



(12) 发明专利

(10) 授权公告号 CN 108564286 B

(45) 授权公告日 2021.01.22

(21) 申请号 201810352544.2  
 (22) 申请日 2018.04.19  
 (65) 同一申请的已公布的文献号  
 申请公布号 CN 108564286 A  
 (43) 申请公布日 2018.09.21  
 (73) 专利权人 天合泽泰(厦门)征信服务有限公司  
 地址 361000 福建省厦门市火炬高新区软件园创新大厦A区10层1002室  
 (72) 发明人 孙晓俊 肖炜 洪倩雯 林佳佳 郭晓凤 李琦薇  
 (74) 专利代理机构 厦门智慧呈睿知识产权代理事务所(普通合伙) 35222  
 代理人 郭福利 魏思凡

(51) Int.Cl.  
 G06Q 10/06 (2012.01)  
 G06Q 40/02 (2012.01)  
 (56) 对比文件  
 CN 106960358 A, 2017.07.18  
 CN 106408413 A, 2017.02.15  
 CN 105259816 A, 2016.01.20  
 CN 106971338 A, 2017.07.21  
 CN 104299105 A, 2015.01.21  
 US 2015/0170270 A1, 2015.06.18  
 审查员 黄超

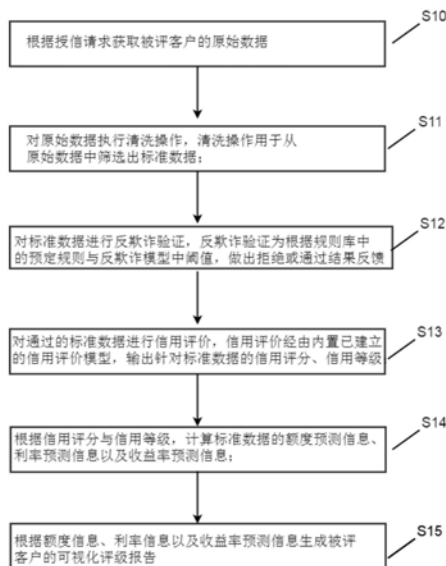
权利要求书4页 说明书13页 附图4页

(54) 发明名称

一种基于大数据征信的人工智能金融风控授信评定方法和系统

(57) 摘要

本发明实施例公开了一种基于大数据征信的人工智能金融风控授信系统和评定方法,其特征在于,该方法由处理器执行,所述方法包括:根据授信请求获取被评客户的原始数据;对所述原始数据执行清洗操作,所述清洗操作用于从所述原始数据中筛选出标准数据;对所述标准数据进行反欺诈验证,所述反欺诈验证为根据规则库中的预定规则与反欺诈模型中阈值,做出拒绝或通过结果反馈;对通过的所述标准数据进行信用评价,所述信用评价经信用评价模型,输出针对所述标准数据的信用评分、信用等级;根据所述信用评分与信用等级,计算所述标准数据的额度测算信息、利率建议信息以及收益率预测信息;根据所述额度测算信息、利率建议信息以及收益率预测信息生成所述被评客户的可视化评级报告。



1. 一种基于大数据征信的人工智能金融风控授信评定方法,其特征在于,该方法由处理器执行,所述方法包括:

根据授信请求获取被评客户的原始数据;

对所述原始数据执行清洗操作,所述清洗操作用于从所述原始数据中筛选出标准数据;

对所述标准数据进行反欺诈验证,所述反欺诈验证为根据规则库中的预定规则与反欺诈模型中的阈值,做出拒绝或通过结果反馈;

对通过的所述标准数据进行信用评价,所述信用评价经由信用评价模型,输出针对所述标准数据的信用评分、信用等级;

根据所述信用评分与信用等级,计算所述标准数据的额度预测信息、利率预测信息以及收益率预测信息;

根据所述额度测算信息、利率测算信息以及收益率预测信息生成所述被评客户的可视化评级报告;

信用评价模型还能够实现自我迭代更新,其包括在信用模型中监控新增数据是否达到定值;通过该监控,一旦新增数据达到了定值,则对该信用评价模型进行重新训练;用重新训练后的信用评价模型更新现有的信用评价模型;

信用评价模块包含模型构建与定时迭代更新2部分,其中模型构建部分包含信用评分卡模型、信用等级模型和风险分类模型,作为客户初始信用分确定、客户群体划分、风险预测的参照,对客户形成一个较为完整的信用评价;模型更新指新增数据累积达到定值,则对模型重新训练,迭代更新;

信用评分卡模型包含评分特征项和对应权重部分;评分特征项中,个人数据维度由基础信息、职业信息、资产及流水信息、信用和借贷信息、模型外调整信息5部分组成;企业数据维度包含7部分,其中法人信息由5部分,企业信息由非财务评价与财务评价组成,权重的设定,采用自适应AHP层次分析模型输出,总信用分由各评分项与权重的乘积所得子分数项相加得来;

信用等级模型是根据借贷历史数据集,包含贷前数据与贷后数据集,数据集中包含信用分特征,做聚类分析;从算法库中选取不少于3个聚类算法模型对数据建模,对建立的模型用留出法进行模型泛化性能检验,再对比不同模型间性能,确定最终使用的模型,并输出该模型返回结果,以每一聚类簇中信用分均值作为信用等级划分边界;

模块中风险分类模型是指对待评价对象进行还款情况预测,根据借贷历史数据集,将通过审批并还款的客户分为两类,逾期客户和正常还款客户分别标记为0、1,代表高风险与低风险客户;从算法库中选取不少于3个分类算法模型对数据建模,对建立的模型进行模型性能检验,对比不同模型间精确率、召回率,确定最终使用的模型;将训练好的模型保存在系统中;同时对于风险分类标记为1的高风险客户,回调其信用等级,下调一级;

对数据进行分类预测建模,调用封装好的随机森林算法包,从原始训练集中使用Bootstrapping方法,即随机有放回采样选出m个样本,共进行n次采样,生成n个训练集;

如果每个样本的特征维度为M,指定一个常数 $m \ll M$ ,随机地从M个特征中选取m个特征子集,每次树进行分裂时,从这m个特征中根据信息增益或信息增益比或基尼指数选择最好的特征进行分裂;

每棵树都尽最大程度的生长,在决策树的分裂过程中不需要剪枝;

将生成的多棵决策树组成随机森林;由于此次模型是分类问题,按多棵树分类器投票决定最终分类结果,此处为‘相对多数投票法’;

预测得票最多的标记,若同时有数个标记同票最高,随机选择一个;模型训练完毕并保存,将多个算法训练的模型进行性能对比,对比精确率、召回率、ROC以及AUC指标;确定一个最优模型,保存模型文件;

所述自适应AHP层次分析模型,由专家打分系统,形成矩阵传输至自适应AHP模型,模型对AHP矩阵进行第一次检查,对没有通过一致性检验的矩阵进行调整,计算偏差矩阵,对其中影响最大的矩阵元素进行微调,再返回一个新的判断矩阵,验证其是否满足一致性检验,循环以上程序直至通过,并最终输出所有评分项对应权重。

2. 根据权利要求1所述的授信评定方法,其特征在于,所述清洗操作包括:

剔除所述原始数据中的异常值、重复值、无效值以及缺失值以获得过滤数据;

对所述过滤数据进行去噪、修复和降维处理,以获得所述标准数据。

3. 根据权利要求1所述的一种基于大数据征信的人工智能金融风控授信评定方法,其特征在于,所述反欺诈验证包括:

判断所述标准数据是否与所述规则库中的预定规则一致;

若不一致,则对所述标准数据运行所述反欺诈模型以输出反欺诈评分,判断所述反欺诈评分是否低于所述阈值;

