Using Sparse Representation to Detect Anomalies in Complex WSNs

XIAOMING LI, Tianjin University
GUANGQUAN XU*, Qingdao Huanghai University, Tianjin University
XI ZHENG, Macquarie University
KAITAI LIANG, EMMANOUIL PANAOUSIS, University of Surrey
TAO LI, Nankai University
WEI WANG, Beijing Jiaotong University
CHAO SHEN, Xi'an Jiaotong University

Abstract:

In recent years, wireless sensor networks (WSNs) have become an active area of research for monitoring physical and environmental conditions. Due to the interdependence of sensors, a functional anomaly in one sensor can cause a functional anomaly in another sensor, which can further lead to the malfunctioning of the entire sensor network. Existing research work has analysed faulty sensor anomalies, but fails to show the effectiveness throughout the entire interdependent network system. In this paper, a dictionary learning algorithm based on a non-negative constraint is developed, and a sparse representation anomaly node detection method for sensor networks is proposed based on the dictionary learning. Through experiment on a specific thermal power plant in China, we verify the robustness of our proposed method in detecting abnormal nodes against four state of the art approaches and proved our method is more robust. Furthermore, the experiments are conducted on the obtained abnormal nodes to prove the interdependence of multi-layer sensor networks and reveal the conditions and causes of a system crash.

CCS Concepts: • **Networks → Network performance evaluation →** Network experimentation

Additional Key Words and Phrases: dependency relationships networks, Sparse Representation, anomaly detection, WSNs

Authors' addresses: X. M. Li, College of Intelligence and Computing, Tianjin University, Tianjin 300350, China; G. Q. Xu (corresponding author, email: losin@tju.edu.cn),   Qingdao Huanghai University, Shandong 250000,China, College of Intelligence and Computing, Tianjin University, Tianjin 300350, China; X. Zheng, Department of Computing, Macquarie University, NSW 2109, Australia; K. T. Liang, E. Panaousis, Surrey Centre of Cyber Security, University of Surrey, Guildford, U.K; T. Li, College of Computer Science, Nankai University, Tianjin 300350, China; W. Wang, Beijing Key Laboratory of Security and Privacy in Intelligent Transportation, Beijing Jiaotong University, Beijing 100044, China; C. Shen, School of Electronic and Information Engineering, Xian Jiaotong University, Xi'an 710048, China.

1 INTRODUCTION

Anomaly detection (also known as outlier detection, novelty detection, or fault detection) refers to the problem of finding patterns of data that do not conform to expected behaviour [1] and has many applications, including intrusion detection [2], fraud detection [3], medical and public health anomaly detection [4], industrial safety detection [5], and image and text processing [6]. Thermal power plant mainly uses coal mill, coal feeder, fan, sensors and other equipments to generate electricity. Effective maintenance of these devices is the key to supply power stably. Traditional planned preventive maintenance (PPM) models are adopted in most thermal power plants. Thousands of wireless sensors, e.g., temperature, pressure, humidity, and speed sensors, are set up to monitor the state of the power plants equipment in real time. Data can be acquired continuously from these sensors, and sensor data monitoring allows to control the plants' status of the equipment. Advanced communication technology has allowed sensor nodes and controllers to be connected and form a network, resulting in greatly improved information data collection. A novel intelligent sensor network based on Fieldbus and Internet has been proposed to enable information exchange between network nodes and servers [7-9].

Analysing and mining sensor data can help detecting anomalies, which is important for the maintenance of complex WSNs with interdependent relationships based on abnormal nodes. In a WSN, detection of sensor anomalies and their dependencies is found to improve the reliability of the system and reduce unplanned downtime by minimizing catastrophic failure. This is achieved by detecting failures early and taking corrective measures before the failure worsens and causes further damage to the system. Mining anomaly information from sensor data is an important step for verifying sensor dependencies. Existing anomaly detection methods mostly rely on embedding abnormal nodes into the graph partition of a normal sensor network [10]. It is assumed that abnormal sensor nodes [11] will be found and embedded into the normal group for comparison. The above research studied a single abnormal node without considering the influence of the node over multiple systems (e.g., the loss of a pressure sensor and a wind sensor can be caused by a malfunctioning temperature sensor) or the interdependency between the nodes in a sensor network [12].

To better understand the robustness of this kind of network system, recent research has focused on multilayer sensor network dependencies. Schneider et al. [13] introduced autonomous nodes to the interdependent network to effectively improve network robustness. [14-15] found that

XX

such a system had the optimal equilibrium solution during the phase change process when the interdependent relationship of the hostile network was introduced; Lee et al. [16] analysed and discussed the vulnerability of an interdependent infrastructure system.

Sparse representation is built upon a rigorous statistical principle, which has been extensively used in applications including neural computing, pattern recognition,and computer vision. This paper is the first of its kind to apply sparse representation in analysing industrial system safety. This paper proposes an anomaly node detection method for complex sensor networks based on sparse representation, and the sensor network is analysed by using anomaly nodes. Through this method, we can detect abnormal sensors in a sensor network and further locate possible abnormal sensors in other networks using the anomalous sensors. Through the network analysis of the anomalous sensors, we found that the multilayer sensor networks are dependent on each other, which means that the robustness of the entire sensor system can be affected when an exception occurs in one sensor.

This paper has the following contributions: (1) sparse representation is used to detect sensor node anomalies in multilayer sensor networks; (2) by evaluating four types of sensor nodes in a thermal power plant, it is proved that our solution is superior to that of the state of the art methods in terms of true positive and false positive rates; and (3) the interdependence of multi-sensor networks is revealed by evaluating abnormal sensors, and the influence of important sensors on multi-sensor networks is studied to improve the robustness of the networks.

The paper is organized as follows: related work is discussed in Section 2; the proposed method is introduced in Section 3; experimental results are presented in Section 4; and Section 5 discusses the findings of the research.

2 RELATED WORK

Characterized by low cost, low power consumption, self-organization, and the ability to be distributed, WSNs can be widely deployed to gain access to real-time information [17-18]. WSNs are commonly used in fields such as smart industry, healthcare, industrial monitoring, and early warning and disaster prevention, which require a high level of data transmission reliability. Data interruption or loss in the transmission process can have catastrophic consequences, including human casualties [19-20]. Many factors affect the reliability of WSNs [21-22]. For example, transmission anomalies can be caused by low processing speed and minimal storage capacity [23-26]. Detecting anomalies in the network is a key to improve WSN performance and promote WSN applications.

2.1 Anomaly Detection

Recently, significant research has been conducted in the area of anomaly detection in various aspects. For instance, if enough labelled data (normal and abnormal samples) can be acquired, supervised machine learning methods can be adopted. Steinway, Hush, and Scovel [27] proposed a learning framework for anomaly detection based on a SVM, advancing the existing state of the art in density-based methods. Menander and SchöLoops [28] proposed a one-class classification (OCC) method for anomaly detection in the image processing domain. Other learning-based methods, such as the isolation forest [29], regularization framework [30], different granularity [31], and artificial neural network (ANN) [32], have been proposed. In most cases, labelled data are unavailable, or only a few datasets are known. Knox and Ng [33] proposed a distance-based method that suggests that an object is considered as an outlier if at least a fraction, p, of all instances have a distance to the object, which is larger than a threshold, d. There are many similar

