Spatial-Temporal Dependency Based Multivariate Time Series Anomaly Detection for Industrial Processes

Qi Sun, Yahui Li, Zhenpeng Hu, Chunjie Zhou(✉), and Lu Liu

Huazhong University of Science and Technology, Wuhan, China
cjiezhou@hust.edu.cn

Abstract. Multivariate time series anomaly detection is crucial for ensuring equipment and systems' safe operation in the industrial process. However, detecting anomalies in multivariate time series is challenging due to the complex temporal and spatial dependencies among variables. To address this issue, we propose a multi-task variational autoencoder for multivariate time series anomaly detection. Structurally, it combines multi-task learning with a variational autoencoder structure to obtain a robust representation of time series with noise. In detail, graph attention networks and selective state space models are utilized to capture spatial and temporal dependencies effectively. Experimental results show that the proposed model outperforms six baselines on three datasets, including an anomaly detection dataset of the catalytic cracking process, achieving F1 scores of 0.9389, 0.8151, and 0.9524. In addition, anomaly scores and a causal graph of variables can provide a highly interpretable analysis of results to assist on-site safety managers in timely handling anomalies.

Keywords: Anomaly detection · Multivariate time series · Selective state space model · Graph attention network · Catalytic cracking process

1 Introduction

Fluid Catalytic Cracking (FCC) is a crucial industrial process in the petroleum industry, which holds immense significance for both nations and societies. It is used to convert heavy petroleum fractions into lighter products that improve the fuel efficiency of car engines and reduce environmental pollution. Ensuring safety in complex environments with high temperatures and pressure is imperative for anomaly detection in the FCC process [1]. However, thousands of measuring points make it difficult to monitor them and take timely actions to prevent malfunctions manually. Simultaneously, the continuous and steady flow of material and energy between devices gives the FCC process variables demonstrating complex temporal and spatial dependencies. Therefore, effectively utilizing multiple sensors' spatial and temporal dependencies to detect anomalies early is the primary challenge for FCC process safety.

Anomaly detection technology for time series has significantly progressed and demonstrated promising outcomes in various fields [2]. Advancements in computing

D.-S. Huang et al. (Eds.): ICIC 2024, LNCS 14874, pp. 212–223, 2024.
https://doi.org/10.1007/978-981-97-5618-6_18

power have increased interest in deep learning-based anomaly detection. Forecasting-based and reconstruction-based methods are representative methods. It trains the model on normal data and subsequently identifies anomalies by prediction errors or reconstruction errors [3]. The neural networks utilized make it effectively recognize a typical patterns or behaviors without requiring feature engineering. The main objective of multivariate anomaly detection is to capture temporal and spatial dependencies effectively. In the past decade, Long short-term memory (LSTM) [2, 4, 5], gated recurrent units [6], and convolutional neural networks [7] are introduced and utilized to capture temporal dependencies of time series. The selective state space model [8] has demonstrated remarkable capabilities in capturing temporal dependencies in sequence modeling domains such as text and image processing. Some scholars have employed graph neural networks [9], graph convolutional networks [10], and graph attention networks (GAT) [6, 11] in terms of spatial feature extraction to obtain rich representations of time series.

Recently, some studies have focused on capturing spatial-temporal dependencies and achieving good performance in multivariate time series anomaly detection [6, 10]. However, some studies have overlooked the importance of spatial dependencies [4, 5]. Ignoring the capture of spatial dependencies entails overlooking the complex constraints among process variables, leading to poorer performance for anomaly detection in the industrial process. In addition, during the production process of FCC, the harsh operating environment results in sensor noise that cannot be avoided. These studies have ignored the effect of noise, which may cause the model to fail to learn the pattern of time series. Therefore, robust capture of spatial-temporal dependencies while accounting for noise is crucial to identifying anomalies within the industrial process.

