A prediction-based inventory optimization using data mining models

Xiaoxiao Guo
School of Information
Renmin University of China
Beijing, P.R. China
for1205@126.com

Chang Liu
School of Information
Renmin University of China
Beijing, P.R. China
zenobia@ruc.edu.cn

Wei Xu
School of Information
Renmin University of China
Beijing, P.R. China
weixu@ruc.edu.cn

Hui Yuan
School of Information
Renmin University of China
Beijing, P.R. China
yuanhui91414@163.com

Mingming Wang
School of Information
Renmin University of China
Beijing, P.R. China
mingmwang@263.net

***Abstract*—As the core of the supply chain management, the inventory management deserves more of our attention, and in the complicated supply chain, especially under the circumstance of spending a long cycle, the inventory management becomes very difficult , which we need to balance the amount of circulating funds used by overmuch inventory and the loss of stock-out. The demand of marketing is viewed as the foundation of the inventory management, so in this paper, we are to adopt this idea and combine it with the information of searching on the web to conduct demand prediction for inventory optimization, and we will use Back-propagation neural network to train the prediction model. Then on the basis of prediction result, we will establish one simple and concise inventory policy. As a comparison result, a traditional inventory policy will be figured out by estimating a normal distribution of demand using the history sales data, and calculate the inventory cost with (s,S) inventory strategy. The result shows that the established inventory policy based on demand prediction has obvious superiority on reducing the total cost of inventory.**

Keywords-Inventory optimization; Back-propagation neutral network; inventory model; (s, S) strategy

I. INTRODUCTION

The inventory is a double-edged sword, on the one hand, although too much inventory can meet the satisfaction of customers, it can also increase the area and storage cost of the warehouse, which can heaven the product cost, thus, the enterprise asset may turn over slowly and insufficiently; on the other hand, too little inventory can also cause some questions. For example, it can easily lead to the stock-out, lost the trust of customers and reduce the benefits of enterprise. So the inventory management is becoming increasingly vital for companies nowadays. However, our market is full of all kinds of uncertainty, so it is considerable important and difficult to keep the inventory management good absolutely. Thus, it attracts more and more experts to pay attention to the research of the inventory optimization domain.

In recent decades, many scholars at home and abroad have been carrying on the extensive research for a variety of inventory system, and the case that contains one warehouse (or central warehouse) and multiple retailers receives special attention [1]. Axsater and his team assumed that the warehouse applies the conventional point stock batch ordering strategy, the homogeneous retailers use the joint complementary strategy and the cost is higher. Besides, Chen studied the layered, distributed, dynamic inventory management (HDDI) scheduling problem using the simulation method [2] and introduced the concept of emergency supplement. Cetinkaya and his team put forward a integrated inventory restocking supplement and delivery a plan model for VMI system, and the commerce applied the (s, S) policy to supplement their inventory, and send the order with the transport joint strategy based on time [3].

Besides, with the advent of web 2.0 and the emerge of the recommender system, the information around us is becoming more and more diverse and beneficial for us to get some potential law of things, so the prediction model is increasingly popular and vital, and many studies have operated it on the determine of inventory indicators. Among them, one of the main branches of the application is neural network. They perform the data simulation for the safety stock inventory with multiple indicators influencing inventory closely. Besides, the Artificial Neural Network is a kind of artificial intelligence technology which developed rapidly in the late 1980s. Many previous studies based on univariate time series analysis attempted to find the inherent policy of the information and data like the application of discrete Grey Correlation Model, ARIMA model, modified Radial Basis Function neural network (RBFNN) and so on [4-5].

From the statements about we can learn that: on the one hand, current studies mainly focus on the aspect of the operational mathematics methods and models for the inventory optimization; on the other hand, the idea of the prediction has been gaining great popularity and has been applied to many research domains. So in our paper we propose a novel idea that establishes an inventory optimization policy using demand prediction values based on the search information on the web.

Then the rest of the paper is constituted as follows: In Section2 we will show you basic theories of prediction methods and inventory management strategies. Then an

978-1-4799-5372-1/14 $31.00 © 2014 IEEE
DOI 10.1109/CSO.2014.118

inventory optimization method combining with the demand prediction result based on historical sales performance and search information is proposed in Section3. And in Section4 we will demonstrate the empirical analysis. Finally in Section 5, conclusions and the future work in this research are presented.

II. INTRODUCTION TO BASIC THEORIES

A. Neural network and Support network regression

Several methods are employed to forecast, generally including two categories: the econometrics models and the machine learning methods [6-8].

