Postprint of WSN Data Fusion Algorithm Based on Anomalous Data Preprocessing and Adaptive Estimation

Authors: Zheng Baozhou, Wu Lili, Li Fuqiang, Yuan Chao

Date: 2018-09-12T00:00:00+00:00

Abstract

To address the issues of limited node energy, low measurement accuracy, and short lifetime in wireless sensor networks (WSN), an abnormal data-preprocessing adaptive estimation weighting fusion (ADAEWF) algorithm is proposed. To enhance algorithm reliability, a data preprocessing mechanism based on abnormal data detection, simple majority principle, and node comprehensive support function is presented; to mitigate the influence of measurement errors on fusion precision, adaptive estimation weighted data fusion is conducted on node measurements based on batch estimation and adaptive theory; subsequently, a WSN simulation model is established, and the mean square error of fusion results and network effective lifetime are obtained for ADAEWF, adaptive forecast weighting data fusion (AFWDF) algorithm, and arithmetic mean method, respectively. Simulation results show that the fusion precision and network effective lifetime of the ADAEWF algorithm outperform those of AFWDF and the arithmetic mean method, demonstrating the superiority of the ADAEWF algorithm in enhancing the effectiveness of fused data, network effective lifetime, and fusion precision.

Full Text

Data Fusion Algorithm Based on Abnormal Data Preprocessing and Adaptive Estimation in WSN

Zheng Baozhou, Wu Lili, Li Fuqiang, Yuan Chao
(College of Sciences, Henan Agricultural University, Zhengzhou 450002, China)

Abstract: To address the problems of limited node energy, low measurement accuracy, and short network lifetime in wireless sensor networks (WSN), this paper proposes a data fusion algorithm based on abnormal data preprocessing and adaptive estimation weighting fusion (ADAEWF). To improve algorithm reliability, we first propose a data preprocessing mechanism based on abnormal data detection, simple majority principle, and node comprehensive support function. To reduce the impact of measurement errors on fusion accuracy, we perform adaptive estimation weighted data fusion on node measurements based on batch estimation and adaptive theory. We then establish a WSN simulation model and obtain the mean square error and network effective lifetime under ADAEWF, adaptive forecast weighting data fusion algorithm (AFWDF), and arithmetic mean method. Simulation results show that ADAEWF outperforms AFWDF and arithmetic mean method in both fusion accuracy and network effective lifetime, demonstrating the superiority of ADAEWF in improving data fusion effectiveness, network lifetime, and fusion accuracy.

Keywords: wireless sensor network; abnormal data preprocessing; adaptive batch estimation; data fusion

0 Introduction

Wireless sensor networks (WSN) consist of numerous sensor nodes with sensing, computing, and wireless communication capabilities, characterized by small node size, low cost, multi-hop self-organization, and large sensing area. WSN can effectively monitor and provide early warnings for environmental information and have been widely applied in smart agriculture, environmental monitoring, national defense, and military fields. WSN is a resource-constrained network whose performance is limited by node energy, computing capacity, and storage space. Typically, a large number of homogeneous sensor nodes are deployed in the sensing area to periodically collect and transmit information. While monitoring information, WSN generates substantial redundant data. On one hand, processing and transmitting redundant data wastes limited energy and network bandwidth; on the other hand, interference factors and sensor measurement accuracy lead to errors in system monitoring results, and randomly occurring faulty nodes reduce system reliability to some extent. How to reduce redundant data, decrease node energy consumption, improve network reliability, and extend its effective lifetime has become a very important issue in WSN research.

Multi-sensor data fusion technology can effectively reduce data transmission, decrease node energy consumption, and improve monitoring result accuracy by fusing data with certain redundancy. Reference [1] uses information entropy to reflect the statistical characteristics of sensor data distribution, determines the upper and lower bounds of data fusion through optimization, and performs two rounds of fusion on node data and intra-cluster data respectively. However, this algorithm' s fusion degree is limited by information entropy values, providing limited energy savings. Reference [2] proposes the AFWDF algorithm, which establishes a prediction model based on temporal correlation of node data. The cluster head calculates data credibility and weights according to eigenvalues and predicted values, and performs weighted fusion on intra-cluster node data. AFWDF requires establishing prediction models on both cluster heads and nodes, involves large computational overhead, and still has room for improvement in fusion accuracy and node energy efficiency. Reference [4] proposes a new inter-node support degree calculation method based on fuzzy evidence theory and fuzzy mathematics principles, converting measured values into corresponding evidence using trust distribution ideas, and obtaining fusion results through evidence combination rules. Existing WSN data fusion algorithms share a common characteristic: nodes collect, send data, and initiate fusion algorithms at fixed periods. In practical applications, monitored objects are mostly slowly changing physical quantities such as environmental temperature and humidity. Without special events, most data collected by nodes are repetitive. However, what the system focuses on is mainly the change of parameters, i.e., whether important events occur in the sensing area. If WSN remains in a high-energy-consumption active state monitoring redundant data for long periods, the practical significance is limited.