To address these issues, we propose a multi-task variational autoencoder (MTVAE-GM) for multivariate time series anomaly detection in the industrial process. Structurally, architectures of multi-task learning and a variational autoencoder are employed to obtain a robust representation of time series and improve tolerance to noise. In detail, graph attention networks and selective state space models are utilized to effectively capture spatial and temporal dependencies. Unlike traditional anomaly detection methods based on autoencoders, anomaly scores are calculated based on both the prediction and reconstruction of time series.

The contributions of our work are summarized as follows:

(1) A multi-task variational autoencoder is proposed for multivariate anomaly detection, which can robustly capture spatial and temporal dependencies under noise.
(2) Based on a simulation system of the FCC process, we construct a multivariate time series anomaly detection dataset CCPAD for the catalytic cracking process.
(3) Experimental validation is carried out, outperforming the baseline methods on all datasets, achieving F1 scores of 0.9389, 0.8151, and 0.9524.

2 Method

In this section, we first provide the formalization of the multivariate anomaly detection task. Subsequently, we provide an introduction to graph attention networks and selective state space models in preliminaries. Following this, the framework of MTVAE-GM is outlined. We will provide a detailed description of extraction components for the spatial-temporal dependencies after the data preprocessing. Finally, we will explain the joint optimization of MTVAE-GM and the principle of anomaly detection.

2.1 Task Formalization

This study focuses on multivariate anomaly detection tasks in the industrial process. Let $x = \{x_1, x_2, \cdots, x_n\}$ denote the time series of variables monitored by sensors, where n is the length of x, denoting the length of the sensor observation sequence. $x_t \in \mathcal{R}^{\mathrm{m}}$ is a vector of length m at time t $(t \leq n)$: $x_t = \left[x_t^1, x_t^2, \cdots, x_t^m\right]$, where m is the quantity of sensor monitor variables. To detect abnormal states, anomaly scores $s = \{s_1, s_2, \cdots, s_n\}$ are calculated for every time t through reconstruction errors and prediction errors. Once the anomaly score s_t at time t exceeds a specific threshold γ, it is considered an anomaly. The goal of the anomaly detection task for the multivariate time series in the industrial process is to determine the presence or absence of anomalies at each time step within the time series x.

2.2 Preliminaries

Graph Attention Network. GAT can model the relationships between arbitrary nodes in the graph and provide interpretability through edge weights. Given a graph $G = \{v1, v2, \ldots, vn\}$ with n nodes, where vi is the feature vector of each node, the graph attention network calculates the representation hi of node i as follows:

$$h_i = \sigma\left(\sum\nolimits_{j=1}^{L} \alpha_{ij} v_j\right) \tag{1}$$

where α_{ij} is the attention score, which measures the contribution of node j to node i, and σ is a nonlinear activation function, such as the sigmoid function. j denotes the nodes adjacent to node i, and L denotes the total number of adjacent nodes to node i. The attention score α_{ij} is calculated as follows:

$$e_{ij} = LeakyReLU\left(w^T \cdot \left(v_i \oplus v_j\right)\right) \tag{2}$$

$$\alpha_{ij} = \frac{exp(e_{ij})}{\sum_{k=1}^{L} exp(e_{ik})} \tag{3}$$

where $\oplus$ represents the concatenation of representations of node i and node j, and $w \in \mathcal{R}^{2\mathrm{n}}$ is a column vector with learnable parameters used to perform the linear transformation of the concatenated features. LeakyReLU is a nonlinear activation function that performs further nonlinear transformations.

