



Машинное обучение и пространственные данные

Дмитрий Стефановский, РГГУ

Что изучать?

Население, дороги, здания, водные системы и открытые пространства в городе совместно создают материальную основу всего города.

К конкретным проявлениям городской жизнеспособности относятся экономическую жизнеспособность, культурную жизнеспособность, качество жизни, социальное управление, качество экологии, и ночная жизнеспособность

Под влиянием человеческого фактора и природных условий эти элементы могут находиться в различных состояниях распределения - агломерации или дисперсии, что приводит к различным городским пространственным структурам.

Источники данных

Показатели качества жизни, выбранные в данной работе, в основном измеряются данными ROI.

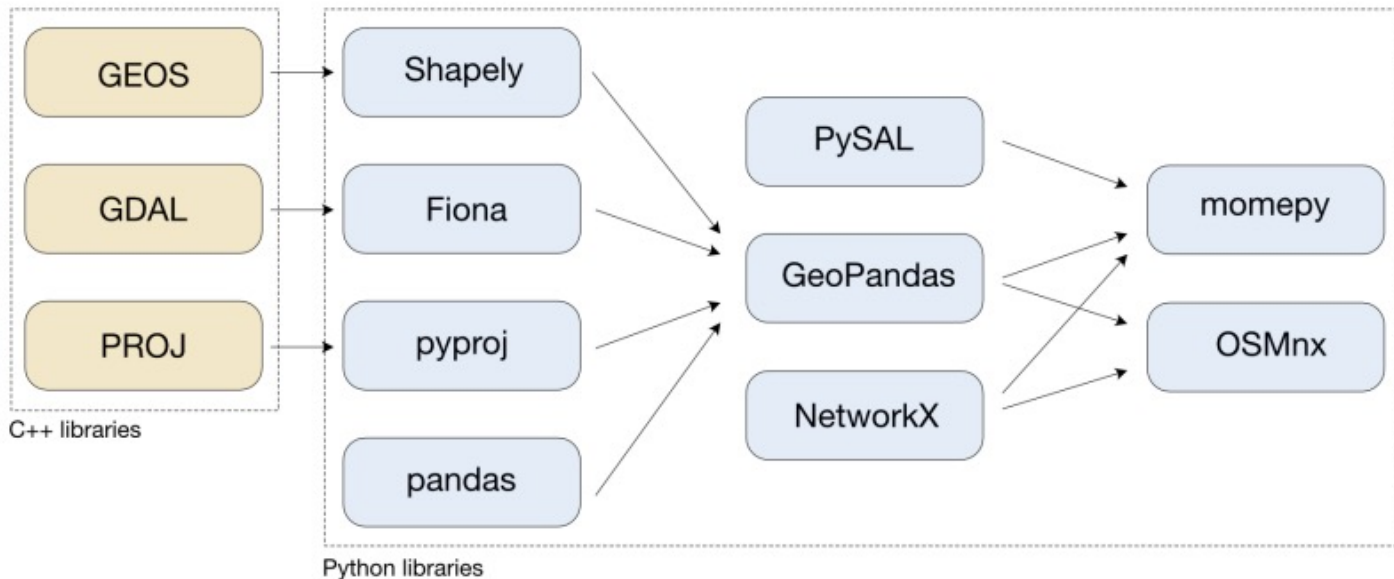
Качество воздуха - это характеристика, отражающая преимущества или недостатки окружающей среды. Получается из государственных сервисов.

Экономическая жизнеспособность включает в себя жизнеспособность потребления, плотность предприятий и цены на жилье. Статистика, экономические исследования

Визуализация

Интегрированные 2D/3D карты:
MapTalks, Openlayers и прочее.

ArcGis, Qgis, Grass, GeoDA - инструменты для
работы с картами



Исследовательские возможности

В изучении пространственной структуры городов и их жизнеспособности есть исследовательские возможности:

Изучение комплексного аспекта жизнеспособности городов;

Рассмотрение факторов пространственной структуры городов не является достаточно полным;

Методы анализа должны учитывать пространственной взаимосвязи между единицами исследования.

Формализация проблемы

Пространственные данные:

Не работают евклидоподобные расстояния.

