ISSN 2616-5775 Vol. 6, Issue 11: 95-101, DOI: 10.25236/AJCIS.2023.061113

Personalized Federated Learning with Attention Mechanisms

Siyuan Zhang[1,a], Guiquan Liu[1,b]

[1]*School of Cyber Security, University of Science and Technology of China, Hefei, China*
[a] *157904415@qq.com,* [b] *gqliu@ustc.edu.cn*

Abstract: *With the rapid development of science and technology, machine learning and deep learning technology have more and more applications in our daily life. At the same time, people pay more and more attention to the protection of their own privacy. And in recent years, various countries have introduced a series of laws and regulations to protect people's data privacy. In this case, the traditional method of pooling the data of each user for model training is no longer applicable, and the data of each user can only be saved in the hands of the user himself, which has a profound impact on the development of artificial intelligence technology. To solve this dilemma, the industry has proposed the concept of federated learning. Due to the data imbalance problem between each user, the prediction ability of the global aggregation model obtained by federated learning on the local specific data of each user needs to be improved. Therefore, personalization is an issue that federated learning needs to pay attention to.In order to improve the personalization ability, this paper proposes a federated transfer learning algorithm with attention mechanism. After each user obtains the global aggregation model of federated learning, the attention module is added to the local model, and then the parameters of the low-level neurons are frozen during training, and only the parameters of the high-level neurons and the attention module are updated. Finally, each user obtains a unique model that is more suitable for local data. In this paper, we conduct experiments to analyze the performance of this algorithm and the federated transfer algorithm and federated learning algorithm in terms of accuracy. The experimental results show that on the convolutional neural network model, the federated transfer learning algorithm applying the attention mechanism has improved the accuracy compared with the federated transfer algorithm and the federated learning algorithm.*

Keywords: *Federated Learning, Transfer Learning, Attention Mechanisms*

1. Introduction

In recent years, with the continuous development of society and the continuous progress of science and technology, machine learning[1-2], artificial intelligence[3-4] and other technologies have gradually been applied in people's life, work, learning and other aspects, such as computer vision[5-6], natural language processing[7-8], recommendation technology[9-10] and other fields. The industry and application innovation related to artificial intelligence technology are constantly developing all over the world.

In real life, in order to obtain large-scale data, the third party often needs to centralize the scattered data distributed in all parties. Only in this way can a good model be obtained. However, with the relevant legislation of privacy security in various countries, and more and more users are concerned about the privacy and security of their own data, the relevant data of each party cannot be simply concentrated on the third party, because it may lead to data leakage. Data leakage may not only violate relevant laws and regulations, but also cause harm to users themselves. In this case, the data can only be saved in the hands of the user itself, which causes the phenomenon of data island[11-12].

In this context, people have tried to design algorithms that do not concentrate the data together, but instead use distributed learning methods. Thus, federated learning algorithm[13] was proposed. The core idea of federated learning method is to ensure the privacy and security of the local data of each user participant, that is, the local data of each user participant does not leave the local, in this case, it is still possible to use the local data of each user participant to aggregate and train an aggregated machine learning model.

However, in the research of federated learning, we find that there is still a problem of personalization[14] in federated learning. In real life, the data of different user participants may be different,

ISSN 2616-5775 Vol. 6, Issue 11: 95-101, DOI: 10.25236/AJCIS.2023.061113

and each user participant may have some unique data, and the model trained by the global dataset is likely to perform poorly on the user-specific dataset.

Based on the research at home and abroad, this paper applies the attention mechanism and federated transfer learning to train the local models specific to each user participant, and improves the prediction effect of the local models of each user participant on the user-specific data set.

2. Related Work

2.1. Federated Learning

Federated learning is a distributed training method, which uses privacy protection technology to exploit data distributed among different user participants. Federated learning In the process of training a model, the information related to the model (such as the specific model structure, the specific parameters of the model, the calculated gradient data, etc.) can be exchanged between different user participants, but the data cannot be exchanged. The data of each user participant is always stored in the local of each user participant. The transmission of model-related information usually uses encrypted transmission. The final trained model can be shared among the various user participants. Compared with traditional machine learning algorithms, federated learning does not need to collect the data of each user participant in the learning process, which effectively protects the privacy security of each user participant.

