Journal of Computer Technology and Software

ISSN:2998-2383

Vol. 4, No. 8, 2025

Research on a Temporal Semantic-Enhanced Transformer-Based Anomaly Awareness Method for Complex Enterprise Data Pipelines

Yingzi Wang[1*]

[1]University of Minnesota, Twin Cities, Minneapolis, USA

*Corresponding author: Yingzi Wang; meetyz.wang@gmail.com

Abstract: Complex enterprise data pipelines are influenced by multiple factors during operation, including changes in data sources, task scheduling, field structures, resource status, and upstream/downstream dependencies. Anomalies often exhibit temporal continuity, semantic relevance, and link propagation. To address the challenge of traditional anomaly detection methods simultaneously characterizing changes in operational state and business semantic constraints, this paper proposes a temporal semantic enhancement Transformer anomaly perception method for complex enterprise data pipelines. This method first jointly encodes multi-source operational observations and node semantic attributes, mapping indicator fluctuations, task status, data quality information, and semantic metadata to a unified representation space to enhance the completeness of pipeline node state representation. Subsequently, temporal position encoding introduces the sequential information of the operational window, enabling the model to perceive the changes in pipeline state over continuous time. Furthermore, the Transformer structure is used to model long-distance temporal dependencies, and a dependency attention mechanism is combined to characterize the relationships between different pipeline nodes, thereby improving the model's ability to identify cross-node anomaly propagation and contextual anomaly patterns. Based on this, a semantic gating enhancement module adaptively adjusts temporal features according to node semantic attributes, ensuring that anomaly judgment considers both numerical changes and business implications. Comparative experimental results show that the proposed method achieves better performance in metrics such as Acc, Pre, Rec, and F1, indicating that the constructed temporal semantic joint modeling framework can effectively improve the accuracy and stability of anomaly detection in complex enterprise data pipelines.

Keywords: Enterprise data pipeline; anomaly detection; temporal semantic enhancement Transformer; semantic gating.

1. Introduction

As enterprises continue their digital transformation, data has become a crucial foundation for business operations, intelligent decision-making, and value creation. Internal data sources within enterprises are increasingly diverse, encompassing multiple stages including business systems, log platforms, transaction systems, customer behavior systems, IoT devices, and third-party service interfaces [1, 2]. Data forms complex data pipelines during collection, cleaning, transformation, loading, synchronization, and analysis. The stable operation of these data pipelines directly impacts the availability of data assets, the continuity of business processes, and the reliability of upper-layer intelligent applications. Issues such as delays, data loss, duplication, formatting errors, field drift, or task failures in the data pipeline can lead to distorted downstream analysis results, ineffective business monitoring, and biased enterprise decision-making. Therefore, research on anomaly detection for complex enterprise data pipelines is of significant practical importance.

Compared to traditional data processing workflows, modern enterprise data pipelines are characterized by long links, numerous nodes, complex dependencies, heterogeneous data types, and dynamically changing operational states [3]. Anomalies are rarely isolated events but are influenced by the combined effects of temporal context, task dependencies, changes in data semantics, and fluctuations in the business cycle. For example, delays in the same type of data may have different meanings at different business stages, and changes in the same field may have different impacts in different upstream and downstream tasks. Methods relying solely on single-indicator thresholds or static rules struggle to accurately identify hidden anomalies in complex environments and fail to characterize the propagation process of anomalies within the pipeline. Therefore, understanding the data pipeline state from both temporal variation and semantic association perspectives has become a key issue in enterprise data governance and intelligent operations and maintenance [4].

The Transformer model possesses strong advantages in sequence modeling and context representation, capable of capturing long-distance dependencies through attention mechanisms, providing a new modeling approach for temporal anomaly perception in complex data pipelines [5]. The large amounts of time-series logs, task states, resource indicators, and data quality indicators generated during the operation of enterprise data pipelines naturally exhibit continuous evolution characteristics. Simultaneously, pipeline nodes, field meanings, task types, business scenarios, and upstream and downstream dependencies contain rich semantic information. Jointly

modeling temporal features and semantic information helps enhance the model's overall understanding of the pipeline's operational state, enabling anomaly detection to move beyond numerical fluctuations and further focus on the underlying business meaning, structural dependencies, and propagation paths.

