

教育经历

爱丁堡大学-人工智能和计算机本科 (QS 世界排名前 20)

2018 年至 2022 年

主要研究方向: 数据分析, 推荐搜索, 机器学习和算法优化; 熟练掌握 Java, Python, SQL; 绩点排名年级前 10%。

爱丁堡大学-人工智能和计算机硕士

2022 年至 2023 年

主要研究方向: 推荐系统, 数据分析。

项目经验

电影推荐系统(基于 TensorFlow)

- 项目描述: 根据用户已经观看电影的评分来预测未观看电影的评分和预测用户喜欢看什么电影, 从而实现电影推荐功能。
- 主要工作: 清洗数据, 提取特征, 协同过滤。通过 PCA 和 normalization 来降低数据维度和解决数据疏离度问题, 使用 CNN 模型进行建模分析。运用方法一: 基于电影的协同过滤, 根据不同的电影种类和平均评分等方面, 结合已有的用户评分数据和观影喜好, 通过建模计算来找出最接近用户喜欢的电影。运用方法二: 基于用户的协同过滤, 根据影评人, 找出影评人对所有评论过的电影的平均分数, 并且由此找出影评人的偏好, 通过比较不同影评人和用户找出最接近用户喜好的影评人, 推荐影评人喜欢的高分电影。
- 项目结果&收获: 预测分数准确率同比增加 10%, 达到了 85% 的准确率。掌握常用的数据清洗方式 PCA 和 normalization, 掌握基础协同过滤方法: K-mean 聚类算法, 线性回归, Euclidean algorithm 等方法。

无人机动态规划飞行路线 (Java)

- 项目描述: 根据无人机位置和空气检测点位置, 找出最佳飞行路径, 以最短的步数走完所有的空气检测点, 确保无人机安全返回起点, 通过 Gjson 画出无人机路线图。
- 主要工作: 数据是无人机坐标和空气检测点坐标, 使用机器学习算法: 贪婪算法, A*算法, 遗传算法, 比较不同算法的优劣。
- 项目难点: 避开禁飞区和逃出禁飞区陷阱。通过增加判断条件来避免无人机由于算法局限性卡住在禁飞区内来回不断飞行消耗步数。改进算法, 避免陷入死循环。
- 项目结果&收获: 所有无人机都能在电量未耗尽前, 读取所有空气检测点并返回初始点。无人机效率提升 25%; 掌握算法优化, 算法改进

自然语言处理 (NLP)

- 项目描述: 通过语料库数据来训练语言模型, 进行词性标注, 句法处理, 涵盖 NLP 的基础模型和算法。
- 主要工作: 使用 Python 导入语料库数据, 进行模型训练, 应用标准序列模型如 hidden Markov Models (HMM), 应用分类模型 (贝叶斯和 MaxEnt), 通过 Smoothing 操作应用解决稀疏数据问题。
- 项目结果&收获: 模型准确率达到 85%, 同比增长 20%。掌握基础 NLP 模型生成和数据处理。

实习经历

京东生鲜-数据分析与推广

2021 年 4 月至 8 月

- 主要工作: 整理优化公司内部需求, 分析用户画像, 挖掘生鲜产品卖点, 跨部门协作与产品经理, 工程师, 区域门店店长合作推进项目, 协助项目落地, 评估投入产出比, 曝光量, 项目 GMV。
- 工作目标: 从京东主会场引流到生鲜会场, 增加京东生鲜会场的曝光量和 GMV, 刺激消费者下单购买, 增加销售额; 协助线下门店的开业的线下引流, 提供流量支持, 推广曝光支持等。
- 难点: 分析京东生鲜主要用户群, 并结合本次项目主打产品的卖点, 针对性地去触达用户感知; 与产品经理, 工程师沟通协调, 开发趣味玩法: 小游戏派发优惠券, 关键词一键触达热卖商品, 提高会场界面的有趣性。
- 结果与收获: 共协作 5 家线下门店开业, 协助推进 4 个生鲜大促项目, 京东生鲜官方视频 TVC 共计 4 条, 采买抖音微博 KOL 超 50 人, 达人视频创作超 20 条, 原创文章 5 篇, 单条视频“广东荔枝”视频全平台播放量超一千万, 总曝光量超 1.5 亿, 同比增加 50%

华泰证券(金工研究)-数据分析与建模

2020 年 6 月至 11 月

- 主要工作: 基于卷积神经网络模型(CNN)进行人工智能选股, 运用多因子选股, 在全 A 股股票池中进行测试和预测, 评估模型表现, 回测时间从 2008 至 2018 年, 最长时间维度 10 年, 最短时间维度 6 个月。
- 工作内容: 将个股的因子数据进行特征提取和特征降维, 形成二维形式, 考虑时间为截面期, 将所有个股放置同一截面, 便可以得到所有个股的因子图片, 滚动训练集和验证训练集, 并且经过交叉认证调参, 测试模型
- 工作结果与收获: 模型进行回测评估, CNN 模型年化超额收益率约为 15%, 最大回撤控制在 7%之内, 优于朴素贝叶斯模型和线性回归模型。通过人工智能算法构建选股策略是历史经验的总结, 存在失效的可能, 数据只具有回溯参考性, 不具有未来的预测性。