Modeling of Main Steam Parameters in Large Thermal Power Plant

Wen Xiaoqiang¹ and Jian Shuguang²

¹simulation center, Northeast Dianli University, Jilin, China
²School of Automation Engineering, Northeast Dianli University, Jilin, China

Abstract. To solve the problem of main steam temperature abnormal of the 300 MW subcritical coal-fired boiler, the main steam temperature model was built based on running data. Firstly, the correlation analysis of the data was carried out based on the SPSS software to determine the main influence factors of the main steam temperature. Then, a nonlinear modeling method based on PLS was proposed, which is used to determine the number of the extracted components. At last, the simulation of the partial least squares regression model of the main steam temperature was realized by using MATLAB software. The partial least squares regression equation of the standard variable and the original variable was constructed by using the training samples and the accuracy and correctness of the model were tested by using the test samples.

1 Introduction

The development of information technology in power plant provides a good platform for data driven operation optimization research. Among them, based on the power plant real time & historical database, the method of modeling of the complex thermal system has gradually become one of the hot topics and it has become the basis of solving the key problems such as operation, control, performance evaluation, fault diagnosis, simulation and so on [1]. However, running data is different from the experimental data, which has many characteristics that are not conducive to modeling. For example, there are multiple correlation among variables, uneven distribution of the working conditions, and the existence of process nonlinearities, etc. All the problems are seriously hindering the development and application of data modeling methods [2]. Therefore, here the main steam temperature prediction model was built based on the Partial Least-squares Regression (PLS) algorithm.

2 Review of research methods and correlation analysis

2.1 Review of research methods

The main steam temperature of boiler has the obvious characteristics of nonlinearity, large inertia and long time-lag and in many cases, it is difficult to establish an accurate mathematical model. To solve the problem, Deng built the on-line fuzzy identification system of steam temperature based on T-S fuzzy rule model and the recursive least-square (RLS) [3]. Zhao developed an advanced Pareto multi-objective optimization with genetic algorithm based on a novel vector module adaptive function, which was applied to the PID controller parameters tuning in the boil superheated steam temperature cascade control system. The simulation results show that the improved multi-objective genetic algorithm (GA) optimization can obtain different optimal controller parameters with respect to the corresponding performance targets [4]. Zhang proposed the super-heater steam temperature model using on-line data to improve the conventional testing method and physical law methods. Based on on-line data, the identifiability of closed-loop control system of super-heater steam temperature could be met and dynamic models of super-heater steam temperature were obtained in the disturbance of super-heater spray flow. Simulations show that the model could reflect the facts of super-heater steam temperature [5]. Luan designed a new cascade control system based on the characteristics of superheated steam temperature of a boiler [6]. To solve the problem of reheater steam temperature (RST) abnormal of the 300MW power station boiler unit, Huang presented a method based on support vector regression (SVR). Based on the data sampled on spot, RST was analyzed using support vector regression method. The RST model was based on the statistical characteristics of the operating parameters and could reflect the potential relationship between RST and the operating parameters. For the units considered here, the RST was low and the temperature of reheater tube wall was high, tilt angles and desupreheater spray, etc. were taken as the tuning parameters and as the features of SVR model. The prediction results on test data with SVR-RST showed high regression coefficient with low complexity.
which means SVR-RST model had excellent robustness, which was important to further optimize the operating parameters for higher efficiency and security[7].

2.2 Correlation analysis

There are many factors that can affect the temperature of the main steam, such as steam parameters, water spray temperature, flue gas temperature and flow rate, ash fouling status and so on. In order to determine the main influencing factors, here the correlation analysis was used.

There are three kinds of calculation of correlation coefficient: Pearson, Spearman and Kendall. Here we selected the Pearson correlation coefficient. The formula of Pearson correlation coefficient is as followed:

\[
r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}}
\]

where the \( r \) is regarded as the correlation coefficient. The \( \bar{x} \) and \( \bar{y} \) are the corresponding mean values and \( x_i \) and \( y_i \) are the ith observe value.

According to the Table 1, the factors whose correlation coefficient was greater than 0.6, were chosen to build the partial least squares regression model.

