Application of Improved Local Models of Large Scale Database-based Online Modeling to Prediction of Molten Iron Temperature of Blast Furnace

Norio KANEKO,1) Shinroku MATSUZAKI,2) Masahiro ITO,2) Haruhisa OOGAI3) and Kenko UCHIDA1)

1) Department of Electrical Engineering and Bioscience, School of Advanced Science and Engineering, Waseda University, 3-4-1 Okubo, Shinjyuku-ku, Tokyo 169-8555 Japan. 2) Environment & Process Technology Center, Technical Development Bureau Nippon Steel Corp., 20-1 Shintomi, Futtsu, Chiba 293-8511 Japan. 3) The Graduate School of Information, Production and Systems, Waseda University, 2-7 Hibikino, Wakamatsu-ku, Kitakyushu, Fukuoka 808-0135 Japan.

(Received on October 29, 2009; accepted on December 24, 2009)

The large scale database-based online modeling called LOM is the one of local modeling method. This method has been developed to apply the just-in-time modeling for the blast furnace by us. In this paper, we propose two new types of local models in LOM to improve the prediction performance. One is used weighted multiple regression model as a linear local model of LOM. The other is used on-line Bayesian learning model as a nonlinear local model of LOM. In order to compare the prediction performance of the two types of local models in LOM, we evaluate the prediction performance by using the real process data of the blast furnace.

KEY WORDS: blast furnace; just-in-time modeling; prediction; process control; weighted multiple regression; mutual information; on-line Bayesian learning; Sequential Monte Carlo.

1. Introduction

The predicting model of the blast furnace operation was variously proposed in the past, and includes the model by a theoretical approach based on the mathematical model of the mass transport,1–3) the reaction, and the thermal transfer based on the one and data. The model based on data includes the AR model.4,5) Moreover, the computer also developed, and the expert model by which the artificial intelligence was used was developed aggressively in the 1980' s.6–9) In addition, the predictive model by the neural network was developed as another use of the artificial intelligence.10,11)

Each model has strengths and limitations. The theoretical model has the limit in the theory though the interpretation of the prediction result is comprehensible. The expert model can pass the hardness of making to the theory. However, making to the rule that is the critical factor for the model is difficult. The maintenance when the operation assumption changes greatly will be difficult.

The model introduces in this paper draws a past predictive model and a line without using all data in making the model, extracting only the data that exists in a similar situation, and making the local model by the data. The neural network has been used for one of the local models. However, such a nonlinear approach became possible by the development of a recent computer.

In case of considering prediction and control for complicated systems, local modeling method is one of efficient techniques. In recent years, “Just-In-Time (JIT) modeling”12) or “Lazy Learning”,13) one of local modeling methods, attract a great deal of attention. That is based on the background that technology advances of computer and database system have made it possible to store large scale process data and search data quickly.

The basic idea of JIT modeling is as follows. When the predictive value is required, we collect the neighboring data that is similar to current information at first. Then a local model is built by using the neighboring data. At last, a predictive value is calculated by assigning the current information to the local model.

We developed “Large scale database-based Online Modeling (LOM)” as a JIT modeling for blast furnace.13,14) LOM was developed mainly to apply JIT modeling to the real data sets of blast furnace online and to improve the prediction performance. The prediction performance of LOM has been improved by using Independent Component Analysis.15) However there is a lot of room to improve the prediction performance. In order to improve the prediction performance of LOM, it is important to consider how to select useful variables for prediction and how to search neighboring data and how to select appropriate local models.

In this paper, we focus on improvement of the local model especially and propose two new types of local models in LOM. One of the proposed local models is weighted multiple regression model and the other is on-line Bayesian learning model. The former model is a linear model so that it is easy to calculate the regression coefficients by using least squares method. This method incorporates the weight
that is influenced by a distance between the current information and the neighboring data. The main advantage of this method is to be able to keep the local model simple. The latter model uses three layer perceptron as a nonlinear model. Therefore we need to pay attention to the dimensionality of input variables and the computation method for the estimation of nonlinear parameters. In order to alleviate the curse of dimensionality, we use mutual information for input variable selection. To estimate the parameters of nonlinear model, we use on-line Bayesian learning scheme with Sequential Monte Carlo. The main advantage of this method is to be able to build the detailed local model.

