一种评分卡模型的调整方法、装置、服务器及存储介质

技术领域

本发明涉及计算机技术领域，尤其涉及一种评分卡模型的调整方法装置、服务器及存储介质。

背景技术

目前，传统评分卡模型在模型建立后各个维度（即变量）、各个维度的系数以及各个维度对应的证据权重（Weight of Evidence，WOE）编码值均是固定不变的，后续无法调整该模型。但对于一些高基数且变量下个各分组的数据变化比较频繁的滚动变量而言，在传统评分卡的模型筛选阶段，很难通过信息量（Information Value，IV）指标将这类滚动变量选入模型，严重影响了评分卡模型风险预测结果的准确性。

发明内容

本发明实施例提供了一种评分卡模型的调整方法、装置、服务器及存储介质，可以将滚动变量选入评分卡模型，并利用该滚动变量对评分卡模型进行调整，有利于提高评分卡模型风险预测结果的准确性。

第一方面，本发明实施例提供了一种评分卡模型的调整方法，所述方法包括:

从所述评分卡模型的多个候选自变量中确定出至少一个高基数变量;

根据预设规则从所述至少一个高基数变量中确定出滚动变量;

获取所述滚动变量下各个分组在预设时间内的参数信息，并根据所述参数信息确定所述各个分组各自对应的证据权重 WOE 值;

根据所述滚动变量下各个分组各自对应的WOE值以及所述滚动变量对所述评分卡模型进行调整。

一个实施例中，所述从所述评分卡模型的多个候选自变量中确定出至少一个高基数变量的具体是实施方式为:

计算所述评分卡模型的多个候选自变量中各个候选自变量各自对应的信

息量 IV 值，并输出所述各个候选自变量各自对应的 IV 值；

获取用户根据所述各个候选自变量各自对应的 IV 值输入的用于确定高基数变量的指示信息；

根据所述指示信息从所述多个候选自变量中确定出至少一个高基数变量。

在一个实施例中，所述从所述评分卡模型的多个候选自变量中确定出至少一个高基数变量的具体实施方式为：

计算所述评分卡模型的多个候选自变量中的各个候选自变量各自对应的 IV 值，并将所述 IV 值大于预设 IV 阈值的变量确定为目标变量，所述目标变量下包括至少一个分组；

获取所述目标变量下各个分组各自对应的 WOE 值；

如果各个第一差值中大于预设 WOE 差值阈值的数量满足预设高基数条件，则将所述目标变量确定为高基数变量，所述第一差值为任意两个分组各自对应的 WOE 值之间的差值。

在一个实施例中，所述高基数变量下包括至少一个分组，所述根据预设规则从所述至少一个高基数变量中确定出滚动变量的具体实施方式为：

获取所述至少一个高基数变量中的各个高基数变量下各个分组在周期内的数据变化信息；如果所述各个分组的数据变化信息满足预设数据变化条件，则将对应的高基数变量确定为滚动变量。

在一个实施例中，所述评分卡模型是基于线性回归模型建立的，所述线性回归模型由至少一个变量以及所述至少一个变量中各个变量各自对应的权重系数组成，所述根据所述各个分组各自对应的 WOE 值以及所述滚动变量对所述评分卡模型进行调整的具体实施方式为：

在所述评分卡模型对应的所述线性回归模型中增加所述滚动变量；

根据所述滚动变量下各个分组各自对应的 WOE 值确定所述滚动变量的值。

在一个实施例中，所述获取所述至少一个高基数变量中的各个高基数变量下各个分组在周期内的数据变化信息的具体实施方式为：

统计所述至少一个高基数变量中的各个高基数变量下各个分组在周期内的数值和/或坏账率；

根据统计结果确定出所述各个高基数变量下各个分组在所述周期内的数量变化信息和/或坏账率变化信息;

基于所述数值变化信息和/或所述坏账变化率信息生成所述各个高基数变量下各个分组在所述周期内的数据变化信息。

在一个实施例中，所述数据变化信息包括以下至少一种：所述高基数变量对应各个分组下数值的变化信息和所述高基数变量对应各个分组的坏账率变化信息，还可以如果所述各个分组下数值变化信息所指示的数值变化率大于或者等于预设数值变化率阈值，或者，所述各个分组下坏账率变化信息所指示的坏账变化率大于或者等于预设坏账变化率阈值，则确定所述各个分组的所述数据变化信息满足预设数据变化条件。

第二方面，本发明实施例提供了一种评分卡模型调整装置，该装置包括：

确定模块，用于从所述评分卡模型的多个候选自变量中确定出至少一个高基数变量;

所述确定模块，还用于根据预设规则从所述至少一个高基数变量中确定出滚动变量，所述滚动变量下包括至少一个分组;

获取模块，用于获取所述滚动变量下各个分组在预设时间内的参数信息;

所述确定模块，还用于根据所述获取模块获取到的所述参数信息确定所述各个分组各自对应的证据权重 WOE 值;

