Transactions on Computational and Scientific Methods | Vo. 4, No. 08, 2024
ISSN: 2998-8780
https://pspress.org/index.php/tcsm
Pinnacle Science Press

Federated Deep Learning with Contrastive Representation for Node State Identification in Distributed Systems

Di Wu
Washington University in St. Louis, St. Louis, USA
inbox.diwu@gmail.com

Abstract: This study proposes a distributed node state identification framework that integrates contrastive learning and federated optimization to address the heterogeneity, latency, and dynamism of distributed systems. At the local node level, the framework introduces contrastive feature constraints by constructing positive and negative sample pairs to achieve self-supervised alignment in the feature space, thereby enhancing representational consistency and discriminative capability within nodes. At the global level, it employs a federated optimization mechanism for parameter aggregation to enable collaborative learning and global consistency across nodes. The framework consists of four core stages: feature encoding, contrastive representation learning, federated aggregation, and consistency regulation, allowing efficient global identification while preserving data privacy. To verify its effectiveness, multidimensional sensitivity experiments were conducted, including analyses of hyperparameters, environmental factors, and data perturbations. The results show that the framework maintains stable performance under varying weight decay coefficients, temperature parameters, communication delays, and noise intensities, achieving significant improvements over traditional centralized and single-node models in accuracy, precision, recall, and F1-score. Further analysis demonstrates that the contrastive learning module effectively suppresses noise interference and feature drift, while the federated optimization mechanism mitigates data distribution bias among heterogeneous nodes, ensuring good convergence under high-latency and unbalanced conditions. This study confirms that the integrated learning strategy can balance feature robustness, model consistency, and computational efficiency, providing a feasible solution for building secure, stable, and efficient distributed intelligent identification systems.

Keywords: Distributed systems; node state determination; comparative learning; federated optimization

1. Introduction

In modern computing architectures, distributed systems play a vital role. With the rapid growth of cloud computing, edge computing, and the Internet of Things, the complexity, heterogeneity, and dynamic nature of system nodes have greatly increased[1]. Communication delays, resource disparities, and uneven task allocation among nodes can lead to performance fluctuations or even system failures. In such environments, accurately identifying and monitoring node states is essential for ensuring stability and optimizing resource scheduling. Traditional centralized monitoring or single-node optimization strategies are insufficient to handle non-independent and identically distributed data and time-varying node states. Node state identification involves not only computational performance but also network load, energy consumption, task congestion, and model drift. Therefore, building an intelligent, adaptive, and collaborative state discrimination framework holds significant theoretical and practical value[2].

The integration of artificial intelligence has brought new opportunities for intelligent distributed systems. Deep learning models can perform nonlinear mappings and temporal dependency modeling in complex multi-dimensional feature spaces, capturing the hidden structure of node behavior patterns. However, directly applying traditional models in distributed environments faces major challenges. First, data heterogeneity and privacy constraints limit the feasibility of centralized training. Second, high communication costs and inconsistent update frequencies among nodes hinder model convergence and may cause performance shifts. Third, frequent node state changes require models with continual learning and dynamic adaptation capabilities. Federated optimization, which enables "local data retention and collaborative model updating," has emerged as a promising solution for distributed learning. It protects data privacy while improving model generalization and stability through collaborative aggregation across nodes[3].

Despite these advantages, federated optimization still faces limitations in real-world applications. Data shifts, gradient conflicts, and inconsistent parameters across nodes make it difficult for a global model to represent local node characteristics accurately, reducing both accuracy and robustness. Moreover, task similarities and differences between nodes often contain valuable structural information. When aggregation relies solely on parameter averaging, critical features may be diluted or lost[4]. Contrastive learning, a leading self-supervised approach, can learn discriminative representations by aligning features between positive and negative sample pairs. Its principle of maximizing similarity between related samples while minimizing the distance between unrelated ones provides a pathway for unified representation learning in heterogeneous distributed data. This idea aligns naturally with the goal of federated optimization and offers a new perspective for addressing conflicts between decentralized features and global consistency[5].

Against this background, combining contrastive learning with federated optimization introduces an innovative paradigm for distributed node state identification. By embedding contrastive constraints within each node, the model learns robust local representations. Through global federated aggregation, it achieves cross-node feature sharing and alignment while maintaining privacy. This mechanism mitigates data heterogeneity among nodes and enhances model stability under concept drift, abnormal node behavior, and communication noise. Furthermore, the framework can be extended to multi-task collaboration and dynamic system management, offering broad adaptability and scalability.

