



Метод кластеризации предпочтений жителей города по перемещению

Автор: Чечеткин И. А.

группа: САПР-2п1

Руководитель: Щербаков М. В.

Волгоград 2016 г.

Целью данной работы являлась разработка метода кластеризации предпочтений жителей, выраженных в паре геораспределенных объектов «отправление-назначение», с учетом городского рельефа.

В данной работе рассматриваются к решению следующие задачи:

- генерация псевдореалистичных данных о перемещениях жителей;
- разработка метрики расстояний, учитывающей рельеф местности;
- модификация и использование существующих алгоритмов для кластеризации геораспределенных данных с разработанной метрикой расстояний;
- представление построенных кластеров на карте.

Актуальность

Изменения в городской среде требуют формирования новых механизмов планирования инфраструктуры города. Для получения эффективных результатов, следует осуществлять принятие решений на основе актуальных данных, отражающих предпочтения жителей.



Научная новизна работы обусловлена:

- предложена автоматизация процесса построения сети остановочных пунктов без участия транспортного инженера;
- предложен метод, основанный на использовании данных о предпочтении жителей для построения сети остановочных пунктов;
- разработан метод кластеризации географических точек с учетом препятствий.

Анализ предметной области

Существующие инструменты для проведения комплексного анализа:

- WEKA;
- RapidMiner;
- R;
- MatLab;
- Scikit-Learn.



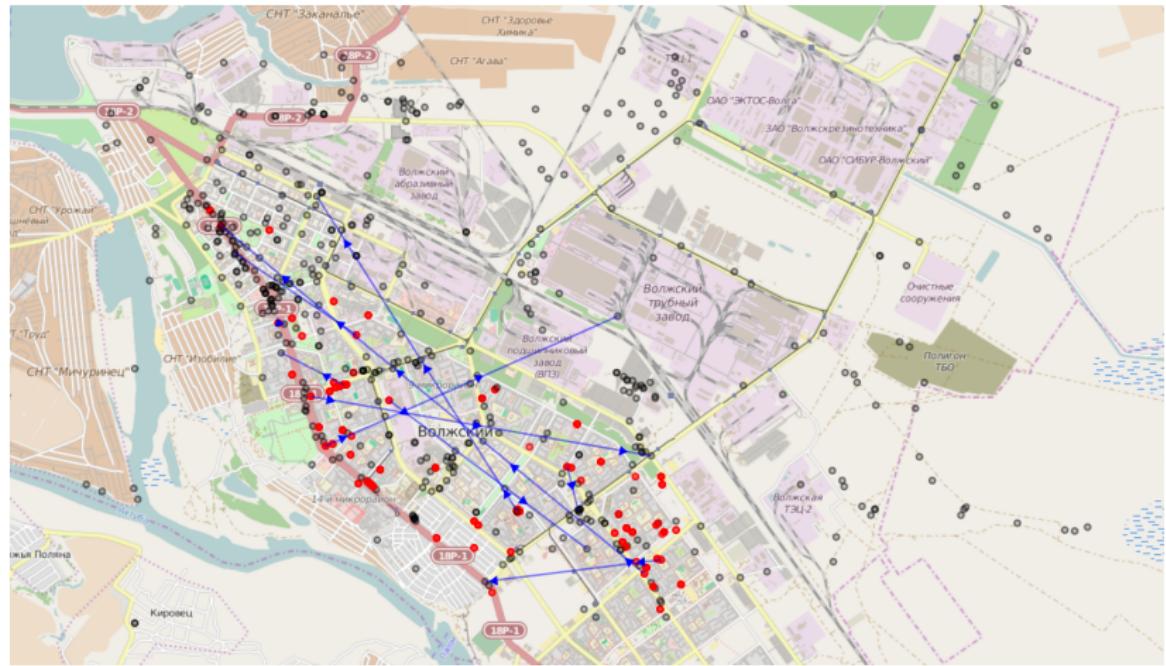
Постановка задачи

Объект исследования – кластеризация предпочтений жителей города по перемещению, выраженных в виде пары точек «Пункт отправления – пункт назначения», каждая из которых содержит две координаты – широту и долготу.

