PENGYU(PERRY) LI

Address:Tsinghua University, Beijing, P.R.China, 100084
Tel:(+86)130-5130-8960 ⋄ *Email*:ryulee2014@gmail.com

EDUCATION

Tsinghua University June 2016(expected)
B.S. in Mathematics & Physics Junior year GPA: 3.8
(A special pilot program aimed at cultivating students' natural science abilities)
Related Courses(Grade):
Introduction to Machine Learning(93), Big Data System(99)[rank 1/120],Numerical Analysis(96), Computer Architecture(87), Computer Information Management(92), Data Structure and Algorithm(85) etc.
Research Interests:
Machine Learning, Data Mining, Information Retrieval, NLP, Data Modeling & Analysis

RESEARCH EXPERIENCE

***Research Intern*, Machine Learning Group, Microsoft Research Asia**
Oct. 2015 - Jul. 2016(expected) Advisor: *Dr. Zhongyuan Wang*

· Designed an approach to introducing explicit knowledge into embed processing with deep learning techniques.
· The work comes under a project called **Probase**. Further work is ongoing.

***Research Intern*, NLP & IR Group, National University of Singapore**
Jul. 2015 - Sep. 2015 Advisor: *Prof. Min-Yen Kan*

· Joined the Web Information Retrieval / Natural Language Processing Group (WING) as a summer intern and built a system called **Wikifier** to do entity-linking in discussion forum of MOOC platform.
· Utilized the LDA model and annotated data to disambiguate the meaning of terms in posts, and found patterns of terms with entities on the webpage such as *lecture, video, slides, etc* linking them automatically.
· My work, as part of group's project, will be submitted to Special Interest Group on Information Retrieval(*SIGIR*) in January 2016.

***Research Assistant*, Lab of Intelligence, Tsinghua University**
Dec. 2014 - Jul. 2015 Advisor: *Prof. Min Zhang*

· Noted the limitations of state-of-the-art collaborative based algorithms like *LFM, BPMF, SVD++ etc*, which only focus on prediction of rating, and designed an ensemble learning framework to combine recommender system models with *learning to rank* algorithms(ranking models in the information retrieval domain) in order to enhance the ranking performance of recommended items.
· Tested framework in different datasets and average performance increase by 30% based on NDCG metric.
· **Manuscript:** Ensemble learning in Recommender System via *Learning to Rank* Methods. (to be submitted to China Conference on Machine Learning as first author)

SELECTED COURSEWORK & PRODUCTS

TUNet, Lab μ May 2015 - Present
Position: *Developer*

- Participated in a group developing campus utilization apps aimed at making students' life in campus more convenient and efficient.
- Programed under MVC framework, implemented functions of domain layer.
- Our mobile application has been downloaded more than 10 thousands times on official pages.

Campus Regulator, (Java Language Course) Aug. 2014 - Sep. 2014
Position: *Developer*

- Developed the web application which is a school regulate system including information regulation, user registration, curriculum enrollment, discussion forum etc.
- Used J2EE and Struts framework to achieve backend logic, and used jQuery, HTML and CSS implement frontend pages.

IR-light, (Data Structure and Algorithms Course) Nov. 2014 - Dec. 2014
Position: *Developer*

- Implemented a Chinese Information Retrieval program called IR-light.
- Extracted News contents from website. Desiged methods to divide Chinese words in certain context. And used divided words find the most relevant web page content stored in high performance data structure(AVL tree).

HONORS & SCHOLARSHIPS

Study Progress Scholarship	2015
Top 5%, 10% in two Kaggle Data Science contests, respectively.	2015
2nd class prize, National Intelligent Urban and Traffic System Competition	2015
Outstanding Student Organization Scholarship	2013, 2014
Outstanding Student Volunteer of Red Cross Campus Organization	2013
2nd class prize, Chinese Mathematical Olympiad Competition	2011
1st class prize, National Olympiad in Informatics Competition	2011

SKILLS

Programming Languages	C++, Java, Python, Scala, R, Matlab
Frameworks	Struts, Django, MVC, Node.js
Databases	MySQL, Oracle, MongoDB, SQLite
Tools	Hadoop, Git, Sublime,LaTeX