

SUMMARY

Seeking a full-time machine learning-related position.

Receiving a master of science degree in music and technology from Carnegie Mellon, with a wide range of programming experience, including **distributed system** and **machine learning**. Has 5+ years of software development experience, and 3+ years of research experience, with solid skills in Java, Python, and Matlab.

EDUCATION

- **Carnegie Mellon University** Pittsburgh, PA
Master of Science in Music and Technology 2014 – 2017
- **Beihang University** Beijing, China
Bachelor of Science in Electrical Engineering 2010 – 2014

PROJECTS

- **A Distributed File System based MapReduce Framework:**
 - **Components:** A Master server to manage metadata of the file system, and assign MapReduce tasks; Worker servers to store real data, and execute MapReduce tasks; and a Client server to submit tasks and data directories.
 - **Fault Tolerance:** Partitioned and replicated data for fault tolerance; listened to server reports to monitor faults.
 - **Serialized Computation:** Assigned map tasks and reduce tasks to multiple servers, and shuffled intermediate results with consistent hashing.
 - **Communication:** Utilized TCP socket in server communication.
 - **Application:** Deployed applications with large dataset such as ‘word count’ and ‘word suggestion with prefix’.
- **Musical Score-Performance Alignment Systems:**
 - **Musical Performance Parsing:** Built a MIDI (Musical Instrument Digital Interface) parser, which takes in musical performance with MIDI binary file format, and extracts musical information such as pitch in Java.
 - **Musical Score Training:** Trained a musical score model based on HMM (Hidden Markov Models) in Matlab.
 - **Alignment:** Implemented two on-line algorithms (forward algorithm and a Bayesian-based algorithm) to estimate current score position of the musical performance.
 - **Result:** Generated musical accompaniment based on score position estimation, and achieved a mean note-level error within 100 ms, between musical performance and accompaniment.
- **Accelerometer Data based Individual Verification:**
 - **Overview:** Attempted multiple classification schemes for individual verification from accelerometer data.
 - **Data:** Utilized raw 3-dimensional pedestrians’ accelerometer data from UCI Machine Learning Repository.
 - **Preprocessing:** Reduced data to 1-D, extracted single steps and found the most common steps with K-means.
 - **Verification:** Tested with correlation filters (MACE and OTSDF), PCA reconstruction and 1-vs-all SVM.
 - **Result:** The SVM gives the best results with 89% true positives and 1.8% false positives.
- **Game Implementations:**
 - **Super Mario:** Implemented the game with Model-View-Controller (MVC) pattern with Python GUI package.
 - **Scrabble:** Implemented the Scrabble word game with GUI following design patterns in Java.

INTERNSHIP

Data Scientist Intern, Musixmatch, 2015:

- **Data:** Crawled Billboard charts in Python, retrieved lyrics with Musixmatch API, organized data with MongoDB.
- **Analysis:** Computed TF-IDF with Python scikit-learn package, to obtain key words of different eras in history.
- **Visualization:** Visualized the history with D3.js.
- **Result:** Summarized the work with a blog post, using HTML, CSS and JavaScript.

SKILLS

- **Languages:** Java, Python, C++, Matlab, JavaScript **Tools:** Git, MongoDB, Tableau, D3.js, scikit-learn
- **Knowledge:** Data structures and algorithms, Object-oriented programming, Distributed system, Machine learning, Pattern recognition, Digital signal processing