Selective State Space Model. The selective state space model is a type of state space sequence model (SSM) that maps a 1-dimensional sequence $x(t) \in \mathcal{R}$ to $y(t) \in \mathcal{R}$. The calculation of it can be represented as follows linear ordinary differential equation:

$$h'(t) = Ah(t) + Bx(t) \tag{4}$$

$$y(t) = Ch(t) \tag{5}$$

where $A \in \mathcal{R}^{N \times N}$ and $B, C \in \mathcal{R}^{N}$ are state matrices, and $h(t) \in \mathcal{R}^{N}$ denotes the implicit latent state. Recently, structured state space sequence models have emerged as a promising class of sequence modeling architectures, demonstrating strong capabilities in various tasks [8]. In [8], Gu et al. designed a simple selection mechanism by parameterizing the SSM parameters based on the input, which enables the model to filter out irrelevant information and remember relevant information for a long time. To give the autoencoder a better ability to capture temporal dependencies, we apply the Mamba to it.

2.3 Framework of MTVAE-GM

To accurately perform anomaly detection in industrial processes, we propose a multi-task variational autoencoder MTVAE-GM. Figure 1 illustrates its framework, which shares the encoder and performs both time series reconstruction and forecasting.

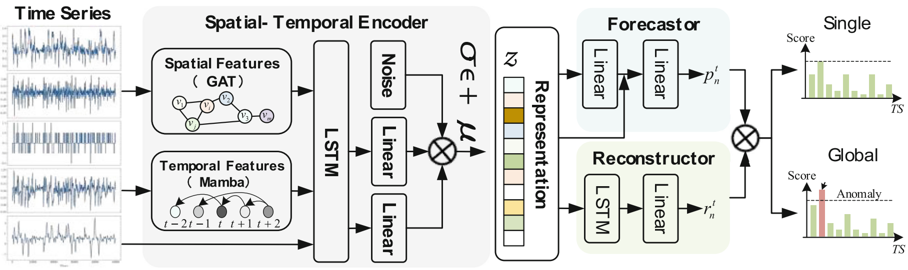

Fig. 1. The framework of MTVAE-GM.

Specifically, to capture temporal and spatial dependencies of multivariate time series, the selective state space model (Mamba) [8] and the graph attention network (GAT) [11] are used for low-dimensional representations. Then, a multi-layer LSTM is applied to further capture the temporal dependencies. In addition, we incorporate the architecture of variational autoencoders to enhance the model's tolerance to noise. Through multi-task learning, MTVAE-GM can learn a more robust representation of time series. Finally, a composite objective function with weight coefficients is proposed to guide the model for end-to-end training.

2.4 Data Preprocessing

Data Normalization. We first normalize the original time seriesto help the model learn the patterns of time series quickly and accurately. To normalize the data, we calculate each variable's minimum and maximum values in the training set. The normalized time series $\tilde{x}$ is calculated as follows:

$$\tilde{x} = \frac{x - min(x)}{max(x) - min(x)} \tag{6}$$

When normalizing the test set, each variable's maximum and minimum values in the training set are calculated. Correspondingly, the final anomaly score will be restricted to a relative range, facilitating anomaly detection.

Causal Graph Construction. A transfer entropy-based approach is utilized to create a causal graph in GAT. To construct the causal graph of industrial process variables, we need to evaluate the transfer entropy for each variable pair. To do this, we determine the value range for each variable and divide it into 100 segments of equal length. Then, we use the segment where the value falls as the discretized value. Let X and Y denote the time series of the two variables after discretization, and then the transfer entropy from X to Y is calculated as follows:

$$T_{X \to Y} = H(Y_t, Y_{t-1:t-L}) - H(Y_t | Y_{t-1:t-L}, X_{t-1:t-L}) \tag{7}$$

where $H(X)$ is Shannon entropy of X. Finally, if the transfer entropy exceeds a certain threshold, we set an edge between the two variables. It will be used as input to the subsequent graph attention network to provide causal relationships.

2.5 Details of MTVAE-GM

In our multi-task variational autoencoder, encoder and decoder are the main components. The shared encoder can learn richer features to improve the robustness of anomaly detection under noisy conditions through reconstruction and forecasting tasks. Correspondingly, two decoders utilize the low-dimensional representation to reconstruct and forecast accurately. Next, we will present the details of MTVAE-GM.

