ISSN 1330-3651 (Print), ISSN 1848-6339 (Online) https://doi.org/10.17559/TV-20230228000387
Preliminary communication

Anomaly Detection in Wireless Sensor Networks Based on Improved GM Model

Hongzhang HAN*, Zhengjun JING

Abstract: Aiming at the problems of poor detection effect, high rate of missed detection and high rate of false detection in traditional methods, an anomaly detection method for wireless sensor networks based on improved GM model is proposed. The multilateral measurement method is used to locate the nodes in the wireless sensor network, and the state tracking of the running track of the located nodes is carried out. According to the tracking results, the self similarity between nodes is measured by Hurst index. Based on the measurement results, the improved GM model is used to predict the abnormal nodes. The abnormal values of the wireless sensor network are calculated by the distance between adjacent points, and the state of the current node is judged, this completes the anomaly detection of wireless sensor networks. The experimental results show that the proposed method is effective in anomaly detection of wireless sensor networks, and the rate of missed detection and false detection is low.

Keywords: hurst index; network anomaly; node positioning; self similarity; wireless sensor

1 INTRODUCTION

Wireless sensor network (WSN) is a self-organizing network composed of multiple sensor nodes according to the network topology. Due to the low cost and intelligence of sensor nodes designed and manufactured today, WSN is widely deployed in more dangerous and complex environments for data collection and transmission, and can also be used for automatic execution of daily tasks. Each node in the wireless sensor network can sense and process the sensor data, and transmit it to users through the base station or transit node. It has the characteristics of large-scale, autonomy, random dynamics [1, 2]. Through the data collection and limited data processing work of WSN, users can obtain the necessary environmental information in the network work area. After later data processing and analysis work, backend users can in real-time grasp the dynamics in the work area, which is conducive to effective decision-making in the future. At present, wireless sensor network has been widely used in environmental monitoring, military reconnaissance, intelligent transportation, health monitoring and other fields due to its advantages of low cost, low power consumption and easy implementation. However, due to the large number and wide distribution of network nodes, communication between nodes is limited by various factors, such as insufficient energy, signal interference, transmission delay, etc. These factors easily lead to various abnormal situations in wireless sensor networks, such as node failure, data loss, communication bottlenecks, etc. These abnormal issues can lead to a decrease in the reliability and communication efficiency of the network, affecting its application performance and service quality [3, 4]. Therefore, researchers must deeply explore the anomaly of wireless sensor networks. In terms of anomaly detection, researchers have proposed some traditional methods, such as statistical models, machine learning models, etc., to discover abnormal nodes or events [5]. These methods can effectively achieve anomaly detection, but they require significant computational resources. At present, this issue has become a relatively important issue in this field, and many experts and scholars have conducted research on it and achieved certain research results.

Wireless sensor network exceptions are caused by sensor node failures. Sensor node failures can be divided into hardware failures and software failures according to their occurrence parts and manifestations. Among them, hardware failure refers to the failure caused by the partial damage of hardware modules during the deployment or operation of sensor nodes. For example, (1) excessive energy consumption of network nodes causes the residual energy of sensor nodes to be exhausted or lower than the critical value of work, forcing them to sleep and suspend service. They can operate normally only after the energy of the node is replenished; (2) Because WSN is usually deployed by ejection or delivery, and its monitoring environment is relatively harsh and complex, the hardware of sensor node is damaged and its operation capability is lost. Once the sensor node has a hardware failure, if you want to ensure the normal operation of the network, you can only troubleshoot it [6]. Software fault refers to the sharp point of abnormal data values collected by sensor nodes, such as data deviation or error, which makes the sensor network unable to maintain normal operation. These failures are all data type failures caused by the sensor node's collection, processing and transmission of data values in an unreasonable range. Only the abnormal detection of the wireless sensor network can ensure the normal and accurate operation of the network. Therefore, in order to ensure the stable operation of each node in WSN, the anomaly detection of wireless sensor networks is studied.

Yu and Xiong [7] proposed a network anomaly detection method based on balanced iterative protocol hierarchical clustering. First, the characteristics of wireless sensor network traffic are extracted through the collection of network traffic, and then the extracted feature dimensions are reduced. Finally, according to the processing results, the normal features and abnormal features of wireless sensor network traffic are clustered using balanced iterative protocol hierarchical clustering

method to detect abnormal network traffic. Yuan et al. [8] extracts the features of the acquired wireless sensor network data, calculates the attribute similarity between adjacent nodes of the wireless sensor network using the graph model, and judges the location of abnormal data in the wireless sensor network according to the calculation results, thus completing the network data anomaly detection. However, the accuracy of the above two methods for network data anomaly detection is low, resulting in poor detection results. Lu et al. [9] proposed a network anomaly detection method based on graph signal processing to extract the network location features, so as to build a K-nearest neighbor graph signal model to obtain the wireless sensor network signal, filter it with a low-pass filter, calculate the statistical inspection quantity, and judge the location of abnormal nodes according to the calculation results. He and Liu [10] first obtain the data of network nodes, and then preprocess them. According to the preprocessing results, support vector machines are used to classify the data of wireless sensor network nodes. Based on the classification results, abnormal values of wireless sensor networks are detected. However, the above two methods for network data anomaly detection have a high rate of missed detection and false detection, resulting in poor detection results.

