Performance Analysis of Model-Contrastive Algorithm

Yuantao Liao *

College of Electronic and Information Engineering, Southwest University, Chongqing, 400715, China

* Corresponding author email: lyt20001020@email.swu.edu.cn

Abstract. Federated learning develops multiple clients with a universal framework platform to train deep learning models with excellent performance. During this process, the local data set is not leaked to the outside, making federated learning a powerful method to protect data security while solving the problem of isolated data islands. However, in some usage scenarios, different clients are holding local data sets which are non- Independent Identically Distribution, i.e., non-IID. There have been significant attempts made to overcome this difficulty, and one of these efforts is the model-contrastive federated learning framework, as known as MOON algorithm. The main concept is to make use of the similarities between the global and local models to adjust the local training. The final global model performs well with non-IID data. The paper aims to replicate and analyze this algorithm to provide data reference for subsequent use. The algorithm is tested on the MNIST data set to get its performance on the handwritten digit classification task. This paper processed the data set to make it non-IID, but the result achieved a high accuracy rate with an acceptable computational cost.

Keywords: Federated Learning; Image Classification; Deep Neural Network; Contrastive Learning; Machine Learning.

1. Introduction

In 2006, Hinton proposed the concept of a deep neural network for the first time and achieved good results in experiments of handwritten font recognition using this method [1]. Deep neural networks have advanced significantly since then. Meanwhile, there have been numerous neural networks proposed with various architectures. One of the most representative deep learning models is convolutional neural network. With the development of graphics processing units and the advances in computing power of all kinds of hardware, deep learning models have increased the accuracy of image classification and significantly decreased the cost of time for feature extraction. Recently, artificial intelligence and machine learning have greatly benefited from extensive research on deep learning. Research on deep learning have made excellent results in a variety of areas, such as computer vision and natural language processing, which has led to advancements in a number of technical tasks.

Supervised learning, one of the most important deep learning methods, has also advanced greatly. However, in addition to the hardware with high computing power mentioned above, supervised learning relies heavily on data. With the iterative development of supervised learning, the neural network models with excellent behavior tend to have complex structures. At the same time, the more complex their structures are, the more data they need. On the one hand, today, people face the problem of isolated data islands. A few large enterprises monopolize a large amount of quality data.

In contrast, other small and medium-sized enterprises lag far behind the large companies in data, which makes the gap between enterprises grow [2]. Meanwhile, it is always difficult for companies of the same size but in different industries to integrate data due to the closed system. This is also a problem people need to solve in China, where data is mainly held by government departments, data operators, and enterprises. Data is stored independently in one of the three isolated islands above, making it quite difficult to share data.

On the other hand, data privacy and information security have attracted worldwide attention. In May 2018, the EU issued General Data Protection Regulation to secure the data management process and strengthen the protection of users' data privacy. China has also introduced regulations to restrict the use of financial data. Increasing privacy concerns and data protection regulations do not allow

companies to send users' private data to the server. Therefore, the server, without enough data, cannot train an ideal model for users. Due to the reasons above, how to ensure the security of data while solving the problem of isolated data islands has become a major problem [2,3].

Federated learning, mooted by Google Research, provides a viable path to solve the problem [4]. The overall framework of federated learning consists of one server and several clients. The main idea is to upload locally trained models instead of local datasets. Specifically, the server provides a globally shared model for the clients, which then downloads the model and trains it with its data set to update the model parameters. In each communication between the server and the client, the server distributes the current model parameters to each client. After the local training of the clients, the renewed model parameters are returned back. In a certain way, the server uses the aggregated model parameters as the updated Server model parameter. During the process, the local data set is not leaked to the outside, which satisfies the need for user privacy protection and data security.

FedAvg, a popular federation learning algorithm, has the same basic framework as above [4]. In each round, the clients provide the updated local models for the server which then aggregate the local models into the global model. No local data set is leaked outside during the learning process.

