CHEMICAL ENGINEERING TRANSACTIONS
VOL. 62, 2017

A publication of
AIDIC
The Italian Association
of Chemical Engineering
Online at www.aidic.it/cet

Guest Editors: Fei Song, Haibo Wang, Fang He

ISBN 978-88-95608- 60-0; **ISSN** 2283-9216

DOI: 10.3303/CET1762135

Application of Data Mining in Chemical Production

Yanpeng Guan

Zhengzhou Institute of Technology, Zhengzhou 454000, China
guanyanpeng1359@163.com

In order to strengthen the safety management of chemical enterprises, improve the quality of operators and management personnel, a dynamic linear modeling method combining data mining with BP neural network is proposed for coal gasification furnace. The coal gasification furnace can indirectly reflect the upstream temperature, the down temperature and the sum of up and down temperature of the gasification layer. On the one hand, the dynamic description of the system model is used to accurately reflect the magnitude and variation trend of the gasification layer temperature. On the other hand, the variables that can be controlled in the system is effectively utilized. The results show that the modeling method proposed in this paper has higher accuracy than other models, and it reduces the training time of neural network. At the same time, the variable of the modeling method is the variable that can be controlled by the gas making system, so the model can be used in practical application. Therefore, it is concluded that the research work in this paper is beneficial to the prediction and control of the coal gasification system, and establishes the prerequisite for the simulation optimization of the control system in the plant.

1. Introduction

At present, coal gasification mainly uses fixed bed intermittent gasification technology. However, it is urgent to solve the problems of energy saving, consumption reduction and output improvement in the fixed bed batch gasifier. (Janusz et al., 201). Therefore, the study of coal gasification system modeling has certain help to improve energy efficiency and optimize coal gasification system. In actual production, the data produced by gasification process are mostly time series data. Therefore, the modeling of time series can be applied to the chemical industry (Khareen et al., 2011).

The data studied in this paper are derived from the gas making workshop of a chemical industry group. Most of the data of gas making process are collected through distributed control system (DCS) (Li et al., 2017). At present, the rule of workshop operation is to infer the temperature of gasification layer according to the change curve of data collected by DCS system. The corresponding adjustment measures are adopted to stabilize the furnace condition to achieve high production. The upstream temperature, the down temperature and the sum of up and down temperature are selected as reference parameters (Khareen et al., 2014). Therefore, it is a new breakthrough to realize the single furnace efficiency and energy saving by using a large amount of gas data to establish the mathematical model of the single furnace (Vazan et al., 2017). Combined with the coal gasification process, the BP neural network modeling method is applied in the coal gasification process (Charoula et al., 2017). The combination of practical application and the dynamic linear model of neural network are put forward.

To strengthen the safety management of chemical enterprises and improve the quality of operators and management personnel, this paper proposed a dynamic linear modeling method combining data mining with BP neural network for coal gasification furnace. The coal gasification furnace can indirectly reflect the upstream temperature, the down temperature and the sum of up and down temperature of the gasification layer. On the one hand, the dynamic description of the system model is used to accurately reflect the magnitude and variation trend of the gasification layer temperature. On the other hand, the variables that can be controlled in the system is effectively utilized.

Please cite this article as: Yanpeng Guan, 2017, Application of data mining in chemical production, Chemical Engineering Transactions, 62, 805-810 DOI:10.3303/CET1762135

2. Linear modeling method combined with BP neural network

BP neural network is a multilayer feedforward network that corrects weights according to the error back-propagation algorithm (Raheela et al., 2017). This kind of network can be used to realize highly nonlinear mapping by combining simple nonlinear functions. At present, it is widely used in nonlinear system modelling. BP neural network is usually composed of input layer, hidden layer and output layer (Francesc et al., 2017). Each layer is connected with each other, and there is no feedback relationship between input and output.

2.1 Multiple linear regression modeling method combined with neural network

The dynamic characteristics between process index and output of complex industrial process are characterized by strong nonlinearity, strong combination and it is difficult to describe with accurate model. Moreover, the data in the process of gasification are time series (Jesús et al., 2017), and most of the variables are autocorrelated. Therefore, only relying on multiple linear regression model cannot describe the gasification process well.

