8-10-2022

A Framework for Model-Centric Cross-Silo Horizontal Federated Machine Learning

Haytham M. Mohamed
Dakota State University, hmmohamed@pluto.dsu.edu

Omar El-Gayar
Dakota State University, omar.el-gayar@dsu.edu

A Framework for Model-Centric Cross-Silo Horizontal Federated Machine Learning

TREO Talk Paper

Haytham Mohamed
Dakota State University
hmmohamed@pluto.dsu.edu

Omar El-Gayar
Dakota State University
omar.el-gayar@dsu.edu

Abstract

Typically, in machine learning applications, the data is combined and centralized in one place alongside a model to train. This imposes the concern of exposing sensitive data and security risks (Mammen, 2021). Federated learning offers a better option to mitigate such challenges. In federated learning, data is distributed across different locations where a machine learning model is trained locally and only the results (not the data) are sent back to a centralized server to aggregate and enhance the trained model. The way in which data is split across the different locations matters in terms of how federated learning is implemented and the practical and technical challenges. The data sets in horizontal (or homogenous) federated learning preserve the same feature space but have different examples (Yang et al., 2019). Furthermore, in cross-silo federated learning, organizations work together to train a global model with their own local data (Figure 1).

Federated learning has two broad challenges: training challenges and security challenges. One of the training-related challenges is the communication overhead during multiple training iterations. This research project aims at tackling this challenge by exploring the design of a communication efficient framework to facilitate training of centralized models using cross-silo federated learning of horizontally distributed data. The aim is to devise a method by which to reduce communication latency while consuming lower computation resources. Our initial investigation led to the identification of potential mechanisms to efficiently connect distributed data locations; transfer local trained models between a centralized location and the distributed locations; and centrally process and aggregate the local trained results to enhance a centralized global model. The proposed approach (Figure 2) leverages recent advancements in the reactive stream processing mechanism and the RSocket protocol to build the framework. RSocket protocol enables flows of byte streams over a TCP transport layer and can establish reliable communications between the distributed components. Additionally, the design uses a cluster of RSocket broker instances to connect remote data locations. We demonstrate the efficacy of the framework by training federally a supervised machine learning model using distributed and homogenous data samples across different locations. We illustrate the efficiency of the framework in terms of computing resource consumption by benchmarking against an available federal learning framework.

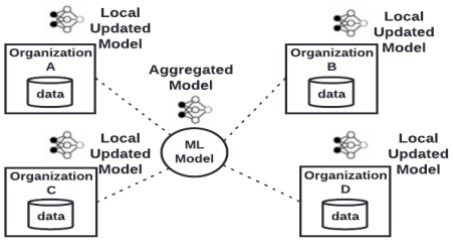

Figure 1 Federated Learning Cross-Silo

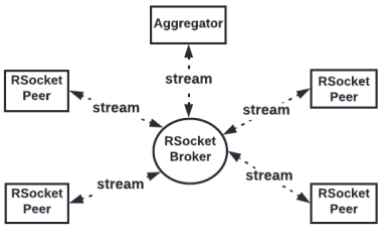

Figure 2 Framework Design Options

References

Mammen, P. M. (2021). *Federated Learning: Opportunities and Challenges*. http://arxiv.org/abs/2101.05428

Yang, Q., Liu, Y., Chen, T., & Tong, Y. (2019). Federated machine learning: Concept and applications. ACM Transactions on Intelligent Systems and Technology, 10(2), 1–19. https://doi.org/10.1145/3298981