

Агрегирование разнородных данных для определения параметров транспортного потока



Е. Л. Хабибуллина,
ассистент кафедры
прикладной математики
ФГБОУ ВО «Липецкий
государственный техниче-
ский университет»
(ФГБОУ ВО «ЛГТУ»)



А. С. Сысоев,
к. т. н, заместитель
декана факультета авто-
матизации и информати-
ки по научной работе,
доцент кафедры
прикладной математики,
ФГБОУ ВО «ЛГТУ»



Н. В. Воронин,
студент 4 курса
факультета инженеров
транспорта
ФГБОУ ВО «ЛГТУ»

В статье рассмотрена возможность использования источников разнородных данных для расчета необходимых характеристик транспортного потока в региональных интеллектуальных транспортных системах. Предложена концептуальная схема модуля агрегирования разнородных данных на основе методов больших данных. Представлена организационно-техническая модель хранилища данных¹.

Создание интеллектуальной системы управления транспортным потоком, или интеллектуальной транспортной системы (ИТС) позволяет решить ряд важных задач: повысить эффективность использования автомобильных дорог, снизить воздействие на окружающую среду, оптимизировать транспортные потоки, уменьшить вероятность ДТП, гарантировать эффективность работы общественного транспорта, в случае ДТП обеспечить оперативное оповещение аварийных служб. Решение приведенных задач возможно при комплексном анализе актуальной информации из различных источников. Это автомобильные датчики, устройства отслеживания на основе спутниковых систем навигации, мобильные телефоны пешеходов, водителей и их пассажиров, камеры видеонаблюдения, радиочастотные идентификационные метки, датчики на дорогах и др. Все участники ИТС представляют собой поставщиков и потребителей данных в рамках единой системы, которую можно рассматривать как область интернета вещей [1]. Указанные источники большей частью гетерогенны не только по природе возникновения, но и по структуре. Данные из некоторых источников не структурированы.

Сегодня при решении задач интеллектуального анализа в транспортных системах наблюдается активный рост объема данных вследствие увеличения количества источников. Согласно результатам исследования Intel [2] только для реализации концепции автономных автомобилей требуется порядка 4000 Гб ежедневно (рис. 1). Таким образом, проблемы больших данных — сбор, обработка

и моделирование неструктурированных данных — актуальны для ИТС.

Следовательно, для поддержки принятия решений, а также для прогнозирования сложных ситуаций необходимо разработать современные модели организации хранения данных, методов и алгоритмов, способных обрабатывать информацию в режиме реального времени.

В статье рассмотрена модель региональной ИТС, представлена концепция согласованных автомобилей и пешеходов, описаны схема агрегирования неоднородных данных и методы, которые можно для этого использовать.

Сбор данных в интеллектуальных транспортных системах

В исследовании [3] была предложена структурная схема организации регионального модуля для интеллектуальной транспортно-логистической системы страны. Региональный центр управления транспортными потоками (РЦУТП) анализирует эколого-экономическую систему региона, систематизирует данные, объединяет их и передает в службу № 1 — «Служба моделирования и управления транспортными потоками пилотируемых транспортных средств» и службу № 2 — «Служба информационной поддержки участников дорожного движения». Служба № 1 разрабатывает имитационные транспортные модели для обеспечения оптимального организационного воздействия на транспортный поток, разрабатывает бесконфликтные и бесперебойные транспортные коридоры, управляет доступом к ним транспортных средств (ТС). Служба № 2 информирует клиентов о текущей и прогнозируемой транспортной ситуации, перегруженности транспортной сети,

¹ Исследование выполнено за счет гранта Российского научного фонда (проект № 18-71-10034).

на основании прогнозов предоставляет информацию о прокладываемых маршрутах с учетом экологической ситуации в регионе [4–6].

