



CCIR 2024

第三十届全国信息检索学术会议

The 30th China Conference on Information Retrieval, CCIR 2024

2024年10月18日~20日武汉



武汉大学
WUHAN UNIVERSITY



微信



Baidu 百度

原子回声
AtomEcho

BaYou
八友科技

目录

关于 CCIR	3
CCIR2024	
会议组织	4
日程一览	5
交通路线	6
会议程序	7
10 月 18 日日程安排	
10 月 19 日日程安排	
10 月 20 日日程安排	
大会报告	11
报告一：辩论式决策机—以人机会诊为例	
报告二：面向稀疏数据的 3D/4D 内容创作和重建	
报告三：通用人工智能的技术进展和典型应用	
多媒体检索专题报告	14
报告一：多模态内容表示与理解	
报告二：基于大小模型协同的视觉语言导航	
报告三：超图计算与立体对象检索	
青年论坛报告	17
报告一：离线强化学习推荐系统	
报告二：降低在线学习算法的探索成本与风险：从单智能体到多智能体	
报告三：离线强化学习中基于扩散模型的数据增强技术：从轨迹拼接的视角	
报告四：基于用户心理奖励模型的序列化推荐对齐方法	
工业论坛报告	21
报告一：微信搜一搜技术简介	

报告二：大语言模型时代推荐系统的发展路径及未来趋势

大会 Tutorial

23

报告一：基于大语言模型的自主智能体

报告二：知识检索增强：范式与关键技术

录用论文列表

25

关于 CCIR

信息检索旨在满足人类在互联网上快速准确地获取信息与知识的需求，研究成果将支撑国家战略决策，推动互联网和人工智能领域的发展，提升整个社会的生产效率，并对社会生活各个领域产生重大影响。全国信息检索学术会议 (CCIR) 由中国中文信息学会 (CIPS) 举办，一路伴随着中国互联网产业的成长，是信息检索领域的旗舰会议。CCIR 已成为全国范围内信息检索领域新学术和技术工作的主要交流平台。

CCIR2024

第三十届全国信息检索学术会议 (The 30th China Conference on Information Retrieval, CCIR 2024) 将于 2024 年 10 月 18-20 日在武汉举行，此次会议由中国中文信息学会信息检索专委会主办、武汉大学承办。

本次大会的主题是“大模型时代的信息检索：挑战与机遇”，会议包含一系列学术活动，除传统的海内外知名学者的大会报告、会议论文报告外，还将组织青年学者论坛以及面向热点研究问题的讲习班等交流环节。

