

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
кафедры 804 МАИ

Московский авиационный институт (НИУ)

1. Введение

Проблема

Периодически на ЖД путях происходят сходы или крушения. Последствия схода могут привести к экологическим, экономическим и логистическим проблемам. В РФ в среднем происходит сход или крушение раз в 5.5 дней.

Задача

По имеющемуся набору данных необходимо построить предсказательную модель количества сошедших подвижных единиц в составе.

2. Введение

Причина схода

Самая частая причина схода – излом боковой рамы



3. Введение



4. Введение

Некоторые обозначения

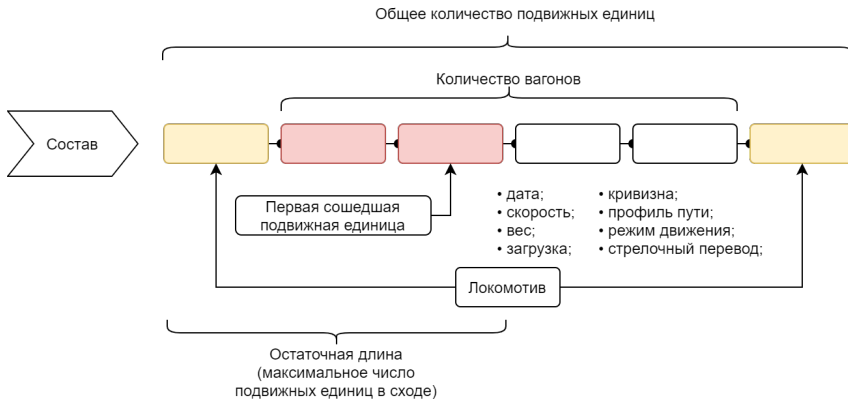
- n – мощность выборки;
- r – количество параметров в обучаемой модели;
- ξ – СВ, характеризующая количество подвижных единиц в сходе;
- $\eta = \xi - 1$;
- θ – вектор обучаемых параметров;
- $AIC_c = 2r - 2 \ln(L) + \frac{2r^2 + 2r}{n - r - 1}$ – скорректированный критерий Акаике;
- L – значение функции правдоподобия для обученной модели.

5. Введение

Некоторые обозначения

- x_i – вектор признаков i -го происхождения;
- y_i – количество подвижных единиц в сходе i -го происхождения;
- L_{best} – значение функции правдоподобия у наилучшей модели по AIC_c ;
- $L_{trivial}$ – значение функции правдоподобия для тривиальной модели.

1. Описание признаков



2. 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%);

3. Корреляция признаков



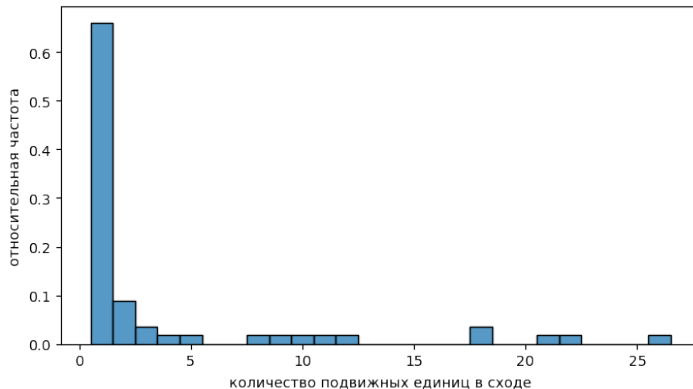
4. Конструирование признаков

Введение новых признаков

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

	target	f_1	f_2	f_3
target	1.0	0.101	-0.286	0.198
f_1	0.101	1.0	-0.086	-0.228
f_2	-0.286	-0.086	1.0	-0.124
f_3	0.198	-0.228	-0.124	1.0

5. Оценка функции вероятности $P(\xi = l)$



1. Пуассоновская регрессия

Функция вероятности

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

Функция логарифмического правдоподобия

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

где $x = \text{col}(x_1, x_2, \dots, x_n)$, $y = \text{col}(y_1, y_2, \dots, y_n)$.

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

$$1 \quad \lambda_1(\theta, x) = e^{\langle \theta, x \rangle};$$

$$2 \quad \lambda_2(\theta, x) = e^{-(\langle \theta, x \rangle)^2};$$

$$3 \quad \lambda_3(\theta, x) = \sqrt{|5^2 - (\langle \theta, x \rangle - 5)^2|} + 1;$$

$$4 \quad \lambda_4(\theta, x) = (\langle \theta, x \rangle - 1)^2;$$

$$5 \quad \lambda_5(\theta, x) = \frac{1}{1 + (\langle \theta, x \rangle)^2};$$

$$6 \quad \lambda_6(\theta, x) = \langle \theta, x \rangle \left(\frac{\pi}{2} + \arctan(\langle \theta, x \rangle) \right) + 1;$$

$$7 \quad \lambda_7(\theta, x) = \ln(1 + (\langle \theta, x \rangle)^2) + 1.$$

3. Геометрическая регрессия

Функция вероятности

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

Функция логарифмического правдоподобия

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

где $x = \text{col}(x_1, x_2, \dots, x_n)$, $y = \text{col}(y_1, y_2, \dots, y_n)$.

