Yong Cao

Huazhong University of Science and Technology, Wuhan, China.
+45 5273-0171 yongcao@di.ku.dk yongcaoplus.github.io update 19th Dec 2023

Research Area and Interests

Natural Language Processing, Cultural Adaptation in LLMs, Multilinguality, Dialogue System, Semantic Representation, Knowledge Base Question Answering

Education

Copenhagen University **Nov 2022 – Apr 2024**
Visiting Ph.D. student, Supervisor: Daniel Hershcovich. *Copenhagen, Denmark*

Huazhong University of Science and Technology **Sep 2018 – Jun 2024**
Ph.D. student of Computer Science, Supervisor: Min Chen. *Wuhan, China*

Chinese University of Hong Kong **Jul 2020 – Nov 2020**
Visiting Ph.D student, Supervisor: Kai Hwang. *Shenzhen, China*

Sichuan University **Sep 2014 – Jun 2018**
Bachelor of Telecommunication Engineering, Rank: 1/60 (1%). *Sichuan, China*

Experience

Ali-DAMO-AI Company **Jun 2022 – Oct 2022**
Intern of NLP Algorithm Researcher *Hangzhou, China*

- Mainly research on the Chinese geographic re-ranking, serving for location-related services such as navigation maps.

Xiaomi-AI Company **Nov 2021 – Jun 2022**
Intern of NLP Algorithm Researcher *Wuhan, China*

- Mainly research on knowledge base question answering(KBQA) optimization via information retrieval method.

Deepwisdom-AI Company **Jan 2021 – Jun 2021**
Intern of NLP Algorithm Engineering *Shenzhen, China*

- Mainly research on AutoML product development and Sequence Tagging algorithm.

Published Paper

1. **Yong Cao**, Yova Kementchedjhieva, Ruixiang Cui, Antonia Karamolegkou, Li Zhou, Megan Dare, Lucia Donatelli, Daniel Hershcovich. Cultural Adaptation of Recipes. Transactions of the Association for Computational Linguistics (TACL).
2. **Yong Cao**, Xianzhi Li, Huiwen Liu, Wen Dai, Shuai Chen, Bin Wang, Min Chen, Daniel Hershcovich, "Pay More Attention to Relation Exploration for Knowledge Base Question Answering.", In Findings of ACL 2023.
3. **Yong Cao**, Li Zhou, Seolhwa Lee, Laura Cabello, Min Chen, Daniel Hershcovich. "Assessing cross-cultural alignment between chatgpt and human societies: An empirical study", In Proceedings of the Workshop on Cross-Cultural Considerations in NLP, EACL 2023.
4. **Yong Cao**, Wei Li, Xianzhi Li, Min Chen, Guangyong Chen, Long Hu, Zhengdao Li, Hwang Kai, "Explore More Guidance: A Task-aware Instruction Network for Sign Language Translation Enhanced with Data Augmentation", In Findings of NAACL 2022.
5. **Yong Cao**, R. Wang, M. Chen, A. Barnawi, "AI Agent in Software-defined Network: Agent-based Network Service Prediction and Wireless Resource Scheduling Optimization", IEEE Internet of Things Journal, DOI: 10.1109/JIOT.2019.2950730, 2019.
6. **Yong Cao**, Ruixue Ding, Boli Chen, Xianzhi Li, Min Chen, Daniel Hershcovich, Pengjun Xie, Fei Huang. "Geo-Encoder: A Chunk-Argument Bi-Encoder Framework for Chinese Geographic Re-Ranking", arxiv 2023, EACL 2024 under review.
7. **Yong Cao**, Min Chen, Daniel Hershcovich, "Bridging Cultural Nuances in Dialogue Agents through Cultural Value Surveys", EACL 2024 under review.
8. Zhengdao Li, **Yong Cao**, Kefan Shuai, Yiming Miao, Kai Hwang, "Rethinking the Effectiveness of Graph Classification Datasets in Benchmarks for Assessing GNNs", ICLR 2024 under review.
9. Andrea Morales-Garzón, **Yong Cao**, Daniel Hershcovich, "Consistency Evaluation of Recipe Cultural Adaptation", LREC-COLING 2024 under review.

