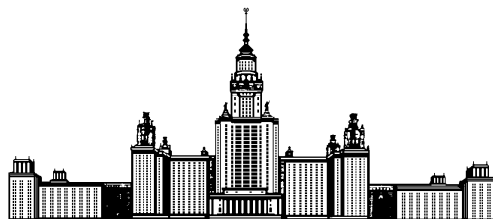


Московский государственный университет имени М.В. Ломоносова



Факультет вычислительной математики и кибернетики
Кафедра методов математического прогнозирования

Дипломная работа

Обнаружение существенно гетерогенных транспортных потоков с помощью методов интеллектуального анализа данных.

Выполнила:
студентка 5 курса 517 группы
Меркулова Татьяна Дмитриевна

Руководитель:
к.ф.-м.н.
Чехович Юрий Викторович

Москва

2014

Содержание

1	Введение	2
2	Обзор области задач	3
3	Как работают агрегаторы сигналов транспортных потоков	4
4	Представление данных	6
5	Постановка прикладной задачи	7
5.1	Причины расщепления транспортного потока	8
6	Общая схема решения задачи	9
7	Обнаружение расщепления на ребре	10
8	Отбор ребер	11
9	Учет распространения волны сгущения	13
10	Экспериментальная оценка	14
11	Заключение	14
	Литература	16

1. Введение

Данная работа посвящена одной из задач анализа транспортных систем, а именно обнаружению расщепления транспортного потока на подпотоки с разными скоростными режимами. На сегодняшний день задачи исследования автомобильных потоков являются актуальными в связи с растущим объемом передвижений в условиях ограниченного ресурса транспортных сетей. Разработано множество моделей, позволяющих учесть различные особенности процесса формирования транспортных потоков.

Широкое распространение мобильных устройств со встроенными датчиками позиционирования ГЛОНАСС/GPS существенно повлияло на область транспортных задач.[3] [1] Появились новые типы задач и новые методы их решения. Так раньше основным средством наблюдения за автомобильными потоками были датчики-радары, измеряющие количество машин, их скорость, насыщенность потока на участке, на котором они установлены. Теперь можно получать данные с мобильных телефонов и навигаторов водителей о местоположении и скорости каждого автомобиля, что позволяет формировать более полное, точное и актуальное представление транспортной ситуации и решать задачи в терминах анализа данных. [7] Одним из таких агрегаторов является портал Яндекс.Пробки[10], данные которого были использованы в работе.

Задача обнаружения скоростного расщепления потока актуальна в условиях многополосных автодорог и магистралей. Она возникает в силу того, что точность современных датчиков далеко не всегда позволяет определить полосу, по которой движется наблюдаемое транспортное средство. Часто в течение дня на некоторых участках дорожной сети возникает резкое разделение транспортного потока на медленно и быстродвижущиеся автомобили. В данной работе рассмотрены предпосылки подобных ситуаций и предложен метод их обнаружения.

Работа выполнена в рамках статистического подхода к распознаванию. Построение алгоритмов учитывает ограничения на сложность, накладываемые высокой вычислительной нагрузкой систем анализа реального времени.

