CHEMICAL ENGINEERING TRANSACTIONS
VOL. 51, 2016
A publication of AIDIC
The Italian Association of Chemical Engineering
Online at www.aidic.it/cet
Guest Editors: Tichun Wang, Hongyang Zhang, Lei Tian
Copyright © 2016, AIDIC Servizi S.r.l.,
ISBN 978-88-95608-43-3; **ISSN** 2283-9216
DOI: 10.3303/CET1651057

Application of an Improved SVM Algorithm for Wireless Sensor Networks in the Prediction of Air Pollution

Jianfen Liu*, Qiming Wang

Collage of Computer Science and Technology, Pingdingshan University, Pingdingshan, Henan, China
2094991521@qq.com

As is known to all, human's daily life can not be separated from the atmospheric environment, and the quality of human life is directly affected by the quality of the atmospheric environment. In recent decades, the rapid development of China's economy, industry, transportation and other industries cause that concentration of the hazardous material in the air is much higher than the identification standard. It not only affects everyone's life and production, but also hinders the development of the whole society. Based on this, with the help of the wireless sensor technology, this paper builds a wireless sensor networks for the acquisition and transmission of various air pollutants data. This network can effectively monitor the current existence of most of the pollutant gas. Secondly, according to the received data, this paper introduces the support vector regression model based on ant colony optimization to forecast the concentration of pollutants in the air. Because the prediction accuracy of support vector machine is largely determined by the selection of parameters, the selection of training parameters is optimized by using ant colony algorithm in order to get the optimized support vector machine prediction model. At last, we use the modified model to predict the concentration of PM2.5 with nonlinear data. Experimental results show that the proposed improved support vector prediction algorithm is effective, and is significantly better than the other two prediction algorithm.

1. Introduction

As is known to all, human's daily life can not be separated from the atmospheric environment, and the quality of human life is directly affected by the quality of the atmospheric environment. In recent decades, the rapid development of China's economy, industry, transportation and other industries cause that concentration of the hazardous material in the air is much higher than the identification standard It not only affects everyone's life and production, but also hinders the development of the whole society(Wu, 2010; Hao et al., 2012; Chen and Li, 2013).

The wireless sensor network technology is introduced in the field of environmental monitoring, and we can play the unique advantages of the wireless sensor network technology to realize the wireless and automation of environmental monitoring, thus greatly improves the environmental monitoring technical level. The spatial interpolation method is a fast and effective method to reflect the whole area pollution situation through the limited monitoring points. It has been widely used in the field of environment (Tian L et al., 2011). However, the research objects are mostly soil element (Zhang et al., 2010) and meteorological element (Yuan F et al.(2008)), and the spatial interpolation method for atmospheric particulate matter is less. This is mainly because the spatial interpolation requires a certain sample. For a long time, due to the monitoring system is not perfect, the high cost of air monitoring equipment and other objective reasons, it is difficult to obtain a complete time series of multiple points simultaneously monitoring of atmospheric particulate matter concentration. Reddy A et al.(2009) propose a smart sensor monitoring system for the study of air quality and comfort of human life, and they use it to monitor the indoor air quality, and achieved good results. Mamidisetty K K et al.(2009) put forward a set of environmental monitoring system which is dedicated to monitor the indoor air quality of subway. Wang Y G et al.(2010) develop a distributed environmental monitoring system, which is also used to monitor the indoor air quality of health care centre.

Please cite this article as: Liu J.F., Wang Q.M., 2016, Application of an improved svm algorithm for wireless sensor networks in the prediction of air pollution, Chemical Engineering Transactions, 51, 337-342 DOI:10.3303/CET1651057

2. Architecture and basic knowledge of Wireless Sensor Networks

Wireless sensor network consists of a large number of sensor nodes, and the network structure is shown in figure 1. They do not need to be processed or pre positioned, and can be densely scattered in the monitored area. Therefore, it is required that the sensor nodes should have the characteristics of self organizing.

