Tech Brief: Apply AON-PRISMA in Vertical Federated Learning

AON-PRISMA, Sept 2023

Vertical Federated Learning (VFL) is a federated learning setting where multiple parties share overlapping data samples while possessing distinct sets of features. The goal is to jointly train a machine learning model by exploiting all features collected by participating parties without exposing private data. For example, an airline company could collaborate with a car rental agency to build a recommendation system for their users.

A general training protocol for VFL consists of two steps: 1) entity alignment; 2) ML model training. In the first step, all parties cooperatively align the common record intersection. This involves finding which data records from different parties are linked to the same entity, and such information is used in the subsequent model training phase. Additionally, we need to make sure we do not reveal any information outside the common entities. One could leverage Private Set Intersection (PSI) for privacy-preserving entity alignment. However, PSI requires that the data records of the same entity from different parties are associated with a common unique identifier, which is used for exact matching during the PSI process. This may be difficult to achieve in real-world scenarios where there are data errors or schema mismatches.

In contrast, AON-PRISMA introduces a more flexible approach to perform private similarity matching among parties. Instead of matching records based on exact identifiers, AON-PRISMA links the same entities based on the similarities in common features. This new approach avoids the need for parties to perform extensive joint data preprocessing. For example, individuals can be aligned based on demographic information such as name, address, DOB, etc., which may contain syntactic or semantic differences. AON-PRISMA achieves this efficiently while preserving data privacy - no information is disclosed other than the matched record indices. We provide an end-to-end demonstration of VFL using AON-PRISMA in our GitHub code repository.

AON-PRISMA is a technology developed by researchers at the University of Waterloo and the National Research Council Canada. To find out more about AON-PRISMA, please visit https://aon-prisma.dev/ or email info@aon-prisma.dev