Distribution-based Active Learning

by

Xiaofeng Cao

Australian Artificial Intelligence Institute
University of Technology Sydney

Supervisor: Prof. Ivor W. Tsang
A thesis submitted for the degree of

Doctor of Philosophy

Sydney, Australia

2021

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

This research is supported by the Australian Government Research Training Program under Grant DP180100106 and DP200101328.

Production Note:

Signature: Signature removed prior to publication.

Date: 30-Apr-2021

Acknowledgements

In a sense, Prof. Ivor W. Tsang brought hope for my research career. The early journey of pursuing the Ph.D. degree in Australia was tortuous and arduous. As the Hong Kong film propagated in my childhood, *"professional"* is the most important impression Ivor gave me. He gradually reversed my dilemma on the academic path using his values. He is a humble scholar who maintains earnest and prudent attitudes, leading me to be a righteous person. I can never forget what he often told me: *"what kind of person you want to be, what you will do"*.

My research went smoothly when I was a master's degree candidate in China. This leads me to ambitiously select some challenging research topics at the beginning of my Ph.D. study. Active learning theory, a fundamental topic in PAC learning, started my Ph.D. journal. Also because of this, it is difficult to understand those theorems and proofs at the beginning, since my master research background was totally about applications. However, Ivor never gave up on me. Continuous encouragement and enthusiasm inspired me to explore the deeper theoretical mysteries behind those complicated formulas and theorems. My theoretical ability thus has been greatly improved. This is something I never thought of. No matter how difficult the time, I always insisted on one truth: *"if you cannot please others, you can please yourself "*. This made me gradually understand why Ph.D. was termed philosophy in early western culture. The harvested values will accompany me all my life.

I met my wife in Australia. Lin is my biggest support in this learning journey. Love and marriage can be selfless. We married at 160 Marsden St, BDM Parramatta, Sydney. Thanks to the wedding witness officer Ms. Aila and my friend Mr. Shuai Cheng. Thanks my co-supervisor Prof. Ling Chen for her help in my study. I want to thank my father-in-law and mother-in-law. They provided strong support in my study whether in funding or life. This allows us to establish a small family at 174/11 Potter Street, Waterloo, Sydney, where provides us with very high-quality study and life. I know they do not care about these expenses. However, it greatly relieved my stress and anxiety during my second half of pursuing the PhD degree. My father-in-law is a farsighted person in Hong Kong. I saw the elite spirits of the Cantonese and Chaoshan. His extraordinary achievements inspired me to work harder and go further. My mother-in-law is an elegant lady who requires Lin to unconditionally support my study in daily life. I want to thank my parents, brothers, and sister in law for their consistent support in my study. Specially thanks to my brother, he supported our family and relieved my stress during this period.

"Be humble, be reasonable, and be respected."

—It is the advice from Ivor before I am leaving UTS and going to start my personal academic career. I will spread this motto to the younger students.

Xiaofeng Cao
Sydney, Australia, 2021.

Abstract

Active learning aims to maximize the learning performance of the current hypothesis by drawing as few labels as possible from an input distribution. To build a near-optimal hypothesis, halfspace learning improved the generalization of a perceptron vector over a unit sphere, presenting model guarantees for the reliable (practical) active learning, in which the error disagreement coefficient controls the hypothesis update via pruning the hypothesis class. However, this update process critically depends on the initial hypothesis and the coefficient. Their improper settings may improve the bounds on the label complexity, which estimates the label demands before achieving a desired error for the hypothesis. One question thus arises: how to reduce the label complexity bounds? In a worse situation, estimating updates of hypothesis using error lacks feasible guarantees, if the initial hypothesis is a null (insignificant) hypothesis. Another question also arises: how to control the hypothesis update without errors, when estimating the error disagreement is infeasible? For error disagreement, most of its generalizations regarding to hypothesis update, either make strong distribution assumptions such as halfspace learning, or else they are computationally prohibitive. How to improve the performance of deep active learning based on the theoretical results of active learning of halfspace?

This thesis tries to answer the three questions from shattering, disagreeing, and matching over distributions. With halfspace learning, the first work presents a novel perspective of shattering the input distribution that, guaranteeing from a lower bound on Vapnik-Chervonenkis (VC) dimension, further reduces the label complexity of active learning. When estimating errors is infeasible, the second work proposes a distribution disagreement graph coefficient, which estimates hypothesis from distribution, yielding a tighter bound on typical label complexity. The constructed hyperbolic model, generalizing distribution disagreement by focal representation, shows effective improvements compared to generalization algorithms of error disagreement. On deep learning settings for active learning, the Bayesian neural network shows expressive distribution matching on the massive training parameters, which allows estimating error disagreement can work effectively. We thus integrate the error and distribution disagreements to establish a uniform framework, which matches the geometric core-set expression of the distribution, interacting with a deep learning model.

Contents

List of Figures

List of Tables