Adjusting Method and Adjusting Device, Server and Storage Medium for Scorecard Model

Field of the invention

[0001] The present invention relates to the technical field of computers, in particular to an adjusting method and an adjusting device, server and storage medium for a scorecard model.

Background of the invention

[0002] At present, after traditional scorecard models are established, various dimensions (namely variables), coefficients of the dimensions and encoded values of weight of evidence (WOE) corresponding to the dimensions are fixed, and the models cannot be adjusted later. However, for some rolling variables with high cardinality and frequent changes in data of various groups of the variables, it is difficult to select such rolling variables into the models through information value (IV) indexes in the screening stage of the traditional scorecard models, so that the accuracy of risk prediction results of the scorecard models is seriously affected.

Summary of invention

[0003] The present invention provides an adjusting method and an adjusting device, server and storage medium for a scorecard model, a rolling variable can be selected into the scorecard model, and the scorecard model can be adjusted by utilizing the rolling variable, so that the accuracy of risk prediction results of the scorecard model is advantageously improved.

[0004] In the first aspect, the present invention provides an adjusting method for a scorecard model, comprising:

[0005] determining at least one high cardinality variable from multiple candidate independent variables of the scorecard model;

[0006] determining a rolling variable from the at least one high cardinality variable according to a preset rule;

[0007] acquiring parameter information of various groups of the rolling variable in a preset time and determining weight of evidence (WOE) values corresponding to the various groups according to the parameter information; and

[0008] adjusting the scorecard model according to the WOE values corresponding to the various groups of the rolling variable and the rolling variable.

[0009] In an embodiment, the specific mode of determining at least one high cardinality variable from multiple candidate independent variables of the scorecard model is:

[0010] calculating information values (IVs) corresponding to various candidate independent variables in the multiple candidate independent variables of the scorecard model, and outputting the IVs corresponding to the various candidate independent variables;

[0011] acquiring instruction information input by a user according to the IVs corresponding to the various candidate independent variables for determining the high cardinality variables; and

[0012] determining at least one high cardinality variable from the multiple candidate independent variables according to the instruction information.

[0013] In an embodiment, the specific mode of determining at least one high cardinality variable from multiple candidate independent variables of a scorecard model is:

[0014] calculating the IVs corresponding to various candidate independent variables in the multiple candidate independent variables of the scorecard model, and determining the variables whose IVs are greater than a preset IV threshold as target variables, wherein each target variable is divided into at least one group;

[0015] acquiring the WOE values corresponding to various groups of each target variable; and

[0016] determining the target variables as the high cardinality variables if the number of first differences greater than a preset WOE difference threshold meets a preset high cardinality condition, wherein each first difference is a difference between the WOE values corresponding to any two groups.

[0017] In an embodiment, each high cardinality variable is divided into at least one group, and the specific mode of determining a rolling variable from the at least one high cardinality variable according to the preset rule is:

[0018] acquiring data change information of various groups of each high cardinality variable in the at least one high cardinality variable in a period; and determining the corresponding high cardinality variable as the rolling variable if the data change information of the various groups meets preset data change conditions.

[0019] In an embodiment, the scorecard model is established based on a linear regression

model, and the linear regression model is composed of at least one variable and weight coefficients corresponding to various variables in the at least one variable. The specific mode of adjusting the scorecard model according to the WOE values corresponding to the various groups and the rolling variable is:

[0020] adding the rolling variable into the linear regression model corresponding to the scorecard model; and

[0021] determining the value of the rolling variable according to the WOE values corresponding to the various groups of the rolling variable.

[0022] In an embodiment, the specific mode of acquiring data change information of various groups of each high cardinality variable in the at least one high cardinality variable in a period is:

[0023] carrying out statistics on values and/or bad debt rates of various groups of each high cardinality variable in the at least one high cardinality variable in the period;

[0024] determining value change information and/or bad debt rate change information of various groups of each high cardinality variable in the period according to statistical results; and

[0025] generating data change information of various groups of each high cardinality variable in the period based on the value change information and/or the bad debt rate change information.

[0026] In an embodiment, the data change information comprises at least one of the following information: the value change information of various groups of each high cardinality variable and the bad debt rate change information of various groups of each high cardinality variable, and if a value change rate indicated by the value change information of the various groups is greater than or equal to a preset value change rate threshold or a bad debt change rate indicated by the bad debt rate change information of the various groups is greater than or equal to a preset bad debt change rate threshold, it is determined that the data change information of the various groups meets preset data change conditions.

[0027] In a second aspect, the present invention provides an adjusting device for a scorecard model, comprising:

[0028] a determining module, used for determining at least one high cardinality variable from multiple candidate independent variables of a scorecard model;

[0029] wherein the determining module is further used for determining a rolling variable from the at least one high cardinality variable according to a preset rule, and the rolling variable is

divided into at least one group;

[0030] an acquiring module, used for acquiring parameter information of various groups of the rolling variable in a preset time;

[0031] wherein the determining module is further used for determining weight of evidence (WOE) values corresponding to the various groups according to the parameter information acquired by the acquiring module; and

[0032] an adjusting module, used for adjusting the scorecard model according to the WOE values corresponding to the various groups of the rolling variable and the rolling variable.

