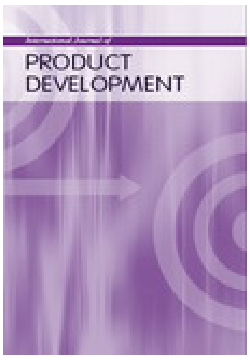


International Journal of Product Development

ISSN online: 1741-8178 - ISSN print: 1477-9056
https://www.inderscience.com/ijpd

Prediction method of product market demand based on Prophet random forest

Chaoyong Jia

DOI: 10.1504/IJPD.2023.10058409

Article History:

Received:	24 April 2022
Last revised:	15 June 2022
Accepted:	03 October 2022
Published online:	05 April 2024

Prediction method of product market demand based on Prophet random forest

Chaoyong Jia

School of Mathematics and Physics,
Bengbu University,
Bengbu, Anhui, China
Email: chaoyjia@mls.sinanet.com

Abstract: This paper proposed a prediction method of product market demand based on Prophet random forest. After analysing the workflow of Prophet model, generate the random forest and its decision-making process and then pre-process the original data of product market through the process of data filling, feature standardisation and feature mapping, providing a reliable data basis for subsequent demand prediction. Then, the optimal subset selection algorithm is used to extract the product market demand characteristics, and the demand characteristics are input into Prophet random forest to realise the prediction of product market demand. The experimental results show that the prediction accuracy of the proposed method is 0.969, and the maximum prediction time in the experiment is only 14.9 s, and the waveform trend of the predicted result is roughly the same as that of the actual value, which highlights the effectiveness of the proposed method.

Keywords: Prophet model; random forest algorithm; optimal subset selection algorithm; product market demand; demand forecast.

Reference to this paper should be made as follows: Jia, C. (2024) 'Prediction method of product market demand based on Prophet random forest', *Int. J. Product Development*, Vol. 28, Nos. 1/2, pp.60–72.

Biographical notes: Chaoyong Jia received his MD degree in Basic Mathematics from Nanjing Normal University in 2009. Currently, he is a Lecturer in the School of Mathematics and Physics of Bengbu University. His research interests include probability theory and mathematical statistics, data analysis and prediction, etc.

1 Introduction

The reform of market economy system has exerted a great influence on the development of enterprises. The decision-makers of enterprises should establish a new management concept, one of which is to change the concept of 'determining production by production' to 'determining production by sales by sales', that is, enterprises produce as much as the market demands (Li et al., 2020; Feng, 2019; Masnadi et al., 2021). For enterprises with many production lines, decision makers timely understand the market demand of various products and make appropriate decisions according to the market demand is the key to the long-term and stable development of enterprises.

With the gradual application of enterprise information management system, many enterprises have stored a large number of historical product sales data in the database (Zhang et al., 2019; Roque et al., 2019). If the computer technology can be used to mine historical data according to the market needs, the market demand of various commodities can be forecast and analysed objectively.

At present, scholars in related fields have carried out a lot of research on demand forecasting methods. In Bandara et al. (2019), a product sales demand prediction method based on long and short-term memory neural network is proposed. Based on the training of long and short-term memory network, the method uses the non-linear demand relationship available in the hierarchy of product classification to construct a forecasting framework and uses the product grouping strategy to supplement the deficiency of long and short-term memory network for different sales patterns. However, each layer of long and short-term memory neural network has four fully connected layers, which greatly increases its time span and computation, and this deficiency is difficult to be improved by a single grouping strategy. Therefore, in practical application, it is found that this method has the problem of long prediction time. In Chu et al. (2019), a logistics product demand prediction method based on GM-Markov model is proposed. This method uses GM (1, 1) prediction model in grey forecasting model to forecast the freight volume in recent years. Then, the prediction accuracy is further optimised according to the posterior error ratio and small error probability. Secondly, the Grey model is corrected by Markov model, and the new model is used to predict the product demand in the future. However, because the order of Markov model is fixed, while the product demand is dynamically time-varying, the correction results obtained by Markov model are still not very accurate to implement the demand modelling, which reduces the final prediction accuracy.

