Mine Working Face Gas Prediction Based on Weighted LS-SVM

Tiezhu Qiao[1] and Meiying Qiao[2]

[1] Institute of Measurement and Control Technology Tanyuan University of Technology Tanyuan, China
[2] School of Electrical Engineering and Automation Henan Polytechnic University Jiaozuo, China
qtz2007@126.com, qiaomy@foxmail.com

Abstract. Because coal and gas outburst prediction are very complex. In recent years, using least square support vector machine (LS-SVM) time series forecasting model to predict mine working gas is proposed. However in the search support vector solution process, inequality constraints become equality constraints in the LS-SVM, its advantage is to improve the algorithm speed, at the same time the sparse of support vectors and robustness to model are loss. In this paper, weighted LS-SVM is proposed to improve sparse and robustness and its time series prediction model is used to analysis short-time mine working face gas emission. Under MATLAB2009b environment, using LS-SVM1.7 toolbox, specific algorithm model is established, further model is verified by Hebi 10th 1113 mine and gas outburst working face time series data. The results showed that: weighted LS-SVM can achieve a better short-time gas prediction than standard LS-SVM; meanwhile its model has a better robustness.

Keywords: weighted LS-SVM, MATLAB, working face gas, time series.

1 Introduction

The disaster caused by mine gas outburst is extremely serious, so its prediction has been widespread concerned around the world. With the development of computer technology and artificial intelligence, method for predicting the mine methane has also been greatly developed, such as the chaotic time series prediction, gray relational analysis, and neural network prediction and so on[1]. However, there are many complex factors which give rise to gas outburst. Furthermore, the data, which are used to predict emission, always contain a lot noise and have certain randomness. So it is very difficult to accurately predict gas outburst using these data [2-3].

This paper presents short-term time series gas emission prediction through using LS-SVM model, which is based on structural risk minimization (SRM) induction principle. Through the following steps to achieve this principle: (1) Input vector is mapped to high-dimensional feature space by utilizing non-linear transformation; (2) In this space, according to the linear decision rules set construct a structure with the formal hyper-plane model. (3) Select the best structure elements and the best function

H. Deng et al. (Eds.): AICI 2011, Part III, LNAI 7004, pp. 639–647, 2011.

of this element in order to minimize the error rate boundary [4]. From the above implementation process can be summed up the advantages of SVM as follows: (1) From low-dimensional space to high-dimensional space non-linear problem is transformed into a linear, and the calculation in the high-dimensional space has been simplified since the introduction of kernel function; (2) Algorithm eventually be transformed into a convex quadratic optimization problem, the final solution is the global optimal point, avoiding the disadvantages that the neural networks fall into local minima, but also to improve the generalization ability[4].

At the same time, the two drawbacks of LS-SVM are brought because of the algorithm improvements. First, the LS-SVM has lost the sparse and robustness of SVM; Second, the error square is as cost loss function in the LS-SVM, however only when the error variable is Gaussian distribution, this assumption is reasonable, if this condition is not satisfied, it will produce large errors [5-6]. In order to overcome these two drawbacks, weighted LS-SVM is proposed. In this paper weighted LS-SVM prediction model is established to analysis Hebi 10th mine gas time-series data. Under MALTAB2009b environment, using LS-SVM1.7 toolbox, which is developed by K. Pelckmans and J.A.K. Suykens et al., time series prediction model is build. The gas outburst time series data from the Hebi 10th mine 1113 working face is used to verify model. At the same time error indicators of weighted LS-SVM and standard LS-SVM are compared. The results showed that: weighted LS-SVM can achieve a better short-time gas prediction than standard LS-SVM, at the same time its model has a better robustness.

2 Basic Principle of LS-SVM and Weighted LS-SVM

LS-SVM is an improved SVM, and is proposed by J.A.K. SUYKENS and J. VANDEWALLE in 1999.It has been used to many engineering fields. LS-SVM retains structural risk minimization principles, and changes inequality constraints into the equation. Thus, the quadratic programming problems change into solving equations problem, which improve the speed of the algorithm.

Given a Training set of N data points (x_i, y_i), $i=1,\cdots,N$, where $x_i \in R^m$ is the i-th input pattern and $y_i \in \{-1,1\}$ is the i-th output pattern.