Электронный блок управления беспилотного ТС собирает информацию из источников данных, связывается с РЦУТП и получает оптимальный маршрут движения, разработанный центром. Управление беспилотными ТС организовано дистанционно и автоматически, по установленным правилам и организованному контролю управления. Таким образом, ошибки могут устраняться на беспилотном ТС или на определенном маршруте.

Концепция «подключенных» транспортных средств и пешеходов. Для решения некоторых задач ИТС необходимо использовать телекоммуникационные технологии концепции «подключенных» ТС и пешеходов, взаимодействующих с ИТС, городской и дорожной инфраструктурой и другими участниками дорожного движения. Интернет ТС (IoV) — это система беспроводной связи и обмена информацией между ТС, ИТС и элементами дорожной инфраструктуры. Рассмотрим ряд интерфейсов в концепции «подключенных» автомобилей и пешеходов.

Интерфейс «ТС–инфраструктура» (vehicle–infrastructure, V2I) дает возможность ТС получать информацию или обмениваться ей с различными объектами дорожной инфраструктуры: дорожными хабами (которые, в свою очередь, обмениваются информацией с региональными модулями ИТС), светофорами и дорожными знаками.

Интерфейс «ТС–ТС» (vehicle–vehicle, V2V) позволяет двум ТС обмениваться информацией. Иными словами, «подключенный» автомобиль может получать данные о скорости, местонахождении или действиях водителя другого автомобиля. Системы V2V, способные обеспечивать безопасность и контроль дорожного движения, помогая управлять автомобилем, могут найти успешное применение в беспилотных ТС.

Интерфейс «ТС–пешеход» (vehicle — pedestrian, V2P) обеспечивает взаимодействие ТС с пешеходами поблизости. Электроника автомобиля определяет диапазон частот смартфонов, что дает возможность оценивать скорость и направление движения пешехода.

Интерфейс «пешеход–инфраструктура» (pedestrian–infrastructure, P2I) позволяет пешеходам обмениваться информацией с ИТС и объектами инфраструктуры. Это необходимо для эффективного



Рис. 1. Ожидаемый поток данных в автономных автомобилях

управления городским общественным транспортом, предоставления пассажирам информации о движении ТС в реальном времени, моделирования пассажиропотоков и прогнозирования востребованности транспорта пассажирами.

Использование концепции согласованных ТС поможет значительно улучшить управление транспортными коридорами и решить другие задачи, поставленные при разработке ИТС. Работа указанных интерфейсов между подключенными автомобилями, пешеходами и ИТС основана на источниках неоднородных данных.

Параметры интеллектуальной транспортной системы и источники разнородных данных. Для обеспечения эффективной работы ИТС необходимо проанализировать параметры транспортного потока: скорость, интенсивность, плотность. Система предназначена для моделирования и управления транспортным потоком, построения оптимального маршрута для пилотируемых ТС, разработки оптимального маршрута для беспилотных ТС, моделирования грузопотоков и планирования перевозок по транспортным коридорам, управления доступом к ним, моделирования потока пассажиров и управления городским общественным транспортом, обеспечения безопасности участников дорожного движения, снижения аварийности на дорогах. Поэтому необходимо собирать и обрабатывать информацию о взаимном расположении ТС, пешеходов, о метеорологических условиях, моделировать перемещения участников дорожного движения и грузов.

Эта информация неоднородна. Данные, собираемые ИТС, могут быть цифровыми или аналоговыми [7]. Данные цифрового характера создаются компонентами дорожной инфраструктуры или компонентами ИТС специально для использования в компьютерной обработке.

Это, например:

- данные местоположения;
- информация с мобильного устройства (состояние, местоположение), подключенного к различным сетям (GSM, Wi-Fi);
- данные от «подключенных» ТС и/или дорожной инфраструктуры;
- информация о транспортных картах и терминалах оплаты в общественном транспорте;
- сообщения электронной почты, SMS, в различных мессенджерах и социальных сетях, метаданные телефонных звонков.

