

YUZHU ZHANG

Phone: (408)610-0380 Email: yuzhuz@cs.cmu.edu

OBJECTIVE

To pursue a summer intern as a software engineering developer, utilizing my expertise in programming and data analytics

EDUCATION

Carnegie Mellon University, School of Computer Science, United States

December 2016

M.S. in Intelligent Information Systems, Language Technologies Institution (QPA:3.7)

Selected Courses: Data Mining*,Texting Mining*,NLP*,Big Data System Practice*, Machine Learning, Search Engine

City University of Hong Kong, Hong Kong

May 2015

B.S. Honors in Computing Mathematics & Minor in Computer Science (CGPA 3.85, CS: 4.0)

Mainland Student First Class Scholarship, Dean's List

TECHNICAL STRENGTHS

Programming Languages

Java, C++, Python, Linux bash shell, SQL,MapReduce

Tools & System

Git,AWS,Sklearn, Jupyter, Hadoop, HBase, HTTPClient, Google Protobuf

Database

MySQL, Sqlite, PostgreSQL

EXPERIENCE & INTERN

Lucene Based Search Engine

September 2015 - December 2015

- Performed object oriented design and implemented in Java a text-based search engine based on Apache Lucene API on corpus of 500,000+ documents from ClueWeb09 dataset
- Evaluated the performance of different retrieval models such as Ranked Boolean, Okapi BM25, Indri
- Designed features and supported Feature Based Retrieval(Learning to rank) using pairwise machine learning ranking algorithm (SVM rank)
- Supported pseudo relevance feedback, query expansion, sequential dependency model (SDM), MMR to improve the retrieval accuracy

Distributed Key Value Store Extension

Oct 2015

- Implemented the PUT/GET request for a distributed key-value store with sharding or replication schemes on AWS platform in Java
- Implemented the coordinator to support strong consistency

Hotbox - Large Scale Machine Learning Database

September 2015-ongoing

- Implemented feature transformations in C++ on large scale data: e.g. using XGBoost to transform and store data into boost tree model
- Extended the parser function to support for CSV data format input feature file in C++ and Python

Kernel Based Principal Component Analysis

September 2014-April 2015

- Employed polynomial and gaussian kernel PCA on USPS handwriting data to reduce data dimensionality in Matlab
- Tested the performance with linear SVM with an accuracy of 89.3%, increase the computing speed by 30% by using C source mex-file

BGI, Inc

May 2013 - August 2013

Software Engineer Intern

Shengzhen, China

- Implemented the RNA Network prediction algorithms on mutual information under Gaussian Mixture model in C++
- Compared the performance of Logistic Regression and Gaussian Mixture Model algorithms using ROC curves
- Improved the computation efficiency and increased speed by 50% with GNU Scientific Library C++ API in Linux