



(12) 发明专利

(10) 授权公告号 CN 111339072 B

(45) 授权公告日 2023. 09. 15

(21) 申请号 202010110684.6

G06F 18/10 (2023.01)

(22) 申请日 2020.02.23

(56) 对比文件

(65) 同一申请的已公布的文献号

CN 104915735 A, 2015.09.16

申请公布号 CN 111339072 A

CN 109856969 A, 2019.06.07

CN 110346725 A, 2019.10.18

(43) 申请公布日 2020.06.26

CN 103060517 A, 2013.04.24

CN 107977754 A, 2018.05.01

(73) 专利权人 中国平安财产保险股份有限公司

地址 518000 广东省深圳市福田区益田路

5033号平安金融中心12、13、38、39、40

层

审查员 陈丽娜

(72) 发明人 张炯文 汪海祥 陈真

(74) 专利代理机构 深圳市沃德知识产权代理事

务所(普通合伙) 44347

专利代理师 高杰 于志光

(51) Int. Cl.

G06F 18/20 (2023.01)

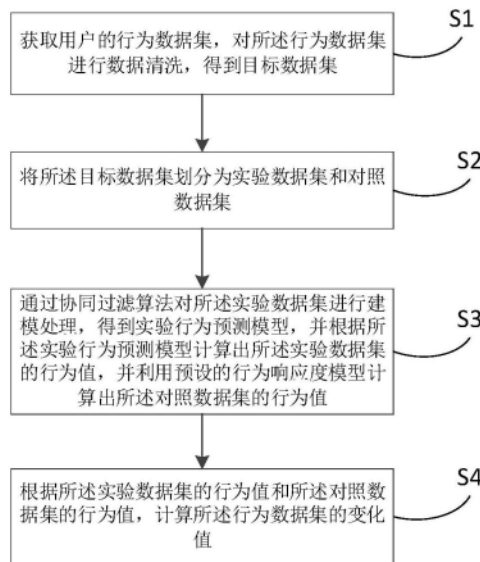
权利要求书3页 说明书10页 附图2页

(54) 发明名称

基于用户行为的变化值分析方法、装置、电子设备及介质

(57) 摘要

本发明涉及大数据技术领域, 揭露了一种基于用户行为的变化值分析的方法, 包括: 获取用户的行为数据集, 对所述行为数据集进行数据清洗, 得到目标数据集; 将所述目标数据集划分为实验数据集和对照数据集; 通过协同过滤算法对所述实验数据集进行建模处理, 得到实验行为预测模型, 并根据所述实验行为预测模型计算出所述实验数据集的行为值; 利用预设的行为响应度模型计算出所述对照数据集的行为值; 根据所述实验数据集的行为值和所述对照数据集的行为值, 计算所述行为数据集的变化值, 并根据所述变化值获取用户的行为值变化轨迹。本发明还提出一种基于用户行为的变化值分析装置、电子设备以及一种计算机可读存储介质。本发明可以实现用户行为的变化值分析。



1. 一种基于用户行为的变化值分析方法,其特征在于,所述方法包括:

获取用户的行为数据集,对所述行为数据集进行数据清洗,得到目标数据集,其中,所述数据清洗包括异常处理和缺失值处理;

将所述目标数据集划分为实验数据集和对照数据集;

通过协同过滤算法对所述实验数据集建立线性回归模型,根据所述实验数据集的变化值调整所述线性回归模型的参数,直至得到的线性回归模型趋于稳定,得到实验行为预测模型,并根据所述实验行为预测模型计算出所述实验数据集的行为值,并根据二叉树构建所述对照数据集的条件概率公式和条件概率的对数似然函数,利用所述条件概率公式和所述对数似然函数计算出所述对照数据集的预测行为学习率,根据所述预测行为学习率,利用所述二叉树计算出所述对照数据集的行为值;

根据所述实验数据集的行为值和所述对照数据集的行为值,计算所述行为数据集的变化值,并根据所述变化值获取用户的行为值变化轨迹;

其中,所述异常处理包括:双边测试剔除和单边测试剔除;所述单边测试剔除包括最小值测试剔除及最大值测试剔除;其中,所述双边测试剔除的计算方法包括:

$$G = \frac{\max|Y_i - \bar{Y}|}{S}$$

其中,G表示异常数据,i属于正整数, \bar{Y} 表示标准数据集的平均值,S表示所述标准数据集的标准差, Y_i 表示所述标准数据集,所述标准数据集是由所述行为数据集进行一致性检查后得到的;

所述最小值测试剔除的计算方法包括:

$$G = \frac{\bar{Y} - Y_{\min}}{S}$$

其中,G表示异常数据, \bar{Y} 表示所述标准数据集的平均值, Y_{\min} 表示所述标准数据集中最小的数据;

所述最大值测试剔除的计算方法包括:

$$G = \frac{\bar{Y} - Y_{\max}}{S}$$

其中,G表示异常数据, \bar{Y} 表示所述标准数据集的平均值, Y_{\max} 表示所述标准数据集中最大的数据;

所述缺失值处理,包括:利用缺失函数检测初始数据集中存在的缺失值,所述初始数据集是由所标准数据集去异常处理后得到的;对所述缺失值利用下述填充方法进行填充,得到所述目标数据集:

$$L(\theta) = L(x_1, x_2, x_3 \dots x_n; \theta) = \prod_{i=1}^n p(x_i | \theta)$$

