Forecasting method of e-commerce cargo sales based on ARIMA-BP model*

Tan Bowen
Chongqing University of Posts and Telecommunications
School of Optoelectronic Engineering City
E-mail: 786633857@qq.com

Zhang Zhe
Chongqing University of Posts and Telecommunications
School of Optoelectronic Engineering City
E-mail: CQUPT_Zhangzhe@163.com

Zhang Yulin
Chongqing University of Posts and Telecommunications
School of Optoelectronic Engineering City
E-mail: 1529965621@qq.com

***Abstract*—In order to prevent merchants from occupying a large amount of cash flow due to unreasonable replenishment and the waste of resources and loss of profits caused by goods in short supply, a reasonable forecast of sales volume is particularly important. This article used the ARIMA-BP non-linear combination model to predict the sale of merchants' goods. Two separate prediction models, the BP neural network and the ARIMA, were used to predict the sales volume in the next 5 days, and then a mean square error model was established to weight the fitting and predictive results of the two single predictions. The weight obtained is brought into the BP neural network model for training, and one is obtained The dynamic weight model is used to predict the dynamic weight of two single prediction models in the next five days. Finally, the test data was substituted into the model, and the normalized error is 0.05363. The research shows that the model can adapt to the problem of forecasting the sales volume of e-commerce goods, more stable, and the error of the combined prediction is much smaller than a single prediction.**

Keywords—ARIMA-BP Nonlinear Combination Forecasting Model, Dynamic Weights, Mean Square Error Model

I. INTRODUCTION

Because of the convenience of online shopping, more and more people use the internet to buy goods, rather than to buy goods offline. In order to make profits and improve the development of e-commerce, it is necessary to predict the sales volume. Actually, if the stock is insufficient, it will result in a phenomenon that the demand exceeds the supply, however, if the stock is excessive, it will cause the phenomenon that the goods is redundant. The problem of e-commerce sales volume is to reasonably predict the future sales volume of the e-commerce through previous sales volume and other data [5-7].

At present, the forecast of e-commerce product sales is mainly solved by establishing a multi-indicator single-item forecasting model. Rong Feiqiong[1] put forward an analysis of online cargo sales analysis based on convolutional neural network, which pointed out that CNN model can get better results in sales forecasting problems. Li Jie[2] proposed a prediction model based on Granger causality test and XGBoost algorithm, which used Granger causality test to extract valid information, and then they used XGBoost algorithm to make predictions. The result showed that the accuracy of prediction can be improved through effective information screening. Wang Jianwei[3] proposed an e-commerce sales model based on e-commerce clustering. That article pointed out that due to the increase of scattered data, product clustering clusters were extracted by using product sales commonality, and then the time series model was used to obtain the prediction results and the hidden Markov prediction model given the probability of the prediction results, which can improve the prediction accuracy.

However, a single-term prediction model only can extract part of the valid information, which will waste much useful information. Hence, some researchers come up with a combination model to solve the problem. Wu Mingshan[4] put the use of the ARIMA-BP model to predict cigarette sales, and achieved good results in the prediction of cigarette sales. Zhang fang[8] constructed a linear combination forecasting model according to the minimum principle of the squared sum of errors of combined forecasting and applied it to sales forecasting, on the basis of two single forecasting models, ARIMA model and BP neural network. The empirical results show that the combined forecasting model can effectively improve the accuracy of sales forecasting and has a good application prospect. Y. Du[9] used the method of ARIMA-BP neural network to predict the future stock price index, which improved the prediction accuracy. Shenjia Ji[10] et al. proposed a new method of power load prediction based on BP neural network and ARIMA combined model, which had better prediction accuracy. Although the combined forecasting model was used, few people considered both return rate and satisfaction as secondary factors to predict sales.

Based on the above researches, in this article, we propose a methodology that using return rate, satisfaction as ARIMA-BP combined model's secondary factors to discuss and analyze the problem of e-commerce sales forecasting, and find that it can be well applied to this problem.

II. PROBLEM DEFINITION

Nowadays, more and more people tend to buy products online rather than offline, because it is more convenient and cheaper. Businesses need to estimate in advance how much demand there will be at that time, and prepare goods in advance to prevent shortages.

