(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 108898479 A
(43)申请公布日 2018.11.27

(21)申请号 201810689255.1

(22)申请日 2018.06.28

(71)申请人 中国农业银行股份有限公司
地址 100005 北京市东城区建国门内大街69号

(72)发明人 赵维平 董晓杰 耿博 刘一阳 李亚琴

(74)专利代理机构 北京集佳知识产权代理有限公司 11227
代理人 钱娜 王宝筠

(51)Int.Cl.
G06Q 40/02(2012.01)
G06F 17/30(2006.01)

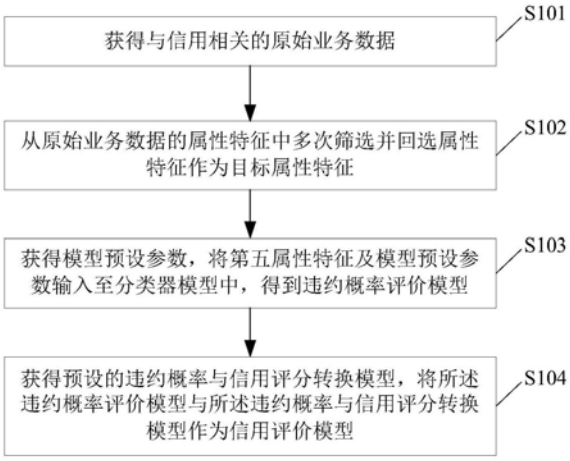
权利要求书4页 说明书12页 附图1页

(54)发明名称

信用评价模型的构建方法及装置

(57)摘要

本申请提供了一种信用评价模型构建方法，该方法可以通过多次筛选及回选的方式，从业务数据的属性特征中选择出对评价信用评分具有影响作用的属性特征，选择出的属性特征用于构建违约概率评价模型，该模型可以计算违约概率，再获得预设的违约概率与信用评分转换模型，该模型可以将违约概率转换为信用评分，因此该两个模型可以作为信用评价模型。另外，本申请还提供了一种信用评价模型构建装置，用以保证所述方法在实际中的应用及实现。



CN 108898479 A

1.一种信用评价模型的构建方法,其特征在于,包括:

获得与信用相关的原始业务数据,所述原始业务数据具有多个初始属性特征,且不同的初始属性特征与信用评价的关联程度不同;

基于机器学习算法使用所述初始属性特征构建分类器模型,得到初始属性特征在所述分类器模型中的重要性值,选择重要性值满足预设条件的初始属性特征作为第一属性特征;

将第一属性特征输入到方差分析算法中得到显著性值,并选择显著性值满足预设条件的第一选择属性特征作为第二属性特征;

使用聚类算法对第二属性特征进行聚类,在同一类型的第二属性特征中选择显著性值满足条件的第二属性特征作为第三属性特征;

使用第三属性特征构建分类器模型,并计算初始属性特征的信息值,选择信息值满足预设条件的属性特征作为回选属性特征;

将回选属性特征依次加入到由第三属性特征构建的分类器模型中,判断每次加入回选属性特征后的分类器模型的分类效果是否提高,并将导致分类效果提高的回选属性特征加入到第三属性特征中,将加入有回选属性特征的第三属性特征作为第四属性特征;

获得模型预设参数,使用第四属性特征及模型预设参数构建分类器模型,得到违约概率评价模型;

获得预设的违约概率与信用评分转换模型,将所述违约概率评价模型与所述违约概率与信用评分转换模型作为信用评价模型。

2.根据权利要求1所述的信用评价模型的构建方法,其特征在于,所述将所述基于机器学习算法使用所述初始属性特征构建分类器模型,得到初始属性特征在所述分类器模型中的重要性值,包括:

对所述初始属性特征的特征值进行线性变换,得到衍生属性特征;

基于机器学习算法使用所述初始属性特征及所述衍生属性特征构建分类器模型,得到初始属性特征及衍生属性特征在所述分类器模型中的重要性值。

3.根据权利要求1所述的信用评价模型的构建方法,其特征在于,在使用第四属性特征及模型预设参数构建分类器模型之前,还包括:

对所述原始业务数据进行比例平衡处理,以使所述原始业务数据中的正业务数据及负业务数据的数量比例达到预设的比例,并获得平衡处理后的原始业务数据的属性特征作为平衡属性特征;

将平衡属性特征及第四属性特征进行合并、去重、聚类操作,得到至少一个属性特征集合,并在每个属性特征集合中选择满足条件的属性特征作为第五属性特征;

则所述使用第四属性特征及模型预设参数构建分类器模型,包括:

使用第五属性特征及模型预设参数构建分类器模型。

4.根据权利要求3所述的信用评价模型的构建方法,其特征在于,在使用第五属性特征及模型预设参数构建分类器模型之前,还包括:

将第五属性特征输入至分类器模型中,并计算第五属性特征的相关性系数及第五属性特征的方差膨胀因子,在第五属性特征中选择第五属性特征在分类器模型中的系数与第五属性特征的相关性系数正负号符号一致以及方差膨胀因子满足预设条件的第五属性特征,

并将选择的第五属性特征作为第六属性特征；

则所述使用第五属性特征及模型预设参数构建分类器模型，包括：

使用第六属性特征及模型预设参数构建分类器模型。

5.根据权利要求4所述的信用评价模型的构建方法，其特征在于，在使用第六属性特征及模型预设参数构建分类器模型之前，还包括：

将第六属性特征进行聚类，在每个类型的第六属性特征中选择部分第六属性特征作为备选第六属性特征；

将备选第六属性特征输入至分类器模型中，并计算备选第六属性特征的相关性系数及备选第六属性特征的方差膨胀因子，在备选第六属性特征中选择备选第六属性特征在分类器模型中的系数与备选第六属性特征的相关性系数正负号符号一致以及方差膨胀因子满足预设条件的备选第六属性特征，将选择的备选第六属性特征作为第六属性特征集合；

在第六属性特征集合中逐个删除备选第六属性特征，并将剩余的第六属性特征集合输入至分类器模型中，判断分类器模型的柯尔莫可洛夫-斯米洛夫值是否下降，如果下降，则将删除的备选第六属性特征重新添加回第六属性特征集合；

将以上步骤得到的第六属性特征集合中的属性特征作为第七属性特征；

则所述使用第六属性特征及模型预设参数构建分类器模型，包括：

使用第七属性特征及模型预设参数构建分类器模型。

6.一种信用评价模型的构建装置，其特征在于，包括：

业务数据获得单元，用于获得与信用相关的原始业务数据，所述原始业务数据具有多个初始属性特征，且不同的初始属性特征与信用评价的关联程度不同；

第一特征筛选单元，用于基于机器学习算法使用所述初始属性特征构建分类器模型，并得到初始属性特征在所述分类器模型中的重要性值，选择重要性值满足预设条件的初始属性特征作为第一属性特征；

