Data-Driven Thermal Efficiency Modeling and Optimization for Multi-fuel Boiler

Jian-Guo Wang, a Jia-Lin Kang, b Shyan-Shu Shieh, c Shi-Shang Jang, b David Shan-Hill Wong b

a School of Mechatronical Engineering and Automation, Shanghai University, Shanghai Key Lab of Power Station Automation Technology, Shanghai 200072, China
b Department of Chemical Engineering, National Tsing-Hua University, Hsin-Chu 30013, Taiwan
c Department of Occupational Safety and Health, Chang Jung Christian University, Tainan 71101, Taiwan

Abstract

A data-driven modeling and optimization scheme for thermal efficiency is developed based on the real operation data from a multi-fuel boiler. The plant input variables are transformed into the derived variables, which are physically and statistically meaningful for modeling. Then, this study proposed an adaptive modeling approach by employing nonnegative garrote variable selection and auto-regression integrated moving average correction. The high accuracy of the proposed modeling approaches makes the implementation of the model-based control of ratio of air to fuel for improving boiler’s efficiency readily practicable.

Keywords: Multi-fuel boiler; thermal efficiency; nonnegative garrote; variable selection; data-driven; statistics analysis

1. Introduction

BFG is the byproduct of manufacturing process in large-scale iron and steel plant, which is a typical low-heating-value gaseous fuel with low flame temperature and weak radiative heat-transfer capability. The supply of BFG is unstable in terms of flow rate and heating value in the batch-type production process. To maintain stable operation of the boiler, the flow rate of coal is adjusted to cope with varying supply of self-produced gases. In fact, the challenge to operate a co-firing boiler in an efficient way in the presence of gases fuel has attracted attention of many researchers from different fields for more than a decade (Hou et al., 2011; Shieh et al., 2011). Artificial neural networks and evolutionary computation can be used to handle nonlinearity, but these methods are operated in a black-box fashion and thus inevitably suffer from over fitting and vague interpretability. Though PCR/PLS/CVA can suppress the sensor variables into a few key latent variables (Jackson, 1991; Höskuldsson, 1988; Shi et al., 2000), they may still include contributions from a large number of predictor variables. This is highly undesirable because operators would fail to understand the physical meanings of the input-output relation and focus on a manageable number of key factors.

Some studies reported that the combustion model built by a selected subset of input variables make more accurate predictions of combustion efficiency than the whole set of variables in the coal-fired boiler. Recently, shrinkage methods, which conduct variable
selection by shrinking or setting some coefficients of a “greedy” model to zero, receive much attention. One of popular forms of these methods is the non-negative garrote (NNG) (Breiman, 1995). This study developed a data-driven modeling and optimization scheme utilizing NNG for thermal efficiency of BFG/coal co-firing boiler.

The rest of the paper is organized as follows. In the next section, system description is introduced. In Section 3, the formation of derived variables and justification of their significance are illustrated. In Section 4, the theoretical framework of an adaptive modeling is presented. The application of the proposed modeling approach is demonstrated in Section 5. A model-based control strategy is given and applied to the efficiency improvement in Section 6. Some remarks are concluded in the last section.

2. System Background description

The studied subject is one of twelve boilers and eight turbines with 410 MW overall rated capacity in a large iron and steel plant in Taiwan. The fluctuating heating value of BFG, as well as its flow rate poses challenges to operate BFG/coal co-firing boilers in an efficient way. The data used in this study were collected from the studied subject from April17 to May11, 2007. The operation data are taken every minute. After removing the abnormal samples, 15000 samples remained.

Among the most relevant input and output parameters of the BFG/coal co-firing boiler, the eight input parameters can be partitioned into three parts, i.e. (1) feed water related: flow rate \( F_{\text{fu}} \), temperature before the economizer \( T_{\text{fu,e}} \), and pressure \( P_{\text{fu}} \); (2) fuel related: coal flow rate \( F_{\text{coal}} \), BFG flow rate \( F_{\text{BFG}} \), and BFG heating value \( H_{\text{BFG}} \); (3) air related: air flow rate \( F_{\text{air}} \). The boiler output parameters include two parts, i.e. (1) main steam related: flow rate \( F_{\text{st}} \), temperature \( T_{\text{st}} \), and pressure \( P_{\text{st}} \); (2) flue gases related: oxygen concentration in flue gas \( O_{\text{FG}} \) and carbon monoxide concentration in flue gas \( C_{\text{FG}} \).