Overall, the distributed node state identification framework that integrates contrastive learning and federated optimization reflects the evolution of intelligent computing from centralized to collaborative and autonomous paradigms. Theoretically, it promotes the integration of distributed intelligence learning paradigms. Practically, it provides new insights for large-scale computing systems, intelligent manufacturing, edge node monitoring, and autonomous network management. This research deepens the understanding of node behavior modeling in distributed systems and supports the development of secure, interpretable, and scalable intelligent systems. As system scale and data complexity continue to increase, this direction is expected to become a key foundation for intelligent infrastructure and efficient, reliable distributed decision-making.

2. Related work

The problem of node state identification in distributed systems has long been a central topic in intelligent computing and system management. Early studies mainly focused on centralized monitoring and statistical modeling. They collected indicators such as CPU utilization, memory usage, and network latency to build system state models for anomaly detection and resource prediction. However, as system scale expanded and heterogeneity increased, such centralized approaches exposed several limitations, including high data transmission overhead, vulnerability to single-point failures, and insufficient real-time performance. To address these challenges, distributed learning and edge intelligence frameworks have been introduced into node state modeling. These frameworks enable each node to perform local, data-driven model updates and feature extraction, thereby reducing dependence on the central node. Despite these advances, heterogeneous feature distributions and non-independent, non-identically distributed data among nodes still cause issues such as performance drift, aggregation distortion, and unstable global convergence. These challenges have become critical bottlenecks that must be overcome in subsequent research[6].

To address the challenges of distributed heterogeneous data, federated learning and federated optimization have become major research focuses. These methods train models locally on each node and then aggregate parameters globally, effectively balancing data privacy and communication efficiency. Subsequent research introduced adaptive weighted aggregation, multi-task collaborative optimization, and personalized federated models to improve generalization and individual performance under node heterogeneity. However, two main limitations remain. First, the global model has limited capacity to preserve local structural information, making it difficult to fully capture semantic differences in the feature space of each node. Second, under conditions of imbalanced data distribution and significant communication delays, gradient aggregation can easily fall into local optima or cause parameter oscillations. Therefore, achieving both representational consistency and discriminative capability under privacy constraints has become a crucial research direction in federated optimization.

Meanwhile, contrastive learning, an important branch of self-supervised learning, has shown great potential in distributed settings. Its core idea is to learn highly discriminative and structured latent representations by contrasting positive and negative sample pairs. Traditional supervised learning relies on large amounts of labeled data, while contrastive learning can uncover hidden semantic structures without labels by maximizing the similarity between related samples and minimizing that between unrelated ones. In recent years, contrastive learning has been widely applied to tasks such as temporal modeling, anomaly detection, cross-domain representation, and multi-task transfer, demonstrating strong generalization and adaptability. In distributed system node state identification, this mechanism can enhance the discriminative power of node representations, reduce feature bias caused by heterogeneous distributions, and improve the model's ability to distinguish node states with greater robustness[7].

Overall, although federated optimization and contrastive learning have achieved remarkable progress in privacy preservation and feature representation, respectively, their integration remains in the exploratory stage. Some existing studies have attempted to embed contrastive learning into federated learning frameworks to improve global aggregation through cross-node feature consistency constraints. However, most approaches still suffer from insufficient local information sharing, difficulty in cross-node feature alignment, and low communication efficiency. In particular, in complex node state identification scenarios, dynamic system changes, non-stationary data distributions, and asynchronous model updates make it challenging to effectively combine contrastive constraints with federated optimization strategies. Therefore, building a distributed node state identification framework that integrates contrastive learning and federated optimization can bridge the gap between feature consistency and global collaboration, offering a promising research pathway toward enhancing the reliability and autonomy of distributed intelligent systems.

3. Proposed Framework

3.1 Method Overview

This study proposes a distributed node state discrimination framework combining contrastive learning and federated optimization to achieve accurate modeling and robust identification of system node states. The framework comprises four core stages: local feature extraction and representation learning, contrastive feature constraints, global federated aggregation and optimization updates, and a consistent adaptive control mechanism. Each node in the system maintains data locality while enhancing the separability of its feature space using a contrastive learning mechanism and achieves consistent updates of the global model through federated aggregation. Let there be N nodes in the system, D_i be the data distribution of node i, θ_i be the corresponding model parameters, and θ_g be the global model parameters. The overall optimization objective can be formalized as:

$$\min_{\theta} \sum_{i=1}^{N} w_i L(D_i, \theta_i),\ subject\ to\ \theta_i \rightarrow \theta_g$$

Here, w_i represents the node weight, reflecting the amount of data or the importance of the node. This optimization process achieves global consistency under privacy constraints, enabling collaborative learning among nodes. The framework as a whole completes the discriminative modeling of distributed node states by alternately executing comparative feature constraints and federated aggregation. Its overall model architecture is shown in Figure 1.

