When Robust Machine Learning Meets Noisy Supervision: Mechanism, Optimization and Generalization

Bo Han

Centre for Artificial Intelligence

University of Technology Sydney

A thesis submitted for the degree of

Doctor of Philosophy

May, 2019

CERTIFICATE OF ORIGINAL AUTHORSHIP

I certify that the work in this thesis has not previously been submitted for a degree nor has it been submitted as part of requirements for a degree except as part of the collaborative doctoral degree and/or fully acknowledged within the text.

I also certify that the thesis has been written by me. Any help that I have received in my research work and the preparation of the thesis itself has been acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

This research is supported by an Australian Government Research Training Program.

Production Note:
Signature removed prior to publication.

Signature of Student: Bo Han
Date: 31/May/2019

I would like to dedicate this thesis to my loving parents and family.

Acknowledgements

I had a wonderful journey at University of Technology Sydney (UTS) for pursuing my PhD degree in the past four years. I am deeply grateful to all those who helped me during the writing of this thesis.

First of all, I would like to express my foremost and deepest gratitude to my advisor, Prof. Ivor W. Tsang, for his visionary, inspiring, and insightful guidance during my PhD period. Honestly speaking, I knew quite little about genuine research when I began my PhD study in UTS. He taught me how to find interesting ideas, how to develop solid algorithms, and how to write papers all from scratch. Besides his knowledge, his research taste encourages me towards the state-of-the-art algorithms, while his work ethic motivates me to become tough and strong. Inspiring by his famous saying "research is like a step function", I am always positive and full of hope in front of my research. Meanwhile, I would like to show my deepest gratitude to my co-advisor, Associate Prof. Ling Chen. She helped me revise my papers, invited me good meals, and often encouraged me, even when I felt very frustrated in my early research. Without their extraordinary support and supervision, I could not achieve what I have right now. I feel extremely grateful for their efforts and friendships.

Moreover, I am greatly indebted to Prof. Chengqi Zhang. Without his endless trust and help, I would never have this valuable opportunity to pursue my PhD research in UTS. Meanwhile, I greatly indebted to the Centre for Artificial Intelligence (CAI) directed by Prof. Jie Lu. In CAI, I got tremendous opportunities to learn from many world-famous experts; I met many colleagues with great enthusiasm for scientific research; and I was also permitted to attend many prestigious international conferences related to my research. Studying in UTS and CAI will be a fantastic memory that I will never forget.

During my PhD period, I am very fortunate to join RIKEN-AIP as a research intern in 1-year "AI Residency" Program, working with Prof. Masashi Sugiyama and Dr. Gang Niu. Prof. Masashi Sugiyama always supported my research ideas, commented all my papers carefully, and approved most of my academic travels financially. Dr. Gang Niu often

brought me excellent ideas, denoted the bar of top-tier research, and handed on many technical details. Our research styles matched very well. Meanwhile, Prof. Mingyuan Zhou from University of Texas Austin taught me Bayesian machine learning; Prof. Xiaokui Xiao from National University of Singapore taught me differential privacy; Dr. Quanming Yao from 4Paradigm Inc. taught me how to write and revise a good paper, and his work ethic won all my respects; Weihua Hu from Stanford University taught me how to evaluate a good idea, and how to follow excellence.

During my PhD period, I have been extremely fortunate to collaborate with three great minds, Jiangchao Yao, Yuangang Pan, and Xingrui Yu, who have prominent impact on my research with their outstanding expertise and profound vision. I will never forget the collaboration experiences with them for catching each deadline. Meanwhile, I must thank to all my wonderful friends, classmates and colleagues, Prof. Yi Yang, Dr. Guodong Long, Dr. Miao Xu, Dr. Liang Zheng, Dr. Zhongwen Xu, Dr. Yan Yan, Dr. Mingming Gong, Dr. Tongliang Liu, Dr. Chen Gong, Dr. Xiyu Yu, Dr. Guoliang Kang, Dr. Shaoli Huang, Dr. Liu Liu, Dr. Jiang Bian, Dr. Jing Chai, Dr. Xiaobo Shen, Dr. Haishuai Wang, Dr. Xun Yang, Dr. Guansong Pan, Dr. Peng Hao, Dr. Qinzhe Zhang, Dr. Fan Zhang, Yiliao Song, Feng Liu, Hehe Fan, Hu Zhang, Yanbing Liu, Linchao Zhu, Pingbo Pan, Zhedong Zheng, Ruiqi Hu, Donna Xu, Yaxin Shi, Bo-Jian Hou, Yao-Xiang Ding, and many others, for every enjoyable moment.

