Data fusion algorithm analysis and realization based on wireless sensor networks

Na Wang*

CAPF Of Engineering University, wujing road NO 1,Weiyang District, Xi'an City, Shaanxi Province, 710086, China

Received 1 March 2014, www.cmnt.lv

Abstract

In the premise of ensuring the veracity of model prediction result as well as simplify the model and prediction algorithm, this paper introduces a wireless sensor network data fusion algorithm based on ARMA time series model. This algorithm aims at reducing the energy consumption of wireless sensor network and improving the accuracy of fusion result. It conducts reliability analysis to node collecting data and removes the abnormal data. By analyzing ARMA model, we find that the construction of prediction model and cost of predicting is related to order of the model. The experiment result shows that this algorithm can not only reduce the network energy consumption but also detect abnormal data. ARMA model that determinates by BIC&F applicability test methods can adapt the wireless sensor network well.

Keywords: wireless sensor networks, data fusion, prediction, time series analysis

1 Introduction

Data Fusion of Wireless Sensor Networks (WSNs) is a kind of technology that eliminates data redundancy transmission by processing data that sensor nodes collects and sends the fusion result to base station [1-3]. By lowering the data traffic in network, data fusion can reduce energy consumption to a large degree and extend network lifetime. Data fusion algorithm based on prediction means predicting future data by modeling analysis of historical data collected by sensor nodes. If the prediction error is within the allowable range of threshold value, the mode won't send the data; otherwise the node will sent the data to fusion node and update prediction model.

When the nodes are affected by human or environmental factors, the collected data are abnormal, which will result in prediction error bigger than threshold [4-6]. Currently, the main processing method is sending the data to fusion node for fusion. This won't influence the final fusion result but can increase network traffic and shorten network lifetime. Besides, the energy, computing power and storage capacity of sensor node itself is limited. Fusion algorithm won't be feasible if the modeling algorithm and prediction algorithm are too complicated. Thus, prediction fusion algorithm asks for simplifying prediction model and prediction algorithm in the premise of guaranteeing the veracity of prediction result. This paper introduces a wireless sensor network data fusion algorithm based on ARMA time series model. This algorithm aims at reducing the energy consumption of wireless sensor network and improving the accuracy of fusion result. By utilizing the correlation of collected data in continuous time, the algorithm analyzes the historical data and predicts the future data. Considering that the deployment strategy of WSN results in the spatial correlation of nodes collected data, the algorithm conducts reliability analysis to collected data and removes the abnormal data. By analyzing the ARMA model, we find that the construction of prediction model and cost of predicting is related to order of the model

2 Data fusion of wireless sensor networks

The core of Data Fusion or the named Information Fusion is gathering multi-level information and extracting specific information from the nodes collecting data as the task request.

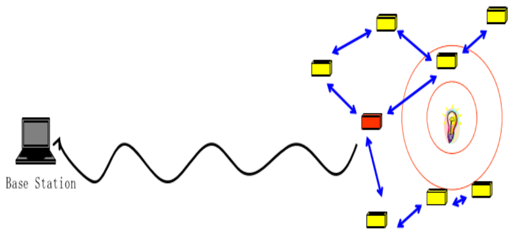

FIGURE 1 Data Fusion of Wireless Sensor Network

Information fusion can process and integrate the information of various sources, patterns and presentations, and finally the descriptors of monitoring target can be got by a more precise way. In order to make up the deficiencies of single sensor nodes in perception precision and data reliability, most of the advanced systems adopt multi-sensor structure at present. Form Figure 2, we can see that adopting the pattern of staged processing in information fusion can not only enhance system reliability effectively but also improve robustness of the system.

