Transactions on Computational and Scientific Methods | Vo. 4, No. 01, 2024
ISSN: 2998-8780
https://pspress.org/index.php/tcsm
Pinnacle Science Press

Prior Enhanced Representation Learning and Adaptive Recognition Method for Backend Anomaly Detection under Weakly Labeled and Few-Shot Conditions

Yilin Sun
University of Pennsylvania, Philadelphia, USA
syl5353@gmail.com

Abstract: To address the challenges of insufficient supervision under weakly labeled conditions, difficulties in learning anomaly patterns in low-sample scenarios, and susceptibility of log sequence representations to noise interference in backend anomaly detection tasks, this paper proposes a prior-enhanced representation learning and adaptive recognition method for weakly labeled and low-sample conditions. This method focuses on backend log sequences, modeling key issues such as insufficient anomaly semantic extraction, inadequate utilization of structural information, and unclear discrimination boundaries. First, heterogeneous input embedding maps the original log sequence to a unified representation space to reduce semantic bias caused by diverse log patterns. Then, a prior-guided encoding mechanism is introduced to integrate prior memory information into the latent representation learning process, enhancing the model's ability to characterize normal operation patterns and anomaly deviation features. Building upon this, a gating fusion strategy is used to adaptively integrate prior information with current sample features, thereby improving the effectiveness of anomaly-related information retention. Furthermore, neighborhood structure modeling and prototype-level metric calibration are combined to constrain the category distribution relationships in the recognition space, enabling the model to form more compact and discriminative anomaly representations even under weak supervision. Finally, an adaptive anomaly recognition mechanism is constructed to achieve stable discrimination of anomaly samples in complex backend log scenarios. Related research shows that the method presented in this paper can effectively improve the representation quality and recognition ability in backend log anomaly detection tasks, and provides a targeted methodological framework for backend anomaly detection under weak labeling and few sample conditions.

Keywords: Log semantic modeling; prior memory guidance; gated feature integration; prototype metric calibration

1. Introduction

With the rapid popularization of cloud computing, microservice architecture, container orchestration, and continuous delivery systems, modern backend systems are continuously evolving towards high concurrency, high coupling, strong dynamism, and large-scale heterogeneous collaboration [1]. The continuous extension of business chains, the deepening of service dependencies, and the uncertainty of the operating environment significantly increase the complexity of system state evolution, causing backend anomalies to exhibit characteristics such as multi-source triggering factors, concealed propagation paths, diversified manifestations, and diffused impact. In this context, backend anomaly detection is no longer a simple fault identification problem, but a crucial foundational link related to platform stability assurance, service continuity maintenance,

resource scheduling optimization, and the construction of intelligent operation and maintenance capabilities [2, 3]. How to accurately identify potential anomaly patterns in complex operating environments has become an important research topic supporting the secure and efficient operation of digital infrastructure [4].

From a practical application perspective, backend systems continuously generate various heterogeneous operation and maintenance data, such as logs, metrics, links, and events, during long-term operation. This data provides a rich information foundation for anomaly detection, but also brings significant challenges to data understanding and modeling [5]. On the one hand, significant differences exist between data from different sources in terms of semantic granularity, temporal scale, and structural form, leading to anomalous signals often being scattered and embedded in multidimensional observations, increasing the difficulty of unified representation and joint modeling. On the other hand, anomalous events in real industrial environments typically exhibit characteristics such as low frequency, suddenness, rapid evolution, and uneven category distribution. Many anomalous patterns lack sufficient and high-quality annotation support, making traditional identification methods relying on large-scale complete supervisory information difficult to apply directly. Therefore, conducting research on backend anomaly detection under conditions of weak annotation and few samples not only has a clear practical need but also significant theoretical value.

Weak annotation and a few samples place higher demands on backend anomaly detection methods. Due to incomplete annotation information, objectively existing label noise, and insufficient sample coverage, models tend to over-rely on local surface features, making it difficult to learn stable, transferable, and discriminative anomaly representations [6]. This results in recognition results exhibiting strong sensitivity to scene changes, business drift, and system evolution. Especially when new anomalies constantly emerge and existing anomaly forms continuously change, relying solely on limited supervisory information is often insufficient to characterize the essential structural features of anomalies. In contrast, prior information can provide additional constraints to representation learning from the perspectives of system mechanisms, operational rules, structural dependencies, and historical knowledge. This helps mitigate learning bias caused by sample scarcity and enhances the model's ability to perceive and represent key anomaly patterns. Therefore, introducing prior reinforcement ideas and constructing a more robust representation learning framework for weakly labeled and low-sample conditions is an important direction for improving the effectiveness of backend anomaly detection.

