

JEREMIAH ZHE LIU

Department of Biostatistics
Harvard University

zhl112@mail.harvard.edu ✉
jereliu.info 🌐
github.com/jereliu 📄
+1 (319) 594-4694 📞

EDUCATION

Harvard University (Boston, MA) *PhD Biostatistics, Minor in Computer Science* Expected 2018
Completed advanced doctoral curriculum. GPA: 3.94/4.00.
Research Keywords: Ensemble Learning, High-dimensional Hypothesis Testing, Nonparametric Bayes, MCMC.

University of Iowa (Iowa City, IA) *BS Statistics, Mathematics, Computer Science* May 2013
magna cum laude, GPA: 3.96/4.00.

PROFESSIONAL EXPERIENCE

Martinos Center for Biomedical Imaging, Mass General Hospital 2016-Pres
Consulting Machine Learning Scientist

- Building reinforcement learning system to aid the development of energy-efficient MRI imaging device.
- Participating in design of manifold-inspired architecture for deep learning system for image reconstruction.

Harvard Clean Air Research Center 2013-2015
Assistant Statistician/Machine Learning Scientist

- Built spatiotemporal prediction system for heavy-metal air pollutants by integrating information from various sources (air monitoring records, meteorological information, etc) under Random Forrest and Kernel Regression.
- Developed Python and R script to extract 3D GIS features from openstreetmap API. Implemented server-end Python program to restructure high-volume wind-trajectory data.
- Implemented automated feature selection for GIS features using a combination of measurement error-based weighting and Ridge-type penalization. Conducted stratified cross validation to assess the model's out-of-sample prediction and the influence of prediction error on the risk estimation in second-stage association studies.
- Developed new methods to conduct semi-supervised clustering of spatial regions to enhance prediction accuracy.

RESEARCH EXPERIENCE

Detecting Nonlinear Interaction between High-dimensional Features using Kernel Ensemble 2015-Present.
Advisor: Dr. Brent Coull & Dr. Xihong Lin

- Proposed testing procedure to detect interaction between groups of high-dimensional features under Tikhonov-regulated Gaussian Kernel Machine.
- Proved theoretical lower bound on the procedure's specificity under mis-specified kernel function.
- Provided guideline in constructing nonparametric kernel functions to guarantee valid hypothesis test.

On convergence of MCMC algorithms for Bayesian Collaborative Filtering 2016-Present.
Advisor: Dr. Lorenzo Trippa & Dr. Sergio Bacallado

- Proposed a unified framework for collaborative filtering models to handle censored continuous and count data.
- Implemented and compared efficacy of different MCMC methods (Gibbs, SGLD, Hamiltonian MC, Particle Filter, etc) in the scenario of latent models.
- Deriving diagnostic method for real-time detection of MCMC convergence.

TECHNICAL SKILLS

- **Analysis & Modelling:** R, Matlab, Python, Stan, BUGS, SAS, Mplus
- **Graphics & Documents:** ggplot2, OpenGL, Shiny, ArcGIS, L^AT_EX
- **High Performance Computing:** C (OpenMP, CUDA, OpenCL)
- **Software Development:** C++, Java, Python, Shell script

OPEN SOURCE PROJECT

GURLS_MKL: Fast Multiple Kernel Learning Library for GURLS Package 2015

- Independently developed multiple kernel learning functionality for *Grand Unified Regularized Least Squares* (GURLS), an state-of-art supervised-learning package developed at MIT
- Extended fast Proximal Forward-Backward Splitting (PFBS) optimization algorithm to allow memory-efficient iteration update with parallel support. Derived boundary conditions on algorithm parameters to guarantee model convergence.

GPU-Accelerated Sampling for Bayesian Normal Conditional Autoregressive Models 2012

PI: Dr. Kate Cowles

- Designed and implemented parallel algorithms in OpenCL for new model computation strategy proposed by Cowles et al.(2012) for Bayesian Normal CAR model.
- Implementation incorporated into R package *CARrampsOcl*.