其中,L(θ)表示填充的数据缺失值, x_i 表示第i个填充的数据缺失值, θ 表示填充的数据缺失值对应的概率参数,n表示初始数据集的数量, $p(x_i | \theta)$ 表示填充的数据缺失值概率;

所述实验行为预测模型包括：

$$E(R|T) = P(0=1|R:T)$$

其中， $E(R|T)$ 表示实验数据集的行为值， P 表示比例符号， T 表示实验数据集， R 表示实验数据集中的实验数据；

所述条件概率公式包括：

$$p(\omega|Context(\omega)) = \prod_{j=2}^{l^\omega} p(d_j^\omega|\omega, \theta_{j-1}^\omega)$$

其中， $p(d_j^\omega|\omega, \theta_{j-1}^\omega)$ 表示二叉树各分支的概率， l^ω 表示联路路径， j 表示分支数量， ω 表示二叉树分支的权值， θ 表示二叉树分支的权重参数， d_j^ω 表示二叉树分支路径；

所述对数似然函数包括：

$$\zeta = \sum_{\omega \in \zeta} p(\omega|Context(\omega))$$

所述预测行为学习率的计算公式包括：

$$V = V(\theta) + \eta \sum_{j=2}^{l^\omega} \frac{\partial \zeta(\omega, j)}{\partial X_\omega}$$

所述对照数据集的行为值记作为：

$$E(R|C) = V(0=1|R:C)$$

其中， $E(R|C)$ 表示对照数据集的行为值， V 表示预测行为学习率， R 表示对照数据集中的对照数据；

所述行为数据集的变化值的计算方法为：

$$Score = |P(0=1|R:T) - V(0=1|R:C)|$$

其中， $Score$ 表示所述变化值， $P(0=1|R:T)$ 表示实验数据集的行为值， $V(0=1|R:C)$ 表示对照数据集的行为值。

2. 如权利要求1所述的基于用户行为的变化值分析方法，其特征在于，所述对所述行为数据集进行数据清洗，得到目标数据集，包括：

执行所述行为数据集的一致性检查处理，得到标准数据集，将所标准数据集进行去异常处理，得到初始数据集，并对所述初始数据集进行缺失值检测，得到所述目标数据集。

3. 一种基于用户行为的变化值分析装置，用于实现如权利要求1至2中任一所述的基于用户行为的变化值分析方法，其特征在于，所述装置包括：

清洗模块，用于获取用户的行为数据集，对所述行为数据集进行数据清洗，得到目标数据集；

划分模块，用于将所述目标数据集划分为实验数据集和对照数据集；

计算模块，用于通过协同过滤算法对所述实验数据集建立线性回归模型，根据所述实验数据集的变化值调整所述线性回归模型的参数，直至得到的线性回归模型趋于稳定，得

到实验行为预测模型,并根据所述实验行为预测模型计算出所述实验数据集的行为值,并根据二叉树构建所述对照数据集的条件概率公式和条件概率的对数似然函数,利用所述条件概率公式和所述对数似然函数计算出所述对照数据集的预测行为学习率,根据所述预测行为学习率,利用所述二叉树计算出所述对照数据集的行为值;

分析模块,用于根据所述实验数据集的行为值和所述对照数据集的行为值,计算所述行为数据集的变化值,并根据所述变化值获取用户的行为值变化轨迹。

4. 一种电子设备,其特征在于,所述电子设备包括:

至少一个处理器;以及,

与所述至少一个处理器通信连接的存储器;其中,

所述存储器存储有可被所述至少一个处理器执行的指令,所述指令被所述至少一个处理器执行,以使所述至少一个处理器能够执行如权利要求1至2中任一所述的基于用户行为的变化值分析方法。

5. 一种计算机可读存储介质,存储有计算机程序,其特征在于,所述计算机程序被处理器执行时实现如权利要求1至2中任一所述的基于用户行为的变化值分析方法。

基于用户行为的变化值分析方法、装置、电子设备及介质

技术领域

[0001] 本发明涉及大数据技术领域,尤其涉及一种基于用户行为的变化值分析的方法、装置、电子设备及计算机可读存储介质。

背景技术

[0002] 在面对用户行为的变化预测场景时,传统的变化预测系统大多使用分割测试和网格搜索的方式来手动调试有限的几种建模组合方式,以实现用户行为变化预测。基于上述的方式,不仅要求工程师有丰富的建模经验,还需要运气,成本和风险都比较高。同时传统机器学习用的大多数模型本质上是线性回归模型,这些模型无法拟合非常复杂的非线性函数,需要人为的通过特征工程把原问题空间转化到一个机器学习模型容易学的表述空间,才能得到好的效果,成本较高且效率低下。

[0003] 因此,需要一种低成本且高效的用户行为变化值的分析方法,以分析用户行为变化。

发明内容

[0004] 本发明提供一种基于用户行为的变化值分析的方法、装置、电子设备及计算机可读存储介质,其主要目的在于采用不同行为预测模型计算出用户的行为值,以分析用户行为变化。

[0005] 为实现上述目的,本发明提供的一种基于用户行为的变化值分析方法,包括:

[0006] 获取用户的行为数据集,对所述行为数据集进行数据清洗,得到目标数据集;

[0007] 将所述目标数据集划分为实验数据集和对照数据集;

[0008] 通过协同过滤算法对所述实验数据集进行建模处理,得到实验行为预测模型,并根据所述实验行为预测模型计算出所述实验数据集的行为值;