[0033] In a third aspect, the present invention provides a server which comprises a processor and a storage device, and the processor and the storage device are connected to each other, wherein the storage device is used for storing a computer program, the computer program includes program instructions, and the processor is configured to call the program instructions to execute the method in the first aspect described above.

[0034] In a fourth aspect, the present invention provides a computer readable storage medium which stores a computer program, the computer program includes program instructions, and a processor is enabled to execute the method in the first aspect described above when the program instructions are executed by the processor.

[0035] In an embodiments of the present invention, the server determines the at least one high cardinality variable from the multiple candidate independent variables of the scorecard model, determines the rolling variable from the at least one high cardinality variable according to the preset rule, acquires the parameter information of various groups of the rolling variable in the preset time, determines the weight of evidence (WOE) values corresponding to the various groups according to the parameter information, and then adjusts the scorecard model according to the WOE values corresponding to the various groups and the rolling variable. By adopting the present invention, the rolling variable can be selected into the model, and the scorecard model can be adjusted by utilizing the rolling variable, therefore the accuracy of score results of the scorecard model is improved.

BRIEF DESCRIPTION OF THE DRAWINGS

[0036] In order to explain the embodiments of the present invention or the technical solutions

in the prior art more clearly, the accompanying drawings required to be used in the embodiments will be briefly introduced below. Apparently, the accompanying drawings in the following description are only some embodiments of the present invention, and those of ordinary skill in the art can obtain other accompanying drawings based on these accompanying drawings without creative efforts.

[0037] Fig. 1 is a flowchart of an adjusting method for a scorecard model provided by an embodiment of the present invention;

[0038] Fig. 2 is a flowchart of another adjusting method for a scorecard model provided by an embodiment of the present invention;

[0039] Fig. 3 is a schematic block diagram of an adjusting device for a scorecard model provided by an embodiment of the present invention; and

[0040] Fig. 4 is a schematic block diagram of a server provided by an embodiment of the present invention.

DETAILED DESCRIPTION OF ILLUSTRATED EMBODIMENTS

[0041] The technical solutions in the embodiments of the present invention will be described clearly and completely in conjunction with the accompany drawings in the embodiments of the present invention below. Apparently, the described embodiments are only a part of the embodiments of the present invention, but not all the embodiments. Based on the embodiments of the present invention, all other embodiments obtained by those of ordinary skill in the art without creative efforts fall within the protection scope of the present invention.

[0042] A scorecard model is used as a prediction method which can be applied to different application scenarios in combination with different business data. Exemplarily, when a scorecard model is a credit scorecard model, the credit scorecard model can describe factors affecting the individual credit level based on the analysis of a large number of credit records of credit cardholders in the past, thereby helping lending institutions to issue consumer credit. The establishment of the credit scorecard model is mainly focused on adopting applicant characteristic variables to predict the default probability of applicants, which requires that the characteristic variables entering the credit scorecard model have high predictive ability.

[0043] In an embodiments of the present invention, an information value (IV) can be adopted

to measure the predictive ability of each variable, wherein the correspondence between the IV and the predictive ability can be as shown in Table 1-1.

[0044] Table 1-1

IV	Predictive ability
<0.03	None
0.03-0.1	Low
0.1-0.2	Medium
0.2-0.3	High
>0.3	Extremely high

[0045] In an embodiment, the scorecard model can be established based on a linear regression model, wherein the linear regression model is equivalent to a relationship established between a dependent variable (y) and one or more independent variables (x) and is expressed as:

[0046] $y = a + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + ... + \beta_n x_n$

[0047] wherein, a represents the intercept, x_n (n is a positive integer) is an independent variable selected into the model, namely an in-model index, and β_n is a coefficient corresponding to each independent variable.

[0048] For the traditional scorecard models, after the models are established, each independent variable x_n, the coefficient β_n corresponding to each independent variable, and a WOE encoded value corresponding to each independent variable are fixed, and the models cannot be adjusted later. However, for some rolling variables with high cardinality and frequent changes in data of various groups of the variables, it is difficult to select such rolling variables into the models through information value (IV) indexes during the model screening stage, but due to the characteristic of frequent change, such rolling variables are often the key variables affecting the risk prediction results. Therefore, the risk prediction results of traditional scorecard models are usually not accurate enough.

[0049] In the present invention, by determining at least one high cardinality variable from multiple candidate independent variables of a scorecard model, determining a rolling variable x_{n+1} from the at least one high cardinality variable according to a preset rule, acquiring parameter information of various groups of the rolling variable in a preset time, determining weight of evidence (WOE) values corresponding to the various groups according to the

parameter information, then selecting the rolling variable x_{n+1} into the scorecard model, and determining a coefficient β_{n+1} corresponding to the rolling variable according to the WOE values corresponding to the various groups of the rolling variable, the accuracy of the risk prediction results of the scorecard model can be improved. Exemplarily, for a credit scorecard model, improving of the accuracy of the risk prediction results can assist lending institutions in issuing consumer credit, and thereby effectively controlling the repayment overdue situation of borrowers.

