Available online at www.sciencedirect.com
ScienceDirect
Neurocomputing 71 (2008) 550–558
NEUROCOMPUTING
www.elsevier.com/locate/neucom

Online prediction model based on support vector machine

Wenjian Wang[a,*], Changqian Men[a], Weizhen Lu[b]

[a]School of Computer and Information Technology, Key Laboratory of Computational Intelligence & Chinese Information Processing of Ministry of Education, Shanxi University, Taiyuan, 030006, PR China
[b]Department of Building and Construction, City University of Hong Kong, Hong Kong, PR China

Available online 29 September 2007

Abstract

For time-series forecasting problems, there have been several prediction models to data, but the development of a more accurate model is very difficult because of high non-linear and non-stable relations between input and output data. Almost all the models at hand are not applicable online, although online prediction, especially for air quality parameters forecasting, has very important significance for real-world applications. A support vector machine (SVM), as a novel and powerful machine learning tool, can be used for time-series prediction and has been reported to perform well by some promising results. This paper develops an online SVM model to predict air pollutant levels in an advancing time-series based on the monitored air pollutant database in Hong Kong downtown area. The experimental comparison between the online SVM model and the conventional SVM model (non-online SVM model) demonstrates the effectiveness and efficiency in predicting air quality parameters with different time series.

Keywords: Air pollutant; Online model; Prediction performance; Support vector machine; Time-series forecasting

1. Introduction

Monitoring and forecasting of air quality parameters are popular and important topics of atmospheric and environmental research today due to the health impact caused by exposure to air pollutants existing in urban air. Accurate models for air pollutant prediction are needed because such models would allow forecasting compliance and non-compliance in both short-term and long-term aspects. At present, monitoring and forecasting air pollutant trends in ambient air involves using a variety of approaches, e.g., on-site measurement, computational fluid dynamics simulations, computational intelligence techniques, etc. Among them, computational intelligence techniques like artificial neural networks (NN) [1,2], genetic algorithms (GA) [3], support vector machines [4–6], etc. are paid more and more attention in environmental time-series prediction researches because they can model non-linear systems well and are robust for the noise data, and so they can produce more accurate results. Among these computational intelligence techniques, NN have become most popular and produced promising results [2,6–13,14,15]. However, the inherent drawbacks in NN such as susceptibility to chaotic behavior, expensive computation on training, local minima, over-fitting problem, topology specification problem etc., have hampered their achieving good prediction performance as effectively and efficiently as expected in engineering applications [5,6]. As an alternative method to NN, the support vector machine (SVM) developed by Vapnik [16,17] can provide an effective approach to improve prediction performance and achieve a global optimization solution simultaneously. SVM has been raised as a powerful tool for solving classification, regression and time-series prediction problems [18–21] in the last few years. It implements the structural risk minimization (SRM) principle, which is an approach to minimize the upper bound risk functional related to the generalization performance; therefore, its solid theoretical basis ensures its possessing more salient advantages than other machine learning methods like NN in generalization and convergence [16]. The practical applications, especially in the environmental prediction domain, suggest that SVM is effective and can produce more accurate prediction results than NN models [4–6].

*Corresponding author. Tel.: +86 351 7010566; fax: 86 351 7018176.
E-mail address: wjwang@sxu.edu.cn (W. Wang).

0925-2312/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.neucom.2007.07.020

For the classical SVM model, learning has to be done from the first data in the case where data are sequentially obtained, that is, when a new observation arrives, learning has to begin from searching with the whole data set. Refs. [22,23] provided some accurate online SVM learning models. It is the case, to some extent, because the kind of models is just an "exact" solution within the model space determined by proceeding observations but not all the present examples. As we know, when new data are available, the proceeding optimal model space may be different from the succeeding optimal model space determined by the new data along with the existing data. Therefore, modification of the model space may be necessary in order to obtain good prediction performance. This is a crucial capability for online active learning scenarios. This paper presents an online SVM model to investigate potential variations of air pollutant, which were measured at the Mong Kok Roadside Monitoring Station during 2000. The performance of the online SVM model is evaluated by comparing with the results produced by the conventional SVM model.

This paper is organized as follows: section 2 reviews the basic idea of SVM for analysis of later sections, and then develops the online SVM model. Section 3 presents the experimental results and discussions. The last section concludes the proposed work.

2. The online SVM model

2.1. Brief review of SVM

The time-series problems can be brought down to support vector regression (SVR) problems [2]. In SVR, the basic idea is to map the data into a higher-dimensional feature space via a nonlinear mapping Φ and then to do linear regression in this space. Therefore, regression approximation addresses the problem of estimating a function based on a given data set $G=\{(x_i,y_i)\}_{i=1}^{l}$ ($x_i \in R^n$ is the input vector, $y_i \in R$ is the desired value). SVM approximates the function with the form

$$f(x)=\sum_{i=1}^{l} w_i\Phi_i(x)+b, \tag{1}$$

where $\{\Phi_i(x)\}_{i=1}^{l}$ are the data in features space, $\{w_i\}_{i=1}^{l}$ and $b$are coefficients. They can be estimated by minimizing the regularized risk function

$$R(C)=C\frac{1}{l}\sum_{i=1}^{l} L_\varepsilon(y_i,f(x_i))+\frac{1}{2}||w||^2, \tag{2}$$

where $L_\varepsilon(y,f(x))$ is the so-called loss function measuring the approximate errors between expected output y_i and calculated output $f(x_i)$, and C is a regularization constant determining the trade-off between the training error and the generalization performance. The second term, $\frac{1}{2}||w||^2$, is used as a measurement of function flatness. Introduction of slack variables ζ,ζ^* leads (2) to the following constrained function

$$\text{Minimize}\quad R(w,\zeta^*)=\frac{1}{2}||w||^2+C^*\sum_{i=1}^{n}(\zeta_i+\zeta_i^*) \tag{3}$$

$$\text{s.t.}\quad w\Phi(x_i)+b-y_i\leqslant\varepsilon+\zeta_i,$$
$$y_i-w\Phi(x_i)-b\leqslant\varepsilon+\zeta_i,$$
$$\zeta,\zeta^*\geqslant 0. \tag{4}$$

