ISSN: 1803-3814 (Printed), 2571-3701 (Online)
https://doi.org/10.13164/mendel.2024.1.051

MENDEL Soft Computing Journal

INTERMITTENT TIME SERIES DEMAND FORECASTING USING DUAL CONVOLUTIONAL NEURAL NETWORKS

Sina Mirshahi[1,2,3]✉, **Patrick Brandtner**[2,3], **Zuzana Komínková Oplatková**[3]

[1]Department of Informatics and Artificial Intelligence, Tomas Bata
[2]Department for Logistics, University of Applied Sciences Upper Austria, Steyr, Austria
[3]Josef Ressel-Centre for Predictive Value Network Intelligence, Steyr, Austria

Mirshahi@utb.cz✉, patrick.brandtner@fh-steyr.at, oplatkova@utb.cz

Abstract

Forecasting intermittent demands is challenging due to their irregular and unpredictable demand pattern. This makes the businesses unprepared for upcoming demands, where the conventional methods often fail to predict the demand occurrence pattern sufficiently. In this paper, we proposed a two-step approach, "UR2CUTE," (Using Repetitively 2 CNN for Unsteady Timeseries Estimation), employing Convolutional Neural Networks (CNNs) specifically designed to handle the unique challenges of intermittent time series. CNNs, known for their effectiveness in capturing spatial and temporal patterns in data, offer a promising area to improve forecast accuracy in predicting time series demand patterns. Our approach presents a combined process for intermittent demand forecasting. A CNN model is initially designed as a binary classifier to determine demand occurrence. Afterward, a distinct CNN model is employed to estimate the magnitude of the demand. This dual-phase approach improves forecasting accuracy in intermittent demands, specifically in predicting the non-demand (Zero-Demand). The suggested approach notably surpasses traditional forecasting techniques, including Croston's method, which is tailored for intermittent demand forecasting. It also outperforms other methods like XGboost, Random Forest, ETR, Prophet, and AutoArima, especially in predicting the lead time demand distribution for sporadic demands. The deployment of dual CNN models facilitates a deeper understanding of intermittent demand dynamics. This, in turn, enhances supply chain management effectiveness and efficiency, offering a robust solution to the complex challenges of intermittent demand forecasting.

Keywords: *Intermittent Demand Forecasting, Time Series Forecasting, Convolutional Neural Networks, Time Series Analysis, Supply Chain Management*

Received: 10 March 2024
Accepted: 27 May 2024
Online: 30 June 2024
Published: 15 July 2024

1 Introduction

Intermittent demand forecasting (IDF) presents a unique challenge in time series analysis, mainly due to demand data's sporadic and unpredictable nature, which often includes long periods of zero demand followed by bursts of activity [8]. This type of demand is commonly encountered in sectors such as supply chain management, spare parts inventory, and online retail, where accurate forecasting is crucial for maintaining optimal inventory levels and meeting customer needs without incurring excessive costs [1].

Croston first introduced the concept of intermittent demand in 1972 [4]. In the forecasting literature, there is a considerable focus on rapidly changing time series data, often favoring causal models when abundant information is available. However, intermittent demand series and their associated models have received comparatively less attention [20]. The intermittence of data is conceptualized as a pattern characterized by the presence of zero values within the dataset [23]. Traditional statistical methods like ARIMA are generally ineffective for intermittent demand due to their reliance on the assumption of stationarity, which is often violated by the high frequency of zero values in the data. Over the past five decades, few forecasting methods have been developed explicitly for intermittent data, leading to a significant gap in the literature and practice [20]. Existing methods, such as Croston's approach and its variants, which focus on modeling the inter-demand intervals and aggregation techniques that smooth out the intermittency, have shown some success but still face limitations in terms of accuracy and applicability [2]. Despite the significant attention given to modern machine learning and AI-based prediction models in this domain, there is still room for improvement in result accuracy [5]. Modern supply chains' growing complexity and dynamism underscore the need for more effective intermittent demand forecasting methods [19, 30].

Recent advancements in deep learning, particularly in applying convolutional neural networks (CNNs), have opened new avenues for tackling the complexities of intermittent demand forecasting. CNNs have demonstrated remarkable success in various domains, including image recognition and natural language processing, due to their ability to capture spatial hierarchies and local patterns in the data. Based on this, researchers have begun exploring using CNNs for time series forecasting, converting time series data into visual representations to leverage these networks' powerful feature extraction capabilities [24].

Addressing this gap, we propose an innovative methodology building on dual convolutional neural networks (i.e., "Using Repetitively 2 CNN for Unsteady Timeseries Estimation", short: "UR2CUTUE"). Our approach aims to enhance the prediction of irregular patterns in time series data. It divides the predictive process into two phases: identifying event occurrences (order prediction) and quantifying their magnitude (quantity prediction). This bifurcated method is specifically designed to address the unique characteristics of intermittent time series, resulting in improved predictive accuracy for both timing and magnitude of sale demands.

