VMD-SVM-based Retail Product Demand Forecasting Model

Aiyuan Zhu*

Sydney Business School, Shanghai University, Shanghai, China

*zhuaiyuan2609@163.com

Abstract

This study addresses the issues of noise and volatility in retail inventory forecasting by proposing a hybrid model based on Variational Mode Decomposition (VMD) and Support Vector Machine (SVM). Existing methods often struggle to handle noise in time-series data, which affects prediction accuracy. To overcome this, we decompose the raw time-series data into multiple modes using VMD, effectively removing noise and capturing the primary trends. The processed data is then modeled and predicted using SVM. The experiments first compare various machine learning models on the raw data, followed by VMD preprocessing to assess its impact on model performance. The results show that VMD significantly improves prediction accuracy, with SVM's R^2 increasing from 0.4039 to 0.9214. This research provides an effective solution for retail inventory forecasting, demonstrating the advantages of VMD in handling complex time-series data.

Keywords

Support vector machine, VMD, time-series prediction, demand management.

1. Introduction

With the development of globalization and digitalization, the retail industry faces complex market demand fluctuations and inventory management challenges. Inventory forecasting, as a core task of supply chain management, is critical for retailers to optimize stock levels and control costs. However, retail product sales are influenced by factors such as seasonal fluctuations, promotional activities, and unexpected events, making accurate inventory forecasting a challenging task. Traditional forecasting methods, such as time-series and regression models, struggle to handle complex nonlinear relationships and high-frequency noise, leading to low prediction accuracy and stability [1][2][3].

The issue of inventory demand forecasting has attracted significant attention, with various studies proposing different forecasting methods. For example, Tian et al. (2018) conducted a comparative analysis of inventory forecasting models, highlighting the impact of seasonality and promotions on demand predictions [4]. Li et al. (2020) introduced a hybrid method combining traditional time-series models with machine learning algorithms to enhance forecasting accuracy [5]. Machine learning techniques, due to their ability to handle complex and nonlinear relationships, have been widely applied in inventory forecasting. Zhang and Li (2017) reviewed the application of machine learning models in inventory management, demonstrating their advantages in demand forecasting [6]. Wang et al. (2019) focused on the impact of promotions on demand forecasts, noting that traditional models struggle to capture demand spikes caused by promotions [7].

In addition, some studies, such as Xu et al. (2021), incorporated external variables like economic conditions and seasonality into machine learning models to improve prediction accuracy [8]. Huang and Wang (2020) explored the application of deep learning models, particularly LSTM networks, in retail demand forecasting, suggesting that deep learning methods can better capture the complexity of demand patterns [9]. Despite the potential of machine learning and deep learning to enhance inventory forecasting accuracy, handling noisy and volatile time-series data remains a challenge. While traditional time-series models are effective at capturing seasonal trends, they perform poorly in the face of high-frequency noise and sudden demand changes. Zhao and Zhang (2019) proposed solutions but noted that model stability and prediction accuracy still require further improvement [10].

This study provides an effective solution to retail inventory forecasting by addressing the noise and volatility inherent in time-series data. The proposed VMD-SVM hybrid model significantly improves forecasting accuracy, especially in environments characterized by high volatility and strong promotional effects, offering an effective approach for the retail industry.

2. VMD-SVM Model Development

2.1. VMD

Variational Mode Decomposition (VMD) is an advanced signal processing method used to decompose complex time-series data into several intrinsic mode functions (IMFs) with different frequency characteristics. Unlike traditional signal decomposition methods, such as Empirical Mode Decomposition (EMD), VMD directly optimizes the frequency components of each mode using variational principles, thereby preserving the inherent features of the signal while removing high-frequency noise. The main advantage of VMD lies in its ability to adaptively decompose the signal based on the data's characteristics, without relying on predefined settings, and it avoids the mode mixing problem.

The core idea of VMD is to iteratively optimize and decompose the raw signal into modes with specific frequency bandwidths (IMFs). Each mode represents the signal's characteristics within a particular frequency range, and together, these modes more accurately reflect the overall trend of the original signal. VMD is implemented by minimizing a variational problem, automatically decomposing the signal into several modes, with each mode capturing the signal's features in a specific frequency band. This approach is particularly effective for handling noisy and volatile signals, as VMD can extract meaningful information from the data and remove unnecessary noise, providing clearer, more stable inputs for subsequent predictive modeling.