[0050] Wherein, the high cardinality variable described in the embodiment of the present invention can be a variable to which multiple groups belong. For example, the variable is a province, and multiple groups belong to the province, such as: Sichuan Province, Guangxi Province, Jiangsu Province, Guangdong Province, Hainan Province and Liaoning Province. In this case, the province variable can be determined as the high cardinality variable. The described rolling variable can be a high cardinality variable with values and/or bad debt rates changing frequently of various groups.

[0051] In an embodiment, each candidate independent variable can be divided into m groups (m is an integer greater than 0), and the IV corresponding to the candidate independent variable satisfies the following formula 1.1:

[0052] $$IV = \sum_i^m IV_i$$

[0053] wherein, i is a positive integer less than m and indicates the i-th group in the m groups; and IV_i indicates the IV corresponding to the i-th group. That is, the IV of each candidate independent variable is obtained by summing the IVs corresponding to various groups of the independent variable. In the embodiment of the present invention, the specific value of IV_i can be determined according to the WOE value (namely WOE_i) of the i-th group and can specifically adopt the following formula 1.2:

[0054] $$IV_i = \left(\left(G_i / G_T \right) - \left(B_i / B_T \right) \right) * WOE_i$$

[0055] wherein, G_i in the above formula represents the number of responding customers in this group, G_T represents the number of all responding customers in samples, B_i represents the number of non-responding customers in this group, and B_T represents the number of all non-responding customers in the samples. It can be seen from the above formula that WOE

actually represents the difference between “the proportion of responding customers in the current group to all responding customers” and “the proportion of non-responding customers in the current group to all non-responding customers”, and the calculation formula of WOE_i can adopt the following formula 1.3:

[0056]
$$WOE = \ln \frac{\left(G_i / G_T \right)}{\left(B_i / B_T \right)}$$

[0057] wherein, the above responding customers refer to individuals whose predictive variable values in the scorecard model are “yes” or “1”. For example, in a risk scorecard model, the above non-responding customers correspond to default customers, which is not specifically limited in the present invention.

[0058] Referring to Fig. 1, Fig. 1 is a flowchart of an adjusting method for a scorecard model provided by an embodiment of the present invention. As shown in Fig. 1, the adjusting method for the scorecard model can comprise:

[0059] 101, determining, by a server, at least one high cardinality variable from multiple candidate independent variables of a scorecard model.

[0060] In an embodiment, the server can calculate information values (IVs) corresponding to various candidate independent variables in the multiple candidate independent variables of the scorecard model, and output the IVs corresponding to the various candidate variables to acquire instruction information which is input by a user according to the IVs corresponding to the various variables for determining high cardinality variables, and then determine at least one high cardinality variable from the multiple candidate independent variables according to the instruction information.

[0061] Wherein, the instruction information is information generated according to user instructions and is used for instructing the server to determine at least one high cardinality variable from multiple candidate independent variables. For example, the server outputs IVs corresponding to j (j is a positive integer) candidate independent variables, that is, the server outputs j IVs (such as IV_1, IV_2, IV_3, …, IV_j). In this case, if a user wants to determine the candidate independent variables corresponding to IV_1 and IV_2 as high cardinality variables after viewing the j IVs, instruction information can be input for IV_1 and IV_2 for instructing the server

to determine the candidate independent variables corresponding to IV_1 and IV_2 as the high cardinality variables. In this case, the server can determine the candidate independent variables corresponding to IV_1 and IV_2 as high cardinality variables after receiving the instruction information.

[0062] Exemplarily, assuming that the scorecard model includes j (j is a positive integer) candidate independent variables, the server can calculate the IV corresponding to each candidate independent variable through formulas 1.1 to 1.3, and display the calculated j IVs (such as IV_1, IV_2, IV_3, ..., IV_j) on a display interface. After viewing the j IVs displayed on the display interface, a user can input instruction information for instructing the server to determine one or more IVs in j IVs as target IVs (such as IV_1 and IV_2). Further, after receiving the instruction information of the user, the server can determine one or more target IVs from the j IVs according to the instruction information and find the candidate independent variables corresponding to the one or more target IVs so as to determine the corresponding candidate independent variables as high cardinality variables.

[0063] In an embodiment, the server can also calculate the IVs corresponding to each candidate independent variable in the multiple candidate independent variables of the scorecard model, determine the variables whose IVs are greater than a preset IV threshold as target variables, and then acquire WOE values corresponding to various groups of each target variable, and if the number of first differences greater than a preset WOE difference threshold meets a preset high cardinality condition, the target variables are determined as the high cardinality variables. Each first difference is a difference between the WOE values corresponding to any two groups.

[0064] In an embodiment, the preset high cardinality condition is that the number of first differences greater than the preset WOE difference threshold is greater than or equal to a preset number threshold r_0 (r_0 is a positive integer), and the scorecard model includes j (j is a positive integer) candidate independent variables. In this case, the server can calculate the IV corresponding to each candidate independent variable through information algorithms represented by formulas 1.1-1.3, that is, the server obtains j IVs (such as IV_1, IV_2, IV_3, ..., IV_j). Further, the j IVs can be compared with a preset IV threshold one by one to determine the IVs greater than the preset IV threshold as IV_1, then the candidate independent variable

corresponding to IV_1 is determined as the target variable, wherein, the target variable comprises r_1 (r_1 is a positive integer) groups. Further, the server can calculate the WOE values corresponding to various groups of the target variable according to formula 1.3 to acquire r_1 WOE values, then further calculate the difference between every two r_1 WOE values (namely the first difference), compare all the acquired first differences with the preset WOE difference threshold, and determine the target variable as a high cardinality variable when it is determined that there are b first differences greater than the preset WOE difference threshold and b is greater than r_0.

