Water Quality Monitoring and Evaluation System Based on Wireless Sensor Network and LS-SVM Method

Wei ZHANG [1,2], **Jian CAO** [2], **Yan YANG** [1,3], **Di WU** [1], **Yong HE** [1*]

1. College of Biosystems Engineering and Food Science，Zhejiang University, Hangzhou, China
2. Mechanical & Electrical Engineering College, Jiaxing University, Jiaxing, China
3. Mathematical Science College, Guangxi Teachers Education University, Nanning, China
Email: yhe@zju.edu.cn

Abstract: Water quality monitoring and evaluation are important practices for water resources management. And the major concerns are remote and low-cost information transmission technology and accurate water quality evaluation strategy. Taking these concerns into consideration, a remote monitoring and evaluation system based on wireless sensor networks (WSN) and LS-SVM method was designed in this paper. The ZigBee-WSN was used to transmit the water quality status information including dissolved oxygen, permanganate index, BOD5, ammonia nitrogen, total phosphorus and total nitrogen And the LS-SVM is selected as water quality level evaluation method. The test results show that the system has many advantages, such as economic, practical, with high communication reliability, fast and accurate evaluation accuracy.

Keyword: water quality; wireless sensor networks; least squares-support vector machine; ZigBee

1 Introduction

As water quality perturbations related to escalating human population growth and industry pressure continue to increase (National Research Council, 2000; World Resources Institute, 2003), effective water quality monitoring and evaluation has become critical for water resources management programs (H.B.Glasgow et al., 2004). Three kinds of technical conditions are included in water-quality determination and evaluation: (1) accurately collecting the water quality status, (2) long-term and low-cost signal transmission technology; (3) accurate water quality evaluation strategy.

For decades, water quality monitoring and evaluation have depended on costly, time- and labor-intensive on-site sampling and data collection, and transport to laboratories for evaluation (Glasgow et al., 2004). While these research and monitoring efforts are episodically intensive, they generally have been limited on temporal and spatial scales. In recent years, in order to solve these problems, advanced in-situ sensors and remote acquisition technologies have been researched. In water-quality determination study field, many index such as dissolved oxygen, permanganate index, BOD_5, ammonia nitrogen, total phosphorus and total nitrogen, can be quickly detected (Kevin G. et al., 2006; Glasgow et al., 2004). In the remote data remote study field, GPRS-based technology has been used widely (Cao Jian et al., 2009), while the communication cost is high. At the same time, it is hard to build precise mathematical model based on multi-sensor information coming from all kinds of water quality sensor, so water quality evaluation method based multi-sensor data fusion is still full of ambiguity and inaccuracy.

The objective of this study was to use wireless sensor networks (WSN) and Least squares-support vector machine (LS-SVM) method to design a economical and accurate water quality monitoring and evaluation system, and the structure, work principle and data-analysis algorithm were introduced in this paper.

2 System General Design

As shown in Figure 1, the system consists of monitoring area, base station, remote monitor center, data server, and transmission network. ZigBee-technology-based network was used in information transmission system, and the nodes in ZigBee network included sensor node, controller node and coordinator node. And *ad hoc* network was selected in the ZigBee network, which has advantages for water monitoring applications, because the mobility and self-configuration are more suitable for a distributed network compared with a wireless local area network (WLAN) (Yunseop Kim et al., 2008).

Water quality detectors and sensor nodes are distributed in the monitoring area. The water quality information, including dissolved oxygen, permanganate index, BOD_5, ammonia nitrogen, total phosphorus and total nitrogen, are converged upon to router nodes through *ad hoc* style. Then, the converged information is transmitted to computer connected with coordinator node in base station through *ad hoc* network. Then the base station computer analyzes the sensor information, calculates water quality level and draw water quality evaluation results. All the information can be displayed in base station computer, shared with remote monitor center and data

