International Journal of Automation and Computing 12(4), August 2015, 402-412
DOI: 10.1007/s11633-014-0847-9

An Isolation Principle Based Distributed Anomaly Detection Method in Wireless Sensor Networks

Zhi-Guo Ding[1,2] Da-Jun Du[1] Min-Rui Fei[1]

[1]Shanghai Key Laboratory of Power Station Automation Technology, School of Mechatronics Engineering and Automation, Shanghai University, Shanghai 200072, China

[2]College of Mathematics, Physics and Information Engineering, Zhejiang Normal University, Jinhua 321004, China

Abstract: Anomaly detection plays an important role in ensuring the data quality in wireless sensor networks (WSNs). The main objective of the paper is to design a light-weight and distributed algorithm to detect the data collected from WSNs effectively. This is achieved by proposing a distributed anomaly detection algorithm based on ensemble isolation principle. The new method offers distinctive advantages over the existing methods. Firstly, it does not require any distance or density measurement, which reduces computational burdens significantly. Secondly, considering the spatial correlation characteristic of node deployment in WSNs, local sub-detector is built in each sensor node, which is broadcasted simultaneously to neighbor sensor nodes. A global detector model is then constructed by using the local detector model and the neighbor detector model, which possesses a distributed nature and decreases communication burden. The experiment results on the labeled dataset confirm the effectiveness of the proposed method.

Keywords: Distributed anomaly detection, isolation principle, light-weight method, ensemble learning, wireless sensor networks (WSNs)

1 Introduction

Wireless sensor networks (WSNs) have been used in various applications, such as environment monitoring, habitat monitoring, health and medical monitoring, industrial monitoring, target tracking, surveillance monitoring, etc.[1−6] In these applications, the most important problem is how to guarantee data reliability and availability. However, the sensor nodes, which are the core component in WSN, are prone to fail. Therefore, anomalous observation is induced by considering the following reasons[7]:

1) The intrinsic characteristics of sensor nodes are low cost and limited resource, such as energy (battery power), memory, computation capability and communication bandwidth, etc.

2) Sensor nodes are usually randomly deployed in a large and hard area. Therefore, the observations are frequently susceptible to the environmental effects.

3) There exist common interference and malicious attack in WSNs. Consequently, the information transmission is unreliable.

The above mentioned issues may lead to the data with noise, errors, etc. This makes the data inaccurate and unreliable, which cannot reflect the real situations and thus affects the final decision. Anomaly detection, as the active monitoring measure tool, plays an important role in ensuring the data quality. Therefore, effectively and efficiently identifying anomalies can help to improve the quality of collected data before transmitting to the station and have correct decision-making of events of interest.

In WSNs, radio transmission among nodes is the main reason of quick energy depletion. The energy consumption for transmission of one bit data is significantly more than to process thousands of bits in sensors network[8]. So the anomaly detection methods developed for WSNs must take into account some inherent characteristics, e.g., massive communication will drain the life of batteries quickly and shorten the life-span of the whole WSNs.

In this paper, considering the constrained resource characteristic in each sensor node and data spatial correlation among neighbor sensor nodes, a light-weight distributed learning technique based on the isolation principle is proposed for the anomalous data detection in WSNs.

The proposed model overcomes the currently existing models′ disadvantages and has some advantages as follows:

1) It is a distributed anomaly detection algorithm and can handle the dataset collected in WSNs.

2) It has high detection accuracy and low false alarm ratio.

3) The proposed detection model does not require any distance or density measurement, which saves computation resource.

4) This algorithm can handle the anomaly and events locally and globally in WSNs.

The main contribution of this paper includes:

Regular paper
Manuscript received October 9, 2013; accepted March 28, 2014
This work was supported by the National High Technology Research and Development Program of China (No. 2011AA040103−7), the National Key Scientific Instrument and Equipment Development Project (No. 2012YQ15008703), the Zhejiang Provincial Natural Science Foundation of China (No. LY13F020015), National Science Foundation of China (No. 61104089), Science and Technology Commission of Shanghai Municipality (No. 11JC1404000), and Shanghai Rising-Star Program (No. 13QA1401600).
Recommended by Guest Editor Xin Sun

1) Considering the spatial correlations of neighbor nodes in WSNs, a new distributed anomaly detection method is proposed based on the ensemble isolation principle.

2) Theoretical analysis and complexity analysis confirm the efficiency of the proposed method.

The rest of this paper is organized as follows. Section 2 summaries the taxonomy of state-of-the-art anomaly detection methods in WSNs and related works. Section 3 describes the problem of anomaly detection in WSNs and proposes distributed anomaly detection method based on isolation principle. Section 4 presents the theory and complexity analysis. In Section 5, the experiments are performed to evaluate the efficiency of the proposed algorithm. Finally, conclusions and future works are presented.

2 Related works of anomaly detection in WSNs

Anomaly detection, or outlier detection, or deviation detection, is an activity to measure an observation (or subset of observations) that appears to be inconsistent with the remainder of that dataset[9, 10]. In the context of WSNs, the anomaly is related to those observations which do not conform to the defined (expected) normal behavior of the data. Based on this definition, the anomalies always correspond to sensor faults, observation data errors, unexpected events and some malicious attacks[5, 11]. To some extent, anomalies also correspond to some interesting events such as sudden changes in the monitored surrounding that result in the unusual observations. In WSNs, according to the scope of anomaly occurrence, anomalies can be at the level of individual measurements with respect to the other measurements at the same sensor node, or at the level of the measurements of one node with respect to other sensor nodes in the network[12].

There are many anomaly detection methods in WSNs. From the perspective of techniques employed, the anomaly detection methods can be categorized into statistical based, nearest neighbor based, clustering based, classification based and spectral decomposition based approaches. Statistical based methods can be further categorized into parametric and non-parametric methods based on how the probability distribution model is built[5]. From the perspective of using the label information of the training dataset during the process of building anomaly detector, these methods can be categorized as supervised, semi-supervised and unsupervised[13]. From the topology structure of WSN deployed, these methods can be classified as hierarchical method and flat method. For the former, the method can be further categorized as statistical based method, data mining and computational intelligence based method, game theory based method and hybrid method. And for the latter, these methods can be categorized as rule based method, statistical based method as well as graph based method. From the perspective of data processing, these methods can be categorized as online or offline methods. From the perspective of node data processing behavior, these methods can be categorized as distributed or centralized, etc. These methods may have some overlap, and interested reader can further review it in [3, 5, 13]. Though these above mentioned methods have gotten successful applications in many fields, there are more or less some disadvantages, e.g., statistic based method requires the knowledge of data distribution in advance or difficult to handle multi-dimensional dataset, some density or distance based methods require to calculate the distance or density among observations which requires massive computation resources. Some machine learning method, such as artificial neural network (ANN) or support vector machine (SVM), requires massive time to train the detector model, or some method cannot handle the high dimensional data in the constrained computation and storage resource, etc.

