Xiaoyue Li, Ph.D. Candidate

CONTACT INFORMATION
+1-650-772-2359
xyrli@ucdavis.edu

SUMMARY

Ph.D. candidate in Statistics with strong technical skills and 5+ years of experience in research, consulting and data analysis. Extensive knowledge of statistical modeling (generalized linear models, change-point detection, covariance estimation, computational statistics), machine learning (network modelling, matrix completion, clustering), optimization (SGD, duality, ADMM, parallel optimization).

EDUCATION

University of California - Davis, CA, United States

Ph.D. Statistics, *Expected:* June 2020 (GPA 3.964/4.0)

- Advisor: Prof. James Sharpnack

Hong Kong University of Science and Technology, Hong Kong

BSc. Risk Management and Business Intelligence, 2015 (GPA: 3.752/4.3) (A: 4.0, A+: 4.3)

- Minor in Information Technology
- Minor in Mathematics

WORK EXPERIENCE

Software Engineer Intern Jul 2018 – Present
Search Ads — Google LLC. Mountain View, CA
C++, Python, TensorFlow

- Built binary classification models for user modeling with TensorFlow models including logistic regression, lattice models, and neural networks
- Improved click through rate prediction model evaluation pipeline and enhanced data representation

Research Scientist Intern Apr 2018 – Jun 2018
Smart Supply Chain and Big Data — JD.COM Mountain View, CA
Python, SQL

- Built prototype models for multi-day sales quantity quantile prediction
- Built and explored various models for missing sales quantities imputation. Among these, a matrix completion approach improved the sales quantity quantile prediction accuracy in MAD by around 10 percent

RESEARCH EXPERIENCE

Compression of Spatio-Temporal Networks via Point-to-Point Process Models Jun 2016 – Present
Python, Julia, SQL

- Work published in Proceedings of International Workshop on Mining and Learning with Graphs in SIGKDD, 2017
- Developed a framework to model the stochastic process of spatio-temporal networks
- Applied the framework to NYC taxi dataset to compress the trip demand spatially and temporally:
  - Cleaned, transformed and analyzed large scale datasets with information of more than 100 million taxi trips
  - Estimated a spatially smoothed community structure and localized temporal change-points for the network
  - Derived and implemented an ADMM optimizer to solve a group-fused LASSO program for a penalized M-estimator

- Visualized spatial clustering of 8000 grid areas with GIS shape files with interactive display of community ID with data cursor
- Extending the algorithm with non-negative matrix factorization to solve for MLE with simultaneous estimation of spatial and temporal clusters, and proving consistency result for the estimator

Projects

Extreme Multi-label Classification Nov 2017 – Dec 2017
Matlab

- Derived and implemented an efficient algorithm based on ADMM to predict true labels for new inputs where the number of labels can be extremely large
- Applied to Bibtex dataset with 1836 features and 159 labels and achieved comparable results with state-of-the-art algorithms
- Scaled the algorithm to a much larger dataset, AmazonCat-13K, containing 203882 features and 13330 labels

Taxi Pickups Near Subway Stations: A GLM Approach Jan 2017 – Mar 2017
Python, R, SQL

- Modeled how taxi pickups near subway entrances during rush hours covary with characteristics of the station, trip date, and subway arrival information
- Fitted a generalized additive partial linear model with negative binomial family after diagnostics and model selection
- Processed complex dataset integrated from 3 data sources involving gtfs-realtime, SQL and web scrapping

Other Experiences

Copula Based Modelling in Geostatistics Sep 2013 – Aug 2015
Extreme value theory, Copula, Variogram, R

- Initiated a copula based method to model the dependence structure in spatial data with the presence of extreme values
- Incorporated a skewed version of t-copula to address asymmetric tail dependence
- Studied various advanced topics including Bayesian estimation, geostatistics, multivariate copula, extreme value theory

Teaching Experience

Teaching Assistant, UC Davis Sep 2015 – Present

- Taught 6 different courses including: Data and Web Technologies for Data Analysis (Python for Statistics), ANOVA, Applied Statistics for Biological Sciences etc.
- Responsibilities: preparing materials, leading discussions, holding office hours and managing teamwork
- Vast majority gave 'excellent' or 'very good' ratings
- Positive student feedback on effectiveness in communication, helpfulness in office hours and enthusiasm in teaching