Моделирование количества подвижных единиц грузового поезда в сходе с рельсов на основе метода максимального правдоподобия

Работу выполнил: Королёв Егор Владимирович, студент группы М8О-401Б-18
Научный руководитель: Игнатов Алексей Николаевич, к.ф.-м.н., доцент

Московский авиационный институт (НИУ)

14 декабря 2021 г.

Введение

Прогресс написания ВКР

1. написано 50 страниц;
2. приведен предварительный анализ данных;
3. построено 96 регрессионных моделей;
4. написана программная реализация ММП;

Введение

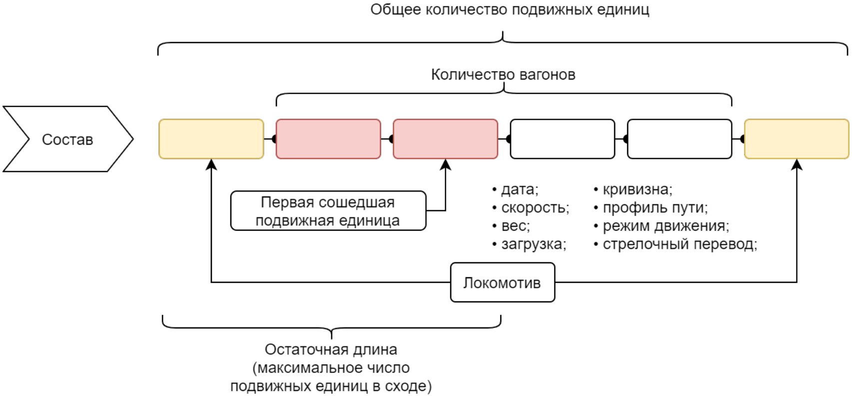
Проблема

В странах с большой ж/д сетью и большим потоком перемещения поездов существует проблема схода составов с рельс. Последствия схода могут привести к экологическим, экономическим и логистическим проблемам. В РФ в среднем происходит 1 авария каждые 4 дня. Поэтому проблема представляет интерес для железнодорожных компаний.

Задача

Для заданного набора данных необходимо построить предсказательную модель количества сошедших подвижных единиц в составе.

Описание признаков

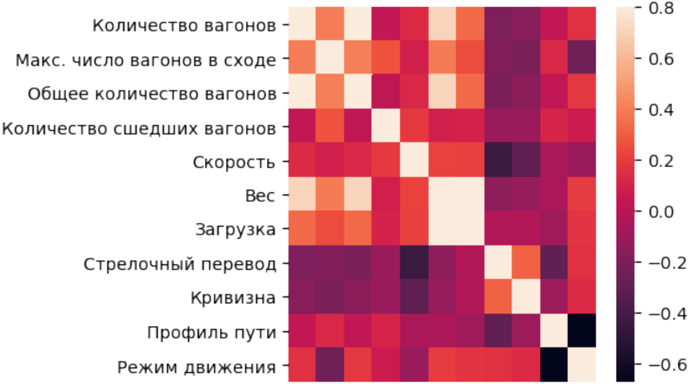


Описание признаков

Разреженность данных

- мощность выборки $n = 56$;
- признак 'Режим движения' имеет 23 пропусков (41%);
- признак 'Профиль пути' имеет 12 пропусков (21%);
- признак 'Кривизна' имеет 10 пропусков (17%);

Корреляция признаков



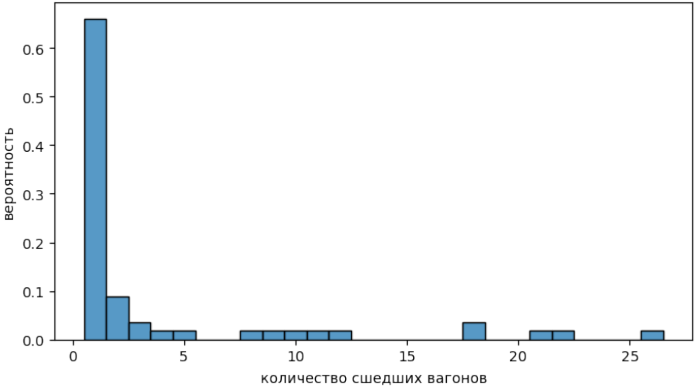
Конструирование признаков

Введение новых признаков

- $f_1 =$ профиль пути $\cdot$ макс. число вагонов в сходе;
- $f_2 = 1 - \frac{\text{макс. число вагонов в сходе}}{\text{общее кол-во вагонов}}$;
- $f_3 =$ скорость $\cdot$ загрузка;

	target	f_1	f_2	f_3
target	1.0	0.101375	-0.286535	0.198847
f_1	0.101375	1.0	-0.086693	-0.228508
f_2	-0.286535	-0.086693	1.0	-0.124420
f_3	0.198847	-0.228508	-0.124420	1.0

