







# Data Reconstruction via Consensus Graph Learning for Effective Anomaly Detection in Industrial IoT

Lin Li , Hongchun Qu , Zhaoni Li , Jian Zheng , Xiaoming Tang , and Ping Liu 

**Abstract**—In the industrial Internet of Things (IIoT), anomaly detection is fundamental to ensuring system safety and product quality, among other things. However, the massive amount of unlabeled, high-dimensional data generated in the IIoT challenges some existing anomaly detection methods as follows: 1) the distance concentration caused by the high dimensionality of the data can cause a decrease in the detection accuracy of some of the methods; and 2) some of the methods fail to explore the intrinsic relationships between the data, resulting in less effective detection of anomalies in the data. To handle the above challenges, a framework named data reconstruction via consensus graph learning (DRCG) and two anomaly score functions are proposed. Specifically, DRCG overcomes the distance concentration problem and explores the intrinsic relationships of the data by integrating projection learning, low-dimensional embedding, and consensus graph learning into a unified objective function. Then, an iterative algorithm is designed to solve the DRCG model. By doing so, DRCG not only drives the reconstruction error of abnormal samples higher than that of normal samples, but also obtains the projection that can effectively extract the intrinsic relationship between the data. Furthermore, to identify anomalies in the data, two anomaly score functions based on the reconstruction error and projection are designed, respectively. To achieve online anomaly detection for streaming data, DRCG with the projection-based anomaly score function is extended into an online version.

The effectiveness and superiority of the proposed methods have been demonstrated on four real-world datasets.

**Index Terms**—Anomaly detection, data reconstruction, graph learning, industrial Internet of Things (IIoT), unsupervised learning.

## I. INTRODUCTION

ADVANCES in technology are making it possible for the industrial Internet of Things (IIoT) to be widespread in the food processing industry [1], smart cities [2], and urban informatics [3], etc. With the increase in devices and applications, a massive amount of complex unlabeled data is generated in industrial cyber-physical systems [4]. Since the relationships between the data are the basis for intelligent automation and decision-making, exploring the relationships between the data plays a crucial role in the IIoT. However, due to the complexity of scale and storage, the occurrence of data anomalies is inevitable. Failure to effectively detect these anomalous data may pose a significant risk to the IIoT applications. Therefore, an effective unsupervised anomaly detection method is urgently needed.

To detect anomalies in data, machine learning techniques are introduced. However, it is still a troublesome task to perform reliable results from the high-dimensional (HD) anomaly data in the IIoT. The reason for the trouble with the above task lies in the following two characteristics of the HD anomaly data. First, as the number of sensors increases, the attributes of each recorded data also increase. Generally, multiple sensors are used in the IIoT, resulting in the recorded data being high dimensional. Second, there are various methods to attack IIoT devices. As technology evolves, these attack methods are constantly being updated. This results in the possibility of different types of abnormal data being recorded.

The characteristics of the industrial anomaly data discussed cause two challenges for some existing anomaly detection methods. The first challenge is to deal with the problem of distance concentration caused by the high dimensionality of the data [5]. As the dimensionality of the data increases, all distances between observations tend to be indiscernible, so that distance or density metrics fail to capture the neighborhood information of the data. This problem causes anomaly detection methods based on distance or density metrics to be unable to detect anomalies in the data. The second challenge is how to effectively preserve the

Manuscript received 13 October 2022; revised 17 February 2023, 6 April 2023, and 2 August 2023; accepted 6 September 2023. Date of publication 4 October 2023; date of current version 23 February 2024. This work was supported in part by the National Natural Science Foundation of China under Grant 61871061, in part by Project of Science and Technology Department of Qinghai Province under Grant 2020-SF-139, and in part by Project of Advanced Scientific Research Institute of CQUPT under Grant E011A2022329. Paper no. TII-22-4272. (Corresponding author: Hongchun Qu.)

Lin Li, Hongchun Qu, and Jian Zheng are with the College of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing 400065, China (e-mail: linligx@126.com; hcchyu@gmail.com; zhengjian.002@163.com).

Zhaoni Li is with the College of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing 400065, China, and also with the College of Computer, Qinghai Normal University, Xining 810008, China (e-mail: lizhaoni01@163.com).

Xiaoming Tang and Ping Liu are with the College of Automation, Chongqing University of Posts and Telecommunications, Chongqing 400065, China (e-mail: txmmyeye@126.com; liuping\_cqupt@cqupt.edu.cn).

Color versions of one or more figures in this article are available at <https://doi.org/10.1109/TII.2023.3316220>.

Digital Object Identifier 10.1109/TII.2023.3316220

intrinsic relationships of normal samples in the data. In particular, the effective retention of such relationships can improve the detection accuracy of the model in the face of various different types of anomalies. However, some subspace-based anomaly detection methods fail to effectively explore the intrinsic relationships of normal samples in the data, which makes these methods less effective in practical applications.

To address the above challenges, a framework of data reconstruction via consensus graph learning (DRCG) is proposed. Specifically, to overcome the first challenge, we use a reconstruction-based method to avoid the dependence on distance calculation. To better solve the second challenge, the reconstructed data is necessary to fully preserve the relationships between the normal samples in the original dataset. The purpose of doing so is to amplify the reconstructed differences between abnormal samples, and to reduce the reconstructed differences of normal samples. To achieve the above purposes, we introduce self-expressiveness technique to preserve the global relationship between the data and introduce the consensus graph learning technique to preserve the local relationship between the data. Notice that the self-expressiveness technique does not rely on distance calculation. Since consensus graph learning techniques require distance computation, it suffers from the first challenge. To suppress the extent of this challenge, the kernel function technique is introduced. However, for different datasets, the appropriate kernel functions might be different. Thus, the selection of an appropriate kernel function is a challenging task. To avoid this problem, we introduce multiple kernel functions in the model to learn multiple graphs, and then obtain a consensus graph. In doing so, during data reconstruction, we can not only reduce the negative impact of the distance concentration problem, but also obtain reconstructed data that preserve the global and local structure of the data. In combination with two newly designed anomaly score functions, our methods can cope with the above challenges to realize accurate and efficient anomaly detection. The main contributions of our article are stated as follows.

- 1) A framework of DRCG is newly designed, which integrates projection learning, low-dimensional embedding, and consensus graph learning into a unified objective function. DRCG is found to result in higher reconstructed differences in abnormal samples than in normal samples.
- 2) Three loss functions are defined to constrain the similarity matrix so that a valid consensus graph can be learned by DRCG. In doing so, the local structure of the data can be preserved by the reconstructed data and projection. Moreover, the design of the approximated low-dimensional embedding loss can be used to avoid the over-fitting problem of low-dimensional embedding learning.
- 3) An iterative algorithm is developed for the DRCG model. Furthermore, two anomaly score functions based on DRCG are carefully designed to identify anomalies. Moreover, DRCG with the projection-based anomaly score function is extended to an online version for online anomaly detection of streaming data. The proposed methods are found to be effective in detecting anomalies in IIoT data.

The rest of this article is organized as follows. In Section II, the related work will be introduced. In Section III, DRCG framework is built. In Section IV, DRCG is used for anomaly detection. In Section V, experiments are set up to demonstrate the effectiveness of the proposed method. Finally, Section VI concludes this article.

## II. RELATED WORK

### A. Anomaly Detection Methods

Currently, mainstream anomaly detection methods that are used to detect data anomalies in the IIoT include distance-based, density-based, ensemble-based and subspace-based methods. In this section, we briefly analyze these methods.

The distance-based methods, such as K-nearest neighbor [6] and local distance-based outlier factor (LDOF) [7], show better performance in the presence of a few anomalies [8]. However, the Euclidean distance between observations is commonly used by most distance-based methods as their similarity metric, which may cause the problem of distance concentration.

The density-based methods detect anomalies by comparing the relative density around an observation with the density around its local neighbors. Local outlier factor (LOF) [9], as one of the density-based methods, which measures the degree of outlying of an observation by comparing the observation relative to its neighbors. The density-based methods were found to be more effective than the distance-based ones. However, the detection accuracy of the density-based methods decreases when dealing with HD data, due to the fact that the accuracy of the density estimation process decreases in HD space.

The motivation for ensemble-based methods is to utilize the advantages of multiple methods for anomaly analysis. Usually, these methods combine the outputs of multiple algorithms or multiple detectors. Isolated Forest (IForest) [10] is one of the ensemble-based methods that builds a collection of trees and identifies an anomaly as an instance with a short path length in the tree. However, IForest suffers from trouble detecting local anomalies in datasets with multiple clusters of normal instances since normal clusters of similar density mask the local anomalies.