第二特征筛选单元，用于将第一属性特征输入到方差分析算法中得到显著性值，并选择显著性值满足预设条件的第一选择属性特征作为第二属性特征；

第三特征筛选单元，用于使用聚类算法对第二属性特征进行聚类，在同一类型的第二属性特征中选择显著性值满足条件的第二属性特征作为第三属性特征；

回选特征筛选单元，用于使用第三属性特征构建分类器模型，并计算初始属性特征的信息值，选择信息值满足预设条件的属性特征作为回选属性特征；

第四特征筛选单元，用于将回选属性特征依次加入到由第三属性特征构建的分类器模型中，判断每次加入回选属性特征后的分类器模型的分类效果是否提高，并将导致分类效果提高的回选属性特征加入到第三属性特征中，将加入有回选属性特征的第三属性特征作为第四属性特征；

违约概率评价模型生成单元，用于获得模型预设参数，使用第四属性特征及模型预设参数构建分类器模型，得到违约概率评价模型；

信用评价模型生成单元，用于获得预设的违约概率与信用评分转换模型，将所述违约概率评价模型与所述违约概率与信用评分转换模型作为信用评价模型。

7.根据权利要求6所述的信用评价模型的构建装置，其特征在于，第一特征筛选单元用于基于机器学习算法使用所述初始属性特征构建分类器模型，并得到初始属性特征在所述

分类器模型中的重要性值，包括：

第一特征筛选单元，具体用于对所述初始属性特征的特征值进行线性变换，得到衍生属性特征；以及基于机器学习算法使用所述初始属性特征及所述衍生属性特征构建分类器模型，得到初始属性特征及衍生属性特征在所述分类器模型中的重要性值。

8.根据权利要求6所述的信用评价模型的构建装置，其特征在于，还包括：

第五特征筛选单元，用于在将第四属性特征及模型预设参数输入至分类器模型中之前，对所述原始业务数据进行比例平衡处理，以使所述原始业务数据中的正业务数据及负业务数据的数量比例达到预设的比例，并获得平衡处理后的原始业务数据的属性特征作为平衡属性特征；以及将平衡属性特征及第四属性特征进行合并、去重、聚类操作，得到至少一个属性特征集合，并在每个属性特征集合中选择满足条件的属性特征作为第五属性特征；

则违约概率评价模型生成单元用于使用第四属性特征及模型预设参数构建分类器模型，包括：

违约概率评价模型生成单元，具体用于使用第五属性特征及模型预设参数构建分类器模型。

9.根据权利要求8所述的信用评价模型的构建装置，其特征在于，还包括：

第六特征筛选单元，用于在将第五属性特征及模型预设参数输入至分类器模型中之前，将第五属性特征输入至分类器模型中，并计算第五属性特征的相关性系数及第五属性特征的方差膨胀因子，在第五属性特征中选择第五属性特征在分类器模型中的系数与第五属性特征的相关性系数正负号符号一致以及方差膨胀因子满足预设条件的第五属性特征，并将选择的第五属性特征作为第六属性特征；

则违约概率评价模型生成单元用于使用第五属性特征及模型预设参数构建分类器模型，包括：

违约概率评价模型生成单元，具体用于使用第六属性特征及模型预设参数构建分类器模型。

10.根据权利要求9所述的信用评价模型的构建装置，其特征在于，还包括：

第七特征筛选单元，用于在将第六属性特征及模型预设参数输入至分类器模型中之前，将第六属性特征进行聚类，在每个类型的第六属性特征中选择部分第六属性特征作为备选第六属性特征；将备选第六属性特征输入至分类器模型中，并计算备选第六属性特征的相关性系数及备选第六属性特征的方差膨胀因子，在备选第六属性特征中选择备选第六属性特征在分类器模型中的系数与备选第六属性特征的相关性系数正负号符号一致以及方差膨胀因子满足预设条件的备选第六属性特征，将选择的备选第六属性特征作为第六属性特征集合；在第六属性特征集合中逐个删除备选第六属性特征，并将剩余的第六属性特征集合输入至分类器模型中，判断分类器模型的柯尔莫可洛夫-斯米洛夫值是否下降，如果下降，则将删除的备选第六属性特征重新添加回第六属性特征集合；以及将以上步骤得到的第六属性特征集合中的属性特征作为第七属性特征；

则违约概率评价模型生成单元用于使用第六属性特征及模型预设参数构建分类器模型，包括：

违约概率评价模型生成单元，具体用于使用第七属性特征及模型预设参数构建分类器

模型。

信用评价模型的构建方法及装置

技术领域

[0001] 本申请涉及数据处理技术领域，更具体地，是信用评价模型的构建方法及装置。

背景技术

[0002] 信贷业务是银行业的核心业务，信贷的利差收入是银行业的主要收入来源，其中，个人客户在信贷业务中具有显著的长尾效应，随着大数据技术的广泛应用，个人客户业务借助信息技术的力量得到迅速发展，成为银行收入的重要来源之一。银行业为了保证良好的运作，不仅需要营销客户来开源，还需要防控风险，避免坏账的发生。其中导致坏账的一个情况是个人客户的违约行为，如逾期不还款。

[0003] 为了减少坏账的发生概率，银行业需要找到信用评价好的个人客户进行产品营销，而为了确定个人客户的信用情况，银行业需要建立信用评价模型，来对个人客户的信用情况进行分析。

发明内容

[0004] 有鉴于此，本申请提供了一种信用评价模型构建方法，用于构建用于评价信用的计算模型。

[0005] 为实现所述目的，本申请提供的技术方案如下：

[0006] 第一方面，本申请提供了一种信用评价模型的构建方法，包括：

[0007] 获得与信用相关的原始业务数据，所述原始业务数据具有多个初始属性特征，且不同的初始属性特征与信用评价的关联程度不同；

[0008] 基于机器学习算法使用所述初始属性特征构建分类器模型，得到初始属性特征在所述分类器模型中的重要性值，选择重要性值满足预设条件的初始属性特征作为第一属性特征；

[0009] 将第一属性特征输入到方差分析算法中得到显著性值，并选择显著性值满足预设条件的第一选择属性特征作为第二属性特征；

[0010] 使用聚类算法对第二属性特征进行聚类，在同一类型的第二属性特征中选择显著性值满足条件的第二属性特征作为第三属性特征；

[0011] 使用第三属性特征构建分类器模型，并计算初始属性特征的信息值，选择信息值满足预设条件的属性特征作为回选属性特征；

[0012] 将回选属性特征依次加入到由第三属性特征构建的分类器模型中，判断每次加入回选属性特征后的分类器模型的分类效果是否提高，并将导致分类效果提高的回选属性特征加入到第三属性特征中，将加入有回选属性特征的第三属性特征作为第四属性特征；