调整模块，用于根据所述滚动变量下各个分组各自对应的 WOE 值以及所述滚动变量对所述评分卡模型进行调整。

第三方面，本发明实施例提供了一种服务器，该服务器包括处理器和存储装置，所述处理器和所述存储装置相互连接，其中，所述存储装置用于存储计算机程序，所述计算机程序包括程序指令，所述处理器被配置用于调用所述程序指令，执行上述第一方面的方法。

第四方面，本发明实施例提供了一种计算机可读存储介质，所述计算机可读存储介质存储有计算机程序，所述计算机程序包括程序指令，所述程序指令当被处理器执行时使所述处理器执行上述第一方面的方法。

本发明实施例中，服务器从评分卡模型的多个候选自变量中确定出至少一个高基数变量，并根据预设规则从至少一个高基数变量中确定出滚动变量，获

取滚动变量下各个分组在预设时间内的参数信息，并根据该参数信息确定各个分组各自对应的证据权重 WOE 值，进而根据各个分组各自对应的 WOE 值以及滚动变量对评分卡模型进行调整。采用本发明，可以将滚动变量选入模型，并利用该滚动变量对模型进行调，有利于提高评分卡模型评分结果的准确性。

附图说明

为了更清楚地说明本发明实施例或现有技术中的技术方案，下面将对实施例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，还可以根据这些附图获得其他的附图。

图 1 是本发明实施例提供的一种评分卡模型的调整方法的流程示意图；

图 2 是本发明实施例提供的另一种评分卡模型的调整方法的流程示意图；

图 3 是本发明实施例提供的一种评分卡模型调整装置的示意性框图；

图 4 是本发明实施例提供的一种服务器的示意性框图。

具体实施方式

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。

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。

评分卡模型为一种预测方法，它可以结合不同的业务数据应用于不同的应用场景。示例性，当评分卡模型为一种信用评分卡模型时，它可以根据过去大量信用卡持卡人的信用记录分析，描述影响个人信用水平的因素，从而帮助贷款机构发放消费信贷。信用评分卡模型的建立主要是利用申请人的特征变量预测其违约概率，进而要求进入信用评分模型的特征变量有较强的的预测能力。

在本发明实施例中，可以采用信息量值（Information Value，IV）来衡量每个变量的预测能力，其中，IV 值与预测能力的对应关系可以如表 1-1 所示。

表 1-1

IV	预测能力
小于 0.03	无预测能力
0.03~0.1	低
0.1~0.2	中
0.2~0.3	高
大于 0.3	极高

在一个实施例中，评分卡模型可以是基于一个线性回归模型建立的，其中，该线性回归模型相当于是在因变量（y）和一个或者多个自变量（x）建立的一种关系，可表示为：

$$y = a + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + ... + \beta_n x_n$$

其中 a 表示截距，x_n（n 为正整数）是被选入模型中的自变量，即为入模指标，β_n 各个自变量对应的系数。

对于传统评分卡模型而言，在模型建立后各个自变量 x_n、各个自变量对应的系数 β_n 以及各自变量对应的 WOE 编码值均是固定不变的，后续无法调整该模型。但对于一些高基数且该变量下个各分组的数据变化比较频繁的滚动变量而言，在模型筛选阶段，很难通过信息量（information value，VI）指标将这类滚动变量选入模型，但这种滚动变量由于存在变化频繁的特性，往往是影响风险预测结果的关键性变量，因此，传统评分卡模型的风险预测结果通常不够准确。

在本发明中，可以通过从评分卡模型的多个候选自变量中确定出至少一个高基数变量，并根据预设规则从至少一个高基数变量中确定出滚动变量 x_{n+1}，并获取滚动变量下的各个分组在预设时间内的参数信息，根据参数信息确定各个分组各自对应的证据权重（Weight of Evidence，WOE）值，进而将滚动变量 x_{n+1} 选入评分卡模型，并根据该滚动变量下各个分组各自对应的 WOE 值确定该滚动变量对应的系数 β_{n+1}，可以提高评分卡模型风险预测结果的准确性。

示例性，对于信用评分卡模型而言，提高风险预测结果的准确性，可以有效帮助贷款机构发放消费信贷，进而有效地控制借贷人的还款逾期。

其中，本发明实施例中所描述的高基数变量可以为该变量下存在多种分组的变量。例如，该变量为省份，该省份下则存在多种分组，如：四川省、广西省、江苏省、广东省、海南省以及辽宁省等等，这种情况下，该省份变量则可以确定为高基数变量。所描述的滚动变量可以为各分组下数值和/或坏账率变化频繁的高基数变量。

在一个实施例中，上述候选自变量下可以包括 m 个分组（m 为大于 0 的整数），候选自变量对应的 IV 值，满足如下公式 1.1:

$$IV = \sum_{i}^{m} IV_i$$

其中，i 为小于 m 的正整数，表示 m 个分组中的第 i 组；IV_i表示第 i 组对应的 IV 值。也即，候选自变量的 IV 值是通过对该自变量下各个分组各自对应的 IV 值进行求和得到的。在本发明实施例中，该IV_i的具体数值可以根据第 i 组的 WOE 值（即WOE_i）来确定，具体可以采用如下公式 1.2:

$$IV_i = \left(\left(G_i / G_T \right) - \left(B_i / B_T \right) \right) * WOE_i$$