schemes, such as the angle-based method [34] and the density-based method [35]. These schemes are advantageous because it is not necessary to know the labels of the data instances, and only the internal structure itself is considered. However, these methods are appropriate for offline data only because they cannot be used to analyse computationally complex data with high dimension, especially in real time. In [36], a BP algorithm is proposed, which is a multilayer neural network that includes an input layer, output layer, and intermediate layer. The middle layer can be extended to many layers, and the algorithm has a high-resolution detection effect. In [37], the authors proposed a decision tree approach using feature selection and an SVM-based method for fault detection in a steam turbine generator. In the field of machine learning, an SVM is a supervised learning model, which is usually used for pattern recognition, classification and regression analysis. Nonlinear principal component analysis (PCA) has also been suggested for feature extraction for reciprocating compressors [38]. In [37], the authors proposed a decision tree approach using feature selection and a support vector machine (**SVM**)-based method for fault detection in a steam turbine generator. Nonlinear principal component analysis (**PCA**) has also been suggested for feature extraction for reciprocating compressors [38]. In [39], the core algorithm for similarity-based modelling (SBM) of a commercial intelligent prognostics and maintenance platform is provided, which can be widely used in various industrial applications, especially in thermal electricity plants. The SBM technology is based on the application of a similarity operation on pairs of observation vectors and the manipulation of a "state" vector matrix, Φ, containing the historical normal training vector, Φ. Mathematically, we define the set of measurements taken at a given time as an exemplar vector X, where X_i is the measurement value from the i-th sensor, L is the number of wireless sensors to be monitored, and M is the number of representative training vectors in Φ. The SBM model is presented below:

$$X = [x_1,\ x_2 \cdots x_L]'$$
$$\Phi = [\phi_1, \phi_2 \cdots \phi_M]$$

In summary, most anomaly detection techniques discussed above focus on certain application domains, and thus dependency-based anomaly detection are not available as they are not suitable for complex industrial wireless sensor network systems.

2.2 Dependency Relationships

An actual sensor network system has multiple dependent relationships. For example, in the power-communication interdependent network, one power station can supply power to multiple communication stations, while one communication station can also control multiple power stations. Shao et al. used numerical analysis to analyse the successive abnormal effects of an interdependent network with dependent relationships and found that dependent relationships can transform successive anomalies from a first-order phase change to a second-order phase change, which can significantly strengthen the robustness of the network [40]. In practice, both intentional attacks and random anomalies may occur. Huang et al. proposed a method for treating a premeditated attack as a random anomaly and provided an analytic proof. They found that, compared to a single scale-free network, even though the nodes could be protected with a high degree of accuracy, the interdependent network was extremely fragile. This conclusion indicates that, under an intentional attack, it is not enough to use a protection strategy that is only effective for a single network to defend an interdependent network [41]. In [42], the influence of multi-

layer network interdependence on community detection is studied. Gao et al. [43] studied the robustness of an interdependent network against various attack strategies, providing a useful reference for the design of a highly robust network. In [44], the repair mechanism for reconnecting an attacked node with a certain probability is studied. Wei et al. analysed the influence of a repair method that partially redistributes the attacked node load in terms of power network robustness [45]. Pahwa et al. found that the repair method that reduces the targeted load can reduce the vulnerability of the network [46]. In [47], the authors specified the importance and position of the interdependent networks in the context of smart sustainable cities and provided a comprehensive investigation of recently developed optimization methods for large-scale networks. Liu et al. [48] found that the robustness of interdependent heterogeneous networks increases, whereas that of interdependent homogeneous networks with strong coupling decreases with in-degree and out-degree correlations. Radicchi et al. [49] demonstrated that percolation transitions in interdependent networks can be understood by decomposing these systems into uncoupled graphs.

Our method shares an approach similar to that of SBM. However, in the field of complex WSNs, we use sparse representation for anomaly detection. Furthermore, there are limited studies on the combination of anomaly detection and dependent networks. Owing to the extensive use of wireless sensors, we can obtain enough labelled data (normal and abnormal samples) for anomaly detection and further study detection methods for dependent sensor networks.

3 PROPOSED METHOD

3.1 Node Anomaly Detection Based on Sparse Representation

Sparse representation is a type of unsupervised learning that is applied to a set of over-complete bases to represent data automatically and linearly. Sparsity means that the weighting vector has few non-zero components. The choice of sparsity, as a desired characteristic of our input representation, is based on the observation that most sensory data, such as natural images, may be described as the superposition of a small number of minute elements, such as surfaces or edges.

The proposed method of detecting anomalies is divided into the following four main stages: (1) data pre-processing, (2) dictionary learning, (3) coding, and (4) anomaly scoring. These stages are not consecutive: the first two processes are performed offline, and the other two processes are performed online. During data pre-processing, the errors and inconsistencies in the samples are removed to enhance the quality of the training data. Moreover, the original data will be normalized to the range [0, 1]in this step. Dictionary learning is an offline training stage that finds a set of bases that can be used to represent input data, x, via linear combinations. While techniques such as PCA allow us to learn a complete set of basis vectors efficiently, we wish to learn an over-complete set of vectors to represent the input vectors $X \in R^M$ (i.e., such that $K > M$, where K is the number of bases). With over-complete basis vectors, we are able to capture the structures and patterns that are inherent in the input data. Coding (or sparse representation) is used to compute the corresponding coefficients for the line representing the input data. In this work, the L1 regularized least squares algorithm is used to implement the sparse code [50]. The last stage of the framework is outlier scoring. In this stage, an anomaly score within the range [0, 1] is assigned to the input data sample based on the cost of the sparse representation of the target function used in the coding stage. A flowchart of the method is outlined in Fig. 1.
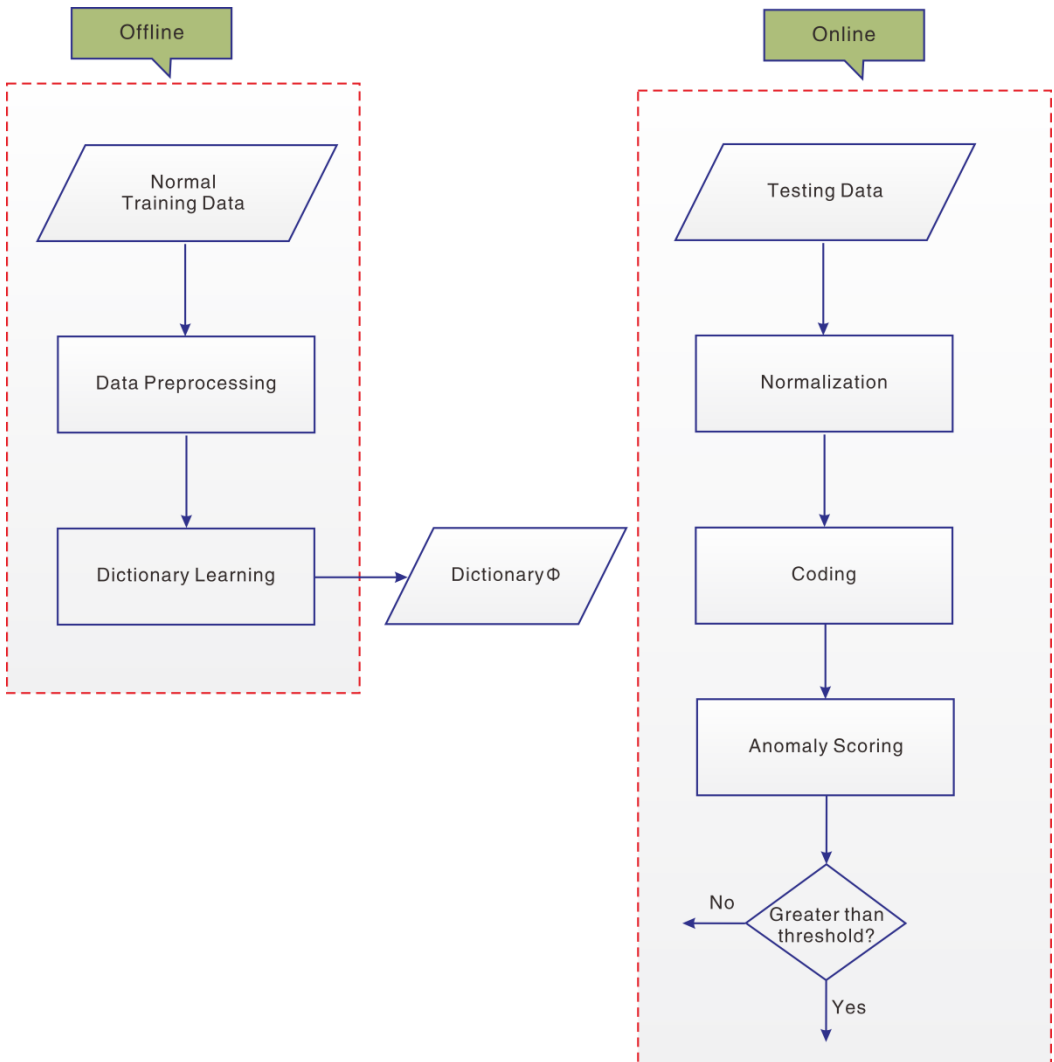
Fig. 1. Flowchart of the proposed framework

