Distributed sequential federated learning

Z. F. Wang[*], X. Y. Zhang[†]   and Y-c I. Chang[‡]

Distributed sequential federated learning

Z. F. Wang[*], X. Y. Zhang[†]

International Institute of Finance, The School of Management,
University of Science and Technology of China, China

and

Y-c I. Chang[‡]
Institute of Statistical Science,
Academia Sinica, Taipei 11529, Taiwan

October 2, 2023

Abstract

The analysis of data stored in multiple sites has become more popular, raising new concerns about the security of data storage and communication. Federated learning, which does not require centralizing data, is a common approach to preventing heavy data transportation, securing valued data, and protecting personal information protection. Therefore, determining how to aggregate the information obtained from the analysis of data in separate local sites has become an important statistical issue. The commonly used averaging methods may not be suitable due to data nonhomogeneity and incomparable results among individual sites, and applying them may result in the loss of information obtained from the individual analyses. Using a sequential method in federated learning with distributed computing can facilitate the integration and accelerate the analysis process. We develop a data-driven method for efficiently and effectively aggregating valued information by analyzing local data without encountering potential issues such as information security and heavy transportation due to data communication. In addition, the proposed method can preserve the properties of classical sequential adaptive design, such as data-driven sample size and estimation precision when applied to generalized linear models. We use numerical studies of simulated data and an application to COVID-19 data collected from 32 hospitals in Mexico, to illustrate the proposed method.

Keywords: Adaptive sampling; Data Communication; COVID-19; Random average; Sequential sampling

---

[*]zfw@ustc.edu.cn

[†]xinyu@amss.ac.cn

[‡]ycchang@sinica.edu.tw

1