其中，上式的G_i是这个组中响应客户的数量，G_T是样本中所有响应客户的数量，B_i是这个组中未响应客户的数量，B_T是样本中所有未响应客户的数量。从上式可以看出，WOE 表示的实际上是“当前分组中响应客户占所有响应客户的比例”和“当前分组中没有响应的客户占所有没有响应的客户的比例”的差异，WOE_i的计算公式可以采用如下公式 1.3:

$$WOE = \ln \frac{\left(G_i / G_T \right)}{\left(B_i / B_T \right)}$$

其中，上述响应客户指的是在评分卡模型中预测变量取值为“是”或者为“1”的个体。例如，在风险评分卡模型中，上述未响应的客户对应的是违约客户，本发明对此不作具体限定。

参见图 1，图 1 是本发明实施例提供的一种评分卡模型的调整方法的流程

示意图，如图所示，该评分卡模型的调整方法可包括:

101、服务器从评分卡模型的多个候选自变量中确定出至少一个高基数变量。

在一个实施例中，服务器可以计算评分卡模型的多个候选自变量中各个候选自变量各自对应的信息量IV值，并输出各个候选变量各自对应的IV值，获取用户根据各个变量各自对应的 IV 值输入的用于确定高基数变量的指示信息，进而根据指示信息从上述多个候选自变量中确定出至少一个高基数变量。

其中，该指示信息为根据用户的指示生成的信息，用于指示服务器从多个候选自变量中确定出至少一个高基数变量。例如，服务器输出了j(j为正整数)个候选自变量各自对应的IV值，也即输出了j个IV值(如IV_1、IV_2、IV_3...IV_j)。这种情况下，用户查看这了j个IV值后，想要将IV_1和IV_2对应的候选自变量确定为高基数变量，则可以针对IV_1和IV_2输入指示信息，用于指示将IV_1和IV_2对应的候选自变量确定为高基数变量。这种情况下，服务器接收到该指示信息后则可以将IV_1和IV_2对应的候选自变量确定为高基数变量。

示例性地，假设评分卡模型包括j(j为正整数)个候选自变量，服务器可以通过公式1.1~1.3计算出每个候选自变量各自对应的IV值，并将计算出的j个IV值(如IV_1、IV_2、IV_3... IV_j)展示在显示界面中。用户查看显示界面中展示的j个IV值后，可以输入指示信息，用于指示将j中的一个或者多个IV值确定为目标IV(如IV_1和IV_2)。进一步地，服务器接收到用户的指示信息后，则可以根据该指示信息从j个IV值中确定出一个或者多个目标IV，并查找到一个或者多个目标IV各自对应的候选自变量，进而将各自对应的候选自变量确定为高基数变量。

在一个实施例中，服务器还可以计算评分卡模型的多个候选自变量中的各个候选自变量各自对应的IV值，并将IV值大于预设IV阈值的变量确定为目标变量，进而获取目标变量下各个分组各自对应的WOE值，如果各个第一差值中大于预设WOE差值阈值的数量满足预设高基数条件，则将目标变量确定为高基数变量。其中，第一差值为任意两个分组各自对应的WOE值之间的差值。

在一个实施例中，上述预设高基数条件为各个第一差值中大于预设WOE

差值阈值的数量大于或者等于预设数量阈值 r_0（r_0 为正整数），评分卡模型包括 j（j 为正整数）个候选自变量。这种情况下，服务器可以利用公式 1.1~1.3 所表征的信息量算法计算出每个候选自变量各自对应的 IV 值，也即获得 j 个 IV 值（如 IV_1、IV_2、IV_3...IV_j）。进一步地，可以将这 j 个 IV 值一一与预设 IV 阈值进行比较，确定出大于预设 IV 阈值的 IV 值为 IV_1，那么则将 IV_1 对应的候选自变量确定为目标变量，其中，该目标变量下包括 r_1（r_1 为正整数）个分组。进一步的，服务器可以根据公式 1.3 计算出目标变量下各个分组各自对应的 WOE 值，获取到 r_1 个 WOE 值，并进一步计算 r_1 个 WOE 值中两两之间的差值（即第一差值），将获取到的所有第一差值与预设 WOE 差值阈值进行比较，确定出存在 b 个第一差值大于预设 WOE 差值阈值，且 b 大于 r_0，则将该目标变量确定为高基数变量。

在一个实施例中，当服务器从评分卡模型的多个候选自变量中确定高基数变量时，还可以直接利用公式 1.3 计算出评分卡模型中任一候选自变量下各个分组各自的 WOE 值，并比较各个 WOE 两两之间的差值，并将该差值大于预设差值阈值的差值确定为目标差值，进步一地，确定目标差值的数量，如果目标差值的数量大于或者等于数量阈值，则可以确定该任一候选自变量为高基数变量。