To efficiently handle HD datasets, fast ABOD (FABOD) [11] was proposed, which is based on the principle of measuring the variance of the angular spectrum of observations to distinguish anomalies. However, FABOD is not applicable to anomaly detection in batch mode since the user has to use all instances to calculate the required angle information. Based on the assumption that anomalies usually exhibit unusual behavior in different low-dimensional subspaces, subspace learning is introduced to handle the anomaly detection problem for HD data. Oversampling principal components analysis (OSPCA) [12] is one of the subspace-based methods. Specifically, the leave-one-out strategy and PCA are used by OSPCA to determine whether an observation is anomalous. Then, OSPCA was extended into online OSPCA (OOSPCA) [13] through an online update technique. Recently, the low-rank and K-nearest neighbor techniques are used in local projection-based outlier detection (LOPD) [5] to extract the local neighborhood information of an observation

for anomaly detection. Graph is excellent in describing the internal relationships between observations in the data [14]. Based on the graph theory, anomaly detection with kernel preserving embedding (ADKE) [15] was proposed. First, the kernel-preserving embedding and dual nuclear norm are used by ADKE to construct a graph that describes the intrinsic relationships of the data. Then, based on the graph, a random walk method is utilized to identify anomalies. With the widespread application of deep learning, the graph is often treated as a regularization term in deep learning methods. For example, robust graph autoencoder (RGAE) [16], a joint AE and graph regularization framework, which is used for hyperspectral anomaly detection. Although the above methods work well, they cannot be easily extended to anomaly detection problems in streaming data.

### B. Online Anomaly Detection Methods

To deal with streaming data, sliding window technology is usually applied. Zhang et al. [17] proposed an unsupervised anomaly detection for streaming data by using the sliding window technology and probability density-based descriptors. However, it is challenging to determine the optimal window size for the window-based approach. To handle high-dimensional streaming data, Kurt et al. in [18] proposed PCA-based online anomaly detection (PCA-OAD). The two phases, offline and online, are used by PCA-OAD to achieve online anomaly detection. Notably, the low-dimensional manifolds extracted by PCA have a direct relationship to the detection accuracy of the PCA-OAD. However, PCA is a global method, thus it cannot extract low-dimensional manifolds that reflect local relationships within the data. Consequently, the application of PCA-OAD is limited.

## III. PROPOSED METHOD

### A. Problem Definition

For a dataset  $X = \{x_1, x_2, \dots, x_n\} \in \mathbb{R}^{m \times n}$ , where  $x_i \in \mathbb{R}^{m \times 1}$  is the  $i$ th sample of  $X$  and  $m$  is the number of dimensions. The purpose of this article is to identify anomalous samples in HD data, i.e., the value of  $m$  is relatively large. However, the high dimensionality of the data can cause distance concentration problem, which leads to low effectiveness of anomaly detection methods based on distance or density metrics. To overcome this challenge, one usually uses dimensionality reduction methods to embed the data  $x_i$  directly or project it through a projection matrix  $P \in \mathbb{R}^{m \times d}$  into a low-dimensional subspace, where  $d$  is the number of dimensions of the low-dimensional subspace. We define the low-dimensional embedding  $Y^T = \{y_1, y_2, \dots, y_n\} \in \mathbb{R}^{d \times n}$ , where  $y_i \in \mathbb{R}^{d \times 1}$  is the  $i$ th sample of  $Y^T$ . Although these dimensionality reduction methods can overcome the above challenge, some of them fail to fully explore the intrinsic relationships between the data (e.g., PCA can only explore the global relationships between the data while ignoring their local relationships), resulting in unsatisfactory results.

To address the above two challenges, we proposed DRCG framework. The objective function of DRCG will be introduced in the next section.

### B. Framework of DRCG

Motivated by dimensionality reduction methods, projection learning is introduced to make connections with low-dimensional embedding to alleviate the distance concentration problem. Ideally,  $y_i$  can be represented by a projection  $P$  as

$$y_i = P^T x_i. \quad (1)$$

To avoid the dependence on distance calculation, a data reconstruction approach is introduced. Under the condition ( $P^T P = I$ ), the raw sample  $x_i$  can be reconstructed as  $\bar{x}_i = P y_i$ , where  $I$  is an identity matrix of appropriate size. Considering that the reconstruction data needs to preserve the global relationship between the raw data, a self-expressiveness technique is introduced. According to self-expressiveness assumption, i.e., each observation is represented as a linear combination of the other observations. Thus, we have

$$\bar{x}_i = P Y^T s_i \quad (2)$$

where  $s_i \in \mathbb{R}^{n \times 1}$  is the representation coefficient vector corresponding to sample  $x_i$ . Thus, the sum of the errors between each sample  $x_i$  and its reconstructed sample  $\bar{x}_i$  is

$$\sum_{i=1}^n \|x_i - \bar{x}_i\|_2^2 = \sum_{i=1}^n \|x_i - P Y^T s_i\|_2^2 = \|X - P Y^T S\|_F^2$$

where  $S = \{s_1, s_2, \dots, s_n\} \in \mathbb{R}^{n \times n}$  denotes as representation graph. Note that the  $ij$ th element of  $S$  is  $s_{ij}$ . To make the reconstructed data close to the raw data, we have

$$\min_{P, Y, S} \|X - P Y^T S\|_F^2, \text{ s.t. } P^T P = I. \quad (3)$$

In model (3), the purpose of optimizing variables  $P$ ,  $Y$  and  $S$  is to amplify the difference between normal and abnormal samples. It is widely known that the matrix  $S$  is crucial to exploit the intrinsic structure of the data. To enable the matrix  $S$  to better describe the relationship between data, we introduce distance similarity, i.e., an approximated low-dimensional embedding loss, to model (3)

$$\begin{aligned} \min_{P, Y, S} & \|X - P Y^T S\|_F^2 + \alpha \sum_{i,j=1}^n \|y_i - f_j\|_2^2 s_{ij} \\ \text{s.t. } & P^T P = I \end{aligned} \quad (4)$$

where  $\alpha$  is a parameter.  $F^T = \{f_1, f_2, \dots, f_n\} \in \mathbb{R}^{d \times n}$  is a predefined low-dimensional embedding, where  $f_i \in \mathbb{R}^{d \times 1}$  is the  $i$ th sample of  $F^T$ . Here,  $F$  is used to avoid over-fitting of the low-dimensional embedding  $Y^T$  in the update process. Notice that the distance  $\|y_i - f_j\|_2^2$  is feasible since it is executed in the low-dimensional subspace. By doing so, the distance concentration problem can be alleviated. However, the matrix  $S$  in problem (4) lacks modeling of the local relationships between the raw data. To better explore the local relationships between the data, an adaptive graph learning model has recently been



proposed as follows:

$$\min_S \sum_{i,j=1}^n \|x_i - x_j\|_2^2 (s_{ij})^r, \text{ s.t. } \sum_{j=1}^h s_{ij} = 1, 0 \leq s_{ij} \quad (5)$$

where  $r$  is set for avoiding trivial solution, and  $h$  is the neighborhood parameter. The model (5) is introduced to select the  $h$  closest samples to  $x_i$  to reconstruct  $\bar{x}_i$ . The constraint, i.e.,  $\sum_{j=1}^h s_{ij} = 1, 0 \leq s_{ij}$ , is to avoid the extreme case where any element of  $s_i$  is zero. To facilitate solving for  $S$ ,  $r$  is usually set to 2 [19]. Note that the graph in problem (5) is designed to the original HD space. By doing so, the distance concentration problem will cause the learned graph to fail in capturing the neighborhood information. Inspired by [20], we introduce a kernel technology to solve the above problem. Specifically, let  $\phi$  be a mapping function that maps the data from the original space to the reproducing kernel Hilbert space  $H$ . Then, we have

$$\min_S \sum_{i,j=1}^n -\text{Ker}(x_i, x_j) (s_{ij})^2, \text{ s.t. } \sum_{j=1}^h s_{ij} = 1, 0 \leq s_{ij} \quad (6)$$

where  $-\text{Ker}(x_i, x_j)$  can present the similarity between  $x_i$  and  $x_j$  in  $H$ . However, for different datasets, the appropriate kernel functions might be different. Thus, the selection of an appropriate kernel function is a challenging task. To avoid this problem, we introduce multiple graphs  $(\{S^k\}_{k=1}^v)$  by varying the mapping function  $\phi$  in the model (6), where  $S^k$  denotes the  $k$ th graph and  $v$  indicates the number of the candidate graphs.

