Journal of Computer Technology and Software

ISSN:2998-2383

Vol. 3, No. 2, 2024

Self-Supervised Anomaly Detection with Knowledge-Enhanced Representation Learning for Distributed System Environments

Ao Zhu
University of Pennsylvania, Philadelphia, USA
fzhu4171@gmail.com

Abstract: To address the challenges of anomaly detection in distributed system environments, such as complex anomaly types, high annotation costs, and limited anomaly sample quantities, this paper proposes a unified modeling method for self-supervised detection and knowledge enhancement, specifically for scenarios with scarce anomalies. This method is based on multi-source operational observations, organizing logs, metrics, and dependencies into a unified representation object. Building upon this, it combines structural modeling and temporal context encoding to learn a latent representation that characterizes the inherent patterns of the system state. To alleviate the limitations of scarce anomaly samples on supervised training, this paper constructs a self-supervised learning objective through mask reconstruction and contrastive constraints, enabling the model to extract discriminative state features from a large amount of unlabeled operational data. Simultaneously, addressing the issue that data-driven representations alone are insufficient to fully express operational semantics in complex distributed environments, this paper further introduces knowledge memory and adaptive fusion mechanisms. Prior knowledge of system structure and fault-related information is injected into the representation learning process, thereby enhancing the model's ability to identify anomaly states and its semantic consistency. Finally, this paper constructs a unified framework for anomaly detection in distributed systems, organically combining multi-source information modeling, self-supervised representation learning, knowledge enhancement constraints, and anomaly scoring. Comparative analysis shows that the proposed method can more effectively improve anomaly detection performance, providing a new research approach for intelligent operation and maintenance analysis in complex digital infrastructure scenarios.

Keywords: Heterogeneous observation fusion; representation constraint learning; operation and maintenance knowledge injection; abnormal state identification

1. Introduction

With the rapid development of technologies such as cloud computing, microservices, container orchestration, and edge computing, distributed systems have become the core infrastructure supporting the operation of the digital economy, widely serving key scenarios such as intelligent manufacturing, financial transactions, e-commerce, industrial internet, and government platforms[1]. These systems typically exhibit significant characteristics such as large node scale, diverse component types, long service call chains, and rapidly changing operating environments. The system state becomes highly complex under conditions of high concurrency, strong coupling, and continuous evolution. Simultaneously, while distributed architecture enhances resource elasticity and business scalability, it also brings practical problems such as concealed anomaly propagation paths, multiple intertwined fault causes, and the rapid amplification of local anomalies into global risks, placing higher demands on system stability assurance. Against the backdrop of digital infrastructure gradually evolving towards intelligence, autonomy, and high reliability, how to detect abnormal states in distributed systems in a timely, accurate, and robust manner has become an important research topic in the fields of intelligent operation and maintenance and system security[2].

However, unlike fault identification tasks in traditional closed environments, anomaly detection in distributed systems has long faced the fundamental challenge of a scarcity of abnormal samples. On the one hand, high-risk anomalies in real production environments generally occur infrequently and are characterized by sporadic occurrence, suddenness, and scenario dependence, resulting in a limited number of anomaly samples available for modeling and an extremely uneven distribution of categories. On the other hand, many anomalies manifest as weak signal disturbances in their early stages, easily confused with normal fluctuations, load jitter, or short-term changes caused by resource contention, increasing the difficulty of characterizing anomaly boundaries[3]. Furthermore, system version iterations, changes in business models, and adjustments to resource scheduling strategies continuously alter the distribution of normal behavior and the forms of anomalies, causing traditional supervised methods trained with sufficient labeled samples to face problems such as insufficient generalization ability, weak transfer adaptability, and high long-term maintenance costs in practical deployments. Therefore, research on anomaly representation learning and detection mechanisms under scarce sample conditions has significant theoretical value and practical necessity.

Against this backdrop, self-supervised learning provides a new research path for anomaly detection in distributed systems.

Compared to supervised paradigms that heavily rely on manual annotation, self-supervised methods can extract intrinsic structural information from massive amounts of unlabeled runtime logs, indicator sequences, call chains, and event contexts[4]. By designing pre-training tasks, they learn the temporal patterns, topological dependencies, and semantic consistency of the system, thereby obtaining representations with greater discriminative and transfer capabilities. For scenarios where anomalous samples are scarce, this modeling approach, using normal data as the primary learning vehicle, helps alleviate the learning bottleneck caused by insufficient annotation and improves the model's sensitivity to unknown anomalies, slight shifts, and complex perturbations. However, self-supervised representations that rely solely on data-driven approaches may still be limited by factors such as incomplete observations, strong heterogeneity of anomaly patterns, and the difficulty in explicitly injecting operational knowledge. Especially when facing multi-source coupled faults, cross-level influence chains, and long-distance dependent anomalies, features learned solely from data distribution are often insufficient to support high-reliability detection.