In recent years, machine learning methods are widely applied in different areas, such as economics, environment and finance [9-11]. In this paper, two kinds of data mining methods are evoked to excavate the relation between the features we select and the historical sales data in order for a more accurate prediction, which are Back-propagation neutral network and support vector regression. Then we will select the better one.

B. Inventory management strategies

From the perspective of academic research, the problem of inventory management is a basic question of the storage theory. The problem needed to be solved by the storage theory can be described as follow: How often does it replenish the inventory? What is the quantity for each replenishment? These are what we need to deal with.

In practical problems, we are more often faced with random demand, so it is necessary to take appropriate measures in accordance with the current stock situation. For random demand, there have been three main types: regular order policy, fixed-point order policy and the (s, S) inventory management policy.

The regular order policy means setting an order cycle. Order is based on the current inventory level at the beginning of each cycle. When inventory is high, order less and vice versa.

The fixed-point order policy means order only when the inventory reduces to the low level s, without considering the time interval. Each order will replenish inventory to high level S, with unchanged quantity.

The (s, S) inventory strategy is regular order integrated with fixed-point order. It means check the inventory with a certain time interval. If the current inventory is higher than low level s, do not order. While lower than s, replenish inventory to high level S.

Note that, when we define the inventory strategy, we need to abstract the practical problem into a mathematical model. firstly, and then during the establishment of the model, we had better to simplify some complex conditions provided that the model can reflect the nature of the practical problem.

III. THE PROPOSED METHOD

A. Structure of the proposed method

The inventory method combined with demand prediction results based on the information from web using prediction models above is described in Figure 1. The process of the proposed method could be summarized as following procedures:

Step 1: historical sales data and the search information data are get from a well-known e-commerce company, for these data, making pre-processing is necessary for the input of the predicting models.

Step 2: set a Normal distribution for the demand function based on the historical sales performance data as the contradistinction method.

Step 3: establish BPNN and SVR model with three input variables: x1(historical sales data), x2(the frequency of searching the commodity), x3(the click volume of the commodity page), train and evaluate these two models to select a better prediction model.

Step 4: achieve inventory control management based on the prediction values and the Normal distribution respectively.

Step 5: compare the inventory cost of two inventory optimization decisions above and then draw the conclusion.

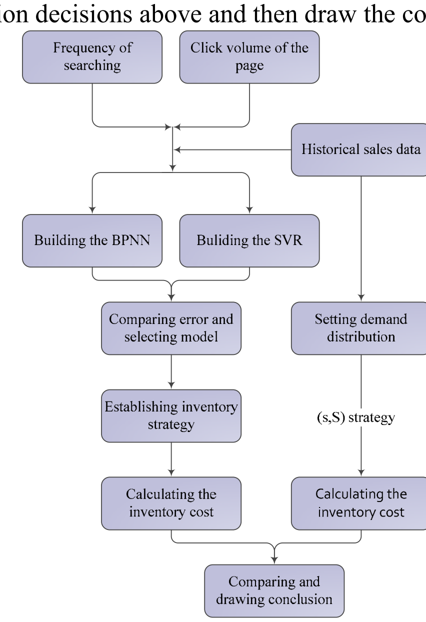

Figure 1.the structure of proposed method

B. Demand prediction using BPNN and SVR

Before the inventory optimization, the demand should be figured out so that the inventory management can be controlled better. Here the demand is closely connected with the sales prediction.

Referring to the features, applied to forecast sales, the historical sales data, which also can be called the historical demand, is always considered in the predicting model, which is to excavate the inherent rule inside the sales data itself. Along with the Web 2.0 boosting, researchers consider more about the web information [12-13].

Based on the decision-making process of consumers, the strategic decision of latent consumers might be influenced after some behavior of searching the detail information. So the features, added in the following model, consist of search information and historical demand data. And here the search information includes the frequency of searching the commodity and the click volume of the commodity page.

The previous part has illustrated the features and the models we employ. Then the modeling process can be presented as follows.

Firstly, aforementioned, three kinds of features are prepared to be input while the real demand data is regarded as the output.

Secondly, two kinds of data mining methods are evoked: back propagation neutral network, support network regression, and SVR are employed with one kernel functions, which is sigmoid kernel function.

Thirdly, the best predicting model can be selected to forecast the demand, and then proposed inventory optimization process can be preceded based on the demand prediction values from the previous steps.

