

# Распределенная программная система автоматического построения схемы городской транспортной сети

Н.А. Пучкова<sup>1,a</sup>, Н.М. Ершов<sup>1,2,b</sup>

<sup>1</sup> ГБОУ ВО Московской области "Университет "Дубна"  
141980, Московская область, г. Дубна, ул. Университетская, 19

<sup>2</sup> Московский государственный университет им. М. В. Ломоносова, факультет ВМК  
119991, Москва, ГСП-1, Ленинские горы, д. 1, стр. 52, факультет ВМК

E-mail: <sup>a</sup> pu4kova.nata@ya.ru, <sup>b</sup> ershov@cs.msu.ru

В работе рассматриваются вопросы, связанные с построением цифрового описания схем движения городского транспорта на основе геолокационных данных, получаемых с мобильных устройств (телефоны, планшеты). Такое описание может в дальнейшем использоваться либо для компьютерного моделирования движения транспорта, например, с целью оптимизации той или иной дорожной развязки, либо для разработки алгоритмов и программ управления автономными транспортными средствами. Одной из проблем, связанных с использованием геолокационной информации, является ее сильная зашумленность. Поэтому на предварительном этапе исследования был разработан ряд алгоритмов, позволяющих по нескольким трекам, соответствующим одному и тому же участку дороги, провести процедуру удаления шума и выделить так называемые «чистые» траектории движения транспорта по данному участку дороги. Серьезным недостатком такого подхода оказалась необходимость в существенной ручной предобработке входных данных, например, в пространственной синхронизации отдельных треков.

В настоящей работе предлагается следующая схема решения исходной задачи. Разрабатывается программа-геотрекер для мобильных устройств, служащая для сбора ими первичной геолокационной информации в непрерывном режиме. Эта информация автоматически пересылается на некоторое облачное хранилище данных, откуда забирается сервером, выполняющим всю статистическую обработку данных. На первом этапе с использованием методов машинного обучения из треков выделяются части, соответствующие движению на транспортных средствах (автомобилях). На втором этапе производится пространственная синхронизация треков, т.е. разбиение их на фрагменты, относящиеся к одинаковым участкам дорог. На третьем этапе выполняется выделение «чистых» траекторий движения транспорта и связывание их в общую систему. Данная работа посвящена разработке алгоритмов решения задач, возникающих на первом и третьем этапах описанной схемы. Приводятся результаты численных исследований, основанных на анализе реальных данных движения транспорта в городе Дубна Московской области.

Ключевые слова: транспортные модели, распределенные системы, геолокация.

Работа выполнена при финансовой поддержке РФФИ (грант № 14-07-00628 А).

© 2016 Наталья Александровна Пучкова, Николай Михайлович Ершов

## Введение

В работе рассматриваются вопросы, связанные с построением цифрового описания схем движения городского транспорта на основе геолокационных данных. Такое описание может в дальнейшем использоваться либо для компьютерного моделирования движения транспорта, например, с целью оптимизации той или иной дорожной развязки, либо для разработки алгоритмов и программ управления автономными транспортными средствами.

Одной из проблем при использовании геолокационной информации является ее сильная зашумленность. Поэтому на первом этапе исследования были разработаны алгоритмы, позволяющие по нескольким трекам, соответствующим одному и тому же участку дороги, провести процедуру удаления шума и выделить «чистые» траектории движения транспорта по данному участку дороги. Серьезным недостатком такого подхода оказалась необходимость в существенной ручной предобработке входных данных.

Предлагается следующая схема решения исходной задачи. Разрабатывается программа-геотрекер для мобильных устройств, служащая для сбора ими первичной геолокационной информации в непрерывном режиме. Эта информация автоматически пересылается на некоторое облачное хранилище данных, откуда забирается сервером, выполняющим всю статистическую обработку данных. На первом этапе с использованием методов машинного обучения из треков выделяются части, соответствующие движению на транспортных средствах (автомобилях). Далее производится разбиение этих треков на фрагменты, относящиеся к одинаковым участкам дорог. На третьем этапе выполняется выделение «чистых» траекторий движения транспорта и связывание их в общую систему.

Настоящая работа посвящена разработке алгоритмов решения задач, возникающих на первом и третьем этапах вышеописанной схемы. Приводятся результаты численных исследований на основе реальных данных движения транспорта в городе Дубна.

## Тестирование существующих решений

Для создания схемы сети дорог нам необходимы координаты этих дорог. В связи с этим было решено использовать мобильный телефон (под управлением ОС Android) с поддержкой GPS и установленным приложением для записи трека.