**Table 1. Correlation analysis results**

| Influencing factors          | Correlation coefficients | Influencing factors          | Correlation coefficients |
|------------------------------|--------------------------|------------------------------|--------------------------|
| Main steam temperature       | 1.000                    | Feed water flow              | -0.129                   |
| Pressure of main steam       | 0.811                    | Front and rear differential pressure of the main feedwater gate | -0.920                   |
| Flow of main steam           | 0.923                    | Speed of feed water pump     | -0.114                   |
| Oxygen content in flue gas   | -0.701                   | Coupler opening of feedwater pump | -0.240                   |
| Total air flow               | 0.834                    | Low pressure side opening    | 0.340                    |
| Inlet baffle opening of primary air fan | 0.686             | High pressure side opening   | 0.332                    |
| Inlet baffle opening of forced draught fan | 0.310             | Outlet temperature of low pressure side | 0.436                    |
| Turbine speed                | 0.092                    | Outlet temperature of high pressure side | -0.148                   |
| Generator power              | 0.165                    | Regulating valve opening of oil fuel atomizing steam | 0.643                    |
| Drum water level             | 0.759                    | Speed of powder feeder       | 0.137                    |

3 Partial least-squares regression

3.1 Partial least-squares regression

The PLS method has a tremendous success in chemometrics and chemical industries for static data analysis. It integrates the principal component analysis (PCA) and canonical correlation analysis (CCA) together naturally and is convenient for the analysis of the multidimensional complexity system. In its general form, PLS creates components by using the existing correlations between different sets of variance while also keeping most of the variance of both sets. Now, we provide some of the formal ingredients of the method [8].

Consider a general setting of the PLS algorithm to model the relation between two data sets. Let \( X = \{x_1, x_2, ..., x_p\} \) denote a \( p \)-dimensional vector of variables in the first block of data, and similarly let \( Y = \{y_1, y_2, ..., y_q\} \) denote a vector of variables from the second set. Observing \( n \) data samples from each block of variables, PLS decomposes \( X = (x_i)_{n \times p} \) and \( Y = (y_j)_{n \times q} \) into the form:

\[
X = TP^T + F
\]

\[
Y = UQ^T + G
\]

where \( T \) and \( U \) are \( n \times r \) matrices of the extracted \( r \) components, the \( p \times r \) matrices \( P \) and \( Q \) represent matrices of projections, and the \( n \times p \) matrices \( F \) and \( G \) are the matrices of residuals. The PLS method, in which the classical form is based on the nonlinear iterative partial least square algorithm, finds projection axes \( w \) and \( c \) such that:

\[
\max(Xw^TYc)
\]

\[
s.t. \quad w^Tw = 1 \quad c^Tc = 1
\]

The solution to this optimization problem is given by the following Eigen value problem:

\[
X^TY^TXw = \xi w
\]

The \( X \)-latent-component vector \( t \) can be obtained as \( t = Xw \), which is also the first Eigen vector about the equation \( XX^TY^TL = \xi L \). The \( Y \)-latent-component vector \( u \) can be estimated as \( u = YY^TL \).

Furthermore, the above algorithm can be repeated \( l \) times with the deflation of \( X, Y \), defined as \( X_{l} \rightarrow X_{l+1}, Y_{l} \rightarrow Y_{l+1} \) then \( T_{l}, U_{l} \), the \( p \times l \) matrix \( W \) and the \( q \times l \) matrix \( C \) can be created, consisting of the columns from the vectors \( \{t_i\}_{i=1}^{p} \), \( \{u_i\}_{i=1}^{q} \), \( \{w_i\}_{i=1}^{p} \), \( \{c_i\}_{i=1}^{q} \), respectively.

The PLS form as regression model can be written in matrix as:

\[
\hat{Y} = XB_{pls}
\]

where coefficient matrix \( B_{pls} \) can be calculated as \( B_{pls} = W(\hat{P}^T W)^{-1}C \), and \( P \) is the \( p \times l \) matrix consisting of loadings vectors \( \{p_i = X^{(l)}(q_i)\}_{i=1}^{p} \).
3.2 Cross validation

Accuracy of PLS regression is verified by cross-validity:

\[ Q_h = 1 - \frac{S_{\text{press},h}}{S_{\text{ss},h-1}} \]  

(7)

where \( S_{\text{press},h} \) is the sum of squares of prediction errors after \( h \) components and \( S_{\text{ss},h-1} \) is sum of squares of errors after \( h-1 \) components. As if \( S_{\text{press},h} \) is smaller than \( S_{\text{ss},h-1} \) to certain extent, adding component \( h \) is considered to be useful for the improvement of regression accuracy, that is, if \( Q_h \leq 0.0975 \), the quality of model could be improved by adding a component. Otherwise, the number of component is enough. The cross-validity \( Q_h \) is the criterion for determining the number of components.

