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, big data analysis, and natural language processing;
  - Comprehensive technical/computing skills includes Python, R, Git, C++, SQL, Hadoop, Spark;
- Excellent problem solving skills in both independent and team environments;
- Skilled presenter of technical materials 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 Infectious Diseases Surveillance App | Python

- Retrieved and cleaned infectious diseases related data from Google Trends, Wikipedia, WordPress etc;
- Developed a regression model and a Multivariate Exponentially Weighted Moving Average (MEWMA) model to detect infectious disease outbreaks using multiple data sources, in collaboration with a mathematician;
- Optimizing data sources on infectious diseases surveillance in different regions (554 time-series) by combining above models and stepwise variable selection algorithms;
- 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 | R

- 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.

PROJECTS

Developing a R Package for Big Data Analysis | R & Rcpp

- Implementing following algorithms in the package: stochastic gradient descent using line search and quasi-Newton methods to determine step size · the lasso · the proximal gradient method · Laplacian smoothing solved by sparse Cholesky/LU, the Gauss-Seidel method, the Jacobi iterative method, and conjugate gradient method · graph fused lasso solved by Alternating Direction Method of Multipliers (ADMM) · sparse matrix factorization.

Predicted the Direction of Exchange-Traded Fund (ETF) movement | Python

- Retrieved nine historical ETF sectors data from Yahoo Finance;
- Implemented Logistic regression, Ridge & Lasso regression, and Artificial Neural Network to predict the direction of ETF movement;
- Achieved an accuracy of 55% $\sim$ 60% for predicting nine ETF sectors movement; and the trading strategy based on my prediction outperforms baseline strategies.

Denoised GPS Data by Applying Kalman Filter | R

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

Predicted Yelp Rating Based on User Review Enhanced Collaborative Filtering | R

- Extracted user opinions from restaurants dataset from Yelp ($\sim$10GB) using Stanford coreNLP tool;
- 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;
- Improved the prediction accuracy by 4.23% compared to the traditional KNN method, and the coverage is 100%.

SKILLS

Programming	Fluency in Python(NumPy, SciPy, Matplotlib, pandas, scikit-learn), R, Git · Familiar with MATLAB, Linux, LaTex · Experience in C++, SQL, Hadoop, Spark
Data Mining & Machine Learning	Regression with regularization · Neural Network · Support Vector Machine · Ensemble Methods · Hidden Markov Model · Clustering · Frequent Pattern Mining
Statistical Modeling	Regression models · Time series and dynamic models · Ordinary differential equations · Model fitting · Network simulation