

Школа - Инженерная школа информационных технологий и робототехники
 Направление подготовки – 01.03.02 Прикладная математика и информатика
 Отделение школы (НОЦ) – Отделение информационных технологий

БАКАЛАВРСКАЯ РАБОТА

Тема работы
Разработка алгоритма динамической кластеризации и его исследование в численном эксперименте

УДК 004.021-047.37:004.93'14:519.6

Студент

Группа	ФИО	Подпись	Дата
8Б51	Кавешников Артем Владимирович		

Руководитель ВКР

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОИТ ИШИТР	Шевелев Геннадий Ефимович	Доцент, к.ф.- м.н.		

КОНСУЛЬТАНТЫ ПО РАЗДЕЛАМ:

По разделу «Финансовый менеджмент, ресурсоэффективность и ресурсосбережение»

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОСГН ШБИП	Подопригора Игнат Валерьевич	Доцент, к.э.н.		

По разделу «Социальная ответственность»

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Ассистент ООТД	Мезенцева Ирина Леонидовна			

ДОПУСТИТЬ К ЗАЩИТЕ:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Руководитель ООП	Шевелев Геннадий Ефимович	Доцент, к.ф.-м.н.		

ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ОБУЧЕНИЯ

Код Результатов	Результат обучения (выпускник должен быть готов)	Требования ФГОС, Критерии АИОР
<i>Профессиональные компетенции</i>		
P1	Применять <i>глубокие математические и профессиональные знания</i> для решения задач научно-исследовательской, проектной, производственной и технологической деятельности в области системного и прикладного программирования.	Требования ФГОС (ОК-11, 12, ПК-3, 10), Критерий 5 АИОР (п. 5.2.1), согласованный с требованиями международных стандартов <i>EUR-ACE</i> и <i>FEANI</i> Требования профессиональных стандартов Ассоциации предприятий компьютерных и информационных технологий Требования работодателей: ФГУП «РФЯЦ-ВНИИТФ им. академика Е.И. Забабахина», ООО НАЦ «Недра», ИХН СО РАН
P2	Умение использовать знания по естественнонаучным дисциплинам при определении задач математического моделирования объектов и явлений в различных предметных областях	Требования ФГОС (ПК-3,9) Критерий 5 АИОР (п.5.2.3), согласованный с требованиями международных стандартов <i>EUR-ACE</i> и <i>FEANI</i> Требования работодателей: ФГУП «РФЯЦ-ВНИИТФ им. академика Е.И. Забабахина», ООО «НАПО им. В.П. Чкалова», ИХН СО РАН

P3	Демонстрировать понимание сущности и значения информации в развитии современного общества, владение основными методами, способами и средствами получения, хранения, переработки информации; использование для решения коммуникативных задач современных технических средств и информационных технологий.	Требования ФГОС (ОК-5, 11, 12,14,15, ПК-2, 6), Критерий 5 АИОР (п. 5.2.2), согласованный с требованиями международных стандартов <i>EUR-ACE</i> и <i>FEANI</i> Требования профессиональных стандартов Ассоциации предприятий компьютерных и информационных технологий Требования работодателей: Контек, ОАО «Газпром переработка», ООО Нижневартовскэнергонефть».
P4	Выполнять <i>инновационные</i> проекты с применением <i>глубоких профессиональных</i> знаний и <i>эффективных</i> методов проектирования для достижения <i>новых</i> результатов, обеспечивающих <i>конкурентные преимущества</i> в условиях экономических, экологических, социальных и других ограничений.	Требования ФГОС (ОК-14, ПК- 7, 9,14), Критерий 5 АИОР (п. 5.2.4), согласованный с требованиями международных стандартов <i>EUR-ACE</i> и <i>FEANI</i> Требования профессиональных стандартов Ассоциации предприятий компьютерных и информационных технологий. Требования работодателей: Контек, ОАО «Газпром переработка», ИХН СО РАН.
P5	<i>Демонстрировать знание</i> о формах организации образовательной и научной деятельности в высших учебных заведениях, <i>иметь навыки</i> преподавательской работы.	Требования ФГОС (ОК-1, 10, 16, ПК-1, 14, 15), Критерий 5 АИОР (п. 5.2.1), согласованный с требованиями международных стандартов <i>EUR-ACE</i> и <i>FEANI</i>
P6	Способность осуществлять организационно-управленческую и социально-ориентированную деятельность с соблюдением профессиональной этики	Требования ФГОС (ОК-5,13,16, ПК-11-13,16) Критерий 5 АИОР (п.5.2.12-13) согласованный с требованиями международных стандартов <i>EUR-ACE</i> и <i>FEANI</i>
Универсальные компетенции		

P7	<i>Активно владеть иностранным языком на уровне, позволяющем работать в интернациональной среде, включая разработку документации и представление результатов инновационной деятельности. Толерантность в восприятии социальных и культурных различий.</i>	Требования ФГОС (ОК-2, 3,4, 7, ПК-8). Критерий 5 АИОР (п. 5.2.11), согласованный с требованиями международных стандартов <i>EUR-ACE</i> и <i>FEANI</i> и Требования профессиональных стандартов Ассоциации предприятий компьютерных и информационных технологий
P8	<i>Эффективно работать индивидуально, в качестве члена и руководителя группы, состоящей из специалистов различных направлений и квалификаций, демонстрировать ответственность за результаты работы и готовность следовать корпоративной культуре организации</i>	Требования ФГОС (ОК-1,4, 6, ПК-8,11,12), Критерий 5 АИОР (пп. 5.2.9,5.2.13), согласованный с требованиями международных стандартов <i>EUR-ACE</i> и <i>FEANI</i> и Требования профессиональных стандартов Ассоциации предприятий компьютерных и информационных технологий. Требования работодателей: Контек, ОАО «Газпром переработка», ООО Нижневартовскэнергонефть».
P9	<i>Самостоятельно учиться и непрерывно повышать квалификацию в течение всего периода профессиональной деятельности.</i> Способность к интеллектуальному, культурному, нравственному и профессиональному саморазвитию.	Требования ФГОС (ОК-8,9,16, ПК-5, 11), Критерий 5 АИОР (5.2.14), согласованный с требованиями международных стандартов <i>EUR-ACE</i> и <i>FEANI</i> . Требования работодателей: Контек, ОАО «Газпром переработка», ООО Нижневартовскэнергонефть».

Министерство науки и высшего образования Российской Федерации
 федеральное государственное автономное
 образовательное учреждение высшего образования
 «Национальный исследовательский Томский политехнический университет» (ТПУ)

Школа - Инженерная школа информационных технологий и робототехники
 Направление подготовки – 01.03.02 Прикладная математика и информатика
 Уровень образования – Бакалавриат
 Отделение школы (НОЦ) – Отделение информационных технологий
 Период выполнения - осенний / весенний семестр 2018 /2019 учебного года

Форма представления работы:

Бакалаврская работа

(бакалаврская работа, дипломный проект/работа, магистерская диссертация)

КАЛЕНДАРНЫЙ РЕЙТИНГ-ПЛАН выполнения выпускной квалификационной работы

Срок сдачи студентом выполненной работы:	
--	--

Дата контроля	Название раздела (модуля) / вид работы (исследования)	Максимальный балл раздела (модуля)
	Основная часть	75
	Финансовый менеджмент, ресурсоэффективность и ресурсосбережение	15
	Социальная ответственность	10

СОСТАВИЛ:

Руководитель ВКР

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОИТ ИШИТР	Шевелев Геннадий Ефимович	Доцент, к.ф.-м.н.		

СОГЛАСОВАНО:

Руководитель ООП

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Руководитель ООП	Шевелев Геннадий Ефимович	Доцент, к.ф.-м.н.		

Министерство науки и высшего образования Российской Федерации
 федеральное государственное автономное
 образовательное учреждение высшего образования
 «Национальный исследовательский Томский политехнический университет» (ТПУ)

Школа - Инженерная школа информационных технологий и робототехники
 Направление подготовки – 01.03.02 Прикладная математика и информатика
 Уровень образования – Бакалавриат
 Отделение школы (НОЦ) – Отделение информационных технологий

УТВЕРЖДАЮ:
 Руководитель ООП ИШИТР
 _____ Шевелев Г.Е.
 (Подпись) (Дата) (Ф.И.О.)

ЗАДАНИЕ на выполнение выпускной квалификационной работы

В форме:

Бакалаврской работы

(бакалаврской работы, дипломного проекта/работы, магистерской диссертации)

Студенту:

Группа	ФИО
8Б51	Кавешников Артем Владимирович

Тема работы:

Разработка алгоритма динамической кластеризации и его исследование в численном эксперименте.

Утверждена приказом директора (дата, номер)

Срок сдачи студентом выполненной работы:

ТЕХНИЧЕСКОЕ ЗАДАНИЕ:

Исходные данные к работе

(наименование объекта исследования или проектирования; производительность или нагрузка; режим работы (непрерывный, периодический, циклический и т. д.); вид сырья или материал изделия; требования к продукту, изделию или процессу; особые требования к особенностям функционирования (эксплуатации) объекта или изделия в плане безопасности эксплуатации, влияния на окружающую среду, энергозатратам; экономический анализ и т. д.).

Предметом исследования является алгоритм динамической кластеризации, работающий с колебательными сигналами. Преимущество данного алгоритма заключается в отсутствии необходимости заранее задавать количество кластеров и простоте работы.

Перечень подлежащих исследованию, проектированию и разработке вопросов <i>(аналитический обзор по литературным источникам с целью выяснения достижений мировой науки техники в рассматриваемой области; постановка задачи исследования, проектирования, конструирования; содержание процедуры исследования, проектирования, конструирования; обсуждение результатов выполненной работы; наименование дополнительных разделов, подлежащих разработке; заключение по работе).</i>	1. Изучить существующие методы группировки; 2. Разработать новый расчетный метод; 3. Протестировать новый метод; Выявить его отличия; 4. Провести оценку устойчивости алгоритма; 5. Сделать выводы об устойчивости алгоритма;
Консультанты по разделам выпускной квалификационной работы <i>(с указанием разделов)</i>	
Раздел	Консультант
Финансовый менеджмент, ресурсоэффективность и ресурсосбережение	Подопригора Игнат Валерьевич, доцент ОСГН ШБИП
Социальная ответственность	Мезенцева Ирина Леонидовна, ассистент ООТД
Названия разделов, которые должны быть написаны на русском и иностранном языках:	
Нет	

Дата выдачи задания на выполнение выпускной квалификационной работы по линейному графику	
--	--

Задание выдал руководитель:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОИТ ИШИТР	Шевелев Геннадий Ефимович	Доцент, к.ф.-м.н.		

Задание принял к исполнению студент:

Группа	ФИО	Подпись	Дата
8Б51	Кавешников Артем Владимирович		

Министерство науки и высшего образования Российской Федерации
 федеральное государственное автономное
 образовательное учреждение высшего образования
 «Национальный исследовательский Томский политехнический университет» (ТПУ)

Школа - Инженерная школа информационных технологий и робототехники
 Направление подготовки – 01.03.02 Прикладная математика и информатика
 Уровень образования – Бакалавриат
 Отделение школы (НОЦ) – Отделение информационных технологий

ЗАДАНИЕ ДЛЯ РАЗДЕЛА «ФИНАНСОВЫЙ МЕНЕДЖМЕНТ, РЕСУРСОЭФФЕКТИВНОСТЬ И РЕСУРСОСБЕРЕЖЕНИЕ»

Студенту:

Группа	ФИО
8Б51	Кавешников Артем Владимирович

Школа	ИШИТР	Отделение (НОЦ)	Информационных технологий
Уровень образования	Бакалавриат	Направление/специальность	Прикладная математика и информатика

Исходные данные к разделу «Финансовый менеджмент, ресурсоэффективность и ресурсосбережение»:

1. Стоимость ресурсов научного исследования (НИ): материально-технических, энергетических, финансовых, информационных и человеческих	Человеческие ресурсы – 2 человека (руководитель и студент-дипломник).
--	---

Перечень вопросов, подлежащих исследованию, проектированию и разработке:

1. Планирование и формирование бюджета научных исследований	Необходимо составить календарный план НИ и рассчитать затраты НИ.
2. Определение ресурсной (ресурсосберегающей), финансовой, бюджетной, социальной и экономической эффективности исследования	Расчет интегральных показателей эффективности исследования, выбор наилучшего исполнения.

Перечень графического материала:

1. Оценка конкурентоспособности технических решений
2. Матрица SWOT
3. График проведения и бюджет НИ
4. Оценка ресурсной, финансовой и экономической эффективности НИ

Дата выдачи задания для раздела по линейному графику	
--	--

Задание выдал консультант:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент	Подопригора Игнат Валерьевич	Кандидат экономических наук		

Задание принял к исполнению студент:

Группа	ФИО	Подпись	Дата
8Б51	Кавешников Артем Владимирович		

Министерство науки и высшего образования Российской Федерации
 федеральное государственное автономное
 образовательное учреждение высшего образования
 «Национальный исследовательский Томский политехнический университет» (ТПУ)

Школа - Инженерная школа информационных технологий и робототехники
 Направление подготовки – 01.03.02 Прикладная математика и информатика
 Уровень образования – Бакалавриат
 Отделение школы (НОЦ) – Отделение информационных технологий

ЗАДАНИЕ ДЛЯ РАЗДЕЛА «СОЦИАЛЬНАЯ ОТВЕТСТВЕННОСТЬ»

Студенту:

Группа	ФИО
8Б51	Кавешников Артем Владимирович

Школа	ИШИТР	Отделение (НОЦ)	Информационных технологий
Уровень образования	Бакалавриат	Направление/специальность	Прикладная математика и информатика

Исходные данные к разделу «Социальная ответственность»:

1. Характеристика объекта исследования (вещество, материал, прибор, алгоритм, методика, рабочая зона) и области его применения.	1. Алгоритм кластеризации. Область применения: климатическое зонирование, распознавание образов.
---	---

Перечень вопросов, подлежащих исследованию, проектированию и разработке:

1. Производственная безопасность 1.1. Анализ выявленных вредных факторов при разработке и эксплуатации проектируемого решения в следующей последовательности. 1.2. Анализ выявленных опасных факторов при разработке и эксплуатации проектируемого решения в следующей последовательности.	1.1 Рассмотрены вредные факторы: Отклонение показателей микроклимата рабочей зоны; повышенный уровень шума на рабочем месте; отсутствие или недостаток естественного света, а также недостаточная освещенность рабочей зоны. 1.2 Рассмотрены опасные факторы: повышенное значение напряжения в электрической цепи, замыкание которой может произойти через тело человека; возгорание эксплуатируемого оборудования.
2. Экологическая безопасность: 2.1 Анализ воздействия объекта на окружающую среду; 2.2 Разработать решения по обеспечению экологической безопасности со ссылками на НТД по охране окружающей среды.	2.1 Рассмотрены негативно влияющие на экологию факторы при эксплуатации компьютера. 2.2 Решения по обеспечению экологической безопасности согласно нормативным документам.

3. Безопасность в чрезвычайных ситуациях: 3.1 Перечень возможных ЧС при разработке и эксплуатации проектируемого решения; 3.2 Разработка действий в результате возникшей ЧС и мер по ликвидации её последствий.	3.1 Перечень возможных ЧС, которые могут возникнуть при работе в помещении офиса. 3.2 Способы защиты от пожара и ликвидация последствий.
4. Правовые и организационные вопросы обеспечения безопасности: 4.1 Организационные мероприятия при компоновке рабочей зоны.	4.1 Организационные мероприятия по обеспечению безопасности трудящихся за персональным компьютером.

Задание выдал консультант:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Ассистент ООТД ШБИП	Мезенцева Ирина Леонидовна			

Задание принял к исполнению студент:

Группа	ФИО	Подпись	Дата
8Б51	Кавешников Артем Владимирович		

РЕФЕРАТ

Пояснительная записка содержит 72 страниц машинописного текста, 18 таблиц, 9 рисунков, 12 список использованных источников.

