

JUNRU LU

Natural Language Processing and Machine Learning enthusiast

@ lj1230@nyu.edu 📍 New York City, USA 🌐 LuJunru 🌐 My World



EDUCATION

M.S. in Applied Urban Science & Informatics

New York University

📅 Sept. 2018 – Sept. 2019 📍 New York City, USA

- **GPA:** 3.63/4.00
- **Project1:** Collect geo-tagged tweets in NYC using Twitter API and create emotional time-spatial maps through sentiment analysis [Codes].
- **Project2:** With RandomForest and LGBM models, use Yelp reviews and related user and business historical infos to predict the reviews' ratings [Codes].

B.Eng. in Information Management and Information System

University of International Relations

📅 Sept. 2014 – July 2018 📍 Beijing, China

- **GPA:** 90.6/100
- **Supervisor:** Dr. Binyang Li
- **Final Thesis:** a two-stage multi-attention Machine Reading Comprehension model (A-Reader). In A-Reader, text representation is realized with self-attention, while semantic interaction between article and question is based on self-attention and bi-attention. "Two-stage" refers to firstly use the final semantic matrix (FSM) within a binary classification model to select a best paragraph, and secondly predict the answer via pointer network with the FSM.

Double Bachelor of Economics

Peking University

📅 Sept. 2015 – July 2018 📍 Beijing, China

- **GPA:** 84.2/100
- **Major courses:** Accounting, Econometrics, Microeconomics, and Finance.

WORK EXPERIENCE

Natural Language Processing Engineer Internship

Beijing iDeepWise Artificial Intelligence Tech Ltd

📅 Mar. 2018 – June 2018 📍 Beijing, China

- **Project1:** Research on two sentence similarity models, TextCNN and Siamese-LSTM. TextCNN is to construct 2D input by calculating W2V-Cosine-Similarity of sentences' sub-units, while Siamese-LSTM is to use Manhattan distance to compare the final hidden states of the sentences encoded with LSTM [Codes].
- **Project2:** Implement a two-stage BiDAF model on Dureader Dataset. The BiDAF model is a traditional Machine Reading Comprehension model, which uses bi-LSTM to realize text representation and bi-attention on semantic interaction. "Two-stage" refers to firstly select the best paragraph with manually selected features, and secondly predict the answer via Pointer Network.

Data Mining Engineer Internship, Dept. of Text Mining

Beijing Baifendian InfoTech Ltd

📅 Oct. 2017 – Mar. 2018 📍 Beijing, China

- **Project:** Develop a single-round Community-based Chinese DeepQA System. The system consists of (Q, A) knowledge database, first-round query engine based on Elasticsearch, second-round selective modules including semantic similarity check on (New Q, Old Q) and pair quality check on (New Q, Old A), and compensatory web crawler for unknown new questions [Codes].

PUBLICATIONS

Journal Articles

- Lu, Junru et al. (2019). "Identifying User Profile by Incorporating Self-Attention Mechanism based on CSDN Data Set". In: *Data Intelligence* 1.2, pp. 160–175.

ACHIEVEMENTS

National 6th position - SMP

- The 6th position in the final-round of the 6th national-level Social Media Processing Contest. The contest asked to realize 3 tasks including keywords extraction from blogs, user interests labeling, and user growth prediction.
- For the 1st task, we used Tf-Idf, Text-rank, LDA, and manual rules combined to find key words. The 2nd task was realized by a stacking classification on document embeddings (a later improvement of 3 % was to use TextCNN and self-attention). The final task was done with a regressive stacking model [Codes].

University Scholarships

University Scholarships of University of International Relations 2015, 2016, 2017 (Top 5%)

Undergraduate Honor

2018 excellent graduation thesis and outstanding graduates of University of International Relations

STRENGTHS & SKILLS

Natural Language Processing

Machine Learning

Web Crawler

Tensorflow

Keras

Sklearn

Numpy

Pandas

PySpark

CODING

Python

Sql

Javascript (2019 Summer)



last updated: June 10th, 2019