

姚开春

☎ 183-2874-6563 ✉ yaokaichun@outlook.com 🏠 中国北京

🎓 教育经历

中国科学院大学		2016.09 – 2019.06
计算机软件与理论	计算机科学与技术学院	博士
西南交通大学		2013.09 – 2016.06
信号分析与处理	牵引动力国家重点实验室	硕士
西南交通大学		2008.09 – 2012.06
机械设计及其自动化	机械工程学院	本科

💼 工作经历

百度 Talent Intelligence Center 算法研究员 2019.07 – 至 今

主要负责人才计算、NLP、图计算等相关算法的研究与业务落地，具体包括：

- NLP 底层算法研发及工具搭建，主要应用于
 - 人才管理场景下的文本信息抽取，人才画像构建；
 - 招聘垂类知识图谱构建及应用研究。
- 图表征学习、知识增强表征学习算法研究，主要应用于
 - 人岗匹配、智能笔面试等智能招聘算法；
 - 多源异构信息下的学术人才评估。
- 文档多模态表征学习，主要应用于
 - 智能简历评估；
 - 文档级信息抽取。

🔧 项目经历

HR 知识中台项目 算法技术负责人 2020.01 – 至 今

该项目针对领域多源异构数据进行挖掘处理，构建结构化底层数据，并开发顶层接口规范，实现算法能力的输出，从而赋能简历画像，人才标签抽取，智能招聘商业化等各类业务。其主要包括以下几个部分：

- 招聘垂类知识图谱
 - 目标：基于多源异构招聘数据，构建招聘垂类知识图谱，推动智能招聘解决方案落地；
 - 覆盖互联网、金融、地产、制造等前沿行业，包含技能、岗位、企业、行业、高校等七类实体，其中实体及关系达到百万量级；
 - 实体抽取：通过人工标注小部分数据，同时利用词库远程监督的方式标注大部分数据，构建基于远程监督的实体抽取模型；
 - 关系判别：设计关系分类模型以及模板匹配的方法来实现实体关系的判别。
- 算法能力
 - 实体抽取接口：针对各类业务场景下的文本数据，抽取实体（包括技能、企业、项目等）等
 - * 采用经典 NER 模型，结合远程监督技术降低对标注数据的依赖；
 - * 采用seq2seq生成模型实现技能实体抽取，通过建模一个技能indicator 和利用层级强化学习优化模型参数从而提高召回率。
 - 智能岗位描述生成接口：根据岗位关键词自动生成岗位需求描述
 - * 训练一个神经主题模型学习岗位职责到岗位技能的映射；
 - * 设计一个自然语言生成模型，结合神经主题模型的技能映射实现技能可感知的岗位需求描述生成。
 - 面试问题生成及推荐：在面试场景中，为面试官推荐合适的面试问题

- * 设计基于远程监督的技能实体识别方法抽取互联网上（知识社区）的技能；
- * 构建 seq2seq 生成模型智能生成风格化的面试问题；
- * 设计图推荐算法，为面试官推荐面试问题。
- 简历解析接口：针对各类形式（PDF, JPEG, PNG, WORD等）的简历，进行结构化数据输出
 - * 基于文档布局信息、视觉信息以及文本信息，设计层级 Transformer 结构模型，引入新颖的自监督目标函数实现模型的预训练，通过少量的监督数据完成多模态简历解析模型的开发，其解析质量与速度行业领先；
 - * 落地百度大脑智能招聘商业化解决方案，入口地址：<https://ai.baidu.com/solution/recruitment>。

智慧校招项目 核心技术开发

2020.03 – 2022.10

该项目服务于百度内部校园招聘平台，通过研发智能招聘算法，摒弃人才招聘过程中的存在的主观偏见与不公平歧视，提升招聘效率，项目入口：<https://talent.baidu.com>。其主要包括以下几个部分：

