

УДК 004.8

ИССЛЕДОВАНИЕ МЕТОДОВ АНАЛИЗА ДАННЫХ ДВИЖЕНИЯ ОБЩЕСТВЕННОГО ТРАНСПОРТА ДЛЯ РЕШЕНИЯ ЗАДАЧИ ПРОГНОЗИРОВАНИЯ ПРИБЫТИЯ ТРАНСПОРТА НА ОСТАНОВКИ

© **Беляев П.И., Белоусов А.А.**

*Самарский национальный исследовательский университет
имени академика С.П. Королева, г. Самара, Российская Федерация*

e-mail: pasha9819@gmail.com

Современный город сложно представить без общественного транспорта. Во многих российских городах применяется спутниковое навигационное оборудование для диспетчеризации движения общественного транспорта. Обладая данными о местоположении каждой единицы транспорта, диспетчер может оперативно реагировать на нештатные ситуации, изменять маршруты движения транспортных средств, оптимизировать скорость их движения. Эти же данные можно использовать для информирования пассажиров о времени прибытия транспорта на остановку. В нашем городе такие данные предоставляет транспортный оператор Самары [1].

Существует несколько групп методов для решения задачи прогнозирования прибытия транспортного средства на остановку: методы, использующие исторические данные; статистические методы; технологии машинного обучения.

Методы, которые основаны на использовании исторических данных, достаточно хорошо прогнозируют движение транспорта при отсутствии заторов и случайных помех. При этом, используя только исторические данные, невозможно улавливать изменения дорожной ситуации в режиме реального времени. В Самаре применяется алгоритм, вычисляющий прогноз прибытия на следующую остановку по ходу движения транспорта, а для всех последующих остановок время прибытия вычисляется, исходя из среднего времени прохождения отрезков маршрута [2].

Технологии машинного обучения позволяют давать наиболее точный прогноз за счет моделирования сложных нелинейных связей между входными и выходными параметрами моделей. Однако использование моделей машинного обучения зачастую требует колоссальных вычислительных затрат для прогнозирования выходных данных.

Статистические методы являются наиболее популярными в рамках решения поставленной задачи. Они аккумулируют в себе достоинства вышеупомянутых групп методов, а также нивелируют их основные недостатки. Зачастую для прогнозирования времени прибытия транспорта на остановку используется модель линейной регрессии.

Для обучения моделей необходимо подготовить данные о движении транспорта. Для этого были разработаны два алгоритма: первый предназначен для накопления исторических данных, второй – для сбора данных в режиме реального времени. Первый алгоритм обрабатывает логики транспортного оператора, которые представляют собой текстовый файл, в который записаны строки с информацией о местоположении транспортных средств (с привязкой ко времени). На выходе алгоритма была получена информация о времени движения транспорта между каждой парой подряд идущих остановок для каждого маршрута города. Вторым алгоритмом предназначен для сбора данных о прогнозируемом времени движения транспорта до конкретной остановки. Данные, полученные на выходе второго алгоритма, будут использованы для оценки погрешности прогнозов.

Транспортная модель города была представлена в виде графа, вершинами которого будут являться остановки общественного транспорта, а ребрами – перегоны

между остановками. Таким образом, рассматриваемую задачу прогнозирования можно свести к вычислению времени прохождения транспортным средством каждого ребра на пути от его текущего местоположения до любой выбранной конечной точки.

Для обучения предсказывающих моделей были сформированы наборы данных, содержащие время, в которое было совершено прохождение ребра, день недели, тип транспортного средства (в виде целого числа), среднее время прохождения этого ребра, время прохождения данного ребра последними пятью транспортными средствами, а также текущее время прохождения.

Для исследования были взяты следующие модели: линейная регрессия, многослойный перцептрон, модель авторегрессии – скользящего среднего, регрессионное дерево (REP Tree), модель скользящего среднего с усреднением на основе исторических данных, усредненная модель (модель, вычисляющая среднее значение между прогнозами всех вышеупомянутых моделей).

После обучения моделей был проведен анализ точности прогнозов каждой модели. Погрешности вычислялись с использованием оценок MAE и MAPE. На рисунке приведен график относительной погрешности.

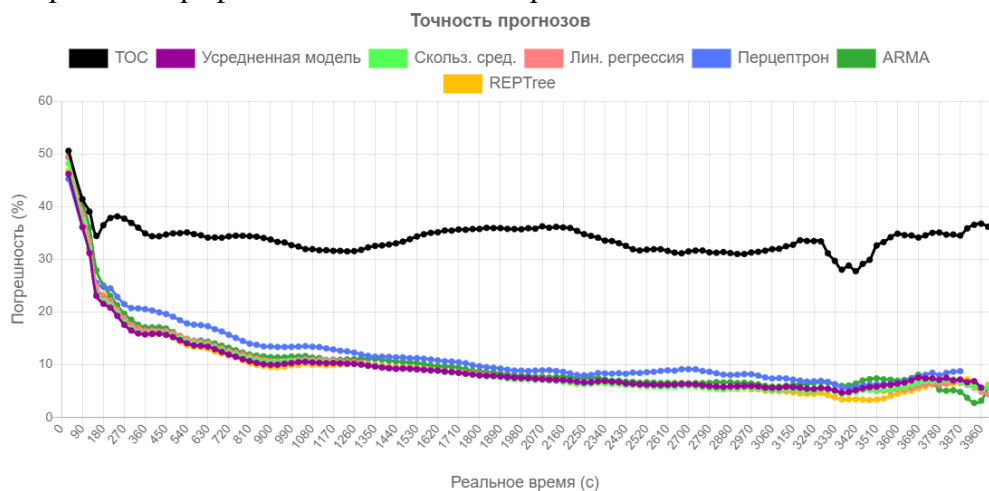


Рис. График зависимости относительной погрешности от реального времени

В ходе анализа полученных результатов было установлено, что все реализованные алгоритмы в среднем дают прогнозы, погрешность которых в 3 раза меньше погрешности прогнозов транспортного оператора Самары.

Библиографический список

1. Транспортный оператор Самары / МП г. о. Самара «Центр организации дорожного движения». [Б. м.], 2019. URL: <http://www.tosamara.ru> (дата обращения: 01.10.2020).
2. Агафонов А.А., Сергеев А.В., Чернов А.В. Прогнозирование параметров движения городского пассажирского транспорта по данным спутникового мониторинга // Компьютерная оптика. 2012. Т. 36, № 3. С. 453–458.