Recently, based on the ensemble learning[14], an isolation based method is proposed[15, 16], which purely considers the intrinsic characteristic of anomalous data to build detector. As we all known, anomalies are the minority consisting of fewer instances and have attribute values that are very different from those of normal instances. In other words, anomalies are "few and different", which make them more susceptible to be isolated than normal data points. This method does not employ any distance or density measure, fundamentally different from most existing model-based methods to detect anomalous data. Fig. 1 demonstrates the initial idea of anomaly detection method based on the isolation principle.

During the procedure of anomaly detection, random partitioning of instances according to the attribute value is repeated recursively until all the instances are isolated or they meet a pre-defined termination criterion. Then those data points that are easily isolated are regarded as the anomalies (black spot x_0) with high possibility. Consequently, the more easily an observation data point is isolated, the more probability it carries to be regarded as the anomaly data point. One of the many advantages of this method is that the anomaly detection doesn′t require any distance or density computations. What′s more, unlike these methods which provided a hard decision for the observation whether it is anomaly or not, isolation based method assigns an anomalous score to each of the testing observation. Local outlier factor (LOF)[17] is also proposed to denote the degree of the object being an anomaly. Unfortunately, this algorithm has complex procedures and needs a high computation cost, the computation complexity reaches the order $O(n^2)$ under the context of high data dimensions. Consequently, isolation based method may be more valuable in the real application. Another advantage is that isolation based anomaly detection employed the ensemble learning strategy[18]. So instead of wasting much time to train the strong detector, it only trains multiple weak detector models and aggregates them into a final ensemble detector with high performance. Though ensemble learning method is not a light-weight method, in our proposed method, a distributed process is employed to overcome this shortcoming successfully.

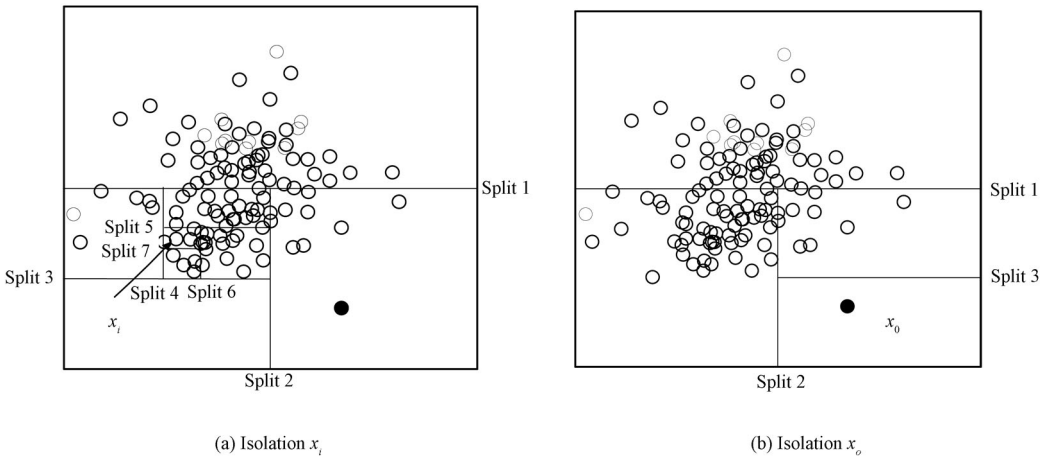


Fig. 1 Anomaly detection based on isolation principle

3 Proposed anomaly detection model for WSNs

3.1 Problem statement

Sensor nodes in WSNs may be arranged in multiple kind of topology for different applications. A commonly used network topology is shown in Fig. 2, which is clustering-based hierarchal network. Generally, there can be multiple levels of cluster heads and sensors in the network. In order to make our proposed algorithm more general and can be adapted to accommodate different contexts as well. We define a sensor sub-network to describe the problems.

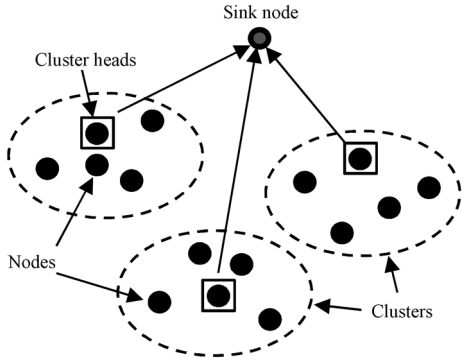


Fig. 2 Topology of typical sensor network

Definition 1. (sensor sub-network). The sensor sub-network is one of the clusters in the whole network where the nodes can communicate directly with each other. Its topology can be modeled as an undirected graph G, G=(N, E), where N denotes the node set and E denotes the edge set. The edge of connecting two nodes exists if the two nodes in the network are within radio transmission range.

To concisely describe our proposed anomaly detection method for WSNs, this relatively small sub-network consists of some sensor nodes which are deployed densely, homogeneously and are time synchronized, where sensor data tends to be correlated in both time and space. An example of such sub-network can be illustrated by Fig. 3[7].

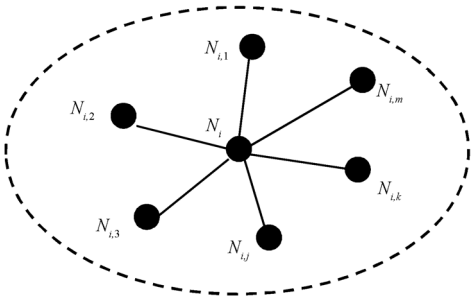


Fig. 3 Example of a sensor sub-network

This sub-network is reasonable because it can be easily extended to other type of networks topologies. For example, if a cluster-head or sink-head is inserted into this network, then a clustering-based network topology can be gained, or if a parent node can be added to take care of his children node, then a hierarchical-based network topology can be made. Consequently, our proposed anomaly detection method for this sub-network can be extended conveniently to multiple kinds of networks.

This sub-network, $N = \{N_i, , N_{i,1}, \cdots, N_{i,m}\}$, which represents a closed neighborhood of node $N_i \in N$, contains node N_i and its m spatially neighboring nodes. The m spatially neighboring nodes are represented by $N_{i,j} = \{N_{i,j}: j = 1, \cdots, m\}$. Each sensor node in the sub-network at every time interval Δt measures a data vector and each data vector is composed of multiple attributes. For the sensor node N_i, the observation is $X' = (x_1^i, x_2^i, \cdots, x_d^i)$, where d denotes the attribute dimension. And for the j-th neighbor node, $N_{i,j}$, the observation is $X_j^i = (x_{j,1}^i, x_{j,2}^i, \cdots, x_{j,d}^i)$. Each node in the sub-network can do the same work as done by N_i, such as collecting data, communication and anomaly detection. Our proposed method is to timely identify ev-

ery new observation of N_i as normal or anomalous based on the detection model which is trained using the collected dataset.

