

**Создание статистической модели определения времени прохождения перегона
поездами по реальным данным**

Е. Ю. Бобарико¹

¹Московский физико-технический институт (государственный университет)

В настоящее время во всем мире и в России в частности актуальна проблема оптимизации железнодорожных перевозок. Цель работы – создание модели определения времени прохождения перегона поездом по статистическим данным. Алгоритмы оптимального пропуска поездов оперируют временами прохода поездов по перегонам. В практике железнодорожных перевозок эти времена определяются из тяговых расчётов. Но проанализировав реальные статистические данные по временам прохода поездов разного веса по перегону 25793 (Рабак) - 25800 (Куеда), мы пришли к заключению, что данные времена хода поездов значительно отличаются от теоретических, которые получаются расчётным путём.

Таким образом, чтобы решить сложившуюся проблему необходимо или привлечь статистический аппарат обработки данных, или внести соответствующие поправки в тяговые расчёты на основе статистических данных. В любом случае, для практических целей результатами тяговых расчётов пользоваться нужно аккуратно, они не совпадают с практикой. Иначе перед нами возникает такая ситуация, что, не имея достоверных времён хода поездов (различной массы, с различными начальными скоростями выхода на перегон, с различными ограничениями максимальной скорости на перегоне, с различными локомотивами и локомотивными бригадами), сделать программу оптимального управления движения поездами невозможно.

Перед нами возникает следующая задача: получить достоверную прогнозную модель для определения времён хода различных поездов по перегонам сети ОАО «РЖД».

В распоряжении имеется выборка данных по временам движения поездов по различным перегонам Горьковской железной дороги за 6 месяцев.

Для разработки модели был выбран Matlab. С помощью него мы обратились к машинному обучению, а конкретно к одному из его разделов - регрессионному анализу.

Для построения регрессионного анализа мы использовали четыре основных метода: линейная регрессия (LR), Лассо (LASSO), эластичная сеть (ELASTICNET), бустинг (BOOST).

Так же была произведена фильтрация данных, чтобы избавиться от «грязных данных», которые были получены в результате непредвиденных обстоятельств, случившихся на железной дороге, и не должны влиять на нашу модель.

Для учета всех ситуаций при построении модели было проведено несколько серий экспериментов с различными критериями отбора входных данных. Построена регрессия:

1. для полной статистики
2. на основе статистики остановок на предыдущей и последующих станциях
3. на основе статистики для разных типов поездов
4. для разных времен года
5. для определенного перегона (25793 (Рабак) - 25800 (Куеда))

В результате было получено следующее: наибольшее влияние на модель будут оказывать: остановка на последующей станции (`stop_to`) (50% влияния), тип поезда (`train`) (25% влияния) и остановка на предыдущей станции (`stop_from`) (12,5% влияния); меньшее воздействие оказывают длина поезда (`long_in_loco`) (3,125% влияния) и тип локомотива (`loco`) (3,125% влияния). Именно все вышеуказанные переменные стоит учитывать при создании нашей модели.

Также во всех экспериментах наименьшее среднее квадратичное отклонение между выходными данными и теоретическими наблюдалось у метода BOOST. Именно его необходимо использовать для создания модели определения времени прохождения поездом перегона.

Стоит отметить, что в последнем исследовании мы показали, что ошибка в предсказанном времени отличается от фактического всего на 0,2%. Это очень хороший показатель, особенно, если сравнивать с результатами модели, которую используют на данный момент РЖД, где ошибка составляет 14%.

Результаты данной работы являются необходимой частью более крупного исследования – создания модели определения времени прохождения перегона поездом по реальным данным. Нам удалось определить наиболее эффективный алгоритм для построения модели, а также найти какие независимые переменные дают основной вклад в вариацию зависимой. Полученный результат дает меньше ошибку, чем используемая в настоящее время модель.

Литература

1. *Баранов Л.А., Ерофеев Е.В., Мелёшин И.С., Чинь Л.М.* Оптимизация управления движением поездов: учебное пособие / под ред. д.т.н., проф. Л.А. Баранова. М.: МИИТ, 2011. – 164 с.
2. *Флах П.* Машинное обучение. Наука и искусство построения алгоритмов, которые извлекают знания из данных/ пер. с англ. А.А.Слинкина.-М.: ДМК Пресс, 2015. - 400с.:ил.