- 简历评估与分发模块
 - 该模块通过智能分析校招候选人简历，评估候选人能力，实现“候选人-岗位”双向智能匹配，助力各部门招最合适的人。
 - 融合简历中文本语义信息、各类实体信息、简历风格信息研发多模态简历评估模型；
 - 从部门发布的岗位信息中抽取部门技能需求画像、从候选人简历中抽取候选人技能池画像；
 - 基于VSM模型对部门画像和候选人画像进行向量表征；
 - 基于余弦相似度以及画像匹配度对候选人和部门进行相似度匹配。
- 岗位推荐模块
 - 该模块为内部招聘平台上上传简历的用户推荐合适的工作岗位；
 - 采用召回+精排的模式为候选人推荐工作；
 - 精排模型：设计了一种知识增强的人岗匹配模型；
- 智能笔试模块
 - 该模块综合候选人技能和岗位需求技能差异，为每个候选人个性化推荐考题，实现千人千面；
 - 基于百度海量的query log 及百度百科数据，利用BiLSTM-CRF模型抽取并筛选技能点；
 - 基于query log中的点击数据，构建实体-链接图，然后基于所构建的图用半监督的标签传播算法进行实体关系抽取；
 - 基于构建的技能关系图，设计并实现基于技能点匹配的个性化笔试题推荐。
- 智能面试模块
 - 打造出行业内第一个可交互的虚拟面试机器人，作为第零面筛选候选者；
 - 主要技术：百度虚拟人+语音（识别+合成）+自适应试题推荐算法（CAT）。

AI学术人才项目 技术负责人

2019.07 – 2021.06

针对高端学术人才信息难以获取、学术人才信息多元异构，构建精准的人才画像和评估方法困难等问题，该项目补充了公司AI 战略下高端人才库的空缺，为公司recruiter提供高端人才sourcing入口；同时构建高端 AI学术人才画像体系和评价方法，为 HR提供人才检索服务；将学术人才搜索无缝接入HR招聘流程，支持发现领域专家人才和新星人才的需求。

- 数据获取：从DBLP、百度学术、微软学术等多个渠道学术论文信息（作者对齐）；
- 数据挖掘：自研学术论文关键词挖掘算法对学术论文领域进行划分，结合论文引用、期刊等级、论文数量构建人才画像和评价机制；
- 人才评估应用：发布[AI华人新星百强榜](#)，[AI华人青年学者榜](#)，引起多家媒体及高校报道。

论文专利

已发表论文：

[1] Kaichun Yao, Jingshuai Zhang, Chuan Qin, Peng Wang, Hengshu Zhu, Hui Xiong, Knowledge Enhanced Person-Job Fit for Talent Recruitment, in ICDE 2022: 3467-3480 (CCF A)

- [2] Chuan Qin, **Kaichun Yao**, Hengshu Zhu, Tong Xu, Dazhong Shen, Enhong Chen, Hui Xiong, Towards Automatic Job Description Generation with Capability-Aware Neural Networks, In IEEE Transactions on Knowledge and Data Engineering (IEEE TKDE), 2022 (CCF A)
- [3] **Kaichun Yao**, Chuan Qin, Hengshu Zhu, Chao Ma, Jingshuai Zhang, Yi Du, Hui Xiong, An Interactive Neural Network Approach to Keyphrase Extraction in Talent Recruitment, in CIKM 2021: 2383-2393 (CCF B)
- [4] **Kaichun Yao**, Libo Zhang, Tiejian Luo, Dawei Du, Yanjun Wu, Non-deterministic and emotional chatting machine: learning emotional conversation generation using conditional variational autoencoders, in Neural Comput. Appl. 33(11): 5581-5589 (2021) (CCF C)
- [5] **Kaichun Yao**, Libo Zhang, Dawei Du, Tiejian Luo, Lili Tao, Yanjun Wu. Dual Encoding for Abstractive Text Summarization, in IEEE Trans. Cybern. 50(3): 985-996 (2020) (CCF B)
- [6] **Kaichun Yao**, Libo Zhang, Tiejian Luo, Lili Tao, Yanjun Wu, Teaching Machines to Ask Questions, in IJCAI 2018: 4546-4552 (CCF A)
- [7] **Kaichun Yao**, Libo Zhang, Tiejian Luo, Yanjun Wu, Deep reinforcement learning for extractive document summarization, in Neurocomputing 284: 52-62 (2018) (CCF C)