* ***Corresponding author's*** e-mail: 2005sissi@sina.com

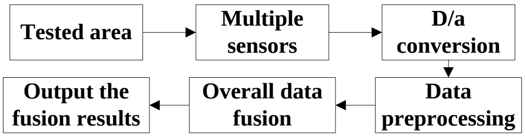

FIGURE 2 Sketch Map of Sensor Data Fusion

One of the important functions of network fusion technology is to combine the data collected by carious nodes. This is the function of fusion function. The specific design of fusion function is closely related to the application program of sensor network. Generally, it can be divided into the following two kinds:

1) Loss and lossless fusion function: fusion function can compress and combine the data of multiple data packet by the means of loss or lossless function. By loss function, the original node-collecting data can't be restored by fusion function; on the contrary, the lossless function can compress data with reserving the original information. This means all data can be reconstructed in the receiving end.

2) Duplicate sensitive and replication insensitive fusion function: Intermediate node can receive multiple duplicates of a same data packet. In this situation, the same information can be considered for many times when fusing. If the fusion function is duplicate sensitive, the final result relies on the consideration times of the same data. If the fusion function is duplicate-insensitive, it will only consider once for a same data packet. For example, as for averaging, this fusion function is duplicate sensitive; but for minimum value, it is insensitive.

The main data fusion algorithm of wireless sensor network includes fusion algorithm based on tree structure, fusion algorithm based on cluster structure, fusion algorithm based on multipath and mixed fusion algorithm. Fusion algorithm based on tree structure combines the data flows from multiple originating nodes to Sink node, namely selecting a few special nodes as the fusion node and defining the path data forwarding following in advance, as shown in Figure 3: increase the gradient of the information first and record the source nodes of the information, and then send the message.

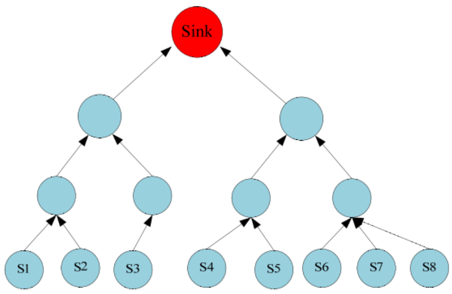

FIGURE 3 Diagram of Tree Structure Data Fusion Algorithm

Because a node might receive many massages, for a node, it may have many father nodes. All these result in many information transmission paths between each source node and the base station. This is the gradient foundation process. In the stage of path enhancing, as to some certain perception data and based on a certain metric function, there will be only path, which will be used for sending perception data. Figure 3 explains these three processes separately from left to right.

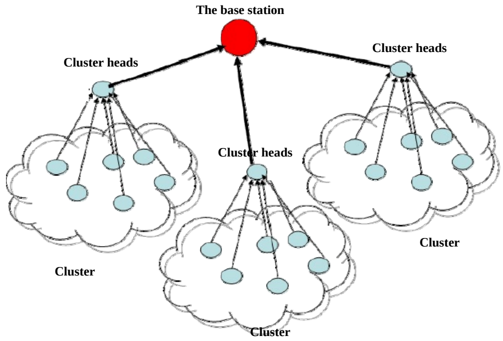

FIGURE 4 Network Topology of Leach Protocol

In the fusion protocol based on cluster structure, all nodes of the wireless sensor network are divided into many clusters. Cluster head is in charge of receiving the information collected by the nodes in this cluster and then send the result after fusing. The cluster head can communicate with Sink node directly if the distance between them is relatively short. On the contrary, they can communicate by the fusion tree constituted by cluster heads in the way of multi-hop from the perspective of energy-saving, as shown in Figure 4.

A new way based on multi-path differs from the fusion tree in that each node can send the perceiving data to multiple neighbor nodes rather than to one father node. In this multi-path fusion protocol, the data can follow many paths from source node to Sink node can fuse in each node. Because of the multi-path transmission of a data packet, this protocol improves the robustness of data transmission so that it can be well applied to the network that losing data packet frequently due to node moving or channel fault, as shown in Figure 5.

