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 github.com/superyong2023 superyong.netlify.app

教育背景

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

华中科技大学 **2018.09 – 2023.06**

计算机系统结构 直博生，成绩排名：*1/15 (前1%)* 武汉，中国

专业课程

- 自然语言处理导论
- 人工智能与机器学习
- 数值分析
- 数据结构与算法
- 分布式计算
- 概率统计
- 高等数学
- 数字信号处理

科研经历

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

- 掌握知识问答国内外的研究进展，熟悉知识问答中的关键技术和实现细节.
- 调研基于知识图谱做问答的benchmarks，state-of-the-art方法以及复现baseline模型.
- 基于检索的方法实现知识问答，并在两个benchmark数据集（WebQSP/ComplexQuestion）上取得最优的效果,撰写并投稿论文一篇.

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

- 实习期间，主要聚焦NLP与产业界的实际需求开发，完整参与项目立项、需求分析、模型训练与部署、测试评估的全开发流程. 工作输出27项(产品/专利/预研/调研/解决方案/宣发/招聘)，其中代码14项、产品5项、专利2项、解决方案3项、论文复现3项、论文分享1项.
- 产品：完整开发序列MT任务1个，支持算子30+，效果对比竞品某大厂AUTONLP自学习平台，胜率100%. 积极承担项目组各项工作：协助接收论文PR以及各类产品调研10+项. 协助算法团队完成招聘工作，面试20+人次.
- 开发WS 7个，已上线4个，支持MT到WS自动化生产. 参与完成解决方案3个，相关产品输出专利申请2项.

香港中文大学（深圳） **2020.07 – 2020.10**

*SDS*学院学术交流 访问生 深圳，中国

- 参与深圳市物联网与智能云研究实验室进行学术交流，了解并熟悉黄铠教授团队的相关研究方向和研究成果.
- 协助管理AI大数据平台. 了解熟悉大数据平台的情况，熟悉云计算平台的硬件架构、软件架构和主要功能. 帮助黄教授管理云计算的课程，协助TA配置学生实验课需要使用到的虚拟机环境，并积极熟悉课程相关情况.
- 基于 AI station 平台和 MEC 平台设计完成相关的实验. 协助完成 AI benchmark 实验，完成4个AI应用场景定义和代码复现工作. 实验在远端云、边缘云和智能终端上进行部署，在端到端、多并发以及多场景下进行性能测试.
- 协助撰写相关论文一篇，一篇在投.

项目经历

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，输出调研报告与可行性分析.
- 开发基于AutoML的序列标注产品，自动实现包括数据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，通过Docker封装部署在公司后台产品库中. 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%，并输出产品测试文档和竞品分析报告.
- 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**

- 调研业界目前简历解析、通用信息抽取成熟的解决方案，输出调研报告与可行性分析.
- 基于公司的简历数据构建PDF解析算子、Word解析算子，基于实体抽取算法与规则匹配进行信息抽取，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，基于准确率/召回率/F1值对模型性能进行评估，算法部署到公司人才管理平台.

文本分类产品算子扩充与性能提升研究 | 深度赋智 **2021.01-2021.02**

- 调研文本分类相关的前沿技术与业界落地现状，输出可行性分析与竞品分析报告.
- 熟悉公司已有的文本分类产品代码，理解代码逻辑与功能.
- 协助完成AutoML文本分类产品的算子扩充工作与debug，并结合细粒度情感分类任务进行算法应用与性能测试.

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.
- 撰写相关研究论文两篇，一篇已发表，一篇在投.

大规模人群异常行为建模与监测 | 横向课题 **2019.12-2022.12**

- 负责项目计划书撰写与前期调研工作，明确算法需求和研发计划，参与需求分析讨论与研究方案制定.
- 主要聚焦与大规模人群异常行为识别算法的研究，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.
- 对接复杂场景下的人脸识别算法，协助完成 Tiny Face 识别任务，基于小规模数据集实现人脸增强与识别，部署算法到实际应用场景中，以实现对监控视频的身份识别与轨迹追踪（进行中）.

发表论文

1. **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, 影响因子 9.515)

2. **Yong Cao**, Wei Li, Xianzhi Li, Min Chen Zhengdao Li, Zhen Huang, Kai Hwang, "BE-SLT: A Novel BERT-Enhanced Network for Sign Language Translation". (AAAI 2022, 在投, AAAI Reviewer)

3. **Yong Cao**, Guoguang Lv, Min Chen, Miao Li, "A Novel Adaptive Semi-Supervised Learning Model for Joint Intent Detection and Slot Filling". (在投)

4. 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, pp. 70-77, 2019. (JCR 一区期刊, 影响因子 7.230, 导师一作)

5. 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.

6. Tianshu Hao, Jianfeng Zhan, Kai Hwang, **Yong Cao**, "Edge AiBench:Scenario-Based AI Benchmarking for Cloud/Edge/Device Computing". (在投)

7. 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.

专业技能

开发基础: 掌握Python，Git，Linux，Pytorch； 熟悉TensorFlow，Keras，Caffe，Docker.
专业知识: 自然语言处理前沿技术，深度学习与机器学习，数字信号处理.
研究兴趣: 低资源场景下的文本分类，多源数据的信息抽取，特定领域的对话系统构建方法.
自我认知: 具备较好的英语、数学和编程能力，熟悉NLP的基础任务和对话系统相关的经典工作和最新论文，想进一步了解工业界现状并将所学技术落地到实际的产品中.