To incorporate the generated multiple graphs into the model (4), consensus graph learning is introduced to obtain a consensus graph that can represent the multiple graphs

$$\begin{aligned} \min_{S,w} \sum_{k=1}^v (w_k)^2 \|S - S^k\|_F^2 \\ \text{s.t. } \sum_{j=1}^h s_{ij} = 1, 0 \leq s_{ij}, s_{ii} = 0, w^T \mathbf{1} = 1, 0 \leq w_k \end{aligned} \quad (7)$$

where the  $i$ th entry of  $w = \{w_1, \dots, w_v\}$  is the reward value of the  $S^i$ , and  $\mathbf{1}$  is a vector of appropriate size and all the elements are 1. Similarly, The constraint, i.e.,  $w^T \mathbf{1} = 1, 0 \leq w_k$ , is also to avoid the extreme case where any element of  $w$  is zero. From model (6), we can know that  $(s_{ij})^2$  is used to describe the similarity between  $x_i$  and  $x_j$ . Thus, we introduced  $s_{ii} = 0$  is to avoid describing the similarity between  $x_i$  and itself. Overall, the model of DRCG can be obtained by combining the model (7) and the model (4) to form the following optimization problem:

$$\begin{aligned} \min_{P,S,Y,w} \|X - PY^T S\|_F^2 + \alpha \sum_{i,j=1}^n \|y_i - f_j\|_2^2 s_{ij} \\ + \beta \sum_{k=1}^v (w_k)^2 \|S - S^k\|_F^2 \\ \text{s.t. } \sum_{j=1}^h s_{ij} = 1, 0 \leq s_{ij}, s_{ii} = 0, Y^T Y = I \\ P^T P = I, w^T \mathbf{1} = 1, 0 \leq w_k \end{aligned} \quad (8)$$

where  $\beta$  is a tradeoff parameter. The orthogonal constraint is imposed on  $Y$  for avoiding trivial solution.

In model (8), the first term is the error reconstruction term, which is used to make the reconstructed sample recover the original sample as much as possible, and to explore the global relationship between the data. The second term is an approximate low-dimensional embedding loss term, which is introduced to explore the relationship between the samples with distance and to avoid over-fitting when updating  $F$ . The third term is the consensus graph learning term, which is introduced to enable the model to explore local relationships between the data. Revisiting model (8), we use reconstruction technique, kernel technique, and a technique for computing distances between samples in low-dimensional space to solve the first challenge (i.e., the distance concentration problem). Further, we effectively explore the local and global intrinsic structure between data by consensus graph learning and self-expressiveness techniques to address the second challenge (i.e., explore intrinsic relationships in the data). Therefore, the benefits of connecting projection learning and consensus graph learning are twofold. First, the two problems investigated in this article can be effectively addressed. Second, projection learning and consensus graph learning can promote higher reconstruction differences produced in abnormal samples than in normal samples. In doing so, the accuracy of anomaly detection can be improved. Furthermore, the learned projection can be used to enable online anomaly detection.

### C. Optimization of DRCG Framework

In this subsection, the problem (8) will be solved by the alternating iterative algorithm developed by us. Specifically, the variables  $P$ ,  $S$ ,  $Y$  and  $w$  in the problem (8) are solved by the following subproblems, respectively.

1) *Update Y*: By fixing the other variables, updating  $Y$  through the following problem:

$$\min_{Y^T Y = I} \text{Tr}(Y^T (\alpha D + S S^T) Y) - \text{Tr}(Y^T (2 S X^T P + 2 \alpha S F)) \quad (9)$$

where  $\text{Tr}(\cdot)$  is the trace operation, and  $D = \text{diag}(d_{11}, d_{22}, \dots, d_{nn})$  is a diagonal matrix with the diagonal element  $d_{ii} = \sum_j s_{ij}$ . The minimization with orthogonality constraint can be fast solved by an efficient algorithm proposed in [21].

2) *Update P*: With the other variables fixed, updating  $P$  via the following problem:

$$\max_{P^T P = I} \text{Tr}(P Y^T S X^T). \quad (10)$$

Due to the condition  $P^T P = I$ , the problem (10) can be transformed into the following problem:

$$\max_{P^T P = I} \text{Tr}(P^T Q) \quad (11)$$

where  $Q = X S^T Y$ . The problem (11) is a Procrustes problem [22] and can be solved as

$$P = J V^T \quad (12)$$

where  $J, V$  are orthogonal matrices from the  $SVD(Q) = J Z V^T$ , and  $SVD(Q)$  represents the singular value decomposition of the matrix  $Q$ .

3) *Update S*: By fixing the variables  $P$ ,  $Y$ , and  $w$ , updating  $S$  via the following problem:

$$\begin{aligned} \min_S & \text{Tr}(-2XS^TY P^T) + \text{Tr}(Y^T S S^T Y) \\ & + \text{Tr}(\alpha K^T S) + \beta \sum_{k=1}^v (w_k)^2 \|S - S^k\|_F^2 \\ \text{s.t.} & \sum_{j=1}^h s_{ij} = 1, 0 \leq s_{ij}, s_{ii} = 0 \end{aligned} \quad (13)$$

where the element  $k_{ij}$  in the matrix  $K$  is  $k_{ij} = \|y_i - f_j\|_2^2$ . To fast solve the problem (13), a two-step fast approximation strategy [20] is used. First, the problem (13) is solved without the constraint to obtain

$$\begin{aligned} S^* &= \left( \beta \sum_{k=1}^v (w_k)^2 I + Y Y^T \right)^{-1} (2Y P^T X \\ &+ \alpha K + 2\beta \sum_{k=1}^v (w_k)^2 S^k). \end{aligned} \quad (14)$$

Then, the problem (13) on  $S$  can be solved as

$$\min_S \|S - S^*\|_F^2, \text{ s.t. } \sum_{j=1}^h s_{ij} = 1, 0 \leq s_{ij}, s_{ii} = 0. \quad (15)$$

The problem (15) requires the projection of a point onto the capped simplex, which can be solved by an efficient iterative algorithm, as proposed in [23].

4) *Update w*: By fixing the variables  $P$ ,  $Y$ , and  $S$ , updating  $w$  via the following problem:

$$\min_w \sum_{k=1}^v (w_k)^2 \|S - S^k\|_F^2, \text{ s.t. } w^T \mathbf{1} = 1, 0 \leq w_k. \quad (16)$$

In the same way as in [24], the solution to problem (16) can be obtained. Algorithm 1 presents the steps of DRCG. Then, based on the outputs of DRCG, i.e., the reconstructed data  $\bar{X} = PY^T S$  and the projection  $P$ , two anomaly score functions based on the reconstructed data and the projection are designed to discriminate the anomalies, respectively.

#### D. Anomaly Score Functions

According to problem (8), it is clear that the reconstruction error of each sample measures the reconstructing quality. Assumption that anomalies are not part of the normal data distribution. Therefore, abnormal samples are more difficult to be represented by other normal samples. As a result, anomalies are likely to exhibit higher reconstruction errors. In other words, the greater the error associated with an observation, the greater the chance that it will be an anomaly. Thus, the reconstruction error of the sample can be used as a distinguishing indicator. When the reconstructed sample  $\bar{X}$  is obtained by the DRCG, the reconstruction error-based anomaly score function  $O_i^I$  of the sample  $x_i$  can be calculated as

$$O_i^I = \|x_i - \bar{x}_i\|_2^2 = \|x_i - PY^T s_i\|_2^2. \quad (17)$$

Since the anomaly score function (17) cannot handle out-of-sample data, it cannot be used for streaming data. To solve this problem, the projection-based anomaly score function is designed. One should note that anomalous points are a small percentage of all the data, and thus the information learned by subspace and projection is more biased towards normal data points. Consequently, anomalous data are expected to deviate from normal data in the subspace. Motivated by PCA, the projection-based anomaly score function  $O_i^{II}$  of the sample  $x_i$  can be modeled as

$$O_i^{II} = \|(I - PP^T)(x_i - \mu)\|_2^2 \quad (18)$$

where  $\mu$  is the global mean of the reconstruction data  $\bar{X}$ .

