(19)中华人民共和国国家知识产权局

(12)发明专利

(10)授权公告号 CN 106933880 B
(45)授权公告日 2020.08.11

(21)申请号 201511028180.5

(22)申请日 2015.12.31

(65)同一申请的已公布的文献号
申请公布号 CN 106933880 A

(43)申请公布日 2017.07.07

(73)专利权人 阿里巴巴集团控股有限公司
地址 英属开曼群岛大开曼资本大厦一座四层847号邮箱

(72)发明人 文镇

(74)专利代理机构 杭州君度专利代理事务所（特殊普通合伙）33240
代理人 诸佩艳

(51)Int.Cl.
G06F 16/953(2019.01)
G06F 16/955(2019.01)

(56)对比文件
CN 103281403 A,2013.09.04,
CN 103237018 A,2013.08.07,
WO 2015030856 A1,2015.03.05,
CN 103870000 A,2014.06.18,
CN 103581883 A,2014.02.12,
CN 104133837 A,2014.11.05,
CN 103581190 A,2014.02.12,
CN 103593465 A,2014.02.19,

审查员 沈晓娟

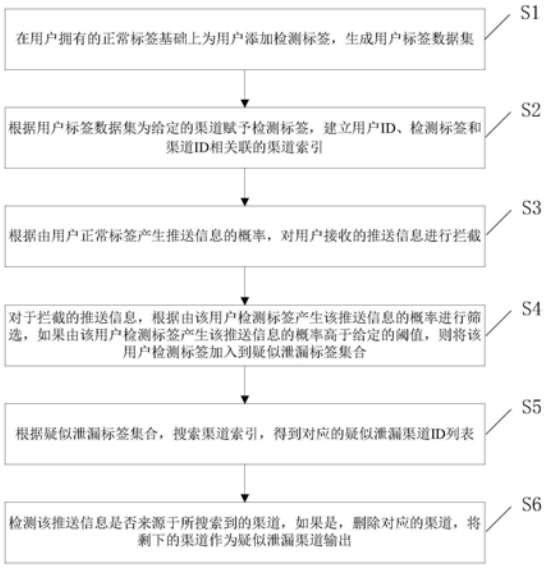
权利要求书2页 说明书7页 附图2页

(54)发明名称

一种标签数据泄漏渠道检测方法及装置

(57)摘要

本发明公开了一种标签数据泄漏渠道检测方法及装置，该方法利用标签在同一用户出现的不同概率，对不同数据使用渠道产生不同的检测标签，然后对检测标签的使用进行间接检测，最后通过海量数据索引和搜索技术有效地检测可能的数据泄露渠道。本发明的装置包括检测标签添加模块、渠道关联模块、拦截模块、拦截信息分析模块、渠道检索模块和输出模块。本发明的检测方法及装置检测效率高，能够处理海量、动态的用户标签数据。



CN 106933880 B

1.一种标签数据泄漏渠道检测方法,用于检测用户标签数据的泄漏渠道,其特征在于,所述检测方法包括:

在用户拥有的正常标签基础上为用户添加检测标签,生成用户标签数据集;

根据用户标签数据集为给定的渠道赋予检测标签,建立用户ID、检测标签和渠道ID相关联的渠道索引;

根据由用户正常标签产生推送信息的概率,对用户接收的推送信息进行拦截;

对于拦截的推送信息,根据由该用户检测标签产生该推送信息的概率进行筛选,如果由该用户检测标签产生该推送信息的概率高于给定的阈值,则将该用户检测标签加入到疑似泄漏标签集合;

根据疑似泄漏标签集合,搜索渠道索引,得到对应的疑似泄漏渠道ID列表;

检测该推送信息是否来源于所搜索到的渠道,如果是,删除对应的渠道,将剩下的渠道作为疑似泄漏渠道输出;

其中,所述在用户拥有的正常标签基础上为用户添加检测标签,包括:

新添加的检测标签与用户现有标签同时出现的概率低于设定的第一阈值;

所述根据由用户正常标签产生推送信息的概率,对用户接收的推送信息进行拦截,包括:

如果推送信息由正常标签产生的概率低于设定的第二阈值,则进行拦截,否则向用户展示该推送信息。

2.根据权利要求1所述的标签数据泄漏渠道检测方法,其特征在于,所述根据用户标签数据集为给定的渠道赋予检测标签,建立用户ID、检测标签和渠道ID相关联的渠道索引,包括:

对于给定的渠道,根据其历史行为计算其可信度;

以该渠道的渠道ID作为变量,从设定的HASH函数集中选取一个HASH函数;

基于渠道可信度抽样用户群;

对抽样得到的用户群中每一个用户,以用户ID作为变量,根据抽取得到的HASH函数从该用户的检测标签中选出该渠道对应的检测标签;

建立[用户ID、检测标签]到渠道ID的渠道索引。

3.根据权利要求1所述的标签数据泄漏渠道检测方法,其特征在于,所述检测方法还包括根据用户正常标签的变化更新用户检测标签的步骤,具体包括:

根据新的正常标签与现有检测标签同时出现概率,删除与用户新的正常标签同时出现概率高的检测标签;

重新为用户添加新的检测标签,新添加的检测标签与用户现有标签同时出现的概率低于第一阈值。

4.根据权利要求3所述的标签数据泄漏渠道检测方法,其特征在于,所述检测方法还包括:

从渠道索引中除去被删除检测标签相关项。

5.一种标签数据泄漏渠道检测装置,用于检测用户标签数据的泄漏渠道,其特征在于,所述检测装置包括:

检测标签添加模块,用于在用户拥有的正常标签基础上为用户添加检测标签,生成用

户标签数据集；

渠道关联模块，用于根据用户标签数据集为给定的渠道赋予检测标签，建立用户ID、检测标签和渠道ID相关联的渠道索引；

拦截模块，用于根据由用户正常标签产生推送信息的概率，对用户接收的推送信息进行拦截；

拦截信息分析模块，用于对于拦截的推送信息，根据由该用户检测标签产生该推送信息的概率进行筛选，如果由该用户检测标签产生该推送信息的概率高于给定的阈值，则将该用户检测标签加入到疑似泄漏标签集合；