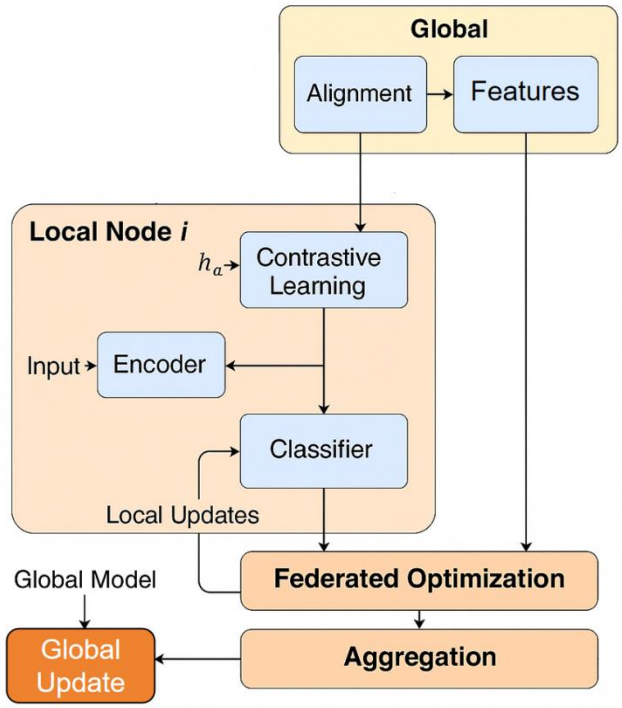

Figure 1. Overall model architecture

3.2 Local node feature modeling and comparison constraint mechanism

In each node, the model first encodes the input sequence $w_i = \{x_{i1}, x_{i2}, \dots, x_{iT}\}$, and then obtains the representation vector h_i through a temporal embedding layer and a nonlinear transformation. This process can be represented as:

$$h_i = f_{enc}(x_i; \theta_i)$$

To enhance the discriminative power of node representations, this study introduces local contrastive learning constraints. Let positive sample pairs $(h_i; h_i^+)$ come from the same node state, and negative sample pairs $(h_i; h_j^-)$ come from different nodes or different states. The temperature-scaled contrastive loss function is defined as follows:

$$L_{con} = -\log \frac{\exp(sim(h_i, h_i^+)/\tau)}{\sum_{j=1}^{K} \exp(sim(h_i, h_i^-)/\tau)}$$

Where $sim(\bullet, \bullet)$ represents cosine similarity and τ is the temperature coefficient. This mechanism can shorten the distribution distance between nodes of the same type and widen the gap between nodes of different types in the feature space, thereby providing a stable feature foundation for subsequent federated aggregation.

3.3 Federated optimization and global aggregation strategies

Global model parameters are aggregated using federated averaging. Let the global model parameters be θ_g, and the local parameters uploaded by the nodes be θ_i. Then, in each round of global updates, the following aggregation operation is performed:

$$\theta_t^{(t+1)} = \sum_{i=1}^{N} \frac{n_i}{\sum_{j=1}^{N} n_j} \theta_t^{(t)}$$

Where n_i is the number of samples at node i. To mitigate aggregation bias caused by node heterogeneity, an adaptive learning rate adjustment term is further introduced, making node updates follow the following form:

$$\theta_t^{(t+1)} = \theta_t^{(t)} - n_i \nabla \theta_i L_{total}(D_i)$$

Where $L_{total} = L_{cls} + \lambda L_{con}$ and λ are the weight coefficients of the comparison constraints. This mechanism ensures that the node model can be differentiated and optimized based on local data characteristics while maintaining global consistency, thereby achieving a more refined state representation.

3.4 Consistent regulation and optimized convergence mechanism

In a multi-node asynchronous update environment, models are prone to parameter drift and feature dispersion. To ensure overall system consistency and convergence stability, a global consistency constraint is introduced to minimize the distance between the local and global models in the feature space:

$$L_{align} = \frac{1}{N} \sum_{i=1}^{N} ||h_i - h_g||_2^2$$

Where A represents the global feature representation. The final joint optimization objective integrates the four loss terms of classification, comparison, alignment, and aggregation to form a complete multi-objective optimization function:

$$L_{align} = L_{align} + \lambda_1 L_{align} + \lambda_2 L_{align} + \lambda_3 L_{align}$$

Where $\lambda_1, \lambda_2, \lambda_3$ is the adjustment coefficient, used to balance the impact of different loss terms. By performing joint optimization updates after each round of communication and aggregation, the system achieves global consistency constraints and accelerated convergence of node-level features, enabling the distributed system to continuously and stably complete node state determination tasks in dynamic environments.