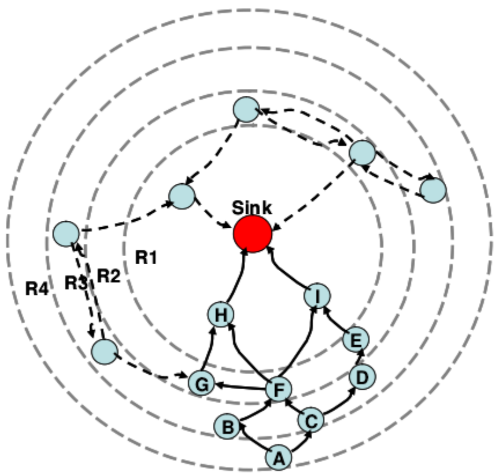

FIGURE 5 Annular Multipath Fusion Structure

In different internet zones, the two kinds of fusion structure operate at the same time. In the zone of low packet loss rate, tree-form fusion structure is more suitable, because it can imply randomized scheduling scheme and compress data effectively. In the zone of high packet loss rate or where the data need to be collected from multiple child node, the robustness of multiple paths make itself more suitable. Thus, in the protocol of mixed fusion structures, the nodes of the sensor network can be divided into two kinds: one transmits data packets by tree form and the other by multiple paths.

3 Universal model of ARMA time series analysis data fusion

3.1 ARMA TIME SERIES ANALYSIS MODEL

Time series analysis is a way of data processing that it analyzes and handles the observed ordered random data by parametric model [7]. ARMA time series analysis model is a kind of time series analysis methods that wildly used in practical time series analysis system. As for time series$\{x_t\}$(t=1,2,…, *N*), if it has the characteristics of stationarity and zero-mean, $\{x_t\}$ can match the stochastic difference equation as:

$$\begin{aligned} x_t &= \varphi_1 x_{t-1} + \varphi_2 x_{t-2} + \ldots + \varphi_n x_{t-n} - \\ &\theta_1 a_{t-1} - \theta_2 a_{t-2} - \ldots - \theta_n a_{t-n} \end{aligned}, \quad (1)$$

x_t is the value of time series $\{x_t\}$at time *t*; $\varphi_i(i=1,2\ldots,n)$ is the parameters of auto-regression(AR); $\theta_j(j=1,2\ldots,m)$ is the parameters of moving average (MA); Order $\{a_t\}$ is the residual sequence of true value and predicted value. If Equation (1) can reflect the system well, $\{a_t\}$ is white noise. The first half of the right of equation is auto-regression and the second half is moving average. This formula is n order auto-regression and m order moving average, denoted as ARMA (*n*, *m*). The orders of auto-aggregation and moving average are separately *n* and *m*.

3.2 ARMA DATA FUSION

3.2.1 Earlier stage processing of ARMA data fusion

The time series ARMA needed should be discrete. As for continuous signal, discrete sampling is needed [6,7]. In this paper, the used data packets are discrete temperatures and deflections collected by sensor nodes deployed in a bridge and they don’t need to discrete. The application backgrounds of sensor network differ and the frequencies of node collecting data differ as well. Sampling interval and data length of ARMA model don’t have to be the true interval and length. Data packet this experiment used is the temperature and deflection time series collected by nodes deployed in a bridge. The nodes collect 120 times data per day in the experiment, namely 12 minutes one time.

ARMA model bases on time series $\{x_t\}$ with the characteristics of stationary, normality and zero-mean. Thus, before ARMA modeling, we should test whether continuous time series satisfy the features of stationary, normality and zero-mean.

If not, the original $\{x_t\}$ series should be normalized and zero out to satisfy the conditions.

3.2.2 Modeling

There are two modeling schemes for ARMA modeling: BOX and (2n, 2n-1). This paper select the (2n, 2n-1) scheme that suitable for engineering application.

(2n, 2n-1) modeling scheme makes n=1, fitting ARMA (2,1) modeling and testing the feasibility. If it is unfit, make n=n+1to fit until find the right model. Then lower the orders of AR or MA to find the most suitable ARMA (n, m). We will introduce the ARMA modeling processes of (2n, 2n-1) modeling scheme with 500 deflection data $\{x_t=x_1, x_2, \ldots, x_{500}\}$:

1) Make n=1. Make least square estimation to x_t for fitting ARMA (n, m) (n=2n, m=2n–1), and then test whether ARMA (2n, 2n–1) is suitable. If not, make n=n+1 to fit ARMA (2n, 2n– 1) until the model is applicable according to BIC test result, then we can enter the next step:

$$\begin{cases} \beta = (x^T x)^{-1} x^T y \\ \varphi = \beta(1:n) \\ \theta = -\beta(n+1:n+m) \end{cases}. \quad (2)$$

2) The parameter φ and θ got from least square method is used for test whether the values of φ_n and θ_m are approaching to zero. If not, we can know that the orders of the model are 2n and 2n-1, and then we can enter the next step.