Цель работы: Разработать новый метод расчета данных в алгоритме, провести оценку устойчивости алгоритма динамической кластеризации.

В данной работе рассмотрены наиболее простые и фундаментальные методы группировки и сферы их применения. Проанализированы методы группировки, выявлены их достоинства и недостатки.

Предложен и обоснован метод расчета типовых сигналов в итерационном алгоритме кластеризации.

Проведена оценка устойчивости алгоритма с разным видом и уровнем шума.

Ключевые слова: кластеризация, методы группировки, оценка устойчивости, разработка метода.

Оглавление

Введение	14
Обзор методов группировки	15
2. Алгоритм динамической группировки	19
2.1 Описание алгоритма	19
2.2 Разработка способа выбора данных при расчете средних в алгоритме динамической группировки	20
3. Исследование алгоритма динамической кластеризации в численном эксперименте	21
3.1. Замкнутый численный эксперимент. Параметры	21
3.1.1 Построение численного исследования алгоритма кластеризации	21
3.1.2. Результаты численного эксперимента	25
4. Социальная ответственность	42
4.1 Правовые и организационные вопросы обеспечения безопасности	42
4.2 Производственная безопасность	46
4.3 Анализ опасных и вредных производственных факторов.	48
4.4 Экологическая безопасность	51
4.5 Обоснование мероприятий по снижению уровней воздействия опасных и вредных факторов на исследователя.	51
4.6 Безопасность в чрезвычайных ситуациях	53
4.7 Выводы и рекомендации	53
5. Финансовый менеджмент, ресурсоэффективность и ресурсосбережение	54
5.1 Оценка коммерческого и инновационного потенциала научных исследований	54
5.1.1 Потенциальные потребители результатов исследований	54
5.1.2 Анализ конкурентных технических решений	55
5.1.3 SWOT – анализ	55
5.2 Определение возможных альтернатив проведения научных исследований	57
5.3 Планирование научно-исследовательских работ	57
5.3.1 Структура работ в рамках научного исследования	57
5.3.2 Определение трудоемкости выполнения работ	58
5.3.3 Разработка графика проведения научного исследования	59
5.4 Бюджет научно-технического исследования (НТИ)	61
5.4.1 Расчет затрат на специальное оборудование для научных (экспериментальных) работ	61
5.4.2 Основная заработная плата исполнителей темы	62
5.4.3 Дополнительная заработная плата	63
5.4.4 Отчисления во внебюджетные фонды (страховые отчисления)	63

5.4.5 Накладные расходы	64
5.4.6 Формирование бюджета затрат научно-исследовательского проекта	64
5.5 Определение ресурсной (ресурсосберегающей), финансовой, бюджетной, социальной и экономической эффективности исследования	65
Заключение.....	68
Список использованных источников.....	69
Приложения	70
Приложение А – последовательность алгоритма	70
Приложение Б – Блок-схема алгоритма	71

Введение

Одной из наиболее значимых и глобальных проблем современности на данный момент можно назвать постоянно растущий объем информации, который требует определенной систематизации, упрощения и выделения её существенной части. С развитием мощностей аппаратного обеспечения и интернет технологий объем цифровых данных растет в огромных масштабах и исчисляется терабайтами. Обрабатывать такие данные вручную представляется труднодоступным, а существующие методы могут оказаться неэффективными. Поэтому для решения подобных задач требуются все более и более новые методы обработки данных. Современные методы должны уметь эффективно анализировать, систематизировать и собирать получаемую информацию в группы, при этом с достаточно высокой точностью.

Методы, позволяющие анализировать большие массивы данных, обладают широким спектром применения. Так в медицине по набору кластерных симптомов можно с достаточно высокой точностью устанавливать диагноз и назначать терапию, в экономике по набору кластерных параметров можно выявлять группы потребителей, их поведение и их потребительскую корзину, в метеорологии кластерный анализ позволяет выявлять климатические зоны и предсказывать их изменение. С помощью алгоритмов кластеризации возможна реализация задачи распознавания образов, также существует достаточно высокая потребность в обработке больших объемов данных в научных исследованиях. Исходя из выше сказанного, можно сделать вывод, что спрос на алгоритмы кластеризации и их исследование достаточно высок.

Обзор методов группировки

Кластер – группа похожих объектов.

Кластеризация – это процесс организации объектов на группы, элементы, которые схожи в некотором роде.

Кластерный анализ – это семейство алгоритмов, разработанных для формирования групп таким образом, чтобы члены группы были наиболее похожими друг на друга и не похожими на элементы, не входящие в группу.

Кластер и группа – синонимы в мире кластерного анализа [1].

Метод k-средних

Сущность метода заключается в том, что процесс классификации начинается с задания некоторых начальных условий (количество образуемых кластеров, порог завершения процесса классификации и т.д.). Название метода было предложено Дж. Мак-Куином в 1967 г. В отличие от иерархических процедур метод k - средних не требует вычисления и хранения матрицы расстояний или сходств между объектами. Алгоритм этого метода предполагает использование только исходных значений переменных. Для начала процедуры классификации должны быть заданы k выбранных объектов, которые будут служить эталонами, т.е. центрами кластеров. Считается, что алгоритмы эталонного типа удобные и быстродействующие. В этом случае важную роль играет выбор начальных условий, которые влияют на длительность процесса классификации и на его результаты. Метод k - средних удобен для обработки больших статистических совокупностей.

Математическое описание алгоритма метода k - средних.

Пусть имеется n наблюдений, каждое из которых характеризуется m признаками $X_1, X_2, \dots, X_m$. Эти наблюдения необходимо разбить на k кластеров. Для начала из n точек исследуемой совокупности отбираются случайным образом или задаются исследователем исходя из каких-либо априорных соображений k точек (объектов). Эти точки принимаются за эталоны. Каждому эталону присваивается порядковый номер, который одновременно является и номером кластера. На первом шаге из оставшихся $(n - k)$ объектов

извлекается точка X_i с координатами $(x_1, x_2, \dots, x_n)$ и проверяется, к какому из эталонов (центров) она находится ближе всего. Для этого используется одна из метрик, например, евклидово расстояние. Проверяемый объект присоединяется к тому центру (эталону), которому соответствует минимальное из расстояний. Эталон заменяется новым, пересчитанным с учетом присоединенной точки, и вес его (количество объектов, входящих в данный кластер) увеличивается на единицу. Если встречаются два или более минимальных расстояния, то i -ый объект присоединяют к центру с наименьшим порядковым номером. На следующем шаге выбираем точку X_{i+1} и для нее повторяются все процедуры. Таким образом, через $(n - k)$ шагов все точки (объекты) совокупности 21 окажутся отнесенными к одному из k кластеров, но на этом процесс разбиения не заканчивается. Для того чтобы добиться устойчивости разбиения по тому же правилу, все точки $X_1, X_2, \dots, X_n$ опять подсоединяются к полученным кластерам, при этом веса продолжают накапливаться. Новое разбиение сравнивается с предыдущим. Если они совпадают, то работа алгоритма завершается. В противном случае цикл повторяется. Окончательное разбиение имеет центры тяжести, которые не совпадают с эталонами, их можно обозначить $C_1, C_2, \dots, C_k$. При этом каждая точка X_i ($i = 1, 2, \dots, n$) будет относиться к такому кластеру (классу) l , для которого расстояние минимально. Возможны две модификации метода k - средних. Первая предполагает пересчет центра тяжести кластера после каждого изменения его состава, а вторая – лишь после того, как будет завершен просмотр всех данных. В обоих случаях итеративный алгоритм этого метода минимизирует дисперсию внутри каждого кластера, хотя в явном виде такой критерий оптимизации не используется.

Многофакторный анализ позволяет создать целостную картину состояния исследуемой системы без большого числа наблюдений: неполнота отдельных наблюдений во многих случаях компенсируется материалами изучения внутренних связей между несколькими варьирующими признаками. Для этого разнородные группы изолированных наблюдений объединяют в

многофакторный комбинационный комплекс той или иной качественной структуры с учетом целей и особенностей объектов исследования. Решение этой задачи предполагает знание специфики объекта исследования, понимание закономерностей происходящих в нем процессов и качественных характеристик конкретных факторов [2].

Метод главных компонент

Метод главных компонент выявляет k компонент — факторов, объясняющих всю дисперсию и корреляции исходных k случайных величин; при этом компоненты строятся в порядке убывания объясняемой ими доли суммарной дисперсии исходных величин, что позволяет зачастую ограничиться несколькими первыми компонентами. Факторный анализ выявляет m ($m < k$) общих для всех исходных величин факторов, объясняя оставшуюся после этого дисперсию влиянием специфических факторов.

Среди прикладных задач, решаемых указанными методами, отметим следующие:

- поиск скрытых, но объективно существующих взаимосвязей между экономическими и социальными показателями;
- проверка гипотез о взаимосвязях этих показателей, выявление природы различий между объектами;
- описание изучаемой системы числом признаков, значительно меньшим числа исходных факторов, при этом выявленные факторы или главные компоненты содержат в среднем больше информации, чем непосредственно зафиксированные на объектах значения исходных факторов;
- построение обобщенных экономических и социальных показателей, таких как качество продукции, размер предприятия, интенсивность ведения хозяйства и т. п.;
- визуализация исходных многомерных наблюдений путем их проецирования на специально подобранную прямую, плоскость или трехмерное пространство;

– построение регрессионных моделей по главным компонентам; в социальных и экономических задачах исходные факторы часто обладают мультиколлинеарностью (т. е. являются коррелированными между собой), что затрудняет построение и интерпретацию регрессионных моделей, не позволяя часто получать сколь-нибудь точные прогнозы; главные компоненты, сохраняя всю информацию об изучаемых объектах, являются некоррелированными по построению;

– классификация по обобщенным показателям; практика показывает, что классификация объектов, проведенная по факторам или по главным компонентам, оказывается более объективной, чем классификация тех же объектов по исходным признакам; по одному — трем факторам или главным компонентам возможно проведение визуальной классификации, в случае большей размерности пространства обобщенных показателей, полученного в результате компонентного или факторного анализа, необходимо привлечение методов многомерной классификации, рассматриваемых во второй главе пособия;

– сжатие исходной информации, значительное уменьшение объемов информации, хранимой в базах данных, без существенных потерь в информативности [3].

2. Алгоритм динамической группировки

2.1 Описание алгоритма

Алгоритм разработан для анализа климатических данных и использовался для работы с квази–периодическими сигналами на примере сигналов, характеризующих изменения среднемесячной температуры.

В качестве признаков группировки использованы амплитуда и фаза сигнал. В основе группировки лежит принцип синхронного поведения сигналов. В качестве критериев отнесения сигнала к группе используется коэф корреляции.

Вычисленные характеристики, фазы или огибающие, образуют множество M^i , номера элементов которого соответствуют метеостанциям с номерами $j = 1, 2, \dots, N$ или $k = 1, 2, \dots, N$, $N = 818$, i – порядковый номер итерации. Процедура кластеризации состоит в вычислении корреляционной матрицы $\|R_{j,k}\|$ элементов множества и последующего разбиения этого множества на кластеры. Каждый кластер формируется из элементов, соответствующих условию $R_{j,k} = \langle M^i_j, M^i_k \rangle \geq r$, где r требуемый уровень взаимной корреляции. Для каждого элемента множества M^i , попавшего в один из кластеров, определяется «текущая» характеристика, которая вычисляется как среднее значение всех элементов вошедших в кластер. Формируется новое множество M^{i+1} , элементами которого являются рассчитанные «текущие» характеристики. Определяется уровень отличия между массивами путем расчета абсолютной разницы между элементами массивов M^i и M^{i+1} . В случае превышения максимального значения абсолютной разницы между элементами заданного уровня точности $\varepsilon = \pm 0,001$, повторяется итерационный процесс, начинающийся с расчета вычисления корреляционной матрицы $\|R_{j,k}\|$ для элементов массива M^{i+1} . Итерационный процесс останавливается при достижении заданного уровня точности ε .

Сходимость итерационного процесса следует из двух условий: конечности выборки (число метеостанций равно N) и справедливости неравенства $|\langle M^i_j \rangle| \leq \max |M^i_j|$. Число групп не может быть больше числа

элементов. В итерационном процессе значения текущих элементов из любого подмножества выборки стремятся к предельному числу кластеров. Количество кластеров зависит от коррелированности участвующих в итерационном процессе исходных и «текущих» элементов, а также от заданного уровня корреляции r .

Полученные «текущие» функции характеризуют климатические закономерности, вокруг которых формируются климатические кластеры. В полученные кластеры включаются климатические станции, для которых уровень корреляций исследуемых характеристик температурных сигналов с «текущей» удовлетворяет заданному критерию r . Кластеры с одинаковым составом станций учитываются один раз.

2.2 Разработка способа выбора данных при расчете средних в алгоритме динамической группировки

В ходе исследования и работы с текущим алгоритмом было повысить вычислительную точность алгоритма. В исходном методе вычисление средних реализовано для текущей итерации из массива средних значений сигналов предыдущей итерации. Вычисление итоговых средних дает погрешность.

Предложено проводить вычисление средних в алгоритме всегда из массива исходных сигналов.

3. Исследование алгоритма динамической кластеризации в численном эксперименте

Для оценки устойчивости реализуется численный эксперимент.

Очень часто процессы, подлежащие исследованию при помощи алгоритма, формируются под воздействием большого числа независимых случайных факторов, при незначительном воздействии каждого отдельного фактора. Основной задачей алгоритма является выделение отдельных кластеров и восстановление исходных сигналов, лежащих в основе изучаемых процессов.

Создание в численном эксперименте имитации влияния независимых случайных факторов, позволит получить достаточные оценки, характеризующие работоспособность численного алгоритма кластеризации.

Реализация численного эксперимента позволит сделать выводы об эффективности, устойчивости и о наличии ограничений для применения предложенного алгоритма.

3.1 Замкнутый численный эксперимент. Параметры

Целью численного эксперимента является оценка точности (ошибки или погрешности) восстановлении сигналов и устойчивости алгоритма при наличии шумов различных видов и уровней.

Основным важным условием для построения численного эксперимента и ограничением является то, что разработанный алгоритм применим только для сигналов, имеющими форму колебательного процесса.

Вторым важным условием для кластеризации является то, что сигналы, принадлежащие к одному кластеру, должны иметь сходство. Все отличия сигналов должны определяться величиной и формой шума.

3.1.1 Построение численного исследования алгоритма кластеризации

Основная задача численного эксперимента — получение результатов работы алгоритма в разных условиях при разных уровнях и видах шума вносимого в исходный (эталонный) сигнал.

Изменяемые параметры: величина шума, вид или способ ввода (аддитивный и мультипликативный) и форма шума (задается шириной спектрального окна), форма исходных (эталонных) сигналов, уровень корреляции, количество сигналов в выборке (на один исходный).

Алгоритм работает с огибающей и фазой сигнала. Данные характеристики имеют отличия, поэтому численный эксперимент проведен отдельно для фазы и огибающей.

Порядок проведения численного эксперимента включает следующие операций:

1. Формирование эталонных сигналов.

Исходный Модельный сигнал состоит из несущего сигнала и имеет модуляцию амплитуды и фазы. Учитывая то, что в большинстве случаев модулирующие сигналы являются сложными функциями времени, их можно представить в виде конечной или бесконечной суммы гармонических составляющих.

В качестве модели эталонного сигнала огибающей использовалась функция следующего вида:

$$A(\tau) = a_n + \sum_{i=1}^N k_i^A a_i(\tau) = a_n + \sum_{i=1}^N K_i^A \cos(2\pi \nu_i \tau) \quad , \quad (3.1)$$

где, τ – время, a_n – амплитуда несущего колебания, K_i^A – коэффициент пропорциональности (отражает глубину амплитудной модуляции), $K_i^A < a_n$, $a_i(\tau)$ – сигнал модулирующий амплитуду, ν_i – частота модулирующего сигнала, i – порядковый номер гармоники, $i=1,2..N$, N – количество модулирующих гармоник.