102、服务器根据预设规则从至少一个高基数变量中确定出滚动变量。

在一个实施例中，服务器确定出至少一个高基数变量后，可以获取任一高基数变量下的一个或者多个分组在某一周期内的数据变化信息，该数据变化信息可以包括各个分组下的数值变化信息和各个分组下的坏账变化率信息中的至少一种。进一步地，服务器可以确定各个分组下的数值变化信息和/或各个分组下的坏账变化率信息是否满足预设数据变化条件，如果满足，则可以确定该任一高基数变量为滚动变量。

103、服务器获取该滚动变量下各个分组在预设时间内的参数信息，并根据参数信息确定各个分组各自对应的证据权重 WOE 值。

104、服务器根据滚动变量下各个分组各自对应的 WOE 值以及滚动变量对评分卡模型进行调整。

其中，该预设时间为一个时间段，该时间段可以对应一个起止日期如 2018

年5月-2018年6月，也可以以当前时间为起始时间，倒推10天、15天或者1个月等。该一段时间可以是系统默认设置的，也可以是根据用户的指示确定的，本发明对此不作具体限定。

在一个实施例中，该参数信息为滚动变量下各个分组在预设时间内的坏账率信息，假设步骤102确定出的滚动变量下包括r_1个分组，该预设时间为2018年5月这一个月。这种情况下，服务器可以获取上述r_1个分组在2018年5月这一个月内的坏账率，并根据该坏账率确定各个分组对应的WOE值，进而利用该滚动变量和各个分组对应的WOE值对评分卡模型进行调整。

在一个实施例中，上述评分卡模型是基于线性回归模型建立的，该线性回归模型由至少一个自变量以及所述至少一个自变量中各个自变量各自对应的权重系数组成。这种情况下，服务器执行步骤104的具体实施方式可以为：在评分卡模型对应的线性回归模型中增加滚动变量，并根据滚动变量下各个分组各自对应的WOE值确定该滚动变量的值，进而实现对线性回归模型的调整，也即实现对评分卡模型的调整。

示例性地，假设评分卡模型用于预测广西省、江苏省、四川省这三个省份贷款用户的还款逾期情况，该评分卡模型是基于线性回归模型$y=a+\beta_1x_1+\beta_2x_2+\beta_3x_3+...+\beta_nx_n$建立的，其中a表示截距，$x_n$（n为正整数）是被选入该模型的自变量，$\beta_n$为各个自变量对应的系数，高基数变量为一个省份变量，该省份变量下包括广西省、江苏省、四川省这3个分组，该预设时间为2018年5月这一个月，2018年5月这一个月内省份变量下各个分组的坏账率信息如表1-2所示，其中，G为坏账的数量，B为非坏账的数量。

表1-2

省份	G	B	合计	坏账比例
广西	400	100	500	20%
江苏	300	200	500	40%
四川	300	200	500	40%
合计	1000	500	1500	33%

进一步地，服务器获取到如表1-2所示的坏账率信息后，则可以根据公式

1.3 确定出省份变量下广西省、江苏省、四川省这 3 个分组的 WOE 值分别为：

$$\ln\frac{\left(400/1000\right)}{\left(100/500\right)}=0.69;\quad \ln\frac{\left(300/1000\right)}{\left(200/500\right)}=-0.287;\quad \ln\frac{\left(300/1000\right)}{\left(200/500\right)}=-0.287;$$

那么，服务器可以将省份这一滚动变量表示为 x_{prov} 并选入线性回归模型，也即上述线性回归模型增加一个 x_{prov} 的滚动变量，增加后的线性回归模型为：$y=a+\beta_1x_1+\beta_2x_2+\beta_3x_3+...+\beta_nx_n+\beta_{n+1}x_{prov}$，其中，当服务器通过该模型预测广西省的还款逾期情况时，则 x_{prov} 的值为广西省对应的 WOE 值 0.69；当服务器通过该模型预测江苏省的还款逾期情况时，则 x_{prov} 的值为广西省对应的 WOE 值 -0.287；当服务器通过该模型预测四川省的还款逾期情况时，则 x_{prov} 的值为四川省对应的 WOE 值-0.287，进而实现对线性回归模型的调整，也即实现对评分卡模型的调整，提高了评分卡模型风险预测结果的准确性。

本发明实施例中，服务器从评分卡模型的多个候选自变量中确定出至少一个高基数变量，并根据预设规则从至少一个高基数变量中确定出滚动变量，获取滚动变量下至少一个分组中各个分组在预设时间内的参数信息，并根据该参数信息确定各个分组各自对应的证据权重 WOE 值，进而根据各个分组各自对应的 WOE 值以及滚动变量对评分卡模型进行调整。采用本发明，可以通过滚动变量对评分卡模型进行调整，进而提高评分卡模型风险预测结果的准确性。