In view of the above problems, in order to improve the accuracy of product demand prediction and shorten the prediction time, this study proposes a prediction method of product market demand based on Prophet random forest. The research ideas are as follows:

First of all, the historical data of product yield, volume growth in trading, annulus compared to historical data, growth, history contract order data, six class data to implement the product life cycle, characteristics of standardisation, such as pretreatment, the original data format neat and complete product market, in improving the quality of the data at the same time, provide reliable data for subsequent demand forecast.

Then, the optimal subset selection algorithm is used to extract the product's market demand characteristics. The traditional forest decision tree will reduce the amount of data every time it splits, and the intentional splitting process will introduce potential errors. Therefore, according to the linear relationship between product market demand characteristic variable and response variable, this study calculates the gap residual between them, and then selects demand characteristic quantity similar to response variable to reduce the basic prediction error and improve the subsequent prediction accuracy.

Once again, bootstrap resampling method was used to extract the training sample set of demand features and establish different branches according to the feature categories. In the process of decision tree growth, the optimal solution is selected according to the minimisation principle of Gini coefficient to form a random forest to avoid the over-fitting problem of decision tree. Then the trend term, period term and holiday term of Prophet model are added to establish Prophet-random forest.

In the end, input the above extracted product market demand characteristics into Prophet random forest, and complete the prediction of product market demand according to random tendency.

2 Analysis of relevant principles

2.1 Prophet model analysis

Prophet model is more inclined to engineering application. Because it encapsulates Python and R language interfaces, it can easily build a predictive analysis environment for time series analysis (Li et al., 2021a; Navratil and Kolkova, 2019; Wu et al., 2019).

The main workflow of Prophet is to combine the two main modules of model and evaluation to conduct rapid iterative optimisation of time series.

Prophet model has the following advantages:

1 There is no need to consider the problem of missing values of data, allowing the input data to have discontinuity in time.

2 The speed of data fitting is greatly improved;

3 It is inclined to engineering practice, and provides a packaged class library. Its parameters can be interpreted strongly, and debugging is convenient and fast.

2.2 Analysis of traditional random forest algorithm

The random forest algorithm takes decision tree as the basic idea and the output classification of the algorithm is obtained by voting the classification of each tree (Zhang and Nie, 2021; Li et al., 2021b).

The forest generation process of random forest is as follows: Firstly, Bagging sampling is conducted on the training set to obtain multiple sub-training sets, and then training is conducted on each sub-training set. In the process of training, the data of each sub-training set can generate a decision tree. Node splitting algorithm is used to generate decision tree nodes, and then continue to split until all nodes become leaf nodes.

The decision-making process of random forest is: for the data of the test set, the predicted value of each test is obtained through the judgment of each test set, and then the result of each test is voted and the one with the highest number of votes is the final predicted result. In this process, the random selection training subset and the random selection splitting feature are introduced to reduce the probability of overfitting problem, and the parallel algorithm is adopted to speed up the running speed of the random forest algorithm.

3 Product market demand forecasting method

3.1 Pre-processing raw data of product market

Firstly, the raw data of product market are pre-processed. The Web Scraper tool is used to collect original product market data in the enterprise research system, including historical product production data, historical trading volume data, year-on-year transaction growth, sequential transaction growth, historical contract order data and product life cycle. The raw data is of type long int and the word length is 32 bits. Because the original data in the product market is missing, omission and other problems, so the original data should be pre-processed first, in order to provide data for the algorithm. The initial data pre-processing of product market includes data filling, feature standardisation, feature mapping, etc. The specific steps are as follows:

(A) *Data filling*: In the original data of the product market, it is inevitable that there will be some abnormal or missing information. If it is not dealt with, the accuracy of the prediction will be affected to some extent. Therefore, these data need to be processed (Du et al., 2020). EM algorithm is used to fill the original data of the product market. The specific description is as follows:

$q = \{q_1, q_2, ..., q_k\}$ is used to represent the division of the original data set of the product market, w is used to represent the original data vector of the missing product market, E is used to represent the original data vector of the complete product market. $r+1$ iteration of the EM filling algorithm is shown as follows:

$$Q_{r+1} = \frac{E_\alpha}{w} \times \left\{ \log R_c(\alpha) \middle| q \right\} \tag{1}$$

In formula (1), α represents the estimated parameter in the $r+1$ iteration, E_α represents the maximum expected value, $R_c(\alpha)$ represents the expected value of the original data filling product market in the $r+1$ iteration.