#### IV. ANOMALY DETECTION BASED ON DRCG

The conceptual of anomaly detection is presented in the Fig. 1. Based on the designed anomaly score functions, anomaly detection methods for the following two scenarios, i.e., detecting anomalies in the input data directly and detecting anomalies online, are described.

##### A. Anomaly Detection With DRCG

As shown in Fig. 1(a)–(c), the method for detecting anomalies in the input data directly has the following three steps. For step 1, the data collected from IIoT is used as an input to Algorithm 1 to train the model (8). For step 2, the iterative algorithm is executed until convergence, and then this algorithm outputs the projection, the low-dimensional embedding, and the consensus graph matrices. Further, its output is utilized to compute  $\bar{X} = PY^T S$  and  $\mu$ . For step 3, performing the procedure (17) or (18) until the  $O_i^I$  and  $O_i^{II}$  are calculated for all samples. Then, ranking the scores,  $\{O_i^I \text{ or } O_i^{II}, i = 1, \dots, n\}$ , in descending order. We mark the samples associated with high abnormal scores as abnormal. The detailed steps of the method refer to Algorithm 1.

##### B. Online Anomaly Detection With DRCG

To achieve online anomaly detection for streaming data, the anomaly score function (18) is used. The detailed steps of the method are presented in Algorithm 2 and Fig. 1(a), (b), and (d). Inspired by [18], the following two phases are used to implement online anomaly detection. In the offline process, a portion of data from an IIoT system is first collected as a training set, then, a subset from the training set is selected for training the model (8) to obtain the projection  $P$  and  $\mu$ , finally,  $P$ ,  $\mu$ , and anomaly score function (18) are used to compute the baseline statistics. In the online detection phase, the anomaly score function (18),  $P$ , and  $\mu$  are also used to compute a summary statistic for each new input sample, and then, the summary statistic and the baseline statistic are used to compute an anomaly score to determine anomalies. Here,  $\theta$  is an adjustable parameter that controls the size of the output anomaly score.  $T$  also is an adjustable parameter, which controls the sensitivity of the system alarm. Since  $T$  can be adjusted according to the anomaly score, the setting of these two parameters can make the system more flexible. Unlike

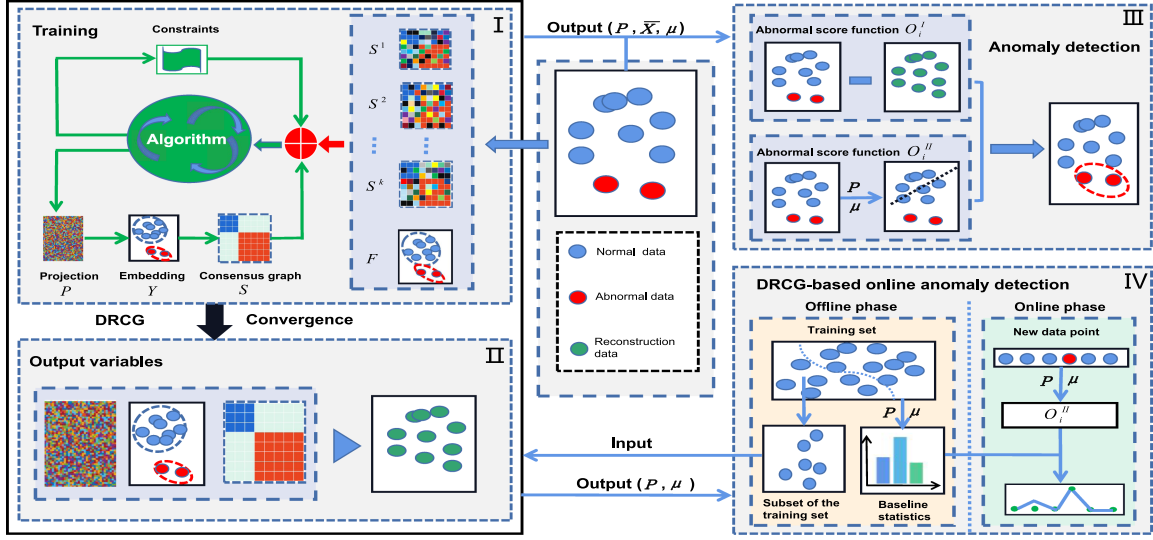


Fig. 1. Overall concept of the proposed methods. (a) Training the model (8) by using IIoT data. (b) When the iterative process of iterative algorithm reaches convergence, the algorithm will output the projection, low-dimensional embedding, and consensus graph matrices. Then, these three matrices are used to reconstruct the data. (c) Recognizing anomalies in the input data. (d) DRCG is used for online anomaly detection.

---

**Algorithm 1: Anomaly Detection With DRCG.**


---

**Input:** The training data  $X$ , the parameters  $\alpha, \beta, h$ , dimensionality of the reduced dimensional space  $d$ , and the number of abnormal samples candidates  $a$ ;  
**Initialization:**  $P$  by performing PCA on  $X$ ,  $F = P^T X$ ,  $S$  by using the problem (5), and  $\{S^k\}_{k=1}^v$  by using problem (6) with different the projection functions.  
**Output:** The  $a$  desired abnormal samples in  $X$ .  
**while** not converged **do**  
 1: Update  $Y$  by computing (9);  
 2: Update  $P$  by computing (12);  
 3: Update  $S$  by computing (13);  
 4: Update  $w$  by computing (16);  
**end while**  
**DRCG-I:**  
 Calculating  $O_i^I$  for  $x_i$  according to (17);  
 Returning top  $a$  observations as desired outliers;  
**DRCG-II:**  
 Calculating  $O_i^{II}$  for  $x_i$  according to (18);  
 Returning top  $a$  observations as desired outliers;

---

PCA-OAD, the projection of DRCG can preserve both local and global relationships between the data. Meanwhile, the  $\mu$  of the Algorithm 2 is obtained via the use of the reconstruction data, which better represents the intrinsic relationship between the data.

### C. Complexity Analysis

To better analyze the Algorithm 2, the time and space complexity of the method is presented. For the time complexity of Algorithm 2, the main computational costs of the offline phase come from computing the variables  $Y, P, S$  in subset  $S_u$ . Specifically, by using the Algorithm in literature [21]

---

**Algorithm 2: DRCG-Based Online Anomaly Detection.**


---

**Input:** The training set  $X_{train}$  with  $N$  data point, the testing set  $X_{test}$ , the parameters  $\alpha, \beta, h, \theta$ , dimensionality of the reduced dimensional space  $d$ , the threshold  $T$ ;  
**Output:** Anomaly score of data point  $x_j$ .  
**Offline phase**  
 1: Choose subset  $S_u$  of  $X_{train}$  with sizes  $N_1$ ;  
 2: Input  $S_u$  into Algorithm 1 to obtain  $P$  and  $\mu$ ;  
 3: Compute baseline statistics  $\{O_i^{II}\}_{i=1}^N$  for any  $x_i \in X_{train}$  by (18);  
**Online anomaly detection**  
 1: **while** new data point  $x_j \in X_{test}$  is received **do**;  
 2: Compute the summary statistic  $O_j^{II}$  for  $x_j$  with (18);  
 3: Compute  $p_j = \frac{1}{N} \sum_{x_j \in X_{train}} cf\{O_i^{II} > O_j^{II}\}$ , where  $cf\{\}$  denotes the counting function;  
 4: Compute anomaly score  $AS_j = \max\{0, \log(\frac{\theta}{p_j})\}$ ;  
 5: If  $AS_j > T$  then mark  $x_j$  as an abnormal sample.