In view of the shortcomings of the above methods, this study proposes an anomaly detection method for wireless sensor networks based on improved GM model. In order to improve the positioning accuracy of wireless sensor network nodes, this study first estimates the location of unknown nodes through the least square method, and then uses the multilateral measurement method to complete the positioning of wireless sensor network nodes. In order to improve the state tracking effect of wireless sensor nodes, the Hurst index is used to calculate the self-similarity between nodes, and then the node classification is completed. Finally, it implements anomaly diagnosis of wireless sensor networks based on the improved GM model to improve the accuracy of its anomaly diagnosis. It is hoped that this research can accurately detect the anomaly of wireless sensor networks and lay a foundation for the security of wireless sensor networks.

2 ANOMALY DETECTION IN WIRELESS SENSOR NETWORKS

2.1 Wireless Sensor Network Node Location

The principle of node positioning in wireless sensor networks is roughly the same as that of GPS positioning. The location coordinates of nodes can be obtained by calculating the distance from a node to four or more other reference nodes. The trilateral measurement method is mostly used in node location, as shown in Fig. 1a. However, there are errors in the process of wireless sensor network node ranging, resulting in inaccurate node location. In order to accurately locate nodes, this paper estimates the location of unknown nodes through the least square method, and uses the multilateral measurement method, the transformation form of the trilateral measurement method, to locate nodes in the wireless sensor network, as shown in Fig. 1b.

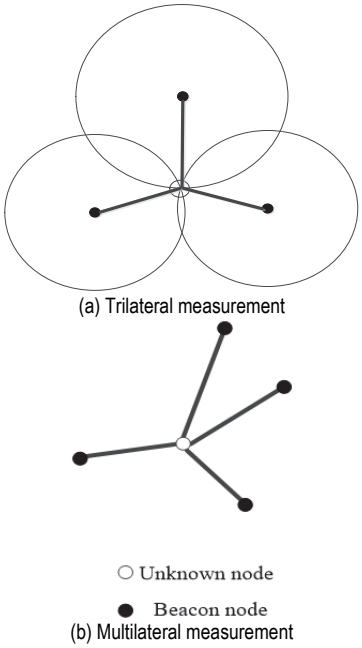

Figure 1 Schematic diagram of trilateral measurement method and multilateral measurement method

Multi-lateral measurement method is used to locate wireless sensor network nodes. First, set an unknown node as u, the coordinate as (x, y), the position coordinate of the i beacon node as (x_i, y_i), and the distance between it and u as h_i, then:

$$\begin{cases} (x_1 - x)^2 + (y_1 - y)^2 = h_1^2 \\ (x_2 - x)^2 + (y_2 - y)^2 = h_2^2 \\ \vdots \\ (x_n - x)^2 + (y_n - y)^2 = h_n^2 \end{cases} \tag{1}$$

where, i and n represent the number of beacon nodes [11].

Then, the above problem is transformed into a linearization problem, and the least square method is used to estimate the node position of wireless sensor network. The expression is:

$$(x', y') = \left(\frac{x_1 + x_2 + ... + x_k}{k}, \frac{y_1 + y_2 + ... + y_k}{k} \right) \tag{2}$$

where, k represents node connectivity coefficient. When using the least square method to estimate the position of a known node, the Error term needs to be eliminated. At this time, the elimination model is established for the double error of node self positioning error and measurement error. Therefore, the true coordinate of the k-th node position is defined as $s_k^0 = \left[x_k^0, y_k^0 \right]^T$, the node coordinate with errors is $s_k = \left[x_k, y_k \right]^T = \left[x_k + \Delta x_k, y_k + \Delta y_k \right]^T$, Δx_k and Δy_k are the errors of the k node position coordinates in the x and y directions, respectively, and the distance difference estimation error is $\Delta d_{kr} = c(n_k - n_r) = cn_{kr}$, then the Error term elimination model at this time is:

$$\left(G_a + \Delta G\right) z = h_a \Delta d_{kr} + \Delta h s_k \tag{3}$$

2.2 Node State Tracking in Wireless Sensor Networks

It tracks the running track of the wireless sensor network node. First of all, suppose that in the set of wireless sensor network nodes, all the running states of the constituent nodes are static [12]. The location coordinate of the node itself is (x_{st}, y_{st}), where t represents the number of network sensors. When the node is within the scanning range of t sensors, the measurement value that the node can track is calculated as follows:

$$z_t\left(k\right) = \sqrt{x_t\left(k\right) {x_{st}}^t + y_t\left(k\right) {y_{st}}^t} \tag{4}$$

where, $z_t\left(k\right)$ represents the measured value of the node, and (x_t, y_t) represents the position coordinate of the key sequence of the node. $\left(x_t\left(k\right), y_t\left(k\right)\right)$ represents the measured value corresponding to the key sequence, and $\left({x_{st}}^t, {y_{st}}^t\right)$ represents the position coordinate when the number of network sensors is t. Suppose that the number of task sensors in the wireless sensor network is L in k period, and all of these task sensors participate in the tracking task of the node [13].