Research on temporal semantic enhancement of Transformer anomaly perception for complex enterprise data pipelines can provide crucial support for the intelligent operation and maintenance of enterprise data platforms. On the one hand, this research helps improve the timeliness and accuracy of anomaly identification in data pipelines, reduces the cost of manual investigation, and minimizes the impact of data quality issues on business and analytics systems. On the other hand, this research can drive the transformation of anomaly detection methods from passive alerts to proactive perception, enabling data platforms to possess stronger state understanding, risk warning, and adaptive governance capabilities. As enterprise data scale continues to expand and data links become increasingly complex, building intelligent methods with time-series modeling, semantic understanding, and anomaly perception capabilities is of great significance for ensuring the stable operation of data infrastructure, improving enterprise data governance, and supporting the construction of intelligent decision-making systems.

2. BackGround

Enterprise data platforms typically handle various task types during long-term operation, including batch processing, stream processing, real-time synchronization, offline computing, and cross-system data exchange. These tasks differ significantly in scheduling cycles, input/output formats, resource consumption, dependencies, and business priorities, causing data pipelines to evolve from linear processing flows into complex systems with multi-level, multi-node, and multi-state collaborative operation. In this process, data pipelines not only need to handle data transmission and format conversion but also ensure data consistency, integrity, timeliness, and traceability [6]. Due to continuous changes in the enterprise business environment, issues such as data source structure adjustments, interface rule changes, field meaning shifts, task scheduling conflicts, and fluctuations in computing resources can all affect pipeline operation, increasing the risk of anomalies arising and spreading.

Existing enterprise data pipeline anomaly monitoring typically relies on manually configured rules, fixed thresholds, log retrieval, or single-indicator statistical analysis. While these methods are usable in simple scenarios, they tend to lack adaptability in complex pipeline environments. Enterprise data pipeline anomalies often exhibit gradual, correlated, and context-dependent characteristics. Changes in values at a single point in time cannot fully reflect the anomaly formation process, and the anomaly behavior of local nodes may be influenced by changes in upstream and downstream task states and business semantics. Therefore, relying solely on independent indicators or local logs is insufficient to meet the needs of complex enterprise data platforms for intelligent monitoring, dynamic identification, and risk warning [7]. Developing anomaly detection methods that integrate temporal evolution patterns with semantic contextual information has become an important research direction for improving the reliability of enterprise data pipelines and data governance capabilities.

Recent intelligent risk and decision models further support the need to connect pipeline anomaly perception with downstream governance. Heterogeneous graph learning for supply chain risk identification demonstrates that multi-entity interaction modeling can capture cross-node risk propagation and structural dependencies in complex business systems [8]. Adaptive stock-trading algorithms jointly driven by multi-agent collaborative decision making and large language models also show how predictive signals can be transformed into dynamic decision policies under uncertain and rapidly changing environments [9]. For enterprise data pipelines, these ideas suggest that anomaly awareness should not stop at detecting abnormal metrics, but should further support interpretable risk localization, cross-link dependency analysis, and adaptive operational decision-making.

3. Methodology

Anomaly awareness in complex enterprise data pipelines depends not only on local fluctuations in operational metrics but also on the joint variation relationships among task links, field semantics, scheduling states, and upstream-downstream dependencies. To avoid simplifying pipeline states into numerical judgments at isolated time points, this paper represents multi-source observations generated during the continuous operation of enterprise data pipelines as temporal semantic state sequences, enabling the model to describe the intrinsic connections among data quality, task execution, resource consumption, and business context within a unified representation space. Considering that different pipeline nodes vary in task type, data granularity, and business meaning, the original observations are first mapped into a high-dimensional representation composed of numerical features, status features, and semantic features, thereby providing a more complete contextual basis for subsequent anomaly awareness. The overall algorithm architecture is presented in this paper, as shown in Figure 1.

Let $\boldsymbol{x}_{t,i}$ denote the raw observation of the i-th pipeline node in the t-th time window, and let $\mathbf{s}_i$ denote its semantic attribute embedding. The fused node state representation is defined as:

$$\boldsymbol{h}_{t,i}^{0} = \phi\left(\boldsymbol{W}_x \boldsymbol{x}_{t,i} + \boldsymbol{W}_s \boldsymbol{s}_i + \boldsymbol{b}_0\right)$$

This representation aims to encode measurable operational variations and business semantics that cannot be directly expressed by numerical metrics into a unified space, where $\phi(\cdot)$ denotes the nonlinear mapping function, $\boldsymbol{W}_x$ and $\boldsymbol{W}_s$ are used to project operational observations and semantic attributes, respectively, and $\boldsymbol{b}_0$ adjusts the bias structure in the representation space. Furthermore, the pipeline state is not determined by a single node alone, but is jointly affected by the collaborative execution relationships among multiple task nodes within the same time window. Therefore, organizing all node states into a global pipeline representation can enhance the overall perception capability of the model for complex link

structures. The time-window state matrix constructed based on node-level representations is expressed as:

$$\boldsymbol{H}_t^0 = \left[\boldsymbol{h}_{t,1}^0; \boldsymbol{h}_{t,2}^0; \cdots; \boldsymbol{h}_{t,N}^0\right]$$

This matrix represents the enterprise data pipeline in each time window as a unified structure containing N node states, where each row corresponds to the comprehensive state of a pipeline node, thereby providing structured input for subsequent temporal modeling.