Данные аналогового характера возникают как следствие физического явления и преобразуются в цифровой сигнал. К таким данным относятся, например:

- записи с камер видеонаблюдения и беспилотных средств, камер фото-, видеофиксации нарушений;
- аудиозаписи голосовых звонков с мобильного телефона;
- данные от акселерометров, компасов, измерителей температуры, инфракрасного излучения, электромагнитного поля, атмосферного давления;
- физические и медицинские данные с носимых устройств синхронизируемого объекта ИТС;
- данные от радаров, приборов, анализирующих RFID-метки.

Таким образом, объемы данных увеличиваются, большей частью они плохо структурированы или вовсе не структурированы (фото, видео, социальные сети). Большие данные обычно включают в себя такие наборы данных, размеры которых выходят за рамки возможностей широко используемых программных инструментов для получения, хранения, управления и обработки информации. Большие данные позволяют получить подробную информацию о пользователях дорог и их поведении,



Рис. 2. Децентрализованный подход к агрегации разнородных данных

но должны быть проанализированы децентрализованным способом, чтобы избежать передачи больших объемов информации по сетям. Поэтому при решении задач ИТС для обработки больших данных предпочтительны облачная и GRID-инфраструктура.

Модели и методы агрегирования разнородных данных

Децентрализованный подход к агрегации разнородных данных. Все источники информации взаимосвязаны, и любой элемент информации может быть доступен любому компоненту ИТС.

Для соблюдения баланса между децентрализованной обработкой данных и затратами на передачу данных/обработанной информации используют методы интеллектуального анализа больших данных (BDPM). Это требует разработки алгоритмов для сбора, агрегирования, обработки и анализа информации, необходимой для принятия оптимальных решений и эффективных стратегий поведения пользователей.

Неоднородные данные можно описать следующими понятиями [7]:

- местоположение (система координат, широта и долгота) передающего устройства или сообщаемого события;
- время передачи данных или действия, о котором сообщается;
- вид источника данных (геопозиционные данные мобильного телефона, GPS, сводки новостей и т. д.);
- формат данных (одно значение, матрица, вектор, текст, изображение и т. д.);
- представление данных (например, единица измерения);
- семантика данных (например, отслеживание ТС или мобильного телефона).

Согласно результатам исследования [8] передача данных — узкое место в концепции облачных вычислений. Поэтому при использовании BDPM очень важен разумный баланс обработки данных на устройствах пользователей (клиентов) и в облачной инфраструктуре. Если устройство клиента достаточно мощное, оно может предварительно обработать данные, а затем передать их в облачную инфраструктуру, что сократит потребление услуг облачных сервисов и сетевой трафик. В противном случае первоначальные данные передаются в облако, где подвергнутся предва-

рительной и последующей обработке. Описываемый процесс децентрализации агрегации и обработки данных представлен на рис. 2.

На начальном этапе сбора данных действуют два механизма фильтрации, ограничивающие объем получаемых данных. Первый фильтр — конструктивные требования (и, следовательно, ограничения), предъявляемые к любому датчику или записывающему устройству. Второй фильтр — ограничения скорости наблюдения регистрируемых или контролируемых событий скоростью записи и передачи данных.

Ценность полученных данных можно не распознать на этапе проектирования системы, поэтому необходимо обеспечить сохранение потенциально полезной информации.

Концептуальная схема агрегирования разнородных данных. Некоторые задачи можно решить на основе анализа отдельных наборов данных, однако реальный потенциал ИТС основан на расширении применения аналитических методологий к многочисленным источникам разнородных данных. Для агрегирования данных их наборы сопоставляются и объединяются на основе общих атрибутов и переменных, каждый составленный набор данных сохраняется.