3. Formation and Incorporation of Derived Variables

3.1. Incorporation of three derived variables and LDA analysis

As mentioned in the previous section, the steam production in the boiler is regulated to serve the varying process demand. Among the boiler input parameters, four variables are loading directly related, namely, \( F_{\text{fu}} \), \( F_{\text{air}} \), \( F_{\text{coal}} \), and \( F_{\text{BFG}} \). The total input energy provided by coal and BFG, denoted as \( E_{\text{in}} \), is also related to boiler loading.

To address the influence of loading variation, two ratio parameters \( F_{\text{fu}}/E_{\text{in}} \) and \( F_{\text{air}}/E_{\text{in}} \) are derived. Moreover, it is obvious that the ratio of BFG providing energy to total fuel is related to boiler’s efficiency. Hence the ratio parameter \( E_{\text{BFG}}/E_{\text{in}} \) is incorporated. Replacing the four loading related variables, the three derived variables and the other four boiler input variables are considered to develop thermal efficiency model.

Linear discriminant analysis (LDA) is utilized to verify whether these three derived variables take important roles in determining efficiency levels. All the data are partitioned into five groups according to their efficiency levels. With descending order, five efficiency levels are denoted as HH, H, M, L, and LL. LDA analysis is conducted on three groups of data with levels of HH, M, and LL. The scattering of LDA projection for these efficiency levels and the weighting factor for all input variables composing projection y1 and y2 are shown in Fig.1 and Fig.2 respectively. The figures reveal that three groups are clearly differentiated by LDA projection and the three most significant variables are the derived variables.
3.2. Incorporation of O$_2$ and PCA Analysis

The oxygen concentration in flue gases $O_{FG}$ is long recognized as an index related to thermal efficiency. Correlation analysis are conducted to investigate relationship for efficiency to $F_{air}/E_{in}$ and $O_{FG}$. The results indicates poor linear relationship, with $14.5\%$ of $R^2$ between efficiency and $F_{air}/E_{in}$, and $2.6\%$ of $R^2$ between efficiency and $O_{FG}$. However, from the experience, we know that $O_{FG}$ could not be irrelevant to thermal efficiency. Principal component analysis (PCA) is further conducted to investigate the relationship of thermal efficiency to $F_{air}/E_{in}$ and $O_{FG}$. The results of correlation analysis and linear fitting of efficiency to 1st and 2nd PC are shown in Fig.3.

Note that the score of the 1st PC explains $77.05\%$ of variations in $F_{air}/E_{in}$ and $O_{FG}$, but it can only explains $1.5\%$ of variations in thermal efficiency. In contrast, the score of the 2nd PC only accounts for $22.95\%$ of total variations in $F_{air}/E_{in}$ and $O_{FG}$, but it can only explains $32.3\%$ of variations in thermal efficiency. The results indicate that the linear combination of $F_{air}/E_{in}$ and $O_{FG}$ taking opposite signs is strongly related to the thermal efficiency of the boiler. Hence, the variable $O_{FG}$ cannot be excluded in building thermal efficiency models and the efficiency model can be given as

$$\eta = f\left(\frac{F_{air}}{E_{in}}, \frac{F_{air}}{E_{in}}, T_{fus}, T_{fb}, P, H_{FG}, O_{FG}\right)$$

(1)

4. Adaptive Modeling Using NNG and ARIMA

4.1. NNG variable selection

The NNG method can be generalized into a two-stage shrinkage method. In the first stage the sign for each variable is determined using ordinary least square (OLS) procedure and in the second stage the corresponding magnitudes are computed by solving a series of constrained quadratic programming.

Suppose being given a set of observation data $\{X, y\}$, where $X \in \mathbb{R}^{n \times p}$ is the input matrix whose columns represent the measured candidate variables, and $y \in \mathbb{R}^n$ is the corresponding vector of response data. Suppose that $X$ and $y$ have been normalized to zero-mean and unitary standard deviation. Let $\hat{\beta} \in \mathbb{R}^{p \times 1}$ be a set of the ordinary least square estimates of the coefficients of the following linear model

$$y = X\hat{\beta} + \epsilon$$

(2)

The second stage shrinkage can be formulated as the following optimization problem:

$$J = \min_{\beta} \sum_{i=1}^{n} (y_i - X_i \beta)^2 \quad \text{s.t.} \quad \sum_{i=1}^{p} \beta_i \leq s \quad (3)$$

As $s$ is decreased and the non-negative garrote is tightened and thus more of the coefficients become zero and the remaining nonzero coefficients are shrunken. A solution path exists with $0 \leq s \leq p$ on which the appropriate shrinkage can be selected. The optimal selection $s^*$ can be determined with v-fold cross-validation (Zhang, 1993).