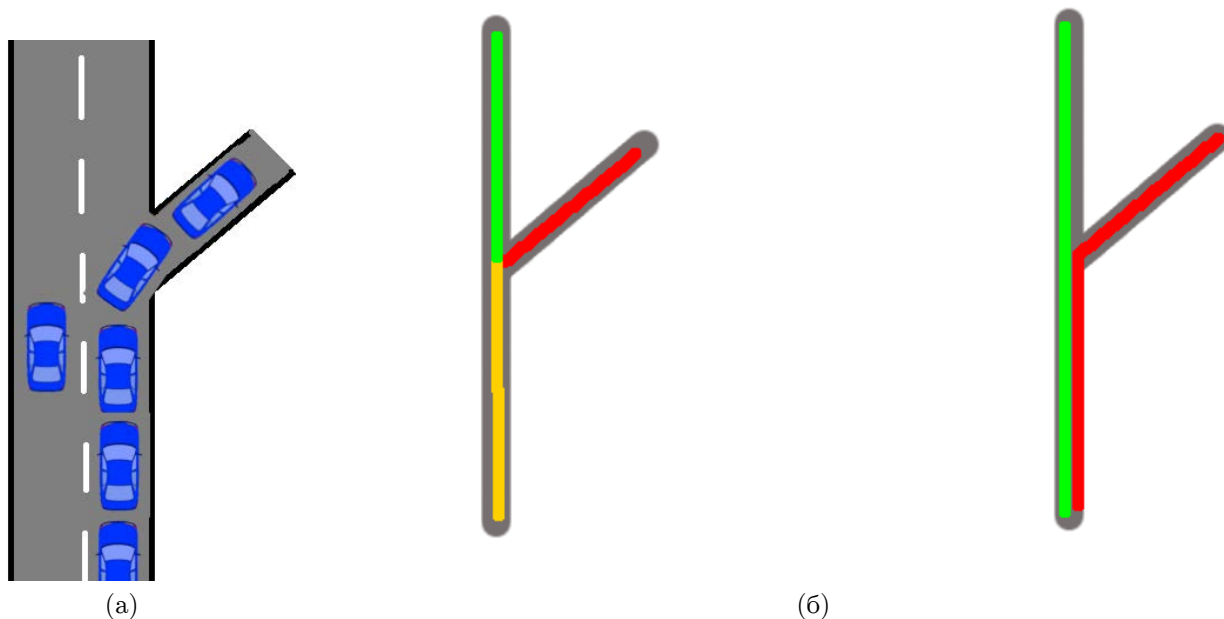


Рис. 1: Пример расщепления потока и представление дорожного графа без и с учетом расщепления

2. Обзор области задач

Транспортные системы являются темами исследований множества работ. Разработана формализация теории транспортного движения в терминах потока, его плотности, скорости. Большинство работ фокусируются на построении теоретических моделей и попытках связать параметры транспортного потока уравнениями в рамках той или иной модели.

С развитием технологий и распространением сенсоров, собирающих данные о транспортной ситуации на определенных участках, стала актуальной область анализа этих данных с целью автоматического измерения скорости и/или плотности потока, обнаружения аварий. Например, работы К. Даганзо[5] или Б. Кофмана[2] описывают схемы для обнаружения аварий на автомагистралях с помощью дорожных детекторов. С. Nanthawichit и Т. Nakatsuji предлагают систему предсказания потока трафика. [6] Главные недостатки подхода -- дороговизна установки и обслуживания сенсоров и недостаточное покрытие сенсорами дорожной сети.

Последние несколько лет большое число мобильных устройств имеют датчики позиционирования ГЛОНАСС/GPS и позволяют передавать показания этих датчиков по сети. Появляются работы исследующие эти данные. От оценки

скорости транспортного средства, передающего свое местоположение[5][2] до оценки состояния участка автотранспортной сети[11], где авторы пытаются автоматически выделить множество состояний методами кластеризации.

Становятся популярны и многие системы автоматического наблюдения за трафиком в режиме реального времени. Например, Waze [8] предлагает сервис слежения за скоростным режимом и происшествиями на дороге, решает задачи оптимизации маршрута для пользователей. Яндекс.Пробки помимо аналогичного функционала предсказывает скоростной режим и оптимизирует маршрут с учетом предсказанной дорожной ситуации. Подобные сервисы автоматически оценивают среднюю скорость на каждом из участков дорожной сети, но для обнаружения характерных транспортных ситуаций: аварий, ограничения передвижения по определенному участку, полагаются на сигналы явно отправляемые активными пользователями сервисов и ассессорами.

3. Как работают агрегаторы сигналов транспортных потоков

Чтобы понять специфику задачи и ограничения на методы ее решения, рассмотрим каким образом функционирует сервис, агрегирующий и анализирующий данные GPS-сигналов автотранспорта.