若所述反欺诈评分低于所述阈值,则将所述标准数据标记为通过。

4. 根据权利要求3所述的一种基于大数据征信的人工智能金融风控授信评定方法,其特征在于,所述标准数据与所述规则库中的所述预定规则一致,则将所述标准数据标记为拒绝,并根据该拒绝结果生成针对所述被评客户的可视化评级报告。

5. 根据权利要求3所述的一种基于大数据征信的人工智能金融风控授信评定方法,其特征在于,在所述判断判断所述反欺诈评分是否低于所述阈值的步骤中,还包括:

若所述反欺诈评分高于所述阈值,则将所述标准数据标记为拒绝,并根据该拒绝结果生成针对所述被评客户的可视化评级报告。

6. 一种基于大数据征信的人工智能金融风控授信系统,其特征在于,所述授信系统包括用于分析客户信用、做信用评估的授信子系统;所述授信子系统包括:

数据获取模块,用于根据授信请求获取被评客户的原始数据;

数据处理模块,用于对所述原始数据执行清洗操作,所述清洗操作用于从所述原始数据中筛选出标准数据;

反欺诈模块,用于对所述标准数据进行反欺诈验证,所述反欺诈验证为根据规则库中的预定规则与反欺诈模型中阈值,做出拒绝或通过结果反馈;

信用评价模块,对通过的所述标准数据进行信用评价,所述信用评价经由内置已建立的信用评价模型,输出针对所述标准数据的信用评分、信用等级;

计算模块,所述计算模块包括:

额度测算模块,用于根据所述标准数据计算所述被评客户所对应的额度信息;

利率建议模块,用于根据所述标准数据计算所述被评客户所对应的利率信息;

收益率预测模块,用于根据所述标准数据计算所述被评客户所能带来的收益率信息;

可视化报告输出模块,用于根据所述额度信息、利率信息以及收益率预测信息生成所述被评客户的可视化评级报告;

信用评价模型还能够实现自我迭代更新,其包括在信用模型中监控新增数据是否达到定值;通过该监控,一旦新增数据达到了定值,则对该信用评价模型进行重新训练;用重新训练后的信用评价模型更新现有的信用评价模型;

信用评价模块包含模型构建与定时迭代更新2部分,其中模型构建部分包含信用评分卡模型、信用等级模型和风险分类模型,作为客户初始信用分确定、客户群体划分、风险预测的参照,对客户形成一个较为完整的信用评价;模型更新指新增数据累积达到定值,则对模型重新训练,迭代更新;

信用评分卡模型包含评分特征项和对应权重部分;评分特征项中,个人数据维度由基础信息、职业信息、资产及流水信息、信用和借贷信息、模型外调整信息等5部分组成;企业数据维度包含7部分,其中法人信息由5部分,企业信息由非财务评价与财务评价组成,权重的设定,采用自适应AHP层次分析模型输出,总信用分由各评分项与权重的乘积所得子分数项相加得来;

信用等级模型是根据借贷历史数据集,包含贷前数据与贷后数据集,数据集中包含信用分特征,做聚类分析;从算法库中选取不少于3个聚类算法模型对数据建模,对建立的模型用留出法进行模型泛化性能检验,再对比不同模型间性能,确定最终使用的模型,并输出该模型返回结果,以每一聚类簇中信用分均值作为信用等级划分边界;

模块中风险分类模型是指对待评价对象进行还款情况预测,根据借贷历史数据集,将通过审批并还款的客户分为两类,逾期客户和正常还款客户分别标记为0、1,代表高风险与低风险客户;从算法库中选取不少于3个分类算法模型对数据建模,对建立的模型进行模型性能检验,对比不同模型间精确率、召回率,确定最终使用的模型;将训练好的模型保存在系统中;同时对于风险分类标记为1的高风险客户,回调其信用等级,下调一级;

对数据进行分类预测建模,调用封装好的随机森林算法包,从原始训练集中使用Bootstrapping方法,即随机有放回采样选出 $m$ 个样本,共进行 $n$ 次采样,生成 $n$ 个训练集;

如果每个样本的特征维度为 $M$ ,指定一个常数 $m \ll M$ ,随机地从 $M$ 个特征中选取 $m$ 个特征子集,每次树进行分裂时,从这 $m$ 个特征中根据信息增益或信息增益比或基尼指数选择最好的特征进行分裂;

每棵树都尽最大程度的生长,在决策树的分裂过程中不需要剪枝;

将生成的多棵决策树组成随机森林;由于此次模型是分类问题,按多棵树分类器投票决定最终分类结果,此处为‘相对多数投票法’;

预测为得票最多的标记,若同时有数个标记同票最高,随机选择一个;模型训练完毕并保存,将多个算法训练的模型进行性能对比,对比精确率、召回率、ROC以及AUC指标;确定一个最优模型,保存模型文件;

所述自适应AHP层次分析模型,由专家打分系统,形成矩阵传输至自适应AHP模型,模型对AHP矩阵进行第一次检查,对没有通过一致性检验的矩阵进行调整,计算偏差矩阵,对其中影响最大的矩阵元素进行微调,再返回一个新的判断矩阵,验证其是否满足一致性检验,循环以上程序直至通过,并最终输出所有评分项对应权重。

7. 根据权利要求6所述一种基于大数据征信的人工智能金融风控授信系统,其特征在

于,所述数据处理模块还包括:

数据过滤模块,用于剔除所述原始数据中的异常值、重复值、无效值以及缺失值以获得过滤数据;

标准数据获取模块,用于对所述过滤数据进行去噪、修复和降维处理,以获得所述标准数据。

8. 根据权利要求6所述一种基于大数据征信的人工智能金融风控授信系统,其特征在于,所述反欺诈模块还包括:

校验模块,用于判断所述标准数据是否与所述规则库中的所述预定规则一致;

识别模块,用于判断所述标准数据是否低于所述反欺诈模型的阈值;

标记模块,用于在所述标准数据与所述预定规则一致时,将所述标准数据标记为拒绝;在所述标准数据与所述预定规则不一致,且所述标准数据高于所述阈值时,将所述标准数据标记为拒绝;在所述标准数据与所述预定规则不一致,且所述标准数据低于所述阈值时,将所述标准数据标记为通过。

9. 一种终端设备,其特征在于,包括处理器、存储器以及存储在所述存储器中且被配置由所述处理执行的计算机程序,所述处理器执行所述计算机程序时实现如权利要求1至5任一项所述的一种基于大数据征信的人工智能金融风控授信评定方法。

## 一种基于大数据征信的人工智能金融风控授信评定方法和系统

### 技术领域

[0001] 本发明涉及一种基于大数据征信的人工智能金融风控授信系统,尤其是一种能够机器学习、深度学习模型的混合构建在智能授信的应用系统。

### 背景技术

[0002] 随着互联网与经济的不断发展,互联网金融、消费金融的崛起也促进了我国小微信贷产业的快速发展,线上借贷的需求增大也带来更多信用风险,目前,国内传统金融风控大部分还都由传统人工进行,从资料提交到审核,需要大量时间和人力物力,为解决人工审核的诸多问题,信用评分技术也应运而生,通过对贷款人各方数据信息的整合,预测其信用分析值,以帮助授信审批者做出决策。

[0003] 随着大数据的普及,数据积累的数量和质量都得到了飞跃,大数据技术推动数据统计模型不断完善,深层次挖掘征信数据,尤其是人工智能模型可以更好的预测未来,更加科学的反映用户的信用状况。