E-commerce sales volume problem is a time series prediction problem, through a certain amount of input, to get the future period of commodity sales. In this paper, through the establishment of ARIMA-BP combination model, satisfaction, return rate and sales volume are taken as data input to obtain the sales volume of e-commerce in the future

Authorized licensed use limited to: UEL - Universidade Estadual de Londrina. Downloaded on November 06,2025 at 18:26:40 UTC from IEEE Xplore. Restrictions apply.

period.

Table I is the notation definition for this article:

TABLE I. NEURAL NETWORK PREDICTS SALES

Symbol	Symbolic meaning
μ_t	ARIMA prediction model stationary sequence
w_{ij}	The connection weight of neuron i and neuron j
a_j	The threshold of neuron j
y_i	Prediction calculation value sequence of ARIMA
x_i	The sequence of calculated values of BP prediction model
y_t	Time series predicted by ARIMA
S_k	The combined forecasting model predicts sales over the next five days
g_1	The coefficient of skewness
g_2	Kurtosis coefficient

III. THE ESTABLISHMENT OF E-COMMERCE GOODS SALES MODEL

A. Establishment of ARIMA model

The ARIMA model treats the data sequence formed by the predicted objects over time as a random sequence, and uses a certain mathematical model to describe this sequence approximately. Once this model is identified, future values can be predicted from past and present values of the time series.

After the times difference of the time series, a stationary sequence is generated, and an model is established:

$$u_t = c + \varphi_p u_{t-p} + \varepsilon_t + \theta_q \varepsilon_{t-q} \tag{1}$$

Where c is a constant, θ is a moving average coefficient, φ is an autoregressive model coefficient, ε_t is a white noise sequence, and p is the order of the autoregressive model.

B. Establishment of BP model

Use BP neural network to make predictions and determine the network structure as 3-20-20-20-1. The input layers are time, return rate and satisfaction, and the output layer is sales volume.

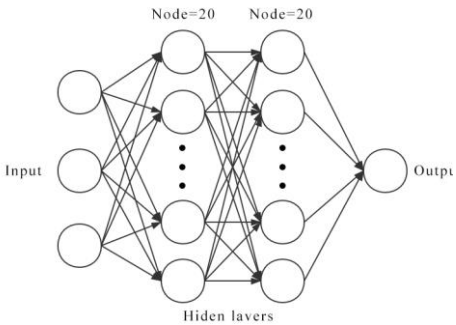

Fig. 1 Neural network model

Constructing the hidden layer transfer function:

$$f(x) = \frac{1}{1+e^{-x}} \tag{2}$$

Artificial neuron output function expression:

$$y_j = f\left(\sum_{i=1}^{n} w_{ij} x_i - a_j\right) \tag{3}$$

Where y_j is the output of neuron j, x_i is the input of neuron i, w_{ij} represents the connection weight of neuron i and neuron j, and a_j represents the threshold of neuron j.

C. Establishment of Nonlinear Combination Model

The above two models explain the connection between sales data to a certain extent, but the single model fails to sufficiently eliminate the missing of the data points, which leads to inaccurate prediction results. Therefore, constructing a nonlinear combination model can overcome the problem of missing data points.

A weight model based on the mean equation is established:

$$a_i = 1 - \frac{(t_i - m_i)^2}{(t_i - m_i)^2 + (t_i - n_i)^2} \tag{4}$$

$$\beta_i = 1 - \frac{(t_i - n_i)^2}{(t_i - m_i)^2 + (t_i - n_i)^2} \tag{5}$$

$$S_i = a_i m_i + \beta_i n_i \tag{6}$$

Where $i = 1, 2, 3 \ldots n$, m_i is a sequence of ARIMA calculated values, n_i is a sequence of BP calculated values, t_i is a time series, a_i, β_i are the weights of two single prediction models, The model shows that prediction models with large errors have less weight, which can further improve prediction accuracy. Bring these two sequences a_i, β_i into the BP neural network for training to gain a dynamic weight model, and then use this model to obtain the weights a_k, β_k, which are the two single prediction models for the target days. The ARIMA model and BP neural network model respectively predict the target days of the sales volume m_k, n_k $(k = 1, 2, 3 \ldots)$. Finally, add the two parts to get the final forecast sales:

Authorized licensed use limited to: UEL - Universidade Estadual de Londrina. Downloaded on November 06,2025 at 18:26:40 UTC from IEEE Xplore. Restrictions apply.

$$S_k = a_k m_k + \beta_k n_k \left(k = 1,2,3...\right) \tag{7}$$

At the same time, the performance test function is constructed to test the prediction error of ARIMA model and BP neural network model by means of mean square deviation.

