

# 刘亚武

手机: 13671241876 | 邮箱: liuyawu1987@126.com

年龄: 36 岁 | 现居: 北京海淀区



## 个人优势

- 工作背景:** 拥有超过 10 年的技术领域工作经验, 涵盖数据库管理、算法开发及 AI 框架构建等多个方向。曾担任 DBA、机器学习工程师和算法工程师等职位, 积累了丰富的数据处理、模型优化及系统集成经验, 能够结合业务需求设计高效的技术解决方案, **具备系统化思维和跨领域协作能力。**
- 专业技能:**
  - ✓ **熟悉传统机器学习算法 (如 KNN、决策树、逻辑回归、SVM、贝叶斯分类、ALS 推荐算法) 及集成学习方法 (如随机森林、Adaboost、GBDT), 掌握聚类算法 (如 Kmeans、DBscan)。** 深入理解卷积神经网络 (CNN)、循环神经网络 (RNN)、长短期记忆网络 (LSTM) 等深度学习模型, 能够使用 TensorFlow、Keras 等框架进行复杂项目开发, 并对 VGG、GoogleNet、ResNet 等先进网络架构有实际应用经验。
  - ✓ **具备 3 年以上大模型相关经验, 熟悉主流大模型 (如 GPT、LLaMA、DeepSeek 等) 的微调和优化, 掌握强化学习、模型蒸馏及 RAG 检索优化技术。** 能够针对 GPU/TPU 算力资源进行调度与性能优化, 熟悉推理优化工具 (如 vLLM、Ollama), 并具备从模型训练到部署全流程的实践经验。**熟练使用 Python、Java 等编程语言完成大模型相关任务,** 结合大数据技术实现高效数据处理与模型迭代。
  - ✓ 熟悉 Hadoop、Hive、Spark SQL、Flink 等大数据计算框架, 掌握 HBase、MongoDB、Redis、ELK 等 NoSQL 数据库技术。精通 MySQL、Oracle 等关系型数据库, 能够熟练编写 SQL 语句并进行性能调优。**熟悉 ETL 工具 (如 Kettle) 对数据进行清洗、抽取及加工, 具备高敏感度的数据分析能力。**
  - ✓ 扎实的 Java 编程基础, 熟悉 Scala、Python 等语言, 能够使用 Numpy、Pandas、Scikit-learn、Matplotlib 等库完成数据分析与可视化。熟悉 Linux 系统操作, 能够使用 Shell 脚本进行基础编程, 搭建并使用 Nginx、Flask、Spring Boot 等 Web 服务框架。
- 综合素质:** 具备较强的学习能力和适应能力, 能够快速掌握新技术和工具, 紧跟 AI 领域的发展趋势。擅长独立分析和解决复杂问题, 能够在多任务环境中高效完成工作。具有良好的团队合作精神和沟通能力, 能够与研发、产品及运营团队紧密协作, 推动项目顺利落地。同时, 具备较强的文档撰写能力, 能够清晰记录技术方案和关键决策, 为团队**提供技术支持和知识传承。**

## 工作经历

- | 2022.02-至今     | 北京智汇云舟科技有限公司  | 算法工程师 |
|----------------|---|-------|
| ● <b>数据关联:</b> | 负责基于 <b>Neo4j 搭建图数据库</b> , 结合公司业务数据完成深度集成, 实现高效的数据关联查询。通过 <b>Django 框架独立部署 Web 应用</b> , 提供知识图谱的查询展示功能, 支持输入 ID 获取网格码三维坐标、时间等多维信息。参与设计和优化图数据库结构。   |       |
| ● <b>语音处理:</b> | 主导实时语音识别项目, 采用 ASRT+CTC 模型实现声音检测、录制、上传及识别的自动化流程。 <b>利用 WebSocket 技术搭建服务端与客户端的实时通信架构</b> , 支持用户通过麦克风输入语音并实时返回识别结果。完成从第一版基础功能到第二版优化版的迭代开发, 提升系统对实时语音数据的处理能力。                                  |       |
| ● <b>图像建模:</b> | 基于 PIFuHD 算法库, 开发从 2D 图像到 3D 人物模型的高分辨率重建流程, 支持多人和单人图像生成 PNG、MP4、OBJ 格式文件。 <b>整合 OpenPose 工具生成人体关键点配置文件, 使用 FFmpeg 工具将 OBJ 模型转换为动画视频。</b> 搭建 Flask 框架网站, 支持异步任务管理, 任务完成后自动生成 ZIP 压缩包并提供下载链接。 |       |
| ● <b>自然语言:</b> | 利用 <b>WebSocketAPI 实现语音识别与虚拟人状态生成的无缝对接</b> , 结合自然语言处理技术优化对话系统的响应逻辑。 <b>通过 PyTorch 和 Mellotron 模型完成 24 种不同声音的合成</b> , 涵盖男声、女声、儿童声等多种类型。构建包含 1.5 亿条常识三元组的知识库, 支持用户输入问题的精准解析与回答, 增强对话系统的交互体验。  |       |