4. Experimental Analysis

4.1 Dataset

This study uses the NSL-KDD network intrusion detection dataset as the primary data source to verify the effectiveness of the distributed node state identification framework under multi-source heterogeneous features. The dataset is an improved version of the classic KDD Cup'99 dataset, with redundant samples and duplicate connection records removed, retaining only high-quality network traffic behavior features. It contains 41 input features, including transport layer protocols, connection duration, service type, error rates, and traffic statistics, which reflect the differences between normal and abnormal node states. The dataset labels include normal connections and four types of attacks (DoS, Probe, R2L, and U2R), which can be mapped to classification tasks representing different node states. This provides rich feature diversity for the joint modeling of contrastive learning and federated optimization.

In the distributed learning setting, the NSL-KDD dataset is divided into multiple node subsets to simulate the heterogeneity of data in multi-source environments. Each node contains distinct traffic patterns and attack ratios. Some nodes include a high proportion of abnormal samples, while others are dominated by normal traffic. This partitioning effectively represents the non-independent and non-identically distributed characteristics of nodes in distributed systems and provides representative experimental conditions for federated optimization strategies. Moreover, since the dataset has been widely used in anomaly detection and system monitoring research, its standardized feature format facilitates fair model comparisons and ensures reproducibility.

In addition, the dataset is preprocessed through normalization and class balancing, mapping features into a fixed-dimensional continuous space for subsequent encoding and contrastive constraint learning. By aligning feature spaces and modeling inter-node distribution differences, the framework achieves consistent feature discrimination across nodes while preserving local data privacy. The complex feature associations, data diversity, and hierarchical labeling structure of NSL-KDD provide a solid foundation for validating the proposed method's ability to identify node states under dynamic, heterogeneous, and non-stationary distributions.

4.2 Experimental Results

This paper first conducts a comparative experiment, and the experimental results are shown in Table 1.

Table 1: Comparative experimental results

Method	Acc	Precision	Recall	F1-Score
MLP[8]	0.8712	0.8425	0.8264	0.8343
1DCNN[9]	0.8847	0.8579	0.8415	0.8496
LSTM[10]	0.8965	0.8683	0.8530	0.8606
CNN-LSTM[11]	0.9081	0.8797	0.8672	0.8734
BILSTM[12]	0.9175	0.8881	0.8766	0.8823
Transformer[13]	0.9263	0.8972	0.8845	0.8908
Ours	0.9487	0.9254	0.9182	0.9218

As shown in Table 1, with the increasing complexity of model structures and the enhancement of feature representation capabilities, the performance of node state identification improves continuously. The traditional MLP model shows relatively low accuracy and recall, mainly because it cannot effectively capture temporal dependencies or dynamic variations among distributed nodes. It can only perform static feature-level discrimination. The 1D-CNN achieves certain improvements in local feature extraction and shows higher feature sensitivity, but it cannot model temporal relationships and global feature consistency across nodes. As a result, its stability under complex distribution conditions remains insufficient.

When temporal structures are introduced, the performance of LSTM and CNN-LSTM models improves significantly, indicating that temporal dependency modeling plays a crucial role in node state identification. The CNN-LSTM combines local convolutional features with temporal memory mechanisms, allowing the model to perceive both spatial correlations among nodes and the evolution of state transitions. This leads to better results in recall and F1-score. However, these models still struggle to maintain global consistency when facing non-independent and non-identically distributed node features and inter-node heterogeneity, which limits their generalization ability in distributed environments.

Furthermore, BiLSTM and Transformer models exhibit stronger robustness in modeling global dependencies. BiLSTM captures both forward and backward dependencies of node states through its bidirectional propagation mechanism, making it more sensitive to state changes. The Transformer, with its multi-head attention mechanism, enables global feature weighting and dynamic interaction, which significantly improves overall accuracy and generalization. However, these models still rely on centralized data synchronization in distributed system scenarios and cannot effectively handle data heterogeneity and communication constraints among nodes.

In contrast, the proposed distributed node state identification framework that integrates contrastive learning and federated optimization achieves the best performance across all metrics. This method enhances the consistency and discriminability of representations through local contrastive feature constraints within nodes. At the global level, it applies federated aggregation to achieve collaborative optimization across nodes. As a result, it improves accuracy, recall, and F1-score simultaneously while preserving data privacy. The outstanding performance demonstrates that the framework can effectively capture implicit relationships among nodes and achieve robust state recognition in heterogeneous and non-stationary distributed systems. It provides a new perspective for adaptive monitoring and optimization in distributed intelligent systems.