Therefore, introducing knowledge enhancement mechanisms into anomaly detection frameworks is of great significance for improving the intelligent diagnostic capabilities of distributed systems[5]. Distributed systems accumulate a wealth of structured and semi-structured knowledge over long-term operation, including system topology dependencies, service call relationships, component functional semantics, alarm rules, fault handling experience, and operational constraints. This knowledge contains a high-level understanding of the mechanisms and propagation patterns of anomalies. By combining domain knowledge with self-supervised representation learning, we can not only bridge the cognitive gap of purely data-driven methods under small sample and weak supervision conditions, but also enhance the model's ability to interpret complex anomaly patterns, its ability to focus on key risk signals, and its adaptability to dynamic scene changes. Research on self-supervised detection and knowledge enhancement methods for distributed systems with scarce anomaly samples will help advance anomaly detection from a passive identification paradigm that relies on a large number of labels to an active intelligent analysis paradigm that balances the utilization of prior knowledge, weak supervision learning, and highly robust perception. This has significant academic and engineering value for ensuring the safe and stable operation of complex digital infrastructure.

2. BackGround

The research foundation for anomaly detection in distributed systems stems from the continuous evolution of system observation data and the ongoing upgrading of operational goals. Early system state awareness primarily relied on single indicators such as resource utilization, service response time, and error code ratios for threshold judgment. While these methods are simple to implement and easy to deploy, they struggle to adapt to the high-dimensional, multi-source, asynchronous, and strongly correlated operational characteristics of modern distributed environments[6]. With the development of cloud-native architectures, service meshes, elastic scheduling, and automated operation and maintenance systems, system observation objects have gradually expanded to include various heterogeneous data sources such as logs, metrics, links, events, configurations, and topologies. The anomaly detection problem has shifted from single-point monitoring to a complex perception task oriented towards global state modeling. Especially in large-scale service dependency scenarios, the same anomaly often exhibits differentiated behavior at different levels, time windows, and in different components[7]. Therefore, anomaly detection is no longer simply about identifying isolated signals deviating from normal distributions; it requires a comprehensive understanding of the system state within temporal correlations, structural dependencies, and semantic context. Table 1 summarizes the relevant content to more clearly summarize the data foundation and task characteristics currently faced by anomaly detection in distributed systems.

Table 1: Main Observation Dimensions and Analytical Features of Distributed System Anomaly Detection

Observation Dimension	Typical Content	Main Features	Role in Detection Tasks
Metric Data	CPU utilization, memory usage, latency, throughput, and error rate	Strong continuous temporal patterns and susceptibility to workload fluctuations	Reflects the operating trends and performance degradation of the system
Log Data	Runtime logs, exception logs, alert text, and status records	Rich semantic information, but often noisy and heterogeneous in format	Provides semantic clues and contextual information for anomalies
Trace Data	Invocation paths, service dependencies, and cross-node request relationships	Strong topological correlations with propagation and cascading characteristics	Supports anomaly propagation analysis and impact localization
Event Data	Container restarts, node migration, elastic scaling, and deployment changes	Clearly triggered discrete events with high temporal sensitivity	Reveals anomaly triggers and critical disturbance moments
Configuration and Topology Data	Service dependency relations, deployment structures, node roles, and resource allocation	Structurally stable but evolving along with version updates	Provides prior structural constraints and knowledge support

At the methodological level, anomaly detection in distributed systems has evolved from rule-driven and shallow statistical analysis to a research direction that emphasizes representation learning, contextual modeling, and knowledge fusion. Because system operational data naturally possesses characteristics such as high dimensionality, sparsity, local coupling, and dynamic drift, relying solely on fixed rules or static patterns is insufficient to cover anomalies in complex environments. This has driven related research to focus on extracting stable representations from unlabeled data and further combining system structural knowledge to improve recognition reliability. From the essence of the problem, anomaly detection involves not only the boundary delineation between normal and anomalies but also the abstract modeling of system behavior mechanisms, that is, how to characterize the internal consistency of normal patterns, how to identify inconsistencies between cross-source data, and how to utilize existing operational knowledge to narrow the search space and improve judgment credibility. Based on this understanding, research on self-supervised detection and knowledge enhancement under the condition of scarce anomaly samples is essentially exploring a modeling approach that better aligns with the cognitive laws of complex systems. This enables detection models to maintain a stable perception of temporal disturbances, structural mutations, and implicit risks under limited explicit supervision, thereby providing a more solid theoretical support for subsequent anomaly localization, root cause analysis, and intelligent operational decision-making.