To compare the prediction performance of the two types of local models in LOM, we evaluate the performance by performing prediction tests based on real process data of blast furnace. This paper is organized as follows. In Sec. 2, we present the improved LOM in detail. In Sec. 3, we apply the proposed LOM to the real data sets of blast furnace and compare the prediction performance of the two types of local models in LOM. Finally, some concluding remarks are made in Sec. 4.

2. Large Scale Database-based Online Modeling

LOM is developed for prediction and control by adapting JIT modeling online to the multidimensional topological space composed of a real process data. LOM consists of the following component steps (Fig. 1) [6]:

(1) Reducing the dimension of topological space by using stepwise method.

(2) Searching the neighboring data on quantized topological space.

(3) Building the local model by using the neighboring data and predicting the system output.

(4) Discarding the model and updating the database.

We consider the blast furnace as a nonlinear dynamic system and assume its dynamic behaviors are given a regression model as follows:

\[ y(t+p) = f(y(t), y(t-1), \ldots, y(t-n_y)) \]
\[ u(t-d), u(t-d-1), \ldots, u(t-d-n_u) \] ........................(1)

where,

- \( u(t) \) is a control input vector of the system at time \( t \),
- \( y(t) \) is an observational output vector of the system at time \( t \),
- \( n_u \) is an order of the control input vector,
- \( n_y \) is an order of the observational output vector,

\( p \) is a time of prediction,
\( d \) is a delay time,
\( f \) is an unknown nonlinear function.

Now the system input vector \( x^i \) and the system output vector \( y^i \) are redefined as the following Eqs. (2) and (3).

\[ x^i = (y(t), y(t-1), \ldots, y(t-n_y)) \]
\[ u(t-d), u(t-d-1), \ldots, u(t-d-n_u)) \] ........................(2)
\[ y^i = y(k+p) \] ..............................(3)

As time progresses, a large number of data, which is composed of the system input vector \( x^i \) and the system output vector \( y^i \), have been stored from the system as a data set \{\( x^i, y^i \)\} \( i = 1, 2, \ldots, n \), where \( k \) is a discrete time.

When we need to predict a system output vector \( y_{ki} \), the current system state \( x_{ki} \) defined as a query is used to search the neighboring data that is similar to the query. Afterwards, the local model is built by using the neighboring data and the system output vector \( y_{ki} \) is calculated by assigning the query to the local model. The details of the component steps of LOM are described as follows.

2.1. Reducing the Number of Topological Space by Using Stepwise Method

For practical effect to estimate with a regression model, stepwise method is a technique that increases or decreases input variables based on an examination standard to aim at lessening a residual sum of squares between the observational value and the predicted one.

For blast furnace, a large number of multidimensional spaces have to be treated because the number of observational data item is extremely huge and its topological variables have to be also considered. So that LOM is studied as a practical modeling method with using stepwise method under the severe examination standard for reducing the number of multidimensional spaces.

The variables selected by stepwise method are used when we need to search the neighboring data.

2.2. Quantizing Topological Space and Searching the Neighboring Data

For the purpose of realizing to search a large number and scale of online process data quickly with efficiently searching and drastically reducing the computational load, LOM stores observational process data in the beforehand database for searching, and searches the neighboring data close to the query in the quantized topological space of the database quantum by quantum. At first, a quantized space \( X^i \) is defined by quantizing the topological space that a vector variable \( x^i \) to which belongs.

\[ X^i = Z(x^i) \] \( i = 1, 2, \ldots, n \) ..............................(4)