Due to the average consistency problem caused by similar network node characteristics in the tracking process of wireless sensor network nodes, node state tracking errors occur. To solve this problem, a consistent hash algorithm can be used to solve the problem, resulting in the following expression:

$$x_t^{k+1} = \frac{x_t^{(k)} \cdot x_j^{(k)}}{\delta} \tag{5}$$

wherein, x_t^{k+1} represents the average tracking value of network nodes with period k+1. $x_t^{(k)}$ and $x_j^{(k)}$ respectively represent the network node state and the next motion state when the number of nodes is t, and δ represents the average motion step of the node, and its expression is:

$$\delta = \frac{\varpi}{\left\| \left(c\right)^T \right\|} \tag{6}$$

where, $\left\| \left(c\right)^T \right\|$ represents the covariance matrix of target tracking nodes, and ϖ represents the average value of discrete nodes [14].

2.3 Self Similarity Measurement Between Nodes

Due to the self-similarity of node attribute characteristics in wireless sensor networks, nodes can be classified according to the similarity of attribute characteristics to lay the foundation for subsequent abnormal node prediction. Therefore, this paper measures the similarity between nodes by Hurst exponent H ($0.5 < H < 1$) according to the above node state tracking results of wireless sensor networks [15].

Set a sliding time window Δt to construct the sensing vector of the wireless sensor network node through this window:

$$D_i\left(t\right) = \left\{d_i\left(t - \Delta t + 1\right), d_i\left(t - \Delta t + 2\right), ..., d_i\left(t\right)\right\} \tag{7}$$

where, $d_i\left(t\right)$ represents the sampling value of the i node at the moment.

The maximum minimum normalization method is used to normalize the sensing vector obtained above, and the expression is:

$$\delta = \frac{\left(d - d_{\min}\right)}{\left(d_{\max} - d_{\min}\right)} \tag{8}$$

where, d is the original value of the numerical attribute of the sensing data instance, and d_{max}, d_{min} is the maximum and minimum value of the attribute in the sensing vector [16]. At this point, the formula for calculating the attribute eigenvalues of wireless sensor network nodes is:

$$W_n = \lambda\left(\delta W_{n-1} + K_n e_n\right) \tag{9}$$

In the equation, K is the gain factor, λ is the forgetting factor, and e_n is the error covariance.

According to the characteristics of wireless sensor network nodes, this paper uses Hurst index to measure the self similarity between wireless sensor network nodes:

$$j_{k,l} = \frac{x_k * x_l}{\left\|x_k\right\|^2 + \left\|x_l\right\|^2 - x_k * x_l} \tag{10}$$

where, k, l represents two adjacent wireless sensor network nodes, and x_k, x_l represents the normalized sensing vector of the node [17].

According to the self similarity between nodes in wireless sensor networks, the nodes are classified, and the expression is:

$$x\left(t\right) = 1 - j_{k,l} D_i\left(t\right) \tag{11}$$

2.4 Prediction of Abnormal Nodes in Wireless Sensor Networks Based on Improved GM Model

(1) GM model fundamentals.

GM (1, 1) model is a grey prediction model, and its establishment steps are as follows.

Step 1: feasibility analysis whether GM (1, 1) model can be modeled. Whether it is possible to build a high-precision GM (1, 1) prediction model, it is necessary to calculate the grade ratio $\sigma^{(0)}\left(k\right)$ of sequence $x^{(0)}$ and judge its interval, that is, whether its battery limit is within the appropriate area [15].

According to the pre inspection criteria, the following can be set:

$$x^{(0)} = \left(x^{(0)}(1), x^{(0)}(2), ..., x^{(0)}(n) \right) \tag{12}$$

The grade ratio $\sigma^{(0)}(k)$ is:

$$\sigma^{(0)}(k) = \frac{x^{(0)}(k-1)}{x^{(0)}(k)} \tag{13}$$

When $\sigma^{(0)}(k) \in \left(e^{-\frac{2}{n+1}}, e^{\frac{2}{n+1}} \right)$, then grey prediction $x^{(0)}$, its accuracy is in line with the requirements. On the contrary, $x^{(0)}$ needs to be transformed to meet the requirements [16-18].

Step 2: Data exchange processing.

The data processing makes the sequence level ratio reach the tolerable range, so that the unqualified sequence can meet the requirements after general selection of data conversion, and can be used for GM (1, 1) model modeling.

The third step: the modeling of GM (1, 1).

According to Eq. (10), set the original sequence:

$$x^{(0)} = (x^{(0)}(1), x^{(0)}(2), ..., x^{(0)}(n)) \tag{14}$$

The modeling of GM (1, 1) is divided into five steps:

(1) Accumulate the original sequence $x^{(0)}$ (1-AGO), and the expression is:

$$x^{(1)} = (x^{(1)}(1), x^{(1)}(2), ..., x^{(1)}(n)) \tag{15}$$

where, $x^{(1)}(k) = \sum_{i=1}^{k} x^{(0)}(i)$.

(2) As an equal weight generating sequence of the adjacent mean value of $x^{(1)}$, let:

$$z^{(1)} = \left(z^{(1)}(1), z^{(1)}(2), ..., z^{(1)}(n) \right) \tag{16}$$

In formula, $z^{(1)}(k) = \frac{1}{2}\left(x^{(1)}(k) + x^{(1)}(k-1) \right)$.

