Towards Robust Federated Learning using Knowledge Distillation Techniques

Arindam Jain
School of Computing & Augmented Intelligence
Arizona State University
ajain243@asu.edu

Kiran Sthanusubramonian
School of Computing & Augmented Intelligence
Arizona State University
ksthanus@asu.edu

I. Introduction

With the onset of improved privacy standards, edge computing capabilities, and large-scale machine learning requirements, Federated Learning has emerged as a privacy-preserving training paradigm for localized devices without data-sharing and aggregation requirements. With privacy-preserving benefits, users of these localized devices (e.g., mobile phones) also benefit from lower latencies in terms of required responses. Potential Federated Learning applications include improved mobile computing [1], healthcare [2], and autonomous vehicles.

In this project, we explore the use of Knowledge Distillation (KD) to create highly accurate and robust Federated Learning paradigms. Knowledge Distillation is conceptualized as a model compression technique in which large models with complex architectures are used to train a single smaller model that can run on devices with lesser computational capabilities while still achieving comparable performance levels. The most common Knowledge Distillation architecture is the Teacher-Student architecture (Fig.1).

The rest of this proposal is arranged as follows: Section II will detail the findings of the Literature Survey (including potential shortcomings), Section III will highlight the Problem Statement, Section IV will discuss the Methodology (including execution plan, datasets to be used, and the metrics to evaluate & validate the methodology), Section V will summarize the objectives and learning outcomes from this project.

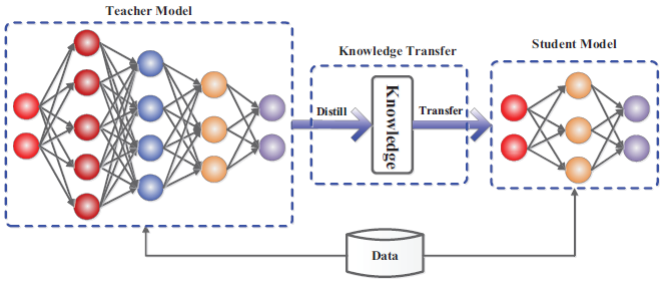

Fig. 1. Teacher-Student Architecture for Knowledge Distillation [3]

II. Literature Survey

A. Knowledge Distillation Techniques for Federated Learning

One of the popular KD Algorithms for Heterogenous Federated Learning is proposed as FedMD (Li & Wang) [4]. Here, the primary assumption is that our local nodes have the computation capability of defining and training their NN models specific to their local datasets - this helps solve the statistical heterogeneity of data across the nodes. Figure 2 clarifies the standing of FedMD with respect to other popular Federated Learning paradigms.

FedMD focuses on utilizing the core idea of KD, i.e., using the Prediction Logits output from local Neural Network (NN) models averaged out as Knowledge to reach a better global consensus. FedMD also employs Knowledge Transfer (using a single giant global dataset) to overcome the issue of small local datasets. However, this paradigm is prone to Byzantine Faults and potential corruption of the global consensus due to adversarial attacks on one or more of the local nodes. Furthermore, the communication efficiency of the overall network architecture reduces as the number of nodes increases.

Related work to Robust Federated Learning is done in [5], where several Client (Teacher) models are fused into a single Server (Student) model, which is further trained using unlabeled datasets to improve robustness. However, this does not defend against potential adversaries (a similar drawback to the algorithm presented in FedMD).

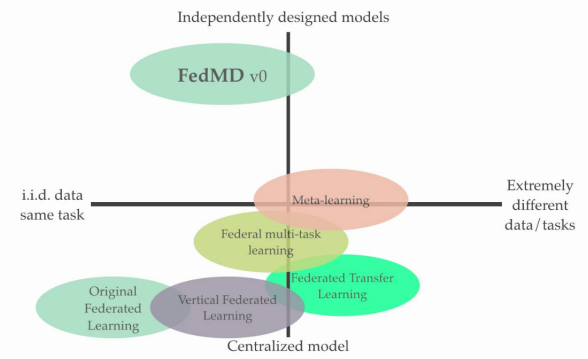

Fig. 2. Overview of Federated Learning Techniques

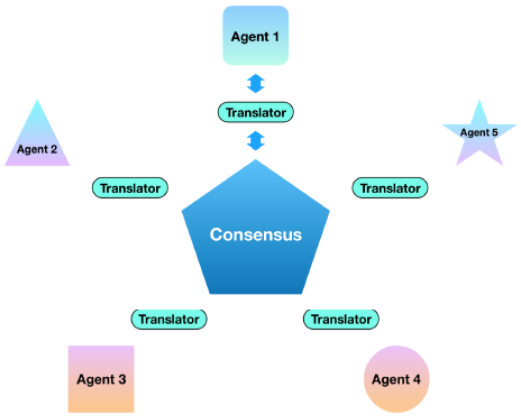

Fig. 3. Heterogenous Federated Learning [4]. Here, the Translator is implemented using Knowledge Distillation in the FedMD Algorithm

B. Improving Fairness and Robustness of Federated Learning Architectures

To improve Federated Learning architectures' fairness, accuracy, and robustness to potential Byzantine faults and adversarial attacks, we take inspiration from the algorithm proposed in Ditto (Tian Li et al) [6]. Here, personalization is used as a technique to balance robustness and fairness requirements. Personalization is mathematically conceptualized through regularization during local model updation with the global consensus calculated. The regularized parameter is tuned with respect to several factors, such as local dataset sample size, number of nodes affected by adversarial attacks, etc.