渠道检索模块，用于根据疑似泄漏标签集合，搜索渠道索引，得到对应的疑似泄漏渠道ID列表；

输出模块，用于检测该推送信息是否来源于所搜索到的渠道，如果是，删除对应的渠道，将剩下的渠道作为疑似泄漏渠道输出；

其中，所述检测标签添加模块在用户拥有的正常标签基础上为用户添加检测标签时，新添加的检测标签与用户现有标签同时出现的概率低于设定的第一阈值；

所述拦截模块在根据由用户正常标签产生推送信息的概率，对用户接收的推送信息进行拦截时，执行如下操作：

如果推送信息由正常标签产生的概率低于设定的第二阈值，则进行拦截，否则向用户展示该推送信息。

6.根据权利要求5所述的标签数据泄漏渠道检测装置，其特征在于，所述渠道关联模块在根据用户标签数据集为给定的渠道赋予检测标签时，执行如下操作：

对于给定的渠道，根据其历史行为计算其可信度；

以该渠道的渠道ID作为变量，从设定的HASH函数集中选取一个HASH函数；

基于渠道可信度抽样用户群；

对抽样得到的用户群中每一个用户，以用户ID作为变量，根据抽取得到的HASH函数从该用户的检测标签中选出该渠道对应的检测标签；

建立[用户ID、检测标签]到渠道ID的渠道索引。

7.根据权利要求5所述的标签数据泄漏渠道检测装置，其特征在于，所述检测标签添加模块还用于根据用户正常标签的变化更新用户检测标签的步骤，具体执行如下步骤：

根据新的正常标签与现有检测标签同时出现概率，删除与用户新的正常标签同时出现概率高的检测标签；

重新为用户添加新的检测标签，新添加的检测标签与用户现有标签同时出现的概率低于第一阈值。

8.根据权利要求7所述的标签数据泄漏渠道检测装置，其特征在于，所述渠道关联模块还用于从渠道索引中除去被删除检测标签相关项。

一种标签数据泄漏渠道检测方法及装置

技术领域

[0001] 本发明属于数据安全技术领域，尤其涉及一种标签数据泄漏渠道检测方法及装置。

背景技术

[0002] 标签是一种互联网内容组织形式，是与对象实体的属性相关性很强的关键字。标签有助于轻松的描述和分类内容，以便于检索和分享。互联网发展中积累了大量以标签来表示的用户偏好数据，这些数据构成了互联网广告、推荐等产品的基础。另一方面，这些数据因为其价值，也和其他用户个人数据(PII)一起成为数据泄露的目标，被违规获取、转卖。现有的数据安全技术利用加密、系统加固、权限控制和审计监控，来防止数据泄露出数据所有者的可控环境。但是在数据合作的业务场景中，数据通常会离开数据所有者的可控环境，进入不可控的合作者的环境中去。在此场景中，传统的数据库水印技术和数据轨迹追踪技术不能解决海量的、动态的用户标签数据的挑战。

[0003] 传统的数据库水印技术和数据轨迹追踪技术不能对用户标签这样缺乏数值型字段的数据有效地产生水印。其次标签数据通常被分散使用，从而使水印检测很困难。另外标签数据具有海量、动态特征，对水印的更新和检测也有很大挑战。标签数据的取值一般很常见，在互联网中进行追踪非常困难。

发明内容

[0004] 本发明的目的是提供一种标签数据泄漏渠道检测方法及装置，以解决现有技术方案标签数据难以跟踪检测的技术问题，能够有效地检测可能的数据泄露渠道。

[0005] 为了实现上述目的，本发明技术方案如下：

[0006] 一种标签数据泄漏渠道检测方法，用于检测用户标签数据的泄漏渠道，所述检测方法包括：

[0007] 在用户拥有的正常标签基础上为用户添加检测标签，生成用户标签数据集；

[0008] 根据用户标签数据集为给定的渠道赋予检测标签，建立用户ID、检测标签和渠道ID相关联的渠道索引；

[0009] 根据由用户正常标签产生推送信息的概率，对用户接收的推送信息进行拦截；

[0010] 对于拦截的推送信息，根据由该用户检测标签产生该推送信息的概率进行筛选，如果由该用户检测标签产生该推送信息的概率高于给定的阈值，则将该用户检测标签加入到疑似泄漏标签集合；