3.2 Isolation based anomaly detection method

The traditional anomaly detection method in the data mining and machine learning community cannot be employed in the context of sensor network well. The main challenge faced by anomaly detection method for WSNs is how to achieve high detection rate and low false alarm rate while consuming the minimum resource of WSNs. Consequently, it is necessary to develop a light-weighted anomaly detection method.

Isolation-based anomaly detection, proposed in [15] assumed that the anomaly data points are always rare and far from the center of the normal clusters. This is a new and effective anomaly detection technique by using the binary tree to build the single detector. Based on the ensemble learning theory, the detection model consists of multiple individual detectors, namely, isolation tree. Here, we give the following definition.

Definition 2 (isolation tree). Let T be a node of an isolation tree, and x be a data point consisted of d attributes in the current subset, T be either an external node with no child, or an internal node with one test and exactly two children nodes (T_l, T_r). A test consists of an attribute q and a split value p such that the test $x.q < p$ can split the current subset into T_l and T_r. Moreover, each node keeps the corresponding intermediate results of the deterministic space partition process. T_l and T_r. are all the isolation trees.

The most valuable advantages of this method are as follow:

1) Isolation based method does not require distance or density measures to detect anomalies. This eliminates major computational cost of distance calculation in all the distance-based methods and density-based methods. It implies that this method is a light-weight method and can save resources of sensor nodes.

2) Isolation based method has linear time complexity with low computational and memory requirement, which implies that this method can detect anomaly fast and save the memory resource effectively.

Here, we give the simple presentation of this method. Building the isolation tree process is intuitive and straightforward, the interested readers can refer to [15] for further detail.

Given data set $X = \{x_1, x_2, \cdots, x_n\}$, $x_i = (x_{i,1}, x_{i,2}, \cdots, x_{i,d})$, to build an isolation tree, we recursively divide X by randomly selecting an attribute q and a split value p. The creation of binary tree procedure is ended if one of the following three principles is satisfied:

1) The tree reaches a pre-defined height limit.

2) $|x'| = 1, x' \subseteq x'$.

3) All the data in X' have the same values, i.e., they are all identical.

The created isolation tree has characters as:

1) It is a proper binary tree, where each node in the tree has exactly zero or two children nodes.

2) If each observation in X is distinguishable, the maximum number of nodes in the isolation tree is $2n$-1, and the number of internal tree nodes is n-1. Consequently, the memory requirement is bounded and only grows linearly with n, namely $O(n)$.

Of course, only a single isolation tree has probably a poor detection performance, isolation based anomaly detection method is an ensemble learning technique in essential, which needs to build multiple isolation trees to gain an aggregate decision. Consequently, ensemble detector is constructed for the given dataset at the training stage, which consists of N individual detectors.

In the anomaly detection phase, the task of anomaly detection is to find a ranking that reflects the degree of anomaly for the new observations. Thus, one way to detect anomalies is to sort data points according to their average path length or anomaly scores. If the threshold u is predefined, then those observations that are ranked at the top of list are defined as anomalies. The anomaly scores can be calculated by

$$\begin{aligned} s(x,n) &= 2^{-\frac{E(h(x))}{c(n)}} \\ E(h(x)) &= \frac{1}{m}\sum_{i=1}^{m} h(x) \\ c(n) &= 2H(n-1) - \frac{2(n-1)}{n} \\ H(n) &= \ln(n) + 0.5772156649 \end{aligned} \tag{1}$$

where $s(x,n)$ denotes the anomaly score of observation x, n is the number of the whole dataset, $E(h(x))$ denotes the expected isolation path length which can be estimated by computing the average of the isolation path length acquired from each isolation tree among the isolation forest, $H(n)$ is the harmonic number which can be estimated by $\ln(n) + 0.5772156649$, the number 0.5772156649 is the Euler′s constant, and $c(n)$ is the average of $h(x)$ given n, which is used to normalized $h(x)$.

3.3 Distributed anomaly detection method based on isolation principle in WSNs considering spatial relationship

In order to avoid massive communication in WSNs, each node builds a local anomaly detector using its m observations which is sub-sampled from the training datasets. The neighbor nodes in the sub network do the similar work. After the local detector is built, each node broadcasts its information of multiple isolated trees, such as tree structures, split attributes and split value in each created tree node, to its spatially neighboring nodes. Consequently, every node in the WSNs receives the local detector information from all of its neighbors and then combines with itself to build the

final global decision isolation forest for the anomaly detection. Another alternative strategy is that if a cluster head node is deployed in the sub network, each network node can transmit its local detector information to the cluster head node. A cluster head combines each node′s local detector information into the final global detector and then broadcast this ensemble detector to its children node. Considering the spatial relationship among the observations in WSNs and the idea of ensemble learning, this model is relatively robust. The model can be seen in Fig. 4, and this method can be described by training phase and testing phase.

3.3.1 Training phase

In the training phase, the normal data samples collected by each sensor node are used to build multiple isolation trees for local anomaly detection, the method of building ensemble anomaly detection model can be seen in Fig. 5.

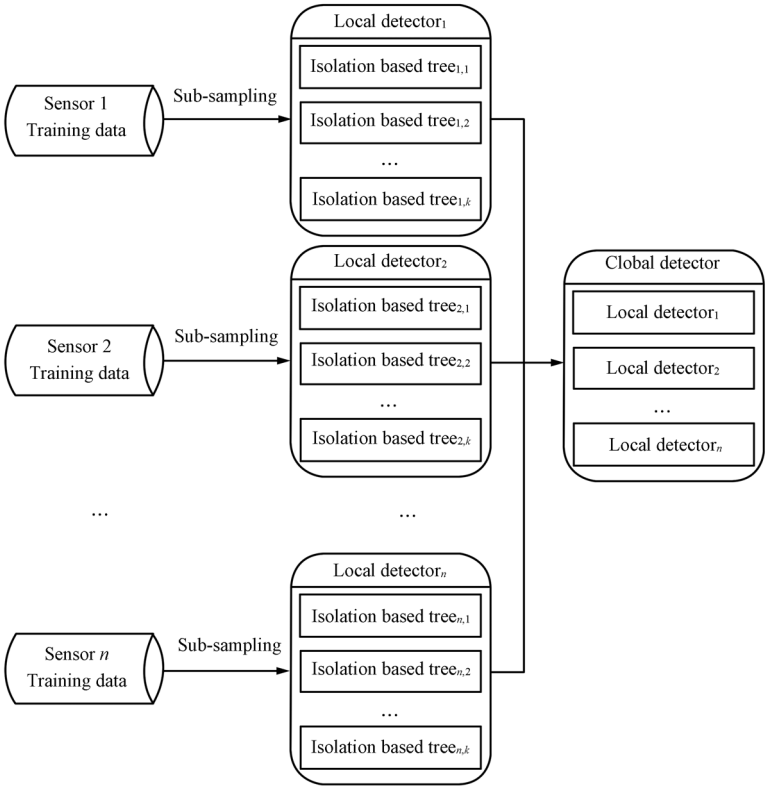


Fig. 4 Diagram of building distributed anomaly detector based on ensemble isolation principle

