SC23 Network Research Exhibition: Demonstration Preliminary Abstract

Distributed Machine Learning over Heterogeneous Edge Computing Resources

Kohei Ichikawa, Nara Institute of Science and Technology, ichikawa@is.naist.jp

Abstract

Federated Learning is a distributed machine learning framework that enables collaborative learning without directly sharing personal or private data. In Federated Learning, each individual's data is used to train models locally on their own computing resources, and only the updated model parameters are aggregated by a central server to build a global model. This way, the original personal data is difficult to extract from the these locally trained models, making Federated Learning a privacy-preserving technique for learning over sensitive datasets. Federated Learning is expected to be applied to edge computing environments that collect and utilize personal data from various sources.

However, the original Federated Learning framework does not consider the heterogeneity of edge computing resources, such as latency, network bandwidth, and computational power. Therefore, if the nodes participating in Federated Learning are distributed over a wide area and have different resource capabilities and/or network connectivity, the training process will not be efficient. This is because the training is synchronous, meaning that it waits for the results from all the nodes before proceeding to the next round.

To address this challenge, we propose hierarchical federated learning, in which the edge computing resources are grouped into clusters based on their locations or other criteria, and the model parameters are aggregated hierarchically from the edge resources and then to the central server. This hierarchical structure reduces the average latency of the entire Federated Learning environment. The hierarchical federated learning is a novel and promising extension of FL that can achieve more efficient and effective learning in a heterogeneous edge computing environment.

In this demonstration, we deploy edge computing resources at several sites in the U.S., Japan, and other Asian regions, and interconnect them by EdgeVPN.io, a P2P-based overlay network software creating scalable virtual Ethernet networks across distributed edge nodes. On top of this overlay network, we perform hierarchical federated learning with different levels of aggregation. We will demonstrate how hierarchical federated learning can achieve higher efficiency and accuracy in a distributed and heterogeneous environment.

Goals

1. Deploy edge computing resources at several sites in the U.S. and Japan, and interconnect them by EdgeVPN.io, and then perform hierarchical federated learning on that environment.
2. Evaluate the performance of the hierarchical federated learning in terms of communication efficiency, learning accuracy, and scalability, and compare it with the original federated learning framework. We also vary the levels of aggregation at different hierarchy levels to show their impact on the performance.
3. Expand the environment to include other Asian regions, and demonstrate how hierarchical federated learning can cope with the increased heterogeneity and latency of the edge computing resources.

Resources

As part of our demonstration, we will deploy a small edge computing resource at the SC23 venue, which will act as one of the sites in our distributed environment. This resource will need an uplink network bandwidth of about 1 Gbps to communicate with the other edge nodes and the cloud server.

Involved Parties

- Kohei Ichikawa, Nara Institute of Science and Technology, ichikawa@is.naist.jp
- Kundjanasith Thonglek, Osaka University, kundjanasith.thonglek.cmc@osaka-u.ac.jp
- Keichi Takahashi, Tohoku University, keichi@tohoku.ac.jp
- Shinji Shimojo, Aomori University, shinji.shimojo@aomori-u.ac.jp
- Renato Figueiredo, Florida University, rjofig@acis.ufl.edu
- Subratie,Kensworth, Florida University, kcratie@ufl.edu
- Vahid Daneshmand, Florida University, vdaneshmand@acis.ufl.edu