We evaluate our model by employing a real-world dataset featuring intermittent time series data on material demand from an ongoing research project focusing on advanced supply chain management analytics. The initial dataset spans over six years and includes daily demand amounts, resulting in over 3.3 million records for more than 5000 materials. The input dataset for training and testing the UR2CUTE model comprises aggregations on a weekly level, covering 326 weeks from 2017 to 2023 for 99 materials showing intermittent demand behavior. We compare UR2CUTE's performance against other established models, including AutoARIMA, Croston, ETR, Prophet, RFR, and XGBoost.

The remainder of this paper is structured as follows: Section 2 provides background information on intermittent demand forecasting and related work—and section 3 details the methodology, including the architecture of our proposed UR2CUTE model. Section 4 presents the evaluation results, comparing our model's performance against other models. Finally, Section 5 concludes the paper, discussing the implications of our findings and potential future research directions.

2 Background

In business decision-making, accurately forecasting demand is vital for ensuring operational efficiency and meeting customer expectations. Intermittent demand, which refers to irregular and sporadic demand patterns, presents unique challenges across various sectors. Unlike steady or seasonal demand, intermittent demand is characterized by unpredictable periods of no activity followed by sudden spikes, making it difficult to apply traditional forecasting techniques effectively. In supply chain management, such demand poses significant challenges, e.g., for inventory management, as traditional forecasting methods need to predict the timing and magnitude of demand accurately. Appropriate approaches to intermittent demand are crucial to avoid overstocking, which can lead to high holding costs, or understocking, which risks stockouts and lost sales. Effective intermittent demand forecasting enables businesses to optimize inventory levels, reduce costs, and improve service levels, ensuring a more resilient and responsive supply chain [3, 34].

The concept of intermittent demand was first defined by Croston in 1972. While Croston's method tends to have a positive bias, it performs well in inventory management. Optimizing smoothing constants at the individual SKU level is less effective compared to optimization across multiple SKUs. In the realm of inventory control, Croston's method consistently outperforms others, even in scenarios where there is a decline in demand [33]. To overcome the poor performance of traditional intermittent forecasting, particularly in addressing the high variation characteristic of demands for aircraft spares, effective methods such as the weighted moving average (WMA) were investigated due to their superior forecasting performance [9]. Wallström and Segerstedt [31] employed principal components analysis (PCA) to show that the presented measures of forecast errors represent different dimensions, which makes it impractical to reduce them to a single dimension without sacrificing information. Therefore, it is suggested that several measures be interpreted to comprehensively evaluate forecasting methods, highlighting that the dimensions vary among different forecasting techniques. Their paper introduced supportive error measures, such as Cumulative Forecast Error (CFE) in conjunction with the Percentage of Inventory Shortages (PIS) and Number of Shortages (NOSp), to trace bias more reliably [31]. New models for intermittent demand forecasting have been introduced by Snyder et al. [27]. They compared the models with established forecasting methods using a database of car parts demands. Their study showed that the traditional static Poisson format is insufficient for products with low-volume intermittent demands. Their results showed the strength of simple exponential smoothing that works well with an unrestricted negative binomial distribution, and a multi-model approach with information criteria does not provide significant advantages [27]. The effect of combining forecasting approaches for intermittent demand using four methods: combining different methods on original data, combining estimates from a single method applied on different frequencies, combining methods on multiple frequencies, and averaging forecasts from multiple aggregation levels has been investigated by Petropoulos and Kourentzes [22]. Combining outputs from multiple methods does not necessarily improve forecasting performance. However, combining forecasts derived from transformed frequencies using the same or multiple methods improves prediction performance [22]. Forecasting of a similar time series with a fuzzy approach has been developed by Novák and Mirshahi [21]. Overcoming the complexity of intermittent demands, Kourentzes investigated two neural network models, NN-Dual and NN-Rate [14]. Proposed models were designed to provide dynamic demand rate forecasts and overcome the limitations of traditional methods like Croston's. They evaluated the effectiveness of their model on inventory simulation with 1000 time series. The models incorporate regularized training and median ensembles to address the challenge of limited fitting samples in intermittent demand. Despite suboptimal accuracy, the NN models, particularly NN-Rate, outperform Croston's method regarding service levels without requiring significant increases in stock holding. The study underscores the inadequacy of conventional forecasting accuracy metrics for evaluating intermittent demand methods, emphasizing the importance of considering inventory metrics [14].