Трудно использовать статистический анализ.

Один и тот же объект на разных масштабах обладает разными свойствами

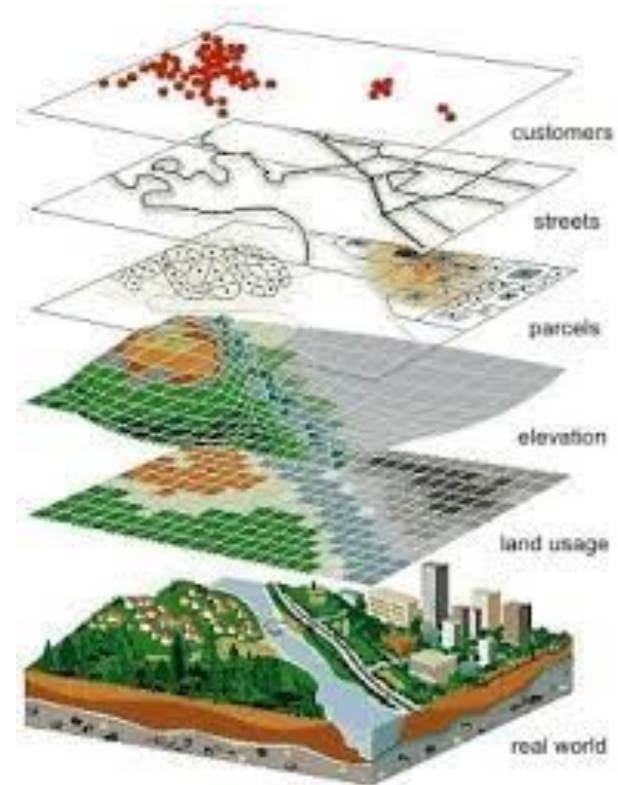
Полностью описать объект невозможно



Формализация проблемы

Объекты с пространственными данными:

- В какую агломерацию входят?
- Какие балансовые модели можно использовать?
- Как сравнивать группы объектов?
- Какую проблему пространственные данные могут показать?
- В чем сильные стороны использования предлагаемого результата обработки пространственных данных?
- В чем уникальность?
- Как мы можем описать пользу от использования результатов обработки данных «в одной фразе»?

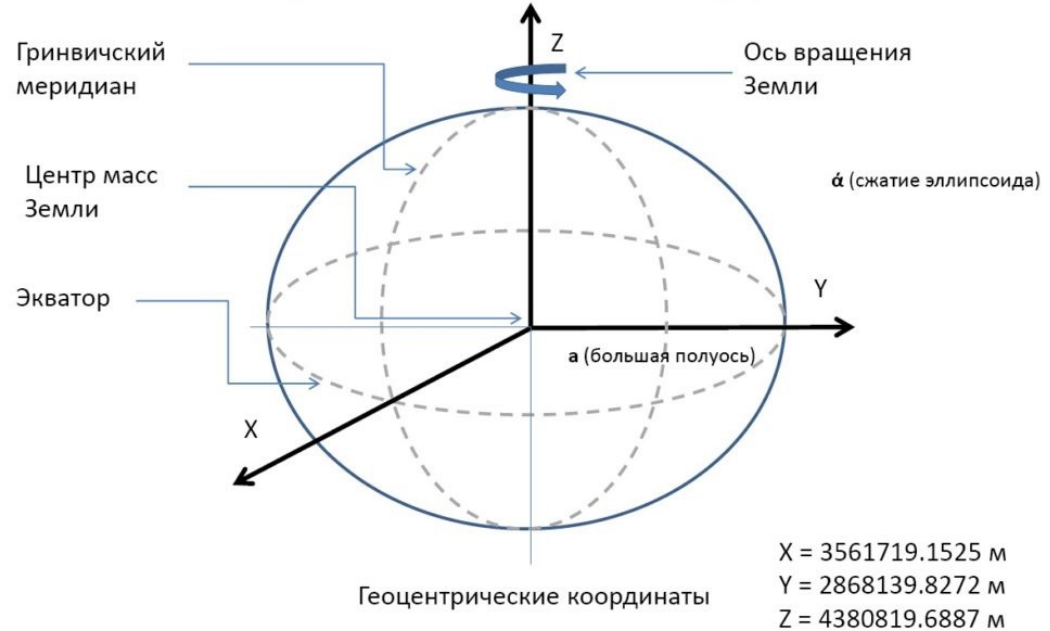


Ось Z — проходящая через точки Зенит и Надир — задает направление на север, для северного полушария.

