Siyu Wang

5002 Sheboygan Ave, Madison, WI, 53705 | (608) 381-5100 | iamwangsiyu@gmail.com | linkedin.com/in/siyu79/

EDUCATION

UNIVERSITY OF WISCONSIN-MADISON *Madison, WI*
Ph.D. student in Statistics *Expected May 2024*
Master of Science in Statistics - Data Science *May 2019*
Coursework: Experimental Design, Statistical Consulting, Causal Inference
Research area: Decision Tree Algorithms, under the supervision of Professor Wei-Yin Loh

UNIVERSITY OF SCIENCE AND TECHNOLOGY OF CHINA *Hefei, China*
Bachelor of Science in Statistics *May 2018*

PROFESSIONAL EXPERIENCE

Inari Medical *Irvine, CA*
Biostatistics Intern *May - Aug 2022*

- Generated an IBM database to the CDISC ADaM dataset pipeline using R and SAS. Took part in the quality check process and identified data issues when combining health records from CRF in real-world registries.
- Teamed with the Clinical Insights department to evaluate the impact of biomarkers on predicting patients' severity. Applied survival models beyond the Statistical Analysis Plan (SAP) for ad-hoc counterfactual analysis and subgroup identification, which led to a 40% decrease in the hazard ratio of the treatment effect.
- Did several methodological research projects, like investigating the nuance between SAS and R in repeated measurements. Reviewed the literature on marginal homogeneity, made an animation using the manim package in Python, and presented it in the monthly department meeting.

RESEARCH

R Package for an Improved Decision Tree Method (LDATree) *Sep 2020 - Now*

- Build a whole decision tree package from scratch in R through a continuous process of missing value imputation, data structure transformation, cross-validation pruning, and final plotting and prediction.
- Integrate linear discriminant analysis (LDA) into the decision tree structure to achieve better interpretability while remaining fast and accurate. Use statistical tools like the ANOVA F-test and the Chi-Squared test to power up this decision tree and run simulations on real data to compare with existing machine learning methods and revise the algorithm. Test many corner cases to make the program more robust and regularly update the code on GitHub to save the progress

PROJECTS

Estimate Yelp Rating Based on Reviews (In Python) *Feb - Apr 2019*

- Used Aspect Term Extraction (ATE) to find the most important terms that affect the restaurant business, fit the MA model to find the seasonal trend, and used non-parametric hypothesis tests to confirm the results.
- Conducted sentiment analysis on the ratings of more than 100K reviews using the Bag of Words Model and LSTM. Applied the Keras package on the Databricks platform for large scale neural network training.

Bootstrap Calibration for Simultaneous Confidence Interval Construction *Sep - Dec 2019*

- Ran thousands of simulations on High Performance Computing Cluster (HPC) using bash and slurm. Applied the idea of bootstrap to find the real size of the type-I error in a time-varying causal model on Ecological Momentary Assessment (EMA).