Encoder: In the encoder, Mamba and GAT are utilized to capture temporal and spatial dependencies from multivariate time series. Let $x \in \mathcal{R}^{n \times m}$ denotes the input time series, and then outputs are calculated as follows:

$$h_{temporal} = Mamba(x) \tag{8}$$

$$h_{spatial} = GAT(x) \tag{9}$$

where n denotes the length of time series and m denotes the number of variables. After capturing temporal and spatial dependencies, the dimensions of $h_{temporal} \in R^{n \times m}$ and $h_{spatial} \in R^{n \times m}$ remain consistent. Further, we combine the features with the input x to obtain richer features to represent the time series as follows:

$$h = x \oplus h_{temporal} \oplus h_{spatial} \tag{10}$$

where $\oplus$ denotes the concatenation of representations and $h \in R^{n \times 3m}$. Next, multi-layer LSTM is used to further capture the temporal dependencies and progressively compress the inputs to finally obtain the last hidden state of the LSTM as a low-dimensional representation of the time series as follows:

$$h_{last} = LSTM(h) \tag{11}$$

where $h_{last} \in \mathcal{R}^c$ and c is the length of h_{last}. To improve the tolerance of the model to noise, we apply the architecture of the variational autoencoder to it. Two linear layers are employed to mimic the computation of mean and standard deviation as follows:

$$\mu = h_{last} W_{\mu}^T + b_{\mu} \tag{12}$$

$$\sigma = h_{last} W_{\sigma}^T + b_{\sigma} \tag{13}$$

$$z = \sigma \cdot \epsilon + \mu \tag{14}$$

where $W_\mu \in \mathcal{R}^{d\times c}$, $b_\mu \in \mathcal{R}^d$, $W_\sigma \in \mathcal{R}^{d\times c}$, and $b_\sigma \in \mathcal{R}^d$ denote the weights and biases of the two linear layers, and $\epsilon \in \mathcal{R}^d$ represents random noise with a mean of 0 and a standard deviation of 1. Finally, the encoder's output $z \in \mathcal{R}^d$ is passed to the decoder as a representation of the time series.

Decoder. The decoder is divided into the forecasting decoder and the reconstruction decoder. The forecasting decoder is a multi-layer perceptron with residual connections. It consists of multiple linear layers connected in series and connected by residuals to accelerate model convergence and prevent overfitting. The activation function GELU is set between the linear layers to realize the nonlinear transformation. In addition, to prevent overfitting, dropout is also added. Ultimately, the forecasting decoder predicts the next step's value for each variable as follows:

$$y = zW_1^T + b_1 \tag{15}$$

$$a = Dropout(GELU(y)) \tag{16}$$

$$p = aW_2^T + b_2 + zW_3^T + b_3 \tag{17}$$

where $W_1 \in \mathcal{R}^{e\times d}$, $b_1 \in \mathcal{R}^e$, $W_2 \in \mathcal{R}^{m\times e}$, $b_2 \in \mathcal{R}^m$, $W_3 \in \mathcal{R}^{m\times d}$, and $b_3 \in \mathcal{R}^m$ denote the weights and biases of the two linear layers and $p \in R^m$ is the output. The reconstruction decoder consists of a multi-layer LSTM connected in series, ultimately linking a linear layer to perform the dimensional transformation. First, we repeat z in the time dimension n times to obtain $z\prime \in \mathcal{R}^{n\times d}$. The calculation process is as follows:

$$o = LSTM\left(z'\right) \tag{18}$$

$$r = oW_o^T + b_o \tag{19}$$

where $o \in \mathcal{R}^{n\times f}$, $W_o \in \mathcal{R}^{m\times f}$, $b_o \in \mathcal{R}^m$, and $r \in \mathcal{R}^{n\times m}$ is the output.