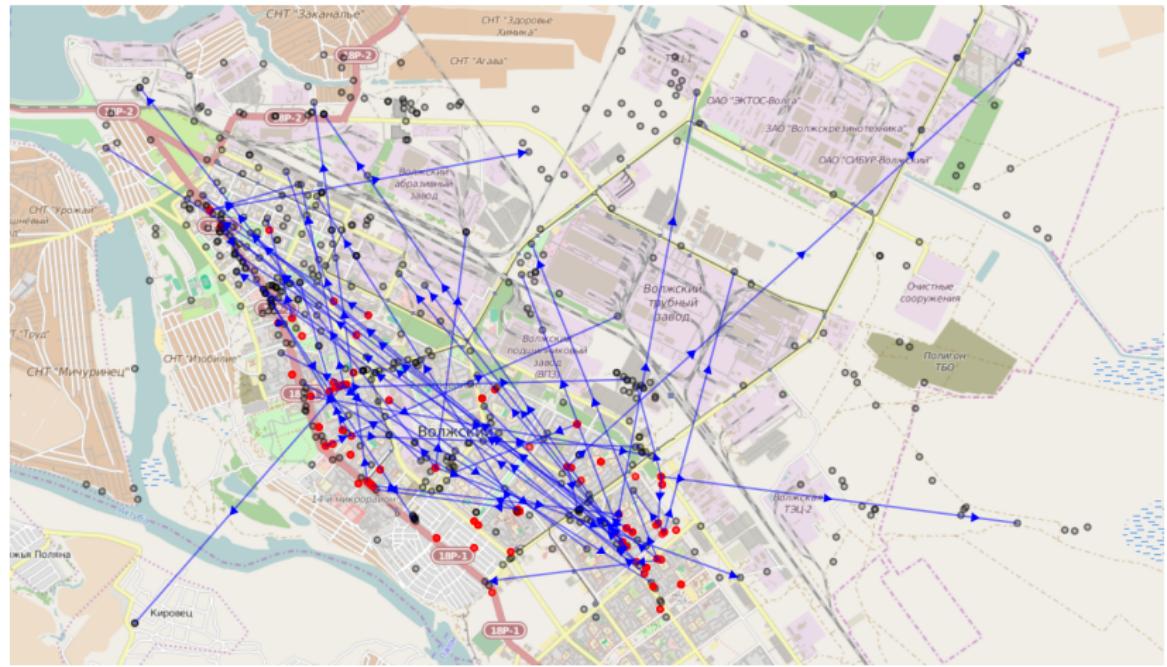
Предмет исследования – разработка и применение методов кластеризации предпочтений жителей, учитывающих элементы рельефа местности.

Гипотеза исследования – Использование данных о перемещениях жителей и автоматизация построения сети остановочных пунктов и маршрутной сети позволяют построить оптимальную сеть маршрутов общественного транспорта.