再请参见图 2，图 2 是本发明实施例提供的另一种评分卡模型的调整方法的流程示意图，如图所示，该评分卡模型的调整方法可包括：

201、服务器从评分卡模型的多个候选自变量中确定出至少一个高基数变量。

其中，步骤 201 的具体实现方式可以参见上述实施例中步骤 101 的相关描述，此处不再赘述。

202、服务器获取至少一个高基数变量中的各个高基数变量对应各个分组在周期内的数据变化信息。

其中，该周期可以为一个时间段，该时间段可以对应一个起止日期如 2018 年 5 月-2018 年 6 月，也可以以当前时间为起始时间，倒推 10 天、15 天或者 1 个月等。该周期对应的具体时间段可以是系统默认设置的，也可以是根据用户

的指示确定的。其中，该数据变化信息可以为高基数变量中各分组下的数值变化信息和/或坏账变化率信息。

在一个实施例中，服务器可以统计周期内至少一个高基数变量中的各个高基数变量对应各个分组下的数值和/或坏账率，并根据统计结果确定出该周期内各个高基数变量对应各个分组下的数值变化信息和/或坏账变化率信息，进而基于数值变化信息和/或坏账率变化信息生成各个高基数变量对应各个分组在周期内的数据变化信息。

具体实现中，服务器可以在周期内按照预设时间间隔去获取至少一个高基数变量中的各个高基数变量对应各个分组下的数值和/或坏账率，也即每个时间间隔对应一个获取的时间节点，进而通过统计该周期内获取的各个时间节点下的上述各个分组下的数值和/或坏账率，确定出该周期内各个高基数变量对应各个分组下的数值变化信息和/或坏账率变化信息，从而基于数值变化信息和/或坏账率变化信息生成各个高基数变量对应各个分组在周期内的数据变化信息。

示例性的，假设上述周期为 2018 年 4 月这一个月，该预设时间间隔为 15 天，评分卡模型用于预测 2018 年 4 月这一个月内任一期存在 60 天以上的逾期情况；至少一个高基数变量中某一个高基数变量 x_1 为贷款人年龄，根据年龄的特征，可将年龄这一高基数变量分为 18-25 岁、25-40 岁、40-65 岁等多组，2018 年 4 月这一个月内总共获取过两次数据，一次为 2018 年 4 月 15 日获取，获取的数据为 x_1 各个分组下的数据统计结果如表 2-1 所示；一次为 2018 年 4 月 30 日获取，获取的数据为 x_1 各个分组下的数据统计结果如表 2-2 所示。

表 2-1

2018 年 4 月 15 日			
年龄	未逾期	逾期	坏账率
18-25 岁	200	300	0.60
25-40 岁	100	400	0.80
40-65 岁	300	200	0.67
合计	600	900	0.60

表 2-2

2018 年 4 月 30 日			
年龄	未逾期	逾期	坏账率
18-25 岁	300	200	0.67
25-40 岁	400	100	0.2
40-65 岁	200	300	0.60
合计	900	600	0.4

服务器获取到如表 2-1 和表 2-2 的数据后，通过分析表 2-1 和表 2-2 记录的数据可以确定出 2018 年 4 月这一个月 18-25 岁、25-40 岁以及 40-65 岁这三个分组下的坏账变化率差值（即坏账率变化信息）分别为 0.07、0.6、0.07，同样地，18-25 岁、25-40 岁以及 40-65 岁这三个分组下逾期的数值变化差值分别为 100、300、400，未逾期的数值变化差值分别为 100、300、100，其中，三个分组下逾期和未逾期的数值变化差值则为这三个分组下的数值变化信息。

203、如果服务器确定上述各个分组的数据变化信息满足预设数据变化条件，则将对应的高基数变量确定为滚动变量。

在一个实施例中，数据变化信息包括以下至少一种：高基数变量对应各个分组下的数值变化信息和高基数变量对应各个分组下的坏账率变化信息。上述预设数据变化条件可以为该数值变化信息所指示的数值变化率大于或者等于预设数值变化率阈值，或者，坏账率变化信息所指示的坏账变化率大于或者等于预设坏账变化率阈值。服务器在执行步骤 203 之前，可以从数据变化信息中获取到上述数值变化信息和/或坏账变化率信息，并根据该数值变化信息确定出高基数变量对应各个分组下的数值变化率，根据该坏账率变化信息确定出高基数变量对应各个分组下的坏账变化率。在一个实施例中，服务器可以在高基数变量对应各个分组下的数值变化率大于或者等于预设数值变化率阈值时，确定上述各个分组的数据变化信息满足预设数据变化条件。在另一个实施例中，服务器可以在高基数变量对应各个分组下的坏账变化率大于或者等于预设坏账变化率阈值时，确定上述各个分组的数据变化信息满足预设数据变化条件。在又一个实施例中，服务器也可以在高基数变量对应各个分组下的数值变化率

