Journal of Computer Technology and Software

ISSN:2998-2383

Vol. 4, No. 5, 2025

Adaptive Gradient Scaling for Federated Learning with Non-IID Data and Privacy Preservation

Elodie Carver
Thompson Rivers University, Kamloops, Canada
elodie.carver998@tru.ca

Abstract: Federated Learning (FL) shows great application potential in distributed data modeling. It can achieve cross-device or cross-organization collaborative training while protecting data privacy. However, the heterogeneous data distribution (Non-IID) in real applications makes traditional federated optimization methods face problems such as slow convergence, model drift, and performance degradation. In addition, there is still a potential risk of privacy leakage during gradient transmission and model update, which affects the security and scalability of federated learning. In response to these challenges, this study proposes an adaptive gradient scaling (AGS) scheme to optimize the convergence of federated learning on non-independent and identically distributed data, and combines differential privacy (DP) and secure aggregation to improve privacy protection capabilities. The experiment is verified based on the LEAF federated learning dataset. The results show that the AGS scheme can effectively improve the model convergence speed, improve the final accuracy, and enhance the training stability. It has achieved significant performance improvements on mainstream federated optimization methods such as FedAvg, FedProx, and Scaffold. In addition, this study further analyzes the adaptability of AGS in different heterogeneous data environments and explores its potential application value in fields such as healthcare, finance, and edge computing. This study provides a new methodology for optimizing federated learning in complex data distribution environments and promotes its efficient deployment in privacy-sensitive scenarios.

Keywords: Federated learning, non-IID, adaptive gradient scaling, privacy protection

1. Introduction

With the increasingly stringent data privacy protection regulations and the rapid development of distributed computing, Federated Learning (FL) has become an important technology to solve the data island problem and privacy protection challenges. In the traditional machine learning paradigm, model training usually relies on centralized data storage. However, this approach faces problems such as data privacy leakage, high communication costs, and high storage pressure. Federated learning distributes computing tasks to various clients, allowing data to be stored locally and model training to be performed, and only sharing model parameters or gradients, thereby effectively protecting data privacy[1]. However, in actual application scenarios, data distribution between different data sources often has significant heterogeneity, including statistical heterogeneity (uneven data distribution), system heterogeneity (differences in computing power and communication bandwidth), and privacy heterogeneity (privacy protection requirements of different nodes). These problems limit the generalization ability and stability of federated learning[2]. Therefore, it is of great theoretical value and application significance to study how to optimize the training effect of federated learning under heterogeneous data distribution and enhance the privacy protection mechanism at the same time[3].

In real-world federated learning scenarios, data heterogeneity is inevitable. For example, in the field of healthcare, the distribution of patient data in different hospitals may vary due to differences in region, population structure and equipment; in financial risk control, the data of each bank or financial institution is affected by factors such as customer groups and business models, resulting in uneven data distribution; in the application of smart mobile devices, the behavioral data of different users also have personalized characteristics. The uneven distribution of these data will make it difficult for traditional federated learning algorithms to converge, and even cause problems such as model performance degradation and unstable client training[4]. Existing federated optimization methods, such as FedAvg, often find it difficult to guarantee the model convergence effect of all clients when facing highly heterogeneous data distribution. Therefore, how to design a more robust and efficient federated optimization strategy to improve model performance in a heterogeneous data environment is an important direction of current federated learning research.

Privacy protection is one of the core advantages of federated learning, but existing methods still face many challenges. Although federated learning avoids the centralized storage of raw data, there may still be a risk of information leakage during parameter transmission and gradient sharing. For example, through the Model Inversion Attack, the attacker can use the published gradient information to infer the original data of the client; Differential Privacy (DP) can effectively

reduce the possibility of privacy leakage, but it may affect the model accuracy under high noise conditions. In addition, in the cross-device or cross-organization federated learning framework, the privacy requirements of different nodes are different, so it is necessary to design an adaptive privacy protection mechanism to achieve a dynamic trade-off between privacy protection and model performance. Current research attempts to combine Secure Multi-Party Computation (SMPC) and Homomorphic Encryption (HE), but these methods often have high computational overhead, which limits their application in resource-constrained environments. Therefore, how to optimize the communication efficiency and computing performance of federated learning while ensuring privacy security is still an urgent problem to be solved[5].

