Feng Yao

+86 138-5203-6209 | yaof20@mails.tsinghua.edu.cn | yaof20.github.io

Education

Tsinghua University — Beijing, China
Master of Computational Law, joint program of CS and Law — *Sep. 2020 - Jul. 2023*
- Research: Natural Language Processing for Legal Intelligence
- Advisors: Prof Weixing Shen & Prof. Zhiyuan Liu

China University of Mining and Technology — Jiangsu, China
Bachelor of Electrical Engineering, Automation Track — *Sep. 2015 - Jul. 2020*
- GPA: 91.82/100, Ranking: 1/165
- Thesis: Question Answering System Based on Deep Learning, Advisor: Liang Zou

Australian National University — Canberra, Australia
Exchange student in CECS, fully sponsored by China Scholarship Council — *Feb. 2018 - Jul. 2018*
- GPA: 6.75/7.0, Courses: Data Structure, Computer Organization, Electronic Systems, Scientific Programming

Research and Industry Experience

Tsinghua NLP Lab — Beijing, China
Research Assistant with Prof. Zhiyuan Liu — *Feb. 2021 – Present*
- **Topic:** Information Extraction & Pretrained Models
- **Outcome 1:** Constructed **LEVEN**, a large-scale legal event detection dataset that has 108 event types and 60K+ sentences, which is fundamental and beneficial to Legal AI and Legal Empirical Studies. (**ACL 2022 Findings**)
- **Outcome 2:** Developed **OmniEvent**, a comprehensive, unified and modular event extraction toolkit, providing sota methods implementations along with unified evaluation. (Targeting at **ACL 2023**)
- **Ongoing:** Working on Efficient Paradigm of Adapting Large Pretrained Models & Legal Event Extraction.

Alibaba DAMO Academy — Beijing, China
Research Intern with Dr. Yating Zhang & Prof. Xiaozhong Liu — *Mar. 2022 – Aug. 2022*
- **Topics:** Information Retrieval & Question Answering
- **Outcome 1:** Introduced the task of Legal Evidence Retrieval, proposed a positive sample construction strategy for unsupervised dense retrieval, and applied it to legal evidence retrieval. (**AAAI 2023**)
- **Outcome 2:** Proposed EQUALS, a real-world Legal Question Answering dataset that targets solving the legal question answering task following retrieve-then-read paradigm. (Under Review)

Institute for AI and Law, Tsinghua University — Beijing, China
Research Assistant with Prof. Weixing Shen & Prof. Zhiyuan Liu — *Feb. 2021 – Dec. 2021*
- **Topic:** Information Retrieval & Legal Intelligence
- **Outcome:** Built a practical similar legal case retrieval system **MSCR** for judges in Datong, Shanxi Province.
- **Method:** Utilized PLMs to predict the critical legal elements and employ them as the guidance of case retrieval.

iFlytek Research Institute — Anhui, China
Research Intern with Dr. Xudong Dai — *Jun. 2020 – Aug. 2020*
- **Topics:** Information Association & Text Classification
- **Outcome:** Designed an ERNIE-based MRC solution to a person-specific information association task in the legal domain with different templates, and achieved **1st** place in a national LegalTech competition.

Publications

- **Feng Yao***, Chaojun Xiao*, Xiaozhi Wang, Zhiyuan Liu, Lei Hou, Cunchao Tu, Juanzi Li, Yun Liu, Weixing Shen, Maosong Sun: *"LEVEN: A Large-Scale Chinese Legal Event Detection Dataset", In Proceedings of Findings of the Association for Computational Linguistics* (**ACL Findings**), 2022 [Paper] [Code] [Poster]
- **Feng Yao**, Jingyuan Zhang, Yating Zhang, Xiaozhong Liu, Changlong Sun, Yun Liu, Weixing Shen: *"Unsupervised Legal Evidence Retrieval via Contrastive Learning with Approximate Aggregated Positive", In Proceedings of the AAAI Conference on Artificial Intelligence* (**AAAI, Main Conference**), 2023 [Paper] [Code]
- **Feng Yao***, Andong Chen*, Xinyan Zhao, Yating Zhang, Changlong Sun, Yun Liu, Weixing Shen, Tiejun Zhao: *"EQUALS: A Real-World Dataset for Legal Questions Answering via Reading Laws",* Under Review, 2023

Projects

OmniEvent: A Comprehensive, Unified and Modular Event Extraction Toolkit [Homepage]

Hao Peng*, **Feng Yao***, Xiaozhi Wang, Zhiyuan Liu, Juanzi Li, Maosong Sun *Apr. 2022 - Present*

- Work in progress and continually updating, preferred venue: **ACL 2023**.
- A powerful Event Extraction toolkit that provides comprehensive implementations of most existing SOTA methods along with unified workflows of data processing and model evaluation. [short-chn-intro]
- Provide off-the-shelf models that simultaneously support Chinese and English event extraction. [Code][Demo]
- Support large pretrained models training and inference with BMTrain.

MSCR: A Multi-View Similar Case Retrieval System [Homepage]

Feng Yao*, Qingquan Li*, Chaojun Xiao, Zhiyuan Liu, Yun Liu, Weixing Shen *Feb. 2021 – Dec. 2021*

- Work in progress and continually updating, preferred venue: **SIGIR 2023**.
- This is a demo of a sub-module in a Key National R&D Program, Chinese Ministry of Science and Technology.
- A professional case retrieval system for the judges that retrieve similar cases using an input case query instead of keywords, with the empowerment of legal element knowledge and NLP techniques.
- Meet the real-world requirements of the judges that retrieves similar cases in multiple perspectives, and has been adopted in Shanxi Datong People's Court.

Contests

Language Intelligence Challenge 2021: Information Extraction Track [Leaderboard]

Organized by CCF, CIPS, and Baidu Brain *Mar. 2021 - May 2021*

- Ranking: 3/2148, Prize: CNY 100,000
- Worked on document-level event extraction task and ranked 1st in this sub-task.
- Extracted the event arguments with MacBERT and grouped them together via core argument roles.

National LegalTech Contest for College Students [Leaderboard]

Organized by Baidu Brain *Jun. 2020 - Aug. 2020*

- Ranking: 1/269, Prize: CNY 50,000
- The task was to associate the legal elements with corresponding suspects in criminal cases.
- Formulated the task in the MRC paradigm and designed different question templates to improve the performance.

Selected Awards & Honors

- National Scholarship for Undergraduate Student (top 1%), Ministry of Education, 2016
- Scholarship for Study Abroad (1 out of 180), China Scholarship Council, 2017
- Special Prize in National University Student Social Practice and Science Contest on Energy Saving & Emission Reduction (10 out of 3190), Ministry of Education, 2017
- 1st Prize in National Undergraduate Electronics Design Contest in Jiangsu Province (top 10%), 2018
- 2nd Prize in National English Competition for College Student (top 30%), 2019

Professional Services

Program Committee Member/Reviewer

- Committee Member of Challenge of AI in Law (CAIL, the most influential Legal AI contest in China): Legal Event Detection Track (2022), Case Label Prediction Track (2021)
- Reviewer of Conference on Empirical Methods of Natural Language Processing (EMNLP), 2022

Skills

- **Programming skills:** Python, Java, C/C++, ARM Assembly
- **Languages tests:** IELTS: 7.0 (Speaking 7.5), CET-6: 649/710, CET-4: 668/710
- **Others:** LaTeX, Markdown, Linux, Shell, Anaconda