Research on Solving Communication Instability and Non-IID

Lan Wang
School of Computer Science and Technology, Henan Polytechnic University, Jiaozuo, Henan Province, 454000, China

Keywords: Conventional Machine Learning, Federated Learning, Model-contrastive Federated Learning, Comparative Learning.

Abstract: Conventional machine learning (ML) methods for load forecasting rely on a central server for ML training. However, this approach has drawbacks as it necessitates transmitting all data collected by diverse devices to the central server. This process poses risks to privacy and security, strains the communication network, and demands significant centralized computing resources. In contrast, federated learning (FL) allows multiple parties to collaboratively train ML models without sharing their local data. An inherent challenge in FL is addressing the diversity in the distribution of local data across participating parties. Despite numerous studies aimed at overcoming this challenge, existing approaches often fall short in achieving satisfactory performance, particularly when dealing with image datasets and deep learning models. Model-contrastive Federated Learning (MOON) presents a straightforward and effective FL framework. MOON's core concept involves leveraging the similarity between model representations to refine individual local training, essentially conducting comparative learning at the model level. Extensive experiments demonstrate that MOON outperforms the most advanced FL algorithms across various image classification tasks.

1 INTRODUCTION

With the popularization of the Internet of Things (IoT), edge computing, and mobile devices, the demand for privacy protection and joint learning of distributed data is increasing. However, the problems of unstable communication and non-independent distributed data make federated learning (FL) face challenges in practical application.

Addressing issues related to unstable communication and non-independent distributed data is crucial for enhancing the efficiency and performance of FL and rendering it more applicable to real-world scenarios, including medical health, finance, and industry. Research topics include but are not limited to FL optimization algorithms under unstable communication, model aggregation methods under dependent and identically distributed data, communication compression and optimization, FL of heterogeneous devices and non-standardized data, etc. Research methods can include theoretical analysis, mathematical modeling, algorithm design, simulation experiments, and actual system construction. At the same time, it can learn from the methods of distributed optimization, communication network optimization, data mining, and privacy protection.

The research goal is to propose effective algorithms and methods to solve the problems of unstable communication and non-independent distributed data, improve the convergence speed, model performance, and data privacy protection level of FL, and promote the wide application of FL in practical applications. By studying the problem of unstable communication and non-independent distributed data, the theoretical basis and practical application technology of FL can be further improved, and the wide application and popularization of FL in various fields can be promoted.

2 OVERVIEW OF THE RESEARCH

2.1 FL Emerged

FL, a distributed machine learning (ML) technology, has emerged recently due to challenges in centrally managing data, privacy security issues, and ML algorithm limitations. The vast data needed for

Wang, L.
Research on Solving Communication Instability and Non-IID.
DOI: 10.5220/0012835800004547

In *Proceedings of the 1st International Conference on Data Science and Engineering (ICDSE 2024)*, pages 168-171
ISBN: 978-989-758-690-3

training Artificial Intelligence (AI) models presents practical challenges, leading to a feasible approach: organizations with data sources train models and communicate on them, ultimately aggregating a global model.

The origin of FL can be traced to the need for analyzing distributed data, spurred by the adoption of technologies like Mobile Internet (MI), IoT, and edge computing. With data increasingly distributed across organizations or devices, traditional centralized processing faces hurdles (Laroui et al. 2021). FL addresses this by enabling local model training and transmitting only encrypted average model parameters to a central server, reducing communication traffic and enhancing efficiency.

Key FL concepts include local model updates, parameter aggregation, and privacy protection. Each participant trains a model locally and shares encrypted average parameters, safeguarding data privacy while enabling collaborative training. FL finds applications in healthcare, finance, IoT, and edge computing.

In summary, FL, as an emerging distributed ML technology, holds significant theoretical significance and practical application value. It offers new approaches and methods for addressing distributed data management and privacy security issues.

2.2 Expect to Solve

FL also entails the following issues and challenges:

- Privacy protection: Since participants only share model parameters rather than raw data, it is essential to ensure the protection of users' privacy data during communication and aggregation processes, to avoid the risks of data leakage and privacy infringement.
- Data imbalance: Different participants may possess varying types or quantities of data, resulting in data imbalance issues. This could potentially affect the training effectiveness and generalization ability of federated learning models.
- Model security: In FL, the central server may become a target for attackers, especially when participants share model parameters. Therefore, measures need to be taken to ensure the security and integrity of models during communication and aggregation processes.
- Computational resource constraints: Participants' local devices may be limited by computational resources, such as memory and processor speed, which could affect the complexity and scale of model training tasks they can perform.