Яндекс.Пробки -- высоконагруженный сервис, предоставляющий информацию о транспортной ситуации: загруженности дорог, скоростях движения на них, появлении заторов, связанных с аварией или ремонтными работами, миллионам пользователей в реальном времени. Он также способен оценить приблизительное время выбранного маршрута с учетом ситуации на дорогах, предсказать загруженность в определенный момент дня.

Основной источник данных -- мобильные телефоны и навигаторы пользователей. Приложения Яндекс.Карты и Яндекс.Навигатор, установленные на мобильных устройствах водителей для навигации, при помощи системы позиционирования GPS определяют местоположение устройства, на котором они запущены, и передают эти сведения вместе с точным временем измерения на сервер. А он в свою очередь обрабатывает, анализирует эти данные и создает картину

пробок, доступную пользователям. Это нетривиальный процесс, требующий применения специальных алгоритмов, учета статистики.

Данные, получаемые системой, неполны: далеко не все участники движения имеют мобильные устройства с запущенными на нем приложениями Яндекс.Карты или Яндекс.Пробки. Кроме того, надежность сигнала, который передает быстро движущееся в городских условиях устройство на сервер не всегда высока.

Данные также зашумлены, так как мобильные устройства пешеходов и водителей, остановивших машину, участвуют в передаче сигналов, если на них запущены упомянутые приложения. Частичный отсев шума производят алгоритмы сервиса.

И, наконец, данные неточны. Технология GPS не позволяет достаточно точно определить положение датчика. В условиях города определить по какой полосе движется автомобиль или выехал ли он на обочину невозможно. Погрешность может составлять несколько метров. Поэтому используется специальный алгоритм, который позволяет выровнять трек автомобиля - протяженный путь следования, учитывая точную электронную схему города, однозначно отнести его к дорожному участку.[9]

Сервис работает в режиме реального времени, обновляет картину скоростного режима транспортной сети каждые 2-4 минуты.

4. Представление данных

Дорожная сеть представлена направленным графом

$$G = (V, E) \quad (1)$$

Его ребра E соответствуют отрезкам дорог. Вершины V соответствуют перекресткам, разветвлениям дорог и подразбивают протяженные участки. Каждое из ребер v содержит информацию о длине соответствующего дорожного участка L_e , широту и долготу инцидентных вершин Lat_e^{start} , $Long_e^{start}$, Lat_e^{end} , $Long_e^{end}$ и идентификаторы смежных ребер.

Обработанные и очищенные от шума данные о трафике на каждом из ребер представлены множеством записей

$$D = \{d_i = (\Delta t, \tau, sessionId)_i\} \quad (2)$$

содержащих время, за которое машина прошла участок Δt , время, в которое автомашина появилась на участке и ее уникальный идентификатор $sessionId$.

Стоит отметить, что этот идентификатор генерируется заново для каждой новой сессии -- подключения к сервису мобильного устройства и не сохраняется между сессиями. В условиях нестабильной интернет связи, характерной для быстрого движения передающего устройства в городской среде, весьма вероятно, идентификатор сменится несколько раз за время прохождения маршрута. Таким образом, нельзя надежно проследить маршрут и скорость конкретной отдельно взятой автомашины на полном треке с информацией о скоростях движения вдоль этого трека от начала движения до прибытия в пункт назначения.

5. Постановка прикладной задачи

На многополосных автодорогах в условиях интенсивного движения часто можно наблюдать ситуацию расщепления транспортного потока. Скоростные режимы движения по соседним полосам в такой ситуации существенно отличаются друг от друга. Так, например, одна полоса может полностью "встать", в то время как проезд по второй будет свободен.

На данный момент сервис Яндекс.Пробки не учитывает гетерогенность и оценивает скорость проезда по участку, усредняя скорости транспортных средств нескольких подпотоков. Конечно, эта оценка неверна и ведет к неправильной оценке общей длительности маршрута, проходящего через подобный участок. В итоге, система может предложить пользователям неоптимальный маршрут.

Стоит отметить, что расщепление -- это динамическая ситуация. Она может иметь некоторую периодичность, но это не является необходимым признаком. Подробнее этот аспект рассмотрен в разделе "Причины расщепления".