In order to improve the adaptability of federated learning in heterogeneous data distribution and privacy protection, this study will focus on two optimization strategies. First, in terms of optimizing heterogeneous data distribution, an adaptive local update mechanism is adopted, combined with personalized model training (Personalized FL) and adaptive gradient adjustment, to improve the convergence and generalization ability of federated learning in non-independent and identically distributed (Non-IID) data environments. Secondly, in terms of privacy protection, a hybrid strategy of differential privacy + federated distillation is introduced to reduce the privacy risk brought by gradient sharing through decentralized knowledge distillation, and at the same time, a variable noise strategy is combined to balance privacy and model accuracy. In addition, this study will also conduct experimental verification in two typical scenarios of cross-device federated learning (Cross-Device FL) and cross-organizational federated learning (Cross-Silo FL) to evaluate the effectiveness of the optimization strategy in different application environments.

The significance of this study is to combine federated optimization with privacy protection technology to improve the applicability of federated learning in heterogeneous data environments and reduce the risk of privacy leakage. The research results not only help to improve the practical application value of federated learning and make it more widely implemented in fields such as medical care, finance, and mobile devices, but also provide new solutions for secure and controllable distributed machine learning. In the future, with the combination of federated learning with blockchain, trusted execution environment (TEE) and other technologies, privacy protection and heterogeneous data optimization will be further developed, providing a more efficient and scalable framework for data sharing and secure computing. This study will provide new theoretical support and technical paths for the application of federated learning in data security and personalized intelligent services[6].

2. Literature Review on Federated Learning in Non-IID and Privacy-Sensitive

Federated learning has emerged as a critical solution to address distributed training in privacy-sensitive environments. To enhance model generalization under data heterogeneity, several works have incorporated temporal and multi-source features using transformer-based architectures, particularly in medical text analysis [7]. In distributed systems, graph neural networks have enabled collaborative perception for adaptive task scheduling [8], while deep neural networks have been employed for proactive fault prediction in time-series learning [9].

Automated feature extraction combined with transformer-based models has demonstrated strong forecasting ability for multivariate time series in non-IID data environments [10]. Reinforcement learning techniques such as TD3 have been utilized for continuous control and dynamic load balancing in federated systems [11], and spatiotemporal learning has further improved memory usage prediction accuracy in cloud servers [12].

In user behavior modeling, deep probabilistic approaches such as mixture density networks have been shown effective for anomaly detection [13], and capsule networks have enabled structured representation learning on complex data [14]. Reinforcement-driven decision-making has also been introduced in distributed scheduling frameworks, allowing topology-aware policies in heterogeneous environments [15]. Lightweight architectures such as MobileNet with compression and edge strategies have provided low-latency solutions in mobile federated applications [16].

Meanwhile, LLM-driven systems have tackled semantic detection and control in adversarial scenarios like phishing detection [17], and fine-grained access control based on semantic context modeling has helped regulate LLM outputs [18]. For knowledge structuring, memory-aware modeling has been employed to enhance data retention in LLMs [19].

Reinforcement learning-controlled subspace ensemble sampling has improved efficiency in handling complex data structures in federated settings [20], and knowledge transfer methods have extended LLM capabilities for low-resource text generation tasks [21]. Reinforcement-based fine-tuning strategies also support structured preference modeling in policy learning frameworks [22].

To improve collaborative learning, policy structuring guided by knowledge has been introduced in multi-agent systems [23]. Parameter coordination has been addressed via spectral decomposition for optimizing fine-tuning processes [24], while retrieval-based context modeling has enhanced generation quality in retrieval-augmented generation (RAG) models [25].