Set the development coefficient of GM (1, 1) model as a and the grey action amount as b, then the original form of GM (1, 1) model is:

$$b = x^{(0)}(k) + az^{(1)}(k) \tag{17}$$

Convert the above equation to albino differential equation:

$$b = \frac{\mathrm{d}x^{(1)}}{\mathrm{d}t} + ax^{(1)} \tag{18}$$

(3) Use the least square method to solve the parameter a, b:

$$a' = (a, b)^T \tag{19}$$

$$a' = (B^T, B)^{-1} B^T Y = \begin{bmatrix} a \\ b \end{bmatrix} \tag{20}$$

where:

$$B = \begin{bmatrix} -z^{(1)}(2) & 1 \\ -z^{(1)}(3) & 1 \\ \vdots & \\ -z^{(1)}(n) & 1 \end{bmatrix} \tag{21}$$

$$Y = \begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \vdots \\ x^{(0)}(n) \end{bmatrix} \tag{22}$$

(4) The time response sequence of formula $b = x^{(0)}(k) + az^{(1)}(k)$ in GM (1, 1) model is:

$$x'^{(1)}(k+1) = \left(x^{(1)}(1) - \frac{b}{a} \right) e^{-ak} + \frac{b}{a} \tag{23}$$

Establish whitening response function:

$$x'^{(1)}(t) = \left(x^{(1)}(1) - \frac{b}{a} \right) e^{-at} + \frac{b}{a} \tag{24}$$

(5) Under the initial condition $x'^{(1)} = x'^{(1)}(1) = x^{(0)}(1)$, it is obtained that:

$$x'^{(1)}(k+1) = \left(x^{(0)}(1) - \frac{b}{a} \right) e^{-ak} + \frac{b}{a} \tag{25}$$

According to the above formula, the cumulative reduction results in:

$$x'^{(0)}(k) = x'^{(1)}(k) - x'^{(1)}(k-1) \tag{26}$$

Step 4: Model inspection.

After the GM (1, 1) model is built, the accuracy of the model needs to be tested. In this paper, the residual test method is used to test the model. The expression is:

$$q(k) = x^{(0)}(k) - x'^{(0)}(k) \tag{27}$$

where, $x^{(0)}(k)$ represents the actual value of point k and $x'^{(0)}(k)$ represents the predicted value.

Step 5: Modeling and forecasting.

The GM (1, 1) model that has passed the test is used for prediction, and the prediction conclusion is drawn [22, 23].

(2) Prediction of abnormal nodes based on improved GM model.

Because the fluctuation of abnormal nodes in wireless sensor networks is a random process of change, and the

rule of change is relatively complex, it is not a simple linear system, but an open complex system composed of nonlinear interaction factors. Its development trend is dynamic and unstable. The GM-ARMA combined model has a high prediction accuracy. It is based on the data processing after the gray system development prediction, and it is a modified judgment of the data error, which is controllable for the prediction. Therefore, the predicted data is basically consistent with the original data, with good accuracy. And it has strong time correlation and real-time dynamics. The prediction can be carried out based on a small amount of data. The data information is easy to find and sort out, and the calculation efficiency is high. It can predict the data development trend of the next several times based on the data of the last several times. The data has strong coupling, strong timeliness and real-time. Because ARMA model prediction is to continuously obtain new prediction value and new residual sequence value through GM (1, 1), calculate a new parameter, and establish new dynamic relationship. The prediction results are alternately followed and pushed forward, and the adjustment of parameters is used to adapt to the fluctuation trend of abnormal nodes in wireless sensor networks, with dynamic updating.

Therefore, for the non-stationary time series information with large fluctuation of abnormal nodes in wireless sensor networks, this paper uses the ARMA model to improve the GM model, and uses the GM-ARMA combination model to predict abnormal nodes in wireless sensor networks. On the contrary, when the volatility of the time series is small and stable, the fitting process is cancelled during the prediction process to reduce the prediction time. The specific steps to improve the GM-ARMA model are as follows:

(1) For the given original time series $X = (x_1, x_2, \ldots, x_n)$ of abnormal nodes in the wireless sensor network, the grey prediction abnormal node sequence is first calculated through the grey GM (1, 1) model:

$$Z' = (z'_1, z'_2, ..., z'_n) \tag{28}$$

where, Z' is X' and the grey prediction value Z'_{n+1} at the next point $n + 1$,

(2) Based on the GM model, the stationarity of the original abnormal node data sequence is tested, and an ARMA model of the grey residual sequence is constructed:

$$\begin{aligned} &Y = X - Z' = \\ &= (x_1 - z'_1, x_2 - z'_2, ..., x_n - z'_n) = \\ &= (y_1, y_2, ..., y_n) \end{aligned} \tag{29}$$

The construction of ARMA model usually involves the following steps:

The first step is to test the stability of the original abnormal node data sequence, and judge whether it meets the requirements by observing its time trend chart. If it does not meet the conditions, it will be subject to differential transformation or other methods. The common method is logarithmic transformation and differential treatment. The values of unknown parameters d, p, q are determined by calculating some characteristic statistics of abnormal node data sequence, and are tested until they meet the requirements. Finally, the required prediction model is constructed to predict and estimate the residual.

The second step is to estimate the parameters of the model built in the previous step, and then test the significance level of the parameters to analyze whether the model is reasonable.

The third step is to diagnose and analyze whether the ARIMA model empirical results are consistent with the characteristics of the original data to determine whether the data can be well fitted. Then, the residual test is performed on the model to determine whether the residual predicted by the model is white noise. If it is white noise, it can be used for data prediction and fitting. On the contrary, repeat the above steps.