2.2. SVM

Support Vector Machine (SVM) is a powerful supervised learning algorithm widely used for classification and regression tasks. The core idea of SVM is to construct a hyperplane that maximizes the margin between classes, ensuring correct classification of new data points. In regression problems, SVM constructs a regression hyperplane to fit the data, minimizing the deviation between predicted and actual values.

The basic logic of SVM involves finding the optimal hyperplane that separates the samples from different classes, maximizing the classification margin. This hyperplane is determined by solving the following optimization problem:

$$\min_{\mathbf{w},b} \frac{1}{2} \parallel \mathbf{w} \parallel^2 \text{ subject to } y_i(\mathbf{w} \cdot \mathbf{x}_i + b) \geq 1, \ \forall i = 1,2,\dots,n \tag{1}$$

Here, $\mathbf{x}_i$ represents the training samples, y_i are the sample labels (class labels for classification target values for regression), $\mathbf{w}$ is the normal vector of the hyperplane, and b is the bias term.

In regression tasks, SVM minimizes the deviation between the predicted values and the actual values by constructing a regression function. The objective function for regression is typically:

$$\min_{\mathbf{w},b} \frac{1}{2} \parallel \mathbf{w} \parallel^2 + C \sum_{i=1}^{n} \xi_i \quad (2)$$

where ξ_i are slack variables measuring the deviation of data points from the regression function, and C is a regularization parameter that balances error and model complexity.

SVM's advantage lies in its ability to handle high-dimensional data and map the original data to a higher-dimensional space through appropriate kernel functions (such as the Radial Basis Function (RBF) kernel), allowing nonlinear problems to be transformed into linear ones in higher dimensions. This enhances the model's ability to fit complex data. SVM's strong generalization capability and robustness make it an excellent choice for various application scenarios.

In the combined VMD-SVM model, VMD plays a crucial role in preprocessing the data by removing noise and decomposing the signal into different frequency components. SVM then uses these clear, stable mode signals to perform regression modeling. This allows SVM to extract underlying patterns from cleaner input data, improving prediction accuracy and generalization. The VMD-SVM model effectively handles the complex nonlinear relationships and noise in time-series data, leading to more accurate predictions.

3. Experimental Procedure

3.1. Data Selection and Splitting

To validate the effectiveness of the proposed VMD-SVM model for retail sales forecasting, we first conducted a comparative experiment with multiple models. The dataset used for the experiment consists of five years of historical sales data from a retail store, focusing on a single product's daily sales. The main objective of the experiment is to evaluate the prediction accuracy of various models and assess the impact of VMD decomposition on the performance of the SVM model. In the experiment, we compared multiple machine learning models, including RF, XGBoost, MLP-BP, SVM, and LSTM. The data was split into training, validation, and test sets using an 8:1:1 ratio.