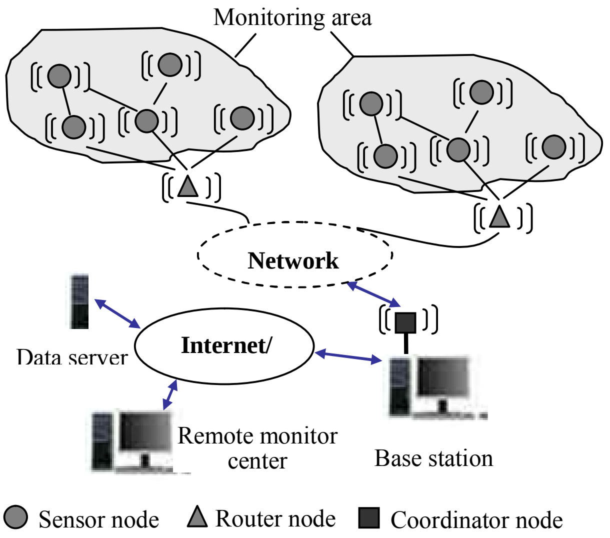

Figure 1. Schematic plan of WSN system

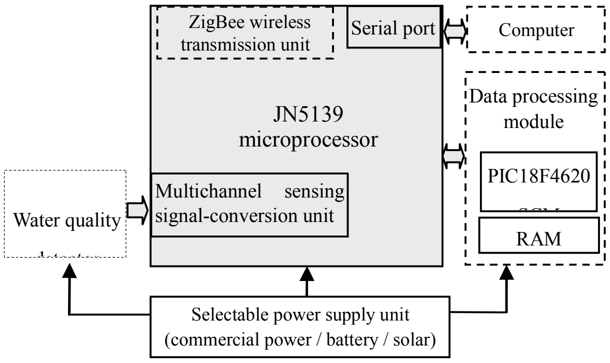

Figure 2. Block diagram of WSN node

server through Internet or GPRS network.

3. Design of WSN Node

The key problem of system hardware is the design of WSN node, which is a kind of miniature embedded system integrating information-processing and communication function (Zhang Wei et al., 2009). In this system, the node was characterized by modularized design consisting of core module and different expansion modules. As illustrated in shaded areas of Fig. 2, the core module was based on JN5139 microprocessor, which was a kind of wireless network processor from Jennic Corporation, and can help users build WSN system based on IEEE802.15.4/ZigBee protocol with saving time and cost (Zhang Wei et al., 2009).

Sensor node consisted of core module and different water quality detectors. The router node and coordinator node consisted of core module and data processing module which was based on PIC18F4620 SCM and used to process more complex data. The power supply unit was divided into three selectable types: commercial power, battery, and solar. i.e., if one kind of water quality detector use commercial power, then the connected sensor node can also use commercial power, while the router node used solar power under normal circumstances.

4. Water Quality Evaluation Model Based on LS-SVM

Least squares-support vector machine (LS-SVM) is a state-of-the-art learning algorithm and capable of dealing with linear and nonlinear multivariate analysis in a relatively fast way. Moreover, SVM is capable of learning in high-dimensional feature space with fewer training data. The model of SVM is shown in Fig. 3, and the standard LS-SVM algorithm was introduced as follows (J. A. K. Suykens et al., 2002; Di Wu, Yong He et al., 2008).

Assume the training set is given as $\{x_k, y_k\}_{K=1}^{N}$, with the input $x_k \in R^N$ and the output $y_k \in R$. The following regression model is constructed using non-linear mapping function $j(\cdot)$, which maps the input data to the higher dimensional feature space:

$$y(x) = w^T j(x) + b$$

where $w \in R^n$ is the weight vector and b is the bias. The new optimization problem is formulated in the case of SRM

$$\min J(w,e) = \frac{1}{2} w^T w + \frac{1}{2} g \sum_{k=1}^{N} e_k^{\,2}$$

subject to the constraints

$$y_k = w^T j(x) + b + e_k, \quad k = 1, \ldots, N$$

where r is the regularization parameter which balances the model's training errors and complexity, and e_k is the random errors. And then, Lagrange function is adopted to solve the following optimization problem

$$L(w,b,e,a) = J(w,e) - \sum_{k=1}^{N} a_k \left\{ w^T j(x_k) + b + e_k - y_k \right\}$$

where a_k is Lagrange multipliers named support value. The solution of the above equation can be obtained by partially differentiating method with respect to each variable.