[0065] In an embodiment, when the server determines the high cardinality variables from the multiple candidate independent variables of the scorecard model, the server can also directly adopt formula 1.3 to calculate the WOE values of various groups of any candidate independent variable in the scorecard model, compare the difference between every two WOE values, and determine the differences greater than the preset difference threshold as target differences and further determine the number of the target differences, and if the number of the target differences is greater than or equal to the number threshold, then the any candidate independent variable can be determined as the high cardinality variable.

[0066] 102, determining, by the server, a rolling variable from the at least one high cardinality variable according to a preset rule.

[0067] In an embodiment, after determining the at least one high cardinality variable, the server can acquire data change information of one or more groups of any high cardinality variable in a certain period. The data change information can comprise at least one of value change information of various groups and bad debt rate change information of various groups. Further, the server can determine whether the value change information of various groups and/or the bad debt rate change information of various groups meet preset data change conditions, and if yes, the any high cardinality variable can be determined as a rolling variable.

[0068] 103, acquiring, by the server, parameter information of various groups of the rolling variable in a preset time and determines weight of evidence (WOE) values corresponding to the various groups according to the parameter information.

[0069] 104, adjusting, by the server, the scorecard model according to the WOE values corresponding to the various groups of the rolling variable and the rolling variable.

[0070] Wherein, the preset time is a time period which can correspond to start and end dates such as May 2018-June 2018, or can start with the current time to push back 10 days, 15 days or 1 month. The time period can be set by default in a system or can be determined according to user instructions, which is not specifically limited in the present invention.

[0071] In an embodiment, the parameter information is bad debt rate information of various groups of the rolling variable in the preset time, assuming that the rolling variable determined in step 102 is divided into r1 groups, and the preset time is the month of May 2018. In this case, the server can acquire bad debt rates of the r1 groups in the month of May 2018, determine the WOE values corresponding to the various groups according to the bad debt rates, and then adjust the scorecard model according to the rolling variable and the WOE values corresponding to the various groups.

[0072] In an embodiment, the above-mentioned scorecard model is established based on a linear regression model which is composed of at least one independent variable and weight coefficients corresponding to various independent variables in the at least one independent variable. In this case, the specific implementation mode of the server performing step 104 can be: adding a rolling variable into the linear regression model corresponding to the scorecard model, determining the value of the rolling variable according to the WOE values corresponding to various groups of the rolling variable, and then adjusting the linear regression model, that is, adjusting the scorecard model.

[0073] Exemplarily, assuming that the scorecard model is used for predicting the repayment overdue situation of borrowers in the three provinces of Guangxi, Jiangsu and Sichuan. The scorecard model is established based on a linear regression model, namely $y = a + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \ldots + \beta_n x_n$, wherein a represents the intercept, x_n (n is a positive integer) is an independent variable selected into the model, and β_n is a coefficient corresponding to each independent variable. A high cardinality variable is a province variable, and the province variable is divided into the three groups of Guangxi, Jiangsu and Sichuan, the preset time is the month of May 2018, the bad debt rate information of various groups of the province variable in May 2018 is shown in Table 1-2, wherein G represents the number of bad debts, and B represents the number of non-bad debts.

[0074] Table 1-2

Province	G	B	Sum	Bad debt rate
Guangxi	400	100	500	20%
Jiangsu	300	200	500	40%
Sichuan	300	200	500	40%
Sum	1000	500	1500	33%

[0075] Further, after acquiring the bad debt rate information shown in Table 1-2, the server can determine the WOE values of the three groups of Guangxi, Jiangsu and Sichuan of the province variable according to formula 1.3 as:

[0076] $\ln\frac{\left(400/1000\right)}{\left(100/500\right)}=0.69;\ \ln\frac{\left(300/1000\right)}{\left(200/500\right)}=-0.287;\ \ln\frac{\left(300/1000\right)}{\left(200/500\right)}=-0.287;$

[0077] Then, the server can express the rolling variable, namely the province as x_{prov} and select the rolling variable into the linear regression model, that is, one rolling variable x_{prov} is added to the above linear regression model. The linear regression model after addition of the rolling variable is: $y=a+\beta_1 x_1+\beta_2 x_2+\beta_3 x_3+...+\beta_n x_n+\beta_{n+1}x_{prov}$, when the server predicts the repayment overdue situation in Guangxi Province through this model, the value of x_{prov} is the WOE value corresponding to Guangxi, namely 0.69; when the server predicts the repayment overdue situation in Jiangsu Province through this model, the value of x_{prov} is the WOE value corresponding to Jiangsu, namely -0.287; when the server predicts the repayment overdue situation in Sichuan Province through this model, the value of x_{prov} is the WOE value corresponding to Sichuan, namely -0.287, thus, the linear regression model is adjusted, that is, the scorecard model is adjusted, and the accuracy of the risk prediction results of the scorecard model is improved.

