

黄杰

电话：13787852186 · 邮箱：2802311325@qq.com

性别：男 · 籍贯：湖南-岳阳

Github：https://github.com/shangxiaabb

Blog：https://www.big-yellow-j.top/



教育背景

中南财经政法大学-统计与数学学院，应用统计，硕士研究生 2023.09 - 2026.06

- 校园经历：1、学院研究生招生办兼职
- 研究方向：文档智能解析、文档多模态模型

湖南工商大学-理学院，数据科学与大数据技术，本科 2019.09 - 2023.06

- 相关课程：文本数据挖掘、机器学习算法、数据挖掘、金融大数据分析

实习经历

iWUDAO 吾道科技 算法工程师 2024.06 - 2024.12

- 项目一、表格线条识别：1、数据集构建：设计并编写 Python 脚本，基于 PDFminer 库对 PDF 文档中的表格线条进行精准识别，处理样本图片约 1 万张。2、算法模型构建：根据已收集的数据集，执行数据增强操作（如图像扭曲、模糊处理、加印章等），然后采用 UNet 算法进行训练，最终实现自主定义的识别效果函数，识别精度达到：0.97。
- 项目二、文档智能解析：1、数据集构建：通过 Python 脚本（基于 PDFminer 库）从金融年报中提取文本内容，共提取约 1000 份 PDF 文件。采用正则匹配方式和 PyPDF 方法筛选可能的标题，并构建文档层次结构。2、算法模型构建：设计并实现融合图像与文本特征的算法模型。通过 ResNet-50 模型提取图像特征，利用 ROI Align 方法进行特征区域对齐；结合 Qwen2.5-0.5B 模型对文本进行编码，综合图像与文本信息通过注意力方式进行融合而后构建文档层次结构，并校验文档阅读顺序，最终识别效果达到：0.91（使用 TEDS 评估指标）。
- 技术栈：1、深度学习框架：Pytorch 分布式训练（数据并行）；2、Python 工具库：PDFminer、PyPDF；3、LLM 微调技术：LoRA（Low-Rank Adaptation）；4、Docker、Linux 基本使用

项目经历

预训练小型百科问答 LLM 2024.10-2025.04

- 模型框架：基于开源 LLM 框架（主体结构 Llama 为主），修改结构：将前馈层替换为混合专家模型（MoE），并通过 Flash-Attention 加速注意力计算；
- 数据集构建：1、预训练数据集：主要来自开源数据集，包括 Wiki 百科、百度百科以及部分竞赛数据。获取数据后，采用 MiniHash 对数据进行去重，随后通过 Qwen 的 Tokenizer 对所有文本进行编码；2、微调训练数据集：一部分数据来自从现有数据集中筛选出的高质量样本，并利用各类 LLM 的 API 生成问题，进一步提高数据质量；另一部分则直接使用相关竞赛的开源数据集。
- 训练方式：为保证模型的正常训练，采用 Deepspeed 框架（Zero-2）进行高效训练；
- 微调过程：在预训练完成后，使用 LoRa（PEFT 框架）对预训练模型进行微调，以提升模型的任务适应性和表现。
- 技术栈：1、深度学习框架：Pytorch 分布式训练（数据并行）、DeepSpeed、PEFT；2、Docker、Linux 基本使用

基于贝叶斯算法的 MCDI 最佳实验条件获取 2022-2023.09

- 数据集获取：MCDI 数据集主要通过实际实验获取，实验与东北电力大学进行合作；算法设计：采用树算法（如 XGBoost、随机森林等）进行建模，并结合贝叶斯优化算法调整超参数。通过训练得到的模型，再次利用贝叶斯优化算法来确定最佳的 MCDI 实验条件。后续，借助 Shap Value 方法解决机器学习算法的可解释性问题，从而提升模型的透明度和可解释性

专业技能

- 编程语言与工具：熟练使用 Python 编程语言，熟悉 PyTorch 框架，了解 SQL 和 C 语言；熟练使用 Linux 系统并掌握基本命令操作；了解 Docker 基本操作。

- **深度学习与算法**：熟悉深度学习核心算法（如 Transformer、GNN 等），相关框架（如 DeepSpeed、FlashAtten 等）以及分布式训练等。

校园经历

卡耐基演讲与口才协会部长 部长 2019.09 - 2020.09

- 任职一年内，在社团内部扮演核心人物，组织 40 人左右的活动，和自己的搭档们连续 2 次获得湖南工商大学社团辩论赛冠军，在辩论赛中多次获得最佳辩手

中南财经政法大学统计与数学学院研究生办公室 助理 2024.03 - 2024.09

- 协助办公室老师整理材料，帮助老师送材料。日常会议时，帮助老师布置会场，并且与其他部门老师进行共同组织活动。

获奖经历

本科阶段

- 湖南工商大学数学建模三等奖
- 全国大学生数学建模湖南省三等奖
- 湖南工商大学统计分析比赛二等奖
- 湖南工商大学校庆优秀演员
- 湖南工商大学优秀团员
- 湖南工商大学社团辩论赛冠军
- 湖南工商大学”芙蓉学子”校级团体奖学金

研究生阶段

- 中南财经政法大学研究生入学奖学金（2023.09）
- 中南财经政法大学研究生三等奖学金（2024.09）
- 第十四届全国大学生市场调查与分析大赛湖北省三等奖（2024.03）
- 第十届全国大学生统计建模大赛湖北省研究生组二等奖（2024.06）
- 第七届全国应用统计专业学位研究生案例大赛湖北省二等奖（2024.05）

专业证书

英语四级、证券从业资格证、英伟达技能培训合格、创业培训合格证、教师资格证高中数学