Bachelor or Master Thesis

Determinants of Successful Federated Learning

Semantic Computing Group
Christoph Düsing
cduesing@techfak.uni-bielefeld.de

Federated Learning (FL) is a novel, privacy-preserving approach towards distributed machine learning. As such, it allows participating parties, so-called clients, to jointly train an arbitrary machine learning model without having to share data among each other. To do so, a central server orchestrates several rounds of training, each consisting of a local training performed by each client and a subsequent aggregation of all model updates. During several rounds of training, a global model is achieved that incorporates knowledge derived from all clients' data silos without directly accessing them.

In spite of its potential, the performance of FL is influenced by a variety of factors, including the data held by clients, their number, drop-out rates, latency, and many others. Accordingly, clients are unable to know to what extent they will benefit from participation prior to committing to it. Ultimately, this might negatively affect their willingness to participate.
To date, data imbalance has been shown to bear the potential to serve as determinant of a client's benefit from participation and contribution towards the federation. Other than that, little research has been conducted to identify determinants of successful FL, especially regarding dynamic client participation during training.

Therefore, this thesis aims to shed some light into potential determinants of successful FL. The thesis might either rely on a literature-based approach towards the topic or make use of existing FL implementations in order to empirically identify them.
The offered thesis does not require prior knowledge of FL. However, familiarity with traditional machine learning, data preprocessing, and python are recommended.

Related literature

1. McMahan, Brendan, et al. "Communication-efficient learning of deep networks from decentralized data." Artificial intelligence and statistics. PMLR, 2017.

2. Düsing, C., & Cimiano, P. (2022, December). On the Trade-off Between Benefit and Contribution for Clients in Federated Learning in Healthcare. In 2022 21st IEEE International Conference on Machine Learning and Applications (ICMLA) (pp. 1672-1678). IEEE.

3. Düsing, C., & Cimiano, P. (2022, December). Towards predicting client benefit and contribution in federated learning from data imbalance. In Proceedings of the 3rd International Workshop on Distributed Machine Learning (pp. 23-29).

The Semantic Computing Group researches and develops methods that enable machines to acquire relevant knowledge as well as linguistic capabilities. Using methods from *natural language understanding* and *machine learning*, we are aiming at machines that are capable of knowledge acquisition by reading unstructured textual data. In particular, the group focuses on methods for information extraction, semantic parsing, ontology learning, sentiment analysis, entity linking, as well as question answering.

More information is available at: http://sc.cit-ec.uni-bielefeld.de.

Interested? @mail to `cduesing@techfak.uni-bielefeld.de`