

# **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.

Production Note:

SIGNATURE: Signature removed prior to publication.

[Your Name]

DATE: 06<sup>th</sup> 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

<b>List of Figures</b>	<b>xiv</b>
<b>List of Tables</b>	<b>xix</b>
<b>List of Publications</b>	<b>xxi</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Private Machine Learning . . . . .	2
1.2 Fair Machine Learning . . . . .	3
1.3 Private and Fair Machine Learning . . . . .	4
1.4 Research Objectives and Challenges . . . . .	5
1.5 Thesis Outline . . . . .	7
<b>2 Background</b>	<b>9</b>
2.1 Differential Privacy in Machine Learning . . . . .	9
2.1.1 Notation . . . . .	9
2.1.2 Differential Privacy . . . . .	9
2.1.3 Differentially Private Machine Learning . . . . .	12
2.2 Fairness in Machine Learning . . . . .	15
2.2.1 Discrimination Sources . . . . .	15
2.2.2 Fairness Metrics . . . . .	17
2.2.3 Discrimination Removal Methods . . . . .	18
2.3 Interaction between Privacy and Fairness . . . . .	21
2.3.1 Simultaneous Enforcement of Privacy and Fairness . . . . .	21
2.3.2 Mutual Impact of Privacy and Fairness . . . . .	21

<b>3</b>	<b>Correlated Differential Privacy: Feature Selection in Machine Learning</b>	<b>22</b>
3.1	Introduction . . . . .	22
3.2	Preliminaries . . . . .	24
3.2.1	Differential Privacy . . . . .	24
3.2.2	Feature Selection . . . . .	25
3.3	Example of Traffic Monitoring . . . . .	25
3.4	The Extent of Data Correlation . . . . .	27
3.4.1	Correlated Degree . . . . .	27
3.4.2	Correlated Sensitivity . . . . .	28
3.5	Correlation Reduction Based on Feature Selection . . . . .	29
3.5.1	Overview of the Method . . . . .	29
3.5.2	The Proposed CR-FS Scheme . . . . .	29
3.5.3	Discussion . . . . .	33
3.6	Experiments . . . . .	34
3.6.1	Experimental Setup . . . . .	34
3.6.2	Experiments for Data Analysis . . . . .	36
3.6.3	Experiments for Data Publishing . . . . .	39
3.6.4	Discussion . . . . .	42
3.7	Summary . . . . .	42
<b>4</b>	<b>Fairness in Semi-supervised Learning:</b>	
	<b>Unlabeled Data Help to Reduce Discrimination</b>	<b>43</b>
4.1	Introduction . . . . .	43
4.2	Background . . . . .	45
4.2.1	Notations . . . . .	45
4.2.2	Fairness Metrics . . . . .	46
4.2.3	Discrimination Sources . . . . .	47
4.2.4	Bias, Variance and Noise . . . . .	48
4.3	The Proposed Method . . . . .	48
4.3.1	Overview of the Fairness-enhanced Sampling Framework . . . . .	48
4.3.2	Where to Sample . . . . .	49
4.3.3	How to Sample . . . . .	50
4.3.4	How to Train the Model . . . . .	51

4.3.5	Discussion . . . . .	53
4.4	Discrimination Analysis . . . . .	53
4.5	Experiment . . . . .	56
4.5.1	Experiments on Real Data . . . . .	56
4.5.2	Experiments on Synthetic Data . . . . .	63
4.5.3	Discussion and Summery . . . . .	66
4.6	Related Work . . . . .	67
4.6.1	Pre-processing Methods . . . . .	67
4.6.2	In-processing Methods . . . . .	67
4.6.3	Post-processing Methods . . . . .	68
4.6.4	Comparison with Other Work . . . . .	68
4.7	Summary . . . . .	69
<b>5</b>	<b>Fairness Constraints in Semi-supervised Learning</b>	<b>70</b>
5.1	Introduction . . . . .	70
5.2	Preliminaries . . . . .	72
5.2.1	Notations . . . . .	72
5.2.2	Graph-base Regularization . . . . .	73
5.2.3	The Proposed Framework . . . . .	74
5.2.4	Fair SSL of Logistic Regression . . . . .	76
5.2.5	A Case of SVM . . . . .	79
5.2.6	Discussion . . . . .	80
5.3	Experiment . . . . .	80
5.3.1	Experimental Setup . . . . .	81
5.3.2	Experimental Results of Disparate Impact . . . . .	82
5.3.3	Experimental Results of Disparate Mistreatment . . . . .	86
5.3.4	Discussion and Summary . . . . .	91
5.4	Related Work . . . . .	92
5.4.1	Fair Supervised Learning . . . . .	92
5.4.2	Fair Unsupervised Learning . . . . .	93
5.4.3	Comparing with Other Work . . . . .	93
5.5	Summary . . . . .	94

