



(12)发明专利申请

(10)申请公布号 CN 111310784 A

(43)申请公布日 2020.06.19

(21)申请号 202010035782.8

(22)申请日 2020.01.14

(71)申请人 支付宝(杭州)信息技术有限公司
地址 310000 浙江省杭州市西湖区西溪路
556号8层B段801-11

(72)发明人 倪翔 汲小溪 王维强

(74)专利代理机构 北京国昊天诚知识产权代理
有限公司 11315
代理人 姚琳洁 朱文杰

(51) Int. Cl.

G06K 9/62(2006.01)

G06Q 10/06(2012.01)

G06Q 40/04(2012.01)

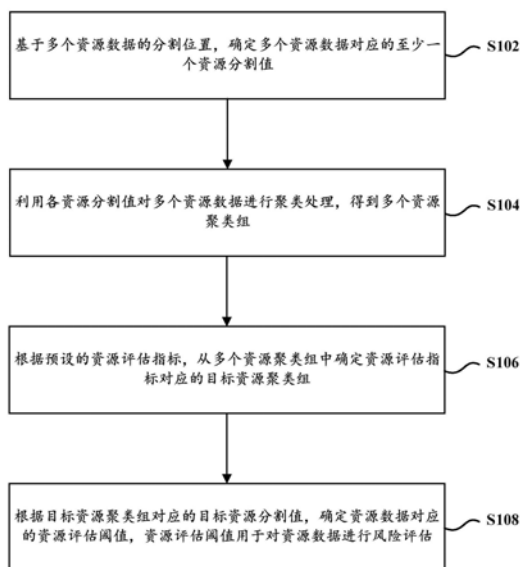
权利要求书3页 说明书11页 附图4页

(54)发明名称

资源数据的处理方法及装置

(57)摘要

本说明书一个或多个实施例公开了一种资源数据的处理方法及装置,用以解决现有技术中数据聚类效率低以及风险管理效率低的问题。所述方法包括:基于多个资源数据的分割位置,确定所述多个资源数据对应的至少一个资源分割值。利用各所述资源分割值对所述多个资源数据进行聚类处理,得到多个资源聚类组。根据预设的资源评估指标,从所述多个资源聚类组中确定所述资源评估指标对应的目标资源聚类组。所述资源评估指标包含对所述资源数据进行风险评估所使用的风险评估参数。根据所述目标资源聚类组对应的目标资源分割值,确定所述资源数据对应的资源评估阈值。所述资源评估阈值用于对所述资源数据进行风险评估。



1. 一种资源数据的处理方法,包括:

基于多个资源数据的分割位置,确定所述多个资源数据对应的至少一个资源分割值;所述资源分割值与所述资源数据的极端值之间的第一差异度小于所述分割位置与所述极端值之间的第二差异度;所述极端值包括所述资源数据中的最大数据值和/或最小数据值;

利用各所述资源分割值对所述多个资源数据进行聚类处理,得到多个资源聚类组;

根据预设的资源评估指标,从所述多个资源聚类组中确定所述资源评估指标对应的目标资源聚类组;所述资源评估指标包含对所述资源数据进行风险评估所使用的风险评估参数;

根据所述目标资源聚类组对应的目标资源分割值,确定所述资源数据对应的资源评估阈值;所述资源评估阈值用于对所述资源数据进行风险评估。

2. 根据权利要求1所述的方法,所述基于多个资源数据的分割位置,确定所述多个资源数据对应的至少一个资源分割值,包括:

确定对所述多个资源数据进行分割的至少一个所述分割位置;

根据各所述分割位置与所述资源分割值之间的映射关系,计算各所述分割位置分别对应的所述资源分割值。

3. 根据权利要求1所述的方法,所述利用各所述资源分割值对所述多个资源数据进行聚类处理,得到多个资源聚类组,包括:

将各所述资源分割值按照预设维度进行排序;所述预设维度包括所述资源数据的数据大小;

基于排序后的各所述资源分割值,将每两个相邻的所述资源分割值确定为一个资源聚类组对应的边界值;

基于各所述资源聚类组分别对应的所述边界值,确定多个所述资源聚类组;

将各所述资源数据分别划分至对应的各所述资源聚类组中。

4. 根据权利要求1所述的方法,所述根据所述目标资源聚类组对应的目标资源分割值,确定所述资源数据对应的资源评估阈值,包括:

确定各所述目标资源分割值分别对应的权重;其中,各所述目标资源分割值分别对应的权重之和为1;

根据各所述目标资源分割值及各所述目标资源分割值分别对应的权重,计算所述资源数据对应的所述资源评估阈值。

5. 根据权利要求4所述的方法,所述风险评估参数包含第一资源分割值;所述目标资源聚类组包含第一目标资源分割值及第二目标资源分割值;

所述确定各所述目标资源分割值分别对应的权重,包括:

确定所述第一资源分割值对应的资源百分位数;

确定所述第一资源分割值对应的资源百分位数为所述第一目标资源分割值对应的权重;

计算所述资源百分位数与1之间的差值的绝对值;确定所述绝对值为所述第二目标资源分割值对应的权重。

6. 根据权利要求5所述的方法,所述根据预设的资源评估指标,从所述多个资源聚类组中确定所述资源评估指标对应的目标资源聚类组,包括:

从所述多个资源聚类组中确定所述第一资源分割值所在的资源聚类组；
将所述第一资源分割值所在的资源聚类组确定为所述目标资源聚类组。

7. 一种资源数据的处理装置, 包括:

第一确定模块, 基于多个资源数据的分割位置, 确定所述多个资源数据对应的至少一个资源分割值; 所述资源分割值与所述资源数据的极端值之间的第一差异度小于所述分割位置与所述极端值之间的第二差异度; 所述极端值包括所述资源数据中的最大数据值和/或最小数据值;

聚类模块, 利用各所述资源分割值对所述多个资源数据进行聚类处理, 得到多个资源聚类组;

第二确定模块, 根据预设的资源评估指标, 从所述多个资源聚类组中确定所述资源评估指标对应的目标资源聚类组; 所述资源评估指标包含对所述资源数据进行风险评估所使用的风险评估参数;

第三确定模块, 根据所述目标资源聚类组对应的目标资源分割值, 确定所述资源数据对应的资源评估阈值; 所述资源评估阈值用于对所述资源数据进行风险评估。

8. 根据权利要求7所述的装置, 所述第一确定模块包括:

第一确定单元, 确定对所述多个资源数据进行分割的至少一个所述分割位置;

计算单元, 根据各所述分割位置与所述资源分割值之间的映射关系, 计算各所述分割位置分别对应的所述资源分割值。

9. 根据权利要求7所述的装置, 所述聚类模块包括:

排序单元, 将各所述资源分割值按照预设维度进行排序; 所述预设维度包括所述资源数据的数据大小;

第二确定单元, 基于排序后的各所述资源分割值, 将每两个相邻的所述资源分割值确定为一个资源聚类组对应的边界值;

第三确定单元, 基于各所述资源聚类组分别对应的所述边界值, 确定多个所述资源聚类组;

划分单元, 将各所述资源数据分别划分至对应的各所述资源聚类组中。

10. 一种资源数据的处理设备, 包括:

处理器; 以及

被安排成存储计算机可执行指令的存储器, 所述可执行指令在被执行时使所述处理器:

基于多个资源数据的分割位置, 确定所述多个资源数据对应的至少一个资源分割值; 所述资源分割值与所述资源数据的极端值之间的第一差异度小于所述分割位置与所述极端值之间的第二差异度; 所述极端值包括所述资源数据中的最大数据值和/或最小数据值;

利用各所述资源分割值对所述多个资源数据进行聚类处理, 得到多个资源聚类组;

根据预设的资源评估指标, 从所述多个资源聚类组中确定所述资源评估指标对应的目标资源聚类组; 所述资源评估指标包含对所述资源数据进行风险评估所使用的风险评估参数;

根据所述目标资源聚类组对应的目标资源分割值, 确定所述资源数据对应的资源评估阈值; 所述资源评估阈值用于对所述资源数据进行风险评估。

11. 一种存储介质,用于存储计算机可执行指令,所述可执行指令在被执行时实现以下流程:

基于多个资源数据的分割位置,确定所述多个资源数据对应的至少一个资源分割值;所述资源分割值与所述资源数据的极端值之间的第一差异度小于所述分割位置与所述极端值之间的第二差异度;所述极端值包括所述资源数据中的最大数据值和/或最小数据值;

利用各所述资源分割值对所述多个资源数据进行聚类处理,得到多个资源聚类组;

根据预设的资源评估指标,从所述多个资源聚类组中确定所述资源评估指标对应的目标资源聚类组;所述资源评估指标包含对所述资源数据进行风险评估所使用的风险评估参数;

根据所述目标资源聚类组对应的目标资源分割值,确定所述资源数据对应的资源评估阈值;所述资源评估阈值用于对所述资源数据进行风险评估。

资源数据的处理方法及装置

技术领域

[0001] 本文件涉及数据处理及风险评估技术领域,尤其涉及一种资源数据的处理方法及装置。

背景技术

[0002] 在风控场景中涉及到的商户主体主要有三种类型:直连商户、间连商户和小程序商户。对这些商户主体的风险管理包含了商户准入、商户风险识别、商户风险运营和商户风险决策者四个主要环节。

[0003] 在商户管理的各个环节中,风控运营、策略和模型都是基于商户主体的各种数据,而实际上,不同商户主体之间存在千丝万缕的联系(例如身份重合关系、资金往来关系、介质共用关系、地理位置相近等),如果能够基于主体之间的联系,将不同的主体聚合在一起,打通不同主体的数据,对每个主体进行分析、管控和建模,则能够有效地提升风险管理的准确性、覆盖率和实效性,实现生态风险联防。

[0004] 聚合计算中一个重要的环节是分位数的计算。分位数的计算在描述性统计中很常见,比如说相较于平均数、中位数不会受到异常值的影响,但分位数的计算过程比较复杂,需要保留所有具体值,排序后取得中间位置的数作为结果。这种复杂的分位数计算方法导致主体聚合效率降低,以及导致主体管理(如主体风险防控等)效率降低等。

发明内容

[0005] 一方面,本说明书一个或多个实施例提供一种资源数据的处理方法,包括:基于多个资源数据的分割位置,确定所述多个资源数据对应的至少一个资源分割值。所述资源分割值与所述资源数据的极端值之间的第一差异度小于所述分割位置与所述极端值之间的第二差异度。所述极端值包括所述资源数据中的最大数据值和/或最小数据值。利用各所述资源分割值对所述多个资源数据进行聚类处理,得到多个资源聚类组。根据预设的资源评估指标,从所述多个资源聚类组中确定所述资源评估指标对应的目标资源聚类组。所述资源评估指标包含对所述资源数据进行风险评估所使用的风险评估参数。根据所述目标资源聚类组对应的目标资源分割值,确定所述资源数据对应的资源评估阈值。所述资源评估阈值用于对所述资源数据进行风险评估。

[0006] 另一方面,本说明书一个或多个实施例提供一种资源数据的处理装置,包括:第一确定模块,基于多个资源数据的分割位置,确定所述多个资源数据对应的至少一个资源分割值。所述资源分割值与所述资源数据的极端值之间的第一差异度小于所述分割位置与所述极端值之间的第二差异度。所述极端值包括所述资源数据中的最大数据值和/或最小数据值。聚类模块,利用各所述资源分割值对所述多个资源数据进行聚类处理,得到多个资源聚类组。第二确定模块,根据预设的资源评估指标,从所述多个资源聚类组中确定所述资源评估指标对应的目标资源聚类组。所述资源评估指标包含对所述资源数据进行风险评估所使用的风险评估参数。第三确定模块,根据所述目标资源聚类组对应的目标资源分割值,确

定所述资源数据对应的资源评估阈值。所述资源评估阈值用于对所述资源数据进行风险评估。

[0007] 再一方面,本说明书一个或多个实施例提供一种资源数据的处理设备,包括:处理器;以及被安排成存储计算机可执行指令的存储器,所述可执行指令在被执行时使所述处理器:基于多个资源数据的分割位置,确定所述多个资源数据对应的至少一个资源分割值。所述资源分割值与所述资源数据的极端值之间的第一差异度小于所述分割位置与所述极端值之间的第二差异度。所述极端值包括所述资源数据中的最大数据值和/或最小数据值。利用各所述资源分割值对所述多个资源数据进行聚类处理,得到多个资源聚类组。根据预设的资源评估指标,从所述多个资源聚类组中确定所述资源评估指标对应的目标资源聚类组。所述资源评估指标包含对所述资源数据进行风险评估所使用的风险评估参数。根据所述目标资源聚类组对应的目标资源分割值,确定所述资源数据对应的资源评估阈值。所述资源评估阈值用于对所述资源数据进行风险评估。

[0008] 再一方面,本说明书一个或多个实施例提供一种存储介质,用于存储计算机可执行指令,所述可执行指令在被执行时实现以下流程:基于多个资源数据的分割位置,确定所述多个资源数据对应的至少一个资源分割值。所述资源分割值与所述资源数据的极端值之间的第一差异度小于所述分割位置与所述极端值之间的第二差异度。所述极端值包括所述资源数据中的最大数据值和/或最小数据值。利用各所述资源分割值对所述多个资源数据进行聚类处理,得到多个资源聚类组。根据预设的资源评估指标,从所述多个资源聚类组中确定所述资源评估指标对应的目标资源聚类组。所述资源评估指标包含对所述资源数据进行风险评估所使用的风险评估参数。根据所述目标资源聚类组对应的目标资源分割值,确定所述资源数据对应的资源评估阈值。所述资源评估阈值用于对所述资源数据进行风险评估。

