Research on Modern Higher Education 2, 01008 (2017)
DOI: 10.24104/rmhe/2017.02.01008

Study on inventory control model based on the B2C mode in big data environment

Zhiping Zhang*
Department of Communication Engineering, Shijiazhuang Post and Telecommunications Technical College, Shijiazhuang, Hebei, China

Meimei Zhang
School of Economics and Management, North China Electric Power University, Baoding, Hebei, China

Zixian Zhang
Network Institute, Shijiazhuang Post and Telecommunications Technical College, Shijiazhuang, Hebei, China

ABSTRACT: The current inventory problem has become the key issue in the enterprise survival and development. In this paper, we take "Taobao" as an example to conduct a detailed study of the inventory of the high conversion rate based on data mining. First, by using a funnel model to predict the conversion of the commodities on the critical path, we capture the factors influencing the consumer decision-making on each key point, and propose corresponding solutions of improving the conversion rate; Second, we use BP neural network algorithm to predict the goods traffic, and then obtain the corresponding weights by the relation analysis and the output of the goods traffic by the input of large data sample goods; Third, we can predict the inventory in accordance with the commodity conversion rate and flow prediction, and amend the predicted results to get accurate and real-time inventory forecast, avoiding the economic loss due to the inaccurate inventory.

Keywords: big data; BP neural network algorithm; conversion; data mining; funnel model; inventory

1 INTRODUCTION

In the field of B2C e-commerce, the merchandise inventory has always been a difficulty for the dealers. In the supply chain operations, the big data from retail terminals, after-sales service providers, distributors, transporters, manufacturer and supplier provide the data environment for business applications in the B2C field[1~5].The big data can be used to correct and optimize inventory.

To conclude the inventory forecast also needs to be based on conversion rate, and goods traffic (traffic) is the basic of goods conversion, so the goods traffic prediction has become another important indicator of stock prediction.

2 CONVERSION FACTORS ANALYSIS UNDER THE B2C MODE

The conversion rate refers to the proportion of persons that click from one page to another page. For example, the number of visitors who enter the home page of a shopping site is 50,000; and the number of these visitors click the commodity detail page through an AD Home page is 50; the conversion rate from home page to this commodity page is 50/50,000 = 0.1%. It is of significance to grasp the conversion rate of the critical path [6~7]. Under the B2C mode the shopping critical path is: (1) browse the product page→ (2) place the product into the shopping cart or the favorites → (3) generate the order → (4) pay for the order → (5) complete the transaction. From the number of each step web access, the statistics of the conversion rate in the 5 steps (one site as an example) are as shown in Table 1:

*Corresponding author: 18932915732@189.cn

Table 1. Calculate the critical path conversion

	Browse Goods	Shopping carts	The order	Payments	Transaction completed
The number of people	3042	788	362	288	284
Conversion in this step	100%	25.9%	45.9%	79.6%	98.6%
The total conversion	100%	25.9%	11.9%	9.5%	9.3%

In general, the relevant factors affecting the conversion rate can be obtained by correlation analysis on the critical path correspondingly and then we can propose subsequent strategies to improve the conversion rate based on the relevant factors.

To improve the conversion rate needs to analyze the relevant factors that affect its conversion. The deals can dig out the factors which affect consumers purchasing decision-making from the product page. The main factors include: shops attraction, goods attraction, quality of service online, and businesses integrity, under which also includes a number of specific factors [8~10].

The weights of the influencing conversion factors should be judged by association rules to analyze the key factors on the critical path. We gathered some samples in Taobao help center to complete the missing corresponding factors data. And the corresponding value of 1 show “Yes”, the corresponding value of 0 shows “No”. The relevant data are as shown in Table 2.

We use SPSS19.0 to make association rules analysis [11~13] for the relevant factors sample data from Table 2. As the conversion rate ultimately reflected in the sales volume, we set the sales volume as dependent variable. And we set store reputation, shop collections and other factors in Table 2 as the variables. With using the vicariate analysis to conduct the correlation analysis between the two factors, we only need to get the correlations between the sales volume and other influencing factors in a symmetric matrix format, thus obtaining the correlations of the influencing factors on the conversion rate. The processing results are as shown in Table 3.

