

УДК 004.62

СРАВНЕНИЕ МЕТОДОВ ИЗВЛЕЧЕНИЯ ОСЕВЫХ ТРАЕКТОРИЙ В ЗАДАЧЕ ОБНОВЛЕНИЯ КАРТ



Н.С. Иванин
Магистрант БГУИР

Инженер-программист ИООО «ЭПАМ СИСТЕМЗ», Республика Беларусь
E-mail: nikivnik@gmail.com

Н.С. Иванин

Окончил Белорусский государственный университет информатики и радиоэлектроники в 2017 году. Магистрант БГУИР. Работает в EPAM Systems в должности инженера-программиста. Проводит научные исследования в области геоинформатики.

Аннотация. В статье рассматривается такой этап решения проблемы обновления карт, как выделение осевой траектории. Обновление карт – это процесс изменения карт путем нанесения новых и удаления уже не существующих объектов, таких как дороги, здания, природные объекты. Качественное выделение осевых траекторий позволит создавать точные системы для автоматического обновления карт и поддерживать карты в актуальном состоянии. В работе сравнивается эффективность методов для выделения осевой траектории, основанных на principal curve, b-spline и kernel density estimation. Эти методы сравниваются с помощью метрик точности, полноты и f-score в зависимости от различного количества входных данных.

Ключевые слова: обновление карт, GPS, транспортные сети, principal curve, сплайн, kernel density estimation

Проблема обновления карт и, в частности, дорожной структуры, является весьма актуальной для различных областей, таких как навигация, управления дорожной инфраструктурой, принятия решений для служб спасения и скорой помощи, мониторинга ситуации о пробках. Система автоматического обновления карт может быть отнесена к классу интеллектуальных транспортных систем, целью которых является автоматизация управления потоком транспорта и предоставление конечным пользователям достоверной информации о дорожной ситуации, а также создание условий для организации более безопасного и удобного использования всей дорожной сети.

Далее дадим определения основных понятий, встречающихся при решении задачи обновления карт.

В процессе обновления карт на карту наносятся новые объекты и изменяются уже существующие. Для нахождения объектов, которые должны будут быть нанесены на карту или изменены применяются различные методы исследования картографической информации. Среди таких методов можно выделить два основных направления[1]:

1. с помощью спутниковых снимков. У этого направления можно выделить 3 основных поднаправления:

а. методы на основе анализа изображений, которые выделяют дорожные сегменты через анализ определенного набора пикселей и некоторой области вокруг них;

- b. методы на основе делений изображений на области, имеющие одинаковые свойства, с последующим извлечением дорожной информации;
 - c. методы, основанные на комбинировании различных источников данных;
2. с помощью анализа GPS данных.

В данной статье будет рассмотрен подход для обновления карт на основе анализа GPS данных. Создание новых дорожных сегментов производится путем генерации отсутствующих путей в дорожной сети на основе коллекции записанных различными устройствами GPS траекторий. Обновление существующих дорожных сетей состоит из процессов нахождения неправильных дорожных сегментов, обновление их геометрии и добавление дополнительных атрибутов.

Таким образом задача обновления карты может быть сформулирована следующим образом: пусть дана дорожная сеть $G(V, E)$ и множество GPS траекторий T , для каждой из которых поставлен в соответствие некоторый путь в дорожной сети. Необходимо обновить дорожную сеть $G(V, E)$, удалив из нее устаревшие дуги и добавив новые.

При решении задачи нахождения осевой траектории на основе коллекции GPS траекторий T необходимо найти кривую, представляющую собой усредненный путь в дорожной сети $G(V, E)$, по которому двигались транспортные средства. Особую сложность представляет тот факт, что точки GPS трека имеют ошибки измерения, которые не позволяют в точности установить геометрию пути.

Для нахождения осевой траектории авторы[2] используют алгоритм polygonal principal curve (PPC), описанный в[3]. PPC – это гладкая кривая, проходящая через «сердину» набора точек в пространстве, являющихся представлением некоторых данных и служащая нелинейной аппроксимацией этих данных. Эти кривые строятся таким образом, что минимизируют ожидаемую квадратичную дистанцию между кривой и точками пространства, которые выбираются случайным образом по заданному распределению. Построенная кривая является ломанной, состоящей из k сегментов, суммарная длина которых равна заданному числу l , причем эта ломанная минимизирует среднюю квадратичную дистанцию до точек.

В работе[4] для нахождения осевых траекторий предлагается использовать метод Kernel density estimation (KDE).

Метод KDE – это метод непараметрической статистики, применяемый для оценки функции плотности вероятности для непрерывной случайной величины. Ядром в этом методе называется специальный тип функции плотности распределения, являющейся неотрицательной, вещественной, четной функцией. Примерами ядер могут служить равномерное и нормальное распределения.