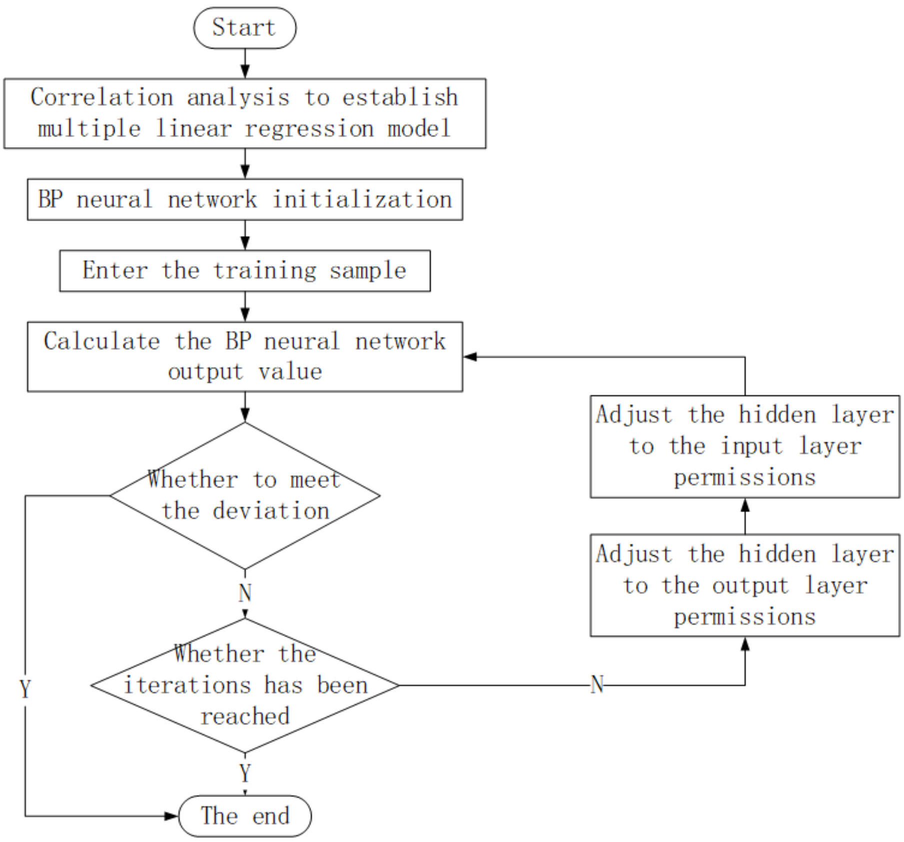

Figure 1: Multiple linear regression model combined with BP neural network flow chart

The combination of multiple linear regression model and neural network can complement each other and improve the accuracy of the model (Wang et al., 2015). Because of the existence of multiple regression model, the dimensionality of neural network input data is reduced, and the training time of neural network is reduced to some extent. Similarly, the existence of the network makes up for the shortcoming that multiple linear regression cannot describe the nonlinearity of the system (Xu et al., 2015). Therefore, the multiple linear regression modeling method combined with neural network is designed. This method takes the difference between the output value of the linear regression model and the actual value of the dependent variable as the expected output of the neural network, and takes the variables other than the linear model as the sample set to train the optimization method.
For BP neural network, When the power values vki, wjk and threshold bi, bj are given initial values, the corresponding output is obtained by a set of inputs. In general, the output Ok of the network does not match the actual data xk well, and the total error E between them is shown in formula (1).

$$E = \frac{1}{2}\sum_{k}(x_k - O_k)^2 \tag{1}$$

The training of neural networks refers to the repeated correction of weights and thresholds. The basic idea is that the weight drops rapidly (Ren et al., 2014) along the negative gradient direction of the function, so as to find the optimal solution. The flow chart of the multiple linear regression model combined with BP neural network is shown in Figure 1.

2.2 Dynamic linear modeling method combined with BP neural network

In the practical application of industrial process, the modeling with controlled variables is beneficial to the optimization and control of industrial system (Yu and Xu, 2014). The basic idea of combining dynamic linear model of the neural network method is: According to the above correlation analysis, we can get the variables most closely related to the dependent variables, and then select controllable variables according to the specific industrial process (Su et al., 2013). Because the industrial data are mostly time series, the dynamic performance of the system should be considered, that is, the influence of the output value of the first few moments on the current time system. A new dynamic linear model is constructed using the above variables, as shown in formula (2).

$$Y(k) + \alpha_0 Y(k-1) + ... + \alpha_m Y(k-m-1) = \beta_0 + \beta_1 X_1(k) + ... + \beta_{p-1} X_{p-1}(k) \tag{2}$$

The difference between the linear model and the real data is used as the expected output of the BP network, and the variables other than the linear model are used as inputs to the BP network. The BP network is trained according to the desired output (Ryan and Paul, 2014). This method has the advantages of BP neural network that can predict the characteristics of unknown samples, and it is applied to the practical industry. This method not only reduces the dimension of BP training samples, but also improves the approximation speed of BP neural network and the accuracy of the model (Jacob et al., 2017). The structure of the dynamic linear model combined with the neural network is shown in Figure 2.

