Fault Diagnosis of Industrial Control System With Graph Attention Network on Multi-view Graph

Dongyue Chen
College of Intelligence and Computing
Tianjin University
City, China
dyue_chen@163.com

Ruonan Liu*
College of Intelligence and Computing
Tianjin University
Tianjin, China
ruonan.liu@tju.edu.cn

Wenlong Yu
College of Intelligence and Computing
Tianjin University
Tianjin, China
751909848@qq.com

Kai Zhang
College of Intelligence and Computing
Tianjin University
Tianjin, China
zhangkai_@tju.edu.cn

Yusheng Pu
College of Intelligence and Computing
Tianjin University
Tianjin, China
pys@tju.edu.cn

Di Cao
College of Intelligence and Computing
Tianjin University
Tianjin, China
cd1147583042@126.com

2021 5th Asian Conference on Artificial Intelligence Technology (ACAIT) | 978-1-6654-2630-5/21/$31.00 ©2021 IEEE | DOI: 10.1109/ACAIT53529.2021.9731197

***Abstract*—In this paper, a fault diagnosis framework for the industrial control systems is proposed based on graph attention networks (GAT). The proposed method models the complex relationship between sensor signals and fuses multivariate sensor signals with relational information. First, to reveal the different relationships between sensor signals, a graph with multi-view is constructed considering the similarity and correlation properties. Second, for each view of the graph, GAT is used to extract the graph feature and adjust the structure information by introducing an attention weight to each edge. Finally, the multi-view graph features are concatenated to obtain the final fault feature of multivariate sensor signals. In addition, the focal loss is introduced to balance the contribution of easy classification samples to the loss function. The superiority of the proposed method is demonstrated by extensive experiments on the three-phase flow facility data set.**

***Index Terms*—Fault diagnosis, graph attention network, multi-view graph, industrial control system**

I. Introduction

With the rapid development of Industry 4.0, fault diagnosis for industrial control systems (ICS) is of utmost importance due to the high efficiency and safety production demand. Monitoring equipment, such as sensors, has been used to collect real-time process parameters from different sub-unit. The monitoring data is characterized by multivariate, highly interactive, and data imbalance of the normal and fault samples, which bring a great challenge to fault diagnosis of ICS. However, due to a large amount of record data, fault diagnosis based on data-driven methods has been widely studied in recent years, such as principal component analysis (PCA) [1], partial least squares (PLS) [2], support vector machine (SVM) [3] and deep learning (DL) based methods [4].

Fault diagnosis consist of two main parts: (1) fault detection, which detects an anomaly event; (2) fault isolation, which locates the faulty unit of the ICS system. The methods, such as PCA [1] and PLS [2] aim to reduce the dimension of multivariate sensor signals and use control limits to detect an anomaly occurrence. While other methods aim to learn the discrimination information to recognize the fault categories. In this paper, we follow the deep learning approach to extract fault features under supervised information. However, these methods always neglect the complex relation and interaction among the sensor signals. Owing to the complex control process in an ICS, the process parameters are highly interactive with each other. Therefore, once a fault occurs on one unit of the system, the fault will propagate to other units, which lead to multiple sensor response. In particular, different faults will have different propagation modes and have different effects on the interaction between two process parameters. Thus, these interactions contain additional fault information that can help to extract fault features.

Recently, the researches transforms the data to graph space and use graph neural networks (GNNs) to extract features to improve the performance on multiple tasks, such as few-shot image classification [5], text classification [6] and molecular properties prediction [7]. If a senor is abstracted as a node and the implicit relationship is abstracted as an edge, the sensors can form a graph. But the input of GNNs is the graph with explicit structure, which has fixed edge connection and is set in advance. Therefore, extending ICS fault diagnosis to the graph recognition problem, we should consider the following challenge: (1) graph construction, which can well represent the interaction of the ICS, and (2) feature extraction model, which not only can extract the graph feature but also can adjust the graph structure.