[0078] In the embodiment of the present invention, the server determines the at least one high cardinality variable from the multiple candidate independent variables of the scorecard model, determines the rolling variable from the at least one high cardinality variable according to a preset rule, acquires the parameter information of various groups in at least one group of the rolling variable in the preset time, determines the weight of evidence (WOE) values corresponding to various groups according to the parameter information, and then adjusts the scorecard model according to the WOE values corresponding to the various groups and the rolling variable. By adopting the present invention, the scorecard model can be adjusted by

utilizing the rolling variable, therefore the accuracy of the risk prediction results of the scorecard model is improved.

[0079] Referring to Fig. 2 then, Fig. 2 is a flowchart of another scorecard model adjusting method provided by an embodiment of the present invention. As shown in Fig. 2, the adjusting method for the scorecard model can comprise:

[0080] 201, determining, by the server, at least one high cardinality variable from multiple candidate independent variables of a scorecard model.

[0081] Wherein the specific mode of step 201 can refer to the related description of step 101 in the foregoing embodiment, which is not described in detail herein.

[0082] 202, acquiring, by the server, data change information of various groups of each high cardinality variable in the at least one high cardinality variable in a period.

[0083] Wherein, the period can be a time period which can correspond to start and end dates such as May 2018-June 2018, or can start with the current time to push back 10 days, 15 days or 1 month. The specific time period corresponding to the period can be set by default by a system or can be determined according to user instructions. Wherein, the data change information can be value change information and/or bad debt rate change information of various groups of each high cardinality variable.

[0084] In an embodiment, the server can carry out statistics on the values and/or the bad debt rates of various groups of each high cardinality variable in the at least one high cardinality variable in the period, determine the value change information and/or the bad debt rate change information of various groups of each high cardinality variable in the period according to statistical results, and then generate the data change information of various groups of each high cardinality variable in the period based on the value change information and/or the bad debt rate change information.

[0085] During specific implementation, the server can acquire the values and/or the bad debt rates of various groups of each high cardinality variable in the at least one high cardinality variable at a preset time interval in the period, that is, each time interval corresponds to an acquiring time node, then determine the value change information and/or the bad debt rate change information of various groups of each high cardinality variable in the period by carrying out statistics on the values and/or the bad debt rates, obtained in the period, of various groups at

each time node, and generate the data change information of various groups of each high cardinality variable in the period based on the value change information and/or the bad debt rate change information.

[0086] Exemplarily, assuming that the above period is the month of April 2018, the preset time interval is 15 days, and the scorecard model is used for predicting the overdue situation of more than 60 days in any period in the month of April 2018; a certain high cardinality variable x_1 in at least one high cardinality variable is the age of borrowers, according to the characteristics of age, the high cardinality variable, namely the age can be divided into multiple groups such as 18-25, 25-40 and 40-65; and data were acquired in the month of April 2018 twice in total, first data were acquired on April 15, 2018, and the acquired data are data statistical results under the groups of x_1 shown in Table 2-1; and second data were acquired on April 30, 2018, and the acquired data are data statistical results under the groups of x_1 shown in Table 2-2.

[0087] Table 2-1

April 15, 2018			
Age	Non-overdue	Overdue	Bad debt rate
18-25	200	300	0.60
25-40	100	400	0.80
40-65	300	200	0.67
Sum	600	900	0.60

[0088]

[0089] Table 2-2

April 30, 2018			
Age	Non-overdue	Overdue	Bad debt rate
18-25	300	200	0.67
25-40	400	100	0.2
40-65	200	300	0.60
Sum	900	600	0.4

[0090] After acquiring the data in Table 2-1 and Table 2-2, the server can determine that the bad debt change rate differences (namely the bad debt rate change information) under the three groups of 18-25, 25-40 and 40-65 in April 2018 are 0.07, 0.6 and 0.07 respectively by analyzing

the data recorded in Table 2-1 and Table 2-2, similarly, overdue value change differences under the three groups of 18-25, 25-40 and 40-65 are 100, 300 and 400 respectively, and non-overdue value change differences are 100, 300, and 100 respectively, wherein, the overdue value change differences and the non-overdue value change differences under the three groups form value change information of the three groups.

[0091] 203, determining, by the server, the corresponding high cardinality variable as a rolling variable if the server determines that the data change information of the various groups meets preset data change conditions.

[0092] In an embodiment, the data change information comprises at least one of the following information: value change information of various groups of each high cardinality variable and bad debt rate change information of various groups of each high cardinality variable. The aforementioned preset data change condition can be that a value change rate indicated by the value change information is greater than or equal to a preset value change rate threshold, or a bad debt change rate indicated by the bad debt rate change information is greater than or equal to a preset bad debt change rate threshold. Before performing step 203, the server can acquire the above-mentioned value change information and/or bad debt rate change information from the data change information, and determine the value change rate of various groups of each high cardinality variable according to the value change information, and determine the bad debt change rate of various groups of the high cardinality variable according to the bad debt rate change information. In an embodiment, the server can determine that the data change information of the various groups meets the preset data change conditions when the value change rate of various groups of the high cardinality variable is greater than or equal to the preset value change rate threshold. In another embodiment, the server can determine that the data change information of the various groups meets the preset data change conditions when the bad debt change rate of various groups of the high cardinality variable is greater than or equal to the preset bad debt change rate threshold. In another embodiment, the server can also determine that the data change information of the various groups meets the preset data change conditions when the value change rate of various groups of the high cardinality variable is greater than or equal to the preset value change rate threshold and the bad debt change rate of various groups of the high cardinality variable is greater than or equal to the preset bad debt change rate threshold.

