哈尔滨工业大学（深圳）

Harbin Institute of Technology ,Shenzhen

Interim Assessment of the Thesis for the Master’s Degree

Name	**Songyue Guo**
Entrance Date	**2021.9.1**

Thesis Title

Federated Learning Algorithm Research for Imbalance Data Distribution

Discipline

School of Computer Science and Technology

Supervisor	**Qing Liao**
Report Date	**2023.5.21**

1. Does the thesis progress according to the research objectives and schedule as stated in the primary report? (at least 100 words)

Yes！

At present, we have proposed two main federated learning framework FedAIM, FedGR for Imbalance Data Distribution. What's more, we have completed their code desgin based on pytorch and taken enough experiments like accuarcy comparison experiments, convergence rounds experiments, superparameter analysis experiments, visiualization experiments to prove the efficiency and effective of FedAIM and FedGR. We have still theoretical analyze their capacity to tackle heteregeous data in federated learning. I also have written two papers and submitted them to CCF T1 *Journal of Computer Research and Development* and CCF B Internaional Conference *DASFAA2023*. These two paper have all been accepted and published yet.

2. The completed work and its related outcomes (at least 1500 words) .

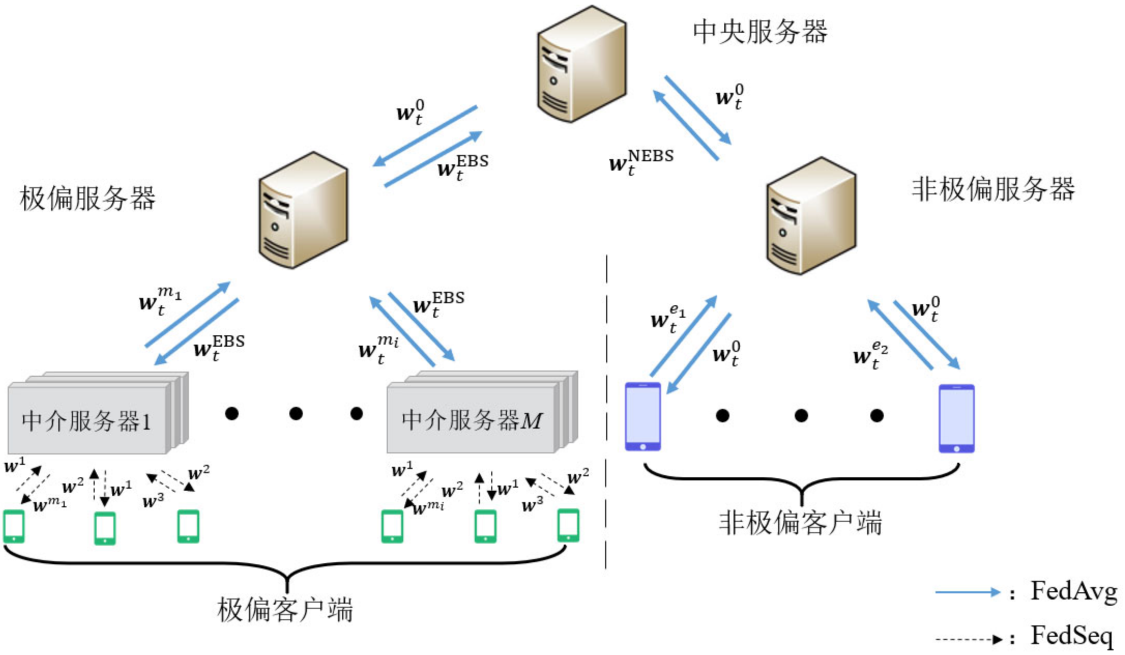
Despite the success of deep learning in numerous fields, a data center training model is typically required. In some real-world applications, individualparticipant data cannot be located on the same device due to data privacy .Federated Learning (FL) is designed for data privacy protection and efficient distributed training.

The advent of FL enables different clients to collectively build a robust globalmodel without broadcasting local private data to the server. FL has demonstrated its ability to facilitate real-world applications in several domains, e.g.,natural language processing, credit card fraud detection and medicalhealthcare. FL, however, also confronts the challenge of imbalance distribution.The imbalance distribution of Non-IID data between clients brings serious performance degradation problems for FL. This performance degradationis attributed to the phenomenon of client drift. Some recent works aim to dealwith this problem, e.g., FedProx included l2 regularizer term to prevent local models from deviating too far from the global model, PerFedAvg utilized contrastive learning ,and MOON used multi-task learning for fast client local adaptation to mitigate the impact of client drift. However, most existing studies focus on single imbalance distribution.

In this work, we focus on double imbalance distribution scenario, which is more common in the real world. We first define the imbalance distribution into two categories:

1) Label imbalance. According to Fig. 1(a), we simulate this scenario with 10 clients on CIFAR-10 dataset. The majority of clients have part labels of the whole, and client's labels are mostly different from others, while the quantity of each label in client is equal. For example, client 1 owns labels 4, 7, 9, but client 2 owns labels 0, 2, 8. Most recent works like FedProx, SCAFFOLD and FedNova only care about label imbalance.