In practice, only partial process parameters can be monitored by the sensors. Thus, the interaction between these sensor signals cannot describe by the control rules. How to represent the implicit relationship between sensor signals is a paramount problem of encoding the interactions into the fault feature. Although common methods, such as the K-nearest neighbor (KNN) graph, can obtain an explicit graph structure, the complex interaction between sensor signals can not be well described. In this paper, different metrics are used to construct the graph structure of the sensor signals. Each metrics can reveal one view of the sensor signals. However, the graph construction based on different metrics will bring edge noise to the graph structure due to the noise in the original sensor signals and this edge noise will burden the feature extraction of GNNs. Therefore, we introduce graph attention neural network (GAT) [8] to extract the fault feature. GAT assigns an importance weight to the edge by using an attention mechanism, in which the learned edge weights are involved in the message passing process of the node feature embedding. The attention mechanism of GAT can reduce the influence of the noise edges by introducing a small weight value to these edges.

Based on the above analysis, an fault diagnosis framework MG-GAT (multi-view graph and graph attention network) is proposed, which considers multiple relationships between sensor signals and constructs the graph in different views (multi-view graph). Then the GAT module is used to extract the fault feature. In particular, we consider the most representative relation between sensor signals, similarity, and correlation. Moreover, data imbalance is one of the common problems of fault diagnosis because of the rarity of the fault event. Therefore, the focal loss function [9] is employed in this fault diagnosis framework, which can balance the training samples by introducing a modulating factor to the cross-entropy loss. The major contributions are summarized as follows:

(1) This paper proposes a fault diagnosis framework MG-GAT based on the graph analysis problem. The multivariate sensor signals are transformed into multi-view graphs, where each sensor signal corresponds to a node and the relationships, similarity and correlation, are corresponding to edges.

(2) The proposed method, using GAT to extract fault features and adjust the edge representation formed by similarity and correlation. In addition, the focal loss function is employed to reduce the influence of sample imbalance on the performance of fault diagnosis.

(3) Extensive experiments are performed on a real-world data set with ground-truth fault categories. The results demonstrate that the proposed method can provide superior fault diagnosis performance.

The remaining parts of this paper are organized as follows. Section II gives the review of related works. Section III addresses the method proposed in this paper. In Section IV the effectiveness of the proposed method is demonstrated on the three-phase flow facility simulation data. Section V gives the conclusion.

II. Related works

A. Deep learning based fault diagnosis

Due to the feature extraction power of deep learning, the performance of the fault diagnosis method has been improved significantly. Jiang et al. [10] propose a multi-scale convolution neural network (CNN) to extract the feature of vibration signal in different scales for gearbox fault diagnosis. Feng et al. [11] introduce zero-shot learning to tackle the sample lankness of certain faults. Yang et al. [12] calculate the correlation matrix and transform the multivariate time series into images and use CNN to learn the fault feature. These works only focus on one view of data properties, either characteristic of sensor signals or relationship between sensor signals. Zhao et al. [13] propose a graph embedded semi-supervised model, named SSGCDBN, for motor bearing fault diagnosis. However, the input signal is single time series, which cannot extend to multivariate time series directly because lacks the relation modeling between sensor signals. In this paper, we tend to fuse the information of data with relational properties in a sensor network.

B. Graph neural network

The graph neural networks (GNNs) aims to extend neural networks for arbitrarily structured graphs, which include but are not limited to graph convolutional networks (GCNs) [14], GraphSAGE [15] and GAT [8]. The core idea of GNNs is to update the node feature via aggregating the features of its neighbors, which can be summarized as a message passing operation. The GNN-based models have successfully applied to many graph-based tasks, such as recommendation system [16], relation prediction for knowledge graphs [17] multi-agent trajectory prediction [18] and graph anomaly detection [19]. Since the superior performance of GNN, a number of applications extend to implicitly structured data. Teney et al [20] improve visual question answer performance by constructing a graph over the question words and scene objects. Hu et al. [21] model the object-object relation via attention mechanism and improve the performance of object detection.

The sensor network of an industrial control system can be seen as a graph with an implicit graph structure. In this paper, we take advantage of rich information in a graph and convert the fault diagnosis problem to a graph classification task.

III. Method

Due to the complex interaction in an ICS, the graph construction considers the most common relationship between different signals: similarity and correlation corresponding to KNN-graph [22] and maximal information-based nonparametric exploration (MINE) graph [23]. Moreover, the structured sensor signals form a heterogeneous graph that has two types of edges. The framework of the proposed method is shown in Fig. 1.