Furthermore, backend anomaly detection not only requires models to be able to identify known anomalies but also to adapt to state changes and pattern transfers in complex environments, achieving more flexible, stable, and practical adaptive recognition [7]. With changes in business load, architecture adjustments, service version iterations, and the superposition of external interference factors, the boundaries and performance of anomaly patterns often change dynamically, making it difficult for static detection mechanisms to maintain stable performance over the long term. Research on prior reinforcement representation learning and adaptive recognition methods can help drive anomaly detection from relying on dense labeling and fixed pattern matching to an intelligent analysis paradigm oriented towards knowledge fusion, structural understanding, and dynamic adaptation. This not only enhances the ability to detect anomalies and warn of risks in complex backend systems, but also provides more robust methodological support for intelligent operation and maintenance, system resilience enhancement, and critical business assurance. Therefore, it has significant academic and engineering application value.

2. Datasets and Dataset Preprocessing

2.1 Dataset Introduction

This paper builds a multi-source representation framework for backend anomaly detection based on the HDFS_v1 open-source log dataset released by LogHub. The dataset originates from the runtime logs of the Hadoop Distributed File System and is widely used in fault analysis and anomaly identification studies for distributed backend systems. Rather than treating each log trajectory as a single homogeneous input, this study reorganizes the dataset from multiple observational perspectives, including event semantic sequences, temporal interval patterns, statistical distribution features, and sequence-level structural dependencies. In this way, the

original HDFS data are transformed into a multi-source backend observation space, enabling the model to jointly characterize anomaly-related semantics, runtime dynamics, and structural variation patterns. Such a representation form is better aligned with the practical needs of backend anomaly detection, since anomalies in real systems are usually reflected through the coupling of multiple correlated signals rather than a single isolated log pattern.

The HDFS_v1 dataset employs a log segmentation method based on block ID, dividing the raw logs into corresponding event trajectories and providing a label of "normal" or "anomaly" for each trajectory. It also provides various preprocessing results, such as event templates, anomaly labels, event sequences, and event statistical matrices, facilitating subsequent representation learning and anomaly identification research. Compared to the ideal data format of fine-grained, fully supervised, and comprehensively defined categories, this dataset is closer to the actual data organization methods in backend anomaly detection, providing supervision information at the log trajectory level rather than the full semantic level. Therefore, it is more suitable for anomaly pattern modeling research under weakly labeled conditions. Furthermore, the organization of its anomaly samples provides a reasonable data foundation for anomaly representation learning and adaptive recognition in scenarios with few samples. Research based on this dataset aligns well with the themes of papers addressing backend anomaly detection tasks under weakly labeled and few-sample conditions.

2.2 Dataset Preprocessing

To ensure the effectiveness of log sequence representation and the consistency of data distribution in the backend anomaly detection task, systematic data preprocessing and statistical analysis were performed on the HDFS log dataset before formal modeling. The preprocessing process mainly included steps such as organizing log segmentation results, aligning anomaly and normal sample labels, filtering samples with abnormal lengths, cleaning up invalid log records, and regularizing the event sequence structure. These steps aimed to reduce the interference of redundant noise, distribution skew, and anomaly fluctuations in the original logs on the subsequent representation learning process.

Figure 1 illustrates the distribution changes of the HDFS dataset across several key statistical dimensions before and after preprocessing. Overall, the data distribution after preprocessing exhibits higher consistency and regularity. The top left figure shows the distribution of event sequence lengths for each block. After preprocessing, samples are more concentrated within a relatively stable length range, indicating that excessively short or long abnormal sequences are effectively cleaned up, resulting in a more uniform sequence structure. The top right figure shows the distribution of time intervals between adjacent log events. After preprocessing, a large number of samples are further concentrated in smaller time interval regions, indicating that highly discrete abnormal time spans are weakened, and the temporal continuity of log events is enhanced. The bottom left figure shows the distribution of log message lengths. After preprocessing, the length distribution is more concentrated in the main intervals, and the long-tail phenomenon is significantly reduced, indicating that redundant text or abnormal format logs are filtered to a certain extent, thus making the message expression more stable. The bottom right figure shows the frequency distribution of different event templates. The horizontal axis represents the event template number, and the vertical axis represents the number of times the corresponding template appears in the data. After preprocessing, the frequency distribution of each template is relatively smoother, indicating that the template statistical structure has been organized, and high-noise, low-consistency template fluctuations have been suppressed. Overall, these four subgraphs illustrate the effectiveness of preprocessing from four aspects: event sequence structure, time dependency features, log text length features, and template statistical features, providing a more stable data foundation for subsequent anomaly representation learning and adaptive recognition.

