UMAD-G : Unsupervised Multi-modal time series Anomaly Detection via Graph

Jinhyeok Lee
Department of Industrial and Management Engineering
Korea University
Seoul, South Korea
plumlee@korea.ac.kr

Jungeol Baek*
Department of Industrial and Management Engineering
Korea University
Seoul, South Korea
jungeol@korea.ac.kr

***Abstract*—In the wake of Industry 4.0, many companies are equipping their systems with more sensors and collecting huge amounts of sensor data. This data, in the form of time series, plays an important role in increasing productivity and implementing smart factories. One of the most important things is to detect anomalies caused by various factors (equipment ageing, sudden stops in the production line, interference from external factors). One of the limitations of existing time series anomaly detection research is that it does not take into account the complex spatial dependencies (inter- and intra-modal correlations) of sensor data. Therefore, we propose a model that can accurately and effectively detect outliers in multi-modal time series collected from manufacturing facilities by utilizing Graph Attention Networks. Our proposed model demonstrates highest performance through performance comparison with six baselines.**

***Index Terms*—multi-modal time series, anomaly detection, sensor, graph attention networks, unsupervised learning**

I. INTRODUCTION

Recently, the industry is witnessing the emergence of groundbreaking technologies and production methodologies concurrent with the advent of the Industry 4.0. Specifically, in the context of burgeoning trends like smart factory and digital transformation, corporations are increasingly adopting systems capable of integrating additional sensors for extensive data collection. Predominantly, this sensor data is being archived in the format of time series data.

This time series data provides insights into the complexity of the production process and is pivotal in the analysis and enhancement of the production system's performance. Nonetheless, the occurrence of anomalies is not uncommon, often attributable to unforeseen equipment malfunctions, abrupt halts in the production line, or perturbations arising from external influences. Swift and precise detection of anomalies in such time series data is vital for the successful functioning and stable production within smart factory system [1].

Given that existing research on time series outlier detection predominantly focused on univariate time series data, the practice of anomaly detection has largely involved engineers setting static thresholds for each sensor and managing them independently. However, this method is labor-intensive and renders accurate anomaly detection challenging, particularly due to the complex interactions characteristic of sensor operations. To address these issues, research is progressing in the field of multi-modal time series (MTS) anomaly detection. MTS is a type of time series data defined when multivariate time series data is viewed as a set of univariate time series data and univariate time series data with topologically different modality (e.g., temperature, speed, current) is included according to the characteristics of each sensor.

Nevertheless, accurate detection of anomalies in MTS is challenging, owing to its intricate spatial dependencies, which include both inter-modality and intra-modality correlations, as well as temporal dependencies. Previous studies have focused on reflecting the temporal dependence of time series. However, these methods are limited to capturing temporal changes and fail to address the spatial dependencies between different time series. In response, models such as MTAD-GAT [2] and GDN [3] have been developed for anomaly detection, leveraging Graph Neural Networks (GNN). These advancements have led to a significantly improved anomaly detection performance compared to existing models. While preceding methodologies have advanced significantly, they are impeded by limitations such as the loss of information among variables arising from the adopted graph structure learning method, commonly termed "Top-k". Another critical limitation is the neglect of the evolving significance of variables over time.

To address these challenges, we propose **UMAD-G** (**U**nsupervised **M**ulti-modal time series **A**nomaly **D**etection via **G**raph), a methodology utilizing Graph Attention Networks [4]. Specifically, through the utilization of an attention-based update algorithm, the proposed method identifies a more precise graph structure, thereby minimizing information loss during the graph representation process and reducing computational costs. Subsequently, it extracts the latent vector by adeptly capturing both spatial and temporal dependencies within each interval of the MTS, achieved through the graph attention module. Anomaly detection is subsequently conducted using the latent vector derived in this manner. Furthermore, the method jointly optimizes both the reconstruction error and the prediction error, thereby amalgamating the strengths of each approach.

The key contributions of the method presented in this paper can be summarized as follows:

- The complex spatial dependencies (inter-modal and intra-modal correlations) that exist within MTS are effectively reflected through GNN.
- By identifying the connection structure of the graph through the attention-based update algorithm, it minimizes the information loss that occurs when representing the time series as a graph structure.
- Accurate anomaly detection is achievable through the graph attention module, which effectively captures and reflects both spatial and temporal dependencies.
- The benefits of each method are synthesized through the joint optimization of the reconstruction error and the prediction error.

II. Methodology

In this section, we outline our proposed methodology, UMAD-G. UMAD-G takes into account both the spatial and temporal dependencies in MTS via a graph attention module, with the graph structure at this juncture being established through updates driven by attention-based update algorithm.

