(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 108140075 A
(43)申请公布日 2018.06.08

(21)申请号 201680044384.4

(22)申请日 2016.07.27

(30)优先权数据
14/810,328 2015.07.27 US

(85)PCT国际申请进入国家阶段日
2018.01.29

(86)PCT国际申请的申请数据
PCT/US2016/044198 2016.07.27

(87)PCT国际申请的公布数据
WO2017/019735 EN 2017.02.02

(71)申请人 皮沃塔尔软件公司
地址 美国加利福尼亚州

(72)发明人 余瑾 雷古纳坦·拉达克里希南 阿尼鲁德·孔达维蒂

(74)专利代理机构 中原信达知识产权代理有限责任公司 11219
代理人 李宝泉 周亚荣

(51)Int.Cl.
G06F 21/31(2006.01)
G06F 21/55(2006.01)
H04L 29/06(2006.01)

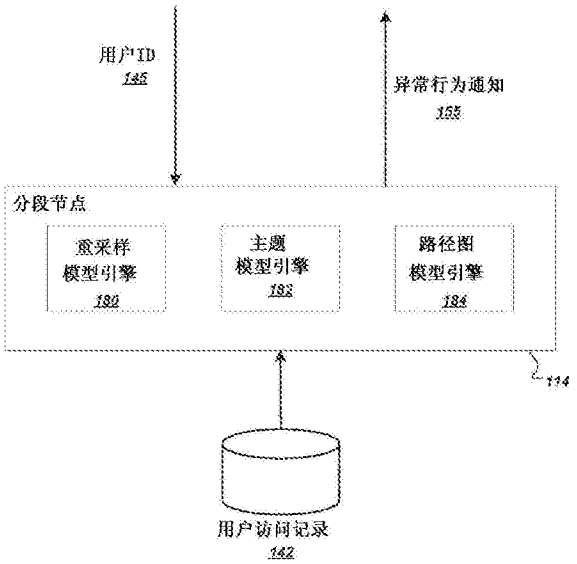
权利要求书3页 说明书12页 附图8页

(54)发明名称

将用户行为分类为异常

(57)摘要

用于将用户行为分类为异常的方法、系统和装置，包括在计算机存储介质上编码的计算机程序。一种方法包括获取表示在主体系统中用户的行为的用户行为数据。从训练数据生成初始模型，初始模型具有训练数据的第一特性特征。根据训练数据和测试时间段的第一表示的多个实例生成重采样模型。计算初始模型和重采样模型之间的差异。基于初始模型和重采样模型之间的差异，测试时间段中的用户行为被分类为异常。



CN 108140075 A

1.一种计算机实现的方法,包括:

获得表示在主体系统中用户的行为的用户行为数据,其中所述用户行为数据指示在所述主体系统中所述用户访问的一个或多个资源,以及对于所述用户访问的每个资源,该资源何时被访问;

从所述用户行为数据生成测试数据,所述测试数据包括在测试时间段期间由所述用户访问的资源的第一表示;

从所述用户行为数据生成训练数据,所述训练数据包括在所述测试时间段之前的多个时间段中的每个时间段中由所述用户访问的资源的相应的第二表示;

从所述训练数据生成初始路径图,其中所述初始路径图包括表示在由所述训练数据表示的一个或多个时间段期间在所述主体系统中由所述用户访问的资源的节点、以及一对或多对节点之间的链接,其中每对节点之间的每个链接表示所述用户从由该对节点的第二节点所表示的第二资源访问该对节点的第一节点所表示的第一资源;

从所述测试数据生成测试路径图,其中所述测试路径图包括表示所述测试时间段期间在所述主体系统中所述用户访问的资源的节点、以及一对或多对节点之间的链接,其中每对节点之间的每个链接表示所述用户从由该对节点的第二节点所表示的第二资源访问该对节点的第一节点所表示的第一资源;

计算所述初始路径图与所述测试路径图之间的差异;以及

基于所述初始路径图和所述测试路径图之间的所述差异,将所述用户在所述测试时间段内的所述用户行为分类为异常。

2.根据权利要求1所述的方法,其中,所述用户行为数据包括用户访问记录,每个所述用户访问记录表示在文件系统中所述用户访问的文件夹或文件。

3.根据权利要求1所述的方法,其中,生成所述初始路径图包括从所述用户的训练数据和在所述主体系统中所述用户的一个或多个对等方的训练数据生成所述初始路径图。

4.根据权利要求3所述的方法,进一步包括:

确定所述主体系统中的一个或多个其他用户,所述一个或多个其他用户在由所述训练数据表示的时间段期间访问与所述用户共同的至少阈值数量的资源;以及

指定所述一个或多个其他用户作为在所述主体系统中所述用户的对等方。

5.根据权利要求1所述的方法,其中,计算所述初始路径图与所述测试路径图之间的差异包括计算所述初始路径图与所述测试路径图之间的Jaccard距离,其中所述Jaccard距离是基于在所述初始路径图和所述测试路径图之间节点的交集基数、以及在所述初始路径图和所述测试路径图之间的节点的并集基数。

6.根据权利要求1所述的方法,其中,计算所述初始路径图与所述测试路径图之间的所述差异包括:

获得与由所述初始路径图和所述测试路径图中的节点表示的资源相关联的权重;以及

计算所述初始路径图与所述测试路径图之间的加权Jaccard距离,其中所述加权Jaccard距离是基于在所述初始路径图与所述测试路径图的交集中出现的所有节点的权重之和、以及在所述测试路径图中出现的所有节点的权重之和。

7.根据权利要求6所述的方法,进一步包括向所述主体系统中的文件夹指派比所述主体系统中的所述文件夹的子文件夹高的权重。

8.根据权利要求6所述的方法，进一步包括向所述主体系统中高于所述资源的层级中的阈值数量的级别的所有资源指派相同的权重。

9.根据权利要求6所述的方法，其中，所述权重是基于所述资源的受欢迎程度的度量。

10.根据权利要求9所述的方法，进一步包括：

生成混合图，其中所述混合图包括表示所述系统中的用户的用户节点和表示所述系统中的资源的资源节点，其中所述混合图包括用户-资源链接和资源-资源链接，其中每个用户-资源链接表示访问所述系统中的资源的相应用户，其中每个资源-资源链接表示所述系统中的资源的结构；

根据所述混合图计算所述系统中的一个或多个资源的受欢迎程度的度量；

选择具有受欢迎程度的最高度量的一个或多个节点；以及

将到具有受欢迎程度的所述最高度量的所述一个或多个节点中的每一个节点的路径添加到针对所述用户的所述初始路径图。

11.一种计算机实现的方法，包括：

获得多个主题，每个主题是表示在个体用户的用户行为数据中频繁同时出现的多个文件类型的数据；

获得表示在主体系统中用户的行为的用户行为数据，其中所述用户行为数据指示在所述主体系统中所述用户访问的文件的文件类型以及所述文件何时被所述用户访问；

从所述用户行为数据生成测试数据，所述测试数据包括根据所述用户行为数据的所述文件类型在测试时间段期间所述用户访问的主题的第一表示；

从所述用户行为数据生成训练数据，所述训练数据包括在所述测试时间段之前的多个时间段的每一个时间段中所述用户访问的主题的相应的第二表示；

从所述测试数据生成初始SVD模型；

从来自在所述测试时间段期间所述用户访问的主题的所述第一表示的多个实例的所述训练数据生成重采样模型；

计算所述初始模型和所述重采样模型之间的差异；以及

基于所述初始模型和所述重采样模型之间的所述差异，将所述测试时间段中的所述用户行为分类为异常。

12.根据权利要求11所述的方法，进一步包括根据在所述主体系统中多个用户访问的文件的文件类型生成所述多个主题。

13.根据权利要求12所述的方法，进一步包括：

使用主题建模过程来生成所述主题，所述主题建模过程包括将每个用户定义为文档并且将每个用户访问的每个文件类型定义为对应文档中的项。

14.根据权利要求13所述的方法，其中，使用所述主题建模过程来生成所述主题包括生成预定数量K个主题。

15.根据权利要求13所述的方法，进一步包括：

对K的多个候选值进行迭代；以及

选择K的特定候选值作为所述预定数量K。

16.一种计算机实现的方法，包括：

获得表示在主体系统中用户的行为的用户行为数据，其中所述用户行为数据指示在所

述主体系统中所述用户访问的一个或多个资源，以及对于所述用户访问的每个资源，该资源何时被访问；

从所述用户行为数据生成测试数据，所述测试数据包括在测试时间段期间由所述用户访问的资源的第一表示；

从所述用户行为数据生成训练数据，所述训练数据包括在所述测试时间段之前的多个时间段中的每个时间段由所述用户访问的资源的相应的第二表示；

从所述训练数据生成初始模型，所述初始模型具有所述训练数据的第一特性特征；

从所述训练数据以及从所述测试时间段的所述第一表示的多个实例生成重采样模型，所述重采样模型具有所述训练数据的第二特性特征和所述测试时间段的所述第一表示的所述多个实例；

计算所述初始模型和所述重采样模型之间的差异，包括比较所述训练数据的所述第一特性特征和所述训练数据的所述第二特性特征以及所述测试时间段的所述第一表示的所述多个实例；以及