Update the estimated value α^{r+1} of α so that the q_{r+1} function of E in the whole parameter space takes the maximum value.

(B) *Standardisation of characteristics*: The value range of product markets for different attributes can vary widely, which requires the standardisation of raw product market data (Zhang et al., 2020). The original data of the product market are converted into standardised data by using the minimum-maximum normalisation method, and the conversion formula is as follows:

$$z = \left[\frac{(z_{\max} - z_{\min})(x - x_{\min})}{(x_{\max} - x_{\min})} \right] + z_{\min} \tag{2}$$

In formula (2), z is the standardised product market processing data, $z_{\max}$ and $z_{\min}$ are the maximum and minimum values of the standardised product market data, x is the original data of the product market, $x_{\max}$ and $x_{\min}$ are the maximum and minimum values of the original data of the product market.

(C) *Feature mapping*: Because some features are represented in text in the raw product market data. Therefore, it must be converted into classified type data. In this paper, a self-organising feature mapping method is used to map data structure features. The specific process is as follows:

The connection weights between input neurons and output neurons were compared, and the neurons adjacent to o were selected to establish p_o group. Then, a new input mode

U is established to calculate the distance between the original input data and each output neuron o:

$$d_o = U - p_o \tag{3}$$

Given a neighbouring domain $p_k(t)$, the weight of the neighbouring neuron of the output neuron is modified in the following process:

$$w_o(t+1) = w_o(t) + \beta \times p_k(t) \tag{4}$$

In formula (4), β is the gain term. The properties of the output raw data are then mapped as:

$$e_o = F(d_o) w_o(t+1) \tag{5}$$

Through the above steps, complete the pre-processing of the original data of the product market, and lay the data foundation for the subsequent feature selection.

3.2 *Select product market demand characteristics*

On the basis of pre-processing the raw data of product market, the optimal subset selection algorithm is adopted to select the product market demand characteristics. Since the traditional forest decision tree will reduce the amount of data every time it splits, at the same time, the intentional splitting process will introduce potential errors. Therefore, the gap residuals between the characteristic variables of market demand and the response variables are firstly calculated in this study, and the demand characteristic quantities similar to the response variables are selected to reduce the base error.

Optimal subset selection algorithm is evolved on the basis of least squares (Cao et al., 2021). It applies the evaluation index of simple linear regression. Simple linear regression refers to an assumed linear relationship between the characteristic variable T and the response variable Y of the market demand for a single product, which is mathematically denoted as follows:

$$Y \approx \gamma_0 + \gamma_1 \times T \tag{6}$$

In formula (6), γ_0 and γ_1 represent two unknown constants. The estimation values of γ_0 and γ_1 can be obtained by using the training set of product market demand characteristics. The estimation values can be obtained as follows:

$$y \approx \gamma'_0 + \gamma'_1 \times t \tag{7}$$

According to the *i*-th value of characteristic variable t of product market demand, formula (7) is used to estimate y. Where, the gap between the y_i value of the *i*-th actual product market demand feature and the y'_i value of the *i*-th product market demand feature obtained by the linear model is the residual, and the calculation process is as follows:

$$a_i = y_i - y'_i \tag{8}$$

The sum of the squares of residuals is:

$$A_S = a_1^2 + a_2^2 + \cdots + a_i^2 \tag{9}$$

The smaller the A_S value is, the smaller the difference between the predicted product market demand eigenvalue and the actual product market demand eigenvalue. On the basis of A_S, δ^2 statistic can be deduced as:

$$\delta^2 = \frac{D_S - A_S}{D_S} = 1 - \frac{A_S}{D_S} \tag{10}$$

In formula (10), $D_S = \sum \left(y_i - y'_i \right)^2$. If the δ^2 statistic is close to 1, it indicates that the product market demand feature selection is more explanatory.

Based on the above analysis, according to the linear relationship between the product market demand characteristic variable and the response variable, the gap residual between them is calculated and then the demand characteristic quantity similar to the response variable is selected.

3.3 Product market demand forecast

Prophet model can not only improve the fitting degree of data, but also allow the discontinuity of input data in time, so as to realise the rapid iteration of time series. Prophet model is used to optimise the random forest, and the two are combined together. In the process of decision tree growth, the optimal solution is selected according to the principle of Gini coefficient minimisation, and the random forest is established to avoid the over-fitting problem caused by decision tree and improve the prediction accuracy.

