http://ansinet.com/itj

ITJ

ISSN 1812-5638

INFORMATION TECHNOLOGY JOURNAL

ANSInet
Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Information Technology Journal 12 (15): 3122-3127, 2013
ISSN 1812-5638 / DOI: 10.3923/itj.2013.3122.3127

Study on Soft-sensing Model of Carbon Content in Fly Ash Based on Support Vector Regression

[1]Bian He-Ying, [2]Li Zeng-Quan and [3]Fang Yan-Jun
[1]College of Electrical and Information Engineering, Pingdingshan University, Henan, 467000, Pingdingshan, China
[2]Henan Quality Polytechnic, Henan, 467000, Pingdingshan, China
[3]Department of Automation, Wuhan University, Hubei, 430072, Wuhan, China

Abstract: A novel soft-sensing model of the carbon content in fly ash is established by applying support vector regression and article swarm algorithm about the problem that the carbon content in fly ash is difficult to measure accurately. The carbon content in fly ash is an important index to judge the quality of boiler operation and coal utilization rate and affects directly the efficiency of the boiler. In this work, carried on data preprocessing, the various variables' correlation analysis and dapted article swarm algorithm to optimize the parameters C and g of the model according to the real time data of a Power Plant 1000MW ultra-supercritical unit. Moreover, the model's accuracy and generalization capability were dentified by using the test data. The simulation results show that soft-sensing model built has a higher prediction accuracy and better generalization ability than the previous measurement methods which provides an effective way for measurement of carbon content in fly ash in power plant.

Key words: Carbon content in fly ash, support vector regression, particle swarm optimization algorithm, soft sensing

INTRODUCTION

The carbon content in fly ash is an important index to judge the quality of boiler operation and coal utilization rate and affects directly the efficiency of the boiler. Correlation research shows that the boiler efficiency can be improved one percentage point when the carbon content in fly ash is decreased about three percent (Han *et al*., 2007). Therefore, timely adjustment of combustion conditions according to the accurate measurement value of carbon content can improve boiler combustion control level and reduce the cost of power generation. But the modern methods of measuring the carbon content in fly ash have some problems such as poor real-time, low accuracy and the carbon content can't be measured when carbon content in fly ash is less than 0.75% (Li *et al*., 2004). Therefore, it is very difficult to establish calculation model or on-line direct accurate measurement because many factors that influence the carbon content are intercoupling and non-linear (Chen and Liu, 2005). To establish soft measurement model of carbon content in fly ash is an effective method which solve this problem (Deng *et al*., 2011; Wang *et al*., 2005). In this study, soft measurement model of carbon content in fly ash is established by applying support vector machine and particle swarm optimization.

SUPPORT VECTOR REGRESSION THEORY

The basic idea of support vector regression is that the non-linear problems of low dimensional space are transformed into linear problems of a high dimensional space by introducing the kernel function $K(x_i, x_j)$.

If $\{(x_1, x_1,),\ldots, (x_n, x_n)\}$ is the training sample set, then regression function is Eq. 1 (Zhang and Bao, 2009):

$$F = \{f \mid f(x) = w^T\phi(x) + b, w \in R^n\} \quad (1)$$

where, w stands for weight vector, b is threshold. R_{emp} is structure risk function:

$$R_{reg}(w) = \frac{1}{2}\|w\|^2 + C \cdot R_{emp}[f(x)] = \frac{1}{2}\|w\|^2 + C \cdot \frac{1}{n}\sum_{i=1}^{n}|y - f(x)| \quad (2)$$

where, $\|w\|^2$ describes the model complexity, the role of the penalty coefficient C is a compromise between experience

Corresponding Author: Bian He-Ying, College of Electrical and Information Engineering, Pingdingshan University, Henan, 467000, Pingdingshan, China

risk and the complexity of the model, n stands for the number of training samples. w can be counted according to Eq. 3:

$$w = \sum_{i=1}^{n} (\alpha_i - \alpha_i^*) \phi(x_i) \quad (3)$$

a_i, a^*_i is solution of R_{reg} minimized, x_i is support vector, n stands for the number of training samples. f(x) can be represented as Eq. 4 according to the Eq. 3:

$$f(x) = \sum_{i=1}^{n} (\alpha_i - \alpha_i^*) [\phi(x_i) \cdot \phi(x)] + b \quad (4)$$

