Shiyu WANG

shiyu.wang@emory.edu / Tel: (203)-390-1645

EDUCATION

Department of Biostatistics and Bioinformatics, Emory University, Atlanta, USA Aug 2019 – May 2022

➢ **PhD in Biostatistics**

➢ Top performer (**first place**) in the second-year theory qualifying exam covering all PhD-level core courses

➢ Core courses: Algorithms (graduate level)/Artificial Intelligence/Probability Theory II/Statistical Inference II/Numerical Analysis/Advanced Statistical Computing/ Advanced Linear Models/Generalized Linear Models

Department of Biostatistics, Yale University, New Haven, USA Aug 2017 - May 2019

➢ **M.S. in Biostatistics**

➢ Core courses: Data Mining and Machine Learning/Optimization Techniques/Data structure and programming techniques/Stochastic Processes/Computational Statistics/Nonparametric Statistical Methods & Applications

School of Pharmacy, Fudan University, Shanghai, China Sep 2013 - June 2017

➢ **B.S. in Pharmaceutical Sciences**

TECHNICAL SKILLS

➢ **Programming languages:** Python, C++, Prolog, Scheme

➢ **Tools:** Shell script (PowerShell)

➢ **Data analysis:** R, SAS, MATLAB

PROFESSIONAL EXPERIENCES

Importance sampling-based parameter estimation | Emory University | Research Assistan Feb 2020 – present

Advisor: **Zhaohui "Steve" Qin**, Professor of Biostatistics

➢ Simulated 500,000 sequencing reads using Python (*NumPy*) for modeling PCR and alignment process

➢ Conducted Poisson regression using R to explore relationship between distance and read counts among molecules in cells

➢ Employed C++ to generate initial 2D molecular coordinates by sequential importance sampling (SIS), improve SIS efficiency by rejection control sampling, update nuisance parameters by Gibbs sampling and recover 2D cellular structure by hybrid Monte Carlo algorithm

Numerical analysis-based regression | Emory University | Research Assistant Feb 2020 – present

Advisor: **Zhaohui "Steve" Qin**, Professor of Biostatistics

➢ Applied numerical-analysis techniques, such as LU decomposition, using R to calculate effects of 50,000 genes on variants

➢ Identified trait-associated pathways by permutation-based gene set enrichment analysis and logistic regression

➢ Compared results against benign group by formulating a chi-square test to reduce FDR

Network analysis on high-throughput data | Yale University | Research Assistant Feb – June 2019

Advisor: **Zuoheng Wang**, Associate Professor of Biostatistics

➢ Built hierarchical clustering trees (*hclust( )* in R) and conducted soft-thresholding power picking (*WGCNA* in R) at each cut height to detect outliers

➢ Conducted weighted correlation network analysis (*WGCNA* in R) and identified modules related with various clinical traits

➢ Performed permutation-based gene set enrichment analysis to explore trait-related pathways

Hidden-Markov-Model Based Topic Model in Keywords Extraction | Yale University | Course project June – July 2018

Course: Stochastic Processes

➢ Formed a Latent Dirichlet Allocation (LDA) model (*tidytext, topicmodels, hunspell, tm* in R) for keywords extraction

➢ Derived and applied a Hidden Markov Model to accommodate hidden topics and extracted keywords using LDA model

➢ Wrote Expectation–Maximization algorithm in R for optimization to update parameters in LDA and achieved better performance than traditional LDA model

Dimension reduction and regression | Yale University | Research Assistant June – Aug 2018

Advisor: **Hongyu Zhao**, Department Chair, Professor of Biostatistics, Professor of Statistics and Data Science

- Formed a linear model with clumping using PLINK (executed on PowerShell) to find significant SNPs related to phenotype
- Conducted PCA in R to capture variations introduced by population stratification
- Formalized a binomial test in R for SNPs without homozygous minor alleles to test Hardy-Weinberg Equilibrium

Immunoassay Data Analysis | WEHI | Research Assistant   Aug - Sep 2016

Advisor: **Terry Speed**, Professor in Bioinformatics, WEHI

- Used R to construct 4- and 5-parameter-logistic models (*nplr* in R) to predict concentrations of cytokines
- Conducted censoring, Mann-Whitney test and Gehan test (*npsm* in R), employed a simulation to test the hypothesis that systemic lupus erythematosus associated with variations of cytokine IL-3 in patients

PUBLICATIONS

- Norah P. Saarman[1], **Shiyu Wang**, Mo Li, Richard Echodu, Elizabeth A. Opiyo, Hongyu Zhao, Serap Aksoy and Adalgisa Caccon. "Genome wide association of trypanosome infection in the tsetse fly (Glossina fuscipes fuscipes) in northern Uganda." Working Paper
- He, Xingchao, **Shiyu Wang**, Jiayi Shi, Zhonglin Sun, Zhentian Lei, Zili Yin, Zigang Qian, Huiru Tang, and Hui Xie. "Genotypic and environmental effects on the volatile chemotype of Valeriana jatamansi Jones." Frontiers in plant science 9 (2018). (**First Author**)
- Pierre Sabatier, Amir Ata Saei, **Shiyu Wang**, Roman A. Zubarev. "Dynamic proteomics reveals high plasticity of cellular proteome: growth‐related and drug‐induced changes in cancer cells are comparable." Proteomics (2018): 1800118.

AWARDS

- Top performer in the second-year theory qualifying exam, Department of Biostatistics, Emory University   2020
- Star Graduate (nominated, **top 20/3000**+ undergraduate, **highest honor for the class of 2017**)   2017
- Outstanding Graduate Award, Shanghai City   2017
- CV Starr Scholarship   2014