

Волгоградский Государственный Технический Университет
Факультет электроники и вычислительной техники
Кафедра САПР и ПК

**Метод кластеризации предпочтений жителей
города по перемещению.**

Исполнитель:
Чечеткин И. А.
Руководитель:
Щербаков М. В.

Волгоград 2016

Актуальность. В настоящее время формирование маршрутов в городской среде осуществляется на основе положений, заложенных в городской план развития. Обычно, эта информация достаточно устаревшая и не учитывает предпочтения жителей. На основе данных о предпочтениях жителей требуется разработать эффективный метод кластеризации предпочтений жителей города по перемещению.

Объект исследования – предпочтения жителей города, выраженные в географических координатах.

Предмет исследования – методы кластеризации предпочтений жителей.

Цель работы – разработка метода кластеризации предпочтений жителей для минимизации дискомфорта перемещения в городе.

Теоретические задачи:

- разработка алгоритма кластеризации;
- разработка метода учета географических особенностей местности;
- разработка критериев для оценки качества кластеризации.

Практические задачи:

- генерация исходных данных;
- реализация разработанных алгоритмов и методов;
- построение полученных результатов на карте;
- оценка качества кластеризации.

- **Кластер** – объединение нескольких однородных элементов, которое может рассматриваться как самостоятельная единица, обладающая определенными свойствами;
- **предпочтение** – пара узлов с определенными координатами и идентификатором пользователя;
- **node (узел)** – точка с указанными координатами и тегами;
- **tag (тег)** – пары «ключ – значение»;
- **дискомфорт** – совокупный параметр, определяющий время перемещения из начального узла в конечный;
- **центроид** – центр тяжести фигуры (геометрический центр);
- **метрика** – функция, определяющая расстояние в метрическом пространстве.

- Data Clustering: A Review
<https://www.cs.rutgers.edu/~mlittman/courses/lightai03/jain99data.pdf>
- Алгоритмы кластеризации на службе Data Mining
<http://www.basegroup.ru/library/analysis/clusterization/datamining/>
- Генетические алгоритмы
<http://mathmod.aspu.ru/images/File/ebooks/GAfinal.pdf>
- Кластеризация: алгоритмы k-means и c-means
<http://habrahabr.ru/post/67078/>
- Лекции по алгоритмам кластеризации и многомерного шкалирования
<http://www.ccas.ru/voron/download/Clustering.pdf>
- Обзор алгоритмов кластеризации числовых пространств данных
<http://habrahabr.ru/post/164417/>
- Современные тенденции в кластерном анализе
<http://www.ict.edu.ru/ft/005638/62315e1-st02.pdf>

Псевдокод алгоритма кластеризации:

ВВОД точки, границы, количество_кластеров,
количество_итераций

итерация = 0, пред._центроиды = []

текущ._центроиды = [ГПСЧ(границы)] * количество_кластеров

ДЕЛАТЬ ПОКА итерация < количество_итераций ||

пред._центроиды != текущ._центроиды:

РАССЧИТАТЬ принадлежность_точек(точки,
 текущ._центроиды)

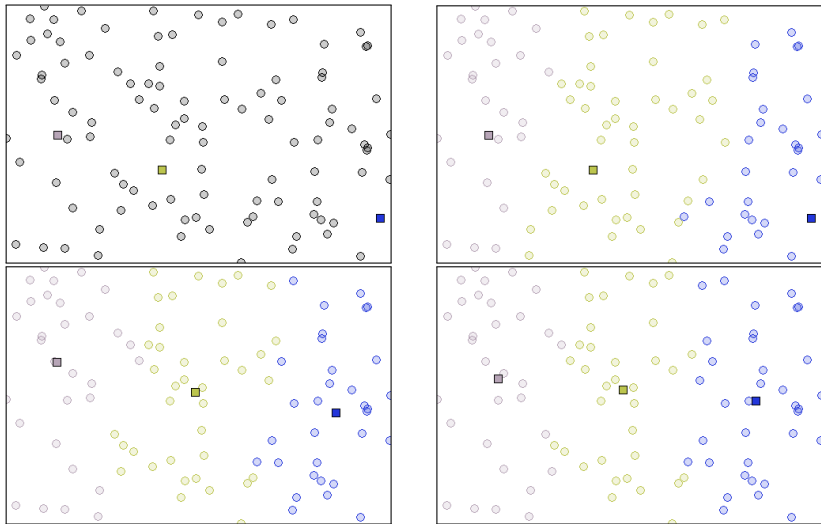
 пред._центроиды = текущ._центроиды

РАССЧИТАТЬ текущ._центроиды(точки, текущ._центроиды)

 итерация += 1

ВЫВОД текущ._центроиды

Прототип



Реализованный алгоритм:

<https://github.com/DSKarramba/clustering>

Результат