Авторы[4] предлагают разделять интересующую область на квадраты размером 1×1 метр. Затем для каждой полученной ячейки рассчитывается сколько раз GPS треки ее пересекали, в результате чего получается двумерная гистограмма. Полученная гистограмма рассматривается как двумерное черно-белое изображение. К этому изображению применяется скелетонизация, чей результат принимается в дальнейшем в качестве осевой траектории.

Поскольку при измерении координат с помощью GPS появляются ошибки измерения, имеющие распределение близкое к нормальному, то авторы[4] предлагают использовать ядро нормального распределения $N(0, \sigma^2)$. Параметр σ выбирается исходя из ожидаемой величины ошибки GPS, а также ширины дороги.

В работе[5] для нахождения осевой траектории используются b-spline. B-spline – это кривая, являющаяся линейной комбинацией контрольных точек p_i и базисных функций $N_{i,k}(t)$. Параметр k определяет порядок кривой (линейная, квадратичная, кубическая).

Авторы[5] отмечают, что для достижения наилучших результатов при получении осевых траекторий следует использовать сплайн степени минимум три. Веса точек определя-

ются обратно пропорционально величине ошибки измерения координат, за которую принимается стандартное отклонение GPS приёмника, поскольку точки с меньшим значением ошибки измерений должны вносить больший вклад в форму кривой.

Для каждой точки p_i необходимо вычислить параметр u_i . Вектор параметров влияет на форму и параметры сплайна. В случае аппроксимации единственной GPS траектории предлагается использовать метод длины хорды. Этот метод позволяет задавать точкам веса пропорционально длине соединяющей их дуги.

Для случая нескольких треков в работе [5] предлагается использовать сегмент дороги, полученный из карты, в качестве начальной аппроксимации осевой траектории. Если такой сегмент отсутствует на карте, то в качестве начальной аппроксимации выбирается одна из траекторий

При решении задач обновления карт и нахождения осевых траекторий необходимо большое количество GPS треков. Треки могут быть записаны самостоятельно, что может потребовать большого количества времени и ресурсов. В настоящее время существуют веб-ресурсы, на которых разные люди делятся своими записанными GPS треками. Примерами таких ресурсов служат strava, runtastic, endomondo, gpsies. С перечисленных ресурсов в качестве тестовых данных были получены 100 записанных в г. Минске GPS треков.

Для проверки результатов из дорожной сети были удалены определенные дуги, после чего был запущен алгоритм поиска осевой траектории. Сегменты полученной осевой траектории сравнивались с удаленными дугами дорожной сети. Мерой схожести сегментов выступает близость местоположений и углов векторов, представляющих направление сегмента. Близость местоположение определяется с помощью расчета расстояния Хаусдорфа. Расстояние Хаусдорфа вычисляется как максимальное значение из минимальных расстояний от точек одной кривой до точек другой кривой.

В работе [6] для оценки качества работы алгоритма для выделения осевых траекторий предлагается использовать метрики precision, recall, f-score. Метрика precision вычисляется по формуле 1. Метрика полноты recall вычисляется по формуле 2. Для оценки работы системы по метрикам precision и recall используется метрика f-score, большее значение которой означает лучшую работу системы по precision и recall. Она вычисляется по формуле 3.

$$Precision = \frac{matched}{matched+spurious} \quad (1)$$

где *matched* – количество правильно построенных сегментов,
spurious – количество неправильно построенных сегментов.

$$Recall = \frac{matched}{matched+empty} \quad (2)$$

где *empty* – количество сегментов карты, которым не соответствует ни один сегмент осевой траектории.

$$F - score = \frac{2*precision*recall}{precision+recall} \quad (3)$$

Далее на рисунках 1,2,3 приведены графики зависимости метрик от количества обработанных треков для каждого из описанных подходов.

