R., Raiffa H. Smart Choices: A Practical Guide to Making Better Decisions. – Crown Business. – 2002.
106. Adair J. Decision Making and Problem Solving (Creating Success). – Kogan Page. – 2013.
107. Kim J. Y. A study on learners' perceptual typology and relationships among the learner's types, characteristics, and academic achievement in a blended e-education environment// Comput. Educ. – 2012. – N 59 (2). – P.304–315.
108. Linder K. E. The Blended Course Design Work book: A Practical Guide. – Workbook edn. – Stylus Publishing. – 2016.
109. Kalberer N., Petendra B., Bohmer C., Schibelbein A., Beck-Meuth E.M. Evaluation process and quality management in a blended-learning bachelor's programme// Procedia Soc. Behav. Sci. – 2016. – N 228. – P.131–137.
110. Belenky A. Developing and Running a Set of Competitive College/University Blended Courses// Data Analysis and Optimization. In Honor of Boris Mirkin's 80th Birthday / Ed. by B. Goldengorin, S. Kuznetzov, Springer. – 2023. – Ch. 2. – P. 23–42,
111. Barak M., Watted A., Haick H. Motivation to learn in massive open online courses: examining aspects of language and social engagement// Comput. Educ. – 2016. – N 94(3). – P.49–80.

112. Sendra-Portero F., Torales-Chaparro O. E., Ruiz-Gomez M. J., Martinez-Morillo M. A pilot study to evaluate the use of virtual lectures for undergraduate radiology teaching// Eur. J. Radiol. – 2013. – N 82(5). – P.888–893.
113. Yilmaz O. The effects of “live online courses” on students’ achievement at distance education// Procedia Soc. Behav. Sci. – 2012. – N 55(5). – P.347–354.
114. Song H., Kim J., Luo W. Teacher-student relationship in online classes: a role of teacher self-disclosure// Comput. Hum. Behav. – 2016. – N 54. – P.436–443.
115. Shukor N. A., Tasir Z., der Meijden H. V. An examination of online learning effectiveness using data mining// Procedia Soc. Behav. Sci. – 2015. – N 172. – P. 555–562.
116. Cho Y. H., Choi H., Shin J., Yu H. C., Kim Y. K., Kim J.Y. Review of research on online learning environments in higher education// Procedia Soc. Behav. Sci. – 2015. – N 191(2). – P. 2012–2017.
117. Beach P. Self-directed online learning: a theoretical model for understanding elementary teachers’ online learning experiences// Teach. Teacher Educ. – 2017. – N 61. – P. 60–72.
118. Rienties B., Brouwer N. Z., Lygo-Baker S. The effects of online professional development on higher education teachers’ beliefs and intentions towards learning facilitation and technology// Teach. Teacher Educ. – 2013. – N 29(1). – P. 122–133.
119. Karaman S., Kucuk S., Aydemir M. Evaluation of an online continuing education program from the perspective of new graduate nurses// Nurse Educ. Today. – 2014. – N 34(5). – P. 836–841.
120. Abdelaziz M., Kamel S. S., Karam O. Abdelrahman A. Evaluation of e-learning program versus traditional lecture instruction for undergraduate nursing students in a faculty of nursing// Teach. Learn. Nurs. – 2011. – N 6(2). – P. 50–58.
121. Pei L., Wu H. Does online learning work better than offline learning in undergraduate medical education? A systematic review and meta-analysis// Med. Educ. Online. – 2019. – N 24(1). – P. 1–13.
122. Mubarak A., Al-Arimi A. K. Distance learning// Procedia Soc. Behav. Sci. – 2014. – N 152. – P. 82–88.
123. Tsiotakis P., Jimoyiannis A. Critical factors towards analyzing teachers’ presence in online learning communities// Internet High. Educ. – 2016. – N 28(1). – P. 45–58.