Although non-linear function Φ is usually unknown all computations related to Φ can be reduced to the form $\Phi(x)^{\mathrm{T}}\Phi(y)$, which can be replaced with a so-called kernel function $K(x,y)=\Phi(x)^{\mathrm{T}}\Phi(y)$ that satisfies Mercer's condition [16,17]. Then, Eq. (1) becomes the explicit form

$$f(x,\alpha_i,\alpha_i^*)=\sum_{i=1}^{l}(\alpha_i-\alpha_i^*)K(x,x_i)+b. \tag{5}$$

In (5), Lagrange multipliers α_i and α_i^* satisfy the equality $\alpha_i\times\alpha_i^*=0$, $\alpha_i\geqslant 0$, $\alpha_i^*\geqslant 0$, $i=1,...,l$. Those vectors with $\alpha_i\neq 0$ are called support vectors, which contribute to the final solution.

2.2. The online SVM model

In this paper, we present an online SVM model whose primary distinction from the conventional SVM model focuses on the manner of data provided, i.e., the data are provided in sequence for the presented online SVM model, while they are supplied in batch for the conventional SVM model. In our model, without loss of generality, suppose that the initial training data set with l samples. Then we select the optimal kernel function with the optimal parameter (say, the optimal model is established based on the current l data). As we know, for any orthogonal basis $(\rho_1,\cdots,\rho_k)$ in Hilbert space H, if $\theta\in H$, then $\Sigma_{i=1}^{k}\cos^2(\rho_i,\theta)=1$ holds. ($\cos(x,y)$ refers to the cosine function of the includes angle between vector x and y, and $\cos(x,y)=x^{\mathrm{T}}y/||x||||y||$). In the presented approach, a vector sequence $\{\alpha_1,\cdots,\alpha_k\}$, $\alpha_k=\Phi(x_i)y_j-\Phi(x_j)y_i$, $k=1,\cdots,l(l-1)/2$, is firstly constructed, and then an orthogonal vector sequence $\{\beta_1,\cdots,\beta_d\}$

$$\beta_i=\frac{\alpha_i-\sum_{j=1}^{d}\beta_j(\beta_j\cdot\alpha_i)}{||\alpha_i-\sum_{y=1}^{d}\beta_j(\beta_j\cdot\alpha_i)||} \tag{6}$$

with $d=$ rank $\{\alpha_1,\cdots,\alpha_k\}$ can be obtained by the well-known Schmidt's orthogonalization procedure. Because $\{W,\beta_1,\cdots,\beta_k\}$ (W is the normal vector of the regression hyperplane) is an orthogonal basis in H and each $\Phi(x_i)$ belongs to H, we have

$$\sum_{j=1}^{d}\cos^2(\beta_j,\Phi(x_i))+\cos^2(W,\Phi(x_i))=1. \tag{7}$$

Hence,

$$||W||=\frac{y_i}{||\Phi(x_i)||\sqrt{1-\sum_{j=1}^{d}\cos^2(\beta_j,\Phi(x_i))}}. \tag{8}$$

Because all computations can be transformed to kernel format, the optimal kernel parameter can be obtained by minimizing $||W||$. When a new observation arrives (we call it the $(l+1)th$ sample), we need to determine whether it works well with the current model. If so, i.e., $\alpha_{l+1} = \sum_{i=1}^{d} \beta_i \alpha_{l+1}^{\mathrm{T}} \beta_i$, the model is not changed, otherwise we select the optimal model under these $l+1$ samples. For the convenience of computations, we calculate $A(d+1,i,j), B(d+1,i), \gamma_{d+1}, C(d+1,i,j), \beta_{d+1}$ in sequence. Here, denote

$$A(l,i,j) = K(x_l, x_i)y_j - K(x_l, x_j)y_i,$$

$$B(l,i) = \beta_l^{\mathrm{T}} \Phi(x_i),$$

$$C(l,i,j) = B(l,i)y_j - B(l,j)y_i,$$

$$\gamma_j = \beta_j^{\mathrm{T}} \Phi(x_{i^*}) = B(j,i^*),$$

where i^* is any index in $\{1,2,\cdots,l\}$.

Because the successive optimal model is obtained on the basis of the proceeding optimal model, we need only one computing step to determine the optimal kernel corresponding to the minimum of an optimization problem. The whole computing cost in each cycle will not increase in comparison with SVM training. Moreover, the collection of data is not continuous, i.e., there is always an interval between measuring two sets of data for real applications. Hence, the presented online model can accomplish optimal model selection and find the optimal solution simultaneously (or, say, by the sequential data, it selects the optimal model space at first, and then finds the optimal approximation in the selected model space, recursively).

The main idea of the online SVM model can be as the following:

Step 1 (Initialization)

(1) Let the initial training data set G have summarized l samples.
(2) Let the class of kernel function

$$Ker(\textstyle\sum) = \{K_1(\textstyle\sum), K_2(\textstyle\sum), \cdots, K_P(\textstyle\sum)\},$$

where $K_i(\sum)$ is the ith type kernel function with continuously adjustable kernel parameters $\sum$.

Step 2 (Optimal kernel selection)

(1) For each $K_{\sum} := K(\sum) \in Ker(\sum)$, solve the following optimization problem

$$\Lambda^* = \mathrm{argmin}_{\Lambda \in \sum} \{F(K_\Lambda)\},$$

where $F(K) = ||W||^2 = \frac{y_{i^*}^2}{K(x_{i^*},x_{i^*}) - \sum_{j=1}^{d} \gamma_j^2}$.

(2) Then the optimal kernel is $K^*(\sum^*) = K_{o^*}(\Lambda^*)$, where $o^* = \arg \min_{1 \leqslant i \leqslant P} (K_i(\Lambda^*))$.