[0013] 获得模型预设参数，使用第四属性特征及模型预设参数构建分类器模型，得到违约概率评价模型；

[0014] 获得预设的违约概率与信用评分转换模型，将所述违约概率评价模型与所述违约概率与信用评分转换模型作为信用评价模型。

[0015] 第二方面，本申请提供了一种信用评价模型的构建装置，包括：

[0016] 业务数据获得单元，用于获得与信用相关的原始业务数据，所述原始业务数据具有多个初始属性特征，且不同的初始属性特征与信用评价的关联程度不同；

[0017] 第一特征筛选单元，用于基于机器学习算法使用所述初始属性特征构建分类器模型，并得到初始属性特征在所述分类器模型中的重要性值，选择重要性值满足预设条件的初始属性特征作为第一属性特征；

[0018] 第二特征筛选单元，用于将第一属性特征输入到方差分析算法中得到显著性值，并选择显著性值满足预设条件的第一选择属性特征作为第二属性特征；

[0019] 第三特征筛选单元，用于使用聚类算法对第二属性特征进行聚类，在同一类型的第二属性特征中选择显著性值满足条件的第二属性特征作为第三属性特征；

[0020] 回选特征筛选单元，用于使用第三属性特征构建分类器模型，并计算初始属性特征的信息值，选择信息值满足预设条件的属性特征作为回选属性特征；

[0021] 第四特征筛选单元，用于将回选属性特征依次加入到由第三属性特征构建的分类器模型中，判断每次加入回选属性特征后的分类器模型的分类效果是否提高，并将导致分类效果提高的回选属性特征加入到第三属性特征中，将加入有回选属性特征的第三属性特征作为第四属性特征；

[0022] 违约概率评价模型生成单元，用于获得模型预设参数，使用第四属性特征及模型预设参数构建分类器模型，得到违约概率评价模型；

[0023] 信用评价模型生成单元，用于获得预设的违约概率与信用评分转换模型，将所述违约概率评价模型与所述违约概率与信用评分转换模型作为信用评价模型。

[0024] 由以上技术方案可知，本申请提供的信用评价模型构建方法，可以通过多次筛选及回选的方式，从业务数据的属性特征中选择出对评价信用评分具有影响作用的属性特征，选择出的属性特征用于构建违约概率评价模型，该模型可以计算违约概率，再获得预设的违约概率与信用评分转换模型，该模型可以将违约概率转换为信用评分，因此该两个模型可以作为信用评价模型。

附图说明

[0025] 为了更清楚地说明本申请实施例或现有技术中的技术方案，下面将对实施例或现有技术描述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申请的实施例，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，还可以根据提供的附图获得其他的附图。

[0026] 图1为本申请提供的信用评价模型构建方法的一种流程图；

[0027] 图2为本申请提供的信用评价模型构建装置的一种结构图。

具体实施方式

[0028] 下面将结合本申请实施例中的附图，对本申请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申请一部分实施例，而不是全部的实施例。基于本申请中的实施例，本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其他实施例，都属于本申请保护的范围。

[0029] 信贷业务作为银行业的核心业务，近几年随着利率市场化和经济下行压力，同业间竞争日益激烈。信贷的利差收入是银行的主要收入来源，各银行不仅需要营销客户以开源，更需要防控风险，避免坏账的发生。其中，个人零售客户在信贷业务中具有显著的长尾效应，随着大数据技术的广泛应用，个人零售业务借助信息技术的力量得以迅速发展，成为银行收入的重要来源之一。

[0030] “好客户”是违约率较低的一类客户，如何通过精准营销争取到此类客户决定了个人零售业务的成功与否。为了找到违约率较低的个人客户，需要构建信用评价模型。信用评价模型用于对个人信用信息进行量化分析，得到违约概率，再将违约概率转化为信用分数。一般地，违约概率越低，信用分数越高。

[0031] 见图1，其示出了本申请提供的一种信用评价模型的构建方法，具体包括步骤S101～S104。

[0032] S101：获得与信用相关的原始业务数据。

[0033] 其中，原始业务数据可以是从大数据平台中获得的业务数据，需要说明的是，由于本申请是需要构建与信用相关的评价模型，因此所获得的业务数据是与信用相关的业务数据。例如，与信用相关的业务数据可以包括：个人身份信息、个人资产信息、个人负债信息、个人贷款信息、个人交易信息等。

[0034] 为了便于与后续经过加工的业务数据区分，从大数据平台获得的业务数据可以称为原始业务数据。

[0035] 执行步骤S102之前，还包含对原始业务数据的预处理步骤。预处理步骤的主要作用是，将不符合业务数据标准的原始业务数据进行特殊处理，以使其符合业务数据标准。例如，将格式异常值转换为格式正常值，为空缺值添加默认值。

[0036] 为了提高业务数据的丰富性，在进行步骤S102之前，还可以基于原始业务数据获得衍生业务数据。衍生的方式可以包括线性变换，线性变换可以包括但不局限于，对数变换、求解平方根、求解立方根等。衍生业务数据与原始业务数据属于相同业务类型，包含的属性特征相同，但属性特征的特征值不同。通过改变属性特征的特征值的分布，可以使属性特征的特征值变得丰富，以探索更加丰富的属性特征是否能够更好地表示出与最终被选择的属性特征之间的关联。

[0037] S102：从原始业务数据的属性特征中多次筛选并回选属性特征作为目标属性特征。

[0038] 其中，原始业务数据具有多个属性特征，为了与后续选择的属性特征区分，可以将该属性特征称为初始属性特征。初始属性特征具有特征值。不同的初始属性特征与信用评价的关联程度不同，例如，个人贷款信息相较于个人资产信息更有助于评价信用情况。

[0039] 具体来讲，在银行系统中，业务数据具有多种多样的属性特征，但并非所有的属性特征都能够影响用户的信用评分，因此需要从用户的业务数据的属性特征中，选择对于评价用户的信用具有帮助作用的属性特征。属性特征也可以称为属性字段、属性变量、影响因子、影响变量。选择出的属性特征可以称为目标属性特征。

[0040] 选择及回选属性特征的具体方式可以包括以下步骤A1～A5。

[0041] A1：将初始属性特征输入至基于机器学习算法构建的分类器模型中得到重要性值，并选择重要性值满足预设条件的属性特征作为第一属性特征。

[0042] 其中，将初始属性特征或者经过处理的初始属性特征输入至分类器模型中。基于机器学习算法构建的分类器模型可以包括但不局限于GBDT(Gradient Boosting Decision Tree，梯度提升决策树)、自提升算法Adaboost、随机森林、逻辑回归模型。

