

Геоинформационные сервисы Яндекса

Гольдштейн В. Б.

15 октября 2010 г

- Поиск адресов, магазинов, заправок и т. п.
- Все что угодно на картах
 - фотографии
 - развязки
 - панорамы, камеры
 - магазины, выставки, кинотеатры
- Пробки
 - текущий момент
 - статистические пробки
 - прогноз пробок
- Маршруты
 - без учета пробок
 - без учета прогноза пробок
 - с учетом прогноза пробок
 - маршрут по статистическим пробкам

По разным оценкам до 20 процентов всех поисковых запросов так или иначе имеют отношение к геопоиску.

- "Колдунчик" в поиске
- Магазины на карте в Яндекс.Маркет
- Поиск магазина, а дальше поиск на карте.
- Магазин, аптека, конкретный товар, кинотеатр, офис рядом с метро, домом и т.п.

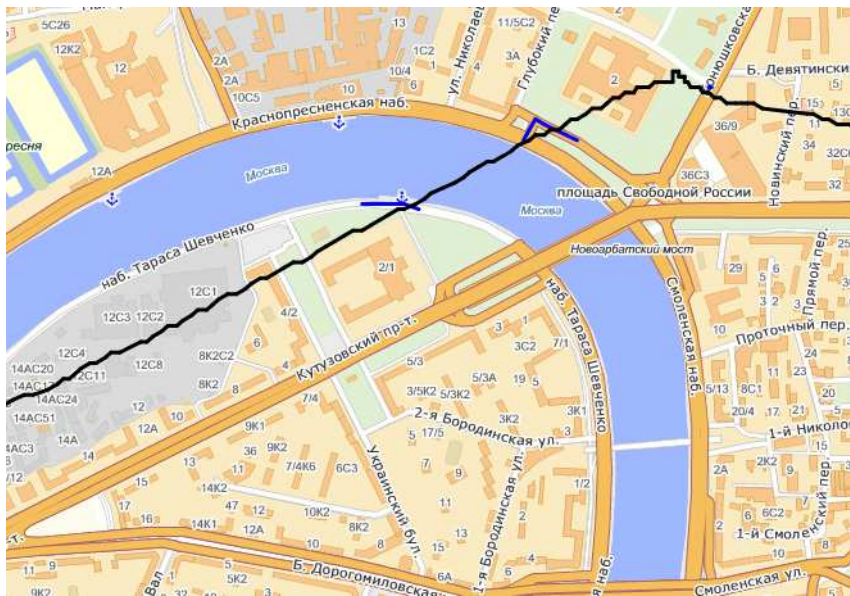
Данные, используемые для пробок

- Данные экспертов
- Партнерские данные о перекрытиях и пробках
- Сигналы от партнеров и пользователей мобильных Яндекс.Пробок

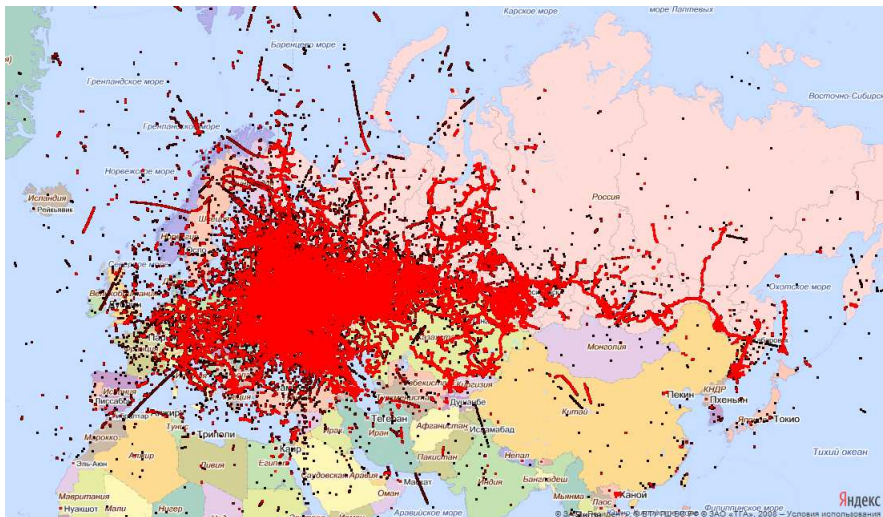
Все данные имеют много ошибок. Люди ошибаются не реже устройств.