[0009] 利用预设的行为响应度模型计算出所述对照数据集的行为值;

[0010] 根据所述实验数据集的行为值和所述对照数据集的行为值,计算所述行为数据集的变化值,并根据所述变化值获取用户的行为值变化轨迹。

[0011] 可选地,所述对所述行为数据集进行数据清洗,得到目标数据集,包括:

[0012] 执行所述行为数据集的一致性检查处理,得到标准数据集,将所标准数据集进行去异常处理,得到初始数据集,并对所述初始数据集进行缺失值检测,得到所述目标数据集。

[0013] 可选地,所述去异常包括:双边测试剔除和单边测试剔除;

[0014] 所述单边测试剔除包括最小值测试剔除及最大值测试剔除;

[0015] 其中,所述双边测试剔除的计算方法包括:

$$[0016] \quad G = \frac{\max|Y_i - \bar{Y}|}{S}$$

[0017] 其中,G表示异常数据,i属于正整数, \bar{Y} 表示所述标准数据集的平均值,S表示所述

标准数据集的标准差, Y_i 表示所述标准数据集;

[0018] 所述最小值测试剔除的计算方法包括:

$$[0019] \quad G = \frac{\bar{Y} - Y_{min}}{S}$$

[0020] 其中, G 表示异常数据, \bar{Y} 表示所述标准数据集的平均值, Y_{min} 表示所述标准数据集中最小的数据;

[0021] 所述最大值测试剔除的计算方法包括:

$$[0022] \quad G = \frac{\bar{Y} - Y_{max}}{S}$$

[0023] 其中, G 表示异常数据, \bar{Y} 表示所述标准数据集的平均值, Y_{max} 表示所述标准数据集中最大的数据。

[0024] 可选地, 所述对所述初始数据集进行缺失值检测, 得到所述目标数据集, 包括:

[0025] 利用缺失函数检测所述初始数据集中存在的缺失值;

[0026] 对所述缺失值利用下述填充方法进行填充, 得到所述目标数据集:

$$[0027] \quad L(\theta) = L(x_1, x_2, x_3 \dots x_n; \theta) = \prod_{i=1}^n p(x_i | \theta)$$

[0028] 其中, $L(\theta)$ 表示填充的数据缺失值, x_i 表示第 i 个填充的数据缺失值, θ 表示填充的数据缺失值对应的概率参数, n 表示初始数据集的数量, $p(x_i | \theta)$ 表示填充的数据缺失值概率。

[0029] 可选地, 所述实验行为预测模型包括下述公式:

$$[0030] \quad E(R|T) = P(O=1|R:T)$$

[0031] 其中, $E(R|T)$ 表示实验数据集的行为值, P 表示比例符号, T 表示实验数据集, R 表示实验数据集中的实验数据。

[0032] 可选地, 所述预设的行为响应度模型包括决策树和二叉树; 及

[0033] 所述利用预设的行为响应度模型计算出所述对照数据集的行为值, 包括:

[0034] 根据所述二叉树构建所述对照数据集的构建条件概率公式和条件概率的对数似然函数;

[0035] 利用所述条件概率公式和对数似然函数计算出所述对照数据集的预测行为学习率;

[0036] 根据所述预测行为学习率, 利用所述决策树决策出所述对照数据集的行为值。

[0037] 为了解决上述问题, 本发明还提供一种基于用户行为的变化值分析装置, 所述装置包括:

[0038] 清洗模块, 用于获取用户的行为数据集, 对所述行为数据集进行数据清洗, 得到目标数据集;

[0039] 划分模块, 用于将所述目标数据集划分为实验数据集和对照数据集;

[0040] 计算模块, 用于通过协同过滤算法对所述实验数据集进行建模处理, 得到实验行为预测模型, 并根据所述实验行为预测模型计算出所述实验数据集的行为值, 并利用预设

的行为响应度模型计算出所述对照数据集的行为值；

[0041] 分析模块,用于根据所述实验数据集的行为值和所述对照数据集的行为值,计算所述行为数据集的变化值,并根据所述变化值获取用户的行为值变化轨迹。

[0042] 为了解决上述问题,本发明还提供一种电子设备,所述电子设备包括:

[0043] 存储器,存储至少一个指令;及

[0044] 处理器,执行所述存储器中存储的指令以实现上述中任意一项所述的基于用户行为的变化值分析方法。

[0045] 为了解决上述问题,本发明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中存储有至少一个指令,所述至少一个指令被电子设备中的处理器执行以实现上述中任意一项所述的基于用户行为的变化值分析方法。

[0046] 本发明实施例获取用户的行为数据集,对所述行为数据集进行数据清洗,得到目标数据集,基于数据清洗减少了后续数据计算的复杂性;将所述目标数据集划分为实验数据集和对照数据集,分别利用不同的行为模型计算出所述实验数据集和对照数据集的对应行为值;根据所述实验数据集的行为值和所述对照数据集的行为值,计算所述行为数据集的变化值,根据所述变化值可以判断出用户的行为值变化轨迹,从而实现了低成本且高效高精度用户行为变化的分析。

附图说明

[0047] 图1为本发明一实施例提供的基于用户行为的变化值分析方法的流程示意图;

[0048] 图2为本发明一实施例提供的基于用户行为的变化值分析方法的模块示意图;

[0049] 图3为本发明一实施例提供的基于用户行为的变化值分析方法的电子设备的内部结构示意图;

