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 Fundamental Mathematics & Physics | the 3rd year GPA: 3.8

Related Courses(score):

Introduction to Machine Learning(93), *Big Data System*(99),*Numerical Analysis*(96) *etc.*

RESEARCH EXPERIENCE

Research Project in Recommender System
Dec.2014 - Jul.2015

Department of Computer Science,**Tsinghua University** — Beijing, China

Position: *Research Assistant* — Advisor: *Prof. Min Zhang*

· Engaged Laboretory of Intelligence Technology in Tsinghua University and did research project about recommeder system under Prof. Min Zhang's supervision.

· Due to common limit of state-of-the-art collaborative based algorithms like *BPMF,LFM,SVD++ etc* which only focus on rated data, designed a ensemble learning framework to combine the different algorithms which have certain advantages in certain area using *learning to rank* algorithms which are ranking models in information retrievel domain.

· Tested framework in different datasets and average perfomance increse by 30% based on top 20 NDCG metric and 22% based on top 20 percision metric.

· **Manuscript:** Ensemble learning in Recommender System via *Learning to Rank* Methods.

Wikifier
Jul.2015 - Sep.2015

School of Computing, **National University of Singapore** — Singapore

Position: *Research Intern* — Advisor: *Prof. Kan Min-Yen*

· Joined in Web Information Retrieval / Natural Language Processing Group (WING) as summer intern. Implemented a system to do entity-linking in discussion forum of MOOC platform. This system called **Wikifier** aim to help MOOC users to reduce time of underderstanding concepts in discussion forum.

· Captured discussion post from *Coursera* webpage using crawler. Analysed the statistic of post data and annotated data for training model. Utilized LDA model and anotated data to disambigute the meaning of term in the post when it is mentioned certain contex such as *algorithm*. Found pattern of term which have entities on webpage such as *lecture,video,slides* etc and linked them or linked with outer knowledge base such as *wikipedia* automatically.

· This work will be submitted on Special Interest Group on Information Retrieval(*SIGIR*) 2016 next January.

Probase
Oct.2015 - Present

Machine Learning Group, **Microsoft Reseach Asia** — Beijing, China

Position: *Intern* — Advisor: *Dr. Zhongyuan Wang*

· Being granted to engage machine learning group in MSRA as research intern recently after highly selective review processing. Doing research independently.

· Tried to design an embedding mapping function to map words in corpus to vector space. Utilized the geometry relation of vectors to help machine understand the text and detect the concepts relationship.

· The whole work is under the project called **Probase**. The further work is ongoing.

SELECTED COURSEWORK & OTHER EXPERIENCE

TUNet
May 2015 - Present

Position: *Android developer*

· Participated in a group to develope campus utilization apps aim to help students' life in campus more convenient and efficent.

· Programed under MVC framework, implemented functions of domain layer.

Campus Regulator
May 2015 - Present

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, used jQuery and HTML, CSS implement frontend pages.

IR-light
May 2015 - Present

Position: *Developer*

· Implemented a Chinese Information Retrievel program called IR-light.

· Extracted information from website. Desiged methods to divide Chinese words of sentence. Try to use divided words find the most relevant web page information stored in database.

HONORS & SCHOLARSHIPS

Study Progress Scholarship

Outstanding Student Organazition Scholarship

Intelligent Urban and Achitecture System Competition 2nd Class Prize

SKILLS

Computer Languages	C++, Java, Python, Scala, R, Matlab
Methods	Data mining, Machine Learning, Distributed Computing
Databases	MySQL, Oracle, MangoDB, SQLite
Tools	Git, Sublime, Hadoop