KAI LIU

kai.liu@utexas.edu • 512.917.6781 • GitHub://KaiUT • LinkedIn://Kai Liu

PROFILE

- Ph.D in Computational Biology (expected August 2018);
- Ample Skills in data science:
  - Experience in data mining, machine learning, statistical inference, and big data analysis;
  - Comprehensive technical/computing skills includes Python, R, Unix, Vim, Git, C++;
- Excellent problem solving skills in both independent and team environments;
- Skilled presenter of technical material to both technical and non-technical audiences;
- Quick, thorough and effective learner.

EDUCATION

Ph.D in Computational Biology ◇ *The University of Texas at Austin, Austin, TX* — Expected August 2018
M.S in Microbiology ◇ *Huazhong Agricultural University, Wuhan, China* — Grad. June 2013
B.S in Biotechnology ◇ *Huazhong Agricultural University, Wuhan, China* — Grad. June 2011

RESEARCH EXPERIENCE

Graduate Research Assistant ◇ *The University of Texas at Austin* — December 2014 - Present

Developing Biosurveillance App in Python

- Retrieved and cleaned infectious diseases related data from Google Trends, Wikipedia, WordPress etc;
- Developed an algorithm to detect infectious disease outbreaks using multiple data sources, in collaboration with a mathematician;
- Improving performance and speed of the algorithm;
- Assessing the algorithms on infectious diseases in different regions (using 552 time series data sources);
- Connecting algorithms with the App back-end and front-end, and integrating the App into Cloud Ecosystem, in collaboration with a front-end engineer.

Assessed Real-time Zika Risk in the State of Texas

- Collaborated with other researchers in developing a branching process model framework that captures variation and uncertainty in Zika case reporting, importations, and transmission;
- Applied the framework to assess county-level epidemic risk throughout Texas.

Developed Mathematical Models for Infectious Diseases

- Developed Ordinary Differential Equations frame for infectious diseases eliminating the assumption that all individuals in a population have the same number of contacts, based on models published previously.
- Simulated infectious diseases spreading on contact networks using multiple algorithms.

COURSE PROJECTS

Denoised GPS Data by Applying Kalman Filter — October 2016 - December 2016

- Implemented Kalman filter in R, and smoothed GPS data collected from a vehicle cruising around campus (814458 samples).

Predicted Yelp Rating Based on User Review Enhanced Collaborative Filtering — September 2015 - December 2015

- Developed a new Collaborative Filtering-based method to improve the accuracy of user's rating prediction and solve the sparseness of dataset by combining item's features and user opinions from all reviews;
- Applied the new method to predict user ratings using restaurants dataset from Yelp (10GB). Its performance is 4.23% better than that of traditional KNN method, and its coverage is 100%.

Forecasted Tourism Earnings of United Kingdom — October 2015 - December 2015

- Predicted Tourism Earnings of UK using a dynamic linear regression model and Forward Filtering and Backward Sampling algorithm in R.

Statistical Modeling — November 2014 - December 2014

- Analyzed a dataset to determine 1) factors related to 12 month weight loss, and 2) whether an intervention was effective in increasing weight loss by applying both frequentist and Bayesian inference methods.

SKILLS

Statistical Modeling	Regression models · Time series and dynamic models · Hypothesis testing and confidence interval · Data fitting · Ordinary differential equations · Network simulation
Big Data Analysis	Online learning · Regularization and sparsity in statistical models · Matrix factorization · Spatial smoothing · Principal component analysis and dimensionality reduction
Data Mining & Machine Learning	Regression · Classification · Clustering · Frequent Pattern Mining
Programming	Fluency in Python, R, Git, Unix · Familiar with Vim, Linux · Experience in C++, MATLAB, LaTex

COURSES

Graduate Courses	Data Mining · Statistical Modeling I · Statistical Models for Big Data · Time Series & Dynamic Models · Regression Analysis
MOOC	Machine Learning · Coding the Matrix: Linear Algebra through Computer Science Applications · Pattern Discovery in Data Mining · R Programming · Getting and Cleaning Data · Exploratory Data Analysis

TEACHING EXPERIENCE

Teaching Assistant ◇ *The University of Texas at Austin* September 2015 - December 2015

September 2014 - December 2014

- Mentored two lab sections (48 students) of an undergraduate genetics course and an undergraduate microbiology course;
- Got 4.3/5.0 in both course evaluations;
- Prepared lab lectures and lab plans;
- Graded quizzes, assignments and exams.