大于或者等于预设数值变化率阈值，且该高基数变量对应各个分组下的坏账变化率大于或者等于预设坏账变化率阈值时，确定上述各个分组的数据变化信息满足预设数据变化条件。

204、服务器获取该滚动变量下各个分组在预设时间内的参数信息，并根据参数信息确定各个分组各自对应的证据权重 WOE 值。

205、服务器根据滚动变量下各个分组各自对应的 WOE 值以及滚动变量对评分卡模型进行调整。

其中，步骤 204 和步骤 205 的具体实现方式可以参见上述实施例中步骤 103 和步骤 104 的相关描述，此处不再赘述。

本发明实施例中，服务器从评分卡模型的多个候选自变量中确定出至少一个高基数变量，获取至少一个高基数变量中的各个高基数变量对应各个分组在周期内的数据变化信息，如果服务器确定上述各个分组的数据变化信息满足预设数据变化条件，则将对应的高基数变量确定为滚动变量，获取该滚动变量下各个分组在预设时间内的参数信息，并根据参数信息确定各个分组各自对应的证据权重 WOE 值，进而根据滚动变量下各个分组各自对应的 WOE 值以及滚动变量对评分卡模型进行调整。采用本发明，可以通过滚动变量对评分卡模型进行调整，进而提高评分卡模型风险预测结果的准确性。

本发明实施例提供了一种评分卡模型调整装置，该装置包括用于执行前述如图 1 或者图 2 所述的方法的模块。具体地，参见图 3，是本发明实施例提供的一种装置的示意性框图。本实施例的装置包括：确定模块 30、获取模块 31 以及调整模块 32，其中：

确定模块 30，用于从所述评分卡模型的多个候选自变量中确定出至少一个高基数变量；

所述确定模块 30，还用于根据预设规则从所述至少一个高基数变量中确定出滚动变量；

获取模块 31，用于获取所述滚动变量下各个分组在预设时间内的参数信息；

所述确定模块 30，还用于根据所述获取模块获取到的所述参数信息确定

所述各个分组各自对应的证据权重 WOE 值;

调整模块 32,用于根据所述滚动变量下各个分组各自对应的 WOE 值以及所述滚动变量对所述评分卡模型进行调整。

在一个实施例中，所述确定模块 30，具体用于:

计算所述评分卡模型的多个候选自变量中各个候选自变量各自对应的信息量 IV 值，并输出所述各个候选自变量各自对应的 IV 值;

获取用户根据所述各个候选自变量各自对应的 IV 值输入的用于确定高基数变量的指示信息;

根据所述指示信息从所述多个候选自变量中确定出至少一个高基数变量。

在一个实施例中，所述确定模块 30，具体用于:

计算所述评分卡模型的多个候选自变量中的各个候选变量各自对应的 IV 值，并将所述 IV 值大于预设 IV 阈值的候选自变量确定为目标变量，所述目标变量下包括至少一个分组;

获取所述目标变量下各个分组各自对应的 WOE 值;

如果各个第一差值中大于预设 WOE 差值阈值的数量满足预设高基数条件，则将所述目标变量确定为高基数变量，所述第一差值为任意两个分组各自对应的 WOE 值之间的差值。

所述确定模块 30，具体用于：获取所述至少一个高基数变量中的各个高基数变量对应的各个分组在周期内的数据变化信息;

如果所述各个分组的数据变化信息满足预设数据变化条件，则将对应的高基数变量确定为滚动变量。

在一个实施例中，所述评分卡模型是基于线性回归模型建立的，所述线性回归模型由至少一个自变量以及所述至少一个自变量中各个自变量各自对应的权重系数组成，所述调整模块 32，具体用于：在所述评分卡模型对应的所述线性回归模型中增加所述滚动变量;根据所述滚动变量下各个分组各自对应的 WOE 值确定所述滚动变量的值。

在一个实施例中，所述获取模块 31，具体用于:

统计周期内所述至少一个高基数变量中的各个高基数变量对应各个分组下的数值和/或坏账率;

根据统计结果确定出所述周期内所述各个高基数变量对应各个分组下的数值变化信息和/或坏账率变化信息;

基于所述数值变化信息和/或所述坏账率变化信息生成所述各个高基数变量对应各个分组在所述周期内的数据变化信息。

在一个实施例中，所述数据变化信息包括以下至少一种：所述高基数变量对应各个分组下的数值变化信息和所述高基数变量对应各个分组下的坏账率变化信息，所述确定模块 30 还用于：如果所述数值变化信息所指示的数值变化率大于或者等于预设数值变化率阈值，或者，所述坏账率变化信息所指示的坏账变化率大于或者等于预设坏账变化率阈值，则确定所述各个分组的所述数据变化信息满足预设数据变化条件。

可以理解的是，本实施例的评分卡模型调整装置的各功能模块、单元的功能可根据上述方法实施例中的方法具体实现，其具体实现过程可以参照上述方法实施例的相关描述，此处不再赘述。