However, in fact, the data held by different clients typically have highly skewed distributions. Specifically, the local data sets of each client are non-IID, which may result in biased local models on the clients and affect the global model's performance [5]. Therefore, it is especially essential to solve the non-IID problem.

One perspective to figure out non-IID problem focuses on the stage of updating the local model by the client to reduce the drift of the local model by adjusting the deviation of the local model from the global model in the parameter space so that the local model can get closer to the global one as well as stabilizing the stage of local training, such as MOON, FedProx, SCAFFOLD [5-7].

In Section 2, this paper introduces the classical federation learning algorithm FedAvg and then presents the MOON algorithm, which was improved based on FedAvg. The MOON algorithm takes inspiration from contrastive learning. It proposes model-contrastive learning, a new learning concept, to reduce the impact caused by the non-IID data problem [6]. In Section 3, to better use the MOON algorithm, this paper replicates and analyzes this algorithm to provide data reference for subsequent use. This experiment uses the MNIST data set with 20 clients and achieves a global accuracy of 93% with about 80 rounds of learning.

2. Methods

FedAvg is a classic federation learning algorithm. In FedAvg, every round of sessions has four steps [4]. First of all, the server will send initial model parameters to each client. The model parameters are updated continuously by the clients using local stochastic gradient descent. When the predetermined number of local training times is reached, the local models are uploaded to the server. Third, the server randomly selects clients and receives model parameters for aggregation. The server updates the global model parameters by weighting the chosen clients' model parameters and adding them to the global model parameters after the previous round of aggregation. Finally, the aggregated new parameters will send to each client again and then go back to step 1 and continue. The above steps are repeated until the set number of communications is reached. During the training process, there is no need to send any local data to the server. Clients must train on local data to get their local models. Meanwhile, the server oversees the aggregation of models for purpose of getting a global model, which converges to a centralized machine learning result after multiple iterations. It effectively reduces many privacy risks associated with the source data aggregation of traditional machine learning.

There are two categories of current research on enhancing the FedAvg algorithm for non-IID problems: to improve the local training and to improve the aggregation. The MOON algorithm belongs to the first one, which concentrates on improving the local training phase by adjusting the deviation of the local models from the global model [6].

2.1 Model-contrastive Algorithm

The Model-contrastive algorithm, i.e., MOON algorithm, is improved based on the FedAvg. The view of the direction improvement is that models training on the whole data set would extract a better feature representation than models training on a skewed subset. In the local training stage, the skewed distribution of local data will result in a worse representation of the model learning. In comparison, the global model can obtain a more excellent feature representation. Consequently, in non-IID data scenarios, not only do people have to reduce the drift, but they also have to bridge the gap between the model trained locally and the global one.

The specific improvement of MOON over FedAvg is the introduction of model-contrast loss derived from contrastive learning [6]. Contrastive learning concentrates on learning similar features among comparable instances and identifying differences between dissimilar ones [8].

In contrast to generative learning, contrastive learning will not concentrate on the boring details of the instances. However, it only needs to be responsible to learn to classify the data on the feature space. It simply ought to learn how to discriminate the data on the feature space, as opposed to generative learning, which must concentrate on the details of the instances. As a result, the model and its optimization are made simpler and more universal. Learning an encoder which encodes identical data of the same sort and makes the encoding outcomes of various data classes as distinct as feasible is the aim of contrast learning. Contrastive learning reaches the true distance of the original spatial distribution by enlarging the distance between the similar samples as well as decreasing the distance between the dissimilar. This makes the gap between positive samples and anchors far less than the gap between negative samples and anchors, which aims to achieve a clustering-like effect.

2.2 SimCLR

Considering the ease of use and effectiveness, the contrastive learning framework chosen for the MOON algorithm is SimCLR [9]. As part of the work on self-supervised learning, SimCLR naturally follows the idea, which is generally to use a great deal of unlabeled data because of the simplicity of obtaining to pre-train the model. Simclr learns the representations by controlling the contrastive loss and maximizing the augmented views of the same data samples as much as possible.