Алгоритмы интеллектуального анализа больших данных отличаются от традиционных: им не нужна модель, описывающая взаимосвязь данных, не требуются специальные запросы, на которых основывается анализ. Алгоритмы используются для обнаружения шаблонов и паттернов, которые не появляются в отдельных наборах данных [7]. На рис. 3 представлена концептуальная схема агрегации данных, основанная на методологии CRISP-DM.

Кроме метода BDPM можно применять подход, основанный на мультиагентных системах (MAS). Система MAS [8] состоит из нескольких автономных программных объектов, называемых агентами. Они воспринимают информацию из источников разнородных данных, создают собственные локальные модели данных, позволяют выбрать наиболее подходящую модель агрегации для каждой конкретной задачи. Агенты взаимозависимы и взаимодействуют на уровне информационных моделей (обмен данными или параметрами модели) или на уровне действий (координация действий, формирование групп).

Организационно-техническая модель хранилища данных. Предлагаемая



Рис. 3. Концептуальная схема агрегации разнородных данных на основе методологии CRISP-DM

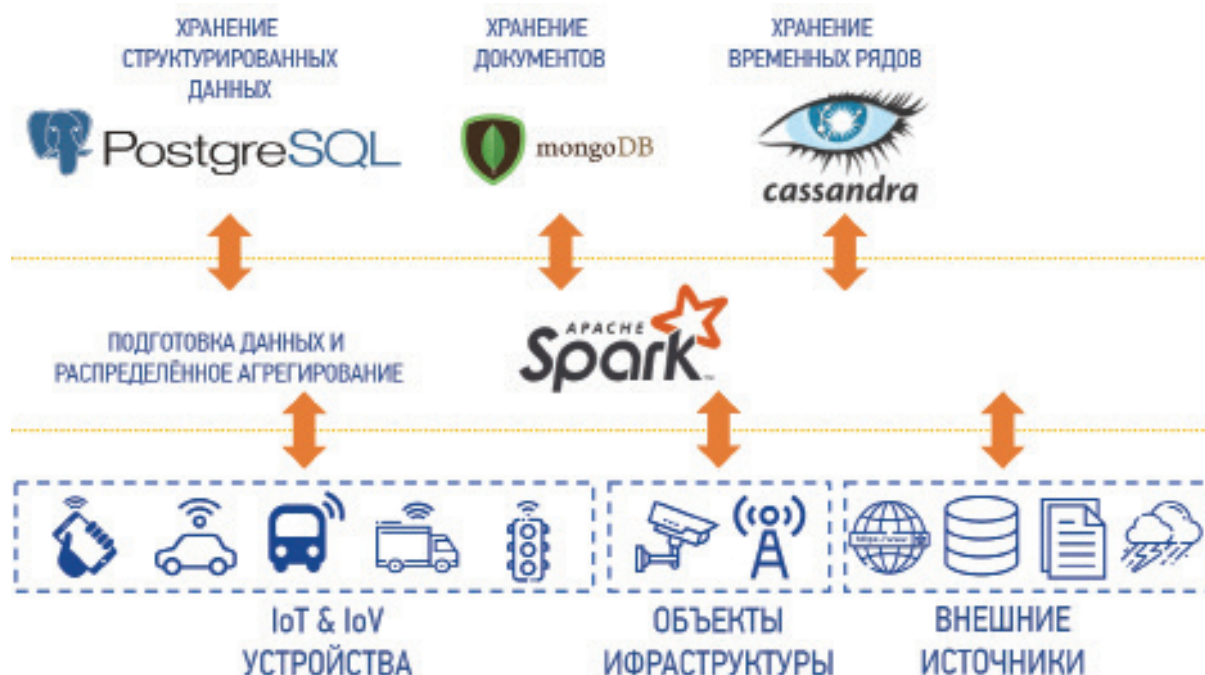


Рис. 4. Организационно-техническая модель хранилища данных

модель (рис. 4) позволит собирать информацию из источников разнородных данных, проводить их предварительную обработку и агрегирование, использовать наиболее подходящую СУБД (реляционную или NoSQL в зависимости от степени структурированности и природы данных).