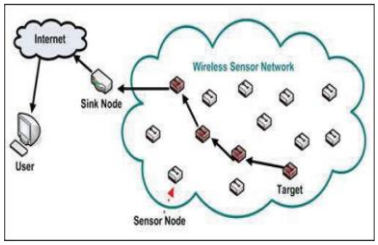

Figure 1: The Wireless sensor network

Node is the basic unit of wireless sensor network, wireless sensor network consists of a large number of small and low cost sensor nodes that have the function with wireless communication and sensor data processing. Wireless sensor node is responsible for information collection and pre-treatment, and is generally composed of the sensor unit, processing unit, wireless transceiver unit and power management unit. Its structure as shown in figure 2.

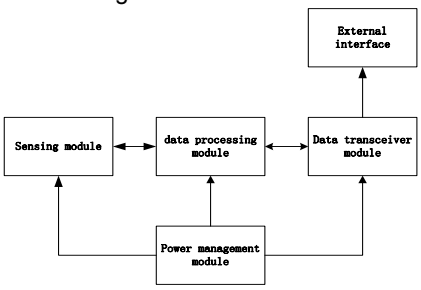

Figure 2: Sensor node hardware structure

1. The sensing unit is responsible for collecting the information in the monitoring area and converting it to digital signal. This paper selects various air quality monitoring sensors, such as semiconductor air quality sensor (TGS2600/TGS260), dust sensor, carbon monoxide sensor (CO-BF), electrochemical sulphur dioxide sensor (SO2-A), and electrochemical NO2 gas sensor (NO2-A1) etc..
2. The processing unit is composed of a processor and a memory, which stores and processes the data collected by the sensor. This paper uses the AVR series microcontroller, which uses the RISC structure, drawing the advantages of PIC and 8051 MCU, and has a wealth of internal resources and external interface. In the aspect of integration, almost all the key components are integrated in the interior.

3. An improved SVM algorithm in WSN

3.1 Basic knowledge of support vector machines

For the meteorological data collected by wireless sensor which is introduced in Section 2, we use an improved support vector machine model for processing, so as to predict the concentration of air pollutants. Next, we will introduce the basic knowledge of support vector machines. In essence, the support vector machine method is a machine learning method. The learning problem can be briefly summarized as follows: for a given data set $T=\{(x_1,y_1),(x_2,y_2),\cdots,(x_n,y_n)\}$, we get a function on the real space *f(x,a)*, which makes the relationship (x,y) in data set T. The *f(x,a)* which is the most consistent with the actual situation is used as a support vector machine learning goal. The optimization problem of LSSVR is as follows:

$$f(x)=\omega^T\varphi(x)+b \tag{1}$$

In formula (1), $\varphi(x)$ is a nonlinear transformation function, ω is the linear regression coefficient, and b is the deviation. At this point, the optimization problem is transformed into the following form:

$$\begin{cases} \min(\frac{1}{2}(\omega^T\omega+\gamma\sum_{i=1}^{n}\xi_i^2)) \\ s.t. \quad y_i=\omega^T\varphi(x)+b+\xi_i \end{cases}, i=1,2,\cdots,n \tag{2}$$

In formula (2), γ is the regularization parameter; ξ_i is the regression error of sample.
In order to eliminate the constraint, the formula (2) is solved by using the *Lagrange* function.

$$L(\omega,b,\xi,a)=\frac{1}{2}(\omega^T\omega+\gamma\sum_{i=1}^{n}\xi_i^2)-\sum_{i=1}^{n}a_i[\omega^T\varphi(x)+b+\xi_i-y_i] \tag{3}$$

In formula (3), a_i is a *Lagrange* multiplier. According to KKT conditions, we can calculate the partial derivative of ω,b,ξ,a, and make them equal to 0. So, we can get the formula (4).

$$\begin{cases} \omega=\sum_{i=1}^{n}a_i\varphi(x) \\ \sum_{i=1}^{n}a_i=0 \\ a_i=\gamma\xi_i \\ \omega^T\varphi(x)+b+\xi_i-y_i=0 \end{cases} \tag{4}$$

To eliminate ω and ξ in formula (4), the optimization problem is transformed into solving the following linear equation:

$$\begin{bmatrix} 0 & \theta^T \\ \theta & \Omega+\gamma^{-1}I \end{bmatrix}\begin{bmatrix} b \\ a \end{bmatrix}=\begin{bmatrix} 0 \\ y \end{bmatrix} \tag{5}$$

