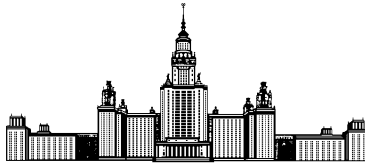


Московский государственный университет имени М. В. Ломоносова



Факультет Вычислительной Математики и Кибернетики

Кафедра Математических Методов Прогнозирования

## КУРСОВАЯ РАБОТА СТУДЕНТА 317 ГРУППЫ

### «Сравнительный анализ методов быстрого поиска ближайших соседей»

Выполнил:

студент 3 курса 317 группы

*Федоров Илья Сергеевич*

Научный руководитель:

д.ф.-м.н., профессор

*Дьяконов Александр Геннадьевич*

Заведующий кафедрой

Математических Методов

Прогнозирования, академик РАН

\_\_\_\_\_ Ю. И. Журавлёв

К защите допускаю

«\_\_\_\_\_» \_\_\_\_\_ 2020 г.

К защите рекомендую

«\_\_\_\_\_» \_\_\_\_\_ 2020 г.

Москва, 2020

# Содержание

<b>1</b>	<b>Введение</b>	<b>3</b>
1.1	Определения и обозначения . . . . .	3
1.2	Обзор литературы . . . . .	4
<b>2</b>	<b>Новые подходы и результаты</b>	<b>4</b>
<b>3</b>	<b>Вычислительные эксперименты</b>	<b>4</b>
3.1	Исходные данные и условия эксперимента . . . . .	5
3.2	Результаты эксперимента . . . . .	5
3.3	Обсуждение и выводы . . . . .	5
<b>4</b>	<b>Заключение</b>	<b>5</b>

## **Аннотация**

Todo

# 1 Введение

Одним из наиболее простых и естественных методов машинного обучения является метод ближайшего соседа. Имея набор данных, представленных в виде точек в некотором многомерном пространстве, целевая величина (будь то класс или вещественное число) прогнозируется по значениям отклика на  $k$  ближайших к запросу точках из исходного набора данных. В то время как существует множество различных подходов к усреднению данных  $k$  значений, наиболее вычислительно затратной частью алгоритмов подобного типа является именно поиск ближайших соседей. Действительно, в современных задачах объемы данных достигают колоссальных размеров, что делает алгоритмы, основанные на полном переборе, неэффективными. Задача поиска ближайших к запросу точек в некотором наборе данных встречается не только в задачах прогнозирования. Примерами приложений могут служить задачи поиска дубликатов в больших объемах данных (или «почти» дубликатов), поиска похожих изображений и текстов. Целью данной работы является обзор современных подходов к решению задачи поиска ближайших соседей и её вариаций, а также сравнительный анализ эффективности тех или иных методов её решения в зависимости от особенностей пространства, в котором расположены данные. В исследовании представлены как классические подходы, основанные на формировании некоторых дополнительных структур данных, так и наиболее современные «приближенные» методы.

## 1.1 Определения и обозначения

Формальная постановка задачи. Для известных понятий желательно придерживаться стандартных обозначений. Общепринятые термины вводятся словом «называется». Термины, придуманные самим автором, вводятся словами «назовём» или «будем называть». Обычно этот раздел заканчивается формальной постановкой задачи. Именно с этого раздела стоит начинать писать работу.

## 1.2 Обзор литературы

Лучше, чтобы название этого подраздела было содержательным, например, общепринятым названием задачи, проблемы или метода, рассматриваемого в данной работе.

Перечисляются подходы, методы, факты, на которые существенно опирается данная работа, но которые могут быть не известны широкому кругу читателей. Здесь ссылки на литературу обязательны. Теоремы только формулируются, но не доказываются.

Данный раздел преследует две цели. Во-первых, сделать работу самодостаточной — дать необходимый минимум информации тем читателям, которые не очень хорошо ориентируются в теме, но желают поближе познакомиться именно с данной работой. Во-вторых, облегчить сопоставление полученных автором результатов с ранее известными.

## 2 Новые подходы и результаты

Название этого раздела обязательно надо заменить на содержательное. В этом разделе, как правило, много подразделов.

В дипломной работе не стоит делать более двух уровней, достаточно разделов и подразделов. Будете писать диссертацию или монографию — сделаете три уровня.

## 3 Вычислительные эксперименты

Цель данного раздела: продемонстрировать, что предложенная теория работает на практике; показать границы её применимости; рассказать о новых экспериментальных фактах.

Чисто теоретические работы могут вообще не содержать раздела экспериментов (не работает, ну и не надо — зато теория красивая). Кстати, теоретики имеют право не догадываться, где, кому и когда их теории пригодятся.

### 3.1 Исходные данные и условия эксперимента

Описывается прикладная задача, параметры анализируемых данных (например, сколько объектов, сколько признаков, каких они типов), параметры эксперимента (например, как производился скользящий контроль).

### 3.2 Результаты эксперимента

Результаты экспериментов представляются в виде таблиц и графиков. Объясняется точный смысл всех обозначений на графиках, строк и столбцов в таблицах.

### 3.3 Обсуждение и выводы

Приводятся выводы: в какой степени результаты экспериментов согласуются с теорией? Достигнут ли желаемый результат? Обнаружены ли какие-либо факты, не нашедшие объяснения, и которые нельзя списать на «грязный» эксперимент?

Обсуждаются основные отличия предложенных методов от известных ранее. В чем их преимущества? Каковы границы их применимости? Какие проблемы удалось решить, а какие остались открытыми? Какие возникли новые постановки задач?

## 4 Заключение

В квалификационных работах последний раздел нужен для того, чтобы конспективно перечислить основные результаты, полученные лично автором.

Результатами, в частности, являются:

- Предложен новый подход к...
- Разработан новый метод..., позволяющий...
- Доказан ряд теорем, подтверждающих (опровергающих), что...
- Проведены вычислительные эксперименты..., которые подтвердили / опровергли / привели к новым постановкам задач.

Цель данного раздела: доказать квалификацию автора. Даже беглого взгляда на заключение должно быть достаточно, чтобы стало ясно: автору удалось решить

актуальную, трудную, ранее не решённую задачу, предложенные автором решения обоснованы и проверены.

Иногда в Заключении приводится список направлений дальнейших исследований.

Список литературы необходим в любой научной публикации. В дипломной работе он обязателен. Дурным тоном считается: ссылаться на работы только одного-двух авторов (например, себя или шефа); ссылаться на слишком малое число работ; ссылаться только на очень старые работы; ссылаться на работы, которых автор ни разу не видел; ссылаться на работы, которые не упоминаются в тексте или которые не имеют отношения к данному тексту.