[0011] 根据疑似泄漏标签集合，搜索渠道索引，得到对应的疑似泄漏渠道ID列表；

[0012] 检测该推送信息是否来源于所搜索到的渠道，如果是，删除对应的渠道，将剩下的渠道作为疑似泄漏渠道输出。

[0013] 进一步地，所述在用户拥有的正常标签基础上为用户添加检测标签，包括：

[0014] 新添加的检测标签与用户现有标签同时出现的概率低于设定的第一阈值。

[0015] 进一步地，所述根据用户标签数据集为给定的渠道赋予检测标签，建立用户ID、检测标签和渠道ID相关联的渠道索引，包括：

[0016] 对于给定的渠道，根据其历史行为计算其可信度；

[0017] 以该渠道的渠道ID作为变量，从设定的HASH函数集中选取一个HASH函数；

[0018] 基于渠道可信度抽样用户群；

[0019] 对抽样得到的用户群中每一个用户，以用户ID作为变量，根据抽取得到的HASH函数从该用户的检测标签中选出该渠道对应的检测标签；

[0020] 建立[用户ID、检测标签]到渠道ID的渠道索引。

[0021] 进一步地，所述根据由用户正常标签产生推送信息的概率，对用户接收的推送信息进行拦截，包括：

[0022] 如果推送信息由正常标签产生的概率低于设定的第二阈值，则进行拦截，否则向用户展示该推送信息。

[0023] 进一步地，所述检测方法还包括根据用户正常标签的变化更新用户检测标签的步骤，具体包括：

[0024] 根据新的正常标签与现有检测标签同时出现概率，删除与用户新的正常标签同时出现概率高的检测标签；

[0025] 重新为用户添加新的检测标签，新添加的检测标签与用户现有标签同时出现的概率低于第一阈值。

[0026] 进一步地，所述检测方法还包括：

[0027] 从渠道索引中除去被删除检测标签相关项。

[0028] 本发明还提出了一种标签数据泄漏渠道检测装置，用于检测用户标签数据的泄漏渠道，所述检测装置包括：

[0029] 检测标签添加模块，用于在用户拥有的正常标签基础上为用户添加检测标签，生成用户标签数据集；

[0030] 渠道关联模块，用于根据用户标签数据集为给定的渠道赋予检测标签，建立用户ID、检测标签和渠道ID相关联的渠道索引；

[0031] 拦截模块，用于根据由用户正常标签产生推送信息的概率，对用户接收的推送信息进行拦截；

[0032] 拦截信息分析模块，用于对于拦截的推送信息，根据由该用户检测标签产生该推送信息的概率进行筛选，如果由该用户检测标签产生该推送信息的概率高于给定的阈值，则将该用户检测标签加入到疑似泄漏标签集合；

[0033] 渠道检索模块，用于根据疑似泄漏标签集合，搜索渠道索引，得到对应的疑似泄漏渠道ID列表；

[0034] 输出模块，用于检测该推送信息是否来源于所搜索到的渠道，如果是，删除对应的渠道，将剩下的渠道作为疑似泄漏渠道输出。

[0035] 进一步地，所述检测标签添加模块在用户拥有的正常标签基础上为用户添加检测标签时，新添加的检测标签与用户现有标签同时出现的概率低于设定的第一阈值。

[0036] 进一步地，所述渠道关联模块在根据用户标签数据集为给定的渠道赋予检测标签时，执行如下操作：

[0037] 对于给定的渠道,根据其历史行为计算其可信度;

[0038] 以该渠道的渠道ID作为变量,从设定的HASH函数集中选取一个HASH函数;

[0039] 基于渠道可信度抽样用户群;

[0040] 对抽样得到的用户群中每一个用户,以用户ID作为变量,根据抽取得到的HASH函数从该用户的检测标签中选出该渠道对应的检测标签;

[0041] 建立[用户ID、检测标签]到渠道ID的渠道索引。

[0042] 进一步地,所述拦截模块在根据由用户正常标签产生推送信息的概率,对用户接收的推送信息进行拦截时,执行如下操作:

[0043] 如果推送信息由正常标签产生的概率低于设定的第二阈值,则进行拦截,否则向用户展示该推送信息。

[0044] 进一步地,所述检测标签添加模块还用于根据用户正常标签的变化更新用户检测标签的步骤,具体执行如下步骤:

[0045] 根据新的正常标签与现有检测标签同时出现概率,删除与用户新的正常标签同时出现概率高的检测标签;

[0046] 重新为用户添加新的检测标签,新添加的检测标签与用户现有标签同时出现的概率低于第一阈值。

[0047] 进一步地,所述渠道关联模块还用于从渠道索引中除去被删除检测标签相关项。

[0048] 本发明提出了一种标签数据泄漏渠道检测方法及装置,利用标签在同一用户出现的不同概率,对不同数据使用渠道产生不同的检测标签。然后对检测标签的使用进行间接检测,最后通过海量数据索引和搜索技术有效地检测可能的数据泄露渠道。检测方法效率高,能够处理海量、动态的用户标签数据。

附图说明

[0049] 图1为本发明标签数据泄漏渠道检测方法流程图;

[0050] 图2为本发明标签数据泄漏渠道检测装置结构示意图。

具体实施方式

[0051] 下面结合附图和实施例对本发明技术方案做进一步详细说明,以下实施例不构成对本发明的限定。

[0052] 用户浏览互联网时,浏览的网页会为用户生成表示其偏好的标签,互联网发展中积累了大量以标签来表示的用户偏好数据。本发明在用户拥有正常标签的基础上,为每一个用户加一定量的检测标签,当发现有检测标签导致的推送信息时,可以根据该推送信息查找用户标签数据泄漏的渠道。本实施例推送信息可以包括广告,推送的网页等,以下以广告为例进行说明。

[0053] 本实施例一种标签数据泄漏渠道检测方法,如图1所示,包括:

[0054] 步骤S1、在用户拥有的正常标签基础上为用户添加检测标签,生成用户标签数据集。

[0055] 本实施例将由用户上网而产生的标识用户偏好的标签称为正常标签,而将通过本步骤为用户生成的用于后续检测的标签称为检测标签,显然检测标签不代表用户的偏好,