Where, \( Z(\cdot) \) is the quantizing operator, \( n \) is the number of the data that belongs to the same quantized space \( X^i \). In the case that a vector variable \( x^i \) has one dimension, a quantized space \( X^i \) is an interval, and in the case that a vector variable \( x^i \) has two dimensions, a quantized space \( X^i \) is a rectangle, and generally a quantized space \( X^i \) is multidimensional prism. Secondarily, a similarity \( s(k, k_j) \) is defined between the quantized space \( X^i \) and \( X^{k_j} \). For example, a similarity is referred as an infinity norm of the quantized space's reciprocal relation.
2.3. Local Model

When the query variable vector \( x^k \) is given, the local model is built using the neighboring data. The system output \( \hat{y}^k \) is calculated by assigning the query variable vector \( x^k \) to the local model. Thus, it is important to build the local model appropriately.

In this paper, we propose two types of local models in LOM. One is weighted multiple regression model and the other is on-line Bayesian learning model. The former model is a linear model so that it is easy to calculate the regression coefficients by using least squares method. The main advantage of this method is to be able to keep the local model simple. The latter model uses three layer perceptron as a nonlinear model. Thus, it may be possible to build the detailed local model. However, we need to pay attention to the dimensionality of input variables and the computation method for the nonlinear parameters. In order to alleviate the curse of dimensionality, we use mutual information for input variable selection. To estimate the parameters of nonlinear model, we use on-line Bayesian learning scheme with Sequential Monte Carlo.

The details of two types of local models are described as follows.

2.3.1. Weighted Multiple Regression Model

One of effective local models for blast furnace operation is the weighted multiple regression model. The multiple regression model assumes that, given some neighboring data \((x_n, y_n), n=1, 2, \ldots, N\), a possibly imperfect relationship between \( x_n \) and \( y_n \). A disturbance term \( \epsilon_n \), which is a random variable too, is added to this assumed relationship to capture the influence of everything else on \( y_n \) other than \( x_n \). Therefore, the multiple regression model takes the following form:

\[
y_n = x_n \beta + \epsilon_n \quad \text{...............(7)}
\]

Where \( \beta \) stands for a regression coefficient vector whose dimension number is as same as the dimension number of \( x_n \).

A weight \( w_n \) for each neighboring data is calculated by using Gaussian function as the following equation:

\[
w_n = \exp(-d_n^2) \quad \text{...............(8)}
\]

Where, \( d_n \) represent a distance between the query variable vector \( x^k \) and the input variable vector of the neighboring data \( x_n \). The distance is calculated by using Euclidian norm and normalized.

The regression coefficient vector \( \beta \) of multiple regression model \( L(x)=x \beta \) are calculated through minimizing the following equation:

\[
\sum_{n=1}^{N} (w_n (y_n - (x_n \beta)))^2 \quad \text{...............(9)}
\]

After the regression coefficient vector is determined, the system output \( \hat{y}^k \) are calculated by \( L(x_\hat{k}) \).

2.3.2. On-line Bayesian Learning Model

The other is on-line Bayesian learning model. In this paper, the nonlinear local model is taken to be a three-layer perceptron, so that we need to reduce the dimensionality of input variable vector as a first step in order to alleviate the curse of dimensionality. To reduce the dimensionality of input variable vector, we use the mutual information for input variable selection.

Let \( x \) and \( y \) be two random variables. The mutual information measures the information content in input variable vector with respect to the output, without making any assumption on the model. The mutual information is calculated by the following equation:

\[
I[x, y] = - \int \int p(x, y) \ln \frac{p(x)p(y)}{p(x,y)} \, dx \, dy \quad \text{...............(10)}
\]

To estimate the mutual information, we use an estimator of the mutual information recently proposed in the literature. We search optimal subset from the variables that is used in case of searching neighboring data.

Next we introduce the on-line Bayesian learning. At first we consider a nonlinear local model with input \( x \), output \( y \) and parameter \( w=(w^1, L, w^0) \), which is described by

\[
y = (x; w) \quad \text{...............(11)}
\]

and assume that, at each step \( n(n=1, 2, L, N) \), an input–output data \( D_n=(x_n, y_n) \) is given. We intend to estimate sequentially the parameter \( w \) at each step \( n \) (denoted \( w_n \)) based on the data \( D_n=(D_1, L, D_N) \) and predict the output when the query variable vector \( x^k \) is given. \( w_n \) and \( w_n \) \((\neq x)\) are assumed statistically mutually independent, and \( x_n \) and \( w_n \) are also assumed statistically mutually independent.

