



X X X

基本信息



电话:

工作年龄: 1.5 年

现居住地: 上海

邮箱:

政治面貌: 共青团员

专业背景

具有 python+mysql 等项目开发经验, 熟悉 tensorflow 框架。熟悉 feed 信息流推荐基本流程, 能够在生产环境下工作 (如 linux 语言环境等); 熟悉常见推荐算法框架并结合业务作出改进。如 USERCF/ITEMCF、SVD/SVD++、FM、LFM、GBDT+LR、FM+FTRL、WIDE&DEEP、DEEPPM 等。熟悉常见深度学习算法及相关实践经验, 如 LSTM、RNN、CNN 等。熟悉 ATTENTION、图神经网络等机制原理及应用。如 TRANSFORMER、BERT、GCN 等。熟悉常见机器学习算法。如逻辑回归、决策树、随机森林、GBDT、LDA 主题模型。熟悉阿里 MAXCOMPUTE, SQL 常见查询及复杂查询语句, UDF 函数, 并熟练使用这些工具进行特征工程及特征变换。熟悉 DOCKER 及 JECKINS 常规工程化操作。熟悉 HBASE 及 REDIS 常规数据存取及常规数据查询操作。

工作经历

2020.04 - 2021.01

推荐算法工程师

工作片段一: 计算广告项目一期

项目简述: 公司已在电商消费领域积累了大量的数据, 该数据可以在广告领域发挥其商业价值, 尤其是针对电商的广告领域, 因此为广告主和流量主提供数据和技术支撑, 进而创造更大的收益。

参与并负责项目整体架构设计。数据特征分析, 预处理及特征工程均在阿里大数据平台 maxcompute 上进行。由于数据主要来自签约店铺订单数据, 缺少用户性别, 年龄等基础人口信息, 通过分析丰富的用户购物行为进而推断出用户的购物性别, 年龄等。为了保证训练目标唯一性, 对订单数据反复清洗, 最终抽取近一年的数据作为训练数据, FM 模型进行训练得到训练集及测试集的 auc 分别为 0.98, 0.6, 出现明显过拟合, 考虑到数据是很突出的长尾分布, 分桶后的连续型数据分布会更加不均衡, 更加单一, 导致模型学习不够充分, 因此重新调整数据分布及编码问题, 并通过正负样本采样解决正样本 auc 小于 0.5 的问题, 最终提取近七天数据作为测试集, 近七天以前的三个月数据作为训练集, FM 模型的 auc 稳定在 0.9 左右。线上服务能达到日均 30w 左右的收益。

工作片段二: 推荐排序一期

项目简述: 细致挖掘用户潜在兴趣焦点, 在小微智能问答窗口中提供更多个性化服务, 进而实现千人千面。

负责梳理规整商品画像底层数据, 将 500 万左右 100 多个字段规范化;

基于订单、用户和商品画像数据进行特征挖掘, 特征分析, 并经过 XGBOOST 进行特征筛选, 提取 TOP30 个特征基于以上特征进行 FM 模型训练, 发现训练集及测试集 AUC 分别为 0.9 以上, 0.7 以下, 为进一步增加模型泛化能力, 加入交叉特征, 模型 AUC 稳定在 0.89 左右; 由于交叉特征不方便线上调用同时为了提取更多抽象特征, 采用 DEEPPM 模型。模型 AUC 最终稳定在 0.90 左右。并将用户及商品画像数据存储于 HBASE 中, 线上购买率提升 0.2 左右;

工作片段三: 智能穿搭一期

项目简述: 为腰部以上服装品牌商家提供智能穿搭服务, 为卖家提升用户购物体验, 为商家提高购物转化率。

基于快速上线的要求, 作为项目主负责人, 项目一期整体分为三个阶段分别为召回, 过滤, 排序。召回阶段中, 由于服装的细分类目存在较多差错, 因此人为总结一套 key_words 及 not_key_words 将服装分为四类-上下衣, 外套及裙子, 准确率达到 98% 以上, 并根据其他业务规则召回出大量可搭配的 pair 对商品; 过滤阶段主要是对年龄, 性别, 季节等运营规则过滤掉不合适的搭配; 排序阶段中, 为了实现业务尝尽中左右滑动单衣变动及上下滑动整套变动, 将 pair 对进行拆分, 训练 doc2vec 模型及 tf-idf 模型根据相似度找到相似单品序列。考虑到用户的兴趣强度会随着展示时间的增长而减少, 对商品热度进行 decay。最终将产生的 match 表和 sim 表存入 hbase 中供线上调用。目前在线服务店铺 cvr 提高 0.2 左右。