基于所述初始模型和所述重采样模型之间的所述差异，将所述测试时间段中的所述用户行为分类为异常。

17.根据权利要求16所述的方法，其中，所述用户行为数据包括用户访问记录，每个所述用户访问记录表示在文件系统中所述用户访问的文件夹或文件。

18.根据权利要求16所述的方法，进一步包括：

生成包括所述训练数据的向量和所述测试数据的相同向量的N个实例的第一矩阵；以及

对所述第一矩阵执行主分量分析以生成所述第一矩阵的第一多个主分量；

从所述训练数据的多个向量生成第二矩阵；以及

对所述第二矩阵执行主分量分析以生成所述第二矩阵的第二多个主分量，

其中，计算所述初始模型和所述重采样模型之间的差异包括计算所述第一多个主分量和所述第二多个主分量中的一个或多个主分量之间的角度。

19.根据权利要求16所述的方法，进一步包括：

生成包括所述训练数据的向量和所述测试数据的相同向量的N个实例的第一矩阵；

对所述第一矩阵执行奇异值分解以生成所述第一矩阵的第一多个主分量；

从所述训练数据的多个向量生成第二矩阵；以及

对所述第二矩阵执行奇异值分解以生成所述第二矩阵的第二多个主分量，

其中，计算所述初始模型和所述重采样模型之间的差异包括计算所述第一多个主分量和所述第二多个主分量中的一个或多个主分量之间的角度。

20.一种系统，包括：一个或多个计算机和存储指令的一个或多个存储设备，所述指令能操作为在由所述一个或多个计算机执行时使得所述一个或多个计算机执行根据权利要求1至19中的任一项所述的方法。

21.一种编码有计算机程序的计算机存储介质，所述程序包括指令，所述指令能操作为在由数据处理装置执行时使所述数据处理装置执行根据权利要求1至19中的任一项所述的方法。

将用户行为分类为异常

背景技术

[0001] 本说明书涉及检测大型数据集中的异常。

[0002] 用于检测大数据集中的异常的技术可以用于数据处理应用的多个领域，包括计算机网络安全和保健。

发明内容

[0003] 本说明书描述了数据处理系统可以如何根据利用指示在一个或多个特定数据处理系统中用户访问的资源的数据的各种技术，将用户行为分类为异常或非异常。即使用户可能有权访问所有访问的资源，系统仍然可以将某些用户的行为归类为可疑的。

[0004] 通常，本说明书中描述的主题的一个创新方面可以体现在包括以下动作的方法：获得表示在主体系统中用户的行为的用户行为数据，其中所述用户行为数据指示所述用户在主体系统中访问的一个或多个资源，以及对于用户访问的每个资源，该资源何时被访问；从用户行为数据生成测试数据，所述测试数据包括在测试时间段期间由所述用户访问的资源的第一表示；从用户行为数据生成训练数据，所述训练数据包括在测试时间段之前的多个时间段中的每个时间段中由用户访问的资源的相应的第二表示；从训练数据生成初始模型，所述初始模型具有训练数据的第一特性特征；从训练数据以及从测试时间段的第一表示的多个实例生成重采样模型，所述重采样模型具有训练数据的第二特性特征和测试时间段的第一表示的多个实例；计算初始模型和重采样模型之间的差异，包括比较训练数据的第一特性特征和训练数据的第二特性特征以及测试时间段的第一表示的多个实例；以及基于初始模型和重采样模型之间的差异，将测试时间段中的用户行为分类为异常。这个方面的其他实施例包括记录在一个或多个计算机存储设备上的对应的计算机系统、装置和计算机程序，每个计算机程序被配置为执行这些方法的动作。对于被配置为执行特定操作或动作的一个或多个计算机系统，意味着该系统上已经安装了软件、固件、硬件或者它们的组合，这些软件、固件、硬件或者它们的组合在操作中使系统执行操作或动作。对于被配置为执行特定操作或动作的一个或多个计算机程序意味着所述一个或多个程序包括在由数据处理装置执行时使装置执行操作或动作的指令。

[0005] 前述和其它实施例可以各自任选地包括单独或组合的一个或多个以下特征。用户行为数据包括用户访问记录，每个用户访问记录表示在文件系统中用户访问的文件夹或文件。该动作包括对第一矩阵执行主分量分析以生成第一矩阵的第一多个主分量；从训练数据的多个向量生成第二矩阵；和对第二矩阵执行主分量分析以生成第二矩阵的第二多个主分量，其中计算初始模型和重采样模型之间的差异包括计算第一多个主分量和第二多个主分量中的一个或多个主分量之间的角度。该动作包括生成包括训练数据的向量和测试数据的相同向量的N个实例的第一矩阵；对第一矩阵执行奇异值分解以生成第一矩阵的第一多个主分量；从训练数据的多个向量生成第二矩阵；和对第二矩阵执行奇异值分解以生成第二矩阵的第二多个主成分，其中计算初始模型和重采样模型之间的差异包括计算第一多个主分量和第二多个主分量中的一个或多个主分量之间的角度。

[0006] 本说明书中描述的主题的另一个创新方面可以体现在在包括以下动作的方法：获得多个主题，每个主题是表示在个体用户的用户行为数据中频繁同时出现的多个文件类型的数据；获得表示在主体系统中用户的行为的用户行为数据，其中所述用户行为数据指示在主体系统中所述用户访问的文件的文件类型以及所述文件何时被所述用户访问；从用户行为数据生成测试数据，所述测试数据包括根据用户行为数据的文件类型在测试时间段期间所述用户访问主题的第一表示；从用户行为数据生成训练数据，所述训练数据包括所述用户在测试时间段之前的多个时间段的每一个时间段中访问的主题的相应的第二表示；从测试数据生成初始SVD模型；从来自在测试时间段期间用户访问主题的第一表示的多个实例的训练数据生成重采样模型；计算初始模型和重采样模型之间的差异；和基于初始模型和重采样模型之间的差异，将测试时间段中的用户行为分类为异常。这个方面的其他实施例包括记录在一个或多个计算机存储设备上的相应的计算机系统、装置和计算机程序，每个计算机程序被配置为执行这些方法的动作。

[0007] 前述和其它实施例可以各自任选地包括单独或组合的一个或多个以下特征。这些动作包括根据在主体系统中多个用户访问的文件的文件类型中生成多个主题。所述动作包括使用主题建模过程来生成所述主题，所述主题建模过程包括将每个用户定义为文档，并且将每个用户访问的每个文件类型定义为对应文档中的项。使用主题建模过程生成主题包括生成预定数量K的主题。这些动作包括对K的多个候选值进行迭代；和选择K的特定候选值作为预定数量K。

[0008] 本说明书中描述的主题的另一个创新性方面可以体现在包括以下动作的方法：获得表示在主体系统中用户的行为的用户行为数据，其中所述用户行为数据指示在所述主体系统中所述用户访问的一个或多个资源，以及对于所述用户访问的每个资源，该资源何时被访问；从用户行为数据生成测试数据，所述测试数据包括在测试时间段期间由所述用户访问的资源的第一表示；从用户行为数据生成训练数据，所述训练数据包括在测试时间段之前的多个时间段中的每个时间段中由用户访问的资源的相应的第二表示；从训练数据生成初始路径图，其中所述初始路径图包括表示在由训练数据表示的一个或多个时间段期间在主体系统中由用户访问的资源的节点、以及一对或多对节点之间的链接，其中每对节点之间的每个链接表示用户从由所述对节点的第二节点所表示的第二资源访问所述对节点的第一节点所表示的第一资源；从测试数据生成测试路径图，其中所述测试路径图包括表示测试时间段期间用户在实主体系统中访问的资源的节点以及一对或多对节点之间的链接，其中每对节点之间的链接表示用户从由所述对节点的第二节点所表示的第二资源访问所述对节点的第一节点所表示的第一资源；计算初始路径图与测试路径图之间的差异；和基于初始路径图和测试路径图之间的差异，将用户在测试时间段内的用户行为分类为异常。

[0009] 前述和其它实施例可以各自任选地包括单独或组合的一个或多个以下特征。用户行为数据包括用户访问记录，每个用户访问记录表示在文件系统中用户访问的文件夹或文件。生成初始路径图包括从用户的训练数据和用户在主体系统中的一个或多个对等方的训练数据生成初始路径图。所述动作包括确定主体系统中的一个或多个其他用户，所述其他用户在由训练数据表示的时间段期间访问与所述用户共同的至少阈值数量的资源；和指定一个或多个其他用户作为所述用户在主体系统中的对等方。计算初始路径图与测试路径图

之间的差异包括计算初始路径图与测试路径图之间的Jaccard距离，其中所述Jaccard距离是基于在初始路径图和测试路径图之间节点的交集基数、以及在初始路径图和测试路径图之间的节点的并集基数。计算初始路径图与测试路径图之间的差异包括获得由与初始路径图和测试路径图中的节点表示的资源相关联的权重；和计算初始路径图与测试路径图之间的加权Jaccard距离，其中所述加权Jaccard距离是基于在初始路径图与测试路径图的交集中出现的所有节点的权重之和、以及测试路径图中出现的所有节点的权重之和。这些操作包括向主体系统中的文件夹指派比所述主体系统中的所述文件夹的子文件夹高的权重。这些操作包括向主体系统中高于所述资源的层级中的阈值数量的级别的所有资源指派相同的权重。权重是基于所述资源的受欢迎程度的度量。该动作包括生成混合图，其中所述混合图包括表示系统中的用户的用户节点和表示系统中的资源的资源节点，其中所述混合图包括用户-资源链接和资源-资源链接，其中每个用户-资源链接表示访问系统中的资源的相应用户，其中每个资源-资源链接表示系统中资源的结构；根据所述混合图计算系统中的一个或多个资源的受欢迎程度的度量；选择具有受欢迎程度的最高度量的一个或多个节点；和将到具有受欢迎程度的最高度量的所述一个或多个节点中的每一个节点的路径添加到针对所述用户的所述初始路径图。

[0010] 可以实现本说明书中描述的主题的特定实施例以实现以下优点中的一个或多个。系统可以将用户访问模式分类为异常，即使之前没有看到过这种模式，而这是基于规则的系统无法做到的。系统可以使用测试数据重采样来使异常检测比先前的方法更加敏感。系统可以为系统中的每个用户生成用户模型，并自动将用户的行为标记为异常。用户对等方的行为可以被纳入分析，以减少异常检测中的误报。系统可以比以前的方法更细粒度地使用数据，例如，它可以使用描述文件夹访问和文件访问的数据。系统可以使用主题建模来检测用户何时访问意料之外的文件类型的组。