<b>6</b>	<b>Balancing Learning Model Privacy, Fairness, and Accuracy with Early Stopping Criteria</b>	<b>95</b>
6.1	Introduction . . . . .	95
6.2	Background . . . . .	97
6.2.1	Notation . . . . .	97
6.2.2	Fairness Metrics . . . . .	98
6.2.3	Differential Privacy . . . . .	99
6.2.4	Early Stopping Criteria . . . . .	101
6.3	Preliminary Studies . . . . .	101
6.3.1	Preliminary Experiments . . . . .	101
6.3.2	Results . . . . .	102
6.4	Early Stopping Criteria Methods . . . . .	104
6.4.1	Stopping Criterion 1 . . . . .	105
6.4.2	Stopping Criterion 2 . . . . .	106
6.5	Theoretical Analysis . . . . .	107
6.5.1	The Impact of DP-SGD on Gradient . . . . .	107
6.5.2	The Impact of DP-SGD on the Training Process . . . . .	109
6.5.3	The Changes of the Discrimination Level in the Validation Set and Test Set . . . . .	110
6.6	Experiment . . . . .	111
6.6.1	Experimental Setup . . . . .	111
6.6.2	Accuracy and Discrimination Level with SGD and DP-SGD . . . . .	113
6.6.3	Effect of Hyperparameters . . . . .	115
6.6.4	The Effect of Stopping Criteria . . . . .	117
6.6.5	Discrimination Levels in Validation and Test Sets . . . . .	118
6.6.6	Positive Prediction Rate and True Positive Rate . . . . .	119
6.6.7	Discussion and Summary . . . . .	121
6.7	Related work . . . . .	124
6.7.1	Differentially Private Machine Learning . . . . .	124
6.7.2	Fair Machine Learning . . . . .	125
6.7.3	Comparison with Other Work . . . . .	126
6.8	Summary . . . . .	126

<b>7</b>	<b>Revisiting Model Fairness via Adversarial Examples</b>	<b>127</b>
7.1	Introduction . . . . .	127
7.2	Related Work . . . . .	130
7.3	Vulnerability of Model Fairness to Adversarial Attacks . . . . .	131
7.3.1	Vulnerability of Individual Fairness . . . . .	131
7.3.2	Vulnerability of Group Fairness . . . . .	133
7.4	Method . . . . .	135
7.4.1	Problem Formulation . . . . .	135
7.4.2	Adversarial Attack on Individual Fairness . . . . .	135
7.4.3	Adversarial Attack on Group Fairness . . . . .	137
7.5	Experiment . . . . .	139
7.5.1	Experimental Setup . . . . .	140
7.5.2	Experimental Results . . . . .	141
7.6	Summary . . . . .	144
<b>8</b>	<b>Conclusion and Future Work</b>	<b>145</b>
8.1	Conclusion . . . . .	145
8.2	Future Work . . . . .	146
	<b>Bibliography</b>	<b>147</b>

## LIST OF FIGURES

<b>FIGURE</b>	<b>Page</b>
1.1 The interaction between accuracy, privacy and fairness in learning models .	5
3.1 Data correlation for different number of features . . . . .	36
3.2 Privacy-Accuracy trade-off in SVM for different datasets . . . . .	37
3.3 Privacy-Accuracy trade-off in LR for different datasets . . . . .	38
3.4 MAE performance for count queries . . . . .	40
3.5 MAE performance for mean queries . . . . .	41
4.1 The three phases of the fairness-enhanced sampling framework: 1) where to sample, 2) how to sample and 3) how to train the model. Step 1 is to generate a new training dataset which consists of the original dataset and the pseudo labeled dataset. Step 2 is to construct multiple fair datasets through re-sampling. Step 3 is to train a model with each of the fair datasets through ensemble learning to produce the final predictions. . . . .	49
4.2 The trade-off between accuracy (Red) and discrimination level (Blue). (a) LR in Health dataset; (b) SVM in Health dataset; (c) LR in Bank dataset; (d) SVM in Bank dataset; (e) LR in Adult dataset; (f) SVM in Adult dataset. The X-axis is the sample ratio $\rho$ , which denotes that the percentage of $\rho$ unlabeled data are sampled from the unlabeled dataset and then pseudo labeled for training. . . . .	59