[0043] 梯度提升决策树GBDT是一种迭代的决策树算法，该算法由多棵决策树组成，所有树的结论累加起来做最终答案。自提升算法Adaboost是一种迭代算法，其核心思想是针对同一个训练集训练不同的分类器(弱分类器)，然后把这些弱分类器集合起来，构成一个更强的最终分类器(强分类器)。

[0044] 逻辑回归模型属于线性分类模型，主要用于二分类问题，也可应用于多分类问题。将输入数据拟合到一个sigmoid函数中，输入可以是负无穷到正无穷，而输出总是[0,1]，并且当输入为0时，输出的值为0.5。逻辑回归模型能够完成对事件发生概率进行预测。

[0045] 分类器模型可以输出多个指标，其中重要性值为其中一个指标，根据重要性值对输入的初始属性特征进行排序，将排序在前的预设数量的初始属性特征选择出来。为了与其他步骤选择出来的属性特征区分，可以将本步骤选择出来的属性特征称为第一属性特征。

[0046] 需要说明的是，本申请各个步骤中的将属性特征输入至分类器模型中，表示的是使用属性特征构建分类器模型。

[0047] A2：将第一属性特征输入到方差分析算法中得到显著性值，并选择显著性值满足预设条件的第一选择属性特征作为第二属性特征。

[0048] 其中，将第一属性特征输入到方差分析算法中，方差分析算法中会输出多个指标，其中一个指标为显著性值，根据显著性值对第一属性特征进行排序，并选择排序在前的预设数量的第一属性特征，为了便于与其他步骤选择出来的属性特征区分，可以将本步骤选择出的属性特征称为第二属性特征。

[0049] 可见，步骤A1及步骤A2是，采用GBDT、Adaboost、随机森林等机器学习算法进行属性特征选择，并结合方差分析，将机器学习算法输出的重要且显著的属性特征保留。其中，方差分析(ANOVA)是指，通过分析研究不同来源的变异对总变异的贡献大小，从而确定属性特征对评价结果影响力的大小。

[0050] A3：使用聚类算法对第二属性特征进行聚类，在同一类型的第二属性特征中选择显著性值满足条件的第二属性特征作为第三属性特征。

[0051] 其中，本步骤是为了对同一类型的第二属性特征进行筛选。在筛选前，首先对第二属性特征按照业务类型进行分类。例如，活期存款是一个业务类型，可以将属于活期存款的属性特征聚类在一起。又如，过去3个月的交易平均值、过去6个月的交易平均值这两个属性特征都是过去一段时间的交易平均值，可以将该两个属性特征划分为同一个业务类型。

[0052] 在同一聚类的多个属性特征属于同一类型，可以选择部分属性特征。选择标准可以是显著性值，即选择显著性值满足条件的属性特征。其中条件可以是但不局限于显著性值最高。

[0053] 需要说明的是，聚类算法可以称为聚类分析。聚类分析：指将物理或抽象对象的集合分组为由类似的对象组成的多个类的分析过程。聚类是搜索簇的无监督学习过程。同一个簇中的对象有很大的相似性，而不同簇间的对象有很大的相异性。聚类算法包括但不局限于KMeans算法。

[0054] 简单来讲，聚类操作可以将属性特征划分为多个类型集合，从每个类型集合中选择一部分属性特征。

[0055] A4：使用第三属性特征构建分类器模型，并计算初始属性特征的信息值，选择信息值满足预设条件的属性特征作为回选属性特征。

[0056] 其中，本步骤及步骤A5是为了回选属性特征，即将步骤A1至A3删除的属性特征选择回来。具体地，最初始的属性特征为数量最多的属性特征，计算这些属性特征的信息值。计算到信息值后，按照信息值的大小进行排序，选择排序在前的预设数量的信息值。或者，选择信息值大于预设信息阈值的信息值。选择信息值后，选择这些信息值对应的属性特征，为了便于与其他属性特征区分，可以将选择出的属性特征称为回选属性特征。

[0057] 需要说明的是，信息值(information value，IV)衡量的是变量所含的信息量，在本申请的应用场景中，属性特征作为变量，则衡量的是属性特征所包含的信息量，信息量是对构建信用评价模型的有用信息量。

[0058] A5：将回选属性特征依次加入到由第三属性特征构建的分类器模型中，判断每次加入回选属性特征后的分类器模型的分类效果是否提高，并将导致分类效果提高的回选属性特征加入到第三属性特征中，将加入有回选属性特征的第三属性特征作为第四属性特征。

[0059] 其中，选择出回选属性特征后，需要循环向由第三属性特征构建的分类器中添加回选属性特征，根据分类器的分类效果来判断是否将该回选属性特征重新选择回来。

[0060] 具体地，每次向分类器模型中添加一个回选属性特征。需要说明的是，当前分类器模型中的属性特征为第三属性特征，由于回选属性特征是从初始属性特征中选择出来的，第三属性特征也是从初始属性特征中选择出来的，则回选属性特征中可能包含第三属性特征，因此在每次向分类器中加入回选属性特征时，如果所加入的回选属性特征已经包含在分类器模型中，则将加入的该回选属性特征删除，重新添加新的回选属性特征。

[0061] 向分类器模型中添加一个回选属性特征后，检查分类器模型的柯尔莫可洛夫-斯米洛夫(Kolmogorov-Smirnov，KS)值是否提升。在添加时可以按照一定的顺序添加，即可以按照回选属性特征的信息值由大到小的顺序依次添加。

[0062] 如果KS值没有提升，则说明分类器模型的分类效果没有提高，进而将添加的该回选属性特征再次从分类器模型中删除，并返回添加新的回选属性特征，再对该新的回选属性特征进行上述判断。

[0063] 如果KS值提升，则获得分类器模型中每一个属性特征在分类器模型中的系数以及自身的相关系数，判断每一个属性特征在分类器模型中的系数与自身的相关系数的正负符号是否一致。

[0064] 如果每一个属性特征在分类器模型中的系数与自身的相关系数的正负符号均一致，则返回添加新的回选属性特征，再对该新的回选属性特征进行上述判断。

[0065] 如果判断出某一个属性特征的符号不一致，则将该某一个属性特征剔除，使用剩余的属性特征重新构建分类器模型，再重新判断重新构建的分类器模型中的每一个属性特征在分类器模型中的系数与自身的相关系数的正负符号是否一致，重复该剔除、重新构建、重复判断，直至每一个属性特征在分类器模型中的系数与自身的相关系数的正负符号一致。需要说明的是，如果剔除的该某一个属性特征正是加入的该回选属性特征，则返回添加