The forecasting accuracy of single-hidden layer neural networks for intermittent demand has been analyzed by Lolli et al. [18]. Comparisons were made with standard methods (CR and SBA) using different input patterns and architectures. Back-propagation demonstrated superior performance in Mean Absolute Percentage Error (MAPE), with the two-input network showing the lowest bias. Although back-propagation outperformed, the computational cost is higher, making the easier-to-implement extreme learning machines a potential area for future research [18]. The application of LSTM in retail has been investigated by Falatouri et al. [8]. The probabilistic Spatial framework for Intermittent Demand Forecasting has been considered by Türkmen [29], drawing on renewal theory and probabilistic neural forecasting. They extend simple models with flexible inter-demand time distributions to capture various demand patterns. The results show improved probabilistic forecasts with renewal processes and further enhancements with RNNs, particularly on larger datasets like the M5 competition [29]. Similarly, the Spatial-Temporal network has been used by Falatouri et al. [7] to predict intermittent demand for distributed maintenance needs. The use of Deep CNN-LSTM for inventory forecasting has been done by Xue et al. [32]. However, tuning the parameters of the network is one of the challenges in this approach. Various heuristic methodologies have been explored for forecasting in different domains. For example, Smejkalová et al. [26] developed a heuristic methodology for forecasting quantities in waste management, which addressed challenges associated with short time series by combining multiple techniques. Tian et al. [28] used the forecasting Markov-combined method (MCM) for two big datasets (Alibaba and JD) with intermittent demands. They claimed their method has a practical, solid application because it combines several simple methods [28]. The ensemble models, such as the modified DeepAR model named rolled DeepAR, with rolling future predictions, have been introduced by Jeon and Seong [11] to improve the final result.

This paper introduces a forecasting algorithm for intermittent demand's cumulative distribution over a fixed lead time. They utilized historical service parts demand data from nine industrial companies; the proposed method demonstrates superior accuracy in estimating lead time demand distribution compared to exponential smoothing and Croston's variant. The new bootstrapping approach offers advantages, including adaptability to variable lead times and enhanced performance, particularly for short lead times, making it an appealing option for practitioners seeking improved forecast accuracy.

3 Method

3.1 Convolutional Neural Networks in UR2CUTE

Convolutional Neural Networks (CNNs) represent a set of deep neural networks that are highly powerful in processing data, such as images and time series, with a grid-like topology [17]. Characterized by their unique architecture, Convolutional Neural Networks (CNNs) utilize convolutional layers to capture spatial hierarchies of features autonomously and adaptively from input data. These models capitalize on three fundamental principles: local receptive fields, weight sharing, and spatial subsampling. Fields shared weights and spatial subsampling [15] enable them to succeed remarkably in task time series analysis. In intermittent demand forecasting, CNNs are particularly valuable for their ability to capture and model temporal patterns and dependencies within the data [16], offering a powerful tool for understanding and predicting complex demand dynamics [13].

The standard architecture of a CNN processes data through a sequence of layers, starting with the input layer, moving through convolutional layers and pooling layers, and concluding with fully connected layers that produce the final output. This structure is pivotal for extracting features and making predictions based on complex data patterns [12] - Fig 1.

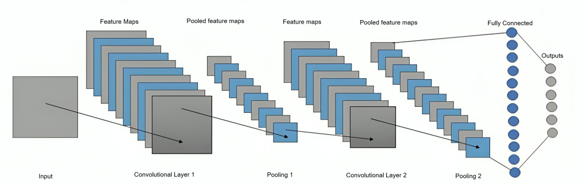

Figure 1: The Generic architecture of CNN

Intermittent demand forecasting poses unique challenges due to the irregular occurrence of demand and the variability in quantities ordered. Our approach (UR2CUTE) addresses these challenges by leveraging the strengths of CNNs. Employing two distinct models for predicting order occurrence and demand quantity, UR2CUTE aims to provide a nuanced understanding of demand dynamics.

The architecture of the CNN models in UR2CUTE is meticulously designed to process time series data efficiently. CNN model core comprises several convolutional layers, which apply filters to the input data to extract and learn temporal features. For layer l, the convolution operation is mathematically represented in Formula (1).

$$F_{l,k}(t) = \phi\left(\sum_{i=1}^{M} W_{l,k,j} \cdot X_{t+j-1} + b_{l,k}\right) \tag{1}$$

Where $F_(l,k)$ (t) is the output of the k-th filter in layer l at time t, $W_(l,k,j)$ denotes the filter weights, M is the filter size, $b_(l,k)$ is the bias, and ϕ is the activation function (e.g., ReLU). A MaxPooling layer comes after the convolutional layers to reduce the data's dimensionality, thereby improving the model's generalization ability. Once the convolutional and pooling layers extract high-level features, these features are flattened and passed into fully connected layers, eventually leading to an output layer customized for the specific prediction task. The order prediction model utilizes a sigmoid activation function [6]. The order occurrence probability is used as a classifier to predict an order's existence and absence. At the same time, the quantity prediction model employs a linear activation for continuous demand quantity predictions. The sequence of convolutional, pooling, and fully connected layers designed to capture temporal patterns in demand forecasting is illustrated in Fig 2.

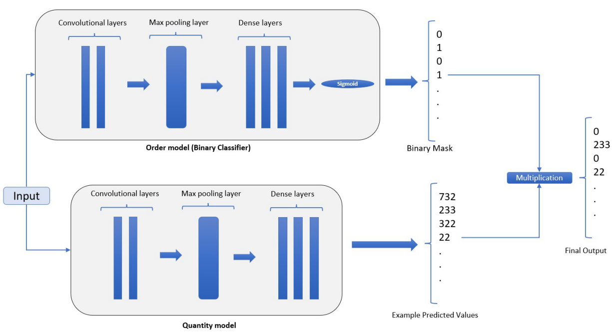

Figure 2: The architecture of the CNN model used in the UR2CUTE approach.