4.3	The impact of ensemble learning on the accuracy (Red) and discrimination level (Blue) on (a) LR in Health dataset; (b) SVM in Health dataset; (c) LR in Bank dataset; (d) SVM in Bank dataset; (e) LR in Adult dataset; (f) SVM in Adult dataset. Initially, there is not an obvious link between accuracy and discrimination level. However, as the ensemble size grows, the accuracy and discrimination levels begin to converge. Each point is an average of 50 times.	60
4.4	The impact of sample size on accuracy (Red) and discrimination level (Blue) on (a) LR in Health dataset; (b) SVM in Health dataset; (c) LR in Bank dataset; (d) SVM in Bank dataset; (e) LR in Adult dataset; (f) SVM in Adult dataset. An increasing in the sampling size leads to an increase in accuracy and may help to reduce discrimination level. . . . .	62
4.5	Comparison with original scheme (ORI), uniform sampling (US) and preferential sample (PS) with (a) LR in Health dataset; (b) SVM in Health dataset; (c) LR in Bank dataset; (d) SVM in Bank dataset; (e) LR in Adult dataset; (f) SVM in Adult dataset. With the fairness-enhanced sampling method (FS), discrimination decreases without much cost of accuracy or accuracy increases without much cost of discrimination. . . . .	64
5.1	The trade-off between accuracy and discrimination in proposed method Semi (Red), FS (Blue), US (Blue cross), PS (Yellow cross) and FES (Green cross) under the fairness metric of disparate impact with LR and SVM in two datasets. As the threshold of covariance $c$ increases, accuracy and discrimination increase. The results demonstrate that our method achieves a better trade-off between accuracy and discrimination than other methods. . . . .	83
5.2	The impact of the amount of unlabeled data in the training set on accuracy (Red) and discrimination level (Blue) under the fairness metric of disparate impact with LR and SVM in two datasets. The X-axis is the size of unlabeled dataset; the left y-axis is accuracy, and right y-axis is discrimination level. .	85
5.3	The trade-off between accuracy and discrimination in proposed method Semi (Red), FS (Blue) with LR and SVM in two datasets under the metric of overall misclassification rate. As the threshold of covariance $c$ increases, accuracy and discrimination increase. The results demonstrate that our method using unlabeled data achieves a better trade-off between accuracy and discrimination.	87

5.4	The trade-off between accuracy and discrimination in the proposed method Semi (Red), FS (Blue) with LR and SVM in two datasets under the metric of false negative rate. As the threshold of covariance $c$ increases, accuracy and discrimination increase. . . . .	88
5.5	The trade-off between accuracy and discrimination in proposed method Semi (Red), FS (Blue) with LR and SVM in two datasets under the metric of false positive rate. As the threshold of covariance $c$ increases, accuracy and discrimination increase. . . . .	90
5.6	The impact of the amount of unlabeled data in the training set on accuracy (Red) and discrimination level (Blue) under the fairness metric of overall mistreatment rate with LR and SVM in two datasets. The X-axis is the size of unlabeled dataset; the left y-axis is accuracy, and the right y-axis is discrimination level. . . . .	91
6.1	Challenges in private and fair deep learning: 1) Privacy comes at the cost of accuracy; 2) Fairness comes at the cost of accuracy; 3) Privacy could affect model fairness. . . . .	96
6.2	Training with SGD (Blue) and DP-SGD (Red) in the Bank dataset. . . . .	103
6.3	Training with SGD (Blue) and DP-SGD (Red) in UTK Face dataset. Figure 3 shows that, with DP-SGD, more variations appear during the training in terms of accuracy and discrimination levels. . . . .	104
6.4	Training with SGD (Blue) and DP-SGD (Red) in the Health dataset. . . . .	114
6.5	Training with SGD (Blue) and DP-SGD (Red) in the IMDB dataset. Figure 6.5 shows that, with DP-SGD, more variations appear during the training in terms of accuracy and discrimination levels. . . . .	115
6.6	Training with SGD (Dark blue) and DP-SGD (other colors; $\sigma \cdot C = 1$ and $C$ varies) in the IMDB dataset. Figure 6.6 shows that clipping bound has an impact on discrimination levels. . . . .	116
6.7	Training with SGD (Dark blue) and DP-SGD (other colors; $C = 1$ and $\sigma$ varies) in the IMDB dataset. Figure 7 shows that the noise scale does not produce a distinctive difference in discrimination levels. . . . .	117



