Modeling of Power Plant Superheated Steam Temperature Based on Least Squares Support Vector Machines

Tian Haijun 1, Wang Jingru 2

1School of Automation Engineering, Northeast Dianli University, Jilin City Jilin province, China
2Institute of Information technology, Beihua University, Jin City Jilin province, China

Abstract

Aiming at the strong nonlinearity and large time-varying characteristics in controlling of super-heater temperature in plant, the method of LS-SVMs (Least Squares Support Vector Machines) based on radial basis function are used to model. Under the condition of modeling approximating to performance, the sparse modeling is gotten by the pruning algorithm. The merits of the algorithm are conforming to the least structural risk in training process and hardly leading to over-fitting. The simulation of a superheating system, in one supercritical concurrent 600MW boiler in one power plant, is taken. The result shows that the controlling system can be adapt to the variation of the object characteristic well with strong nonlinearity and large time-varying characteristics rapidly.

© 2012 Published by Elsevier Ltd. Selection and/or peer-review under responsibility of Hainan University.
Open access under CC BY-NC-ND license.

Keywords: support vector machine, radial basis function, pruning, Sparse, least squares

1. Introduction

The objects of all level overheated steam in large units in power plants are nonlinear, which are generally high in both inertia time constant and lag time. In boiler operation, there are many factors influencing steam temperature at the over heater outlet, such as steam flow, combustion condition, boiler water supply temperature, steam enthalpy at over heater entrance, temperature, flow, speed of smoke steam running through over heater and slagging condition of boiler heating surface and etc. And the main influential factors are steam flow, smoke steam heat transfer and water spray. The technology based on linear models is maturely applied in the thermal process control. However, characterization of high nonlinear process with linear dynamic models is inadequate and difficult as well [1]. With the improvement of the complexity of production process and the higher and higher requirements for the running performance, the research on the basis of nonlinear models has been promoted.

Neural network has been proved to be the effective method in the establishment of nonlinear process models [2]. The neural network reflects the input space onto the output space through a connection model, which enables the approximation of nonlinear function with arbitrary precision. The dynamic model built as per this property can be used in the prediction and optimization of process control. Therefore to a great extent the nonlinear prediction model based on the neural network solves the problem of the establishment
of process models. Whereas some suspensive problems still remain in the modeling of neural network, such as the training problem in neural network, local minimum problem and the determination of neural network structure. Of these problems, the design of the network structure is a problem due to the extraordinary rich dynamic behaviors of nonlinear systems, which is nothing like the linear system where a explicit and uniformed general model can be used to express the structure. In recent years, there emerges a new learning machine-support vector machine which automatically studies the structure of the model in question as per risk minimization principle of the structure. It is truly a learning machine to be able to choose the structure automatically. Consequently this thesis studies the overheated steam temperature system with strong nonlinearity in power plants to establish the prediction model using support vector machine according to the specificity and complexity of the object. Meanwhile pruning algorithm is adopted in the treatment of support vector.

2. LS-SVM algorithm and modeling

Vapnik [3] proposes a theory of machine learning rules aimed at small sample conditions on the basis of statistic learning theories. This theory establishes a set of new theoretical system-support vector machine (or SVM) aimed as small sample conditions. SVM seeks the best compromise between the complexity and learning ability of models in accordance with the limited sample information so as to achieve the best generalization. Compared with the traditional neural network, SVM algorithm will eventually transforms into a quadratic optimization[4]. Theoretically, the global optimum point will be acquired so that the inevitable local minimum problem in the neural network can be solved. The topology structure of SVM is determined by support vectors so that the inevitable demands for test pieces in the traditional neural network can be avoided [5].

The “over fitting problem” must exist since the algorithm using neural network as prediction models is based on the empirical risk minimization principle[6, 7]. But SVM method takes into account both the empirical risk of learning algorithm and promotion capacity to fit nonlinear function. Suykens increased the error sum of squares in SVM objective function in 1999, putting forward the measure of least squares support vector machines (LS-SVM). It solved the problems of robustness, sparsity and large-scale computing. LS-SVM needs no solution to quadratic programming problem in SVM due to its adoption of equality constraints.

A. LS-SVM algorithm and modeling

If, train set \( \{x_t, y_t\}_{i=1}^N \), \( x_t \in \mathbb{R}^n \) is the input mode of the t ordinal sample, \( x_t \in \mathbb{R}^n \) the desired output corresponding to the t ordinal sample, N sample number, LS-SVM is to predict the unknown function with the following forms:

\[
y(x) = w^T \varphi(x) + b
\]  

(1)

In the above, the nonlinear function of \( \varphi(\Pi): \mathbb{R}^n \rightarrow \mathbb{R}^{n_k} \) reflects the input space to the high dimension feature space, where the dimensionality of \( w \) should not be pre-designated (with possibility of infinite dimension). In LS-SVM, the objective function is characterized as:

\[
\min J(w,e) = \frac{1}{2} w^T w + \frac{1}{2} \sum_{i=1}^N e_i^2
\]  

(2)

To satisfy constraints:

\[
y(x) = w^T \varphi(x_i) + e_i, t = 1 \cdot N
\]  

