Data Mining on Prices of Highway Project Main Material Based on WEB

Jie He   Yikai Chen   Wen Hang   Zhiguo Qi

Transportation College, Southeast University, Nanjing 210096, P. R. China

Abstract

The management of main material prices of provincial highway project quota has problems of lag and blindness. Framework of provincial highway project quota data MIS and main material price data warehouse were established based on WEB firstly. Then concrete processes of provincial highway project main material prices were brought forward based on BP neural network algorithmic. After that standard BP algorithmic, additional momentum modify BP network algorithmic, self-adaptive study speed improved BP network algorithmic were compared in predicting highway project main prices. The result indicated that it is feasible to predict highway main material prices using BP NN, and using self-adaptive study speed improved BP network algorithmic is the relatively best one.

Key Words: Highway project main material, Data mining, WEB, BP neural network

1. Introduction

Low managing efficiency, long updating cycle and difficult in decision support are general problems in highway project management in China. For highway project management in province, these problems are especially prominent, becoming one of the most difficult problems in project fund management. These problems also have indispensable negative effects on projects' economic and social benefits [1]-[2]. Since that, this thesis analyzed the framework of provincial highway project integrate MIS and the construction of database, data warehouse at first, based on which studied the application and implement of data mining in highway project main material prediction. Finally overcome the lag and blindness of highway project main material management, enhanced highway project finance management efficiency.

2. Data Collection and Management Based on WEB

2.1. Framework of data management system based on WEB

The whole structure of provincial highway project quota integral MIS' transfer system could be constructed by a combination of B/S and C/S [3]-[5]. The structure shows in Fig.1.

Firstly, design companies, building companies, supervision companies, construction companies and public who have certain qualifications were connected to INTERNET directly. People could access system through WEB browser, they could visit distributed database and provincial quota station's data warehouse to complete general operations such as quota and project information query, data mining result query, supplementary quota report, already-built MIS utilization, etc. This part used B/S structure. Secondly, inside city's quota station, city's quota database servers were connected to city quota station's administrator computers as a part of distributed database. This part could complete any operations besides system maintenance. It used C/S structure.

The structure in provincial quota station was more complicated: On one hand, provincial quota station database was connected to other databases in distributed database system; On the other hand, it was connected to system administrator computers. Besides these, it also could receive data from already-built MIS. Distributed database system and already-built MIS were connected to provincial data warehouse as information resources for data mining. The structure of provincial quota station used C/S too.