В данной работе требовалось построить алгоритм обнаружения участков с расщеплением транспортного потока, применимый в режиме реального времени.

Так для описанной структуры данных: дорожного графа и непрерывно поступающих записей о посещении ребер необходимо:

- а) Выделить $E_1 \in E$, подмножество ребер, где происходит расщепление.
- б) Для каждого $e \in E_1$ определять $\tau_{start}, \tau_{stop}$ - границы интервала расщепления потока по непрерывно поступающим записям $\{d_i = (\Delta t, \tau, sessionId)_i\}$.

5.1. Причины расщепления транспортного потока

Причины исследуемой гетерогенности разнообразны. Наиболее частые из них:

- Затруднение на съезде с магистрали. На некотором протяжении до точки съезда соответствующая полоса будет иметь существенно медленную среднюю скорость.
- Затрудненный выезд на основную дорогу. В этом случае выезжающие машины тормозят прилегающую крайнюю полосу.

Замечено, что вероятность возникновения расщепления потока на отрезке дороги неодинакова для различных участков. Какие-то из них склонны образованию описанной гетерогенности.

Стоит отличать исследуемую ситуацию от мультимодальности, обусловленной регулярными влияниями светофора на режим движения участка. Здесь также наблюдается бимодальность распределения по скоростям, но скорость прохождения автомобилем участка зависит лишь от фазы светофора, но не от выбора полосы движения.

6. Общая схема решения задачи

С целью оптимизации вычислений разобьем задачу на две подзадачи.

Первая -- подготовительная фаза, поиск участков транспортной сети: склонных к образованию гетерогенных потоков. На этом этапе мы используем исторические данные за некоторый продолжительный период времени. Этап может регулярно выполняться для обновления множества "интересных" ребер независимо от общей системы мониторинга и текущей транспортной ситуации. Полученная разметка, используется для оптимизации второй фазы.

Вторая -- мониторинг выявленных участков с целью обнаружения начала ситуации расщепления и ее завершения. Этот алгоритм уже обрабатывает данные, непрерывно поступающие от участников движения. Выделение подмножества ребер, на которых производятся дополнительные расчеты, необходимо для снижения нагрузки на вычислительные системы.

Конечно, при подобном разбиении теряются некоторые расщепления, обусловленные авариями или другими слабопредсказуемыми ситуациями. Здесь можно предложить дополнительные эвристики, как например, временное включение в множество расчетных ребер участков, прилегающих к участку, на котором произошла авария в момент поступления информации о событии. В данной работе эти дополнения рассматриваться не будут.

7. Обнаружение расщепления на ребре

Для каждого ребра доступны данные об автосредствах, проехавших соответствующий дорожный участок за последние M минут $M = 20$ мин.

1. Если машин, проехавших участок меньше N_{min} , расщепления нет. Проезд либо свободен, либо существенно затруднен по всем полосам.
2. Определим долю "медленных" машин P_{slow} , скорости на участке $< w$.
3. Определим долю "быстрых" машин P_{fast} , скорости на участке $> W$.
4. Если обе доли существенны: $P_{slow} * P_{fast} \geq P1$, считаем поток локально гетерогенным.

Параметрами алгоритма здесь являются $N_{min}, w, W, P1$.

Определим эти параметры.

N_{min} положим равным константе. Вообще говоря, этот параметр зависит от M и полосности дороги, но последний параметр нам недоступен. Поэтому произведем отсечение по константе. $N_{min} = 50$.

w и W определим исходя из определения состояний дорожного движения сервисом Яндекс.Пробки: система определяет участок, как участок с "пробкой", если средняя скорость на нем не превышает 10 км/ч. И участок "со свободным проездом", если средняя скорость на нем выше 25 км/ч. Именно эти значения и примем за пороговые.

Для определения $P1$ разметим вручную несколько последовательностей данных, содержащих как участки расщепления, так и участки без него.