3) Lower the orders of AR and MA to fit ARMA (2n-1, 2n-2). Conduct F-test to the former fitted ARMA (2n, 2n-1) and present ARMA (2n-1, 2n-2). If F< F_a, the ARMA (2n-1, 2n-2) is suitable, then we can lower the order of MA to m=2n-2 and conduct F-test to the former fitted ARMA (2n, 2n-1) until F>F $_a$, when m is the order of MA. Or we have to enter the next step.

4) Conduct F-test to the former fitted ARMA (2n, 2n-1) and ARMA (2n-1, 2n-2). If F> F_a, the ARMA (2n-1, 2n-2) is unsuitable. Thus we don’t need to lower the order of AR but the order of MA: m. We further fit ARMA (2n, m) and conduct F-test with the former ARMA (2n, 2n-1) until F is efficacious. Thus we got the final model ARMA (2n, m).

We have to conduct applicability test after fitting time series x_t, namely testing whether the fitted ARMA (n, m) accords with the assumptions. The essence of model applicability test is to determine the order. According to the differences in testing form, the present model application test can be divided into the following categories: white noise test criterion, residual sum of squares test criterion, Akaike test criterion etc. The common test criterions are F-test, AIC information criteria and BIC information criteria.

3.2.3 Model Data Prediction and Update

ARMA time series model can predict the future value of sensor node by establishing prediction algorithm. But the predicted value may have deviation with the physical truth over time. At this time, we should update predicted value. Thus, the present prediction-based wireless sensor network data fusion algorithm will send the truth data to aggregation node when finding there are large differences between generated value and predicted value and update its prediction model. However, abnormal data can be larger than error threshold, when the generated data shouldn't be sent to aggregation node but should be removed. In most of the application environment of wireless sensor network, sensor node will generate unreliable or inaccurate data when energy lowers or the node is affected by other nodes or the surrounding environment. If sensor node sends abnormal data to aggregation node for fusion, the network data traffic won't increase but the fusion result will affect the accuracy. Thus, the abnormal data source node generated should be tested and removed.

Test the stationarity, normality and zero-mean of the original data $x_t^{(0)}$ firstly. Pretreat the data and get $y(t) = (x_t^{(0)} - \overline{u_x}) / \overline{\sigma_x}$. Then determine the order of $\{y_t\}$: ARMA (n, m) by (2n, 2n-1）modeling process and the applicability test, and got the model parameter φ and θby least square estimation. Afterword, we can compute the predicted value of time series $\{y_t\}$ at time t by the following equation:

$$\overline{y}_t = \varphi_1 y_1 + \varphi_2 y_{t-2} + \ldots + \varphi_m y_{t-m} - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \ldots - \theta_m a_{t-m}, \quad (3)$$

At last, restore the predicted value at time t. in most of the WSN environment monitoring applications, sensors are densely deployed into the monitoring environment to make the monitoring more effective. Thus, the sensing data of nodes in a small area vary little and the data variation tendencies are also similar. In mathematical statistics, correlation coefficient is used for expressing the relations of variable x, y or the relation of signal. Correlation coefficient is the mathematical expectation of the product of the two random variables.

Only when the predicted value of most modes in an area (mainly judging by the percentage of the total nodes) differ from the actual situation a lot. We can suspect that the applicability of the model and update the prediction model in the area, or we believe the node-collecting data are affected if the error of predicted value and collection value is larger than the threshold value.

