Haocheng An

anhaocheng@hotmail.com +1 (512) 431-0893 https://www.cs.utexas.edu/~henear

EDUCATION

The University of Texas at Austin *Aug 2017 - Aug 2019(Expected)*
Master of Science in Computational Science, Engineering and Mathematics GPA 3.7/4.0

The University of Texas at Austin *Jan 2014 - May 2018*
Bachelor of Science in Computer Science, CS GPA 3.9/4.0
Bachelor of Science in Mathematics, Mathematical Sciences

RELEVANT COURSEWORK

(CS) Big Data Programming, Database, Machine Learning, Neural Networks, NLP, Programming Language
(Math) Regression, Stats, Probability, Stoc process, Numerical Analysis, Math Modeling, MCMC

SKILLS

Programming Language Java (Hadoop, Spark)/C **Script Language** MATLAB/SQL/Python/HTML

WORK EXPERIENCE

Incoming Software Development Engineer I, Amazon.com, Seattle, WA *Sept 2019-*

Software Development Intern, Oracle Corporation, Boston, MA *May 2018-Aug 2018*
- Augment 12 extra metrics from meter report, remove duplicates walker result and post on Grafana
- Automate meter report to DB process by Python and write SQL queries to track server/user behavior
- Predict user number& aggregation file size using LSTM, regression and Bollinger Bands with error ~2%

Software Development Intern, Cisco Systems, Inc, Dallas, TX *Jun 2017-Aug 2017*
- Query more than 9000 result counts for each 4 nodes and 5 service ID from Kibana using Elasticsearch
- Predict count's normal interval for sparse count cases using statistical methods and ARIMA thoughts
- Develop Python micro service to alarm engineers when anomaly occur and deploy code to the Docker

Research Intern, Institute for Computational Engineering and Sciences, Austin, TX *Jun 2016-Aug 2016*
- Implement condition number estimation of matrices with dimensions 500~10000 using C and BLIS
- Improve the Matrix-Matrix multiplication from 17 GFlops to 23GFlops in C using advanced kernel
- Co-Plot the performance and accuracy of the estimations by MATLAB and compare with LAPACK

RELEVANT PROJECTS

Call Log Classification Hackathon
- Extract 5 features for all 400k call logs and digitalize it to a 400k dimension matrix
- Develop a CNN through tensorflow to analyze the matrix and achieve 97% accuracy on classification

MapReduce
- Implement Inverted Index, User Sessions by using Hadoop and Spark respectively
- Utilize AVRO Files and Hadoop to characterize the behavior of users by analyzing user session data
- Use Spark MySQL interface to get average, min, max of large data set

HONORS AND AWARDS

Early membership (Junior Elected) of Phi Beta Kappa *Nov 2016*
Nominee of Unrestricted Endowed Presidential Scholarship by Department of Mathematics *Feb 2016*