[0050] 本发明目的的实现、功能特点及优点将结合实施例,参照附图做进一步说明。

具体实施方式

[0051] 应当理解,此处所描述的具体实施例仅仅用以解释本发明,并不用于限定本发明。

[0052] 本发明提供一种基于用户行为的变化值分析的方法。参照图1所示,为本发明一实施例提供的基于用户行为的变化值分析方法的流程示意图。该方法可以由一个装置执行,该装置可以由软件和/或硬件实现。

[0053] 在本实施例中,基于用户行为的变化值分析的方法包括:

[0054] S1、获取用户的行为数据集,对所述行为数据集进行数据清洗,得到目标数据集。

[0055] 在本发明的至少一个实施例中,所述用户的行为数据集指的是基于在实际业务场景中,用户参与活动所产生的数据。例如,对于一个产险业务的场景中,用户的购买记录、用户的咨询记录以及用户的后续购买记录等。

[0056] 由于获取到的行为数据集中包含大量无用的数据,因此,本发明实施例通过对所述行为数据集进行数据清洗,以实现用户行为的精确定位。

[0057] 详细地,所述对所述行为数据集进行数据清洗,得到目标数据集,包括:

[0058] 执行所述行为数据集的一致性检查处理,得到标准数据集,将所述标准数据集进行去异常处理,得到初始数据集,并对所述初始数据集进行缺失值检测,得到所述目标数据

集。

[0059] 所述行为数据集的一致性检查处理是根据每个数据的合理取值范围和相互关系,检查数据是否合乎要求,发现并删除超出正常范围、逻辑上不合理或者相互矛盾的数据。例如,用1-7级量表测量的变量出现了0值,体重出现了负数等。

[0060] 所述缺失值检测包括:完全随机缺失、随机缺失以及非随机缺失。所述完全随机缺失指的是某一变量缺失值不依赖于其他任何原因的完全随机缺失;所述随机缺失指的是某一变量的缺失与其他变量相关但与该变量本身的数值不相关;所述非随机缺失指的是某一变量的缺失和该变量本身的数值相关。

[0061] 较佳地,本发明实施例通过missmap function缺失函数检测所述初始数据集中存在的缺失值,对所述缺失值利用下述填充方法进行填充,得到所述目标数据集:

$$[0062] \quad L(\theta) = L(x_1, x_2, x_3 \dots x_n; \theta) = \prod_{i=1}^n p(x_i | \theta)$$

[0063] 其中, $L(\theta)$ 表示填充的数据缺失值, x_i 表示第*i*个填充的数据缺失值, θ 表示填充的数据缺失值对应的概率参数, n 表示初始数据集的数量, $p(x_i | \theta)$ 表示填充的数据缺失值概率。本发明实施例中,所述去异常包括:双边测试剔除和单边测试剔除,其中,所述单边测试剔除包括最小值测试剔除及最大值测试剔除;

[0064] 其中,所述双边测试剔除的计算方法包括:

$$[0065] \quad G = \frac{\max |Y_i - \bar{Y}|}{S}$$

[0066] 其中, G 表示异常数据, i 属于正整数, \bar{Y} 表示所述标准数据集的平均值, S 表示所述标准数据集的标准差, Y_i 表示所述标准数据集;

[0067] 所述最小值测试剔除的计算方法包括:

$$[0068] \quad G = \frac{\bar{Y} - Y_{min}}{S}$$

[0069] 其中, G 表示异常数据, \bar{Y} 表示所述标准数据集的平均值, Y_{min} 表示所述标准数据集中最小的数据;

[0070] 所述最大值测试剔除的计算方法包括:

$$[0071] \quad G = \frac{\bar{Y} - Y_{max}}{S}$$

[0072] 其中, G 表示异常数据, \bar{Y} 表示所述标准数据集的平均值, Y_{max} 表示所述标准数据集中最大的数据。

[0073] S2、将所述目标数据集划分为实验数据集和对照数据集。

[0074] 在本发明的至少一个实施例中,将所述目标数据集划分为实验数据集和对照数据集,即将所述目标数据集拷贝一份,其中一份作为实验数据集,另外一份作为对照数据集,分别利用不同的行为模型计算出所述实验数据集和对照数据集的对应行为值,以实现用户行为变化的精准定位。

[0075] S3、通过协同过滤算法对所述实验数据集进行建模处理,得到实验行为预测模型,并根据所述实验行为预测模型计算出所述实验数据集的行为值,并利用预设的行为响应度模型计算出所述对照数据集的行为值。

[0076] 本发明实施例中,所述通过协同过滤算法包括:slope one算法,所述slope one以一种简单的线性回归模型进行行为预测。详细地,利用所述slope one算法对所述实验数据集建立线性回归模型,根据实验数据集的变化值调整所述线性回归模型的参数,直至得到的线性回归模型趋于稳定,得到所述实验行为预测模型。

[0077] 进一步地,本发明实施例根据所述实验行为预测模型计算出所述实验数据集的行为值,详细地,将用户购买成交记为1,未购买成交记为0,于是,所述实验行为预测模型包括:

$$[0078] \quad E(R|T) = P(0=1|R:T)$$

[0079] 其中, $E(R|T)$ 表示实验数据集的行为值, P 表示比例符号, T 表示实验数据集, R 表示实验数据集中的实验数据。

[0080] 进一步地,本发明实施例中,所述预设的行为响应度模型包括:决策树和二叉树。

[0081] 详细地,所述利用预设的行为响应度模型计算出所述对照数据集的行为值,包括:

[0082] 根据所述二叉树构建所述对照数据集的构建条件概率公式 $p(\omega | \text{Context}(\omega))$ 和条件概率的对数似然函数 ζ ,利用所述条件概率公式和对数似然函数计算出所述对照数据集的预测行为学习率,根据所述预测行为学习率,利用所述决策树决策出所述对照数据集的行为值。

[0083] 其中,所述条件概率公式包括:

$$[0084] \quad p(\omega | \text{Context}(\omega)) = \prod_{j=2}^{l^\omega} p(d_j^\omega | \omega, \theta_{j-1}^\omega)$$

[0085] 其中, $p(d_j^\omega | \omega, \theta_{j-1}^\omega)$ 表示二叉树各分支的概率, l^ω 表示联路路径, j 表示分支数量, ω 表示二叉树分支的权值, θ 表示二叉树分支的权重参数, d_j^ω 表示二叉树分支路径。

[0086] 所述对数似然函数包括:

$$[0087] \quad \zeta = \sum_{\omega \in \zeta} p(\omega | \text{Context}(\omega))$$

[0088] 所述预测行为学习率的计算公式包括:

$$[0089] \quad V = V(\theta) + \eta \sum_{j=2}^{l^\omega} \frac{\partial \zeta(\omega, j)}{\partial X_\omega}$$

[0090] 进一步地,本发明实施例可以将所述对照数据集的行为值记作为:

$$[0091] \quad E(R|C) = V(0=1|R:C)$$

[0092] 其中, $E(R|C)$ 表示对照数据集的行为值, V 表示预测行为学习率, T 表示对照数据集, R 表示对照数据集中的对照数据。

[0093] S4、根据所述实验数据集的行为值和所述对照数据集的行为值,计算所述行为数据集的变化值,并根据所述变化值获取用户的行为值变化轨迹。

[0094] 在本发明的至少一个实施例中,所述行为数据集的变化值的计算方法为:

$$[0095] \text{Score} = |P(0=1|R:T) - V(0=1|R:C)|$$

[0096] 其中,Score表示所述变化值,P(0=1|R:T)表示实验数据集的行为值,V(0=1|R:C)表示对照数据集的行为值,根据所述变化值可以精确识别出用户的行为变化,以实现用户行为变化的高精度定位。

[0097] 如图2所示,是本发明基于用户行为的变化值分析装置的功能模块图。

[0098] 本发明所述基于用户行为的变化值分析装置100可以安装于电子设备中。根据实现的功能,所述基于用户行为的变化值分析装置可以包括清洗模块101、划分模块102、计算模块103以及分析模块104。本发所述模块也可以称之为单元,是指一种能够被电子设备处理器所执行,并且能够完成固定功能的一系列计算机程序段,其存储在电子设备的存储器中。

[0099] 在本实施例中,关于各模块/单元的功能如下:

[0100] 所述清洗模块101,用于获取用户的行为数据集,对所述行为数据集进行数据清洗,得到目标数据集;

[0101] 所述划分模块102,用于将所述目标数据集划分为实验数据集和对照数据集;

[0102] 所述计算模块103,用于通过协同过滤算法对所述实验数据集进行建模处理,得到实验行为预测模型,并根据所述实验行为预测模型计算出所述实验数据集的行为值,并利用预设的行为响应度模型计算出所述对照数据集的行为值;

[0103] 所述分析模块104,用于根据所述实验数据集的行为值和所述对照数据集的行为值,计算所述行为数据集的变化值。

[0104] 详细地,所述基于用户行为的变化值分析装置各模块的具体实施步骤如下:

[0105] 所述清洗模块101获取用户的行为数据集,对所述行为数据集进行数据清洗,得到目标数据集。

[0106] 在本发明的至少一个实施例中,所述用户的行为数据集指的是基于在实际业务场景中,用户参与活动所产生的数据。例如,对于一个产险业务的场景中,用户的购买记录、用户的咨询记录以及用户的后续购买记录等。

[0107] 由于获取到的行为数据集中包含大量无用的数据,因此,本发明实施例通过对所述行为数据集进行数据清洗,以实现用户行为的精确定位。

[0108] 详细地,所述对所述行为数据集进行数据清洗,得到目标数据集,包括:

[0109] 执行所述行为数据集的一致性检查处理,得到标准数据集,将所标准数据集进行去异常处理,得到初始数据集,并对所述初始数据集进行缺失值检测,得到所述目标数据集。

[0110] 所述行为数据集的一致性检查处理是根据每个数据的合理取值范围和相互关系,检查数据是否合乎要求,发现并删除超出正常范围、逻辑上不合理或者相互矛盾的数据。例如,用1-7级量表测量的变量出现了0值,体重出现了负数等。

[0111] 所述缺失值检测包括:完全随机缺失、随机缺失以及非随机缺失。所述完全随机缺失指的是某一变量缺失值不依赖于其他任何原因的完全随机缺失;所述随机缺失指的是某一变量的缺失与其他变量相关但与该变量本身的数值不相关;所述非随机缺失指的是某一变量的缺失和该变量本身的数值相关。

[0112] 较佳地,本发明实施例通过missmap function缺失函数检测所述初始数据集中存在的缺失值,对所述缺失值利用下述填充方法进行填充,得到所述目标数据集:

$$[0113] \quad L(\theta) = L(x_1, x_2, x_3 \dots x_n; \theta) = \prod_{i=1}^n p(x_i | \theta)$$

