

石心 Shi Xin

15995721202 | deerspiritx@gmail.com



教育经历

苏州大学 211 双一流

2018年09月 - 2022年06月

信息与计算科学 本科 数学科学学院

主修课程：数学分析、高等代数、数据结构、计算机网络、计算机组成、算法设计、数据库原理

苏州大学 211 双一流

2022年09月 - 2025年06月

人工智能 硕士 计算机科学与技术学院

研究方向：自然语言处理、计算机视觉

实习经历

第四范式(北京)科技有限公司

2024年05月 - 2024年09月

NLP算法工程师

实习期间熟练并掌握Linux命令和docker的使用，掌握对LLM进行SFT的方法(lora, qlora, 全量)，并使用lmdeploy, vllm框架对LLM进行加速推理和部署，同时也掌握LLM结合RAG技术解决实际问题。并独立或合作完成以下项目：

1. 多轮对话问题理解

项目介绍：本项目旨在提高对多轮对话问题理解的准确程度。给出一组多轮对话，判断最后的提问和是否需要改写，若需要则给出改写后的问题，要求改写后的问题跟上下文语义连贯且无歧义。改写的F1和不改写的F1均值为评价指标，并要求TP99<1s。

大小模型相结合策略介绍：第一步使用双向LSTM模型判断是否改写，第二步进行改写，使用qwen2-1.5B大模型进行改写。

模型微调：首先对网上获取的多轮对话数据集进行清洗，并采用GPT4生成更多训练数据来增强数据集，是否改写的数据比例是1:1。然后使用清洗后的数据集对LSTM模型进行训练用于判断是否改写，采用需要改写的数据集全量微调qwen2-1.5B大模型用于改写。

推理：为了加快推理速度，满足TP99<1s的业务要求，采用了LMdeploy框架对qwen2大模型进行加速推理。

成果：在众多参与该项目的算法工程师中，使用微调后的大模型策略效果排名第一。

2. 大模型基础能力

项目介绍：测试挖掘各种大模型的语数英以及小语种的选择题和写作能力,用于实际业务。

策略介绍：采用COT+FEWSHOT的方式设计了多种prompt，测试了qwen2、glm4、internlm2.5、gemma2等大模型在10B、20B、80B参数量下的选择题与写作能力。最佳选择题策略为多次投票（TTA）策略：prompt1提示大模型给出题目的思考过程（重复多次）+prompt2提示大模型结合问题和多次的思考过程给出答案,最后进行答案集成确定最终答案，我们还在以上策略的基础上增加了RAG技术增强模型能力，通过引用外部数据库提高模型对问题的理解和回答准确度，并使用vllm和lmdeploy框架加速推理。

成果：在众多参与该项目的算法工程师中，策略效果在10B、20B、80B参数量赛道上都排名第一。

项目经历

2023多模态情绪识别挑战赛（MER 2023）

2023年03月 - 2023年06月

问题：情感分析需要进行量化的操作，而量化的体现则是离散和连续的表达。该挑战要求参与者预测离散和连续维度情绪。

方法：①A模型使用训练数据集进行训练，并从测试数据集中选择高置信度的样本；②使用该高置信度测试数据集样本训练B模型，目的是选择与测试集特征分布对齐的训练集样本；③将从②的训练中选择的训练数据集样本用于训练最终的C模型；④采用模型融合的投票法与平均法选择A、B、C模型中的推理结果；⑤采用当前较为流行的Bert模型进行文本编码，采用Wav2Vec2进行声音编码。

结果：在23支参赛队伍中最终排名为第6名。

篇章感知的情感原因蕴含（已中ICONIP 2023）

2022年09月 - 2023年02月

问题：对话篇章结构表示话语之间的交互关系，其中包含情感子句与诱因子句的潜在关联。本研究分别通过隐式和显式的融合篇章结构信息来编码这种潜在联系，增强模型捕获诱因子句的能力。

方法：①引入篇章分析作为辅助任务来隐式的加入篇章信息，使编码器可以感知对话篇章的结构和关系；②使用GNN编码对话，以话语为节点、结构关系为边，显式融合对话内篇章、话者、相对距离等结构性信息，提升模型对对话结构的感知能力。

结果：在Dailydialog数据集中Macro F1分别提升了1.6%和1.3%，联合起来总共提升了2.9%。

技能证书

考级证书：CET-6笔试、CET-4口语B等级、普通话水平测试二甲等。

荣誉证书：专利《一种电商、旅游平台用户评论回复生成方法、装置及设备》；2023年研究生二等奖学金；2023大学生创新创业大赛三等奖；2021大学生蓝桥杯省赛三等奖；2021年建模美赛S奖；2021苏州大学优秀少数民族专项奖学金；2020年中国大学生计算机设计大赛（游戏设计）全国二等奖；2020年度志愿时长优秀成员；2018大学生新生军事技能训练优秀学员。