In this method, the input data are mapped into a higher dimensional feature space by using non-linear mapping $\Phi(x)=[\varphi_1(x),\varphi_2(x),\cdots\varphi_n(x)]$, and an optimal separating hyper-plane can be constructed in this space.

$$\sum_{i=1}^{N} \omega^T \varphi(x_i) + b = y_i \tag{1}$$

Optimal hyper-plane geometry as a measure of the standard optimal hyper-plane is will be determined by the maximum geometry interval. So that the problem can be transformed as follows.

$$\min \frac{1}{2}\omega^T \omega \tag{2}$$

The statement is hold in the linear separable case, in order to make the algorithm more fault-tolerant nature, the introduction of slack variables ζ_i and penalty factors γ. In the LS-SVM is used in second-order soft margin classification. Therefore, LS-SVM optimization problem become as follows.

$$\min \quad J(\omega, b, \zeta) = \frac{1}{2}\omega^T \omega + \frac{1}{2}\gamma \sum_{i=1}^{N} \zeta_i^2 \tag{3}$$

$$s.t. \quad y_i[\omega^T \varphi(x_i) + b] = 1 - \zeta_i (i = 1,2 \cdots n) \tag{4}$$

n is the number of samples in here.

One can using Lagrange optimization rule to solve the above problems, the introduction of Lagrange multipliers constructor as follows.

$$L(\omega, b, \zeta, \alpha) = J(\omega, b, \zeta) - \sum_{i=1}^{N} \alpha_i \{y_i[\omega^T \varphi(x_i) + b] - 1 + \zeta_i\} \tag{5}$$

α_i is Lagrange multipliers in here.

According to Kuhn-Tucker condition, one can obtain as follows

$$\begin{gathered}
\frac{\partial L(\omega, b, \zeta, \alpha)}{\partial \omega} = 0 \to \omega = \sum_{i=1}^{N} \alpha_i y_i \varphi(x_i) \\
\frac{\partial L(\omega, b, \zeta, \alpha)}{\partial b} = 0 \to \sum_{i=1}^{N} \alpha_i y_i = 0 \\
\frac{\partial L(\omega, b, \zeta, \alpha)}{\partial \zeta_i} = 0 \to \alpha_i = \gamma \zeta_i \\
\frac{\partial L(\omega, b, \zeta, \alpha)}{\partial \alpha_i} = 0 \to y_i[\omega^T \varphi(x_i) + b] - 1 + \zeta_i = 0
\end{gathered} \tag{6}$$

In order to solve the equations, some middle variables are introduced as follows [7].

$$Z = [\varphi(x_1)^T y_1; \cdots; \varphi(x_N)^T y_N], Y = [y_1; \cdots; y_N] \ \vec{1} = [1; \cdots; 1], \quad \zeta = [\zeta_1; \cdots; \zeta_N];$$

$$\begin{bmatrix} I & 0 & 0 & -Z^T \\ 0 & 0 & 0 & Y^T \\ 0 & 0 & \gamma I & -I \\ Z & Y & I & 0 \end{bmatrix} \begin{bmatrix} \omega \\ b \\ \zeta \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vec{1} \end{bmatrix} \tag{7}$$

The solution is also given by equation (8).

$$\begin{bmatrix} 0 & Y^T \\ Y & \Omega + \gamma^{-1} I \end{bmatrix} \begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ \vec{1} \end{bmatrix} \tag{8}$$

Here, $\Omega = ZZ^T$ Mercer's condition can be applied to this matrix.

$$\Omega_{ij} = y_i y_j \varphi(x_i)^T \varphi(x_j) = y_i y_j K(x_i, x_j) \tag{9}$$

Given a kernel function, one can give the solution to determine the optimal classification surface by solving the linear set of Equations (7)-(9) instead of quadratic programming.

In the regression case, the objective function formula (4) is only rewritten as follows.

$$s.t. \quad y_i = \omega^T \varphi(x_i) + b - \zeta_i (i = 1,2 \cdots n) \tag{10}$$

n is the number of samples in here[8].

One can get the solution of least-squares regression equation by repeating the above calculation process.

$$\begin{bmatrix} 0 & \vec{1} \\ \vec{1} & \Omega + \gamma^{-1} I \end{bmatrix} \begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ Y \end{bmatrix} \tag{11}$$

Thus, the standard LS-SVM regression model can be drawn as follows.

$$y(x) = \sum_{i=1}^{N} \alpha_i K(x, x_i) + b \tag{12}$$

In order to improve the robustness and sparse of standard LS-SVM, a weight value v_i is multiplied to error variable in the standard LS-SVM. So the optimization problem can be written as follows:

$$\min_{\omega^*, b^*, e^*} J(\omega^*, e^*) = \frac{1}{2} \omega^{*T} \omega^* + \frac{1}{2} \gamma \sum_{i=1}^{N} v_i e_i^{*2} \tag{13}$$

$$s.t. \quad y_i = \omega^{*T} \varphi(x_i) + b^* + e_i^*, \qquad (i = 1,2 \cdots N) \tag{14}$$

Note: the unknown variables with "*" are used in weighted LS-SVM optimization problem. When weighting coefficient is 1, the weighted LS-SVM becomes common LS-SVM.