Система Apache Spark будет использоваться для реализации распределенной предварительной обработки (фильтрации и агрегирования) неструктурированных и слабоструктурированных данных, которые поступают из гетерогенных источников в разных форматах.

Сегодня система PostgreSQL входит в число наиболее удобных СУБД для работы с хорошо структурированными данными, которые описываются в терминах реляционных баз данных. MongoDB позволяет эффективно хранить и использовать данные о транспортном потоке, полученные из источников разнородных данных. IoT и IoV — это сети из миллионов устройств, генерирующих данные в формате временных рядов. Для хранения и обработки таких данных можно использовать реляционные СУБД, но специализированные базы данных (например, Cassandra) обеспечивают масштабируемость работы с большими объемами данных и предоставляют специальные функции для обработки временных рядов.

Таким образом, в ИТС должна использоваться информация из многих возможных источников, хотя данные, получаемые

из них, разнородны. Эти источники информации порождают большие объемы данных. Для сбора, хранения и обработки разнородных данных рекомендуется использовать облачную инфраструктуру с разделением обработки данных на устройстве клиента и в облаке. Для качественного агрегирования разнородных данных необходимо исследовать в них скрытые зависимости, анализировать различные варианты моделей агрегации с использованием BDPM и интерпретировать их согласно задачам ИТС.

К сожалению, далеко не все участники дорожного движения захотят предоставить для ИТС информацию с персональных устройств даже при обеспечении требуемого уровня анонимности. Кроме того, сложно разделить информацию с носимых и мобильных устройств и личность субъекта, который использует предоставляющее информацию устройство, и обеспечить гарантии безопасности полученных данных. ■

Литература

1. Guerreiro G., Figueiras P., Silva R. et al. An Architecture for Big Data Processing on Intelligent Transportation Systems An application scenario on highway traffic flows// Proceed. Abstr. IEEE 8th Int. Conf. on Intelligent Systems. 2016. P. 66–72.
2. Krzanich B., Intel. Data is the New Oil in the Future of Automated Driving. 2016. URL: [\[automated-driving/#gs.8q7ekx\]\(https://newsroom.intel.com/editorials/krzanich-the-future-of-automated-driving/#gs.8q7ekx\) \(Дата обращения 24.04.2019\).](https://newsroom.intel.com/editorials/krzanich-the-future-of-

</div>
<div data-bbox=)

3. Sysoev A., Kadasev D., Lyapin S. et al. Conceptual Scheme Of Regional Module For Intelligent Transportation And Logistics System // Proceed. Abstr. Int. conf. on traffic and transport engineering. Belgrade, 2018.
4. Корчагин В. А., Антропов В. А., Ляпин С. А. и др. Экологические проблемы развития России // Вестн. Урал. гос. ун-та путей сообщения. 2016. № 3 (31). С. 128–138.
5. Новиков А. Н., Катунин А. А., Семкин А. Н. Современные подходы к управлению перевозками грузов автомобильным транспортом // Мир трансп. и технол. машин. 2015. № 1 (48). С. 119–126.
6. Korchagin V. A., Lyapin S. A., Rizayeva Yu. N. et al. Subsystem of Road Accident Consequences Elimination. Methodology of Subsystem Efficiency Improvement// Transp. Res. Proced. 2017. Vol. 20. № 3. P. 316–321.
7. International Transport Forum. Big Data and Transport Understanding and assessing options: Report, Corporate Partnership Board, Leipzig, Germany, 2015. — URL: https://www.itf-oecd.org/sites/default/files/docs/15cpb_bigdata_0.pdf (Дата обращения 24.04.2019).
8. Fiosina J., Fiosins M., Müller J. P. Big Data Processing and Mining for Next Generation Intelligent Transportation Systems // J. Teknol. (Sci. & Engineer.). 2013. № 3 (63). P. 23–38.