

Лебедева Ольга Анатольевна,

к.т.н., доцент, ФГБОУ ВО «Ангарский государственный технический университет»,
e-mail: kravhome@mail.ru,

Ерофеев Ермак Владимирович,

студент гр. ИТБ-23-1, ФГБОУ ВО «Ангарский государственный технический университет»,
e-mail: ermak.080@mail.ru

Баянов Семен Владимирович,

студент гр. ТТП-23-1, ФГБОУ ВО «Ангарский государственный технический университет»,
e-mail: cyja642@yandex.ru

КОНТЕКСТНО-ЗАВИСИМОЕ МОДЕЛИРОВАНИЕ ДАННЫХ В РАБОТЕ ОБЩЕСТВЕННОГО ТРАНСПОРТА

Lebedeva O.A., Erofeev E.V., Bayanov S.V.

METHODOLOGY FOR INTEGRATION OF CONTEXTUAL ROAD TRAFFIC DATA IN THE FRAMEWORK OF PUBLIC TRANSPORT PERFORMANCE ANALYSIS

Аннотация. В работе рассматривается методология интеграции ситуационного контекста в анализ данных о работе общественного транспорта. Анализ включает: онлайн-консолидацию и маркировку разнородных источников контекста; статистическое моделирование ожидаемых характеристик транспорта на основе факторов сезонной и суточной загруженности; и интегративное отображение дорожного движения и его ситуационного контекста, сопровождаемое пространственно-временной навигацией и возможностями масштабирования. Предварительные результаты, полученные на основе данных о сети, показывают актуальность этого направления исследования для поддержки принятия решений с учетом контекста.

Ключевые слова: моделирование, общественный транспорт, ситуационный контекст, данные о пассажиропотоках.

Abstract. This paper examines a methodology for integrating situational context into the analysis of public transport data. The analysis includes: online consolidation and labeling of heterogeneous context sources; statistical modeling of expected transport behavior based on calendar load; and an integrated display of road traffic and its situational context, accompanied by spatiotemporal navigation and zoom capabilities. Preliminary results obtained using network data demonstrate the relevance of this research direction for supporting context-aware decision making.

Keywords: modeling, public transport, situational context, passenger flow data.

В последние годы актуальность приобретает сбор всех доступных данных о дорожном движении для решения задачи устойчивой мобильности. Данные включают общественные мероприятия (спортивные и культурные мероприятия, конгрессы), запреты (плановые строительные работы, аварии и уведомления от населения), карты городского планирования, а также другие виды деятельности, потенциально влияющие на мобильность [1]. Интегративный анализ контекстных данных предоставляет уникальные возможности для понимания динамики дорожного движения и, при наличии информации о предстоящих мероприятиях или запланированных событиях, для краткосрочного и долгосрочного планирования работы общественного транспорта [2]. Несмотря на актуальность контекстно-зависимого анализа данных о дорожном движении, в нем наблюдаются и недостатки. Во-первых, контекст отсутствует в анализе данных о дорожном движении. Во-вторых, источники данных либо являются частными,

либо разрозненными, либо недоступны. В-третьих, данные о дорожном движении могут анализироваться с использованием нескольких временных, пространственных и модальных разрешений, что приводит к объемным и труднопригодным результатам. Наконец, моделирование транспортного потока для изучения значимых отклонений в движении транспорта не всегда адекватно [3]. Эти проблемы препятствуют всестороннему, действенному и оперативному пониманию динамики движения транспорта, что приводит к неэффективности системы мобильности. Для решения этих задач в исследовании разработана система поддержки принятия решений, основанная на пяти основных принципах:

1) консолидация данных о транспорте от перевозчиков с широким спектром источников о ситуационном контексте;

2) онлайн-аналитика, основана на обновляемом поиске и автоматической маркировке (в соответствии с пространственным, временным и модальным охватом) контекстных данных;

3) качественное моделирование транспортного поведения с использованием статистических моделей с календарным планированием для многомерных временных рядов данных;

4) интегративное отображение (как ожидаемого, так и наблюдаемого) транспортного потока и его ситуационного контекста, сопровождаемое временной навигацией и возможностями масштабирования;

5) иерархические представления данных о дорожном движении (агрегированные данные доступны на уровне остановочных пунктов, маршрутов) для поддержки оперативных решений.

