Research on Forecasting Model of New Retail Sales Volume Based on BP Artificial Neural Network and RBF Neural Network Algorithm

Longhao Pang [1,*]

[1] *School of Mathematics and Statistics, Wuhan University, Wuhan, Hubei, 430072*
[*] *Corresponding author. Email: panglonghao1027@163.com*

ABSTRACT
With the continuous development of China's consumer market, the production mode of new retail enterprises is gradually moving forward to multi-variety and small batch. Therefore, the importance of accurate demand forecast for sales categories is self-evident. First of all, this paper makes an analysis and judgment on the main factors affecting sales volume through Pearson correlation analysis and grey correlation analysis. Then, considering the requirements of some factors, the BP artificial neural network model is used to predict the monthly sales of the target class accurately and quickly. Finally, under the requirements of comprehensive and in-depth consideration of various factors, the sales volume of the target subcategory is refined, and the weekly sales of all skc in the subcategory are predicted by the PBF model. In this paper, an analysis model of influencing factors of sales volume based on Pearson correlation analysis and grey correlation analysis is established, and the significance of the results is tested. Then the artificial neural network model is established, and the BP algorithm is used to predict the sales volume of the target subclass. Finally, the RBF neural network model is used to predict the sales volume of skc in a specified period of time. The two elements of subclass level and skc level are comprehensively considered, which is consistent with the expected goal.

Keywords: *Pearson correlation analysis, grey correlation analysis, BP artificial neural network, RBF neural network*

1. INTRODUCTION

Consumer market refers to the exchange field or place of consumer goods (including labor services). It is the sum of the relationship between supply and demand of consumer goods. It is based on commodity production and commodity exchange. With the continuous development of China's consumer market, the production mode of new retail enterprises is gradually moving forward to multi-variety and small batch. Commodity producers and consumers have to exchange goods and services through the market in order to meet their own material and cultural needs. Therefore, this paper makes an in-depth discussion on the new retail model and enterprises.

2. PEARSON CORRELATION ANALYSIS AND GREY CORRELATION ANALYSIS OF SALES VOLUME INFLUENCE MODEL

2.1. Data preprocessing

This paper collects the relevant data sets of the sales process (merging the same number of skc and summing), and deletes the unsuccessful data due to the lack of time in the process of data statistics. And the data (label price, average price, discount, average price) are averaged before grey relational analysis[1].

$$x_i(k) = \frac{x_i(k)}{x_i}, k = 1,2,\ldots,n, i = 0,1,2,\ldots,m \quad (1)$$

k corresponds to the time period, and i corresponds to a certain component of the comparison sequence.

2.2. Data preprocessing

The correlation coefficient is expressed as:

$$r_{X,Y} = \frac{\Sigma\left(X-\bar{X}\right)\left(Y-\bar{Y}\right)}{\sqrt{\Sigma\left(X-\bar{X}\right)^2 \Sigma\left(Y-\bar{Y}\right)^2}} \quad (2)$$

X, Y is the random variable obtained through initial valuation and reflects the original information of the sample; $\bar{X},\bar{Y}$ is the random variable obtained after the variable is averaged and reflects the average value of the sample. $\Sigma\left(X-\bar{X}\right)\left(Y-\bar{Y}\right)$ represents the covariance of two variables. R is the correlation coefficient between variables, and the value range is [-1,1].

The statistical formula of significance test of correlation analysis is as follows:

$$t = r \cdot \sqrt{\frac{n-2}{1-r^2}} \quad (3)$$

n is the data sample size, r is the calculated Pearson correlation coefficient, and t is the T distribution conforming to the degree of freedom n-2, which is used for significance test to test whether there is correlation between variables.

2.3. Solution of model

Using the selected four parameters and sales volume data, in the SPSS environment, the scatter diagram of sales volume and influencing factors was first made to investigate the linear relationship between the two, and the normality test of the selected parameters was carried out. The results met the requirements of correlation analysis, and then the correlation coefficient between each variable was obtained by inputting five groups of data . Significance tests were then performed to ensure that the data were correlated.