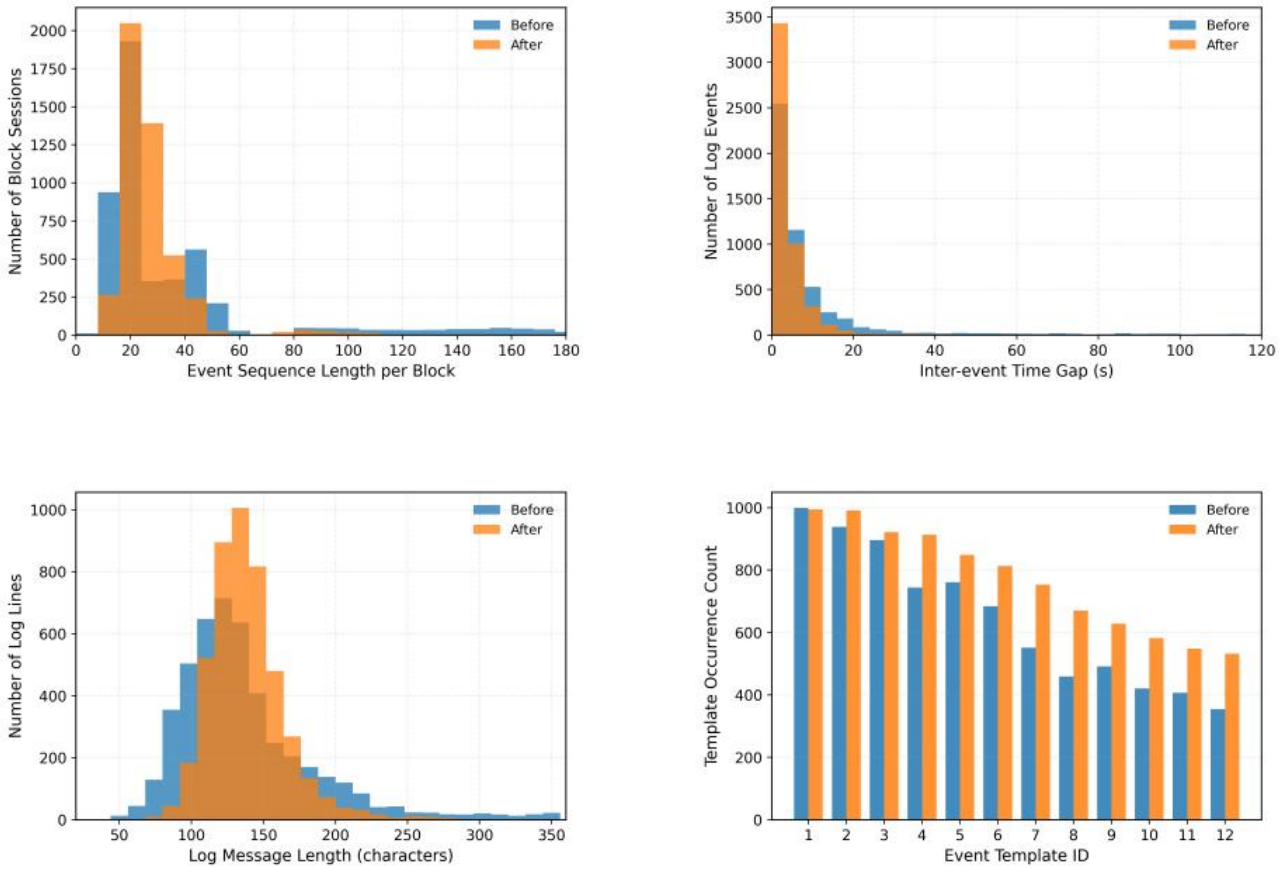

Figure 1. Comparison of data before and after preprocessing

3. Methodology

To address the coupled difficulties caused by weak supervision, scarce abnormal samples, and highly heterogeneous backend observations, this study develops a prior-enhanced representation learning and adaptive recognition framework that transforms raw backend traces into structurally constrained latent descriptions before anomaly decision making. This paper presents the overall model architecture, as shown in Figure 2.