- Не известен источник данных
 - от самолетов, вертолетов
 - от лодок
 - медленных машин (грузовики, поливалки, автобусы)
 - от сломанных устройств
- ошибка gpr
- ошибки в графе

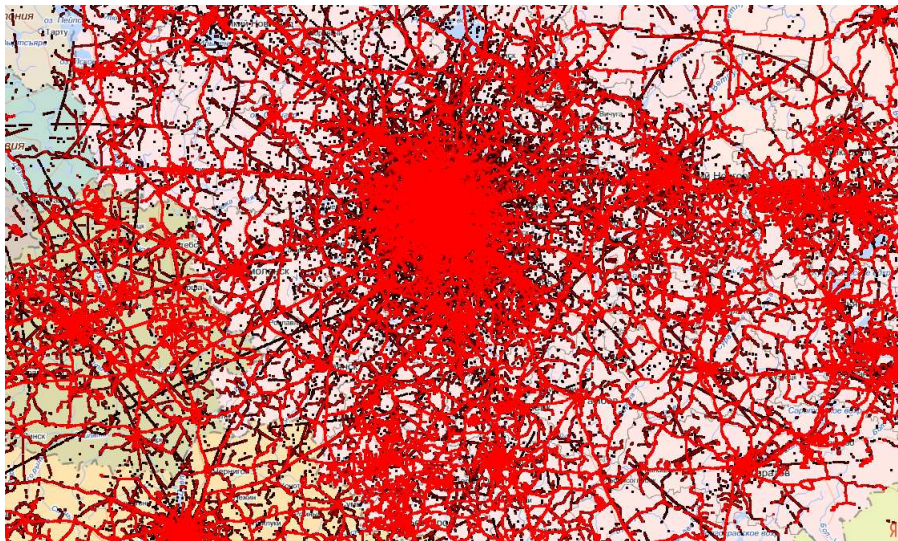
Ошибка gps



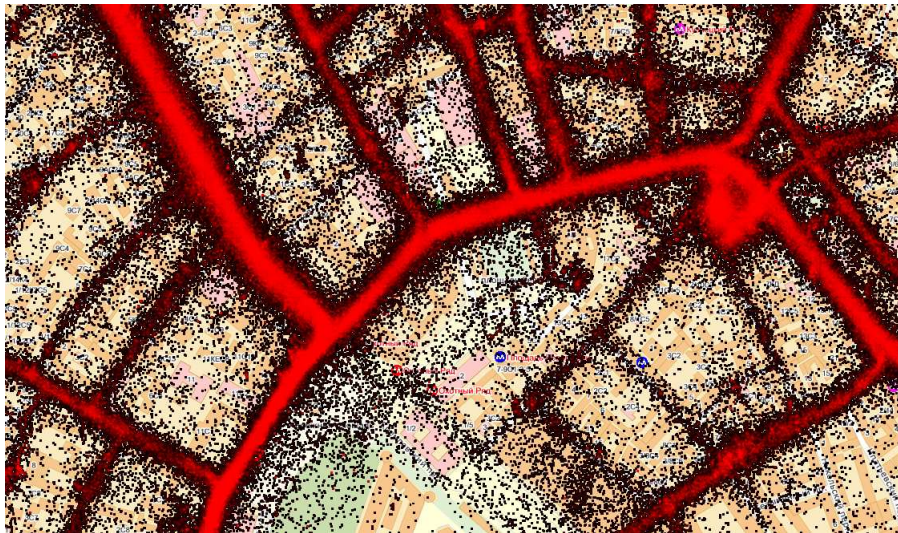
Данные по миру



Данные по Москве

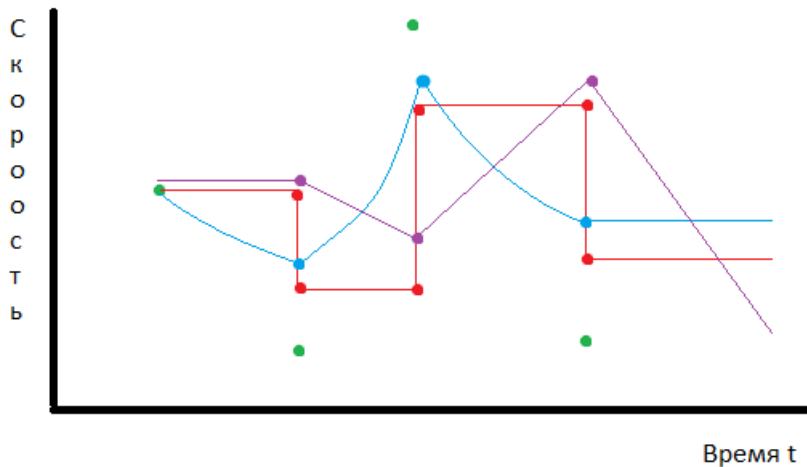


Дорога в Москве



- Хорошо масштабируемая система 60М сигналов в день, в пиковое время около 1000 сигналов в секунду
- Привязка треков к графу.
- Определение скорости потока по точечным данным
- Определение поведения пользователя

