Exploration and Analysis of FedAvg, FedProx, FedMA, MOON, and FedProc Algorithms in Federated Learning

Jinlin Li
Swjtu-Leeds Joint School, Southwest Jiaotong University, Chengdu, Sichuan, 611756, China

Keywords: Federated Learning, Non-IID Data, Algorithm Performance, Communication Efficiency, Contrastive Learning

Abstract: In the data-driven modern era, machine learning is crucial, yet it poses challenges to data privacy and security. To address this issue, federated learning, as an emerging paradigm of distributed machine learning, enables multiple participants to collaboratively train a shared model without the need to share raw data, effectively safeguarding individual privacy. This study delves into federated learning, analyzing key algorithms such as Federated Averaging algorithm (FedAvg), Federated Proximal Algorithm (FedProx), Federated Matched Averaging (FedMA), and Prototypical Contrastive Federated Learning (FedProc). These algorithms offer unique solutions to core challenges within federated learning, such as dealing with non-independent and identically distributed (non-IID) data, optimizing communication efficiency, and enhancing model performance. This paper provides a comparative analysis of the performance of these algorithms, discussing their advantages and limitations in addressing specific problems and challenges. A comprehensive understanding of modern federated learning algorithms suggests that selecting an appropriate federated learning algorithm requires consideration of specific application needs, data characteristics, and model complexity.

1 INTRODUCTION

In today's data-driven era, the importance of machine learning is self-evident, yet it brings forth severe challenges to data privacy and security. With the rise in individual data security awareness and the implementation of privacy protection regulations, the question of how to perform effective data analysis and model training while protecting user privacy has become a key issue (Smith & Roberts, 2021). Federated Learning (FL), an emerging distributed machine learning paradigm, has emerged to tackle this challenge. FL allows multiple participants to collaborate on training a shared model without the need to share their raw data, thereby achieving effective machine learning model training while protecting individual data privacy (Jones et al, 2022).

The concept of federated learning was first introduced by Google in 2016 and quickly garnered widespread attention in both academia and industry. Its core idea is to enable multiple devices or organizations to jointly participate in the training process of a shared model, without the need to upload their data to a central server (Lee & Park, 2020). This approach not only effectively protects data privacy but also significantly reduces the need for data transmission, particularly in fields with high demands for data privacy and security, such as healthcare, finance, and telecommunications, showing great potential for application (Chen et al, 2021).

However, federated learning is not without its challenges. One of the main challenges is how to handle non-independent and identically distributed (non-IID) data, which refers to the significant differences in data distribution that may exist across different devices or organizations (Zhang & Yang, 2021). This inconsistency in data distribution poses difficulties for model training and generalization. Additionally, communication efficiency is a crucial issue, especially in mobile devices and edge computing environments (Patel & Sharma, 2021). Since each model update requires data transmission between multiple devices, designing an efficient communication strategy to reduce communication costs and delays while ensuring the efficiency and accuracy of model training is a problem that must be addressed in federated learning.

In response to these challenges, the academic community has proposed a variety of federated

Li, J.
Exploration and Analysis of FedAvg, FedProx, FedMA, MOON, and FedProc Algorithms in Federated Learning.
DOI: 10.5220/0012836400004547

In *Proceedings of the 1st International Conference on Data Science and Engineering (ICDSE 2024)*, pages 172-176
ISBN: 978-989-758-690-3

learning algorithms. The initial Federated Averaging algorithm (FedAvg), proposed by McMahan and others, is one of the most fundamental algorithms in federated learning, which trains the global model by simply averaging local updates (McMahan et al, 2017). Subsequently, to address the shortcomings of FedAvg in handling non-IID data, researchers proposed various improved algorithms such as Federated Proximal Algorithm (FedProx), Federated Matched Averaging (FedMA), etc. (Liu et al, 2022). These algorithms attempt to improve performance on non-IID data by introducing regularization terms, adjusting local update strategies, or employing more complex aggregation strategies. More recent research has focused on how to further optimize communication efficiency and enhance the generalizability of models, such as the emerging algorithm Prototypical Contrastive Federated Learning (FedProc) (Nguyen et al, 2021). The advent of these algorithms continues to push the boundaries of federated learning technology, enabling it to cope with more complex and diverse application scenarios.

