Coal Calorific Value Prediction Based on Projection Pursuit Principle

JING Yuan*, FU Zhongguang, QI Minfang
North China Electric Power University, Beijing, China
No.2, Beinong Road, Changping District, Beijing 102206
*corresponding author, e-mail: wuque211@163.com

Abstract
The calorific value of coal is an important factor for the economic operation of coal-fired power plant. However, calorific value is tremendous difference between the different coal, and even if coal is from the same mine. Restricted by the coal market, most of coal fired power plants can not burn the designed-coal by now in China. The properties of coal as received are changing so frequently that pulverized coal firing is always with the unexpected condition. Therefore, the researches on the prediction of calorific value of coal have a profound significance for the economic operation of power plants. Aiming at the problem of uncertainty of coal calorific value, establish a soft measurement model for calorific value of coal based on projection pursuit principle combined with genetic algorithm to optimize parameters, and support vector machine algorithm. It is shown by an example that the model has a stronger objectivity, effective and feasible for avoiding the disadvantage of the artificially decided weights of feature indexes. The model could provide a good guidance for the calculation of the coal calorific value and optimization operation of coal-fired power plants.

Keywords: calorific value of coal, prediction, projection pursuit principle, genetic algorithm, support vector machine

Copyright © 2012 Universitas Ahmad Dahlan. All rights reserved.

1. Introduction
According to the distribution of coal resources in China, the way of coal-fired power will play a main role in electricity generation at present and in the future. For the coal-fired power plant, the coal quality indexes will affect its economic operation directly. The calorific value of coal is the most important index for economic operation of the power plant. First of all, in the process of combustion and gasification, it could calculate the thermal equilibrium, coal consumption and the thermal efficiency of plant according to the calorific value of coal, and then consider trying to achieve the maximum heat efficiency by improving the operational approach and the technical process; secondly, for the boiler, the calorific value of coal determines the excess air coefficient, waste gas density and the required theoretical combustion temperature, etc. in the process of burning [1]; finally, in the work of boiler's design calculation, it also needs to know the calorific value of coal to determine a series of parameters, such as the type of boiler, combustion method and the material balance in the during of burning and so on.

Before then, there were some empirical formulas between industrial composition and the calorific value of coal [2-4] in the published literature. But the equations could only apply to a certain mine, so they were lack of generality. Fan Li-jun [5], Huang Ping [6] et al. also did some researches on the coal calorific value by establishing equations of bivariate linear regression and ternary linear regression. But the formulas fitted out by the regression analysis could only apply to the specific coal seam, and the prediction model always need to revise, so the application scope of regression equations still has a lot of limitations. Therefore, establish a concise and accurate soft measurement model for calorific value of coal has an important significance to provide a good guidance for the economical efficiency evaluation and optimization operation of coal-fired power plants.

In view that projection pursuit principle has the characteristic of overcoming the noise of variables with little relationship of structures and features of data, so it could realize the regression analysis of multi-index samples and not depend on the determinants decided artificially at the same time. Therefore, it could predict the calorific value of coal by establishing
the soft measurement model for coal calorific value based on projection pursuit principle combined with genetic algorithm to optimize parameters. It provides a new method of calculating the coal calorific value of coal-fired power plants.

2. The Proposed Method: Projection Pursuit Principle

Projection pursuit principle is a new statistical method presented by American scientist Kruskal [7], which could analyse and deal with the high-dimensional observation data, especially the high-dimensional data of non-linear or non-normal. It is an interdiscipline subject of statistics, applied mathematics and computer technology, which belongs with a frontier field at present. Projection pursuit method has the characteristics of high robustness, strong antijamming capability and high degree of accuracy. The method could transform the high-dimensional data to the low-dimensional subspace, and find out the projection which could reflect the structures or features of original high-dimensional data so as to do the research and analysis of high-dimensional data.

The essence of projection pursuit method is to construct a quantitative projection-indexes function \( Q(a) \), and optimize it by numerical method in order to get the optimal projection-direction \( a^* \). Use the optimal projection-direction \( a^* \) to do the dimensionality reduction of multi-index data to the one-dimensional projection-value \( z \). The projection and dimensionality reduction of the high-dimensional data could realize the analysis of required questions. The specific steps are as follows:

1. Normalization of sample sets. Suppose the sample index sets are \( \{ x^*(i,j) | i=1, 2, ..., n; j=1, 2, ..., p \} \), among them, \( x^*(i,j) \) represents the index-\( j \) of sample-\( i \), and \( n, p \) represent the number of samples (i.e. the capacity of samples) and the number of indexes respectively. In order to eliminate the dimension and unify the variation range of each index-value, do some operations of normalization with the following formulas:

   For the index-- the bigger the better:
   \[
   x(i, j) = \left[ \frac{x^*(i, j) - x_{\text{min}}(j)}{x_{\text{max}}(j) - x_{\text{min}}(j)} \right] 
   \]

   (1)

   For the index-- the smaller the better:
   \[
   x(i, j) = \left[ \frac{x_{\text{max}}(j) - x^*(i, j)}{x_{\text{max}}(j) - x_{\text{min}}(j)} \right] 
   \]

   (2)

   Among them, \( x_{\text{max}}(j) \), \( x_{\text{min}}(j) \) represent the maximum and minimum of the index-\( j \) respectively, \( x(i,j) \) represents the normalized data.