Step 3 (Online learning loop)

When a new sample, $(l+1)th$, arrives,

(1) If $\alpha_{l+1} \neq \sum_{i=1}^{d} \beta_i \alpha_{l+1}^{\mathrm{T}} \beta_i$, then calculate $\{A(d+1,i,j), B(d+1,i), \gamma_{d+1}, C(d+1,i,j), \beta_{d+1}\}$ and then go to step 2.
(2) Otherwise, the optimal kernel and corresponding online SVM model are not changed.

3. Simulation results and discussions

The simulation programs are constructed using Matlab 6.5. The most commonly used kernel, Gaussian kernel $K(x,y) = \exp(-|x-y|^2/(2\sigma^2))$, is specified at first, while ε and C are set to 0.01 and 100, respectively.

3.1. Original data set

The available air quality database measured at the Mong Kok Roadside Monitoring Station in 2000 is selected as the original data set. The database includes seven major air pollutants, i.e., carbon monoxide (CO), nitric oxide (NO), nitrogen dioxide (NO_2), sulphur dioxide (SO_2), nitrogen oxides (NO_x), ozone (O_3), and respirable suspended particulate (RSP), and five meteorological parameters, i.e., indoor and outdoor temperature (IT and OT), solar radiation (SR), wind direction (WD) and wind speed (WS), which were hourly measured at the said location.

In prediction experiments, the recorded levels of RSP, NO_x, and SO_2 in January and June are selected as original samples. The reason to choose the data in these two months is because January and June represent two different seasons in Hong Kong, i.e., January corresponds to dry, cold weather and is normally accompanied by the prevailing north-eastern wind and the highest pollutant levels (i.e., local vehicle exhausts combining with migration of industrial pollutants from Mainland China), while June corresponds to hot, wet season and often undergoes the dominant south-eastern wind and the lowest pollutant concentrations (i.e., local vehicle pollution dominates). Hence, the robustness of the online SVM and the conventional SVM models can be verified by seasonal variation. In simulations, the data of the first 10 days (240 data points) in each month are used as training data. Two SVM models, the presented online SVM and the conventional SVM models, are then used to predict the pollutant levels in different time series, i.e., 1 day and 1 week predictions for coming periods. Thus, the simulation results have either 24 test points corresponding to the hourly measurements on the 11th day of selected month, or 168 test points representing the hourly data for the week of the 11th–17th day of each month. Note, the ways of data provided are different between the two models, i.e., data arrive in sequence for the online SVM model and in batch for the conventional SVM model.

Here, the mean absolute error (MAE), root mean squared error (RMSE) and Willmott's index of agreement (WIA, which denotes the close degree between the computed results and original data) are used as measurements of derivation between observed and predicted values. They are defined in the following, respectively:

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|(a_i - t_i)\right|, \tag{9}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(a_i - t_i)^2}, \tag{10}$$

$$\mathrm{WIA} = \frac{\sum_{i=1}^{n}(a_i - t_i)^2}{\sum_{i=1}^{n}(|a_i'| + |t_i'|)^2}, \tag{11}$$

where

$$a_i' = a_i - \bar{a}, t_i' = t_i - \bar{a}, \bar{a} = \frac{1}{n}\sum_{i=1}^{n} a_i, \tag{12}$$

and a_i and t_i denote the predicted result and measured value, respectively.

3.2. Performance of the online SVM model

Taking RSP levels as training examples, we compare the recovery performances between the online SVM and the conventional SVM models on the training set (i.e., 240 data of 1–10 December 2000) and the testing set (i.e., 168 data of 11–17 December 2000). Fig. 1 illustrates the recovery performances on the training data using both methods. It can be observed that for the online SVM model the predicted results are identical to the original data and the maximum deviation is 73.509 μg/m^3, while for the conventional SVM model more deviating points are observed and the maximum deviation is 75.9565 μg/m^3. The MAE, RMSE and WIA are 10.4466 μg/m^3, 14.4858 μg/m^3 and 0.9236 for the online SVM model, and 12.8997 μg/m^3, 17.1204 μg/m^3 and 0.8908 for the conventional SVM model, respectively. In general, both models show good recovery performances on the training data except for some individual deviating points observed, more in the conventional SVM model than in the online SVM model.

The comparisons of prediction results on the testing data and residual error between the two models are shown in

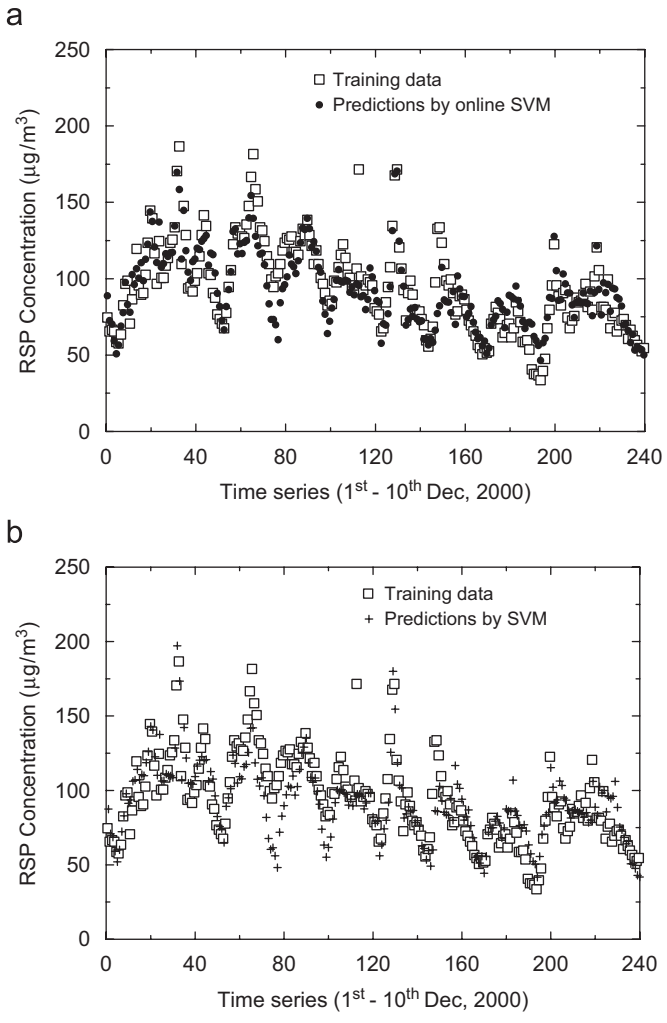

Fig. 1. Recovery performances of two models on training data.

