

Кластеризация путей абонентов мобильного оператора в евклидовом пространстве

Федоров Роман, 371 группа

Научный руководитель: Константин Невоструев

Цели работы

Постановка задачи

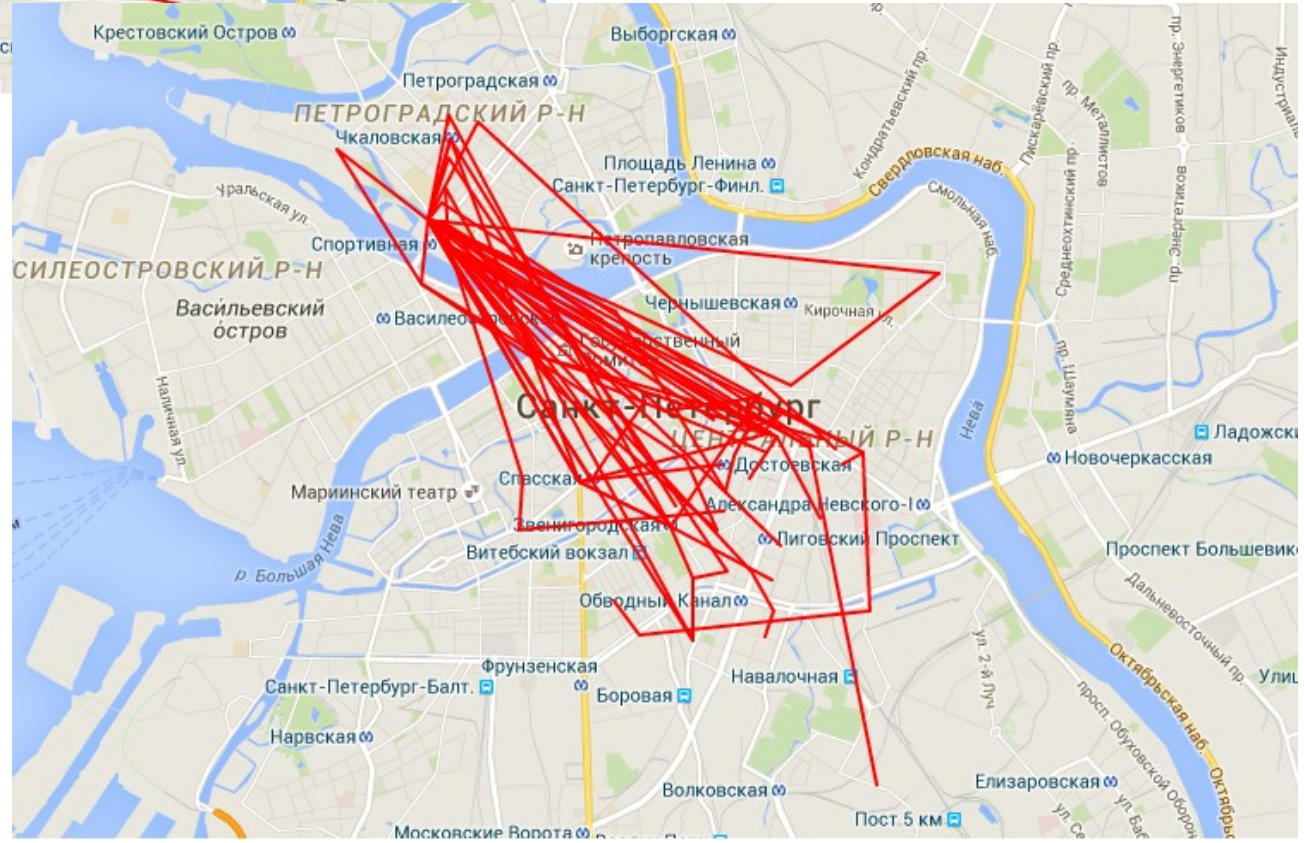
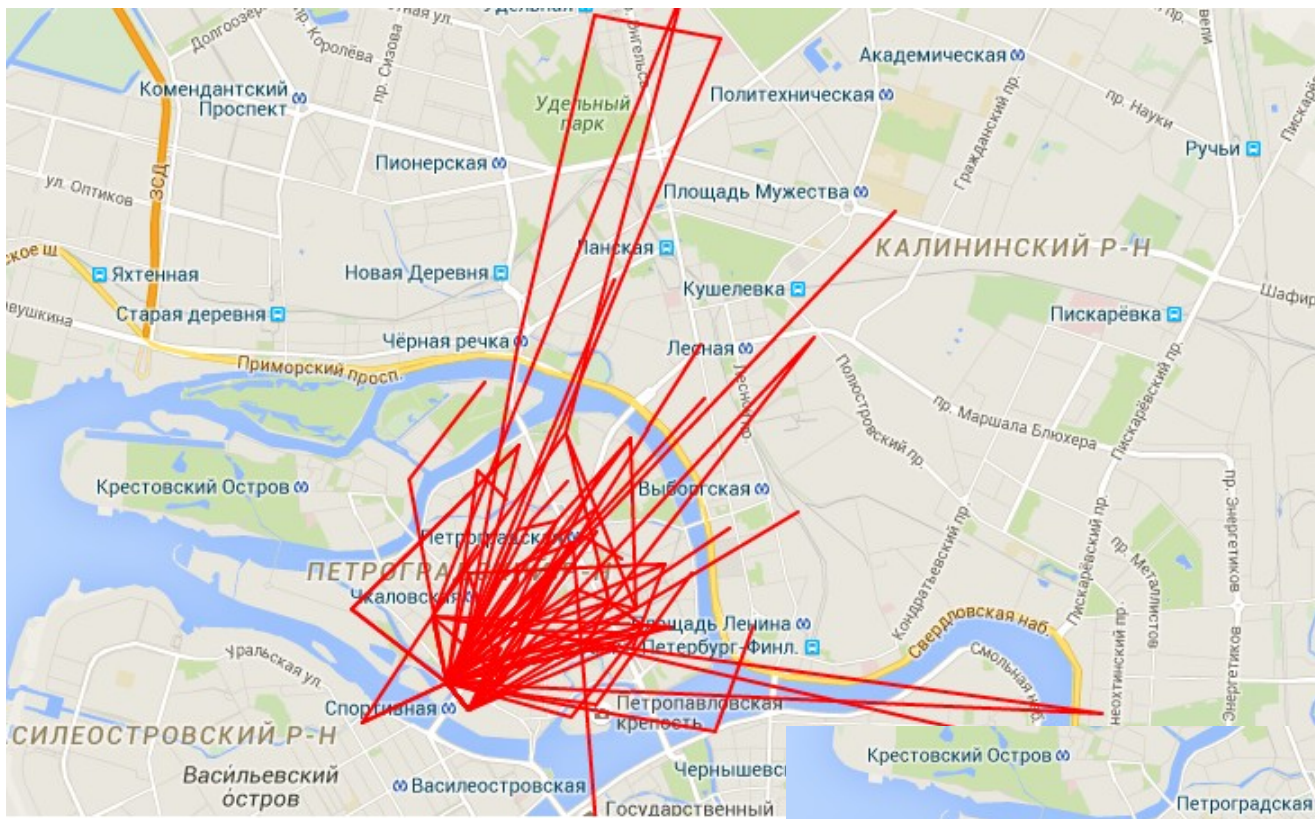
- Используя данные об активностях абонентов на базовых станциях мобильного оператора, выделить маршруты абонентов
- Объединить схожие маршруты в группы, используя методы кластеризации
- Выделить наиболее популярные маршруты пользователей
- Возможное применение: эффективное размещение рекламы

