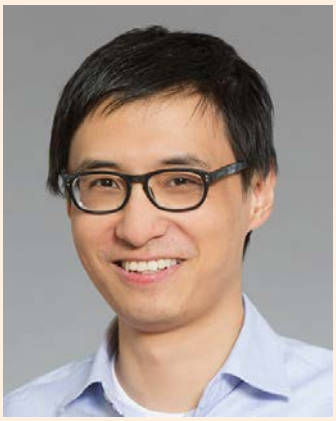
The Hong Kong University of Science and Technology

Department of Information Systems, Business Statistics and Operations Management
Financial Technology Thrust Area, Society Hub, HKUST(GZ)

Joint Seminar Announcement



Learning from Crowdsourced Multi-Labeling – A Variational Bayesian Approach

by

Prof. Junming YIN
Assistant Professor
University of Arizona

Date : **Tuesday, 2 February 2021**
Time : **11:00 am - 12:15 pm (Hong Kong Time)**
Zoom Details : **Click here to join Zoom**
Meeting ID: 980 4923 8100
Passcode: 854264

Abstract: Microtask crowdsourcing has emerged as a cost-effective approach for obtaining large-scale labeled data. Crowdsourcing platforms, such as MTurk, provide an online marketplace where task requesters can submit a batch of microtasks for a crowd of workers to complete for a small monetary compensation. As the information collected from a crowd can be prone to errors, additional algorithmic techniques are needed to infer the ground truth labels and estimate heterogeneous worker quality from crowdsourced annotations.

In this project, we present a variety of new Bayesian approaches for modeling label dependency and worker quality in the general context of multi-label crowdsourcing. Efficient collapsed and Laplace variational inference algorithms are then developed to jointly infer ground truth labels and worker quality. Extensive simulation and MTurk experiments show that the models based on integrating Bernoulli mixtures and shared structure of worker quality achieve a significant improvement. The proposed approach can also be generalized beyond the MTurk context and has great potential to be applied to a much broader range of domains, such as education and healthcare, in which different opinions need to be combined to measure multiple aspects of an object.

Bio: Junming Yin is an assistant professor in the Department of MIS at Eller College of Management and a faculty member of the Statistics Graduate Interdisciplinary Program at the University of Arizona. He received his Ph.D. in EECS and M.A. in Statistics from the University of California Berkeley. His research interests focus on statistical learning and its applications in business, such as information systems, FinTech, and marketing. His works have been published in top-tier journals and conferences of various fields, including ISR, JMLR, Bioinformatics, NeurIPS, ICML, and UAI. He is a recipient of the Best Paper Award at WITS' 18, Amazon AWS Machine Learning Research Award, Adobe Digital Experience Research Award, Ray and Stephanie Lane Fellowship from Carnegie Mellon University, and Max Planck Society Fellowship.