附图说明

[0009] 为了更清楚地说明本说明书一个或多个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说明书一个或多个实施例中记载的一些实施例,对于本领域普通技术人员来讲,在不付出创造性劳动性的前提下,还可以根据这些附图获得其他的附图。

[0010] 图1是根据本说明书一实施例的一种资源数据的处理方法的示意性流程图;

[0011] 图2是根据本说明书另一实施例的一种资源数据的处理方法的示意性流程图;

[0012] 图3是根据本说明书一实施例的一种资源数据的处理装置的示意性框图;

[0013] 图4是根据本说明书一实施例的一种资源数据的处理设备的示意性框图。

具体实施方式

[0014] 本说明书一个或多个实施例提供一种资源数据的处理方法及装置,用以解决现有技术中数据聚类效率低以及风险管理效率低的问题。

[0015] 为了使本技术领域的人员更好地理解本说明书一个或多个实施例中的技术方案,下面将结合本说明书一个或多个实施例中的附图,对本说明书一个或多个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说明书一部分实施例,而不是

全部的实施例。基于本说明书一个或多个实施例，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，都应当属于本文件保护的范围。

[0016] 图1是根据本说明书一实施例的一种资源数据的处理方法的示意性流程图，如图1所示，该方法包括：

[0017] S102，基于多个资源数据的分割位置，确定多个资源数据对应的至少一个资源分割值。

[0018] 其中，资源分割值与资源数据的极端值之间的第一差异度小于分割位置与极端值之间的第二差异度。极端值包括资源数据中的最大数据值和/或最小数据值。例如，将资源数据按照数据大小的顺序进行排列，那么极端值即为排在最前的资源数据和排在最后的资源数据。资源分割值一般位于0~1之间，其表示对资源数据的分割位置。例如，资源分割值为20%，则表示在资源数据的20%位置处进行分割。

[0019] 资源分割值与资源数据的极端值之间的第一差异度，即为资源分割值与资源数据的的极端值的接近程度。分割数据与资源数据的极端值之间的第二差异度，即为分割数据与资源数据的的极端值的接近程度。

[0020] S104，利用各资源分割值对多个资源数据进行聚类处理，得到多个资源聚类组。

[0021] S106，根据预设的资源评估指标，从多个资源聚类组中确定资源评估指标对应的目标资源聚类组。

[0022] 其中，资源评估指标包含对资源数据进行风险评估所使用的风险评估参数。风险评估参数例如：低风险数据在所有资源数据中的占比。例如，风险评估参数为低风险数据在所有资源数据中的占比90%，则说明预设的资源评估指标希望对资源数据进行风险评估时低风险数据的占比为90%。当然，还可预设其他资源评估指标，例如低风险数据的总数目或高风险数据的总数目等。

[0023] S108，根据目标资源聚类组对应的目标资源分割值，确定资源数据对应的资源评估阈值，资源评估阈值用于对资源数据进行风险评估。

[0024] 采用本说明书一个或多个实施例的技术方案，能够基于多个资源数据的分割位置确定多个资源数据对应的至少一个资源分割值，并利用各资源分割值对多个资源数据进行聚类处理，得到多个资源聚类组。可见，对资源数据进行分割时并不是基于原始的分割位置分割的，而是通过计算后的与资源数据的极端值之间的差异度更小的资源分割值进行分割，由于实际应用中人们更关心处于极端位置的资源数据，因此可使得资源数据的聚类结果更加精准、快速。并且通过根据预设的资源评估指标从多个资源聚类组中确定资源评估指标对应的目标资源聚类组，进而根据目标资源聚类组对应的目标资源分割值确定资源数据对应的资源评估阈值，使得资源评估阈值能够基于预设的资源评估指标来确定，不仅提升了资源评估阈值的精确度，且使得资源数据的风险评估结果更加准确。

[0025] 上述实施例提供的资源数据的处理方法中，对资源数据的聚类可利用Tdigest算法。TDigest是一个简单、快速、精确度高、可并行化的近似百分位算法，TDigest算法的核心思想是Sketch(素描)，是一种抽象、简化的思想，一种处理问题的方式。以下详细介绍如何利用Tdigest算法对资源数据进行聚类。

[0026] 在一个实施例中，执行S102时，可先确定对多个资源数据进行分割的至少一个分割位置；进而根据各分割位置与资源分割值之间的映射关系，计算各分割位置分别对应的

资源分割值。

[0027] 其中,分割位置可以是百分数的形式,分割位置对应的资源分割值可以为资源数据的质心。分割位置与资源分割值之间的映射关系(即百分数与质心之间的映射关系)如下述公式(1)。

$$[0028] \quad k(q,\delta) = \delta \left(\frac{\sin^{-1}(2q-1)}{\pi} + \frac{1}{2} \right) \quad (1)$$

[0029] 其中,q表示百分数;k表示质心; δ 为常量,通常取比较小的值,如0.01。 δ 的取值影响质心k的大小。

[0030] 本实施例中,通过各分割位置与资源分割值之间的映射关系即可将分割位置转化为资源分割值,由于资源分割值与资源数据的极端值之间的差异度更小,因此能够使资源数据的聚类结果更加精准、快速。

[0031] 在一个实施例中,执行S104时,可首先将各资源分割值按照预设维度进行排序,预设维度包括资源数据的数据大小。然后基于排序后的各资源分割值,将每两个相邻的资源分割值确定为一个资源聚类组对应的边界值。再基于各资源聚类组分别对应的边界值,确定多个资源聚类组,将各资源数据分别划分至对应的各资源聚类组中。

[0032] 假设资源分割值为资源数据的质心,对各资源分割值按照资源数据的大小进行排序,即按照质心大小对各质心进行排序。然后将每两个相邻的质心确定为一个资源聚类组对应的边界值,再基于每个资源聚类组分别对应的边界值确定出多个资源聚类组,进而将各资源数据分别划分至对应的各资源聚类组中。

[0033] 将各资源数据分别划分至对应的各资源聚类组中时,可将各资源数据按照数据大小进行排序,然后将各资源数据与各资源聚类组分别对应的边界值进行比对,以确定出各资源数据落入哪一个资源聚类组,并将各资源数据划分至各自所落入的资源聚类组内。