This paper also presents an experiment on the sensitivity of the weight decay coefficient to the accuracy of node state discrimination, and the experimental results are shown in Figure 2.

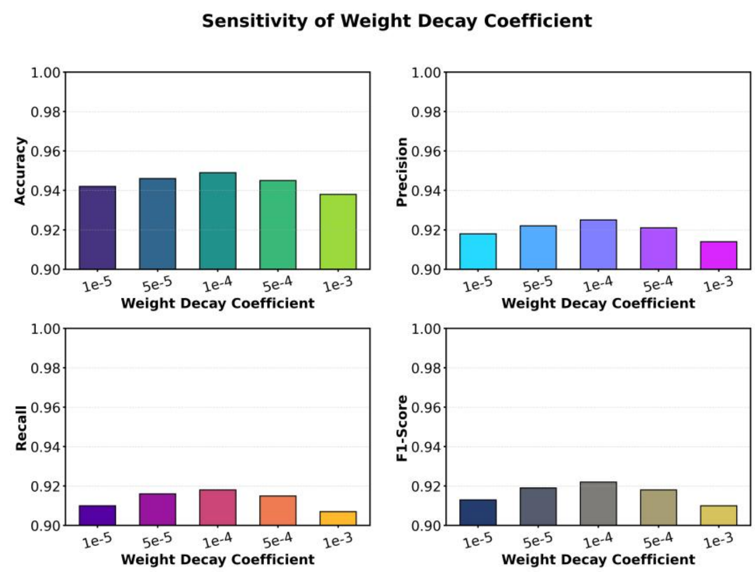

Figure 2. Sensitivity experiment of the weight decay coefficient to node state discrimination accuracy

As shown in Figure 2, the weight decay coefficient has a certain sensitivity to the overall performance of node state identification. When the weight decay is too small, the model tends to overfit, resulting in poor generalization across global node distributions. When the weight decay is too large, the model update is overly constrained, weakening its ability to learn local features and making it difficult to capture subtle dynamic differences among nodes. The experimental results show that when the weight decay coefficient is set to 1×10^{-4} , all evaluation metrics reach a relatively optimal balance. At this point, the model avoids overfitting while maintaining strong feature representation capability and stable convergence performance.

In terms of accuracy, the model shows a gradual improvement in the smaller decay range (1×10^{-5} to 5×10^{-5}). It reaches its peak performance in the moderate decay range, after which accuracy slightly decreases as the constraint becomes stronger. This indicates that a moderate parameter penalty helps the model form a more discriminative embedding space during the contrastive learning stage. As a result, the state features among the nodes remain more consistent in the global representation. However, an excessively large decay restricts gradient updates and limits the model's ability to adapt to non-stationary node state changes.

From the variation trends of precision and recall, both metrics show strong synchronization near the optimal point. With a reasonable increase in weight decay, the model's ability to distinguish positive and negative samples improves. This suggests that contrastive learning enhances positive sample aggregation and negative sample separation within local nodes. When the decay becomes too large, the model becomes overly smoothed in the local feature space, reducing its ability to discriminate boundary samples, which in turn decreases recall. This result confirms that the proposed framework effectively balances feature constraints and global consistency.

Considering the F1-score results, the model achieves its best overall performance when the weight decay coefficient is 1×10^{-4} . This indicates that under this configuration, the collaborative effect between federated optimization and contrastive learning modules is strongest. The model not only maintains feature discriminability among nodes but also achieves a unified representation space through global aggregation. The overall findings demonstrate that an appropriate weight decay effectively balances model complexity and generalization, enabling stable and high-accuracy node state identification in heterogeneous distributed environments.

This paper further analyzes the impact of the temperature parameter in contrastive learning on model performance, as shown in Figure 3. Specifically, it examines how different temperature values influence feature similarity distribution, optimization stability, and representation discrimination within the contrastive space. The results demonstrate that the temperature parameter plays a critical role in balancing intra-class compactness and inter-class separability, directly affecting the accuracy and robustness of node state discrimination in the distributed learning framework.