В качестве модели сигнала полной фазы использовалась функция вида:

$$\Phi(\tau) = \varphi_n(\tau) + \sum_{i=1}^N I_i^\Phi \varphi_i(\tau) = \varphi_n(\tau) + \sum_{i=1}^N I_i^\Phi \cos(2\pi \nu_i \tau) \quad , \quad (3.2)$$

где τ – время, $\varphi_n(\tau)$ – фаза несущего колебания, I_i^Φ – индекс фазовой модуляции (определяет максимальное отклонение фазы), $\varphi_i(\tau)$ – сигнал

модулирующий фазу, ν_i – частота сигнала модулирующего фазу, i – порядковый номер гармоники, $i = 1, 2..N$, N – количество модулирующих гармоник.

Для реализации алгоритма кластеризации при построении модели эталонного сигнала фазы из полной фазовой функций (2) вычитается фаза несущего колебания:

$$\Phi'(\tau) = \Phi(\tau) - \varphi_n(\tau) = \sum_{i=1}^N I_i^\Phi \cos(2\pi \nu_i \tau) \quad (3.3)$$

Форма эталонного модулирующего сигнал (фазы или огибающей) варьировалась через изменение количество гармоник N . Были сформированы сигналы с $N = 2, 4, 6, 12, 24$. Частоты гармоник задавались в интервале от 6 до 30 частотных полос.

2. Формирование сигнала с шумом

Шум задавался в частотной области в интервале от 6 до 30 частотных полос. Интервал частот наложения шума выбран равным интервалу частот основных гармонических составляющих эталонного сигнала. Для генерации шума применялся датчик случайных чисел программы «Mathcad», который создавал вектор из заранее заданного количества случайных чисел имеющих нормальное распределение.

Сигнал с шумом формировался как исходный сигнал, содержащий аддитивную или мультипликативную составляющую шума.

Уровень шума определялся как отношение СКО шума к СКО исходного сигнала.

$$P_A = \sqrt{\frac{\sum_{i=1}^N (S_i - \bar{S})^2}{\sum_{i=1}^N (A_i - \bar{A})^2}} \cdot 100\% \quad , \quad P_\Phi = \sqrt{\frac{\sum_{i=1}^N (S_i - \bar{S})^2}{\sum_{i=1}^N (\Phi'_i - \bar{\Phi}')^2}} \cdot 100\% \quad (3.4)$$

где

A, Φ' – эталонная огибающая, эталонная фаза,

$\bar{A}, \bar{\Phi}'$ – среднеарифметическое значение эталонной огибающей, фазы,

S – «сигнал-шум»,

$\bar{S}$, – среднеарифметическое значение для «сигнал-шум»

3. Формирование выборки и расчет ошибки

Количество эталонных сигналов для одновременной реализации алгоритма кластеризации выбрано равным 4. Для каждого эталонного сигнала формировалась выборка из 25 эталонных сигналов шумом. Таким образом, общий массив данных формировался из 100 сигналов с шумом.

Алгоритм динамической кластеризации, описанный ранее в параграфе 3, применяется одновременно ко всему общему массиву сигналов.

В результате вычислений получаем восстановленные сигналы огибающих и (или) фаз. Ошибки восстановления рассчитывались как оценки нормированных среднеквадратических отклонений по следующим формулам:

$$\varepsilon_A = \sqrt{\frac{\sum_{i=1}^N (\Delta A_i - \bar{\Delta A})^2}{\sum_{i=1}^N (A_i^e - \bar{A}^e)^2}} \cdot 100\% \quad , \quad \varepsilon_\Phi = \sqrt{\frac{\sum_{i=1}^N (\Delta \Phi_i - \bar{\Delta \Phi})^2}{\sum_{i=1}^N (\Phi_i^e - \bar{\Phi}^e)^2}} \cdot 100\% \quad (3.5)$$

где

A, Φ – эталонная огибающая, фаза,

$\Delta A, \Delta \Phi$ – разность между эталонной и восстановленной огибающей,

$\Delta A = A^T - A$, фазой $\Delta \Phi = \Phi^T - \Phi$,

$\bar{\Delta A}, \bar{\Delta \Phi}$ – среднеарифметическое значение разности огибающих, фаз,

$\bar{A}, \bar{\Phi}$ – среднеарифметическое значение эталонной огибающей, фазы,

N – количество отсчётов.

Оценка нормированной среднеквадратической ошибки восстановления фазы рассчитывалась при изменении отношения шума к сигналу от 0 до 100 процентов.

4. Параметры численного эксперимента

В ходе численного эксперимента изменялись следующие параметры: способ наложения шума (аддитивно или мультипликативно), уровень шума (в процентах от эталонного сигнала), интервал частот наложения шума, форма эталонного сигнала (варьировалось число гармоник), размер выборки (задавалось количество сигналов с шумом приходящееся на один эталонный сигнал), уровень корреляции между сигналами. Численный эксперимент проведен для фаз и огибающих отдельно.

Исследование влияния уровня шума на оценку нормированной среднеквадратической ошибки восстановления фазы и огибающей проводилось отдельно для мультипликативного и аддитивного шума.

Одновременно задавалось четыре эталонных сигнала разной формы. Выборочный ансамбль состоял для всех экспериментов из 100 разных реализаций сигналов для каждого уровня шума.

Отдельные контрольные эксперименты с большим объемом выборки равной 225 реализациям сигнала показали, что среднее и дисперсия не меняются. Это было расценено как наличие статистической устойчивости.

При данном объеме выборки в силу центральной предельной теоремы среднее арифметическое $\langle \varepsilon_{\phi} \rangle$ и $\langle \varepsilon_A \rangle$ будет распределено по закону, близкому к нормальному. При проведении статистического эксперимента представляют интерес случаи с малыми значениями СКО, что означает устойчивость алгоритма оценивания фазы и огибающей к входным параметрам, вследствие этого средние арифметические ε_{ϕ} и ε_A как и оценки будут близки к оценкам максимального правдоподобия.

3.1.2. Результаты численного эксперимента

Результаты численного эксперимента разделены на несколько групп в зависимости от изменяемых параметров и вида исследования.

Устойчивость работы алгоритма оценивалась по количеству восстановленных «типовых» функций. При заданных четырех эталонных сигналах устойчивой работе алгоритма соответствует результат, при котором

получены ровно четыре восстановленных «типовых» функции при низком уровне ошибки (3.5). Число восстановленных функций большее или меньшее, чем число заданных эталонных сигналов, однозначно оценивается, как неустойчивая работ алгоритма независимо от уровня ошибки.

В результате работы алгоритма формируются «типовые» восстановленные сигналы. Соответствие между восстановленными сигналами и эталонными неизвестно, поэтому проводится сравнение каждого восстановленного сигнала с каждым эталонным, для каждого варианта рассчитывается ошибка. Наименьшее значение ошибки показывает принадлежность восстановленного сигнала к эталонному.

При отсутствии устойчивости алгоритма, таким способом, можно определить восстановленные «типовые» сигналы наиболее близкие к эталонным. На графиках зависимостей, отражающих результаты исследований, значения ошибок при неустойчивой работе алгоритма обозначены штриховой линией.

Результаты численного эксперимента восстановления фазовых функций при аддитивном шуме

1. Исследование влияния размера выборки на ошибку восстановления фазы.

Постоянные параметры эксперимента: шум аддитивный, ширина окна шума 25 частот в интервале от 5 до 30, количество эталонных сигналов 4, количество гармонических составляющих в эталонных сигналах 4, минимальное значение коэффициента корреляции 0,8.

Изменяемый параметр – размер выборки. В численном эксперименте на один эталонный сигнал задавалось 5, 10, 15, 25.сигналов с шумом. Таким образом, выборка состояла из 20, 40, 60, 100 сигналов с шумом.

Оценка нормированной среднеквадратической ошибки восстановления фазы рассчитывалась при изменении отношения шума к сигналу от 0 до 100 процентов.

Из полученных данных, представленных на рисунке 3.1, $a - z$, видно, что при всех размерах выборки с увеличением уровня шума растет среднее значение нормированной ошибки восстановления фаз.

При этом, чем больше сигналов в выборке, тем ниже среднее значение нормированной ошибки, вычисленное для области устойчивой работы алгоритма (на графиках – сплошная линия). (ср знач ош 35,99%, СКО 3,71%)

Устойчивая работа алгоритма сохраняется при размере выборки 20, 40, 60 сигналов до уровня шума 70%. При размере выборки равно 100 сигналам алгоритм устойчив до уровня шума 60%.

Вывод: при увеличении размера выборки ошибка вычисления снижается и падает устойчивость алгоритма к повышению уровня шума.

За пределами устойчивой работы алгоритма при шуме более 70% ошибка вычисления значительно увеличивается.

2. Исследование влияния формы эталонного сигнала на ошибку восстановления фазы.

Постоянные параметры эксперимента: шум аддитивный, ширина окна шума 25 частот в интервале от 5 до 30, количество эталонных сигналов 4 (их всегда 4), выборка состояла из 100 сигналов с шумом, минимальное значение коэффициента корреляции 0,8.

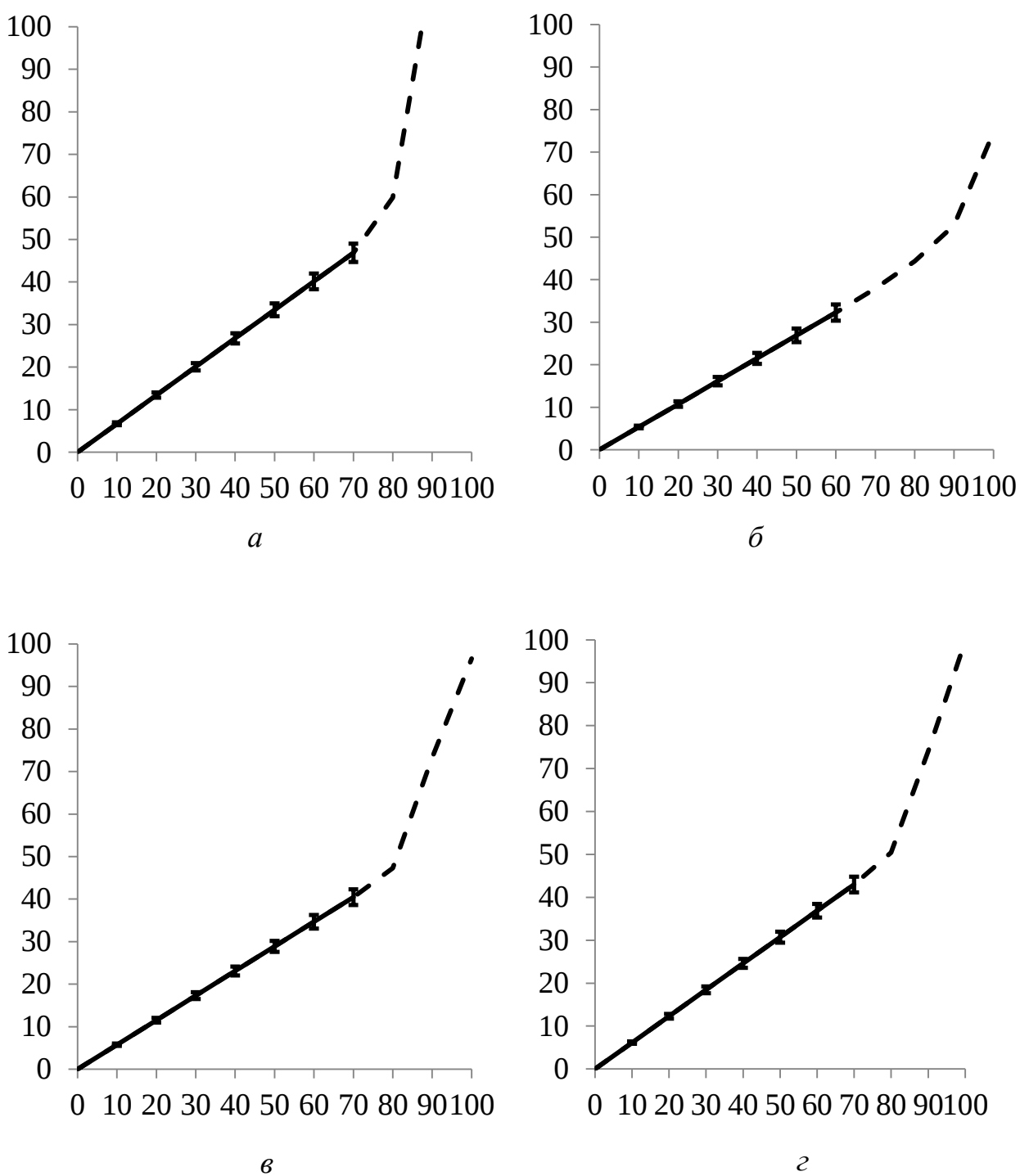


Рис. 3.1. Оценки нормированной среднеквадратической ошибки восстановления фаз алгоритмом динамической кластеризации при разном размере выборки: *а* – 20 , *б* – 40, *в* – 60 , *г* – 100 сигналов с шумом

Изменяемый параметр – количество гармонических составляющих в эталонных сигналах (так изменяется форма). В численном эксперименте для каждого эталонного сигнала задавалось количество гармоник равное 3, 6, 12, 24. Частоты гармонических составляющих выбирались из интервала от 5 до 35.

Оценка нормированной среднеквадратической ошибки восстановления фазы рассчитывалась при изменении отношения шума к сигналу от 0 до 100 процентов.

Из полученных данных, представленных на рисунке 3.2, $a - z$, видно, что при всех формах сигнала с увеличением уровня шума растет среднее значение нормированной ошибки.

Устойчивая работа алгоритма сохраняется до уровня шума 70%. при количествах гармоник в каждом эталонном сигнале 3, 12, 24, при количестве гармоник равном 6 алгоритм устойчив до уровня шума 60%.

Значение нормированной ошибки, вычисленное для области устойчивой работы алгоритма, при уровне шума 60%, изменяется незначительно от -1,35% до +1,35%. (-СКО, а среднее значение ошибки равно 32,74%).

Вывод: Независимо от формы сигнала точность вычисления сохраняется при величине шума от 0 до 60%. За пределами устойчивой работы алгоритма, при шуме более 70%, ошибка вычисления значительно увеличивается.

3. Исследование влияния формы шума на ошибку восстановления фазы.

Постоянные параметры эксперимента: шум аддитивный, количество эталонных сигналов 4, выборка состояла из 100 сигналов с шумом, количество гармонических составляющих в эталонных сигналах 4, минимальное значение коэффициента корреляции 0,8.