3. Methods

Distributed systems generate heterogeneous observations whose temporal fluctuation, cross-service dependency, and semantic inconsistency jointly determine whether an operational state should be regarded as normal or suspicious. To capture this property under anomaly-scarce conditions, the proposed method builds a unified event graph over a sliding window and treats self-supervised structure discovery as the primary learning principle rather than relying on dense anomaly labels.

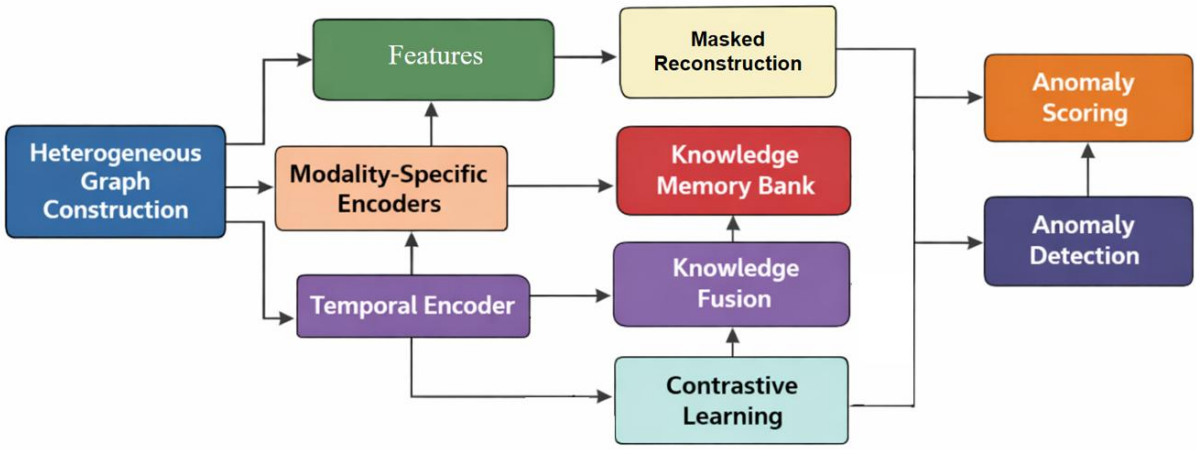

Figure 1. Overall model architecture

Let the observation window at time step t be represented as a heterogeneous graph composed of service instances, log events, metric segments, and dependency relations, where topology is used to preserve propagation paths that are often lost in isolated sequence modeling:

$$\mathcal{G}_t = (\mathcal{V}_t, \mathcal{E}_t, \boldsymbol{X}_t, \boldsymbol{A}_t)$$

Here, $\mathcal{V}_t$ denotes the node set, $\mathcal{E}_t$ is the edge set induced by invocation, deployment, and temporal co-occurrence relations, $\boldsymbol{X}_t \in \mathbb{R}^{|\mathcal{V}_t| \times d}$ contains multimodal node attributes, and $\boldsymbol{A}_t$ is the adjacency matrix encoding structural connectivity. Since raw observations arise from different sources and scales, each node first undergoes modality-aware projection so that metric values, log semantics, and categorical states can be aligned into a comparable latent space.

$$\boldsymbol{H}_t^{(0)} = \phi\left(\boldsymbol{X}_t \boldsymbol{W}_0 + \boldsymbol{b}_0\right)$$

In this expression, $\phi(\cdot)$ is a nonlinear transformation, while $\boldsymbol{W}_0$ and $\boldsymbol{b}_0$ map heterogeneous observations into a shared representation space. Structural context is then injected through relational message passing, because anomalies in distributed systems rarely remain local and instead emerge through dependency distortion, delayed response chains, and correlated deviations across components:

$$\boldsymbol{H}_t^{(l+1)} = \sigma\left(\widetilde{\boldsymbol{D}}_t^{-\frac{1}{2}} \widetilde{\boldsymbol{A}}_t \widetilde{\boldsymbol{D}}_t^{-\frac{1}{2}} \boldsymbol{H}_t^{(l)} \boldsymbol{W}_l\right)$$

where $\widetilde{\boldsymbol{A}}_t = \boldsymbol{A}_t + \boldsymbol{I}$ augments self-loops, $\widetilde{\boldsymbol{D}}_t$ is the corresponding degree matrix, and $\sigma(\cdot)$ denotes nonlinear activation. A graph-level summary is further extracted to describe the holistic operational state of the window, which is critical because many faults can only be recognized after aggregating weak local disturbances into a system-level signal:

$$\boldsymbol{g}_t = READOUT\left(\boldsymbol{H}_t^{(L)}\right)$$

Beyond structural encoding, temporal continuity is explicitly modeled to distinguish transient noise from semantically meaningful state transitions. Instead of assuming that adjacent windows are equally informative, the method constructs context-aware dependencies so that abrupt but explainable workload shifts are not confused with genuine anomalies. For this purpose, graph summaries from consecutive windows are fed into a temporal encoder that preserves both short-range fluctuation and long-range evolution:

$$\boldsymbol{s}_t = GRU\left(\boldsymbol{g}_t, \boldsymbol{s}_{t-1}\right)$$

Such a recurrent state $\mathbf{s}_t$ serves as a compact descriptor of evolving system dynamics. On top of this dynamic state, a masked reconstruction task is introduced to force the encoder to recover missing node information from the surrounding structure and temporal context, thereby encouraging the model to learn normal operational regularities without requiring manual labels:

$$\widehat{\boldsymbol{X}}_t^m = f_{rec}\left(\boldsymbol{H}_t^{(L)}, \boldsymbol{s}_t\right)$$

Because reconstruction alone may overemphasize easy local patterns, a contrastive consistency objective is added to preserve invariant semantics across perturbed views of the same system state while separating unrelated windows:

$$\mathcal{L}_{con} = -log\frac{exp\left(sim\left(\boldsymbol{z}_t, \boldsymbol{z}_t^+\right)/\tau\right)}{exp\left(sim\left(\boldsymbol{z}_t, \boldsymbol{z}_t^+\right)/\tau\right) + \sum_{j=1}^{K} exp\left(sim\left(\boldsymbol{z}_t, \boldsymbol{z}_j^-\right)/\right.}$$

where $\mathbf{z}_t$ is the anchor representation, $\boldsymbol{z}_t^+$ is a positive view generated from the same window, $\boldsymbol{z}_j^-$ denotes negative samples, $\mathrm{sim}(\cdot,\cdot)$ is cosine similarity, and τ is the temperature parameter controlling distribution sharpness. Reconstruction and contrast are complementary, since the former emphasizes recoverability of operational patterns while the latter improves discriminability under scarce abnormal evidence:

$$\mathcal{L}_{ssl} = \lambda_1 \|\boldsymbol{X}_t^m - \widehat{\boldsymbol{X}}_t^m\|_2^2 + \lambda_2 \mathcal{L}_{con}$$

System knowledge is subsequently incorporated to reduce the ambiguity that often arises when purely data-driven learning encounters unseen failure modes or unstable observation quality. Practical distributed platforms contain abundant prior information, including service dependency graphs, component functionality, deployment constraints, and known fault associations; these signals do not merely supplement the learned representation but also constrain what kinds of deviations should be considered operationally meaningful. To exploit such knowledge, a semantic prototype bank is constructed from curated domain rules and topology-aware priors, then aligned with the self-supervised state representation through attention-based fusion:

$$\boldsymbol{q}_t = Attn\left(\boldsymbol{s}_t, \boldsymbol{M}_k\right)$$

In this formulation, $\boldsymbol{M}_k \in \mathbb{R}^{P\times d}$ is the knowledge memory with P learnable prototypes, and $\boldsymbol{q}_t$ denotes the knowledge-aware contextual response relevant to the current operational state. Since not every prior pattern should dominate the decision process, an adaptive gate is used to modulate how much external knowledge should influence the final representation under different system conditions:

$$\boldsymbol{u}_t = \gamma_t \odot \boldsymbol{s}_t + \left(1-\gamma_t\right) \odot \boldsymbol{q}_t$$

$$\gamma_t = sigmoid\left(\boldsymbol{W}_g\left[\boldsymbol{s}_t; \boldsymbol{q}_t\right] + \boldsymbol{b}_g\right)$$

Here, $[\mathbf{s}_t; \mathbf{q}_t]$ denotes vector concatenation, $\mathbf{W}_g$ and $\mathbf{b}_g$ are trainable parameters, and $\odot$ represents element-wise multiplication. This design ensures that stable normal patterns remain mostly data-dominated, whereas ambiguous states can borrow explanatory support from prior knowledge. To maintain semantic compatibility between the learned system dynamics and expert-informed prototypes, a knowledge alignment term is introduced so that the latent state stays close to its most relevant prior manifold:

$$\mathcal{L}_{kn} = \|\boldsymbol{u}_t - \boldsymbol{m}_t^*\|_2^2, \quad \boldsymbol{m}_t^* = \underset{\boldsymbol{m}\in\boldsymbol{M}_k}{arg\,max}\, sim\left(\boldsymbol{u}_t, \boldsymbol{m}\right)$$