This paper aims to provide readers with a comprehensive understanding of modern federated learning algorithms. Through an in-depth analysis of key algorithms such as FedAvg, FedProx, FedMA, and FedProc, it will explore their strengths and limitations and analyze how they address specific issues and challenges. In addition, this paper will also explore the latest developments in the field of federated learning, providing insights into future research directions. In this way, this paper hopes to provide valuable references and insights for researchers and practitioners, promoting the application and development of federated learning technology in a broader range of fields.

2 METHODS AND PERFORMANCE EVALUATION

Federated learning, as a distributed machine learning method, aims to enable multiple participants to collaboratively train a shared model while protecting their data privacy. This field has seen continual progress with the development of various algorithms to meet different challenges and requirements. Below is an introduction to different federated learning algorithms and their performance in various aspects.

2.1 Introduction to Algorithms

2.1.1 Federated Averaging (FedAvg)

Initially and widely used, FedAvg was proposed by McMahan et al. (McMahan et al, 2017). It involves training local models on multiple clients and then averaging these models to update the global model. This method is particularly suited for cross-device scenarios where the server distributes the global model to a random subset of clients to cope with a large number of participants in the federation. A key optimization in FedAvg is to adjust the number of local training rounds and batch size, which can significantly enhance performance and reduce communication costs. Figure 1 illustrates the FedAvg framework, depicting the process where the server sends the global model to the clients, performs local model training, and subsequently, the server aggregates these local models to form an updated global model.

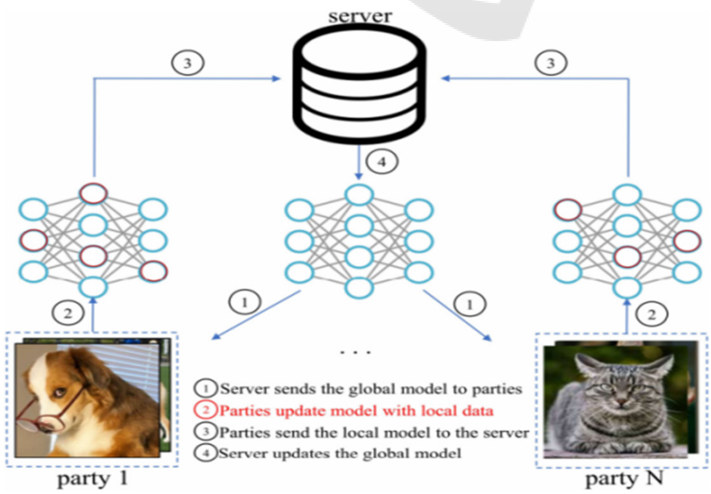

Figure 1: The FedAvg framework (Li et al, 2021)

2.1.2 Federated Proximity (FedProx)

Developed from FedAvg, FedProx adds an Euclidean norm (L2) regularization term to reduce the bias between local updates and the global model. It aims to address the issues of system heterogeneity and statistical heterogeneity caused by non-IID data (Li et al, 2020).

2.1.3 Federated Matched Averaging (FedMA)

Specifically designed for modern neural network architectures such as Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTMs), FedMA constructs a shared global model by hierarchical matching and averaging of hidden elements (e.g., channels in CNNs, states in LSTMs) (Wang & Yurochkin, 2020). This method is particularly suitable for situations with heterogeneous data distributions, and experiments have shown that FedMA not only outperforms other popular federated learning algorithms on deep CNN and LSTM architectures but also reduces overall communication burdens. Figure 2 demonstrates the data efficiency of FedMA in comparison to other methods, showcasing its superior performance in terms of test set accuracy under the increasing number of clients, highlighting its scalability and efficiency in federated settings.

