Azerbaijan Journal of High Performance Computing, Vol 6, Issue 2, 2023, pp. 153-162
https://doi.org/10.32010/26166127.2023.6.2.153.162

Application of AHP for Weighting Clients in Federated Learning

Samir Aliyev

Azerbaijan State Oil and Industry University, Baku, Azerbaijan, aliyev.samir@asoiu.edu.az

*Correspondence: Samir Aliyev, Azerbaijan State Oil and Industry University, Baku, Azerbaijan, aliyev.samir@asoiu.edu.az

Abstract

Federated Learning is a branch of Machine Learning. The main idea behind it, unlike traditional Machine Learning, is that it does not require data from the clients to create a global model, so clients keep their data private. Instead, clients train their model on their own devices and send their local model to the server, where the global model is aggregated and sent back to clients. In this research work, the Federated Averaging algorithm is modified so that clients get their weights by the Analytical Hierarchal Process. Results showed that applying AHP for weighting performed better than giving clients weights solely based on their dataset size, which the Federated Averaging algorithm does.

Keyword: Federated Learning, AHP, Geometric Mean, Client Weighting, Federated Averaging

1. Introduction

Federated Learning is a newly emerging machine learning technique that emphasizes decentralized model training and privacy. Sensitive data is kept local and is not sent to a central server in this collaborative paradigm because models are trained across several decentralized devices or servers. Through iterative learning, each participating device improves its local model; only the updated models are transmitted to the central server. This allows for ongoing development without jeopardizing the privacy of personal information. Federated learning is beneficial when data is dispersed among multiple sources, including IoT, edge, and mobile devices. Applications in healthcare, finance, and other industries where data privacy is a concern benefit greatly from its ability to increase efficiency and decrease the need for large-scale data transfers. One of its main advantages is its ability to handle non-IID (non-Independently and Identically Distributed) data distributions, which addresses real-world scenarios where data characteristics may vary considerably.

Aggregation is a crucial phase in the cooperative model training process in federated learning. Once individual servers or devices have trained on their respective datasets to create local model updates, these updates are combined at a central server to create the aggregated model. Calculating the weighted average of the model parameters across all participating devices is a common step in the aggregation process. Depending on the particular needs of the FL system, various aggregation techniques, such as simple or weighted averaging, can be used. One of these methods is Federated Averaging

(FedAVG). In the FedAVG model, parameters are weighted and averaged. Weighting is respective to the size of the dataset clients used for training. The main goal of this work is to propose different methods to give weights to clients using other facts such as computing capabilities and distribution of classes in the clients using expert systems.

Section 2 is dedicated to related research on model aggregation and client weighting. Section 3's central methodology is about how our approach works for aggregation in FedAVG using the Analytical Hierarchal Process (AHP). The results, comparisons, and training curves are demonstrated in section 4.

2. Literature Review

Several works have been devoted to aggregating models' weights in federated learning. Federated averaging is one of the earliest and most cost-effective methods used in federated learning (T. Sun, D. Li and B. Wang, 2023). However, it is a very naive approach that only uses the training set size to give the weights to the client. In this case, clients with small data set sizes can hardly influence the model. Considering that some clients may have "healthier" data even at small sizes, assigning more weights can potentially improve the performance of the global model (Li, Y., Guo, Y., Alazab, M., Chen, S., Shen, C., & Yu, K., 2022). To overcome this difficulty, several works have been carried out (Qi, P., Chiaro, D., Guzzo, A., Ianni, M., Fortino, G., & Piccialli, F., 2023).

One such work is done by applying forgetting (Xu, C., Hong, Z., Huang, M., & Jiang, T., 2022). Forgetting in Federated means that one observation is classified correctly on the local model, but the global model should have classified that observation. In this case, the new models will likely forget previous observations because the global model aggregates the local model. Because of this local model, performance decreases when new batches are tested. To overcome this difficulty, Federated Weighted Averaging (FedWAvg) was proposed (Hong, M., Kang, S. K., & Lee, J. H., 2022). It gives clients weights based on the local forgettable examples. The clients with more forgotten examples. This rebalances the global model and makes clients with forgetful examples less affected by global updates. Experiments showed that the proposed approach performed better than the previous algorithm.

