Yutong Dai

217.721.9646 · yutongd3@illinois.edu · 1627 E Florida Avenue, Urbana, IL 61802 · roth.rbind.io

Education

University of Illinois at Urbana-Champaign (UIUC), IL, USA — Sept.2017 – May.2019
M.S. in Statistics — GPA: 4.00/4.00

Sichuan University (SCU), Sichuan, China — Sept.2013 – Jun.2017
B.S. in Mathematics with honors (concentration in Statistics) — GPA: 3.68/4.00

Skills & Tools

- Skills: Machine Learning Fundamentals, Scientific Data Visualization, Deep Learning
- Programming Language & Software : R, Python, SQL, LaTeX, Rmarkdown, Rshiny, Microsoft Office

Experiences & Research

Internship: **Anheuser-Busch InBev** — Jan.2018 – Present
Data Scientist Urbana, IL

Global Environment Pattern Modeling

- Revised clustering algorithms with agronomists' on field knowledge to formulate a global barley production environment model that accounts for complex weather (El Niño/La, Niña,...) and soil systems
- Explored new potential barley production regions based on the global barley production environment model

Best Management Package Recommendation

- Proposed an inverse predictive model based on the Random Forest to suggest optimal management packages (variety, fertilizer, fungicide, crop rotation...) that help growers to hit highest barley yield
- Designed Smart Barley UI/UX prototype in Rshiny to dynamically visualize analytic results, like growers' production performance and highest yield management packages, and delivered it to agronomist teams
- Collaborated with computer scientists to scale up analytics results and put them into production environments

Anomaly Detection (in progress)

- Modified the Isolation Forest algorithm to provide an automatic anomaly detection mechanism to improve the Smart Barley database quality

Internship: **Brightech International** — Dec.2016-Jan. 2017
Research Assistant Chengdu, China

- Performed data cleaning and exploratory data analysis in SAS to support research projects
- Conducted clinical data analysis to assist protocol writing

Research: **Convergence Rate Analysis of Parallel Block Coordinate Descent Method**[Link] — Dec.2016 – Jun.2017
Advisor: Prof. Yang Weng — **Accepted by** *Journal of System Science and Complexity*

- Proposed synchronous parallel block coordinate descent algorithms for minimizing a genre of composite functions with sub-linear convergence rate
- Implemented algorithms to solve large scale logistic regression with ℓ_1 norm penalty

Research: **Attention Data Provides Insights into the Financial Market Bubble**[Link] — Aug.-Nov.2016
Advisor: Prof. Yang Weng — Submitted to *PlOS ONE*

- Utilized the search volume index as the novel and leading measurement of public attention to study the stock market bubble generating feedback hypothesis
- Tested the hypothesis on Nasdaq Composite, Dow Jones Industrial Average, and S&P 500 Index

Project: **Show and Tell: Neural Image Caption**[Link] — Nov.2017 – Dec.2018
Advisor: Prof. Justin A. Sirignano

- Fine-tuned the 101 layered Residual Network pre-trained on the ImageNet as an image encoder to interpret image contents
- Designed and trained a Recurrent Neural Network with 3 layers of Long short-term memory(LSTM) cells as an image decoder to convert visual information into texts
- Calculated vocabulary scores from the image decoder's outputs and generated captions using the beam search method to obtain high quality captions (achieved 0.33 BLEU score)
- Designed a simple UI for general audience to explore the prototype

Project: **Real Estate Market Data Analysis**[Link] — Dec.2017
Advisor: Prof. Feng Liang

- Performed data querying from website and exploration data analysis
- Explored how amenities and geo-location influence listing prices
- Developed a web-based data product incorporating with machine learning algorithms to help 1)Airbnb hosts to determine listing prices; 2)Airbnb customers to gain insight into Airbnb Boston rental markets

Project: **Predicting Box Office Revenue and IMDb rating for Movies** Dec.2017

Advisor: Prof. Ruoqing Zhu

- Developed a revenue prediction model based on Gradient Boosting, Random Forest and Elastic Net
- Designed a classifier to predict the IMDb ratings based on SVM and Sparse Logistic Regression

Project: **Optimal Values of Variables Based on Computer Experiments**[Link] Apr.-May.2016

Advisor: Prof. Yongdao Zhou

- Performed uniform designs, orthogonal designs and Latin hypercube designs to select suitable experimental points.
- Built models via regression and neural network with data from simulated experiments points.

Project: **Sufficient Boarding Strategy Based on Cellular Automata**[Link] Dec.2015

Advisor: Prof. Rui Zhu

- Designed Cellular Automata and Monte Carol methods to explore how seats and luggages influence the boarding/deboarding time
- Proposed the optimal boarding/deboarding strategy with minimal time

Contest: **Interdisciplinary Contest in Modeling**[Link] Feb.2016

Group Leader Advisor: Prof. Rui Zhu **Awards**: Meritorious Winner (Top 8%)

- Addressed the problem of ranking the shortage degree by using systematic cluster analysis
- Modified Cobb-Douglas production function to find the mechanism of how social and environmental drivers influence supply and demand
- Devised an intervention strategy based on the existing Lesotho Highlands Water Project

Related Courses

Statistical Learning, Linear and Generalized Linear Models, Design of Experiments, Multivariate Statistical Analysis, Statistical Computing, Time Series Analysis, Probability Theory, Statistical Inference, Convex optimization

Honors and Awards

- *Bachelor's Thesis: Best Paper Award* (0.58%) 2017
- *Dean's List* (Top 10) 2014-2017
- 2^{nd} *Prize*, National College Students' Mathematical Modeling Competition 2015
- *Excellent Student Leader*, Sichuan University 2014