Sicheng Jia

23 Stewart Cir, Centereach, NY 11720

(631) 560-4048 | jsicheng@hotmail.com | jsicheng.github.io | jsicheng | jsicheng

Education

University of California, Los Angeles *Sept 2020 - Present*

Master of Science in Computer Science *GPA: 4.0*

University of California, Los Angeles *Sept 2016 - June 2020*

Bachelor of Science in Computer Science *GPA: 3.62*

Skills

Languages: Python, C/C++, SQL, HTML/CSS/Javascript

Interests: Machine Learning, Software Development, Deep Learning, Natural Language Processing, Data Mining, Bioinformatics.

Frameworks: Numpy, Pandas, Scikit-learn, PyTorch, Tensorflow, Matplotlib, Plotly, Django

Experience

Teaching Assistant *October 2020 - Present*

University of California, Los Angeles

- Teaching assistant for Operating Systems, Logic Design of Digital Systems, and Intro to Computer Science.
- Taught students in weekly discussion sections, assisted students in office hours, and created/graded assignments and exams.

Software Engineering Intern *June 2021 - Sept 2021*

North Atlantic Industries, Bohemia, NY

- Utilized BERT, Whoosh, and Django to construct a natural language search engine for Jira issues and Confluence pages.
- Extracted and processed test data from the module board SQL server to create a data visualization and analysis tool.
- Collaborated with various teams to develop software for comparing and analyzing development/production firmware files.
- Generated extensive documentation for each application using Sphinx and Read the Docs.
- Hosted and maintained web-based applications on the company's internal network.

Projects

Ancestry Inference in Low Dimensional Space Using Autoencoders

- Assessed various dimensionality reduction techniques for mapping high dimensional genome data for ancestry visualization and inference.
- Designed novel autoencoders using convolutional and recurrent layers which outperformed traditional dimensionality reduction methods.

GATMC: Graph Attention Network Matrix Completion

- Designed a novel recommender system architecture in PyTorch using graph autoencoders and neural networks.
- Analyzed dropout effects in the cold start setting, and outperformed the baseline from GCMC on the Movielens 100k dataset.

Seed-based Weakly Supervised Named Entity Recognition

- Built, trained and tested a seed-based weakly supervised machine learning model to automatically tag named entities in text.
- Applied automatic phrase mining, discriminate topic mining, BERT and K-Means to generate, expand, and group seed words.
- Seeded and trained a Bi-LSTM-CRF model using distant supervision on the COVID-19 Twitter stream and CORD-19 dataset.

Evaluating Machine Learning Approaches for Super-Resolution MRI

- Surveyed the effectiveness of various deep neural network models for image super-resolution against traditional interpolation techniques.
- Utilized data augmentation techniques on 8,500 brain MRI images, and evaluated models using the PSNR and SSIM metrics.

Socipolicy: Predicting Covid-19 Policy Compliance using Twitter Data

- Developed a website using Python, HTML, and Django to predict a Twitter user's influence on mask wearing and vaccine acceptance likelihood.
- Exploited Tweepy/snscrape to extract COVID-19 related Tweets, and constructed a pipeline to process CMU Delphi's COVIDcast data.
- Predicted policy compliance likelihood using a Random Forest Regression model trained on Tweets, COVID-19 surveys, and Google Trends data.

Wander.io: Automatic Itinerary Recommendation Website

- Leveraged machine learning and Django to create a website that generates an itinerary based on location and time inputs from the user.
- Utilized the Google Places API to obtain attractions ratings and hotel distances, then applied K-Means to produce an optimal travel itinerary.
- Mapped out optimal routes to various attractions using the Google Maps API, and displayed the travel schedule using FullCalendar.