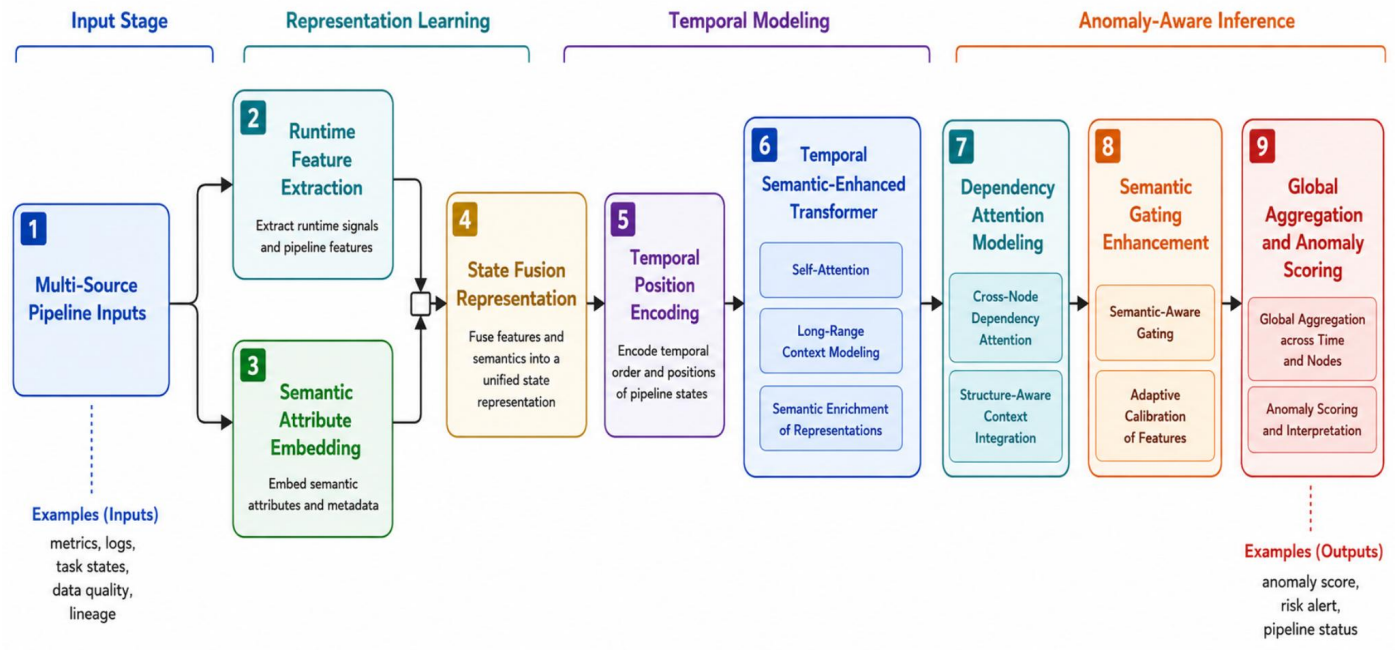

Figure 1. Overall Algorithm Framework

Long-running data pipelines usually exhibit obvious periodicity, latency, and accumulation effects, and abnormal states may gradually emerge across multiple time windows rather than appearing suddenly at a single moment. In response to this characteristic, this paper introduces temporal positional encoding on top of the node state representation, allowing the model to distinguish the actual meanings of the same semantic state at different operational stages and capture the dynamic evolution process from latent disturbances to explicit failures. Let $\boldsymbol{p}_t$ denote the temporal encoding corresponding to the t-th time window. The enhanced input sequence representation is formulated as:

$$\boldsymbol{Z}_t^0 = \boldsymbol{H}_t^0 + \mathbf{1}_N \boldsymbol{p}_t^\top$$

This design enables temporal order information to directly participate in state modeling, where $\mathbf{1}_N$ is used to extend the temporal encoding to all pipeline nodes, and $\boldsymbol{p}_t$ has the same hidden-space dimension as the node states, ensuring that temporal context can be naturally integrated with pipeline semantic representations. Subsequently, the temporal semantic-enhanced Transformer models the dependencies among different time windows and different pipeline nodes through a multi-head self-attention mechanism, enabling the anomaly awareness process to focus not only on the current state but also to trace historical patterns and identify potential propagation paths. For the m-th attention head in the l-th layer, the query, key, and value are obtained through linear transformations of the previous-layer representation, and the attention output is defined as:

$$\boldsymbol{A}_m^l = softmax\left(\frac{\boldsymbol{Q}_m^l \left(\boldsymbol{K}_m^l\right)^\top}{\sqrt{d}}\right)\boldsymbol{V}_m^l$$

This computation adaptively assigns attention weights according to the correlations among states, where d denotes the hidden-space dimension, and $\boldsymbol{Q}_m^l$, $\boldsymbol{K}_m^l$, and $\boldsymbol{V}_m^l$ carry the information of the state to be matched, the reference state, and the state to be aggregated, respectively, thereby enhancing the model's ability to represent long-range dependencies and cross-node anomaly correlations. The architecture of the multi-head attention mechanism is shown in Figure 2.

Semantic information plays a key role in anomaly awareness for enterprise data pipelines, because the same numerical fluctuation may correspond to completely different risk levels under different task types, field meanings, and business scenarios. To prevent the model from making mechanical judgments based solely on numerical patterns, this paper further constructs a semantic gating mechanism, allowing semantic context to dynamically regulate the proportion of effective information in temporal representations. Specifically, the semantic gating weight is generated jointly from the current temporal hidden state and the node semantic representation, and is used to control the retention intensity of anomaly-related features. This process can be expressed as:

$$\boldsymbol{g}_{t,i} = \sigma\left(\boldsymbol{W}_g\left[\boldsymbol{z}_{t,i}^{l}; \boldsymbol{s}_i\right] + \boldsymbol{b}_g\right)$$

The semantic gate constrains the formation process of anomaly representations through business context, where $\sigma(\cdot)$ denotes the Sigmoid activation function, $\boldsymbol{z}_{t,i}^{l}$ represents the hidden state of node i in time window t at the l-th layer, $\boldsymbol{s}_i$ provides the semantic prior of the node, and $\boldsymbol{W}_g$ and $\boldsymbol{b}_g$ are used to learn the semantic modulation relationship. After gating modulation, the model obtains the semantic-enhanced anomaly-aware representation, which is calculated as:

$$\widetilde{\boldsymbol{z}}_{t,i}^{l} = \boldsymbol{g}_{t,i} \odot \boldsymbol{z}_{t,i}^{l} + \left(1 - \boldsymbol{g}_{t,i}\right) \odot \boldsymbol{s}_i$$

This fusion strategy establishes an adaptive balance between temporal states and semantic priors, where $\odot$ denotes element-wise multiplication. A higher gating weight emphasizes the variation information of the current operational state, whereas a lower gating weight strengthens the constraint imposed by the inherent semantics of the node on state judgment. Unlike methods that rely only on log rules or fixed thresholds, this mechanism enables the model to identify complex anomalies such as field drift, abrupt task behavior changes, and upstream-downstream semantic mismatches from the perspective of semantic consistency.

The final anomaly awareness stage needs to transform multi-layer temporal semantic representations into anomaly scores for risk judgment while preserving the representation capability for the overall pipeline state. Based on the output of the last Transformer layer, this paper adopts a global aggregation strategy to compress the hidden states across different time windows and different nodes, thereby obtaining a comprehensive representation that reflects the current operational risk of the pipeline. This comprehensive representation is defined as:

$$\boldsymbol{u}_t = Pool\left(\widetilde{\boldsymbol{Z}}_t^{L}\right)$$

The purpose of the aggregation operation is to preserve the dominant anomaly patterns in the pipeline state, where $\widetilde{\boldsymbol{Z}}_t^{L}$ denotes the semantic-enhanced state matrix output by the L-th layer, and $\mathrm{Pool}(\cdot)$ is used to extract global contextual features from the node dimension. Subsequently, the anomaly score is generated by a nonlinear mapping function, enabling the model to convert complex hidden-space states into a continuous risk measure and thus support fine-grained characterization of different anomaly degrees. The anomaly awareness function is expressed as:

$$\alpha_t = \sigma\left(\boldsymbol{w}_a^{\top} ReLU\left(\boldsymbol{W}_a \boldsymbol{u}_t + \boldsymbol{b}_a\right)\right)$$

This score reflects the probability of an anomaly occurring in the data pipeline within the t-th time window, where $\boldsymbol{W}_a$, $\boldsymbol{w}_a$, and $\boldsymbol{b}_a$ are used to learn the risk mapping relationship, and $\mathrm{ReLU}(\cdot)$ enhances the nonlinear representation capability of the anomaly decision boundary. Through the above modeling process, operational states, temporal dependencies, and semantic relationships in complex enterprise data pipelines are unified within the anomaly awareness framework, allowing the model to understand the anomaly formation mechanism from three perspectives, namely local fluctuations, cross-node correlations, and business semantic consistency, while providing methodological support for proactive monitoring and intelligent governance in enterprise data platforms.

