Time Series Analytics - Practice and Research

Department of Computer Science & Department of Statistics

Prof. Dr. Emmanuel Müller,
Jan Aust and Nadine Ott

fakultät für informatik

fakultät statistik

technische universität dortmund

Business Case

Objective

- a retailer plans to restock its inventory every other week and only keep in stock the items that it has actually sold during that period
- create machine learning model to predict the demand
  - based on a six months training data
  - for every product
  - over a two-week period
- some products will be promoted for a limited period of time
- no response to price changes

Special Challenges of the Task

- unusual $\rightarrow$ promotion feature is missing for the training period
- orders not already aggregated and predict the next 14 days
- unbalanced evaluation metric:

$$dmcscore(y, p, w) = \begin{cases} pw & \text{for } y \geq p \\ (\frac{8}{5}y - \frac{3}{5}p)w & \text{sonst} \end{cases}$$

In this case y is the real value, p is the value of the prediction, w is the simulation price of the itemID

First Machine Learning Pipeline

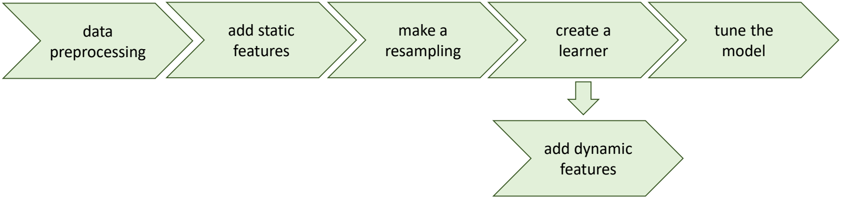
Prepocessing
- create promotion feature
- aggregate orders by two weeks

Features
- log-transformation of the target value
- dynamic and static features

XGBoost Regressor
- custom objective function
- predict demand of the next biweek

Tuning
- Bayesian optimization algorithm
- parallelize

Second Machine Learning Pipeline

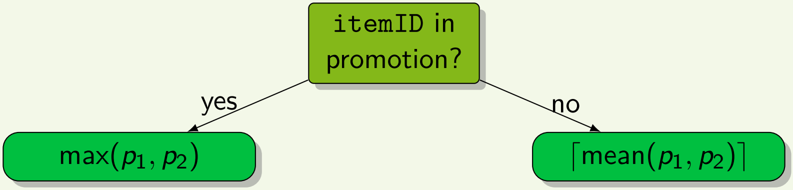


mlr mlr3 v0.3.0

Final Model

We combined 2 XGBoost models because the single models tend to underestimate the demand

Max-Model



p_1 is the prediction aggregated per biweek and p_2 is the prediction aggregated per day

Bachelor thesis - Online anomaly detection - Problem

PhD-Student Bin Li - Department of Computer Science Chair IX

Sensor 1	30°	25°	25°	...	...	...	...	...	...	...	...	...	...	...
Sensor 2	40°	30°	30°	...	...	...	...	...	...	...	...	...	...	...
Sensor 3	50°	45°	45°	...	...	...	...	...	...	...	...	...	...	...
Sensor 4	25°	20°	20°	...	...	...	...	...	...	...	...	...	...	...

time →

Bachelor thesis - Online anomaly detection - Problem

Sensor 1	30°	25°	25°	20°	20°	20°	...	...	...	...	...	...	...	...
Sensor 2	40°	30°	30°	25°	25°	25°	...	...	...	...	...	...	...	...
Sensor 3	50°	45°	45°	40°	40°	40°	...	...	...	...	...	...	...	...
Sensor 4	25°	20°	20°	15°	15°	5°	...	...	...	...	...	...	...	...

time →

Bachelor thesis - Online anomaly detection - Problem

Sensor 1	30°	25°	25°	20°	20°	20°	15°	15°	15°	...	...	...	...	...
Sensor 2	40°	30°	30°	25°	25°	25°	25°	25°	25°	...	...	...	...	...
Sensor 3	50°	45°	45°	40°	40°	40°	40°	40°	40°	...	...	...	...	...
Sensor 4	25°	20°	20°	15°	15°	5°	10°	10°	10°	...	...	...	...	...