(3)

Provided the equality is constrained and the loss function of the optimized objective being the second norm of the error \( e_i \), it will simplify the solution to the question. Thereby, Lagrange Function is defined as
\[ L(w,b,e,\alpha) = J(w,e) - \sum_{i=1}^{N} \alpha_i \{ w^T \varphi(x_i) + b + e_i - y_i \} \]  

(4)

Of the above, \( \alpha_i \) being Lagrange multiplier, from Karush-Kuhn-Tucker(KKT) of the optimal, condition

\[
\begin{align*}
\frac{\partial L}{\partial w} &= 0 \rightarrow w = \sum_{i=1}^{N} \alpha_i \varphi(x_i) \\
\frac{\partial L}{\partial b} &= 0 \rightarrow \sum_{i=1}^{N} \alpha_i = 0 \\
\frac{\partial L}{\partial e_i} &= 0 \rightarrow \alpha_i = y_i \\
\frac{\partial L}{\partial \alpha_i} &= 0 \rightarrow w^T \varphi_i + b + e_i - y_i = 0, t = 1 \cdots N
\end{align*}
\]  

(5)

Eliminate \( e_i \) and win formula (5) with \( \alpha_i, b \) to get the following linear equation:

\[
\begin{bmatrix} 0 & 1^T \\ 1 & \varphi(x_i)^T \varphi(x_i) + D \end{bmatrix} \begin{bmatrix} b \\ a \end{bmatrix} = \begin{bmatrix} 0 \\ y \end{bmatrix}
\]  

(6)

The following matrix reversible:

\[
\varphi = \begin{bmatrix} 0 & 1^T \\ 1 & \varphi(x_i)^T \varphi(x_i) + D \end{bmatrix}
\]  

(7)

Then the LS-SVM algorithm optimization is transformed into the linear equation of solution (6) with least squares:

\[
\begin{bmatrix} b \\ a \end{bmatrix} = \varphi^{-1} \begin{bmatrix} 0 \\ y \end{bmatrix}
\]  

(8)

B. Establishment of LS-SVM model

Substitute formula (7) to the optimal condition KKT to get w, consequently obtaining the nonlinear approximation of training data set:

\[ y(x) = \sum_{i=1}^{N} \alpha_i k(x,x_i) + b \]  

(9)

In the above,

\[ k(x_i,x_k) = \varphi(x_i)^T \varphi(x_k) \]  

(10)

Being kernel function, it can be any symmetric function satisfying MerCer theorem. This thesis uses the most common radial basis functioning (RBF) as kernel function,

\[ K(x,x_i) = \exp\left\{-\frac{||x-x_i||^2}{2\sigma^2}\right\} \]  

(11)

\( \sigma \) in the equation is positive constant. we can obtain the following nonlinear model of the controlled objective from (8)-(10):
\[ y(x) = \sum_{i=1}^{N} \alpha_i \exp \left\{ -\|x - x_i\|^2 / 2\sigma^2 \right\} + b \]  

(12)

**C. Pruning algorithm based on LS-SVMn**

As for the modeling process of the complex systems, the data set used to train models should contain information as much as possible. If the systematic information contained in the training data set is not adequate enough, the generalization capacity of the model shall be improved to guarantee the model performance. From formula (2), the objective function of LS-SVM includes the information in two fields. The first field information is used to improve the model generalization, the second used to reduce the desired model risk. Therefore a conspicuous advantage of LS-SVM is that it is easy to improve the generalization of the model for it is easy to adjust itself between the improvement of generalization and reduction of the desired risk. So this thesis treats the data set using pruning algorithm on the basis of the guarantee of function approximation and model generalization. From \( \frac{\partial L}{\partial e_i} = 0 \rightarrow \alpha_i = ye_i \) in KKT condition of equation (5), we know that in LS-SVM all the elements of \( \alpha \) are non-zero and that all the data vectors \( x_i \) are support vectors, which shows LS-SVM no longer possesses the support vector sparsity in the standard SVM\[6\]. The pruning algorithm is as follows:

- Suppose the train set is N, train LS-SVM.
- Eliminate a small amount of the least significant data from the data set (For example, the least data in \( \alpha_i \) is chosen, which covers 10% of the training set).
- Use the rest data vectors to re-estimate function, i.e., to re-train LS-SVM.
- Repeat the above mentioned two steps till the performance of function drops dramatically

**3. Simulation study of overheated steam temperature control system based on LS-SVM**

**D. modeling of overheated steam temperature control system based on LS-SVM**

The overheated steam temperature in thermal power plants features dynamically nonlinear and slow time varying. Besides, in different operating points its dynamic features vary far differently. This thesis fits the uniform SVM nonlinear model in different working conditions resulting from the dynamic features of different working points based on LS-SVM method according to the characteristics of overheated steam temperature system. The thesis [8] points out that the dynamic feature of overheated steam temperature objective is the function of main steam flow \( D \), main steam pressure \( P \) and temperature \( T \). Of the above, the variation of the main steam flow affects the objective dynamic features the most. The temperature influence can be overlooked when the direct boiler runs normally. The following table 1 shows the dynamic features of the steam temperature against spraying water disturbance under 100%, 75% and 50% load for the super heater of the supercritical 600MW direct boiler of certain power plant.