First, suppose there is an image x, called the original image, and SimCLR does data enhancement on it, which includes three simple enhancements in sequence: random color distortion, random cropping and then returning back, random Gaussian blur, then get two enhanced images x_i, x_j. After that, the enhanced images x_i and x_j are input to a neural network-based encoder, generating two feature vectors. Then, the feature vectors are travelled through a neural network projection header to acquire z_i, and z_j represents the space which applies the contrastive loss. The feature vectors often contain more information about data enhanced transforming such as color and orientation, while the output z removes this extra information and restores the essence of the data. SimCLR maximizes the consistency of different enhanced views from the same instance by maximizing the z_i, z_j of the same image.

The contrastive loss makes a comparison to the representations of different instances, while model-contrast loss compares the representations learned by different models. MOON algorithm focuses on supervised learning and uses it to propose model-contrastive learning to compare the representations learned by different models for dealing with non-IID data distributions between parties in federation learning.

Convolutional Neural Network (CNN), which have great performance on image processing, is the network architecture used in this article. CNN is a feed-forward neural network containing artificial neurons. First, there are two 5x5 convolutional layers for feature extraction and filtering, where the first one has 32 channels while the next one has 64 channels. Both convolutional layers contain no padding, have a stride of 1, and have ReLu activation and 2x2 maximum pooling. After that, it is a fully connected layer which has 512 channels and ReLu activation used for weighted summation. Finally, there is a linear regression output layer.

3. Experimental Results and Analysis

This paper is dedicated to replicating and analyzing this algorithm on the image classification problem. The data set used in the experiments is the MNIST data set [10]. The data set contains 70,000 handwritten black and white photographs labeled with ten numbers from 0-9, each number of figures ranging from 6900 to 7300. In the field of image recognition, the MNIST data set is one of the most widely used data sets.

This experiment sets up a total of 20 clients. In practice, each client should have unique data, but in the simulation, the data set is manually divided and assign it to every client. The data on each client may be either IID or non-IID. Two different scenarios correspond to other allocation methods. The experiment first breaks the data set for the IID case and then assign 3500 samples to each client. Moreover, in regard to the non-IID scenarios, the data set is first sorted according to the data labels, which means the images in the data set are sorted by the label, then divided into 20 groups of data slices of different sizes [4]. Each client will be assigned two slices from two figures. In practice, due to the differences in user habits and duration, the size of the data set differs from one client to another, so the scale of the data allocated to each client is not the same. In this experiment, the amount of data owned by the client is only 1000 at the minimum and 12000 at the maximum, which corresponds to the inactive and active users. The difference in the number of data finally assigned to each client in the case, coupled with the small quantity of clients and the large offset between the data on every client and the complete data, enables the exploration of the extent to the accuracy of the algorithm on highly non-IID data.

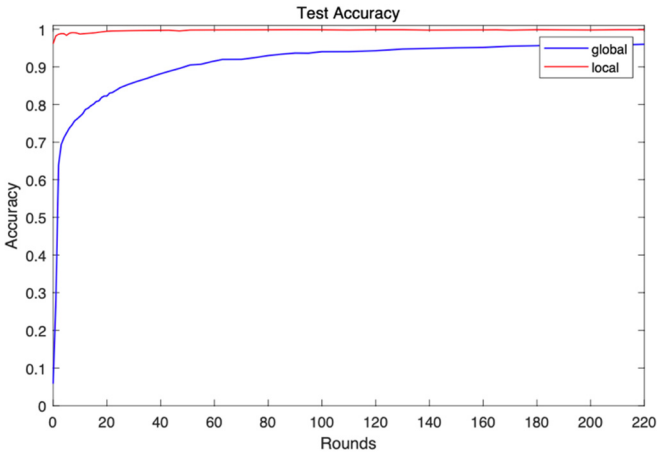

Figure 1. The accuracy growing over time

In this paper, the entire algorithm is implemented by PyTorch. This paper sets the learning rate for local training and the batch size to 0.005 and 10 respectively. The number of clients is set to 20. Based on the small quantity of clients, default the client drop rate to 0 and the joining ratio to 1. The time threshold is set to 10000, which means each client will complete the local model training and sends it to the server. This experiment set the temperature parameter of the model-contrastive loss in the MOON algorithm to 0.5. The communication rounds are set to 250, which provides a larger view of the algorithm's performance.