3.2 Data Pre-processing

Wireless sensors are installed on a number fixed equipment in the thermal power plant to capture the continuous (analogue or discrete) data of the equipment. Such data can be used to estimate the symptom, feature or state of the equipment. Using the notation that, at time T_j, the data from sensor X_i are denoted as Y_{ij} such that all the sensor data constitutes a feature vector $(Y_{1j}, Y_{2j}, \ldots Y_{nj})$, where n represents the number of wireless sensors installed on the equipment. In real-world cases, a very large number of wireless sensors are used to monitor equipment, and a large amount of data is captured from these sensors; handling such a large volume of data requires fast and efficient schemes, such as those proposed in [51-52].

The data pre-processing stage involves the following two tasks: data cleaning and data normalization. Often, real-world data contain information that is incomplete, noisy, and inconsistent. Data cleaning, which is the process through which errors and inconsistent data samples are removed from the dataset to enhance the quality of the training data, is performed as follows: (1) data that was acquired when the equipment was under load instability (overloading or underloading) are remove; (2) data spikes are removed; (3) data acquired when the equipment was faulty are removed; and (4) data are smoothed using classical Gaussian filter to eliminate the influence of noise. Moreover, to avoid the influence of scalability, the training data are normalized in the range [0, 1] using min-max normalization, as expressed by formula (1):

$$v'_{ij} = \frac{v_{ij} - min_i}{max_i - min_i} \tag{1}$$

where max_i and min_i are the maximum and minimum of the i-th attribute value, respectively, v_{ij} is the value of the i-th attribute of the j-th object, and v'_{ij} is the normalized value.

3.3 Sparse Representation

Given a normalized test sample, $x \in \square^m$, and a dictionary, $\Phi \in \mathbb{R}^{m \times k}$ (where m<k), $\Phi = (d_1, d_2 \cdots d_k)$ is an over-complete normal set (each column is a basic vector; learning these columns will be detailed later). Generally, x* is an approximation of x that could be reconstructed by a sparse linear combination of Φ:

$$x^* = \Phi \alpha^* \tag{2}$$

$$\alpha^* = \arg \min_{\alpha} J(x, \alpha, \Phi) \tag{3}$$

$$J(x, \alpha, \Phi) \overset{\Delta}{=} \frac{1}{2} \|x - \Phi\alpha\|_2^2 + \lambda\varphi(\alpha) \tag{4}$$

where α^* is the reconstruction coefficient, $\lambda \geq 0$ is a predefined coefficient that balances reconstruction accuracy and sparsity, and φ is the sparse regularization item.

Since x* is reconstructed from the normal dictionary, Φ, it could be seen as a theoretical normal value of the wireless sensors in the current environment. Usually, the predictive values can be calculated by using an auto-regressive model based on historical data from the sensor to estimate the output value of the sensor. The predicted value is different from the calculated value, and the theoretical value is used to estimate the output value. J(x, α, Φ) is called the sparse representation cost (SRC), which is comprised of the reconstruction error and the sparse regularization item. In this paper, the learned dictionary was expected to represent the normal output state of the wireless sensors. Therefore, for a routine data sample, the reconstruction error will be small. In contrast, a larger reconstruction error implies that sample x is not well represented by the customary dictionary, Φ, and can thus be considered an anomaly.

The second term in formula (4) is the sparse regularization term, and λ is a regularization parameter. Rather than applying the sparsity penalty, which is a pseudo-norm referring to the number of coefficients that are not equal to zero, the convex l1 norm is used. L1 norm regularization produces sparse coefficients and can be more robust for dictionary learning. The purpose of the sparse regularization term is to guarantee that a test sample, x, could be reconstructed in a concise way by the dictionary, Φ. In other words, although the abnormal sample achieves a small reconstruction error, it would require a large number of normal bases of

the dictionary, Φ. Thus, routine events are likely to get more sparse reconstruction coefficients, while abnormal events get more dense representations, which solves the sparse representation problem.

3.4 Dictionary Learning

The dictionary learning stage is an offline training stage that aims to find a set of bases such that the sample input data, X, can be represented as a linear combination of the basis vectors. A finite training set, $X = [x_1, x_2 \ldots x_n] \in \mathbb{R}^{m \times n}$, is required to build a dictionary, $\Phi (\in \mathbb{R}^{m \times k})$, where each column represents a basis vector. Considering pressure in thermal plant monitoring, which is monitored by 32 wireless sensors, it is possible to acquire 32 values from all the sensors every minute. By using data collected over one year, the number of training vectors is m = 32. The learning process is required to build over-complete dictionaries, which means that k>m (Due to $k \ll n$; here, we set k to 256). Dictionary learning can be represented by a formal expression as follows:

$$
\begin{aligned}
\Phi &= \underset{\Phi, \alpha_i}{\arg \min} \sum_{i=1}^{n} J\left(x_i, \alpha_i, \Phi\right) \\
&= \underset{\Phi, \alpha_i}{\arg \min} \sum_{i=1}^{n} \left(\frac{1}{2} \left\| x_i - \Phi \alpha_i \right\|_2^2 + \lambda \left\| \alpha_i \right\|_1 \right)
\end{aligned}
\tag{5}
$$

It is also assumed that $\lambda \geq 0$, and $\left\| * \right\|_F$ is the Fresenius norm. Formula (5) requires explanation.

Spatiotemporal smoothing: Dictionary learning attempts to find a set of over-complete basis vectors that can be used to best approximate the data to be analysed. Here, if X_i and X_j are similar in terms of some metrics, then a_i and a_j, i.e., the sparse representation of the two vectors, should be similar. This assumption, which is known as the local invariance assumption, is commonly used in pattern recognition and machine learning. The similarity between input vectors x_i and X_j can then be defined as

$$
W(x_i, x_j) = S(x_i, x_j) * T(x_i, x_j)
\tag{6}
$$

Non-negative constraint: The main purpose of this stage is to learn a dictionary, Φ, that characterizes the "normal" or "desirable" operating conditions of the monitored equipment. During pre-processing, all the training data are normalized to the range [0, 1]; thus, a non-negative constraint must be added to acquire the bases. We adopt the idea proposed by [36], which is to combine sparse representation and non-negative matrix factorization (NMF). NMF is a group of algorithms that factorize an input matrix, $X_{m \times n}$, into two matrices, $\Phi_{m \times k}$ and $A_{k \times n}$, where all three matrices have no negative elements; NMF is widely used in dimensionality reduction and clustering algorithms. Different from classic NMF, where k <<m, for over-complete bases, k > m.

$$W(x_i, x_j) = S(x_i, x_j) * T(x_i, x_j) \tag{7}$$

$$S(x_i, x_j) = \frac{x_i^T x_j}{\|x_i\|\|x_j\|}, T(x_i, x_j) = \exp(-\frac{\left\| t_{x_i} - t_{x_j} \right\|^2}{\sigma})$$

Thus, the following terms can be used to quantify the smoothness of the sparse representation:

$$
\begin{aligned}
R &= \frac{1}{2} \sum_{i,j=1}^{N} \left\| \alpha_i - \alpha_j \right\|^2 W_{ij} \\
&= \sum_{i=1}^{N} \alpha_i^T \alpha_i D_{ii} - \sum_{i=1}^{N} \alpha_i^T \alpha_j W_{ij} \\
&= \mathrm{Tr}(ADA^T) - \mathrm{Tr}(AWA^T) \\
&= \mathrm{Tr}(ALA^T)
\end{aligned}
\tag{8}
$$

where Tr (*) denotes the trace of a matrix, and D is a diagonal matrix whose entries are column sums (or row sums, given that W is symmetric) of W.

$$DD_{ii} = \sum_j W_{ij} \quad L = D - W$$

Combined with spatiotemporal smoothing and the non-negative constraint, we redefined formula (9) as follows:

$$\min_{\Phi, A} \frac{1}{2} \left\| X - \Phi A \right\|_F^2 + \lambda \sum_{ij} A_{ij} + \gamma \mathrm{Tr}(ALA^T) \tag{9}$$

subject to λ≥0, γ≥0, Aij≥0, and Φij≥0.

