Hao Li

857-891-9461 Leo.t5@outlook.com vida42.github.io

EDUCATION

Northeastern University Boston, MA

Candidate for Master of Science in Data Analytics Engineering May 2020

- Relevant Courses: Supervised Machine Learning, Computer Networking

Alibaba Research Center for Complexity Science Hangzhou, China

Master of Science in Computer Science June 2018

- Focused Areas of Study: Network Science, Linked Prediction

Xidian University Xi'an, China

Bachelor of Science in Mathematics and Applied Mathematics July 2015

SKILLS

- Programming Languages: Python, SQL, R, C, JavaScript, HTML, CSS
- Tools: Tensorflow, Spark, Hadoop MapReduce, Nginx, Redis, Scrapy, Linux, Git

PROFESSIONAL EXPERIENCE

Orimuse Xi'an, China

Systems Engineer March - May 2014

- Developed a T-shirt customizing mobile application which has been thrived for 5 years
- Designed the database schema and split data into three shards by using horizontal partitioning
- Built MySQL Master/Slave replication, copied data on slaves and achieved Read/Write splitting
- Utilized Nginx as a load balancer to split system horizontally to run multi-backend

PROJECTS AND ACADEMIC PUBLICATIONS

Sentiment Analysis of GitHub Commit Comments Feb. - March 2019

- Measured emotions expressed in different projects' commit comments by using VADER, one lexical sentiment analysis method, wrote a module by using sqlite3 to connect to SQLite and run query in Python
- Built one analyzer using numpy, pandas, texttable and so on in Python to study the relationship between the expressed emotions and different factors such as used programming language

Movie Recommendation System Feb. 2019

- Predicted ratings for movies with Tensorflow Content-based filtering and Spark Collaborative filtering, then recommended movies to users

Adult Income Prediction Nov. - Dec. 2018

- Implemented Classification tree, Boosting tree, Random forests and Neural network by Python to achieve our goal in classifying whether an individual's annual salary is more than $50,000
- Estimated imbalanced data with synthetic sampling technique, SMOTE, by Python and improved accuracy of all models to 86% and higher

Raw Sockets *HTTPGET* Application Oct. 2018

- Created a Socket program in Python that takes a URL on the command line and downloads the associated file and saves it to the current directory
- Rebuilt the system's TCP/IP stack in socket, implemented all features of IP/TCP packets

Measuring Diversity of Music Tastes in Online Musical Society Oct. - Dec. 2017

- Produced a true diversity measurement by Python to better capture diversity of users' musical tastes
- Discovered factors (levels of economic development and so on) that greatly impact users' music tastes, paper accepted by International Journal of Modern Physics C as first author

Distributed WebCrawler and User Behavior Analytics for Xiami Music Sept. - Dec. 2017

- Established Python-based web crawler engine by combining Scrapy with BeautifulSoup and Requests in Python and collected information of more than three million users
- Designed MySQL database to manage 2GB data, generated six relational tables under 3NF