3.2 Model Training

Practical feature engineering is crucial for enhancing the predictive capability of CNNs in time series forecasting. In the UR2CUTE approach, we employ a multi-step feature engineering process to prepare the input data for our models, ensuring they can capture and learn from the complex patterns in demand data. As mentioned, the original dataset spans over six years and includes daily demand amounts, resulting in over 3.3 million records for more than 5000 materials. The input dataset comprises aggregations on a weekly level, covering 326 weeks from 2017W01 to 2023W13 for 99 materials with intermittent demand behavior. This input dataset was split into training and testing data with a ratio of 85% for training (i.e., 277 weeks for 99 materials) and 15% for testing (i.e., 49 weeks for 99 materials).

The initial phase of our feature engineering process entails aggregating demand data [10], which can be performed over various intervals, such as daily, weekly, or monthly, depending on the specific requirements and characteristics of the dataset. In our approach, we opt for weekly aggregation. This decision to consolidate daily observations into weekly summaries effectively minimizes the noise and variability inherent in the data, thereby enabling the model to concentrate on discerning more substantial, long-term trends. This aggregation is accomplished through a custom function designed to cumulate demand for each week, thereby transforming the time series data into a format that more accurately represents weekly demand patterns. This strategic choice underscores our method's adaptability, allowing users to tailor the aggregation interval to best suit their analysis needs and the peculiarities of their demand data.

Following the aggregation, we generate lagged features to capture the historical demand trends effectively. For demand X_t at time t, we create a series of lagged features $X_(t-1), X_(t-2), ..., X_(t-n)$, where n represents the number of prior periods we consider. This approach allows the model to learn from the temporal dependencies in demand, providing a structured format that is highly amenable to CNN analysis. Including these lagged features ensures that our models can accurately leverage past demand patterns to forecast future demand.

An essential component of our feature engineering process is input data normalization [25]. Given the variability in demand data and the potential influence of outlier values, we apply normalization techniques to scale the features into a consistent range. This step, implemented with the MinMaxScaler, assures that all input variables have an equal impact on the model's learning process, avoiding the dominance of any single feature due to varying scales. Normalization is crucial for models like CNNs, which are sensitive to the scale and distribution of input data.

Through these feature engineering steps—weekly data aggregation, generation of lagged features, and normalization—we create a robust dataset optimized for CNN analysis. This structured approach to preparing the input data significantly enhances the models' ability to learn from historical data, ensuring that our CNNs can effectively capture and predict the complex dynamics of intermittent demand.

Distinct loss functions are utilized for each model to cater to their specific forecasting objectives. The order prediction model minimizes binary cross-entropy loss, which is ideal for binary classification tasks in Formula (2).

$$L_{\text{binary}} = -\frac{1}{N}\sum_{i=1}^{N}\left[y_i \log(\hat{y}_i) + (1-y_i)\log(1-\hat{y}_i)\right] \tag{2}$$

With N representing the sample size, y_i the true label, and $\hat{y_i}$ the predicted order occurrence probability for the sample i.

For the quantity prediction model, a custom loss function combines mean squared error (MSE) and mean absolute error (MAE) to reflect the demand quantity's magnitude accurately. This combination allows the model to balance the importance of significant errors—which can significantly impact the overall performance—against the need to minimize smaller but potentially more frequent errors. The combined loss function is justified by its ability to provide a more holistic view of the model's performance across various demand quantities, leading to more accurate and robust demand forecasts. The selection of the hyperparameter α In the combined loss function, it is critical to balance the contributions of MSE and MAE to the overall loss. To optimize this and other hyperparameters, we employ the Optuna framework, a hyperparameter optimization library that systematically searches through predefined parameter space to find the most effective values. This process ensures that the model's performance is maximized by fine-tuning the balance between precision and sensitivity to errors in Formula (3).

$$\begin{aligned} L_{\text{quantity}} &= \alpha \cdot \text{MSE} + (1-\alpha)\cdot \text{MAE} \\ &= \alpha \cdot \frac{1}{N}\sum_{i=1}^{N}(q_i - \hat{q}_i)^2 + (1-\alpha)\cdot \frac{1}{N}\sum_{i=1}^{N}|q_i - \hat{q}_i| \end{aligned} \tag{3}$$

Where q_i and $\hat{q}_i$ are the true and predicted quantities, respectively, and α balances the two error metrics.

4 Result and evaluation

To evaluate our proposed model, we conducted a comparative analysis alongside six other prominent models, including AutoARIMA, Croston, ETR, Prophet, RFR, and XGBoost. For this purpose, we have selected materials from our practical use case that exhibit intermittent behaviors. Based on Croston [4], we have chosen materials with an Average Demand Interval (ADI) greater than 1.32 and a coefficient of variation (CV2) less than 0.49. Applying these criteria, 99 materials remained, and we utilized our model, along with the other mentioned models, to predict their sales values on unseen data (test dataset).