A. Problem definition

In existing studies, anomaly detection was typically conducted on multivariate time series, $X = [x_1, x_2, \cdots, x_t] \in \mathbb{R}^{N \times T}$, where *N* means the number of variables and *T* means the length of the entire time series. However, multivariate time series collected at manufacturing domain possess varying topological characteristics dependent on the sensor type. As these characteristics alter the spatial dependence between the time series, it becomes imperative to conduct anomaly detection while reflecting these changes. Therefore, this paper aims to conduct anomaly detection on multi-modal time series. The definition of multi-modal time series employed here is consistent with that outlined in Introduction section.

B. Graph structure representation

UMAD-G represents the MTS in a graph structure, where each variable is a node and their correlations are edges. A critical consideration in representing MTS as a graph structure is the definition of edges that signify the correlation between nodes. Existing methods, such as the fully-connected approach [2] and the top-k approach [3], encounter issues with high computational costs and information loss, respectively. Consequently, in this paper, we propose a balanced compromise between these existing approaches. Assume a graph structure $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with *N* nodes. Let $\mathcal{V}$ denote the node representation in which each variable in the MTS is embedded, and $\mathcal{E}$ denote the connectivity between nodes. The edges of each node are represented in a matrix called Adjacency matrix $\mathcal{A}$. Since time series data natively has no information about edges, we utilize attention score that can be obtained from graph attention network (GATv2) [5]. GATv2 is an algorithm that takes into account the importance of variables in a weighted form as the model is trained, fetching information from other nodes connected to the target node. The weight utilized in this instance is known as the Attention score α_{ij}, and the expression related to the learning parameters of GATv2 is as follows:

$$e(\boldsymbol{h}_i, \boldsymbol{h}_j) = \boldsymbol{a}^\top \mathrm{LeakyReLU}(\boldsymbol{W} \cdot [\boldsymbol{h}_i \parallel \boldsymbol{h}_j]) \tag{1}$$

$$\alpha_{ij} = \mathrm{softmax}_j(e(\boldsymbol{h}_i, \boldsymbol{h}_j)) = \frac{\exp(e(\boldsymbol{h}_i, \boldsymbol{h}_j))}{\sum_{j' \in \mathcal{N}_i} \exp(e(\boldsymbol{h}_i, \boldsymbol{h}_{j'}))} \tag{2}$$

$$\boldsymbol{h'}_i = \sigma(\sum_{j \in \mathcal{N}_i} \alpha_{ij} \cdot \boldsymbol{W} \boldsymbol{h}_j) \tag{3}$$

1) Attention-based update algorithm: Attention score is appropriate for identifying edges as it signifies the relevance among MTS variables. For its application, the attention score is determined through a fully-connected graph structure during the initial k epochs. Nodes exhibiting an attention score exceeding a specified threshold are then identified as connected nodes. The adjacency matrix $\mathcal{A}$, defined in this manner, undergoes updates throughout the learning phase commencing from the k+1 epoch. The adequately trained matrix $\mathcal{A}$ is then employed during the testing process.

C. Graph attention module

For the training MTS X, we use the sliding window with length w to generate inputs of fixed length. We define $\tilde{X}$ as the input of Graph attention module at time t:

$$\tilde{\boldsymbol{X}} = [\boldsymbol{x}_{t-w+1}, \boldsymbol{x}_{t-w+2}, \cdots, \boldsymbol{x}_t] \in \mathbb{R}^{N \times w} \tag{4}$$

The proposed graph attention module consists of three attention modules. The multi-head attention module captures the global spatial relationships of MTS, while the intra-modal attention module is designed to model the relationships between different modalities. In the temporal attention module, temporal dependencies are modeled, and the feature vectors produced from each of the three modules are concatenated into a single feature vector.

1) Multi-head attention module: The multi-head attention module interacts with all nodes during the feature vector update process to reflect the global spatial relationships, which is formulated as:

$$\boldsymbol{h'}_i = \sigma(\frac{1}{K} \sum_{k=1}^{K} \sum_{j \in \mathcal{N}_i} \alpha_{ij} \cdot \boldsymbol{W} \boldsymbol{h}_j) \tag{5}$$

where, K is the number of heads in the multi-head attention module. The remaining formula is the same as that typically used for updating feature vectors in GATv2.

2) Intra-modal attention module: The intra-modal attention module introduced to reflect the modalities of MTS.The process of updating the feature vector is similar to that in the multi-head attention module. However, it differs in that when constructing the adjacency matrix used for learning, neighboring nodes are restricted to variables with different modalities.

3) Temporal attention module: Additionally, the temporal attention module models temporal dependencies. For this purpose, the input data X is transposed to X^T, which is then used as input data.