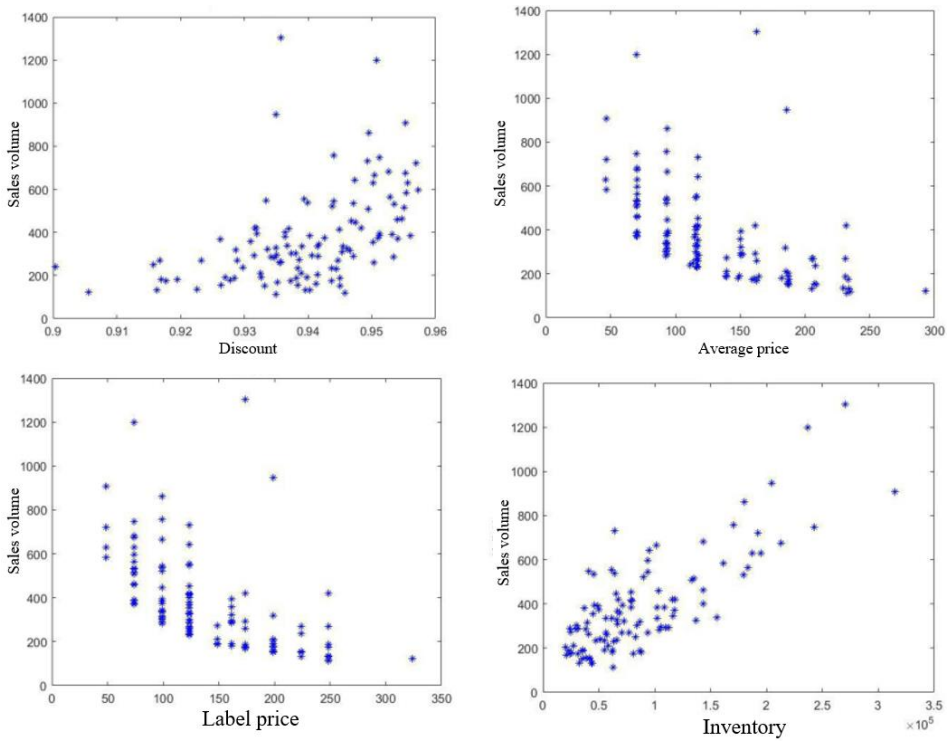

Figure 1 Correlation analysis sales volume -- influencing factors scatter chart

As can be seen from the aggregation of data points on the scatter chart, discount, inventory and sales volume have a good positive linear correlation, while average price, tag price and sales volume have a good negative linear correlation[2].

Then the correlation coefficient and significance level of each factor were obtained, and it was found that the sig value was all at 0.01 level, indicating that the correlation was significant. The absolute value of the correlation coefficient is basically above 0.5, indicating that there is a strong correlation between the influencing factors and sales volume. The correlation coefficients between sales volume and label price, inventory, discount and average price were-0.554, 0.794, 0.468 and 0.553, respectively.

In this paper, correlation analysis is carried out to obtain the qualitative relationship between influencing factors and sales volume, and grey correlation analysis is used to describe and compare the degree of influence quantitatively. The reference sequence and comparison sequence are extracted, and the correlation coefficients of each index data are calculated after the comparison sequence is averaged, and then the grey correlation degree of each influencing factor is obtained.