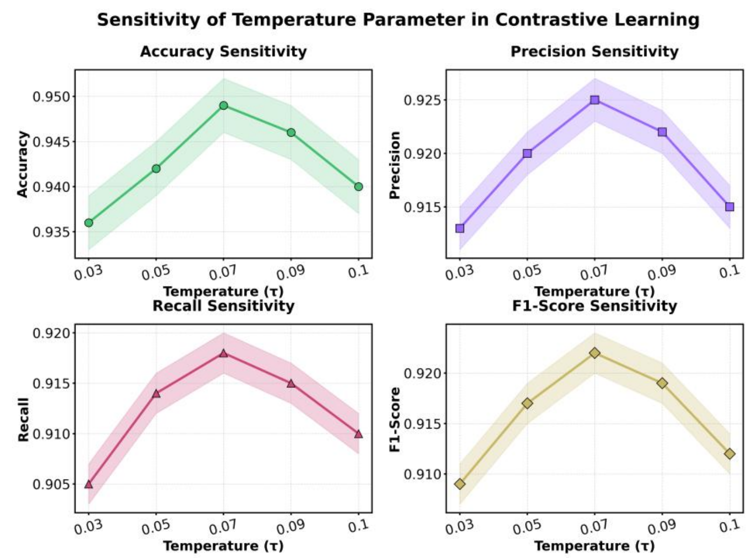

Figure 3. Compare the effects of temperature parameters on experimental results

As shown in Figure 3, the temperature parameter in contrastive learning has a significant impact on the overall performance of node state identification. When the temperature is too low (for example, 0.03), the model forces samples too close together in the feature space. This reduces inter-class separability and affects both classification accuracy and feature stability. As the temperature increases appropriately, the contrast tension among features becomes stronger, allowing the model to form clearer boundaries between similar samples. This leads to improved performance across all metrics. When the temperature reaches 0.07, both accuracy and F1-score achieve their highest values, indicating that the balance between positive and negative sample similarity is optimal. At this point, the model maintains compact feature representations while achieving good global separability.

From the trends of precision and recall, it can be seen that temperature adjustment plays a key role in balancing model overfitting and excessive feature separation. A lower temperature causes positive and negative pairs to cluster too tightly in high-dimensional space, making it easy for the model to confuse boundary samples. Conversely, an excessively high temperature introduces noisy features, leading to insufficient aggregation of positive samples. The results show that within the range of 0.05 to 0.09, the precision and recall curves rise and fall almost synchronously. This indicates that within this range, the model can effectively balance feature discrimination and generalization, thereby improving the stability of node state recognition.

Overall, an appropriate temperature parameter can significantly enhance the discriminative power of the contrastive learning module, enabling more stable representation learning in distributed environments. When the temperature is too low, the feature distribution becomes blurred, making it difficult to distinguish node differences. When the temperature is too high, the consistency of the feature structure is disrupted, reducing the generalization ability. The experiments demonstrate that selecting a suitable temperature parameter strengthens the model's balance between global consistency and local diversity under heterogeneous node features, thereby improving the overall accuracy and robustness of node state identification.

This paper also presents the impact of communication delay on the experimental results, which are shown in Figure 4.

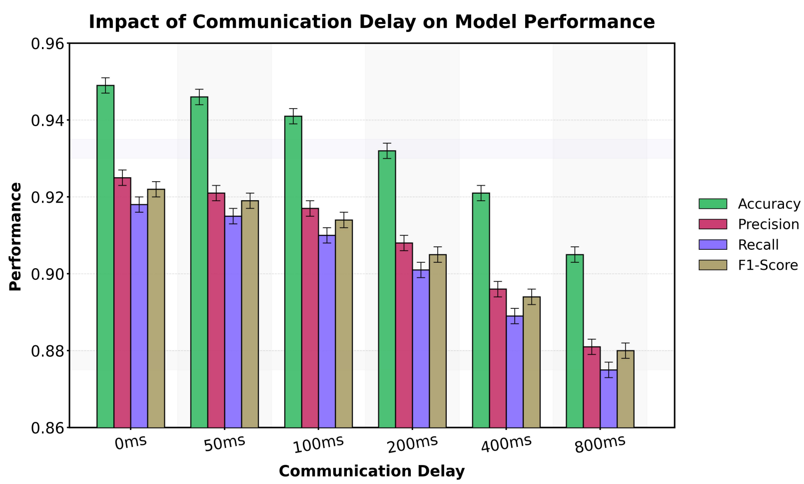

Figure 4. The impact of communication delay on experimental results

As shown in Figure 4, communication delay has a clear negative impact on the performance of distributed node state identification. As the delay increases, all evaluation metrics show a downward trend. When the delay rises from 0 ms to 200 ms, both accuracy and F1-score decline significantly. This indicates that asynchronous communication causes instability in parameter aggregation and delays in local updates, weakening the consistency of the global model. When the delay exceeds 400 ms, the performance drop becomes more severe, suggesting that federated optimization can no longer effectively synchronize gradient information across nodes. As a result, the global feature representation deviates from the optimal distribution. This phenomenon confirms the dual impact of communication delay on the convergence speed and stability of model training in distributed systems.

