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

УДК 338.24.01
Рег. № НИОКТР 122040800216–2
Рег. № ИКРБС

УТВЕРЖДАЮ
Проректор НИУ ВШЭ,
канд. физ.-мат. наук
_____Д.А. Дагаев
«___» _____ 2022 г.

ОТЧЕТ
О НАУЧНО-ИССЛЕДОВАТЕЛЬСКОЙ РАБОТЕ

ТЕОРЕТИЧЕСКИЕ И ПРИКЛАДНЫЕ ЗАДАЧИ АНАЛИЗА ДАННЫХ И ПРИНЯТИЯ РЕШЕНИЙ В СОЦИАЛЬНО-ЭКОНОМИЧЕСКОЙ И ФИНАНСОВОЙ СФЕРАХ

(заключительный)

Руководитель НИР,
Директор, д-р техн. Наук, проф. _____ Ф.Т. Алескеров

Москва 2022

СПИСОК ИСПОЛНИТЕЛЕЙ

Руководители НИР:

Директор, д-р техн. наук, проф.	_____ подпись, дата	Ф.Т. Алескеров (введение, заключение)
Гл. науч. сотр., PhD	_____ подпись, дата	Э. Маскин (подраздел 2.3)

Исполнители:

Зам. директора центра, д-р физ.- мат. наук, проф.	_____ подпись, дата	А.Е. Лепский (подраздел 3.2)
Ведущий науч. сотр., д-р техн. наук, проф.	_____ подпись, дата	А.С. Беленький (подраздел 2.3)
Ведущий науч. сотр., д-р физ.-мат. наук, проф.	_____ подпись, дата	А.А. Лазарев (подраздел 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 1.1)
Ст. науч. сотр., канд. экон. наук	_____ подпись, дата	Д.С. Карабекян (подраздел 1.4)
Ст. науч. сотр., канд. экон. наук	_____ подпись, дата	А.В. Карпов (подраздел 1.2)
Ст. науч. сотр., канд. наук по прикл. мат.	_____ подпись, дата	Т.А. Ломоносов (подраздел 6.3)
Ст. науч. сотр., канд. физ.-мат. наук	_____ подпись, дата	В.С. Молоствов (подраздел 2.1)
Ст. науч. сотр., канд. техн. наук, доц.	_____ подпись, дата	А.Л. Мячин (подраздел 5.3, подраздел 6.1)
Ст. науч. сотр., канд. физ.-мат. наук	_____ подпись, дата	А.Н. Субочев (подраздел 1.3)
Ст. науч. сотр., канд. наук по прикл. мат.	_____ подпись, дата	С.В. Швыдун (подраздел 4.2)
Ст. науч. сотруд., канд. техн. наук	_____ подпись, дата	В.И. Якуба (подраздел 1.4)

Науч. сотр., PhD	_____ подпись, дата	Э. Доган (подраздел 1.5)
Науч. сотр., PhD	_____ подпись, дата	С. Дутта (подраздел 2.1)
Науч. сотр., канд. наук по прикл. мат.	_____ подпись, дата	М.А. Марон (подраздел 2.2)
Науч. сотр.	_____ подпись, дата	Н.Г. Мещерякова (подраздел 4.1)
Мл. науч. сотр., канд. физ.-мат. наук	_____ подпись, дата	Ю.А. Веселова (подраздел 1.4)
Мл. науч. сотр.	_____ подпись, дата	С.С. Демин (подраздел 4.4)
Мл. науч. сотр., канд. физ.-мат. наук	_____ подпись, дата	Л.Г. Егорова (подраздел 5.4)
Мл. науч. сотр.	_____ подпись, дата	Ю.В. Зонтов (подраздел 1.4)
Стажёр-исследователь	_____ подпись, дата	А.П. Гусейнова (подраздел 3.2)
Стажёр-исследователь	_____ подпись, дата	И.В. Дунилов (подраздел 6.2)
Стажёр-исследователь	_____ подпись, дата	Д.С. Егоров (подраздел 5.2)
Стажёр-исследователь	_____ подпись, дата	А.А. Иванов (подраздел 1.4)
Стажёр-исследователь	_____ подпись, дата	К.Е. Зиновьева (подраздел 5.1)
Стажёр-исследователь	_____ подпись, дата	А.И. Казачинская (подраздел 5.1)
Стажёр-исследователь	_____ подпись, дата	Д.А. Кобец (подраздел 5.3)
Стажёр-исследователь	_____ подпись, дата	А.С. Семина (подраздел 5.1)

Стажёр-исследователь	_____ подпись, дата	А.К. Степочкина (подраздел 4.3)
Стажёр-исследователь	_____ подпись, дата	М.А. Тарасова (подраздел 5.3)
Стажёр-исследователь	_____ подпись, дата	Д.С. Ткачев (подраздел 1.6)
Стажёр-исследователь	_____ подпись, дата	М.Е. Торопова (аннотация на англ.)
Стажёр-исследователь	_____ подпись, дата	Д.А. Чубарова (подраздел 5.3)
Стажёр-исследователь	_____ подпись, дата	Е.В. Шахалиева (аннотация на рус.)
Стажёр-исследователь	_____ подпись, дата	А.В. Юдина (подраздел 1.3)
Нормоконтроль	_____ подпись, дата	М.М. Земцова

РЕФЕРАТ

Отчет 252 с., 1 кн., 43 рис., 33 табл., 273 источн., 1 прил.

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

Объектами исследования были методы и приложения сетевого и кластерного анализа данных, задачи и методы многокритериального принятия решений, методы и задачи теории свидетельств, коллекции текстов, базы данных торговли продовольствием, базы данных регионального развития и пр.

Целью проекта были разработка новых моделей, методов и алгоритмов в областях оптимизационного моделирования, сетевого и кластерного анализа данных, многокритериального принятия решений, теории коллективного выбора, теории функций доверия. Прикладная часть проекта направлена на решение ряда задач анализа данных и принятия решений в социально-экономической и финансовой сферах.

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

Область применения полученных результатов: управление образованием, антиковидное регулирование, управление региональным развитием, системы выборов на всех уровнях, принятие решений во внешней торговле, управление природопользованием и защитой окружающей среды, теория коллективного выбора, сетевой анализ данных, многокритериальное принятие решений и др.

СОДЕРЖАНИЕ

ВВЕДЕНИЕ 11

1 Исследования по анализу и выбору решений 18

1.1 Анализ непротиворечивости качественной информации о важности групп критериев 18

1.1.1 Введение 18

1.1.2 Математическая модель и сведения из теории важности критериев 19

1.1.3 Отношения важности критериев 22

1.1.4 Качественная вероятность 23

1.1.5 Непротиворечивость информации о важности 25

1.1.6 Коэффициенты важности критериев 26

1.1.7 Отношения предпочтения 26

1.1.8 Об одном претенденте на аналитическое решающее правило 28

1.1.9 Некоторые выводы 29

1.2 Исследование вычислительных свойств сепарабельных предпочтений 30

1.2.1 Постановка задачи 30

1.2.2 Вычислительная сложность 33

1.2.3 Генерирование сепарабельных предпочтений 33

1.2.4 Некоторые выводы 34

1.3 Исследование турнирных способов выбора оптимальных вариантов на основе минимального покрывающего множества 35

1.3.1 Введение 35

1.3.2 Общие теоретические сведения 36

1.3.3 Свойства версий минимального покрывающего множества и их связь с другими турнирными решениями 41

1.3.4 Эмпирическое сравнение методов ранжирования 43

1.3.5 Некоторые выводы 50

1.4 Исследование манипулируемости правил коллективного выбора 51

1.4.1 Введение 51

1.4.2 Индивидуальное манипулирование при полной информации. Расчёт и анализ индексов ... 51
1.4.3 Неполнота информации и одновременное манипулирование ... 60
1.4.4 Некоторые выводы ... 69
1.5 Исследования по теории паросочетаний ... 69
1.5.1 Проблема паросочетаний адвокатов и судебных дел ... 69
1.5.2 Алгоритмы паросочетаний, максимизирующие доход ... 73
1.6 Исследование аксиоматики линейных функций интенсивности коалиций ... 76
1.6.1 Введение ... 76
1.6.2 Основная часть ... 77
1.6.3 Некоторые выводы ... 82
2 Исследование оптимизационных моделей ... 83
2.1 Нахождение структуры многовалютных депозитов, оптимальных по гарантированному риску ... 83
2.1.1 Введение ... 83
2.1.2 Постановка задачи диверсификации вклада ... 84
2.1.3 Анализ функции риска и аналитическое решение в частном случае ... 85
2.1.4 Вычислительная среда и результаты расчетов ... 87
2.1.5 Некоторые выводы ... 92
2.2 Формирование гибких бизнес-правил для улучшения процесса закупки компонентов для систем управления ... 93
2.2.1 Введение ... 93
2.2.2 Анализ проблемы ... 94
2.2.3 Математическая постановка проблемы ... 94
2.2.4 Решение задачи ... 95
2.2.5 Численный пример ... 97
2.2.6 Некоторые выводы ... 97
2.3 Математическое моделирование формирования оптимального набора учебных курсов по выбору ... 98

2.3.1 Введение 98

2.3.2 Основная часть 103

2.3.3 Некоторые выводы 117

2.4 Исследования по теории расписаний 121

3. Исследования по развитию методов кластерного анализа данных 122

3.1 Обобщенное описание коллекции текстов с использованием таксономии предметной области 122

3.1.1 Введение 122

3.1.2 Задача оптимального подъема нечеткого тематического кластера в таксономии: модель и метод 124

3.2 Кластерная декомпозиция тел свидетельств на основе оптимизации мер конфликта131

3.2.1 Введение 131

3.2.2 Некоторые сведения из теории свидетельств 132

3.2.3 Кластеризация тела свидетельств 133

3.2.4 Оценивание внутреннего конфликта тела свидетельств на основе его кластеризации 142

3.2.5 Некоторые выводы 143

4. Исследования по теории и приложениям сетевого анализа данных 144

4.1 Аксиоматика мер центральности 144

4.1.1 Введение 144

4.1.2 Описание аксиом мер центральности 145

4.1.3 Некоторые выводы 147

4.2 Модели схожести сетевых структур 147

4.2.1 Введение 147

4.2.2 Основная часть 148

4.2.3 Некоторые выводы 153

4.3 Сетевой анализ публикаций в области научных исследований болезни Паркинсона154

4.3.1 Введение 154

4.3.2 Сетевой анализ ... 155
4.3.3 Некоторые выводы ... 160
4.4 Сценарный анализ ущерба экологии от разлива нефти в Арктическом регионе на основе имитационной сетевой модели ... 160
4.4.1 Введение ... 160
4.4.2 Основная часть ... 162
4.4.3 Некоторые выводы ... 163
5. Анализ социально-экономических и наукометрических данных ... 165
5.1 Анализ цитируемости и библиометрических показателей международных бизнес-журналов ... 165
5.1.1 Введение ... 165
5.1.2 Основная часть ... 166
5.1.3 Некоторые выводы ... 174
5.2 Обработка зеркальных данных торговли продовольственными товарами ... 174
5.2.1 Введение ... 174
5.2.2 Основная часть ... 175
5.2.3 Некоторые выводы ... 178
5.3 Динамический анализ структурной близости регионов ... 179
5.3.1 Введение ... 179
5.3.2 Сбор, описание и обработка данных ... 180
5.3.3 Выбор показателей ... 181
5.3.4 Описание данных ... 182
5.3.5 Нормировки данных ... 183
5.3.6 Алгоритм кластеризации ... 185
5.3.7 Выбор параметров алгоритма ... 185
5.3.8 Динамический паттерн-анализ ... 187
5.3.9 Некоторые выводы ... 188
5.4 Анализ предпочтений студентов относительно вакцинации ... 189

5.4.1 Введение 189
5.4.2 Основная часть 190
5.4.3 Некоторые выводы 200
6. Исследование по проблематики COVID-19 и другим моделям распространения 201
6.1 Моделирование прироста числа случаев заболевания COVID-19 для разных групп стран 201
6.1.1 Введение 201
6.1.2 Исходные данные 202
6.1.3 Методология разбиения 203
6.1.4 Составление регрессионных моделей 205
6.1.5 Некоторые выводы 210
6.2 Исследование математических моделей распространения эпидемии COVID-19 ... 211
6.2.1 Введение 211
6.2.2 Основная часть 211
6.2.3 Некоторые выводы 214
6.3 Сравнение условий L2-диссипативности и спектрального условия фон Неймана на схемах на разнесённых сетках для уравнений газовой динамики 214
6.3.1 Введение 214
6.3.2 Основная часть 219
6.3.3 Некоторые выводы 222
ЗАКЛЮЧЕНИЕ 223
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ 226
ПРИЛОЖЕНИЕ А. Таблицы к подразделу 5.2 246

ВВЕДЕНИЕ

В данном отчете приведены результаты выполнения проекта «Теоретические и прикладные задачи анализа данных и принятия решений в социально-экономической и финансовой сферах» в 2022 году сотрудниками Международного центра анализа и выбора решений (МЦАВР) НИУ ВШЭ.

Основными научными направлениями исследований были:

– модели анализа и выбора решений;

– оптимизационные модели;

– методы и прикладные вопросы кластерного анализа данных;

– модели сетевого анализа данных;

– анализ социально-экономических и наукометрических данных;

– модели распространения (в том числе COVID-19).

В этом проекте были продолжены исследования 2021 г. по выполнению НИР «Анализ данных и выбор решений в задачах исследования социально-экономических и политических систем» [1]. В частности, в 2021 г. были получены следующие результаты:

– исследованы некоторые классы ПН-средних величин, основанные на применении методов многокритериального принятия решений; исследована устойчивость таких средних к малым изменениям исходных данных;

– в рамках исследования структурированных дихотомических предпочтений в модели с полной информацией получены характеристики доменов через запрещенные подпрофили; определено может ли неполный профиль допускать возможное или необходимое завершение в определенных ограниченных доменах;

– в рамках исследования турнирных способов выбора оптимальных вариантов дано содержательное описание существенного множества, как обобщения двухпартийного множества; описаны важные свойства турнирных решений на основе такого множества; рассмотрена задача агрегирование противоречивых рейтингов научных журналов по менеджменту;

– в рамках исследования манипулируемости в теории коллективного выбора разработана и проанализирована модель, учитывающая одновременно и неполноты информации, и неопределенность относительно действий других избирателей на манипулируемость; разработаны и реализованы эффективные алгоритмы вычисления точных значений сильной и слабой манипулируемости;

– в рамках исследования индексов влияния, учитывающих предпочтения участников по созданию коалиций, найдена аксиоматика 4 функций интенсивности связи, основанных на функциях минимума и максимума;

– предложена и исследована теоретико-игровая модель задачи организации переработки твердых бытовых отходов города; установлены проверяемые необходимые и достаточные условия точек равновесия Нэша в такой игре;

– исследована задача оптимальной диверсификации мультивалютных депозитов с неопределенными будущими обменными курсами как проблема минимизации упущенной из-за неопределенности выгоды;

– предложена модификация метода расчёта приоритетов стратегических целей по таблице парных сравнений их значимости; разработано соответствующее программное обеспечение;

– разработан новый программный пакет для таксономического контент-анализа коллекций текстовых документов, реализующий методологию обобщения в таксономии предметной области;

– разработана вычислительно экономичная модификация алгоритма диффузионно-инвариантной паттерн-кластеризации на основе сортировки данных;

– предложена новая модификация дихотомии графов на основе разработанного ранее частотного алгоритма декомпозиции; с помощью модифицированного алгоритма проанализировано функционирования рынка S&P-500 в период 1990–2019 гг.;

– разработан и исследован двухэтапный алгоритм кластеризации тел свидетельств в рамках теории функций доверия; функцию доверия, соответствующую кластеризованному телу свидетельств, можно рассматривать как аппроксимацию исходной функции доверия;

– проанализированы механизмы изменения климата и его текущего воздействия на социально-экономическое развитие Арктики;

– проанализированы среднесрочных стратегий развития Арктического региона в условиях риска;

– проанализированы основные причины нехватки водных ресурсов в странах Северной Африки и механизмы, которые позволили бы смягчить негативные последствия нехватки воды;

– рассмотрена модификация интервального метода оболочечного анализа данных; с помощью этой модификации оценена эффективность карантинных мер в разных странах мира;

– методами регрессионного анализа выявлены основные факторы, определяющие предпочтения респондентов относительно формата онлайн-обучения;

– проанализировано влияние перехода на основанный на внутренних рейтингах (ПВР) подход на котировки банковских акций в рамках исследования банковского регулирования;

– предложена параметризация регулирования достаточности капитала исламских банков как функция от макроэкономических параметров;

– проведено сетевое исследование международного рынка заимствований за 2005–2020 гг. с точки зрения анализа структуры сети, выделения наиболее важных элементов и сообществ, анализа устойчивости сети и её динамического изменения;

– проведено исследование сети цитирования 466 российских экономических журналов РИНЦ за период 2016–2020 гг. с использованием индекса центральности Коупленда, индекс группового влияния BI и индекс влияния по ключевым вершинам;

– построена сетевая структура международной торговли продовольствием с учетом неполноты (неточности, неопределенности) и противоречивости данных;

– в рамках исследования моделей распространения COVID-19 выполнен анализ эффективности антиковидных мер в странах мира, выявлены группы стран со схожими показателями интенсивности введенных мер, законопослушности, уровня образования, вакцинационной политики;

– проведен анализ изменения экономических показателей Великобритании с изменением интенсивности заболеваемостью COVID-19 и с принимаемыми антиковидными мерами;

– разработан программный комплекс моделирования клеточного апоптоза для изучения поведения клеток в различных условиях.

Исследования 2022 года частично продолжают исследования 2021 года. В частности, это исследования по оцениванию степени манипулируемости правил агрегирования предпочтений, исследования по тематическому моделированию, по сетевому анализу данных и др.

Некоторые результаты исследований этого года носят предварительный характер — их предполагается продолжить в следующем году.

Методологической базой исследования данного проекта были современные методы принятия решений, теории коллективного выбора, теории вероятностей, теории функций доверия, методы оптимизации, методы сетевого и кластерного анализа данных и др.

Ряд исследований имеют прикладной характер. При решении некоторых прикладных задач было использовано прикладное программное обеспечения и разработаны комплексы программ.

Отчет состоит из 6 разделов.

В первом разделе представлены результаты исследований по анализу и выбору решений.

В подразделе 1.1 приведены результаты исследования задачи принятия решения при наличии информации о важности групп критериев. В частности, найдены способы проверки непротиворечивости информации о важности и способы построения введенных отношений предпочтения. Кроме того, выявлена взаимосвязь качественной значимости и качественной вероятности.

В подразделе 1.2 приведены результаты исследования вычислительной сложности задач коллективного выбора для случая, когда предпочтения избирателей являются сепарабельными. В частности, показано, что правило Янга и правило Чемберлина − Куранта являются NP-трудными. Кроме того, разработан алгоритм генерирования сепарабельных предпочтений.

Подраздел 1.3 посвящен исследованиям турнирных способов выбора оптимальных вариантов на основе обобщений минимального покрывающего множества. Кроме того, в этом подразделе рассмотрена задача агрегирования международных рейтингов вузов на основе указанных обобщений и других методов ранжирования на основе турнирных решений.

В подразделе 1.4 приведены результаты продолжающегося в отчетном году исследования манипулируемости правил коллективного выбора. В частности, разработаны алгоритмы, позволяющие получить точные значения индексов манипулируемости для случая трех альтернатив. Проанализированы значения индексов индивидуального манипулирования для числа участников от 3 до 100, для различных процедур голосования и для 4 типов расширенных предпочтений.

Подраздел 1.5 отчета посвящен результатам исследований по теории и приложениям паросочетаний. В частности, исследована стабильность паросочетаний между государственными адвокатами и судебными делами, в которых истцы и/или ответчики не могут позволить себе правовых услуг. Рассмотрены алгоритмы паросочетаний, максимизирующие доход на рынке бракосочетаний.

В подразделе 1.6 приведены результаты исследования аксиоматики линейных функций интенсивности связи для кардинальных индексов влияния.

Во втором разделе отчета приведены результаты исследования оптимизационных моделей.

Так, подраздел 2.1 посвящен разработке программного комплекса в среде Excel для решения задачи о нахождении структур трехвалютного депозита, оптимальных по гарантированному риску, в условиях неизвестных будущих обменных курсов валют. Кроме того, проведено численное моделирование зависимости минимального гарантированного риска от максимальных возможных будущих значений курсов валют.

В подразделе 2.2 предложена методика формирования гибких бизнес-правил для улучшения процесса закупки компонентов для систем управления на основе критерия минимума максимального риска.

Подраздел 2.3 посвящен математическому моделированию формирования набора учебных курсов по выбору, который был бы оптимальным в некотором смысле, как для студента с точки зрения администрации, так и для самих студентов.

Наконец в подразделе 2.4 приведены некоторые результаты исследований по теории расписаний.

В третьем разделе отчета приведены результаты исследований в рамках развития методов кластерного анализа.

Так в подразделе 3.1 приведены результаты разработки и тестирования метода максимального правдоподобия для решения задачи обобщения нечетких кластеров в таксономиях предметных областей.

В подразделе 3.2 приведены результаты по разработке методов кластеризации тел свидетельсвт в рамках теории функций доверия. В частности, описаны два новых метода кластеризации тел свидетельств. Первый — на основе вычисления функции плотности конфликта между фокальными элементами, а второй — на основе минимизации среднего внутрикластерного конфликта.

Раздел 4 отчета посвящен результатам исследования по разработке методов сетевого анализа данных и их применению к анализу конкретных сетевых структур.

Так в подразделе 4.1 проанализированы аксиоматические свойства ряда мер центральности сетевых структур.

В подразделе 4.2 приведены результаты сравнительного анализа существующих моделей оценки схожести сетевых структур.

Подраздел 4.3 посвящен описанию результатов сетевого анализа научных публикаций в области исследований болезни Паркинсона за период 2015–2021 гг.

В подразделе 4.4 приведены результаты разработки имитационной модели разлива нефти для специфических условий Арктики (низкие температуры, лед на поверхности моря и т. д.).

Раздел 5 посвящен результатам анализ социально-экономических и наукометрических данных.

Так в подразделе 5.1 исследованы библиометрические и наукометрические показатели международных журналов по бизнесу, менеджменту и бухгалтерскому учету из базы данных Scopus.

В подразделе 5.2 приведены результаты по разработке методологии полуавтоматического устранения расхождения зеркальных сетевых данных торговли продовольственными товарами. Эта методология была апробирована на реальных данных мировой торговли пшеницей и рисом за последние четыре года.

В подразделе 5.3 описаны результаты анализа регионального развития методом динамической паттерн-кластеризации. В частности, выявлены схожие траектории развития регионов, найдены группы регионов со схожим развитием, проанализирована устойчивость регионального развития.

Подраздел 5.4 посвящен выявлению ключевых факторов, влияющих на принятие решений относительно вакцинации против COVID-19 молодыми людьми. Это исследование выполнено с помощью обработки результатов опроса, проведенного среди студентов вузов методом эксперимента дискретного выбора.

Раздел 6 отчета посвящен исследованию различных по характеру моделей распространения (эпидемий, информации, тепла и т. д.).

Так в подразделе 6.1 приведены результаты разработки методологии 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).

Подраздел 6.2 посвящен описанию процедуры унификации и оптимизации мониторинга текущей ситуации с COVID-19, прогнозирования роста заражений с использованием моделей кластеризации и линейной регрессии, анализа эффективности с помощью метода оболочечного анализа данных. Эта процедура была реализована на языке Python.

Наконец, в подразделе 6.3 рассмотрены результаты численного сравнения условий диссипативности и спектрального условия фон Неймана на схемах на разнесённых сетках для уравнений газовой динамики.

Представленные в отчете результаты были апробированы на семинарах и конференциях различного уровня. Результаты исследований нашли отражение в ряде публикаций или находятся в печати.

1 Исследования по анализу и выбору решений

Этот раздел отчета посвящен результатам исследований в области анализа и выбора решений. Основными направлениями исследований в этой области знаний были:

− анализ непротиворечивости качественной информации о важности групп критериев;

− исследование вычислительных свойств сепарабельных предпочтений;

− исследование турнирных способов выбора оптимальных вариантов на основе минимального покрывающего множества;

− исследование манипулируемости правил коллективного выбора;

− исследования по теории паросочетаний;

− исследование аксиоматики линейных функций интенсивности коалиций.

1.1 Анализ непротиворечивости качественной информации о важности групп критериев

1.1.1 Введение

Принципиальным недостатком известных методов анализа многокритериальных задач принятия решений, использующих оценки важности критериев, является то, что само понятие важности критериев формально не определяется и полагается, что человек будет исходить из своего интуитивного понимания, что такое важность [2, 3]. Поэтому для оценивания коэффициентов важности человеку предлагается отвечать на «лобовые» вопросы типа «Во сколько раз один из критериев важнее другого?» или «Какая доля общей важности всех критериев приходится на рассматриваемый критерий?». Проблема состоит в том, что невозможно установить точный смысл, вкладываемый конкретным человеком в ответы на указанные вопросы [4], а потому и нельзя корректно использовать такую информацию о важности для анализа решений и выработки обоснованных рекомендаций.

В России была создана и продолжает активно развиваться математическая теория важности критериев – ТВК (историю и библиографию см. в [5]). Она опирается на строгие определения понятий равенства критериев в важности и превосходства в важности одного критерия над другим (качественная важность) и превосходства в важности одного критерия над другим в h раз (количественная важность). В этой теории созданы корректные методы получения информации о важности критериев и разработаны решающие правила, задающие отношения предпочтения на основе качественной или же количественной информации о важности критериев с учетом информации об изменении предпочтений вдоль их шкалы.

Хотя определения равенства и превосходства в важности для групп, или (под)множеств критериев были даны ещё в [6], почти во всех работах по ТВК рассматрива-

лись только такие многокритериальные задачи, в которых информация о важности касалась лишь отдельных критериев. Исключение составляют несколько публикаций, однако в них изучались только отдельные частные случаи специфических групп критериев. Так, в [7] рассматривались задачи с несколькими упорядоченными по важности, но попарно не пересекающимися группами критериев. В [8] рассматривались равноважные группы, состоящие из равноважных критериев. В статье [9] введено определение понятия степени превосходства в важности одной из групп критериев над другой. Статьи [10, 11] посвящены задачам с иерархической критериальной структурой, в которой в роли групп критериев выступают критерии более высоких уровней иерархии по отношению к критериям нижнего уровней. Сложившееся положение можно объяснить сложностью анализа задач, в которых имеются упорядоченные по важности группы критериев, хотя актуальность исследований таких задач, часто встречающихся на практике, несомненна.

В настоящем подразделе отчета рассмотрены вопросы анализа непротиворечивости качественной информации о важности групп критериев, получения и обработки такой информации и построения на её основе решающих правил, позволяющих сравнивать варианты решений по предпочтительности. При этом используется аналогия с качественной вероятностью.

1.1.2 Математическая модель и сведения из теории важности критериев

Дальнейшее изложение опирается на следующую математическую модель ситуации принятия индивидуального решения в условиях определенности:

$$< X,\ \tau, f, Z_0, R >, \tag{1.1}$$

где X – множество вариантов (альтернатив, планов, стратегий, …); τ – тип постановки задачи (выбрать один наилучший или несколько лучших вариантов, упорядочить все варианты по предпочтительности и т.д.); $f = (f_1, f_2, \ldots, f_m)$ – векторный критерий, состоящий из $m \geq 2$ частных критериев f_i; $Z_0 \subseteq \mathrm{Re} = (-\infty, +\infty)$ – область значений критериев, или множество шкальных оценок (шкала критериев); R – отношение нестрогого предпочтения лица, принимающего решение (ЛПР). Под *критерием* f_i понимается функция, определенная на X и принимающая значения из Z_0. Каждый вариант x из множества X характеризуется своей *векторной,* или *критериальной оценкой* $y(x) = f(x) = (f_1(x), f_2(x), \ldots, f_m(x))$. Поэтому сравнение вариантов по предпочтительности сводится к сопоставлению их критериальных оценок. Множество всех критериальных оценок есть $Z = Z_0^m$. Далее полагается, что шкала критериев порядковая и

известно лишь, что предпочтения вдоль их шкалы возрастают: чем больше оценка $z \in Z_0$, тем она предпочтительнее.

Предпочтения ЛПР моделируются на Z при помощи *отношения нестрогого предпочтения* R, так что yRz означает, что векторная оценка y не менее предпочтительна, чем z. Отношение R является (*частичным*) *квазипорядком*, т.е. оно рефлексивно и транзитивно, и порождает *отношение безразличия* I и (строгого) *предпочтения* P следующим образом: $yIz \Leftrightarrow yRz \wedge zRy$; $yPz \Leftrightarrow yRz \wedge \neg zRy$ (т.е. zRy неверно). Отношение R неизвестно и подлежит восстановлению на основе информации о предпочтениях ЛПР, состоящей из сведений об относительной важности критериев и характере изменения предпочтений на Z_0. При отсутствии такой информации на множестве векторных оценок Z можно задать лишь отношение строгого предпочтения $P^{\varnothing}$ (отношение Парето):

$$yP^{\varnothing}z \Leftrightarrow (y_i \geq z_i,\ i = 1, \ldots, m,\ y \neq z) \qquad (1.2)$$

Пусть $R^{\varnothing}$ есть объединение $P^{\varnothing}$ и отношения равенства векторов.

Для расширения отношения $P^{\varnothing}$ требуется дополнительная информация Ω о предпочтениях ЛПР, в роли которой будут выступать сведения о важности критериев, например, сообщение типа «Одна группа критериев важнее другой». Приведем базовые определения равенства и превосходства в важности для непересекающихся групп критериев. Далее для простоты записи критерии f_i будем обозначать также и их номерами i, так что, например, под i будем понимать критерий f_i, а под $A = \{i_1, i_2, \ldots, i_q\}$ будем понимать группу (множество) критериев с номерами $i_1, i_2, \ldots, i_q$. Пусть $M = \{1, 2, \ldots, m\}$.

Пусть A и B – непересекающиеся непустые множества номеров, или группы критериев, в векторной оценке y все компоненты y_i с номерами из A равны между собой и все компоненты y_i с номерами из B равны между собой (т. е. $y_i = y_j$ при $i, j \in A$ и при $i, j \in B$). Векторные оценки такой структуры будем обозначать y^{AB}. Обозначим через $y^{A \leftrightarrow B}$ векторную оценку, полученную из векторной оценки y с указанной структурой путем замены каждой компоненты y_j, $j \in B$, на (любую) компоненту y_i, $i \in A$, и замены каждой компоненты y_i, $i \in A$, на (любую) компоненту y_j, $j \in B$.

Определение 1.1. Группы критериев *A и B равноважны*, или *одинаково важны* (такое сообщение обозначается $A \sim B$), когда всякая векторная оценка y^{AB} и векторная оценка $y^{A \leftrightarrow B}$ одинаковы по предпочтительности (безразличны).

Определение 1.2. Группа критериев *A важнее* группы критериев *B* (такое сообщение обозначается $A \succ B$), когда всякая векторная оценка y^{AB}, в которой $y_i > y_j$ при $i \in A$ и $j \in B$, предпочтительнее векторной оценки $y^{A \leftrightarrow B}$.

Согласно этим определениям, сообщения $A \sim B$ и $A \succ B$ задают на множестве Z соответственно отношения безразличия $I^{A \sim B}$ и предпочтения $P^{A \succ B}$:

$$yI^{A \sim B}z \Leftrightarrow y = y^{AB}, z = y^{A \leftrightarrow B}, y_i \neq y_j, i \in A, j \in B;$$
$$yP^{A \succ B}z \Leftrightarrow y = y^{AB}, z = y^{A \leftrightarrow B}, y_i > y_j, i \in A, j \in B. \quad (1.3)$$

Каждое сообщение ω из накопленной информации Ω о важности задает на множестве векторных оценок Z, согласно (1.3), соответствующее отношение R^{ω}, а именно отношение предпочтения P^{ω} или безразличия I^{ω}. В соответствии с фрагментарным подходом, разработанным в ТВК (см. также [12]) на основе этой информации на множестве векторных оценок Z определяется отношение нестрогого предпочтения R^{Ω} (рефлексивное и транзитивное отношение) как транзитивное замыкание объединения всех отношений R^{ω} и отношения $R^{\varnothing}$:

$$R^{\Omega} = \mathrm{TC}[(\bigcup_{\omega \in \Omega} R^{\omega}) \bigcup R^{\varnothing}], \quad (1.4)$$

где ТС – символ операции транзитивного замыкания бинарного отношения. Согласно (1.4), соотношение $yR^{\Omega}z$ верно тогда и только тогда, когда существует цепочка вида

$$yR^{\omega_1}u^1, u^1R^{\omega_2}u^2, \ldots, u^{r-1}R^{\omega_r}z, \quad (1.5)$$

где u^k – векторные оценки, а ω_k – сообщения из Ω (так что R^{ω_k} есть P^{ω_k} или I^{ω_k} в зависимости от смысла ω_k) или же символ $\varnothing$. Отношение нестрогого предпочтения R^{Ω} порождает указанным выше образом отношения предпочтения P^{Ω} и безразличия I^{Ω}. Отношение R^{Ω} индуцирует аналогичное по смыслу отношение R_{Ω} на множестве вариантов:

$$x' R_{\Omega} x'' \Leftrightarrow f(x') R^{\Omega} f(x'') \quad (1.6)$$

Отношение R_{Ω} непосредственно используется для формирования решения поставленной задачи требуемого типа.

1.1.3 Отношения важности критериев

Информация Ω задает на множестве 2^M подмножеств множества (номеров) критериев $M = \{1, 2, \ldots, m\}$ (бинарные) отношения *равноважности*, или одинаковой важности $\sim^\Omega$ и *превосходства в важности*, или большей важности $\succ^\Omega$: $A\sim^\Omega B$, когда $A\sim B\in\Omega$ или $B\sim A\in\Omega$; $A\succ^\Omega B$, когда $A\succ B\in\Omega$.

Определение 1.3. Информация Ω *слабо непротиворечива,* если соответствующие отношения $\sim^\Omega$ *и* $\succ^\Omega$ не пересекаются, а отношение $\succ^\Omega$ иррефлексивно и асимметрично.

Далее полагаем, что Ω слабо непротиворечива, так что не может быть одновременно $A\sim^\Omega B$ и $A\succ^\Omega B$, а также $A\succ^\Omega A$ и одновременно $A\succ^\Omega B$ и $B\succ^\Omega A$. Для случая сравнения по важности только отдельных критериев расширение отношения $\sim^\Omega$ по транзитивности не расширяет отношения R^Ω, ибо когда $i\sim j\in\Omega$ и $j\sim k\in\Omega$, то, как легко проверить, если $yI^{i\sim j}z$ и $zI^{j\sim k}u$, то и $yI^{i\sim k}u$. Для групп критериев аналогичное утверждение неверно.

Однако для случая сравнения по важности только отдельных критериев расширение отношения $\succ^\Omega$ по транзитивности расширяет отношение R^Ω, что было отмечено еще в [13].

Пусть $\succcurlyeq^\Omega$ – объединение отношений $\sim^\Omega$ и $\succ^\Omega$. Если принять естественное допущение, что отношения важности должны быть транзитивными, то тогда целесообразно ввести в рассмотрение транзитивное замыкание объединения отношений $\succcurlyeq^\Omega$ и отношения равенства множеств =. Однако для эффективного использования качественной информации о важности этого недостаточно. Вначале обобщим определения 1 и 2 на случай пересекающихся групп критериев. Пусть группы критериев A и B не пусты ($A \neq \varnothing$, $B \neq \varnothing$) и находятся в общем положении: $C = A\setminus B \neq \varnothing$ и $D = B\setminus A \neq \varnothing$. Тогда будем полагать, что $A\sim B$ или же что $A\succ B$, если выполнены условия из определения 1 или же 2 соответственно, т.е. (см. (1.3)):

$$yI^{A\sim B}z \Leftrightarrow y = y^{CD},\ z = y^{C\leftrightarrow D},\ y_i \neq y_j,\ i\in C,\ j\in D; \tag{1.7}$$

где $C = A\setminus B$ и $D = B\setminus A$.

Пусть группы критериев A, B и C не пусты, причем A и B не пересекаются с C, т. е. $A\cap C = \varnothing$ и $B\cap C = \varnothing$. Естественно полагать, что если $A\sim B$ или же $A\succ B$, то и, в соответствии с расширенными определениями 1 и 2, верно также $A\cup C\sim B\cup C$ или же $A\cup C\succ B\cup C$ соответственно; и наоборот: если верно $A\cup C\sim B\cup C$ или же $A\cup C\succ B\cup C$, то верно также и $A\sim B$ или же $A\succ B$ соответственно. Указанное свойство называется *аддитивностью*. Наконец, если $A \supset B$,

то разумно принять, что $A \succ B$, причем и в том случае, когда $B = \varnothing$ (в частности, верно $M \succ \varnothing$). Отметим, что соотношению $A \succ \varnothing$ соответствует отношение $P^{A \succ \varnothing}$, определяемое для векторных оценок y^A и z^A, каждая из которых имеет равные компоненты с номерами из A, следующим образом: $y^A P^{A \succ \varnothing} z^A \Leftrightarrow y_i^A > z_i^A$, $i \in A$.

Расширим упомянутое выше отношение $\succcurlyeq^{\Omega}$ путем внесения всевозможных соотношений между группами критериев за счет использования свойств транзитивности и аддитивности, а также всех соотношений $i \succ \varnothing$ и отношения равенства множеств, до отношения $\succcurlyeq^{at\Omega}$, называемого его *аддитивно-транзитивным замыканием*. По построению, отношение $\succcurlyeq^{at\Omega}$ обладает следующими свойствами:

1) рефлексивность: $A \succcurlyeq^{at\Omega} A$ для любого $A \in M$;

2) транзитивность: если $A \succcurlyeq^{at\Omega} B$ и $B \succcurlyeq^{at\Omega} C$, то $A \succcurlyeq^{at\Omega} C$ для любых A, B, $C \in M$;

3) аддитивность: если $A \cap C = B \cap C = \varnothing$, то соотношения $A \succcurlyeq^{at\Omega} B$ и $A \cup C \succcurlyeq^{at\Omega} B \cup C$ выполнены или не выполнены одновременно для любых $A \neq \varnothing$, $B \neq \varnothing$, $C \neq \varnothing$;

4) существенность: для любого $A \in M$ выполнено $A \succcurlyeq^{at\Omega} \varnothing$.

Отметим, что отношение, обладающее свойствами 1, 2 и 3, называется *частичным аддитивным квазипорядком*, который можно назвать *частичной качественной важностью*. Если для любых групп критериев A и B верно $A \succcurlyeq^{at\Omega} B$ или $B \succcurlyeq^{at\Omega} A$, то имеем полный *аддитивный квазипорядок*, или *полную качественную важность*. Свойства 1–4 показывают, что качественная важность формально близко к качественной вероятности.

1.1.4 Качественная вероятность

В интересах дальнейшего изложения приведем сведения о качественной вероятности [14; 15]. Пусть $\Lambda = \{\lambda^1, \lambda^2, \ldots, \lambda^n\}$ – конечное множество произвольной природы. Его элементы λ^j называются элементарными событиями, а его подмножества A, B, C – событиями.

Определение 1.4. Бинарное отношение $\succsim$ на множестве 2^{Λ} называется (*полной*) *качественной вероятностью*, если оно обладает следующими свойствами:

А1. *Сравнимость*: для любых A и B выполнено по крайней мере одно из соотношений: $A \succsim B$ или $B \succsim A$.

А2. *Транзитивность*: *из соотношений* $A \succsim B$ *и* $B \succsim C$ *следует* $A \succsim C$.

А3. *Существенность*: $\Lambda \succ \varnothing$ *и для любого* A *выполнено* $A \succsim \varnothing$.

А4. *Аддитивность*: если $A \cap C = B \cap C = \varnothing$, то соотношения $A \succsim B$ и $A \cup C \succsim B \cup C$ выполнены или не выполнены одновременно.

Из аксиом А1 и А2 следует, что качественная вероятность рефлексивна и потому является связным квазипорядком. Поэтому его асимметричная часть $\succ$ есть (частичный) порядок, а симметричная часть $\sim$ есть эквивалентность.

Пусть задана количественная вероятность Pr рядом распределения: $\Pr(\lambda^1) = p_1$, …, $\Pr(\lambda^n) = p_n$, так что вероятность события A равна: $\Pr A = \sum_{j:\, \lambda^j \in A} p_j$. Говорят, что количественная вероятность (*численно) представляет* (полную) качественную вероятность $\gtrsim$, или что количественная вероятность *согласована* с качественной, если выполнено условие:

$$A \gtrsim B \Leftrightarrow \Pr A \geqslant \Pr B \text{ для любых } A, B. \tag{1.8}$$

Из (1.8) следует, что $A \succ B \Leftrightarrow \Pr A > \Pr B$, $A \sim B \Leftrightarrow \Pr A = \Pr B$ для любых A, B.

Если задана количественная вероятность Pr, то порождаемое ею согласно (1.8) отношение $\gtrsim$ удовлетворяет всем аксиомам А1 – А4, т. е. является (полной) качественной вероятностью.

Ещё де Финетти в 1949 г. предположил, что всякую качественную вероятность можно представить некоторой количественной. Однако [16] построили контрпример для $n = 5$. В итоге оказалось, что лишь при $n \leqslant 4$ для всякой качественной вероятности существует согласованная с ней количественная вероятность. Ниже приводится этот контрпример в иных обозначениях.

Определение 1.5. Бинарное отношение $\gtrsim$ на множестве 2^Λ называется частичной качественной вероятностью, если оно удовлетворяет аксиомам А2 – А4 и аксиоме

А6. *Рефлексивность*: для любого A выполнено $A \gtrsim A$.

Отметим следующие свойства частичной качественной вероятности:

1) Если $A \cap C = B \cap C = \varnothing$, то оба соотношения $A \sim B$ и $A \cup C \sim B \cup C$, а также оба соотношения $A \succ B$ и $A \cup C \succ B \cup C$ выполнены или не выполнены одновременно.

2) Если $A \supset B$, то $A \gtrsim B$ и $A \succ B \Leftrightarrow A \setminus B \succ \varnothing$.

3) Если $A \gtrsim B \cup C$, то $A \gtrsim B$. Если при этом $A \succ B \cup C$, то $A \succ B$.

Оказывается, что частичная качественная вероятность обладает также рядом неочевидных и своеобразных свойств, в том числе таких:

1) При $n \geqslant 5$ не всякая частичная качественная вероятность может быть продолжена до полной [17].

2) если для частичной качественной вероятности события A и B не сравнимы, т. е. неверно ни $A \gtrsim B$, ни $B \gtrsim A$, и она может быть продолжена до полной неединственным образом,

то не все из трех возможностей $A \succ B$, $B \succ A$, $A \sim B$ могут быть реализованы хотя бы для одного из продолжений. Т. е., например, если A и B не сравнимы, то для одних продолжений может оказаться, что имеет место $A \succ B$, для других – что справедливо $A \sim B$, но ни для одного продолжения не будет верно $B \succ A$. Соответствующие контрпримеры были построены в [18].

Для качественной важности выполнение $i \sim^{at\Omega} \varnothing$ означает учет критерия, не имеющего вообще никакой важности. Но если набор критериев сформирован с выполнением известных требований к его формированию [19], то такой результат нужно рассматривать как следствие нарушения указанных требований. Если же дальнейший анализ покажет, что указанный результат следует признать верным, то это будет означать, что нужно скорректировать математическую модель проблемной ситуации, удалив соответствующий критерий из состава векторного критерия. Поэтому для качественной важности естественно принять требование, что $i \succ^{\Omega} \varnothing$ для каждого критерия.

1.1.5 Непротиворечивость информации о важности

Отметим вначале, что если $A \sim^{\Omega} B$, т. е. $A \sim B \in \Omega$, то $A \sim^{at\Omega} B$.

Определение 1.6. Информация Ω непротиворечива, если $\succ^{\Omega} \subseteq \succ^{at\Omega}$ и для каждого критерия i верно $i \succ^{at\Omega} \varnothing$. В противном случае она противоречива.

Согласно этому определению, непротиворечивость информации Ω означает, в частности, что если $A \succ^{\Omega} B$, т.е. $A \succ B \in \Omega$, то должно быть верным и $A \succ^{at\Omega} B$.

Теорема 1.1. *Информация* Ω *непротиворечива тогда и только тогда, когда невозможно составить цикл вида* $A \succ^{\Omega} B$, $B \succcurlyeq^{at\Omega} A$ или $i \succ \varnothing$, $\varnothing \succcurlyeq^{at\Omega} i$.

В теореме 1.1 вместо $B \succcurlyeq^{at\Omega} A$ и $\varnothing \succcurlyeq^{at\Omega} i$ можно записать соответственно $B \sim^{at\Omega} A$ и $\varnothing \sim^{at\Omega} i$.

Пример 1.1. В трехкритериальной задаче $\Omega = \{1 \succ 2, 23 \succ 13\}$. Имеем $1 \succcurlyeq^{at\Omega} 2$ и $23 \succcurlyeq^{at\Omega} 13$, так что верно $2 \sim^{at\Omega} 1$, а не $1 \succ^{at\Omega} 2$. Информация Ω противоречива. Имеем цикл: $1 \succ^{\Omega} 2$, $2 \sim^{at\Omega} 1$.

Пример 1.2. В пятикритериальной задаче $\Omega = \{15 \succ 235, 23 \succ 134\}$. Здесь $15 \succcurlyeq^{at\Omega} 235 \Rightarrow 15 \succ^{at\Omega} 25 \Rightarrow 1 \succcurlyeq^{at\Omega} 2$, а также $23 \succcurlyeq^{at\Omega} 134 \Rightarrow 2 \succcurlyeq^{at\Omega} 14 \Rightarrow 2 \succcurlyeq^{at\Omega} 1$, так что $1 \sim^{at\Omega} 2$. Далее, $15 \succcurlyeq^{at\Omega} 235 \Rightarrow 1 \succcurlyeq^{at\Omega} 23$, откуда с учетом $23 \succcurlyeq^{at\Omega} 134$ получаем $1 \succcurlyeq^{at\Omega} 134$. Из $1 \succcurlyeq^{at\Omega} 134$ по аддитивности $\varnothing \succcurlyeq^{at\Omega} 34$, $\varnothing \succcurlyeq^{at\Omega} 3$, $\varnothing \succcurlyeq^{at\Omega} 4$. В итоге имеем: $1 \sim^{at\Omega} 2$, $3 \sim^{at\Omega} \varnothing$, $4 \sim^{at\Omega} \varnothing$ (критерий 5 не сравним по $\succcurlyeq^{at\Omega}$ с первыми четырьмя критериями). Информация Ω противоречива. Имеется цикл: $15 \succ^{\Omega} 235$, $235 \succcurlyeq^{at\Omega} 15$, так как $235 \succcurlyeq^{at\Omega} 25 \sim^{at\Omega} 15$.

1.1.6 Коэффициенты важности критериев

Аналогом вероятностей элементарных событий p_i являются *коэффициенты важности* (отдельных) *критериев* α_i – положительные в сумме равные 1 числа, обладающие для полной качественной важности $\succcurlyeq^{at\Omega}$ следующим свойством:

для любых групп критериев A и B: $$A\succcurlyeq^{at\Omega}B \Leftrightarrow \sum_{i:\, i\in A}\alpha_i \geqslant \sum_{i:\, i\in B}\alpha_i. \qquad (1.9)$$

Коэффициенты важности для полной качественной важности могут и не существовать (см. пример 1.1). Для частичной качественной важности $\succcurlyeq^{at\Omega}$ *коэффициентами важности* называются положительные в сумме равные 1 числа, обладающие свойством (1.9), в котором знак $\Leftrightarrow$ заменен на $\Rightarrow$, а также таким свойством:

для любых групп критериев A и B: $$A\succ^{at\Omega}B \Rightarrow \sum_{i:\, i\in A}\alpha_i > \sum_{i:\, i\in B}\alpha_i. \qquad (1.10)$$

Из (1.9) следует, что если для любых групп критериев A и B выполнено $A\sim^{at\Omega}B$, то справедливо равенство $\sum_{i:\, i\in A}\alpha_i = \sum_{i:\, i\in B}\alpha_i$. Понятно, что коэффициенты важности для частичной качественной важности могут и не существовать.

Так, в примерах 1.1, 1.2 коэффициенты важности не существуют и информация Ω противоречива. С другой стороны, существование коэффициентов важности для частичной качественной важности не гарантирует возможность ее продолжения до полной и тем более существование для нее коэффициентов важности.

1.1.7 Отношения предпочтения

При сравнении векторных оценок по предпочтительности нужно учитывать не только те соотношения из $\succcurlyeq^{\Omega}$, которые прямо следуют из Ω, но и те, которые получаются в результате аддитивно-транзитивного замыкания, т. е. учитывать отношение $\succcurlyeq^{at\Omega}$.

Пример 1.3. Пусть в четырехкритериальной задаче $\Omega = \{1\succ 2,\ 24\succ 34\}$. Эта информация непротиворечива. При такой информации о важности $y = (2, 1, 1)$ и $z = (1, 1, 2)$ не сравнимыми по предпочтительности, даже если попытаться расширить Ω за счет транзитивного замыкания отношения $\succcurlyeq^{\varnothing} = \{(1, 2), (24, 34)\}$. Однако, с учетом аддитивности, верно $2\succ^{at\Omega}3$, а по транзитивности с учетом $1\succ^{at\Omega}2$ получаем $1\succ^{at\Omega}3$. Следовательно, нужно считать, что y предпочтительнее, чем z.

Пополним множество Ω, добавив для него соответственно сообщение $A \succ B$, если $A \succ^{at\Omega} B$, и $A \sim B$, если $A \sim^{at\Omega} B$. Обозначим его $\overline{\Omega}$. Это множество порождает на Z отношение $R^{\overline{\Omega}}$ согласно (2), где вместо Ω стоит $\overline{\Omega}$. По построению, $R^{\Omega} \subseteq R^{\overline{\Omega}}$. Отметим, что $I^{\Omega} \subseteq I^{\overline{\Omega}}$.

Теорема 1.2. Информация Ω непротиворечива если и только если $P^{\Omega} \subseteq P^{\overline{\Omega}}$.

Согласно теореме 1.1 информация Ω противоречива тогда и только тогда, когда можно построить цикл вида (1.5), где $z = y$ и каждое ω_k есть соотношение вида $A \succ B$ или $A \sim B$ из $\overline{\Omega}$ или же символ $\varnothing$, причем хотя бы раз встречается $P^{A \succ B}$ или $P^{\varnothing}$. Теорема 1.2 показывает, что определение непротиворечивости информации о важности полностью согласовано с определениями непротиворечивости, принятыми в теории важности критериев [5, 6, 12].

Теперь пополним множество Ω так: добавим в него соответственно сообщение $A \succ B$, если $A \succ^{at\Omega} B$, и $A \sim B$, если $A \sim^{at\Omega} B$, только если множества A и B не пусты и не пересекаются. Обозначим его через $\widehat{\Omega}$. Это множество порождает на Z отношение $R^{\widehat{\Omega}}$ согласно (1.4), где вместо Ω стоит $\widehat{\Omega}$.

Теорема 1.3. Справедливо равенство $R^{\overline{\Omega}} = R^{\widehat{\Omega}}$.

Построение отношений R^{Ω}, $\succcurlyeq^{at\Omega}$, $R^{\overline{\Omega}}$ и $R^{\widehat{\Omega}}$ основано на операции транзитивного или аддитивно-транзитивного замыкания бинарных отношений. Если размерность задачи «не очень велика», то указанные отношения можно построить, если использовать матричное представление бинарных отношений.

Бинарное отношение ρ на конечном множестве $H = \{h^1, h^2, \ldots, h^q\}$ можно задать квадратной булевой матрицей смежности $B(\rho) = \|b_{ij}(\rho)\|$ порядка q, где

$$b_{ij}(\rho) = \begin{cases} 1 \text{ при } (h^i, h^j) \in \rho, \\ 0 \text{ при } (h^i, h^j) \notin \rho. \end{cases} \quad (1.11)$$

Для булевых матриц определены операции: транспонирование ^{T}, сложение +, поэлементное умножение $*$, умножение $\times$, образование асимметричной разности \:

$$\|a_{ij}\| \setminus \|b_{ij}\| = \|c_{ij}\|, \text{ где } \|c_{ij}(\rho)\| = \begin{cases} 1 & \text{при } a_{ij} = 1 \text{ и } b_{ij} = 0, \\ 0 & \text{в остальных случаях.} \end{cases} \quad (1.12)$$

Операциям над бинарными отношениями – взятия обратного отношения $^{-1}$, пересечения отношений $\cap$, объединения отношений $\cup$, разности отношений \, умножения отношений

$\circ$ и построения транзитивного замыкания $\hat{\rho}$ соответствуют операции над их матрицами смежности [20]:

$$B(\rho^{-1}) = B(\rho)^T,\ B(\rho^1 \cap \rho^2) = B(\rho^1) \circ B(\rho^2),\ B(\rho^1 \cup \rho^2) = B(\rho^1) + B(\rho^2),$$
$$B(\rho^1 \setminus \rho^2) = B(\rho^1) \setminus B(\rho^2),\ B(\rho^1 \circ \rho^2) = B(\rho^1) \times B(\rho^2),\ B(\hat{\rho}) = B(\rho) + B^2(\rho) + \ldots + B^n(\rho). \quad (1.13)$$

Поэтому $B(\mathrm{Sym}\ \rho) = B(\rho) \circ B(\rho)^{\mathrm{T}}$, $B(\mathrm{As}\ \rho) = B(\rho) \setminus B(\mathrm{Sym}\ \rho)$, где Sym и As – символы операций выделения симметричной и асимметричной частей бинарного отношения ρ соответственно: $\mathrm{Sym}\ \rho = \rho \cap \rho^{-1}$, $\mathrm{As}\ \rho = \rho \setminus \mathrm{Sym}\ \rho$. Для построения транзитивного замыкания существуют эффективные алгоритмы [21]. Заметим еще, что

$$\rho^1 \subseteq \rho^2 \Leftrightarrow B(\rho^1) \leqq B(\rho^2),\ \rho^1 \subset \rho^2 \Leftrightarrow B(\rho^1) \leq B(\rho^2), \quad (1.14)$$

где знак матричного неравенства $\leqq$ для двух матриц означает, что каждый элемент первой матрицы не больше соответствующего элемента второй матрицы, а знак $\leq$ показывает, что верно указанное условие, но матрицы не равны: в первой матрице найдется элемент, который меньше соответственного элемента второй матрицы.

Пример 1.4. Полную качественную вероятность $\gtrsim$ из примера 1.1 можно интерпретировать как отношение важности $\succcurlyeq^{\Omega}$ на множестве групп критериев, причем $\succcurlyeq^{\Omega} = \succcurlyeq^{\mathrm{at}\Omega}$. Информация Ω непротиворечива. При помощи специальной компьютерной программы для этого отношения важности для случая, когда $Z_0 = \{1, 2\}$, было построено отношение $R^{\overline{\Omega}}$. Оно оказалось полным (аддитивным) квазипорядком. Но уже для $Z_0 = \{1, 2, 3\}$ квазипорядок $R^{\overline{\Omega}}$ оказывается лишь частичным (см. пример 1.5).

Пример 1.5. Частичную качественную вероятность $\succcurlyeq$ из примера 2 можно интерпретировать как отношение важности, полученное путем аддитивно-транзитивного замыкания для информации $\Omega = \{13 \succ 24,\ 14 \succ 23\}$. Информация Ω непротиворечива. Здесь $\overline{\Omega} = \Omega$. При помощи специальной компьютерной программы для случая, когда $Z_0 = \{1, 2\}$, было построено отношение $R^{\overline{\Omega}}$. Оно оказалось частичным (аддитивным) квазипорядком.

1.1.8 Об одном претенденте на аналитическое решающее правило

Пусть y и z – две произвольные векторные оценки. Пусть $W(y, z)$ – множество, состоящее из попарно различных компонент этих векторов, элементы которого перенумерованы в порядке их возрастания:

$$W(y, z) = \{y_1\} \cup \{y_2\} \cup \ldots \cup \{y_m\} \cup \{z_1\} \cup \{z_2\} \cup \ldots \cup \{z_m\} = \{w_1, w_2, \ldots, w_q\}, \quad (1.15)$$

где $q \leqslant m$ и зависит от y и z, $w_1 < w_2 < \ldots < w_q$. Пусть, далее, $M_k(y) = \{i \in M \mid y_i \geqslant w_k\}$, $k = 2, 3, \ldots, q$. Под $\succcurlyeq^{\bar{\Omega}}$ будем понимать $\succ$ или $\sim$ из соотношений между группами критериев в $\bar{\Omega}$.

Теорема 1.4. Справедливо утверждение: $yR^{\bar{\Omega}}z \Rightarrow M_k(y) \succcurlyeq^{at\Omega} M_k(z)$, $k = 2, 3, \ldots, q$.

Утверждение в теореме 1.4 напоминает решающее правило для количественной важности, когда имеются коэффициенты важности для всех отдельных критериев [5]). Однако, в отличие от случая отдельных критериев, теорема 4 дает только лишь необходимые, но не достаточные условия для справедливости соотношения $yR^{\bar{\Omega}}z$. Это доказывает следующий контрпример.

Пример 1.6. Пусть в пятикритериальной задаче $Z_0 = \{1, 2, 3, 4\}$ и $\Omega = \{23 \succ 4, 13 \succ 45, 1 \succ 4\}$. Эта информация непротиворечива. Так как в сообщениях из Ω критерии 1, 2 и 3 находятся слева, а 4 и 5 справа, то с помощью аддитивно-транзитивного замыкания в $\bar{\Omega}$ не удаётся ввести никаких новых сообщений вида $A \succ B$ с непересекающимися A и B. Сравним векторные оценки $y = (4, 2, 3, 1, 2)$ и $z = (2, 1, 1, 4, 3)$. Все соотношения между $M_k(y)$ и $M_k(z)$ из теоремы 4 выполняются:

при $k = 2$: $M_2(y) = \{1, 2, 3, 5\} \succ^{at\Omega} M_2(z) = \{1, 4, 5\}$, ибо $23 \succ^{\bar{\Omega}} 4$;

при $k = 3$: $M_3(y) = \{1, 3\} \succ^{at\Omega} M_3(z) = \{4, 5\}$, ибо $13 \succ^{\bar{\Omega}} 45$;

при $k = 4$: $M_4(y) = \{1\} \succ^{at\Omega} M_4(z) = \{4\}$, ибо $1 \succ^{\bar{\Omega}} 4$.

Однако, как показали расчеты, выполненные при помощи компьютерной программы с использованием матричного представления бинарных отношений, векторные оценки y и z несравнимы по $R^{\bar{\Omega}}$.

1.1.9 Некоторые выводы

Представленные в отчете результаты позволяют анализировать прикладные многокритериальные задачи, когда имеется информация об относительной важности не только отдельных критериев, но и групп критериев с порядковой шкалой. Они являются базой для дальнейшего развития теории важности критериев для задач с упорядоченными по важности группами критериев, имеющих более совершенные, чем порядковая, шкалы.

Актуальной остается проблема разработки достаточно простого аналитического правила, которое позволило бы эффективно сравнивать по предпочтительности варианты с учетом информации о важности в задачах с большим числом критериев и вариантов.

Поставленные перед началом исследований задачи выполнены полностью. Исследования проводились совместно с к.ф.-м.н. Нелюбиным А.П. Полученные результаты о средних опубликованы в статьях [22, 23].

1.2 Исследование вычислительных свойств сепарабельных предпочтений

1.2.1 Постановка задачи

Данный подраздел описывает результаты, полученные в [24]. В данном подразделе исследуются вычислительные свойства задач коллективного выбора для случая, когда предпочтения избирателей являются сепарабельными. В этом мы следуем устоявшемуся направлению исследований, в котором авторы предполагают, что предпочтения избирателей хорошо структурированы — например, являются однопиковыми [25] или имеют единственное пересечение [26, 27] — и показывают, что различные NP-трудные задачи, связанные с выборами либо становятся разрешимыми за полиномиальное время, либо остаются вычислительно неразрешимыми (см., например, обзор [28]). До сих пор сепарабельным предпочтениям уделялось ограниченное внимание в литературе.

Рассматривается ординальная модель предпочтений, то есть предполагаем, что каждый избиратель ранжирует кандидатов от наиболее до наименее привлекательных. В общем, неограниченном случае каждый избиратель волен сообщить о любом таком порядке предпочтения. Предполагая, что порядок предпочтений каким-то образом структурирован, мы требуем, чтобы избиратели следовали некоторым общим, рациональным критериям. Например, Блэк [25] ввел понятие однопиковых предпочтений, когда все избиратели соглашаются с некоторым порядком кандидатов по заданному критерию (этот порядок называется социальной осью; на политических выборах он может, например, соответствовать стандартному принципу "слева направо"), каждый избиратель имеет свою позицию на оси (т. е. каждый избиратель волен выбрать любого из кандидатов в качестве наиболее предпочтительного), и, рассматривая двух кандидатов, которые находятся на одной стороне оси по отношению к любимому кандидату избирателя, избиратель предпочитает того, кто ближе к любимому кандидату. Например, если избиратель ставит центристского кандидата на первое место, то он или она вынуждены предпочесть умеренного левого кандидата крайне левому. Предпочтения с единственным пересечением, введенные Миррлисом [26] и Робертсом [27], следуют аналогичному принципу, но с акцентом на упорядочивании избирателей. В частности, можно упорядочить избирателей таким образом, чтобы при рассмотрении избирателей в таком порядке относительный рейтинг каждых двух кандидатов менялся не более одного раза. Например, рассмотрим ситуацию, когда избиратели распределены от самого левого к самому правому, и рассмотрим двух кандидатов, x и y, где x более левый, а y более правый. Первый

избиратель предпочел бы x y, последний избиратель предпочел бы y x, и по мере того, как мы переходили от первого избирателя до последнего, мы бы наблюдали только одну точку, где изменился относительный рейтинг x и y.

Сепарабельные предпочтения, введенные Инадой [29, 30], также отражают определенное предположение о рациональности, но совершенно другое. Формально мы говорим, что избиратели имеют сепарабельные предпочтения, если каждое подмножество A кандидатов, $A \geq 2$, может быть разделено на две части, A и A', таким образом, что каждый избиратель предпочитает все элементы одного из этих наборов всем элементам другого. Одна из интерпретаций этого заключается в том, что существует иерархия атрибутов. Из-за последовательного разделения кандидатов групповые предпочтения также называются последовательно дихотомическими (этот термин используется в литературе о случайных назначениях, см., например, работу [31]).

Чтобы проиллюстрировать вышеизложенную интуицию, давайте рассмотрим пример, в котором избиратели отдают предпочтение пяти напиткам: кофе, зеленому чаю, черному чаю, соку и газировке. Есть три избирателя со следующими предпочтениями (мы пишем $x \succ y$, чтобы указать, что избиратель ценит x выше y):

И1 :кофе $\succ$ черный чай $\succ$ зеленый чай $\succ$ соды $\succ$ сок,

И2 :зеленый чай $\succ$ черный чай $\succ$ кофе $\succ$ сок $\succ$ соды,

И3 :соды $\succ$ сок $\succ$ черный чай $\succ$ зеленый чай $\succ$ кофе.

Чтобы убедиться, что эти предпочтения можно разделить на группы, мы делаем следующие наблюдения. Во-первых, если мы рассмотрим набор всех напитков, то каждый избиратель либо предпочитает все горячие напитки всем холодным, либо наоборот. Во-вторых, среди горячих напитков все избиратели либо предпочитают кофе обоим видам чая, либо наоборот. Другими словами, все избиратели согласны с тем, что напитки могут быть классифицированы так, как показано на рисунке 1.1, и соответственно формируют свои предпочтения (такие деревья, как на рисунка 1.1 уже обсуждались [29]). В частности, ни один избиратель не поставил бы кофе выше сока, чем зеленый чай, потому что это поставило бы холодный напиток между двумя горячими.

Разбиение на группы приводит к иерархическому дереву, в котором листьями являются альтернативы. В теории графов такие деревья называются деревьями Шредера. Пример иерархическое дерева альтернатив представлен на рисунке 1.1.

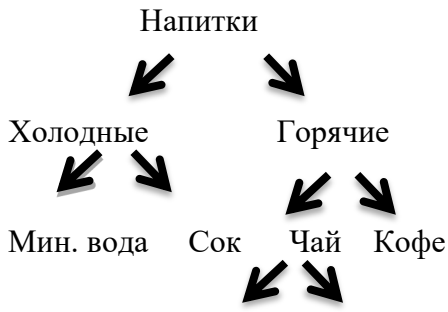


Рисунок 1.1 − Иерархическое дерево альтернатив

Иерархическое дерево альтернатив можно построить как для профиля предпочтений, так и для домена. В последнем случае дерево будет бинарным. Алгоритм построения дерева подробно описан в [32].

Отметим, что дерево, представленное на рисунке 1.1, по сути, является деревом декомпозиции клонов для профиля предпочтений; действительно, все кандидаты в данной ветви дерева являются клонами в том смысле, что все избиратели ранжируют их последовательно. Предположение, что у избирателей есть предпочтения с одним пиком (или с одним пересечением, или структурированные каким-либо другим привлекательным способом), имеет различные положительные последствия.

Например, если число избирателей нечетное, то всегда существует победитель Кондорсе (т. е. кандидат с, которому большинство избирателей отдает предпочтение перед любым другим кандидатом). Кроме того, многие NP-трудные проблемы, связанные с выборами, становятся разрешимыми за полиномиальное время. Хотя хорошо известно, что сепарабельные предпочтения ведут себя одинаково хорошо с точки зрения их нормативных свойств — и, в частности, они также гарантируют существование победителя Кондорсе — мы показываем, что их вычислительные свойства более важны. С этой целью мы рассматриваем проблемы определения победителя в соответствии с правилом голосования с несколькими победителями Чемберлина – Куранта [33] и правилом голосования с одним победителем Янга [34], а также два вида проблем контроля за выборами. Мы выбрали эти задачи, потому что все они являются NP-трудными в общих условиях, но становятся разрешимыми за полиномиальное время, если предпочтения избирателей однопиковы (или являются предпочтениями с единственным пересечением).

Многие (хотя и не все) NP-трудные проблемы, связанные с выборами, остаются неразрешимыми, даже если у избирателей есть сепарабельные предпочтения. Тем не менее, часто эти проблемы могут быть решены более эффективно (часто за полиномиальное время), если дерево декомпозиции клонов для данного входного выбора имеет умеренную высоту.

Мы также показываем алгоритм полиномиального времени для выборки сепарабельных профилей предпочтений (с заданным числом кандидатов и избирателей) равномерно случайным образом. Такой алгоритм очень полезен, если кто-то заинтересован в проведении численных экспериментов, касающихся сепарабельных предпочтений.

1.2.2 Вычислительная сложность

В данном пункте мы исследуем вычислительную сложность классических проблем вычислительной теории коллективного выбора при сепарабельных предпочтениях.

Согласно правилу Янга победителем является альтернатива, которая побеждает все остальные альтернативы в попарном мажоритарном сравнении при исключении минимального количества агентов, при этом количество исключенных агентов, приводящих к победе альтернативы называется баллами Янга.

Теорема 1.5 [24]. *Задача нахождение баллов Янга при сепарабельных предпочтениях принадлежит* классу *NP-трудных*.

Правило Чемберлина − Куранта выбирает комитет размера k с наименьшей суммой рангов согласно индивидуальным предпочтениям, где 1 соответствует наилучшей альтернативе в предпочтении агента. Задача *CC-SCORE* «Существует ли комитет размера k с суммой рангов не более t» принадлежит к классу NP-полных в общем случае, но упрощается до полиномиальной сложности в случае однопиковых предпочтений. В случае сепарабельных предпочтений такого упрощения не происходит.

Теорема 1.6 [24]. *Задача CC-SCORE при сепарабельных предпочтениях принадлежит классу NP-трудных.*

Кроме того, задачи манипулирования результатом голосования с помощью добавления/удаления кандидатов, добавления/удаления агентов являются NP-трудными. Для всех упомянутых в данном разделе задач существует алгоритм полиномиальной сложности если ограничить разнообразие предпочтений, задав максимально возможную высоту иерархического дерева, представляющего профиль предпочтений.

1.2.3 Генерирование сепарабельных предпочтений

Мы приводим процедуру выборки групповых выборов равномерно случайным образом. Другими словами, наша процедура реализует модель беспристрастной культуры, но ограничивается сепарабельными выборами. Такая процедура существует, например, для однопиковых выборов [35], но пока отсутствует для выборов с одним пересечением (см., например, работу [36] для обсуждения обходного пути). Выборка сепарабельных выборов желательна, если, например, кто-то хочет провести численные эксперименты в отношении этой области предпочтений. Описание процедуры:

Шаг 1 (порядок предпочтения v_1). Мы выбираем порядок предпочтений v_1 равномерно случайным образом.

Шаг 2 (количество внутренних узлов). Мы выбираем количество r внутренних узлов дерева декомпозиции, которое будет иметь наша выборка. Мы отбираем каждое значение r с вероятностью:

$$\Pr(r) = \frac{\binom{m-1}{r}\binom{m-1+r}{m}(2^{n-1}-1)^{r-1}}{\sum_{j=1}^{m-1}\binom{m-1}{j}\binom{m-1+j}{m}(2^{n-1}-1)^{j-1}}. \quad (1.16)$$

Шаг 3 (дерево декомпозиции). Мы выбираем (равномерно случайным образом) дерево декомпозиции для m кандидатов и с r внутренними узлами. То есть мы выбираем упорядоченное корневое дерево с r внутренними узлами и m листьями. Такие деревья известны как деревья Шредера, и алгоритм для их выборки был предоставлен Алонсо и др. [32, 37].

Шаг 4 (выборы). Мы выбираем порядок предпочтений для избирателей v_2, ..., v_n, используя кодировку, описанную в [24].

1.2.4 Некоторые выводы

До сих пор большинство алгоритмических исследований ограниченных областей были сосредоточены на положительных результатах и обеспечивали эффективные алгоритмы. Мы показали, что сепарабельные предпочтения несколько отличаются.

Известно, что они обладают хорошими нормативными свойствами, но для получения эффективных алгоритмов необходимо заглянуть глубже в его структуру. В частности, высота дерева декомпозиции клонов для данных выборов сильно связана с тем, насколько сложно решить ряд проблем, связанных с этими выборами. Кроме того, мы также предоставили алгоритм для выборки групповых выборов равномерно случайным образом, и как часть нашей процедуры, мы предоставили алгоритм для выборки выборов с заданным деревом декомпозиции. Мы считаем, что эти алгоритмы облегчат экспериментальные исследования групповых выборов.

Существует ряд открытых проблем, которые вытекают из нашей работы. Действительно, существует много других NP-трудных проблем с выборами, которые мы не рассматривали и сложность которых в предположении о сепарабельности было бы интересно исследовать. Такие проблемы включают, например, проблему вычисления балла кандидата на выборах Доджсона и проблему вычисления победившего комитета по правилу пропорционального представительства при одобряющем голосовании.

1.3 Исследование турнирных способов выбора оптимальных вариантов на основе минимального покрывающего множества

1.3.1 Введение

Проблема оптимального выбора заключается в выборе в некотором смысле лучших альтернатив из множества, предъявленного выбирающему. Важным частным случаем является выбор, основанный на результатах попарного сравнения вариантов. Подобную проблему можно представить как задачу выбора победителя или победителей в спортивном турнире.

В литературе предложено довольно много различных решений данной задачи, а именно турнирных решений. Свойства и теоретико-множественные отношения этих решений хорошо исследованы в случае, когда орграф, соответствующий турниру, является полным. В практических задачах особенно важно наличие у турнирного решения устойчивости. Данным свойством обладают лишь некоторые турнирные решения, что придаёт им особую ценность. Благодаря наличию устойчивости одним из кандидатов на роль наилучшего турнирного решения является так называемое минимальное покрывающее множество (minimal covering set), введённое в [38].

В практических задачах часты ситуации, когда в предъявлении есть пары равноценных альтернатив (несравнимые альтернативы без потери общности можно считать равноценными). В связи с этим возникает необходимость в обобщении известных турнирных решений, на случай выбора победителя или же победителей в турнирах, чей орграф неполон (в слабых турнирах).

Рассматриваются три различные версии минимального покрывающего множества, предложенные в качестве обобщений уже упомянутого минимального покрывающего множества.

Основными задачами исследования, были, во-первых, проверка сохранения рассматриваемыми версиями ряда важных свойств минимального покрывающего множества (устойчивости, монотонности и вычислительной простоты), во-вторых, установление теоретико-множественных отношений с обобщениями других турнирных решений, и в-третьих, проверка на реальных данных качества ранжирований, полученных с помощью этих решений.

В результате проведённого исследования получено достаточно полное описание свойств рассматриваемых обобщений минимального покрывающего множества и описана их связь с различными обобщениями других турнирных решений. Также в ходе исследования был создан комплекс программ, позволяющий производить парсинг данных с сайтов, вычислять рассматриваемые турнирные решения, ранжировать альтернативы, создавать и выво-

дить аналитическую отчетность. Полученная программа применялась, во-первых, для проверки примеров и поиска контрпримеров для доказательства теорем, во-вторых, для реализации эмпирической части исследования.

В пункте 1.3.2 рассматривается описание теоретической модели, приводятся основные определения. Формулировки заимствованы в основном из [39].

В пункте 1.3.3 сформулированы и доказаны теоремы, показывающие наличие или отсутствие у рассматриваемых обобщений важных свойств турнирных решений, определённых в пункте 1.3.2. Также в ней приводится таблица, содержащая в себе информацию о связи версий минимального покрывающего множества с обобщениями других известных турнирных решений.

В пункте 1.3.4 приведен результат сравнения сортировки с помощью рассматриваемых решений с пятью другими методами построения ранжирований, основанных на турнирных решениях. Методы сравниваются на основе корреляционного анализа ранжирований высших учебных заведений (далее для обозначения данного понятия будет использоваться общепринятая аббревиатура «вуз»). Методология проведения анализа заимствована из [40].

