On the Relevance of Demand Pattern Categorization

Florian Grimm
Reutlingen University
florian.grimm@reutlingen-university.de

Clemens van Dinther
Karlsruhe Institute of Technology
clemens.dinther2@kit.edu

Tim Straub
Reutlingen University
tim.straub@reutlingen-university.de

Günter Bitsch
Reutlingen University
guenter.bitsch@reutlingen-university.de

Abstract

The application of transfer learning to predict sales demand is an emerging topic that has been attracting more and more attention recently. However, the selection of data to be used for the learning process is not trivial. Data resources are usually scarce and often anonymized to a certain extent, so their usability for successful training is not guaranteed. One solution is to use already developed categorization schemes that group time series based on certain calculated parameters, but the derived categories do not necessarily capture the process of time series formation. This research addresses the question of whether categorization schemes are beneficial for transfer learning approaches by conducting an experiment in which Syntetos', Boylan's and Croston's categorization scheme is used in combination with two deep learning architectures for the transfer learning process. The results show that similar patterns are indeed beneficial for prediction, but that models using all available data perform quite similarly.

Keywords: demand pattern, forecasting, time series categorization, transfer/cross learning, sales prediction.

1. Introduction

Predicting sales demand is a challenging task that companies and business-to-business retailers devote enormous resources to researching. Not only does it influence a company's production planning, inventory control or marketing decisions (Kerkkänen et al., 2009, Ma et al., 2016), it also gives them decisive advantages over competitors when carried out accurately (Seyedan and Mafakheri, 2020). It contributes to lower storing costs (Carbonneau et al., 2008) and leads to better planning reliability. The range of prediction algorithms already developed is huge (Fildes et al., 2022) and research on this topic is still ongoing. However, precise forecasts require high computing capacities and a great amount of detailed knowledge about the prediction area (Ingle et al., 2021), and many companies do not have the resources to handle that.

To overcome these limitations, there are several ways to tackle the problem, a selection of which is outlined here. From a computational point of view, a solution is to concept more efficient algorithms that require fewer resources and enable faster predictions. Another option is the design of more powerful hardware. These are valid approaches, but it is very unlikely that companies will develop artificial intelligence (AI) capabilities or their own deep learning devices from scratch overnight. Another approach is to outsource the problem by hiring external service providers. However, this option involves high costs and not all companies feel comfortable handing over their data. A more sensible way would be the use of generic models that can handle different domains without prior knowledge, where companies can simply apply existing algorithms to their problem, but as mentioned earlier, this is not so straightforward (Ingle et al., 2021).

Nonetheless, there are attempts to develop generic approaches for specific domains (e.g. sales forecasting), that have explored different ways of dealing with the problem and have been successful in vastly improving predictions. This research distinguishes between the following groups: *time series categorization*, *model recommendation*, *and transfer or cross learning*.

Time Series Categorization attempts to cluster time series that exhibit similar structural behavior. The aim is to find parameters that represent the time series, and boundaries that divide them into

URI: https://hdl.handle.net/10125/109017
978-0-9981331-8-8

groups in which certain prediction or classification network architectures are sufficiently precise so that the computational costs exceed the actual benefit when optimizing models for each time series individually. Within the forecasting domain, there are several categorization schemes developed, such as (Williams, 1984, Eaves and Kingsman, 2004, A. A. Syntetos et al., 2005, Kostenko and Hyndman, 2006 or Bauer, 2023).

Model Recommendation tries to predict the best suitable algorithm for a specific forecasting problem. Instead of finding boundaries between different time series categories, a complete feature vector with statistical components (e.g. standard deviation, median, mean, etc.), categorization parameters and other characteristic attributes is created to train a classification algorithm that determines the best prediction method for each time series individually. First approaches within this field are investigated by (Montero-Manso et al., 2020, Kiefer, Bauer, and Grimm, 2021, Bauer, 2023).