工作片段四: 智能穿搭二期

二期主体流程框架不变, 依旧分为召回, 过滤和排序阶段。在召回阶段, 由于一期规则较少且宽泛, 导致大量召回后的 pair 对多样性极低, 重新加入更细致的服饰专家业务规则, 同时通过 cv 手段提取分割后商品主图的颜色, 最终准确率和多样性提高 30 以上。由于规则过于死板, 泛化能力及延展性远不如模型, 因此召回阶段加入了模型, 对比神经网络, dssm+cv+nlp, 基于 gcn 的 fashion compatibility 等算法, 从实用性及扩展性考虑, 最终选择基于 gcn 的 fashion compatibility 构建智能穿搭算法。为了得到优质的训练数据, 整合 cv, nlp 及推荐相关领域的知识主导并推进构建了训练数据库, 得到模型需要的推荐规则, 图片及商品名称。通过图像组已训练好的 resnet 模型提取图像 embedding, 通过 nlp 组的 bert 模型提取文本向量, 将两者 concat 后保存后, 使用 fashion compatibility 模型构建智能穿搭算法, 最终模型准确率达到 0.98 左右。

2019.12 - 2020.04

算法工程师

工作片段一：智能推荐项目

为提高模型泛化能力，构造时间穿越特征以及向量特征搭建 deepfm 模型，线上转化率提高 3.5%左右。为更高效提取交叉特征，训练 fm 模型并提取二阶项，最终 auc 提高千分之五，线上转化率提高 1.5%左右。加入向量召回分支-word2vec，后为避免因文本序列长短不一造成的问题，改用 node2vec，auc 提高千分之三，线上转化率提高 1.2%。离线计算存储，线上召回时调取候选集。使用 LR 作为基准模型，离线训练数据集，为融入深层交叉特征信息同时减轻手工特征负担，运用 wide&deep 算法，线上转化率提高 8.5%。为减少模型耗时，wide 部分优化算法使用 ftrl。为提高模型准确率和覆盖率，有效添加客户情感分析模块，通过 lstm 神经网络引入情感倾向的 embedding 特征，并将 wdl 模型替换为 deepfm 模型，线上转化率成功提高 8%左右。为减少负采样对原始分布的影响，修改损失函数进行补偿。

工作片段二：视频推荐项目

主责推荐排序。使用具有较好拟合能力的 xgboost 算法进行排序，auc 为 0.75 左右。为了加快模型训练及提高自动交叉特征能力，使用 lightgbm+FM 组合模型，由于以上组合模型未挖掘深层交叉特征，后使用 deepfm 模型，并通过 din 中的 attention 机制学习用户兴趣 embedding 向量输入 deepfm 中，auc 提高至 0.83 左右。

工作片段三：智能推荐项目

除基本用户画像以外，又以店铺销售价格作为核心指标进行用户画像探索，按区域、一级类目、二级类目等分别进行探索，并借助 GBDT,RANDOMFOREST 等模型进行特征优选，最终筛选前 20 个特征作为模型训练指标，在召回阶段，基于 lmf 使用 item2vec+faiss 和 user2vec+faiss 进行向量化召回，当用户行为序列不够长时，word2vec 召回效果会下降，而 node2vec 会进行广度和深度搜索，构建关系网络，可避免此问题，因此使用 node2vec 进行召回，最终线上效果提升了 2%左右，由于 word2vec 和 node2vec 仅仅使用了 id 类特征，导致召回模型不够精细，因此我们尝试使用 DNN 模型进行深度召回，将用户侧、店铺侧、浏览时间等特征，用 ball tree 算法确定召回列表，最终 auc 提升千分之八。本着快速上线原则，以 LR 模型作为 baseline，而上线后出现覆盖率低，推荐效果差等问题，经过分析，为了平衡用户长短期兴趣我们使用了阿里的 DIN 模型，经过 shuffle 和 rerank 后，最后上线效果明显改善。

教育背景

校园经历

2018.09 - 2019.07 老年手机评论情感分析研究

以京东老年手机评论作为切入点，深层挖掘老年人对智能机需求特点。项目职责：主责模型训练、优化及技术文档编写。使用 RNN 网络进行情感倾向训练，产生梯度弥散等问题，因此选用具有特殊门结构的 LSTM 神经网络。对 LSTM 神经网络进行 FINE TUNE，取前三层并加入输出层，将通过 BERT 预训练好的语料作为输入得到正负情感样本，最终准确率提高至 99.8%，AUC 提高

实践经历

《基于机器学习的老年人手机用户中文情感分析-以京东数据为例》、《大数据背景下中小型服装电商地域销售特征的研究-以淘宝汉服的数据挖掘为例》。

证书荣誉

自我评价

有扎实的专业知识，有良好的理解能力和实践能力，能高效率地检索海量数据并提炼出核心论点，用于解决实际问题；本人性格乐观开朗，工作自主性强，思维敏捷，有亲和力，富有责任心，具备良好的沟通协调能力和敬业精神和职业道德操守。适应能力强，能迅速的适应新环境并融入其中。擅于发现疑难点并积极主动地去处理各类棘手的问题。具有创新意识，擅于观察和总结，敢于尝试新鲜事物，敢于挑战自己，追求完美，注重细节。团队合作意识强。执行力强，以结果为导向设立各类目标，能接受加班，服从领导安排。