

# A Survey of Developments in Federated Meta-Learning

Yong Zhang, Mingchuan Zhang

Henan University of Science and Technology, Luoyang 471000, China

**Abstract:** Federated meta-learning is a widely used few-shot learning method and has a very good development prospect. Federated meta-learning combines the characteristics of federated learning and meta-learning. It can not only use the data of each client while protecting its privacy to a certain extent, but also solve the problem of data volume that requires a large amount of data for model training in machine learning. With the rise of big data technology and edge computing, federated meta-learning technology has become a research hotspot in machine learning. In this paper, we provide an overview of the development of federated meta-learning and point out the relationship between federated learning, meta-learning and federated learning. Finally, some existing problems in federated meta-learning are pointed out, which provides ideas for the subsequent research on federated meta-learning.

**Keywords:** Federated Learning; Meta-Learning; Federated Meta-Learning.

## 1. Introduction

In recent years, machine learning has been extremely widely used in many scientific fields with the continuous development of artificial intelligence technology. Language recognition, face recognition, and autonomous driving are all used in daily life. And the development of machine learning is driving the development of modern science and technology. Some scholars have proposed that machine learning is playing an increasingly important supporting role in People's Daily life and scientific research[1]. AlphaGo[2], which is well known for defeating the Go world champion, Lee Sedol, is aGo artificial intelligence program developed by using deep learning in machine learning. The University of Alberta team's DeepStack beat the pros in a two-player no-limit Texas Hold 'em game[3]. In everyday life, the development process of machine learning provides convenience for people's daily life. Devlin et al[4]. proposed a natural language processing method that enables better interaction with machines. The current self-driving technology was implemented by the Waymo team[5] using machine learning and tested in a variety of traffic conditions. These examples fully show that the development process of machine learning has facilitated people's lives and promoted the development of science and technology.

Machine learning has achieved success in many fields, and the foundation of these successes is the support of large amounts of data. But with the development of society, people pay more and more attention to their privacy. Since data often contains individual or group privacy, people do not participate in data sharing in most cases for privacy protection. At the same time, due to the rise of the Internet of Things and edge computing, data is distributed on different devices, resulting in the problem of data island. To solve this problem, some scholars have proposed a federated learning framework[6]. In federated learning, each client trains a local model using local data and sends the model parameters to the server. After the server obtains the model parameters of each client, it averages all the model parameters and sends the model parameters to the client, and the client uses the averaged model parameters for the next round of model update. This update method can not only make better use of the large amount of distributed data obtained by big data and cloud computing technology,

but also protect the user's privacy to a certain extent while using the user's local data. Since the characteristics of federated learning meet the needs of actual production environment, this method has received extensive attention after being proposed.

## 2. Federated Learning

Federated learning is a framework first proposed by McMahan et al. It mainly performs local updates on the client, and then realizes a flexible and efficient training mode, but does not provide convergence guarantees for federated learning in the research. Since federated learning was proposed, it has attracted the attention of many scholars, and has been optimized in different aspects. Later, in the study of federated learning, not only the analysis of convergence rate is provided, but also the federated learning is optimized. Wang et al[7]. improved the convergence rate of the model in federated learning by adding momentum to the server. Khanduri et al[8]. simultaneously added momentum to the central agent and the edge agent to improve the convergence speed. Gupta et al[9] improved the model generalization ability of federated learning by combining federated learning with game theory to learn causal features in the client to reach Nash equilibrium at training time. On the other hand, Yoon et al[10]. proposed a FedMix federated learning framework to improve the generalization ability of global models by using mixup, a data augmentation technique, in federated learning. We present the federated learning framework in Figure 1.

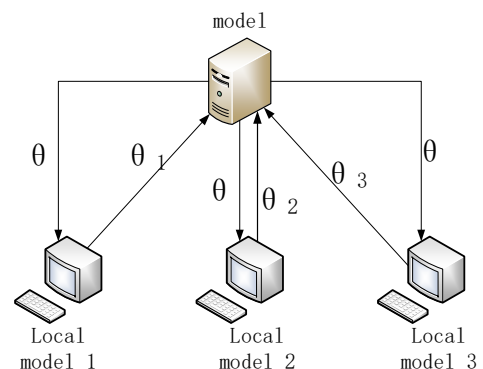


Figure 1. Federated Learning Framework

### 3. Meta-Learning

Meta-learning was introduced in 1998 by Thrun et al.[11]. After meta-learning was proposed, it attracted the interest of many scholars, who optimized meta-learning in different aspects. Luke Metz et al.[12]. investigate an unsupervised update rule that can be used between different tasks, enabling it to generalize to different neural network architectures, datasets, and data modalities. Andrychowicz et al.[13] proposed the use of gradient descent to enable the algorithm to learn by exploiting the structure in the problem of interest in an automatic way. Finn et al.[14] proposed model-agnostics based meta-learning, which learns features to adapt to new tasks more easily. At the same time, in order to solve the problem of sparse data in some fields, such as rare diseases, it is difficult to support federated learning to obtain a good model. At the same time, in order to make the model have better personalization ability, federated meta-learning framework[15] has been proposed. This is an optimization algorithm that incorporates meta-learning into federated learning. This paper mainly considers that meta-learning only needs to use a small number of samples to learn unknown tasks quickly. Model-Agnostic Meta-Learning[16] (MAML) is one of the most popular meta-learning methods, which can match any model trained using gradient descent algorithms and can be applied to a variety of different problems, such as classification, regression, and reinforcement learning[14].