Transfer Learning is the process of learning weights of a source domain and transfer the knowledge gains to a target domain (Tan et al., 2018). It is thereby not necessary, that the distributions of the training data and the testing data are equal (Pan and Yang, 2009). Originally from the field of image classification, initial approaches in the area of time series prediction were explored, such as (Kiefer et al., 2022), who trained a forecasting model with Walmart demand data and fine-tuned it to the spare part sales of a German business-to-business retailer for prediction. A detailed overview of further approaches is provided by (Weber et al., 2021). **Cross Learning** can be seen as subarea of transfer learning. Instead of adapting features across different tasks, it focuses on related domains and uses all available time series within this domain for the training process, to then forecast them all individually (Semenoglou et al., 2021).

It is the last procedure to which this study is dedicated. Assuming time series categorization and capturing the structural behavior of demand patterns brings an advantage for prediction, and assuming cross learning algorithms benefit from an extended database from the same domain, the derived assumption would then state that cross learning applied to categorized time series would bring an added value compared to a fully trained prediction model. This study has the objective to investigate whether this is the case.

1.1. Research Question and Positioning

As categorization scheme, the proposed procedure by Syntetos, Boylan and Croston (SBC) is chosen (A. A. Syntetos et al., 2005), being one of the most recent presented and well-known approaches of time series categorization for demand forecasting in literature. The application of their scheme to a different dataset, specifically the m5 dataset (Makridakis et al., 2022), leads to the first research question: **(RQ1) How does the SBC categorization scheme behave on a different dataset and what observations can be drawn?**

The categorized data is used to conduct an experiment which investigates the influence of the structural composition of the time series on the prediction quality of transfer and cross learning algorithms. Assuming that similarities in the structure are captured by means of categorization, this results in the second research question: **(RQ2) Can transfer or cross learning benefit from similarity provided by categorization and based only on the structure of a time series?**

With the stated research questions, this study contributes to the research fields of sales prediction and the categorization of time series or demand patterns. A particular focus is on transfer and cross learning algorithms in combination with categorization schemes, as these are promising steps towards more generic prediction approaches without prior knowledge.

1.2. Structure of this Paper

This paper is divided into six Sections. Section 1 introduces the general problem, describes the current state-of-the-art and outlines the resulting research questions. In Section 2, the experiment design is explained, subdividing the work into the categorization of the used time series (Section 3) and the behavioral analysis of transfer and cross learning algorithms with respect to the structural similarity of time series (Section 4). Results are presented and discussed in Section 5. The research concludes with a short summary and an outlook on future investigations (Section 6).

2. Experiment Design

For all experiments, the m5 dataset is taken (Makridakis et al., 2022), which is described more thoroughly in Section 3. In order to answer the first research question, the application of the SBC categorization scheme (A. A. Syntetos et al., 2005) on this dataset will be carried out and evaluated. After the categorization, an experiment is conducted to investigate the influence of the different categories on the training process of the prediction algorithms. For this purpose, two simple deep learning architectures, one multi-layer-perceptron (MLP) and one long-short-term-memory network

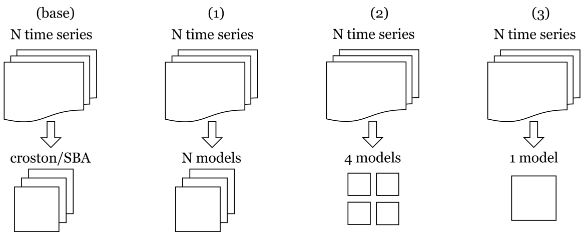

Figure 1. Proposed experiment design. The four models of step (2) correspond to the four categories of Syntetos, Boylan and Croston.

(LSTM, Hochreiter and Schmidhuber, 1997), are used to predict each time series individually. The specific architectures are presented in Section 4. All time series are then grouped by category, and a model is trained for each category using the same deep learning architecture as before. To investigate the relevance of the categories and their influence on the forecasting performance, a final model is trained with all available time series, again using the same model architecture. An exemplary overview of the proposed stages can be seen in Figure 1.