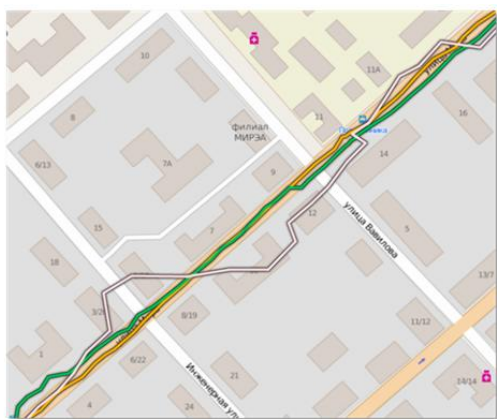


Рис. 1. Примеры зашумленных треков



Рис. 2. Трек с разрывами в записи

Для тестирования существующих решений на Google Play Market было выбрано семь приложений, из которых четыре было отброшено сразу в силу низкого качества выдаваемых ими данных. Оставшиеся три приложения были протестированы на записи реальных траекторий

(треков) движения легкового и общественного транспорта в городе Дубна Московской области. Результаты тестирования выявили две основные проблемы, связанные с полученными геолокационными данными: сильная зашумленность этих данных (рисунок 1) и разрывы в записи (например, из-за потери связи со спутниками, рисунок 2). Поэтому первая поставленная нами задача была связана с решением этих двух указанных проблем – имея серию зашумленных треков с возможными разрывами в записи выделить из них ту реальную траекторию, по которой происходит движение транспорта.

## Автоматическое усреднение треков

Предполагается, что имеются записи нескольких треков, соответствующих одной и той же траектории движения транспорта, в частности, все треки начинаются в одной точке и в одной точке заканчиваются. Требуется выделить из этих данных «чистую», т.е. не зашумленную траекторию, по которой реально движется транспорт. Было предложено и реализовано два алгоритма решения поставленной задачи.

Первый алгоритм (Forces) основан на моделировании сил упругости и гравитации. Для получения усредненного трека выполняются следующие шаги: 1) выбирается вспомогательный трек без разрывов; 2) на этом треке равномерно расставляются узлы; 3) к узлам применяются силы: со стороны соседних узлов этого трека действует сила упругости, со стороны точек всех остальных треков – сила притяжения; 4) производится суммирование сил для каждого узла; 5) после нахождения действующих сил для всех узлов, производится перемещение каждого узла на новое место. Все шаги выполняются до тех пор, пока система не придет в равновесие.

Второй алгоритм (Median) основан на выполнении параметризации координат всех треков относительно (нормированного) времени  $t$ . Затем по времени  $t$  строится новая равномерная сетка и во всех узлах этой сетки вычисляется средняя точка среди всех траекторий, причем, в качестве средней берется медиана, что позволяет эффективно бороться с отдельными значительными «выбросами».

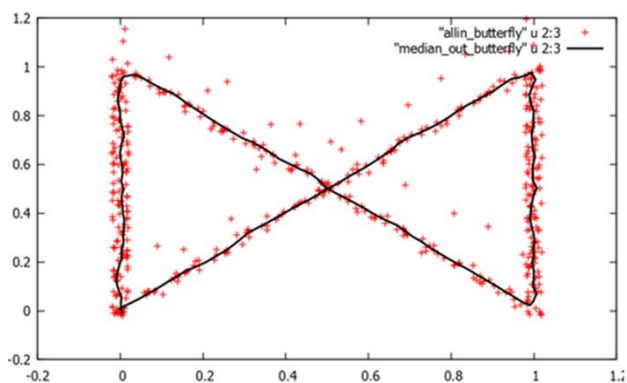


Рис. 3. Пример работы алгоритма Median

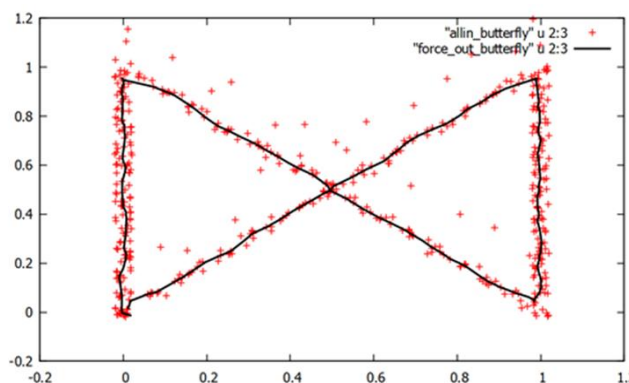


Рис. 4. Пример работы алгоритма Forces

Данные методы усреднения треков были численно протестированы на трех разных искусственных траекториях: гладкая кривая (парабола), ломаная линия без самопересечений (траектория типа «П») и ломаная линия с самопересечениями («Бабочка»). Были созданы по 10 зашумленных вариантов каждой из этих кривых, по которым и выполнялось усреднение. На рисунках 3 и 4 показаны результаты тестирования для траектории «Бабочка».

