

EDUCATION

- **Columbia University** New York, NY
 • **Master of Science in Computer Science** Aug. 2018 – Dec. 2019
- **Wuhan University** Wuhan, CN
 • **Bachelor of Science in Statistics (Major GPA: 3.86)** Sep. 2014 – June. 2018

EXPERIENCE

- **China Galaxy Securities Co., LTD.** Beijing, CN
 • **Database Developer Intern** Jul. 2017 - Oct. 2017
 - **Database Optimization:**
 Optimized a data integration project for financial services group that consolidated information from third-party market data vendors' flat files into SQL Server.
 - **API for China Galaxy Mobile App:**
 Redesigned the API of database for mobile app to deliver data in a better format, and updated APIs to improve the speed and the maintainability.
- **NLP Lab, Wuhan University** Wuhan, CN
 • **Research Intern** Feb. 2017 - Jul. 2017
 - **BioCreative VI Track 5: Text mining chemical-protein interactions:**
 Developed a system to automatically detect and extract the relations between compounds/drug and genes/proteins, using Python and a deep learning framework Pytorch to process the annotated text and build a CNN-based classifying model, and reached a precision value at 60.78%, closed to the results of the best performing teams.
- **Institute of Computing Technology, Chinese Academy of Sciences** Beijing, CN
 • **Software Engineering Intern** Dec. 2016 - Feb. 2017
 - **Gait Analysis:**
 Developed a gait analysis app for Android system, which can count steps and recognize posture and action as sitting, standing, walking or running. The classifier is based on SVM. Also created a light version that can be embedded in wearable devices like smart watch or band.
 - **Teaching Assistant:**
 Served as teaching assistant for ICT Machine Learning Winter Course. Created syllabus for new course offering collaborating with Prof. Jiao Shuai, taught basic statistics knowledge, mainly about linear regression. Also offer advising for student projects.

PROJECTS

- **National Innovation Practice Project: Context-Based Sentiment Analysis in Weibo:**
 Built a model for classifying the short text based on the sentiment using Pytorch and CRF++. This process included designing network structure and loss function, and tuning parameter. Also implementing data cleaning and annotation before classification. Analyzed and visualized the affective status of net users based on the classification result using R and ggplot2, and created a report about our analysis.
- **Students Grade System:**
 Developed a desktop program that can read, transform, store and display information about students, courses and grades. Fulfilled by C++ and SQL.
- **Kaggle: Gender Recognition by Voice:**
 Created a classifier to recognize a voice record as male or female. Adopted multiple methods and compared their performance by dimensionality reduction and visualization using t-SNE algorithm and ggplot2; finally reached a precision high at 100%/99% for male/female respectively.

PROGRAMMING SKILLS

- **Languages:** C++, Python, R, Java, C, SQL, Javascript, LaTeX
- **Technologies:** Pytorch, RCpp, ggplot2, NLTK, Hadoop, React, CRF++