For the data \( D_n \) and the parameter \( w_n \), we assume the following probability densities:

\[
D_n; p(D_n | w_n) \quad \text{...............(12)}
\]

\[
w_n; p(w_n | w_{n-1}) \quad \text{...............(13)}
\]

These densities contain hyperparameters that control the distribution of the parameter \( w_n \).

According to, it is known that adjusting the hyperparameters often makes it possible to improve the prediction performance. Thus, we consider that the hyperparameters are random parameters and adjust them together with the parameter \( w_n \). We use the likelihood of \( w_n \) and the transition densities of \( w_n \) \( \alpha_n, \beta_n \), which are formulated by.

The likelihood function is defined by

\[
p(D_n | w_n, \beta_p) = p(y_n | x_n, w_n, \beta_p)p(x_n) \quad \text{...............(14)}
\]

where
\[ p(y_n | x_n, w_n, \beta_n) = \frac{\beta_n}{2\pi} \exp\left(-\frac{\beta_n^2}{2} (y_n - f(x_n, w_n))^2\right) \]

...(15)

where \( \beta_n \) is a hyperparameter which stands for the uncertainty level of the observation noise.

The transition density of the parameter \( w_n \) is defined as the following equation.

\[ p(w_n | w_{n-1}, \alpha_n) = \left(\frac{\alpha_n}{2\pi}\right)^{w_n/2} \exp\left(-\frac{\alpha_n}{2} \left| w_n - w_{n-1} \right|^2\right) \]

...(16)

where \( \alpha_n \) is also a hyperparameter which controls the parameter \( w_n \).

The transition densities of the hyperparameter \( \alpha_n, \beta_n \) is defined as follows:

\[ p(\alpha_n | \alpha_{n-1}) = \frac{1}{\sqrt{2\pi} \sigma_n} \exp\left(-\frac{1}{2\sigma_n^2} (\log\alpha_n - \log\alpha_{n-1})^2\right) \]

...(17)

\[ p(\beta_n | \beta_{n-1}) = \frac{1}{\sqrt{2\pi} \sigma_n} \exp\left(-\frac{1}{2\sigma_n^2} (\log\beta_n - \log\beta_{n-1})^2\right) \]

...(18)

where, \( \sigma_n \) is assumed to be positive constant.

We can calculate the posterior and predictive distribution by using the formulation given above.

The prior distribution is given by

\[ p(w_n, \alpha_n, \beta_n | D_{t=n-1}) = \frac{1}{\pi^{w_n/2} \Gamma(n+n-n-1)\sigma_n} \exp\left(-\frac{1}{2\sigma_n^2} \sum_{i=1}^{n+n-n-1} (x_n - x_i)^2\right) \]

...(19)

The posterior distribution is given by

\[ p(w_n, \alpha_n, \beta_n | D_{t=n-1}) = \frac{1}{\pi^{w_n/2} \Gamma(n+n-n-1)\sigma_n} \exp\left(-\frac{1}{2\sigma_n^2} \sum_{i=1}^{n+n-n-1} (x_n - x_i)^2\right) \]

...(20)

As soon as the all learning data is given, we can calculate the predictive distribution by using the query variable vector \( x^q \).

\[ p(w_q, \alpha_q, \beta_q | D_{t=n}) = \frac{1}{\pi^{w_q/2} \Gamma(n+n-n-1)\sigma_n} \exp\left(-\frac{1}{2\sigma_n^2} \sum_{i=1}^{n+n-n-1} (x_n - x_i)^2\right) \]

...(21)

The transition density of the parameter \( w_n \) is defined as the following equation.

\[ p(w_n, \alpha_n, \beta_n | D_{t=n}) = \int \int p(w_n | w_{n-1}, \alpha_n) \int p(\alpha_n | \alpha_{n-1}) \int p(\beta_n | \beta_{n-1}) \]

...(22)

The predict value is calculated as a mean value of the predictive distribution.

\[ \hat{y}^q_t = \int y \cdot p(y | x^q, D_{t=n}) dy \]

...(24)

In case of using on-line learning scheme, complicated high-dimensional integral is typically required to calculate the posterior and predictive distribution. In this paper, we use Sequential Monte Carlo to overcome such a problem of on-line learning scheme. Sequential Monte Carlo consists of iterating importance sampling and resampling, which are given below.

