

张龙辉

✉ longhuizhang97@gmail.com · ☎ 15040150349 · 男 · 28 岁 · 河南

🎓 教育背景

| | |
|--|-------------|
| 哈尔滨工业大学（深圳），广东 | 2022 – 2026 |
| 在读博士研究生 计算机科学技术，导师：李晶教授（海外优青）、张民教授（国家杰青） | |
| 东北大学, 辽宁 | 2019 – 2022 |
| 硕士研究生 计算机技术，导师：任飞亮教授 | |
| 沈阳化工大学, 辽宁 | 2015 – 2019 |
| 学士 计算机科学与技术 | |

💼 实习经历

| | |
|-------------|-------------|
| 阿里通义实验室, 浙江 | 2022 – 2024 |
|-------------|-------------|

🔬 研究方向

- 代码大模型，信息检索，信息抽取

📄 学术成果

Function-to-Style Guidance of LLMs for Code Translation

- 会议: ICML 2025 (CCF A 类会议, 一作)
- 研究方向: 代码翻译、代码大模型
- 论文内容: 针对当前代码大模型在代码翻译任务上存在正确性低下和风格糟糕的问题，提出了从功能到风格的两阶段大模型训练策略。

Speed Up Your Code: Progressive Code Acceleration Through Bidirectional Tree Editing

- 会议: ACL 2025 (CCF A 类会议, 一作)
- 研究方向: 代码优化、代码大模型
- 论文内容: 针对当前代码大模型在代码优化任务上存在的优化路径不清晰的问题，提出通过双向树编辑策略构建高质量指令微调的数据，并通过课程学习提高模型在代码优化任务上的表现。

A Two-Stage Adaptation of Large Language Models for Text Ranking

- 会议: ACL 2024 Findings (CCF A 类会议, 一作)
- 研究方向: 信息检索、大模型微调
- 论文内容: 针对当前大模型在信息检索任务上探索不足的问题，提出了两阶段的训练策略，通过弱监督数据预训练和有监督数据上的微调，增强大模型的信息检索能力。
- 成果应用: 研究成果已成功部署于阿里云多个重要产品线，包括但不限于阿里云的“百炼”与“灵积”大模型平台、智能搜索引擎、NLP 服务平台，有效赋能了通义 Qwen APP、通义智文、通义听悟、通义星辰等一系列企业创新应用。

Progressive Adaptation of Large Language Models for Multilingual Text Ranking

- 期刊: TOIS (CCF A 类期刊, 一作, 大修投稿中)
- 研究方向: 多语言信息检索、大模型微调
- 论文内容: 针对当前多语言文本重排序研究匮乏的问题，提出利用大模型进行高质量的多语言重排序数据合成，并通过模型蒸馏策略缓解多语言冲突问题。

Chinese Sequence Labeling with Semi-Supervised Boundary-Aware Language Model Pre-training

- 会议: COLING 2024 (CCF B 类会议, 一作)
- 研究方向: 语言模型的预训练

- 论文内容: 针对 BERT 在中文场景下能力受限的问题, 通过半监督学习策略将中文词汇的边界信息注入预训练模型, 强化 BERT 的中文理解能力。

A Simple but Effective Bidirectional Framework for Relational Triple Extraction

- 会议: WSDM 2022 (清华大学计算机推荐 A 类会议, CCF B 类会议, 与导师共同一作)
- 研究方向: 信息抽取
- 论文内容: 针对级联的关系三元组抽取模型对 subject 抽取效果的高度依赖性问题进行解决。

Hybrid Retrieval and Multi-stage Text Ranking Solution at TREC 2022 Deep Learning Track

- 会议: TREC 2022
- 研究方向: 信息检索
- 论文内容: 基于国际信息检索竞赛 TREC 2022 的解决方案发表论文。

A Novel Global Feature-Oriented Relational Triple Extraction Model based on Table Filling

- 会议: EMNLP 2021 (CCF B 类会议, 与导师共同一作)
- 研究方向: 信息抽取
- 论文内容: 指明了现有的实体关系联合抽取模型忽略元组间关联的事实, 并利用 Transformer 捕获关系三元组间的全局信息, 进行三元组间的验证和补充。
- 学术影响: 谷歌引用 100+, 2022 CCF 大数据与计算智能大赛 (CCF BDCI) “工业知识图谱关系抽取任务”官方基线。

A Conditional Cascade Model for Relational Triple Extraction

- 会议: CIKM 2021 (CCF B 类会议, 与导师共同一作)
- 研究方向: 信息抽取
- 论文内容: 以现有的关系三元组抽取模型普遍具有的类别不平衡问题为出发点, 通过优化损失函数和目标抽取流程, 将标注的样本规模由平方空间缩减到了线性空间, 解决类别失衡问题。

An Effective System for Multi-format Information Extraction

- 会议: NLPCC 2021 (CCF C 类会议, 共同一作)
- 研究方向: 信息抽取
- 论文内容: 基于 2021 语言与智能技术竞赛的解决方案发表论文。

BSLRel: 基于二元序列标注的级联关系三元组抽取模型

- 期刊: 中文信息学报 (CCF T1 类期刊, 一作)
- 研究方向: 信息抽取
- 论文内容: 基于 2020 语言与智能技术竞赛的解决方案发表论文。

🏆 荣誉及比赛经历

辽宁省优秀硕士学位论文 2022

TREC 2022 2022/06 - 2022/09

- 排名: 文本排序任务第一
- 任务介绍: 基于 MS MARCO 训练集以及 TREC 2022 信息检索测试集进行模型优化。

2021 语言与智能技术竞赛: 多形态信息抽取任务, 队长 2021/03 - 2021/05

- 排名: 4/2148, 三等奖
- 任务介绍: (1) 关系抽取任务: 针对关系三元组间的实体重叠和传统模型面临的标签不平衡、暴漏偏差等问题, 设计了一种基于表填充的多槽值实体关系联合抽取模型, 同时对实体搜索算法进行优化, 有效地解决了嵌套实体的识别问题。(2) 事件抽取任务: 针对事件间的实体重叠问题, 设计了一种基于集合预测的句子级和篇章级事件抽取模型。

2020 语言与智能技术竞赛: 关系抽取任务, 队长 2020/03 - 2020/05

- 排名: 5/1166, 优胜奖
- 任务介绍: 针对 DuIE 数据集中对 Object 进行的复杂化扩展, 设计了一种基于 pipeline 的多槽值实体关系联合抽取模型 BSLRel, 该模型将多槽值三元组解耦合为一个单槽值三元组和一个被 Object 实体触发的事件, 可以有效的对不同类型的多槽值三元组进行完全抽取。