In formula (5), $y=[y_1,y_2,\cdots,y_n]^T$, θ is a diagonal matrix, $a=[a_1,a_2,\cdots,a_n]^T$, Ω is a matrix $\Omega_{ij}=k(x_i,x_j)=\varphi(x_i)^T\varphi(x_j)$ composed of Ω_{ij}, and $k(x_i,x_j)$ is a kernel function. The classification decision function of LSSVM and the kernel function is obtained by calculating the formula (6) and formula (7):

$$f(x)=\sum_{i=1}^{n}a_ik(x,x_i)+b \tag{6}$$

$$k(x_i,x_j)=\exp(-\frac{\|x_i-x_j\|^2}{2\sigma^2}) \tag{7}$$

3.2 Using ant colony algorithm to optimize the parameters of SVM

At this point, the parameters that need to be determined are the regularization parameter γ and kernel function parameter σ, and we use the ant colony algorithm to optimize the above two parameters. Ant colony algorithm is a global search algorithm which combines the positive feedback, distributed computing and greedy heuristic search. It can effectively solve the problem of easy to fall into local extremes. Based on this, we take the minimum mean square error MSE as the objective function *F* of the optimization problem, through the ant colony algorithm to search the optimal objective function value, so as to obtain the optimal parameter set (γ,σ). Its objective function is:

$$\begin{cases} \min F=\min MSE=\frac{1}{m}\sum_{i=1}^{m}(y_i-\hat{y}_i)^2 \\ s.t. \quad \gamma_{\min}\le\gamma\le\gamma_{\max} \\ \sigma_{\min}\le\sigma\le\sigma_{\max} \end{cases} \tag{8}$$

In the formula (8), y_i and $\hat{y}_i$ are the real values of monitoring samples and the predicted values calculated by LSSVR, respectively.

1. The parameter initialization. Set the basic parameters of ant colony algorithm initialization, such as population size *n*, the proportion factor of pheromone update ρ, pheromone heuristic factor τ, etc.. Then, the termination condition of ant colony search operation is determined.
2. The iterative search. First, set the initial time *t=0*, and n ants are located at the starting point. After the search starts, according formula (9) to calculate the transition probability P_{ij}^k of each ant *k(1,2,…,m)* from the layer L_i to the layer L_{i+1}. Then, the roulette wheel method is used to select a node on the L_{i+1} layer, and it is transferred to that node.

$$P_{ij}^k = \begin{cases} \dfrac{\tau_{ij}^{\alpha}(t)\eta_{ij}^{\beta}(t)}{\sum_{s\in A_k}\tau_{is}^{\alpha}(t)\eta_{is}^{\beta}(t)} & s \in A_k \\ 0 & s \notin A_k \end{cases} \tag{9}$$

Where, α represents the relative importance of residual information, β represents the relative importance of expectations, and A_k represents a collection of all the target nodes that may be accessed. In order to avoid repeated visits to the same node, each ant has a list of forbidden *tabu (k)*, which is used to record the nodes that have been visited so far.
3. Update pheromone. After each ant completes the access to all the nodes, it is necessary to update the information in the process. At the same time, the pheromone concentration must be updated.

$$\tau_{ij}(t+n) = \rho_1 \tau_{ij}(t) + \sum_{k=1}^{m} \tilde{\tau}_{ij}^k \tag{10}$$

Where, ρ_1 represents the retention portion of the residue information, $\tilde{\tau}_{ij}^k$ represents the pheromone concentration of residual information on the path of nodes *i* and *j* between the time *t* and *t+n*.
4. To determine the termination conditions. Check whether the algorithm satisfies the iteration termination conditions. If not satisfied, return to Step 2. Otherwise, the algorithm ends, and the outputs the optimal parameter set (γ,σ). The specific process is shown in Figure 3:

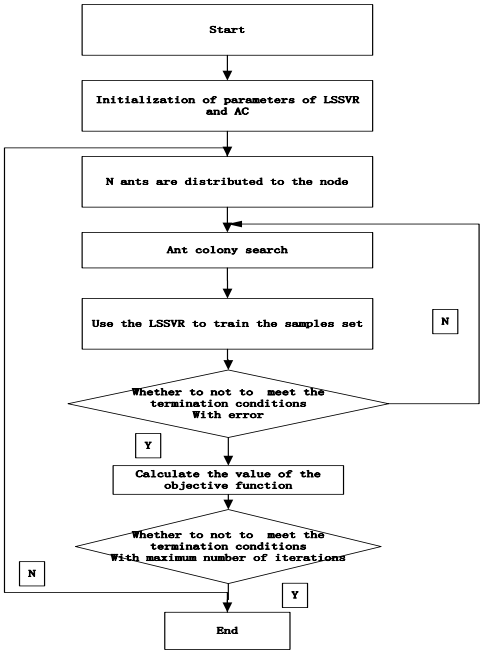

Figure 3: Specific process of the GA algorithm

4. Simulation experiment and result analysis

4.1 Environment and parameters

In this experiment, we use the monitoring data of PM2.5 of Beijing city in 2015 as an example. Air detection sensors are distributed in the city to collect and summarize these data which are used as the training set of

our algorithm. Then, we predict the concentration of PM2.5 of the next 30 days, so as to achieve the purpose of early warning of air pollutants. The parameters of the model are set as follows: the range of γ is $(0.01,100)$, the range of σ is (0.01,100), the ant population size is n=150, the maximum iteration number is N_{max}=500, and the proportion factors of pheromone update are ρ=0.75, α=2 and α=2.

4.2 The experiment steps and the result analysis

Firstly, we show the monitoring data of PM2.5 in 2015. As can be seen from Figure 4, the PM2.5 value of Beijing city is higher in spring and winter, but lower in summer and autumn. This shows that the weather conditions are the main factors that affecting the PM2.5 seasonal changes.

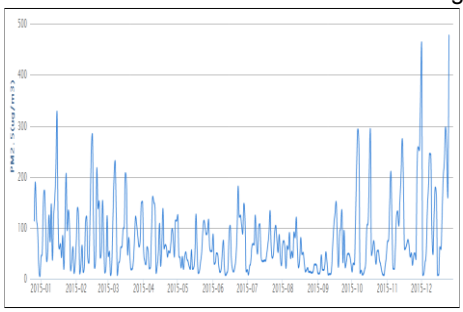

Figure 4: PM2.5 monitoring data of Beijing city in 2015

Secondly, five kinds of meteorological factors are selected in this study. They are temperature (T), relative humidity (RH), pressure (P), wind speed (F), rainfall (R). From Figure 5 we can see that the annual temperature has a close relationship with PM2.5. In spring and winter, when the temperature is low, the PM2.5 value is more than 300. And when the temperature is higher, the PM2.5 value is less than 100. This shows that the air quality is excellent.

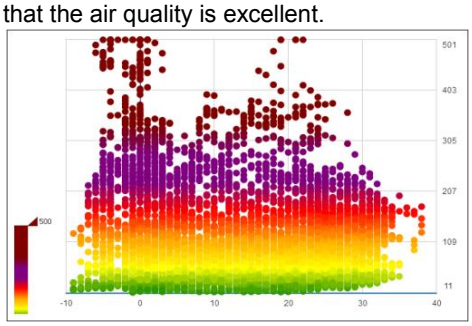

Figure 5: Relationship between temperature and PM2.5

From Figure 6 we can see that the annual humidity has a close relationship with PM2.5. At the top right of the figure, there are a lot of brown spots. This shows that when the air is more humid, the probability of a higher concentration of PM2.5 is very large. On the contrary, when the air is dry, the air quality is usually better. In addition, the air pressure, wind speed and rainfall also have a very close relationship with the PM2.5. Therefore, this paper selects these five factors as sample data, and inputs them to the improved support vector regression model.