2.6 Joint Optimization

To enable MTVAE-GM to learn a robust representation of time series, we propose a composite objective function for end-to-end training. It comprises three components:

the reconstruction loss, the prediction loss, and the Kullback-Leibler divergence loss. Let $V = \{v_0, v_1, \ldots, v_n\}$ denote the original time series, the prediction is $P = \{p_n\}$, and the reconstruction $R = \{r_1, r_2, \ldots, r_n\}$. The objective function is calculated as follows:

$$L_{rec} = \frac{1}{nm} \sum_{i=1}^{n} \sum_{j=1}^{m} \left| v_i^j - r_i^j \right| \tag{20}$$

$$L_{pre} = \frac{1}{m} \sum_{j=1}^{m} \left| v_n^j - p_n^j \right| \tag{21}$$

$$L_{kld} = -\frac{1}{2} \sum_{i=1}^{d} \left(1 + log(\sigma_i) - \mu_i^2 - \sigma_i\right) \tag{22}$$

$$L_{all} = \alpha \cdot L_{rec} + \beta \cdot L_{pre} + \gamma \cdot L_{kld} \tag{23}$$

When the model is fed with normal data, its corresponding reconstruction and prediction losses will be small. In contrast, when the model is fed with abnormal data, it reveals a significant error in reconstruction and prediction because spatial-temporal dependencies do not conform to the normal pattern. Moreover, the Kullback-Leibler divergence loss is directly affects the model's adaptation to noise. MTVAE-GM's objective function optimizes forecasting and reconstruction tasks by summing three loss terms with weighting coefficients that determine the model's focus on a particular task.

2.7 Inference

Anomaly Score Calculation. In this method, time series with anomaly scores exceeding a threshold are treated as an anomaly. The MTVAE-GM performs time series reconstruction and forecasting simultaneously, so we include both reconstruction and prediction for the time series in calculating anomaly scores. Let $V = \{v_0, v_1, \ldots, v_n\}$ denote the original time series, $P = \{p_n\}$ denote the prediction, and $R = \{r_1, r_2, \ldots, r_n\}$ denote the reconstruction, then the anomaly score S is calculated as follows:

$$S = \frac{1}{m} \sum_{i=1}^{m} \left| v_n^i - r_n^i \right| + \frac{1}{m} \sum_{i=1}^{m} \left| v_n^i - p_n^i \right| \tag{24}$$

Anomaly Detection. The autoencoder learns the patterns of variable changes under normal conditions. Therefore, abnormal inputs will result in a higher anomaly score, which allows for the identification of anomalies. Identifying all the points with anomaly scores greater than a given threshold makes it possible to discern all possible anomalies. Let $SEG = \{seg_1, seg_2, \ldots, seg_n\}$ denote all time series segments and $S = \{s_1, s_2, \ldots, s_n\}$ denote the anomaly scores corresponding to each segment, and γ denotes the detection threshold, then the anomaly detection is as follows:

$$r(s_i) = \begin{cases} 0, s_i < \gamma \\ 1, s_i \geq \gamma \end{cases} \tag{25}$$

If the value of $r(s_i)$ is 1, the time series segment seg_i is identified as an anomaly. On the other hand, if the value of $r(s_i)$ is 0, the time series seg_i is considered normal.

3 Experiments

3.1 Datasets and Metrics

Datasets. We employ three datasets to assess the performance of our model in anomaly detection tasks, namely CCPAD (Catalytic Cracking Process Anomaly Detection), MSL (Mars Science Laboratory rover) [12], and SWaT (Secure Water Treatment) [13]. The CCPAD dataset is a multivariate time series anomaly detection dataset constructed by the simulation system of the catalytic cracking process based on SUPCON VxOTS.