Изменяемый параметр – размер частотного (спектрального) окна для внесения шума (так изменяется форма шума). В численном эксперименте

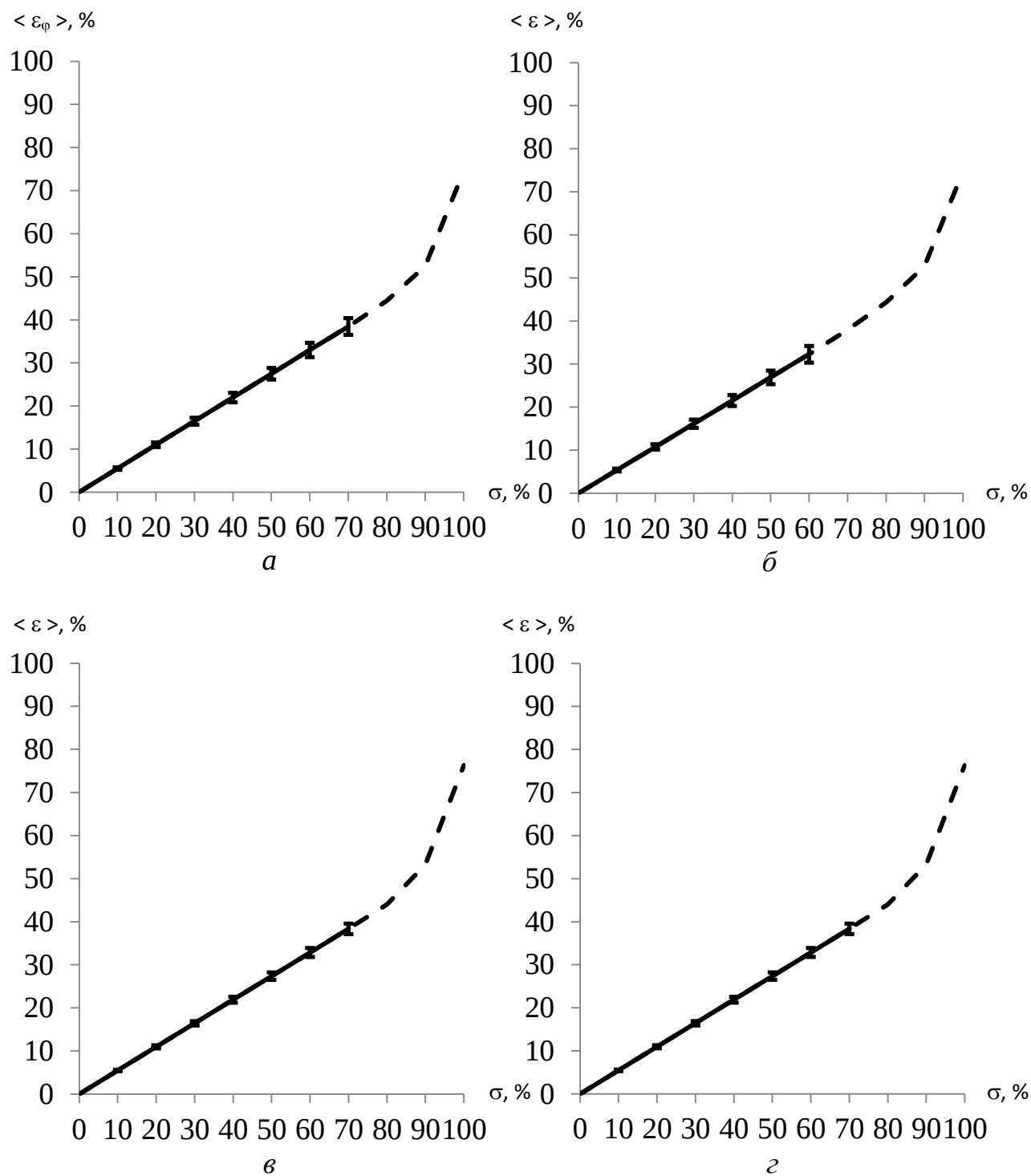


Рис. 3.2. Оценки нормированной среднеквадратической ошибки восстановления фаз алгоритмом динамической группировки для сигналов с разным количеством гармонических составляющих: а – 3 , б – 6, в – 12 , г – 24

Ширина спектра шума задавалась в размере 25, 50, 75, 100 частот в интервале от 0 до 100.

Оценка нормированной среднеквадратической ошибки восстановления фазы рассчитывалась при изменении отношения шума к сигналу от 0 до 100 процентов.

Из полученных данных, представленных на рисунке 3.3, *a – z*, видно, что при всех формах сигнала с увеличением уровня шума растет среднее значение нормированной ошибки.

При этом, чем больше ширина спектра шума, тем ниже среднее значение нормированной ошибки, вычисленное для области устойчивой работы алгоритма.

Устойчивая работа алгоритма сохраняется при ширине спектра шума 25, 50 частот до уровня шума 70%. При ширине спектра шума 75, 100 частот алгоритм устойчив до уровня шума 60%.

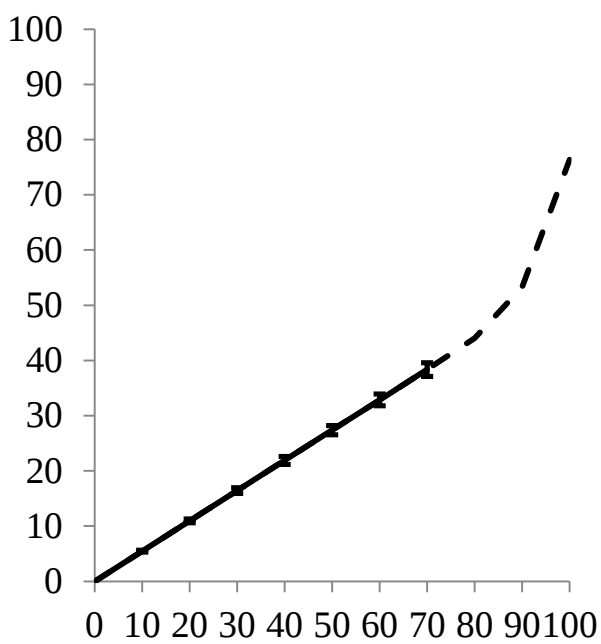
Вывод: при увеличении ширины спектра шума ошибка вычисления снижается. Устойчивость алгоритма снижается с увеличением ширины шума.

За пределами устойчивой работы алгоритма при шуме более 70% ошибка вычисления значительно увеличивается.

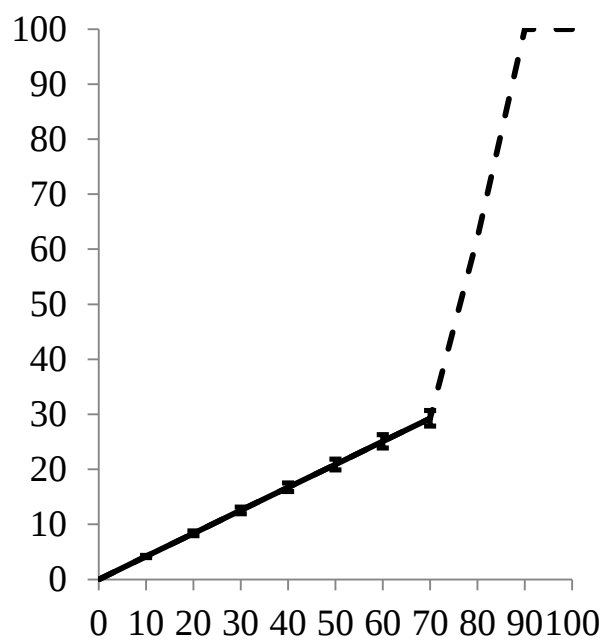
Исследование влияния уровня корреляции на ошибку восстановления фазы.

Постоянные параметры эксперимента: шум аддитивный, ширина спектрального окна шума 25 частот в интервале от 5 до 30, количество эталонных сигналов 4 (их всегда 4), выборка состояла из 100 сигналов с шумом, количество гармонических составляющих в эталонных сигналах 4,

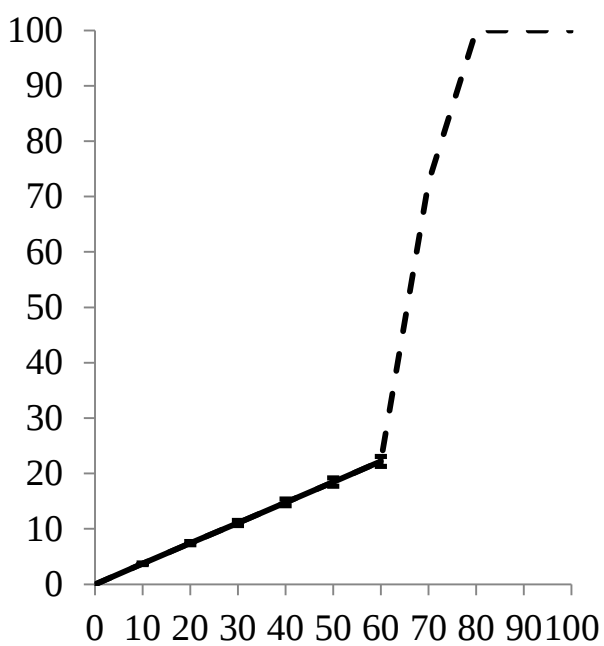
Изменяемый параметр – минимальное значение коэффициента корреляции. В численном эксперименте минимальное значение коэффициента корреляции задавалось равным: 0,3, 0,4, 0,5, 0,6, 0,7, 0,8, 0,9.



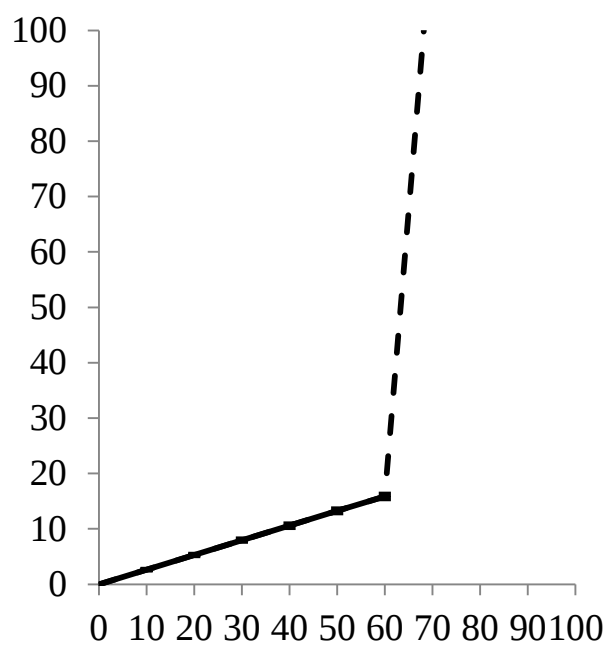
a



б



в



г

Рис. 3.3. Оценки нормированной среднеквадратической ошибки восстановления фаз алгоритмом динамической группировки при разной ширине спектра шума: а – 25 , б – 50, в – 75 , г – 100

Максимальный уровень корреляции между эталонными рядами в данном численном эксперименте составил 0,015.

Оценка нормированной среднеквадратической ошибки восстановления фазы рассчитывалась при изменении отношения шума к сигналу от 0 до 100 процентов.

Из полученных данных, представленных на рисунках 3.4 и 3.5 видно, что при всех формах сигнала с увеличением уровня шума растет среднее значение нормированной ошибки.

При этом, ошибка вычисления остается неизменной для каждого отдельного уровня шума при разных уровнях корреляции в зоне устойчивости алгоритма.

Устойчивая работа алгоритма сохраняется при минимальном уровне корреляции 0,3 и 0,7 до уровня шума 90%, при минимальном уровне корреляции 0,4 до уровня шума 80%, при минимальном уровне корреляции 0,8 до уровня шума 70%, при минимальном уровне корреляции 0,9 до уровня шума 40%.

Минимальный уровень корреляции равный 0,5 и 0,6 обеспечивает устойчивость алгоритма до уровня шума равного 100%.

Вывод: Алгоритм имеет одинаковые ошибки вычисления при разных уровнях шума независимо от уровня корреляции в зоне его устойчивой работы.

Существуют уровни корреляции при прочих равных условия (форма сигнала, и т.д.) при которых алгоритм имеет максимальную устойчивость при уровнях шума от 0 до 100 %.

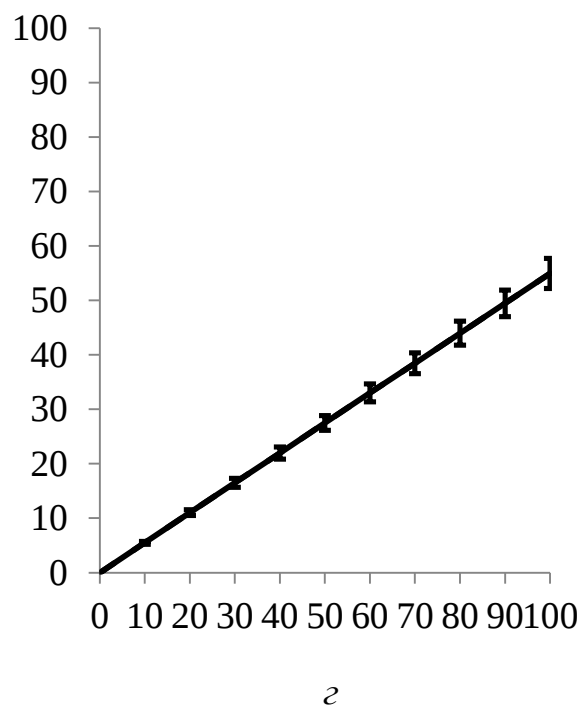
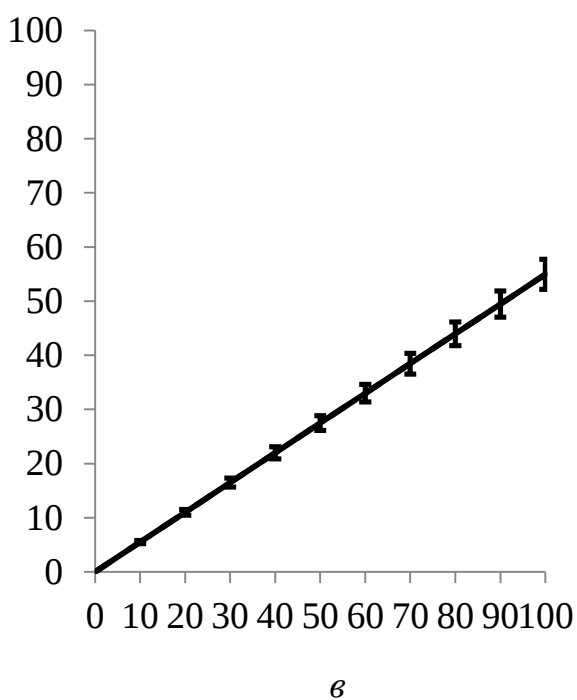
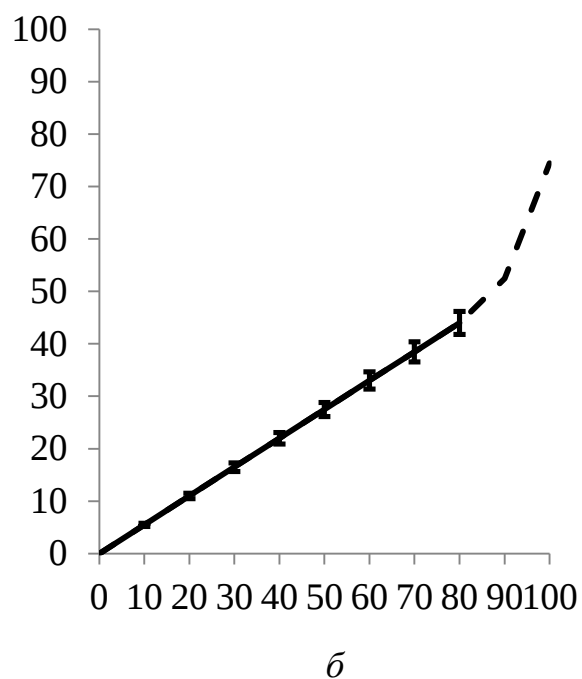
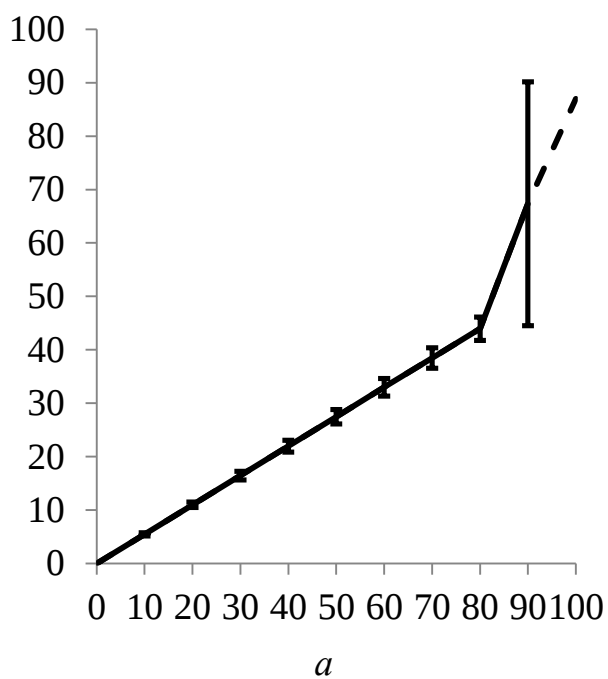