3 RELATED WORK

Jakub Koneˇcný proposed a method to reduce the uplink communication cost in FL called "Client-to-Server FL Communication". This approach involves transmitting only locally calculated model updates from the client to the central server, rather than sending the complete local model. While effective in reducing communication costs, this method may not fully address the requirements of complex business applications (Konečný et al. 2016).

To tackle the dual challenge of minimizing both uplink and downlink communication expenses while seamlessly integrating with existing methods, a new approach is proposed. It involves implementing lossy compression on the global model transmitted from the server to the client, along with utilizing Federated Dropout (FD) techniques. FD allows users to efficiently perform local training on a smaller subset of the global model, thereby reducing both communication costs from the client to the server and local computation requirements (Caldas et al. 2018).

In traditional machine learning setups, data is typically stored centrally, allowing ML models to access all data. However, in Federated Learning, data is distributed across local devices, resulting in inconsistent data distributions (Li et al. 2022).

In summary, these three related works offer innovative solutions to the challenges of communication cost reduction and data distribution inconsistency in Federated Learning. Jakub Koneˇcný's method focuses on reducing uplink communication costs, while the proposed approach incorporates lossy compression and FD techniques to address both uplink and downlink communication costs. These methods represent significant advancements in the field and provide valuable insights for future research in Federated Learning.

4 SOLUTION

4.1 Reduce Communication Cost

Federated Semi-supervised Hierarchical Learning via Proxy Global Model (FetchSGD) utilizes Count Sketch for compressing model updates and leverages the mergeability of Sketches to combine model updates from different clients. A key challenge in

FetchSGD's design stems from the linearity of Count Sketch, allowing for momentum and error accumulation within it. This characteristic enables the method to transfer momentum and error accumulation from clients to the central server, ensuring high compression rates and good convergence despite sparse client participation. The complete FetchSGD method, illustrated in Figure 1, involves local gradient calculation at the client: (1), sending gradient sketches to the central server; (2), central server gradient aggregation; (3), momentum and error accumulation; (4, 5), approximate top-k value extraction; (6), and central server sparse value update to participating client devices for the next round of training (7) (Fekri et al. 2022).

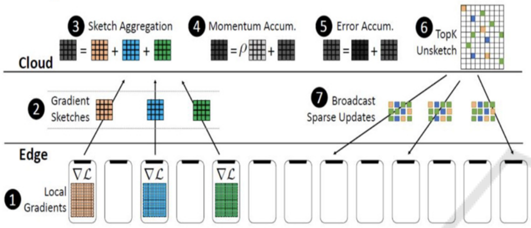

Figure 1. FetchSGD Complete Method (Konečný et al. 2016).

4.2 Solve the Non-Independent and Identically Distributed (Non-IID) Problem

During local training, Model-contrastive Federated Learning (MOON) corrects the update direction by introducing model-contrastive loss. Since the global model from the server typically produces superior features compared to the locally updated model, the model contrast loss aims to minimize the discrepancy between the features generated by the current updated model and those produced by the global model, while maximizing the gap between the features generated by the current model and those generated by the previous model.

Similar to the contrast loss, the model contrast loss is defined as follows.

$$l_{con} = -\log \frac{\exp(sim(z, z_{glob})/T)}{\exp(sim(z, z_{glob})/T) + \exp(sim(z, z_{prev})/T)} \tag{1}$$

Where, z is the feature representation generated by the current updated model on the local device; z_{glob} is the feature representation of the globally generated model by the server; z_{prev} is the feature representation of the model before the current update; $sim(\square)$ represents the similarity function between two feature representations, such as cosine similarity or dot product; T is a temperature parameter used to control the scale of the logits, typically scaling the similarities before applying the softmax function.

And the model contrastive loss l_{con} corrects the update direction by introducing the model-contrastive loss.

Three models are considered in Moon. The first one is the received global model, which is given by the server. The second is the local model uploaded in the last round. The third is after this round of training. With the above formula of comparative learning, moon's purpose is to maximize the model after this round of training. And the distance between the feature vectors and the global model received from the server. At the same time, a parameter μ will be used to determine the proportional relationship between the two loss functions. The first one is normal, when there is supervised learning. The loss function that will be used, the second is a loss function brought by comparative learning (Li et al. 2021).