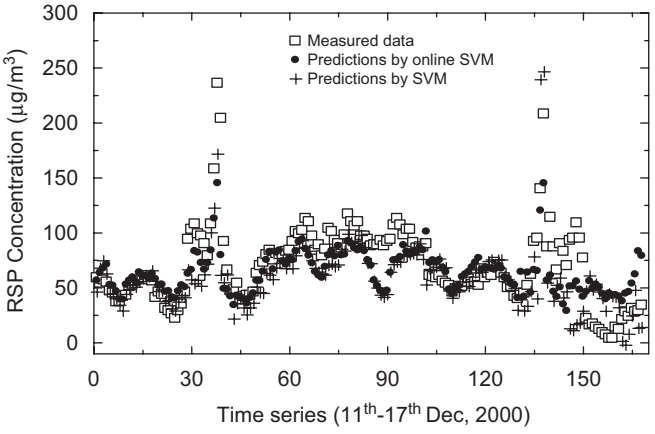

Fig. 2. Predictions of two models on testing data.

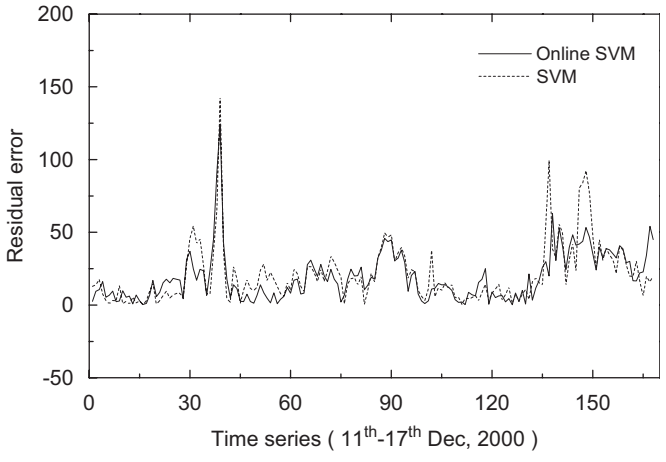

Fig. 3. Residual errors of two models on testing data.

Table 1
Comparisons of two models on training and testing stages

	Maximum deviation ($\mu g/m^3$)		MAE ($\mu g/m^3$)		RMSE ($\mu g/m^3$)		WIA	
	Online SVM	SVM	Online SVM	SVM	Online SVM	SVM	Online SVM	SVM
Training data	73.5090	75.9565	10.4466	12.8997	14.4858	17.1204	0.9236	0.8908
Testing data	124.3689	142.6556	19.2902	20.9723	25.8993	29.3482	0.7880	0.7616

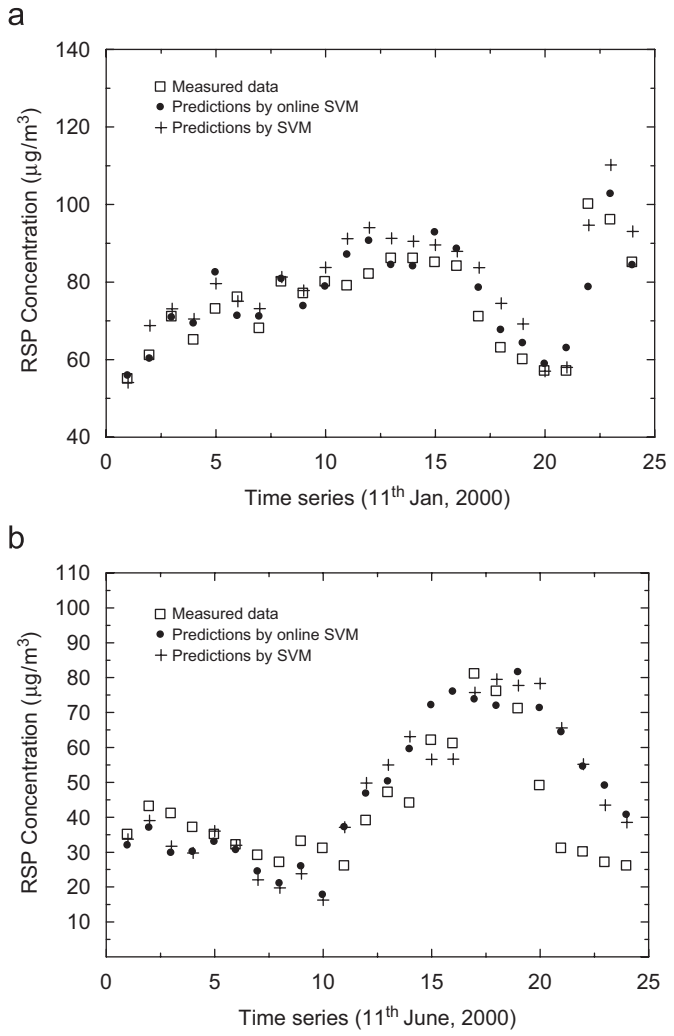

Fig. 4. Predictions of two models for 24-h period: (a) in January and (b) in June.