[0004] 在现有技术中,传统信用评分技术存在准确率较低和效率较差的技术问题。

[0005] 基于机器学习、深度学习的评分技术能在一定程度提高准确率,但结果单一,多只显示评价对象的信用评分,对审批者的指导建议有限。

### 发明内容

[0006] 本发明目的在于克服现有技术的不足,提供一种基于大数据征信的金融风控授信系统,系统从数据的获取到授信建议的输出形成一套完整的智能金融风控系统流程,为有授信需求的平台提供完整数据采集、量化、分析服务,能够有效、有针对性地对被评个人、企业进行科学的信用分析,快速智能输出授信建议,提高工作效率、降低违约风险,推进信贷市场健康发展。

[0007] 为实现上述目的:

[0008] 第一方面,本发明实施例提供一种基于大数据征信的人工智能金融风控授信评定方法,该方法由处理器执行,方法包括:

[0009] 根据授信请求获取被评客户的原始数据;

[0010] 对原始数据执行清洗操作,清洗操作用于从所原始数据中筛选出标准数据;

[0011] 对标准数据进行反欺诈验证,反欺诈验证为根据规则库中的预定规则与反欺诈模型中阈值,做出拒绝或通过结果反馈;

[0012] 对通过的标准数据进行信用评价,信用评价经信用评价模型,输出针对标准数据的信用评分、信用等级;

[0013] 根据信用评分与信用等级,计算标准数据的额度测算信息、利率建议信息以及收益率预测信息;

[0014] 根据额度测算信息、利率建议信息以及收益率预测信息生成被评客户的可视化评

级报告。

[0015] 进一步地,清洗操作包括:

[0016] 剔除原始数据中的异常值、重复值、无效值、缺失值以获得过滤数据;

[0017] 对过滤数据进行去噪、修复和降维处理,以获得标准数据。

[0018] 进一步地,反欺诈验证包括:

[0019] 判断标准数据是否与规则库中的预定规则一致;

[0020] 若不一致,则对标准数据运行反欺诈模型以输出反欺诈评分,判断反欺诈评分是否低于所述阈值;

[0021] 若反欺诈评分低于所述阈值,则将标准数据标记为通过。

[0022] 更进一步地,标准数据与规则库中的预定规则一致,则将标准数据标记为拒绝,并根据该拒绝结果生成针对被评客户的可视化评级报告。

[0023] 更进一步地,在判断所述标准数据是否低于反欺诈模型的阈值的步骤中,还包括:

[0024] 若标准数据高于所述阈值,则将标准数据标记为拒绝,并根据该拒绝结果生成针对被评客户的可视化评级报告。

[0025] 第二方面,本发明实施例提供一种基于大数据征信的人工智能金融风控授信系统,授信系统包括用于分析客户信用、做信用评估的授信子系统;

[0026] 授信子系统包括:

[0027] 数据获取模块,用于根据授信请求获取被评客户的原始数据;

[0028] 数据处理模块,用于对原始数据执行清洗操作,清洗操作用于从原始数据中筛选出标准数据;

[0029] 反欺诈模块,用于对标准数据进行反欺诈验证,反欺诈验证为根据规则库中的预定规则与反欺诈模型中阈值,做出拒绝或通过结果反馈;

[0030] 信用评价模块,对通过的标准数据进行信用评价,信用评价经由内置已建立的信用评价模型,输出针对标准数据的信用评分、信用等级;

[0031] 计算模块,所述计算模块包括:

[0032] 额度测算模块,用于根据标准数据计算被评客户所对应的额度信息;

[0033] 利率建议模块,用于根据标准数据计算被评客户所对应的利率信息;

[0034] 收益率预测模块,用于根据标准数据计算被评客户所能带来的收益率信息;

[0035] 可视化报告输出模块,用于根据额度测算信息、利率建议信息以及收益率预测信息生成被评客户的可视化评级报告。

[0036] 进一步地,授信子系统还包括:

[0037] 数据过滤模块,用于剔除原始数据中的异常值、重复值、无效值、缺失值以获得过滤数据;

[0038] 标准数据获取模块,用于对过滤数据进行去噪、修复和降维处理,以获得标准数据。

[0039] 进一步地,所述授信子系统还包括:

[0040] 校验模块,用于判断标准数据是否与规则库中的预定规则一致;

[0041] 识别模块,用于判断标准数据是否低于反欺诈模型的阈值;

[0042] 标记模块,用于在标准数据与预定规则一致时,将标准数据标记为拒绝;在标准数

据与预定规则不一致,且标准数据高于所述阈值时,将标准数据标记为拒绝;在标准数据与  
所述预定规则不一致,且标准数据低于阈值时,将标准数据标记为通过。

[0043] 第三方面,本发明实施例提供一种终端设备,包括处理器、存储器以及存储在所述  
存储器中且被配置由所述处理执行的计算机程序,所述处理器执行所述计算机程序时实现  
上述任一项所述的一种基于大数据征信的人工智能金融风控授信评定方法。

[0044] 实施本发明实施例,具有如下有益效果:

[0045] 本发明提供了一种基于大数据征信的人工智能金融风控授信评定方法和系统,将  
评价对象的相关数据流程化,通过多种渠道获取被评级对象相关性数据,并不局限于被评  
级对象主动提供的信息;数据清洗采用多种方式进行,避免由于一两种单一清洗方法造成  
的局限;数据建模采用多种模型结合,选取最适合被评级对象数据特征的模型,增加评级效  
果的科学性;将被评级对象的分析结果进行输出,结合其他基本信息完成报表制作,最后呈  
现一份对被评级对象的科学合理的评级报告。

[0046] 本发明为有授信需求的平台提供从平台整体运作监测到专业工作人员审批工作  
流程化的完整系统,按照系统提供的数据处理与应用方案对被评价个人/企业的相关信息  
进行专业化处理,最终通过结果输出得到科学合理的授信建议报告。

[0047] 本发明综合运用人工智能技术和信息技术,同时实现信用分评估/授信额度建议/  
反欺诈提示,结合定性指标/定量指标/专家经验,采用静态分析和动态分析的全面评估,集  
数据采集、信息筛选、评分审批于一体。

[0048] 本发明所应用的相关人工智能模型具有能够考虑参数的时间动态属性、具备拓展  
参数维度的能力、具备定性和定量的数据处理能力、能够纳入客户的履约/违约信息反馈、  
实现自适应,达到合理优化信用评级的目标等优势。

[0049] 本发明所设计的授信系统,不止于考虑贷前风险,更关注整体收益,比传统的众多  
信用评价模型更具优势。

## 附图说明

[0050] 为了更清楚地说明本发明的技术方案,下面将对实施方式中所需要使用的附图作  
简单地介绍,显而易见地,下面描述中的附图仅仅是本发明的一些实施方式,对于本领域普  
通技术人员来讲,在不付出创造性劳动的前提下,还可以根据这些附图获得其他的附图。

[0051] 图1是本发明的一种基于大数据征信的人工智能金融风控授信评定方法的第一个  
实施例的流程示意图;

[0052] 图2是本发明的一种数据清洗操作的流程示意图;

[0053] 图3是本发明提供的一种标准数据判断流程示意图;

[0054] 图4本发明的一种基于大数据征信的人工智能金融风控授信系统的第二实施例的  
流程示意图;

[0055] 图5是本发明提供的一种数据处理模块的结构示意图;

[0056] 图6是本发明提供的一种反欺诈模块的结构示意图;

[0057] 图7是本发明提供的一种模型构建示意图;

## 具体实施方式

[0058] 下面将结合本发明实施例中的附图,对本发明实施例中的技术方案进行清楚、完整地描述,显然,所描述的实施例仅仅是本发明一部分实施例,而不是全部的实施例。基于本发明中的实施例,本领域普通技术人员在没有作出创造性劳动前提下所获得的所有其他实施例,都属于本发明保护的范围。

[0059] 本发明第一实施例:

[0060] 参照图1,图2和图3,图1是本发明的一种基于大数据征信的人工智能金融风控授信评定方法的一个实施例的流程示意图,图2是本发明的一种数据清洗操作的流程示意图,图3是本发明提供的一种标准数据判断流程示意图。将评价对象的相关数据流程化,通过多种渠道获取被评级对象相关性数据,并不局限于被评级对象主动提供的信息;数据清洗采用多种方式进行,避免由于一两种单一清洗方法造成的局限;数据建模采用多种模型结合,选取最适合被评级对象数据特征的模型,增加评级效果的科学性;将被评级对象的分析结果进行输出,结合其他基本信息完成报表制作,最后呈现一份对被评级对象的科学合理的评级报告,该人工智能金融风控授信系统评定方法具体包括以下步骤:

[0061] S10,根据授信请求获取被评客户的原始数据。

[0062] 在本发明实施例中,根据有授信需求的平台获取被评客户,资源平台整体运作监测被评客户的原始数据信息,其来源方面可以包括:消费分期平台提供的该被评客户的个人基本信息,包括年龄、性别、手机号等信息,此部分信息为待评客户主动提供信息;第三方企业信用评级机构通过自身的专业技术进行网络数据爬虫,获取该公司的一些其他相关信息,如法院诉讼信息、失信惩戒信息等信息;根据被评客户签订授权同意书,从银行获取被评客户个人征信报告;将以上通过各种渠道获取的信息作为该被评客户的原始数据信息。

[0063] S11,对原始数据执行清洗操作,清洗操作用于从所原始数据中筛选出标准数据。

[0064] 在本发明实施例中,对根据授信请求而获取的被评客户的原始数据进行数据清洗,整理出该原始数据中的“脏”数据,以得到可直接分析的标准数据存储到数据存储数据库中。数据清洗主要针对网络数据爬虫获得的信息以及网站后台导出的数据信息,因为网络爬虫得到的信息格式多样。基于此情况,例如,如图2所示,数据清洗操作可以包括:

[0065] B10,剔除原始数据中的异常值、重复值、无效值、缺失值以获得过滤数据。

[0066] 异常值指样本中的个别值,其数值明显偏离它(或他们)所属样本的其余观测值,也称异常数据,离群值。缺失值是指现有数据集中某个或某些属性的值是不完全的。重复值是指数据行中完全一样的数据。

[0067] B11,对过滤数据进行去噪、修复和降维处理,以获得标准数据。

[0068] 降维一般使用主成分分析法:主成分分析法是一种数学变换的方法,它把给定的一组相关变量通过线性变换转成另一组不相关的变量,这些新的变量按照方差依次递减的顺序排列。在数学变换中保持变量的总方差不变,使第一变量具有最大的方差,称为第一主成分,第二变量的方差次大,并且和第一变量不相关,称为第二主成分。依次类推,I个变量就有I个主成分。进行主成分分析后,还可以根据需要进一步利用K-L变换(霍特林变换)对原数据进行投影变换,达到降维的目的。

[0069] S12,对标准数据进行反欺诈验证,反欺诈验证为根据规则库中的预定规则与反欺诈模型中阈值,做出拒绝或通过结果反馈。

[0070] 经过数据处理之后的获得可应用的标准数据,命中规则库任一预定规则,则标记为拒绝,并根据拒绝信息反馈直接输出可视化报告,拒绝该被评用户的授信。未命中规则库中的任一预定规则,则继续运行反欺诈模型进行反欺诈验证,以输出反欺诈评分。在反欺诈验证中,根据事先设定的相应阈值 $\theta$ 进行判定。若经反欺诈模型处理之后的标准数据高于阈值 $\theta$ ,则将其标记为拒绝,并根据拒绝信息反馈直接输出可视化报告,拒绝该被评用户的授信。若经反欺诈模型处理之后的标准数据低于阈值 $\theta$ ,则将其标记为通过。例如,如图3所示,反欺诈验证方法可以包括:

[0071] A11,判断标准数据是否与规则库中的预定规则一致;

[0072] A12,若不一致,则对标准数据运行反欺诈模型以输出反欺诈评分,判断所述反欺诈评分是否低于所述阈值;

[0073] A13,若反欺诈评分低于所述阈值,则将标准数据标记为通过;

[0074] A20,标准数据与规则库中的预定规则一致,则将标准数据标记为拒绝;

[0075] A30,若反欺诈评分高于所述阈值,则将标准数据标记为拒绝。

[0076] 更具体地,反欺诈模型可以包括评分特征项和对应权重部分,由特征维度包含个人违规信息等多种验证项,分别赋予不同权重。权重的设定,采用自适应AHP层次分析模型输出。总反欺诈评分由各评分项与权重的乘积所得子分数项相加得来。例如设反欺诈模型共包含特征维度基本信息违规核查,不良信息扫描,关联人信息扫描,客户行为检测四项,分数分别为25分、30分、30分、15分,通过AHP得出其权重分别为0.25、0.25、0.3、0.2,则反欺诈总分为: $25*0.25+30*0.25+30*0.3+15*0.2=25.75$ 。

[0077] S13,对通过的标准数据进行信用评价,信用评价经信用评价模型,输出针对标准数据的信用评分、信用等级。

[0078] 信用评价是授信建议的核心部分,其主要目的是反馈申贷客户的风险情况。其中,信用评价模型可以包括信用评分卡模型、信用等级模型、风险分类模型等。一般的过程是:在对数据进行清洗和降维/拓展之后,所有的数据转换为表格形式,将这些表格数据灌入机器学习模型,用 $X$ 表示数据, $y$ 表示标签。

[0079] 如风险分类模型中,客户的年龄性别学历等一系列特征为 $X$ ,是否逾期的结果为 $Y$ ,其中逾期标记为0,正常还款标记为1。

[0080] 把数据分成训练集和测验集两个部分,用训练集进行模型训练,测验集用来检测正确率等指标,选定算法之后,进行算法参数调整,得到最终模型。

[0081] 更具体地说,例如,信用评分卡模型可以包括评分特征项和对应权重部分,在评分特征项中,个人数据维度由基础信息、职业信息、资产信息及流水信息、信用和借贷信息、模型外调整信息等5个部分组成;企业数据维度则可以包括基础信息、职业信息、资产信息及流水信息、信用和借贷信息、模型外调整信息、非财务评价以及财务评价等7个部分组成。权重的设定,采用自适应AHP层次分析模型输出。总信用分由各评分项与权重的乘积所得子分数项相加得来。

[0082] 个人信用分示例,如表1:

[0083]	工作年限	①[5,x)年, 得 5 分
		②[3-5)年, 得 3 分
		③[1-3)年, 得 2 分
		④[0.5-1) 年, 得 1 分
		⑤[0,6)个月, 得 0 分

[0084] 表1

[0085] 工作年限权重为0.05,某客户已参加工作7年,那么该项得分为0.25.