A. Graph construction

The multivaite sensor signals can be denoted as $\mathbf{S} = \{s_i \mid i = 1, \ldots, N\}$. These sensor signals are cut into segments $\mathbf{x}_i = \left(s_i^t, \ldots, s_i^{t-m+1}\right)$ as the input of graph construction, where m is the segment size. Therefore, the graph constructed by sensor signals can be defined as $\mathcal{G} = (\mathcal{V}, \mathcal{E})$. $\mathcal{V} = \{v_i \mid i = 1, \ldots, N\}$ is a set of vertices corresponding to the set of sensors with feature $\mathbf{X} \in \mathbb{R}^{n \times m}$. $\mathcal{E}$ is the set of edges with two types.

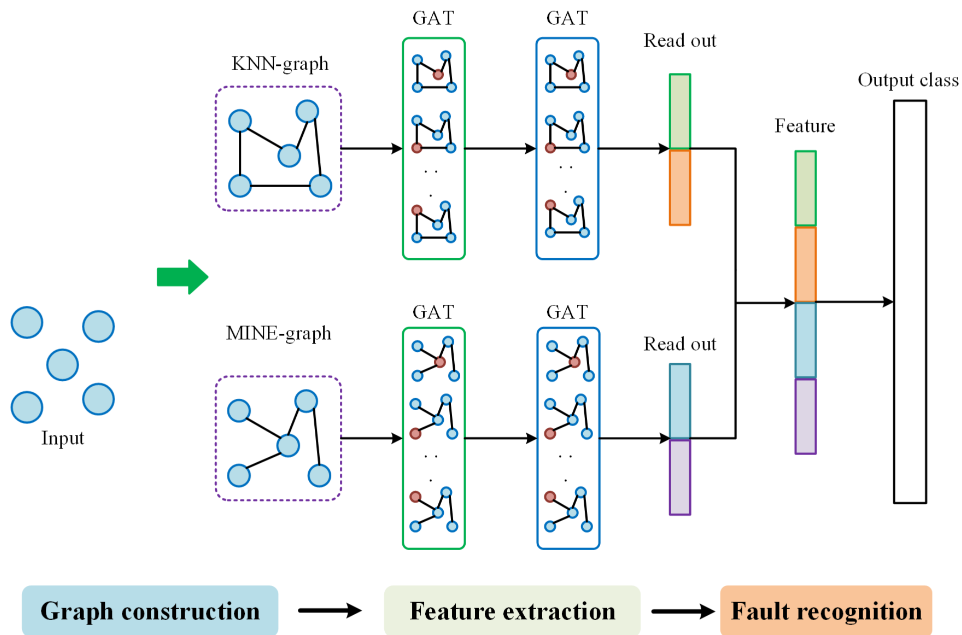

Fig. 1. The framework of the proposed method.

1) KNN-graph construction: The k-nearest neighbor graph is a graph in which each node has k closest neighbors. For node v_i, if v_j is the top k nearest node of v_i, a connection is established between v_i and v_j. The neighbor set of node v_i can be defined as follows:

$$\psi\left(v_i\right)=\left\{v_i^j\right\}_{j=1}^k, \quad \text{if Distance}\left(x_i, x_i^j\right) \text{ is } k\text{-th samllest} \tag{1}$$

where x_i and x_i^j is the node features corresponding to the sensor signal segments and $\text{Distance}(\cdot)$ is the Euclidean distance metric.

2) MINE-graph construction: To obtain the MINE-graph, the maximal information coefficient (MIC) of each node pair is calculated. The MIC belongs to maximal information-based nonparametric exploration (MINE), which is designed to measure the correlation strength between paired variables. The MIC not only catch both linear and nonlinear relationships, while other commonly used correlation metric method, such as Person's correlation coefficient, describe only linear association. The definition of MIC is given as follows:

Definition: Given a set of ordered pair $D=\{x, y\}$. The x-value and y-value are divided into s bins and t bins, respectively. Then, a s-by-t grid G is obtained. The maximum mutual information $I^*(D, x, y)=\max_G I\left(D|_G\right)$ is taken over all grid G. Therefore, the MIC is given as follows:

$$\operatorname{MIC}(x, y \mid D)=\max_{xy<B(|D|)} \frac{I^*(D, x, y)}{\log_2 \min\{x, y\}} \tag{2}$$

where B is a growing function. In [23] the authors suggest that $B(n)=n^{0.6}$ and n is the sample size. Higher MIC value indicate that the two variables x and y are highly dependent. The MIC value of arbitrary variable pairs falls between $[0,1]$ and the MIC is also symmetry, in which $\operatorname{MIC}(x, y)=\operatorname{MIC}(y, x)$.