2.2. Data warehouse of highway project main material prices

The main tasks of provincial highway project quota data warehouse could be classified into two aspects: Prediction of manpower, material, machinery One-shift expense quota and prediction of highway maintenance project quota expense. According to this, we designed three subject fields for data warehouse: project information subject, material and equipment price quota subject and maintenance history subject. Among these, project information subject was composed by project information fact table and many dimensional tables, the primary key of fact table was project ID; The public key of material and equipment price quota subject was project ID too; The public key of maintenance history subject was maintenance ID. Because of its simpleness, we put manpower expense quota into the subject information subject, not list it separately.
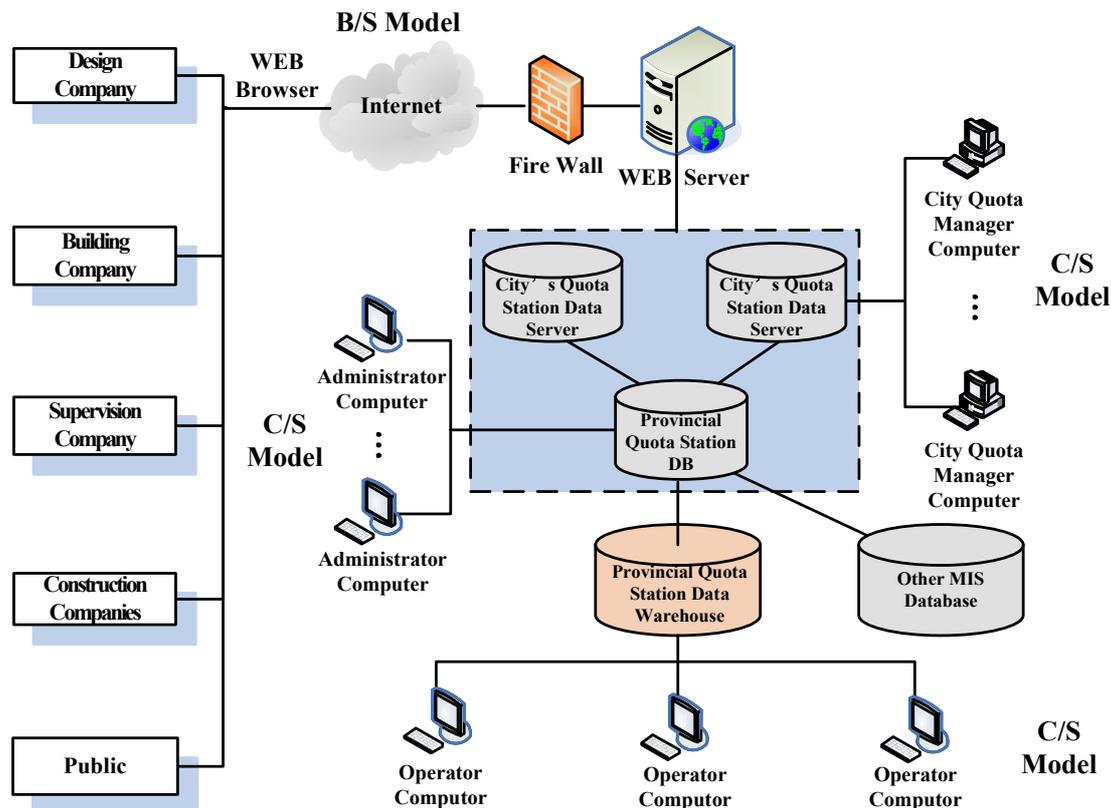
Fig. 1: Information transfer system of provincial highway project quota integral MIS.

Provincial highway project quota data warehouse' concept model is shown in Fig.2. Material and equipment price quota subject was composed of five properties. They were estimate quota, budget estimate quota, budget quota, construction quota and highway project quota.

The aim to establish material and equipment price subject was to predict and analyze all kinds of material prices and equipment prices. So its data quantity was the largest and its structure was the most complicated. The relation model is shown in Fig.3.

