ClusterPredict: Enhancing Federated Clustering by Combining Global and Local Parameters

MingliangNi, Chaochao Sun

School of Computer Science and Technology, Shanghai University of Electric Power,China

nmlnml@mail.shiep.edu.cn,pop-scc@163.com

***Abstract*—This paper proposes a novel federated learning algorithm and framework for multicentre clustering, which addresses the challenges of non-IID data distribution. The distributed machine learning technique of federated learning, which is currently being employed in an expanding array of domains, represents a novel approach to data management and analysis. The technique permits the training of models on local devices, with only the parameters of the model being updated, rather than the raw data, being transmitted to and from a server. This aggregation of updates allows for the improvement of model training efficiency while ensuring the protection of user privacy. Although federated learning provides an efficient distributed machine learning framework, traditional federated learning methods are prone to inefficiencies and model performance degradation when dealing with non-iid data. In the context of data science, non-independent and identically distributed (non-iid) data refers to a situation where the data samples are not independent from each other and their distributions are not identical. A variety of personalised federated learning methods have been developed to address the challenges posed by non-IID data. Among these, clustered federated learning algorithms represent a particularly robust strategy. However, general clustering federation learning algorithms are unable to effectively address the cases of new clients joining the training set and some clients participating in the training. Consequently, based on clustering federation learning, we also consider the above cases and propose a new multi-centre federation learning algorithm using the FedPredict mechanism, which enables the new mechanism to effectively address the cases of some clients participating in the training set as well as new clients joining the training set. The efficacy of the algorithm is validated through experimentation with disparate datasets.**

***Keyword*—federated learning, non-iid data, distributed, personalization, clustering**

Mingliang Ni is a master's student currently studying for his master's degree in the School of Computer Science and Technology at Shanghai University of Electric Power. His research interests include federated learning algorithms and privacy protection. Focusing on communication efficiency improvement by federated learning.

Chaochao Sun received his Ph.D. degree from the School of Computer Science and Technology, Fudan University, in 2020, and he is currently a lecturer at the School of Computer Science and Technology, Shanghai University of Electric Power. His research interests include spectrum sensing, cognitive wireless networks, and federated learning.