Based on the correlation analysis we can obtain Pearson correlation between each factor, and then we aggregate the correlation coefficient results to determine the weight of each factor. With adding up the absolute value of the correlation coefficients, we can get:

W=0.617+0.512+0.723+0.861+0.578+0.120+...+0.51

+0.471+0.411=10

Then we calculate the correlation of each factor, and obtain the weight of the factor. For example, the calculation process of the weight of the store credit is as shown below:

W_1=0.617/10=6%, it means that adjustment coefficient of shop reputation is 6, which accounts for 6% of all the influencing factors.

The correlation proportions of other factors can be calculated by this method correspondingly. In this paper, the influencing factors weighs over 5% will be used to predict the goods traffic. With the prediction of goods traffic under the prediction model, then we can implement the inventory gross prediction.

3 THE REALIZATION OF GOODS INVENTORY MODEL

3.1 *Goods traffic prediction*

3.1.1 *Brief introduction of BP network algorithm*

The basic unit of information processing in neural network is the neuron, which is also the basis of neural network design. The neuron model is mainly composed of three elements with a synapse, adder and the activation function [14]. In general, the neuron output range is [0, 1] and [-1, 1], Beyond this region will show abnormal. External offset *j* is another important indicator of neuron model. If the corresponding offset is positive, the input of network will increase to meet the stress activated function validation constraints, and if the corresponding offset is negative, the input of network will decrease to comply with the limits of the activation function verification, which can be described by the formula (1).

$$\begin{aligned} U_k &= \sum_{j=1}^{n} W_{kj} X_j \\ y_k &= f(u_k - \theta_j) \end{aligned} \tag{1}$$

In this formula, X_1, $X_2 \dots X_n$ is neuronal input signal, W_{k1}, $W_{k2} \dots W_{kn}$ is the synaptic weights of neuron K; θ_j is the limit value, $f(u_k-\theta_j)$ is the used activation function . The methods which are commonly used to activate the function in neural network prediction are: limit function, piecewise linear function, s-shaped function.

3.1.2 *BP algorithm design* [15~19]

Assuming that commodity projections of BP neural network has *n* neurons with arbitrary multi-layers, then the corresponding input layer neurons and their input and output are the same. What’s more, the operating characteristics of the hidden layer neurons and their output layer neurons are as shown in formula (2), (3).

$$\begin{cases} net_{jk} = \sum_{i=1}^{n} w_{ji} o_{ik} \\ o_{ik} = f(net_{jk}) \end{cases} \tag{2}$$

$$E_k = \frac{1}{2} \sum_{j=1}^{n} (y_{jk} - O_{jk}) \tag{3}$$

Table 2. Relational data samples

Shop reputation	4	12	8	6	15	4.9	Delivery speed	4.9	4.8	4.7	4.5	4.9	4.9
Shop set as favorites	1823	26839	16433	3278	31687	4.6	Speed of logistics	4.3	4.8	4.2	5	4.6	4.6
Graphic details	376	16548	896	635	3264	3.8	required amount of consumption with free gifts	3.8	4.5	3.9	4.2	3.8	3.8
Decoration templates of the well-managed shops	71	468	328	763	846	8	Marketing Packages	5	6	3	5	8	8
Dynamic score of the shops	8743	5693	4682	1633	15436	6	Credit card payment	1	1	0	0	1	1
Collection popularity of the items	4.8	5	4.6	4.9	5	5	VIP discount	2	3	3	4	5	5
Cumulative sales of merchandise	4.5	5	5	4.6	4.8	5	coupons	4.6	4.66	4.83	4.7	5	5
Goods visited	4.82	4.86	4.73	4.64	4.9	5	VIP store	3	4	5	4	5	5
Various activities in Taobao	3	4	2	4	6	6	Wangwang online	0	1	1	1	1	1
No reason refund in 7 days	1	1	1	0	1	1	Cash on delivery	1	1	1	1	1	1
Shopping Mall	0	1	0	0	1	1	Protection for consumers	1	1	0	1	1	1
Shop service	4.9	4.8	4.6	4.7	5	1	Fake one compensate three	1	1	1	1	1	1