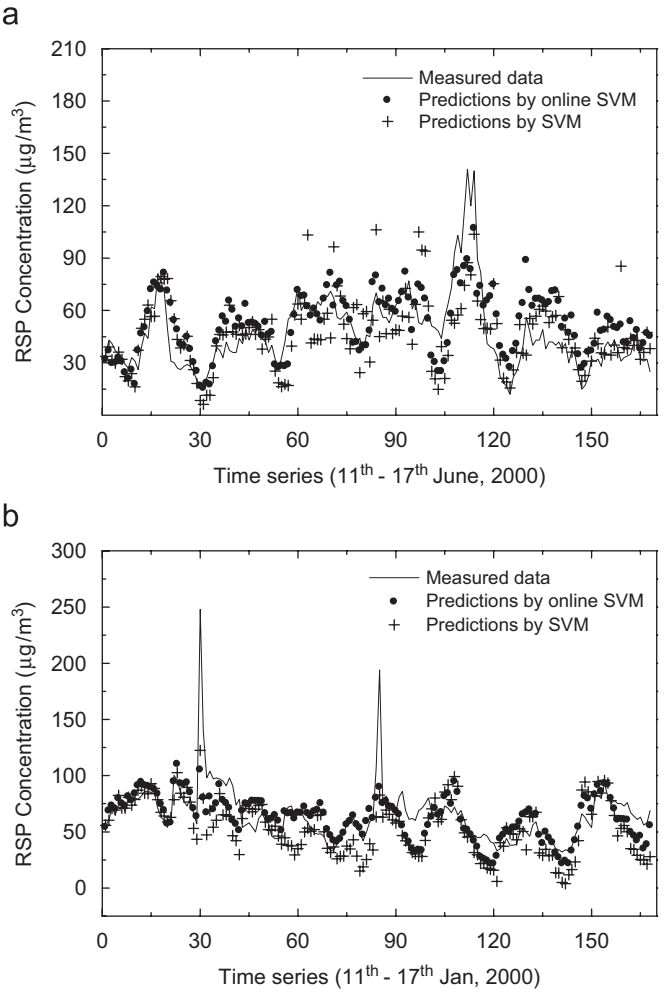

Fig. 5. Predictions of two models for 1-week period: (a) in January and (b) in June.

Figs. 2 and 3. It can be observed from Fig. 2 that, except for a few deviation points, both models present good performances on simulating the testing data. The maximum deviations are 124.3689 $\mu g/m^3$ for the online SVM model and 142.6556 $\mu g/m^3$ for the conventional SVM model. The MAE, RMSE and WIA are 19.2902 $\mu g/m^3$ and 25.8993 $\mu g/m^3$ and 0.7880 for the former, and 20.9723 $\mu g/m^3$, 29.3482 $\mu g/m^3$ and 0.7616 for the latter. Either for the individual case or for the average case, the online SVM model shows better prediction performance than the conventional SVM model. Fig. 3 also demonstrates the same conclusion. Although, for the two models, most computing errors keep within a small range (no greater than 60 $\mu g/m^3$), the error produced by the conventional SVM model increases sharply at several testing stages, while the error created by the online SVM model is increased only at one testing stage. Almost always the residual error of the online SVM model is lower than that of the conventional SVM model. Hence, it can be concluded that the online SVM model has better prediction performance than the conventional SVM model on the testing process. Table 1 lists the comparisons of predicting performance between the online SVM model and the conventional SVM model on training and testing stages.

3.3. Predictions of pollutant levels in different time series

The robustness and tolerance of both online SVM and conventional SVM models are inspected and discussed under the impact of meteorological factors such as temperature, humidity, wind speed and direction, and solar condition in different seasons. Figs. 4 and 5 compare the RSP concentration levels predicted by the two models with two time periods, i.e., 24 h and 1-week advancing, in January and June of 2000. It can be seen that, for the 24 h period, both models produce generally good results for the selected months, but the results produced by the online SVM model are slightly closer to the measured data than those by the conventional SVM model (see Figs. 4a and b), while for 1-week period the online SVM model expresses great advantages over the conventional SVM model. The predictions produced by the online SVM model are generally close to the measured data in both months. The maximum absolute errors of the online SVM model are 125.6623 μg/m^3 in January and 51.3749 μg/m^3 in June. The results created by the conventional SVM model fluctuate and, at certain points, deviate from the measured data points in a great range (see Fig. 5b). The maximum absolute errors by the conventional SVM model are 142.8467 μg/m^3 in January and 54.3749 μg/m^3 in June. The online SVM model performs better than the conventional SVM model does either for a special case or for an average case. Hence, it can be concluded that, although the impact of meteorological variables exists, the online SVM model still possesses superior advantages to the conventional SVM model and it can produce good prediction performance due to its special feature of dynamic optimal modeling.

Considering the characteristics of each pollutant, e.g., accumulation of RSP matter, physical and chemical complexity of SO_2 and ON_x, etc., the prediction performance of the online SVM model can be further verified by forecasting the other two pollutant levels, i.e., SO_2 and NO_x. Figs. 6 and 7 describe the predictions of hourly SO_2 and ON_x levels in 24-h and 1-week advancing time series in January and June of 2000. It is observed that, for the 24-h