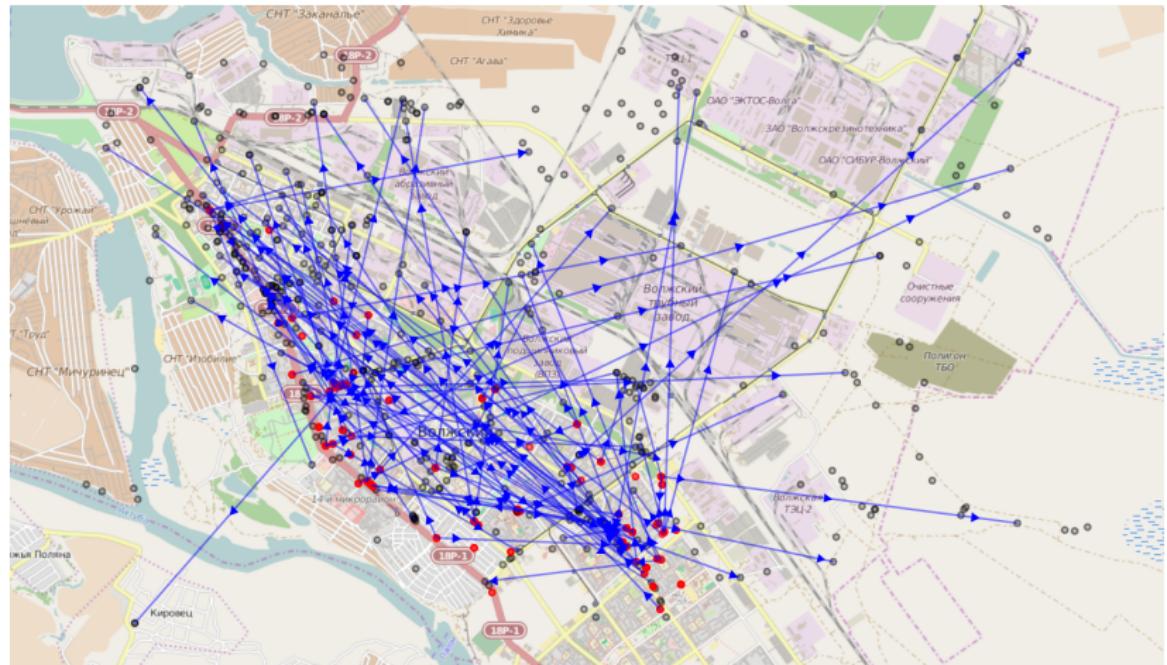
Входные данные



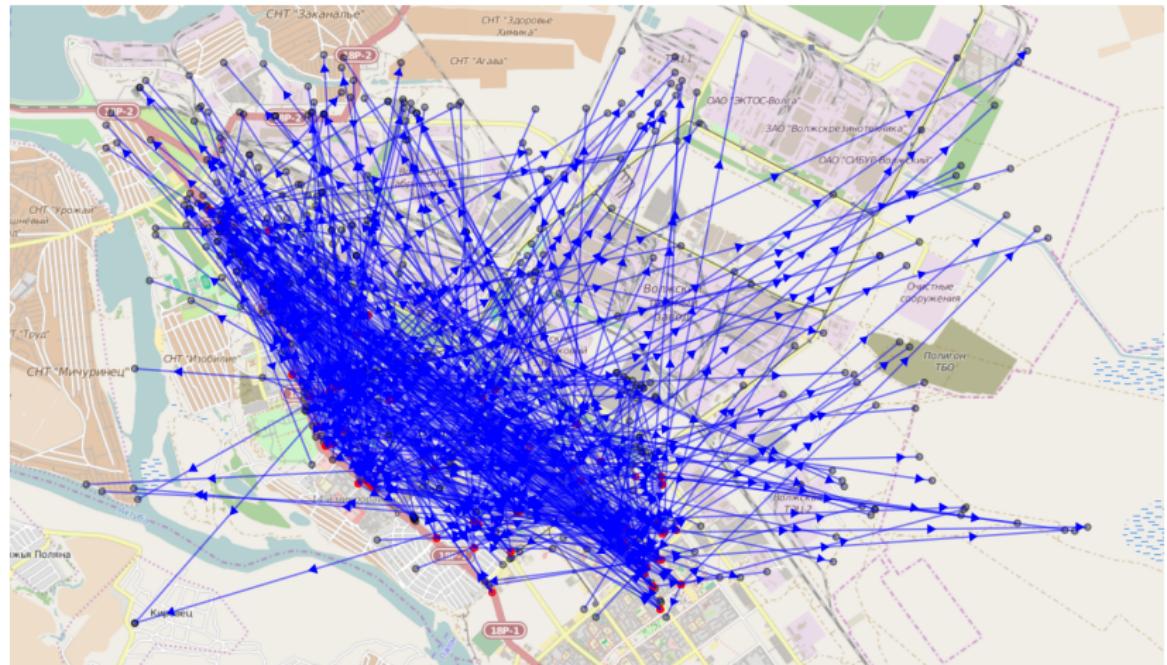
Входные данные



Входные данные