2. Construct projection-direction \( a \). Construct a \( p \)-dimensional unit vector \( a = \{ a_1, a_2, ..., a_p \} \), and transform the sample data to the one-dimensional projection-value \( z(i) \) with \( a \):

   \[
   z(i) = \sum_{j=1}^{p} a(j) \cdot x(i, j) 
   \]

   (3)

   Among them, \( i=1, 2, ..., n \).

3. Construct projection-indexes function \( Q(a) \). Aimed at the specific question, construct projection-indexes function \( Q(a) \) of high-dimensional data.

4. Optimize the projection-indexes function \( Q(a) \) and get the optimal projection-direction \( a^* \). Different projection-directions reflect different data structures and data features, and the optimal projection-direction is the direction which could reflect the characteristic structure of high-dimensional data as largely as possible. Therefore, solving the maximization problem of the projection-indexes function could get the optimal projection-direction \( a^* \):

   \[
   \text{Max}: Q(a) 
   \]

   Constraint condition:

   \[
   \sum_{j=1}^{p} a(j) = 1 
   \]

(4)
3. Modeling based on projection pursuit principle for coal calorific value

3.1. Feature variables selection of model and data preprocessing

Previous researches have shown that the calorific value of coal is closely related to many parameters by the elemental analysis and proximate analysis of coal. For example, the higher the values of coal moisture and ash content are, the lower the calorific value of coal is [4]. Therefore, the selection of the feature variables is based on the parameters by the proximate analysis of coal. Among them, the sample index variables based on projection pursuit method are these four groups of variables, including the moisture in the whole (Mt), the moisture as air dry basis (Mad), the ash content as received basis (Aar), and the volatile matter as dry ash-free basis (Vdaf), and then do the prediction research for the calorific value of coal with them.

The data this paper used comes from a coal analysis report of the operation department chemical profession fuel lab of a 300000 kW unit. Since the data are detected by standard instrumental, they are all useable. Among them, 20 groups are used as the model training data, and the remaining 5 groups as the model test data. The normalized sample index sets data of model are shown in Table 1.

<table>
<thead>
<tr>
<th>Point-in-time</th>
<th>Normalized sample index sets</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mt*</td>
</tr>
<tr>
<td>1</td>
<td>0.7010</td>
</tr>
<tr>
<td>2</td>
<td>0.3164</td>
</tr>
<tr>
<td>3</td>
<td>0.4790</td>
</tr>
<tr>
<td>4</td>
<td>0.4311</td>
</tr>
<tr>
<td>5</td>
<td>0.7040</td>
</tr>
<tr>
<td>6</td>
<td>0.5617</td>
</tr>
<tr>
<td>7</td>
<td>0.3861</td>
</tr>
<tr>
<td>8</td>
<td>0.5893</td>
</tr>
<tr>
<td>9</td>
<td>0.6531</td>
</tr>
<tr>
<td>10</td>
<td>0.5152</td>
</tr>
<tr>
<td>11</td>
<td>0.8970</td>
</tr>
<tr>
<td>12</td>
<td>0.2612</td>
</tr>
<tr>
<td>13</td>
<td>0.7446</td>
</tr>
<tr>
<td>14</td>
<td>0.4064</td>
</tr>
<tr>
<td>15</td>
<td>0.7968</td>
</tr>
<tr>
<td>16</td>
<td>0.6720</td>
</tr>
<tr>
<td>17</td>
<td>0.3672</td>
</tr>
<tr>
<td>18</td>
<td>0.5733</td>
</tr>
<tr>
<td>19</td>
<td>0.3237</td>
</tr>
<tr>
<td>20</td>
<td>0</td>
</tr>
</tbody>
</table>

Among them, Mt* presents the normalized moisture in the whole, Mad* presents the normalized moisture as air dry basis, Aar* presents the normalized ash content as received basis, Vdaf* presents the normalized volatile matter as dry ash-free basis.

3.2. Optimizing to get the optimal projection-direction a*

3.2.1 Construct projection-indexes function Q(a)

In order to concentrate the local projection-points to be some point-groups, and at the same time, all point-groups are separated as possible as they could, construct projection-indexes function Q(a) as follows:

\[ Q(a) = S_z \cdot D_z \]  

Among them, \( S_z \) is the standard deviation of projection-values \( z(i) \), \( D_z \) is the local density of projection-values \( z(i) \):

Coal Calorific Value Prediction Based on Projection Pursuit Principle (JING Yuan)
\[ S(z) = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} [z(i) - E(z)]^2} \]

\[ D_z = \sum_{i=1}^{n} \sum_{j=1}^{n} [R - r(i, j)] \cdot u[R - r(i, j)] \]

Among them, \( E(z) \) is average value of distance between the sample \( i \) and \( j \), \( r(i, j) \) is the distance between the sample \( i \) and \( j \), \( u(t) \) is a unit step function: when \( t \geq 0 \), \( u(t) = 1 \); and when \( t < 0 \), \( u(t) = 0 \).