C. The construction of the inventory optimization

i. inventory optimization using demand perediction values

With the prediction of products demands, instead of those previous vague and hypothetical demand distribution, we can get a more approximate value and its error with the actual demands performance based on some demand-related factors using predicting models, then we are to use this value to determine the optimal inventory as following steps:

Assume the prediction value is P_U, the mean absolute percentage error(MAPE) is ξ, and we can get its value from the formula (1).

$$\xi = \frac{1}{\mathrm{n}}\left(\sum \frac{A_t - F_t}{A_t}\right) \tag{1}$$

Where A_t represents the actual value; F_t means the predicting value; n is the number of the predicting data.

Now giving the prediction value P_U and mean absolute percentage error σ, according to the definition of σ ,we can calculate the range of the demand: $\left[\frac{P_U}{1+\xi}, \frac{P_U}{1-\xi}\right]$, that is to say the actual demand is between these two values, then we set Q as the variable of optimal inventory, and it will meet the formula (2).

$$\left(Q - \frac{P_U}{1+\xi}\right) \times b = \left(\frac{P_U}{1-\xi} - Q\right) \times c \tag{2}$$

Where b represents the storage fee for a unit of goods for one period; c means the loss fee by lacking of one unit of goods.

ii.inventory optimization based on Normal distribution.

In practice, the market usually has a stochastic and successive demand for the products. Although the fluctuations in demand may often make people feel that the randomness is so much that it is difficult to find the varying regulation. In fact, if the product is in a mature stage of the life cycle, we could observe its demands during a very long-term period. Then you can find that the stochastic fluctuation of the actual demand is in line with a statistical distribution. And a great many of experimental studies have shown that many features reflected from the actual demand are submitted to the Normal distribution. Therefore, as a comparison, for the previous research, we assume the demand trend is accordance with the Normal distribution and set the parameter of the Normal distribution as follows:

μ: the average value of the historical demands in a given period of time.

$$\mu = \frac{1}{\mathrm{n}}\sum_{i=1}^{n} x_i \tag{3}$$

σ :the standard deviation of the past demands in a given period of time.

$$\sigma^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2 \tag{4}$$

Usually, the inventory can be classified to be two types of the continuous or the periodical, and they also can be used in a joint system and in the continuous inspection strategy, the levels of inventory are examined constantly. And when a particular condition is satisfied, the corresponding products are ordered, also, there is a famous and often-used continuous inspection strategy: the (s, S) inventory strategy [14].

After determining the characteristics of the demand distribution, the next problem to be solved is to clear the inventory service level in the (s, S) inventory policy, that is to clear the value of s and S, and the concept and solving method for those two variables has been shown concisely in the past section.

Here is the general construction of (s, S) inventory strategy and symbols are defined as follows:

u shows the demand of a period, and values in $[0, +\infty]$. Beyond that, as a normally distributed random variable, u satisfies formula (5):

$$u \sim N(\mu, \sigma^2) = \int_0^{+\infty} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(u-\mu)^2}{2\sigma^2}\right) \mathrm{d}u = 1 \tag{5}$$

b shows the storage fee for a unit of goods for one period.

c shows the loss by lacking of one unit of goods.

s shows the low level of inventory, S shows the high level of inventory. At the beginning of each period, check the inventory. If the current inventory quantity is less than s, then replenish it to S immediately.

e shows the unit price of goods.

a shows the fixed cost during every purchase.

m shows the quantity of goods in the inventory at the beginning.

$\phi(u)$ shows the density function of demand.

Note that, in reality, the order lead time is usually a random variable (that is the length between the time when the companies sent order to their suppliers and the time when the ordered suppliers delivered goods to the business location), however, in this situation, in order to simplify the difficulty of the question, we assume in advance that it is zero.

Then we will determine the inventory service level separately with the Normal distribution based on prediction and the Normal distribution in traditional situation. And then we are to calculate a cost index to make a comparison between those two levels, and the detail will be expanded in the empirical analysis of the next section.

IV. EMPIRICAL ANALYSIS

A. Data description and evaluation criteria

The data of the historical demands and the search information on web we use are obtained from the JingDong e-business, and it is from Feb.2013 to May.2013, and here, to increase the sample data and its prediction accuracy, we set a week as a cycle. Then we could get products' sales performance for 16 weeks. The former ten weeks' data is used to train the prediction model while the rest data is applied to test the predicting accuracy. And then we conduct the inventory optimization using demand prediction values and the inventory optimization based on Normal optimization introduced in Section 3.

