

Использование статистических методов и машинного обучения в задаче исследования и анализа данных автомобильных аварий на автомагистрали

Александра Малькова

Научные руководители J. Arbel, M.L. Delle Monache, V. Strijov

МФТИ, ФУПМ,
Кафедра интеллектуальных систем

17 апреля 2019 г.

Цель работы

Распознать паттерны возникновения пробок в результате дорожно-транспортных происшествий

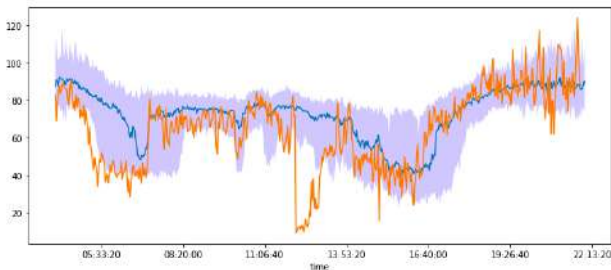
Проблема

Данные о потоке зашумлены, ненадёжны или частично отсутствуют

Исследование влияния аварий на распределение средней скорости в течение суток

Пробки и заторы являются проблемами на дороге. Главные причины:

- Часы-пик
- Аварии
- Погодные условия, и т.д.



Медианное значение скорости за все понедельники 2016г. (голубая кривая), и распределение скорости в течении дня с аварией (оранжевая кривая). В случае аварии скорость падает в среднем.

Поток – количество транспортных средств (ТС), проходящих через поперечное сечение дороги в определенном направлении в единицу времени.

Плотность потока – количество ТС, проходящих единицу расстояния в определённом направлении.

Критическая плотность – плотность, при которой движение меняется со свободного (нет взаимодействия между ТС) на перегруженное.

Пропускная способность – максимальный поток для данного дня.

Главными переменными трафика являются плотность ρ (veh/km), поток q (veh/h), пропускная способность q_c (veh/h), скорость v (km/h).

- 1 M. J. Lighthill and G. B. Whitham. *On kinematic waves II. //A theory of traffic flow on long crowded roads*, 1955
- 2 P. Richards. *Shock waves on the highway*, 1956
- 3 N. Polson and V. Sokolov. *Bayesian analysis of traffic flow on interstate I-55. // Annals of Applied Statistics*, 2014
- 4 Morbidi, F. Ojeda, L. L. Kibangou, A. Y. Bellicot, I. De Wit, C. C. and P. Bellemain. *Grenoble traffic lab: An experimental platform for advanced traffic monitoring and forecasting. //IEEE Control Systems*, 35(3):23–39, 2015.

Постановка задачи

Дано

Пятёрка $(s_i, t_i, v_i, \rho_i, q_i)$, где для каждого сенсора s_i и каждого момента времени t_i известны значения основных переменных трафика; тройка $(t_j^{\text{start}}, t_j^{\text{end}}, x_j)$, j - номер события, где известны время начала, время конца и место аварии.

Требуется:

- Исследовать параметры $\psi = (q_c, \rho_c, \rho_{\max})$ модели транспортного потока за день при наличии и отсутствии аварии.
- Построить модель, определяющую пропускную способность по причине аварии.

- Метод главных компонент (PCA)
- Свёрточная нейронная сеть (CNN)
- Многочастотный фильтр (Particle filtering)

Модель потока Lighthill-Whitham-Richards (LWR)

Уравнение непрерывности^{1,2}:

$$\begin{cases} \partial_t \rho(t, x) + \partial_x q(\rho(t, x)) = 0, \\ \rho(0, x) = \rho_0(x) \end{cases}$$

где $(t, x) \in \mathbb{R}^+ \times \mathbb{R}$.

