Rongrong Zhang

www.linkedin.com/in/rongrongz

CONTACT INFORMATION	2252 Sagamore PKWY W APT F3 West Lafayette, IN 47906	765-637-8717 zhan1602@purdue.edu

OBJECTIVE Looking for full-time opportunities in data scientist positions in fall 2018.

EDUCATION

PhD, *Statistics*, **Purdue University**, West Lafayette, IN., USA

> **GPA: 3.97**, Advisor: *Prof.* Michael Yu Zhu *Expected:* **May** 2018

MS, *Mathematics,* **Peking University**, Beijing, China

> Partial differential equations, Advisor: *Prof.* Baoxiang Wang **June** 2013

BS, *Mathematics and Statistics*, **Lanzhou University**, Lanzhou, China **June** 2010

EXPERIENCE

Statistic Consultant **June** 2014 — **August** 2016

- *Statistical consulting service* at Purdue University

Proactively engaging with over **30** clients from Purdue University among various disciplines. The prompt and detailed procedures or statistic implementations are provided to clients by communicating their data and requirements through emails or face-to-face meetings. To help client achieve the goals in time, I helped them with statistical modeling, experiment designs and survey data analysis, including power calculations, repeated measure mixed effects models, linear regressions, prediction, classification using a wide variety of statistical computing programs, including R, SAS, SPSS, and Minitab.

Research Assistant **August** 2015 — **Present**

- *Using Deep Neural Networks to Automate Large Scale Statistical Analysis for Big Data Applications*, advised by *Prof. Michael Zhu.* Presented in 9th Asian Conference on Machine Learning (ACML17), Seoul, Korea, 2017.

Statistical analysis (SA) is a complex process to deduce population properties from analysis of data. It usually takes a well-trained analyst to successfully perform SA, and it becomes extremely challenging to apply SA to big data applications. We propose to use deep neural networks to automate the SA process. In particular, we propose to construct convolutional neural networks (CNNs) to perform automatic model selection and parameter estimation, two most important SA tasks. We refer to the resulting CNNs as the neural model selector and the neural model estimator, respectively, which can be properly trained using labeled data systematically generated from candidate models. Simulation study shows that both the selector and estimator demonstrate excellent performances. The idea and proposed framework can be further extended to automate the entire SA process and have the potential to revolutionize how SA is performed in big data analytics.

- *Inferring Spatial Organizations of Chromosomes via Helix Models*, advised by *Prof. Michael Zhu.* Presented in Joint Statistical Meeting 2016 in Chicago.

Working with *Prof. Michael Yu Zhu* and *Prof. Ming Hu* from NYU to understand the 3D spatial organizations of chromosomes and functional implications of such structure. We developed and applied statistical and computational methods in analyzing Hi-C data. We proposed new parsimonious, easy to interpret piecewise helical curve models for reconstructing 3D chromosomal structure from Hi-C data. Markov Chain Monte Carlo (MCMC) method is used in parameters estimation and is implemented using C++ programming. Please see this website for details: *https://rsquared1427.github.io/phm/*

- *Differential Gene Expression Pattern with Exon-level Count Data using Compound Symmetry Multivariate Negative Binomial Distribution*, advised by *Prof. Michael Zhu.*

We proposed to use multivariate negative binomial distribution with certain covariance structure to model RNA-Seq exon count data, and two likelihood ratio tests were also proposed to decide whether a gene is differentially expressed and whether there is any changes in the alternative splicing events. The expression of a gene is characterized by a distribution rather than a single value. The major difficulty in applying multivariate count distribution lie in the computational burden. Our method provides a modeling scheme for the RNA-Seq count data when counts are correlated.

Graduate Student Instructor at Purdue University   January 2016 — Present

- Acting as the course web-master for Introduction to Statistics class (STAT 350).

  Responsible for preparing and administering weekly quizzes,reviewing SAS, R and lab notes. holding weekly office hours, and proctoring examinations.

Research Assistant at Peking University   August 2011 — June 2013

- *Nonlinear partial differential equations, Harmonic analysis*, advised by *Prof. Baoxiang Wang*

COURSE PROJECTS

Loan Payoff Prediction (Data Mining Course Project)
Purdue University   **March** 2016 — **May** 2016

- Decide if the bank should or should not lend money to these costumers who are asking for personal loans. Applied NBC, SVM, Logistic regression, Random Forest to determine whether these costumers can payoff their loans or not. Built and tested models via cross-validation method.

Convolutional Neural Network for Facial Attractiveness Rating
Purdue University   **Sep** 2016 — **Dec** 2016

- We proposed a deep leaning method and constructed a convolutional neural network (CNN) to predict how attractive you are from a photo, which is implemented using MXNET.

GRADUATE COURSEWORK

- Statistical modeling:

  Generalized Linear Model
  Advanced Stat Methodology
  Nonparametrics & Machine Learning
  Artificial Intelligence

- Computing:

  Computational Statistics I, II
  Visualizing Large Complex Data
  Data Mining
  Computing For Big Data Analysis (Spark)

AWARDS

- The Founder Scholarship of Peking University, 2010-2011
- Merit Student in Peking University (1759 out of 22854, 7.70%), 2010-2011
- National Scholarship,(142 out of 19213 students in Lanzhou University, 0.74%), Ministry of Education of the People's Republic of China, 2008-2009
- First Prize of Lanzhou University Undergraduate Mathematical Contest in Modeling, 2007-2008

SKILLS

Machine Learning: parametric, non-parametric, resampling techniques (bootstrap, cross validation), classification, clustering, bagging, boosting
Programming: R, SAS, Python, C++, SQL, LaTeX, Linux, Apache Spark.
Deep Learning packages: Caffe, MXNET, Tensorflow

PUBLICATIONS

- S. Mitsuru, B. Wang, and **R. Zhang**. Local Well-Posedness for the DaveyStewartson Equation in a Generalized Feichtinger Algebra. *Journal of Fourier Analysis and Applications*, 21.5 (2015): 1105-1129.
- **R. Zhang**, M. Hu, Y. Zhu, Z. Qin, K. Deng and J Liu. Inferring Spatial Organizations of Chromosomes via Piecewise Helical Models. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, Accepted.
- **R. Zhang**, W. Deng, Y. Zhu. Using Deep Neural Networks to Automate Large Scale Statistical Analysis for Big Data Applications. Proceedings of the 9th Asian Conference on Machine Learning (ACML17), Seoul, Korea, 2017.