[0034] 本实施例中,由于依据资源分割值(即质心)对资源数据进行聚类,且资源分割值更加靠近资源数据的极端值,因此聚类结果更加精确、更加符合实际应用中人们对极端位置资源数据的关心需求。

[0035] 上述实施例中指出,风险评估参数可以为高风险数据在所有资源数据中的占比。基于此,若风险评估参数包含第一资源分割值,即以第一资源分割值表征高风险数据在所有资源数据中的占比。则执行S106时,可从多个资源聚类组中确定出第一资源分割值所在的资源聚类组,并将第一资源分割值所在的资源聚类组确定为目标资源聚类组。

[0036] 在一个实施例中,根据目标资源聚类组对应的目标资源分割值确定资源数据对应的资源评估阈值时,可首先确定各目标资源分割值分别对应的权重,进而根据各目标资源分割值及各目标资源分割值分别对应的权重,计算资源数据对应的资源评估阈值。其中,各目标资源分割值分别对应的权重之和为1。目标资源聚类组对应的目标资源分割值即为目标资源聚类组的两个边界值。

[0037] 本实施例中,若风险评估参数包含第一资源分割值,即以第一资源分割值表征高风险数据在所有资源数据中的占比。目标资源聚类组包含第一目标资源分割值及第二目标资源分割值(即目标资源聚类组的两个边界值)。则可按照如下方式确定各目标资源分割值分别对应的权重:

[0038] 首先,确定第一资源分割值对应的资源百分位数。

[0039] 其次,确定第一资源分割值对应的资源百分位数为第一目标资源分割值对应的权重。

[0040] 再次,计算资源百分位数与1之间的差值的绝对值,确定该绝对值为第二目标资源分割值对应的权重。

[0041] 上述实施例中,通过根据预设的资源评估指标从多个资源聚类组中确定资源评估指标对应的目标资源聚类组,进而根据目标资源聚类组对应的目标资源分割值确定资源数据对应的资源评估阈值,使得资源评估阈值能够基于预设的资源评估指标来确定,不仅提升了资源评估阈值的精确度,且使得资源数据的风险评估结果更加准确。

[0042] 上述实施例提供的资源数据的处理方法可应用于各类资源数据的风险评估场景中。例如,对交易金额的风险评估、业务数据的风险评估等。下面以对交易金额的风险评估场景为例进行说明。

[0043] 图2是根据本说明书另一实施例的一种资源数据的处理方法的示意性流程图。该实施例中,资源数据为某一时间段内的多笔交易金额。如图2所示,该方法包括:

[0044] S201,确定对多笔交易金额进行分割的多个百分数。

[0045] 其中,百分数即为多笔交易金额的分割位置,例如,确定对多笔交易金额进行分割的百分数为20%、40%、60%和80%。百分数的确定可由用户指定,也可由计算机按照预设规则来确定。预设规则可以是以下任一种规则:从0~1之间随机确定百分数、从0~1之间按照预设间隔均匀确定百分数、从0~1之间均匀确定N个百分数,等等。

[0046] S202,基于确定的多个百分数,确定多笔交易金额对应的多个金额分割值。

[0047] 其中,金额分割值相当于多笔交易金额的质心。已知百分数,则可通过上述公式(1)来确定出每个百分数对应的金额分割值。该步骤确定出的金额分割值一般位于0~1之间,其表示对交易金额的分割位置。例如,金额分割值为20%,则表示在交易金额的20%位置处进行分割。

[0048] 本实施例中,质心具有以下特征:质心与交易金额的极端值之间的第一差异度小于S201中确定的百分数与交易金额的极端值之间的第二差异度。其中,交易金额的极端值包括所有交易金额中的最大金额值和最小金额值。即,将百分数转化为质心对交易金额进行分割,可使得交易金额的分割位置更靠近交易金额的极端值。

[0049] S203,对多个金额分割值按照金额大小进行排序,并将每两个相邻的金额分割值作为一个金额聚类组对应的边界值,得到多个金额聚类组。

[0050] S204,将各笔交易金额分别划分至对应的金额聚类组中。

[0051] 该步骤中,可将各笔交易金额按照金额大小进行排序,然后看每笔交易金额落入哪一个金额聚类组,再将每笔交易金额划分至各自所落入的金额聚类组中。

[0052] S205,确定交易金额评估指标。

[0053] 其中,交易金额评估指标包括对交易金额进行风险评估所使用的风险评估参数,例如,低风险交易金额在所有交易金额中的占比。

[0054] S206,从多个金额聚类组中确定交易金额评估指标对应的目标金额聚类组。

[0055] 假设交易金额评估指标为低风险交易金额在所有交易金额中的占比90%。再假设S202确定出的多个金额分割值分别为:0.2、0.8、0.95。则可确定交易金额评估指标对应的目标金额聚类组为由金额分割值0.8和0.95组成的金额聚类组,即将金额分割值0.8和0.95

做为两个边界值的金额聚类组。

[0056] S207,根据交易金额评估指标以及目标金额聚类组对应的目标金额分割值,并利用插值法计算出金额评估阈值。

[0057] 该步骤中,利用插值法计算出金额评估阈值的具体方式如下:首先,确定金额百分位数(即交易金额评估指标)。其次,确定金额百分位数为第一目标金额分割值(即目标金额聚类组对应的较小的目标金额分割值)对应的权重,以及,计算金额百分位数与1之间的差值的绝对值,确定该绝对值为第二目标金额分割值(即目标金额聚类组对应的较大的目标金额分割值)对应的权重。最后,基于各目标金额分割值及各目标金额分割值分别对应的权重,计算交易金额对应的金额评估阈值。

[0058] 假设金额百分位数为90%,则第一目标金额分割值为a,其对应的权重为0.9;第二目标金额分割值为b,其对应的权重为0.1。金额评估阈值为: $a*0.9+b*0.1$ 。

[0059] S208,基于金额评估阈值对交易金额进行风险评估。

[0060] 例如,若交易金额大于金额评估阈值,则可确定交易金额属于高风险交易金额;若交易金额小于或等于金额评估阈值,则可确定交易金额属于低风险交易金额。

[0061] 本实施例中,能够基于与交易金额的极端值之间的差异度更小的各金额分割值对多笔交易金额进行聚类处理,得到多个金额聚类组。由于实际应用中人们更关心处于极端位置的交易金额的风险情况,因此可使得交易金额的聚类结果更加精准、快速。并且通过根据预设的交易金额评估指标从多个金额聚类组中确定交易金额评估指标对应的目标金额聚类组,进而根据目标金额聚类组对应的目标金额分割值确定交易金额对应的金额评估阈值,使得金额评估阈值能够基于预设的交易金额评估指标来确定,不仅提升了交易金额评估阈值的精确度,且使得交易金额的风险评估结果更加准确。