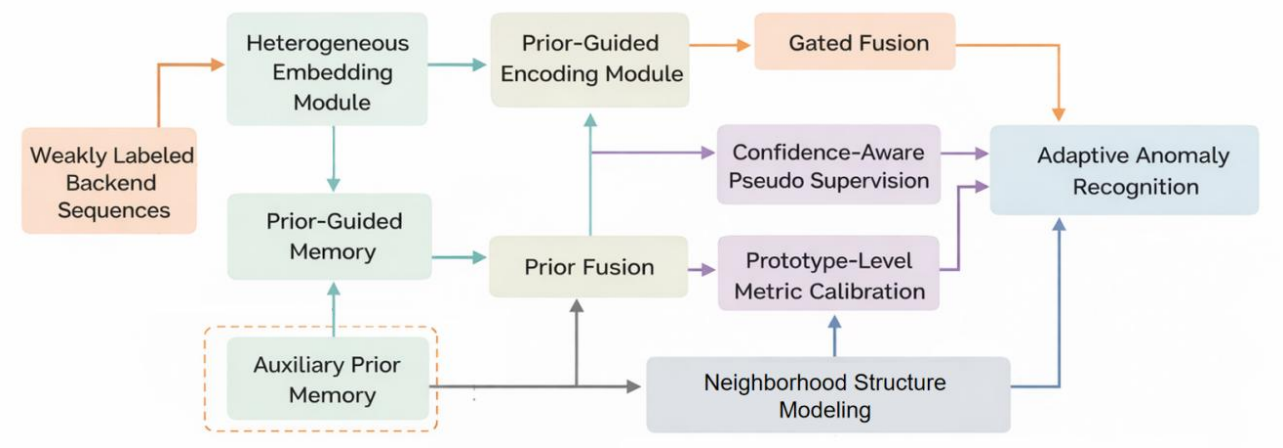

Figure 2. Overall model architecture

Given an input sample set denoted by $\mathcal{D} = \{(\boldsymbol{x}_i, \tilde{y}_i)\}_{i=1}^{N}$, each $\boldsymbol{x}_i \in \mathbb{R}^{T\times d}$ represents a multivariate backend sequence composed of log-derived semantic events, metric fluctuations, and temporal context cues, while $\tilde{y}_i$ corresponds to a weak label that may be incomplete or noisy. Rather than directly optimizing a detector over such imperfect supervision, the proposed method first introduces a prior-guided encoding process so that latent features preserve both discriminative anomaly cues and stable structural regularities shared across backend states. The initial embedding of each sample is defined as:

$$\boldsymbol{H}_i^{(0)} = \phi_{emb}(\boldsymbol{x}_i)$$

where $\phi_{\mathrm{emb}}(\cdot)$ maps heterogeneous backend observations into a unified feature space and reduces the semantic inconsistency among different observation channels. Since weak labels alone are insufficient to constrain the latent geometry, an auxiliary prior memory is further constructed to summarize recurring normal patterns and sparse abnormal prototypes, which improves representation stability under limited annotations. The memory-enhanced prior is formulated as:

$$\boldsymbol{P}_i = Attn\left(\boldsymbol{H}_i^{(0)}, \boldsymbol{M}\right)$$

in which $\boldsymbol{M} \in \mathbb{R}^{K\times d}$ denotes the learnable prior bank and $\mathrm{Attn}(\cdot)$ measures the correspondence between the current sequence representation and stored backend priors. Through this design, the encoder is encouraged to focus on essential operational dependencies rather than overfitting accidental noise patterns, which is particularly important when anomaly samples are rare and label reliability is uneven across instances.

Building upon the prior response, the latent representation is refined through a gated fusion mechanism that selectively injects stable structural knowledge into the current sample while preserving its instance-specific abnormal dynamics. Such a strategy is meaningful because backend anomalies often manifest as subtle deviations from regular service evolution rather than completely isolated patterns, and therefore, the detector must jointly model common operational priors and local irregularity signals. The fusion process is expressed as:

$$\boldsymbol{Z}_i = \sigma\left(\boldsymbol{W}_g\left[\boldsymbol{H}_i^{(0)} \parallel \boldsymbol{P}_i\right]\right) \odot \boldsymbol{P}_i + \left(1 - \sigma\left(\boldsymbol{W}_g\left[\boldsymbol{H}_i^{(0)} \parallel \boldsymbol{P}_i\right]\right)\right) \odot \boldsymbol{H}_i^{(0)}$$

where $\boldsymbol{W}_g$ is a learnable projection, $\sigma(\cdot)$ is the sigmoid activation, and $\odot$ denotes element-wise modulation. Afterward, structural consistency is imposed in the latent space so that samples with similar backend operating semantics remain close even when their observed weak labels are ambiguous. A neighborhood-preserving regularization objective is therefore introduced as:

$$\mathcal{L}_{str} = \sum_{i=1}^{N}\sum_{j=1}^{N} s_{ij} \left\|\boldsymbol{Z}_i - \boldsymbol{Z}_j\right\|_2^2$$

where s_{ij} quantifies structural affinity derived from temporal co-occurrence, event dependency, or sequence similarity. By constraining the latent manifold in this manner, the method improves robustness against fragmented supervision and helps maintain coherent backend state organization under complex operational fluctuations.

