Federated Extremely Random Forest with Local Differential Privacy

陈明鑫 †2019200562

October 2019

1 Background

The burgeoning demand of multiple big data analytics and the strict constraints of data privacy and access have made it risky and sometimes illegal for a stand-alone user or institution to exposure it's owned data for a joint model. To tackle this dilemma, federated learning[2] was proposed to train a collaborative model with privacy-preserving algorithms(e.g., DP, MPC and CE) while keeping the private data locally.

2 Purpose

The aim of this paper is to propose a privacy-preserving extremely random forest model with local differential privacy and make a balance between the utility and privacy.

3 Method

Based on the work of Geurts[1], we extend the stand-alone Extra Trees to a distributed setting. In distributed Extra Trees, the best split point on a specific feature of all data nodes will be determined by the comparison of the perturbed information collected from the correspond data provider. And the comparison process and perturbing process will be explained detailedly in the following formal papers.

Reference

[1] Pierre Geurts, Damien Ernst, and Louis Wehenkel. Extremely randomized trees. *Machine learning*, 63(1):3–42, 2006.

[2] H Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, et al. Communication-efficient learning of deep networks from decentralized data. *arXiv preprint arXiv:1602.05629*, 2016.