A Survey about Federated Learning in Financial Services

Jiakun Li, Jifeng Chen*

I. Introduction

Deep Learning is often used in solving prediction or classification problems. Suppose we want a deep learning model to have a good performance. The model must be prepared with a big dataset. However, the data is sensitive in many scenarios, and we don't want the raw data to be transmitted over the Internet. Under this circumstance, federated learning is developed, which is transmitting the model weights instead of the raw data.

A typical scenario is a video recommendation. When we watch videos in apps, the apps will automatically recommend the next video according to our preferences. There will be a local model on each mobile device fitting your preference. However, the local dataset, which is also the user's watching history, needs to be bigger to train a local model. The lack of local experience also led to the overfitting of the local model. With the help of Federated Learning, the local models will aggregate into a global model. Every local model can share the experience through the global model. Even if a local model is lack of experience, it can use the weights in the global model to initialize.

Federated Learning is one of the cutting-edge approaches to realizing distributive machine learning. Unlike conventional methods where data is collected to a central server for training, federated learning operates decentralized, preserving data privacy and security while executing the collaborative model training across multiple devices or edge servers. Federated learning doesn't transmit the raw dataset in the distributive machine learning procedure. Instead, it allows machine learning models to be trained across a distributed network of devices while keeping data localized.

Financial services is a potential area for federated learning because different providers often provide financial services. On the one hand, financial services providers need the data to train the models to recognize financial crime or money laundering. On the other hand, personal account information and transaction information are sensitive, so they shouldn't be shared directly. With the help of Federated Learning, the models from different financial services providers can share their experiences with each other without sharing the raw dataset.

In this paper, we are going to discuss several financial use cases and the state-of-the-art in Federated Learning. After that, we will discuss the Federated Learning-assisted financial scenario and use experiments to verify the features of the state-of-the-art Federated Learning methods.

The contribution of this paper can be summarized as follows:

- Discussing the financial service scenario and discuss the potential of federated learning assisted financial service
- Concluding and discussing the state-of-the-art federated learning methods and techniques.
- Conducting experiments to verify the feature and evaluate the performance of the latest federated learning methods.

II. Use case Scenario

The Federated Learning is often used to solve the machine learning problem on distributive clients. When a group of clients want to train a model together, they might need to send the raw data set back to the centralized server for training. This method not only increase the chance of leaking data when transmitting but also increase the workload of the server.

Under this circumstance, the Federated Learning was introduced by McMahan *et ali.* [1] in 2016. A classic scenario of the Federated Learning is the word prediction of the mobile input method. There is the model on each mobile phone to predict the next typing word. However, the typing words and the replies might be quite different from the standard language dataset. Under this circumstance, the models can be trained with local datasets on the mobile devices. After the model is trained on local device, it will be aggregated on global servers. This behavior can utilise the customized data input on the edge devices and avoid over-fitting.

Machine Learning methods are also applied to financial services these days. The machine learning can be helpful to identify the financial crime behavior, such as the money laundering, loan fraud, fake account, and so on.

The accounts and the transactions can be considered as the nodes and the edges in the graph neural network. Under this circumstance, the behavior of the account or the transactions will be the node features and the edge features in the graph neural network. Then the invalid account identification or the transaction identification will be turned into node classification and edge classification problems in the graph neural network.

Although there are some methods that can handle the edge-sensitive classification problem now, it's not easy to handle the financial scenario directly. Because each bank wants to keep their account information and the transactions as an internal secrets. But they want to handle the crime identification at the same time. Under this circumstance, a machine learning method which can keep the dataset locally as a secret and learn the feature of the dataset on each node at the same time.

*Corresponding Author
J. Chen was with the Faculty of Sci Eng and Built Env, Deakin University. E-mail: jifeng.chen@deakin.edu.au
J. Li was with the Faculty of Sci Eng and Built Env, Deakin University.

