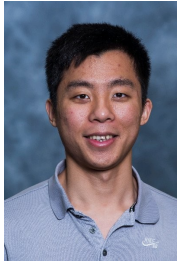


# Shan Zhong

1-435-327-9155 | [zhongs@email.sc.edu](mailto:zhongs@email.sc.edu) | [linkedin.com/in/shan-zhong-sc/](https://www.linkedin.com/in/shan-zhong-sc/) | [shanlearning.github.io](https://shanlearning.github.io)

## EDUCATION

---



### University of South Carolina

*Doctor of Philosophy in Statistics, 3.6/4.0*

Columbia, SC

*Aug.2018 – Present*

### Columbia University

*Master of Science in Actuarial Science, 3.5/4.0*

New York, NY

*Jan.2017 – May.2018*

### Southern Utah University

*Bachelor of Science in Mathematics and Economics, 3.3/4.0*

Cedar City, UT

*Aug.2012 – Dec.2016*

## PUBLICATION&PHD PROJECTS

---

Published(First Author): **Shan Zhong**, & David B. Hitchcock. (2021). S&P 500 Stock Price Prediction Using Technical, Fundamental and Text Data. *Statistics, Optimization & Information Computing*, 9, 769-788. In this paper we summarized historical prices, finance reports, and text information from news items associated with 518 different common stocks issued by current and former S&P 500 large-cap companies, from 2000 to 2020. A 66.18% accuracy in S&P 500 index directional prediction and 62.09% accuracy in individual stock directional prediction was achieved by combining different machine learning models such as Random Forest and LSTM together into state-of-the-art ensemble models. Our study's innovation includes utilizing deep language models (word embedding and BERT) to categorize and infer financial news item sentiment, and overcoming the insufficient data problem for machine learning models in time series by fusing different models containing different combinations of variables and stocks to jointly make predictions.

Nearly completed (First Author): Bayesian function data clustering. We identified and extracted the emotions of the main characters in nearly 100 world famous novels (Entity Sentiment), and performed a functional cluster analysis on the time function curve between the sentiment scores corresponding to the characters and the percentage of novels read, to discover the different patterns authors have in their writing. We warped the function curves using Bayesian registration to align the appearance times and spans of different characters. At the same time, we also carried out various visualizations relating to word frequency, sentence pattern, and sentiments after mining into the data.

Two other papers in preparation: Reinforcement learning stock trading policy with limited action and states. Dropout and deep ensemble based neural network confidence interval for stocks.

## EXPERIENCE

---

### Teaching and Research Assistant

Aug 2018 – Now

*University of South Carolina*

*Columbia, SC*

- Teach three sections of STAT 201 Lab, 72 people each. Wrote blogs style statistical learning tutorials.
- Course project: using max pooling and CNN to do image classification, Gibbs and Metropolis-Hastings sampling for simulation, EM algorithm to solve likelihood function, and time series Nonparametric model for weather forecasting.
- Completed four large-scale final projects: Bayesian stochastic clustering of social media friends circle, high-dimensional LASSO regression for medical cost prediction, MCMC story generation, and time series block bootstrap. Demo of the four projects: <https://github.com/Shanlearning/past-coursework#past-coursework>

### Actuarial Consultant

Jun 2019 – Sep 2019

*Yu Chun Business Consulting*

*Shanghai, China*

- Automated web data scraping from 91 life insurance companies for over 20,000 Critical Illness, health, and accident insurance products via python Scrapy.
- Use regular expressions and TF-IDF word frequency language model to extract keywords from documents and construct databases.
- Product valuation in terms of net present value (NPV), construct recommendation system.

### School Integrated Project

May 2017 – Aug 2017

*Guy Carpenter & Columbia University*

*New York, NY*

- Developed Poisson regression models to predict actual claim count, with past 30 years of data. Calculated adjusted loss ratio volatility and correlation between 15 different lines of business, for 1400 companies.

- Drafted weekly research reports to present to chief actuary *Mr. Steve White*.
- Used SQL language to manipulate raw data and construct Access databases.

### **Math Modeling Intern**

Oct 2016 – Dec 2016

*Casino Game Maker, Inc.*

*Cedar City, UT*

- Evaluated winning strategy for different 52-deck poker games and designed new game variants for use in gaming devices.
- Used excel VBA Monte Carlo decision tree simulations to validate game programs.
- Design algorithm, and calculate the odds ratio according to the weight change of possible outcomes.

### **Research Assistant**

2015-01 – 2016-05

*Southern Utah University*

*Cedar City, UT*

- Used three different databases, to build models to assess the family assets of incoming freshmen based on their home addresses filled for over 10,000 students.
- Utilized Excel's vlookup function to input and clean data.
- Grade other students' exams and assignments.

### **Actuarial Intern**

Jun 2014 – Aug 2014

*New China Life Insurance Co., Ltd.*

*Shanghai, China*

- Used excel to process about 70,000 rows of premium income data of around 200,000 people, filtered and categorized the data.
- Calculated and analyzed accident insurance claims data from multiple regions.
- Processed hundreds of accident insurance underwriting documents, review and seal.

## **PERSONAL WEBSITE AND QUALIFICATION CERTIFICATE**

---

- Personal website: [shanlearning.github.io](http://shanlearning.github.io)
- Actuarial Exam: P, FM, IFM, SRM, VEE all Filled

## **TECHNICAL SKILLS**

---

**Python:** tensorflow, pytorch, sklearn, pandas, matplotlib

**R:** ggplot, data.table, various statistical packages

**Other:** Latex, HPC, SQL query, node.js webpack