### 4. Federated Meta-Learning

Federated learning and meta-learning are both gaining attention because their integration is an inevitable development. Federated meta-learning, which combines federated learning and meta-learning, was first proposed by Chen et al[17]., to obtain faster convergence rate and higher accuracy. Federated meta-learning uses MAML method to train the local meta model on the client, and then passes the meta model parameters to the server for model parameter aggregation. The server aggregates the model parameters and sends them to the client as the initial model parameters for the next update. Federated meta-learning can make use of data distributed on different clients while protecting its privacy, and better train metamodels suitable for multiple tasks. It is also possible to use the feature of meta-learning to quickly adapt to new tasks to provide different models for different users.

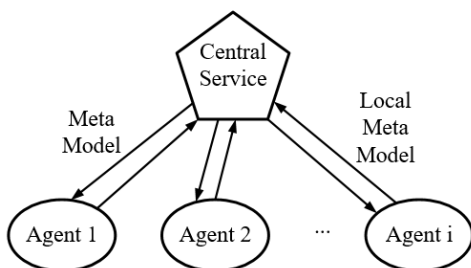


Figure 2. Federated Meta-Learning Framework

Since federated meta-learning was proposed, the combination of federated learning framework and meta-learning has received extensive attention. Khodak et al.[18] proposed a meta-learning method with adaptive learning rate and applied it to federated learning. Fallah et al.[19] proposed a personalized federated meta-learning method, and they proposed an algorithm for all agents to share the starting

model. Kayaalp et al.[20] study a decentralized federated meta-learning. This method can make the system more extensible and flexible, and avoid the communication bottleneck in the central processor. In figure 2, the structural framework of federated meta-learning is shown.

In federated learning, especially federated meta-learning, the generalization ability of models faces more challenges than the generalization ability of models derived from centralized algorithms. For example, the data heterogeneity problem between different clients, the two-layer structure of federated meta-learning leads to a more complex loss landscape. Aiming at the generalization problem of federated learning algorithms with semi-distributed structure, Mendieta et al.[21] solved the problem of data heterogeneity by using standard regularization techniques. Caldarola et al.[22] used experiments to verify that sharpness aware minimization can be used as a local optimizer to improve the generalization ability of federated learning. Since it has been proved that sharp minima will have a significant impact on the generalization ability of the model. In order to avoid finding sharp minima when the model converges, the generalization ability of the model is improved by starting from the sharp points of the model. Chaudhari et al.[23]proposed an Entropy-SGD algorithm, which uses the local extremum in the loss landscape in which a large proportion of eigenvalues in the Hessian matrix are 0 and only a very small number of positive and negative eigenvalues.

In this way, an objective function based on local entropy is constructed, which can find the extreme value with good generalization ability in the flat area of the loss landscape, and avoid the extreme value with poor generalization ability located in the sharp valley. Izmailov et al.[24] proposed simple averaging of multiple points along the SGD trajectory, which can find a flatter optimal value than SGD with almost no additional computational overhead while improving the generalization ability. And the sharpness-aware minimization algorithm proposed by Foret et al.[25] proposes to find the flat optimal value by calculating the sharpness in terms of loss landscape. We present a diagram of machine learning, federated learning, meta-learning, and federated meta-learning in Figure 3.

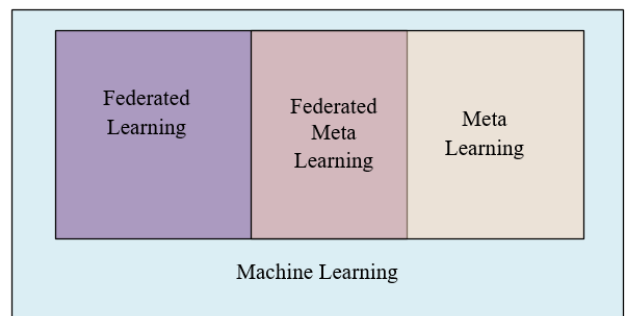


Figure 3. Logical graph

### 5. Summary

Federated meta-learning is an optimization algorithm which aims to solve the problems existing in the actual production. Federated meta-learning not only solves the problem of data islands, but also can use the fast learning ability of meta-learning to provide different models for different users with a small amount of data. Despite the success of federated meta-learning, it still has many shortcomings, such as limited generalization ability,

communication bottleneck, vulnerability to adversarial attacks, distribution shift and so on. This points out a certain direction for the subsequent research on federated meta-learning.

