

# 基本信息

- 石振强
- 男，92 生人
- 中共党员
- 13655422693(微信同号)
- 已婚，现居青岛

# 教育背景

山东科技大学 本硕

- 2011年9月至2015年7月 本科 计算机科学与技术
- 2015年9月至2018年7月 研究生 人工智能-数据挖掘

研究生课题介绍

以教育数据挖掘为课题，针对在线教学平台的教育数据进行知识挖掘。主要分析平台上学习者的测评结果数据，详细挖掘了各类学习者的行为和产生此行为的原因。进一步针对不同类型的学习者的做题情况，提出了一种题目类型的描述方式,即由学习者的不同类别来对题目进行描述。通过对题目类型进行描述,我们可以更好的体现学习者与题目之间的关系。相对于仅把难度作为题目的分类标准，此次分类对于教学提供了极大的帮助。以此为基础，结合学习者的特征、做题情况、登录情况等，实现了不同阶段考试（期中、期末）的成绩等级预测，不同于以往的学习者成绩预测分析，研究学习者的特征时,加入了题目类型的因素,而且采用了多种分类算法进行组合。最后设计对比试验,验证了在学习者的特征中加入题目类型因素,可以提高模型的精度。以此，教学者可以根据预测结果，更早的预判学习者掌握情况，最终可以有有的放矢，因材施教。

涉及模型算法包含聚类算法（kmeans）、集成学习算法(RandomTree, GBDT等)，模型精确度达90%以上。

在校期间成绩优异，多次获得校奖学金。曾获美国数学建模二等奖，软件著作权一项。

# 工作经历

- 2017年8月-2017年11月 青岛智能研究院 研发岗 实习
- 2017年11月-2018年2月 高校信息产业股份有限公司 大数据算法实习岗 参与大数据平台的设计和数据挖掘算法的实现
- 2018年7月-至今 青岛中瑞汽车服务有限公司 大数据 金融风控

# 项目经历

现担任PaaS工程组组长，负责大数据金融风控模型、客户私有云定制、车联网数据分析等TOB产品。

金融风控数据系统

金融风控评分卡模型是一个对贷中车辆风险评估的一个模型，主要以车联网数据为核心，为客户（金融保险、汽车租赁等客户）提供车辆的风险分数及等级，帮助客户实现有效的风险预判，及时规避风险。金融风控评分卡模型是一个基于多规则的混合模型，其输入是，全平台设备（900w+）、多种预警（7+）、各类特征(30+)、业务相关数据等，输出是每辆车的一个你风险等级和风险分数。每日计算结果留存Hbase并推送全量设备等级及分数至Kafka（部分为接口，部分为RocketMQ），由各业务部门消费并应用，现已平稳运行三年。

模型难点:依赖服务多、计算链路长、涉及数据源杂且数据量大。

#### 负责内容

##### 1、相关核心开发

聚集服务：帮助运营排查可疑风险聚集点（如黑区、拆车点、白名单等）。设计实现了全量聚集(MapReduce)计算（T+1d）及小批量实时聚集查询(redis-geo)。

里程计算：全平台所有车量里程数值(MapReduce)。

风控评分卡模型（金融数据系统核心）：Python实现设计开发，完成所有预警数据的处理、提取、分析及最终的模型计算。

通过使用MapReduce计算框架，实现部分大批量计算，提高了时效性；对各个独立模块进行流程划分、关键节点监控及错误重试，可以对链路长的数据链路进行快速定位和定时任务重启不影响数据准确性（用于异常因素的数据补充）；梳理复杂的计算规则，将规则分为过滤、计算、二次计算及匹配等类型，使得计算逻辑模块可以按需调整应对模型规则变更。

##### 2、私有云部署

负责风控评分模型所有的依赖服务（20+）的客户定制化私有云部署，已经为平安金融租赁、吉致、安吉等公司部署完成。主要负责所有服务的依赖梳理、代码规范、配置规范、链接规范及个性化定制开发内容。通过前期沟通、规范设置、客户需求与服务梳理，使得团队人员快速交付客户的私有云部署。

##### 3、其他

多部门数据流转沟通及确定、应用部门(saas)数据对接、数据流转设计等。

## 亿咖通车联网数据分析项目

吉利下属子公司的车辆数据分析项目，也是金融风控一个行业应用的典型案例。通过阿里云DataWorks构建实时数据仓库和离线数仓来推动大数据业务化运营，使用MaxCompute离线计算，其中spark做离线计算，hive做离线存储。使用flink进行实时事件侦测并推送kafka，flink-sql将实时事件落库到实时数仓（hologres），业务系统接口调用hologres进行应用。