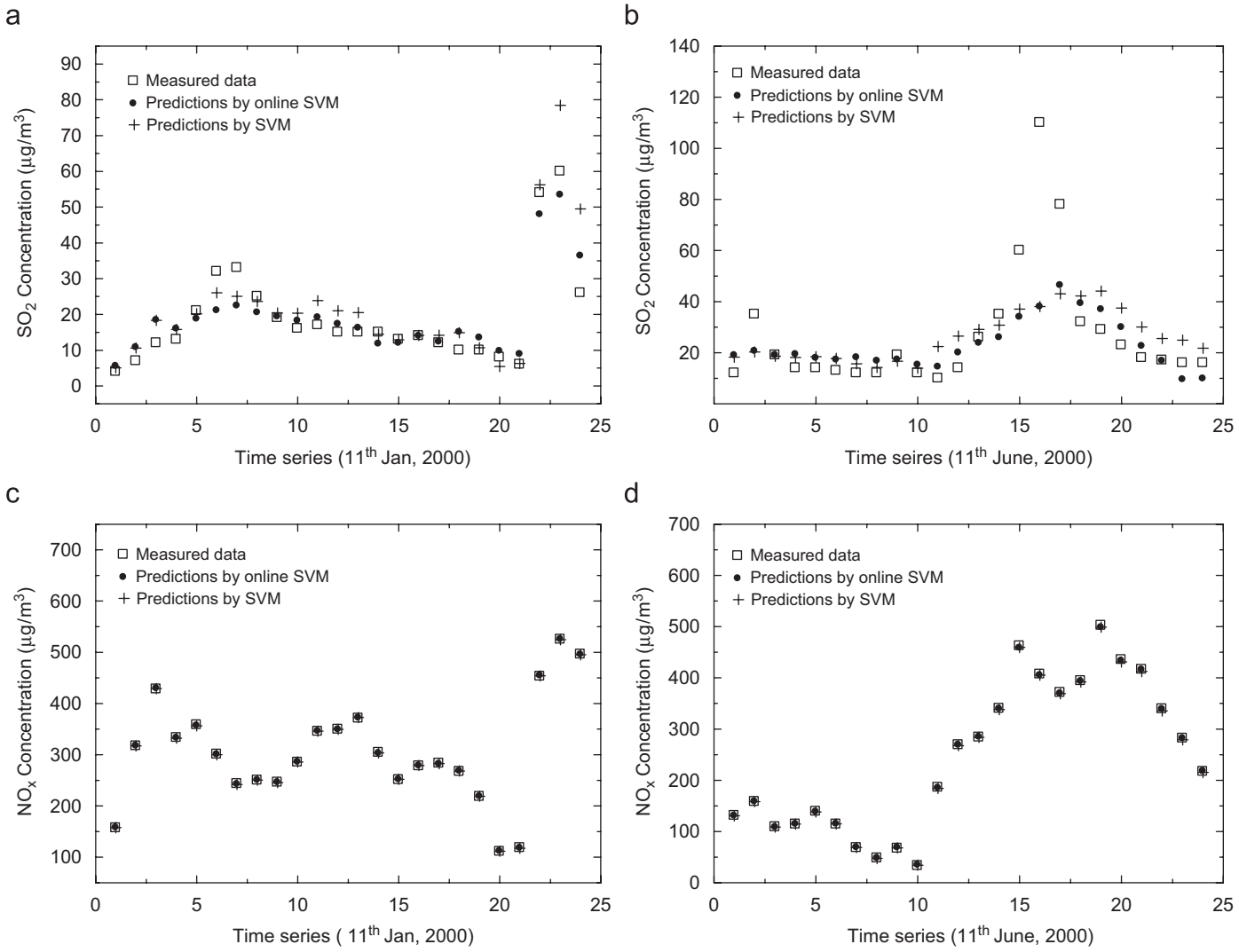

Fig. 6. Prediction comparison between two models for 24-h period: (a, b) for SO_2 and (c, d) for NO_x.

period, both models produce good predicting results for SO_2 and NO_x in the selected months. The two models perform almost ideally for predicting NO_x level (see Figs. 6c and d), but for predicting SO_2 level the online SVM model is slight by better than the conventional SVM model (see Figs. 6a and b), although for the 1-week period, the two models still perform well for predicting NO_x level (see Fig. 7c and d) in the selected two months, but they behave differently in predicting SO_2 level (see Figs. 7a and b) in the selected months. Intuitionist observations illustrate that the predictions generated by the online SVM model are much better than those produced by the conventional SVM model in both months. Especially in June, more prediction results by the conventional SVM model are away from the measured points, but only individual prediction results by the online SVM model are deviated from the measured ones. Both the maximum absolute error and MAE produced by the online SVM model are smaller than those obtained by the conventional SVM model. Hence, the same conclusion, that the online SVM model possesses better prediction performance than the conventional SVM model, can be obtained.

Table 2 shows predicting error comparisons between the online SVM model and the conventional SVM models for three pollutants in 24-h and 1-week time series. It can be seen that, for three pollutants, both MAE and RMSE produced by the online SVM model are smaller than those created by the conventional SVM model in the two selected months, except in one case (labeled in italics, i.e., MAE of RSP in June for 24-h advancing prediction), while for WIA the values by the online SVM model are greater than those by the conventional SVM model for all the prediction cases. Additionally, the prediction results in 24-h advancing time series are better than those in 1-week advancing time series for both models except for an individual case (labeled in bold, i.e., MAE of SO_2 for 24-h prediction is greater than that for 1-week prediction in June). For the prediction performance of the two models in different seasons, it is better in January than in June for 24-h predictions, while it is contrary for 1-week predictions, i.e., the performance is better in June than in January except for a few special cases (labeled by underlines).

Based on the above experiments, it can be concluded that the online SVM model is superior to the conventional SVM model, and possesses good, robust predicting performance.