Функция потока

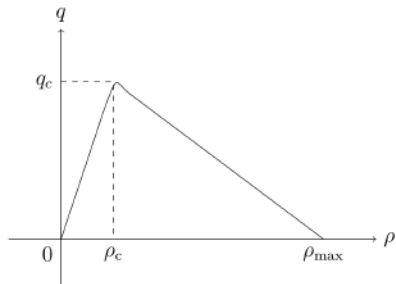
$$q(\rho) = \rho \cdot v$$

¹M. J. Lighthill and G. B. Whitham. On kinematic waves ii. a theory of traffic flow on long crowded roads, 1955

²P. Richards. Shock waves on the highway., 1956

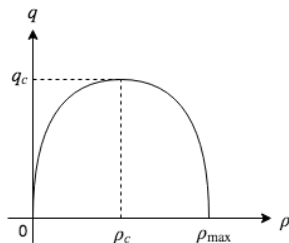
Фундаментальная диаграмма

Набор $\psi = (q_c, \rho_c, \rho_{\max})$ является параметрами модели.



Треугольная ФД,

$$q(\rho) = \begin{cases} \frac{q_c}{\rho_c} \cdot \rho, & \rho < \rho_c, \\ q_c \cdot \frac{\rho_{\max} - \rho}{\rho_{\max} - \rho_c}, & \rho \geq \rho_c. \end{cases}$$



ФД Гриншилда,

$$q(\rho) = \rho \cdot \left(1 - \frac{\rho}{\rho_{\max}}\right) \cdot 4 \cdot \frac{q_c}{\rho_{\max}}.$$

Решающий алгоритм Римана задачи измерения параметров модели

ρ_0^n	ρ_1^n	ρ_{M-1}^n	ρ_M^n
------------	------------	-----	-----	----------------	------------

Сегмент дороги и связанные значения плотности

При данных $\rho_l, \rho_r \in [0, \rho_{\max}]$, представим задачу Римана для уравнения неразрывности с начальными условиями

$$\rho(0, x) = \begin{cases} \rho_l, & \text{если } x < 0, \\ \rho_r, & \text{если } x > 0, \end{cases}$$

где ρ_l и ρ_r – значения плотности потока слева и справа от разрыва.

Условие Rankine-Hugoniot о прыжке для транспортного потока:

$$\lambda = \frac{q(\rho_r) - q(\rho_l)}{\rho_r - \rho_l}$$

Если $\rho_l < \rho_r$, то шоковая волна:

$$\rho(t, x) = \begin{cases} \rho_l, & \text{if } x < \lambda \cdot t, \\ \rho_r, & \text{if } x > \lambda \cdot t. \end{cases}$$

Если $\rho_l > \rho_r$, что равносильно $q'(\rho_l) < q'(\rho_r)$, то волна разрежения:

$$\rho(t, x) = \begin{cases} \rho_l, & \text{if } x < q'(\rho_l) \cdot t, \\ (q')^{-1}\left(\frac{x}{t}\right), & \text{if } q'(\rho_l) \cdot t < x < q'(\rho_r) \cdot t, \\ \rho_r, & \text{if } x > q'(\rho_r) \cdot t. \end{cases}$$

Рекурсивно вычисляет значения плотности потока в ячейках участка дороги:

$$\rho_i^{n+1} = \rho_i^n + \frac{\tau}{s}(g(\rho_{i-1}^n, \rho_i^n) - g(\rho_i^n, \rho_{i+1}^n)).$$

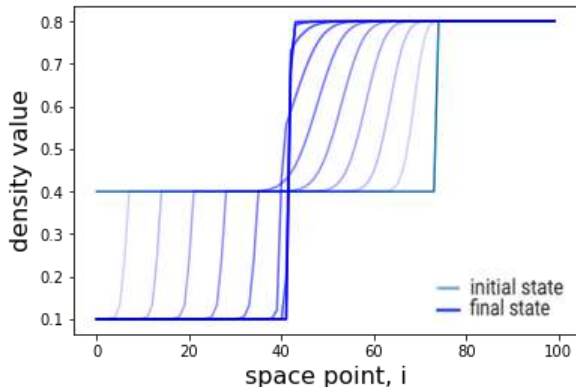
Значение функции g вычисляется следующим образом:

$$g(\rho_l, \rho_r) = \begin{cases} q(\rho_c), & \rho_r \leq \rho_c \leq \rho_l, \\ q(\rho_l), & \rho_r \leq \rho_l \leq \rho_c, \\ q(\rho_r), & \rho_c \leq \rho_r < \rho_l, \\ \min(q(\rho_l), q(\rho_r)), & \rho_l < \rho_r. \end{cases}$$