124. van Rooij S. W., Zirkle K. Balancing pedagogy, student readiness and accessibility: a case study in collaborative online course development// Internet High. Educ. – 2015. – N 28(1). – P. 1–7.

125. Markova T., Glazkova I., Zaborova E. Quality issues of online distance learning// Procedia Soc. Behav. Sci. – 2017. – N 237. – P. 685–691.
126. Cohen A., Baruth O. Personality, learning, and satisfaction in fully online academic courses// Comput. Hum. Behav. – 2017. – N 72. – P. 1–12.
127. Boling E. C., Hough M., Krinsky H., Saleem H., Stevens M. Cutting the distance in distance education: perspectives on what promotes positive, online learning experiences// Internet High. Educ. – 2012. – N 15(2). – P. 118–126.
128. Moore J. L., Dickson-Deane C., Galyen K. E-learning, online learning, and distance learning environments: are they the same?// Internet High. Educ. – 2011. – N 14(2). – P. 129–135.
129. Sanders-Smith S., Smith-Bonahue T., Soutullo O. Practicing teachers' responses to case method of instruction in an online graduate course// Teach. Teacher Educ. – 2016. – N 54(2). – P. 1–11.
130. Vo H. M., Zhu C., Diep N.A. The effect of blended learning on student performance at course-level in higher education: a meta-analysis// Stud. Educ. Eval. – 2017. – N 53. – P. 17–28.
131. Ocak M. A. Why are faculty members not teaching blended courses? Insights from faculty members// Comput. Educ. – 2011. – N 56(3). – P. 689–693.
132. Lust G., Elen J., Clarebout G. Regulation of tool-use with in a blended course: Student differences and performance effects// Comput. Educ. – 2013. – N 60(1). – P. 385–395.
133. Ellis R. A., Pardo A., Han F. Quality in blended learning environments significant differences in how students approach learning collaborations// Comput. Educ. – 2016. – N 102. – P. 90–102.
134. Boelens R., DeWever B., Voet M. Four key challenges to the design of blended learning: a systematic literature review// Educ. Res. Rev. – 2017. – N 22. – P. 1–18.
135. Dziuban C., Moskal P. A course is a course is a course: factor in variance in student evaluation of online, blended and face-to-face learning environments// Internet High. Educ. – 2011. – N 14(4). – P. 236–241.
136. Nazarenko A. L. Blended learning vs traditional learning: what works? (a case study research)// Procedia Soc. Behav. Sci. – 2015. – N 200. – P. 77–82.
137. Slechtova P., Hana V., Jan V. Blended learning: promising strategic alternative in higher education// Procedia Soc. Behav. Sci. – 2015. – N 171. – P. 1245–1254.
138. Sophonhiranrak S., Suwannathachote P., Ngudgratoke S. Factors affecting creative problem solving in the blended learning environment: a review of the literature// Procedia Soc. Behav. Sci. – 2015. – N 174. – P. 2130–2136.

139. Havnes A., Christiansen B., Bjork I. T., Hessevaagbakke E. Peer learning in higher education: patterns of talk and interaction in skills centre simulation// Learn. Cult. Soc. Interaction. – 2016. – N 8. – P. 75–87.
140. Brunello G., DePaola M., Scoppa V. Residential peer effects in higher education: does the field of study matter?// Econ. Enquiry. – 2010. – N 48(3). – P. 621–634.
141. Sacerdote B. Peer effects in education: How might they work, how big are they, and how much do we know thus far?// In: Handbook of the economics of Education. – 2010. – N 3. – P. 249–277.
142. Kashefi H., Ismail Z., Yusof Y. M. Overcoming students obstacles in multivariable calculus through blended learning: a mathematical thinking approach// Procedia Soc. Behav. Sci. – 2012. – N 56. – P. 579–586.
143. Savenye W. C., Olina Z., Niemczyk M. So you are going to be an online writing instructor: issues in designing, developing and delivering an online course// Comput. Compos. – 2001. – N 18(4). – P. 371–385.