6.8	Discrimination level of demographic parity (a) and equal odds (b) on validation set (Blue) and test set (Red) in the Bank dataset. Figure 6.8 shows the changes of discrimination level is similar in validation and test set in the Bank dataset. . . . .	119
6.9	Discrimination level of demographic parity (a) and equal odds (b) on validation set (Blue) and test set (Red) in the Health dataset. Figure 6.9 shows the changes of discrimination level is similar in validation and test set in the Health dataset. . . . .	120
6.10	Discrimination level of demographic parity (a) and equal odds (b) on validation set (Blue) and test set (Red) in the UTK Face dataset. Figure 6.10 shows the trend of discrimination level is similar in validation and test set. . . . .	120
6.11	Discrimination levels on the validation set (Blue) and test set (Red) in the IMDB dataset. Figure 6.11 shows the trend of discrimination levels is similar in both sets. . . . .	121
6.12	Training with SGD (Blue) and DP-SGD (Red) in the Health dataset. (a) Positive predication rate in Group 1 (Solid line) and Group 2 (Dotted line) (b) True positive rate in Group 1 (Solid line) and Group 2 (Dotted line). Figure 6.12 shows how PPR and TPR change in two groups in the Health dataset. . . . .	122
6.13	Training with SGD (Blue) and DP-SGD (Red) in the Bank dataset. (a) Positive prediction rate in the protected group (solid line) and unprotected group (Dotted line) (b) True positive rate in protected group (solid line) and unprotected group (Dotted line). Figure 6.13 shows how PPR and TPR change in two groups in the Bank dataset. . . . .	122
6.14	Training with SGD (Blue) and DP-SGD (Red) in the IMDB dataset. (a) Positive prediction rate in Group 1 (Solid line) and Group 2 (Dotted line) (b) True positive rate in Group 1 (Solid line) and Group 2 (Dotted line). Figure 6.14 shows how PPR and TPR change in two groups in the IMDB dataset. . . . .	123
6.15	Training with SGD (Blue) and DP-SGD (Red) in the UTK Face dataset. (a) Positive predication rate in Group 1 (Solid line) and Group 2 (Dotted line) (b) True positive rate in Group 1 (Solid line) and Group 2 (Dotted line). Figure 6.15 shows how PPR and TPR change for two groups in the UTK Face dataset. . . . .	123

---

7.1	Two examples in the group 1 with negative predicted labels are perturbed into adversarial examples. These original examples are similar to their adversarial examples but are being treated differently by the model. This violates individual fairness, and furthers skews the demographic information in group fairness. . . . .	129
7.2	An example of group adversarial bias. . . . .	134
7.3	Group adversarial bias and individual group adversarial bias on the German and Bank datasets with increasing perturbation scale $\epsilon$ . Results show that adversarial examples in our proposed methods can significantly skew individual fairness and group fairness more than baselines. . . . .	143
7.4	The mean of the perturbation norm on the German and Bank datasets with an increasing number of adversarial examples $\epsilon$ in Fig. 7.4 (a) and 7.4 (b). The result of the group adversarial bias on the German and Bank datasets with the growing number of adversarial examples in Fig. 7.4 (c) and 7.4 (d). . . . .	143

## LIST OF TABLES

<b>TABLE</b>	<b>Page</b>
2.1 Notations . . . . .	10
3.1 Users' locations at different times . . . . .	26
3.2 The sum counts of users' locations . . . . .	26
3.3 Number of features in different stages . . . . .	35
4.1 Two discriminatory datasets tested on the discriminatory test dataset in ORI method and the proposed fairness-enhanced method (FS) with LR and SVM. We show accuracy (Acc), discrimination level (Dis) and the number of data points of each group in the discriminatory test dataset after classification. . . . .	65
4.2 Two discriminatory datasets tested on the fair test dataset in ORI method and the proposed fairness-enhanced method (FS) with LR and SVM. We show accuracy (Acc), discrimination level (Dis) and the number of data points of each group in the fair test dataset after classification. . . . .	65
5.1 The impact of fairness constraints on different datasets in terms of accuracy (Acc) and discrimination level (Dis) under the fairness metric of disparate impact with LR in the Health dataset. . . . .	84
5.2 The impact of fairness constraints on different datasets in terms of accuracy (Acc) and discrimination level (Dis) under the fairness metric of disparate impact with LR in the Titanic dataset. . . . .	84
5.3 The impact of fairness constraints on different datasets in terms of accuracy (Acc) and discrimination level (Dis) under the fairness metric of overall misclassification rate with LR in the Bank dataset. . . . .	89

5.4	The impact of fairness constraints on different datasets in terms of accuracy (Acc) and discrimination level (Dis) under the fairness metric of overall misclassification rate with LR in the Titanic dataset. . . . .	89
6.1	Accuracy, discrimination levels and training epoch for the two early stopping criteria with different parameters on the UTK Face dataset . . . . .	118
6.2	Accuracy, discrimination level and training epoch for the two early stopping criteria with different parameters on the IMDB dataset . . . . .	119
7.1	Description of datasets . . . . .	140
7.2	Results for all datasets and all methods concerning fairness metrics and perturbation metrics . . . . .	142

## 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*.