YUFANG LIU

+01 7627282278 | yl13924@uga.edu

EDUCATION

Southern University of Science and Technology (SUSTech) Shenzhen, China
Bachelor of Science in Statistics Aug. 2018- June 2022
- Cumulative GPA: 3.7 / 4.0
- 1st year GPA: 3.49/4.0; 2nd year GPA: 3.77/4.0; 3rd year GPA: 3.84/4.0; 4th year GPA: 3.9/4.0

National University of Singapore (NUS) Singapore
Data Story App Development with R, Grade: A May 2021-Aug.2021

University of Georgia (UGA) Athens, United States of America
PhD student in Statistics Aug. 2022- present
- Major GPA: 3.9 / 4.0

RESEARCH

Multimodality Integration Analysis for Spatial Transcriptomics Data Using Partial Linear Regression March 2024-Present
- Literature review: Address the challenges of integrating spatial information in the analysis of spatial transcriptomics datasets.
- Method derivation: Apply partial linear regression to each ADT to model the relationships between ADTs, RNA, and the spatial locations of cells. This involves adding a thin plate spline term to the spatial coordinates to extract spatial features of each cell.
- Interpretability: Quantify the contributions of RNA, ADTs, and spatial information to cell type clustering through an analysis of variance of the partial linear model.

Extracting True Virus SERS (Surface-enhanced Raman Spectroscopy): Concentration-Dependent Measurements and Data Augmentation for Species Classification and Quantification Sep. 2023-Present
- Literature review: Recall the importance of using virus SERS spectra for infectious disease diagnostics, and challenges that no true virus spectra could be measured, as well as dilution cause both the virus concentration and medium concentration to change simultaneously.
- Method derivation: Assume that the measured mixture spectra of pure virus and background signals is the linear combination of targeted pure virus spectra and measured background signals spectra. Construct two fully connected neural networks (FCNNs) with linear combination assumptions to extract virus spectra; repeat the same procedure for a total of 13 different types of viruses and result in 13 pure virus spectra
- Downstream analysis: Perform data augmentation skills to generate 65,000 mixture spectra using 13 extracted pure virus spectra; build XGboost model for classification of concentration as well as virus types based on augmented mixture spectra; construct FCNNs regression model in which concentration is response and augmented mixture spectra are predictors

Orthogonal Multimodality Integration and Clustering in Multimodal Single-cell Data Sep. 2022-March 2024
- Literature review: Recall the challenges of interpretability of critical predictive features, quantification of prediction power and computational speed in multimodality integration of single-cell analysis
- Method derivation: Normalize and scale the feature RNA and ADT features; perform ADT projections on RNA and Orthogonal RNA space; cluster cell types based on RNA and ADT projections; explore substantial predictive power of cell types by identifying differentially expressed RNAs and ADTs
- Performance on CITE-Seq Dataset: Compare the Adjusted Rand Index value of the clustering cell types of our method to WNN method and MOFA+ method; show that our method can pertain a high accuracy while spend less time in computing
- Manuscript draft: Accepted by 'BMC Bioinformatics'

Response-adaptive Treatment Randomization for Clinical Trials with Survival Outcome June 2021-June 2022
- Literature review: Recall the ethnic and hypothesis testing objective of clinical trials
- Method derivation: Assume Piecewise Constant Hazard Function which generalize traditional constant hazard or monotone hazard function and weaken distribution assumptions on survival data
- Extensive simulation study: use numeric method to obtain parameter & variance estimator from partial likelihood function; solve the adaptive optimal allocation and apply Doubly adaptive Biased Coin Design procedure to realize allocation process
- Won Best Research Paper Award of SUSTech

Restaurants' Levels Prediction, NUS Summer Workshop May 2021-July 2021
- Teamed with 2 classmates to collect 60,000 sets of data regarding Restaurants' location, menu and other information in Bangalore, India, from Kaggle
- Applied ggplot to draw the restaurants heatmap on the google map; represented average consumption level using different shade; cleanse and filter data with KNN; constructed logistic regression model, random forest model, and advanced multi-logit model with R to classify restaurants
- Predicted the level of restaurants; developed an application with R to check restaurants' location, customer feedback, ratings and provide suggestions on location, menu to open a restaurant in India
- Won Best Teamwork Award of 150 Singapore Dollar and A performance

PROJECT

Survey on SUSTech Students' Online Learning Situation April 2021
- Designed questionnaire to explore the influence of online learning ability, and the relationship between gender, years at university, majors and the preference of online learning

- Emailed questionnaire to 1,000 students and collected data of 190 students
- Applied descriptive analysis, linear regression and clustering to find influential factors on online learning preference; summarized the learning ability and online learning preference distribution among undergraduates with different majors and genders
- Utilized Mann Whitney U test and comparison test to check the difference of online learning ability among undergraduates; plotted the density function of online learning ability in genders; found male students have a larger variance than females

Analysis on Symptomatic Factors of Stroke April 2021

- Explored the relationship of the stroke and symptomatic factors with 3000 sets of individual data from kaggle; made suggestions to avoid the stroke
- Made descriptive analysis on 11 variables in the data; applied logistic regression using SAS
- Constructed and compared 4 advanced models- GLM, LDA, Random Forests and Naïve Bayes using python; found that GLM has the best accuracy and Naïve Bayes has the best specificity of the prediction

Analysis of Housing Price in Shenzhen with R Dec. 2020

- Collected the data of apartments in Xiamen, Guangzhou, Shenzhen about the administrative district, population, area, floors, facing direction, traffic convenience and nearby schools
- Cleansed data; generated dynamic bubble chart of housing prices changes with population and area from 1980 to 2010 using R
- Used linear regression and ANOVA to find that district, construction year, nearby schools and subway stations had significantly influential factors on apartment prices, and Shenzhen had the biggest price differences between apartments in city center and suburban areas

ACTIVITIES

Volunteer, SUSTech Sports Meeting 2 Open Days Nov. 2018
Member of Shuren College Students' Union, SUSTech Sep. 2018- Sep.2020
Member, Soccer Team of Shuren College, SUSTech Sep. 2018- June 2022

- Won championship of 6 competitions on campus

Poster Presentation, Georgia Statistics Day, Georgia Institute of Technology Oct. 2023

- Present a poster on the topic: 'Spectra Decomposition in Surface-enhanced Raman Spectroscopy (SERS): Extract Pure Spectra Using Deep Learning Method'

HONORS & AWARDS

Selected students (5%) in 2 Honor classes (Calculus & Advanced Algebra)
Mathematical department of SUSTech, Spring 2019
Excellent Student Scholarship (3rd Prize) of SUSTech (20%),Oct. 2020, Oct. 2021
Best Research Paper Award of SUSTech (10%) for 'Response-adaptive Treatment Randomization for Clinical Trials with Survival Outcome', June 2022
Honorable Mention: Student Employee of the Year team award of University of Georgia, March 2024

Publications

Yufang Liu, Yongkai Chen, Haoran Lu, Wenxuan Zhong, Guocheng Yuan & Ping Ma. **Orthogonal multimodality integration and clustering in single-cell data**. *BMC Bioinformatics* 25, 164 (2024). https://doi.org/10.1186/s12859-024-05773-y

Susan Wu, Xilin Gong, Yongkai Chen, **Yufang Liu**, Wenxuan Zhong, Ping Ma. **Fisher Contrastive Learning: A Robust Solution to the Feature Suppression Effect**. Open Review.

SKILLS

R, JAVA, Python, SAS