Описание работы (1)

Подготовка данных для анализа:

- ✓ Перевод из табличного представления в последовательности базовых станций для каждого абонента – пути
- ✓ Фильтрация путей по количеству базовых станций
- ✓ Фильтрация выбросов

Выравнивание путей и первоначальная кластеризация методом k-средних.



Описание работы (2)

Перевод из абсолютных координат в относительные.

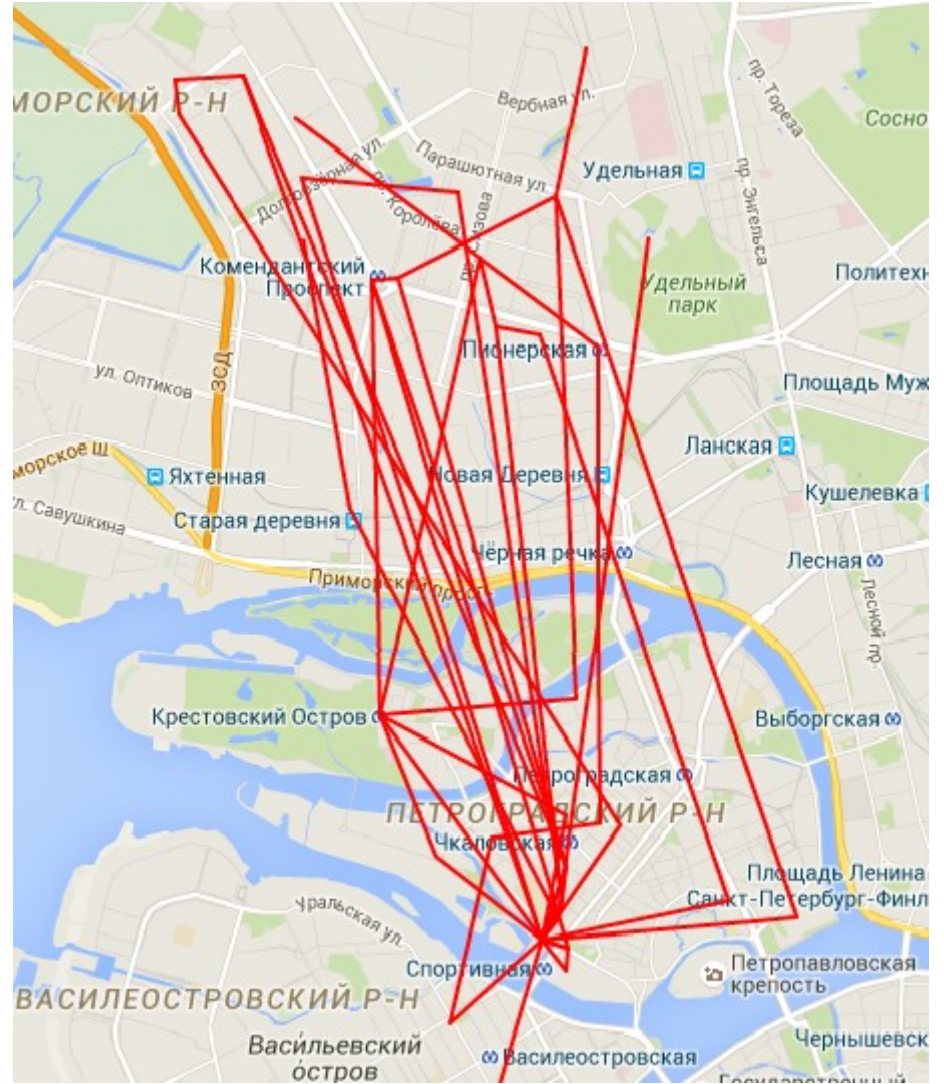
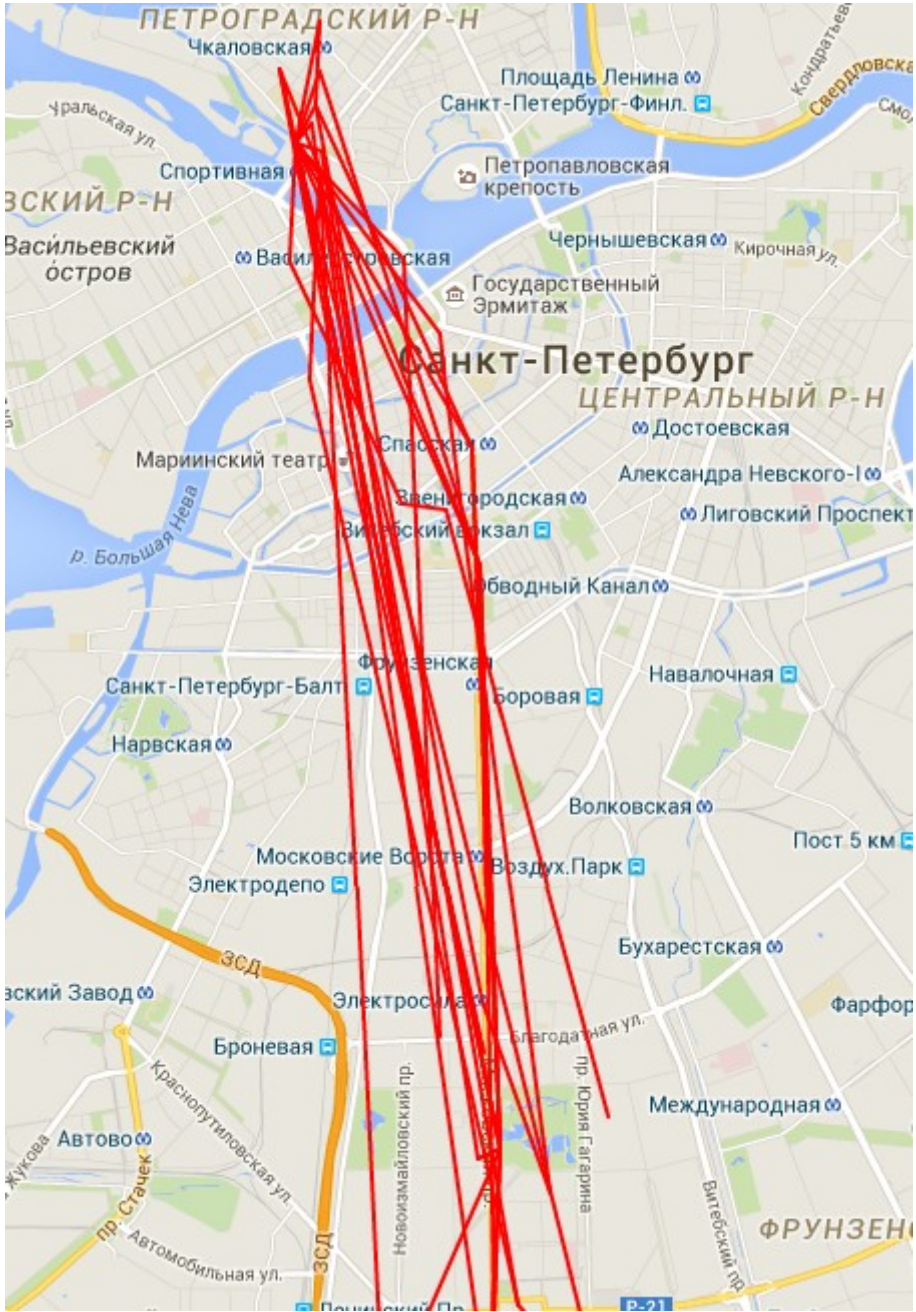
Применение различных функций к координатам.