4 Efficient WSN data fusion algorithm

In the fusion algorithm of ARMA prediction model, the energy loss of source node mainly lies in constructing ARMA model, data prediction and transmitting model parameters; the energy of fusion node mainly contributes to data prediction, model parameter receiving, data fusing and fusion result transmitting to base station. In the aspect of computing processing capacity, source node has certain requirements for the node's processing capability in ARMA model constructing or data predicting; cluster head should restore the first n data (given the order of Auto-aggregation is n)for predicting except for storing the model parameters.

The energy consumption, processing capability, storage space of ARMA prediction algorithm is related to the order of ARMA. The higher the order is, the more energy, the better processing capability and bigger storage space will be needed. Thus, the order of ARMA prediction model needs to be lowered to optimize in the premise of guaranteeing the accuracy. In this paper, we make contrastive analysis to the adaptability of model order to search for adaptability that is applicable to WSN for model optimizing. We also analyze the performance of the optimized model from the perspective of energy consumption, computing complexity and storage space. At last, the optimized ARMA time series model is applied to wireless sensor network for WSN data fusion algorithm design.

Processes of the algorithm in member nodes is shown in Figure 6.

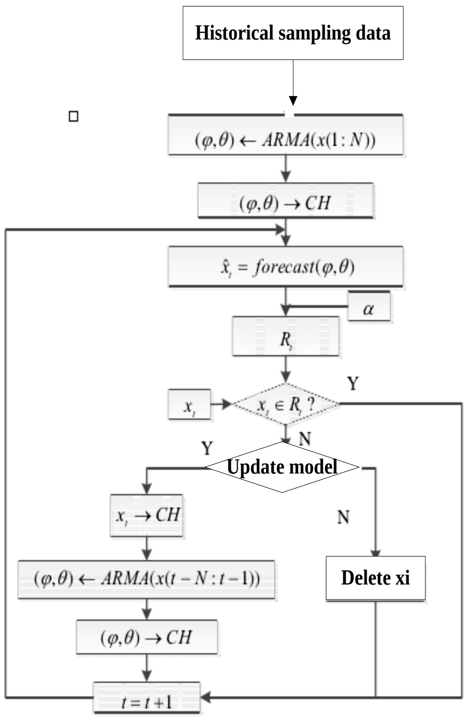

FIGURE 6 Algorithm Flows of Member Nodes

1) Member nodes establish the ARMA (n, m) model by the historical sampling value sequence.

2) Member nodes send the model parameters to the cluster head of their cluster and predict the next sampling value by historical data and model parameters.

3) Member nodes compute the range R_t of the new collecting data by the given confidence level and predicted value.

4) If x isn't in R, move to step 5; Otherwise move to step 8.

5) If most of the nodes in the cluster believe the prediction model isn't applicable, move to step 6; Otherwise move to step 7.

6) Member node sends real collecting value to cluster and updates the model by the N updated data, and then it sends the new model parameters to the cluster.

7) Member node removes the collecting value x_t。.

8) Make t=t+1 and move to step 2 to process the next round.

Cluster heads are in charge of fusing the predicated value and node-collecting data and send them to base station. Cluster head receives the model parameters m, n, φ, θ sent by nodes in this cluster in the beginning and then predicts the next collecting value by the parameters. If cluster head doesn't receive the collecting value source node sending within the stipulated time, it will fuse the predicted value as the true value; if cluster head receives the data sent by the source node, it will fuse the received data and replace the former parameter with the received one for the next prediction. At last, it sends the fusion result to the base station. Process of the algorithm in cluster head is shown in Figure 7:

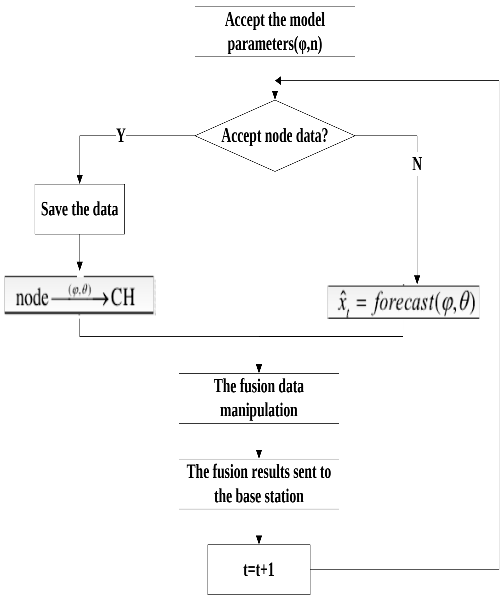

FIGURE 7 Algorithm Flows of Cluster Head

1) Cluster head receives the model parameters sent by member nodes.