Ось X — лежит в плоскости Экватора, задает направление на Нулевой меридиан (осевой меридиан зоны в проекции Гаусса — Крюгера).

Ось Y — лежит в плоскости Экватора и дополняет систему до правой.

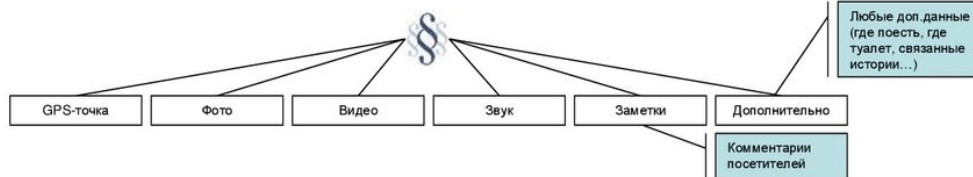
Общеземные системы координат



POI обязательно имеют определённые географические координаты (широту и долготу), которые определяют положение данной точки на карте.



Каждый текст (история) делится на параграфы. Для каждого параграфа доступны, но не обязательны:



POI могут содержать дополнительную информацию, например:

название

категория

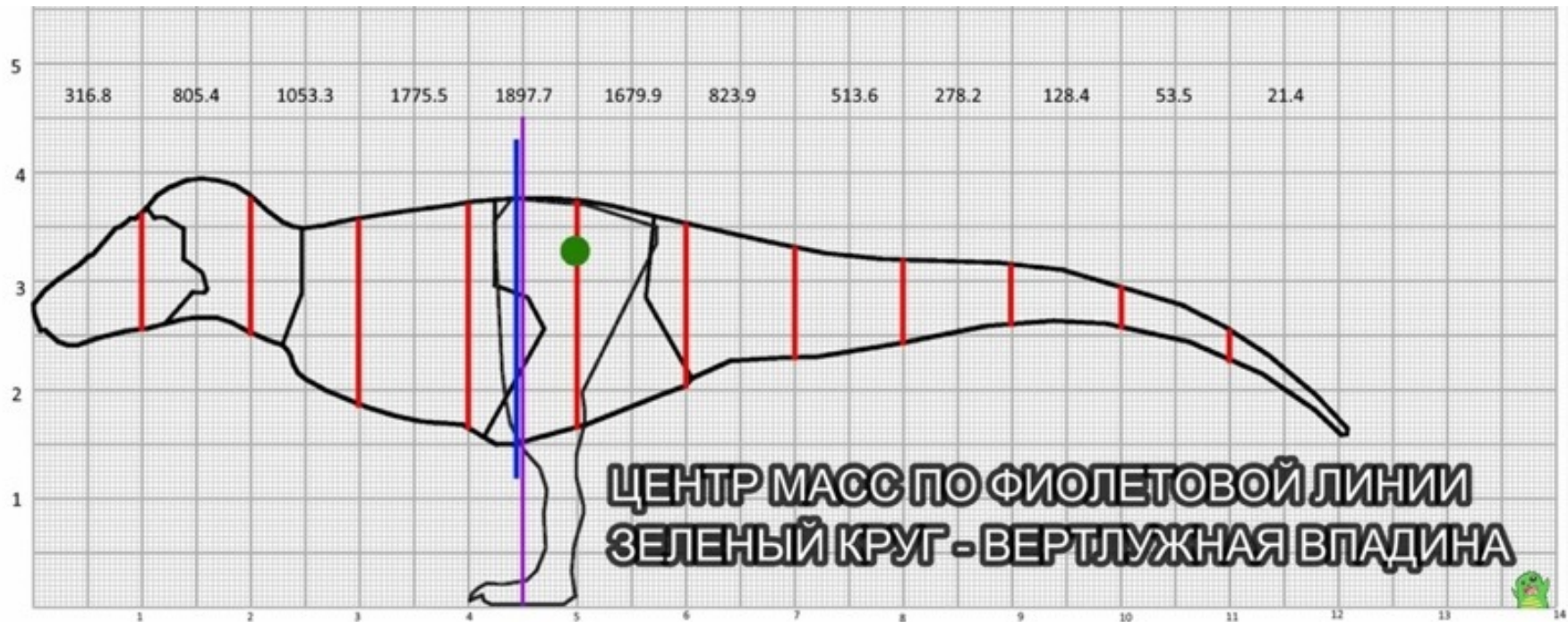
адрес

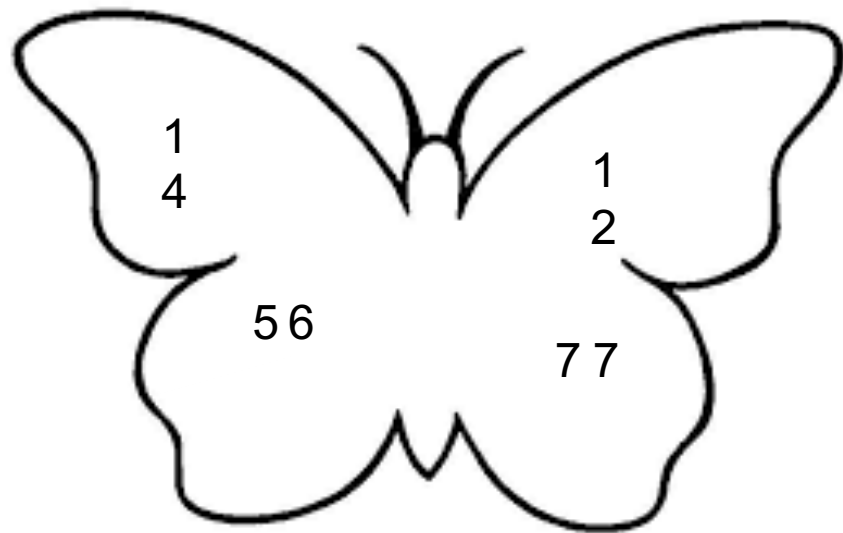
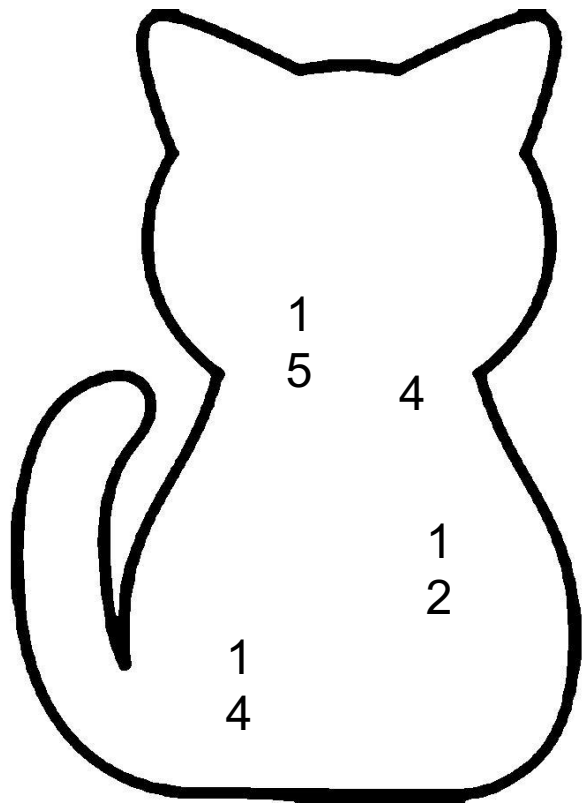
часы работы

сайт

Как правило, POI сгруппированы по типам объектов (например, «проживание», «питание», «культура» и т. д.) и отображаются на карте с помощью соответствующих иконок.

Центр масс точка, через которую должна проходить линия действия силы, чтобы под действием этой силы тело двигалось поступательно (не вращалось).