Non-linear support vector regression function f(x) can be defined according to the kernel function $K(x_i, x_j) = f(x_i).f(x_j)$, it is shown in Eq. 5 (Mohandes *et al*., 2004):

$$f(x) = \sum_{i=1}^{n} (\alpha_i - \alpha_i^*) K(x_i, x) + b = \sum_{i=1}^{n} w_i \cdot K(x_i, x) + b \quad (5)$$

where, w_i is the coefficient of support vector, x_i is support vector, $K(x_i, x)$ is kernel function, b is defined as Eq. 6:

$$b = y_i - w.\phi(x_i) \quad (6)$$

Introduction of slack variable e can increase the robust of vector regression. Concrete expression is shown in Eq. 7:

$$|y - f(x)|_\varepsilon = \begin{cases} 0 & |y - f(x)| \le \varepsilon \\ |y - f(x)| - \varepsilon & \end{cases} \quad (7)$$

METHOD OF ESTABLISHING MODEL BY SUPPORT VECTOR REGRESSION

Parameter selection: In this study, the radial basis kernel function $(\exp(-g \times |x_i - x_j|^2))$ is selected as Kernel function of support vector regression model. Training parameters of the support vector regression model have four parameters; they are kernel function parameter g, penalty coefficient C, maximum allowable error e and insensitive loss function e. Parameters ε and ε are controlled by the human's behavior, they can often be ignored because of little impact on model prediction ability (Ren and Feng, 2002). Parameters $_C$ and g directly affect the modeling calculation process, the properties of the model and great impact on the prediction precision and generalization ability (Shen *et al*., 2005). This research selects the particle swarm optimization algorithm to optimize parameters C and g in order to obtain good parameters.

Method of PSO-SVR model: In this research, the method that support vector machine is combined with optimization algorithm is known as PSO-SVR model method.

Final prediction effect is evaluated by using Mean Square Error (MSE) and squared correlation coefficient r^2, expression equations of MSE and r^2 are Eq. 8-9. The value of MSE is the smaller, prediction precision of the model is the better; Square correlation coefficient is the more close to 1, the result of regression fitting is the better (Li *et al*., 2007):

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (f(x_i) - y_i)^2 \quad (8)$$

$$r^2 = \frac{\left(n\sum_{i=1}^{n} f(x_i) y_i - \sum_{i=1}^{n} f(x_i) \sum_{i=1}^{n} y_i \right)^2}{\left(n\sum_{i=1}^{n} f(x_i)^2 - \left(\sum_{i=1}^{n} f(x_i) \right)^2 \right) \left(n\sum_{i=1}^{n} y_i^2 - \left(\sum_{i=1}^{n} y_i \right)^2 \right)} \quad (9)$$

where, $f(x_i)$ stands for prediction value of the output, y_i stands for actual value of the output.

PSO-SVR training steps are as follows:

Step 1: To choose the training data collected from field, reduce noise and normalize data

Step 2: Optimal parameters C and g are obtained by applying particle swarm optimization algorithm to identify training data. Detailed process is as follows:

- **Initialize the particle population:** Set the number of iterations is 200, particle number is 20 and randomly generate a set of initial position and velocity of the particles
- Evaluation fitness by using the fitness function MSE
- Replace the previous position with the current best position 'pbest' if the current fitness value of particles is better than the previous best position 'pbest'; Replace the previous global best position with the current best global position 'gbest' if the current fitness value of particles is superior to the global best position experienced; Where, 'pbest' stands for the current best position of particle, 'gbest' stands for the current best position of group
- Update the particle's speed and position according to the Eq. 10-11:

$$V_i = w_v \times V_i + c_1 \times Rand() \times pBest[i] - X_i + c_2 \times Rand() \times (pBest[g] - X_i) \quad (10)$$

$$X_i = X_i + V_i \tag{11}$$

where, the constant c_1 and c_2 is called learning factors, Rand () and Rand () are random number in [0, 1], w_v is inertia weight that control the degree which the past speed effect on the current speed, is particle position, is particle speed:

- The iteration stop and output optimal solution when the number of iterations achieve target value, otherwise, jump to step 1

Step 3: To perform the SVR algorithm, train training data and get the SVR model

Step 4: To verify model's accuracy and generalization capability by using the test data

SOFT-SENSING MODEL OF CARBON CONTENT IN FLY ASH BASED ON PSO-SVR

Data preprocessing: The reliability and accuracy of the data affect directly the quality of the soft sensor model and affect indirectly the process control and optimization operation of the boiler. Field data is error because the measuring instruments that obtained data are affected by the measuring method, precision of instruments and production environment. Error data must be removed in order to ensure the accuracy of the soft measurement model. The error output curve of the carbon content in fly ash which is forecasted by applying the soft-sensing model under the data is not preprocessed is shown in Fig. 1.