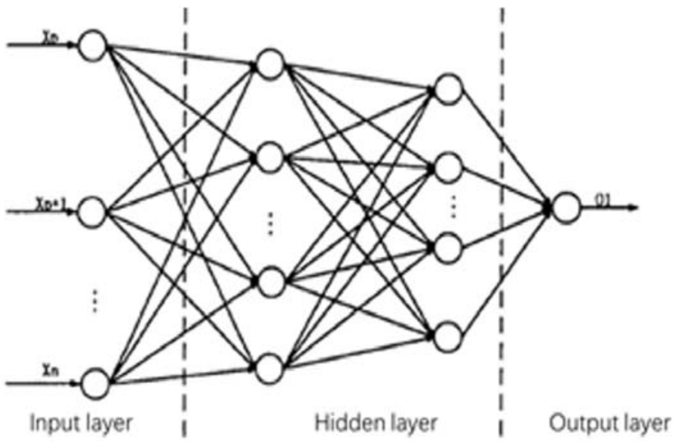

Figure 2: Dynamic linear model combined with BP neural network

3. Application of linear modeling method combined with neural network in coal gasification system

The data processed in this paper are derived from a coal gasification system in a chemical plant. The partial continuous intervals data of the furnace from May 12, 2016 to May 23 are used for simulation, and the data interval is 6 minutes (Yuan and Eric, 2017). The experimental data are processed by the data compensation algorithm based on hierarchical clustering. Firstly, the historical data of missing data are clustered and analyzed to find out the nearest group of data which is the closest to the missing data, and then the compensation is made according to the interpolation estimation (Wang et al., 2017).

Taking the descending temperature x1, west ash temperature x2, furnace speed x3, top blowing number x4, jacket temperature x5 and the furnace steam flow x6 as linear model variable, according to the linear modeling algorithm combined with BP neural network, the ascending temperature is modeled. The input layer contains 11 nodes, and the number of nodes in the middle-hidden layer is [4,4,4] with 3 layers. The number of nodes in the output layer is 1, which represents the ascending temperature. (Xie, 2017). After constructing the network, the network is trained by MATLAB7.1 experimental platform. The parameters of the experiment are as follows: the training step is 50000, the target precision is 0.001, and the learning rate is set to 0.15. The simulation results are shown in Figure 3 and Figure 4.

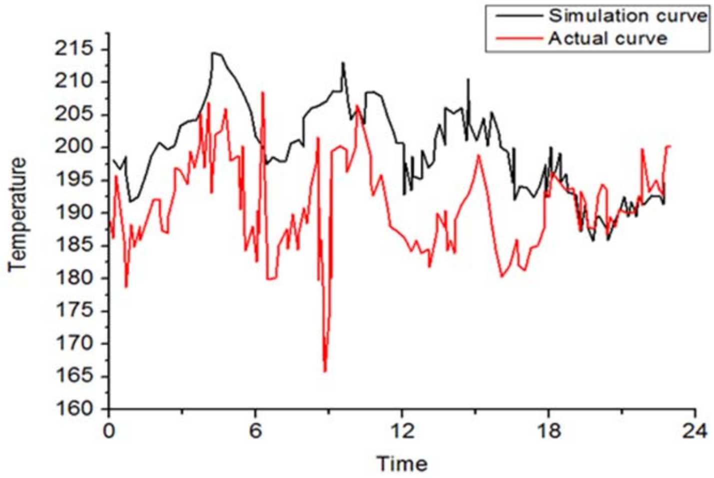

Figure 3: Simulation Results of Multivariate Linear Regression Model with Uplink Temperature Combined with BP Neural Network