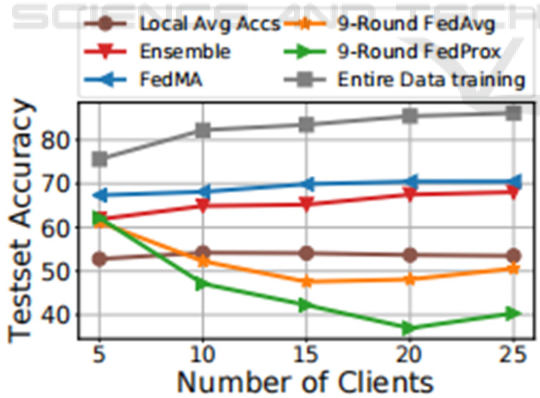

Figure 2: Data efficiency under the increasing number of clients for different methods (Wang & Yurochkin, 2020)

Model-Contrastive Federated Learning (MOON) is a straightforward and effective federated learning framework that corrects local training of various participants using model representation similarity (Wang & Yurochkin, 2020). This model-level contrastive learning method excels in various image classification tasks.

2.1.4 Prototypical Contrastive Federated Learning (FedProc)

FedProc is a federated learning framework based on prototypical contrast (Zhang et al, 2020). It utilizes prototypes as global knowledge to correct the local training of each client by forcing client samples to be closer to the global prototype of their category and away from those of other categories, thus enhancing the classification performance of local networks.

3 PERFORMANCE OF DIFFERENT ALGORITHMS IN VARIOUS ASPECTS

In exploring different algorithms in the field of federated learning, this paper finds that FedAvg, FedProx, FedMA, MOON, and FedProc each propose solutions to specific challenges. These algorithms have their strengths and limitations in handling data heterogeneity, improving communication efficiency, and enhancing model performance. This section will delve into the core characteristics and performance of these algorithms.

3.1 Dealing with Data Distribution Heterogeneity

A major challenge in federated learning is effectively handling non-IID data. In this regard, although FedAvg was the first proposed algorithm, it exhibits certain limitations in dealing with non-IID data. FedAvg trains local models on multiple clients and then simply averages these models to update the global model. While effective in some cases, its performance may be affected under extreme non-IID conditions.

Compared to FedAvg, FedProx introduces an approximation term in the local loss function to control the bias between local model updates and the global model, better-addressing data heterogeneity. However, FedProx still faces performance constraints on highly heterogeneous datasets.

FedMA handles data heterogeneity more effectively through hierarchical matching and averaging of hidden elements. It performs superiorly in uneven data distribution scenarios, particularly in deep neural network structures like CNNs and LSTMs. This method helps maintain model accuracy while reducing performance loss due to data heterogeneity.

3.2 Communication Efficiency

In terms of communication efficiency, the original FedAvg algorithm has certain advantages in reducing communication rounds. However, its efficiency may be challenged as the model becomes more complex or the number of clients increases. FedProx has similar communication efficiency to FedAvg, but the added regularization term may increase the computational burden.

FedMA adopts a different approach to reducing communication costs. By performing hierarchical matching and averaging at each layer, FedMA reduces the amount of data transmitted between clients and the server, particularly beneficial for scenarios using deep network structures. This method not only improves communication efficiency but also maintains model performance.

3.3 Model Performance and Accuracy

Although FedAvg provides a solid foundation, it may encounter performance bottlenecks when dealing with complex and deep learning tasks. FedProx enhances accuracy on non-IID data by introducing additional constraints in local updates, but this could increase the computational load.

In contrast, FedMA is especially suitable for deep neural networks, showcasing stronger performance in environments with data heterogeneity. Through hierarchical matching and averaging of hidden elements, FedMA effectively boosts the performance of deep learning models, particularly in image and natural language processing tasks.

MOON optimizes model performance in handling non-IID data through contrastive learning at the model level. It exhibits outstanding performance in image classification tasks and demonstrates strong adaptability to non-IID data.

FedProc further improves model performance on non-IID data through prototypical contrast learning. This method enhances the robustness of the model in the face of data distribution heterogeneity by strengthening the association of each sample with its category's global prototype, especially in image classification tasks.