В пункте 1.3.5 кратко резюмируются основные результаты и даётся оценка рассмотренных обобщений минимального покрывающего множества как инструментов оптимального выбора.

1.3.2 Общие теоретические сведения

1.3.2.1 Выбор и предпочтения

Объекты выбора составляют *универсальное множество* A всех потенциально доступных для выбора альтернатив, $|A| > 2$. *Предъявлением* называется конечное непустое множество X, $X \subseteq A$. В задаче выбора X является переменной величиной и представляет из себя множество альтернатив, актуально доступных для выбора. Выбор есть разделение предъявления на два подмножества – подмножество выбранных и подмножество отвергнутых альтернатив. *Функция выбора* $S(X)$ – это функция, которая каждому предъявлению ставит в соответствие подмножество выбранных альтернатив $S \subseteq X$.

Выбор из пары однозначно определяет на A бинарное отношение $\succsim$.

Асимметричную часть $\succ$ этого отношения составляют упорядоченные пары (a, b) элементов A такие, что при выборе из $X = \{a, b\}$ выбирается a и отвергается b, то есть вариант a строго предпочитается варианту b. Из этого определения следует, что отношение $\succ$ является полным, $\forall a, b \in A,\ a \succ b \vee b \succ a$, антирефлексивным, $\forall a \in A, \neg a \sim a$, и асимметричным $\forall a, b \in A,\ a \succ b \Rightarrow \neg b \succ a$. Оно репрезентирует строгие предпочтения субъекта выбора.

Симметричную часть $\sim$ отношения $\succsim$ составляют упорядоченные пары (a, b) элементов A такие, что при выборе из $X = \{a, b\}$ альтернативы a и b рассматриваются как равноценные. Без потери общности можно считать, что в этом случае выбираются обе альтернативы.

Из этих определений следует, что отношение $\succsim$ является полным, $\forall\, a, b \in A,\ a \succsim b \vee b \succsim a$, рефлексивным, $\forall\, a \in A,\ a \sim a$, и антисимметричным, $\forall\, a, b \in A,\ (a \succsim b \wedge b \succsim a) \Rightarrow a = b$. Отношение $\succsim$ представляет нестрогие предпочтения выбирающего.

Задача оптимального выбора из предъявления, в котором больше двух альтернатив, имеет естественное решение в том случае, когда у сужения отношения $\succsim$ на данное предъявление есть *максимальный элемент* (или максимальные элементы), то есть альтернатива, которая не хуже, чем любая другая альтернатива, доступная для выбора. Однако хорошо известно, что в задаче коллективного выбора может получаться отношение $\succsim$ без максимальных элементов, что является следствием наличия циклов в его асимметричной части $\succ$. Данный результат известен как парадокс Кондорсе. Следствием данного парадокса является необходимость определить иной, нетривиальный способ выбора наилучших альтернатив, каковым и является любое турнирное решение.

1.3.2.2 Турнир и турнирные решения

Турниром T называют пару $(X, \succ_X)$, где $\succ_X$ обозначает сужение отношения $\succ$ на предъявление X. Турнир $T(X, \succ_X)$ можно представить полным ориентированным графом, вершины которого представляют элементы множества X, а дуги – элементы множества $\succ_X$. *Слабым турниром* W называют пару $(X, \succsim_X)$. Граф слабого турнира, в отличие от графа турнира, может быть неполным в случае, если $\exists (a, b \in X \wedge a \neq b)\colon a \sim b$.

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

1) $\forall T,\ S(T) \neq \emptyset$ (*непустота*);
2) $\forall T,\ MAX(\succsim_X) \subseteq S(T)$, где $MAX(\succsim_X)$ – множество максимальных элементов $\succsim_X$ (*согласованность по Кондорсе*);
3) $S(\pi(X)) = \pi(S(X))$, где $\pi(X)$ – произвольная перестановка альтернатив из X (*нейтральность*).

1.3.2.3 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 $\succsim_X$ и $\succsim'_X$ совпадают для всех пар, не содержащих альтернативу a: $\succsim_{X\setminus\{a\}} = \succsim'_{X\setminus\{a\}}$. Пусть $\forall b \in X\setminus\{a\}, (a \succsim b \Rightarrow a \succsim' b) \wedge (a \succ b \Rightarrow a \succ' b)$. Тогда решение S является монотонным, если из того, что $a \in S(\succsim_X)$ следует, что $a \in S(\succsim'_X)$.

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:

$$S(Y, \succsim_Y) = S(Z, \succsim_Z) = X \Leftrightarrow X = S(Y \cup Z, \succsim_{Y\cup Z}). \quad (1.17)$$

Это утверждение можно рассматривать как конъюнкцию прямой и обратной импликаций. Прямая импликация в работе [41] называется свойством $\hat{\gamma}$, а обратная - свойством $\hat{\alpha}$ (в литературе у свойства $\hat{\alpha}$ есть множество различных названий, в частности, обобщённая независимость от посторонних альтернатив по Нэшу, сильное свойство надмножеств или SSP (от англ. strong superset property) [42], свойство отбрасывания [43].

3) Композиционная согласованность.

Орграф 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

$$S(T) = \bigcup_{i\in S(\tilde{T})} S(T_i). \quad (1.18)$$

В данном исследовании рассматривается минимальное покрывающее множество и его обобщения на случай слабого турнира. Остальные решения рассматриваются только в связи с этим множеством и его обобщениями.

1.3.2.4 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) $\forall x \notin Y, x \notin UC(Y \cup \{x\})$

В связи с тем, что субъект выбора обычно предпочитает сузить множество выбираемых альтернатив, оптимальным решением будет являться минимальное по включению покрывающее множество. Можно показать, что такое множество будет единственным. Получившее турнирное решение было названо минимальным покрывающим множеством МС [38].

Известно, что минимальное покрывающее множество является монотонным, устойчивым и композиционно-согласованным, а также обладает вычислительной простотой [39].

Стоит отметить, что *MC* обладает важной теоретико-игровой интерпретацией. В [44] было доказано, что данное турнирное решение совпадает со слабой седловой точкой [45] соответствующей турнирной игры.

Как видно из определения, минимальное покрывающее множество связано с непокрытым множеством, что приводит к мысли о возможном обобщении покрывающего множество на случай слабых турниров через уже известные версии непокрытого множества (покрывающие по Миллеру и Фишберну множества были введены в [46]).

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

1) $UC_V(Y) = Y$;

2) $\forall x \notin Y, x \notin UC_V(Y \cup \{x\})$, где $V = \{M, IM, F, IF, McK, IMcK, D, ID\}$.

К сожалению, далеко не все версии непокрытого множества приводят к решению, пригодному для использования. Так в слабых турнирах возможно существование сразу нескольких минимальных покрывающих по Миллеру или Фишберну множеств. В случае же версии Миллера решение вообще может оказаться пустым [46]. Также можно показать, что минимальные покрывающие множества, определенные через слабую версию отношений покрытия по МакКельви и Дуггану, могут быть пустыми.

В связи с описанными выше недостатками далее не будут рассматриваться обобщения минимального покрывающего множества, связанные с версиями $C_{WM}, C_M, C_{WF}, C_F, C_{WMcK}, C_{WD}$ отношения покрытия.

С оставшимися версиями (C_{McK}, C_D) ситуация лучше: минимальное покрывающее множество однозначно определяется как (единственное) минимальное по вложению покрывающее по соответствующей версии отношения покрытия множество. Известно, что MC_{McK} и MC_D являются устойчивыми, а MC_{McK} – монотонным [47] и вычислительно простым решением [46].

1.3.2.5 Другие турнирные решения и их обобщения

Для сравнения рассмотрим иные турнирные решения и некоторые их обобщения на случай слабых турниров.

Непокрытое множество (UC) основано на отношении покрытия, для определения которого, требуются понятия верхнего и нижнего срезов и горизонта данной альтернативы.

Верхний срез $D(x) = \{y \in X | y \succ x\}$. Нижний срез $L(x) = \{y \in X | x \succ y\}$. Горизонт $H(x) = \{y \in X | x \sim y\}$.

Существует 15 различных версий определения отношения покрытия [48], а следовательно, и непокрытого множества. Следует отметить, что первые три версии (Ia, Ib, Ic) не принято рассматривать в рамках изучения турнирных решений, так как они могут приводить к пустому выбору, что противоречит определению турнирного решения. Все версии отношения покрытия определяются в [48], в данном исследовании рассмотрение ограничивается версиями, предложенными Миллером, Фишберном, МакКельви и Дугганом, а также их слабыми аналогами (UC_M, UC_F, UC_{McK}, UC_D, UC_{IM}, UC_{IF}, UC_{IMcK}, UC_{ID}).

Оптимальный выбор можно определить как выбор альтернатив из «хорошего» подмножества предъявления. Множество называется минимальным относительно некоторого свойства, если само множество имеет данное свойство, но ни одно его строгое подмножество его не имеет.

Можно определить ряд турнирных решений как объединение минимальных подмножеств предъявления, обладающих определенным «хорошим» свойством. В литературе представлено несколько турнирных решений, определяемых обсуждаемым способом. В исследовании рассматриваются следующие из них: объединение минимальных доминирующих множеств D, объединение минимальных самозащищающихся множеств SP, объединение минимальных слабоустойчивых множеств WS [49], объединение минимальных внешне устойчивых множеств ES [50, 51, 52, 53], объединение минимальных слабо внешне устойчивых множеств RES [48, 54], слабый и сильный максимальный цикл (WTC и STC) [55, 56, 57, 58], незапертое множество UT, незахваченное множество UC_p [59].

Для определения двухпартийного множества требуется понятие максимальной лотереи. Лотереей $\boldsymbol{p}$ будем называть распределение вероятностей $p(x)$ над альтернативами из предъявления X, $\boldsymbol{p} = \{p(x) \in \mathbb{R} \mid p(x) \geq 0 \wedge x \in X \wedge \sum_{x \in X} p(x) = 1\}$. Определим функцию сравнения $g: X \times X \to \mathbb{R}$

$$g(x, y) = \begin{cases} 1, & x \succ y, \\ -1, & y \succ x, \\ 0, & x \sim y. \end{cases} \tag{1.19}$$

На множестве лотерей можно ввести бинарное отношение $\succcurlyeq$. Так $\mathbf{p}_1 \succcurlyeq \mathbf{p}_2 \Leftrightarrow \sum_{x\in X}\sum_{y\in X} p_1(x)g(x,y)p_2(y) \geq 0$. Максимальной лотереей называется максимальный элемент отношения $\succcurlyeq$. Двухпартийное множество BP определяется как носитель единственной максимальной лотереи $\boldsymbol{p}$ турнира [60].

В [61] было введено существенное множество E, являющееся обобщением двухпартийного множества на случай слабого турнира. Как известно, в слабом турнире, максимальная лотерея существует, но может не быть единственной. Существенное множество E определяется как объединение носителей всех максимальных лотерей слабого турнира.

Еще одним естественным обобщением любого турнирных решения S на случай слабого турнира является консервативное обобщение, введенное в [62], и обозначающееся $[S]$.

$$[S](W) = \bigcup_{T\in[W]} S(T). \tag{1.20}$$

1.3.3 Свойства версий минимального покрывающего множества и их связь с другими турнирными решениями

1.3.3.1 Свойства версий минимального покрывающего множества

В [62] показано, что консервативное обобщение наследует многие свойства исходного турнирного решения, из чего можно сделать вывод о том, что консервативное обобщение минимального покрывающего множества $[MC]$ удовлетворяет особо интересным свойствам монотонности и устойчивости, а также является композиционно-согласованным.

Несмотря на хорошие теоретические свойства, возможности использования данного турнирного решение сильно ограничены, так как до сих пор неизвестно, можно ли построить алгоритм, позволяющий вычислять $[MC]$ за полиномиальное время.

Как уже говорилось ранее, о MC_{McK} известно, что оно является устойчивым и монотонным [47]. Для версии Дуггана отношения покрытия была доказана лишь устойчивость.

Теорема 1.7. Минимальное покрывающее по Дуггану множество MC_D удовлетворяет свойству монотонности.

В [46] приведен полиномиальный алгоритм вычисления MC_{McK} через существенное множество. К сожалению, в данной статье не рассматривалась версия Дуггана отношения покрытия.

Теорема 1.8. Минимальное покрывающее по Дуггану множество MC_D обладает свойством вычислительной простоты.

Устойчивость, монотонность и вычислительная простота исследуемых турнирных решений дают основание считать их разумными инструментами выбора.

1.3.3.2 Теоретико-множественные отношения с другими турнирными решениями

Рассмотрим теоретико-множественные отношения минимального покрывающего по МакКельви множества, минимального покрывающего по Дуггану множества и консервативного обобщения минимального покрывающего множества с обобщениями других турнирных решений на случай слабых турниров.

Теорема 1.9.

1) MC_{McK} всегда является подмножеством MC_D, UC_{McK}, UC_D, $[D]$, $[ES]$, UCp, WTC. MC_{McK} всегда является надмножеством E. MC_{McK} не включено в $[BP]$, $[MC]$, UC_M, UC_{IF}, UC_F, D, SP, WS, ES, RES, STC, UT, но их пересечение всегда непусто. MC_{McK} не включено в UC_{IM}, причем их пересечение может быть пустым.

2) MC_D всегда является подмножеством UC_D, $[D]$, $[ES]$, WTC. MC_D всегда является надмножеством E, $[BP]$, MC_{McK}, $[MC]$. MC_D не включено в UC_{IM}, UC_M, UC_{IF}, UC_F, UC_{McK}, D, SP, WS, ES, RES, UCp, STC, UT, но их пересечение всегда непусто.

3) $[MC]$ всегда является подмножеством MC_D, UC_D, $[D]$, $[ES]$, WTC. $[MC]$ всегда является надмножеством $[BP]$. $[MC]$ не включено в MC_{McK}, UC_{IM}, UC_M, UC_{IF}, UC_F, UC_{McK}, D, SP, WS, ES, RES, UCp, STC, UT, но их пересечение всегда непусто.

Таблица 1.1 − Теоретико-множественные соотношения между турнирными решениями

	E	$[BP]$	MC_{McK}	MC_D	$[MC]$	UC_{IM}	UC_M	UC_{IF}	UC_F	UC_{McK}	UC_D	D	SP	$[D]$	WS	ES	RES	$[ES]$	UCp	STC	UT	WTC
MC_{McK}	$\cup^1$	$\supset$	$=$	$\subset$	$\supset$	$\varnothing$	$\supset^{\square}$	$\supset$	$\supset^{\square}$	$\subset^{\square}$	$\subset^{\square}$	$\supset$	$\supset$	$\subset$	$\supset$	$\supset$	$\supset$	$\subset$	$\subset^{\square}$	$\supset^{\square}$	$\supset^{\square}$	$\subset^{\square}$
MC_D	$\cup$	$\cup$	$\cup$	$=$	$\cup^2$	$\supset$	$\supset$	$\supset$	$\supset$	$\supset$	$\subset$	$\supset$	$\supset$	$\subset$	$\supset$	$\supset$	$\supset$	$\subset$	$\supset$	$\supset$	$\supset$	$\subset$
$[MC]$	?	$\cup$	$\supset$	$\subset^3$	$=$	$\supset$	$\supset$	$\supset$	$\supset$	$\supset$	$\subset$	$\supset$	$\supset$	$\subset$	$\supset$	$\supset$	$\supset$	$\subset$	$\supset$	$\supset$	$\supset$	$\subset$

В Таблице 1.1 символ “$\subset$” означает, что турнирное решение, расположенное в строке, всегда содержится в турнирном решении, расположенном в соответствующем столбце. Знаком “$\varnothing$” обозначены независимые турнирные решения, то есть решения, пересечение которых может оказаться пустым, а символ “$\supset$” обозначает турнирные решения, пересечение которых всегда непусто. Знаком вопроса “?” обозначена пара решений, для которых пока не найдено доказательство вложенности или примера, показывающего ее отсутствия. Равенством обозначаются эквивалентные решения.

Заметим, что часть утверждений уже были доказаны ранее. Так отношения с верхним индексом 1 доказаны в [61], с индексом 2 – в [63] и с индексом 3 – в [62].

1.3.4 Эмпирическое сравнение методов ранжирования

В предыдущем разделе показано, что с точки зрения теории разумно использовать минимальное покрывающее по МакКельви множество и минимальное покрывающее по Дуггану множество как инструменты выбора. Однако для оценки реальной полезности методов следует сравнить результаты их применения в практической задаче с результатами, которые дают другие процедуры выбора.

Одной из самых больших проблем в жизни практически любого ученика старшей школы является выбор вуза для поступления и дальнейшего обучения. Международные рейтинги призваны помочь будущим студентам в решении этой задачи. К сожалению, проблема не является тривиальной, что связано с разнообразием используемых методик построения рейтингов. Так ежегодно огромное количество организаций проводят совершенно разное оценивание вузов, из-за чего на выходе получается множество рейтингов, не согласующихся между собой.

В статье [40], лежащей в основе данной части исследования, была показана целесообразность использования турнирных решений в задаче построения агрегированных ранжирований применительно к библиометрическим данным. В данном исследовании используется схожий подход для построения агрегированных ранжирований шести международных рейтингов вузов.

Для решения поставленной задачи была написана программа на языке программирования Python, в функционал которой входят все этапы анализа от парсинга данных с сайтов и вычисления турнирных решений до вывода отчетности. Все реализованные в программе алгоритмы для нахождения турнирных решений полиномиальные, что дает возможность использовать ее на любых объемах данных.

1.3.4.1 Описание используемых международных рейтингов вузов

Помимо трех самых популярных и авторитетных рейтингов (QS, THE и ARWU) и их агрегата ARTU рассматриваются и появившиеся сравнительно недавно CWUR и RUR. Такой выбор был сделан с целью внесения максимального разнообразия в начальные данные.

Почти все существующие рейтинги представляют из себя некоторые семейства, включающие общий международный рейтинг, предметные рейтинги, рейтинги молодых университетов и т. п. Интересен только первый из перечисленных. Стоит отметить, что даты публикаций рейтингов не совпадают, из-за чего год в названии рейтинга может не отражать реальной ситуации. В Таблице 1.2 приведены сокращениям и понимаемые под ними официальные названия рейтингов вузов.

Таблица 1.2 − Источники данных

Сокращенное название	Полное название	Дата публикации
QS	QS 2022 World University Ranking	Июнь 2021
THE	THE 2022 World University Ranking	Сентябрь 2021
ARWU	ARWU 2021 World University Ranking	Август 2021
ARTU	ARTU 2021 World University Ranking	Ноябрь 2021
CWUR	CWUR 2021-2022 World University Ranking	Апрель 2021
RUR	RUR 2021 World University Ranking	Апрель 2021

В данном пункте будут приведены методики построения каждого из обсуждаемых рейтингов, а также отмечены некоторые их особенности.

1.3.4.1.1 Рейтинг QS

QS является рейтингом, ежегодно составляемым компанией Quacquarelli Symonds. С 2004 года до 2009 данный рейтинг публиковался совместно с рейтингом Times Higher Education (THE), но в 2010 году методология формирования последнего была изменена, из-за чего QS и THE стали независимыми. Методология составления рейтинга описано в [64].

Несмотря на свою высокую популярность, рейтинг довольно часто подвергается критике, связанной с используемым методом оценивания университетов, основанным на различных опросах. Далеко не всех экспертов устраивает излишний субъективизм в проставляемых оценках.

1.3.4.1.2 Рейтинг Times Higher Education (THE)

THE – ежегодно публикуемый одноименным британским журналом международный рейтинг университетов. Как уже говорилось ранее, самостоятельная история рейтинга началась в 2010 году после его отделения от QS.

THE оценивает вузы по пяти группам индикаторов: преподавание, исследования, цитирования, международные перспективы, доход от индустрии. Окончательная оценка получается в результате линейной свертки показателей с весами $0.3, 0.3, 0.3, 0.075, 0.025$ соответственно. Вузы сортируются в порядке убывания результата, перевод оценки в стобалльную шкалу не производится [65].

Как и в случае QS, высокая популярность и авторитетность рейтинга не означает отсутствия критики. Эксперты выражают недовольство так называемым репутационным исследованием, проводимым в виде опросов и имеющим довольно большой вес в группах «преподавание» и «исследования». Естественно, вопросы возникают к обоснованности субъективизма в оценках.

1.3.4.1.3 Academic Ranking of World Universities (ARWU)

Academic Ranking of World Universities, более известный как Шанхайский рейтинг, публикуется с 2003 года и является одним из самых значимых рейтингов высших учебных заведений в мире. Методология построения рейтинга приведена в [66].

К сожалению, данному рейтингу не удалось избежать критики. Сомнению подвергаются довольно экзотические индикаторы, сильно завязанные на международные премии. Кроме того, долгое время шли споры о воспроизводимости результатов рейтинга с помощью официальной методики его формирования [67, 68].

1.3.4.1.4 Рейтинг Aggregate Ranking of Top Universities (ARTU)

Рейтинг ARTU с 2019 года ежегодно предоставляется университетом Нового Южного Уэльса (UNSW Sydney). ARTU агрегирует три самых авторитетных рейтинга: QS, THE, ARWU. Правило, по которому производится агрегирование, является очень простым и очевидным: берется сумма от мест, на которые университет определил каждый рейтинг, после чего вузы располагаются в порядке увеличения результата [69].

$$ARTU\ score = QS + THE + ARWU. \tag{1.21}$$

Мотивацией для рассмотрения данного рейтинга послужил интерес в сравнении результатов наивного агрегирования с показателями, полученными с помощью турнирных решений.

1.3.4.1.5 Рейтинг Center for World University Rankings (CWUR)

Рейтинг CWUR берет свое название от публикующей его организации, штаб-квартира которой находится в Объединенных Арабских Эмиратах. CWUR ежегодно публикуется с 2012 года.

Согласно методологии составления CWUR, альтернативы оцениваются по следующим индикаторам: количество выпускников, получивших престижные академические награды по отношению к размеру вуза, количество выпускников, занимавших руководящие должности в крупных компаниях относительно размера университета, количество научно-педагогических работников, получивших престижные академические награды, общее количество научных статей, количество научных статей, опубликованных в ведущих журналах, количество научных работ, опубликованных в влиятельных журналах, количество высокоцитируемых научных статей. Для линейной свертки показатели берутся с весами $0.25, 0.25, 0.1, 0.1, 0.1, 0.1, 0.1$ соответственно. Вузы расставляются в порядке убывания результата линейной свертки [70].

Из-за высокого веса показателей академических достижений отличительной особенностью этого рейтинга является наличие в нем исследовательских организаций, не являющихся университетами.

1.3.4.1.6 Round University Ranking (RUR)

RUR является первым международным рейтингом вузов, составляемым в России. С 2013 года данный рейтинг ежегодно публикуется одноименной компанией.

Рейтинг строится на основании 20 показателей, разделенных на четыре группы: преподавание, исследования, международное разнообразие и финансовая стабильность. В линейную свертку группы входят с весами $0.4, 0.4, 0.1, 0.1$ соответственно. Вузы ранжируются по убыванию получившейся оценки [71].

Методология составления RUR несколько отличается от других, тем, что учитывает довольно много финансовых показателей, например, доход вуза на одного студента и доля дохода от исследований в общем доходе университета.

1.3.4.2 Описание данных

В качестве начальных данных были взяты шесть ранжирований, основанных на международных рейтингах высших учебных заведений, опубликованных в 2021 году (см. Таблица 1.2), полученных с помощью парсинга данных с официальных сайтов представляющих компаний (рейтинги QS, THE, ARWU и ARTU были собраны с сайта ARTU из соображений удобства). Из собранных таблиц были исключены вузы, которые не встречаются хотя бы в одном рейтинге, а ранги оставшихся смещены с целью исключения пропусков. После чего была собрана итоговая таблица вузов и их мест согласно каждому из шести исходных рейтингов. Данная таблица приведена в [72].

Итоговая таблица содержит в себе 239 университетов. Согласно ей в используемых данных имеются вузов, находящихся в отношении эквивалентности, то есть данная ситуация представляет собой проблему ранжирования в слабом турнире.

1.3.4.3 Корреляционный анализ

Для проведения корреляционного анализа используются коэффициенты ранговой корреляции τ_b-Кендалла и r [40]. На рисунках 1.2 и 1.3 приведены значения коэффициентов τ_b и r, рассчитанные для отобранных вузов.

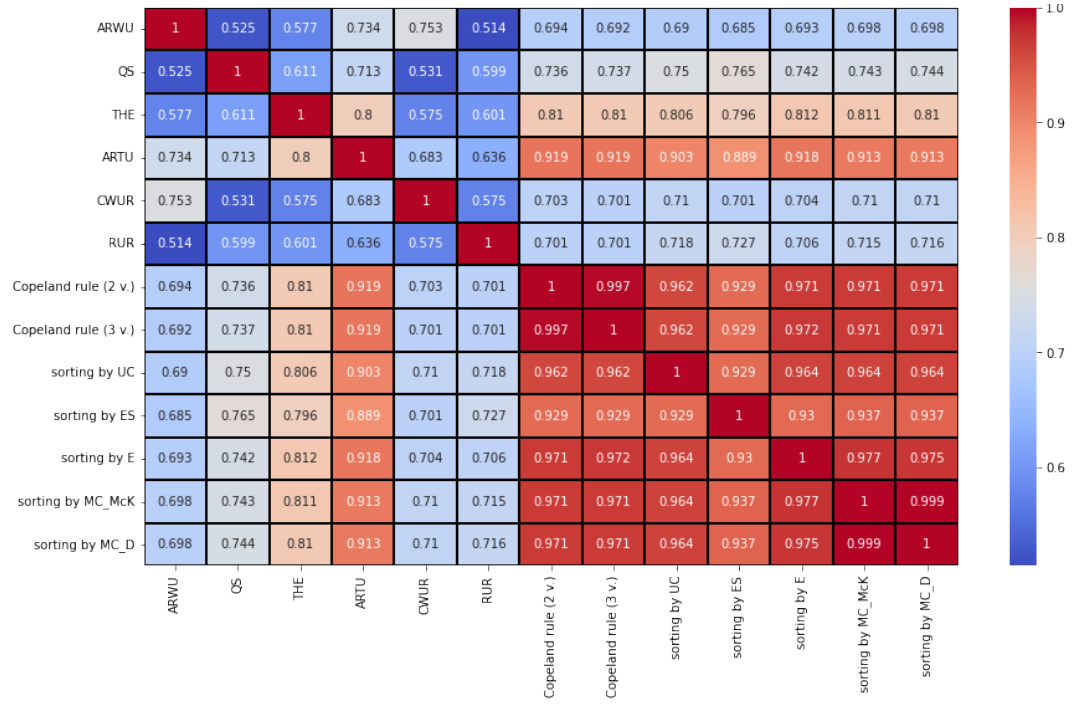


Рисунок 1.2 – Значения коэффициента τ_b

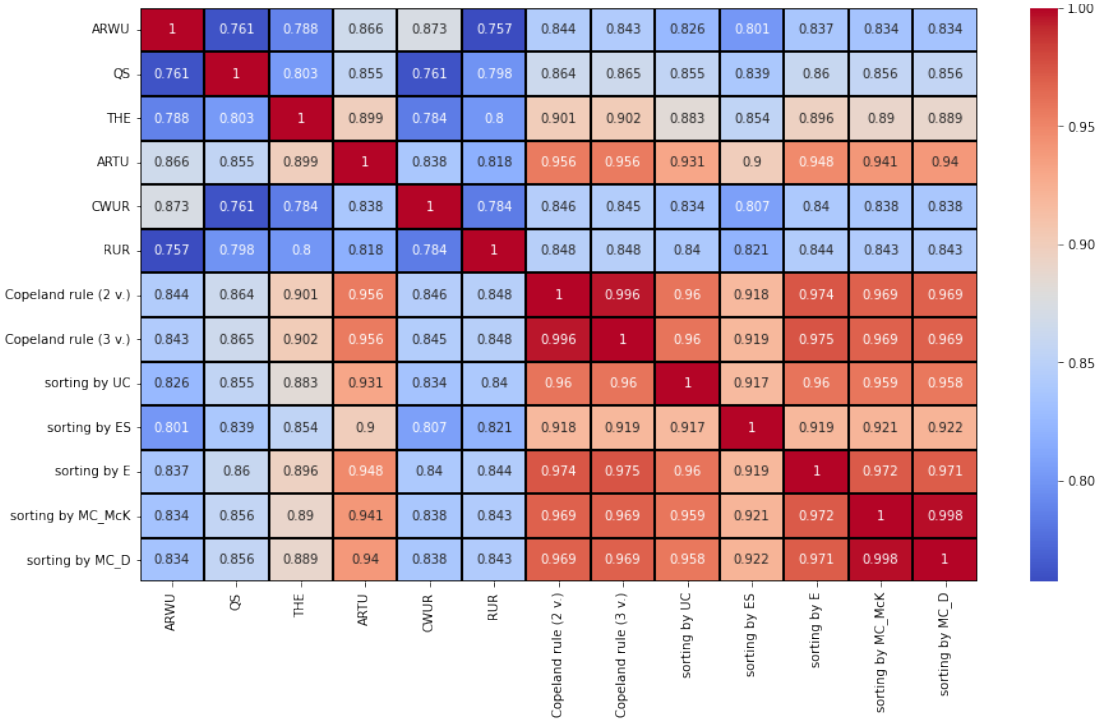


Рисунок 1.3 – Значения коэффициента r

1.3.4.4 Сравнение ранжирований

Ранжирования можно сравнить попарно и упорядочить следующим образом. В данном исследовании предполагается считать, как предлагают авторы статьи [40], что одно ранжирование I репрезентирует множество агрегируемых рейтингов лучше ранжирования II, если ранжирование I коррелирует с большинством из агрегируемых рейтингов сильнее, чем

ранжирование II. Будем считать, что ранжирование I коррелирует с ранжированием III сильнее согласно, например, коэффициенту τ_b, чем ранжирование II, если значение τ_b для пары $\{I, III\}$ больше значения τ_b для пары $\{II, III\}$. Попарное сравнение определяет на множестве ранжирований мажоритарное отношение. Матрицы полученных мажоритарных отношений для коэффициентов τ_b и r приведены в Таблицах 1.3 и 1.4 соответственно.

Таблица 1.3 − Матрица строгого мажоритарного отношения (τ_b)

	ARWU	QS	THE	ARTU	CWUR	RUR	Copeland rule (2	Copeland rule (3	sorting by UC	sorting by ES	sorting by E	sorting by MC_{McK}	sorting by MC_D	**Copeland score**
ARWU	0	0	0	0	0	0	0	0	0	0	0	0	0	0
QS	1	0	0	0	1	1	0	0	0	0	0	0	0	3
THE	1	1	0	0	1	1	0	0	0	0	0	0	0	4
ARTU	1	1	1	0	1	1	0	0	0	0	0	0	0	5
CWUR	1	0	0	0	0	1	0	0	0	0	0	0	0	2
RUR	1	0	0	0	0	0	0	0	0	0	0	0	0	1
Copeland rule (2 v.)	1	1	1	1	1	1	0	1	1	1	0	0	0	9
Copeland rule (3 v.)	1	1	1	1	1	1	0	0	1	1	0	0	0	8
sorting by UC	1	1	1	1	1	1	0	0	0	1	0	0	0	7
sorting by ES	1	1	1	1	1	1	0	0	0	0	0	0	0	6
sorting by E	1	1	1	1	1	1	1	1	1	1	0	0	0	10
sorting by MC_{McK}	1	1	1	1	1	1	1	1	1	1	1	0	0	11
sorting by MC_D	1	1	1	1	1	1	1	1	1	1	1	1	0	12

Таблица 1.4 − Матрица строгого мажоритарного отношения (r)

	ARWU	QS	THE	ARTU	CWUR	RUR	Copeland rule (2	Copeland rule (3	sorting by UC	sorting by ES	sorting by E	sorting by MC_{McK}	sorting by MC_D	**Copeland score s2**
ARWU	0	0	0	0	0	0	0	0	0	0	0	0	0	0
QS	1	0	0	0	1	1	0	0	0	0	0	0	0	3
THE	1	1	0	0	1	1	0	0	0	0	0	0	0	4
ARTU	1	1	1	0	1	1	0	0	1	1	0	0	0	7
CWUR	0	0	0	0	0	0	0	0	0	0	0	0	0	0
RUR	1	0	0	0	1	0	0	0	0	0	0	0	0	2
Copeland rule (2 v.)	1	1	1	1	1	1	0	1	1	1	1	1	1	12
Copeland rule (3 v.)	1	1	1	1	1	1	0	0	1	1	1	1	1	11
sorting by UC	1	1	1	0	1	1	0	0	0	1	0	0	0	6
sorting by ES	1	1	1	0	1	1	0	0	0	0	0	0	0	5
sorting by E	1	1	1	1	1	1	0	0	1	1	0	1	1	10

Продолжение таблицы 1.4

sorting by MC_{McK}	1	1	1	0	1	1	0	0	1	1	0	0	1	8
sorting by MC_D	1	1	1	0	1	1	0	0	1	1	0	0	0	7

Для ранжирования имеющихся ранжирований применим к полученному мажоритарному отношению вторую версию правила Коупланда [40]. Значения оценок Коупланда приведены в Таблицах 2.3 и 2.4, а полученные на их основе ранжирования методов построения ранжирований – в Таблицах 1.5 и 1.6.

Таблица 1.5 − Ранжирование ранжирований (τ_b)

Ранг	Метод
1	Сортировка с помощью MC_D
2	Сортировка с помощью MC_{McK}
3	Сортировка с помощью E
4	Правило Коупланда (2 версия)
5	Правило Коупланда (3 версия)
6	Сортировка с помощью UC
7	Сортировка с помощью ES
8	ARTU
9	THE
10	QS
11	CWUR
12	RUR
13	ARWU

Таблица 1.6 − Ранжирование ранжирований (r)

Ранг	Метод
1	Правило Коупланда (2 версия)
2	Правило Коупланда (3 версия)
3	Сортировка с помощью E
4	Сортировка с помощью MC_{McK}
5	Сортировка с помощью MC_D/
6	ARTU
7	Сортировка с помощью UC
8	Сортировка с помощью ES
9	THE
10	QS
11	RUR
12	ARWU/
13	CWUR

На основании результатов сравнения можно сказать, что минимальные покрывающие по версиям МакКельви и Дуггана множества представляют исходное множество агрегируемых ранжирований лучше, чем любое другое рассматриваемое ранжирование. На данной за-

даче отлично видно причину некоторой критики решений: уже MC_{McK} разбивает вузы на довольно крупные группы, а MC_D и вовсе является наименее дискриминирующим турнирным решением из всех используемых.

Стоит отметить, что несмотря на то, что наивный агрегат трех самых популярных рейтингов ARTU показал себя в задаче довольно неплохо, он все равно хуже любого турнирного решения по τ_b, что говорит о целесообразности использования более сложных методов агрегирования.

1.3.5 Некоторые выводы

В ходе исследования рассмотрены обобщения такого важного турнирного решения, как минимальное покрывающее множество. Было дано достаточно полное описание минимального покрывающего по МакКельви и минимального покрывающего по Дуггану множества, а также консервативного обобщения минимального покрывающего множества. Были описаны свойства рассматриваемых турнирных решений, а именно устойчивость, монотонность и вычислительная простота. Удовлетворение рассматриваемых обобщений данным свойствам делает их разумными инструментами выбора.

Была исследована связь минимального покрывающего по МакКельви и минимального покрывающего по Дуггану множества и консервативного обобщения минимального покрывающего множества с многими другими известными обобщениями турнирных решений на случай слабых турниров, в частности, с обобщениями двухпартийного множества.

В ходе исследования также была рассмотрена практическая задача агрегирования несогласующихся международных рейтингов вузов, основанных на различных индикаторах эффективности университетов в областях науки, образования и финансов. Стоит отметить, что задача решалась на актуальных данных, опубликованных в 2021 году. В результате проведенного корреляционного анализа выяснилось, что обсуждаемые обобщения минимального покрывающего множества, представляют исходную совокупность агрегируемых рейтингов лучше, чем каждый из них, и в случае коэффициента ранговой корреляции τ_b-Кендалла лучше, агрегированные ранжирования, построенные с помощью иных процедур выбора.

Для проведения исследования была создана программа на языке программирования Python. В ней были реализованы полиномиальные алгоритмы для вычисления всех рассматриваемых турнирных решений, вычислительная простота которых доказана. Так полиномиальный алгоритм для вычисления минимального покрывающего по МакКельви и минимального покрывающего по Дуггану множества был взят из [46]. Разработанное программное решение позволило оптимизировать ряд рутинных исследовательских задач, таких как сбор данных с сайтов, расчет турнирных решений, построение с их помощью ранжирований, про-

ведение корреляционного анализа, поиск примеров и контрпримеров, построение и вывод аналитической отчетности.

1.4 Исследование манипулируемости правил коллективного выбора

1.4.1 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. Как было доказано в работах [73] и [74], не существует недиктаторской неманипулируемой процедуры голосования. Позднее такой же вывод был получен в [75] 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 неманипулиемых процедур голосования, можно ли найти наименее манипулируемую процедуру голосования? Для решения этой задачи наиболее часто используется в литературе вероятностный подход, при котором степень манипулиремости рассчитывается как доля профилей предпочтений, в которых возможно манипулирование. В рамках проекта МЦАВР в 2022 году было продолжено изучение и развитие моделей, предложенных в исследованиях предыдущих лет, а именно: индивидуальное манипулирование при полной информации, при неполной информации и одновременном манипулировании других избирателей.

1.4.2 Индивидуальное манипулирование при полной информации. Расчёт и анализ индексов

За последние десятилетия был опубликован целый ряд работ в области исследования процедур голосования на предмет манипулируемости. Такие работы затрагивали исследование как индивидуальной [76], [49], так и коалиционной манипулируемости [77]. 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. Несмотря на то, что такая генерация позволяет добиться определённой

точности, остаются вопросы: если одно правило менее манипулируемо, чем другое, не может ли это являться причиной погрешности вследствие генерации лишь части профилей, а не полного набора?

В рамках исследований мы разработали алгоритмы, позволяющие произвести расчёт точных значений индексов манипулируемости для случая 3 альтернатив для моделей Impartial Culture и Impartial Anonymous Culture.

1.4.2.1 Описание модели

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 $C(P') \succ C(P)$ для манипулирующего участника или группы участников, где знак $\succ$ 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, $a \succ b \succ c$) 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 худшей альтернативами?

Для ответа на этот вопрос в литературе используются расширенные предпочтения [78], 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 $a \succ b \succ c$).

1) Лексимин

Результаты процедур агрегирования сравниваются лексикографически по наихудшей альтернативе. Выбор, у которого худшая альтернатива лучше, является более предпочтительным. Расширенные предпочтения Лексимин:

$$\{a\} \succ \{a,b\} \succ \underline{\{b\} \succ \{a,c\} \succ \{a,b,c\}} \succ \{b,c\} \succ \{c\}. \quad (1.22)$$

2) Лексимакс

Результаты процедур агрегирования сравниваются лексикографически по наилучшей альтернативе. Выбор, у которого лучшая альтернатива лучше, является более предпочтительным. Расширенные предпочтения Лексимакс:

$$\{a\} \succ \{a,b\} \succ \underline{\{a,b,c\} \succ \{a,c\} \succ \{b\}} \succ \{b,c\} \succ \{c\}. \quad (1.23)$$

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:

$$\{a\} \succ \{a,b\} \succ \underline{\{a,c\} \succ \{a,b,c\} \succ \{b\}} \succ \{b,c\} \succ \{c\}. \quad (1.24)$$

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:

$$\{a\} \succ \{a,b\} \succ \underline{\{b\} \succ \{a,b,c\} \succ \{a,c\}} \succ \{b,c\} \succ \{c\}. \quad (1.25)$$

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.

1.4.2.2 Индексы манипулируемости

В литературе известно несколько индексов манипулируемости. Самый популярный — индекс Нитцана − Келли (NK-index). Он был введён в работах [79, 80]. Индекс Нитцана − Келли — это доля манипулируемых профилей в общем количестве профилей.

$$NK = \frac{\text{количество манипулируемых профилей}}{\text{общее количество профилей}}. \quad (1.26)$$

Алгоритм вычисления индекса NK.

1) Для каждого профиля сгенерировать все возможные попытки манипулирования.

2) Если хотя бы одна попытка манипулирования является успешной, то есть $C(P') \succ C(P)$ 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.

Позднее были введены и другие индексы манипулируемости. Например, в работе [49] были предложены индексы свободы манипулирования: индексы I_1^+, I_1^0 и I_1^-.

Для их расчёта классифицируем все возможные попытки манипулирования на три категории. Предположим, что всего существует p возможных попыток манипулирования. Тогда p попыток разделяются на:

− успешные попытки манипулирования (p^+). Таковыми считаются такие изменения предпочтений, при которых результат после манипулирования стал лучше, чем был при записывании в бюллетень искренних предпочтений;

− неуспешные попытки манипулирования, которые не изменяют результат процедуры агрегирования (p^0). Таковыми считаются попытки манипулирования, при которых результат после манипулирования остаётся таким же, каким был и до манипулирования;

− неуспешные попытки манипулирования, которые ухудшают результат (p^-). Таковыми считаются попытки манипулирования, при которых результат после манипулирования стал хуже, чем был до манипулирования.

По определению, $p = p^+ + p^0 + p^-$. Далее мы рассчитываем индексы I_1^+, I_1^0 и I_1^- как доли попыток, которые ведут к улучшению выбора, такому же выбору и худшему выбору, соответственно.

$$I_1^+ = \frac{p^+}{p};\ I_1^0 = \frac{p^0}{p};\ I_1^- = \frac{p^-}{p}. \tag{1.27}$$

В сумме эти индексы дают единицу: $I_1^+ + I_1^0 + I_1^- = 1$.

Разработанные нами алгоритмы позволяют при необходимости рассчитать и другие индексы манипулируемости.

1.4.2.3 Вероятностные модели профилей и методы

Чему равны вероятности различных профилей? В литературе используются две наиболее популярные модели вероятностей профилей: Impartial Culture (IC) и Impartial Anonymous Culture (IAC).

В модели IC предпочтения участников равновероятны. Для случая n участников и m альтернатив у каждого участника возможно одно из $m!$ предпочтений, значит, общее количество профилей равно $(m!)^n$.

Можно заметить, что количество профилей для случая IC растёт очень быстро. Для случая 3 альтернатив и 100 участников всего будет $6.5 \cdot 10^{77}$ различных профилей. Генерация такого количества профилей, очевидно, невозможна даже на современных суперкомпьютерах.

В модели IAC предполагается, что профили равновероятны с учётом анонимности, то есть два профиля считаются одинаковыми, если их можно получить переименованием участников. В модели IAC «единогласные профили», то есть профили, когда все участники имеют одинаковые или почти одинаковые предпочтения, будут более вероятны, чем в модели IC.

Количество профилей в модели IAC для случая n участников и m альтернатив равно $C_{m!+n-1}^n$. Для случая $n = 100$ и $m = 3$ общее количество профилей в IAC составит 96.5 мил-

лионов. Такое количество профилей возможно сгенерировать на современной ЭВМ, равно как и профили для случаев меньшего количества участников.

Мы разработали алгоритмы для расчёта точных значений индексов манипулируемости как для модели IAC, так и для модели IC. Механизм расчётов для модели IC, для которой невозможно сгенерировать все возможные профили, строится на генерации всех профилей для модели IAC с дальнейшим пересчётом количеств таких профилей в модели IC и домножением на соответствующие коэффициенты.

Теорема 1.10. Пусть профиль P состоит из $p_1, \dots, p_6$ участников ($p_1 + \dots + p_6 = n$) с каждым из $3! = 6$ различных предпочтений, где p_1 — количество участников с предпочтениями $a \succ b \succ c$, … p_6 — количество участников с предпочтениями $c \succ b \succ a$. Тогда для каждого такого профиля из модели IAC будет существовать $\frac{n!}{p_1!p_2!p_3!p_4!p_5!p_6!}$ профилей в модели IC.

Тогда при расчётах мы можем сначала для каждого сгенерированного профиля в модели IAC рассчитать индексы манипулируемости, затем посчитать по формуле количество таких профилей в модели IC и получить значение индексов манипулируемости для модели IC.

Таким образом, механизм расчёта для моделей IC и IAC следующий.

1) Генерируем все возможные профили (от 56 для случая 3 участников до 96.5 млн для случая 100 участников) для модели IAC.

2) Для каждого профиля проверяем, является ли он манипулируемым

3) Считаем индексы манипулирования для модели IAC.

4) Умножаем индексы для этого профиля на количество таких профилей в модели IC, чтобы получить индексы манипулирования для модели IC.

Как результат, генерируя все возможные профили для случая IAC мы делаем необходимый пересчёт количества таких профилей в модели IC, чтобы получить точные значения индексов манипулируемости для обеих вероятностных моделей: IAC и IC.

1.4.2.4 Схема расчётов

Стандартный механизм расчёта манипулирумости состоит из генерации профилей и проверки их на манипулируемость. Его можно найти, например, в [81]:

1) Генерация профилей.

2) Для каждого профиля — генерация всех возможных попыток манипулирования, то есть всех возможных изменений предпочтений манипулирующим участником или группой участников.

3) Подсчёт нового результата коллективного выбора для каждой попытки манипулирования.

4) Расчёт индексов I_1 на основе этих результатов (стал ли новый выбор лучше, таким же или хуже для манипулирующего участника или коалиции).

5) Расчёт индекса NK: если хотя бы одна попытка манипулирования привела к более хорошему результату, то профиль считается манипулируемым, и неманипулируемым в противном случае.

При таком подходе возникают две проблемы. Первая – количество профилей. Для случая IC их количество равно $(m!)^n$, поэтому в предыдущих исследованиях в этой области, например, в [78], генерируются не все профили, а 1 миллион случайных профилей. Мы же генерируем все возможные профили для случая IAC, затем домножаем на их количества в модели IC, получая тем самым генерацию всех возможных профилей и точные результаты индексов.

Вторая проблема — высокая вычислительная сложность проверки всех профилей на манипулируемость. На третьем шаге требуется посчитать новый результат коллективного выбора для каждой возможной попытки манипулирования. Как следствие, вычислительная сложность равна количеству всех профилей умноженному на количество всех возможных попыток манипулирования и на алгоритмическую сложность вычисления результата процедуры голосования для профиля. В [81] была показана оценка сложности для метода с генерацией 1 миллиона случайных профилей, и для случая 3 альтернатив и 100 участников время работы алгоритма составляло 63 часа. Если предположить, что расчёты производились для 96.5 млн профилей (общее число профилей в модели IAC для случая 100 участников), то расчёт только одного случая со 100 участниками занял бы более 6000 часов (более 8 месяцев).

Далее используется важное наблюдение, позволяющее снизить сложность расчётов. Идея состоит в том, что при прямолинейном подходе, когда для каждого профиля P генерируются все возможные попытки манипулирования и для всех возможных профилей P' (с неискренними предпочтениями) происходят подсчёты нового результата процедуры $C(P')$, многие расчёты повторяются. Вычисление $C(P')$ может происходить для профилей, для которых результат процедуры голосования на этих профилях уже считался ранее.

Мы достигаем оптимизации за счёт того, что результат процедуры голосования считается на каждом профиле ровно один раз. Для этого мы разбиваем процесс вычислений на два этапа. В рамках первого этапа мы генерируем все возможные профили и считаем выбор для каждого правила на этом профиле. Вычислительная сложность данного этапа равна количе-

ству профилей ($C^n_{m!+n-1}$) умноженному на вычислительную сложность процедуры голосования.

В рамках второго этапа мы проводим перебор всех возможных попыток манипулирования в каждом профиле. Ключевая идея в том, что на втором этапе нам не приходится каждый раз пересчитывать результат правила голосования, мы берём его из посчитанного один раз массива данных.

Если предположить, что в каждом профиле есть p' возможных попыток манипулирования, а вычислительная сложность правила составляет t операций, то прямолинейная модель расчётов предполагает сложность $C^n_{m!+n-1}p't$ операций (соответствует 6000 часов по оценкам выше), а модель с разбиением вычислений на два этапа предполагает сложность первого этапа в $C^n_{m!+n-1}t$ (подсчёт результатов правила голосования для всех профилей) и второго этапа в $C^n_{m!+n-1}p'$ (просчёт всех возможных попыток манипулирования), то есть суммарно получается не $C^n_{m!+n-1}p't$ операций, а $C^n_{m!+n-1}t + C^n_{m!+n-1}p' = C^n_{m!+n-1}(t + p')$. Поскольку и вычислительная сложность правила (t), и количество попыток манипулирования в профиле (p') могут оказаться большими числами, например, для случая многораундовых правил и коалиционного манипулирования, эта оптимизация имеет большое значение, и как минимум на два порядка увеличивает скорость расчётов.

Однако используемый метод является более затратным с точки зрения использования памяти. На первом этапе расчётов генерируются все возможные $C^n_{m!+n-1}$ профилей, для каждого профиля высчитывается результат коллективного выбора по каждой процедуре, и это сохраняется в памяти. Суммарный размер полученных файлов для случаев от 3 до 100 участников и для разных процедур голосования составляет примерно 260 Гигабайт.

На втором этапе этот массив данных используется как входные параметры. Программа считывает все профили для заданного количества участников, после чего производится расчёт индексов манипулируемости. За счет посчитанных заранее результатов коллективного выбора для каждого профиля, на этом этапе после генерации каждой попытки манипулирования можно мгновенно получить новый коллективный выбор для профиля с неискренними предпочтениями, и также мгновенно определять, стал ли выбор при попытке манипулирования лучше или хуже.

Отметим, что максимальная нагрузка на память программы составляет не более 700 Мегабайт: для максимального случая (100 участников) нужно хранить в памяти данные о 96,5 млн профилей, для каждого профиля хранится сам профиль (6 байт) и результат правила коллективного выбора по данному профилю (1 байт), т. е. 96,5 млн * 7 байт = 675.5 Мегабайт. Здесь возможна и дополнительная оптимизация при хранении данных в памяти, напри-

мер, хранить результат коллективного выбора не как 1 байт, а как 3 бита, однако поскольку для современных ЭВМ размер памяти в 700 Мегабайт не является проблемой, мы реализовали хранение в виде более простого подхода, где один параметр хранится в ячейке в один байт.

1.4.2.5 Результаты

Разработанные алгоритмы позволяют получить точные значения индексов манипулируемости для случая 3 альтернатив. Мы провели расчеты 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 мы приводим пример результатов для самого известного правила голосования — Правила относительного большинства — для случая индивидуального манипулирования для иллюстрации результатов (рисунок 1.4).



Рисунок 1.4 − Индекс NK для расширенных предпочтений Лексимин

Начиная с определенного количества участников, степень манипулируемости начинает снижаться. Это соответствует известным в литературе результатам.

Для проверки результатов мы использовали работы [49, 78], в которых расчет проходил с помощью генерации набора случайных профилей. Сравнение результатов показало, что расхождение по индексу NK между точным расчётом и методом генерации случайных профилей составляло в пределах 0,001 (одной тысячной).

Теперь приведем график для индексов свободы манипулирования (I_1). Самым тёмным цветом, внизу графика, обозначена доля попыток, ведущих к улучшению результата (I_1^+), немного светлее, вверху графика, обозначена доля попыток, ведущая к ухудшению результата коллективного выбора для манипулирующего участника (I_1^-), посередине, самым светлым

серым — доля попыток манипулирования, ведущих к неизменности результата (I_1^0) (рисунок 1.5):



Рисунок 1.5 − Индексы свободы манипулирования

Чаще всего запись в бюллетень неискренних предпочтений приводит к тому, что коллективный выбор не меняется. Это интуитивно понятный вывод, так как особенно на большом количестве участников голос одного человека имеет всё меньшее значение. Количество случаев, когда изменение предпочтений ведет к ухудшению результата превышает количество случаев, когда изменение предпочтений приводит к улучшению результата. Это означает, что манипулирующему участнику нужно быть осторожным, чтобы случайно не получить более плохой для себя результат.

Главный результат данной работы — алгоритмы, позволяющие получить точные оценки индексов манипулируемости для случая трёх альтернатив. Мы рассмотрели, как возможно посчитать точные значения как для IAC, так и для IC, сделав это вплоть до случая 100 участников за разумное время. Были приведены примеры результатов для Правила относительного большинства, была произведена сверка с известными в литературе результатами, полученными на основе подхода случайной генерации профилей. При этом расчеты были сделаны и для ряда других правил.

Дальнейшее направление исследования — анализ индексов манипулируемости, полученных данных для различных правил. В частности, исследование вопроса, можно ли выделить правила голосования, которые Парето-доминируют другие правила в терминах манипулируемости, то есть являются всегда менее манипулируемыми, чем другие.

1.4.3 Неполнота информации и одновременное манипулирование

В рамках проекта МЦАВР в 2022 году было также продолжено изучение модели манипулирования при неполной информации, в которой манипулирующие избиратели учиты-

вают действия других участников голосования. Так, в 2021 году был произведен сравнительный анализ так называемого наивного манипулирования (Модель 1), при котором стимул к манипулированию у данного избирателя не зависит от наличия стимула к манипулированию у других, и модели одновременного манипулирования (Модель 2). Последняя модель предполагает, 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. Вычислены были точные значения индекса манипулируемости для шести правил коллективного выбора при 4 типах публичной информации. В 2022 году исследование было дополнено Моделью 3, в которой избиратель учитывает, что манипулировать могут не все, кто имеет к этому стимул, и потому неопределенность для принимающего решение избирателя становится еще выше, чем в Модели 2. Представленные в данном отчете результаты касаются теоретического исследования всех трех моделей и вычислений для новой Модели 3.

1.4.3.1 Определения и обозначения

Пусть $N = \{1, 2, ..., n\}$ — множество избирателей, имеющих предпочтения на множестве альтернатив X, $|X| = m$. Бинарное отношение $P_i \subset A \times A$ — предпочтения избирателя i, являющиеся линейным порядком, $L(X)$ — множество всех линейных порядков. Профиль предпочтений всех избирателей обозначается за $\vec{P} = (P_1, ..., P_n)$, а профиль предпочтений всех избирателей, кроме i — за $\vec{P}_{-i}$. Множество всех профилей предпочтений есть $L(X)^N$, включающее $(m!)^n$ элементов. Отображение $F : L(X)^N \to 2^X \setminus \varnothing$ называется правилом коллективного выбора. Если результат правила включает в себя более одной альтернативы, то применяется алфавитное правило устранения множественности выбора, $T : 2^X \setminus \varnothing \to X$. Согласно ему, на множестве X предполагается заданным некоторый линейный порядок P_T, например, $aP_TbP_Tc...$, 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. $T(F(\vec{P}))$ обозначается как $V(\vec{P})$. Пусть $v(a, \vec{P})$ — вектор распределения рангов альтернативы a, j-й элемент которого, $v_j(a, \vec{P})$, равен количеству избирателей, в предпочтениях которых альтернатива a стоит на j-ом месте.

Обозначим мажоритарное отношение: $a_k \mu a_l$, если $|\{i \in N : a_k P_i a_l\}| > |\{i \in N : a_l P_i a_k\}|$, а матрицу мажоритарного отношения — за $MG(\vec{P})$, где

$$MG(\vec{P})_{kl} = \begin{cases} 1, \text{ если } a_k P_M a_l, \\ -1, \text{ если } a_l P_M a_k, \\ 0, \text{ иначе.} \end{cases} \quad (1.28)$$

Пусть $r_i(a, P_i) = |\{b \in A \mid aP_i b\}|$ — ранг альтернативы a в предпочтениях P_i.

1.4.3.2 Моделирование неполной информации

Предполагается, что перед голосованием избиратели располагают некоторой информацией о профиле предпочтений $\vec{P} = (P_1, ..., P_n)$. 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) π. Таким образом, у каждого из них имеется множество возможных профилей предпочтений, согласующихся с полученной избирателем информацией – информационное множество.

В исследовании рассматривали следующие виды ФПИ:

ФПИ-Profile, $\pi(\vec{P}) = \vec{P}$.

− ФПИ-Rank, упорядочение альтернатив, построенное согласно правилу.

− ФПИ-Winner, $\pi(\vec{P}) = F(\vec{P})$.

− ФПИ-1Winner, $\pi(\vec{P}) = V(\vec{P})$.

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

$$W_i^{\pi(\vec{P})} = \{\vec{P}'_{-i} \in L(X)^{N \setminus \{i\}} : \pi(P_i, \vec{P}'_{-i}) = \pi(\vec{P})\}. \quad (1.29)$$

1.4.3.3 Модели манипулирования

Модель 1

Определение 1.7. Избиратель i имеет стимул манипулировать в Модели 1 при ФПИ π в профиле $\vec{P}$ при правиле коллективного выбора F, если существует $\tilde{P}_i$ такое, что 1) либо $V(\tilde{P}_i, \vec{P}'_{-i}) = V(\vec{P})$, либо $V(\tilde{P}_i, \vec{P}'_{-i}) P_i V(\vec{P})$ для всех $\vec{P}'_{-i} \in W_i^{\pi(\vec{P})}$; 2) существует $\vec{P}''_{-i} \in W_i^{\pi(\vec{P})}$ такой, что $V(\tilde{P}_i, \vec{P}''_{-i}) P_i V(\vec{P})$.

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

Эту модель манипулирования, Модель 1, мы назовем «наивной моделью». В противоположность ей рассмотрим Модель 2, в которой избиратель проверяет, будет ли стратегия, с которой он имеет стимул манипулировать в Модели 1, более предпочтительна, чем истинные предпочтения, если все остальные избиратели из $\Pi(\vec{P})\setminus\{i\}$ решают манипулировать.

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

Модель 2

Определение 1.8. Избиратель i имеет стимул манипулировать в Модели 2 при ФПИ π в профиле $\vec{P}$ при правиле коллективного выбора F, если существует $\tilde{P}_i$ такое, что 1) избиратель i имеет стимул манипулировать в Модели 1 при π в $\vec{P}$ и при правиле F; 2) либо $V(\tilde{P}_i,\tilde{\vec{P}}'_{-i})=V(P_i,\tilde{\vec{P}}'_{-i})$, либо $V(\tilde{P}_i,\tilde{\vec{P}}'_{-i})P_iV(P_i,\tilde{\vec{P}}'_{-i})$ для всех $\vec{P}'_{-i}\in W_i^{\pi(\vec{P})}$.

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

Модель 3

Определение 1.9. Избиратель i имеет стимул манипулировать в Модели 3 при ФПИ π в профиле $\vec{P}$ при правиле коллективного выбора F, если существует $\tilde{P}_i$ такое, что 1) избиратель i имеет стимул манипулировать в Модели 1 при π в $\vec{P}$ и при правиле F; 2) либо $V(\tilde{P}_i,(\tilde{\vec{P}}'_K,\vec{P}'_{-K-i}))=V(P_i,(\tilde{\vec{P}}'_K,\vec{P}'_{-K-i}))$, либо $V(\tilde{P}_i,(\tilde{\vec{P}}'_K,\vec{P}'_{-K-i}))P_iV(P_i,(\tilde{\vec{P}}'_K,\vec{P}'_{-K-i}))$ для всех $\vec{P}'_{-i}\in W_i^{\pi(\vec{P})}$ и для всех $K\subseteq\Pi(\vec{P})\setminus\{i\}$.

Определение 1.9 означает, что, если применение стратегии манипулирования избирателем i приводит к худшему результату, чем неманипулирование при условии, что манипу-

лирует кто-то из $\Pi(\vec{P})\setminus\{i\}$, стимул к манипулированию у избирателя i в Модели 3 отсутствует.

1.4.3.4 Правила коллективного выбора

1) Правило относительного большинства. Выбор по данному правилу определяется как множество альтернатив, являющихся наилучшими для наибольшего числа избирателей, т. е.

$$c \in F(\vec{P}) \Leftrightarrow c \in \arg\max_{a\in X}\left(\left\langle s_{Pl}, v(a,\vec{P})\right\rangle\right), \tag{1.30}$$

где $s_{Pl} = (1,0,...,0)$, а $\left\langle s_{Pl}, v(a,\vec{P})\right\rangle$ — скалярное произведение.

2) Правило вето (обратное правило простого большинства). Выбирается альтернатива, занявшая наименьшее число последних мест в предпочтениях избирателей.

3) Правило Борда. По данному правилу для каждой альтернативы a вычисляется ранг Борда $r(a,\vec{P})$ как сумма рангов $r_i(a,P_i)$ по всем избирателям. Или, с помощью вектора очков $s_B = (m-1, m-2, ..., 1, 0)$,

$$c \in F(\vec{P}) \Leftrightarrow c \in \arg\max_{a\in X}\left(\left\langle s_{B}, v(a,\vec{P})\right\rangle\right). \tag{1.31}$$

4) Двухступенчатая мажоритарная система. Выбирается альтернатива, являющаяся наилучшей больше, чем для половины избирателей. Если такой альтернативы нет, то проводится второй тур голосования, в котором участвуют две альтернативы, набравшие наибольшее количество голосов в первом туре.

5) Правило передачи голосов (Процедура Хара). Выбирается альтернатива, имеющая более 50% первых мест в предпочтениях избирателей. Если таковой нет, то из профиля исключается альтернатива, имеющая наименьшее число первых мест. Процедура повторяется до тех пор, пока не будет выбрана какая-либо альтернатива.

6) Правило Коупленда. Выбирается альтернатива, для которой количество побед минус количество поражений по мажоритарному отношению наибольшее, т. е.

$$CS(a,\vec{P}) = |\{b \in X \mid a\mu b\}| - |\{b \in X \mid b\mu a\}|,$$
$$c \in F(\vec{P}) \Leftrightarrow c \in \arg\max_{a\in X} CS(a,\vec{P}). \tag{1.32}$$

7) Максиминная процедура. Для каждой альтернативы a вычисляется минимальное количество избирателей, которые предпочитают эту альтернативу некоторой другой, а затем выбирается альтернатива, для которой это значение максимально

$$c \in F(\vec{P}) \Leftrightarrow c \in \arg\max_{a \in X} \min_{b \in X} |\{i \in N : aP_i b\}|. \tag{1.33}$$

8) Процедура Болдуина (обратная процедура Борда). Для каждой альтернативы подсчитывается ранг Борда $r(a, \vec{P})$, затем из профиля удаляется альтернатива, имеющая наименьший ранг Борда. Процедура исключения альтернатив продолжается до тех пор, пока не останутся неисключаемые альтернативы, которые и являются множеством выбора в процедуре Болдуина.

9) Правило Нансона. Рассчитываются значения ранга Борда для каждой альтернативы, затем из профиля удаляются все альтернативы, имеющие ранг Борда ниже среднего ранга,

$$\overline{r} = \frac{1}{m} \sum_{a \in X} r(a, \vec{P}). \tag{1.34}$$

10) Процедура Блэка. Выбирается победитель Кондорсе — альтернатива, недоминируемая по мажоритарному отношению, т. е. $a \in F(\vec{P}) \Leftrightarrow \neg \exists b \in A:\ b \mu a$, если он существует, в противном случае победитель определяется по правилу Борда.

11) Правило Кемени. Введем сначала функцию расстояния по Кемени между линейными порядками P_1 и P_2. Пусть $\{d_{ij}^{(k)}\}$ — матрица бинарного отношения P_k: если $aP_k b$, то элемент этой матрицы $d_{ab}^{(k)} = 1,\ d_{ba}^{(k)} = -1$. Расстояние между двумя предпочтениями вычисляется как сумма модулей разности элементов матриц

$$d(P_1, P_2) = \sum_{a \in X} \sum_{b \in X} | d_{ab}^{(1)} - d_{ab}^{(2)} |. \tag{1.35}$$

Результат правила Кемени — линейный порядок, для которого сумма расстояний до предпочтений всех избирателей минимальна, т. е.

$$P_K = \arg\min_{P \in L(X)} \sum_{i \in N} d(P, P_i). \tag{1.36}$$

Победитель — кандидат, занимающий первое место в найденном упорядочении.

12) Правило порогового агрегирования. Альтернативы упорядочиваются на основании сравнения векторов распределения рангов. Строится бинарное отношение R: aRb, если $v_m(a,\vec{P}) < v_m(b,\vec{P})$ или если существует такое $k \le m$, что $v_i(a,\vec{P}) = v_i(b,\vec{P})$, $i = k-1,...,m$ и $v_k(a,\vec{P}) < v_k(b,\vec{P})$. Альтернативы, недоминируемые в R, являются победителями: $a \in F(\vec{P}) \Leftrightarrow \neg\exists b \in A:\ bRa$.

1.4.3.5 Результаты

Пусть $I^M(m,n,\pi,F)$ — доля профилей предпочтений, в которых хотя бы один избиратель имеет стимул манипулировать в модели M при ФПИ π и правиле F. Сформулируем и докажем несколько простых утверждений об индексах манипулируемости в различных моделях.

Утверждение 1.1. Для любой ФПИ π и для любого правила коллективного выбора F, если $I^{M1}(m,n,\pi,F) = 0$, то и $I^{M2}(m,n,\pi,F) = 0$.

Утверждение 1.2. Для любой ФПИ π и для любого правила коллективного выбора F, если $I^{M1}(m,n,\pi,F) = 0$, то и $I^{M3}(m,n,\pi,F) = 0$.

Следующее утверждение — наследование неманипулируемости при переходе от Модели 2 к Модели 3.

Утверждение 1.3. Для любой ФПИ π и для любого правила коллективного выбора F, если $I^{M}(m,n,\pi,F) = 0$, то и $I^{M3}(m,n,\pi,F) = 0$.

Утверждение 1.4. Пусть π' не менее информативна, чем π''. Тогда для любого $M \in \{M1, M2, M3\}$, для любого правила F, числа избирателей n и числа альтернатив m, если $I^{M}(m,n,\pi',F) = 0$, то $I^{M}(m,n,\pi'',F) = 0$.

1.4.3.6 Вычисление индекса манипулируемости в Модели 3

В результате вычислительных экспериментов в среде MATLAB были получены значения $I^{M3}(m,n,\pi,F)$ для $m = 3$ и для всех четырех рассматриваемых ФПИ и 12 правил коллективного выбора. Значения приведены в Таблицах 1.7–1.10.

В Таблице 1.7 приведены значения разности индекса $I^{M2}(m,n,Profile,F)$ и $I^{M3}(m,n,Profile,F)$, позволяющие оценить то, насколько влияет неопределенность относительно действий других избирателей в Модели 3 при полной информации. Как следует из определений 2 и 3, значения этой разности неотрицательны, т. е. индекс манипулируемости

в Модели 3 не может увеличиться по сравнению с Моделью 2. Наибольшая разность индексов наблюдается для правила относительного большинства.

Таблица 1.7 − Значения $I^{M2}(3,n,Profile,F)-I^{M3}(3,n,Profile,F)$

Profile	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
Plurality	0,11	0,15	0,19	0,17	0,20	0,22	0,21	0,23	0,24	0,23	0,24	0,24	0,24	0,24	0,24	0,24	0,24
Veto	0,20	0,20	0,18	0,18	0,17	0,15	0,14	0,13	0,12	0,11	0,10	0,09	0,09	0,08	0,07	0,07	0,06
Borda	0,24	0,27	0,25	0,22	0,21	0,20	0,18	0,17	0,16	0,15	0,14	0,13	0,13	0,12	0,11	0,11	0,10
Run-off	0,11	0,09	0,09	0,10	0,11	0,09	0,10	0,09	0,09	0,08	0,09	0,08	0,08	0,08	0,08	0,07	0,08
STV	0,11	0,15	0,09	0,12	0,11	0,11	0,10	0,08	0,10	0,10	0,08	0,09	0,08	0,07	0,08	0,08	0,07
Copeland	0,11	0,26	0,14	0,19	0,14	0,16	0,13	0,14	0,12	0,13	0,11	0,12	0,11	0,11	0,10	0,11	0,09
Maximin	0,11	0,14	0,14	0,15	0,14	0,14	0,13	0,13	0,12	0,12	0,11	0,11	0,11	0,11	0,10	0,10	0,09
Baldwin	0,11	0,14	0,14	0,15	0,13	0,14	0,12	0,13	0,10	0,12	0,09	0,11	0,09	0,10	0,08	0,09	0,07
Nanson	0,24	0,14	0,18	0,15	0,11	0,13	0,07	0,10	0,05	0,08	0,04	0,07	0,03	0,06	0,02	0,05	0,02
Black	0,11	0,14	0,14	0,15	0,13	0,14	0,12	0,13	0,10	0,12	0,09	0,10	0,08	0,09	0,07	0,09	0,07
Kemeny	0,11	0,14	0,14	0,15	0,14	0,14	0,13	0,13	0,12	0,12	0,11	0,11	0,11	0,11	0,10	0,10	0,09
Threshold	0,19	0,28	0,23	0,16	0,23	0,19	0,16	0,18	0,14	0,14	0,14	0,10	0,11	0,12	0,08	0,09	0,10

Рассмотрим теперь различные типы неполной публичной информации. При ФПИ-Rank избирателям известно упорядочение, построенное на основе правила коллективного выбора. В правилах подсчета очков это будет ранжирование, слабый порядок, в котором более высокое место будет занимать альтернатива с большим количеством очков.

По сравнению случаем полной информации, значения индекса манипулируемости при ФПИ-Rank меняются наиболее заметно для правила вето, правила Нансона (обращаются в нуль в большинстве случаев) и для двухступенчатой мажоритарной системы (возрастают до 0,3–0,5). Эффект возрастания значений индекса манипулируемости в случае неполной информации объясняется возрастанием мощности информационного множества избирателя. В Таблице 1.8 представлены результаты вычисления индекса манипулируемости для ФПИ-Winner. По сравнению с ФПИ-Rank, случаев неманипулируемости стало гораздо больше, однако большинство из них наследовано из Модели 2. Ненулевой индекс при правиле относительного большиснтва, правиле Борда и правиле порогового агрегирования означает, что некоторые избиратели все же имеют стимул к искажению предпочтений, несмотря на неполноту информации и неопределенность относительно действий других избирателей. Это наблюдение противопоставляет данные правила остальным, где неманипулируемость возникает начиная с определенного числа избирателей или с чередованием.

Таблица 1.8 − Значения $I^{M3}(m,n,Rank,F)$

Rank	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
Plurality	0,00	0,07	0,12	0,12	0,18	0,23	0,25	0,36	0,31	0,33	0,43	0,37	0,39	0,47	0,40	0,42	0,50
Veto	0,03	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
Borda	0,28	0,19	0,08	0,08	0,07	0,06	0,07	0,08	0,09	0,10	0,10	0,11	0,11	0,12	0,12	0,12	0,13
Run-off	0,11	0,35	0,28	0,52	0,29	0,42	0,36	0,54	0,27	0,42	0,24	0,26	0,10	0,21	0,00	0,09	0,00
STV	0,00	0,07	0,52	0,24	0,00	0,72	0,23	0,22	0,74	0,37	0,00	0,58	0,27	0,00	0,00	0,21	0,00
Copeland	0,19	0,25	0,21	0,21	0,25	0,22	0,27	0,22	0,29	0,23	0,29	0,23	0,30	0,23	0,30	0,23	0,30
Maximin	0,06	0,21	0,02	0,16	0,17	0,11	0,07	0,00	0,07	0,00	0,07	0,00	0,07	0,00	0,07	0,00	0,06
Baldwin	0,19	0,19	0,30	0,32	0,19	0,28	0,31	0,18	0,15	0,23	0,13	0,22	0,14	0,23	0,14	0,24	0,15
Nanson	0,28	0,17	0,19	0,30	0,00	0,18	0,25	0,05	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
Black	0,06	0,22	0,02	0,13	0,02	0,09	0,03	0,09	0,03	0,09	0,03	0,01	0,04	0,01	0,04	0,01	0,04
Kemeny	0,19	0,16	0,15	0,21	0,36	0,31	0,42	0,28	0,45	0,31	0,48	0,33	0,49	0,34	0,50	0,35	0,51
Threshold	0,06	0,17	0,11	0,19	0,24	0,22	0,30	0,34	0,32	0,37	0,40	0,38	0,42	0,44	0,42	0,45	0,46

Таблица 1.9 − Значения $I^{M3}(m,n,Winner,F)$

Winner	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
Plurality	0,00	0,10	0,12	0,14	0,26	0,28	0,25	0,33	0,33	0,30	0,37	0,36	0,34	0,39	0,37	0,36	0,40
Veto	0,03	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
Borda	0,22	0,31	0,15	0,15	0,16	0,17	0,19	0,20	0,22	0,23	0,24	0,25	0,26	0,27	0,27	0,28	0,28
Run-off	0,22	0,38	0,71	0,70	0,37	0,59	0,00	0,37	0,00	0,08	0,00	0,00	0,00	0,00	0,00	0,00	0,00
STV	0,00	0,10	0,71	0,29	0,00	0,60	0,00	0,00	0,00	0,15	0,00	0,00	0,00	0,00	0,00	0,00	0,00
Copeland	0,06	0,37	0,00	0,29	0,00	0,32	0,00	0,35	0,00	0,36	0,00	0,38	0,00	0,38	0,00	0,39	0,00
Maximin	0,06	0,22	0,02	0,13	0,43	0,09	0,20	0,00	0,23	0,00	0,25	0,00	0,27	0,00	0,28	0,00	0,29
Baldwin	0,06	0,22	0,64	0,13	0,00	0,30	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
Nanson	0,22	0,22	0,21	0,25	0,00	0,18	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
Black	0,06	0,22	0,64	0,13	0,00	0,09	0,00	0,09	0,00	0,08	0,00	0,00	0,00	0,00	0,00	0,00	0,00
Kemeny	0,06	0,22	0,02	0,13	0,43	0,09	0,20	0,00	0,23	0,00	0,25	0,00	0,27	0,00	0,28	0,00	0,29
Threshold	0,06	0,22	0,32	0,14	0,19	0,18	0,22	0,25	0,23	0,27	0,29	0,27	0,30	0,31	0,29	0,32	0,32

Таблица 1.10 − Значения $I^{M3}(m,n,1Winner,F)$

1Winner	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
Plurality	0,00	0,15	0,00	0,17	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
Veto	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
Borda	0,31	0,27	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
Run-off	0,22	0,52	0,71	0,57	0,37	0,32	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
STV	0,00	0,15	0,71	0,29	0,00	0,22	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
Copeland	0,17	0,21	0,00	0,15	0,00	0,20	0,00	0,23	0,00	0,26	0,00	0,28	0,00	0,29	0,00	0,30	0,00
Maximin	0,17	0,18	0,21	0,57	0,26	0,20	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
Baldwin	0,17	0,18	0,43	0,57	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
Nanson	0,31	0,18	0,00	0,57	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
Black	0,17	0,18	0,43	0,57	0,00	0,20	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
Kemeny	0,17	0,18	0,21	0,57	0,26	0,20	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
Threshold	0,11	0,44	0,62	0,13	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00

Небольшое различие в качестве публичной информации между ФПИ-Winner и ФПИ-1Winner, а именно то, что победитель голосования сообщается уже после устранения множе-

ственности выбора, приводит к отсутствию стимула к манипулированию начиная с некоторого n для всех правил, кроме Коупленда. Для последнего сохраняется чередование ненулевой манипулируемости и нуля, причем амплитуда колебаний возрастает на рассматриваемой области значений n. Несмотря на это, по результатам вычислений можно сделать вывод, что манипулирование в условиях неполной информации представляет собой довольно трудоемкую задачу с неизбежным риском оказаться в проигрыше, если учитывать то, что и другие участники могут искажать свои предпочтения. Следовательно, манипулирование уже не является однозначно выигрышной стратегией, так что в некоторых случаях можно говорить об отсутствии стимулов к манипулированию у всех избирателей, т. е. о неманипулируемости.