Sales volumes among the selected materials are significantly different; some have higher sales volumes than others; we had to employ appropriate metrics across different materials to evaluate the performance of our models. To this end, we utilized the following metrics. It is essential to mention that metrics such as MAPE may not be appropriate for our analysis. This is because we evaluate intermittent demand, which inherently includes zero values in our dataset [19].

Mean Absolute Error % (MAE %)

$$\text{MAE }\% = \frac{\text{MAE}}{\bar{y}} = \frac{1}{n}\sum_{t=1}^{n}\frac{|\hat{y}_t - y_t|}{\bar{y}} \tag{4}$$

Root Mean Square Error (RMSE%)

$$\text{RMSE }\% = \frac{\text{RMSE}}{\bar{y}} = \sqrt{\frac{1}{n}\sum_{t=1}^{n}\frac{(\hat{y}_t - y_t)^2}{\bar{y}}} \tag{5}$$

Coefficient of Determination (R^2)

$$R^2 = 1 - \frac{\sum_{t=1}^{n}(\hat{y}_t - y_t)^2}{\sum_{t=1}^{n}(y_t - \bar{y})^2} \tag{6}$$

To gain a better overview and understanding of our model's performance compared to other models, we have used a whisker plot to represent the distribution of evaluation metrics across materials Fig 3. These plots visually compare critical statistics such as median and quartiles, helping to identify performance variability across materials. Additionally, they highlight the consistency of our model compared to others and assist in detecting outliers.

Closer proximity of the box elements indicates that the model performs reliably and uniformly across all materials. For two metrics, MAE% and RMSE%, lower values indicate better accuracy, while higher values indicate more accurate prediction for R-squared.

It's worth noting that, as acknowledged in our previous discussions, accurate predictions can be challenging due to the intermittent nature of demand. As shown in Fig 3, the evaluation metrics for such cases may not perform as well as expected, as most of the materials have high MAE% and RMSE% and negative R-squared values, especially when compared to forecastable materials. However, UR2CUTE outperforms other models across all three metrics, indicating that for two metrics, MAE% and RMSE%, the lower and more compact boxes, and for R-squared, higher and more compact boxes, compared to the other models.

For MAE%, although the median is not significantly different from other models where 93.6% compared to the best 89.6%), almost 75 percent of materials (below the 75th percentile) have MAE% less than 100%, and 50 percent of materials with the range between the 25th and 75th percentile have MAE% between 78% and 100%. In contrast, other models exhibit a wider range of performance. XGBOOST performs the best among the other models, ranging from 93% to 185% for the interquartile range.

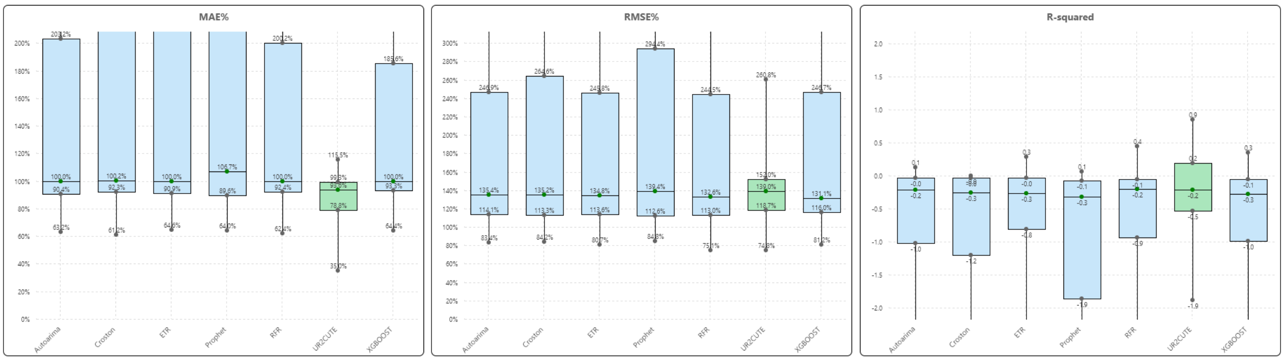

Figure 3: Distribution of evaluation metrics across materials

Regarding RMSE%, similar to MAE%, although the median is not significantly different from other models (139% compared to the best 131%), our model outperforms other models, as almost 75% of the data (below the 75th percentile) have RMSE% less than 152%, compared to the best performer, RFR, which has RMSE% less than 244.5%, indicating significant improvement once again.

Finally, for R-squared, again, our model performed better, particularly regarding the data below the 75th percentile.

Our analysis of the model yielded another valuable insight: adding the classifier to the model improved performance in some materials. Specifically, it predicted zero values when other models typically predicted a fixed value during extended periods of no sales activity. To illustrate this, we included the results of predictions and actual values for a specific material in Fig 4. As shown in the figure, after week 35 (indicated by the dashed vertical line), our model (UR2CUTE) predicted zero values when there were no actual sales values for the remainder of the period. In contrast, other models maintained a fixed amount, significantly deviating from the actual zero sales value.

