

工作经历

2019 ~ 2023

网易云音乐

职位: 资深推荐算法工程师 所在部门: 音乐事业部

职责业绩:

- 1、结合业务需求,对音乐歌曲、歌单、视频等多个推荐场景数据建模
- 2、负责前沿机器学习算法的调研及实现,并应用到召回、粗排、精排阶段,探索更高效商业模式
- 3、在线推荐服务的稳定性监控,优化线上代码性能,构建完善报警体系全面监控线上机器
- 4、4篇专利发表(2篇一作,2篇二作)

2017 ~ 2019

杭州海康威视数字技术股份有限公司

职位: 大数据算法工程师 所在部门: 海康威视研究院

职责业绩:

- 1、多维物联网时空知识图谱构建与安防综合研判模型训练
- 2、一站式多维物联网大数据挖掘平台设计与研发
- 3、大数据挖掘项目框架设计、特征工程构建、算法训练、模型部署
- 4、3篇专利发表(均一作)

2015 ~ 2016

天津神舟通用数据技术有限公司

职位: 数据挖掘工程师 所在部门: 商务智能与数据分析事业部

职责业绩:

- 1、参与数据挖掘项目工作(需求调研、算法训练、模型部署及分析报告撰写)
- 2、研发并沉淀业务通用算法及算子(数据挖掘、文本挖掘等)

教育经历

学校: 南开大学 专业名称: 管理科学与工程(数据挖掘方向) 学历: 硕士 时间: 2013 ~ 2015

学校: 西北师范大学 专业名称: 计算机科学与技术 学历: 本科 时间: 2008 ~ 2012

技术栈

语言: Java、Scala、Python、SQL、C++

算法: 深度学习、机器学习、自然语言处理、图挖掘

项目经历

首页歌单推荐项目

2022-2023

项目职务: 探索首页推荐歌单新模式,建模提升业务转化率

项目描述: 探索首页歌单第一坑位流式推荐模式,拓宽用户消费深度,填充首页资源位紧张缺陷;构建云音乐通用歌单推荐能力和架构,支持新业务快速上线。应用spark、hadoop、hive等分布式技术完成特征样本构建及存储;利用tensorflow深度学习框架,搭建神经网络完成召回、排序多个阶段的指标优化。

项目职责:

- 1、分析首页下滑率和首页各位置转化率指标,上线多目标粗排模型,优化流式歌单模式,提升用户转化率,模块整体点击率提升7.5%以上;红心率提升15%。
- 2、利用用户、歌单、长短序等多种特征,优化点击率、用户播放时长、用户收藏等多个指标,构建多

塔多目标排序模型；线上推理C++算子开发，提升推理效率，rt下降40%；机器资源节省50%。

3、探索长尾item冷启动分发策略，构建歌单多级流量扶持体系，提升独家自制内容播放效率。

4、完成通用歌单推荐从0到1的推荐框架的搭建，开发实时、离线多种召回策略提升候选物料多样性；支持流式+列表式多种推荐模式，已支持场景歌单、情绪歌单、歌单广场多个场景上线。

单曲-视频跨域推荐项目

2021-2022

项目职务：从0到1实现视频-单曲跨域推荐场景落地

项目描述：负责视频推歌、边听边看多个视频-单曲跨域推荐算法研发，基于视频分发孕育爆款单曲，提升云音乐视频用户渗透率，增加用户亮屏时间，为商业化变现引流。

项目职责：

- 1、预训练item、user的embedding特征，应用到召回、排序等多个阶段；对新老用户应用不同模型提升优化ctr、前端观看时长、人均VV等指标。
- 2、构建单曲-视频知识图谱，开发并上线多种图挖掘算法，完成单曲-视频离线跨域召回；开发长短行为序列向量召回、双塔等多种模型对用户进行实时召回。
- 3、利用用户在单曲场景行为特征，探索跨域学习模型在音乐推荐场景的落地，实现丰富单曲域知识迁移到视频域，持续提升用户体验。

单曲个性化推荐项目

2019-2021

项目职务：解决私人fm长期遗留问题，优化业务指标，提升效率

项目描述：负责音乐推荐私人fm场景的指标优化，解决长期遗留问题

项目职责：

- 1、推荐资源池日益扩充，但精排模型对排序数量敏感，通过加入播放pv、收藏率等多目标优化排序模型解决问题，完成指标提升收藏率提升10%；播放时长提升5%。
- 2、由于fm流式播放对实时特征的敏感性，对上述粗排模型实现增量学习；开发flink实时任务，实现实时样本落盘，且label归因正确率达到97%以上。实现每1h更新一次线上模型，收藏率提升5%，播放时长提升3%。
- 3、在fm已有精排模型基础上添加丰富实时特征、序列特征；利用快照技术保证线上线上特征一致性，构建并上线复杂精排模型。
- 4、fm场景因上线年代久远，代码版本老旧，对推荐主流程代码进行并行性改造；开发c++算子改造传统Java推理模式，rt降低30%。

