The Hong Kong Polytechnic University
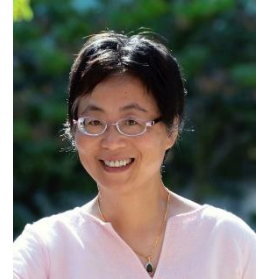## Department of Applied Mathematics

AMA Distinguished Seminar Series in Data Science and Machine Learning

Query-augmented Active Metric Learning

By

Prof. Peiyong QU
University of California, Irvine

Abstract

We propose an active metric learning method for clustering with pairwise constraints. The proposed method actively queries the label of informative instance pairs, while estimating underlying metrics by incorporating unlabeled instance pairs, which leads to a more accurate and efficient clustering process. In particular, we augment the queried constraints by generating more pairwise labels to provide additional information in learning a metric to enhance clustering performance. Furthermore, we increase the robustness of metric learning by updating the learned metric sequentially and penalizing the irrelevant features adaptively. Specifically, we propose a new active query strategy that evaluates the information gain of instance pairs more accurately by incorporating the neighborhood structure, which improves clustering efficiency without extra labeling cost. In theory, we provide a tighter error bound of the proposed metric learning method utilizing augmented queries compared with methods using existing constraints only. Furthermore, we also investigate the improvement using the active query strategy instead of random selection. Numerical studies on simulation settings and real datasets indicate that the proposed method is especially advantageous when the signal-to-noise ratio between significant features and irrelevant features is low.

Biography

Dr. Annie Qu is Chancellor's Professor, Department of Statistics, University of California Irvine. She received her Ph.D. degree in Statistics from the Pennsylvania State University. 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 joins the UC Irvine, Dr. Qu is 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: 30 May 2022 (Monday)
Time: 10:00-11:00 (Hong Kong Standard Time GMT +8)
Venue: Online Talk via Zoom (Meeting ID: 916 8009 1418)
Speaker: Prof. Peiyong Qu, University of California, Irvine
Host: Prof. Xingqiu Zhao, The Hong Kong Polytechnic University
Click to join:
https://polyu.zoom.us/j/91680091418?pwd=QmZuUHU5RytuU1Q0VEpyUzh3bXVkZz09

Click to join (Zoom)

* * * ALL ARE WELCOME * * *
For enrolment, please send your name and email to wai-yan.moon@polyu.edu.hk on or before 29 May 2022