The m5 dataset consists of three groups of sales data categories from different Walmart stores in the USA: *Food*, *Household* and *Hobbies*. As it is not exactly clear which products are behind which time series, the approach can be seen as a mixture of both, transfer and cross learning: All available time series are used for the training process in order to subsequently forecast them individually, whereby certain time series originate from the same area and others are of a different domain.

The mean-squared-error (MSE) method according to the SBC categorization paper is chosen as the evaluation standard. This gives the opportunity to compare the results with their research findings. There are a number of other metrics that have been developed specifically for predicting demand, such as SPEC (Martin et al., 2020) or RMSSE (Makridakis et al., 2022). However, as (Kiefer, Grimm, et al., 2021) noted, the choice of a suitable metric is not trivial and can significantly influence the selection of a prediction algorithm. Therefore, in order to draw comparable conclusions, the same metric as in (A. A. Syntetos et al., 2005) is used to answer RQ1.

The experimental setup outlined here is expected to provide information on the influence of structurally similar time series, which are captured by the categorization scheme, on transfer or cross learning algorithms. The relevance of categorization in relation to time series prediction is highlighted by comparing the results of stages (2) and (3).

Syntetos, Boylan and Croston derived in their paper *on the categorization of demand patterns* (A. A. Syntetos et al., 2005) specific boundaries for a categorization of time series, based on a theoretical comparison of the MSE of three different forecasting algorithms: Exponentially Weighted Moving Average (EWMA), Croston (Croston, 1972) and the Syntetos-Boylan-Approximation (SBA, A. Syntetos and Boylan, 1999). According to their research, each time series can be represented by two parameters, the *squared coefficient of variance* (CV^2) and the *average inter-demand interval* (ADI). With cut-off values equal to 1.32 for the ADI and 0.49 for the CV^2, four quadrants can be created, representing the categories *intermittent*, *lumpy*, *erratic* and *smooth* (see Figure 2). To validate the scheme, a test is carried out with $3,000$ real demand series from the automotive sector. In the paper, it is noted that the available data is relatively short with a length of 24 periods per series. The results show that for all forecasting methods the derived bounds are independent of varying smoothing constants and that for demand patterns that are considered smooth, Croston's method is superior, while the remaining three categories are dominated by SBA (see Figure 2).

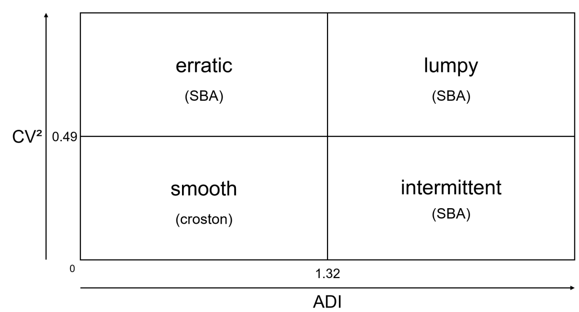

Figure 2. Derived categorization scheme by Syntetos, Boylan and Croston.

3. Dataset Categorizing

To investigate their categorization scheme on a different dataset, the sales data of the m5-competition (Makridakis et al., 2022) is chosen, consisting of $40,820$ time series representing the demand of $30,490$ products of the retail company Walmart. The data is recorded on a daily basis over a time period of approximately 5.5 years. During the competition, participants were asked to create point forecasts for the next 28 days of each time series, whereby the previous sales of the last $1,913$ days served as training period. For the investigation, all daily sales of the training period are taken and categorized according to Syntetos, Boylan and Croston (A. A. Syntetos et al., 2005). It has been

