

Hang Zhao

✉ hangz1@andrew.cmu.edu • 🌐 hsparrow.github.io

Eduation

Carnegie Mellon University

Master of Science Candidate, Information Technology

Renmin University of China

Bachelor of Engineering, Information Management and Information System

City University of Hong Kong

Exchange program, Department of Information Systems

Pittsburgh, U.S.

Sep 2017 -

Beijing, China

Sep 2010 - June 2014

Hong Kong

Sep 2012 - Dec 2012

Project Experience

Data Visualization Framework for Sentiment Analysis

Oct 2018 - Nov 2018

- Built a black-box framework visualizing sentiment analysis results via Google Natural Language API
- Designed a scalable plugin mechanism supporting multiple data sources and data visualizations
- Implemented 3 data plugins and 3 visualization plugins for reviews from IMDB, Twitter, Amazon

Question Answering System

Oct 2018 - Dec 2018

- Developed a QA system to generate 100-150 questions or to answer given question for a specific Wiki passage.
- Questions are generated by the syntactic parse tree matching with the pre-defined question format trees. The wh-question word is determined by the classification on the NER results.
- The answering system sorts the tokenized sentences by the correlation to the question, which is defined by the cosine similarity, Bag of Words(BOW) with tf-idf weights given to words.
- Applied Seq2Seq on question generation, training on the SQuAD dataset with the word vectors from Conceptnet Numberbatch, beam search as decoder, and Tensorflow as the backend.

Second-order Hidden Markov Model for PoS tagging

Nov 2018 - Dec 2018

- Implemented a second-order HMM training on the Penn Treebank.
- Applied deleted interpolation method to avoid the zero probability with unseen trigram in the test data.
- Implemented second-order Viterbi algorithm and the Backward algorithm for decoding.
- Evaluated on data in English, Japanese, and Bulgarian. Reduced 3.4% sentence-level error on English data compared to baseline model.

Twitter Data Analytics System

Jan 2018 - May 2018

- Developed a high performance, fault-tolerant web service analyzing over 1 TB Twitter data and supporting different queries
- Processed data with an ETL pipeline from AWS S3 to data warehouse via MapReduce
- Implemented the system with Java Vert.x framework and AWS load balancer for the front end and both MySQL and HBase as back end database
- Applied different methods to profile and optimize the database and achieved an average of 13000 QPS

Analytics Engine on Cloud

Jan 2018 - May 2018

- Built web service for predicting input text and auto-completing query by analyzing 27GB Wikipedia data using MapReduce
- Built Twitter social graph on Spark and applied a modified PageRank algorithm to identify the most influential user, and optimized the graph analysis using GraphX by considering the second-degree influential score
- Developed a streaming pipeline with Apache Kafka and Samza for New York cab matching with GPS data

Topic modeling on Mobile Applications

Nov 2017 - Dec 2017

- Modeled 180,000 Android applications from Android app store with spaCy
- Applied clustering to discover latent topics and optimized their categories with LDA model
- Visualized the document-topic proportion matrix in 173-Dimensions into 2-Dimensions using t-SNE

Face Identification System for CMU Reception (Performance Rank 1/21)

Nov 2017 - Dec 2017

- Built a reception system with OpenCV, recognizing and identifying faces with real-time tracking. It could correctly identify 4 faces at the same time in the same frame.
- Developed the system in an MVC style with professional UI implemented by JavaFX.

Work Experience

Gas Turbine Establishment of China(GTE)

Sichuan, China

Assistant Aeroengineer

July 2014 - Apr 2016

- o Measured and monitored aero-pneumatic and aero-acoustic signals in different aero-engine tests over 1,000 hours, and processed acquired signals to detect the anomaly status like stall and surge
- o Designed and tested phased microphone arrays and tunnel acoustic mode experiments on combustion chamber and compressor, successfully identify the acoustic sources
- o Solved a workbench noise problem by recognizing the noise resonance and the source of the noise, saving millions budget to rebuild it
- o Wrote a literature review on aero-engine acoustic testing for Journal of International Aviation
- o Designed and assembled an apparatus to measure the angular displacement of aero-engine guide vanes with angle encoders, which decreased the system error compared to previous angle adjustment apparatus

Skills

Programming: Python, Java, C/C++, Scala, Bash

Tools: Deep Learning (PyTorch/Keras/Tensorflow), Cloud Computing (Spark/Hadoop/Samza/Kafka), Cloud Platform (AWS/Azure/GCP), NoSql (HBase/MongoDB), Docker (Kubernetes), Software Development (Gradle/Maven/Travis CI/UML)