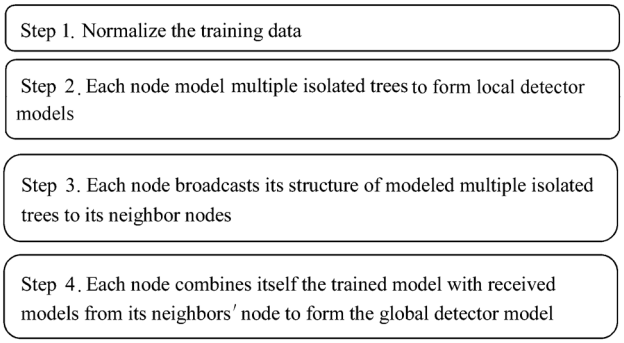


Fig. 5 Training phase of modeling distributed anomaly detector based on ensemble isolation principle

3.3.2 Testing phase

When a new observation x arrives at node N_i, its isolation depth in each tree is calculated and finally average isolation depth is got. Then an anomalous score is calculated by (1). Three principles are used to detect whether an observation is anomalous or not.

1) If anomaly score of an observation is very close to 1, then it is definitely an anomaly.

2) If anomaly score of an observation is much smaller than 0.5, then it is quite safe to be regarded as normal one.

3) If anomaly score of an observation is approximately equal to 0.5, then whether it is an anomaly or not, it is hard to determine.

However, the anomalous score of many observations on the fringe of data distribution is near 0.5 and hard to identify. Therefore, in the real applications, some prior knowledge may be used to assist the anomaly detection, e.g., anomaly rate, which is predefined by the domain experts. In most cases, obtaining the real anomaly rate is difficult. However, giving a vague scope of anomaly rate may be a relatively easy job. Consequently, the anomaly detection is performed based on the concept of anomaly score and anomaly rate. For a given observation, the procedure of identifying that whether it is an anomaly or not is described by Fig. 6.

After an observation was identified as an anomaly locally, it was represented as an error or event based on the rule: If the anomaly is spatially unrelated to its neighbor nodes, it is an error, otherwise it is an event.

4 Theoretical and complexity analysis

4.1 Theoretical analysis of isolation principle

In this section, we give the theoretical analysis for isolation principle based on the probabilistic explanation. The more detail can be found in [15, 19].

We firstly assume that the data is one-dimensional. It is noted that this assumption is mainly for clear presentation of the isolation principle rather than a limitation of the theoretical analysis, which can be generalized to multi-dimensional dataset.

Let $X=\{x|x \in \mathbf{R}\}$, for each data point x, $h(x)$ is the path length traveled from the tree root to x, and the expected path length $E(h(x))$ is a summation of a series of the possible path lengths with probabilistic components. It can be calculated by

$$E(h(x)) = \sum_l P(h(x) = l) \times l$$
$$\sum_l P(h(x) = l) = l \tag{2}$$

where $P(\cdot)$ denotes the probability function of path length and $E(\cdot)$ is the expectation function. For any fringe point which may have high probability as an anomalous point, the generalized possible path length is $h(x) \in [1, |X| - 1]$. For any non-fringe point, the generalized possible path length is $h(x) \in [2, |X| - 1]$.

Assuming that each possible tree structure (generated

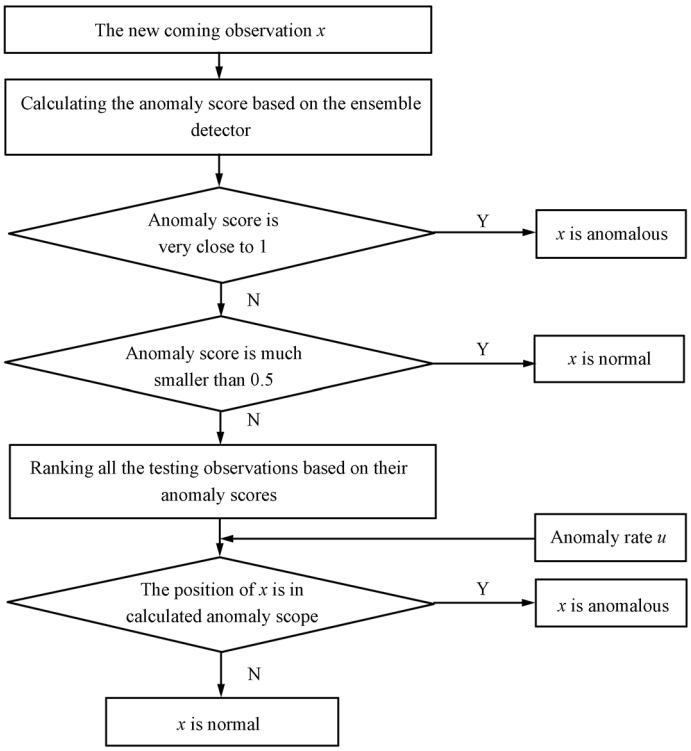


Fig. 6 The testing procedure for an observation x

randomly) is equally probable, i.e., uniform distribution, the term $P(h(x)=l)$ can be estimated by

$$P(h(x)=l)=\frac{t_{l,m,j}}{c_j}$$
$$t_{l,m,j}=\sum_{u=0}^{n}\left(\begin{array}{c} l \\ u \end{array}\right)C_{(m-u)(m-l)}C_{(j-m-l+u)(n-m-l)}$$
$$C_p=\left(\begin{array}{c} 2p \\ p \end{array}\right)-\left(\begin{array}{c} 2p \\ p-1 \end{array}\right)$$
$$C_{pr}=\left(\begin{array}{c} p+r \\ p \end{array}\right)-\left(\begin{array}{c} p+r \\ p-1 \end{array}\right) \tag{3}$$

where $t_{l,m,j}$ is the total number of possible trees that have $h(x_m)=l(m\in\{0,\cdots,j\})$ with j internal nodes. The average path length of a data point x_m can be calculated by

$$E(h(x))=\frac{h_{m,j}}{C_j}$$
$$h_{m,j}=2\left(\begin{array}{c} 2m \\ m \end{array}\right)\left(\begin{array}{c} 2j-2m \\ j-m \end{array}\right)\frac{(2m+1)(2j-2m+1)}{(j+1)(j+2)}-C_j \tag{4}$$

where $h_{m,j}$ denotes the sum of path lengths for x_m in all possible trees with j internal nodes. For uniformly distributed data points, the average path length $E(h(x))$ of randomly generated binary trees has a dome shape. Consequently, the dome shape reveals that the fringe points have much lower expected path lengths than those core points. And these fringe points may be easily isolated and have the high probability as anomalous data points.

