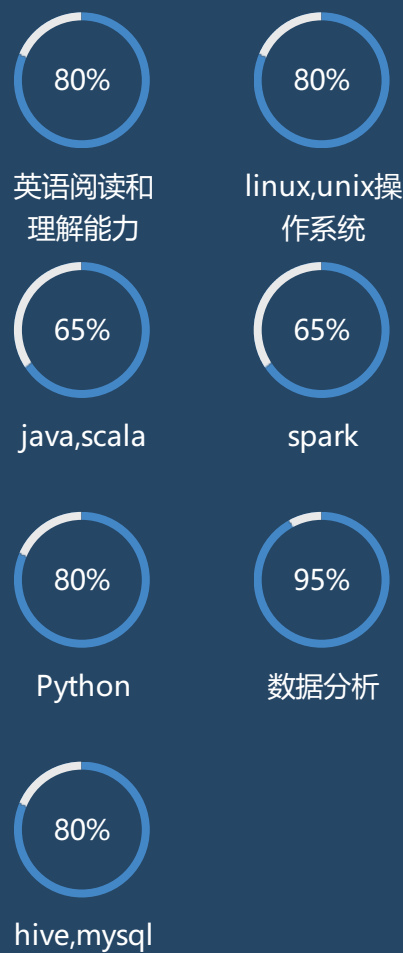




技能特长



奖项荣誉

荣获IT总部大数据中心2017年度优秀员工

2016年第四届全国大学生统计建模大赛全国二等奖

2015年度全国大学生数学建模大赛全国三等奖

2013-2016年分别获得全日制硕士统计专业二等奖两次，一等奖

总想看的更高，望得更远，所以我选择不断攀登

求职意向

机器学习与自然语言处理 1周内到岗

教育背景

2013.9-2016.7

统计学（硕士）

研究生阶段选择应用统计为主要方向，期间主要从事数据挖掘与机器学习相关的项目开发工作。以非参数和半参数混合效应回归为研究主题发表核心论文两篇，并在毕业论文评选中，获评优秀论文。

2009.9-2013.7

数学（本科）

本科主要学习数学分析、空间几何、概率论与数理统计、线性代数、实变函数、常微分与偏微分方程、计算机技术、matlab等课程。成绩名列学院前5，荣获优秀毕业生称号。

工作经验

2016.7-2018.3

算法工程师

主要以标签开发和预测两大业务为产品线进行研发，并以此为基础展开算法研究与优化工作。标签主要以商品和会员两个维度展开，商品属性标签的挖掘，主要通过规则设计以及正则表达式提取来实现落地，会员标签则是结合关联规则，聚类，逻辑回归等算法展开研发。

围绕物流，供应链，财务等业务需求延展开的销量、作业量与GMV预测，主要以离线计算为核心，通过对历史数据的收集，展开特征维度的横向构建，结合集成学习与提升算法模型，通过不断调参，迭代，特征优化来提升预测准确率。

以商品和会员属性为载体，构建尺寸，城市，季节，年龄，偏好等2000多个标签维度的零售知识库。在供应链补货，仓库作业量等业务主线基础上搭建以量预测为核心目标的预测系统，以特征库为引擎支撑，提供可视化的前端效果。

一次

2013年度中南民族大学优秀毕业生

2009-2013分别获得全日制本科
数学与统计学专业二等奖两次，
一等奖两次

自我评价

虽然经验不多，见识也不够，但一路走来的沉淀与积累是我不断追求进步和更高层次有力见证。

虽然天赋不是最高，能力不是最强，但我觉得现在的我就是目前最好的我，意识到不足，而不断追寻更好的自己。

虽然不是出类拔萃，受众人喜爱的那位，但凭借坚定不移的决心，强而不衰的责任心，做好自己的本职，团结大家是实现当前价值的最佳定位。

2018.4-至今

数据挖掘部经理

处理来自爬虫引擎丰富的文本数据，以企业销售案例为种子文件，行业领域，主营产品信息为主要特征，构建相似搜索引擎，放大目标客户群体，将销售线索提供给需求客户。

基于业务拓展和产品推广，搭建公司主站后台nlp系统并做前端可视化展示，主要负责词向量，文本相似，主题模型，主体识别以及ocr接口开发，通过丰富公司的业务方向，对外展示我们在文本处理与挖掘上的能力与成果。

开发人机聊天系统，助力社交模式。通过丰富的语料以及规则逻辑，搭建宠物社交聊天引擎，以宠物为纽带，基于宠物的特征，属性以及成长历程，将主人的社交属性辐射开，匹配契合度高的交友对象，推荐给用户以达到合理，科学，快速交友的目标。

项目经验

2018.12-2019.2

宠物社交聊天引擎

算法研发负责人

作为建立人际交往圈的桥梁，寻找切入点的破冰者，以宠物作为媒介实现人际交往过程的准确切入，合理引导。基于微博，趣味论坛，步行街等丰富的且实效性强的对话式信息作为训练语料，第一层嵌套规则逻辑，主要以宠物，宠物主人以及用户提问式信息作为匹配对象的话题场景，提取用户输入主干信息，根据句式结构，匹配答案集结果；第二层以训练语料做支撑，搭建句子主干完全匹配，关键词匹配，以及句向量相似匹配等搜索引擎模块；第三层备选答案集处理无匹配结果情形，并抛出新的话题，作为新一轮对话的切入。

