WHEN WILL FEDERATED LEARNING CHANGE FROM GENERALIZATION TO PERSONALIZATION?

Anonymous authors
Paper under double-blind review

ABSTRACT

In order to improve the performance of federated learning models in heterogeneous situations, personalized federated learning aims to build a personalized model for each client. First, improve the generalization performance of the global model, and then improve the personalization performance of the personalized model. This is an important federated learning strategy. When will it change from generalization to personalization? The paper proposes starts personalized processing ahead of the global model convergence to improve personalized performance.This method can obtain the global model knowledge without being misled by the global model.

1 INTRODUCTION

The standard setting of federated learning aims to train a global model. The training of this model is based on the combination of data from all clients. It can perform well in general data distribution. However, due to the high Non-IID of user data and the inconsistency of users' requirements for model performance, a single global model is difficult to meet the needs of all participants. Therefore, we need to adopt a personalized method to optimize the global model for each user. This is personalized federated learning, which no longer adheres to a single global model. It tries to build a personalized model for each client to adapt to the heterogeneity of the client.The client model of local learning may not be well generalized due to insufficient data. So train the global model, and then personalize each client through local adaptation steps. This approach is an important federal personalization strategy. Because the personalization performance directly depends on the generalization performance of the global model, many personalized federation methods aim to improve the generalization performance of the global model in heterogeneous situations, and then improve the personalization performance.When did federated learning begin to change from generalization to personalization? The existing research is to start personalized training when the accuracy of the global model does not change at each client. That is, personalization starts when the global model converges. However, such operation faces a problem, and the personalized effect is not significant.

In this work, we propose to start the personalization operation before the global model converges, which is helpful to improve the personalization performance. It can obtain global knowledge without being misled by the global model. We use the accuracy of the model to evaluate the performance of the model. Therefore, we set a threshold for the accuracy of the global model. When the accuracy reaches the threshold, the training changes from generalization to personalization.

2 IMPLEMENTATION

We need the accuracy threshold of the global model. When the accuracy reaches the threshold, the training changes from generalization to personalization. Therefore, we need to determine the number of communication rounds required for the accuracy rate to reach the threshold and the real-time change of the accuracy rate with the number of communication rounds. We train the global model using the standard federated learning algorithm FedAvgMcmahan et al. (2016). Each user constructs a personalized model by performing gradient descent on his own data. So that the model can adapt to its own local data set.

Table 1: Local average accuracy

	Accuracy(%)
Our approach	90.62
FedAvg	86.83
FedProx	89.25
Scaffold	82.52

3 EXPERIMENTS AND RESULTS

In this section, we have experimentally studied the impact of personalization start time on personalization performance when the data is non-iid. We use MNIST as benchmark dataset in our study. And the MNIST data set is sampled according to the Dirichlet distribution to obtain the non-IID data set. Use the symbol $Dir(\alpha)$ To represent different Dirichlet distributions controlled by α. The smaller the value of α, the higher the heterogeneity of the data. This experiment takes α= 0.5. We use the same CNN architecture as FedAvg. And we use the SGD optimizer with a learning rate of 0.01, and set the SGD momentum to 0.9. The local batch size is set to B=50, and the local epoch is set to E=5. In this experiment, we set 1 central server and 100 clients.

In order to prove the effectiveness of our proposed method, we analyze the performance of the personalized model. The performance of the personalized model is evaluated using the average value of the local model's performance on each online client. We use FedAvg, FedProxLi et al. (2018) and SCAFFOLDKarimireddy et al. (2019) as the baseline algorithms. Based on prior knowledge, Fedavg algorithm converges in the 56th round of Mnist data set. Therefore, we choose 40-50 wheels to start steering for personalized operation. We use FedAvg to train a global model for 40 rounds of communication and distribute it to each client as an initial model. Use client data for personalized training. Compare the accuracy with the baseline algorithm.

Through experiments, Table 1 shows that the performance of our proposed method is improved compared with the baseline algorithm. This means that our method provides better personalized performance, avoids the misleading of the global model, and obtains reliable knowledge from the global model.

4 CONCLUSIONS AND FUTURE WORK

For the question of when federated learning should change from generalization to personalization, we propose to start personalization before the global model converges. Our experiments show that the accuracy of personalized model can be improved by starting personalized processing ahead of the global model convergence. In the future, we will continue to study how to determine the specific threshold. And research on which clients need personalized operation and how to select clients.

REFERENCES

S. P. Karimireddy, S. Kale, M. Mohri, S. J. Reddi, S. U. Stich, and A. T. Suresh. Scaffold: Stochastic controlled averaging for federated learning, 2019.

T. Li, A. K. Sahu, M. Zaheer, M. Sanjabi, A. Talwalkar, and V. Smith. Federated optimization in heterogeneous networks. 2018.

H. B. Mcmahan, E. Moore, D. Ramage, S. Hampson, and Bay Arcas. Communication-efficient learning of deep networks from decentralized data. 2016.