4.2 Algorithm complexity analysis

Complexity analysis is another factor to be taken into account for evaluating the performance of algorithm, especially in the application of WSNs. We evaluate our proposed algorithm in terms of communication overhead and computation and memory complexity.

The communication complexity of our distributed techniques depends on the local transmission of the isolated tree structure parameter information in each sensor node. Because the isolation tree is a binary tree, supposing that this binary tree has m leaf nodes in the worst case, then each isolation tree has approximately $2m$-1 nodes totally. For each sensor node which owns k isolation trees, the maximum communication complexity is O (km).

The computational complexity in our proposed algorithm is mainly composed of two parts: one is computation cost of creating the k isolation trees for each sensor node, the other is that each observation calculates the isolation depth and has the average depth. For the first part, the maximum of computation complexity is O(km) for each sensor node. For the second part, acquiring the observation depth is equal to travel each isolation tree, so the maximum computation complexity is $O(\log_2^m)$. Therefore, the maximum computational complexity at each node of creating and traveling K isolation trees is $\mathrm{O}(km+K\log_2^m)$ $(K=\sum k)$, the final computational complexity is $\mathrm{O}(km)$.

The memory complexity in our proposed algorithm is mainly dominated by K isolation trees for each sensor node, then the maximum memory requirement is $\mathrm{O}(km)$.

In the real application, when the values of k and m are fixed, then the communication complexity is constant. Consequently, our proposed method may have the promising advantage to fit for the resource constrained WSNs applications.

5 Experiment and result analysis

The proposed anomaly detection method is evaluated by using the labeled dataset[12]. Because isolation based method is an un-supervised method and does not need the category attributes during the procedure of anomaly detection, the anomalies′ labels are only used in evaluating the final anomaly detection performance. Our experiment goal is to test the effectiveness of our proposed distributed anomaly detector based on isolation principle in WSNs. The experiments are conducted on a personal PC with Intel® Core™ 2 Duo CPU, P7450@2.13GHZ and 4 GB memory. The operating system is Windows 7 Professional. The algorithms described in Section 3 are programmed by the C++ language with the Visual C++ software platform and the dataset preprocessing as well as result analysis are done using Matlab 2010 platform.

5.1 Labeled dataset

This dataset is a real humidity-temperature sensor data which is collected using TelosB motes in a single-hop WSNs[12]. This dataset is designed and has a controlled anomalous data, means labeled anomalies. There are two indoor sensor nodes and two outdoor sensor nodes, respectively. The data consists of temperature and humidity measurements collected over a period of 6 hours in intervals of 5 seconds. The sensors 1 and 2 are deployed in indoor as well as 3 and 4 outdoor, respectively. In order to simulate the anomaly occurrence, it is designed by using a hot water kettle which increases the temperature and the humidity simultaneously. The detailed information of this dataset is described in Table 1.

Table 1 Detailed information of labeled dataset

Setting	Node	Total sample	Normal	Anomaly	Training	Testing
Indoor	1	4417	4300	117	2917	1500
	2	4417	4417	0	2917	1500
Outdoor	3	5039	5039	0	3339	1700
	4	5041	5009	32	3341	1700

In this dataset, the anomalous observations are introduced at nodes 1 and 4, while nodes 2 and 3 have no anomalous data and represent normal observations. The data distributions for each of these sensor nodes are depicted in Fig. 7.

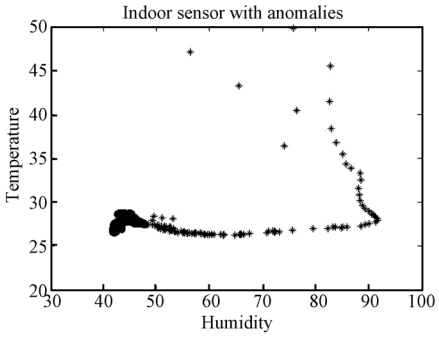
5.2 Data pre-processing

Consider the fact that sensor observations collected by different types of sensors may have different scales in the real application. These observations are pre-processed firstly[4]. There are some commonly used data preprocessing methods in the data mining and machine learning community. For a data vector $x_i \in X$, its auto-scaled value is formulated as

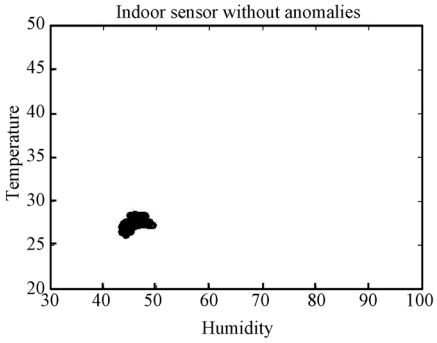
$$x_i' = \frac{x_i - u}{\sigma}$$

where u is the mean of the attribute values and σ is the corresponding standard deviation. However, the auto-scaled values may be sensitive to anomalies in the field of anomaly detection, and the common method is replacing the arithmetic mean by the median and replacing the standard deviation by the median absolute deviation (MAD). This is because the median and MAD are more robust than mean and standard deviation against extreme high or low values. The MAD is calculated by $\text{MAD} = \text{median}\{|x_i - \text{median}(X)|\}$, where $|.|$ denotes the operator of absolute values. Consequently, for a data vector $x_i \in X$, its modified auto-scaled value is formulated as

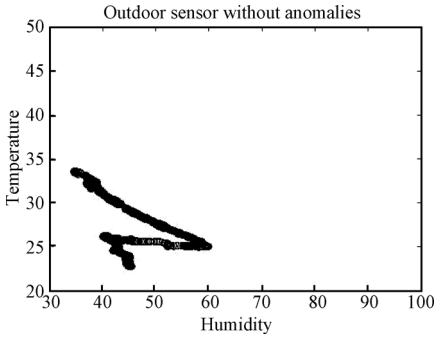
$$x_i' = \frac{x_i - \text{median}(X)}{\text{MAD}}.$$



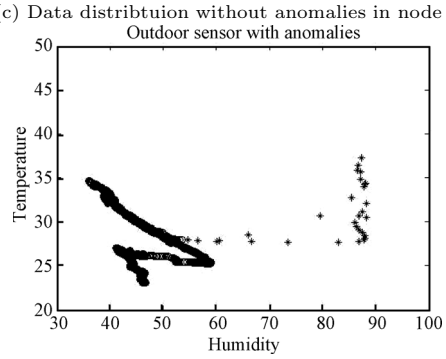
(a) Data distribtuion with anomalies in node 1