Рис. 3.4. Оценки нормированной среднеквадратической ошибки восстановления фаз алгоритмом динамической группировки при разном минимальном уровне корреляции: а – 0,3 , б – 0,4, в – 0,5 , г – 0,6

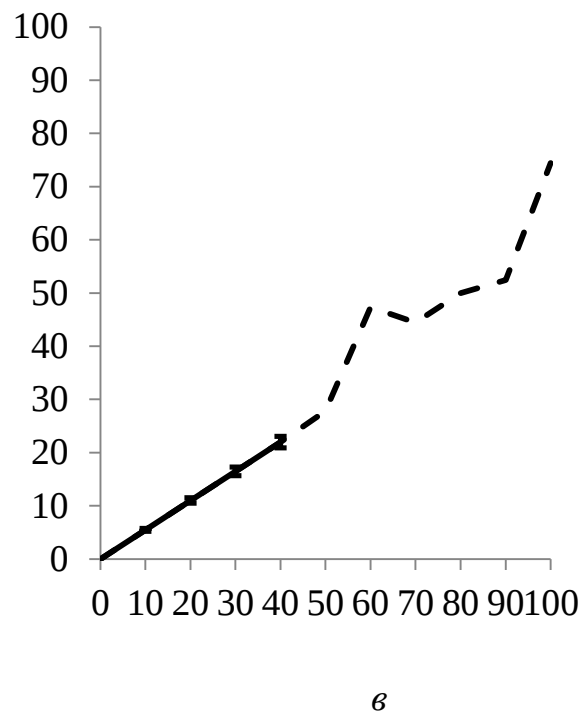
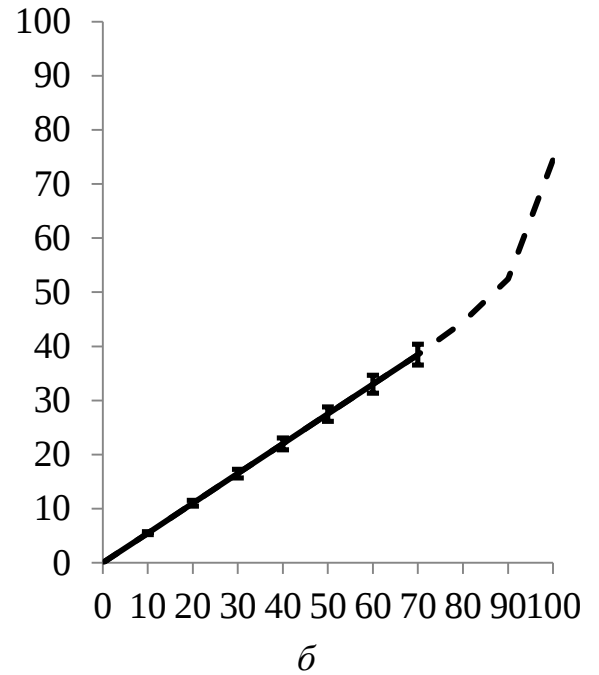
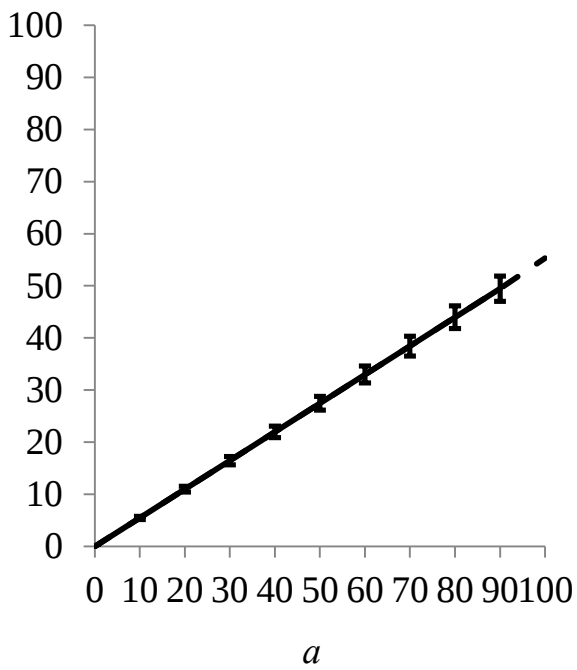


Рис. 3.5. Оценки нормированной среднеквадратической ошибки восстановления фаз алгоритмом динамической группировки при разном минимальном уровне корреляции: а – 0,7 , б – 0,8, в – 0,9

Существование уровней с высокой устойчивостью к шуму (до 100% включительно) говорит о наличии оптимальных уровней корреляции, для предлагаемой выборки сигналов.

За пределами зоны устойчивой работы алгоритма ошибки вычисления значительно увеличивается.

Результаты численного эксперимента восстановления огибающих функций при аддитивном шуме.

5. Исследование влияния уровня корреляции на ошибку восстановления огибающей.

Постоянные параметры эксперимента: шум аддитивный, ширина спектрального окна шума 25 частот в интервале от 5 до 30, количество эталонных сигналов 4 (их всегда 4), выборка состояла из 100 сигналов с шумом, количество гармонических составляющих в эталонных сигналах 4,

Изменяемый параметр – минимальное значение коэффициента корреляции. В численном эксперименте минимальное значение коэффициента корреляции задавалось равным: 0,4, 0,5, 0,6, 0,7, 0,8, 0,9.

Максимальный уровень корреляции между эталонными рядами в данном численном эксперименте составил 0,41.

Оценка нормированной среднеквадратической ошибки восстановления фазы рассчитывалась при изменении отношения шума к сигналу от 0 до 100 процентов.

Из полученных данных, представленных на рисунках 3.6 и 3.7 видно, что при всех формах сигнала с увеличением уровня шума растет среднее значение нормированной ошибки.

При этом ошибка вычисления остается неизменной для каждого отдельного уровня шума при разных уровнях корреляции в зоне устойчивости алгоритма.

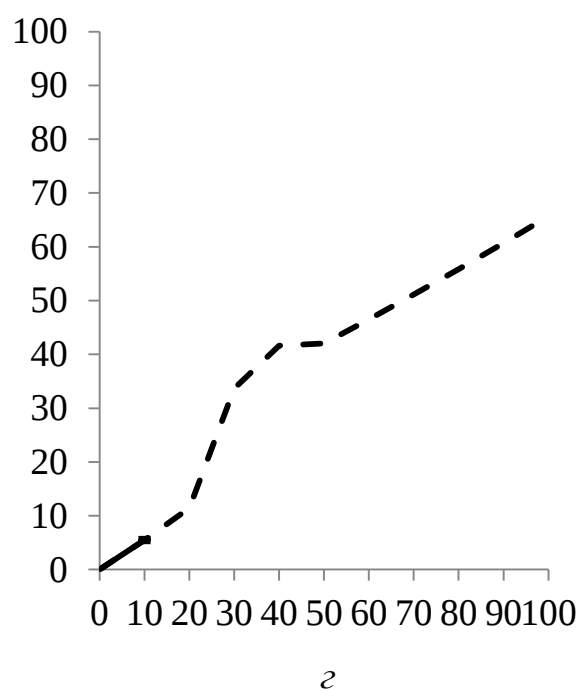
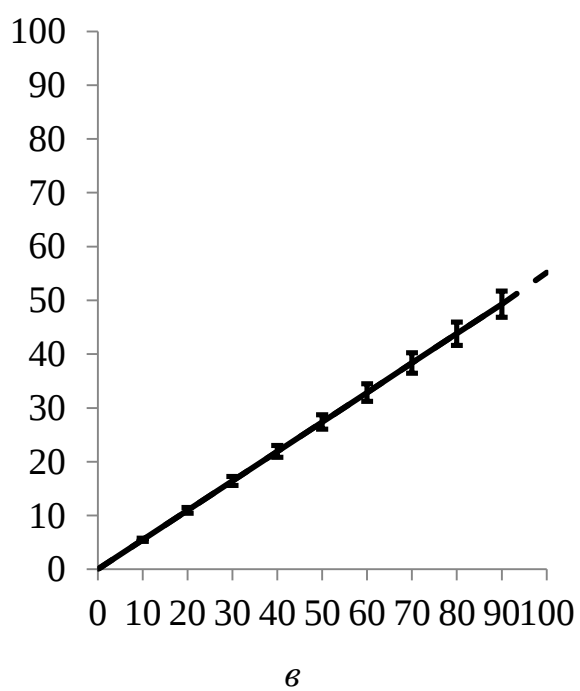
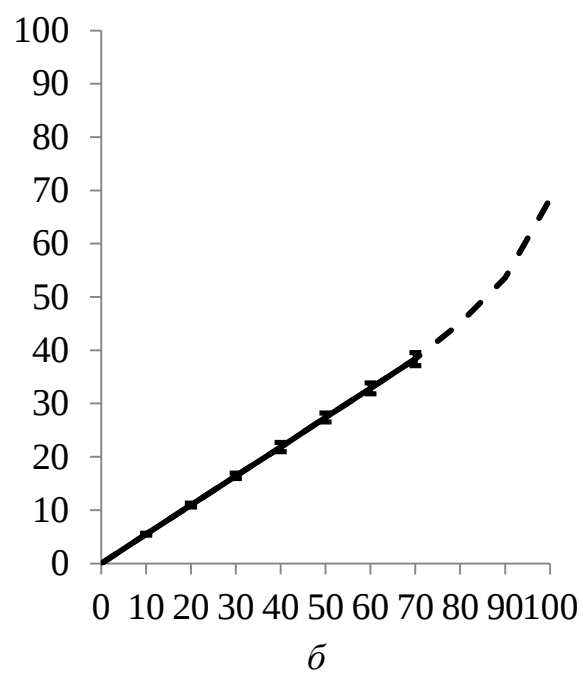
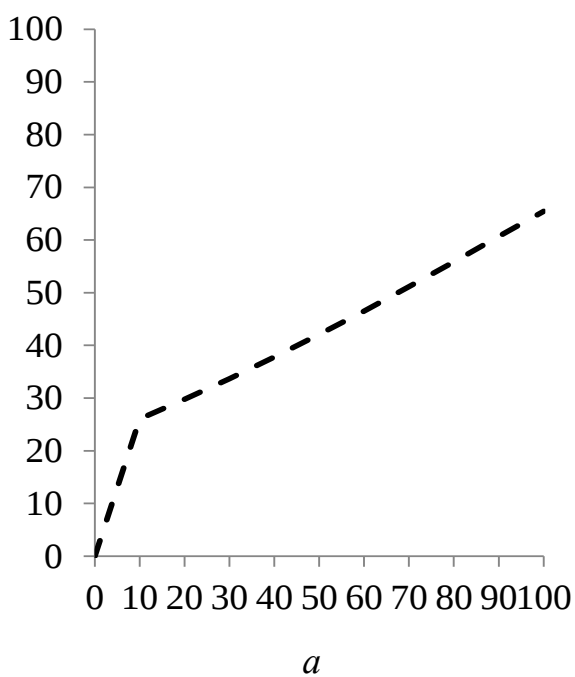


Рис. 3.6. Оценки нормированной среднеквадратической ошибки восстановления огибающей алгоритмом динамической группировки при разном минимальном уровне корреляции: а – 0,4 , б – 0,5, в – 0,6 , г – 0,7

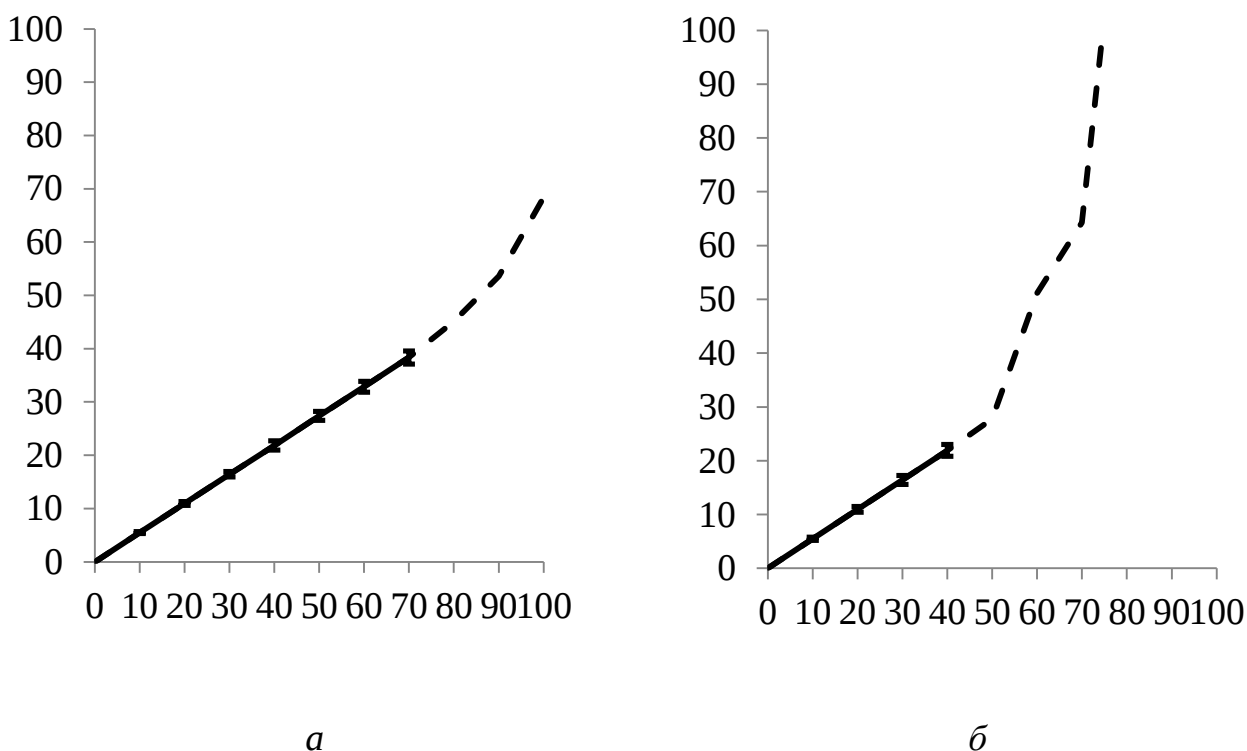


Рис. 3.7. Оценки нормированной среднеквадратической ошибки восстановления огибающей алгоритмом динамической группировки при разном минимальном уровне корреляции: а – 0,8 , б – 0,9

Устойчивая работа алгоритма сохраняется при минимальном уровне корреляции 0,5 до уровня шума 10%, при минимальном уровне корреляции 0,6 до уровня шума 60%, при минимальном уровне корреляции 0,7 до уровня шума 90%, при минимальном уровне корреляции 0,8 до уровня шума 70%, при минимальном уровне корреляции 0,9 до уровня шума 40%.

При минимальном уровне корреляции равном 0,4 алгоритм неустойчив в силу того, что максимальный уровень корреляции между эталонными рядами в данном численном эксперименте составил 0,41, что больше чем заданный минимальный уровень корреляции. В результате кластеризации две группы сливаются в одну. Высокий максимальный уровень корреляции между эталонными рядами в данном численном эксперименте задан умышленно для проверки устойчивости алгоритма. В сравнении с результатами исследования влияния уровня корреляции на ошибку восстановления фазы для огибающей при минимальных уровнях корреляции равных: 0,7, 0,8, 0,9 получены одинаковые

результаты, как по уровню устойчивости, так и по величине ошибки. Это является подтверждением того, что форма сигнала, а именно фаза или огибающая, не влияет на ошибку восстановления при реализации алгоритм.