The overall coefficients can be determined by solving the optimization problem:

$$c^* = \arg\min_{c} \sum_{i=1}^{n} (y_i - X_i c)^2 \quad \text{s.t.} \quad 0 \leq \sum_{j=1}^{p} c_j \leq s \quad (4)$$

NNG regression coefficient can be given as

$$\hat{\beta}_{\text{NNG}} = c^* \hat{\beta}$$

(5)

4.2. ARIMA model correction

Given a set of measured input data $X_k$, the prediction of $y_k$ can be obtained

$$\hat{y}_k = X_k \hat{\beta}_{\text{NNG}}$$

(6)

Although the NNG prediction accuracy has been enough for ordinary usage, it is not satisfactory for efficiency improvement. When conducting time series correlation analysis for the NNG prediction error of efficiency and $O_{FG}$ denoted as $e_k = y_k - \hat{y}_k$. 

$$\alpha_k = y_k - \hat{y}_k$$
it is discovered that there is high correlation between \( \alpha_k \) and \( \alpha_{k-1} \) and the time series of prediction error is non-stationary. Hence, ARIMA(1,1,1) model is adopted to improve the prediction accuracy as follows
\[
y_k = \hat{y}_k + a_k = \frac{1-\theta B}{(1-\alpha B)(1-B)} \epsilon_k
\]
where \( B \) is back shift operator, \( \epsilon_k \sim N(0, \delta^2) \) is a zero mean white noise with variance \( \delta^2 \). \( \theta \) and \( \varphi \) are two coefficients to be identified. The following two equations are obtained:
\[
\hat{\epsilon}_k = y_k - \hat{y}_k - (1 + \varphi) \hat{a}_{k-1} + \varphi \hat{a}_{k-2} + \theta \epsilon_{k-1}
\]
\[
\epsilon_k = y_k - \hat{y}_k - (1 + \varphi) \hat{a}_{k-1} + \varphi \hat{a}_{k-2} + \theta \epsilon_{k-1}
\]
Given \( y \) and corresponding NNG prediction \( \hat{y} \) with length \( n \), the coefficients \( \theta \) and \( \varphi \) can be estimated by minimizing the objective function
\[
\min_{\theta, \varphi} \sum_{k=1}^{n} \epsilon_k^2
\]
After the ARIMA(1,1,1) model is obtained, the new prediction can be given as
\[
y_k = \hat{y}_k + (1 + \varphi) a_{k-1} - \varphi a_{k-2} - \theta \epsilon_{k-1}
\]
Adaptive strategy is often used to deal with time-varying characteristics of industrial processes. In this paper, the NNG regression and ARIMA correction methods is implemented in moving window manner. It should be noted that adaptive modeling and prediction should be conducted for both efficiency and \( O_{\text{FG}} \) to implement efficiency control.

5. Efficiency and Modeling and Prediction

5.1. Efficiency modeling and prediction
Setting the length of observation window as 5000 and the length of prediction horizon 500, hence there are 20 observation windows for modeling and 20 prediction horizons. The adaptive modeling method with NNG and ARIMA was applied to thermal efficiency modeling and prediction. Fig. 4 shows the NNG regression coefficients of the eight variables in modeling for each of 20 observation windows. The average \( R^2 \) in 20 NNG regressions is 92.4\% and the root-mean-square of prediction error (RMSE) is 0.56\%. However, the model is considered to be too rough to be incorporated to the model-based control in the case of boiler operation. To further improve modeling reliability, ARIMA(1,1,1) model was adopted. Fig. 5 shows the prediction results for interval [5000, 15000], a large scale, and [10100, 10150], a close look. The RMSE of the prediction error throughout the whole prediction horizon is 0.026\%.

