Analysis Models of Technical and Economic Data of Mining Enterprises Based on Big Data Analysis

Jian Ming, Lingling Zhang, Jinhai Sun
State Key Laboratory of High-Efficient Mining and Safety of Metal Mines (University of Science and Technology Beijing)
Ministry of Education
Beijing, China
e-mail: mingjian.ustb@163.com

Yi Zhang
School of Management and Economics
Beijing Institute of Technology
Beijing, China
e-mail: zhangyi_bjtu@163.com

***Abstract*—Characteristics of the technical and economic data of mining enterprises are multi-dimensionality and nonlinearity. The sales price data of mineral products is an important economic indicator of mining enterprises, and the geological data is an important technical data. The analysis method of the technical and economic data is researched using technologies of big data analysis and data mining. The fluctuation pattern and influencing factors of the mineral products price are analyzed. The prediction model of the mineral products price is established using artificial neural network. The results show that the practicability of the prediction model is strong, and the prediction accuracy is high. During the process of mineral development, due to the limitation of technical conditions and equipment conditions, lots of geological data have been lost, which reduces the accuracy of the orebody shape and that of reserves estimation. Based on techniques of geostatistics and artificial neural network, the prediction model of the geological missing data is established. By using the model, the regularity of geological data of single borehole, the regularity of geological data of group boreholes and the regularity of geological data of all boreholes is discussed and analyzed. It has been proved that most of the geological missing data can be predicted and interpolated, and results of prediction and interpolation are reliable.**

Keywords-mining enterprises; technical and economic data; BP neural network; prediction models

I. INTRODUCTION

In consequence of the long production cycle, mining enterprises find it difficult to adapt the changes in supply and demand of the mineral products. The operations plan of the enterprise can't be made quickly to follow the trend of mining industry globalization [1], [2]. The making of the mine enterprise operations plan is important parts of the operations management of mining enterprise [3]. At present, large amounts of data resources generated and collected by information systems of the mining enterprise are not analyzed scientifically [4]. These data failed to provide adequate support to processes of production management and decision-making of the mining enterprise [5], [6]. Thus, the establishment of efficient analysis and prediction models of mine technical and economic data is of great significance to the mining enterprise. At present, large amounts of data resources generated and collected by information systems of the mining enterprise are not analyzed scientifically. These data failed to provide adequate support to processes of production management and decision-making of the mining enterprise. Thus, the establishment of efficient analysis and prediction models of mine technical and economic data is of great significance to the mining enterprise.

Many studies are carried out about the analysis and prediction methods of the technical and economic data [7], [8], [9]. Reference [10] analyzes the prediction and interpolation methods of missing economic data of the mining enterprise, such as the mean method, the weighted average method, the linear regression method, the maximum expected method and the multiple imputation method. Reference [11] collects the borehole data and discovers the global trend and the aeolotropism existing in the data. The data is transformed to normal distribution, and the aeolotropism of the data is rejected, thus the interpolation precision is improved.

In order to improve the ability of data analysis of mining enterprises, basing on the characteristics of the technical and economic data of mining enterprises, the prediction mode of mineral products price and the prediction and interpolation model of the geological missing data is established respectively by means of techniques of geostatistics and artificial neural network.

II. THE ANALYSIS AND PREDICTION METHOD OF TECHNICAL AND ECONOMIC DATA

A. Types of Prediction Techniques

The price prediction method of mineral products price and the prediction and interpolation method of the geological missing data are important parts of the mine technical and economic data analysis model. Then short term price prediction data can be calculated by using the price prediction model. The new geological database which has more accurate and abundant data is established. At the present time, there are two main kinds of quantitative prediction models, namely, the casual prediction models and the extrapolation prediction models [12], [13]. The input data

of the casual prediction models consist of the data of forecast objects themselves and their influencing factors [14].