The same construct Lagrange function to solve the above equation, and then eliminate the middle variables ω^* and e^*.

$$\begin{bmatrix} 0 & \vec{1}^T \\ \vec{1} & \Omega + V_\gamma \end{bmatrix} \begin{bmatrix} b^* \\ \alpha^* \end{bmatrix} = \begin{bmatrix} 0 \\ y \end{bmatrix} \tag{15}$$

Where $V_r = diag\ (1/\gamma v_1, \cdots, 1/\gamma v_N)$, the selection of the weight vector estimate is based on the following formula [9]:

$$v_i = \begin{cases} 1 & if\ |e_k / \hat{s}| \le c_1 \\ \dfrac{c_2 - |e_k / \hat{s}|}{c_2 - c_1} & if c_1 \le |e_k / \hat{s}| \le c_2 \\ 10^{-4} & other \end{cases} \tag{16}$$

Where $\hat{s}$ is a robust estimates value of standard LS-SVM error, it is calculated as follows:

$$\hat{s} = \frac{IQR}{2 \times 0.6745} \tag{17}$$

Where IQR the interquartile range of the error variable, this value is used to estimate how many error variables deviate from the Gaussian distribution. Since the cost loss function of the standard LS-SVM is set only error variable is the Gaussian distribution, then the weight vector is added in the formula (16) can increase robustness of the system. So even the error variable does not obey the Gaussian distribution, a better analysis results can be gained by using this model. Usually constant $c_1 = 2.5, c_2 = 3$, they can be gain from the density estimation of the error variable [9].

The radial basis function (RBF) is used in this paper, its expression as follows.

$$K(x, x_i) = \exp(-\frac{(x - x_i)^2}{\sigma^2}) \tag{18}$$

The penalty factor γ and nuclear width σ^2 are two parameters which need to be adjusted in this model [5].

In this article 10 fold cross-validation commands in LS-SVM1.7 toolbox is used to determine these two parameters.

Wherever Times is specified, Times Roman or Times New Roman may be used. If neither is available on your word processor, please use the font closest in appearance to Times. Avoid using bit-mapped fonts if possible. True-Type 1 or Open Type fonts are preferred. Please embed symbol fonts, as well, for math, etc.

3 Prediction Model Algorithm Steps

Nonlinear time series prediction is currently based on the Takens' phase space reconstruction theory. The key of reconstruction phase space is how to determine the delay time τ and embedding dimension m . In this paper minimum differential Entropy ratio method is used to determine those two parameters [10-11].

Given a set of nonlinear time-series data $\{x_t\}_{t=1}^n$, firstly embedding dimension m and delay time τ can be determined by some method, Entropy ratio method is used in this paper. Then on this basis phase space can be reconstructed, LSSVM training and prediction are carried out in this space.

Reconstructed phase space as follows:

$$\left(X_1, x(t_{2+(m-1)\tau})\right), \left(X_2, x(t_{3+(m-1)\tau})\right), \cdots \left(X_{M-1}, x(t_n)\right) \tag{19}$$

Here $X_i = (x(t_i), x(t_{i+\tau}), \cdots x(t_{i+(n-1)\tau}))$ is the input of time series prediction model, and $x(t_{i+1+(m-1)\tau})$ is the output.

According to the Takens' phase space reconstruction theory, the state variables between exist function relationship.

$$x(t_{i+1+(n-1)\tau}) = f(X_i) \qquad i = 1,2,\cdots M-1 \tag{20}$$

Here $f(\cdot)$ is the state transfer function of the system, and this transfer function is one step time series forecasting model. Multi-step prediction model can gain by repeat these steps.