Anomaly scoring is finally formulated as a joint deviation measurement that combines predictive inconsistency, representational irregularity, and prior mismatch, because no single signal is sufficient for complex distributed environments. A state that cannot be well reconstructed, fails to remain consistent with neighboring temporal context, and simultaneously deviates from known operational priors should receive a higher anomaly intensity than one exhibiting only isolated fluctuation:

$$a_t = \alpha\|\boldsymbol{X}_t^m - \widehat{\boldsymbol{X}}_t^m\|_2^2 + \beta\left(1 - sim\left(\boldsymbol{u}_t, \boldsymbol{u}_{t-1}\right)\right) + \eta\|\boldsymbol{u}_t - \boldsymbol{m}_t^*\|_2^2$$

A higher score a_t indicates stronger evidence that the current window departs from normal operational dynamics. Training minimizes the self-supervised learning term and the knowledge alignment term jointly, which encourages the representation space to be both structurally informative and semantically grounded:

$$\mathcal{L} = \mathcal{L}_{ssl} + \lambda_3 \mathcal{L}_{kn}$$

Through this objective, the method learns to characterize normal behavior from abundant unlabeled observations, strengthens representation reliability with system knowledge, and produces anomaly scores that remain meaningful even when explicit abnormal samples are highly limited. The overall design, therefore, addresses scarcity not by amplifying rare labels but by improving the quality, interpretability, and operational faithfulness of the latent state space itself.

4. Experimental Results and Analysis

4.1 Dataset

This paper selects the Multi-Source Distributed System Data for AI-Powered Analytics as the research dataset. This dataset is an open-source data resource publicly released for intelligent operation and maintenance tasks of distributed systems. Published on the Zenodo platform and licensed under CC BY 4.0, it supports reproduction and expansion in academic research. The data originates from the actual operation of the complex distributed system OpenStack, covering three core observable signals: distributed traces, application logs, and metrics. It also provides workload scripts, fault scripts, and a rally report that can serve as ground truth. The dataset is further divided into two subsets: sequential data and concurrent data, corresponding to sequential request load scenarios and concurrent request load scenarios, respectively. This configuration effectively preserves the dynamic behavioral characteristics of distributed systems under different operational pressures.

This dataset is highly compatible with research on self-supervised detection of distributed systems oriented towards anomaly-scarce samples. On the one hand, multi-source heterogeneous observation data can support graph structure modeling, cross-modal representation learning, and temporal context modeling, providing a rich information foundation for the design of self-supervised tasks. On the other hand, the dataset contains both fault scripts and synchronously acquired multimodal observation signals, making it suitable for studying representation shifts, cross-source correlations, and knowledge enhancement fusion under anomalous conditions. Notably, the data description explicitly emphasizes the time synchronization requirement between logs, metrics, and traces, a feature that makes it more suitable for building anomaly detection frameworks for distributed systems with complex dependencies and multi-source evidence fusion. Therefore, compared to datasets containing only a single log or metric, this open-source dataset better aligns with the paper's thematic focus on distributed systems, self-supervised detection, and knowledge enhancement modeling.

4.3 Experimental setup

The experimental setup was configured to meet the stable training and representation learning requirements of anomaly detection in distributed systems. Considering the research focus on self-supervised detection and knowledge-enhanced modeling under conditions of scarce anomaly samples, a unified windowed input method and mini-batch iteration strategy were adopted during the training phase to ensure that temporal context, structural dependencies, and multi-source observation information could be learned collaboratively within the same training framework. During model optimization, the encoding layer dimension, time window length, number of graph propagation layers, contrastive learning temperature coefficient, knowledge memory unit size, and weights of each loss term were all set according to the principle of balancing expressive power and training stability. A lower initial learning rate helps avoid drastic oscillations in the representation space during multi-objective joint optimization; a moderate dropout ratio mitigates the risk of overfitting in scenarios with scarce anomaly samples; and knowledge-enhanced parameters are used to balance the influence between data-driven representation and domain prior constraints. For ease of explanation, the specific hyperparameter configurations are shown in Table 2.