144. Mo Z. Research on the teaching model application tendency of blended learning in the displayed design course// IERI Procedia. – 2012. – N 2. – P. 204–208.
145. McGee P., Reis A. Blended course design: a synthesis of best practices// J. Asynchronous Learn. Netw. – 2012. – N 16(4). – P. 7–22.
146. Vai M., Sosulski K. Essentials of Online Course Design: A Standards-Based Guide (Essentials of Online Learning). – 2nd edn. Routledge. – 2015.
147. Belenky A. Analyzing the potential of a firm: an operations research approach// Autom. & Remote Control. – 2002. – N 35(13). – P. 1405–1424.
148. Hu T. C., Kahng A. B. Linear and Integer Programming Made Easy. – 1st edn. Springer. – 2016.
149. Belenky A. Two classes of games on polyhedral sets in systems economic studies// In: Network Models in Economics and Finance. – Springer. – 2014. – P. 35–80.
150. Гермейер Ю. Б. Введение в теорию исследования операций. – Москва: Наука. – 1971.
151. Savage L. Y. The Foundation of Statistics. – New York: Willey. – 1954.
152. Черемных Ю. Н. Микроэкономика. – Москва: ИНФРА-М. – 2008.
153. Zhukovskiy V. I., Molostvov V. S., Topchishvili A. L. Multicurrency Deposit Diversification – Three Possible Approaches to Risk Accounting// International Journal of Operations and Quantitative Management. – 2014. – V 20. – P. 1–14.

154. Molostvov V. S. Multiple-criteria optimization under uncertainty: concepts of optimality and sufficient conditions// Theory and Practice of Multiple Criteria Decision Making. North-Holland. – 1983. – P. 91–105.
155. Molostvov V. S. Savage's principle for non-cooperative games under uncertainty// Proceedings of the international scientific conference “Problems of control and power engineering”. – Tbilisi. – 2004. – P.38–39.
156. Molostvov V. S. Multiple criteria optimization for stochastic systems with uncertain parameters// Model Assisted Statistics and Applications. – 2011. – V 3. – P. 231–237.
157. Molostvov V. S. Three-Currency Deposit Diversification: Savage's Principle Approach// Advances in Systems Science and Applications. – 2022. – V 22(3). – P. 135–146.
158. Wald A. Contribution to the theory of statistical estimation and testing hypothesis// Annals Math. Statist. –1939. – N 10. – P. 299–326.
159. Лабскер Л. Г., Ященко Н. А. Экономические игры с природой. – Москва: Кнорус. – 2015.
160. Воробьев Н. Н. Теория игр для экономистов-кибернетиков. – Москва: Наука. – 1985.
161. Петросян Л. А., Зенкевич Н. А., Семина Е. А. Теория игр. – Москва: Книжный дом «Университет». – 1998.
162. Markowitz H. M. Portfolio Selection// Journal of Finance. – 1952. – V 7(1). – P 77–91.
163. Берзон Н. И., Дорошин Д. И. Особенности применения показателей эффективности финансовых инвестиций// Финансы и кредит. – 2012. – N 14. – P 21–33.
164. Markowitz H. M. Mean Variance Analysis in Portfolio Choice and Capital Markets. – Basil: Blackwell. – 1990.
165. Подиновский В. В. Парето-оптимальные решения многокритериальных задач. – Москва: ФИЗМАТЛИТ. – 2007.
166. Шапкин А. С. Экономические и финансовые риски. Оценка, управление, портфель инвестиций. – Москва: «Дашков и Ко». – 2003.
167. Молоствов В. С. Решение Сэвиджа в задаче диверсификации трехвалютного депозита// Журнал Новой Экономической Ассоциации. – 2023 (в печати).
168. Aleskerov F., Khutorskaya O., Yakuba V., Stepochkina A., Zinoveva K. Network analysis of publications on studies of Parkinson Disease //Procedia Computer Science. – 2023. – N 219. – P. 1380–1387.
