

蔡银琼

✉ caiyinqiong@163.com · ☎ (+86) 159-1081-3282 · 🏠 主页

🎓 教育经历

中科院计算所 硕博连读（保送），计算机应用技术, 导师: 郭嘉丰 2018.09 – 2024.01
大连理工大学 学士, 网络工程, 排名: 3/120 2014.09 – 2018.06

📄 代表性工作

【模型架构】基于混合专家的多路检索模型架构

- 构建基于混合专家的多路检索系统，提出竞争学习机制协调各路检索模型“分而治之”地解决多样相关模式建模的难题。
- Yinqiong Cai, Yixing Fan, Keping Bi, Jiafeng Guo, Wei Chen, Ruqing Zhang, Xueqi Cheng.** CAME: Competitively Learning a Mixture-of-Experts Model for First-stage Retrieval. CIKM 2023 在投. CCF-B.

【模型训练】基于不完备标注数据的检索模型训练

- 从无偏学习角度解决人工标注检索数据中存在的未标注正例（伪负例）问题。基于反事实学习理论提出“对偶估计技术”，联合建模相关性和选择倾向性，实现无偏检索模型学习。
- Yinqiong Cai, Jiafeng Guo, Yixing Fan, Qingyao Ai, Ruqing Zhang, Xueqi Cheng.** Hard Negatives or False Negatives: Correcting Pooling Bias in Training Neural Ranking Models. CIKM 2022. CCF-B.

【模型生长】基于无标注新增数据的检索模型终身学习

- 首次提出并定义检索模型的终身学习任务。提出基于记忆机制的检索模型持续学习方法，实现对于无标注新增数据的模型适配更新，并保证文档表达的后向兼容性，避免索引重建的高时间代价。
- Yinqiong Cai, Keping Bi, Yixing Fan, Jiafeng Guo, Wei Chen, Xueqi Cheng.** L²R: Lifelong Learning for First-stage Retrieval with Backward-Compatible Representations. CIKM 2023 在投. CCF-B.

面向召回阶段的语义检索模型综述

- 首篇针对召回阶段检索模型（尤其是稠密向量检索模型）的全面综述。论文引用量 **47+**，Github 资源库获得 Star **250+**。
- Jiafeng Guo, **Yinqiong Cai**, Yixing Fan, Fei Sun, Ruqing Zhang, Xueqi Cheng. Semantic Models for the First-stage Retrieval: A Comprehensive Review. TOIS 2021. CCF-A. (学生一作)

🏆 项目经历

检索增强大语言模型 SearchGPT / 核心成员 2023.03 – 至今

- 依托大规模实时数据、结构化知识库等，提高大语言模型信息处理的准确性、实时性和可靠性
- 负责检索模块的设计与开发实现，初版 Demo 已在第十九届科学院公众开放日对外展出

深度文本匹配工具 MatchZoo / 核心成员 2019.03 – 2019.12

- 通用的开源文本匹配工具包，在信息检索社区具有较大影响力，获得 Star 4000+，Fork 1000+
- Pytorch 版本的核心开发者之一，进行项目基础架构和经典排序模型的开发

网络数据引擎 WDE / 开发人员 2017.10 – 2018.04

- 负责监控与统计子平台的设计与开发实现，对采集数据进行多维度文本分析，以便进行及时预警

🏆 获奖情况

WI-IAT 2022 志愿者奖（共 2 人） 2022
所长冠名奖-易方达金融科技硕士生奖（6%） 2020
中国科学院大学三好学生 2019, 2020, 2023
大连市优秀毕业生 2018

≡ 发表文章列表

- [1] L²R: Lifelong Learning for First-stage Retrieval with Backward-Compatible Representations
Yinqiong Cai, Keping Bi, Yixing Fan, Jiafeng Guo, Wei Chen, Xueqi Cheng. 在投 *The 32nd ACM International Conference on Information and Knowledge Management (CIKM)*, 2023.
- [2] CAME: Competitively Learning a Mixture-of-Experts Model for First-stage Retrieval.
Yinqiong Cai, Yixing Fan, Keping Bi, Jiafeng Guo, Wei Chen, Ruqing Zhang, Xueqi Cheng. 在投 *The 32nd ACM International Conference on Information and Knowledge Management (CIKM)*, 2023.
- [3] Hard Negatives or False Negatives: Correcting Pooling Bias in Training Neural Ranking Models
Yinqiong Cai, Jiafeng Guo, Yixing Fan, Qingyao Ai, Ruqing Zhang, Xueqi Cheng. *The 31st ACM International Conference on Information and Knowledge Management (CIKM)*, 2022.
- [4] Pre-training Methods in Information Retrieval
Yixing Fan*, Xiaohui Xie*, **Yinqiong Cai**, Jia Chen, Xinyu Ma, Xiangsheng Li, Ruqing Zhang, Jiafeng Guo. *Foundations and Trends in Information Retrieval (FnTIR)*, 2022.
- [5] Semantic Models for the First-stage Retrieval: A Comprehensive Review
Jiafeng Guo, **Yinqiong Cai**, Yixing Fan, Fei Sun, Ruqing Zhang, Xueqi Cheng. *ACM Transactions on Information Systems (TOIS)*, 2021.
- [6] A Discriminative Semantic Ranker for Question Retrieval
Yinqiong Cai, Yixing Fan, Jiafeng Guo, Ruqing Zhang, Yanyan Lan, Xueqi Cheng. *The 11th International Conference on the Theory of Information Retrieval (ICTIR)*, 2021.
- [7] 基于多表达的第一阶段语义检索模型
蔡银琼, 范意兴, 郭嘉丰, 张儒清, 程学旗. 计算机工程与应用, 2021.