A) Importance Sampling

(i) Obtain \( w_n^{(i)}, \alpha_n^{(i)}, \beta_n^{(i)} \) which are i.i.d samples \( w_n, \alpha_n, \beta_n \) from proposal distribution \( Q(\cdot) \). Where \( i \) stands for the index of samples. We use the prior distribution as the proposal distribution.

(ii) Calculate the normalized importance weight \( \hat{\Omega}(w_n^{(i)}, \alpha_n^{(i)}, \beta_n^{(i)}) \) as the following equation by using \( w_n^{(i)}, \alpha_n^{(i)}, \beta_n^{(i)} \) and \( D_{t=n} \).

\[ \hat{\Omega}(w_n^{(i)}, \alpha_n^{(i)}, \beta_n^{(i)}) = \frac{\Omega(w_n^{(i)}, \alpha_n^{(i)}, \beta_n^{(i)})}{\sum_{i=1}^{N} \Omega(w_n^{(i)}, \alpha_n^{(i)}, \beta_n^{(i)})} \]

...(25)

where

\[ \Omega(w_n^{(i)}, \alpha_n^{(i)}, \beta_n^{(i)}) = \int p(w_n^{(i)} | x_n, D_{t=n}) d\alpha_n d\beta_n \]

...(26)

...(27)

The Eq. (26) which is the update equation of the importance weight can be described simply because we consider the prior distribution as the proposal distribution.

B) Resampling

Draw the samples \( w_n^{(i)}, \alpha_n^{(i)}, \beta_n^{(i)} \) by using the normalized importance weight.

3. Simulation

3.1. Concrete Procedure of Calculation

In this paper, the large scale database is constructed by the process data of the No. 3 blast furnace in Nagoya works, Nippon Steel Corp. The number of data items is 275, the data sampling time is 1 h and the data stored period is from January 2004 to August 2006. Items of data are main data of the blast furnace operations of the blast volume, the blast pressure, and the reduction agent ratio etc.; and the sensor data of the stave temperature, the shaft pressure, and the gas composition, etc.; and the operation management indices calculated from the operation data and the sensor data of K value and the thermal heat ratio, etc.
We adopt the molten iron temperature as an example of the estimated data item. If LOM is used, all the data items can be forecast. However, the predictive accuracies of the molten iron temperature that was a very important item for the blast furnace operation were compared in this paper.

Concrete computing procedure is shown below.

**Step 1. Selection of Variable Used for Model**

With applying LOM, variables are selected by calculating “variable proportion $F$” against the molten iron temperature after 1 h from a present time with stepwise method. For example, adopting a general test criterion on stepwise method or, 415 variables are selected at first, but many noise variables still exist in the selected ones, so that the estimation accuracy for the molten iron temperature is not enough. Therefore adopting a more severe test criterion, or, 35 variables shown in the Table 1 are selected.

However, the variable that went back as a variable used for the retrieval before not only the value but also $N$ time (8 h ago usually) now is treated as independent variable. The selected data is different according to the data base (time etc. that go back to the period and the past) used and forecast variables.

