



### Литература

1. Кириленко А.Г., Кузнецов Ю.В., Фоминов Д.А. Исследование работы рельсовых цепей на участках с тяжеловесным движением // Автоматика, связь, информатика. 2012. № 10. С. 14 – 17.
2. Шаманов В.И. Динамика асимметрии переменного тягового тока в рельсовых линиях на двухпутных перегонах / Шаманов В.И. // Электротехника. 2016. №10. С. 74 – 79.
3. Бушуев А.В. Рельсовые цепи: теоретические основы и эксплуатация : монография / А.В. Бушуев, В.И. Бушуев, С.В. Бушуев. – Екатеринбург : Изд-во УрГУПС, 2014. – 311 с.
4. Исайчева А.Г. К вопросу мониторинга асимметрии тягового тока / А.Г. Исайчева, В.Г. Волик, М.В. Башаркин И.С. Яшин, А.С. Белоногов // Наука и образование транспорту : материалы XII Международной научно-практической конференции (2019, Самара). Международная научно-практическая конференция «Наука и образование транспорту», 2019 г. Том 1 / редкол.: И. К. Андрончев [и др.] – Самара : СамГУПС, 2019. С. 324 – 326.
5. Исайчева А.Г. Применение информационно-измерительного устройства мониторинга асимметрии тягового тока в рельсовой линии / А.Г. Исайчева, В. Б. Тепляков, М. В. Башаркин, Л. В. Корытин // Вестник транспорта Поволжья. 2019. - № 5 (77). С. 71 – 77.
6. Андрончев И.К. Влияние неисправного состояния сборных токопроводящих стыков на работу рельсовых цепей / И. К. Андрончев, Е. М. Тарасов, А. Г. Исайчева, В. Б. Тепляков, А. А. Булатов // Вестник транспорта Поволжья. 2018. - № 6 (72). С. 40 – 45.

Д.С. Климова

### ПРИМЕНЕНИЕ АЛГОРИТМА RANDOM FOREST В ПРОГНОЗИРОВАНИИ ДОРОЖНОГО ТРАФИКА НА ПРИМЕРЕ ГОРОДА САМАРА

(Самарский университет)

Предоставление информации о загруженности автобусной сети играет важную роль в процессе повышения эффективности системы общественного транспорта и является одной из приоритетных задач в изучении управления пассажирским автотранспортом.

Большинство муниципальных и коммерческих автобусов оснащены системой GPS, что позволяет в режиме реального времени собирать и сохранять данные о времени и местонахождении транспортного средства на маршрутной линии. Эти данные отражают структуру трафика пассажирских перевозок и могут использоваться для различного рода прогностических исследований.



В статье «Online Traffic Congestion Prediction Based on Random Forest»[1] авторы описали свой опыт прогнозирования дорожных заторов на основе GPS-данных. Исследование показало, что применение алгоритма Random Forest для данной цели повышает точность прогнозов, и метод в целом является более стабильным по сравнению с такими алгоритмами, как модель авторегрессии скользящего среднего (ARMA), прогностическая модель нейронной сети, метод опорных векторов и модель непараметрического регрессионного прогнозирования.

В настоящем исследовании рассматривается применение алгоритма Random Forest для прогнозирования времени прибытия автобуса на следующие остановки. В эксперименте используются данные за период с 1 февраля 2020 года по 15 марта 2020 года, предоставляемые Интернет-порталом Транспортного оператора Самары [2]. Данные включают идентификатор GPS, идентификатор маршрута следования, отметку о времени и местоположении автобуса (географические широта и долгота). Записи сохраняются каждые 20 секунд, всего исследуется 62 маршрута.

Обработка данных состоит в нахождении пройденного расстояния, затраченного времени в пути и средней скорости по двум последовательным записям GPS для каждого автобуса. Для нахождения расстояния используется формула гаверсинусов:

$$\Delta\delta = 2\arcsin\left\{\sqrt{\sin^2\left(\frac{\varphi_2 - \varphi_1}{2}\right) + \cos\varphi_1 \cos\varphi_2 \sin^2\left(\frac{\Delta\lambda}{2}\right)}\right\}, (1)$$

где  $\varphi_i$  - широта двух точек в радианах,  $\Delta\lambda$  - разница координат по долготе,  $\Delta\delta$  – угловая разница.

Для перевода углового расстояния в метрическое, нужно угловую разницу умножить на радиус Земли (6372795 метров), единицы конечного расстояния будут равны единицам, в которых выражен радиус (в данном случае - метры) [3].

На рисунке 1 проиллюстрировано распределение средней скорости пассажирского транспорта за выбранный период. Перекос графика влево объясняется дорожными заторами.