2019.05 - 2022.01

北京华宇信息技术有限公司

机器学习工程师/算法工程师

- **算法开发**: 负责安防和广告机项目的**算法调研、技术路线设计及模型构建**, 完成从**特征工程到模型训练的全流程开发**。通过深度学习框架实现模型调优, 并对算法进行验证和迭代改进, 支持项目关键模块的功能实现。参与数据预处理流程的设计与优化, 解决实际业务场景中的复杂问题。
- **数据管理**: 负责 ETL 系统的 **BUG 修复**, 快速定位并解决多个关键问题, 优化**数据抽取、转换和加载流程**。完善数据查询及资源管理系统, **改进查询算法和数据库结构**, 提升数据处理效率。配合团队完成数据清洗、存储及分析的相关工作。
- **系统集成**: 配合产品研发部完成产品的功能测试、联调及维护工作, **提供算法层面的技术支持**。参与多模块的集成测试, 分析并解决跨系统协作中的技术难点, 推动产品从研发到上线的顺利过渡。**撰写算法技术文档**, 记录开发过程中的关键决策和技术细节。
- **模型应用**: 参与机器学习及深度学习项目, 运用专业知识解决实际业务问题, 重点**负责模型选择、训练过程优化及性能评估**。结合具体应用场景设计实验方案, 完成模型的部署与调试, 支持算法在安防、广告等领域的落地应用。

2017.01 - 2019.04

中化能源科技有限公司

DBA/图像处理算法工程师

- **数据管理**: 负责数据库系统的架构设计、开发及优化, 涵盖 MySQL、SQL Server 和 Oracle 等多种数据库。完成业务系统数据库的日常维护。制定并实施数据库备份与监控策略, 设计存储过程, 审核开发工程师的数据库设计方案, 指导数据存取方法。
- **算法创新**: 负责算法调研、技术路线设计及模型构建, 完成从算法实现到调优的全流程开发。参与图像处理相关算法的设计与优化, 解决实际业务场景中的复杂问题。

2011.11 - 2016.12

碧海蓝天(北京)科技有限公司

DBA

- **数据运维**: 负责数据库备份策略的制定与实施, 定期检查备份集的完整性, 确保数据存储的可靠性。审核开发人员的 SQL 语句, 参与初期表结构建模, 规划分区表的使用与约束条件。
- **故障应对**: 处理数据库异常宕机事件, 分析问题原因并采取紧急恢复措施, 保障业务系统的连续性。通过日志分析与性能监控, 定位潜在问题并制定优化方案, 支持数据库的稳定运行与长期维护。

## 📁 项目经历

### 知识图谱构建与应用

- **项目描述**: 搭建了一个基于 Neo4j 的图数据库, 用于构建和存储公司的知识图谱。该系统整合了公司的业务数据, 并利用 Django 框架在服务器上部署了一个 Web 应用, 实现了一个独立的查询展示平台, 允许用户通过输入 ID 检索相关的网格码、三维坐标和时间信息。
- **项目职责**:
  - ✓ **数据建模**: 负责知识图谱图数据库的整体架构设计, 定义节点、关系及属性结构, 确保数据模型能够准确反映业务实体间的关系。
  - ✓ **数据集成**: 实施数据集成策略, 将业务数据映射到图数据库中, 完成数据清洗、转换及加载 (ETL) 流程的设计与实现。
  - ✓ **系统实现**: 开发 Web 应用前端界面, 提供用户友好的查询入口; 后端服务负责处理复杂查询请求, 与 Neo4j 进行高效的数据交互, 支持多维数据的实时检索。
  - ✓ **权限控制**: 设计并实现系统的安全性机制, 包括权限控制、数据访问审计及安全防护措施, 保障业务数据的安全性与合规性。
- **项目成果**:
  - 1) 实现了对大规模图数据的快速检索, 查询响应时间从原来的平均 800ms 缩短至 480ms, 效率提升约 40%。
  - 2) Web 应用的部署显著提升了用户访问知识的便捷性, 用户满意度调查显示满意度提升了 25%。
  - 3) 系统上线后, 日均查询量达到 10,000 次以上, 高峰期支持并发查询量超过 500 次, 系统稳定性达到 99.9%。
  - 4) 数据集成覆盖公司核心业务数据, 累计存储节点数超过 100 万, 关系数超过 500 万, 为业务决策提供了全面的知识支持。

