

Machine Learning for Precision Medicine

Date: 16 November 2022 (Wednesday)

Time: 10:00am - 11:00am

Seminar link: <https://cityu.zoom.us/j/97509153818>



ABSTRACT

The vision of precision medicine is to diagnose patients more accurately and treat them more effectively taking into account their individual genomic, life-style, and environmental factors. Machine learning is expected to play a major role in implementing this vision, using the capability of machine learning methods to learn from very complex training datasets and to produce consistent, repeatable predictions (diagnoses, prognoses, treatment recommendations) for test cases.

In this talk, I will focus on three of our own works in this area. Machine learning models typically exploit strong correlations between input and output variables but scientists want to discover causal mechanisms. We have explored the task of discovering causal relationships from observational data, employing data mining methods to generate candidate causal relationships and adopting quasi-experimental design to test the significance of the candidates. Our method HUME discovers single causes in the application domain of pharmacogenomics, using network analysis for candidate generation. Biomedical datasets tend to be small and high-dimensional. Fortunately, related public datasets are available, and transfer learning is a promising approach to increase the effective dataset size. We have used drug response prediction as our driving application, where we want to transfer from large pre-clinical datasets to small clinical datasets. We have proposed AITL, which adjusts for the discrepancies in both the input and the output space, employing adversarial domain adaptation and multi-task learning. Machine learning models for precision medicine need to be explainable. Our method DBKANN adopts a knowledge-based approach that employs the available biological knowledge on how proteins form complexes and act together in pathways to form the architecture of a deep neural network. BDKANN does not only achieve high accuracy but also enables meaningful explanations of individual predictions and the discovery of novel connections in the biological network.

In the last part of the talk, I will introduce single-cell RNA sequencing, a new technology which creates gene expression profiles at single-cell resolution, and enables new types of deeper biomedical analyses such as the discovery of cell-types and their relevance for phenotype prediction. I will sketch some of our ongoing work in the area of single cell data science.



Prof Martin ESTER GUEST SPEAKER'S PROFILE

Martin Ester received a PhD in Computer Science from ETH Zurich, Switzerland, in 1989. He has been working for Swissair developing expert systems before he joined University of Munich as an Assistant Professor in 1993. Since November 2001, he has first been an Associate Professor and now a Full Professor at the School of Computing Science of Simon Fraser University. From May 2010 to April 2015, he has served as the School Director. Dr Ester has published extensively in the top conferences and journals of his field such as ACM SIGKDD, WWW, ACM RecSys, ISMB, and PSB. According to Google Scholar, his publications have received more than 51,000 citations, and his h-index is 69. He received the KDD 2014 Test of Time Award for his paper on DBSCAN, was elected as a Fellow of the Royal Society of Canada in 2019, and was appointed Distinguished Professor at Simon Fraser University in 2021. Martin Ester's research interests are in the area of data mining and machine learning, with a current focus on transfer learning, causal discovery and inference, explainable machine learning, and clustering. Many of the driving applications of his research are in the biomedical field, and he has an honorary appointment as Senior Research Scientist at the Vancouver Prostate Center.