Контекстно-зависимая аналитика имеет решающее значение для поддержки работы общественного транспорта, рекомендаций по времени ожидания, устойчивого планирования транспорта, оповещений в режиме реального времени.

Общедоступные данные о дорожном движении сложны, часто многомерны, пространственны, временны и разрежены. Данные о городском движении – будь то автомобильное, железнодорожное, речное или пешеходное – зависят от контекста ситуации, включая спортивные/культурные мероприятия, ограничения движения, городское планирование, погодные условия и другое. Для понимания моделей городской мобильности такие контекстуализирующие факторы должны быть интегрированы в анализ данных о дорожном движении. Прогресс достигается как за счет специализированного анализа данных о дорожном движении, так и за счет контекстно-обогащенных данных о дорожном движении. В обоих направлениях исследований вклад можно классифицировать в соответствии с целевой задачей: описательный, прогностический и/или предписывающий. Описательный анализ направлен на обобщение и извлечение нетривиальных, значимых и статистически важных связей из городских данных, включая отклоняющиеся, периодические и возникающие пространственно-временные закономерности; прогностический анализ направлен на прогнозирование динамики мобильности; а предписывающий анализ направлен на под-

держку принятия решений с использованием оптимизационных и управляющих подходов.

В области описательной аналитики вклад в моделирование поведения транспортного потока делится на три основных направления: классическое статистическое моделирование, моделирование на основе машинного обучения и гибридное моделирование. В первом направлении популярны авторегрессионные модели, модели скользящего среднего и модели сглаживания, используемые с 1970-х годов, с разнообразными приложениями для моделирования (и прогнозирования) данных об объеме и скорости транспортного потока. Их главный недостаток – необходимость значительного объема исторических данных для построения качественной модели. Второе направление, моделирование на основе машинного обучения, включает в себя вклад в анализ временных закономерностей, многомерный анализ временных рядов и кластеризацию подпространств. Несмотря на свою актуальность, эти модели, как правило, дают локальное представление о поведении транспортного потока. Исследования третьего направления, гибридного моделирования, изучают синергию между подходами двух предыдущих направлений. В дополнение к этому, были предложены системы визуальной аналитики для решения различных проблем мобильности, таких как анализ закономерностей заторов на основе мобильных данных, динамика дорожного движения на основе данных счетчиков движения и социальной активности людей. Помимо описательной аналитики, в исследованиях в области прогнозирования наблюдается активность в области глубокого обучения (многоагентные системы, таких как обучение с подкреплением), при этом все больше попыток позволяют исследовать присущую данным о дорожном движении пространственную и временную природу.

Следует отметить, что в некоторых исследованиях анализировались различные типы ситуационного контекста, такие как погодные данные, сведения об авариях, полученные из социальных сетей с помощью методов обработки, а также запланированные мероприятия, например, фестивали и спортивные соревнования. В рамках этих работ подчеркивается важность интеграции разнородных источников информации, включая сочетания запланированных событий и метеоданных, а также указывается, что разнообразие таких данных является ключевым фактором в развитии «умных городов». Область контекстно-зависимой аналитики дорожного движения менее изучена из-за связанных с ней проблем (идентификация и поиск релевантного контекста, обучение на основе разнородных данных). Включение этой информации имеет первостепенное значение для объяснения поведения мобильности и даже для создания более совершенных описательно-прогностических моделей и процедур оптимизации. В отличие от представленного подхода, ни одна из этих работ не поддерживает анализ отклоняющегося поведения на основе исторических данных и не представляет контекстно-обогащенное отображение поведения дорожного движения с использованием средств временной и пространственной навигации.

Для автоматического и комплексного отображения данных о дорожном движении и их контекста должна быть реализована периодическая процедура считывания новой информации из разнообразных и разнородных источников данных о ситуационном контексте. Кроме того, полученные события необходимо автоматически аннотировать в соответствии с их категорией, а продолжительность события указывается в соответствии с доступной информацией или дополняется при наличии правил пользователя.