4 Partial least squares regression model

According to the correlation analysis above, the main influencing factors of main steam temperature are pressure of main steam \( x_1 \), flow of main steam \( x_2 \), oxygen content in flue gas \( x_3 \), total air flow \( x_4 \), inlet baffle opening of primary air fan \( x_5 \), drum water level \( x_6 \), front and rear differential pressure of main water gate \( x_7 \) and regulating valve opening of oil fuel atomizing steam \( x_8 \).

According to the partial least squares regression algorithm, the modeling process is as follows:

- Extracting the 1st principal component \( t_1 \)
  
  Combination coefficients (weights), vector:
  \[ w_1 = (0.4346, 0.4403, -0.3566, 0.3153, -0.0898, -0.4031, -0.4686, -0.0556) \]
  
  Regression coefficients, scalar:
  \[ r_1 = 0.3348 \]
  
  while \( Q_1 = 0.6124 > 0.0975 \), continue to calculate.

- Extracting the 2nd principal component \( t_2 \)
  
  Combination coefficients (weights), vector:
  \[ w_2 = (0.0869, 0.2067, 0.1948, -0.3340, 0.8322, -0.0483, -0.3108, 0.0933) \]
  
  Regression coefficients, scalar:
  \[ r_2 = 0.6749 \]
  
  while \( Q_2 = 0.8035 > 0.0975 \), continue to calculate.

- Extracting the 3rd principal component \( t_3 \)
  
  Combination coefficients (weights), vector:
  \[ w_3 = (0.1440, 0.2067, 0.1948, -0.3340, 0.8322, -0.0483, -0.3108, 0.0933) \]
  
  Regression coefficients, scalar:
  \[ r_3 = 0.1146 \]
  
  while \( Q_3 = 0.0854 < 0.0975 \), stop calculating.

Thus, when the extraction of 2 components, the model's predictive ability is best, and the standardized variable regression equation is:

\[ y^* = 0.5169 x_1^* + 0.1682 x_2^* - 0.4432 x_3^* - 0.0353 x_4^* \]  

(8)

For the Eq. 8, the original partial least squares regression equation is:

\[ y = 492.1938 + 0.8351 x_1 + 0.0139 x_2 - 0.1339 x_3 - 0.0588 x_4 + 0.7045 x_5 + 0.0151 x_6 - 3.8666 x_7 + 0.0426 x_8 \]

5 Conclusion

Here, the model of main steam temperature was built based on partial least-squares regression. The main influencing factors of main steam temperature were determined through correlation analysis. The prediction results showed that the model has high accuracy and can predict the main steam temperature in large thermal power plant.

Acknowledgement

This project Supported by the National Natural Science Foundation of China (51176028), the KEY Scientific, Technological Project of Jilin Province of China (20140204006SF) and Science and technology development plan of Jilin City (201464061)

References

1. Zhang X. T., Ni W. D., Li Z., Zheng S., Identifiability of building thermal system models

For the Eq. 8, the original partial least squares regression equation is:

\[ y = 492.1938 + 0.8351 x_1 + 0.0139 x_2 - 0.1339 x_3 - 0.0588 x_4 + 0.7045 x_5 + 0.0151 x_6 - 3.8666 x_7 + 0.0426 x_8 \]
using on-line data. J. Tsinghua Univ. (Sci & Tech), 44, (2004).
2. Jin T., Fu Z. G., Liu G., Yang Y. P., Application of reversed modeling method in power plant critical parameters modeling. J. Power Engineering, 29, (2009).
3. Deng L C, Wang G J, Chen H. On-line Fuzzy Identification of the Steam Temperature Object of Boiler, Proceedings of the CSEE, 26, (2006).
4. Zhao L., Ju G., Lv J. H., An improved genetic algorithm in multi-objective optimization and its application, Proceedings of the CSEE, 28, (2008).
5. Zhang X T, Ni W D, Li Z. Dynamic modeling study of super heater steam temperature based on principal component analysis method and online data. Proceedings of the CSEE, 25, (2005).
6. Luan X. C., Li S. Y., A study on multiple models predictive control for superheated steam temperature based on neural network models, Proceedings of the CSEE, 24, (2004)
7. Huang J T., Ma L. H., Mao J. B., Qian J. X., Modeling research of the reheat steam temperature of 300MW boiler based on support vector regression. Proceedings of the CSEE, 26, (2006).
8. Wen X Q. Analysis of carbon in coal-fired based on partial least squares regression algorithm, J. Northeast Dianli University, 32, (2012).