2) If cluster head receives the data, move to step 3, otherwise move to step 4.

3) Cluster head save the data sent by the nodes and receive the new model parameters sent the nodes, and then move to step 5.

4) Cluster head regards the predicted value getting from the model parameters of each node as the true collecting value.

5) Cluster head fuses the data collected by each node and transmits them to the base station.

6) Make t=t+1 and move to step 2 for another round.

By improving ARMA prediction model, the algorithm lowers its orders n and m, therefore lowers the cost of ARMA construction and data prediction, reduces the demand for computing capability and energy loss of source node and cluster head, and extends the network lifetime. Order lowering reduces the needed historical data when predicting and the demand for storage space. The simplifying of model construction and prediction algorithm lowers the demands for node processing capability.

5 WSN data fusion algorithm simulation experiment analysis

This stimulation experiment proceeds on Microsoft Windows 7 + MATLAB R2010a. Data packet is the collected temperature and deflection. Parameter data number for model building is N=100; confidence level is α=0.1; the percentage of mistakes is β=0.8. The experiment selects 2000 continuous data from 6 neighbor nodes to ensure the space correlation consumption of the collecting data, as shown in Figure 8.

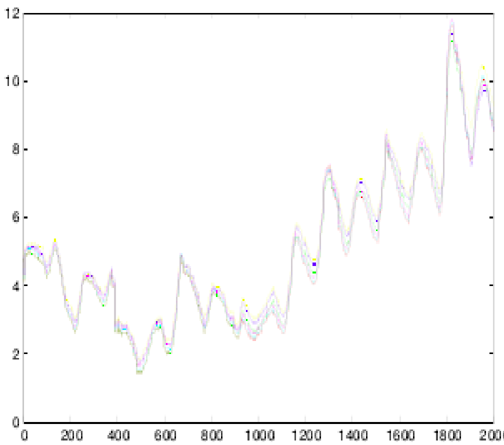

FIGURE 8 Data Variation Tendency Chart of Adjacent Nodes

We will contrast the gray prediction fusion algorithm mentioned in the literature with ARMA prediction fusion algorithm in the following. The experiment conducts average fusion to the data collected by 6 sensor nodes at fusion nodes. Figure 9 shows the different results of gray prediction fusion algorithm and ARMA prediction fusion algorithm. In the chart, red line is the fusion result of real data; blue-green line and blue line are separately the fusion result of gray prediction fusion algorithm and ARMA prediction fusion algorithm. We can see that the two algorithms are all close to the fusion result of sensing data, but ARMA model is more close to the reality.

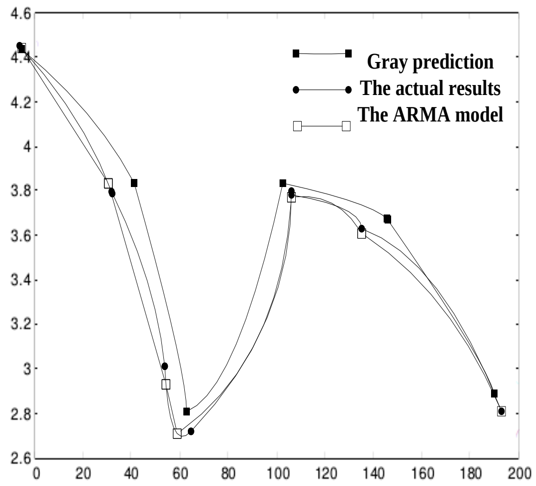

FIGURE 9 Comparison Diagram of Gray Prediction and ARMA Prediction

As shown in Figure 10, ARMA dynamic data fusion algorithm has improved a lot in system operation cycle comparing with static fusion algorithm when the first node dying. However, with energy consuming of the node, it improves little in extending system lifetime when half of the nodes die and the last node dies. For one thing, all applications are data related, namely the nodes collect temperatures and select the maximum one. For another, the energy consumption of the fusion, part of fusion function is ignored. Thus, the experiment result cannot embody the superiority of dynamic data fusion algorithm well.

