Shubo Sun

(814) 852-8517 | sxs3935@miami.edu | personal homepage | github.com/shubosun0113

Education

PhD in Management Science Fall 2022 – Present

University of Miami

- GPA: 4.0
- **Courses Taken**: Generalized Linear Models, Statistical Machine Learning, Advanced Econometrics, Logitudinal and Multilevel Data Analysis

MSc in Statistics Fall 2020 – Fall 2021

University of Illinois at Urbana–Champaign

- GPA: 3.96
- **Courses Taken**: Mathematical Statistics, Computational Statistics, Applied Bayesian Analysis, Stochastic Processes, Advanced Data Analysis, Statistical Consulting, Deep Learning, Advanced Regression

Dual BSc in Mathematics and Statistics Fall 2016 – Fall 2019

The Pennsylvania State University (University Park, PA)

- Minor: Korean
- GPA: 3.71 (Major GPA: 3.85)
- Dean's List: Fall 2016 – Fall 2018
- **Courses Taken**: Discrete Mathematics, Real Analysis, Linear Algebra, Numerical Computations, Calculus I, II, III, Ordinary and Partial Differential Equations, Probability Theory, ANOVA, Time Series, Linear Regression, Game Theory, Mathematical Statistics, Programming in SAS

Working Papers

Sun, S., Zhao, Z., Jiang, F. & Shao, X. (2023+) An R Package for Change-Point Analyses and Time Series Segmentation via Self-Normalization (Currently Working)

Relevant Skills

Programming Skills: R (Expert), Python (Intermediate), SAS (Intermediate), C++ (Beginner)
Writing Skills: LaTeX(Intermediate), Microsoft (Intermediate)
Language Skills: English (Fluent), Chinese (Native), Korean (Intermediate)

Research Experience

R Package SNSeg Compilation and Paper Writing May 2021 – Present

Graduate Research Assistant

- Created functions in R to conduct change-point estimation with Self-Normalization (SN) based algorithms
- Implemented algorithms in C++ and seamlessly integrated them with R through Rcpp
- Included graphics for time series and SN test statistics segmentation plots with examples plots
- Offered *SNSeg*: R Package Manual and academically written manuscripts at github.com/shubosun0113/SNSeg
- **Advisor 1**: Dr. Zifeng Zhao, Department of Business Analytics and Statistics, Notre Dame
- **Advisor 2**: Dr. Xiaofeng Shao, Department of Statistics, UIUC

Investigation on SN for Change-Points Estimation for US COVID-19 Dec 2020 – Feb 2022

Graduate Research Assistant

- **Project 1**: Research on COVID-19 Infection Change-point Analyses for Various Countries Using Piecewise Linear Quantile Regression Model with Self-Normalization Procedures
- **Project 2**: Research on COVID-19 Infection, Death and Recovered Cases Change-point Analyses and Future Value Predictions Using Penalized Piecewise Polynomial Regression Model with Self-Normalization Procedures
- **Advisor**: Dr. Xiaofeng Shao, Department of Statistics, UIUC

Analysis of Effects of Loading on Bone Mechanical Properties and Strain Jan 2021 – May 2021

Statistical Consulting for R in research

- Studied the loads on tibia bone at the normal state during different basketball maneuvers using ANOVA

- Compared the loads at the fatigue and the bracing state by adding a fatigue or bracing factor with the normal state through K-means clustering
- Performed model selections for loads prediction using multiple supervised learning models
- Offered a project report to the client with feasible suggestions on ways to improve athletes' bone qualities

Penn State BOAST R Shiny Program May 2019 – Aug 2019

Undergraduate Research Assistant

- Developed R Shiny Application at sites.psu.edu/shinyapps within a team to supplement undergraduate statistical courses
- Improved the functions and the visualization for previous apps, and created new apps for advanced contents
- Collaborated in group meetings for discussions of apps, and documented codes through GitHub for apps' writing at github.com/EducationShinyAppTeam/12-Stochastic_Processes
- Reported the apps and findings through poster presentation during the Eberly College of Science Exhibition

Teaching Experience

Penn State MATH 251 Ordinal and Partial Differential Equations Jan 2019 – May 2019

Undergraduate Teaching Assistant – Prof. Bharath Narayanan

- Collaborated with the professor to grade assignments, and prepare solutions with feedback
- Held review sessions for students before midterm and final exams

Penn State MATH 230 Calculus and Vector Analysis Aug 2018 – Dec 2018

Undergraduate Teaching Assistant – Prof. Lisa Melanson

- Collaborated with the professor to grade assignments and exams, and prepare solutions with feedback
- Held review sessions for students before midterm and final exams

Penn State STAT 200 Elementary Statistics Aug 2018 – Dec 2018

Undergraduate Lab Assistant – Prof. Cecil Shelton

- Instructed students to complete their lab problems and answered students' questions
- Taught students the basic grammar of R and basic statistical inference (e.g., t-test and linear regression)

Work Experience

HD Education Oct 2019 – Present

Senior Tutor

- Held online public and 1-to-1 courses to teach upper-level undergraduate Math & Statistics courses for Chinese students enrolled in US universities, and helped them to improve their grades in their real courses
- Tutored courses include Introduction to Probability, Calculus, Linear Regression, Stochastic Processes and Mathematical Statistics
- Held review sessions before midterms and final exams
- Helped around 10+ out of 40+ students to get an A in their real courses

Project Experience

Google reCAPTCHAs v2 Solver | *Deep Learning* Oct 2021 – Dec 2021

- Designed a reCAPTCHAs v2 using cv2 template matching, EasyOCR and YOLOv5
- Model architecture: cv2 template matching to pass the "I'm not a robot" checkbox, YOLOv5 to identify 3x3 or 4x4 reCAPTCHAs and object detecions in images, and EasyOCR for text detections to identify the object classes
- Trained the model with over $4,000$ images and achieved a decent mAP of 0.89 for some object classes
- Provided a project report at github.com/shubosun0113/STAT-430

Text CAPTCHAs Solver | *Deep Learning* Oct 2021 – Dec 2021

- Designed a text-based Captcha's solver using a 40 layers' convolutional and recurrent neural network (CRNN)
- Trained the CRNN with $100,000$ generated captchas using CTC loss and ADAM optimizer
- Achieved an 87.5% prediction accuracy on the testing data
- Provided a project report with source codes at github.com/shubosun0113/STAT-430

Analysis of Region Differences of Education Resources in China | *Bayesian Analysis* Mar 2021 – May 2021

- Fitted a Bayesian hierarchical linear model for high school enrollment rate in 6 different regions of China

- Applied MCMC using JAGS to achieve the target distribution of enrollment rates
- Ranked the 6 regions based on posterior probabilities and provided a project report with sources codes

Analysis of Food Nutrient Score in terms of Nutrient Ingredients | *Machine Learning* Oct 2020 – Jan 2021

- Performed dimension reduction using PCA and exploratory factor analysis
- Conducted NLP to categorize the food based on the most frequent 150 additives
- Compared different tuned models using ROC Curve
- Achieved a 84.4% high accuracy on binary classification considering missing data
- Documented codes and the report at github.com/shubosun0113/STAT-542

Wikipedia Page Views Prediction | *Time Series* Oct 2018 – Dec 2018

- Forecasted Wikipedia daily page views for the word "Thanksgiving"
- Developed a SARIMA model based on five years' data in November using AIC and MLE
- Guaranteed an over 80% prediction accuracy in a 30-days period, making our group 1^{st} place among all 23 groups