3. Predictions on Highway Project Main Material Prices Based on BP NN

3.1. Dada mining procedure of highway project main material prices
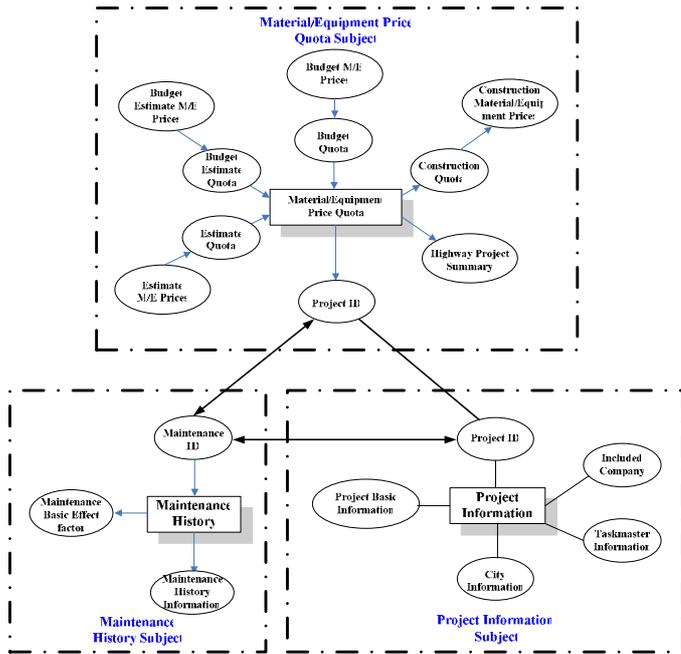
Fig. 2: Concept model of Provincial highway project quota data warehouse.
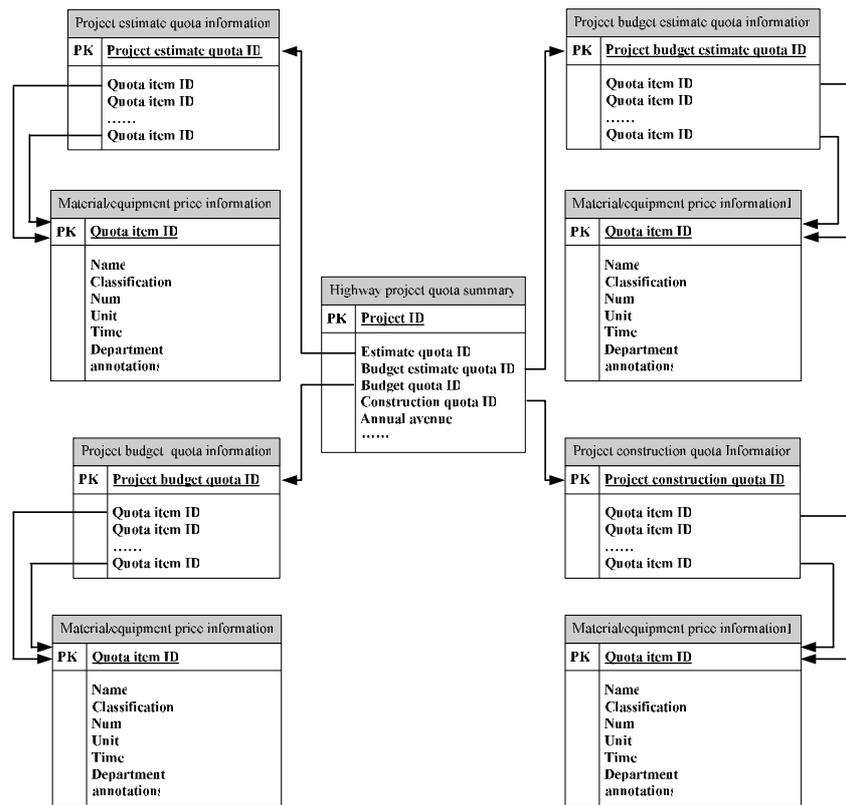
Fig. 3: Relation model of material/equipment price ration subject.

The change of highway project main material prices are influenced by many factors. These factors would also influence each other to present a complex nonlinear relation which makes predictions on highway project main material prices very complex problems. As a powerful tool, BP NN was adopted to dealing with such complex nonlinear problem [6]-[8]. According to the basic procedure of DM, this paper suggested DM procedure of the provincial

highway project main material prices based on BP NN.
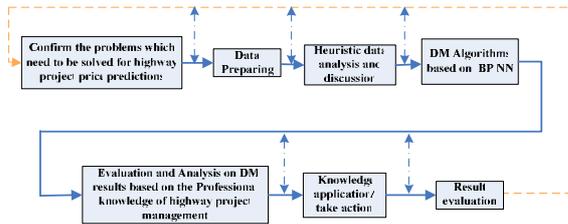
Fig. 4: Basic DM procedure of highway project main material prices.

3.2. Analysis and determination of influencing factors

Under the condition of Market Economy, Factors that influence the price consists of cost, supply-demand relationship, market competition status, industry policy and so on [8]-[9]. According to this, factors that influence highway project main material prices can be summarized as follows: ① National industry policy: Highway development policy, revenue policy, monetary policy, labor policy, etc. ② Regional Economic Development status: GDP, industry development level, highway investment level, average income level, etc. ③ Regional transportation condition: Highway transportation condition, railroad transportation condition, waterway transport- ation condition, pipeline transportation condition, etc. ④ Market changing status: home market status, international market status, etc.