5 Discussion and Conclusion

The UR2CUTE methodology, through its two-step use of CNNs and meticulous data preprocessing and model training, offers a sophisticated solution to the challenges of intermittent demand forecasting. By capturing the complex temporal patterns and dependencies inherent in in-demand data, UR2CUTE provides accurate predictions for the occurrence and quantity of future demand. It provides a way for more effective and efficient supply chain management.

Traditional models, such as Croston's method and its variations, while foundational in the field of intermittent demand forecasting, have notable limitations in accuracy, particularly when faced with highly irregular demand patterns [27]. The results of our UR2CUTE methodology exhibit a significant improvement in the forecasting accuracy of intermittent demand compared to traditional methods. Our model's superior performance, indicated by lower Mean Absolute Error (MAE%) and Root Mean Square Error (RMSE%) values, along with higher R-squared values, underscores the efficacy of using Convolutional Neural Networks (CNNs) to capture the temporal dependencies inherent in intermittent demand data.

Previous studies confirmed that combining multiple methods' outputs does not necessarily improve forecasting performance [22]. In the case of UR2CUTE, we combined two CNNs – one for frequency and one for magnitude of demand. Our results indicate a significant improvement and show how combining multiple methods can improve forecasting accuracy for intermittent demand.

Convolutional Neural Networks (CNNs) have been widely applied in time series prediction across various domains because they automatically extract and learn relevant features from raw data. Generic applications of CNNs in time series forecasting include financial market analysis, weather prediction, and anomaly detection in industrial processes. These methods typically involve transforming time series data into formats suitable for CNN processing, such as sequences of observations or image-like representations that capture temporal dependencies.

However, generic CNN applications in time series data are not specifically optimized for the unique challenges posed by intermittent demand forecasting. For instance, in the study by Semenoglou et al. [24], a method called "ForCNN" transforms time series data into visual representations, leveraging architectures like VGG-19 and ResNet-50 for forecasting. The strengths of ForCNN include the ability to leverage well-established image processing techniques and the potential for high accuracy in diverse forecasting tasks. While this approach is innovative and effective for general time series data, it may need to address the challenges of intermittent demand patterns fully. In contrast to ForCNN, UR2CUTE offers several distinct advantages, particularly for intermittent demand forecasting. First, UR2CUTE handles time series data directly, eliminating the need to transform time series into images, as ForCNN requires. This direct processing approach simplifies preprocessing and avoids the

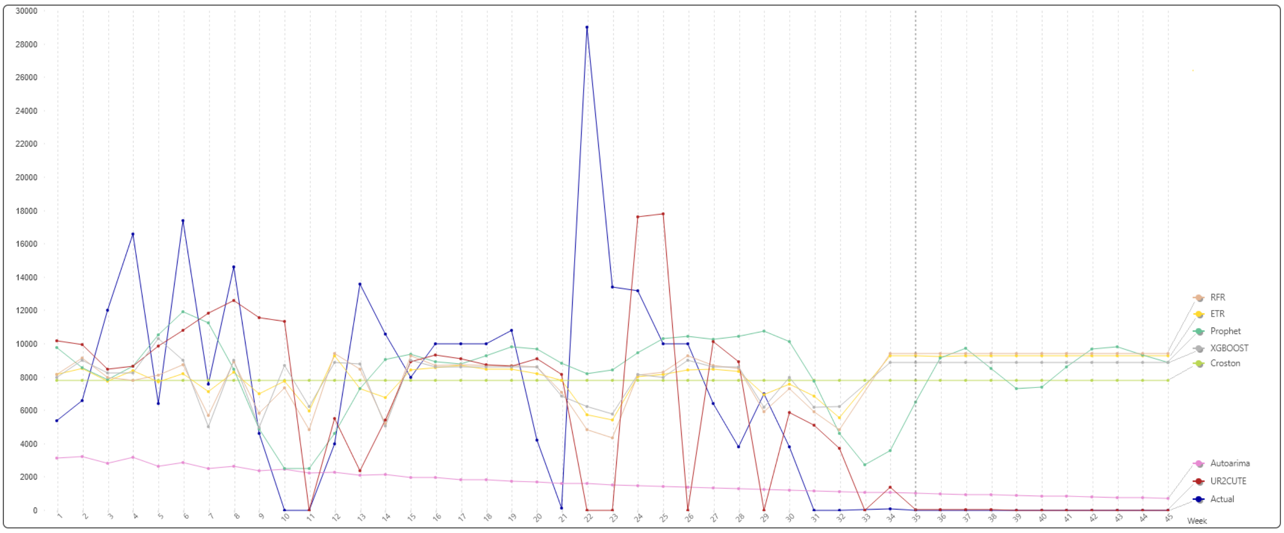

Figure 4: Comparison of Predicted and Actual Sales Values for a Specific Material: Impact of Classifier Integration

risk of information loss during transformation. Second, UR2CUTE is tailored explicitly for intermittent demand scenarios, utilizing dual CNNs to predict the occurrence and magnitude of demand separately. In contrast, ForCNN employs a more generalized approach, which may not effectively capture the unique characteristics of intermittent time series. Lastly, UR2CUTE's architecture is optimized for intermittent time series, resulting in a less computationally intensive model compared to the more complex structures used in ForCNN. This optimization leads to faster training times and reduced computational costs, enhancing its suitability for intermittent demand forecasting.