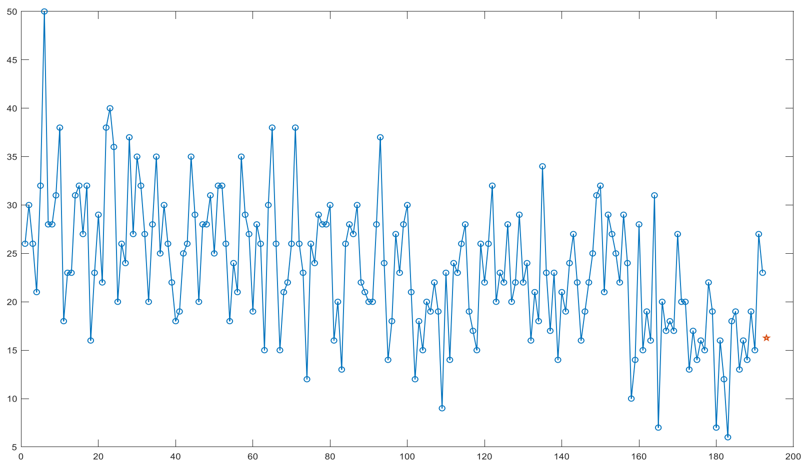

Figure 1. Product Inventory Demand Time-Series Trend

As shown in Figure 1, the raw demand data exhibits significant volatility and irregularity. While periodic fluctuations are clearly visible, the data is also affected by noise and random factors. These unstable components can significantly degrade model performance if the raw data is used directly. Thus, addressing the noise in the data becomes critical to improving model performance. To this end, we applied Variational Mode Decomposition (VMD) to decompose the raw time-series data into multiple modes with different frequency characteristics, effectively removing noise and extracting the primary patterns in the data. This provides smoother and clearer inputs for subsequent modeling.

3.2. Experimental Framework

To thoroughly assess the effectiveness of the proposed VMD-SVM model for retail sales forecasting, we designed a two-phase experimental framework. In the first phase, we compared various machine learning models using the raw data, investigating the prediction performance of each model without any preprocessing. This phase served as a baseline for comparison before introducing VMD decomposition. In the second phase, VMD was applied to preprocess the raw time-series data by extracting different frequency components and removing noise, aimed at enhancing the model's stability and prediction accuracy. The models were then trained and tested on the VMD-processed data to further evaluate the improvement in forecasting accuracy due to VMD decomposition.

In this study, we selected five commonly used machine learning models: Random Forest (RF), XGBoost, MLP-BP neural networks, Support Vector Machine (SVM), and Long Short-Term Memory (LSTM) networks. Each model was trained and tested on both the raw dataset and the VMD-processed dataset to comprehensively assess the effect of different data preprocessing methods on model performance. The final prediction performance of each model was quantitatively compared using standard regression evaluation metrics, ensuring that the accuracy and stability of the models were effectively reflected.

4. Experimental Results and Analysis

4.1. Multi-Model Forecasting with Raw Data

In this section, we first conducted forecasting experiments using various machine learning models with the raw time-series data. The dataset was split into training, validation, and test sets using an 8:1:1 ratio. The training set was used for model training, the validation set for hyperparameter tuning and model selection, and the test set for final performance evaluation. The models selected for this experiment included RF, XGBoost, MLP-BP, SVM, and LSTM, each trained and tested on the data. The experimental results are shown in Table 1.

Table 1. Multi-Model Forecasting Results with Raw Data

Model	**Dataset**	**MAE**	**MAPE**	**MSE**	**RMSE**	**R^2**
RF	Training	2.4872	0.1537	10.4173	3.2276	0.7575
	Validation	4.5313	0.2392	33.1571	5.7582	0.3529
	Test	4.3571	0.2221	27.8073	5.2733	0.3679
XGBoost	Training	2.8281	0.1745	13.3704	3.6566	0.6888
	Validation	4.7507	0.2509	35.0392	5.9194	0.3162
	Test	4.4383	0.2257	28.7912	5.3657	0.3455
MLP-BP	Training	3.9818	0.2517	25.3178	5.0317	0.4107
	Validation	4.5639	0.2394	33.3514	5.7751	0.3491
	Test	4.3558	0.2223	28.5206	5.3405	0.3517
SVM	Training	3.9528	0.2397	25.6132	5.0609	0.4039
	Validation	4.4112	0.2321	31.7277	5.6327	0.3808
	Test	4.4802	0.2277	30.0602	5.4827	0.3167
LSTM	Training	3.8937	0.2392	24.5889	4.9587	0.4277
	Validation	4.5004	0.2356	32.6026	5.7099	0.3637
	Test	4.3453	0.2193	28.5657	5.3447	0.3507

In the multi-model forecasting experiments with raw data, all models exhibited noticeable fluctuations due to the presence of significant noise and irregular patterns in the data. These disturbances interfered with model training and prediction, affecting accuracy and stability. For instance, the LSTM model achieved a MAPE of 0.2392 and RMSE of 4.9587, indicating that it struggled to capture the underlying trends in the raw data. Specifically, the noise caused instability in model performance across different subsets of data, especially between the training and test sets, highlighting that the raw data did not effectively capture meaningful features, thus impacting the model's generalization and stability. In models such as XGBoost and SVM, the performance gap between the training and test sets was substantial, suggesting overfitting to noise and anomalies in the training data, which led to lower prediction accuracy on the test set. This indicates that multi-model forecasting with raw data presents certain limitations, such as lower prediction accuracy and model instability.

