

Министерство образования Республики Беларусь
Учреждение образования
«Белорусский государственный университет
информатики и радиоэлектроники»

УДК 004.62

Иванин
Никита Сергеевич

**ОБРАБОТКА ДАННЫХ GPS С ПОМОЩЬЮ СКРЫТЫХ
МАРКОВСКИХ МОДЕЛЕЙ**

АВТОРЕФЕРАТ

магистерской диссертации на соискание степени
магистра информатики и вычислительной техники

по специальности 1-40 81 04 – Обработка больших объемов информации

Научный руководитель
Волорова Н.А.
Кандидат технических наук,
доцент

Минск 2019

ВВЕДЕНИЕ

В настоящее время практически вся картографическая информация находится в цифровом виде, что значительно упрощает процесс ее изменения. Эта причина послужила толчком к развитию систем автоматического изменения карт. Основными задачами, решаемые подобными системами являются задачи сопоставления и обновления карт.

Сопоставление карт – это процесс определения дороги на карте, по которой двигалось транспортное средство на основе записанного GPS трека. Дороги на большинстве современных карт представлены в виде ломанных(сегментов). GPS трек также представляется в виде ломанной, точки которой являются координатами на местности. Координаты ломанной получаются специальным устройством – GPS приёмником, который запрашивает координаты через определенные интервалы времени, связываясь с GPS спутником.

Результатом сопоставления карт является последовательность дорожных сегментов реальной карты, по которым двигалось транспортное средство. Наиболее хорошо зарекомендовавшим себя подходом для нахождения такой последовательности сегментов является применение скрытых марковских моделей. В этом случае задача сопоставления карт решается как задача поиска скрытых состояний СММ.

Результаты сопоставления карт могут быть использованы затем для обновления карт. Для этого анализируются последовательности точек, не соотнесённые с дорожными сегментами на карте, а также параметры скрытой марковской модели после обработки GPS трека.

Обновление карт – это процесс изменения карт путем нанесения новых и удаления уже не существующих объектов, таких как дороги, здания, природные объекты. В магистерской диссертации решалась задача удаления и добавление дорог, таких как автомобильные, пешеходные, велосипедные дороги.

Решение задач сопоставления и обновления карт было бы невозможно без путей движения транспортных средств, записанных с использованием системы GPS. В настоящее время GPS приемники устанавливаются во все смартфоны и любой человек может выступать поставщиком данных о своих траекториях движения.

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы исследования

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. Программа, выполняющая автоматическое обновление 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.

Интеллектуальные транспортные системы активно развиваются в настоящее время и некоторые из них действуют на просторах г.Минска. Так, например, на многих остановках присутствуют информационные табло, сообщающие о времени прибытия следующего трамвая или автобуса. Дорожную ситуацию возможно отследить с помощью различных сервисов, например, Яндекс пробки, пробки ГдеТут.by. Большой потенциал имеет использование радаров для регулировки дорожного трафика, а также использование систем электронной оплаты проезда для изменения количества транспорта на маршруте.

Другим толчком к развитию системы автоматического обновления карт является развитие беспилотных автомобилей. В настоящее время многие производители, исследовательские организации, ИТ-компании занимаются разработкой полностью автономных беспилотных автомобилей. Таким автомобилям необходимы очень точные карты, содержащие все последние изменения дорожной структуры, поскольку неточности могут привести к различным аварийным ситуациям.

Многие цифровые карты имеют плохое покрытие вне автомобильных дорог и крупных населенных пунктов, большинство пешеходных дорог и вовсе не нанесены. Поэтому использование записанных GPS треков для последующего обновления карт является весьма перспективным.

Степень разработанности проблемы

В работе П. Ньюсона и Дж. Крума была впервые рассмотрена возможность применения скрытых марковских моделей для решения задачи сопоставления карт как решение задачи поиска скрытых состояний модели на основе последовательности наблюдений. Теоретические основы решения задачи поиска скрытых состояний были заложены в работе Л. Рабинера.

Среди всех исследований, посвящённых улучшению выбора параметров скрытой марковской модели, необходимо отметить работы Г. Гоха, И. Ванга, Т. Ву, Ж. Шэна. В магистерской диссертации дано дальнейшее развитие применению СММ для сопоставления карт в виде их применения для обработки и сопоставления с картами GPS треков, записанных велосипедистами и пешеходами. Подобные треки зачастую проходили по областям с низкой полнотой и детальностью карт.

Авторами работ, посвящённых кластеризации точек GPS треков, являются И. Ванг, Т. Ву, Ж. Шэн. В магистерской диссертации

усовершенствована кластеризация точек путем применения метода кластеризации DBSCAN и разработки собственного метода поиска расстояния между точками GPS трека.

В работах Кегля, Биаджиони, Шхрэдля рассматриваются различные подходы для нахождения осевой траектории по коллекции точек. В диссертации был использован подход на основе аппроксимации множества точек с помощью b-spline и впервые использован метод KDE для оценки весов точек b-spline.