(b) Data distribtuion without anomalies in node 1



(c) Data distribtuion without anomalies in node 3



(d) Data distribtuion with anomalies in node 4

Fig. 7 Data distribution of the labeled dataset (black star means normal observations, black circle means anomalous observations)

5.3 Result evaluation metrics

To evaluate our proposed algorithm, four performance metrics are selected. The first is the accuracy rate (ACC), which is a traditional metric used to evaluate the classifier performance in the community of data mining and machine learning. ACC represents the percentage of right prediction. The second is the true positive rate (TPR), which represents the percentage of anomalies that are correctly detected, i.e., the ratio between the number of correctly detected anomalies and the total number of anomalies. The third is the false alarm rate (FAR) or false positive rate (FPR), which represents the percentage of normal data that are incorrectly considered as anomalies, i.e., the ratio between the number of normal data detected as anomalies and the total number of normal data. ACC, TPR and FAR can be calculated by

$$\begin{aligned} \text{ACC} &= \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}} \\ \text{TPR} &= \frac{\text{TP}}{\text{TP} + \text{FN}} \\ \text{FAR} &= \frac{\text{FP}}{\text{FP} + \text{TN}} \end{aligned} \tag{5}$$

where TP, FP, TN and FN denote the number of true positive sample points, the number of false positive sample points, the number of true negative sample points and the number of false negative sample points, respectively.

The 4th metric is selected to evaluate our method, which is the area under cover (AUC)[15, 20] . Compared to the TPR and FAR, AUC is widely used to measure the overall

performance of anomaly detector regardless of the threshold between the true positive and true negative. The calculation formula is described as [21].

$$S = \sum_{i=1}^{n_a} r_i$$

$$AUC = \frac{S - \frac{n_a^2 + n_a}{2}}{n_a \times n_n} \quad (6)$$

where r_i is the rank of the i-th anomalous point list which is sorted by the anomaly score in ascending order, S is the sum of the ranks of the actual anomalous points, n_a and n_n denote the number of true anomalies and the number of true normal points, respectively.

5.4 Experiment and result analysis

Isolation based anomaly detection method has three critical parameters which affects the result performance, i.e., the sub-sampling size, the ensemble size and the estimated anomaly rate. According to the result in [15], when the sub-sampling size is set as 256 and the ensemble size is set as 100, most datasets have an acceptable detecting performance. In order to select the acceptable parameter combination, extensive experiments are performed based on different parameter combination. For sub-sampling size, its value is set as 64, 128, 256, 512 and 1024, respectively. For ensemble size, its value is set as 20, 30, 40, 50 and 60 for each local detector model, respectively. For indoor observations, the anomaly rate is 2.6% (117/4417≈0.026). However, in the real application, deciding the true anomaly rate is always unknown and is not a trivial problem, and the expert usually gives a rough estimation range based on their prior knowledge. In our experiment, the anomaly rate is set as 1%, 2% and 3%, respectively. For outdoor sensor, the anomaly rate is 0.6% (32/5041≈0.006), so the anomaly rate is set as 0.3%, 0.6% and 1%, respectively.

The dataset was divided into two parts. For each sensor dataset, about 66% of the each sensor dataset was used for training the local detection model, the remaining part was used for evaluating our proposed method. The experiment results for node 1 can be seen in Table 2 which is based on the ensemble model trained by indoor nodes 1 and 2.

From the results presented in Table 2, it can seen clearly that the estimation metric, ACC, may be not appropriate for the anomaly detection performance evaluation because it cannot reflect the effects induced by different parameter setting, such as ensemble size, sampling size and anomaly rate for the extreme distribution unbalance in the dataset. However, for other evaluation metrics, TPR, FAR and ACC can reflect the different parameter setting. It is clear that TPR increases and FAR decreases with the ensemble size and sample size increasing. Similarly, the value of AUC increases gradually. It can be also seen that when the ensemble size reaches to a value, such as 40, or sampling size reaches to 256, the values of different evaluation metrics vary unnoticeably and become gradually stable. Then the detecting performance is acceptable too. Therefore, there is no need to increase the sampling size and ensemble size further because it only increases the processing time and memory requirement as well as communication cost without any improvement of the detection performance.

However, for the estimated anomaly rate, different values may have significant effect on the results. For our proposed method, the probability of an observation regarded as an anomaly is based on its anomaly score and its ranking in the whole testing dataset. If the value that is set lower than its true anomaly rate, only part of true anomalous observations is identified as anomaly. Meanwhile, these true anomalies are regarded as normal observations. As a result, the FAR is low. On the contrary, when the value of anomaly rate is set higher than its true anomaly rate, then

Table 2 Experiment result on labeled dataset (indoor) for different parameter combinations