[0086] 个人信用分总分由各个子项的分数和权重相乘累加得到,企业信用分以此类推。

[0087] 信用等级模型可以是根据借贷历史数据集,包含贷前数据与贷后数据集,数据集中包含信用分特征,做聚类分析。从算法库中选取不少于3个聚类算法模型(包含k-means算法、DBSCAN、GMM、SOM等)对数据建模,对建立的模型用留出法进行模型泛化性能检验,再对比不同模型间性能,确定最终使用的模型,并输出该模型返回结果,以每一聚类簇中信用分均值作为信用等级划分边界。风险分类模型可以是指对待评价对象进行还款情况预测,根据借贷历史数据集,将通过审批并还款的客户分为两类,逾期客户和正常还款客户分别标记为0、1,代表高风险与低风险客户。从算法库中选取不少于3个分类算法模型(包含BP神经网络、随机森林、SVM、xgboost等)对数据建模,对建立的模型进行模型性能检验,对比不同模型间精确率、召回率等,确定最终使用的模型。将训练好的模型保存在系统中。同时对于风险分类标记为1的高风险客户,回调其信用等级,下调一级。

[0088] 例如:

[0089] 首先选择数据,包含客户贷前数据:年龄/性别/收入/住房/信用分等维度,贷后数据:是否逾期,逾期时间等;

[0090] 对数据进行清洗和特征选择(降维);

[0091] 对数据进行聚类分析,调用算法封装函数包,此处示例kmeans算法:

[0092] 1随机选取k个中心点

[0093] 2遍历所有数据,将每个数据划分到最近的中心点中

[0094] 3计算每个聚类的平均值,并作为新的中心点

[0095] 4重复直到这k个中心点不再变化,即已收敛,或执行了足够多的迭代则结束

[0096] K值需要事先设定,且对簇中心初始化较敏感。使用随机分区方法,随机为每个观测值分配一个簇,随后进行更新,簇的随机分配点的质心就是计算后得到的初始平均值。设此处信用等级划分为5个等级A~F,则K=4,查看收敛结束之后的4个簇中信用分均值,其中信用分范围为0~1000,假若聚类结果中信用分均值分别为250,400,550,680,850则F级为(0,250分],E为(250,350],A为(850,1000],以此类推。

[0097] 使用多个聚类算法对以上进行重复计算,进行模型效果评估,评估指标选择适用于实际类别信息未知的轮廓系数:

[0098] 对于单个样本,设a是与它同类别中其他样本的平均距离,b是与它距离最近不同类别中样本的平均距离,其轮廓系数为:  $s = \frac{b-a}{\max(a,b)}$ , 整个数据样本集的轮廓系数是单个样本的均值。轮廓系数的取值范围是[-1,1],同类别样本距离越相近不同类别样本距离越远,分数越高。

[0099] 选择其中效果最佳的模型,以该模型的结果作为最终信用等级结果。

[0100] 此外,根据初始信用分确定、客户群体划分、风险预测的参照,可以对客户形成一个较为完整的信用评价。待评价客户数据可以经由内置已训练好的信用评价模型,输出信用评分,信用等级的结果数据。

[0101] 其中,在本实施例中,信用评价模型还能够实现自我迭代更新,其包括在信用模型中监控新增数据是否达到定值。通过该监控,一旦新增数据达到了定值,则对该信用评价模型进行重新训练。用重新训练后的信用评价模型更新现有的信用评价模型。

[0102] 同时,上述提及的自适应AHP层次分析,是本发明中自主改进的AHP算法,由专家对评分特征项打分,形成矩阵传输至自适应AHP模型,模型对AHP矩阵进行第一次检查,对没有通过一致性检验的矩阵进行调整,计算偏差矩阵,对其中影响最大的矩阵元素进行微调,再返回一个新的判断矩阵,验证其是否满足一致性检验,循环以上程序直至通过,并最终输出所有评分项对应权重。通过对各指标之间进行两两对比之后,然后按9分位比率排定各评价指标的相对优劣顺序,依次构造出评价指标的判断矩阵A。如表2,

[0103]

职业信息	职业	年限	职务
职业	1	3	1/5
年限	1/3	1	1/7
职务	5	7	1

[0104] 表2

[0105] 有9种取值,分别为1/9、1/7、1/5、1/3、1/1、3/1、5/1、7/1、9/1,分别表示i要素对于j要素的重要程度由轻到重,其中行作为i要素,列作为j要素,如表3:

[0106]

标度	含义
----	----

[0107]	1	表示两个因素相比，具有相同重要性
	3	表示两个因素相比，前者比后者稍重要
	5	表示两个因素相比，前者比后者明显重要
	7	表示两个因素相比，前者比后者强烈重要
	9	表示两个因素相比，前者比后者极端重要
	2,4,6,8	表示上述相邻判断的中间值
	倒数	若因素 i 和因素 j 的重要性之比为 $a_{ij}$ ,那么因素 j 与因素 i 重要性之比为 $a_{ji}=1/a_{ij}$

[0108] 表3

[0109] 关于判断矩阵权重计算的方法有两种,即几何平均法(根法)和规范列平均法(和法)。

[0110] (1) 几何平均法(根法)

[0111] 计算矩阵A各个元素的乘积,得到一个n行一列的矩阵B;

[0112] 计算矩阵B中每个元素的n次方根得到矩阵C;

[0113] 对矩阵C进行归一化处理得到矩阵D;

[0114] 该矩阵D即为所求权重向量。

[0115] (2) 规范列平均法(和法)

[0116] 矩阵A每一列归一化得到矩阵B;

[0117] 将矩阵B每一行元素的平均值得到一个一列n行的矩阵C;

[0118] 矩阵C即为所求权重向量。

[0119] 上述提及的算法库内置了机器学习/深度学习的诸多算法,调用所述数据存储数据库中清洗后的数据,根据数据特征状况选取适当的算法进行模型加以调参建立模型。模型验证评价装置提供了模型性能的验证和对比,通过多次验证模型,修正模型可信度。

[0120] S14,根据信用评分与信用等级,计算标准数据的额度预测信息、利率预测信息以及收益率预测信息。

[0121] 在此步骤中,额度预测信息可以依据借贷历史数据集,筛选无违约按期还贷的客户数据,划分训练集与测试集,在原特征中加入信用分与信用等级两个维度,以历史授信额度作为Y值,进行授信额度预测,模型建立及选择方法同‘风险分类’,将训练好的模型保存在系统中。

[0122] 利率建议模块是依据上述信用分与信用等级的输出结果,映射至真实大样本违约率,得出该客户的预测违约率,将预测违约率代入核心利率计算公式,输出该客户建议授信利率。

[0123] 收益率预测模块依据借贷历史数据集,包含全量贷前特征与贷后数据,以最终收益的IRR值作为Y值,以LR等回归算法预测新增客户将为平台带来的IRR,模型建立及选择方法类同‘风险分类’,将训练好的模型保存在系统中。

[0124] S15,根据额度信息、利率信息以及收益率预测信息生成所述被评客户的可视化评级报告。

[0125] 可视化评级报告,用于最后输出对象的总体评价和授信建议并以可视化报告方式展示。

[0126] 本发明第二实施例:

[0127] 在第一个实施例的基础上,参照图4,图5,图6和图7,图4本发明的一种基于大数据征信的人工智能金融风控授信系统的第二实施例的流程示意图;图5是本发明提供的一种数据处理模块的结构示意图;图6是本发明提供的一种反欺诈模块的结构示意图;图7是本发明提供的一种模型构建示意图;本发明人工智能金融风控授信系统由授信子系统20和平台监测子系统30构成,其中授信子系统包含数据处理模块21、数据获取模块22、反欺诈模块23、信用评价模块24、计算模块28、可视化报告输出模块29;平台监测子系统30包含统计指标计算模块31和可视化报告模块32。

[0128] 请参阅图4,该人工智能金融风控授信系统具体包括以下:

[0129] 数据获取模块,21,用于根据授信请求获取被评客户的原始数据;

[0130] 数据处理模块22,用于对原始数据执行清洗操作,清洗操作用于从原始数据中筛选出标准数据;

[0131] 数据处理模块为数据库中的与受评对象相关的各类数据进行清洗处理,包含数据导入,数据特征工程等流程。数据导入是指使用该系统的平台将待评价的个人/企业对象的相关维度数据按规定参数格式录入传输至系统。数据特征工程是指将原始数据所具有的数据缺失、信息冗余、数据无法直接使用、不同量纲、数据稀疏等问题予以处理转换,以便为下一步的数据分析和建模做准备,

[0132] 例如,如图5,数据处理模块还包括:

[0133] 数据过滤模块40,用于剔除原始数据中的异常值、重复值、无效值、缺失值以获得过滤数据;

[0134] 标准数据获取模块50,用于对过滤数据进行去噪、修复和降维处理,以获得标准数据。

[0135] 反欺诈模块23,用于对标准数据进行反欺诈验证,反欺诈验证为根据规则库中的预定规则与反欺诈模型中阈值,做出拒绝或通过结果反馈;

[0136] 反欺诈模块作为防火墙,为平台拦下具有欺诈性质的高风险客户,是第一道重要屏障,反欺诈模型包含评分特征项和对应权重部分,由特征维度包含个人违规信息等多种验证项,分别赋予不同权重,权重的设定,采用自适应AHP层次分析模型输出。总反欺诈分由各评分项与权重的乘积所得子分数项加总得来。

[0137] 参见如图6,反欺诈模块23还包括:

[0138] 校验模块60,用于判断标准数据是否与规则库中的预定规则一致;

[0139] 识别模块61,用于判断标准数据是否低于反欺诈模型的阈值;

[0140] 标记模块62,用于在标准数据与预定规则一致时,将标准数据标记为拒绝;在标准

数据与预定规则不一致,且标准数据高于所述阈值时,将标准数据标记为拒绝;在标准数据与所述预定规则不一致,且标准数据低于阈值时,将标准数据标记为通过。

[0141] 信用评价模块24块,对通过的标准数据进行信用评价,信用评价经由内置已建立的信用评价模型,输出针对标准数据的信用评分、信用等级;

[0142] 信用评价模块包含模型构建与定时迭代更新2部分,其中模型构建部分包含信用评分卡模型91、信用等级模型92、风险分类模型93,作为客户初始信用分确定、客户群体划分、风险预测的参照,对客户形成一个较为完整的信用评价。模型更新指新增数据累积达到定值,则对模型重新训练,迭代更新。

[0143] 信用评分卡模型91包含评分特征项和对应权重部分。评分特征项中,个人数据维度由基础信息、职业信息、资产及流水信息、信用和借贷信息、模型外调整信息等5部分组成;企业数据维度包含7部分,其中法人信息由5部分(与个人数据维度相同),企业信息由非财务评价与财务评价组成,权重的设定,采用自适应AHP层次分析模型输出,总信用分由各评分项与权重的乘积所得子分数项加总得来。

[0144] 信用等级模型92是根据借贷历史数据集,包含贷前数据与贷后数据集,数据集中包含信用分特征,做聚类分析。从算法库中选取不少于3个聚类算法模型(包含k-means算法、DBSCAN、GMM、SOM等)对数据建模,对建立的模型用留出法进行模型泛化性能检验,再对比不同模型间性能,确定最终使用的模型,并输出该模型返回结果,以每一聚类簇中信用分均值作为信用等级划分边界。

[0145] 模块中风险分类模型93是指对待评价对象进行还款情况预测,根据借贷历史数据集,将通过审批并还款的客户分为两类,逾期客户和正常还款客户分别标记为0、1,代表高风险与低风险客户。从算法库中选取不少于3个分类算法模型(包含BP神经网络、随机森林、SVM、xgboost等)对数据建模,对建立的模型进行模型性能检验,对比不同模型间精确率、召回率等,确定最终使用的模型。将训练好的模型保存在系统中。同时对于风险分类标记为1的高风险客户,回调其信用等级,下调一级。例如:

[0146] 首先选择数据,包含客户贷前数据:年龄/性别/收入/住房/职业等众多维度,贷后数据:是否逾期,逾期时间等;将通过审批并还款的客户分为两类,逾期客户和正常还款客户分别标记为0、1,代表高风险与低风险客户。

[0147] 对数据进行清洗和降维;

[0148] 对数据进行分类预测建模,调用封装好的算法包,此处示例随机森林算法:

[0149] 1.从原始训练集中使用Bootstrapping方法,即随机有放回采样选出m个样本,共进行n次采样,生成n个训练集;

[0150] 2.如果每个样本的特征维度为M,指定一个常数 $m \ll M$ ,随机地从M个特征中选取m个特征子集,每次树进行分裂时,从这m个特征中根据信息增益/信息增益比/基尼指数选择最好的特征进行分裂,此处示例基尼指数:

[0151] 
$$\text{Gini}(D) = 1 - \sum_{k=1}^{|Y|} p_k^2$$
 Gini(D) 越小则数据集D的纯度越高。

[0152] 3.每棵树都尽最大程度的生长,在决策树的分裂过程中不需要剪枝。

[0153] 4.将生成的多棵决策树组成随机森林。由于此次模型是分类问题,按多棵树分类器投票决定最终分类结果,此处示例‘相对多数投票法’:

[0154]  $H(x) = C_{\arg \max} \sum_{j=1}^T h_i^j(x)$  其中  $h_i$  为学习器,  $h_i$  在样本  $x$  上的预测输出表示为  $N$

为向量  $(h_i^1(x); h_i^2(x); \dots; h_i^N(x))$ , 其中  $h_i^j(x)$  是  $h_i$  在类别标记  $c_j$  上的输出。

[0155] 预测为得票最多的标记, 若同时有数个标记同票最高, 随机选择一个。

[0156] 5. 模型训练完毕并保存。

[0157] 将多个算法训练的模型进行性能对比, 对比精确率/召回率/ROC/AUC等指标。确定一个最优模型, 保存模型文件。诸如神经网络等算法是黑盒子模型, 无法输出规则与具体算法过程内核, 训练完毕之后新增客户数据传输进该模型, 将输出一个预测结果, 若输出为0, 则表示该客户有逾期风险, 回调其信用等级, 下降一级。

[0158] 上述提及的自适应AHP层次分析模型, 是本发明中自主改进的AHP算法, 由专家打分系统82, 形成矩阵传输至自适应AHP模型, 模型对AHP矩阵进行第一次检查, 对没有通过一致性检验的矩阵进行调整, 计算偏差矩阵, 对其中影响最大的矩阵元素进行微调, 再返回一个新的判断矩阵, 验证其是否满足一致性检验, 循环以上程序直至通过, 并最终输出所有评分项对应权重。

[0159] 计算模块28, 所述计算模块包括:

[0160] 额度测算模块26, 用于根据标准数据计算被评客户所对应的额度信息;

[0161] 利率建议模块27, 用于根据标准数据计算被评客户所对应的利率信息;

[0162] 收益率预测模块28, 用于根据标准数据计算被评客户所能带来的收益率信息;

[0163] 可视化报告输出模块29, 用于根据额度信息、利率信息以及收益率预测信息生成被评客户的可视化评级报告。

[0164] 统计指标计算模块31, 平台上授信客户总数, 每月/日申请客户数量相关统计指标, 客户信用分、风险指标, 平台违约率、收益率变动相关统计指标等。由内置统计算法依次计算出上述指标。

[0165] 可视化报告模块32, 提供了报告生成装置, 内置评级报告底稿模板, 在数据流转过所有模型之后, 将各个模型输出结果, 以可视化报表的方式展示在前端, 自动生成评级报告, 使授信审批人员能够更直观地接受并据以做出授信决策。

[0166] 本发明第三实施例提供一种基于大数据征信的人工智能金融风控授信系统, 包括处理器、存储器以及存储在所述存储器中且被配置由所述处理执行的计算机程序。所述处理器执行所述计算机程序时实现上述任一项所述的一种基于大数据征信的人工智能金融风控授信评定方法的实施例中的步骤, 例如图1所示的步骤S10。或者, 所述处理器执行所述计算机程序时实现上述系统实例中的功能, 例如图4所示的数据获取模块21。

