Zhipeng Hou

929-215-1015 | zhipenghou2022@u.northwestern.edu | 2006 Sheridan Road, Evanston, IL 60208

EDUCATION

Northwestern University **Evanston, IL**
PhD Candidate in Statistics 09/2017 – Present

Teacher's College, Columbia University **New York, NY**
Master of Science in Applied Statistics 02/2017
Teacher's College Scholarship Recipient.

Shandong University, School of Mathematics **Jinan, China**
Bachelor of Science in Information Security 06/2014
Outstanding Student Scholarship Recipient.

OPEN SOURCE EXPERIENCE

Open Source Developer (Natural Language Processing) 10/2020 – 12/2020
- Wrote and tested a training script to train SPECTER (natural language processing model for scientific literature processing developed by researchers at Allen Institute for AI) via PyTorch.
- The trained model is available on Hugging Face Transformers and has been downloaded more than 5,000 times in a month by users.

INTERNSHIP EXPERIENCE

Allen Institute for AI **Seattle, WA**
Research Intern 06/2021 – 09/2021
- Advised by Dr. Doug Downey who is a leading expert in information extraction, artificial intelligence, and machine learning.
- Working on advanced research concerning the intersection of information extraction, artificial intelligence, and machine learning.

Northwestern University, Research Computing Services **Evanston, IL**
Summer Intern 07/2020 – 09/2020
- Provided high performance computer cluster and cloud computing technical assistance to Northwestern University faculty and administration.
- Conducted high performance computing (HPC) analyses and developed technical reports for Research Computing Services leadership and end users detailing HPC performance.

RESEARCH EXPERIENCE

Northwestern University **Evanston, IL**
Research on Probabilistic Literature Screening Framework for Systematic Reviews 06/2019 – Present
- Processed and analyzed scientific publications "big data" via natural language processing (NLP) models.
- Developed interactive learning algorithms that order the relevance of scientific publications with respect to specified characteristics.
- Developing estimators to measure the variation and bias introduced by respective natural language processing models and interactive learning algorithms.

Research on Cloud-based Federated Forecast System 06/2018 – 06/2019
- Designed and built a cloud-based system for managing multi-party optimization experiments, such as time series forecasting, data analysis, and machine learning.
- Designed and implemented robust forecasting algorithms based on model ensembles.

Teachers College, Columbia University **New York, NY**
Research on Causal Inference 12/2015 – 06/2017
- Investigated the performance of traditional variance estimators of causal effects after model selection.

- Developed accurate estimators via resampling-based and sandwich-type methods and examined their efficacy through simulation.

Research on Meta-analysis and Robust Variance Estimation

- Studied the implementation of robust variance estimation on meta-analysis, along with its small sample correction and application on fixed effects models in economics.

Chinese Academy of Sciences, Shenzhen Institute of Advanced Technology **Shenzhen, China**

Visiting Student on Kinematics Data Analysis 09/2014 – 03/2015

- Conducted a data analysis study on medical devices for elderly people and cerebral palsy patients.
- Designed randomized experiments for data collection based on biomechanics knowledge and statistical principles.
- Designed discriminant algorithms for fall detection and cerebral palsy rehabilitation.

Chinese Academy of Sciences, China National Supercomputing Center **Shenzhen, China**

One-month Technical Training: Special Topics on **Management of Big Data** in Data Centers. 09/2014

- Trained on big data manipulation, including technical issues, design choices, and problem-solving skills in distributed computing and systems to support big data in large-scale data centers.

Shandong University, Department of Mathematics **Jinan, China**

Research on Big Data Manipulation Algorithms 09/2012 – 07/2013

- Conducted research on the application and improvement of large integer factorization algorithms.
- Implemented algorithms to process big data and adopted distributed computing techniques to handle factorization tasks with high complexity.

MATHEMATICS MODELING AND DATA ANALYSIS EXPERIENCE

2016 Atlantic Causal Inference Conference Competition (05/2016) **New York, NY**

- Performed causal inference analysis with methods based on variable selection, counterfactual model, and inverse probability weighting.

China Undergraduate Mathematical Contest in Modeling (09/2012) **China**

- Developed a wine-ranking model based on hundreds of physicochemical indexes.
- Applied machine learning and statistical methods, including cluster analysis, principal component analysis and BP neural network, to extract valuable features and build a reliable ranking model from large unstructured data.

LEADERSHIP AND COMMUNICATION

Columbia University, School of Social Work **New York, NY**

Lead Teaching Assistant, Introduction to Statistics 08/2016 – 12/2016

- Lead a team of five teaching assistants.
- Developed numerous tables and interactive statistical graphs for students via R shiny and ggplot.

PROGRAMMING SKILLS

- **Python:** highly proficient in python programming; experienced in building cloud-based real time forecasting systems.
- **Cloud Computing:** experienced in building cloud-based apps with Docker and Kubernetes.
- **Full Stack programming:** experienced in developing data intensive web apps with Flask/Tornado, Vue.js, and MongoDB.
- **R:** highly proficient in R programming; experienced in R Shiny (for interactive data visualization), R Stan and R markdown; competent in parallel computing, optimizing R code and building R packages.
- **SPSS, HLM:** experienced in multivariate analysis and other statistics applications.
- **Matlab, C, and C++:** proficient in programming statistical learning applications.
- **Linux OS:** highly proficient with over 1.5 years' work experience as a network administrator.

R Package

Fisher, Z., Tipton, E., & Hou, Z. (2016). Robumeta: Robust Variance Meta-regression. R package (Version 1.8)

CONFERENCE PRESENTATION

Keller, B. & Hou, Z. (03/2017). On the Standard Errors of Causal Effect Estimators after Variable Selection. Presented at the 2017 Society for Research on Educational Effectiveness Spring Conference, Washington, DC.
Keller, B. & Hou, Z. (05/2017). Causal Inference after Model Selection: Simulation Results. Presented at the Columbia University Causal Inference Conference, New York, NY.