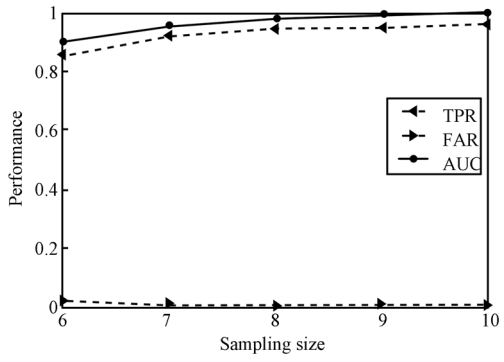
		Ensemble size																			
		20				30				40				50				60			
Anomaly rate	Sampling size	ACC	TPR	FAR	AUC	ACC	TPR	FAR	AUC	ACC	TPR	FAR	AUC	ACC	TPR	FAR	AUC	ACC	TPR	FAR	AUC
1%	64	0.9740	0.6066	0.0149	0.9579	0.9667	0.6598	0.0086	0.9646	0.9635	0.6000	0.0102	0.9703	0.9635	0.6000	0.0102	0.9561	0.9635	0.6000	0.0102	0.9573
1%	128	0.9635	0.7900	0.0102	0.9700	0.9703	0.7282	0.0067	0.9684	0.9785	0.8821	0.0026	0.9859	0.9808	0.9248	0.0014	0.9898	0.9785	0.8821	0.0026	0.9849
1%	256	0.9758	0.8308	0.0040	0.9733	0.9753	0.8222	0.0042	0.9847	0.9785	0.8821	0.0026	0.9898	0.9817	0.9419	0.0009	0.9926	0.9776	0.8650	0.0030	0.9862
1%	512	0.9731	0.7795	0.0053	0.9849	0.9780	0.8735	0.0028	0.9901	0.9785	0.8821	0.0026	0.9889	0.9771	0.8564	0.0033	0.9897	0.9735	0.8880	0.0051	0.9863
1%	1024	0.9767	0.8479	0.0035	0.9910	0.9731	0.7795	0.0053	0.9850	0.9749	0.8137	0.0044	0.9856	0.9749	0.8137	0.0044	0.9886	0.9744	0.8751	0.0247	0.9674
2%	64	0.9649	0.7137	0.0167	0.9237	0.9590	0.7026	0.0177	0.8612	0.9826	0.8470	0.0156	0.9272	0.9712	0.8333	0.0114	0.9786	0.9699	0.8077	0.0121	0.9705
2%	128	0.9726	0.7890	0.0124	0.9246	0.9826	0.9470	0.0156	0.9808	0.9663	0.8993	0.0140	0.9476	0.9762	0.9274	0.0088	0.9831	0.9703	0.8162	0.0119	0.9749
2%	256	0.9726	0.8290	0.0107	0.9448	0.9771	0.9444	0.0094	0.9827	0.9694	0.9515	0.0123	0.9762	0.9821	0.9385	0.0058	0.9897	0.9708	0.8848	0.0116	0.9822
2%	512	0.9726	0.8590	0.0107	0.9690	0.9722	0.9504	0.0109	0.9799	0.9776	0.9530	0.0080	0.9886	0.9803	0.9443	0.0067	0.9903	0.9789	0.9786	0.0074	0.9890
2%	1024	0.9776	0.9530	0.0081	0.9911	0.9744	0.9832	0.0098	0.9844	0.9776	0.9530	0.0081	0.9884	0.9753	0.9603	0.0093	0.9863	0.9803	0.9743	0.0067	0.9905
3%	64	0.9513	0.7553	0.0267	0.9490	0.9645	0.7032	0.0200	0.9683	0.9744	0.8912	0.0149	0.9832	0.9703	0.8143	0.0170	0.9760	0.9631	0.8375	0.0207	0.9646
3%	128	0.9640	0.7946	0.0202	0.9721	0.9740	0.9227	0.0151	0.9752	0.9740	0.8827	0.0151	0.9778	0.9789	0.9467	0.0126	0.9847	0.9740	0.8427	0.0151	0.9773
3%	256	0.9749	0.8697	0.0147	0.9843	0.9862	0.9434	0.0118	0.9939	0.9726	0.9370	0.0158	0.9794	0.9735	0.9441	0.0153	0.9851	0.9799	0.9538	0.0121	0.9876
3%	512	0.9812	0.9094	0.0114	0.9874	0.9799	0.9938	0.0121	0.9880	0.9785	0.9681	0.0128	0.9859	0.9794	0.9852	0.0123	0.9897	0.9785	0.9281	0.0128	0.9905
3%	1024	0.9744	0.8912	0.0149	0.9852	0.9776	0.9510	0.0133	0.9880	0.9835	0.9721	0.0102	0.9913	0.9780	0.9596	0.0130	0.9882	0.9821	0.9965	0.0109	0.9898

more observations are identified as anomalies. As a result, the FAR is high, which can be seen from our experiment results in Table 2.

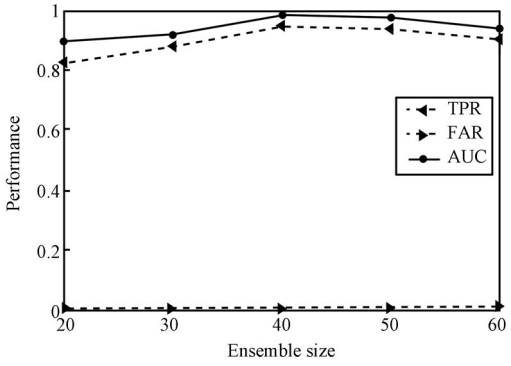
The same experiment is repeated for the outdoor labeled dataset. The global ensemble detector is trained by the nodes 3 and 4. Similar results can be obtained for node 4. Here, for the space limitation, we only gave partial results induced by the different local ensemble size, sampling size and anomaly rate, respectively, which can be seen in Fig. 8.

In order to evaluate the performance affected by the different sampling sizes, Fig. 8 (a) presents the performance of TPR, FAR, AUC for testing dataset collected by node 4. The sampling sizes are set as 64, 128, 256, 512 and 1024, respectively, and other two parameters are arranged as fixed values. For example, the local ensemble size is set as 50 and the anomaly rate is set as 0.006. Similarly, in order to evaluate the performance affected by different ensemble sizes, Fig. 8 (b) presents the performance of TPR, FAR and AUC for testing dataset collected by node 4. The local ensemble sizes are set as 20, 30, 40, 50 and 60, respectively, and other two parameters are arranged as fixed values. For example, the sampling size is set as 256 and the anomaly rate is set as 0.006. In order to evaluate the performance affected by different anomaly rates. Fig. 8 (c) presents the performance of TPR, FAR, AUC for testing dataset collected by node 4. The sampling sizes are set as 64, 128, 256, 512 and 1024, respectively, and the other two parameters are arranged as fixed values. For example, the local ensemble size is set as 50 and the anomaly rate is set as 0.006. From the above presented results, it can be clearly seen that anomaly detection ratio increases and the false alarm rate decreases gradually with the increasing of sampling size and ensemble size. Similarly, the value of AUC increases gradually. Considering the resources constraints and communication burden, the local ensemble size is set as 40 or 50, and 256 may be the appropriate value of the sampling size.

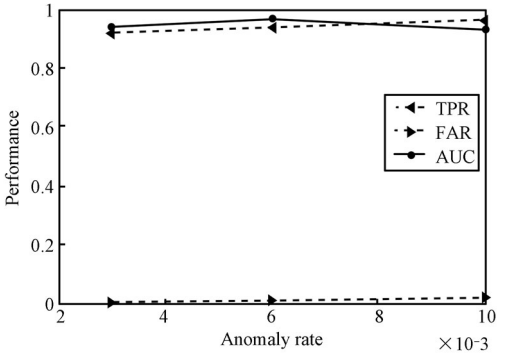
For the testing dataset collected from nodes 2 and 3, no one as anomalous exists, the anomaly score is used to decide that whether it is anomaly or not. For the space limitation, we omit the detailed procedure, and only the final results are presented in Table 3, where sampling size and local ensemble size are set as 256 and 50, respectively.



(a) The performance (y-axis) versus different sampling sizes in $\log_2$ scale (x-axis)



(b) The performance (y-axis) versus different local ensemble sizes (x-axis)



(c) The performance (y-axis) versus different anomaly rates (x-axis)

Fig. 8 Experiment results for different local ensemble sizes, sampling sizes and anomaly rates

Table 3 Experiment results about nodes 2 and 3 for labeled dataset

Setting	Node	TPR	FAR	AUC
Indoor	2	0.9437	0.0054	0.9763
Outdoor	3	0.9386.	0.0057	0.9548

6 Conclusions

The anomaly detection in WSN can improve the reliability and availability of the collected data. Considering the spatial correlation among the neighbor sensor data, a distributed anomaly detection method has been proposed based on the isolation principle. A local detector in each node is built, and an ensemble strategy among different nodes is employed. It is worthy to note that instead of communication raw data between sensors, only the structure and parameters of the trained detector on each sensor are broadcasted in its communication range. The experiment results demonstrate that the proposed method has high detection accuracy and low false prediction.

