题 目：租赁业务风险控制安全中台的设计与实现

学院： 软件学院 专业： 软件工程 学生姓名： 学号：

|  |  |  |  |
| --- | --- | --- | --- |
| 1. **项目背景** 2. 项目来源   本人在自2021年9月在北京自如科技有限公司实习，期间参与了该公司实际需求项目“租赁业务风险控制安全中台的设计与实现”。   1. 项目意义   随着互联网的发展，有一批黑色产业也在随之兴起，他们依靠互联网为媒介，利用非法手段从中获益，称之为网络黑产。2018年5月发布的《数字金融反欺诈白皮书》显示，2017年我国黑产从业人员超过150万人，年产值达千亿元级别。网络黑产遍布在互联网的各个领域，他们通过恶意注册账号开展网络诈骗、在平台发布恶意违法内容、利用平台交易渠道进行虚假交易、薅取营销活动和新人礼包优惠等方式，对平台进行攻击，从而从中获益。2019年，黑灰产团伙通过过期的优惠券漏洞攻击拼多多，获利数千万元。为了对抗网络黑产，减少公司损失，保障用户利益，风险控制产品也应运而生。  本人所在公司以租赁业务为主，涉及到的业务包括租房、收房、装修、搬家、保洁、维修、电商等。本项目所设计的业务安全风险控制中台，以公司业务为基础，通过建立公司范围的风险控制体系，设定一系列风险控制策略和规则，在黑灰产攻击平台时迅速识别，进行拦截或者反攻击，从而保障平台安全。   1. 项目的研发现状（目前同类产品或解决方案中存在哪些不足，或现实情况下，存在哪些问题，例如：效率低下、手工操作等）   目前，公司内部的风险控制系统建设刚刚起步，曾经设置过公司层面的风险控制规则，小部分业务领域有自己的风险控制策略，但是存在以下问题：  （1）虽然小部分业务领域有自己的风险控制策略，但是较为简单，如设置黑名单、限制同一用户某一操作单天次数上限等，易被绕过，且各个业务领域之间的风险控制策略和结果无法共享。  （2）曾经的公司层面风险控制规则，仅每天针对当天的用户操作数据，按照特定的规则进行统计，将统计数据导出为excel进行人工筛查。这种方式十分机械，筛查为纯手工操作，效率低下。规则简单、单一，所有规则为硬编码方式实现，不方便进行修改。并且规则和结果均受到时间限制，针对具有时间跨度的数据无法进行分析，不同时间产生的结果也不方便进行横向比较。由于这种方式过于机械，目前已不再使用。  （3）公司层面没有针对风险控制形成体系，即公司整体风险情况以及每个用户的风险情况无法进行评估和可视化展示。 | | | |
| **二、项目内容、项目目标与拟解决的关键问题**  **1. 项目的内容（范围）**  租赁业务风险控制安全中台是搭建在用户和公司各个子系统之间的的风险控制平台，以租赁业务和其余衍生业务为前提，接收用户操作，根据中台内部的规则体系对用户操作进行分析，向子系统提供防控建议。  其中，规则为逻辑表达式，可抽象为：“特征-运算符-阈值->结果”，如“年龄小于18，拒绝请求”，规则可由多个条件表达式组成，对表达式结果再进行逻辑运算。若某一条数据触发了结果，称之为规则命中。  系统主要内容包括：  （1）**规则配置**。系统通过脚本对常见特征、运算符、结果进行预定义，业务人员通过规则模板，针对不同事件和业务，使用可视化前端界面进行规则配置，将规则按一定分支及流程编排形成决策流和规则组。  （2）**规则解析、执行与展示**。当有事件触发时，系统对决策树进行解析，加载用户数据进行执行，使用大数据处理框架减少规则执行时间，并根据执行结果向业务方反馈用户风险系数和防控建议。系统根据规则模型执行情况动态调整用户风险分数，对规则命中情况和平台风险情况进行实时展示。  （3）**规则评价和阈值自动调整**。系统对前一段时间的用户操作数据进行分析，通过模型对未来一段时间内的规则阈值进行预测和动态调整。系统通过规则评价体系分析规则的有效性和合理性，若指标超出正常范围，向业务人员发送告警信息。   1. **项目目标（即预期交付物）**   设计并实现一个租赁业务风险控制安全中台，系统功能包括：  （1）**系统为业务人员提供规则配置功能**。系统根据业务逻辑设置不同事件，将常用特征、运算符和结果通过提前定义的方式写入系统，可通过API的方式进行修改。  系统提供决策树和规则组配置，业务人员可在可视化界面上，将特征、运算符、阈值、结果拼接为规则，将多规则通过逻辑运算（与或非）进行组合形成决策流，或将多条规则进行罗列形成规则组。  （2）**系统能自动解析并执行所配置的规则。**当事件发生时，系统自动根据事件和业务寻找对应的规则，对决策流进行解析，转化为可执行代码，同时加载用户数据进行执行。