

Dongrong Li

Tel: (+852)96790474

Email: DongrongLi@link.cuhk.edu.hk

EDUCATION

Chinese University of Hong Kong

Doctor of Philosophy in Statistics

September 2020 --- June 2024 (Expected)

- **Selected Courses:** Advanced Statistical Inference, Bayesian Methods in Computational Biology, Advanced Statistical Computing, Statistical Learning Theory
- **Research Interest:** Statistical Machine Learning, Parallel Computing, Scalable Bayesian Inference, Medical Data Analysis, Sequential Data Analysis, Missing Data Problem.

Sun Yat-sen University

Bachelor of Statistics

August 2016 --- April 2019

- GPA: 3.8/5.00
- **Selected Courses:** C++ Programming, Data Structure, Multivariate statistics, numerical analysis, Probability, Mathematical Statistics, Mathematical Analysis,

University of California, Berkeley

As an exchange student

August 2018 --- December 2018

- **Selected Courses:** Linear Modelling(A+), Stochastic Process(B+), Abstract Algebra(A)

TEACHING EXPERIENCE

STAT 3210 Statistical Techniques in Life Science

- Worked as the teaching assistant, designed & prepared tutorial sessions
- Delivered tutorials related to computing in statistics and their application to life science

RESEARCH EXPERIENCE

Visiting Student,

Shenzhen Institute of Advanced Technology, Chinese Academy of Science

July 2019 --- September 2019

- Joined a research group focusing on developing ectoskeleton (a wearable robot that helps paralyzed patients to walk).
- Studied several advanced models in Graph Neural Networks, including Graph Attention Networks(GATs) and Graph Convolutional Networks.

- Helped evaluate and analyze the time complexity of different neural networks on graph. The result of the analysis matched the outcomes of the experiment.
- Analyzed the data acquired from patients. Used clustering analysis to label the walking gestures(which is an unsupervised problem).

Research Assistant

Peking University, Chinese Biographical Database Project (CBDB Project)

July 2020 --- July 2021

- The CBDB Project is supported by Harvard University and Peking University. The aim of this project is to conduct machine learning research on ancient textual data (mainly sequences of biographical data).
- Used deep Language models (including LSTM and BERT) to complete text mining and transfer learning tasks on ancient biographies.
- Conducted Entity Recognition (ER) task on ancient biographical data. Developed language models that could catch high-order contextual features and tag corresponding entities according to partially tagged ancient texts. The tagging results achieved a high accuracy rate according to the evaluation of experts.
- Also helped construct the data analysis and preprocessing pipeline for the project.

Research Project: Modelling Disease Progression on Partially Observed Medical Data

Chinese University of Hong Kong, Department of Statistics

July 2021 --- Now

- Electronic Medical Record (EMR) are sequential health records of patients. We wished to detect the trace of disease progression according to EMR data. However, the EMR data for patients are only partially observed and contain missing data because they visit doctors irregularly.
- Built latent variable models that could reconstruct the disease progression trace under high missingness. Scalable and parallelizable algorithms were developed. Probabilistic bounds on trace reconstruction error were derived to ensure medical reliability.
- The model is currently applied to EMR data from CUHK Medical Center to detect the development and progression from cirrhosis to hepatic carcinoma.

Research Project: Scalable Bayesian Inference with Latent Variables

Chinese University of Hong Kong, Department of Statistics

April 2022 --- Now

- Bayesian Models with latent variables (such as mixture model, missing data model and generative model) are of great interest. Though scalability is a desired property in big data era, latent variable models cannot be easily scaled to large datasets.
- Developed mini-batch Markov Chain Monte Carlo (MCMC) algorithms that scales well on large-scale Bayesian latent variable models with convergence guarantee.
- The algorithm can be applied to advanced models with large training dataset and complicated structures, including Latent Dirichlet Allocation, Large-scale Matrix Factorization and Deep Generative Model.

Leadership & Honors

- **S Prize in the Mathematical Contest in Modelling, 2018**
- **Scholarship for Excellent Undergraduates in College of Mathematics**
- **Captain of the debate team of the Department of Mathematics**

SKILLS AND INTEREST

- TOEFL Test: 110/120
- Computer Programming and Software Skills: Advanced Proficiency in Matlab, C++, Python and R, Basic Proficiency in Latex.
- Interest: Computer Programming, Tennis and reading.