Наблюдаемое и смоделированное поведение транспортного потока может быть оценено с использованием:

1) различных методов: Холт-Винтерс, SARIMAX и рекуррентные нейронные сети (проведем сравнительный анализ методов в сводной таблице 1);

Таблица 1 – Ключевые характеристики методов [4]

Характеристика	Холт-Винтерс (Holt-Winters)	SARIMAX	Рекуррентные нейронные сети (RNN)
тип подхода	классический статистический	классический статистический	машинное обучение / глубокое обучение
интерпретируемость	высокая (компоненты тренда и сезонности наглядны)	высокая (интерпретация коэффициентов AR, MA, I)	низкая (сложно понять, как именно модель делает прогноз)
учет сезонности	поддерживается	поддерживается (встроенная сезонная компонента)	сеть извлекает закономерности, если имеются данные
экзогенные переменные	не поддерживаются	поддерживается (возможность включать внешние факторы)	поддерживается (сеть может обучаться на многомерных данных)
сложность и скорость	низкая, обучается быстро	средняя, требует тщательной настройки параметров	высокая, требует больших данных и мощных вычислений
объем данных	эффективна на малых и средних данных	качественно работает на малых и средних данных	требует больших объемов исторических данных для обучения

2) календарно-независимых и специфичных (рабочий/ выходной, день недели, недифференцированные даты);

3) небольших и длинных разбиений (от количества точек),

4) различной временной детализации (окна от минут до нескольких часов) и пространственной детализации (специфичные для остановочного пункта, региона, маршрута и всей сети).

Метод моделирования/прогнозирования Холт-Винтерса оказался наиболее конкурентоспособным, поскольку способен использовать тройное экспо-

ненциальное сглаживание для учета ключевых компонентов временного ряда: уровня, тренда и сезонности (аддитивной или мультипликативной).

Для всестороннего анализа и выявления лучших статистических моделей на основе многомерных данных временных рядов необходимо учитывать данные о средней абсолютной ошибке (MAE), среднеквадратичной ошибке (RMSE) и средней абсолютной процентной ошибке (MAPE).

Рассмотренная методология онлайн-анализа данных о дорожном движении в контексте конкретной ситуации предоставляет уникальные возможности, по сбору исторических и перспективных данных. С этой целью рассматривается инструмент, способный собирать и маркировать эти события в режиме реального времени и интегрировать их отображение с анализом данных о дорожном движении.

Критическая важность контекстно-чувствительного анализа и его дополнение надежно смоделированным эталонным поведением транспорта, а также географическими и временными средствами навигации для поддержки как краткосрочных, так и долгосрочных решений в области планирования доказана посредством анализа литературы, и определены основные направления дальнейших исследований: 1) расширение предложенной методологии для много-модального анализа данных о дорожном движении; 2) контекстно-чувствительное моделирование ожидаемого поведения на основе выборочного воспроизведения исторических данных с аналогичным контекстом; 3) автоматическое обнаружение аномального поведения; и 4) контекстно-зависимое прогнозирование данных о дорожном движении при наличии запланированных событий.

ЛИТЕРАТУРА

1. Транспортное планирование в развитии Иркутской агломерации / А. Левашев, М. Шаров, О. Лебедева [и др.] // Проект Байкал. – 2021. – Т. 18, № 70. – С. 94-98. – DOI 10.51461/projectbaikal.70.1896.

2. **Полтавская, Ю. О.** Методы сбора данных о продолжительности движения на маршруте и требования к объему выборки / Ю. О. Полтавская // Вестник Ангарского государственного технического университета. – 2018. – № 12. – С. 192-195. – DOI 10.36629/2686-777X-2018-1-12-192-195.

3. **Лебедева, О. А.** Поиск оптимального проложения автобусного маршрута в транспортной сети города / О. А. Лебедева // Информационные технологии в науке, управлении, социальной сфере и медицине : Сборник научных трудов V Международной научной конференции. В 2-х частях, Томск, 17–21 декабря 2018 года / Под редакцией О.Г. Берестневой, А.А. Мицеля, В.В. Спицына, Т.А. Гладковой. Том 1. – Томск: Национальный исследовательский Томский политехнический университет, 2018. – С. 92-94.

4. **Leite, I.** Context-sensitive modeling of public transport data / I. Leite, A. Finamore, R. Henriques // – URL: <https://fenix.tecnico.ulisboa.pt/downloadFile/1689244997259528/81328-ines-elaterium.pdf> (дата обращения: 13.01.2026).