短期记忆功能扩展—采集用户与宠物的实时会话信息，通过句式分析，主干提取，对应标签化数据存储，做本地化缓存，生命周期为会话开启到结束，将该信息作为第一层和第二层逻辑的信息补充。

本期测试效果为：通过20个线程，500轮对话评估，回复合理性或准确率接近70%，可提升的空间和后期方向：1、在新词发现方向扩充自定义词典，减少未登录词带来的匹配干扰；2、搭建生成式seq2seq模型，通过迁移学习或自定义神经网络结构，将闲聊型模块以更具泛化效果的呈现。

2018.7-2018.10

nlp后台系统

算法研发经理

构建多功能主站nlp引擎，围绕文本相似，主体识别，主题生成和概要抽取等核心功能点，铺设业务线。在线索推荐功能模块，通过构建隐藏层数目为2，节点数512的双向RNN神经网络实现的主体识别功能，将大量TMT行业描述性新闻中公司信息提取出来，关联公司业务动态，主营方向，丰富线索特征，在保持80%以上的精确率与召回率的基础上将线索转化率提高5个百分点左右。

文本相似和概要提取丰富线索匹配特征维度，补充线索池功能性支撑，更有效的进行后台线索维度的标签化管理。在腾讯ocr接口的基础上，添加业务逻辑和匹配规则进行再封装，用于商标日期标识识别，将准确率提高至90%以上。

2018.3-2018.6

销售线索推荐

算法研发经理

以行业属性，产品信息为特征依托，辐射出多元化企业关联维度，依据文本描述信息，从文本包含信息的丰富程度构建相似搜索引擎，基于短文本，高质量的词表征信息，通过simhash建模查找相似；

长文本以预处理为关键环节，通过主干提取，将词特征转化为句向量表征，从多维向量的余弦距离计算相似性，基于相似结果，做离线化补充，整个相似度计算过程的前提依赖贝叶斯行业分类算法的类别划分。

算法轮询过程，从用户上传的种子案例，特征化处理，行业归集，再到相似搜索，最后沉淀线索，反馈给用户线索信息。通过离线化的结果的补充，将销售线索转化率提高了约15个百分点。

2018.11-2018.12

sem优化与反欺诈点击

算法研发经理

通过前端埋点数据，分析用户使用paperpass论文查重的行为轨迹，基于漏斗式转化路径，将用户基于搜索引擎关键词进入主站到最终跳出一系列的行为事件进行切分，汇总，归集，提供给归因模型计算关键词间接转化价值。并基于关键词转化周期和转化成本分类，一类是快速转化型，通过比较排名竞价算法提供最高出价和建议排名策略；另一类则采用自动竞价修正算法，针对不同的价值的关键词：高价值、低价值和消耗词提供不同的出价计算逻辑。

面向多渠道扩展，通过近一个月以周为轮次的AB测，各渠道在相同效益的前提下，优化后的成本均有下降，总体评估结果，每个月可以为paperpass在搜索广告营销投入节省近5万的成本。

2017.8-2017.12

仓库作业量预测

算法研发工程师

基于物流的库存优化和货架位置摆放诉求，对仓库SKU作业量展开量预测，以精准的量结果为导向，提高工作人员拣货分货效率。分三个阶段进行：

第一阶段，业务数据接入，特征设计，筛选，数据清理。历史销量数据采用ma3,ma5,ma10,var3,var5,var10构造，并以折扣粒度来计算促销提供的未来信息，共构建出140多个特征。

第二阶段，模型搭建，效果调试以及数据接入点选择。初期尝试在R服务器进行模型调试，以品类类目和销量相似度进行聚类归集，对每个类分别采用GBDT,Xgboost等提升算法，以及尝试神经网络如：LSTM，DNN等，进行效果比对，调参优化。

第三阶段，上线部署与运维监控。基于运维成本和监控便利性考虑，将整个模型的训练和预测任务部署在Spark集群上。准确率效果反馈：线上单品平均预测准确率达到67%，重点单品（长尾商品，热销商品）能达到85%，线下和天猫平均预测准确近90%。

算法研发工程师

依据商品适用年龄属性推导客户小孩年龄，推荐适龄小孩商品，丰富会员年龄标签信息，为精准化营销和个性化推荐提供数据支持。标签覆盖主要围绕母婴类目的SKU标签和会员偏好属性展开。

关联基于小孩商品衍生出来的会员行为记录，以商品适用年龄属性为基础特征对小孩年龄展开区间估计，规则化购买周期和商品保质期，预估小孩的个数。以丰富的会员行为数据为支撑，精确率达到：75%，召回率85%左右。

偏好设计上，短期偏好部分，通过会员—商品Item关系，构建依托行为次数的关系矩阵。结合商品记忆随时间指数衰减模型赋予的权重，求解出偏好得分，最后根据得分高低筛选出偏好前3的品牌（商品组）。而长期偏好模型部分则通过主成分分析获取基于每个会员行为商品所产生的次数作为特征的方差贡献度，给出相应的主成分得分，根据得分高低输出偏好商品。

尺寸，适用城市等标签，则是依赖规则以及正则提取，作规范化标准输出。城市标签准确率100%，覆盖率接近100%；商品尺码准确率98%以上，覆盖率92.8%。