time

Bachelor thesis - Online anomaly detection - Problem

Sensor 1	30°	25°	25°	20°	20°	20°	15°	15°	15°	10°	10°	10°	...	...
Sensor 2	40°	30°	30°	25°	25°	25°	25°	25°	25°	20°	20°	20°	...	...
Sensor 3	50°	45°	45°	40°	40°	40°	40°	40°	40°	36°	38°	36°	...	...
Sensor 4	25°	20°	20°	15°	15°	5°	10°	10°	10°	5°	5°	5°	...	...

time →

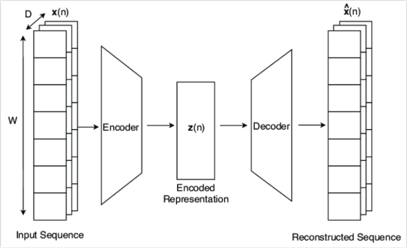
Bachelor thesis - Online anomaly detection - Problem

Concept Drift

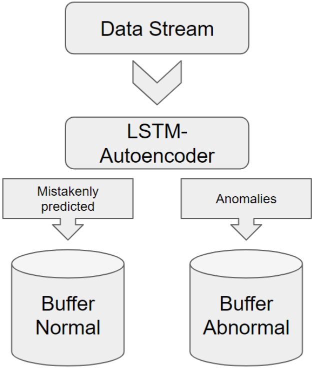
Sensor 1	30°	25°	25°	20°	20°	20°	15°	15°	15°	10°	10°	0°	...	...
Sensor 2	40°	30°	30°	25°	25°	25°	25°	25°	25°	20°	20°	20°	...	...
Sensor 3	50°	45°	45°	40°	40°	40°	40°	40°	40°	36°	38°	36°	...	...
Sensor 4	25°	20°	20°	15°	15°	5°	10°	10°	10°	5°	5°	5°	...	...

time

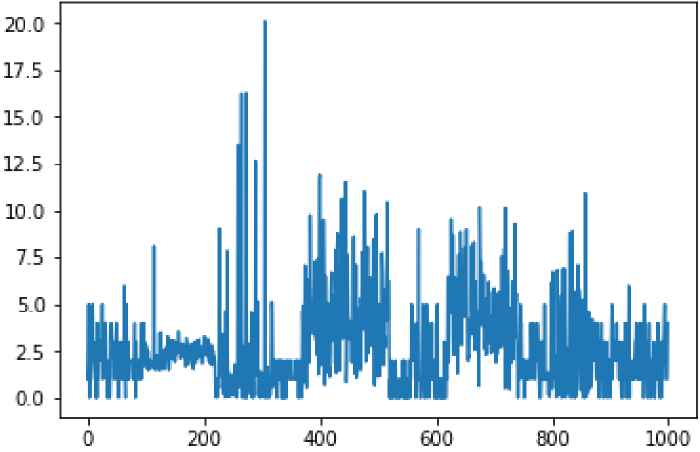
Bachelor thesis - Online anomaly detection - Solution



Trinh, Hoang Duy & Zeydan, Engin & Giupponi, L. & Dini, Paolo. (2019). Detecting Mobile Traffic Anomalies through Physical Control Channel Fingerprinting: a Deep Semi-supervised Approach. IEEE Access. PP. 1-1. 10.1109/ACCESS.2019.2947742.



Research project - Change point detection



Research project - Change point detection

- together with Erik Scharwächter (Department of Computer Science)
- evaluate existing algorithms for change point detection in consideration of the location of the change points
- H1: The relative position of the change points has an effect on the difficulty of the detection problem
- if we can reject H0, we may be able to better estimate in which scenarios which algorithm works well