For robust classification, transformer architectures with dual-loss strategies have demonstrated improvements in few-shot learning settings [26]. DeepSORT-based visual tracking methods have supported gesture recognition in interactive systems [27], and dynamic low-rank fine-tuning has enabled flexible model adaptation for few-shot scenarios [28]. Financial fraud detection has benefited from fusion frameworks leveraging LLMs [29].

Sampling methods with contextual awareness have improved data acquisition in intelligent systems using DQN strategies [30], while pre-trained language models and few-shot learning approaches have contributed to accurate medical entity extraction [31]. Hierarchical term relationship investigation has further enriched LLM interpretability in structured knowledge systems [32].

Joint retrieval and external knowledge incorporation have strengthened harmful text detection frameworks [33], and smarter low-rank adaptation techniques have improved LLM fine-tuning efficiency under the LoRA framework [34]. Contextual modeling with BERT-BiLSTM has been used for malicious comment detection [35], and LongFormer-based methods have improved the summarization of long medical texts [36].

3. Adaptive Optimization Strategies for Federated Learning

In view of the challenges of federated learning in heterogeneous data distribution and privacy protection, this study proposes an adaptive optimization strategy to improve the convergence of the model in a non-independent and identically distributed (Non-IID) data environment and enhance the privacy protection mechanism[37]. The model architecture is shown in Figure 1.

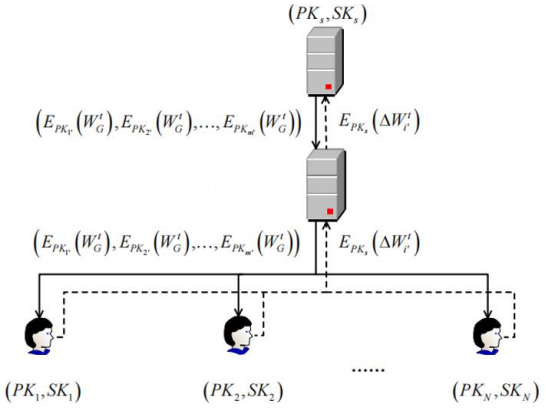

Figure 1. Federated Learning Model Architecture

First, in order to solve the model bias problem caused by data heterogeneity, we introduced an adaptive local update mechanism based on the traditional Federated Averaging (FedAvg) training framework. In the standard FedAvg, the core formula for updating the federated model is:

$$w_{t+1} = w_t - \eta \sum_{i=1}^{N} \frac{n_i}{N} \nabla F_i(w_t)$$

Among them, w_t represents the parameters of the global model, η is the learning rate, N is the total number of clients, n_i is the amount of data of the i-th client, and $\nabla F_i(w_t)$ represents the gradient calculated by client i. However, under heterogeneous data distribution, the loss function F of different clients may be quite different, resulting in excessive influence of some clients in the global update process. To this end, we introduced an adaptive gradient scaling (AGS) strategy to enable dynamic adjustment of gradient updates of different clients:

$$g'_i = \frac{\nabla F_i(w_t)}{\| \nabla F_i(w_t) \| + \varepsilon} \cdot a_i$$

Among them, g'_i is the normalized gradient, ε is the smoothing term, and a_i is the adaptability coefficient of the client. It is adaptively adjusted according to the convergence speed of the local model to ensure a more balanced gradient update and improve the robustness of the model on heterogeneous data[38].