Определение скорости потока при точечных данных



- Зеленые точки — информация о замерах скорости
- Красным показана простейшее усреднение
- Фиолетовым – более гладкое усреднение
- Голубым — идеальное усреднение, учитывающее будущее.

Данные

Задана последовательность GPS сигналов (с сопутствующей информацией) одного транспортного средства.

Требование

Найти последовательность участков дорог, по которым двигалось ТС, а так же скорость его движения в каждый момент.

Подход к решению

Для привязки используется метод наибольшего правдоподобия. Для каждого сигнала необходимо определить на каком участке дороги находилось ТС при измерении своего местоположения. В качестве кандидатов для каждого сигнала рассмотрим несколько ближайших участков. Из всех вариантов выберем наиболее правдоподобный.

Перебирать все варианты вычислительно слишком затратно, поэтому при вычислении правдоподобия используется только геометрическое расположение и правдоподобие пути между двумя последовательными сигналами.

Правдоподобие вычисляется из следующих принципов

- Геометрическая вероятность: двумерное нормальное распределение $\frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{dx^2+dy^2}{2}}$
- Правдоподобие кривизны пути $\frac{length}{distance}$
- Правдоподобие скорости движения по пути $\frac{length/speed}{dtime}$ и $\frac{dtime}{driving_time}$
- Правдоподобие количества маневров $\frac{cnt}{dtime}$
- Правдоподобие траектории пути, основанная на расстоянии Фреше.

Расстояние Фреше — это расстояние между двумя кривыми. Пусть A и B — параметризованные кривые, а α и β — их параметризации.

$$F(A, B) = \inf_{\alpha, \beta} \max_{t \in [0, 1]} \left\{ d(A(\alpha(t)), B(\beta(t))) \right\}$$

Расстояние Фреше — минимальная длина поводка, который необходим, чтобы хозяин и собака прошли по соответствующим кривым от начала и до конца.

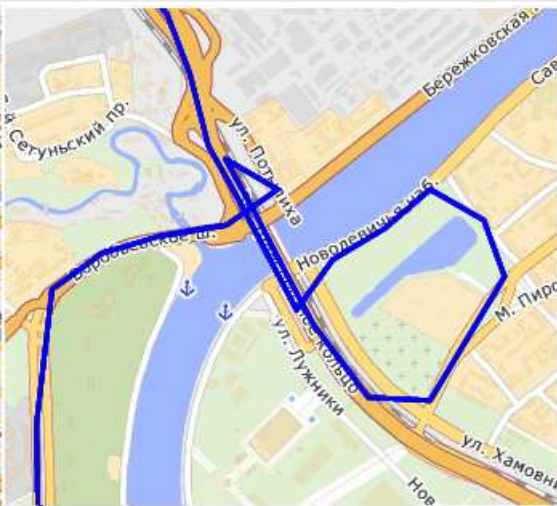
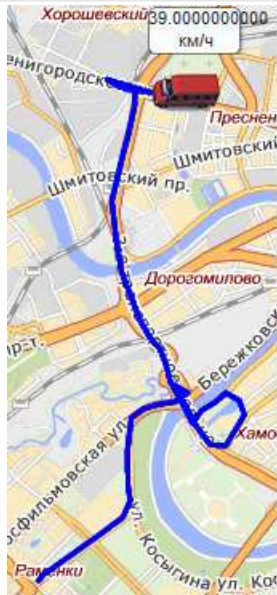
Для каждого из правдоподобий приведена простейшая формула, которую можно усложнять. Формулу необходимо подбирать, подставляя различные коэффициенты.