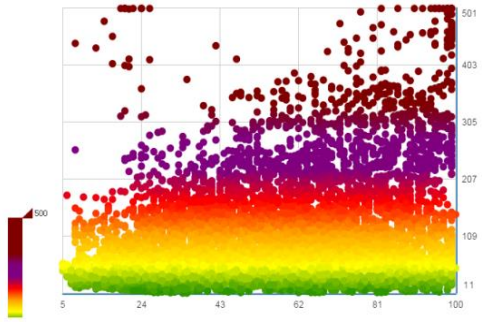

Figure 6: Relationship between humidity and PM2.5

Thirdly, we use three different methods to predict the value of PM2.5. Then, by calculating the percentage of successful times and the number of samples, we can judge the success rate of the prediction algorithm. The improved SVM algorithm, BP neural network method and grey prediction method are analyzed and compared in the case of error threshold value between 1 $\mu g/m^3$ and 5$\mu g/m^3$. Figure 7 shows the average success rate of the three prediction algorithms that predict the PM2.5 of Beijing in January 2016.

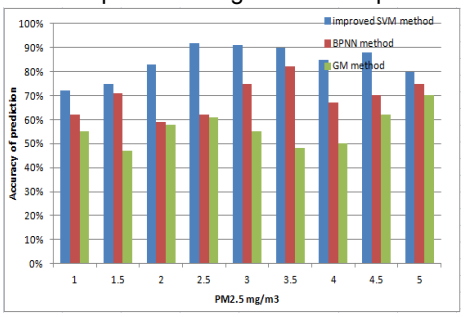

Figure 7: Comparison of the average success rate of the three prediction methods

As can be seen from Figure 7, the improved support vector prediction algorithm is significantly better than the other two algorithms. It shows that the optimization of continuous space based on ant colony algorithm is very important for the important parameters (γ,σ) which affect the prediction accuracy of support vector machines. It avoids the defects of the experience when the parameters are chosen, and the prediction accuracy is obviously improved.

5. Conclusion

This paper builds a wireless sensor networks for the acquisition and transmission of various air pollutants data. This network can effectively monitor the current existence of most of the pollutant gas. Secondly, according to the received data, this paper introduces the support vector regression model based on ant colony optimization to forecast the concentration of pollutants in the air. Because the prediction accuracy of support vector machine is largely determined by the selection of parameters, the selection of training parameters is optimized by using ant colony algorithm in order to get the optimized support vector machine prediction model. At last, we use the modified model to predict the concentration of PM2.5 with nonlinear data. Experimental results show that the proposed improved support vector prediction algorithm is effective, and is significantly better than the other two prediction algorithm.

References

Chen J.P., Li Z.J., 2013, The situation and existing problems of air pollution control in China and some policy suggestions [J]. Development research, 10: 4-14.

Hao J.M., Cheng Z., Wang S.X., 2012, Research on the present situation and control measures of atmospheric environmental pollution in China [J]. Environmental protection, 09: 17-20.

Mamidisetty K.K., Duan M.L., Sastry S., 2009, Multipath dissemination in regular mesh topologies[J]. IEEE Transactions on Parallel and Distributed Systems, 20(8): 1188-1201. DOI: 10.1109/TPDS.2008.164

Reddy A., Kumar A., Janakiram D., 2009, Wireless sensor network operating systems: a survey[J]. International journal of sensor networks, 5(4): 236-255. DOI: 10.1504/IJSNET.2009.027631

Tian L., Dong D.M., Wei Q, et al., 2011, Comparison of three spatial interpolation methods for statistical processing of lead monitoring data in road dus t[J]. Jilin University Sci Ed, 49(5): 964-970.

Wang Y.G., Yin X.G., You D., 2010, Application of wireless sensor networks in Smart Grid [J]. Power System Technology, 34(5): 7-11.

Wu M.Y., 2010, Suggestions on Revision of air pollution control law in China[J]. Knowledge economy, 16: 31-32.

Yuan F., Bai X.Y., Zhou T.F., et al., 2008, Comparison between methods for interpolation of studying spatial distribution of elements: a case study of soil heavy metals in Tongling area [J]. Earth Sci Front, 15(5): 103-110.

Zhang T.C., Chang Q.R., Liu J., 2010, Comparison of spatial interpolation methods for soil nutrient elements: a case study of Yuyang County [J]. Agric Res Arid Areas, 28(2): 177-180.