In terms of privacy protection, this study uses differential privacy (DP) combined with federated distillation (FD) to reduce the risk of information leakage during gradient transmission. Under the traditional DP mechanism, noise is added when the parameters of the federated model are updated to limit attackers from inferring client data through gradients. The formula is as follows:

$$w_{t+1} = w_t - \eta \sum_{i=1}^{N} \frac{n_i}{N} (\nabla F_i(w_t) + N(0, \sigma^2))$$

$N(0, \sigma^2)$ represents Gaussian noise, which is used to protect client data privacy. However, high noise levels may affect model performance, so we further combine federated distillation to reduce the reliance on gradients through knowledge distillation. Specifically, each client trains a lightweight sub-model S_i using local data, and then sends its soft labels to the server instead of the original gradients:

$$T_i = \text{Softmax}(\frac{S_i(x)}{\tau})$$

Among them, T_i is the knowledge distillation information sent by client i, and τ is the distillation temperature parameter, which can adjust the smoothness of knowledge transfer. The server uses the distillation information of all clients for integration, thereby reducing the risk of privacy leakage while maintaining the stability of the global model.

In addition, to further optimize the adaptability of federated learning in heterogeneous computing environments, we introduced a hierarchical federated optimization strategy (HFO) to reduce communication costs and improve computing efficiency. In the traditional federated learning framework, all clients communicate directly with the server, resulting in high communication overhead in large-scale scenarios. To this end, we adopt a two-level aggregation mechanism, first performing local model aggregation on the local device side (such as mobile devices):

$$w_t^{local} = \sum_{j \in Ci} \frac{n_j}{N_i} w_j$$

Among them, C_j is the local client cluster that device i participates in, and N_i is the total data volume of the cluster.

Then, the server performs global aggregation from all local models:

$$w_{t+1}^{global} = \sum_{j \in Ci} \frac{N_j}{N} w_i^{local}$$

This can reduce the communication burden on the server side and improve the scalability of federated learning in large-scale scenarios. The experimental part of this study will evaluate the effectiveness of this optimization strategy in a cross-device and cross-organizational federated learning environment.

4. Experimental Setup and Dataset Description

This study uses the LEAF (A Benchmark for Federated Settings) dataset for experiments. This dataset is specifically used to evaluate the performance of federated learning algorithms in non-independent and identically distributed (Non-IID) environments. The LEAF dataset covers multiple sub-datasets, including FEMNIST (handwritten character recognition), Sent140 (social media text analysis), Shakespeare (natural language processing), CelebA (face attribute recognition), etc., which can simulate federated learning tasks in different scenarios. Since the data in the LEAF dataset comes from multiple independent clients and the data distribution has significant heterogeneity, it can well reflect the challenges of federated learning in the real world, such as different user data distributions on different devices and uneven client computing power.

This study mainly selects FEMNIST and Sent140 as experimental data. FEMNIST (Federated Extended MNIST) is an extension of the classic MNIST dataset, in which handwritten characters are collected from multiple different users, and the data distribution of each user is different, which conforms to the non-independent and identically distributed (Non-IID) characteristics of federated learning. Sent140 is a text sentiment analysis dataset based on Twitter, in which each client corresponds to a user's tweet data, which contains obvious personalized features and is suitable for studying personalized federated learning (Personalized FL) and privacy protection mechanisms. Due to the strong distribution heterogeneity of these datasets, the optimization strategy proposed in this study can be effectively evaluated in terms of dealing with data imbalance, personalized training, and privacy protection[39].

During the experiment, we performed standardized preprocessing on the dataset and used data partitioning to simulate different types of non-independent and identically distributed (Non-IID) environments. Specifically, we used three partitioning strategies: label imbalance, number imbalance, and feature distribution imbalance to construct federated learning tasks with different data distribution characteristics. In addition, in order to evaluate the effectiveness of the privacy protection mechanism, we introduced differential privacy (DP) perturbations on some client data to test the impact of privacy protection on model performance. Ultimately, the experiments on this dataset will help verify the applicability and robustness of adaptive federated optimization strategies and privacy protection methods in real-world non-independent and identically distributed data environments.

5. Evaluation and Result Analysis

This paper first compares the convergence of different federated optimization strategies (FedAvg, FedProx, Scaffold) in a heterogeneous data environment. The experimental results are shown in Figure 2.

