韩竞骁

13624298637 | hanjingxiao@bupt.edu.cn | 北京
微信：bupt_ee_hjx | 岗位：机器学习/技术研究

教育经历

北京邮电大学 双一流 2023年08月 - 2026年06月
通信工程 硕士 信息与通信工程学院 北京
一等学业奖学金 GPA: 90.5/100

哈尔滨工业大学 985 2018年08月 - 2022年06月
通信工程 本科 电子与信息工程学院 哈尔滨
一等人民奖学金 GPA: 89.5/100

实习经历

北京腾讯信息技术有限公司 2025年05月 - 至今
技术研究实习生 腾讯广告商业AI部 北京
一、工作概述：
使用强化学习提升大模型的广告营销文案生成的多样性，提升服务行业覆盖率和文案采纳率。
二、工作内容：
1. 偏好数据定义和生产：
系统推荐文案和用户针对其修改后提交的文案作为偏好数据。建立了偏好数据、历史文案、资产信息整合流水线。
2. 基于主动学习的数据迭代：
- 过滤SFT模型采样多样性弱或奖励标准差低的样本；过滤RLHF前后奖励分数均值上升过快的样本；
- 基于奖励作弊现象（特定句式、符号、词汇、长度）合成负例数据，补充到偏好数据集。

3. 高质量偏好数据筛选：
规则+LLM的两阶段筛选策略：
- 启发式规则：包括基于编辑距离和偏好对BLEU过滤、根据资产信息去重、过滤沿用历史文案的样本；
- LLM筛选：采用Qwen3-32B从偏好对描述产品的一致性、文案可信度、用户修改动机及其实际修改效果等角度进行分析。最后留下1%的高质量偏好数据。

4. 多维度奖励系统设计：
- 奖励模型：先将文案中资产信息做哈希编码，训练标量奖励模型；针对RM过于关注文案而忽视提示上下文的问题，加入注意力均衡辅助损失；
- 规则奖励：长度惩罚，限制文案长度在可用区间；格式惩罚，要求模型严格按照提示词设置的模板输出内容；流畅性奖励，采用标准化困惑度；
- 相关性奖励：模型输出文案作为难负样本，历史文案作为正样本，微调bge-embedding，输出相似度作为相关性奖励；
- 生成式奖励：从资产信息正误、广告合规性、典型优劣文案对等方面合成带有是否可用标签的数据。GRPO训练Qwen2.5-7B判断广告语是否可用。

5. 文案生成模型强化精调：
5.1. SFT模型生成文案和优秀历史文案组成文案对，先做DPO微调；
5.2. 在6k广告文案生成任务数据上，做GRPO强化精调:
- 维度奖励系统：自定义奖励函数和奖励模型的加权平均作为最终奖励；
- 推理加速：把参考模型和奖励模型在训练环境外部部署多个实例，实现并行采样和打分；参考模型采样和流畅性奖励推理过程合并；生成式奖励4-bit AWQ量化；优化提示词，将固定规则前提充分利用前缀缓存；
- 超参优化：离线策略提高数据利用率；动态高温采样保证多样性；根据minHash LSH和长度阈值做动态采样保证采样多样性和可用性。

三、工作结果：
相比于前期SFT和DPO模型，输出多样性和可用性均提升；上线覆盖率从40%增加到55%；采纳率提升2.4个百分点。

理想汽车 2024年08月 - 2024年12月
大模型算法实习生 空间AI-语言智能 北京
一、工作概述：
参与理想大模型基座迭代，包括车机知识退火实验、高质量数据筛选和方言理解能力提升。
二、工作内容：
1. 方言指令理解：
- 测试方面，结合IP和方言词汇从线上日志捞取对应方言数据，采用LLM+人工专家的方式制作包含2k条数据的方言测试基准；
- 训练方面，采用翻译模型把普通话数据翻译为方言样本，扩充数据集。