Выборки линейно разделяются при $P1 = 0.028$

Если рассмотреть скоростные профили следующих за рассматриваемым участком, то различия истинного расщепления и влияния светофора очевидны. В случае расщепления поток разделяется на унимодальные подпотоки, причем их средняя скорость соответствует модам входящего потока.

В случае регулируемой ситуации, бимодальность сохраняется.

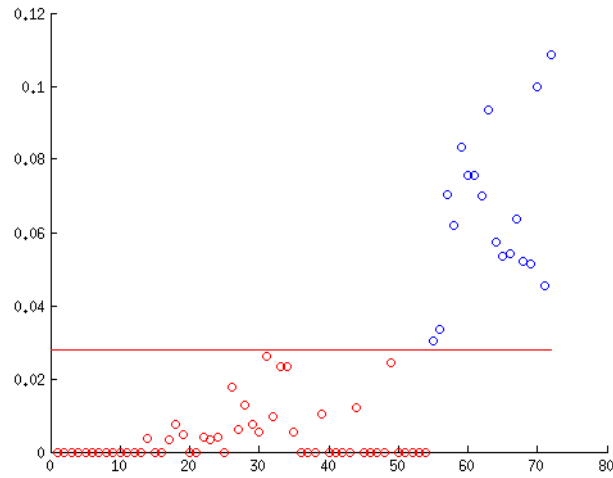


Рис. 2: Разметка тестовой выборки

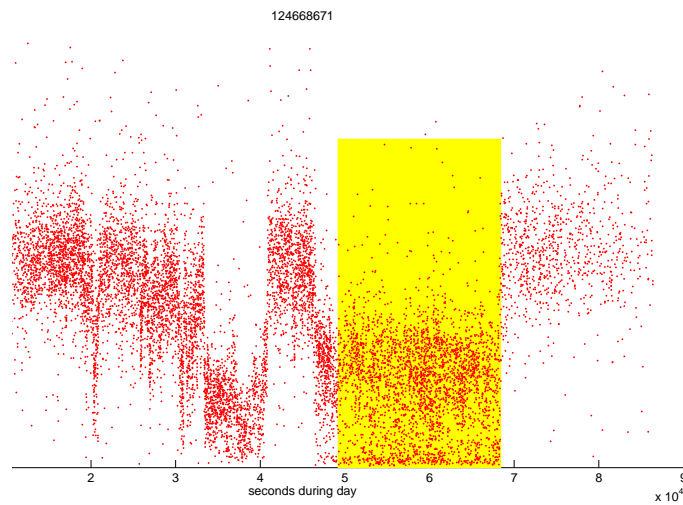


Рис. 3: Дневной скоростной профиль участка с чистым расщеплением с выделением обнаруженного расщепления

8. Отбор ребер

Допустим, что существует процедура, определяющая наличие мультимодальности скоростной гистограммы транспортного потока на дорожном отрезке в каждый момент времени. Необходимо выделить ребра, склонные к образованию расщепления. Кроме того, необходимо отделить те ребра, на которых мультимодальность задается фазой регулирующего светофора.

1. Рассмотрим длительность ситуаций мультимодальности на участках в течение дня.

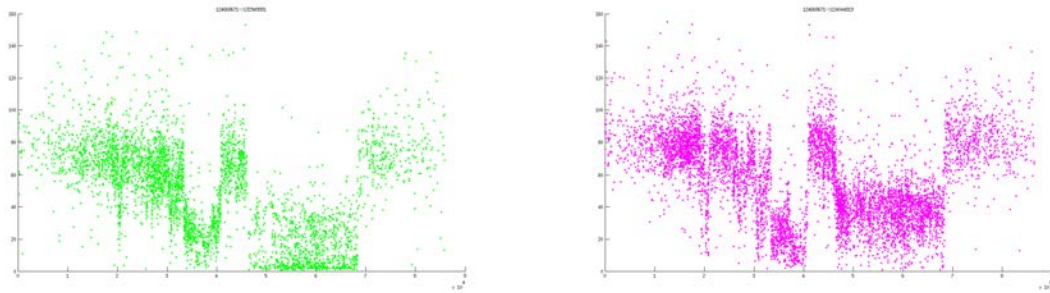


Рис. 4: Разделение потока на прилегающих участках.