2.2. Federated Learning Algorithm

Federated averaging algorithm (FedAvg) is a classical model aggregation algorithm, which is mainly applied to lateral federated learning systems. The main idea of federated averaging is that it assigns a weight to each party, and the larger the local dataset, the larger the weight will be assigned to each party.

$$w_{t+1} \leftarrow \sum_{k=1}^{K} \frac{n_k}{n} w_{t+1}^k$$

n is the total number of datasets for all clients of federated learning. n_k is the number of datasets owned by the client. w_{t+1}^k is the model parameters learned from the client. w_{t+1} is the aggregated model parameter.The larger the dataset used by the client to train the local model, the larger the weight of the client in the aggregation.

2.3. Transfer Learning

Transfer learning[15], as the name implies, transfer learning is that we can transfer the previous learning results to the new learning, in order to achieve twice the result with half the effort. In the field of artificial intelligence and machine learning, transfer learning represents a specific learning mode. As an important part of machine learning, transfer learning focuses on transferring the knowledge that has been learned in machine learning to new problems, so as to enhance the ability of machine learning to solve new problems. Specifically, in the field of artificial intelligence and machine learning, transfer learning is a type of learning that takes advantage of the similarity between data, tasks, or models to apply the knowledge learned in the old domain to a new problem.

In the research content and experimental analysis of this paper, we use transfer learning to train the global model in federated learning into a local model that is more suitable for user participants.

2.4. Attention Mechanis

Attention is a complex cognitive function of human brain, which is very important for human beings. Attention represents the ability of humans to pay attention to part of the information and ignore the rest when we receive it instead of paying attention to all of it. In human daily life, human beings get a lot of sensory information all the time, such as visual information, auditory information, tactile information and so on. However, the human brain is still orderly when dealing with these large and complicated external information, because the human brain can choose some of the most useful information from these complex external input information to focus on processing, and then ignore the other information. This ability is called attention.

In machine learning and artificial intelligence, neural networks often input a large amount of data. At

ISSN 2616-5775 Vol. 6, Issue 11: 95-101, DOI: 10.25236/AJCIS.2023.061113

this time, we can refer to the attention mechanism of the human brain to generate a resource allocation scheme. Through the attention mechanism, the neural network will first choose to accept the key information in the data, and then the neural network will give priority to processing these key information, and for other information. The neural network will ignore. In this way, the efficiency of the neural network is improved.

In the research content and experimental analysis of this paper, we use the attention mechanism[16] to enhance the classification ability of the local model in federated learning.

3. Proposed Method

3.1. Federated Transfer Learning

The model obtained by federated averaging algorithm often does not have good results on the specific data of each user's local data. Therefore, we use the idea of transfer learning. After obtaining the aggregated global model by federated averaging algorithm, each user's local data is used to train the model again, and the parameters of the lower layer neurons are not updated during training. Instead, only the parameters of the high-level neurons are updated to increase the predictive power of the local model trained by each user party on the local data specific to each user party.

We can explain that for the neural network, its shallow neurons are responsible for learning the general features related to the task, and the high-level neurons of the neural network are responsible for learning the special features related to the task. That is to say, in the process of neural network training, the learning representation of the network gradually changes from the general features to the special features.

In the following experiments, we design a model that, after training with the federated averaging algorithm, we obtain a joint trained global model. Each user party does not update the parameters of the shallow neurons, but only updates the parameters of the deep neurons in the process of retraining with local data.

3.2. Convolutional Block Attention Module(CBAM)

CBAM[17] is a special attention mechanism module, it is a simple and high school attention module. Given a feature map, our feature map in this CBAM module, in turn, passes through two separate dimensions, generates the attention map we need in these two dimensions, and then we multiply the generated attention map by the input feature map for adaptive feature refinement. Because the CBAM attention mechanism is individual modules, it is very lightweight and simple, so it can be seamlessly integrated into any architecture.

The structure of the CBAM attention mechanism is shown in Figure 1. Next, let's explain the operation principle of the CBAM attention mechanism in detail.

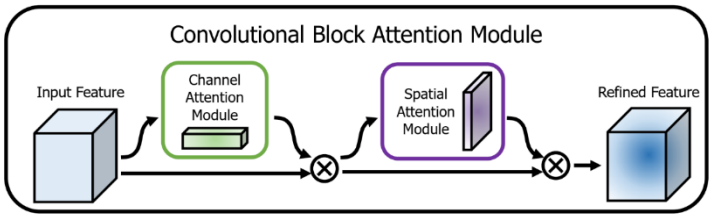

Figure 1: Convolutional Block Attention Module.