[0093] 204, acquiring, parameter information of various groups of the rolling variable in the preset time and determines weight of evidence (WOE) values corresponding to the various groups according to the parameter information.

[0094] 205, adjusting, by the server, the scorecard model according to the WOE values corresponding to the various groups of the rolling variable and the rolling variable.

[0095] Wherein, the specific implementation modes of step 204 and step 205 can refer to the related description of step 103 and step 104 in the foregoing embodiment, which is not described in detail herein.

[0096] In the embodiment of the present invention, the server determines the at least one high cardinality variable from the multiple candidate independent variables of the scorecard model, acquires the data change information of various groups of each high cardinality variable in the at least one high cardinality variables in the period, then determines the corresponding high cardinality variable as the rolling variable if the server determines that the data change information of the various groups meets the preset data change conditions, acquires the parameter information of various groups of the rolling variable in the preset time, determines the weight of evidence (WOE) values corresponding to the various groups according to the parameter information, and then adjusts the scorecard model according to the WOE values corresponding to the various groups of the rolling variable and the rolling variable. By adopting the present invention, the scorecard model can be adjusted by utilizing the rolling variable, and therefore the accuracy of the risk prediction results of the scorecard model is improved.

[0097] An embodiment of the present invention provides an adjusting device for a scorecard model, and the device comprises modules for executing the method described in Fig. 1 or Fig. 2. Specifically, Fig. 3 is a schematic block diagram of a device according to an embodiment of the present invention. The device of the embodiment comprises: a determining module 30, an acquiring module 31 and an adjusting module 32, wherein:

[0098] the determining module 30 is used for determining at least one high cardinality variable from multiple candidate independent variables of a scorecard model;

[0099] the determining module 30 is further used for determining a rolling variable from the at least one high cardinality variable according to a preset rule;

[00100] the acquiring module 31 is used for acquiring parameter information of various groups

of the rolling variable in a preset time;

[00101] the determining module 30 is further used for determining weight of evidence (WOE) values corresponding to the various groups according to the parameter information acquired by the acquiring module; and

[00102] the adjusting module 32 is used for adjusting the scorecard model according to the WOE values corresponding to the various groups of the rolling variable and the rolling variable.

[00103] In an embodiment, the determining module 30 is specifically used for:

[00104] calculating information values (IVs) corresponding to various candidate independent variables in the multiple candidate independent variables of the scorecard model, and outputting the IVs corresponding to the various candidate independent variables;

[00105] acquiring instruction information input by a user according to the IVs corresponding to the various candidate independent variables for determining the high cardinality variables; and

[00106] determining at least one high cardinality variable from the multiple candidate independent variables according to the instruction information.

[00107] In an embodiment, the determining module 30 is specifically used for:

[00108] calculating the IVs corresponding to various candidate variables in the multiple candidate independent variables of the scorecard model, and determining the candidate independent variables whose IVs are greater than a preset IV threshold as target variables, and each target variable is divided into at least one group;

[00109] acquiring the WOE values corresponding to various groups of each target variable; and

[00110] determining the target variables as the high cardinality variables if the number of first differences greater than a preset WOE difference threshold meets a preset high cardinality condition, wherein each first difference is a difference between the WOE values corresponding to any two groups.

[00111] The determining module 30 is specifically used for: acquiring data change information of various groups of each high cardinality variable in the at least one high cardinality variable in a period; and

[00112] determining the corresponding high cardinality variable as a rolling variable if the data change information of the various groups meets preset data change conditions.

[00113] In an embodiment, the scorecard model is established based on a linear regression

model, and the linear regression model is composed of at least one independent variable and weight coefficients corresponding to various independent variables in the at least one independent variable. The adjusting module 32 is specifically used for: adding the rolling variable into the linear regression model corresponding to the scorecard model; and determining the value of the rolling variable according to the WOE values corresponding to various groups of the rolling variable.

[00114] In an embodiment, the acquiring module 31 is specifically used for:

[00115] carrying out statistics on values and/or bad debt rates of various groups of each high cardinality variable in the at least one high cardinality variable in the period;

[00116] determining value change information and/or bad debt rate change information of various groups of each high cardinality variable in the period according to statistical results; and

[00117] generating data change information of various groups of each high cardinality variable in the period based on the value change information and/or the bad debt rate change information.

[00118] In an embodiment, the data change information comprises at least one of the following information: the value change information of various groups of the high cardinality variable and the bad debt rate change information of various groups of the high cardinality variable. The determining module 30 is further used for: determining that the data change information of the various groups meets the preset data change conditions if the value change rate indicated by the value change information is greater than or equal to a preset value change rate threshold or the bad debt change rate indicated by the bad debt rate change information is greater than or equal to a preset bad debt change rate threshold.