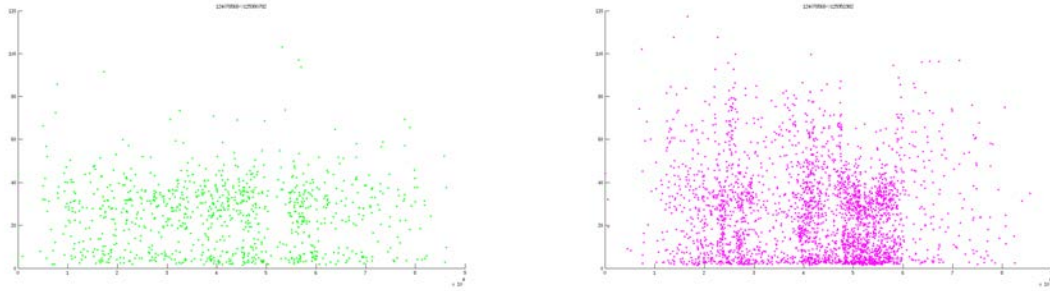


Рис. 5: Участки, прилегающие к перекрестку со светофором.

2. Проанализируем длительность и частоту возникновения мультимодальности на участках в течение недели или месяца.

Выделим участки, на которых за время долгого промежутка появлялась мультимодальность распределения скоростей. Отсечем их подмножество, на котором ее причина не расщепление потока, а результат фаз работы светофора.

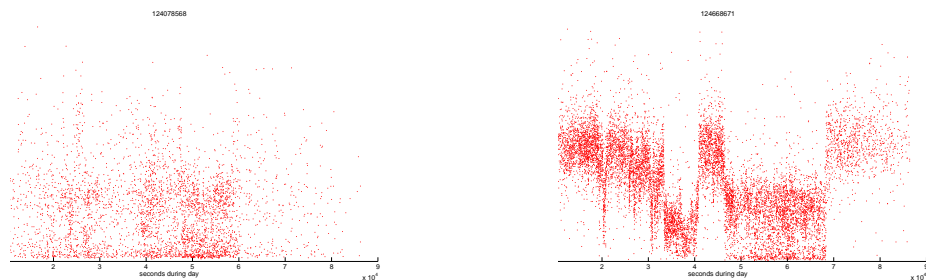


Рис. 6: Дневной скоростной профиль участка с влиянием светофора и участка с чистым расщеплением

Критерий наличия влияния светофора -- периодичность мультимодальности (с характерными периодами 24 часа и 7 дней) и существенная длительность наличия медленной компоненты, которая сохраняется в течение суток.

Если мультимодальность регулярно присутствует в течение существенной части дня -- она обусловлена в первую очередь средствами регулирования

движения, светофорами. Если же мультимодальность возникает неперiodично и длится относительно недолго, это и есть расщепление.

9. Учет распространения волны сгущения

Из эмпирических наблюдений и математических моделей транспортных потоков известно [12] [4], что после образования затрудненного движения, появляется волна плотного потока, распространяющаяся в направлении против потока. Это значит, что с момента появления гетерогенности на ребре дорожного графа, ей подвержены и ребра, входящие в него.

Учтем это явление в нашей модели, динамически изменяя множество $E1$ наблюдаемых ребер.

Если на ребре $e \in E1$ детектировано расщепление, включим в $E1$ все ребра e_1, \dots, e_k , непосредственно предшествующие e .

10. Экспериментальная оценка

Выделим базу ребер для тестирования:

Чтобы автоматически выделить периоды расщепления на ребрах рассмотрим ребра, входящие в вершины соответствующие перекрестку дорог. Вручную удалим ребра, соответствующие участкам дорог при светофорах.

Используя данные треков со смежных участков, разметим дневной профиль ребер на промежутки с существующим расщеплением и промежутки без него.