$$\begin{cases} \dfrac{\partial L}{\partial w} = 0 \rightarrow w = \sum_{k=1}^{N} a_k j(x_k) \\ \dfrac{\partial L}{\partial b} = 0 \rightarrow \sum_{k=1}^{N} a_k = 0 \\ \dfrac{\partial L}{\partial e_k} = 0 \rightarrow a_k = g e_k, \quad k = 1, \ldots, N \\ \dfrac{\partial L}{\partial a_k} = 0 \rightarrow w^T j(x_k) + b + e_k - y_k = 0, \quad k = 1, \ldots, N \end{cases}$$

If the variable w and e is removed, the equation can be rewritten as the linear function group

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

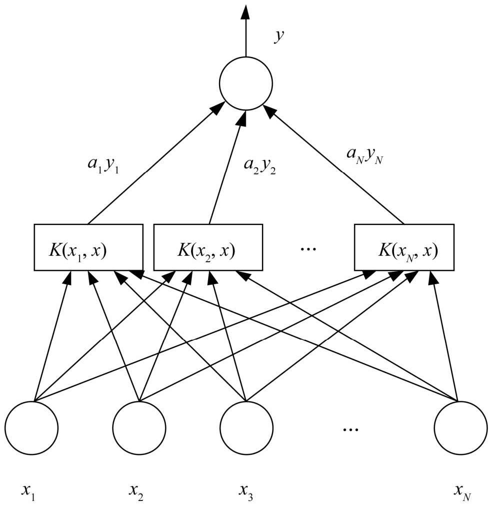

Figure 3. Model of support vector machines

If the variable w and e is removed, the equation can be rewritten as the linear function group

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

with

$$\begin{cases} y=[y_1,\dots,y_N] \\ \vec{1}=[1,\dots,1] \\ \alpha=[\alpha_1,\dots,\alpha_N] \\ \Omega=\{\Omega_{kj}\,|\,k,l=1,\dots,N\} \end{cases}$$

and $\Omega_{kl}=\varphi(x_k)^T\varphi(x_l)=K(x_k,x_l)$, $k,l=1,\dots,N$,

where $K(x_i,x_j)$ is the kernel function, and should follow Mercer's theory. The common examples of kernel function conclude linear, polynomial, radial basis function (RBF) kernel, multi-layer perceptron (MLP) and so on. In the water quality evaluation system, RBF kernel was selected as the kernel function as

$$K(x,x_k)=\exp(-\|x-x_k\|^2/\sigma^2)$$

The LS-SVM model can be obtained as

$$y(x)=\sum_{k=1}^{N}\alpha_k K(x,x_k)+b$$

When using LS-SVM, the determination of the optimal input feature subset, proper kernel function, and the best kernel parameters are three crucial problems need to be solved. In this paper, RBF kernel was used as the kernel function of LS-SVM, considering that it was a nonlinear function and a more compact supported kernel, and could reduce the computational complexity of the training procedure while giving good performance under general smoothness assumptions (Cao Jian et al., 2009). According to the *Environmental quality standards for surface water (GB 3838-2002)*, the input feature subset contains six variables including dissolved oxygen, permanganate index, BOD$_5$, ammonia nitrogen, total phosphorus and total nitrogen. And the water quality level (Ⅰ, Ⅱ, Ⅲ, Ⅳ and Ⅴ) is selected as out variable. In addition, proper parameter setting plays a crucial role in building a good LS-SVM classification model with high prediction accuracy and stability. In this paper, we employed Grid-search Technique to find out the optimal parameter values such as regularization parameter *gam* (γ) and the RBF kernel function parameter *sig2* (σ^2) which is the bandwidth in the case of the RBF kernel.