本发明实施例中，确定模块 30 从评分卡模型的多个候选自变量中确定出至少一个高基数变量，并根据预设规则从至少一个高基数变量中确定出滚动变量，获取模块 31 获取滚动变量下各个分组在预设时间内的参数信息，确定模块 30 根据获取模块获取到的参数信息确定各个分组各自对应的证据权重 WOE 值，调整模块 32 根据滚动变量下各个分组各自对应的 WOE 值以及滚动变量对评分卡模型进行调整。采用本发明，可以通过滚动变量对评分卡模型进行调整，进而提高评分卡模型风险预测结果的准确性。

参见图 4，是本申请实施例提供的一种服务器的示意性框图。如图所示的本实施例中的服务器可以包括：一个或多个处理器 401；一个或多个存储装置 402。上述处理器 401、存储装置 402 通过总线连接。存储装置 402 用于存储计算机程序，计算机程序包括程序指令，处理器 401 用于执行存储装置 402 存储的程序指令。其中，处理器 401 被配置用于调用程序指令执行:

针对所述评分卡模型选取第一因变量和第二因变量，所述第一因变量和所述第二因变量属于同一维度;

从所述评分卡模型的多个候选自变量中确定出至少一个高基数变量;

根据预设规则从所述至少一个高基数变量中确定出滚动变量;

获取所述滚动变量下各个分组在预设时间内的参数信息,并根据所述参数信息确定所述各个分组各自对应的证据权重 WOE 值;

根据所述滚动变量下各个分组各自对应的WOE值以及所述滚动变量对所述评分卡模型进行调整。

在一个实施例中,处理器 401,可以用于计算所述评分卡模型的多个候选自变量中各个候选自变量各自对应的信息量 IV 值,并输出所述各个候选自变量各自对应的 IV 值;获取用户根据所述各个候选自变量各自对应的 IV 值输入的用于确定高基数变量的指示信息;根据所述指示信息从所述多个候选自变量中确定出至少一个高基数变量。

在一个实施例中,处理器 401,还可以用于计算所述评分卡模型的多个候选自变量中的各个候选变量各自对应的 IV 值,并将所述 IV 值大于预设 IV 阈值的候选自变量确定为目标变量,所述目标变量下包括至少一个分组;获取所述目标变量下各个分组各自对应的 WOE 值;如果各个第一差值中大于预设 WOE 差值阈值的数量满足预设高基数条件,则将所述目标变量确定为高基数变量,所述第一差值为任意两个分组各自对应的 WOE 值之间的差值。

在一个实施例中,处理器 401,还可以用于获取所述至少一个高基数变量中的各个高基数变量对应的各个分组在周期内的数据变化信息;如果所述各个分组的数据变化信息满足预设数据变化条件,则将对应的高基数变量确定为滚动变量。

在一个实施例中,所述评分卡模型是基于线性回归模型建立的,所述线性回归模型由至少一个自变量以及所述至少一个自变量中各个自变量各自对应的权重系数组成,处理器 401 还可以用于,在所述评分卡模型对应的所述线性回归模型中增加所述滚动变量;根据所述滚动变量下各个分组各自对应的 WOE 值确定所述滚动变量的值。

在一个实施例中,处理器 401,还可以用于统计周期内所述至少一个高基数变量中的各个高基数变量对应各个分组下的数值和/或坏账率;根据统计结果确定出所述周期内所述各个高基数变量对应各个分组下的数值变化信息和/或坏账率变化信息;基于所述数值变化信息和/或所述坏账率变化信息生成所

述各个高基数变量对应各个分组在所述周期内的数据变化信息。

在一个实施例中，所述数据变化信息包括以下至少一种：所述高基数变量对应各个分组下的数值变化信息和所述高基数变量对应各个分组下的坏账率变化信息，所述处理器401，还可以如果所述数值变化信息所指示的数值变化率大于或者等于预设数值变化率阈值，或者，所述坏账率变化信息所指示的坏账变化率大于或者等于预设坏账变化率阈值，则确定所述各个分组的所述数据变化信息满足预设数据变化条件。

应当理解，在本申请实施例中，所称处理器 401 可以是中央处理单元(Central Processing Unit，CPU)，该处理器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。通用处理器可以是微处理器或者该处理器也可以是任何常规的处理器等。

该存储装置 402 可以包括只读存储器和随机存取存储器，并向处理器 401 提供指令和数据。存储装置 402 的一部分还可以包括非易失性随机存取存储器。例如，存储装置 402 还可以存储设备类型的信息。

具体实现中，本申请实施例中所描述的处理器 401 可执行本申请实施例图 1 和图 2 提供的评分卡模型调整方法的实施例和图 3 所描述的评分卡模型调整装置的实现方式，在此不再赘述。

本发明实施例中还提供一种计算机可读存储介质，所述计算机可读存储介质存储有计算机程序，所述计算机程序包括程序指令，所述程序指令被处理器执行时，可执行上述如图 1 或者图 2 所述方法实施例中服务器所执行的步骤。

