Pengyu(Perry) Li

Address: 2410 Shakespeare St. #113 Houston TX U.S. 77030
Tel:(+1)346-813-6257 ⋄ *Email*:ryulee2014@gmail.com

SUMMARY OF QUALIFICATIONS

Master of Computer Science student at Rice University, quick and avid learner, with diverse development experience both in industrial and research settings. Actively seeking a software engineer intern position in summer 2017.

EDUCATION

Rice University | *Master of Computer Science* May 2018(Expected) | Houston, TX
Courses: Web Development, Graduate OOP, Algorithms, Data Mining and Statistical Learning.

Tsinghua University | *B.S. in Mathematics & Physics* Jun 2016 | Beijing, China
GPA of Computer Science related courses: 3.8/4.0
Courses: Big Data System(Distributed System), Introduction to Machine Learning, Data Mining, Data Structure and Algorithms, Computer Architecture, Information Management, Computer Networks, Introduction to Artificial Intelligence.

TECHNICAL SKILLS

Languages:	C, C#, C++, Java, Python, Javascript, R, Matlab, HTML/CSS
Databases:	MySQL, Oracle, SQLite, MongoDB
Tools:	Hadoop, LaTeX, Unix/Linux
Misc.:	Data Analysis, Statistical Inference, Machine Learning

PROFESSIONAL EXPERIENCE

Microsoft Research Asia Beijing, China
Intern, Machine Learning Group Nov 2015 - Jul 2016

- Extracted data for specific NLP tasks from Bing snapshots. Deployed tasks on Microsoft's internal distributed system.
- Implemented neural language models (RNN, LSTM, etc.) with Theano framework. Built c# wrapper to integrate trained models which was original Python code into a Chinese healthcare chat robot project.
- Designed an framework to understand semantical information of short text, especially queries on search engine, by incorporating external concept clusters, which improved by 7% on NDCG metric.
- Recieved certificate of " Stars of Tomorrow", award of excellence.

School of Computing, National University of Singapore Singapore
Intern, NLP & IR Group Jun 2015 - Sep 2015

- Built a Chrome extension called Wikifier powerd by Django to do entity-linking on discussion forum of MOOC platform.
- Analyzed Coursera courses schema and crawled them down from web by designed Scrapy crawling pipeline.
- Utilized context-based probability models and annotated data which was crawled from Coursera pages to disambiguate meaning of terms in posts. Found patterns of terms indicating entities of courses schema such as *lecture, video, slides, etc* , then linked post terms with their real resources automatically.

National Key Laboratory of Intelligence, Tsinghua University Beijing, China
Undergraduate Research Assistant Dec 2014 - Jun 2015

- Noted the limitations of state-of-the-art collaborative filtering based algorithms, and designed an ensemble learning framework to improve the ranking performance of recommender system by incorporating L2R models.
- Testing performance improved by 20% on MOVIES and IMDB datasets based on NDCG metric.

SELECTED PROJECT EXPERIENCE

FTP Protocol Nov 2015

- Wrote a File Transfer Protocol(FTP) by C++, based on Socket, implementing functions(put, get, cd, ls, mkdir, etc.) required by IEEE RFC 959.
- Developed client and thread-safe server end for multi-users.

IR-light Apr 2015

- Developed a Chinese news retrieval prototype to resolve XML syntax and search from crawled HTML/XML files.
- Implemented Chinese word segmentation function and AVL tree(Java) to store content data.

Campus Regulator Aug 2014

- Developed a full-stack web app to handle daily regulation tasks on campus, including student management, class enrollment, news feed, email notification, etc.
- Wrote front-end based on Bootstrap and JQuery, and back-end based on Spring framework. Designed database schema to implement interaction between client and server . Created thread pool to handle high concurrency issues.