基于非独立同分布和长尾数据的双解耦联邦学习

王朝晖，李红娇，李晋国，胡仁豪，王宝金
上海电力大学计算机科学与技术学院，中国上海市，201306

摘要：联邦学习(FL)作为一种最前沿的分布式机器学习训练范式，旨在通过协作训练客户端模型生成全局模型，且不泄露本地私有数据。然而客户端数据同时呈现出非独立同分布(non-IID)和长尾分布时会严重影响全局模型准确率，从而对联邦学习造成根本性挑战。针对非独立同分布和长尾数据，提出一种通过模型和逻辑校准的双解耦联邦学习(FedDDC)框架。该模型具有3个特点。首先，解耦全局模型为特征提取器和分类器，从而微调受异构数据影响的组件。针对有偏特征提取器，提出客户端置信度重加权算法辅助校准，该算法为每个客户端分配最优权重。针对有偏分类器，采用分类器再平衡方法进行微调。其次，校准并集成经过客户端重加权和分类器重平衡的逻辑，从而得到无偏逻辑。最后，首次在非独立同分布和长尾分布下的联邦学习中使用解耦知识蒸馏，通过提取无偏模型知识提高全局模型准确率。大量实验表明，在非独立同分布和长尾数据上FedDDC优于最先进的联邦学习算法。

关键词：联邦学习；非独立同分布；长尾数据；解耦学习；知识蒸馏
https://doi.org/10.1631/FITEE.2300284