



The Hong Kong Polytechnic University Department of Applied Mathematics

Colloquium

Crowdsourcing Utilizing Subgroup Structure of Latent Factor Modeling

By

Prof. Annie QU University of California, Irvine



Abstract

Crowdsourcing has emerged as an alternative solution for collecting large scale labels. However, the majority of recruited workers are not domain experts, so their contributed labels could be noisy. In this paper, we propose a twostage model to predict the true labels for multicategory classification tasks in crowdsourcing. In the first stage, we fit the observed labels with a latent factor model and incorporate subgroup structures for both tasks and workers through a multi-centroid grouping penalty. Group-specific rotations are introduced to align workers with different task categories to solve multicategory crowdsourcing tasks. In the second stage, we propose a concordance-based approach to identify high-quality worker subgroups who are relied upon to assign labels to tasks. In theory, we show the estimation consistency of the latent factors and the prediction consistency of the proposed method. The simulation studies show that the proposed method outperforms the existing competitive methods, assuming the subgroup structures within tasks and workers. We also demonstrate the application of the proposed method to real world problems and show its superiority.

Biography

Dr. Annie Qu is Chancellor's Professor, Department of Statistics, University of California Irvine. She received her Ph.D. in Statistics from the Pennsylvania State University. Prof. Qu's research focuses on solving fundamental issues regarding structured and unstructured large-scale data, and developing cutting-edge statistical methods and theory in machine learning and algorithms on personalized medicine, text mining, recommender systems, medical imaging data and network data analyses for complex heterogeneous data. The newly developed methods are able to extract essential and relevant information from large volume high-dimensional data. Her research has impacts in many fields such as biomedical studies, genomic research, public health research, social and political sciences.

Before she joined the UC Irvine, Dr. Qu was a Data Science Founder Professor of Statistics, and the Director of the Illinois Statistics Office at the University of Illinois at Urbana-Champaign. She was awarded as Brad and Karen Smith Professorial Scholar by the College of LAS at UIUC, a recipient of the NSF Career award in 2004-2009. She is a Fellow of the Institute of Mathematical Statistics, a Fellow of the American Statistical Association, and a Fellow of American Association for the Advancement of Science. She is also a recipient of Medallion Award and Lecturer in 2024.

Date: 6 July 2023 (Thursday) Time: 15:00-16:00 (Hong Kong Standard Time GMT +8) Venue: HJ303 Speaker: Prof. Peiyong Qu, University of California, Irvine Host: Prof. Xingqiu Zhao, The Hong Kong Polytechnic University