For the dataset collected from WSNs, it has the characteristic of spatio-temporal correlation, and future research can integrate the temporal correlation into the proposed method.

References

[1] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, E. Cayirci. Wireless sensor networks: A survey. *Computer Networks*, vol. 38, no. 4, pp. 393–422, 2002.

[2] J. Y. Huang, I. E. Liao, Y. F. Chung, K. T. Chen. Shielding wireless sensor network using Markovian intrusion detection system with attack pattern mining. *Information Sciences*, vol. 231, pp. 32–44, 2013.

[3] S. Rajasegarar, C. Leckie, M. Palaniswami, Anomaly detection in wireless sensor networks. *IEEE Wireless Communications*, vol. 15, no. 4, pp. 34–40, 2008.

[4] Y. Zhang. Observing the Unobservable: Distributed Online Outlier Detection in Wireless Sensor Networks, Ph. D. dissertation, University of Twente, The Netherlands, 2010.

[5] Y. Zhang, N. Meratnia, P. Havinga. Outlier detection techniques for wireless sensor networks: A survey. *IEEE Communications Surveys & Tutorials*, vol. 12, no. 2, pp. 159–170, 2010.

[6] W. J. Zhou, M. R. Fei, H. Y. Zhou, Z. Li. A fast detection method for bottle caps surface defect based on sparse representation. *Intelligent Computing for Sustainable Energy and Environment, Communications in Computer and Information Science*, Berlin, Germany: Springer, pp. 76–84, 2013.

[7] Y. Zhang, N. Meratnia, P. Havinga. An online outlier detection technique for wireless sensor networks using unsupervised quarter-sphere support vector machine. In *Proceedings of IEEE Intelligent Sensors, Sensor Networks and Information Processing*, IEEE, Sydney, NSW, Australia, pp. 151–156, 2008.

[8] J. Hill, R. Szewczyk, A. Woo, S. Hollar, D. Culler, K. Pister. System architecture directions for networked sensors. In *Proceedings of International Conference on Architectural Support for Programming Language and Operating Systems*, ACM, Cambrige, Massachusetts, USA, vol. 34, pp. 93–104, 2000.

[9] V. Chandola, A. Banerjee, V. Kumar. Anomaly detection: A survey. *ACM Computing Surveys*, vol. 41, no. 3, pp. 1–58, 2009.

[10] V. Hodge, J. Austin. A survey of outlier detection methodologies. *Artificial Intelligence Review*, vol. 22, no. 2, pp. 85–126, 2004.

[11] Y. Zhang, N. A. S. Hamm, N. Meratnia, A. Stein, M. V. D. Voort, P. J. M. Havinga. Statistics-based outlier detection for wireless sensor networks. *International Journal of Geographical Information Science*, vol. 26, no. 8, pp. 1373–1392, 2012.

[12] S. Suthaharan, M. Alzahrani, S. Rajasegarar, C. Leckie, M. Palaniswami. Labelled data collection for anomaly detection in wireless sensor networks. In *Proceedings of the 6th International Conference on Intelligent Sensors, Sensor Networks and Information Processing*, IEEE, Brisbane, QLD, Australia, pp. 269–274, 2010.

[13] M. Xie, S. Han, B. M. Tian, S. Parvin. Anomaly detection in wireless sensor networks: A survey. *Journal of Network and Computer Applications*, vol. 34, no. 4, pp. 1302–1325, 2011.

[14] L. J. Zhao, T. Y. Chai, D. C. Yuan. Selective ensemble extreme learning machine modeling of effluent quality in wastewater treatment plants. *International Journal of Automation and Computing*, vol. 9, no. 6, pp. 627–633, 2012.

[15] F. T. Liu, K. M. Ting, Z. H. Zhou. Isolation-based anomaly detection. *ACM Transactions on Knowledge Discovery from Data*, vol. 6, no. 1, Article 3, 2012.

[16] Z. G. Ding, M. R. Fei. An anomaly detection approach based on isolation forest algorithm for streaming data using sliding window. In *Proceedings of the 3rd IFAC International Conference on Intelligent Control and Automation Science*, IFAC, Chengdu, China, vol. 3, pp. 12–17, 2013.

[17] M. M. Breunig, H. P. Kriegel, R. T. Ng, J. Sander. LOF: Identifying density-based local outliers. In *Proceedings of ACM SIGMOD Record on Management of Data*, ACM, Dallas, USA, vol. 29, pp. 93–104, 2000.

[18] W. G. Yi, J. Duan, M. Y. Lu. Double-layer bayesian classifier ensembles based on frequent itemsets. *International Journal of Automation and Computing*, vol. 9, no. 2, pp. 215–220, 2012.

[19] D. E. Knuth. *Art of Computer Programming*, vol. 4, fascicle 4: *The Generating All Trees–History of Combinatorial Generation*, AddisonWesley Professional, pp. 83–98, 2006.

[20] S. C. Tan, K. M. Ting, T. F. Liu. Fast anomaly detection for streaming data. In *Proceedings of the 22nd International Joint Conference on Artificial Intelligence*, vol. 2, pp. 1511–1516, 2011.

[21] D. J. Hand, R. J. Till. A simple generalisation of the area under the ROC curve for multiple class classification problems. *Machine Learning*, vol. 45, no. 2, pp. 171–186, 2001.

Zhi-Guo Ding received the B. Sc. and M. Sc. degrees in computer science and technology from the Shanxi Normal University and Shanghai University, China in 2001 and 2007, respectively. He is currently a Ph. D. candidate in Shanghai University, China.

His research interests include intelligent control, big data processing and application.

E-mail: dingzhiguo@shu.edu.cn (Corresponding author)
ORCID iD: 0000-0001-8103-5130

Da-Jun Du received the B. Sc. and M. Sc. degrees both from the Zhengzhou University, China in 2002 and 2005, respectively, and the Ph. D. degree in control theory and control engineering from Shanghai University, China in 2010. He has been an associate professor at Shanghai University, China

His research interests include networked control systems, system modeling and identification.

E-mail: ddj@shu.edu.cn
ORCID iD: 0000-0003-2979-1507

Min-Rui Fei received the B. Sc. and M. Sc. degrees in industrial automation from Shanghai University of Technology, China in 1984 and 1992, respectively, and the Ph. D. degree in control theory and control engineering from Shanghai University, China in 1997. Since 1998, he has been a professor and doctoral supervisor at Shanghai University, China. He is the vice-chairman of Chinese Association for System Simulation, and Standing Director of China Instrument & Control Society.

His research interests include intelligent control, complex system modeling, networked control systems, and field control systems.

E-mail: mrfei@staff.shu.edu.cn