Новые методы кластеризации:

- ✓ Агломеративная кластеризация
- ✓ BIRCH
- ✓ DBSCAN

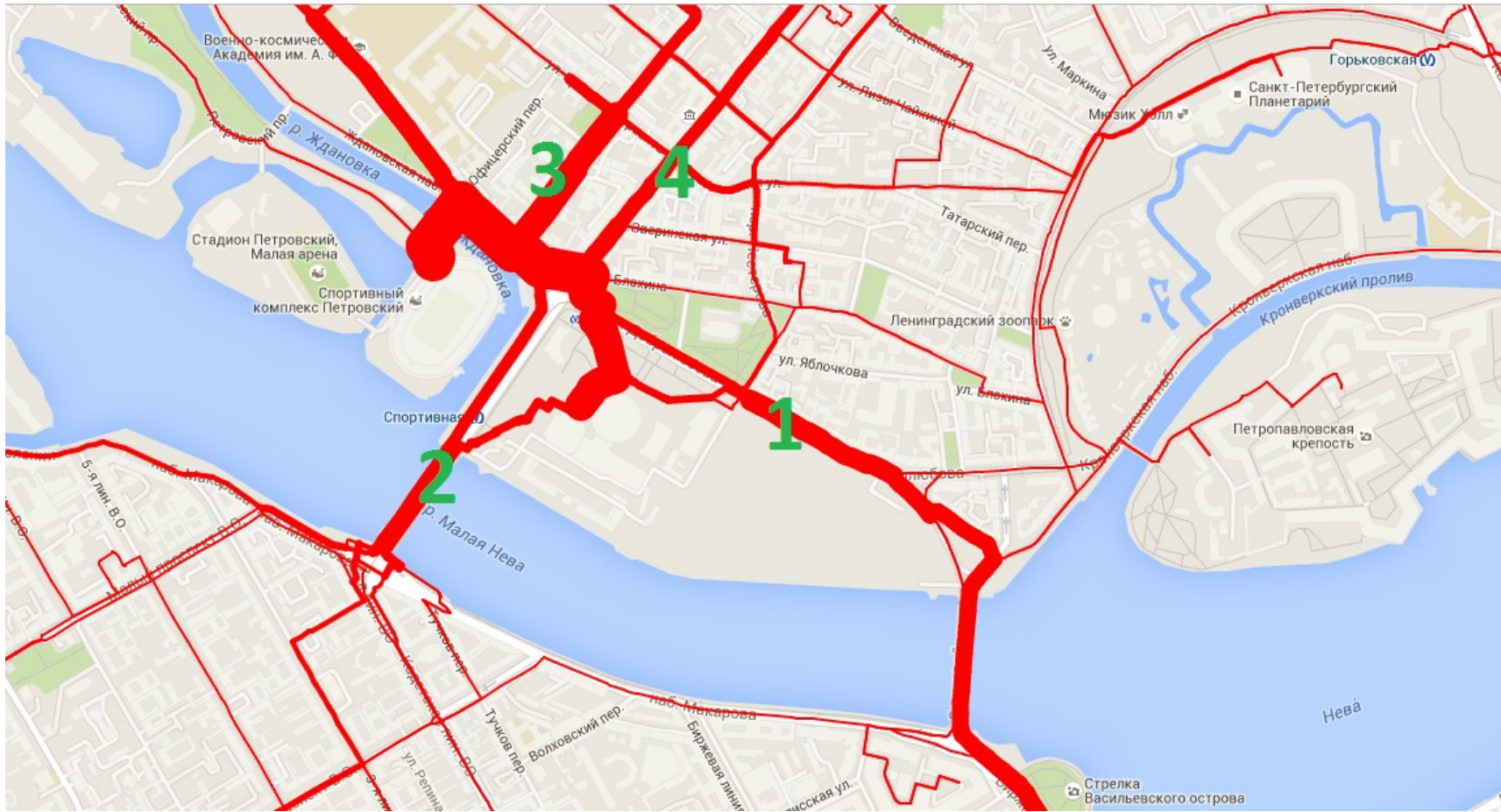
Оценка кластеризации:

- ✓ Визуальная
- ✓ Коэффициент силуэта
- ✓ Диаметр кластера



Описание работы (3)

- Визуализация популярных маршрутов, популярных отрезков маршрутов по кластерам и по всем данным.
- Сравнение результатов на данных за разные дни.
- Оценка популярных отрезков маршрутов с точки зрения эффективного размещения рекламы с помощью Daily Effective Circulation.



Количество отрезков	1	2	3	4
Нормированный D.E.C.	24.2%	44.5%	59.7%	69.6%

ИТОГИ

- Разработаны и применены различные способы сравнения путей абонентов
- Изучены и применены различные алгоритмы кластеризации, сделан перебор параметров алгоритмов для достижения наилучших оценок качества
- Разработаны средства для визуализации результатов кластеризации
- Пути абонентов кластеризованы, выделены различные группы абонентов, найдены наиболее популярные пути, проведена их оценка