Figure 1 shows that the simulation results are not ideal that maximum of relatively error is over 40%, the Mean Square Error (MSE) is 0.0825208 and the squared correlation coefficient (r^2) is 0.590623.Therefore, the data must be pretreated before the ideal soft measurement model is established. Where,, the gross error is eliminated by using Pauta criterion and random error is eliminated by using linear smoothing.

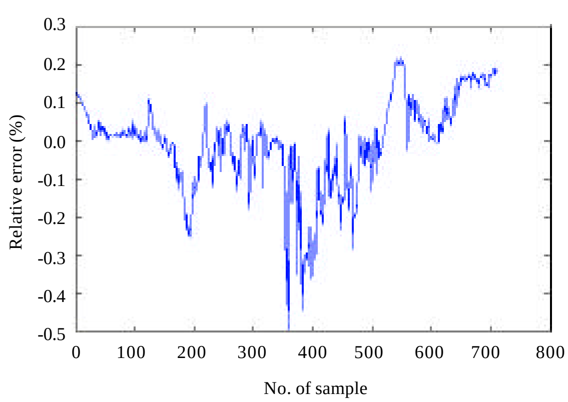

Fig. 1: Error output curve of the carbon content in fly ash

Pauta criterion: $y_1, y_2, \ldots, y_n$ is assumed to be a set of sample data, $\bar{y}$ is average value of the sample data, deviation of a single sample data is $v_i = y_i - \bar{y}$ ($i = 1, 2, \ldots, n$), standard deviation of sample data $_s$ can be calculated according to Eq. 12:

$$\sigma = \left(\sum_{i=1}^{n} v_i^2 \Big/ n-1\right)^{1/2} = \left\{\left[\sum_{i=1}^{n} y_i^2 - \left(\sum_{i=1}^{n} y_i^2\right)\Big/ n\right]\Big/(n-1)\right\}^{1/2} \tag{12}$$

Linear smoothing: $\{x_1, x_2, \ldots, x_n\}$, $x_i = s_i + n_i$ is assumed to be a group of sample data, $_{ni}$ is truth value, $_{si}$ is noise value, then linear smoothing data $_{yi}$ can be calculated by using Eq. 13:

$$y_i = \sum_{r=-q}^{q} a_r x_{i+r} = \sum_{r=-q}^{q} a_r (s_{i+r} + n_{i+r}) = \sum_{r=-q}^{q} a_r s_{i+r} + \sum_{r=-q}^{q} a_r n_{i+r} = \overline{s_i} + \overline{n_i} \tag{13}$$

$\{a_r\}$ is weight value and:

$$\sum_{r=-q}^{q} a_r = 1$$

in Eq. 13.

The five point linear smoothing and seven point linear smoothing are usually applied in engineering. The oxygen content data in flue gas is processed as an example; processing effect of these data by applying the two methods is shown in Fig. 2.

Figure 2 shows the data smoothing degree are all improved every method of the two but the five point

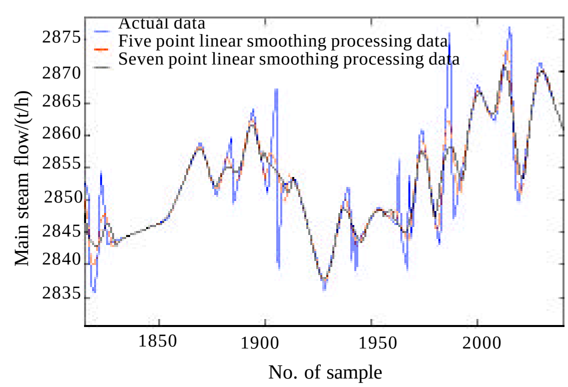

Fig. 2: Processing effect of the oxygen content data in flue gas

Table 1: Correlation value between the carbon content in fly ash and the other influening variables

Auxiliary variables	Correlation value	Auxiliary variables	Correlation value
Carbon content	1.0000	Main steam flow	0.8371
Coal quantity of feeder	0.8968	Total air volume	0.8829
Oxygen content	0.8136	Motor current of coal mills	0.8720
Flue gas temperature	0.9208	Entrance flow of the economizer	0.8255
Main steam pressure	0.7562	Outlet pressure of the economizer	0.6022
Main steam temperature	0.6417	Entrance pressure of the economizer	0.5963

linear smoothing processing data is closer to the original data than the seven point linear smoothing data. Therefore, the other variable data is processed by using five point linear smoothing method in this study.

