MING-CHANG CHIU

3572 S Budlong Ave, Los Angeles, CA 90007 ⋄ mingchac@usc.edu ⋄ (702) 209 6629
https://www.linkedin.com/in/eric-chiu ⋄ http://charismaticchiu.github.io/

EDUCATION

University of Southern California *May 2022 (expected)*
Ph.D. in Computer Science, Advisor: Shrikanth S. Narayanan

University of Southern California *May 2018*
M.S. in Electrical Engineering, Data Science track
Overall GPA: 3.78

National Tsing Hua University *June 2015*
B.S. in Computer Science & Electrical Engineering
Last 60 GPA: 4.05 / 4.30; Cumulative GPA: 3.82 / 4.30

RELATED COURSEWORK

Natural Language Processing, Machine Learning, Robotics, Parallel and Distributed Computation, Analysis of Algorithms, Probability Theory, Cloud Programming, Digital Signal Processing

EXPERIENCE

Information Retrieval and Data Science Group GitHub Sept. 2017 - Present
Researcher [Project: TensorFlow-trained Byte Histograms for MIME Detection] *Los Angeles, CA*

- Developing TensorFlow CNN, MLP models to extract fingerprints of file types for network security in TREC-DD data and evaluating neural network models with Apache Tika default
- Generating byte frequency analysis signatures for particular MIME types and accumulating this signature for 93 file types in the 80GB TREC-DD polar dataset
- Integrating a command-line interface that can be run on the TREC-DD-Polar data

Institute for Creative Technology Jan. 2017 - Dec. 2017
Researcher [Project: Conversation Quality Assessment] *Los Angeles, CA*

- One paper submitted to NAACL 2018
- Trained Long Short Term Memory (LSTM) deep learning model and word embeddings for behavioral modeling based on the Fisher dataset and alcoholism treatment data
- Predicted new conversation and treatment session quality reaching 75% accuracy
- Generated conversation snapshots on top of Fisher dataset and created Amazon Mechanical Turk jobs to collect objective assessments from people

Illumina, Inc. May 2017 - Aug. 2017
DevOps Applications Intern *San Diego, CA*

- Built a machine learning cron for analyzing jobs in the High Performance Computing cluster to identify "destined to fail" jobs using Sci-kit learn, reaching 95% accuracy
- Developed a deep learning daemon that constantly loads new samples from database (Hive) to train a classification model using Mini-batch update technique with TensorFlow, achieving 78% accuracy
- Maintained enterprise Atlassian Jira and Confluence; created customer workspaces for internal clients

TECHNICAL STRENGTHS

Computer Languages Python, C/C++, MATLAB, Ocaml, Java, SQL, Bash, JavaScript
Technologies & APIs TensorFlow, Pandas, NLTK, OpenMP, SK-learn, OpenCV, Hadoop, Docker, Spark

PROJECT

Kaggle: Porto Seguro's Safe Driver Prediction GitHub Nov. 2017

- Won silver medal (top 4% out of 5,332 teams)
- Interpreted the foundations of machine learning theory to teammates, and coordinated teammates' works
- Implemented XGBoost, Neural Network, lightGBM algorithms for predicting the probability that a driver will initiate an auto insurance claim in the next year
- Incorporated hierarchical interpolation and boosting techniques to produce better models

Kaggle: German Credit Risk GitHub May 2017

- Used the Python Pandas library to implement a reproducible pre-processing function for raw text data
- Designed Python Scikit-learn pipeline to automate the machine learning grid search and model selection on 20 variables
- Applied SVMs, Neural Network, Random Forrest, Dimension Reduction, etc. algorithms and achieved 75% accuracy, outperforming the baseline by 5%

Markov Chain Monte Carlo (MCMC) for optimization GitHub Nov. 2016

- Implemented MCMC Simulated Annealing procedure and 3 cooling schedules to find global minimum of Schwefel function; found best cooling schedule, achieving performance at least 10% better than the others
- Improved variance of estimation by applying 3 variance reduction methods, with each at least 5 times outperforming pure MC and best variance close to 0 (nearly perfect)
- Applied Metropolis-Hastings Algorithm to sample from arbitrary high dimensional spaces and reduce corresponding variances

The World is Changing: Finding Changes on the Street GitHub Feb. 2015 - Sep. 2015

- Constructed an image change detection model using the SIFT algorithm in MATLAB, successfully detected street view mismatches in Dash camera images with respect to preprocessed Google Street View (GFV) to provide updated information
- Applied RANSAC to re-outline the areas of mismatches in the original GFV images with an accuracy outperforming the baseline by 46%
- Devised a reusable manual labeling software and data types that recorded ground truth mismatch areas to help data collection

Dictionary Search Engine GitHub Feb. 2015 - April 2015

- Implemented PageRank and TF-IDF algorithm for Apache Hadoop in Java and constructed a search engine which prioritizes relevant links
- Deployed the MapReduce framework on 8-node distributed computers allowing massive dataset to be processed
- Devised a file system database for dictionary content retrieval by applying Apache Hbase and Hive

HONORS & AWARDS

- Silver Medal (top 4% out of 5,332 teams), Porto Seguro's Safe Driver Prediction, Kaggle, 2017
- Attending Award, Celebrating the Viterbi Algorithm Through Art, 2017
- Honorary Member of Phi Tau Phi Scholastic Society, 2015
- Excellent Study Group Award, NTHU, 2015
- National Tsing Hua University International Exchange Scholarship; Amount: $10K, 2014