

Privacy-preserving Blockchain Based Global Data Sharing for Federated Learning with Non-IID Data

Zhuotao Lian*

Qingkui Zeng[†]

Chunhua Su*

Keywords: Blockchain, data sharing, federated learning, non-IID data, privacy-preserving.

Abstract

Federated learning, as a privacy-enhanced distributed machine learning method, is a very hot research topic recently. It solves the direct problem of private data sharing and global model training because each participant uses their own data for training locally, and only shares the training results (such as model weights, gradients, etc.). But a practical problem it faces is the non-independent and identical distribution of data, which means that the local data of each participant is highly inconsistent, both in terms of quantity and distribution. Moreover, there is a lack of research related to the efficiency and privacy issues in the pre-training process. Therefore, in this paper, we propose a novel solution that uses blockchain technology to realize small-scale global data sharing which is used for pretraining. Simulation experiments verify that our method not only guarantees data security but also greatly improves performance in terms of training speed and accuracy.

* Zhuotao Lian and Chunhua Su are with the Division of Computer Science, University of Aizu, Aizuwakamatsu, Japan
zhuotaolian@ieee.org
chsu@u-aizu.ac.jp

[†] Qingkui Zeng is with the School of Electronics and Information Engineering, Nanjing University of Information Science & Technology, Nanjing, China
zenghuh1996@gmail.com