---

to compute  $Y$ , the computational cost is  $O(N_1 d^2)$ . It takes  $O(d^3)$  to compute the matrix  $P$  to perform the singular value decomposition of the matrix  $Q \in \mathbb{R}^{m \times d}$ . However, the main computational cost of computing  $S$  is  $O(N_1^3)$  with the inverse of the matrix  $(\beta \sum_{k=1}^v (w_k)^2 I + YY^T) \in \mathbb{R}^{N_1 \times N_1}$ . Thus, the main time complexity of the offline phase is  $O(\tau(N_1 d^2 + d^3 + N_1^3))$ , with  $\tau$  is the number of iterations. For the online phase, the main computational costs of it come from computing  $O_j^{II}$ . For a new data point  $x_j \in \mathbb{R}^{m \times 1}$ , the main time complexity of the offline phase is  $O(m^2)$ . For the space complexity of Algorithm 2, the offline phase requires the storage of  $Y, P, S$  for iterative computing and  $(I - PP^T), \{O_i^{II}\}_{i=1}^N$  for the online phase. Thus, the main space complexity of the offline phase is  $O(m^2 + dN_1 + N_1^2 + N)$ . The online phase requires the

**TABLE I**  
ATTRIBUTES OF THE USED DATASETS

Data set	Sample size	Dimension	Anomaly class
HAI 21.03	12628	79	attack (0.05%)
N-BaIoT	43423	115	attack (0.69%)
UNSWNB	17910	42	attack (0.25%)
CRCAF	14492	29	attack (0.04%)

storage of  $(I - PP^T)(x_i - \mu)$  to obtain  $Q_j^{II}$ . Thus, the main space complexity of the online phase is  $O(m)$ . From the above analysis, it is clear that the complexity of the offline phase is higher. For large-scale datasets, one can solve the problem with higher complexity by selecting a subset of small size for training or choosing devices with better computing power.

## V. EXPERIMENTS

The feasibility of the proposed anomaly detection methods is verified by the following two experiments. For the first experiment, we use the dataset with a few anomaly samples to verify the anomaly detection capability of Algorithm 1. In this experiment, commonly used assessment metrics [26], such as Accuracy, F1-score, and Kappa, are used to assess the accuracy of the results of abnormality detection. High values of all the indicators indicate better detection performance. For the second experiment, streaming data is used to verify the advantages of the DRCG-based online anomaly detection (DRCG-OAD). In this experiment, receiver operating characteristic (ROC), and area under curve (AUC) [27] are used to evaluate the performance of DRCG-OAD. Before performing the experiments, the mapping function  $\phi$  should be built, which consists of five candidates (i.e.,  $v = 5$ ): three gaussian function  $Ker(x_i, x_j) = \exp(-\|x_i - x_j\|_2^2 / 2\theta^2)$ , where  $\theta$  varies from  $\{0.1, 1, 10\}$ ; and two polynomial function  $Ker(x_i, x_j) = (x_i^T x_j + u)^2$ , where  $u$  varies from  $\{0, 1\}$ . The experiments are performed on a Win10 laptop with an Intel Core i5-8250 U 1.6 GHz CPU. The experimental tool used is the 2019a version of MATLAB.

### A. Datasets

Four real-world datasets were used for comparative experiments. Table I presents some information about the datasets. The details are described below.

1) **HAI 21.03**: The HAI 21.03 [28] dataset was obtained from an industrial control system (ICS) testbed of a realistic power generation system, i.e., the hardware-in-the-loop-based augmented ICS security (HAI) testbed. This testbed has 59 sensors. The dataset was formed by collecting all the sensor measurements every second for four consecutive days. During this period, 28 attacks were performed to inject into the system. The following four points in the system, namely, setpoints, process variables, control outputs, and control parameters, are affected by the attacks. In this article, a subset of HAI 21.03 was used as experiments.

2) **N-BaIoT**: The network-based IoT botnet attack detection (N-BaIoT) [29] dataset was gathered from an IoT network under normal and attack conditions. The IoT network consists of the

following nine devices connected via Wi-Fi, i.e., a baby monitor, a thermostat, two doorbells, a webcam, and four security cameras. Each sample records the time interval, packet size, and count of packets arriving in real network traffic for each IoT device. The data were obtained for each device under normal operating conditions and under several different attacks executed by the BASHLITE botnet [29]. In this article, the attributes of the attack conditions recorded in the data are the traffic statistics generated by the BASHLITE botnet's spam attacks on the IoT network.

3) **UNSWNB**: The UNSW-NB15 (UNSWNB) dataset,<sup>1</sup> created by the Australian Cyber of Security Centre, has real modern normal and abnormal network traffic records, containing a total of 2.5 million records, 13% of which are abnormal. Each record in this dataset has 42 attributes. In this article, a subset of this data is selected for experimentation.

4) **CRCAF**: The credit card fraud (CRCAF) dataset<sup>2</sup> was collected from credit card fraud incidents. This dataset records transactions made by European cardholders with their credit cards over a two-day period in September 2013. The anomaly category of this dataset represents 0.17% of all the samples, where each sample has 29 attributes. As shown in Table I, a subset of this dataset was also selected for the experiments.

### B. Anomaly Detection Capability of Algorithm 1

1) **Experimental Setup**: In this part, the corresponding subsets of all the datasets were used for this experiment, respectively. Specifically, the subset of HAI 21.03 is formed by randomly selecting 3000 normal samples and 150 abnormal samples from the HAI 21.03 dataset. The same operation was performed for the other three datasets. Before training, we normalize each row of the data matrix to have a unit norm. Then, seven anomaly detection methods, such as OSPCA [12], LDOF [7], IForest [10], FABOD [11], LOPD [5], RGAE [16], and ADKE [15], were used for comparison. Notice that LDOF is a distance-based method. Subspace learning methods (e.g., OSPCA, LOPD), and the methods related to graph learning (e.g., ADKE, RGAE) are commonly used to detect anomalies in HD data. Before doing the experiment, the parameters of all the anomaly detection methods should be set. For the proposed DRCG-I and DRCG-II, the two parameters (i.e.,  $\alpha, \beta$ ) in the DRCG need to be adjusted. It is an open problem to choose the best parameters for different datasets. In this study, we used a grid search method to find the best combination of the two parameters from the candidate set  $\{10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 1, 10^1, 10^2\}$ . Fig. 2 shows the classification performance versus different combinations of the parameters  $\alpha$  and  $\beta$ . From Fig. 2, we can easily adjust the parameters  $\alpha$  and  $\beta$ . The parameter  $d$  is set to the dimension corresponding to the best detection performance. For convenience,  $h$  is uniformly set to 8. The parameters of the other compared methods were set to those recommended in their corresponding

<sup>1</sup>[Online]. Available: <https://research.unsw.edu.au/projects/unsw-nb15-dataset>

<sup>2</sup>[Online]. Available: <https://www.kaggle.com/datasets/mlg-ulb/creditcard-fraud/discussion>



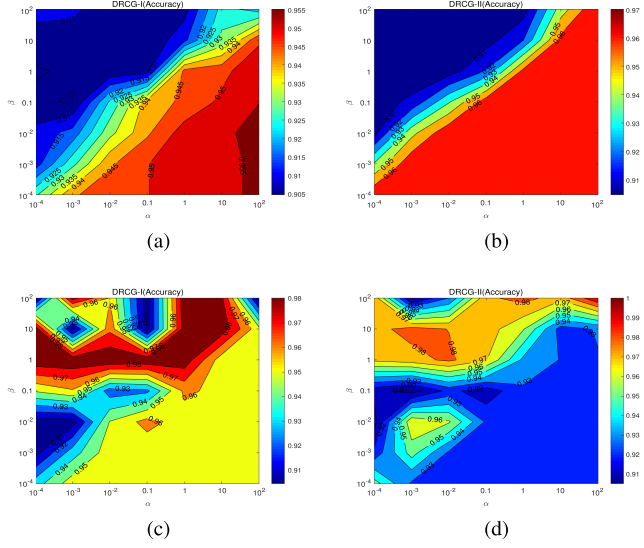


Fig. 2. Classification performance versus different combinations of the parameters  $\alpha, \beta$  on the (a), (b) HAI 21.03 and (c), (d) N-BaloT datasets.

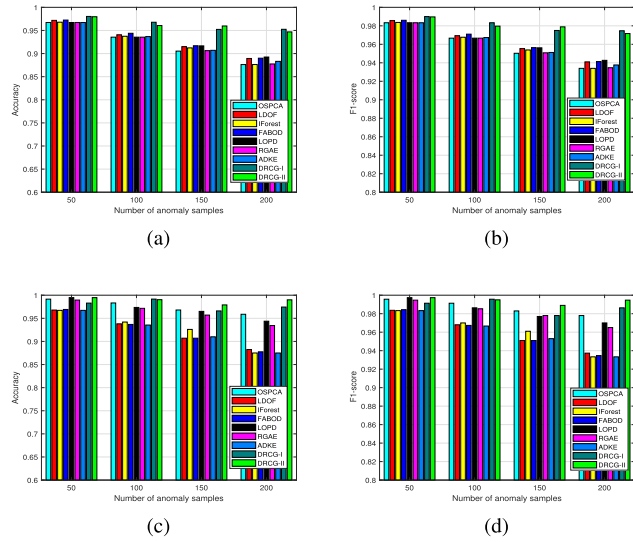


Fig. 3. Classification performance versus different number of anomaly samples on the (a), (b) HAI 21.03 and (c), (d) N-BaloT datasets.

papers. By using ten randomly divided subsets, all experiments were performed ten times, and then the average classification results and the corresponding standard deviations were reported in Table II. To explore whether the subset contains different numbers of anomaly samples would have an impact on all the anomaly detection methods, four randomly divided subsets, each containing 50, 100, 150, and 200 anomalous samples, were used for the experiments. Then, the experimental results are presented in Fig. 3.