Пусть e - рассматриваемое ребро

$E_{next} = \{e_1, \dots, e_l\}$ - ребра, следующие непосредственно за e .

Если в период $T \exists e_i, e_j \in E_{next} : i \neq j; |medw_i - medw_j| \geq 15$ км/ч, где $med w_i$

- соответствующая медиана скоростей прохождения ребра e_i за период T , то помечаем ребро e как имеющее расщепление в период T .

Если $\forall e_i, e_j \in E_{next} |medw_i - medw_j| \leq 5$ км/ч, отмечаем участок, как имеющий гетерогенный поток за период T .

Проведем процедуру обнаружения расщепления на выбранных ребрах (20) дорожного графа, подразбив их дневные скоростные профили на промежутки.

	Расщепление	Нет расщепления
Обнаружено расщепление	356	7
Не обнаружено расщепления	27	1050

Таким образом, $FalsePositiveRate = 0.7\%$ $FalseNegativeRate = 7\%$

11. Заключение

Предложены методы и алгоритмы обнаружения участков транспортной сети, склонных к образованию гетерогенных потоков, и обнаружению факта расщепления по непрерывно поступающим с участка данным в режиме реального времени.

Для существенной оптимизации расчетов предполагается разбиение анализа на две фазы. Первая - анализ данных о скоростях участников движения в транспортной сети за длительный период от недели с целью обнаружения отрезков дорожного графа, где возможно появление гетерогенного потока в течение дня. Эти расчеты могут периодически выполняться на предварительно сохраненных исторических данных без привязки к мониторингу и анализу ситуации в режиме реального времени. Вторая фаза - мониторинг выделенных участков с целью выявления факта начала расщепления потока и завершения. С учетом высокой нагрузки на системы реального времени, особенно в пиковые моменты, предложенный алгоритм нетребователен к вычислительным ресурсам и применим к непрерывному потоку поступающих данных. Исследование проводилось на данных, предоставленных агрегирующим и анализирующим транспортные данные сервисом Яндекс.Пробки.

Проведен эксперимент по обнаружению гетерогенных участков потока, показывающий применимость разработанного метода.

Практическая значимость состоит в следующем: предложенные методы одобрены руководством сервиса Яндекс.Пробки и переданы на внутреннюю разработку с целью интеграции в существующий функционал сервиса.

Список литературы

- [1] Martin Treiber Arne Kesting. Online traffic state estimation based on floating car data. 2010.
- [2] B. Coifman. Identifying the onset of congestion rapidly with existing traffic detectors. 2003.
- [3] S. Amin et al. Mobile century-using gps mobile phones as traffic sensors: a field experiment. In *15th World congress on ITS, New York, NY*, 2008.
- [4] Newell G.F. A moving bottleneck. 1998.
- [5] Wei-Hua Lin and Carlos F. Daganzo. A simple detection scheme for delay-inducing freeway incidents. 1997.
- [6] C. Nanthawichit T. Nakatsuji and H. Suzuki. Application of probe-vehicle data for real-time traffic-state estimation and short-term travel-time prediction on a freeway. *transportation research record*. 2003.
- [7] Martin Treiber, Arne Kesting, and R. Eddie Wilson. Reconstructing the traffic state by fusion of heterogeneous data. *Comp.-Aided Civil and Infrastruct. Engineering*, 26:408--419, 2011.
- [8] Waze. Main page of waze service. <https://www.waze.com/>.
- [9] Yandex. Технология обработки gps треков сервисом Яндекс.Пробки. <http://company.yandex.ru/technologies/yaprobki/trek>.
- [10] Yandex. Описание сервиса Яндекс.Пробки. <http://help.yandex.ru/maps/traffic/traffic-info.xml>.
- [11] J. Yoon, B. Noble, and M. Liu. Surface street traffic estimation. pages 220--232, 2007.
- [12] Е.А.Нурминский Я.А.Холодов Н.Б. Шамрай А.В. Гасников, С.Л.Кленов. Введение в математическое моделирование транспортных потоков.