Soft Instrument for the Flue Gas Oxygen of Power Plant Based On Improved SMO Algorithm

Yong-jie Zhai [1] Hong Qiao [2] Hai-li Li [1] Guo-rui Ji [2] Pu Han [1]

[1] School of Control Science and Engineering,
North China Electric Power University, Baoding 071003, China
[2] School of Energy and Power Engineering,
North China Electric Power University, Beijing 102206, China

Abstract

As to the problem that normal SVM algorithm has a high computational complexity with large scale data and the method of selecting parameters of the study machine is complexity,we improved the SMO algorithm in two aspects of structure and parametric selection to increase operational speed and efficiency of modeling. It used grey theory to select the auxiliary variables and build a model of soft instrument for the flue gas oxygen content in power plant. The simulation with historical data measured by plant show that compared with the normal SMO algorithm the improved algorithm is better in performance.

Keywords: Sequential Minimal Optimization (SMO); Support vector regression; Grey; Oxygen; Soft-sensing

1. Introduction

With the hard work of peak-shaving in power system, frequent load-changing occurs in thermal power plant. The operation personnel have to make many parameters adjusted for the boiler furnace safety and economical operation, such as air outlet velocity and air flow rate of burner and air volume in furnace, and then optimize boiler burning condition by changing oxygen volume. It's important for boiler burning efficiency to measure the flue gas oxygen content timely and accurately. At present, oxygen volume analyzers based on zirconia oxygen sensors or thermomagnetic oxygen analyzers are widely used[1]. However, burning efficiency is often influenced owing to many disadvantages,such as, low precision,high investment, short service life, long time-delay in measure and not favorable to monitor on-line and to offer real-time feedback signal to closed-loop control.

Recently, the algorithm of SVR has widely used in soft-sensing modeling of industrial object, and has been proved to be good at test and measure. The main idea of SMO algorithm which was put forward by Platt is to make one big optimization problem into a series of optimization problems only including two variables. SMO algorithm is only used to classification firstly, but then SMO algorithm of training regression SVM which is the analogy and extension of Platt's algorithm was put forward by Smola and Scholkopf , and made it possible for SVM to process mass data regression problem. Now SMO with the advantages of rapid speed, low memory owing to no matrix operation and easy to realize has been one of the more effective SVR training algorithm. Therefore, it has bright future to apply it to soft-sensing modeling.

The result of this paper which made some improvements on SMO algorithm and combined with gray theory shows that the improved algorithm has fast convergence and the soft-sensing model has high precision.

2. SVR and adaptively selecting of parameters

2.1. Principle of SVR algorithm

Aimed at non-linear regression problem, SVM processes nonlinear transformation by defining proper kernel function to transform input space to high-dimensional space, and then searches for support vectors by linear regression in the new space. Optimal separating hyperplane is constructed by training samples which are in the hyperplane paralleling with optimal separating hyperplane and nearest from it. SVM can achieve much better generalization than neural network model and fuzzy model.

The basic idea of support vector regression is to map the input vector x into high dimensional feature space by nonlinear mapping function Φ and then to perform linear regression in the feature space. This transformation is realized by Kernel function $K(x_i,x_j)=\Phi(x_i)\cdot\Phi(x_j)$. It can be written as follows[2]:

$$f(x) = w\cdot\Phi(x)+b \tag{1}$$

Φ : $\chi \to H$, $w\in H$, b is threshold value.