Вообще говоря есть следующие параметры:

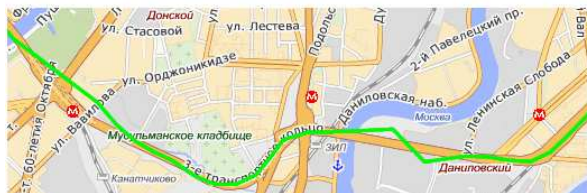
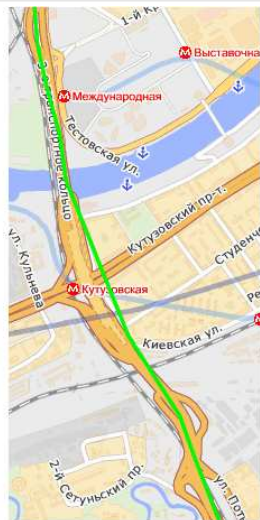
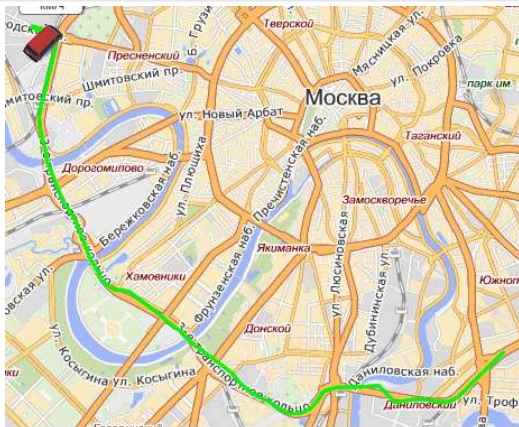
- dx , dy — отклонение от дороги
- длина пути
- расстояние между сигналами
- скорость в сигналах
- разрешенная и средняя скорость на дорогах
- количество маневров
- расстояние Фреше
- возможно многое другое

По этим параметрам необходимо подобрать оптимальную формулу для правдоподобия.

Трек 1



Трек 2



- Построение карт дорог
- Исправление ошибок в картах
- Детектирование и исправление ошибок в запрещенных маневрах.
- Моделирование транспортных потоков.
- Нахождение предпочитаемых маршрутов.

Задача о предсказании пробок на небольшой промежуток времени вперед <http://imat2010.yandex.ru/>

Данные

- Граф дорог. Перекрестки Москвы соответствуют вершинам графа, а отрезки улиц — дугам
- Данные о пробках (загруженности дорог) Наблюдения охватывают 31 день. Для первых 30 дней в файле содержится информация о скорости движения потока автотранспорта с 16:00 до 22:00; для последнего дня — с 16:00 до 18:00.

Оценка

$$Q = \frac{1}{n} \sum k_t k_l |v^* - v|$$

, где

- n — общее количество предсказаний,
- k_l — "коэффициент длины": длина улицы, отнесенная к средней длине улиц (120 м),
- k_t — "временной коэффициент": $1 + 0.1 \cdot \text{порядковый номер четырехминутного интервала, считая от 18:00}$ (например, для 18:56 $k_t = 1 + 0.1 \cdot 14 = 2.4$),
- v^* — наблюдаемая скорость,
- v — предсказанная скорость.

Данные

Данные о треках транспортных средствах за месяц в районе Москвы Строгино.

- Координата GPS (долгота и широта)
- Направление движения (азимут)
- *Мгновенная скорость*
- *Точное время получения данных*
- *Идентификатор трека транспортного средства*

Требования

Требуется восстановить геометрическое расположение автомобильных дорог. Направление движение, количество полос и т. п. в данной задаче восстанавливать не требуется. Результатом должен быть набор ломанных на земном шаре, образующих сеть дорог.

Оценка

Оценка будет проходить с использованием эталонной карты дорог (абсолютная точность которой также не гарантируется) по следующему алгоритму:

- Каждый отрезок дорог обеих карт будет разбит на кусочки не более 10 метров (он будет делиться пополам, пока не станет меньше 10 метров).
- Центры каждого из получившихся отрезков будут выбраны как опорные точки.
- Найдем паросочетание (соответствие) между опорными точками эталонного и проверяемого графа. Погрешность — суммарное расстояние между соответствующими точками, а так же "штраф" 500 метров за вершины без пары.
- Паросочетание будет находиться жадным алгоритмом — на каждом шаге в пару будут объединяться две ближайшие точки.

Особенности при построении карт дорог

- Конкретное устройство может давать систематическую ошибку
- В конкретном месте может быть систематическая ошибка
- Ошибки gpr могут создавать иллюзию несуществующей дороги.
- Сложность отличить мнимую дорогу от редко используемой.
- Большой разброс рядом с тунелями и на развязках

Вопросы?

Гольдштейн Виталий, Яндекс.Пробки
mailto: goldvitaly@yandex-team.ru