执行过程中，借助大数据处理框架加快规则运算速度，并根据规则命中情况生成惩罚措施集合，反馈给对应的子系统，同时动态修改所涉及的用户风险分数。子系统执行惩罚，向系统反馈该惩罚措施是否验证通过。  系统提供规则执行及规则命中情况的统计查看功能，通过图表形式显示每条规则的触发次数、命中次数、命中率、准确率，以及具体命中名单。  系统提供平台整体风险情况查看，可通过图表显示风险分在每个分数段的分布情况。通过系统可查看所有用户的具体风险分数以及在系统内的具体操作。系统根据风险分和用户标签判断用户风险程度，根据风险分、风险程度对所有用户进行筛选、排序，以及根据用户id和手机号进行搜索。  （3）**系统能自动调整规则阈值。**系统要求业务人员为每一条规则设定期望命中率和波动范围，针对每一条规则，根据前一段时间的数据进行分析，计算满足期望命中率的阈值，并通过模型对接下来一段时间内的规则阈值进行预测，以保证阈值的合理性，并将预测结果应用于接下来一段时间内的规则执行中。若阈值波动较大，向业务人员发送告警，提示检查系统是否受到攻击。  系统根据子系统反馈的惩罚措施执行情况，通过模型定时计算每一条规则的准确率和合理性，若结果超出正常范围，则向业务人员发送告警信息，提示对规则进行修改。   1. **拟解决的关键问题**   （1）如何**设计规则配置机制**是规则配置的关键问题：  根据初步统计，目前已有不同规则特征数百个。若所有规则特征均通过预定义的方式写入系统，开发工作量大，一点小小的改动就需要重新编写特征，造成冗余。如何设计并实现一套机制，以满足规则动态配置和方便调整的需求是规则配置的关键问题。  （2）如何**解析决策树**是规则解析的关键问题：  决策树为向业务人员展示规则逻辑的工具，但如何将其中的逻辑表达式进行解析，是其中的关键  （3）如何**设计有效的数据处理算法**是规则执行中的关键问题：  规则的执行基础是数据，而随着规则体系与业务领域的不断扩大，需要记录、分析的数据量也是庞大的。根据统计，系统单日接收数据量上百万条。在如此大量的数据处理过程中，如何将规则执行导致的延迟控制在用户和业务方可接受范围内，是规则执行的关键问题。  （4）如何**处理频繁读取数据库导致的处理延迟**是规则执行的关键问题：  在系统运行过程中，需要反复读取数据库以获取数据进行计算，关系型数据库的处理速度无法应对该系统过于频繁的请求。  （5）如何**设计预测模型**是自动调整阈值的关键问题。  规则的阈值受到时间、营销活动等的影响，不是一成不变的。如何设计模型，根据已有的数据预测未来一段时间内的合理阈值是阈值调整的关键问题。 | | | |
| **三、拟采取的解决方案及可行性分析**  **1. 理论依据（可选）**  **2. 拟采用的技术路线**  系统后端采用springBoot框架进行搭建，前端采用vue框架，数据库采用mysql+redis实现，通过redis进行缓存提高读写速度。数据处理部分采用flink大数据框架。  针对关键问题，下面给出拟采用技术方案。  （1）针对**规则配置**的关键问题，拟采用的技术路线描述如下：  有时间跨度的规则特征可通过抽象提炼出一部分共性特征，称之为原子特征，特征由原子特征和时间窗口构成。由于该系统所需要覆盖的业务问题仅为租赁业务和其他租赁衍生业务，因此原子特征和处罚措施是有限的，据分析，原子特征个数约为15个，处罚措施个数约为6个。  如对于两个特征“同一用户1分钟的登录次数大于5”和“同一用户5分钟的登录次数小于10”，特征分别为1分钟登陆次数和5分钟登陆次数，可抽象出的原子特征为1分钟的登录次数，这两条规则所对应的时间窗口分别为1分钟和5分钟。  因此，仅将原子特征通过预先定义写入系统，将时间窗口作为特征属性进行配置，通过这种方法实现规则的自由配置与动态调整。  （2）针对**规则解析中解析决策树**的关键问题，拟采用的技术路线如下：  决策树通过转化，加载复杂特征计算方式和操作符，可生成条件表达式，将条件表达式转化为AST抽象语法树进行语法解析，利用栈计算的方式执行。在实际应用中，可利用aviator表达式求值引擎辅助计算，aviator支持算术运算符、关系运算符、逻辑运算符等构成的表达式运算，具有性能高和轻量级的优点。  （3）针对**规则执行中数据处理算法**的关键问题，拟采用的关键技术描述如下：  将规则执行与原子特征的计算进行分离，采用flink大数据框架完成原子特征计算。业务方通过MQ消息队列将用户数据传入系统，通过flink订阅数据，按照原子特征的计算方式完成原子粒度的聚合，并将计算得到的指标写入数据库。在规则执行过程中，根据特征对应的原子特征和时间窗口从数据库中读取需要的数据，再根据操作符和阈值进行判断是否命中。