Metrics. The precision, recall, and F_1 scores are utilized as metrics to evaluate the performance of different models. They are calculated as follows:

$$Precision = \frac{TP}{TP+FP} \tag{26}$$

$$Recall = \frac{TP}{TP+FN} \tag{27}$$

$$F_1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{28}$$

where *TP* is the number of abnormal samples identified, *FP* is the number of samples identified as abnormal that are normal, *FN* is the number of abnormal samples not identified, and *TN* is the number of samples identified as normal. These evaluation metrics range from 0 to 1, with higher values indicating better performance. Among them, the F_1 score indicates the overall performance.

3.2 Setup

To evaluate the performance of the proposed MTVAE-GM, we select LSTM-NDT [12], MTAD-GAT [6], USAD [14], OmniAnomaly [3], DAGMM [15] and TranAD [16] as baseline for multivariate time series anomaly detection. We use the AdamW optimizer with $betas$ = (0.9,0.98) to train our model for 20 epochs with an initial learning rate of 0.001, and a learning rate decay strategy with $lr\_decay = 0.99$ is employed. For all models, we set the sliding window size to 50. We employ the same POT (Peaks-Over-Threshold) threshold selection method as described in [17] to choose the threshold for computing the evaluation metrics.

3.3 Performance Comparison

As shown in Table 1, MTVAE-GM outperforms six baselines on three datasets, achieving the highest F1 scores of 0.9389, 0.8151, and 0.9524. The best performance on the CCPAD, SWaT, and MSL datasets improved by 2.80%, 0.08%, and 0.29%. It can be observed that MTVAE-GM exhibits high recall on all three datasets, which shows that it can detect all real anomalies as much as possible. In addition, the MTVAE-GM also shows a high precision, which indicates that only a small percentage of normal data is recognized as anomalies. These features align closely with the demands of industrial anomaly detection, guaranteeing that every irregular occurrence is identified.

Table 1. Performance of different models on three datasets.

Method	CCPAD			SWaT			MSL		
	P	R	F1	P	R	F1	P	R	F1
LSTM-NDT	0.7429	0.7891	0.7653	0.9655	0.6957	0.8087	0.6288	1.0000	0.7721
USAD	0.8712	0.9042	0.8874	0.9977	0.6879	0.8143	0.7949	1.0000	0.8857
OmniAnomaly	0.8853	0.8973	0.8912	0.9782	0.6957	0.8131	0.7848	1.0000	0.8794
DAGMM	0.8809	0.8328	0.8562	0.9955	0.6879	0.8136	0.7029	1.0000	0.8256
TranAD	0.7019	1.0000	0.8248	0.9977	0.6879	0.8143	0.9038	1.0000	0.9495
MTAD-GAT	0.8367	1.0000	0.9109	0.9888	0.6879	0.8114	0.9931	0.7024	0.8228
MTVAE-GM	0.8848	1.0000	**0.9389**	1.0000	0.6879	**0.8151**	0.9091	1.0000	**0.9524**

It is noteworthy that there are no complex spatial dependencies between their variables in the SWaT and MSL datasets. In that case, the temporal dependencies are only needed to be considered. Therefore, these methods before MTAD-GAT performed well on the two datasets. However, spatial dependencies are equally crucial for multivariate time series anomaly detection. It leads to poor performances on the CCPAD dataset with complex spatio-temporal dependencies. Therefore, the last two models consider both temporal and spatial dependencies. Thus MTAD-GAT also achieves a good performance on the CCPAD dataset. Specifically, MTVAE-GM consistently outperforms OmniAnomaly across all three datasets, with F1 scores increasing by 4.77% and 7.30% in CCPAD and MSL.

In addition, among the methods targeting temporal and spatial feature extraction, graph attention networks are utilized to capture spatial dependencies. However, the transfer entropy is employed by MTVAE-GM to integrate causal information among different variables. It enables the model to integrate spatial causal information between variables. Compared to MTAD-GAT, we apply the novel selection state space model Mamba and variational autoencoder architecture to achieve a robust representation of time series, yielding superior results. In contrast, F1 scores improved by 2.80%, 0.37%, and 12.96% on three datasets, indicating significant enhancements. Meanwhile, multi-task learning has also brought significant benefits to the model.