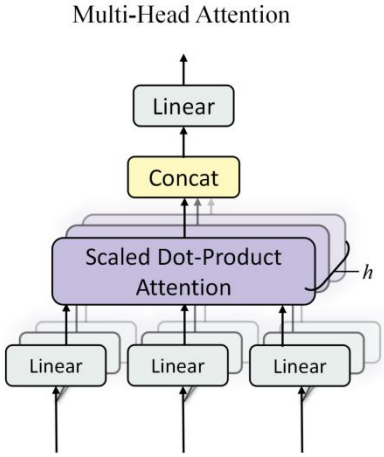

Figure 2. Multi-head attention mechanism architecture

4. Experimental Results and Analysis

4.1 Dataset

This paper selects the 2018 AIOps KPI Anomaly Detection dataset as the open-source benchmark dataset for anomaly detection tasks in complex enterprise data pipelines. This dataset is publicly released in the NetManAIOps KPI-Anomaly-Detection repository, which contains two data parts: Preliminary_dataset and Finals_dataset, and provides data files such as phase2_train.csv.zip and phase2_ground_truth.hdf.zip. It is designed for timing anomaly detection tasks of key performance indicators in AIOps scenarios, reflecting the state fluctuations and anomaly risks of enterprise-level service systems during continuous operation. Each sample in the dataset is based on information such as timestamp, KPI identifier, observation value, and anomaly label, encompassing both continuous time-series operational changes and characterizing the state differences of multi-source pipeline nodes through different KPI sequences. Therefore, it is suitable for simulating problems such as task latency, indicator mutations, service fluctuations, and data quality anomalies in enterprise data pipelines. Based on the proposed temporal semantic enhancement Transformer anomaly detection method, KPI numerical sequences can be used as operational status inputs, KPI identifiers and their corresponding sequences can be used as semantic attribute information, and anomaly annotations can be used to characterize the risk status of pipeline operation. Thus, this dataset can better support temporal modeling, semantic enhancement, and anomaly detection research in complex enterprise data pipeline scenarios.

4.2 Experimental setup

The experiments were conducted in a unified hardware and software environment to ensure the stability and reproducibility of the model training and testing process. All input sequences were segmented according to fixed time windows, and numerical KPI observations were standardized. KPI identifiers

were also used as semantic attributes input to the model. During model training, the AdamW optimizer was used to update parameters, combined with weight decay and Dropout to suppress overfitting. To meet the temporal dependency modeling requirements of enterprise data pipeline anomaly detection tasks, a Transformer encoding layer was used to capture long-distance state changes, a semantic gating module was used to enhance the semantic distinguishability between different KPI sequences, and finally, the risk probability for each time window was output through an anomaly scoring layer. The experimental setup is shown in Table 1.

Table 1. Experimental setup.

Category	Setting
Operating System	Ubuntu 20.04
Programming Language	Python 3.10
Deep Learning Framework	PyTorch 2.1
CUDA Version	CUDA 11.8
GPU	NVIDIA A100 40GB
Optimizer	AdamW
Learning Rate	1e-4
Weight Decay	1e-4
Batch Size	32
Training Epochs	100
Sequence Length	96
Hidden Dimension	128
Attention Heads	4
Transformer Layers	4
Dropout Rate	0.1

4.3 Experimental Results and Analysis

To ensure consistency between the comparison objects and enterprise data pipeline anomaly detection tasks, this paper selects related works on multivariate time series anomaly detection, system operation status monitoring, cloud environment anomaly identification, and deep sequence modeling as baseline methods for comparison. The experimental results are shown in Table 2. These methods all focus on state shifts, context dependencies, and anomaly risk identification in continuously running data, providing a reference for anomaly detection performance analysis in complex enterprise data pipeline scenarios.