[0114] 其中, $L(\theta)$ 表示填充的数据缺失值, x_i 表示第*i*个填充的数据缺失值, θ 表示填充的数据缺失值对应的概率参数, n 表示初始数据集的数量, $p(x_i | \theta)$ 表示填充的数据缺失值概率。本发明实施例中,所述去异常包括:双边测试剔除和单边测试剔除,其中,所述单边测试剔除包括最小值测试剔除及最大值测试剔除;

[0115] 其中,所述双边测试剔除的计算方法包括:

$$[0116] \quad G = \frac{\max|Y_i - \bar{Y}|}{S}$$

[0117] 其中, G 表示异常数据, i 属于正整数, \bar{Y} 表示所述标准数据集的平均值, S 表示所述标准数据集的标准差, Y_i 表示所述标准数据集;

[0118] 所述最小值测试剔除的计算方法包括:

$$[0119] \quad G = \frac{\bar{Y} - Y_{min}}{S}$$

[0120] 其中, G 表示异常数据, \bar{Y} 表示所述标准数据集的平均值, Y_{min} 表示所述标准数据集中最小的数据;

[0121] 所述最大值测试剔除的计算方法包括:

$$[0122] \quad G = \frac{\bar{Y} - Y_{max}}{S}$$

[0123] 其中, G 表示异常数据, \bar{Y} 表示所述标准数据集的平均值, Y_{max} 表示所述标准数据集中最大的数据。

[0124] 所述划分模块102将所述目标数据集划分为实验数据集和对照数据集。

[0125] 在本发明的至少一个实施例中,将所述目标数据集划分为实验数据集和对照数据集,即将所述目标数据集拷贝一份,其中一份作为实验数据集,另外一份作为对照数据集,分别利用不同的行为模型计算出所述实验数据集和对照数据集的对应行为值,以实现用户行为变化的精准定位。

[0126] 所述计算模块103通过协同过滤算法对所述实验数据集进行建模处理,得到实验行为预测模型,并根据所述实验行为预测模型计算出所述实验数据集的行为值,并利用预设的行为响应度模型计算出所述对照数据集的行为值。

[0127] 本发明实施例中,所述通过协同过滤算法包括:slope one算法,所述slope one以一种简单的线性回归模型进行行为预测。详细地,利用所述slope one算法对所述实验数据集建立线性回归模型,根据实验数据集的变化值调整所述线性回归模型的参数,直至得到的线性回归模型趋于稳定,得到所述实验行为预测模型。

[0128] 进一步地,本发明实施例根据所述实验行为预测模型计算出所述实验数据集的行为值,详细地,将用户购买成交记为1,未购买成交记为0,于是,所述实验行为预测模型包

括：

[0129] $E(R|T) = P(0=1|R:T)$

[0130] 其中, $E(R|T)$ 表示实验数据集的行为值, P 表示比例符号, T 表示实验数据集, R 表示实验数据集中的实验数据。

[0131] 进一步地, 本发明实施例中, 所述预设的行为响应度模型包括: 决策树和二叉树。

[0132] 详细地, 所述利用预设的行为响应度模型计算出所述对照数据集的行为值, 包括:

[0133] 根据所述二叉树构建所述对照数据集的构建条件概率公式 $p(\omega | \text{Context}(\omega))$ 和条件概率的对数似然函数 ζ , 利用所述条件概率公式和对数似然函数计算出所述对照数据集的预测行为学习率, 根据所述预测行为学习率, 利用所述决策树决策出所述对照数据集的行为值。

[0134] 其中, 所述条件概率公式包括:

$$[0135] \quad p(\omega | \text{Context}(\omega)) = \prod_{j=2}^{l^\omega} p(d_j^\omega | \omega, \theta_{j-1}^\omega)$$

[0136] 其中, $p(d_j^\omega | \omega, \theta_{j-1}^\omega)$ 表示二叉树各分支的概率, l^ω 表示联路路径, j 表示分支数量, ω 表示二叉树分支的权值, θ 表示二叉树分支的权重参数, d_j^ω 表示二叉树分支路径。

[0137] 所述对数似然函数包括:

$$[0138] \quad \zeta = \sum_{\omega \in \zeta} p(\omega | \text{Context}(\omega))$$

[0139] 所述预测行为学习率的计算公式包括:

$$[0140] \quad V = V(\theta) + \eta \sum_{j=2}^{l^\omega} \frac{\partial \zeta(\omega, j)}{\partial X_\omega}$$

[0141] 进一步地, 本发明实施例可以将所述对照数据集的行为值记作为:

[0142] $E(R|C) = V(0=1|R:C)$

[0143] 其中, $E(R|C)$ 表示对照数据集的行为值, V 表示预测行为学习率, T 表示对照数据集, R 表示对照数据集中的对照数据。

[0144] 所述分析模块104根据所述实验数据集的行为值和所述对照数据集的行为值, 计算所述行为数据集的变化值。

[0145] 在本发明的至少一个实施例中, 所述行为数据集的变化值的计算方法为:

[0146] $\text{Score} = |P(0=1|R:T) - V(0=1|R:C)|$

[0147] 其中, Score 表示所述变化值, $P(0=1|R:T)$ 表示实验数据集的行为值, $V(0=1|R:C)$ 表示对照数据集的行为值, 根据所述变化值可以精确识别出用户的行为变化, 以实现用户行为变化的高精度定位。

