SIHAN MIN

2250 North Triphammer Road. Apt M3F, Ithaca, NY 14850, (+1) 650-933-6688, sm2379@cornell.edu

EDUCATION

Cornell University, Ithaca, NY GPA:3.883/4.0 *Aug 2021 - Dec 2022*
M.Eng. in Financial Engineering (with Financial **Data Science** Certificate)
University of California, Los Angeles, Los Angeles, CA GPA:3.788/4.0 *Sep 2016 - Mar 2021*
B.S. in **Computer Science** & Applied Mathematics, *Cum Laude*

Related **CS** Courses: Database Systems, Distributed System, Operating System, Data Structures and Algorithms, Object-Oriented Design, Computer Networks

Related **DS** Courses: Machine Learning, Advanced Data Mining, Computational Statistics with R, Machine Learning for Medical Imaging, Probability Theory and Mathematical Statistics, Data Analysis and Regression, Big Data Technologies, Linear Algebra, Applied Numerical Methods, Optimization, Game Theory

Related **FE** Courses: Investment and Portfolio Management, Global Financial Markets and Instruments, Derivatives Securities, Data Driven Marketing

TECHNICAL SKILLS

Programming Languages	**Java**, Python, C/C++, SQL, PHP, JavaScript/HTML/CSS, R, MATLAB, Shell
Databases	MySQL, MongoDB, Memcached, Redis
Tools and Frameworks	SpringBoot, Django, Hibernate, Spark, Hadoop, React, Maven, Docker Elasticsearch, Unix/Linux, AWS EC2, GCP
Data Science	PyTorch, TensorFlow, Scikit-Learn, Keras, OpenCV, Statistics, Optimization

WORK EXPERIENCE

UCLA ScAi Lab, Node Embedding for Large-Scaled Knowledge Graph Feb 2020 – Apr 2020
Research Assistant *Los Angeles, CA*

- Discovered a better embedding method for nodes from a large-scale, real-world Knowledge Graph named YAGO, and completed various downstream tasks such as link prediction and node classification using **PyTorch** framework
- Initialized nodes with **TransE embedding**, sampled balanced subgraph based on edge-type and neighborhood, and fed subgraphs into mini batches to train a **Relational Graph Convolutional Network (RGCN)** with embedding output

UCLA WiNG Lab, Android Chrome RRC Request Latency Measurement June 2018 – Aug 2018
Research Assistant *Los Angeles, CA*

- Calculated the latency and frequency of **RRC connection** setup during Google Chrome users' daily web browsing on Android phones with information in **JSON** format extracted from low-level network communication packages
- Analyzed connection pattern together with download bytes for different types of browsing via Excel and R
- **Decreased latency** in some web pages' loading and reloading by 0.2s by setting up RRC connection

PROJECTS

Twitter Stream Real-time Data pipelines (Spark, Kafka, Flink)

- Implemented a real-time processor with **Spark** for popular Twitter hashtags
- Designed and implemented positive/negative word monitor with **Kafka** and **Spark** (60 Tweets per second)
- Optimized the processing with **Flink** with better efficiency and suitability
- Visualized the results with **Ajax** and **Javascript** chart for 1% of all public Tweets

Configurable Web Server on Google Cloud (team of 4) Mar 2020 - June 2020
Capstone project *UCLA, Los Angeles, CA*

- Built a configurable and **scalable** web server with real-time logging in **Object-Oriented programming** via C++ and Shell.
- Developed different classes for server configuration parsing, HTTP request parsing, and multiple types of request handling with **Boost library**
- Wrote **unit and integration tests** with more than 80% test coverage.
- Deployed on **Google Cloud** for public access with robust request echoing, file serving, and status checking functionalities.

Skull Stripping Using Semi-Supervised Deep Learning Apr 2019 - June 2019
Capstone project *UCLA, Los Angeles, CA*

- Developed an automatic segmentation solution for brain MRI to keep the essential tissue with deep learning models.
- Employed OSVOS model with Tensorflow using **fully convolutional neural network (FCN)**: pre-trained parent network for basic foreground segmentation; fine-tuned the network on ground truth image pairs to minimize pixel-wise cross entropy loss.
- Used less than 20 sets of segmented MRI, and got over **90% accuracy** in pixel-wise comparison on testing sets

Movie Recommendation by Rating Prediction (team of 5) Jan 2019 - Mar 2019
Capstone project *UCLA, Los Angeles, CA*

- Built personalized **recommendation system** using user movie ratings with multiple machine learning methods in Python.
- After experiments on different model combinations, used ensemble of two best models to predict user rating: **SVD** model on user movie pairs' ratings and user-based Linear Regression with movie tags information.
- Achieved root MSE of 0.822 on over 4,000,000 user-movie pairs in final rating prediction.

Political Sentiments Analysis on Reddit Text Apr 2018 - June 2018
Capstone project *UCLA, Los Angeles, CA*

- Aggregated people's attitudes towards the two Parties and Donald Trump by **NLP** on Reddit posts and comments.
- Fit tokenized and lemmatized sentences from Reddit text into Machine Learning model (Logistic Regression) in Python, which learns to label sentiments of positive/negative towards two parties and Donald Trump.
- Combined queries to MySQL database, and visualized clear political sentiments fluctuation over states in time series graph with R.

LEADERSHIP AND ACTIVITY

Co-Founder of Free Space Dance, a fast-growing all-level dance club with over a hundred members at Cornell University