$$E\left(x\right) = \int_{-\infty}^{\infty} xf\left(x\right)dx \tag{8}$$

$$D\left(x\right) = E\left[x - E\left(x\right)\right]^2 \tag{9}$$

$$MSE\left(\hat{x}\right) = D\left(\bar{x}\right) + \left[E\left(\hat{x} - x\right)\right]^2 \tag{10}$$

IV. Simulation and Prediction Forecast

A. BP neural network forecast

Add up the sales of all the goods in each day, which are the sales of the goods on that day. Normalize time and sales:

$$n_k = \frac{n_k - n_{\min}}{n_{\max} - n_{\min}}\left(k = 1,2,3...\right) \tag{11}$$

Where $x_{\min}$ is the minimum value in the data series and $x_{\max}$ is the maximum value in the data series.

Through the training of sample data, the network weights and thresholds are continuously modified to make the error function continuously decrease to a negative gradient, thereby approaching the expected output.

Use MATLAB to get the predicted sales volume and corresponding errors in the next 5 days, as shown in Table I.

TABLE II. Neural network predicts sales

Number of days	The neural network predicts sales
1	199.6784427
2	344.8687954
3	344.8687954
4	7656.115797
5	1047.350682

The comparison between the BP neural network fitted image and the original image is shown in Figure 2.

It can be seen from Figure 2 that the function image fitted by BP neural network is not accurate enough with the original image. By calculating formulas (8) (9) (10), the error of fitting the BP model can be calculated as 0.56415.

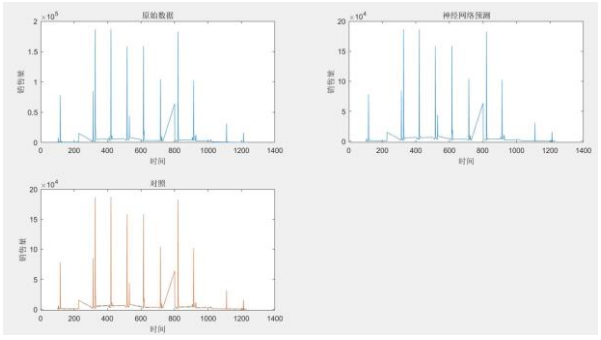

Fig .2 Comparison of neural network fitted image and original image

B. ARIMA Forecast

In an autoregressive model such as ARIMA, the model needs the stability of time series data. Therefore, it is necessary to perform a stationary test on the data or the nth order difference of the data, and use the autocorrelation graph and partial autocorrelation graph of the data to select partial $\left(p,q\right)$ for AIC verification. When $p = 5, q = 5$ the AIC is the smallest, and the error is the most accurate. Finally, the predicted sales volume for the next 5 days is shown in Table II.

TABLE III. ARIMA Forecast Sales

Number of days	ARIMA predicts sales
1	295
2	307.4836
3	328.1062
4	7768.689
5	1007.833

Fit the data by using the ARIMA(5,0,5) model as shown in Figure 3.

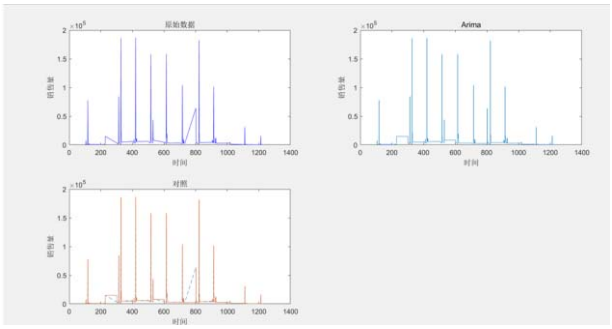

Fig. 3 Comparison between ARIMA fitting image and original image

According to formulas (8) (9) (10), the error of the ARIMA prediction model is 1.67211.

C. ARIMA-BP combination forecast

Take the sequences of a_i , β_i into the BP neural network for training to obtain a dynamic weight model, and then use this model to obtain two single prediction models' the weights a_k , β_k in the next 5 days. The results are shown in Table III.

Authorized licensed use limited to: UEL - Universidade Estadual de Londrina. Downloaded on November 06,2025 at 18:26:40 UTC from IEEE Xplore. Restrictions apply.

