UTS UNIVERSITY OF TECHNOLOGY SYDNEY

Privacy-Preserving and Fairness in Machine Learning

by Tao Zhang

Thesis submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy

under the supervision of Tianqing Zhu

University of Technology Sydney
Faculty of Engineering and Information Technology
January 2022

CERTIFICATE OF ORIGINAL AUTHORSHIP

I, Tao Zhang 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 Engineering and Information at the University of Technology Sydney.

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.

[Your Name]

DATE: 06th January, 2021

PLACE: Sydney, Australia

ACKNOWLEDGMENTS

First and foremost, I am extremely grateful to my supervisors, Prof. Tianqing Zhu and Prof. Wanlei Zhou, for their invaluable advice, continuous support, and patience during my PhD study. Their immense knowledge and plentiful experience have encouraged me all the time in my research and daily life. In particular, I am really grateful to Prof. Tianqing for giving me a lot of guidance on how to do research, giving me the space to work on the research that I am interested in, and showing support for my internship.

I would also like to thank Dr Dayong Ye for his technical support on my research and for his invitation to coffee chat almost every week. I would like to thank Dr Angela Huo for being my co-supervisor. I would like to thank Prof. Philip Yu, who gave me suggestions for research and helped me polish papers. I want to thank all external collaborators: Ping Xiong, Zahir Tari, and Philip Yu. Thanks for the insightful discussion and valuable advice from them. Additionally, I would like to express gratitude to all staff in the School of Computer Science for helping PhD students solve issues.

I am fortunate to have been a part of our research group - Cyber Privacy and Safety. I also owe plenty of thanks to my research friends who helped me during my research and daily life during my PhD study: Steven Cheng, Sheng Shen, Mengde Han, Tingting Liao, Congcong Zhu, Yuan Zhao, Chi Liu, Zhuowei Wang, Yanbin Liu, and so many others. It is such an honour to work with all of you. Special thanks to Jing Li; he is always willing to discuss research with me, and he has given me a lot of inspiration.

It is important to strike a balance with life outside the hard work of the lab. I would like to thank my tennis buddies: Jack, Neo, Kenyo, Jason, Jianqiao, Alexander, Huipeng Xue and many others. Special thanks to my friend Steven Holley, who let me understand many aspects of Australia. Most importantly, I am grateful for my family's unconditional, unequivocal, and loving support.

Tao Zhang
Sydney, Australia
January 2022

Dedication

To myself . . .

Abstract

Machine learning is widely deployed in society, unleashing its magic in a wide range of applications following the progress of big data and computing power. However, society is beginning to realize that machine learning models designed to help human beings in various tasks may also have a negative impact on human beings, especially in terms of privacy and fairness. In terms of privacy, data are increasingly collected from human beings, and when these data are used for machine learning, data privacy might be compromised. In terms of fairness, machine learning, as a useful decision-making tool, is widely used to allocate resources and opportunities for humans. Many studies have shown that decisions made by these models may be biased against certain populations. Machine learning has passed the stage of only considering model performance, and ethical issues have a decisive impact on the use of machine learning. This thesis mainly studies how to design a fair and private machine learning model to foster private and fair machine learning and develops methods broadly covering different aspects of privacy and fairness to enhance the trade-off between fairness, privacy and model accuracy. Specifically, it makes the following contributions.

- We propose a correlation reduction scheme with feature selection - selecting features considering data correlation and utility. The proposed scheme involves five steps to manage the extent of data correlation, preserve privacy, and support accuracy in the model outputs.
- We present a framework of fair semi-supervised learning in the pre-processing phase, including pseudo labeling, re-sampling, and ensemble learning to improve accuracy and decrease discrimination. We also propose a framework of fair semi-supervised learning in the in-processing phase. The objective function includes a loss for both the classifier and label propagation and fairness constraints over

labeled and unlabeled data.

- We study the balance between accuracy, privacy and fairness in deep learning by designing two different early stopping criteria to help analysts choose when to stop training a model to achieve their ideal trade-off.
- We investigate how adversarial examples will skew model fairness. We formulate the problem as an optimization problem: maximizing the model bias with the constraint of the number of adversarial examples and the perturbation scale.

Keywords: Machine learning, Differential privacy, Algorithmic fairness

Table of Contents

List of Figures

List of Tables

List of Publications

Published Papers :

1. **Tao Zhang**, Tianqing Zhu, Kun Gao, and Wanlei Zhou. 2021. "Balancing Privacy, Fairness, and Accuracy with Early Stopping Criteria", in IEEE Transactions on Neural Networks and Learning Systems, doi: 10.1109/TNNLS.2021.3129592.

2. **Tao Zhang**, Tianqing Zhu, Jing Li, Mengde Han, Wanlei Zhou and Philip Yu, "Fairness in Semi-supervised Learning: Unlabeled Data Help to Reduce Discrimination," in IEEE Transactions on Knowledge and Data Engineering, doi: 10.1109/TKDE.2020. 3002567.

3. **Tao Zhang**, Tianqing Zhu, Ping Xiong, Huan Huo, Zahir Tari and Wanlei Zhou, "Correlated Differential Privacy: Feature Selection in Machine Learning," in IEEE Transactions on Industrial Informatics, vol. 16, no. 3, pp. 2115-2124, March 2020, doi: 10.1109/TII.2019.2936825.

4. **Tao Zhang**, Dayong Ye, Tianqing Zhu, Tingting Liao, and Wanlei Zhou, 2020. "Evolution of cooperation in malicious social networks with differential privacy mechanisms". Neural Computing and Applications, pp.1-16.

5. **Tao Zhang**, Tianqing Zhu, Renping Liu, and Wanlei Zhou, 2020. "Correlated data in differential privacy: Definition and analysis". Concurrency and Computation: Practice and Experience, p.e6015.

6. Chen, Xin, **Tao Zhang**, Sheng Shen, Tianqing Zhu, and Ping Xiong. "An Optimized Differential Privacy Scheme with Reinforcement Learning in VANET." Computers and Security (2021): 102446.

7. Minghao Wang, Tianqing Zhu, **Tao Zhang**, Jun Zhang, Shui Yu, and Wanlei Zhou, 2020. "Security and privacy in 6G networks: New areas and new challenges". Digital Communications and Networks, 6(3), pp.281-291.

8. Xiuting Gu, Tianqing Zhu, Jie Li, **Tao Zhang**, and Wei Ren. "The Impact of Differential Privacy on Model Fairness in Federated Learning." In International Conference on Network and System Security, pp. 419-430. Springer, Cham, 2020.

SUBMITTED PAPERS :

1. **Tao Zhang**, Tianqing Zhu, Mengde Han, Jing Li, Wanlei Zhou and Philip Yu, 2020. "Fairness Constraints in Semi-supervised Learning". arXiv preprint: 2009.06190.

2. **Tao Zhang**, Tianqing Zhu, Jing Li, and Wanlei Zhou. 2021. "Revisiting Model Fairness via Adversarial Examples", submitted to IEEE Transactions on Neural Networks and Learning Systems.

3. Xin Chen, **Tao Zhang**, Ping Xiong, Sheng Shen. 2021. "Trajectory privacy-preserving with multiple obfuscation over road networks in VANET", submitted to Journal of Information Security and Applications.