#### 负责内容

##### 1、设计开发

负责架构落地，开发Flink实时计算模块。根据重要程度、耦合度、逻辑变更频率、上线影响范围等因素，主要将事件侦测（26+）分组为三大类，节省FLink资源的同时，减少因变更产生的风险。

核心事件:车辆的点火、启动、驻车、拖车等重要事件;

故障侦测类: 机油、漏气、报警行驶、加油等9个故障事件侦测;

维护侦测类: 轮胎充气、换机油、里程等维保类事件侦测。

应项目紧急，FLink人员紧缺，经短暂学习Flink，并上手开发核心内容，因数据异常问题过多，多次自主试错反馈产品完善需求。通过使用Flink时间窗口模块及状态设计解决了事件实时侦测的计算，项目上线后运行平稳，验收过程顺利，所需资源及设计均符合架构要求。

##### 2、交付

负责整个数据流转的文档编写、数据格式定义、部署及阶段性交付演示和最终交付演示等工作。

## 在线分析平台

这是本人主导且参与全程设计(模型生命周期、前后端对接、前端页面设计及流转)的一个内部系统，因发现产品与开发在工作中的痛点，想要进一步解决数据整合、特征加工、规则（算法）模型、模型监控、模型治理等问题。目的是规范流程、沉淀过程、积累数据资产。

#### 负责研发及系统设计

- 1、自定义规则模型
- 2、模型监控
- 3、后台表设计（mongo）及前端页面流转及部分UI图标提供。

亮点：

- 1、数据集及特征字段后台分析，前端页面统一格式，新增数据特征不新增页面，展示格式由接口(Flask)定义。
- 2、数据集的统计图，便于了解数据集的信息及各个特征的统计信息。
- 3、规则格式分类，采用python进行规则解析转换为mongo查询语句，并根据规则进行模型计算。此过程全部异步(Celery框架)进行，前端可显示任务状态，并可进行终止、编辑等操作。
- 4、模型效果比对，可以展示不同客户、不同省份、不同门店下的数据准确度。

通过此系统，原来需要产品或运营工作人员两天建立的模型，可在系统中五分钟建立完成并且可不断修改，可以通过模型效果展示页面及时展示模型的效果及比对结果，从而积累了数据、模型、特征、规则等重要成果。

## 车主画像

为服务运营部门的业务，方便精准车主营销和分群，要对平台车主进行画像。

担任项目经理，梳理不同数据源的特征计算，由更为熟悉特定数据及相应计算方式的同事担任该数据源特征计算，设计数据整合流程。通过配合开发完成200+数据特征的计算。

特征的存储采用mongo，基于特征的二次处理，形成画像特征。比如，驶过的城市个数大于3，那么其画像特征为多城活动。诸如此类的画像特征都是基于数据特征的二次挖掘及分析。设计并开发车主画像系统，展示车主的画像。进一步进行标签统计，统计各类标签下的车主群像。通过比对不同标签下的车主可以帮助产品及运营挖掘更多的业务场景。如不同城市的早高峰情况，不同标签下车主的偏好及用车行为等。提出自定义标签及自定义分类，可以由系统页面进行手工批量划分，这样可以对进一步车主画像的闭环有极大的帮助。

涉及中间件：

- Python
- Mongo(最新画像特征)
- redis(慢查询快照)
- hbase(历史数据查询接口)
- hdfs(文件存储)
- PostgreSQL(历史数据查询批量分析)

## UBI车险模型

基于驾驶行为以及使用车辆相关数据相结合的个性化保险产品，不同于简单的里程计算和出险次数的车险评估，本模型是通过挖掘车辆定位数据，进一步挖掘车主驾驶行为。

负责客户需求解读、UBI调研、特征挖掘，最后通过决策树模型得到一个较好的模型效果。其中，通过研究发现，早高峰的定位点密集程度作为一个有效特征，提高了模型的召回率。最终与产品开发的规则模型相结合，提高了模型的鲁棒性。

如上模型均采用Python 的开源机器学习包-sklearn。

## 自我评价

- 数据挖掘思维
- 善于分析总结
- 责任心强
- 有一定的学习能力及沟通能力