To obtain a sparse graph structure, an edge will be established between x_i and x_j, if the MIC value is over a threshold. In this paper, the threshold is set to 0.3.

In this paper, the graphs constructed by KNN and MINE are converted to a heterogeneous graph with multi-view, in which there will be two types of edge between every node pair. However, in order to maintain the independence of each sub-graph, we use paralleled backbones (GAT) to extract the fault feature.

B. Graph Attention network

The sub-graph construction based on KNN and MINE can reveal the relationship between different nodes. However, all sensor signal comes from the same ICS and share the same control rules. Therefore, the graph structure of different faults will have similar parts. In order to adjust the graph structure information of different fault categories, we introduce the attention-based graph neural network (GAT) to learn the fault feature.

The self-attention mechanism is employed to replace the statically normalized graph convolution operation. First, a feature transform layer is used to produce a new set of node features.

$$z_i^{(l)}=W^{(l)} h_i^{(l)}, \tag{3}$$

where h_i^l is the lower layer feature embedding and the initial input h_i^0 is one senor signal segment, $W^{(l)}$ is the learnable weight matrix, and $z_i^{(l)}$ is the transformed node feature.

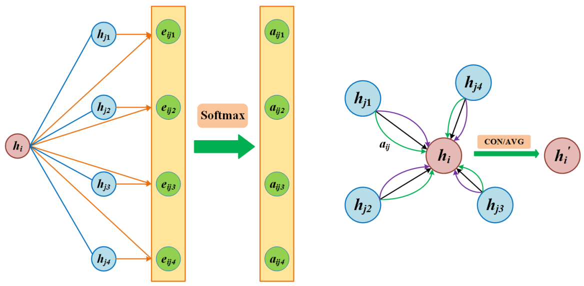

Fig. 2. Simplified schematic of GAT.

Then a shared attention weight vector $\overrightarrow{\boldsymbol{a}}^{(l)}$ is introduced to computes attention coefficients $e_{ij}^{(l)}$.

$$e_{ij}^{(l)} = \text{LeakyReLU}\left(\vec{a}^{(l)^T}\left(z_i^{(l)} \| z_j^{(l)}\right)\right). \tag{4}$$

In order to compare the attention coefficient across different nodes, a softmax is used to normalize the attention coefficient:

$$\alpha_{ij}^{(l)} = \frac{\exp\left(e_{ij}^{(l)}\right)}{\sum_{k \in \mathcal{N}(i)} \exp\left(e_{ik}^{(l)}\right)} \tag{5}$$

The learned normalized attention coefficients introduce a weight to the edge constructed by KNN and MINE, which distinguish the importance of the neighbor node to the target node according to the supervision information. At last, the node feature h_i^{l+1} is updated by aggregating the neighbor node and edge information, which is shown in Fig. 2

$$h_i^{(l+1)} = \sigma\left(\sum_{j \in \mathcal{N}(i)} \alpha_{ij}^{(l)} z_j^{(l)}\right) \tag{6}$$

GAT also introduces multi-head attention to stabilize the learning process and advance the model capability. There are two ways to merge the outputs of each head, concatenation and average. It is suggested to use concatenation operation for intermediary layer and use average layer for the final layer.

$$\text{Concatenation: } h_i^{(l+1)} = \|_{k=1}^{K} \sigma\left(\sum_{j \in \mathcal{N}(i)} \alpha_{ij}^k W^k h_j^{(l)}\right), \tag{7}$$

$$\text{Average: } h_i^{(l+1)} = \sigma\left(\frac{1}{K} \sum_{k=1}^{K} \sum_{j \in \mathcal{N}(i)} \alpha_{ij}^k W^k h_j^{(l)}\right), \tag{8}$$

where k is the number of attention heads and W^k is the corresponding learnable feature transform weight matrix.