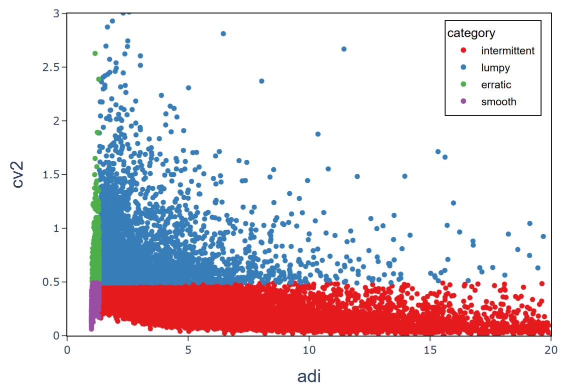

Figure 3. Categorization according to Syntetos, Boylan and Croston. The plot has been cropped to a CV² < 3 and an ADI < 20.

observed that all categories are represented to varying degrees (intermittent $22,168$ - lumpy $5,528$ - erratic 880 - smooth $1,914$). The resulting plot can be seen in Figure 3.

To determine the best performing algorithm per category in average, all before mentioned forecasting methods except for the EWMA are applied to each time series, since the EWMA is considered to perform worse than the other methods, at least theoretically (A. A. Syntetos et al., 2005). The forecast is done by predicting the absolute demand on the 28 days of the evaluation set. In Table 1, the MSE for each method per category is listed.

It turns out that the SBA algorithm performs best for all categories, using the complete sales data of the m5 dataset as input. A discussion of the results follows in Section 5.

Table 1. Results of m5 dataset prediction per category and method. The error is measured using the MSE, lowest errors are underlined.

	croston	**SBA**
intermittent	2.25	2.04
lumpy	11.87	10.64
erratic	41.65	31.57
smooth	33.70	21.72

4. Impact Analysis of Categorized Data

The analysis of the structural behavior influence of categorized time series on transfer and cross learning algorithms is divided into three stages: (1) *single predictions*, (2) *transfer or cross learning predictions*, and (3) *total predictions* (see Figure 1).

In **(1) single predictions**, all available time series are predicted using only the single time series itself. The amount of trained models per algorithm is equal to the amount of available time series. For the **(2) transfer or cross learning approach**, the determined categories of Section 3 are taken into account, resulting in four trained models, one for each category. For the prediction process, these pre-trained models are fine-tuned using the individual time series by incorporating all learned weights except for the last layer. In **(3) total predictions**, all available time series are used to pre-train a single model per algorithm, neglecting the corresponding categories. As with the transfer or cross learning approach, the individual forecast is performed by fine-tuning the model to the time series.

Two deep learning models, a MLP and a LSTM, are selected as prediction methods (see Figure 4). Their architecture is chosen randomly, which means that no optimization task regarding the amount of layers or the number of nodes was performed. Considering the experiment design, it is assumed that the research questions formulated are independent of the architecture, as the architecture remains static in all described stages. According to the m5 challenge, the training history spans $1,913$ entries and the prediction period 28 days. All deep learning trainings are performed using a rolling window forecast, creating $1,885$ training vectors with length of 28 plus label. The final prediction is performed by taking the last 28 days of the training period and iteratively predicting the next day. Each newly predicted day is appended to the time series and used as part of the input vector for the next iteration. Due to the enormous training times (see Table 4), a subset of the m5 dataset is taken by randomly selecting 150 time series of each category. In this way, the dataset is balanced and the results between the categories become comparable.

During the training process, there were extreme outliers in some of the predictions. As the forecast results are of crucial importance for corporate planning, it is important that they are handled with care. Forecasts that are not credible should be treated by means of a firmly defined rule. For this reason, a risk management is introduced and applied to all forecasting models used, including the croston and SBA baselines. Regarding the type of sales prediction, it is obvious that sales cannot be negative. Therefore, all sales forecasts that are below zero are truncated to zero. Choosing a suitable value for the upper limits, on the other hand, is more complicated. In many business applications, outlier detection is carried out using the *3-sigma* rule, which means that with a normal distribution, approximately $\approx 99.73\%$ of the data is captured within this interval. However, this would not allow predictions equal or higher to the

