

Baode GAO

(2/23/2024)

7900 Cambridge, Houston, TX 77054

(+1)917-470-7259 bg2715@caa.columbia.edu

EDUCATION

School of Public Health, The University of Texas Health Science Center at Houston, US Aug. 2023 – May 2028

- Ph.D. Candidate in Biostatistics and Data Science
- 8 years of programming experience: solid in R lme4, tidyverse, caret, NetworkToolbox, etc., and Python Pandas, NumPy, PyTorch & Keras frameworks, Linux, able to quickly start projects

Mailman School of Public Health, Columbia University in the City of New York, US Sept. 2021 – May 2023

- Master's in Biostatistics (Theory and Methods Track)
- Worked as a paid Student Research Worker
- Core Courses: Statistical Inference / Topics in Statistical Learning and Data Mining / Causal Inference / Deep Learning

Department of Mathematical Science, Xi'an Jiaotong-Liverpool University, China Sept. 2016 – July. 2020

- Bachelor's in Applied Mathematics
- Contributed an article (as the first author) to ICMLC
- Core Courses: Optimization Theory / Analysis I & II / Complex Analysis / Partial Differential Equations

RESEARCH EXPERIENCES

R Package Developing in Network Analysis Aug. 2023 – Mar. 2024

UTH | School of Public Health | Graduate Research Assistant

Advisor: Kayo Fujimoto, Professor and Distinguished Professorship in Social Determinants of Health, UTH

- Coded for exposure measure in Network Exposure, Affiliation Network Exposure and Decomposed Network Exposure in R.
- Preprocessed to linear network autocorrelation models and reproduced VF-brokerage, EV-brokerage in R.
- Developed R code for susceptibility, influence and integrated the above functions into a R package.

LASSO improves the true detection rate of connectome-based predictive modeling June 2022 – Dec. 2022

Columbia University | School of Nursing | Student Research Worker

Advisor: Yihong Zhao, Professor of Data Sciences at Columbia University Medical Center, Columbia University

- Discovered the deficiency of the Connectome-based Prediction Model by the simulation study of fMRI data.
- Established penalized LASSO by features in networks to select the independent variable.
- Achieved higher true detection rate and prediction accuracy by variable selection methods in simulation study.

Covid-19 Infection Percentage Estimation by Transfer Learning Mar. 2022 – May 2022

Columbia University | The Fu Foundation School of Engineering and Applied Science | Research Assistant

Advisor: Iddo Drori, Adjunct Associate Professor at the Department of Computer Science, Columbia University

- Split CT scans by subjects to avoid data leakage and enhanced model robustness by data augmentation – horizontal flip, rotation, random cropping, etc.
- Employed 4-fold cross-validation to adapt the uneven positive and negative covid images and small dataset.
- Reduced bias by halving the learning rate for every five epochs.
- Improved training speed and accuracy of results (MAE: 6.938) by establishing the pre-trained ResNet50 and freezing parameters of the feature extraction part during training.

Subgroup Analysis Under Cure Rate Model June 2019 – Sept. 2019

XJTLU | Department of Mathematical Science | Research Assistant

Advisor: Xiaojun Zhu, Assistant Professor at the Department of Mathematical Science, XJTLU

- Deduced the expectations of latent variables, the approximation of the marginal likelihood, and the estimation of the baseline survival function of the subgroup cure rate model.
- Employed cox proportional-hazards model and logistic model to obtain the initial value of parameters in the survival function, cure rate function, and grouping function.
- Calculated expectations of hidden variables and estimated baseline survival function in E-step and estimated parameters by maximum likelihood estimation of the marginal likelihood in M-step.
- Compared our approach (AIC 1083.046) with Peng's Cure Rate Model (AIC 1090.103) (Likelihood Ratio Test,

ratio 13.057; critical value 7.81; p-value 0.0045).

Learning Linear Mixed Model for Group Recommendation System

June 2018 – Nov. 2018

XJTLU | Department of Mathematical Science | Research Assistant

Advisor: Shengxin Zhu, Associate Professor at Department of Mathematical Science, Beijing Normal University

- Estimated the coefficients implying the latent association between items' attributes and users' characteristics.
- Verified the choice of fixed effects and random effects with ANOVA test and compared nested models.
- Achieved better prediction accuracy compared with Fusion Heterogeneous Information Network and Rating Matrix Reconciliation Method.
- Contributed a first-author paper to the International Conference on Machine Learning and Computing 2019.

PUBLICATIONS

- **Gao, B.**, Zhan, G., Wang, H., Wang, Y., & Zhu, S. (2019, February). Learning with Linear Mixed Model for Group Recommendation Systems. In *Proceedings of the 2019 11th International Conference on Machine Learning and Computing* (pp. 81-85). ACM.
- Kayo Fujimoto, Lizhong Liu, Jacky Kuo, **Baode Gao**, Michelle Carr, Armand Brown, Xi Luo, Justin Bahl. Multi-layer Graph Attention Fusion Network for COVID-19 Prediction within Complex Social Interactions. (To be submitted to PLOS complex networks)

WORKING EXPERIENCES

City of Houston | Health Department | Intern | Houston

Aug. 2023 – Nov. 2024

- Preprocessed and linked survey data.
- Developed graph-based machine learning approaches to improve prediction accuracy.
- Employed explainable AI to identify risk factors and provide actionable insights for public health strategies.
- Visualized results in Python.

Irving Medical Center | Nursing | Student Research Worker | New York

June 2022 – Dec. 2022

- Detected defects in existing models.
- Promoted model performances by exploring new structures of model.

AWARDS AND ACHIEVEMENTS

- Honorable title of XJTLU Outstanding Student (2017-18, top 2.5%)
- Conference Travel Fellowship (2019, XJTLU)