香港浸會大學
HONG KONG BAPTIST UNIVERSITY

DEPARTMENT OF
COMPUTER SCIENCE
計算機科學系

TMLR Young Scientist SEMINAR

2022 SERIES

Trustworthy Machine Learning and Reasoning Group

Mr. Wenxiao Wang

Ph.D. student
CS Department,
University of Maryland.

Date: 15 December 2022 (Thursday)

Time: 09:00 – 10:00 (HKT)

Zoom: https://meeting.tencent.com/dm/XBT2e5b8OFCy

Lethal Dose Conjecture: From Few-shot Learning to Potentially Near Optimal Defenses against Data Poisoning

ABSTRACT

Data poisoning considers an adversary that distorts the training set of machine learning algorithms for malicious purposes. In this talk, I will first introduce aggregation-based certified defenses against general data poisoning. Then I will present a conjecture targeting the fundamentals of robustness against data poisoning, namely Lethal Dose Conjecture. This conjecture relates the optimal robustness against data poisoning to few-shot learning with clean distributions. I will provide theoretical results verifying the conjecture in multiple cases. I will also show why the conjecture is important: If the conjecture is true for a given task, aggregation-based defenses will be (asymptotically) optimal—if we have the most data-efficient learner, we can turn it into one of the most robust defenses against data poisoning, essentially reducing data poisoning defenses to few-shot learning.

BIOGRAPHY

Wenxiao Wang is currently a CS Ph.D. student at University of Maryland, working with Prof. Soheil Feizi. His research interests include machine learning robustness, privacy-preserving machine learning and self-supervised representation learning. Lately he has been working on the mitigation of data poisoning. Wenxiao received his B.S. degree in Computer Science from Yao Class, Tsinghua University in 2020. He was a research intern at Bytedance (summer 2022); a research assistant at IIIS, Tsinghua University (2020-2021), working with Prof. Hang Zhao in his MARS Lab; a visiting student researcher at UC Berkeley (2019), working with Dr. Xinyun Chen, Prof. Ruoxi Jia and Prof. Dawn Song; an intern in Bytedance AI Lab (2018), working with Dr. Yi He and Prof. Lei Li.

ENQUIRY

Email: bhanml@comp.hkbu.edu.hk