Decentralized Federated Learning: An Introduction and the Road Ahead

Reza M. Parizi
Kennesaw State University
rparizi1@kennesaw.edu

Ali Dehghantanha
University of Guelph
adehghan@uoguelph.ca

Qi Zhang
IBM Thomas J. Watson Research
q.zhang@ibm.com

Katrin Franke
Norwegian University of Science and Technology
katrin.franke@ntnu.no

Abstract

Machine learning (ML) has exhibited great potential to transform the computing field. Huge interest has developed over the past years in applying machine learning-assisted approaches in the Internet of Things, healthcare, transportation, and security space, to name a few. However, the assumption in many current solutions is that big training data is widely available and transferable to a centralized server without much considering data privacy concerns. A new framework for machine learning has emerged, referred to as Federated Learning (FL), that advocates the AI-on-edge principle. The main objective of federated learning is to provide privacy-by-design training with decentralized data among local machines at the edge layer. In federated learning, a central server just coordinates with local clients to aggregate the model's updates without requiring the actual data (i.e., zero-touch). However, given the fresh nature of the FL, it is important to keep improving and design innovative solutions to mitigate its shortcomings and identify its best applications. This is the focus of the 'Decentralized Federated Learning: Applications, Solutions, and Challenges' mini-track. In this introduction article, we will describe the topic and the accepted paper(s) contributed by researchers.

1. Introduction

The idea of decentralization received tremendous popularity with the introduction of cryptocurrency and its backbone technology, blockchain [1], [2], to disperse the control away from centralized parties. Soon after, decentralization found its way into the machine learning space in the world of data engineering. Federated learning (FL) [3], [4] has been the realization of decentralized machine learning in the Artificial Intelligence (AI) realm that builds upon decentralized data and training that brings learning to the edge layer of networks or simply on-device, i.e., where data is generated. This paradigm came to light mainly for two reasons [3]: (1) The unavailability of sufficient (big) data to reside centrally on the server-side (as opposed to traditional machine learning) for training models; and (2) Data privacy protection by preventing actual (local) data to be transferred from edge devices, i.e., clients, to the server owned by companies or organizations. FL enables AI benefits to the domains with sensitive data and heterogeneity in the loop. Preserving data privacy provides feasibility to leverage AI benefits enabled through machine learning models efficiently across multiple domains (e.g., [5]). FL can be used to test and train not only on smartphones and tablets but on all types of devices. For instance, FL makes it possible for autonomous vehicles to train on decentralized driver behavior across the globe, or hospitals to strengthen diagnostics without breaching the privacy of their patients. This new research area often referred to as a new dawn in AI, is however in infancy and the introduction of its enabling technology has arguably required more profound research into its confirmation, particularly with its use-cases, scalability, performance, and security aspects.

In order to advance the state of the research in this area and to realize extensive utilization of the FL paradigm and its mass adoption in practice, this mini-track promotes new research from both academia and industry, with a particular emphasis on open-source solutions, applications, and industrial-strength tools to pave the way for the future. In its first occurrence, we accepted one paper contributed by Robin Hirt; Akash Srivastava; Carlos Berg; Niklas Kühl titled Sequential Transfer Machine Learning in Networks: Measuring the Impact of Data and Neural Net Similarity on Transferability.

URI: https://hdl.handle.net/10125/71471
978-0-9981331-4-0

2. Concluding Remarks

With growing concerns around people's data privacy in intelligent digital platforms, the importance of decentralized machine learning would be beneficial to our digital society in years to come. Currently, federated learning has shown to be a viable solution to mitigate issues and risks in this space. However, this new technology is in its infancy and still has a long way to its maturity for mass adoption. As pointed by the authors in [3], federated learning has a set of challenges that need further research (readers are encouraged to refer to). From a close observation in this field, one of the major obstacles by researchers is the lack of scalable simulation tools/frameworks to support federated-based approaches' implementation for evaluation purposes.

The topics for future research that we may suggest include:

- Developing on-device FL techniques for vision, audio, speech, and natural language processing
- Developing scalable frameworks and APIs for the implementation of federated learning-based solutions
- Exploring blockchain integration for industrial-strength federated learning
- Exploring new federated-based solutions for improved security and privacy in 5G-enabled Internet of Things (IoT), Internet of Medical Things (IoMT), Drones, and Autonomous Vehicles
- Developing new aggregation and data selection techniques driving FL
- Creating FL-specific datasets for federated learning research

References

[1] P. J. Taylor, T. Dargahi, A. Dehghantanha, R. M. Parizi, K-K. R. Choo, "A systematic literature review of blockchain cyber security," *Digital Communications and Networks*, vol. 6, pp. 147-156, 2020.