4.2. VMD Data Preprocessing

In this study, Variational Mode Decomposition (VMD) was applied to preprocess the raw time-series data, extracting different frequency components and effectively removing noise. Figure 1 displays the time-domain signals of the four modes (IMF1 to IMF4) and the residual (res) obtained through VMD. These decomposed modes reveal the characteristics of different frequency components over time. IMF1 and IMF2 exhibit strong volatility, reflecting the high-frequency components of the data, typically containing rapid noise or short-term fluctuations. In contrast, IMF3 and IMF4 are relatively stable, representing low-frequency components, mainly capturing long-term trends or cyclical variations. The residual (res) includes the part of the original data not explained by other modes, typically containing long-period trends.

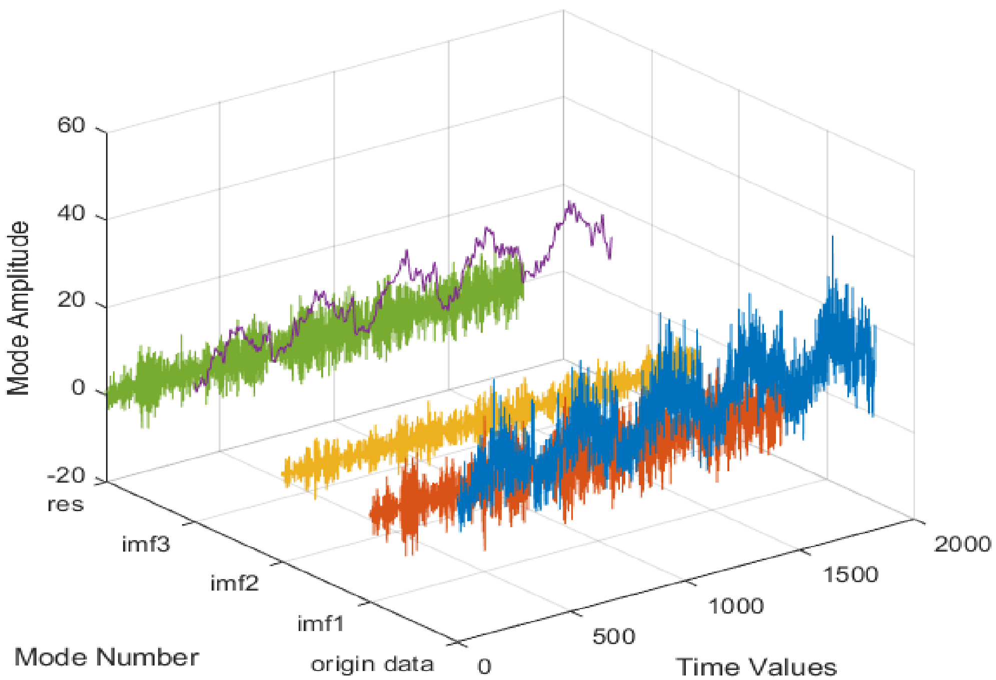

Figure 2. 3D Visualization of VMD Decomposed Modes

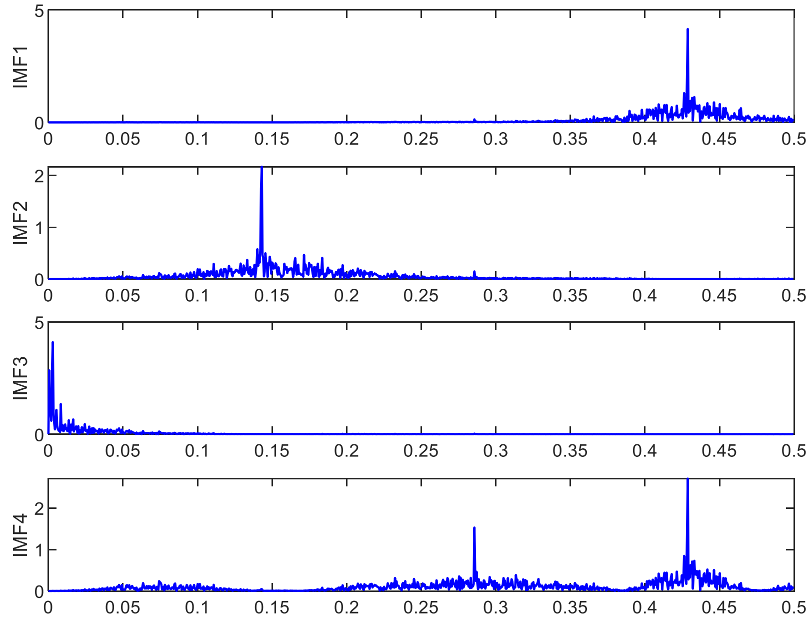

Figure 3. Frequency Spectrum of VMD Modes