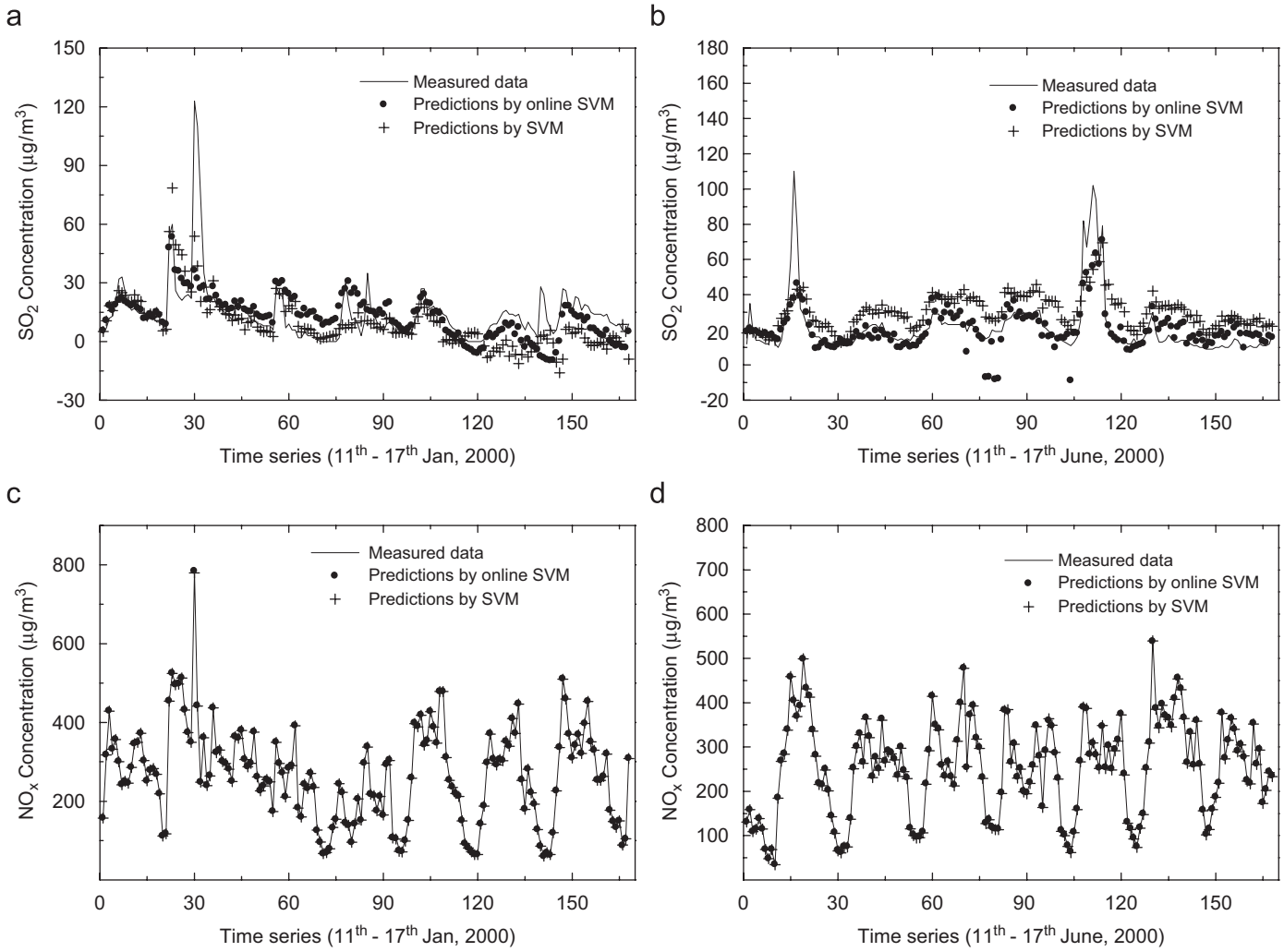

Fig. 7. Prediction comparison between two models for 1-week period: (a, b) for SO_2 and (c, d) for NO_x.

Table 2
Comparisons of two models in predicting performance

Pollutant	Time series	Month	MAE		RMSE		WIA	
			Online SVM	SVM	Online SVM	SVM	Online SVM	SVM
RSP	24-h	January	4.7033	6.1880	6.4105	7.6586	0.9194	0.9094
		June	***10.9513***	***10.5630***	13.4547	13.6678	0.8517	0.8431
	1-week	January	12.9691	18.1527	20.1765	24.7101	0.7506	0.7209
		June	12.8602	14.0857	15.8035	18.3700	0.8267	0.7679
NO_x	24-h	January	0.6862	0.8018	0.7776	0.8631	1.0000	1.0000
		June	1.0506	1.6267	1.4490	1.9704	1.0000	1.0000
	1-week	January	2.3142	2.4202	2.6573	2.8491	0.9999	0.9999
		June	1.2703	2.4104	1.8536	2.6760	0.9999	0.9999
SO_2	24-h	January	3.8305	4.4683	4.9501	7.1658	0.9589	0.9421
		June	**10.0673**	11.7641	17.8720	18.8042	0.6798	0.6270
	1-week	January	8.0950	**8.8526**	12.9677	**13.4863**	0.7490	0.7094
		June	**6.337**	**13.0626**	10.9011	**15.3413**	0.6818	0.6331

4. Conclusions

Forecasting of air pollutant trends has received much attention in recent years, and the requirement of the online predicting models is practical in real-world applications. This paper develops an online SVM model to predict air quality parameter concentration levels, and it can provide promising prediction results. Compared with the conventional SVM model, it can not only receive data in sequence and determine dynamically the optimal prediction model, possess a good prediction performance as well. Another outstanding advantage of the presented method is that the optimal prediction model can be determined before SVM learning unlike in general prediction models. (e.g., For a one-hidden-layer neural network, it needs to predefine the number of hidden nodes and the type of active function. Even these terms can be set in advance through some heuristic approaches such as genetic algorithm, simulating algorithm etc. There is no way at hand to test whether the so-obtained network model is optimal. The comparison between SVM and NN for air quality parameter forecasting can be found in Ref. [6]. The obtained SVM model can lead to high prediction performance; therefore, the applications of the online SVM method in an environmental aspect is a good, interesting attempt, and it may be worthy to test its value in more areas. Additionally, the computational problem in the proposed approach with the numerical optimization in a high-dimensional space may suffer from the course of dimensionality. How to solve this problem will be our future research work.

Acknowledgments

The work described in this paper was partially supported by the National Natural Science Foundation of China (no. 60673095, 70471003), Hi-Tech R&D (863) Program (No. 2007AA01Z165), Program for New Century Excellent Talents in University (NCET), Project for Young Learned Leader, Program for Science and Technology Development in University (No. 200611001), and Program for Selective Science and Technology Development Foundation for Returned Overseas of Shanxi Province. The provision of original data from the Environmental Protection Department, HKEPT, is also appreciated.