C. Read out

The node feature is updated through the GAT layer by aggregating the information of neighbor nodes with edge weights. In order to obtain the global representation of a graph, a read out function is employed. In this paper, we consider the global mean and max characteristic of a graph. Thus, the final embedding of each view of the graph is obtained by the average-pooling and max-pooling read out function, which is shown as follows.

$$\mathbf{r}_{mean} = \frac{1}{N} \sum_{n=1}^{N} \mathbf{H}_n^L, \tag{9}$$

$$\mathbf{r}_{max} = \max_{n=1}^{N} \mathbf{H}_n^L, \tag{10}$$

In this paper, a multi-view graph is used to real the properties of the sensor signals, and thus the multi-view graph embedding is obtained by concatenating all views of the graph, which is defined as follows.

$$\mathbf{r} = \text{CONCAT}\left\{\mathbf{r}_{mean}^1, \mathbf{r}_{max}^1, \mathbf{r}_{mean}^2, \mathbf{r}_{max}^2\right\}. \tag{11}$$

where $\mathbf{r}^1$ is the read out feature of the first view of graph and $\mathbf{r}^2$ is the read out feature of the second view of graph.

D. Focal loss

In practice, the fault event is rare compared to the normal state and the duration of the fault type is varied based on the fault type, which will lead to a class imbalance problem. In the training phase, not only the class with a large number of samples but also the easily classified samples comprise the majority of the loss and influence the gradient.

We first give the definition of the estimated probability p_t of a class with label $y = 1$:

$$p_\text{t} = \begin{cases} p & \text{if } y = 1 \\ 1 - p & \text{otherwise,} \end{cases} \tag{12}$$

Focal loss introduces a modulating factor $(1 - p_t)^\gamma$ to the cross entropy loss, where the focusing parameter $\gamma \geq 0$ can be tuned as a hyperparameter. The modulating factor focus on balancing the influence of easy classification samples. Therefore, the focal loss is defined as follows:

$$\text{FL}(p_\text{t}) = -\alpha_t (1 - p_\text{t})^\gamma \log(p_\text{t}) \tag{13}$$

where α_t is a balance parameter.

IV. Experiment

We evaluate the fault diagnosis performance of the proposed method on a three-phase flow facility (TFF) data set. Three problems are considered in this paper: (1) The comparison with the state-of-the-art fault diagnosis method. (2) The effectiveness of the attention-based GAT. (3) The effectiveness of multi-view of graph construction and effectiveness of focal-loss.

A. Data description

The three-phase flow facility data set collected by Cranfield University is a benchmark case for variant health monitoring tasks [24]. The sketch of the TFF is shown in Fig. 3. The TFF data set contain one normal case and six fault case under steady or changing operation. Each fault case starts by health state, and the faults are introduced after a certain time. The simulator outputs 24 process measurements and the details can be found in [24]. All normal and fault case

TABLE I
FAULT TYPES IN THE TFF DATA SETS

No.	Fault types	Sample number train set	valid set	test set
1	Air line blockage	127	23	44
2	Water line blockage	121	16	35
3	Top separator input blockage	307	35	91
4	Open direct bypass	153	20	53
5	Slugging conditions	75	11	19
6	Pressurization of the 2" line	64	13	19
7	Normal	478	72	118

TABLE II
THE MODEL STRUCTURE OF MG-GAT

Layer name	Input size	Output size	Heads number
GAT layer1	50	128	3
GAT layer2	384	128	1
Fc Layer1	256	256	-
Fc Layer2	256	128	-
Fc Layer3	128	7	-

data were normalized by max-min normalization, where $\mathbf{x} = (\mathbf{x} - x_{min}) / (x_{max} - x_{min})$. The normal data are removed from the fault cases, and the fault data are cut into segments with a size of 50. The samplings frequency is 1 Hz. Both normal samples and fault samples are mixed together and randomly divided into the training set, valid set, and test set with the ratio 70%, 10%, and 20%, respectively. Table I gives the number of samples and corresponding fault categories.