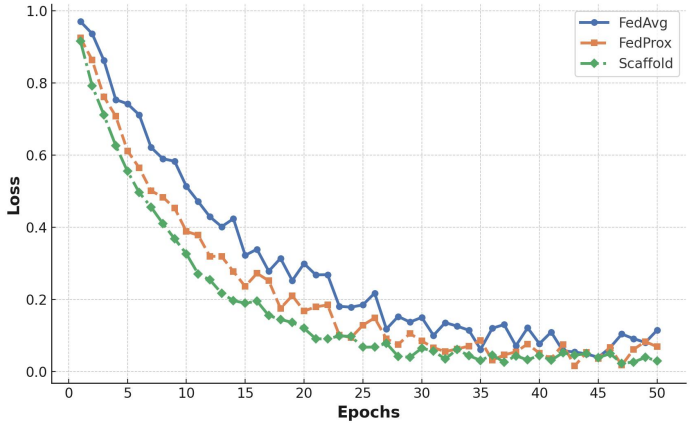

Figure 2. Comparison of Federated Optimization Strategies in Heterogeneous Data Environments

From the experimental results, it can be seen that different federated optimization strategies have obvious differences in convergence performance in heterogeneous data environments. Among them, FedAvg has the slowest loss decline rate and large convergence fluctuations, indicating that it is difficult to stably optimize in non-independent and identically distributed (Non-IID) data scenarios. Since FedAvg adopts a simple global parameter averaging strategy, when the data distribution of each client is uneven, the updates of some clients may have too much impact on the overall model, resulting in unstable model training. In addition, in the early stage of training, FedAvg declines rapidly, but with the increase of training rounds, its convergence speed slows down significantly, and there are large fluctuations in the later stage, indicating that its adaptability to heterogeneous data is weak.

In contrast, FedProx performs more stably throughout the training process and converges faster than FedAvg. FedProx constrains the model update amplitude of the client by adding regularization terms in the local optimization process, so that the global model can be optimized more smoothly between different data distributions. Therefore, its loss curve declines faster than FedAvg and maintains lower volatility in the later stage. In addition, FedProx can more effectively reduce the local optimal problem caused by the heterogeneity of client data, which significantly improves the convergence of the global model.

Scaffold achieved the best convergence effect, with the fastest loss decrease and finally convergence to the lowest value. This result shows that the control variable method

adopted by Scaffold can effectively reduce the model drift problem caused by uneven data distribution, making the client update closer to the global optimal direction. Its loss decline curve is not only steeper, but also maintains the minimum volatility in the later stage of training, indicating that this method has stronger robustness when dealing with heterogeneous data. Therefore, in a non-independent and identically distributed environment, Scaffold is a better federated optimization strategy than FedAvg and FedProx, which can more efficiently improve model performance and accelerate convergence.

Secondly, this paper evaluates the performance of adaptive gradient scaling (AGS) on non-independent and identically distributed data. The experimental results are shown in Table 1.

Table 1: Experimental results

Model	Epochs	Final accuracy	Convergence stability
Standard FedAvg	45	72.3	0.085
FedAvg + AGS	38	75.8	0.063
FedProx	40	74.1	0.072
FedProx + AGS	34	77.2	0.058
Scaffold	32	78.6	0.049
Scaffold + AGS	28	80.4	0.036

From the experimental results, in the non-independent and identically distributed (Non-IID) data environment, adaptive gradient scaling (AGS) can effectively improve the convergence speed and final model performance of federated learning. Standard FedAvg requires 45 rounds of training to converge, and the final accuracy is only 72.3%, and the convergence stability is poor, with a loss fluctuation of 0.085. This instability is mainly due to the fact that FedAvg adopts a simple global average strategy. When the data distribution is uneven, the model update of some clients may have a greater impact on the global model, resulting in an unstable optimization process. In contrast, FedAvg + AGS only needs 38 rounds to converge, and the accuracy is increased to 75.8%, and the convergence stability is also significantly improved (the fluctuation range is reduced to 0.063). This shows that AGS effectively alleviates the training instability problem caused by data heterogeneity by dynamically adjusting the gradient contribution of different clients.