新的回选属性特征。

[0066] 如果在对某次构建的分类器模型进行判断时发现，每一个属性特征在该某次构建的分类器模型中的系数与自身的相关系数的正负符号一致，但是，还需要进一步判断该某次构建的分类器模型的KS值相较于未添加回选属性特征时是否提升，如果提升，则返回添加新的回选属性特征。如果未提升，则说明虽然添加了回选属性特征，但可能剔除了一些重要的回选属性特征，从而导致该某次构建的分类器模型的KS值降低了，因此还是将添加的该回选属性特征从分类器模型中剔除，返回添加新的回选属性特征。

[0067] 如果没有新的回选属性特征，则将分类器模型中的属性特征(可能加入有回选属性特征，也可能没有加入有回选属性特征)作为第四属性特征。

[0068] 通过上述具体实现方式可以看出，分类器模型的分类效果是否提高通过KS值是否提升及分类器模型中的属性特征的系数符号是否一致两个因素来判断。如果KS值没提升，则直接确定分类器模型的分类效果没有提高，如果KS值有提升，则进一步判断分类器模型中的属性特征的系数符号是否一致，只有在全部一致的情况下才确定分类效果提高。

[0069] KS值是累计坏占比曲线和累计好占比曲线差值的最大值。KS值表示了模型将正样本和负样本区分开来的能力。KS值越大，模型的预测准确性越好。

[0070] S103:获得模型预设参数，将第四属性特征及模型预设参数输入至分类器模型中，得到违约概率评价模型。

[0071] 其中，获得为违约概率评价模型设置的模型预设参数，参数是可以调整的，具体调整方式为根据正负业务数据样本比例调节目标变量类别权重参数，较高比例的样本集具有较高的类别权重，使得输出的违约概率值具有sigmoid函数的分布特点，从而达到业务要求。

[0072] 或者，可以使用KS值或AUC(Area Under Curve，曲线下的面积)值来作为模型评价标准，根据评价标准调整模型预设参数的值。

[0073] 需要说明的是，本步骤中的分类器模型可以包括但不局限于逻辑回归模型、GBDT模型或Adaboost等模型。

[0074] 以上具有第四属性特征及模型预设参数的分类器模型，可以用来计算任何一个未知用户的违约概率，因此可以将该模型称为违约概率评价模型。

[0075] S104:获得预设的违约概率与信用评分转换模型，将所述违约概率评价模型与所述违约概率与信用评分转换模型作为信用评价模型。

[0076] 其中，违约概率与信用评分转换模型是预设的模型，用于将上述违约概率评价模型得到的违约概率输出的违约概率转换为信用评分。

[0077] 例如，违约概率与信用评分转换模型可以是:Y＝A+B*LOG((1-q)/q)，其中Y为信用评分，A和B为具有预设值的参数，q为使用违约概率评价模型得到的违约概率。将违约概率输入至该模型中，便可以得到信用评分。

[0078] 因此信用评价模型可以包括两个模型，一个是用于得到违约概率，一个是用于将违约概率转换为信用评分。需要说明的是，以上步骤S101～S103构建违约概率评价模型的过程是，为了从业务数据的属性特征中，选择出一些属性特征，这些属性特征可以用来评价信用评分，从而可以作为违约概率评价模型的变量。信用评价模型可以应用于各营销和风控系统中，营销或风险管理人员依据客户的信用评分，在实际工作中做出有利的经营决策。

[0079] 由以上技术方案可知，本申请提供的信用评价模型构建方法，可以通过多次筛选及回选的方式，从业务数据的属性特征中选择出对评价信用评分具有影响作用的属性特征，选择出的属性特征用于构建违约概率评价模型，该模型可以计算违约概率，再获得预设的违约概率与信用评分转换模型，该模型可以将违约概率转换为信用评分，因此该两个模型可以作为信用评价模型。

[0080] 需要说明的是，本申请在构建违约概率评价模型时，对删除的属性特征进行回选，降低了因抽样可能导致的属性特征丢失问题，增强了所构建的模型的稳定性。

[0081] 另外，本申请的构建方法灵活可变。该方法基于机器学习技术，以计算机理论为基础，较传统的信用评分工具和统计学理论，算法更加丰富，可调节的参数更多，因此建立的模型具有更强的灵活性，适用性更强。

[0082] 再者，本申请的构建流程自动化。该方法实现了从数据加载到客户信用评分的全流程自动化，几乎无需人工干预，大大减少了工作量和主观判断，较传统的建模流程更便捷和客观。

[0083] 目前，信用评价模型的构建方式中，基于统计学构建方法选择属性特征。其中，统计学构建方法主要包括向前选择法、向后剔除法、逐步回归法。

[0084] 向前选择法，是从逻辑回归模型中最显著的预测开始，循环向逻辑回归模型中添加属性特征。在添加前确定添加的标准，添加过程中按属性特征的贡献程度从大到小依次加入到逻辑回归模型中，每添加一个属性特征，需要重新计算剩余属性特征的贡献程度，直至模型外所有属性特征无法达到标准为止。属性特征一旦添加至模型中，便不会被删除。

[0085] 对于向前选择法，Y对每一个变量作直线回归，对偏回归平方和最大的变量进行F检验，p值满足要求则进入模型。每次循环做回归并检验，因为过程中不再对引入的变量做删除，可能出现的问题是后续变量的引入可能会使得先进入模型的变量变得不重要或者出现共线性。

[0086] 向后剔除法，是将所有属性特征作为逻辑回顾模型的变量集合，每一次循环删除变量集合中最小显著性的属性特征。类似向前选择法，事先确定一个剔除属性特征的标准，按照属性特征的贡献程度从小到大依次剔除。每剔除一个属性特征，则需要重新计算剩余属性特征的贡献，直至集合内所有属性特征无法达到剔除标准为止。属性特征一旦被剔除，便不会被加入到模型的变量集合中。

[0087] 对于向后剔除法，Y对每一个变量作直线回归，对偏回归平方和最小的变量进行F检验，p值超过阈值则从模型中删除，并重复上述过程。可能出现的问题在于若自变量高度相关，可能得不到正确的结果。

[0088] 逐步回归法，结合了向前选择法及向后剔除法，每一次循环既增加属性特征，也删除属性特征。

[0089] 对于逐步回归法，如果自变量间的共线性较强，改变变量的顺序，则得到的结果也会不一样，因此是一种不稳定变量选择方法。此外，自变量进入模型的顺序并不反映它们的重要程度，不利于建模人员进行调优。