Table 2: Detailed hyperparameter settings

Parameter Category	Parameter Name	Setting Value
Data Processing	Time Window Length	20
Data Processing	Sliding Step Size	5
Data Processing	Batch Size	32
Representation Learning	Input Feature Dimension	128
Representation Learning	Hidden Layer Dimension	256
Representation Learning	Number of Graph Propagation Layers	2
Representation Learning	Number of Temporal Encoding Layers	1
Representation Learning	Dropout	0.3
Self-Supervised Learning	Mask Ratio	0.15
Self-Supervised Learning	Contrastive Learning Temperature Coefficient	0.2
Knowledge Enhancement	Number of Knowledge Memory Units	32
Loss Function	Reconstruction Loss Weight	1.0
Loss Function	Contrastive Loss Weight	0.5
Loss Function	Knowledge Alignment Loss Weight	0.3
Optimization Settings	Optimizer	AdamW
Optimization Settings	Initial Learning Rate	0.0005
Optimization Settings	Weight Decay	0.0001
Training Control	Number of Training Epochs	100

4.4 Experimental Results and Analysis

To ensure a high degree of consistency between the comparison objects and the research topic of this paper in terms of task attributes, data sources, and methodological approaches, this paper selects relevant research as comparative references. These works mainly focus on distributed systems, microservice logs, call chains, performance indicators, and self-supervised or unsupervised anomaly detection, and can comprehensively reflect the representative research paths in this direction. Their experimental results are shown in Table 3.

Table 3: Experimental results compared with other models

Method	ACC	PRE	REC	AUC
Zhang et al.[8]	91.24	90.87	90.31	95.42
Xie et al.[9]	92.08	91.56	91.12	96.03

Xie et al.[10]	92.41	92.05	91.67	96.28
Chen et al.[11]	93.15	92.74	92.33	96.91
Panahandeh et al.[12]	93.68	93.21	92.89	97.24
Guo et al.[13]	94.02	93.66	93.28	97.53
Yamanaka et al.[14]	94.37	94.05	93.71	97.86
Ours	95.48	95.12	94.83	98.44

Overall, the proposed method achieves state-of-the-art performance across all evaluation metrics, demonstrating its ability to learn more stable and discriminative anomaly representations under complex operating conditions in distributed systems. Since this task involves multi-source observation information modeling, temporal context characterization, and boundary identification under anomaly scarcity conditions, relying solely on a single signal or shallow patterns often fails to adequately reflect the fine-grained differences between system states. In contrast, the current method enhances its ability to model normal operating patterns through self-supervised representation learning and further compresses the confusion space between anomalous and normal states by incorporating a knowledge enhancement mechanism. Therefore, it exhibits stronger comprehensive performance in terms of identification accuracy, prediction precision, anomaly recall, and overall discriminative ability.

This result demonstrates the effectiveness of integrating structural dependence, temporal evolution, and prior knowledge constraints into a unified detection framework. For distributed systems, anomalies are often not isolated local disturbances but rather gradually emerge alongside cross-component propagation, imbalances in upstream and downstream relationships, and inconsistencies among multi-source signals. The proposed method effectively preserves this complex correlation and maintains high recognition quality even with a limited number of anomalous samples, indicating that the learned representation possesses better robustness and task adaptability. Furthermore, the consistently leading performance across various metrics reflects that the method does not merely improve on a single dimension but rather achieves a relatively balanced advantage in overall detection capability. This provides a more reliable foundation for subsequent anomaly localization, risk warning, and intelligent operation and maintenance analysis.

The introduction of contrastive learning constraints essentially enhances the structural separability of the representation space under conditions of anomaly scarcity, enabling clearer boundaries between normal behavior patterns and anomalous perturbation patterns in the latent space. For distributed systems, complex coupling relationships often exist between multi-source observations, and relying solely on reconstruction consistency is insufficient to fully characterize fine-grained state differences. Therefore, it is necessary to examine the distance distribution characteristics of positive and negative samples in the representation space. Based on this, visualizing the distance changes in the representations of positive and negative samples helps to verify the rationality of the proposed method in terms of representation constraints and anomaly discrimination from a geometrical perspective. The experimental results are shown in Figure 2.

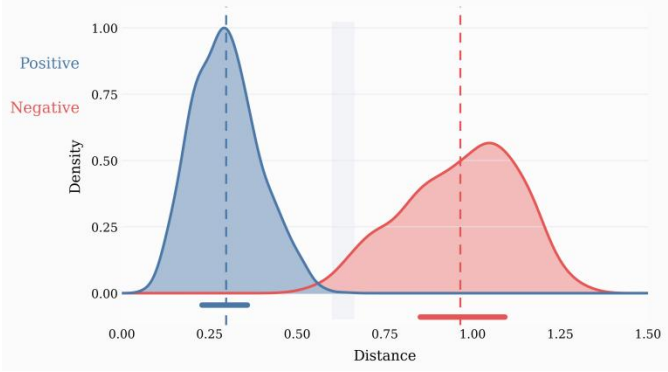

Figure 2. Visualization experiment of the distribution of positive and negative sample representation distance changes under contrastive learning constraints

From the perspective of the distribution pattern of the representation space, the proposed method can effectively compress the semantic distance between positive samples and maintain sufficient distinguishing boundaries for negative samples in the latent space, indicating that the state representation learned by the current method has strong structural constraint capabilities. In distributed system scenarios where anomalies are scarce, if the model cannot stably identify the internal consistency of normal behavior, it is easily affected by local perturbations and multi-source noise, thereby weakening the reliability of anomaly detection. Current visualization results show that the proposed method has good compactness and separability at the representation level, which means that it can not only preserve the core structure of the normal operation mode, but also more clearly separate anomaly-related perturbations from the shared representation.