Table 2. Comparative results of related methods and the proposed method

Method	Acc	Pre	Rec	F1
Tuli et al. [10]	92.18	91.35	90.74	91.04
Audibert et al. [11]	90.62	89.91	89.35	89.63
Munir et al. [12]	87.46	86.92	85.77	86.34
Malhotra et al. [13]	86.31	85.68	84.92	85.30
Hundman et al. [14]	88.27	87.54	86.83	87.18
Bashar et al. [15]	89.13	88.40	87.76	88.08
Wielgosz et al. [16]	85.72	84.96	84.25	84.60
Ours	95.64	94.83	94.26	94.54

The results in Table 2 show that the proposed algorithm outperforms other algorithms in terms of Acc, Pre, Rec, and F1, indicating that the method can more accurately identify abnormal states during the operation of enterprise data pipelines and maintain a good balance between capturing abnormal samples and distinguishing normal samples. This result is mainly due to the joint modeling of pipeline state evolution, node semantic attributes, and cross-node dependencies by the temporal semantic enhancement Transformer. This allows the model to not only focus on local changes in operational metrics but also to judge abnormal risks by combining business semantics and contextual relationships, thereby improving the anomaly perception capability in complex enterprise data pipeline scenarios.

To further validate the necessity of the key structural design in our proposed method, Table 3 presents ablation experiments focusing on the core components of the temporal semantic enhancement Transformer anomaly detection framework. These experiments examine the impact of Temporal Position Encoding, Dependency Attention Modeling, and Semantic Gating Enhancement on the anomaly detection capabilities of complex enterprise data pipelines. All settings maintain the same data processing methods, training strategies, and evaluation criteria, ensuring that performance variations under different structural configurations more comprehensively reflect the roles of each module in temporal dependency modeling, cross-node relationship perception, and semantic calibration.

Table 3. Ablation test results

Ablation Setting	Acc	Pre	Rec	F1
w/o Temporal Position Encoding	92.84	91.97	91.42	91.69
w/o Dependency Attention Modeling	93.26	92.44	91.88	92.16
w/o Semantic Gating Enhancement	93.71	92.85	92.37	92.61
Ours	95.64	94.83	94.26	94.54

As shown in Table 3, the proposed algorithm achieves more stable anomaly detection performance under the complete structure, indicating that temporal position encoding, dependency attention modeling, and semantic gating enhancement all have a positive impact on model performance. Temporal position encoding helps the model understand the continuous evolution of the enterprise data pipeline state, dependency attention modeling enhances the expressive ability of the relationships between different pipeline nodes, and semantic gating enhancement further improves the model's ability to filter business semantics and anomaly features. The complete model can uniformly incorporate temporal changes, node dependencies, and semantic information into the anomaly judgment process, thus exhibiting better recognition ability and overall stability in complex enterprise data pipeline scenarios.

Figure 3 shows that the loss function variation indicates that the proposed algorithm maintains good optimization stability during training, with both training and validation losses gradually decreasing and entering a relatively stable convergence state in the later stages. The moderate fluctuations in the training curve reflect the model's ability to continuously adapt to different anomaly patterns and semantic relationships

during the temporal state learning process of complex enterprise data pipelines. As shown in Figure 3, the synchronous decrease in validation loss indicates that the model does not simply memorize training samples but possesses good generalization ability. These results further demonstrate that the temporal semantic enhancement Transformer can effectively integrate operational metrics, node semantics, and dependency information, thereby improving the representation learning quality and overall recognition performance in anomaly perception tasks of complex data pipelines.

Figure 3. Image showing how the loss function changes with the epoch

The confusion matrix results demonstrate that the proposed algorithm can accurately distinguish between normal and abnormal states, maintaining good stability in both types of samples. For normal samples, the model effectively identifies the regular operating patterns of enterprise data pipelines, reducing the misclassification of normal fluctuations as anomalies, as illustrated in Figure 4. For abnormal samples, the model also effectively captures risk-reflecting signals such as sudden changes in operating metrics, state transitions, and semantic inconsistencies. These results indicate that the temporally semantically enhanced Transformer, after fusing operational features, node semantics, and dependencies, can form more discriminative state representations, thereby improving the classification reliability of anomaly detection tasks in complex enterprise data pipelines.