Another study by Xue et al. [32] presented a hybrid CNN-LSTM model for inventory forecasting, which combines the strengths of CNNs in capturing local trend features and LSTMs in learning long-term dependencies. These models are effective in handling highly nonlinear and non-stationary inventory data. UR2CUTE's dual CNN approach focuses specifically on intermittent demand by using two specialized CNNs for occurrence and magnitude predictions. In contrast, hybrid CNN-LSTM models are more complex, requiring careful architecture design and parameter tuning. Furthermore, the evolutionary algorithms for optimizing CNN-LSTM architectures can be computationally intensive and require substantial computational resources. UR2CUTE, with its simpler architecture, offers a more efficient solution while maintaining high accuracy.

In summary, image-based time series forecasting methods like ForCNN and hybrid models like CNN-LSTM have shown that sophisticated architectures can enhance forecasting accuracy. However, UR2CUTE demonstrates that directly addressing the unique characteristics of intermittent demand through specialized CNN architectures can provide even more significant improvements. This positions UR2CUTE as a robust and effective tool for modern supply chain management, capable of adapting to the dynamic and complex nature of intermittent demand.

References

[1] Anggraeni, W., Yuniarno, E. M., Rachmadi, R. F., Sumpeno, S., Pujiadi, P., Sugiyanto, S., Santoso, J., and Purnomo, M. H. 2024. A hybrid EMDGRNN-PSO in intermittent time-series data for dengue fever forecasting. Expert systems with applications 237, 121438.

[2] Belmiro,M. and Oktariani, F. 2024. Graph Neural Networks in Intermittent Time-Series Forecasting. In Proceedings of the 2nd International Conference on Neural Networks and Machine Learning 2023 (ICNNML 2023), I. H. Agustin, Ed. Advances in Intelligent Systems Research. Atlantis Press International BV, Dordrecht, 108–117. DOI=10.2991/978-94-6463-445-7 13.

[3] Brandtner,P. 2023. Enhancing Decision-Making In SCM: Investigating The Status Quo And Obstacles Of Advanced Analytics In Austrian Companies. Hannover : publish-Ing.

[4] Croston,J. D. 1972. Forecasting and Stock Control for Intermittent Demands. Journal of the operational research society 23, 3, 289–303.

[5] Darbanian,F., Brandtner, P., Falatouri, T., and Nasseri, M. 2024. Data Analytics in Supply Chain Management: A State-of-the-Art Literature Review. OSCM: An Int. Journal, 1–31.

[6] Ding,B., Qian, H., and Zhou, J. 2018. Activation functions and their characteristics in deep neural networks. In 2018 Chinese Control And Decision Conference (CCDC). IEEE, 1836–1841. DOI=10.1109/CCDC.2018.8407425.

[7] Falatouri, T., Brandtner, P., Nasseri, M., and Darbanian, F. 2023. Maintenance Forecasting

Model for Geographically Distributed Home Appliances Using Spatial-Temporal Networks. Procedia Computer Science 219, 495–503.

[8] Falatouri, T., Darbanian, F., Brandtner, P., and Udokwu, C. 2022. Predictive Analytics for Demand Forecasting – A Comparison of SARIMA and LSTM in Retail SCM. Procedia Computer Science 200, 993–1003.

[9] Ghobbar, A. A. and Friend, C. H. 2003. Evaluation of forecasting methods for intermittent parts demand in the field of aviation: a predictive model. Computers & Operations Research 30, 14, 2097–2114.

[10] Hoffmann, M., Kotzur, L., Stolten, D., and Robinius, M. 2020. A Review on Time Series Aggregation Methods for Energy System Models. Energies 13, 3, 641.

[11] Jeon, Y. and Seong, S. 2022. Robust recurrent network model for intermittent time-series forecasting. International Journal of Forecasting 38, 4, 1415–1425.

[12] Khan, A., Sohail, A., Zahoora, U., and Qureshi, A. S. 2020. A survey of the recent architectures of deep convolutional neural networks. Artif Intell Rev 53, 8, 5455–5516.

[13] Koprinska, I., Wu, D., and Wang, Z. 2018. Convolutional Neural Networks for Energy Time Series Forecasting. In 2018 International Joint Conference on Neural Networks (IJCNN 2018). Rio de Janeiro, Brazil, 8-13 July 2018. IEEE, Piscataway, NJ, 1–8. DOI=10.1109/IJCNN.2018.8489399.

[14] Kourentzes, N. 2013. Intermittent demand forecasts with neural networks. International Journal of Production Economics 143, 1, 198–206.

[15] Lawrence, S., Giles, C. L., Tsoi, A. C., and Back, A. D. 1997. Face recognition: a convolutional neural-network approach. IEEE Trans. Neural Netw. 8, 1, 98–113.