1.4.4 Некоторые выводы

В рамках проекта было проведено масштабное исследование манипулируемости правил коллективного выбора. Разработаны алгоритмы, позволяющие получить точные значения индексов манипулируемости для случая 3 альтернатив. Вычислены значения индексов индивидуального манипулирования для числа участников от 3 до 100 участников для различных процедур голосования для 4 типов расширенных предпочтений (РП): Лексимин, Лексимакс, Рискофил, Рискофоб для случая IAC. Возможное дальнейшее направление исследования — анализ индексов манипулируемости, полученных данных для различных правил. В частности, исследование вопроса, можно ли выделить правила голосования, которые Парето-доминируют другие правила в терминах манипулируемости, то есть являются всегда менее манипулируемыми, чем другие. Кроме того, предложена и исследована модель, сочетающая в себе предпосылку о неполной информации и неопределенность относительно действий других избирателей, развивающая и дополняющая предложенную ранее модель.

Результаты исследований нашли отражение в работах [82, 83].

1.5 Исследования по теории паросочетаний

В этом подразделе отчета приведены результаты исследований по теории паросочетаний. Исследованы две задачи:

– проблема паросочетаний адвокатов и судебных дел;

– алгоритмы паросочетаний, максимизирующие доход.

1.5.1 Проблема паросочетаний адвокатов и судебных дел

1.5.1.1 Введение

Распределение государственных адвокатов неимущим клиентам является стандартной практикой во многих странах по всему миру, включая США и страны Европы. Судебное дело включает в себя двух экономических агентов: истца и ответчика. В большинстве случаев только неимущие ответчики в уголовном деле имеют право на государственного адвоката, в

то время как истцом является государство, представляемое прокурором. В странах Европы судебные системы различаются, однако во многих странах как истец, так и ответчик имеют права на государственного адвоката даже в гражданских делах. Споры о “стабильности” подобных распределений начались в США в 1970-х. В знаменитом деле Флита Драмго подал в суд Верховный Суд Марин Каунти за отказ назначить ему Ричарда Ходжа в качестве государственного адвоката, в то время как Ричард Ходж был готов вести его дело.

Хотя Драгмо и проиграл дело, но судебное решение не было анонимным и один из судей не соглашался с решением суда, аргументируя это тем, что система уголовного правосудия может лишиться легитимности, если неимущим ответчикам не будет разрешено выбирать своих представителей из группы адвокатов, желающих их представлять.

В Румынии, если ответчик или истец хочет, чтобы его представлял не назначенный ему, а другой государственный адвокат, и этот адвокат готов его представлять, то суд обязан удовлетворить этот запрос.

Общий признак всех этих примеров заключается в том, что при рассмотрении их как паросочетаний в них имеются блокирующие пары и устранение этих блокирующих пар либо обязательно, либо остаётся на усмотрение суда. Более того, касательно тех систем, в которых это не обязательно, ведутся споры о том, нет ли необходимости эти системы изменить. В данной работе мы представляем новую модель паросочетаний адвокатов для обеих сторон судебных дел с неимущими агентами с каждой из сторон и анализируем стабильность данных паросочетаний. Оказывается, что чтобы данная модель была в состоянии описать рассматриваемые условия недостаточно стандартного анализа паросочетаний для двух сторон или любого стандартного расширения этого анализа на случаи большего количество сторон. Причина этого заключается в том, что элементами паросочетаний являются тройки вида (адвокат истца, дело, адвокат ответчика). Однако, в каждом деле участвуют четыре активных агента: два адвоката, истец и ответчик. Более того, потенциальные блокирующие пары могут возникать на любой из сторон паросочетаний – истец может образовать блокирующую пару с другим адвокатом истца с одной стороны, ответчик может образовать блокирующую пару с другим адвокатом ответчика с другой стороны. Таким образом, никакие результаты для паросочетаний с двумя сторонами, представленные в литературе, не могут быть расширены на наш случай.

Очевидно, что ответчик (как и истец) хочет быть представленным сильным адвокатом, в то же время он хочет, чтобы истец (ответчик) был представлен слабым адвокатом. Также логично предположить, что адвокаты хотят выигрывать дела, либо иметь дело с делами, требующими меньше усилий при их ведении. Например, если есть явное доказательство

правоты их клиента, это будет означать, что адвокату не будет необходимости прилагать слишком много усилий для ведения данного дела. Более того, молодой адвокат, в большей степени заботящийся о карьерном росте, или более старый адвокат, заботящийся о репутации, могут предпочесть более крупные дела с большей публичностью, чем менее значимые дела. Даже если паросочетание стабильно с одной из сторон (например, на стороне ответчика), на другой стороне могут существовать потенциальные блокирующие адвокат и истец. В силу того, что предпочтения адвокатов зависят от силы адвоката на противоположной стороне, то если мы удовлетворяем запросу блокирующей пары, это может привести к нестабильности и на стороне ответчика. Это явно указывает на то, что мы находимся в условиях, в которых имеется два паросочетания двух сторон, однако паросочетание на одной из сторон имеет влияяние на паросочетание на другой стороне.

Данная работа вносит вклад в литературу по теме паросочетаний в тех условиях, в которых данный конфликт релевантен. Примерами подобных работ являются [84, 85, 86]. Однако, в данных работах проблема рассматривается как проблема двух сторон – ответчиков и государственных адвокатов, и игнорируется возможный эффект от полученных паросочетаний адвокатов истцов и адвокатов ответчиков. Паросочетания с подобными внешними эффектами являются в текущий момент развивающейся темой в литературе, по которой уже имеется достаточное количество интересных и значимых публикаций. В [87, 88] предлагают концепцию стабильности, в которой агенты формируют гипотезы об эффекте их пар на общий рынок до принятия того или иного решения о формировании блокирующей пары. Наша концепция стабильности накладывает большие ограничения, так как разрешает агентам и клиентам игнорировать реакцию противоположной стороны, формируя блокирующую пару. В [89] была предложена крайне абстрактная модель для работы над паросочетаниями с внешними эффектами. Однако, их результаты ограничены в силу слишком абстрактной модели, не подходящей под особенности прикладных сценариев, анализируемых в данной работе.

1.5.1.2 Модель и результаты

Имеется две непересекающиеся группы адвокатов размера n; адвокатов истцов назовём левыми адвокатами (L), адвокатов ответчиков назовём правыми адвокатами (R). Имеется n судебных дел, состоящих из истца (левый клиент) и ответчика (правый клиент). Имеется объективное ранжирование адвокатов с каждой стороны. Каждый клиент хочет быть представленным сильным адвокатом, и каждый агент на одной из сторон хочет, чтобы с противоположной стороны был слабый оппонент. Имеется объективный порядок дел для L-адвокатов, характеризующийся преимуществами, определяемыми имеющимися у истцов до-

казательствами. Чем больше преимущество истца, тем соответственно меньше преимущество ответчика. Более того, имеется разбиение, распределяющее наборы дел в различные кластеры с учетом какого-либо другого критерия. Стандартным примером подобного разбиения может служить разбиение, которое относит судебные дела к одной из двух категорий: мелкие/незначимые (low-profile) и крупные/резонансные (high-profile) дела. Адвокаты имеют субъективные предпочтения и единственным ограничением на эти предпочтения является то, что имея определенного адвоката на противоположной стороне, в каждом из кластером разбиения, адвокат предпочитает дело клиента с большим преимуществом, чем дело клиента с меньшим преимуществом. Паросочетанием для данной задачи называем тройку следующего вида: (L-адвокат, дело, R-адвокат), где каждый адвокат и дело присутствуют ровно в одной тройке. Паросочетание стабильно, если не имеется блокирующей пары ни на одной из сторон.

В описанных выше условиях мы доказываем, что если разбиение состоит из одного элемента (дела различаются только преимуществом истца) или двух элементов (присутствует разбиение на мелкие и крупные дела), то существует стабильное паросочетания для любого профиля предпочтений. Если разбиение состоит из по меньшей мере трех элементов (имеются более сложные критерии отличия судебных дел), то стабильного паросочетания может не существовать для некоторого профиля предпочтений. Для случаев, когда разбиение состоит из одного элемента (то есть, по сути, нет никакого разбиения) или из двух элементов, мы представляем быстрый алгоритм, который всегда находит стабильное паросочетание. Мы также доказываем, что для любого разбиения и любого профиля предпочтений все стабильные паросочетания имеют определенное свойство: в каждом кластере разбиения паросочетание между адвокатами и делами, ограниченное на данном кластере, позитивно ассортативно (более сильные адвокаты получают более легкие дела), при этом паросочетание между адвокатами, ограниченное на данном кластере, негативно ассортативно (более сильные адвокаты получают более слабых адвокатов в качестве оппонентов). Более того, для любого разбиения и для любого паросочетания, удовлетворяющего данному свойству, существует профиль предпочтений, для которого данное паросочетание стабильно. Таким образом, данные паросочетания – единственные рационализируемые. Это важный критерий, который необходимо проверять, если рассматриваемое паросочетание может быть потенциально нестабильно или когда предпочтения неизвестны. С точки зрения анализа общего благосостояния мы показываем, что если стабильное паросочетание применяется, то эффективность так же будет гарантироваться. Хотя в данном случае релевантной является наша концепция стабильности, для удовлетворения теоретического любопытства мы проверили, есть

ли какая-либо логическая связь между этой концепцией и стабильностью с точки зрения ядра. Выяснилось, что данные концепции независимы.

1.5.1.3 Некоторые выводы

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. Описаны стандартные условия, в которых стабильное паросочетание гарантированно существует. Разработан быстрый алгоритм, который всегда находит стабильное паросочетание в данных условиях. В других условиях стабильного паросочетания может не существовать. Описаны общие характеристики всех стабильных паросочетаний независимо от того, находимся ли мы в рамках стандартных условий. Показано, что данное определение стабильности и стандартное определение ядра логически независимы при нашей постановке задачи. Однако, все стабильные паросочетания эффективны.

1.5.2 Алгоритмы паросочетаний, максимизирующие доход

1.5.2.1 Введение

Стандартное предсказание в литературе двусторонних рынков паросочетаний говорит о том, что такие рынки должны продуцировать стабильные паросочетания без или практически без каких-либо поисковых помех. Онлайн платформы могут предоставить среду, значительно уменьшающую поисковые помехи. Основываясь на этих предсказаниях, в литературе рассматривают формирование цен платформ-монополистов, максимизирующих этот доход, а также о результирующей членской базе, формирующейся исходя из предположения, что платформы используют стабильные паросочетания. Примерами таких исследований могут служить [90, 91] и другие. Алгоритм паросочетаний, используемый на онлайн платформах, зачастую использует “черный ящик”. Например, онлайн платформы для свиданий имеют стимулы сохранять своих членов достаточно долго на платформе, чтобы получать больший доход. В связи с этим моментальное нахождение “идеальной пары” для участников такой платформы не обязательно будет оптимальным для самой платформы. В данной работе мы исследуем алгоритм, максимизирующий доход, для платформ паросочетаний. Мы пытаемся выяснить, что происходит, если мы добавляем дополнительный инструмент, позволяющий платформам получать доход, помимо выбора цен. Более того, наш анализ не ограничен только алгоритмом монополиста, но также включает в себя и рассмотрение алгоритма и ценообразования на рынке дуополии.

В нашей модели имеется n мужчин и k женщин ($k \leq n$). Агенты на одной из сторон имеют одинаковые порядковые предпочтения, включая объективное ранжирование мужчин

или женщин. Однако, мы допускаем в модели и гетерогенность в кардинальных (численных) предпочтениях относительно противоположной стороны. Мы работаем в домене монотонных предпочтений, мужчины (женщины), имеющие более высокий ранг, получают (строго) большую полезность, чем те, кто имеет более низкий ранг, при паросочетании с любой фиксированной женщиной (любым фиксированным мужчиной) на другой стороне. Мы называем такие предпочтения (сильно) монотонными, и эти предпочтения включают в себя предпочтения, задаваемые “pizzazz” значениями [90, 92], и пересекаются с супермодулярными функциями прибыли паросочетаний [93, 94]. Здесь имеется единственное стабильное паросочетание, являющее позитивно ассортативным, то есть наилучшему мужчине дают в пару наилучшую женщину, второму – вторую и так далее.

Сначала мы анализируем случай монополистической платформы, которая анонсирует цены, зависящие от гендера, а затем следует этом анонсу, а также алгоритм, который указывает паросочетание для каждой потенциальной группы подписчиков (членской базы). Мы работаем в предположении, что предпочтения являются общим знанием. Решение агента о подписке основывается на ex-ante сравнении его/ее ожидаемого дохода от подписки и цены на нее. Ожидаемый доход зависит от алгоритма, который был анонсирован, и от решений о подписках остальных агентов. Платформа выбирает алгоритм и цены оптимально с точки зрения 1) стимулирования нужной группы в решении о подписке, 2) создания наибольшей для данной группы прибыли от паросочетания и 3) извлечения наибольшей части данной прибыли. Предполагая такие же предпочтения и схожие стратегические взаимодействия, мы анализируем случай с двумя конкурирующими платформами. И последнее, мы ослабляем предположение о полной информации, разрешая платформам предлагать случайные паросочетания в случаях, когда мы работаем с “pizzazz” значениями в качестве предпочтений. В монополистическом случае, в отличие от того, что предполагается в литературе, мы показываем, что стабильное паросочетание не максимизирует доход платформы, а значит платформа не обязательно считает выгодным аннулировать все поисковые помехи. Если же платформы конкурируют, то они будут предлагать стабильные паросочетания, несмотря на то что это может не принести им положительного дохода.

1.5.2.2 Модель и результаты

Сначала мы работаем над доменом монотонных предпочтений и монополистической платформой. Платформа одновременно анонсирует алгоритм паросочетания (паросочетания для всех потенциальных членских баз) и цены для мужчин и женщин. После этого, все мужчины и женщины одновременно решают подписываться ли им на платформу или нет. Мы показываем, что негативно ассортативное паросочетание для членской базы, состоящей из

одинакового количества мужчин и женщин, имеющих наиболее высокие ранги, является “pairwise coalition-proof subgame perfect equilibrium” результатом этой последовательной игры, в которой мы разрешаем отклонения пар вида мужчина-женщина. Данный результат единственный, если предпочтения строго монотонны, а значит стабильное паросочетание в данном случае никогда не является оптимальным. Однако, при “pizzazz” значениях все паросочетания дают одинаковую прибыль, а значит, платформа не имеет сильного стимула предлагать стабильное паросочетание даже в этом случае. Мы расширяем наш анализ на случай конкуренции двух платформ паросочетаний. Это расширение позволяет нам рассматривать различные внешние опции участников каждой из платформ. Мы показываем, что “pairwise coalition-proof subgame perfect equilibrium” результат типа Бертрана с нулевой прибылью всегда существует. В таком равновесии результат всегда является стабильным паросочетанием, в котором участники максимальной симметричной членской базы на одной из платформ распределены позитивно ассортативно. Однако, может быть такое, что конкурирующие платформы могут косвенно координироваться через предложения вне равновесий. Над этим вопросом мы все еще ведем работу.

И последнее, мы ослабляем предположение о том, что предпочтения являются полной общеизвестной информацией. Предпочтения задаются “pizzazz” значениями. Платформе известно лишь распределение этих значений для каждого из типов - мужчин и женщин. Более того, платформам разрешается анонсировать случайные паросочетания. Мы показываем, что доход уникально максимизируется на членской базе, и на платформу подписываются только лучшие k мужчин и лучшие k женщин, а также что равномерное случайное паросочетание, где все возможные пары участников реализуются с одинаковой вероятностью, является оптимальным. Более того, у платформы есть возможность генерировать максимальную общую прибыль и полностью ее извлекать. Крайне интересно, что этот результат идет в контрасте с механизмами, рассматриваемыми в литературе. Так в [93, 94, 95] в рамках строгих предпосылок и/или в более узком сетапе показано, что платформа реализует стабильное паросочетание для максимизации дохода, используя ценовую дискриминацию. В нашем случае платформа может извлекать всю общую прибыль, не используя сильный инструмент ценовой дискриминации, а лишь равномерно случайно распределяя людей.

1.5.2.3 Некоторые выводы

Рассмотрены алгоритмы паросочетаний, максимизирующие доход на рынке бракосочетаний. Показано, что если платформа является монополией и то, как она распределяет подписчиков является переменной выбора, то распределение не обязательно будет аннулировать все поисковые помехи на рынке. Однако, если имеется по крайней мере две конкури-

рующие платформы, то существуют равновесия с нулевым доходом. Для всех таких равновесий все члены общества входят в паросочетание, и общее паросочетание в обществе стабильно. Могут ли конкурирующие платформы косвенно координировать свои действия друг с другом через предложения вне равновесий, является открытым вопросом. Все упомянутые результаты получены в рамках модели полной информации и при допустимости только детерминированных паросочетаний. Также проанализирован случай, при котором платформы знают только распределения предпочтений людей в обществе и могут предлагать случайные паросочетания. Если предпочтения описываются “pizzazz” значениями, то платформа может извлекать наибольшую общую прибыль при помощи генерации равномерно распределенных случайных паросочетаний.

1.6 Исследование аксиоматики линейных функций интенсивности коалиций

1.6.1 Введение

Индексы влияния используются для изучения избирательных органов и баланса власти [96, 97, 98]. В качестве примера, рассмотрим индекс влияния Банцафа (Bz), который был введен в [99].

Пусть N — конечное множество агентов. Всякое подмножество множества N называется коалицией. Коалиция является выигрывающей, если количество голосов данной коалиции не меньше квоты q. Агент i называется ключевым в выигрывающей коалиции A, если коалиция $A\backslash\{i\}$ не выигрывающая. Множество коалиций, в которых агент i является ключевым, обозначается W_i.

Индекс Банцафа для агента i определяется следующей формулой

$$Bz_i = \frac{|W_i|}{\sum_{j\in N}|W_j|}. \tag{1.37}$$

Например, рассмотрим законодательный орган, в котором присутствуют три партии — A, B, C. Предположим, что партия A имеет 30 голосов, партия B 45 голосов, партия C 25 голосов, квота q = 51 (правило простого большинства). Тогда, индекс Банцафа для партии A равен $Bz_A = \frac{|\{\{A,B\},\{A,C\}|}{|\{\{B,A\},\{B,C\}|+|\{\{A,B\},\{A,C\}|+|\{\{C,B\},\{C,A\}|} = \frac{1}{3}$. Для партий B, C, $Bz_B = Bz_B = \frac{1}{3}$.

При вычислении индекса Банцафа и других часто используемых индексов влияния предполагается, что все коалиции могут быть образованы. Предположим, что в рассмотренном раннее примере партия B предпочтет не создавать коалицию с партиями A, C. Тогда, коалиции $\{B,A\},\{B,C\}$ и $\{A,B,C\}$ не могут быть образованы. Если вычислить индекс Банцафа для данного случая, мы получим для партий A, C $Bz_A = Bz_C = \frac{1}{2}$, для B $Bz_B = 0$. Партия B

имеет больше голосов чем партии A, C, однако влияние B в рассматриваемом законодательном органе равно 0. Отметим, что данные значения не удовлетворяют аксиоматике индекса Банцафа [100, 101]. Подобные ситуации имеют место в реальных политических системах [102].

Индексы влияния, учитывающие предпочтения участников по созданию коалиций были введены в [96]. Данные индексы вычисляются с помощью функций интенсивности связи агентов. В [96] введены 16 функций интенсивности связи, построена аксиоматика для одной функции интенсивности и поставлена проблема аксиоматизации остальных.

1.6.2 Основная часть

1.6.2.1 Постановка задачи

Индекс влияния агента i, учитывающий предпочтения по созданию коалиций, вычисляется с помощью функции интенсивности связи агента i с остальными агентами коалиции w $f(i,w)$. Для каждого агента $i \in w$ значение $f(i,w)$ является действительным числом, отражающим предпочтения агентов в коалиции w. Другими словами, $f: N \times (2^N \backslash \emptyset) \to \mathbb{R}$.

В [96] для каждого агента $i \in N$ вводится значение

$$\chi_i = \sum_{w \in W_i} f(i,w). \tag{1.38}$$

Таким образом, значение χ_i является агрегированной интенсивностью связи агента i с остальными агентами в выигрывающих коалициях, в которых i — ключевой.

Индекс влияния, учитывающий предпочтения по созданию коалиций, вводится в [96] по аналогии индекса Банцафа (модификации для остальных индексов влияния были рассмотрены в [103]) в соответствии со следующей формулой

$$\alpha_i = \frac{\chi_i}{\sum_{j \in N} \chi_j}. \tag{1.39}$$

Таким образом, индекс влияния агента i α_i является нормированным значением χ_i.

В [96] предполагается, что предпочтение агента i образовать коалицию с агентом j является действительным значением $p_{ij} \in [0,1]$, $\forall\, i \in N$ $\sum_{j \in N} p_{ij} = 1$. Значение p_{ij} называется интенсивностью связи агента i с агентом j. В общем случае, $p_{ij} \neq p_{ji}$. Индексы влияния, в которых предпочтения определяются значениями p_{ij} называются кардинальными [96].

Рассмотрим следующие линейные функции интенсивности связи для кардинальных индексов [96]:

– средняя интенсивность связи агента i с остальными агентами коалиции w

$$f^+(i,w) = \frac{\sum_{j\in w} p_{ij}}{|w|-1}; \tag{1.40}$$

– средняя интенсивность связи агентов коалиции w с агентом i

$$f^-(i,w) = \frac{\sum_{j\in w} p_{ji}}{|w|-1}; \tag{1.41}$$

– средняя интенсивность для агента i

$$f_{avg}(i,w) = \frac{1}{2}(f^+(i,w) + f^-(i,w)); \tag{1.42}$$

– средняя положительная интенсивность в w

$$f^+(w) = \frac{\sum_{i\in w} f^+(i,w)}{|w|}; \tag{1.43}$$

– средняя отрицательная интенсивность в w

$$f^-(w) = \frac{\sum_{i\in w} f^-(i,w)}{|w|}; \tag{1.44}$$

– средняя интенсивность в w

$$f_{avg}(w) = \frac{\sum_{i\in w} f_{avg}(i,w)}{|w|}. \tag{1.45}$$

Отметим, что три последние функции не зависят от агента i. Таким образом, для данных функций все агенты имеют одинаковые значения интенсивности связи.

Заметим, что

$$\frac{\sum_{i\in w} f^+(i,w)}{|w|} \equiv \frac{\sum_{i\in w} f^-(i,w)}{|w|} \equiv \frac{\sum_{i\in w} f_{avg}(i,w)}{|w|} \Rightarrow f^+(w) = f^-(w) = f_{avg}(w). \tag{1.46}$$

Таким образом, вместо всех трех функций $f^+(w), f^-(w), f_{avg}(w)$ будем рассматривать $f_{avg}(w)$.

В [96] введена аксиоматика для функции интенсивности $f^+(i,w)$.

Аксиома 1. Для любого набора, состоящего из m значений $(p_{i1}, \dots, p_{im})$ существует функция $f(i,w)$, $0 \le f(i,w) \le 1$, где $f(i,w)$ – непрерывно дифференцируемая функция для каждого своего аргумента.

Аксиома 2. Если $p_{ij} = 0$ для любого j, тогда $f(i,w) = 0$.

Аксиома 3 (Монотонность). Значение функции $f(i,w)$ возрастает (убывает) тогда и только тогда, когда любое значение p_{ij} возрастает (убывает). Более того,

$$\frac{\partial f(i,w)}{\partial p_{ij}} = \mu_i \text{ для любого } j \tag{1.47}$$

и

$$\frac{\partial f(i,w)}{\partial p_{lj}} = 0 \text{ для любого } l \neq j. \tag{1.48}$$

Теорема 1.11 [96]. Функция $f(i,w)$ удовлетворяет аксиомам 1–3 тогда и только тогда, когда $f(i,w) = f^+(i,w)$.

В данном исследовании мы вводим аксиоматику для функций $f^-(i,w)$, $f_{avg}(i,w)$, $f_{avg}(w)$.

1.6.2.2 Аксиоматика линейных функций интенсивности

Рассмотрим следующий набор аксиом.

Аксиома 1. Для любого набора, состоящего из m значений $(p_{i1}, \dots, p_{im})$ существует функция $f(i,w)$, $0 \le f(i,w) \le 1$, где $f(i,w)$ – непрерывно дифференцируемая функция для каждого своего аргумента (Функция интенсивности связи существует для любого агента и любого набора предпочтений).

Аксиома 1a. Для любого набора, состоящего из k значений $(p_{1i}, \dots, p_{ki})$ существует функция $f(i,w)$, $0 \le f(i,w) \le 1$, где $f(i,w)$ – непрерывно дифференцируемая функция для каждого своего аргумента.

Аксиома 1b. Для любого набора, состоящего из $2s$ значений $(p_{1i}, p_{i1}, \dots p_{si}, p_{is})$ существует функция $f(i,w)$, $0 \le f(i,w) \le 1$, где $f(i,w)$ – непрерывно дифференцируемая функция для каждого своего аргумента.

Аксиома 1c. Для любого набора, состоящего из $r(r-1)$ значений $(p_{12},\dots,p_{1r},p_{21},\dots,p_{r\,1},\dots,p_{r\,r-1})$ существует функция $f(w)$, $0 \le f(w) \le 1$, где $f(w)$ – непрерывно дифференцируемая функция для каждого своего аргумента.

Аксиома 2. Если $p_{ij}=0$ для всех p_{ij} из $(p_{i1},\dots,p_{im})$, тогда $f(i,w)=0$ (Если агент i состоит в плохих отношениях с остальными агентами коалиции, то интенсивность связи агента i с ними равна 0).

Аксиома 2a. Если $p_{ji}=0$ для всех p_{ji} из $(p_{1i},\dots,p_{ki})$, тогда $f(i,w)=0$ (Если агенты коалиции w плохо относятся к i, то интенсивность связи агента i с остальными агентами коалиции равна 0).

Аксиома 2b. Если $p_{ij}=0$, $p_{ji}=0$ для всех p_{ji}, p_{ji} из $(p_{1i},p_{i1},\dots p_{si},p_{is})$, тогда $f(i,w)=0$.

Аксиома 2c. Если $p_{lm}=0$ для всех p_{lm} из $(p_{12},\dots,p_{1r},p_{21},\dots,p_{r\,1},\dots,p_{r\,r-1})$, тогда $f(w)=0$.

Аксиома 3 (Монотонность 1) Значение $f(i,w)$ возрастает (убывает) тогда и только тогда, когда любое значение p_{ij} возрастает (убывает). Кроме того,

$$\frac{\partial f(i,w)}{\partial p_{ij}} = \mu_i \text{ для любого } p_{ij} \text{ из } (p_{i1},\dots,p_{im}) \tag{1.49}$$

и

$$\frac{\partial f(i,w)}{\partial p_{lj}} = 0 \text{ для любого другого случая.} \tag{1.50}$$

Аксиома 3a (Монотонность 2) Значение $f(i,w)$ возрастает (убывает) тогда и только тогда, когда любое значение p_{ji} возрастает (убывает). Кроме того,

$$\frac{\partial f(i,w)}{\partial p_{ji}} = \mu_i \text{ для любого } p_{ji} \text{ из } (p_{1i},\dots,p_{mi}) \tag{1.51}$$

и

$$\frac{\partial f(i,w)}{\partial p_{jl}} = 0 \text{ для любого другого случая.} \tag{1.52}$$

Аксиома 3b (Монотонность 3) Значение $f(i,w)$ возрастает (убывает) тогда и только тогда, когда любые значения p_{ij}, p_{ji} возрастают (убывают). Кроме того,

$$\frac{\partial f(i,w)}{\partial p_{ij}} = \frac{\partial f(i,w)}{\partial p_{ji}} = \mu_i \text{ для любых } p_{ij}, p_{ij} \text{ из } (p_{1i}, p_{i1}, \dots p_{si}, p_{is}) \quad (1.53)$$

и

$$\frac{\partial f(i,w)}{\partial p_{il}} = \frac{\partial f(i,w)}{\partial p_{jl}} = 0 \text{ для любого другого случая.} \quad (1.54)$$

Аксиома 3c (Монотонность 4) Значение $f(w)$ возрастает (убывает) тогда и только тогда, когда любое значение p_{lm} возрастает (убывает). Кроме того,

$$\frac{\partial f(i,w)}{\partial p_{lm}} = \mu_w \text{ для } p_{lm} \text{ из } (p_{12}, \dots, p_{1r}, p_{21}, \dots, p_{r\,1}, \dots, p_{r\,r-1}) \quad (1.55)$$

и

$$\frac{\partial f(i,w)}{\partial p_{lm}} = 0 \text{ для любого другого случая.} \quad (1.56)$$

Теорема 1.12.

1) $f(i,w)$ удовлетворяет Аксиомам 1, 2, 3 тогда и только тогда, когда $f(i,w) = f^{+}(i,w)$;

2) $f(i,w)$ удовлетворяет Аксиомам 1a, 2a, 3a тогда и только тогда, когда $f(i,w) = f^{-}(i,w)$;

3) $f(i,w)$ удовлетворяет Аксиомам 1b, 2b, 3b тогда и только тогда, когда $f(i,w) = f_{avg}(i,w)$;

4) $f(i,w)$ удовлетворяет Аксиомам 1c, 2c, 3c тогда и только тогда, когда $f(w) = f_{avg}(w)$.

Отметим, что рассмотренные линейные функции интенсивности связи могут быть представлены тремя аксиомами с незначительными изменениями в условиях. Список условий аксиом представлен в Таблице 1.11, где «+» означает, что функция удовлетворяет соответствующей аксиоме.

Таблица 1.11 − Список аксиом

функция/аксиома	**1**	**1*a***	**1*b***	**1*c***	2	**2*a***	**2*b***	**2*c***	3	**3*a***	**3*b***	**3*c***
$f^{+}(i,w)$	+				+				+			
$f^{-}(i,w)$		+				+				+		
$f_{avg}(i,w)$			+				+				+	
$f_{avg}(w)$				+				+				+

1.6.3 Некоторые выводы

В данной работе рассматриваются линейные функции интенсивности связи для кардинальных индексов влияния. Построена аксиоматика для каждой линейной функции интенсивности связи и доказана теорема. Показано, что каждая линейная функция определяется одним набором из трех аксиом с незначительными изменениями в условиях. Результаты проведенного исследования нашли отражение в [104].

2 Исследование оптимизационных моделей

В этом разделе отчета приведены результаты разработки и исследований оптимизационных моделей. Работа в рамках этой области знаний осуществлялась в следующих научных направлениях:

− нахождение структуры многовалютных депозитов, оптимальных по гарантированному риску;

− формулирование гибких бизнес-правил для улучшения процесса закупки компонентов для систем управления;

− математическое моделирование формирования оптимального набора учебных курсов по выбору;

− исследования по теории расписаний.

2.1 Нахождение структуры многовалютных депозитов, оптимальных по гарантированному риску

2.1.1 Введение

При определении оптимальной структуры многовалютного вклада приходится сталкиваться с неопределенностью будущих значений экономических параметров, в частности, с неизвестными величинами обменных курсов используемых валют. Типичной является ситуация так называемой существенной неопределенности, при которой нет не только точных значений неопределенных параметров, но и каких-либо статистических характеристик. Известны лишь границы их возможных значений. Это приводит к возникновению риска, понимаемого как отклонение получаемого результата от желательного или ожидаемого результата.

Одним из подходов к измерению и минимизации риска является концепция минимаксного сожаления Сэвиджа [105]. В соответствии с ней риск интерпретируется как потеря от незнания неопределенных факторов. К функции риска применяется минимаксный подход, позволяющий вычислить наименьший гарантированный риск.

Концепция оптимальности по Сэвиджу и другие подходы к задачам с неопределенными параметрами рассматривались в публикациях [106, 107, 108]. В работе [109] исследовалось решение минимаксного сожаления Сэвиджа в задаче диверсификации депозита по двум видам валют. Статья [110] 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.

Дальнейшие исследования заключались в разработке вычислительных средств для решения задачи нахождения оптимальных по гарантированному риску структур многовалютных депозитов и проведении вычислительных экспериментов на различных данных.

2.1.2 Постановка задачи диверсификации вклада

В начале временного периода вклада Лицо, Принимающее Решение (ЛПР – депозитор) 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, оставляя $1 - x_d - x_e$ рублей 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 соответствии с этими правилами итоговый доход в рублевом выражении можно представить в следующем виде:

$$f(x, y) = (1 + r)(1 - x_d - x_e) + x_d \frac{1+d_d}{K_d} y_d + x_e \frac{1+d_e}{K_e} y_e. \quad (2.1)$$

Данный показатель желательно максимизировать путем выбора стратегии $x = (x_d, x_e) \in X = \{x_d + x_e \leq 1,\ x_i \geq 0\ (i = d, e)\}$, причем ЛПР должен учитывать возможность реализации любых значений неопределенности $y = (y_d, y_e) \in Y = [a_d, b_d] \times [a_e, b_e]$.

В рамках максиминного подхода Вальда [111] эта задача может рассматриваться как антагонистическая игра ЛПР с «Природой». Решением ее является полученная по принципу наилучшего гарантированного результата «осторожная» максиминная стратегия. Отметим, что в рассматриваемой задаче она имеет довольно тривиальный вид: следует вычислить доходность каждого вида валют при самом неблагоприятном возможном сочетании финальных курсов $y = (a_d, a_e)$ и все средства вложить в ту валюту, для которой этот показатель максимален. Если максимум достигается на нескольких видах валют, весь рубль распределяется между таковыми произвольным образом. Этот результат верен для любого числа валют [109].

Как известно, критерий Вальда пессимистичен и ориентирован на самый худший случай, на «катастрофу». Он не допускает риска и зачастую дает крайне заниженные показатели. С другой стороны, многие ЛПР принимают риск и нуждаются в инструментарии для его минимизации. Одним из них является принцип минимаксного сожаления (критерий Сэвиджа). В нем также используется понятие наилучшего гарантированного результата, но применяемое не к исходному показателю качества $f(x, y)$, а к так называемой функции риска (сожаления)

$$\Phi(x,y) = \max_{z \in X} f\,(z,y) - f(x,y)\,. \qquad (2.2)$$

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 применить к функции риска концепцию наилучшего гарантированного результата (принцип Вальда) [111].

Эти соображения приводят к следующему определению (принцип минимаксного сожаления Сэвиджа [105].

Определение 2.1. Назовем решение x^* гарантированным по риску решением (ГРР) если

$$\Phi^* = \max_{y \in Y} \Phi(x^*, y) = \min_{x \in X} \max_{y \in Y} \Phi(x, y), \qquad (2.3)$$

где функция риска (потерь) $\Phi(x, y)$ определена в (2.2).

Будем называть также оптимальными решение x^* и соответствующее ему минимальное значение гарантированного риска Φ^*.

2.1.3 Анализ функции риска и аналитическое решение в частном случае

Явный вид функции риска получен в работе [110]:

$$\Phi[x] = \max\{\alpha_d x_d + \alpha_e x_e\,, \beta_d(1 - x_d) + \alpha_e x_e,\ \alpha_d x_d + \beta_e(1 - x_e)\}\,, \qquad (2.4)$$

где производные параметры $\propto_d, \propto_e, \beta_d, \beta_e$ определяются следующим образом:

$$\alpha_d = \left[1 + r - \frac{1+d_d}{K_d} a_d\right], \qquad \alpha_e = \left[1 + r - \frac{1+d_e}{K_e} a_e\right], \qquad (2.5)$$

$$\beta_d = \left[\frac{1+d_d}{K_d} b_d - (1 + r)\right], \quad \beta_e = \left[\frac{1+d_e}{K_e} b_e - (1 + r)\right]. \qquad (2.6)$$

Коэффициенты $\propto_d, \propto_e, \beta_d, \beta_e$ зависят от 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 β_d (β_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).

Рассмотрим частный случай, когда $\propto_d = \propto_e = 0$. Это означает, что при минимальных финальных значениях курсов доллара и евро все три валюты равнодоходны. Тогда, с точки зрения принципа Вальда, любой допустимый план диверсификации дает один и тот же (и, стало быть, оптимальный) результат [109]. Это объясняется тем, что данный принцип учитывает только наихудшее значение $y = (y_d, y_e) \in Y$ для каждой стратегии $x = (x_d, x_e) \in X$. Таким в нашем случае, очевидно, для всех x будет $y = (a_d, a_e)$. В отличие от принципа Вальда, оптимальное по Сэвиджу решение находится с учетом как нижних, так и верхних границ неопределенности b_d, b_e и зависящих от них коэффициентов β_d, β_e.

Функция гарантированного риска (2.4) принимает вид

$$\Phi[x] = \max\{0,\ \beta_d(1-x_d), \beta_e(1-x_e)\}, x \in X. \tag{2.7}$$

Из условия $\propto_d = \propto_e = 0$ и определения коэффициентов $\propto_d, \propto_e, \beta_d, \beta_e$ следует, что $\beta_d > 0, \beta_e > 0$. Кроме того, $(1-x_d) \geq 0,\ (1-x_e) \geq 0$ для $x \in X$. Поэтому

$$\Phi[x] = \max\{\beta_d(1-x_d), \beta_e(1-x_e)\}, x \in X. \tag{2.8}$$

Напомним, что множество допустимых решений является двумерным симплексом $X = \{x = (x_d, x_e) | x_d + x_e \leq 1, x_i \geq 0\ (i = d, e)\}$.

Очевидно, что решение $x = (x_d, x_e)$, такое, что $x_d + x_e < 1$, не может быть точкой минимума функции (2.8), так как «слегка» увеличив обе переменные, получим уменьшение функции. Поэтому точка минимума на множестве X лежит на отрезке $\{x_d + x_e = 1, x_d \geq 0, x_e \geq 0\}$, где $x = (1-x_e, x_e)$. Соответствующая задача рассматривалась в [110] без дополнительного предположения о равной доходности валют. В частном случае $\propto_d = \propto_e = 0$ с учетом условия $\beta_d > 0, \beta_e > 0$ из Утверждения 3 в [110] получаем следствие:

Следствие 2.1. Если $\propto_d = \propto_e = 0$, то функция гарантированного риска $\Phi[x]$ достигает минимального на множестве X значения в некоторой точке из множества $\{(1,0), (0,1), x_e^{23} = (\beta_d/(\beta_e + \beta_d)\,, \beta_e/(\beta_e + \beta_d)\}$.

При выводе следствия учтено, что некоторые точки из списка в Утверждении 3 в [110] здесь дублируются и что условие $x_e^{23} \in X$ выполнено в данном случае автоматически.

Для окончательного выбора решения вычислим значение функции гарантированного риска (2.8) в найденных точках-кандидатах:

$$\Phi[(1,0)] = \max\{\beta_d(1-1), \beta_e(1-0)\} = \beta_e\,,\ \Phi[(0,1)] = \beta_d\,,\ \Phi[(x_d^{23}, x_e^{23})] = \max\{\beta_d\beta_e/(\beta_e + \beta_d), \beta_e\beta_d/(\beta_e + \beta_d)\} = \beta_d\beta_e/(\beta_e + \beta_d). \quad (2.9)$$

Так как последнее вычисленное значение меньше значений в двух других точках, гарантированное по риску решение в случае равной доходности валют единственно и имеет вид:

$$x^* = (\beta_d/(\beta_e + \beta_d)\,, \beta_e/(\beta_e + \beta_d)). \quad (2.10)$$

Оно обеспечивает минимальный гарантированный риск

$$\Phi^* = \Phi[x^*] = \beta_d\beta_e/(\beta_e + \beta_d). \quad (2.11)$$

Таким образом, в случае рублевой равной доходности всех 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 $x^* = (0.5, 0{,}5)$, оптимальный риск $\Phi^* = \beta/2$. Безрисковая валюта (рубль) в распределении не участвует.

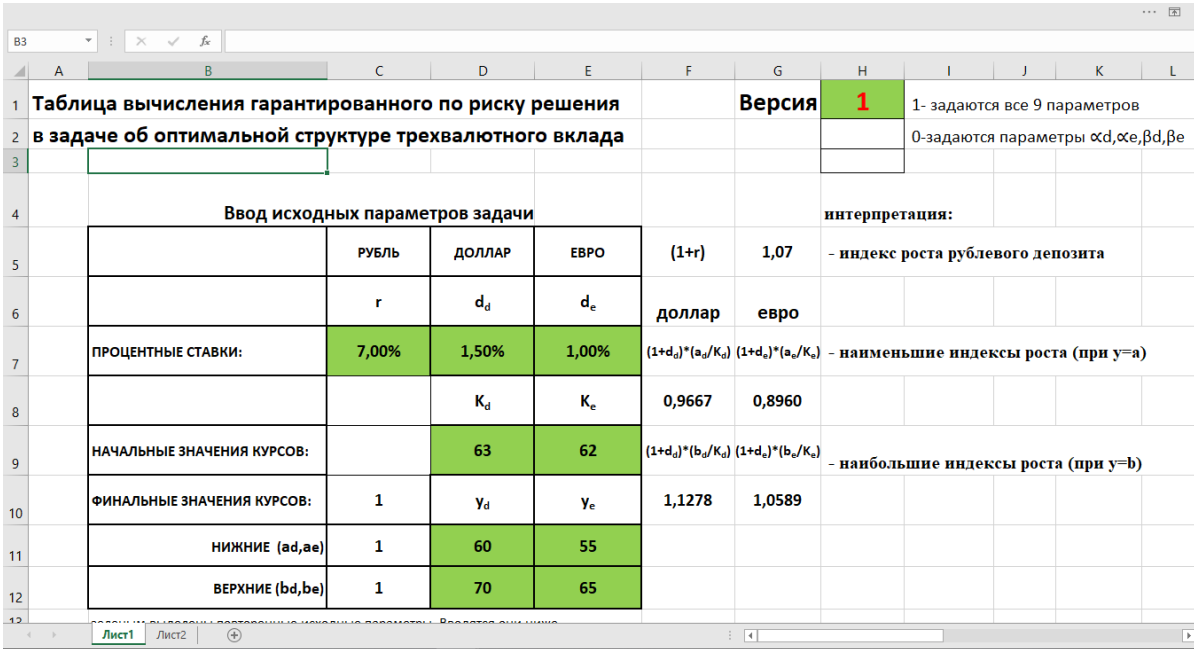
В общем случае аналитический поиск оптимального решения задачи диверсификации затруднителен и требуется привлечение численных методов и алгоритмов.

2.1.4 Вычислительная среда и результаты расчетов

Предложенный выше метод реализован в среде Excel 2016 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 последующему анализу автоматически полученных результатов.

Электронная таблица здесь согласована с Таблицей 1 из [110], содержит приведенные в ней расчетные формулы для точек, из которых хотя бы одна является искомым решением.

На рисунке 2.1 воспроизведена область для ввода исходных данных. Входные ячейки отмечены тонировкой. Справа приведены для сведения значения индексов роста валют при наименьших и наибольших значениях курсов.



1	**Таблица вычисления гарантированного по риску решения**					**Версия**	**1**	1- задаются все 9 параметров
2	**в задаче об оптимальной структуре трехвалютного вклада**							0-задаются параметры ∝d,∝e,βd,βe
3								
4		**Ввод исходных параметров задачи**					**интерпретация:**	
5		**РУБЛЬ**	**ДОЛЛАР**	**ЕВРО**	**(1+r)**	**1,07**	**- индекс роста рублевого депозита**	
6		**r**	**d_d**	**d_e**	**доллар**	**евро**		
7	**ПРОЦЕНТНЫЕ СТАВКИ:**	**7,00%**	**1,50%**	**1,00%**	**$(1+d_d)*(a_d/K_d)$**	**$(1+d_e)*(a_e/K_e)$**	**- наименьшие индексы роста (при y=a)**	
8			**K_d**	**K_e**	**0,9667**	**0,8960**		
9	**НАЧАЛЬНЫЕ ЗНАЧЕНИЯ КУРСОВ:**		**63**	**62**	**$(1+d_d)*(b_d/K_d)$**	**$(1+d_e)*(b_e/K_e)$**	**- наибольшие индексы роста (при y=b)**	
10	**ФИНАЛЬНЫЕ ЗНАЧЕНИЯ КУРСОВ:**	**1**	**y_d**	**y_e**	**1,1278**	**1,0589**		
11	**НИЖНИЕ (ad,ae)**	**1**	**60**	**55**				
12	**ВЕРХНИЕ (bd,be)**	**1**	**70**	**65**				

Рисунок 2.1 – Ввод исходных данных. Въодные ячейки

На рисунке 2.2 представлена часть листа Excel с расчетами, соответствующими приведенным исходным параметрам из Рисунка 2.1. Столбец B содержит текст расчетных формул, в столбцах C и D производятся соответствующие вычисления. Результаты проверяются с использованием логических операторов на выполнение условия $x \in X$ (расшифровка условий в колонках J и K). При невыполнении условия в ячейке столбца E записывается текст «НЕ КАНДИДАТ», иначе – результат вычисления функции гарантированного риска в соответствующей точке. Далее вычисляется минимальный из этих результатов. Соответствующая ячейка (или ячейки в случае неединственности) средствами условного форматирования выделяется цветом, слева от нее расположено оптимальное решение. По значениям целевой функции в остальных допустимых точках-кандидатах можно оценить увеличение риска при уменьшении числа используемых валют.

A	B	C	D	E	F	G	H	I	J	K
		ТАБЛИЦА ДЛЯ ПРОВЕРОК, ИСКЛЮЧЕНИЯ "ЛИШНИХ" ТОЧЕК И ОТБОРА ТОЧКИ МИНИМУМА								
	ФОРМУЛА ДЛЯ РАСЧЕТА	Xd	Xe	ИТОГ ПРОВЕРКИ	Ф	Ф1	Ф2	Ф3	ЧТО ПРОВЕРЯЕМ	
1	Вершина O(0,0) (только рубли)	0,000	0,000	0,058	0,058	0,000	0,058	-0,011	ПРОВЕРКА НЕ НУЖНА	
2	Вершина A(1,0) (доллары и рубли)	1,000	0,000	0,103	0,103	0,103	0,000	0,092	ПРОВЕРКА НЕ НУЖНА	
3	Вершина B(0,1) (евро и рубли)	0,000	1,000	0,232	0,232	0,174	0,232	0,000	ПРОВЕРКА НЕ НУЖНА	
4	$x_d^{OA} = \frac{\beta_d - \beta_e(sign(\beta_e)+1)/2}{\alpha_d + \beta_d}$	0,359	0,000	0,037	0,037	0,037	0,037	0,026	1>xd>0	
5	$x_e^{OB} = \frac{\beta_e - \beta_d(sign(\beta_d)+1)/2}{\alpha_e + \beta_e}$	0,000	-0,423	НЕ КАНДИДАТ	-0,016	-0,074	-0,016	-0,016	1>xe>0	
6	$x_e^{12} = \frac{\alpha_d}{\alpha_d + \beta_d}$	0,359	0,641	0,149	0,149	0,149	0,149	0,033	xd,xe>0	
7	$x_e^{13} = \frac{\beta_e}{\alpha_e + \beta_e}$	1,068	-0,068	НЕ КАНДИДАТ	0,099	0,099	-0,016	0,099	xd,xe>0	
8	$x_e^{23} = \frac{\alpha_d + \beta_e}{\alpha_e + \alpha_d + \beta_e + \beta_d}$	0,715	0,285	0,123	0,123	0,123	0,066	0,066	xd,xe>0	
9	$x^* = (x_d^*, x_e^*) = (\frac{\beta_d}{\alpha_d + \beta_d}, \frac{\beta_e}{\alpha_e + \beta_e})$	0,359	-0,068	НЕ КАНДИДАТ	0,025	0,025	0,025	0,025	xd,xe>0,xd+xe<1	
	ТОЧКА ОПТИМУМА:	0,359	0,000		0,037057	0,037	0,037	0,026	xd,xe>0,xd+xe<2	
Ответ:	Минимальный гарантированный риск Ф*=			0,037057471						

Рисунок 2.2 – Расчёт оптимального решения

Исследовательский вариант содержит ряд дополнительных возможностей. Помимо использования исходных данных, в нем возможно прямое задание производных параметров $\alpha_d, \alpha_e, \beta_d, \beta_e$, также полностью определяющих решение. Так, расчеты на Рисунке 2.3 соответствуют рассмотренному выше аналитически случаю $\alpha_d = \alpha_e = 0$ при $\beta_d = 0{,}2,\ \beta_e = 0{,}3$.

	ТАБЛИЦА ДЛЯ ПРОВЕРОК, ИСКЛЮЧЕНИЯ "ЛИШНИХ" ТОЧЕК И ОТБОРА ТОЧКИ МИНИМУМА								
ФОРМУЛА ДЛЯ РАСЧЕТА	xd	xe	ИТОГ ПРОВЕРКИ	Ф	Ф1	Ф2	Ф3	ЧТО ПРОВЕРЯЕМ	
Вершина O(0,0) (только рубли)	0,000	0,000	0,300	0,300	0,000	0,200	0,300	ПРОВЕРКА НЕ НУЖНА	
Вершина A(1,0) (доллары и рубли)	1,000	0,000	0,300	0,300	0,000	0,000	0,300	ПРОВЕРКА НЕ НУЖНА	
Вершина B(0,1) (евро и рубли)	0,000	1,000	0,200	0,200	0,000	0,200	0,000	ПРОВЕРКА НЕ НУЖНА	
$x_d^{OA} = \frac{\beta_d - \beta_e(sign(\beta_e)+1)/2}{\alpha_d + \beta_d}$	-0,500	0,000	НЕ КАНДИДАТ	0,300	0,000	0,300	0,300	1>xd>0	
$x_e^{OB} = \frac{\beta_e - \beta_d(sign(\beta_d)+1)/2}{\alpha_e + \beta_e}$	0,000	0,333	0,200	0,200	0,000	0,200	0,200	1>xe>0	
$x_e^{12} = \frac{\alpha_d}{\alpha_d + \beta_d}$	1,000	0,000	НЕ КАНДИДАТ	0,300	0,000	0,000	0,300	xd,xe>0	
$x_e^{13} = \frac{\beta_e}{\alpha_e + \beta_e}$	0,000	1,000	НЕ КАНДИДАТ	0,200	0,000	0,200	0,000	xd,xe>0	
$x_e^{23} = \frac{\alpha_d + \beta_e}{\alpha_e + \alpha_d + \beta_e + \beta_d}$	0,400	0,600	0,120	0,120	0,000	0,120	0,120	xd,xe>0	
$x^* = (x_d^*, x_e^*) = (\frac{\beta_d}{\alpha_d + \beta_d}, \frac{\beta_e}{\alpha_e + \beta_e})$	1,000	1,000	НЕ КАНДИДАТ	0,000	0,000	0,000	0,000	xd,xe>0,xd+xe<1	
Минимальный гарантированный риск Ф*=			0,12						

Рисунок 2.3 – Случай равной доходности валют: $\alpha_d = \alpha_e = 0;\ \beta_d = 0{,}2,\ \beta_e = 0{,}3$

Однако прямое задание производных параметров не может быть произвольным, так как их значения должны быть согласованы с исходными параметрами (см. формулы (5) и

(6)). Например, как отмечалось выше, если $\propto_d = \propto_e = 0$,то параметры β_d, β_e не могут быть отрицательными.

В программной реализации алгоритма предусмотрены средства проведения множественных расчетов, при которых варьируется один или два параметра. В частности, можно:

− изучить зависимость оптимального риска от верхних пределов курсов b_d, b_e при фиксированных остальных исходных данных;

− построить таблицу зависимости оптимальной структуры депозита и оптимального риска от процентной ставки для рублевого вклада;

− построить таблицу зависимости оптимального риска от значений процентных ставок доллара и евро;

− представить результаты в графическом виде.

Возможны и другие постановки экспериментов.

В Таблице 2.1 приведены результаты множественных расчетов оптимального риска при всех сочетаниях значений верхних границ курсов b_d, b_e от 66 до 85 с шагом 1.

Таблица 2.1 – Зависимость оптимального риска от верхних границ курсов валют b_d, b_e. В верхней строке – значения курса доллара, в левом столбце – евро

0	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85
66	0,0050	0,0091	0,0215	0,0312	0,0389	0,0452	0,0505	0,0550	0,0588	0,0621	0,0650	0,0676	0,0698	0,0719	0,0737	0,0754	0,0769	0,0782	0,0795	0,0807
67	0,0191	0,0201	0,0247	0,0359	0,0448	0,0520	0,0581	0,0632	0,0676	0,0714	0,0748	0,0777	0,0803	0,0827	0,0848	0,0867	0,0884	0,0900	0,0914	0,0928
68	0,0310	0,0327	0,0356	0,0405	0,0506	0,0588	0,0657	0,0715	0,0764	0,0808	0,0845	0,0879	0,0908	0,0935	0,0958	0,0980	0,0999	0,1017	0,1034	0,1049
69	0,0412	0,0435	0,0473	0,0511	0,0564	0,0656	0,0733	0,0797	0,0853	0,0901	0,0943	0,0980	0,1013	0,1042	0,1069	0,1093	0,1115	0,1135	0,1153	0,1170
70	0,0501	0,0528	0,0574	0,0621	0,0667	0,0724	0,0808	0,0880	0,0941	0,0994	0,1040	0,1081	0,1118	0,1150	0,1180	0,1206	0,1230	0,1246	0,1259	0,1271
71	0,0578	0,0610	0,0663	0,0717	0,0770	0,0824	0,0884	0,0962	0,1029	0,1087	0,1138	0,1183	0,1223	0,1258	0,1280	0,1297	0,1313	0,1328	0,1342	0,1356
72	0,0647	0,0682	0,0742	0,0801	0,0861	0,0921	0,0981	0,1045	0,1118	0,1181	0,1236	0,1284	0,1312	0,1333	0,1352	0,1371	0,1389	0,1405	0,1421	0,1437
73	0,0707	0,0746	0,0811	0,0877	0,0942	0,1008	0,1073	0,1139	0,1206	0,1274	0,1327	0,1352	0,1376	0,1399	0,1421	0,1441	0,1461	0,1479	0,1497	0,1514
74	0,0762	0,0803	0,0874	0,0944	0,1015	0,1085	0,1156	0,1226	0,1297	0,1353	0,1383	0,1411	0,1437	0,1462	0,1485	0,1508	0,1529	0,1550	0,1569	0,1587
75	0,0811	0,0855	0,0930	0,1005	0,1080	0,1155	0,1230	0,1305	0,1371	0,1404	0,1436	0,1466	0,1494	0,1521	0,1547	0,1571	0,1594	0,1617	0,1638	0,1658
76	0,0855	0,0901	0,0981	0,1060	0,1139	0,1218	0,1297	0,1376	0,1416	0,1452	0,1486	0,1518	0,1549	0,1578	0,1605	0,1631	0,1657	0,1680	0,1703	0,1725
77	0,0895	0,0944	0,1027	0,1110	0,1192	0,1275	0,1358	0,1419	0,1460	0,1498	0,1534	0,1568	0,1600	0,1631	0,1661	0,1689	0,1716	0,1741	0,1766	0,1790
78	0,0932	0,0983	0,1069	0,1155	0,1241	0,1328	0,1413	0,1458	0,1500	0,1541	0,1579	0,1615	0,1650	0,1682	0,1714	0,1744	0,1772	0,1800	0,1826	0,1851
79	0,0966	0,1018	0,1107	0,1197	0,1286	0,1376	0,1447	0,1494	0,1539	0,1581	0,1622	0,1660	0,1696	0,1731	0,1764	0,1796	0,1827	0,1856	0,1884	0,1910
80	0,0997	0,1051	0,1143	0,1235	0,1328	0,1420	0,1480	0,1529	0,1576	0,1620	0,1662	0,1703	0,1741	0,1778	0,1813	0,1846	0,1879	0,1909	0,1939	0,1967
81	0,1025	0,1081	0,1176	0,1271	0,1366	0,1456	0,1510	0,1562	0,1610	0,1657	0,1701	0,1743	0,1784	0,1822	0,1859	0,1894	0,1928	0,1961	0,1992	0,2022
82	0,1052	0,1109	0,1206	0,1304	0,1401	0,1483	0,1539	0,1593	0,1644	0,1692	0,1738	0,1782	0,1824	0,1865	0,1903	0,1940	0,1976	0,2010	0,2043	0,2074
83	0,1076	0,1135	0,1234	0,1334	0,1434	0,1509	0,1567	0,1622	0,1675	0,1725	0,1773	0,1819	0,1863	0,1905	0,1946	0,1984	0,2022	0,2057	0,2092	0,2125
84	0,1099	0,1159	0,1261	0,1362	0,1464	0,1533	0,1593	0,1650	0,1705	0,1757	0,1807	0,1855	0,1901	0,1944	0,1986	0,2027	0,2066	0,2103	0,2139	0,2173
85	0,1121	0,1181	0,1285	0,1389	0,1491	0,1556	0,1618	0,1677	0,1734	0,1788	0,1839	0,1889	0,1936	0,1982	0,2025	0,2067	0,2108	0,2147	0,2184	0,2220

Полученные данные удобно анализировать графическими средствами Excel. Так, на рисунке 2.4 представлены графики для первых 10 строк Таблицы 2.1. Для остальных 10 значений b_e картина аналогична. Подобным же образом выглядят и графики для столбцов таб-

лицы. Как и следовало ожидать, оптимальный риск является нелинейной возрастающей функцией по каждому параметру b_d, b_e.

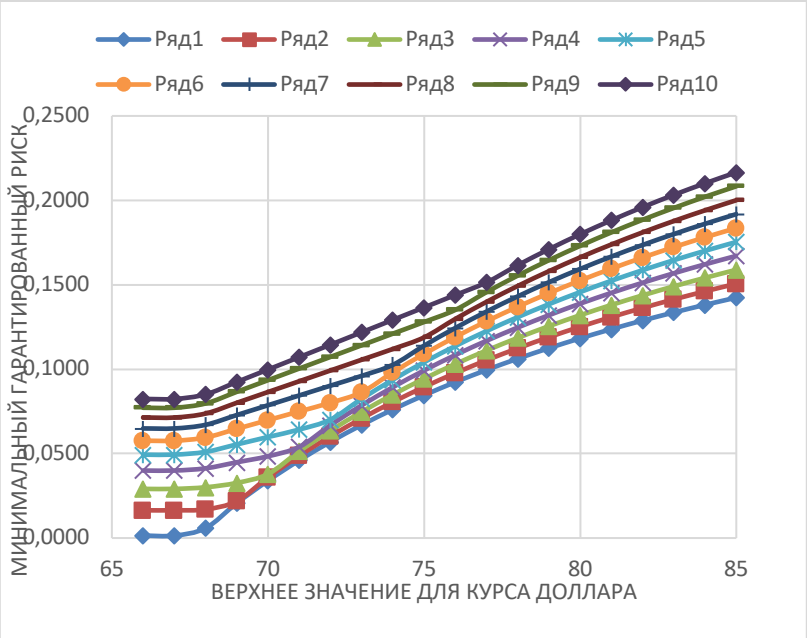


Рисунок 2.4 – Графики зависимости оптимального риска от параметра b_d при различных значениях параметра b_e

В Таблицу 2.2 приведены оптимальные структуры вклада при изменении процентной ставки для рубля от 1% до 10% с шагом 0,5%. Прочие параметры фиксированы на следующих значениях: ставка по доллару – 1,5%, по евро – 1%, начальные курсы – 63 руб./долл. и 62 руб./евро, границы курса доллара – от 60 до 70, евро – от 55 до 65. В последней строке приведены соответствующие значения оптимального гарантированного риска. Как видно, он монотонно убывает с ростом ставки рубля. Начиная со ставки 1,5%, доллар замещается на евро, далее доля рубля монотонно возрастает.

Рисунок 2.5 дает наглядное представление о динамике структуры.

Таблица 2.2 – Зависимость структуры депозита от процентной ставки по рублю

Ставка	1,0%	1,5%	2,0%	2,5%	3,0%	3,5%	4,0%	4,5%	5,0%	5,5%	6,0%	6,5%	7,0%	7,5%	8,0%	8,5%	9,0%	9,5%	10,0%
доллар	0,715	0,428	0,428	0,428	0,428	0,428	0,428	0,428	0,428	0,428	0,421	0,390	0,359	0,328	0,297	0,266	0,234	0,203	0,172
евро	0,285	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
рубль	0,000	0,572	0,572	0,572	0,572	0,572	0,572	0,572	0,572	0,572	0,579	0,610	0,641	0,672	0,703	0,734	0,766	0,797	0,828
риск	0,066	0,065	0,062	0,059	0,056	0,053	0,050	0,047	0,045	0,042	0,039	0,038	0,037	0,035	0,034	0,031	0,029	0,026	0,023

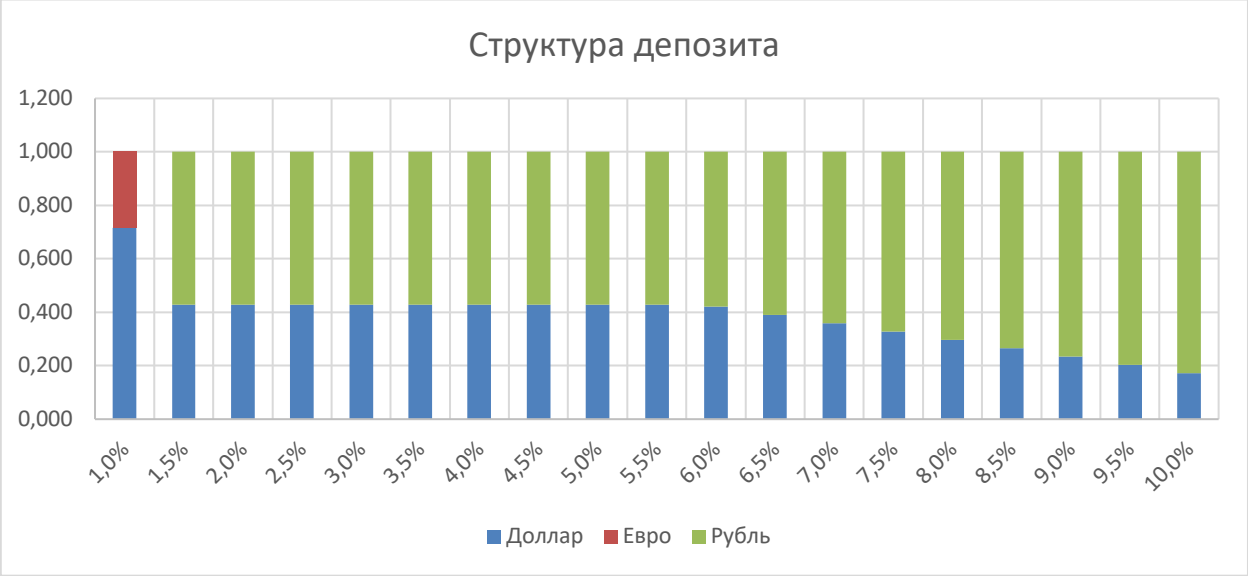


Рисунок 2.5 – Оптимальная структура депозита

В следующей Таблице 2.3 приведены результаты вычислений минимального гарантированного риска при изменении процентных ставок доллара и евро в диапазоне от -1% до 3,75% с шагом 0,25%. Прочие параметры фиксированы на следующих значениях: начальные курсы – 63 руб./долл. и 62 руб./евро, границы курса доллара – от 60 до 75, евро – от 55 до 70.

Таблица 2.3 – Оптимальный гарантированный риск при различных процентных ставках доллара (верхняя строка) и евро (левый столбец)

0	-1,00%	-0,75%	-0,50%	-0,25%	0,00%	0,25%	0,50%	0,75%	1,00%	1,25%	1,50%	1,75%	2,00%	2,25%	2,50%	2,75%	3,00%	3,25%	3,50%	3,75%
-1,00%	0,0806	0,0814	0,0822	0,0830	0,0837	0,0843	0,0849	0,0854	0,0858	0,0862	0,0865	0,0867	0,0869	0,0871	0,0871	0,0872	0,0871	0,0870	0,0869	0,0867
-0,75%	0,0819	0,0828	0,0836	0,0844	0,0851	0,0858	0,0863	0,0869	0,0873	0,0877	0,0881	0,0884	0,0886	0,0888	0,0889	0,0889	0,0889	0,0889	0,0887	0,0885
-0,50%	0,0832	0,0841	0,0850	0,0858	0,0865	0,0872	0,0878	0,0884	0,0889	0,0893	0,0897	0,0900	0,0903	0,0905	0,0906	0,0907	0,0907	0,0907	0,0906	0,0904
-0,25%	0,0845	0,0854	0,0863	0,0872	0,0880	0,0887	0,0893	0,0899	0,0904	0,0909	0,0913	0,0917	0,0920	0,0922	0,0924	0,0925	0,0925	0,0925	0,0924	0,0923
0,00%	0,0858	0,0868	0,0877	0,0886	0,0894	0,0901	0,0908	0,0914	0,0920	0,0925	0,0929	0,0933	0,0936	0,0939	0,0941	0,0942	0,0943	0,0943	0,0943	0,0942
0,25%	0,0871	0,0881	0,0891	0,0900	0,0908	0,0916	0,0923	0,0930	0,0936	0,0941	0,0946	0,0950	0,0953	0,0956	0,0958	0,0960	0,0961	0,0962	0,0962	0,0961
0,50%	0,0884	0,0894	0,0904	0,0914	0,0922	0,0931	0,0938	0,0945	0,0951	0,0957	0,0962	0,0966	0,0970	0,0973	0,0976	0,0978	0,0979	0,0980	0,0980	0,0980
0,75%	0,0897	0,0908	0,0918	0,0928	0,0937	0,0945	0,0953	0,0960	0,0967	0,0973	0,0978	0,0983	0,0987	0,0990	0,0993	0,0995	0,0997	0,0998	0,0999	0,0999
1,00%	0,0910	0,0921	0,0932	0,0942	0,0951	0,0960	0,0968	0,0975	0,0982	0,0988	0,0994	0,0999	0,1003	0,1007	0,1010	0,1013	0,1015	0,1016	0,1017	0,1018
1,25%	0,0923	0,0934	0,0945	0,0956	0,0965	0,0974	0,0983	0,0991	0,0998	0,1004	0,1010	0,1015	0,1020	0,1024	0,1028	0,1031	0,1033	0,1035	0,1036	0,1036
1,50%	0,0936	0,0948	0,0959	0,0970	0,0980	0,0989	0,0998	0,1006	0,1013	0,1020	0,1026	0,1032	0,1037	0,1041	0,1045	0,1048	0,1051	0,1053	0,1054	0,1055
1,75%	0,0949	0,0961	0,0972	0,0983	0,0994	0,1004	0,1013	0,1021	0,1029	0,1036	0,1043	0,1048	0,1054	0,1058	0,1063	0,1066	0,1069	0,1071	0,1073	0,1074
2,00%	0,0962	0,0974	0,0986	0,0997	0,1008	0,1018	0,1027	0,1036	0,1044	0,1052	0,1059	0,1065	0,1071	0,1076	0,1080	0,1084	0,1087	0,1090	0,1092	0,1093
2,25%	0,0975	0,0987	0,1000	0,1011	0,1022	0,1033	0,1042	0,1051	0,1060	0,1068	0,1075	0,1081	0,1087	0,1093	0,1097	0,1101	0,1105	0,1108	0,1110	0,1104
2,50%	0,0988	0,1001	0,1013	0,1025	0,1037	0,1047	0,1057	0,1067	0,1075	0,1084	0,1091	0,1098	0,1104	0,1110	0,1115	0,1119	0,1123	0,1126	0,1120	0,1114
2,75%	0,1001	0,1014	0,1027	0,1039	0,1051	0,1062	0,1072	0,1082	0,1091	0,1099	0,1107	0,1114	0,1121	0,1127	0,1132	0,1137	0,1141	0,1136	0,1129	0,1123
3,00%	0,1014	0,1027	0,1041	0,1053	0,1065	0,1077	0,1087	0,1097	0,1106	0,1115	0,1123	0,1131	0,1138	0,1144	0,1149	0,1154	0,1150	0,1144	0,1138	0,1132
3,25%	0,1027	0,1041	0,1054	0,1067	0,1080	0,1091	0,1102	0,1112	0,1122	0,1131	0,1139	0,1147	0,1154	0,1161	0,1167	0,1163	0,1158	0,1153	0,1147	0,1141
3,50%	0,1040	0,1054	0,1068	0,1081	0,1094	0,1106	0,1117	0,1128	0,1138	0,1147	0,1156	0,1164	0,1171	0,1178	0,1175	0,1171	0,1166	0,1161	0,1156	0,1150
3,75%	0,1053	0,1067	0,1082	0,1095	0,1108	0,1120	0,1132	0,1143	0,1153	0,1163	0,1172	0,1180	0,1188	0,1186	0,1182	0,1178	0,1174	0,1169	0,1164	0,1158

2.1.5 Некоторые выводы

Разработаны вычислительные средства для решения задачи о нахождении структур трехвалютного депозита, оптимальных по гарантированному риску, в условиях неизвестных будущих обменных курсов валют.

Применение в качестве вычислительной среды Excel позволило разработать вычислительные средства, сочетающие простоту использования для лиц, имеющих минимальные

навыки работы в электронных таблицах, с обширными средствами анализа данных для продвинутых пользователей.

Метод реализован в двух вариантах – пользовательском и исследовательском. Первый рассчитан на лиц, имеющего минимальные навыки работы в простой готовой таблице Excel. Работа в нем сводится к вводу упомянутых выше исходных данных и содержательному анализу автоматически полученных результатов. Второй, исследовательский вариант содержит ряд дополнительных возможностей: прямое задание производных параметров доходности отдельных валют, множественные расчеты с варьированием одного или двух параметров задачи, графическую визуализацию получаемых результатов.

Проведены отладочные и экспериментальные расчеты. В частности, случай равной гарантированной доходности всех трех валют (при минимальных будущих курсах) исследован полностью аналитически и использовался для отладки программы.

Проведено численное моделирование зависимости минимального гарантированного риска от максимальных возможных будущих значений курсов валют. Выявлено монотонное возрастание риска по каждой из указанных переменных.

Построены таблицы зависимости оптимальной структуры депозита от процентной ставки по рублю (при фиксированных остальных параметрах), а также таблицы зависимости оптимального риска от процентных ставок по доллару и евро.

Реализованный алгоритм предоставляет опытному пользователю Excel и иные возможности анализа рассматриваемой задачи средствами раздела «Анализ «что, если» и другими.

Полученные результаты изложены в статьях [110, 112]. Они могут быть использованы при анализе проблем управления финансами в условиях неполной информации.

2.2 Формирование гибких бизнес-правил для улучшения процесса закупки компонентов для систем управления

2.2.1 Введение

Рассмотрим компанию, которая производит блоки для системы управления движением поездов. Блоки производятся в большом количестве. Они устанавливаются на локомотивах. Для производства блоков необходимы микроконтроллеры. Базовая цена на микроконтроллер установлена в одной из резервных валют. Имеются возможные поставщики. Они находятся в разных странах. Покупатель должен учитывать, что конвертация валюты его страны в конечную валюту покупки будет проходить в два этапа. Сначала конвертация в валюту страны поставщика. Затем, конвертация валюты поставщика в валюту покупки. Кроме того, цена микроконтроллера у каждого поставщика может отличаться от базовой. Компания

знает сколько микроконтроллеров она должна закупить в месяц для выполнения плана производства. Закупку на месяц можно проводить в два этапа: в начале месяца и в середине. На начала месяца курсы конвертации известны. Курсы, которые будут в середине месяца точно неизвестны. Однако, можно предсказать диапазоны их изменения. Компания желает потратить как можно меньше валюты своей страны на закупку. Для этого необходимо правильно выбрать моменты закупки и поставщика на мировом рынке. Актуальной задачей является создание гибких правил для осуществления такого выбора. Эта задача решена в данной работе на основании критерия Сэвиджа – критерия минимума максимального риска. Работа [113] является основной работой, на базе которой получено решение.

2.2.2 Анализ проблемы

Большое число работ в сфере закупок посвящена переходам от одного инструмента к другому. В большинстве необходимость перехода основывается на предсказания об изменении стоимости оборудования [114]. Другие ссылаются на финансовые кризисы и как на это отреагировал рынок [115]. Обычно мы используем технический анализ, фундаментальный анализ или их комбинации для прогнозирования. Технический анализ в данном случае это попытка спрогнозировать будущие значения основываясь на прошлых значениях учитывая массовое поведение игроков на рынке.

