Multi-Scale Dynamic Graph Learning for Time Series Anomaly Detection (Student Abstract)

Yixuan Jin[1], Yutao Wei[1], Zhangtao Cheng[*] [1,3], Wenxin Tai[1,3], Chunjing Xiao[2], Ting Zhong[1,3]

[1]University of Electronic Science and Technology of China, Chengdu, Sichuan 610054, China
[2]Henan University, Kaifeng, Henan 475000, China
[3]Kashi Institute of Electronics and Information Industry, Kashgar 844000, China
yx.jin@std.uestc.edu.cn, yutaowei@std.uestc.edu.cn, zhangtao.cheng@outlook.com, wxtai@std.uestc.edu.cn, chunjingxiao@gmail.com, zhongting@uestc.edu.cn

Abstract

The success of graph neural networks (GNNs) has spurred numerous new works leveraging GNNs for modeling multivariate time series anomaly detection. Despite their achieved performance improvements, most of them only consider static graph to describe the spatial-temporal dependencies between time series. Moreover, existing works neglect the time- and scale-changing structures of time series. In this work, we propose MDGAD, a novel multi-scale dynamic graph structure learning approach for time series anomaly detection. We design a multi-scale graph structure learning module that captures the complex correlations among time series, constructing an evolving graph at each scale. Meanwhile, an anomaly detector is used to combine bilateral prediction errors to detect abnormal data. Experiments conducted on two time series datasets demonstrate the effectiveness of MDGAD.

Introduction

Multivariate time series (MTS) data exist in many real-world industrial and business systems and are often associated with spatial information. For example, the water treatment plant has a lot of sensors monitoring various important indicators such as water level and equipment voltage, producing massive spatial MTS data. Detecting the anomalies from the MTS data is of paramount importance for such systems to ensure security and avoid economic losses (Xiao et al. 2023).

Recently, graph neural networks (GNNs) have shown great potential for capturing the complex spatial-temporal dependencies between times series and have been widely applied to detect anomalies in MTS data. For example, GDN (Deng and Hooi 2021) proposed a graph structure learning module and a graph attention network to make forecasting on MTS. However, most GNN-based studies only utilized static graphs for MTS anomaly detection.

Existing MTS learning models mainly confront the following two notable drawbacks. (1) The graph structure of time series varies over time. Works focused on static and fixed graph structures ignore the structural dynamics in MTS data. (2) The graph structure also changes at different extraction scales along with temporal dimensions. Little attention has been paid to the structural relationships among time series at different scales. Moreover, static graph structure is insufficient to capture the long-term variation in MTS data, which we believe is critical for MTS anomaly detection.

In this work, we proposed a **M**ulti-scale **D**ynamic **G**raph structure learning approach for multivariate time series **A**nomaly **D**etection (**MDGAD**). We design a multi-scale graph structure learning module that converts MTS data into multi-scale dynamic graphs and apply the bilateral forecast method for predicting the masked nodes in graphs. At last, we use the prediction errors to detect anomalies in MTS. Extensive experiments conducted on real-world MTS datasets validated the effectiveness of MDGAD in comparison with state-of-the-art MTS models.

Methodology

Problem Definition. Given a set of N time series associated with C-dimensional features $\mathbf{X} = \{X_1, X_2, \ldots, X_T\}$, $\mathbf{X} \in \mathbb{R}^{N \times C \times T}$. Here X_t indicates a single observation at time t. The MTS anomaly detection problem aims to identify whether a given X_t is anomalous or not without using ground-truth labels.

Overall Framework. The proposed MDGAD consists of a multi-scale graph structure learning module and an anomaly detector. In the graph structure learning module, we first design a dilated inception layer-based temporal convolution network to extract the multi-scale variable representations from the MTS data. It can capture variable features with different scales by adopting dilation factors. Second, we present a scale-specific dynamic graph structure learner to model the variable relationships and establish the dynamic graph. The latest information of the dynamic graph are updated by considering both spatial and temporal dependencies from the MTS data at each scale. At last, the anomaly detector equipped with a bilateral forecast is used to infer the values of the masked points. The inferred values are then combined with the prediction errors as the final anomaly score for MTS anomaly detection.

Multi-Scale Dynamic Graph Structure Learning. The complex industrial scenarios, the correlations between variable features in MTS data can be dynamic and long-term.

[*]corresponding author.

MTS interactions may also evolve at different time scales, making their representation learning extremely challenging. Motivated by this, we design a multi-scale dynamic graph structure learning module to encode multi-scale variable features by leveraging dilation factors and multiple filters with different sizes. Specifically, the dilation factors served as a convolution layer distancing controllers, and the filers at different scales as convolution kernel constructors. By stacking multiple convolution layers, the obtained variable feature representations can capture MTS characteristics at different scales along with temporal dimension. These features are cut and merged as $\xi \in \mathbb{R}^{N \times T \times C_\delta}$.