[0148] 如图3所示, 是本发明实现基于用户行为的变化值分析的方法的电子设备的结构示意图。

[0149] 所述电子设备1可以包括处理器10、存储器11和总线, 还可以包括存储在所述存储器11中并可在所述处理器10上运行的计算机程序, 如基于用户行为的变化值分析程序。

[0150] 其中,所述存储器11至少包括一种类型的可读存储介质,所述可读存储介质包括闪存、移动硬盘、多媒体卡、卡型存储器(例如:SD或DX存储器等)、磁性存储器、磁盘、光盘等。所述存储器11在一些实施例中可以是电子设备1的内部存储单元,例如该电子设备1的移动硬盘。所述存储器11在另一些实施例中也可以是电子设备1的外部存储设备,例如电子设备1上配备的插接式移动硬盘、智能存储卡(Smart Media Card,SMC)、安全数字(Secure Digital,SD)卡、闪存卡(Flash Card)等。进一步地,所述存储器11还可以既包括电子设备1的内部存储单元也包括外部存储设备。所述存储器11不仅可以用于存储安装于电子设备1的应用软件及各类数据,例如基于用户行为的变化值分析程序的代码等,还可以用于暂时地存储已经输出或者将要输出的数据。

[0151] 所述处理器10在一些实施例中可以由集成电路组成,例如可以由单个封装的集成电路所组成,也可以是由多个相同功能或不同功能封装的集成电路所组成,包括一个或者多个中央处理器(Central Processing unit,CPU)、微处理器、数字处理芯片、图形处理器及各种控制芯片的组合等。所述处理器10是所述电子设备的控制核心(Control Unit),利用各种接口和线路连接整个电子设备的各个部件,通过运行或执行存储在所述存储器11内的程序或者模块(例如基于用户行为的变化值分析程序等),以及调用存储在所述存储器11内的数据,以执行电子设备1的各种功能和处理数据。

[0152] 所述总线可以是外设部件互连标准(peripheral component interconnect,简称PCI)总线或扩展工业标准结构(extended industry standard architecture,简称EISA)总线等。该总线可以分为地址总线、数据总线、控制总线等。所述总线被设置为实现所述存储器11以及至少一个处理器10等之间的连接通信。

[0153] 图3仅示出了具有部件的电子设备,本领域技术人员可以理解的是,图3示出的结构并不构成对所述电子设备1的限定,可以包括比图示更少或者更多的部件,或者组合某些部件,或者不同的部件布置。

[0154] 例如,尽管未示出,所述电子设备1还可以包括给各个部件供电的电源(比如电池),优选地,电源可以通过电源管理装置与所述至少一个处理器10逻辑相连,从而通过电源管理装置实现充电管理、放电管理、以及功耗管理等功能。电源还可以包括一个或一个以上的直流或交流电源、再充电装置、电源故障检测电路、电源转换器或者逆变器、电源状态指示器等任意组件。所述电子设备1还可以包括多种传感器、蓝牙模块、Wi-Fi模块等,在此不再赘述。

[0155] 进一步地,所述电子设备1还可以包括网络接口,可选地,所述网络接口可以包括有线接口和/或无线接口(如WI-FI接口、蓝牙接口等),通常用于在该电子设备1与其他电子设备之间建立通信连接。

[0156] 可选地,该电子设备1还可以包括用户接口,用户接口可以是显示器(Display)、输入单元(比如键盘(Keyboard)),可选地,用户接口还可以是标准的有线接口、无线接口。可选地,在一些实施例中,显示器可以是LED显示器、液晶显示器、触控式液晶显示器以及OLED(Organic Light-Emitting Diode,有机发光二极管)触摸器等。其中,显示器也可以适当的称为显示屏或显示单元,用于显示在电子设备1中处理的信息以及用于显示可视化的用户界面。

[0157] 应该了解,所述实施例仅为说明之用,在专利申请范围上并不受此结构的限制。

[0158] 所述电子设备1中的所述存储器11存储的基于用户行为的变化值分析程序12是多个指令的组合,在所述处理器10中运行时,可以实现:

[0159] 获取用户的行为数据集,对所述行为数据集进行数据清洗,得到目标数据集;

[0160] 将所述目标数据集划分为实验数据集和对照数据集;

[0161] 通过协同过滤算法对所述实验数据集进行建模处理,得到实验行为预测模型,并根据所述实验行为预测模型计算出所述实验数据集的行为值,并利用预设的行为响应度模型计算出所述对照数据集的行为值;

[0162] 根据所述实验数据集的行为值和所述对照数据集的行为值,计算所述行为数据集的变化值。

[0163] 具体地,所述处理器10对上述指令的具体实现方法可参考图1对应实施例中相关步骤的描述,在此不赘述。

[0164] 进一步地,所述电子设备1集成的模块/单元如果以软件功能单元的形式实现并作为独立的产品销售或使用,可以存储在一个计算机可读取存储介质中。所述计算机可读取介质可以包括:能够携带所述计算机程序代码的任何实体或装置、记录介质、U盘、移动硬盘、磁碟、光盘、计算机存储器、只读存储器(ROM,Read-Only Memory)。