The fourth step is to use the final model to make prediction analysis. The autoregressive ARMA model is built to predict the prediction value y'_{n+1} of the next point $n + 1$ based on the abnormal node data of the grey residual sequence Y, then:

$$X = Y + Z' = (y_1 + z'_1, y_2 + z'_2, ..., y_n + z'_n) \tag{30}$$

(3) The prediction of abnormal nodes in wireless sensor networks based on the improved GM model can be expressed as:

$$x'_t = y'_t + z'_t \tag{31}$$

2.5 Network Anomaly Detection

According to the prediction results of the above abnormal nodes, the wireless sensor network anomaly is detected. The distance between adjacent points is used to judge the abnormal value of wireless sensor network. If the distance between the current wireless sensor network node and the node at the previous time exceeds a certain range, the node is considered as an abnormal point. The schematic diagram of abnormal values is shown in Fig. 2.

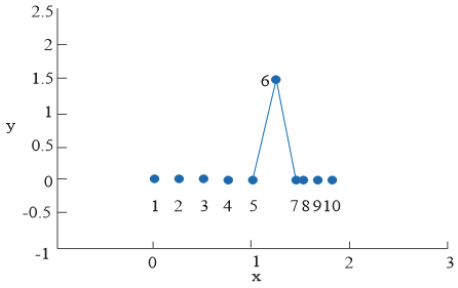

Figure 2 Schematic diagram of abnormal values

As shown in Fig. 2, nodes 1, 2, 3, 4 and 5 are normal nodes, and node 6 is obviously an abnormal node. However, the distance between 6 and 7 is almost the same as that between 6 and 5. If an abnormal node is judged according to the distance between adjacent points, 7 will also be detected as abnormal. However, as shown in Fig. 2, node 7 is a normal node, which obviously has a false positive. To avoid suspicion of this phenomenon, the following definitions are made.

Define 1 (Normal Hop Distance) Node *obj*(*k*) as the node captured by the sensor at time *t*. Suppose that among

the wireless sensor network nodes captured before time t, the node that is closest to $obj(k)$ in time has been detected as normal, and the Mahalanobis distance from node $obj(k)$ to node $obj(i)$ is $dist(obj(k), obj(i))$, then the normal hop distance of current node $obj(k)$ is:

$$NHD(k) = \frac{dist(obj(k), obj(i))}{k - i} \tag{32}$$

Definition 2 (Outlier Factor): If the node captured by the wireless sensor at t moment is $obj(k)$ and its corresponding dynamic threshold is $\sigma(k)$, then the anomaly factor of the current node $obj(k)$ is:

$$OF(k) = \frac{NHD(k)}{\sigma(k)} \tag{33}$$

According to the size of node anomaly factor, each wireless sensor network node is divided into three states:

Normal state: if the abnormal factor $OF(k) \in [0, 1]$ of node $obj(k)$;.

Critical state: if the abnormal factor $OF(k) \in 1$ *trans value* , of node $obj(k)$.

Abnormal state: if the abnormal factor $OF(k) \in (trans\ value, +\infty)$ of node $obj(k)$.

where, *trust value* is a parameter greater than 1.

For wireless sensor network node $obj(k)$, the anomaly factor $OF(k)$ can be obtained when the corresponding threshold value $\sigma(k)$ is known, and the current node's state can be judged based on this. If $obj(k)$ is in an abnormal state, it is judged to be an abnormal value, indicating that the wireless sensor network is abnormal; On the contrary, it is judged to be a normal value, indicating that the wireless sensor network is normal.

3 SIMULATION EXPERIMENT ANALYSIS

In order to verify the effectiveness of the anomaly detection method proposed in this paper based on the improved GM model in practical applications, a wireless sensor network platform is built.

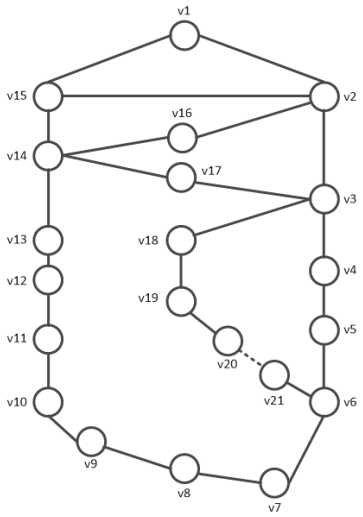

Figure 3 Wireless sensor network topology

In the Ubuntu environment, deploy the Open Flow controller Flood Light. Install Virt Box, create three virtual machines, and build the Mini Net platform respectively. After the construction is successful, enter the Mini Net platform. In this way, the controller Flood Light can connect three Mini Net switches to form a large wireless sensor network, there are 21 types of network attribute nodes. The network topology is shown in Fig. 3.

The hardware and software configurations related to the experimental platform are shown in Tab. 1.

Table 1 Operation configuration of communication network information system

name	model	quantity
backbone switch	UGS-1224T/F	1
Floor access main switch	UGS2402N	30
Wireless controller	ULM-302T	2
Wireless access point	ULM-302GB	90
Wireless gateway	Ruijie Network RG-EG580-W	3
Multimode optical interface module	GLC SX MM,850nm,500m	60
Single mode optical interface module	GLC LH SM,1310nm,10km	15
Stacking module	UFS3264I	15
Broadband access server	SE800400	5
Antivirus software	Kaspersky PURE V13.0.2.558	1
Lenovo Desktop	AMD R7-4800U 16G 512G SSD	10
Lenovo portable terminal	i5-1135G7 16G 512G 100%sRGB	5
Disk enclosure hard disk expansion	Huawei OceanStor 5300 V3, 2.5 inch	10

In the above experimental environment, the methods in this paper, Yu et al. [4, 5] were first used to detect the localization effect of wireless sensor network nodes. Assuming an initial number of 20 beacon nodes and a ranging radius of 20 m, verify the positioning accuracy of the three methods for wireless sensor network nodes. The test results are shown in Fig. 4.