From the variations in precision and recall, it can be observed that under low-delay conditions, information exchange between nodes is sufficient. The model can accurately distinguish node states and maintain high classification precision. As the communication delay increases, some nodes update more slowly, leading to instability in the global model when identifying boundaries and noisy samples. This imbalance in asynchronous updates causes overfitting on certain nodes and underfitting on others, reducing the overall recall. These results show that in asynchronous environments, delay affects not only gradient synchronization but also the structural balance of the feature space, which further impacts the model's generalization performance.

Overall, this experiment demonstrates that communication delay is a key factor affecting the stability of federated optimization. Under moderate delay, the model can still maintain high performance, but when the delay exceeds the system's adaptive threshold, global consistency drops sharply. The proposed framework maintains a relatively stable performance curve under these conditions, benefiting from the enhanced local feature robustness provided by the contrastive learning module and the adaptive update mechanism in global aggregation. By incorporating consistency regulation and delay-aware optimization mechanisms, the framework can further improve robustness and generalization in high-delay distributed environments, providing practical guidance for real-world system deployment.

This paper also presents an experiment on the sensitivity of noise injection intensity to the experimental results, and the experimental results are shown in Figure 5.

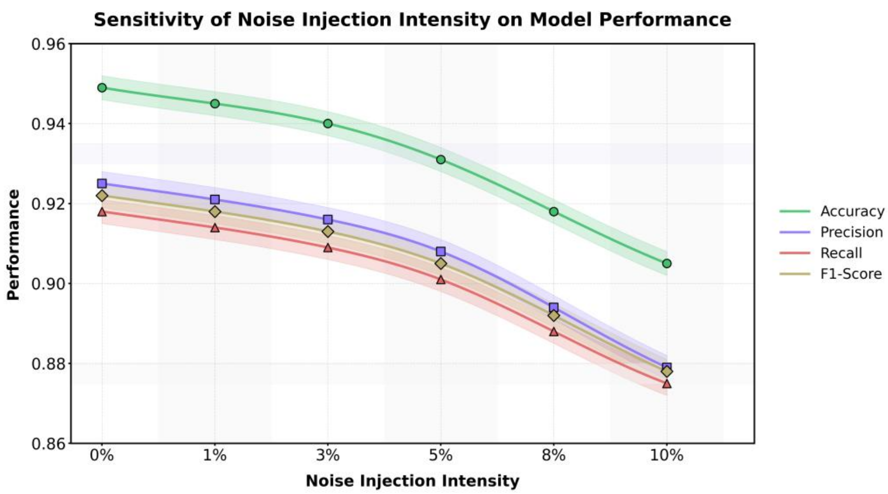

Figure 5. Sensitivity experiment of noise injection intensity to experimental results

As shown in Figure 5, the intensity of noise injection has a clear negative impact on the performance of node state identification. As the noise ratio increases, all performance metrics show a downward trend, with accuracy and F1-score declining most significantly. When the noise level is low (0%–3%), the model maintains high stability and classification accuracy, indicating that the contrastive learning mechanism can effectively suppress feature deviation caused by noise disturbance at this stage. However, when the noise intensity exceeds 5%, the feature space becomes disrupted, the semantic consistency between nodes weakens, and the model can no longer maintain clear boundaries between normal and abnormal states. As a result, overall performance drops sharply.

From the trends of precision and recall, it can be observed that noise injection causes instability in the feature distribution, reducing the model's ability to distinguish between positive and negative samples. When the noise level is low, feature representations remain relatively compact, and recall is slightly higher than precision, meaning the model can still capture most abnormal nodes. In high-noise environments, excessive disturbance blurs the feature clustering boundaries, and precision decreases significantly, indicating a higher false detection rate when identifying abnormal nodes. This feature drift reflects the destructive effect of noise on the structure of the contrastive learning embedding space. It also highlights the importance of data quality at distributed nodes for the convergence of the global model.

Overall, this experiment shows that noise intensity is one of the key factors affecting the accuracy of distributed node state identification. Low-intensity noise can be absorbed by the model's adaptive regularization, with limited impact on performance. High-intensity noise, however, leads to local feature distortion and global aggregation deviation, weakening the stability of federated optimization. The proposed framework maintains a relatively smooth decline in performance even under high-noise conditions, benefiting from the robustness of contrastive feature learning and the effectiveness of the consistency regulation mechanism. These results indicate that the model retains strong resistance to perturbations and good generalization when dealing with noisy and heterogeneous distributed data.