Граничные условия для предсказания следующего состояния:

$$\begin{aligned} \rho_0^{n+1} &= \rho_0^n + \frac{\tau}{s}(f^{\text{in}} - g(\rho_0^n, \rho_1^n)), \\ \rho_M^{n+1} &= \rho_M^n + \frac{\tau}{s}(g(\rho_{M-1}^n, \rho_M^n) - f^{\text{out}}). \end{aligned}$$

Пример численного моделирования плотности потока



Заданы

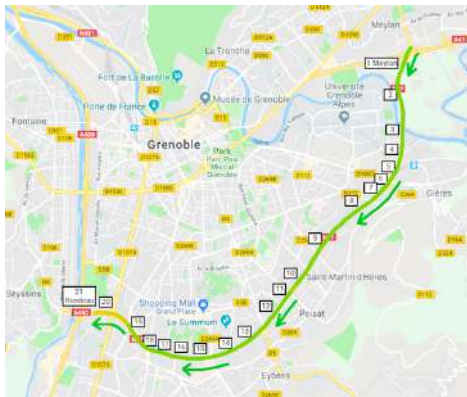
$l = 100$ ячеек,
 $f^{\text{in}}, f^{\text{out}} = 0.1$ ТС.

Начальное условие ρ^0 :

$$\rho_i^0 = \begin{cases} 0.4, & i < 75 \\ 0.8, & i \geq 75 \end{cases}$$

Последовательное вычисление вектора
плотности потока

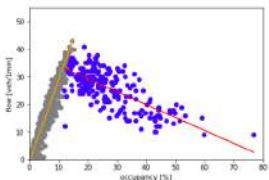
Цель эксперимента: найти параметры модели и выявить закономерности их изменения из-за аварий



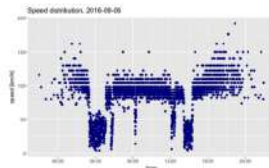
- каждые 15s
- все сенсоры (66)
- 2014-2017 гг.



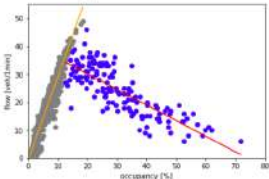
Сравнение фундаментальных диаграмм и графиков распределения скоростей для дней с аварией и без неё



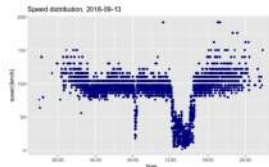
(a) with accident



(b) with accident



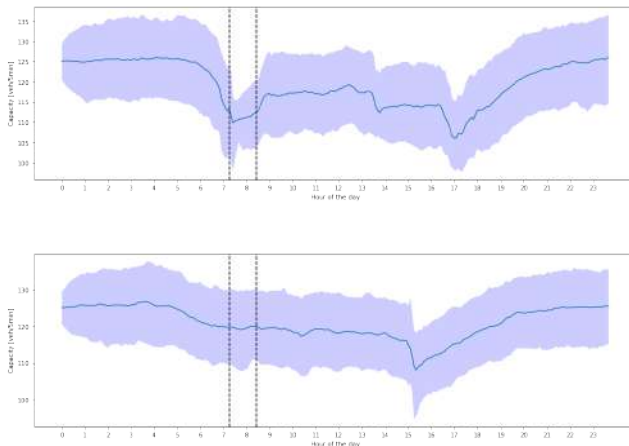
(c) without accident



(d) without accident

Слева: ФД. Справа: Скорость VS время. Первая строка: День с аварией. Вторая строка: День без аварии. Небольшое отличие в параметрах ФД, в то время как на графике распределения скорости видно падение значения.

Сравнение пропускной способности для дней с аварией и без



Полученная с помощью многочастотного фильтра пропускная способность в течение дня. *Сверху*: День с аварией. *Снизу*: День без аварии.

- Проанализированы данные сенсоров о потоке и данные про аварии.
- Была численно решена задача Римана с использованием схемы Годунова для уравнения модели LWR.
- Был реализован алгоритм нахождения пропускной способности транспортного потока с помощью многочастотного фильтра и улучшен с помощью вычисления фактора Байеса.

В нашем случае алгоритм гарантировал, что фундаментальная диаграмма (ФД) имеет треугольную форму и показал распознавание падения пропускной способности во время аварии.