The coefficients w and b are estimated by minimizing

$$E(w)=C\frac{1}{N}\sum_{i=1}^{N}\left|y_i-f(x_i,w)\right|_\varepsilon+\frac{1}{2}\|w\|^2 \tag{2}$$

$C\in R^+$, which determines the trade-off between the empirical risk and the regularization term $\frac{1}{2}\|w\|^2$. $|\cdot|_\varepsilon$ is the ε-insensitive loss function given by (3)

$$|x|_\varepsilon=\begin{cases}0 & if \quad |x|<\varepsilon\\ |x|-\varepsilon & \text{else}\end{cases} \tag{3}$$

In order to obtain the estimations of w and b , Eq. (2) is transformed to Eq. (4) as optimal function, by introducing the positive slack variables ξ_i and ξ_i^* as follows:
Minimize

$$E(w)=C\sum_{i=1}^{n}(\xi_i+\xi_i^*)+\frac{1}{2}\|w\|^2$$

Subjected to：

$$\begin{cases}y_i-f(x_i,w)\le\varepsilon+\xi_i\\ f(x_i,w)-y_i\le\varepsilon+\xi_i^*\\ \xi_i,\xi_i^*\ge 0\end{cases} \tag{4}$$

Slack variables ξ_i and ξ_i^* can be introduced when data can't be estimated by the function f under the precise ε [3].

Introducing Lagrange multipliers and according to Karush-Kuhn-Tucker conditions, Eq. (4) can be transformed into the form as follows:

Minimize:

$$L_p(\alpha^*,\alpha)=\varepsilon\sum_{i=1}^{N}(\alpha_i^*+\alpha_i)-\sum_{i}^{N}y_i(\alpha_i^*-\alpha_i)+\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}(\alpha_i^*-\alpha_i)(\alpha_j^*-\alpha_j)K(x_i,x_j) \tag{5}$$

Subject to:

$$\sum_{i=1}^{N}(\alpha_i^*-\alpha_i)=0,$$
$$0\le\alpha_i^*,\alpha_i\le C, i=1,2,\cdots,N$$

In Eq. (5), α_i , α_i^* are Lagrange multipliers.The model output is given by Eq. (6)

$$f(x,\alpha)=\sum_{i=1}^{N}(\alpha_i^*-\alpha_i)K(x_i,x)+b \tag{6}$$

2.2. Adaptive optimized select of SVR training Parameter

When non-linear modeling is implemented based on ε-SVR with fixed kernel function (for example, radial basis function kernel),it has to set three training parameters (kernel, C, and ε).They directly affect complexity and precision of the final model. The Cut-and-Try Method in tradition requires modeling personnel more experienced and makes the selection process a black box[4]. A parameter selection method based on training data is put forward by Cherkassky and Ma, in which C, and ε are obtained based on training data not by sample selection. It can be shown in Eq. (7) to Eq. (9).

$$C=\max(\left|\bar{y}+3\delta_y\right|,\left|\bar{y}-3\delta_y\right|) \tag{7}$$

$$\varepsilon=3\delta\times\left[\ln \mathrm{n}\ \times \mathrm{n}^{-1}\right]^{-1/2} \tag{8}$$

$$\hat{\delta}=\sqrt{\frac{n^{1/5}k}{n^{1/5}k-1}\frac{1}{n}\sum_{i=1}^{n}(y_i-\hat{y}_i)^2} \tag{9}$$

Where $2\le k\le 6$

In Eq. (7), $\bar{y}$ is mean value of y in training data, and δ_y is standard deviation. In Eq. (8), ε is obtained form Central Limit Theorem. δ is standard deviation of input noise, while n is the number of training samples. In Eq. (9), δ is unknown at first, but can be estimated by the method based on the idea of k-neighborhood and calculated by processing linear regression to training data. K is the low bias/ high variance estimators, and $\bar{y}$ is the predicted value of y.

3. Algorithm of SMO and its improvement

3.1. Principle and implement of SMO

SMO algorithm can solve the QP problem of SVM rapidly. According to Osuna theory, the QP problem of SVM can be parted into several sub-problems when convergence is ensured. As to the standard SVM optimization problem, only two Lagrange multipliers are selected to be optimized each time, and then SVM can be updated by these two multipliers. The computing process won't end until all the multipliers are selected. Meanwhile, all the multipliers may accord with the KKT condition and the objective function will win the best[5].

