Research on Sales Forecast Based on Prophet-SARIMA Combination Model

Jiantao Zhao, Chunwei Zhang

School of Control and Computer Engineering, North China Electric Power University, Beijing 102206, China
zhangchunwei@eplian.com

Abstract: The stable development of e-commerce enterprises cannot be separated from scientific replenishment planning, and forecasting the sales volume of products helps e-commerce to make a reasonable replenishment plan, so as to speed up cash flow and reduce inventory pressure. Based on the in-depth study of Prophet model and seasonal differential autoregressive moving average model, as well as the analysis of daily sales volume of an e-commerce product from January 2015 to March 2018, a Prophet-SARIMA combination model for sales volume prediction is proposed. The principle of the combination model is to use prophet and SARIMA model to forecast the sales data respectively, and then weighted combination to get the final forecast results. Analyze and compare the prediction effect of single model and combination model. The first mock exam shows that the Prophet-SARIMA combination model has higher accuracy and stability in sales volume forecasting than single model. The sales volume forecast results of Prophet SARIMA combination model can provide an important basis for the e-commerce to supplement orders scientifically.

1. Introduction
In e-commerce work, replenishment is an unavoidable problem for each seller, and the reasonableness of replenishment will directly affect the occupancy ratio of cash flow and inventory pressure[1]. Accurate prediction of product sales can help sellers to make a scientific replenishment plan and provide a reference for the capital and inventory allocation of enterprises.

At present, the prediction methods of sales volume mainly include traditional time series model such as Arima, neural network model and combination prediction model. Ge Na used ARIMA model to predict the sales volume of footwear products of an enterprise, and achieves good results[2]. Chang Xiaohua used random forest algorithm to model and forecast the sales volume of a medical device company[3]. Liu Lu established a multi-dimensional time series model of tobacco sales volume based on SVM, and achieved high prediction accuracy. In the first mock exam[4], Zhou Yanfu used the grey neural network model optimized by fruit fly algorithm to predict monthly sales volume of new energy, which has higher accuracy and stability than single model[5].

In order to better capture the composite characteristics of sales data, this paper proposes a Prophet-SARIMA combined forecasting model which can simultaneously fit the periodic and holiday effects. First of all, use prophet model to model and forecast the sales data of the enterprise. Then, the prediction result of SARIMA model is superposed with the result of Prophet model according to the corresponding weight, and the prediction result of combined model is obtained. Finally, a comparative evaluation experiment is designed to realize the prediction of combination model and single model. The experimental results show that the Prophet-SARIMA combined forecasting model can predict the
sales volume of products more accurately and provide more scientific basis for the supplement planning of e-commerce enterprises.

2. Research Theory and Method

2.1. Prophet Model

Prophet is a time series prediction model developed by Facebook, which is very simple and practical. Prophet model can not only fit very fast, but also automatically fill in the missing value, without a lot of data preprocessing. In addition, prophet model can adjust periodicity flexibly, which is suitable for time series data with outliers.

The prediction model of Prophet time series is composed of four parts: trend item, cycle item, festival item and error item. The composition of the model is as follows:

\[ y(t) = g(t) + s(t) + h(t) + e \]  

Formula: \( g(t) \) is the trend function, which is used to fit the non periodic changes in the time series; \( s(t) \) is a periodic term function that fits the periodicity of a week or a year; \( h(t) \) represents the impact of special days such as holidays; \( e \) is the error term, representing the error effect not considered.

\[ g(t) = \frac{C}{1 + e^{-k(t-m)}} \]  

Formula: \( C \) is the saturation value; \( k \) is the growth rate; \( m \) is the bias parameter.

\[ s(t) = \sum_{n=1}^{N} \left[ a_n \cos\left(\frac{2\pi nt}{P}\right) + b_n \sin\left(\frac{2\pi nt}{P}\right) \right] \]  

Formula: \( t \) represents period and \( P \) represents regular period length of time series.

\[ h(t) = Z(t)\kappa_i, \kappa_i \sim \text{Normal}(0, \sigma) \]
\[ Z(t) = [1(t \in D_1),...,1(t \in D_i),...,1(t \in D_L)] \]  

Formula: \( i \) represents holidays; \( D \) represents the collection of past and future holidays; \( K \) represents the impact of each holiday on the forecast.

2.2. SARIMA Model

SARIMA model is also called seasonal difference autoregressive moving average model. SARIMA model is based on ARMA model. Its basic idea is: firstly, test the stability of time series; secondly, eliminate the regularity of non-stationary series with trend and seasonality through several phase by phase and seasonal differences.