### 3.2.2 Determination of window radius of local density \( R \)

In the step of constructing projection-indexes function with the above-mentioned method, the only concerned parameter is the window radius of local density \( R \). The selection of \( R \) need to consider that the number of samples in the window should not be too small lest the moving average deviation of samples is too large; at the same time, \( R \) should not increase too much with the increase of the sample size. Therefore, the reasonable range of \( R \) is:

\[ r_{\text{max}} + p/2 \leq R \leq 2p, \quad \text{and} \quad r_{\text{max}} = \max(r(i, j)), \quad -p \leq z(i) \leq p. \]

After calculations, get the values: \( r_{\text{max}} = 0.969, \; p = 1.054 \). Therefore, \( 1.496 \leq R \leq 2.108 \), and determine the value of \( R \) is 2.

### 3.2.3 Optimizing search of the optimal projection-direction \( a^* \)

The essence of optimizing the projection-indexes function \( Q(a) \) and getting the optimal projection-direction \( a^* \) is a complex nonlinearity optimization problem based on the optimizing variables \( \{a(j) \mid j=1,2,\ldots,p\} \). Therefore, solve the high-dimensional optimizing problem combined with genetic algorithm, which has perfect global optimization ability.

Define the projection-indexes function \( Q(a) \) as the fitness function. After 476 generations of optimization search, \( Q(a) \) get the maximum. And then, get the optimal projection-direction \( a^* \), \( a^*=(2.2195e^{-6}, \; 9.0170e^{-7}, \; 0.9622, \; 0.2725) \). The trace of the best value by genetic algorithm is shown in Figure 1.

\[ \text{Figure 1. Curve: trace of the best value by genetic algorithm} \]

### 3.3. Modeling the coal calorific value soft measurement model

Calculate the projection-values \( z^*(i) \) of each sample combined with the optimal projection-direction \( a^* \) and Eq. (3), \( z^*(i)=(1.2012, \; 1.0676, \; 1.0623, \; 0.7384, \; 0.8670, \; 0.8959, ...) \).
Coal Calorific Value Prediction Based on Projection Pursuit Principle (JING Yuan)

0.7507, 0.2419, 0.6774, 0.5663, 0.1152, 1.0480, 0.5180, 0.6321, 0.5597, 0.5754, 0.4697, 0.5336, 0.3852, 0.5985). Considering that the algorithm of Support Vector Machine (SVM) has good regression prediction ability of small sample, so modeling the SVM model for coal calorific value. Among them, input variable is the projection-values $z^*(i)$, and output variable is the net calorific value of coal ($Q_{net, ar}$).

4. Results and Discussion

Test the feasibility of model with the reserved 5 groups of sample data as the test data. The verification results based on projection pursuit principle and the verification results based on BP neural network [8], which is the most popular machine learning method are shown in Table 2.

<table>
<thead>
<tr>
<th>Point-in-time</th>
<th>Q_{net, ar} (MJ/Kg)</th>
<th>Model by projection pursuit method</th>
<th>Test result (MJ/Kg)</th>
<th>Relative error (%)</th>
<th>Model by BP neural network method</th>
<th>Test result (MJ/Kg)</th>
<th>Relative error (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>17.50</td>
<td>17.1997</td>
<td>17.8735</td>
<td>-2.1343</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>17.28</td>
<td>17.2486</td>
<td>18.3324</td>
<td>-6.0903</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>17.50</td>
<td>17.4239</td>
<td>16.9304</td>
<td>3.2549</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>17.34</td>
<td>17.3265</td>
<td>16.0231</td>
<td>7.5946</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>16.48</td>
<td>16.8502</td>
<td>17.2247</td>
<td>-4.5188</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

As shown in Table 2, the test results and the true values of the 5 groups of test samples data show that the coal calorific value model based on projection pursuit principle has high accuracy. Therefore, the model is effective and feasible.

5. Conclusion

The projection pursuit method is applied and combined with genetic algorithm to get the optimal projection-direction of prediction for coal calorific value. And then, construct the SVM model for coal calorific value based on projection pursuit principle. Projection pursuit method has the characteristic of overcoming the noise of variables which have nothing to do with structures and features of data, so it could avoid the disadvantage of the artificially decided weights of feature indexes. The model has a higher objectivity.

It is shown by an example of test results that the model has high accuracy. Therefore, the model is effective and feasible. It provides a new method of calculating the coal calorific value, and a good guidance for optimization operation of coal-fired power plants.

Acknowledgements

Supported by "Natural Science Foundation of China (51036002, 50776029)"

References