5.2. Modeling and prediction
As given in equation (1), \( O_{\text{FG}} \) is treated as one of model inputs in building the thermal efficiency model. \( O_{\text{FG}} \) model is constructed as follows
\[
O_{\text{FG}} = g_{\text{FG}} \frac{F_{\text{air}}}{E_{\text{in}}} \frac{F_{\text{fuel}}}{E_{\text{in}}} \frac{E_{\text{net}}}{E_{\text{in}}} \frac{T_{\text{fuel}}}{T_{\text{in}}} \frac{T_{\text{hot air}}}{T_{\text{in}}} \frac{T_{\text{hot air}}}{T_{\text{in}}} \frac{P_{\text{fuel}}}{P_{\text{in}}} \frac{H_{\text{net}}}{H_{\text{in}}} \frac{H_{\text{net}}}{H_{\text{in}}}(12)
\]
The same adaptive modeling method was applied to \( O_{\text{FG}} \). It is found that all the coefficients of \( F_{\text{air}}/E_{\text{in}} \) have positive sign, which accord with the fact that the increase of air supply will result in the higher \( O_{\text{FG}} \). Note that all the signs of regression coefficients of variable \( F_{\text{air}}/E_{\text{in}} \) are positive in Fig. 4. The positive coefficients mean that we can increase \( F_{\text{air}}/E_{\text{in}} \) to improve boiler efficiency. In contrast, all the signs of regression coefficients of variable \( O_{\text{FG}} \) are negative, which mean that when \( O_{\text{FG}} \) being increased, the boiler efficiency will degrade. Hence, the effect of boiler efficiency improvement from the increase of \( F_{\text{air}}/E_{\text{in}} \) is constrained, since the increase of \( F_{\text{air}}/E_{\text{in}} \) results in the higher \( O_{\text{FG}} \), which may lead to the efficiency degradation as the result of
heat loss in flue gases. It can be seen that the sign of regression coefficients has explicit physical meaning and is in accord with the actual operation experience.

Fig. 1. Scattering of LDA projection

Fig. 2. Weighting factor for input variables

Fig. 3. Correlation analysis and linear fitting

Fig. 4. NNG regression coefficients

Fig. 5. Prediction with NNG and ARIMA

Fig. 6. Efficiency improvement results

6. Model-based Efficiency Control
Taking a conservative approach, the corrective actions taken for both $F_{air}/E_{in}$ and $O_{FG}$ are limited by the upper and lower bound, which are set by the historical maximum and minimum. The overall efficiency control strategy can be formulated as:

$$\max_{E_{in}, F_{air}, E_{air}} \left( \frac{F_{air}}{E_{in}}, \frac{E_{air}}{E_{in}}, \frac{E_{air}}{E_{in}}, \frac{T_{in}}{T_{air}}, \frac{F_{air}}{F_{air}}, \frac{F_{air}}{F_{air}}, \frac{H_{FG}}{H_{FG}}, \frac{O_{FG}}{O_{FG}} \right)$$

s. t.(13)
Apply the efficiency control strategy to determine the optimal \( \frac{F_{\text{air}}}{E_{\text{in}}} \). In the same moving window manner as in modeling and prediction, the efficiency in interval [5001 15000] could be controlled based on the obtained model. Fig. 6 shows the results after implementing the model-based control for the thermal efficiency. The original operation data of \( \frac{E_{\text{BFG}}}{E_{\text{in}}}, \frac{F_{\text{air}}}{E_{\text{in}}} \) and \( O_{\text{FC}} \) are indicated in blue color, while the adjusted \( \frac{F_{\text{air}}}{E_{\text{in}}} \), \( O_{\text{FC}} \) and the improvement in thermal efficiency in the red color.

The actual operation curves shown revealed that the more BFG share was in fuel, the less is needed. The trend is in accord with the combustion theory that coal as the solid fuel needs more amount of air to achieve complete combustion than BFG in term of unit heating value. Comparing the data of \( \frac{F_{\text{air}}}{E_{\text{in}}} \) between the real operation and the proposed adjustment, the actions taken by the operator were correct, but were not accurate enough to reach optimal air supply. The results revealed that higher \( \frac{F_{\text{air}}}{E_{\text{in}}} \) throughout the interval is needed to improve thermal efficiency in our studied case, especially in the case of high BFG input.

The increase of \( \frac{F_{\text{air}}}{E_{\text{in}}} \) on average is 8.73\%, and the improvement of thermal efficiency averages at 0.32\%. The benefit/cost ratio is significant enough when considering almost no fixed cost is needed in the implementation of model-based control in this case for any real boiler operation.

7. Conclusion

A data-driven scheme of thermal efficiency modeling and optimization for a BFG/coal co-firing boiler was investigated based on the real operation data. The derived variables are found more physical meaningful than some plain variables in constructing the model of thermal efficiency in the case of the co-firing boiler. The models of thermal efficiency and oxygen concentration in the flue gas were built by the proposed adaptive algorithm based on NNG variable selection and ARIMA correction. The implementation of model-based control was virtually studied based on the real operation data of a BFG/coal co-firing boiler. The arriving of the high accuracy prediction models would convince the operator to shift the present practice to model-based control and explore the room for energy conservation with high benefit/cost ratio.

References