Given an input feature map $F \in R^{C\times H\times W}$, $C\times H\times W$ is the dimension of the input feature map. CBAM attention mechanism passes through Channel attention module and Spatial attention module respectively. The feature map of dimension $C\times H\times W$ is passed through the channel attention module to generate a channel attention map $M_c \in R^{C\times H\times W}$ F. The dimension of the channel attention map M_c is $C\times 1\times 1$. The feature map of degree $C\times H\times W$ is passed through the spatial attention module to generate a spatial attention map $M_s \in R^{1\times H\times W}$. The overall running process of the attention module can be summarized as follows.

ISSN 2616-5775 Vol. 6, Issue 11: 95-101, DOI: 10.25236/AJCIS.2023.061113

$$F' = M_c(F) \otimes F$$

$$F'' = M_c(F') \otimes F$$

Where $\otimes$ represents element-by-element multiplication. When two feature maps of different dimensions are multiplied, the element values on the feature map of the smaller dimension are copied, and the dimension of the feature map of the smaller dimension is expanded to match the feature map of the larger dimension. According to the above equation, we can see that the input is a feature map of dimension $C \times H \times W$, and the output is still a feature map of dimension $C \times H \times W$.

3.3. Federated Transfer Learning with Attention

The model learned by the federated averaging algorithm is a general model. When each user party has some specific data that is unique to them, the model often does not have ideal results on these specific data. Therefore, we use the idea of transfer learning, and each user party trains the aggregated model again locally with its own local data. During the training process, we do not update the parameters of the low-level neurons, but only the parameters of the high-level neurons, so that the new model improves the prediction ability on the local data of each user party.

We note that the plug-and-play feature of the CBAM attention mechanism, after the processing of the CBAM attention mechanism, the input dimension and output dimension of the data are the same. Therefore, we can easily add the CBAM attention mechanism to any part of the model without changing the overall structure of the model. The addition of CBAM attention mechanism can make the model pay more attention to the important part of the data and improve the prediction ability of the model.

Therefore, federated transfer learning with attention mechanism is that after obtaining the aggregated model trained by the federated average algorithm, each user participant adds the CBAM attention mechanism to the local model, and then trains with the local data of each user participant. During training, we only update the parameters of the high-level neurons and the attention mechanism. Instead of updating the parameters of the low-level neurons, after training, each user participant obtains its own new local model with the attention mechanism.

In the following experiments, we design a convolutional neural network model, and after training with the federated averaging algorithm, we obtain a global model that is jointly trained. Then, each user participant added the CBAM attention mechanism to the local model, and then each user participant used the local data to train again. In this process, each user participant only updated the parameters of the deep neurons of the model and the CBAM attention mechanism, but did not update the parameters of the shallow neurons.

4. Experiments

4.1. Dataset and Dataset Processing

In our experiments, we use CIFAR-10 dataset and CIFAR-100 dataset. The CIFAR-10 dataset is an image dataset. It was put together by Hinton's students Alex Krizhevsky and Ilya Sutskever. The images in the CIFAR-10 dataset are RGB color images with dimensions of 3×32×32. In the CIFAR-10 dataset, images are divided into 10 categories. These are airplane, automobile, bird, cat, deer, dog, frog, horse, ship, and truck. The training set in the CIFAR-10 dataset has a total of 50,000 images, with 5,000 images per class. The test set in the CIFAR-10 dataset has a total of 10000 data points, and there are 1000 data points for each category of images.

In order to verify the effect of our personalized federated transfer learning algorithm applying the attention mechanism on the local specific data of each user participant, we first need to preprocess the data. Data preprocessing is to simulate the data imbalance situation that federated learning mechanism may encounter in real life, that is, each user party may have some unique data in the local data of one user party, and these data are not unique to the local data of other users.

As shown in Figure 2, we divide the CIFAR-10 dataset into three different datasets, representing client A, client B, and client C. We dealt with a total of 3 different classification methods(CIFAR-10_1, CIFAR-10_2 and CIFAR-10_3).