Based on the above analysis and on the basis of extracting product market demand characteristics, this study input reliable market demand characteristics into Prophet random forest and select the optimal solution according to the minimisation principle of Gini coefficient in the process of decision tree growth to establish a random forest to avoid over-fitting problems due to decision tree. Then output the product market demand forecast results, realise the product market demand forecast.

Firstly, Bootstrap resampling method was used to extract the training sample set of demand features from the original data set and establish different branches according to the feature categories. In the process of decision tree growth, the optimal solution is selected according to the minimisation principle of Gini coefficient to form a random forest. Then the trend term, period term and holiday term of Prophet model are added to establish Prophet-random forest. Input the above extracted product market demand characteristics into Prophet random forest, and complete the prediction of product market demand according to random tendency.

The construction process of Prophet-random forest algorithm is as follows:

Step 1: Bootstrap resampling method was used to extract g_{tree} training sample sets of product market demand characteristics from the original data set G of product market.

Step 2: Design a model based on weak learner, and randomly select j_{try} features from the market demand features of J products for analysis. Then, according to the minimisation principle of Gini coefficient, the optimal characteristic solution is selected and decomposed into:

$$\mu = G\left(J\left(g_{tree} \times j_{try}\right)\right) \tag{11}$$

Step 3: As the decision tree grows, each decision tree follows Step 2 and keeps the j_{try} value unchanged.

Step 4: Construct a random forest by g_{tree} decision tree generated in the previous 3 steps.

Step 5: Prophet model consists of trend term, cycle term and holiday term and its basic decomposition mode is as follows:

$$z(t) = x(t) + c(t) + v(t) + b(t) \tag{12}$$

In formula (12), $x(t)$ is the trend term, representing the change function of the non-periodic part of the time series, including assumptions of different degrees and parameters that can adjust the smoothness. $c(t)$ is the periodic term, which is constructed by the Fourier series in the Prophet model. $v(t)$ is a holiday item, which refers to the abnormal effect caused by holidays or special circumstances. $b(t)$ is a residual term with normal distribution, which represents the random tendency that is not predicted in the Prophet model. Among them:

(A) *Trend term is a logical function:*

$$x(t) = \frac{H}{1 + e^{\left(-\theta\left(\varepsilon - f\right)\right)}} \tag{13}$$

In formula (13), H represents Prophet model capacity, θ represents growth rate and ϵ represents offset. With the increase of ε, $x(t)$ gradually approaches H.

(B) *The calculation process of the period term is as follows:*

$$c(t) = \sum_{n=1}^{N}\left(\cos\frac{2\pi\vartheta}{G} + \sin\frac{2\pi\vartheta}{G}\right) \tag{14}$$

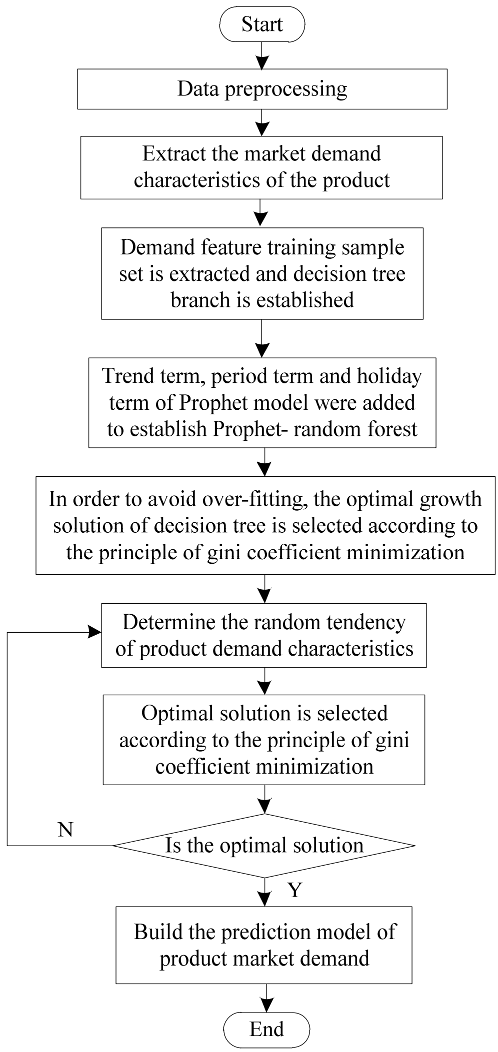
In formula (14), G represents a fixed period, and ϑ represents the expected number of cycles in the Prophet model.