The object formula in (9) is not convex in both Φ and A combined. It is not easy to acquire the global minimum. Considering the Lagrange function L:

$$
\begin{aligned}
L &= \frac{1}{2} \left\| X - \Phi A \right\|_F^2 + \lambda \sum_{ij} A_{ij} + \gamma \mathrm{Tr}(ALA^T) + \mathrm{Tr}(\psi \Phi^T) + \mathrm{Tr}(\varphi A^T) \\
&= \frac{1}{2} \mathrm{Tr}(XX^T) - \mathrm{Tr}(XA^T \Phi^T) + \frac{1}{2} \mathrm{Tr}(\Phi AA^T \Phi^T) + \lambda \sum_{ij} A_{ij} + \gamma \mathrm{Tr}(ALA^T) \\
&\quad + \mathrm{Tr}(\psi \Phi^T) + \mathrm{Tr}(\varphi A^T)
\end{aligned}
\tag{10}
$$

Using the necessary condition for the existence of the extreme value (i.e., the KKT conditions $\psi\Phi^T = \mathbf{0}$; $\varphi A^T = \mathbf{0}$):

$$
\begin{cases}
\left(XA^T\right)_{ik} \mu_{ik} - \left(\Phi AA^T\right)_{ik} \mu_{ik} - \lambda \mu_{ik} = 0 \\
\left(X^T A^T\right)_{jk} \upsilon_{jk} - \left(A^T \Phi^T \Phi\right)_{jk} \upsilon_{jk} - \gamma \left(LA^T\right)_{jk} \upsilon_{jk} = 0
\end{cases}
\tag{11}
$$

Many iterative methods can be used to solve the problem of dictionary learning. Gradient descent, which is a commonly used method, adopts additive update rules and is easy to implement;

however, it is difficult to set the step size while maintaining the non-negativity of the data, and convergence may be very slow. Other methods, such as conjugate gradients, have faster convergence but are more complicated to apply. The multiplicative update rules proposed by [53-55] were adopted in this work. This iterative rule has been proven to be monatomic and convergent. Details of the derivations are available in [56], and only the concrete iterative formulas are shown in Table 1.

Table 1. Algorithm for Dictionary Learning.

Input: Normal sample X.

Output: Dictionary Φ.

Step 1: Compute similarity $W(X_i, X_j)$ for all pairs of samples in X using (7).

Step 2: Set constant γ and λ, initialize Φ and A as dense positive random matrices.

Step 3: Repeat:

  For t = 1... N do

     Set

$$\Phi_{ik}^t \leftarrow \Phi_{ik}^t \left(\frac{\left(XA^T\right)_{ik}}{\left(\Phi AA^T\right)_{ik} + \lambda} \right) \tag{12}$$

$$A_{ik}^t \leftarrow A_{ik}^t \left(\frac{\left(X^T A^T + \gamma WA^T\right)_{ik}}{\left(A^T \Phi^T \Phi + \gamma DA^T\right)_{ik}} \right)$$

  end For

   Until convergence.

3.5 Anomaly Scoring

As discussed in subsection 3.3, given a newly observed vector, X', and a learned dictionary, Φ, we can acquire the optimal representation coefficient, α', through sparse representation. We define the abnormal score value of vector X' as a function of the representation cost as

$$AS(X') = 1 - \exp\left(-J(X', \alpha', \Phi)\right) \tag{13}$$

The abnormal score of vector X' is in the range from 0 to 1. It can be understood that the greater the score is, the higher the probability that the equipment being monitored fails is. Therefore, X' is detected as an anomaly if the score is greater than a predefined threshold, ε, which balances the false positive rate (FPR) and the false negative rate (FNR) and controls the sensitivity of the algorithm.

4 Experimental Results

4.1 Abnormal Node Detection Method

All experiments are conducted on a computer using the Windows 7 operating system with a 3.10 GHz processor and 32.00 GB of RAM

We implement the proposed approach on four pieces of equipment (i.e., the temperature, the pressure, the humidity and speed sensors) in a specified thermal power plant in China. We can backup the sensor data from the equipment from a DCS (distributed control system) at run-time. For a fixed sensor, (continuous or discrete) numerical data were gathered every ten minutes. Table 2 displays the characteristics of the datasets, including the equipment name, data size, related sensors, and data composition. In the experiment, every equipment dataset is divided into 2 parts as follows: 20% for training (dictionary learning) and 80% for testing. During the dictionary learning stage, the coefficient of sparsity regularized item λ is set to 2.0, and the coefficient of smooth item γ is set to 0.9. There is no mature theory to guide the value selection of parameter k, which is the number of bases of dictionary Φ. Considering the computational complexity and accuracy, after many tests, we set k to 256.

Table 2. Dataset Characteristics

Equipment	Edges (Sensor connections)	Nodes (wireless sensors)	Data composition
Temperature sensor	57824/2688	28	Normal 92%
			Abnormal 8%
Pressure sensor	57824/3260	32	Normal 89%
			Abnormal 11%
Humidity sensor	578243589	34	Normal 95%
			Abnormal 5%
Speed sensor	57824/2560	26	Normal 87%
			Abnormal 23%

To evaluate the performance of our proposed approach, we compare it with existing methods such as SVM, PCA+DT, BP, and SBM; the results are shown in Fig. 2. The first three methods are parametric classification-based methods that consider a two-class problem (normal or abnormal). The differences between these methods involve the feature extraction method and the construction of the classifier. Our proposed method is a non-parametric model that requires no

assumptions or prior knowledge of the equipment to be monitored; in this sense, it is a purely data-driven model.

We use the ROC graph to indicate the performance of five algorithms under multiple thresholds. Fig. 2. shows the ROC diagram of the relationship between the TPR and FPR under different threshold Settings. The TPR (also called recall or sensitivity) measures the proportion of actual positives (anomalies) that are correctly identified. The FPR (also called fallout) refers to the probability that negative results are incorrectly marked as positive samples during the test.

Fig. 2. illustrates the ROC curves of the SVM, PCA+DT, BP and SBM methods based on the temperature, pressure humidity and speed sensors in a thermal power plant. It is noted that the proposed method performs better than the other methods, i.e., it could obtain a higher TPR and a lower FPR.