Table 3. SPSS Correlation

		Cumulative sales of merchandise	Shop reputation	Shop set as favorites	collection popularity of the items	Goods visited	Graphic details	decoration templates of the well managed shops	dynamic score of the shops
	Pearson Correlation	1	0.617	0.512	0.723	0.861	0.578	0.12	0.405
		Cumulative sales of merchandise	Cash on delivery	Credit card payment	Shopping mall	Taobao shop service	Delivery speed	Logistics speed	required amount of consumption with free gifts
Cumulative sales of merchandise	Pearson Correlation	1	-0.617	0.145	0.465	0.171	0.171	0.641	0.184
		Cumulative sales of merchandise	Marketing packages	Coupons	The VIP store	Protection for consumers	No reason refund in 7 days	Fake one compensate three	Wangwang online
	Pearson Correlation	1	0.589	0.627	0.572	-0.61	0.51	-0.471	0.411

E_k is the error between the sample outputs of the k-th commodity and its corresponding vector; y_{jk} is the corresponding expectation value of the sample output neuron of s the j-th commodity; O_{jk} is the corresponding actual value of the sample output neuron of the j-th commodity. Suppose $E_k=\sum E_k$ is the sum of the commodity in the training set to predict the output error for all samples generated. ΔkW_{ji} is adjusted value for W_{ji} , and order

$$\Delta k W_{ji} = -\frac{\partial E}{\partial W_{ji}} \tag{4}$$

by (2),

$$\frac{\partial net_{jk}}{\partial w_{ii}} \frac{\partial}{\partial w_{ii}} (\sum w_{jb} O_{bk}) = O_{ik} \tag{5}$$

$$\frac{\partial E_k}{\partial w_{ji}} = \frac{\partial E_k}{\partial net_{jk}} \frac{\partial net_{jk}}{\partial w_{ji}} \tag{6}$$

Suppose σ_{jk} is a definition of the error signal, which is calculated as follows:

$$\sigma_{jk} = -\frac{\partial E_k}{\partial net_{jk}} \tag{7}$$

Then,

$$\frac{\partial E_k}{\partial W_{ji}} = -\sigma_{jk} O_{ik} \tag{8}$$

To decrease the error E along with the gradient direction, we need to adjust the weights in accordance with the (9

$$\Delta k W_{ji} = \eta \sigma_{jk} O_{ik} \tag{9}$$

To get the most available differential equations, we do not take the change of σ_{jk} into consideration and calculated by the partial differential equation:

$$\sigma_{jk} = -\frac{\partial E_k}{\partial net_{jk}} = -\frac{\partial E_k}{\partial O_{jk}} \frac{\partial O_{jk}}{\partial net_{jk}} \tag{10}$$

From the (2), it can be concluded that

$$\frac{\partial O_{jk}}{\partial net_{jk}} = f(net_{jk}) \tag{11}$$

And the (10) in the first item $\partial E_k/\partial O_{jk}$ must be divided into two kinds of situations:

If j is the associated network output neuron, from the definition of E_k, the conclusion can be obtained:

$$\frac{\partial E_k}{\partial O_{jk}} = y_{jk} - O_{jk} \tag{12}$$

Then,

$$\sigma_{jk} = (y_{jk} - O_{jk}) f(net_{jk}) \tag{13}$$

If j is related to the hidden layer neurons

$$\sum_m \frac{\partial E_k}{\partial net_{mk}} \frac{net_{mk}}{\sigma O_{jk}} = \sum_m \frac{\partial E_k}{\partial net_{mk}} \frac{\partial}{\partial O_{jk}} \{W_{mi} O_{jk}\} = \sum_i \sigma_{mk} W_{mj} \tag{14}$$

Put the results into (10), we can draw that

$$\sigma_{jk} = f(net) \sum_m \sigma_{mk} W_{mj} \tag{15}$$

Put the (13) into (14), $f(net_{jk})$ will be expressed as a function of O_{jk}, put the (11) into (13), we can get the differential equations required by BP algorithm network training on the computer:

$$f(net_{jk}) = O_{jk}(1 - O_{jk}) \tag{16}$$

3.1.3 *Measure of prediction accuracy*

Indicators to measure the accurate prediction are as follows: the average error MD, Standard error of S, absolute percentage error.

