香港浸會大學
HONG KONG BAPTIST UNIVERSITY

DEPARTMENT OF COMPUTER SCIENCE
計算機科學系

35 YEARS OF EXCELLENCE

DEPARTMENT OF COMPUTER SCIENCE

SEMINAR

2023 SERIES

Model-based Deep Embedding for the Analysis of Single-cell RNA Sequencing Data

DATE & TIME

28 NOV 2023 (TUE) 2:30 PM – 3:30 PM

VENUE

Mr. and Mrs. Lee Siu Lun Lecture Theatre (WLB205),
The Wing Lung Bank Building for Business Studies, Shaw Campus

PROF. ZHI WEI

Professor
Department of Computer Science
Ying Wu College of Computing
New Jersey Institute of Technology

ABSTRACT

Methods to sequence the DNA and RNA of single cells are poised to transform many areas of biology and medicine. Single-cell RNA sequencing (scRNA-seq) promises to provide high resolution of cellular differences. Clustering transcriptomes profiled by scRNA-seq has been routinely conducted to reveal cell heterogeneity and diversity, followed by differential expression (DE) analysis to identify marker genes accounting for cellular differences. However, the analysis of scRNA-seq data remains a statistical and computational challenge, due to the pervasive dropout events, which obscure the high-dimensional data matrix with many 'false' zero count observations. Most existing methods focus on dimension reduction, often followed by simple clustering using, for example, k-means. Such a divided strategy is suboptimal for clustering, as we demonstrate. Furthermore, subsequent differential expression analysis after clustering incurs the so-called "double use of data" problem, which will compromise type 1 error control for standard statistical tests. In this talk, I will introduce model-based deep autoencoders to address these issues. The proposed approaches leverage the most recent developments in feature representation learning in deep learning and feature selection in statistical learning, as well as prior information from domain scientists. This development provides a good example of how computer science, statistics, and domain science integrate to yield the optimal solution. Extensive experiments on both simulated and real datasets demonstrate that the proposed methods can boost clustering performance significantly while effectively filtering out most irrelevant genes. Our methods can generate more biologically meaningful clusters, enhancing interpretability as desired by biologists.

SPEAKER'S BIOGRAPHY

REGISTER NOW

Enquiries: 3411-2385   Email: comp@comp.hkbu.edu.hk   Website: https://bit.ly/bucs-events