Abnormal data has important application significance in WSN monitoring systems. Abnormal data refers to measurement data that significantly deviates from the normal working pattern of sensors. The generation of abnormal data indicates that important monitoring events have occurred in the sensing area, or that abnormal factors such as external interference or node failures have caused sensor data anomalies. Whether caused by important events or faulty nodes, abnormal data should receive sufficient attention. For abnormal data caused by external events, the system should timely and accurately collect, fuse, and report monitoring results; for abnormal data caused by external interference and node failures, the system should identify, detect, and process them accordingly to avoid affecting fusion result accuracy.

Based on data-driven thinking, this paper proposes an adaptive estimation weighted fusion algorithm based on abnormal data preprocessing (ADAEWF). Simulation results show that the ADAEWF algorithm can effectively reduce WSN redundant data volume, extend network effective lifetime, and achieve high fusion accuracy.

1 Network Model and Algorithm Concept

As shown in Figure 1 [Figure 1: see original paper], WSN consists of a monitoring center PC, gateway, and multiple wireless sensor nodes. The monitoring system uses a first-order radio model as the WSN energy consumption model and employs the Energy Balanced Adaptive Clustering Algorithm (EBACA) to divide the network into several clusters, each containing a cluster head and several sensor nodes. Sensor nodes are responsible for data collection, detection, transmission, and monitoring of node status and external events. Cluster head nodes collect data from intra-cluster sensor nodes, run the ADAEWF fusion algorithm, and upload fusion results to the gateway. The gateway forwards cluster data fusion results to the monitoring center PC. If users issue data query commands, the gateway and cluster heads sequentially receive and forward query commands to each sensor node.

The ADAEWF algorithm concept is as follows: In automatic monitoring mode, effective abnormal data serves as the driver for the fusion algorithm. Sensor nodes periodically collect data and detect abnormal data through the Sliding Window Local Event Detection (SW-LED) algorithm, sending persistent abnormal data to the cluster head. The cluster head judges the validity of abnormal data based on the simple majority principle and node data comprehensive support degree. If the quantity of abnormal data reaches the stimulation intensity threshold, the cluster head triggers the adaptive estimation weighted data fusion algorithm. Finally, the cluster head uploads the fusion result to the gateway and monitoring center PC. In user query mode, ADAEWF treats user query requests as important monitoring events. When users issue a "data query" command, the gateway and cluster heads broadcast a "forced data transmission" command to the network. Intra-cluster sensor nodes collect current data and send it to the cluster head, which unconditionally triggers the adaptive estimation weighted data fusion algorithm and finally uploads the data fusion result to the gateway and monitoring center PC. The ADAEWF algorithm flow is shown in Figure 2 [Figure 2: see original paper].

2 ADAEWF Algorithm

2.1 Abnormal Data Preprocessing

Sensor node-generated abnormal data is divided into three types: (a) instantaneous abnormal data caused by sudden external interference or transient faults; (b) faulty abnormal data caused by insufficient node residual energy, hardware failures, or software defects; and (c) effective abnormal data where environmental information changes significantly due to external events, causing "abnormal" real data collected by multiple nodes. The processing of different abnormal data types is described below.

2.1.1 Instantaneous Abnormal Data Preprocessing Instantaneous abnormal data occurs frequently but cannot accurately reflect real environmental information. If involved in data fusion, it reduces fusion result accuracy. The system uses the SW-LED algorithm to exclude instantaneous abnormal data from fusion operations.