Step 6: Use demand training sample set to establish Prophet random forest algorithm and build product market demand prediction process as follows:

$$\rho = \left(\frac{g_{tree}}{G} + \frac{j_{try}}{J}\right) \times \mu + z(t) \tag{15}$$

Through the above steps, the product market demand forecast results can be output. The specific product market demand forecasting process is shown in Figure 1.

Figure 1 Product market demand forecasting flow chart



4 Experimental simulation and analysis

In order to verify the effectiveness of prediction method of product market demand based on Prophet random forest, the following experiment is designed.

4.1 Setting up the experiment environment

Set the experimental hardware platform as Intel Core I3-2350M CPU@2.30 GHz processor, memory as 8 GB, software platform as Window 7 operating system and experimental simulation software as MATLAB 2016a. Based on Mintel Global New

Product Database (GNPD), this paper selects the product market demand characteristic time series as experimental data, uses Prophet random forest to establish a prediction model for market demand characteristic variables, and obtains the final actual data and forecast data.

In order to highlight the comparison of experimental results, Bandara et al. (2019) method, Chu et al. (2019) method and the proposed method were used for comparison.

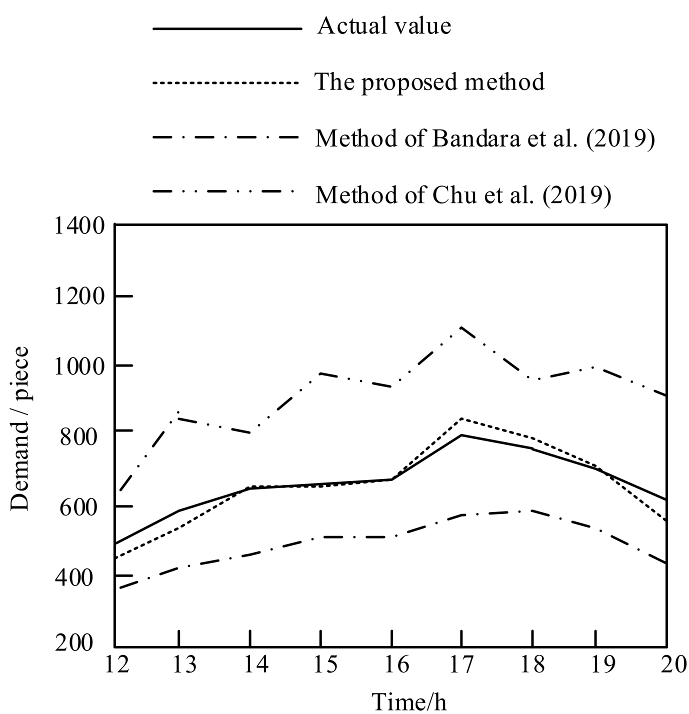
4.2 Contrast indicators

In the experiment, the correlation between demand prediction result waveform and actual waveform, prediction accuracy and prediction efficiency were taken as indicators to verify the prediction effect and efficiency of different methods.

4.3 Forecasting effect of product market demand

In order to verify the predicted effect of the proposed method on product market demand, based on the test sample set of product market demand characteristics, Bandara et al. (2019) method and Chu et al. (2019) method were used to compare with the proposed method. Third-party software was used to output actual and predicted data of product market demand, and the prediction effects of different methods were compared. The result is shown in Figure 2.

