Enhancing Blockchain Based Federated Learning

Federated learning (FL) allows to train models from multiple distributed data pools without the need to ever collect the data at a central location. This can be useful when there are mechanisms in place that make sharing or moving data impractical, e.g. data privacy protection. Using FL enables the data pool owners to contribute to (and in some cases profit from) a shared model trained in conjunction with other owners (that hold similar, but different data) in an alliance without the necessity to share the raw data with anyone. FL can be realized in many different ways. One prominent way is to train local models on side of the data pool storage locations, send the resulting model parameters to a central location and merge them together [1]. This enables the creation of a model on multiple data pools without ever exposing the data to other data owners. However, this introduces additional problems regarding privacy, trust and stability.

As soon as an alliance member is not training the model completely on their own, but rather incorporate other's training results into a shared model, the members need to **trust** that other's training results are correct and are contributed with good intentions. A prominent way to establish trust between multiple parties without the use of a trusted authority is the **blockchain [2]**. It builds an open record of every transaction and can be verified and audited by anyone in the network. On top of that there can be **smart contracts [4]**, a way of writing and executing code on the blockchain.

Recently we have developed a platform for federated machine learning using the Ethereum [3] blockchain. Following is the Github repository link of the project:

https://github.com/FMorsbach/DecFL

This framework was developed by Mr. Felix Morsbach during his master thesis. Felix has recently defended his thesis and the report will soon be available online. Within the scope of the project course, we would like to extend the functionality of the framework in the following two directions:

1 – **(Compulsory)** In the current implementation, consensus to make commits in the chain is based on contributions from all the alliance members. This approach increases the trust but the cost of a single commit is high. The task is to develop a smart policy based on smaller group of alliance members that are allowed to verify the contributions and make commits on the chain.

2 – **(Optional)** The framework adheres a simple federated averaging scheme that requires all alliance members to send model weights from their sides. This approach has high overhead. An alternate approach is to use incremental federated averaging scheme. The scheme is fairly straight forward however, to integrate it within the blockchain settings will be a challenge.

Project Supervisors

Dr. Salman Toor, salman.toor@it.uu.se
Mr. Felix Morsbach , Felixjohannes.Morsbach.4921@student.uu.se

References

1 - McMahan, H. Brendan, Eider Moore, Daniel Ramage and Blaise Agüera y Arcas. “Federated Learning of Deep Networks using Model Averaging.” ArXiv abs/1602.05629 (2016): n. pag.

2 – Blockchain Technology: https://blockgeeks.com/guides/what-is-blockchain-technology/

3 - Etherum White Paper: https://ethereum.org/en/whitepaper/

4 – Smart Contracts: https://cointelegraph.com/ethereum-for-beginners/what-are-smart-contracts-guide-for-beginners