Robust weighted LS-SVM time series forecasting model algorithm steps are as follows:

- Processing the sample data, such as handling the exception samples, normalized samples. Gain the processed data samples $\{x_t\}_{t=1}^{n}$. According to minimum differential Entropy ratio method to gain the embedding dimension m and delay time τ .On this basis, using phase space reconstruction technique gain the sample data. Part of the samples as training samples, part of the sample as the test samples.
- Parameter initialization, according to training samples to adjust the standard LS-SVM algorithm model, where (γ,σ^2) is gained by the 10 fold grid search and cross-validation, here the search range is [0.1 10000; 0.1 10000].
- e_k is gained by using $e_k = \alpha_k / \gamma$,and $\hat{s}$ is gained from distribution of e_k . Weight vector v_k is calculated by e_k and $\hat{s}$, then robust weighted LS-SVM model can be obtained according to the formula (15).
- Train sample is used to train weighted LS-SVM. One step prediction model $\hat{x}(t_{i+1+(n-1)\tau}) = f(X_i)$ can be gained when end of the training [12].
- According to this prediction model can predict value $\hat{x}(t_{n+1})$ of $x(t_{n+1})$ time can be gained, repeat the above steps predict value $\hat{x}(t_{n+2})$ of $x(t_{n+2})$ time can be gained. Multi-step prediction can be gained by repeat multi step predict above step multi-step prediction model can be drawn.
- When the start points and predicted number are given, the regression model will give the prediction result. At the same time, the evaluation criteria, including of the root mean square error and maximum error percentage, will be given.

4 Example Analyses

In this paper, the gas outburst data is come from 1113 working face of the10th mine in Hebi. The time series data are the percentage of gas emission concentration. These data are used for the learning samples and the training samples. KJ93 environmental

monitoring system was used for monitoring Face tunneling head gas emission. The amplitude of the gas concentration is in the range of 0~4% using this system measurement.

There are 120 time-series data. Each interval between two points is 5 minutes. Embedding dimension and delay time are obtained by using minimum differential Entropy ratio method.

Fig. 1 shows the two parameters values in differential Entropy ratio map. Vertical axis is time-series differential Entropy rate, and two horizontal axes are respectively the embedding m and delay time τ . Minimum differential Entropy rate points corresponding to the two horizontal axes values are the optimal m and. In this example, "★" position show $m = 3, \tau = 1$.

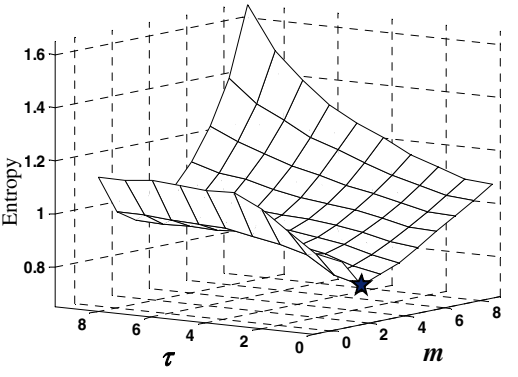

Fig. 1. Face gas time-series Entropy ratio

In the MATLAB2009b environment, LS-SVMlab1.7 package is used to compile the algorithm program. The model is trained by using given learning samples and training samples, then 20 point data will be forecasted through this learning machine, where each time interval is five minutes.

In this paper, Root Mean Square Error ($RMSE$) and Mean Absolute Percentage Error ($MAPE$) are used to evaluate the prediction results. The expression is as follows:

$$RMSE = \sqrt{\frac{1}{l}\sum_{i=1}^{l}\left(\frac{Y_i^{'} - Y_i}{Y_i}\right)^2} \tag{21}$$

$$MAPE = \frac{1}{l}\sum_{i=1}^{l}\left|\frac{Y_i^{'} - Y_i}{Y_i}\right| \tag{22}$$

Where $Y_i^{'}$ is the i-th predictive value. Y_i is the i-th true value. l is the predicted points number.

70 data points is selected as training samples. According to above algorithm step, 20 steps are predicted by using standard LS-SVM and the weighted LS-SVM, respectively. Coal and gas outburst time is included in the 20 points. Prediction results of LSSVM and weighted LSSVM are shown in Fig.2 and Fig.3.