This paper also introduces essential robustness and fairness metrics, a crucial step toward defining evaluation parameters for Federated Learning research in real-world scenarios.

III. Problem Statement

Our problem statement is to increase the robustness of Knowledge Distillation techniques for Federated Learning. We intend to combine the idealogy presented in the FedMD algorithm with the Ditto Personalization Algorithm. We will define this combined implementation as a baseline for further investigating the effect of different Knowledge Distillation techniques, such as Ensemble, Cross-Modal & Data-Free Distillation, to design robust, accurate & communication-effective Federated Learning paradigms.

IV. Methodology

A. Execution Plan

Our plan of action can be broken down into the following stages:

1) Gain a Complete Theoretical & Practical Understanding of Knowledge Distillation - specific to the Multi-Teacher to Student Architecture.
2) Gain a Complete Theoretical & Practical Understanding of general Federated Learning Paradigms such as FedAvg [7].
3) Implement the FedMD Algorithm for applying Knowledge Distillation to Federated Learning settings.
4) Integrate the Ditto Algorithm step for Personalized Regularization to the created FedMD Algorithm.
5) Simulate Adversarial attacks on a randomized node selection to verify the overall robustness (test accuracy) & fairness of the network. Code & visualize the evaluation metrics. FedMD + Ditto will be defined as a baseline implementation for other experiments.
6) Integrate the Ditto Algorithm Personalization technique to any one of the other Knowledge Distillation techniques - Ensemble Distillation Algorithm, Cross-Modal, or Data-Free Distillation.

B. Datasets

We strictly define our problem space to Image Classification tasks. Implementing the FedMD algorithm requires a general public dataset & local private datasets.

In the first instance, a subset of the FEMNIST serves as the private data while the MNIST serves as the public data. We consider the non-.i.i.d. case and the i.i.d. scenario, where each private dataset is chosen randomly from FEMNIST.

In the second instance, the private dataset is a subset of the CIFAR100, which comprises 100 subclasses and 20 superclasses, and the public dataset is the CIFAR10. For example, the big carnivores category includes the bear, leopard, lion, tiger, and wolf. Each node's task in the i.i.d. scenario is to classify the test pictures correctly. It is more difficult to categorize generic test data in the non-i.i.d. scenario since each participant only has data from one subclass per superclass during training. For instance, it is predicted that a person who has only seen wolves throughout training will properly identify lions as huge carnivores. As a result, it must rely on the information shared by other nodes

C. Evaluation Metrics

We will primarily evaluate the Fairness & Robustness of the complete network and compare it with the results produced using the original FedMD algorithm & the original Ditto algorithm. The evaluation metrics used are based on the same used in the Ditto paper [6], as explained below.

The algorithm's robustness directly corresponds to the test accuracy in various adversarial settings. We will evaluate robustness for different percentages of adversarially-affected (malicious) nodes.

The algorithm's fairness corresponds to the variance of test accuracy for multiple iterations of simulations under the same setting. The minimization of variance corresponds to improved Fairness of the algorithm.

V. Objectives & Learning Outcomes

The primary objective of this project is to acquire foundational knowledge on Federated Learning, which is a massive breakthrough in deploying large-scale machine learning models in the real world while still preserving user data privacy.

With our work in this project, we wish to establish a baseline for robust & fair distillation techniques for improved personalization for heterogeneous devices and resilience to common adversarial attacks that can affect the whole system. Further developments can include:

1) Extending robustness & fairness to different distillation techniques for Federated Learning.
2) Further improving Privacy settings of local datasets using Differential Privacy.

REFERENCES

[1] W. Y. B. Lim et al., "Federated Learning in Mobile Edge Networks: A Comprehensive Survey," in IEEE Communications Surveys & Tutorials, vol. 22, no. 3, pp. 2031-2063, third quarter 2020.

[2] Xu, J., Glicksberg, B.S., Su, C. et al. "Federated Learning for Healthcare Informatics," J Healthc Inform Res 5, 1–19 (2021).

[3] Gou, J., Yu, B., Maybank, S.J. et al. "Knowledge Distillation: A Survey." Int J Comput Vis 129, 1789–1819 (2021).

[4] Li, Daliang and Junpu Wang. "FedMD: Heterogenous Federated Learning via Model Distillation." ArXiv abs/1910.03581 (2019)

[5] Tao Lin, Lingjing Kong, Sebastian U. Stich, and Martin Jaggi. 2020. "Ensemble distillation for robust model fusion in federated learning." In Proceedings of the 34th International Conference on Neural Information Processing Systems (NIPS'20). Curran Associates Inc., Red Hook, NY, USA, Article 198, 2351–2363.

[6] Tian Li, Shengyuan Hu, Ahmad Beirami & Virginia Smith. (2021). "Ditto: Fair and Robust Federated Learning Through Personalization."

[7] McMahan, H.B., Moore, E., Ramage, D., Hampson, S., & Arcas, B.A. (2017). "Communication-Efficient Learning of Deep Networks from Decentralized Data." AISTATS.