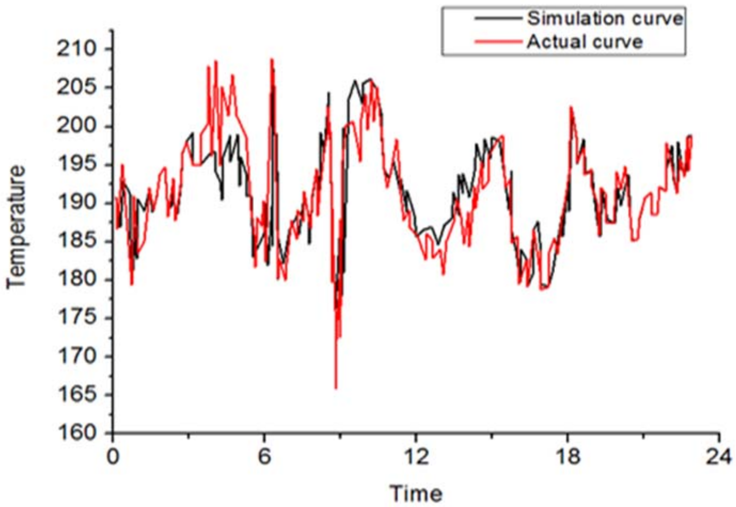

Figure 4: Dynamic Linear Model Simulation Results of Uplink Temperature Based on Similarity Analysis Based on BP Neural Network

The mean square error of the model is 139.2734 by using the multiple linear regression graph combined with the neural network. The design of the neural network is unchanged. The mean square error of the model obtained by the dynamic linear method based on similarity analysis is 22.7316.

Because the variables of industrial system are not all controllable, the application oriented dynamic linear model combined with BP neural network is adopted. The setting of neural network is unchanged. The furnace speed x1 and furnace steam flow x2 are taken as the independent variable of linear model (Ma et al., 2017). The simulation results are shown in Figure 5, and the mean square error is 22.0318.

The simulation results show that the dynamic linear model combined with BP neural network proposed in this paper has good effect. Compared with the use of neural network alone, it has the advantages of high model accuracy, accurate trend prediction and short training time. Compared with the multiple linear regression model combined with BP neural network and the dynamic neural network model combined with neural network based on similarity analysis, it is closer to the actual industrial production, and is a kind of model for practical application.

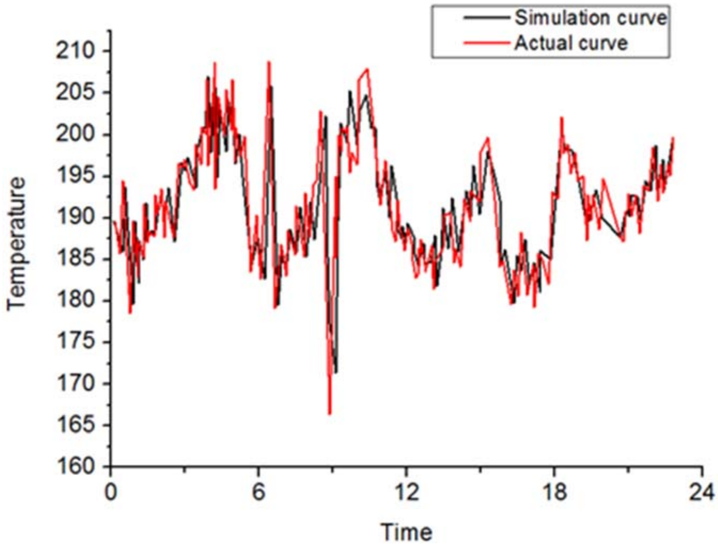

Figure 5: The simulation results obtained by using the dynamic linear modeling method of BP neural network with uplink temperature

4. Conclusion

A modeling method based on BP neural network and multiple linear regression is designed, and the operation steps of the algorithm are introduced. At the same time, in order to apply the system model to the actual industrial production and consider the dynamic performance of the system, a dynamic linear model combined with BP neural network is proposed, while the idea and calculation steps of the algorithm are introduced in detail. In addition, based on the historical data of a chemical plant, the ascending temperature of coal gasification system is modeled. The results show that the dynamic linear model combined with BP neural network is close to the actual value, and the model accuracy is high. The experimental results verify the superiority of the proposed algorithm. It is concluded that the research work in this paper is beneficial to the prediction and control of the coal gasification system, and establishes the prerequisite for the simulation optimization of the control system in the plant.