**2) Experimental Results and Analysis:** From Table II, it is clearly evident that the proposed DRCG-II method obtained attractive classification results on all the datasets. In particular, in the HAI 21.03 dataset, the values of DRCG-I and DRCG-II in terms of Kappa are much higher than the competing methods.

TABLE II  
CLASSIFICATION RESULTS (MEAN  $\pm$  STD) ON ALL THE DATASETS

Data set	Methods	Accuracy	F1-score	Kappa
HAI 21.03	OSPCA	0.905 $\pm$ 0.001	0.950 $\pm$ 0.001	0.042 $\pm$ 0.005
	LDOF	0.915 $\pm$ 0.002	0.955 $\pm$ 0.001	0.063 $\pm$ 0.021
	IForest	0.912 $\pm$ 0.006	0.954 $\pm$ 0.003	0.056 $\pm$ 0.048
	FABOD	0.917 $\pm$ 0.002	0.956 $\pm$ 0.001	0.084 $\pm$ 0.025
	LOPD	0.917 $\pm$ 0.002	0.956 $\pm$ 0.001	0.080 $\pm$ 0.017
	RGAE	0.907 $\pm$ 0.002	0.951 $\pm$ 0.001	0.031 $\pm$ 0.011
	ADKE	0.906 $\pm$ 0.001	0.951 $\pm$ 0.001	0.026 $\pm$ 0.017
	DRCG-I	0.952 $\pm$ 0.006	0.975 $\pm$ 0.003	0.474 $\pm$ 0.059
	DRCG-II	<b>0.960<math>\pm</math>0.006</b>	<b>0.979<math>\pm</math>0.003</b>	<b>0.556<math>\pm</math>0.068</b>
N-BaloT	OSPCA	0.968 $\pm$ 0.019	0.983 $\pm$ 0.011	0.722 $\pm$ 0.047
	LDOF	0.907 $\pm$ 0.001	0.951 $\pm$ 0.001	0.023 $\pm$ 0.008
	IForest	0.926 $\pm$ 0.032	0.961 $\pm$ 0.017	0.343 $\pm$ 0.300
	FABOD	0.907 $\pm$ 0.001	0.951 $\pm$ 0.001	0.022 $\pm$ 0.010
	LOPD	0.965 $\pm$ 0.003	0.977 $\pm$ 0.009	0.665 $\pm$ 0.054
	RGAE	0.957 $\pm$ 0.002	0.978 $\pm$ 0.001	0.530 $\pm$ 0.017
	ADKE	0.910 $\pm$ 0.015	0.953 $\pm$ 0.008	0.091 $\pm$ 0.126
	DRCG-I	0.966 $\pm$ 0.038	0.978 $\pm$ 0.021	0.678 $\pm$ 0.150
	DRCG-II	<b>0.979<math>\pm</math>0.009</b>	<b>0.989<math>\pm</math>0.005</b>	<b>0.812<math>\pm</math>0.096</b>
UNSWNB	OSPCA	0.915 $\pm$ 0.004	0.956 $\pm$ 0.002	0.067 $\pm$ 0.046
	LDOF	0.919 $\pm$ 0.002	0.958 $\pm$ 0.001	0.110 $\pm$ 0.024
	IForest	0.921 $\pm$ 0.014	0.959 $\pm$ 0.008	0.153 $\pm$ 0.098
	FABOD	0.924 $\pm$ 0.003	0.960 $\pm$ 0.002	0.162 $\pm$ 0.035
	LOPD	0.924 $\pm$ 0.004	0.960 $\pm$ 0.003	0.166 $\pm$ 0.048
	RGAE	0.927 $\pm$ 0.003	0.962 $\pm$ 0.002	0.192 $\pm$ 0.035
	ADKE	0.915 $\pm$ 0.009	0.956 $\pm$ 0.005	0.093 $\pm$ 0.077
	DRCG-I	0.931 $\pm$ 0.003	<b>0.965<math>\pm</math>0.001</b>	0.271 $\pm$ 0.028
	DRCG-II	<b>0.934<math>\pm</math>0.003</b>	0.964 $\pm$ 0.002	<b>0.299<math>\pm</math>0.017</b>
CRCAF	OSPCA	0.905 $\pm$ 0.001	0.950 $\pm$ 0.001	0.041 $\pm$ 0.005
	LDOF	0.911 $\pm$ 0.002	0.953 $\pm$ 0.001	0.019 $\pm$ 0.020
	IForest	0.912 $\pm$ 0.005	0.954 $\pm$ 0.002	0.046 $\pm$ 0.039
	FABOD	0.917 $\pm$ 0.001	0.951 $\pm$ 0.001	0.022 $\pm$ 0.009
	LOPD	0.923 $\pm$ 0.0031	0.957 $\pm$ 0.001	0.024 $\pm$ 0.011
	RGAE	0.942 $\pm$ 0.003	<b>0.969<math>\pm</math>0.002</b>	0.356 $\pm$ 0.037
	ADKE	0.910 $\pm$ 0.008	0.953 $\pm$ 0.004	0.091 $\pm$ 0.074
	DRCG-I	0.940 $\pm$ 0.011	0.962 $\pm$ 0.006	0.350 $\pm$ 0.077
	DRCG-II	<b>0.944<math>\pm</math>0.009</b>	0.968 $\pm$ 0.005	<b>0.359<math>\pm</math>0.094</b>

The bold entities implies the best classification results obtained by the anomaly detection methods under the different datasets.

Although OSPA, IForest, LOPD, and RGAE achieved high accuracy and F1-score on some of the datasets (e.g., OSPA obtained high accuracy on the N-BaloT dataset and RGAE obtained high F1-score on the CRCAF dataset), their advantages on indicator Kappa became less obvious on the HAI 21.03 dataset. One should note that Kappa is used to check whether the model prediction results and the actual classification results are consistent. That is, although most abnormalities can be detected by OSPA, IForest, LOPD, and RGAE, these methods have the risk of recognizing some normal samples as abnormal or some abnormal samples as normal. LDOF did not obtain better classification results because the distance-based measures suffer from the distance concentration problem fail to capture the neighborhood information in the HD space. As shown in Fig. 3, as the number of anomaly samples contained in the training set increases, the classification performance of all the anomaly detection methods decreases. The reasons for this phenomenon are twofold. 1) These methods failed to learn the intrinsic relationship between normal samples in the dataset, and thus these methods mistake more abnormal samples for normal ones. 2) These methods failed to solve the distance concentration problem well, which increases the risk that these methods mistake anomalous features for normal ones. However, the degradation of the classification performance of DRCG-I and DRCG-II is slower compared to the other comparative



TABLE III  
INITIAL VALUES OF THE SIMULATION PARAMETERS

Data set	$\alpha$	$\beta$	$d$	$h$	$\theta$	T
HAI 21.03	0.01	10	25	8	0.2	0
N-BaIoT	0.0001	10	29	8	0.2	0
UNSWNB	10	1	6	8	0.2	0
CRCAF	10	10	19	8	0.2	0

methods. In summary, the reason why competing methods do not always produce better results in all the datasets is that they suffer from distance concentration problem or fail to explore more of the intrinsic structure between the data. However, the better abnormality detection performances obtained by the proposed methods are attributed to the following two advantages: 1) DRCG can effectively attenuate the negative impact of the distance concentration problem on anomaly detection; and 2) DRCG effectively exploits the local and global intrinsic structure among the data.