[16] Li, Y., Li, K., Chen, C., Zhou, X., Zeng, Z., and Li, K. 2022. Modeling Temporal Patterns with Dilated Convolutions for Time-Series Forecasting. ACM Trans. Knowl. Discov. Data 16, 1, 1–22.

[17] Li, Z., Liu, F., Yang, W., Peng, S., and Zhou, J. 2022. A Survey of Convolutional Neural Networks: Analysis, Applications, and Prospects. IEEE transactions on neural networks and learning systems 33, 12, 6999–7019.

[18] Lolli, F., Gamberini, R., Regattieri, A., Balugani, E., Gatos, T., and Gucci, S. 2017. Single-hidden layer neural networks for forecasting intermittent demand. International Journal of Production Economics 183, 116–128.

[19] Nasseri, M., Falatouri, T., Brandtner, P., and Darbanian, F. 2023. Applying Machine Learning in Retail Demand Prediction—A Comparison of Tree-Based Ensembles and Long Short-Term Memory-Based Deep Learning. Applied Sciences 13, 19, 11112.

[20] Nikolopoulos, K. 2021. We need to talk about intermittent demand forecasting. European Journal of Operational Research 291, 2, 549–559.

[21] Novák, V. and Mirshahi, S. 2021. On the Similarity and Dependence of Time Series. Mathematics 9, 5, 550.

[22] Petropoulos, F. and Kourentzes, N. 2015. Forecast combinations for intermittent demand. Journal of the operational research society 66, 6, 914–924.

[23] Petropoulos, F., Makridakis, S., Assimakopoulos, V., and Nikolopoulos, K. 2014. 'Horses for Courses' in demand forecasting. European Journal of Operational Research 237, 1, 152–163.

[24] Semenoglou, A.-A., Spiliotis, E., and Assimakopoulos, V. 2023. Image-based time series forecasting: A deep convolutional neural network approach. Neural networks : the official journal of the International Neural Network Society 157, 39–53.

[25] Singh, D. and Singh, B. 2020. Investigating the impact of data normalization on classification performance. Applied Soft Computing 97, 105524.

[26] Smejkalova, V., Somplak, R., and Nevrly, V. 2019. Heuristic Methodology for Forecasting of Production in Waste Management. mendel 23, 1, 185–192.

[27] Snyder, R. D., Ord, J. K., and Beaumont, A. 2012. Forecasting the intermittent demand for slow-moving inventories: A modelling approach. International Journal of Forecasting 28, 2, 485–496.

[28] Tian, X., Wang, H., and E, E. 2021. Forecasting intermittent demand for inventory management by retailers: A new approach. Journal of Retailing and Consumer Services 62, 102662.

[29] Türkmen, A. C., Januschowski, T., Wang, Y., and Cemgil, A. T. 2021. Forecasting intermittent and sparse time series: A unified probabilistic framework via deep renewal processes. PloS one 16, 11, e0259764.

[30] Udokwu, C., Brandtner, P., Darbanian, F., and Falatouri, T. 2022. Proposals for Addressing Research Gaps at the Intersection of Data Analytics and Supply Chain Management. JAIT 13, 4.

[31] Wallström, P. and Segerstedt, A. 2010. Evaluation of forecasting error measurements and techniques for intermittent demand. International Journal of Production Economics 128, 2, 625–636.

[32] Xue, N., Triguero, I., Figueredo, G. P., and Landa-Silva, D. 2019. Evolving Deep CNN-LSTMs for Inventory Time Series Prediction. In 2019 IEEE Congress on Evolutionary Computation (CEC). IEEE, 1517–1524. DOI=10.1109/CEC.2019.8789957.

[33] Zied Babai, M., Syntetos, A., and Teunter, R. 2014. Intermittent demand forecasting: An empirical study on accuracy and the risk of obsolescence. International Journal of Production Economics 157, 212–219.

[34] Zimmermann, R. and Brandtner, P. 2024. From Data to Decisions: Optimizing Supply Chain Management with Machine Learning-Infused Dashboards. Procedia Computer Science 237, 955–964.

Appendix 1

Approach Process and Pseudo Code

To clarify the process of the UR2CUTE approach, we detail the steps taken from data preprocessing to model training and prediction as follows:

\# Data Preprocessing
1. Aggregate demand data weekly.
2. Generate lagged features for the time series data.

\# Model Training
3. Split the data into training and testing sets.
4. Normalize the features (if required).
5. Reshape data for the neural network.
6. Set up early stopping based on validation loss.

\# Order Prediction Model
7. Build the CNN model for order prediction.
8. Compile the model with binary cross-entropy loss.
9. Train the model using the training data.

\# Quantity Prediction Model
10. Build the CNN model for quantity prediction.
11. Compile the model with a custom loss function (MSE and MAE).
12. Train the model using the training data.

\# Model Prediction
13. Make predictions using the trained models.
14. Combine predictions to estimate order quantities.

Github repository, has created for code for detailed code review: FH-Prevail. *UR2CUTE*. GitHub, https://github.com/FH-Prevail/UR2CUTE.git.