Collaborative Privacy-Preserving Training of Decision Trees

Y. El-Zein, R. Delgado-Gonzalo, K. Huguenin •

As data generation becomes more ubiquitous, datasets are increasingly spread across several entities. Exploiting such distributed datasets simultaneously, by training machine learning models on them, can be beneficial to many fields, but also raises serious privacy concerns. In this work, a scalable protocol is designed, implemented, and evaluated, for training decision tree models collaboratively across mutually distrustful data providers, while ensuring privacy of each of the stakeholders' data.

Training accurate and robust machine learning (ML) models requires access to large datasets. Often, subsets of records of these datasets are held by different data providers (DPs). Due to privacy concerns and regulations, these records cannot always be shared among DPs, nor centralized in one common location, hindering the training of ML models on them. Privacy-preserving ML, more particularly federated learning, has emerged as a solution to this problem. However, federated learning solutions are mainly geared towards the training of neural network models. In this work, the focus is turned to privacy-preserving training of decision tree models. Contrary to neural networks, decision trees have the advantage of being interpretable white-box models. In domains such as medicine or finance, where every decision must be supported by a clear rationale, model interpretability is of utmost importance. To this end, a scalable protocol is designed, implemented, and evaluated for training decision tree models in a collaborative manner across different DPs, while maintaining the privacy of their data records.

In our implementation, *N* data providers are considered, each holding a subset of a horizontally partitioned dataset, whereby each DP holds a different set of records (of comparable size), with the same attributes as the others. Their goal is to collaboratively train a decision tree model that benefits from their joint data records, without disclosing the latter to each other. The training protocol also calls for a central coordinator (CC), who orchestrates the training process without contributing any data to training the model. In terms of threat model, we consider all involved entities to be honest-but-curious. An overview of the considered system model is given in Figure 1.

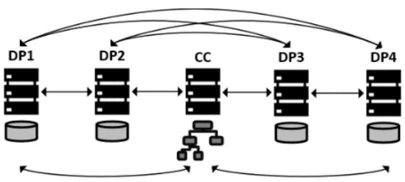

Figure 1: Overview of the different entities of the considered system model, and the different communication channels between them.

The protocol is designed to fit the following requirements. (1) The raw data of any DP is never shared with nor leaked to any other entity. (2) None of the exchanged messages between entities should leak information about the data distribution of a data provider's private dataset. The protocol is composed of two phases: the pre-processing phase and the training phase.

The goal of the pre-processing phase is to privately determine the optimal depth of the tree. Briefly, each DP computes, on their local private dataset, the optimal tree depth and engages with the other DPs in an additive secret sharing subprotocol to compute the optimal tree depth. Following the pre-processing phase, the goal of the training phase is to select the best attribute to further split a given node in the decision tree. Each DP computes locally the information gain associated with the different attributes in their dataset and selects the best attribute for splitting the node accordingly. If the attribute is continuous, a numerical threshold for splitting is also computed by each DP. Using a proposed adapted variation of additive secret sharing, the attribute selected by most DPs is obtained and used to further split the node, without revealing the preferred attribute of individual DPs.

The protocol was implemented in Python and evaluated on 3 public datasets from the UCI (University of California, Irvine) machine learning repository [1]. It achieves privacy with a modest loss in accuracy compared to the privacy-breaching model, trained on all the data records centralized in one location, and consistently produces a decision tree model with better accuracy than the local models, trained separately by each DP. Figure 2 summarizes the model accuracy results.

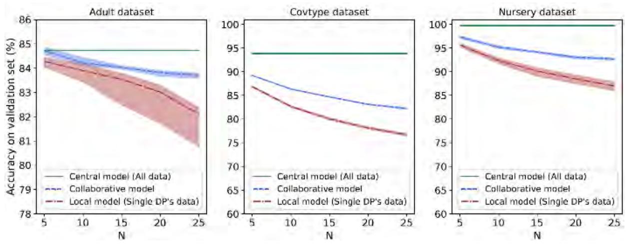

Figure 2: Accuracy of model trained on three datasets using the proposed protocol, compared to centralized and local models' accuracy.

Moreover, compared to existing solutions, the protocol is efficient in terms of communication and computation costs. For instance, training on one dataset, composed of 30'162 data records and 14 attributes, across 4 DPs each running on an AMD 2950X CPU at 3.5 GHz and using 128 GB of RAM, generates only 331.44 MB of exchanged messages, and requires 141.39 seconds of training time, including both CPU and communication time. This is an estimated 63× speedup compared to the current state of the art for collaborative training of decision trees [2].

The presented work showcases an alternative to centralized training, that is privacy-aware, in the context of white-box models. The proposed architecture is scalable and easy to deploy, which can accelerate the introduction of ML into heavily regulated sectors (e.g., healthcare), where the data is considered extremely sensitive due to its nature.

• University of Lausanne, Switzerland

[1] D. Dua, C. Graff, "UCI Machine Learning Repository", University of California, School of Information and Computer Science (2017) [http://archive.ics.uci.edu/ml]

[2] M. Abspoel, D. Escudero, N. Volgushev, "Secure training of decision trees with continuous attributes", Proceedings on Privacy Enhancing Technologies (2021).