5. Conclusion

This study proposes an intelligent identification framework that integrates contrastive learning and federated optimization to address the complexity and heterogeneity of node state identification in distributed systems. The method introduces feature contrast constraints within local nodes to enhance model consistency and robustness. At the global level, it applies a federated aggregation mechanism to achieve collaborative optimization across multiple nodes, thus balancing privacy protection and performance improvement. Experimental results show that the proposed framework demonstrates superior stability and generalization under various sensitivity conditions. It maintains convergence and accuracy of the global model effectively under communication delay, noise interference, and hyperparameter variations. The findings confirm that the integrated learning mechanism significantly improves the system's adaptability to dynamic node features under heterogeneous and non-stationary data distributions.

From an application perspective, the proposed method holds significant value for intelligent computing, cloud-edge collaboration, and adaptive system management. By combining the feature constraints of contrastive learning with the collaborative update mechanism of federated optimization, the model achieves global information sharing under data isolation while maintaining efficient state recognition performance with privacy protection. This design can be widely applied to tasks such as distributed anomaly detection, edge device health monitoring, financial risk identification, intelligent manufacturing networks, and dynamic scheduling in multi-agent systems. It provides a scalable technical pathway for stable operation and resource optimization in large-scale complex systems. Moreover, the framework offers new theoretical insights for building self-learning and adaptive feature selection mechanisms, laying the foundation for collaborative modeling among distributed intelligent agents.

Future research can further extend this work in the directions of model lightweighting, adaptive communication strategies, and cross-domain collaborative learning. On one hand, it is possible to reduce communication and computation costs while maintaining model performance to improve real-time capability in large-scale distributed scenarios. On the other hand, dynamic temperature adjustment and multi-level feature contrast mechanisms can be explored to enhance model adaptability under non-stationary and drifting environments. In addition, integrating this framework with reinforcement learning and graph neural networks can enable more fine-grained node dependency modeling and cross-network collaborative optimization. With the rapid development of distributed intelligent systems, the concepts and outcomes of this study are expected to provide strong support for the next generation of robust, interpretable, and adaptive distributed learning systems.

References

[1] Li Q, He B, Song D. Model-contrastive federated learning[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021: 10713-10722.

[2] Nardi M, Valerio L, Passarella A. Anomaly detection through unsupervised federated learning[C]//2022 18th International Conference on Mobility, Sensing and Networking (MSN). IEEE, 2022: 495-501.

[3] Sáez-de-Cámara X, Flores J L, Arellano C, et al. Clustered federated learning architecture for network anomaly detection in large scale heterogeneous IoT networks[J]. Computers & Security, 2023, 131: 103299.

[4] Cai J, Zhang Y, Fan J, et al. Lg-fgad: An effective federated graph anomaly detection framework[C]//Proceedings of the International Joint Conference on Artificial Intelligence. 2024: 3760-3769.

[5] Tan X, Cheng J, Li H, et al. Contrastive Learning for Network Intrusion Detection: A Comprehensive Survey[C]//Proceedings of the 2024 2nd International Conference on Computer, Internet of Things and Smart City. 2024: 160-166.

[6] Li, Q., He, B., & Song, D. (2021). Model-contrastive federated learning. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 10713-10722).

[7] Karimireddy, S. P., Kale, S., Mohri, M., Reddi, S., Stich, S. and Suresh, A. T., "SCAFFOLD: Stochastic controlled averaging for federated learning", Proceedings of the International Conference on Machine Learning, pp. 5132-5143, November 2020.

[8] Diro, A. A. and Chilamkurti, N., "Distributed attack detection scheme using deep learning approach for Internet of Things", Future Generation Computer Systems, vol. 82, pp. 761-768, 2018.

[9] Qazi E U H, Almorjan A, Zia T. A one-dimensional convolutional neural network (1D-CNN) based deep learning system for network intrusion detection[J]. Applied Sciences, 2022, 12(16): 7986.

[10] Yang, X. S., "Firefly algorithms for multimodal optimization", Proceedings of the International Symposium on Stochastic Algorithms, pp. 169-178, October 2009.

[11] Gueriani A, Kheddar H, Mazari A C. Enhancing iot security with cnn and lstm-based intrusion detection systems[C]//2024 6th International Conference on Pattern Analysis and Intelligent Systems (PAIS). IEEE, 2024: 1-7.

[12] Jouhari M, Guizani M. Lightweight cnn-bilstm based intrusion detection systems for resource-constrained iot devices[C]//2024 International Wireless Communications and Mobile Computing (IWCMC). IEEE, 2024: 1558-1563.

[13] Manocchio L D, Layeghy S, Lo W W, et al. Flowtransformer: A transformer framework for flow-based network intrusion detection systems[J]. Expert Systems with Applications, 2024, 241: 122564.