B. Experimental setup

The accuracy and Macro F1 score are used to evaluate the performance of the proposed method. We conduct a number of experiments for all comparison methods and MG-GAT and give the best results by tuning the hyperparameters. The learning rate is 0.0005. The hyperparameter γ of focal loss equals 2, α for normal class is tuned in set of $\{1.12, 1.32, 1.8\}$ and α for other class equals 1. The maximum epoch equals 300. The structure of MG-GAT is shown in Table II. The proposed method was implemented using PyTorch geometric on a PC server with NVIDIA RTX 2080Ti and Xeon Silver 4212 CPU.

C. Comparison with baseline methods

In order to demonstrate the superior performance of the proposed method, the comparison methods are list as follows:

1) SR-CNN: SR-CNN [12] transform the multivariate signals into 2D images by calculating the Spearman rank correlation matrix. Then a CNN model is used to extract the fault feature and classify the fault categories.
2) RC: RC classifier [25] is a hybrid generative-discriminative approach, which learns the low-dimensional embedding of multivariate time series in an unsupervised way. And Then, a decoder (e.g. SVM, MLP) is employed to classify the embedding.

The comparison result is shown in Table III. The MG-GAT achieves the best performance compared with the baseline methods. The baseline methods, SR-CNN focus on extracting the correlation between different sensor signals, and RC focuses on extracting the feature of sensor signals. In consequence, both the SR-CNN model and RC model only consider a certain view of the sensor signals. In contrast, MG-GAT fuses the data properties with the relation between the sensor signals, and thus MG-GAT can obtain better fault recognition results.

TABLE III
COMPARISON BETWEEN BASELINE METHODS AND MG-GAT

Method	Valid-set acc	Valid-set F1	Test-set acc	Test-set F1
SR-CNN	73.16	63.07	71.24	63.13
RC	90.00	88.90	88.39	87.86
MG-GAT	95.78	94.30	93.93	90.92

D. Effects of multi-view graph and focal loss

In this paper, the multi-view graphs with respect to similarity and correlation are constructed to reveal different properties of the ICS. To demonstrate the effectiveness of the multi-view graph construction, a comparison between different graph construction methods is given. The comparison methods are given as follows. Meanwhile, to validate the positive effect of focal loss, the comparison of CE loss is also given in this section.

1) Fully connected (FC) graph. The connection exists between all the nodes, except the self-connection.
2) K-Nearest Neighbor (KNN) graph. Only one-view graph is constructed by the KNN method.
3) Maximal information-based nonparametric exploration (MINE) [23] graph. Only one-view graph is constructed by calculating the MIC value.
4) Multi-view graph: Graph construction with two views, KNN-graph and MINE-graph.

The comparison results of the valid set and test set are shown in Table IV. It is can be seen that the Multi-view graph achieves the best performance, the results of the KNN-graph are better than the MINE graph, and the FC graph achieves no competitive results. The results indicated that (1) it is effective to construct a graph structure in advance, (2) different graph construction methods reveal variant properties of data and have different contributions to the fault recognition, and (3) multi-view graph can provide support information to each other and obtain better performance of fault diagnosis.

E. Effects of attention-based GAT

In order to differentiate the graph structure according to fault categories, attention-based GAT is employed in this paper. This section gives the comparison between GAT and the basic graph convolution method GCN. The results are shown in Fig. 4, which also include the results of the FC, KNN, and MINE graph. All results of GAT are better than GCN. First, although, the FC graph does not reveal any properties of the system, the GAT can also achieve results close to the MINE graph with the GAT backbone. This validates that the attention

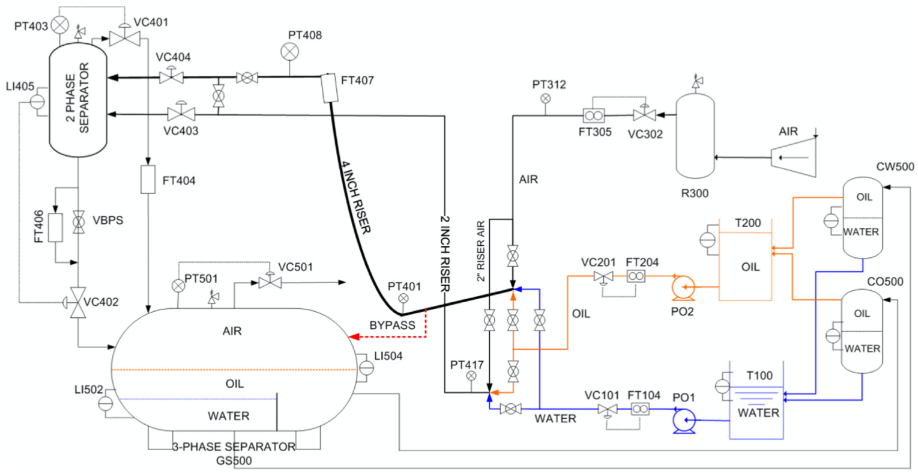

Fig. 3. Sketch of three-phase flow facility [24].