[0090] 与此同时，如果数据集中的正负样本比例差距较大，统计学模型中无法根据样本比例通过合适的参数进行调节，只能被动接受模型的输出，容易影响模型的效果。本申请并非简单地使用向前剔除法及向后剔除法进行变量选择，可以避免出现上述问题。

[0091] 为了进一步提高步骤S103中输入至分类器模型中的属性特征(即第四属性特征)的准确度,可以继续对该第四属性特征进行筛选。如下所示,在步骤A5之后增加的筛选步骤可以包括:A6～A8。

[0092] 需要说明的是,步骤A6可以称为初步筛选、步骤A7可以称为二次筛选、步骤A8可以称为三次筛选。A6～A8可以并非一次性全部添加至流程中,可以分别添加一个步骤,两个步骤及三个步骤,以分别形成三个信用评价模型的构建流程。

[0093] A6:对原始业务数据进行比例平衡处理,以使原始业务数据中的正业务数据及负业务数据的数量比例达到预设的比例,并获得平衡处理后的原始业务数据的属性特征作为平衡属性特征;将平衡属性特征及第四属性特征进行合并、去重、聚类操作,得到至少一个属性特征集合,并在每个属性特征集合中选择满足条件的属性特征作为第五属性特征。

[0094] 具体来讲,原始业务数据包括但不局限于:交易数据、资产数据、用户信息数据等等。原始业务数据经过预处理加工以后得到的业务数据样本,可以分为正业务数据样本及负业务数据样本。正业务数据样本为符合信用标准的样本,如按期还款的用户的业务数据,反之,不符合信用标准的业务数据样本为负业务数据样本,如逾期还款的用户的业务数据。当然,信用标准可以是根据实际业务需求而定义的其他标准。

[0095] 在实际应用中,负业务数据样本的数量相较于正业务数据样本而言较少,因此业务数据样本中,正负业务数据样本比例不平衡,使用比例不平衡的业务数据样本筛选属性特征,会导致某些较为重要属性特征的遗漏,从而导致所最终选择的属性特征不够准确,因此,需要对业务数据样本进行平衡处理。

[0096] 平衡处理方式可以是但不局限于以下方式:

[0097] 针对正业务数据样本按照预设比例欠抽样,得到抽样后的正业务数据样本。使用负业务数据样本进行新的负业务数据样本的合成,将合成的负业务数据样本加入到负业务数据样本中,得到平衡后的负业务数据样本。这个过程可以称为过采样,该过程需要满足要求,即抽样后的正业务数据样本,以及平衡后的负业务数据样本的数量比例可以达到预设的比例要求。其中预设的比例根据逻辑回归模型的KS值确定。

[0098] 其中,对负业务数据样本的合成方法可以使用但不局限于smote(Synthetic Minority Oversampling Technique,合成少数类过采样技术)算法。具体地,smote算法随机过采样算法的改进算法,由于随机过采样采取简单复制样本的策略来增加少数类样本,这样容易产生模型过拟合的问题,即使得逻辑回归模型学习到的信息过于特别而不够泛化。smote算法的基本思想是对少数类样本进行分析并根据少数类样本人工合成新样本添加到数据集中。或者说,smote算法利用特征空间中现存少数类样本之间的相似性来建立人工数据的。具体地,smote算法根据样本集合S生成子集S_{min},对于每一个样本$x_i \subset S_{min}$使用K-近邻法得到新的样本,并将得到的新的样本加入到样本集合S中,其中K是某些制定的整数。K-近邻被定义为:子集S_{min}中的K个样本与样本x_i的欧氏距离在n维特征空间X中表现为最小幅度值的样本。

[0099] 平衡后的业务数据样本,可以进行属性特征筛选。具体地,可以将平衡后的业务数据样本输入到基于机器学习算法构建的分类器中,输出的属性特征具有重要性排名,从而可以选择排名在前预设名次的属性特征。或者,可以在将平衡后的业务数据样本输入到分类器之前进行抽样,抽样的样本输入到分类器中,从而选择出属性特征,如此循环抽样N次,

在所选择的属性特征中查找出现次数为L次的属性特征，将查找的属性特征作为最终筛选的属性特征。为了便于将此处筛选出的属性特征与其他步骤选择的属性特征区分，可以将此处筛选出的属性特征称为平衡属性特征。

[0100] 在得到平衡属性特征后，将平衡属性特征与第四属性特征进行合并、去重，再使用聚类算法对去重后的属性特征进行分类，从每个分类中选择部分满足条件的属性特征。选择的条件可以是，如果一个分类中只有一个属性特征，则选择该属性特征；如果一个分类中包含多个属性特征，则选择显著性值较小的预设数量的属性特征。

[0101] 为了便于描述，将选择的属性特征作为第五属性特征。得到第五属性特征后，一种方式是可以直接将第五属性特征替换掉步骤S103第四属性特征，将第五属性特征及模型预设参数输入至分类器模型中，得到违约概率评价模型，另一种方式是可以继续对第五属性特征进行下述步骤A7的处理。

[0102] A7：将第五属性特征输入至分类器模型中，并计算第五属性特征的相关性系数及第五属性特征的方差膨胀因子，在第五属性特征中选择第五属性特征在分类器模型中的系数与第五属性特征的相关性系数正负号符号一致以及方差膨胀因子满足预设条件的第五属性特征，并将选择的第五属性特征作为第六属性特征。

[0103] 具体地，将第五属性特征输入至分类器模型后，需要判断两个方面的特征是否满足要求，一是，第五属性特征的相关性系数与该第五属性特征在分类器模型中的系数正负号符号是否一致，二是第五属性特征的方差膨胀因子是否满足预设条件，预设条件是方差膨胀因子小于预设阈值，为了便于与其他阈值区分，可以将该预设阈值称为预设因子阈值。

[0104] 方差膨胀因子(Variance Inflation Factor，VIF)：指解释变量之间存在多重共线性时的方差与不存在多重共线性时的方差之比。例如，VIF大于预设阈值X时，说明解释变量之间存在较强的共线性，易造成模型不稳定，因此要求VIF小于预设阈值X。在本申请的应用场景中，解释变量指的是属性特征。

[0105] 需要说明的是，将第五属性特征添加至分类器模型中，根据第五属性特征构建分类器模型，判断所构建的分类器模型中的每个属性特征(即第五属性特征)是否满足上述两个要求。如果以上两个方面的判断结果均为是，如果有一个方面不满足要求，则将不满足要求的属性特征从分类器模型中剔除。

[0106] 再使用剩余的属性特征重新构建分类器模型，再重新判断重新构建的分类器模型中的每一个属性特征是否满足上述两个要求，重复该剔除、重新构建、重复判断，直至每一个属性特征是否满足上述两个要求，将最后构建的分类器模型中的第五属性特征称为第六属性特征。

