Algorithmic Foundation of Federated Learning with Sequential Data

Mingrui Liu

Department of Computer Science, George Mason University
4400 University Drive, Fairfax, VA 22030, USA
mingruil@gmu.edu

Abstract

The current analysis of federated optimization algorithms for training deep neural networks assumes that the data is non-sequential (e.g., images), which incurs a smooth loss objective. In contrast, edge devices generate lots of *sequential data* every day, where these sequences exhibit significant *sequential correlation* at different time stamps (e.g., text messages). In order to learn from such sequential data, people typically use a class of neural networks that is inherently nonsmooth, with a potentially unbounded smoothness parameter. Examples include recurrent neural networks, long-short-term memory networks, and transformers. It remains unclear how to design provably efficient algorithms for training these neural networks to learn from sequential data. My goal is to lay the algorithmic foundation of federated learning with sequential data, which contributes novel algorithms for learning from a range of real-world sequential data (e.g., natural language, electronic health record, transportation, time series, etc.) using state-of-the-art deep neural networks.

In this talk, I will first motivate the problem by showing that the transformer, which is widely used for sequential data learning, has an unbounded smooth landscape. Then, I will introduce provably efficient federated deep learning algorithms in the presence of unbounded smoothness. In particular, I will introduce a few efficient algorithms for various settings of federated learning, including homogeneous data, heterogeneous data, and partial client participation. The main result is twofold. First, we show that the designed algorithms provably small computational and communication complexities. Second, we establish fundamental hardness results in the unbounded smoothness setting. Ultimately, I will discuss the future challenges of extending our research framework from medium-scale neural networks to large language models.

Federated Learning with Sequential Data. This talk involves the line of work (Liu et al. 2022; Crawshaw, Bao, and Liu 2023a,b), which focus on federated optimization algorithms for autoregressive models. These models, including transformer, are shown to have unbounded smoothness parameter (Crawshaw et al. 2022).

Biography. Dr. Mingrui Liu is an assistant professor at the Department of Computer Science at George Mason University since August 2021. Before that he was a postdoctoral fellow at Boston University from 2020-2021. He received his Ph.D. in Computer Science at The University of Iowa in August 2020. His research interests include machine learning, mathematical optimization, statistical learning theory, and deep learning. He has served as an area chair for NeurIPS/AISTATS/IJCAI. Dr. Liu has published over 20 papers on leading machine learning venues such as NeurIPS, ICML, ICLR and JMLR. He has co-organized the CVPR 2022 tutorial on AUC maximization (Yang et al. 2022). His work (Liu et al. 2022; Ji et al. 2022; Guo et al. 2020; Hao, Gong, and Liu 2024) received spotlights at NeurIPS and ICLR.

References

Crawshaw, M.; Bao, Y.; and Liu, M. 2023a. EPISODE: Episodic Gradient Clipping with Periodic Resampled Corrections for Federated Learning with Heterogeneous Data. In *The Eleventh International Conference on Learning Representations*.

Crawshaw, M.; Bao, Y.; and Liu, M. 2023b. Federated Learning with Client Subsampling, Data Heterogeneity, and Unbounded Smoothness: A New Algorithm and Lower Bounds. *Advances in Neural Information Processing Systems*.

Crawshaw, M.; Liu, M.; Orabona, F.; Zhang, W.; and Zhuang, Z. 2022. Robustness to Unbounded Smoothness of Generalized SignSGD. *Advances in Neural Information Processing Systems*.

Guo, Y.; Liu, M.; Yang, T.; and Rosing, T. 2020. Improved Schemes for Episodic Memory-based Lifelong Learning. *Advances in Neural Information Processing Systems*, 33.

Hao, J.; Gong, X.; and Liu, M. 2024. Bilevel Optimization under Unbounded Smoothness: A New Algorithm and Convergence Analysis. In *The Twelfth International Conference on Learning Representations*.

Ji, K.; Liu, M.; Liang, Y.; and Ying, L. 2022. Will Bilevel Optimizers Benefit from Loops. *Advances in Neural Information Processing Systems*.

Liu, M.; Zhuan, Z.; Lei, Y.; and Liao, C. 2022. A Communication-Efficient Distributed Gradient Clipping Algorithm for Training Deep Neural Networks. In *Advances in Neural Information Processing Systems*.

Yang, T.; Ying, Y.; Narasimhan, H.; and Liu, M. 2022. Deep AUC Maximization: From Algorithms to Practice. *IEEE / CVF Computer Vision and Pattern Recognition Conference*.