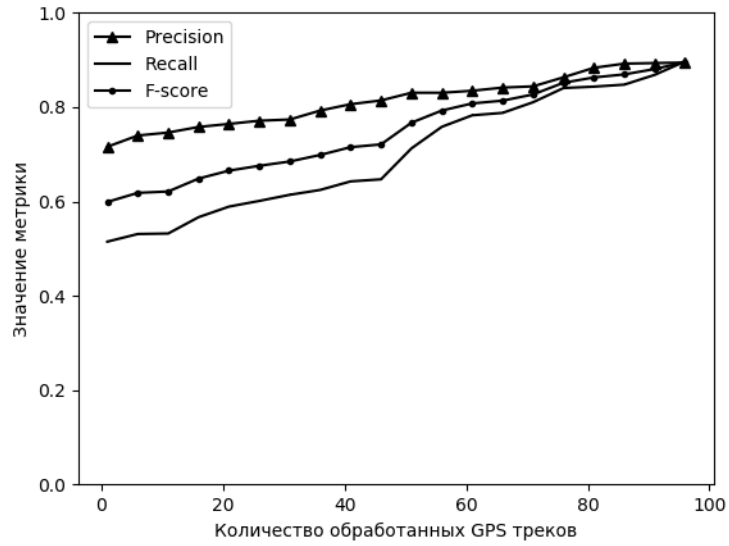


Рисунок 1. Значение метрик для b-spline

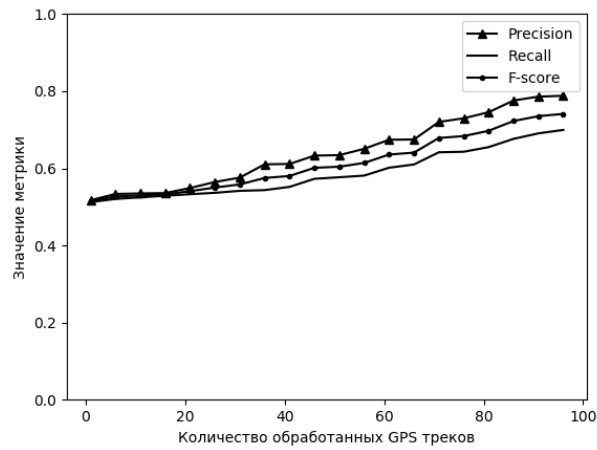


Рисунок 2. Значение метрик для KDE

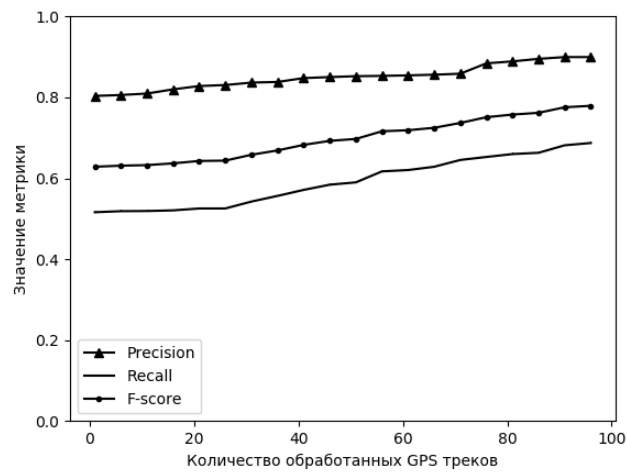


Рисунок 3. Значение метрик для PPC

Как видно из графиков точность работы всех алгоритмов возрастает с ростом количества обработанных GPS треков. Самым чувствительным к количеству данных является метод KDE, он показывает слабые результаты на малом количестве GPS треков. Метод на основе использования PPC продемонстрировал слабые результаты по метрике recall, что может быть связано с тем, что сегменты полученных осевых траекторий имели большую длину, зачастую превосходящую длину сегментов дорог на карте. Метод на основе b-spline продемонстрировал хорошие результаты, на что повлиял удачный выбор весов точек. Все рассмотренные алгоритмы демонстрируют рост точности работы с ростом входных данных, однако имеют сложности на малом количестве данных, что может вызывать сложности в регионах с малым покрытием GPS треками. Рост качества работы алгоритмов для нахождения осевых траекторий может быть осуществлен за счет создания эффективных методов фильтрации «зашумленных» точек.

Литература

- [1]. Wu, T. Updating road networks by local renewal from GPS trajectories / T. Wu, L. Xiang, J. Gong // ISPRS International Journal of Geo-Information, vol. 5, no. 9. – 2016. – № 5(9). – С. 163.
- [2]. Crowdatlas: self-updating maps for cloud and personal use / Y. Wang [и др.] // In: Proceedings 11th international conference mobile systems, applications and services. – 2013. – С. 27-40.
- [3]. Learning and design of principal curves / B. Kegl [и др.] // IEEE Transactions on Pattern Analysis and Machine Intelligence. – 2000. – № 22(3). – С. 281-297.
- [4]. Biagioni, J. Map inference in the face of noise and disparity / J. Biagioni, J. Eriksson // Proceedings of the 20th International Conference on Advances in Geographic Information Systems. – 2012. – С. 79-88.
- [5]. Mining GPS traces for map refinement / S. Schroedl [и др.] // Data Mining and Knowledge Discovery. – 2004. – № 9(1) – С. 59-87.
- [6]. Yang, W. A Method for Extracting Road Boundary Information from Crowdsourcing Vehicle GPS Trajectories/ W. Yang, T. Ai, W. Lu // Sensors(Basel). – 2018. – № 18(4). – С. 1261.

COMARISON OF CENTERLINE EXTRACTION METHODS IN THE MAP UPDATE PROBLEM

N.S. IVANIN
Master student at BSUIR

Software engineer at EPAM Systems, Republic of Belarus
E-mail: nikivnik@gmail.com

Abstract. The article considers centerline extraction step of the map update problem. Map update is a process of map changing through adding new and removal of non-existent objects, such as roads, buildings, natural objects. Qualitative methods of centerline extraction will lead to creation of automatic map update systems with high accuracy and keeping maps up to date. The paper compares efficiency of methods based on principal curve, b-spline and kernel density estimation. These methods are compared using precision, recall and f-score metrics for different amount of input data.

Keywords: map update, GPS, transport networks, principal curve, spline, kernel density estimation