[0011] 在下面的附图和描述中阐述了本说明书的主题的一个或多个实施例的细节。本主题的其它特征、方面和优点将从描述，附图和权利要求中变得显而易见。

附图说明

[0012] 图1A是示例异常检测系统的图。

[0013] 图1B是分段节点的图。

[0014] 图2是使用重采样模型将用户访问记录分类为异常的示例过程的流程图。

[0015] 图3是使用主题模型将用户行为分类为异常的示例过程的流程图。

[0016] 图4是使用路径图将用户行为分类为异常的示例过程的流程图。

[0017] 图5A图示了初始路径图。

[0018] 图5B图示了示例测试路径图。

[0019] 图5C图示了另一个示例测试路径图。

[0020] 图6是用于确定主体系统中最受欢迎的资源的示例过程的流程图。

[0021] 图7图示了示例混合图。

[0022] 在各个附图中相同的附图标号和标记指示相同的元件。

具体实施方式

[0023] 图1A是示例异常检测系统100的图。异常检测系统100是可以用于检测异常用户行为的计算系统的示例。通常，异常检测系统100包括用户设备102，主节点110和多个分段节点114a、114b到114n。

[0024] 待检测的异常用户行为通常是与异常检测系统100不同的主体系统中的用户的行为。例如，主体系统可以是属于公司的计算机网络。

[0025] 用户设备102的用户可以通过与主节点110通信来访问存储在异常检测系统100中的数据。用户设备102可以是个人计算机、智能手机或用户可以与之交互的任何其他类型的基于计算机的设备。例如，用户可以针对主节点查询在特定时间段(例如前一天或前一周)期间发生的异常用户行为。然后主节点110可以与分段节点114a-n通信以获得在该特定时间段期间的行为可疑的用户的识别，然后主节点110可以向用户设备102通信。

[0026] 主节点110和每个分段节点114a-n被实现为安装在一个或多个物理计算机上的软件，或者作为在一个或多个物理计算机上安装为一个或多个虚拟机的软件，或两者。另外，每个分段节点114a n可以在分段节点内执行多个分段过程。例如，分段节点可以是多核心计算机，其中每个分段过程在不同的核心上执行。在一些实现中，每个物理分段节点具有8到12个分段过程。

[0027] 主节点110例如通过一个或多个通信网络(例如局域网或互联网)或通过直接连接而连接到每个分段节点114a-n。另外，每个分段节点114a-n可以连接到一个或多个其他分段节点。主节点110指派每个分段节点以对存储在异常检测系统100中的数据的一部分进行操作。

[0028] 每个数据部分通常是主体系统中的用户的用户行为数据的集合。为了利用分段节点114a-n的并行处理，可以将每个不同用户的所有用户行为数据存储在单个部分中。然而，分段节点114a-n也可以彼此通信以共享信息，使得单个分段节点可以获得特定用户的所有用户行为数据。

[0029] 用户行为数据是表示主体系统中的用户访问资源的数据。例如，数据可以表示用户访问主体系统中的服务器、网站、网页、文件、目录、数据库或任何其他可访问资源的次数。

[0030] 用户访问资源的每个实例例如通过访问记录在用户行为数据中表示。访问记录可以包括描述资源、用户以及访问资源的日期和时间的信息。用户行为数据还可以包括聚集的访问记录。例如，对于每个用户，用户行为数据可以包括表示在特定时间段期间每个资源被访问多少次的数据。

[0031] 系统100可以以任何适当的格式存储数百万或数十亿的访问记录。例如，系统可以将每个访问记录存储为文件系统中的文件或文件系统中的文件中的一行或一行数据或数据库中的记录。访问记录可以被索引。

[0032] 主节点110可以划分在N个分段节点(例如，分段节点114a-n)之间的处理。分段节点可以通过与底层分布式存储系统中的数据节点(例如，实现Hadoop文件系统(HDFS)的数据节点)通信来获得访问记录。数据通常在多个存储设备之间进行分区，并且可以由任何适当的键值存储子系统来组织。例如，数据部分可以是分布在多个存储设备之间的关系数据库的表分区，例如作为大规模并行处理(MPP)数据库的一部分。数据部分也可以作为分布式非关系数据库的一部分存储，例如存储在Hadoop数据库(HBase)中，该Hadoop数据库

(HBase)通过不同列家族中的键值对来组织数据并分布在多个存储设备上。数据部分也可以被分区以由分段节点114a-n本地存储。

[0033] 在图1A所示的示例异常检测系统100中，主节点110已经被指派了分段节点114a，以对存储在底层分布式存储系统的第一存储子系统132a中的第一组用户的访问记录142a进行操作。类似地，主节点110已经被指派了分段节点142b以对存储在第二存储子系统142b中的访问记录142b进行操作，并且主节点110已经被指派了分段节点114n来对存储在第N个存储子系统132n中的访问记录142n进行操作。

[0034] 图1B是分段节点114的图。系统中的每个分段节点114a-n并行计算每个用户模型。换句话说，系统为主体系统的每个用户生成一个或多个不同的模型。

[0035] 每个分段节点114运行安装在分段节点114上的接收由主节点指派的用户ID 145的异常检测软件。然后，异常检测软件从底层存储子系统获得与用户ID 145对应的用户的用户访问记录142，并确定哪些访问记录142是训练数据而哪些访问记录是测试数据。一些个人访问记录可以用作训练数据和测试数据两者。

[0036] 通常，测试数据包括用户在最近时间段内访问的资源的表示，并且训练数据包括在测试数据的时间段之前的多个时间段内访问的资源的表示。例如，如果时间段是一年中的几周，则测试数据可以包括最近一周期间访问的资源的表示，并且训练数据可以包括在前一个月或一年中访问的资源的表示。对应于测试数据的时间段可以被称为测试时间段。

[0037] 然而，测试数据不一定表示最近的时间段。例如，系统可以使用任何适当时间段的访问记录作为测试数据，以便识别过去发生的异常行为。

[0038] 安装在分段节点114上的异常检测软件利用安装在每个分段节点上的一个或多个建模引擎180，182和184来确定用户的访问记录142是否反映用户的异常行为。所有建模引擎180，182和184或者仅建模引擎180，182和184中的一些可以已经被安装在任何特定的分段节点上。

[0039] 分段节点114可以使用重采样模型引擎180，其重采样一些测试数据作为训练数据。下面参照图2更详细地描述重采样模型。分段节点114还可以使用主题模型引擎182，其基于由用户访问的文件类型生成主题模型。以下参照图3更详细地描述主题模型。分段节点114还可以使用路径图模型引擎184，其从训练数据和测试数据建立路径图以确定异常行为。路径图在下面参照图4-7更详细地描述。

[0040] 系统可以使用建模引擎180，182和184将用户访问记录142的测试数据分类为异常或非异常。如果测试数据被分类为异常，则分段节点114可以生成异常行为通知155，并将通知155提供给系统中的另一个节点，例如提供回主节点110。主节点110然后可以将该通知传播回到用户设备102。

[0041] 图2是使用重采样模型将用户访问记录分类为异常的示例过程的流程图。通常，系统确定在多次重采样时多少测试数据会影响用户访问行为的初始统计模型的特征。示例过程将被描述为由一个或多个计算机的适当编程的系统执行。

[0042] 系统获得用户的访问记录(210)。如上所述，访问记录指示在多个时间段中的每个时间段期间用户访问哪些资源。

[0043] 系统生成用户的访问记录的表示(220)。在一些实现中，该表示是向量或矩阵，并且系统为多个时间段中的每一个生成向量。向量中的每个位置表示主体系统中的资源，并

且向量中的每个值表示用户访问主体系统中与向量中的值的位置相对应的资源的次数。

[0044] 系统使用训练数据生成初始模型(230)。如上所述，训练数据包括先前时间段的用户访问记录的表示。

[0045] 系统可以生成初始模型作为任何适当的统计模型来表示数据集的特性特征。在一些实现中，系统将数据表示为矩阵，并使用任何适当的矩阵分解技术(例如奇异值分解(SVD)、主分量分析(PCA)或非负矩阵分解(NMF))来生成用于用户的训练数据的特性特征的表示。

[0046] 例如，系统可以从训练数据生成访问记录向量的矩阵X。系统然后可以执行SVD来生成表示X的主分量的矩阵T。

[0047] 系统从多次采样的训练数据和测试数据生成重采样模型(240)。对测试数据进行多次重新采样，具有放大训练数据与测试数据之间差异的效果。

[0048] 例如，如果使用SVD生成重采样模型，则系统可以使用训练数据的所有向量和测试数据的向量的N个实例。换句话说，系统可以生成包括训练数据的向量和测试数据的向量的N个实例的矩阵X'。通常，对于初始模型，矩阵X'将包括比矩阵X更多的列。系统然后可以执行SVD以生成表示X'的主分量的矩阵T'。

[0049] 系统比较初始模型和重采样模型(250)。系统可以使用任何适当的比较方法，例如通过计算模型的特性特征之间的距离来确定初始模型和重采样模型之间不同的程度。如果使用SVD，则系统可以计算初始模型T的主分量与重采样模型T的主分量之间的角度。

[0050] 系统基于比较将测试时段中的用户行为分类为异常或非异常(260)。只要相对于初始模型测试数据对重采样模型有更显著的影响，那么初始模型和重采样模型之间的差异就会很大。因此，当差异较大时，测试数据更可能是异常的。