本领域普通技术人员可以理解，以上所揭露的仅为本发明较佳实施例而已，当然不能以此来限定本发明之权利范围，因此依本发明权利要求所作的等同变化，仍属本发明所涵盖的范围。

权 利 要 求

1、一种评分卡模型的调整方法，其特征在于，包括:

从所述评分卡模型的多个候选自变量中确定出至少一个高基数变量;

根据预设规则从所述至少一个高基数变量中确定出滚动变量;

获取所述滚动变量下各个分组在预设时间内的参数信息，并根据所述参数信息确定所述各个分组各自对应的证据权重 WOE 值;

根据所述滚动变量下各个分组各自对应的WOE值以及所述滚动变量对所述评分卡模型进行调整。

2、根据权利要求 1 所述的方法，其特征在于，所述从所述评分卡模型的多个候选自变量中确定出至少一个高基数变量，包括:

计算所述评分卡模型的多个候选自变量中各个候选自变量各自对应的信息量 IV 值，并输出所述各个候选自变量各自对应的 IV 值;

获取用户根据所述各个候选自变量各自对应的 IV 值输入的用于确定高基数变量的指示信息;

根据所述指示信息从所述多个候选自变量中确定出至少一个高基数变量。

3、根据权利要求 1 所述的方法，其特征在于，所述从所述评分卡模型的多个候选自变量中确定出至少一个高基数变量，包括:

计算所述评分卡模型的多个候选自变量中的各个候选变量各自对应的 IV 值，并将所述 IV 值大于预设 IV 阈值的候选自变量确定为目标变量，所述目标变量下包括至少一个分组;

获取所述目标变量下各个分组各自对应的 WOE 值;

如果各个第一差值中大于预设 WOE 差值阈值的数量满足预设高基数条件，则将所述目标变量确定为高基数变量，所述第一差值为任意两个分组各自对应的 WOE 值之间的差值。

4、根据权利要求 1-3 任一项所述的方法，其特征在于，所述根据预设规

则从所述至少一个高基数变量中确定出滚动变量，包括:

获取所述至少一个高基数变量中的各个高基数变量对应的各个分组在周期内的数据变化信息;

如果所述各个分组的数据变化信息满足预设数据变化条件，则将对应的高基数变量确定为滚动变量。

5、根据权利要求 1 所述的方法，其特征在于，所述评分卡模型是基于线性回归模型建立的，所述线性回归模型由至少一个自变量以及所述至少一个自变量中各个自变量各自对应的权重系数组成，所述根据所述滚动变量下各个分组各自对应的 WOE 值以及所述滚动变量对所述评分卡模型进行调整，包括:

在所述评分卡模型对应的所述线性回归模型中增加所述滚动变量;

根据所述滚动变量下各个分组各自对应的 WOE 值确定所述滚动变量的值。

6、根据权利要求 4 所述的方法，其特征在于，所述获取所述至少一个高基数变量中的各个高基数变量对应的各个分组在周期内的数据变化信息，包括:

统计周期内所述至少一个高基数变量中的各个高基数变量对应各个分组下的数值和/或坏账率;

根据统计结果确定出所述周期内所述各个高基数变量对应各个分组下的数值变化信息和/或坏账率变化信息;

基于所述数值变化信息和/或所述坏账率变化信息生成所述各个高基数变量对应各个分组在所述周期内的数据变化信息。

7、根据权利要求 4 或 6 所述的方法，其特征在于，所述数据变化信息包括以下至少一种: 所述高基数变量对应各个分组下的数值变化信息和所述高基数变量对应各个分组下的坏账率变化信息，所述方法还包括:

如果所述数值变化信息所指示的数值变化率大于或者等于预设数值变化率阈值，或者，所述坏账率变化信息所指示的坏账变化率大于或者等于预设坏

账变化率阈值，则确定所述各个分组的所述数据变化信息满足预设数据变化条件。

8、一种评分卡模型调整装置，其特征在于，包括:

确定模块，用于从所述评分卡模型的多个候选自变量中确定出至少一个高基数变量;

所述确定模块，还用于根据预设规则从所述至少一个高基数变量中确定出滚动变量;

获取模块，用于获取所述滚动变量下各个分组在预设时间内的参数信息;

所述确定模块，还用于根据所述获取模块获取到的所述参数信息确定所述各个分组各自对应的证据权重 WOE 值;

调整模块，用于根据所述滚动变量下各个分组各自对应的 WOE 值以及所述滚动变量对所述评分卡模型进行调整。

9、一种服务器，其特征在于，包括处理器和存储装置，所述处理器和所述存储装置相互连接，其中，所述存储装置用于存储计算机程序，所述计算机程序包括程序指令，所述处理器被配置用于调用所述程序指令，执行如权利要求 1-7 任一项所述的方法。

10、一种计算机可读存储介质，其特征在于，所述计算机可读存储介质存储有计算机程序，所述计算机程序包括程序指令，所述程序指令当被处理器执行时使所述处理器执行如权利要求 1-7 任一项所述的方法。