在投论文:

- [1] **Kaichun Yao**, Jingshuai Zhang, Chuan Qin, Peng Wang, Hengshu Zhu, Hui Xiong, ResuFormer: Semantic Structure Understanding for Resumes via Multi-modal Pre-training, in ICDE 2022: 3467-3480 (CCF A)
- [2] Chuan Qin, Hengshu Zhu, **Kaichun Yao**, Peng Wang, and Hui Xiong. Automatic Skill-oriented Question Generation and Recommendation for Intelligent Interview. In TOIS, 2021 (CCF A)
- [3] Chengguang Du, **Kaichun Yao**, Hengshu Zhu, Hui Xiong, MPT: A Heterogeneous Graph based Meta-Path Transformer for Intelligent Recruitment, in WWW 2022: 3467-3480 (CCF A)
- [4] Chengguang Du, **Kaichun Yao**, Hengshu Zhu, Fuzheng Zhuang, Hui Xiong, Mining Technology Trends in Scientific Publications: A Graph Propagated Neural Topic Modeling Approach, in TKDE 2022: 3467-3480 (CCF A)

专利:

- [1] 姚开春, 张敬帅, 祝恒书, 秦川, 马超, 王鹏. 用于处理数据的方法、装置、设备和计算机可读存储介质, 申请号: CN202110349452.0, 申请日: 2021/03/31
- [2] 姚开春、秦川、祝恒书、马超、张敬帅, 关键词生成方法、装置、设备及存储介质, 申请号: CN202011505008.5, 申请日: 2020/12/18
- [3] 秦川, 姚开春, 祝恒书, 马超, 申大忠. 一种招聘岗位描述文本的生成方法、装置、设备及介质, 申请号: CN202010381686.9, 申请日: 2020/05/08
- [4] 张敬帅, 马超, 祝恒书, 姚开春. 技能词评估方法及装置、电子设备、计算机可读介质, 申请号: CN202010598970.1, 申请日: 2020/06/28
- [5] 马超, 张敬帅, 黄启帆, 姚开春, 王鹏, 祝恒书. 模型训练和词向量确定方法、装置、设备、介质和产品, 申请号: CN202110277972.5, 申请日: 2021/03/15
- [6] 马超, 张敬帅, 黄启帆, 姚开春, 王鹏, 祝恒书. 用于文本匹配的方法、装置、设备、存储介质和程序产品, 申请号: CN202110282699.5, 申请日: 2021/03/16
- [7] 张敬帅, 黄启帆, 马超, 祝恒书, 王鹏, 姚开春, 王晶, 用于匹配数据的方法、装置、设备和计算机可读存储介质, 申请号: CN202110262413.7, 申请日: 2021/03/09
- [8] 李满伟, 秦川, 申大忠, 祝恒书, 张敬帅, 姚开春. 联合能力模型的训练方法、能力需求匹配方法和装置, 申请号: CN202210878392.6, 申请日: 2022/07/25
- [9] 董政, 王鹏, 宋欣, 祝恒书, 姚开春. 社群发现方法、装置、电子设备和存储介质, 申请号: CN202210062376.X, 申请日: 2022/11/04
- [10] 李满伟, 秦川, 张乐, 张敬帅, 祝恒书, 王鹏, 王晶, 姚开春. 一种人岗匹配检测方法、装置、设备以及存储介质, 申请号: CN202211378386.0, 申请日: 2022/01/19
- [11] 王鹏, 祝恒书、董政, 姚开春, 秦川. 对象关系网络的构建方法、装置和电子设备, 申请号: CN202210051806.8, 申请日: 2022/01/17

其他:

[1] “百度智能招聘-智能模拟面试软件V1.0” 软件著作权

研究兴趣

- 自然语言处理包括信息抽取，对话系统，文本摘要
- 知识图谱、图神经网络、强化学习
- 数据挖掘包括人才计算，推荐算法
- 语言与视觉结合的多模态技术