会议组织

大会主席： 吴丹（武汉大学）
郭嘉丰（中国科学院计算技术研究所）
兰艳艳（清华大学）

程序委员会主席： 何向南（中国科学技术大学）
任昭春（荷兰莱顿大学）
唐睿明（华为）

最佳论文评审主席： 程学旗（中国科学院计算技术研究所）

组委会主席： 李晨亮（武汉大学）
张帆（武汉大学）

宣传主席： 白婷（北京邮电大学）

出版主席： 陈旭（中国人民大学）

网络主席： 艾清遥（清华大学）

青年论坛主席： 张伟楠（上海交通大学）
陈佳伟（浙江大学）

评测主席： 朱小飞（重庆理工大学）
刘伟东（中国移动通信研究院）
许静芳（微信）

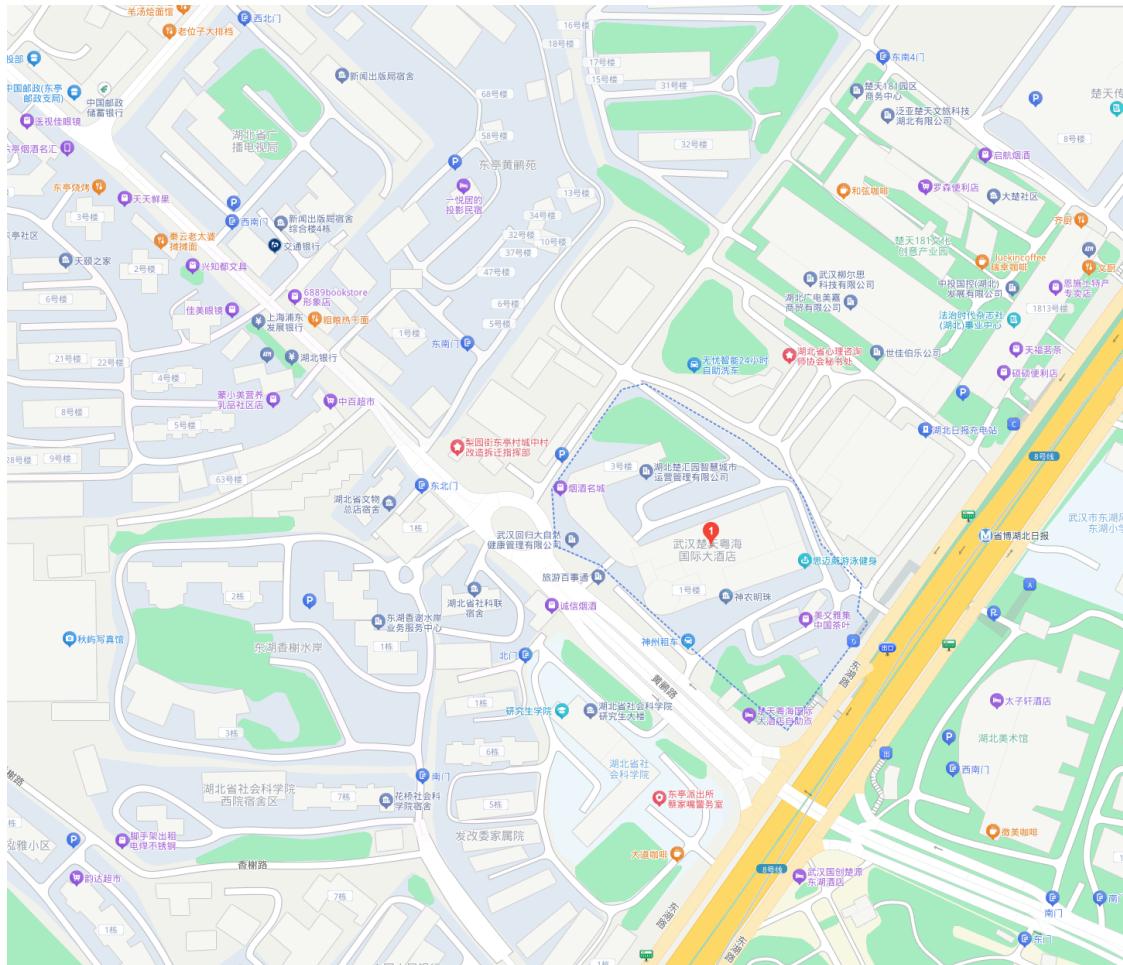
赞助主席： 王仲远（北京智源人工智能研究院）
马为之（清华大学）

财务主席： 孙培杰（南京邮电大学）

日程一览

时间	时长	10月18日 星期五
14:00-15:30	90分钟	专题讲习班1 基于大语言模型的自主智能体
15:50-17:20	90分钟	专题讲习班2 知识检索增强：范式与关键技术
17:00-19:00	120分钟	专委会工作会议
19:00-20:00	60分钟	师生互动论坛
时间	时长	10月19日 星期六
8:30-9:00	30分钟	大会开幕式
9:00-10:10	70分钟	30届纪念活动
10:10-10:25	15分钟	合影
10:25-11:25	60分钟	大会报告1
11:25-12:25	60分钟	大会报告2
12:25-13:30	65分钟	午休
13:30-15:30	120分钟	多媒体检索专题报告与圆桌交流
15:30-16:00	30分钟	Poster交流
16:00-17:30	90分钟	大会候选最佳论文报告
18:00-20:00	120分钟	晚宴
时间	时长	10月20日 星期日
8:30-9:30	60分钟	大会报告3
9:30-11:30	120分钟	青年论坛与圆桌交流
11:30-12:10	40分钟	工业论坛
12:10-13:30	80分钟	午休
13:30-15:30	120分钟	新委员论坛
15:30-16:30	60分钟	评测报告
16:30-17:00	30分钟	大会闭幕式

交通路线



会场地址：楚天粤海国际大酒店（武汉市武昌区东湖路 181 号）

- 距离天河国际机场 T2 航站楼，35 公里，约 60 分钟车程，打车预计 65 元
- 距离武汉站，12 公里，约 25 分钟车程，打车预计 25 元
- 距离武昌站，12 公里，约 35 分钟车程，打车预计 25 元
- 距离汉口站，15 公里，约 40 分钟车程，打车预计 35 元

* 注意，武汉高峰时段可能需要更长的时间，以上信息仅供参考

附近地铁站：8 号线中南医院（202 米）、4 号线青鱼嘴（924 米）

会议程序

10月18日 星期五

专题讲习班（主持人：张伟楠，陈佳伟）	
14:00-17:20	1. 基于大语言模型的自主智能体，陈旭，中国人民大学 2. 知识检索增强：范式与关键技术，王昊奋，同济大学
17:00-19:00	专委会工作会议（主持人：兰艳艳）
19:00-20:00	师生互动论坛（主持人：艾清遥，毛佳昕）

10月19日 星期六

大会开幕式（主持人：李晨亮）	
8:30-9:00	1. 武汉大学校领导致辞 2. 中文信息学会领导致辞 3. 信息检索专委会领导致辞 4. 大会主席致辞 5. 程序委员会主席致辞
9:00-10:10	30届纪念活动
10:10-10:25	合影
大会报告 1（主持人：刘奕群）	
10:25-11:25	辩论式决策机—以人机会诊为例，刘挺，哈尔滨工业大学
大会报告 2（主持人：郭嘉丰）	
11:25-12:25	面向稀疏数据的 3D/4D 内容创作和重建，田奇，华为
12:25-13:30	午休
多媒体检索专题报告（主持人：聂礼强）	
13:30-13:50	1. 多模态内容表示与理解，赵洲，浙江大学
13:50-14:10	2. 基于大小模型协同的视觉语言导航，刘偲，北京航空航天大学
14:10-14:30	3. 超图计算与立体对象检索，高跃，清华大学
14:30-15:30	多媒体交叉圆桌

15:30-16:00	Poster 交流
大会最佳候选论文报告（主持人：何向南）	
16:00-16:15	1. 孙文奇, 谢若冰, 张君杰, 赵鑫, 康战辉, 文继荣. 基于矩阵与混合机制双流长短期记忆网络的终身长序列推荐方法
16:15-16:30	2. 程佳玮, 朱小飞, 尹雅博. 基于会话有向图卷积和模态有偏融合的多模态会话情感识别
16:30-16:45	3. 蓝恒志, 黄乐, 白婷. 基于多粒度记忆 RAG 的个性化角色扮演框架
16:45-17:00	4. 严明时, 陈慧临, 程志勇. MB-HGCN: 基于层次图卷积的多行为推荐方法
17:00-17:15	5. 张开元, 叶子逸, 艾清遥, 刘奕群, 王晶, 张敏, 马少平. 面向脑电信号情绪识别任务的预训练模型设计
17:15-17:30	6. 张伟超, 张儒清, 郭嘉丰, 范意兴. 基于长距离上下文的预训练数据检测方法
18:00-20:00	晚宴