169. Nascimento S., Casca S., Mirkin B. A seed expanding cluster algorithm for

deriving upwelling areas on sea surface temperature images// Computers and Geosciences. – 2015. – N 85. – P.74–85.

170. Миркин Б. Г. Введение в анализ данных. – М.: Юрайт. – 2014.
171. Rodin I., Mirkin B. Supercluster in statics and dynamics: An approximate structure imitating a rough set// In: Polkowski, L., Yao, Y., Artiemjew P., Ciucci D., Liu D., Śleżak D., Zielosko B. (Eds.). Rough Sets. – Springer Int. Publ.: Cham. – 2017. – P. 576–586.
172. Nascimento S., Martins A., Relvas P., Luís J. F., Mirkin B. Core–shell clustering approach for detection and analysis of coastal upwelling// Computers & Geosciences. – 2023. – N 179. – A. 105421.
173. Алекскеров Ф. Т., Белоусова В. Ю., Егорова Л. В., Миркин Б. Г. Анализ паттернов в статике и динамике, часть 1: Обзор литературы и уточнение понятия // Бизнес-информатика. – 2013. – 3(25). – С. 3–18.
174. Алекскеров Ф. Т., Белоусова В. Ю., Егорова Л. В., Миркин Б. Г. Анализ паттернов в статике и динамике, часть 2: Примеры применения к анализу социально-экономических процессов. – 2013. – № 4 (26). – С. 3–20.
175. Алекскеров Ф. Т., Белоусова В. Ю., Сердюк М. Ю., Солодков В. М. Стереотипы поведения российских банков// Банковское дело. – 2008. – N 7. – С.44–50.
176. Aleskerov F. T., C. Alper E. C. A clustering approach to some monetary facts: a long-run analysis of cross-country data// The Japanese Economic Review. – 2000. – N 51(4) – P. 555–567.
177. Алекскеров Ф. Т., Гохберг Л. М., Егорова Л. Г., Мячин А. Л., Сагиева Г. С. Анализ данных науки, образования и инновационной деятельности с использованием методов анализа паттернов – препринт WP7/2012/07 // НИУ «Высшая школа экономики». – М.: Изд. дом Высшей школы экономики. – 2012.
178. Алекскеров Ф. Т., Солодков В. М., Челнокова Д. С. Динамический анализ паттернов поведения коммерческих банков России // Экономический журнал ВШЭ. – 2006. – №1 – С. 48–61.
179. Global Electricity Review // Ember. – 2022.
180. Statistical Review of World Energy, 71<sup>st</sup> edition // BP. – 2022.
181. Electricity Market Report // International Energy Agency. – 2022.
182. Pineau P.-O., Hira A., Froschauer K. Measuring international electricity integration: a comparative study of the power systems under the Nordic Council, MERCOSUR, and NAFTA // Energy policy 32. – Elsevier. – 2004. – P. 1457–1475.
183. Newbery D., Strbac G., Viehoff I. The benefits of integrating European electricity markets // Cambridge working papers in economics. – Cambridge University. – 2015.

184. Abrell J., Rausch S. Cross-County Electricity Trade, Renewable Energy and European Transmission Infrastructure Policy // Economics Working Paper Series – No. 16/229. – 2016.
185. Antweiler W. Cross-border trade in electricity // Journal of International Economics – 2016. – N 101. – P. 42–50.
186. Murshed M. Can regional trade integration facilitate renewable energy transition to ensure energy sustainability in South Asia? // Energy Reports – 2021. – N 7. – P. 808–821
187. Oseni M. O. Get rid of it: To what extent might improved reliability reduce self-generation in Nigeria? – University College London. – 2016.