<table>
<thead>
<tr>
<th>Load</th>
<th>Anterior guidance C/(kg/s)</th>
<th>Inert zone C/C</th>
<th>Steam flow D kg/s</th>
</tr>
</thead>
<tbody>
<tr>
<td>50%</td>
<td>( \frac{-3.067}{(1+25s)^2} )</td>
<td>( \frac{1.119}{(1+42.1s)^2} )</td>
<td>242.2</td>
</tr>
<tr>
<td>75%</td>
<td>( \frac{-1.657}{(1+20s)^2} )</td>
<td>( \frac{1.202}{(1+27.1s)^2} )</td>
<td>404.7</td>
</tr>
<tr>
<td>100%</td>
<td>( \frac{0.815}{(1+18s)^2} )</td>
<td>( \frac{1.276}{(1+18.4s)^2} )</td>
<td>527.8</td>
</tr>
</tbody>
</table>
The uniform nonlinear model is built as per method in episode two given the dynamic features in table 1 of the 2 different working point inert zones with different loads. The input and output orders of the model is taken respectively as n=3, m=4. Use sinusoidal signal of the white noise and the different frequencies as the input to generate 400 input/output samples. The training set of \( \{ x_k, y_k \}_{k=1}^{400} \) is thus made, of which 200 of them will be used to train SVM, another 200 used for the generalization testing of the trained SVM. Take \( y = 10, \sigma = 2.0 \), the simulation results are shown in figure 1.

![Figure 1. Fitting testing set](image)

**E. Usage of pruning algorithm to treat sparsity and select parameters**

In order to get sparse SVM, here the method in section II.C is used to prune the support vectors. Meanwhile the following performance indicators are defined to measure the variation of function approximation performance before and after the pruning and the changes of output errors. It is shown in figure 2.

![Figure 2. SVM data set fitting after pruning](image)

\[
J(Y, Y_m) = 100 \times (1 - \frac{Y^2 - Y_m^2}{Y^2})
\]

\( Y \): real output; \( Y_m \): model output

\[
RSMS = \sqrt{\frac{\sum_{k=4}^{N} (Y_k - Y_{mk})^2}{(N - 4)}}
\]

\( Y_k \): real output; \( Y_{mk} \): model output
Different numbers of support vectors are gained during the process of LS-SVM modeling and pruning algorithm. By analyzing LS-SVM modeling process, table 2 function approximation performance and output errors changes, the following conclusions can be drawn.

### Table II. Comparison of approximation performance of different number of SVM and different parameters

<table>
<thead>
<tr>
<th>SVM number</th>
<th>Function approximation performance</th>
<th>( J )</th>
<th>( \sigma )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>( \gamma=10 )</td>
<td>( \gamma=50 )</td>
</tr>
<tr>
<td>397</td>
<td></td>
<td>92.6456</td>
<td>96.9697</td>
</tr>
<tr>
<td>358</td>
<td></td>
<td>92.6280</td>
<td>96.9801</td>
</tr>
<tr>
<td>318</td>
<td></td>
<td>92.6620</td>
<td>97.0592</td>
</tr>
<tr>
<td>278</td>
<td></td>
<td>92.8130</td>
<td>97.1446</td>
</tr>
<tr>
<td>239</td>
<td></td>
<td>93.4393</td>
<td>97.8437</td>
</tr>
</tbody>
</table>

- Influence of parameter \( \gamma \) to the model: High-precision model is gained when \( \gamma \) is bigger, i.e., the higher function approximation and the lower output errors are gained with the reduction of generalization capacity. Suppose the train set is \( N \), train LS-SVM.
- The network complexity degree of neural network sometimes rests on the numbers and sizes of the network weight numbers to a great extent. In order not to get serious “over fitting” phenomenon, a training sample 30 times weight is needed [9]. For instance, if a neural network of \( 10 \times 20 \times 1 \) is weighted \( 10+200+1=211 \), then 3660 training samples are needed. The high-quality training samples of such big numbers are hard to get in the field. While LS-SVM is supposed to improve the generalization of the learning machine as much as possible according to the principle of structure risk minimization, i.e., to gain the minimum errors as per limited training samples, it still guarantees the small errors for independent testing set.
- SVM algorithm is highlighting optimization, so the local optimization solution must be the global optimization solution.
- The use of pruning algorism can not only reduce the support vectors numbers in need, but reduce the computing time of the system.

### 4. Conclusions

Thesis innovation: LS-SVM method is used to generally model the overheated steam temperature system in power plants. Because of following structure risk minimization principle in the train process of LS-SVM, both the empirical risk and the promotion ability of the model can be taken into account. Over fitting phenomenon is hard to take place. Then pruning algorithm is used to prune the support vectors of the gained model so as to decrease the numbers of support vectors. The experiment results show the effectiveness of this algorithm. The modeling measure processes of LS-SVM is easy to follow and convenient to adjust. In addition, the generalization capacity of the gained model is pretty strong.

### References