10月20日 星期日

大会报告 3 (主持人: 李晨亮)	
8:30-9:30	通用人工智能的技术进展和典型应用, 王士进, 讯飞研究院
青年学者论坛 (主持人: 张伟楠, 陈佳伟)	
9:30-9:50	1. 离线强化学习推荐系统, 辛鑫, 山东大学
9:50-10:10	2. 降低在线学习算法的探索成本与风险: 从单智能体到多智能体, 孔芳, 南方科技大学
10:10-10:30	3. 离线强化学习中基于扩散模型的数据增强技术: 从轨迹拼接的视角, 龙婷, 吉林大学
10:30-10:50	4. 基于用户心理奖励模型的序列化推荐对齐方法, 赵梦辰, 华南理工大学
10:50-11:30	圆桌交流
工业论坛 (主持人: 马为之)	
11:30-11:50	1. 微信搜一搜技术简介, 张帆, 微信
11:50-12:10	2. 大语言模型时代推荐系统的发展路径及未来趋势, 唐睿明, 华为
12:10-13:30	午休
新委员论坛 (主持人: 兰艳艳)	
13:30-14:00	林政 (中科院信工所), 陈昊 (香港理工大学), 陈旭 (中国人民大学), 程明月 (中国科学技术大学), 杜一 (中科院计算机网络信息中心), 冯福利 (中国科学技术大学), 高宸 (清华大学)
14:00-14:30	官赛萍 (中科院计算所), 郭志强 (清华大学), 胡莎 (西南大学), 黄瑾 (荷兰阿姆斯特丹大学), 雷文强 (四川大学), 李志宇 (上海算法创新研究院), 刘泽阳 (山东大学)
14:30-15:00	马登豪 (北京信息科技大学), 裴红斌 (西安交通大学), 钱泓锦 (智源研究院), 荣欢 (南京信息工程大学), 苏越阳 (中科院计算所), 陶重阳 (北京航空航天大学), 王希廷 (中国人民大学)
15:30-15:30	吴宾 (郑州大学), 徐博 (大连理工大学), 张海楠 (北京航空航天大学), 张晓堃 (香港城市大学), 周雨佳 (清华大学), 朱倩男 (北京师范大学), 邹杰 (电子科技大学)
评测报告 (主持人: 朱小飞)	
15:30-15:45	1. 评测总体情况介绍
15:45-15:50	2. 赛题一介绍
15:50-16:00	3. 赛题一冠军团队报告
16:00-16:05	4. 赛题二介绍

16:05-16:15	5. 赛题二冠军团队报告
16:15-16:20	6. 赛题三介绍
16:20-16:30	7. 赛题三冠军团队报告
大会闭幕式（主持人：兰艳艳）	
16:30-17:00	1. 评测颁奖 2. 下一届举办地宣讲 3. 大会总结

大会报告

报告一：辩论式决策机—以人机会诊为例

讲者：刘挺，哈尔滨工业大学

时间：10月19日 10:25-11:25



个人介绍 刘挺，教授，哈尔滨工业大学党委常委、副校长、自然语言处理研究所所长。国家级高层次人才，黑龙江省政协教科卫体专委会副主任，国家重点研发计划“智能机器人”重点专项指南专家。“认知智能与内容安全”教育部重点实验室主任，黑龙江省中文信息处理重点实验室主任，黑龙江省“人工智能”头雁团队带头人。中国计算机学会会士、中国中文信息学会副理事长、黑龙江省计算机学会理事长。主要研究方向为人工智能、自然语言处理、大模型、具身智能和社会计算。主持研制的“语言技术平台 LTP”、“大词林”等科研成果被业界广泛使用，主持研制了“活字”、“本草”、“巧板”等中文大模型。曾获国家科技进步二等奖、省科技进步一等奖等。

报告摘要 人工智能已由认知智能时代进入到决策智能时代，人机协同的智能决策无论是在科学研究还是实践应用都有广阔的探索空间。当前，大模型展现出强大的通用智能和泛化能力，但是在将其应用于复杂问题的决策任务时仍面临巨大的挑战。为此，本次报告将介绍我们团队在辩论式决策机的相关思考和规划，包括辩论式决策机概念及原理，以及辩论式决策机在人机协同会诊方面做的初步实践，并展望辩论式决策机未来在司法、舆情等领域的应用。

报告二：面向稀疏数据的 3D/4D 内容创作和重建

讲者：田奇，华为

时间：10 月 19 日 11:25-12:25

个人介绍 田奇教授目前是华为终端 BG 首席科学家，2018.06-2024.04 先后担任华为诺亚方舟实验室计算视觉首席科学家和华为云人工智能首席科学家，2020 年 3 月。田奇教授高中毕业于成都七中(1987)，本科毕业于清华大学电子工程系(1992)，后赴美国伊利诺伊大学香槟分校学习，师从 Thomas S. Huang 教授获博士学位 (2002)。后历任美国德克萨斯大学圣安东尼奥分校计算机系助理教授、副教授、和正教授 (2002-2019)，2010 年获 Google Faculty Research Award，2017 年 UTSA 校长杰出研究奖、2016 年获评多媒体领域 10 大最具影响力的学者，并于 2018 年入选国家级领军人才创新项目。田教授也是教育部长江讲座教授、国家自然科学基金海外杰青，中国科学院海外评审专家。他是国际欧亚科学院院士 (2021)、IEEE Fellow (2016)，CAAI Fellow (2022)，CCF Fellow(2023)，获得 2021 年中国人工智能学会吴文俊人工智能领域杰出贡献奖。田教授是 IEEE TMM, TCSVT, TNNLS, ACM TOMM、Multimedia Systems Journal 等多个期刊的 Associate Editor。他拥有多项美国专利，在计算机视觉及多媒体方向顶级期刊及会议如 IEEE TPAMI, IJCV, TIP, TMM, CVPR, ICCV, ECCV, ACM MM 上发表文章约 760+ 余篇，谷歌学术引用次数 75400+，h 指数为 119，有 9 篇论文获最佳论文奖或者最佳学生论文包括 ACM Multimedia 等。田教授的盘古气象大模型论文（唯一通信作者）发表于《Nature》杂志正刊，盘古气象大模型被《Science》杂志评为 2023 年度十大突破，也被评为 2023 年中国科学十大进展，位列第一。