An anomaly is a relative concept; its sensitivity to a threshold is predefined. In thermal power plants, the threshold can be set according to the functional demands of the users. The FPR is a very important index that is usually guaranteed to be greater than 90%. Usually, we will sacrifice precision to improve the FPR.
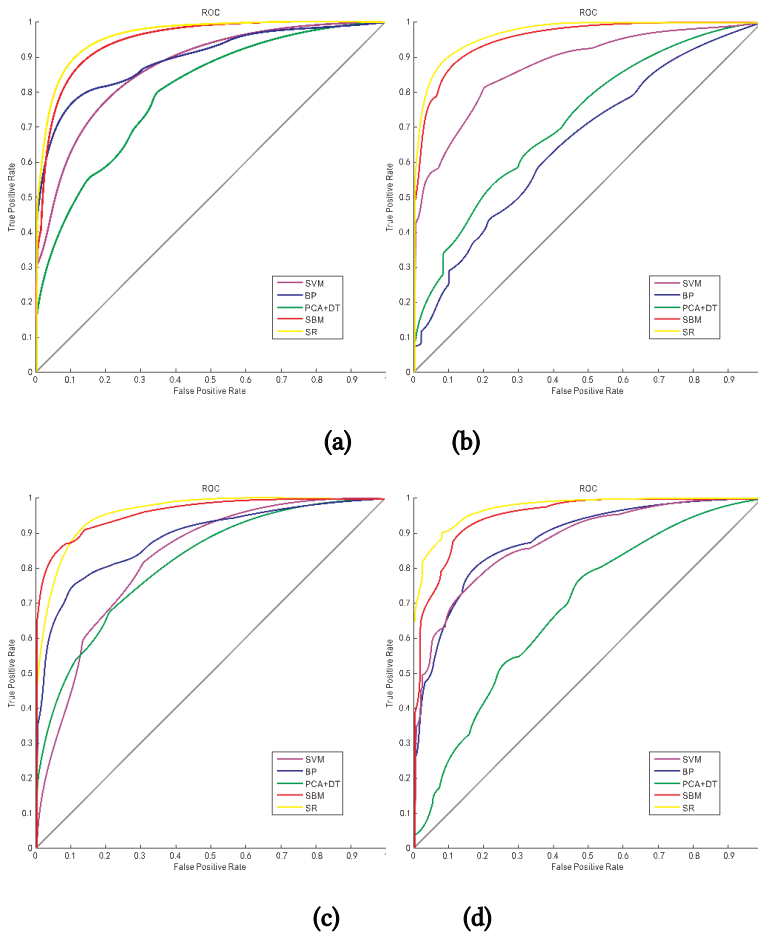

Fig. 2. Comparison of ROC curves: a) temperature sensor, b) pressure sensor, c) humidity sensor, d) speed sensor
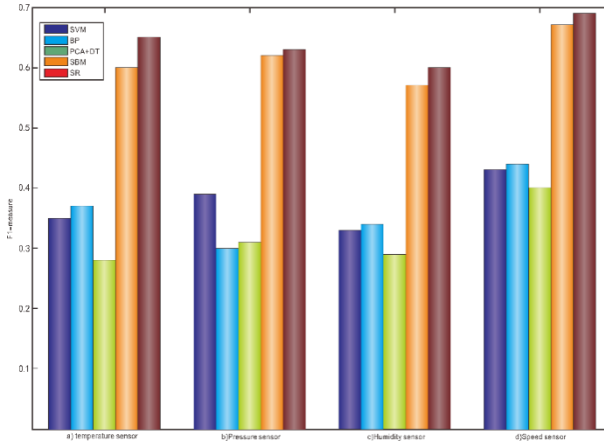
Fig. 3. F1-measure comparison

Fig. 3. presents an F1-measure comparison of the five methods. The F1-measures of the SVM, BP and PCA+DT methods, which are classification-based methods, are similar. The SVM and BP methods perform slightly better for the temperature, pressure and humidity sensors than the PCA+DT method. For different parameter sets (e.g., the PCA and the layer settings for BP), the performance of these methods will vary slightly. As a non-parametric model, the proposed method has a higher F1-measure than the other methods.

4.2 Sensor Network Dependencies

According to the abnormal nodes of a sensor network obtained in 4.1, the interdependence of multilayer sensor networks in thermal power plants is studied. Fig. 4. shows the schematic diagram of the multi-layer sensor network (to better compare the inter-layer dependencies, we assume the existence of the convergence sensor). We suppose that there are connections between multiple different sensors in a thermal power plant. Each sensor is a node in the network, and the connections between the same kind of sensors constitutes one layer of the sensor network. Separate sensor connections constitute the edges between sensor layers. It is further assumed that the multilayer sensor network converges to the same layer.
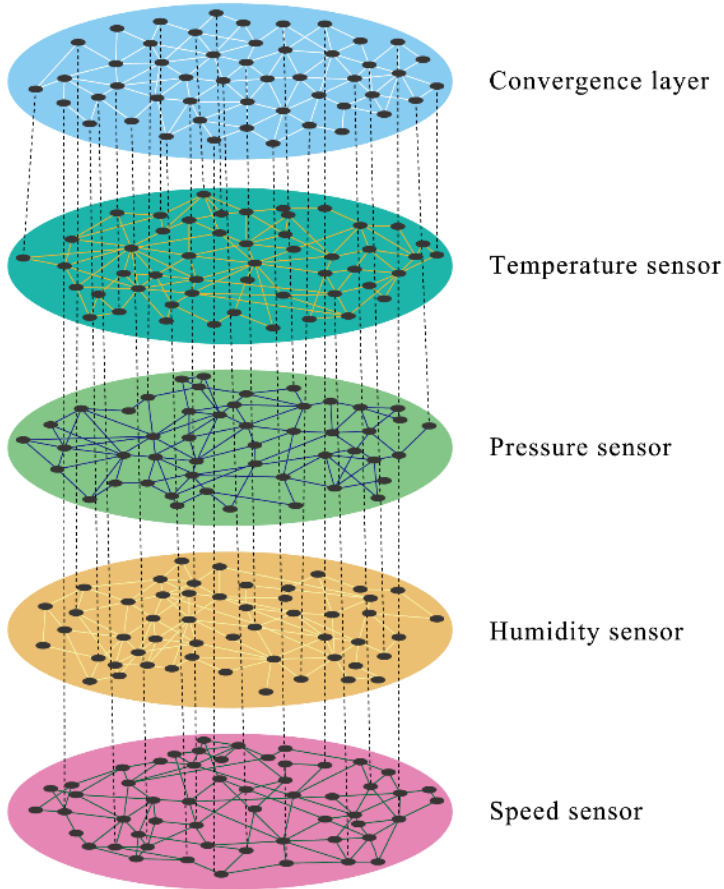
Fig. 4. Schematic diagram of a multilayer sensor network in a thermal power plant

Under the initial conditions, to better verify the sensor network dependencies, two operating sensor systems are used in two different networks: A (the convergence layer) and B (the other sensor layer). Each of the networks contains m_0 convergence sensors or pressure sensors and n_0 connection edges between sensors. In the operating system, in each time step, t_0, two new sensors will be added to operating systems A and B, and for the new sensor added to network A, $(1-q_A)m_A$ edges are connected to sensor network A using the method of preferential connections. The probability of a preferential connection depends on the connectivity of the existing nodes in sensor network A. In addition, the new sensor randomly or preferentially connects q_A and m_A connection edges to sensor network B as dependent edges. When another new node is added to sensor network B, an addition process is executed. In this process, the new node has $(1-q_n)m_B$ connection edges to be randomly connected to sub-sensor network B, and q_B m_B dependent edges are randomly or preferentially connected to sub-sensor network A.

Q_A and q_B are defined as the dependence intensity between two sensor networks. The higher q_A or q_B is, the more dependent edges there are in the sensor network or the more interdependent the two sensor networks are. In this paper, the abnormal node obtained according to formula (14) is used as the original node to obtain the dependence intensity of the two-layer sensor network.

$$q_{\alpha,\beta} = \frac{\left(k_i^{[\alpha]}k_i^{[\beta]}\right) - \left(k_i^{[\alpha]}\right)\left(k_i^{[\beta]}\right)}{\sigma_{i^{[\alpha]}}\sigma_{i^{[\beta]}}} \tag{14}$$

The square brackets represent an operation which averages all the nodes on the two layers, and

$$\sigma_{k^{[\alpha]}} = \sqrt{\left(k_i^{[\alpha]}k_i^{[\alpha]}\right) - \left(k_i^{[\alpha]}\right)^2} \tag{15}$$

When the size of the sensor network expands to N, the process of adding new sensor nodes will end. The two dependency relationships between dependent nodes are represented by the degree distribution, P, of the dependent edges. One type of dependency relationship follows an exponential distribution, which is expressed as

$$P\left(k_{dep}\right) = \frac{1}{mq+1}\left(\frac{mq}{mq+1}\right)^{k_{dep}-mq} \tag{16}$$

where $m1 \geq 1$ and $m_A = m_B = m$, $q_A = q_B = q$, which corresponds to the random dependency relationship between sensor networks.