Definition 1: For node data $x(t)$, if $|x(t)-x(t-d)| \geq \varepsilon$, then $x(t)$ is abnormal data. Where $x(t-d)$ is the last abnormal data sent by the node to the cluster head, and ε is the abnormal data determination threshold, whose value depends on the actual application requirements of the monitoring system.

According to Definition 1, sensor nodes perform binary judgment on $x(t)$: if abnormal data, the conversion result is recorded as 1, otherwise 0. If the conversion result is 1, the system further counts the sum of the previous $W-1$ data conversion results $\sum_{j=t-W+1}^{t-1} y(j)$. If $\sum_{j=t-W+1}^{t-1} y(j) < 0.5W$, meaning less than half of the measurement values in the sliding window are abnormal, $x(t)$ is considered instantaneous abnormal data and requires no further processing. If

$\sum_{j=t-W+1}^{t-1} y(j) \geq 0.5W$, meaning more than half of the measurement values in the sliding window are abnormal, $x(t)$ is considered persistent abnormal data, and the node transmits $x(t)$ to the cluster head.

2.1.2 Faulty Abnormal Data and Effective Abnormal Data Preprocessing As the system operates, node residual energy continuously decreases, and node software and hardware are affected by various uncertain factors, causing random faulty nodes in WSN. Faulty nodes continuously generate abnormal data in each collection cycle, called faulty abnormal data. On one hand, although faulty abnormal data can pass SW-LED algorithm detection, it cannot reflect real environmental information, and participating in fusion algorithms reduces fusion result accuracy. On the other hand, persistently existing faulty abnormal data frequently triggers fusion algorithms, causing other nodes to passively participate in data fusion and wasting node energy and network bandwidth. Therefore, cluster heads should identify faulty abnormal data and send faulty node ID and status information to the gateway and monitoring center PC. Faulty nodes often appear randomly, and the probability of multiple adjacent nodes failing simultaneously is very small. However, if important monitoring events occur in the sensing area, measurement values of adjacent nodes will change significantly. For example, wind and rain can cause large changes in farmland temperature and humidity, leading to abnormal sensor measurement values. Abnormal data caused by monitoring events truly reflects environmental information changes and is called effective abnormal data.

In summary, for faulty abnormal data, cluster heads should report node ID and status information to the gateway and monitoring center PC; for effective abnormal data, cluster heads should initiate the fusion algorithm to process all node data.

Definition 2: Cluster head stimulation intensity threshold γ, representing the cluster head excitation threshold. Let the number of intra-cluster nodes be m. According to the simple majority principle, γ is set to $0.4m$. Let the cluster head receive abnormal data from l nodes. If $l \geq \gamma$, the cluster head is strongly stimulated by abnormal data, considers that important monitoring events have occurred in the sensing area, enters an excited state, and triggers the data fusion algorithm. If $l < \gamma$, the cluster head considers that only a small number of nodes generate faulty abnormal data and sends faulty node information to the gateway.

2.1.3 Data Validity Judgment The group support degree method is introduced to judge data validity: a support degree is introduced for each node data, and the support degree size reflects the support level of neighboring nodes for that data' s validity. When multiple nodes measure the same parameter, the higher the validity of node i' s measurement value x_i, the higher the support degree from other nodes, and the higher the probability that x_i is valid data. This paper uses the support function $sup(a, b)$ to represent the support degree of data a for data b.

Yager proposed that the support function must satisfy three conditions: 1. $sup(a,b) \in [0,1]$; 2. $sup(a,b) = sup(b,a)$; 3. If $|a-b| < |x-y|$, then $sup(a,b) > sup(x,y)$.

Yager also provided a Gaussian support function:

$$sup(a,b) = K \times e^{-\beta(a-b)^2}$$

where $K \in [0,1]$ and $\beta \geq 0$.

The Gaussian support function can realistically reflect the support degree relationship between data, but exponential operations consume substantial hardware resources and are not suitable for resource-constrained WSN. This paper adopts a new support function based on grey relational nearness to describe inter-data support degree.

Definition 3: The support degree between data x_i and x_j is:

$$s_{ij} = \frac{D_{ij}}{1 + \beta \times K \times (x_i - x_j)^2}$$

where $D_{ij} = \frac{1}{1+|x_i - x_j|}$, $K \in [0,1]$, and $\beta \geq 0$. The K value determines the amplitude of the support function; the larger the β value, the faster the support function decays, called the decay factor. K and β can be adjusted according to actual monitoring requirements. Equation (2) satisfies Yager' s three necessary conditions, well approximates the Gaussian support function curve, and saves node energy and hardware resources by avoiding exponential operations.

