

Qiaohui (Chelsea) Lin

2317 Speedway, Austin, Texas 78712 <https://github.com/qiaohuilin> Email: qiaohui.lin@utexas.edu Cell: 919-433-6310

EDUCATION

University of Texas at Austin Austin, Texas
PhD student in Statistics and Data Sciences Expected Jun., 2022
· GPA: 4.00/4.00
· Relevant Courses: Deep Probabilistic Modeling, Statistic Modeling in Big Data, etc.
· Research Interest: Bayesian Nonparametrics, Variational Inference, Network Analysis

Duke University Durham, NC
M.A. in Economics, GPA: 3.80/4.00, MA. Merit Scholar May, 2018
· Relevant Courses: Bayesian Statistics, Machine Learning, Data Mining, Probability/Measure Theory.

Fudan University Shanghai, China
B.A. in Economics, GPA: 3.81/4.00, Outstanding Graduate of Fudan, 2016 Jun., 2016

RESEARCH AND PUBLICATION

Current Projects Austin, Texas
Subsample Network Analysis (Supervised by Purna Sarka)
- Develop consistent variance estimator with a confidence bound for large scale sparse graph in social networks using subsampling methods, which is otherwise either intractable or computationally expensive.

Collaborative Filtering Recommender System Under Bayesian Double Feature Allocation
- In a recommender system, develop a Bayesian nonparametric model, a double feature allocation on both users and items to predict a user's rating/ranking to unseen items. (Supervised by Peter Müller)

Publications

On the Theoretical Properties of the Network Jackknife
ICML 2020, proceedings of international conference on machine learning
- Jackknife estimate method for network variance.

Prediction of Appointment No-shows using Electronic Health Records
Journal of Applied Statistics, In Press 2019
- Bayesian hierarchical model predicting hospital no-shows.

WORK EXPERIENCE

Amazon LLC Seattle, Washington
Applied Scientist Intern Jun., 2020 – Aug. 2020
· Delivery Experience Machine Learning Team in optimizing delivery experience via feature engineering.

Homeaway Inc, Expedia Group Austin, TX
Data Science Intern Jun., 2019 – Aug. 2019
· Search Engine Optimization: Keyword Clustering and Performance Prediction
· Cluster Search Keywords based on NLP methods and use a Gaussian Process for time series analysis of clicks and conversions in each cluster.

Duke Clinical Research Institute Durham, NC
Research Assistant Apr., 2017 – May, 2018
· Paper published on topic of hospital no-shows mentioned above
· Presented on Women in Machine Learning Conference 2017

SKILLS

Coding: Proficient in R, Python and Git;
Language: English, Chinese