Inventory	Discount	Average price	List price
0. 846998662	0. 350888268	0. 37497219	0. 375122707
0. 949066184	0. 467794985	0. 549246046	0. 549914306
0. 821939038	0. 917680743	0. 671351649	0. 668177668
0. 530875741	0. 589341774	0. 56069019	0. 557924449
0. 787002929	0. 380491046	0. 334687809	0. 334027469
0. 90454777	0. 510671957	0. 458469159	0. 456960537
0. 70120654	0. 657624604	0. 622217109	0. 619337859
0. 986396831	0. 571120122	0. 506940497	0. 505895039
0. 810579779	0. 917508739	0. 918972271	0. 914604732
0. 642040074	0. 73894525	0. 694886683	0. 694014663
0. 578672665	0. 742867358	0. 698622853	0. 699834157
0. 872937507	0. 832434487	0. 560909278	0. 557365277
0. 703903603	0. 638511909	0. 558877223	0. 556492976
0. 872048468	0. 965524213	0. 929727313	0. 925857391
0. 876486704	0. 903182655	0. 703200681	0. 698581625

Figure 2 Grey correlation analysis of influencing factors correlation coefficient

From the results of correlation analysis combined with correlation analysis, it can be seen that inventory and discount are positively correlated, label price and average price are negatively correlated with sales volume, and the degree of influence is inventory, discount, average price, label price[3].

3. NATURAL GAS LOAD FORECASTING MODEL BASED ON BP ARTIFICIAL NEURAL NETWORK MODEL

3.1. Data preprocessing

In the input layer, normalize the value of sales volume to [0.1,0.9]

$$x'_{input} = 0.1 + 0.8\frac{x_{input} - x_{\min}}{x_{\max} - x_{\min}} \tag{4}$$

In the output layer, the sales volume is inversely normalized as follows:

$$x_{output} = x_{\min} + \frac{(x_{\max} - x_{\min})(x_{output} - 1)}{0.8} \tag{5}$$

3.2. Establishment of the model

Artificial neural network structures include input layer, hidden layer and output layer, each layer has a number of neurons, including any neurons and by weighting the next layer of arbitrary neurons and using BP (error back propagation) algorithm, according to the error of the network output and the actual output, to adjust the weights between neurons, eventually making minimum output.

(a) The connection weights of hidden nodes and output nodes are given to random small quantities in the interval (-1,1).

(b) Take a sample from $(x_j, y_j), j = 1,2,\cdots,P$, training sample set, input its signal x_j into network learning training, and indicate its expected output y_j.

(c) Activation function (Sigmoid-type function) $f = \frac{1}{1+e^{-x-v}}$ is performed on both the hidden node and the output node, and the actual output value of the network is finally obtained from the output node, where V is the threshold value of the activation function.

(d) Repeat the above steps for each sample in the training sample set until all learning is completed, and calculate the squared error.

(e) If $E < \delta$ (specify precision), the adjusted weight value will be output at the end of training, otherwise, the next training will be carried out.

3.3. Solution of model

Table 1. Input and output neuron structure scheme of artificial neural network

Input neuron	Corresponding input value
1	Forecast sales in the month before the forecast period
2	Inventory forecast for the previous month
3	The average price of the month before the forecast period
4	Forecast period prior to the month discount
5	Forecast sales in the previous month
6	Monthly sales of forecast period
7	Forecast period monthly inventory
8	Forecast period monthly average price
9	Forecast period monthly discount
Output neuron	Corresponding output value
1	Forecast period sales

The BP artificial network model with momentum-adaptive learning rate adjustment algorithm is satisfactory in both the convergence speed and the convergence accuracy, and it is a reliable model from the relative error table of the final simulation prediction. However, because the model is a black box model, the internal relationship of data is unknown. As a result, the model can only be improved from the data specifications and THE BP algorithm itself, and it is difficult to modify the model according to the facts, so it is difficult to make a new qualitative leap after the model reaches a certain accuracy.

4. SKC SALES FORECASTING MODEL BASED ON RBF ARTIFICIAL NEURAL NETWORK MODEL UNDER SMALL CLASS

4.1. Initialization Set

Without loss of generality, the initial value of the center component of each neuron in the hidden layer is changed with equal spacing from small to large, so that the weak input information can produce strong effects near the small center, and the center parameters of each neuron are initialized as follows [5].

$$c_{ji} = \min i + \frac{\max i - \min i}{2p} + (j-1)\frac{\max i - \min i}{p} \tag{6}$$

Where P is the total number of neurons in the hidden layer, mini is the minimum value of all input information of the first feature in the training set, and maxi is the maximum value of all input information of the first feature.

