

Hanchao(Ted) Zhang

Tel: (646) 206-6662 | Email: Hanchao.Zhang@aol.com | 334 E 25th st. APT 112, NY, NY, 10010

EDUCATION

New York University, Grossman School of Medicine

New York, NY

Doctor of Philosophy in Biostatistics

9/18 – 7/29

Research Interest: Machine Learning, Functional Data Analysis, Missing Data, and Microbiome

Teaching: Biostatistics I and R Lab at Weill Cornell Medicine

Johns Hopkins University, Whiting School of Engineering

Baltimore, MD

Master of Science in Applied Mathematics and Statistics

9/18 – 7/29

Courseworks: Probability Theory, Statistical Theory, Stochastic Process, Intro Algorithms, Deep Learning

Research Assistant of Dr. Michael Rosenblum and Dr. Stephanie Hicks

Cornell University, School of Medical Sciences

New York, NY

Master of Science in Biostatistics & Data Science | **Cumulative GPA:** 4.1/4.3

9/17 – 7/18

Courseworks: Categorical Data Analysis, Advanced Biostatistics, Biomedical Applications, NLP, and Statistical Learning

Research Assistant of Dr. Kalar V. Ballman and Dr. Ivan Diaz

Honor: Best Capstone Project Award

07/18

University of Michigan, Ann Arbor

Ann Arbor, MI

Exchange student in ICPSR | **Cumulative GPA:** 4.1/4.3

07/15 - 09/15

Courseworks: Regression Analysis, and Advanced Time Series Analysis

Capital University of Economics and Business (CUEB)

Beijing, CN

Bachelor of Science in International Finance, Minor in Statistics | **Cumulative GPA:** 3.8/4.0

9/13 – 6/17

PROFESSIONAL EXPERIENCES

Johns Hopkins University Bloomberg School of Public Health

Baltimore, MD

Research Assistant/Department of Biostatistics

10/18 – present

- Worked with Dr. Stephanie Hicks on her case studies in public health topics and course materials in advanced data science
- Contributed case studies to JHU case studies Github repository (<https://github.com/opencasestudies>)
- Used Empirical Efficiency Maximization to improve Covariate Adjustment built a best prediction model and evaluated the Intention-to-Treat Parameter with Dr. Rosenblum

Technical Consulting & Research, Inc.

New York, NY

Statistical Data Mining Consultant/Department of Data Consulting

6/18 – 9/18

- Extracted meaningful data (age and location of patients) from doctors' scripts using Regular Expression by R (gsub series)
- Visualized data using Tableau to build website for National Library of Medicine to help patients with specific diseases find hospitals running clinical trials for the same

AIG

New York, NY

Data Analyst Intern/Department of Investment

4/18 – 6/18

- Collected and managed dataset of U.S. dollar index, treasury bonds, effective federal fund rate, and commodities price using R
- Performed exploratory data analysis with histograms, correlation heat maps, and time series plots using R to detect data outliers and characteristics(ggplot2, ggpubr, corrplot, tidyverse)
- Utilized bootstrapping to randomly select 10,000 time-periods with different lengths to perform regression analysis; found distribution of coefficient on return rate of the ten-year treasury rate with USD Index
- Applied time series analysis and visualized past behavior patterns. Found more than 100% appreciation of effective federal rate 300 days prior to each past recession

Weill Cornell Medicine

New York, NY

Statistician Intern/Department of Healthcare Policy and Research

12/17 – 3/18

- Met with physicians at Weill Cornell Medicine and New York Presbyterian Hospital regularly
- Designed experiments, set up sample size, conducted power analysis and perform statistical analysis

ACADEMIC PROJECTS AT NEW YORK UNIVERSITY

"An Outliers Detection Algorithm for Functional Data Analysis"

02/20 – 05/20

- Developed a outliers detection algorithm for Electroencephalography (EEG) data using a random consecutive window method
- Applied the developed outliers detection algorithm on the real EEG data set, and improved the AUC from 88% to 96%
- Built an spline model on the data preprocessed by the outliers detection algorithm, and improved prediction accuracy by 12%

"Association between maternal prenatal stress and human fetal brain development"

08/19 – 10/19

- Acquired the fMRI and questionnaire data, preprocess and winsorized the survey data and fMRI data
- Applied EDA for previewing the data, and obtained the association between cortisol value and multiple stress scores