ISSN 2616-5775 Vol. 6, Issue 11: 95-101, DOI: 10.25236/AJCIS.2023.061113

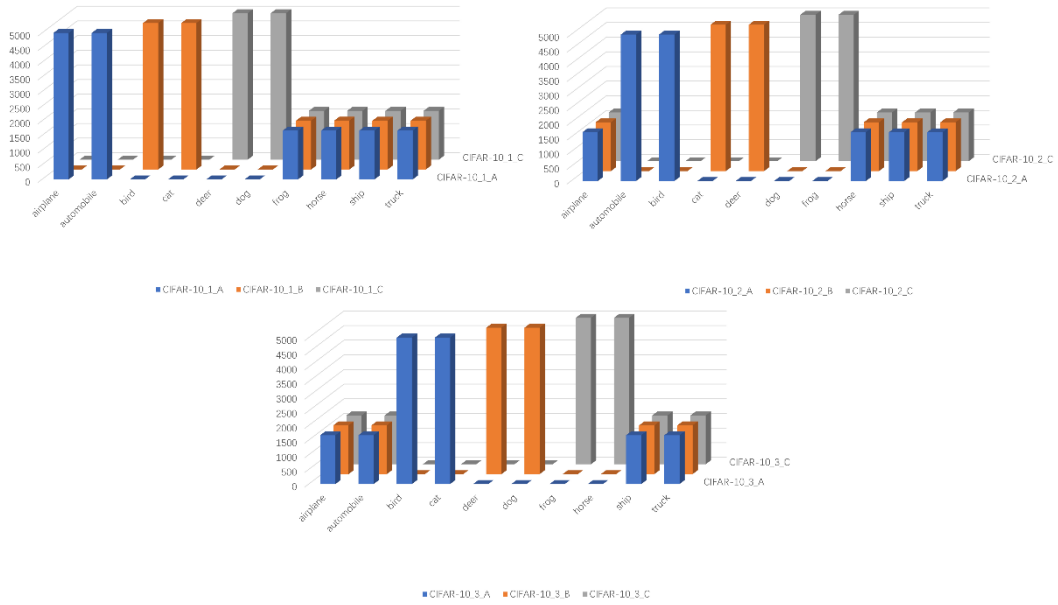

Figure 2: Dataset Processing.

4.2. Experimental Design and Settings

We designed a convolutional neural network for experiments. Convolutional neural networks have one input layer, one output layer(3×32×32), two convolutional layers(5×5×64 and 5×5×128), two pooling layers(2×2), two fully connected layers(1024 and 256), and one output layer(10).

4.3. Experimental Design and Settings

After training the model through the federated average algorithm, we obtain a joint model. During the training of federated transfer learning, each user party trains the joint model again with local data respectively, and only updates the model parameters of the following fully connected layer 1, fully connected layer 2 and output layer. Without updating the parameters of the previous input layer, convolutional layer 1, pooling layer 1, convolutional layer 2 and pooling layer 2, a new federated transfer learning multilayer perceptron model is obtained after training.

After training the model with federated averaging algorithm, we obtain a joint model. When training federated transfer learning with attention mechanism, we first add a CBAM attention module before the fully connected layer 1. Because the input and output feature maps of the CBAM attention module have the same latitude, we can plug the CBAM attention module directly into the neural network. Then each user participant trains the joint model again with local data. During the training process, we only update the model parameters of hidden layer 2 and output layer and the parameters of CBAM attention module, but do not update the parameters of input layer and hidden layer 1. After training, a new federated transfer learning multilayer perceptron model applying attention mechanism is obtained.

In order to test the effect of the federated transfer learning algorithm applying the attention mechanism on the multilayer perceptron model, we compare the performance of the federated average algorithm, the federated transfer algorithm and the federated transfer learning algorithm applying the attention mechanism on different datasets. On different data sets, after learning by the federated average algorithm, federated transfer algorithm and federated transfer learning algorithm with attention mechanism, each of the three participants has its own local model. The accuracy of each participant's local model on the test set is shown in Table 1, Table 2 and Table 3

Table 1: Accuracy of different clients on locale-specific datasets on the CIFAR-10_1 dataset.

	Client A	Client B	Client C
FedAvg	50.1%	54.1%	54.8%
Fed_T	53.2%	54.8%	55.6%
Fed_T_A	54.2%	55.7%	55.8%

ISSN 2616-5775 Vol. 6, Issue 11: 95-101, DOI: 10.25236/AJCIS.2023.061113

Table 2: Accuracy of different clients on locale-specific datasets on the CIFAR-10_2 dataset.

	Client A	Client B	Client C
FedAvg	55.8%	56.1%	57.2%
Fed_T	56.2%	57.2%	57.9%
Fed_T_A	56.6%	57.9%	58.4%

Table 3: Accuracy of different clients on locale-specific datasets on the CIFAR-10_3 dataset.