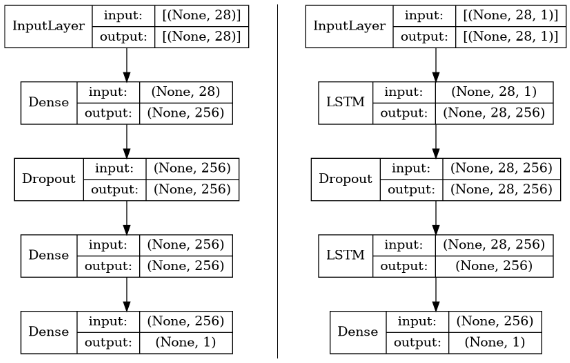

Figure 4. Architecture of the used deep learning algorithms. On the left, the MLP can be seen, the right shows the LSTM.

previous maximum of the sales. Therefore, and in order to minimize the influence of risk management on the experiment, a relatively high upper limit is selected, which corresponds to the maximum of the time series' training period added to its average and three times the standard deviation:

$$\text{upper bound} = \max(TS) + \text{mean}(TS) + 3\sigma(TS) \quad (1)$$

This allows predictions above the previous maximum, but caps unforeseen breakouts of the deep learning algorithms. In total, the rule for the upper limit was applied 8 times, all for the LSTM algorithm. To ensure transparency, more detailed information can be found in the Appendix B.

The experiment's results are presented in Table 2. It can be seen, that the SBA baseline performs relatively best and is only beaten once by the approach of stage (2) with a MLP transfer or cross learning algorithm, and is on same level with the method of stage (3) using one large model trained on all time series. Regarding the transfer or cross learning approach, the performance of the deep learning algorithms is superior to their corresponding single forecast implementation of stage (1).

A closer look at the results of stage (2) shows that the deep learning approaches perform better than the baselines for the intermittent and lumped time series. However, the SBA method dominates in the two

Table 2. Results of all experiments and the baselines. The error is measured using the MSE, lowest errors are underlined.

	croston	SBA	MLP	LSTM
(1) single	22.9	19.17	20.97	20.83
(2) trans/cross	22.9	19.17	19.03	19.19
(3) total	22.9	19.17	19.17	21.25

remaining categories, erratic and smooth (see Table 3 All results are discussed thoroughly in Section 5.

Table 3. Detailed results of stage (2), comparing the baselines against the transfer or cross learning approaches. The error is measured using the MSE, lowest errors are underlined.

	croston	SBA	MLP	LSTM
intermittent	2.47	2.88	1.68	1.66
lumpy	10.68	9.92	8.29	8.41
erratic	42.29	37.73	39.79	39.03
smooth	36.16	26.15	26.38	27.63

5. Results and Discussion

In this section, all results are recaptured and outlined in detail. A thorough discussion is provided, to interpret the findings and answer the stated research questions.

5.1. Evaluation of the M5 Categorization

The analysis of the categorization of the m5 dataset shows that the SBA method introduced by Syntetos and Boylan (A. Syntetos and Boylan, 1999) is superior to Croston's method (Croston, 1972) in all categories. Looking at the results listed in Table 1, the deviations within the intermittent and lumpy categories are the smallest, averaging 0.21 and 1.23, respectively. For the other two categories, erratic and smooth, the average MSE difference is greater than 10. This is surprising and is in contrast to the results of Syntetos, Boylan and Croston (A. A. Syntetos et al., 2005), according to which Croston performs best with smooth time series.

One possible explanation for this deviation is the difference between the two datasets. While Syntetos, Boylan and Croston had less data available, namely only $3,000$ samples with 24 time units each, the m5 dataset contains ten times the number of time series and over $1,900$ data points per sales recording. Furthermore, the results are highly dependent on the underlying distribution of the time series. For the experiment conducted in Section 4, both algorithms, croston and SBA, are evaluated on a randomly sampled subset. With this new dataset, which is evenly distributed across all categories, the results for the predictions have changed. With a value by 0.31 regarding the MSE, croston is superior to SBA for intermittent time series. Looking at the other categories, the MSE difference between both algorithms shrinks on average across all time series.