B. The Artificial Neural Network Prediction Models

The artificial neural network model is a casual prediction model [15]. The influencing factors data of the forecast object can be fully analyzed and utilized by using these models. The influence of different factors on the change of the forecast object can be described. Its basic theories are that after the prediction accuracy is determined, the functional relationship between input and output is/will be established using training samples to train the network. In the paper, the artificial neural network model of the backpropagation algorithm is used to predict the mineral products price and predict and interpolate the geological missing data. The model of three-layer neural network is applied to the establishment of these/above prediction and interpolation models.

Let $u = 1,2,\ldots,s$ be the number of training samples. Let $v = 1,2,\ldots,w$ be the number of the input unit of the network and $t = 1,2,\ldots,q$ be the number of the output unit of the network. Let $j = 1,2,\ldots,p$ be the number of hidden layer neurons.

In the input layer, the set of input signals of the hidden neurons $\{s_j(u)\}$ is gotten by using the input vector O_u .and the set of weights which are from input layer neurons to hidden layer neurons $\{\omega_{vj}(u)\}$. Then, $\{s_j(u)\}$ is used to calculate the set of inputs of the hidden neurons $\{b_j(u)\}$ by using the transfer function between the input layer and the hidden layer. The equation for $\{b_j(u)\}$ is

$$b_j(u) = f(s_j(u)) = f\left(\sum_{v=1}^{w} \omega_{vj} \cdot x_v(u) - \phi_j\right) \tag{1}$$

where $\{\phi_j\}$ is the set of thresholds of the hidden layer, $\{\varphi_j\}$ is the set of thresholds of the output layer.

The set of input signals of the output neurons $\{l_t(u)\}$ is gotten by using the set of output signals of the hidden neurons and the set of weights which are from hidden layer neurons to output layer neurons $\{\upsilon_{jt}(u)\}$.Then, $\{l_t(u)\}$ is used to calculate the set of responses of the output neurons $\{c_t(u)\}$ by using the transfer function between the hidden layer and the output layer. The equation for $\{c_t(u)\}$ is

$$c_t(u) = f(l_t(u)) = f\left(\sum_{j=1}^{p} \upsilon_{jt} \cdot b_j(u) - \varphi_t\right) \tag{2}$$

The set of the error signals of the output neurons $\{d_t(u)\}$ and that of the hidden neurons $\{e_j(u)\}$ are gotten by using the desired output vector O_u, the set of weights $\{\upsilon_{jt}(u)\}$which is from hidden layer neurons to output layer neurons, the set of output signals of output neurons $\{c_t(u)\}$ and the set of output signals of hidden neurons $\{b_j(u)\}$. The equations for $\{d_t(u)\}$ and $\{e_j(u)\}$ are

$$d_t(u) = (y_t(u) - c_t(u)) \cdot c_t(u) \cdot (1 - c_t(u)) \tag{3}$$

$$e_j(u) = \left(\sum_{t=1}^{q} d_t(u) \cdot \upsilon_{jt}(u)\right) \cdot b_j(u) \cdot (1 - b_j(u)) \tag{4}$$

$\{d_t(u)\}$ and $\{b_j(u)\}$ are used to adjust $\{\upsilon_{jt}(u)\}$ .. $\{e_j(u)\}$ and $\{x_v(u)\}$ are used to adjust $\{\omega_{vj}(u)\}$. The equations for $\{\upsilon_{jt}(N+1)\}$ and $\{\omega_{ij}(N+1)\}$ are

$$\upsilon_{jt}(N+1) = \upsilon_{jt}(N) + \alpha \cdot d_t(u) \cdot b_j(u) \tag{5}$$

$$\omega_{vj}(N+1) = \omega_{vj}(N) + \beta \cdot e_j(u) \cdot x_v(u) \tag{6}$$

In (5) and (6), α is the learning rate of weights between hidden neurons and output neurons, β is the learning rate of weights between input neurons and hidden neurons. In sequence, each input vector is used to train the neural network. If all s input vectors have been used to train the network, an input vector will be not randomly selected from s input vectors until the global error function of network E is less than the minimum value set in advance.

