



姓名：纪杰周



电话：17353951520



邮箱：jiezhoujif@126.com



求职意向：NLP/AIGC/LLMs 算法岗



现居地址：辽宁省大连市



微信：17353951520



学历背景

2017.09—2021.6

青岛理工大学

工程造价

本科

2021.09—2024.6

大连理工大学

应用数学

硕士研究生

- 研究方向：基于强化学习方法的演化博弈论。
- 主修课程：应用常微分方程、深度学习与算法设计、机器学习理论、强化学习、现代图论、高等组合学等。
- 《The influence of own historical information and environmental historical information on the evolution of cooperation》，目前已被 Applied Mathematics and Computation 录用 (SCI 检索，第一作者，方向为演化博弈论)。
- 《A Survey of Text Quality Evaluation for LLMs Generation》已投至 ACL，目前处于审稿中 (第二作者，方向为生成文本评估)。

专业技能

- 熟悉 NLP 基础子任务以及模型构建，包括命名实体识别、文本分类、机器翻译等。
- 熟悉 BERT 系列、GPT 系列、Encode-Decode、MoE 系列以及目前主流的 LLMs 基座框架。
- 熟练使用 Pytorch 框架和 Linux 操作系统，熟练使用 TorchRun、DeepSpeed 和 Megatron-lm 等辅助分布式训练框架。
- 熟悉 LLMs 增量预训练、SFT 与人类偏好对齐技术，熟练使用 Lora 及 Q-Lora 等 LLMs 微调方法。

实习经历

2023.8—2024.1

美团

交互应用技术中心 / LLMs 算法实习生

主要工作：数据工程实验与业务场景下 LLMs 的推理能力建设。具体工作：

- ◎ 针对招聘场景构建 Text2SQL 任务。简化招聘部门提供的 SQL 代码并按查询类型进行分类，实现了查询需求的自动化生成并结合 GPT4 构造微调指令数据集。以 Ziya-Llama-13B 为基座模型，微调后其在评测集上的 Pass@1 为 0.8，随后通过拒绝采样生成部分人工评估较高的样本添加至训练集，最终使得微调模型在评测集上的 Pass@1 达到了 0.95。
- ◎ 优化数据质量过滤 Pipeline。在原数据质量过滤基础上考虑指令遵循难度 (IFD) 与平均交叉熵损失，在 Ziya-Llama-13B 数据实验中实现了微调数据压缩 (100W(Old)→30W(New))，相同数据量微调，后者在业务与公开榜单上表现更好。
- ◎ 将业务场景中各类任务大致分成“重过程”和“重结果”两种，分别参考 Orca1 与 Orca2 来构造业务场景下的增强数据，后对 Ziya-Llama-13B 进行解释微调，同时设置 Magic-Prompt 激发模型针对特定问题的推理能力。通过分析错误 Case 的推理过程不断修正增强数据集，最终使得微调后的 Ziya-Llama-13B 在业务榜单提升约 10 个 pp，且推理能力提升明显。
- ◎ 调研了“思维链”与“自由文本合理化”文本质量评估的主流方法。给出了不同业务场景下生成文本量化评估方法以及错误类型并构建了 Critical 数据。以 Ziya-Llama-13B 为基座训练了业务场景下的 Critical-Model-MT，实现了业务场景下生成文本质量的有效评估，GPT4 与人工评估下，Critical-Model-MT 较其他 Critical-Model 优势更加明显。

2023.5—2023.8

MetaApp

AI 研发部 / AIGC 算法实习生

主要工作：构建基于 TypeScript 语言的代码生成模型与代码补全模型。具体工作如下：

- ◎ 负责开源代码数据与业务代码数据的收集与预处理，并创建业务榜单 Benchmark。
- ◎ 分别对 ChatGLM-6B/10B、Llama-7B/13B/30B、StarCoderBase-15B、CodeGen25-7B 制定训练策略进行消融试验，综合考虑选 StarCoderBase-15B 作为基准模型（其在 MxEval 和自建 Benchmark 上的 Pass@1 分别达到 gpt3.5 的 90% 和 85%）。在此基础上训练 Llama-7B 作为 RM 模型并对四种输出类型进行评分，结合 PPO 算法对模型输出做对齐，使 StarCoderBase-15B 在 MxEval 和自建 Benchmark 上的 Pass@1 分别达到 gpt3.5 的 110% 和 120%。
- ◎ 重构 Llama-7B 的 Loss 计算并构建双塔架构，有效缓解其在 Full-FineTuning 过程中出现的灾难性遗忘问题 (15%→2%)。
- ◎ 对比 CodeGen-350m、CodeT5_base、PolyCoder 等轻量代码生成模型在 CPU 上的推理速度与代码生成质量，最终在 CodeGen-350m 基础上整合业务代码对其进行增量训练，使其具备代码补全能力并供游戏开发部门使用。

2023.2—2023.5

千里马网信科技有限公司

数据研究院 / NLP 算法实习生

主要工作：标书资质要求分类、基于 ChatGLM-6B 的 Lora 方法微调。具体工作如下：

- ◎ 构建标文中筛选资质要求的正则表达式，利用筛选后的文本对预训练的 BERT 进行微调。通过参考验证集结果修改所构建的正则表达式，循环往复最终显著提高了预训练 BERT 在测试集上分类的 F1 值 (F1=0.9223)。
- ◎ 构建若干指令数据 (任务为标文生成和命名实体识别)，基于 Lora 的方法对 ChatGLM-6B 进行指令微调。实体识别方面 F1 值较之前提升 30%，标书生成方面人工评估认可度相对较高。

竞赛荣誉

- 第十三届全国大学生数学竞赛全国一等奖
- 第十八届华为杯研究生数学建模全国三等奖
- DataCastle 文本情感分类，名次 6/487，F1=0.942
- DataFountain 中文命名实体识别，名次 4/476，F1= 0.713

自我评价

学习能力强，会定期阅读 NLP/LLMs 方向最新进展文献，并及时总结。沟通能力强，实习期间积极参与团队探讨项目最新进展与模型评估。执行能力强，工作认真负责，踏实稳重，主动接受新事物，敢于迎接挑战。