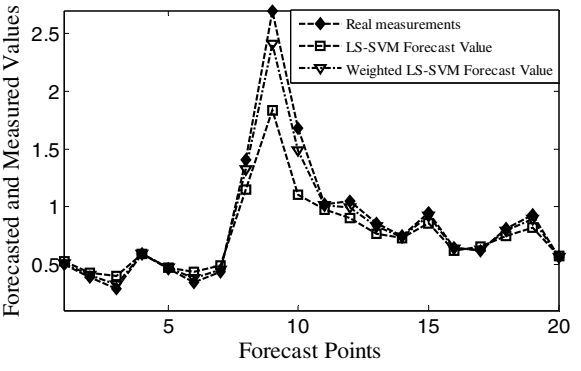

Fig. 2. Forecast curse of LS-SVM and weighted LS-SVM

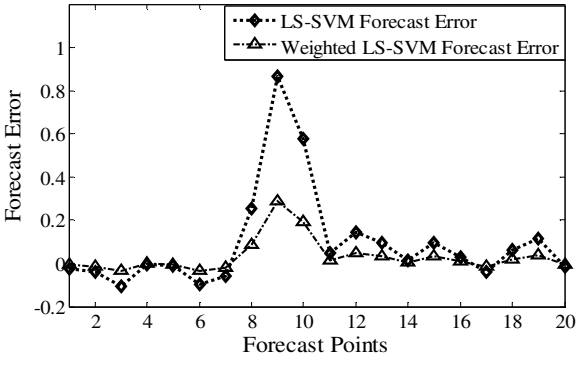

Fig. 3. Weighted LS-SVM and LS-SVM Forecast Error

Table 1. Error indicators comparison of two models

	RMSE	*MAPE*
LS-SVM	0.1678	0.1271
Weighted LS-SVM	0.0559	0.0424

Fig. 3 shows prediction error values at each point of the two models. Table 1 shows the predict effect evaluation index of two models.

5 Conclusions

In this paper, the basic theory of LS-SVM is discussed in detail based on the good generalization performance. Because of the two shortcoming of standard LS-SVM, a

weighted LS-SVM is proposed in this paper. Under MALTAB2009b environment, using LS-SVM1.7 toolbox, which is developed by K. Pelckmans and J.A.K. Suykens et al., time series prediction model is build. The gas outburst time series data from the Hebi 10th mine 1113 working face is used to verify model. At the same error indicators of weighted LS-SVM and standard LS-SVM are compared. The results showed that: weighted LS-SVM can achieve a better short-time gas prediction than standard LS-SVM, at the same time its model has a better robustness.

Acknowledgment. The author would like to thank to Liu Yanwei providing the gas time-series data during writing the paper.

References

1. Cui, W.Z., Zhu, C.C., Bao, W.X.: Prediction of the chaotic time series using support vector machines for fuzzy rule-based modeling. Acta Physica Sinica 54(7), 2221–2230 (2009) (in Chinese)
2. Yang, M., Wang, Y.J., Chen, Y.P.: Improved differential evolution neural network and it s application in prediction of coal and gas outburst. Journal of China University of Mining & Technology 38, 439–444 (2009) (in Chinese)
3. Wang, G.H., Wang, J.G., Wang, H.F.: Prediction of coal and gas outburst based on neural network and grey theory. China Coal Bed Methane 16, 27–30 (2009) (in Chinese)
4. Vapnik (ed.), Xu, J.H., Zhang, X.G.:Trans. Statistical Learning Theory. Publishing House of Electronics Industry (2004)
5. Vapnik, V.N., Xu, J., Zhang, X. (translation): Statistical learning theory. Electronic Industry Press, Beijing (2004)
6. Deng, N.Y., Tian, Y.J.: Support Vector Machines - theory, algorithms and expansion. Science Press (2009)
7. Ji, L.L., Lin, Z.S., Wang, C.Y.: Experiments on nonlinear time series prediction with least square support vector regression machine. Journal of PLA University of Science and Technology (Natural Science Edition) 110, 92–97 (2009)
8. Suykens, J.A.K., Vandewalle, J.: Least squares support vector machine classifiers. Neural Processing Letters 9, 439–444 (1999)
9. Suykens, J.A.K., De Brabanter, J., Lukas, L., et al.: Weighted least squares support vector machines: robustness and sparse approximation. Neurocomputing 48, 85–105 (2002)
10. Gautama, T., Mandic, D.P., Van Hulle, M.M.: A differential entropy based method for determine the optimal embedding parameters of a signal. In: Proc. of the Int. Conf. an Acoustic, Speech and Signal Processing, Hongkong, pp. 29–32 (2003)
11. Guo, X.H., Ma, X.P.: Coal washery daily water consumption short-term prediction based on least squares support vector machines. Journal of China Coal Society 32, 1093–1097 (2007) (in Chinese)
12. Zhao, X.H., Wang, G., Zhao, K., et al.: On-line least squares support vector machine algorithm in gas prediction. Mining Science and Technology 19, 194–198 (2009) (in Chinese)