4. Определение функций $p(\theta, x)$

$$1 \quad p_1(\theta, x) = e^{\langle \theta, x \rangle};$$

$$2 \quad p_2(\theta, x) = e^{-(\langle \theta, x \rangle)^2};$$

$$3 \quad p_3(\theta, x) = \frac{1}{1 + e^{-\langle \theta, x \rangle}};$$

$$4 \quad p_4(\theta, x) = \frac{1}{1 + (\langle \theta, x \rangle)^2};$$

$$5 \quad p_5(\theta, x) = \langle \theta, x \rangle \left(\frac{\pi}{2} + \arctan(\langle \theta, x \rangle) \right) + 1.$$

5. Признаковые пространства

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

1. Программная реализация ММП

Конструктор класса

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

2. Численный эксперимент

Пуассоновская регрессия

- наилучшая модель (по признаковым пространствам и функциям связи): $(\lambda_1, features_6)$.
 $AIC_c = 356.87$, $\hat{\theta} = (0.92, -105.33, 34.34, 0.02, -2.24)$, $\frac{\ln(L_{best})}{\ln(L_{trivial})} = 0.62$;
- $AIC_c \in [356.87, 567.74]$;
- в условиях (а) $\arg \max_k \hat{P}(\xi = k) = 2$.

Геометрическая регрессия

- наилучшая модель: $(p_4, features_6)$. $AIC_c = 153.65$, $\hat{\theta} = (-2.01, 29.36, 226.20, -0.03, 3.85)$,
 $\frac{\ln(L_{best})}{\ln(L_{trivial})} = 0.57$;
- $AIC_c \in [153.65, 918.43]$;
- в условиях (а) $\arg \max_k \tilde{P}(\xi = k) = 1$.

(а). профиль пути = 0.001555, кривизна = 0.001875, скорость = 67, загрузка = 0.87, макс. число вагонов в сходе = 21, общее кол-во вагонов = 67

1. Демонстрация работы

Метод максимального правдоподобия

Войти или зарегистрироваться

Датасет (текущая страница)


Целевой признак

Признаковые пространства

Функции связи

Оптимизация

Вычисление оценок



Перетащите файл сюда
или

Выберите файл

Загрузите файл с расширением csv не более 50 МБ

Далее

2. Демонстрация работы

Метод максимального правдоподобия

Войти или зарегистрироваться

Датасет

Целевой признак (текущая страница)

Признаковые пространства

Функции связи

Оптимизация

Вычисление оценок

Целевой признак:



Дата

Распределение:

Пуассоновское

Далее

3. Демонстрация работы

Метод максимального правдоподобия

[Войти или зарегистрироваться](#)

Датасет

[Целевой признак](#)[Признаковые пространства \(текущая страница\)](#)[Функции связи](#)[Оптимизация](#)[Вычисление оценок](#)

Добавленные признаковые пространства:



first:

11 13 14 15 16

second:

13 15 16

[Далее](#)

Новое признаковое пространство:

1. ☐ Дата
2. ☐ Количество вагонов
3. ☐ Макс. число вагонов в сходе
4. ☐ Общее количество вагонов
5. ☐ Количество сшедших вагонов
6. ☐ Скорость
7. ☐ Вес
8. ☐ Загрузка
9. ☐ Стрелочный перевод
10. ☐ Кривизна
11. ☐ Профиль пути
12. ☐ Режим движения
13. ☐ n_feature_1
14. ☐ n_feature_2
15. ☐ n_feature_3
16. ☐ intercept

[Добавить](#)

4. Демонстрация работы

Метод максимального правдоподобия

[Войти или зарегистрироваться](#)

Датасет

Целевой признак

Признаковые пространства

Функции связи (текущая страница)

Оптимизация

Вычисление оценок

Функции связи:



```
link_funs = [  
    exp(t),  
    exp(-t * t),  
]  
  
log_link_funs = [  
    t,  
    -t * t,  
]
```

[Далее](#)

5. Демонстрация работы

Метод максимального правдоподобия

Войти или зарегистрироваться

Датасет

Целевой признак

Признаковые пространства

Функции связи

Оптимизация (текущая страница)

Вычисление оценок

Метод оптимизации: ⓘ

SHGO

Далее

6. Демонстрация работы

Метод максимального правдоподобия

Войти или зарегистрироваться

Датасет

Целевой признак

Признаковые пространства

Функции связи

Оптимизация

Вычисление оценок (текущая страница)

Идут вычисления. Пожалуйста, подождите.



7. Демонстрация работы

Метод максимального правдоподобия

[Войти или зарегистрироваться](#)[Датасет](#)[Целевой признак](#)[Признаковые пространства](#)[Функции связи](#)[Оптимизация](#)[Вычисление оценок \(текущая страница\)](#)100%

Скачать в формате:

[tex](#)[json](#)[txt](#)[pdf](#)

8. Схема приложения




Заключение

Были построены предсказательные модели числа сошедших подвижных единиц. Их можно разделить на 2 класса: модели пуассоновской регрессии и модели геом. регрессии. Для каждого класса были рассмотрены различные функции связи и признаковые пространства. Были проведены численные эксперименты.

Был написан веб-сервис, реализующий метод максимального правдоподобия. Сервис позволяет задать собственный набор данных, целевой признак и его распределение, указать признаковые пространства, функции связи, а также выбрать один из методов оптимизации. Скачать результаты вычислений можно в нескольких форматах.

Спасибо за внимание!

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. – С.1-15.