This phenomenon further illustrates that contrastive learning constraints play a key role in the framework of this paper, enabling the model to no longer be limited to local reconstruction of consistency during self-supervised learning, but to strengthen the distinguishing boundaries between different states from a global geometric perspective. For distributed systems, which have complex dependencies and multimodal observation characteristics, state differences are often reflected in joint offsets across components, time, and signal sources. Therefore, only by forming a clearer semantic organization in the representation space can the stability and reliability of subsequent anomaly detection be supported. The current results demonstrate that the proposed method has established a relatively reasonable geometric structure during the representation learning stage, providing a higher-quality potential representation foundation for anomaly detection tasks.

The learning rate determines the parameter update step size and the stability of the optimization trajectory; therefore, it is necessary to examine the changes in the model's discriminative ability under different learning rate settings. For the method presented in this paper, self-supervised representation learning, knowledge reinforcement constraints, and anomaly scoring mechanisms are jointly optimized, and the learning rate setting

directly affects the organization quality of the latent space and the shaping process of anomaly boundaries. Based on this, visualizing the AUC changes under different learning rates helps to present the training adaptability and parameter sensitivity characteristics of the proposed method from the perspective of optimization behavior.

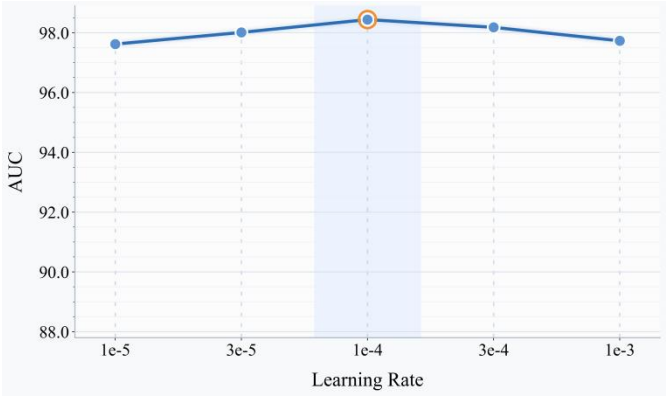

Figure 3. The impact of the learning rate on the AUC experimental results

As shown in Figure 3, the proposed method achieves better discriminative ability within a moderate learning rate range, indicating that the joint optimization process of the current model is highly sensitive to the step size setting. If the learning rate is too low, the parameter update amplitude is limited, making it difficult to fully promote a more effective collaboration between the self-supervised representation space and knowledge reinforcement constraints; if the learning rate continues to increase, it is easy to disrupt the stable organization of the latent representation, causing perturbation to the learning of anomaly boundaries. The current results show that an appropriate learning rate can better balance the sufficiency of representation learning and the stability of the training process, thus enabling the proposed method to maintain good recognition quality and optimization adaptation ability in anomaly detection tasks.

5. Conclusion

This paper addresses the key challenges of anomaly detection in distributed systems under conditions of scarce anomaly samples, systematically studying a unified modeling method combining self-supervised detection and knowledge enhancement. To overcome the shortcomings of traditional methods, such as strong label dependency, ambiguous anomaly boundaries, insufficient utilization of cross-source information, and limited generalization ability in complex system scenarios, this paper starts from the intrinsic structure of multi-source observation data in distributed systems. It organically combines heterogeneous graph modeling, temporal context characterization, self-supervised representation learning, and domain knowledge fusion to construct an anomaly detection framework for complex operating environments. This research not only addresses the core issue of how to improve detection capabilities under conditions of scarce anomaly samples at the methodological level but also further illustrates that, in intelligent operation and maintenance scenarios, relying solely on a single data-driven paradigm is insufficient to meet the actual needs of highly reliable and complex systems. A closer synergy must be established between representation learning capabilities and the utilization of prior knowledge.