[0051] 但是，如果初始模型和重采样模型之间的差异很小，那么测试数据对仅从训练数据生成的初始模型具有最小的影响。因此，测试数据不太可能是异常的。

[0052] 因此，系统可以确定模型之间的差异是否满足阈值，并且如果差异满足阈值，则将用户行为分类为异常。

[0053] 图3是使用主题模型将用户行为分类为异常的示例过程的流程图。在这个示例过程中，系统根据用户访问的相关文件类型的组而不是根据用户访问的资源来表示用户在主体系统中的行为。如果测试数据指示用户在测试时段访问了基本上不同的文件类型，则相关文件类型的组可以被表示为主题，并且系统可以将用户的行为分类为异常。该过程将被描述为由一个或多个计算机的适当编程的系统执行。

[0054] 系统从主体系统中的文件生成主题(310)。系统可以生成主题，其中每个主题表示在个人用户的用户访问记录中频繁同时出现的文件类型的组。通常，系统使用来自许多不同用户的用户访问记录来生成主题。

[0055] 在一些实现中，系统使用文件的扩展来指示文件的类型。然而，系统可以使用关于系统中的文件的其他元数据来确定文件类型。

[0056] 系统可以通过将每个用户视为文档，并将用户访问的每个文件类型视为文档中出现的术语来使用任何适当的主题建模技术。系统可以在表示主体系统中用户访问文件的所有用户访问记录上使用主题建模技术。因此结果是一些各表示频繁发生的文件类型的主题。系统可以为每个发现的主题指派唯一的标识符。

[0057] 例如，系统可以使用隐含狄利克雷分配(LDA)生成K个主题。LDA将K个主题作为输入参数，并为K个主题中的每个生成概率分布。每个概率分布为访问指派给该主题的文件类型的用户所访问的特定文件类型指派似然率。

[0058] 系统可以通过对K的候选值进行迭代并计算模型的困惑度来为K选择值。系统可以为K选择值来平衡模型中的主题数量和模型的困惑度。

[0059] 系统获得用户的访问记录(320)。访问记录可以指示用户访问的电子文件和用户访问的文件的文件类型信息。

[0060] 系统生成用户的访问记录的表示(330)。系统可以为几个时间段中的每一个生成向量。向量的每个元素表示K个主题中的一个，并且向量中的每个值表示用户访问属于每个对应主题的文件类型的次数。在一些实现中，每个元素表示用户访问属于每个对应主题的文件类型的每个时间段中的天数。

[0061] 系统使用从训练数据生成的初始SVD模型来重建测试数据(340)。如上所述，系统可以使用任何适当的统计模型(例如SVD或PCA)来表示训练数据和测试数据的特性特征。

[0062] 系统可以类似地使用上面参照图2描述的重采样技术来确定当测试数据被多次添加到训练数据时初始SVD模型相对于重采样模型如何改变。

[0063] 在一些实现中，系统可以使用奇异值分解(SVD)来比较模型。例如，系统可以从训练数据中获得矩阵X，其中每列表示训练数据中的时间段，并且每行表示K个主题中的一个。然后系统可以执行SVD以生成矩阵Y。然后系统可以从Y中选择前k个右奇异列向量V作为在训练时间段期间表示用户的行为。

[0064] 然后系统可以根据以下公式通过来算距离D来比较测试数据，所述测试数据被表示为训练数据的向量：

[0065] $D=||x_{t+1}-V\times(V^{T}\times x_{t+1})||$

[0066] 系统基于比较将测试时段中的用户行为分类为异常或非异常(350)。如果差异满足阈值，则系统可以将用户行为分类为异常。否则，系统可以将用户行为分类为非异常。

[0067] 图4是使用路径图将用户行为分类为异常的示例过程的流程图。路径图是用户在相关时间段期间如何导航到主体系统中的资源的表示。如果路径图在测试时段显著变化，则系统可以将用户行为分类为异常。该过程将被描述为由一个或多个计算机的适当编程的系统执行。

[0068] 系统使用训练数据生成初始路径图(410)。路径图表示用户在主体系统中访问的资源之间的关系。

[0069] 路径图包括表示用户在主体系统中访问的资源的节点。例如，路径图的节点可以表示文件系统中的文件夹和文件。路径图的节点也可以表示由主体系统维护的网页。

[0070] 路径图包括两个节点之间的链接以表示访问一个节点而不是另一个节点的用户。换句话说，路径图包括表示访问由第一节点表示的第一资源，然后访问由第二节点表示的第二资源的用户的链接。因此，链接可以表示文件夹和子文件夹之间的关系、网页之间的链接、文件系统中的符号链接或快捷方式，或者用于访问一个资源而不是另一个的任何其他合适的方法。

[0071] 图5A图示了初始路径图。在此示例中，路径图的节点表示文件系统中的文件夹，链接表示从父文件夹访问子文件夹的用户。

[0072] 初始路径图具有表示“家”目录的根节点510。初始路径图还包括表示“家”目录的子文件夹的其他节点520，522和530。

[0073] 节点510和节点520之间的链接表示用户访问“家”目录，然后从“家”目录访问“文件夹B”。类似地，节点520和节点530之间的链接表示用户从“文件夹”目录访问“子文件夹C”目录。

[0074] 因此，当系统使用训练数据生成初始路径图时，作为结果的初始路径图包括表示用户访问的资源的节点以及表示用户如何导航到这些资源的链接。

[0075] 当生成初始路径图时，系统还可以包括来自用户对等方的数据。在一些情况下，使用具有来自用户对等方的数据的初始路径图可以减少异常行为的误报检测。

[0076] 用户的对等方通常是主体系统中与该用户在访问的资源方面有显著重叠的用户。例如，用户的对等方可以是组织内同一团队中的其他成员，也可以是同一部门、地点或公司中的其他员工。

[0077] 在一些实施方式中，系统通过识别具有至少阈值量的资源重叠的其他用户来确定用户的对等方。换句话说，系统使用主体系统中的所有用户的训练数据来计算哪个其他用户访问了与所考虑的用户至少共同的阈值量的资源，例如至少10％、50％或80％的相同资源。

[0078] 系统还可以为拥有该主体系统的组织使用组织数据。例如，系统可以将属于同一个团队或部门的用户指定为对等方。系统还可以指定组织内具有相同或相似角色的用户作为对等方。

[0079] 在识别用户的对等方之后，系统可以使用训练数据为用户和所有用户的对等方生成初始路径图。

[0080] 如图4所示，系统使用测试数据生成测试路径图(420)。测试路径图是从测试数据生成的路径图。如上所述，测试数据可以表示用户在最近时间段期间访问的资源。因此，测试路径图表示用户在由测试数据表示的时间段期间如何导航到主体系统中的资源。

[0081] 图5B图示了示例测试路径图。测试路径图包括两个新节点，节点540和542以及由虚线表示的对应的新链接。新节点540和542表示在测试时间段期间但不在训练时间段期间由用户访问的资源。

[0082] 如图4所示，系统比较初始路径图和测试路径图(430)。系统可以使用多种方法来比较初始路径图和测试路径图。通常，系统计算初始路径图与测试路径图之间的重叠度量。显著重叠初始路径图的测试路径图表示正常的用户行为。另一方面，具有许多不与初始路径图重叠的节点和边缘的测试路径图表示异常的用户行为。

[0083] 例如，系统可以根据下式计算初始路径图G1和测试路径图G2之间的Jaccard距离D：

[0084] $$D=1-\frac{|G1\cap G2|}{|G1\cup G2|},$$

[0085] 其中|G1∩G2|表示G1中的节点集合与G2中的节点集合的交集的基数，并且|G1∪G2|表示G1中节点集合与G2中节点集合的并集的基数。

[0086] 在一些实现中，系统根据对资源的权重来计算加权Jaccard距离。系统可以根据各种因素为资源指派权重。例如，系统可以为包含敏感信息的资源(例如，敏感的公司或员工

数据)指派更高的权重。因此,异常行为的检测对于访问具有较高权重的文件夹的用户变得更加敏感。

[0087] 系统还可以根据资源的层级关系指派权重。例如,如果资源表示文件夹和子文件夹,则系统可以为文件夹指派比该文件夹的子文件夹高的权重。这使得异常行为的检测对用户仅访问该用户已经访问过的文件夹的新子文件夹的情况较不敏感。在一些实现中,系统将第一权重指派给层级中的阈值数量级别之上的所有资源,并将更小的第二权重指派给所有其他资源。例如,系统可以为文件系统的根目录以及根目录下最多三级的所有目录指派第一权重。对于所有其他子文件夹,系统可以指派第二权重。

[0088] 系统也可以基于系统资源的年龄指派权重。在某些情况下,异常行为比旧资源更可能涉及新创建的资源。因此,系统可以增加指派给新资源的权重并随着资源变老减少该资源的权重。

[0089] 系统还可以基于主体系统中资源的受欢迎程度来指派权重。例如,系统可以降低指派给主体系统中许多用户访问的受欢迎资源的权重。系统可以类似地降低受欢迎资源的所有子资源(例如,流行文件夹的子文件夹)的权重。

[0090] 在为系统中的资源指派权重后,系统可以根据以下公式计算加权Jaccard距离WD:

[0091] $$WD = 1 - \frac{\sum_{i \in G1 \cap G2} w_i}{\sum_{i \in G2} w_i},$$

[0092] 其中分子项表示在G1和G2的交集中出现的所有节点的权重之和,而分母项表示在G2中出现的所有节点的权重之和。

[0093] 系统基于比较将测试时段中的用户行为分类为异常或非异常(440)。如果初始路径图和测试路径图之间计算的距离大,则用户行为更可能是异常的。如果计算的距离很小,用户行为不太可能是异常的。因此,如果计算的距离满足阈值,则系统可以将用户行为分类为异常。

[0094] 异常事件通常需要由主体系统的取证团队(forensic team)跟进。因此,系统可以基于团队的预期可用性调整每个测试时间段的阈值,以调查异常情况。