Beyond structural regularization, adaptive anomaly recognition is realized by explicitly reducing the bias introduced by weak labels through confidence-aware pseudo-supervision and prototype-level metric calibration. This stage is necessary because backend anomaly categories are often underrepresented, causing a standard classifier to become dominated by frequent normal patterns and to produce unstable boundaries for hard minority cases. For this reason, each latent feature is projected onto a compact recognition space and compared against class prototypes whose positions are iteratively updated during training. The prototype-centered metric is defined as:

$$d_{ic} = \|\psi(\boldsymbol{Z}_i) - \boldsymbol{q}_c\|_2^2$$

with $\psi(\cdot)$ denoting the recognition projection and $\boldsymbol{q}_c$ denoting the prototype of class c. On this basis, the adaptive posterior probability is estimated by:

$$p_{ic} = \frac{\exp(-d_{ic}/\tau_i)}{\sum_{c'=1}^{C} \exp(-d_{ic'}/\tau_i)}$$

where the temperature τ_i varies across samples so that uncertain weakly labeled instances receive smoother decision distributions, whereas high-confidence instances contribute sharper supervisory signals. Owing to this adaptive scaling, the recognition process becomes less sensitive to sparse abnormal categories and more capable of identifying subtle deviations embedded in highly imbalanced backend data.

Finally, the overall optimization objective integrates weakly supervised classification, structural preservation, and prior-consistent representation shaping into a unified learning framework, thereby enabling the model to align semantic abstraction with adaptive anomaly discrimination. In practical backend scenarios, this joint formulation is beneficial because it prevents the encoder from drifting toward purely discriminative but fragile solutions and instead promotes latent patterns that remain meaningful under evolving workloads and incomplete labels. The final loss is written as:

$$\mathcal{L} = \mathcal{L}_{\text{wcls}} + \lambda_1 \mathcal{L}_{\text{str}} + \lambda_2 \mathcal{L}_{\text{prior}} + \lambda_3 \mathcal{L}_{\text{adapt}}$$

where $\mathcal{L}_{\text{wcls}}$ denotes the weakly supervised recognition loss, $\mathcal{L}_{\text{prior}}$ enforces alignment between latent representations and prior memory responses, $\mathcal{L}_{\text{adapt}}$ penalizes unreliable prototype assignments, and λ_1, λ_2, and λ_3 balance the contribution of each term. As a result, the proposed method does not merely treat anomaly detection as a direct label-fitting task, but instead establishes a representation-to-recognition pipeline in which prior knowledge, structural continuity, and adaptive decision mechanisms cooperate to improve reliability under weakly labeled and few-shot backend anomaly detection conditions.

4. Experimental Results and Analysis

4.1 Experimental Setup

To ensure the consistency and reproducibility of the proposed method's evaluation in weakly labeled and few-shot backend anomaly detection tasks, experiments were conducted in a unified software and hardware environment, with standardized settings for data partitioning, training strategies, and optimization parameters. Specifically, training, validation, and test sets were constructed based on the HDFS log dataset. The training set was primarily used for model parameter learning, the validation set for hyperparameter tuning and model selection, and the test set for final performance evaluation. Considering the paper's focus on weakly labeled and few-shot scenarios, only a limited proportion of labeled information was retained during the training phase, and anomaly modeling was completed through prior augmented representation learning and adaptive recognition mechanisms. During optimization, AdamW was used as the parameter update method, combined with fixed batch size, learning rate decay, and weight decay strategies to improve training stability. Simultaneously, all input samples underwent uniform sequence preparation, template encoding, and length normalization before entering the model, thereby reducing the interference of original log noise on the recognition results. Specific experimental settings are shown in Table 1.

Table 1. Experimental setup parameters

Item	Setting
Dataset	HDFS log dataset
Task Type	Backend anomaly detection
Scenario Setting	Weakly labeled and few-shot setting

Data Split	Training set 70%, validation set 10%, test set 20%
Input Form	Log event sequences
Sequence Normalization Length	128
Embedding Dimension	128
Batch Size	32
Optimizer	AdamW
Initial Learning Rate	0.001
Weight Decay	0.0001
Training Epochs	100
Learning Rate Scheduler	Cosine Annealing
Experimental Platform	Python 3.10, PyTorch 2.1, NVIDIA RTX 4090

4.2 Experimental Results Compared with Other Models

To verify the recognition capability of the proposed method in backend log anomaly detection scenarios, representative methods with similar research directions were selected for horizontal comparison, covering techniques such as deep sequence modeling, semi-supervised learning, semantic representation learning, and prototype metric modeling. The experimental results are shown in Table 2.

