Yong Cao

Huazhong University of Science and Technology, Wuhan, China.
+45 5273-0171 yongcao@di.ku.dk yongcaoplus.github.io update 6th Nov 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**, 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.
5. **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.
6. **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.
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. 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.
14. Tianshu Hao, Jianfeng Zhan, Kai Hwang, **Yong Cao**, "Edge AiBench:Scenario-Based AI Benchmarking for Cloud/Edge/Device Computing", IEEE Transactions on Computers, 2023.
15. 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, ACL Rolling Reviewer.
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**

- Understand and be familiar with the research status of AutoML and Sequence Tagging algorithm.
- 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. Online prediction services are deployed based on Sanic framework, which can process users' input sequences in real-time.
- 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

- 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 Cadres of SCU.
- 2016 Excellent Paper Award in the National Mathematical Modeling Challenge.
- 2016 First prize in Mathematical Modeling Competition of SCU.
- 2015 Outstanding Students of SCU.

曹 勇

2018级直博生，华中科技大学计算机科学与技术学院，武汉.
136-6715-1764 yongcao_epic@hust.edu.cn yongcaoplus.github.io Update：2023-11-06

研究方向

自然语言处理，大语言模型中的文化适应，多语言，对话系统，语义表征，知识问答等。

教育背景

哥本哈根大学 **2022.10 – 2023.10**
*NLP*组 *CSC*博士联培，导师*:Daniel Hershcovich* 哥本哈根，丹麦

华中科技大学 **2018.09 – 2024.06**
计算机系统结构 直博生，成绩排名*: 1/15 (*前*1%)*，导师：陈敏 武汉，中国

香港中文大学（深圳） **2020.07 – 2020.10**
*SDS*学院 学术访问，导师：黄铠 深圳，中国

四川大学 **2014.09 – 2018.06**
通信工程 工学学士，成绩排名*: 1/60 (*前*1%)* 四川，中国

科研经历

阿里巴巴-达摩院 **2022.06– 2022.10**
语义匹配组 研究型实习生 杭州，中国

- 主要研究中国的地址重排序算法，为导航地图等与位置相关的服务提供支持，在中文地址数据集上取得最优效果.

小米-AI公司 **2021.11– 2022.06**
知识图谱组 知识问答实习生 北京，中国

- 主要研究基于信息检索框架的知识库问答（KBQA）的优化算法，在两个benchmark数据集上取得最优的效果.

深度赋智-AI公司 **2021.01– 2021.06**
自然语言处理算法部 实习生 深圳，中国

- 主要研究自动机器学习的产品开发和序列标记算法，相关产品输出专利1项.

论文发表

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, NLP顶刊).
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 (ACL, NLP顶会).
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**, 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.
5. **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 (NAACL, NLP顶会).
6. **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（JCR一区期刊；IF: 10.6）.
7. **Yong Cao**, Min Chen, Daniel Hershcovich, "Bridging Cultural Nuances in Dialogue Agents through Cultural Value Surveys", EACL 2024 under review.
8. 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（(导师一作，JCR一区期刊; IF: 9.3）.

9. 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.
10. Andrea Morales-Garzón, **Yong Cao**, Daniel Hershcovich, "Consistency Evaluation of Recipe Cultural Adaptation", LREC-COLING 2024 under review.
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. 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.
14. Tianshu Hao, Jianfeng Zhan, Kai Hwang, **Yong Cao**, "Edge AiBench:Scenario-Based AI Benchmarking for Cloud/Edge/Device Computing", IEEE Transactions on Computers, 2023.
15. 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.

学术活动

审稿: AAAI 2022, 2023, 2024, ACL Rolling Reviewer.
组织: Cross-Cultural Considerations in NLP workshop @ ACL 2024.
宣传: "Assessing Cross-Cultural Alignment between ChatGPT and Human Societies: An Empirical Study" 论文获得多家媒体报道：Politiken, Børsen, Ekstra Bladet, P1 Morgen, TV 2 (丹麦) 以及 Science et Avenir (法国).

项目经历

大型语言模型中的文化适应 | 哥本哈根大学 **2022.11-2024.4**

- 研究大型语言模型（LLMs）中的文化差异，并基于人类社会调查问卷研究LLMs与人类社会之间的文化对齐差距.
- 构建文化适应基准数据集，并提出在特定领域（如食谱）和通用领域（如对话）上研究文化适应的优化算法.

大规模人群异常行为建模与监测 | 与*KAUST*合作 **2019.12-2022.12**

- 收集并标注Hajj场景中的大规模人群异常行为基准数据集。
- 基于光流法和生成对抗模型学习正常人群运动模式，并通过阈值判别法进行异常行为识别和跟踪。

基于AutoML框架的序列标注产品开发与部署 | 深度赋智 **2021.03-2021.05**

- 调研和学习AutoML、序列标注算法的基础知识与研究现状.
- 开发AutoML序列标注产品，实现数据EDA、离线训练、离线测试与在线推理四个阶段的任务. 算法基于PyTorch框架与Keras框架实现，基于Docker部署算法、基于Sanic框架部署在线推理服务，可实时处理用户输入序列.
- 扩充模型基础算子30+，并基于10+ benchmark对产品性能进行评估，对比国内某AutoNLP平台胜率100%.
- 基于AutoML序列标注算法开发NLP基础SaaS服务7个，均合并到公司产品库并上线：分词、词性标注、关键词抽取、（通用）实体识别、（金融）实体识别、（医疗）实体识别、合同实体抽取.

多源数据的简历解析原型设计与实现 | 深度赋智 **2021.05-2021.06**

- 基于公司的简历数据构建PDF、Word解析算子，基于实体抽取算法与规则匹配构建简历解析抽取原型系统.
- 对特定字段抽取（eg.教育背景、工作经历）进行特定的规则优化，算法部署到公司人才管理平台.

抑郁症复发的预警体系建立和综合干预策略研究 | 国家重点研发计划 **2018.11-2021.06**

- 与武汉大学人民医院合作申请，负责项目多模态数据集的采集、存储方案制定与实施，全程参与项目沟通与协调工作.
- 基于脑电信号、近红外信号、视频、音频多模态数据建立抑郁症诊断模型，基于决策融合实现抑郁症诊断干预.
- 开发智能随访系统，实现心理咨询与干预机器人的开发与研究，负责NLP相关工作.

获奖及荣誉

- 2022年CSC奖学金
- 2021年之江国际青年人才基金
- 2020年华中科技大学知行奖学金
- 2019年华中科技大学三好研究生
- 2018年四川大学优秀毕业生
- 2017年四川大学优秀学生干部
- 2016, 2017年国家奖学金
- 2016年深圳杯全国数学建模挑战赛论文优秀奖
- 2016年四川大学数学建模比赛一等奖
- 2015年四川大学优秀学生
- 2015年CPA一星级志愿者证书
- 英语水平：CET4 593 / CET6 565