Lang Cao

E-mail: langcao2@illinois.edu
Mobile: (+1) 217-621-0532

Research Interests: Natural Language Processing, Information Extraction

EDUCATION

- **University of Illinois at Urbana-Champaign** — Urbana, USA
  Master of Computer Science — Sept. 2022 - Dec. 2023(expected)
- **Wuhan University of Technology (WHUT)** — Wuhan, China
  Bachelor of Engineering in Software Engineering — Sept. 2018 - June 2022
  GPA: 93.51/100, Rank: 1/79

EXPERIENCE

- **Bioinformatics Innovation Lab (Text Group), WHUT** — Jan. 2021 - May 2021
  Causality Extraction from Unstructured Financial Text — Wuhan, China
  Research Intern, Advised by Prof. Jing Peng
  - Manually extracted causal element data from texts for pre-processing, and aggregated into a high-quality large-scale dataset
  - Introduced conditional layer normalization and distance embedding, combined with pattern extraction to build the model that can extract five kinds of causality elements, which effectively leverages center words information
  - Adopted FGM adversarial training and semi-supervised learning for unlabeled data to further optimize the model
  - Developed a center word-based BERT-CRF with pattern extraction (CBCP) method; without further improvement, the performance on previous datasets reaches an F1 accuracy of **25.75** for causality relations extraction and of **59.37** for causality entities, which is much better than traditional BERT-CRF method with a relative improvement of **13.5%** and **30.4%**
  - Built a graph of financial event evolution demo based on the Neo4j platform, and programmed to implement the financial event evolution Q&A system
- **Bioinformatics Innovation Lab (Text Group), WHUT** — Mar. 2022 - June 2022
  Argument Mining on Non-tree Argument Structure Text — Wuhan, China
  Research Intern, Advised by Prof. Jing Peng
  - Introduce argument component self-attention mechanism to improve the model by better capturing the relationship between argument components
  - Introduce Threshold Argmax soft classification method to alleviate the problem of uneven argument relation samples in sparse argument graph
  - Developed a general argument extraction model that can analyze non-tree argument structure, which reaches $F1_{ACTC}$ of **83.7**, $F1_{ARI}$ of **69.8**, $F1_{ARTC}$ of **68.7** on the CDCP dataset, especially raising the state-of-the-art performances with a relative improvement of **3%** on $F1_{ARI}$
- **Smart Car Technology R&D Division in iFLYTEK CO. LTD.** — June 2021 - Aug. 2021
  Natural Language Processing Algorithm Engineer — Hefei, China

NLP Algorithm Engineer Intern, Mentored by Shen'an Li

- Wrote a crawler using Selenium and Requests libraries, and developed a data analysis aid that can assist in automatic data annotation and analysis
- Developed and maintained an efficient data analysis model for a large-scale unknown data optimization system which can improve other or follow-up tasks in our whole speech intelligent platform

PUBLICATIONS

- **Lang Cao**, Shihua Zhang and Juxing Chen. 2021. "CBCP: A Method of Causality Extraction from Unstructured Financial Text". In 2021 5th International Conference on Natural Language Processing and Information Retrieval (NLPIR 2021). [Paper][Github]
- Yujing Xue and **Lang Cao**. 2020. "Clustering of Functionally Related Genes Using Machine Learning Techniques". In 2021 5th International Conference on Compute and Data Analysis (ICCDA 2021). [Paper]
- Lingfei Xu, Jiaming Zhang*, **Lang Cao**, and Xinyu Hu. 2021. "Intelligent Cross-sensing Sensor Based on Deep Learning". In 2021 6th IEEE International Conference on Signal and Image Processing (ICSIP2021). [Paper]

HONORS & AWARDS

- Silver Medal, **top 5%** in Kaggle Common Lit Readability Prize (2021.8)
- **Top 2%** in Alibaba Tianchi NLP Chinese Pre-training Model Generalization Ability Challenge (2021.1)
- 2nd Prize in National College Computer Ability Challenge - Artificial Intelligence Application Contest (2021.1)
- 3rd Prize in China University Computer Contest - Network Technology Challenge (2021.7)
- 3rd Prize in Service Outsourcing Innovation & Entrepreneurship Competition (SOIEC) for Chinese Students (2021.6)
- 3rd Prize in CCF China Soft-National College Green Computing Design Competition (2020.11)
- Silver Award in China College Students' "Internet Plus" Innovation and Entrepreneurship Contest (2020.9)
- National Scholarship, WHUT (2020); Merit Student Model Honor, WHUT (2020); First-class Scholarship, WHUT (2021); Merit Student Honor, WHUT (2021); Outstanding Graduate (2022.6); Outstanding Graduation Thesis (2022.6)

SKILLS

- Programming: C/C++, Python, Java, JavaScript
- Techniques: PyTorch, PyTorch-Lightning, TensorFlow, Huggingface Transformers, Scikit-learn, NumPy, Pandas
- Others: LaTeX, Markdown, Git, SQL, Linux, Vue.js, Node.js, Django, Flask