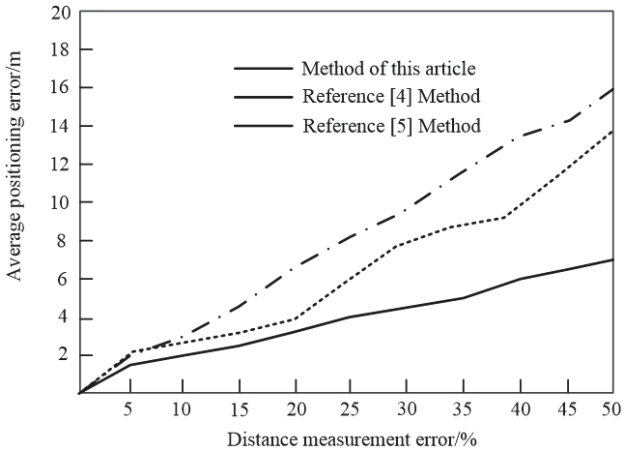

Figure 4 Comparison of three methods for locating wireless sensor network nodes

As shown in Fig. 4, under the same conditions, the positioning error of the proposed method is smaller than that of the two comparison methods, and the positioning error of the proposed method gradually decreases as the ranging error increases. This proves that the positioning effect of the proposed method is better. Then, abnormal nodes in the wireless sensor network are detected to verify the detection accuracy of the three methods. The test results are shown in Fig. 5.

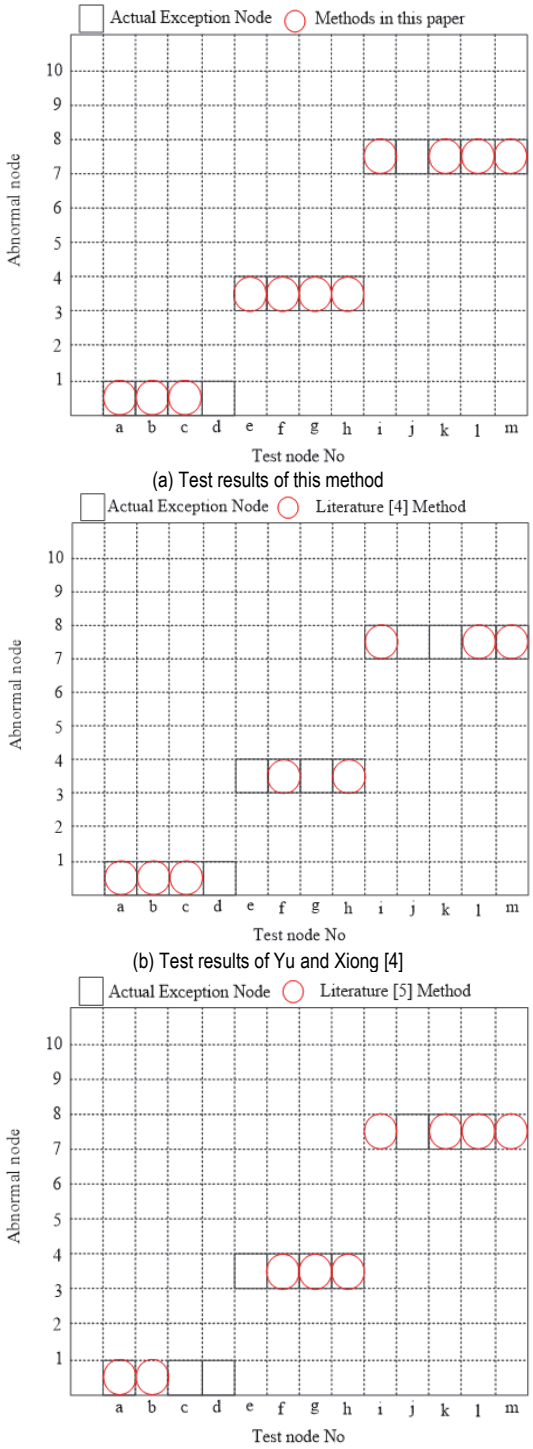

Figure 5 Comparison of abnormal node detection results of three methods in wireless sensor networks

According to Fig. 5, the results of abnormal node detection in wireless sensor network by this method are consistent with the actual test results. The results of abnormal node detection in wireless sensor network by using the methods in Yu and Xiong [7] and Yuan et al. [8] differ greatly from the actual test results, indicating that the detection accuracy of this method is high and the detection effect is good.

The methods in this paper, Yu and Xiong [7] and Yuan et al. [8] are used to detect abnormal nodes in wireless sensor networks, and the missed detection rate and false detection rate of the three methods are compared. The comparison results are shown in Fig. 6 and Fig. 7.