[0095] 例如,图5A所示的初始路径图和图5B所示的测试路径图之间的Jaccard距离是相对较低的0.333。因此,系统可能不会考虑用户行为是异常的。

[0096] 图5C图示了另一个示例测试路径图。测试路径图包括六个新节点540,542,544,550,552和554以及由虚线表示的对应新链接。

[0097] 图5A中所示的初始路径图和图5C所示的测试路径图之间的Jaccard距离是比较高的0.6。因此,系统可能会认为用户行为是异常的。

[0098] 图6是用于确定主体系统中最受欢迎的资源的示例过程的流程图。当做出用户行为异常与否的确定时,系统可以考虑哪些资源受欢迎。如果用户的行为是正常的,但是访问本来是流行的资源,则系统避免将该用户行为标记为异常。该过程将被描述为由一个或多个计算机的适当编程的系统执行。

[0099] 系统生成混合用户/资源图(610)。混合图有两种类型的节点,表示用户的用户节点和表示主体系统中资源的资源节点。混合图还具有两种类型的对应链接,表示主体系统

中的资源结构的资源-资源链接以及表示访问主体系统中的资源的用户的用户-资源链接。

[0100] 图7图示了示例混合图。混合图具有与图5A所示的示例图相同的资源结构，具有表示文件系统中的文件夹的四个资源节点710，720，722和730。

[0101] 混合图在资源节点之间具有资源-资源链接，其表示系统中资源的结构。在这个示例中，资源-资源链接表示目录包含。

[0102] 混合图还包括表示系统中不同用户的两个用户节点760和762。混合图具有用户-资源链接，其表示每个用户访问哪些资源。

[0103] 在此示例中，用户资源链接很可能表示主文件夹比其他文件夹更受欢迎，因为主文件夹比其他文件夹被更多用户访问。

[0104] 如图6所示，系统根据混合图计算系统中资源的分值(620)。通常，该分值表示基于图所表示的关系在系统中资源的流行程度。因此，被更多用户访问的资源将具有更高的分值，并且被更少用户访问的资源将具有更低的分值。

[0105] 在一些实现中，系统计算具有表示用户通过在节点处结束的资源-资源链接执行随机导航的似然率的第一分量和表示用户从由子节点的父节点表示的资源到达由子节点表示的资源的似然率的第二分量。

[0106] 系统可以根据以下方程，迭代计算每个节点S(i)的分值：

[0107] $$S(i)=\frac{(1-d)}{N}+d\cdot\sum_{j}\frac{S(j)}{out(j)},$$

[0108] 其中每个节点j表示另一个用户节点或另一个具有到节点i的链接的资源节点，N是混合图中节点的数目，并且其中d是阻尼因子。对于没有任何传出边缘的节点，系统可以在图中所有N个节点之间平均分配它们的分值。

[0109] 系统选择具有最高分值的资源节点(630)。系统可以根据计算得分对资源节点进行排名，并选择最高排名的资源节点作为系统中最受欢迎的节点。系统可以选择预定数量的最高排名的资源节点，或者备选地，系统可以选择具有满足阈值的分值的所有资源节点。

[0110] 系统将所选资源节点的路径添加到所有初始路径图(640)。在确定最受欢迎的节点之后，系统可以将所有最受欢迎的节点的路径添加到主体系统中所有用户的初始路径图中。如此，系统将每个用户视为该用户访问了每个最受欢迎的的文件夹。当使用基于对等方的方法时，系统将每个用户的对等方视为他们已经访问了每个最受欢迎的文件夹。

[0111] 通过将路径添加到最受欢迎的文件夹中，系统可以减少由于用户访问他们不频繁访问但是在系统中的用户中本来是受欢迎的文件夹而产生的误报数量。

[0112] 本说明书中所描述的主题和功能操作的实施例能够用数字电子电路、用有形地体现的计算机软件或固件、用包括本说明书中所公开的结构及其结构等同物的计算机硬件、或者用它们中的一个或多个的组合来实现。本说明书中所描述的主题的实施例能够作为一个或多个计算机程序(即，在有形非暂时性程序载体上编码以用于由数据处理设备执行或者控制数据处理设备的操作的一个或多个计算机程序指令模块)被实现。替选地或附加地，可以将程序指令编码在为对信息进行编码以便发送到适合的接收器设备以由数据处理设备执行而生成的人工生成的传播信号(例如，机器生成的电、光学或电磁信号)上。计算机存储介质可以是机器可读存储设备、机器可读存储基底、随机或串行存取存储器设备，或它们中的一个或多个的组合。然而计算机存储介质不是传播信号。

[0113] 术语“数据处理装置”包含用于处理数据的所有类型的装置、设备和机器，作为示例包括可编程处理器、计算机、或多个处理器或计算机。所述装置可以包括专用逻辑电路，例如，FPGA(现场可编程门阵列)或ASIC(专用集成电路)。所述装置除了包括硬件之外，还可以包括为所述的计算机程序创建执行环境的代码，例如，构成处理器固件、协议栈、数据库管理系统、操作系统或它们中的一个或多个的组合的代码。

[0114] 计算机程序(其还也可以被称为或者描述为程序、软件、软件应用、模块、软件模块、脚本或代码)可以用任何形式的编程语言(包括编译或解释语言、或者描述性或过程语言)来编写，并且它可以被部署为任何形式(包括作为独立程序或者作为适合于在计算环境中使用的模块、组件、子例程或其它单元)。计算机程序可以但不必对应于文件系统中的文件。可以在保存其它程序或数据(例如，存储在标记语言文档中的一个或多个脚本)的文件的一部分中、在专用于所述程序的单个文件中、或在多个协同文件(例如，存储一个或多个模块、子例程、或代码的部分的文件)中存储程序。可以将计算机程序部署成在一个计算机上或者在位于一个站点处多个计算机上、或在跨越多个站点分布并且通过通信网络互连的多个计算机上执行。

[0115] 如本说明书中所使用的，“引擎”或“软件引擎”是指提供与输入不同的输出的软件实现的输入/输出系统。引擎可以是功能性的编码块，诸如库、平台、软件开发套件(“SDK”)或对象。每个引擎可以被实现在包括一个或多个处理器和计算机可读介质的任何适当类型的计算设备上，所述计算设备例如服务器、移动电话、平板计算机、笔记本计算机、音乐播放器、电子书阅读器、膝上型或台式计算机、PDA、智能电话或其它固定或便携式设备。附加地，所述引擎中的两个或更多个可以被实现在同一计算设备上或者在不同的计算设备上。

[0116] 本说明书中所描述的过程和逻辑流程可以由执行一个或多个计算机程序的一个或多个可编程计算机来执行以通过对输入数据进行操作并且生成输出来执行功能。还可以由专用逻辑电路来执行过程和逻辑流，并且装置还可以被实现为专用逻辑电路(例如FPGA(现场可编程门阵列)或ASIC(专用集成电路))。

[0117] 作为示例，适合于执行计算机程序的计算机可以基于通用微处理器或专用微处理器或两者，或任何其它类型的中央处理单元。一般地，中央处理单元将从只读存储器或随机存取存储器或两者接收指令和数据。计算机的必要元件是用于依照指令执行或者实行指令的中央处理单元以及用于存储指令和数据的一个或多个存储器设备。一般地，计算机还将包括用于存储数据的一个或多个大容量存储设备(例如，磁盘、磁光盘或光盘)或者与所述于一个或多个大容量存储设备操作地耦合，以从其接收数据或者向其转移数据或两者。然而，计算机不必具有所述设备。而且，可以将计算机嵌入在另一设备(例如，移动电话、个人数字助理(PDA)、移动音频或视频播放器、游戏控制台、全球定位系统(GPS)接收器或便携式存储设备(例如，通用串行总线(USB)闪存驱动器))等中。

[0118] 适合于存储计算机程序指令和数据的计算机可读介质包括所有形式的非易失性存储器、介质以及存储器设备，作为示例包括半导体存储器设备，例如，EPROM、EEPROM和闪存设备；磁盘，例如，内部硬盘或可移除磁盘；磁光盘；以及CD ROM和DVD-ROM盘。处理器和存储器可以由专用逻辑电路补充或者并入专用逻辑电路。

[0119] 为了提供与用户的交互，可以将本说明书中所描述的主题的实施例实现在具有用于向用户显示信息的显示设备(例如，CRT(阴极射线管)或LCD(液晶显示)监视器)，或OLED

显示器以及用于向计算机提供输入的输入设备(例如键盘、鼠标)或呈现敏感显示器或其他表面。其它类型的设备也可以被用来提供与用户交互;例如,提供给用户的反馈可以是任何形式的感觉反馈,例如,视觉反馈、听觉反馈或触觉反馈;并且可以以包括声输入、语音输入或触觉输入的任何形式接收来自用户的输入。此外,计算机可以通过向由用户使用的设备发送资源并且从由用户使用的设备接收资源来与用户交互;例如,通过响应于从web浏览器接收到的请求而向用户的客户端设备上的web浏览器发送web页面。

[0120] 本说明书中所描述的主题的实施例可以被实现在包括后端组件的计算系统中(例如,作为数据服务器),或者被实现在包括中间件组件的计算系统中(例如,应用服务器),或者被实现在包括前端组件的计算系统中(例如,具有用户可以用来与本说明书中所描述的主题的实施方式交互的图形用户界面或Web浏览器的客户端计算机),或可以被实现在包括一个或多个这种后端组件、中间件组件或前端组件的任何组合的计算系统中。系统的组件可以通过任何形式或介质的数字数据通信(例如,通信网络)来互连。通信网络的示例包括局域网("LAN")和广域网("WAN")(例如,互联网)。

[0121] 计算系统可以包括客户端和服务器。客户端和服务器一般地彼此远离并且典型地通过通信网络交互。客户端和服务器的关系借助于在相应的计算机上运行并且彼此具有客户端-服务器关系的计算机程序而产生。

