

QIUYUN ZHU

Contact Information

TEL: +1 608-9605525 (US) +86 185-29153812 (China)
Email: qiuyunz@yahoo.com rachyun94@gmail.com
Current Address: 2nd Floor, Cisco, INDIGO, Chaoyang District, Beijing, China

Education

Master of Science in Statistics, Data Science *Sep.2015-May.2017*
& Visiting International Student Program

Department of Statistics
University of Wisconsin-Madison, Madison, WI
Overall GPA (excluding Piano Lessons): 3.72/4.00

Bachelor of Science in Mathematics, Statistics *Sep.2012-June.2016*

Department of Mathematics
Nanjing University, Jiangsu, China
Overall GPA: 4.27/5.00

Related Coursework

University of Wisconsin-Madison:

- STAT 733: Theory of Probability I. (Probability Theory and Examples, Fourth Edition. *Rick Durrett*)
- CS 760: Machine Learning
- STAT 771: Statistical Computing. (Computational Statistics, Second Edition. Geof H. Givens and Jennifer A. Hoeting)
- STAT 601&602: Statistical Methods. (Statistical Principles for the Design of Experiments. R. Mead, S. G. Gilmour and A. Mead)
- STAT 609&610: Statistical Inference. (Statistical Inference, Second Edition. George Casella and Roger L. Berger)
- STAT 679: Data Science in R. (ISLR. Gareth James, Daniela Witten, Trevor Hastie and Robert Tibshirani)
- STAT 679: Data Science Computing Project
- STAT 679: Graphical Models
- STAT 456: Applied Multivariate Analysis. (Applied Multivariate Statistical Analysis. Richard A. Johnson and Dean W. Wichern)
- STAT 349: Introduction to Time Series. (Time Series Analysis with Applications in R. Jonathan D. Cryer and Kung-Sik Chan)
- STAT 351: Intro Nonparametric Statistics
- STAT 327: Data Analysis with R

Nanjing University:

- *Mathematics*: Real Analysis, Higher Algebra, Functional Analysis, Numerical Methods and Experiments, Mathematical Analysis, Ordinary Differential Equations, Partial Differential Equations, Functions of one Complex Variable, Abstract Algebra, Analytic Geometry, Discrete Mathematics, Actuarial Mathematics
- *Statistics and Computer Science*: Foundations of Probability Theory, Mathematical Statistics, Applied Stochastic Processes, Visual C++ and Computer Programming

Research Experience

A Comparison among Classification Algorithm

Mar.2017-Apr.2017

Faculty adviser: Professor David Pages, University of Wisconsin-Madison

- This project mainly focuses on the most common classification algorithms by comparing error rate, training time and some parameters (e.g. size of the tree). Five datasets are selected from different real-life domains with diverse properties.

The Effect of Road Condition on Fatal accidents in United States

Sep.2016-Dec.2016

Faculty adviser: Professor Karl Rohe, University of Wisconsin-Madison

- Selected control variables out of 50 features and estimated the quality of the road by bridge condition. After data cleaning, fitted a linear model with R between number of accidents by county and selected features. By visualizations with ggplot in R, analyzed the pattern of the residuals.

Analysis Based on Generalized Measure of Correlation (GMC) and Simple Linear Model to Investigate Which Genes are Related to Tumor Protein p53

Sep.2015-Dec.2015

Faculty adviser: Professor Zhengjun Zhang, University of Wisconsin-Madison

- Applied a transformation based on the nonparametric estimation of the empirical cumulative distribution function to the original data to solve the bimodal phenomenon, then used three criteria (AIC, BIC and Mallows Cp) to select variables to find the most reliable linear model with R.
- Used GMC and Lasso as the criteria to select variables, considering the nonlinearity between response and variables. For each transformation of the original data, came out with a new algorithm to evaluate variable selection.

Work Experience

Cisco (Contractor)

Sep.2017-present

Data Scientist (Machine Learning and Data Mining)

- Working on prognostic problems of bearings' life duration. I am using Hilbert-Huang Transform to do feature extraction (including denoising) and feature selection. Then I use classification method such as SVM to do segmentation on the time series to find degradation threshold. After, fitting a regression model on the selected indicator data to predict the bearings' behavior, thus getting the remaining useful life.
- Used linear regression on processed data to predict the average removal rate of wafer in Spark.

University of Wisconsin-Madison (Part-time)

Sep.2016-Dec.2016

Grader of Math 521 (Analysis I)

Amazon Information Service (Summer Internship)*July.2016-Aug.2016**Software Development Engineer II*, International Technology Department

- Extracted and analyzed big data about “ops” and click counts classified by websites and slots in the UK market by SQL, then analyzed the reason why some websites and slots had the most clicks or “ops” by Excel and R to improve the slots and websites.
- Analyzed data about “ops” and clicks in the UK market using feature analyzer, page analyzer and journey analyzer, and investigated customer behaviors.
- Obtained the promotion contributions to “ops” by SQL, and established a predictive model on how to allocate the promotion quantities for each promotion type to make the most profits.
- Used content-symphony to design slots of landing pages in China for A/B testing.

Skills Summery

- Proficient in R, Python, SQL, Latex, Word, Excel, PowerPoint
- Familiar with C++ and C, Spark, HTCondor (High-Throughput Computing), Tableau, Linux Shell Script, Emacs Editor
- National Computer Rank Examination Certificate of Level 2 in C