Научный аппарат технического анализа очень разнообразен. В нем используются как простые визуальные методы, так и сложнейшие численные методы последовательных приближений — нейронные сети. В первую очередь технический анализ применяется при игре на бирже с акциями и на валютных рынках, где имеются большие массивы исторических данных. Многие исследователи скептически относятся к техническому анализу. Они ставят под сомнение возможность предсказания будущего по прошлому без объяснения структуры этого явления. Фундаментальный анализ базируется на экономической теории и объясняет причины изменения стоимости инструментов, но точность этих прогнозов не всегда достаточна для принятия обоснованных решений. Методы теории игр также применяются в первую очередь в теории формирования оптимального портфеля инвестиций. Классический подход здесь — теория Марковица.

Отличие рассматриваемой задачи от задач, решаемых ранее, состоит в том, что обмен валюты по заданной цене в краткосрочной перспективе обязателен. Соответственно нужно выполнить его наилучшим образом.

2.2.3 Математическая постановка проблемы

Компания должна закупить определённое количество микроконтроллеров. На международном рынке имеется n возможных поставщиков. Известна сумма S_i в валюте покупки,

которая потребуется для закупки всей партии контроллеров у поставщика i (i = 1, 2, …, n). Закупку можно проводить в два этапа: в начале месяца – момент t_0 и в середине месяца – момент t_1. Компания платит в валюте своей страны. Конвертация проходит в два этапа. Сначала конвертация в валюту страны поставщика. Затем, конвертация валюты поставщика в валюту покупки. В момент t_0 для покупки единицы валюты страны поставщика i надо продать a_{0i} единиц валюты страны покупателя. Для покупки в момент t_0 единицы конечной валюты, по которой закупается микроконтроллер, поставщик i должен продать b_{0i} единиц валюты своей страны. Оба курса конвертации в момент t_1 неизвестны. Однако, можно предполагать, что они находятся в промежутках $[a_i; A_i]$ и $[b_i; B_i]$, соответственно.

Требуется построить гибкие бизнес-правила для рационального определения поставщика и моментов закупки у него.

2.2.4 Решение задачи

Для решения задачи надо определить критерий оптимальности.

В этом случае для решения задачи целесообразно применять методы теории игр, а именно методы принятия решений в условиях неопределенности, которая не может быть оценена с помощью методов теории вероятностей. Существует два основных критерия принятия решения в таких ситуациях: критерий Вальда [116] и критерий Сэвиджа хотя иногда используются и другие критерии как, например, критерий Гурвица. Согласно критерию Вальда оптимальна стратегия, при которой минимальный выигрыш максимален. Это критерий «бедного пессимиста», который считает, что условия проведения операции будут для него худшими, и в то же время необходимо получить небольшой, но гарантированный выигрыш.

Критерий Сэвиджа основан на понятии риска. Риск, по определению Сэвиджа, есть разница между тем выигрышем, который мы могли бы получить, если бы заранее знали, какими будут условия проведения операции, и выигрышем, который будет получен от применения выбранной стратегии. По критерию Сэвиджа оптимальная стратегия, при которой максимальный риск минимален. Это критерий пессимиста, но пессимист слишком богат, учитывая, что самый страшный исход — это упустить выигрыш [105]. В нашем случае решение, основанное на критерии Сэвиджа, не допустит значительных отклонений от минимально возможной суммы, необходимой для покупки [117]. Воспользуемся критерий Сэвиджа в качестве критерия оптимальности для решения поставленной задачи.

В [113] решена следующая задача. Сегодня (в момент t_0) у компании есть деньги в евро. Сегодняшний курс продажи евро за рубли равен E_0 [RUB/EUR]. Завтра (в момент t_1)

компании необходимо заплатить сумму S в рублях. Завтрашний курс неизвестен. Однако есть основания предполагать, что она будет меняться в пределах от E_{min} до E_{max} и

$$E_0 \ \epsilon \ [E_{min};\ E_{max}]. \tag{2.12}$$

Требуется определить, какую часть инструментов в евро продать сегодня, а какую завтра для получения требуемой суммы денег в рублях.

Возьмем x количество евро, проданное сегодня по курсу E_0. Доказано, что минимум максимального риска достигается в точке x, где

$$x\left(1-\frac{E_0}{E_{min}}\right)+S\left(\frac{1}{E_{min}}-\frac{1}{E_0}\right)=\ x\left(1-\frac{E_0}{E_{max}}\right). \tag{2.13}$$

Оптимальная точка

$$x^*=\frac{E_{max}(E_0-E_{min})}{E_0^2(E_{max}-E_{min})}S. \tag{2.14}$$

Применительно к нашему случаю при покупке через поставщика i

$$E_0 = a_{0i}b_{0i};\ E_{min} = a_ib_i;\ E_{max} = A_iB_i;\ S = S_i. \tag{2.15}$$

Подставим (2.14) в (2.3). Учтём (2.15). Получим следующее выражение для минимума максимального риска покупки микроконтроллеров через поставщика i

$$R_i=\frac{(a_{0i}b_{0i}-a_ib_i)(A_iB_i-a_{0i}b_{0i})}{a_{0i}^2b_{0i}^2(A_iB_i-\ a_ib_i)}S_i. \tag{2.16}$$

Наиболее сложным для выбора является случай, когда для всех i выполняется условие аналогичное условию (2.12)

$$a_{0i}b_{0i}\ \epsilon\ [a_ib_i;\ A_iB_i]. \tag{2.17}$$

Полученные результаты позволяют сформулировать гибкие бизнес-правила выбора поставщика и моментов закупки для этого важнейшего случая. Эти правила таковы.

1) Определить поставщика, для которого значение R_i минимально.

2) Допустим это поставщик с номером k. В начале месяца потратить на закупку микроконтроллеров сумму

$$x_k^* = \frac{A_k B_k (a_{0k} b_{0k} - a_k b_k)}{a_{0k}^2 b_{0k}^2 (A_k B_k - a_k b_k)} S_k. \quad (2.18)$$

В середине месяца докупить на оставшеюся необходимую сумму по фактическим курсам на дату покупки.

2.2.5 Численный пример

Имеются два возможных поставщика. Один — в Китае. Один — в Индии. Конечная цена в долларах США (USD) предлагаемая ими: 1 200 000 USD и 1 050 000 USD, соответственно. Покупка через первого поставщика (i=1) предполагает предварительную конвертацию российских рублей (RUR) в юани (CNY). Покупка через второго поставщика (i=2) предполагает предварительную конвертацию российских рублей в индийские рупии (INR).

Курсы на начало месяца: a_{01} = 0,0611 [CNY/RUR], a_{02} = 0,7353 [INR/RUR], b_{01} = 0,1575 [USD/CNY], b_{02} = 0,0131[USD/INR]. Ожидаемые диапазоны изменения к середине месяца: CNY/RUR − [0,0525;0,0688], CNY/USD − [0,1567;0,1575], INR/RUR − [0,5495;0,9009], INR/USD − [0,0131;0,0132].

Требуется определить наилучшего поставщика и суммы в рублях, которые следует потратить на закупку в начале месяца.

Прежде всего убеждаемся, что условие (2.17) выполняется для обоих поставщиков.

Воспользуемся предложенным гибким бизнес – правилом.

1) По формуле (2.16) определяем минимаксные риски покупки у первого и второго поставщиков. Имеем R_1 = 8 397 017 RUR; R_2 = 13 264 368 RUR. Принимаем решение закупать микроконтроллеры через поставщика из Китая.

2) По формуле (2.18) определяем сумму, которую следует потратить на закупку в начале месяца. Она составит $x_1^* = 75\ 068\ 418\ RUR$.

2.2.6 Некоторые выводы

Наиболее важным в полученном результате представляется следующее. Предлагается принимать решение на основании предсказания диапазонов изменения курсов. Это практически осуществимо, в отличие, от прогноза значений курсов. Более того, результат устойчив, т.е. небольшие изменения границ предсказанных диапазонов не влияют резко на результат!

Полученные результаты изложены в статье [117]. Результат представляет особый интерес для менеджеров, ответственных за закупки комплектующих на международном рынке.

2.3 Математическое моделирование формирования оптимального набора учебных курсов по выбору

2.3.1 Введение

Качество государственного образования в колледжах/университетах всегда интересовало органы государственного управления разных уровней в каждой стране как “большую систему” обслуживания населения. По этой причине существующие в любой стране намерения улучшить это качество являются вполне естественными. Указанная система обслуживания должна предлагать способ приобретения знаний и навыков каждому из миллионов ее клиентов, которые хотят быть значимыми членами общества, и помогать всем им в достижении этих устремлений. Аккредитованные университеты и колледжи должны предлагать набор курсов для изучения предметов, а подмножество курсов для изучения, предлагаемых конкретному студенту, должно зависеть от будущей специализации (или профессии), выбранной студентом. Поэтому каждая страна заинтересована в таких колледжах/университетах, которые привлекали бы новых студентов: а) предоставляя образовательные услуги, интересующие студентов, и б) предлагая при этом такие услуги на уровне, признанном в мире [118].

Среди основных образовательных услуг, которые колледж/университет может предоставить для достижения этой цели, следует отметить

− предложение разнообразных наборов предметов для изучения в рамках каждой дисциплины в рамках каждой специализации, выбранной студентом;

− преподавание предметов, представляющих интерес для студента, таким образом, чтобы сделать изучение этих предметов увлекательным и обеспечивающим возможность глубокого их понимания каждым заинтересованным студентом;

− развитие способности к восприятию знаний у каждого заинтересованного студента;

− обучение студентов решению иллюстративных задач, близких к тем реальным задачам, с которыми они могут столкнуться в практической деятельности, после окончания колледжа/университета;

− оказание помощи каждому студенту в разработке оптимального учебного плана, исходя из существующей на данный момент способности студента к получению знаний вообще, и к изучению предлагаемых конкретных предметов, в частности, что сделало бы академические годы студента в колледже/университете увлекательным путешествием;

− оказание помощи каждому желающему студенту, успешно окончившему колледж/университет, в трудоустройстве по окончании учебы в нем.

Одна из услуг в указанном списке — оказание помощи каждому студенту колледжа/университета в разработке оптимальной учебной программы, исходя из (выявленной) существующей способности студента к получению знаний вообще, и к изучению предлагаемых конкретных предметов, в частности, - заслуживает особого внимания. Она должна представлять интерес, по крайней мере, для тех подразделений систем государственного управления, которые занимаются проблемами образования, и которые определяют порядок и размеры государственного финансирования учебных заведений страны.

1) Как известно, курсы, изучаемые студентом колледжа/университета, образуют две группы курсов, которые этот студент должен изучать в каждый конкретный период времени в течение каждого учебного года [119].

Первая группа состоит из так называемых обязательных (или основных) курсов. Эти курсы включают те, которые должен изучать каждый студент колледжа/университета, и успешное изучение каждого из них является обязательным для студента, независимо от того, в какой области этот студент планирует специализироваться (и получить диплом по окончании этого колледжа/университета). Обязательные курсы хорошо известны в каждом колледже/университете, хотя наборы конкретных обязательных курсов могут различаться для разных специальностей даже в одном и том же колледже/университете [120]. Для каждого такого обязательного курса администрация колледжа/университета устанавливает определенное количество часов, которые каждый студент должен провести в аудиторных занятиях с профессорами и преподавателями, ведущими курс (лекции, контрольные работы, лабораторные работы и т. д.), и количество часов, которые этот студент должен потратить на внеаудиторные занятия (чтобы подготовиться к каждому аудиторному занятию). Все вышеперечисленные часы, связанные с изучением курса, определяют количество так называемых кредитов, которые администрация колледжа/университета приписывает курсу и которые должен “заработать” студент, изучающий этот курс (хотя, как правило, он может заработать эти кредиты с разными оценками) [121]. Общее количество кредитов, которое должен “заработать” каждый студент колледжа/университета, изучая обязательные курсы во время учебы, полностью определяется администрацией колледжа/университета и известно заранее (хотя могут быть разными для разных колледжей/университетов).

Вторую группу образует подмножество так называемых курсов по выбору, т. е., курсов, которые студент может изучать по своему выбору из определенного набора специальных курсов, предлагаемых администрацией колледжа/университета для изучения в рамках каждой конкретной специализации [121]. Точно так же, как это происходит для каждого обя-

зательного курса, администрация колледжа/университета устанавливает для каждого курса по выбору:

а) количество часов, которые каждый студент должен проводить в аудиторных занятиях — с профессорами и преподавателями, ведущими курс по выбору (лекции, контрольные работы, лабораторные работы и др.) — и которые он должен потратить на внеаудиторные занятия (чтобы подготовиться к каждому аудиторному занятию);

б) количество кредитов, которые администрация колледжа/университета приписывает курсу и которые должен “заработать” каждый студент, успешно изучивший этот курс (хотя, как правило, студент может заработать эти кредиты с разными оценками).

В каждом колледже/университете существуют экспертные оценки, связанные с каждым курсом по выбору, предлагаемым студентам для обучения, в частности, экспертные оценки (минимального и/или среднего) количества часов, которые они должны потратить на успешное изучение каждого из разделов каждого курса по выбору, а также общее количество часов для успешного изучения курса по выбору в целом.

Как правило, студент колледжа/университета выбирает для изучения курсы по выбору, исходя из личного впечатления о том,

− что ей/ему было бы интересно изучать;

− что она/он способна (способен) успешно изучать вместе с обязательными курсами, которые ей/ему будут там предложены для изучения;

− насколько успешно она/он сдаст экзамены по этим курсам по выбору.

Это впечатление часто основывается на имеющейся у студента информации о каждом курсе от другого студента колледжа/университета (или от группы студентов), которые успешно изучили этот курс, а также от студентов, которые не смогли этого сделать. Именно это впечатление и оценки, которые студент ожидает (или, по крайней мере, надеется) получить, являются основными факторами, влияющими на принятие студентом решения о выборе курсов из второй группы. Однако при таком подходе, как правило, студент принимает решение по этому вопросу, рассматривая только лишь благоприятные сценарии изучения выбранных курсов из второй группы. Серьезные проблемы возникают, если на каком-либо этапе изучения какого-либо уже выбранного конкретного курса по выбору студент понимает, что не способен успешно изучить это курс и сдать требуемый экзамен. Такое положение дел становится еще более угрожающим, если число таких «проблемных» курсов из второй группы превышает единицу. (При этом следует иметь в виду, что если столь серьезные проблемы с изучением курса по выбору возникают у студента, то к их успешному решению подключается и администрация колледжа/университета.)

2) Предположим, что в распоряжении студента колледжа/университета имеется система принятия решений (например, такая, как предложена в рамках настоящего исследования) о включении курса по выбору в учебный план в любой момент в течение обучения в этом колледже/университете. Хотя в этом случае шансы на неуспешное изучение того или иного курса по выбору (или даже нескольких отдельных курсов) из любой группы курсов по выбору могут казаться меньше, чем без использования такой системы, любая количественная оценка этих шансов может быть сделана лишь на основе имеющейся информации о студенте, которую этот студент вводит о себе в систему сам. В любом случае, все оценки предлагаются системой на основе собственной оценки студентом своих способностей к получению знаний вообще и/или тех способностей, которые могут быть студентом отнесены к изучению конкретного курса или даже нескольких конкретных курсов, на момент получения оценки от системы.

Однако, если бы администрация колледжа/университета использовала ту же самую или аналогичную систему, то оценка способностей студента по всем курсам по выбору производилась бы профессионально, специалистами в области образования, и эта оценка могла бы основываться на результатах соответствующих тестов. Хотя выбор «курсов по выбору студента» из набора предлагаемых колледжем/университетом остался бы прерогативой каждого студента, в этом случае этот студент консультировался бы с администрацией колледжа/университета. Эта возможность, вероятно, повлияла бы на ее/его шансы принять правильное окончательное решение о наборе курсов по выбору, которые будут включены в ее/его учебную программу на каждый год и на весь период обучения в колледже/университете. Можно также ожидать, что предоставление студентам рекомендаций по выбору набора курсов по выбору для изучения (в любой конкретный период их обучения) наряду с оптимальным (с точки зрения администрации колледжа/университета) набором курсов по выбору для каждого студента, вероятно, снизило бы шансы студентов колледжа/университета потерпеть неудачу в изучении выбранных ими курсов.

3) Ясно, что:

а) предоставление студентам рекомендаций по принятию решения о выборе курсов из имеющегося набора курсов по выбору, предлагаемых администрацией колледжа/университета (для каждой конкретной специализации), и:

б) анализ результатов вышеуказанных тестов на способность студентов к успешному усвоению знаний вообще и к успешному изучению любого отдельного курса из множества предлагаемых курсов по выбору, позволило бы администрации колледжа/университета лучше определить окончательный состав наборов курсов по выбору, предлагаемых каждому

студенту по всем студенческим специализациям. Это положительно сказалось бы на репутации колледжа/университета и способствовало бы повышению качества образования в нем, что должно быть одной из целей каждой государственной администрации, управляющей территорией, на которой расположен колледж/университет. Кроме того, это, вероятно, положительно повлияло бы на бюджет колледжа/университета. (Последнее важно и для государственных, и для частных колледжей/университетов, с точки зрения получения финансовой поддержки от государства в любой форме.)

Как следует из вышеизложенного, представляется, что на качество образования в любом колледже/университете можно было бы положительно повлиять, если бы его администрация предложила обоснованные рекомендации по составу и структуре оптимальных подмножеств курсов по выбору каждому студенту в рамках каждой специализации, (по которой этот колледж/университет имеет право выпускать дипломированных специалистов). Однако разработка систем, способных формировать оптимальные подмножества такого типа для любого колледжа/университета, требует от него дополнительных затрат, которые в большинстве случаев не могут быть покрыты без государственных субсидий, выделяемых в сферу образования исключительно государственными органами.

Целью настоящего исследования является:

− создание системных проработок, направленных на привлечение внимания государственных администраций к уникальной проблеме высшего образования, одинаково интересной, как для колледжей/университетов, так и для студентов этих учебных заведений;

− разработка возможного подхода к решению этой проблемы.

Для этого предлагается

а) математическая модель в виде системы линейных неравенств с булевыми переменными для формализации задачи выбора состава обязательных курсов и курсов по выбору, который администрация колледжа/университет считает оптимальным для каждого своего студента в рамках каждой конкретной специализации;

б) другая математическая модель (в виде системы линейных неравенств с булевыми переменными, имеющими определенную структуру) для формализации задачи выбора такого состава курсов по выбору, который может показаться оптимальным студенту на основе его собственных соображений.

С помощью этих моделей,

а) задача нахождения оптимального (с точки зрения администрации колледжа/университета) подмножества курсов по выбору из всего набора курсов по выбору, предлагаемого колледжем/университетом в рамках каждой конкретной специализации, формули-

руется как оптимизационная задача дробного программирования с линейными ограничениями и булевыми переменными [122, 123];

б) задача поиска оптимального (теперь уже с точки зрения студента того же самого колледжа/университета) подмножества курсов по выбору из всего множества курсов по выбору (предлагаемых колледжем/университетом в рамках каждой конкретной специализации) формулируется как задача дискретной оптимизации специальной структуры с булевыми переменными (задача о рюкзаке с условиями предшествования) [124, 125];

в) задача оптимизации финансовых затрат колледжа/университета на организацию проведения курсов по выбору, достаточных для достижения студентами значения GPA ("grade point average" [126]) не ниже приемлемого для администрации этого колледжа/университета (в рамках каждой специализации, по которой этот колледж/университет имеет право выпускать дипломированных специалистов), формулируется как задача билинейного программирования со смешанными переменными [127].

Все эти задачи могут эффективно решаться с использованием стандартных пакетов прикладных программ для решения задач дискретной оптимизации, которые можно найти, например, в [128, 129].

2.3.2 Основная часть

2.3.2.1 Постановка задач

Для определенности далее рассматриваются только студенты программ бакалавриата университетов и студенты четырехлетних колледжей. Однако интуитивно ясно (и это будет показано далее), что все рассуждения остаются справедливыми и для магистерских программ. Таким образом, эти рассуждения применимы как для колледжей, присуждающих только степени бакалавра, так и для университетов и колледжей, присуждающих и степени бакалавра и степени магистра. Нетрудно убедиться (и это обсуждается в заключительном разделе представленного далее материала), что предлагаемые математические модели и сформулированные на их основе задачи имеют одинаковую структуру как для студентов бакалавриата, так и для студентов магистратуры, а отличаются эти модели только значениями некоторых их параметров.

Рассмотрим ситуацию, когда студент бакалавриата желает выбрать:

− подмножество курсов из множества курсов по выбору, предлагаемых для изучения колледжем/университетом (в дополнение к набору обязательных курсов, которые этот студент должен изучать в рамках выбранной специализации);

− разумную последовательность изучения выбранных курсов по выбору, исходя из последовательности изучения обязательных курсов, которая определяется администрацией колледжа/университета.

С точки зрения студента, реализация этих пожеланий позволила бы ему успешно окончить этот колледж/университет за четыре года со степенью бакалавра, присуждаемой колледжем/университетом в выбранной (этим студентом) области специализации. Этот студент считает свой выпуск успешным, если получает:

а) определенное количество кредитов;

б) максимально возможный GPA, что может повлиять (и, как правило, влияет) на шансы этого студента получить хорошее образование, хорошо оплачиваемую работу, и/или продолжить обучение в магистратуре в колледжах/университетах, в частности в тех, которые имеют право на государственную финансовую помощь.

Таким образом, в течение своих четырех академических лет в колледже/университете студент хотел бы изучить а) обязательные курсы и б) курсы по выбору для изучения (вместе с обязательными курсами) как можно более успешно. Каждый из этих двух типов курсов из учебного плана студента дает студенту определенное количество кредитов (при условии, что этот студент успешно сдает все соответствующие промежуточные тесты, относящиеся к этому курсу, и итоговый экзамен по этому курсу).

Общее количество кредитов, которое студент должен набрать в колледже/университете для получения желаемой степени бакалавра, не может быть меньше определенного числа, и это число, как правило, разное для разных специальностей, как внутри колледжа/университета, так и для разных колледжей.

Хотя студент не обязательно должен изучать какие-либо конкретные предметы по выбору в течение какого-либо конкретного учебного года в колледже/университете, предполагается, что существуют определенные приоритеты изучения (по крайней мере) некоторых из этих курсов. Такое положение вещей представляется естественным, так как для каждого учебного предмета (охватываемого одним курсом по выбору или даже более чем одним курсом по выбору) существует перечень курсов, которые подлежат изучению (или, по крайней мере, рекомендованы администрацией колледжа/университета для изучения (рассмотрения) перед изучением этого курса. (В рамках обязательных курсов также существуют условия первоочередности изучения курсов в рамках определенного круга изучаемых предметов.)

Как в обязательных курсах, так и в курсах по выбору, каждый конкретный курс, который предстоит пройти студенту в рамках изучения того или иного предмета, разбит на конечное число тем (разделов). Изучение каждой из этих тем конкретным студентом предпола-

гает а) аудиторную деятельность, включающую выполнение коллективных и индивидуальных заданий, б) выполнение только индивидуальных заданий (домашних работ), в) как успешное выполнение этим студентом промежуточных контрольных работ, так и сдачу этим студентом итогового экзамена. Выполнение перечисленных заданий предполагает решение конкретных задач, каждая из которых имеет один из трех уровней сложности, а именно: а) обязательный (первый) уровень, б) продвинутый (второй) уровень, в) повышенной сложности (третий) уровень.

В рамках каждого предмета, изучение которых требуется для получения конкретной специализации (и степени бакалавра), администрация колледжа/университета предлагает несколько обязательных (основных) курсов и дополнительных курсов (которые, обычно, можно выбрать из числа курсов по выбору). Для каждого из курсов, помимо лекций по предмету, которые обычно являются частью курса, лектор курса может рекомендовать студенту (или студентам), которые должны (для обязательных курсов) или планируют изучать этот курс (для курсов по выбору), рассмотреть возможность изучения одного или даже нескольких дополнительных курсов. Изучение каждого дополнительного курса (по мнению лектора курса) поможет студенту глубже освоить определенную тему (или даже несколько конкретных тем) из основного курса. Однако студент может принять решение ограничить свое знакомство с рекомендуемыми дополнительными курсами, выполнив лишь несколько заданий из этих курсов, без глубокого изучения этих дополнительных курсов. Так, в частности, может произойти, если студент считает, что выполнения только выбранных заданий достаточно для повышения шансов на успешную сдачу экзамена по изучаемому курсу (обязательному курсу или курсу по выбору). Следует отметить, что при вычислении количественных оценок, рассматриваемых в рамках настоящего исследования, естественно предполагать, что эти шансы могут быть оценены (количественно) для каждого такого задания (причем, для каждого уровня сложности этого задания).

Предварительно выбрав:

− предмет для изучения из набора предметов, предлагаемых колледжем/университетом студенту для изучения в рамках выбранной этим студентом области специализации;

− базовый курс по выбору (или базовые курсы по выбору) для изучения предмета;

− (если возможно) лектора (или лекторов) для базового курса по выбору (или для базовых курсов по выбору),

студент может решить пройти тест (или тесты и контрольные опросы), чтобы определить свою способность успешно изучать как выбранный курс по выбору (курсы по выбору),

так и выбранный предмет. Каждый такой тест, как и каждый такой контрольный опрос, в частности, поможет определить:

− среднее время, необходимое студенту для выполнения заданий по каждому разделу предмета (и курса по выбору);

− набор дополнительных курсов к курсам по выбору для изучения (или, по крайней мере, для просмотра) студентом из числа дополнительных курсов, рекомендуемых студентам для изучения преподавателями, преподающими (выбранные студентом) курсы по выбору.

Результаты каждого пройденного теста (каждого из пройденных тестов и контрольных опросов) могут помочь студенту определить ожидаемую оценку, которую этот студент может получить за каждый изучаемый курс (в рамках изучения предмета) и накопить за изучение всего предмета (если изучение всего предмета включает в себя изучение более одного курса), которую студент будет стараться максимизировать.

На основании информации (о времени, которое предполагается потратить на изучение конкретных курсов, и об ожидаемых оценках, которые могут получены в результате изучения этих курсов), студент может эффективнее выбрать подмножество курсов из набора курсов по выбору для изучения в колледже/университете (из всего набора курсов по выбору, предлагаемых администрацией колледжа/университета в рамках своей специализации). Также, студент может лучше спланировать последовательность график изучения всех отобранных курсов по выбору в рамках каждого учебного года в колледже/университете.

Решение администрацией колледжа/университета следующих задач может помочь студенту в достижении сформулированных выше целей.

Задача 1. Как найти оптимальный для конкретного студента набор (оптимальные наборы) курсов по выбору для изучения (из числа всех курсов по выбору, предлагаемых администрацией колледжа/университета, в котором студент учится, в рамках его специализации) и оптимальную последовательность изучения курсов по выбору в рамках оптимального набора этих курсов:

а) в текущем учебном году;

б) в течение всех четырех лет обучения в колледже/университете (вместе взятых), который максимизирует ожидаемое значение GPA, которое этот студент может достичь, “заработав” (по крайней мере) наименьшее приемлемое количество кредитов (для выбранных курсов по выбору)?

Задача 2. Предположим, что за определенный год обучения в колледже/университете конкретный студент:

а) не проходит тест (или набор тестов), или

б) сдает тест (или набор тестов) со слишком низкими оценками, или

в) чувствует, что определенную тему из любого курса, пройденного в рамках изучения определенного предмета, не удалось изучить на приемлемом уровне.

Принимая во внимание:

− текущий академический статус студента;

− курсы, выбранные студентом (из всего набора курсов по выбору, предлагаемых администрацией колледжа/университета для изучения в рамках специализации студента);

− общее количество кредитов, которые уже были “заработаны”;

− время, которое уже было потрачено на учебу в колледже/университете;

− время, оставшееся до (планируемого) выпуска студента,

какова наилучшая стратегия продолжения образовательного процесса, чтобы максимизировать GPA, который студент может получить в течение а) текущего учебного года, и б) оставшейся части четырех учебных лет обучения в колледже/университете, предполагая, что студент наберет хотя бы минимально допустимое количество кредитов, необходимое для получения желаемой степени бакалавра?

С учетом максимального значения GPA, которое студент может достичь (определяемого в результате решения Задачи 1), для администрации колледжа/университета представляет интерес решение следующей задачи.

Задача 3. Каковы минимальные затраты колледжа/университета на организацию и проведение набора курсов по выбору, достаточного для достижения студентами значения GPA не ниже некоторого, приемлемого для администрации этого колледжа/университета, для каждой специализации, которую предлагает колледж/университет?

2.3.2.2 Математическая формулировка задачи нахождения администрацией оптимального учебного плана для студента

Предложенная в этом разделе отчета математическая модель позволяет формализовать Задачу 1 для администрации колледжа/университета и разработать прототип упомянутой выше системы поддержки принятия решений, использование которой может позволить этой администрации рекомендовать своим студентам оптимальную (с точки зрения администрации) учебную программу.

Для студента (потенциального, который еще только планирует учиться в колледже/университете, или для того, который в настоящее время уже учится там с целью получения степени бакалавра в той или иной области), пусть:

i – номер курса в наборе обязательных курсов, которые студент должен изучать $i \in \overline{1,n}$;

$n+j$ – номер курса по выбору в наборе курсов по выбору, которые интересны студенту и из которых им выбираются курсы по выбору для изучения, $j \in \overline{1,r}$;

Q_i – количество тем (разделов), составляющих обязательный курс $i, i \in \overline{1,n}$;

q_i – номер темы (раздела) в рамках курса i из набора n обязательных курсов, $q_i \in \overline{1,Q_i}$, $i \in \overline{1,n}$;

Q_{n+j} – общее количество тем (разделов), составляющих курс по выбору $n+j$ из набора r курсов по выбору для изучения, $j \in \overline{1,r}$;

q_{n+j} – номер темы (раздела) в в рамках курса по выбору $n+j$ из набора r курсов по выбору, $q_{n+j} \in \overline{1,Q_{n+j}}$, $j \in \overline{1,r}$;

M_{iq_i} – общее количество заданий по теме $q_i \in \overline{1,Q_i}$ из обязательного курса $i, i \in \overline{1,n}$;

m_{iq_i} – номер задания в рамках темы q_i из обязательного курса $q_i \in \overline{1,Q_i}$, $m_{iq_i} \in \overline{1,M_{iq_i}}$, $i \in \overline{1,n}$;

$M_{(n+j)q_{n+j}}$ – общее количество заданий по теме $q_{n+j} \in \overline{1,Q_{n+j}}$ из курса по выбору $n+j$ в наборе r курсов по выбору, $j \in \overline{1,r}$;

$m_{(n+j)q_{n+j}}$ – номер задания в рамках темы q_{n+j} из курса по выбору $n+j$ в наборе r курсов по выбору, $q_{n+j} \in \overline{1,Q_{n+j}}$, $m_{(n+j)q_{n+j}} \in \overline{1,M_{(n+j)q_{n+j}}}$, $j \in \overline{1,r}$;

l – уровень сложности заданий по теме основных курсов или курсов по выбору, $l \in \overline{1,K}$;

$t^l_{m_{iq_i}}$ – время, требуемое для выполнения задания m_{iq_i} уровня сложности l из обязательного курса, i, $q_i \in \overline{1,Q_i}, i \in \overline{1,n}$, $m_{iq_i} \in \overline{1,M_{iq_i}}$, $l \in \overline{1,K}$;

$t^l_{m_{(n+j)q_{n+j}}}$ – время, требуемое для выполнения задания $m_{(n+j)q_{n+j}}$ уровня сложности l, $q_{n+j} \in \overline{1,Q_{n+j}}$ из курса по выбору $n+j$ в наборе r курсов по выбору, $j \in \overline{1,r}$, $m_{(n+j)q_{n+j}} \in \overline{1,M_{(n+j)q_{n+j}}}$, $l \in \overline{1,K}$;

γ_i – количество кредитов, которое получает студент за успешное завершение обязательного курса i, $i \in \overline{1,n}$;

γ_{n+j} – количество кредитов, которое получает студент за успешное завершение курса по выбору $n+j$ из набора r курсов по выбору, $j \in \overline{1,r}$;

β_i – число кредитов, отводимых администрацией колледжа/университета на обязательный курс i, которое получает студент, успешно прошедший этот курс, $i \in \overline{1,n}$;

β_{n+j} – число кредитов, отводимых администрацией колледжа/университета на курс по выбору $n+j$, которое получает студент, успешно прошедший этот курс, $j \in \overline{1,r}$;

p_i – количество заданий в обязательном курсе i, $i \in \overline{1,n}$, подлежащих обязательному выполнению;

p_{n+j} – количество заданий в курсе по выбору $n+j$, подлежащих обязательному выполнению, $j \in \overline{1,r}$;

T – время, которое студент может потратить на выполнение заданий по всем изучаемым курсам (обязательным курсам и курсам по выбору), включаемым в учебную программу студента, в период обучения в колледже/университете (в дополнение ко времени, которое будет потрачено студентом на посещение лекций и участие в аудиторных занятиях, проводимых профессорами и преподавателями, преподающими курсы, включенные в учебную программу студента);

w – минимальное общее количество кредитов, которое необходимо набрать студенту, чтобы получить высшее образование в колледже/университете с желаемой степенью (с желаемым дипломом);

J_{n+j} – количество дополнительных курсов, рекомендованных студенту для изучения (преподавателем курса по выбору $n+j$) из тех дополнительных курсов, которые предлагаются в колледже/университете (до изучения курса по выбору $n+j$ из r курсов по выбору), $j \in \overline{1,r}$;

$\tau_{(n+j)\xi}$ – время, требуемое для изучения дополнительного курса $\xi \in \overline{1,J_{n+j}}$, рекомендованного (преподавателем курса по выбору $n+j$) студенту для изучения из дополнитель-

ных курсов, предлагаемых в колледже/университете (до изучения курса по выбору $n+j$ из r курсов по выбору), $j \in \overline{1,r}$.

Пусть: $x_{n+j} = \begin{cases} 1, \text{ если курс по выбору } n+j \text{ из } r \text{ курсов по выбору выбран студентом} \\ \text{для изучения}, j \in \overline{1,r}, \\ 0, \text{ в противном случае} \end{cases}$

$$y^l_{m_{iq_i}} = \begin{cases} 1, \text{ если задание } m_{iq_i} \in \overline{1, M_{iq_i}} \text{ уровня сложности } l \\ \text{является частью обязательного курса } i, i \in \overline{1,n}, \ q_i \in \overline{1,Q_i}, l \in \overline{1.K}, \\ 0, \text{ в противном случае,} \end{cases}$$

$$y^l_{m_{(n+j)q_{n+j}}} = \begin{cases} 1, \text{ если задание } m_{(n+j)q_{n+j}} \in \overline{1, M_{(n+j)q_{n+j}}} \text{ уровня сложности } l \\ \text{является частью курса по выбору } n+j \text{ из } r \text{ курсов по выбору} \\ , j \in \overline{1,r}, \ q_{n+j} \in \overline{1,Q_{n+j}}, l \in \overline{1,K}, \\ 0, \text{ в противном случае,} \end{cases} \quad (2.19)$$

$$z_{(n+j)\xi} = \begin{cases} 1, \quad \text{курс } \xi \text{ из } \overline{1,J_{n+j}} \text{ курсов рекомендуется студенту для изучения} \\ \quad (\text{преподавателем, ведущим курс по выбору } n+j \text{ из} \\ \quad r \text{ курсов по выбору) до изучения курса по выбору,} \\ n+j \text{ , выбранного студентом для изучения}, j \in \overline{1,r}, \\ 0, \quad \text{в противном случае.} \end{cases}$$

Далее считается, что выполняются следующие три предположения.

Предположение 1. Суммарная оценка за курс рассчитывается по формуле

$$O_{cummulative} = 0.25\, O_{inaudit} + 0.25\, O_{hw} + 0.5\, O_{exam}. \quad (2.20)$$

Здесь $O_{inaudit}$ − оценка за занятия в аудитории, O_{hw} − оценка за домашние задания и O_{exam} − оценка за экзамен.

Предположение 2. Все оценки $O_{inaudit}$, O_{hw} и O_{exam} являются независимыми дискретными случайными величинами с известными функциями распределения, и каждая из этих функций может быть определена путем статистической обработки имеющихся данных.

Предположение 3. Оценки за выполнение заданий разного уровня сложности являются независимыми дискретными случайными величинами с известными функциями распределения, и каждая из этих функций может быть определена путем статистической обработки имеющихся данных.

При этих предположениях математическая модель выбора оптимальной схемы изучения всех обязательных курсов и всех курсов по выбору в колледже/университете, которую администрация колледжа/университета считает оптимальной для студента, может быть представлена в виде следующей системы линейных ограничений с булевыми переменными:

$$y(l)_{m_{(n+j)q_{n+j}}}^{(n+j)q_{n+j}} \le x_{n+j},\ m_{(n+j)q_{n+j}} \in \overline{1, M_{(n+j)q_{n+j}}},\ q_{n+j} \in \overline{1, Q_{n+j}}\ \ j \in \overline{1, r},\ l \in \overline{1, K}, \quad (2.21)$$

(если курс по выбору $n+j$ из r курсов по выбору не выбран студентом для изучения, то никакое задание из любого раздела любого уровня сложности этого задания по этому курсу по выбору не может быть частью учебной программы студента),

$$\sum_{l=1}^{K} \sum_{q_{n+j}=1}^{Q_{n+j}} \sum_{m_{(n+j)q_{n+j}}=1}^{M_{(n+j)q_{n+j}}} y(l)_{m_{(n+j)q_{n+j}}}^{(n+j)q_{n+j}} \ge x_{n+j}, \quad j \in \overline{1, r}, \quad (2.22)$$

(если курс по выбору $n+j$ из r курсов по выбору выбран студентом для изучения, то хотя бы одно из заданий какого-либо уровня сложности из числа заданий, связанных с этим курсом по выбору, должно быть включено в учебный план студента),

$$y(l)_{m_{iq_i}}^{iq_i} \le y(k)_{m_{iq_i}}^{iq_i}, i \in \overline{1, n},\ \ q_i \in \overline{1, Q_i},\ m_{iq_i} \in \overline{1, M_{iq_i}},\ \ \forall l \ge k,\ l, k \in \overline{1, K}, \quad (2.23)$$

(если задание m_{iq_i} уровня сложности k, связанное с разделом обязательного курса q_i, не входит в учебную программу студента, то задание с таким же номером из того же раздела того же самого обязательного курса любой более высокой сложности l не может быть частью этого учебного плана),

$$y(l)_{m_{(n+j)q_{n+j}}}^{(n+j)q_{n+j}} \le y(k)_{m_{(n+j)q_{n+j}}}^{(n+j)q_{n+j}}, \quad (2.24)$$

$j \in \overline{1, r},\ \ q_{n+j} \in \overline{1, Q_{n+j}},\ \ m_{(n+j)q_{n+j}} \in \overline{1, M_{(n+j)q_{n+j}}},\ \ \forall l \ge k \quad l, k \in \overline{1, K},$ (если задание $m_{(n+j)q_{n+j}}$ уровня сложности k, связанное с разделом курса по выбору q_{n+j} из r курсов по выбору, не входит в учебную программу студента, то задание с таким же номером из того же раздела

того же самого обязательного курса любой более высокой сложности l не может быть частью этого учебного плана),

$$\sum_{l=1}^{K}\sum_{q_i=1}^{Q_i}\sum_{m_{iq_i}=1}^{M_{iq_i}} y(l)_{m_{iq_i}}^{iq_i} \geq p_i, i \in \overline{1,n}, \tag{2.25}$$

(общее количество заданий, связанных с обязательным курсом i, включенным в учебную программу студента, не может быть меньше обязательного их количества, которое требуется выполнить в рамках изучения этого обязательного курса)

$$\sum_{l=1}^{K}\sum_{q_{n+j}=1}^{Q_{n+j}}\sum_{m_{(n+j)q_{n+j}}=1}^{M_{(n+j)q_{n+j}}} y(l)_{m_{(n+j)q_{n+j}}}^{(n+j)q_{n+j}} \geq p_{n+j}, \ j \in \overline{1,r}, \tag{2.26}$$

(общее количество заданий, связанных с курсом по выбору $n+j$ из набора r курсов по выбору, включенному в учебную программу студента, не может быть меньше обязательного их количества, которое требуется выполнить в рамках изучения этого курса по выбору),

$$\sum_{l=1}^{K}\sum_{m_{iq_i}=1}^{M_{iq_i}} y(l)_{m_{iq_i}}^{iq_i} \geq y(l)_{m_{i(q_i+\omega)}}^{i(q_i+\omega)}, \ i \in \overline{1,30}, \ l \in \overline{1,K}, \ q_i+\omega \in \overline{q_i+1,Q_i}, \ \omega \geq 1, \tag{2.27}$$

(никакое задание любой сложности по какому-либо разделу $q_i+\omega$ обязательного курса i, не включается в учебный план студента, если в этот учебный план не входит задание по разделу q_i, изучение которого предшествует изучению раздела $q_i+\omega$ этого обязательного курса),

$$\sum_{l=1}^{K}\sum_{m_{(n+j)q_{n+j}}=1}^{M_{(n+j)q_{n+j}}} y(l)_{m_{(n+j)q_{n+j}}}^{(n+j)q_{n+j}} \geq y(l)_{m_{(n=j)(q_{n+j}+\omega)}}^{(n=j)(q_{n+j}+\omega)}, q_{n+j}+\omega \in \overline{q_{n+j}+1,Q_{n+j}}, \omega \geq 1, l \in \overline{1,K}, \tag{2.28}$$

(никакое задание любой сложности по какому-либо разделу $q_{n+j}+\omega$ курса по выбору $n+j$ из набора r курсов по выбору не включается в учебный план студента, если в этот учебный план не входит задание по разделу q_{n+j}, изучение которого предшествует изучению раздела $q_{n+j}+\omega$ этого курса по выбору),

$$\sum_{i=1}^{n}\gamma_i+\sum_{j=1}^{r}x_{n+j}\gamma_{n+j}\geq w, \tag{2.29}$$

(общее количество кредитов, ожидаемое студентом к получению в результате изучения n основных курсов и курсов по выбору из набора r курсов по выбору, не может быть меньше, чем минимальное общее количество кредитов, которое требуется от студента получить для успешного окончания колледжа/университета с желаемой степенью бакалавра),

$$z_{(n+j)\xi}\leq x_{n+j},\ \xi\in\overline{1,J_{n+j}},\ j\in\overline{1,r}, \tag{2.30}$$

(если курс по выбору $n+j$ из набора r курсов по выбору не включен в учебную программу студента, то никакие дополнительные курсы, рекомендуемые преподавателем этого курса по выбору, не изучаются студентом в дополнение к курсу по выбору $n+j$ и не могут рассматриваться как дополнительный курс по выбору в учебной программе студента в принципе),

$$\sum_{l=1}^{K}\sum_{i=1}^{n}\sum_{q_i=1}^{Q_i}\sum_{m_{iq_i}=1}^{M_{iq_i}}t(l)_{m_{iq_i}}^{iq_i}\,y(l)_{m_{iq_i}}^{iq_i}+\sum_{l=1}^{K}\sum_{j=1}^{r}\sum_{q_{n+j}=1}^{Q_{n+j}}\sum_{m_{(n+j)q_{n+j}}=1}^{M_{(n+j)q_{n+j}}}t(l)_{m_{(n+j)q_{n+j}}}^{(n+j)q_{n+j}}\,y(l)_{m_{(n+j)q_{n+j}}}^{(n+j)q_{n+j}}+$$

$$+\sum_{j=1}^{r}\sum_{\xi=1}^{J_{n+j}}\tau_{(n+j)\zeta}z_{(n+j)\zeta}\leq T, \tag{2.31}$$

(время, которое студент может потратить на выполнение заданий по всем изучаемым курсам, включенным в учебную программу студента в период обучения — дополнительно ко времени, затрачиваемому на посещение лекций, на участие во всех аудиторных мероприятиях, проводимых преподавателями (преподающими курсы, включенные в учебную программу студента), и для изучения дополнительных курсов, рекомендованных преподавателями, преподающими курсы по выбору, для изучения этих дополнительных курсов, рекомендованных для изучения перед изучением курсов по выбору — не может превышать T).

Пусть далее:

Δ_i − математическое ожидание оценки успешности изучения (grade) студентом обязательного курса i, $i\in\overline{1,n}$;

Δ_{n+j} − математическое ожидание оценки успешности изучения (grade) студентом курса по выбору $n+j$ из набора r курсов по выбору.

Методика вычисления значений Δ_i и Δ_{n+j} на основе естественных теоретико-вероятностных предположений приведена в [130].

Тогда, целевая функция рассматриваемой задачи может быть записана, например, в виде

$$\frac{\sum_{i=1}^{n}\Delta_i\beta_i+\sum_{j=1}^{r}x_{n+j}\Delta_{n+j}\beta_{n+j}}{\sum_{i=1}^{n}\beta_i+\sum_{j=1}^{r}x_{n+j}\beta_{n+j}}, \tag{2.32}$$

и эта функция булевых переменных $x_{n+j},\ j\in\overline{1,r}$ подлежит максимизации при выполнении системы ограничений (2.21) − (2.31) на переменные

$$x_i,\ x_{n+j},\ y^l_{m_{iq_i}},\ y^l_{m_{(n+j)q_{n+j}}},\ z_{(n+j)\xi},\ i\in\overline{1,n},\ \ j\in\overline{1,r},\ l\in\overline{1,K},\ \xi\in\overline{1,J_{n+j}}. \tag{2.33}$$

Так, проблему выбора оптимального (с точки зрения только администрации колледжа/университета) набора курсов по выбору (из набора *r* курсов по выбору) — рекомендуемого для изучения конкретному студенту этого колледжа/университета в рамках его специализации, наряду с изучением всех обязательных предметов, для достижения максимального ожидаемого GPA за время обучения—удается формализовать, например, в виде задачи дробного программирования с булевыми переменными (2.21) − (2.32). Методы решения задач булева программирования такого типа можно найти, например, в [131, 132] а пакеты программ для решения задач булева программирования, можно найти, в частности, в [133].

Заметим, что в задаче (2.20) − (2.32) можно рассматривать любой период времени (например, семестр, учебный год, количество учебных лет в рамках бакалавриата в колледже/университете). Для этого следует правильно перенумеровать все курсы, разделы внутри этих курсов, задания внутри этих разделов и т. д. Однако, при проведении конкретных вычислений может оказаться удобнее ввести дополнительные индексы к переменным модели, например индекс, учебного года для каждой переменной.

Как упоминалось ранее во Введении, хотя задача (2.20) − (2.32) решается для оценки влияния выбора набора курсов по выбору на ожидаемое значение GPA по результатам обу-

чения в бакалавриате колледжа/университета, ее можно легко переформулировать для конкретного учебного года в процессе этого четырехлетнего обучения, а также для любого двухлетнего обучения в магистратуре любого университета.

2.3.2.4 Математическая формулировка задачи поиска оптимального учебного плана для студента колледжа/ университета, решаемая непосредственно студентом этого колледжа/университета

Рассмотрим теперь задачу формирования подмножества курсов по выбору из множества r курсов по выбору (интересных для студента, в частности, в смысле размера кредита, который ассоциирован с каждым из этих курсов) — предлагаемых колледжем/университетом своим студентам для изучения в течение учебного года вместе со всеми обязательными курсами - которое (подмножество курсов) может выглядеть оптимальным с точки зрения конкретного студента.

В дополнение к обозначениям, введенным при формулировке *Задачи 1*, пусть:

θ_i − время, которое студент рассчитывает потратить на изучение обязательного курса i, $i \in \overline{1,n}$;

θ_{n+j} − время, которое студент рассчитывает потратить на изучение курса по выбору $n+j$ из набора r курсов по выбору, $j \in \overline{1,r}$;

h_{n+j} − оценка, которую студент предполагает получить в результате успешного прохождения как всех промежуточных тестов курса по выбору $n+j$ из набора r курсов по выбору, так и итогового экзамена по курсу по выбору $n+j$, $j \in \overline{1,r}$ и пусть выполняется система ограничений

$$\sum_{j=1}^{r} \theta_{n+j} x_{n+j} + \sum_{\xi=1}^{J_{n+j}} \tau_{(n+j)\xi} z_{(n+j)\xi} \le T - \sum_{i=1}^{n} \theta_i, \tag{2.34}$$

$$z_{(n+j)\xi} \le x_{n+j}, \qquad \xi \in \overline{1, J_{n+j}}, j \in \overline{1,r}. \tag{2.35}$$

Рассмотрим задачу максимизации функции

$$\sum_{j=1}^{r} h_{n+j} \beta_{n+j} x_{n+j} \to \max \tag{2.36}$$

при ограничениях (2.34) − (2.35).

Эта задача является задачей булева программирования специальной структуры (задачей о рюкзаке с условиями предшествования) [124, 125], и пакеты программ, реализующие алгоритмы решения такого типа задач, можно найти в [134].

2.3.2.5 Математическая формулировка задачи оптимизации финансовых затрат колледжа/университета на организацию и проведение курсов по выбору

Решая задачу (2.20) − (2.32) для каждой специализации бакалавра каждого из факультетов колледжа/университета, администрация колледжа/университета получает наилучшую, в принципе возможную оценку GPA студента, обучающегося на этой специализации, независимо от того каких (в частности, финансовых) затрат может потребовать достижение этой оценки от колледжа/университета. В то же время, ясно, что бюджет любого колледжа/университета ограничен (даже с учетом возможных государственных и частных субсидий). Поэтому, естественной является постановка задачи об отыскании всего набора курсов по выбору по всем специализациям (по которым эти курсы по выбору предлагаются на всех факультетах колледжа/университета) — в предположении, что GPA по каждой специализации должен быть не ниже некоторого значения, приемлемого для каждой специализации бакалавра колледжа/университета (которое, однако, может быть ниже наилучшего значения, определенного для специализации, например, из решения Задачи 1) — требующего минимальных финансовых затрат на их организацию и проведение.

Пусть далее:

N − число специализаций, по которым колледж/университет присуждает степень бакалавра;

B^{χ} − система ограничений вида (2.21) − (2.31) для специализации $\chi \in \overline{1,N}$, по которой колледж/университет присуждает степень бакалавра;

$c_{n^{\chi}+j^{\chi}}$ − стоимость организации и проведения курса по выбору $n^{\chi}+j^{\chi}$ из набора курсов по выбору r^{χ}, предлагаемых в рамках специализации $\chi \in \overline{1,N}$ в колледже/университете;

$\varepsilon_{n^{\chi}+j^{\chi}}$ − приемлемое для администрации колледжа/университета значение GPA студента, обучающегося по специализации $\chi \in \overline{1,N}$.

Тогда, задача оптимизации финансовых затрат колледжа/университета на организацию и проведение курсов по выбору, достаточных для достижения студентами по каждой специализации значения GPA не ниже некоторого, приемлемого для администрации этого

колледжа/университета ($\varepsilon_{n^\chi+j^\chi}$), может быть сформулирована как задача билинейного программирования со смешанными переменными $B^\chi, \chi \in \overline{1,N}$,

$$\frac{\sum_{i=1}^{n^\chi} \Delta_{i^\chi} \beta_{i^\chi} + \sum_{j=1}^{r^\chi} x_{n^\chi+j^\chi} \Delta_{n^\chi+j^\chi} \beta_{n^\chi+j^\chi}}{\sum_{i^\chi=1}^{n^\chi} \beta_{i^\chi} + \sum_{j^\chi=1}^{r^\chi} x_{n^\chi+j^\chi} \beta_{n^\chi+j^\chi}} \geq \varepsilon_{n^\chi+j^\chi}, \quad \chi \in \overline{1,N}, \tag{2.37}$$

$$\sum_{\chi=1}^{N} \sum_{j^\chi=1}^{n^\chi} c_{n^\chi+j^\chi} x_{n^\chi+j^\chi} \to \min, \tag{2.38}$$

методы решения которых приведены, в частности, в [135].

Здесь B^χ, $\chi \in \overline{1,N}$ − условное обозначение совокупности N систем линейных ограничений (в скалярной форме) вида (2.21) − (2.31) на переменные

$$x_{i^\chi}\ x_{n^\chi+j^\chi}, y^l_{i^\chi q_{i^\chi}}, y^l_{m_{(n^\chi+j^\chi)q_{n^\chi+j^\chi}}},\ z_{(n^\chi+j^\chi)\xi n^\chi+j^\chi}\ i^\chi \in \overline{1,n^\chi},\ j^\chi \in \overline{1,r^\chi},, l \in \overline{1,K}, \xi_{n^\chi+j^\chi} \in \overline{1,J_{n^\chi+j^\chi}} \tag{2.39}$$

где $\chi \in \overline{1,N}$, каждое значение параметра $\varepsilon_{n^\chi+j^\chi}$ — не превосходящее значения целевой функции в задаче (2.21) − (2.31), решаемой для специализации $\chi \in \overline{1,N}$ — выбирается администрацией колледжа/университета, переменные $x_{n^\chi+j^\chi}$ являются таковыми в системе ограничений задачи (2.21) − (2.32), формулируемой для специализации $\chi \in \overline{1,N}$, а числа $\Delta_{i^\chi}, \beta_{i^\chi}$, $\Delta_{n^\chi+j^\chi}, \beta_{n^\chi+j^\chi}$ являются параметрами в этой же системе ограничений, а переменные $c_{n^\chi+j^\chi}$ удовлетворяют какой-либо совокупности из $N+1$ систем линейных ограничений (одной системе для каждого значения $\chi \in \overline{1,N}$ и одной системе для всех N значений χ).

2.3.3 Некоторые выводы

1) Предметом настоящего исследования являлась лишь конкретная часть проблемы формирования учебного плана колледжем/университетом для своих студентов. Именно, рассматривались две постановки задачи выбора оптимального подмножества курсов из множества курсов по выбору (наряду с обязательными курсами) для изучения студентом колледжа/университета в рамках каждой специализации, по которой этому колледжу/университету дано право выпускать дипломированных специалистов. Первая задача отражает сложившее-

ся (в конкретном колледже/университете) понимание его администрацией того какое подмножество является оптимальным для обучающегося там конкретного студента или даже для какой-либо группы студентов в рамках каждой конкретной специализации, в любом конкретном периоде времени в течение всего четырехлетнего периода обучения. Вторая задача отражает понимание оптимальности студентом того же самого колледжа/университета, который разрабатывает свой учебный план на определенный период времени, комбинируя обязательные курсы с курсами по выбору, доступными для изучения также любом конкретном периоде времени в течение всего четырехлетнего периода обучения.

В обеих формулировках предполагалось, что все курсы по выбору по любой дисциплине, которая может быть интересна для изучения студенту, курсы по выбору предлагаются колледжем/университетом таким образом, чтобы отрезки времени проведения этих курсов не пересекались. Например, если предлагается h курсов по выбору, связанных с изучением дисциплины A, и, если каждый курс по выбору преподается в форме лекций, других аудиторных занятий и предполагает выполнение домашних занятий, которые предлагаются на еженедельной основе, все они предлагаются в разные временные отрезки суток, и эти временные отрезки согласованы. Как только становится известно число студентов, заинтересованных в изучении определенного набора курсов по выбору по каждой специализации, можно сформулировать и решить проблему поиска оптимального учебного плана для проведения этих курсов по выбору (которая аналогична простейшему из вариантов задачи о выборе оптимального портфеля заказом и очередности их выполнения [136]).

Здесь оптимальность может пониматься в каком-либо определенном смысле — например, в смысле общего количества пар лекция/аудиторные занятия, включая все курсы по выбору, отобранные студентом для изучения. Ее конкретное понимание может отражать предпочтения преподавателей, назначенных (и/или приглашенных) для преподавания этих курсов. (Однако, заметим, что проблема придания учебной программе студента какого-либо определенного смысла оптимальности как самим студентом, так и администрацией колледжа/университета, где этот студент обучается, не является предметом настоящего исследования.)

2) Следует иметь в виду, что некоторые студенты могут предпочесть полностью полагаться на советы администрации колледжа/университета по разработке своего оптимального учебного плана. Эта возможность еще раз подчеркивает, что рассматриваемая в настоящем исследовании проблема, формализованная в виде Задачи 1 и Задачи 2 — это даже в большей степени проблема администрации каждого конкретного колледжа/университета, чем про-

блема обучающихся в нем студентов, и должна рассматриваться как одна из проблем, связанных с управлением высшим образованием в стране.

3) Любые рекомендации администрации колледжа/университета своим студентам о том, какие курсы по выбору следует изучать конкретному студенту (или даже группе конкретных студентов), могут быть особенно ценны, когда этот студент (или некоторые или даже все студенты из этой группы) убеждается (убеждаются) в том, что она/он (они) не могут успешно пройти необходимые тесты и экзамены, чтобы получить ранее запланированные ими кредиты. Более того, это убеждение может возникнуть в любой момент в ходе изучения того или иного курса по выбору, после того как это изучение уже началось. В любом случае любые рекомендации о том, как действовать, должны быть, прежде всего, направлены на помощь в выяснении того, какая альтернатива имеет место, а именно:

а) возможна ли замена конкретного курса по выбору другим (или другими курсами по выбору), без уменьшения суммарного количества кредитов, которое, как ожидалось, получит студент (каждый из студентов из группы) для любого определенного временного сегмента ее/его (их) обучения в колледже/университете, а также какими такие замены могут быть, если они возможны в принципе;

б) следует ли исходить из того, что любые замены, вызывающие приемлемые для студента (студентов из группы) уменьшения этого общего количества кредитов и GPA, которое, как ожидалось, получит студент, невозможны.

Кроме того, целесообразно выяснить, могут ли

− дополнительные занятия с преподавателями колледжа/университета в какой-либо форме или

− дополнительные консультации лектора, ведущего конкретный курс по выбору, и/или преподавателя, ведущего аудиторные занятия в рамках этого курса по выбору (если это разные лица), или

− частные уроки, относящиеся к проблемам, изучаемым в курсе по выбору, изменить ситуацию, т. е. позволить студенту (студентам группы) успешно сдать курсовые зачеты и экзамен по курсу с приемлемыми оценками, независимо от того, что этот студент (эти студенты) предполагают относительно возможности успешного прохождения этого конкретного курса по выбору в момент появления у нее/него (них) вышеуказанного убеждения.

Перечисленные выводы можно сделать, решив задачу (2.20) − (2.32), в которой, однако, значения всех переменных

$$y^{l}_{m_{iq_i}},\ i, i \in \overline{1,n},\ \ q_i \in \overline{1,Q_i},\ m_{iq_i} \in \overline{1,M_{iq_i}}, l \in \overline{1.K}, \qquad (2.40)$$

а также значения некоторых из переменных

$$y^{l}_{m_{(n+j)q_{n+j}}}, j \in \overline{1,r},\ \ q_{n+j} \in \overline{1,Q_{n+j}},\ m_{(n+j)q_{n+j}} \in \overline{1,M_{(n+j)q_{n+j}}},\ l \in \overline{1.K}, \qquad (2.41)$$

принимаются равными тем значениям, которые они имели при предыдущих расчетах, при выполнении которых предполагалось, что студент (каждый из студентов группы) изучал обязательные курсы и курсы по выбору из соответствующих им наборов.

При этом администрация колледжа/университета может предложить конкретному студенту (конкретным студентам из группы) пройти дополнительные тесты, чтобы оценить ее/его (их) шансы на успешное изучение тех курсов по выбору, которые должны заменить один из этих курсов (те курсы), которые беспокоят студента (студентов группы), если это ее/его (их) решение о замене изучения того или иного курса по выбору (или курсов по выбору) принимается (принимаются) этим студентом (студентами из группы) в принципе.

4) Любые тесты, определяющие способность студента к успешному изучению курса по выбору в колледже/университете, должны проводиться в форме, близкой к той, в которой студентам там предлагаются для изучения обязательные курсы, причем способность успешно работать над заданиями и домашними заданиями должна быть проверена в первую очередь. С этой целью, студенту следует предложить:

− несколько упрощенных заданий, требующих от учащегося демонстрации способности к логическому анализу вообще и к анализу заданий, касающихся разделов, которые будут изучаться в рамках курса по выбору, в частности;

− упрощенное домашнее задание, требующее от студента продемонстрировать свое понимание сути того, что должно быть сделано, чтобы успешно выполнить это домашнее задание;

− несколько вопросов, ответы на которые должны продемонстрировать видение студентом подходов к организации своей работы над проблемами, которые должны быть решены в этом домашнем задании.

Количество таких упрощенных заданий должно варьироваться в зависимости от информации обо всех предыдущих занятиях студента в колледже/университете и, в определенных случаях, даже в зависимости от всех ее /его школьных оценок, которыми располагает администрация колледжа/университета. Все эти упрощенные задания должны быть разрабо-

таны теми, кто будет вести курс по выбору — лектором и преподавателем, назначенным вести аудиторные занятия, которые также должны разработать набор вопросов, и ответы на эти вопросы должны продемонстрировать способность студента (или не более чем ограниченную способность студента) решать основные вопросы общего характера, которые авторы курса по выбору считают важными для понимания материала курса. Структура вопросов должна быть аналогична таковым из тестов и экзамена, который студент должен пройти в конце курса по выбору, и вопросы не обязательно должны быть связаны только с темами, охватываемыми курсом по выбору.

5) При постановке задач, являющихся предметом настоящего исследования, предполагалось, что математические ожидания оценок успешного изучения каждого обязательного курса и каждого курса по выбору, в том числе успешного прохождения всех домашних заданий, всех курсовых заданий и итоговых экзаменов, относящиеся к этому курсу, могут быть рассчитаны на основе доступной статистической информации. Однако, при проведении конкретных вычислений, студентам следует иметь в виду, что эта информация, обычно, доступна только администрации их колледжа/университета, и эта администрация может не предоставить ее студентам по их просьбе.

Результаты проведенного исследования нашли отражение в [130].

2.4 Исследования по теории расписаний

Для всех известных (псевдо) и полиномиально разрешимых случаев NP-трудной в сильном смысле задачи минимизации максимального временного смещения удалось построить две матрицы, с помощью которых выражаются эти разрешимые случаи. Продолжается работа над гипотезой, что все полиномиально разрешимые случаи теории расписаний могут быть выражены системой линейных неравенств с конечным числом матриц.

Для задачи минимизации максимального временного смещения найдена мера неразрешимости, равная $1/\sqrt{3}$.

Для задач железнодорожного планирования для двух станций построены алгоритмы решения на основе метода динамического программирования для любых (коммутативных и ассоциативных) функций штрафа. Продолжаются работы для построения плана формирования поездов (ПФП) на базе Горьковской железной дороги.

Эти результаты нашли отражение в [137, 138, 139, 140, 141, 142].

3. Исследования по развитию методов кластерного анализа данных

В этом разделе отчета приведены результаты исследований по развитию методов кластерного анализа данных, которые осуществлялись в следующих направлениях:

− обобщенное описание коллекции текстов с использованием таксономии предметной области;

− кластерная декомпозиция тел свидетельств на основе оптимизации мер конфликта.

3.1 Обобщенное описание коллекции текстов с использованием таксономии предметной области

3.1.1 Введение

В данном подразделе отчета приведены результаты исследований одной из основных задач искусственного интеллекта — автоматизации анализа содержания текстов. Эта тематика привлекает серьезное внимание исследователей. Одним из популярных направлений исследований является так называемый «контент-анализ» [143], метод, разработанный в доцифровую эпоху, прежде всего как метод социологического исследования. В настоящее время он используется для анализа коллекций текстов в любой области. Используются методы многомерного анализа данных, прежде всего, так называемое тематическое моделирование [144]. Модель тематического моделирования применяется к текстовым документам, представленным в виде «мешков слов». Она представляет собой так называемую факторизацию вероятностей появления слов документов через вероятности воображаемых «скрытых», «латентных» тем и условные вероятности слов в «темах». В результате применения метода получаются списки слов, соответствующих отдельным темам, а смысл тем восстанавливается вручную из анализа общего значения слов, входящих в них. Типичное исследование по тематическому моделированию описано в работе [145]. Для научного направления «экологическая социология» авторы сформировали набор из 815 научных публикаций и проанализировали содержание аннотаций к этим публикациям методом тематического моделирования. Они определили, используя так называемый индекс перплексии, что наилучшее количество скрытых тем для данной выборки будет от 25 до 31, из которых они выбрали 25, потому что полученные темы лучше всего интерпретируются. Таковы, например, «Экологическая озабоченность» (окружающая среда, озабоченность, отношение, результат); «Ценности» (социальные ценности, исследования, концепции, отношения), «Политика и управление» (общество, менеджмент, группа, решение, план). Темы, формируемые тематическим моделированием, в отличие от классического контент-анализа, определяются эмпирически, а не теоретически, что создает определенные проблемы на пути теоретического осмысления результатов.

Мы развиваем другой подход, включающий теоретический компонент явно: мы основываемся на заранее разработанной таксономии предметной области, представляющей собой основное знание о ней. Таксономия – это древовидная иерархия основных понятий. Тот факт, что A является «родителем» B в таксономии означает, что отношение «B является частью или конкретным случаем A».

Листовые темы таксономии представляют категории для последующего анализа. Согласно нашему подходу, последовательно выполняются следующие три этапа:

− расчет матрицы релевантности листовых тем и текстов коллекции;

− получение нечетких кластеров листовых тем по структуре матриц релевантности и со-релевантности;

− оптимальный подъем каждого из полученных кластеров.

На первом этапе вычисляется степень представления листовых тем таксономии в текстах коллекции. Наша мера релевантности между текстом и ключевой фразой категории основана на использовании механизма так называемых аннотированных суффиксных деревьев (АСД). Формально говоря, АСД вычисляет среднее значение условной вероятности правильного предсказания букв ключевой фразы на основе знания фрагментов текста, с которым происходит сопоставление. Эта характеристика средней условной вероятности используется в качестве меры релевантности между ключевой фразой и текстом. Данный метод, представленный в работах [146, 147], показал свою эффективность в ряде экспериментов и проектов. Этот метод особенно хорошо работает в тех довольно распространенных ситуациях, когда рассматриваемые тексты включают темы таксономии не дословно, а в свободных выражениях. Формирование матрицы релевантности между текстами с одной стороны, и темами с другой, завершает первый этап нашего метода.

Второй этап — поиск нечетких кластеров листовых тем. Здесь можно применять разнообразные методы формирования нечетких кластеров. Мы применяем, в основном, хорошо себя зарекомендовавший аддитивно-спектральный метод FADDIS [148]. Чтобы сделать кластерную структуру матрицы более «резкой», каждый текст взвешивается обратным значением количества соответствующих ему тем. Кроме того, применяется псевдо-лапласова нормализация из спектрального кластерного анализа. Метод FADDIS извлекает кластеры последовательно, один за другим, так что всегда можно остановить процесс получения кластеров, и таким образом определить само количество кластеров. Этому способствуют характеристики вклада кластеров в разброс данных.

Третий этап — поиск обобщений полученных кластеров. В отличие от других подходов, где определение обобщения кластера является предметом воображения и изобретатель-

ности авторов, метод, предложенный в [149], делает это автоматически с использованием специального алгоритма, который находит оптимальный набор «головных тем» для данного нечеткого кластера листовых тем в таксономии. Набор «головных тем» (одной или нескольких) содержит концепты таксономии высокого ранга, которые включают (то есть «покрывают») темы из данного кластера, возможно, с ошибками двух типов: небольшим количеством «пробелов» и «выбросов». «Пробел» — это тема дерева таксономии, которая покрывается головной темой, но не принадлежит рассматриваемому исходному кластеру, а «выброс» — это тема из кластера, которая не охвачена головной темой. Во многих случаях такой оптимальный подъем переводит рассматриваемую структуру, кластер листовых тем, на некоторый, значительно более «грубый», уровень грануляции, что позволяет увидеть такие идеи и связи, которые не очевидны при более интуитивных подходах.

Эффективность этой гибридной методологии была продемонстрирована нами в нескольких публикациях. В [149] методология была применена для извлечения исследовательских тенденций из коллекции исследовательских работ около 18000 научных статей, опубликованных в журналах Springer по Науке о Данных за 20 лет. Мы использовали таксономию науки о данных (TDS), основанную на классификации вычислительных систем международной Ассоциации вычислительной техники (ACM-CCS 2012) [150].

К сожалению, наша методология имеет очевидное слабое место. В ней используется эвристический критерий наибольшей экономии, минимизирующий общее количество введенных элементов — головных тем, пробелов и выбросов, суммируемых с определенными весовыми коэффициентами, выбор числовых значений которых имеет довольно шаткие основания (в работе [149] мы использовали коэффициенты 1, 0.1, 0.9, для количеств введенных головных тем, пробелов и выбросов, соответственно). В данном году мы сосредоточили усилия на том, чтобы освободиться от этого недостатка путем перехода к критерию максимального правдоподобия, основанному не на эвристических весовых коэффициентах, а на натурально возникающих вероятностях возникновения и потери головных тем в отдельных вершинах таксономии. В дальнейшем тексте мы опишем соответствующие построения и результаты.

3.1.2 Задача оптимального подъема нечеткого тематического кластера в таксономии: модель и метод

Математически, таксономия — это корневое дерево, чьи узлы аннотированы темами таксономии. Мы будем рассматривать следующую задачу: дано нечеткое множество S листьев таксономии, необходимо найти узел h(S) более высокого ранга в таксономии, который покрывает множество S наиболее “тесным” образом. Такая задача “подъема” — возможная

математическая формализация человеческой способности делать обобщения, то есть, формировать “концептуальную” форму представленного явления, в данном случае, нечеткого множества листьев дерева.

Задача не так проста, как могло бы показаться.

Рассмотрим, для простоты, четкое множество S, представленное пятью черными прямоугольниками на фрагменте дерева на рисунке 3.1. Рисунок 3.2 иллюстрирует ситуацию, при которой это множество обобщено тем, что «поднято» в корень A данного фрагмента таксономии, что показано закрашиванием соответствующего ему прямоугольника, как и его потомков.

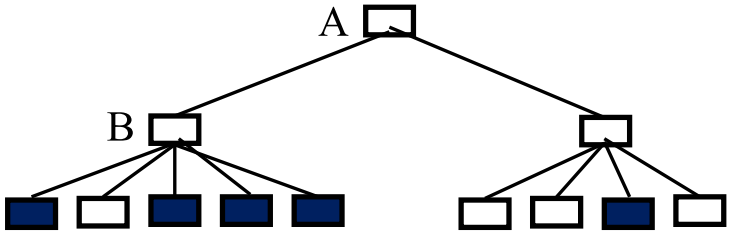


Рисунок 3.1 − Фрагмент дерева

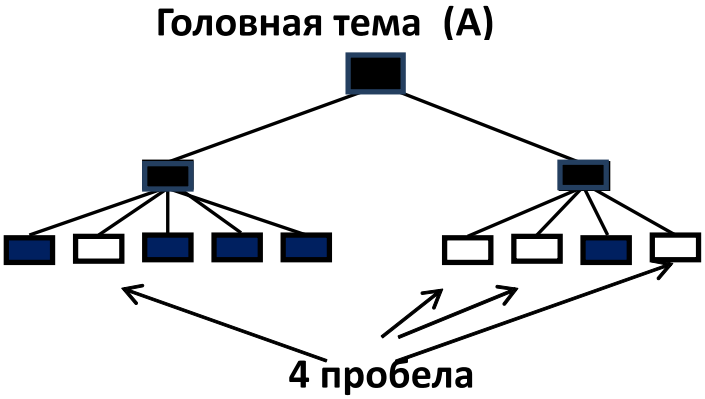


Рисунок 3.2 – Подъем кластера в корень A

Такое обобщение приведет к тому, что темой A будет покрыто и четыре белых прямоугольника, которые попадут в то же обобщение, что и S, при том, что они не принадлежат множеству S. Такую ситуацию мы называем появлением пробелов. Очевидно, что вариант подъема, при котором возникают пробелы, должен штрафоваться каким-либо образом.

В целом, число концептуальных элементов, введенных для обобщения S, здесь равно 1 головной вершине (то есть корню A), который мы теперь присвоили переменной S, и 4-м образовавшимся пробелам. Другой возможный вариант подъема показан на рисунке 3.3: здесь заданное множество поднимается только до корня левой ветви дерева, B.

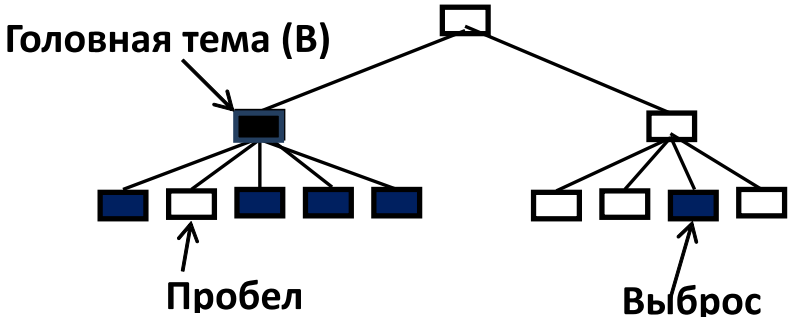


Рисунок 3.3 – Подъем кластера в корень B

Число пробелов на рисунке 3.3 резко сократилось — до 1. Однако возник еще один тип ошибки при подъеме: мы видим, что черный прямоугольник справа, принадлежащий S, не покрыт вершиной B, в которую отображается множество S. Этот тип ошибки будем называть выбросом. При варианте подъема рисунке 3.3 появляются три новых объекта: одна головная вершина, один выброс и один пробел.

Это меньше, чем количество объектов, появившихся при поднятии множества к корню (одна головная вершина и четыре пробела, то есть совокупно — пять), что делает его более предпочтительным. Конечно, этот вывод верен только в том случае, если относительный штраф за один выброс меньше суммарного штрафа за три пробела.

Нас интересует, можно ли обобщить нечеткое множество *S* в узел *t*, принадлежащий более высокому рангу таксономии, так, чтобы можно было считать, что *S* попадает в покрытие, охватываемое узлом *t*. Если это будет сделано напрямую, например, путем выбора последнего общего предка для узлов в *S* получение желаемого охвата будет маловероятным. Чтобы сделать процесс обобщения более адекватным, следует допустить наличие в итоговом обобщении не только пробелов, но и выбросов, являющихся листьями, принадлежащими обобщению, но не исходному множеству.