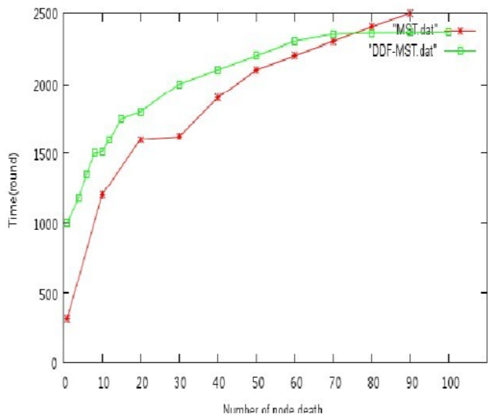

FIGURE 10 System Operation Cycle Comparison When Node Dies Based on ARMA Static and Dynamic Algorithm

6 Conclusions

Data fusion makes the original row data into less refining data to transmit to users, decreasing the traffic of WSN as well as saving node energy and extending network lifetime. Prediction-based WSN data fusion technology establishes the prediction model by node-collecting historical data to reducing data transmitting. After observing the historical data sensor node collected and analyzing the practical application of WSN, we find that the collecting data of neighboring nodes are correlated on time and space. To simplify the ARMA prediction model and the prediction algorithm, we lower the parameter of the model in the premise of keeping the result accuracy. The experiment result shows that the algorithm in this paper can predict and detect abnormal data and save energy to extend lifetime.

References

[1] Sun L, Li J, Chen Y 2005 Wireless Sensor Network *Tsinghua University Press* 2005.5: 3-23/259-27 *(in Chinese)*

[2] Fu H 2009 Balanced-energy Clustering Scheme for Wireless Sensor Networks *Computer Application Research* **26**(4) 1494-6 *(in Chinese)*

[3] Zhou P 2010 Node Energy Aggregation Mechanism Based on Forecast Model in WSNs *Computer engineering* **36**(1) 110-1 *(in Chinese)*

[4] Xiang M 2008 Energy Saving Algorithm Based on Cluster Head Prediction for Wireless Sensor Networks. Computer engineering **34**(18) 27-9 *(in Chinese)*

[5] Yao L, Zhao L 2010 A Wireless Sensor Network Data Fusion Algorithm Based on Neural Network *Computer Knowledge and Technology* 2010.06.18

[6] Yang S, Wu Y, Xuan J 2007 Engineering Application of Time Series Analysis *Huazhong University of Science and Technology Press (in Chinese)*

[7] Wu H, Zhang Z, Li W 2011 Energy-balanced Routing Protocol for Wireless Sensor Networks in Coal Mine *Computer Science* **38**(4) 145-9 *(in Chinese)*

[8] Xiang M, Shi W, Jiang C, Luo Z 2008 Energy Saving Algorithm Based on Cluster Head Prediction for Wireless Sensor Networks *Computer Engineering* **34**(18) 27-9 *(in Chinese)*

[9] Zheng W, Liu S, Kou X 2009 Energy Efficient Routing Algorithm Based on Ant Colony Optimization for Wireless Sensor Network *Systems Engineering and Electronics* **31**(8) 10-6 *(in Chinese)*

[10] Lang W, Yang K, Wu S 2011 Security in Wireless Sensor Network. *Journal of PLA University of Science and Technology* **12**(1) 42-7 *(in Chinese)*

Author

Na Wang, July1985, Shannxi Xi'an, China.

Current position, grades: teacher, associate professor.
University studies: CAPF of Engineering University.
Scientific interests: electronic information engineering.
Publications: 15.