After dealing with the influencing factors by using relative analysis method, this paper choose several influencing factors which were much more relative with highway project main material prices and more easily to be quantized. Firstly, considering the Demand, choose GDP, infrastructure payout and highway route mileage. Secondly, according to the supply, take 7 factors, such as average income level of on guard employee, freight volume, water price, oil price etc, and some historical data as input vector.

3.3. The network structure determination

Practice suggested that the increase of hidden layer units could enhance the Nonlinear Mapping Ability of BP NN, whereas if the number of the units exceeded a certain value, the Network performed worse. Therefore, this paper

adopted a BP NN with one hidden layer, which mean that the BP NN consisted of an input layer, an output layer and a hidden layer.

The neurons number of input layer and output layer was confirmed by the dimension of the input vector and output vector. The dimension of the input vector was the number of influencing factors added to the number of historical data. For example, we forecasted the main material price of the current quarter in some city according to the price of the first 3 quarters and the influencing factors. That was to say, the neurons number of input layer was 10, while there was only one output vector. The neurons number of hidden layer directly influenced the nonlinear forecasting performance of the network. According to the Kolmogorov theorem, i.e. the neurons number of input layer is n, the neurons number of hidden layer was 2n+1, the neurons number of hidden layer was 21.

3.4. Analysis and processing of sample data

It is generally considered that the lack of sample might make the network not perform sufficiently, while the overfull sample might present a redundancy situation which both adds the burthen of network training and lets the network over-fitting [7]-[9]. Anyway, the samples should be representative, balanced and be consonant with the main material price in Highway Engineering.

The sample data required by predicting highway project main material prices contained historical data of main material price, which could be acquired from the administration section, and the data of influencing factors which could be obtained from the local government or the statistical annual. Generally speaking, as the data collected might had different order of magnitude, after getting the input and output data, we should did normalization to make all the value of the data in the section [0, 1] in order to raise BP NN training speed.

3.5. Example analysis

This paper took the 3rd quarter stone chip price in Zhengzhou as an example. The input vector consisted of the three quarters' quarter stone chip price before the forecasted season, and 7 influencing factors of the forecasted season, such as GDP, infrastructure payout and highway route mileage, freight volume, average income level of

on guard employee, water price, gasoline price, etc. That was to say, the neurons number of input layer was 10. The object vector was the stone chip price of the forecasted season, i.e. the neurons number of output layer was only one. According to the Kolmogorov theorem, the neurons number of hidden layer was 21. The transfer function of the hidden layer was tansig, while the transfer function of the output layer was logsig. The data which was chosen as the training samples were the prices of stone chip, water and gasoline from the 1st season of 2004 to the 1st season of 2006 as well as GDP, infrastructure payout and highway route mileage, freight volume, average income level of on guard employee.

As the standard algorithm has a low identifying speed and easy to encounter local minimization, this paper adopted the general gradient descent algorithm with the training function traingd, momentum algorithm amendment algorithm for the BP network with the training function traingdm, adaptive algorithm of learning rate for BP NN with the training function traingda and the adaptive learning rate momentum algorithm amendment algorithm with the training function traingdx. According to the algorithms mentioned above, a model based on MATLAB NN toolbox was created to forecast the 3rd quarter stone chip price in Zhengzhou. In the model, the value of target error was setup to 0.01, the training steps were 7000, and the preliminary learning speed of the momentum algorithm amendment algorithm for the BP network was 0.05, while the values of other algorithms were 0.01. The forecasting result indicated that there were some differences in convergence rate and precision of different algorithms, as is shown in table 1.

training functions of BP algorithms	Training steps	Predicting results	actual results	error
traingd	3590	21.41	20.5	4.44%
traingdm	3244	21.23	20.5	3.56%
traingda	163	21.39	20.5	4.34%
traingdx	164	21.29	20.5	3.85%

Table. 1: Predicting results of the stone chip price based on NN.

