

Санкт-Петербургский государственный университет
математико-механический факультет
кафедра системного программирования

Разработка кроссплатформенного приложения для кластерного анализа данных на основе рандомизированных алгоритмов

Дипломная работа
студента 544 группы

Морозкова Михаила
Андреевича

Научный руководитель

к. ф.-м. н.

Шалымов Д. С.

Рецензент

д. ф.-м. н., проф.

Граничин О. Н.

2010 г

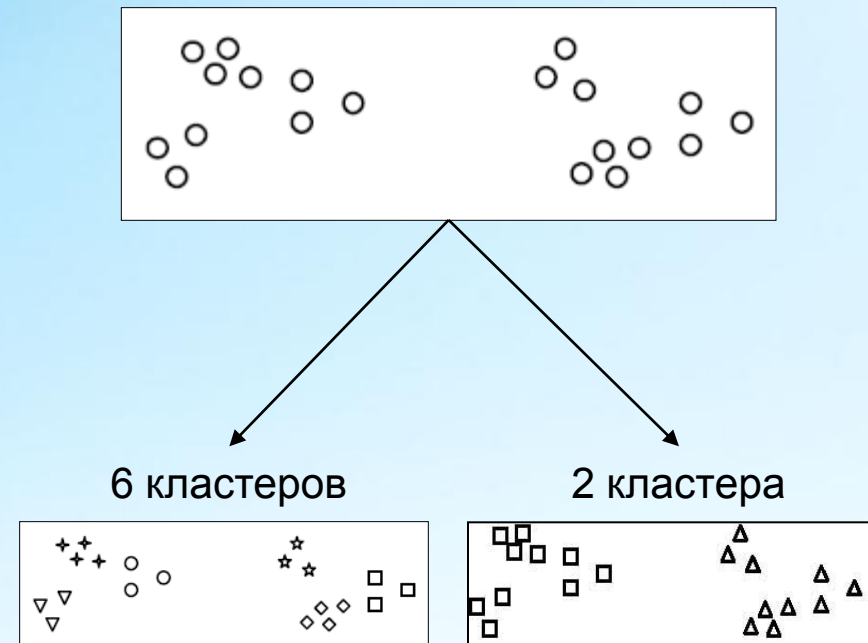
Введение

Проблемы:

- Качество кластеризации
- Трудоёмкость задачи
- Обоснованность

Решение:

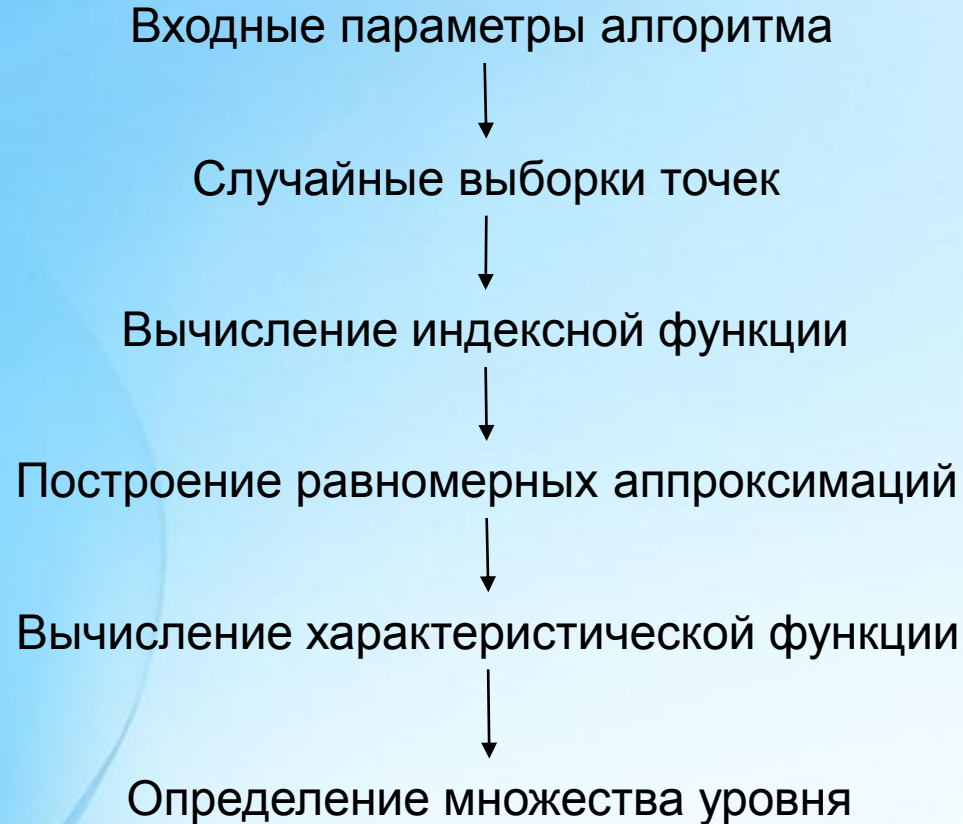
- Рандомизированные алгоритмы
- Сценарный подход