TABLE IV
COMPARISON BETWEEN DIFFERENT GRAPH CONSTRUCTION METHOD UNDER CE LOSS AND FOCAL LOSS

Method	CE loss				Focal loss			
	valid-set acc	valid-set F1	Test-set acc	Test-set F1	valid-set acc	valid-set F1	Test-set acc	Test-set F1
FC	80.00	74.24	76.51	66.90	84.74	74.78	81.00	69.00
MINE	86.32	78.61	82.85	74.69	86.31	78.86	83.37	74.88
KNN	91.05	87.48	86.81	82.93	93.16	90.50	88.39	85.03
KNN-MINE	94.21	92.39	90.76	86.95	95.79	94.30	93.93	90.92

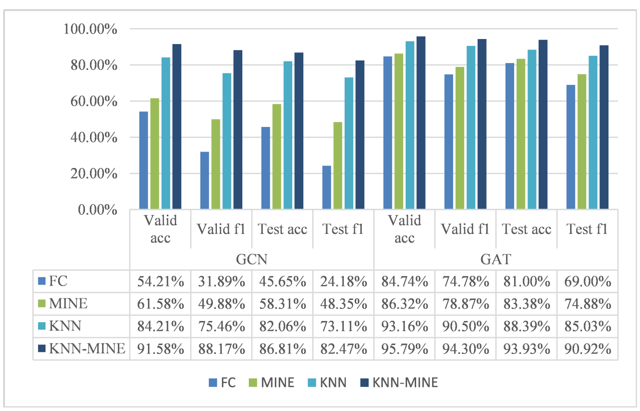

Fig. 4. The comparison between GCN and GAT.

mechanism can adjust the structure information and promote the performance of fault diagnosis. Second, the GCN results of multi-view graphs are also superior to other graph construction methods with GCN backbones, which also demonstrate the effectiveness of multi-view graph construction.

V. CONCLUSION

Considering fusing the multivariate sensor signals with relational information, this paper proposes a framework MG-GAT for fault diagnosis of an industrial control system. MG-GAT can model a multi-view graph by exploring similarity and correlation relationships between sensor signals. Moreover, the GAT module and focal loss are used to extract fault features and solve the sample imbalance problem. Experiments on the three-phase flow facility show that the proposed MG-GAT achieves superior performance on fault diagnosis. The effect of multi-view graph, GAT feature extraction module, and focal loss are further investigated, and the results show the rationality of the proposed model. Future work will focus on exploring the relationship between sensor signals in a learning way and introducing industrial background knowledge to the fault diagnosis method.

REFERENCES

[1] Y. Tao, H. Shi, B. Song, and S. Tan, "A novel dynamic weight principal component analysis method and hierarchical monitoring strategy for process fault detection and diagnosis," *IEEE Transactions on Industrial Electronics*, vol. 67, no. 9, pp. 7994–8004, 2020.

[2] G. Wang and J. Jiao, "A kernel least squares based approach for nonlinear quality-related fault detection," *IEEE Transactions on Industrial Electronics*, vol. 64, no. 4, pp. 3195–3204, 2017.

[3] S. Mahadevan and S. L. Shah, "Fault detection and diagnosis in process data using one-class support vector machines," *Journal of Process Control*, vol. 19, no. 10, pp. 1627–1639, 2009.

[4] C.-L. Liu, W.-H. Hsaio, and Y.-C. Tu, "Time series classification with multivariate convolutional neural network," *IEEE Transactions on Industrial Electronics*, vol. 66, no. 6, pp. 4788–4797, 2018.