In the FedProx solution, the addition of AGS also brings significant optimization. The convergence number of standard FedProx is 40 rounds, the final accuracy is 74.1%, and the stability fluctuation is 0.072, which is more stable than FedAvg. FedProx + AGS further accelerates the convergence, and only 34 rounds are needed to achieve 77.2% accuracy, while the loss fluctuation is reduced to 0.058, indicating that its global model optimization direction is more stable. Since FedProx introduces regularization terms in the local optimization process, it reduces the impact of data heterogeneity on the global model. Combined with AGS, it further optimizes the balance of gradient contributions of different clients, making training more efficient.

Scaffold + AGS achieved the best experimental results, converging in only 28 rounds, with a final accuracy of 80.4% and optimal stability (loss fluctuation 0.036). Compared with the standard Scaffold (32 rounds of convergence, 78.6% accuracy), the introduction of AGS further enhances the adaptability of the global model in a heterogeneous data environment. This shows that the gradient scaling strategy combined with the control variable method can effectively reduce the model drift problem caused by data distribution differences, accelerate convergence and improve the final performance. Overall, AGS can improve the convergence speed, stability and final accuracy in all optimization strategies, proving its applicability and effectiveness in non-independent and identically distributed data environments.

6. Conclusion

This study proposes an adaptive gradient scaling (AGS) strategy for the optimization problem of federated learning in a non-independent and identically distributed (Non-IID) data environment to improve the model convergence speed, enhance training stability, and improve the final accuracy. Experimental results show that after introducing AGS on the basis of the standard FedAvg, FedProx and Scaffold schemes, the number of convergence rounds of all methods is significantly reduced, the final accuracy is improved, and the training stability is enhanced. Among them, the Scaffold + AGS scheme performs best, converging in only 28 rounds, with a final accuracy of 80.4%, an increase of 1.8% over the standard Scaffold, and the loss fluctuation is also minimized. This shows that in a heterogeneous data environment, dynamically adjusting the client gradient contribution can effectively alleviate the negative impact of data imbalance on model optimization, making federated learning more adaptable in a wider range of application scenarios.

In addition, the experimental results of this study also show that the AGS scheme can bring consistent performance improvements in different federated optimization strategies (FedAvg, FedProx, Scaffold), especially in the FedAvg and FedProx frameworks, where the convergence speed is more obvious. Since the standard FedAvg is unstable in training under highly heterogeneous data environments, the introduction of AGS can significantly reduce gradient deviation and improve the robustness of the global model. At the same time, the FedProx + AGS scheme can further optimize the local training process of the client, while reducing communication costs, so that the model can still converge stably under uneven data distribution. This shows that AGS has strong versatility and is suitable for different types of federated optimization methods. It can improve the overall performance of the federated learning model while ensuring privacy protection.

Future research can further explore the combination of AGS with other federated optimization strategies, such as personalized federated learning (Personalized FL), federated distillation (Federated Distillation), asynchronous federated learning (Asynchronous FL), etc., to improve its adaptability in complex scenarios. In addition, combining privacy protection technologies such as differential privacy (DP) and homomorphic encryption (HE) to optimize the application potential of AGS in scenarios with high security requirements

is also a direction worthy of further research. Ultimately, the results of this study can provide new optimization ideas for federated learning applications involving privacy data protection in medical, financial, and mobile devices, and promote the efficient deployment and implementation of federated learning in large-scale, heterogeneous environments.

References

[1] Anderson, J., & Roberts, M. (2021). "Optimizing Federated Learning in Heterogeneous Environments: A Gradient Scaling Approach." Journal of Distributed AI Systems, 18(2), 45-58.

