Federated Learning with Tsetlin Machine

Shannon How Shi Qi*, Jagmohan Chauhan*, Geoff V Merrett†, Jonathan Hare*
School of Electronics and Computer Science, University of Southampton, Southampton, UK
{s.how,j.chauhan,jsh}@soton.ac.uk*, gvm@ecs.soton.ac.uk†

I. Introduction

Federated Learning (FL) has gained significant traction in recent years across diverse domains like Automatic Speech Recognition (ASR) and Image Recognition [1]. FL addresses the critical issue of privacy, particularly concerning user data like images and personal information. FL tackles this challenge by allowing data to remain decentralized on client devices rather than being centralized on a server for training.

With the increasing number of network-connected devices, including smartphones and Internet-of-Things (IoT) devices, there is a growing need for FL models that are not only communication-efficient but also capable of accommodating edge devices' constraints. Additionally, such models must prioritize data privacy while ensuring robust learning outcomes.

In FL, Deep Neural Networks (DNN) are typically employed, requiring intensive arithmetic computations for gradient descent. This involves multiple iterations of learning on individual client devices, followed by the transmission of DNN parameters across the network for aggregation on a central server [2]. Given the resource-intensive nature of DNNs, existing FL methodologies may struggle to align with the vision of deploying FL on edge devices. Compounded by challenges such as high communication overheads, data heterogeneity , and scalability, FL at the edge remains challenging [1].

II. FedTM: Federated Learning with Tsetlin Machine

Motivated by the logic-based formulation and bit-based representation of the Tsetlin Machine (TM) [3], we propose FedTM, a novel FL framework that leverages the TM for memory and communication-efficient FL. Unlike traditional neural networks, where the server aggregates floating-point parameters using FedAvg [2], FedTM requires a two-step aggregation approach to independently aggregate clause weights and states. Our two-step aggregation approach with TopK and AverageCW address challenges such as aggregating bit-based representations, handling data heterogeneity, and accommodating varying client participation ratios.

III. Results

We compared the performance of FedTM with baseline models with FedAvg that uses the Convolutional Neural Network (FA (CNN)) and Binary Neural Networks (BiFL-BiML) as the base model. We simulate experiments with 100 clients and vary data heterogeneity on three commonly used image datasets for FL, MNIST, FashionMNIST and EMNIST. Our results demonstrate that FedTM outperforms BiFL-BiML in every FL setting while providing $1.37 - 7.6\times$ reduction in communication costs and $2.93 - 7.2\times$ reduction in runtime memory on the evaluated datasets. Compared to FA (CNN), on average, FedTM provides a substantial reduction in communication costs by $30.5\times$ and $36.6\times$ reduction in storage memory footprint. The base model used in FedTM is also considerably smaller than the neural network models with reductions of $8.24\times$, $35.7\times$, and $6.6\times$ across the evaluated datasets.

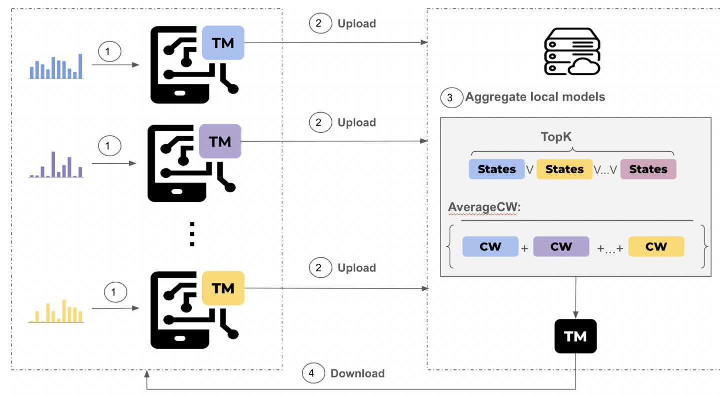

Fig. 1: Overview of FedTM

IV. Conclusion

Through extensive experiments, we demonstrated that FedTM achieves a compelling balance of efficiency and performance compared to the baseline models, and outperform BiFL-BiML while reducing average communication costs across our evaluated datasets by $3.57\times$. While our experiments showed promising results, particularly with the image datasets, we plan to extend FedTM's capabilities to constrained devices and to other domains such as audio and Natural Language Processing (NLP) to prove its generality. Our work presents an exciting and promising avenue to explore the potential of TM in the context of FL, opening up new possibilities for efficient FL on edge devices.

References

[1] A. Imteaj, U. Thakker, S. Wang, J. Li, and M. H. Amini, "A Survey on Federated Learning for Resource-Constrained IoT Devices," *IEEE IoT-J*, vol. 9, no. 1, pp. 1–24, 2022.

[2] H. B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-Efficient Learning of Deep Networks from Decentralized Data," in *AISTATS*, 2016.

[3] O.-C. Granmo, "The Tsetlin Machine - A Game Theoretic Bandit Driven Approach to Optimal Pattern Recognition with Propositional Logic," 2021.