

The Research of ANN Forecasting Mode Based on Cloud Platforms*

Jiang Xuesong**, Wei Xiumei, Geng Yushui, and Wang Xingang

Shandong Polytechnic University, School of information,
250353, ShanDong Jinan, China
jxs@spu.edu.cn

Abstract. With the rapid development of Cloud computing, data mining and ANN technology, we try to combined cloud data mining and ANN technologies, Using the Capacity of these technologies such as data acquisition, data storage and analyzing large amounts of complicated non-linear data, To construct a forecasting model for massive data real-time accurately which is difficult to complete on the traditional mode. This new model utilizes an enterprise sales data to test. The result proves that the model is feasible and effective. According to this model's prediction result, we can provide scientific and real-time decision support for the enterprise management.

Keywords: cloud computing, Ann, RBF, forecasting, data mining.

1 Introduction

Along with the development of computer, internet, communication technologies, and the automation of operational processes in all, almost all business generates tens of hundreds of GB or even TB of historical date[1]. How to forecast based on the massive data? The traditional forecast systems can not meet the new requirements any more, which is mainly manifested in the following points: data forecasting involved the dealing with massive data, so the traditional forecast systems can not satisfied the requirements of Operating efficiency, Computing performance, accuracy and storage space; the data have much incompleteness, noise and inconsistency which leads forecasting to chaos[2]. In this case, we should use data mining and cloud platform to store and compute the massive data. Data mining is the process of computers extracting implicit unknown information or mode with potential applications [5].

Cloud computing is the development of distributed computing, Parallel Computing and grid Computing [3]. It is a resource pool with a large number of Virtual Resources such as Development Platform, hardware and I/O services. Depending on the load, these virtual resources can be dynamically reconfigured, so that it can improve resource utilization. Cloud storage is achieved by using distributed storage to ensure

* Fund project: Development of information industry of ShanDong province special fund project (2008R00041).

** Corresponding author.

high reliability, availability and economy[4]. Therefore cloud computing system can satisfy the requirement of massive data storage, operation and computing more cheaply. This system can provide data forecasting based on massive data services for users.

Now popular forecasting methods are as following: Statistical methods based on probabilistic model, Decision tree based on machine learning and ANN and so on.

ANN methods are robust, Self-organizational, adaptive, parallel processing. ANN can analyse a large number of complex data, and exact extremely complex modes. So ANN methods are very suitable to resolve the data forecasting.

2 Construction of ANN Forecasting Model

ANN forecasting model based on cloud platform is shown as figure 1:

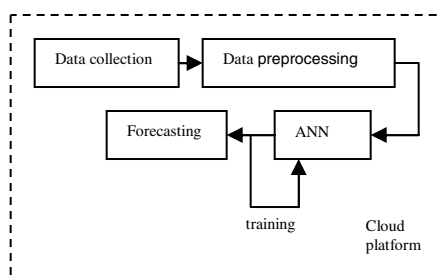


Fig. 1. ANN forecasting model based on cloud platform

Forecasting model based on RBF neural networks and BP neural networks are the most commonly used methods. Both RBF and BP neural networks are mature technology. So they are used at data forecasting successfully. Both RBF and BP neural networks are feed forward artificial neural network. There are some similarities and differences between them.[6][7].

3 The Result of Experiment Verify the Forecasting Effect

This paper take the various manufacturing enterprise product sales forecast as an example to verify the effect of using BP and RBF for data forecasting.

Sales forecast is one of the most important factors in the enterprise marketing management. And it also is the key link of enterprise supply chain.[1]

Based on the analysis and research affecting the market supply and demand of the many factors and the present and past sales data, sales forecast uses scientific method to estimate and speculate the future market products supply and demand development trend. According to the sales forecast results, the enterprise may set up a rational purchasing plan, production plan, stock plan and marketing plan.

3.1 Evaluation Standard of Data Forecasting

To compare the BP model and RBF model predictive power, we adopted mean absolute error percentage MAPE to evaluate accuracy of prediction.

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{f}_i}{y_i} \right| \times 100\% \quad (1)$$

In the expressions (1), y_i is the actual output of the i node, \hat{f}_i is the prediction of the i node.[2]

3.2 Data Collection

Collecting the data from web, we need to use web data mining technology. Web data mining uses data mining technology to find and retrieve information from web documents and services. Because there are lots of information available on Internet, we need distributed operation in clouds platform.

Considering the various factors affecting the product sales, we choose the sum of investments in fixed assets used, producer price index, output of major industrial products, purchasing price indices of raw material, fuel and power, the related industrial index, gross domestic product, price of production sales, quality of products, type of production etc. 11 main influencing factors as the input of the model.

So the number of input cells is 11. The output of the model is the forecasted sales after neural network model calculation. The number of output cells is 1.

The sample data of this model is from the National Bureau of Statistics of China website. And the products data is from a glass fiber manufacturing enterprise's production and sales statistics in 2006, 2007, and 2008. Data of 2006 and 2007 is the training data. After training the algorithm and adjusting parameters, we determine the final neural network model. Data of 2008 is the test data to provide the forecasting effect of the ANN model.

3.3 Data Pretreatment

Because we collect the mass of data, and some of them are useless or damaged information, we must preprocess the data before analysis. How to preprocessing the data?