2) Quantity imbalance. We still use 10 clients on CIFAR-10 to describe this imbalance distribution. Based on Fig. 1(b), each client owns an entire set of labels, but the quantity of each label in client varies, i.e., client 1 has 10 labels, and the number of label 9 is about 450, but the number of label 0 and 1 is approximately 0.



In this study, we focus on double imbalance distribution like Fig. 1(c), each client owns a partition of entire labels, and the quantity of each class in client varies, e.g., client 1 only possesses labels 4, 7, 9, and each class' sample number is imbalanced. It is clear that the double imbalance distribution scenario is more in line with reality than any single imbalance scenario. However, existing works mostly omit the real scenario of double imbalance distribution scenario.

In order to investigate the performance of existing IID and Non-IID FL algorithms for double imbalance scenario, we use client distribution as Fig. 1 to design an observation experiment. We use TFCNN6 as client's base model and implement all compared FL methods with the same model for a fair comparison. All performance results are expressed by accuracy of an average of five times. The experiment results are summarized in Fig. 2. From the performance results shown, we can find that no matter what FL algorithms gets a significant performance loss on the double imbalance, compared to the left two scenes. For example, the accuracy of FedAvg declines by about 16%. Apparently, the double imbalance scenario brings a new challenge for existing FL algorithms, which is the goal of our study tries to solve.

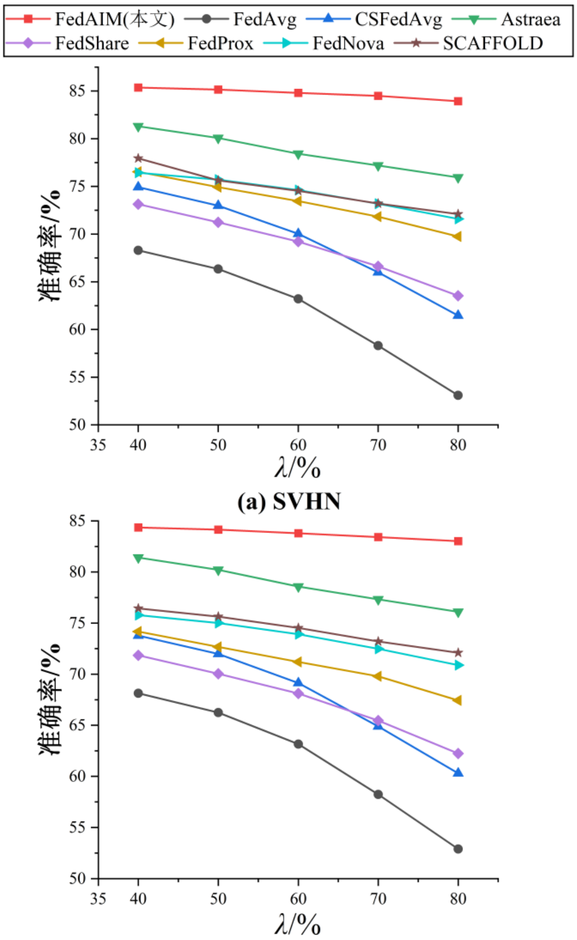
Motivated by the above observation experiment of double imbalance distribution, we propose a novel FL algorithm called Federated Learning withGravitation Regulation (FedGR) to deal with this problem. We define a novel softmax function called unbalanced softmax to balance the importance of classes under quantity imbalance in clients. In addition, we propose an efficient gravitation regularizer to deal with label imbalance among clients by encouraging collaboration among clients. Combining these two components, we can correct the gradient of traditional loss function of typical FL methods. The contributions of this paper can be summarized as follows:

－ We propose a novel federated learning method FedGR to effectively deal with the performance degradation problem caused by double imbalance distribution scenario.

－ We design an unbalanced softmax function, which can solve the problem of unbalanced number of samples within the client by adjusting the forces of positive and negative samples on classes.

Table 2-2 Accuracy Comparison with SOTA %

Method	**MNIST**	**CIFAR-10**	**Fashion-MNIST**	**SVHN**	**FEMNIST**
FedAIM(ours)	**98.54**	**82.57**	**84.34**	**85.35**	**82.13**
Astraea	97.92	80.11	81.36	81.41	79.46
FedShare	95.25	72.64	73.12	71.84	68.97
FedAvg	70.13	65.47	68.38	68.13	58.49
CSFedAvg	98.27	75.83	74.97	73.78	70.72
FedProx	96.65	74.49	76.54	74.17	75.69
FedNova	97.23	79.47	80.73	75.78	76.65
SCAFFOLD	94.42	75.73	77.94	76.44	78.54



(a) SVHN

(b) Fashion-MNIST

It can be seen that FedGR achieves the highest performance in terms of accuracy and F1 with different degrees of double imbalance. The highest performance improvement of FedGR compared to baseline occurs on the CIFAR-10 dataset where the client has only two labels (about 4.53%, 4.91% accuracy and the next best result for F1). pFedMe achieves the lowest performance in most cases of double imbalance, even lower than FedAvg. The possible reason is that meta-learning is not useful for dealing with severe quantitative imbalances on the client side is not useful. FedRS achieves the next best result in most scenarios because it limits the erroneous updates of missing classes, but quantitative imbalances are still a challenge for it.