[0062] 上述对本说明书特定实施例进行了描述。其它实施例在所附权利要求书的范围内。在一些情况下,在权利要求书中记载的动作或步骤可以按照不同于实施例中的顺序来执行并且仍然可以实现期望的结果。另外,在附图中描绘的过程不一定要求示出的特定顺序或者连续顺序才能实现期望的结果。在某些实施方式中,多任务处理和并行处理也是可以的或者可能是有利的。

[0063] 以上为本说明书一个或多个实施例提供的资源数据的处理方法,基于同样的思路,本说明书一个或多个实施例还提供一种资源数据的处理装置。

[0064] 图3是根据本说明书一实施例的一种资源数据的处理装置的示意性框图。如图3所示,资源数据的处理装置包括:

[0065] 第一确定模块310,基于多个资源数据的分割位置,确定多个资源数据对应的至少一个资源分割值;资源分割值与资源数据的极端值之间的第一差异度小于分割位置与极端值之间的第二差异度;极端值包括资源数据中的最大数据值和/或最小数据值;

[0066] 聚类模块320,利用各资源分割值对多个资源数据进行聚类处理,得到多个资源聚类组;

[0067] 第二确定模块330,根据预设的资源评估指标,从多个资源聚类组中确定资源评估指标对应的目标资源聚类组;资源评估指标包含对资源数据进行风险评估所使用的风险评估参数;

[0068] 第三确定模块340,根据目标资源聚类组对应的目标资源分割值,确定资源数据对

应的资源评估阈值;资源评估阈值用于对资源数据进行风险评估。

[0069] 在一个实施例中,第一确定模块310包括:

[0070] 第一确定单元,确定对多个资源数据进行分割的至少一个分割位置;

[0071] 计算单元,根据各分割位置与资源分割值之间的映射关系,计算各分割位置分别对应的资源分割值。

[0072] 在一个实施例中,聚类模块320包括:

[0073] 排序单元,将各资源分割值按照预设维度进行排序;预设维度包括资源数据的数据大小;

[0074] 第二确定单元,基于排序后的各资源分割值,将每两个相邻的资源分割值确定为一个资源聚类组对应的边界值;

[0075] 第三确定单元,基于各资源聚类组分别对应的边界值,确定多个资源聚类组;

[0076] 划分单元,将各资源数据分别划分至对应的各资源聚类组中。

[0077] 在一个实施例中,第三确定模块340包括:

[0078] 第四确定单元,确定各目标资源分割值分别对应的权重;其中,各目标资源分割值分别对应的权重之和为1;

[0079] 计算单元,根据各目标资源分割值及各目标资源分割值分别对应的权重,计算资源数据对应的资源评估阈值。

[0080] 在一个实施例中,风险评估参数包含第一资源分割值;目标资源聚类组包含第一目标资源分割值及第二目标资源分割值;

[0081] 第四确定单元还用于:

[0082] 确定第一资源分割值对应的资源百分位数;

[0083] 确定第一资源分割值对应的资源百分位数为第一目标资源分割值对应的权重;

[0084] 计算资源百分位数与1之间的差值的绝对值;确定绝对值为第二目标资源分割值对应的权重。

[0085] 在一个实施例中,第二确定模块330包括:

[0086] 第五确定单元,从多个资源聚类组中确定第一资源分割值所在的资源聚类组;

[0087] 第六确定单元,将第一资源分割值所在的资源聚类组确定为目标资源聚类组。

[0088] 采用本说明书一个或多个实施例的装置,能够基于多个资源数据的分割位置确定多个资源数据对应的至少一个资源分割值,并利用各资源分割值对多个资源数据进行聚类处理,得到多个资源聚类组。可见,对资源数据进行分割时并不是基于原始的分割位置分割的,而是通过计算后的与资源数据的极端值之间的差异度更小的资源分割值进行分割,由于实际应用中人们更关心处于极端位置的资源数据,因此可使得资源数据的聚类结果更加精准、快速。并且通过根据预设的资源评估指标从多个资源聚类组中确定资源评估指标对应的目标资源聚类组,进而根据目标资源聚类组对应的目标资源分割值确定资源数据对应的资源评估阈值,使得资源评估阈值能够基于预设的资源评估指标来确定,不仅提升了资源评估阈值的精确度,且使得资源数据的风险评估结果更加准确。

[0089] 本领域的技术人员应可理解,上述资源数据的处理装置能够实现前文所述的资源数据的处理方法,其中的细节描述应与前文方法部分描述类似,为避免繁琐,此处不另赘述。

[0090] 基于同样的思路,本说明书一个或多个实施例还提供一种资源数据的处理设备,如图4所示。资源数据的处理设备可因配置或性能不同而产生比较大的差异,可以包括一个或一个以上的处理器401和存储器402,存储器402中可以存储有一个或一个以上存储应用程序或数据。其中,存储器402可以是短暂存储或持久存储。存储在存储器402的应用程序可以包括一个或一个以上模块(图示未示出),每个模块可以包括对资源数据的处理设备中的一系列计算机可执行指令。更进一步地,处理器401可以设置为与存储器402通信,在资源数据的处理设备上执行存储器402中的一系列计算机可执行指令。资源数据的处理设备还可以包括一个或一个以上电源403,一个或一个以上有线或无线网络接口404,一个或一个以上输入输出接口405,一个或一个以上键盘406。

[0091] 具体在本实施例中,资源数据的处理设备包括有存储器,以及一个或一个以上的程序,其中一个或者一个以上程序存储于存储器中,且一个或者一个以上程序可以包括一个或一个以上模块,且每个模块可以包括对资源数据的处理设备中的一系列计算机可执行指令,且经配置以由一个或者一个以上处理器执行该一个或者一个以上程序包含用于进行以下计算机可执行指令:

[0092] 基于多个资源数据的分割位置,确定所述多个资源数据对应的至少一个资源分割值;所述资源分割值与所述资源数据的极端值之间的第一差异度小于所述分割位置与所述极端值之间的第二差异度;所述极端值包括所述资源数据中的最大数据值和/或最小数据值;

[0093] 利用各所述资源分割值对所述多个资源数据进行聚类处理,得到多个资源聚类组;

[0094] 根据预设的资源评估指标,从所述多个资源聚类组中确定所述资源评估指标对应的目标资源聚类组;所述资源评估指标包含对所述资源数据进行风险评估所使用的风险评估参数;

[0095] 根据所述目标资源聚类组对应的目标资源分割值,确定所述资源数据对应的资源评估阈值;所述资源评估阈值用于对所述资源数据进行风险评估。

[0096] 可选地,计算机可执行指令在被执行时,还可以使所述处理器:

[0097] 确定对所述多个资源数据进行分割的至少一个所述分割位置;