Figure 2 The forecasting effect of different methods on product market demand

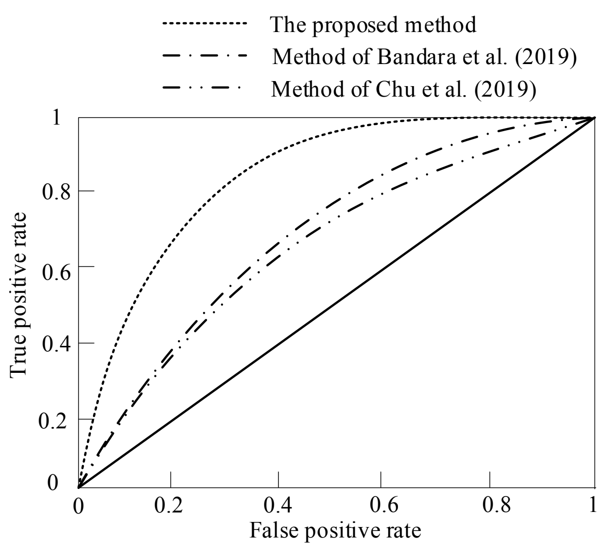


As can be seen from Figure 2, the waveform of product market demand prediction result of Bandara et al. (2019) method is relatively smooth, and the prediction result is relatively small. The waveform of product market demand prediction result of Chu et al. (2019) method fluctuates greatly, and the prediction result is relatively large. The waveform trend of the predicted product market demand of the proposed method is roughly the same as that of the actual value, and the predicted result is close to the actual value. It can be seen that the proposed method has a good forecast effect on product market demand.

4.4 Product market demand forecasting accuracy

The proposed method further verified the accuracy of product market demand prediction, and took ROC curve as an evaluation index. Where, AUC value refers to the area under the ROC curve. The closer AUC value is to 1, it indicates that the authenticity of product market demand prediction is higher and the accuracy of product market demand prediction is higher. The product market demand prediction accuracy of different methods is shown in Figure 3.

Figure 3 Different methods of product market demand forecasting accuracy



As can be seen from Figure 3, AUC value of Bandara et al. (2019) method is 0.785, that of Chu et al. (2019) Method is 0.724 and that of the proposed method is 0.969. Therefore, compared with Bandara et al. (2019) method and Chu et al. (2019) method, AUC value of the proposed method is closer to 1, indicating that the proposed method has a higher authenticity in product market demand prediction. This is because the method in this paper uses the Prophet model to optimise the random forest and combines the two. In the process of the growth of the decision tree, the optimal solution is selected according to the minimisation principle of the Gini coefficient to construct the random forest, so as to avoid the over-fitting problem caused by the decision tree and improve the prediction accuracy.

4.5 Product market demand forecast efficiency comparison

The efficiency of product market demand prediction of the proposed method was verified, and the prediction time was taken as an evaluation index. The forecasting time of product market demand by different methods is shown in Table 1.

Table 1 Different methods of product market demand forecast time

Iterations/times	*The proposed method/s*	*Bandara et al. (2019) method/s*	*Chu et al. (2019) method/s*
100	6.8	9.9	12.2
200	8.3	12.5	15.8
300	10.5	15.2	18.5
400	12.7	18.3	23.3
500	14.9	21.1	26.2

As can be seen from Table 1, with the increase of iterations, the time of product market demand prediction of different methods increases. When the number of iterations is 500, the prediction time of product market demand of Bandara et al. (2019) method is 21.1 s, and that of Chu et al. (2019) method is 26.2 s. The proposed Method only predicted the product market demand within 14.9 s. Therefore, it can be seen that the proposed Method has a short forecasting time for product market demand, which proves that it can effectively improve the forecasting efficiency of product market demand. This is because the method in this paper firstly pre-processes the relevant data, such as filling and feature standardisation, so that the original product market data format is regular and without leakage. It not only provides a reliable data basis for the follow-up demand prediction, but also avoids the marketing of problem data to the prediction process, thus improving the prediction efficiency.

5 Conclusions

In this paper, a Prophet-random forest-based product market demand forecasting method is proposed. The method pre-processes the raw data of product market through data filling, feature standardisation and feature mapping. Then, on the basis of Prophet model, the optimal subset selection algorithm is used to extract the product market demand characteristics, and the demand characteristics are input into Prophet-random forest to realise the prediction of product market demand. In this study, the innovation work is to use Prophet model to optimise the random forest and combine the two. In the process of decision tree growth, the optimal solution is selected according to the minimisation principle of Gini coefficient to build the random forest, so as to avoid over-fitting problems caused by decision tree and improve prediction accuracy.