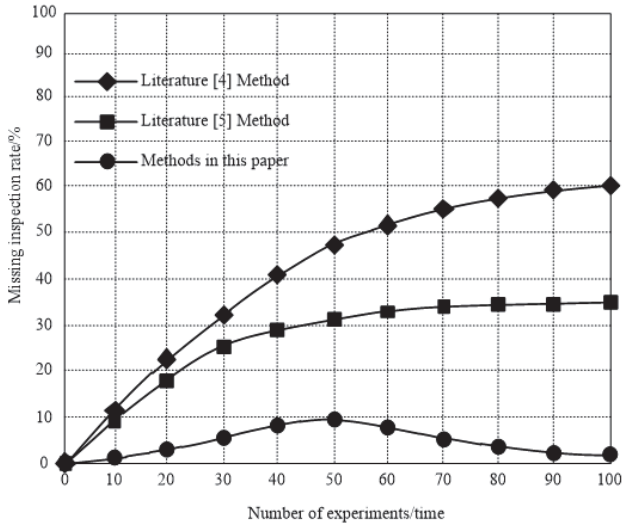

Figure 6 Test results of leakage rate

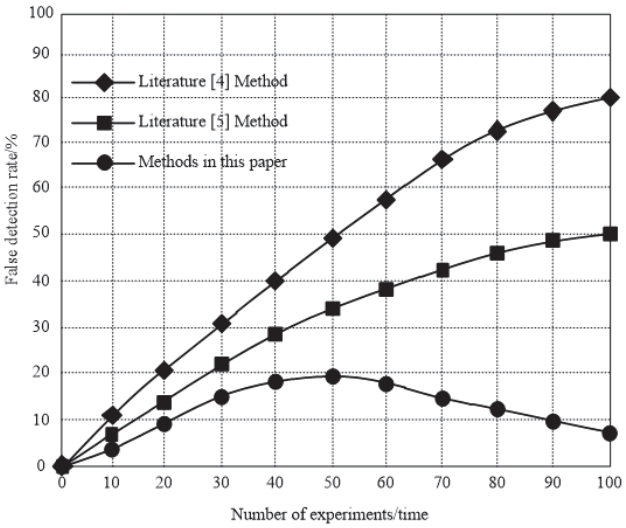

Figure 7 Test results of false detection rate

According to Fig. 6 and Fig. 7, the undetected rate of abnormal node detection in wireless sensor network using this method is within 10%, and the false detection rate is within 20%. The undetected rate of abnormal node detection in wireless sensor network is less than 60% and the false detection rate is less than 80% by using the method in Yu and Xiong [7]. The undetected rate of abnormal node detection in wireless sensor network is less than 35% and the false detection rate is less than 50% using the method in Yuan et al. [8]. This shows that the method in this paper has the lowest rate of missed detection and false detection in the detection of abnormal nodes in wireless sensor networks, indicating that the method in this paper has the best detection effect. The reason for this phenomenon is that the multilateral measurement method is used to locate the wireless sensor network nodes in this study. When anomaly detection is performed, the ARMA model is used to improve the GM model, which greatly improves the model's ability to analyze the fluctuation time series, quickly excavates the periodicity and trend of abnormal node fluctuations in the wireless sensor network,

and greatly improves the prediction accuracy, so its detection effect is the best.

The methods in this paper, Yu et al. [4] and Yuan et al. [5] are used to detect abnormal nodes in wireless sensor networks, and the missed detection rate and false detection rate of the three methods are compared. The comparison results are shown in Tab. 2.

Table 2 Comparison of detection time under different methods/s

number of times	Method of this article	Yu et al. [4] Method	Yuan et al. [5] Method
100	10.5	50.6	63.2
200	15.6	71.4	69.5
300	20.4	76.9	70.6
400	25.7	80.1	76.3
500	30.0	93.5	80.6

From Tab. 2, it can be seen that under a certain number of experiments, the detection time is about 20.44 seconds when using the method proposed in this paper, and about 74.5 seconds when using the method proposed in reference [4]. When using the method in reference [5], the detection time is about 72.04 seconds. The above three methods have shown an increase in detection time as the number of detections increases, but their overall detection time has decreased by 54.06 s and 52 s compared to the methods in reference [4, 5], respectively, with a slow growth rate and certain advantages.

4 CONCLUSIONS

In order to improve the effect of anomaly detection in wireless sensor networks, an improved GM model is used to study anomaly detection methods for wireless sensor networks. The multilateral measurement method is used to locate wireless sensor network nodes and track their operation status. Based on the tracked wireless sensor network node status, Hurst index is used to calculate the self similarity between wireless sensor network nodes. According to the calculation results, the ARMA model is used to improve the GM model. The GM-ArMA combined model is used to predict the abnormal nodes of the wireless sensor network, and the distance between adjacent points is used to judge the abnormal nodes of the wireless sensor network, so as to complete the abnormal detection of the wireless sensor network. Experimental results show that the detection results of abnormal nodes in the wireless sensor network by the proposed method are consistent with the actual test results, and the missed detection rate of abnormal nodes in the wireless sensor network is within 10%, which is 50% and 25% lower than that of the methods in Yu et al. [4, 5] respectively. However, the false detection rate of abnormal nodes in wireless sensor networks applied by the proposed method is less than 20%, which is 60% and 30% lower than that of the methods in Yu et al. [4, 5] respectively. Therefore, the proposed method has the best detection effect. Although this method can reduce the missed detection rate and false detection rate of abnormal node detection in wireless sensor networks, there are still some shortcomings. In the future, it will be further optimized to diagnose the causes of abnormal nodes in wireless sensor networks, and then improve its anomaly detection effect.