5 SUGGESTION

5.1 Communication Instability Problem

- Communication optimization: By compressing traffic, reducing communication frequency, and adopting incremental updating, the influence of communication instability can be reduced. In addition, asynchronous FL can be used to allow devices to update models and transmit parameters at different times, to reduce communication competition and conflict (Li et al. 2019).
- Anomaly detection: An anomaly detection mechanism is introduced to monitor and deal with the anomalies in the communication process, to reduce the impact of communication instability (Zhu et al. 2021).

5.2 Dependent Identically Distributed Data Problem

To address the issue of dependent identically distributed data, several solutions can be implemented:

- Data resampling: Non-independent and non-identically distributed data can be resampled to enhance uniformity and independence, thereby mitigating their impact (Li et al. 202).
- Clustering and hierarchical aggregation: Devices exhibiting similar data distributions can be grouped into clusters, allowing for local training and subsequent global aggregation within each cluster. This approach helps alleviate the effects of dependent and identically distributed data (Bendiab et al. 2019).
- Meta-learning and transfer learning: Utilizing meta-learning and transfer learning techniques enables the acquisition of an improved global model in FL, better suited to handle situations involving dependent and identically distributed data.

In the local training process, the introduction of model-contrastive loss aids in resolving issues associated with dependent identically distributed data.

6 CONCLUSION

Through the investigation, it is found that Federated Averaging (FedAvg) outperforms Federated Stochastic Gradient Descent (FedSGD) in terms of accuracy while requiring fewer communication rounds. Both FedAvg and FedSGD update local models on respective devices and then transmit the average values of model parameters to the central server to enhance communication efficiency and reduce traffic. This communication strategy effectively mitigates communication costs and enhances model performance in federated learning tasks. To ensure privacy, mechanisms like differential privacy are integrated into the communication process, safeguarding users' private data. This lays a foundation for the widespread adoption of FL in practical applications and enhances its scalability across large-scale heterogeneous devices.

In various image classification tasks, MOON demonstrates superiority over other advanced FL methods. The MOON algorithm has yielded promising results in handling non-IID, thereby enhancing the applicability of FL in real-world scenarios. By dynamically weighting and rescaling dependent identically distributed data, the MOON algorithm contributes to improving the performance of the FL model in such cases.

Further efforts are directed towards enhancing the algorithm to bolster the protection of user privacy data, including the application of technologies like differential privacy and homomorphic encryption. This paper explores methods to better adapt to heterogeneous devices and non-standardized data, thereby enhancing the practical applicability of FL in real-world scenarios.

Encouraging interdisciplinary collaboration and integrating methodologies from various fields such as distributed optimization, communication network optimization, data mining, and privacy protection will further promote the role of federated learning in a broader range of application scenarios.

REFERENCES

G. Bendiab, S. Shiaeles, S. Boucherkha, et al. Computers & Security, 86, 270-290, (2019).

H. Zhu, J. Xu, S. Liu, et al. Neurocomputing, 465, 371-390, (2021).

J. Konečný, H. B. McMahan, F. X. Yu, et al. arXiv Preprint arXiv:1610.05492, (2016).

M. Laroui, B. Nour, H. Moungla, et al. Computer Communications, 180, 210-231, (2021).

M. N. Fekri, K. Grolinger, S. Mir. International Journal of Electrical Power & Energy Systems, 137, 107669, (2022).

Q. Li, B. He, D. Song. "Model-contrastive FL". In P*roceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, (2021), pp. 10713-10722.

Q. Li, Y. Diao, Q. Chen, et al. "FL on non-iid data silos: An experimental study". In *2022 IEEE 38th International Conference on Data Engineering (ICDE)*, (2022), pp. 965-978.

Q. Li, Y. Diao, Q. Chen, et al. "FL on non-iid data silos: An experimental study". In *2022 IEEE 38th International Conference on Data Engineering (ICDE)*, (2022), pp. 965-978.

S. Caldas, J. Konečny, H. B. McMahan arXiv Preprint arXiv:1812.07210, (2018).

X. Li, K. Huang, W. Yang, et al. arXiv Preprint arXiv:1907.02189, (2019).