References

[1] M. Boznar, M. Lesjak, P. Mlakar, A neural netwok-based method for short-term predictions of ambient SO_2 concentrations in highly polluted industrial areas of complex terrain, Atmos. Environ. 27 (2) (1993) 221–230.
[2] A.C. Comrie, Comparing neural networks and regression models for ozone forecasting, J. Air Waste Manage. 47 (1997) 653–663.
[3] G. Nunnari, L. Bertucco, Modeling air pollution time-series by using wavelet function and genetic algorithms, in: Proceedings of the International Conference on Artificial Neural Network and Genetic Algorithms, Prague, 2001, pp. 489-492.
[4] W. Z. Lu, W. J. Wang, H. Y. Fan, A. Y. T. Leung, Z. B. Xu, S. M. Lo, Air pollutant parameter forecasting using support vector machines, in: IJCNN'2002, vol. 1–3, IEEE, 2002, 630–635.
[5] W.Z. Lu, W.J. Wang, Potential assessment of the support vector Machine method in forecasting ambient air pollutant trends, Chemosphere 59 (5) (2005) 693–701.
[6] W.J. Wang, Z.B. Xu, W.Z. Lu, Three improved neural network models for air quality forecasting, Eng. Comput. 20 (2) (2003) 192–210.
[7] M.W. Gardner, S. R . Dorling, Artificial neural networks (the multilayer feed-forward neural networks)— a review of applications in the atmospheric science, Atmos. Environ. 30 (14/15) (1998) 2627–2636.
[8] W.Z. Lu, W.J. Wang, H.Y. Fan, A.Y.T. Leung, S.M. Lo, Z.B. Xu, J.C.K. Wong, Prediction of pollutant levels in Causeway Bay area in Hong Kong using an improved neural network model, ASCE J. Environ. Eng. 128 (12) (2002) 1146–1157.
[9] W.Z. Lu, W.J. Wang, H.Y. Fan, A.Y.T. Leung, Z.B. Xu, S.M. Lo, J.C.K. Wong, Using improved neural network model to analyze RSP, NO_x and NO_2 levels in urban air in Mong Kok, Hong Kong, Environ. Monit. Assess. 87 (2003) 235–254.
[10] P. Perez, A. Trier, J. Reyes, Prediction of PM2.5 concentrations several hours in advance using neural networks in Santiago Chile, Atmos. Environ. 34 (2000) 1189–1196.
[11] S.L. Reith, D.R. Gomez, L.E. Dawidowski, Artificial neural network for the identification of unknown air pollution sources, Atmos. Environ. 33 (1999) 3045–3052.

[12] M. Roadknight, G.R. Balls, G.E. Mills, B.D. Palmer_Brown, Modelling complex environmental data, IEEE T. Neural Netw. 8 (4) (1997) 852–861.
[13] X.H. Song, P.K. Hopke, Solving the chemical mass balance problem using an artificial neural network, Environ. Sci. and Technol. 30 (2) (1996) 531–535.
[14] W.J. Wang, W.Z. Lu, X.K. Wang, A.Y.T. Leung, Prediction of maximum daily ozone level using combined neural network and statistical characteristics, Environ. Int. 29 (5) (2003) 555–562.
[15] J. Yi, R. Prybutok, A neural network model forecasting for prediction of daily maximum ozone concentration in an industrialized urban area, Environ. Pollut. 92 (3) (1996) 349–357.
[16] V. Vapnik, The Nature of Statistical Learning Theory, Springer, New York, 1995.
[17] V. Vapnik, Statistical Learning Theory, Wiley, New York, 1998.
[18] C.J.C. Burges, A tutorial on support vector machines for pattern recognition, Data Min. Know. Disc. 2 (2) (1998) 121–167.
[19] H. Drucker, C.J.C. Burges, L. Kaufman, A. Smola, V. Vapnik, Support vector regression machines, in: M. Mozer, M. Jordan, J. Petsche (Eds.), Advances in Neural Information Processing Systems, vol. 9, MIT, Cambridge, MA, 1997, pp. 155–161.
[20] D. Mattera, S. haykin, Support vector machines for dynamic reconstruction of a chaotic system, in: B. Schölkopf, C.J.C. Burges, A. Smola (Eds.), Advances in Kernel Methods—Support Vector Learning, MIT, Cambridg, MA, 1999, pp. 211–242.
[21] K.R. Müller, A. Smola, G. Ratsch, B. Scholkopf, J. Kohlmorgen, V. Vapnik, Predicting time series with support vector machines, in: W. Gerstner, A. Germond, M. Hasler, J.D. Micord (Eds.), Artificial Neural Networks, Berlin, 1997, pp. 999–1004.
[22] M. Junshui, P. Simon, Accurate online support vector regression, Neural Comput. 15 (2003) 2683–2703.
[23] M. Martin, Online support vector machine regression (2002) http://www.lsi.upc.es/dept/techreps/html/R02-11.html.

Wenjian Wang received the B.S. degree in computer science from Shanxi University, China, in 1990, the M.S. degree in computer science from Hebei Polytechnic University, China, in 1993, and Ph. D. degree in applied mathematics from Xi'an Jiao Tong University, China, in 2004. She worked as a research assistant at the Department of Building and Construction, The City University of Hong Kong from May 2001 to May 2002. She has been with the Department of Computer Science at Shanxi University since 1993, where she was promoted as Associate Professor in 2000 and as Full Professor in 2004, and now serves as a Ph.D. supervisor in Computer Application Technology and System Engineering. She has published more than 40 academic papers on machine learning, computational intelligence, and data mining. Her current research interests include neural networks, support vector machines, machine learning theory and environmental computations.

Changqian Men received the B.S. and M.S. degrees in computer science and technology from Shanxi University, China, in 2003 and 2006, respectively. Now he is a Ph.D. student in Key Laboratory of Computational Intelligence & Chinese Information Processing of Ministry of Education at Shanxi University, China. His current research interests include support vector machines, pattern recognition and machine learning theory.

Weizhen Lu received B.Sc. and M.Eng. in Xi'an Jiaotong University in 1982 and 1985. From 1990 to 1993, she worked as a research scientist at the National Engineering Laboratory, UK. She joined a consultancy project of multi-phase flow for UK Oil Companies. In 1993, she worked as a Research Officer at the De Montfort University, UK. She finished her Ph.D. project in October 1995, took Lecturer post at the Dept. of Building Services Engineering, Hong Kong Polytechnic University in November 1995 and worked actively in teaching and research. She joined the City University of Hong Kong in 1996 as a research Fellow, was promoted as Assistant Professor in October 1998 and Associate Professor in February 2004. Her main research interests include air quality, air infiltration, HVAC system, wind effect on high-rise buildings, application of Computational Fluid Dynamics and computation intelligence in various engineering disciplines including building engineering, chemical & environmental engineering, mechanical engineering, power engineering, wind engineering, etc.