Yanjie He

Website: https://yanjiehe.github.io/
LinkedIn: Yanjie He

Email : heyanjie0@outlook.com
Phone: (202)733-7796
Github: YanjieHe

Education

- **The George Washington University** — Washington D.C., USA
  Master of Science in Data Analytics (Computer Science Track) — *Anticipated July. 2020*
- **Shanghai University of International Business and Economics** — Shanghai, China
  Bachelor of Arts in Economics; GPA: 3.71/4.00 — *Sept. 2013 – June. 2017*

Skills

- **Programming Languages**: C, C++, C#, Java, Scala, Python, R, SQL, Scheme/Racket, Html/CSS
- **Analytical Skills**: Data Analysis, Social Network Analysis, Natural Language Processing
- **Technologies**: Qt5, Linux, MySQL, Spring MVC, Spark, Recommender System, Gephi

Work Experience

- **George Washington University** — Washington D.C., USA
  Computational Social Scientist — *Sept. 2018 - Anticipated April. 2019*
  - **Social Network Analysis and Data Analysis**:
    - Working with Professor Vontrese Pamphile at the George Washington University School of Business.
    - Helping manage ongoing research projects related to social network analysis.
    - Applying mathematical and statistical techniques to novel data, including SIENA model.
  - **Programming**: Reviewing academic papers and used Python and R to run the models.
- **Kantar Media CIC** — Shanghai, China
  Data Engineer Intern — *July 2016 - Feb 2017*
  - **Data Collection**: Designed data collection and data cleaning solution for Chanel APAC project.
  - **Data Visualization**: Developed Data Visualization Solution for GroupM television show, which can generate Venn Graph according to given input numbers automatically.
  - **Text Mining System**: Developed text mining system with team members. The system was used by more than 50 data analysts in the company to make data analysis solution for L'Oréal, Chanel, Volkswagen, and Dell. Saved more than 10,000 dollars of outsourcing cost.
    - The system contains a rule parser and an evaluator, which are able to let the data analysts define the patterns that they want to match.
    - The system can provide basic-level functions for sentiment analysis.
    - The system processed millions of text data records every month.
  - **Sentiment Analysis**: Improved the accuracy of sentiment analysis by 17% compared to old-version tools used in the company.
    - Used jieba package in Python to cut the Chinese content into separated words.
    - Selected words as features that are important for classification by χ^2 test.
    - Applied the logistic regression to classify the sentiment of user comments.

Selected Projects

- **Compiler and Virtual Machine**: Developed a compiler for a statically typed language, a bytecode disassembler, and a virtual machine. And also, designed a bytecode instruction list, which is similar to the JVM instructions. The compiler and disassembler are written in C++ and the virtual machine is written in ANSI C.
- **Reversi A.I.**: Reversi is a board game. Due to the huge amount of possible states of the game, it is next to impossible for a computer to enumerate every situation. Therefore, I implemented the minimax algorithm with alpha-beta pruning. The project is written in C++ and has a GUI which is build using Qt5 Framework.
- **Text Co-occurrence Network Analsis for The Hunger Games**: Completed a text analysis project where implemented text co-occurrence network to visualize the relationship between the main characters in the novel *The Hunger Games*. Written the program in Python and used packages including nltk, pandas, matplotlib and wordcloud. Plotted the network by using Gephi.
- **Landscape Image Clustering based on Color Histogram**: Completed a computer vision project.
  - Wrote a web crawler to download the landscape images from the Internet.
  - Used C++ to extract color histogram as image features based on OpenCV framework
  - Finally run the K-Means model on Apache Spark using scala to cluster the data.