In work (Ye, R., Xu, M., Wang, J., Xu, C., Chen, S., & Wang, Y., 2023), the authors considered local and global distribution through the FedDisco algorithm. The results showed that there are better ways than just using the size of the dataset to weight clients. FedDisco algorithm outperformed the FedAvg accuracy score by about 17%. In another work (Tang, Z., Shao, F., Chen, L., Ye, Y., Wu, C., & Xiao, J., 2021), weights are assigned to clients dynamically. Instead of giving them static weights, the authors proposed a method that assigns weights based on the contribution in each round. The experiments were carried out in CIFAR-10 and MNIST. The proposed approach increased performance. The disadvantage of this approach is that the weights must be calculated in each round of contributions.

Expert systems were also used to assign weights to the customer. One of these works was carried out to overcome the problem of non-IID data (Wilbik, A., & Grefen, P., 2021). In another work, the authors used various criteria, such as computing capacity and network resources, to give weights to clients (Du, Z., Wu, C., Yoshinage, T., Zhong, L., & Ji, Y., 2022). Fuzzy Inference Systems carried it out. In order to test the different calculation capabilities, different devices were used for training. The FIS system was also applied in federated learning in work by (Aliyev, S. & Ismayilova; N., 2023). This work also took into account the distribution of classes for each customer. Computational capacity, class distribution, and dataset size were passed to the Mamdani-type FIS as input, and the clients' weights were the output. The results showed that applying these criteria increased the performance of the global model. This research work is dedicated to the application of the Analytical Hierarchal Process for weighting the clients.

3. Methodology

3.1. AHP

Analytical Hierarchal Process is a multicriteria decision making method developed by Saaty (Saaty, R. W., 1987). In AHP, decision-makers must determine and understand the problem and its goal. Factors that can influence the decisions are evaluated and compared pairwise. These comparisons are given some value based on the importance of one factor over the importance of another. These values are evaluated by experts based on the state-of-art results, problems, experience from previous cases, and learning. The importance of the factor is taken as a value in the range of 1 and 9. One such scale adopted by Saaty is shown in Table 1 (Saaty, R. W., 1987).

Table 1. Importance scale

Value	Meaning	Description
1	Equal	Two factors influence the decision equally
3	Somewhat more important	One factor is slightly more important than the other for decision
5	Much more important	One factor is strongly preferred over the other in order to make a decision
7	Very much more important	One factor is much more preferred than the other. Its importance is already observed in practice
9	Absolute importance	The importance of one factor over another is evaluated in the highest possible validity.
2,4,6,8	Intermediate values	When compromises are needed.

Each value of the matrix is given value as:

$$A_{ij} = \frac{1}{A_{ji}}$$

AHP usually consists of three steps.

Decomposition: This step decomposes the problem in hierarchal form. Objectives and criteria are detected. If the hierarchy goes deeper, sub-criterions are also taken into consideration.

Pairwise comparison: In this step, the experts create a pairwise comparison matrix using Saaty's 9-point scale. Giving values to each comparison is the central concept in AHP. The priority vector is calculated using this matrix. Several methods for estimating the priority vector include geometric mean, fuzzy geometric mean, eigenvector method, etc.

Composition of priorities: In this step, priorities, starting from the lowest level to the highest level, are synthesized. The goal is to obtain overall priority that reflects the importance of each alternative.

3.2. Geometric Mean

Geometric Mean is one of the prioritization methods of AHP (Yadav, A., & Jayswal, S. C., 2013). The steps of this prioritization method are as follows:

Find the geometric mean of each row

$$GM_i = \left(\prod_{j=1}^{n} a_{ij}\right)^{\frac{1}{n}} \quad i = 0,1,\dots,n$$

Sum them up

$$The\ sum\ of\ Rows = \sum_{i=1}^{n} GM_i$$

Normalize each row by dividing by the sum of rows to obtain the priority vector

$$Vector_i = \frac{GM_i}{Sum\ of\ rows}$$

Calculate the CR:

$$CI = \frac{\lambda_{max} - n}{n - 1}$$

$$CR = \frac{CI}{RCI}$$

Where λ_{max} is the maximum eigenvalue of the matrix, and RCI is provided by:

If CR is less than 0.1, then the matrix is acceptable. Note that if CR is greater than 0.1, it does not necessarily mean the matrix is inconsistent.