In conclusion, to answer **(RQ1)**, it can be said that the SBC categorization scheme can be applied to the m5 dataset. It was observed that the performance of the croston and SBA algorithms differed from the original study, but that the underlying distribution of the dataset had a major influence on this prediction result.

5.2. Evaluation of the Impact Analysis

The results of the conducted experiment show that the baseline with the SBA as a prediction method is superior overall to the other approaches investigated (see Table 2). Although the deviation in stage (2) was undercut by an average of 0.14 by the MLP network, the single forecast using the MLP has a higher average by the value 1.8.

Considering the computational costs (see Table 4), a company's current direct choice would be the SBA algorithm, as it delivers top values while taking only one minute and 28 seconds for the prediction process. However, taking the detailed results of the transfer or cross learning into account, a better way would be an ensemble of SBA for the categories erratic and smooth, as well as a MLP for intermittent and lumpy time series. Although the LSTM approach is slightly better by 0.02 compared to the MLP in the intermittent category, the execution times are by at least factor 15 higher.

Regarding the performance of the transfer or cross learning algorithms compared to the single forecasts of stage (1) a clear improvement can be seen of a, at least, 1.64 lower MSE in average. This is a strong indicator, that the structural behavior of the categorized time series has a positive impact on a transfer or cross learners training process.

Looking at the MLP network and taking into account that the predictions were not capped by the upper limit of risk management, the results of the predictions on the entire dataset in stage (3) are remarkable. Although the SBA algorithm is significantly faster and produces the same outcome, a neural network could benefit from further training data in comparison and achieve better results in the long term. Furthermore, it should be noted that, as described in Section 4, no optimization was performed with regard to the network architecture - this would very likely increase the quality of the prediction and make the transfer or cross learning algorithms more competitive against the baselines.

In summary, to answer **(RQ2)**, the results show that deep learning algorithms such as MLPs and LSTMs benefit from categorized data and that regardless of the real product behind the time series within a group, the structural behavior captured by the SBC categorization scheme leads to an improvement in prediction error for transfer or cross learning algorithms compared to individual predictions without information sharing. In addition, indications were observed that the principle "the more the better" also applies to the training of forecasting models, since the MLP model of stage (3) is amongst others the second best performing approach using all available data regardless of its structural similarity.

Table 4. Training times for all prediction approaches. Unit is provided.

model	single	transfer	total
croston	64.87 [s]	64.87 [s]	64.87 [s]
SBA	88.17 [s]	88.17 [s]	88.17 [s]
MLP	11.16 [h]	14.42 [h]	10.79 [h]
LSTM	171.29 [h]	196.82 [h]	163.05 [h]

6. Conclusion and Outlook

In this study, the influence of structural behavior of time series on deep learning algorithms in context with transfer or cross learning is investigated. To capture similar structured demand sales, the categorization scheme from Syntetos, Boylan and Croston (A. A. Syntetos et al., 2005) has been selected and applied to the m5 dataset, a demand sales dataset of Walmart introduced by (Makridakis et al., 2022). A three staged experiment is conducted to explore the differences between single forecasts, transfer or cross learning predictions and a third approach, consisting of a model trained on all available time series of the dataset.

Results show, that the structural behavior captured by the categorization scheme is beneficial for the prediction performance. However, well established algorithms, such as SBA currently outperform simple deep learning algorithms overall in accuracy and performance, when taking the computational costs into account. MLP networks are partially better, but require significantly more time. A closer look on the performance within the different categories, on the other hand, reveals the supremacy of deep learning for intermittent and lumpy time series.

Interesting indications concerning the possibility of building a generic forecast model using all available time series for the training process are gathered: The second best performing model over all time series in average was a MLP using all data as input and competing with the transfer approaches and SBA.