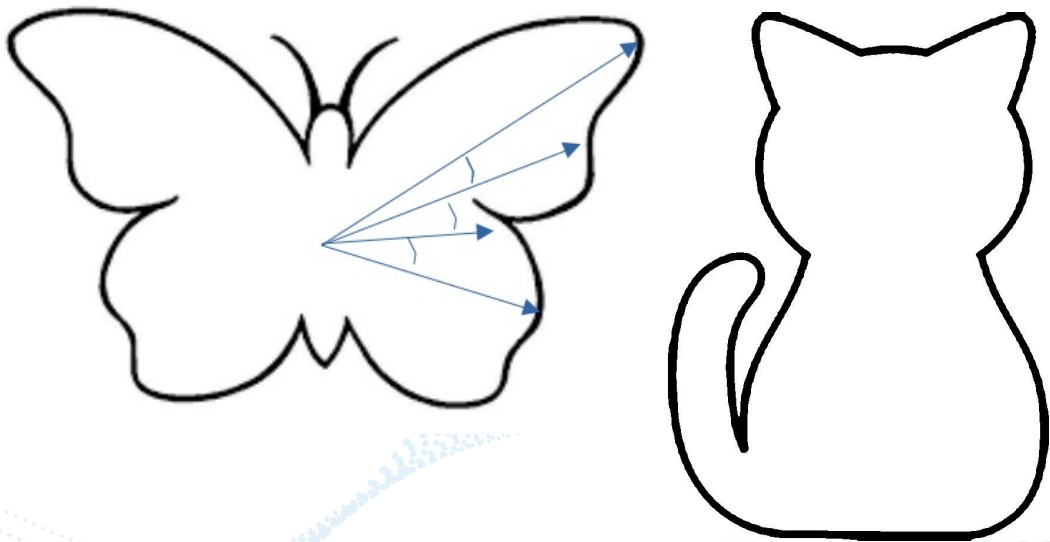




Бабочка: 6,14,12,7

Кошечка 14,15,4,12

Контур как набор чисел



Шаг 1. Находим центр масс контура как набора точек. Формулу не приводим

Шаг 2. Строим через шаг набор векторов с началом в центре масс и окончанием на точке контура. Величина шага определяется угловой мерой

Для каждой фигуры повторяем эти два шага. Получаем два вектора(списка чисел) одинаковой размерности.

Получена таблица для которой можно использовать алгоритмы машинного обучения без учителя: кластеризация, метод главных и метод независимых компонент.