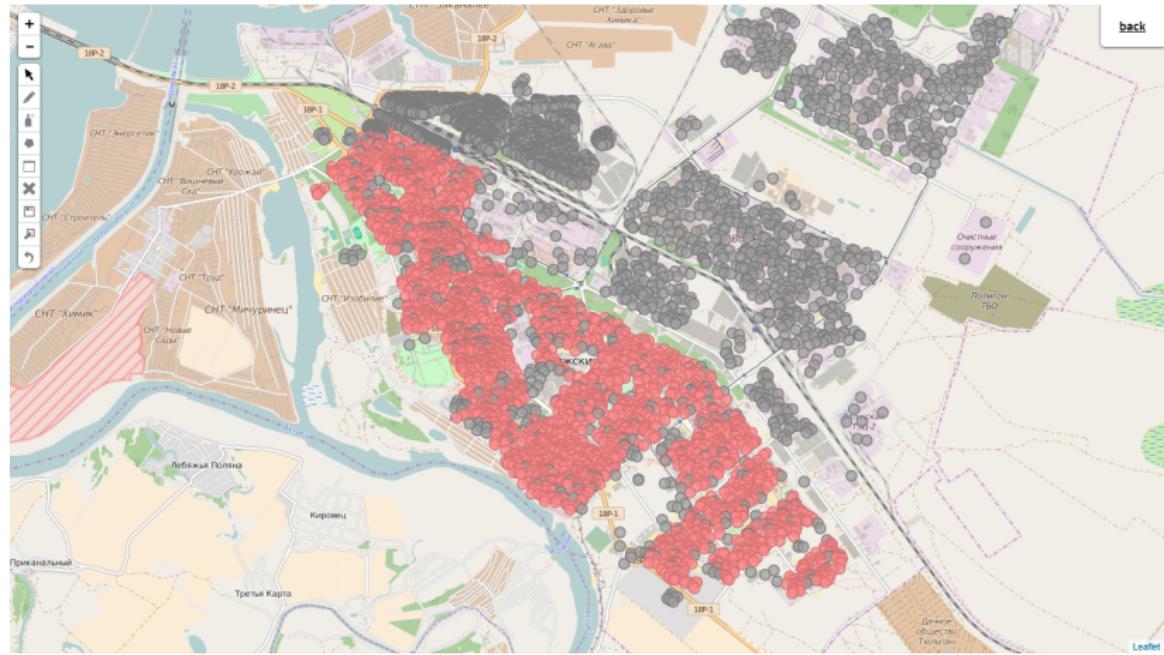


Входные данные

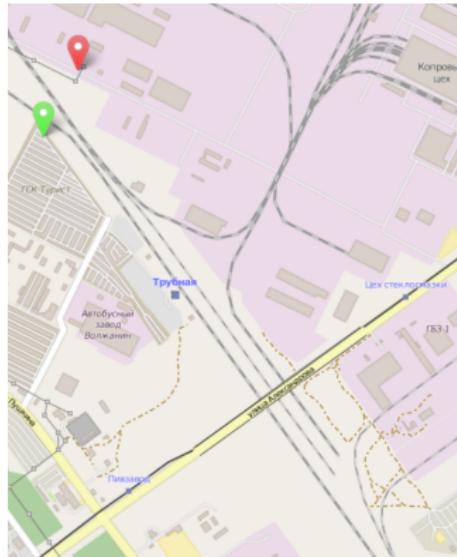
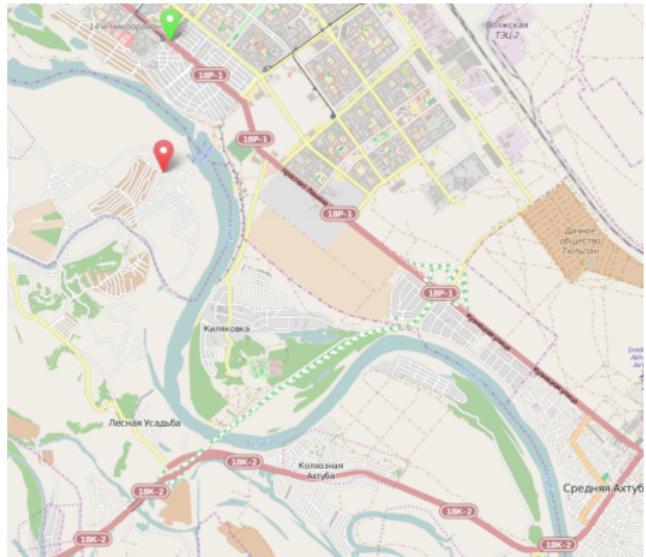