При минимальных уровнях корреляции равных: 0,4, 0,5, 0,6, результаты отличаются, как по величине ошибке восстановления, так и по уровню устойчивости. Данные результаты показывают, что заданный высокий уровень корреляции между эталонными рядами значительно понижает устойчивость работы алгоритма.

Вывод: Алгоритм имеет одинаковые ошибки вычисления при разных уровнях шума независимо от уровня корреляции в зоне его устойчивой работы.

Существуют уровни корреляции, при которых алгоритм имеет высокую устойчивость (до уровня шума 90% при величине ошибки 49,28%). Установлено, что значительное влияние на ошибку кластеризации оказывает величина корреляции между эталонными сигналами.

6. Исследование влияния формы шума на ошибку восстановления огибающей.

Постоянные параметры эксперимента: шум аддитивный, количество эталонных сигналов 4 (их всегда 4), выборка состояла из 100 сигналов с шумом, количество гармонических составляющих в эталонных сигналах 4, минимальное значение коэффициента корреляции 0,8.

Изменяемый параметр – размер спектрального окна для внесения шума.

В численном эксперименте ширина спектра шума задавался в размере 25, 50, 75, 100 частот в интервале от 0 до 100.

Из полученных данных, представленных на рисунке 4.8, $a - z$, видно, что при всех формах сигнала с увеличением уровня шума растет среднее значение нормированной ошибки.

При этом, чем больше ширина спектра шума, тем ниже среднее значение нормированной ошибки, вычисленное для области устойчивой работы алгоритма.

Устойчивая работа алгоритма сохраняется при ширине спектра шума 25, 50 частот до уровня шума 70%. При ширине спектра шума 75, 100 частот алгоритм устойчив до уровня шума 60%. Результаты для огибающей повторяют результаты для полученные для фазы.

Вывод: при увеличении ширины спектра шума ошибка вычисления огибающей снижается. Устойчивость алгоритма снижается с увеличением ширины шума. За пределами устойчивой работы алгоритма при шуме более 70% ошибка вычисления значительно увеличивается.

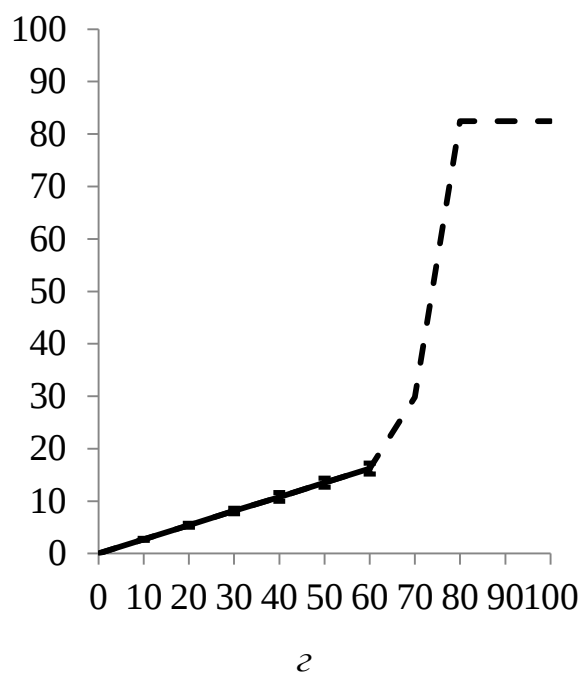
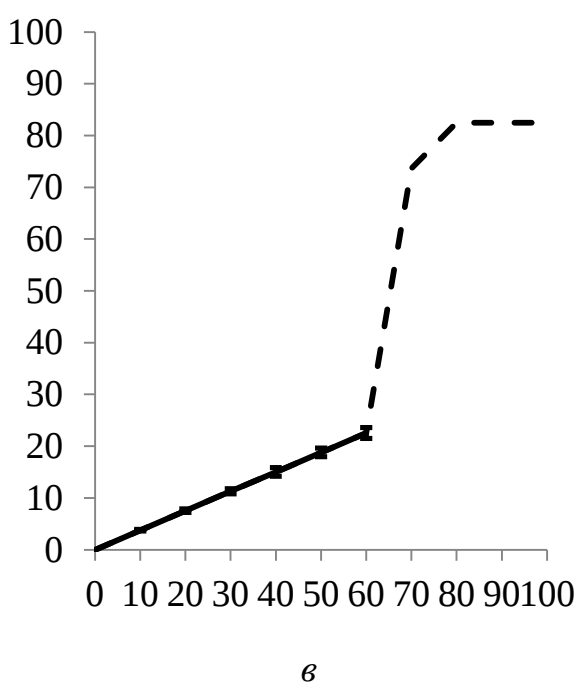
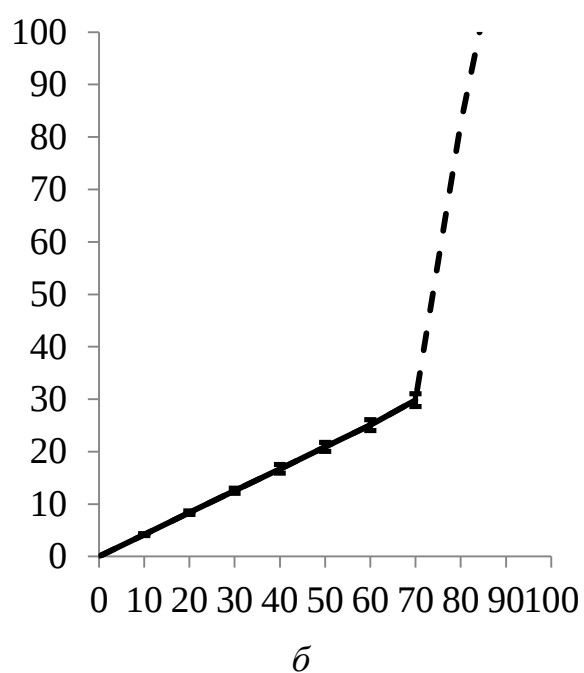
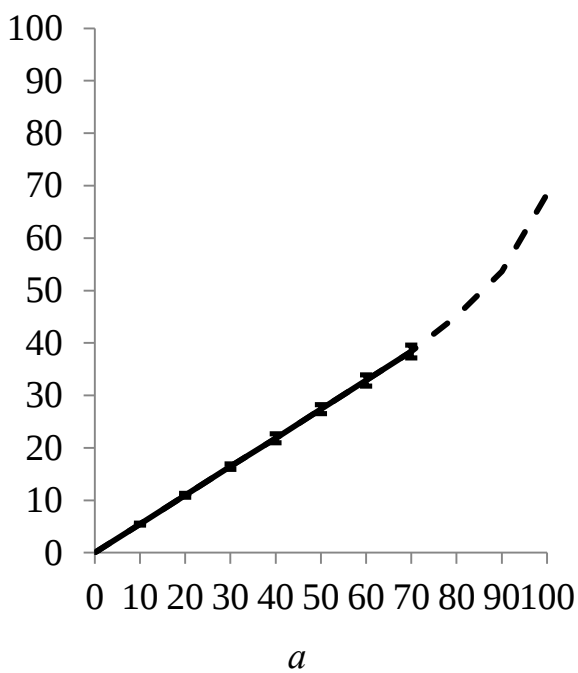


Рис. 3.8. Оценки нормированной среднеквадратической ошибки восстановления огибающей алгоритмом динамической группировки при разной ширине спектра шума: а – 25 , б – 50, в – 75 , з – 100

4. Социальная ответственность

Выполненная работа направлена на построение алгоритма, который будет решать задачи кластеризации и группировки большого объема данных, направленных на выявление групп, определенных по некоторому признаку.

Информация о пространственно-временной изменчивости границ климатических классов может быть востребована в сферах экономики и рационального природопользования.

Все действия, описанные в данной выпускной квалификационной работе, были совершены в городе Томск и на территории Томского Академгородка.

4.1 Правовые и организационные вопросы обеспечения безопасности

Сотруднику, постоянно работающему за компьютером, очень нелегко долгое время концентрировать взгляд на экране монитора. Устают глаза, шея, спина, затекают ноги. Для предотвращения негативных последствий воздействия на организм необходимо делать перерывы в работе. Законодательно установление перерывов закреплено в статье 109 Трудового Кодекса РФ. Правда, указано, что виды работ, предусматривающих предоставление работникам в течение рабочего времени специальных перерывов, а также их продолжительность устанавливаются правилами внутреннего трудового распорядка. Конкретно про перерывы при работе за компьютером ничего не сказано.

Однако, по типовой инструкции по охране труда при работе на персональном компьютере ТОО Р-45-084-01 всё же вменяется установление перерывов. В целом продолжительность непрерывной работы за компьютером не должна превышать 2-х часов.

Следует обратить внимание на то, что основная работа за компьютером предусматривает не менее 50 % времени в течение рабочей смены или рабочего дня нахождения за ним. Время перерыва зависит от вида и сложности осуществляемой работы путем деления на группы. Выделяют 3 группы: А (работа по считыванию информации с экрана компьютера с предварительным

запросом), Б (работа по вводу информации), В (творческая работа в режиме диалога с компьютером).

В зависимости от сложности работы, установление числа и длительности перерывов происходит следующим образом:

1) для группы А (не свыше 60000 считываемых знаков за смену) перерыв составляет 15 минут, предоставляется два раза – через два часа после начала работы и перерыва на обед;

2) для группы Б (не свыше 40000 вводимых знаков за смену) перерыв составляет 10 минут через каждый трудовой час;

3) для группы В (не свыше шести 6 часов за смену) перерыв составляет 15 минут через каждый трудовой час.

В целях обеспечения прав и свобод человека и гражданина работодатель и его представители при обработке персональных данных работника обязаны соблюдать следующие общие требования:

1) обработка персональных данных работника может осуществляться исключительно в целях обеспечения соблюдения законов и иных нормативных правовых актов, содействия работникам в трудоустройстве, получении образования и продвижении по службе, обеспечения личной безопасности работников, контроля количества и качества выполняемой работы и обеспечения сохранности имущества;

2) при определении объема и содержания обрабатываемых персональных данных работника работодатель должен руководствоваться Конституцией Российской Федерации, настоящим Кодексом и иными федеральными законами;

3) все персональные данные работника следует получать у него самого. Если персональные данные работника возможно получить только у третьей стороны, то работник должен быть уведомлен об этом заранее и от него должно быть получено письменное согласие. Работодатель должен сообщить работнику о целях, предполагаемых источниках и способах получения персональных данных, а также о характере подлежащих получению персональных данных и последствиях отказа работника дать письменное согласие на их получение;

4) работодатель не имеет права получать и обрабатывать сведения о работнике, относящиеся в соответствии с законодательством Российской Федерации в области персональных данных к специальным категориям персональных данных, за исключением случаев, предусмотренных настоящим Кодексом и другими федеральными законами;

5) работодатель не имеет права получать и обрабатывать персональные данные работника о его членстве в общественных объединениях или его профсоюзной деятельности, за исключением случаев, предусмотренных настоящим Кодексом или иными федеральными законами;

6) при принятии решений, затрагивающих интересы работника, работодатель не имеет права основываться на персональных данных работника, полученных исключительно в результате их автоматизированной обработки или электронного получения;

7) защита персональных данных работника от неправомерного их использования или утраты должна быть обеспечена работодателем за счет его средств в порядке, установленном настоящим Кодексом и иными федеральными законами;

8) работники и их представители должны быть ознакомлены под роспись с документами работодателя, устанавливающими порядок обработки персональных данных работников, а также об их правах и обязанностях в этой области;

9) работники не должны отказываться от своих прав на сохранение и защиту тайны;

10) работодатели, работники и их представители должны совместно вырабатывать меры защиты персональных данных работников.

Согласно ТК РФ Статья 135 об установлении заработной платы, заработная плата работнику устанавливается трудовым договором в соответствии с действующими у данного работодателя системами оплаты труда.

Системы оплаты труда, включая размеры тарифных ставок, окладов (должностных окладов), доплат и надбавок компенсационного характера, в том числе за работу в условиях, отклоняющихся от нормальных, системы доплат и надбавок стимулирующего характера и системы премирования, устанавливаются коллективными договорами, соглашениями, локальными нормативными актами в соответствии с трудовым законодательством и иными нормативными правовыми актами, содержащими нормы трудового права.

Локальные нормативные акты, устанавливающие системы оплаты труда, принимаются работодателем с учетом мнения представительного органа работников.

Условия оплаты труда, определенные трудовым договором, не могут быть ухудшены по сравнению с установленными трудовым законодательством и иными нормативными правовыми актами, содержащими нормы трудового права, коллективным договором, соглашениями, локальными нормативными актами. Условия оплаты труда, определенные коллективным договором, соглашениями, локальными нормативными актами, не могут быть ухудшены по сравнению с установленными трудовым законодательством и иными нормативными правовыми актами, содержащими нормы трудового права.

Что же касается нормы выработки, времени, нормативы численности и других норм – они устанавливаются в соответствии с достигнутым уровнем техники, технологии, организации производства и труда. Нормы труда могут быть пересмотрены по мере совершенствования или внедрения новой техники, технологии и проведения организационных либо иных мероприятий, обеспечивающих рост производительности труда, а также в случае использования физически и морально устаревшего оборудования. Достижение высокого уровня выработки продукции (оказания услуг) отдельными работниками за счет применения по их инициативе новых приемов труда и совершенствования рабочих мест не является основанием для пересмотра ранее установленных норм труда.

Конкретные размеры повышения оплаты труда устанавливаются работодателем с учетом мнения представительного органа работников в порядке, установленном статьей 372 настоящего Кодекса для принятия локальных нормативных актов, либо коллективным договором, трудовым договором.

Работодателю необходимо предоставить обязательное социальное страхование работников от несчастных случаев на производстве и профессиональных заболеваний. ИП и организации, являющиеся работодателями, обязаны с выплат сотрудникам, работающим по трудовым договорам, ежемесячно перечислять страховые взносы на пенсионное, медицинское и социальное страхование в ФНС РФ.

Под проектированием рабочего места понимается целесообразное пространственное размещение в горизонтальной и вертикальной плоскостях функционально взаимосвязанных средств производства (оборудования, оснастки, предметов труда и др.), необходимых для осуществления трудового процесса. При проектировании рабочих мест должны быть учтены освещенность, температура, влажность, давление, шум, наличие вредных веществ, электромагнитных полей и другие санитарно-гигиенические требования к организации рабочих мест.

4.2 Производственная безопасность

Опасным производственным фактором (ОПФ) называется такой производственный фактор, воздействие которого на работающего мгновенно приводит к травме или летальному исходу. Травма – это повреждение тканей организма и нарушение его функций внешним воздействием.

Вредным производственным фактором (ВПФ) называется такой производственный фактор, воздействие которого на работающего в определенных условиях приводит к заболеванию или снижению трудоспособности.

Таблица 1 - Возможные опасные и вредные факторы

Факторы (ГОСТ 12.0.003-2015)	Этапы работ		Нормативные документы
	Разработка	Эксплуатация	
1.Отклонение показателей микроклимата	+	+	СанПиН 2.2.4.548-96 Гигиенические требования к микроклимату производственных помещений
2. Превышение уровня шума		+	СН 2.2.4/2.1.8.562-96. Санитарные нормы. Шум на рабочих местах, в помещениях жилых, общественных зданий и на территории жилой застройки.
3.Отсутствие или недостаток естественного света	+	+	СанПиН 2.2.1/2.1.1.1278-03 Гигиенические требования к естественному, искусственному и совмещенному освещению жилых и общественных зданий.
4.Недостаточная освещенность рабочей зоны		+	СП 52.13330.2016 Естественное и искусственное освещение. Актуализированная редакция СНиП 23-05-95*[59].
5.Повышенное значение напряжения в электрической цепи, замыкание которой может произойти через тело человека	+	+	СанПиН 2.2.2/2.4.1340-03 "Гигиенические требования к персональным электронно-вычислительным машинам и организации работы" (с изменениями на 21 июня 2016 года)

4.3 Анализ опасных и вредных производственных факторов.