Цель нахождения интерпретируемого обобщения для S в рамках таксономии может быть формализована как нахождение одного или нескольких “головных тем” для покрытия S с минимальным количеством новых элементов всех типов, получившихся в процессе обобщения: главных 150 вершин, пробелов и выбросов. Такая постановка цели реализует принцип максимальной экономии (Maximum Parsimony, MP) при описании данного явления, популярный в некоторых областях, таких как биоинформатика. Другая постановка нацелена на максимизацию вероятности порождения данного нечеткого множества листьев, исходя из оптимального множества головных тем *H*.

Рассмотрим корневое дерево T, представляющее иерархическую таксономию, такую, что его узлы снабжены ключевыми фразами, обозначающими различные понятия. Обозначим множество его листьев через I. Соотношение между узлами в иерархии обычно выража-

ется с использованием генеалогических терминов: каждый узел $t \in T$ называется родителем узлов, непосредственно ответвляющихся от t, его детей. Будем использовать χ(t) для обозначения множества детей узла t. Предполагается, что каждый внутренний узел $t \in T - I$, то есть узел, не являющийся листом таксономии, соответствует понятию, обобщающему темы, соответствующие листьям I(t), выходящим из t, то есть, листья поддерева T(t) имеют общий корень в узле t, называющийся кластером листьев. Нечеткое множество на I — это отображение u из множества I в множество неотрицательных действительных чисел, которое присваивает значение функции принадлежности (или поддержку) $u(i) \geq 0$ каждому узлу $i \in I$. Мы называем множество, где $S_u \subset I$, где $S_u = \{i \in II: u(i) > 0\}$, поддержкой u. Никаких других предположений о виде функции u не делается. Отметим только то, что для удобства целесообразно наложить такое ограничение, чтобы ее значение не превышало единицу. Классические, или четкие множества соответствуют бинарным функциям принадлежности — таким, что u(i)=1, если $i \in S_u$ и u(i)=0 в противном случае. Рассматривая нечеткое множество запроса u, определенное на листьях I дерева T, можно считать u проекцией понятия более высокого ранга, “головной темы”, на соответствующий кластер листьев, возможно, сделанной с погрешностями. В этом предположении должен существовать головной узел h среди внутренних узлов дерева T — такой, что его кластер листьев I(h) в определенной степени совпадает (с небольшими ошибками) с S_u. Этот головной узел – то самое обобщение u, которое и необходимо найти. Возможны два типа ошибок, связанных с поиском головного узла, если он точно не покрывает поддержку, которые представляют собой ложные срабатывания и ложные отрицания, уже упомянутые ранее как пробелы (gaps) и выбросы (offshoots) — соответственно, иллюстрируются на рисунках 3.2 и 3.3. Общее количество головных узлов, пробелов и выбросов должно быть как можно меньше. Для этого мы вводим штраф за каждый из этих элементов. Предполагая, для простоты, что все черные прямоугольники на рисунке 1 имеют значения функции принадлежности, равные единице, можно легко увидеть, например, что общее значение штрафа, например, для подъема на рисунке 3.3 равно 1 + γ + λ , где γ, λ и 1 — штрафы за выброс, пробел и головную тему соответственно. Рассмотрим узел-кандидат h в T и его значение принадлежности в нечетком множестве u. Назовем h-пробелом такой узел g из T(h), отличный от h, в котором произошла потеря значения головного понятия, то есть g является максимальным u-иррелевантным узлом в том смысле, что его родитель не является u-иррелевантным. И наоборот, установление узла h в качестве головного можно рассматривать как приобретение значения/смысла u в узле. Множество всех h-выбросов обозначим как G(h). Чтобы правильно определить идею использования пробелов, введем понятие u-иррелевантных узлов. Узел t называется u-иррелевантным, если его кластер листьев I(t)

не пересекается с поддержкой u. Очевидно, что, если узел принадлежит u, все его потомки также принадлежат u. h-выброс — это лист $i \in S_u$, который не покрывается h, то есть $i \notin I(h)$. Множество всех h-выбросов — это $S_u - I(h)$. Если задано нечеткое тематическое множество u над I, множество узлов H будет называться u-покрытием, если: (а) H покрывает S_u, и (б) узлы в H не связаны в дереве T. Внутренние узлы H будут называться головными темами, а листовые узлы — выбросами, поэтому множество выбросов в H равно H ∩ I. Множество пробелов в H является объединением G(h) по всем головным темам $h \in H - I$.

Для формализации критерия максимального правдоподобия определим понятие сценария, ассоциированного с любым u-покрытием, то есть, отображением, приписывающим каждому узлу t ∈ T событие, связанное с общим рассматриваемым понятием (концепцией) — Обретение (Gain), или Потеря (Loss), или Пас (Pass) (событие, в котором ничего не происходит). При заданном сценарии все узлы таксономии можно разделить на два класса — те, в которых данное понятие присутствует, обозначим это множество N_P, и те, в которых оно отсутствует, N_A.

Тогда все дети узлов из N_P могут быть помечены как унаследовавшие данное общее понятие, а дети узлов из N_A — как не унаследовавшие данное общее понятие. Это приводит к возникновению четырех классов узлов, проиллюстрированных на рисунке 3.4:

(1) множество NGn таких узлов, которые не унаследовали понятие, но приобрели его согласно рассматриваемому сценарию;

(2) множество NƎn таких узлов, которые не унаследовали понятие, и не приобрели его согласно рассматриваемому сценарию;

(3) множество NLi таких узлов, которые унаследовали понятие, но потеряли его согласно рассматриваемому сценарию; и

(4) множество N⅂I таких узлов, которые унаследовали понятие и не потеряли его согласно рассматриваемому сценарию.

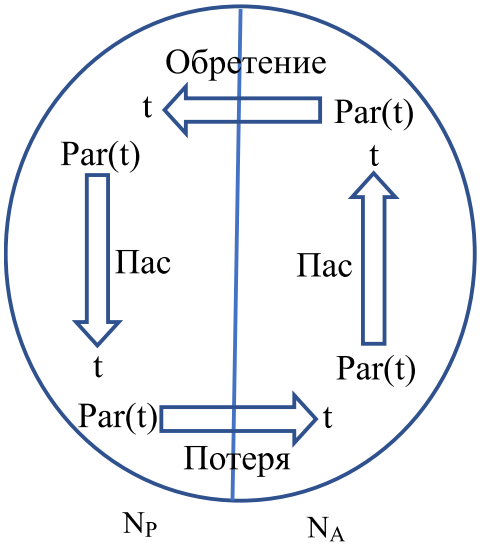


Рисунок 3.4 – Четыре класса узлов

В этих обозначениях множество H состоит из тех узлов в N_P, чьи родители находятся в N_A, то есть $H= N^G_n$.

Вероятность данного сценария во введенных обозначениях будет равна

$$P(H)= \prod_{t\in N_n^G} p_G(t) \prod_{t\in N_n^{\bar{O}}}(1-p_G(t)) \prod_{t\in N_i^L} p_L(t) \prod_{t\in N_i^{\bar{l}}}(1-p_L(t)). \quad (3.1)$$

В этом определении p_G и p_L — заранее заданные вероятности обретения и потери общего понятия в данном узле, соответственно; *t* – узел, *i* – индекс для унаследования, *n* — индекс для не унаследования, N с индексами — вышеопределенные множества узлов, *u(h)* — функция принадлежности нечеткого множества запроса.

Задача состоит в том, чтобы найти сценарий, максимизирующий функцию правдоподобия (3.1). В силу аддитивности логарифмической формы критерия (3.1), его максимизация может быть достигнута, практически тем же алгоритмом, что мы разработали для задачи отыскания максимально экономного множества головных тем в [149]. Алгоритм начинает работу с листьев дерева таксономии и рекурсивно поднимается с уровня на уровень. Вычисления заканчиваются при достижении корня дерева. Для каждого узла t данный алгоритм вычисляет два набора — H(t) и L(t), содержащие такие узлы в T(t), в которых, соответственно, происходят приобретения и потери головных узлов (включая выбросы). Соответствующая часть вероятности обозначается p(t) и вычисляется в соответствии с формулой (3.1). В каждом пройденном алгоритмом узле t выбирается то событие из трех — Пас, Обретение, Потеря — которое имеет максимальную вероятность. Предположение алгоритма состоит в том, что после потери узла не может быть последующего ее приобретения. Следовательно,

H(t) и L(t) вычисляются при условии, что головной узел не был ни приобретен (ни потерян) ни у одного из предков t. Как и в случае максимизации экономии, для каждого узла рассматривается два сценария: «головная тема возникла в t» и «головная тема возникла выше». Выбирается тот из них, вероятность которого выше.

Для определения вероятностей $p_G(t)$ и $p_L(t)$, вероятно, можно использовать такие общие депозитарии научных публикаций как Гугл-Академия. Однако, нам не удалось сформулировать критерии для определения соответствующих событий. Поэтому мы решили определять эти вероятности по статистике решений по критерию максимальной экономии при случайных значениях параметров. Например, мы брали 5000 случайных статей из рассматриваемой выборки и случайные значения из интервала [0.05, 0.3] для λ и интервала [0.6,0.9] для γ и применяли к ним метод максимальной экономии. После 300 применений этой процедуры мы получали достаточно устойчивые частоты для определения соответствующих вероятностей. Величина pL(t) определяется, тогда как количество случаев, когда t наследовало, но теряло понятие, отнесенное к общему числу случаев, когда у родителя t данное понятие имелось в наличии.

Аналогично, $p_G(t)$ определяется как количество случаев, когда t не наследовало понятие, но приобретало его, отнесенное к общему числу случаев, когда у родителя t данного понятия не имелось в наличии.

Применив разработанный алгоритм максимально правдоподобного обобщения к таксономии Науки о Данных из [151] и коллекции 17685 публикаций из 6, мы нашли 6 нечетких кластеров, головные темы которых, согласно данному алгоритму:

− «Information retrieval»,

− «Machine learning»,

− «Clustering»,

− «Modeling methods»,

− «Markov networks» and «network modeling»,

− «Reinforcement learning».

Первые три совпадают с теми, что получены в [147]. Данные шесть кластеров и их обобщения представляют Науку о данных в значительно более сбалансированном виде, чем это удалось сделать в [147]. Материалы этого раздела опубликованы в работах [152, 153].

3.2 Кластерная декомпозиция тел свидетельств на основе оптимизации мер конфликта

3.2.1 Введение

В прикладных задачах теории функций доверия тело свидетельств может иметь довольно сложную структуру: состоять из большого числа фокальных элементов со сложной структурой пересечения. Такие свидетельства трудно интерпретировать. Кроме того, поскольку многие операции теории функций доверия (например, правила комбинирования), являются вычислительно трудными, то применение этих операций к телам свидетельств с большим числом фокальных элементов становится также вычислительно трудной задачей.

Поэтому актуальными являются задачи:

1) анализ структуры множества фокальных элементов A тела свидетельств $F=(A,m)$ (m — функция масс);

2) нахождение укрупненной (упрощенной) структуры множества фокальных элементов $\tilde{A}$;

3) перераспределение масс фокальных элементов множества A на фокальные элементы из $\tilde{A}$ – в результате получим новую функцию масс $\tilde{m}$ и др.

Эти задачи можно решать на основе кластеризации множества фокальных элементов. При этом мы предполагаем, что сложная структура тела свидетельств часто является результатом агрегирования разнородной информации. Эта информация, полученная из разных источников, может быть противоречивой (конфликтной). Поэтому общий подход к кластеризации тела свидетельств может быть следующий. В полученном разбиении на кластеры исходного множества фокальных элементов противоречивость внутри кластеров должна быть минимальной, а между кластерами — максимальной. Этот подход аналогичен принципу компактности в кластерном анализе — расстояния между элементами одного кластера должны быть минимальны, а между кластерами — максимальны. Но в качестве функционала близости между кластерами вместо метрики используются мера внешнего конфликта (противоречивости) между телами свидетельств [154] и/или внутреннего конфликта самих тел свидетельств [155].

Идея аппроксимации тела свидетельств F более простым телом свидетельств $\tilde{F}$ путем иерархической кластеризации множества фокальных элементов была предложена в [156, 157] и развита в [158]. Причем кластеризация осуществлялась путем объединения и пересечения «близких» фокальных элементов и суммирования их масс. Мера близости двух фокальных элементов определялась с помощью некоторого функционала приращения информативности (взвешенного расстояния Хэмминга) после объединения/приращения этих двух

элементов. В [159] был предложен алгоритм кластеризации, основанный на понятии плотности конфликта.

В общем случае в результате кластеризации множества фокальных элементов A мы получим его разбиение (или покрытие) $\{A_1,\ldots,A_l\}$, которому можно поставить в соответствие множество тел свидетельств $\{F_1,...,F_l\}$, где $F_i=(A_i,m_i),\ i=1,...,l$.

Кроме выявления структуры и упрощения тела свидетельств, кластеризация может быть использована для оценивания внутреннего конфликта $Con_{in}(F)$ исходного тела свидетельств $F=(A\ ,m)$: $Con_{in}(F)=Con(F_1,...,F_l)$, где Con — мера внешнего конфликта.

В данной части отчета описана разработка двух алгоритмов кластеризации тела свидетельств. Первый алгоритм основан на использовании функции плотности распределения конфликта. Второй алгоритм является аналогом алгоритма k-means, но вместо метрики используется мера внешнего конфликта.

3.2.2 Некоторые сведения из теории свидетельств

Напомним коротко основные понятия теории свидетельств [160, 161]. Пусть X — конечное множество, 2^X — множество всех подмножеств на X. Телом свидетельств на X называют пару $F=(A\ ,m)$, где A — множество непустых подмножеств (фокальных элементов) из X, $m:2^X\to[0,1]$ — функция масс, удовлетворяющая условиям: $m(A)>0\Leftrightarrow A\in A$, $\sum_A m(A)=1$. Пусть $F\ (X)$ – множество всех тел свидетельств на X.

Частными случаями тел свидетельств являются: категоричное свидетельство $F_A=(\{A\},1)$, бессодержательное свидетельство $F_X=(\{X\},1)$, простое свидетельство $F_A^\alpha=\alpha F_A+(1-\alpha)F_X,\ \alpha\in[0,1]$. Произвольное тело свидетельств $F=(A\ ,m)$ можно представить в виде: $F=\sum_{A\in A} m(A)F_A$.

С телом свидетельств $F=(A\ ,m)$ связаны некоторые функции множеств, которые также однозначно представляют свидетельстсво: функция доверия $Bel(A)=\sum_{B\subseteq A} m(B)$, функция правдоподобия $Pl(A)=1-Bel(\neg A)=\sum_{A\cap B\neq\varnothing} m(B)$ и др.

Важной совместной характеристикой двух и более тел свидетельств является их противоречивость (конфликтность). Под (внешним) конфликтом двух тел свидетельств $F_1=(A_1,m_1)$ и $F_2=(A_2,m_2)$ понимают величину $Con(F_1,F_2)$ (обычно из промежутка [0,1]), которая принимает тем большее значение, чем больше и с большими массами есть пар непересекающихся фокальных элементов (A,B) двух тел свидетельств: $A\in A_1,\ B\in A_2$,

$A \cap B = \varnothing$. Обзор по мерам внешнего конфликта можно найти в [154]. Ниже будем использовать каноническую меру внешнего конфликта, связанную с правилом Демпстера: $Con(F_1, F_2) = \sum_{A \cap B = \varnothing} m_1(A) m_2(B)$.

Наряду с мерой внешнего конфликта между несколькими телами свидетельств рассматривают и противоречивость информации, предоставляемой одним телом свидетельств. Таковыми, например, могут быть тела свидетельств, которые получены в результате комбинирования нескольких тел противоречивых тел свидетельств. Противоречивость тела свидетельств $F = (A, m)$ оценивается с помощью некоторой меры внутреннего конфликта $Con_{in}(F) : F\ (X) \to [0,1]$ [155].

3.2.3 Кластеризация тела свидетельств

3.2.3.1 Сужение и продолжение функции масс

Пусть имеется тело свидетельств $F = (A, m)$, где A — множество фокальных элементов этого свидетельства. Рассмотрим некоторое подмножество $A' \subseteq A$. Сужением функции масс m на $A' \subseteq A$ назовем функцию множеств $m' : 2^X \to [0,1]$, $m'(A) = m(A)\ \forall\ A \in A'$ и $m'(A) = 0\ \forall\ A \notin A'$. В общем случае функция масс m' не удовлетворяет условию нормировки $\sum_A m'(A) = 1$. Поэтому пара (A', m') не определяет некоторое тело свидетельств. Чтобы построить тело свидетельств F', отражающее распределение m', необходимо продолжить функцию множеств m' до функции масс $\tilde{m}'$. Это можно сделать разными способами. Примеры продолжений:

1) пропорциональное продолжение: $\tilde{m}'(A) = m'(A) \Big/ \sum_{B \in A'} m'(B)\ \forall\ A \in A'$;

2) несодержательное продолжение: $\tilde{m}'(A) = m'(A)\ \forall\ A \in A' \setminus X$, $\tilde{m}'(X) = m'(X) + 1 - \sum_{B \in A'} m'(B)$.

Заметим, что различные продолжения функции множеств до функции масс некоторого тела свидетельств используются в правилах комбинирования. Например, в конъюнктивных правилах пропорциональное продолжение применяется в правиле Демпстера [160], а несодержательное — в правиле Ягера [162].

Если зафиксировано некоторое правило продолжения функции масс, то для подмножества $A' \subseteq A$ новое тело свидетельств $F' = (A', \tilde{m}')$ будет однозначно определяться по исходному телу свидетельств $F = (A, m)$ множеством A'. Поэтому такое тело свидетельств будем обозначать как $F(A') = (A', \tilde{m}')$.

В частности, если используется несодержательное продолжение, то для любого множества $A \in A$ тело свидетельств $F(\{A\}) = F_A^{m(A)} = m(A)F_A + (1-m(A))F_X$ будет простым.

3.2.3.2 Постановки задач кластеризации тела свидетельств на основе оптимизации конфликта

В основе оценивания внутреннего конфликта методом кластерной декомпозиции тела свидетельств лежит предположение о том, что тело свидетельств с большим внутренним конфликтом является неоднородным как по структуре множества фокальных элементов, так и по распределению функции масс на этом множестве. При этом возможны разные постановки задачи кластеризации. Отметим некоторые из них.

Пусть имеется тело свидетельств $F = (A, m)$. Требуется найти такое разбиения множества всех фокальных элементов A на подмножества $\{A_1,\ldots,A_l\}$, чтобы:

1) максимизировать внешний конфликт между свидетельствами-кластерами: $Con(F(A_1),\ldots,F(A_l)) \to \max$, где Con — мера внешнего конфликта;

2) минимизировать суммарный внутренний конфликт внутри кластеров $\sum_{i=1}^{l} Con_{in}(F(A_i)) \to \min$, где Con_{in} — мера внутреннего конфликта;

3) минимизировать суммарный конфликт между центрами кластеров и телами свидетельств, построенных на фокальных элементах этих кластеров $\sum_{i=1}^{l}\sum_{B\in A_i} Con\left(F(\{B\}), C_i\right) \to \min$, где C_i — эталонное тело свидетельств, соответствующее i-му кластеру, $i = 1,...,l$.

В более общей постановке можно искать не разбиение, а покрытие множества фокальных элементов A .

3.2.3.3 Кластерная декомпозиция свидетельств на основе функций плотности конфликта

Можно предложить алгоритм кластерной декомпозиции свидетельств, основанный на вычислении «плотности» распределения фокальных элементов.

3.2.3.3.1 Функция плотности распределения конфликта

В [159] было введено понятие плотности конфликта. Эта плотность использовалась для выделения в множестве фокальных элементов A небольшого подмножества $A' \subseteq A$ фокальных элементов, каждый из которых вносит наибольший вклад в конфликт с другими фокальными элементами (т. е. имеют большую массу и высокую плотность конфликта). После чего находилась такая функция масс m', которая минимизировала бы некоторую метрику между телами свидетельств $F = (A, m)$ и $F' = (A', m')$.

Напомним понятие плотности конфликта [159]. Пусть $F=(A,m)$ — тело свидетельств. Функцией плотности распределения конфликта тела свидетельств F называют отображение $\psi_F: 2^X \to [0,1]$, которое удовлетворяет условиям:

1) $\psi_F(A)=0$, если $B\cap A\neq\varnothing \ \ \forall B\in A$;

2) $\psi_F(A)=1$, если $B\cap A=\varnothing \ \ \forall B\in A$;

3) $\psi_{\alpha F_1+\beta F_2}=\alpha\psi_{F_1}+\beta\psi_{F_2} \ \ \forall F_1,F_2\in F\ (X)$, где $\alpha+\beta=1,\ \alpha\geq 0,\ \beta\geq 0$.

Нетрудно показать [159], что функция множеств, удовлетворяющая условиям 1) − 3), равна $\psi_F(A)=\sum_{B:A\cap B=\varnothing} m(B)=1-Pl(A)$. Отсюда видно, что значение $\psi_F(A),\ A\in A$ должна быть тем большим, чем больше фокальных элементов $B\in A$ с большими массами не пересекаются с A. Заметим, что функция ψ_F рассматривалась в [163] и называлась функцией противоречивости (inconsistency function).

Основная идея алгоритма кластеризации множества фокальных элементов A на основе функции плотности конфликта состоит в том, что «центры» кластеров должны иметь большое значение плотности конфликта.

Для нахождения таких «центров» вместо непосредственно функции плотности ψ_F будем использовать функцию $\varphi_F(A)=m(A)\psi_F(A),\ A\in A$, которая будет принимать большие значения на тех фокальных элементах, которые имеют не только большую плотность, но и большую массу.

3.2.3.3.2 Алгоритм кластерной декомпозиции свидетельств на основе функций плотности конфликта

Теперь алгоритм кластерной декомпозиции свидетельств на основе функций плотности конфликта будет состоять из следующих шагов.

Алгоритм 1.

1) Вычисляем значения функции множеств $\varphi_F(A),\ A\in A$. Если $\varphi_F(A)=0$ для всех $A\in A$, то алгоритм останавливаем. В этом случае мы имеем неконфликтность тела свидетельств A : $B\cap A\neq\varnothing \ \ \forall A,B\in A$. Поэтому кластеризации не будет.

2) Если же существуют $A\in A$, для которых $\varphi_F(A)>0$, то упорядочиваем такие фокальные элементы по убыванию значений функции φ_F: $\varphi_F(A_1)\geq\varphi_F(A_2)\geq....$ По характеру убывания последовательности $\{\varphi_F(A_i)\}$ выбираем число кластеров l, начальные кластеры будут совпадать с выбранными фокальными элементами: $A_i^{(0)}=\{A_i\},\ i=1,...,l$.

3) Остальные фокальные элементы перераспределяем по кластерам $A_1^{(0)},...,A_l^{(0)}$ по принципу максимизации конфликта между кластерами-свидетельствами – относим фокальный элемент $B \in A \setminus \left\{A_1^{(0)},...,A_l^{(0)}\right\}$ к тому кластеру $A_i^{(0)}$, при включении в который достигается максимум меры конфликта: $A_i^{(0)} = \arg\max\limits_{j:B\in A_j^{(0)}} Con\left(F\left(A_1^{(0)}\right),...,F\left(A_j^{(0)} \cup \{B\}\right),...,F\left(A_l^{(0)}\right)\right)$.

Если получаются равные максимальные значения конфликта при отнесении B разным кластерам $A_j^{(0)}$, $j \in J$, то включаем B во все эти кластеры, а значение массы $m(B)$ равномерно распределяем по обновляемым кластерам, т. е. B будет входить в каждый кластер $A_j^{(0)}$, $j \in J$ с весом $m(B)/|J|$. В результате получим покрытие $\{A_1,\ldots,A_l\}$ множества всех фокальных элементов A . Функции масс m_i на A_i, $i = 1,...,l$ строятся с помощью процедур сужения и продолжения.

В этом алгоритме на шаге 3 применяется некоторое фиксированное правило продолжения функции масс.

Пример 3.1. Пусть $X = \{1,2,3\}$ и на X задано тело свидетельств $F = 0.3F_{\{1\}} + 0.2F_{\{2\}} + 0.3F_{\{1,3\}} + 0.2F_{\{2,3\}}$, т. е. $A = \{\{1\},\{2\},\{1,3\},\{2,3\}\}$.

Шаг 1: найдем значения функции φ_F: $\varphi_F(\{1\}) = \varphi_F(\{2\}) = 0.12$, $\varphi_F(\{1,3\},\{2,3\}) = 0.06$.

Шаг 2: выберем $l = 2$ кластера и $A_1^{(0)} = \{\{1\}\}$, $A_2^{(0)} = \{\{2\}\}$.

Шаг 3: распределим оставшиеся два фокальных элемента по кластерам.

Для $B = \{1,3\}$ имеем. Если $B \in A_1$, то получим: $F\left(\{B\} \cup A_1^{(0)}\right) = 0.3F_{\{1\}} + 0.3F_{\{1,3\}} + 0.4F_X$, $F\left(A_2^{(0)}\right) = 0.2F_{\{2\}} + 0.8F_X$. Тогда $Con\left(F\left(\{B\} \cup A_1^{(0)}\right), F\left(A_2^{(0)}\right)\right) = 0.12$. Если же $B \in A_2$, то получим: $F\left(A_1^{(0)}\right) = 0.3F_{\{1\}} + 0.7F_X$, $F\left(\{B\} \cup A_2^{(0)}\right) = 0.2F_{\{2\}} + 0.3F_{\{1,3\}} + 0.5F_X$ и $Con\left(F\left(A_1^{(0)}\right), F\left(\{B\} \cup A_2^{(0)}\right)\right) = 0.06$. Таким образом, относим $B = \{1,3\}$ к кластеру A_1.

Для $B = \{2,3\}$ имеем. Если $B \in A_1$, то получим $F\left(\{B\} \cup A_1^{(0)}\right) = 0.3F_{\{1\}} + 0.2F_{\{2,3\}} + 0.5F_X$, $F\left(A_2^{(0)}\right) = 0.2F_{\{2\}} + 0.8F_X$ и $Con\left(F\left(\{B\} \cup A_1^{(0)}\right), F\left(A_2^{(0)}\right)\right) = 0.06$. Если же $B \in A_2$, то получим: $F\left(A_1^{(0)}\right) = 0.3F_{\{1\}} + 0.7F_X$, $F\left(\{B\} \cup A_2^{(0)}\right) = 0.2F_{\{2\}} + 0.2F_{\{2,3\}} + 0.6F_X$ и $Con\left(F\left(A_1^{(0)}\right), F\left(\{B\} \cup A_2^{(0)}\right)\right) = 0.12$. Таким образом, относим $B = \{2,3\}$ к кластеру A_2

и получим окончательную кластеризацию фокальных элементов $A_1 = \{\{1\},\{1,3\}\}$, $A_2 = \{\{2\},\{2,3\}\}$.

Замечание 3.1. На шаге 2 алгоритма, кроме учета плотности конфликта (функции φ_F) можно учитывать, как это делалось в [159], и расстояние между выбранными фокальными элементами. Выбираются фокальные элементы в порядке убывания функции φ_F. Если A' — множество уже выбранных фокальных элементов, то очередной элемент A_k добавляется к этому множеству при условии, что $\min_{A \in A'} d(F(\{A\}), F(\{A_k\})) > h$, где d — некоторая метрика на множестве тел свидетельств [164], h — пороговое значение.

3.2.3.4 Алгоритм k-means для тела свидетельств

Пусть имеется тело свидетельств $F = (A, m)$. Требуется найти такое покрытие множества всех фокальных элементов A подмножествами (кластерами) $C = \{A_1, \ldots, A_l\}$, которое минимизировало бы внутрикластерный конфликт. По аналогии с классическим алгоритмом k-means будем использовать понятие центров кластеров фокальных элементов. Под центром i-го кластера A_i будем рассматривать некоторое тело свидетельств C_i, построенное по паре (A_i, m_i), где m_i — сужение функции масс m на $A_i \subseteq A$, $i = 1, \ldots, l$.

В качестве минимизируемого функционала по аналогии с алгоритмом k-means рассмотрим суммарный конфликт между центрами кластеров и телами свидетельств, построенных на фокальных элементах этих кластеров:

$$\Phi = \sum_{i=1}^{l} \sum_{B \in A_i} Con\left(F(\{B\}), C_i\right), \tag{3.2}$$

где $F(\{B\})$ — свидетельство, построенное по множеству $\{B\}$ с помощью процедур сужения и продолжения. В частности, если выбрано несодержательное продолжение, то $F(\{B\}) = m(B)F_B + (1 - m(B))F_X$.

В этом алгоритме число тел свидетельств, на которые декомпозируется свидетельство $F = (A, m)$, будем считать фиксированным (оно определяется из некоторых других эвристических соображений). Также заранее заданным считается способ продолжения функции масс.

Предположим, что покрытие $C=\{A_1,\ldots,A_l\}$ фиксировано, а центр i-го кластера имеет вид

$$C_i=\sum_{A\in A_i}\alpha_i(A)F_A, \tag{3.3}$$

где $\boldsymbol{\alpha}_i=(\alpha_i(A))_{A\in A_i}\in S_{|A_i|}$, $S_k=\left\{(t_1,...,t_k):t_i\ge 0,i=1,...,k,\sum_{i=1}^k t_i=1\right\}$ — k-мерный симплекс. Тогда для несодержательного продолжения (в частности, $F(\{B\})=m(B)F_B+(1-m(B))F_X$, если $B\in A$) имеем

$$\Phi=\sum_{i=1}^{l}\sum_{B\in A_i}Con(F(\{B\}),C_i)=\sum_{i=1}^{l}\sum_{B\in A_i}m(B)\sum_{\substack{A\in A_i:\\ A\cap B=\varnothing}}\alpha_i(A)=$$

$$=\sum_{i=1}^{l}\sum_{B\in A_i}m(B)\left(1-\sum_{\substack{A\in A_i:\\ A\cap B\neq\varnothing}}\alpha_i(A)\right)=k_C-\sum_{i=1}^{l}\sum_{A\in A_i}\alpha_i(A)\sum_{\substack{B\in A_i:\\ A\cap B\neq\varnothing}}m(B)=k_C-\sum_{i=1}^{l}Q_i(\boldsymbol{\alpha}_i), \tag{3.4}$$

где $k_C=\sum_{i=1}^{l}\sum_{B\in A}m(B)\ge 1$ ($k_C=1\Leftrightarrow C=\{A_1,\ldots,A_l\}$ — разбиение множества фокальных элементов), $Q_i(\boldsymbol{\alpha}_i)=\sum_{A\in A_i}\alpha_i(A)Pl_{A_i}(A)$ и $Pl_{A_i}(A)=\sum_{\substack{B\in A_i:\\ A\cap B\neq\varnothing}}m(B)$ — сужение функции правдоподобия на множество A_i. Тогда при фиксированном покрытии $C=\{A_1,\ldots,A_l\}$ минимум функционала Φ будет достигаться тогда, когда линейные функции $Q_i(\boldsymbol{\alpha}_i)$ будут достигать максимумов на симлексах $S_{|A_i|}$, $i=1,\ldots,l$. Но

$$\max_{\boldsymbol{\alpha}\in S_{|A_i|}}Q_i(\boldsymbol{\alpha})=\max_{A\in A_i}Pl_{A_i}(A),\ i=1,\ldots,l. \tag{3.5}$$

Пусть $\overline{A_i}=\left\{A\in A_i:A=\arg\max_{A\in A_i}Pl_{A_i}(A)\right\}$. Если

$$C_i=\sum_{A\in\overline{A_i}}\alpha_i(A)F_A,\ \boldsymbol{\alpha}_i=(\alpha_i(A))_{A\in\overline{A}_i}\in S_{|\overline{A_i}|}\ i=1,\ldots,l, \tag{3.6}$$

то при таком выборе центров кластеров функционал Ф будет достигать минимума при фиксированном покрытии $C=\{A_1,\ldots,A_l\}$. Этот минимум при фиксированном покрытии $C=\{A_1,\ldots,A_l\}$ будет равен

$$\min \Phi = k_C - \sum_{i=1}^{l} \max_{A\in A_i} Pl_{A_i}(A) \qquad (3.7)$$

и не будет зависеть от выбора параметров $\boldsymbol{\alpha}_i = \left(\alpha_i(A)\right)_{A\in\overline{A}_i} \in S_{|\overline{A_i}|}$, $i=1,\ldots,l$.

Тогда по аналогии с классическим алгоритмом k-means алгоритм кластеризации тела свидетельств будет следующим.

Алгоритм 2.

1) Выбираем и фиксируем число кластеров l. Назначаем начальные центры кластеров – тела свидетельств $C_i^{(0)}$, $i=1,...,l$. Фиксируем порог максимального конфликта внутри кластеров $Con_{\max}\in[0,1]$. Полагаем $s=0$.

2) Фокальные элементы перераспределяем по кластерам по принципу минимизации конфликта между кластерами-свидетельствами и центрами кластеров: $B\in A$ относим к тому кластеру $A_i^{(s)}$, для которого достигается $\min_i Con\left(F(\{B\}),C_i^{(s)}\right)\le Con_{\max}$. Если же $\min_i Con\left(F(\{B\}),C_i^{(s)}\right)>Con_{\max}$, то фокальный элемент B назначается центром нового кластера. В результате получим кластеры $A_i^{(s)}$, $i=1,...,l$.

3) Вычисляем новые центры кластеров по формуле (3.7). Наращиваем $s\leftarrow s+1$.

4) Шаги 2 и 3 повторяем до стабилизации кластеров (или их центров).

Утверждение 3.1. Алгоритм 2 сходится за конечное число шагов.

Замечание 3.2. На 1-м шаге алгоритма в качестве начальных центров кластеров могут быть выбраны тела свидетельств $C_i^{(0)}=F_{A_i}$, где фокальные элементы A_i, $i=1,...,l$ выбираются произвольным образом или, например, с помощью алгоритма максимизации плотности.

Замечание 3.3. Центры кластеров могут зависеть от параметров $\boldsymbol{\alpha}=\left(\alpha(A)\right)_{A\in\overline{A}_i}\in S_{|\overline{A_i}|}$ (см. формулу (3.7)). В этом случае на 2-м или 3-м шагах алгоритма необходимо использовать дополнительные процедуры выбора параметров $\boldsymbol{\alpha}=\left(\alpha(A)\right)_{A\in\overline{A}_i}\in S_{|\overline{A_i}|}$. В качестве критериев выбора могут рассматриваться:

1) минимизация покрытия, т. е. параметры на 2-м шаге алгоритма выбираем такими, чтобы покрытие $C = \{A_1,\ldots,A_l\}$ «было ближе» к разбиению. Например, $\sum_{i=1}^{l} |A_i| \to \min$.

2) минимизация неопределенности свидетельств-центров кластеров C_i, $i = 1,...,l$. Эта процедура применяется на 3-м шаге алгоритма. Неопределенность свидетельства может быть оценена с помощью одного из индексов неточности [165], например, обобщенной меры Хартли $H(C_i) = \sum_{A \in \overline{A}_i} \alpha_i(A) \ln |A|$.

3) минимизация расстояния между центрами кластеров и исходным телом свидетельств относительно некоторой метрики d между телами свидетельств [164]: $d(C_i, F) \to \min$, $i = 1,...,l$;

4) максимизация расстояния между центрами кластеров $d(C_i, C_j) \to \max$ или максимизации конфликта $Con(C_i, C_j) \to \max$, $i, j = 1,...,l$ $(i \neq j)$

и т. д.

Замечание 3.4. Один из способов оценивания внутреннего конфликта [155] тела свидетельств F на X связан с нахождением максимума контурной функции [166]: $Con(F) = 1 - \max_{x \in X} Pl(x)$. Тогда формулу (3.8) можно интерпретировать как суммарный внутрикластерный внутренний конфликт.

Замечание 3.5. Можно искать центры кластеров C_i, $i = 1,...,l$ в виде (3.6), минимизирующие функционал (3.5) при фиксированном покрытии $C = \{A_1,\ldots,A_l\}$ и при условии ограниченности неопределенностей центров кластеров C_i:

$$H(C_i) = \sum_{A \in A_i} \alpha_i(A) \ln |A| \le u_i,\ i = 1,...,l, \tag{3.8}$$

где u_i, $i = 1,...,l$ — некоторые пороговые значения. Тогда задача минимизации функционала (3.5) при фиксированном покрытии $C = \{A_1,\ldots,A_l\}$ сводится к решению l задач линейного программирования:

$$\sum_{A \in A_i} \alpha_i(A) Pl_{A_i}(A) \to \max \text{ при } \boldsymbol{\alpha}_i = \left(\alpha_i(A)\right)_{A \in A_i} \in S_{|A_i|},\ \sum_{A \in A_i} \alpha_i(A) \ln |A| \le u_i,$$
$$i = 1,...,l. \tag{3.9}$$

Пример 3.2. Для тела свидетельств из Примера 3.1 ($X=\{1,2,3\}$, $F=0.3F_{\{1\}}+0.2F_{\{2\}}+0.3F_{\{1,3\}}+0.2F_{\{2,3\}}$) Алгоритм 2 для двух кластеров и несодержательного продолжения даст следующий результат.

Шаг 1: выберем число кластеров $l=2$; пусть начальные центры кластеров $C_1^{(0)}=F_{\{1\}}$, $C_2^{(0)}=F_{\{2\}}$; $s=0$.

Шаг 2: имеем

$$Con\left(F(\{1\}),C_1^{(0)}\right)=Con\left(F(\{1,3\}),C_1^{(0)}\right)=0,$$

$$Con\left(F(\{2\}),C_1^{(0)}\right)=Con\left(F(\{2,3\}),C_1^{(0)}\right)=0.2,$$

$$Con\left(F(\{1\}),C_2^{(0)}\right)=Con\left(F(\{1,3\}),C_2^{(0)}\right)=0.3,$$

$$Con\left(F(\{2\}),C_2^{(0)}\right)=Con\left(F(\{2,3\}),C_2^{(0)}\right)=0$$

(например, $Con\left(F(\{1,3\}),C_2^{(0)}\right)=Con\left(0.3F_{\{1,3\}}+0.7F_X,F_{\{2\}}\right)=0.3$).

Поэтому по принципу минимизации конфликта между кластерами-свидетельствами и центрами кластеров начальная кластеризация будет иметь вид $A_1{}^{(0)}=\{\{1\},\{1,3\}\}$, $A_2{}^{(0)}=\{\{2\},\{2,3\}\}$.

Шаг 3: вычисляем новые центры кластеров по формуле (3.7): $Pl_{A_1^{(0)}}(\{1\})=0.3+0.3=0.6$, $Pl_{A_1^{(0)}}(\{1,3\})=0.3+0.3=0.6$, $Pl_{A_2^{(0)}}(\{2\})=0.2+0.2=0.4$, $Pl_{A_2^{(0)}}(\{2,3\})=0.2+0.2=0.4$.

Поэтому $C_1^{(1)}=\alpha F_{\{1\}}+(1-\alpha)F_{\{1,3\}}$ и $C_2^{(1)}=\beta F_{\{2\}}+(1-\beta)F_{\{2,3\}}$, $\alpha,\beta\in[0,1]$.

Если потребовать минимизации обобщенной меры Хартли (см. Замечание 3.2), то $C_1^{(1)}=\arg\min\limits_{0\le\alpha\le1}H\left(\alpha F_{\{1\}}+(1-\alpha)F_{\{1,3\}}\right)=F_{\{1\}}$, $C_2^{(1)}=\arg\min\limits_{0\le\beta\le1}H\left(\beta F_{\{2\}}+(1-\beta)F_{\{2,3\}}\right)=F_{\{2\}}$ и алгоритм остановит свою работу, так как центры кластеров не изменились. Если же мы применим правило минимизации покрытия (см. Замечание 3.3), то перейдем к следующему шагу.

Шаг 4: перераспределяем фокальные элементы по критерию наименьшего конфликта с новыми центрами:

$$Con\left(F(\{1\}),C_1^{(1)}\right)=Con\left(F(\{1,3\}),C_1^{(1)}\right)=0,$$

$$Con\left(F(\{2\}),C_1^{(1)}\right)=0.2,\ Con\left(F(\{2,3\}),C_1^{(1)}\right)=0.2\alpha$$

$$Con\left(F(\{1\}),C_2^{(1)}\right)=0.3,\ Con\left(F(\{1,3\}),C_2^{(1)}\right)=0.3\beta,$$

$$Con\left(F(\{2\}),C_2^{(1)}\right)=Con\left(F(\{2,3\}),C_2^{(1)}\right)=0. \quad (3.10)$$

Применяя правило минимизации покрытия (см. Замечание 3.3), получим $A_1^{(1)}=\{\{1\},\{1,3\}\}$, $A_2^{(1)}=\{\{2\},\{2,3\}\}$. Кластеры стабилизировались. Стоп алгоритма.

После выполнения процедуры кластерной декомпозиции мы получим, по сути, новое свидетельство, определенное на базовом множестве A, фокальными элементами этого нового свидетельства будут найденные множества покрытия (кластеры) $C=\{A_1,\ldots,A_l\}$, функция масс $m(A_i)=\sum_{B\in A_i} m(B)/n(B)$, где $n(B)=\left|\{A_i : B\in A_i\}\right|$ (число кластеров, содержащих множество B). Такое свидетельство можно считать свидетельством второго порядка, которое отражает укрупненную структуру исходного свидетельства.

3.2.4 Оценивание внутреннего конфликта тела свидетельств на основе его кластеризации

Предположим, что тем или иным способом получено кластерное покрытие (в частном случае разбиение) $C=\{A_1,\ldots,A_l\}$ тела свидетельств $F=(A,m)$. Тогда можно предложить следующие способы оценивания внутреннего конфликта этого тела свидетельств с помощью некоторой меры внешнего конфликта Con:

1) $Con_1(F)=Con(F(A_1),\ldots,F(A_l))$;

$Con_2(F)=Con(C_1,\ldots,C_l)$, где $C_1,\ldots,C_l$ — тела свидетельств — центры кластеров $A_1,\ldots,A_l$ соответственно.

Например, для тела свидетельств, рассмотренного в примере, несодержательного продолжения и канонической меры внешнего конфликта, имеем:

$$A_1=\{\{1\},\{1,3\}\},\ A_2=\{\{2\},\{2,3\}\} \text{ и } Con_1(F)=Con(F(A_1),F(A_2))=0.18,$$
$$Con_2(F)=Con(C_1,C_2)=\alpha+(1-\alpha)\beta,\ \alpha,\beta\in[0,1]. \quad (3.11)$$

Утверждение 3.2. Верно равенство

$$Con_1(F)=Con(F(A_1),\ldots,F(A_l))=\sum_{A_1\in A_1,\ldots,A_l\in A_l} Con(F(\{A_1\}),\ldots,F(\{A_l\})). \quad (3.12)$$

3.2.5 Некоторые выводы

Разработаны два метода кластеризации тела свидетельств. Каждый из этих методов предполагает, что одному кластеру должны принадлежать слабо конфликтные фокальные элементы, а разным кластерам — сильно конфликтные. Это требование аналогично основному принципу компактности в кластерном анализе – расстояния между элементами одного кластера должны быть минимальны, а между кластерами — максимальны.

Первый алгоритм основан на использовании функции плотности конфликта. Второй алгоритм реализует идею метода k-means: формируются оптимальным образом тела свидетельств, которые выступают в качестве центров кластеров, после чего перераспределяются фокальные элементы по принципу минимизации конфликта с центрами кластеров.

Показано, как кластеризация может быть использована для оценивания внутреннего конфликта тела свидетельств. Результаты этой части отчета нашли отражение в [167].

4. Исследования по теории и приложениям сетевого анализа данных

В этом разделе отчета приведены результаты по развитию и исследованию методов сетевого анализа данных, а также по приложению этих методов к анализу конкретных сетевых систем. Эти исследования осуществлялись в следующих научных направлениях:

– аксиоматика мер центральности;

– модели схожести сетевых структур;

− сетевой анализ публикаций в области научных исследований болезни Паркинсона;

− сценарный анализ ущерба экологии от разлива нефти в Арктическом регионе на основе имитационной сетевой модели.

4.1 Аксиоматика мер центральности

4.1.1 Введение

Сетевая структура состоит из вершин, которые обозначают объекты исследования, и рёбер, которые обозначают связи между рассматриваемыми объектами. При изучении сетевых структур возникает задача выделения её ключевых элементов. В данном разделе мы сфокусируемся на выделении наиболее значимых вершин, в то время как в литературе также известны задачи по выделению важных рёбер в сети.

Одним из самых известных и классических подходов к задаче выделения ключевых элементов в графе является применение индексов центральности. В литературе известен целый ряд индексов центральности, различающихся интерпретацией заложенной в них модели. Ниже мы приведём описание некоторых основных классических мер центральности.

Самой популярной и простой мерой центральности является степенная центральность, оценивающая общее число соседей для каждой из вершин [168]. Существует взвешенная версия данной центральности, а также степенная центральность, учитывающая направление рёбер в графе.

Другая группа центральностей сводится к вычислению собственного вектора для рассматриваемой сети. В данную группу центральностей входят центральность по собственному вектору [169], PageRank [170], центральность по Боначичу [171], центральность Катца [172], Hubs and Authorities [173] и многие другие. Данные центральности учитывают не только количество соседей, но и их собственную центральность.

Ещё одной группой центральностей являются меры, учитывающие кратчайшие пути между вершинами в графе, например, центральность по близости [174] и центральность по промежуточности [175].

Известны также многие другие меры центральности, чьё описание можно найти, например, в работе [176].

С практической точки зрения все эти меры имеют разную интерпретацию и, соответственно, зачастую дают разный результат, выделяя те или иные ключевые элементы. С теоретической точки зрения встаёт вопрос о сравнении данных мер центральностей между собой, а именно об их валидности и состоятельности.

В задаче выделения ключевых элементов в сети не существует правильного ответа на вопрос, кто действительно является ключевым объектом (ground truth), и мы не можем говорить о качестве мер центральности с точки зрения их близости или отклонения от истинных результатов. Соответственно, сравнивать меры между собой можно, например, проверяя выполнение тех или иных свойств. Так, в литературе известен ряд аксиом для мер центральности, которые дают понимание о работе рассматриваемых индексов и о том, будет ли выявлен тот или иной ключевой элемент, если граф имеет определённую структуру. Ниже мы приведём описание некоторых известных аксиом, а также приведём список индексов центральности, для которых эти аксиомы выполняются.

4.1.2 Описание аксиом мер центральности

В данном пункте мы рассмотрим некоторые аксиомы мер центральности, а также опишем, для каких индексов они выполняются.

Одной из самых простых аксиом является аксиома анонимности, представленная в работе [177]. Аксиома анонимности выполняется, если центральность вершины не зависит от её имени, а именно, центральность вершины совпадает с соответствующей центральностью вершины в изоморфном графе. Данная аксиома выполняется для всех центральностей, в которых не учитываются атрибуты вершин.

Другой аксиомой в данной работе является аксиома роста конечной точки. Для центральности выполняется данная аксиома, если при добавлении ребра между двумя вершинами значения их центральностей не уменьшаются: $\forall G,\ \forall u, v \in G\ \ C_u(G + \{u, v\}) \geq C(G)$ и $C_v(G + \{u, v\}) \geq C(G)$, где G – некоторый граф, u, v – вершины графа, $C_i(G)$ значение центральности C вершины i в графе G. Данная аксиома выполняется, например, для степенной центральности, центральности близости, центральности Катца, но не выполняется для центральности по собственному вектору, центральности по промежуточности и некоторых других мер [178].

В работе [177] также приведена аксиома монотонности, которая говорит о том, что при добавлении в граф ребра значения центральности вершин не уменьшается: $\forall G,$ $\forall u, v, w \in G\ \ C_u(G + \{v, w\}) \geq C_u(G)$. Данная аксиома выполняется для таких центральностей, как степенная, центральность по близости, центральности Катца, но не выполняется

для центральности по собственному вектору, Pagerank, центральности по промежуточности и ряда других мер [178].

Ещё одной аксиомой в данной работе является аксиома центральной вершины. Если вершина в графе имеет наибольшее значение центральности, то при добавлении к этой вершине инцидентного ребра, она также будет иметь наибольшее значение центральности: $\forall G$ и u: $\forall v \in G\ C_u(G) \geq C_v(G)$, то $\forall w, t \in G\ C_u(G + \{u, w\}) \geq C_t(G + \{u, w\})$. В частности, данная аксиома выполняется для степенной центральности, но не выполняется для центральности близости и промежуточности, а также для центральности по собственному вектору [178].

В работе [179] приведена аксиома справедливости. Аксиома выполняется, если при добавлении ребра между вершинами значение их центральности меняется на одинаковую величину: $\forall G,\ \forall u, v \in G\ \ C_u(G + \{u, v\}) - C_u(G) = C_v(G + \{u, v\}) - C_v(G)$. Данная аксиома выполняется, например, для степенной центральности, но не выполняется для центральности близости, промежуточности и группы центральностей по собственному вектору.

В другой работе Майерсона [180] приведен принцип равнозначного вклада для значений Шэпли. Позже данный принцип был аксиоматизирован в работе [181]. Для центральности будет выполняться данная аксиома, если $\forall u, v \in G\ \ C_u(G) - C_u(G \backslash E_v) = C_v(G) - C_u(G \backslash E_u)$, где E_i — множество рёбер, инцидентных вершине i. Другими словами, удаление всех рёбер у вершины u влияет на значение центральности вершины v в той же мере, как удаление всех рёбер у вершины v влияет на значение центральности вершины u. Данная аксиома выполняется, например, для степенной центральности, но не выполняется для центральности близости, промежуточности, центральностей по собственному вектору и других [178].

Стоит отметить, что некоторые аксиомы сформулированы для графов только определённой структуры. Так, в работе [182] наряду с аксиомой монотонности для любых графов представлены аксиома размера и аксиома плотности для графов определённого типа.

Рассмотрим граф $S_{k,p}$ (образованного k-кликой и циклом длины p). Центральность удовлетворяет аксиоме размера, если $\forall k\ \exists P_k: \forall p \geq P_k$ в $S_{k,p}$ выполняется, что центральность любой вершины в p-цикле будет больше, чем центральность вершины в k-клике, и $\forall p\ \exists K_p: \forall k \geq K_p$ в $S_{k,p}$ выполняется, что центральность любой вершины в k-клике будет больше, чем центральность вершины в p-цикле. Иными словами, для каждой клики найдётся достаточно большой цикл, что все вершины цикла будут иметь значение центральности больше, чем в клике, и наоборот. Данная аксиома со стороны клики выполняется, например, для степенной центральности и центральности Катца, а со стороны цикла для центральности

промежуточности. Аксиома не выполняется для центральности по близости, PageRank и некоторых других.

Аксиому плотности можно сформулировать следующим образом. Рассмотрим граф $D_{k,p}$ (образованного k-кликой и циклом длины p, $k, p \geq 3$), с ребром между вершинами x и y, где x — вершина клики, а y — вершина цикла ($x \leftrightarrow y$ — мост). Центральность удовлетворяет аксиоме плотности, если для $k = p$ центральность x больше центральности y. Иными словами, если клика и цикл содержат одно и то же количество вершин и соединены между собой мостом, то значение центральности конечной вершины моста, которая находится в клике, будет больше, чем у той вершины, которая находится в цикле. Данная аксиома выполняется, например, для степенной центральности, центральности Катца, PageRank, Hubs and Authorities и не выполняется для центральности близости, промежуточности и некоторых других.

Некоторые другие аксиомы и их выполнение для различных мер центральности можно найти в [178].

4.1.3 Некоторые выводы

Для понимания работы той или иной меры центральности можно обратиться к аксиоматическому подходу и понять, каким свойствам рассматриваемая мера удовлетворяет, а каким не удовлетворяет. В данном подразделе мы привели некоторые аксиомы мер центральности для сетей общей структуры, а также для графов определённой структуры. Для каждой из описанных аксиом мы также привели список нескольких мер центральности, для которых эти аксиомы всегда выполняются или всегда не выполняются. В дальнейшем мы планируем расширить аксиоматику мер центральности новыми аксиомами, основанными на принципах теории функции выбора, и проверить их выполнение на графах простой структуры.

4.2 Модели схожести сетевых структур

4.2.1 Введение

В последнее время все больше внимания уделяется анализу сетевых структур, описывающих множество социальных, экономических, технологических, биологических и других сложных систем. Поскольку большинство структур изменяется со временем, возникает необходимость в отслеживании основных изменений сети, а также выявлении периодов, в рамках которых система находится в одинаковом состоянии. В связи с этим методы сетевого анализа играют важнейшую роль в исследовании сложных систем и изучении того, как различные элементы системы функционируют как по отдельности, так и как система в целом.

К сожалению, оценка схожести сетевых графов является плохо определенной задачей. Поскольку понятие сходства сетевых структур не имеет четкого определения, в литературе было предложено множество моделей, которые позволяют оценить меру схожести сетей по

тому или иному заранее заданному критерию. К сожалению, большинство таких моделей являются локальными и не позволяют учитывать глобальную структуру сети. Свойства ряда других моделей до сих пор не были изучены. Кроме того, многие модели схожести были разработаны только для ненаправленных графов, что ограничивает их применение к ряду существующих систем. Таким образом, возникает необходимость в исследовании свойств существующих моделей и их сравнении на примере различных сетевых структур. В рамках данного исследования проводится детальное изучение современных моделей оценки схожести сетевых структур с целью выявления их основных преимуществ и недостатков от их применения.

4.2.2 Основная часть

На данный момент в литературе известно множество моделей, с помощью которых можно оценить степень схожести или различия двух сетевых структур. Наиболее простые модели основаны на вычислении таких характеристик, как коэффициент корреляции между матрицами связности каждой сети, расстояние Хэмминга, либо коэффициента Жаккарда, который определяет долю связей, одновременно присутствующих в обеих сетевых структурах. Другим примером модели оценки схожести сетей является вычисление так называемого расстояния редактирования графа (Graph Edit Distance), который заключается в нахождении минимального числа изменений, необходимых для преобразования одной сети к другой [183]. Тем не менее, данные модели являются локальными и не учитывают непрямые взаимодействия между вершинами. В [184] рассматривается другой подход, который представляет собой усовершенствованную версию приведенных выше моделей. Вместо сравнения двух исходных матриц связности, предложенная модель (алгоритм называется DELTACON) производит сравнение двух новых матриц, построенных на основе оценки «близости» между вершинами сети с учетом ее структуры (например, оценивается расстояние между любой парой вершин). Отметим, что все перечисленные модели зависят от нумерации вершин, что ограничивает их применение для сравнения графов различной природы или решения задачи изоморфизма графов.

В последнее время активно развиваются модели, основанные на применении методов теории информации, в частности, вычислении энтропии. В частности, известны модели, основанные на вычислении дивергенции Йенсена − Шеннона (Jensen − Shannon entropy) между двумя вероятностными распределениями, построенными для каждой сети. Например, в [185] схожесть сетей оценивается на основе сравнения двух вероятностных распределений: распределения значений индекса центральности, а также распределения длин путей между вершинами. Наконец, в [186] предложен метод оценки схожести сетевых структур на основе

спектральной теории, который заключается в сравнении спектра матриц Лапласа, построенных по каждой сети. Стоит отметить, что большинство приведенных моделей основаны на сравнении общих характеристик графов, поэтому данные модели не чувствительны к конкретному названию вершин. Кроме того, свойства данных моделей до сих пор до конца не изучены. Наконец, перечисленные модели основаны на сравнении двух распределений, при этом они не учитывают, какие конкретно вершины являются наиболее важными в сетевых структурах.

Известны также модели, которые оценивает сходство сетей на основе нескольких параметров. Например, в работе [187] была предложена модель, которая позволяет производить оценку того, как изменяется сеть с точки зрения своей структуры, а также набора ключевых элементов. Изменение структуры оценивается с точки зрения попарного взаимоотношения узлов сети, в то время как ключевые элементы сравниваются на основе интервальных порядков. В [187] также показано применение данной модели к сети международной миграции, в результате чего было изучено изменение миграционных процессов между странами в различные годы.

В рамках данного исследования рассматриваются 39 различных моделей оценки схожести сетевых структур. Полный список исследуемых моделей, а также информация об их применимости к направленным, ненаправленным и взвешенным графам приведена в Таблице 4.1. Описание большинства моделей приведено в [188, 189]. Стоит отметить, что в Таблице 4.1 приведено только 29 моделей, поскольку остальные меры представляют собой модификацию приведенных 29 моделей с другими значениями параметров. Например, модель *k-hop NN* оценивает среднюю схожесть набора соседей каждой вершины, расположенных на расстоянии не более *k*. Соответственно, при значении *k=1* происходит сравнение только прямых соседей, при значении *k=2* сравниваются все соседи вершины, находящиеся на расстоянии не более 2, а при *k=3* рассматриваются только вершины, связанные через менее 3 посредников в графе.

Таблица 4.1 − Модели оценки схожести/различия графов

№	Название модели	Тип графа				Не чувствителен к названию вершин?
		Ненаправленный	Направленный	Невзвешенный	Взвешенный	
1	Jaccard Index (JI)	Да	Да	Да	Да	Нет
2	Graph Edit Distance (GED)	Да	Да	Да	Нет	Нет
3	Vertex/Edge Overlap (VEO)	Да	Да	Да	Нет	Нет

Продолжение таблицы 4.1

4	k-hop Nodes Neighborhood (k-hop NN)	Да	Да	Да	Нет	Нет
5	Maximum Common Subgraph Distance (MCS)	Да	Нет	Да	Нет	Нет
6	Frobenius distance (FRO)	Да	Да	Да	Да	Нет
7	Vector Similarity Algorithm (VS)	Да	Да	Да	Да	Нет
8	DELTACON	Да	Да	Да	Да	Нет
9	Polynomial Dissimilarity (POL)	Да	Да	Да	Нет	Нет
10	Graph Diffusion Distance (GDD)	Да	Да	Да	Да	Нет
11	Resistance Perturbation (RP)	Да	Да	Да	Да	Нет
12	Vertex Ranking (VR)	Да	Да	Да	Да	Нет
13	Degree Jenson-Shannon divergence (degreeJSD)	Да	Да	Да	Нет	Да
14	Portrait Divergence (POR)	Да	Да	Да	Нет	Да
15	Communicability Sequence Entropy (CSE)	Да	Да	Да	Нет	Да
16	NetSimile measure	Да	Да	Да	Да	Да
17	Onion Spectrum (OnionS)	Да	Нет	Да	Нет	Да
18	dk-series	Да	Нет	Да	Нет	Да
19	λ-distances	Да	Да	Да	Нет	Да
20	Ipsen–Mikhailov (IM) distance	Да	Да	Да	Нет	Да
21	Hamming–Ipsen–Mikhailov (HIM) distance	Да	Да	Да	Нет	Да
22	NetLSD measure	Да	Да	Да	Нет	Да
23	Non-Backtracking Spectral Distance (NBD)	Да	Да	Да	Нет	Да
24	Distributional Non-Backtracking Spectral Distance (d-NBD)	Да	Да	Да	Нет	Да
25	Quantum Jensen-Shannon divergence (QJSD)	Да	Да	Да	Нет	Да
26	Signature similarity (SimHash)	Да	Да	Да	Да	±
27	D-measure	Да	Да	Да	Да	Да
28	Layer Difference (LD)	Да	Да	Да	Нет	Нет
29	SLRIC similarity	Да	Да	Да	Да	Нет

Сравнение приведенных моделей осуществлено в несколько этапов. Во-первых, осуществлено сравнение моделей на простых графовых структурах. На втором этапе сравнение моделей оценки схожести графов проводится на случайных графах. Отметим, что все вычислительные эксперименты проведены на языке программирования Python, где доступна программная реализация большинства исследуемых моделей [190].

1) Сравнение моделей схожести сетевых структур на простых графах.

Рассмотрим несколько простых сетевых структур для понимания свойств исследуемых моделей. Для этого мы рассмотрим графы, приведенные в работах [184, 185] (см. рисунок 4.1). Отметим, что все рассматриваемые графы являются ненаправленными и невзвешенными, что позволяет производить сравнение всех исследуемых моделей.

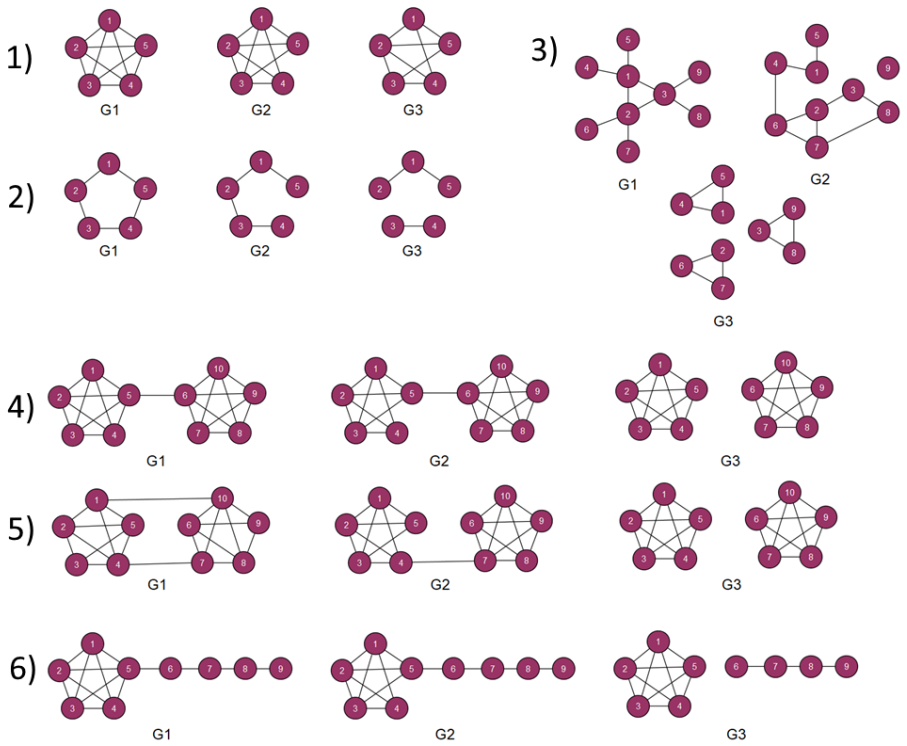


Рисунок 4.1 − Пример простых графовых структур

Интуитивно, графы G1 и G2 более похожи друг на друга, чем графы G1 и G3 для всех приведенных сетевых структурах (гипотеза A). Аналогичным образом можно предположить, что расстояние между графами G1 и G2 должно быть везде меньше, чем расстояние между графами G2 и G3 (гипотеза B). Произведем вычисление всех исследуемых моделей и проверим, согласуется ли выдвинутые гипотезы с результатами вычислений.

Таблица 4.2 − Сравнение моделей схожести на простых графах

Модель	Проверка гипотез на наборе графов											
	1		2		3		4		5		6	
	A	B	A	B	A	B	A	B	A	B	A	B
JI	+	+	+	+	–	–	–	+	–	–	–	+
GED	+	+	+	–	–	–	–	+	–	–	–	+
GED norm	+	+	+	+	–	–	–	+	–	–	–	+
VEO	+	+	+	+	–	–	–	+	–	–	–	+
1-hop NN	+	–	+	–	–	–	–	+	–	–	+	+
2-hop NN	–	–	+	+	+	+	+	+	+	+	+	+
3-hop NN	–	–	+	+	+	+	+	+	+	+	+	+
MCS	–	–	+	+	+	+	+	+	+	+	+	+
FRO	+	+	+	–	–	–	–	+	–	–	–	+
VS	+	+	+	+	–	+	+	+	–	–	+	+
deltacon	+	+	+	+	+	+	+	+	+	+	+	+
POL	+	–	+	–	–	–	–	+	–	–	+	+
GDD	–	+	+	+	+	+	+	–	–	+	+	+
RP	+	+										
VR	+	–	+	–	–	+	+	+	+	+	+	+
degreeJSD	+	–	+	–	+	–	+	+	+	+	–	–

Продолжение 4.2

POR	+	–	+	–	+	–	+	+	+	+	+	+
CSE	+	–	+	+	+	+	+	+	+	+	+	+
NetSimile	+	+	+	–	+	+	+	+	+	+	+	+
OnionS	–	–	–	–	–	–	–	–	+	–	–	+
dk-series	+	–	+	+	–	–	+	+	+	+	–	–
λ-d Adj.	+	–	+	–	+	+	+	+	+	+	–	+
λ-d Lap.	+	–	+	–	+	+	+	+	+	+	–	+
λ-d N.L.	+	–	+	–	+	+	+	+	+	+	–	+
Lap.JS	+	+	–	–	+	+	–	–	+	–	+	+
IM	+	+	+	+	+	+	+	+	+	+	+	+
HIM	+	+	+	–	+	–	+	+	–	–	–	+
NetLSD	+	+	+	+	+	–	+	+	+	+	+	+
NBD	+	–					+	+	+	–		
d-NBD	+	–			+	–	+	+	–	–	–	–
QJSD	+	+	+	–	–	–	–	+	–	–	+	+
SS	+	+	–	–	+	–	+	+	+	+	+	+
d-measure	+	–										
LD-measure	+	–										
SLRIC-sim	–	–	+	+	–	–	+	+	+	–	+	+

Результаты проведенного эксперимента приведены в Таблице 4.2. Пустые ячейки соответствуют случаям, когда гипотеза не была проверена из-за невозможности вычислить меру сходства графов с использованием существующих программных реализаций. Согласно полученным результатам в Таблице 4.2, только две модели схожести сетевых структур (deltacon, IM) подтвердили все выдвинутые гипотезы для приведенных графов. Модели CSE, NetSimile, NetLSD успешно подтвердили 11 из 12 гипотез. Отметим, что локальные меры, которые определяют схожесть графов только на основе сравнения прямых связей (JI, GED, VEO, FRO), не подтвердили истинность большинства гипотез. Наконец, набор графов №3 и №5 являются примерами, на которых большинство существующих моделей не подтвердили выдвинутые гипотезы.

2) Сравнение моделей схожести сетевых структур на случайных графах.

Произведем сравнение исследуемых моделей схожести на примере более сложных сетевых структур. Для этого сгенерируем 1000 случайных графов, состоящих из 100 вершин ($n = 100$), связи между которыми появляются с вероятностью $p = 2/(n-1)$. Далее для каждого рассматриваемого графа произведем последовательное добавление и удаление нескольких ребер (общее число таких итераций − 100). Таким образом, мы получим 1000 случайных графов, изменяющихся во времени.

Произведем сравнение исследуемых моделей на сгенерированных графах. Для оценки согласованности моделей схожести была использована мера корреляции Кендалла [191]. Результаты сравнений моделей схожести приведены на рисунке 4.2.

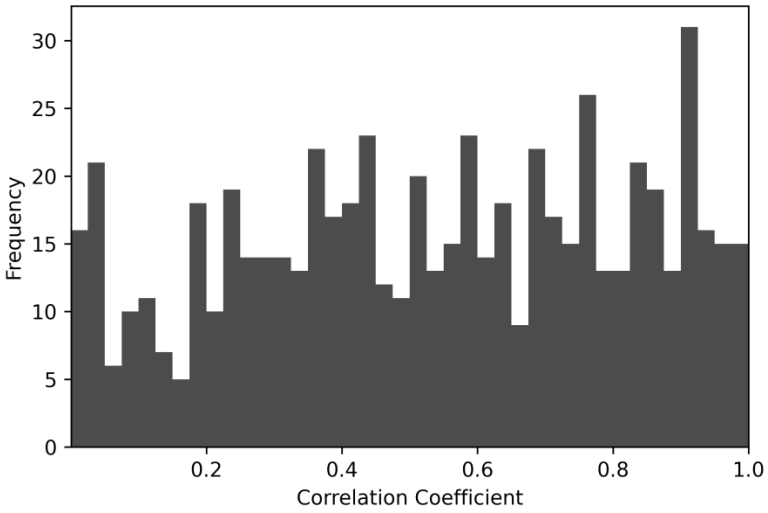


Рисунок 4.2 − Распределение коэффициента корреляции Кендалла между различными моделями схожести

В рамках проведенного эксперимента было выявлено, что среди 630 пар моделей схожести сетей, 144 пар показали высокий коэффициент корреляции (>0.8) на 1000 случайных графах. Таким образом, можно предположить, что хотя многие модели имеют совершенно различную интерпретацию понятия “схожести” сетевых структур, ряд моделей хорошо согласуются друг с другом с точки зрения итоговых результатов. Кроме того, было выявлено, что ряд простых моделей высоко коррелирует с некоторыми моделями, имеющими высокую вычислительную сложность.

4.2.3 Некоторые выводы

Оценка схожести графов является одной из важнейших задач в анализе сетевых структур, позволяющих изучать динамику сложных систем и выявлять возникающие аномалии в графах. В рамках данной работы было осуществлено детальное изучение 39 существующих моделей оценки схожести сетевых структур и проведено их сравнение на примере простых и сложных сетевых структур. Несмотря на то, что большинство моделей основано на совершенно различных подходах к оценке схожести графов, результаты проведенных экспериментов показали, что ряд моделей хорошо согласуется как друг с другом, так и с интуитивно понятными примерами. Отметим, что полученные результаты могут быть использованы как для выбора подходящей модели схожести, так и для разработки новых подходов к оценке различия между сетевыми структурами.

4.3 Сетевой анализ публикаций в области научных исследований болезни Паркинсона

4.3.1 Введение

Данная работа посвящена сетевому анализу публикаций в области исследований болезни Паркинсона. Методы сетевого анализа центральности позволяют оценить важность, как и отдельных работ и тем исследований, так и журналов, в которых они публикуются и авторов, которые их пишут.

Более 10 миллионов людей во всем мире болеют болезнью Паркинсона в настоящее время. Болезнь Паркинсона (БП) — это второе по распространенности после болезни Альцгеймера нейродегенеративное заболевание. Заболевание является прогрессирующим и в основном характерно для людей старшего возраста. Оно связано с уменьшением выработки дофамина, что приводит к гибели нейронов мозга в системах, отвечающих за двигательные функции человека, а в последствии и к дисфункции других систем мозга.

Основными симптомами болезни являются гипокинезия, тремор, мышечная ригидность, также постуральные нарушения — неустойчивость, сгорбленная поза, семенящая походка, шаркающие шаги. Нарушение сна, вегетативные и когнитивные нарушения относятся к немоторным симптомам.

Причины заболевания до сих пор неизвестны, несмотря на определенный прогресс в исследовании болезни. Ученые публикуют тысячи статей по данной тематике, поэтому им необходимо учитывать предыдущий опыт исследователей и понимать, какие направления являются перспективными в изучении болезни. Сетевой анализ цитирований позволяет выделить ключевые исследования, журналы, в которых они публикуются.

Основным инструментом сетевого анализа цитирований является сеть цитирований, являющаяся направленным графом, где вершины — статьи, а ребра — цитирования между ними. Такая же сеть цитирований была смоделирована и для журналов, в ней граф является взвешенным: вес ребра — количество цитирований одного журнала другим в определенный год.

Показателями важности элементов являются индексы центральности. В данной работе сравниваются четыре классических индекса (In-degree, Eigenvector, PageRank и Betweenness) и два новых (Bundle и Pivotal), введенных в [192]. Новые индексы позволяют выделять как группы влиятельных вершин, как для всего сообщества, так и в узкоспециализированных областях.

Целью данной работы является проведение сетевого анализа публикаций и журналов в области научных исследований БП.

4.3.2 Сетевой анализ

4.3.2.1 Подготовка данных

Для последующего анализа были выбраны статьи с 2015 по 2022 год из системы Microsoft Academic. Условие попадания статьи в выборку: слова “parkinson” и “disease” содержатся в названии и аннотации. Было скачано 70119 статей. Из первоначального набора данных удалены публикации, которые не имели DOI, аннотацию или информацию о журнале. В итоговой выборке получилось 45940 статей. Распределение по годам представлено на рисунке 4.3.

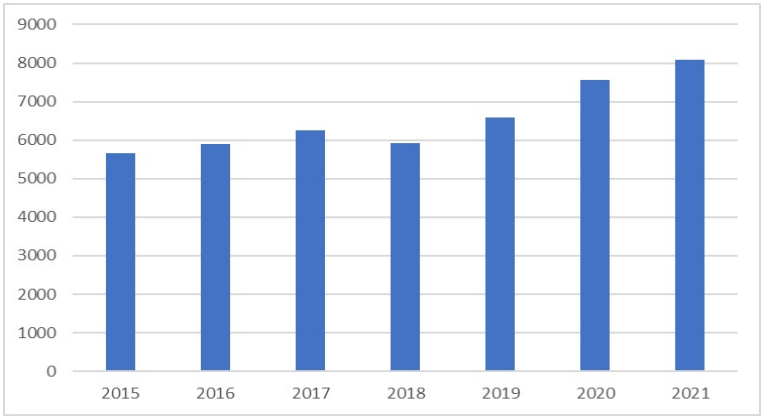


Рисунок 4.3 − Распределение статей с DOI, аннотацией и журналом по годам

4.3.2.2 Построение сети цитирований по статьям

Сеть цитирований строится на основе статей, которые содержатся в итоговом наборе данных. Она представляет собой ориентированный граф. Ребро от статьи ID1 к статье ID2 появляется в случае, когда статья ID1 цитирует статью ID2. Изолированные вершины в сеть не включаются. В данной сети вес дуг равен 1, он указан в колонке ‘weight’. В колонке ‘Y’ указан год ссылающейся статьи. Ребра, ведущие на статьи не из исходного списка удалены. Также удалены статьи, относящиеся к синдрому Вольфа-Паркинсона-Вайта (содержат слова ‘wolff’ и ‘parkinson’, ‘wpw’, ‘wpws’ или ‘danon’), и ребра из прошлого в будущее, которые являлись ошибочными в базе данных (год ссылающийся статьи меньше года статьи, на которую она ссылается). На рисунке 4.4 представлен фрагмент сети цитирований. В сети получилось 39811 вершин и 310829 ребер.

ID1	ID2	weight	Y
2754967293	2339791932	1	2017
2112455323	1920030402	1	2015
2177834950	2123627348	1	2015
2558041282	1934236512	1	2016
2558041282	1947901277	1	2016
2558041282	1564387586	1	2016
2558041282	1903888485	1	2016

Рисунок 4.4 − Фрагмент сети цитирований статей

4.3.2.3 Построение сети цитирований по журналам

Далее построена сеть цитирований по журналам. Ребро появляется в случае цитирования статьи, опубликованной в журнале JId1, статьи из журнала JId2. Вес ребра — это количество ссылок за определенный год. Если статья из JId ссылается на другую статью из этого же журнала, то получается самоцитирование. На рисунке 4.5 представлен фрагмент сети цитирований по журналам.