[0107] 得到第六属性特征后，一种方式是可以直接将第六属性特征替换掉步骤S103第四属性特征，将第六属性特征及模型预设参数输入至分类器模型中，得到违约概率评价模型，另一种方式是可以继续对第六属性特征进行下述步骤A8的处理。

[0108] A8：将第六属性特征进行聚类，在每个类型的第六属性特征中选择部分第六属性特征作为备选第六属性特征；将备选第六属性特征输入至分类器模型中，并计算备选第六属性特征的相关性系数及备选第六属性特征的方差膨胀因子，在备选第六属性特征中选择备选第六属性特征在分类器模型中的系数与备选第六属性特征的相关性系数正负号符号一致以及方差膨胀因子满足预设条件的备选第六属性特征，将选择的备选第六属性特征作

为第六属性特征集合;在第六属性特征集合中逐个删除备选第六属性特征,并将剩余的第六属性特征集合输入至分类器模型中,判断分类器模型的柯尔莫可洛夫-斯米洛夫值是否下降,如果下降,则将删除的备选第六属性特征重新添加回第六属性特征集合;将以上步骤得到的第六属性特征集合中的属性特征作为第七属性特征。

[0109] 具体地,选择备选第六属性特征可以是人工方式,即人工选择同类第六属性特征中的一些属性特征,或者可以是随机方式,即随机选择同类第六属性特征中的一些属性特征,或者可以使用其他方式选择同类第六属性特征。为了便于描述,可以将选择的第六属性特征称为备选第六属性特征。

[0110] 得到备选第六属性特征后,可以按照步骤A7中的方式筛选备选第六属性特征,与步骤A7中的筛选方式不同的是,本步骤中用于检验方差膨胀因子是否满足条件的预设阈值X_2,比步骤A7中的预设阈值X_1要更严格一些,从而可以更严格地筛选出满足条件的属性特征。其中更严格表现为预设阈值X_2比预设阈值X_1更小一些。

[0111] 为了便于描述,可以将选择的备选第六属性特征作为第六属性特征集合。然后对第六属性特征集合中的每个属性特征进行依次筛选,筛选方式是判断这些属性特征是否是必要的,是否为必要的判断方式是将这些属性特征依次从第六属性特征集合中删除,检验分类器模型的KS值是否下降。如果是必要的,则将删除的属性特征重新添加回来。

[0112] 为了便于描述,将以上步骤得到的第六属性特征集合中的属性特征称为第七属性特征。得到第七属性特征后,可以直接将第七属性特征替换掉步骤S103第四属性特征,将第七属性特征及模型预设参数输入至分类器模型中,得到违约概率评价模型。

[0113] 在实际应用中,以上使用递归特征消除(Recursive feature elimination,RFE)算法反复构建分类器模型,从中选取最好的属性特征。

[0114] 见图2,其示出了本申请提供的一种信用评价模型的构建装置,包括:

[0115] 业务数据获得单元201,用于获得与信用相关的原始业务数据,所述原始业务数据具有多个初始属性特征,且不同的初始属性特征与信用评价的关联程度不同;

[0116] 第一特征筛选单元202,用于基于机器学习算法使用所述初始属性特征构建分类器模型,并得到初始属性特征在所述分类器模型中的重要性值,选择重要性值满足预设条件的初始属性特征作为第一属性特征;

[0117] 第二特征筛选单元203,用于将第一属性特征输入到方差分析算法中得到显著性值,并选择显著性值满足预设条件的第一选择属性特征作为第二属性特征;

[0118] 第三特征筛选单元204,用于使用聚类算法对第二属性特征进行聚类,在同一类型的第二属性特征中选择显著性值满足条件的第二属性特征作为第三属性特征;

[0119] 回选特征筛选单元205,用于使用第三属性特征构建分类器模型,并计算初始属性特征的信息值,选择信息值满足预设条件的属性特征作为回选属性特征;

[0120] 第四特征筛选单元206,用于将回选属性特征依次加入到由第三属性特征构建的分类器模型中,判断每次加入回选属性特征后的分类器模型的分类效果是否提高,并将导致分类效果提高的回选属性特征加入到第三属性特征中,将加入有回选属性特征的第三属性特征作为第四属性特征;

[0121] 违约概率评价模型生成单元207,用于获得模型预设参数,使用第四属性特征及模型预设参数构建分类器模型,得到违约概率评价模型;

[0122] 信用评价模型生成单元208，用于获得预设的违约概率与信用评分转换模型，将所述违约概率评价模型与所述违约概率与信用评分转换模型作为信用评价模型。

[0123] 在一个示例中，第一特征筛选单元用于基于机器学习算法使用所述初始属性特征构建分类器模型，并得到初始属性特征在所述分类器模型中的重要性值，包括：

[0124] 第一特征筛选单元，具体用于对所述初始属性特征的特征值进行线性变换，得到衍生属性特征；以及基于机器学习算法使用所述初始属性特征及所述衍生属性特征构建分类器模型，得到初始属性特征及衍生属性特征在所述分类器模型中的重要性值。

[0125] 在一个示例中，信用评价模型的构建装置还包括：

[0126] 第五特征筛选单元，用于在将第四属性特征及模型预设参数输入至分类器模型中之前，对所述原始业务数据进行比例平衡处理，以使所述原始业务数据中的正业务数据及负业务数据的数量比例达到预设的比例，并获得平衡处理后的原始业务数据的属性特征作为平衡属性特征；以及将平衡属性特征及第四属性特征进行合并、去重、聚类操作，得到至少一个属性特征集合，并在每个属性特征集合中选择满足条件的属性特征作为第五属性特征；

[0127] 则违约概率评价模型生成单元用于使用第四属性特征及模型预设参数构建分类器模型，包括：

[0128] 违约概率评价模型生成单元，具体用于使用第五属性特征及模型预设参数构建分类器模型。

[0129] 在一个示例中，信用评价模型的构建装置还包括：

[0130] 第六特征筛选单元，用于在将第五属性特征及模型预设参数输入至分类器模型中之前，将第五属性特征输入至分类器模型中，并计算第五属性特征的相关性系数及第五属性特征的方差膨胀因子，在第五属性特征中选择第五属性特征在分类器模型中的系数与第五属性特征的相关性系数正负号符号一致以及方差膨胀因子满足预设条件的第五属性特征，并将选择的第五属性特征作为第六属性特征；