The weight initialization method from the hidden layer to the output layer is given by referring to the center initialization ground method. The weight initialization method of each unit connection weight is given.

$$W_{kj} = \min k + j\frac{\max k - \min k}{q+1} \tag{7}$$

mink is the minimum value of all expected outputs in the KTH output neuron in the training set. maxk is the maximum value of all expected outputs in the KTH output neuron in the training set.

The width vector that affects the scope of the neuron's action on the input is initialized as follows:

$$d_{ji} = d_f\sqrt{\frac{1}{N}\sum_{k=1}^{N}\left(x_i^k - c_{ji}\right)} \tag{8}$$

4.2. Establishment of the model

The structure of artificial neural network includes input layer, hidden layer and output layer. Each layer has several neurons, and any neuron is connected with any neuron of the next layer by weight. RBF network training process is divided into unsupervised learning to determine the weight Cj and width vector Dj between the input layer and the hidden layer, and supervised learning to determine the weight Wj between the hidden layer and the output layer. According to the error between network output and actual output, the weight between each neuron is adjusted to make the mean square error root RMS minimum.

4.3. Solution of model

Under the small target class hierarchy all SKC sales forecast, sales characteristics has four as independent variables, as the dependent variable has a sales, so RBF neural network model of the number of neurons in the input is 4, the number of neurons in the output of 1, and the number of neurons in hidden layer will be the middle boundary is adaptively in the process of training[6].

Using MATLAB neural network toolbox, a radial basis neural network is established, gaussian function is selected as the basis function, k-means clustering method is used to determine the basis function center, and the width of the basis function is taken as the average distance between the basis function center and the sample mode in the subsample set.

5. CONCLUSION

First of all, this paper analyzes the factors affecting sales based on Pearson correlation analysis and grey correlation analysis, and tests the significance of the results. The artificial neural network model and BP algorithm are used to predict the sales volume of the target subclass. Finally, in view of the huge amount of data to be processed for all skc under the target subclass, and considering that the structure of RBF neural network model is simple, the training is simple, the learning convergence is fast, and the approximation ability and learning speed are better than BP neural network, the model is used to predict skc sales in a specified period of time. The prediction accuracy of sales volume has been improved, which is consistent with the actual situation. In this paper, the sales volume of new retail goods is predicted through the exploration of the influencing factors of sales volume and the prediction of CP algorithm, and the supply and marketing data with large amount of data can be further optimized and improved based on convolution neural network.

REFERENCES

[1] Mei Zhengyang, HAN Zhibin. Mathematical Modeling Tutorial [M]. Beijing: Science Press,2012.

[2] SUN Xiaojun, Jie Kewei. Research on the influencing factors of college students' well-being: Based on Pearson Correlation Analysis and Cumulative Logistic Regression Model [J].Journal of capital normal university,10(5):14-19,2018.

[3] Huang yue-rong, xu xiao-zhong, zhang yi-ming,etal. Prediction of Urban Gas Load Based on ARIMA and Neural Network Model [J]. Computer Application and Software,28(12):206-209,2011.

[4] JIN Renjie. Performance Comparison and Application of Momentum-adaptive Learning Rate BP Algorithm and BP Algorithm in Neural Networks [J].

[5] Si Shoukui, Sun Zhaoliang. Mathematical Modeling Algorithm and Application [M]. Beijing: National Defense Industry Press,2015.

[6] JIANG Jingqing, SONG Chuyi, LIU Naren, ZHANG Pinyi. Training Method and Analysis of RBF Neural Network [J]. Journal of Inner Mongolia University for Nationalities,18(4):301-303,2003.