[00119] It is understandable that the functions of functional modules and units of the scorecard model adjusting device of the embodiment can be achieved specifically according to the method in the above method embodiments, and the specific implementation process can refer to the related description of the above method embodiments, which is not described in detail herein.

[00120] In the embodiment of the present invention, the determining module 30 determines the at least one high cardinality variable from the multiple candidate independent variables of the scorecard model and determines the rolling variable from the at least one high cardinality variable according to the preset rule, the acquiring module 31 acquires the parameter information of various groups of the rolling variable in a preset time, the determining module 30 determines

the weight of evidence (WOE) values corresponding to the various groups according to the parameter information acquired by the acquiring module, and the adjusting module 32 adjusts the scorecard model according to the WOE values corresponding to the various groups of the rolling variable and the rolling variable. By adopting the present invention, the scorecard model can be adjusted by utilizing the rolling variable, therefore the accuracy of the risk prediction results of the scorecard model is improved.

[00121] Fig. 4 is a schematic block diagram of a server provided by an embodiment of the present invention. The server shown in Fig. 4 in the embodiment can comprise: one or more processors 401 and one or more storage devices 402. The aforementioned processor 401 and storage device 402 are connected by a bus. The storage device 402 is used for storing a computer program, and the computer program includes program instructions, and the processor 401 is used for executing the program instructions stored in the storage device 402. Wherein, the processor 401 is configured to call the program instructions to:

[00122] select a first dependent variable and a second dependent variable for the scorecard model, wherein the first dependent variable and the second dependent variable belong to the same dimension;

[00123] determine at least one high cardinality variable from multiple candidate independent variables of the scorecard model;

[00124] determine a rolling variable from the at least one high cardinality variable according to a preset rule;

[00125] acquire parameter information of various groups of the rolling variable in a preset time, and determine weight of evidence (WOE) values corresponding to the various groups according to the parameter information; and

[00126] adjust the scorecard model according to the WOE values corresponding to the various groups of the rolling variable and the rolling variable.

[00127] In an embodiment, the processor 401 can be used for calculating the information values (IVs) corresponding to various candidate independent variables in the multiple candidate independent variables of the scorecard model, and outputting the IVs corresponding to the various candidate independent variables; acquiring instruction information input by a user according to the IVs corresponding to the various candidate independent variables for

determining the high cardinality variables; and determining the at least one high cardinality variable from the multiple candidate independent variables according to the instruction information.

[00128] In an embodiment, the processor 401 can also be used for calculating the IVs corresponding to various candidate variables in the multiple candidate independent variables of the scorecard model, and determining the candidate independent variables whose IVs are greater than a preset IV threshold as target variables, wherein each target variable is divided into at least one group; acquiring the WOE values corresponding to various groups of each target variable; and determining the target variables as the high cardinality variables if the number of first differences greater than a preset WOE difference threshold meets a preset high cardinality condition, wherein each first difference is a difference between the WOE values corresponding to any two groups.

[00129] In an embodiment, the processor 401 can further be used for acquiring the data change information of various groups of each high cardinality variable in the at least one high cardinality variable in the period; and determining the corresponding high cardinality variable as the rolling variable if the data change information of the various groups meets preset data change conditions.

[00130] In an embodiment, the scorecard model is established based on a linear regression model, and the linear regression model is composed of at least one independent variable and weight coefficients corresponding to various independent variables in the at least one independent variable. The processor 401 can further be used for: adding the rolling variable into the linear regression model corresponding to the scorecard model; and determining the value of the rolling variable according to the WOE values corresponding to various groups of the rolling variable.

[00131] In an embodiment, the processor 401 can further be used for carrying out statistics on values and/or bad debt rates of various groups of each high cardinality variable in the at least one high cardinality variable in the period; determining value change information and/or bad debt rate change information of various groups of each high cardinality variable in the period according to statistical results; and generating data change information of various groups of each high cardinality variable in the period based on the value change information and/or the bad debt

rate change information.

[00132] In an embodiment, the data change information comprises at least one of the following information: the value change information of various groups of the high cardinality variable and the bad debt rate change information of various groups of the high cardinality variable. The processor 401 can further determine that the data change information of the various groups meets preset data change conditions if value change rate indicated by the value change information is greater than or equal to a preset value change rate threshold or bad debt change rate indicated by the bad debt rate change information is greater than or equal to a preset bad debt change rate threshold.

[00133] It should be understood that in the embodiment of the present invention, the processor 401 can be a central processing unit (CPU), and the processor can also be other general-purpose processors, digital signal processors (DSP), application specific integrated circuits (ASIC), field-programmable gate arrays (FPGA) or other programmable logic devices, discrete gates or transistor logic devices, discrete hardware components and the like. The general-purpose processors can be microprocessors, or the processors can also be any conventional processors or the like.

[00134] The storage device 402 can comprise a read-only memory and a random access memory, and provide instructions and data to the processor 401. A part of the storage device 402 can also comprise a non-volatile random access memory. For example, the storage device 402 can also store device type information.