Reference

Charoula A., Sarah K.H., Jun M., Jie Y., Paul A.K., 2017, Data mining in educational technology classroom research: Can it make a contribution? Computers & Education, 113, 226-242. DOI: 10.1016/j.compedu.2017.05.021

Francesc V., Ernesto R., David F., Ricardo T.K., Sergi V., Nuria M., 2017, Urban data and urban design: A data mining approach to architecture education, Telematics and Informatics, DOI: 10.1016/j.tele.2017.09.015

Jacob P.M., Yamin P., Perez-Storey C., Hopgood M., Lapkin A.A., 2017, Towards automation of chemical process route selection based on data mining, Green Chemistry, 19, 140-152, DOI: 10.1039/C6GC02482C

Janusz J.S., Joy H., Markus K., 2017, Blockchain technology in the chemical industry: Machine-to-machine electricity market, Applied Energy, 195, 234-246, DOI: 10.1016/j.apenergy.2017.03.039

Jesús P., Alejandro M., Manuel M., 2017, Application of Data Mining techniques to identify relevant Key Performance Indicators, Computer Standards & Interfaces, 54, 76-85, DOI: 10.1016/j.csi.2016.11.006

Khareen S., Christian I., Christopher O., Jane P., Nikolaos V., 2011, Developing a screening method for the evaluation of environmental and human health risks of synthetic chemicals in the mining industry, International Journal of Mineral Processing, 101(1), 1-20, DOI: 10.1016/j.minpro.2011.07.014

Khareen S., Christopher O., Jane P., Nikolaos V., 2014, Undisclosed chemicals — implications for risk assessment: A case study from the mining industry, Environment International, 68, 1-15, DOI: 10.1016/j.envint.2014.02.012

Li J., Zhang Y., Du D., Liu Z., 2017, Improvements in the decision making for Cleaner Production by data mining: Case study of vanadium extraction industry using weak acid leaching process, Journal of Cleaner Production, 143, 582-597, DOI: 10.1016/j.jclepro.2016.12.071

Ma D.L., Zhou T., Chen J., Qi S., Shahzad M.A., 2017, Supercritical water heat transfer coefficient prediction analysis based on BP neural network, Nuclear Engineering and Design, 320, 400-408, DOI: 10.1016/j.nucengdes.2017.06.013

Raheela A., Agathe M., Syed A.A., Najmi G.H., 2017, Analyzing undergraduate students' performance using educational data mining, Computers & Education, 113, 177-194, DOI: 10.1016/j.compedu.2017.05.007

Ren C., An N., Wang J., Li L., Hu B., Shang D., 2014, Optimal parameters selection for BP neural network based on particle swarm optimization: A case study of wind speed forecasting, Knowledge-Based Systems, 56, 226-239, DOI: 10.1016/j.knosys.2013.11.015

Ryan S.B., Paul S.I., 2014, Educational Data Mining and Learning Analytics, Learning Analytics, 61-75, DOI: 10.1007/978-1-4614-3305-7_4

Su L., Zha Z., Lu X., Shi T., Liao G., 2013, Using BP network for ultrasonic inspection of flip chip solder joints, Mechanical Systems and Signal Processing, 34(1), 183-190, DOI: 10.1016/j.ymssp.2012.08.005

Vazan P., Janikova D., Tanuska P., Kebisek M., Cervenanska Z., 2017, Using data mining methods for manufacturing process control, IFAC-PapersOnLine, 50(1), 6178-6183, DOI: 10.1016/j.ifacol.2017.08.986

Wang J., Wen Y., Ye Z., Jian L., Chen H., 2017, Convergence analysis of BP neural networks via sparse response regularization, Applied Soft Computing, 61, 354-363, DOI: 10.1016/j.asoc.2017.07.059

Wang Y., Lu C., Zuo C., 2015, Coal mine safety production forewarning based on improved BP neural network, International Journal of Mining Science and Technology, 25(2), 319-324, DOI: 10.1016/j.ijmst.2015.02.023

Xie L., 2017, The Heat Load Prediction Model based on BP Neural Network-markov Model, Procedia Computer Science, 107, 296-300, 10.1016/j.procs.2017.03.108

Xu Y., You T., Du C., 2015, An integrated micromechanical model and BP neural network for predicting elastic modulus of 3-D multi-phase and multi-layer braided composite, Composite Structures, 122, 308-315, DOI: 10.1016/j.compstruct.2014.11.052

Yu F., Xu X., 2014, A short-term load forecasting model of natural gas based on optimized genetic algorithm and improved BP neural network, Applied Energy, 134, 102-113, DOI: 10.1016/j.apenergy.2014.07.104

Yuan Y., Eric M., 2017, Life-Cycle Modeling Framework for Generating Energy and Greenhouse Gas Emissions Inventory of Emerging Technologies in the Chemical Industry, Journal of Cleaner Production, DOI: 10.1016/j.jclepro.2017.10.125