If the mean absolute percentage error is between 18% and 48%, the prediction can be considered feasible, if less than 18%, the prediction can be considered good, and the results can be used to estimate the sample values [20].

3.1.4 *Data collection and processing*

Take Taobao brand "oak Philippines" for example, methods for data acquisition are: First, write an access code to access API on Taobao Open Platform, then Cumulate the sales of single items, the existing inventory, store reputation, dynamic evaluation scores and other data can be obtained; Second, write similar Web crawler tool to crawl the front page data which is not open, such as collection frequency, the number of orders for the unconverted and the number of transactions. Third, describe the ranking rules of Taobao shop and commodities, and other related weighting factors with the help of Taobao help documentation.

3.1.5 *MATLAB realize goods traffic forecast*

Through the analysis of the data Taobao, we can draw the score of various factors which affecting the flow of goods ratings as the training set *X*train. While the corresponding initial weights W will be learned repeatedly to obtain the squared error after the weight correction. Then observe whether the squared error is below the desired standard deviation, if so, end the training, if not, continue to learn until the squared error is below the desired standard deviation. Now we take the following five indicators of 13 Philippine rubbers stored in Taobao Mall for training input: they are shop grade, cumulative sales, merchandise collections popularity, existing inventory and dynamic scoring.

This paper conduct the BP neural network simulation study with MATLAB [21-23].Assume that the first layer nodes is 20, the second layer nodes is 40, the measure accuracy is 0.0001, training times are 50000. Final prediction: *O* represents the predicted value, + indicates the actual value. The codes to realize BP neural network algorithm with MATLAB simulation are as follows.

```
>>%===Five factors data input of 13 commodities ===
P= [8 13539 11 6902 4.73; 3 24262 36 5264 4.85; 4 279 1 251 4.66; 4 89 5 803 4.76…
1 53 1 192 4.76; 4 863 90 75 4.80; 4 806 82 106 4.70; 4 55 8 216 4.79; 3 71 22 945 4.78…
2 23 3 136 4.64; 4 102 38 481 4.81; 4 123942 82 951 4.86; 4 3672 34 421 4.82]';
%===desired output Views of 13 commodities ===
t=[352772 427427 11515 3597 467 689 348 467 1205 237 1495 46985 895];
ptest-[1 106 41 539 4.78]';% Predictive input
[pn,minp,maxp,tn,mint,maxt]=premnmx(p,t);%normalize the data
NodeNum1=20;% The first layer of hidden nodes
NodeNum2=40;% The second layer of hidden nodes
TypeNum=1;% Output dimension
TF1='tansig';
TF2='tansig';
TF3='tansig';
net=newff(minmax(pn),[NodeNum1,NodeNum2,TypeNum],{TF1TF2TF3},' traingdx');%Network creation trained
net.trainParam.show=50;
net.trainParam.epochs=50000;%Setthe number oftraining
net.trainParam.goal=le-5;%achieve the accuracy of training net.trainParam.lr=0.01;% Learning rate
net=train(net,pn.tn);
p2n=tramnmx(ptest,minp,maxp);%Testdatanormalization
an=sim(net,p2n);
[a]=postmnmx(an,mint,maxt);%Anti-normalized data, That is, ultimately desired results
plot(1:length(t),t,' o' ,1:length(t)+1,a,' ⊕ ' );
title( 'O represents the predicted value---⊕ indicates the actual value' );%put out the prediction
grid on
```

```
m=length(a);% The length of the vector a
tl=[t,a(m)];
error=tl-a;% Error vector
figure
plot(1:length(error),error,'-.');
title('Error variation')
```

Sample sets BP neural network algorithm can predict the flow of goods in the 14th samples of merchandise is O=785.4156.

3.2 *Inventory achievement based on the conversion rate*

3.2.1 *Rough inventory forecasting*

It can obtain the differential equations required by BP algorithm for network training from the (16):

$$\begin{cases} net_{jk} = \sum_{i=1}^{n} w_{ji} o_{ik} \\ o_{ik} = f(net_{jk}) \end{cases} \tag{17}$$

$$f(net) = \frac{2}{1+\exp(-\lambda net)},\ f(net) \in (0,1) \tag{18}$$

In this formula, O_{jk} is the actual value related to the Sample output neurons of the j-th commodity.