**Step 2. Quantization of Data**

The data base composed of the selected variable is made a quantum, and it preserves it in the quantized data base.

| Table 1. Selected variables for molten iron temperature after 1 h from a present time. |
|---------------------------------|-----------------|-----------------|
| Number | Selected variables by stepwise method | F value |
| 1      | Molten iron temperature (present)       | 28879          |
| 2      | Tapping velocity (present)              | 706            |
| 3      | Molten iron Ti concentration (present)  | 416            |
| 4      | Molten iron temperature (before 4 hours)| 222            |
| 5      | Flame temperature (before 3 hours)      | 160            |
| 6      | Tapping velocity (before 1 hour)        | 134            |
| 7      | Slag Al2O3 (present)                    | 100            |
| 8      | Pulverized Coal Ratio (present)         | 94.6           |
| 9      | Heat Load (present)                     | 90.6           |
| 10     | Slag TiO2 (present)                     | 80.1           |
| ...    | ...                                        | ...            |
| 35     | #10 tuyere blast volume (present)       | 20.6           |

**Fig. 2.** Example of calculation of similarity.
Step 3. Retrieval of Neighborhood Data

The distance of the query and the neighborhood data is calculated with $\| \cdot \|$ norm. The calculated distance is called a similarity. Data that the similarity falls below a constant value is selected as neighborhood data.

Figure 2 is one example of calculating of similarity. Time when it wants to look for the past similar example data (query) is 0 o’clock August 11, 2004. The similarity calculated from the variable that related to the temperature of the molten iron temperature was shown in the first step. The vertical line shines at the time judged to be past similar example data. One example of the variable used the temperature of the hot pig iron and to forecast is shown since the second step. It is understood not only the temperature of the hot pig iron but also to show similar movement to other variables about the place in which it was judged a similar example.

Step 4. Calculation of Estimated Data

The predictive data of the query point will be forecast by supplementing the past similar data with using the local model. When estimation, past similar data below a constant value is selected from the calculated similarity in the above-mentioned step, and the local model is made. In this paper, two models of the weighted multiple regression model and on-line Bayesian learning model were compared as a local model.

3.1. Comparison of Local Models

We apply the two types of local models in LOM shown in the previous section to the real data sets of blast furnace and compare the prediction performance of the two new types of local models. The performance of the molten iron temperature prediction is evaluated by the correlation coefficient between the predicted molten iron temperature and the actual molten iron temperature. We picked up 200 data as the query variable vector $x_k^q$ from the database and predicted the molten iron temperature. In case of weighted multiple regression, we use 400 data as the neighboring data. In case of on-line Bayesian learning, we use 100 data as the neighboring data.

As an illustrative example, we show the results of 1, 2, 4, 6 h prediction of molten iron temperature. Figure 3, Fig. 4, Fig. 5 and Fig. 6 correspond to the result of 1 h prediction, 2 h prediction, 4 h prediction and 6 h prediction by using the weighted multiple regression model in LOM. Figure 7, Fig. 8, Fig. 9 and Fig. 10 correspond to the result of 1 h prediction, 2 h prediction, 4 h prediction and 6 h prediction by using the on-line Bayesian learning model in LOM. Table 1 summarizes the results of 1, 2, 4, 6 h prediction of two types of local models.

From Table 2, we see that the prediction performance in the two models is worse as time of prediction gets longer. In case that time of prediction is short, the prediction performance of the weighted multiple regression model is bet-
ter than that of on-line Bayesian learning model. The prediction performance of the on-line Bayesian learning is better than that of multiple regression model as the time of prediction gets longer.

4. Conclusion

In this paper, we have proposed the two new types of local models in LOM. One of the proposed local models was the weighted multiple regression model and the other was the on-line Bayesian learning model. The former model incorporated the weight between query variable vector and the neighboring data in multiple regression model to improve the prediction performance. The latter model was used three layer perceptron as a nonlinear model. In order to alleviate the curse of dimensionality, we used mutual information for input variable selection. To estimate the parameters of nonlinear local model, we used on-line Bayesian learning scheme with Sequential Monte Carlo.

The two new types of local models in LOM were applied to the real data sets of blast furnace. As a result, the better prediction performance of the weighted multiple regression model was demonstrated in case that time of prediction was short, and the prediction performance of the on-line Bayesian learning model was better than that of the multiple regression model as time of prediction got longer.

REFERENCES