The advantages of SMO lie with that only two Lagrange multipliers, which can be obtained by using the method of analysis, are needed each step. And thus the complex numerical solution method can be avoided[6]. Meanwhile, the two multipliers to be optimized are selected by using the heuristics method, and that makes sure the algorithm is efficient [7]. Comparing to other algorithms, SMO has the advantages of higher computation speed, less memory, and easy to implement. SMO includes two parts: one is to optimize the coupled Lagrange multipliers by using the method of analysis. The other is to select the multipliers to be optimized by heuristics. And the algorithm can be carried out by using two loops in the program： The nested block is used to select another sample to match the sample which doesn't satisfy the KKT condition (make Lagrange multipliers to be optimized). The main idea of selecting the second sample is to make the optimization step length the longest. Platt uses

$\left|E_1 - E_2\right|$ as the length of one Lagrange multiplier, and choose the sample with maximum of $\left|\mathrm{E}_1 - \mathrm{E}_2\right|$ as the second multiplier. If there is none to be optimized, another sample can be found to match the first one all over the non-bound samples. If we also can't find it, all the samples should be returned to. The last two methods both start at random locations to avoid certain direction's deviation. The outer layer iterate through all non-bound samples or all the samples: First we choose to go over all non-bound samples, and make samples violate the KKT condition adjusted until all the samples satisfy it. If none of non-bound samples need to be adjusted, we can go over all the samples. If some samples are optimized, all non-bound samples should be gone over again. Do things above over and over again until all the training subsets satisfy KKT condition.

3.2. Improvement of SMO

After the research on principle and implement of SMO, some improvements can be made. We can optimize step length and its threshold to get operation speed and precision improved. Concrete procedures are followed.

- $\left|(E_1 - E_2)/\eta)\right|$ can be used as the most optimized step length in the nested block to avoid problems aroused by rough step length $\left|E_1 - E_2\right|$ (for example, when $\eta = 0$,quadratic form of minimal optimization problem is just Semi-definite).
- Threshold can be introduced as the standard of adjusting step length. In the nested block, there is no need to adjust the step length when the Lagrange multiplier is less than a certain value, which is called the threshold ξ .Different sample sets have different optimized Lagrange multiplier, thus only defining threshold a constant is unreasonable. This text introduces two concepts, absolute threshold ξ_{a} and relative threshold $\xi_{\mathrm{r}}(\lambda_{\mathrm{b}} + \lambda_{\mathrm{b}}^{*})$. The threshold can be represented as $\xi_{\mathrm{r}}(\lambda_{\mathrm{b}} + \lambda_{\mathrm{b}}^{*} + \xi_{\mathrm{a}})$ .It requires Lagrange multiplier to beyond a absolute constant and has an obvious change simultaneously. So we can avoid the neglect of important adjustment resulted form the improper threshold, and also not spend time on unnecessary adjustment.

4. Flue Gas Oxygen Grey Soft-sensing Modeling

4.1. Selection of Auxiliary Variable Based on Grey Relational Analysis

Generally speaking, for a Soft-sensing Object, there are a big number and many more types of original auxiliary variable based on initial theory analysis, and there is greater degree of the coupling between them, so it's necessary to execute dimension reducing processing on input variable in order to improve the performance and accuracy of the model. In the power plant operation, air gas system is a typical Grey System which partial information is knowable, it's difficult to get all the mechanistic information for the air gas system, and it's hard to effectively obtain the optimal result selected by auxiliary variable only through mechanism analysis. Grey Theory's basic content is Grey Relational Analysis whose basic idea is according to the similarity degree between curves to judge the correlation degree between factors, which has no high requirements on sample amount, which needs no typical distribution law, and which the results of quantitative analysis are in general concordance with the results of qualitative analysis[7]. Therefore, we can combine mechanism analysis with grey relational analysis to select auxiliary variable of Soft-sensing Model so as to develop the performance and accuracy of model.