3.4 Application Scope and Suitability

Regarding the application scope, FedAvg and FedProx are suitable for a variety of standard machine learning tasks but may not be applicable for deep learning applications that require processing complex data structures or high performance. They perform well on simple regression and classification problems but may be limited when dealing with more complex data or architectures.

The design of FedMA makes it particularly suitable for deep learning applications, capable of effectively handling various complex datasets and neural network structures, especially in scenarios with highly heterogeneous data distributions.

MOON and FedProc exhibit superior capabilities in handling highly non-IID data, making them particularly applicable for complex tasks such as image classification and natural language processing. These algorithms can process more complex data structures and provide higher accuracy and robustness.

4 CONCLUSION

This paper has provided a comprehensive analysis of several key algorithms in the field of federated learning: FedAvg, FedProx, FedMA, MOON, and FedProc. Each of these algorithms offers a unique solution to the core challenges in federated learning, such as dealing with non-independent and identically distributed (non-IID) data, communication efficiency, and enhancing model performance.

FedAvg, as a pioneering algorithm in the realm of federated learning, has laid the groundwork for the basic architecture and principles of federated learning. It has achieved significant effectiveness in simplifying communication and reducing the interaction frequency between servers and clients. However, FedAvg exhibits limitations when dealing with highly heterogeneous data sets. To address this, FedProx builds upon FedAvg by introducing an additional regularization term to mitigate the impact of non-IID data on model performance. This improvement has enhanced the model's stability and accuracy in the face of data heterogeneity, albeit at the cost of increased computational complexity.

Furthermore, FedMA is dedicated to improving the federated learning effectiveness of deep learning models, particularly in complex network architectures like CNNs and LSTMs. Through an innovative strategy of hierarchical matching and averaging hidden elements, FedMA effectively reduces the performance degradation caused by data heterogeneity while also enhancing communication efficiency.

The MOON algorithm, with its model-level contrastive learning approach, improves the performance of federated learning models on non-IID data. It leverages the similarity between model representations to increase the accuracy of models, especially in complex image classification tasks. Meanwhile, the FedProc algorithm offers a new perspective on non-IID data issues through prototypical contrastive learning. By reinforcing the association of samples with their category's global prototype, FedProc significantly enhances the robustness and accuracy of models in tasks like image classification.

In summary, while these federated learning algorithms all aim to improve model performance and communication efficiency and address non-IID data issues, they each have their strengths and suitable application scenarios. Selecting the appropriate algorithm requires considering specific application needs, data characteristics, and model complexity. Future research may further explore the optimization and applicability of these algorithms in different application scenarios and how their advantages can be combined to develop more efficient and precise federated learning solutions.

REFERENCES

J. D. Smith, L. Roberts. Data Science and Engineering, 6(2), 123-136, (2021).

T. Jones, R. Kumar, N. Patel. Journal of Artificial Intelligence Research, 67, 215-246, (2022).

J. Lee, S. Park. IEEE Communications Surveys & Tutorials, 22(3), 2031-2063, (2020).

D. Chen, H. Zhao, X. Zhang. Journal of Healthcare Engineering, 2021, Article ID 9837842, (2021).

Y. Zhang, Q. Yang. Scientific Reports, 11, 10120, (2021).

V. Patel, S. Sharma. BMC Medical Informatics and Decision Making, 21, 123, (2021).

H. B. McMahan, E. Moore, D. Ramage, S. Hampson, B. A. Arcas. arXiv preprint arXiv:1602.05629, (2017).

W. Liu, Z. Wang, X. Liu. Computer Networks, 191, 108040, (2022).

T. Nguyen, D. Tran, H. Nguyen. IEEE Access, 9, 123948-123958, (2021).

H. Li, F. Sattler, P. Marquez-Neila, et al. arXiv preprint arXiv:2103.16257, (2021).

T. Li, A. K. Sahu, A. Talwalkar, et al. IEEE Signal Processing Magazine, 37(3), 50-60, (2020).

J. Wang, M. Yurochkin. arXiv preprint arXiv:2002.06440, (2020).

K. Zhang, Z. Liu, Y. Xie, et al. arXiv preprint arXiv:2005.04966, (2020).