Last but not the least, my gratitude also extends to all my family members who have been consistently supporting, encouraging and caring for me all of my life! Most importantly, I feel strongly indebted to my parents Junxiao Han and Jianping Wang. I would never have gone so far without their unconditional love, accompany and support. I would also express my special thanks to my aunt Jianhua Wang, uncle Zhisheng Si and cousin Wei Si for their great help in many years.

Abstract

Modern machine learning is migrating to the era of complex models (i.e., deep neural networks), which requires a plethora of well-annotated data. Crowdsourcing is a promising tool to achieve this goal, since a plethora of labels that can be efficiently collected from crowdsourcing services at very low cost. However, existing crowdsourcing approaches barely acquire a sufficient amount of high-quality labels. This brings the *first* question: How to design the robust mechanism to improve the label quality?

Without such robust mechanism, labels annotated by crowdsourced workers are often noisy, which inevitably degrades the performance of large-scale optimizations, including the prevalent stochastic gradient descent (SGD). Specifically, these noisy labels adversely affect updates of the primal variable in conventional SGD. This bring the *second* question: How to optimize the training model robustly under noisy labels?

Without such robust optimization, it is challenging to train deep neural networks robustly with noisy labels, as the learning capacity of deep neural networks is so high that they can totally memorize and over-fit on these noisy labels. This brings the *third* question: How to acquire the robust model with good generalization under noisy labels? Therefore, in this thesis, we aim to develop a series of robust machine learning approaches, so that they can perfectly handle the difficult from noisy supervision. Our works are summarized as follows:

Chapter 2 answers the first question. Motivated by the "Guess-with-Hints" answer strategy from the Millionaire game show, we introduce the hint-guided approach into crowdsourcing to deal with this challenge. Our approach encourages workers to get help from hints when they are unsure of questions. Specifically, we propose a hybrid-stage setting, consisting of the main stage and the hint stage. When workers face any uncertain question on the main stage, they are allowed to enter the hint stage and look up hints before making any answer. A unique payment mechanism that meets two important design principles is developed. Besides, the proposed mechanism further encourages high-quality workers less using hints, which helps identify and assigns larger possible payment to them.

Experiments are performed on Amazon Mechanical Turk, which show that our approach ensures a sufficient number of high-quality labels with low expenditure and detects high-quality workers.

Chapter 3 answers the second question. We propose a robust SGD mechanism called PrOgressive STochAstic Learning (POSTAL), which naturally integrates the learning regime of curriculum learning (CL) with the update process of vanilla SGD. Our inspiration comes from the progressive learning process of CL, namely learning from "easy" tasks to "complex" tasks. Through the robust learning process of CL, POSTAL aims to yield robust updates of the primal variable on an ordered label sequence, namely from "reliable" labels to "noisy" labels. To realize POSTAL mechanism, we design a cluster of "screening losses", which sorts all labels from the reliable region to the noisy region. We derive the convergence rate of POSTAL realized by screening losses. Meanwhile, we provide the robustness analysis of representative screening losses. Experiments on benchmark datasets show that POSTAL using screening losses is more effective and robust than several existing baselines.

Chapter 4 answers the third question. Motivated by the memorization effects of deep networks, which shows networks fit clean instances first and then noisy ones, we present a new paradigm called "*Co-teaching*" combating with noisy labels. We train two networks simultaneously. First, in each mini-batch data, each network filters noisy instances based on memorization effects. Then, it teaches the remained instances as the useful knowledge to its peer network for updating the parameters. Empirical results on three benchmark datasets demonstrate that, the robustness of deep learning models trained by Co-teaching approach is much superior than that of state-of-the-art methods.

Contents

List of Figures

List of Tables