III. THE PREDICTION MODEL OF MINERAL PRODUCTS PRICE

Due to wide fluctuation of the mineral products price and long production cycle of mineral products, the mining enterprise usually finds it difficult to adapt the changing market. The information chain and the value chain among operations departments of the mining enterprise are near blocked and strategic decisions of the enterprise can't keep pace with the changing trend of supply and demand of the mineral products.

A. The Establishment of the Model

The mineral products prediction model of neural network is established. Its network framework is a single hidden layer three-layered neural network with 5 input neurons and 1 output neuron. Based on the training samples and the empiric formula, the number of hidden layer neurons is 12. The transfer function between the input layer and the hidden layer is 'tansig', and 'logsig' is used as the transfer function between the hidden layer and the output layer. The network training function is the 'traingdx' function.

B. The Training and Prediction of the Model

Let o_u be the price of the mineral product of the period u . The correlation analysis of the price data of the mineral product is conducted. The analysis results can show that o_u for the period $u - x$ has strong correlations with the average monthly spot price of the mineral product imports, the

monthly amount of the mineral product imports, PPIs for mining and quarrying industry, BCI and USDX for the period $u-x$. Let i_1, i_2, i_3, i_4, i_5 be these influencing factors of the price of the mineral product respectively. The data, as the input of the back-propagation neural network model, are used to establish the input vector $I_u = (i_{u1}, i_{u2}, \ldots, i_{us})$. In this model, $u = 1,2,\ldots,s$ is the number of training samples and $v = 1,2,\ldots,w$ is the number of the input unit of the network. The data of the mineral product price during the observation period are used to establish the desired output vector $O_u = (o_{u1}, o_{u2}, \ldots, o_{uq})$. In this model, $t = 1,2,\ldots,q$ is the number of the output unit of the network.

The network is trained and tested with the historical data of influencing factors of the mineral products price. The prediction results show that the prediction model practicability of the prediction model is strong and the prediction accuracy of the prediction model is high.

IV. The Prediction and Interpolation Model of the Geological Missing Data

During the process of mineral development, a large number of boreholes are constructed for mineral exploration. By analyzing the core samples collected from boreholes, the ore grade of different depths of the core samples can be obtained. The data of core samples are logged to determine the ore grade of the formation surrounding the borehole. Based on these geological data, the specific grade of ore for an area where there are only a few known sample values is estimated by using geostatistics method [16], [17]. Thus, the boundary of orebody can be determined and the 3D model of the orebody can be established [18]. Orebody delineation is the basis for accurately describing the spatial distribution pattern of the ore body and plays an important role in reserves calculation and mining design. Properly constraining the shape and size of an orebody requires a complete database of geophysical and geological information derived from both surface and borehole data. Due to the limitation of technical conditions and equipment conditions, lots of basic geological data have been lost [19]. Then it makes it impossible to provide the complete and accurate geological data for the modeling process of the orebody, which reduces the accuracy of the orebody shape and that of reserves estimation.

A. The Estabishment of the Model

The geological missing data prediction model of neural network is established. The geological coordinate data, as the input of the back-propagation neural network model, are used to establish the input vector. The corresponding ore grade data is used to establish the output vector. Its network framework is a single hidden layer three-layered neural network with 3 input neurons and 1 output neuron. Based on the training samples and the empiric formula, the number of hidden layer neurons is 10. The transfer function between the input layer and the hidden layer is 'tansig', and 'purelin' is used as the transfer function between the hidden layer and the output layer. The network training function is the 'trainlm' function. In view of the large amount of the sample data, in order to improve the learning speed and the effectiveness of function approximation, the network structure is properly adjusted in the application process.