From a research significance perspective, the value of this work lies not only in the task of anomaly detection in distributed systems itself, but also in its role in promoting intelligent operation and maintenance, system security assurance, and the management paradigm of complex digital infrastructure. With the widespread deployment of cloud computing platforms, microservice architectures, edge collaboration systems, and industrial-grade digital platforms, system state evolution is increasingly exhibiting high dynamism, high coupling, and high uncertainty. Traditional anomaly identification methods relying on human experience and static rules are no longer sufficient to support real-time monitoring and risk warning in large-scale scenarios. The method proposed in this paper emphasizes extracting stable representations from large amounts of unlabeled operational data while leveraging structural knowledge, operational knowledge, and system dependencies to enhance anomaly semantic discrimination. This enables the detection model to not only improve the identification quality of complex anomaly patterns but also enhance the understandability and scenario adaptability of the results to a certain extent. For application areas such as cloud service governance, critical business continuity assurance, fintech infrastructure operation and maintenance, industrial internet platform monitoring, and high-reliability computing environment management, this research approach has strong practical application value, helping to promote the evolution of anomaly detection from passive alerting to proactive analysis, from local monitoring to global perception, and from experience-driven to intelligent collaboration.

Future research still has considerable room for expansion. First, as distributed systems continue to expand in scale and business models evolve, anomaly detection models need to further enhance their adaptability to open environments, unknown anomaly types, and continuous distribution drift. Therefore, self-supervised anomaly detection mechanisms oriented towards continuous learning and online updates deserve in-depth research. Second, the knowledge enhancement component can continue to develop to higher levels, such as introducing more granular system topology semantics, service dependency constraints, fault handling logic, and operation and maintenance rule bases, to achieve a further extension from anomaly identification to anomaly interpretation, impact analysis, and root cause inference. Third, for real-world deployment scenarios, how to ensure detection accuracy while considering inference efficiency, resource consumption, and engineering feasibility is also an important and unavoidable direction for the future. Overall, this research provides a new methodological foundation and theoretical perspective for anomaly detection in distributed systems under anomaly scarcity conditions, and lays a valuable research foundation for building a more intelligent, reliable, and adaptive modern operation and maintenance analysis system.

References

[1] Xie Y, Zhang H, Babar M A. Logsd: Detecting anomalies from system logs through self-supervised learning and frequency-based masking[J]. Proceedings of the ACM on Software Engineering, 2024, 1(FSE): 2098-2120.

[2] Kohyarnejadfard I, Aloise D, Azhari S V, et al. Anomaly detection in microservice environments using distributed tracing data analysis and NLP[J]. Journal of Cloud Computing, 2022, 11(1): 25.

[3] Nobre J, Pires E J S, Reis A. Anomaly detection in microservice-based systems[J]. Applied Sciences, 2023, 13(13): 7891.

[4] D'Angelo A, d'Aloisio G. Grammar-based anomaly detection of microservice systems execution traces[C]//Companion of the 15th ACM/SPEC International Conference on Performance Engineering. 2024: 77-81.

[5] Xie Z, Pei C, Li W, et al. From point-wise to group-wise: A fast and accurate microservice trace anomaly detection approach[C]//Proceedings of the 31st ACM Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering. 2023: 1739-1749.

[6] Liu Y, Ren S, Wang X, et al. Temporal logical attention network for log-based anomaly detection in distributed systems[J]. Sensors, 2024, 24(24): 7949.

[7] He Y, Zhang X, Wang P, et al. Microservice traces anomaly detection based on structural clustering and graph attention autoencoder[C]//2024 7th International Conference on Advanced Algorithms and Control Engineering (ICAACE). IEEE, 2024: 977-981.

[8] Zhang C, Peng X, Sha C, et al. Deeptralog: Trace-log combined microservice anomaly detection through graph-based deep learning[C]//Proceedings of the 44th international conference on software engineering. 2022: 623-634.

[9] Xie X, Jian S, Huang C. Logtracead: anomaly detection from both logs and traces with graph representation learning[C]//Proceedings of the 2023 2nd International Conference on Networks, Communications and Information Technology. 2023: 116-121.

[10] Xie Z, Xu H, Chen W, et al. Unsupervised anomaly detection on microservice traces through graph vae[C]//Proceedings of the ACM Web Conference 2023. 2023: 2874-2884.

[11] Chen J, Liu F, Jiang J, et al. TraceGra: A trace-based anomaly detection for microservice using graph deep learning[J]. Computer Communications, 2023, 204: 109-117.

[12] Panahandeh M, Hamou-Lhadj A, Hamdaqa M, et al. ServiceAnomaly: An anomaly detection approach in microservices using distributed traces and profiling metrics[J]. Journal of Systems and Software, 2024, 209: 111917.

[13] Guo H, Yuan S, Wu X. Logbert: Log anomaly detection via bert[C]//2021 international joint conference on neural networks (IJCNN). IEEE, 2021: 1-8.

[14] Yamanaka Y, Takahashi T, Minami T, et al. Logelectra: Self-supervised anomaly detection for unstructured logs[J]. arXiv preprint arXiv:2402.10397, 2024.