Kai Liu

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

EDUCATION

The University of Texas at Austin
Ph.D in Computational Biology
Expected August 2018 | Austin, TX

Huazhong Agricultural University
M.S in Microbiology
Grad. June 2013 | Wuhan, China

B.S in Biotechnology
Grad. June 2011 | Wuhan, China

SKILLS

Statistical Modeling
Regression models • Time series and dynamic models • Hypothesis testing and confidence interval • Data fitting (Least square, maximum likelihood, Bayesian methods) • Ordinary differential equations (ODEs) • 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 and Machine Learning
Regression • Classification • Clustering • Frequent Pattern Mining

Programming
Fluency in Python, R • Fluency in Git, Vim, Linux • Familiar with C++, MATLAB, LaTex

COURSEWORK

Graduate
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

RESEARCH PROJECTS

Developing Surety BioEvent App
December 2015 – Present

- Retrieved and cleaned Influenza- and Dengue-related data from Athena Health, Google Trends, Wikipedia etc;
- Collaborated with a mathematician in developing a new Early Event Detection (ED) algorithm to detect infectious disease outbreaks using multiple data sources;
- Assessing the Situational Awareness (SA) and ED algorithms on Influenza in US (using 452 time series data sources) and Dengue in Puerto Rico (using 100 time series data sources);
- Improving performance and speed of the ED algorithm;
- Collaborating with a front-end engineer to connect algorithms with the App back-end and front-end, and integrate the App into Biosurveillance Ecosystem.

Assessed Real-time Zika Risk in the State of Texas
March 2016 – May 2016

- 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 Edge-based Mathematical Models for Infectious Diseases
December 2014 – November 2015

- Developed susceptible-exposed-infectious-recovered mathematical models for infectious diseases eliminating the assumption that all individuals in a population have the same number of contacts, based on models published previously;
- Compared results from these models with that from infectious diseases spreading simulations on contact networks.

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 police 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 (5GB). 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.