На отклонение показателей микроклимата влияют влажность, запыленность, температура воздуха. Если температура выше нормы, кровеносные сосуды расширяются и теплоотдача в окружающую среду возрастает, если ниже, то сосуды сужаются, приток крови к телу замедляется и теплоотдача уменьшается. Слишком высокая влажность (более 85 %) затрудняет терморегуляцию, а слишком низкая (менее 20 %) вызывает пересыхание слизистых, причем не только дыхательных путей, но и глаз. Чем выше влажность в помещении, тем слабее влияние электростатических и электромагнитных полей, уровень излучения которых в помещении, где установлен компьютер, всегда повышен. Чтобы узнать необходимые параметры микроклимата помещения, нужно знать категорию интенсивности энергозатрат организма. Исполнители данной работы относятся к Ia-категории с интенсивностью энергозатрат до 120 ккал/ч (до 139 Вт), так как представленный вид работ выполняется сидя и сопровождается незначительным физическим напряжением. Выполненная работа проводилась в теплый период года. Исходя из заданных параметров, температура воздуха должна составлять 22-24 °С, температура поверхностей: 21-25 °С, относительная влажность воздуха – 60-40%, а скорость движения воздуха – 0.1 м/с. Допустимые нормы: перепад температуры воздуха по высоте должен быть не более 3° С; перепад температуры воздуха по горизонтали, а также ее изменения в течение смены не должны превышать 4° С; при температуре воздуха на рабочих местах 25° С и выше максимально допустимые величины относительной влажности воздуха не должны выходить за пределы 70% - при температуре воздуха 25°С. Допустимые величины интенсивности теплового облучения работающих от источников излучения, нагретых до белого и красного свечения (раскаленный или расплавленный металл, стекло, пламя и др.) не должны превышать 140 Вт/кв.м. При этом облучению не должно подвергаться более 25% поверхности тела и обязательным является

использование средств индивидуальной защиты, в том числе средств защиты лица и глаз.

1) Источниками шума являются сами вычислительные машины, центральная система вентиляции и кондиционирования воздуха и другое оборудование. Работающие в условиях длительного шумового воздействия люди испытывают раздражительность, головокружение, снижение памяти, повышенную утомляемость. Под воздействием шума снижается концентрация внимания, нарушаются физиологические функции, ухудшается речевая коммутация. Все это снижает работоспособность человека и его производительность, качество и безопасность труда. При выполнении основной работы уровень шума на рабочем месте не должен превышать 50 дБА. Дополнительными мероприятиями по шумогашению в машинных залах могут быть устройство подвесного потолка, который служит звукопоглощающим экраном, использование звукопоглощающих материалов с максимальными коэффициентами звукопоглощения в области частот 63-8000 Гц для отделки помещений, уменьшение площади стеклянных ограждений и оконных проемов, так же можно использовать звукогасящие экраны на рабочем месте и однотонные занавески из плотной ткани, подвешенные в складку на расстоянии 15-20 см от ограждения. Ширина занавеси должна быть в 2 раза больше ширины окна.

2) Отсутствие оконных проемов влечет за собой отсутствие или недостаток естественного освещения. Отсутствие естественного света на рабочем месте влечет за собой снижение общего тонуса организма и ослаблению иммунитета. Более того, отсутствие солнечных лучей негативно влияет на нервную систему и обменные процессы. Для оценки использования естественного света введено понятие коэффициента естественной освещенности (КЕО) и установлены минимальные допустимые значения КЕО — это отношение освещенности E_v внутри помещения за счет естественного света к наружной освещенности E_n от всей полусферы небосклона, выраженное в процентах: $КЕО = (E_v / E_n) 100\%$.

По характеристике зрительской работы труд учащихся можно отнести ко второму разряду работы, и при боковом естественном освещении в аудитории, лаборатории на рабочих столах и партах должен обеспечиваться $KEO = 1,5 \%$. К средствам нормализации освещения производственных помещений и рабочих мест относятся: источники света, осветительные приборы, световые проемы, светозащитные устройства, светофильтры.

3) Недостаточное освещение влияет на функционирование зрительного аппарата, то есть определяет зрительную работоспособность, на психику человека, его эмоциональное состояние, вызывает усталость центральной нервной системы, возникающей в результате прилагаемых усилий для опознания четких или сомнительных сигналов. Известно, что при длительной работе в условиях недостаточной освещенности и при нарушении других параметров световой среды зрительное восприятие снижается, развивается близорукость, болезнь глаз, появляются головные боли. Норма освещенности, согласно нормативным документам, должна быть не менее 200-300 Лк. Средства нормализации освещения производственных помещений и рабочих мест: источники света; осветительные приборы; световые проемы; светозащитные устройства; светофильтры.

4) Перегрузки в работе систем, а так же несоблюдение правил по электробезопасности могут вызвать повышенное значение напряжения в электрической цепи. Это может повлечь за собой местные поражения организма человека электрическим током (ожоги, механические повреждения и т.п.) или электрический удар. Чтобы избежать короткого замыкания, а значит, возникновения пожара и получения электротравмы, помещения, где размещаются рабочие места с компьютерами, должны быть оборудованы защитным заземлением (занулением) в соответствии с техническими требованиями по эксплуатации.

4.4 Экологическая безопасность

Охрана окружающей среды – это комплексная проблема и наиболее активная форма её решения – это сокращение вредных выбросов промышленных предприятий через полный переход к безотходным или малоотходным технологиям производства.

С точки зрения потребления ресурсов компьютер потребляет сравнительно небольшое количество электроэнергии, что положительным образом сказывается на общей экономии потребления электроэнергии в целом.

Основными отходами при выполнении данной бакалаврской работы являются черновики бумаги, отработавшие люминесцентные лампы и картриджи. Израсходованная бумага была направлена на утилизацию в пункт приема макулатуры, люминесцентные лампы – в упакованном виде на ртутьперерабатывающий завод, расположенный в г. Северск. Составляющие израсходованных картриджей (стальные винты, алюминиевые каркасы, пластик, углеродистый порошок) по отдельности в упакованном виде также были направлены на утилизацию в соответствии с требованиями ГОСТ 30775-2001.

4.5 Обоснование мероприятий по снижению уровней воздействия опасных и вредных факторов на исследователя.

Организационные мероприятия при компоновке рабочей зоны Большое значение для профилактики статических физических перегрузок имеет правильная организация рабочего места человека, работающего с ПЭВМ. Рабочее место должно быть организовано в соответствии с требованиями стандартов, технических условий и (или) методических указаний по безопасности труда. Оно должно удовлетворять следующим требованиям:

- 1) обеспечивать возможность удобного выполнения работ;
- 2) учитывать физическую тяжесть работ;
- 3) учитывать размеры рабочей зоны и необходимость передвижения в ней работающего;
- 4) учитывать технологические особенности процесса выполнения работ.

Невыполнение требований к расположению и компоновке рабочего места может привести к получению работником производственной травмы или развития у него профессионального заболевания. Рабочее место программиста должно соответствовать требованиям СанПин 2.2.2/2.4.1340-03. Конструкция оборудования и рабочего места при выполнении работ в положении сидя должна обеспечивать оптимальное положение работающего, которое достигается регулированием высоты рабочей поверхности, высоты сидения, оборудованием пространства для размещения ног и высотой подставки для ног. Схемы размещения рабочих мест с персональными компьютерами должны учитывать расстояния между рабочими столами с мониторами: расстояние между боковыми поверхностями мониторов не менее 1,2 м, а расстояние между экраном монитора и тыльной частью другого монитора не менее 2,0 м. Клавиатура должна располагаться на поверхности стола на расстоянии 100-300 мм от края, обращенного к пользователю. Быстрое и точное считывание информации обеспечивается при расположении плоскости экрана ниже уровня глаз пользователя, предпочтительно перпендикулярно к нормальной линии взгляда (нормальная линия взгляда 15 градусов вниз от горизонтали). Рабочие места с ПЭВМ при выполнении творческой работы, требующей значительного умственного напряжения или высокой концентрации внимания, рекомендуется изолировать друг от друга перегородками высотой 1,5 - 2,0 м.

Разрабатываемый в ходе выполнения ВКР программный продукт используется для анализа численных данных и наборов значений. При работе с данным продуктом не требуется постоянное его использование, так как программа анализирует входные данные и получает необходимый результат, который, непосредственно, будет использован в дальнейшем. Работнику необходимо контролировать параметры входных данных и фиксировать выходные параметры. Следовательно, преимуществом данного продукта является практически полная автономность.

4.6 Безопасность в чрезвычайных ситуациях

В Томске преобладает континентально-циклонический климат. Природные явления (землетрясения, наводнения, засухи, ураганы и т. д.) отсутствуют. Возможными ЧС могут являться сильные морозы.

Для Сибири в зимнее время года характерна очень низкая температура воздуха. Достижение критически низких температур приведет к авариям систем теплоснабжения и жизнеобеспечения, приостановке работы, обморожениям и даже жертвам среди населения. В случае переморозки труб должны быть предусмотрены запасные обогреватели. Их количества и мощности должно хватать для того, чтобы работа на производстве не прекратилась. Кроме того, необходимо иметь альтернативные источники тепла, электроэнергии и транспорта.

Чрезвычайные ситуации, возникающие в результате диверсий, возникают все чаще. Зачастую такие угрозы оказываются ложными. Но случаются взрывы и в действительности.

4.7 Выводы и рекомендации

Проанализировав условия труда на рабочем месте, где была разработана бакалаврская работа, можно сделать вывод, что помещение удовлетворяет необходимым нормам и в случае соблюдения техники безопасности и правил пользования компьютером работа в данном помещении не приведет к ухудшению здоровья работника. Само помещение и рабочее место в нем удовлетворяет всем нормативным требованиям. Кроме того, действие вредных и опасных факторов сведено к минимуму, т.е. микроклимат, освещение и электробезопасность соответствуют требованиям, предъявленным в соответствующих нормативных документах.

Относительно рассмотренного вопроса об экологической безопасности можно сказать, что деятельность в ходе выполнения выпускной квалификационной работы не представляет опасности окружающей среде.

5. Финансовый менеджмент, ресурсоэффективность и ресурсосбережение

Целью написания раздела «Финансовый менеджмент, ресурсоэффективность и ресурсосбережение» является проектирование и создание конкурентоспособных разработок, отвечающих современным требованиям в области ресурсоэффективности и ресурсосбережения.

5.1 Оценка коммерческого и инновационного потенциала научных исследований

5.1.1 Потенциальные потребители результатов исследований

Выполненная работа направлена на разработку и реализацию алгоритма, который будет решать задачи кластеризации и группировки большого объема данных, направленных на выявление групп, по некоторому признаку.

Информация о пространственно-временной изменчивости границ климатических классов востребована в сферах экономики и рационального природопользования.

В таблице приведены основные сегменты рынка по следующим критериям: размер компании-заказчика и направление деятельности. Анализ рынка проводился на основе следующих компаний: DPD (крупная компания), Томские Транспортные Линии (средняя компания), ПЭК Томск (мелкая компания).

Таблица 1 – Карта сегментирования рынка

		Направление деятельности	
		Потребности	Запасы
Размер компании	Крупная	+	+
	Средняя	+	-
	Мелкая	+	-

Согласно карте сегментирования рынка, выберем следующие сегменты: построение кластеров изображений для крупных компаний.

5.1.2 Анализ конкурентных технических решений

Анализ конкурентных технических решений позволяет провести оценку сравнительной эффективности научной разработки и определить направления для ее будущего повышения. Основным методом используемым в данной работе является метод нечеткой кластеризации. Оценочная карта для сравнения конкурентных технических решений представлена в таблице 2.

Таблица 2 – Оценочная карта для сравнения конкурентных технических решений

Критерии оценки	Вес	Баллы			Конкуренто- способность		
		Б _{осн}	Б _{см}	Б _{пп}	Б _{осн}	Б _{см}	Б _{пп}
Технические критерии оценки ресурсоэффективности							
Эффективность	0,3	4	4	4	1,2	1,2	1,2
Устойчивость	0,2	5	5	5	1	1	1
Временные затраты	0,2	5	4	4	1	0,8	0,8
Новизна метода	0,1	5	3	3	0,5	0,3	0,3
Простота реализации	0,1	5	4	5	0,5	0,4	0,5
Универсальность	0,1	4	4	4	0,5	0,4	0,4
Итого	1	29	24	25	4,7	4,1	4,2

По полученным результатам можно сделать вывод, что разрабатываемый алгоритм для оценки информативности является по конкурентоспособности наиболее эффективным.

5.1.3 SWOT – анализ

SWOT-анализ – это эффективный инструмент стратегического менеджмента. SWOT-анализ заключается в исследовании внешней и внутренней сред проекта.

Факторы разделяются на четыре категории:

- Strengths (сильные стороны),
- Weaknesses (слабые стороны),
- Opportunities (возможности),
- Threats (угрозы).

Разработанная для алгоритма поиска информативного атрибута матрица SWOT представлена в таблице 3.

Таблица 3 – SWOT- анализ

	Сильные стороны: С1. Алгоритм самостоятельно определяет количество кластеров. С2. Разработанный алгоритм обладает достаточно высокой точностью.	Слабые стороны: Сл1. При пограничных условиях возможна долгая сходимость.
Возможности: В1. Расширение функционала. В2. Автоматизация алгоритма.	При некоторых модификациях алгоритм может справляться с решением многих задач. Например, с задачей выделения климатических групп или задачей распознавания образов.	Автоматизация алгоритма позволит значительно ускорить его работу.
Угрозы: У1. Схожесть с алгоритмом k-means может повлечь за собой низкий спрос алгоритма на рынке. У2. В некоторых ситуациях возможна слабая устойчивость алгоритма.	При всей схожести алгоритма с k-means он обладает существенным отличием и преимуществом, а именно отсутствием необходимости задавать количество кластеров заранее, что делает работу с ним простой и удобной.	Неустойчивость алгоритма может спровоцировать низкий спрос на продукт.

5.2 Определение возможных альтернатив проведения научных исследований

Для выявления возможных альтернатив разработки проекта и доработки результатов был использован морфологический подход. Он основан на подборе возможных решений для отдельных частей задачи и последующем систематизированном получении их сочетаний.

В таблице 4 в виде матрицы представлены возможные варианты реализации разработки.

Таблица 4 – Морфологическая матрица

	1	2	3
А. Количество языков программирования, для которых проводится анализ	1	2	Больше 2
Б. Платформа для анализа кода	Mathcad	C++	–

Путём комбинации различных параметров были определены два наиболее оптимальных варианта исполнения:

- А1Б1;
- А1Б2.

Для данной матрицы наиболее оптимальным вариантом исполнения является А1Б1.

5.3 Планирование научно-исследовательских работ

5.3.1 Структура работ в рамках научного исследования

Трудоемкость выпускной квалификационной работы определяется опытным путем в человеко-днях и имеет вероятностный характер, так как зависит от трудно учитываемых факторов. Для реализации проекта необходимо два человека: студент-дипломник (СД) и научный руководитель (НР). Основные этапы выполнения ВКР приведены в таблице 5.

