

郑行 算法工程师（机器学习/NLP/推荐系统）

电话：18696069951 邮箱：zheng_xing@aliyun.com

政治面貌：中共党员 年龄：24 居住地：湖北武汉

实习 offer：蚂蚁金服-机器学习算法；今日头条-推荐算法；360-智慧城市算法

SCI 论文：IEEE Transactions on Signal Processing 无线通信方向论文在审

奖励证书：英语六级 542，计算机等级 2-4 级，国家励志奖学金，知行，优干



教育经历

教育背景 2017.09 ~ 2020.06 华中科技大学 电子信息与通信学院 信息通信工程 硕士 5%
2013.09 ~ 2017.06 华中科技大学 电子信息与通信学院 电子信息工程 本科 10%
学生工作 2017.09 至今 电子信息与通信学院 通信系党支部书记，研究生会副主席

专业技能

1. 熟练使用 Java, Python, 熟悉常用数据结构及算法。
2. 熟悉机器学习算法与深度学习 CNN, RNN 经典算法模型，熟悉 Scikit-learn 与 TensorFlow 等算法架构；熟悉常见推荐系统 CTR 预估算法如 GBDT+LR, FFM, AFM, Wide & Deep, Deep FM, xDeepFM 等，了解图表征学习算法。
3. 熟悉 NLP 相关领域，如实体抽取、文本生成、语义匹配等，熟悉经典的预训练模型 ELMO, GPT, BERT, GPT-2 以及 XLNet 等。

实习经历

暑期实习：蚂蚁金服 大安全 反洗钱产品技术部 算法工程师 时间：2019.06 ~ 2019.08

项目一：基于命名实体识别的企业名制裁扫描匹配算法

职责描述：面向蚂蚁金服自研制裁扫描引擎 WISE 的多语种企业名制裁扫描算法优化，解决现有算法中存在的非关键匹配，简称匹配以及误写匹配等问题。在保证制裁名单全覆盖时，降低制裁业务的打扰率。

主要工作：将命名实体识别算法引入公司名匹配场景，基于经典的实体识别模型 LSTM+CRF 改进出 Transformer + CRF 模型及后续优化，模型上线将站内（蚂蚁+AE）企业名制裁打扰率由 0.2%降低到 0.01%，输出站点由 2%降低到 0.05%。

项目二：基于差异表征的可疑交易文本风险定位算法研究

职责描述：面向蚂蚁金服反洗钱相关可疑交易业务，旨在挖掘并表征可疑文本与正常交易文本的差异，以辅助定位可疑文本中的风险来源，降低运营审理压力，并提供一定的可解释性。

主要工作：基于词嵌入实现词汇级的文本差异定位算法落地；基于语义匹配模型 Match Pyramid，利用 Attention 机制以及 Saliency map 进行语义级差异定位。模型基于赌博可疑交易场景上线，缩短运营平均审理时间约 30%。

比赛经历

https://tianchi.aliyun.com/home/science/scienceDetail?spm=5176.12281920.0.0.372b3f746ZKZdh&use_rId=1095279390141（天池主页）

比赛一：（推荐系统）阿里移动推荐算法 比赛成绩：92/10399 时间：2018.12 ~ 2019.03

项目描述：以阿里巴巴电商平台的真实用户-商品行为数据为基础，同时提供移动时代特有的位置信息，赛题要求利用一个月内用户对商品全集上的行为数据集构造用户集合对于商品子集中的推荐模型，并输出用户在接下来一天对商品购买行为的预测结果。评价指标：精确度、召回率和 F1 值。

项目难点：2 万用户，百万级商品；时间以及位置特征的合理利用；特征组合；商品冷启动。

主要贡献：特征工程：特征族，多维度二次组合特征，滑动窗口截取时间相关特征；

模型与算法：LR, GBDT, Wide & Deep, DeepFM;

物品冷启动：相似性挖掘——kd 树与 LSH 以及 Exploitation & Exploration。

比赛二：(深度学习) 异烟酸合成工艺优化 **比赛成绩： 86/2682** **时间： 2019.01 ~ 2019.02**

项目描述：异烟酸合成过程每个步骤会受到温度、时间、压强等各方面因素的影响，造成异烟酸收率的不稳定。项目要求以异烟酸生产过程中的各参数，包括各主要步骤的时间、温度、压强等参数为基础，设计精确智能的优秀算法，以提升异烟酸的收率。评价指标：预测收率均方误差，最优产率。

项目难点：特征含义未知；异常值敏感；训练集规模小(1397)，易过拟合；模型最优参数求解复杂度高。

主要贡献：特征处理：缺失填充，异常检测、分箱、One-hot 编码，时间点与时间段特征处理等；

模型算法：NN，决策树，集成学习 RF、XGBoost、Light-GBM 及 Stacking；

参数优化：梯度下降法，坐标下降法，粒子群算法。