COURSEWORK

Statistical Machine Learning

- Kernel Method Theory • Advanced Regression & Learning
- Semi-parametric Inference • Computation Intensive Statistics

Theoretical Statistics

- Probability Theory • Advanced Bayesian Inference • Theory of Hierarchical Linear Models
- Environmental & Spatial Statistics • Analysis of Genetic Association Studys
- Causal Inference

Mathematics

- Linear Algebra & Multivariate Calculus • Real Analysis & Measure Theory • Matrix Theory
- Numeric Analysis • Ordinary Differential Equation • Nonlinear Optimization

Computer Science

- High Performance & Parallel Computing • Data Structure • Algorithm • Foundation in GIS

CONSULTING EXPERIENCE

Causal Networks for Retarded Bone Growth in HIV-infected Adolescents 2014-2015.

Advisor: Dr. Brent Coull, Dr. Jane Lindsey & Dr. Denise Jacobson

- Trained regularized probabilistic network to model the association between biomarkers and bone growth measures in prenatally HIV-infected adolescents in PACTG 1045 study.
- Wrote interface between M-plus and R to deploy computation-intensive hypothesis generation and testing.
- Identified sub-collection of biomarkers robustly associated with retarded bone growth in HIV-affected teenagers.

Advisor: Dr. Jeffery Dawson

- **Objective** To analyse patients' neuromuscular reflexion pattern during weight-bearing task.
- **Role** Applied and compared two-level hierarchical model and semi-parametric GEE model in the analysis of temporal pattern, and used Tukey HSD to test for level of response between treatment groups.
- **Status** Presented results in Dept. of Biostatistics' Doctoral Student Seminar, and independently finished an abstract of analysis to advisor.

High-level Air Pollution and Health in Taiyuan, China: from Birth to Death

2011-2012

- **Objective** To assess the association between air pollution (PM10, SO2, NO2, CO) and series of health endpoints (Negative Birth Outcome, Children Respiratory Symptom, Mortality) in highly air polluted city in China.
- **Role** Modelled Concentration-response Functions (CRF) between exposure and health outcome through Generalized Additive Model with cubic splines. Wrote C++ program to detect change points in nonlinear CRFs in order to estimate the dose-specific relative risk of air pollutants.
- **Status** Accepted to ISEE Conference and series of journal papers are in preparation.

Standards for Birthweight by Gestational Age in Northern China, 2010

2010

- **Objective** To construct centile charts of birthweight, body length, head and chest circumference from 2006-2010 Taiyuan birth surveillance data containing 200,000 newborn records. (Negative Birth Outcome, Children Respiratory Symptom, Mortality) in highly air polluted city in China.
- **Role** Combined Two-component Mixture Models with LOWESS to obtain robust percentile regression for noisy data.
- **Status** Submitted to local Dept. of Health to release as diagnosis reference for Small-for- Gestational-Age (SGA) Newborn, and journal paper accepted by *Early Human Development*

PUBLICATIONS

Liu Z, Coull B. *Testing for Nonlinear Interaction between High-dimensional Feature Groups using Targeted Kernel Ensemble*. In Progress

Liu Z, Lindsey J, Coull B, Jacobson D. *Biomarkers and bone growth across Tanner stages in perinatally HIV- exposed youth in PACTG 1045*. AIDS Submitted

Zhang J, Gao J, Liu Z, et al. *Short-term Effects of the Beijing Haze Episode on Local Hospital Outpatient and Emergency Room Visits*. Journal of Environment and Health. To appear

Liu Z, Zhang J, Zhao B, et al. *Population-based reference for birth weight for gestational age in northern China*. Early Human Development 2014;90(4):177-87.

HONORS & AWARDS

HSPH Central Grant, Department of Biostatistics, Harvard School of Public Health, 2014-2015

Department Scholarship, Department of Statistics & Act. Sci., University of Iowa, 2013

ISEE Conference Student Scholarship, International Society of Environmental Epidemiology, 2012-2014

Annual Departmental Scholarship, Department of Statistics & Act. Sci., University of Iowa, 2012

Phi Beta Kappa, Alpha of Iowa Chapter, CLAS, University of Iowa, 2012