仅用作后续的检测。用户标签数据集包括正常标签和检测标签。

[0056] 为了后续分析方便，每个用户需有足够多的检测标签，以便对应不同的渠道。为此，当用户没有足够多的检测标签时，为用户生成检测标签，使用户的检测标签达到设定的数量。

[0057] 例如，用户U1有两个正常标签，分别为：看电视、垃圾快餐，而本实施例要求的检测标签为两个，则为其产生两个检测标签，例如为：蔬菜、登山鞋。

[0058] 具体生成用户检测标签的过程如下：

[0059] 判断用户标签数据集中是否有指定数量的检测标签，如果已经达到指定的数量则结束，否则进入下一步；

[0060] 生成一个与用户现有标签同时出现概率低于设定的第一阈值的标签，将该标签作为用户的检测标签加入到用户标签数据集。

[0061] 其中，在生成新的检测标签时，需要在常见的标签中找到一个与用户现有的正常标签、现有的检测标签同时出现概率较低的标签，即新生成的检测标签与用户标签集中现有标签均不相似，具有差异性，同时出现的概率低。

[0062] 步骤S2、根据用户标签数据集为给定的渠道赋予检测标签，建立用户ID、检测标签和渠道ID相关联的渠道索引。

[0063] 对于给定的渠道，可以根据其历史行为计算其可信度。本实施例渠道是指使用用户数据的渠道，例如一个网络平台将自己的用户数据提供给一个广告商，该广告商就是网络平台的客户，也是使用用户数据的一个渠道。渠道的可信度是指该渠道根据用户数据发送广告的可信度，如果该渠道不是基于用户数据来推送广告，而是将用户不感兴趣的广告推送给用户则不可信。并且可以利用该渠道的唯一ID作为变量key，从一个设定的Hash函数集里面选取Hash函数H1。接下来基于渠道可信度抽样用户群，可信度高的渠道抽样人群可以小一些。然后对抽样人群的每一个用户，以用户ID为key，用H1函数从该用户的检测标签集中选出该渠道对应的检测标签。

[0064] 例如对于给定的渠道1，抽样用户中包括用户U1，通过H1函数与用户U1的用户ID计算得到一个随机值，根据该随机值从用户U1的所有检测标签中选择一个检测标签赋予给渠道1。例如渠道1，通过H1函数计算得到的随机值为1，则根据用户1检测标签的排序，选择第一个检测标签赋予给渠道1。假设将用户U1的检测标签“蔬菜”赋予给渠道1。

[0065] 同样地，将用户U1的检测标签“登山鞋”赋予给渠道2。

[0066] 这样就可以建立[用户ID，检测标签]到渠道ID的渠道索引，即在渠道索引中建立一条记录，例如建立如表1所示的渠道索引：

[0067]

序号	[用户ID，检测标签]	渠道ID
1	[U1，蔬菜]	渠道1
2	[U1，登山鞋]	渠道2

[0068] 表1

[0069] 在用户标签数据集中加入检测标签，仅将与渠道对应的检测标签赋予给对应的渠道，例如将[U1，登山鞋]赋予给渠道2。如果渠道2根据用户标签数据集来推送广告，无论是根据正常标签还是检测标签[U1，登山鞋]发送的广告都认为是安全的。而非法的用户获得泄漏的用户标签数据后，也向用户发送登山鞋之类的广告，根据渠道索引发现该非法渠道

不是渠道索引中的渠道2时，则认为用户标签数据发生了泄漏。

[0070] 步骤S3、根据由用户正常标签产生推送信息的概率，对用户接收的推送信息进行拦截。

[0071] 一般情况下，由于用户上网的终端一般都在用户一侧，因此用户接收到的广告是反映在用户的终端上的，对于广告的检测首先可以在用户终端上的客户端上进行。例如现在很多个人电脑和智能手机上都安装了安全助手，可以直接采用现有的安全助手在用户终端上进行广告拦截。当然也可以开发特定的客户端，用于在用户终端上进行广告检测。