Table 2-3 Accuracy Comparison of FedGR and SOTA

Algorithms	CIFAR-10 (2)		CIFAR-10 (3)		CIFAR-10 (20)		CIFAR-100 (30)	
	Acc(%)	F1(%)	Acc(%)	F1(%)	Acc(%)	F1(%)	Acc(%)	F1(%)
FedAvg	50.36	48.27	53.79	49.42	36.15	34.10	42.19	40.42
FedProx	48.84	46.96	54.94	53.85	36.24	34.42	42.21	41.09
FedNova	56.33	54.59	68.63	66.09	38.63	37.72	45.35	45.59

SCAFFOLD	57.37	54.53	67.32	62.44	38.43	37.76	46.82	45.44
PerFedAvg	44.67	42.56	54.87	53.73	35.98	34.76	40.14	40.33
pFedMe	45.81	44.35	50.18	50.24	35.36	33.59	40.18	40.54
FedOpt	62.37	60.68	70.63	69.79	42.37	40.68	49.63	49.79
MOON	61.45	60.71	72.91	70.45	40.53	41.46	47.91	48.76
FedRS	63.22	60.13	73.56	70.13	42.76	42.21	50.73	50.31
FedGC	62.91	60.35	72.11	70.64	42.11	40.35	50.21	50.46
FedGR(ours)	**67.84**	**65.62**	**77.86**	**75.32**	**45.44**	**44.85**	**53.16**	**53.32**

We also compare the convergence speed with the comparison methods. We chose FedAvg's accuracy after one thousand rounds as the standard and then compared the number of communication rounds required by the other methods to reach this accuracy. The number of communication rounds required by the other methods to achieve this accuracy is then compared. The results are shown in Table 2-4. We removed PerFedAvg and pFedMe from Table 2-4 because these two methods could not achieve the precision of FedAvg. According to Table 2-4, FedGR can get the minimum number of communication rounds (from 300 to 390 rounds) to achieve FedAvg accuracy on all data sets, which is about 2 times faster than FedAvg. 2 times faster than FedAvg. The reason why FedGR requires fewer rounds is that the gravitational regularizer encourages effective cooperative collaboration among different clients, which is not taken into account by existing FL methods.

Table 2-4 Convergence Rounds Comparison

Algorithms	CIFAR-10 (3)		CIFAR-100 (30)	
	#rounds	speedup	#rounds	speedup
FedAvg	1000	1x	1000	1x
FedProx	800	1.25x	850	1.17x
FedNova	600	1.67x	650	1.53x
SCAFFOLD	860	1.16x	830	1.20x
Fedopt	390	2.56x	450	2.22x
FedRS	350	2.85x	430	2.32x
FedGC	590	1.69x	650	1.53x
FedGR(ours)	**330**	**3.03x**	**390**	**2.56x**

3. The work to be completed and its schedule .

The subsequent work to be completed is mainly thesis writing, and revise the thesis according to the supervisor's guidance and suggestions to complete the thesis defense.

The specific schedule is as follows:

May 2023: Finish the empirical content that has been completed and complete the mid-term defense. Organize the mid-term inspection review recommendations and complete the writing of the first draft of the paper.

June 2023: Ask the supervisor for suggestions on revising the first draft of the thesis, and complete the thesis revision.

August 2023: Continue to revise the thesis and prepare for defense.

4. The existing or expected difficulties and problems.

(1) Firstly, the subject requires a lot of mathematical knowledge to be learned to carry out, involving non-convex optimization, multi-objective optimization, and optimization solving. Secondly, the experimental design aspect of this topic requires careful consideration and full analysis to measure the performance of the algorithm.

In view of the current problems and difficulties, the following solutions are proposed:

In order to improve the speed and efficiency of the progress of the subject and to ensure the quality of learning, the method of learning and re-application while applying is adopted for the systematic learning of the main knowledge, and more hands-on practical operation to provide hands-on ability and also lay a good foundation outside the later stage. Also adopt the following programs:

In the design of experiments, read more other top meeting top journal papers, learn how to write papers, learn to optimize mathematical knowledge, and enhance the ability to code

5. The considerations on the possibility of completing the thesis on-time (at least 100 words).

In view of the current problems and difficulties, the following solutions are proposed:

(1) Ask the tutor for advice on the construction of possible paper modification, actively explore previous references, and find more effecitve method as much as possible to complete the algorithm desgin of this paper.

(2) Strive to prove the convergence analysis on algorithm theoretical analysis.

At present, all the empirical content of the paper and the preliminary preparation of the paper have been completed, and the results are also in line with all the assumptions proposed in the early stage of this paper.

I am currently writing the content of the paper.

In summary, the thesis can be completed on schedule and achieve certain research results.