

徐鸣远 | 算法日常实习生 · 搜索 / 推荐 / 广告 随时到岗 实习时长：3个月以上

电话：+86 18164267192 | 邮箱：mingyuanxu.unsw@gmail.com | GitHub：MasterpieceXu

教育背景

中国地质大学 | 地下水科学与工程 | 工学学士 | 2020.09 - 2024.06

新南威尔士大学 UNSW | 计算机科学与工程学院 | 信息技术硕士 (AI) | 2025.05 - 2027.05 (GPA 4.0/4.0, 前10%)

实习经历

中国地质大学产学研基地 & 武汉 | 数据分析实习生 | 2023.07 - 2024.05

- **数据处理与建模**：使用 Python 对地下水监测数据进行清洗与可视化分析。构建 Random Forest 回归模型，对水质关键指标进行拟合与预测。
- **结果分析**：基于模型输出的特征重要性 (Feature Importance) 识别主要影响因子，绘制数据分析图表，相关结果直接用于辅助科研项目报告撰写。

项目经历

MovieLens-32M 生成式推荐系统 (Generative Recommender) | T5, RQ-VAE, PyTorch, Gin-Config

- **生成式检索架构落地**：基于 TIGER 范式重构传统的“召回-排序”两阶段漏斗，实现端到端的 Next-Item Generation 推荐。系统支持千万级数据的 **5阶段流水线** (清洗-量化-序列化-训练-评测)，通过 Gin-Config 实现实验配置解耦与可复现性。
- **语义索引与显存极致优化**：针对 ID 类特征显存占用大的痛点，引入 RQ-VAE (Residual Quantized VAE) 进行残差量化。将 70k+ 电影的高维 Text Embedding 压缩为 **128x128 分层离散语义 ID**，在保留语义信息的同时将索引显存压缩至 **<100MB**，成功在单卡 L4 环境下解锁全量训练。
- **冷启动与长尾攻坚**：利用语义 ID 的 Content-aware 特性，使得新物品无需交互即可获得具备语义含义的 ID。实验表明，模型在 Zero-shot 场景下的 **Recall@50 达到 0.18+**，显著解决了长尾物品 (Long-tail) 在传统协同过滤模型中无法被召回的难题。
- **工业级工程实践**：设计模块化代码架构，集成 OneRec-lite 多目标生成策略与 DPO (Direct Preference Optimization) 偏好对齐接口，提升推荐结果的多样性与用户满意度。

十万级全链路新闻推荐系统 (News Recommender System) | PyTorch, FAISS, Pandas, YouTubeDNN

- **工业级多路召回架构落地**：重构传统单一召回链路，独立设计并开发新闻资讯个性化推荐系统。构建集特征提取、模型训练、多路召回、加权融合于一体的数据流水线，并通过完善的本地数据存储与缓存机制 (Cache)，极大提升了特征构建与实验调参的效率。
- **深度表征学习与毫秒级检索**：针对用户与物品交互特征稀疏的痛点，基于 PyTorch 构建 YouTubeDNN 双塔召回模型，精准提取用户定长 (Top-30) 历史序列与物品的高维 Embedding 表征。引入 FAISS 向量数据库构建 IndexFlatIP 索引，将高维 Embedding 的近线相似度检索耗时降至毫秒级。
- **多维度策略协同与长尾攻坚**：针对资讯推荐的时效性与覆盖率问题，设计 4 路并行召回策略 (启发式 ItemCF、基于向量求交的 ItemCF-Emb 与 UserCF-Emb，及 YouTubeDNN)。设计 Min-Max 归一化机制与动态置信度权重配置，综合挖掘行为轨迹与全局共现性，有效保障推荐候选集的多样性。
- **内存优化与工程化实践**：设计高度模块化的在线预测 (Online Predict) 接口，支持动态加载预训练模型库。针对大规模计算场景，采用动态批量采样技术 (Dynamic Sampling) 计算 Embedding 相似度矩阵，在单机环境下显著降低内存开销，实现算法工程化的平滑落地。

其他技能

- **编程语言**：Python 熟练，SQL，C 了解
- **深度学习 / 机器学习**：PyTorch, Transformers, LightGBM, XGBoost, Scikit-learn, Pandas, NumPy, Faiss
- **推荐 / 评估与工具**：召回-排序架构, AUC, Recall@K ; Flask, Git, Linux, Azure, Colab, $\$R^2\$$