## Acknowledgments

This work was supported in part by the Key Technologies R & D Program of Henan Province under Grant No. 242102 210102 and 222102210080, in part by the joint Funds for Science and Technology Research and Development Plan of Henan Province under Grant No. 222103810031.

## References

- [1] Mjolsness E, DeCoste D. Machine learning for science: state of the art and future prospects [J]. *Science*, 2001, 293(5537): 2051-2055.
- [2] Silver D, Huang A, Maddison C J, et al. Mastering the game of Go with deep neural networks and tree search[J]. *Nature*, 2016, 529(7587): 484-489.
- [3] Moraveik M, Schmid M, Burch N, et al. DeepStack: Expert-level artificial intelligence in heads-up no-limit poker[J]. *Science*, 2017, 356(6337): 508-513.
- [4] Devlin J, Chang M W, Lee K, et al. Bert: Pre-training of deep bidirectional transformers for language understanding[J]. *arXiv preprint arXiv:1810.04805*, 2018.
- [5] Sun P, Kretschmar H, Dotiwalla X, et al. Scalability in perception for autonomous driving: Waymo open dataset[C]. *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2020: 2446-2454.
- [6] Konečný J, McMahan H B, Ramage D, et al. Federated optimization: Distributed machine learning for on-device intelligence[J]. *arXiv preprint arXiv:1610.02527*, 2016.
- [7] Wang J, Tantia V, Ballas N, et al. Slowmo: Improving communication-efficient distributed sgd with slow momentum [J]. *arXiv preprint arXiv:1910.00643*, 2019.
- [8] Khanduri P, Sharma P, Yang H, et al. Stem: A stochastic two-sided momentum algorithm achieving near-optimal sample and communication complexities for federated learning[J]. *Advances in Neural Information Processing Systems*, 2021, 34: 6050-6061.
- [9] Gupta S, Ahuja K, Havaei M, et al. FL Games: A federated learning framework for distribution shifts[J]. *arXiv preprint arXiv:2205.11101*, 2022.
- [10] Yoon T, Shin S, Hwang S J, et al. Fedmix: Approximation of mixup under mean augmented federated learning[J]. *arXiv preprint arXiv:2107.00233*, 2021.
- [11] Thrun S, Pratt L. *Learning to learn: Introduction and overview*[M]. Learning to learn. Boston, MA: Springer US, 1998: 3-17.
- [12] Metz L, Maheswaranathan N, Cheung B, et al. Meta-learning update rules for unsupervised representation learning[J]. *arXiv preprint arXiv:1804.00222*, 2018.
- [13] Andrychowicz M, Denil M, Gomez S, et al. Learning to learn by gradient descent by gradient descent[J]. *Advances in neural information processing systems*, 2016, 29.
- [14] Finn C, Abbeel P, Levine S. Model-agnostic meta-learning for fast adaptation of deep networks[C]. *International conference on machine learning*. 2017: 1126-1135.
- [15] Donahue J, Jia Y, Vinyals O, et al. Decaf: A deep convolutional activation feature for generic visual recognition[C]. *International conference on machine learning*. 2014: 647-655.
- [16] Obamuyide A, Vlachos A. Model-agnostic meta-learning for relation classification with limited supervision[C]. *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. 2019: 5873-5879.
- [17] Chen F, Luo M, Dong Z, et al. Federated meta-learning with fast convergence and efficient communication[J]. *arXiv preprint arXiv:1802.07876*, 2018.
- [18] Khodak M, Balcan M F F, Talwalkar A S. Adaptive gradient-based meta-learning methods[J]. *Advances in Neural Information Processing Systems*, 2019, 32.
- [19] Fallah A, Mokhtari A, Ozdaglar A. Personalized federated learning: A meta-learning approach[J]. *arXiv preprint arXiv:2002.07948*, 2020.
- [20] Kayaalp M, Vlaski S, Sayed A H. Dif-MAML: Decentralized multi-agent meta-learning[J]. *IEEE Open Journal of Signal Processing*, 2022, 3: 71-93.
- [21] Mendieta M, Yang T, Wang P, et al. Local learning matters: Rethinking data heterogeneity in federated learning[C]. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2022: 8397-8406.
- [22] Caldarola D, Caputo B, Ciccone M. Improving generalization in federated learning by seeking flat minima[C]. *European Conference on Computer Vision*. Cham: Springer Nature Switzerland, 2022: 654-672.
- [23] Chaudhari P, Choromanska A, Soatto S, et al. Entropy-sgd: Biasing gradient descent into wide valleys[J]. *Journal of Statistical Mechanics: Theory and Experiment*, 2019, 2019(12): 124018.
- [24] Izmailov P, Podoprikin D, Garipov T, et al. Averaging weights leads to wider optima and better generalization[J]. *arXiv preprint arXiv:1803.05407*, 2018.
- [25] Foret P, Kleiner A, Mobahi H, et al. Sharpness-aware minimization for efficiently improving generalization[J]. *arXiv preprint arXiv:2010.01412*, 2020.