Multi-Center Federated Learning to Cluster Clients with non-IID data

A thesis submitted in fulfilment of the requirements
for the degree of

Doctor of Philosophy
in
Analytics

by

Ming Xie

to

School of Computer Science
Faculty of Engineering and Information Technology
University of Technology Sydney
NSW - 2007, Australia

August 2020

CERTIFICATE OF ORIGINAL AUTHORSHIP

I, *Ming Xie* declare that this thesis is submitted in fulfilment of the requirements for the award of Doctor of Philosophy, in the *School of Computer Science*, *Faculty of Engg. & IT* at the University of Technology Sydney, Australia.

This thesis is wholly my own work unless otherwise referenced or acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

This document has not been submitted for qualifications at any other academic institution.

This research is supported by the Australian Government Research Training Program.

SIGNATURE: Production Note: – Signature removed prior to publication.

[Ming Xie]

DATE: 30th August, 2022

PLACE: Sydney, Australia

Abstract

Federated learning (FL) is a new machine learning paradigm to collaboratively learn an intelligent model across many clients without uploading local data to the server. Non-IID data across clients is a significant challenge for the FL system because its inherited distributed machine learning framework is designed for the scenario of IID data across clients. Clustered FL is a type of FL method to solve non-IID challenges using a client clustering method in the FL context. However, even adopts a client clustering FL method still facing minor problems such as unstable against client-wise outliers and the drop of model performance with model poisoning attack.

To face the aforementioned challenges, the main research objective of the thesis is to study that how to make FL effectively, seamlessly solved non-IID data across clients in horizontal clients partition settings.

The main research objective has been studied from four coherently linked perspectives: (i) how to make FL to address the non-IID distribution of data across different clients in a effective and scalable manner so that they can be applied to real world cases which consists of thousands of client and varies type of devices, (ii) how to make cluster FL methods more robust to client-wise outliers, (iii) how to make better balance between the performance of global models and the extent of personalisation of local models, (iv) how to make FL training more robust to model poisoning attack by density methods.

This thesis proposes a novel FL framework with robust clustering algorithm and secure the models to tackle client-wise outliers as well as model poisoning in the FL system. Specifically, we will develop a robust federated aggregation operator using a bootstrap median-of-means mechanism that can produce a higher breakdown point to tolerate a larger proportion of outliers. All work experiments on three benchmark datasets have demonstrated the effectiveness of the proposed method that outperforms other baseline methods in terms of evaluation criteria.

In short, we develop a original, effective clustered FL baseline algorithm which can improve FL performance in horizontal clients partition settings. We compared proposed work against several state-of-the-art FL algorithms using both synthetic and real-world data.

Dedication

dedicate my thesis to my mother and my wife for their love

ACKNOWLEDGMENTS

I, Ming Xie, acknowledge Dr. Chandranath Adak for providing this amazing thesis template.I am truly grateful to my supervisor, Prof. Guodong Long, for his great guidance through every facet of the research world. In the past few years, now only I acquired critical technical knowledge in Federated Learning research topic, but also how to clearly write and communicate those my ideas formally and accurately. In addition, I realize the importance of prepare presentations and engage audience plays a vital role in research. Likewise, interact with other academics and build research collaborations. All of this was crucial to my future work and life.

A very special gratitude goes to A/Prof. Lu Qin, who has provided many life changing ideas and advice not only on academic but also on personal life. Words cannot express my emotions in this regard.

Thanks to all those who shared endless discussion and conversation in UTS building 2. It was great spending time in the building and in the laboratory with all of you.

I am also grateful to all the following people, who have helped and supported me along the way: Prof Chengqi Zhang, Dr. Jing Jiang, Dr. Xueping Peng, Mr. Wensi Tang, Mr. Jie ma, Dr. Shen Tao, Dr. Xubo Wang, Dr. Wentao Li, Mr. Yang Wang.

I am indebted to Faulty of Engineering & IT of the University of Technology Sydney for the wonderful, inclusive and productive work environment, to LINKAGE scholarship for initially funding my PhD. Thank you all the staff, admins and proof readers of my Faulty, they are very professional. I am extremely grateful for living in Sydney, even though I moved out from time to time, it is a city full of culture diversity, beautiful beach and I love to work here.

Finally, I would like to thank my family, for all the obstacles they quietly removed so I could realize my dreams.

List of Publications

Related to the Thesis :

1. Ming Xie, Jie Ma, Guodong Long, "Personalized Federated Learning with Robust Clustering against Model Poisoning", Advanced Data Mining and Applications, 2022 [CORE rank B] (Accepted by 10 Aug 2022)

2. Ming Xie, Jie Ma, Guodong Long, Chengqi Zhang, "Robust Clustered Federated Learning", The Asia Pacific Web (APWeb), 1-15, 2022 [CORE rank B] (Accepted by 7 June 2022)

3. Ming Xie, Jing Jiang, Tao Shen, Yang Wang, Leah Gerrard, Allison Clarke, "A Green Pipeline for Out-of-Domain Public Sentiment Analysis", International Conference on Advanced Data Mining and Applications, 190-202, 2021, DOI: https://doi.org/10.1007/978-3-030-95405-5_14, [CORE rank B] (**Best Student Paper Award**)

4. Guodong Long*, Ming Xie*, Tao Shen, Tianyi Zhou, Xianzhi Wang, Jing Jiang, "Multi-center Federated Learning: Client Clustering for Better Personalization", World Wide Web Journal, June 2022, DOI: https://doi.org/10.1007/s11280-022-01046-x [**CORE rank A**]

   Notes: * indicates equal contributions.

Table of Contents

List of Figures

List of Tables