According to mechanism analysis, considering that coal quality variation, air leakage of furnace, incomplete combustion and other factors have some effect upon Flue Gas Oxygen, we can preliminary select main steam flow, main steam pressure, main steam temperature, feed water flow, fuel quantity, fuel air-opening degree, low heating value, coal-feeder speed, primary air flow, secondary air flow, induced draft fan current, feeding draft fan current, unit load, furnace temperature and other technological parameters as related parameter set, then successively analyze the potential relation between each variable and Flue Gas Oxygen using the method of Grey Relational Analysis, and we base on the relation to select out the ultimate auxiliary variable. As Table 1 shows, association analysis and selection are given to auxiliary variable using Deng's Correlation Degree, appointing threshold value based on correlation is 0.76, well then the Grey Correlation Degree between main steam flow, main steam pressure, main steam temperature, feed water flow, coal-feeder speed, primary air flow, feeding draft fan current, fuel quantity wait 8 technological parameters and oxygen content meet the demands, then this paper selects this parameters as auxiliary variables.

Table 1: Analyse of related Technological Parameters and Deng's Grey Correlation Degree

Technological Parameter	Correlation Degree	Technological Parameter	Correlation Degree
main steam flow	0.81	primary air flow	0.76
main steam pressure	0.76	secondary air flow	0.74
main steam temperature	0.77	induced draft fan current	0.74
feed water flow	0.76	feeding draft fan current	0.77
fuel air-opening degree	0.69	fuel quantity	0.76
low heating value	0.69	unit load	0.70
coal-feeder speed	0.78	Furnace temperature	0.75

4.2. Soft-sensing Modeling and Simulation Research

Aiming to the demand that oxygen content of boiler exhaust fume should be online measured in practical power plant operation, considering there are none mature technology and methods for the practical application up to now, in this paper, the improved algorithm of SMO based, combining foregoing Grey Uniform Relevance Theory, the study that improved algorithm of SMO based Grey Soft-sensing Modeling was performed. Use the auxiliary variables selected in the part 4.1 as the input of modeling, and the oxygen content as output. Simulation Data is the 400 groups' historical data that Tianjin Panshan Power plant 600MW Power Unit worked for 24 hours normally, steadily, continuously. In order to get the better modeling effect, curve fitting and de-noising of filtering were performed on Simulation Data before modeling, and standardized processing of Range Standardization was operated on the data, the numerical value after standardized between 0 and 1.

Then, model-training 200 groups' data, calculate out model parameters, C is 6.3244, ε is 0.094, using SVR Parametric Adaptive Method; and model-checking the remaining 200 groups' data, when determine RBF Kernel Parameter σ is 0.4 finally, the mean of square testing errors of the model is optimal. The final mean of square testing errors of the model E is 0.018, but E which is less than 0.03 can meet the need of practical application. The test result of Model Simulation based on Soft-sensing as Figure 1 shows.

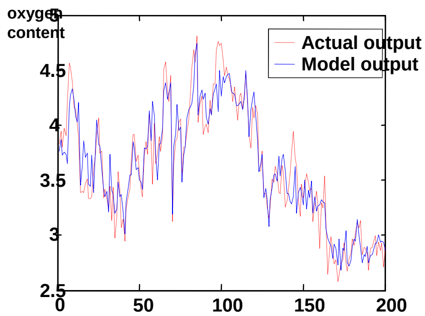

Figure 1: The Testing Curves of Improved Algorithm of SMO Based Soft-sensing Model

Feedforward Neural Networks can approach a nonlinear function using arbitrary precision, and has a better generalization ability; therefore, the learning algorithm has been widely used in modeling and controlling of industrial process and get better effects[6]. BP Neural Network is the most widely used network model in Feedforward Neural Networks, and in this paper, compare Grey Oxygen Soft-sensing Model used BP Neural Network with the Testing Curves of Soft-sensing Based on Improved Algorithm of SMO. Simulation results as Figure 2 shows, and the mean of square testing errors of the model is 0.033.