Grey correlation analysis: There are many factors which influence the carbon content in fly ash, but the impact degree can be judged by the correlation value between the carbon content in fly ash and the other influencing variables. The factors with the small correlation value are discarded that will help to improve the calculation speed of the soft-sensing model built. Here 11 influencing factors are identified according to the mechanism analysis, such as coal quantity provided by feeder, flue gas temperature, oxygen content in flue gas, main steam temperature, main steam pressure, main steam flow, total air volume, entrance flow of the economizer, entrance pressure of the economizer, outlet pressure of the economizer and coal mills. Correlation values are shown in Table 1 between the carbon content in fly ash and the 11 auxiliary variables.

Zheng and Hou (2012) said that the correlation was strong when the correlation value is bigger than 0.65. Therefore, the 8 auxiliary variables in table 1 are chosen according to if the correlation value is greater than 0.65, soft-sensing model of carbon content in fly ash is established by the method that the 8 auxiliary variables are as input of soft-sensing model and carbon content values measured with carbon instrument are used as the output of the model.

Establishment of SVR model: To establish soft-sensing model according to PSO-SVR algorithm introduced in section 2.2. Simulation data is from the historical data of a 1000MW ultra-supercritical unit of Datang Chaozhou power station, the sampling period of these data is 5 sec.

Twenty four hours data of some day is chosen as data sample of model, 18 h data of data sample is as training data and 6 h data of data sample is as test data. Modeling data is preprocessed and normalized in order to guarantee the output good effect and accuracy of the soft-sensing model.

Initialization parameters of the particle swarm algorithm are as follows: learning factor $c_1 = 1.5$, $c_2 = 1.7$,

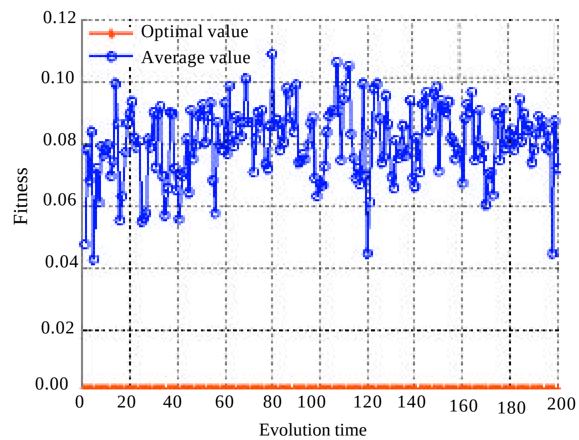

Fig. 3: Fitness curve of SVR optimized by PSO

population quantity pop = 20, maximum generation maxgen = 200. The fitness curve of the training sample set acquired by applying PSO-SVR is shown in Fig. 3.

Figure 3 shows that the soft-sensing model built has 76 support vectors, the kernel function parameter g = 28.6892 and constant b = -0.0537. Based on these parameters, the decision function of the SVR model is obtained by using MATLAB which is shown in Eq. 14:

$$f(x) = \sum_{i=1}^{n} w_i \exp\left(-g\left|x_i - x\right|^2\right) + b \tag{14}$$

where,, w_i is support vector coefficient, x_i is support vector of the model, x is the data for forecasting, g is the kerael function parameter, b is constant and n is the number of support vectors.

VERIFY THE VALIDITY OF THE SOFT SENSING MODEL OF CARBON CONTENT IN FLY ASH

To verify the validity of the soft-sensing model build in the third section by test data. The test results show that MSE is 0.02538, r^2 is 0.99847, the accuracy of the model and the regression fitting curve can meet the practical requirements. Simulation output curve of the carbon content in fly ash by using matlab is shown in Fig. 4.

