The Hong Kong University of Science and Technology

Department of Mathematics

PhD THESIS EXAMINATION

Computational and Statistical Methods for Data Integration and Causal Inference

By

Miss Jia ZHAO

ABSTRACT

Data integration and causal inference are two important tasks for effective utilization of big data. Data integration is especially critical in biological studies. In single-cell studies, there is a pressing need for data integration methods to assemble numerous datasets originating from multiple sources into a single comprehensive cell atlas. However, the task of single-cell data integration poses challenges due to the heterogeneity of diverse data sources and the extremely large scale of datasets to be integrated. Causal inference, which aims to infer the causal relationship between a risk factor and an outcome of interest, is also essential in biomedical research. In recent years, Mendelian randomization (MR) has gained increasing attention, as it can take GWAS summary statistics to perform causal inference. However, existing MR analysis often rely on strong assumptions that are often not satisfied in practice.

In this thesis, we propose two computational and statistical methods to address the above challenges in single-cell data integration and MR for causal inference. For comprehensive integration of atlas-level single-cell datasets, we propose a deep learning-based method named Portal. Viewing datasets from different studies as distinct domains with domain-specific effects, Portal achieves data integration through a unified framework of uniquely designed domain translation networks. Through experiments conducted using heterogeneous collections of atlas-level single-cell datasets, we show Portal's superior accuracy and computational efficiency compared to other state-of-the-art single-cell integration algorithms. For inferring causal relationships among traits, we propose a statistical method named MR-APSS. MR-APSS relaxes strong MR assumptions by accounting for two major confounding factors, pleiotropy and sample structure, simultaneously. We validate MR-APSS using comprehensive simulations and negative controls, and apply MR-APSS to study the causal relationships among a collection of diverse complex traits.

Date : **19 July 2023, Wednesday**

Time : **10:00 a.m.**

Venue : **Room 4472 (Lift 25/26)**

Zoom ID: **922 1092 6449 (passcode: 867559)** *~ EE opted via online mode.*
https://hkust.zoom.us/j/92210926449

***Thesis Examination Committee*:**

Chairman	**:**	**Prof. Yong HUANG, CHEM/HKUST**
Thesis Supervisor	**:**	**Prof. Can YANG, MATH/HKUST**
Member	**:**	**Prof. Xinzhou GUO, MATH/HKUST**
Member	**:**	**Prof. Ke WANG, MATH/HKUST**
Member	**:**	**Prof. Angela Ruohao WU, LIFS/HKUST**
External Examiner	**:**	**Prof. Shuqin ZHANG,** **School of Mathematical Sciences/ Fudan University**

(Open to all faculty and students)

The student's thesis is now being displayed on the reception counter in the General Administration Office (Room 3461).