188. World Trade Organization [Электронный ресурс]. – 2022 – URL: <https://stats.wto.org> (дата обращения 15.11.2022).
189. World Bank [Электронный ресурс]. – 2023 – URL: <https://data.worldbank.org> (дата обращения 13.01.2023).
190. The Observatory of Economic Complexity [Электронный ресурс]. – 2023 – URL: <https://oec.world/en/profile/hs/electricity> (загружено 21.01.2023).
191. Energy Information Administration [Электронный ресурс]. – 2023 – URL: <https://www.eia.gov/international/data/world/electricity> (дата обращения 13.01.2023).
192. Our World in Data based on BP Statistical Review of World Energy (2022) [Электронный ресурс]. – 2023 – URL: <https://ourworldindata.org> (дата обращения 13.01.2023).
193. База данных Всемирного банка [Электронный ресурс]. – 2023 – URL: <https://databank.worldbank.org> (дата обращения 13.01.2023).
194. Aleskerov F. Egorova L., Gokhberg L., Myachin A., Sagieva G. Pattern analysis in the study of science, education and innovative activity in Russian regions // Procedia Computer Science. – 2013. – N 17. – P. 687-694.
195. Алескеров Ф.Т, Белоусова В.Ю., Егорова Л.Г., Миркин Б.Г. Анализ паттернов в статике и динамике, часть 2: Примеры применения к анализу социально-экономических процессов // Бизнес-информатика. – 2013. – N 4(26). – С. 3–20.
196. Global Electricity Review // Ember. – 2021.
197. Chen W. Differential Operator Method of Finding a Particular Solution to an Ordinary Nonhomogeneous Linear Differential Equation with Constant Coefficients // arXiv:1802.09343v1. – 2018.
198. Spiegel M. R. Schaum's Outline of Theory and Problems of Advanced Mathematics for Engineers and Scientists. – McGraw-Hill. – 2002.

199. Fecenko J. Differential Operator Method of Finding a Particular Solution to a Nonhomogeneous Linear Ordinary Differential Equation with Constant Coefficients // arXiv: 2101.02037v1. – 2020.
200. Эльсгольц Л. Э. Дифференциальные уравнения и вариационное исчисление. – М.: Наука. – 1969.
201. Степанов В. В. Курс дифференциальных уравнений. – М.: Наука. – 1959.
202. Понtryгин Л. С. Курс дифференциальных уравнений. – М.: Наука. – 1959.
203. Zill D.G. A First Course in Differential Equations with Modeling Applications. – 10<sup>th</sup> edition. – Brooks/Cole. – 2011.
204. Boyce W. E., DiPrima R.C. Elementary Differential Equations and Boundary Value Problems. – 10<sup>th</sup> edition. – Wiley. – 2012.
205. Nagle R., Saff E. B., Snider A. D. Fundamentals of Differential Equations. – 8<sup>th</sup> edition. – AddisonWesley. – 2012.
206. Романко В. К. Курс разностных уравнений. – М.: Физматлит. – 2012.
207. Гельфонд А. О. Исчисление конечных разностей. – М.: Физматлит. – 1959.
208. Lomonosov T. A. Algebraic Modification of the Method of Undetermined Coefficients for Solving Nonhomogeneous Linear Difference Equations // arXiv: 2307.07182. – 2023.
209. Bardos C., Lebeau G., Rauch J. Sharp scientific conditions for the observation, control and stabilization of wave from the boundary // SIAM. Journal on Optimization and Control. – 1992. – 30(5). – P. 1024– 1065.
210. Rosier L., Zhang B. Y. Control and stabilization of the nonlinear Schrodinger equation on rectangles// M3AS: Math. Models Methods. Sci. – 2010. – 20(12). – P. 2293– 2347.
211. Jaffard S. Contrôle interne exact des vibrations d'une plaque rectangulaire// Portugal. Math. – 1990. – 47(4). – P. 423.
212. Romanov I., Shamaev A. Controllability to Rest for the “Plate Oscillation” Equation on the Torus in the Case of Local Force Action / Пер. с рус. // Mathematical notes. – 2023. – 113(4). – P. 598–600.