Помимо этого на тех же искусственных примерах была рассчитана погрешность работы данных методов (рисунок 5). Было выяснено, что 1) при увеличении числа точек трека уменьшается размер ошибки; 2) погрешность обоих методов примерно одинаковая. Но т.к. алгоритм Median работает намного быстрее, чем алгоритма Forces, было решено использовать именно алгоритм Median для дальнейшего тестирования уже на реальных данных.

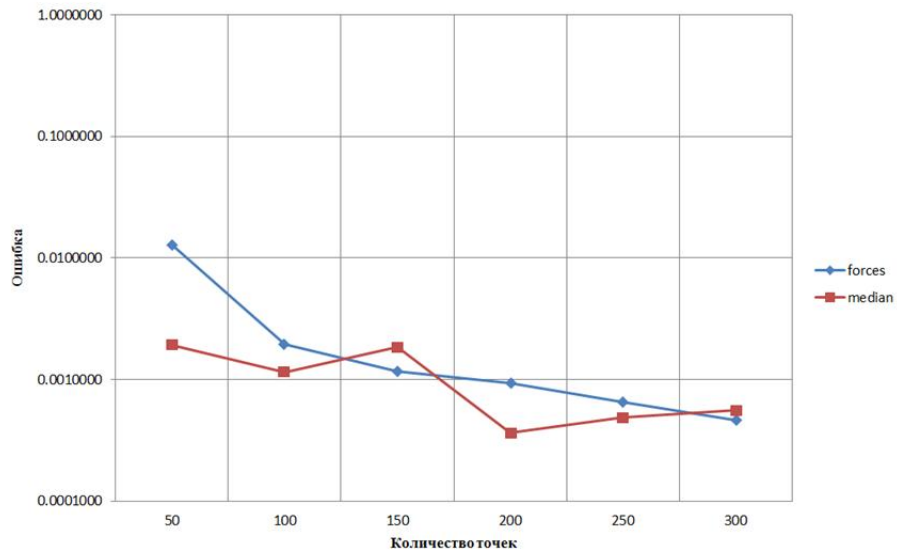


Рис. 5. Зависимость погрешности алгоритмов усреднения треков от числа точек измерения

На рисунке 6 показаны два примера применения алгоритма Median к анализу реальных треков, записанных при движении на общественном транспорте (автобусы и маршрутные такси) в городе Дубна Московской области.



Рис. 6. Примеры усреднения реальных треков

## Архитектура системы

Серьезным недостатком описанного выше подхода оказалась необходимость в существенной ручной предобработке входных данных, например, в пространственной синхронизации различных треков. Поэтому была предложена следующая архитектура программной системы автоматической генерации схем дорожного движения (рисунок 7).