报告摘要 3D/4D 内容创建在各个领域的研究和应用发挥着越来越重要的作用，如混合现实 (MR)、影视/游戏制作、世界模型构建、设计仿真等，然而 3D/4D 内容获取一直以来是一个巨大的挑战，往往需要从采集到的稀疏数据重建或生成。本报告将探讨如何从不同稀疏程度的数据创建高质量的 3D/4D 内容，并介绍我们的研究工作，通过充分利用重建和生成模型的先验，从文本、图像、视频到已有的 3D/4D 场景，均可创作或重建出符合需求的 3D/4D 内容，极大程度的降低用户采集和创建的门槛。

报告三：通用人工智能的技术进展和典型应用

讲者：王士进，讯飞研究院

时间：10月20日 8:30-9:30



个人介绍 王士进博士，现任讯飞研究院执行院长、认知智能全国重点实验室副主任。他主要从事信号处理、大模型、行业智能等研究，获国家科技进步一等奖、省科学技术奖一等奖、吴文俊人工智能科技进步奖一等奖、中国科协求是杰出青年成果转化奖，带领团队获得 SQuAD、GLUE、VCR 等十几项国际比赛冠军。他主导了“讯飞星火大模型技术及应用”项目，发布了国内领先的星火大模型，并持续赋能多个行业应用。

报告摘要 本报告首先提出当前以认知大模型为代表的通用人工智能技术引发全球广泛关注，然后还分析了从认知大模型到多模态大模型的技术特性、发展趋势及应用价值。其次，报告汇报了科大讯飞研发星火大模型的成果和研发经历，最后重点介绍了大模型服务多行业的探索经验。

多媒体检索专题报告

报告一：多模态内容表示与理解

讲者：赵洲，浙江大学

时间：10月19日 13:30-13:50

个人介绍 赵洲，浙江大学计算机学院教授、博士生导师、国家青年人才、科技部2030重点研发计划青年科学家。主要研究方向为自然语言理解、计算机视觉计算和生成式模型，在国际期刊TPAMI和会议NeurIPS, ICML, ICLR, CVPR等上发表50余篇论文，被谷歌学术引用1.5万余次，相关技术被应用于微软、字节、Stability AI、华为等公司，获2022年度教育部科技进步一等奖、2021年度中国电子学会科技进步一等奖、2022年浙江大学竺可桢学院十佳专业导师，入选“全球前2%顶尖科学家榜单”和“中国高被引学者”，指导的多位学生获华为天才少年奖和浙江大学竺可桢奖学金。



报告摘要 多模态统一表征旨在通过整合来自不同模态（如文本、图像、声音、视频等）的信息，以提高机器学习模型的性能和泛化能力。本报告首先介绍跨模态泛化Uni-Code模型，通过实现了不同模态间的细粒度对齐和统一表示，使模型在仅有单一模态标记时能对其他模态实现零样本泛化。其次针对多模态预训练模型的泛化介绍DG-SCT注意力机制，通过动态调整模型参数以适应多模态输入特征，提升大规模预训练模型在音视频下游任务中的性能。针对多模态数据对不足的问题，设计C-MCR方法来通过语义增强的内-外MCR连接方法，增强了不同模态间嵌入的语义一致性和完整性。进一步提出FreeBind方法，通过“空间键”将多模态表示空间视为基本单元，自由扩展预训练的统一空间，整合额外专家空间的知识，显著提升了多模态理解与生成的性能。此外，本报告介绍一种基于持续学习的跨模态统一表示方法，通过使用一种模态作为中介，逐步将其他模态映射到共享的离散字典中，形成动态增长的统一表示空间。通过引入了CMoE-Adapter模块和伪模态重放(PMR)机制，实现模型以适应新模态的特征并实现多模态统一表示。

报告二：基于大小模型协同的视觉语言导航

讲者：刘偲，北京航空航天大学

时间：10月19日 13:50-14:10



个人介绍 刘偲，教授，博导，北京航空航天大学人工智能学院副院长，中国图象图形学会理事、副秘书长。研究方向是具身智能和多模态内容理解。发表CCF-A类论文100余篇，含IEEE T-PAMI 14篇，谷歌引用14000余次。主持国家优青项目，企业联合重点项目，担任科技创新2030重大项目课题负责人。获CSIG自然科学奖一等奖(1/5)。担任IEEE T-MM、IEEE T-CSVT、CVIU编委(AE)。

报告摘要 对大语言模型的不断探索为增强具身智能决策的可解释性和可控性开辟了新的途径。然而，基于大模型的规划器仍然面临着重大挑战，包括大量的资源消耗和极长的推理时间，这对实际部署构成了重大障碍。鉴于这些挑战，引入了异步大模型增强闭环决策框架，解耦了大模型和小模型规划器的推理过程。分别在无人机视觉语言导航与自动驾驶场景进行了验证，取得了优异的导航性能。

报告三：超图计算与立体对象检索

讲者：高跃，清华大学

时间：10月19日 14:10-14:30



个人介绍 高跃，清华大学长聘副教授、博士生导师。国家青年特聘专家、DeepTech 2022 年中国智能计算科技创新人物，主要研究领域为人工智能、计算机视觉及医学图像处理，特别是超图计算基础理论、方法及应用，在 IEEE TPAMI 等国际期刊及会议发表论文百余篇，论文引用万余次，出版《Hypergraph Computation》等英文专著，多次入选 Clarivate 高被引科学家和爱思唯尔中国高被引学者，担任 International Journal of Computer Vision 和 Medical Image Analysis 等期刊编委，曾获得中国电子学会自然科学一等奖、福建省科技进步一等奖等。