Table 2. Comparison with representative related methods

Method	ACC	PRE	REC	F1
Du et al. [8]	95.21	94.87	94.63	94.75
Meng et al. [9]	96.08	95.74	95.51	95.62
Zhang et al. [10]	96.41	96.02	95.89	95.95
Yang et al. [11]	96.93	96.48	96.31	96.39
Guo et al. [12]	97.16	96.85	96.67	96.76
Li et al. [13]	97.34	97.02	96.88	96.95
Huang et al. [14]	97.51	97.23	97.04	97.13
Ours	98.27	98.01	97.86	97.93

The proposed algorithm demonstrates superior comprehensive recognition capabilities in backend log anomaly detection tasks, indicating that the method can effectively learn more discriminative anomaly representations under conditions of complex log semantics, weak annotation constraints, and limited sample size. Its advantages are not only reflected in the improved overall recognition accuracy but also in the simultaneous enhancement of the sufficiency and discriminative stability of anomaly sample detection. This demonstrates that prior reinforcement representation learning and adaptive recognition mechanisms can effectively alleviate the limitations of traditional methods in areas such as insufficient log structure modeling, inadequate anomaly semantic capture, and unclear category distinction boundaries. Overall, the proposed method is more suitable for log anomaly detection tasks in real-world backend scenarios, exhibiting strong robustness and practical application value.

4.3 Ablation Test Results

To further verify the actual contribution of each key component of the proposed method to the backend anomaly detection performance, an ablation study was conducted under a unified experimental setup, focusing on configuration analysis of three core aspects: prior guidance coding, gating fusion mechanism, and prototype-level metric calibration. The relevant results are shown in Table 3.

Table 3. Ablation experiment results

Ablation Setting	ACC	PRE	REC	F1
w/o Prior-Guided Encoding Module	96.74	96.31	96.08	96.19
w/o Gated Fusion	97.02	96.68	96.44	96.56
w/o Prototype-Level Metric Calibration	97.21	96.87	96.63	96.75
Ours	98.27	98.01	97.86	97.93

The results demonstrate that the model's performance improvement does not rely on a single local module, but rather on the synergistic effect of multiple components to form a more complete and stable anomaly representation capability. Especially under weak labeling and limited sample conditions, the proposed method can more fully extract key anomaly information from log sequences and maintain good discriminative ability and recognition consistency in complex backend scenarios.

Furthermore, the prior guided coding module, the gated fusion mechanism, and the prototype-level metric calibration all play crucial roles in the final recognition performance. Prior guided coding helps enhance the model's ability to model potential anomaly semantics and structural dependencies; gated fusion improves the targeting and stability of effective information integration; and prototype-level metric calibration further strengthens the ability to distinguish category boundaries and the reliability of recognition. The combined effect of these three mechanisms enables the proposed method to demonstrate higher levels of performance in anomaly representation extraction, key feature preservation, and discriminative decision formation, thus exhibiting strong robustness and practical application value.

4.4 The Impact of the Gating Fusion Coefficient on Acc

To further examine the role of the gating fusion mechanism in our proposed method, it is necessary to conduct a targeted analysis of the values of the gating fusion coefficients. This parameter directly affects the weighting of prior information and current sample representation during the fusion process, and is therefore closely related to the degree of anomaly feature preservation, the strength of structural information injection, and the quality of the final recognition space formation. Conducting sensitivity studies on this parameter will help to more clearly reveal the practical significance of the gating fusion design in backend log anomaly detection tasks and provide a basis for model parameter configuration.

As shown in Figure 3, a reasonable gating fusion coefficient can more effectively coordinate the relationship between prior guiding information and the original representation, enabling the model to maintain a good balance between anomaly semantic modeling and discriminative feature extraction. When the coefficient is within a moderate range, the proposed method can form more stable and discriminative feature representations, thus exhibiting superior recognition capabilities. However, when the coefficient is too low or too high, the utilization of key information in the fusion process will be limited, thereby weakening the overall modeling effect. This indicates that gating fusion is not a simple information superposition process, but plays a crucial role in regulating the preservation of anomaly patterns and the injection of prior structures during the representation learning stage, and is therefore a key component supporting the performance improvement of the proposed method.