[5] V. Garcia and J. Bruna, "Few-shot learning with graph neural networks," Jan. 2018. 6th International Conference on Learning Representations, ICLR 2018 ; Conference date: 30-04-2018 Through 03-05-2018.

[6] H. Peng, J. Li, Y. He, Y. Liu, M. Bao, L. Wang, Y. Song, and Q. Yang, "Large-scale hierarchical text classification with recursively regularized deep graph-cnn," in *Proceedings of the 2018 world wide web conference*, pp. 1063–1072, 2018.

[7] J. Klicpera, J. Groß, and S. Günnemann, "Directional message passing for molecular graphs," in *International Conference on Learning Representations (ICLR)*, 2020.

[8] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Liò, and Y. Bengio, "Graph Attention Networks," *International Conference on Learning Representations*, 2018. accepted as poster.
[9] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 2, pp. 318–327, 2020.
[10] G. Jiang, H. He, J. Yan, and P. Xie, "Multiscale convolutional neural networks for fault diagnosis of wind turbine gearbox," *IEEE Transactions on Industrial Electronics*, vol. 66, no. 4, pp. 3196–3207, 2019.
[11] L. Feng and C. Zhao, "Fault description based attribute transfer for zero-sample industrial fault diagnosis," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 3, pp. 1852–1862, 2021.
[12] D. Yang, Y. Pang, B. Zhou, and K. Li, "Fault diagnosis for energy internet using correlation processing-based convolutional neural networks," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 49, no. 8, pp. 1739–1748, 2019.
[13] X. Zhao, M. Jia, and Z. Liu, "Semisupervised graph convolution deep belief network for fault diagnosis of electormechanical system with limited labeled data," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 8, pp. 5450–5460, 2021.
[14] T. N. Kipf and M. Welling, "Semi-Supervised Classification with Graph Convolutional Networks," in *Proceedings of the 5th International Conference on Learning Representations*, ICLR '17, 2017.
[15] W. L. Hamilton, R. Ying, and J. Leskovec, "Inductive representation learning on large graphs," NIPS'17, (Red Hook, NY, USA), p. 1025–1035, Curran Associates Inc., 2017.
[16] X. Wang, X. He, Y. Cao, M. Liu, and T.-S. Chua, "Kgat: Knowledge graph attention network for recommendation," in *KDD*, 2019.
[17] S. Mai, S. Zheng, Y. Yang, and H. Hu, "Communicative message passing for inductive relation reasoning," *arXiv preprint arXiv:2012.08911*, 2020.
[18] J. Li, F. Yang, M. Tomizuka, and C. Choi, "Evolvegraph: Multi-agent trajectory prediction with dynamic relational reasoning," in *Advances in Neural Information Processing Systems* (H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin, eds.), vol. 33, pp. 19783–19794, Curran Associates, Inc., 2020.
[19] Y. Liu, Z. Li, S. Pan, C. Gong, C. Zhou, and G. Karypis, "Anomaly detection on attributed networks via contrastive self-supervised learning," *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–15, 2021.
[20] D. Teney, L. Liu, and A. van Den Hengel, "Graph-structured representations for visual question answering," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1–9, 2017.
[21] H. Hu, J. Gu, Z. Zhang, J. Dai, and Y. Wei, "Relation networks for object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3588–3597, 2018.
[22] W. Liao, D. Yang, Y. Wang, and X. Ren, "Fault diagnosis of power transformers using graph convolutional network," *CSEE Journal of Power and Energy Systems*, vol. 7, no. 2, pp. 241 – 249, 2021.
[23] D. N. Reshef, Y. A. Reshef, H. K. Finucane, S. R. Grossman, G. McVean, P. J. Turnbaugh, E. S. Lander, M. Mitzenmacher, and P. C. Sabeti, "Detecting novel associations in large data sets," *Science*, vol. 334, no. 6062, pp. 1518–1524, 2011.
[24] C. Ruiz-Cárcel, Y. Cao, D. Mba, L. Lao, and R. Samuel, "Statistical process monitoring of a multiphase flow facility," *Control Engineering Practice*, vol. 42, pp. 74–88, 2015.
[25] F. M. Bianchi, S. Scardapane, S. Løkse, and R. Jenssen, "Reservoir computing approaches for representation and classification of multivariate time series," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 5, pp. 2169–2179, 2021.