通过这种方法减小规则执行过程中数据处理的时间，提升系统性能。  （4）针对**规则执行中频繁读取数据库导致处理延迟**的关键问题，拟采用的技术路线如下：  项目采用mysql+redis的形式解决读写延迟问题。需要写入数据时，首先将数据写入mysql数据库，并刷新redis缓存数据。读取数据时，首先从redis中进行查询，若查询不到结果再前往mysql数据库进行查找，并把内容更新进入redis。由于redis的读写速度远远高于mysql，使用redis进行缓存可以提高读写速度。  （5）针对**规则阈值自动调整中预测模型设计**的关键问题，拟采用的技术路线如下：  拟采用三次指数平滑法对阈值进行预测。三次指数平滑法适用于有趋势性和季节性的时间序列预测，具有理论简单、预测效果好的特点。三次指数平滑法首先通过公式对历史数据进行平滑，计算得出一、二、三次指数平滑值，再根据三个平滑值，通过公式计算得出未来的预测值。  在计算过程中需要用到自定义平滑参数，该参数拟通过梯度下降算法进行确定，以历史数据为基础，采用误差平方和作为损失函数进行求解。  **3. 可行性分析**  （1）前期基础：  ①公司前期已经完成对特定操作的数据埋点，如登陆注册、下单、抢券等，这些数据均可直接使用。  ② 公司内部有完整的MQ消息队列技术，可进行系统间的数据传输，接口完善可以直接调用。  ③ 公司内部有完整的权限管理平台，不同角色在本系统上的权限控制可以直接在权限管理平台上进行配置。  （2）理论基础：  本项目所用的主要技术包括spring boot、vue、mysql+redis、flink框架均为工业生产中非常成熟的技术。风险控制场景是flink主要应用场景之一，且基于 Apache Flink 和规则引擎的实时风险控制解决方案在业内也有先例，因此技术实现方面具有可行性。  （3）已有的积累  本人在校期间多次接触web项目，对于使用的主要技术有所了解，熟悉mysql和redis的相关操作。在大数据处理方面，曾经接触过spark框架，对于大数据处理有一定经验。 | | | |
| 主要参考文献：  [1] 陈亚军,肖海斌.基于Spring Boot和Vue框架的水库雨水情监测系统设计与实现[J]. 软件 2021,42(09),29-31+41.  [2] 刘广轩,黄山,胡佳丽,段晓东.面向Flink流处理框架的主动备份容错优化[J/OL].浙江大学学报(工学版):1-9[2022-02-23].  [3] 封宇,周杰. 基于Flink的景区人数监测平台设计[J]. 电子技术. 2021(11).  [4] 马云,七七.当互联网业务安全来到反欺诈风口[J].上海信息化,2017(09):63-65.  [5] 常昊. 互联网金融风控系统贷中风控和预警功能的设计与实现[D].南京大学,2020.DOI:10.27235/d.cnki.gnjiu.2020.001604.  [6] 曹邦兴.自适应动态三次指数平滑法在交通预测中的应用[J].吉首大学学报(自然科学版),2019,40(05):13-17  [7] 张鲁男,常宝岗,梅利.基于规则引擎及智能阈值的实时业务风控系统[J].通信技术,2019,52(11):2720-2724. | | | |
| 毕业设计（论文）进度安排： | | | |
| 序号 | 毕业设计（论文）各阶段内容 | 时间安排 | 备注 |
| 1 | 参与完成整个系统的需求分析； | 2021.10.20至2021.10.31 |  |
| 2 | 参与完成系统的总体设计； | 2021.11.1至2021.11.31 |  |
| 3 | 设计并实现规则配置模块； | 2021.12.1至2021.12.20 |  |
| 4 | 设计并实现规则解析模块 | 2021.12.21至2021.1.14 |  |
| 5 | 设计并实现规则执行和结果展示模块 | 2021.1.15至2021.2.19 |  |
| 6 | 设计并实现规则评价和阈值自动调整模块 | 2022.2.20至2022.3.15 |  |
| 7 | 进行集成测试、确认测试、系统测试； | 2022.3.15至2022.3.31 |  |
| 8 | 撰写毕业论文 | 2022.4.1至2022.4.15 |  |
| 指导教师意见：  填写说明：查阅资料是否全面，提出的研究方案和计划进度是否可行，还有什么需要注意和改进的方面，是否同意按学生提出的计划进行等。（**填写后请删除该说明）** | | | |

指导教师（审核签名）： 审核日期： 年 月 日