B. The Training and Prediction of the Model

1) The regularity of geological data of all boreholes

The x, y, z values of the samples centroid of all boreholes as the input of the back-propagation neural network model, are used to establish the input vector, and corresponding ore grade data are used to establish the output vector. The results of training and test show that MSE is small and closes to the target value, but the goodness of fit R value between the prediction result and the prediction object is small, which indicates there is no significant regularity in the geological data of all boreholes.

2) The regularity of geological data of single borehole

In order to study and analyze the grade data characteristic of boreholes, boreholes A, B, C, D, E, F, G and H are randomly selected to train and test the network respectively. By using the geological data of these boreholes, the network models are trained in different directions (XYZ, x, y, z) respectively, and some data are used to test the network. The results of training and test show that the training effectiveness of each borehole in at least one direction is good. The data of boreholes A, B, C, D are less affected in z coordinate direction than in other directions. The data of the borehole E are less affected in y coordinate direction than in other directions. The data of the boreholes F, G, H are significant affected by each coordinate direction.

Based on the similarity of the data of boreholes and the training effectiveness of each borehole, the data of boreholes are divided into different group, and the data of boreholes of the same group are trained together. For example, the x, y, z values of the samples centroid of boreholes A,B,C,D, as the input of the back-propagation neural network model, are used to establish the input vector, and corresponding ore grade data are used to establish the output vector. The training results show that R is 0.8497 and MSE is 2.26. The training effectiveness is good.

The neural network models trained by the samples data are used to predict the missing data of 8 boreholes. The analysis of prediction results of boreholes A, B, C, D shows that R values are more than 0.9226, and MSE values are less than 0.12. The R values of prediction results of boreholes F, G, H are more than 0.9443, and MSE values are less than 0.19. Therefore, it is scientific that the data of boreholes with same characteristics are trained together.

3) The regularity of geological data of group boreholes

Based on the data characteristics of boreholes, the continuity of ore body and the distance from the main fault, all boreholes are divided into five groups. The network models are trained by using five groups of the geological data. The number of samples in the first group is 414, and the analysis of the training results shows that R value is 0.9623. The number of samples in the second group is 1433, and the analysis of the training results shows that R value is 0.8791.

The number of samples in the third group is 10760, and the analysis of the training results shows that R value is 0.7449. Numbers of samples in the fourth group and fifth group are over forty thousand, and the analysis of the training results shows that R values are close to 0. The training effectiveness of the first group samples, the second group samples and the third group samples are much better than that of the fourth group samples and the fifth group samples. The analysis results show that the regularity of the data is not significant due to the amount of samples is too large. The results obtained by mathematical statistics show that the kurtosis values of the ore grade data of the fourth group and the fifth group are much more than 0. The regularity of the data is not significant, which make the training effectiveness of the data is worse. Therefore, if distributions of a large sample of the geological data are not normally distributed, the training effectiveness of the BP neural network will be not good.

The results obtained by mathematical statistics show that the kurtosis values of natural logarithm values of ore grades of the fourth group samples and the fifth group samples are 0.697 and 0.540 respectively. Due to the kurtosis values are close to 0, the ore grade data after taking natural logarithm are approximate normally distributed. The network models are trained by using the ore grade data after taking natural logarithm. The analysis of the training results of the fourth group samples and the fifth group samples shows that R values are 0.5338 and 0.5702 respectively. From the comparison of the train results, it was known that the better fitting effect can be gotten using the ore grade data after taking natural logarithm to train network.

V. CONCLUSIONS

1) Based on the characteristics of the technical and economic data of mining enterprises, the prediction and interpolation methods of the technical and economic data are analyzed.

2) The prediction model of the mineral products price is built using BP neural network. The prediction results show that the prediction model practicability of the prediction model is strong and the prediction accuracy of the prediction model is high.

3) The prediction and interpolation model of the geological missing data is built using techniques of geostatistics and BP neural network. It has been proved that most of the geological missing data can be predicted and interpolated by using the model, and the results of prediction and interpolation are reliable.