Let n nodes measure the same parameter. The support degree matrix S is calculated from Equation (2):

$$S = \begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1n} \\ s_{21} & s_{22} & \cdots & s_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ s_{n1} & s_{n2} & \cdots & s_{nn} \end{bmatrix}$$

Define the comprehensive support degree $s_i = \sum_{j=1}^{n} s_{ij}$, representing the degree to which data x_i is comprehensively supported by other nodes. The larger the s_i, the closer x_i is to the measurement values of most other nodes, and the higher the validity of x_i; conversely, the more x_i deviates from most node data, the lower its validity.

Definition 4: Data validity determination threshold δ. If node data x_i' s comprehensive support degree $s_i \geq \delta$, then x_i is valid data; otherwise, x_i is considered invalid data.

According to Definition 4, all node data are judged: valid data participate in data fusion, while invalid data do not participate in data fusion, otherwise they would reduce fusion result accuracy.

2.2 Batch Estimation Fusion Algorithm

The batch estimation fusion algorithm originates from recursive estimation. By fusing multiple measurement values of the same parameter, it can effectively reduce the impact of measurement errors on fusion results and obtain accurate measurement results.

Let n nodes monitor the same environmental parameter. According to the principle of grouping adjacent nodes differently, nodes are divided into k groups, with each group containing the same or different numbers of nodes. The data collected by the p-th group of nodes is denoted as $x_{p1}, x_{p2}, \dots, x_{pn_p}$, where n_p is the number of nodes in this group. The average value $\hat{x}_p$ and standard deviation σ_p of the node data are:

$$\hat{x}_p = \frac{1}{n_p} \sum_{i=1}^{n_p} x_{pi}, \quad p = 1, 2, \dots, k$$

$$\sigma_p = \sqrt{\frac{1}{n_p} \sum_{i=1}^{n_p} (x_{pi} - \hat{x}_p)^2}$$

From batch estimation theory, the fusion result is shown in Equation (7):

$$\hat{x} = \left(\sum_{p=1}^{k} H_p^T R_p^{-1} H_p \right)^{-1} \sum_{p=1}^{k} H_p^T R_p^{-1} \hat{x}_p$$

where $\hat{x}$ is the average value matrix, $\hat{x}_p$ is the last data fusion result, σ is the variance of the fusion result, R is the measurement noise covariance matrix, and H is the coefficient matrix of the measurement equation. The expressions of each parameter are shown in Equations (8)-(11):

$$H = [1 \ 1 \ \cdots \ 1]^T$$

$$x = [x_1 \ x_2 \ \cdots \ x_k]^T$$

$$R = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_k^2 \end{bmatrix}$$

$$\hat{x}_p = \frac{1}{n_p} \sum_{i=1}^{n_p} x_{pi}$$

Substituting Equations (9)-(11) and H into Equations (7)-(8), the variance and batch estimation fusion result can be obtained as Equations (12)-(13):

$$\sigma^2 = \left(\sum_{p=1}^{k} \frac{1}{\sigma_p^2} \right)^{-1}$$

$$\hat{x} = \sigma^2 \sum_{p=1}^{k} \frac{\hat{x}_p}{\sigma_p^2}$$

2.3 Adaptive Estimation Weighted Fusion Algorithm

To reduce the impact of measurement errors on fusion accuracy, adaptive theory is combined with the batch estimation fusion algorithm. The relative variance between node data and batch estimation fusion results is calculated, and node data weights are adjusted according to the relative variance for adaptive estimation weighted data fusion. Figure 3 [Figure 3: see original paper] shows the fusion algorithm flow. The algorithm steps are briefly described as: first, nodes are grouped, and each group' s data mean $\hat{x}_p$, standard deviation σ_p, and fusion weight w_p are calculated to further solve the batch estimation fusion result $\hat{x}$; then the relative variance s_p^2 between node data x_{pi} and $\hat{x}$ is calculated, and the correction factor a_p for each group' s data is calculated based on s_p^2, σ_p, and w_p; finally, the corrected weight w_p' is used for secondary weighted fusion of data to obtain the final fusion result x^+.