B. Experimental results

1) BPNN model and SVR model

In the demand predicting process, the input is historical sales performance, the frequency of searching the commodity and the click volume of the commodity page. Then we use all data of these three features from the first week to the tenth week to train our prediction model: BPNN and SVR, and we set a cycle (a week) as the lag period; later we use the last five six weeks' data to test these models, here, to make the problem simple and intuitive, we select 10 products and calculated their mean demand prediction values, which are used to be compared with the actual demands for the certain period. Following are the fitting performance of BPNN and SVR. It can be obviously observed that the BPNN method get more accurate prediction values.

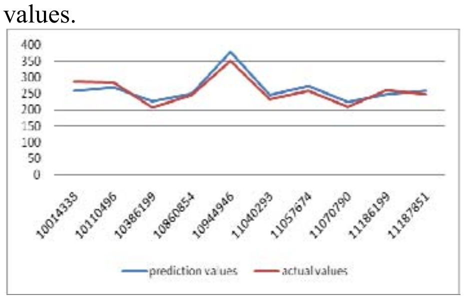

Figure 1. The performance of BPNN

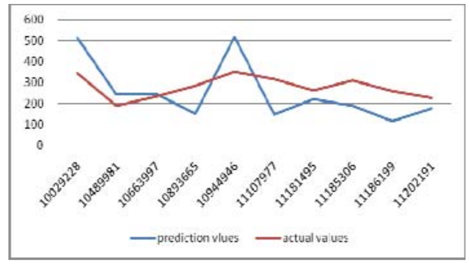

Figure 2. The performance of SVR

Then with those prediction values and according to the formula(1), we can get the parameter ξ is 0.06.

2) Inventory strategy with demand prediction

With the introduction of our proposed method before and the information of demand prediction values last part, we are to conduct the inventory optimization. In order to make the experiment more convincing, we decide to observe four continuous weeks' data from the 11th week to the 15th week and calculate its corresponding indictors.

Then, we randomly select one product (SKU: 11167073) to calculate its optical inventory during the four periods according to the formula (2), besides, we set the storage fee b for a unit of goods for one period as 1(Yuan), the loss fee c by lacking of one unit of goods as 29(Yuan), the fixed cost a during every purchase as 50(Yuan), the unit price of products e as 10(Yuan). And the calculation results of their optical inventory and inventory cost in every week are shown as the Table 1.

Table 1.the results of optical inventory and cost

No.(week)	12th	13th	14th	15th
Prediction values	282	244	282	282
Optical inventory	311	269	311	311
Inventory cost	3221	2815	2378	2605

3) (s,S)strategy with Normal distribution in traditional situation

To begin with this part, we use the formula (3) and (4) to get parameters' value of our Normal distribution.

Then to solve this demand distribution, we adopt following idea:

Firstly, the S should satisfy, the cost expectation can reach the minimal value when inventory high level is S.

Then among the entire process, the purchase payments:

$$a + e(S - m) \tag{6}$$

The expectation of storage fee:

$$\int_0^S b(S-u)\phi(u)\,\mathrm{d}u \tag{7}$$

The expectation loss of shortage is:

$$\int_S^{+\infty} c(u\text{-}S)\phi(u)\,\mathrm{d}u \tag{8}$$

So thirdly the total cost C(S) is:

$$C(S) = a + e(S-m) + \int_0^S b(S-u)\phi(u)\,\mathrm{d}u + \int_S^{+\infty} c(u-S)\phi(u)\,\mathrm{d}u \tag{9}$$

And S satisfies:

$$\frac{\mathrm{d}C(S)}{\mathrm{d}S} = e + b\int_0^S \phi(u)\,\mathrm{d}u - c\int_S^{+\infty} \phi(u)\,\mathrm{d}u = 0 \tag{10}$$

that is:

$$\int_0^S \phi(u)\mathrm{d}u = \frac{c-e}{b+c} \tag{11}$$

Finally, S should meet the formula above, and the order quantity is $S - m$;

The same to s, it reaches the minimum expectation cost at s. When the inventory is s and do not replenish the inventory, the likely storage fees and loss of shortage:

$$C(s)_1 = b\int_0^s (s-u)\phi(u)\mathrm{d}u + c\int_s^{+\infty} (u-s)\phi(u)\mathrm{d}u \tag{12}$$

While decide to replenish the inventory, the costs:

$$C(s)_2 = a + e(S-s) + b\int_0^S (S-u)\phi(u)\mathrm{d}u + c\int_S^{+\infty} (u-S)\phi(u)\mathrm{d}u \tag{13}$$

It is obvious that the expectation cost of replenishing is less than that of not replenishing when the current inventory is small; however, the expectation cost of replenishing is more when the current inventory is big. If the current inventory is equal to s, the expectation cost of replenishing is same to not replenishing. And this is the s at critical condition.

