Zifan (Fred) Yu

University of Tennessee – Knoxville, TN

📱 +1 2406443948 • ✉ zyu20@uthsc.edu • 🌐 learningMalanya.github.io

Education

PhD Student in Data Science and Engineering — TBD/4.0
University of Tennessee, Knoxville — *2023-Present*

Master of Science in Biostatistics — 3.7/4.0
University of Washington (UW), Seattle — *2020-2022*

Bachelor of Science in Mathematics — 3.8/4.0
University of Maryland (UMD), College Park — *2016-2020*

Research Highlights

Fast applications of linear mixed model for multiple-trait association scans
Work supervised by Saunak Sen, University of Tennessee Health Science Center (UTHSC) — *June 2022 - June 2023*

○ Investigated the classic pipelines for evaluating the linear mixed effects models in the context of genome-wide association studies.

○ Developed new algorithms to perform linear mixed model association scans over many quantitative traits efficiently, where the efficiency is achieved by grouped computations on clusters of traits and vectorized operations for each cluster. Implemented the methods as the software package named *BulkLMM* in the Julia language.

○ Experimented the performances of our proposed methods on the department's high-performance server. On the BXD data for scanning over 32k traits, our methods achieved runtimes 150 to 20k times faster than our main competitor, GEMMA, with insignificant difference in the outputs.

Applications of nowcasting methods to notifiable disease surveillance
Master's capstone project mentored by Ian Painter, Washington Department of Health — *Sep. 2021 - Mar. 2022*

○ Conducted a literature review on the nowcasting methods for disease surveillance.

○ Performed data cleaning to the data of Washington state COVID-19 hospitalization from June 2020 to December 2021; adapted the nowcasting methods to analyze the data and predict hospitalizations.

○ By comparing the predicted count retrospectively with the actual case count, we show that the results we obtained precede the WADOH approach for about eight days to give accurate predictions.

Other Research Experience

Sparse matrix linear models with the Elastic-net regularization
Intership project mentored by Gregory Farage & Saunak Sen, UTHSC — *June 2021 - Sep. 2021*

○ Based on prior work of Liang and Sen, extended the estimation framework of the matrix linear models under the Elastic-net regularization, using proximal gradient methods.

○ Implemented the extended functionalities in the Julia package MatrixLMnet.

Variable Importance for Fixed Effects in Linear Mixed Models
Student paper of WNAR conference 2022, co-authored with Yongzhe Wang & Lingbo Ye, UW — *March 2022*

○ Adapted the Shapley value approach for the study of variable importance of linear mixed model fixed effects.

○ Through simulation, we show that our proposed method of measurement is more robust in capturing the true variable importance under the case of high muti-collinearity, compared to p-values.

Approximation Methods for Large Sparse Precision Matrices
Independent reading project, UW — *March 2022*

○ Reviewed the Blocked Rao-Blackwellized Monte Carlo methods, which approximate the large precision matrices based on the idea of Gaussian-Markov Random Field; studied its algorithmic complexity and compared with the classic approach by the Takahashi equations.

Presentations

Real-time linear mixed models for association studies on multiple quantitative traits
Oral Presentation, Complex Trait Community Annual Meeting 2023, Virtual — *October 11, 2023*

Real-time linear mixed models for association studies on multiple quantitative traits
Invited speaker, UTHSC Biostatistics Seminar Series, Virtual *September 18, 2023*

Real-time linear mixed models for association studies on multiple quantitative traits
Oral presentation, ENAR 2023, Nashville *March 21, 2023*

Applications of nowcasting methods to notifiable disease surveillance
Invited speaker, Washington Department of Health, Virtual *March 23, 2022*

About the Keplan Meier estimator
Contributed talk, UW Biostatistics Survival Analysis Reading Group, Virtual *July 21, 2021*

Honors

Meritorious Winner Award
Mathematical Contest in Modeling (MCM) 2020

Software

- **BulkLMM.jl**: https://github.com/senresearch/BulkLMM.jl
- **MatrixLMnet.jl**: https://github.com/senresearch/MatrixLMnet.jl

Skills

- **Programming:** Julia, Python, R, Matlab, Java, high-performance computing, distributed version control (Git), package development, LaTeX
- **Highlighted coursework:** Statistical inference, machine learning, mathematical modeling, data structures and algorithms, object-oriented programming
- **Others:** effective communication, collaboration in a team, oral presentation, technical writing.