- Utilized unsupervised clustering methods (PCA and T-SNE) to cluster the patients, identify subgroups of the patients for further analysis

ACADEMIC PROJECTS AT JOHNS HOPKINS UNIVERSITY

“Covariates Adjustment for Selecting Best Model for Psychosis Improvement”

10/18 – present

- Utilized the randomized trial data in NIMH with outcome variable quality of life score
- Built best logistic model with covariates adjustment, evaluate the improvement of accuracy and intention -to-treat parameter
- Built machine learning models such as random forest, comparing it with the logistic model in improvement of accuracy
- Transformed the Model to other dataset representing different population, evaluating the difference of the intention -to-treat parameter

“Deep Learning U-Net Model in Segmentation of Brain MR Images”

11/18 – present

- Utilized the MRI data released for the MICCAI 2012 Grand Challenge on Multi-Atlas Segmentation
- Established a baseline network for two basic segmentation tasks: brain/non-brain and grey matter/white matter/cerebrospinal fluid
- Provided the network full 2D axial slices of the MR volumes for training and testing
- Interpreted the changes that are being made to the network design by exploring intermediate feature maps created in the alternate networks, looking for differences to organization of features

ACADEMIC PROJECTS AT CORNELL UNIVERSITY

“Risk Prediction and Evaluation of the Effect of Myocardial Infarction on Stroke”

03/18-07/18

- Performed data & feature engineering over 1.7 M healthcare data(Strokes) including extraction, manipulation, feature generation & selection
- Applied Survival Analysis and machine learning methods in computer clusters, including Cox Regression, Accelerated Failure model Survival Random Forest. Selected the Cox Regression without interaction as the best prediction model by Cross-Validation using R (survival, randomForestSRC, and pec) (Dr. Diaz’s Lab)

“Robust Regression for Outliers in Laboratory Data”

03/18-04/18

- Applied three robust regression methods, M-estimation, S-estimation, and MM-estimation to evaluate association between urinary PGE-M level and urinary PGD-M level in obese and lean mice, with and without celecoxib by R (MASS, ggplot2); Concluded that M-estimation is slightly better than other two robust regressions
- Improved certainty of variance analysis by leveraging each data point with weight and new loss function

“Transportation and Electronic Health Record (EHR) Database Construction”

03/18-04/18

- Established database schema and populated tables based on electronic health record and New York City transportation data from U.S. Department of transportation on MySQL server
- Applied Logistic Regression to search association of transportation preference, concluding that number of patients having Benign Essential Hypertension was associated with patients’ location in New York City and that Hyperlipidemia, and Atrial Fibrillation, was associated with patient’s transportation preference using R (stat, ggplot2, randomForest)

ACADEMIC PROJECTS AT CAPITAL UNIVERSITY OF ECONOMICS AND BUSINESS

“Optimization Model for Investment for Charity Foundation Investing in Universities”

02/14-04/14

- Built an index model used to identify American universities to invest in over five years for a variety of areas
- Led and supervised the activities of other team members for model building and client engagement purposes
- Initiated and applied for the Subjective Analytic Hierarchy Process and Objective Variation Coefficient Method to calculate the synthesis weight and know students’ performance to build the overall model
- Used our models to identify the top eleven schools which were invested in by charitable organizations
- Methods we developed were given an honorable mention (2nd place) by the organization

“SIR and Transportation model for Ebola disease”

02/15-04/15

- Built a SIR model using transportation networks to analyze infections rates of different Ebola virus strains
- Created a general wellness model that used each country’s estimated life expectancy to estimate infection spread rates in different countries.
- Analyzed optimal medicine and treatment delivery methods using transportation models.
- Given an honorable mention (2nd place) by the organization

TECHNICAL SKILLS

Programming: R (dplyr, tidyr, tibble), Python (Numpy, Pandas, NLTK, TensorFlow), SQL, SAS (Advanced Certified), Linux

Statistics: Classification (Logistic Regression, Tree-Based Methods, KNN), Regression (Linear Regression, Robust Regression, Shrinkage Regression), Clustering (K-Means, Hierarchical Clustering), Feature Engineering (PCA), and Bayes

Languages: Mandarin and English