4.3 Cascading Failure of Interdependent Sensor Network

During the initialization of the cascading failure iteration process, according to formula (16), the abnormal sensor nodes and all the connection edges are removed from sensor network A according to the proportion 1-P. When there is a failure of nodes in sensor network A, the dependent nodes in sensor network B will fail with a certain probability. In particular, assume that when only the sensors of one dependent edge in the entire sensor network can maintain normal functioning, there is further node failure of sensor network A. This dynamic process of iterative circulation will end only when there is no failure of sensor nodes in the system. A schematic diagram of the cascading failure process of a small interdependent network consisting of N=7 nodes is shown in Fig. 5.
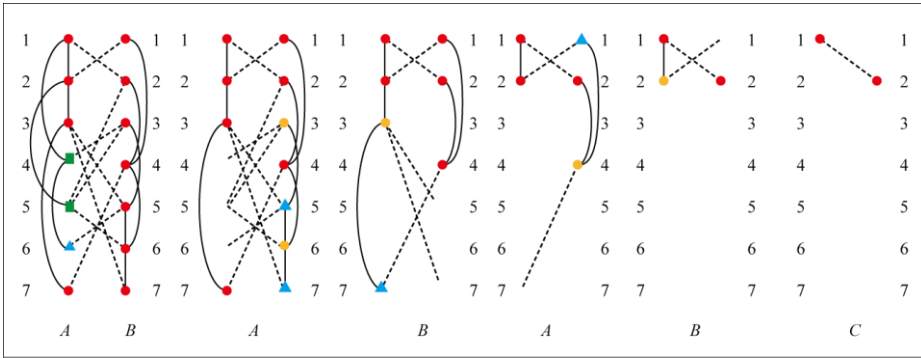
Fig. 5. Schematic diagram of the cascading failure process of an interdependent network

A solid line represents the connection edge to a layer, and a dotted line represents a dependent edge to a layer. A square represents an abnormal sensor. A triangle represents a sensor that does not belong to the normal network. A solid circle represents a sensor with no dependent edge. A hollow circle represents a normally functioning sensor. After two stages, the interdependent network will achieve a stable state. There will be no failure of a sensor node at this moment, and there will be only nodes 1 and 2 in the entire network.

To better illustrate the dependencies in multi-layer sensor networks, we will, based on the convergence layer and the temperature sensor layer, remove the abnormal nodes from the network and express the dependencies between the two layers through the exponential distribution dependence and the dependence in the form of the power law. The influence of these two dependencies on the cascading failure results of multilayer sensor networks is presented in Fig. 6. and Fig. 7. When the dependency intensity approaches zero, the system degenerates to the prototype stage (i.e., the critical penetration threshold of the monolayer scale-free network is achieved). When the dependence intensity, C, reaches 0.2 or so, the two different dependencies both correspond to the same critical value, indicating that when the dependence intensity is given, different sensor networks correspond to the same diagnostic value. However, when the dependence intensity, q_B, is greater than 0.2, the critical value under the exponential distribution dependence is greater than the critical value under the power law dependence. The results show that when the intensity of dependence is given, the intensity of dependence under the exponential distribution is less than that under the power law, which makes the robustness of the exponential distribution-dependent system weaker. In addition, as the correlation strength increases, the determination point value of multi-layer sensor failure increases gradually, indicating that the robustness of the system is weakening. The foremost reason is that the total number of connected and dependent edges of each sensor is fixed. The stronger the dependency strength is, the lower the number of edges connected by the current sensor is and the smaller the average connection strength is. The higher the number of helpless edges each sensor has in each subnet, the stronger the interdependence of the multi-layer sensor network is. This strong interdependence leads to an acceleration of the cascading failures that render the entire system vulnerable. This is in line with the introduction of other multi-layer network-related nodes. It was found that a coupled network with positively correlated dependent nodes was always more robust than a randomly paired network [57-58].
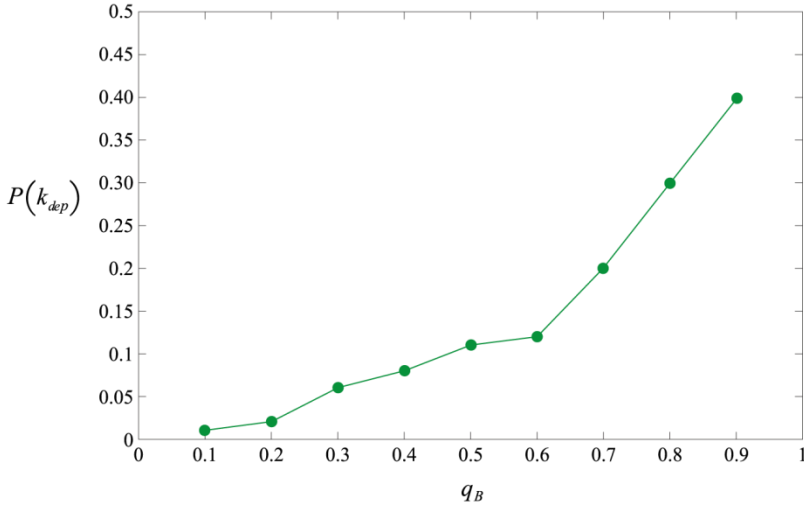
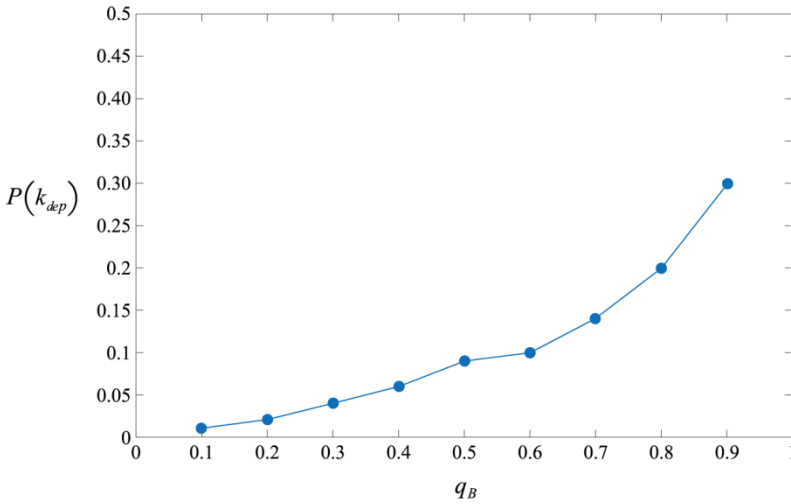Fig. 6. The change trend of P(K$_{dep}$) with intensity q$_B$ under exponential distribution dependence

Fig. 7. The change trend of P(K$_{dep}$) with intensity q$_B$ under power law distribution dependence

The influence of the interdependencies between discrete sensor networks and the convergence layer on system cascading failures is shown in Fig. 8. For weak dependence intensity, several different dependency relationships correspond to the same critical threshold value. Among them, the humidity sensor network and the pressure sensor network are more dependent on the convergence layer. The potential cause for this result is that when the dependence intensity is near a critical threshold value, it is close to the degree in which the node depends on the convergence layer, and when this degree is higher, the possibility of network cascading failure is also higher.