Атрибуты сети цитирований по журналам:

− JId1 — Id ссылающегося журнала;

− JId2 — Id цитируемого журнала;

− Y — год публикации из журнала JId1;

− weight — вес ребра, то есть количество цитирований журнала JId1 журналом JId2 в год Y.

AfId1	AfId2	Y	weight
4605	138006243	2017	1
9507	9507	2021	1
9507	35928602	2021	2
9507	40120149	2021	1
9507	177639307	2021	1

Рисунок 4.5 − Фрагмент сети цитирований журналов

4.3.2.4 Результаты сетевого анализа по статьям

На рисунке 4.6 представлено распределение количества цитирований статей за весь рассматриваемый период. Большая часть публикаций цитировалась менее 20 раз. 9626 статей не цитировались совсем, остальные имели от 1 до 500 цитирований, максимальное число цитирований 1563 у одной статьи.

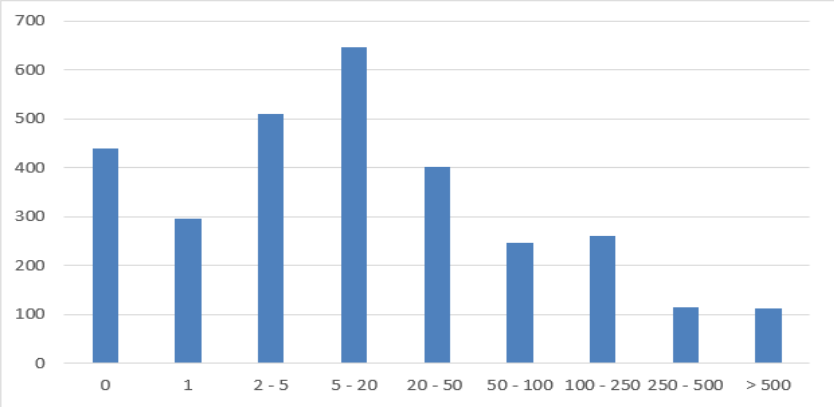


Рисунок 4.6 – Распределение количества цитирований статей

Классические и новые индексы центральности были рассчитаны по сети публикаций. Для новых индексов [192] максимальный размер критического множества был выбран равным 3, то есть 3 вершины могут одновременно влиять на другую вершину. Для новых индексов рассмотрены следующие значения квоты в процентах от общего числа ссылок на каждую статью:

– 0.1% ссылок;

– 0.5% ссылок;

– 1% ссылок.

Большие значения квот не приведены, потому что вес ребер равен 1 и сумма весов входящих ребер из критического множества не может превысить квоту.

В Таблице 4.3 представлено топ-10 статей по индексу In-degree. Это самые цитируемые статьи. Среди них первое и четвертое место занимают диагностические критерии, № 2, 8, 10 являются статьями по биохимическим и генетическим исследованиям, остальные содержатся статьи по ранней диагностике и статистическому анализу.

Таблица 4.3 – Топ-10 статей по индексу In-degree

Название	DOI	In-degree index	Eigenvector ранг	PageRank ранг
1. MDS clinical diagnostic criteria for Parkinson's disease	10.1002/MDS.26424	0.005	15	1
2. Gut Microbiota Regulate Motor Deficits and Neuroinflammation in a Model of Parkinson’s Disease	10.1016/J.CELL.2016.11.018	0.0016	62	8
3. Epidemiology of Parkinson's disease	10.1007/S00702-017-1686-Y	0.00156	1224	5
4. MDS research criteria for prodromal Parkinson's disease	10.1002/MDS.26431	0.00152	13	3

Продолжение таблицы 4.3

5. The epidemiology of Parkinson's disease: risk factors and prevention	10.1016/S1474-4422(16)30230-7	0.0015	705	4
6. Non-motor features of Parkinson disease	10.1038/NRN.2017.62	0.00148	764	14
7. Global, regional, and national burden of Parkinson's disease, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016	10.1016/S1474-4422(18)30295-3	0.00148	3585	7
8. The Roles of PINK1, Parkin and Mitochondrial Fidelity in Parkinson's Disease	10.1016/J.NEURON.2014.12.007	0.00147	4	2
9. α-Synuclein strains cause distinct synucleinopathies after local and systemic administration	10.1038/NATURE14547	0.0014	1	6
10. Gut microbiota are related to Parkinson's disease and clinical phenotype	10.1002/MDS.26069	0.0013	2	10

В Таблице 4.4 представлены статьи по Total индексу (TI) с квотой q = 0.1%, который является линейной комбинацией индексов In-degree, Bundle (BI), Pivotal (PI) с равными коэффициентами. Видно, что ранжированный список статей повторяется со списком по In-degree индексу, потому что квота достаточно мала.

Таблица 4.4 − Топ-10 статей по TI, q = 0.1%

Название	DOI	In-degree	BI, q=0,1%	PI, q=0,1%	TI, q=0,1%
1. MDS clinical diagnostic criteria for Parkinson's disease	10.1002/MDS.26424	0,005	0,6307	0,9404	0,5254
2. Gut Microbiota Regulate Motor Deficits and Neuroinflammation in a Model of Parkinson’s Disease	10.1016/J.CELL.2016.11.018	0,0016	0,0206	0,000096	0,0075
3. Epidemiology of Parkinson's disease	10.1007/S00702-017-1686-Y	0,00156	0,0188	0,000093	0,0068
4. MDS research criteria for prodromal Parkinson's disease	10.1002/MDS.26431	0,00152	0,0175	0,000091	0,0064
5. The epidemiology of Parkinson's disease: risk factors and prevention	10.1016/S1474-4422(16)30230-7	0,0015	0,01704	0,00009	0,0062
6. Non-motor features of Parkinson disease	10.1038/NRN.2017.62	0,00148	0,016	0,0000884	0,00584
7. Global, regional, and national burden of Parkinson's disease, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016	10.1016/S1474-4422(18)30295-3	0,00148	0,016	0,0000884	0,00584

Продолжение таблицы 4.4

8. The Roles of PINK1, Parkin and Mitochondrial Fidelity in Parkinson's Disease	10.1016/J.NEURON.2014.12.007	0,00147	0,0159	0,0000882	0,00581
9. α±-Synuclein strains cause distinct synucleinopathies after local and systemic administration	10.1038/NATURE14547	0,0014	0,014	0,000085	0,0052
10. Gut microbiota are related to Parkinson's disease and clinical phenotype	10.1002/MDS.26069	0,0013	0,0119	0,00008	0,0044

В Таблице 4.5 представлено топ-10 статей по TI q = 1%. Список статей уже совершенно другой, нет самых цитируемых статей. Статьи № 1, 2, 4, 6, 9, 10 относятся к ранней диагностике и клиническим исследованиям. Остальные являются биохимическими исследованиями.

Таблица 4.5 − Топ-10 статей по TI, q = 1%

Название	DOI	In-degree	BI, q=1%	PI, q=1%	TI, q=1%
1. Vagotomy and subsequent risk of Parkinson's disease	10.1002/ANA.24448	0,0009	0,0204	0,068	0,0298
2. The clinical symptoms of Parkinson's disease	10.1111/JNC.13691	0,00089	0,0202	0,0673	0,0295
3. Oxidative stress and Parkinson’s disease	10.3389/FNANA.2015.00091	0,00087	0,0183	0,061	0,0267
4. Neuroinflammation in Parkinson’s disease and its potential as therapeutic target	10.1186/S40035-015-0042-0	0,00086	0,0177	0,0589	0,0258
5. Short chain fatty acids and gut microbiota differ between patients with Parkinson's disease and age-matched controls	10.1016/J.PARKRELDIS.2016.08.019	0,00082	0,0156	0,052	0,0228
6. The Onset of Nonmotor Symptoms in Parkinson's disease (The ONSET PD Study)	10.1002/MDS.26077	0,000814	0,015	0,0501	0,022
7. Transneuronal Propagation of Pathologic α-Synuclein from the Gut to the Brain Models Parkinson's Disease	10.1016/J.NEURON.2019.05.035	0,000811	0,0149	0,0496	0,0217
8. T cells from patients with Parkinson’s disease recognize α-synuclein peptides	10.1038/NATURE22815	0,00079	0,0135	0,04497	0,0197

Продолжение таблицы 4.5

9. Prediagnostic presentations of Parkinson's disease in primary care: a case-control study	10.1016/S1474-4422(14)70287-X	0,00078	0,0132	0,0439	0,0193
10. New Clinical Subtypes of Parkinson Disease and Their Longitudinal Progression: A Prospective Cohort Comparison With Other Phenotypes	10.1001/JAMANEUROL.2015.0703	0,00077	0,012676	0,0422	0,0186

Сети цитирований статей и журналов и значения всех индексов представлены на странице https://www.hse.ru/DeCAn/parkinson.

4.3.3 Некоторые выводы

Сетевой анализ публикаций позволяет выделять ключевые исследования, относящиеся к болезни Паркинсона, с помощью классических и новых индексов центральности. Были подсчитаны новые индексы Pivotal Index и Bundle Index для почти 40 тысяч публикаций, журналов, в которых они были опубликованы, их авторов и аффилиаций. С помощью новых индексов появляется возможность учитывать групповое влияние и выделять группы ссылающихся друг на друга исследователей. Также анализ центральности позволяет определить главные тенденции в научной области. Например, для исследований о болезни Паркинсона в 2017–2019 годах самыми популярными являлись биохимические статьи, а к 2021 большое влияние получили статистические исследования.

Также это подход может быть применен и к другим научным областям. Это позволит ученым и исследователям выделять популярные области и оценивать научный вклад предыдущих исследований.

4.4 Сценарный анализ ущерба экологии от разлива нефти в Арктическом регионе на основе имитационной сетевой модели

4.4.1 Введение

В последнее время в Арктическом регионе активно развивается добыча и последующая транспортировка нефти. Кроме того, Северный морской путь также активно используется в международной торговле. Несмотря на некоторое снижение транзитных перевозок по этой транспортной артерии, общий грузооборот всё равно растет.

Например, интенсивность транспортировки нефти через Печорское море растет из-за увеличения добычи на Приразломном и Новопортовском месторождениях. Также в период до 2030 года запланированы геологоразведочные работы на ещё нескольких участках. В ре-

зультате ожидается увеличение объемов транспортировки нефти через этот регион в 2–4 раза (более 60 миллионов тонн в год).

Вместе с увеличением добычи и перевозки нефти, а также других транспортных потоков растут риски технологических катастроф, связанных с разливами вредных веществ. В свою очередь, загрязнение моря и воздуха наносит ущерб фауне во всех регионах [193, 194], особенно хрупким арктическим экосистемам, где ценность каждого вида флоры и фауны возрастает в условиях низкого разнообразия по сравнению с южными широтами. При этом важно учесть, что арктические экосистемы характеризуются низкой способностью к самовосстановлению и самоочищению, что делает их еще более уязвимыми к загрязнению.

Однако действующее российское законодательство не обязывает компании, ответственные за разлив нефти, принимать меры по спасению животных.

Например, в 2020 году в Норильске произошла страшная техногенная катастрофа. Из-за проседания свай фундамента в одной из цистерн произошло разгерметизирование, и дизельное топливо вылилось наружу. В общей сложности было разлито около 21 000 тонн нефтепродуктов, из которых 6000 тонн попали в землю, остальное — в реку Амбарная и ее приток Далдыкан, после чего нефть утекла через реку Пясина в Карское море.

Общая площадь загрязнения составила 180 000 квадратных метров. Предельно допустимые концентрации загрязняющих веществ в реках, куда попало дизельное топливо, были превышены в десятки тысяч раз.

Учитывая этот фактор, управляющим организациям важно хотя бы оценивать возможные риски от потенциальных катаклизмов, чтобы заранее обнаруживать наиболее опасные участки.

На данный момент существует два основных подхода к моделированию разливов нефти. Первый основан на системе дифференциальных уравнений. Основными уравнениями для этого подхода являются уравнения Навье − Стокса, а также уравнения поверхностной динамики нефти. Основным преимуществом является точность моделирования, поскольку большинство специфических факторов распространения нефти может быть учтено путем добавления новых уравнений: адвекция, волны на поверхности моря [195], дисперсия нефти [196] и многое другое.

Альтернативный подход к моделированию разлива нефти основан на идее сетевого моделирования, которая позволяет получить менее точные результаты, но с меньшими вычислительными затратами. Учитывая тот факт, что для правильного управления операциями нам нужно знать только основные направления и приблизительную скорость распространения разлива нефти, такой подход становится более привлекательным для этой цели.

Одно из первых применений этой идеи для моделирования разливов нефти было предложено Жиляковой [197]. В этой работе было предложено разделить морскую поверхность на маленькие секторы, так называемые элементарные территории одинакового размера. После этого все пары смежных территорий были соединены ребрами, чтобы построить сеть возможного распространения нефти по поверхности моря.

После построения сети происходит моделирование распространения нефтяного пятна путем многократного переноса нефти из загрязненных секторов на соседние территории.

4.4.2 Основная часть

В предложенной модели используется сетевая модель, как в работе [197]. При этом на каждой итерации для расчета объема переносимой на прилегающие территории нефти используется экспоненциальная функция со взвешенной суммой скорости ветра и морского течения в аргументе ($Q_{distributed}$ — объем перемещаемой нефти, Q_0 — количество нефти на загрязненной территории). Эта функция позволяет нам настраивать модель, изучая весовые коэффициенты для интенсивности ветра и морского потока (V_{wind} и V_{flow} соответственно). Кроме того, экспоненциальная функция гарантирует неотрицательное количество распределяемого масла

$$Q_{distiruted} = Q_0 e^{a+bV_{flow}+cV_{wind}}, \qquad a, b, c \in R. \tag{4.1}$$

Учитывая, что мы моделируем разлив нефти в Арктическом регионе, важно учитывать еще один критический фактор — наличие льда на поверхности моря. Лед может остановить или, по крайней мере, замедлить дальнейшее распространение разливов нефти. Поэтому было добавлено условие, которое блокирует распространение нефти по территории со льдом на ее поверхности:

$$Q_{distiruted} = \begin{cases} Q_0 e^{a+bV_{flow}+cV_{wind}}, & \text{без льда,} \\ 0, & \text{элементарная территория со льдом,} \end{cases} \quad a, b, c \in R. \tag{4.2}$$

Для апробации разработанной модели были смоделированы несколько вариантов потенциального разлива нефти в разных частях Арктики. Для всех расчетов были использованы карты течений и ветров, полученные в рамках Программы мониторинга и оценки Арктики (1997).

Для моделирования мы решили зафиксировать объём нефтяного разлива — 20 000 тонн. Это около 350 000 баррелей, что равно половине вместимости судна Aframax, являющегося одним из основных типов нефтяных танкеров.

Всего было рассмотрено три локации для потенциальной аварии: берега Шпицбергена, северо-восточное побережье Гренландии и Чукотское море вблизи острова Врангеля.

Кроме того, для оценки ущерба экологии прогнозировалось количество погибших животных из-за утечки нефти. В качестве примера использовались данные о среде обитания трех видов животных: белухи, нарвала и моржей.

В результате для каждой локации была получена модель разлива нефти во времени с прогнозом погибших животных в результате катастрофы. Например, для разлива нефти в Гренландском море к юго-западу от Шпицбергена были получены следующие результаты (рисунок 4.7; погибнет 1000 белух, 23 000 нарвалов и 4000 моржей). Для более ясного и точного понимания ситуации на рисунках используется постепенное окрашивание загрязненной области в зависимости от количества нефти — более темный оттенок означает большее количество масла.

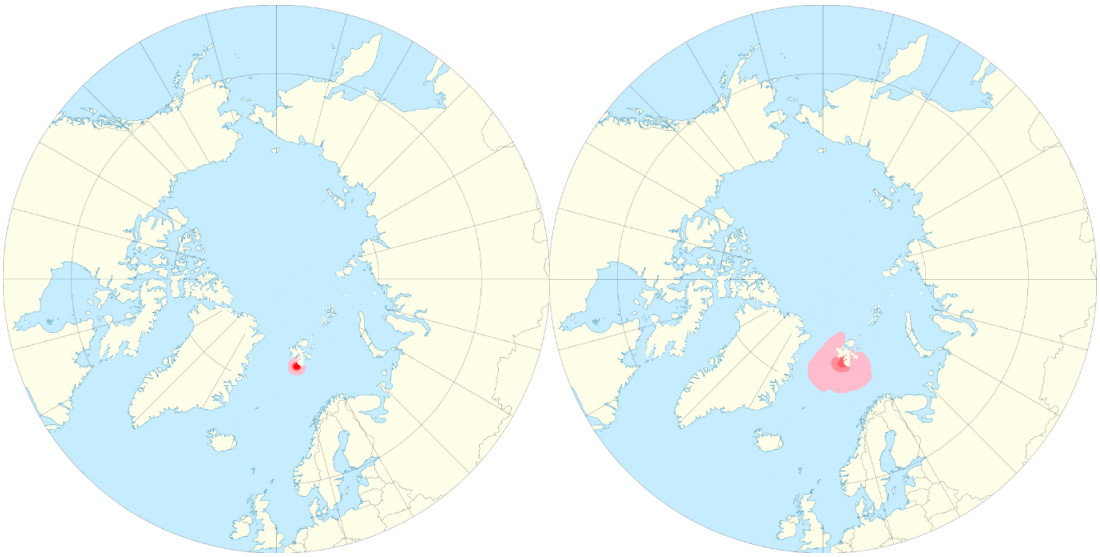


Рисунок 4.7 − Разлив нефти у берегов Шпицбергена

4.4.3 Некоторые выводы

По итогам работы можно сказать, что поставленные задачи выполнены в полном объёме. Во-первых, была улучшена модель разлива нефти. При этом сохранились основные преимущества модели — гибкость. Сетевая структура моря может быть инициализирована поразному, и поэтому может моделировать как базовые сценарии катаклизма (локальный одномоментный разлив или продолжающийся в течение определенного периода времени прокол в нефтепроводе), так и различные комбинации базовых сценариев [198].

Во-вторых, был добавлен автоматический расчет ущерба хрупкой экосистеме Арктического региона путем расчета приблизительного количества пострадавших животных.

5. Анализ социально-экономических и наукометрических данных

В этом разделе отчета приведены результаты по анализу социально-экономических и наукометрических данных. Исследования осуществлялись в следующих направлениях:

− анализ цитируемости и библиометрических показателей международных бизнес-журналов;

− обработка зеркальных данных торговли продовольственными товарами;

− динамический анализ структурной близости регионов;

− анализ предпочтений студентов относительно вакцинации.

5.1 Анализ цитируемости и библиометрических показателей международных бизнес-журналов

5.1.1 Введение

В настоящее время базы данных цитирования научных журналов и публикаций очень важны, поскольку они размещают информацию, которая служит для оценки ценности и влияния журнала, например, различные наукометрические и библиометрические показатели, используемые в управлении наукой. Такие базы данных также полезны для исследователей, которые изучают и составляют различные рейтинги журналов, анализируют цитируемость и структуру различных научных областей [199]. По мере того, как роль различных метрик — индекса Хирша, импакт-фактора и других – растет, все больше и больше людей в этой области используют их для манипулирования данными в своих собственных целях [200, 201]. Это даже привело к появлению нового термина – хищнические журналы [202] и, как следствие, к расхождениям между экспертными и наукометрическими рейтингами журналов [203, 204]. Хищнические журналы считаются очень плохим явлением, поскольку они принимают статьи к публикации без проведения обещанных проверок качества на предмет таких проблем, как плагиат или этическое одобрение [205].

Мы решили провести исследование международных бизнес-журналов, представленных в базе данных Scopus [206]. В прошлом году мы провели аналогичное исследование российских экономических журналах и опубликовали статью на эту тему [207]. Эта работа является логическим продолжением предыдущей. Наша цель — изучить журналы из базы данных Scopus, их цитируемость и библиометрические показатели, такие как CiteScore, SNIP и SJR, которые являются показателями базы данных Scopus.

Целью нашего исследования является проведение анализа научных журналов из базы данных Scopus [206], которые связаны с предметной областью бизнеса, менеджмента и бухгалтерского учета (Business, Management and Accounting), их цитированием и показателями.

5.1.2 Основная часть

5.1.2.1 Предобработка данных

Сначала мы загрузили и проанализировали все журналы, относящиеся к теме бизнеса, менеджмента и бухгалтерского учета. Их количество составило 1758 журналов. Однако, данные требовали предварительной обработки. В массиве данных было 63 дубликата — журналы с одинаковым названием и издателем. Мы удалили все дубликаты, и количество журналов уменьшилось до 1695.

Затем мы удалили журналы, которые имели пустые значения одновременно в следующих столбцах: «CiteScore», «2017–2020 Citations», «2017–2020 Documents», «% citation», «SNIP», «SJR», – потому что в таких журналах нет информации для анализа. Мы оставили журналы, в которых было всего несколько пустых значений, и, конечно же, те, в которых были нули. Итого, получилось 284 журнала с пустыми значениями, которые мы удалили. Все 284 журнала представлены на сайте Международного центра анализа и выбора решений [208].

Наконец, у нас получился набор данных с 1411 журналами без дубликатов и без журналов с большим количеством пропущенных значений, которые мы теперь анализировали далее.

Также стоит отметить, что полученные данные подходят только для библиометрического анализа. Для сетевого анализа, над которым сейчас проводится работа, мы загрузили данные из базы Microsoft Academic Graph [209], так как взять необходимые данные по кроссцитирования из базы Scopus уже невозможно из-за ограничения доступа. В базе данных Microsoft Academic Graph [209] есть не все из проанализированных нами журналов из базы Scopus [206], поэтому мы будем анализировать 850 совпадающих журналов. Возможно, мы расширим анализ журналами, которые есть только в базе Microsoft Academic Graph [209], но тоже относятся к теме бизнесе.

5.1.2.2 Описание данных

Окончательный набор данных представлен в файле “journals_data.xslx” на сайте Международного центра анализа и выбора решений [208]. Он содержит все 1411 журналов с их названиями и показателями, в частности «CiteScore», «2017–2020 Citations», «2017–2020 Documents», «% citation», «SNIP», «SJR» и другие параметры, в частности «Publisher», «Percentile», «Position», «Category». Давайте теперь опишем каждый показатель. Показатели «2017–2020 Citations» и «2017–2020 Documents» подсчитывают цитаты и документы [210] журнала за период 2017–2020 годов соответственно. CiteScore, SNIP, SJR — это индексы цитируемости журнала, созданные Scopus [211, 212, 213].

Параметры «Percentile» и «Position» связаны с параметром «Category», который является категорией журнала. Эти параметры рассчитываются путем сравнения журналов в одной категории. Перцентиль показывает процент журналов, которые находятся ниже в категории, чем выбранный журнал, например, если у журнала перцентиль равен 89%, это означает, что журнал занимает более высокое положение, чем 89% журналов в категории. Позиция показывает место журнала относительно других журналов в категории, например, если позиция журнала равна 5/325, это означает, что журнал является 5-м из 325 журналов, имеющих ту же категорию. Столбцы «Journal's name» и «Publisher» — это полные названия журнала и издателя журнала соответственно.

Список всех издателей и количества журналов, которые они выпускают, представлены на листе «Publishers» в файле на сайте [208]. Общее число издателей составляет 385. Максимальное количество журналов для издателя составляет 183. Только 5 издательств выпускают более 100 различных серий, и это следующие: Emerald, Taylor & Francis, Springer Nature, Wiley-Blackwell и Elsevier. Среднее количество журналов, которые выпускает один издатель, составляет примерно 366 журналов, что означает, что цифры, близкие к 100, являются экстраординарными. Более того, 368 издательств, что составляет 95,58% от всех издателей, выпускают 5 или менее серий, а 323 издательства (84%) — только 1 журнал.

Также мы составили список всех категорий журналов с указанием количества журналов в каждой категории. Список представлен на листе «Categories» в файле на сайте [208]. Общее количество категорий составляет 96. Самая популярная категория — Бизнес и международный менеджмент (Business and International Management), в ней представлено 128 журналов. Среднее количество журналов в категории составляет 14,55. Существует 71 категория, в которой журналов меньше, чем среднее значение, 60 категорий с 5 журналами или меньше и 30 категорий только с 1 журналом.

5.1.2.3 Анализ данных

После обработки и изучения данных, мы приступили к их анализу. CiteScore – это основной индекс Scopus. Топ-5 журналов, оцениваемых по этому показателю, представлены в Таблице 5.1. Лидеры по индексу CiteScore также имеют остальные показатели на высоком уровне, и все они входят в топ-1 в своей категории. Есть 35 журналов, у которых CiteScore равен 0, и 336 журналов, у которых CiteScore меньше 1. Средний показатель цитируемости составляет 2,96.

Таблица 5.1 − Топ-5 по CiteScore

Journal's name	Cite Score	2017-20 Citations	2017-20 Documents	% citation	SNIP	SJR	Publisher	Position	Category
Academy of Management Annals	24.7	2620	106	89	9.32	18.32	Academy of Management	1/399	Business and Int. Management
Annual Review of Organiz. Psychology and Organiz. Behavior	22.3	1720	77	96	8.89	9.807	Annual Reviews Inc.	1/227	Applied Psychology
Journal of Management	21.4	7935	371	96	5.76	7.491	SAGE	1/440	Strategy and Management
Int. Journal of Management Reviews	20.8	2327	112	92	5.55	4.475	Wiley-Blackwell	1/248	Management of Technology and Innovation
Int. Journal of Information Management	18.1	10563	583	93	4.83	2.77	Elsevier	1/235	Library and Information Sciences

Топ-5 журналов по SNIP представлены в Таблице 5.2. Первые два журнала одинаковы, журнал, занявший третье место по версии CiteScore, сейчас находится на пятом месте. Есть два «новых» журнала: журнал, занявший третье место по SNIP, был девятым по CiteScore, а тот, что сейчас на четвертом месте, раньше занимал тридцать восьмое место. Все топ-5 журналов по версии SNIP являются лидерами в своих категориях.

Таблица 5.2 − Топ-5 по SNIP

Journal's name	Cite Score	2017-20 Citations	2017-20 Documents	% citation	SNIP	SJR	Publisher	Position	Category
Academy of Management Annals	24.7	2620	106	89	9.32	18.32	Academy of Management	1/399	Business and International Management
Annual Rev. of Organiz. Psychol. and Organiz. Behav.	22.3	1720	77	96	8.89	9.807	Annual Reviews Inc.	1/227	Applied Psychology
Brookings Papers on Econ. Activity	15.6	640	41	88	7.67	7.134	Brookings Institution Press	1/218	General Business, Management and Account.
Journal of Finance	11.2	3114	278	86	6.93	18.15	Wiley-Blackwell	1/155	Accounting

Продолжение таблицы 5.2

Journal of Management	21.4	7935	371	96	5.76	7.491	SAGE	1/440	Strategy and Management

По индексу SJR в топ-5 вошли: «Academy of Management Annals» (1-е место по рейтингу CiteScore), «Journal of Finance» (38-е место по рейтингу CiteScore), «Review of Financial Studies» (67-е место по рейтингу CiteScore), «Journal of Financial Economics» (56-е место по рейтингу CiteScore), «Academy of Management Journal» (12-е место по версии CiteScore). Журнал, который раньше был на втором месте как по версии CiteScore, так и по версии SNIP — «Annual Review of Organizational Psychology and Organizational Behavior» — теперь находится на седьмом месте. Список этих журналов представлен в Таблице 5.3. Средний SJR составляет 0,83. Нет журналов с SJR, равным 0, но есть пять журналов с SJR, равным 0,1, и 49 журналов, в которых отсутствует SJR.

Таблица 5.3 − Топ-5 по SJR

Journal's name	Cite Score	2017-20 Citations	2017-20 Documents	% citation	SNIP	SJR	Publisher	Position	Category
Academy of Management Annals	24.7	2620	106	89	9.32	18.32	Academy of Management	1/399	Business and Intern. Management
Journal of Finance	11.2	3114	278	86	6.93	18.15	Wiley-Blackwell	1/155	Accounting
Review of Financial Studies	9.2	4214	459	84	5.07	12.8	Oxford Univ.Press	3/155	Accounting
Journal of Financial Economics	9.6	4957	518	83	5.66	11.67	Elsevier	2/155	Accounting
Academy of Management Journal	14.2	4629	327	92	4.23	11.19	Academy of Management	2/218	General Business, Management and Account.

Давайте теперь сравним рейтинги по этим трем показателям с рейтингами по количеству цитирований и документов. Максимальное количество документов в журнале составляет 15505, и это «Journal of Cleaner Production», издаваемый Elsevier. Этот журнал также входит в топ-1 по цитируемости с числом цитирований, равным 203300. Вторым журналом по цитируемости является «Knowledge-Based Systems» от Elsevier, и количество его цитирований значительно ниже — 21478, что почти в десять раз меньше, чем количество цитирований первого журнала. Среднее количество цитирований составляет 802,86. Есть 22 журнала, которые не имеют цитирования, и 95 журналов, которые имеют 10 цитирований или меньше.

Топ-5 журналов по цитируемости полностью отличаются от любого из топ-5 по показателям. Он представлен в Таблице 5.4. Стоит отметить, что другие параметры этих журналов по-прежнему высоки, но не самые высокие. Как мы можем видеть, четыре из пяти журналов имеют Elsevier, которому принадлежит Scopus, в качестве издателя, и Elsevier является одним из издателей, который выпускает более 100 различных серий.

Таблица 5.4 − Топ-5 по цитированиям

Journal's name	Cite Score	2017-20 Citations	2017-20 Documents	% citation	SNIP	SJR	Publisher	Position	Category
Journal of Cleaner Production	13.1	203300	15505	89	2.48	1.937	Elsevier	7/440	Strategy and Management
Knowledge-Based Systems	11.3	21478	1894	82	2.89	1.587	Elsevier	6/114	Management Inform. Systems
Journal of Business Research	9.2	18769	2036	84	2.85	2.049	Elsevier	14/185	Marketing
Technol. Forecasting and Social Change	12.1	17797	1472	85	3.04	2.226	Elsevier	11/399	Business and Intern. Management
Intern. Journal of Production Research	10.8	17590	1636	90	2.4	1.909	Taylor & Francis	5/166	Management Sc. and Operations Research

Вероятно, существует сильная зависимость между количеством цитирований и количеством документов, как это следует из сравнения информации во всех таблицах. Давайте проверим это. Топ-5 журналов по количеству документов представлены в Таблице 5.5. Они почти совпадают с журналами топ-5 по количеству цитирований, четыре из пяти журналов снова публикуются Elsevier.

Таблица 5.5 − Топ-5 по документам

Journal's name	Cite Score	2017-20 Citations	2017-20 Documents	% citation	SNIP	SJR	Publisher	Position	Category
Journal of Cleaner Production	13.1	203300	15505	89	2.48	1.937	Elsevier	7/440	Strategy and Management

Продолжение таблицы 5.5

Eastern-European Journal of Enterprise Technologies	2.2	4756	2157	52	0.67	0.268	PC Technology Center	237/548	Applied Mathematics
Journal of Business Research	9.2	18769	2036	84	2.85	2.049	Elsevier	14/185	Marketing
Knowledge-Based Systems	11.3	21478	1894	82	2.89	1.587	Elsevier	6/114	Management Information Systems
Journal of Manufacturing Processes	6.6	12301	1871	76	2.08	1.387	Elsevier	18/166	Management Science and Operations Research

Судя по топ-5 журналов по каждому параметру, между ними может быть корреляция. Кроме того, существует сильная зависимость между количеством цитирований и количеством документов. Мы измеряли корреляции Пирсона между показателями: CiteScore, SNIP, SJR, цитированиями на 2017–2020 годы и документами на 2017–2020 годы [214]. Результат представлен в Таблице 5.6.

Таблица 5.6 − Матрица корреляции Пирсона

	CiteScore	SNIP	SJR	2017-20 Citations	2017-20 Documents
CiteScore	1	0.899	0.738	0.254	0.22
SNIP	0.899	1	0.825	0.175	0.143
SJR	0.738	0.825	1	0.128	0.096
2017-20 Citations	0.254	0.175	0.128	1	0.825
2017-20 Documents	0.22	0.143	0.096	0.825	1

Таким образом, индикаторы CiteScore и SNIP больше похожи, чем любой из них и индикатор SJR. Кроме того, CiteScore коррелирует с количеством цитирований немного больше, чем другие показатели, хотя корреляция все еще низкая. Самая низкая корреляция с количеством цитирований наблюдается у SJR. Вероятно, это связано с тем, что метрика SJR является наиболее взвешенной. Кроме того, существует корреляция между количеством цитирований и количеством документов. Другие параметры не коррелируют с количеством документов.

5.1.2.4 Анализ топ-10 авторов

После анализа данных мы проанализировали топ-10 авторов в сфере бизнеса, которые написали наибольшее количество статей за выбранный период (2017–2020). Схема представ-

лена на рисунке 5.1. Чтобы проверить гипотезу о том, публикуются ли ведущие авторы в журналах с наибольшим количеством документов за период 2017–2020 годов, мы нашли список этих авторов. Затем мы просмотрели топ-10 статей каждого автора и проверили, в каких журналах они были опубликованы. Существует 38 таких журналов, и только 6 из них не включены в наш список бизнес-журналов. Этими 6 журналами являются «Journal of Petroleum Technology», "Drapers, Global Transitions», «Process Safety and Environmental Protection», «Computers in Industry», «Journal of Applied Psychology». Полный список этих журналов представлен в файле на сайте [208].

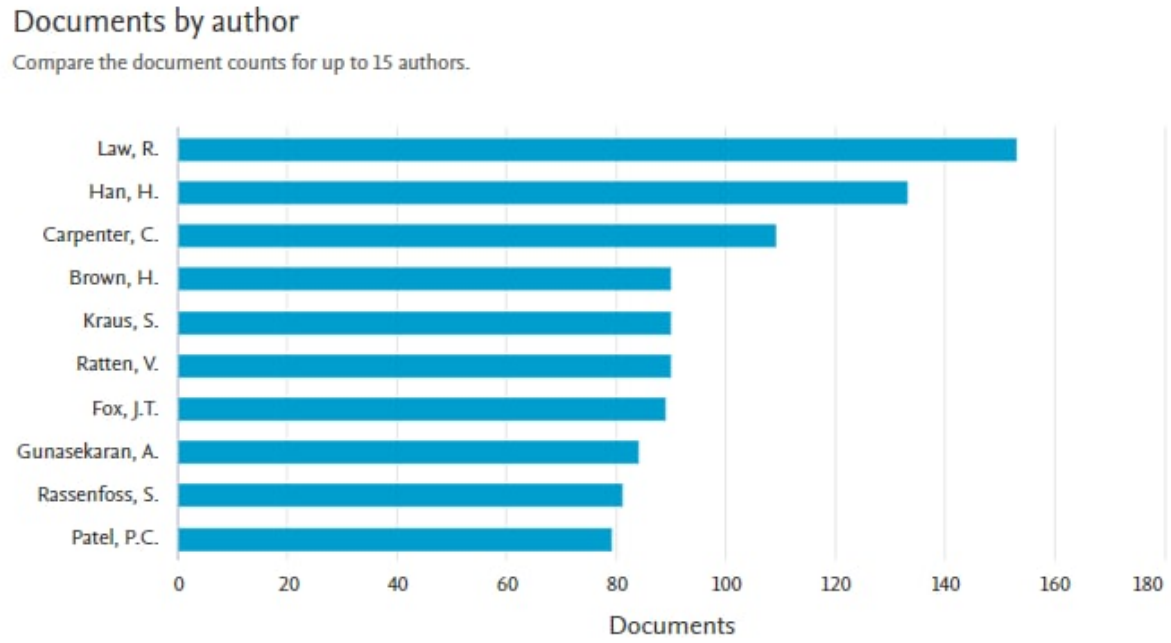


Рисунок 5.1 − Топ-10 авторов по документам

Мы хотели бы сравнить оставшиеся журналы с топ-10 журналов по количеству документов в нашем списке, который представлен в Таблице 5.5. Ни один из журналов с топ-10 авторами не входит в топ-10 по документам. Хотя некоторые из этих журналов входят в топ-10 по версии CiteScore: «Journal of Management» (3-е место), «International Journal of Management Reviews» (5-е место), «International Journal of Information Management» (6-е место), «Tourism Management» (10-е место). Это может указывать на то, что авторы, которые активно публикуются, с большей вероятностью будут опубликованы в журналах с высоким рейтингом.

5.1.2.5 Черные списки журналов

И последнее, но не менее важное: мы хотели бы написать о черных списках журналов и издателей и о том, почему исследователи включают их в такие списки. Термин «хищниче-

ская публикация» был впервые введен Джеффри Биллом [215]. Этот термин описывал публикации, которые были опубликованы в журналах без должной проверки. Появились издатели, которые выпускали журналы только для собственной материальной выгоды и не заботились о деловой, исследовательской и издательской этике — трех столпах научного издательства. Поэтому возникла необходимость в специальном черном списке с такими хищническими журналами и хищными издателями. Одним из самых известных авторов таких списков является сам Джеффри Билл. Он использовал несколько критериев для добавления журналов или издателей в список, это основные из них:

− два или более журналов имеют дублирующие редакционные коллегии, то есть существует одна и та же редакционная коллегия более чем для одного журнала;

− издатель ложно утверждает, что его контент индексируется в законных службах ссылок и индексации, или утверждает, что его контент индексируется на ресурсах, которые не являются службами ссылок и индексации.

Стоит отметить, что этот список подвергся критике за то, что он ориентирован только на издателей с открытым доступом [216]. Кроме того, хотя публикации, включенные в список, подвергаются критике и ограничиваются в цитировании в течение многих лет, доказательств проверки было мало. Например, издательство MDPI попало в список, составленный Джеффри Биллом, но позже это решение было успешно обжаловано, а издательство исключено из списка.

Крупные научные базы данных, такие как Scopus и Web of Science, обращают внимание на список Билла, но не сосредотачиваются на нем полностью. После того, как публикации Джеффри Билла прекратились, начали появляться новые списки с аналогичной идеей. Например, одним из списков, который заменил список Билла, был платный ресурс Cabell's Blacklist, содержащий более 4000 спорных и хищнических научных журналов.

Большинство списков включают только модель открытого доступа, в то время как модель издателей с платным доступом также вызывает вопросы. В статье [217] Roberto Dell'Annoa, Rocco Caferrab и Andrea Morone проверяют адекватность принятия статей в различные журналы с платным доступом. Они написали спорную статью и разослали ее издателям с разными моделями (открытый доступ и платный доступ). Целью их работы было понять, поможет ли платная публикация в экономическом журнале избежать попадания в хищнический журнал, или Джеффри Билл и его последователи незаслуженно отказались от этой модели. В результате авторы пришли к выводу, что традиционная модель, когда читатель платит за доступ, более устойчива к публикациям, которые не внушают доверия, поскольку они стремят-

ся к более точным статьям, тогда как журналы с моделью открытого доступа приняли их спорную статью более чем в половине журналов.

Университеты часто открыто или в частном порядке публикуют списки журналов, разрешенных к публикации, а также черные списки с потенциальным запретом. Однако журналы с хорошей репутацией не всегда попадают в белый список. Например, 84 журнала из списка «потенциальных, возможных или вероятных» хищнических журналов Билла были включены в список Комиссии по университетским грантам Индии [218, 219].

5.1.3 Некоторые выводы

Мы загрузили 1758 журналов предметной области Бизнес, менеджмент и бухгалтерский учет из базы данных Scopus, обработали их, и их количество сократилось до 1411.

Мы определили, что общее число издателей составляет 385. Максимальное количество журналов для издателя составляет 183, однако только 5 издательств выпускают более 100 различных серий, и это следующие: Emerald, Taylor & Francis, Springer Nature, Wiley-Blackwell и Elsevier. Более того, более 95% всех издательств выпускают 5 или менее серий, а 84% выпускают 1 журнал. Кроме того, почти у каждого журнала есть дополнительная категория, например, категория «Право», а самая популярная категория — «Бизнес и международный менеджмент», в которой 128 журналов, в то время как среднее количество журналов в категории составляет 14,55. Общее количество категорий составляет 96.

Мы изучили три библиометрических показателя — CiteScore, SNIP и SJR – и такие параметры, как количество цитирований и количество документов. Показатели имеют зависимости между собой, и индикатор CiteScore и индикатор SNIP коррелируют больше, чем любой из них и индикатор SJR. Мы не обнаружили каких-либо очевидных линейных зависимостей между каким-либо из показателей и количеством цитирований или количеством документов. Однако существует корреляция между количеством цитирований и количеством документов. Стоит отметить, что журналы, которые имеют наибольшее количество цитирований, не имеют самых высоких показателей.

Мы изучили, как работают библиометрические показатели Scopus, и можем утверждать, что они зависят не только от количества цитирований и могут быть подходом к оценке ценности журнала.

5.2 Обработка зеркальных данных торговли продовольственными товарами

5.2.1 Введение

Для изучения торговых отношений, определения наиболее влиятельных торговых партнёров и исследования последствий экономических шоков часто экспортно-импортные отношения между экономиками представляют в виде взвешенного направленного графа –

продовольственной сети. У такого графа веса рёбер характеризуют торговые потоки между странами. В зависимости от целей исследования у этих графов изучают их различные характеристики.

Торговые потоки можно характеризовать стоимостью товаров (например, в долларах США) или весом экспортируемых/импортируемых товаров (например, в тоннах или килограммах).

Страны ведут собственную статистику торговли товарами и услугами, однако сводная статистика, полученная от разных стран, может содержать значительные расхождения, которые необходимо учитывать при построении единой глобальной продовольственной сети.

5.2.2 Основная часть

5.2.2.1 Источник данных

Источником информации о мировой торговли может служить база данных Организации Объединенных Наций UN [220], которая ведётся сектором статистики торговли статистического отдела ООН. База данных UN Comtrade содержит детальную статистику импорта и экспорта различных товаров (как стоимости, так и веса) от статистических служб различных стран и территорий начиная с 1962 г.

В настоящей работе эта база данных используется в качестве основной. Для выгрузки данных использовался стандартный функционал веб-сайта UN Comtrade Database (согласно политике пользования базой данных всем пользователям веб-сайта предоставляется бесплатный неограниченный доступ к базе данных), при использовании данного способа накладывается техническое ограничение на одновременное включение в выборку большого числа записей в рамках одного запроса к базе данных; для обеспечения стабильной работы сервиса существует ограничение со стороны UN Comtrade Database в 100 запросов к базе данных за один час на один IP адрес. Для получения данных за последние четыре года (2018–2021 гг.) и раздельному включению отдельных категорий продовольственных товаров при формировании запросов изменялись соответствующие параметры выборки, запросы к базе данных выполнялись последовательно. В результате были получены файлы отдельно для разных категорий продовольственных товаров за 2018, 2019, 2020 и 2021 годы.

Следующей подзадачей стало объединение этих файлов в единую csv-таблицу и удаление неподходящих записей. В качестве торговых партнеров в базе данных содержатся объединения стран, например, Ассоциация государств Юго-Восточной Азии (ASEAN), которые необходимо было исключить из рассмотрения.

Для объединения всех файлов в четыре файла (согласно рассматриваемому году) была написана программа на языке программирования Python.

Была сделана выборка данных за последние пять лет, в ней представлена информация о стране, поставляющей определённый товар, и стране, в которую данный товар поставляется, а также суммарная стоимость товара в долларах США и название категории продовольственного товара. Каждая запись формирует направленное ребро графа, моделирующего продовольственную сеть. Все рассматриваемые страны составляют вершины графа. Всего в графе 201 вершина.

Использовалась международная стандартная торговая классификация четвёртой (последней) редакции (Standard International Trade Classification, SITC) для рассмотрения отдельных категорий товаров. Четвёртая редакция SITC часто используется для анализа движения продовольственных товаров и некоторых других категорий товаров (см. Таблицу 5.7). В настоящем исследовании обработку и корректировку зеркальных данных UN Comtrade опишем на примере экспортируемых и импортируемых товаров с кодами 041–042 согласно SITC. Эти категории включаются в более крупную категорию товаров: 04 “Зерновые и продукты из них”.

Таблица 5.7 − Фрагмент таблицы SITC

Код	Описание
041	Пшеница (включая полбу) и суржик, немолотые
042	Рис
043	Ячмень, немолотый
044	Кукуруза (в зерне), немолотая
045	Зерновые, немолотые (кроме пшеницы, риса, ячменя и кукурузы)

В идеальной системе торговли стоимость экспорта из страны А в страну Б будет равна стоимости импорта в Б из страны А. Применительно к выборке из базы данных для каждой пары стран А, Б имеем две записи, в одной графа “reporter” задана значением «А», а другой записи — «Б». Эти две записи в дальнейшем будем называть зеркальными данными о маршруте поставки товара. Однако на реальных данных легко заметить отличия в несколько раз для некоторых пар стран, поэтому при инициализации весов графов потребуется учитывать указанное несоответствие. Среди причин, которые приводят к искажению на 10% и более, можно отметить учёт страхования товаров и сопряжённых их транспортировке расходов. Более существенные отличия могут иметь место, если одна из экономик из пары допускает неполное следование всем рекомендациям ООН при предоставлении отчёта о торговли.

Для некоторых прикладных аналитических задач вместо стоимости экспорта/импорта зерновых продуктов правильней рассмотреть вес нетто всех экспортирован-

ных/импортированных товаров. База данных UN Comtrade содержит отдельный столбец, числа в котором обозначают вес нетто в килограммах. Однако некоторые страны публикуют данные только о стоимости экспорта/импорта товаров, в таком случае графа “Netweight, kg” остаётся пустой.

Для проведения корректировки зеркальных данных будем использовать алгоритм, предложенный в [221]. Этот алгоритм изначально разработан для корректировки зеркальных данных для стоимости товара, поэтому его необходимо адаптировать для работы с весом товара.

Опишем, какие результаты получены при следовании этому алгоритму с учётом использования веса товара, а не его стоимости.

5.2.2.2 Алгоритм корректировки

В 2006 году ООН провела опрос для выявления, какие экономики следуют рекомендациям организации при предоставлении данных. В данной работе будем считать, что если экономика в данном опросе имеет более 70% правильных ответов и менее 30% неправильных ответов, то её данным можно доверять. Таких стран всего 46 (были исключены страны, которые не предоставили статистику за 2018−2021 гг.): Антигуа и Барбуда, Австралия, Азербайджан, Бельгия, Боливия, Бразилия, Канада, Китай, Хорватия, Куба, Фиджи, Финляндия, Грузия, Германия, Греция, Венгрия, Ямайка, Казахстан, Кения, Кувейт, Латвия, Малайзия, Мальта, Маврикий, Монголия, Марокко, Намибия, Северная Македония, Норвегия, Польша, Республика Корея, Республика Молдова, Румыния, Российская Федерация, Сент-Винсент и Гренадины, Сенегал, Сербия, Словения, Испания, Швеция, Швейцария, Турция, Украина, Великобритания, Узбекистан и Вьетнам.

Дополнительно построим рейтинг для всех оставшихся стран, которые имеют ненулевой торговый оборот с доверенными странами.

Сами шаги алгоритма выполняются отдельно и независимо для разных категорий товаров и временных промежутков:

1) Сначала можно убрать из рассмотрения значения, не превышающие некоторого порога q от 10% усечённого среднего экспорта (импорта). Этот шаг может помочь уменьшить плотность получаемого графа, так как часть рёбер убирается из рассмотрения. В этой работе мы пропустили первый шаг, так как для некоторых небольших стран даже малые значения импорта могут стать критически важными.

2) Если зеркальных данных у партнера нет (или они были убраны из рассмотрения на шаге 1), то берём данные только от одного партнера и добавляем соответствующее ребро в граф.

3) Рассматриваем все такие пары, в которых обе экономики являются доверенными. Добавляем ребро с весом, соответствующим данным от экономики, предоставившей большее число правильных ответов в опросе ООН.

4) Рассматриваем все такие пары, в которых только одна экономика является доверенной. Берём данные от доверенной экономики.

5) Рассматриваем все пары, в которых ни одна экономика не является доверенной, но для всех из них определен рейтинг. Берём данные от экономики, у которой рейтинг выше.

6) Рассматриваем все пары, в которых ни одна экономика не является доверенной и не торгует с доверенными странами. Такие пары рассматриваются отдельно.

Когда в данных есть только стоимость экспорта/импорта, но нет данных о весе товара, мы применяем описанный алгоритм и для этих данных с некоторым фиктивным весом. Если после применения алгоритма фиктивное значение веса не было заменено значением веса от партнёра, то эта пара стран рассматривается отдельно. Все такие случаи и способы корректировки веса для них перечислены в Таблицах А.1, А.2 в Приложении А. Для расчёта примерной цены 1 кг пшеницы (1 кг риса) использовалось 15% усечённое среднее. На основании общей стоимости пшеницы (риса) и вычисленной средней цены определялся примерный вес экспортируемого/импортируемого товара.

Можно использовать данные, предоставляемые Продовольственной и сельскохозяйственной организацией ООН (ФАО). Отметим несколько особенностей данных ФАО. Во-первых, вместо килограммов в ФАО используются тонны. При проверке случаев, отмеченных выше, по базе данных ФАО тонны были конвертированы в килограммы. Во-вторых, некоторые данные, которые представлены в базе данных UN Comtrade, отсутствуют в базе данных ФАО, поэтому использование базы данных UN Comtrade как основной предпочтительнее из-за её полноты.

Для реализации представления данных в виде графа были проанализированы возможности современных библиотек для Python. В настоящее время предоставляется возможность выбора между библиотеками igraph и NetworkX, была выбрана библиотека NetworkX, предоставляющая удобные методы для работы с вершинами, ребрами, весами ребер по отдельности. Для инициализации направленного взвешенного графа использовался класс DiGraph библиотеки NetworkX.

5.2.3 Некоторые выводы

В данном исследовании рассмотрена проблема расхождения зеркальных данных при построении продовольственных сетей. Расхождение данных, полученных от каждой из двух

стран (торговые партнёры), не позволяет корректно задать вес ребра — число, характеризующее товарный поток между этой парой стран.

Предлагаемый системный подход к анализу несоответствия в данных глобальной торговли не зависит от типа экспортируемого/импортируемого товара и может быть применён для препроцессинга данных любого уровня агрегирования.

В работе приводится иллюстрация применения алгоритма для реальных данных за последние четыре года для двух категорий продовольственных товаров (пшеница и рис), отдельно показываются случаи, когда определить точное значение веса ребра невозможно, рассматриваются подходы для приближенной оценки значения веса для таких рёбер.

Построение более точной модели текущего и прошлого состояния торговли для разного уровня агрегирования даёт более широкое представление о глобальной торговле и может быть полезным при проведении исследований в сфере международной торговой политики.

5.3 Динамический анализ структурной близости регионов

5.3.1 Введение

Динамический паттерн-анализ представляет из себя анализ поведения объектов с течением времени: наблюдение и исследование перехода объектов из одного паттерна в другой и выделение объектов со схожим поведением. Анализ такого поведения для регионов за 2016–2021 гг. позволит выделить группы регионов, схожих по поведению, а также посмотреть, как вели себя регионы в период пандемии.

Целью данной работы является проанализировать структурную близость регионов в каждый момент времени, а также их близость относительно структурных изменений с течением времени.

Были сформулированы следующие задачи:

− собрать данные, на основе которых будет проводиться анализ;

− реализовать алгоритм кластеризации;

− кластеризовать регионы в каждый момент времени для выбранного периода;

− провести динамический анализ регионов;

− выделить регионы со схожим поведением.

В результате план работы выглядит следующим образом:

− сбор данных с сайта Росстата [222];

− подсчет квартальных данных для каждого региона;

− подсчет коэффициента корреляции для каждого региона;

− выбор слабо-коррелированных показателей;

− написание и реализация алгоритма на языке Python;

− проведение кластеризации регионов для периода с 2016 по 2021 года;

− выделение регионов с похожим поведением изменения паттернов с течением времени;

− анализ получившихся групп.

Теоретической и методической базой исследования были работы [223, 224, 225]. В [223] приведено подробное описание и значения термина «паттерн», а также основные методы кластеризации при анализе данных и их сравнение. Приведен пример анализа данных известного дата сета цветка ириса. В [224] реализован алгоритм из [223] на примере кластеризации округов по поведению. Для каждого из трех округов уровень поддержки партий был отображен в параллельных координатах и разбиты на две группы по форме кривой. В [225] приведен паттерн-анализ поведения банков Турции и России. Эта работа подтверждает необходимость изучения математических моделей, строящих паттерны без непосредственного предварительного выделения кластеров. Наибольший интерес представляет работа [225]. В ней производится динамический паттерн-анализ банков России. Сначала банки были объединены в группы по схожести внутренней структуры, далее банки были разделены на группы и отслеживалось изменение группы для каждого банка с течением времени. Устойчивыми были признаны банки, не меняющие группу, полуустойчивыми были признаны банки, меняющие паттерн один раз. Таких банков было выделено чуть более 50%. Однако сама математическая модель не описывается.

В результате проведенного анализа литературы были определены основные этапы работы по динамическому анализу структурной близости регионов:

− описание математической модели выделения паттернов;

− описание модели динамического анализа;

− анализ структурной близости регионов России с помощью разработанной модели на основе показателей, характеризующих промышленность, сельское хозяйство и обороты рынков.

5.3.2 Сбор, описание и обработка данных

Для анализа структурной близости регионов были выделены десять показателей, которые характеризуют отраслевую деятельность регионов:

− индекс промышленного производства (месячные данные);

− производство продуктов животноводства в хозяйствах всех категорий (квартальные данные);

− объем работ, выполненных по виду деятельности «Строительство» (месячные данные);

− объем инвестиций в основной капитал (квартальные данные);

− обороты розничной торговли (месячные данные);

− обороты оптовой торговли организаций оптовой торговли (месячные данные);

− обороты рынка платных услуг населению (месячные данные);

− оборот розничной торговли пищевыми продуктами.

Далее из месячных данных были получены квартальные данные: для индекса промышленного производства был посчитан средний показатель по трем месяцам квартала, для остальных показателей квартальные данные были почитаны, как сумма по трем месяцам. Данные взяты из доклада «Социально-экономическое положение России» [222].

5.3.3 Выбор показателей

Для выявления зависимых показателей был выбран коэффициент корреляции Пирсона. Ниже в Таблице 5.8 представлены результаты подсчета корреляции между показателями.

Таблица 5.8 – Корреляция показателей

	ИПП	Сельск_хоз_скот	Сельск_хоз_молоко	Сельск_хоз_яйца	Строительство	Инвестиции	ОбщПит	ОптТорг	ПлатУсл	РозТорг
ИПП	1.00	0.02	-0.01	0.02	0.05	0.07	0.11	0.06	0.06	0.07
Сельск_хоз_скот	0.02	1.00	0.43	0.34	0.12	0.06	0.10	-0.01	0.07	0.12
Сельск_хоз_молоко	-0.01	0.43	1.00	0.39	0.26	0.14	0.25	0.02	0.20	0.26
Сельск_хоз_яйца	0.02	0.34	0.39	1.00	0.13	0.09	0.09	-0.01	0.07	0.09
Строительство	0.05	0.12	0.26	0.13	1.00	0.92	0.82	0.70	0.77	0.82
Инвестиции	0.07	0.06	0.14	0.09	0.92	1.00	0.82	0.76	0.77	0.80
ОбщПит	0.11	0.10	0.25	0.09	0.82	0.82	1.00	0.84	0.91	0.94
ОптТорг	0.06	-0.01	0.02	-0.01	0.70	0.76	0.84	1.00	0.94	0.92
ПлатУсл	0.06	0.07	0.20	0.07	0.77	0.77	0.91	0.94	1.00	0.97
РозТорг	0.07	0.12	0.26	0.09	0.82	0.80	0.94	0.92	0.97	1.00

Как видно из таблицы, все коэффициенты корреляции в правом нижнем углу превышают 0.75 (за исключением коэффициента корреляции между показателями «Строительство» и «Оптовая торговля», который составляет 0.7). Все остальные коэффициенты корреляции не превосходят 0.5.

Проанализируем эти сильно коррелированные показатели.

1) Объем работ по виду деятельности «Строительство» — это суммарные затраты на выполнение всех работ всех строительных компаний в стране собственными силами на основании договоров с заказчиками.

2) Объем инвестиций в основной капитал — это суммарные затраты, направленные на создание нового оборудования, строительство новых объектов, приобретение машин, реконструкцию действующих объектов с целью получения прибыли.

3) Обороты общественного питания — это выручка от продажи собственной кулинарной продукции и покупных товаров без кулинарной обработки населению для потребления, главным образом, на месте, а также организациям и индивидуальным предпринимателям для организации питания различных контингентов населения.

4) Обороты оптовой торговли — это выручка от реализации товаров, приобретенных ранее на стороне, в целях перепродажи юридическим лицам и индивидуальным предпринимателям для профессионального использования (переработки или дальнейшей продажи).

5) Объем платных услуг населению м это показатель, характеризующий платное обслуживание населения, сформирован на основании данных, предоставляемых юридическими лицами по формам федерального статистического наблюдения, и оценки скрытой и неформальной деятельности на рынке платных услуг по утвержденной методике.

6) Оборот розничной торговли — выручка от продажи товаров населению для личного потребления или использования в домашнем хозяйстве за наличный расчет или оплаченных по кредитным карточкам, расчетным чекам банков, по перечислениям со счетов вкладчиков, по поручению физического лица без открытия счета, посредством платежных карт.

Последние четыре показателя характеризуют исключительно обороты определенных рынков товаров и услуг. Объем работ по виду деятельности «Строительство» характеризуют работу исключительно строительных компаний. Объем инвестиций же — более общий показатель, так как он охватывает и работу строительных компаний и характеризует обороты рынков.

Поэтому из шести рассмотренных сильно коррелированных показателей целесообразно оставить «Объем инвестиций в основной капитал».

5.3.4 Описание данных

Данные подаются на вход в виде списка векторов $v \in R^n$, каждый из которых представляет собой значения показателей для регионов в какой-то из фиксированных годов.

Каждый вектор представляется в виде параллельных координат и для любых двух отрезков, соединяющих точки, соответствующие значениям двух соседних показателей в параллельных координатах, вычисляется угол между этим отрезком и осью Ox.

Максимальный угол отклонения α — угол, на который может отклониться от оси Ox отрезок, соединяющий точки, соответствующие значениям двух соседних показателей в параллельных координатах. Так, при первой и второй нормировках данные распределены на

отрезке $[0, 1]$, а при третьей нормировке на всей прямой $[-\infty, +\infty]$, значит, в первых двух случаях $\alpha = 45°$, а в третьем $\alpha = 90°$.

Коэффициент отклонения ϵ — процент от α, соответствующий максимально допустимой разнице между соответственными углами в одном кластере.

5.3.5 Нормировки данных

Были рассмотрены три вида нормировка данных:

– минимаксная нормировка $x_i^{new} = \frac{x_i - min\,(x)}{(x) - min\,(x)}$ (рисунок 5.2);

– деление каждого показателя на сумму всех показателей $x_i^{new} = \frac{x_i}{sum(x)}$ (рисунок 5.3);

– стандартная нормировка $x_i^{new} = \frac{x_i - E(x)}{\sigma(x)}$ (рисунок 5.4).

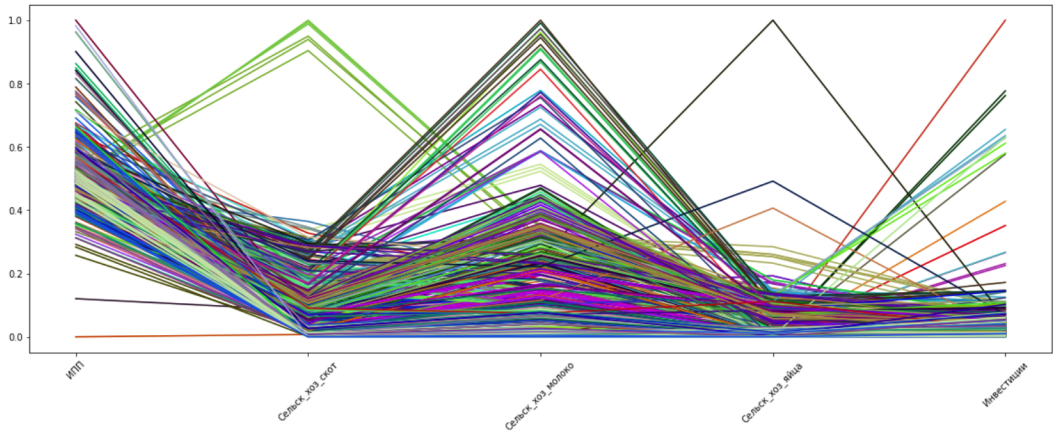


Рисунок 5.2 – Минимаксная нормировка

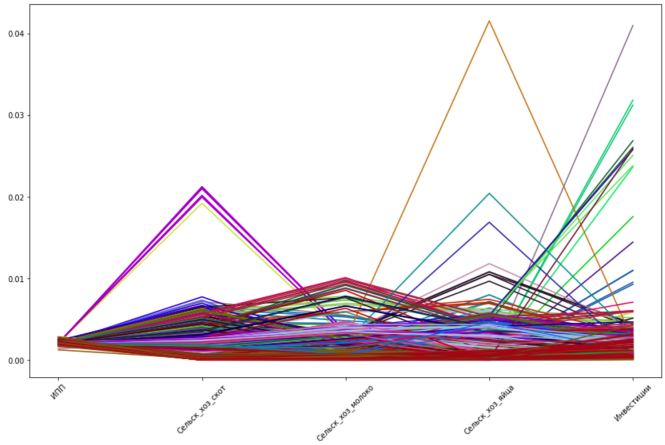


Рисунок 5.3 – Нормировка на сумму показателей

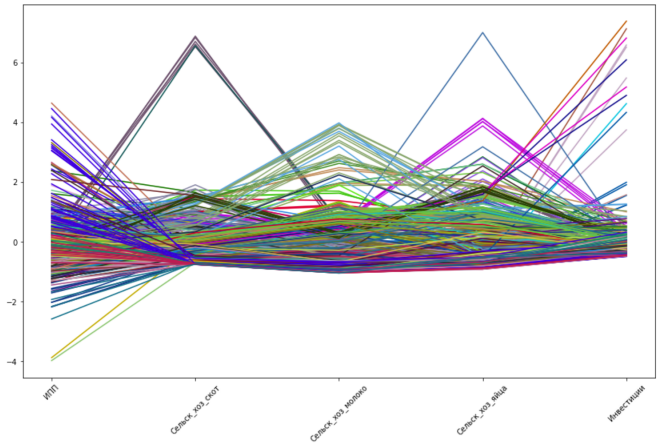


Рисунок 5.4 – Стандартная нормировка

К данным были применены три вышеописанные нормировки и применен нижеописанный алгоритм.

В первом случае алгоритм выделяет крупные кластеры, размер которых отличается между собой не более, чем в три раза, причем в них попадает порядка 40% всех объектов при коэффициенте отклонения 12%.

Во втором случае выделяются один крупный кластер, несколько средних кластеров, которые в 2–3 раза меньше крупного кластера и мелкие кластеры, состоящие из 1–3 элементов. При чем при коэффициенте отклонения в 0.2% крупный кластер составляет 25% от всей выборки данных. Это может происходить из-за наличия выбросов, то есть регионов, значения показателей в которых сильно превышают значения показателей в остальных регионах. Поэтому основная масса регионов, где все значения показателей достаточно малы, объединяется в один крупный кластер.

В третьем случае алгоритм выделяет крупные кластеры, отличающиеся между собой по количеству входящих элементов не более, чем в три раза, при этом в эти кластеры входит порядка 36% объектов при коэффициенте отклонения 50%. Кроме того, в результате такой нормировки полученные данные распределены нормально. И, хотя вероятность попадания случайной величины за пределы отрезка [−3,3] не превышает 1%, случайная величина, распределенная нормально, может принимать любые значения, поэтому максимальный угол отклонения равен 90 градусам. Учитывая вышеуказанный коэффициент отклонения, максимальная разница соответсвенных углов в одном кластере при третьей нормировке значительно больше, чем при первой нормировке.

В итоге была выбрана первая нормировка. Вторая нормировка выделяет один очень крупный кластер, что не позволяет характеризовать и отличать между собой значительное по размеру множество элементов выборки. Третья же нормировка (по сравнению с первой) при

большой допустимой разнице между соответственными углами выделяет недостаточно много объектов в крупные кластеры, размер наибольшего из которых не сильно отличается от размера наименьшего.

5.3.6 Алгоритм кластеризации

Алгоритм кластеризации предложен в [224], ниже будет приведено его описание.

1) На вход алгоритма подаются $v \in R^n$, α и ϵ.

2) Каждый вектор отображается в системе параллельных координат.

3) Каждому вектору $v \in R^n$ сопоставляется вектор углов $v_\alpha \in R^{n-1}$ между осью Ox и $n-1$ отрезком, соединяющим точки, соответствующие значениям показателей, расположенных на графике рядом (рисунки 5.2–5.4).

4) Кластеризация производится t раз, пока есть вектора, для которых еще не определён кластер:

− выбирается случайный вектор, для которого кластер еще не определен;

− выбираются все вектора, для которых

$$|u_{\alpha i} - v_{\alpha i}| < \alpha \frac{\epsilon}{200} \quad \forall i \le n-1; \tag{5.1}$$

− эти вектора объединяются в один кластер.

Среди t вариантов кластеризации выбирается тот вариант, при котором заранее заданный функционал принимает минимальное значение.

Выбранный функционал:

$$\sum_i^{k_t} \sum_j^{k_t} \rho(Ev_i, Ev_j) - \sum_{i=1}^{k_t} E\rho(Ev_i, v_i), \tag{5.2}$$

где k_t — количество кластеров после итерации t, $Ev_i \in R^n$ — вектор средних значений показателей для i-го кластера, $\rho(a,b) = \sum_{l=1}^{n}(a_l - b_l)^2$, $\rho(Ev_i, Ev_j)$ — расстояние между двумя векторами средних значений i-го и j-го кластеров, $E\rho(Ev_i, v_i)$ — величина, характеризующая плотность кластера.

5.3.7 Выбор параметров алгоритма

Коэффициент отклонения ϵ подбирается по следующим признакам:

– количество элементов выборки, попавших в самые крупные кластеры;

– на сколько близки между собой элементы получившихся кластеров;

– объединяются ли близкие элементы выборки в один кластер.

Первый признак численный — определяется, как процентное соотношение объектов, попавших в наиболее крупные кластеры, то есть тех, которые не более чем в три раза отличаются от самого крупного кластера. Второй и третий признаки опираются на анализ графиков регионов и паттернов в параллельных координатах.

Были рассмотрены различные коэффициенты отклонения. При изменении коэффициента отклонения на 1–2 процента первый параметр меняется не значительно. Так при $\epsilon = 11\%$ количество элементов, попадающих в кластеры, не более чем в 3 раза отличающиеся от самого крупного кластера, составляют 29.4%, а при $\epsilon = 13\%$ — 27.9%.

При уменьшении максимально допустимой разницы между соответственными углами отклонения увеличивается количество похожих кластеров, которые должны быть объединены в один. А при уменьшении максимальной допустимой разницы между углами отклонения увеличивается разброс элементов внутри каждого кластера. Были проанализированы разные коэффициенты отклонения и выделены два его значения для данной выборки при первой нормировке.

При $\epsilon = 13\%$ появляются паттерны, элементы которых далеки друг от друга по структуре (рисунок 5.5).

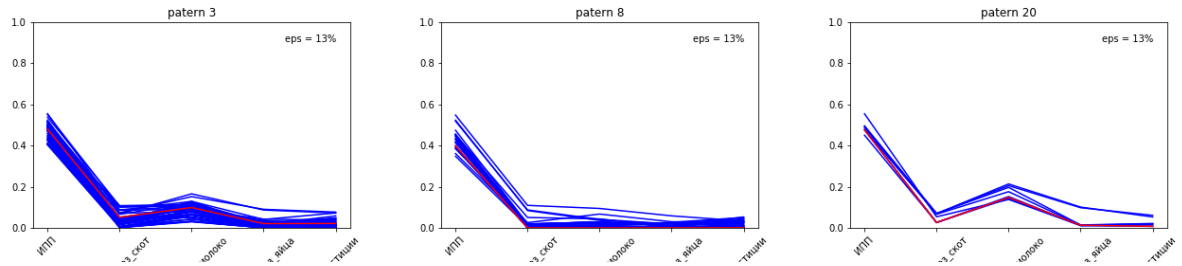


Рисунок 5.5 – Паттерны при $\epsilon = 13\%$

При $\epsilon = 11\%$ появляются различные паттерны, близкие между собой по структуре (рисунки 5.6–5.9)

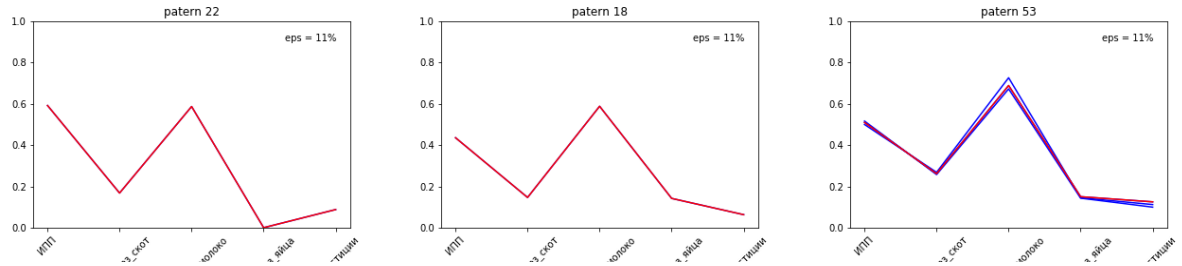


Рисунок 5.6 – Паттерны при $\epsilon = 11\%$, структура 1

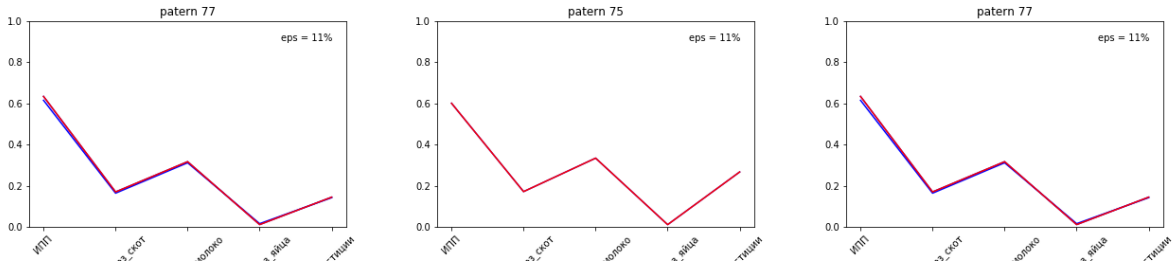


Рисунок 5.7 – Паттерны при $\epsilon = 11\%$, структура 2

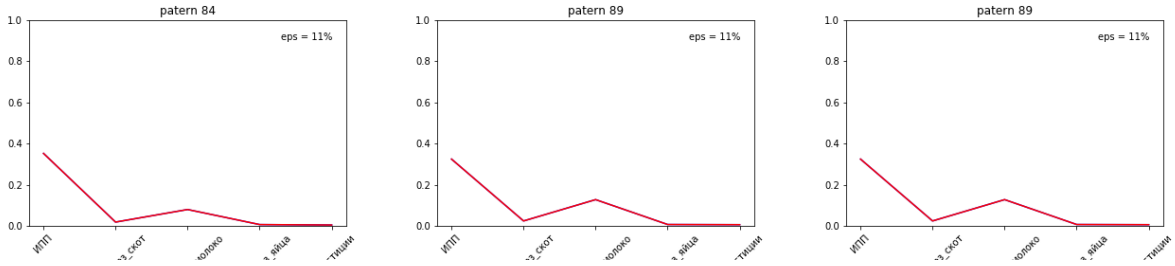


Рисунок 5.8 – Паттерны при $\epsilon = 11\%$, структура 3

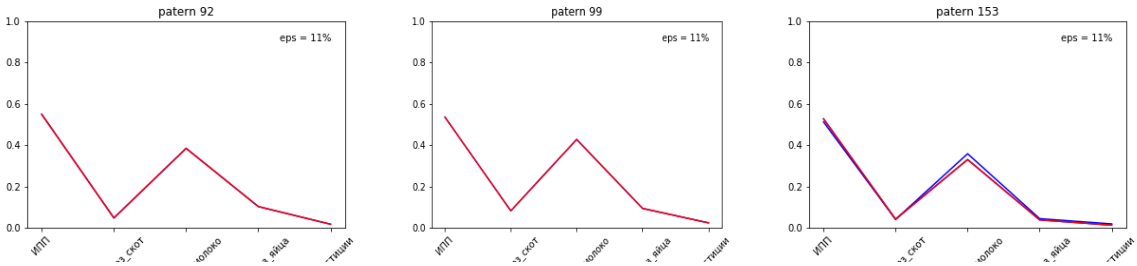


Рисунок 5.9 – Паттерны при $\epsilon = 11\%$, структура 4

Поэтому был выбран коэффициент отклонения, равный 12%.

5.3.8 Динамический паттерн-анализ

Так как при формировании паттернов опорные вектора выбирались случайным образом, номера паттернов (их порядок относительно друг друга) не несут в себе информации. Поэтому при кластеризации регионов не следует учитывать относительную нумерации паттернов, которые соответствуют каждому региону в каждый конкретный год.

Каждый вектор $v_r \in R^6$, соответствующий какому-то региону, был переведен в вектор $u_r \in \{0,1\}^5$, где $u_{ri} = 1$, если в год i паттерн региона такой же, как в год $i+1$, и $u_{ri} = 0$ иначе.