ACKNOWLEDGMENT

Supported by “Key Projects in the National Science & Technology Pillar Program during the Twelfth Five-year Plan Period (Grant No. 2012BAB01B04)” and “The Fundamental Research Funds for the Central Universities (Grant No. FRF-BD-16-001A)”.

REFERENCES

[1] Ming Jian, Hu Nailian, “The Enterprise Performance Management System for Mining Enterprises”, Advanced Materials Research ,Vols.433-440, pp.3276-3283,2012.

[2] Ming Jian, Hu Nailian, “The operations plan management system of metallurgical mining enterprise based on Business Intelligence”, 2nd International Conference on Information Science and Engineering, ICISE2010, Vol.I, pp.742-747, 2010.

[3] Wu Qi, Yan Hongsen, “Forecasting Model of Product Sales Based on the Chaotie v-Support Vector Machine”. Journal of mechanical engineering, Vol. 46, No.7, pp.102-103, April 2010.

[4] Hu Nailian, Zu Jianhua,”Mining planning of underground mine by computer”, Nonferrous Metals (Mining Section), No.6, pp.32, 1993.

[5] He Xichun, Hu Nailian, “Study of Mining Enterprise DSS Based on Data Warehouse”, China Mining Magazine, Vol.16, No.4, pp.4-7, April 2007.

[6] Dong Weijun, “Intelligent Decision System for Mining Plan”, Metal Mine, No.309, pp.10-16, March 2002.

[7] Peng Huaiwu, Liu Fangrui, Yang Xiaofeng, “Short term wind speed forecast based on conmbined prediction”, Acta energiae solaris sinica, Vol.32, No.4, pp.543-547, April 2011.

[8] Pan Guihao, Hu Nailian, Liu Huanzhong, et al., “Empirical analysis of gold price based on ARMA-GARCH model”, Gold, vol. 31, no. 1, p. 5, 2010.

[9] Hu Nailian, Liu Jiangtian, “Application of grey system theory in gold cost forecasting”, China Mining Magazine, Vol.8, No.3, pp.65, 1999.

[10] Zen Jiemei, “Lack of data processing in the ‘green mine’ applicaiton”, Anhui Univeristy of Technology, 2012.

[11] Zhang Lei, “Research on the interpolation algorithm for the geological space visualization -Take the urban district of ZheJiang-JinHua as example”, Nanjing Normal Univeristy, 2006.

[12] Robert H. Shumway, David S. Stoffer, Time series analysis and its applications:with R examples, 2nd ed., New York: Springer, 2006.

[13] Lu Qi, Gu Peiliang, Qiu Shiming. “The construction and application of combination forecasting model in Chinese energy consumption system”, Systems Engineering-Theory & Practice, No.3, pp.24, 2003.

[14] Tan Pangning, Michael Steinbach, Vipin Kumar, Introduction to data mining, Boston: Pearson Addison Wesley, 2006.

[15] Simon Haykin, Neural Networks and Learning Machines, 3rd ed., London: Pearson Academic, 2008.

[16] Peter K.Fullagar, Binzhong Zhou, Gary N.Fallon, “Automated interpretation of geophysical borehole logs for orebody delineation and grade estimation”, Mineral Resources Engineering,Vol.8, No.3, pp.269-284, 2010.

[17] GN Fallon,PK.Fullagar, B Zhou, “Towards grade estimation via automated interpretation of geophysical boreholelogs”, Exploration Geophysics,Vol.31, No.2, pp.236-242, 2000.

[18] Cheng Daogui, “Research on optimum control technology of kriging variance and its applicaiton” , Univeristy of Science and Technology Beijing, 2010.

[19] Zhang Lingling,Li Guoqing, Kang Kuangsong, Li Wei, et al., “A BP neural network based method for geological missing data processing”, Gold Science and Technology,Vol.23, No.5, pp.53-58, 2015.