Acknowledgements

Supported by the National Natural Science Foundation of China,Project No.:61602216;Project Name:Study on Guarantee Protocols of Cross-layer QoS for Low Duty-Cycled Wireless Sensor Networks.

5 REFERENCE

[1] Sun, X. (2020). Node identification of abnormal energy loss in wireless sensor networks. *Computer Engineering and Design*, *41*(10), 2724-2728.

[2] Xu, L. & Li, G. H. (2019). Multi-protocol layer intrusion detection method for wireless sensor networks based on trust mechanism. *Chinese Journal of Sensors and Actuators*, *32*(5), 739-748.

[3] Zhang, X. (2021). Method for detecting data anomaly of wireless sensor network of martin system. *New Generation of Information Technology*, *4*(5), 31-38. https://doi.org/10.1186/s13638-021-01902-w

[4] Yu, B. & Xiong, J. (2022). Anovel WSN traffic anomaly detections chemebased on BIRCH. *Journal of Electronics & Information Technology*, *44*(1), 305-313.

[5] Yuan, J., Wang, X., & Pan, Z. (2019). Anomaly detection in wireless sensor networks. *Electronic Technology & Software Engineering*, *24*, 10-11. https://doi.org/10.3390/molecules24112104

[6] Lu, G., Zhou, L., Lyu, S., Shi, C., & Su, K. (2020). Out lier node detection algorithmin wireless sensor net works based on graph signal processing. *Journal of Computer Applications*, *40*(3),783-787.

[7] He, Y. & Liu, Y. (2021). Outlier detection algorithm based on machine learning in wireless sensor networks. *lectronic Technology & Software Engineering*, *2*, 40-41.

[8] Feng, Q. (2020). Anomaly detection and analysis of sensor network databased on hypersphere support vector machine. *Microcomputer Applications*, *36*(10), 174-176.

[9] Wang, L. & Tai, Q. (2021). Outlier detection in wireless sensor networks based on high-order Markovchain. *Journal of Heilongjiang Universityof Technology (Comprehensive Edition)*, *8*, 93-97.

[10] Lin, T., Fu, C. & Ji, M. (2021). Sensor fault pre-detection method based on improved self-attention with long short-term memory. *Journal of Computer Applications*, *41*(S01), 31-35.

[11] Zhang, Y., Cao, J. & Kan, Z. (2020). Network intrusion detection method based on CPS system. *Transducer and Microsystem Technologies*, *39*(10), 126-128.

[12] Monshizadeh, M., Khatri, V., Gamdou, M., Kantola, R., & Yan, Z. (2021). Improving data generalization with variational auto encoders for network traffic anomaly detection. *IEEE Access*, *9*, 56893-56907. https://doi.org/10.1109/ACCESS.2021.3072126

[13] Wang, Z., Luo, N., & Zhou, P. (2020). Guard Health: Blockchainempoweredsecuredatamanagementand Graph Convolutional Network enabled anomaly detection in smart healthcare. *Journal of Parallel and Distributed Computing*, *142*, 1-12. https://doi.org/10.1016/j.jpdc.2020.03.004

[14] Tang, Z., Chen, Z., Bao, Y., & Li, H. (2019).Convolutional neural network-based data anomaly detection method using multiple information for structural health monitoring. *Structural Control and Health Monitoring*, *26*(1), e2296. https://doi.org/10.1002/stc.2296

[15] Bao, Y., Tang, Z., Li, H., & Zhang, Y. (2019). Computer vision and deep learning-based data anomaly detection

method for structural health monitoring. *Structural Health Monitoring*, *18*(2), 401-421. https://doi.org/10.1177/1475921718757405

[16] Nakashima, M., Sim, A., Kim, Y., Kim, J., & Kim, J. (2021). Automated feature selection for anomaly detection in network traffic data. *ACM Transactions on Management Information Systems(TMIS)*, *12*(3), 1-28. https://doi.org/10.1145/3446636

[17] Venskus, J., Treigys, P., Bernatavičienė, J., Tamulevičius, G., & Medvedev,V. (2019). Real-time maritime traffic anomaly detection based on sensors and history data embedding. *Sensors*, *19*(17), 3782. https://doi.org/10.3390/s19173782

[18] Cai, Y., Shyu, M. L., Tu, Y. X., Teng, Y. T., & Hu, X. X. (2019). Anomaly detection of earth quake precurs or data using long short-term memory networks. *Applied Geophysics*, *16*, 257-266. https://doi.org/10.1007/s11770-019-0774-1

[19] Deng, X. (2019). Research on the anomaly detection method in intelligent patrol based on big data analysis. *Journal of Computer and Communications*, *7*(8), 1-7. https://doi.org/10.4236/jcc.2019.78001

[20] Zhao, L. & Shi, G. (2019). Maritime anomaly detection using density-based clustering and recurrent neural network. *The Journal of Navigation*, *72*(4), 894-916. https://doi.org/10.1017/S0373463319000031

Contact information:

Hongzhang HAN
(Corresponding author)
School of Computer Engineering, Jiangsu University of Technology,
Changzhou 213001, China
E-mail: hhz@jsut.edu.cn

Zhengjun JING
School of Computer Engineering, Jiangsu University of Technology,
Changzhou 213001, China
E-mail: jzjing@jsut.edu.cn