The p-th group node data fusion weight w_p, batch estimation fusion result $\hat{x}$, and relative variance s_p^2 are solved by Equations (14)-(16):

$$w_p = \frac{1/\sigma_p^2}{\sum_{i=1}^{k} 1/\sigma_i^2}, \quad p = 1, 2, \dots, k$$

$$s_p^2 = \frac{1}{n_p} \sum_{i=1}^{n_p} (x_{pi} - \hat{x})^2$$

From multi-sensor weighted data fusion algorithm theory, to eliminate measurement errors, fusion weights should be inversely proportional to estimation variance, and the sum of all weights should always be 1. Therefore, w_p and a_p satisfy three constraint conditions shown in Equations (17)-(19):

$$\sum_{p=1}^{k} w_p = 1$$

$$\sum_{p=1}^{k} w_p a_p = 1$$

$$w_p a_p \geq 0$$

The batch estimation fusion algorithm fuses node current measurement values without considering historical monitoring information, assuming historical data standard deviation $\sigma_{-1} = \infty$. Based on the above analysis, the correction factor a_p is shown in Equation (20):

$$a_p = \frac{1/s_p^2}{\sum_{i=1}^{k} 1/s_i^2}$$

The adjusted weight w'_p is substituted into Equation (15) to obtain the adaptive estimation weighted fusion result shown in Equation (21):

$$x^+ = \sum_{p=1}^{k} w'_p \hat{x}_p = \sum_{p=1}^{k} w_p a_p \hat{x}_p$$

2.4 ADAEWF Operating Mechanism

Nodes with different roles have different functional behaviors. To facilitate description of ADAEWF algorithm operating mechanism, we define three state parameters: abnormal data generation flag mTag, cluster head excitation flag hTag, and data forced transmission flag send.

Definition 5: Abnormal data generation flag mTag, reflecting whether cluster member nodes generate non-instantaneous abnormal data, where 0 means no and 1 means yes.

Definition 6: Cluster head excitation flag hTag, reflecting whether the cluster head is in an excited state. hTag=0 means the cluster head receives fewer abnormal data than γ, and the cluster head is in an inhibited state; hTag=1 means the number of abnormal data exceeds γ, and the cluster head is in an excited state; hTag is reset to 0 at the end of the collection period.

Definition 7: Data forced transmission flag send, reflecting whether nodes receive the "forced data transmission" command. send=1 means nodes receive the "forced data transmission" command and will send measurement values to the cluster head; send=0 means nodes do not receive the transmission command.

Cluster member node actions are shown in Figure 4 [Figure 4: see original paper]: Upon receiving the "forced data transmission" command, the node is in an excited state, sets send=1 and mTag=1, sends measurement value $x_i(t)$ to the cluster head, and updates cached data; If send=0, the node judges according to Definition 1: if $x_i(t)$ is instantaneous abnormal data, sets mTag=0 and discards $x_i(t)$ without processing; otherwise, sets mTag=1, sends $x_i(t)$ to the cluster head, and updates cached data; At the end of the data collection period, resets mTag and send to initial value 0.

Cluster head actions are shown in Figure 5 [Figure 5: see original paper]: a) If receiving the "data query" command forwarded by the gateway, the cluster head is in an excited state and sets hTag=1; b) According to Definition 2, if the number of abnormal data $l \geq \gamma$, the cluster head is in an excited state and sets hTag=1; c) If $l < \gamma$, the cluster head is in an inhibited state, sets hTag=0, and sends faulty node information to the gateway; d) If hTag=1, the cluster head broadcasts the "forced data transmission" command, each member node sends data, and after receiving data, the cluster head initiates the fusion algorithm and sends the fusion result to the gateway, then restores hTag to the initial value 0.

3 Simulation Results and Analysis

To analyze and compare fusion algorithm performance, this paper conducts simulation experiments based on the NS2 platform, comparing fusion errors and network effective lifetime of ADAEWF, AFWDF, and arithmetic mean method. The experimental background is set as: 200 sensor nodes are uniformly deployed in a 200m×400m farmland to monitor farmland temperature changes; node row and column spacing is 10m, gateway node coordinates are (100,200); node failures, artificial irrigation, weather changes, etc., can all cause abnormal data generation; the energy-balanced EBACA algorithm is used for network node clustering, and the first-order radio model is selected as the WSN network energy consumption model; other simulation parameters are shown in Table 1 .

