Enhancing Product Demand Forecasting Accuracy through Clustering

Problem definition

The current industry standard employs statistical models to facilitate demand prediction, often framed as a time series forecasting task. With the rapid advancement of machine learning (ML), there exists a significant opportunity to enhance accuracy through the application of ML models in time series forecasting. Predicting product demand becomes particularly challenging when dealing with a large number of products exhibiting diverse characteristics. A simple first approach to integrating ML models into time series forecasting is utilizing a global ML model, which incorporates historical demand data of all forecastable products. However, these products may exhibit distinct behavioral patterns, thereby limiting the applicability of such a model. One way of mitigating this problem is integrating an auxiliary clustering-based model to categorize products into distinct classes to facilitate the consecutive application of appropriate ML models for time series forecasting. Notably, this approach has been successfully demonstrated in the banking industry with time series data (Bandara et al., 2020), showcasing an improved performance of Long Short-Term Memory (LSTM) models through clustering.

This thesis aims to apply the aforementioned clustering approach to a different industry while extending its applicability to various ML models. The study will utilize the fifth M-Competition (M5) dataset, comprising 100,000 real-life time series derived from Walmart retail sales data, which presents a suitable foundation for this investigation. This thesis will implement a 2-stage model for product demand prediction. Initially, the dataset will be partitioned into multiple classes via clustering techniques, followed by the training of suitable ML models to forecast time-series demand. Various architectures, including LightGBM and N-beats, will be examined. Subsequently, the overall forecast accuracy will be assessed and compared with the performance of a global model.

Aims and scope of the thesis

It is subject of this thesis to evaluate the effectiveness of applying clustering approach to product demand forecasting. This comprises the following research tasks:

- Literature review on time series forecasting algorithms and clustering methods with a special focus on ML applications
- Development and implementation of different ML demand prediction model and previous clustering methods
- Creation of global model as benchmark sets and comparison of the developed clustering model with the benchmark

Related Research

- Bandara, K., Bergmeir, C., Smyl, S. (2020). Forecasting across time series databases using recurrent neural networks on groups of similar series: A clustering approach. Expert Systems with Applications, 140, 112896. https://doi.org/10.1016/j.eswa.2019.112896
- Bontempi, G., Ben Taieb, S., Le Borgne, YA. (2013). Machine Learning Strategies for Time Series Forecasting. In: Aufaure, MA., Zimányi, E. (eds) Business Intelligence. eBISS 2012. Lecture Notes in Business Information Processing, vol 138. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-36318-4_3
- Parmezan, A. R. S., Souza, V. M. A., Batista, G. E. A. P. A. (2019). Evaluation of statistical and machine learning models for time series prediction: Identifying the state-of-the-art and the best conditions for the use of each model. Information Sciences, 484, 302–337. https://doi.org/10.1016/j.ins.2019.01.076

Begin: March 1st, 2024

Advisor: Christina Johanna Liepold