报告摘要 许多生物、社会和技术系统的复杂性源于系统单元之间丰富性的相互作用。社会交流、化学反应、生命系统到生态系统中相互作用通常发生在三个或更多对象的组中，并不能简单地用二元关系来描述。有效地利用这些高阶关联结构能够增强对各类系统的建模能力，帮助理解和预测其动态行为，特别在多媒体检索等应用场景中可以发挥重要作用。超图是建模此类高阶交互、即高阶关联的天然结构。实际应用中通常也面临数据不足和关联复杂难题。本报告主要介绍面向高阶关联的超图计算理论、方法及在立体对象检索中的应用。具体而言，从由数据到高阶关联的映射、由高阶关联到语义表示的映射及高阶关联语义领域泛化三个维度介绍高阶关联的超图结构建模、数据和高阶关联协同的超图语义计算及面向立体对象检索的相关进展。

青年论坛报告

报告一：离线强化学习推荐系统

讲者：辛鑫，山东大学

时间：10月20日 9:30-9:50



个人介绍 辛鑫现为山东大学计算机科学与技术学院信息检索实验室助理教授，人工智能学院院长助理。在此之前，他在格拉斯哥大学获得计算机科学博士学位，导师是 Joemon Jose 教授。他的研究兴趣包括推荐系统、强化学习、图学习、自然语言处理等。他已在 SIGIR、WSDM、WWW、CIKM、ACL 等顶级机器学习与信息检索会议上发表了 40 多篇论文，并获得了 WSDM 2024 年最佳论文提名奖。

报告摘要 在序列推荐中，通常需要考虑多个推荐因素，例如用户的长期参与度、点击、购买等多种的用户行为等。然而，当前的监督学习方法不能很好地建模这些因素。将序列推荐任务转化为强化学习问题是一个可能的解决方案。强化学习方法的一个重要部分是通过与环境的交互来训练推荐智能体。然而，由于训练中的探索阶段需要向用户展示不相关的推荐商品，通过在线算法训练推荐智能体往往存在问题。因此，如何从历史的隐式反馈中学习推荐策略至关重要，但由于离线训练的设置和缺乏负反馈信息，该方法同样面临挑战。在本次报告中，我们将介绍使用离线强化学习的推荐系统，包括将强化学习与监督学习和对比学习相结合，考虑多种推荐奖励信号，以及使用 decision transformer 进行推荐等。

报告二：降低在线学习算法的探索成本与风险：从单智能体到多智能体

讲者：孔芳，南方科技大学

时间：10月20日 9:50-10:10



个人介绍 孔芳，现担任南方科技大学统计与数据科学系助理教授、副研究员、博士生导师。她于 2024 年在上海交通大学取得博士学位，导师为李帅副教授，于 2020 年在山东大学取得学士学位。孔芳的主要研究方向为在线学习和强化学习理论，在 SODA、COLT、ICML 等理论计算机和机器学习领域顶级会议上发表研究成果十余篇，于多智能体顶级会 AAMAS 上给出在线学习的教程，并担任 ICML、NeurIPS、TPAMI 等学术会议和期刊的审稿人。曾获得百度奖学金和多次国家奖学金。

报告摘要 在线学习算法在诸如推荐系统、实验平台等实际应用中扮演着日益重要的角色。算法通过与环境的持续迭代交互，不断更新对未知知识的了解，以期最大化累积收益并最小化累积懊悔。在线学习算法面临的核心挑战之一在于探索与利用的平衡：过度探索虽然有助于发现更多信息，但将导致高昂的成本和资源消耗；而探索不足则可能使算法无法充分了解环境，从而难以找到全局最优解。本报告首先讨论了如何利用动作的特征信息、动作之间的反馈关联等辅助信息，来减少探索成本，将累积懊悔对动作集合大小的依赖降低为对更低维特征维度的依赖，从而提高算法的效率和可扩展性。此外，实际应用中反馈机制和决策空间可能无法与理想假设保持一致，算法在探索过程中往往面临潜在风险。本报告进一步探讨了如何利用主动监测和被动适应等技术来降低这些风险，以确保算法在不完美信息下依然具备稳健性。最后，本报告将视角延伸至多智能体问题，分析了多智能体之间的合作、通信与竞争对探索进程管理带来的新挑战，并探讨了应对这些挑战的策略。

报告三：离线强化学习中基于扩散模型的数据增强技术：从轨迹拼接的视角

讲者：龙婷，吉林大学

时间：10 月 20 日 10: 10-10: 30



个人介绍 龙婷，现担任吉林大学人工智能学院副教授，研究员，博士生导师。她的主要研究方向为信息检索，强化学习。在 SIGIR, KDD, AAAI, ICML 等顶级学术会议上发表多篇研究成果，并担任 KDD, WWW, WSDM, AAAI 等学术会议的审稿人。

报告摘要 离线强化学习旨在使用一组固定的经验数据来优化策略。因此，现有离线强化学习算法的性能通常受制于离线数据集的质量。然而，在大多数场景下，收集大量高质量离线数据用于策略学习具有一定挑战。本次报告探讨了离线强化学习中的数据增强方法，重点介绍如何使用扩散模型从轨迹拼接的角度生成数据。通过将离线数据集中低回报和高回报的子轨迹拼接在一起，以创建更多高质量的样本，丰富原始数据集，从而帮助离线强化学习算法进行高效学习。本次报告讨论的新颖的离线强化学习数据增强策略，在高回报数据有限的情况下具有较大的应用潜力。