[0167] 示例性地, 所述计算机程序可以被分割成一个或多个模块, 所述一个或者多个模块被存储在所述存储器中, 并由所述处理器执行, 以完成本发明。所述一个或多个模块可以是能够完成特定功能的一系列计算机程序指令段, 该指令段用于描述所述计算机程序在所述实现一种基于大数据征信的人工智能金融风控授信评定方法的执行过程。

[0168] 所称处理器可以是中央处理单元 (Central Processing Unit, CPU), 还可以是其他通用处理器、数字信号处理器 (Digital Signal Processor, DSP)、专用集成电路 (Application Specific Integrated Circuit, ASIC)、现成可编程门阵列 (Field-

Programmable Gate Array, FPGA) 或者其他可编程逻辑器件、分立门或者晶体管逻辑器件、分立硬件组件等。

[0169] 所述存储器可用于存储所述计算机程序和/或模块,所述处理器通过运行或执行存储在所述存储器内的计算机程序和/或模块,以及调用存储在存储器内的数据,实现一种基于大数据征信的人工智能金融风控授信评系统的各种功能。所述存储器可主要包括存储程序区和存储数据区,其中,存储程序区可存储操作系统、至少一个功能所需的应用程序(比如声音播放功能、文字转换功能等)等;存储数据区可存储根据手机的使用所创建的数据(比如音频数据、文字消息数据等)等。此外,存储器可以包括高速随机存取存储器,还可以包括非易失性存储器,例如硬盘、内存、插接式硬盘,智能存储卡(Smart Media Card, SMC),安全数字(Secure Digital, SD)卡,闪存卡(Flash Card)、至少一个磁盘存储器件、闪存器件、或其他易失性固态存储器件。

[0170] 本系统包含两个数据库。数据存储数据库,是服务器群进行存储调用信息的载体,存储数据包括导入的原始数据、清洗后的整理数据、模型数据、报表数据、评级结果数据等内容;CRM数据库主要存储受评客户评级结果展示信息,为客户关系管理的统一存储数据库。

[0171] 其中,所述实现一种基于大数据征信的人工智能金融风控授信系统的模块如果以软件功能单元的形式实现并作为独立的产品销售或使用,可以存储在一个计算机可读存储介质中。基于这样的理解,本发明实现上述实施例方法中的全部或部分流程,也可以通过计算机程序来指令相关的硬件来完成,所述的计算机程序可存储于一个计算机可读存储介质中,该计算机程序在被处理器执行时,可实现上述各个方法实施例的步骤。其中,所述计算机程序包括计算机程序代码,所述计算机程序代码可以为源代码形式、对象代码形式、可执行文件或某些中间形式等。所述计算机可读介质可以包括:能够携带所述计算机程序代码的任何实体或装置、记录介质、U盘、移动硬盘、磁碟、光盘、计算机存储器、只读存储器(ROM, Read-Only Memory)、随机存取存储器(RAM, Random Access Memory)、电载波信号、电信信号以及软件分发介质等。需要说明的是,所述计算机可读介质包含的内容可以根据司法管辖区内立法和专利实践的要求进行适当的增减,例如在某些司法管辖区,根据立法和专利实践,计算机可读介质不包括电载波信号和电信信号。

[0172] 需说明的是,以上所描述的装置实施例仅仅是示意性的,其中所述作为分离部件说明的单元可以是或者也可以不是物理上分开的,作为单元显示的部件可以是或者也可以不是物理单元,即可以位于一个地方,或者也可以分布到多个网络单元上。可以根据实际的需要选择其中的部分或者全部模块来实现本实施例方案的目的。另外,本发明提供的装置实施例附图中,模块之间的连接关系表示它们之间具有通信连接,具体可以实现为一条或多条通信总线或信号线。本领域普通技术人员在不付出创造性劳动的情况下,即可以理解并实施。

[0173] 需说明的是,以上所描述的装置实施例仅仅是示意性的,其中所述作为分离部件说明的单元可以是或者也可以不是物理上分开的,作为单元显示的部件可以是或者也可以不是物理单元,即可以位于一个地方,或者也可以分布到多个网络单元上。可以根据实际的需要选择其中的部分或者全部模块来实现本实施例方案的目的。另外,本发明提供的装置实施例附图中,模块之间的连接关系表示它们之间具有通信连接,具体可以实现为一条或

多条通信总线或信号线。本领域普通技术人员在不付出创造性劳动的情况下,即可以理解并实施。

[0174] 以上所述是本发明的优选实施方式,应当指出,对于本技术领域的普通技术人员来说,在不脱离本发明原理的前提下,还可以做出若干改进和润饰,这些改进和润饰也视为本发明的保护范围。