Оценка функции вероятности $P(\xi = x)$



Пуассоновская регрессия

Функция вероятности

$$P(\xi = k) = \frac{\lambda^k}{k!} e^{-\lambda}$$

Функция логарифмического правдоподобия

$$\log(L(\theta, x, y)) = \sum_{i=1}^{n} (-\lambda(x_i, \theta) + y_i \ln(\lambda(x_i, \theta)) - \ln(y_i!))$$

Определение функций $\lambda(\theta, x)$

1. $\lambda_1(\theta, x) = e^{(\theta \cdot x)}$;
2. $\lambda_2(\theta, x) = e^{-(\theta \cdot x)^2}$;
3. $\lambda_3(\theta, x) = \sqrt{|5^2 - ((\theta \cdot x) - 5)^2|} + 1$;
4. $\lambda_4(\theta, x) = ((\theta \cdot x) - 1)^2$;
5. $\lambda_5(\theta, x) = \frac{1}{1+(\theta \cdot x)^2}$;
6. $\lambda_6(\theta, x) = (\theta \cdot x)(\frac{\pi}{2} + \arctan(\theta \cdot x)) + 1$;
7. $\lambda_7(\theta, x) = \log(1 + (\theta \cdot x)^2) + 1$.

Геометрическая регрессия

Функция вероятности

$$P(\xi = k) = (1 - p)^k p$$

Функция логарифмического правдоподобия

$$\log(L(\theta, x, y)) = \sum_{i=1}^{n} (y_i \ln(1 - p(x_i, \theta)) + \ln(p(x_i, \theta)))$$

Определение функций $p(\theta, x)$

1. $p_1(\theta, x) = e^{(\theta \cdot x)}$;
2. $p_2(\theta, x) = e^{-(\theta \cdot x)^2}$;
3. $p_3(\theta, x) = \frac{1}{1+e^{-(\theta \cdot x)}}$;
4. $p_4(\theta, x) = \frac{1}{1+(\theta \cdot x)^2}$;
5. $p_5(\theta, x) = (\theta \cdot x)(\frac{\pi}{2} + \arctan(\theta \cdot x)) + 1$.

Признаковые пространства

1. $features_1$: (кривизна);
2. $features_2$: (кривизна, профиль пути);
3. $features_3$: (кривизна, профиль пути $\cdot$ макс. число вагонов в сходе);
4. $features_4$: (кривизна, $1 - \frac{\text{макс. число вагонов в сходе}}{\text{общее кол-во вагонов}}$);
5. $features_5$: (кривизна, профиль пути, скорость $\cdot$ загрузка);
6. $features_6$: (кривизна, профиль пути, скорость $\cdot$ загрузка, $1 - \frac{\text{макс. число вагонов в сходе}}{\text{общее кол-во вагонов}}$);
7. $features_7$: (кривизна, скорость $\cdot$ загрузка, $1 - \frac{\text{макс. число вагонов в сходе}}{\text{общее кол-во вагонов}}$);
8. $features_8$: (скорость $\cdot$ загрузка, $1 - \frac{\text{макс. число вагонов в сходе}}{\text{общее кол-во вагонов}}$).

Программная реализация ММП

Конструктор класса

```
def MLM(log_likelihood_fun,
          const_log_likelihood_fun,
          optim_method, borders,
          predict_fun, param_fun,
          features, target, df)
```

Численный эксперимент

Пуассоновская регрессия

- наилучшая модель: $(\lambda_1, features_7)$.
  $AIC_c = 400.13, \quad \hat{\theta} = (0.92, -105.35, 34.34, 0.02, -2.24)$;
- диапазон значений: $AIC_c \sim 400.13 \div 617.6$;

Геометрическая регрессия

- наилучшая модель: $(p_4, features_7)$.
  $AIC_c = 153.79, \quad \hat{\theta} = (-0.68, 89.66, -113.05, -0.05, 3.31)$;
- диапазон значений: $AIC_c \sim 153.79 \div 352.91$;

Выводы

В данной работе были построены предсказательные модели числа сошедших вагонов. Глобально их можно разделить на 2 класса: модели Пуассоновской регрессии и модели геометрической регрессии. Для каждого класса были рассмотрены различные параметрические виды и признаковые пространства.

Список литературы

Замышляев А.М., Игнатов А.Н., Кибзун А.И., Новожилов Е.О. Функциональная зависимость между количеством вагонов в сходе из-за неисправностей вагонов или пути и факторами движения // Надежность. 2018. Т. 18, № 1. С. . . . . DOI: 10.21683/1729-2646-2018-18-1. . .

Спасибо за внимание!