Federated Learning in Wireless Security

Ekansh Singh
Department of Electrical Engineering
ee1200490@iitd.ac.in

Tanish Singh Tak
Department of Electrical Engineering
ee1200560@iitd.ac.in

Dr. Harshan Jagadeesh
Department of Electrical Engineering
jharshan@ee.iitd.ac.in

***Abstract*—The advent of data privacy regulations and concerns has led us to a new era of secure data sharing. Faced with a scenario where a third party is unwilling to share data due to privacy constraints, we try to look for an innovative solution using Federated learning and cryptographic methods.In this report we will present to you the problem statement which we would like to solve, our work and results until now and our next course of action**

***Index Terms*—Federated Learning, privacy, neural networks, ensemble learning**

I. Introduction

This mid-term report explores the innovative application of federated learning, a privacy-preserving machine learning paradigm. Federated learning is a machine learning approach that allows multiple parties to collaboratively train a shared machine learning model while keeping their data decentralized and private.

In Federated Learning, merging multiple models to boost accuracy is a clear objective. Ensemble learning plays a crucial role in this context, harnessing collective data from diverse models to enhance overall performance and reliability. Ensemble learning includes various techniques like (Weighted) Average Method, Bagging, Boosting, Stacking, tailored to specific aspects of model collaboration. We used the (Weighted) Average Method in our initial work to gain some insights on how merging of models can increase the accuracy. We would try to quantify how diverse is the third party dataset compared to ours so as to achieve more diversity without any compromise to privacy of data.

II. Our Work

We constructed a neural network utilizing the MNIST dataset for the purpose of image classification. In order to investigate the impact of weights and bias sharing between two distinct models, our approach involved the training of one model on data comprising images falling within the range of zero to four, while concurrently training the second neural network on a distinct dataset comprising images falling within the range of five to nine. Subsequently, we conducted testing on both of our models employing two distinct datasets: one comprising images within the range of zero to four, and the other comprising images within the range of five to nine. Following this initial evaluation, we merged our two models using the (Weighted) Average Method. Subsequently, we subjected this updated model to testing using the earlier test datasets, observing the impact on accuracy of the model. The observations made are given in the following Table 1.

TABLE I
Accuracy Comparison

Model Type	Test dataset Type	Accuracy (in %)
Model_04	Dataset_04	98.988
Model_59	Dataset_59	98.478
Model_04	Dataset_04	0.0
Model_59	Dataset_59	0.0
Model_merged	Dataset_04	51.683
Model_merged	Dataset_59	62.292

[1]

We initially noticed high accuracy rates in the first two cases due to the similarity between training and testing datasets. In contrast, accuracy dropped to zero in the subsequent two cases, where the testing dataset differed from the training dataset, highlighting the absence of relevant training data. However, in the final two cases, when testing the model with merged weights and bias, we observed a substantial accuracy improvement, rising from 0% to approximately 50-60% for both datasets.

III. Next Course of Action

In the next half of the project, we would try to build upon the results we have till now and initially come up with a method which could quantify the diversity of a dataset without explicitly having access to the dataset. One common issue with this is the large size of dataset and excessive computation resources to be consumed in the process. To tackle this, once we achieve a method to measure diversity by correlating the original data with the encrypted one recieved by the third party, we would try to replicate the same with weights and biases of a neural network.

IV. Conclusion

In an era characterized by growing concerns about data privacy and the increasing demand for secure, collaborative data analysis, federated learning has emerged as a powerful and promising solution which we would like to achieve by our work and proposed next couse of action.

References

[1] Yang, K., Jiang, T., Shi, Y. and Ding, Z., 2020. "Federated learning via over-the-air computation". IEEE transactions on wireless communications, 19(3), pp.2022-2035.

[1]Model_04 and Model_59 denote models trained on numbers 0-4 and 5-9, respectively, with Model_merged indicating the merged model. Dataset_04 and Dataset_59 represent datasets for numbers 0-4 and 5-9, respectively.