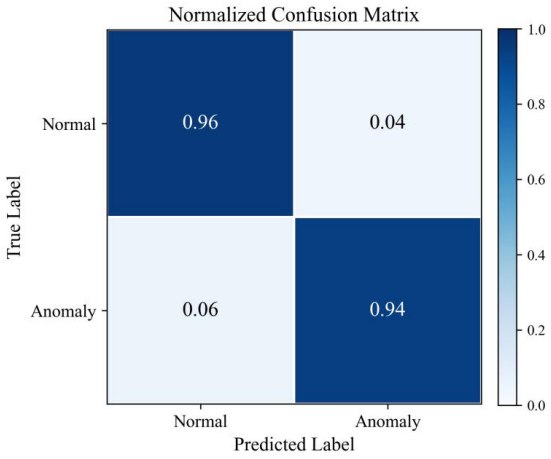

Figure 4. Confusion matrix experimental results

When the learning rate is low, the model parameters update slowly, resulting in insufficient learning of temporal state changes, node semantic relationships, and dependency structures within the enterprise data pipeline, thus limiting the overall recognition capability. As the learning rate gradually increases, the model can more effectively extract anomaly-related features and form a more stable state representation, as shown in Figure 5. However, further increases in the learning rate and excessively large parameter update amplitudes can weaken the stability of the optimization process and affect the model's fine-grained modeling of complex pipeline states. Overall, a moderate learning rate is more conducive to maintaining good convergence and classification reliability of the temporally semantically enhanced Transformer model in complex enterprise data pipeline anomaly detection tasks.

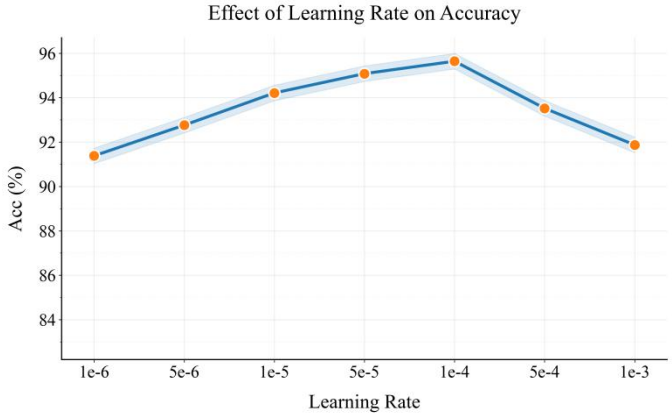

Figure 5. The impact of learning rate on Acc

5. Conclusion

This paper addresses the anomaly detection problem in complex enterprise data pipelines by proposing a temporal semantic enhancement Transformer method. This method integrates operational metrics, node semantics, task states, and dependencies into a unified modeling framework. Instead of judging anomalies solely from the perspective of single numerical fluctuations, this method characterizes the continuous evolution of pipeline states through temporal position encoding, captures the relationships between different nodes through dependency attention modeling, and adaptively fuses task semantics and operational features through a semantic gating enhancement mechanism. As a result, local anomalies, link propagation anomalies, and semantic inconsistency anomalies in complex data pipelines can be more fully expressed in the same representation space, enhancing the anomaly detection process's contextual understanding and structural modeling capabilities.

From the perspective of actual operational needs of enterprise data platforms, the proposed method has significant application value. Complex enterprise data pipelines typically carry critical stages such as data acquisition, cleaning, transformation, synchronization, computation, and analysis. Their stability directly impacts upper-level applications such as business reports, risk monitoring, intelligent recommendations, operational analysis, and automated decision-making. Temporal semantic enhancement of Transformers enables more

granular identification of anomalies and risks during pipeline operation, helping to reduce manual investigation costs, shorten anomaly localization time, and mitigate the impact of data latency, field drift, task failure, and upstream/downstream inconsistencies on business systems. For data governance, intelligent operation and maintenance, and enterprise-level data quality management, this method provides effective support for the shift from passive alerting to proactive perception.

Comparative experimental results show that the proposed algorithm exhibits good classification ability and overall stability in complex enterprise data pipeline anomaly detection tasks. This demonstrates that the combination of temporal semantic information and the Transformer representation learning mechanism effectively improves the model's ability to identify anomalies, especially suitable for enterprise data scenarios with multi-source input, multi-node dependencies, and dynamic state changes. Compared to anomaly detection methods that rely solely on local operational indicators or static rules, this method can more fully understand the temporal correlations and business semantics behind anomalies, making the model output more consistent with the actual risk changes during enterprise data pipeline operation.

