



TMLR Young Scientist SEMINAR

2023 SERIES

Trustworthy Machine Learning and Reasoning Group



Mr. Jiaheng Wei

Ph.D. Candidate, Computer Science and Engineering, UC Santa Cruz.

Date: 13 April 2023 (Thursday)
Time: 15:00 – 16:00 (HKT)
Zoom: https://meeting.tencent.com/dm/MDwJzAnNQeIS

Learning with Noisy Labels: Theoretical Approaches and Empirical Studies

ABSTRACT

A machine learning system continuously observes noisy training annotations, and it remains challenging to perform robust training in such scenarios. In this talk, I will briefly introduce earlier and classical approaches to mitigate the impacts of noisy labels, which rely on accurate estimations of the noise/error rates of the labels. Then I will move on to discuss my two approaches of robust training methods under synthetic noisy labels with theoretical guarantees (Robust f-divergence loss functions and Negative Label Smoothing). Synthetic label noise, though has clean structures which greatly enable statistical analyses, often fails to model real-world noise patterns. To better understand real-world label noise, we collect human-annotated real-world noisy labels on CIFAR datasets via Amazon Mechanical Turk. We show indeed the real-world noise patterns impose new and outstanding challenges as compared to synthetic label noise, in the view of noise transition vectors, hypothesis testing, and memorization effect of model predictions. These observations require us to rethink the treatment of noisy labels.



Jiaheng Wei (he/him) is a PhD candidate of Computer Science and Engineering at UC Santa Cruz (2019 - present). He was a student researcher at Google Research, Brain Team (2022). He received the M.S. degree (Data Science) at Brown University and B.S. degree in Honors Science (Mathematics and Applied Mathematics) and Honors Youth (Gifted Young) from Xi'an Jiaotong University. His research interests mainly include robust learning under real-world constraints (i.e., label noise in human-generated data, class-imbalanced learning, group distributional robustness, fairness). His recent work NLS is selected for long presentation in ICML 2022. Jiaheng has also presented multiple talks about weakly supervised learning, relevant teams and organizations include Google Brain, Bosch AI, University of Toronto, AI Time, ICML (long-presentation), and invited talks from workshops in WSDM, ACML, etc. He coorganized the first competition on ``Learning and Mining with Noisy Labels" at IJCAI 2022. Email: jiahengwei@ucsc.edu. Web page: https://weijiaheng.github.io/.

ENQUIRY

Email: bhanml@comp.hkbu.edu.hk