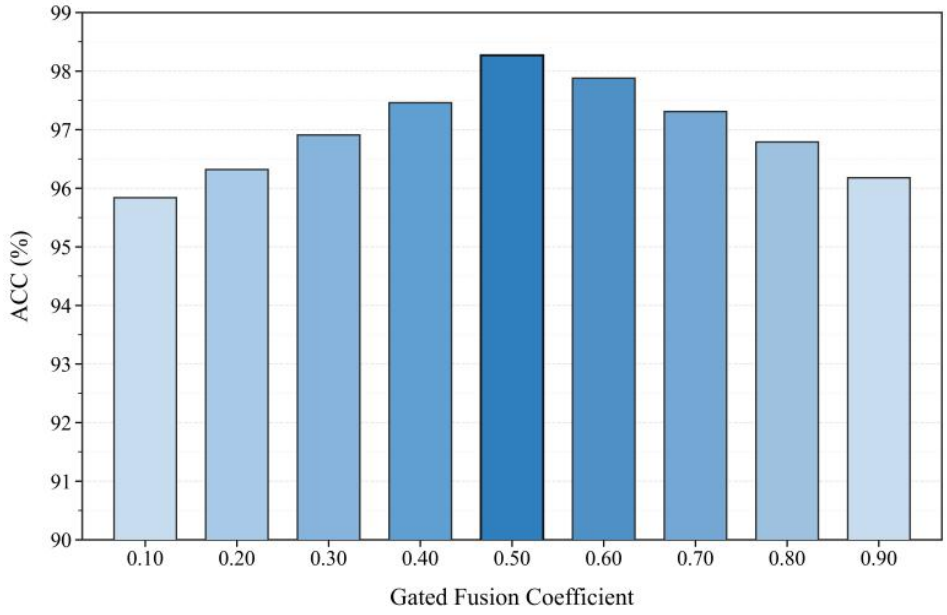

Figure 3. The impact of the gating fusion coefficient on Acc

4.5 The Impact of Latent Representation Dimension Setting on Acc

The latent representation dimension determines the model's ability to represent the semantic structure of backend logs, anomaly pattern differences, and the compression of prior information; it is one of the important factors affecting the overall recognition quality. If the dimension is set too low, key anomaly features and structural dependencies may not be fully preserved; if the dimension is set too high, redundant information is easily introduced, and the instability of the representation space is increased. Conducting sensitivity analysis around the latent representation dimension helps to more clearly characterize the balance between representational compactness and discriminative sufficiency in the proposed method, and the experimental results are shown in Figure 4.

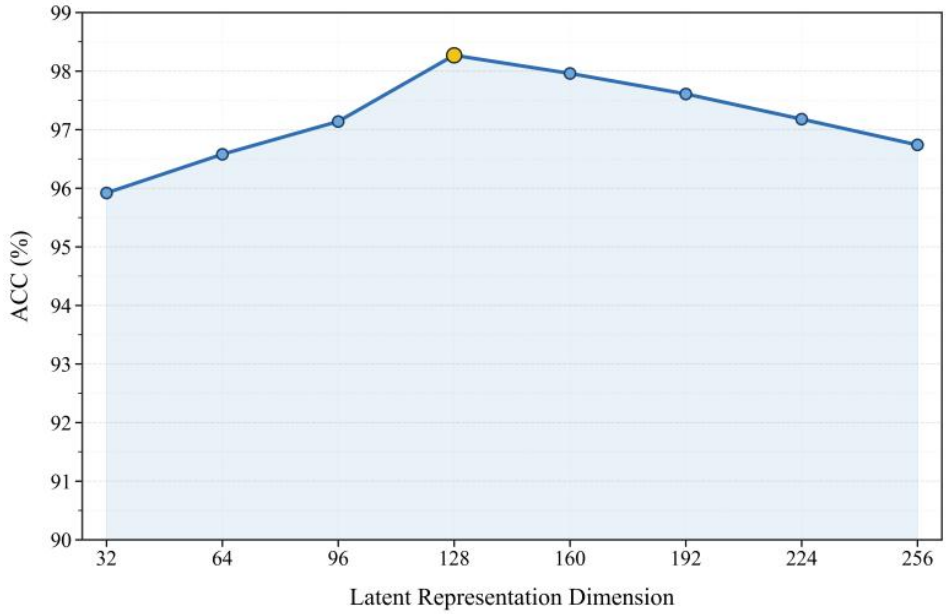

Figure 4. The impact of latent representation dimension setting on Acc

The appropriate setting of the latent representation dimension plays a crucial role in the quality of anomaly representation and the stability of recognition in our proposed method. With a suitable latent space configuration, the algorithm we present can more fully preserve the key semantic information and structural dependencies in the backend logs, enabling clearer distinction between anomalous and normal patterns in the representation space, thus demonstrating better recognition performance. This indicates that prior reinforcement representation learning not only needs to focus on the ability to extract anomalous information but also needs to consider the compactness and effectiveness of the representation space. Our proposed method demonstrates strong adaptability and practical application value in this regard.

5. Conclusion

This paper addresses the problem of backend anomaly detection under weak labeling and limited sample conditions, proposing a priori augmented representation learning and an adaptive recognition method. This aims to alleviate the problems faced by traditional anomaly detection models, such as unstable representations, ambiguous discrimination boundaries, and insufficient generalization ability, in scenarios with insufficient labeling, scarce anomaly samples, and complex log semantics. Focusing on the structural features and anomaly semantics of backend log sequences, the research unfolds from three levels: prior information guidance, representation space construction, and recognition mechanism optimization, forming a relatively complete anomaly modeling framework. This method organically combines prior-guided encoding, gating fusion mechanisms, and prototype-level metric calibration, enabling the model to learn more discriminative and stable latent representations even under weak supervision constraints, thereby improving the ability to recognize complex backend anomaly patterns. The results show that the proposed method can adapt well to the characteristics of strong log data heterogeneity, complex anomaly types, and uneven sample distribution in real-world backend environments, providing a research approach for backend log anomaly detection that has both theoretical significance and engineering value.