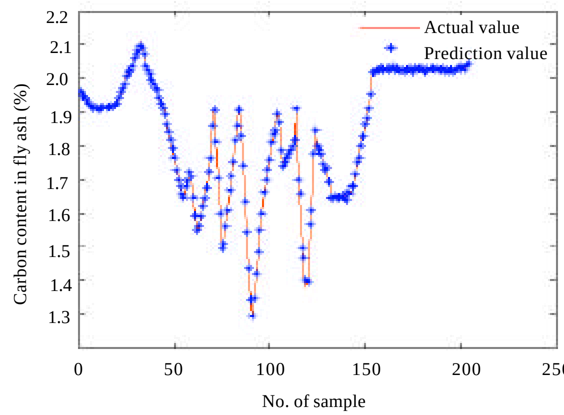

Fig. 4: Soft-sensing measurement output curve of carbon content in fly ash

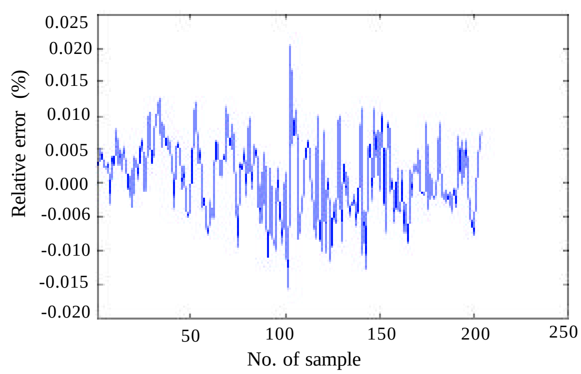

Fig. 5: Relative error curve of carbon content in fly ash

The dotted line represents the predicted value of the carbon content in fly ash by applying the soft-sensing model, the solid line represents measured value by carbon measuring instrument. Figure 4 shows the same trend of the rise and fall both the predicted value and measured value, only amplitude of variation is slightly different in the local range which verify the build model has high prediction accuracy and good generalization ability. Figure 5 is the relative error curve corresponding to Fig. 4.

The relative error is controlled within -0.5~1%, variance is 0.0033, as are shown from Fig. 5.

CONCLUSION

A soft sensor model of the carbon content of fly ash is established by applying support vector regression and article swarm lagorithm about the problem that the carbon content of fly ash is difficult to measure accurately. To optimize the parameters C and g of the model by dapting article swarm algorithm. Prediction accuracy and generalization capability of the model are improved by preprocessing acquisition data and verified by using test data. The simulation results show that the prediction value relative error of the soft-sensing model established is controlled within -0.5~1%, variance is only 0.0033.In summary, prediction accuracy and generalization capability of the soft-sensing model established can meet the practical requirements and provide an effective way for the measurement of carbon content of the boiler fly ash in power plant.

ACLNOWLEDGMENT

The authors would like to thank for the support by Natural Science Foundation of China under the Grant No. 61170024.

REFERENCES

Chen, M.S. and D.P. Liu, 2005. Optimized control of carbon content in utility boilers fly ash. Power Eng., 25: 545-549.

Deng, S., S.Y. Lin and W.L. Chang, 2011. Application of multiclass support vector machines for fault diagnosis of field air defense gun. Expert Syst. Appl., 38: 6007-6013.

Han, P., H. Qiao, D.F. Wang and Y. Zhai, 2007. Development and current status of thermal parameter soft-sensing technique in power plant. Chin. J. Sci. Instr., 28: 1139-1146.

Li, Z., J.J. Cai and H. Guo, 2004. Soft measurement system based on neural network for the unburned carbon in fly ash from utility boilers. Energy Conserv. Technol., 22: 6-7.

Li, R., H.N. Wang, H. He, Y.M. Cui and Z.L. Du, 2007. Support vector machine combined with K-nearest neighbors for solar flare forecasting. Chin. J. Astronomy Astrophys., Vol. 7. 10.1088/1009-9271/7/3/15

Mohandes, M.A., T.O. Halawani, S. Rehman and A.A. Hussain, 2004. Support vector machines for wind speed prediction. Renewable Energy, 29: 939-947.

Ren, B. and Z.P. Feng, 2002. Improved genetic algorithm and particle swarm optimization as well as comparison between them. J. Nanjing Normal Univ. (Eng. Technol.), 2: 14-20.

Shen, Y., B. Guo and T.X. Gu, 2005. Particle swarm optimization algorithm and comparison with genetic algorithm. J. Univ. Electr. Sci. Technol. China, 34: 696-699.

Wang, C.L., H. Zhou, Z.H. Zhou, Z.Q. Ling, G.N. Li and K.F. Cen, 2005. Support vector machine modeling on the unburned carbon in fly ash. Proc. Chin. Soc. Electr. Eng., 25: 72-76.

Zhang, G. and L. Bao, 2009. Soft-sensing modeling of the carbon content in fly ash based on information fusion for thermal power plant. Proceedings of the International Conference on Mechatronics and Automation, August 9-12, 2009, Changchun, China, pp: 3860-3865.

Zheng, J. and Y.Y. Hou, 2012. Soft measurement system based on hybrid model for the unburned carbon in fly ash from utility boilers. Sci. Technol. Inform., 11: 77-78.