4. Comparisons and Analysis of the Predicting Results from Different BP Algorithms

Conclusions drawn from the results were given as follows:

- With the lowest converging velocity, the standard BP algorithm had big error.
- As the training steps reduced 9.64%, the converging speed of momentum amendment algorithm for the BP network was faster than that of the standard BP algorithm. What was more, compared with the error of the later algorithm, the error of the former one was significantly reduced by 0.88%.
- Compared with the value of the BP network, the converging speed of adaptive algorithm of learning rate for BP NN was significantly reduced by 95.46%, which means the later algorithm was much faster. However, the error of the later algorithm only reduced 0.1%.
- If adopted the adaptive learning rate momentum algorithm and amendment algorithm at the same time, the converging speed was much faster than both standard BP algorithm and momentum amendment algorithm. Compared to the two algorithms, the training steps of the former algorithm reduced 95.43% and 94.94% separately.
- In a word, adopted the adaptive learning rate momentum algorithm and amendment algorithm at the same time was the most effective amendment algorithm for the BP NN.

The conclusions indicate that the errors of all different BP algorithms are small, which means predicting highway project main material prices is feasible. According to the results, the adaptive learning rate momentum amendment algorithm with the training function traingdx performs well in both converging speed and error, so that it is regarded as the best of the four algorithms.

5. Conclusions

This paper created a Web-based framework for the quota database management system of provincial highway project and a date warehouse model of the main material prices. Then brought forward the DM procedure of highway project main material prices based on BP NN. However, with the limitation of the paper length, some problems, such as the standardizing processing for the data collection, the relation of the DDB and the date warehouse, the adaptive learning rate momentum amendment algorithm with the training function traingdx, were not in details. If readers have further demands, please refer to the

report "Study on Highway Project Quota Online MIS" [10], which was written by the author.

References

[1] Y.L. Niu, Y.K. Chen, J. He, X.H. Li, Analysis and Design on Highway Project Ration Distributed Database of Henan Province. *Computer and Commuication,* 25(134):132-135,2007.

[2] Y.L. Niu, W. Hang, J. He, X.H. Li, H.J. Mao, Algorithm of Time Quota Estimation Using Unsupervised General Fuzzy Min-max Artificial Neural Network. *Journal of Highway and Transportation Research and Development*, 24(2):155-158,2007.

[3] K. Nassar, A. Al-Mohaisen. Simplified Approach to Probabilistic Cost Benefit Analysis Architectural Lighting Example. *Cost Engineering*, 48(1): 13-18, 2006.

[4] M. E. Georgy, L.M. Chang, Lei Zhang. Engineering Performance in the US Industrial Construction Sector. *Cost Engineering*, 47(1): 27-36, 2005.

[5] K.R. Molenaar. Framework for Comparing Project Delivery Costs. *Cost Engineering*, 46(11): 24-32, 2004.

[6] X.Q. Wang, J. Kang, A Fast-Quotation Method in International Construction Project Based on the Historic Data. *Journal of Hebei University of Technology*, 31(3): 49-53, 2002.

[7] K.W. Wu, J. Chen, C.Y. Ge, Fuzzy Neural Network's Calculating Method of Building Price. *Journal of Donghua University, Natural Science*, 27(3): 85-88, 2001.

[8] J.L. Peng, Q.S. Zhang, Study on the Model of Cost Estimate of Highway Engineering. *Journal of East China Jiaotong University*, 22(5): 41-51, 2005.

[9] H. Ren, Q.M. Zhou, Application of Neural Network for Quick Estimation of Engineering Construction Cost and Main Quantities. *China Civil Engineering Journal*, 38(8): 135-138, 2005.

[10] Traffic Transportation Task Group of Southeast University. Study on Highway Project Quota Online MIS. 2007.