[0122] 虽然本说明书包含许多特定实施方式细节,但是这些不应该被解释为对任何发明的范围或可能要求保护的范围构成限制,而是相反被解释为可能对特定发明的特定实施例的而特有的特征的描述。在本说明书中在分离实施例的上下文中所描述的特定特征还可以在单个实施例中组合地实现。相反地,在单个实施例上下文下所描述的各种特征还可以分离地在多个实施例中或在任何适合的子组合中实现。另外,尽管特征可以被以上描述为在特定组合中行动并且因此甚至最初要求保护如此,但是来自要求保护的组合的一个或多个特征可以在一些情况下被从组合中删除,并且所要求保护的组合可以被导向子组合或子组合的变体。

[0123] 类似地,虽然按特定次序在附图中描绘操作,但是这不应该被理解为为了实现所希望的结果,要求所述操作被以所示出的特定次序或以顺序次序执行,或者要求执行所有图示的操作。在特定情况下,多任务处理和并行处理可能是有利的。此外,在上面所描述的实施例中各种系统模块和组件的分离不应该被理解为在所有实施例中要求这种分离,而应该理解的是,所描述的程序组件和系统通常可以被一起集成在单个软件产品中或者封装到多个软件产品中。

[0124] 已经描述了本主题的特定实施例。其它实施例在以下权利要求的范围内。例如,权利要求中所记载的动作可以按照不同次序被执行并且仍然实现所希望的结果。作为一个示例,附图中所描绘的过程未必要求所示出的特定次序或顺序次序以实现所希望的结果。在某些实施方式中,多任务处理和并行处理可能是有利的。

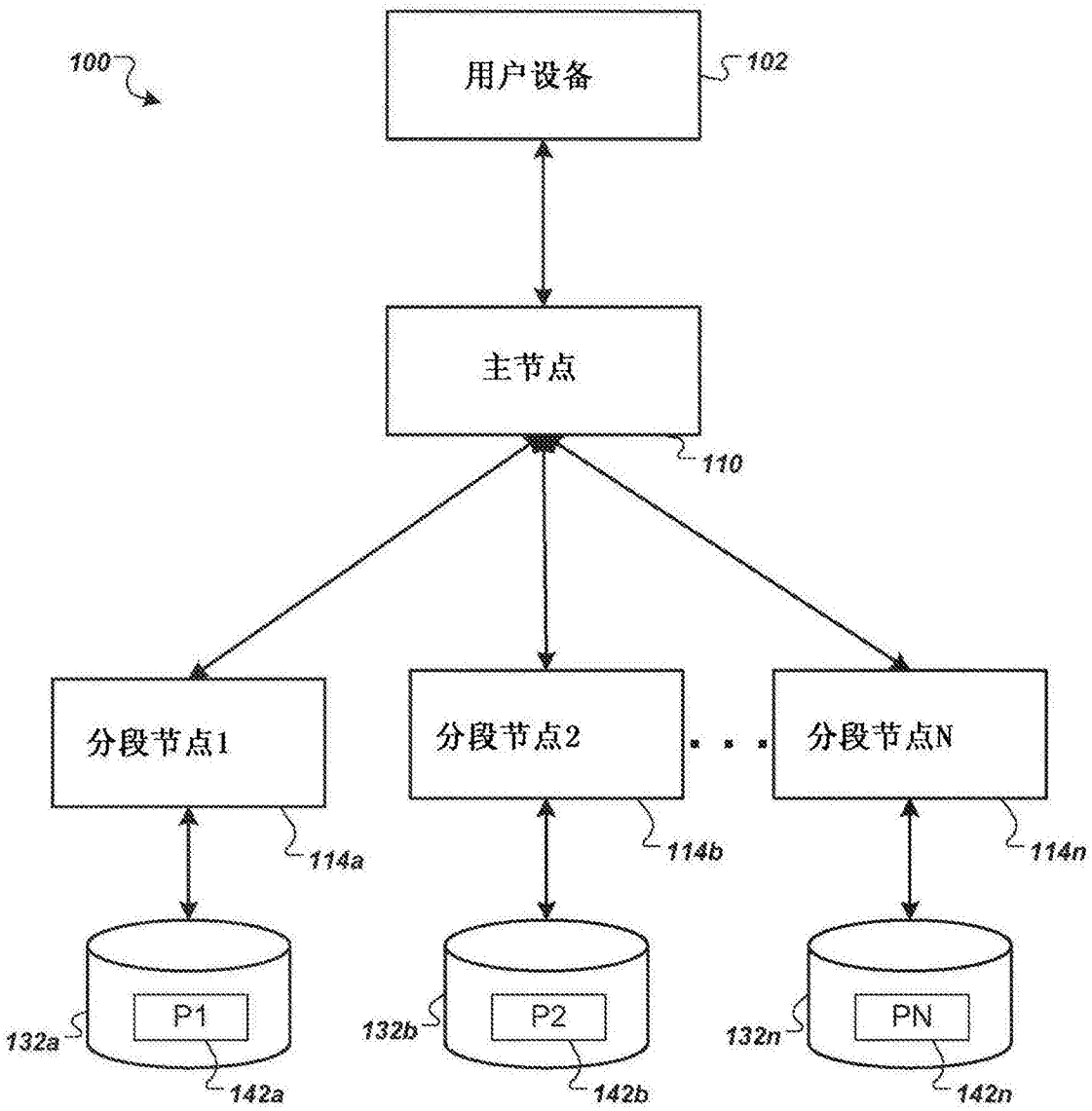

100
用户设备
102
主节点
110
分段节点1
分段节点2
分段节点N
114a
114b
114n
P1
P2
PN
132a
132b
132n
142a
142b
142n