The selected prediction value in the prediction process of the goods traffic is as shown in Table 4. Based on MATLAB, the simulation results O = 785.4156.

Table 4. Predictive value parameter values

Store level	Collection popularity of the items	Cumulative sales	Existing stocks	Dynamic score	Goods visited
1	106	41	539	4.78	O=785

In the ideal model, the conversion rate of each path can be obtained from the following formula.

$$TR = \prod_{i=1}^{n} R_i = 9.3\% \tag{19}$$

According to the neural network model, the flow of goods O =785; Amount of inventory forecast= the predicted conversion× prediction of the commodity views.

$$SKU = TR \times O = 9.3\% \times 785 = 73 \tag{20}$$

Compared with the actual sales the value of goods 41, prediction result is close to 73, to a certain extent, error is inevitable, and O = 785 is bigger than the actual value 440 also partial. The places which may produce the error in analysis of the prediction process are:

(1) Sample data quantity is too small, not enough to reflect the authenticity of the data.

(2) Taobao, the authenticity of the data itself has been lost.

(3) Without full consideration of other influencing factors, the factors can impact the prediction data are rare.

In summary, the construction of inventory control model based on the conversion of the B2C mode by (17), (18), (19), and (20) are as follows:

$$SKU = TR \times O = \begin{cases} TR = \prod_{i=1}^{n} R_i \\ O_{ik} = f(net_{jk}) \\ net_{jk} = \sum_{i=1}^{n} w_{ji} O_{ik} \\ f(net) = \frac{2}{1+\exp(-\lambda net)} \end{cases} \tag{21}$$

3.2.2 *Correcting the inventory prediction*

The inventory prediction of the goods based on the number of weekly inventory is a rough prediction method, because it does not fully consider the factors such as seasonal fluctuations, holiday law, promotional activities and so on [24~29]. The concept of the weights of weekly sales index need to be introduced for more accurate prediction of the goods sales.

Due to the flexibility of e-commerce, we need to take several steps including: drawing up real-time replenishment strategy and establishing correspondence between related products; making statistical analysis of the sales data and establishing a basic prediction model with the use of regression methods if it is simple, and triple exponential smoothing methods if it is more complex; finding out the factors that can influence the forecast based on the weekly sales monitoring, and correcting the forecast timely; involving in the operations and sales decisions and getting to understand the impact of each activity on the sales; analyzing the major factors such as holidays and other important factors[30] and so on.

Because of the complexity of other factors and the limited time and experience, the authors do not make in-depth research. But our thought is that: added with the correction factor θ, the revised inventory is $SKU\times\theta$, and then we can obtain the inventory cycle of the goods $KOI= SKU\times\theta/T$. Finally, businesses can make purchases of the goods in accordance with the inventory cycle. The purchase quantity is equal to the current revised inventory forecast $SKU\times\theta$.

4 CONCLUSION

In the background of big data, this article focus on the data mining of commodity conversion rate, the prediction research of goods traffic and the analysis of BP neural network model. According to the existing domestic and international research, and combined with the characteristics of e-commerce users' prediction, this article established a prediction model based on artificial neural network, moreover, applied the model in the scale prediction of e- commerce users. The results show that the application of the neural network on the scale prediction of the e-commerce users is better than that of the mainstream method. Thus we can conclude that the result of the prediction is good, and the method is simple and practical.

ACKNOWLEDGEMENT

This article is supported by the Fundamental Research Funds for the Central Universities and Hebei Province Science and Technology Agency Fund (NO. 9100613021).

REFERENCES

[1] Yu Gang, E-commerce data Times, http://www.100ec.cn/detail-6085.html.

[2] Chen Yunhai & Huang Lanqiuz. 2013. Study on impact of big data processing electronic commerce. *Telecom Science*, (3): 17-21.

[3] Wang Juan. 2015. Big data research and analysis of electronic commerce application value. *Science and Technology Economic Market*, (2): 154-155.