10. M. Chen, **Yong Cao**, R. Wang, Y. Li, D. Wu, Z. Liu, "DeepFocus: Deep Encoding Brainwaves and Emotions with Multi-scenario Behavior Analytics for Human Attention Enhancement", IEEE Network, Vol. 33, No. 6, 2019.
11. Rui Wang, **Yong Cao**, Adeeb Noor, Thamer A.Alamoudi, Redhwan Nour. "Agent-enabled task offloading in UAV-aided mobile edge computing", Computer Communications 149, 324-331.
12. M. Chen, Y. Jiang, **Yong Cao**, A. Y. Zomaya, "CreativeBioMan: Brain and Body Wearable Computing based Creative Gaming System", IEEE Systems, Man, and Cybernetics Magazine, Vol. 6, No. 1, pp. 14-22, Jan. 2020.
13. Li Zhou, Wenyu Chen, **Yong Cao**, etc, "MLPs Compass: What is learned when MLPs are combined with PLMs?", ICASSP 2024.
14. Tarik Alfif, Bander Alzahrani, **Yong Cao**, Reem Alotaibi, Ahmed Barnawi, Min Chen, "Generative Adversarial Network BasedAbnormal Behavior Detection in Massive CrowdVideos: A Hajj Case Study".Journal of Ambient Intelligence and Humanized Computing, 2021: 1-12.
15. Tianshu Hao, Jianfeng Zhan, Kai Hwang, **Yong Cao**, "Edge AiBench:Scenario-Based AI Benchmarking for Cloud/Edge/Device Computing", IEEE Transactions on Computers, 2023.
16. Jun Yang, Jiayi Lu, Yiming Miao, Lu Wang, Yiting Zhao, **Yong Cao**, "The Effective Recycling of Crashed Drone Based on Machine Intelligence", 14th International Wireless Communications & Mobile Computing Conference (IWCMC), 2018.

Academic Activities

Program Committee: AAAI 2022, 2023, 2024, EACL 2023, NAACL 2024, ACL 2024.
Co-Organizer: Cross-Cultural Considerations in NLP workshop @ ACL 2024.
Media: "Assessing Cross-Cultural Alignment between ChatGPT and Human Societies: An Empirical Study" was picked up by Politiken, Børsen, Ekstra Bladet, P1 Morgen, TV 2 (Denmark) and Science et Avenir (France).

Projects

Cultural Adaption in LLMs | *Copenhagen University* **Nov 2022 - Apr 2024**

- Consider the cultural difference in large language models (LLMs), and study the cultural alignment between LLMs and human societies based on human society surveys.
- Construct cultural adaption benchmark datasets, and propose solutions to the cultural adaptation on specific domains (e.g. recipe) and general domains (e.g. dialogue).

Modeling of Abnormal Behavior of Large-Scale Crowd | *Collaboration with KAUST* **Dec 2019 - Dec 2022**

- Collect large-scale crowd abnormal behavior benchmark dataset on hajj scenario.
- Design abnormal behavior classification algorithm based on the optical flow algorithm and GAN model.

Product of Sequence Tagging Based on AutoML | *Deepwisdom* **Mar 2021 - May 2021**

- Develop sequence tagging product based on AutoML, and automatically construct the task pipeline in four stages: Data EDA, offline training, testing and online prediction. The algorithm is implemented based on PyTorch and Keras framework and is deployed through Docker.
- Expand the basic operators of the model to 30+, and evaluate the product performance based on 10+ benchmark.
- Seven basic NLP SaaS services are developed based on this product.

Resume Parsing System for Multi-Source Unstructured Data | *Deepwisdom* **May 2021 - Jun 2021**

- Build PDF parsing and word parsing operator based on the company's resume data, extract information based on entity extraction algorithm and rule matching, and establish a resume parsing prototype system.
- Optimize specific rules for specific field extraction (eg. educational background and work experience), and merge the algorithm into the company's platform.

Early Warning System and Intervention Strategy for Depression | *National key R&D plan* **Nov 2018 - Jun 2021**

- Develop collection and storage scheme of multi-modal dataset from real depression patients and normal volunteers.
- The depression diagnosis model was established based on EEG signal, near-infrared signal, video and audio data, and the depression diagnosis result was realized based on prediction fusion.
- Develop intelligent follow-up system, and construct the development of psychological counseling and intervention robot.

Honors and Awards

- 2023 DAAD AInet Fellow on Human-centered AI.
- 2023 Outstanding PhD Scholarship of HUST.
- 2021 International Youth Talent Fund by Zhejiang Lab, Hangzhou, China.
- 2020 Zhixing Scholarship of HUST.
- 2019 Outstanding Student of HUST.
- 2018 Outstanding Graduates of SCU.
- 2017,2016 National Scholarship.
- 2017 Outstanding student cadre of SCU.
- 2016 Excellent Paper Award in the National Mathematical Modeling Challenge.
- 2015 Outstanding Students of SCU.