В результате множество регионов будет разбито на группы по траекториям изменения паттернов. В одну группу попадут те регионы, у которых одинаковые траектории. Введем понятия устойчивых, полуустойчивых и неустойчивых регионов. Устойчивые регионы —

регионы, ни разу не поменявшие свой паттерн, полуустойчивые регионы — регионы, поменявшие свой паттерн 1 раз, неустойчивые регионы — все прочие.

В результате кластеризации были выделены три крупных паттерна, объединяющие в себе 41% от всех элементов выборки.

Самый крупный паттерн состоит из 14 регионов: Владимирская область, Костромская область, Курская область, Республика Карелия, Республика Коми, Архангельская область, Краснодарский край, Республика Башкортостан, Самарская область, Ульяновская область, Тюменская область, Забайкальский край, Новосибирская область, Камчатский край. Этот паттерн составляет 17% от всей выборки и его траектория 1-1-1-0-0, что говорит об изменчивости регионов этой группы. Первые три года паттерн менялся, а затем в последние два года, во время распространении пандемии в стране, оставался фиксированным. Эти регионы можно назвать неустойчивыми.

Второй по величине паттерн состоит из 11 регионов: Ивановская область, Ярославская область, Ленинградская область, Республика Адыгея, Карачаево-Черкесская Республика, Саратовская область, Челябинская область, Республика Хакасия, Иркутская область, Амурская область, Чукотский автономный округ. Размер этого паттерна составляет 14% от всей выборки. Сюда входят регионы, не менявшие свой паттерн в течение всего выбранного периода, то есть устойчивые регионы.

И третий по величине паттерн состоит из 8 регионов: Калининградская область, Новгородская область, Республика Калмыкия, г. Севастополь, Республика Марий Эл, Республика Тыва, Томская область, Приморский край. Элементы этого паттерна составляют 10% от всей выборки. Сюда входят регионы, поменявшие свой паттерн 1 раз: в 2019 году. Поэтому эти регионы можно назвать полуустойчивыми.

Три самых крупных кластера составляют 41% от всей выборки. Так же были выделены средние кластеры размером от 2 до 4 регионов. Таких паттернов получилось 14. И их суммарный размер составляет 48% от всей выборки. Оставшиеся 9 регионов (это 11% от выборки) были выделены каждый в отдельный кластер.

5.3.9 Некоторые выводы

Использование динамической модели паттерн-кластеризации в задаче анализа регионального развития позволило:

− найти схожие траектории развития регионов;

− выделить группы регионов со схожим развитием;

− выделить устойчивые, неустойчивые и полуустойчивые группы развития регионов.

В результате проведенного анализа выделены три крупные группы, состоящие из 14, 11 и 8 регионов соответственно (41% от всей выборки), которые ведут себя динамически-идентично. Кроме того, выделены средние группы (от 2 до 4 регионов), которые в совокупности составляют 48% выборки.

5.4 Анализ предпочтений студентов относительно вакцинации

5.4.1 Введение

Пандемия COVID-19 оказала огромное влияние на жизнь и здоровье миллионов людей на всей планете, вызвала экономическую рецессию и политическую напряженность во всем мире.

Известно, что одной из эффективных мер по предотвращению вирусного заболевания является массовая вакцинация, создающая коллективный иммунитет и препятствующая распространению вируса. Уже к марту 2021 года некоторые страны завершили разработку вакцины против вируса SARS-CoV-2, однако уровень уверенности в эффективности и безопасности вакцины против COVID-19 был низким. Более того, были зафиксированы летальные случаи после вакцинации, например, в Норвегии 23 человека умерли после вакцинации. Такие ситуации подрывают доверие к вакцинам и к государству. Нерешительность в отношении вакцинации приводит к распространению дезинформации о предполагаемой безопасности и эффективности вакцин, влияет на спрос на вакцины и может привести к высокому риску заболевания населения.

Всероссийский опрос ВЦИОМ представил данные об отношении жителей России к вакцине от коронавируса. Так, 64% россиян считают, что призыв людей к отказу от вакцинации приведет к смертельным рискам (среди них 68% опрошенных являются жителями Москвы). 35% граждан РФ привиты, 31% респондентов согласились бы привиться отечественной вакциной и 30% предпочли бы не вакцинироваться вовсе. Были проанализированы 1600 респондентов в возрасте от 18 лет.

Основной целью данной работы является анализ отношения людей к вакцинации на примере студенческой и молодежной среды. Целевой аудиторией была выбрана молодежь, поскольку молодые люди (20–30 лет) обладают еще не до конца сформированными и твердыми убеждениями, и государству и обществу было бы крайне важно понять установки молодых людей и суметь повлиять на их решение о вакцинации.

Для изучения предпочтений относительно вакцинации от COVID-19 был проведен опрос по методу экспериментов дискретного выбора (DCE) для выявления факторов, влияющих на принятие решений относительно вакцинации против COVID-19.

5.4.2 Основная часть

5.4.2.1 Методология и обзор литературы

Главной идеей, которая обеспечивает работу DCE, можно считать предпосылку о рациональности действий индивида, позволяя ему в гипотетической ситуации при условии наличия некоторой достоверной информации выбирать альтернативу с наивысшим уровнем полезности. Гипотетической ситуацией (кейсом) в экспериментах дискретного выбора считается каждый отдельный вопрос, о выборе между предложенными альтернативами который заключает определенную информацию о предлагаемых альтернативах и подразумевает однозначный ответ от индивида (выбор строго одной из предоставленных опций).

В DCE предполагается, что информация об альтернативах может быть представлена через атрибуты. Атрибут — это характеристика альтернативы, сигнализирующая респонденту о ее качестве по конкретному параметру. Например, атрибутами вакцины можно считать эффективность, частоту и силу побочных эффектов, длительность защиты и другие. Различия между альтернативами по каждому параметру ограничены уровнями атрибутов. Уровень – это фиксированный набор значений, который может принимать атрибут. Например, эффективность вакцины может быть задана как 50%, 70% или 90% в рамках одного эксперимента. Как правило, информация об атрибутах и уровнях, соответствующая каждой альтернативе, считается справедливой. Качественные характеристики альтернатив, не включенные в атрибуты, признаются одинаковыми для всех.

DCE особенно удобен в ситуациях, когда потенциальный продукт требует больших затрат на НИОКР и запуск в производство и (или) времени на разработку. Метод активно применяется для исследования отношения к лекарственным препаратам, так как они попадают под указанные условия и к тому же зачастую имеют недостаточно рыночных аналогов, чтобы анализировать выявленные предпочтения. В 2014 году, до пандемии COVID-19, было проведено исследование по методу DCE с целью выяснить, как характеристики инфекций и вакцин влияют на процесс вакцинации и долю населения, готового вакцинироваться [226]. Было опрошено 1083 человека, и 536 ответов было включено в анализ после обработки первичных данных. В этом исследовании предлагаемые вакцины имели 5 атрибутов: 1) эффективность вакцины – это доля реципиентов, которые после вакцинации приобретут иммунитет к заражению (этот атрибут имеет 4 уровня); 2) безопасность вакцины – это наличие/отсутствие у вакцины побочных эффектов (2 уровня); 3) совет по вакцине – отношение разных социальных групп к вакцинации продуктом (6 уровней); 4) мнение о вакцине в медиа – это отношение разных СМИ к вакцинации продуктом (4 уровня); 5) стоимость вакцинации

продуктом описывает готовность респондента заплатить за вакцинацию продуктом (willingness to pay, WTP; 3 уровня).

Данное исследование не посвящено конкретной пандемии COVID-19. Несколько кейсов объединялись единым вступлением, описывающим фиктивную пандемию с установленными параметрами. В каждом кейсе респонденту предлагалось 3 альтернативы: выбрать вакцину A, выбрать вакцину B или отказаться от вакцинации предложенными вакцинами. Вакцины A и B описывались комбинацией уровней всех атрибутов, благодаря чему исследователи рассчитывали предоставить всю важную информацию.

В эксперименте дискретного выбора индивиду предлагается выбирать между гипотетическими альтернативами; вакцины A и B не существуют на рынке, но доступны как опции в кейсе. Следовательно, если выбрать для опции K атрибутов и сформировать уровни $m_k \in \mathbb{N}, k = \overline{1,K}$, то общее число опций составит $N = \prod_{k=1}^{K} m_k$ и возможное число паросочетаний опций в кейсах будет $C_N^2 = \frac{N*(N-1)}{2}$. Например, в опросе авторы могли выбирать из 576 гипотетических вариантов вакцин для составления 165600 вопросов с разными опциями 2 и 3. Вдобавок, учитывая 3 уровня заразности и 4 уровня тяжести инфекции, исследователи имели 12 сценариев пандемии. Использование всех возможных вариантов именуется full factorial design и позволяет получить наиболее точные представления о полезностях респондентов, однако на практике такой подход редко применим.

В целях соблюдения разумных ограничений по времени и сложности принято переходить от full factorial design к fractional factorial design — подвыборке из множества альтернатив и сценариев с выполнением (двойного) условия: минимизировать число альтернатив/сценариев (чаще, довести до установленного значения) при минимизации качественных потерь. Как правило, используется специальное программное обеспечение (например, Ngene) для генерирования эффективного дизайна. В работе [226] из 165600 возможных вопросов было использовано 48: по 16 вопросов в 3 опросниках, распространенных среди респондентов случайным образом. Каждый респондент участвовал только в одном опросе.

В дополнение к DCE-кейсам опросник включает в себя блок общих вопросов для сбора информации о респондентах. Традиционно в него входят вопросы про пол, возраст, национальность/гражданство, достаток и другие. Однако содержание может разниться от целей исследования, так как в дальнейшем данные используются как ковариаты и контрольные переменные. В частности, в указанном исследовании респондентов просили ответить на вопросы о степени религиозности и получение прививки от гриппа в течение последнего года.

По результатам анализа ответов респондентов были выявлены факторы, влияющие на предпочтения респондентов относительно вакцинации. В первую очередь внимание уделя-

лось эффективности вакцины (41%). Затем респондентов интересовала безопасность вакцины и наличие рекомендаций от социальных групп. Больше всего респонденты прислушивались к рекомендациям врача (45%), международных организаций (27%) и правительства Нидерландов (26%), в то время как семью и друзей указали только 2% участников. Наименьший приоритет оказался у стоимости вакцинации и мнения в медиа, причем 90% респондентов отметило, что мнение в традиционных медиа (радио, газеты, телевидение…) интересует их больше, чем социальные сети и блоги.

При помощи панельной версии модели латентных классов была изучена гетерогенность в предпочтениях. Исследовав два класса, было установлено, что пол и общее отношение к практике вакцинации являются двумя определяющими параметрами при классификации. Мужчины и одобряющие практику вакцинации имели высочайшие шансы попасть в класс 1 (средняя вероятность попасть в класс по выборке 0.63), в то время как женщины и противники вакцинации чаще относились к классу 2 (средняя вероятность по выборке 0.37). Другие социально-демографические переменные не оказывали значительного влияния на распределение по классам. Эффективность вакцин и рекомендация от правительственных учреждений оказывали значительно большее влияние на респондентов класса 1. Для класса 2 важнее были стоимость вакцинации и мнение в медиа, а также рекомендация врача. В обоих классах более высокий уровень эффективности и ожидание отсутствия побочных эффектов имели позитивное влияние на вероятность вакцинироваться, в то время как обратный эффект оказывал рост стоимости вакцинации.

Пандемия COVID-19 послужила новым поводом для исследования отношения к вакцинации с помощью метода DCE в Китае [227]. В опросе было четыре атрибута – эффективность, продолжительность защиты, побочные действия и частота инъекций. В опросе приняли участие 1694 человека, из которых 458 ответов были исключены из анализа вследствие ошибок или незаполненных ответов.

Результаты показали, что ключевым атрибутом была эффективность. Выбор программ вакцинации против коронавирусной инфекции был прямо пропорционален длительности эффекта защиты вакцинации, но обратно пропорционален к росту числа побочных эффектов и частоты инъекций. К тому же место производства влияло на предпочтения респондентов — иностранная вакцина имеет более высокий балл полезности. В среднем участники готовы заплатить 446 юаней за вакцину с эффективностью 90% и продолжительностью защиты 18 месяцев и только 98 юаней за одну инъекцию, а не за три. Вакцина с более высокой эффективностью является более ценной для респондентов женщин, которые жили в сельской местности, воспитывали детей и ранее были вакцинированы.

В Индии [228] было проведено исследование с помощью DCE, в котором приняли участие 1371 респондентов. В опросе DCE для описания вакцин были использованы такие атрибуты, как эффективность, продолжительность защиты, побочные эффекты, цена и локация, где вводили инъекций (в государственной больнице, в частном госпитале или дома). Оказалось, что эффективность вакцины и побочные эффекты являются важнейшими факторами, которые определяют предпочтения респондентов в отношении вакцины против COVID-19. Вероятность получения вакцины с 90-процентной или более высокой эффективностью в 1,3 раза больше, чем у вакцины с 80-процентной эффективностью. Люди также имеют более высокую вероятность выбора вакцины без побочных эффектов, а также они предпочитают прививку более длительным защитным действием.

В работе [228] DCE использовали для определения влияния ряда факторов на выбор альтернатив в европейских странах: Бельгии, Великобритании, Франции. Исследовательской группой были выделены 6 атрибутов и определены уровни для каждого из них: 1) эффективность вакцины – процент защиты среди числа вакцинированных (2 уровня); 2) тяжесть болезни – степень тяжести болезни к антигену которого вакцина обеспечивает иммунитет (4 уровня); 3) побочные эффекты – частота проявления побочных эффектов (2 уровня); 4) доступность – экономическая доступность препарата (2 уровня); 5) локальный охват – процент вакцинированных среди знакомых (3 уровня); 6) охват населения – процент вакцинированного населения (3 уровня). Кроме того, респонденты были распределены по двум отдельным опросам: группа «я» (без ограничения по распределению) и группа «с ребенком» (респонденты с хотя бы одним ребёнком в возрасте до 18 лет).

В результате анализы были выявлены наиболее значимые параметры при выборе альтернатив. Эффективность вакцины является наиболее важным атрибутом в Великобритании. В Бельгии и Франции данный атрибут является относительно важным, 80% и 60% респондентов дали соответствующий ответ. При этом эффективность вакцины оценивалось равноценно важным фактором среди группы «с ребенком» в сравнении с группой «я». Во Франции и Бельгии более важным фактором оказалась доступность вакцины, в Бельгии также очень значим атрибут «тяжесть болезни».

5.4.2.2 Описание анкеты

Эксперименты дискретного выбора моделируют ситуации выбора, с которыми человек сталкивается в реальной жизни, и оценивают значимость отдельных характеристик изучаемого объекта в предпочтениях людей. Они проводятся в форме опроса, выполненного в формате мини-кейсов принятия решений, в которых респонденту предлагается выбрать между несколькими альтернативами. Ключевым преимуществом такого подхода является тот

факт, что DCE позволяет запрашивать предпочтения о вакцинах, которые на самом деле не существуют и разработка которых может стоить времени и больших денежных вливаний.

На первом этапе проведения исследования была создана анкета для проведения опроса. Она состояла из нескольких блоков, в одном из которых и проводились эксперименты дискретного выбора (DCE). Эксперименты DCE требуют составления специальных вопросов о выборе среди предложенных альтернатив с разными атрибутами (характеристиками).

Сначала нами были определены атрибуты (характеристики изучаемого объекта) и зафиксированы их уровни (то есть несколько возможных значений). Выбор атрибутов для данного исследования относительно предпочтений по вакцинации основан на обзоре источников, которые были детально описаны ранее. В Таблице 5.9 представлена подробная информация об атрибутах, включенных в исследование. Каждый атрибут имеет по два уровня.

Таблица 5.9 − Атрибуты объектов в DCE

Атрибуты	Уровни
Эффективность вакцины	70%; 90%
Побочные эффекты	Сильные (лихорадка); слабые (легкое недомогание)
Разработка	Отечественная; импортная
Отзывы близких людей	Одобрение; нейтральное отношение
Мнение в СМИ	Позитивное; негативное

В ходе опроса в блоке DCE респондентам задавалось 8 вопросов, в каждом из которых необходимо было выбрать одну из двух предложенных вакцин (A или B) или же они могли воздержаться от вакцинации и выбрать третью опцию ответа: «Я не буду вакцинироваться предложенными вакцинами». Такой ответ был добавлен для людей, затрудняющихся в выборе вакцины или же являющихся противниками вакцинации. Анализ полученных ответов показал, что третья опция не привела к существенному искажению данных.

Дизайн анкеты для DCE был создан с использованием ортогональных массивов, заменивших меру D-efficiency, максимизация которой требует работы с лицензированным платным ПО. Основная характеристика ортогональных массивов — это равенство количества нулей количеству единиц в каждом столбце (соответствуют двум уровням каждого атрибута), что обеспечивает равный вклад каждого уровня каждого атрибута и способствует эффективной генерации опций (вакцин). Между столбцами отсутствует мультиколлинеарность, что также обеспечивает достижение цели. Ниже приведен использованный в работе сокращенный код Хэмминга, который состоит из 5 столбцов (число атрибутов) и 8 строк (количество создаваемых опций-вакцин). Имея 5 атрибутов с 2 уровнями в каждом, общее число вопросов респонденту могло составить 496, включая попарное сравнение всех возможных вакцин (Full Factorial Design). В проведенном опросе респонденту было предложено 8 вопросов, и каждая вакцина встречалась только 2 раза.

Опрос состоял из трех разделов. В первом блоке респондент отвечал на вопросы социально-демографические характера, а также на добавленные специфические вопросы:

− имеет ли респондент опыт работы в медицине или нет;

− как часто респонденту приходится взаимодействовать с маленькими детьми и людьми пожилого возраста.

Вопросы ставили целью позднее узнать, как эти независимые переменные повлияют на принятие или отторжение вакцины.

Второй раздел опроса являлся основным для проведения эксперимента дискретного выбора и включал в себя разработанные 8 вопросов. Для примера один из вопросов представлен в Таблице 5.10. Респонденту предлагалось выбрать одну из вакцин (A или B) или отказаться от вакцинирования обеими.

Таблица 5.10 − Пример вопроса второго блока анкеты, в котором проводится DCE

Вопрос 1	Вакцина A	Вакцина B
Эффективность	90% 9 из 10 не заразятся	70% 7 из 10 не заразятся
Побочные эффекты (не гарантированы и равновероятны)	Сильные Лихорадка	Слабые Легкое недомогание
Разработка	Импортная	Импортная
Отзывы близких людей	Нейтральны	Нейтральны
Мнение в СМИ (ТВ, социальные сети, сайт ВОЗ, интернет…)	Позитивное	Негативное

Третий и последний блок был посвящен вопросам о COVID-19. На фоне текущей пандемии необходимо знать наиболее актуальное мнение респондентов по данному вопросу в целях интерпретации результатов.

Анализ полученных данных после проведения опроса опирается на модель случайной полезности (random utility maximization, RUM), которая основана на оценке разности полезности альтернатив, включенных в множество выбора в одном вопросе. Предполагается, что выбор осуществлен респондентом в соответствии с его функцией полезности, которую, однако, исследователь не знает и наблюдает с некоторой ошибкой. Поэтому реальное значение полезности V_{ik} альтернативы i для респондента k может быть представлено следующим образом:

$$V_{ik} = v_{ik}(Z_i, y_k - p_i) + \varepsilon_{ik}, \tag{5.3}$$

где Z_i — вектор значений характеристик альтернативы i, y_k — доход респондента, p_i — стоимость альтернативы i, ε_{ik} — случайная ошибка.

В этом случае вероятность выбрать альтернативу i из множества C других предложенных в вопросе равна

$$P_{ik} = Pr\big(\forall j \in C \;\; v_{ik}(Z_i, y_k - p_i) + \varepsilon_{ik} > v_{jk}\big(Z_j, y_k - p_j\big) + \varepsilon_{jk}\big). \qquad (5.4)$$

Дальнейшие расчеты требуют предположений о виде функции полезности (простейшим видом является линейная функция) и виде распределения случайной ошибки. Для дальнейшей оценки параметров функции полезности, как правило, используются мультиномиальные логит-модели.

5.4.2.3 Краткие результаты

Онлайн-опрос был проведен в период с марта по апрель 2022 года. К этому моменту COVID-19 перестал быть одной из самых остро обсуждаемых тем, что могло, в свою очередь, уменьшить искажение в откликах под влиянием третьей стороны. Итоговая выборка состоит из 306 респондентов. Анализ отвечавших показал, что в опросе участвовали 65,4% женщин и 34,6% мужчин. По материальному благосостоянию: 46,4% в целом обеспечены, но на покупку автомобиля нужно долго копить или брать кредит, и только 6% хорошо обеспечены, могут легко позволить себе покупку автомобиля или дорогостоящий отдых. На вопрос о взаимодействии 53,3% респондентов ответили, что несколько раз в месяц или реже взаимодействуют с детьми и людьми пожилого возраста, 30,1% ежедневно и 17,6% два-три раза в неделю. Также было отмечено, что 86,6% респондентов из общей выборки 306 респондентов не имеют опыта работы в медицинской организации. Большая часть отвечавших является студентами (58,6%), из них 26,5% — это студенты НИУ ВШЭ. Из 306 респондентов 64,1% переболели COVID-19 лично и почти у всех (более 95%), кто участвовал в анкетировании, переболели члены семьи, родственники, друзья и (или) коллеги. Детально эта информация отражена в Таблице 5.11.

Таблица 5.11 − Описание данных

Внешние факторы	Опции (доля)
Количество респондентов	N= 306
Пол	- Женский (0,65) - Мужской (0,35)

Продолжение таблицы 5.11

Возраст	- <19 (0,17) - 20-23 (0,39) - 24-30 (0,18) - 31-45 (0,16) - 46-60 (0,06) - >61 (0,04)
Гражданство	- РФ (0,81) - Иностранное (0,19)
Завершенная ступень образования	- Среднее общее (0,38) - Бакалавриат (0,35) - Магистратура (0,27)
Семейное положение	- Без супруга(-и) (0,66) - С супругом(-ой) (0,34)
Рабочий статус	- Работаю (0,68) - Не работаю (0,32)
Студент	- Студент НИУ ВШЭ (0,26) - Студент другого вуза (0,32) - Нет, не студент (0,42)
Болел COVID-19	- Да (0,64) - Нет (0,36)
Текущий доход	- Затрудняюсь ответить (0,08) - Живем крайне экономно (0,1) - На еду и одежду хватит (0,29) - В целом обеспечены (0,46) - Хорошо обеспечены (0,07)
Взаимодействие с детьми и пожилыми	- Несколько раз в месяц или реже (0,52) - Ежедневно (0,3) - Два-три раза в неделю (0,18)
Отношение к вакцинации от коронавируса	- Положительное (0,49) - Нейтральное (0,42) - Негативное (0,09)
Медицинский работник	- Да (0,13) - Нет (0,87)
Опыт вакцинации (за последние 3 года)	- Был (0,73) - Нет (0,27)

Для анализа данных в программе STATA были добавлены такие переменные, как choiceset, alt, choice и personid. Choiceset указывает на набор вариантов из онлайн-анкетирования. Alt указывает альтернативу в каждом наборе вариантов. В блоке 2 опросника у каждого вопроса было по 3 варианта ответа, поэтому переменная alt принимает 3 значения. Personid — это параметр, который идентифицирует каждого респондента по-отдельности, т. е. респондент отвечает на 8 вопросов, но при этом на каждый вопрос ему предлагают по 3 варианта ответ. Поэтому в конечном наборе данных идентификационная переменная растягивается на 24 строки. Choice – это зависимая переменная, которая отражает выбор респон-

дента. Для полученных данных, представленных одним набором выбора, выбор = 1 записывается в той строке, в которой была выбрана вакцина, и выбор = 0 для других строк, где ответа не было. Это отражает тот факт, что респондент выбрал один вариант из 3 альтернатив как наиболее предпочтительный. Ниже представлена Таблица 5.12, которая была перенесена в Stata в числовом виде. Атрибуты являются независимыми переменными.

Таблица 5.12 − Набор данных для анализа

personid	obsid	alt	choiceset	eff	se	developer	fam advise	mass edia	choice	const(A=1,B=2, не буду=0)
1	1	1	1	90	2	2	1	2	0	1
1	1	2	1	70	1	2	1	1	0	2
1	1	3	1	0	0	0	0	0	1	0
1	2	1	2	70	2	1	2	2	0	1
1	2	2	2	70	1	1	1	2	1	2
1	2	3	2	0	0	0	0	0	0	0
1	3	1	3	90	1	2	2	2	0	1
1	3	2	3	90	2	1	1	1	1	2
1	3	3	3	0	0	0	0	0	0	0
1	4	1	4	70	2	2	2	1	0	1
1	4	2	4	70	1	2	1	1	1	2
1	4	3	4	0	0	0	0	0	0	0
1	5	1	5	90	2	1	1	1	1	1
1	5	2	5	70	2	1	2	2	0	2
1	5	3	5	0	0	0	0	0	0	0
1	6	1	6	90	2	2	1	2	1	1
1	6	2	6	90	1	1	2	1	0	2
1	6	3	6	0	0	0	0	0	0	0
1	7	1	7	70	2	2	2	1	0	1
1	7	2	7	70	1	1	1	2	1	2
1	7	3	7	0	0	0	0	0	0	0
1	8	1	8	90	1	2	2	2	1	1
1	8	2	8	90	1	1	2	1	0	2
1	8	3	8	0	0	0	0	0	0	0

На этих данных была построена модель регрессии Mixed Logit, которая является значимой (см. скриншот на рисунке 5.10). Тест отношения правдоподобия показывает, что полученная модель имеет смысл, а $P - value$ для всех атрибутов значим (меньше 0,05).

Mixed logit model — Number of obs = 6,975; LR chi2(5) = 1136.70; Log likelihood = -1661.2105; Prob > chi2 = 0.0000

choice	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
Mean						
constA1B2неб~0	-1.147632	.0921923	-12.45	0.000	-1.328325	-.9669382
se	-2.658572	.1823871	-14.58	0.000	-3.016044	-2.3011
eff	.0558892	.0049439	11.30	0.000	.0461993	.0655792
developer	-.3806223	.152561	-2.49	0.013	-.6796364	-.0816082
famadvise	.2152679	.0873801	2.46	0.014	.0440059	.3865298
massedia	1.501224	.1209752	12.41	0.000	1.264117	1.738331
SD						
se	1.44772	.1526818	9.48	0.000	1.148469	1.746971
eff	.0501202	.0047837	10.48	0.000	.0407443	.0594962
developer	1.764454	.1975226	8.93	0.000	1.377317	2.151591
famadvise	-.1685383	.1983037	-0.85	0.395	-.5572065	.2201299
massedia	.9424598	.1494156	6.31	0.000	.6496106	1.235309

The sign of the estimated standard deviations is irrelevant: interpret them as being positive

Рисунок 5.10 – Скриншот STATA построение модели регрессии Mixed Logit

По знакам коэффициентов регрессии видно, что рост эффективности вакцины (eff) оказывает строго положительное влияние на вероятность вакцинации (значение больше нуля на всем доверительном интервале), но влияние совсем невелико. Чуть большее позитивное влияние оказывает мнение родственников, друзей и близких (famadvise) с более широким доверительным интервалом. Наконец, очень высокий положительный коэффициент у влияния СМИ (massedia), которым респонденты доверяют. В то же время отрицательный коэффициент перед разработчиком вакцины (developer) отражает тот факт, что отвечающие скорее будут готовы вакцинироваться отечественной разработкой. Отчасти на этот коэффициент влияет текущая мировая обстановка. Самым большим по модулю и отрицательным по направлению является параметр побочных эффектов (se). Вероятности вакцинироваться при переключении уровней атрибутов могут быть выявлены с использованием следующей функции:

$$change_k = -100\phi\left(\frac{Mean_k}{SD_k}\right), \tag{5.5}$$

где $change_k$ — процентное изменение в склонности при переключении уровня в рамках одного атрибута, ϕ — значение из кумулятивного нормального распределения, $Mean_k$ — коэффициент среднего значения атрибута и SD_k — коэффициент среднего значения стандартного отклонения.

5.4.3 Некоторые выводы

В данном исследовании были изучены факторы, определяющие предпочтения относительно вакцинации от COVID-19 среди молодых людей. Эти результаты полезны для понимания настроений в целевой аудитории и для разработки эффективных кампаний о необходимости вакцинирования.

Результаты проведенных исследований нашли отражение в [229, 230].

6. Исследование по проблематики COVID-19 и другим моделям распространения

В этом разделе отчета приведены некоторые результаты, которые условно можно отнести к группе моделей распространения (эпидемий, технологий, информации, инноваций, тепла и т. д.). Эти исследования осуществлялись в следующих научных направлениях:

− моделирование прироста числа случаев заболевания COVID-19 для разных групп стран;

− исследование математических моделей распространения эпидемии COVID-19;

− сравнение условий L2-диссипативности и спектрального условия фон Неймана на схемах на разнесённых сетках для уравнений газовой динамики**.**

6.1 Моделирование прироста числа случаев заболевания COVID-19 для разных групп стран

6.1.1 Введение

В связи с новыми проблемами, вызванными начавшейся в 2019 году пандемией, быстро выросло число различных исследований по данной тематике. При запросе в ресурсе Google Academy по ключевому слову «COVID-19» будут представлены около 4.820.000 результатов, что, учитывая относительно небольшой период изучения, является весьма значительным числом. Работы посвящены различным аспектам изучения данного заболевания: сбору и визуализации данных [231], клиническим симптомам [232], описанию конкретных случаев заболевших [233]. Множество статьей посвящены исследованиям влияния COVID-19 на здоровье человека, а также возможным сопутствующим заболеваниям [234, 235, 236]. Однако, несмотря на обширную статистическую базу, относительно небольшое число работ, насколько известно авторам, посвящены использованию количественных методов для составления краткосрочных и среднесрочных прогнозов.

Настоящее исследование посвящено двум основным задачам. Во-первых, рассмотрена методология разбиения 66 стран на группы со схожими структурами по следующей базовой системе показателей: агрегированное значение еженедельных принятых в стране мер и ограничений, общий уровень образования, а также экспертное мнение относительно законопослушности граждан. Разумеется, большое внимание было уделено выбору методологии разбиения исследуемой выборки стран и базовой системе показателей. Во-вторых, составление модели прогноза прироста числа случаев заболевания COVID-19. Поскольку большую сложность представляет составление единой модели для всех стран, выявление кластеров (или групп) позволяет существенным образом повысить конечную точность прогноза. Тестирование происходило на данных за 1 февраля 2020 – 1 октября 2020, а также сопоставлено с дан-

ными за 2022 год. В большинстве случае, полученный коэффициент детерминации превышал значение 0.8.

6.1.2 Исходные данные

Поскольку конечной целью исследования является построение модели, позволяющей в краткосрочной перспективе прогнозировать прирост числа заболевших в отдельных странах мира, предполагается использование базовой системы показателей, позволяющей определить различия в поведении граждан при наличии ограничительных мер, а также непосредственно тренд заболевания. Для определения поведения граждан предполагалось использование как общего уровня образования и законопослушности, так и строгость принятых в стране ограничительных мер. Работа с данными по образованию не представляет особой сложности: имеются общедоступные базы данных о количестве граждан с разным уровнем образования [237]. Однако, большой вопрос составляет операционализация понятия «законопослушность» в данной задаче и выбор характеризующих показателей. Среди возможных вариантов, предлагалось использование различных агрегированных индексов, в том числе [238], или частичное использование показателей некоторых исследований (к примеру, Public Trust in Politicians из [239]). Однако, данные индексы характеризуют скорее уровень доверия, чем законопослушность. В связи с этим, было решено перейти к экспертным оценкам, проведенным специалистами из Национального исследовательского университета «Высшая школа экономики» и Института проблем управления им. В. А. Трапезникова Российской Академии Наук. Экспертные оценки потребовали определенных ограничений для исходной выборки, в связи с чем в нее включены 66 стран.

Рассмотрим работу с выбранной системой показателей подробнее:

1) Еженедельные данные о количестве заболевших COVID-19 в стране, взятые из [240]. Поскольку сравнение абсолютного числа значений в ряде случае не всегда репрезентативно, в настоящем исследовании используется понедельный прирост. Все значения логарифмируются.

2) Агрегированный показатель, характеризующий уровень принятых в стране ограничительных мер для населения в связи с ухудшением эпидемиологической ситуации [241]. Принятые меры нормируются по шкале [0,1].

3) Агрегированный показатель, характеризующий уровень образования, на основе данных, представленных в [237]. В качестве исходных данных выступает процент населения no education, incomplete primary, primary, lower secondary, upper secondary and post secondary. Уровень образования нормируется в шкале [0,1].

4) Общий уровень законопослушности в стране. Показатель рассчитан на основе экспертных оценок.

6.1.3 Методология разбиения

В работе предполагалось повышение точности прогноза за счет составления не единой модели для всех стран, а после предварительно их разбиения на схожие группы. С этой целью рассматривалось использование методов анализа паттернов [242, 243], классификации и кластеризации. Результаты, полученные на основе паттерн-анализа, оказались трудно интерпретировать, а для классификации весьма сложно составить тестовую выборку. В связи с этим, для дальнейшего анализа использовались классические методы кластеризации. Разумеется, большой вопрос представляет как выбор конкретной методологии, так и обоснованность данного выбора. В исследовании критерием выбора являлось интерпретируемость конечных результатов и точность прогноза.

Результаты, позволяющие повысить точность прогнозных моделей, получены при помощи классического метода k-means. Критерий оценки качества кластеризации для определения количества кластеров служили силуэтные оценки (в совокупности с коэффициентом детерминации регрессионных моделей для определения точности прогноза). Однако, следует отметить, что в связи с использованием динамических данных, полученные кластеры могут отличаться как по составу, так и по количеству, в зависимости от исследуемой недели, на которой рассматривается принятые в стране меры, а также делается прогноз о приросте числа заболевших. Рассмотрим пример на 18 июля 2020. Разбиение на кластеры представлено на рисунке ниже.

Рассмотрим рисунке 6.1 подробнее. Ось X характеризует принятые меры на 18 июля 2020, ось Y — законопослушность населения в разных странах. Цветовая гамма отображает нормированный показатель, характеризующий уровень образования.

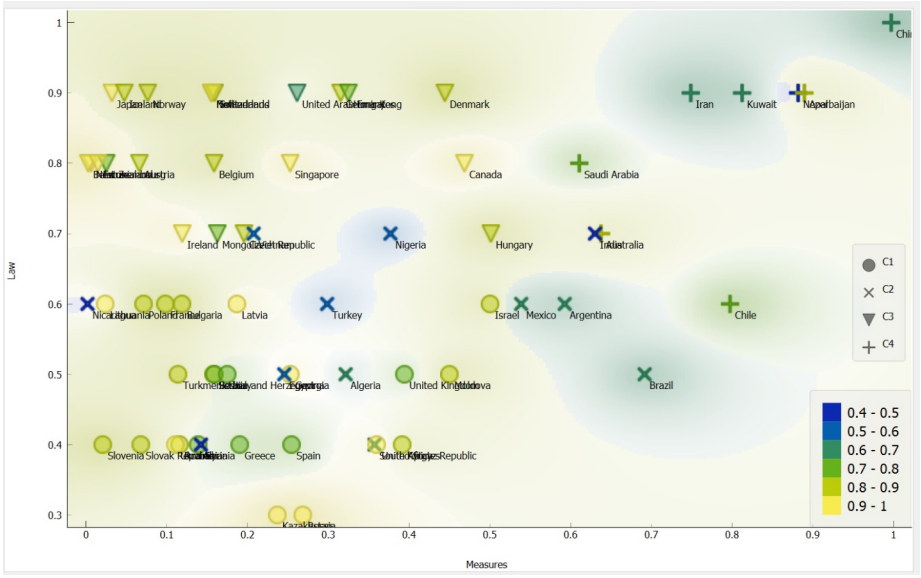


Рисунок 6.1 − Разбиение исходной выборки стран на группы (по состоянию на 18 июля 2020)

Разумеется, состав групп несколько отличается в зависимости от рассматриваемой даты (хотя, в ряде случаев, и незначительно). К примеру, приведем альтернативное разбиение, представленное на 19 сентября 2020.

На приведенных выше рисунках 6.1 и 6.2 наглядно показана разница между полученными кластерами. В ряде случаев отличается не только количество кластеров, но и их состав (хотя при близких датах изменения, как правильно, не столь ощутимы). Однако, подобный результат не противоречит исходной задаче, а именно — составлению регрессионных моделей для прогноза прироста заболеваемости. Подобный подход, наоборот, позволяет корректировать результаты в динамике, что позволяет получать более высокую точность при составлении краткосрочных прогнозов.

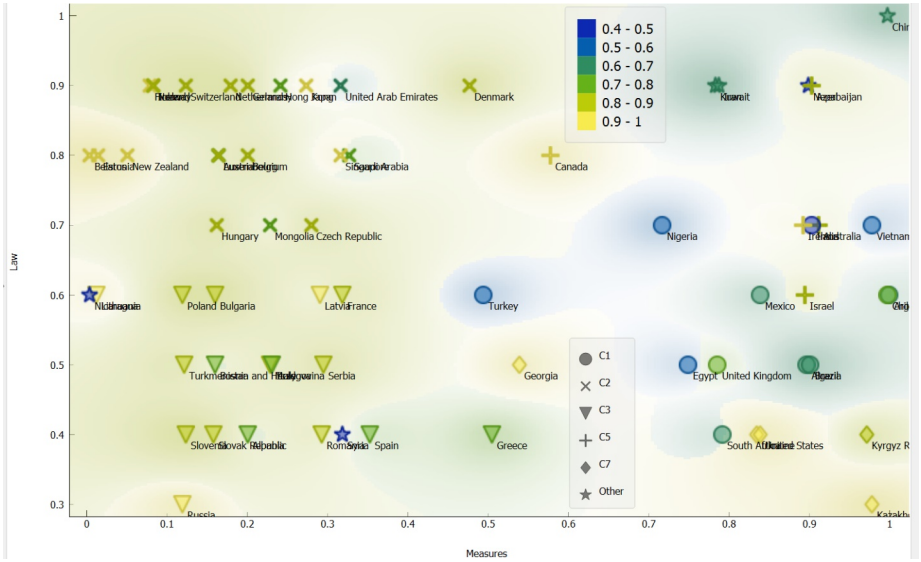


Рисунок 6.2 − Разбиение исходной выборки стран на группы (по состоянию на 19 сентября 2020)

Для сопоставления результатов приведем аналогичный пример разбиения для января 2022 года (рисунок 6.3).

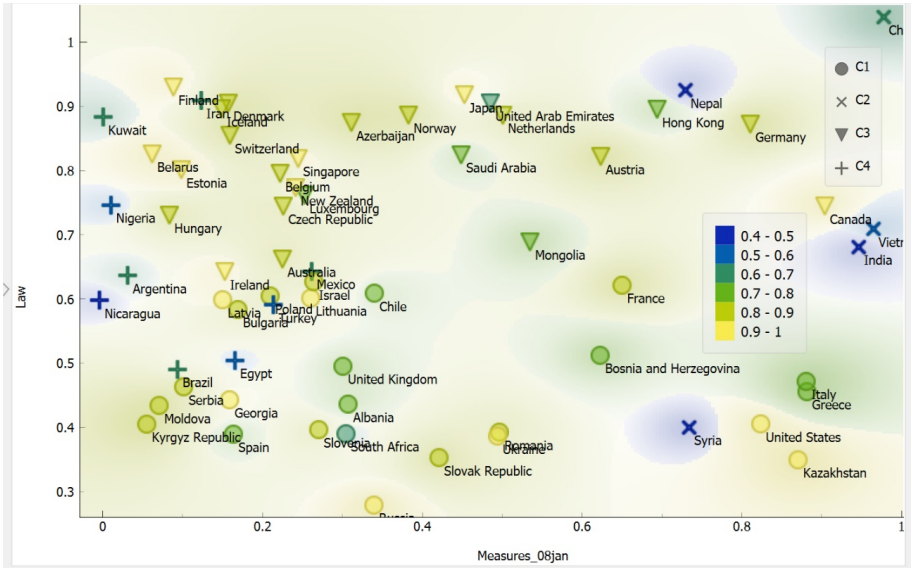


Рисунок 6.3 − Разбиение исходной выборки стран на группы (по состоянию на 9 января 2022)

6.1.4 Составление регрессионных моделей

Следующим этапом исследования являлось построение регрессионных моделей с целью прогнозирования в краткосрочном периоде прироста заболеваемости. Разумеется, подобные модели не могут учитывать все возможные факторы, однако являются весьма полезными как при составлении общих рекомендаций для населения, так и для принятия необходимых решений. Как было отмечено ранее, общая модель дает недостаточно точный прогноз

(сопоставление приведено в Таблице 6.1), в связи с чем представлены отдельные модели для каждого кластера. Для начала приведем примеры моделей 2020 года.

Кластер 1 (рисунок 6.4) состоит из 24 стран: Albania, Bosnia and Herzegovina, Bulgaria, France, Georgia, Greece, Israel, Italy, Kazakhstan, Kyrgyz Republic, Latvia, Lithuania, Moldova, Poland, Romania, Russia, Serbia, Slovak Republic, Slovenia, Spain, Turkmenistan, Ukraine, United Kingdom, United States. Характерным являются относительно средние показатели как принятых на 18 июля 2020 мер, так и значения экспертных оценок относительно законопослушности, а также высокие показатели образования.

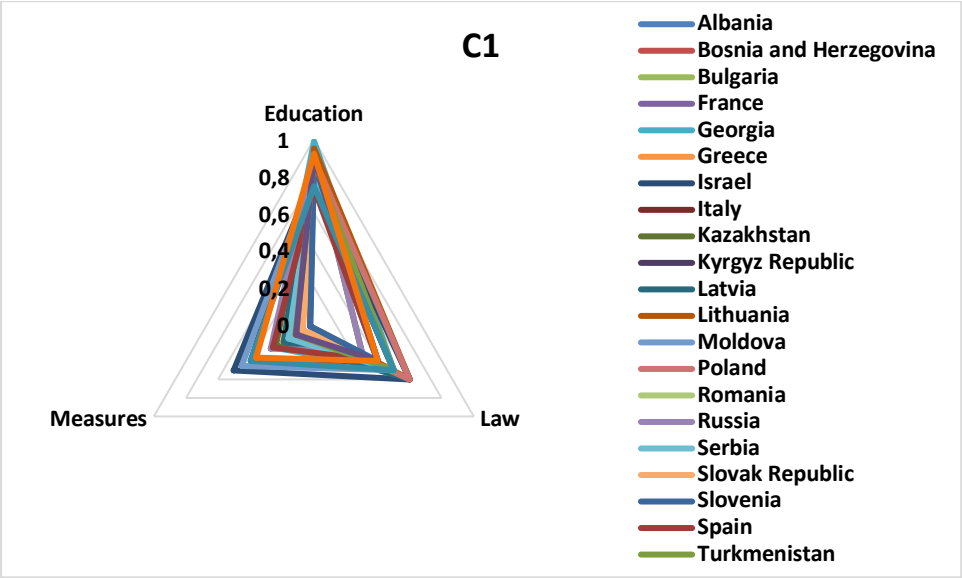


Рисунок 6.4 − Визуализация кластера 1 (по состоянию на 18 июля 2020)

Модель множественной регрессии в данном кластере выглядит следующим образом:

$$N_t(Cluster\ 1) = 0{,}015 + 0{,}523N_{t-1} - 0{,}715N_{t-2} - 0{,}009N_{t-3} + \\ +0{,}868N_{t-4} + 0{,}078N_{t-5}. \qquad (6.1)$$

Таким образом, прирост заболеваемости на 25 июля 2020 года прогнозируется на основании кластеризации, проведенной за неделю до (т. е. на 18 июля 2020). Для точного прогноза используются показали 5 предшествующих недель. При этом коэффициент детерминации равен 0.911. Для сравнения, при составлении единой регрессионной модели для 66 стран на тех же данных (и для того же периода) коэффициент детерминации равен 0.678. Таким образом, можно заметить существенное повышение качества прогноза за счет разбиения стран на группы (полное сравнение приведено в Таблицах 6.1–6.2).

Во второй кластер (рисунок 6.5) вошли 12 стран: Algeria, Argentina, Brazil, Egypt, India, Mexico, Nicaragua, Nigeria, South Africa, Syria, Turkey, Vietnam. В данном случае, ха-

рактерен относительно высокий показатель уровня образования, а также законопослушности.

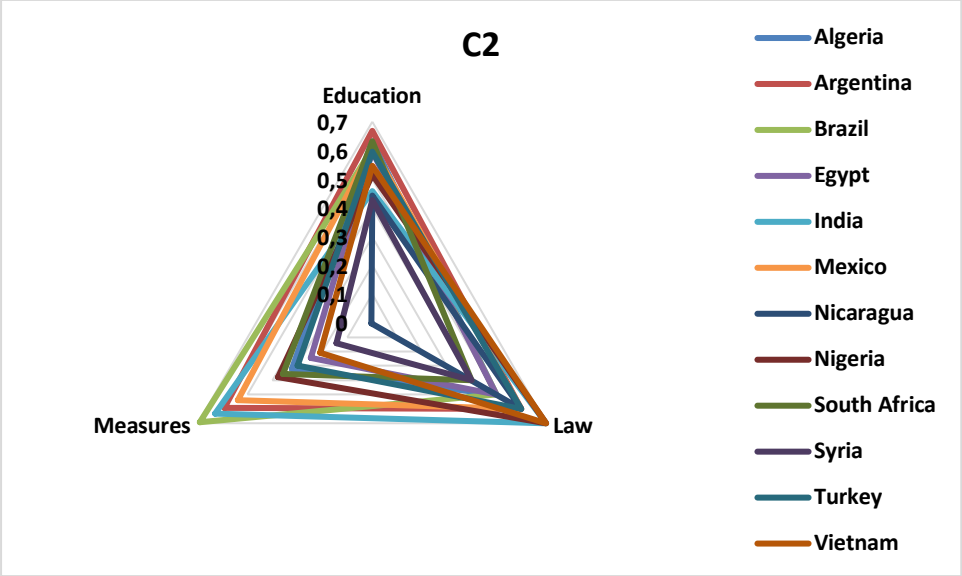


Рисунок 6.5 − Визуализация кластера 2 (по состоянию на 18 июля 2020)

Уравнение регрессии выглядит следующим образом:

$$N_t(Cluster\ 2) = 0{,}036 - 0{,}356N_{t-1} - 0{,}328N_{t-2} - 0{,}461N_{t-3} - \\ -0{,}44N_{t-4} + 1{,}737N_{t-5}. \qquad (6.2)$$

Коэффициент детерминации наименьший среди всех полученных кластеров: на 5 неделях он составляет 0.812. Однако, следует отметить, что в данном кластере значение R-squared относительно высокое и при 2 неделях, и далее меняется незначительно.

Третий кластер (рисунок 6.6) характерен для 22 стран: Austria, Belarus, Belgium, Canada, Czech Republic, Denmark, Estonia, Finland, Germany, Hong Kong, Hungary, Iceland, Ireland, Japan, Luxembourg, Mongolia, Netherland, New Zealand, Norway, Singapore, Switzerland, United Arab Emirates. В данном кластере видны относительно высокие показатели образования, уровня законопослушности, а также относительно невысокие значения агрегированного показателя принятых на 18 июля 2020 мер.

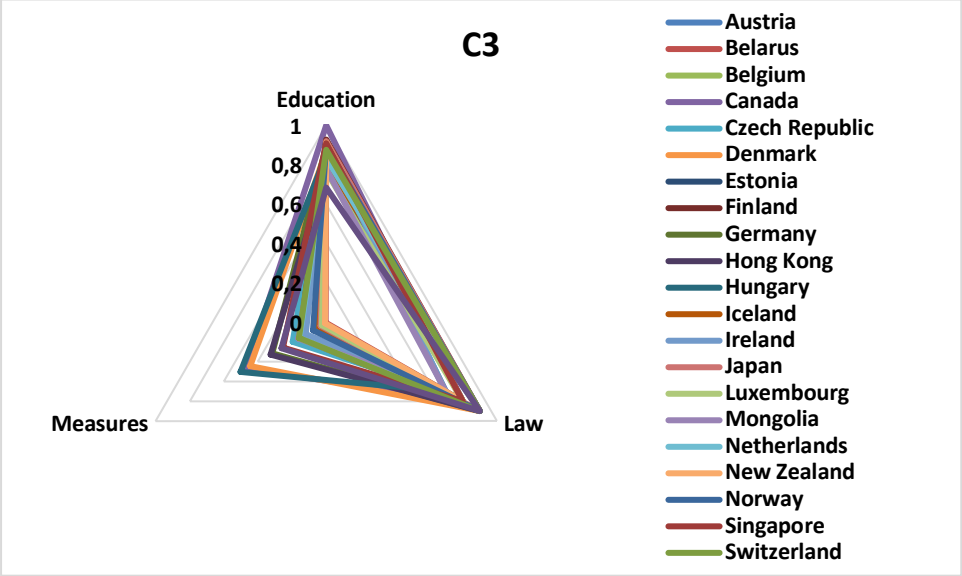


Рисунок 6.6 − Визуализация кластера 3 (по состоянию на 18 июля 2020)

Регрессионное уравнение для данного кластера:

$$N_t(Cluster\ 3) = 0{,}001 - 1{,}209N_{t-1} + 0{,}587N_{t-2} - 0{,}923N_{t-3} + \\ +3{,}363N_{t-4} - 0{,}259N_{t-5}. \quad (6.3)$$

Интересной особенностью кластера 3 является существенный прирост значения коэффициента детерминации от 1 к 2 неделям: 0.355 и 0.776 соответственно. Для 5 недель R-squared равен 0.857, что, несмотря на высокое значение, превосходит лишь значение кластера 2. Однако, отметим, что данное значение можно повысить (хотя и не очень значительно) при увеличении количества рассматриваемых периодов.

Кластер 4 (рисунок 6.7) описывается 8 странами: Australia, Azerbaijan, Chile, China, Iran, Kuwait, Nepal, Saudi Arabia. На рисунке 6.7 видны характерные для данного кластера черты: средние и относительно высокие значения для всех рассматриваемых показателей: уровня образования, законопослушности и принятых по состоянию на 18 июля 2020 года мер.

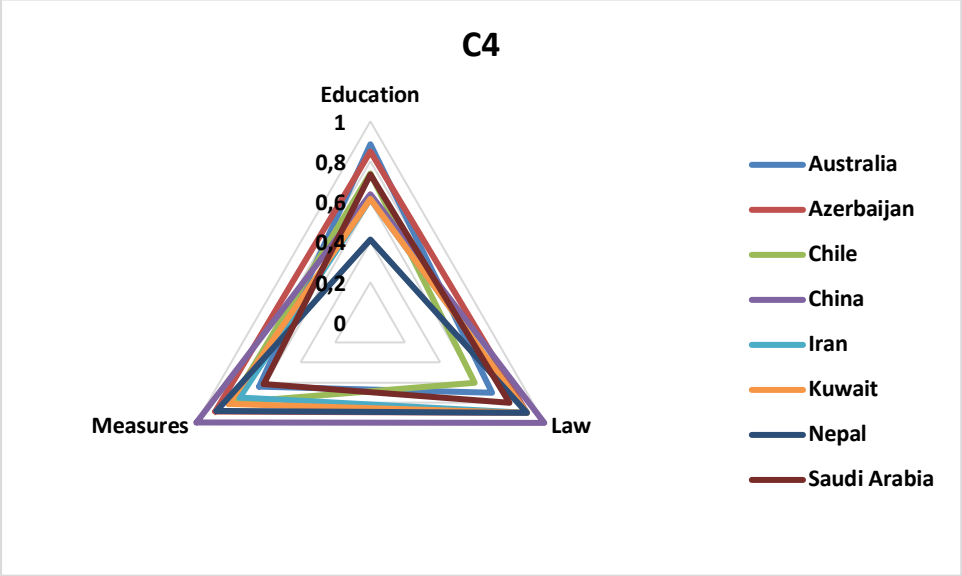


Рисунок 6.7 − Визуализация кластера 4 (по состоянию на 18 июля 2020)

Уравнение регрессии для данного кластера:

$$N_t(Cluster\ 4) = 0{,}014 + 0{,}048N_{t-1} - 0{,}529N_{t-2} + 0{,}848N_{t-3} - \\ -0{,}922N_{t-4} + 1{,}293N_{t-5}. \quad (6.4)$$

Характерным является весьма высокое значение R-squared, и его постепенное незначительное увеличение при включении в модель большего числа недель: от 0.955 при 2 неделях до 0.991 при 5. Следует отметить, что, поскольку значение коэффициента детерминации в данном кластере изначально очень высокое, прирост — наименее значимый (при сопоставлении с другими полученными кластерами).

Далее, приведем общее сравнение результатов. В качестве примера — построение регрессионной модели для кластера 1 на 18 июля 2020 с последующим прогнозом прироста заболеваемости на 25 июля 2020. Точность построенной модели, как и ранее, будет оценена на основании стандартного показателя – коэффициента детерминации. Результаты представлены в Таблице 6.1 ниже.

Таблица 6.1 − Сравнение результатов на 25 июля 2020 для кластера 1 со всей выборкой

Кол-во недель для прогноза	Коэффициент детерминации для кластера 1	Коэффициент детерминации для всей выборки
1	0.487	0.346
2	0.893	0.644
3	0.896	0.655
4	0.896	0.679
5	0.911	0.678

В Таблице 6.1 очевидно преимущество использования данных одного кластера от всей выборки, за исключением краткосрочного прогноза на основе однонедельных данных. Далее, в Таблице 6.2 приведем результаты других полученных в исследовании кластеров.

Таблица 6.2 − Сравнение результатов на 25 июля 2020 для кластеров 2–5

Кол-во недель для прогноза	Кластер 2	Кластер 3	Кластер 4
1	0.763	0.355	0.883
2	0.763	0.776	0.955
3	0.766	0.832	0.977
4	0.789	0.844	0.985
5	0.812	0.857	0.991

Результаты для других рассматриваемых периодов показали аналогичный результат: при предварительном разбиении стран на кластеры (на определенную дату) точность прогноза увеличивается. При этом, состав кластеров, как и их количество, может меняться за разные периоды, что обусловлено использованием динамического показателя — принятых ограничительных мерах в каждой стране за определенный период времени.

Для сравнения результатов представлены регрессионные модели для прогнозирования числа случаев 2022 году. Модели построены для 4 кластеров, полученных на основе k-средних.

Для всех полученных кластеров коэффициент детерминации не ниже 0,95:

$$N_t(Cluster\ 1) = 0{,}001 + 0{,}075N_{t-1} - 0{,}108N_{t-2} + 0{,}248N_{t-3} - \\ -1{,}3N_{t-4} + 1{,}99N_{t-5},$$

$$N_t(Cluster\ 2) = 0{,}025 + 0{,}523N_{t-1} - 1{,}75N_{t-3} + 1{,}25N_{t-5},$$

$$N_t(Cluster\ 3) = 0{,}009 - 0{,}46N_{t-1} + 1{,}59N_{t-2} - 0{,}97N_{t-3} - \\ -1{,}05N_{t-4} + 1{,}68N_{t-5},$$

$$N_t(Cluster\ 4) = 0{,}001 + 0{,}149N_{t-1} - 0{,}239N_{t-2} - 0{,}007N_{t-3} \\ -0{,}319N_{t-4} + 1{,}217N_{t-5}. \qquad (6.5)$$

6.1.5 Некоторые выводы

Заболевание COVID-19 представляет серьезную угрозу для здоровья всего населения планеты. Ограничительные меры были введены во многих странах, а последствия данного заболевания ещё предстоит изучить. В связи с этим, весьма актуальным является создание математических моделей прогнозирования распространения числа заболевших.

В работе приведена методология повышения точности прогноза прироста числа заболевших в 66 странах мира при предварительном разбиении стран на группы с использовани-

ем методов кластерного анализа на основании трех показателей: уровень образования в стране, законопослушность граждан и принятые ограничительные меры. Использование динамических данных позволяет корректировать как число кластеров, так и их состав, что увеличивает и качество прогноза. Высокие значения R-squared для всех полученных кластеров говорит о целесообразности проведения дальнейших исследований.

6.2 Исследование математических моделей распространения эпидемии COVID-19

6.2.1 Введение

В январе-феврале 2020 года такой показатель, как количество заражений COVID-19 в день, достиг своего исторического максимума, который оказался в разы больше любого предыдущего пика. Летом этого же года данный показатель достиг второго по высоте исторического максимума [244]. Эти данные указывают на актуальность проблемы спустя годы с начала эпидемии.

Однако с начала эпидемии произошли некоторые изменения: менялась реакция и подходы государств к борьбе с вирусом, появились вакцины, появлялись разные штаммы вируса. Подобные перемены обусловили необходимость к внесению изменений в модели. Например, с распространением вакцинации, данные о ней были внедрены в используемые модели, но, когда вакцинация стала широко доступна в большинстве анализируемых стран, использование этих данных более не давало улучшений результатов регрессионной модели, поэтому было решено убрать их из модели. Учитывая, что исследование разбито на три основные составляющие (анализ текущей ситуации, прогнозирование роста заболеваемости и анализ эффективности мер), основной задачей стало объединение связанных с этими направлениями процессов и унификация используемых данных и инструментов. Решение данной задачи позволяет быстро и просто передавать информацию между моделями, легче вносить изменения в модели, экспериментировать и быстро получать результаты.

6.2.2 Основная часть

6.2.2.1 Выбор инструментария

В каждом из трех перечисленных ранее направлений для работы моделей используется свой инструментарий, то есть это три разных процесса, между которыми необходимо передавать некоторые общие данные. Так как большинство расчетов производится при использовании различных библиотек на языке программирования Python — было решено объединить расчеты для всех моделей в едином отчете, реализованном на данном языке программирования.

6.2.2.2 Описание процесса

Как и раньше, первым делом из источника «Oxford COVID-19 Government Response Tracker» [245] выгружаются следующие ежедневные данные по странам:

− уровень строгости используемых противовирусных мер;

− кумулятивное число заболеваний;

− доля вакцинированного населения.

На данном этапе главным нововведением является использование «индекса строгости» [246] для оценки введенных противовирусных мер. Изначально для этого брались данные по 8 разным категориям введенных мер и агрегировались пороговым методом в один показатель. «Индекс строгости» — это уже заранее рассчитанный в источнике показатель, на основе тех же 8 категорий мер и еще одной, которая не использовалась при пороговом агрегировании: закрытие школ; закрытие рабочих мест; отмена массовых мероприятий; ограничения на массовые собрания; закрытие общественного транспорта; требования оставаться дома; кампании по информированию общественности («новая» категория); ограничения внутренних перемещений; контроль за международными поездками. Преимуществом использования данного показателя является повышение скорости и воспроизводимости дальнейших расчетов. С другой стороны, использование «индекса строгости» не позволяет нам выбрать конкретные категории мер для расчетов, если, например, нам будет нужен показатель, отражающий только меры направленные на ограничение передвижения. Так как «индекс строгости» и агрегированные пороговым методом меры, по сути, отражают одно и то же, то и динамика у них схожая, что можно заметить на примере графика по данным Великобритании (рисунок 6.8).

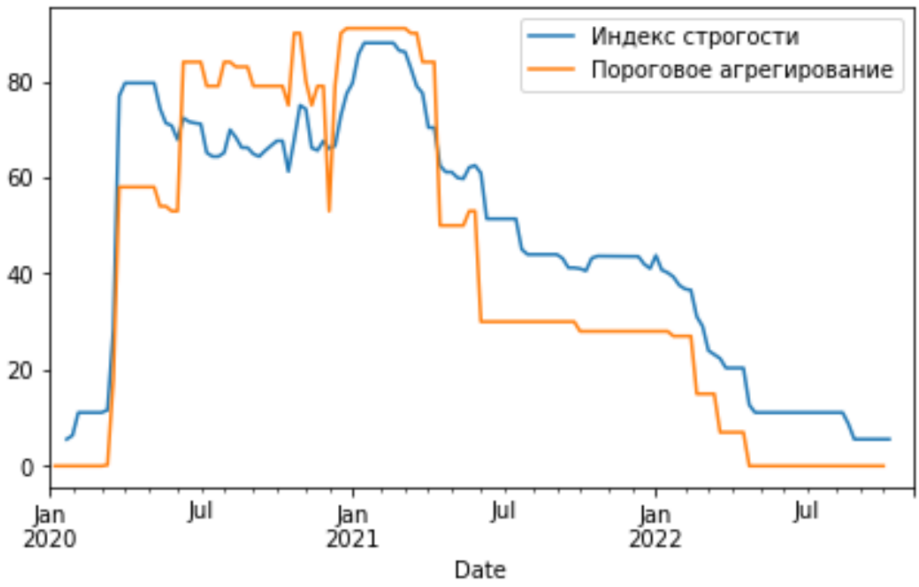


Рисунок 6.8 − Сравнение показателя строгости противовирусных мер из источника и показателя, рассчитанного пороговым методом на примере Великобритании

При необходимости, в процесс вычислений несложно внедрить показатель строгости мер, рассчитанный пороговым (или другим) методом, и использовать вместе или вместо «индекса строгости». Такая гибкость процесса является важным нововведением.

Следующим шагом является агрегирование ежедневных данных в еженедельные. При этом для «индекса строгости» и доли вакцинированного населения рассчитывается среднее значение за неделю (для каждой страны за каждую неделю; отчетный день — суббота); а для количества заболеваний берется значение за отчетный день недели (суббота), т. к. исходные данные кумулятивные.

На данном этапе у нас уже есть все данные для анализа текущей ситуации – происходит построение графиков по интересующим странам с данными о количестве заболеваний, строгости противовирусных мер и вакцинации.

Далее в любом порядке можно выполнять расчеты по двум оставшимся направлениям: прогнозирование прироста заболеваемости и анализ эффективности мер (в конкретном процессе расчеты ведутся в перечисленном порядке). Так как анализ эффективности мер изначально реализован полностью на Python, особых изменений в данную часть процесса не было внесено — все расчеты происходят аналогично тому, как это было в прошлом году, а код интегрирован в основной скрипт.

Самые значительные изменения касаются процесса расчетов в рамках модели прогнозирования. Данный процесс из полностью ручного был перестроен в полностью автоматизированный. На вход берутся ранее приведенные к еженедельным данные, которые затем

фильтруются, оставляя выборку только из интересующих нас стран. Далее происходит поиск наиболее актуальной недели, за которую по большинству стран есть статистика о строгости противовирусных мер (обычно эта статистика появляется с запозданием). Берется самая актуальная неделя, за которую только у девяти или менее стран отсутствует статистика, пропуски заполняются значениями за прошлые недели. Для каждой интересующей нас страны берутся данные о строгости мер за выбранную неделю, об уровне образования и законопослушности, затем происходит масштабирование этих данных, а на основе результата происходит кластеризация стран. Выбор количества кластеров основывается на «оценке силуэта» (silhouette score): строится несколько моделей К-средних с разным количеством кластеров (от 3 до 8 включительно), для каждой модели определяется данная метрика, выбирается модель с наивысшей оценкой. Далее каждой стране присваивается определенный номер кластера и формируется выборка с данными о приросте заболеваний за неделю (в процентном соотношении). Эта выборка используется для регрессионного анализа: выбранная ранее неделя и 15 предыдущих используются как независимые переменные, а следующая неделя после выбранной – как зависимая переменная. Для каждого кластера строится несколько регрессионных моделей: сначала со всеми 16 независимыми переменными, затем только с 15 последними неделями и т. д., вплоть до 1 последней недели. Для каждой модели рассчитывается коэффициент детерминации (R-squared) и выбирается модель с максимальным коэффициентом неравным 1. На выходе получаем регрессионную модель для каждого кластера.

6.2.3 Некоторые выводы

Результатом проведенных исследований были оформлены в виде кода на языке программирования Python, который выполняет расчеты по всем трем упомянутым направлениям — от процесса сбора данных до получения готовых моделей, графиков, расчетов. Программа позволяет быстро вносить изменения на любом из этапов и тестировать новые гипотезы, легко передавать данные между моделями; повышается воспроизводимость расчетов и потенциальная частота их обновления. Возможно реализовать автоматическое обновление данных и построение актуальных моделей в любое время по запросу.

6.3 Сравнение условий L2-диссипативности и спектрального условия фон Неймана на схемах на разнесённых сетках для уравнений газовой динамики

6.3.1 Введение

Численные методы решения газовой динамики относятся к важнейшим методам вычислительной математики. Им посвящена обширная литература, в частности, см. [247, 248, 249, 250] и ссылки внутри этих источников. Условия устойчивости таких методов, включая методы, явные по времени, представляют большой теоретический и прикладной интерес.

Среди таких численных методов существует класс методов, основанный на предварительной регуляризации уравнений газовой динамики. Они включают явные методы с симметричной аппроксимацией по пространству, основанные на так называемых квазигазодинамической (КГД) и квазигидродинамической (КГДД) регуляризациях, которые успешно применяются для решения широкого круга прикладных задач, например, см. [251, 252, 253, 254]. Эти методы были использованы для различных проблем в баротропной постановке, например, см. [255, 256, 257, 258]. До сих пор всегда использовали разностные схемы на неразнесённых сетках, в которых основные искомые функции определены на одинаковых пространственных сетках. Анализ L^2-диссипативности таких схем в линеаризованной баротропной постановке недавно был проведён в [259, 260, 261, 262]. Заметим, что в последние годы, для построения численных методов решения уравнений газовой динамики было разработано множество других подходов, основанных на различных регуляризациях соответствующих уравнений, например, см. [263, 264, 265, 266].

В вычислительной гидродинамике также известны схемы на разнесённых сетках [267], и спектральный анализ устойчивости, включая условие фон Неймана, для таких схем был проведён, в частности, в [268, 269, 270]. В [271] впервые была построена и успешно апробирована явная схема на разнесённых сетках с КГДД регуляризацией для 3D уравнений Навье – Стокса – Кана − Хилларда. Критерий (необходимое и достаточное условие) L^2-диссипативности этой схемы, линеаризованной на постоянном решении в случае 1D баротропных уравнений Эйлера и Навье − Стокса сжимаемого газа был позже выведен в [272] (среди прочих результатов) для нулевого фонового числа Маха M=0. Однако, тем не менее, на практике, этот результат применим лишь для малого числа Маха.

Результатом данного исследования являются численные эксперименты в оригинальной нелинейной постановке, которые показывают, что изучаемое на протяжении долгого времени свойство L^2-диссипативности по сравнению с условием фон Неймана является более тонким инструментом для избегания сильных осцилляций решения или существенного их уменьшения.

КГДД 1D-баротропная система уравнений вязкого сжимаемого потока газа без учёта внешних сил состоит из следующих уравнений баланса массы и импульса:

$$\partial_t \rho + \partial_x j = 0, \partial_t \left(\rho u\right) + \partial_x \left(ju\right) + \partial_x p(\rho) = \partial_x \Pi, \tag{6.6}$$

где ρ и u — искомые плотность газа и скорость в зависимости от (x,t), $x \in \mathbf{R}$, $t > 0$, $\partial_t = \dfrac{\partial}{\partial t}$ и $\partial_x = \dfrac{\partial}{\partial x}$. Также $p(\rho)$ — давление, $p(\rho) \in C^2(0,+\infty)$ и $p'(\rho) > 0$.

Регуляризованные поток массы j, регуляризующие скорость и вязкое напряжение задаются формулами

$$j = \rho\left(u - \hat{w}\right), \hat{w} = \tau\left[u\partial_x u + \frac{1}{\rho}\partial_x p(\rho)\right], \quad \Pi = \mu\partial_x u + \Pi^\tau, \quad \Pi^\tau = \rho u\hat{w}. \qquad (6.7)$$

Здесь $\mu\partial_x u$, Π^τ и Π — вязкие напряжения Навье − Стокса и регуляризующее, τ — параметр регуляризации, $\mu=\mu_{ph}+\mu_{art}$, μ_{ph} и $\mu_{art} = \tau\alpha_s p'(\rho)$ — коэффициенты полной, физической и искусственной вязкости соответственно, которые могут зависеть от ρ и u. Кроме того, α_s — число Шмидта; его также можно использовать в качестве параметра численного метода.

Пусть $\nu = \dfrac{\mu}{\rho} = \nu_{ph} + \tau\alpha_s p'(\rho)$и $\nu_{ph} = \dfrac{\mu_{ph}}{\rho}$ — соответствующие коэффициенты кинематической вязкости.

Определим сетку ω_h с узлами $x_k=kh$ и сдвинутую сетку ω_h^* с узлами $x_{k-1/2} = (k-1/2)h, k \in \mathbf{Z}$ с шагом $h>0$, а также сетку $\bar{\omega}^{\Delta t}$ с узлами $t_m = m\Delta t$ и шагом $\Delta t > 0$.

Введём операторы на функциях v, w и y, определённых, соответственно, на сетках ω_h, ω_h^* и $\bar{\omega}^{\Delta t}$:

$$sv_{k-1/2} = \frac{v_{k-1} + v_k}{2}, \delta v_{k-1/2} = \frac{v_k - v_{k-1}}{h}, \delta^\circ v_{k-1/2} = \frac{v_{k+1} - v_{k-1}}{2h}, \qquad (6.8)$$

$$s^* w_k = \frac{w_{k-1/2} + w_{k+1/2}}{2}, \delta^* w_{k-1/2} = \frac{w_k - w_{k-1}}{h}, \delta^{\circ *} w_{k-1/2} = \frac{w_{k+1} - w_{k-1}}{2h}, \qquad (6.9)$$

$$\delta_t y = \frac{y^+ - y}{\Delta t}, y^{+,m} = y^{m+1}, \qquad (6.10)$$

где $v_k = v(x_k), w_{k-1/2} = w(x_{k-1/2})$ и $y^m = y(t_m)$. Заметим, что $\delta^{\circ} = \delta^* s = s^* \delta$ и $\delta^{\circ^*} = s\delta^* = \delta s^*$.

Введём известную первообразную функцию, энтальпию:

$$P_1(\rho) = \int_{r_0}^{\rho} \frac{p'(r)}{r} dr, \rho > 0, \tag{6.11}$$

где r_0>0. Явная двухслойная по времени 3D разностная схема на разнесённых сетках с КГД-регуляризацией была построена в [271], и в 1D постановке (6.6) − (6.7) принимает следующий вид:

$$\delta_t \rho + \delta j = 0 \text{на } \omega_h^*, \quad \delta_t \left[\left(s^* \rho \right) u \right] + \delta^* \left[\left(sj \right) su \right] + \left(s^* \rho \right) \delta^* P_1(\rho) \text{ на } \omega_h, \tag{6.12}$$

$$j = \left(s^* \rho \right) \left(u - \hat{w} \right), \hat{w} = \tau \left[us^* \delta u + \delta^* P_1(\rho) \right] \text{на } \omega_h, \tag{6.13}$$

$$\Pi = \mu \delta u + \Pi^{\tau}, \Pi^{\tau} = s \left[u \left(s^* \rho \right) \hat{w} \right] \text{ на } \omega_h^*, \tag{6.14}$$

на $\overline{\omega}^{\Delta t}$. Здесь □ и u заданы на разнесённых сетках пространственных сетках, соответственно, ω_h^* и ω_h.

Заметим, что нестандартная аппроксимация $\partial_x p(\rho) \approx \left(s^* \rho \right) \delta^* P_1(\rho)$ была предложена в [273] для неразнесённых сеток. В этой схеме пространственная дискретизация не является консервативной по импульсу, но она является диссипативной по полной энергии.

Выписанную схему линеаризуем на постоянном решении $\rho_* = const > 0$ и $u_* = const$. Для этого выпишем её решение в виде $\rho = \rho_* + \rho_* \tilde{\rho}, u = u_* + c_* \tilde{u}$, подставим его в уравнения схемы, опустим слагаемые второго порядка малости относительно безразмерных возмущений $\tilde{\rho}$ и $\tilde{u}$. Тогда получим линеаризованную схему

$$\delta_t \tilde{\rho} + u_* \delta^{\circ^*} \tilde{\rho} + c_* \delta \tilde{u} - \tau_* u_* c_* \delta \delta^{\circ} \tilde{u} = 0, \tag{6.15}$$

$$\delta_t \tilde{u} + c_* \delta^* \tilde{\rho} + u_* \delta^{\circ} \tilde{u} - \tau_* c_*^2 u_* \delta^* \delta^{\circ^*} \tilde{\rho} - \tau_* u_*^2 \left(\delta^{\circ} \right)^2 \tilde{u} - \nu_* \delta^* \delta \tilde{u} = 0, \tag{6.16}$$

где $\tau_* = \tau(\rho_*, u_*), c_* = \sqrt{p'(\rho_*)}, \nu_* = \nu_{ph*} + \tau_* \alpha_s c_*^2$ и ν_{ph*} — фоновые параметр регуляризации, скорость звука и кинематические вязкости. Отметим, что линеаризованная схема не меняется, если слагаемое $\delta^* P_1(\rho)$ заменить на более стандартное $\left(s^* \rho\right)^{-1} \delta^* p(\rho)$ в (6.3) − (6.4). Ниже мы опускаем тильды над возмущениями $\tilde{\rho}$ и $\tilde{u}$ для упрощения нотации.

Приведём формулировки теорем о спектральном условии фон Неймана и условии L_2-диссипативности схемы, поскольку в основной части будем на них ссылаться.

Теорема 6.1. Пусть $\mu_{ph} = 0$ и M=0. Спектральное условие фон Неймана для схемы (6.15) − (6.16) выполнено тогда и только тогда, когда

$$\beta \le \beta_{vN} := \begin{cases} \dfrac{(\alpha_s + 1)\alpha}{4\alpha_s \alpha^2 + 1}, & |\alpha_s - 1|\,\alpha \le 1, \\ \dfrac{1}{(\alpha_s + 1)\alpha + \sqrt{(\alpha_s - 1)^2 \alpha^2 - 1}}, & \text{иначе.} \end{cases} \tag{6.17}$$

Теорема 6.2.

