Yunyi Li

+1 (319)-519-3209   Yunyi.Li@mccombs.utexas.edu

Education

The University of Texas at Austin, Austin, TX, USA   2020 - 2025 (Expected)
- Ph.D. in Information, Risk, and Operations Management

The University of Iowa, Iowa City, IA, USA   *2018 - 2020*
- M.S. in Informatics-Information Science

The University of Iowa, Iowa City, IA, USA   *2015 - 2018*
- B.B.A. in Business Analytics with **Honors**
  - Large Data Analysis Certificate

Selected Publications

1. **Li, Y.**, De-Arteaga, M., and Saar-Tsechansky, M., "More Data Can Lead Us Astray: Active Data Acquisition in the Presence of Label Bias." Proceedings of the Tenth AAAI Conference on Human Computation and Crowdsourcing (HCOMP), 2022.

2. Wang, T., Yang, J., **Li Y.**, and Wang, B., "Partially interpretable estimators (PIE): black-box-refined interpretable machine learning." Submitted, 2022.

3. Srivastava, S., Xu Z., **Li Y.**, Street, W. N., and Gilbertson-White S., "Gaussian Process Regression and Classification using International Classification of Disease Codes as Covariates." Submitted, 2021.

4. Gilbertson-White S., Srivastava, S., **Li, Y.**, Laures, E., Saeidzadeh, S., Yeung, C., and Chae, S., "Multimorbidity, Cancer, and Symptoms: Using Electronic Health Record Data to Cluster Patients in Multimorbidity Phenotypes." Journal of Pain and Symptom Management (JPSM), 2019.

5. Wang, Y., Wang A., Liu, Z., Thurman, A., Powers, L., Zou M., Hefel, A., **Li, Y.**, Zabner, J., and Au, K.F., "Single-molecule Long-read Sequencing Reveals the Chromatin Basis of Gene Expression." Genome Research, 2019.

6. **Li, Y.**, and Wang, T., "Next Hit Predictor - Self-exciting Risk Modeling for Predicting Next Locations of Serial Crimes." AI for Social Good NeurIPS Workshop, 2018.

Research Experiences

McCombs School of Business, University of Texas at Austin: Research Assistant   *Aug 2020 – Present*
Topic: *More Data Can Lead Us Astray: Active Data Acquisition in the Presence of Label Bias*
- Identified different types of label bias in the context of supervised learning systems
- Conducted empirical experiments using five active data acquisition algorithms for income prediction and offensive language detection
- Found three *more data can lead us astray* patterns

Department of Statistics, University of Iowa: Research Assistant   *Dec 2018 – May 2020*
Topic: *Statistical learning the Electronic Health Record (EHR) to Cluster Patients with Cancer Based on Multimorbidity Profiles*
- Identified the most common clusters of chronic conditions co-occurring in patients with advanced solid tumor cancer and describing differences across these clusters
- Designed a kernel distance function for cancer patients with Electronic Health Record
- Employed advanced statistical learning such as Gaussian Process and Gibbs Sampling to solve regression and classification problems of cancer patients

Carver College of Medicine, University of Iowa: Research Internship   *May 2018 – Sep 2018*
Topic: *Single-molecule Long-read Sequencing Reveals the Chromatin Basis of Gene Expression*
- Created visualizations using R for Chromatin accessibility and Nuclearsome positioning
- Drew Rcircos Landscape to display multiple features in a whole genome scale
- Applied Hidden Markov Model and Dynamic Programming to analyze the signal distribution
- Actively participated in the project group meetings by brainstorming to solve problems, and helped with quality control

Tippie College of Business, University of Iowa: Research Assistant   *Dec 2016 – Jul 2018*
Topic: *Next Hit Predictor - Self-exciting Risk Modeling for Predicting Next Locations of Serial Crimes*
- Built a novel spatial-temporal crime prediction model for Cambridge, MA, police department
- Formulated a convex learning objective considering pairwise rankings and used SGD for training
- Paper accpeted by NeurIPS 2018 Workshop & INFORMS 2018
- Presented in NeurIPS 2018 Workshop, Iowa Research Open House, Data Mining Iowa Group, and Iowa Dare to Discover Campaign Showcase

Teaching Experiences

Teaching Assistant: *MIS 304 Intro to Problem Solving and Programming*   *Fall 2020 & Fall 2022*

Teaching Assistant: *MSCI:6070:Data Science,* University of Iowa   *Spring 2020*

Teaching Assistant: *MSCI:4220:Advanced Database Management & Big Data.* University of Iowa   *Spring 2020*

Teaching Assistant: *MSCI:6050:Data Management and Visual Analytics,* University of Iowa   *Fall 2018*

Teaching Assistant: *Big Data Summer School,* University of Iowa   *Summer 2017 & Summer 2018*

Leadership and Professional Development

2nd Place, USITCC Business Analytics Competition   *Apr 2018*
U.S. Information Technology Collegiate Conference, San Antonio, TX
Annual National Undergraduate Information Technology Competition
- Competed in 2 matches: got 2nd place for Business Analytics and honorable mention for Database Management

Team Leader: MinneMUDAC Competition   *Aug 2017 – May 2018*
MinneAnalytics – Twin Cities Big Data, Data Sciences, and Analytics Competition, Eden Prairie, MN
Annual Midwest Undergraduate Data Analytics Competition
- Led and managed a group of four business students to apply analytical skills in the competition
- Applied machine learning to discover the relationship between water quality and housing price

President of the Public Relations and Sponsorship   *Sep 2014 – Sep 2015*
Organization: Chinese Students and Scholars Association (CSSA), The University of Iowa
- Led, planned, and organized members to get sponsorship from local merchant
- Hosted public seminar to welcome new students, and made financial budget for every event

Study Abroad
National Taiwan University of Science and Technology   *Summer 2015*

Selected Awards

UT Austin Machine Learning Laboratory Research Grant	2021
Graduate College Presentation Travel Fund	*2019*
NeurIPS Travel Award	*2018*
Honor in Major	*May 2018*
Big Data Grant from National Science Foundation (NSF)	*Fall 2017 – Spring 2018*
Carol Fethke Honors Scholarship	*2018*
Iowa Center for Research by Undergraduates Research Fellowship	*Summer 2017*
Dean's List	*2015 – 2017*

Computer Skills

Python, R, C++, Tensorflow, SQL, Linux, MATLAB, SAS, Tableau, QGIS, Optimization Solver, LaTeX, and Microsoft Access.