[00135] During specific implementation, the processor 401 described in the embodiment of the present invention can execute the embodiments of the scorecard model adjusting method provided in Figs. 1 and 2 and the implementation of the scorecard model adjusting device described in Fig. 3, which is not described in detail herein.

[00136] An embodiment of the present invention further provides a computer readable storage medium, the computer readable storage medium stores a computer program, the computer program includes program instructions, and the steps executed by a server in the method embodiments described in Fig. 1 or Fig. 2 can be executed when the program instructions are executed by a processor.

[00137] Those of ordinary skill in the art can understand that only the preferred embodiments of

the present invention are disclosed above, and certainly cannot limit the scope of claims of the present invention. Therefore, equivalent changes made according to the claims of the present invention still fall within the scope of the present invention.

Claims

1. An adjusting method for a scorecard model, characterized by comprising:

determining at least one high cardinality variable from multiple candidate independent variables of the scorecard model;

determining a rolling variable from the at least one high cardinality variable according to a preset rule;

acquiring parameter information of various groups of the rolling variable in a preset time and determining WOE values corresponding to the various groups according to the parameter information; and

adjusting the scorecard model according to the WOE values corresponding to the various groups of the rolling variable and the rolling variable.

2. The method according to claim 1, characterized in that the step of determining at least one high cardinality variable from multiple candidate independent variables of the scorecard model comprises:

calculating IVs corresponding to various candidate independent variables in the multiple candidate independent variables of the scorecard model, and outputting the IVs corresponding to the various candidate independent variables;

acquiring instruction information input by a user according to the IVs corresponding to the various candidate independent variables for determining the high cardinality variables; and

determining at least one high cardinality variable from the multiple candidate independent variables according to the instruction information.

3. The method according to claim 1, characterized in that the step of determining at least one high cardinality variable from multiple candidate independent variables of the scorecard model comprises:

calculating the IVs corresponding to various candidate variables in the multiple candidate independent variables of the scorecard model, and determining the candidate independent variables whose IVs are greater than a preset IV threshold as target variables, wherein each target

variable is divided into at least one group;

acquiring the WOE values corresponding to various groups of the target variables; and

determining the target variables as the high cardinality variables if a number of first differences greater than a preset WOE difference threshold meets a preset high cardinality condition, wherein the first difference is a difference between the WOE values corresponding to any two groups.

4. The method according to any one of claims 1 to 3, characterized in that the step of determining a rolling variable from the at least one high cardinality variable according to a preset rule comprises:

acquiring data change information of various groups of each high cardinality variable in the at least one high cardinality variable in a period; and

determining the corresponding high cardinality variable as the rolling variable if the data change information of the various groups meets preset data change conditions.

5. The method according to claim 1, characterized in that the scorecard model is established based on a linear regression model, and the linear regression model is composed of at least one independent variable and weight coefficients corresponding to various independent variables in the at least one independent variable, and the step of adjusting the scorecard model according to the WOE values corresponding to various groups of the rolling variable and the rolling variable comprises:

adding the rolling variable into the linear regression model corresponding to the scorecard model; and

determining the value of the rolling variable according to the WOE values corresponding to various groups of the rolling variable.

6. The method according to claim 4, characterized in that the step of acquiring data change information of various groups of each high cardinality variable in the at least one high cardinality variable in a period comprises:

carrying out statistics on values and/or bad debt rates of various groups of each high cardinality variable in the at least one high cardinality variable in the period;

determining value change information and/or bad debt rate change information of various groups of each high cardinality variable in the period according to statistical results; and

generating data change information of various groups of each high cardinality variable in the period based on the value change information and/or the bad debt rate change information.

7. The method according to claim 4 or 6, characterized in that the data change information comprises at least one of the following information: the value change information of various groups of each high cardinality variable and the bad debt rate change information of various groups of each high cardinality variable, and the method further comprises:

determining that the data change information of the various groups meets a preset data change condition if a value change rate indicated by the value change information is greater than or equal to a preset value change rate threshold or a bad debt change rate indicated by the bad debt rate change information is greater than or equal to a preset bad debt change rate threshold.

8. An adjusting device for a scorecard model, characterized by comprising:

a determining module, used for determining at least one high cardinality variable from multiple candidate independent variables of the scorecard model;

wherein the determining module is further used for determining a rolling variable from the at least one high cardinality variable according to a preset rule;

an acquiring module, used for acquiring parameter information of various groups of the rolling variable in a preset time;

wherein the determining module is further used for determining WOE values corresponding to the various groups according to the parameter information acquired by the acquiring module; and

an adjusting module, used for adjusting the scorecard model according to the WOE values corresponding to various groups of the rolling variable and the rolling variable.

9. A server, characterized by comprising a processor and a storage device, wherein the processor and the storage device are connected to each other, the storage device is used for storing a computer program, the computer program includes program instructions, and the processor is configured to call the program instructions to execute the method according to any one of claims 1-7.

10. A computer readable storage medium, characterized in that the computer readable storage medium stores a computer program, the computer program includes program instructions, and a processor is enabled to execute the method according to any one of claims 1-7 when the program instructions are executed by the processor.