Let formula

$$b\int_0^s (s-u)\phi(u)\mathrm{d}u + c\int_s^{+\infty} (u-s)\phi(u)\mathrm{d}u$$

equals to formula

$$a + e(S-s) + b\int_0^S (S-u)\phi(u)\mathrm{d}u + c\int_S^{+\infty} (u-S)\phi(u)\mathrm{d}u$$

then get the s. If there are several s which can meet the above function, choose the minimal one.

Finally, the results are demonstrated in Table 2.

Table 2.the results of Normal distribution

No.(week)	12th	13th	14th	15^{th}
μ	54.64	70.92	79.62	91.29
σ	3478.05	6342.45	6797.59	8181.45
S	74.17	97.43	107.36	121.02
s	53.04	73.14	82.55	95.04
optical inventory	74	97	107	121
inventory cost	5894	3543	5064	3145

V. CONCLUSION

In our paper, we firstly combine the prediction theory with the inventory optimization that simply focused on the optional research model and mathematical model previously. In the process of demand prediction, we propose to adopt the search information on the web to conduct the prediction. And based on the predicting results, we come up with the novel inventory strategy that combining with the prediction model, instead of the Normal distribution in traditional way. And the results in the forth section shows that the proposed method can optimize the inventory management and reduce the inventory cost better.

However, the study also has a lot of problems for future research. First, for the prediction part, perhaps there are other factors that can influence the prediction accuracy; secondly, due to the limit capacity, the data we collect is not so adequate, so some errors may exist and further deep study based on more data is necessary; finally, the model solution method can be improved to achieve the inventory management optimization.

ACKNOWLEDGMENTS

This research work was partly supported by 973 Project (Grant No. 2012CB316205), National Natural Science Foundation of China (Grant No. 71001103, 91224008, 91324015), Humanities and Social Sciences Foundation of the Ministry of Education (No. 14YJA630075), Beijing Social Science Fund (No. 13JGB035), Beijing Natural Science Foundation (No. 9122013), Beijing Nova Program (No.Z131101000413058), and Program for Excellent Talents in Beijing.

REFERENCES

[1] Baganha M P, Cohen M A. "The stabilizing effect of inventory in supply chains", Operations Research, Vol. 46, No. 3, pp. 572-583, 1998.

[2] Chen L R, Ghosh S, "Modeling and simulation of a hierarchical, distributed, dynamic inventory management scheme", Simulation, Vol. 68, No. 6, pp. 340-362, 1997.

[3] Çetinkaya S, Lee C Y, "Stock replenishment and shipment scheduling for vendor-managed inventory systems", Management Science, Vol. 46, No. 2, pp. 217-232, 2000.

[4] Zhen D, and Feng S, "A novel DGM (1, 1) model for consumer price index forecasting", Grey Systems and Intelligent Services (GSIS), pp. 303-307, 2009.

[5] Wang Y, and Li X D, "Prediction and Analysis of Chinese CPI Based on RBF neural network", Information Technology and Applications, Vol. 3, pp. 530-533, 2009.

[6] Granger C W J, "Investigating causal relations by econometric models and cross-spectral methods. Econometrics", Journal of the Econometric Society, Vol. 37, No. 3, pp. 424-438, 1969.

[7] Pindyck R S, Rubinfeld D L, "Econometric models and economic forecasts", New York: McGraw-Hill, 1981.

[8] Bishop C M.,"Pattern recognition and machine learning", New York: Springer,pp.740, 2006.

[9] Schapire R E, "The boosting approach to machine learning: An overview", Nonlinear estimation and classification, pp. 149-171,2003.

[10] Muggleton S, King R D, Stenberg M J E, "Protein secondary structure prediction using logic-based machine learning", Protein Engineering, Vol.5,No.7,pp.647-657,1992.

[11] Weiss S M, Indurkhya N, "Rule-based machine learning methods for functional prediction", Journal of Artificial Intelligence Research,Vol.3,pp.383-403, 1995.

[12] Spertus E, "ParaSite: Mining structural information on the Web", Computer Networks and ISDN Systems, Vol. 29, No. 8, 1997.

[13] Chang C H, Kayed M, Girgis M R, Shaalan K F, "A survey of web information extraction systems", IEEE Transactions on Knowledge and Data Engineering, Vol.18, No.10, pp.1411-1428, 2006.

[14] Horenbeek A V, Buré J, Cattrysse D, Pintelon L, Vansteenwegen P, "Joint maintenance and inventory optimization systems: A review", International Journal of Production Economics, 2013, 143(2): 499-508.