D. Joint optimization

As previously mentioned, we train our model through joint optimization. This is to integrate the advantages of forecasting-based methods and reconstruction-based methods. The forecasting-based model detects anomalies by comparing predicted values with actual values. It is sensitive even to slight pattern variations, making it particularly effective in detecting point anomalies. On the other hand, the reconstruction-based model performs anomaly detection using reconstruction errors and tends to reflect the overall distribution of the data well. This attribute makes it robust to local changes. These advantages are integrated by jointly optimizing both the Forecasting loss L_f and the Reconstruction loss L_r. The output from the graph attention module is utilized as the input for each module.

1) Forecasting module: To reduce the complexity of the model in the Forecasting module, a simple multi-layer perceptron (MLP) is employed, and the output of the graph attention module serves as the input to this MLP. The forecasting loss can be formulated as Root Mean Square Error (RMSE):

$$L_f = \frac{1}{T-w}\sqrt{\sum_{n=1}^{N}(x_{n,t+1} - \hat{x}_{n,t+1})^2} \tag{6}$$

where $x_{n,t+1}$ denotes the ground truth of nth time series at time $t+1$, and $\hat{x}_{n,t+1}$ is the output from the forecasting module of nth time series at time $t+1$

2) Reconstruction module: For the Reconstruction module, a Variational Auto-Encoder (VAE) [6] is adopted. The reconstruction loss is formulated as:

$$L_r = \mathbb{E}_{z\sim q_\phi(z|x)}[\log p_\theta(x|z)] - \mathrm{KL}(q_\phi(z|x)\|p(z)) \tag{7}$$

where, $\mathbb{E}_{z\sim q_\phi(z|x)}[\log p_\theta(x|z)]$ denotes the log-likelihood expectation of x. KL represents the KL divergence.

E. Anomaly detection

As previously described, the Forecasting module and the Reconstruction module each yield a predicted value, $\hat{\boldsymbol{x}}_{\boldsymbol{i}}$ and a reconstruction probability, p_i, respectively. $\hat{\boldsymbol{x}}_{\boldsymbol{i}}$ represents the predicted value of the ith time series at each timestamp and p_i represents the probability of reconstruction of the ith time series at each timestamp. The final anomaly score is defined as the sum of the results at each timestamp and is calculated as follows:

$$score = \sum_{i=1}^{N}\frac{(1-p_i)+\gamma(\boldsymbol{x}_i - \hat{\boldsymbol{x}}_{\boldsymbol{i}})^2}{1+\gamma} \tag{8}$$

where, γ is a hyperparameter that adjusts the weighting of the two modules, and we set its value to 0.5.

TABLE I
STATISTICS OF DATASETS

Datasets	Features	Train	Test	Anomalies rate
SWaT	51	47515	44986	11.97%
WADI	127	118795	17275	5.99

III. EXPERIMENTS

A. Dataset

Since there is a lack of real-world datasets with labeled ground truth for manufacturing factories as our domain, we conducted this experiment on two benchmark datasets: SWaT, WADI. Secure Water Treatment (SWaT) dataset [7] originates from a Testbed that is a scaled-down version of a real water treatment plant. This Testbed was developed by the iTrust Cyber Security Research Centre at the Singapore University of Technology and Design. It serves as a miniature version of a modern Cyber-Physical system, enabling the study of both normal sensor operations and the impacts of various cyber-attacks. Water Distribution (WADI) dataset [8] is an extension of the SWaT, providing data encompassing both normal operations and deliberate attack scenarios. The statistics of both datasets have been summarized and presented in Table 1.

B. Baselines

We compare the performance of our proposed method with six popular anomaly detection models, including DAGMM [9], Omni-Anomaly [10], MAD-GAN [11], USAD [12], MTAD-GAT, and GDN. For a more comprehensive comparison, we included two representative GNN-based anomaly detection models in the baselines.

C. Evaluation metric

We use the Area Under the Receiver Operating Characteristic Curve (AUROC) as the performance metric for comparing the effectiveness of different models. AUROC allows for a threshold-independent assessment of performance, as it is calculated for all possible threshold values. Specifically, the AUROC is the calculated area under the ROC curve. A model is considered to have better performance the closer the AUROC value is to 1.

The ROC curve itself is a graphical representation showing the variation of the False Positive Rate (FPR) and True Positive Rate (TPR) at different threshold settings. FPR refers to the proportion of cases that are actually negative but are incorrectly predicted as positive by the model. On the other hand, TPR indicates the proportion of actual positive cases that the model correctly predicts as positive.

Most time series anomaly detection models commonly use the F1 score as a performance metric. However, due to the ambiguity in setting the threshold and the recently highlighted issue of overestimation with the point adjustment method [13], this paper solely employs AUROC for performance comparison.

