Review of IEEE Edge 22

Review 1:

2: (accept)
Handling data heterogeneity in federated learning has received considerable attention as the data distribution may not be IID in all participators. In this paper, the authors argue that the widely used FEDAVG, as a centralised solution, is not suitable to be deployed in edge computing systems as it is a resource-intensive solution. Meanwhile, the authors also state that the existing data heterogeneity techniques are either resource-intensive or privacy breaching. To address these problems, they propose a lightweight algorithm that only uses the local training loss of FL clients for auto-weighting the model aggregation. The algorithm is presented with the pseudo-code, overall idea and functions description. One could argue that the theoretical analysis falls short, but as an "idea-testing" work is with enough details for the audience. The experimental results look promising. As indicated in the future work, the authors still have a long way to go to provide a practical version for handling data heterogeneity in federated learning. I believe it is still good timing to share this work with the community and carry the work forward.

The authors are also expected to investigate the communication efficiency of their solution as the edge device is not only limited by computation capacity but also energy (lifetime-related). Several typos are found in the paper, e.g. auto-weighed(in the abstract, it should be auto-weighted). Please carefully check the text when doing the revision.

Review 2:

2: (accept)
This paper proposes a weight aggregation scheme for federated learning with clients reporting bad models. I enjoyed reading this manuscript: since this is well-organized, I could understand what approaches exist, why something matters, and why they propose a new scheme step by step. In particular, I like the experiment-based demonstration of the problem situations and the logical flow in this paper to address the problems.

But, I have some questions or recommendations as follows:

1. It would be better to explain more about the rationale behind the approach. It might be straightforward that the training loss larger than some threshold indicates the corresponding model is not good for updating the model. But I am guessing that it depends on the distribution of client data (see question 2).

2. Looking at the setting for experiments where data are uniformly divided, I assume that clients have training data containing all classes. What if some clients have training data with many classes and others have training data with a few ones, i.e., non-IID distribution. Assuming the former is more difficult to train than the latter, it might be possible that the former yields a larger training loss, thus, is likely to be excluded from the model update even though it can provide meaningful information. In this case, FedASL might not work as we want. Is it right? Then any countermeasure?

Review 3:

0: (borderline paper)
This paper proposed a FedASL algorithm that is FL with auto-weighted aggregation based on standard deviation of training loss using only the local training loss of FL clients for auto-weighting the model aggregation. In performance evaluation section, three datasets and data corruption scenarios are evaluated, and computation cost of the FedASL is much lower than those of existing approaches. This paper still has some problems to improve readability.

The main contribution is not well presented, the main idea and some details of the proposed algorithm FedASL is introduced in the contribution part. This paper is not well written and organized, and the organization is out of order. The readers is quite difficult to get the contribution. The phraseology and expressions need to be much condensed to improve readability and clarity.

Review 4:

2: (accept)
This paper is very well written and timely. It addressed key issues in Federated computing. This paper manages to address the FL's heterogeneity issue by providing a new approach named: FedASL, which can automatedly aggregate the clients' updates based on its loss. Some suggestions below to enhance the paper in the final version.

Firstly, this paper lacks consistency and obfuscates the updates heterogeneity because of non-iid and byzantine failure. In 'Motivation,' the study identifies the heterogeneity that comes from the various trading data of different clients but consequently compares with Krum, Trimmed mean FL methods, which are proposed to relieve the byzantine failure. Thus, it is important to clarify what the target issue is focusing on, the non-iid issue or byzantine failure.

Secondly, the proposed FedASL method highly relies on the Median loss while lacking median is unbiased. In other words, the median is biased when all malicious clients manipulate the updates toward one direction and subsequently draw a biased good region.

Thirdly, the performance of FedASL is just shown somehow in experiments that achieve an average of 4% higher global accuracy than the existing FL methods. Moreover, the experiments could include more state of art attacks and FL methods.

Some typo mistakes: idd clients should be iid clients on page 1 right side.

Review 5:

0: (borderline paper)
This paper studies model aggregation in federated leaning. It proposes a novel auto-weighted method for model aggregation which can reduce the computation for statistical analysis of clients' model updates. Specifically, the server determines the aggregation weights of clients' local models based on only clients' local training losses using "a good region". Extensive simulations show that the proposed algorithms offer better or similar model accuracy compared to existing techniques while avoiding the heavy computation.

+It looks novel to use local training loss of FL clients for auto-weighting to avoid extensive computation.
+The prior work are clearly discussed and compared with the proposed algorithms.
+The simulations are comprehensive.

-Some related works on fault-tolerant FL should be discussed and compared.
-The paper does not provide any theoretical analysis and performance guarantee of the proposed approach.
-The rationality of setting the good region based on the median of losses in (6) should be better justified.
-It is mentioned that tunable parameters can discriminate clients outside the good region. It is not clear at all how the weights decided by tunable parameters and local training losses will benefit the model training.
-Why does Algorithm 1 always select K agents at random in T rounds?
-Please discuss more about the criteria for selecting tunable parameters.
-There are some grammatical mistakes, e.g. on page 2, the sentence "the local datasets for of the k-th client is D_k" needs to be corrected.