We should classify, screen, and preserve them in a standard format. So that it is easy for us to input analyze module. After preprocessing the data, we can get a set input cells (x_1, x_2, \dots, x_n) to describe the users' behavior characteristics.

Because collecting data is not in the same order of magnitude, they must be mapped to [0,1] interval. After pretreatment the data are normalized to neural network's input.

Because the data for pretreatment of is numerous, and normalized calculation is complex, we should use clouds platform to calculate to improve computational speed and efficiency. [6][7]

Take X_{ij} for the training data set, which contains a large amount of data element changes, the goal is to make the change between elements less than 1. First calculate the average of all elements X :

$$p_0 = \frac{1}{m * n} \sum_{i=1}^m \sum_{j=1}^n x_{ij} \quad (2)$$

In the expressions (2), $i=1,2, \dots, m$; $j=1,2, \dots, n$.

And then calculate the standard deviation of all the elements of X:

$$p_1 = \sigma(x_{ij}) \quad (3)$$

In the expressions (3), σ --Standard deviation.

So the normalized X is:

$$x_{ij} = \frac{x_{ij} - p_0}{p_1} \quad (4)$$

The normalized data as the neural network model of the input data, the data of 2006, 2007 and 2008 are stored in three tables.

Table 1. Normalized data

No	x1	x2	x3	...	x11	Actual sales
1	0.16221	0.60714	0.4902	...	0.54098	0.15678
2	0.04097	0.57143	0.43137	...	0.38883	0.52347
3	0	0.5	0.2549	...	0.31325	0.23654
4	0.055184	0.35714	0.13725	...	0.06135	0.36983
...

The data in Table 1 is the result of normalization, so all data are in [0, 1] interval.

3.4 The Experimental Results of BP Network and RBF Network

In this paper we use matlab6.5 to train the BP network and RBF network. First, we train the BP network. After repeated testing, we select the number of hidden layer neurons is 25. BP network for each input and output sample value, each power affects the output of the network. During the training, each power needs to be adjusted through the back error propagation, thus learning is slowly.

In Table 1 take x1, x2, ... , x11 as input data of the BP network, to train the network, and then get the output data, ie the forecast to the sales. Compare the predicted values to actual sales in Table 1, adjust the parameters of the network repeatedly, then we will get the final model. The BP network took 7.2650 seconds, 120 times training to the goal error of 0.01, and then the training stops.[7]

But when we use BP network, during the training, not converge condition often occurs.

The test sample set data of 2008, input to the trained BP network model to predict the effect of the model. After the BP network computing, we get the output data. Compare the predicted data to the actual sales data in 2008; we can see the final prediction results of the BP network.

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{f}_i}{y_i} \times 100\% \right| = 2.22\% . \tag{5}$$

After training the RBF neural network model and adjusting the parameters, we get the ultimate model of RBF neural network. The RBF network took 0.254 seconds to the goal, and stop the training. not converge condition did not occur.

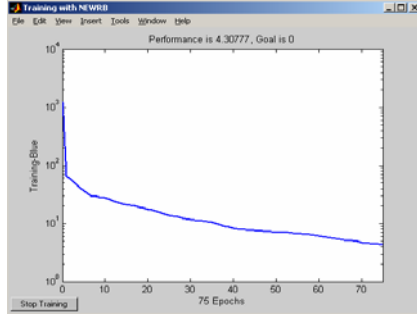


Fig. 2. The training of RBF network

From fig 2, we can see that the convergence speed is relatively fast.

Input the test sample set data of 2008 to the trained RBF network model to predict the effect of the model. After the RBF network computing, get the output data. Comparing the predicted data to the actual sales data in 2008, we can see the final prediction results of the RBF network.

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{f}_i}{y_i} \times 100\% \right| = 0.82\% . \tag{6}$$

4 Conclusion

Through the experiment, it can be proved that RBF neural network has better forecasting effect than BP neural network. But the results of RBF are affected by various parameters, so the prediction accuracy will reduce with the passage of time. So the RBF network does not use on long-term forecast. We can predict after period of time, and then put the new data as the training data on network model to training, in order to achieve higher precision of prediction.

References

1. Bo, C., Zhu, O.: Prediction of winter wheat evapotranspiration based on BP neural networks. Transactions of the CSAE 26(4), 81–86 (2010)
2. Chang, F., Dean, J., Chema Wat, S.: BigTable: A distributed storage system for structured data. ACM Transactions on Computer System 26(2), 1–26 (2008)

3. Cai, X.-w., Duan, M.-b.: Research about Cloud Computing and Using at Data Mining. *Computer Knowledge and Technology* 6(22), 6272–6274 (2010)
4. Gui, B.-x., He, J.: Distributed Data Mining Approach with High Performance Cloud. *Computer Engineering* 36(5), 76–78 (2010)
5. Li, L.: Research of sale forecast based on data mining. Example of Application 8, 74–76 (2009)
6. Wei, W.: *Neural Networks compute*, pp. 41–46. Higher Education Press
7. Yan, P.-f., Zhang, C.-s.: *Artificial neural network and Simulated evolutionary computation*, pp. 46–48. Tsinghua University Press