Ruoxi (Anna) Shang

Seattle, WA | (206) 369-9637 | rxshang@uw.edu | ruoxishang.com

Education

UNIVERSITY OF WASHINGTON
- PhD Candidate in Human-centered Design and Engineering
- Expected Aug 2025

UNIVERSITY OF CALIFORNIA, BERKELEY
- BA in Applied Mathematics, Statistics
- Aug 2016 – May 2020
- Coursework: *Abstract Algebra, Numerical Analysis, Complex Analysis, Linear Modeling, Deep Learning, Data Structures, Causal Inference, Optimization Models in Engineering, Machine Learning, Natural Language Processing, Stochastic Process*

Experience

RESEARCH ASSISTANT

Human-centered data science lab
Sep 2020 – Present
- Distributed Mentoring in Online FanFiction Community with Human-centered NLP
  - Apply Network Analysis on review interaction among authors in the online FanFiction Community
  - Perform Regression Analysis on features of review reciprocity
  - Fine-tune various state-of-art NLP models (e.g. BERT, RoBERTa, XLNet) to optimize classification performance on review categories
  - Apply Task Adaptive Pre-training on RoBERTa to boost performance

RESEARCH INTERN

Daylight Security Research Lab | Center for Long-Term Cybersecurity | UC Berkeley
May 2019 – May 2020
- Cybersecurity Imagery
  - Performed exploratory data analysis on the Cybersecurity Imagery Dataset with the purpose to identify trend, abnormity and other features from findings to provide future researchers, policymakers, and the general public with a baseline understanding of how cybersecurity is depicted overtime
  - Applied deep learning CV models pre-trained on ImageNet to extract the feature distribution from the Cybersecurity Dataset (two years of Google Image Search results).
  - Analyzed the color clustering of the dataset and detected anomalies and homogeneity in the color scheme used and objects depicted in the image representation of cybersecurity.

RESEARCH INTERN

Berkeley Institute for Data Science | UC Berkeley

Jan 2019 – Dec 2019

- Rapid Gamma-Ray Burst Localization with Deep Learning
    - Implemented a deep learning 3D convolutional neural network architecture with TensorFlow based on the idea of VoxNet to improve the data analysis pipeline for Compton telescopes (e.g. COSI, AMEGO).
    - Proposed improvements for the 3D convolution network layout and achieved an 82% decrease in RMS Angular Deviation for prediction accuracy.
    - Applied a similar methodology to localize the origin of Compton-scattered gamma-rays from the Gamma-Ray Bursts with the detection of gravitational waves.

DATA SCIENCE ASSOCIATE

Bio Tesseract

Jan 2018 – Aug 2020

- Machine Learning Powered Biogas Production Optimization
    - Trained a set of ML models (e.g. KNN, SVM, Random Forest, XGBoost, etc.) on waste input and biogas output data to build a predictive model that help biogas facilities enhance their operational productivity. model was applied to an open-source Python-based web application.
    - Collaborated with new researchers to train and optimize the previously proposed machine learning predictive system with larger available data for better results.

Publications

- De Clercq, Djavan, Ruoxi Shang et al. Machine learning powered software for accurate prediction of biogas production: A case study on industrial-scale Chinese production data. Journal of Cleaner Production, 218 (2019): 390-399.
- Ruoxi Shang, A. Zoglauer, Rapid gamma-ray burst localization aboard the e-Astrogam satellite using a 3D convolutional neural network. Presented at Bay Area Machine Learning Symposium 2019, Oct 16, San Francisco, CA.
- De Clercq, Djavan, Zongguo Wen, Fan Fei, Luis Caicedo, Kai Yuan, and Ruoxi Shang. Interpretable machine learning for predicting biomethane production in industrial-scale anaerobic co-digestion. Science of The Total Environment (2019): 134574.