Ming-Chang (Eric) Chiu

3572 S Budlong Ave, Los Angeles, CA, 90007 • http://charismaticchiu.github.io/ • +1 702-209-6629
mingchac@usc.edu • https://www.linkedin.com/in/eric-chiu • https://github.com/charismaticchiu

EDUCATION

University of Southern California (USC) — Los Angeles, CA

Ph.D. in Computer Science — Aug 2018 – May 2022 (Expected)
- Advisor: Professor Shrikanth S. Narayanan
- Areas of Specialty: Natural Language Processing & Speech Recognition

University of Southern California (USC) — Los Angeles, CA

Master of Science (M.S.) in Electrical Engineering — Aug 2016 – May 2018 (Expected)
- Advisor: Professor Stefan Scherer
- Areas of Specialty: Data Science & Digital Signal Processing
- GPA: 3.78

National Tsing Hua University (NTHU) — Hsinchu, Taiwan

Bachelor of Science (B.S.) in Computer Science & Electrical Engineering — Sep 2011 – Jun 2015
- Advisors: Professor Min Sun and Jerry Chou
- Last 60 GPA: 4.05 / 4.30; Cumulative GPA: 3.82 / 4.30

University of Minnesota — Twin Cities, MN

Exchange Student in Computer Science and Engineering — Fall 2014
- GPA: 4.0

Tsinghua University — Beijing, China

Summer Exchange Student in Computer Science — Summer 2012

WORK EXPERIENCES

Information Retrieval and Data Science Group, USC Link Github — Los Angeles, CA

Research Assistant [Project: TensorFlow-trained Byte Histograms for Better MIME Detection] — Sep 2017 – Present
- 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 120 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 Technologies — Los Angeles, CA

Researcher [Project: Conversation Quality Assessment] — Jan 2017 – Dec 2017
- One paper submitted to NAACL-HLT 2018
- Trained Long Short Term Memory (LSTM) deep learning model and word embeddings for behavioral modeling based on Fisher 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. — San Diego, CA

DevOps Applications Intern — May 2017 – Aug 2017
- 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

Vision Science Lab, NTHU Github — Hsinchu, Taiwan

Research Assistant [Project: The World is Changing: Finding Changes on the Street] — Feb 2015 – Sep 2015
- Constructed image change detection model using 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 accuracy outperforming baseline by 46%
- Devised a reusable manual labeling software and data types that recorded ground truth mismatch areas to help data collection

PUBLICATION

- "Learning Dyadic Attention Networks to Predict Outcomes of Motivational Interviewing" X. Huang, L. Liu, M.-C. Chiu, J. Woolley, S. Scherer and B. Borsari. Submitted to NAACL 2018
- "Automatic and Scalable Content-based MIME Type Identification via Deep Learning" M.-C. Chiu, C. Mattmann. In Progress

SELECTED PROJECTS

Kaggle Competition: Porto Seguro's Safe Driver Prediction GitHub — Nov 2017
- Won Silver medal (top 4% out of 5,332 teams) in the biggest Kaggle competition in history as of December 2017
- 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 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 15+ variables
- Applied SVMs, Neural Network, Random Forrest, Dimension Reduction, etc. algorithms and achieved 75% accuracy, outperforming baseline by 5%

Markov Chain Monte Carlo (MCMC) for optimization Github — Nov 2016 – Dec 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

Dictionary Search Engine Github — Feb 2015 – Apr 2015
- Implemented PageRank and TF-IDF algorithms for Apache Hadoop in JAVA and constructed a search engine which prioritizes relevant links
- Deployed MapReduce framework on 8-node distributed computers allowing massive dataset to be processed
- Devised file system database for dictionary content retrieval by applying Apache Hbase and Hive

Autonomous Robotic Convoy System Design Github — Oct 2014 – Dec 2014
- Proposed an algorithm that allows rovers to move toward one and only one target even when doing sharp turns using C++
- Detected object distances using ultrasonic sensors and translated information into 2D surface using gnuplot as human computer interface
- Devised a paradigm to discern the original moving object while multiple static and moving objects are present

Online Task Re-scheduling using Machine Learning — Jul 2013 – Aug 2014
- Automated Hadoop benchmark (HiBench) to test performance of processing 8 types of computing job on 2 heterogeneous clusters using Perl
- Designed testing environment settings using Linux BASH shell scripts and analytically found suitable disk for certain computing job types
- Applied machine learning algorithm on Linux resource usage to discern types of computing and then moved the data to either Hard Disk or Solid-State Disk to proceed and so reduce power consumption

ACADEMIC EXPERIENCE

Grader, University of Southern California — Spring 2018
CS599 Special Topics: Content Detection and Analysis for Big Data

SKILLS

Languages: Python, C/C++, MATLAB, Ocaml, Java, HTML/CSS, PHP, SQL, JavaScript
Technologies % API: Hadoop, Docker, TensorFlow, OpenMP, OpenCV, Spark

RELATED COURSEWORK

Natural Language Processing, Machine Learning, Robotics, Parallel and Distributed Computation, Analysis of Algorithms, Pattern Recognition, Probability Theory, Simulation Methods for Stochastic Systems, Digital Signal Processing

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
Attended Citadel SC The Data Open Datathon (80 out of 600 contestants)	2017
Honorary Member of Phi Tau Phi Scholastic Society (only 1 in each department a year)	2015
Awarded Excellent Study Group Award, NTHU	2015
Awarded National Tsing Hua University International Exchange Scholarship; Amount: $10K	2014

LEADERSHIP & CAMPUS ACTIVITIES

SC Ballroom & Latin Dance Team, USC
Member — Aug 2017 – Present

NTHU Orchestra, National Tsing Hua University
Vice President — Jul 2012 – Jun 2013

NTHU Student Council, National Tsing Hua University
Counselor — Jul 2012 – Jun 2013

MILITARY SERVICE

Infantry Battalion, Dongyin Area Branch, Republic Of China (Taiwan) Army — Dongyin, Matsu
Soldier — Sep 2015 – Aug 2016

REFERENCES

Professor Stefan Scherer
University of Southern California
Institute for Creative Technologies
Email: scherer@ict.usc.edu

Professor Keith Jenkins
University of Southern California
Ming Hsieh Department of Electrical Engineering
Email: jenkins@sipi.usc.edu

Mr. Robert Suarez
Illumina, Inc.
Director, Scientific Computing
Email: rsuarez@illumina.com

Professor Min Sun
National Tsing Hua University
Department of Electrical Engineering
Email: sunmin@ee.nthu.edu.tw

Professor Jerry Chou
National Tsing Hua University
Department of Computer Science
Email: jchou@cs.nthu.edu.tw

Professor Stergios Roumeliotis
University of Minnesota, Twin Cities
Department of Computer Science and Engineering
Email: stergios@cs.umn.edu