As far as the outlook is concerned, further steps include research into more sophisticated forecasting methods that can make use of transfer or cross learning. In addition to exploring algorithms such as XGBoost (Chen and Guestrin, 2016), temporal convolutional neural networks (TCNN, Pandey and Wang, 2019) and transformers (Vaswani et al., 2017), parameter determination and optimization of the MLP used is also promising. A significance study of the results will also be carried out as part of this process.

The investigation of other categorization schemes as well is an interesting approach. A detailed comparison of all existing methods and their applicability to different datasets is still missing and necessary.

Last but not least, the use of a real-world categorization scheme is the logical next step after this research to investigate the question of whether time series with a degree of similarity based on real-world categories provide comparable or even better added value.

These suggestions will be considered in future work.

Appendices

A. Model Parameters

In Table 5, further information about the model training parameters are listed. All calculations were performed on a Nvidia RTX2080TI and an Intel i0-10900X accompanied by 128 GB RAM.

Table 5. Details about the used model parameters for deep learning algorithm training.

parameters	value
optimizer	adam
learning rate	0.001
loss	MSE
activation functions	relu
dropout	0.2
validation split	0.2
epochs	300
shuffle	false
scaling	standardized
random seed	42

B. Risk Management

This section provides details on the application of risk management and the frequency. The upper bound is only applied to the LSTM model, a total of 8 times. The lower bound is applied to both deep learning methods. In the case of the LSTM model, negative forecasts occurred in 157 time series, in the case of the MLP model this was the case 177 times. In Tables 6, 7, and 8 the influence of the risk management application is specified.

All results are once shown without any risk management, with risk management for lower bounds and with risk management for upper bounds. It can be observed that error deviations are sometimes gigantic, which is due to outbreaks within the prediction process.

Table 6. Details about the applied risk management - table with no correction.

	MLP	LSTM
(1) single	21.21	30.85
(2) trans/cross	19.15	6055.16
(3) total	11643134597203.12	34154.43

Table 7. Details about the applied risk management - table with lower bound correction.

	MLP	LSTM
(1) single	20.97	30.65
(2) trans/cross	19.03	112.12
(3) total	11643134597203.12	34134.4

Table 8. Details about the applied risk management - table with upper bound correction.

	MLP	LSTM
(1) single	21.21	21.03
(2) trans/cross	19.03	5952.23
(3) total	19.17	41.27

C. Exemplary Prediction Plot

In Figure 5, an exemplary prediction plot of one time series can be seen.

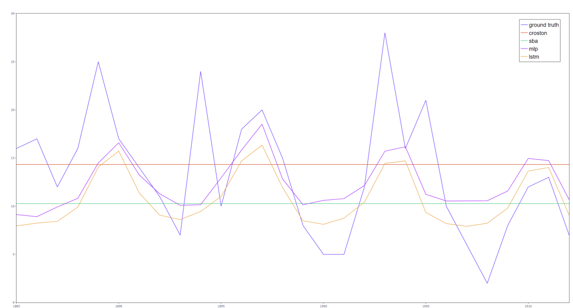

Figure 5. Example prediction curve of the test period of one of the 600 predicted time series. The legend shows the model used for the corresponding colored curve.

References

Bauer, M. (2023). *Improving demand forecasting: The challenge of forecasting studies comparability and a novel approach to hierarchical time series forecasting* [Doctoral dissertation, Karlsruher Institut für Technologie (KIT)]. Karlsruher Institut für Technologie (KIT). https://doi.org/10.5445/IR/1000161070

Carbonneau, R., Laframboise, K., & Vahidov, R. (2008). Application of machine learning techniques for supply chain demand forecasting. *European journal of operational research*, *184*(3), 1140–1154.

Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 785–794.

Croston, J. D. (1972). Forecasting and stock control for intermittent demands. *Journal of the Operational Research Society*, *23*(3), 289–303.

Eaves, A. H., & Kingsman, B. G. (2004). Forecasting for the ordering and stock-holding of spare parts. *Journal of the Operational Research Society*, *55*(4), 431–437.