According to the experiment result shows that the method of product market demand prediction accuracy is 0.969, among the highest in the experiment to predict time of just 14.9 s, and the trend of waveform prediction results and the actual value is roughly same, explain product market demand prediction effect of this method is good, can effectively improve the product market demand prediction accuracy and efficiency.

However, this method does not consider the nature of the product market itself and does not rely on the product attributes of historical sales. Therefore, in the following research, multiple combination attempts can be made on different product market attributes to find the optimal product combination, so as to further improve the prediction accuracy of product market demand.

Acknowledgement

This work was supported by Key Natural Science Foundation of Anhui Education Commission (No. KJ2021A1128).

References

Bandara, K., Shi, P., Bergmeir, C., Hewamalage, H., Tran, Q. and Seaman, B. (2019) 'Sales demand forecast in e-commerce using a long short-term memory neural network methodology', *International Conference on Neural Information Processing*, Springer, Cham, Vol. 19, No. 9, pp.462–474.

Cao, D., Fan, S. and Xia, K. (2021) 'Comparison of machine learning methods for short-term demand forecasting of shared bicycles', *Computer Simulation*, Vol. 38, No. 1, pp.92–97.

Chu, Y., Lian, W. and Wang, Y. (2019) 'Prediction of logistics demand in Jing-Jin-Ji region based on GM-Markov Model', *Mathematics in Practice and Theory*, Vol. 49, No. 21, pp.262–270.

Du, G., Xie, J., Fan, Z., Geng, J. and Shen, Z. (2020) 'Research on the prediction method of comprehensive energy demand based on fuzzy evaluation', *Automation and Instrumentation*, Vol. 59, No. 12, pp.223–226.

Feng, W. (2019) 'Market demand, government participation and capacity utilization – an empirical analysis of Chinese Industries', *Journal of Dalian University of Technology (social sciences)*, Vol. 40, No. 5, pp.26–35.

Li, F., Xv, L., Zhu, R., Luo, H. and Chen, G. (2021) 'Borrowing demand forecast of shared bikes based on XGBoost algorithm', *Journal of Wuhan University of Technology (Transportation Science and Engineering)*, Vol. 45, No. 5, pp.880–884.

Li, Y., Yang, Y., Zhu, K. and Zhang, J. (2021) 'Clothing sale forecasting by a composite GRU–Prophet Model with an attention mechanism', *IEEE Transactions on Industrial Informatics*, Vol. 17, No. 12, pp.8335–8344.

Li, Z., Yang, W. and Liu, X. (2020) 'A strategic analysis of dual-channel supply chain with stochastic market demand and different game status', *Industrial Engineering Journal*, Vol. 23, No. 4, pp.36–42+52.

Masnadi, M.S., Benini, G., El-Houjeiri, H.M., Milivinti, A., Anderson, J.E., Wallington, T.J., Kleine, R., Dotti, V., Jochem, P. and Brandt, A.R. (2021) 'Carbon implications of marginal oils from market-derived demand shocks', *Nature*, Vol. 59, No. 3, pp.80–84.

Navratil, M. and Kolkova, A. (2019) 'Decomposition and forecasting time series in the business economy using prophet forecasting model', *Central European Business Review*, Vol. 8, No. 4, pp.26–34.

Roque, L., CFernandes, C.A.C. and Silva, T. (2019) 'Optimal combination forecasts on retail multi-dimensional sales data', *Preprint*, Vol. 22, No. 8, pp.478–486.

Wu, H., Zhang, L., Liu, D. and Yang, B. (2019) 'Research on forecasting method of port logistics demand trend – analysis of Chongqing port logistics demand trend based on combined forecasting model', *Price: Theory and Practice*, Vol. 32, No. 9, pp.75–78.

Zhang, F., Cao, G., Liu, S. and Zhu, Y. (2020) 'Research on user requirements acquisition and prediction method based on web and patent statistical analysis', *Machinery Design and Manufacture*, Vol. 43, No. 8, pp.59–63+67.

Zhang, X. and Nie, W. (2021) 'Forecast of shared bicycle demand based on improved random forest', *Computer and Digital Engineering*, Vol. 49, No. 9, pp.1860–1865.

Zhang, Y., Long, J. and Chen, Y. (2019) 'Contingency police and operation simulation of supply chain disruption with changing demand', *Journal of Systems and Management*, Vol. 28, No. 6, pp.1202–1210.