Xinyue Chang

xchang225@gmail.com | chxyself25.github.io | chxyself25 | chxyself25

- *Trained applied statistician with 5+ years' experience in data-driven research and statistical programming, 1 year's intern experience in consulting and technology.*
- *Strategic, proactive data scientist with background in statistics and good understanding of machine learning algorithms, capable of independently conducting end-to-end analysis and modeling for large-scale data.*
- *Positive, deliberative collaborator with strong communication skills, having engaged with researchers from machine learning, geology and business.*

Technical Skills

Programming R, Python, SQL, SAS, Matlab, C++, parallel programming
Reporting R Shiny, HTML, LaTeX
System Linux, macOS, Windows

Education

Iowa State University (ISU)

Ames, IA

Ph.D. in Statistics

Aug. 2016 – Dec. 2020

- GPA 3.95/4.0 including courses: Modern Multivariate Statistical Learning, Machine Learning, Nonparametric Methods in Statistics, Advanced Spatial Statistics, Theory and Applications of Sample Surveys.
- Research Interest: Functional/longitudinal data, remote sensing data, spatial statistics, data applications on geological and environmental science.

University of Minnesota, Duluth (UMD)

Duluth, MN

M.S. in Applied and Computational Mathematics

Aug. 2014 – May 2016

- Minor in Computer Science, Overall GPA: 3.97/4.0, Major GPA: 4.0/4.0

Harbin Institute of Technology (HIT)

Harbin, China

B.Eng. in Automation

Aug. 2010 – Jun. 2014

- Overall GPA: 89.1/100, Major GPA: 90.72/100, Rank: 8% (14/179)

Work Experience

Center for Survey Statistics and Methodology

Ames, IA

Research Assistant

Jun. 2018 – PRESENT

- Developed and implemented a class of sparse functional change-point methodology based on FPCA, CUSUM and ensemble to do hypothesis testing and estimation of change-point for urbanization process, which achieves higher accuracy than the existing regression-based method and BEAST algorithm.
- Improved spatiotemporal satellite images imputation algorithm on estimations for non-stationary pixel-wise annual mean function and spatial effect of images with completely missing data, which supports the downstream task of water classification in localized regions.

Uber Technologies Inc.

San Francisco, CA

Data Scientist Intern

May 2019 – Aug. 2019

- Worked at marketplace forecasting team, focused on machine learning model evaluation and diagnosis.
- Conducted error analysis based on statistical testing and feature ranking, identified significant features in explaining large errors of tree-based models and improved holiday model performance.
- Built a monitoring & alerting system prototype in Python to accelerate machine learning model development cycle, which includes extracting data by SQL queries, metrics threshold calculation, detecting daily alerts in city level, and automatic root cause analysis.

After Inc.

Norwalk, CT

R Shiny Developer (Remote) *Aug. 2017 – May 2018*

- Generated R shiny reports by templates, created customized reports and participated in developing new web apps: early warning and survival analysis.
- Developed templates in Python for heatmap and serial range diagnostics reports, maintained the reporting repository.
- Designed and completed a Python-written framework for doing the same modification/editing in multiple R files.

Iowa State University

Ames, IA

Teaching Assistant *Aug. 2016 – May 2017*

- Assisted with answering R questions in lab session once a week, instructed students finish their assignments.
- Held office hours, graded lab assignments and homework timely.

University of Minnesota, Duluth

Duluth, MN

Teaching Assistant *Aug. 2014 – May 2016*

- Led and managed discussion sections twice a week, explained example questions to the class.
- Held office hours and graded quizzes every week.

Relevant Projects

Regression and Variable Selection in Asynchronous Longitudinal Data

Ames, IA

PhD Thesis *Jun. 2018 – PRESENT*

- Proposed using sparse FPCA techniques to calibrate asynchronous longitudinal data, proved asymptotic normality of the single regression estimator and achievability of the LASSO estimator selecting both time-varyinging and time-invariant variables.
- Demonstrated advantages over other available methods under various simulation scenarios, applied to "Study of Women's Health Across the Nation" dataset through a subsampling manner for obtaining stability paths and p-values.

Change-point Detection for Sparse Multivariate Functional Data

Ames, IA

PhD Thesis *Nov. 2018 – Jun. 2020*

- Developed both multivariate and univariate approaches for detecting and estimating the change-point in sparse multivariate functional data, which are verified in theory and simulation study.
- Implemented proposed methods for detecting urbanized changes and estimating urbanization years by using LANDSAT data in 503 locations across United States.

Don't Overfit! II

Ames, IA

Kaggle Competition *Apr. 2019 – May 2019*

- Led team to develop a Bayesian logistic regression model for the binary classification using 300 continous features, with only 250 training samples aimed for 19750 cases in test set.
- Defined the problem and preliminary analysis by implementing classic machine learning algorithms: logistic regression with L1 regularization, neural network, random forest, support vector machine.
- Organized results and tracked progress, implemented the ensemble of predictive models based on cross validation.

Publications

Chang, X., Dai, X., Zhu, Z. (2020). Functional Change-point Detection for Sparse Multivariate Functional Data with Application to Urban Dynamics. In preparation, link.

Chang, X., Zhu, Z., Hobbs, J., Dai, X. (2020). A Geospatial Functional Model for OCO-2 data with Applications on Imputation and Land Fraction Estimation. In preparation, link.

Labuzzetta, C., Zhu, Z., **Chang, X.**, Zhou, Y. (2020). A Methodological Pipeline for Submonthly and Highly Sensitive LANDSAT Surface-water Classifications via Gap-fill Imputation and Random Forest classifiers. *Remote Sensing of Environment*. To be sumitted.

Zhuang, Y., Lee, Y., **Chang, X.**, Kim, R. (2019). Entrepreneurial Orientation and Corporate Social Responsibility Performance: An Empirical Study of State- and Privately-Controlled Firms in China. *Corporate Social Responsibility and Environmental Management* 27(1), 383-392. DOI: 10.1002/csr.1872

Presentations

May 2020, **Functional Change-point Detection for Sparse Multivariate Functional Data with Application to Urban Dynamics**, Survey Working Group Seminar, Ames IA, link.

July 2019, **Functional Change-point Detection for Multivariate Sparse Functional Data with Application on Urban Dynamics**, Joint Statistical Meetings (JSM), Denver CO, link.

Selected Honors & Awards

2019	**Top 13%, 300/2330**, Kaggle Competition: Don't Overfit! II, Team leader	*ISU*
2018	**Meritorious Research Award**, Advanced Spatial Statistics Class Course Project, Team leader	*ISU*
2017	**Award for Experiential Development**, Department of Statistics	*ISU*
2016	**Outstanding Graduate Student Award**, Department of Mathematics & Statistics	*UMD*
2015	**Comprehensive Examination Distinction**, Department of Mathematics & Statistics	*UMD*
2012	**8841 Continuous Influence Scholarship**, Department of Control Science and Engineering	*HIT*

Leadership & Service

2019-2020	**Member**, Graduate College Emerging Leadership Academy	*ISU*
2019	**Organizer**, Survey Working Group Seminar at Department of Statistics	*ISU*
2019	**Membership**, American Statistical Association	*U.S.*
2017-2019	**Senator**, Graduate and Professional Student Senate	*ISU*
2011-2012	**Co-founder**, Magic Club	*HIT*