报告四：基于用户心理奖励模型的序列化推荐对齐方法

讲者：赵梦辰，华南理工大学

时间：10月20日 10:30-10:50

个人介绍 赵梦辰，现任华南理工大学软件学院准聘副教授，博士生导师，教育部大数据与智能机器人重点实验室成员，CCF 人工智能与模式识别专委会多智能体学组执行委员。博士毕业于新加坡南洋理工大学计算机科学与工程学院，师从安波教授。本科毕业于中山大学应用数学系。曾先后在腾讯广告和华为诺亚方舟实验室担任高级研究员。主要研究方向是强化学习，多智能体系统，推荐系统以及 AI agents 在软件工程领域的应用。近年来，在人工智能顶级会议 (ICML, ICLR, NeurIPS, AAAI, IJCAI, CVPR, WWW) 和期刊 (IEEE Trans. on Cybernetics, Trans. on Dependable and Secure Computing, IEEE Intelligent Systems) 上发表了 20 余篇学术论文。长期担任 ICML, ICLR, NeurIPS, AAAI, IJCAI, CVPR, AAMAS 等顶级会议和期刊的 Committee Member 和审稿人。曾获南洋理工大学杰出博士论文奖、Microsoft Malmo Collaborative AI 挑战赛冠军。



报告摘要 推荐模型通常基于用户隐式反馈进行训练，例如点击行为。然而，用户隐式反馈在许多场景中可能是不准确的。例如，用户可能因为标题吸引而点击了一则新闻，但在阅读内容后感到不适。这种不准确的点击信号可能导致推荐的项目与用户的真实偏好之间产生显著的偏差。然而，由于准确收集明确的用户反馈成本较高且难以实现，现有的工作通常忽略了这种偏差。但是，我们注意到用户对历史交互的感受会显著影响他们在推荐系统中的后续行为。这促使我们从用户行为轨迹中推断他们的真实偏好。在本报告中，我们提出了一种新的序列推荐框架，称为 MTRec。我们将用户的决策过程建模为马尔可夫过程，其中用户每次对商品的行为都会获得一个反映其行为满意度的心理奖励。为了捕捉心理奖励的随机性，我们提出了一种逆强化学习分布以学习心理奖励模型。然后，使用心理奖励模型提供的监督信号，我们提出了一个辅助学习目标，其指导推荐模型与用户的真实偏好对齐。实验表明，MTRec 在 CTR 模型和交互式推荐模型上带来了显著的性能提升。最后，我们将 MTRec 部署在一个工业级短视频推荐平台上，带来了用户观看时长 7% 的提升。

工业论坛报告

报告一：微信搜一搜技术简介

讲者：张帆，微信

时间：10月20日 11:30-11:50



个人介绍 北京大学硕士，2018年加入微信搜一搜，先后参与并负责小程序搜索、相关搜索词条推荐、搜索发现等业务的算法研发，2023年至今主要负责表情搜索整体算法研发工作。

报告摘要 微信搜一搜，将用户和微信生态内容高效地连接在一起，近年来用户量和搜索量保持高速增长。本次报告将介绍微信搜一搜所承接的微信生态搜索能力，以及如何使用各种不同的信息检索方法、深度学习算法、大模型技术等在庞大的微信内容生态中检索到满足用户需求的多元内容。

报告二：大语言模型时代推荐系统的发展路径及未来趋势

讲者：唐睿明，华为

时间：10月20日 11:50-12:10



个人介绍 唐睿明，华为诺亚方舟实验室资深研究员，任推荐与搜索实验室主任，负责推荐和检索领域的前沿研究和技术落地，研究方向包括机器学习、推荐系统、深度学习、强化学习、AutoML、图神经网络等，谷歌学术引用 7000 余次。他于 2009 年在中国东北大学获得学士学位，并在 2014 年从新加坡国立大学计算机专业获得博士学位，2014 年底加入华为诺亚方舟实验室。

报告摘要 推荐系统在满足用户多样化信息需求和提供个性化内容方面发挥着至关重要的作用。在本次报告中，我们将推荐系统置于更广泛的发展背景中，系统性地梳理了推荐系统的技术演进趋势：从早期的深度学习方案，逐步演进到结合大语言模型的方案，并进一步发展为个性化智能体。通过归纳梳理，我们明确了现代推荐系统的两种关键形态，即列表式推荐和对话式推荐。这两种形态最终都朝着具有长期记忆、反思和工具智能的个性化智能体方向发展汇聚。在这个发展过程中，我们发现，随着对用户理解的不断深入，推荐信息的有效性逐步提升，同时用户获取信息的成本逐渐降低。我们列举对未来个性化技术和交互形态发展至关重要的挑战，并展望未来的发展趋势和前景。

大会 Tutorial

报告一：基于大语言模型的自主智能体

讲者：陈旭，中国人民大学

时间：10月18日 14:00-15:30



个人介绍 陈旭，中国人民大学高瓴人工智能学院准聘副教授，主要研究大语言模型，因果推断和推荐系统等。曾在 TheWebConf、NeurIPS、AIJ、KDD、ICLR 等著名国际会议/期刊发表论文 60 余篇。他带领团队撰写了国际上早期的大语言模型智能体综述《A survey on large language model based autonomous agents》、构建了首个基于 LLM Agent 的用户行为模拟环境“RecAgent”等。他主持/参与十余项国家自然科学基金以及企业合作项目，相关成果在多家企业落地。

报告摘要 基于大语言模型的自主智能体研究近年来受到了人们的广泛关注，本次报告围绕该领域的基础知识展开，具体介绍如何构建大语言模型驱动下的自主智能体，包括如何刻画智能体，如何设计记忆机制和规划手段，如何让智能体执行动作等。同时报告人也将介绍其开发的基于自主智能体的用户行为模拟智能体。通过该报告，听者可以对大语言模型驱动下的自主智能体这一研究方向及其关键问题有较为系统的了解。

