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

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

B.E. in Software Engineering
Award: People's scholarship

Experience

Worcester Polytechnic Institute, Research Assistant May. 2019 – Present

Conducted data analyses of music features and explored Deep Learning models for various Music Information Retrieval topics. (**Tensorflow, Keras, Librosa**)

- Developed **pitch correction** model via Phase-Vocoder method and neural network models to make karaoke vocal in tune. Smoothing out the borders between pitches to make the voice sounds more realistic.
- Extracted music mel-spectrogram feature and built a **music genre** classifier via CNN model

Northwestern University, Research Assistant 2017 – 2018

Participated in the Liver Fibrosis project and conducted statistical analyses. (**Python, Scikit-Learn**)

- Built the correlation of the independent common features between donors and recipients.
- Implemented an algorithm to match every donor with the potential recipients.
- Built models of predicting patients' life to optimize the process of matching liver donors with recipients.

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

Developed and maintained backend API of the hospital's e-commerce system via **SSM** framework and implemented the QR code payment module with Alipay SDK.

Projects

SQuAD QuestionAnswering Challange 2019

Given a paragraph and a question about it. Trained a system to answer the question correctly by highlighting a section of the corresponding paragraph. (**Pytorch, NLTK**)

- Trained a neural network model based on BiDAF and R-net, which contains embedding, attention, modeling, and output layer.
- Tuned the hyperparameters, added more components like Self-Attention to the network and gained a 72.5% F1 score.

Surgery Blood Cell Prediction 2017

In collaboration with USC's Keck Hospital, built a Machine Learning system that optimizes the amount of blood required for surgical procedures. (**Scikit-Learn, Python, Numpy, Pandas**)

- Pre-processed the dataset by filling in missing data, detecting outliers and cleaning the data
- Conducted feature engineering and statistical analysis including label encoding, log transformation, data visualization, and feature selection
- Built and tuned a Random Forest model to increase the MAE by 102% compared to the benchmark.

Byte Cup Machine Learning Competition 2016

International Machine Learning Competition hosted by IEEE-China and Toutiao in which teams attempt to build the most accurate predictive models of community-based questions and answers.

- Constructed a 2-layer Stacking model in which the first layer used **Factorization Machine (FM), LR**, and **XGBoost** as the base model, which were merged and then generated the meta-features.
- The second layer extracted **SVD, TSNE, NMF** dimension reduction information from the first layer FM model and combined this with the meta-features from the first-layer to train an XGBoost model.

Movie Recommender System 2016

- Built a movie recommendation system system based on 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 2016

For the USC course Operating Systems taught by Dr. Bill Cheng, built a **Unix-like OS kernel** written in **C** in a Linux Environment.

- Implemented key components—Process Management, drivers for terminals and hard drives, Virtual File System and page-based Virtual Memory

Tap News, Real-Time News Scraping and Recommendation System 2015

- Implemented a data pipeline which monitors, scrapes and dedupes latest news (**MongoDB, Redis, RabbitMQ**)
- Implemented a click event log processor which collects users' click logs, then updates a news preference model for each user (**NLP, TF-IDF**)
- Designed and built an offline training pipeline for news topic modeling (**Tensorflow, CNN, NLP**)

Skills

- Programming Languages: Python, Java, Javascript, C, Scala.
- Tools: Scikit-Learn, Tensorflow, Pytorch, Keras, Numpy, Pandas, Hadoop, Spark, Matplotlib, Seaborn, Librosa