其他经历及荣誉

- 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
- 2016年四川攀枝花盐边小学支教
- 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

Yong Cao

Huazhong University of Science and Technology, Wuhan, China.
136-6715-1764 yongcao_epic@hust.edu.cn github.com/superyong2023 superyong.netlify.app

Education

Sichuan University **Sep. 2014 – May 2018**
Bachelor of Communication Engeneering, Rank: 1/60 (1%) *Sichuan, China*

Huazhong University of Science and Technology **Sep. 2018 – May 2023**
Ph.D of Computer System Structure, Rank: 1/15 (1%) *Wuhan, China*

Relevant Coursework

- NLP Introduction
- Artificial Intelligence
- Numerical Analysis
- Data Structure
- Distribution Computing
- Digital Signal Processing
- Probability and Statistics
- Advanced Mathematics

Experience

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

- Mainly research on the construction of question and answer system based on knowledge Graph.
- Establish a question-answering model based on knowledge graph. The model is constructed by retrieval method, encoding the relationship and entity into low-dimension embedding and the reasoning is carried by Graph Neural Network, Variational Graph Auto-Encoders and Transformer.
- The model is validated on two benchmarks and SOTA results are achieved.

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

- During intership, 27 work are finished (Products / Patents / Industry-Solutions / Recruitment), including 14 codes, 5 industry-solutions, 2 patents applications, 3 model reproduction.
- Product: Complete MT task, supporting 30 + operators. Compared to another AutoNLP Product, we 100% win based on 10+ benchmark. Assist in receiving 10 + paper PR and various product researches. Help the team to complete the recruitment work, with 20 + person of interview.
- Mainly focus on Natural Language Processing applications, including named entity recognition, information extraction, auto machine learning.

Chinese University of Hong Kong **Jun 2020 – Nov 2020**
Visiting students of SDS Institute for academic exchange *Shenzhen, China*

- Mainly research on benchmark establish in different scenarios and reproducing four AI applications.The experiment is deployed on remote cloud, edge cloud and intelligent terminal to complete end-to-end, multi-concurrency and multi-scenario performance tests.
- Assist in writing one related paper.

Projects

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

- Understand and be familiar with research status of AutoML and Sequence Tagging algorithm.
- Develop sequence tagging product based on AutoML, and automatically realize the task 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 user's input sequences in real-time.
- Expand the basic operators of the model to 30+, evaluate the product performance based on 10+ benchmark.
- Six 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.

Optimization on Text Classification Product | *Deepwisdom* **Jan 2021 - Feb 2021**

- Investigate the cutting-edge technologies related to text classification and the research status of the industry.
- Assist in operator expansion and debug of text classification products. Conduct algorithm application and performance test combined with fine-grained emotion classification tasks.

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

- Develop collection and storage scheme of project multi-modal dataset from real depression patients.
- 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 decision fusion.
- Develop intelligent follow-up system, realize the development of psychological counseling and intervention robot.
- Write two related research papers, one has been published and one is under review.

Modeling of Abnormal Behavior of Large-Scale Crowd | *Transverse Project* **Dec 2019 - Dec 2022**

- Design large-scale crowd abnormal behavior classification algorithm based on optical flow and GAN model.
- Research on face recognition algorithms, assisting in deploying algorithms to practical application scenarios to realize identity recognition and trajectory tracking of surveillance video.
- One paper has been published.

Published Paper

1. **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.
2. **Yong Cao**, Wei Li, Xianzhi Li, Min Chen Zhengdao Li, Zhen Huang, Kai Hwang, "BE-SLT: A Novel BERT-Enhanced Network for Sign Language Translation". (NAACL 2022, Review stage)
3. **Yong Cao**, Guoguang Lv, Min Chen, Miao Li, "A Novel Adaptive Semi-Supervised Learning Model for Joint Intent Detection and Slot Filling". (Review stage)
4. 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, pp. 70-77, 2019.
5. 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.
6. Tianshu Hao, Jianfeng Zhan, Kai Hwang, **Yong Cao**, "Edge AiBench:Scenario-Based AI Benchmarking for Cloud/Edge/Device Computing". (Review stage)
7. 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.

Skills

- **Technologies**: Python, Git, Linux, Pytorch, TensorFlow, Keras, Caffe, Docker.
- **Direction**: Natural language processing, deep learning and machine learning, digital signal processin.
- **Interests**: Text classification in low resource scenarios, question-answering based on knowledge graph and dialogue system in specific domain.
- **Self-Evaluation**: Be well at English, mathematics and programming and familiar with the classic and the state-of-the-art papers related to NLP domain and basic tasks. Sincerely hope to further make some ideas into practice.

Other Experiences and Awards

- 2021 International youth talent fund by Zhijiang Lab, Hangzhou, China
- 2020 Zhixing Scholarship of HUST
- 2019 Outstanding Student of HUST
- 2018 Outstanding Graduates of SCU
- 2017 National Scholarship
- 2017 Outstanding student Cadres of SCU
- 2016 Excellent Paper Award in the National Mathematical Modeling Challenge
- 2016 National Scholarship
- 2016 First prize of Mathematical Modeling Competition of SCU
- 2016 Rural Supporting Education Experience
- 2015 Outstanding Students of SCU