[0098] 根据各所述分割位置与所述资源分割值之间的映射关系,计算各所述分割位置分别对应的所述资源分割值。

[0099] 可选地,计算机可执行指令在被执行时,还可以使所述处理器:

[0100] 将各所述资源分割值按照预设维度进行排序;所述预设维度包括所述资源数据的数据大小;

[0101] 基于排序后的各所述资源分割值,将每两个相邻的所述资源分割值确定为一个资源聚类组对应的边界值;

[0102] 基于各所述资源聚类组分别对应的所述边界值,确定多个所述资源聚类组;

[0103] 将各所述资源数据分别划分至对应的各所述资源聚类组中。

[0104] 可选地,计算机可执行指令在被执行时,还可以使所述处理器:

[0105] 确定各所述目标资源分割值分别对应的权重;其中,各所述目标资源分割值分别对应的权重之和为1;

[0106] 根据各所述目标资源分割值及各所述目标资源分割值分别对应的权重,计算所述资源数据对应的所述资源评估阈值。

[0107] 可选地,所述风险评估参数包含第一资源分割值;所述目标资源聚类组包含第一目标资源分割值及第二目标资源分割值;

[0108] 计算机可执行指令在被执行时,还可以使所述处理器:

[0109] 确定所述第一资源分割值对应的资源百分位数;

[0110] 确定所述第一资源分割值对应的资源百分位数为所述第一目标资源分割值对应的权重;

[0111] 计算所述资源百分位数与1之间的差值的绝对值;确定所述绝对值为所述第二目标资源分割值对应的权重。

[0112] 可选地,计算机可执行指令在被执行时,还可以使所述处理器:

[0113] 从所述多个资源聚类组中确定所述第一资源分割值所在的资源聚类组;

[0114] 将所述第一资源分割值所在的资源聚类组确定为所述目标资源聚类组。

[0115] 本说明书一个或多个实施例还提出了一种计算机可读存储介质,该计算机可读存储介质存储一个或多个程序,该一个或多个程序包括指令,该指令当被包括多个应用程序的电子设备执行时,能够使该电子设备执行上述资源数据的处理方法,并具体用于执行:

[0116] 基于多个资源数据的分割位置,确定所述多个资源数据对应的至少一个资源分割值;所述资源分割值与所述资源数据的极端值之间的第一差异度小于所述分割位置与所述极端值之间的第二差异度;所述极端值包括所述资源数据中的最大数据值和/或最小数据值;

[0117] 利用各所述资源分割值对所述多个资源数据进行聚类处理,得到多个资源聚类组;

[0118] 根据预设的资源评估指标,从所述多个资源聚类组中确定所述资源评估指标对应的目标资源聚类组;所述资源评估指标包含对所述资源数据进行风险评估所使用的风险评估参数;

[0119] 根据所述目标资源聚类组对应的目标资源分割值,确定所述资源数据对应的资源评估阈值;所述资源评估阈值用于对所述资源数据进行风险评估。

[0120] 上述实施例阐明的系统、装置、模块或单元,具体可以由计算机芯片或实体实现,或者由具有某种功能的产品来实现。一种典型的实现设备为计算机。具体的,计算机例如可以为个人计算机、膝上型计算机、蜂窝电话、相机电话、智能电话、个人数字助理、媒体播放器、导航设备、电子邮件设备、游戏控制台、平板计算机、可穿戴设备或者这些设备中的任何设备的组合。

[0121] 为了描述的方便,描述以上装置时以功能分为各种单元分别描述。当然,在实施本说明书一个或多个实施例时可以把各单元的功能在同一个或多个软件和/或硬件中实现。

[0122] 本领域内的技术人员应明白,本说明书一个或多个实施例可提供为方法、系统、或计算机程序产品。因此,本说明书一个或多个实施例可采用完全硬件实施例、完全软件实施例、或结合软件和硬件方面的实施例的形式。而且,本说明书一个或多个实施例可采用在一个或多个其中包含有计算机可用程序代码的计算机可用存储介质(包括但不限于磁盘存储器、CD-ROM、光学存储器等)上实施的计算机程序产品的形式。

[0123] 本说明书一个或多个实施例是参照根据本申请实施例的方法、设备(系统)、和计算机程序产品的流程图和/或方框图来描述的。应理解可由计算机程序指令实现流程图和/或方框图中的每一流程和/或方框、以及流程图和/或方框图中的流程和/或方框的结合。可提供这些计算机程序指令到通用计算机、专用计算机、嵌入式处理机或其他可编程数据处理设备的处理器以产生一个机器,使得通过计算机或其他可编程数据处理设备的处理器执行的指令产生用于实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能的装置。

[0124] 这些计算机程序指令也可存储在能引导计算机或其他可编程数据处理设备以特定方式工作的计算机可读存储器中,使得存储在该计算机可读存储器中的指令产生包括指令装置的制品,该指令装置实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能。

[0125] 这些计算机程序指令也可装载到计算机或其他可编程数据处理设备上,使得在计算机或其他可编程设备上执行一系列操作步骤以产生计算机实现的处理,从而在计算机或其他可编程设备上执行的指令提供用于实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能的步骤。

[0126] 在一个典型的配置中,计算设备包括一个或多个处理器(CPU)、输入/输出接口、网络接口和内存。

[0127] 内存可能包括计算机可读介质中的非永久性存储器,随机存取存储器(RAM)和/或非易失性内存等形式,如只读存储器(ROM)或闪存(flash RAM)。内存是计算机可读介质的示例。

[0128] 计算机可读介质包括永久性和非永久性、可移动和非可移动媒体可以由任何方法或技术来实现信息存储。信息可以是计算机可读指令、数据结构、程序的模块或其他数据。计算机的存储介质的例子包括,但不限于相变内存(PRAM)、静态随机存取存储器(SRAM)、动态随机存取存储器(DRAM)、其他类型的随机存取存储器(RAM)、只读存储器(ROM)、电可擦除可编程只读存储器(EEPROM)、快闪记忆体或其他内存技术、只读光盘只读存储器(CD-ROM)、数字多功能光盘(DVD)或其他光学存储、磁盒式磁带,磁带磁磁盘存储或其他磁性存储设备或任何其他非传输介质,可用于存储可以被计算设备访问的信息。按照本文中的界定,计算机可读介质不包括暂存电脑可读媒体(transitory media),如调制的数据信号和载波。

[0129] 还需要说明的是,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、商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、商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,并不排除在包括所述要素的过程、方法、商品或者设备中还存在另外的相同要素。