Цель и задачи исследования

Целью диссертации является разработка программного продукта, позволяющего решать задачи обновления и сопоставления карт на основе GPS треков, записанных пешеходами и велосипедистами для обновления областей карт, имеющих низкую полноту и детальность.

Для выполнения поставленной цели в работе были сформулированы **следующие задачи**:

- разработать модуль для взаимодействия и получения объектов карт для сервиса OpenStreetMaps;
- найти и получить данные GPS треков, записанных пешеходами и велосипедистами для г. Минска;
- найти параметры скрытой марковской модели, позволяющие качественно решать задачу сопоставления карт для GPS треков, записанных пешеходами и велосипедистами;
- построить СММ, одновременно учитывающую специфику траектории движения и расположения существующих объектов на карте;
- верифицировать разработанную модель, используя метрики полноты, точности, f-score;
- выявить признаки, свидетельствующие о существовании неточностей в существующих картах;
- разработать методику автоматического нахождения подмножества тестовых данных, позволяющего произвести анализ найденной неточности;
- провести анализ возможности применения различных методов кластеризации для кластеризации точек GPS траекторий;
- разработать методику расчета расстояний между элементами кластера, учитывающую природу этих элементов;
- провести анализ возможности применения различных методов для нахождения осевой траектории с учетом особенностей источников этих траекторий;
- построить модель, решающую задачу сопоставления карт;
- верифицировать разработанную модель для обновления карт с помощью метрик точности, полноты, f-score.

Объектом исследования является процесс обработки GPS данных.

Предметом работы выступают методы обработки GPS данных, использующиеся для сопоставления и обновления карт.

Область исследования. Содержание диссертационной работы соответствует образовательному стандарту высшего образования второй ступени (магистратуры) специальности 1-40 81 04 «Обработка больших объемов информации».

Теоретическая и методологическая основа исследования

Для получения теоретических результатов в задаче сопоставления карт были применены скрытые марковские модели, чьи параметры рассчитывались на основе GPS треков, записанных в г. Минске. Задача нахождения кластеров GPS точек в задаче обновления карт решалась с помощью применения алгоритма DBSCAN. Задача нахождения осевой траектории была решена с использованием b-spline, чьи коэффициенты были рассчитаны через оценку распределения точек методом kernel density estimation.

Программный продукт был написан на языке python. Данные хранились в базе данных MongoDB. Визуализация графиков была осуществлена с помощью библиотеки matplotlib. Визуализация карт, треков и GPS точек осуществлялась javascript библиотекой leaflet.

Информационная база исследования для решения задач сопоставления и обновления карт сформирована с использованием GPS треков, хранящихся в сервисе strava.

Научная новизна диссертационной работы заключается в разработке и верификации метода расчета расстояния между точками GPS трека, а также применения метода KDE для вычисления весов точек b-spline.

Основные положения, выносимые на защиту

1. Модель рекурсивного парсера для OpenStreetMap XML.
2. Модель для сопоставления карт, построенная с применением скрытых марковских моделей, учитывающая особенности движения велосипедов и пешеходов.
3. Модель для кластеризации точек, основанная на применении алгоритма DBSCAN.
4. Методика расчета расстояния между точками GPS трека, которая учитывает их географическое расстояние, а также угол между элементарными участками траектории, в которые они входят.
5. Модель нахождения осевой траектории на основе множества GPS точек путем аппроксимации методом b-spline, веса точек которого учитывают предполагаемое распределение этих точек.

Теоретическая значимость диссертации заключается в том, что в ней предложен подход к вычислению расстояния между точками GPS траекторий, которая учитывает их географическое расстояние, а также угол между элементарными участками траекторий. Для задачи нахождения осевой траектории был предложен метод на основе использования b-spline, веса точек которого учитывают предполагаемое распределение этих точек.

Практическая значимость диссертации состоит в том, что с помощью

предложенной модели возможно осуществлять автоматической обновление карт для различных источников данных и поставщиков карт.

Апробация и внедрение результатов исследования

Результаты исследования были представлены на 54-й научной конференции аспирантов, магистрантов и студентов, проходившей 23 – 27 апреля 2018 г. в Белорусском государственном университете информатики и радиоэлектроники.

Публикации

Основные положения работы и результаты диссертации изложены в одной опубликованной работе общим объемом 1,5 п.л. (авторский объем 1,5 п.л.).

Структура и объем работы. Структура диссертационной работы обусловлена целью, задачами и логикой исследования. Работа состоит из введения, четырех глав и заключения, библиографического списка и приложения. Общий объем диссертации – 86 страниц. Работа содержит 4 таблицы, 48 формул, 25 рисунков. Библиографический список включает 36 наименований.

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

Во **введении** рассмотрено современное состояние задач сопоставления и обновления карт, приведены предпосылки появления интереса к этим задачам, определены основные направления исследований, а также сделано обоснование актуальности темы диссертационной работы.

В **первой главе** рассматриваются принципы работы системы глобального позиционирования GPS, приводятся сведения о стандарте WGS 84, описывающий механизм измерения координат, осуществляется постановка задач сопоставления и обновления карт.