	Client A	Client B	Client C
FedAvg	56.3%	56.6%	55.9%
Fed_T	56.1%	57.7%	56.9%
Fed_T_A	56.7%	58.1%	57.4%

It can be seen that for the convolutional neural network model constructed in this experiment, federated transfer learning with attention mechanism improves the accuracy compared with federated learning and federated transfer learning on the local datasets of each participant. This indicates that the personalization ability has improved.

5. Conclusion

This paper first introduces the idea of transfer learning into federated learning, and introduces the CBAM attention module. Then this chapter gives the design of federated averaging algorithm, federated transfer learning and federated transfer learning with attention mechanism, and compares them in experiments. The model used in the experiment is a convolutional neural network model. In addition, this paper used multiple data sets to conduct experiments on the model, and compared the accuracy of the generated local model on the data sets specific to each participant.

The experimental results show that on the convolutional neural network model, the local model generated by each participant using federated transfer learning of the attention mechanism improves the prediction accuracy for the unique data owned by each participant. The personalization ability of federated learning is enhanced.

References

[1] Zhang B, Anderljung M, Kahn L, et al(2021). Ethics and governance of artificial intelligence: Evidence from a survey of machine learning researchers[J]. Journal of Artificial Intelligence Research, 71: 591–666-591–666.

[2] Jokar M, Semperlotti F(2021). Finite element network analysis: A machine learning based computational framework for the simulation of physical systems[J]. Computers & Structures,247: 106484.

[3] Fuketa H, Uchiyama K(2021). Edge artificial intelligence chips for the cyberphysical systems era[J]. Computer, 54(1): 84-88.

[4] Ossewaarde M, Gulenc E(2020). National varieties of artificial intelligence discourses: Myth, utopianism, and solutionism in West European policy expectations[J]. Computer, 53(11): 53-61.

[5] Servadei L, Mosca E, Zennaro E, et al(2020). Accurate cost estimation of memory systems utilizing machine learning and solutions from computer vision for design automation[J]. IEEE Transactions on Computers, 69(6): 856-867.

[6] Stutz D, Geiger A(2020). Learning 3d shape completion under weak supervision[J]. International Journal of Computer Vision, 128: 1162-1181.

[7] Villamizar D A, Muratore D G, Wieser J B, et al(2021). An 800 nW switched-capacitor feature extraction filterbank for sound classification[J]. IEEE Transactions on Circuits and Systems I: Regular Papers, 68(4): 1578-1588.

[8] Koyama S, Brunnström J, Ito H, et al(2021). Spatial active noise control based on kernel interpolation of sound field[J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 29: 3052-3063.

[9] Wang S, Pasi G, Hu L, et al(2020). The Era of Intelligent Recommendation: Editorial on Intelligent Recommendation with Advanced AI and Learning[J]. IEEE Intell. Syst., 35(5): 3-6.

[10] Chang L, Chen W, Huang J, et al(2021). Exploiting multi-attention network with contextual influence for point-of-interest recommendation[J]. Applied Intelligence,51: 1904-1917.

[11] Inkster N(2018). China's cyber power[M]. Routledge.

ISSN 2616-5775 Vol. 6, Issue 11: 95-101, DOI: 10.25236/AJCIS.2023.061113

[12] Voigt P, Von dem Bussche A(2017). The eu general data protection regulation (gdpr)[J]. A Practical Guide, 1st Ed., Cham: Springer International Publishing, 10(3152676): 10-5555.
[13] McMahan B, Moore E, Ramage D, et al(2017). Communication-efficient learning of deep networks from decentralized data[C]//Artificial intelligence and statistics. PMLR,1273-1282.
[14] Duan M, Liu D, Chen X, et al(2019). Astraea: Self-balancing federated learning for improving classification accuracy of mobile deep learning applications[C]//2019 IEEE 37th international conference on computer design (ICCD). IEEE, 246-254.
[15] Torrey L, Shavlik J(2010). Transfer learning[M]//Handbook of research on machine learning applications and trends: algorithms, methods, and techniques. IGI global, 242-264.
[16] Niu Z, Zhong G, Yu H(2021). A review on the attention mechanism of deep learning[J]. Neurocomputing, 452: 48-62.
[17] Woo S, Park J, Lee J Y, et al(2018). Cbam: Convolutional block attention module[C]//Proceedings of the European conference on computer vision (ECCV). 3-19.