3.4 Ablation Study

To further explore the effectiveness of each module in MTVAE-GM, we performed ablation experiments on the CCPAD dataset as Table 2. As seen from the results, all three modules show essential contributions. Compared to models not utilizing Mamba, those incorporating Mamba demonstrate respective increases of 0.53%, 1.17%, and 1.72% in F1 scores. Furthermore, models capturing spatial dependencies through GAT exhibit respective increases of 0.66%, 1.30%, and 2.73% in F1 scores. Moreover, adopting VAE architecture leads to an average enhancement of 0.74% in F1 scores. Our model seamlessly integrates these modules, ultimately achieving outstanding performance in anomaly detection.

Table 2. Performance of the ablation experiments on CCPAD dataset.

Mamba	GAT	VAE	Precision	Recall	F1
×	×	×	0.8278	1.0000	0.9058
×	√	×	0.8389	1.0000	0.9124
√	×	×	0.8367	1.0000	0.9111
√	√	×	0.8589	1.0000	0.9241
×	×	√	0.8362	1.0000	0.9108
×	√	√	0.8547	1.0000	0.9217
√	×	√	0.8376	1.0000	0.9116
√	√	√	0.8848	1.0000	0.9389

3.5 Case Study

Interpretable analysis of results is crucial for anomaly detection in catalytic cracking processes, as it empowers safety managers to address anomalies promptly. As shown in Fig. 2, we exemplify the interpretability of anomaly detection results using the outlet temperature of the primary riser as a case study. The model effectively learns the patterns of normal data, enabling it to generate predictions and reconstructions with significant errors for abnormal data. Anomaly scores are significantly higher during abnormal events, indicating a substantial deviation from normal data. Therefore, assigning appropriate thresholds to each variable can also achieve variable-level anomaly detection. Furthermore, integrating the causal relationships among variables can infer the root cause of the abnormal event. It indicates that the model can provide a highly interpretable analysis of results for on-site safety managers, demonstrating strong usability.

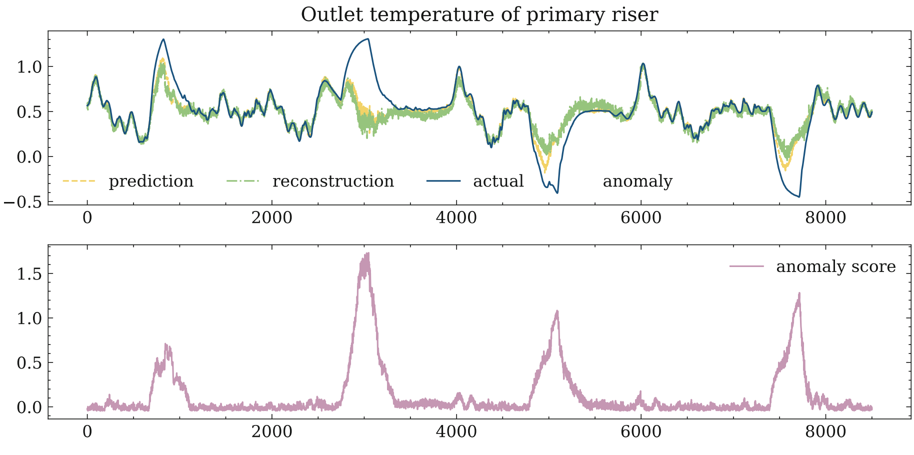

Fig. 2. Interpretability analysis of anomaly detection results.