Fildes, R., Ma, S., & Kolassa, S. (2022). Retail forecasting: Research and practice. *International Journal of Forecasting*, *38*(4), 1283–1318.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, *9*(8), 1735–1780.

Ingle, C., Bakliwal, D., Jain, J., Singh, P., Kale, P., & Chhajed, V. (2021). Demand forecasting: Literature review on various methodologies. *2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, 1–7.

Kerkkänen, A., Korpela, J., & Huiskonen, J. (2009). Demand forecasting errors in industrial context: Measurement and impacts. *International Journal of Production Economics*, *118*(1), 43–48.

Kiefer, D., Bauer, M., & Grimm, F. (2021). Univariate time series forecasting: Machine learning prediction of the best suitable forecast model based on time series characteristics. *International Conference on Human-Centered Intelligent Systems*, 152–162.

Kiefer, D., Grimm, F., Bauer, M., & van Dinther, C. (2021). Demand forecasting intermittent and lumpy time series: Comparing statistical, machine learning and deep learning methods. https://doi.org/10.24251/HICSS.2021.172

Kiefer, D., Grimm, F., & van Dinther, C. (2022). Artificial intelligence in supply chain management: Investigation of transfer learning to improve demand forecasting of intermittent time series with deep learning. https://doi.org/10.24251/HICSS.2022.205

Kostenko, A. V., & Hyndman, R. J. (2006). A note on the categorization of demand patterns. *Journal of the Operational Research Society*, *57*, 1256–1257.

Ma, S., Fildes, R., & Huang, T. (2016). Demand forecasting with high dimensional data: The case of sku retail sales forecasting with intra-and inter-category promotional information. *European Journal of Operational Research*, *249*(1), 245–257.

Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2022). The m5 competition: Background, organization, and implementation. *International Journal of Forecasting*, *38*(4), 1325–1336.

Martin, D., Spitzer, P., & Kühl, N. (2020). A new metric for lumpy and intermittent demand forecasts: Stock-keeping-oriented prediction error costs. *arXiv preprint arXiv:2004.10537*.

Montero-Manso, P., Athanasopoulos, G., Hyndman, R. J., & Talagala, T. S. (2020). Fforma: Feature-based forecast model averaging. *International Journal of Forecasting*, *36*(1), 86–92.

Pan, S. J., & Yang, Q. (2009). A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, *22*(10), 1345–1359.

Pandey, A., & Wang, D. (2019). Tcnn: Temporal convolutional neural network for real-time speech enhancement in the time domain. *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 6875–6879.

Semenoglou, A.-A., Spiliotis, E., Makridakis, S., & Assimakopoulos, V. (2021). Investigating the accuracy of cross-learning time series forecasting methods. *International Journal of Forecasting*, *37*(3), 1072–1084.

Seyedan, M., & Mafakheri, F. (2020). Predictive big data analytics for supply chain demand forecasting: Methods, applications, and research opportunities. *Journal of Big Data*, *7*(1), 53.

Syntetos, A., & Boylan, J. (1999). Correcting the bias in forecasts of intermittent demand. *19th International Symposium on Forecasting*.

Syntetos, A. A., Boylan, J. E., & Croston, J. (2005). On the categorization of demand patterns. *Journal of the operational research society*, *56*, 495–503.

Tan, C., Sun, F., Kong, T., Zhang, W., Yang, C., & Liu, C. (2018). A survey on deep transfer learning. *Artificial Neural Networks* *and Machine Learning–ICANN 2018: 27th International Conference on Artificial Neural Networks, Rhodes, Greece, October 4-7, 2018, Proceedings, Part III 27*, 270–279.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, *30*.

Weber, M., Auch, M., Doblander, C., Mandl, P., & Jacobsen, H.-A. (2021). Transfer learning with time series data: A systematic mapping study. *Ieee Access*, *9*, 165409–165432.

Williams, T. M. (1984). Stock control with sporadic and slow-moving demand. *Journal of the Operational Research Society*, *35*, 939–948.