[4] Wei Qunfeng. 2015. E-commerce platform of the era of large data model research. *Economic Research Guide*, (9): 228-229.

[5] Ni Ning. 2014. Study on electronic commerce platform the era of big data-taking www. Taobao.com as an example. *Jiangsu Business Theory*, (5): 13-14.

[6] Zhong Maiying & Tang Bingyong, 2002. Inventory control problem research under electronic commerce. *Journal of Systems Engineering*, 17(3): 257-260.

[7] Kaplan, S. E. & Nieschwietz, R. J. 2003. A web assurance services model of trust for B2C e-commerce. *International Journal of Accounting Information Systems*, (4): 95-114.

[8] Yuan Juanjuan. 2010. Improve the enterprise e- commerce website conversion rate analysis of the factors. *China's Trade*, (10): 107-108.

[9] LI Liang. 2012. The key indicators of e-commerce operations. *Modern Appliances*, (2): 41-42.

[10] Ling Shouxing. 2012. The influence of customer perceived value on site customer conversion rate empirical analysis. *E-business Journal*, (9): 47-49.

[11] Hou Haigui. 2014. Statistical analysis about the content classification in using SPSS analysis method to explore. *Economist*, (5): 72-75.

[12] Li Xiaosheng, Chen Zhenzhen, 2010. Correctly using SPSS software for principal components analysis. *Statistical Research*, (8):26-29.

[13] Chen Minqiong. 2015. The discriminate analysis using SPSS several problems. *Modern Compute*, (2): 34-40.

[14] Li Xiaofeng. 2000. Improved artificial neural network BP algorithm and its application. Journal of Sichuan University (*Engineering Sciences*), 32(2): 105-109.

[15] JIiang Ling & Ye Mingliang. 2011. Study of data mining method based on BP neural network algorithm. *Computer Programming Skills and Maintenance*, (18): 44-45.

[16] Huang Qingbin, The improvement of BP algorithm and its application research. Chengdu: Southwest Jiaotong University.

[17] Kong Fanye & Geng Ye. 2013. The supermarket goods inventory control model based on BP neural network. Journal of Liaoning Technical University (*Natural Science*), (6): 817-821.

[18] Pang Ningtan, Michael Steinbach & Vipin Kumar. 2011. *Introduction to Data Mining*. Beijing: China Machine Press.

[19] Gao Yuefang, Liang Yongsheng, Tang Fei & Ou Zhiwei. 2011. Forecasting model based on neural network. *Computer Engineering and Applications*, 47(22): 240-243.

[20] Zhang Xiaohui, 1994. *An Introduction to Neural Network and the Neural Computer*, Xi'an: Northwestern Polytechnic University Press, pp: 45-62.

[21] Su Gaoli & Deng Fangping. 2011. Improved algorithm of BP neural network based on MATLAB language. *Report of Science and Technology*, 19(2): 130-135.

[22] Luo Chenghan. 2004. Realization of BP network based on neural network toolkit in MATLAB. *Computer Simulation*, (5): 109-111.

[23] Liu Jingjing, 2012. Study on safety stock forecasting for chain retail enterprises based on BP neural network. *Logistics Technology*, (11): 325-329.

[24] Jansen B J & Resnick M. 2012. An examination of searcher's perceptions of no sponsored and sponsored links during ecommerce Web searching. *Journal of the American Society for Information Science and Technology*, 57(14): 1949-1961.

[25] Tu Jianbo, 2015. The influence factors of e-commerce platform user purchase intention. *Management Observer*, (2):169-170.

[26] Fang Mingzhu, Lu Runde & Wu Weiping, 2010. A review on the influence factors of consumer trust based on e-commerce. *Journal of Intelligence*, (6): 195-198.

[27] Jiang Hong. 2012. From Taobao to Taobao Mall, *Chinese Advertising*, (2): 53.

[28] Chen Bing. 2011. Taobao Marketing Strategy, Xiamen: Xiamen University.

[29] Liu Tiemin & Sun Ju. 2014. B2C e-commerce factors affecting consumer behavior research, *China Market*, (10): 67-68.

[30] Tan Xiongpai, 2011. Architecture: big data challenge, present situation and prospect, *Journal of Computer*, 34(10): 171-172.