物联网时空知识图谱构建及应用

2017-2019

项目职务：图谱构建+综合研判数据挖掘

项目描述：基于安防行业特点，创新性的采用规则驱动细粒度继承时空动态本体构建技术，搭建混合存储架构，支持多数据源，实现知识图谱快速构建。基于图谱，实现公安领域风控预警、案件预测、嫌疑人推荐等综合研判数据挖掘场景，最终沉淀行业通用场景模型库。

项目职责：

- 1、通过图谱规则驱动技术，抽象出算法规则、公式公理，采用动态本体继承机制，构建图谱。
- 2、数据探索及预处理，融合实体、关系、事件等信息，构建面向分析主题的特征工程。
- 3、算法建模方案设计与实现，基于stacking算法框架不断优化算法效果
- 4、基于知识图谱的综合研判功能，沉淀嫌疑人积分模型、案件地图、时空碰撞等安防业务决策能力

项目职务： 架构设计+算法研发

项目描述： 深入分析现有机器学习平台特点，结合海康物联网数据特性，完成平台需求调研、技术选型、架构设计、接口设计、数据库设计、可行性论证，攻克 DAG workflow、算法并行化、分布式算法框架、高可用微服务化扩展等技术难点，提供一站式的可视化数据分享挖掘能力。

项目职责： 1、实现 DAG 驱动的工作流调度引擎，训练流程接入丰富的数据处理环节
2、设计多类型组件扩展框架，具有良好的扩展性和通用型，结合行业需求实现包括预处理、分类、回归、特征工程在内的算法组建，有效支撑行业场景
3、深度整合 spark 计算框架，优化继承体系，抽象公用算子，提升算法能力，提高开发效率，同时抽象体系提供跨语言算法应用能力
4、深入分析产品形态，抽象包括 DAG、CMS等核心微服务模块，为平台的高可用、高并发、横向扩展提供了良好的基础

中科院自动化所专家职称评定系统

2016-2017

项目职务： 方案设计+模型构建

项目描述： 根据专家的研究成果和影响度，对全国专家学者进行“长江学者”、“青年拔尖人才计划”的评审。系统实现查询专家在 SCI、EI、CSCD 文献数据库论文的收录与引用情况，对专家进行评审。

项目职责： 1、编程实现 SCI、EI、CSCD 论文爬取，根据不同的检索条件获取专家论文列表，统计不同专家论文的收录与引用，并将论文的收录与引用写入本地文件
2、将不同文献数据库爬取信息建立索引并存入数据库，实现专家成果的离线检索
3、由于同一篇论文被不同论文引用，引用格式繁多，通过文本相似度计算，精确得到不同专家的引用成果
4、建立利用论文引用数据，建立人物关系网络，实现基于社交关系的信息推送

信通院模块化成果库信息系统建设

2016-2017

项目职务： 方案设计+模型构建

项目描述： 经过多年的积累，信通院积累了内容丰富的软科学成果资源。通过利用信息抽取、非结构化数据挖掘、知识图谱等技术，构建信通院模块化成果库工作平台；开展模块化成果分析与挖掘，绘制科研知识全景图、领域知识图谱，实现科研成果的可视化。

项目职责： 1、实现 word、PPT、PDF 多种格式的研究成果和软科学出版社的解析，建立模块化知识的思维导图，绘制信通院研究领域知识全景图。
2、利用文本分词、词性标注、特征词提取、命名体识别、文本相似性计算等文本挖掘算法，实现模块化知识的拆解和领域知识图谱的绘制
3、将撰写研究报告的专家通过词共现算法挖掘人物关系，展示人物关系网络图，实现基于社交关系的知识成果的推荐
4、采用技术：Java、检索、文本挖掘算法（命名体识别、文本分词、文本相似性计算、特征词提取、词共现）

项目职务： 方案设计+模型构建

项目描述： 利用文本挖掘技术，辅助国家审计署对全国审计工作进行审计。本项目主要负责项目申报文档的解析，对存在重复项目申报现象进行审计，帮助客户降低工作量，提高审计精度；从互联网爬取与审计有关信息，及时发现和揭露违法违纪问题和重大风险隐患；从申报项目中进行新词识别，提取审计词语，并写入审计专用词典，为后期文本挖掘提供支持。

项目职责： 1、实现项目申报文档的解析，提取关键词，并进行文本聚类实现文本信息进行相似性比较，发现潜在重复申报项目的情形

2、实现网络信息的爬取，利用百度、bing、google、搜狗、雅虎等搜索引擎，通过基础 url+ 输入关键词进行多线程检索，并解析外链，获取审计相关的舆情数据

3、实现审计词典管理，对申报信息进行新词识别，并写入新词词典

4、采用技术：爬虫、检索、文本挖掘（关键词提取、新词识别、文本聚类）

证书及荣誉

- 1、大学生英语四、六级证书
- 2、普通话等级证书
- 3、英语托业证书
- 4、个人专利7篇
- 5、海康威视研究院年度优秀员工