Table 2 RCI table

N	RCI
1	0
2	0
3	0.5799
4	0.9
5	1.12
6	1.25

7	1.33
8	1.39

3.3. Steps

The following preference matrix used in this work is described in Figure 1:
Geometric means of each row:

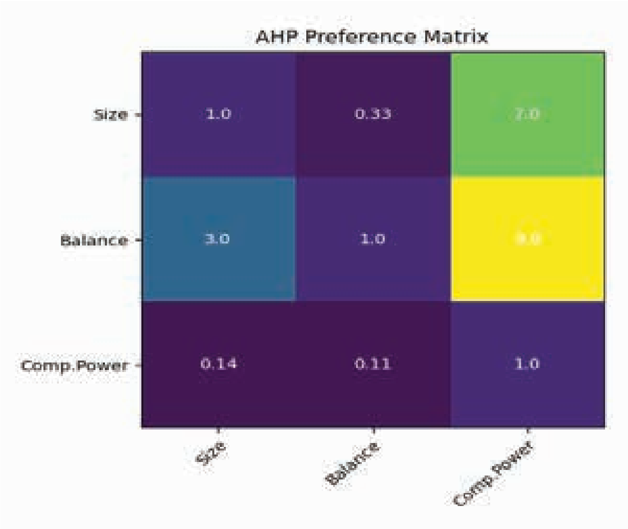

Fig. 1: Preference matrix used in this work

$$\sqrt[3]{1 * 0.33 * 7} = 1.32$$
$$\sqrt[3]{3 * 1 * 9} = 3$$
$$\sqrt[3]{0.14 * 0.11 * 1} = 0.25$$

The sum of them: 1.32+3+0.25 = 4.57

Priority vector is: $[\frac{1.32}{4.57}, \frac{3}{4.57}, \frac{0.25}{4.57}] = [0.29,\ 0.66, 0.05]$

Maximum eigenvalue is 3.08. Then CR is:

$$CI = \frac{3.08 - 3}{3 - 1} = 0.04$$
$$CR = \frac{0.04}{0.5799} = 0.068$$

CR is less than 0.1, which implies the matrix is consistent and can be applied for decision making.

3.4. Logistic Regression

Logistic Regression is a parametric linear machine learning algorithm used for classification tasks. It uses the sigmoid function to predict (Roberts, G., Rao, N. K., & Kumar, S., 1987). Sigmoid is an "S" shaped function. The threshold value controls the classification boundary. Because the distribution of classes among clients is not known in federated learning, 0.5 was used as a threshold for all clients. Sigmoid function can be defined as:

$$\sigma(x) = \frac{1}{1 + e^{-z}}$$

where z is:

$$z = \beta_0 + \sum_{i=1}^{k} \beta_i x_i$$

Where vector x is input and vector β are the model parameters.

The binary cross entropy function was used as a loss function and optimized by gradient descent algorithm.

3.5. Federated Averaging

Federated Averaging is one of the original approaches for aggregating model parameters. Federated averaging can be defined as:

$$For\ each\ client\ k, w_{i+1}^{k} \leftarrow w_i^k - \alpha g_i$$

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

where α is learning rate, g_i is gradient in i^{th} iteration for each client. n_k is the size of a training set of client k, and w_{i+1}^{k} is the updated weights in iteration i+1 for each client k.

The parameters of the model can be updated for several iterations before sending it to the server, reducing communication costs and making parallelization easier.

3.6. Our Approach

Our approach changes the FedAVG algorithm so clients get weights based on AHP results. So the overall formula is defined as:

$$For\ each\ client\ k, w_{i+1}^{k} \leftarrow w_i^k - \alpha g_i$$

$$w_{i+1} \leftarrow \sum_{i=1}^{k} A_k w_{i+1}^{k}$$

where the A_k is the weight of the client k obtained by AHP

Results

Accuracy, Precision, Recall, and F1 score were used as evaluation criteria. Results of original Federated Averaging were compared with the approach where AHP estimated the client weights. These criteria are calculated as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$Recall = \frac{TP}{TP + FP}$$

$$Precision = \frac{TP}{TP + FN}$$

$$F1\ Score = \frac{2 * Recall * Precision}{Recall + Precision}$$

F1-score was taken as the primary evaluation criterion since the training set for some clients is more imbalanced. After each iteration, weights were sent back to clients to evaluate their test dataset.

Experiments

For experiments, the MAGIC Gamma Telescope Dataset was used. This dataset consists of 19020 rows and 11 columns with two classes: gamma and hydron. For experimental purposes, the dataset is distributed among clients with different sizes and imbalances. The summary of clients is demonstrated in Table 3.