Table 1. The accuracy at different rounds

Rounds	1	5	10	20	50	100	150	200	220
Global average accuracy	5.83%	72.47%	76.88%	82.26%	90.10%	94.03%	95.06%	95.72%	96.00%
Local average accuracy	96.13%	98.96%	98.72%	99.45%	99.77%	99.83%	99.78%	99.78%	99.85%

For the experimental results in this paper, after the first round of MOON, the accuracy of the global model is 5.83%, and the accuracy of each local model reaches 96.13%. Meanwhile, the accuracy of the global model remains below the one of the local models. After five rounds of MOON, the global accuracy reached 72.47%, while the local accuracy reached 98.86%. The accuracy of both models improved, and the gap between them decreased.

With the number of training rounds increasing, the local accuracy has stabilized. However, the global accuracy continues to improve, and the rate of improvement also becomes smaller. As shown in figure 1, the global model finally reaches about 93% accuracy at about 80 rounds, which meets part of the usage scenarios. After that, if the training rounds increase, the accuracy of the global model will continue to improve, but it will also consume computational memory and communication costs, which are less meaningful.

4. Conclusion

Currently, federation learning has become a meaningful path to make data privacy protected as well as solving the problem of isolated data islands. Meanwhile, solving non-IID data is a fundamental problem to improve the performance of federation learning. On this basis, this paper set experiments to replicate the MOON algorithm, which is one of the feasible ways to mitigate the effect of the non-IID data problem. This algorithm combines federation learning and contrastive learning. It proposes model-contrastive learning, which can powerfully reduce the drift of the local model by adjusting the deviation of the local model from the global model in the parameter space to make the distance between two models smaller. The innovation effectively decreases the influence of the non-IID data problem. This paper tries to provide data references for subsequent use by replicating and analyzing this algorithm.

References

[1] Hinton, G. E., Osindero, S., & Teh, Y. W. (2006). A fast-learning algorithm for deep belief nets. Neural computation, 18(7), 1527-1554.

[2] Wang, J. Z., Kong, L. W., Huang, Z., et al. (2020). Research review of federated learning algorithms. Big Data Reserch, 6(6), 70-88.

[3] Kairouz, P., McMahan, H. B., Avent, B., et al. (2021). Advances and open problems in federated learning. Foundations and Trends® in Machine Learning, 14(1–2), 1-210.

[4] McMahan, B., Moore, E., Ramage, D., et al. (2017, April). Communication-efficient learning of deep networks from decentralized data. In Artificial intelligence and statistics (pp. 1273-1282).

[5] Karimireddy, S. P., Kale, S., Mohri, M., et al. (2019). SCAFFOLD: Stochastic Controlled Averaging for On-Device Federated Learning.

[6] Li, Q., He, B., & Song, D. (2021). Model-contrastive federated learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 10713-10722).

[7] Li, T., Sahu, A. K., Zaheer, M., et al. (2020). Federated optimization in heterogeneous networks. Proceedings of Machine Learning and Systems, 2, 429-450.

[8] He, K., Fan, H., Wu, Y., et al. (2020). Momentum contrast for unsupervised visual representation learning. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 9729-9738).

[9] Chen, T., Kornblith, S., Norouzi, M., et al. (2020, November). A simple framework for contrastive learning of visual representations. In International conference on machine learning (pp. 1597-1607).

[10] Deng, L. (2012). The mnist database of handwritten digit images for machine learning research [best of the web]. IEEE signal processing magazine, 29(6), 141-142.