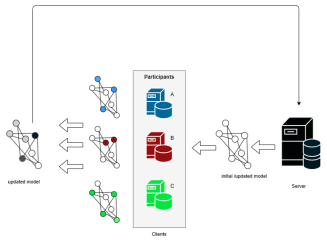

Fig. 1. Federated Learning

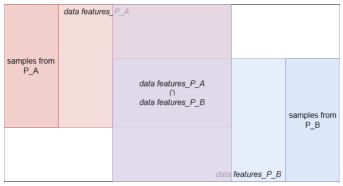

Fig. 2. Horizontal Federated Learning

III. Federated Learning

Federated Learning is a cutting-edge technology in machine learning that enables the models training across decentralized devices while keeping data localized. In Federated Learning, the edge devices are sending trained models instead of the raw data to the central server. The Federated Learning is composed of two stages, which are the local training stage and the global aggregation stage. The procedure and the architecture of the Federated Learning is shown in the Fig. 1.

In the local training stage, the local clients will firstly download the global aggregated model. Then it will train the local model with the local dataset which is observed from the environment. After the model is trained, the local training stage is finished.

In the global aggregation stage, the edge nodes will send the trained model back to the server for aggregation. As the distribution of the dataset could have some bias on each edge node, the server will have many different kind of aggregation strategies to keep the local model's accuracy and the global model's accuracy at the same time.

A. Main kinds of the Federated Learning

The Federated Learning can be generally divided into three main kinds:

1) Horizontal Federated Learning: The Horizontal Federated Learning is the Federated Learning with Sample Partitioning, which is applicable to the data of participants in federated learning has overlapping data features, and the data samples owned by participants are different. The data feature for Horizontal Federated Learning can be described as Fig. 2

For example, when the participants in federated learning are two banks serving different regional markets, although they may only have a few overlapping customers, their customer data may have very similar feature spaces due to similar business models. This means that the overlap of users between these two banks is small, while the overlap of data features is large. These two banks can collaborate to establish a machine learning model through horizontal federated learning.

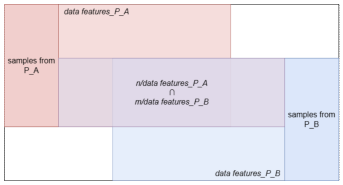

Fig. 3. Vertical Federated Learning

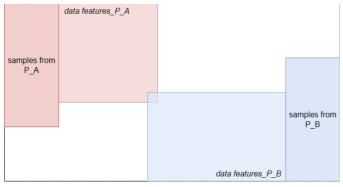

Fig. 4. Federated Transfer Learning

2) Vertical Federated Learning: The Vertical Federated Learning is the Federated Learning Divided by Features, which is applicable to the Participants in federated learning have overlapping data samples, but differ in data features. The data feature for Vertical Federated Learning can be described as Fig. 3

When two companies (such as a bank and an e-commerce company) offer different services but have a significant intersection in their customer base, they can collaborate in their respective feature spaces to obtain a better machine learning model for each. In other words, if there is a significant overlap between users and a small overlap between data features, these two companies can collaborate to train machine learning models through vertical federated learning. It uses deep neural networks on top of vertical federated learning. That is to say, segmentation learning mainly uses the settings of federated learning and trains DNN on vertically partitioned datasets.

3) Federated Transfer Learning: The Federated Transfer learning is applicable to situations where the data samples and features of the participants overlap very little. The data feature for Federated Transfer Learning can be described as Fig. 4

When the overlap between the user and data features of the dataset owned by the participants in federated learning is relatively small, they can collaborate to train machine learning models by using federated transfer learning.

B. Literature Review

Different methods are applied to increase the accuracy of the global model in Federated Learning through two aspects, which are designing a new scheme for global aggregation or adding a new regularization term in the local training procedures.

Considerable studies are focusing on proposing a novel aggregation scheme. Takayuki *et al.* [2] proposed a method solving the heterogeneity problem, which selects the clients with the highest model iteration efficiency for the global aggregation, and aggregates as much client updates as possible to accelerate the performance improvement in ML models. Yang. *et al.* [3] proposed an online model compression method. The model compression will use the model quantization to save

the size of the model need to be transmitted. Lin. *et al.* [4] proposed a method which compresses the models before the transmission. According to their experiments, their method can compress the model 600 times without an accuracy loss on the global model. Ye. *et al.* [5] proposed a method which treat the Personalized FL System as a Collaboration Graph. The aggregation will depend on this weight of each client.For aggregation methods, they are trying to use weights or other strategies to customized the ratio of the contribution from the local models to the global models. They can also use model quantization to compress the model before transmission. These kind of methods are good at solving the computational heterogeneity or communication heterogeneity problems.

