Songzhu (Sean) Zheng

23 Acorn Lane, Stony Brook, NY
zheng.songzhu@stonybrook.edu - (631)710-9042 - www.linkedin.com/in/songzhu-zheng-18132abb

Qualifications

- PhD student focusing on robust deep learning against label noise and data poisoning attack
- Experienced with machine learning modeling and statistical inference
- Problem solving, focus, persistence, excellent execution

Education Background

Stony Brook University Jun. 2022 (Expected)
PhD. in Applied Math and Statistics (GPA: 3.95/4.0)
Rice University Mar. 2017
MA. in Statistics (GPA: 3.76/4.3)
Communication University of China Jul. 2015
BS. in Statistics (GPA: 3.80/4.0, rank 2/56)

Experiences

Research Assistant Stony Brook University Aug. 2018—Present
- Design scalable, noise robust, and theoretically guaranteed deep learning models
- Develop detection method against to a generic family of trojan attack
- Perform spatial analysis to explore the association between tumor and immune cells
- Conduct hypothesis testing to support the hypothesis of the research team

Statistics Lecture Instructor Stony Brook University Jul. 2018—Dec. 2018
- AMS-102 elementary statistics: classic probability and basic hypothesis testing
- AMS-315 data analysis: hypothesis testing, ANOVA, linear regression, and R programming

Data Analyst Internship Rice Kinder Institution Jun. 2016—Aug. 2016
- Build pipelines that integrate R, SQL, ArcGIS, and geocode API to match database instances
- Develop matching algorithms that achieve over 95% instances matching accuracy

Projects and Competitions

Label Noise in DNN Research Project Aug. 2018—Present
- Propose a likelihood ratio testing procedure to correct noisy labels
- Propose a topology-based filtering procedure to collect clean data
- Establish theorems that guarantee the performance of these methodologies
- Develop algorithms that are highly scalable, easy to tune, and achieve state-of-the-art performance on both public and massive industrial level data sets

Deep Fake Detection Challenge Kaggle Competition Apr. 2020
- Extract human face images from .mp4 files that are polluted by deep fake using MTCNN
- Develop a metric learning method to detect if a video contains deep-fake frames

XTX Forecasting Challenge XTX Competition Sep. 2019
- Develop machine learning models to forecast the movement of an equity using TAQ data
- Run massive model selection procedure to pick the best classifier
- Deploy a XGBoost regression tree with DART booster and rank 76 out of 4000 submissions

Logistic Regression in Hadoop Course Project Dec. 2018
- Write MapReduce to query data from HDFS and to finish the linear algebra calculation
- Deploy logistic LASSO regression to forecast equity movement in the IBM Ambari cloud
- Implement the iteratively reweighted least square algorithm to accelerate the optimization

Publications

- “A Topological Filter for Learning with Label Noise.” Neural Information Processing Systems (NeurIPS), 2020 (*equal contribution)
- “Error-Bounded Correction of Noisy Labels.” International Conference on Machine Learning (ICML), 2020
- “A Movie Customer Satisfaction Index Model Based on Structural Equation Model.” International Journal of Arts and Technology (IJART), 2019
- “Measure Audiences' Satisfaction through User Generated Content: Satisfaction Research in Motion Picture Industry.” International Journal of Arts and Technology (IJART), 2017

Awards

- Merit Student (highest level within the university) Jun. 2015
- Outstanding Student Leader (university level) Jun. 2015
- CUMCM Math Modeling Contest (second prize in Beijing) Nov. 2014
- CUMCM Math Modeling Contest (first prize in Beijing) Nov. 2013
- Outstanding Student Scholarship (university level) Nov. 2012 and Nov. 2013

Programming Skills and Certificate

- Python (numpy, pandas, matplotlib, scikit-learn, pytorch, keras, tensorflow)
- R (ggplot2, tidyverse, spatstat, xgboost, glmnet, e1071, shiny, rhadoop, rmarkdown)
- C++, SQL, MATLAB, Microsoft Office, Linux
- CFA (level-I passed)