With the increase in dependence intensity, because the nodes in sensor network B have more dependent edges to sensor network A, more nodes can maintain their normal functions; thus, the proportion of nodes under two dependency relationships increases. Fig. 9. shows the influence of the other layers of the sensor network when the nodes in the convergence network layer are abnormal. When nodes on layer A become abnormal, the sensor will inaccurately estimate a parameter, and this kind of partial negative influence would spread to the entire network through a series of neighbouring sensors, which may result in the collapse of the distribution mechanism of the entire network.
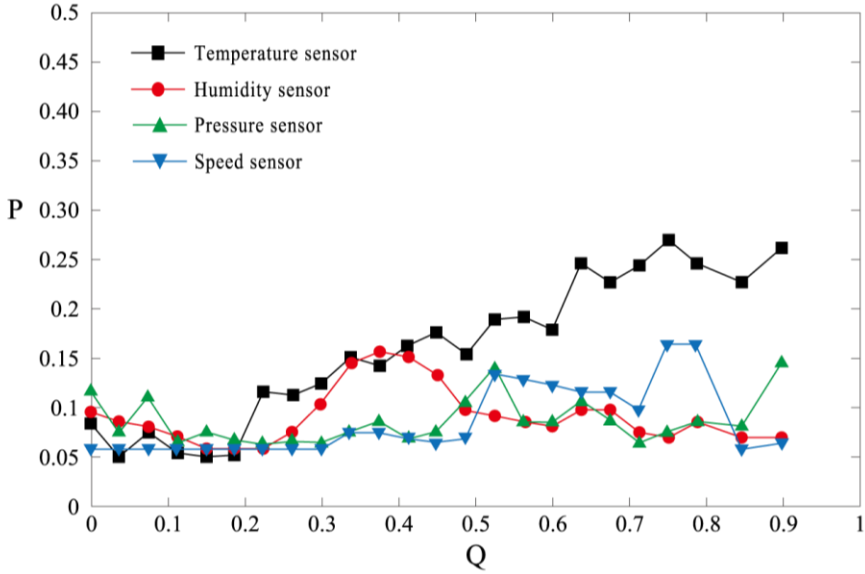

Fig. 8. Change trend according to the dependence intensity under two different dependency relationships
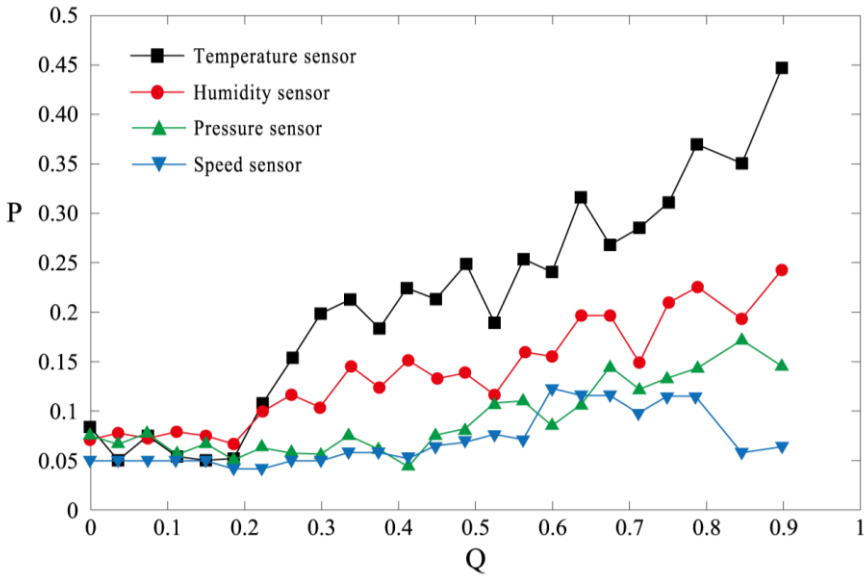
Fig. 9. Change trend of the nodes on the convergence layer according to the dependence intensity

5 CONCLUSION AND FUTURE WORK

A sparse representation-based detection model has been proposed in this paper to find abnormal nodes from the wireless sensor network of a thermal plant. A comparison has also been made between normal and anomalous data in the relationship database. Our proposed model considers the relationships that may exist between nodes. The raw data are preprocessed to convert relational data into a learned dictionary, and the abnormal nodes in a sensor network are found. The proposed method obtains a higher F1-value than the state-of-the-art methods. Because the sensor network is an interdependent network, removing a sensor affects the robustness of the entire system. In this paper, through experimentation with the convergence layer and the sensor network layer, we found that with the increase of dependence intensity, the critical point value gradually increases, indicating that system robustness is weakening. In addition, with the intensification of interdependency between sensor networks, the occurrence of a cascading failure accelerates. Furthermore, when there is an anomaly in the convergence layer, it is transmitted to the entire sensor network.

Although the proposed method (which is tested experimentally on actual equipment) has good detection results, there are still some open issues that need to be addressed in the future, such as the determination of the threshold value used to detect anomaly in the given WSN equipment. Also over time, the relationships between wireless sensors may change and the dictionary learned during training may become obsolete. In a sensor network, it is challenging and necessary to determine whether a sensor node is abnormal and to understand its influence across the network by dynamically adding nodes. We will consider these as our future work..

REFERENCES

[1]   Chandala V, Banerjee A, Kumar V. Anomaly Detection: A Survey, ACM Computing Surveys[J]. University of Minnesota, 2009.

[2]   Lazarevic A, Ertoz L, Kumar V, et al. A comparative study of anomaly detection schemes in network intrusion detection[C]//Proceedings of the 2003 SIAM International Conference on Data Mining. Society for Industrial and Applied Mathematics, 2003: 25-36.

[3]   Chang-Tien K Y L, Sirirat S. Survey of Fraud Detection Techniques in Networking, Sensing and Control[C]//IEEE International Conference. 2004, 2: 749-754.

[4]   Aggarwal C C. On anomaly detection in spuriously populated data streams[C]//Proceedings of the 2005 siam international conference on data mining. Society for Industrial and Applied Mathematics, 2005: 80-91.

[5]   Widodo A, Yang B S. Support vector machine in machine condition monitoring and fault diagnosis[J]. Mechanical systems and signal processing, 2007, 21(6): 2560-2574.

[6]   Kratz L, Nishino K. Anomaly detection in extremely crowded scenes using spatio-temporal motion pattern models[J]. 2009.

[7]   Akyildiz I F, Su W, Sankarasubramaniam Y, et al. A survey on sensor networks[J]. IEEE Communications Magazine, 2002, 40(8):102-114.

[8]   Xu G, Li W, Xu R, et al. An algorithm on fairness verification of mobile sink routing in wireless sensor network[J]. Personal and ubiquitous computing, 2013, 17(5): 851-864.

[9]   Radicchi F, Castellano C, Cecconi F, et al. Defining and identifying communities in networks[J]. Proceedings of the National Academy of Sciences, 2004, 101(9): 2658-2663.

[10] Cai D, Shao Z, He X, et al. Mining hidden community in heterogeneous social networks[C]//Proceedings of the 3rd international workshop on Link discovery. ACM, 2005: 58-65.

[11] Gao J, Buldyrev S V, Stanley H E, et al. Percolation of a general network of networks[J]. Physical Review E, 2013, 88(6): 062816.

[12] Cellai D, López E, Zhou J, et al. Percolation in multiplex networks with overlap[J]. Physical Review E, 2013, 88(5): 052811.

[13] Schneider C M, Yazdani N, Araújo N A M, et al. Towards designing robust coupled networks[J]. Scientific reports, 2013, 3: 1969.

[14] Boccaletti S, Bianconi G, Criado R, et al. The structure and dynamics of multilayer networks[J]. Physics Reports, 2014, 544(1): 1-122.

[15] Wang S, Hong L, Chen X, et al. Review of interdependent infrastructure systems vulnerability analysis[C]//Intelligent Control and Information Processing (ICICIP), 2011 2nd International Conference on. IEEE, 2011, 1: 446-451.