[2] A. Yazdinejad, G. Srivastava, R. M. Parizi, A. Dehghantanha, K. -K. R. Choo and M. Aledhari, "Decentralized Authentication of Distributed Patients in Hospital Networks Using Blockchain," *IEEE Journal of Biomedical and Health Informatics*, vol. 24, pp. 2146-2156, 2020.

[3] V. Mothukuri, R. M. Parizi, S. Pouriyeh, Y. Huang, A. Dehghantanha, G. Srivastava, "A survey on security and privacy of federated learning", *Future Generation Computer Systems*, vol. 115, pp. 619-640, 2021.

[4] M. Aledhari, R. Razzak, R. M. Parizi and F. Saeed, "Federated Learning: A Survey on Enabling Technologies, Protocols, and Applications," *IEEE Access*, vol. 8, pp. 140699-140725, 2020.

[5] Y. Chen, X. Qin, J. Wang, C. Yu and W. Gao, "FedHealth: A Federated Transfer Learning Framework for Wearable Healthcare," *IEEE Intelligent Systems*, vol. 35, pp. 83-93, 2020.

Mini-Track Chairs Biography

Ali Dehghantanha is a leader in the field of cybersecurity and threat intelligence. He has been an invited or keynote speaker in many national and international conferences, is a well-known media contributor in cybersecurity and has published more than 100 peer-reviewed papers in top journals and conferences in the field. Dr. Dehghantanha is a professor of Cybersecurity at the University of Guelph, Ontario, Canada, the director of Cyber Science Lab (https://cybersciencelab.org/) and the director of the Master of Cybersecurity and Threat Intelligence program. He was is awarded a Tier 2 Canada Research Chair in Cybersecurity and Threat Intelligence in 2020 and a prestigious EU Marie Curie International Incoming Research Fellowship in "Privacy Respecting Digital Forensics" in 2015. His lab is continually offering positions to talented Ph.D. and postdoctoral candidates in cybersecurity.

Reza M. Parizi is the director of the Decentralized Science Lab (dSL) in the College of Computing and Software Engineering at Kennesaw State University, GA, USA. He is a consummate AI technologist and cybersecurity researcher with an entrepreneurial spirit with experiences working on industry projects from Oracle and SunTrust. He is a senior member of the IEEE, IEEE Blockchain Community, and ACM. Prior to joining KSU, he was with the New York Institute of Technology. He received a Ph.D. in Software Engineering in 2012 and M.Sc. and B.Sc. degrees in Computer Science respectively in 2008 and 2005. His research interests are R&D in decentralized AI, federated learning, blockchain systems, smart contracts, and emerging issues in the practice of secure software-run world applications.

Qi Zhang received a Ph.D. degree in computer science from Georgia Institute of Technology, Atlanta, USA. He is currently a research scientist at IBM Thomas J. Watson Research Center. His research interests include Blockchain systems, Cloud computing, and Big Data processing. He published research articles in refereed journals and conference proceedings such as International Conference on Blockchain, IEEE Blockchain NewsLetter, IEEE TC,

IEEE TSC, ACM CSUR, VLDB, SC, HPDC, IEEE ICDCS, IEEE ICWS, IEEE CLOUD. He served as the PC chair of the International Conference on Blockchain and the program committee member for many blockchain conferences, such as IEEE Blockchain, IEEE International Conference of Blockchain and Cryptocurrency, and International Workshop on Blockchain and Data Management. He is also a frequent reviewer for international research journals such as IEEE TSC, IEEE TCC, IEEE TDSC, and international conferences such as ICDCS, SIGMOD, and Middleware. Dr. Zhang is a co-inventor of more than 20 filed US patents and a recipient of First Patent Application Award from IBM in 2018. He has also received Top 5 Picks Award from IEEE ICWS 2017, Outstanding Paper Award in IEEE Blockchain 2019, IBM Research Outstanding Accomplishment Award and Best Paper Award from the 29th Annual International Conference on Computer Science and Software Engineering in 2019.

Katrin Franke is a professor in computer science within the information security environment at NTNU in Gjøvik. In 2007 she joined the Norwegian Information Security Lab (NISlab) with the mission to establish research and education in digital and computational forensics. In this context she was instrumental in setting up the partnership with the Norwegian police organisations as part of the Center for Cyber and information Security (CCIS). Dr. Franke is now leading the NTNU Digital Forensics group. Dr. Franke has 20+ years experiences in basic and applied research for financial services & law enforcement agencies (LEAs) working closely with banks and LEAs in Europe, North America and Asia.