孙晓洁

1999.10 | 136-2202-0972 | sunxj1999@126.com

教育背景

中国科学院计算技术研究所 计算机技术 硕士 (**保送**) *2021 年 08 月 - 2024 年 07 月*

- **研究方向:** 信息检索; **实验室:** 网络数据重点实验室; **导师:** 郭嘉丰;
- **荣誉奖项:** **研究生国家奖学金、所长冠名奖-易方达金融科技硕士奖 (6%)**、中国科学院大学三好学生等;

南开大学 软件工程 学士 *2017 年 08 月 – 2021 年 07 月*

- **专业排名:** **3/135**（前 **2.2**%）; **学生工作:** 班长、校学生会体育部副部长;
- **荣誉奖项:** **本科生国家奖学金 (2/135)、 南开大学优秀毕业生 (3/135)** 等;

实习经历

蚂蚁金服 · 搜索组 *学术合作实习生, 2022 年 08 月 - 至今*

- **多属性稠密检索:** 针对结构化数据探索其多属性信息的有效使用方式，产出**两篇专利和两篇论文**: (1) 提出了一种属性和内容文本交互预测的预训练方法; (2) 在多粒度视角下建模属性值间的细粒度语义联系。
- **检索增强大语言模型:** 围绕信息检索与大语言模型的关系展开调研和相关研究。

字节跳动 · Tiktok 电商 *推荐算法实习生, 2022 年 02 月 - 2022 年 07 月*

- **倒排索引:** 完成 sellervideo、author2video 倒排全流程构建，结合用户的点击、购买等交互行为召回 trigger，利用 ctr、cvr、gpm 探索合理的视频排序公式。在英国、印度等多国全量上线, cvr+3.2%, gmv+5.5%。
- **SSL 自监督学习:** 探索短视频召回阶段的对比学习应用。(1) 利用 feature mask 构建正样本, 优化视频表征，在多国线上测试收益正向; (2) 学习用户多行为序列间可迁移的知识，完成实现和基本训练。
- **全网行为召回模型:** 基于用户全网商品交互行为，完成离线数据流构建，结合 logQ 纠偏等技术实现模型。

竞赛经历

国际顶级会议竞赛 WSDM CUP 2023 (3/215 Teams) *2022 年 12 月 - 2023 年 01 月*

- 对搜索引擎中海量的用户搜索点击日志进行去偏预训练，在无偏排序学习和互联网搜索预训练模型两项任务中均获得第三名，受邀投稿两篇论文，并在会议上进行报告。

2022 语言与智能技术竞赛-段落检索 (6/916 Teams) *2022 年 05 月 - 2022 年 06 月*

- 针对首个大规模中文段落检索数据集 DuReader 完成全阶段检索任务，负责精排阶段的优化，结合动态难负例挖掘、多阶段 rerank 等技巧，在难负例挖掘能力层面进行集成，mrr@10 得分 0.81953，获得第六名。

学术成果

[1] A Multi-Granularity-Aware Aspect Learning Model for Multi-Aspect Dense Retrieval
Xiaojie Sun, Keping Bi, Jiafeng Guo, Sihui Yang et al. (**WSDM 2024, CCF-B & 清华 A**)

[2] Reproducibility Analysis and Enhancements for Multi-Aspect Dense Retriever with Aspect Learning
Keping Bi, **Xiaojie Sun**, Jiafeng Guo, Xueqi Cheng. (**ECIR 2024**)

[3] Pre-training with Aspect-Content Text Mutual Prediction for Multi-Aspect Dense Retrieval
Xiaojie Sun, Keping Bi, Jiafeng Guo, Xinyu Ma, Fan Yixing et al. (**CIKM 2023, CCF-B**)

[4] Feature-Enhanced Network with Hybrid Debiasing Strategies
Lulu Yu*, Yiting Wang*, **Xiaojie Sun***, Keping Bi, Jiafeng Guo. (**WSDM CUP 2023, CCF-B**)

[5] Ensemble Ranking Model with Multiple Pretraining Strategies
Xiaojie Sun*, Lulu Yu*, Yiting Wang*, Keping Bi, Jiafeng Guo. (**WSDM CUP 2023, CCF-B**)