Постановка задачи

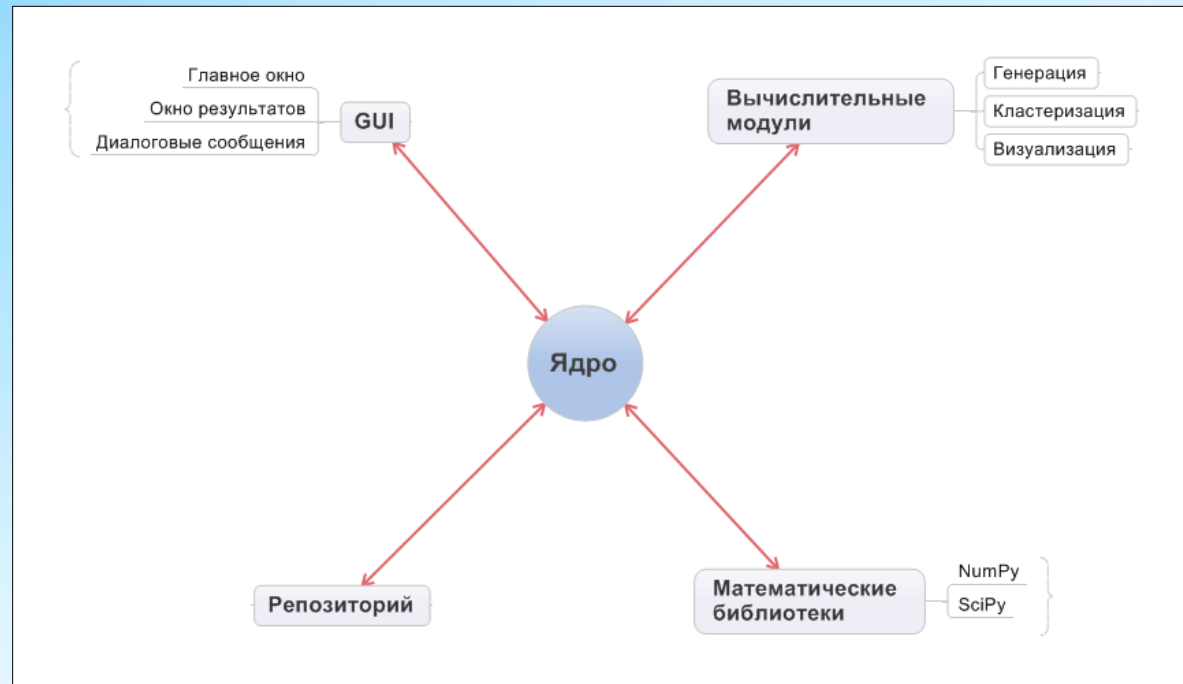
- Реализация алгоритма
- Апробация
- Сравнительный анализ
- Исследование зависимости от параметров