TABLE II
CONFUSION MATRIX

	Predicted: No	**Predicted: Yes**
Actual: No	True Negative (TN)	False Positive (FP)
Actual: Yes	False Negative (FN)	True Positive (TP)

TABLE III
RESULTS OF EXPERIMENTS

Model	**SWaT**	**WADI**
DAGMM	0.8436	0.6721
OmniAnomaly	0.8167	0.8198
MAD-GAN	0.8363	0.8026
USAD	0.836	0.7723
MTAD-GAT	0.8464	0.5202
GDN	<u>0.8781</u>	<u>0.8326</u>
UMAD-G	**0.8904**	**0.8428**

D. Results and analysis

The experiment was conducted by repeating each comparison model 10 times. The batch size was set at 64 for the SWaT dataset and 128 for the WADI dataset. Table 2 summarizes the comparison between UMAD-G and the baselines from the perspective of AUROC. Bold indicates the highest performance, underlined indicates the second highest performance. The results demonstrate that UMAD-G achieved superior performance over the baselines on both datasets. Notably, the proposed model's superiority is evident in its slightly better performance compared to GDN, a known GNN-based anomaly detection model renowned for its robust performance.

IV. CONCLUSION

In this paper, we propose UMAD-G, a model designed for effectively detecting anomalies by reflecting the spatial and temporal dependencies in multi-modal time series. UMAD-G models these dependencies through three attention modules designed to extract different features from the MTS. The proposed model integrates the advantages of each by jointly optimizing the reconstruction error and prediction error. Performance comparison experiments on benchmark datasets (i.e., SWaT, WADI) show that UMAD-G outperforms Baselines in terms of AUROC.

Future work in this paper can be done on two fronts. First, our model performed anomaly detection without any domain knowledge about the correlations between variables. If domain knowledge can be acquired in advance, it can help the model's anomaly detection performance. Second, current anomaly detection research assumes a situation after enough data has been collected. Extending this to real-time anomaly detection will result in a more practical model.

ACKNOWLEDGMENT

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) (NRF-2022R1A2C2004457). This work was also supported by Samsung Electronics Co., Ltd. (IO201210-07929-01) and Brain Korea 21 FOUR.

REFERENCES

[1] G. A. Susto, M. Terzi, and A. Beghi, "Anomaly detection approaches for semiconductor manufacturing," *Procedia Manufacturing*, vol. 11, pp. 2018–2024, 2017.

[2] H. Zhao, Y. Wang, J. Duan, C. Huang, D. Cao, Y. Tong, B. Xu, J. Bai, J. Tong, and Q. Zhang, "Multivariate time-series anomaly detection via graph attention network," in *2020 IEEE International Conference on Data Mining (ICDM)*. IEEE, 2020, pp. 841–850.

[3] A. Deng and B. Hooi, "Graph neural network-based anomaly detection in multivariate time series," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 35, no. 5, 2021, pp. 4027–4035.

[4] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Lio, and Y. Bengio, "Graph attention networks," *arXiv preprint arXiv:1710.10903*, 2017.

[5] S. Brody, U. Alon, and E. Yahav, "How attentive are graph attention networks?" *arXiv preprint arXiv:2105.14491*, 2021.

[6] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," *arXiv preprint arXiv:1312.6114*, 2013.

[7] J. Goh, S. Adepu, K. N. Junejo, and A. Mathur, "A dataset to support research in the design of secure water treatment systems," in *Critical Information Infrastructures Security: 11th International Conference, CRITIS 2016, Paris, France, October 10–12, 2016, Revised Selected Papers 11*. Springer, 2017, pp. 88–99.

[8] C. M. Ahmed, V. R. Palleti, and A. P. Mathur, "Wadi: a water distribution testbed for research in the design of secure cyber physical systems," in *Proceedings of the 3rd international workshop on cyber-physical systems for smart water networks*, 2017, pp. 25–28.

[9] B. Zong, Q. Song, M. R. Min, W. Cheng, C. Lumezanu, D. Cho, and H. Chen, "Deep autoencoding gaussian mixture model for unsupervised anomaly detection," in *International conference on learning representations*, 2018.

[10] Y. Su, Y. Zhao, C. Niu, R. Liu, W. Sun, and D. Pei, "Robust anomaly detection for multivariate time series through stochastic recurrent neural network," in *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, 2019, pp. 2828–2837.

[11] D. Li, D. Chen, B. Jin, L. Shi, J. Goh, and S.-K. Ng, "Mad-gan: Multivariate anomaly detection for time series data with generative adversarial networks," in *International conference on artificial neural networks*. Springer, 2019, pp. 703–716.

[12] J. Audibert, P. Michiardi, F. Guyard, S. Marti, and M. A. Zuluaga, "Usad: Unsupervised anomaly detection on multivariate time series," in *Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining*, 2020, pp. 3395–3404.

[13] S. Kim, K. Choi, H.-S. Choi, B. Lee, and S. Yoon, "Towards a rigorous evaluation of time-series anomaly detection," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 36, no. 7, 2022, pp. 7194–7201.