Second, we feed the variable features ξ into the scale-specific dynamic graph structure learner to generate the spatial-temporal graph. The variable features ξ are separated into m segments along the temporal dimension, each segment has a time length of d. All segments are fed into an aggregator and sent to GRU to obtain the node representations. Next, we concatenate the node representations and adopt dense layers for learning graph structures. The value of the learned graph structure is denoted by adjacency matrices:

$$A^{(l)} = \left[A^{(l,1)}, A^{(l,2)}, \ldots, A^{(l,m^l)}\right], \tag{1}$$

where m^l is the number of segments at the l-th layer.

At last, we update the latest spatial-temporal graph by exploiting mix-skip propagation and information selection. The mix-skip propagation maintains a part of the previous skip step node representations, which are implemented by the graph convolution operation between the obtained prior skip step graph representation and the learned graph structure. Meanwhile, we add up the input variable features by using a hyperparameter λ that controls the ratio between variable features and different skip information. The mix-skip propagation is defined as:

$$E_{(\phi)} = \lambda\xi + (1-\lambda)AE_{(\phi-1)}, \tag{2}$$

where $E_{(\phi)}$ is the representation at skip ϕ and we set $E_{(0)} = \delta$. The information selection is used to selectively retain more important node representations by attaching $E_{(\phi)}$ with different weights $W_{(\phi)}$ adaptively:

$$P' = \sum_{\phi=0}^{\Phi} W_{(\phi)} E_{(\phi)}, \tag{3}$$

where Φ is the depth of the propagation and we leave out the l and m for simplicity. The graph representation $P^{(l+1)}$ of time series at the l-th layer is obtained by residual connection between $P'^{(l+1)}$ and $P^{(l)}$.

Anomaly Detector. Bilateral estimation errors are utilized in the anomaly detector at time t to detect whether x_t^i is anomalous or not. Specifically, the predictions at time t are made by the times from 1 to $t-1$ in the forward direction and from T to $t+1$ in the backward direction. Then we attached the forward and backward forecasting errors with different weights as anomaly scores, which are defined as follows:

Method	**SWaT**			**WADI**		
	P	R	F1	P	R	F1
AE	72.63	52.63	61.03	34.35	34.35	34.35
LSTM-VAE	96.24	59.91	73.85	87.79	14.45	24.82
MAD-GAN	98.97	63.74	77.54	41.44	33.92	37.30
GDN	99.35	68.12	80.82	**97.50**	40.19	56.92
MDGAD	**99.47**	**71.40**	**82.27**	82.47	**46.58**	**60.10**

Table 1: Performance comparison between proposed MDGAD and baselines on two MTS datasets.

$$Z^{\text{fwd}} = \sum_{i=1}^{N} \|x_t^{i,\text{fwd}} - \hat{x}_t^{i,\text{fwd}}\|^2, \tag{4}$$

$$\text{AnomalyScore}(x_t) = \eta Z^{\text{fwd}} + (1-\eta) Z^{\text{bwd}}, \tag{5}$$

where η denotes a balancing weight between 0 and 1, Z^{fwd} and Z^{bwd} indicate the forward and backward prediction errors, respectively. In the end, we combine them to determine whether x_t is an anomalous point.

Experiments & Conclusion

Datasets & Baselines. We conducted experiments on two real world MTS datasets for anomaly detection: Secure Water Treatment (SWaT) and Water Distribution (WADI). We compare our proposed MDGAD model with four strong baselines: Autoencoder (AE), LSTM-VAE , MAD-GAN (Li et al. 2019) and GDN (Deng and Hooi 2021).

Performance Comparison & Discussion. In Table 1, we show the MTS anomaly detection results of MDGAD and baselines, in terms of Precision (P), Recall (R), and F1 score. The results showed that MDGAD achieved the best MTS anomaly detection performance on both datasets for most of the metrics (except for Precision on WADI dataset). Specifically, we have the following observations: (1) Dynamic graph structure construction along with temporal dimension is important for GNN-based MTS anomaly detection. (2) The relationships between MTS data at different scales play a significant role in MTS graph structure learning.

Acknowledgments

This work was supported in part by National Natural Science Foundation of China (Grant No.62176043 and No.62072077).

References

Deng, A.; and Hooi, B. 2021. Graph neural network-based anomaly detection in multivariate time series. In *AAAI*.

Li, D.; Chen, D.; Jin, B.; Shi, L.; Goh, J.; and Ng, S.-K. 2019. MDD-GAN: Multivariate anomaly detection for time series data with generative adversarial networks. In *ICANN*.

Xiao, C.; Gou, Z.; Tai, W.; Zhang, K.; and Zhou, F. 2023. Imputation-based Time-Series Anomaly Detection with Conditional Weight-Incremental Diffusion Models. In *SIGKDD*, 2742–2751.