Figure 2 further illustrates the frequency spectra of the decomposed modes. From the spectral analysis, IMF1 predominantly occupies the higher frequency range, while IMF3 and IMF4 concentrate in the lower frequency bands. This validates VMD's ability to separate high-frequency noise from the useful signal components, providing cleaner and more stable input for predictive models. Overall, VMD significantly improves data quality, offering a more reliable foundation for model training.

4.3. Multi-Model Forecasting based on VMD

Table 2. Multi-Model Forecasting Results with VMD Data Decomposition

Model	Dataset	MAE	MAPE	MSE	RMSE	R^2
LSTM	Training	1.7354	0.1022	4.744	2.1781	0.8896
	Validation	1.9771	0.0998	6.4045	2.5307	0.8750
	Test	1.9956	0.0972	6.0417	2.458	0.8627
MLP-BP	Training	1.4241	0.0812	3.2245	1.7957	0.9250
	Validation	1.8453	0.0942	5.2912	2.3003	0.8967
	Test	1.9265	0.0938	5.8711	2.423	0.8665
XGBoost	Training	1.1813	0.0700	2.1955	1.4817	0.9489
	Validation	2.2394	0.1170	7.9541	2.8203	0.8448
	Test	2.1834	0.1064	7.3496	2.711	0.8329
SVM	Training	1.4529	0.0829	3.3765	1.8375	0.9214
	Validation	1.679	0.0866	4.4061	2.0991	0.9140
	Test	1.7114	0.0836	4.4762	2.1157	0.8983
RF	Training	1.3439	0.0777	2.9722	1.724	0.9308
	Validation	2.6317	0.1353	10.8866	3.2995	0.7875
	Test	2.4602	0.1195	9.3382	3.0558	0.7877

