JUNRU LU

Natural Language Processing and Machine Learning enthusiast

junru.lu@warwick.ac.uk | Coventry, UK | github.com/LuJunru | lujunru.github.io

EDUCATION

MPhil/Ph.D. candidate in Computer Science

University of Warwick

Nov. 2019 – Nov. 2023 | Coventry, UK

- **Supervisor**: **Prof. Yulan He**
- **Department**: Computer Science
- **Research Interests**: Text representation, question answering and controllable text generation. Currently, my work is around question answering under real world scenarios.

M.S. in Applied Urban Science & Informatics

New York University

Sept. 2018 – Sept. 2019 | New York City, USA

- **GPA**: 3.67/4.00
- **Center**: Center for Urban Science + Progress
- **Department**: Tandon School of Engineering
- **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]**.
- **Capstone project**: Using DID and Bayesian Network to infer causality from the increase of Uber & Lyft on NYC's parking violation. **[Codes]**.

B.Eng. in Information Management and Information System

University of International Relations

Sept. 2014 – July 2018 | Beijing, China

- **GPA**: 90.6/100
- **Supervisor**: **Prof. Binyang Li**
- **Department**: School of Information Science and Technology
- **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
- **Department**: National School of Development
- **Major courses**: Accounting, Econometrics, Microeconomics, and Finance.
- **Notes**: The **Double B.Eco.** is a program for non-economics undergraduates from PKU and other universities who are interested in economics since 2003.

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, Textrank, 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 Honor

- University Scholarships of University of International Relations 2015, 2016, 2017 (Top 5%).
- 2018 excellent graduation thesis and outstanding graduates of University of International Relations.

Enactus Word Cup

Third prize of national final and first prize of refional semi-final in 2018 Enactus World Cup.

Like a Boss Product Design Contest

Global champion, awarded with $30,000. This contest is hosted by Boss Zhipin App, aiming to invite students to improve product design for them. **[Official Website]**

STRENGTHS & SKILLS

Natural Language Processing | Machine Learning | Web Crawler

Pytorch | PySpark | SkLearn

CODING

Python
Sql
Javascript

RESEARCH & WORK EXPERIENCE

AI Algorithm Research Internship, Dept. of AI platform

Shanghai Bilibili Group

Nov. 2022 – Present Shanghai, China

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

Conference Proceedings

- [Accepted by Findings of EMNLP 2022] Lu, Junru et al. (2022). "Event-Centric Question Answering via Contrastive Learning and Invertible Event Transformation". In: *The 2022 Conference on Empirical Methods in Natural Language Processing*.
- [Phase 2 reviewing of AAAI 2023] Lu, Junru et al. (2022). "Event Knowledge Incorporation with Posterior Regularization for Event-Centric Question Answering". In: *The 37th AAAI Conference on Artificial Intelligence*.
- [Submission of EACL 2023] Lu, Junru et al. (2022). "NapSS: Paragraph-level Medical Text Simplification via Narrative Prompting and Sentence-matching Summarization". In: *The 17th Conference of the European Chapter of the Association for Computational Linguistics*.
- Cai, Junjie et al. (2020). "Estimating the effect of Uber & Lyft on parking violation in NYC". in: *Transportation Research Board Annual Meeting*.
- Lu, Junru, Gabriele Pergola, et al. (2020). "Chime: Cross-passage Hierarchical Memory Network for Generative Review Question Answering". In: *The 28th International Conference on Computational Linguistics*.