[16] Lee II E E, Mitchell J E, Wallace W A. Restoration of services in interdependent infrastructure systems: A network flows approach[J]. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 2007, 37(6): 1303-1317.

[17] De Gante A, Aslan M, Matrawy A. Smart wireless sensor network management based on software-defined networking[C]//Communications (QBSC), 2014 27th Biennial Symposium on. IEEE, 2014: 71-75.

[18] Xiong N, Vasilakos A V, Yang L T, et al. Distributed Explicit Rate Schemes in Multi-Input–Multi-Output Network Systems[J]. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 2010, 40(4): 448-459.

[19] Islam K, Shen W, Wang X. Wireless sensor network reliability and security in factory automation: A survey[J]. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 2012, 42(6): 1243-1256.

[20] Xu G, Shen W, Wang X. Applications of WSN in marine environment monitoring: A survey[J]. Sensors, 2014, 14(9): 16932-16954.

[21] Li X M, Yuan L, Liu C C, et al. An Efficient Critical Incident Propagation Model for Social Networks Based on Trust Factor[C]//International conference on collaborative computing: Networking, applications and worksharing. Springer, Cham, 2017: 416-424.

[22] Venkatesan L, Shanmugavel S, Subramaniam C. A survey on modeling and enhancing reliability of wireless sensor network[J]. Wireless Sensor Network, 2013, 5(03): 41.

[23] Wang W, Li X, Jiao P, et al. Exploring Intracity Taxi Mobility during the Holidays for Location-Based Marketing[J]. Mobile Information Systems, 2017, 2017.

[24] Xu G, Zhang Y, Sangaiah A K, et al. CSP-E2: An abuse-free contract signing protocol with low-storage TTP for energy-efficient electronic transaction ecosystems[J]. Information Sciences, 2019, 476: 505-515.

[25]   Zhu X, Lu Y, Han J, et al. Transmission reliability evaluation for WSN[J]. International Journal of Distributed Sensor Networks, 2016, 12(2): 1346079.

[26]   Sangaiah A K, Medhane D V, Han T, et al. Enforcing Position-Based Confidentiality with Machine Learning Paradigm through Mobile Edge Computing in Real-Time Industrial Informatics[J]. IEEE Transactions on Industrial Informatics, 2019.

[27]   Steinwart I, Hush D, Scovel C. A classification framework for anomaly detection[J]. Journal of Machine Learning Research, 2005, 6(Feb): 211-232.

[28]   Muandet K, Schölkopf B. One-class support measure machines for group anomaly detection[J]. arXiv preprint arXiv:1303.0309, 2013.

[29]   Chen C, Zhang D, Castro P S, et al. iBOAT: Isolation-based online anomalous trajectory detection[J]. IEEE Transactions on Intelligent Transportation Systems, 2013, 14(2): 806-818.

[30]   Yin J, Lo W, Deng S, et al. Colbar: A collaborative location-based regularization framework for QoS prediction[J]. Information Sciences, 2014, 265: 68-84.

[31]   Yin Y, Deng S. Analysing and determining substitutability of different granularity Web services[J]. International Journal of Computer Mathematics, 2013, 90(11): 2201-2220.

[32]   Mukkamala S, Janoski G, Sung A. Intrusion detection using neural networks and support vector machines[C]//Neural Networks, 2002. IJCNN'02. Proceedings of the 2002 International Joint Conference on. IEEE, 2002, 2: 1702-1707.

[33]   Knox E M, Ng R T. Algorithms for mining distancebased outliers in large datasets[C]//Proceedings of the international conference on very large data bases. Citeseer, 1998: 392-403.

[34]   Kriegel H P, Zimek A. Angle-based outlier detection in high-dimensional data[C]//Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2008: 444-452.

[35]   Breunig M M, Kriegel H P, Ng R T, et al. LOF: identifying density-based local outliers[C]//ACM sigmod record. ACM, 2000, 29(2): 93-104.

[36]   Leung H, Haykin S. The complex backpropagation algorithm[J]. IEEE Transactions on signal processing, 1991, 39(9): 2101-2104.

[37]   Chen K Y, Chen L S, Chen M C, et al. Using SVM based method for equipment fault detection in a thermal power plant[J]. Computers in industry, 2011, 62(1): 42-50.

[38]   Shao J, Buldyrev S V, Havlin S, et al. Cascade of failures in coupled network systems with multiple support-dependence relations[J]. Physical Review E, 2011, 83(3): 036116.

[39]   Wegerich S. Similarity-based modeling of vibration features for fault detection and identification[J]. Sensor Review, 2005, 25(2): 114-122.

[40]   Danziger M M, Bashan A, Berezin Y, et al. An introduction to interdependent networks[C]//International Conference on Nonlinear Dynamics of Electronic Systems. Springer, Cham, 2014: 189-202.

[41]   Dong G, Gao J, Tian L, et al. Percolation of partially interdependent networks under targeted attack[J]. Physical Review E, 2012, 85(1): 016112.

[42]   Li X M, Tian Q, Tang M, et al. Local community detection for multi-layer mobile network based on the trust relation[J]. Wireless Networks, 2019: 1-13.

[43]   Gao J, Buldyrev S V, Stanley H E, et al. Networks formed from interdependent networks[J]. Nature physics, 2012, 8(1): 40.

[44]   Luo X S, Zhang B. Analysis of cascading failure in complex power networks under the load local preferential redistribution rule[J]. Physica A: Statistical Mechanics and its Applications, 2012, 391(8): 2771-2777.

[45]   Crucitti P, Latora V, Marchiori M. Model for cascading failures in complex networks[J]. Physical Review E, 2004, 69(4): 045104.

[46]   Kinney R, Crucitti P, Albert R, et al. Modeling cascading failures in the North American power grid[J]. The European Physical Journal B-Condensed Matter and Complex Systems, 2005, 46(1): 101-107.

[47]   Amini M H, Boroojeni K G, Iyengar S S, et al. Sustainable Interdependent Networks II[J]. 2018.

[48]   Liu X, Stanley H E, Gao J. Breakdown of interdependent directed networks[J]. Proceedings of the National Academy of Sciences, 2016, 113(5): 1138-1143.

[49]   Radicchi F. Percolation in real interdependent networks[J]. Nature Physics, 2015, 11(7): 597.

[50]   Rifkin R, Yeo G, Poggio T. Regularized least-squares classification[J]. Nato Science Series Sub Series III Computer and Systems Sciences, 2003, 190: 131-154.

[51]   Angiulli F, Pizzuti C. Outlier mining in large high-dimensional data sets[J]. IEEE transactions on Knowledge and Data engineering, 2005, 17(2): 203-215.

[52]   Zeng X, Xu G, Zheng X, et al. E-AUA: An Efficient Anonymous User Authentication Protocol for Mobile IoT[J]. IEEE Internet of Things Journal, 2018.

[53] Chandola V, Banerjee A, Kumar V. Anomaly detection: A survey[J]. ACM computing surveys (CSUR), 2009, 41(3): 15.

[54] Lee D D, Seung H S. Algorithms for non-negative matrix factorization[C]//Advances in neural information processing systems. 2001: 556-562.

[55] Cai D, He X, Wu X, et al. Non-negative matrix factorization on manifold[C]//Data Mining, 2008. ICDM'08. Eighth IEEE International Conference on. IEEE, 2008: 63-72.

[56] Bresar M. Centralizing Mappings and Derivations in Prime Rings[J]. Journal of Algebra, 1993, 156(2):385-394.

[57] Parshani R, Rozenblat C, Ietri D, et al. Inter-similarity between coupled networks[J]. EPL (Europhysics Letters), 2011, 92(6): 68002.

[58] Cho W, Goh K I, Kim I M. Correlated couplings and robustness of coupled networks[J]. arXiv preprint arXiv:1010.4971, 2010..