

HAOJIE PAN

+86 1781 6855 741 ◇ myscarletpan@gmail.com ◇ github ◇ blog

EDUCATION

Bachelor of Engineering in Computer Science

Sept. 2014 - June 2018.

Zhejiang University (ZJU), Hangzhou, China. GPA: 3.94/4.0, RANK: 3/137

Prof. Deng Cai

Enrolled in He Class, a special pilot CS program led by Prof. Zhijun He for talented students

Master of Philosophy in Computer Science and Engineering

Aug. 2018 - Recent

Hong Kong University of Science and Technology (HKUST), Kowloon, Hong Kong.

Prof. Yangqiu Song

Received two years' full postgraduate studentship (PGS) award.

INTERNSHIP & EXPERIENCE

Tencent, Inc.

Nov. 2018 - Current

Algorithm Engineer Intern, Wechat

Shenzhen, China

- Explored content-based feed recommendation with social information for wechat app "Top Stories" (Kanyikan).
- Cleaned and constructed Wechat FeedRec dataset, which contains 67 million articles and 4 million users. In order to run experiment on a single-server environment, I used the idea of data partition and multi-process to extract keywords/user history and optimized the data structure for storage/access to save disk and memory.
- I achieved **65.17%** AUC score (**8.4%** improvement compared with the baseline) on the subset of this dataset only using article features from user himself and his friends.

Alibaba Inc.

Jan. 2018 - June 2018

Algorithm Engineer Intern, IDST(DAMO Academy) NLP Group

Hangzhou, China

- Mined ten billion product reviews and built a weakly supervised reviews-to-summary dataset using existing annotation for aspect extraction. It saved money for the company since annotating summaries is expensive.
- Built an aspect-aligned model based on Seq-to-seq framework and aspect attention for review summarization. It beats the baseline on human evaluation score by **7%**. The system is likely to be deployed to production in Taobao, a renowned Chinese E-commerce platform which serves 600 million customers.
- Used limited space to build an ensemble of several deep text classification models for Judicial crime prediction and law prediction. We obtained 2^{rd} in Chinese Law Cup - Judicial Artificial Intelligence Challenge.

NetEase, Inc.

Sept. 2017 - Jan. 2018

Algorithm Engineer Intern, NetEase Fuxi AI Lab

Hangzhou, China

- Built the first automatic Resume evaluation system with stable API call in every recruitment period.
- Conducted feature engineering and built model using sklearn/xgboost, reporting scores of each aspects' features by interpreting the decision paths.
- Cooperated with colleagues to deploy an API. Feedbacks from HRs in latest Recruitment are very positive, and we gain **85.7%** Accuracy (Out of our expectation) from 6359 resumes.

State Key Laboratory of CAD&CG@ZJU

Oct. 2016 - June 2018

Research Assistant

Hangzhou, China

- Built a VQA system automatically generating answers for questions related to images, based on deep learning.
- Built a PC software using QT, an Android app and deployed the system on a server. Obtained many awards.

OTHER PROJECTS

Situation Frame Extraction on Low Resource Languages (organized by NIST) July 2017 - Sept. 2017

- Built a system to automatically translate a disaster news in Tigrinya and Oromo into English, extracted situation frame sentences and then used classifiers to determine whether there's any "need" and any "issue".
- Used heuristic keywords matching to build issue classifier, almost doubled the final results of one evaluation.

SKILLS & INTERESTS

Programming Language	Python, C/C++, Java, Javascript, Matlab, SQL
Machine Learning	PyTorch, Sklearn, Pandas, Numpy, Xgboost, d3.js
Interests	Text Mining, Natural Language Processing, Data Mining

PUBLICATIONS & MANUSCRIPTS

Haojie Pan (First author), Manuscript, Long Paper, *WWW*, 2019 (submitted)

Haojie Pan (First author), Manuscript, Long Paper, *NAACL*, 2019 (submitted)

Stephen Mayhew, Chase Duncan, Mark Sammons, Mark Sammons, Chen-Tse Tsai, Xin Li, **Haojie Pan**, Sheng Zhou, Jennifer Zou, Yangqiu Song, Dan Roth. Illinois CCG LoReHLT 2017 Named Entity Recognition and Situation Frame Systems. *LoReHLT*. 2017 [pdf]

Haojie Pan, Yang Liu, Junpei Zhou, Jialin Lu, Method for Constructing Visual Question Answering System for the Disabled. *Patent of mainland China*, NO. 201710128118.6 [link]

COMPETITIONS

Paraphrase Identification

- Alibaba Ant Financial AI Competition - Financial Intelligence NLP Service [link] Team, Top 5%, 21/449
- Paipaidai 3rd Mojing Cup Competition [link] Team, Top 3%, 16/576

Text Classification

- Chinese Law Cup - Judicial Artificial Intelligence Challenge [link] Team, Top 1%, 2/202

Tabular Data Mining

- Kaggle: Two Sigma Connect: Rental Listing Inquiries [link] [code][post] Solo, Top 5%, 105/2488