Robust Rank Aggregation and its Applications

Yuangang Pan

Faculty of Engineering and Information Technology
University of Technology Sydney

This dissertation is submitted for the degree of
Doctor of Philosophy

September 2019

Certificate of Original Authorship

I hereby declare that this thesis, is submitted in fulfilment of the requirements for the award of Doctor of Philosophy, in the Faculty of Engineering and Information Technology at the University of Technology Sydney.

This thesis is wholly my own work unless otherwise reference or acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis. The contents of this dissertation are original and have not been submitted in whole or in part 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.

Date:

24 Dec 2019

I would like to dedicate this thesis to my family who, through thick and thin, has been there for me. Their support and drive are what have made me who I am.

Acknowledgements

First and foremost, my utmost gratitude goes to my supervisor Professor Ivor W. Tsang, for his cordiality and generousness in offering me this research opportunity in his group. Life has not been easy for me, especially at the beginning of this adventure. He has always been there for me, serving as a friend and a mentor. His illuminating instructions and insightful guidance helped me to avoid being unnecessarily stuck and kept me focused on the nature of science itself. His consistent support and encouragement helped me to survive the many struggles I experienced on my Ph.D. journey. His depth and breadth of knowledge, insightful vision, and wisdom have not only guided me through my Ph.D. research but also shaped what I have achieved in this dissertation. His dedication and professionalism also set a great example from which I can learn and benefit from in my future career.

During my time in Sydney, I have had the privilege of meeting many excellent people. Being more senior to me, I have appreciated how they have generously shared many insightful thoughts on life and research. Having them as companions in my daily life makes me feel life brighter. Their enthusiasm in research, diligent attitude, depth and breadth of knowledge guided me through my Ph.D. research and shaped what I have achieved. My warmest gratitude, therefore, goes to Dr. Qiong Wang, Dr. Yali Du, Dr. Jiangchao Yao, Yiming Jiang, Licheng Feng, Prof. Weijie Chen, Prof. Weiwei Liu, Dr. Muming Zhao, Dr. Donna Xu, Yueming Lyu, Xiaoqi Jiang and Xu Chen for many illuminating conversations on life and research.

Thankyou also to the University of Technology Sydney for offering me this rare opportunity. My appreciation also goes to those fine minds in our "Eating Group". Your participation has made this journey for me so memorable. I would like to thank especially Dr. Yan Zhang, Yaxin Shi, Yinghua Yao, Jing Li, Xiaowei Zhou, Xingrui Yu. Thank you all!

Last but not least, I owe the deepest debt of gratitude to my family for their unbounded love and unselfish support. Those miseries my parents suffered for me will be buried in my heart forever.

Abstract

Rank aggregation (RA) refers to the task of recovering the total order over a set of items, given a collection of preferences over the items. The flexible collection of preferences enables successful application of RA in various fields, e.g., image rating and bioinformatics. A basic assumption underlying the vanilla RA is that all preferences are provided by homogeneous users. However, this assumption is rarely satisfied in real applications, due to the complex real situation. Therefore, RA usually suffers from model misspecification, namely the inconsistency between the collected preferences and the homogeneity assumption. Another challenge associated with RA is the scalability issue. In particular, RA usually involves ranking over tens of thousands of items, leading to an exponential volume of preferences for aggregation. Therefore, an inappropriate inference method would limit the application of the proposed model.

This thesis considered RA under model misspecification in the following three scenarios:

- In a crowdsourcing scenario, sufficient annotations from each user are available, which enables exploration of user heterogeneity to account for model misspecification. Therefore, I proposed a reliable CrowdsOUrced Plackett-LucE (COUPLE) model, which introduces an uncertainty vector to make a fine-grained categorization of users. Meanwhile, a general Bayesian Moment Matching (OnlineGBMM) was proposed, to ensure an analytic Bayesian update with an almost twice differentiable likelihood function.

- In a general setting, typical model augmentation methods would cause overfitting, because insufficient annotations from each user are available. Inspired by the distributional robust literature, I proposed CoarsenRank, which performs regular RA over a neighborhood of preferences. The resultant inference would enjoy robustness against model misspecification. To this end, I first defined a neighborhood of the rank dataset using relative entropy. Then, I instantiated CoarsenRank with three popular probability ranking models and discussed the optimization strategies.

- RA for mental fatigue monitoring. Common practices for mental fatigue monitoring refer to predicting the reaction time (RT) by aggregating the EEG signal from multiple heterogeneous EEG channels. Let us consider the RT as the item score and view each EEG channel as a user. The mental fatigue monitoring task could be formulated as RA under model misspecification, particularly in a crowdsourcing scenario. To address this problem, a Self-Weight Ordinal REgression (SWORE) model with Brain Dynamics table (BDtable) is proposed. The SWORE model could give a reliable evaluation of brain dynamics preferences from multiple channels, while the BDtable is employed to calibrate the SWORE model by utilizing the proposed online generalized Bayesian moment matching (OGMM) algorithm.

Table of contents

List of figures

List of tables