### C. Anomaly Detection Capability of Algorithm 2

1) *Experimental Setup*: The proposed DRCG-OAD is used to detect anomalies in streaming data. Before the experiment, we normalize each column of the data points to have a unit norm. The following anomaly detection methods, including OOSPCA [13], LOPD [5], and PCA-OAD [18], are used for comparison. Furthermore, sparse PCA (SPCA) [25] is a dimensionality reduction method related to DRCG has also been used to compare. Notice that SPCA and LPOD do not implement online anomaly detection for the streaming data. By adopting a similar approach to DRCG-OAD, SPCA-OAD, and LOPD-OAD are yielded. Furthermore, the sliding window technique needs to be used by LOPD-OAD to calculate baseline statistics and summary statistics. The detailed configuration of the parameters of the method is placed in Table III. One should note that the parameters ( $\alpha$ ,  $\beta$ ,  $d$ , and  $h$ ) in this experiment were set as in the first experiment. The parameters  $\theta$  and  $T$  can be flexibly adjusted according to the abnormal score of the online output. For convenience, we set  $\theta$  and  $T$  uniformly as 0.2 and 0, respectively. For the HAI 21.03 dataset, we randomly selected 8000 normal samples as the training set for offline training and the rest of the samples were used for online testing. From these 8000 samples, we randomly selected 6000 data points for training the corresponding models. Then, the entire training set was used to constitute the summary statistics. For the remaining three datasets, the same way was used. Fig. 4 shows the ROC curves and AUC of all the methods. To visually compare the effectiveness of all the methods, we randomly select 100 samples from the N-BaIoT dataset, including 85 normal samples and 15 abnormal samples, to simulate a continuous time series. Fig. 5 reports the anomaly scores for all methods that were adjusted to 0 and 1 intervals. The red and black dots indicate the true labels of abnormal and normal samples in the dataset, respectively. The blue line indicates the score of the abnormal point.

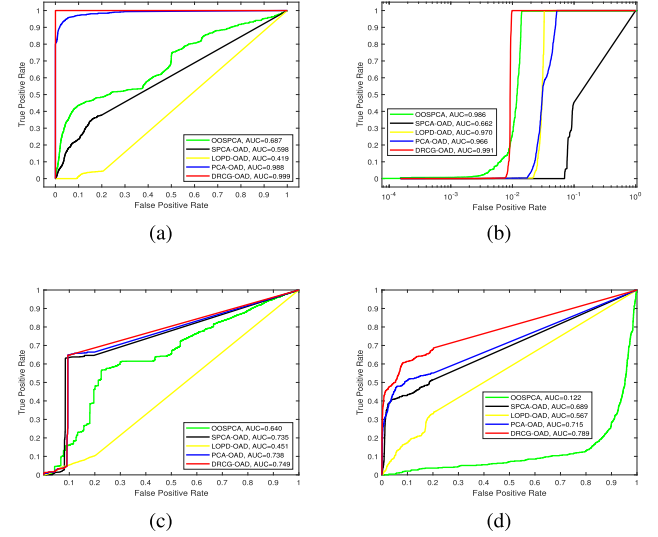


Fig. 4. ROC and AUC analysis for all the methods in the (a) HAI 21.03, (b) N-BaIoT, (c) UNSWNB, and (d) CRCAF datasets.

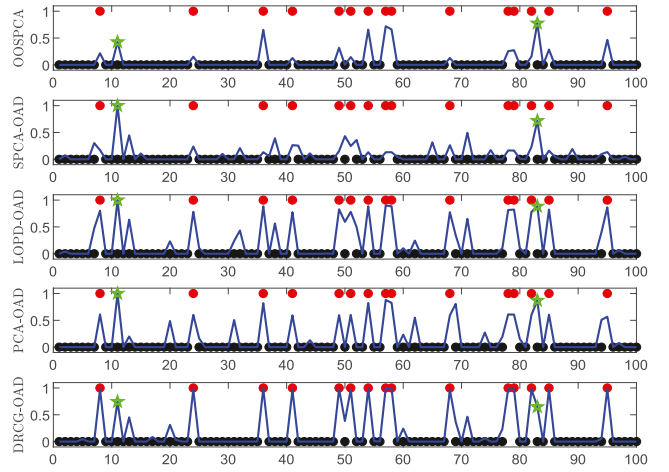


Fig. 5. Capability of anomaly detection on the N-BaIoT dataset.

2) *Experimental Results and Analysis*: As shown in Fig. 4, we can see that the performances of DRCG-OAD and PCA-OAD outperform the other comparison methods on the HAI 21.03, UNSWNB, and CRCAF datasets. In comparison with PCA-OAD, which focuses only on exploring the global relationships between data, the proposed DRCG-OAD obtains excellent results from the fact that DRCG focuses on the exploration of both local and global relationships in the data. Poor performance of SPCA-OAD means that not all data reconstruction-based methods can be used for anomaly detection. From Fig. 5, we can observe that all the methods, except SPCA-OAD, can identify a large number of anomalies. However, LOPD-OAD and PCA-OAD have the risk of mistaking normal points for abnormal points (e.g., the green star points in Fig. 5), which increases the false alarm rate of real-time abnormality check measurements. Although OOSPCA obtains relatively good AUC values in the N-BaIoT dataset, it is unsuitable for working with

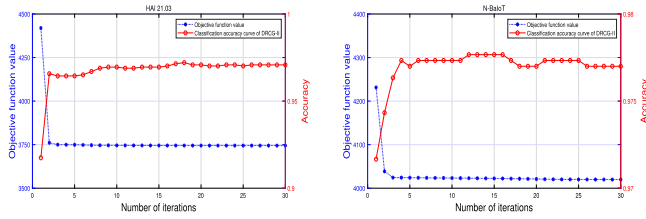


Fig. 6. Convergence curve of DRCG-II on the HAI 21.03 and N-BaIoT datasets.

the CRCAF dataset. In the above experiment results, it is clearly that DRCG-OAD obtained satisfactory results in all the datasets. The reason for this phenomenon is that DRCG learned the projection that can effectively extract features describing the relationships within the data. Also, The above results show that the proposed DRCG-OAD can be used to detect anomalies in IIoT.

#### D. Convergence Analysis

In this part, we analyze the convergence of the proposed iterative algorithm for solving problem (9). Fig. 6 shows the curves of DRCG-II on the HAI 21.03 and N-BaIoT datasets. Specifically, the blue and red curves show the trend of the objective function value of problem (9) and classification accuracy with the number of iterations, respectively. It can be clearly observed that the objective function value decreases rapidly with an increasing number of iterations and then reaches stability within 15 iterations. Obviously, the classification accuracy value of DRCG-II changes relatively steadily when the convergence curve is stable. This implies that the iterative algorithm can converge efficiently.

### VI. CONCLUSION

In this article, we proposed a framework named DRCG, which integrates projection learning, low-dimensional embedding, and consensus graph learning into a unified objective function. Then, an iterative algorithm was designed to solve the DRCG model. By doing so, DRCG can drive the reconstruction error of abnormal samples higher than that of normal samples. Furthermore, two anomaly score functions based on the reconstruction error and projection were proposed. Furthermore, to achieve online anomaly detection for streaming data, DRCG with the projection-based anomaly score function was extended into an online version. The experimental results demonstrated the effectiveness of the proposed methods and their usability in detecting data anomalies in IIoT. One should note that the computational complexity of DRCG is  $O(\tau(N_1 d^2 + d^3 + N_1^3))$ , which means it takes more time for training. Therefore, we will develop faster reconstruction-based anomaly detection methods in the future.