图1A

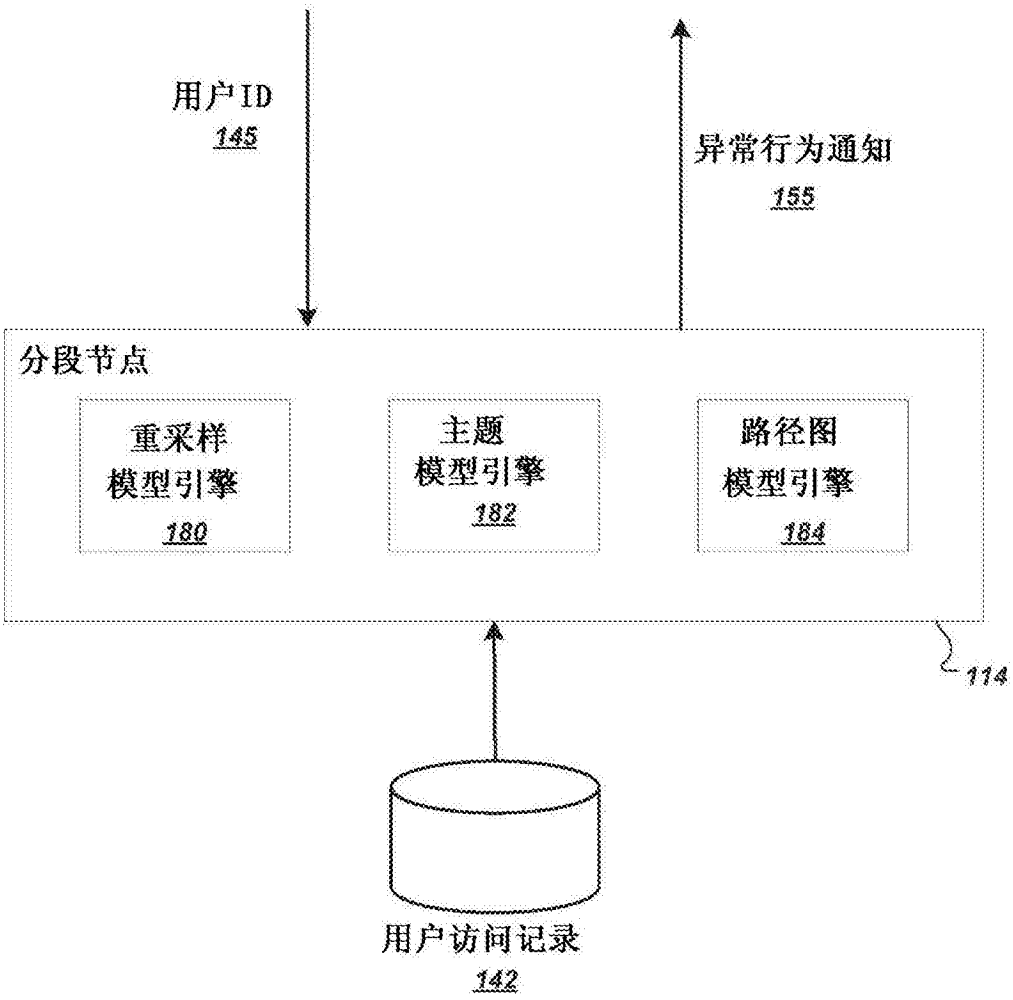

用户ID
145
异常行为通知
155
分段节点
重采样
模型引擎
180
主题
模型引擎
182
路径图
模型引擎
184
114
用户访问记录
142

图1B

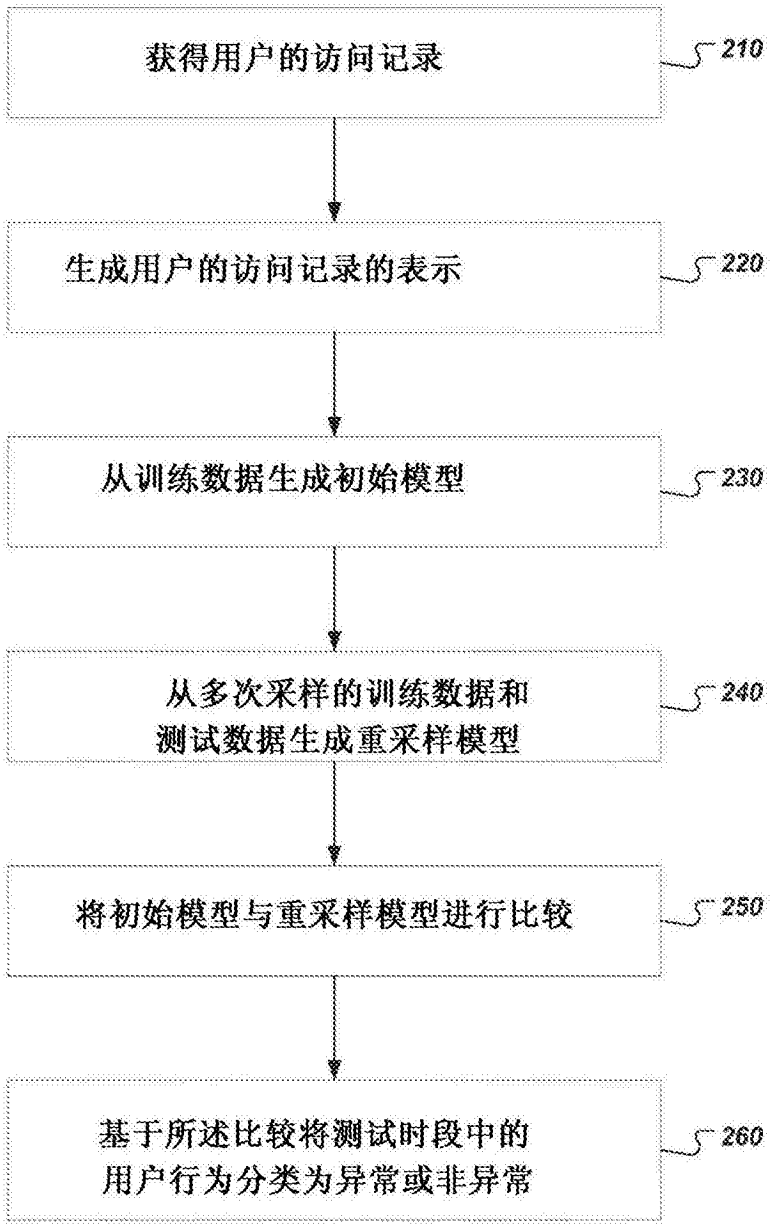

获得用户的访问记录
210
生成用户的访问记录的表示
220
从训练数据生成初始模型
230
从多次采样的训练数据和
测试数据生成重采样模型
240
将初始模型与重采样模型进行比较
250
基于所述比较将测试时段中的
用户行为分类为异常或非异常
260

图2

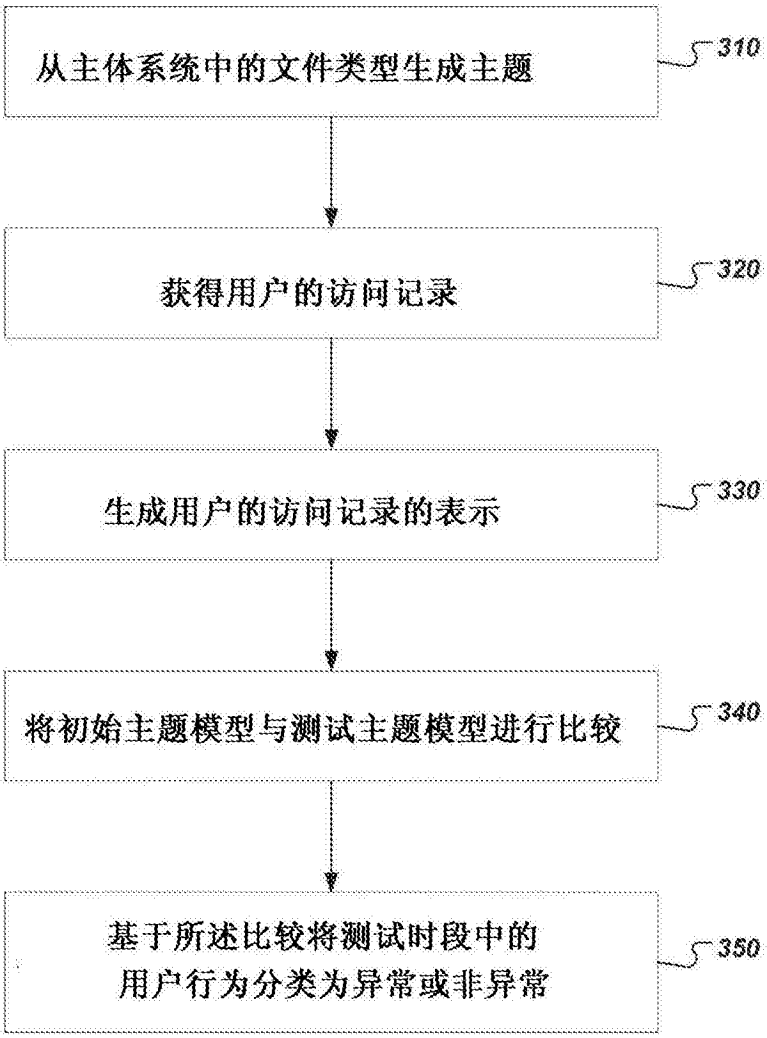

从主体系统中的文件类型生成主题
310
获得用户的访问记录
320
生成用户的访问记录的表示
330
将初始主题模型与测试主题模型进行比较
340
基于所述比较将测试时段中的
用户行为分类为异常或非异常
350