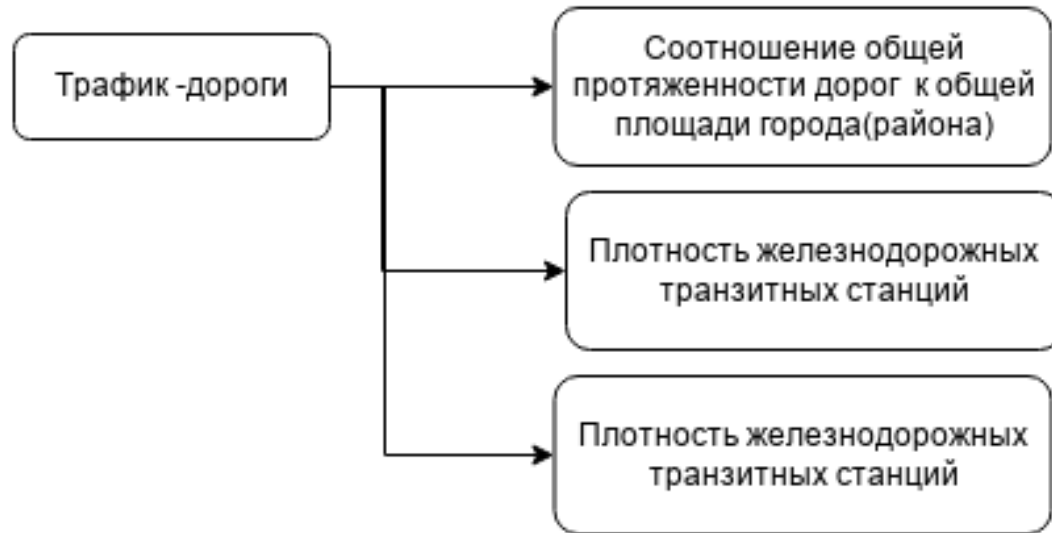
Какие параметры можно изучать



Какие параметры можно изучать



Какие параметры можно изучать



GPS -треки

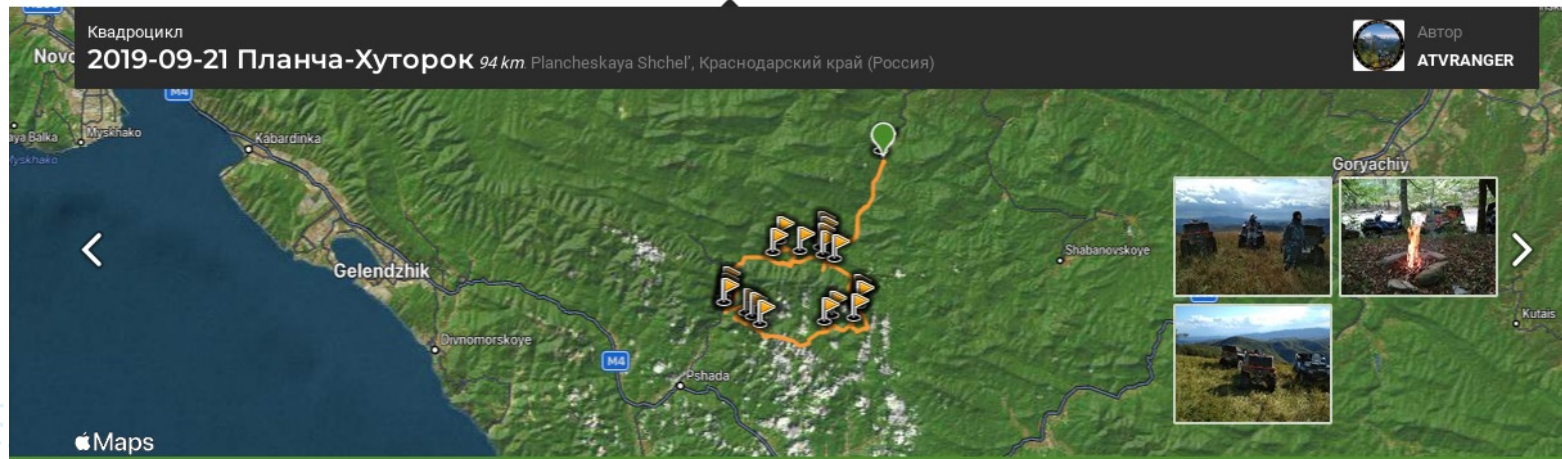


Маршруты мира

Войти

Зарегистрироваться

Wikiloc — это место, где можно открывать лучшие маршруты для пеших походов, велопоходов и многих других видов походов, а также рассказывать о них



Маршрут - это последовательность точек, через которые планируется пройти или прошли.

Что может хранить GPS -трек

Атмосф. давление

Скорость

Восход солнца

Общее время в движении с момента последнего сброса данных.

Общее время без движения с момента последнего сброса данных.

Расчётное время до попадания на следующую маршрутную точку на маршруте.

Время по секундомеру на последнем пройденном круге.

Время в пути

Время до цели

Текущее время суток

Текущее положение

Общее пройденное расстояние с момента последнего сброса данных.

Температура воздуха.

Температура воды.

Точность GPS

Текущая скорость

Текущий уровень заряда батарей.

Мощность спутникового сигнала GPS.

Количество оборотов шатуна педали или шагов в минуту.

Нормирование GPS-треков

Путь есть множество gps треков.

Фактически это GPS-координатов, скорости данные датчиков

Как кластеризовать треки?

Как кластеризовать треки? Разное время разная скорость . Нужен прием позволяющий нормировать

Если добавляете POI данные то возникает описание например экскурсии.

Алгоритм:

Шаг 1. Разбиваем все треки на одинаковые временные промежутки

Шаг 2. Для каждого промежутка определяем описательную статистику по параметрам GPS-трека.

В результате двух шагов все треки оцениваются одинаковым по размерности данными.

Спасибо !