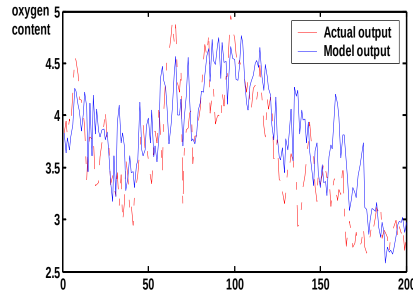

Figure 2: The Testing Curves of BP Algorithm Based Oxygen Soft-sensing Model

To verify the differences between Improved Algorithm of SMO Based Grey Soft-sensing Model and Ordinary Algorithm Model in modeling accuracy and convergence rate, I made a experimental group as Table 2 shows (use the 14 technological parameters showed in Table 1 as auxiliary variables of SMO modeling). The results shows that Grey Model Based on Improved Algorithm of SMO is better than Ordinary Soft-sensing Model Based on SMO Algorithm in modeling accuracy, and the Improved Algorithm of SMO also has obvious advantages in convergence time.

Table 2: Effect comparison between SMO andimproved SMO

Training sample number	Testing sample number	Improved SMO		SMO	
		mean of square testing errors	Conergence time/s	mean of square testing errors	convegence time /s
100	300	0.035	14.22	0.037	15.02
150	250	0.026	43.85	0.029	46.33
200	200	0.018	65.53	0.019	76.64

5. Conclusion

This paper studied SMO Algorithm of SVR training, improved the learning parameters and algorithm convergence structure, and this method combining improved Algorithm with Grey Theory is applied to Flue Gas Oxygen Soft-sensing Modeling of large power plant:

- The method of data acquisition based on SVR learning can effectively decrease the demand of the modeling engineers' experience, reduce the influence of data noise on model accuracy, and improve the modeling efficiency and the model accuracy.

- The optimization of Interior Cycle Iterative Optimal Step-size of SMO Algorithm and the introduction of assessment rules of threshold value can effectively improve algorithm performance and efficiency.
- Using Grey Correlation Analysis to select the auxiliary variable reduces the dimensionality and the coupling degree of the model's input, improves the Soft-sensing Model's accuracy and has a certain application potential.

6. Acknowledge

The research is support by Science and Technology Foundation of North China Electric Power University, No.200814003.

7. References

[1] Zhao Zheng，Zeng De-liang，Tian Liang， et al． Research on Soft-Sensing of Oxygen Content Based on Data Fusion [J]．Proceedings of the CSEE，2005,25 (7)：7－12．

[2] XiongZhi-hua,ShaoHui-he ． SVM-Based Soft Sensor Applied to O2 Content in Flue Gas in the Power Plant [J].Test Control Technology，2004，23 (8)：15－16．

[3] Platt .J . Fast training of support vector machines using sequential minimal optimization[M]. In B .Schölkopf , C .Burges , and A .Smola ,editors , Advances in Kernel Methods-Support vector Learning ,MIT Press,1998.

[4] Steve G. Support vector Machines Classification and Regression[R]. ISIS Technical Report. Image Speech & Intelligent Systems Group, University of Southampton. May 10, 1998

[5] V. Cherkasske & Y. Ma, 2004. "Practical selection of SVM parameters and noise estimation for SVM regression,"Neural Networks, Vol. 17(1), pp. 113-126

[6] Zhai Yong-jie，Yang Jin-fang，et al．The prediction of coordination system based on the algorithm of SMO [J]． Power Engineering，2005，25 (6)：849－854．

[7] Huang Fu-chen ， Yang Yong-guo．The Evaluation of Power Coal Quality Based on Grey Relational Analysis Method [J]．Energy Technology，2007，28 (1)：11－13．