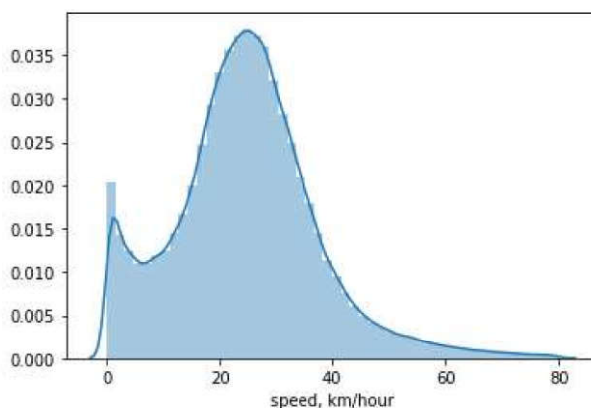


Рис. 1. Распределение средней скорости



После подготовки набора данных осуществляется создание и обучение модели при помощи библиотеки Scikit-learn языка программирования Python. Наилучший результат отмечается при использовании параметра «количество деревьев» равным 1000.

При таком варианте значение средней абсолютной ошибки (mae) равно 25,12. На рисунке 2 изображен график, построенный по набору данных за 2 февраля с полученными прогнозами. График показывает, что модель имеет тенденцию к занижению прогнозируемых величин.

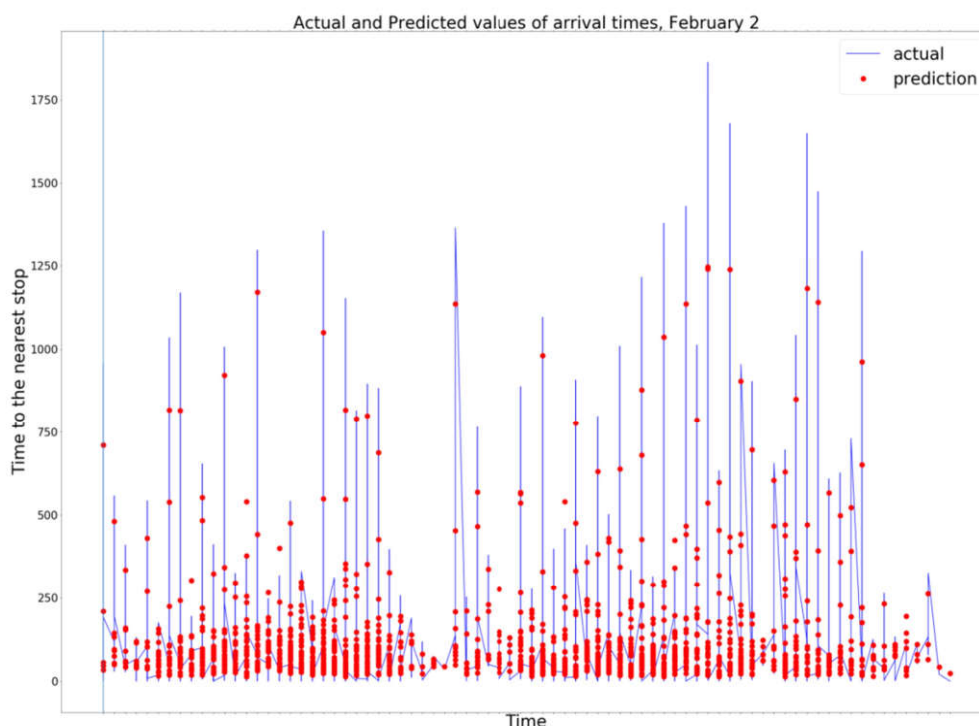


Рис. 2. Полученные прогнозы

Таким образом, реализована модель прогнозирования дорожного трафика пассажирских транспортных средств. В среднем, модель ошибается на 25,12 секунд, что является приемлемым результатом.

### Литература

1. Xiao Han Online Traffic Congestion Prediction Based on Random Forest / Xiao Han, Yijie Shi // 2015 4th International Conference on Mechatronics, Materials, Chemistry and Computer Engineering. - Paris : Atlantis Press, 2015. - P. 2683-2689.
2. Программные интерфейсы (API) // Транспортный оператор Самары URL: <https://tosamara.ru/api/> (дата обращения: 08.04.2020).
3. Learning transportation mode from raw gps data for geographic applications on the web / Yu Zheng, Like Liu, Longhao Wang, Xing Xie // Proceedings of the 17th international conference on World Wide Web. - New York : ACM, 2008. - P. 247-256.