[2] Harrison, D., & Lewis, T. (2022). "Federated Learning on Non-IID Data: A Comparative Study of FedAvg, FedProx, and Scaffold." Proceedings of the International Conference on Edge AI (ICEAI), 112-124

[3] Martinez, R., & Evans, K. (2020). "Privacy-Preserving Federated Learning: Evaluating the Impact of Differential Privacy on Model Performance." Journal of Artificial Intelligence and Security, 25(3), 67-80.

[4] Baker, S., & Thompson, L. (2021). "Personalized Federated Learning: A Framework for Handling Client Data Diversity." International Workshop on Secure Distributed Learning (WSDL), 78-90.

[5] Williams, P., & Brown, J. (2023). "Adaptive Gradient Scaling for Federated Learning with Heterogeneous Clients." Journal of AI-Driven Optimization, 12(4), 90-105.

[6] Miller, J., & Adams, T. (2022). "Adaptive Gradient Scaling for Federated Learning in Non-IID Environments." Journal of Distributed AI and Privacy, 14(3), 112-126.

[7] Wang, X. (2024). Time-Aware and Multi-Source Feature Fusion for Transformer-Based Medical Text Analysis. Transactions on Computational and Scientific Methods, 4(7).

[8] Zhu, W., Wu, Q., Tang, T., Meng, R., Chai, S., & Quan, X. (2025). Graph Neural Network-Based Collaborative Perception for Adaptive Scheduling in Distributed Systems. arXiv preprint arXiv:2505.16248.

[9] Wang, Y., Zhu, W., Quan, X., Wang, H., Liu, C., & Wu, Q. (2025). Time-Series Learning for Proactive Fault Prediction in Distributed Systems with Deep Neural Structures. arXiv preprint arXiv:2505.20705.

[10] Cheng, Y. (2025). Multivariate Time Series Forecasting through Automated Feature Extraction and Transformer-Based Modeling. Journal of Computer Science and Software Applications, 5(5).

[11] Duan, Y. (2024). Continuous Control-Based Load Balancing for Distributed Systems Using TD3 Reinforcement Learning. Journal of Computer Technology and Software, 3(6).

[12] Aidi, K., & Gao, D. (2025). Temporal-Spatial Deep Learning for Memory Usage Forecasting in Cloud Servers.

[13] Dai, L., Zhu, W., Quan, X., Meng, R., Cai, S., & Wang, Y. (2025). Deep Probabilistic Modeling of User Behavior for Anomaly Detection via Mixture Density Networks. arXiv preprint arXiv:2505.08220.

[14] Lou, Y. (2024). Capsule Network-Based AI Model for Structured Data Mining with Adaptive Feature Representation. Transactions on Computational and Scientific Methods, 4(9).

[15] Wang, B. (2025). Topology-Aware Decision Making in Distributed Scheduling via Multi-Agent Reinforcement Learning. Transactions on Computational and Scientific Methods, 5(4).

[16] Zhan, J. (2024). MobileNet Compression and Edge Computing Strategy for Low-Latency Monitoring. Journal of Computer Science and Software Applications, 4(4).

[17] Wang, R. (2025). Joint Semantic Detection and Dissemination Control of Phishing Attacks on Social Media via LLama-Based Modeling.

[18] Peng, Y. (2024). Semantic Context Modeling for Fine-Grained Access Control Using Large Language Models. Journal of Computer Technology and Software, 3(7).

[19] Peng, Y. (2024). Structured Knowledge Integration and Memory Modeling in Large Language Systems. Transactions on Computational and Scientific Methods, 4(10).

[20] Liu, J. (2025). Reinforcement Learning-Controlled Subspace Ensemble Sampling for Complex Data Structures.

[21] Deng, Y. (2024). Transfer Methods for Large Language Models in Low-Resource Text Generation Tasks. Journal of Computer Science and Software Applications, 4(6).

[22] Zhu, L., Guo, F., Cai, G., & Ma, Y. (2025). Structured Preference Modeling for Reinforcement Learning-Based Fine-Tuning of Large Models. Journal of Computer Technology and Software, 4(4).