Table 1 Simulation Parameters | Parameter | Parameter Values | |——-|————| | E_{elec} | 50 nJ/bit | | ε_{fs} | 10 pJ/(m²·bit) | | ε_{mp} | 0.0013 pJ/(m ·bit) | | E_{init} | 2 J | | d_0 | 87 m | | l | 4000 bits | | f | 20 Hz | | T | 15 + 15sin(2 ft) °C | | ε | 5 °C |

The temperature change curve $T = 15 + 15\sin(2\pi f t)$ is superimposed with zero-mean white noise signal with variance of 5 to simulate farmland temperature changes and the impact of electromagnetic interference, zero drift, and other factors on node sensor measurement values. A certain proportion of nodes are randomly selected, and node data is set to 0 to simulate randomly occurring faulty nodes in the network. Adjusting ε can change the frequency of abnormal data occurrence; larger ε means higher abnormal data frequency.

3.1 Fusion Error Comparative Analysis

Under conditions of $\varepsilon = 5$°C and faulty node proportion of 5%, ADAEWF, AFWDF, and arithmetic mean method are used to perform data fusion on node data for 80 consecutive cycles. The fusion curves are shown in Figure 6 [Figure 6: see original paper]. Simulation results show that AFWDF and arithmetic mean method have relatively large fluctuations in fusion results, while ADAEWF algorithm has obvious advantages. The reason is that ADAEWF' s abnormal data preprocessing mechanism excludes instantaneous and faulty abnormal data generated by sensor nodes in the farmland temperature wireless monitoring system from participating in data fusion, effectively improving fusion result reliability. On the other hand, by adjusting fusion weights based on the relative variance between node data and initial batch estimation fusion results, and performing adaptive estimation weighted data fusion, the impact of sensor measurement errors on fusion results is further reduced, improving fusion accuracy and reliability.

Under the above simulation conditions, the mean square error curves of fusion results for the three algorithms varying with faulty node proportion are shown in Figure 7 [Figure 7: see original paper]. When the faulty node proportion is below 10%, the mean square errors of all three algorithms are less than 0.7°C. When the faulty node proportion exceeds 10%, the mean square errors of

AFWDF and arithmetic mean method gradually increase, reaching 2.3°C and 3.4°C respectively when approaching 40%. During the process of faulty node proportion increasing from 0% to 40%, ADAEWF' s mean square error remains below 0.8°C. Comparative analysis shows that in farmland temperature wireless monitoring systems with faulty nodes, ADAEWF algorithm has smaller fusion result errors and higher reliability.

3.2 Network Lifetime Comparative Analysis

Under conditions of $\varepsilon = 5$°C and faulty node proportion of 10%, the curve of network surviving node count versus algorithm operation rounds is shown in Figure 8 [Figure 8: see original paper]. AFWDF and arithmetic mean method experience their first node death at 850 and 760 rounds respectively, while ADAEWF experiences its first node death at 1300 rounds. When 70% of nodes die, AFWDF and arithmetic mean method operate for 1230 and 1110 rounds respectively, while ADAEWF operates for 1670 rounds. In farmland temperature wireless monitoring systems, when 70% of nodes die, the system cannot effectively complete monitoring tasks, and the network is considered to have entered the death period. In summary, ADAEWF' s network effective lifetime is 150% of arithmetic mean method and 135% of AFWDF, indicating higher energy efficiency. This is because AFWDF and arithmetic mean method periodically perform data collection, transmission, and fusion, consuming more energy, while ADAEWF only performs data collection and transmission when users query data or when farmland abnormal temperature data is detected; at other times, nodes remain in a low-power inhibited state, and cluster heads only trigger the fusion algorithm when stimulated by sufficient abnormal data. Additionally, the abnormal data preprocessing mechanism excludes instantaneous and faulty abnormal data from participating in data fusion, effectively improving fusion accuracy. Simulation results show that in the farmland temperature wireless monitoring system simulation model, compared with arithmetic mean method and AFWDF algorithm, ADAEWF algorithm has lower mean square error and longer effective lifetime.