### 中文情感分析

- **项目描述:** 基于酒店评论数据集的情感分类项目, 目标是将用户评论分为正面评价和负面评价两类。训练集包含 3000 条正样本和 3000 条负样本, 测试集包含 7000 条正样本和 3000 条负样本。该模型可用于客户问答体系、留言系统等内容的筛选与分类。
- **项目职责:**
  - ✓ **文本清洗:** 使用 jieba 对文本进行分词, 并通过停用词过滤、特殊字符删除 (如 “\r”、 “\n”) 及字符长度判断等策略完成数据初筛。
  - ✓ **特征提取:** 利用 TfidfVectorizer 对文本进行词频-逆向文件词频 (TF-IDF) 向量化, 调整参数包括 ngram\_range=(1,5)和 max\_features=12000, 以捕捉更丰富的语义信息。
  - ✓ **模型训练:** 对比多种机器学习算法 (如朴素贝叶斯、SVM、随机森林、AdaBoost、XGBoost) 的效果, 最终选择 RandomForest 模型进行训练与测试, 调整参数包括 max\_depth=130、n\_estimators=600, 并设置 n\_jobs=-1 以实现多核并行计算。
  - ✓ **优化策略:** 将训练集中的正样本数量翻倍, 使正负样本比例达到 4:1, 缓解类别不平衡问题; 同时删去降维环节 (PCA、LDA), 保留原始特征空间以提升模型性能。
- **项目成果:**
  - 1) 训练集准确率达到 99.9%, 测试集准确率达到 96.4%, 有效验证了模型的泛化能力。
  - 2) 在实际应用中, 模型能够高效处理大规模文本数据, 单次预测耗时平均降低至 20ms, 支持每秒处理约 50 条用户评论。
  - 3) 模型被成功应用于客户问答体系和留言系统, 日均处理评论量超过 10,000 条, 错误分类率低于 3.6%。

AI 框架生成三维人体图像

- **项目描述:** 开发了一款基于 PIFuHD 算法的 AI 框架, 专门用于从高清 2D 图片生成 3D 人物模型。该项目通过先进的深度学习技术, 实现了人体重建的自动化, 能够将平面图像转化为高分辨率的三维人体模型, 为虚拟现实、增强现实及游戏动画等领域提供高质量的内容支持。
- **项目职责:**
  - ✓ **算法研究:** 负责 AI 框架的核心算法研究与开发, 专注于 PIFuHD 技术的应用与优化。设计并实现从 2D 图像到 3D 模型的完整转换流程, 包括图像预处理、特征提取、深度估计和三维重建等关键步骤。
  - ✓ **性能优化:** 针对模型的计算效率进行优化, 调整网络结构和参数配置, 减少计算资源的消耗。在保持高分辨率输出的同时, 提升模型的推理速度, 确保其在实际应用中的高效性。
  - ✓ **工具整合:** 整合 OpenPose 工具生成人体关键点配置文件, 并使用 FFmpeg 工具将生成的 OBJ 模型转换为 MP4 格式动画视频。搭建 Flask 框架网站, 支持任务上传、异步处理及结果下载功能。
  - ✓ **部署实施:** 完成 AI 框架的构建与部署, 支持多用户并发任务处理。优化任务管理模块, 确保任务完成后自动生成 ZIP 压缩包并提供前端下载链接, 提升用户体验。
- **项目成果:**
  - 1) 优化后的算法将单张图片生成 3D 模型的时间从原来的平均 20 分钟缩短至 5 分钟, 效率提升约 75%。
  - 2) 系统上线, 日均处理任务量超过 200 个, 高峰期支持并发任务数达到 50 个, 模型生成成功率稳定在 98%以上。
  - 3) 该技术已成功应用于虚拟现实和游戏动画领域, 累计生成 3D 模型超过 10,000 个, 客户满意度调查显示模型质量评分达到 4.8/5。

📁 教育背景

2020-2024	东北林业大学	产品设计   本科
2008-2012	北京人文大学	计算机及应用   本科