Во **второй главе** приводится теоретическое описание скрытых марковских моделей, производится анализ основных проблем СММ и алгоритмов их решения.

В **третьей главе** приведен анализ современного состояния и тенденций развития алгоритмов для решения задач сопоставления и обновления карт. Рассматривается вопрос применения скрытых марковских моделей в задаче сопоставления карт, описываются различные подходы для вычисления параметров СММ. Рассматриваются основные этапы решения задачи обновления карт: нахождение новых и ошибочных сегментов, кластеризация точек GPS, нахождение осевой траектории. Для каждой из этих задач осуществляется описание и анализ наиболее хорошо зарекомендовавших алгоритмов ее решения.

В **четвертой главе** представлена описана архитектура реализованного приложения. В ней рассматриваются различные форматы хранения

картографических данных, модуль для работы с картами OpenStreetMap, модули для решения задач сопоставления и обновления карт. Для всех модулей приводится подробное описание и анализ принятых решений по реализации, проверяются качественные и количественные показатели их работы, которые затем представляются в виде графиков.

В **приложении** приведен исходный код полученного программного продукта.

ЗАКЛЮЧЕНИЕ

В рамках магистерской диссертации было разработано приложение для обработки GPS данных с различных устройств. Обработка GPS данных заключалась в решении трех основных задач:

1. сопоставления записанного GPS трека с реальной дорожной сетью. Путь транспортного средства по реальной дорожной сети находился с помощью применения скрытых марковских моделей;

2. нахождение различий между дорожной сетью и коллекцией записанных GPS треков. Было осуществлено обнаружение таких видов различий, как геометрические (изменение дорожной структуры), топологические (при которых происходит изменение соединений между дорогами в местах перекрестков), семантические (изменение различных свойств дорожных сегментов);

3. добавление найденных изменений в оригинальную дорожную сеть.

Все перечисленные задачи были успешно решены. В ходе исследования областей сопоставления и обновления карт на основе коллекции GPS треков были встречены и решены следующие проблемы:

1. слабое покрытие многих участков дорог. Это означает, что популярность (количество записанных треков, проходящих через определенный сегмент) разных дорожных сегментов сильно отличается;

2. слабая доступность картографической информации. Картографические коммерческие сервисы не предоставляют числовую информацию о карте (координаты, типы объектов). В то время как открытые источники информации предоставляют всю необходимую информацию, но зачастую она является устаревшей или неточной;

3. низкая топологическая точность. Большинство разработанных алгоритмов демонстрируют низкую топологическую точность, что происходит из-за наличия шума в GPS данных.

Разработанная библиотека показала удовлетворительные результаты на тестовых данных, являющихся записанными различными людьми GPS треками. Качество обнаруженных неточностей карт проверялось специальными метриками.

Для используемых моделей были найдены значения параметров, позволяющие успешно работать с GPS треками, записанными велосипедистами и пешеходами. Именно эта область представляла наибольший интерес, поскольку участки карты, не посещаемые автомобилями, имеют наименьшее покрытие картами.

В магистерской диссертации были предложены методы по улучшению кластеризации точек GPS трека и выделения осевой траектории с помощью b-spline. Для алгоритма кластеризации DBSCAN была разработана собственная мера расстояния между точками, которая учитывает их географическое расстояние, а также угол между элементарными участками траекторий, в которые они входят. При выделении осевой траектории для расчета весов точек сплайна был применен метод KDE, позволивший получить высокую точность выделенных траекторий.

В данный момент разработанный программный продукт имеет мало прямых аналогов из-за новизны области обработки GPS данных. Толчок к ее развитию был дан широким распространением портативных устройств с GPS приемниками и созданием веб-сервисов, позволяющих любому желающему делиться записанными GPS треками и выступать поставщиком данных. По этим причинам эта область является активно развивающейся и в ней постоянно предлагаются более точные алгоритмы.

В результате цель магистерской диссертации была достигнута. Была исследована область и создана библиотека для решения поставленной задачи. Помимо решения задачи сопоставления карт и обнаружения неточностей в них разработанный программный продукт может быть расширен для решения таких задач, как предложения оптимального маршрута между точками (в том числе по новым, не посещенным до этого дорогам), анализ активности велосипедистов и бегунов на различных участках.

Список опубликованных работ

1. Иванин, Н. С. Применение Скрытых Марковских моделей для сопоставления карт / Н. С. Иванин // Компьютерные системы и сети: материалы 54-й научной конференции аспирантов, магистрантов и студентов, Минск, 23 – 27 апреля 2018 г. / Белорусский государственный университет информатики и радиоэлектроники. – Минск, 2018. – С. 166 – 167.

2. Иванин, Н. С. Реализация шифрования с элементами подписи на основе эллиптических кривых / Н. С. Иванин // Технические средства защиты информации: тезисы докладов XVI Белорусско-российской научно – технической конференции, Минск, 5 июня 2018 г. – Минск: БГУИР, 2018. – С. 42.