[0165] 在本发明所提供的几个实施例中,应该理解到,所揭露的设备,装置和方法,可以通过其它的方式实现。例如,以上所描述的装置实施例仅仅是示意性的,例如,所述模块的划分,仅仅为一种逻辑功能划分,实际实现时可以有另外的划分方式。

[0166] 所述作为分离部件说明的模块可以是或者也可以不是物理上分开的,作为模块显示的部件可以是或者也可以不是物理单元,即可以位于一个地方,或者也可以分布到多个网络单元上。可以根据实际的需要选择其中的部分或者全部模块来实现本实施例方案的目的。

[0167] 另外,在本发明各个实施例中的各功能模块可以集成在一个处理单元中,也可以是各个单元单独物理存在,也可以两个或两个以上单元集成在一个单元中。上述集成的单元既可以采用硬件的形式实现,也可以采用硬件加软件功能模块的形式实现。

[0168] 对于本领域技术人员而言,显然本发明不限于上述示范性实施例的细节,而且在不背离本发明的精神或基本特征的情况下,能够以其他的具体形式实现本发明。

[0169] 因此,无论从哪一点来看,均应将实施例看作是示范性的,而且是非限制性的,本发明的范围由所附权利要求而不是上述说明限定,因此旨在将落在权利要求的等同要件的含义和范围内的所有变化涵括在本发明内。不应将权利要求中的任何附关联图标记视为限制所涉及的权利要求。

[0170] 此外,显然“包括”一词不排除其他单元或步骤,单数不排除复数。系统权利要求中陈述的多个单元或装置也可以由一个单元或装置通过软件或者硬件来实现。第二等词语用来表示名称,而并不表示任何特定的顺序。

[0171] 最后应说明的是,以上实施例仅用以说明本发明的技术方案而非限制,尽管参照较佳实施例对本发明进行了详细说明,本领域的普通技术人员应当理解,可以对本发明的技术方案进行修改或等同替换,而不脱离本发明技术方案的精神和范围。

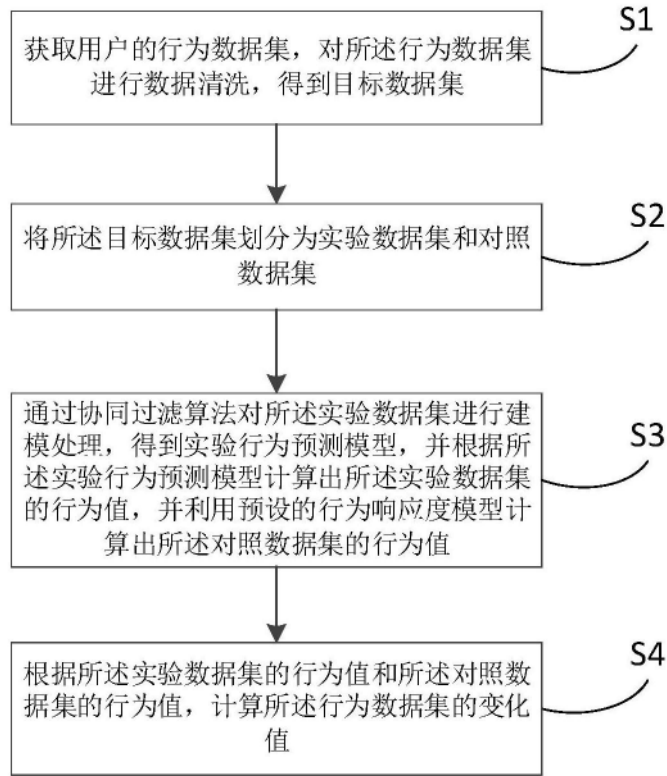


图1



图2

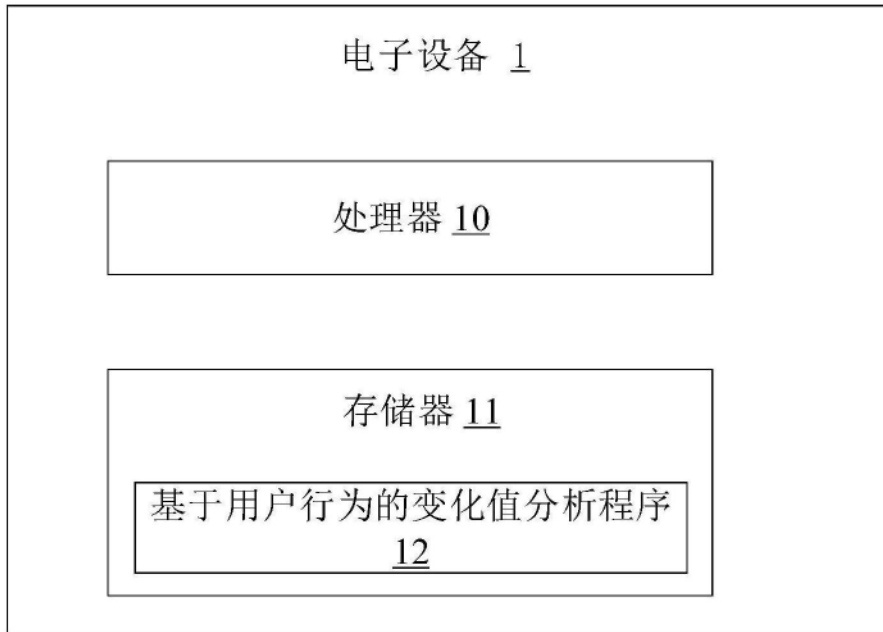


图3