Future research can be further developed in three directions: interpretability, cross-scenario transfer, and automated governance. On the one hand, an explanation mechanism oriented towards nodes, fields, and task chains can be further built on the anomaly score, enabling the system not only to determine whether an anomaly has occurred, but also to explain the source, scope of impact, and potential propagation path of the anomaly. On the other hand, as enterprise data platforms gradually evolve towards multi-cloud architecture, real-time computing, and cross-organizational data collaboration, anomaly perception models need to possess stronger cross-business scenario adaptability and incremental learning capabilities to cope with constantly changing data structures and operating environments. Furthermore, in the future, anomaly perception results can be combined with automated repair strategies, scheduling optimization mechanisms, and data quality feedback systems to form an integrated intelligent data pipeline governance system encompassing perception, diagnosis, handling, and feedback, thereby further improving the reliability, maintainability, and intelligence level of enterprise data infrastructure.

References

[1] J. Xu, H. Wu, J. Wang, et al., "Anomaly Transformer: Time Series Anomaly Detection with Association Discrepancy," arXiv preprint arXiv:2110.02642, 2021.

[2] Y. Yang, C. Zhang, T. Zhou, et al., "DCdetector: Dual Attention Contrastive Representation Learning for Time Series Anomaly Detection," in Proceedings of the 29th Association for Computing Machinery Special Interest Group on Knowledge Discovery and Data Mining Conference on Knowledge Discovery and Data Mining, 2023, pp. 3033-3045.

[3] J. Song, K. Kim, J. Oh, et al., "Memto: Memory-Guided Transformer for Multivariate Time Series Anomaly Detection," Advances in Neural Information Processing Systems, vol. 36, pp. 57947-57963, 2023.

[4] H. Kang and P. Kang, "Transformer-Based Multivariate Time Series Anomaly Detection Using Inter-Variable Attention Mechanism," Knowledge-Based Systems, vol. 290, p. 111507, 2024.

[5] Z. Liu, X. Huang, J. Zhang, et al., "Multivariate Time-Series Anomaly Detection Based on Enhancing Graph Attention Networks with Topological Analysis," in Proceedings of the 33rd Association for Computing Machinery International Conference on Information and Knowledge Management, 2024, pp. 1555-1564.

[6] Q. He, G. Wang, H. Wang, et al., "Multivariate Time-Series Anomaly Detection via Temporal Convolutional and Graph Attention Networks," Journal of Intelligent & Fuzzy Systems, vol. 44, no. 4, pp. 5953-5962, 2023.

[7] Y. Chen, C. Zhang, M. Ma, et al., "ImDiffusion: Imputed Diffusion Models for Multivariate Time Series Anomaly Detection," arXiv preprint arXiv:2307.00754, 2023.

[8] Y. Xu, "Intelligent Supply Chain Risk Identification via Heterogeneous Graph Learning and Multi-Entity Interaction Modeling," 2025.

[9] S. Sun, "Adaptive Stock Trading Algorithms Jointly Driven by Multi-Agent Collaborative Decision Making and Large Language Models in Complex Financial Market Environments," 2025.

[10] S. Tuli, G. Casale, and N. R. Jennings, "TranAD: Deep Transformer Networks for Anomaly Detection in Multivariate Time Series Data," arXiv preprint arXiv:2201.07284, 2022.

[11] J. Audibert, P. Michiardi, G. Guyard, et al., "USAD: UnSupervised Anomaly Detection on Multivariate Time Series," in Proceedings of the 26th Association for Computing Machinery Special Interest Group on Knowledge Discovery and Data Mining International Conference on Knowledge Discovery & Data Mining, 2020, pp. 3395-3404.

[12] M. Munir, S. A. Siddiqui, A. Dengel, et al., "DeepAnT: A Deep Learning Approach for Unsupervised Anomaly Detection in Time Series," IEEE Access, vol. 7, pp. 1991-2005, 2018.

[13] P. Malhotra, A. Ramakrishnan, G. Anand, et al., "LSTM-Based Encoder-Decoder for Multi-Sensor Anomaly Detection," arXiv preprint arXiv:1607.00148, 2016.

[14] K. Hundman, V. Constantinou, C. Laporte, et al., "Detecting Spacecraft Anomalies Using LSTMs and Nonparametric Dynamic Thresholding," in Proceedings of the 24th Association for Computing Machinery Special Interest Group on Knowledge Discovery and Data Mining International Conference on Knowledge Discovery & Data Mining, 2018, pp. 387-395.

[15] M. A. Bashar and R. Nayak, "TAnoGAN: Time Series Anomaly Detection with Generative Adversarial Networks," in 2020 Institute of Electrical and Electronics Engineers Symposium Series on Computational Intelligence, 2020, pp. 1778-1785.

[16] M. Wielgosz et al., "Recurrent Neural Networks for Anomaly Detection in the Post-Mortem Time Series of LHC Superconducting Magnets," arXiv preprint arXiv:1702.00833, 2017.