There are also considerable studies are foucsing on proposing a novel training scheme. These schemes are introducing regularization terms into the local training procedure. They are also trying to envolve the concept of Personalized Federated Learning (PFL). When the aggregated model is sending back from the global server, the local model on each node could be different and fit into the local dataset distribution. Xu *et al.* [6] proposed a Personalized Federated Learning Algorithm called FedPAC. The clients can collaborate with each other sharing the feature extractor. Their method is aimed to solve the non-IID problem in the Federated Learning. Wu *et al.* [7] proposed a method called FedGMM. This method uses the Gaussian Mixture Model to fit the input distribution on each client, and combine it with the Expectation Maximum (EM) algorithmto solve the non-IID problem. Zhang *et al.* [8] proposed a method called FedCR. This method is doing the feature alignment by minimizing the discrepancy between local and global conditional mutual information. Nguyen *et al.* [9] proposed a method called FedSR. This algorithm introduce two regulators, which are the L2 regulation and the CMI. For regulation methods, they are introducing different regulators in the Personalized Federated Learning. These kinds of methods are good at training a local personalized model and develop a global model at the same time. It can deal with the scenario where the distribution and the classes of the input dataset is varying from node to node.

C. Advantages and Disadvantages of the Federated Learning

We have discussed the Federated Learning with different kinds of methods and the problems they are aiming to solve. The advantages and the disadvantages can be summarized as follow:

The advantages of the Federated Learning:

- The data is kept on local nodes as secrets. The raw data won't be transmitted between different nodes.
- The computational resources on the edge nodes can be fully utilized and reduce the training burden on the centralized server.
- The latest Federated Learning techniques can deal with the computational heterogeneity and the communicational heterogeneity problem.
- The latest Federated Learning techniques can solve the non-IID data input problem. Each node can keep a local model for fitting the local data input and utilize global experience through global aggregation at the same time.

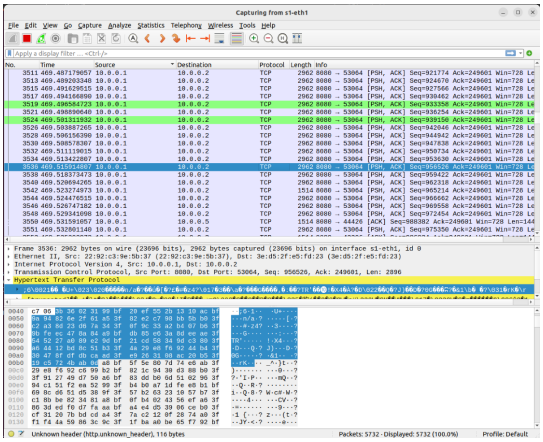

Fig. 5. Server to Client Transmission

The disadvantages of the Federated Learning:

- The accuracy of the global model may vary a lot due to the distribution of the input dataset for some methods
- The security problem still needs to be take into consideration. Because the distribution of dataset on edge nodes might be inferred from the weight of the parameters of the model.
- The models and the feature extractors need to be designed carefully to deal with the architectural heterogeneity on different edge nodes.

D. Experiment Analysis

We have built a small demo to valid the state-of-the-art researches of the Federated Learning. For this demo, we have used the flower framework [10] as the Federated Learning clients and servers, which is one of the state-of-the-art federated learning framework. We also used the mininet as the network simulator to simulate a Federated Learning Topology. In this experiment, we use a star topology with 1 server and 10 clients. The bandwidth of each link is 100 Mbps with different delay and packet loss.

The model we will be used in the Federated Learning will be a standard CNN provided by the python official website. Then this model will do the classification on the Cifar10 dataset.

When we are providing a financial service, we care about the transmitting security and its robustness facing with non-IID data. Because the data distribution, including the account information and the transaction preferences may vary a lot according to different service providers. In the following experiments, we will validate these features of the Federated Learning.