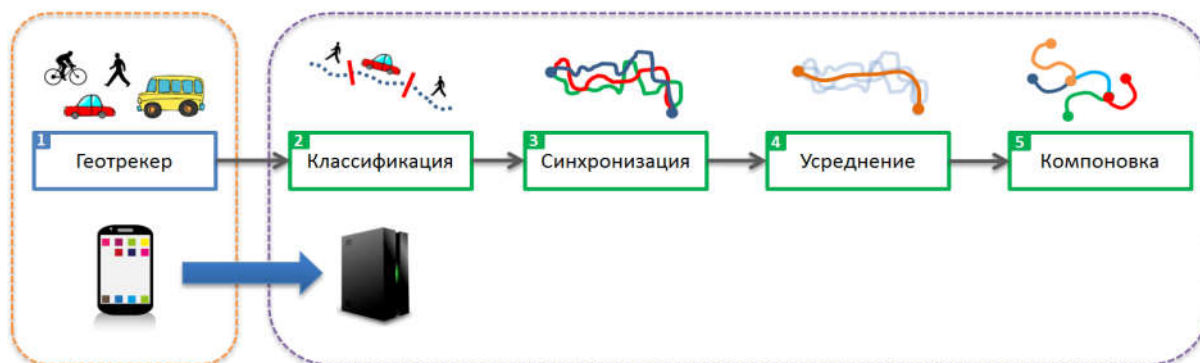


Рис. 7. Архитектура разрабатываемой программной системы

- 1) Разрабатывается программа-геотрекер для мобильных устройств, служащая для сбора ими первичной геолокационной информации в непрерывном режиме. Эта информация автоматически пересылается на некоторое облачное хранилище данных, откуда забирается сервером, выполняющим всю статистическую обработку данных.
- 2) На следующем этапе с использованием методов машинного обучения выполняется классификация треков, из них выделяются части, соответствующие движению на транспортных средствах (автомобилях).
- 3) На третьем этапе производится пространственная синхронизация треков, т.е. разбиение их на фрагменты, относящиеся к одинаковым участкам дорог.
- 4) На четвертом этапе с использованием вышеописанных алгоритмов усреднения выполняется выделение «чистых» траекторий движения транспорта.
- 5) Наконец, на последнем этапе все выделенные траектории связывают в общую систему.

В настоящее время разработан пилотный вариант алгоритма для автоматической классификации треков по типу используемого транспорта, основанный на методе логистической регрессии. Ведется работа по решению остальных задач.

## Список литературы

*Hauert J.-H., Budig B.* An algorithm for map matching given incomplete road data // Proceedings of the 20th International Conference on Advances in Geographic Information Systems (SIGSPATIAL '12). ACM, New York, NY, USA, 2012, Pages 510-513.

*Петров М.А., Ершов Н.М.* Модель сети дорожного движения на основе расширенной сети Петри // Современные проблемы прикладной математики и информатики: Тезисы докладов международной конференции (Дубна, 25-29 августа 2014 г.). — Дубна: ОИЯИ, 2014. — С. 103–106.

*Petrov M., Ershov N.* Model seti dorozhnogo dvizheniya na osnove rasshirennoy seti Petri [Simulation of the city traffic by using extended Petri networks] // Modern problems of applied mathematics and computer science, International conference for young scientists, Dubna, August 25 – 29, 2014. — Dubna: JINR, 2014. — P. 103-106 (in Russian).

# Distributed program system for automatic construction of the city traffic scheme

N. A. Puchkova<sup>1,a</sup>, N. M. Ershov<sup>1,2,b</sup>

<sup>1</sup> Moscow Region State Educational Institution for higher professional education University “Dubna”  
19, Universitetskaya str., Dubna, Moscow region, 141980, Russia

<sup>2</sup> Lomonosov Moscow State University, the Faculty of Computational Mathematics and Cybernetics  
MSU, Faculty of Computational Mathematics and Cybernetics,  
1-52, Leninskiye Gory, GSP-1, Moscow, 119991, Russia

E-mail: <sup>a</sup> pu4kova.nata@ya.ru, <sup>b</sup> ershov@cs.msu.ru

The paper discusses issues related to the construction of a digital description of the city traffic schemes on the basis of geo-location data, obtained from mobile devices (phones, tablets and so on). This description may later be used for computer modeling of traffic, for example, to optimize a particular junction. Or it can be used to the development of algorithms for autonomous vehicle control programs. One of the problems associated with the use of geolocation information that it is very noisy. Therefore, on the preliminary phase of the study we developed a number of algorithms that allow us to extract «pure» tracks from the set of noisy records corresponding to the same section of road. A serious drawback of this approach was the need for significant manual pre-processing of the input data, for example, for spatial synchronization of individual records.

In this paper, we propose the following scheme for solving the original problem. Special software (geo-tracker) for mobile devices is written, which serves to collect in continuous mode primary geolocation information. This information is sent to a cloud data storage, where is taken from by the server, which performs all the statistical data processing. In the first stage we extract from the given records their parts corresponding to the movement by use of vehicles. We do it by using machine learning methods. At the second stage the spatial synchronization of tracks (i.e. splitting them into fragments referring to the same road sections) should be performed. In the third final phase, the extraction of «pure» traffic paths and tying them into the overall system should be done. This work is devoted to the development of algorithms for solving problems arising in the first and third stages of this scheme described above. The results of numerical studies based on the analysis of real data traffic in the city of Dubna (Moscow Region) are presented.

Keywords: traffic modeling, distributed systems, geolocation

The work was supported by Russian Foundation for Basic Research (RFBR Project 14-07-00628).

© 2016 Natalia A. Puchkova, Nikolay M. Ershov