[0131] 则违约概率评价模型生成单元用于使用第五属性特征及模型预设参数构建分类器模型，包括：

[0132] 违约概率评价模型生成单元，具体用于使用第六属性特征及模型预设参数构建分类器模型。

[0133] 在一个示例中，信用评价模型的构建装置还包括：

[0134] 第七特征筛选单元，用于在将第六属性特征及模型预设参数输入至分类器模型中之前，将第六属性特征进行聚类，在每个类型的第六属性特征中选择部分第六属性特征作为备选第六属性特征；将备选第六属性特征输入至分类器模型中，并计算备选第六属性特征的相关性系数及备选第六属性特征的方差膨胀因子，在备选第六属性特征中选择备选第六属性特征在分类器模型中的系数与备选第六属性特征的相关性系数正负号符号一致以及方差膨胀因子满足预设条件的备选第六属性特征，将选择的备选第六属性特征作为第六属性特征集合；在第六属性特征集合中逐个删除备选第六属性特征，并将剩余的第六属性特征集合输入至分类器模型中，判断分类器模型的柯尔莫可洛夫-斯米洛夫值是否下降，如果下降，则将删除的备选第六属性特征重新添加回第六属性特征集合；以及将以上步骤得到的第六属性特征集合中的属性特征作为第七属性特征；

[0135] 则违约概率评价模型生成单元用于使用第六属性特征及模型预设参数构建分类器模型，包括：

[0136] 违约概率评价模型生成单元，具体用于使用第七属性特征及模型预设参数构建分类器模型。

[0137] 需要说明的是，本说明书中的各个实施例均采用递进的方式描述，每个实施例重点说明的都是与其他实施例的不同之处，各个实施例之间相同相似的部分互相参见即可。

[0138] 还需要说明的是，在本文中，诸如第一和第二等之类的关系术语仅仅用来将一个实体或者操作与另一个实体或操作区分开来，而不一定要求或者暗示这些实体或操作之间存在任何这种实际的关系或者顺序。而且，术语“包括”、“包含”或者其任何其他变体意在涵盖非排他性的包含，从而使得包括一系列要素的过程、方法、物品或者设备不仅包括那些要素，而且还包括没有明确列出的其他要素，或者是还包括为这种过程、方法、物品或者设备所固有的要素。在没有更多限制的情况下，由语句“包括一个……”限定的要素，并不排除在包括上述要素的过程、方法、物品或者设备中还存在另外的相同要素。

[0139] 对所公开的实施例的上述说明，使本领域专业技术人员能够实现或使用本申请。对这些实施例的多种修改对本领域的专业技术人员来说将是显而易见的，本文中所定义的一般原理可以在不脱离本申请的精神或范围的情况下，在其它实施例中实现。因此，本申请将不会被限制于本文所示的这些实施例，而是要符合与本文所公开的原理和新颖特点相一致的最宽的范围。

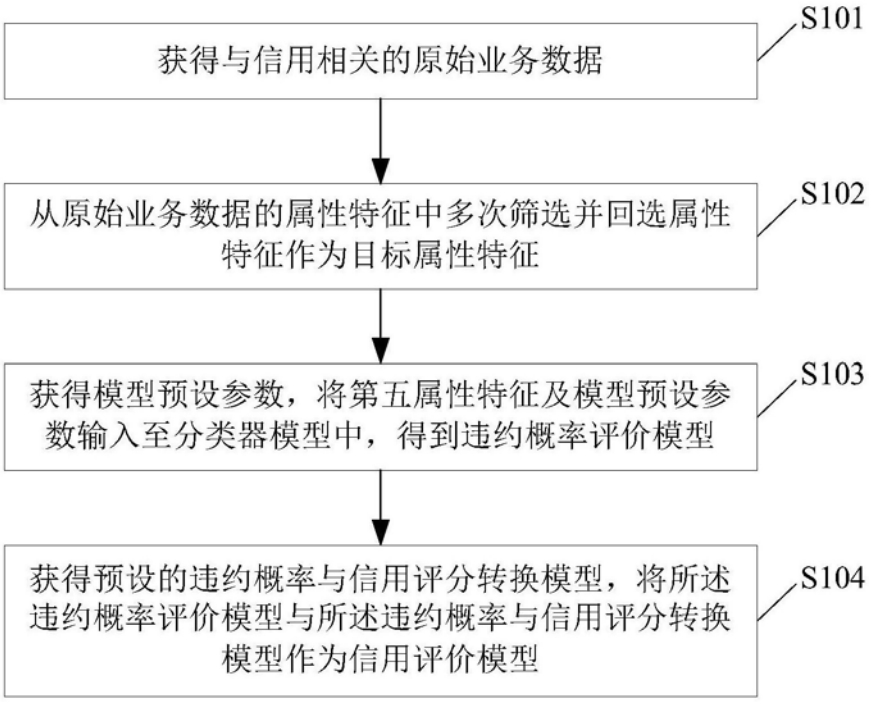

S101
获得与信用相关的原始业务数据
S102
从原始业务数据的属性特征中多次筛选并回选属性特征作为目标属性特征
S103
获得模型预设参数，将第五属性特征及模型预设参数输入至分类器模型中，得到违约概率评价模型
S104
获得预设的违约概率与信用评分转换模型，将所述违约概率评价模型与所述违约概率与信用评分转换模型作为信用评价模型

图1

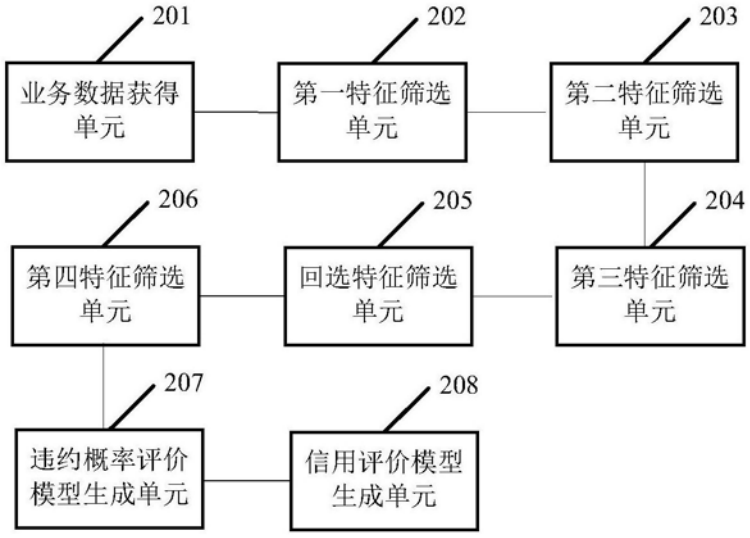

201
业务数据获得单元
202
第一特征筛选单元
203
第二特征筛选单元
204
第三特征筛选单元
205
回选特征筛选单元
206
第四特征筛选单元
207
违约概率评价模型生成单元
208
信用评价模型生成单元

图2