TABLE IV. COMPARISON OF ARIMA FITTED IMAGE WITH ORIGINAL IMAGE

Number of days	BP forecast model weights	ARIMA predictive model weights
1	0.960331629	0.339668371
2	0.958528286	0.041471714
3	0.970019666	0.029980334
4	0.986029772	0.013970228
5	0.99625639	0.0037461

The combined model predicts sales as shown in Table IV.

TABLE V. THE COMBINED MODEL PREDICTS SALES

Number of days	The combination model predicts sales
1	203.4597
2	343.3184
3	344.3662
4	7657.688
5	1047.203

The function chart of the combined prediction model is shown in Figure 4.

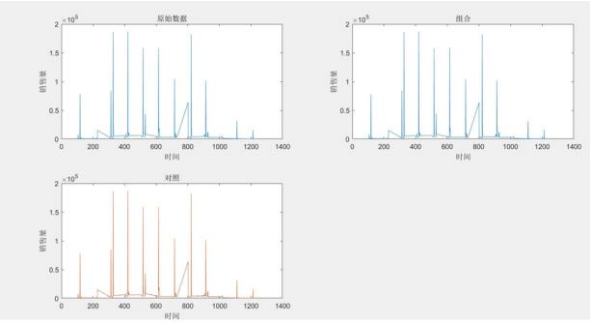

Fig. 4 Comparison of the combined model fitted image and the original image

As can be seen from Figure 4, compared with the ARIMA model and the BP neural network model, the combined prediction model has a more accurate prediction result, and the error of the combined prediction model is calculated as 0.05363, the experimental results are better, which provides a reliable solution for the prediction of e-commerce sales volume Conclusion.

V. CONCLUSION

In this paper, ARIMA-BP combined forecasting model is used to analyze the future sales volume of e-commerce, and it is found that it can be well applied to the future sales volume prediction of e-commerce. The two single models lack complementarity and can effectively extract and analyze feature information. The test data was substituted into the training model, and the error was found to be 0.05363. Finally, the calculation results can well prove the applicability of the ARIMA-BP combination model in the e-commerce sales forecast problem.

REFERENCES

[1] Rong Feiqiong, Guo Mengfei. Research on Online Product Sales Forecast Analysis Based on Convolutional Neural Network [J]. Journal of Northwest University for Nationalities (Philosophy and Social Sciences Edition), 230(02):21-32.

[2] Li Jie, Wang Yuxia, Zhao Xudong. Prediction Methods for Commodity Sales of E-commerce Enterprises [J]. Statistics and Decision, 2018 (22).

[3] Wang Jianwei. E-commerce sales forecast based on commodity clustering [J]. Journal of Computer System Applications, 2016, 25 (10).

[4] Wu Mingshan, Wang Bing, Kiyaning, et al. Research on Cigarette Sales Combination Forecasting Model [J]. China Tobacco Journal, 2019 (3).

[5] D. Wei, P. Geng, L. Ying and L. Shuaipeng, "A prediction study on e-commerce sales based on structure time series model and web search data," The 26th Chinese Control and Decision Conference (2014 CCDC), Changsha, 2014, pp. 5346-5351.

[6] Bandara K , Shi P , Bergmeir C , et al. Sales Demand Forecast in E-commerce using a Long Short-Term Memory Neural Network Methodology[J]. 2019.

[7] Li M , Ji S , Liu G . Forecasting of Chinese E-Commerce Sales: An Empirical Comparison of ARIMA, Nonlinear Autoregressive Neural Network, and a Combined ARIMA-NARNN Model[J]. Mathematical Problems in Engineering, 2018, 2018:1-12.

[8] Zhang Fang, Zhou Zongfang. Research on Sales Portfolio Forecast Based on ARIMA Model and BP Neural Network [J]. Management Scientist: Academic Edition (04):38-43+80.

[9] Y. Du, "Application and analysis of forecasting stock price index based on combination of ARIMA model and BP neural network," 2018 Chinese Control And Decision Conference (CCDC), Shenyang, 2018, pp. 2854-2857.

[10] Shenjia Ji, Hongyan Yu, Yinan Guo, and Zongrun Zhang. 2016. Research on sales forecasting based on ARIMA and BP neural network combined model. In Proceedings of the 2016 International Conference on Intelligent Information Processing (ICIIP '16). Association for Computing Machinery, New York, NY, USA, Article 41, 1–6.

Authorized licensed use limited to: UEL - Universidade Estadual de Londrina. Downloaded on November 06,2025 at 18:26:40 UTC from IEEE Xplore. Restrictions apply.