Table 3. Client information

Clients	Size	Class 'g'	Class 'h'	Gini	Comp. Power (GHz)
1	4000	2000	2000	0.5	4.5
2	6300	4500	1800	0.4	3.0
3	3500	2000	1500	0.48	1.5
4	1100	500	600	0.49	4.5
5	4120	3332	788	0.31	3.0

AHP returned the following weights for each client:

- Client 1 – 0.242
- Client 2 – 0.199
- Client 3 – 0.171
- Client 4 – 0.222
- Client 5 – 0.164

The following table demonstrates the accuracy, recall, precision, and f1 score of both the ahp approach and the original FedAVG:

Table 4 Results and comparisons. Numbers in bold show the better result between the original and our approach

Device	Original				Our approach			
	Acc	Rec	Pre	F1	Acc	Rec	Pre	F1
1	0.70	0.77	0.54	0.64	0.73	0.75	0.61	0.67
2	0.76	0.60	0.52	0.56	0.76	0.57	0.60	0.59
3	0.68	0.64	0.47	0.55	0.71	0.66	0.59	0.62
4	0.68	0.83	0.53	0.65	0.69	0.80	0.58	0.67
5	0.81	0.47	0.57	0.51	0.79	0.43	0.66	0.52
Aver-age	0.72	0.66	0.53	0.58	0.73	0.64	0.61	0.61

From the results, recall score is generally better in original weighting. However, except for one client, the AHP approach outperforms in Precision, F1 score, and accuracy.

F1 curves are demonstrated in the following figure. For each iteration, we can see that the AHP approach demonstrates better results and converges faster. It can be caused by taking computation power into account.