4 Conclusion

This paper proposes a novel multi-task variational autoencoder MTVAE-GM based on GAT and Mamba for multivariate time series anomaly detection in the industrial process. By leveraging strong capturing capabilities of temporal and spatial dependencies, the multi-task learning paradigm, and variational autoencoder architecture, our model outperforms other baselines on three datasets. Furthermore, our model offers high interpretability in the analysis of results, enabling on-site safety managers to swiftly diagnose the root causes of anomalies, thereby mitigating severe accidents. In the future, we will investigate the integration of textual knowledge pertaining to the industrial process into our model for applications in safety protection, such as anomaly detection and root cause analysis.

Acknowledgment. This work is supported by the National Natural Science Foundation (NNSF) of China under Grants 62127808.

References

1. Yang, F., Xu, M., Lei, W., Lv, J.: Artificial intelligence methods applied to catalytic cracking processes. Big Data Mining Analy. **6**, 361–380 (2023)
2. Wu, D., Jiang, Z., Xie, X., Wei, X., Yu, W., Li, R.: LSTM learning with Bayesian and Gaussian processing for anomaly detection in industrial IoT. IEEE Trans. Ind. Inf. **16**, 5244–5253 (2019)
3. Su, Y., Zhao, Y., Niu, C., Liu, R., Sun, W., Pei, D.: 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, pp. 2828–2837 (2019)
4. Wang, Y., Du, X., Lu, Z., Duan, Q., Wu, J.: Improved LSTM-based time-series anomaly detection in rail transit operation environments. IEEE Trans. Ind. Inf. **18**, 9027–9036 (2022)
5. Wei, Y., Jang-Jaccard, J., Xu, W., Sabrina, F., Camtepe, S., Boulic, M.: LSTM-autoencoder-based anomaly detection for indoor air quality time-series data. IEEE Sens. J. **23**, 3787–3800 (2023)
6. Zhao, H., et al.: Multivariate time-series anomaly detection via graph attention network. In: 2020 IEEE International Conference on Data Mining (ICDM), pp. 841–850. IEEE (2020)
7. Lea, C., Flynn, M.D., Vidal, R., Reiter, A., Hager, G.D.: Temporal convolutional networks for action segmentation and detection. In: proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 156–165 (2017)
8. Gu, A., Dao, T.: Mamba: Linear-time sequence modeling with selective state spaces. arXiv preprint arXiv:2312.00752 (2023)
9. Deng, A., Hooi, B.: Graph neural network-based anomaly detection in multivariate time series. In: Proceedings of the AAAI Conference on Artificial Intelligence, pp. 4027–4035 (2021)
10. Chen, Z., Chen, D., Zhang, X., Yuan, Z., Cheng, X.: Learning graph structures with transformer for multivariate time-series anomaly detection in IoT. IEEE Internet Things J. **9**, 9179–9189 (2021)
11. Velickovic, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., Bengio, Y.: Graph attention networks. Stat **1050**, 10–48550 (2017)

12. Hundman, K., Constantinou, V., Laporte, C., Colwell, I., Soderstrom, T.: Detecting spacecraft anomalies using LSTMS and nonparametric dynamic thresholding. In: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, pp. 387–395 (2018)
13. Mathur, A.P., Tippenhauer, N.O.: SWaT: a water treatment testbed for research and training on ICS security. In: 2016 International Workshop on Cyber-Physical Systems for Smart Water Networks (CySWater), pp. 31–36. IEEE (2016)
14. Audibert, J., Michiardi, P., Guyard, F., Marti, S., Zuluaga, M.A.: USAD: unsupervised anomaly detection on multivariate time series. In: Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, pp. 3395–3404 (2020)
15. Zong, B., et al.: Deep autoencoding Gaussian mixture model for unsupervised anomaly detection. In: International Conference on Learning Representations (2018)
16. Tuli, S., Casale, G., Jennings, N.R.: Tranad: Deep transformer networks for anomaly detection in multivariate time series data. arXiv preprint arXiv:2201.07284 (2022)
17. Siffer, A., Fouque, P.-A., Termier, A., Largouet, C.: Anomaly detection in streams with extreme value theory. In: Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 1067–1075 (2017)