5. Results and Discussion

First, the WSN network including eighteen sensor nodes and related water quality index detectors, three router nodes, one coordinator node, is distributed in surface wastewater monitoring region which is about 0.15 square kilometer. The reliability in communication is tested and verified; the result showed that the bit error rate is smaller than 2%.

Second, one hundred water samples were collected from five kinds of region, including municipal wastewater, industry wastewater, mixed domestic and industrial wastewater, reclaimed water and well water region. The samples were randomly separated into training set (80 samples) and validation set (20 samples), and LS-SVM classification algorithm was used to build training model, then the model was employed for the prediction of the validation set samples. The result shows that 19 prediction values are right, namely, the accuracy rate is 95%. And the only exception is that showed as follows:

Table 1. The only wrong prediction sample

Dissolved oxygen	3.65
Permanganate index	4.82
BOD$_5$	4.72
Ammonia nitrogen	0.70
Total phosphorus	0.284
Total nitrogen	1.02
Actual water quality level	Ⅳ
Prediction water quality level	Ⅴ

6. Conclusions

1) A water quality monitoring and evaluation system based on WSN/ZigBee and LS-SVM method was designed. The system is economic and practical, and can reach real-time water quality index detecting and automatic water quality evaluation.

2) The WSN/ZigBee node design was introduced in this paper. The test results, coming from a wastewater surface water monitoring experiment, show that the ZigBee-based remote monitoring network had sufficient reliability in communication.

3) The LS-SVM method was introduced in detail, and the test results show that the LS-SVM-based quality evaluation method has high prediction and classification accuracy.

Acknowledgements

This work was supported by the Natural Science Foundation of Zhejiang Province of China (No. Y107761), the Major Science & Technology Program of Ningbo City of Zhejiang Province of China (NO. 2007C10034), and the Science & Technology Program of Jiaxing City of Zhejiang Province of China (NO. 2009AY2016).

References

[1] World Resources Institute, "A Guide to the global environment-environmental change and human health", The World Resources Institute, The United Nations Environmental Programme, The United Nations Development Program, and The World Bank, New York, NY, pp. 1-369, 2003.

[2] Burkholder, J.M., "Critical needs in harmful algal bloom research", National Research Council, Opportunities for Environmental Applications of Marine Biotechnology. National Academy of Sciences Press, pp. 126–149, 2000.

[3] H.B. Glasgow et al., "Real-time remote monitoring of water quality- a review of current applications, and advancements in sensor, telemetry, and computing technologies", Journal of Experimental Marine Biology and Ecology, Vol. 300, pp. 409-448, Feb. 2004.

[4] Kevin G. R., Vera D. C., Young-Je P. "Seaborne measurements of near infrared water-leaving reflectance: the similarity spectrum for turbid waters", Limnol. Oceanogr. Vol. 51, No.2, pp. 1167-1179, 2006.

[5] Cao Jian, Qian suxiang, Hu hongsheng. "Water-quality on-line evaluation system based on RBF-ANN". Mechanical and manufacturing, Vol.47, No.533, pp. 53-55, January 2009.

[6] Yunseop Kim, Robert G.E., and William M.I. "Remote Sensing and Control of an Irrigation System Using a Distributed Wireless Sensor Network", IEEE Transactions on Instrumentation and Measurement, Vol. 57, No. 7, pp. 1379-1386, July 2008.

[7] Zhang Wei, He Yong, Qiu Zhengjun, et al. "The design of precision irrigation system based on wireless sensor network and fuzzy control". Transactions of the CSAE, Vol. 25, Supp. 2, pp. 7-12, Oct. 2009.

[8] J. A. K. Suykens et al., Least Squares Support Vector Machines, World Scientific Publishing, 2002.

[9] Di Wu, Yong He, Shuijuan Feng, et al. "Study on infrared spectroscopy technique for fast measurement of protein content in milk powder based on LS-SVM", Vol. 84, Issue 1, pp. 124-131, January 2008.

Biographies

Zhang Wei, male, born in 1980, from Shandong Province, Ph. D. Candidate of Zhejiang University, interested in the research on the application of information technology in agriculture.