Таблица 5 – Основные этапы выполнения ВКР

№	Описание этапа	Исполнитель и	Загруженность исполнителей
1	Составление и утверждение задания	НР	НР – 100%
2	Анализ предметной области	С, НР	С – 50%, НР – 50%
3	Разработка календарного плана	С, НР	С – 10%, НР – 90%
4	Обзор литературы и интернет - источников	С, НР	С – 50%, НР – 50%
5	Обзор задач и методов решения задач ЛП	С	С – 100%
6	Математическая постановка задачи	С, НР	С – 10%, НР – 90%
7	Выбор метода решения поставленной задачи	С, НР	С – 10%, НР – 90%
8	Разработка алгоритмов решения задачи	С	С – 100%
9	Программная реализация компьютерной модели	С	С – 100%
10	Проведение исследований на модельных данных	С, НР	С – 80% НР – 20%
11	Проведение исследований на реальных данных	С, НР	С – 60% НР – 40%
12	Расчет экономической эффективности научно-технической продукции	С	С – 100%
13	Оценка социальной ответственности проекта	С	С – 100%
14	Написание пояснительной записки	С, НР	С – 90% НР – 10%

5.3.2 Определение трудоемкости выполнения работ

Трудовые затраты в большинстве случаев образуют основную часть стоимости разработки, поэтому важным моментом является определение трудоемкости работ каждого из участников научного исследования.

Ожидаемая продолжительность этапа рассчитывается по формуле:

$$t_{ож} = \frac{3 \cdot t_{\min} + 2 \cdot t_{\max}}{5}, \quad (1)$$

где $t_{\min}$ – предположительно минимальная продолжительность этапа в рабочих днях, определяемая методом экспертной оценки;

t_{max} – предположительно максимальная продолжительность этапа в рабочих днях, определяемая методом экспертной оценки.

Продолжительность каждого этапа рассчитывается по формуле:

$$t_{paб} = t_{ож} \cdot k_{\partial}, \quad (2)$$

где $t_{paб}$ – длительность этапов в рабочих днях;

k_{∂} – коэффициент, учитывающий дополнительное время на консультации и согласование работ, $k_{\partial} = 1.2$.

5.3.3 Разработка графика проведения научного исследования

В данном пункте необходимо построить диаграмму Ганта – горизонтально-ленточный график, на котором работы представляются протяженными во времени отрезками, характеризующимися датами начала и окончания выполнения приведенных работ.

Линейный график строится на основании полученных значений $t_{paб}$, предварительно переведенных в календарные дни по формуле:

$$t_{\kappa} = t_{paб} \cdot K_{\kappa}, \quad (3)$$

где t_{κ} – длительность этапов работ в календарных днях;

K_{κ} – коэффициент календарности.

Коэффициент календарности рассчитывается по формуле:

$$K_{\kappa} = \frac{T_{\kappa}}{T_{\kappa} - T_{\partial\partial} - T_{\partial\partial}}, \quad (4)$$

где T_{κ} – количество календарных дней, $T_{\kappa} = 365$,

$T_{\partial\partial}$ – количество выходных дней, $T_{\partial\partial} = 108$,

$T_{\partial\partial}$ – количество праздничных дней, $T_{\partial\partial} = 10$,

$$K_{\kappa} = \frac{365}{365 - 108 - 10} = 1.478. \quad (5)$$

Все расчеты сведены в таблицу 6.

Таблица 6 – Временные показатели проведенных работ

№	Исполнители	Продолжительность работ в днях			Трудоемкость	
		t_{min}	t_{max}	$t_{ож}$	$t_{раб}$	t_k
1	НР	1	2	1,4	1,68	2,48
2	С, НР	2	3	2,4	2,88	4,26
3	С, НР	1	2	1,4	1,68	2,48
4	С, НР	3	5	3,8	4,56	6,74
5	С	2	5	3,2	3,84	5,68
6	С, НР	3	6	4,2	5,04	7,45
7	С, НР	5	8	6,2	7,44	11,00
8	С	3	6	4,2	5,04	7,45
9	С	4	6	4,8	5,76	8,51
10	С, НР	2	5	3,2	3,84	5,68
11	С, НР	2	4	2,8	3,36	4,97
12	С	5	7	5,8	6,96	10,29
13	С	4	5	4,4	5,28	7,80
14	С, НР	9	12	10,2	12,24	18,09
Итого				58	69,6	102,87

На основании таблицы 6 построим диаграмму Ганта (Рис. 1).

№ раб от	Вид работ	Исполнители	T _{к_и} , кал. дн.	Продолжительность выполнения работ													
				февр		март			апрель			май			июнь		
				2	3	1	2	3	1	2	3	1	2	3	1	2	
1	Выбор направления исследований	Руководитель, инженер	3, 3														
2	Составление и утверждение научного задания	Руководитель	5														
3	Подбор и изучение материалов по теме	Руководитель, инженер	11, 11														
4	Календарное планирование работ по теме	Инженер	5														
5	Проведение теоретических расчетов и обоснований	Инженер	58														
6	Анализ полученных результатов	Руководитель, инженер	8, 8														

■ – инженер

■ – руководитель

Рис. 1. Диаграмма Ганта

5.4 Бюджет научно-технического исследования (НТИ)

5.4.1 Расчет затрат на специальное оборудование для научных (экспериментальных) работ

В данную статью включают все затраты, связанные с приобретением специального оборудования, необходимого для проведения работ по конкретной теме.

В ходе выполнения НТИ имеющееся компьютерное оборудование (ПК) было улучшено дополнительной оперативной памятью с целью ускорения работы т.к. в процессе выполнения алгоритм проходит множество итераций, в ходе которых происходит пересчет большого количества данных, представленных в виде огромных матриц.

Таблица 7 – Расчет бюджета затрат на приобретение спецоборудования для
научных работ

Наименование	Единица измерения	Количество	Цена за единицу, руб.	Затраты на материалы, руб.
Оперативная память DDR3-L	шт.	1	4800	4800
Итого				4800

5.4.2 Основная заработная плата исполнителей темы

Основная заработная плата руководителя рассчитывается на основании отраслевой оплаты труда (оклад, стимулирующие выплаты, районный коэффициент). Для студента-дипломника основную заработную плату составляет государственная стипендия с учетом районного коэффициента.

Для расчета основной заработной платы (табл. 9) необходимо привести действительный годовой фонд рабочего времени руководителя и студента (табл. 8).

Таблица 8 – Баланс времени

Показатели рабочего времени	Руководитель	Студент-дипломник
Календарное число дней	365	365
Количество нерабочих дней (выходные дни и праздничные дни)	52 и 10	108 и 10
Потери рабочего времени (отпуск, больничные), дни	48	24
Действительный годовой фонд рабочего времени, дни	255	223

Таблица 9 – Расчет основной заработной платы

Исполнитель	Тарифная заработная плата, руб.	Районный коэффициент	Месячный должностной оклад работника, руб.	Среднедневная заработная плата, руб.	Продолжительность работ, дни	Заработная плата основная, руб.
Руководитель	18745	1,3	24368,5	981,4	45	44163

Итого:	44163
---------------	-------

5.4.3 Дополнительная заработная плата

Затраты по дополнительной заработной плате исполнителей темы учитывают величину предусмотренных Трудовым кодексом РФ доплат за отклонение от нормальных условий труда, а также выплат, связанных с обеспечением гарантий и компенсаций (при исполнении государственных и общественных обязанностей, при совмещении работы с обучением, при предоставлении ежегодного оплачиваемого отпуска и т.д.).

Расчет дополнительной заработной платы ведется по следующей формуле:

$$Z_{доп} = k_{доп} * Z_{осн} = 0.15 * 44163 = 6624,5, \quad (6)$$

где $Z_{осн}$ – основная заработная плата, $k_{доп}$ – коэффициент дополнительной заработной платы на стадии проектирования, принимается равным 0,15.

5.4.4 Отчисления во внебюджетные фонды (страховые отчисления)

В данной статье расходов отражаются обязательные отчисления по установленным законодательством Российской Федерации нормам органами государственного социального страхования (ФСС), пенсионного фонда (ПФ) и медицинского страхования (ФФОМС) от затрат на оплату труда работников.

Величина отчислений во внебюджетные фонды определяется исходя из следующей формулы:

$$Z_{внеб} = k_{внеб} \cdot (Z_{осн} + Z_{доп}), \quad (7)$$

где $k_{внеб}$ – коэффициент отчислений на уплату во внебюджетные фонды (пенсионный фонд, фонд обязательного медицинского страхования и пр.).

Для учреждений, осуществляющих образовательную и научную деятельность в 2019г., водится пониженная ставка 28% (п. 6 ч. 1 ст. 58 Закона 212-ФЗ).

Расчет отчислений во внебюджетные фонды приведен в таблице 10.

Таблица 10 – Отчисления во внебюджетные фонды

Исполнитель	Основная заработная плата, руб.	Дополнительная заработная плата, руб.
Руководитель проекта	44163	6624,5
Коэффициент отчисления во внебюджетные фонды	12365,64	1854,8
Итого		14220,5

5.4.5 Накладные расходы

Накладные расходы учитывают прочие затраты организации, не попавшие в предыдущие статьи расходов: печать и ксерокопирование материалов исследования, оплата услуг связи, электроэнергии и т.д.

Их величина определяется по следующей формуле:

$$Z_{\text{накл}} = (\text{сумма статей } 1 \div 4) \cdot k_{\text{ир}}, \quad (8)$$

где $k_{\text{ир}}$ – коэффициент, учитывающий накладные расходы, примем равным 16%.

5.4.6 Формирование бюджета затрат научно-исследовательского проекта

Рассчитанная величина затрат научно-исследовательской работы является основой для формирования бюджета затрат проекта, который при формировании договора с заказчиком защищается научной организацией в качестве нижнего предела затрат на разработку научно-технической продукции.

Определение бюджета затрат на научно-исследовательский проект приведено в таблице 11.

Таблица 11 – Расчет бюджета затрат

Наименование статьи	Сумма, руб.
1. Затраты на специальное оборудование	4 800
2. Затраты по основной заработной плате исполнителей проекта	44 163
3. Затраты по дополнительной заработной плате исполнителей проекта	6624
4. Отчисления во внебюджетные фонды	14 220
5. Накладные расходы	9650
Бюджет затрат НТИ	79 457

5.5 Определение ресурсной (ресурсосберегающей), финансовой, бюджетной, социальной и экономической эффективности исследования

Определение эффективности происходит на основе расчета интегрального показателя эффективности научного исследования. Его нахождение связано с определением двух средневзвешенных величин: финансовой эффективности и ресурсоэффективности.

Интегральный финансовый показатель разработки определяется как:

$$I_{\text{финр}}^{\text{исп}} = \frac{\Phi_{pi}}{\Phi_{\text{max}}}, \quad (9)$$

где Φ_{pi} – стоимость варианта исполнения, Φ_{max} – максимальная стоимость исполнения научно-исследовательского проекта.

Т.к. стоимость всех вариантов исполнения одинакова, интегральные финансовые показатели также будут одинаковы и равны 1.

Интегральный показатель ресурсоэффективности вариантов исполнения объекта исследования можно определить следующим образом:

$$I_{pi} = \sum a_i \cdot b_i, \quad (10)$$

где a_i – весовой коэффициент варианта исполнения разработки, b_i – балльная оценка варианта исполнения разработки.

Интегральный показатель эффективности вариантов исполнения разработки определяется на основании интегрального показателя ресурсоэффективности и интегрального финансового показателя по формулам:

$$\begin{aligned} I_{исп.1} &= \frac{I_{р-исп.1}}{I_{финр}}, \\ I_{исп.2} &= \frac{I_{р-исп.2}}{I_{финр}}. \end{aligned} \quad (11)$$

Так как интегральные финансовые показатели одинаковы и равны 1, то интегральные показатели эффективности вариантов исполнения разработки равны соответствующим интегральным показателям ресурсоэффективности.

Сравнение интегрального показателя эффективности вариантов исполнения разработки позволит определить сравнительную эффективность проекта и выбрать наиболее целесообразный вариант из предложенных.

Сравнительная эффективность проекта:

$$\mathcal{E}_{cp} = \frac{I_{исп.i}}{I_{исп.1}}. \quad (12)$$

В пункте 4.2 было рассмотрено два варианта исполнения алгоритма. На основании этого необходимо провести сравнительную характеристику вариантов исполнения (табл. 12).

Таблица 12 – Сравнительная оценка характеристик вариантов исполнения проекта

Критерии	Весовой коэффициент	Исп. 1	Исп. 2
1. Повышение производительности труда	0,3	5	5
2. Удобство в эксплуатации	0,2	4	3
3. Удобство в считывании исходных данных	0,2	5	2
4. Скорость работы	0,1	4	3
5. Простота эксплуатации	0,1	4	3
6. Техническая поддержка платформы	0,1	5	3
I_{pi}		4,6	3,4

На основании полученных показателей выполним сравнение интегрального показателя эффективности вариантов исполнения разработки (табл. 13).

Таблица 13 – Сравнительная эффективность разработки

Показатели	Исп.1	Исп.2
Интегральный финансовый показатель разработки	1	1
Интегральный показатель ресурсоэффективности разработки	4,6	3,4
Интегральный показатель эффективности	4,6	3,4
Сравнительная эффективность вариантов исполнения	1	0,74

С позиции финансовой и ресурсной эффективности на основании таблицы, первый вариант исполнения системы наиболее выгодный. Данный вариант исполнения и используется в выпускной квалификационной работе.

Заключение

Целью данной выпускной квалификационной работы является разработка метода и исследование алгоритма динамической кластеризации.

Для этого были решены следующие задачи:

- Предложен метод расчета средних значений из исходного массива данных;
- Проведен анализ существующих и актуальных на данный момент методов группировки. Выявлены их преимущества и недостатки;
- Реализован метод расчета средних значений из исходного массива данных. Данный метод повышает вычислительную точность алгоритма;
- Подготовлен и организован численный эксперимент;
- Проведена оценка устойчивости алгоритма.
- Сделаны выводы по устойчивости алгоритма. В большинстве случаев алгоритм сохраняет устойчивость вплоть до 70%, исходя из чего можно сделать вывод о том, что алгоритм обладает высокой устойчивостью.
- Были выявлены факторы, оказывающие наибольшее влияние на устойчивость алгоритма: значительное влияние на ошибку оказывает величина корреляции между эталонными сигналами, так как величина корреляции напрямую влияет на поведение алгоритма на первых итерациях. Ошибка вычислений также росла при увеличении ширины шума.

Список использованных источников

1. Метод k-средних [Электронный ресурс]. – URL: <http://datascientist.one/k-means-algorithm/> (дата обращения 25.04.19).
2. Н.Н. Буреева. Многомерный статистический анализ с использованием ППП “STATISTICA”. Нижний Новгород: «ННГУ», 2007. 112 с.
3. В.Н. Калинина, В. И. Соловьев. Введение в многомерный статистический анализ: Учебное пособие / ГУУ. М., 2003. 66 с.
4. СН 2.2.4/2.1.8.562–96. Шум на рабочих местах, в помещениях жилых, общественных зданий и на территории застройки.
5. СанПиН 2.2.548-96. Гигиенические требования к микроклимату производственных помещений.
6. СанПиН 2.2.2/2.4.1340-03 Гигиенические требования к персональным электронно-вычислительным машинам и организации работы.
7. Гигиенические требования к видеодисплейным терминалам, персональным электронно-вычислительным машинам и организации работы. Выписки из СанПиН 2.2.2.542-96.
8. СНиП 23-05-95 Естественное и искусственное освещение.
9. ГОСТ 12.1.019 (с изм. №1) ССБТ. Электробезопасность. Общие требования и номенклатура видов защиты.
10. Федеральный закон Российской Федерации от 22 июля 2008 г. N 123-ФЗ "Технический регламент о требованиях пожарной безопасности".
11. Федеральный классификационный каталог отходов [Электронный ресурс] – 2013. – Режим доступа: <http://www.ecoguild.ru/>, свободный. – Загл. с экрана.
12. ТОИ Р-45-084-01. Типовая инструкция по охране труда при работе на персональном компьютере.

Приложения

Приложение А – последовательность алгоритма

Алгоритм динамической группировки

1. Расчет взаимной корреляции
2. Выделение групп по коэффициенту взаимной корреляции
3. Построение типовых функции
4. Получение единого массива типовых функций
5. Массив типовых функций вычитается поэлементно из массива фазовых значений
6. Повторение итераций

Приложение Б – Блок-схема алгоритма