1) Federated Learning Privacy Protection Analysis: We will first validate the privacy protection ability of the Federated Learning first. We will start the Federated Learning first, then use the Wireshark, which is an open-source packet analyzer to analyze the content of the packets being transmitted.

We start the Federated Learning first then start the WireShark to capture the packages being transmitted from the client 1 to the server. The result is shown in the Fig. 6. In this figure, we can find out that all the clients are transmitting the trained models to the server instead of the raw data set.

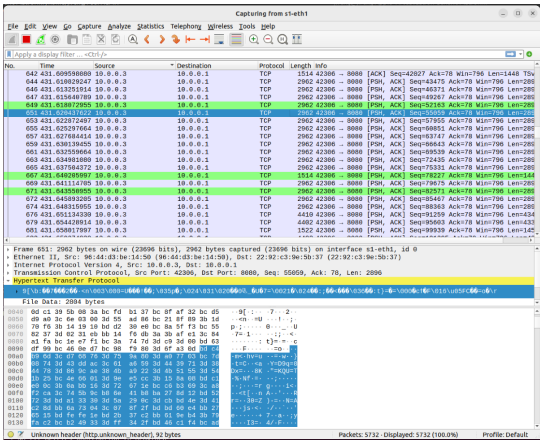

Fig. 6. Client to Server Transmission

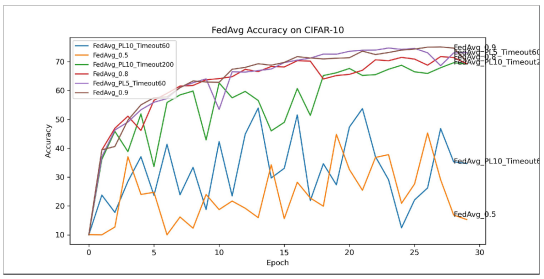

Fig. 7. FedAvg on non-IID with heterogeneity scenario

Then we use the WireShark to analyze the data transmitted from the server to the clients. In the Federated Learning, the server will send the aggregated models to the client, the result is shown in the Fig. 5. In this figure, we can find out that the server is sharing the global experience with clients by sending the aggregated model. There won't be any data being transmitted in this procedure.

2) Federated Learning VS. Personalized Federated Learning Performance: Both the inter-client and inter-organization Federated Learning will be used in the Financial Service scenario. In this scenario, not all the clients can communicate with the centralized server in every global aggregation. Besides, the data distribution could be also different for different edges.

In the financial services scenario, there will be a heterogeneity on each client and the dataset distribution should be non-IID. To illustrate the potential of applying the Federated Learning method into the financial services scenario, we have tested the FedAvg [1] and the FedPAC [6] methods.

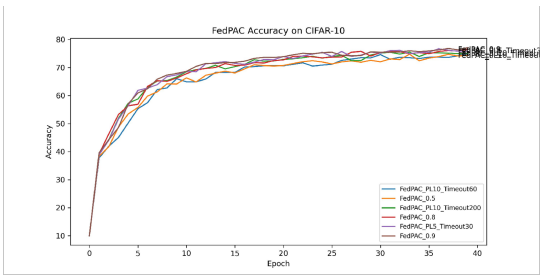

Fig. 8. FedPAC on non-IID with heterogeneity scenario

FedAvg [1] is a typical Federated Learning method. Fig. 7 is the result of the FedAvg performance under this scenario. From this result we can see that the accuracy and the convergence will drop a lot when the dataset is not following the same distribution and there's a heterogeneity. This result shows that the FedAvg is not suitable for the financial service scenario.

The reason behind this result is because the FedAvg taking an average weight of the aggregated models. When the clients are training with different hardware and datasets with different distribution, the training results is likely to conflict with each other. That's the reason why we can see the accuracy is always fluctuating and making FedAvg a bad choice for the Financial Service Scenario. For the conventional Federated Learning methods, they will all sharing these kinds of features and influence the performance in the Financial Service Scenario.

FedPAC [6] is a typical Personalized Federated Learning method. Fig. 8 is the result of the FedPAC performance under the financial service scenario. From the result we can see the accuracy is growing stable with a good convergence. This is because the FedPAC can coordinate with the result come from nodes. For these Personalized Federated Learning methods, they will have a good performance under this scenario, which show the potential of application in financial service scenarios.