From an application perspective, this research has strong practical significance for intelligent operation and maintenance, cloud platform service assurance, distributed system stability maintenance, and risk warning for complex business systems. With the continuous development of cloud-native architecture, microservice collaboration, and automated operation and maintenance systems, the operating environment of backend systems is becoming increasingly dynamic and complex. Anomaly detection methods not only need higher recognition accuracy but also stronger adaptability and deployment value. The effectiveness of the proposed method under weak labeling and few-sample conditions provides valuable technical support for reducing manual labeling costs, improving anomaly detection efficiency, and enhancing the reliability of complex systems. Future research can further integrate multi-source heterogeneous operation and maintenance data to conduct unified modeling studies, achieving deeper fusion of logs, metrics, links, and event information to enhance the comprehensive perception of complex anomaly scenarios. Simultaneously, it can explore more adaptable, self-evolving, and interpretable anomaly identification mechanisms for novel anomalies, continuously evolving anomalies, and cross-system migration anomalies in open environments, thereby promoting the development of backend anomaly detection technology towards a more intelligent, robust, and universal direction.

References

[1] S. Hashemi and M. Mäntylä, "SiaLog: Detecting Anomalies in Software Execution Logs Using the Siamese Network," Automated Software Engineering, vol. 29, no. 2, Art. no. 61, 2022.

[2] J. Bogatinovski, G. Madjarov, S. Nedelkoski et al., "Leveraging Log Instructions in Log-Based Anomaly Detection," Proceedings of the 2022 IEEE International Conference on Services Computing, pp. 321-326, 2022.

[3] Y. Xie and K. Yang, "Log Anomaly Detection by Adversarial Autoencoders With Graph Feature Fusion," IEEE Transactions on Reliability, vol. 73, no. 1, pp. 637-649, 2023.

[4] C. Zhang, T. Jia, G. Shen et al., "MetaLog: Generalizable Cross-System Anomaly Detection From Logs With Meta-Learning," Proceedings of the IEEE/ACM 46th International Conference on Software Engineering (ICSE), pp. 1-12, 2024.

[5] S. Hashemi and M. Mäntylä, "OneLog: Towards End-to-End Software Log Anomaly Detection," Automated Software Engineering, vol. 31, no. 2, Art. no. 37, 2024.

[6] A. Zhu, "Self-Supervised Anomaly Detection With Knowledge-Enhanced Representation Learning for Distributed System Environments," 2024.

[7] Q. Shi, Z. Yang, A. Haidar et al., "Anomaly Transformer-Based System Log Anomaly Detection," Proceedings of the 2024 Cyber Awareness and Research Symposium, pp. 1-6, 2024.

[8] M. Du, F. Li, G. Zheng et al., "DeepLog: Anomaly Detection and Diagnosis From System Logs Through Deep Learning," Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security, pp. 1285-1298, 2017.

[9] W. Meng, Y. Liu, Y. Zhu et al., "LogAnomaly: Unsupervised Detection of Sequential and Quantitative Anomalies in Unstructured Logs," Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence (IJCAI), vol. 19, no. 7, pp. 4739-4745, 2019.

[10] X. Zhang, Y. Xu, Q. Lin et al., "Robust Log-Based Anomaly Detection on Unstable Log Data," Proceedings of the 2019 27th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering, pp. 807-817, 2019.

[11] L. Yang, J. Chen, Z. Wang et al., "Semi-Supervised Log-Based Anomaly Detection via Probabilistic Label Estimation," Proceedings of the 2021 IEEE/ACM 43rd International Conference on Software Engineering (ICSE), pp. 1448-1460, 2021.

[12] H. Guo, S. Yuan and X. Wu, "LogBERT: Log Anomaly Detection via BERT," Proceedings of the 2021 International Joint Conference on Neural Networks (IJCNN), pp. 1-8, 2021.

[13] X. Li, P. Chen, L. Jing et al., "SwissLog: Robust and Unified Deep Learning Based Log Anomaly Detection for Diverse Faults," Proceedings of the 2020 IEEE 31st International Symposium on Software Reliability Engineering (ISSRE), pp. 92-103, 2020.

[14] S. Huang, Y. Liu, C. Fung et al., "HitAnomaly: Hierarchical Transformers for Anomaly Detection in System Log," IEEE Transactions on Network and Service Management, vol. 17, no. 4, pp. 2064-2076, 2020.