[23] Ma, Y., Cai, G., Guo, F., Fang, Z., & Wang, X. (2025). Knowledge-Informed Policy Structuring for Multi-Agent Collaboration Using Language Models. Journal of Computer Science and Software Applications, 5(5).

[24] Zhang, H., Ma, Y., Wang, S., Liu, G., & Zhu, B. (2025). Graph-Based Spectral Decomposition for Parameter Coordination in Language Model Fine-Tuning. arXiv preprint arXiv:2504.19583.

[25] He, J., Liu, G., Zhu, B., Zhang, H., Zheng, H., & Wang, X. (2025). Context-Guided Dynamic Retrieval for Improving Generation Quality in RAG Models. arXiv preprint arXiv:2504.19436.

[26] Han, X., Sun, Y., Huang, W., Zheng, H., & Du, J. (2025). Towards Robust Few-Shot Text Classification Using Transformer Architectures and Dual Loss Strategies. arXiv preprint arXiv:2505.06145.

[27] Zhang, T., Shao, F., Zhang, R., Zhuang, Y., & Yang, L. (2025). DeepSORT-Driven Visual Tracking Approach for Gesture Recognition in Interactive Systems. arXiv preprint arXiv:2505.07110.

[28] Cai, G., Kai, A., & Guo, F. (2025). Dynamic and Low-Rank Fine-Tuning of Large Language Models for Robust Few-Shot Learning. Transactions on Computational and Scientific Methods, 5(4).

[29] Gong, J., Wang, Y., Xu, W., & Zhang, Y. (2024). A Deep Fusion Framework for Financial Fraud Detection and Early Warning Based on Large Language Models. Journal of Computer Science and Software Applications, 4(8).

[30] Huang, W., Zhan, J., Sun, Y., Han, X., An, T., & Jiang, N. (2025). Context-Aware Adaptive Sampling for Intelligent Data Acquisition Systems Using DQN. arXiv preprint arXiv:2504.09344.

[31] Wang, X., Liu, G., Zhu, B., He, J., Zheng, H., & Zhang, H. (2025). Pre-trained Language Models and Few-shot Learning for Medical Entity Extraction. arXiv preprint arXiv:2504.04385.

[32] Cai, G., Gong, J., Du, J., Liu, H., & Kai, A. (2025). Investigating Hierarchical Term Relationships in Large Language Models. Journal of Computer Science and Software Applications, 5(4).

[33] Yu, Z., Wang, S., Jiang, N., Huang, W., Han, X., & Du, J. (2025). Improving Harmful Text Detection with Joint Retrieval and External Knowledge. arXiv preprint arXiv:2504.02310.

[34] Wang, Y., Fang, Z., Deng, Y., Zhu, L., Duan, Y., & Peng, Y. (2025). Revisiting LoRA: A Smarter Low-Rank Approach for Efficient Model Adaptation. arXiv preprint arXiv: not available.

[35] Fang, Z., Zhang, H., He, J., Qi, Z., & Zheng, H. (2025). Semantic and Contextual Modeling for Malicious Comment Detection with BERT-BiLSTM. arXiv preprint arXiv:2503.11084.

[36] Sun, D., He, J., Zhang, H., Qi, Z., Zheng, H., & Wang, X. (2025). A LongFormer-Based Framework for Accurate and Efficient Medical Text Summarization. arXiv preprint arXiv:2503.06888.

[37] Garcia, M., & Foster, G. (2021). "Federated Optimization for Imbalanced Data: A Case Study on Financial Risk Assessment." Journal of Applied Federated Computing, 17(1), 78-92.

[38] Bennett, L., & Carter, T. (2021). "Towards Robust Federated Learning: A Study on Model Drift and Adaptive Regularization." International Symposium on Decentralized AI (ISDAI), 155-170.

[39] Gomez, R., & Clark, S. (2023). "Federated Learning for Healthcare: Addressing Data Distribution Challenges with Adaptive Methods." Journal of Intelligent Systems in Medicine, 12(2), 130-145.