2. 车机知识退火实验：
- 退火实验：学习率decay阶段学习率和性能关系实验；采用re-warmup的方式抬高起点学习率。
- 课程学习：采用专业词汇占比、文本困惑度加权的方式对领域数据打分，根据分数制定样本训练顺序。

3. 数据去重：
- 精准匹配：去除空格、标点、数字之后，基于MD5的文本去重；
- 文本去重：MinHash+滑窗把文本压缩为低维向量，通过LSH分桶去重；
- 语义去重：BGE模型得到语义向量后先将数据分片，再对每个分片的语义向量用DBSCAN聚类去重。

4. 三方面加权的高质量数据选择：
- 指令遵循难度：采用IFD对指令遵循难度打分（分数过高者滤除）；
- 指令丰富度：对指令打分类标签，对多标签的指令和涉及低资源标签的指令打高分；
- 文本质量：使用Qwen2对文本是否涉及深度分析和细节描述进行判断，对于有深度有细节的对话数据给予高分。

三、工作结果：
经历1次模型上线更新，团队在方言和车机知识增量训练方面的工作结果刊登公司公众号“有个理想”。

科大讯飞研究院 2025年02月 - 2025年03月
助理算法工程师 核心研发平台 北京
工作描述：微调多模态大模型，提升其理解文档图片的能力。
工作内容：
1. 两阶段训练InternVL2.5-8b（书生万象）：
- 阶段1-视觉文本和文本对齐：冻结LLM，增量训练ViT和MLP在表格解析、文本定位、理解化学方程式等任务；
- 阶段2-下游任务指令微调：解冻LLM，冻结其他。视觉QA、要点总结、文本阅读等多任务学习。

2.在SFT训练集的OCR任务中加入特殊标识（<row-span>或<col-span>）解决跨行跨列表格不能转markdown代码的问题。
3.设计多页文档的文本定位和解析特定页文字的任务，提升模型区分多页文档的能力。

比赛经历

WWW2025-阿里天池-多模态对话意图理解挑战赛 2024年12月 - 2025年01月
赛题背景：根据用户与客服的多模态对话，判断用户意图。
赛题难点：1.分类类别多 2.呈现长尾分布 3.大量未标注数据 。
解决方案：1.Qwen2-VL-7b隐藏层接分类头，不生成文字，直接输出类别。
2.采用两阶段的训练策略，提升长尾类别的识别性能。
- 阶段1：均匀采样数据，训练LLM和分类器，目的是学习到长尾数据集上的最佳特征表示。
- 阶段2：逆采样数据，得到平衡子集。冻结LLM，训练分类器。

3. 伪标签法：阶段1模型在未标注数据中筛选高置信度的尾部类别数据，将其添加到阶段2的数据集中。
比赛结果：初赛7th

科研经历

中国电信研究院合作-运营商网络流量预测算法研究 2023年09月 - 2024年07月
项目描述: 使用深度学习方法，围绕运营商业务特点挖掘序列特征，提升流量预测模型精度。
核心贡献：
1. 网络拓扑建模：在网络连接关系保密的情况下，采用图卷积神经网络建模网络交换节点的关系；
2. 流量序列聚类：互相关系数作为相似性度量，做k-means聚类；
3. 长时序列周期建模：自相关函数确定周期长度，采用滑动平均获得周期序列。
项目成果:
1. 投稿2篇一作论文：一篇**TII**；一篇已发表在ICCIP24，获**best paper**。
2. 预测算法部署到北京电信城域网运维系统中，预测误差5%范围内节点占比98%。
3. 评选为北邮研究生科研创新 A 级项目。

个人总结

1. 实习比赛：3段大模型实习，一段多模态比赛。
2. 专业知识：在强化学习、分布式训练、多模态有实践经历。
3. 科研经历：从事时序预测研究。一篇TII在投，一篇ICCIP24最佳论文。
4. 学业学工：本硕均一等奖学金；本科辅修计算机。哈工大优秀本科生党员；北邮实验室纵向党支部宣传委员。