IV. Generalized Federated Learning Assisted Financial Service Scenario

For the financial services scenario we have discussed before have two significant features. Firstly, the data will be used in the Machine Learning for this scenario are all sensitive data. Secondly, the data distributed on different clients (organizations) might be quite different from each other. Under this circumstance, the Federated Learning method may potentially contribute a lot to solve this problem.

Firstly, the Federated Learning and the model quantization/compression technology can save the transmission bandwidth and keep the security at the same time. The data transmitting between different organizations should only be the parameters of the model instead of the raw data. It can keep sensitive data like the account details or the transaction information locally.

Secondly, the Personalized Federated Learning Technology can solve the non-IID problem. The different financial organizations may have different targeted users, which also lead to the different distribution of the data input. The Personalized Federated Learning Technology can keep a personalized model locally and learn the experience from the global model from the global aggregation at the same time.

Thirdly, the machine learning objectives in the financial services scenarios are generally edge feature-sensitive classification problems or node feature-sensitive classification problems. There are some latest Federated Learning methods that can optimize the model training for graph neural network problems.

V. Conclusion

In conclusion, in this paper we have discussed the Federated Learning Privacy Protection technology and the Financial services scenario which have the potential to be improved with the Federated Learning technology. Then we have concluded the Federated Learning Technology and their advantages and the disadvantages. After that, we have concluded the specific techonologies can be applied to the federated learning scenarios which we have discussed before. For this paper, we investigated the Federated Learning and discussed the potential of applying Federated Learning into Financial services.

References

[1] H. B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. Agüera y Arcas, "Communication-Efficient Learning of Deep Networks from Decentralized Data," *arXiv e-prints*, p. arXiv:1602.05629, Feb. 2016.

[2] T. Nishio and R. Yonetani, "Client selection for federated learning with heterogeneous resources in mobile edge," in *ICC 2019 - 2019 IEEE International Conference on Communications (ICC)*, 2019, pp. 1–7.

[3] T.-J. Yang, Y. Xiao, G. Motta, F. Beaufays, R. Mathews, and M. Chen, "Online model compression for federated learning with large models," 2022.

[4] Y. Lin, S. Han, H. Mao, Y. Wang, and W. J. Dally, "Deep gradient compression: Reducing the communication bandwidth for distributed training," 2020.

[5] R. Ye, Z. Ni, F. Wu, S. Chen, and Y.-F. Wang, "Personalized federated learning with inferred collaboration graphs," in *ICML 2023*, June 2023. [Online]. Available: https://www.microsoft.com/en-us/research/publication/personalized-federated-learning-with-inferred-collaboration-graphs/

[6] J. Xu, X. Tong, and S.-L. Huang, "Personalized federated learning with feature alignment and classifier collaboration," in *The Eleventh International Conference on Learning Representations*, 2023. [Online]. Available: https://openreview.net/forum?id=SXZr8aDKia

[7] Y. Wu, S. Zhang, W. Yu, Y. Liu, Q. Gu, D. Zhou, H. Chen, and W. Cheng, "Personalized federated learning under mixture of distributions," 2023.

[8] H. Zhang, C. Li, W. Dai, J. Zou, and H. Xiong, "Fedcr: Personalized federated learning based on across-client common representation with conditional mutual information regularization," in *Proceedings of the 40th International Conference on Machine Learning*, ser. ICML'23. JMLR.org, 2023.

[9] A. T. Nguyen, P. Torr, and S.-N. Lim, "FedSR: A simple and effective domain generalization method for federated learning," in *Advances in Neural Information Processing Systems*, A. H. Oh, A. Agarwal, D. Belgrave, and K. Cho, Eds., 2022. [Online]. Available: https://openreview.net/forum?id=mrt90D00aQX

[10] D. J. Beutel, T. Topal, A. Mathur, X. Qiu, J. Fernandez-Marques, Y. Gao, L. Sani, H. L. Kwing, T. Parcollet, P. P. d. Gusmão, and N. D. Lane, "Flower: A friendly federated learning research framework," *arXiv preprint arXiv:2007.14390*, 2020.