### REFERENCES

- [1] A. Corradi et al., "Smart appliances and RAMI 4.0: Management and servitization of ice cream machines," *IEEE Trans. Ind. Inform.*, vol. 15, no. 2, pp. 1007–1016, Feb. 2019.
- [2] X. Zhou, W. Liang, S. Shimizu, J. Ma, and Q. Jin, "Siamese neural network based few-shot learning for anomaly detection in industrial cyber-physical systems," *IEEE Trans. Ind. Inform.*, vol. 17, no. 8, pp. 5790–5798, Aug. 2021.
- [3] Y. Zhang, Z. Dong, W. Kong, and K. Meng, "A composite anomaly detection system for data-driven power plant condition monitoring," *IEEE Trans. Ind. Inform.*, vol. 16, no. 7, pp. 4390–4402, Jul. 2020.
- [4] X. Hu, Y. Li, L. Jia, and M. Qiu, "A novel two-stage unsupervised fault recognition framework combining feature extraction and fuzzy clustering for collaborative AIoT," *IEEE Trans. Ind. Inform.*, vol. 18, no. 2, pp. 1291–1300, Feb. 2022.
- [5] H. Liu, X. Li, J. Li, and S. Zhang, "Efficient outlier detection for high-dimensional data," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 48, no. 12, pp. 2451–2461, Dec. 2018.
- [6] M. Chehreghani, "K-nearest neighbor search and outlier detection via min-max distances," in *Proc. SIAM Int. Conf. Data Mining*, 2016, pp. 405–413.
- [7] K. Zhang, M. Hutter, and H. Jin, "A new local distance-based outlier detection approach for scattered real-world data," in *Proc. 13th Pac.-Asia Conf. Knowl. Discov. Data Mining*, 2009, pp. 813–822.
- [8] J. Zheng, H. Qu, Z. Li, L. Li, and X. Tang, "An irrelevant attributes resistant approach to anomaly detection in high-dimensional space using a deep hypersphere structure," *Appl. Soft Comput.*, vol. 116, 2022, Art. no. 108301.
- [9] M. Breunig, H. Kriegel, R. Ng, and J. Sander, "LOF: Identifying density-based local outliers," in *Proc. ACM SIGMOD Int. Conf. Manag. Data*, 2000, pp. 93–104.
- [10] F. T. Liu, K. M. Ting, and Z. H. Zhou, "Isolation forest," in *Proc. IEEE Int. Conf. Data Mining*, Pisa, Italy, 2008, pp. 413–422.
- [11] H. Kriegel, M. Schubert, and A. Zimek, "Angle-based outlier detection in high-dimensional data," in *Proc. 14th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, Las Vegas, NV, USA, 2008, pp. 444–452.
- [12] Y. Yeh, Z. Lee, and Y. Lee, "Anomaly detection via oversampling principal component analysis," in *Proc. 1st KES Int. Symp. Intell. Decis. Technol.*, 2009, pp. 449–458.
- [13] Y. Lee, Y. Yeh, and Y. Wang, "Anomaly detection via online oversampling principal component analysis," *IEEE Trans. Knowl. Data Eng.*, vol. 25, no. 7, pp. 1460–1470, Jul. 2013.
- [14] L. Li, H. Qu, Z. Li, J. Zheng, and F. Guo, "Discriminative projection learning with adaptive reversed graph embedding for supervised and semi-supervised dimensionality reduction," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 12, pp. 8688–8702, Dec. 2022.
- [15] H. Liu, E. Li, X. Liu, K. Su, and S. Zhang, "Anomaly detection with kernel preserving embedding," *ACM Trans. Knowl. Discov. Data*, vol. 15, no. 5, May 2021, Art. no. 91.
- [16] G. Fan, Y. Ma, X. Mei, F. Fan, J. Huang, and J. Ma, "Hyperspectral anomaly detection with robust graph autoencoders," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–14, 2022.
- [17] L. Zhang, J. Zhao, and W. Li, "Online and unsupervised anomaly detection for streaming data using an array of sliding windows and PDDs," *IEEE Trans. Cybern.*, vol. 51, no. 4, pp. 2284–2289, Apr. 2021.
- [18] M. Kurt, Y. Yilmaz, and X. Wang, "Real-time nonparametric anomaly detection in high-dimensional settings," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 7, pp. 2463–2479, Jul. 2021.
- [19] F. Nie, Z. Wang, R. Wang, and X. Li, "Adaptive local embedding learning for semi-supervised dimensionality reduction," *IEEE Trans. Knowl. Data Eng.*, vol. 34, no. 10, pp. 4609–4621, Oct. 2022.
- [20] Z. Ren, M. Mukherjee, J. Lloret, and P. Venu, "Multiple kernel driven clustering with locally consistent and selfish graph in industrial IoT," *IEEE Trans. Ind. Inform.*, vol. 17, no. 4, pp. 2956–2963, Apr. 2021.
- [21] Z. Wen and W. Yin, "A feasible method for optimization with orthogonality constraints," *Math. Prog.*, vol. 142, no. 1, pp. 397–434, Dec. 2013.
- [22] H. Zou, T. Hastie, and R. Tibshirani, "Sparse principal component analysis," *J. Comput. Graph. Statist.*, vol. 15, no. 2, pp. 265–286, 2006.
- [23] F. Nie, X. Wang, M. I. Jordan, and H. Huang, "The constrained Laplacian rank algorithm for graph-based clustering," in *Proc. 30th AAAI Conf. Artif. Intell.*, 2016, pp. 1969–1976.
- [24] X. Li, M. Chen, F. Nie, and Q. Wang, "Locality adaptive discriminant analysis," in *Proc. 26th Int. Joint Conf. Artif. Intell.*, 2017, pp. 2201–2207.
- [25] A. Seghouane, N. Shokouhi, and I. Koch, "Sparse principal component analysis with preserved sparsity pattern," *IEEE Trans. Image Process.*, vol. 28, no. 7, pp. 3274–3285, Jul. 2019.
- [26] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks," *Inform. Process. Manag.*, vol. 45, no. 4, pp. 427–437, Jul. 2009.

- [27] H. Lu, T. Wang, X. Xu, and T. Wang, "Cognitive memory-guided autoencoder for effective intrusion detection in Internet of Things," *IEEE Trans. Ind. Inform.*, vol. 18, no. 5, pp. 3358–3366, May 2022.
- [28] H. Shin, W. Lee, J. Yun, and B. Min, "Two ICS security datasets and anomaly detection contest on the HIL-based augmented ICS testbed," in *Proc. Cyber Secur. Experimentation Test Workshop*, 2021, pp. 36–40.
- [29] Y. Meidan et al., "N-BaIoT: Network-based detection of IoT botnet attacks using deep autoencoders," *IEEE Pervas. Comput.*, vol. 17, no. 3, pp. 12–22, Third Quarter 2018.



learning.

**Lin Li** received the B.S. degree in mathematics and applied mathematics and M.S. degree in electronic science and technology from Chongqing Three Gorges University, Chongqing, China, in 2015 and 2019, respectively, and the Ph.D. degree in computer science and technology from Chongqing University of Posts and Telecommunications, Chongqing, China, in 2023.

His research interests include dimensional-ity reduction, anomaly detection, and machine



the Institute of Ecological Safety, Forecasting and Control. His research interests include computer simulation and modeling, machine learning, and data science.

**Hongchun Qu** received the Ph.D. degree in computer science and technology from Chongqing University, Chongqing, China, and Iowa State University, Ames, IA, USA, in 2009.

During 2010 and 2012, he was a Postdoctoral Researcher with the University of Tübingen. Since 2013, he has been a Professor with the College of Automation and College of Computer Science, Chongqing University of Posts and Telecommunications (CQUPT), Chongqing, China, where he also serves as the Director of



**Zhaoni Li** received the B.S. degree in computer science and technology from Shaanxi University of Science and Technology, Shaanxi, China, in 2006, the M.S. degree in computer system architecture from Shaanxi Normal University, Shaanxi, in 2014. She is pursuing the Ph.D. degree in computer science and technology from Chongqing University of Posts and Telecommunications, Chongqing, China.

Her research interests include methods and applications of data mining and forecasting.



**Jian Zheng** received the M.S. and Ph.D. degrees in computer science and technology from the Chongqing University of Posts and Telecommunications, Chongqing, China, in 2013 and 2023, respectively.

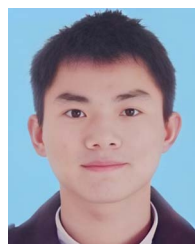
He currently serves as a Lecturer with the College of Artificial Intelligence, Chongqing Technology and Business University, Chongqing, China. His research interests include anomaly detection and machine learning.



**Xiaoming Tang** received the B.S. degree in automation from the College of Information and Electrical Engineering, Panzhihua University, Panzhihua, China, in 2008, and the Ph.D. degree in control theory and control engineering from the College of Automation, Chongqing University, Chongqing, China, in 2013.

From 2016 to 2017, he was a Postdoctoral Researcher with the University of Texas, Arlington, TX, USA. He is currently a Professor with the College of Automation, Chongqing University of Posts and Telecommunications, Chongqing.

His current research interests include model predictive control and networked control systems.



**Ping Liu** received the B.S. degree in automation from North China Electric Power University, Baoding, China, in 2012, and the Ph.D. degree in control science and engineering from Zhejiang University, Hangzhou, China, in 2017.

He is currently an Associate Professor with the College of Automation, Chongqing University of Posts and Telecommunications, Chongqing, China. His research interests include optimal control and dynamic optimization of complex systems and trajectory optimization.