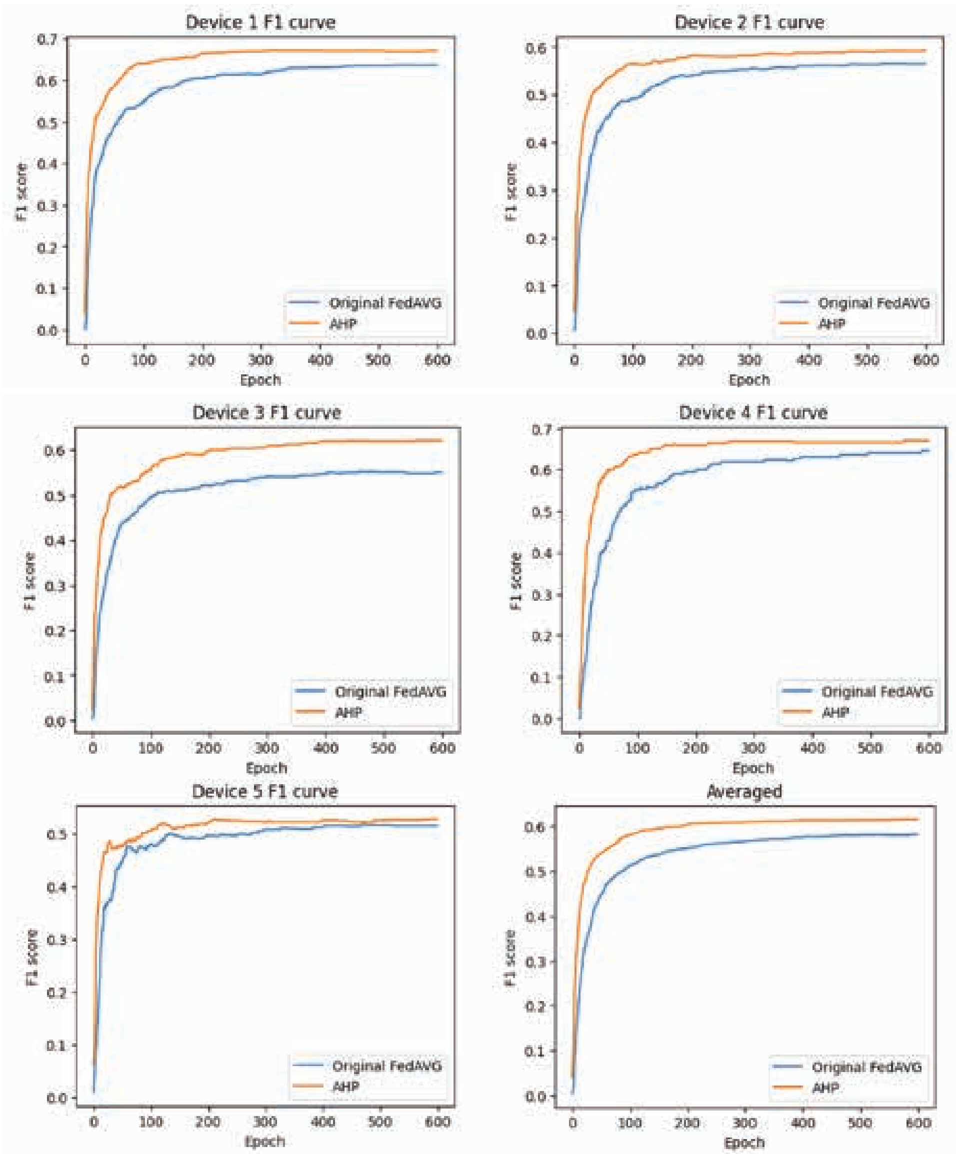

Fig. 5: F1 curves. (a),(b),(c),(d),(e) shows the curves of clients respectively. The curve in the (f) represents the average F1 score of clients.

Conclusion and Future Works

Results show that adding other criteria to deciding the weights performed better than the FedAVG algorithm, which only considers the dataset size. Using AHP also made the global model converge faster. The Only downside was the recall score decline, whereas other metric measures were better.

Work dedicated to increasing the recall score can be done in the future. Other open research questions are how it will perform if the number of clients is bigger. The preference matrix discussed in the Methodology section is open to be modified. It can be done by changing preferences or adding other criteria. Applying this to bigger models, such as neural networks, is another work that can be an extension of this research.

References

Aliyev, S., & Ismayilova, N. (2023, October). FL2: Fuzzy Logic for Device Selection in Federated Learning. In *2023 IEEE 17th International Conference on Application of Information and Communication Technologies (AICT)* (pp. 1-6). IEEE.

Du, Z., Wu, C., Yoshinage, T., Zhong, L., & Ji, Y. (2022, October). On-device federated learning with fuzzy logic based client selection. In *Proceedings of the Conference on Research in Adaptive and Convergent Systems* (pp. 64-70).

Hong, M., Kang, S. K., & Lee, J. H. (2022). Weighted averaging federated learning based on example forgetting events in label imbalanced non-iid. *Applied Sciences, 12(*12), 5806.

Li, Y., Guo, Y., Alazab, M., Chen, S., Shen, C., & Yu, K. (2022). Joint optimal quantization and aggregation of federated learning scheme in VANETs. *IEEE Transactions on Intelligent Transportation Systems, 23*(10), 19852-19863.

Qi, P., Chiaro, D., Guzzo, A., Ianni, M., Fortino, G., & Piccialli, F. (2023). Model aggregation techniques in federated learning: A comprehensive survey. *Future Generation Computer Systems.*

Roberts, G., Rao, N. K., & Kumar, S. (1987). Logistic regression analysis of sample survey data. *Biometrika, 74*(1), 1-12.

Saaty, R. W. (1987). The analytic hierarchy process—what it is and how it is used. *Mathematical modelling, 9*(3-5), 161-176.

Sun, T., Li, D., & Wang, B. (2022). Decentralized federated averaging. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 45*(4), 4289-4301.

Tang, Z., Shao, F., Chen, L., Ye, Y., Wu, C., & Xiao, J. (2021). Optimizing federated learning on non-IID data using local Shapley value. In *Artificial Intelligence: First CAAI International Conference, CICAI 2021, Hangzhou, China, June 5–6, 2021, Proceedings, Part II 1* (pp. 164-175). Springer International Publishing.

Wilbik, A., & Grefen, P. (2021, July). Towards a federated fuzzy learning system. In *2021 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)* (pp. 1-6). IEEE.

Xu, C., Hong, Z., Huang, M., & Jiang, T. (2022). Acceleration of federated learning with alleviated forgetting in local training. *arXiv preprint arXiv:2203.02645.*

Yadav, A., & Jayswal, S. C. (2013). Using geometric mean method of analytical hierarchy process for decision making in functional layout. *International Journal of Engineering Research and Technology (IJERT), 2*(5).

Ye, R., Xu, M., Wang, J., Xu, C., Chen, S., & Wang, Y. (2023). FedDisco: Federated Learning with Discrepancy-Aware Collaboration. *arXiv preprint arXiv:2305.19229.*

Submitted 19.09.2023
Accepted 07.11.2023