Guangbo Yu

guangboyu@alumni.usc.edu · 213-422-4717 · GuangboYu · Los Angeles

Education

University of Southern California (USC), Los Angeles 2015 – 2017

M.S. in Computer Science (GPA: 3.2/4.0)
Relevant Coursework: Machine Learning, Data Mining, Analysis of Algorithms, Applied Probability.

University of Electronic Science and Technology of China, Chengdu, China 2011 – 2015

B.E. in Software Engeering (GPA: 3.7/4.0, Top 5%)

Projects

Byte Cup Machine Learning Competition. Aug. 2016 – Dec. 2016

- An international Machine Learning Competition of match community raised questions with domain experts held by IEEE-China and Toutiao
- Constructed a 2-layer **Stacking** model, the first layer used **Factorization Machine (FM)**, **LR** and **XGBoost** as the base model, trained, merged and generated meta-features
- The second layer extracted **SVD, TSNE, NMF** dimension reduction information using the features of the first layer **FM** model and combined the meta-features from the first-layer to train **XGBoost** model.
- Combined the Stacking and **Collaborate Filter** model with a weight of 2:1 to generate the final model and improved the result by 9%

Surgery Blood Cell Prediction Jan. 2017 – May. 2017

- Preprocessed the dataset including filling missing data, outlier detection and data cleaning
- Conducted feature engineering including label encoding, log transformation to minimize the skewness
- Visualized the data and conducted feature selection
- Built and tuned a **Random Forest** model to increased the MAE by 102% compared to Benchmark

Movie Recommender System Aug. 2017 – Sep. 2017

- Built a recommender system to recommend movies based on an adapted Netflix user dataset via **Hadoop**
- Computed top 5 recommendations for each user (**Item-based Collaborative Filtering, JAVA**)
- Processed 1GB data by Hadoop MapReduce jobs in the environment set up by **Docker**

Weenix OS Jan. 2016 – May. 2016

- Built a Mini Unix-like OS kernel written in **C** in Linux Environment
- Implemented key components including process management, drivers for terminals and hard disks, VFS, and page-based virtual memory
- Implemented system calls including fork, waitpid, execve, open, read/write, mmap, sbrk, etc

Twitter Sentiment Analysis Jan. 2016 – May. 2016

- Extracted live streaming twitter by **Python** and **Oauth2** library
- Implemented the module calculating **term frequency** and **document frequency**
- Used **Apache Spark** and **Scala** to derive Tweet Sentiment Score

Skills

- Programming Languages: Python, Java, Javascript, C, Scala.
- Data & Machine Learning: Hadoop, Spark, Data Mining, Machine Learning, Numpy, Pandas, Matplotlib.

EXPERIENCE

Software Developer, New Beast Corporation Aug. 2017 – Jul. 2018

Designed and Built a data pipeline which monitors, scrapes and dedupes latest news. (**MongoDB, Redis, RabbitMQ, TF-IDF**)

- Built a single-page web application for users to browse news (**React, Node.js, RPC, SOA, JWT**)
- Implemented a click event log processor which collects users' click logs, then updates a news preference model for each user (**NLP**)
- Designed and built an offline training pipeline for news topic modeling (**Tensorflow, DNN, NLP**)
- Deployed an online classifying service for news topic modeling using the trained model

Java Software Developer Intern, Luzhou Hospital Jan. 2015 – May. 2015

Responsible for developing backend API of an E-commerce System via **SSM** framework

- Implemented QR code payment module. (**Alipay SDK**)
- Designed and implemented database schema for persistence and implemented Vo, Dao, Pojo, Mapper. (**MySQL, MyBatis**)
- Deployed and configured Nginx, Tomcat, Ftp Server and Iptables. (**Alibaba Cloud**)