报告二：知识检索增强：范式与关键技术

讲者：王昊奋，同济大学

时间：10月18日 15:50-17:10

个人介绍 王昊奋，上海交大博士、上海市优秀博士，同济大学特聘研究员。其研究兴趣及专长是人工智能、知识图谱、自然语言处理、对话式用户交互、智能内容生成。他是全球最大的中文开放知识图谱联盟 OpenKG 发起人之一。出版了《自然语言处理实践》《知识图谱方法、实践与应用》等著作。他还受邀在世界人工智能大会等诸多国际与国内智能峰会上担任讲者，并在自然语言处理国内顶级会议 NLPCC 多次担任知识图谱方向主席，长期作为 ISWC, WWW, AAAI 等人工智能国际顶级会议程序委员会委员。负责参与 10 余项国家自然科学基金、863 重大专项、科技部科技支撑、经信委和科委等 AI 相关项目，共发表 100 余篇 AI 领域高水平论文。长期在一线人工智能公司担任 CTO 之职，荣获徐汇区学科带头人人才计划。他构建了全球首个可交互养成的虚拟偶像——“琥珀·虚颜”；所构建的智能客服机器人已累计服务用户超过 10 亿人次，并成功入选中国人工智能产业发展联盟最佳 AI 应用案例。



报告摘要 知识检索增强技术为大语言引入了额外的知识源，有效地缓解了幻觉问题与知识时效性问题，迅速成为优化大模型实践中的一项关键技术。在技术迭代的过程中，RAG (Retrieval Augmented Generation) 与结构索引优化、知识图谱、向量数据库、大模型微调、提示工程等多项技术深度融合，众多功能模块相继被提出，这对研究人员全面理解 RAG 构成了挑战。本次报告将从范式、关键技术与应用发展等角度对 RAG 进行全面梳理和分析，旨在从更高层面把握技术发展趋势和未来方向。通过对当前研究现状的综合分析，我们提出了模块化 RAG 与 RAG Flow 的研究范式，总结了 6 大功能模块，包含 50 余个算子操作，并从百余篇论文中凝练出了 7 种典型的 RAG Flow 设计模式，为 RAG 系统的设计提供指导。基于这些范式，我们进一步推进了 OpenRAG 系列的开源工作，构建了 OpenRAG Knowledge Base，全面覆盖了 RAG 研究人员和开发者所需的信息，并提供了支持高度自定义的多维度分析视图；同时，我们搭建了 OpenRAG Playground，协助研究人员和工程师快速搭建前沿的基线方法，并在公开或自定义数据集上进行快速验证以及不同 RAG Flow 效果的比较。

录用论文列表

1. MB-HGCN：基于层次图卷积的多行为推荐方法. 严明时, 陈慧临, 程志勇
2. 基于矩阵与混合机制双流长短期记忆网络的终身长序列推荐方法. 孙文奇, 谢若冰, 张君杰, 赵鑫, 康战辉, 文继荣
3. 面向脑电信号情绪识别任务的预训练模型设计. 张开元, 叶子逸, 艾清遥, 刘奕群, 王晶, 张敏, 马少平
4. 基于会话有向图卷积和模态有偏融合的多模态会话情感识别. 程佳玮, 朱小飞, 尹雅博
5. 基于多粒度记忆 RAG 的个性化角色扮演框架. 蓝恒志, 黄乐, 白婷
6. 基于长距离上下文的预训练数据检测方法. 张伟超, 张儒清, 郭嘉丰, 范意兴
7. 结合潜在关系信息增强策略和全局对应矩阵的古汉语关系三元组联合抽取. 胡益裕, 康文军
8. 融合代码文本和得分信息的编程知识追踪. 陈浩飞, 付广成, 曾雪强, 左家莉, 王明文
9. 面向多模态大模型的检索增强生成算法研究. 邓梦洁, 窦志成
10. 基于宽度优先搜索策略的工具文档增强方法. 张振, 高莘, 施政良, 陈竹敏, 任昭春
11. 从表情到情感：基于情绪监督信号的生成式对话摘要. 王富瑞, 朱振方, 刘培玉
12. ECPE-Qwen：微调大语言模型进行零样本情绪原因对抽取. 王琪尧, 杨亮, 许侃, 林原, 林鸿飞
13. 面向司法文书聚类的预训练文本表示模型研究. 杨越千, 窦志成
14. 基于大语言模型的跨语言会话式商品推荐设计与实现. 李奕杉, 孙培杰, 张敏
15. 基于检索增强和知识蒸馏的汉越跨语言查询扩展方法. 李京, 朱恩昌, 高盛祥, 余正涛, 周帅帅, 黄于欣
16. 基于多任务指令微调的少样本事件抽取. 孙大鹏, 张绍武, 李森森, 王伊可, 张益嘉
17. PPMEP：基于最大熵原理的插用式中文拼写纠错方法. 唐劲松, 任俊, 黄瑞章, 任丽娜
18. 基于大语言模型多智能体协作的查询扩展方法. 张顺平, 涂新辉, 黄彬彬
19. 基于预训练模型的仇恨言论检测研究. 林原, 张亚, 于蒙, 许侃, 林鸿飞
20. 基于问题导向提示学习和多路推理总结的检索增强生成问答. 王昱婷, 陈波, 闫强, 范意兴, 余智华, 郭嘉丰
21. SSEM：基于跨语义增强的中文拼写纠错方法. 王浩宇, 任丽娜, 黄瑞章, 马灿
22. 基于预训练语言模型和图神经网络的学科划分. 杨柳, 徐晨, 毛佳昕, 徐君, 任超, 文继荣
23. 基于自监督就诊相似性网络的药物推荐. 何亦琛, 董守斌, 郑潇柔, 胡金龙
24. 基于案例三元组的法律判决预测模型. 占婉妍, 孙媛媛, 李志廷, 林国凯, 王勇, 徐博,