1) Пусть $\mu_{ph} = 0,\ \alpha_s > 0$ и $M \ne 0$. Для L_2-диссипативности схемы необходимо выполнение неравенства

2)

$$\frac{1}{\beta} \ge \frac{1}{\beta_{nec}} := \max\left\{ \bar{b}_1 \alpha, \frac{\bar{b}_2}{4\alpha_s \alpha} \right\} \tag{6.18}$$

и достаточно выполнение неравенства

$$\frac{1}{\beta} \ge \frac{1}{\beta_{suf}} := 2\bar{b}_1 \alpha + \frac{\bar{b}_2}{4\alpha_s \alpha}, \tag{6.19}$$

где

$$\bar{b}_1 = \bar{b}_1\left(\alpha_s, M^2\right) := \begin{cases} \alpha_s + 1, & M^2 \le \alpha_s + 1 \\ \dfrac{1}{4M^2}\left(M^2 + \alpha_s + 1\right)^2, & \text{иначе} \end{cases}$$

и

$$\overline{b}_2 = \overline{b}_2\left(\alpha_s, M^2\right) := \begin{cases} \alpha_s + 1, & \alpha_s \le 2 \wedge M^2 \le 2 - \alpha_s \\ \left(M^2 - 1\right)^2 + \alpha_s\left(M^2 + 1\right), & \text{иначе} \end{cases}.$$

6.3.2 Основная часть

Ниже представлен результат исследования. Проведённые численные эксперименты в целом хорошо соответствуют теоретическому анализу. Для линеаризованной задачи с M=0, уже простые эксперименты показывают, что даже малейшее нарушение условия устойчивости фон Неймана обычно приводит к распаду численного решения. С другой стороны, они подтверждают, что это условие не обеспечивает L_2-диссипативности, как и предсказывает теория (детали для краткости опустим). Хотя последний пункт не так критичен для линеаризованной задачи, тем не менее, выполнение этого свойства оказывается важным в исходной нелинейной постановке для лучшего качества численного решения с целью предотвращения или существенного уменьшения сильных численных осцилляций, хотя обычно лишь самые высокоамплитудные приводят к полному распаду решения.

В нелинейной постановке положим $p(\rho) = \rho^\gamma$ с $\gamma = 1.4$. Возьмём

$$\Delta t = \frac{\hat{\beta} h}{\max\limits_{k \in \mathbf{Z}} \left(c_k + |u_k|\right)}, \tau = \frac{\alpha h}{c}, c = \sqrt{p'\left(s^* \rho\right)}. \tag{6.20}$$

Рассмотрим две задачи Римана.

Пример 6.1. Возьмём начальные данные

$$\rho_0(x) = \begin{cases} 1.4, & x < 0 \\ 1, & x > 0 \end{cases}, u_0(x) = 0, \tag{6.21}$$

и решим задачу до момента времени $t_{fin} = 0.04$. Тогда число Маха $|M|$ меняется от 0 до ≈0.1777 и, тем самым, достаточно мало.

Положим $\alpha = 0.25, \hat{\beta} = 0.05$ и $h = 1/150$. На рисунке 6.9 приведены пространственные графики решения □ и u в момент времени $t = t_{fin}$, а также график его общей пространственной вариации $V(\rho, u)$ во времени, где

$$V(\rho,u)=\sum_{k\in\mathbf{Z}}\left(\left|\rho_{k+1/2}-\rho_{k-1/2}\right|+\left|u_k-u_{k-1}\right|\right). \tag{6.22}$$

При $\alpha_s=1$ решение вычисляется без видимых осцилляций, а $V(\rho,u)$, хотя слегка и немонотонна, но сходится к некоторому пределу с ростом t. Заметим, что в этом случае $\beta_{suf}=0.32$. Однако, в случае $\alpha_s=0$ (рисунок 6.10), осцилляции в ρ,u и $V(\rho,u)$ становятся заметными и, несмотря на то, что $\hat{\beta}=0.05$ много меньше $\beta_{vN}=0.25$, см. (6.17), качество численного решения низко.

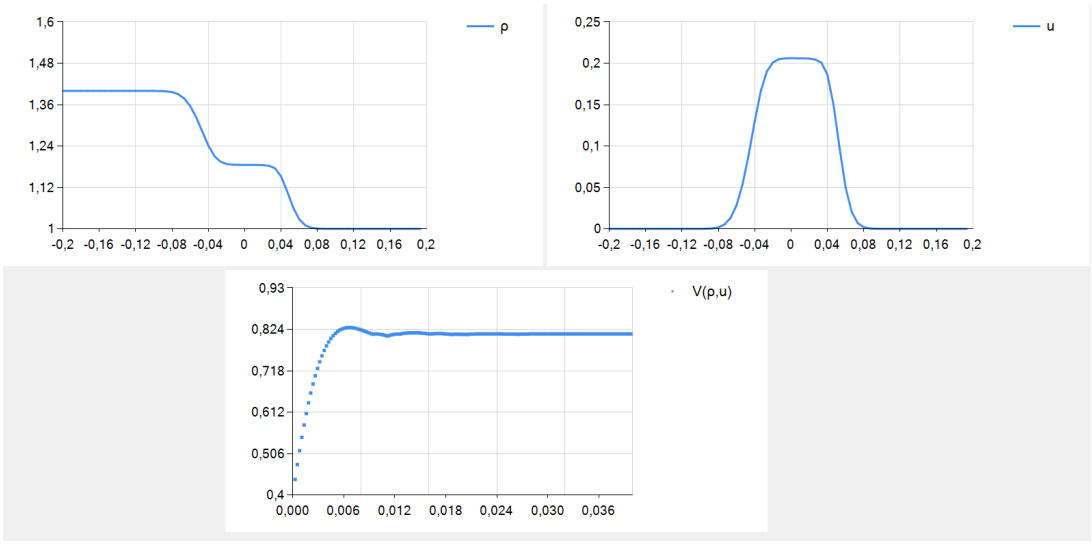


Рисунок 6.9 − Пример 1. Численное решение по пространству в момент времени $t_{fin}=0.04$ и его общая вариация во времени при $\alpha_s=1$ в случае $\beta_{suf}=0.32$

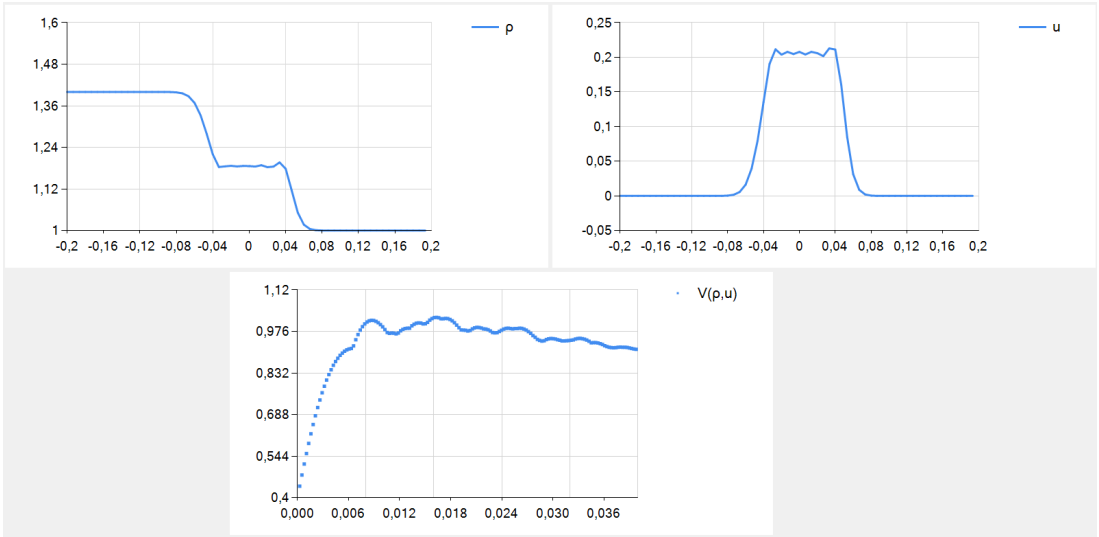


Рисунок 6.10 − Пример 1. Численное решение по пространству в момент времени $t_{fin}=0.04$ и его общая вариация во времени при $\alpha_s=0$

Пример 6.2. В этом примере используем другие начальные данные

$$\rho_0(x)=\begin{cases}\rho_L=1, & x<0\\ \rho_R=1.1, & x>0\end{cases}, u_0(x)=\begin{cases}-0.5\sqrt{p'(\rho_L)}\approx -0.5916, & x<0\\ 0.5\sqrt{p'(\rho_L)}\approx 0.6021, & x>0\end{cases} \tag{6.23}$$

и то же t_{fin}. Тогда число Маха меняется от 0 до 0.5 и уже не мало.

Возьмём те же α_s и h, но $\hat{\beta}=0.3$. На рисунке 6.11, при $\alpha_s=1$ нет осцилляций в ρ и u, а $V(\rho,u)$ монотонно возрастает и сходится к некоторому пределу с ростом t.

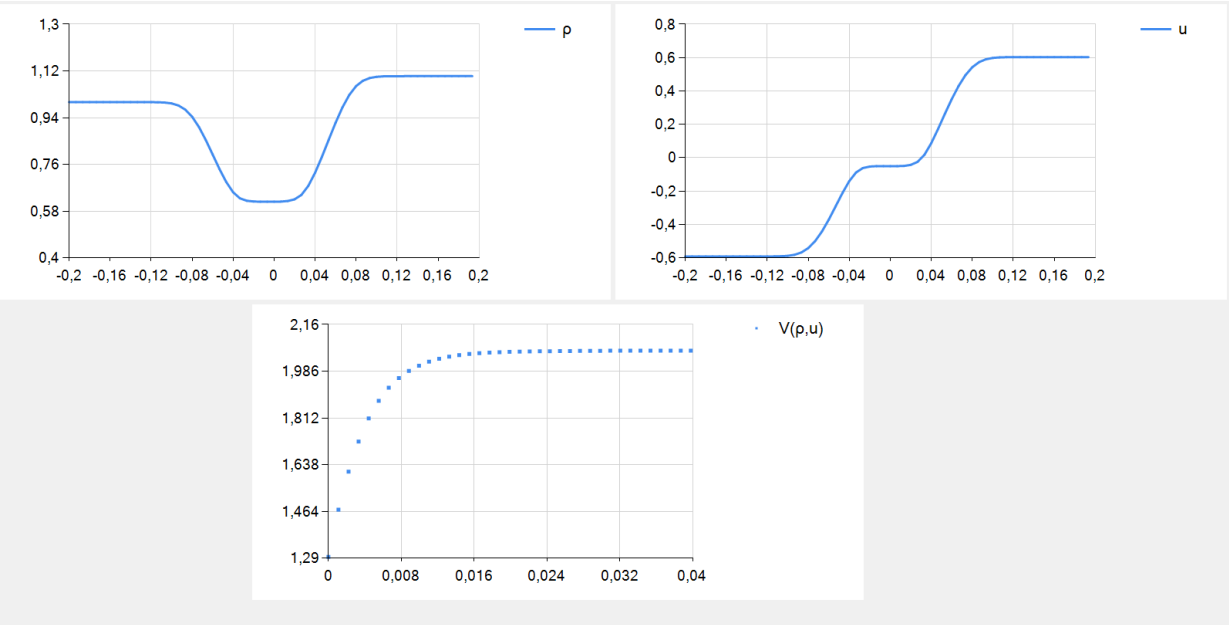


Рисунок 6.11 − Пример 2. Численное решение по пространству в момент времени t_{fin} и его общая вариация во времени при $\alpha_s=1$

Напротив, при $\alpha_s=0$ (рисунок 6.12) каждая из функций ρ,u и $V(\rho,u)$ осциллирует, и, следовательно, последний выбор гораздо хуже, что в точности предсказывает проведённый ранее анализ диссипативности в контраст условию фон Неймана.

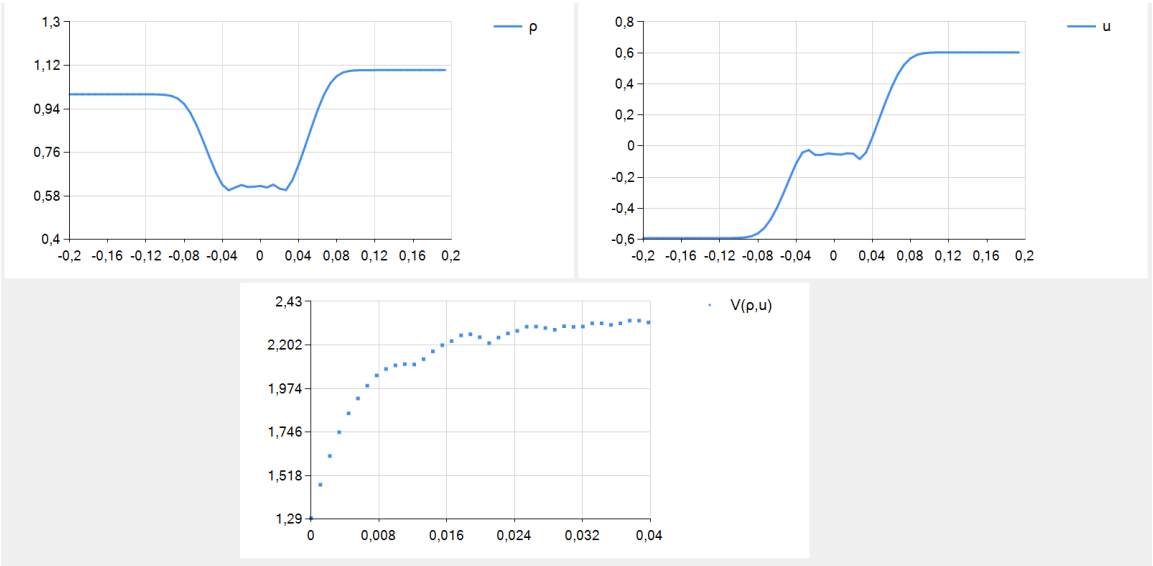


Рисунок 6.12 − Пример 2. Численное решение по пространству в момент времени t_{fin} и его общая вариация во времени при $\alpha_s = 0$

В некоторых численных экспериментах с ударными волнами также наблюдалось, что значение β_{suf} позволяло отделять достаточно слабо осциллирующие или вовсе не осциллирующие решения от осциллирующих; детали для краткости опустим.

6.3.3 Некоторые выводы

Проведены численные эксперименты в нелинейной постановке, целью которых являлось сравнение выведенного теоретического спектрального условия диссипативности с классическим спектральным условием фон Неймана. Численные эксперименты показали, что условие диссипативности обеспечивает не только устойчивость в L_2-норме, но и лучшее качество численного решения, а условие фон Неймана служит лишь необходимым условием этого свойства и, само по себе, не гарантирует устойчивости ни в какой норме. Более того, проведённые в нелинейной постановке численные эксперименты показали, что L_2-диссипативность является более тонким инструментом по сравнению с условием фон Неймана для избегания сильных осцилляций или существенного их уменьшения.

ЗАКЛЮЧЕНИЕ

В 2022 году в рамках выполнения проекта были получены следующие основные 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 на рынке 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 зависи-

мости 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 сравнительный анализ существующих моделей оценки схожести 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;

− разработана методология полуавтоматического устранения расхождения зеркальных сетевых данных торговли продовольственными товарами, которая апробирована на реальных данных мировой торговли пшеницей и рисом за последние четыре года;

− выполнен анализа 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 унификация и оптимизация мониторинга текущей ситуации с COVID-19, прогнозирование роста заражений с использованием моделей кластеризации и линейной регрессии, анализа эффективности с помощью метода оболочечного анализа данных;

− проведено численное сравнение условий диссипативности и спектрального условия фон Неймана на схемах на разнесённых сетках для уравнений газовой динамики.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Отчет МНУЛ анализа и выбора решений, ТЗ-45 за 2021г.
2. Trends in multiple criteria decision analysis / M. Ehrgott, J. Figueira, S. Greco (Eds.). – New York: Springer. – 2010. – 412 c.
3. Multiple criteria decision analysis: State of the art surveys / J. Figueira, S. Greco, M. Ehrgott (Eds.). V. 1. Second edition. – New York: Springer, 2016. – P. 467 – 496.
4. Roy B., Mousseau V. A theoretical framework for analyzing the notion of relative importance of criteria // Journal of multi-criteria decision analysis. – 1996. – V. 5. – P. 145 – 159.
5. Подиновский В.В. Идеи и методы теории важности критериев в многокритериальных задачах принятия решений. – М.: Наука, 2019. – 103 с.
6. Подиновский В.В. Аксиоматическое решение проблемы оценки важности критериев в многокритериальных задачах принятия решений // Современное состояние теории исследования операций / Под ред. Н.Н. Моисеева. – М.: Наука, 1979. – С.117 – 145.
7. Яшина Н.П. Упорядочение множества векторных оценок при наличии информации о сравнительной важности критериев // Депонированная рукопись № 2903-В86. – М.: ВИНИТИ, 1986.
8. Подиновская О.В., Подиновский В.В. Информация о важности групп критериев в многокритериальных задачах принятия решений. I. Качественная информация. Равноважные группы критериев равной важности // Информационные технологии моделирования и управления. – 2014. – № 1 (85). – С. 58 – 67.
9. Подиновская О.В., Подиновский В.В. Информация о важности групп критериев в многокритериальных задачах принятия решений. II. Количественная важность // Информационные технологии моделирования и управления. – 2014. – № 3(87). – С. 238 – 247.
10. Подиновская О.В., Подиновский В.В. Анализ иерархических многокритериальных задач принятия решений методами теории важности критериев// Проблемы управления. – 2014. – № 6. – С. 2 – 8.
11. Podinovskaya O.V., Podinovski V.V. Criteria importance theory for multicriterial decision making problems with a hierarchical structure // European journal of operational research. – 2017. – V. 258. – P. 983 – 992.
12. Гафт М.Г., Подиновский В.В. О построении решающих правил в задачах принятия решений // Автоматика и телемеханика. – 1981. – № 6. – С. 128 – 138.

13. Подиновский В.В. Многокритериальные задачи с упорядоченными по важности однородными критериями// Автоматика и телемеханика. – 1976. – № 11. – С. 118 – 127.

14. Foundation of measurement / Krantz D.H., Luce R.D., Suppes P., Tverski A. Vol. 1. – New York: Academic Press, 1971. – 577 c.

15. Наумов Г.Е., Подиновский В.В., Подиновский Вик. В. Субъективная вероятность: способы представления и методы получения // Известия АН СССР. Техническая кибернетика. – 1991. – № 1. – С. 94 – 109.

16. Kraft C.H., Pratt J.W., Seidenberg A. Intuitive probability on finite sets // Annals of mathematical statistics. – 1959. – V. 30. – P. 408 – 419.

17. Fishburn P.C. Linear extensions of additive partial orders // Order. – 1997-1998. – V. 14. – P. 153 – 169.

18. Подиновский Вик. В. К вопросу о продолжении частичных бинарных отношений в моделях принятия решений // Вычислительные системы и вопросы принятия решений: сборник / под. ред. Л.Н. Королева, П.С. Краснощекова. – М., МГУ, 1991. – С. 166 – 168.

19. Кини Р., Райфа Х. Принятие решений при многих критериях: предпочтения и замещения / Пер. с англ. – М.: Радио и связь, 1981. – 560 с.

20. Шрейдер Ю.А. Равенство, сходство, порядок. – М.: Наука. – 1971. – 256 с.

21. Warshall S. A theorem on Boolean matrices // Journal of the association for computing machinery. – 1962. – V. 9. – P. 11 – 12.

22. Нелюбин А.П., Подиновский В.В. Многокритериальные задачи с упорядоченными по важности группами критериев // Автоматика и телемеханика. – 2022. – № 7. – С. 119 – 136.

23. Nelyubin A.P., Podinovski V.V. Multicriteria problems with importance-ordered criteria groups // Automation and Remote Control. – 2022. – No 7. – P. 1108 – 1122.

24. Faliszewski P., Karpov A., Obraztsova S. The complexity of election problems with group-separable preferences// Auton. Agents Multi-Agent Syst. – 2022. – N 36. 18.

25. Black D. The theory of committees and elections. – Cambridge University Press. – 1958.

26. Mirrlees J. An exploration in the theory of optimal income taxation// Rev. Econ. Stud. – 1971. – N 38. – P. 175–208.

27. Roberts K. Voting over income tax schedules// J. Pub. Econ. – 1977. – N 8(3). – P. 329–340.

28. Elkind E., Lackner M., Peters D. Preference restrictions in computational social choice: Recent progress// In Proceedings of the 25th international joint conference on artificial intelligence (IJCAI-16). – 2016. – P. 4062–4065.

29. Inada, K. A note on the simple majority decision rule// Econometrica. – 1964. – N 32. – P. 525–531.

30. Inada K. The simple majority decision rule// Econometrica. – 1969. – N 37(3). – P. 490–506.

31. Liu P. Random assignments on sequentially dichotomous domains// Games Econ. Behav. – 2020. – N 121. - P. 565–584.

32. Alonso L., Rémy J.-L., Schott, R. A linear time algorithm for the generation of trees// Algorithmica. – 1997. – N 17. – P. 162–182.

33. Chamberlin B., Courant P. Representative deliberations and representative decisions: Proportional representation and the Borda rule// Am. Polit. Sci. Rev. – 1983. – N 77(3). – P. 718–733.

34. Young H. Extending Condorcet's rule// J. Econ. Theor. – 1977. – N 16(2). – P. 335–353.

35. Walsh, T. Generating single peaked votes/ Technical Report arXiv: 1503. 02766. – 2015.

36. Szufa S., Faliszewski P., Skowron P., Slinko A., Talmon N. Drawing a map of elections in the space of statistical cultures/ In Proceedings of the 19th international conference on autonomous agents and multiagent systems (AAMAS-20). – 2020. – P. 1341–1349.

37. Alonso L., Rémy J.-L., Schott, R. Uniform generation of a Schröder tree// Inf. Process. Lett. – 1997. – N 64. – P. 305–308.

38. Dutta B. Covering sets and a new Condorcet choice correspondence // Journal of Economic Theory. – 1988. – N 44. – P. 63–80.

39. 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.

40. 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.

41. Brandt F., Harrenstein P. Set-rationalizable choice and self-stability // Journal of Economic Theory. – 2011. – N 146(4). – P. 1721–1731.

42. Bordes G. Some more results on consistency, rationality and collective choice // In: Aggregation and revelation of preferences – Amsterdam: North-Holland / Elsevier. – 1979. – P. 175–197.

43. Aizerman M., Aleskerov F. Theory of Choice – Amsterdam: North-Holland / Elsevier. – 1995. – 324 c.

44. Duggan J., Le Breton M. Dutta's minimal covering set and Shapley's saddles // Journal of Economic Theory. – 1996. – N 70. – P. 257–265.

45. Shapley L. Some topics in two-person games // In: Advances in Game Theory – Princeton: Princeton University Press. – 1964. – N 52. – P. 1–29.

46. Brandt F., Fischer F. Computing the minimal covering set // Mathematical Social Sciences. – 2008. – N 56(2). – P. 254–268.

47. Peris J.E., Subiza B. Condorcet choice correspondences for weak tournaments // Social Choice and Welfare. – 1999. – N 16(2). – P. 217–231.

48. 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.

49. Aleskerov F. Kurbanov E. Degree of manipulability of social choice procedures // In: Current Trends in Economics: Theory and Applications – N.Y.: Springer-Verlag, 1999. – P. 13–27.

50. Subochev A. Dominant, Weakly Stable, Uncovered Sets: Properties and Extensions // Working paper WP7/2008/03 – Moscow: Higher School of Economics Publ. House. – 2008. – 28 c.

51. Subochev A. Dominating, Weakly Stable, Uncovered Sets: Properties and Extensions // Automation and Remote Control. – 2010. – N 1. – P. 130–143.

52. 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.

53. 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.

54. Aleskerov F., Subochev A. Modeling optimal social choice: Matrix-vector representation of various solution concepts based on majority rule // Journal of Global Optimization. – 2013. – N 56(2). – P. 737–756.

55. Good I.J. A note on Condorcet sets // Public Choice. – 1971. – N 10. – P. 97–101.

56. Schwartz T. Rationality and the Myth of the Maximum // Noûs. – 1972. – N 6. – P. 97–117.

57. Schwartz T. Collective choice, separation of issues and vote trading // The American Political Science Review. – 1977. – N 71(3). – P. 999–1010.

58. Smith J. Aggregation of Preferences with Variable Electorates // Econometrica. – 1973. – N 41(6). – P. 1027–1041.

59. Duggan J. A systematic approach to the construction of nonempty choice sets // Social Choice and Welfare. – 2007. – N 28. – P. 491–506.

60. Laffond G., Laslier J.-F., Le Breton M. The bipartisan set of a tournament game // Games Economic Behavior. – 1993. – N 5. – P. 182–210.

61. Dutta B., Laslier J.-F. Comparison functions and choice correspondences // Social Choice and Welfare. – 1999. – N 16(4). – P. 513–532.

62. Brandt F., Brill M., Harrenstein P. Extending tournament solutions // Social Choice and Welfare. – 2018. – N 51(2). – P. 193–222.

63. Duggan J. Uncovered sets // Social Choice and Welfare. – 2013. – N 41(3). – P. 489–535.

64. QS Methodology // QS World University Rankings – Methodology [Электронный ресурс]. – 2021 – URL: https://www.topuniversities.com/qs-world-university-rankings/methodology (дата обращения: 05.06.2022).

65. THE Methodology // World University Rankings 2022: methodology [Электронный ресурс]. – 2021 – URL: https://www.timeshighereducation.com/world-university-rankings/world-university-rankings-2022-methodology (дата обращения: 05.06.2022).

66. ARWU Methodology // ShanghaiRanking's Academic Ranking of World Universities Methodology 2021 [Электронный ресурс]. – 2021 – URL: https://www.shanghairanking.com/methodology/arwu/2021 (дата обращения: 05.06.2022).

67. Docampo D. Reproducibility of the Shanghai academic ranking of world universities // Scientometrics. – 2013. – N 94(2). – P. 567–587.

68. Florian R. Irreproducibility of the results of the Shanghai academic ranking of world universities // Scientometrics. – 2007. – N 72(1). – P. 25–32.

69. ARTU Methodology // Methodology of UNSW’s Aggregate Ranking Of Top Universities (ARTU) [Электронный ресурс]. – 2021 – URL: https://research.unsw.edu.au/artu/methodology (дата обращения: 05.06.2022).

70. CWUR Methodology // Methodology [Электронный ресурс]. – 2021 – URL: https://cwur.org/methodology/world-university-rankings.php (дата обращения: 05.06.2022).

71. RUR Methodology // Round University Ranking Methodology [Электронный ресурс]. – 2021 – URL: https://roundranking.com/methodology/methodology.html (дата обращения: 05.06.2022).

72. Итоговая таблица ранжирований вузов [Электронный ресурс]. – 2022 – URL: https://github.com/AngelinaYudina/Universities/blob/d260c108d113697f4cf746056ac9db8cb41df4d3/Result.xlsx (дата обращения: 04.11.2022).

73. Gibbard A. Manipulation of voting schemes// Econometrica. – 1973. – V. 41. – P. 587–601.

74. Satterthwaite M. Strategy-proofness and Arrow's conditions: existence and correspondence theorems for voting procedures and social welfare functions// Journal of Economic Theory. – 1975. – V. 10. – P. 187–217.

75. Duggan J., Schwartz T. (2000), Strategic manipulability without resoluteness or shared beliefs: Gibbard–Satterthwaite generalized// Social Choice and Welfare. – 2000. – V. 17. – P. 85–93.

76. Chamberlin J. R. An investigation into the relative manipulability of four voting systems// Behavioral Science. –1985. – N 30(4). – P.195–203.

77. 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.

78. Aleskerov F., Karabekyan D., Sanver R., Yakuba V. On manipulability of positional voting rules // SERIEs: Journal of the Spanish Economic Association. – 2011. – V. 2(4). – P. 431 – 446.

79. Nitzan S. The vulnerability of point-voting schemes to preference variation and strategic manipulation// Public Choice. – 1985. – V. 47. – P. 349 – 370.

80. Kelly J. Almost all social choice rules are highly manipulable, but few aren't// Social Choice and Welfare. – 1993. – V. 10. – P. 161–175.

81. Иванов А.А. Эффективные вычислительные схемы расчета манипулируемости процедур агрегирования// Информационные технологии и вычислительные системы. – 2020. – N 2. – С. 38–50.

82. Aleskerov F, Ivanov A, Karabekyan D, Yakuba V. On the Individual and Coalitional Manipulability of q-Paretian Social Choice Rules. Advances in Collective Decision Making ed. Sascha Kurz, Nicola Maaser, Alexander Mayer. — 2023. (в печати)

83. Карабекян Д.С. Об устойчивости результатов для правил агрегирования// Журнал Новой Экономической Ассоциации. — 2022. (в печати)

84. Kamali Shahdadi B. Matching with moral hazard: Assigning attorneys to poor defendants// Amer. Econ. J.: Microecon. – 2018. – V. 10(3). – P. 1–33.

85. Agan A., Freedman M., Owens E. Is your lawyer a lemon? Incentives and selection in the public provision of criminal defense// Rev. Econ. and Stat. – 2021. – V. 103(2). – P. 294–309.

86. Shem-Tov Y. Make-or-buy? The provision of indigent defense services in the US// Rev. Econ. and Stat. – 2020. P. 1–27.

87. Sasaki H., Toda M. Two-sided matching problems with externalities// J. Econ. Theory. – 1996. – V. 70(1). P. 93–108.

88. Hafalir I. E. Stability of marriage with externalities// Int. J. Game Theory. – 2008. – V. 37(3). – P. 353–369.

89. Chade H., Eeckhout J. Competing teams// The Rev. of Econ. Stud. – 2020. – V. 87(3). – P. 1134–1173.

90. Bloch F., Ryder H. Two-sided search, marriages, and matchmakers// Int. Econ. Rev. – 2000. – V. 41(1). – P. 93–116.

91. Marx P., Schummer J. Revenue from matching platforms// Theory Econ. – 2021. – V. 16(3). P. 799–824.

92. Burdett K., Coles M.G. Marriage and class// Q. J. Econ. – 1997. – V. 112(1). – P. 141–168.

93. Damiano E., Li H. Price discrimination and efficient matching// Econ. Theory. – 2007 – V. 30(2). – P. 243–263.

94. Johnson T.R. Matching through position auctions// J. Econ. Theory. – 2013. – V. 148(4). – P. 1700–1713.

95. Aoyagi M., Yoo S.H. Matching strategic agents on a two-sided platform// Games Econ. Behav. – 2020. – V. 135. – P. 271–296.

96. Aleskerov F. Power indices taking into account agents' preferences// Mathematics and democracy. – Springer, Berlin, Heidelberg, 2006. – P. 1–18.

97. Brams S. Game Theory and Politics// The Free Press, New York. – 1975.

98. Felsenthal D., Machover M. The Measurment of Voting Power: Theory and Practices, Problems and Paradoxes// Edgar Elgar Publishing House. – 1998.

99. Banzhaf III J. F. Weighted voting doesn't work: A mathematical analysis// Rutgers L. Rev. – 1964. – V. 19. – P. 317.

100. Dubey P., Shapley L. S. Mathematical properties of the Banzhaf power index// Mathematics of Operations Research. – 1979. – V. 4. – N. 2. – P. 99–131.

101. Laruelle A., Valenciano F. Shapley-Shubik and Banzhaf indices revisited// Mathematics of operations research. – 2001. – V. 26. – N. 1. – P. 89–104.

102. Aleskerov F., Holler M. J., Kamalova R. Power distribution in the Weimar Reichstag in 1919–1933// Annals of Operations Research. – 2014. – V. 215. – N. 1. – P. 25–37.

103. Соколова А. Модифицированные индексы влияния, учитывающие предпочтения участников по коалиционированию// Моделирование в социально-политической сфере. – 2009. – Т. 3.

104. Ткачев Д. С. An axiomatization for linear functions of coalescence intensity// Издательский дом ВШЭ. Series WP7 "Математические методы анализа решений в экономике, бизнесе и политике". – 2022.

105. Savage L.Y. The Foundation of Statistics. – New York: Willey, 1954.

106. 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.

107. 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.

108. Molostvov V.S. Multiple criteria optimization for stochastic systems with uncertain parameters// Model Assisted Statistics and Applications. – 2011. – V. 3. – P. 231–237.

109. 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.

110. Molostvov V.S. Three-Currency Deposit Diversification: Savage's Principle Approach// Advances in Systems Science and Applications. – 2022. – V. 22(3). – P. 135–146.

111. Wald A. Contribution to the theory of statistical estimation and testing hypothesis// Annuals Math. Statist. –1939. – N 10. – P. 299–326.

112. Молоствов В.С. Решение Сэвиджа в задаче диверсификации трехвалютного депозита// Журнал Новой экономической ассоциации. – 2022. (в печати).

113. Maron A., Maron M. Determination the Different Categories of Buyers Based on the Jaynes' Information Principle// International Journal of Economics and Business Administration. – 2019. – V. 7(4). – P. 336–342.

114. Dichev I.L., Tang. Earnings Volatility and Earnings Predictability // Journal of Accounting and Economics, VW – 2009. – V. 47. – P. 160–181.

115. Thalassinos I.E., Thalassinos Y. Financial Crises and e-Commerce: How Are they related?// 2018. Available at SSRN: https://ssrn.com/abstract=3330169.

116. Sniedovich M. Wald's Maximin Model: A Treasure in Disguise// J. Risk Finance. – 2008. – N 9. – P. 287–291.

117. Maron A., Maron M. Formulation of Agile Business Rules for Purchasing Control System Components Process Improvement// in: Model-Driven Organizational and Business Agility. MOBA 2022. Lecture Notes in Business Information Processing, v. 457 / Ed. by E. Babkin, J. Barjis, Pavel Malyzhenkov, V. Merunka. Issue 457: Model-Driven Organizational and Business Agility. MOBA 2022. Switzerland: Springer. – 2022. – P. 41–49.

118. Chediak A. Thriving at College: Make Great Friends, Keep Your Faith, and Get Ready for the Real World! – Tyndale House Publishers, 2011.

119. Belidor M. Mathematics & Statistics 101: 260+ Textbooks on Core and Elective Mathematics and Statistics Courses// Independently published, 2021.

120. Schnider E. MBA Core and Elective: Elective Courses Derived From Ten Years of Professional Experience (Kindle Edition).

121. Oxford English Dictionary, Oxford University Press, 2022.

122. Bajalinov E. B. Linear-Fractional Programming: Theory, Methods. – Kluwer Academic Publishers, Dordrecht, 2003.

123. Chen D. et.al. Applied Integer Programming: Modeling and Solution. – John Wiley and Sons. – 2010.

124. Nemhauser G., Wosley L. Integer and combinatorial optimization., Wiley, 1988.

125. Belen'kii A., Levner E., Application of models and methods of scheduling theory in optimal scheduling of freight transportation// Automation and Remote Control. – 1989. – N 50(1). – C. 1–56.

126. Foster G., Turban S. Your Relationship GPA: Lessons from Harvard students on how to make time for what matters most// Paperback, CreateSpace Independent Publishing Platform. – 2016.

127. Adams W., Sherali H. Mixed-integer bilinear programming problems// Mathematical programming. – 1993. – N 59(1). – P. 279–305.

128. Nickel S. Decision Optimization with IBM ILOG CPLEX Optimization Studio: A Hands-On Introduction with Optimization Programming Language (OPL), (Grade Texts in Operations Research), Springer. – 2022.

129. Nodet X. CPLEX Optimization Studio 20.1 is available, Dec. 11, 2020, IBM Community, 2022.

130. Belenky A., Voznesenskaja T. Three Boolean programming problems of finding optimal sets of electives to study for a college/university student"// (To be submitted to Journal of Learning Analytics in November 2022).

131. Li H. A global approach for general 0-1 fractional programming// European Journal of Operations Research. – 1994. – N 73. – P. 590–596.

132. Granot D., Granot F., On integer and mixed integer fractional programming problems// Annals of Discrete Mathematics. – 1977. – N 1. – P. 221–231.

133. Messac A. Optimization in Practice with MATLAB: For Engineering Students and Professsionals. – 1st Edition, Cambridge University Press 1st edition). – 2014.

134. Kanna S. Genetic algorithm for bin packing problem// Paperback, Independently published, December 31, 2016.

135. Kaufman A. Integer and Mixed Programming: Theory and Applications. – Academic Press. – 2012.

136. Belenky A. A Boolean Programming Problem of Choosing an Optimal Portfolio of Projects and Optimal Schedules for Them by Reinvesting Within the Portfolio the Profit From Project Implementation// Applied Mathematics Letters. – 2012. – N 25(10). – P. 1279–1284.

137. Lazarev A.A., Lemtyuzhnikova D.V., Somov M.L. Decomposition of the Knapsack Problem for Increasing the Capacity of Operating Rooms // Mathematics. – 2022. –V. 10(5). – P. https://www.mdpi.com/2227-7390/10/5/784.

138. Lazarev A., Lemtyuzhnikova D., Tyunyatkin A., Battaia O. A new interpolation-based polynomial algorithm for estimating lateness in single machine scheduling problem // IFAC-PapersOnLine. 2022. V. 55(10). P. 2881-2886.

139. Lazarev A.A., Lemtyuzhnikova D.V., Somov M.L. Approaches to solving the problem for increasing the capacity of operating rooms // Proceedings of the 10th IFAC Conference on Manufacturing Modelling, Management and Control (Nantes, France, 2022). Nantes, France: MIM-2022. – 2022. – V. 55(10). – P. 2569–2574

140. Гришин Е.М., Правдивец Н.А., Морозов Н.Ю., Лазарев А.А., Коровкин Д.М., Тюленев И.Д. Comparison of mathematical programming models for optimization of transshipment point seaport – railway // Proceedings of the 10th IFAC Conference on Manufacturing Modelling, Management and Control (Nantes, France, 2022). Nantes, France: MIM-2022. – 2022. – V. 55(10). – P. 2561–2566.

141. Cheng T.C.E., Lazarev A., Lemtyuzhnikova D. A Metric Approach for the Two-Station Single-Track Railway Scheduling Problem// Proceedings of the 10th IFAC Conference on

Manufacturing Modelling, Management and Control (Nantes, France, 2022). Nantes, France: MIM-2022. – 2022. – V. 55(10). – P. 2875–2880.

142. Zinder Y., Lazarev A., Musatova E. Single Track Transportation in a Two-Machine Production System Cheng// Proceedings of the 10th IFAC Conference on Manufacturing Modelling, Management and Control (Nantes, France, 2022). Nantes, France: MIM-2022. – 2022. – V. 55(10). – P. 2893–2897.

143. Krippendorff K. Content analysis// E. Barnouw, G. Gerbner, W. Schramm, T. L. Worth, L. Gross (Eds.). – International encyclopedia of communication. – V. 1. – New York, NY. – Oxford University Press. – 1989. – P. 403–407.

144. Blei D. Probabilistic topic models// Communications of the ACM. – N 55(4). – 2012. – P. 77–84.

145. Bohr J., Dunlap R.E. Key Topics in environmental sociology, 1990- 2014: results from a computational text analysis// Environmental Sociology. – 2018. – N 3-4(2). – P. 181–95.

146. Артемова Е. Л., Миркин Б. Г. Меры релевантности строка-текст в проблеме рубрикации научных статей// Бизнес-информатика. – 2014. – № 2. – С. 51-62.

147. Дубов М. С., Миркин Б. Г., Шаль А. А. Система автоматической обработки русскоязычных текстов // Открытые системы. СУБД. – 2014. – № 10. – С. 15–17.

148. Mirkin B., Nascimento S. Additive spectral method for fuzzy cluster analysis of similarity data including community structure and affinity matrices // Information Sciences. – 2012. – N 183(1). – P. 16–34.

149. Frolov D., Nascimento S., Fenner T., Mirkin B. Parsimonious Generalization of Fuzzy Thematic Sets in Taxonomies Applied to the Analysis of Tendencies of Research in Data Science // Information Sciences. 2020. – N 512. – P. 595–615.

150. The 2012 ACM Computing Classification System// [Электронный ресурс]. – 2012. URL: http://www.acm.org/about/class/2012 (Accessed 2022, 30 October).

151. Frolov D., Mirkin B., Nascimento S., Fenner T. Finding an appropriate generalization for a fuzzy thematic set in taxonomy // Working paper WP7/2019/01. – Moscow, Higher School of Economics Publ. House. – 2018. – 60 P.

152. Айрапетян Ж. С., Фролов Д. С., Миркин Б. Г. Метод максимального правдоподобия для обобщения нечетких множеств в таксономиях //ББК 22.1, 32.97 я43 В85. – 2021. – С. 96.

153. Hayrapetyan Z., Nascimento S., Fenner T., Frolov D., Mirkin B. Modeling Generalization in Domain Taxonomies Using a Maximum Likelihood Criterion // In World Conference on Information Systems and Technologies. – 2022. – Springer, Cham. – P. 141–147.

154. Lepskiy A. E. Analysis of Information Inconsistency in Belief Function Theory. Part I: External Conflict// Control Sciences. – 2021. – N 5. P. 2–16.

155. Lepskiy A. Analysis of Information Inconsistency in Belief Function Theory. Part II: Internal Conflict // Control Sciences. – 2021. – N 6. P. 2–12.

156. Harmanec D. Faithful approximations of belief functions// In K. B. Laskey and H. Prade, editors, Uncertainty in Artificial Intelligence 15 (UAI99), Stockholm, Sweden, 1999.

157. Petit-Renaud S., Denoeux T. Handling different forms of uncertainty in regression analysis: a fuzzy belief structure approach// In A. Hunter and S. Pearsons, editors, Symbolic and quantitative approaches to reasoning and uncertainty (ECSQARU'99). – P. 340–351. – London, June. – 1999. – Springer Verlag.

158. Denoeux T. Inner and outer approximation of belief structures using a hierarchical clustering approach// Int. J. of Uncertainty, Fuzziness and Knowledge-Based Systems. – 2001. – V. 9(4). – P. 437–460.

159. Bronevich A., Lepskiy A. Measures of conflict, basic axioms and their application to the clusterization of a body of evidence// Fuzzy Sets and Systems. – 2022. – N 446. – P. 277–300.

160. Dempster A.P. Upper and lower probabilities induced by multivalued mapping// Ann. Math. Statist. – 1967. – 38. – P. 325–339.

161. Shafer G. A mathematical theory of evidence. – Princeton University Press. – Princeton, N.J. – 1976.

162. Yager R. On the Dempster-Shafer Framework and New Combination Rules// Information Sciences. – 1987. – N 41. – P. 93–137.

163. Pichon F., Jousselme A.L., Abdallah N.B. Several shades of conflict// Fuzzy Sets Syst. – 2019. – N 366. – P. 63–84.

164. Jousselme A.-L., Maupin P. Distances in Evidence Theory: Comprehensive Survey and Generalizations // International Journal of Approximate Reasoning. – 2012. – V. 53. – P. 118–145.

165. Bronevich A., Lepskiy A. Imprecision indices: axiomatic, properties and applications// International Journal of General Systems. – 2015. – V. 44(7-8). – P. 812–832.

166. Daniel M. Properties of Plausibility Conflict of Belief Functions // Rutkowski, L., Korytkowski, M., Scherer, R., et al. (Eds.). – ICAISC 2013, Part I, LNAI 7894. – Springer-Verlag. – 2013. – P. 235–246.

167. Lepskiy, A. Cluster Decomposition of the Body of Evidence// In: Le Hégarat-Mascle, S., Bloch, I., Aldea, E. (eds) Belief Functions: Theory and Applications. BELIEF 2022. Lecture Notes in Computer Science. – V. 13506. – P. 163–173. – 2022. – Springer, Cham,

168. Freeman L.C. Centrality in social networks: conceptual clarification// Social Networks. – 1979. – P. 215–239.

169. Newman M. E. J. The mathematics of networks// The new palgrave encyclopedia of economics. – 2008. – N 2. – P. 1–12.

170. Brin S., Page L. The anatomy of a large-scale hypertextual web search engine// Computer networks and ISDN systems. – 1998. – V. 30. – N. 1-7. – P. 107–117.

171. Bonacich P. Power and centrality: A family of measures// American journal of sociology. – 1987. – V. 92. – N. 5. – P. 1170–1182.

172. Katz L. A new status index derived from sociometric analysis// Psychometrika. – 1953. – V. 18. – N. 1. – P. 39–43.

173. Kleinberg J. M. Authoritative sources in a hyperlinked environment// Journal of the ACM. – 1999. – V. 46. – N. 5. – P. 604–632.

174. Bavelas A. Communication patterns in task-oriented groups// The journal of the acoustical society of America. – 1950. – V. 22. – N. 6. – P. 725–730. https://doi.org/10.1121/1.1906679

175. Freeman L. C. A set of measures of centrality based on betweenness// Sociometry. – 1977. – P. 35–41.

176. Aleskerov F. T., Shvydun S., Meshcheryakova N. New Centrality Measures in Networks: How to Take into Account the Parameters of the Nodes and Group Influence of Nodes to Nodes// Boca Raton: CRC Press. – 2022.

177. Sabidussi G. The centrality index of a graph// Psychometrika. – 1966. – V. 31(4). – P. 581–603.

178. Centrality Measures: from Theory to Applications. [Электронный ресурс]. – URL: https://centrality.mimuw.edu.pl/ (Дата обращения: 2.11.2022)

179. Myerson R. B. Graphs and cooperation in games// Mathematics of operations research. – 1977. – V. 2. – N. 3. – P. 225–229.

180. Myerson R. B. Conference structures and fair allocation rules// International Journal of Game Theory. – 1980. – V. 9. – N. 3. – P. 169–182. https://doi.org/10.1007/BF01781371

181. Skibski O., Michalak T. P., Rahwan T. Axiomatic characterization of game-theoretic centrality// Journal of Artificial Intelligence Research. – 2018. – V. 62. – P. 33–68. https://doi.org/10.1613/jair.1.11202

182. Boldi P., Vigna S. Axioms for centrality// Internet Mathematics. – 2014. – V. 10. – N. 3-4. – P. 222–262.

183. Sanfeliu A., Fu K. S. A distance measure between attributed relational graphs for pattern recognition// IEEE transactions on systems, man, and cybernetics. – 1983. – N. 3. – P. 353–362.

184. Koutra D., Vogelstein J. T., Faloutsos C. Deltacon. A principled massive-graph similarity function// Proceedings of the 2013 SIAM international conference on data mining. – Society for Industrial and Applied Mathematics, 2013. – P. 162–170.

185. Schieber T. A. et al. Quantification of network structural dissimilarities// Nature communications. – 2017. – V. 8. – N. 1. – P. 1–10.

186. De Domenico M. et al. Structural reducibility of multilayer networks// Nature communications. – 2015. – V. 6. – N. 1. – P. 1–9.

187. Aleskerov F., Shvydun S. Stability and similarity in networks based on topology and nodes importance //International Conference on Complex Networks and their Applications. – Springer, Cham, 2018. – P. 94–103.

188. Hartle H. et al. Network comparison and the within-ensemble graph distance// Proceedings of the Royal Society A. – 2020. – V. 476. – N. 2243.

189. Papadimitriou P., Dasdan A. and Garcia-Molina, H. Web graph similarity for anomaly detection// J Internet Serv Appl 1. – 2010. – P. 19 – 30.

190. McCabe S. et al. netrd: A library for network reconstruction and graph distances// Journal of Open Source Software. – 2021. doi: 10.21105/joss.02990. arXiv: 2010.16019. arXiv preprint arXiv:2010.16019.

191. Kendall M. Rank correlation methods. – 4th Edition, Griffin, London. – 1970.

192. Aleskerov F., Yakuba V. Matrix-vector approach to construct generalized centrality indices in networks. 2020 // SSRN 3597948. Available at: https://ssrn.com/abstract=359794

193. Soussi A., Bersani C., Sacile R., Bouchta D., El Amarti A., Seghiouer H., Nachite D., Al Miys J. Coastal Risk Modelling for Oil Spill in The Mediterranean Sea// Advances in Science, Technology and Engineering Systems Journal. – 2020. – V. 5(4). – P. 273–286.

194. Hole L. R., Aguiar V., Dagestad K.-F., Kourafalou V. H., Androulidakis Y., Kang H., Henaff M., Calzada A. Long term simulations of potential oil spills around Cuba// Marine Pollution Bulletin. – 2021. – Vol. 167. – P. 11–13.

195. Wang J. H., Shen Y. M. Development of an integrated model system to simulate transport and fate of oil spills in seas// Sci. China Tech. Sci. – 2010. – V. 53. – P. 2423–2434.

196. Imanian H., Kolahdoozan M., Zarrati A. R. Vertical Dispersion in Oil Spill Fate and Transport Models// Journal of Hydrosciences and Environment. – 2017. – V. 1(2). – P. 21–33.

197. Жилякова Л. Ю. Применение ресурсных сетей для моделирования распространения веществ в водной среде// Пробл. управл. – 2011. – N. 2. – С. 46–51.

198. Aleskerov F. T., Demin S. S. Modelling Possible Oil Spills in the Barents Sea and their Consequences// Dynamics of Disasters. – Ed. by P. M. Pardalos, I. S. Kotsireas, A. Nagurney. Springer International Publishing. – 2018. – V. 2. – P. 47–56.

199. Waltman L. A review of the literature on citation impact indicators// Journal of informetrics. – 2016. – V. 10. – N. 2. – P. 365–391.

200. Biagioli M, Lippman A. Gaming the metrics: Misconduct and manipulation in academic research. – Cambridge: MIT Press. – 2020.

201. Baccini A., De Nicolao G., Petrovich E. Citation gaming induced by bibliometric evaluation: A country-level comparative analysis// PLoS One. – 2019. – V. 14. – N 9.

202. Marina T., Sterligov I. Prevalence of potentially predatory publishing in Scopus on the country level// Scientometrics. – 2021. – N 126(6). – P. 5019–5077.

203. Рубинштейн А.Я. Ранжирование российских экономических журналов: научный метод или «игра в цыфирь»?// Журнал Новой Экономической Ассоциации. – 2016. – № 30(2) – С. 162–175.

204. Балацкий Е.В., Екимова Н.А. Консенсусный рейтинг российских экономических журналов: идеология и опыт составления// Институциональный анализ науки. – 2018 – № 10(1) – С. 93–106.

205. Grudniewicz A., Moher D., Cobey K. and 32 co-authors. Predatory journals: no definition, no defence// Nature. – 2019. – V. 576.

206. Scopus Preview [Электронный ресурс]. – URL: https://www.scopus.com/ (Дата обращения: 03.11.2022)

207. Алескеров Ф.Т., Карабекян Д.С., Казачинская А.И., Семина А.С., Якуба В.И. Российские экономические журналы, их характеристики и сетевой анализ// Журнал Новой Экономической Ассоциации. – 2021– № 50(2). – С. 170–182.

208. An analysis of international business journals by bibliometric indicators [Электронный ресурс]. – 2022. – URL: https://www.hse.ru/DeCAn/business-journals (Дата обращения: 03.11.2022)

209. Microsoft Academic Graph [Электронный ресурс]. – URL: https://zenodo.org/record/6511057#.YuudRb1ByHs (Дата обращения: 03.11.2022)

210. Research Intelligence. Scopus. Content Coverage Guide [Электронный ресурс]. – URL: https://www.elsevier.com/?a=69451 (Дата обращения: 03.11.2022)

211. How are CiteScore metrics used in Scopus? [Электронный ресурс]. – URL: https://service.elsevier.com/app/answers/detail/a_id/14880/supporthub/scopus/ (Дата обращения: 03.11.2022)

212. How is SNIP (Source Normalized Impact per Paper) used in Scopus? [Электронный ресурс]. – URL: https://service.elsevier.com/app/answers/detail/a_id/14884/supporthub/scopus/related/1/ (Дата обращения: 03.11.2022)

213. How is SJR (SCImago Journal Rank) used in Scopus? [Электронный ресурс]. – URL: https://service.elsevier.com/app/answers/detail/a_id/14883/supporthub/scopus/related/1/ (Дата обращения: 03.11.2022)

214. Subochev A., Aleskerov F., Pislyakov V. Ranking journals using social choice theory methods: A novel approach in bibliometrics// Journal of Informetric. – 2018 – N 12(2). – P. 416–429.

215. Beall J. What I learned from predatory publishers// Biochem Med (Zagreb). – 2017. – N 27. – P. 273–278.

216. Beall, J. Predatory publishers are corrupting open access// Nature. – 2012. – N 489.

217. Dell'Annoa R., Caferrab R., Moroneba A. A “Trojan Horse” in the peer-review process of fee-charging economic journals// Journal of Informetrics. – 2020. – N 3(14).

218. Sudhira H. The precarious prevalence of predatory journals. [Электронный ресрус] // Research Matters. – 2018. – URL: https://researchmatters.in/news/precarious-prevalence-predatory-journals (Дата обращения: 03.11.2022)

219. Ramani S., Prasad R. UGC’s approved journal list has 111 more predatory journals. [Электронный ресурс]// The Hindu. – 2017. – URL: https://www.thehindu.com/sci-tech/science/ugcs-approved-journal-list-has-111-more-predatory-journals/article19677830.ece#comments_61473824 (Дата обращения: 03.11.2022) +

220. United Nations Comtrade Database [Электронный ресурс]. – URL: https://comtrade.un.org/data/ (Дата обращения: 03.11.2022)

221. Meshcheryakova N. Network analysis of bilateral trade data under asymmetry// 2020 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM). – IEEE, 2020. – P. 379-383.

222. Доклад «Социально-экономическое положение России» [Электронный ресурс] / Федеральная служба государственной статистики – электрон. текст. дан. – URL: https://rosstat.gov.ru/compendium/document/50801 (дата обращения 05.06.22)

223. Алескеров 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. – N. 3(25). – С. 3–19.

224. Алескеров 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// Бизнес-информатика. – 2013. – N. 4(26). – С. 3–20.

225. Алескеров Ф. Т., Белоусова В. Ю., Сердюк М. Ю., Солодков В. М. Стереотипы поведения российских банков// Банковское дело. – 2008. – N. 7. – С.44–50.

226. Determann D., Korfage I.J., Lambooij M.S., Bliemer M., Richardus J., Steyerberg E., de Bekker-Grob E. Acceptance of Vaccinations in Pandemic Outbreaks: A Discrete Choice Experiment // PLoS ONE. – 2014. – 9(7). – e102505.

227. Dong D., Xu R.H., Wong E.L., Humg C.T., Feng D, Feng Z., Yeoh E., Wong S. Public preference for COVID-19 vaccines in China: A discrete choice experiment // Health expectations: an international journal of public participation in health care and health policy. – 2020. – N 23(6). – P. 1543–1578.

228. Bansal P., Raj A., Mani Shukla D., Sunder N. COVID-19 vaccine preferences in India // Vaccine. – 2021. – N 40(15). – P. 2242–2246.

229. Афанасьев А.С., Бейсенбаева Л., Егорова Л.Г. Анализ предпочтений студентов ВШЭ по отношению к вакцинации// 2022 (готовится к публикации).

230. Афанасьев А.С., Бейсенбаева Л., Егорова Л.Г. Выявление предпочтений молодежи по отношению к вакцинации с помощью экспериментов дискретного выбора// 2022 (готовится к публикации).

231. Dong E., Du H., Gardner L. An interactive web-based dashboard to track COVID-19 in real time// The Lancet infectious diseases. – 2020. – V. 20(5). – P. 533–534.

232. Zu Z. Y. et al. Coronavirus disease 2019 (COVID-19): a perspective from China// Radiology. – 2020. – P. E15–E25.

233. Zhang Y. et al. Coagulopathy and antiphospholipid antibodies in patients with Covid-19// New England Journal of Medicine. – 2020. – V. 382(17). – P. e38.

234. Fang L., Karakiulakis G., Roth M. Are patients with hypertension and diabetes mellitus at increased risk for COVID-19 infection?// The Lancet. Respiratory Medicine. – 2020. – V. 8(4). – P. e21.

235. Li G., Fan Y., Lai Y., Han T., Li Z., Zhou P., Pan P., Wang W., Hu D., Liu X., Zhang Q., Wu J. Coronavirus infections and immune responses// Journal of medical virology. – 2020. – V. 92(4). – P. 424–432.

236. Zheng Y. Y. et al. COVID-19 and the cardiovascular system// Nature Reviews Cardiology. – 2020. – V. 17(5). – P. 259–260.

237. The World Bank Education Statistics – All Indicators. – https://databank.worldbank.org/databases/education

238. Edelman Trust Barometer. – 2020. –https://www.edelman.com/trustbarometer

239. The Global Competitiveness Report 2017-2018. –http://reports.weforum.org/global-competitiveness-index-2017-2018/

240. Johns Hopkins University. – https://coronavirus.jhu.edu/map.html

241. Coronavirus Government Response Tracker. – https://www.bsg.ox.ac.uk/research/research-projects/coronavirus-government-response-tracker

242. 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. – V. 17. – P. 687–694.

243. Myachin A. L. Pattern analysis in parallel coordinates based on pairwise comparison of parameters// Automation and Remote Control. – 2019. – V. 80(1). – P. 112–123.

244. COVID-19 Dashboard: Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU) [Электронный ресурс]. – 2022. – URL: https://www.arcgis.com/apps/dashboards/bda7594740fd40299423467b48e9ecf6 (дата обращения 04.11.2022).

245. Oxford COVID-19 Government Response Tracker [Электронный ресурс]. – 2022. – URL: https://www.bsg.ox.ac.uk/research/covid-19-government-response-tracker (дата обращения 04.11.2022).

246. Calculation of the Stringency Index: Oxford COVID-19 Government Response Tracker [Электронный ресурс]. – 2022. – URL: https://www.bsg.ox.ac.uk/sites/default/files/Calculation%20and%20presentation%20of%20the%20Stringency%20Index.pdf (дата обращения 04.11.2022).

247. Куликовский А.Г., Погорелов Н.В., Семенов А.Ю. Математические вопросы численного решения гиперболических систем уравнений. – 2-е изд. – М.: Физматлит. – 2012.

248. LeVeque R.J. Finite volume methods for hyperbolic problems. –Cambridge: Cambridge Univ. Press. – 2004.

249. Wesseling P. Principles of computational fluid dynamics. – Berlin: Springer. – 2009.

250. Abgrall R., Shu C.-W., eds. Handbook of numerical methods for hyperbolic problems: basic and fundamental issues. Handbook of Numerical Analysis. V. 17. – Amsterdam: North Holland. – 2016.

251. Четверушкин Б.Н. Кинетические схемы и квазигазодинамическая система уравнений. – М.: МАКС Пресс. – 2004.

252. Елизарова Т.Г. Квазигазодинамические уравнения и методы расчета вязких течений. – М.: Научный мир. – 2007.

253. Шеретов Ю.В. Динамика сплошных сред при пространственно-временном осреднении. – М., Ижевск: Регулярная и хаотическая динамика. – 2009.

254. Елизарова Т.Г., Широков И.А. Регуляризованные уравнения и примеры их использования при моделировании газодинамических течений. – М.: МАКС Пресс. – 2017.

255. Булатов О.В., Елизарова Т.Г. Регуляризованные уравнения мелкой воды и эффективный метод численного моделирования течений в неглубоких водоемах // Ж. вычисл. матем. матем. физ. – 2011. – Т. 51. – № 1. С. 170–184.

256. Balashov V., Zlotnik A., Savenkov E. Analysis of a regularized model for the isothermal two-component mixture with the diffuse interface // Russ. J. Numer. Anal. Math. Model. – 2017. – V. 32. – N 6. – P. 347–358.

257. Елизарова Т.Г., Злотник А.А., Истомина М.А. Гидродинамические аспекты формирования спиральновихревых структур во вращающихся газовых дисках // Астрон. журн. – 2018. – Т. 95. – № 1. – С. 11–21.

258. Balashov V., Zlotnik A. An energy dissipative spatial discretization for the regularized compressible NavierStokes-Cahn-Hilliard system of equations // Math. Model. Anal. – 2020. – V. 25. – N 1. – P. 110–129.

259. Сухомозгий А.А., Шеретов Ю.В. Анализ устойчивости одной разностной схемы решения уравнений Сен–Венана в теории мелкой воды // В сб.: Прилож. функц. анализа в теории приближений. Тверь: ТвГУ. – 2013. – С. 48–60.

260. Zlotnik A., Lomonosov T. On conditions for weak conservativeness of regularized explicit finite-difference schemes for 1D barotropic gas dynamics equations // In: Differential and Difference Equations with Applications. Springer Proc. in Math. and Stat. – 2018. – V. 230. – P. 635–647.

261. Злотник А.А., Ломоносов Т.А. Условия -диссипативности линеаризованных явных разностных схем с регуляризацией для уравнений 1D баротропной газовой динамики // Ж. вычисл. матем. матем. физ. – 2019. – Т. 59. – № 3. – С. 481–493.

262. Zlotnik A. On conditions for -dissipativity of an explicit finite-difference scheme for linearized 2D and 3D barotropic gas dynamics system of equations with regularizations // Symmetry. – 2021. – V. 13. – N 11. – Article 2184.

263. Попов И.В., Фрязинов И.В. Метод адаптивной искусственной вязкости численного решения уравнений газовой динамики. – М.: Красанд. – 2015.

264. Guermond J.-L., Popov B., Tomov V. Entropy–viscosity method for the single material Euler equations in Lagrangian frame // Comput. Meth. Appl. Mech. Eng. – 2016. – V. 300. – P. 402–426.

265. Feireisl E., Luká ová-Medvidová M., Mizerová H. A finite volume scheme for the Euler system inspired by the two velocities approach // Numer. Math. – 2020. – V. – 144. – N 1. – P. 89–132.

266. Dolejší V., Svärd M. Numerical study of two models for viscous compressible fluid flows // J. Comput. Phys. – 2021. – V. 427. – Article 110068.

267. Роуч П. Вычислительная гидродинамика. – М.: Мир. – 1980.

268. Van der Heul D.R., Vuik C., Wesseling P. Stability analysis of segregated solution methods for compressible flow // Appl. Num. Math. – 2001. – V. 38. – P. 257–274.

269. Bauer A.L., Loubere R., Wendroff B. On stability of staggered schemes // SIAM J. Numer. Anal. – 2008. – V. 46. – N 2. – P. 996–1011.

270. Konangi S., Palakurthi N.K., Ghia U. Von Neumann stability analysis of first-order accurate discretization schemes for one-dimensional (1D) and two-dimensional (2D) fluid flow equations // Comput. Math. Appl. – 2018. – V. 75. – P. 643–665.

271. Balashov V., Zlotnik A. An energy dissipative semi-discrete finite-difference method on staggered meshes for the 3D compressible isothermal Navier–Stokes–Cahn–Hilliard equations // J. Comput. Dynamic. – 2020. – V. 7. –N 2. – P. 291–312.

272. Злотник А.А., Ломоносов Т.А. диссипативность разностных схем для регуляризованных 1D баротропных уравнений движения газа при малых числах Маха // Матем. моделир. – 2021. – Т. 33. – № 5. – С. 16– 34.

273. Злотник А.А. Пространственная дискретизация одномерной баротропной квазигазодинамической системы уравнений и уравнение энергетического баланса // Матем. моделирование. – 2012. – Т. 24. – № 10. – С. 51–64.

ПРИЛОЖЕНИЕ А ТАБЛИЦЫ К ПОДРАЗДЕЛУ 5.2

Таблица А.1 − Отдельно рассмотренные случаи при применении алгоритма корректировки для категории товаров: 041 “Пшеница (включая полбу) и суржик, немолотые”

	Год	Откуда	Куда	Вес, кг	Комментарий
1	2018	Albania	United Arab Emirates	966	Усечённое среднее по всему экспорту/импорту ОАЭ
2	2018	Belarus	Egypt	55448410	Цена Египет ← РФ, 2018
3	2018	China	Norway	3166	Цена 2019
4	2018	Czechia	Nigeria	37933352	Усечённое среднее по всему экспорту/импорту Чехии
5	2018	Mongolia	Russian Federation	12404.0	Усечённое среднее по всему экспорту/импорту РФ
6	2018	Rwanda	Nigeria	10105916	Усечённое среднее по всему экспорту/импорту Нигерии
7	2018	South Africa	Cameroon	862	Усечённое среднее по всему экспорту/импорту ЮАР
8	2018	USA	Thailand	697557793.0	Данные от США
9	2019	Argentina	Angola	10521998	Цена 2021
10	2019	Australia	Chile	545	Усечённое среднее по всему экспорту/импорту Австралии
11	2019	Belarus	Egypt	61580728	Цена Египет ← РФ, 2018
12	2019	Brazil	Egypt	3177365	Цена 2021
13	2019	China	Finland	379	Цена 2018
14	2019	Cyprus	USA	6762	Усечённое среднее по всему экспорту/импорту Кипра
15	2019	Denmark	Czechia	106	Цена 2018
16	2019	Finland	Portugal	5012415.0	Данные от Португалии
17	2019	Finland	Morocco	26004473.0	Цена 2020
18	2019	France	Finland	3398568.0	Данные от Франции
19	2019	Germany	French Polynesia	325	Усечённое среднее по всему экспорту/импорту Германии
20	2019	Hungary	Czechia	805829	Цена 2018
21	2019	India	France	980427.0	Данные от Индии

Продолжение таблицы А.1

22	2019	Italy	Portugal	24943.0	Данные от Италии
23	2019	Japan	Philippines	1537	Усечённое среднее по всему экспорту/импорту Филиппин
24	2019	Latvia	Finland	68397610.0	Данные от Латвии
25	2019	Malaysia	Finland	200	Усечённое среднее по всему экспорту/импорту Малайзии
26	2019	Morocco	Belgium	97	Усечённое среднее по всему экспорту/импорту Бельгии
27	2019	Namibia	Angola	298	Усечённое среднее по всему экспорту/импорту Анголы
28	2019	Netherlands	Egypt	841408.0	Цена 2021
29	2019	Rep. of Moldova	Netherlands	1647050.0	Данные от Нидерландов
30	2019	Rep. of Moldova	Egypt	7866644	Цена 2021
31	2019	Rep. of Moldova	Philippines	1123480.0	Данные от Филиппин
32	2019	Rep. of Moldova	Malaysia	42184940.0	Данные от Малайзии
33	2019	Rep. of Moldova	Bangladesh	4245000.0	Данные от ФАО
34	2019	Singapore	Nigeria	72105	Усечённое среднее по всему экспорту/импорту Сингапура
35	2019	Singapore	Philippines	230.0	Данные от Сингапура
36	2019	Switzerland	Qatar	1638	Усечённое среднее по всему экспорту/импорту Швейцарии
37	2019	USA	Finland	135000.0	Данные от США
38	2019	United Arab Emirates	Australia	102057	Цена 2020
39	2019	United Kingdom	India	25000.0	Данные от Индии
40	2019	United Kingdom	Iraq	4300	Усечённое среднее по всему экспорту/импорту Великобритании
41	2019	Uruguay	Chile	346	Цена 2020

Продолжение таблица А.1

42	2020	Argentina	Angola	23638145	Цена 2021
43	2020	Czechia	Nigeria	49033472	Усечённое среднее по всему экспорту/импорту Чехии
44	2020	Denmark	Greece	3790	Усечённое среднее по всему экспорту/импорту Греции
45	2020	Estonia	Turkey	76998900.0	Данные от Эстонии
46	2020	Finland	Morocco	64518300.0	Данные от Марокко
47	2020	Finland	Poland	16196880.0	Данные от Польши
48	2020	Finland	Portugal	6050110.0	Данные от Португалии
49	2020	France	Switzerland	149944966.0	Данные от Франции
50	2020	France	Finland	2700278.0	Данные от Франции
51	2020	Hungary	Czechia	169555	Цена 2021
52	2020	Hungary	Nigeria	21670547	Усечённое среднее по всему экспорту/импорту Венгрии
53	2020	Ireland	Italy	603	Усечённое среднее по всему экспорту/импорту Италии
54	2020	Israel	United Kingdom	17336	Цена 2019
55	2020	Italy	Finland	1330.0	Данные от Италии
56	2020	Liberia	Angola	19865141.0	Усечённое среднее по всему экспорту/импорту Анголы
57	2020	Luxembourg	Romania	21927.0	Цена 2021
58	2020	New Zealand	United Arab Emirates	8	Усечённое среднее по всему экспорту/импорту Новой Зеландии
59	2020	New Zealand	Australia	8497.0	Данные от Новой Зеландии
60	2020	Serbia	Austria	515.0	Данные от Австрии
61	2020	Spain	Norway	673	Усечённое среднее по всему экспорту/импорту Испании
62	2020	Switzerland	Netherlands	3093	Цена 2019

Продолжение таблицы А.1

63	2020	USA	Czechia	5849	Цена 2021
64	2020	USA	Kenya	38999.0	Цена 2019
65	2020	USA	Finland	301000.0	Данные от США
66	2020	United Arab Emirates	Ethiopia	101808	Усечённое среднее по всему экспорту/импорту ОАЭ
67	2020	United Arab Emirates	USA	572	Цена 2021
68	2020	United Kingdom	Canada	3070.0	Данные от Канады
69	2020	United Kingdom	Israel	153386	Цена 2021
70	2020	United Kingdom	New Zealand	227	Цена 2019
71	2020	Zambia	Angola	1000.0	Данные от Замбии
72	2021	Argentina	Gambia	34293925	Усечённое среднее по всему экспорту/импорту Аргентины
73	2021	Argentina	Nigeria	491231186.0	Усечённое среднее по всему экспорту/импорту Аргентины
74	2021	Czechia	Nigeria	5809138	Усечённое среднее по всему экспорту/импорту Чехии
75	2021	Denmark	Saudi Arabia	11164.0	Усечённое среднее по всему экспорту/импорту Дании
76	2021	Egypt	Romania	769	Усечённое среднее по всему экспорту/импорту Румынии
77	2021	Estonia	United Arab Emirates	2224	Усечённое среднее по всему экспорту/импорту Эстонии
78	2021	Finland	Ireland	5020910	Цена 2020
79	2021	Italy	Finland	50367	Цена 2020
80	2021	Italy	Portugal	115.0	Данные от Италии
81	2021	Lithuania	Angola	41255000.0	Данные от Литвы
82	2021	Singapore	Nigeria	601973	Усечённое среднее по всему экспорту/импорту Сингапура
83	2021	Thailand	South Africa	1233	Усечённое среднее по всему экспорту/импорту Таиланда
84	2021	Turkey	Dominican Rep.	1354	Усечённое среднее по всему экспорту/импорту Турции

Продолжение таблицы А.1

85	2021	United Arab Emirates	Romania	1931	Усечённое среднее по всему экспорту/импорту ОАЭ
86	2021	United Arab Emirates	Ghana	123	Усечённое среднее по всему экспорту/импорту ОАЭ
87	2021	United Kingdom	Sweden	5450.0	Данные от Швеции
88	2021	United Kingdom	Thailand	966.0	Данные от Таиланда
89	2021	United Kingdom	Slovenia	13.0	Данные от Словении
90	2021	United Kingdom	Australia	6025	Цена 2020
91	2021	United Kingdom	China, Hong Kong SAR	38822.0	Усечённое среднее по всему экспорту/импорту Великобритании
92	2021	United Kingdom	Netherlands	29762710	Цена 2020
93	2021	United Kingdom	Turkey	9445	Цена 2020

Таблица А.2 − Отдельно рассмотренные случаи при применении алгоритма корректировки для категории товаров: 042 “Рис”

	Год	Откуда	Куда	Вес, кг	Комментарий
1	2018	China	Dominican Rep.	92481	Цена 2019
2	2018	Dominican Rep.	Anguilla	3122	Цена 2019
3	2018	Gambia	Senegal	44832	Это полностью реэкспорт, цена, по которой ранее импортировался этот товар
4	2018	Paraguay	El Salvador	1586	Усечённое среднее по всему экспорту/импорту Парагвая
5	2018	Qatar	Egypt	116	Усечённое среднее по всему экспорту/импорту Египта
6	2018	United Kingdom	Egypt	741	Цена 2019
7	2019	Brunei Darussalam	Egypt	152496	Усечённое среднее по всему экспорту/импорту Брунея
8	2019	Brunei Darussalam	Malaysia	3427.0	Данные от Малайзии
9	2019	Egypt	Finland	152183	Цена 2018

Продолжение таблицы А.2

10	2019	Greece	Egypt	26768	Цена 2018
11	2019	Latvia	Montenegro	9	Цена 2018
12	2019	Qatar	Egypt	705	Усечённое среднее по всему экспорту/импорту Египта
13	2019	Russian Federation	Finland	107902	Усечённое среднее по всему экспорту/импорту РФ
14	2019	Switzerland	Dominican Rep.	6	Цена 2019
15	2019	United Kingdom	Israel	120612.0	Данные от Израиля
16	2020	Germany	Central African Rep.	38351	Усечённое среднее по всему экспорту/импорту Германии
17	2020	Ireland	Belgium	1139	Цена 2021
18	2020	Pakistan	Finland	318000	Данные от Пакистана
19	2020	Russian Federation	Finland	142100	Усечённое среднее по всему экспорту/импорту РФ
20	2021	Brazil	Switzerland	51883981.0	Данные от Швейцарии
21	2021	Cambodia	Finland	1278286	Цена 2020
22	2021	China, Hong Kong SAR	Samoa	321003	Цена 2019
23	2021	Colombia	Costa Rica	59551	Усечённое среднее по всему экспорту/импорту Колумбии
24	2021	France	Finland	39408.0	Данные от Франции
25	2021	Greece	Finland	413470.0	Данные от Греции
26	2021	Lithuania	Finland	25150.0	Данные от Литвы
27	2021	North Macedonia	Finland	23552	Усечённое среднее по всему экспорту/импорту Северной Македонии за 2020 г.
28	2021	Pakistan	Finland	128000	Данные от Пакистана
29	2021	Russian Federation	Finland	180044	Усечённое среднее по всему экспорту/импорту РФ
30	2021	Senegal	Belgium	83	Цена 2020

Продолжение таблицы А.2

31	2021	Singapore	Belgium	124	Усечённое среднее по всему экспорту/импорту Сингапура
32	2021	Spain	Denmark	184389.0	Данные от Дании
33	2021	Thailand	Finland	2340404.0	Данные от Таиланда
34	2021	United Kingdom	Cyprus	82159.0	Данные от Кипра