Входные данные

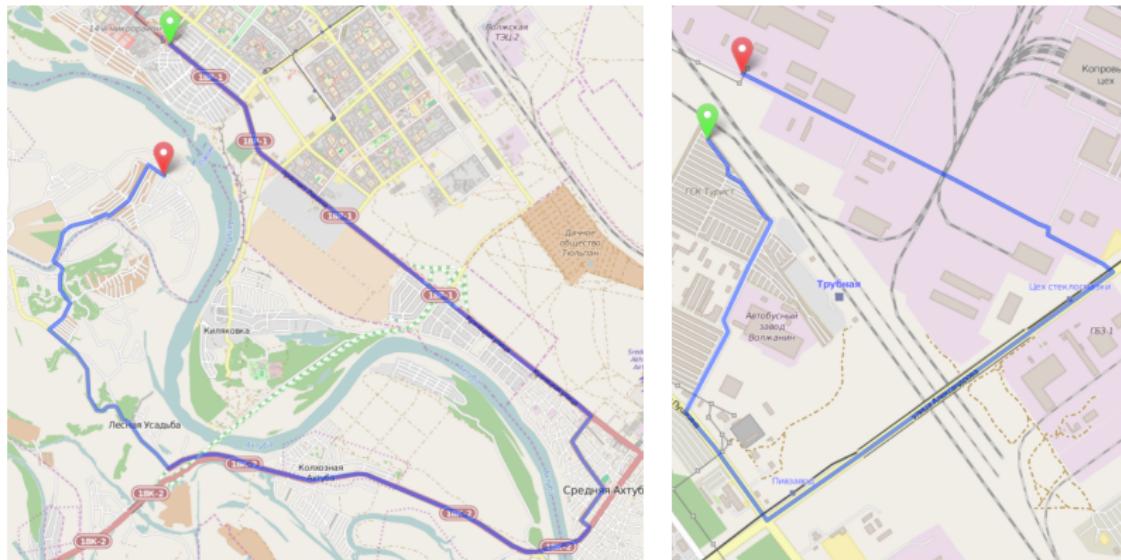
Сгенерированные данные



Метрика

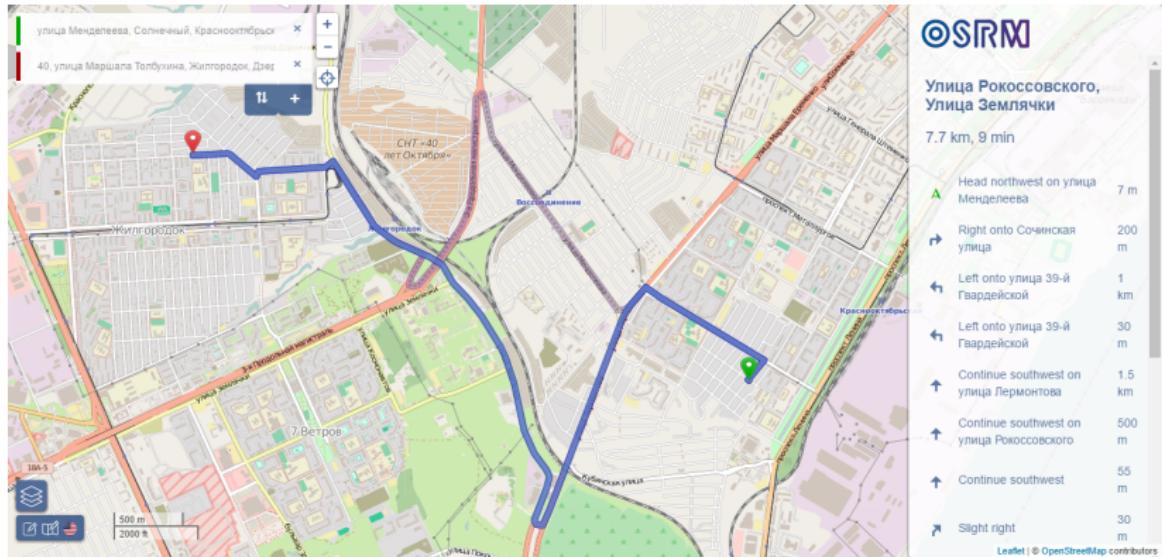


Метрика

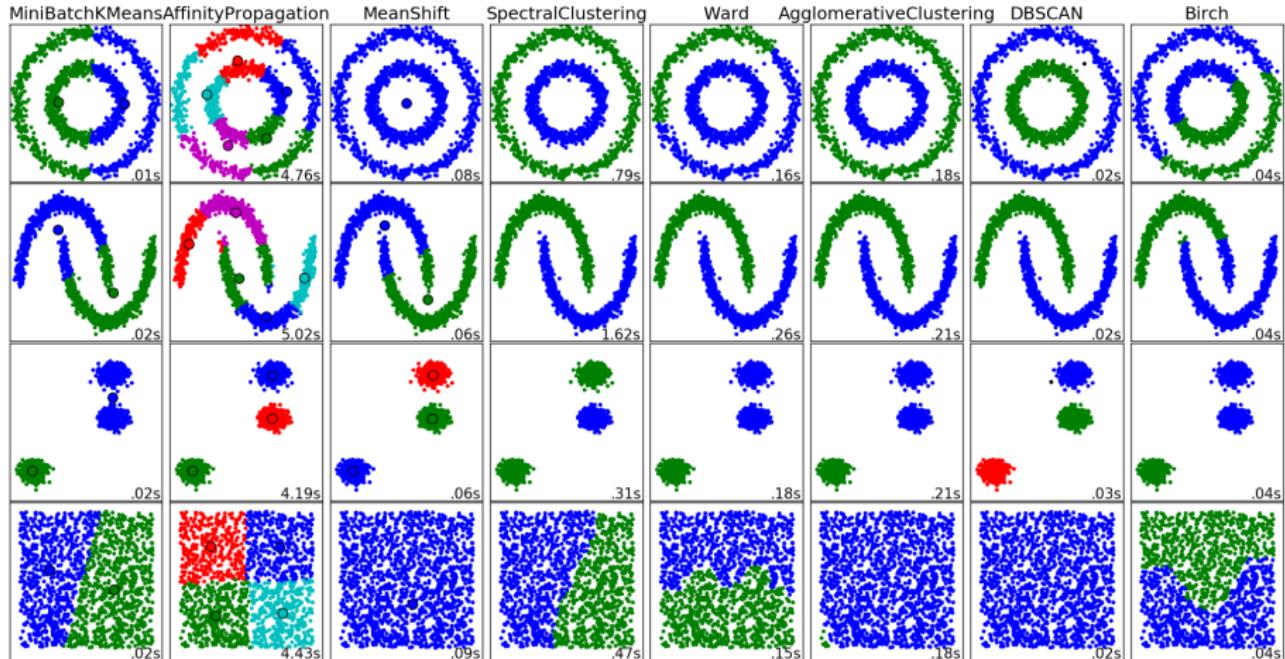


Метрика

Пример работы OSRM



Алгоритмы кластеризации



Алгоритмы кластеризации

Алгоритм k-means

Исходные параметры: $X, k, C_0, I.$

Результат: $C, L.$

1. $C = C_0, L_0 = \emptyset, i = 0;$
2. **До тех пор, пока** $L \neq L_0$ **и** $i < I$ **выполнять**

1. $L_0 = L, A = \emptyset, \mu = \emptyset;$
2. **Для каждого** $x \in X$ **выполнить**
 1. $r = \emptyset;$
 2. **Для каждого** $c \in C$ **выполнить**
 1. Рассчитать $r_{xc} = \rho(x, c);$
 2. $r = r + \{r_{xc}\};$

конец цикла

3. $L[x] = C[\min(r)];$
4. $A[C[\min(r)]] = A[C[\min(r)]] + \{x\};$

конец цикла

3. **Для каждого** $c \in C$ **выполнить**

1. $\mu[c] = \sum(A[c]) / |A[c]|;$
2. $c = \mu[c];$

конец цикла

4. $i = i + 1;$

конец цикла

Проектирование программного продукта

Используемые технологии:

- Python 3 (<https://www.python.org/>)
- OSRM (<http://project-osrm.org/>)
- OpenStreetMap (<https://www.openstreetmap.org>)
- Leaflet (<http://leafletjs.com/>)

Исходный код доступен по следующей ссылке:

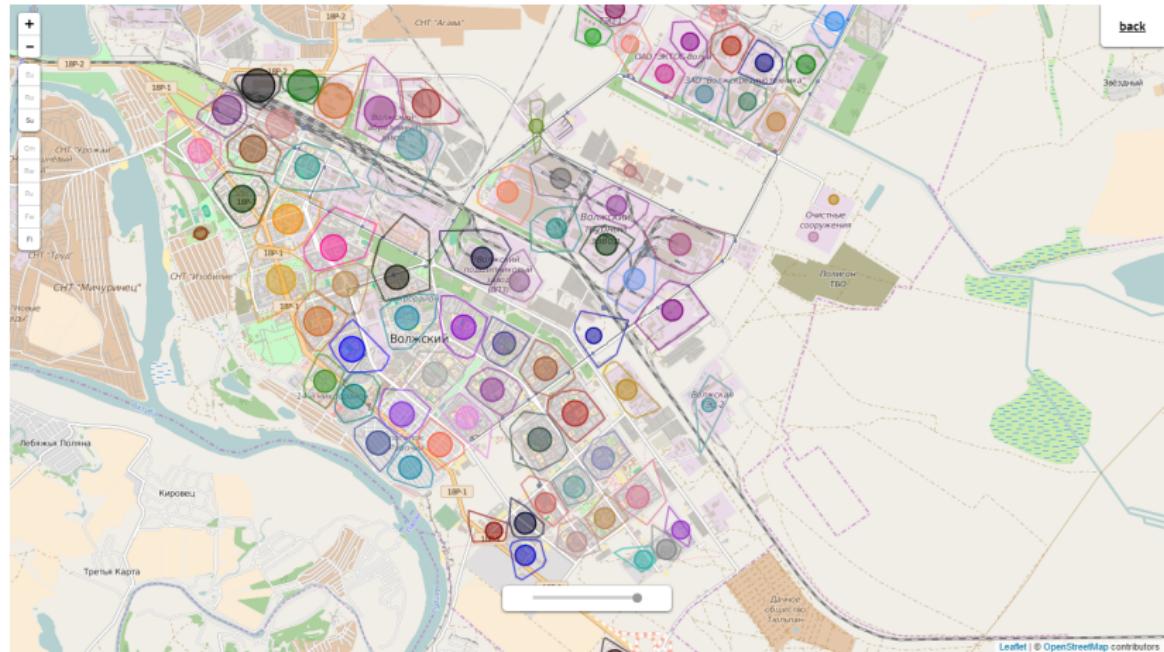
<https://github.com/vstu-cad-stuff/clustering/tree/master>

Результаты работы реализованных алгоритмов:

<https://vstu-cad-stuff.github.io/clustering/>

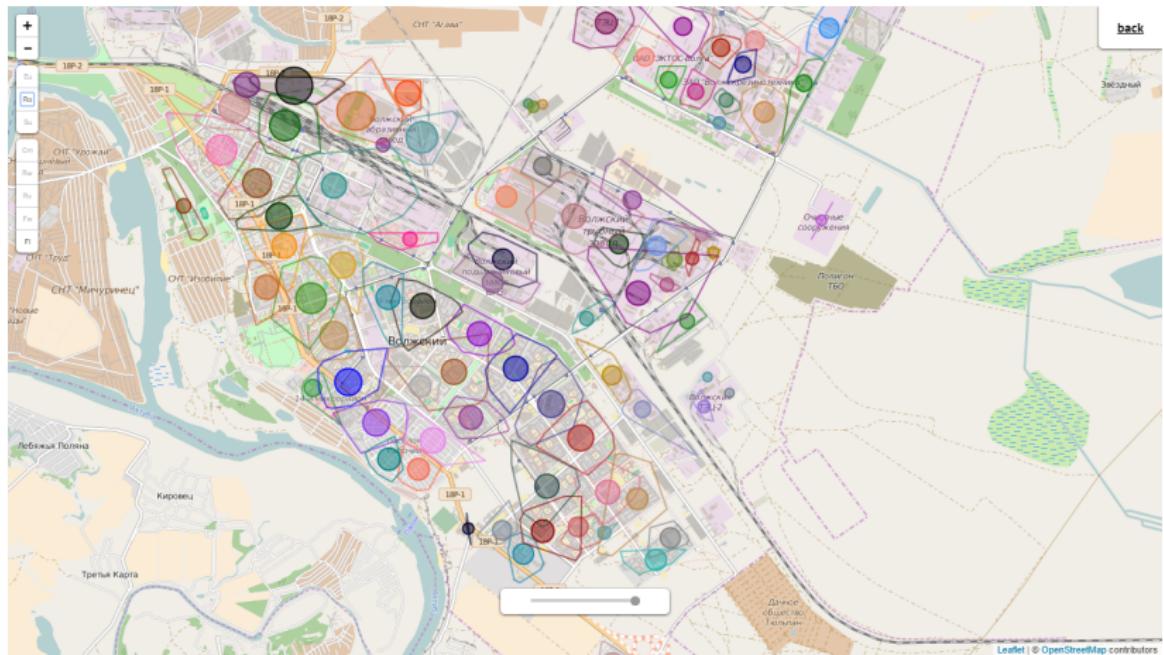
Результат

Евклидова метрика в проекции Меркатора



Результат

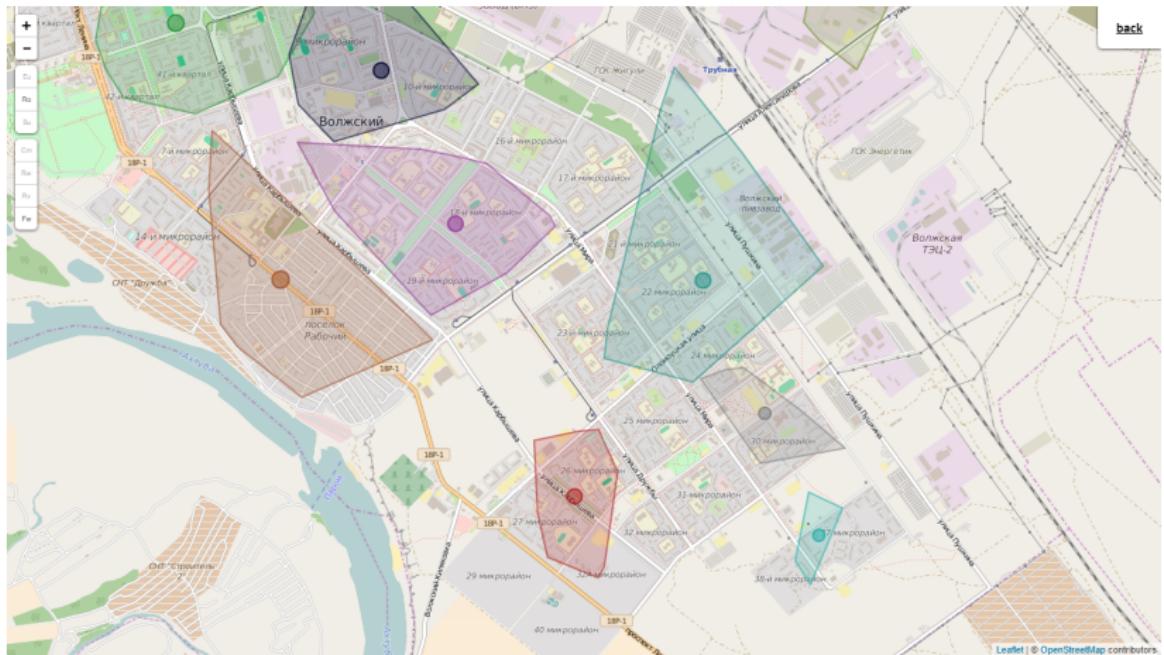
Построение по дорожной сети



<http://vstu-cad-stuff.github.io/clustering/kmeans/>

Результат

Сдвиг центров кластеров к участкам дорожной сети



<http://vstu-cad-stuff.github.io/clustering/terminals/>

Основные результаты работы

В ходе проведения научной работы были получены следующие результаты:

- предложен алгоритм формирования сети остановочных пунктов на основе кластеризации данных о предпочтениях жителей;
- изучены алгоритмы, используемые для кластеризации распределенных данных;
- реализована метрика расстояний, основанная на расчете расстояния по графу дорог;
- реализован алгоритм формирования сети остановочных пунктов на основе кластеризации данных о предпочтениях жителей;
- проверена эффективность реализованного алгоритма на сгенерированных наборах данных.

Публикации

-  Strategway: web solutions for building public transportation routes using big geodata analysis / Golubev A., Chechetkin I., Solnushkin K.S., Sadovnikova N., Paraygin D., Shcherbakov M., Brebels A. // Proceedings of The 17th International Conference on Information Integration and Web-based Applications & Services (iiWAS2015) (December 11 - 13, 2015 Brussels, Belgium) ACM New York, New York pp. 665 - 668
-  Комплекс инструментов интеллектуального анализа данных strategway для поддержки принятия решений по управлению развитием инфраструктуры города / Садовникова Н.П., Щербаков М.В., Парыгин Д.С., Солнушкин К.С., Голубев А.В., Чечеткин И.А. // В сборнике: Развитие средних городов: замысел, модели, практика Материалы III Международной научно-практической конференции. Волгоград, 2015. С. 147-150
-  Автоматизация поддержки принятия решений по разработке маршрутов общественного транспорта на основе анализа данных о корреспонденциях жителей / М. В. Щербаков, Н. П. Садовникова, Д. С. Парыгин, А. В. Голубев, И. А. Чечеткин // Вестник компьютерных и информационных технологий. – М. : Издательский дом «Спектр», 2016. – В печати.

Вопросы



Чечеткин Илья Александрович,
САПР-2п1, illaech@gmail.com