Рандомизированный алгоритм

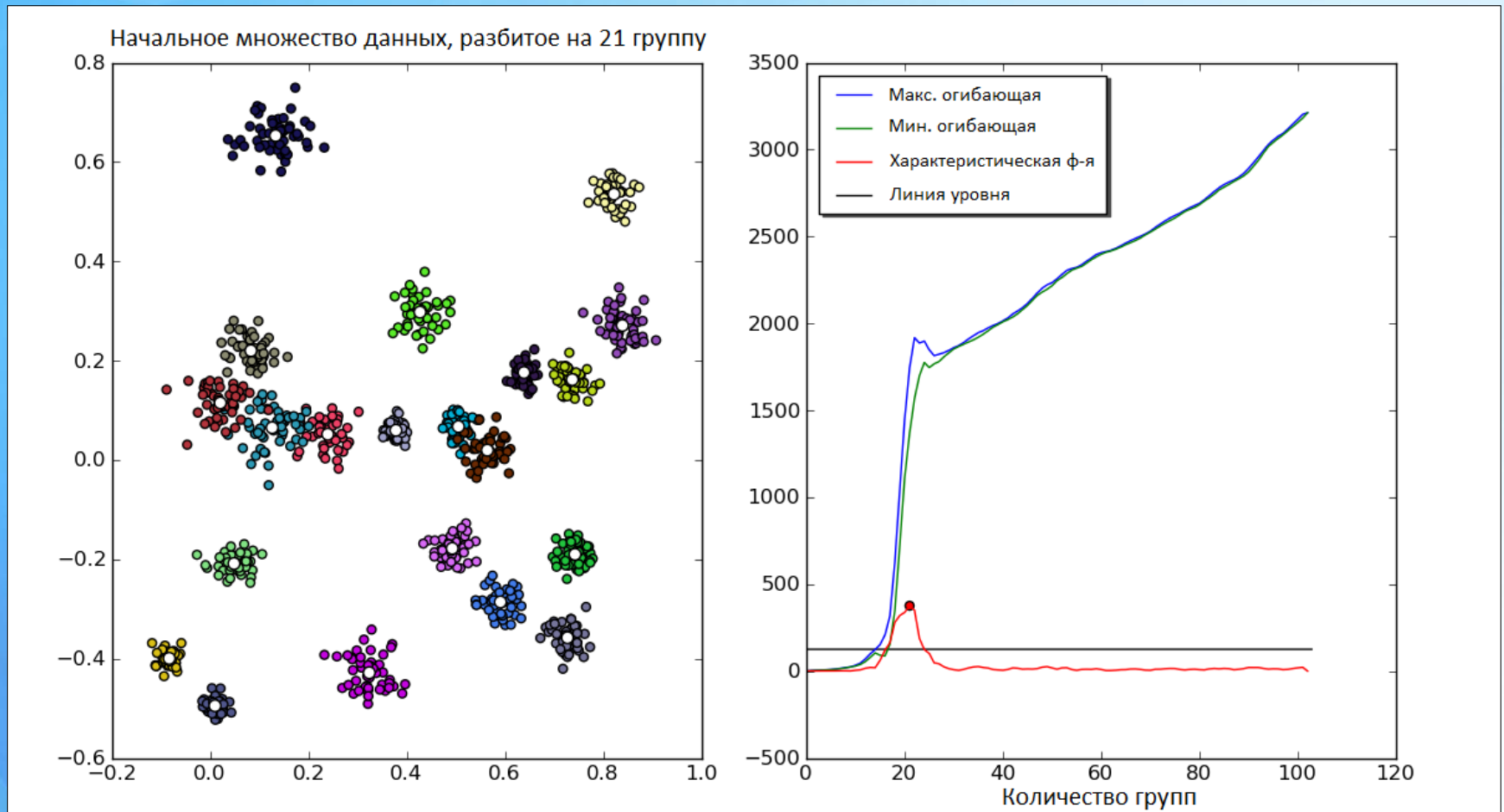


Реализация

- Python
- QT & PyQt
- NumPy, SciPy
- Matplotlib



Результаты моделирования



Сравнительный анализ

| Тест | Алгоритм | Среднее время, сек | % правильных ответов |
|---|--------------------------|--------------------|----------------------|
| Тест №1 (16 четырехмерных кластеров) | Рандомизированный | 66 | 81 |
| | Сьюгер-Джеймс | 156 | 91 |
| | Хартиган | 161 | 65 |
| | Кржановский-Лай | 153 | 35 |
| Тест №2 (16 десятимерных кластеров) | Рандомизированный | 56 | 83 |
| | Сьюгер-Джеймс | 144 | 96 |
| | Хартиган | 140 | 100 |
| | Кржановский-Лай | 144 | 80 |
| Тест №3 (22 четырехмерных кластера, объединенных в 3 надкластера) | Рандомизированный | 90 | 81 |
| | Сьюгер-Джеймс | 211 | 93 |
| | Хартиган | 214 | 76 |
| | Кржановский-Лай | 211 | 71 |

Результаты

- Разработано кроссплатформенное приложение для кластерного анализа данных
- Реализован новый рандомизированный алгоритм
- Подтверждены оценки состоятельности
- Сделаны выводы о зависимости работоспособности от параметров