3. Experiment and Result Analysis

3.1. Data Overview and Display

The data set used in this paper is the product sales data provided by an e-commerce enterprise, among which the daily sales data of camera products are relatively representative, so the daily sales data of the camera of the enterprise from January 1, 2015 to March 13, 2018 is used. The sales time series chart can more intuitively reflect the data characteristics, and display the sales data according to the granularity of days, as shown in Fig.1. Fig.1 shows that with the passage of time, the camera sales volume of the e-commerce enterprise has a significant upward trend, and the sales volume will increase significantly on the dates of June 18, double 11 and double 12 every year.
3.2. Prophet Model Building

The prediction process of Prophet model is a cycle system which combines analyst and automation process. This combination greatly increases the scope of application of the model and improves the accuracy of the model. The prediction process of Prophet model is shown in Fig.2.

First of all, after analyzing the enterprise camera sales data, this paper selects the piecewise linear growth mode to fit the trend item of the sales data. Then, prophet model can model the impact of holidays. According to the characteristics of e-commerce sales, special dates such as six one eight, double eleven and Double Twelve are added to holiday item h(t).

In order to accurately evaluate the accuracy of the model, the time series data of sales volume from January 2015 to December 2017 is used as the training set, the data from January 2018 to March 2018 is used as the test set, and the prophet model is used to train the training data set.

Fig.3 shows the individual analysis results of the four components in the addition model. It can be seen from Fig.3(a) that the sales volume increased rapidly from 2016 to 2017; Fig.3(b) shows that the sales volume fluctuated significantly during the six one eight, double eleven and Double Twelve activities; Fig.3(c) and fig.3(d) show the annual and weekly changes of sales data respectively.
Comparing the sales forecast data and test data of the best prophet model from January 2018 to March 2018, it is concluded that the RMSE of the model is 3.469 and the MAE is 2.359, and the effect is shown in Fig.4.
3.3. Combination Model Construction

In order to solve the limitations of a single prediction model and improve the accuracy of sales volume prediction, this paper proposes a group and prediction model based on Prophet-SARIMA. The forecasting steps of the combined forecasting model are as follows:

1. Analyze the original sales data and observe the data characteristics;
2. Use the sales data to build the prophet model, by adjusting the parameters, the best prophet model is obtained, and the prediction result $P(t)$ is output;
3. The SARIMA model is constructed by using the sales volume data, and the best single model is obtained by parameter search, and the forecast result $A(t)$ is output;
4. Finally, the prediction value $P(t)$ of Prophet model and $A(t)$ of SARIMA model are added according to different weights to get the prediction result $Y(t)$ of the combined model.

In order to find the best weight distribution of Prophet SARIMA combination model, the weight coefficients are $[0.1,0.9]$, $[0.2,0.8]$, $[0.3,0.7]$, $[0.4,0.6]$, $[0.5,0.5]$, $[0.6,0.4]$, $[0.7,0.3]$, $[0.8,0.2]$ and $[0.9,0.1]$, respectively multiplied by the sales volume prediction results of Prophet model and SARIMA model, and then the results of the two models are added correspondingly to get the final prediction results. RMSE and Mae were selected as the criteria to evaluate the prediction effect of Prophet SARIMA combination model corresponding to each weight coefficient. The results showed that when the weight coefficient was $[0.3,0.7]$, the combination model had the best prediction effect. The RMSE and Mae of the prediction results were 1.961 and 1.478 respectively. The prediction effect diagram of the best combination model is shown in Fig.5.

4. Conclusion

The experimental results show that the RMSE of the best combination model is 1.508 lower than that of Prophet model, and the MAE is 0.881 lower than that of Prophet model. Compared with single model, prophet SARIMA combined forecasting model can effectively improve the prediction accuracy of single model, and has stronger applicability, more suitable for the product sales forecast of e-commerce enterprises, so that enterprises can formulate more scientific replenishment planning, to speed up cash flow and reduce inventory pressure.

In this paper, only the combination model which is weighted by two single models is realized, and different combination methods can be designed in the future. For example, take the fitting data of one single model as the input of another single model, or use two single models to model and predict the
data of different periods. In the future research, we can explore other model combination methods to further improve the prediction accuracy.

Acknowledgments
This work was supported by the Institute of Network and Information Systems of North China Electric Power University.

References
[1] Li Tuo, Xu Chengqian, et al. Sales forecast based on BP neural network[J]. China New Telecommunication, 2020,22(01):137(in Chinese).
[2] Ge Na, Sun Lianying, et al. Forecast and analysis of sales volume based on ARIMA time series model[J]. Journal of Beijing Union University, 2018,32(04):27-33(in Chinese).
[3] Chang Xiaohua, Xiong Ao. Application of Adaboost based random forest algorithm in medical sales forecast[J]. Computer Systems & Applications, 2018,27(02):202-206(in Chinese).
[4] Liu Lu, Ding Fuli, et al. Tobacco sales forecast based on SVM[J]. Software Guide, 2016,15(11):134-137(in Chinese).
[5] Zhou Yanfu, Wang Honglei. Research on monthly sales forecast model of new energy vehicles in China[J]. Software Guide, 2019,18(08):149-153(in Chinese).