图3

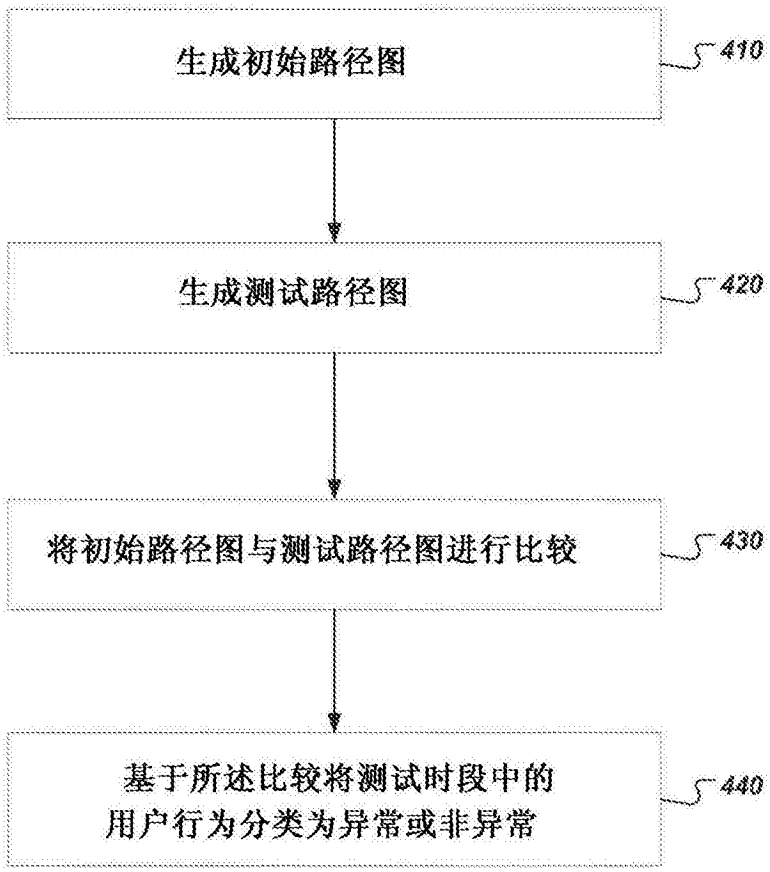

生成初始路径图
410
生成测试路径图
420
将初始路径图与测试路径图进行比较
430
基于所述比较将测试时段中的
用户行为分类为异常或非异常
440

图4

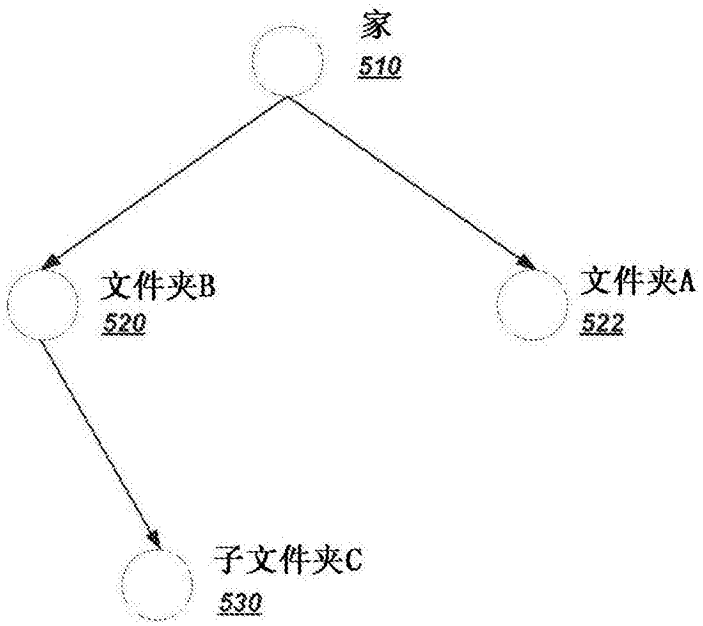

家
510
文件夹B
520
文件夹A
522
子文件夹C
530

图5A

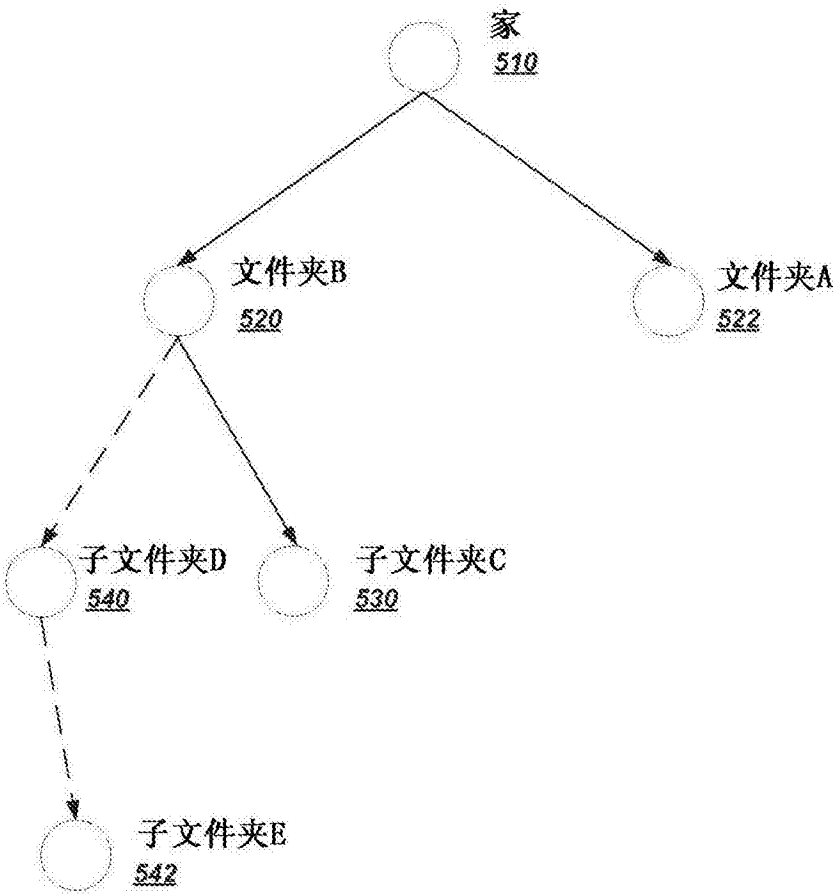


图5B

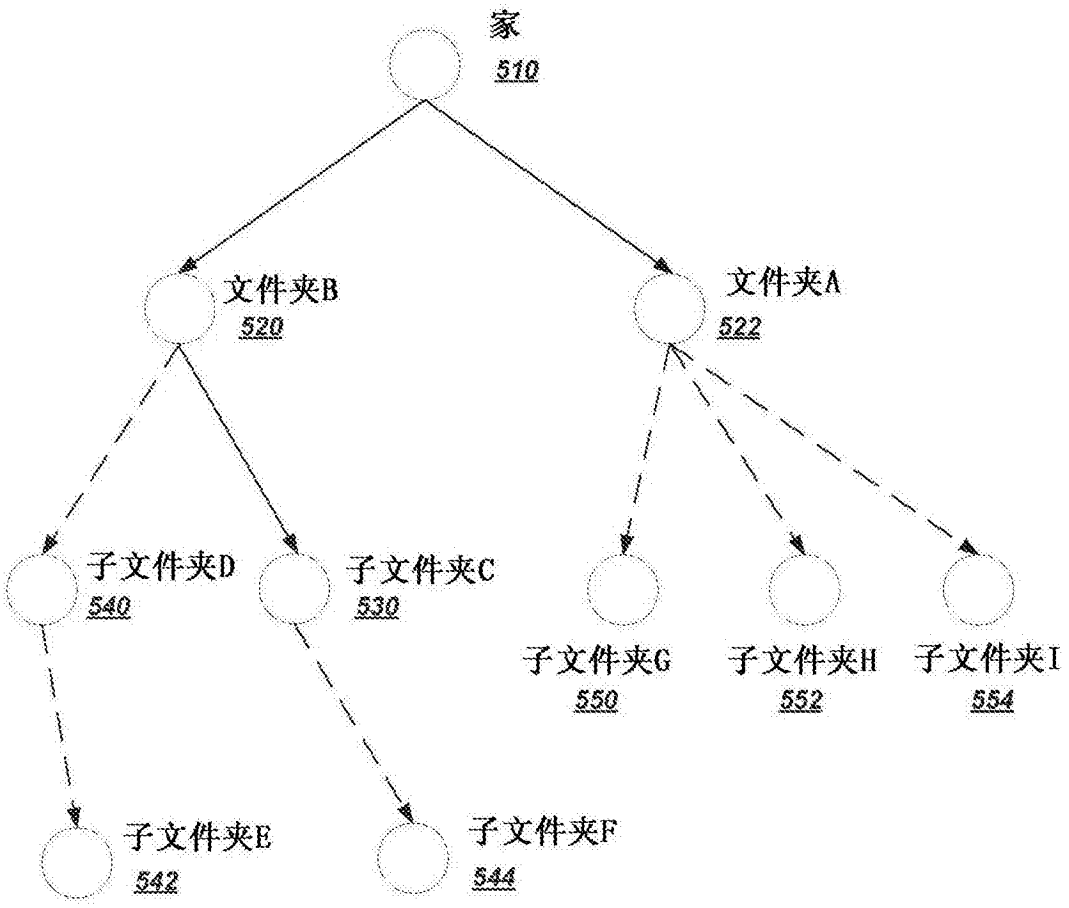

家
510
文件夹B
520
文件夹A
522
子文件夹D
540
子文件夹C
530
子文件夹G
550
子文件夹H
552
子文件夹I
554
子文件夹E
542
子文件夹F
544

图5C

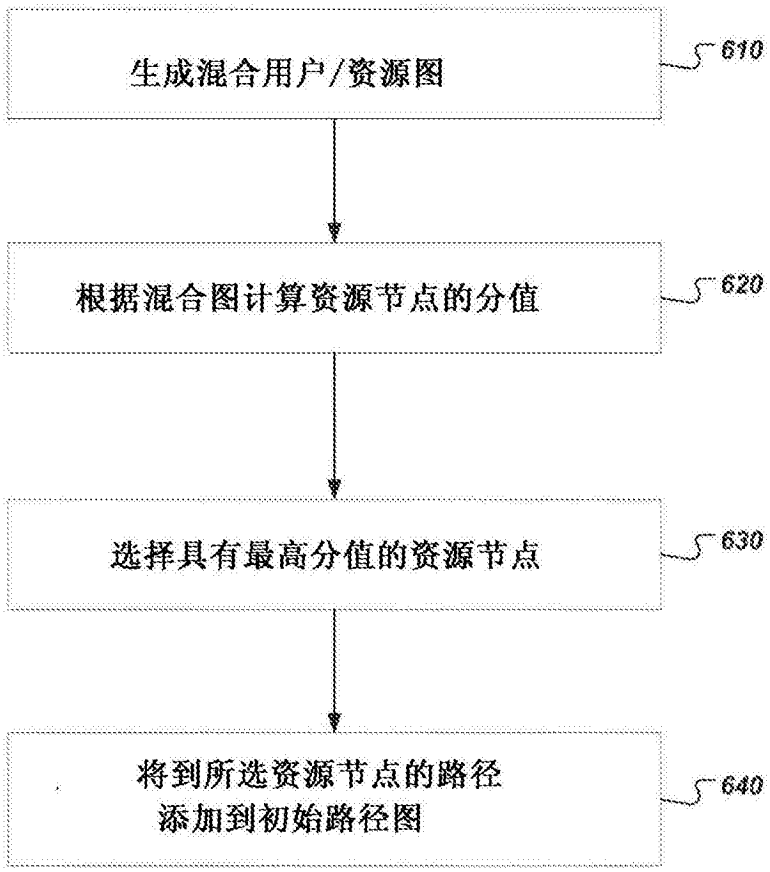

生成混合用户/资源图
610
根据混合图计算资源节点的分值
620
选择具有最高分值的资源节点
630
将到所选资源节点的路径
添加到初始路径图
640

图6

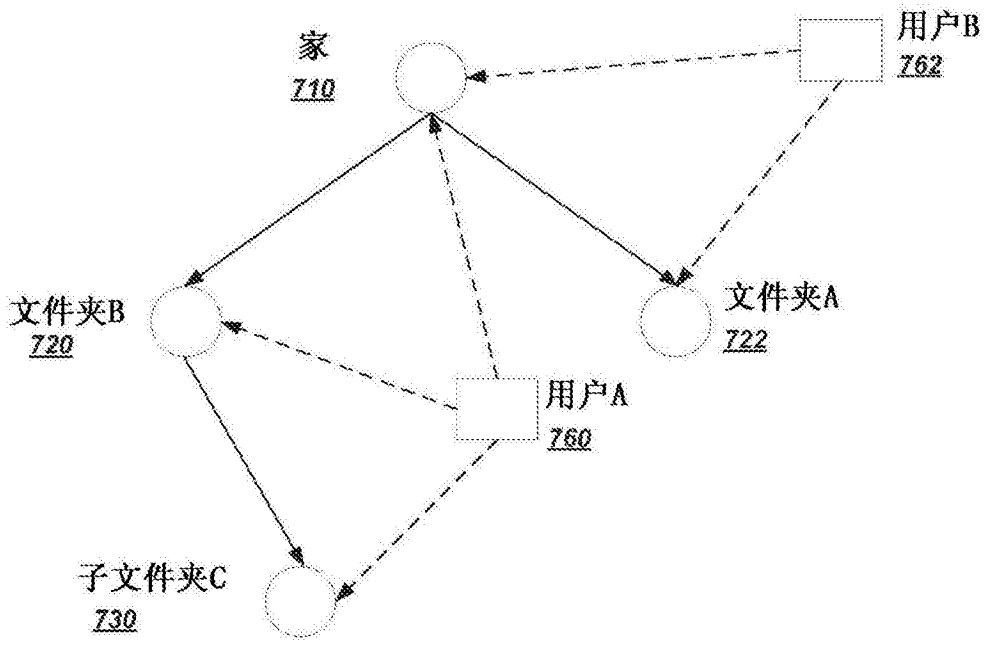

家
710
用户B
762
文件夹B
720
文件夹A
722
用户A
760
子文件夹C
730

图7