[0130] 本说明书一个或多个实施例可以在由计算机执行的计算机可执行指令的一般上下文中描述,例如程序模块。一般地,程序模块包括执行特定任务或实现特定抽象数据类型的例程、程序、对象、组件、数据结构等等。也可以在分布式计算环境中实践本申请,在这些分布式计算环境中,由通过通信网络而被连接的远程处理设备来执行任务。在分布式计算环境中,程序模块可以位于包括存储设备在内的本地和远程计算机存储介质中。

[0131] 本说明书中的各个实施例均采用递进的方式描述,各个实施例之间相同相似的部

分互相参见即可,每个实施例重点说明的都是与其他实施例的不同之处。尤其,对于系统实施例而言,由于其基本相似于方法实施例,所以描述的比较简单,相关之处参见方法实施例的部分说明即可。

[0132] 以上所述仅为本说明书一个或多个实施例而已,并不用于限制本说明书。对于本领域技术人员来说,本说明书一个或多个实施例可以有各种更改和变化。凡在本说明书一个或多个实施例的精神和原理之内所作的任何修改、等同替换、改进等,均应包含在本说明书一个或多个实施例的权利要求范围之内。

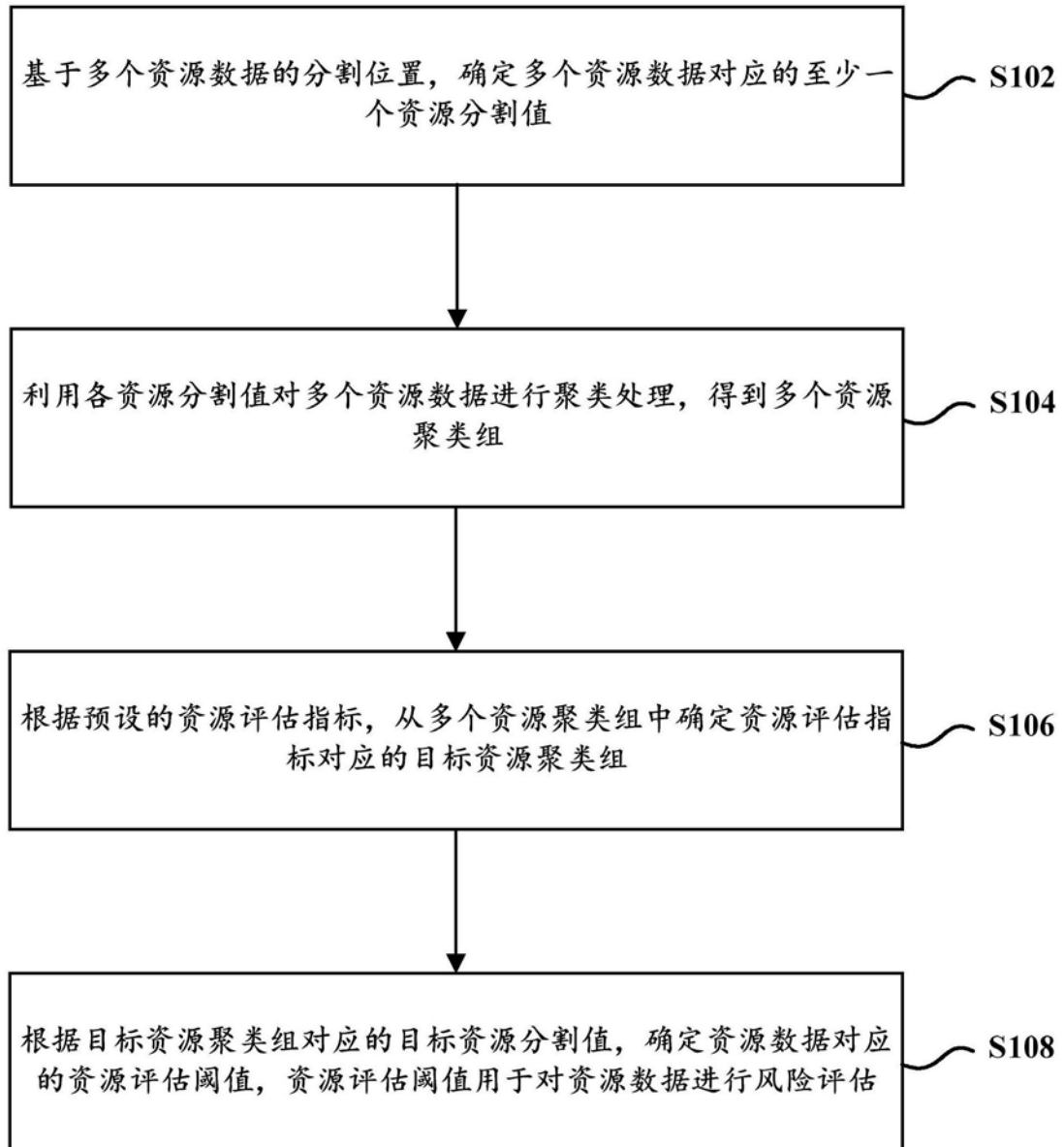


图1

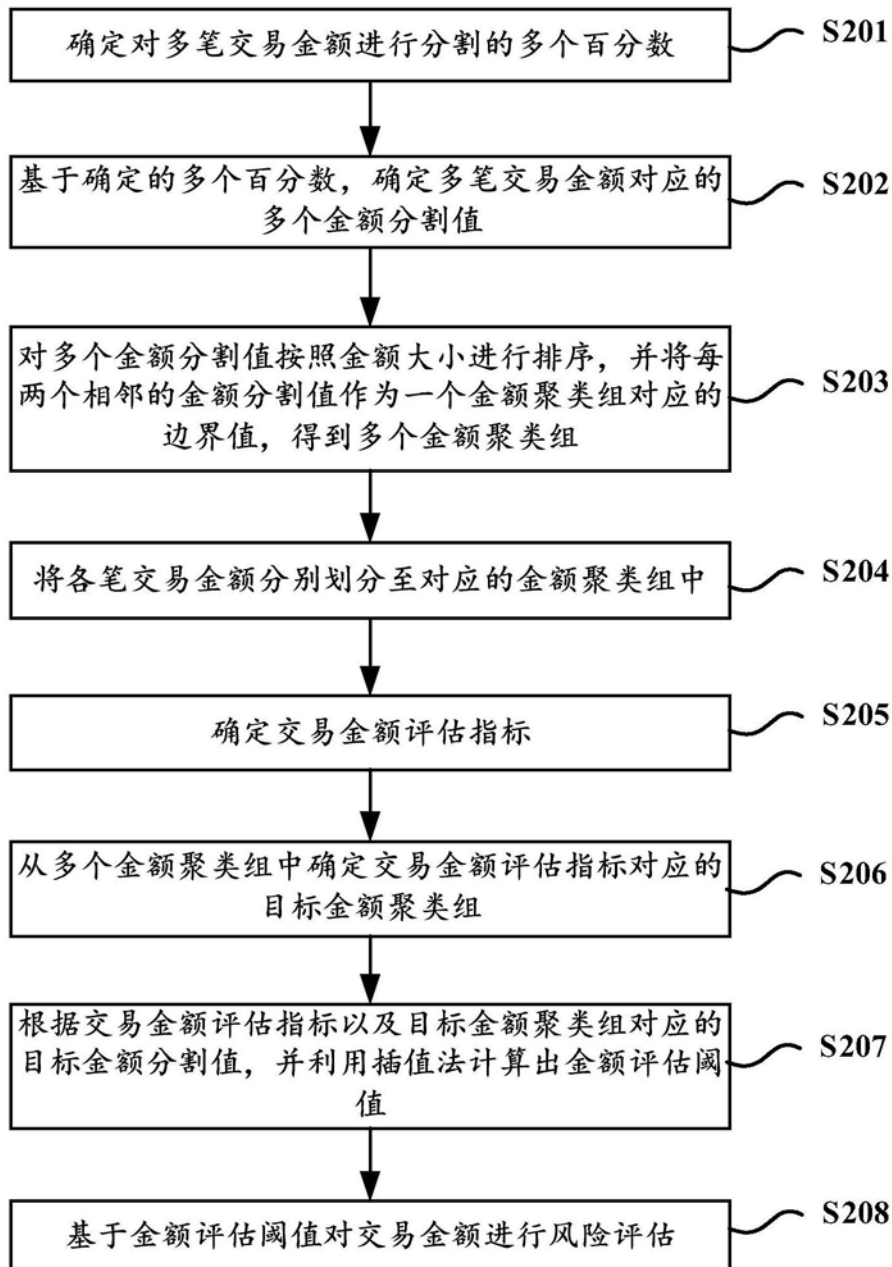


图2

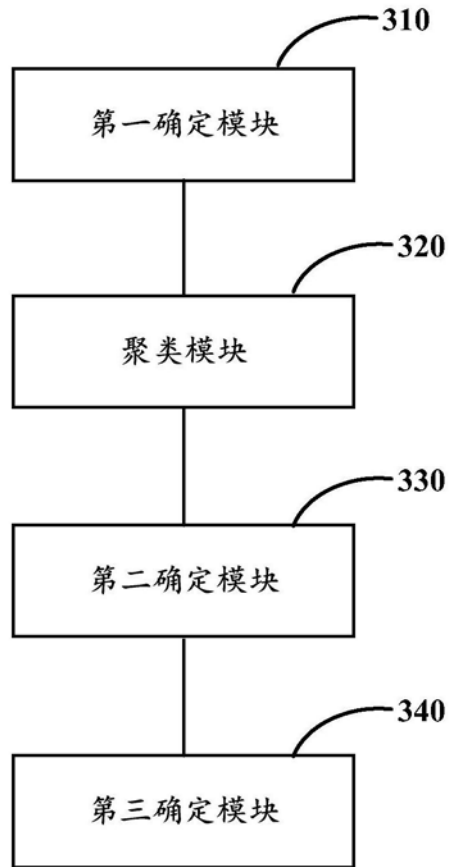


图3

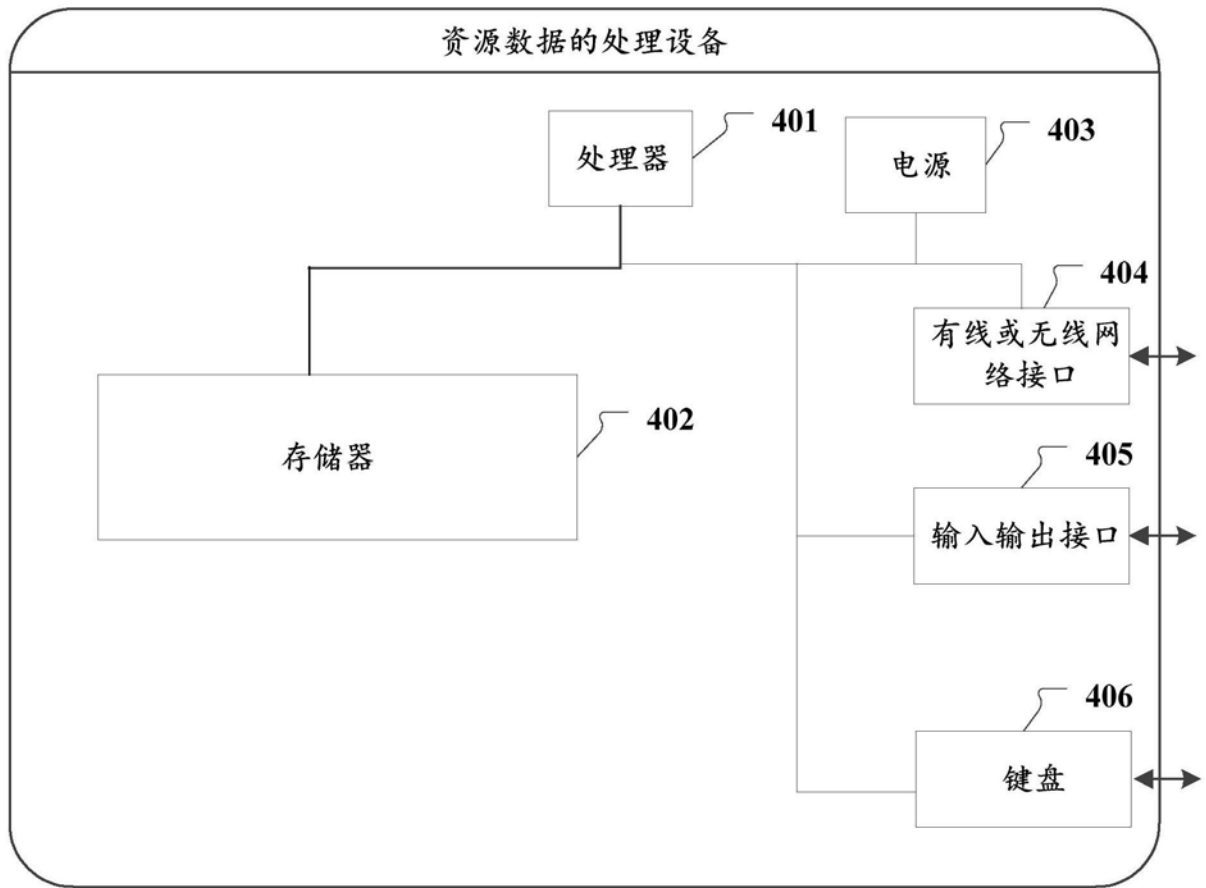


图4