杨亮, 杨志豪

25. 分解和学习: 基于双重表征和自监督学习策略的多模态情感分析方法. 孟静, 朱振方, 刘培玉
26. 一种具有均匀性约束和重要性均衡的基于原型的矩阵分解方法. 曹玉祥, 廉涛, 王龙, 荆星博, 窦浩铖
27. 结合边界信息的对比学习嵌套命名实体识别方法. 范锦涛, 陈艳平, 黄瑞章, 杨采薇, 林川
28. 金融文本情感极性分析算法比较研究. 柯姝言, 李辉, 纪荣嵘
29. 基于双向文本扩展的信息检索重排方法. 郭聪, 涂新辉, 宗宇航
30. 项目属性信息增强的临时群组长短期兴趣融合建模方法. 王智玄, 庞继芳, 王智强
31. 基于预训练语言模型和与知识蒸馏的轻量化序列推荐方法. 李莉, 程明月, 刘芷町, 章浩, 刘淇, 陈恩红
32. 基于混合检索重排序策略的大模型增强方法. 张健, 唐晋韬, 李莎莎, 王挺
33. 重新审视检索系统在教育场景下的应用. 兰天, 马骜桀, 刘进宇, 李林瀚, 毛先领
34. 从无限维到无限层: 转移矩阵视角下的课程推荐算法. 周云垚, 刘海峰, 周俊生
35. 基于多模型论辩的跨领域成分句法分析. 田海燕, 张栋, 朱苏阳, 李寿山, 周国栋
36. 基于大语言模型的搜索用户模拟器研究. 张颖, 吴之璟, 梁怡栋
37. 融合关键概念和潜在概念的冗长查询缩略方法. 朱铭洋, 黄于欣, 余正涛
38. 基于法条检索的生成式法律问答研究. 李明达, 邱洪波, 孙媛媛, 王艳华, 杨志豪, 林鸿飞
39. 融合 VAD 知识的情感分布增强的细粒度情绪识别方法. 李春阳, 万中英, 曾雪强, 左家莉, 王明文
40. 基于主题多视图表示的零样本实体检索的方法. 齐丹丹, 王长征, 郭少茹, 闫智超, 胡志伟, 苏雪峰, 马博翔, 李时钊, 李茹
41. 融合边界交互信息的命名实体识别方法. 何安康, 陈艳平, 扈应, 黄瑞章, 秦永彬
42. 基于社交媒体的伪健康信息检测数据集构建与研究. 张桐瑄, 张李洋, 宋志腾, 郭凤羽, 张桂芸, 徐博, 林鸿飞
43. 基于掩码注意力与多特征卷积网络的关系抽取方法. 卢展跃, 陈艳平, 杨卫哲

赞助单位

白金赞助



金牌赞助



银牌赞助



铜牌赞助



微信搜一搜 x CCIR

关于我们

微信搜索依托独特的微信公众号、视频号和小程序生态，在微信各个场景中通过搜索技术、推荐技术等构建新一代的搜索产品，实现人与信息、人与服务、人与人之间的精准连接。独特的场景、海量的用户、前沿的技术支撑业务的高速发展。我们拥有优秀的AI算法、工程技术和产品团队，分布在广州、北京、深圳多地，欢迎你的加入。



用户量持续增长

8亿

MAU增长至

54%

搜索QV增长

332%

视频QV增长

用户需求不断演进

+895%

视频分发量

+26%

小程序分发量

+43%

公众号文章分发量

多元载体，融合统一

经典载体：图文搜索



新兴媒介：视频搜索



微信特色：服务/账号搜索



问答社区：问一问



贴近用户，有用有趣

表情搜索与合成



搜索x微信状态联动



图片搜转excel



2012诺亚方舟实验室：华为AI算法研究中心

诺亚方舟实验室介绍：

诺亚方舟实验室（Noah's Ark Lab）是华为公司从事人工智能基础研究的实验室，秉持理论研究与应用创新并重的理念，致力于推动人工智能领域的技术创新和发展，并为华为公司的产品和服务提供支撑。

在理论研究方面，我们拥有来自全球顶尖高校、优秀企业的高水平研究员，同时我们与世界顶尖创新资源保持连接，包括10+国家、25+大学、50+项目、1000+研究人员。2023年诺亚方舟实验室在顶会中成绩斐然，其中Integral Neural Networks获得计算机视觉顶会CVPR 2023 Best Paper Award候选之一（12/9155），首创全连续表征的积分神经网络，代替传统的离散表征范式，为Once-for-All场景开拓全新的技术路径；Confidence-guided Adaptive Memory Efficient Optimization（CAME）获得自然语言处理顶会ACL 2023 Outstanding Paper Award，构筑华为盘古模型训练的核心竞争力。

在应用创新方面，我们致力于推动领先的人工智能技术在各场景、包括网络智能、企业智能、终端智能、自动驾驶等新兴领域的应用，帮助公司解决商业场景中的挑战难题。面对智能时代大模型的机遇和挑战，在昇腾集群系统下，创新盘古基础模型构建万亿级高质量训练数据，实现4千卡训练近线性扩展，解决训推加速、长序列等技术难题，支撑Mate60发布，使能华为云等业务。



招聘领域:



招聘要求:

- 对研究工作，特别是算法研究兴趣浓厚，有相关领域内的研究成果；
- 业务抽象能力强，具备创造性思维，能够将全新想法转化为工程应用；
- 具备较强的编程实现能力，精通主流编程语言，如python/C++/Java等；
- 在高水平国际会议或学术期刊发表多篇论文，国内外知名竞赛获奖者；
- 对研究工作充满热情，具备良好的团队合作精神，较强的沟通能力，有团队协作经验优先。