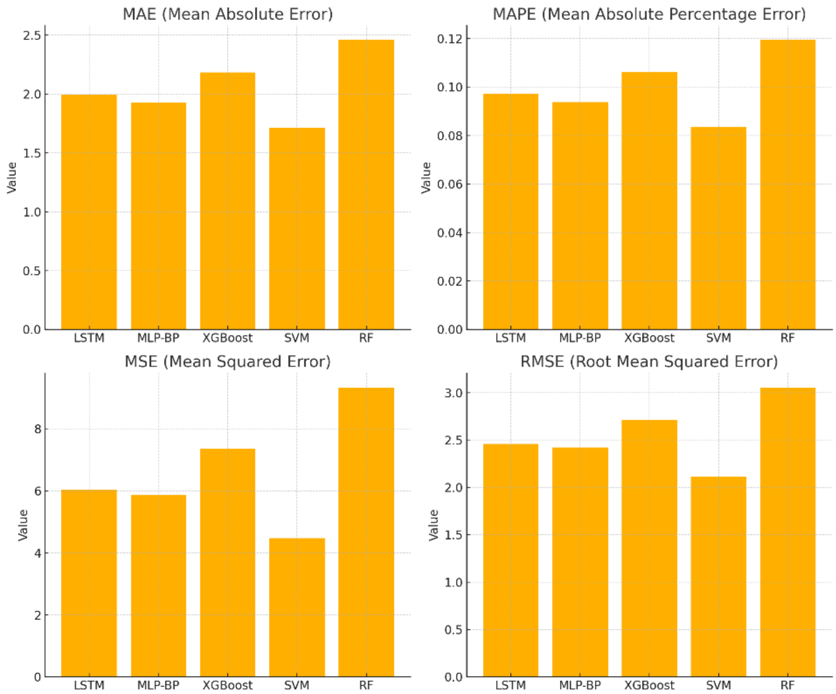

Figure 4. Multi-Model Comparison on Test Data After VMD Decomposition

The results indicate that using VMD-decomposed data significantly improves model prediction performance. This improvement is primarily due to VMD's effective removal of high-frequency noise, enhancing the stability and clarity of the data, which provides a more stable input for the models. Specifically, the models show a noticeable reduction in prediction errors (such as MAE and MAPE) and an increase in R^2 values. For example, the R^2 for the RF model improved from 0.7575 to 0.9308, and for LSTM, it increased from 0.4277 to 0.8896. This demonstrates that VMD decomposition improves the quality of the data, allowing models to better capture long-term trends and periodic patterns.

The effective noise removal by VMD also improves the stability of the models. Noise significantly affects model predictions when using raw data, particularly for complex models like LSTM and MLP-BP, where noise interference can lead to high prediction variance. However, after VMD processing, these noise components are removed, improving the model's adaptability and generalization ability. For instance, LSTM's MAPE decreased from 0.2392 to 0.0972, and its RMSE reduced from 4.9587 to 2.1781, showing a notable increase in prediction accuracy.

Furthermore, VMD helps mitigate overfitting. In models such as XGBoost and SVM, the raw data could lead to overfitting to noise and anomalies, causing significant performance discrepancies between the training and test sets. After VMD processing, the performance gap between training and testing data was significantly reduced, indicating that VMD effectively helped the model identify underlying patterns in the data, rather than memorizing noise. This suggests that VMD not only enhances data usability but also substantially improves model stability and accuracy.

5. Conclusion

In this study, we proposed a hybrid forecasting model combining Variational Mode Decomposition (VMD) and Support Vector Machine (SVM) to address noise and volatility issues in retail inventory forecasting. Experimental results show that VMD preprocessing effectively removes high-frequency noise from time-series data, enhancing data stability and clarity, which significantly improves SVM model prediction accuracy and stability. Compared to other classical models, the VMD-SVM model outperformed in terms of prediction accuracy and generalization ability, demonstrating its reliability and effectiveness in complex time-series forecasting tasks. Future research could extend the application of VMD to other fields, such as finance and weather forecasting, to further validate its versatility and advantages in time-series data analysis.

References

[1] Liu, X., & Zhang, L. (2016). "Challenges in inventory forecasting under complex conditions". Journal of Retailing and Consumer Services, 29, 1-10.

[2] Zhang, Q., & Xu, J. (2017). "Improvement of traditional forecasting methods in retail inventory management". International Journal of Production Economics, 193, 56-70.

[3] Wang, T., & Yang, F. (2018). "Analysis of demand forecasting models for retail inventory systems". International Journal of Forecasting, 34(2), 270-283.

[4] Tian, J., et al. (2018). "Impact of seasonal and promotional variations on demand forecasting for retail inventory". Operations Research Perspectives, 5, 83-92.

[5] Li, H., et al. (2020). "A hybrid model combining time series and machine learning for inventory demand forecasting". Computers & Industrial Engineering, 139, 106184.

[6] Zhang, L., & Li, F. (2017). "Application of machine learning models in retail inventory management". Journal of Retailing and Consumer Services, 34, 275-285.

[7] Wang, X., et al. (2019). "Promotions and demand forecasting in retail: A machine learning approach". European Journal of Operational Research, 278(3), 923-937.

[8] Xu, Y., et al. (2021). "Incorporating economic and seasonal factors in machine learning models for retail demand forecasting". Journal of Retailing, 97(4), 461-474.

[9] Huang, X., & Wang, L. (2020). "Deep learning approaches to retail inventory forecasting: A comparison of LSTM and other models". Decision Support Systems, 128, 113151.

[10] Zhao, H., & Zhang, J. (2019). "Addressing noise and volatility in time series data for retail demand forecasting". Computers & Operations Research, 112, 107-116.