4 Conclusion

This paper studies wireless sensor network data fusion and proposes an abnormal data preprocessing mechanism and adaptive estimation weighted data fusion algorithm. Different from continuous or periodically triggered fusion algorithms in existing literature, ADAEWF algorithm only operates when the system detects a sufficient quantity of effective abnormal data, which can effectively reduce algorithm operation frequency. Simulation results show that ADAEWF algorithm can effectively reduce WSN redundant data volume, extend network effective lifetime, and achieve high fusion accuracy.

References

[1] Li Huaijun, Zhang Xuexi. Research on data fusion method in cluster for wireless sensor network based on 2D entropy theory [J]. Application Research of Computers, 2014, 31(7): 2171-2174.

[2] Yu Xiuwu, Fan Feisheng, Zhou Lixing, et al. Adaptive Forecast Weighting Data Fusion Algorithm for Wireless Sensor Network [J]. Chinese Journal of Sensors and Actuators, 2017, 30(5): 772-776.

[3] He You, Wang Guohong, Peng Yingning, et al. Multi-sensor Information Fusion and Application [M]. Beijing: Publishing House of Electronics Industry, 2000.

[4] Qu Jinglei, Li Shaobo, Zhang Chenglong. Multisensor Data Fusion Algorithm Based on Fuzzy Evidence Theory [J]. Instrument Technique and Sensor, 2017(10): 118-122.

[5] Chhabra S, Singh D. Data fusion and data aggregation//summarization techniques in wsns: a review [J]. International Journal of Computer Applications, 2015, 121(19): 21-30.

[6] Awang A, Agarwal S. Data aggregation using dynamic selection of aggregation points based on rssi for wireless sensor networks [J]. Wireless Personal Communications, 2015, 80(2): 611-633.

[7] Rout R R, Ghosh S K. Adaptive data aggregation and energy efficiency using network coding in a clustered wireless sensor network: An analytical approach [M]. Elsevier Science Publishers B. V. 2014, 40: 65-75.

[8] Yeganeh M H, Yousefi H, Alinaghipour N, et al. RDAG: a structure-free real-time data aggregation protocol for wireless sensor networks [C]//Proc of the 17th IEEE International Conference on Embedded and Real-Time Computing Systems and Applications. 2011, 1: 51-60.

[9] Hou Xin, Zhang Dongwen, Zhong Ming. Data aggregation of wireless sensor network based on event-driven and neural network [J]. Chinese Journal of Sensors and Actuators, 2014, 27(1): 142-148.

[10] Qiu Lida, Liu Tianjian, Fu Ping. Data fusion in wireless sensor network based on sparse filtering [J]. Journal of Electronic Measurement and Instrumentation, 2015, 29(3): 352-357.

[11] Tan Dekun, Fu Xuefeng, Zhao Jia, et al. A data aggregation method based on abnormal data-driven in clusters of wireless sensor networks [J]. Chinese Journal of Sensors and Actuators, 2017, 30(2): 306-312.

[12] Lyu Tao, Zhu Qingxin, Zhu Yuyu. Energy-balanced adaptive clustering algorithm for wireless sensor network [J]. Journal of Computer Applications, 2012, 32(11): 3107-3111.

[13] Wan Yejing, Ye Jihua, Jiang Aiwen. An improved wsn energy saving strategy based on spatio-temporal correlation and anomaly detection [J]. Chinese Journal of Sensors and Actuators, 2017, 30(8): 1267-1273.

[14] Fei Huan, Li Guanghui. Abnormal Data Detection Algorithm for WSN Based on K-means Clustering [J]. Computer Engineering, 2015, 41(7): 124-128.

[15] Ding M, Chen D, Xing K, et al. Localized Fault-Tolerant Event Boundary Detection in Sensor Networks [C]//Proc of the 24th Annual Joint Conference of the IEEE Computer and Communications Societies, 2005, 2: 902-913.

[16] Liu Sifeng, Xie Naiming, FORREST Jeffery. On new models of grey incidence analysis based on visual angle of similarity and nearness [J]. Systems Engineering-Theory & Practice, 2010, 30(5): 881-887.

[17] Jiang Ting, Teng Zhaosheng, Gu Hongyan, et al. Fast fusion method of dynamic weighing based on EMD and batch estimation [J]. Chinese Journal of Scientific Instrument, 2015, 36(6): 1406-1414.

Note: Figure translations are in progress. See original paper for figures.