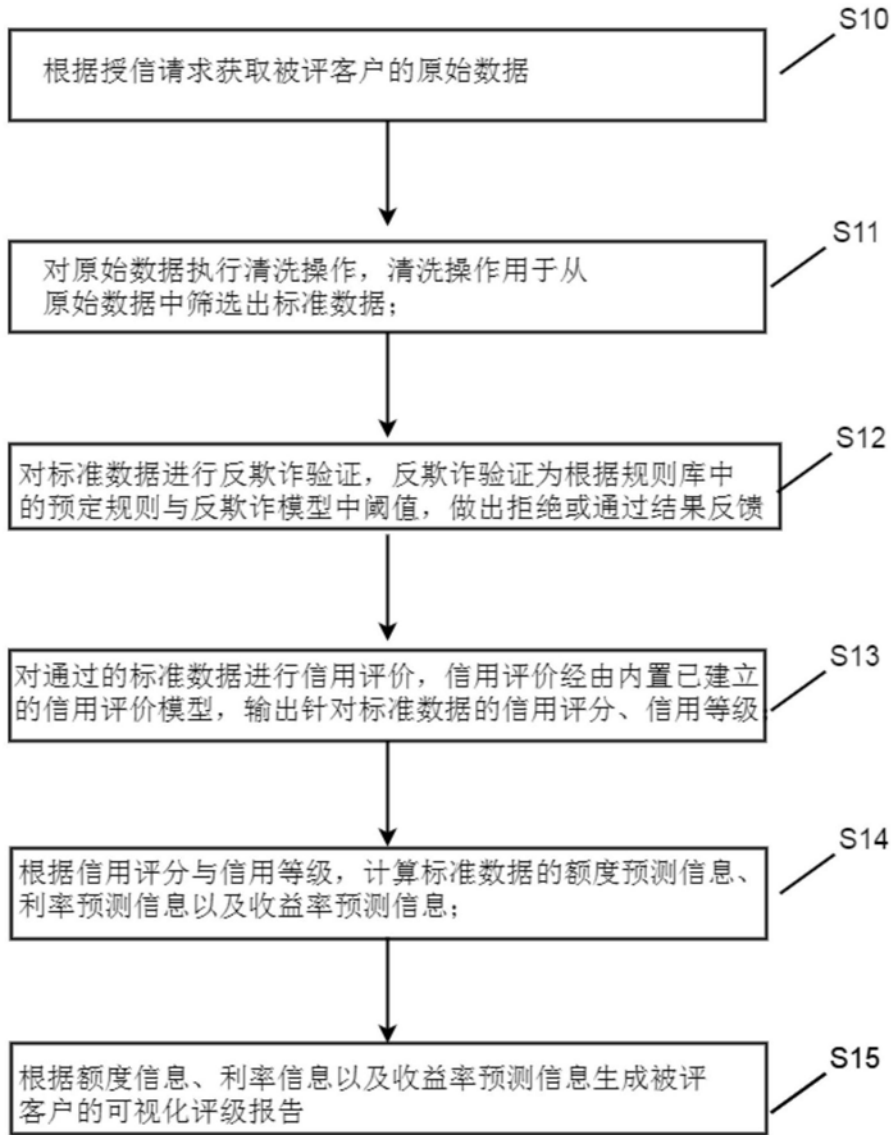


图1

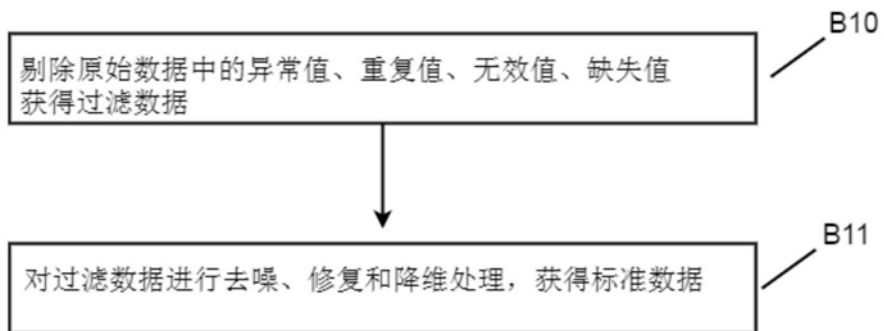


图2

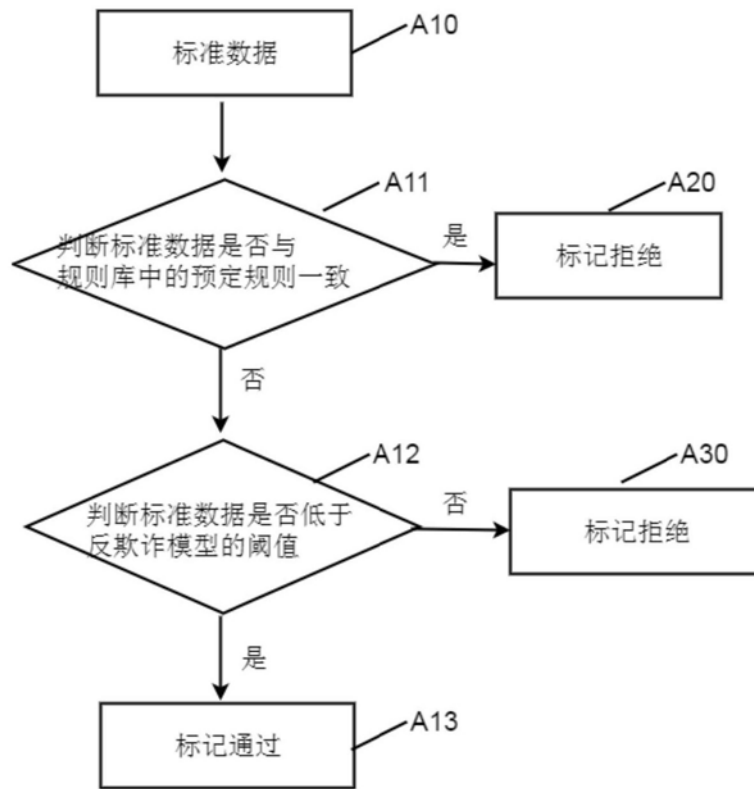


图3

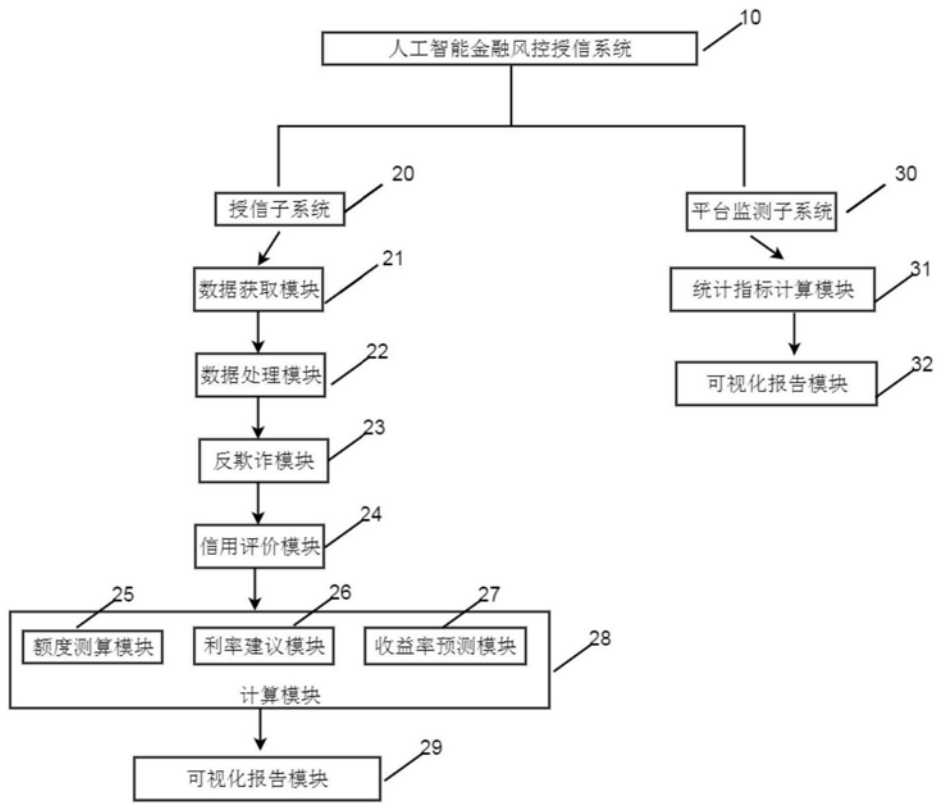


图4

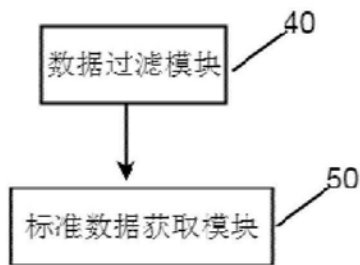


图5

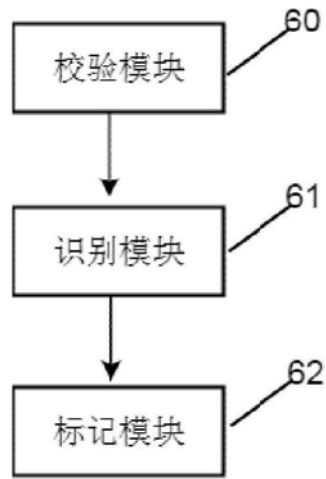


图6

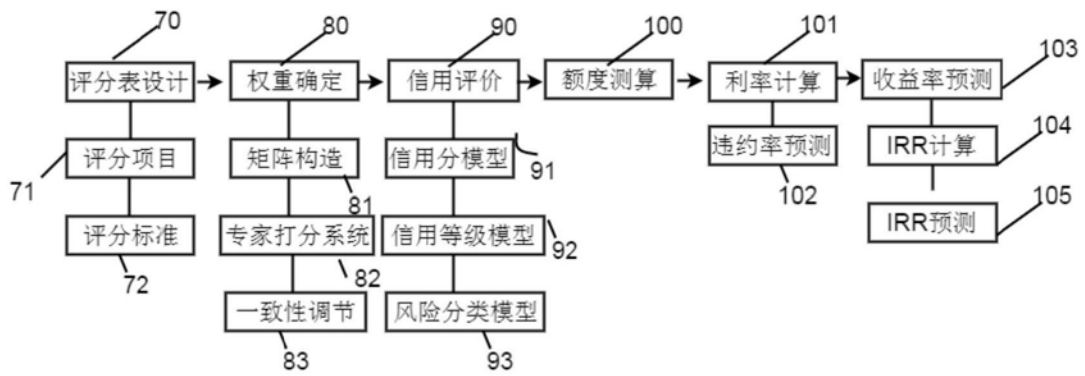


图7