[0072] 在进行广告拦截时，如果广告由正常标签产生的概率低于设定的第二阈值，则进行拦截，否则向用户展示该广告。

[0073] 容易理解的是，如果采用用户终端现有的安全助手，在步骤S2中，对于用户标签数据集，首先要过滤掉其中没有安装安全助手的人群。即仅对安装了安全助手的人群进行抽样，对于没有安装安全助手的用户不予考虑。这样可以不需要额外开发客户端，直接采用用户的安全助手来进行用户终端一侧的广告过滤。

[0074] 具体地，对广告进行过滤，即根据由用户正常标签产生广告的概率，对用户接收的广告进行拦截，如果该广告由正常标签产生的概率低于设定的阈值，则进入下一步处理，否则向用户展示该广告。

[0075] 需要说明的是，用户正常标签需要同步到该用户的用户端安全助手中，以便安全助手根据正常标签产生该广告的概率来进行拦截。根据正常标签产生该广告的概率，一般由安全助手根据该广告来源与用户正常标签的匹配程度来计算，这里不再赘述。对于由正常标签产生的概率低于设定的阈值的广告，进行拦截并发送到专门的后台服务器端进行下一步的处理。

[0076] 步骤S4、对于拦截的推送信息，根据由该用户检测标签产生该推送信息的概率进行筛选，如果由该用户检测标签产生该推送信息的概率高于给定的阈值，则将该该用户检测标签加入到疑似泄漏标签集合。

[0077] 对于发送到后台服务器端的广告，进一步根据由该用户检测标签产生该广告的概率进行筛选。如果由某一用户检测标签产生该广告的概率高于给定的阈值，则将该用户检测标签加入到疑似泄漏标签集合。

[0078] 例如一个发送到用户U1的登山杖的广告，根据正常标签“看电视”、“垃圾快餐”产生的概率比较低，被发送到后台服务器端。然而对于用户U1的检测标签“登山鞋”，由“登山鞋”产生该广告的概率却比较高，因此[用户U1，登山鞋]被加入到疑似泄漏标签集合。

[0079] 步骤S5、根据疑似泄漏标签集合，搜索渠道索引，得到对应的疑似泄漏渠道ID列表。

[0080] 接下来，从疑似泄漏标签集合中取出疑似标签，并在渠道索引中进行搜索，得到有可能的渠道ID排序列表。

[0081] 例如前面这个例子中，从疑似泄漏标签集合中取出疑似泄漏标签[用户U1，登山鞋]，在渠道索引中因为渠道2的检测标签有“登山鞋”，将渠道2加入到疑似泄漏渠道ID列表。

[0082] 步骤S6、检测该推送信息是否来源于所搜索到的渠道，如果是，删除对应的渠道，将剩下的渠道作为疑似泄漏渠道输出。

[0083] 最后，需要检测该用户终端的广告来源是否是渠道2，如果是的话则表明是合规情况，从渠道列表中删除。

[0084] 最终渠道列表中包括了所有可能的标签数据泄露渠道。对这些渠道，可以采取更多的调查手段收集证据，例如在合作数据中加入可监控的诱饵(蜜罐)数据，结合线下调查等手段。

[0085] 进一步地，由于用户的正常标签经常得到更新，在更新了用户的正常标签后，需要更新该用户的检测标签。本实施例用户检测标签更新的过程如下：

[0086] 根据新的正常标签与现有检测标签同时出现概率，删除与用户新的正常标签同时出现概率高的检测标签；

[0087] 重新为用户添加新的检测标签，新添加的检测标签与用户现有标签同时出现的概率低于第一阈值。

[0088] 对应地，还需要对渠道索引进行更新：

[0089] 从渠道索引中除去被删除检测标签相关项。

[0090] 从而更新了渠道索引，以便再次拦截广告时，采用新的渠道索引来检测疑似泄漏渠道。

[0091] 如图2所示，一种标签数据泄漏渠道检测装置，用于检测用户标签数据的泄漏渠道，该检测装置包括：

[0092] 检测标签添加模块，用于在用户拥有的正常标签基础上为用户添加检测标签，生成用户标签数据集；

[0093] 渠道关联模块，用于根据用户标签数据集为给定的渠道赋予检测标签，建立用户ID、检测标签和渠道ID相关联的渠道索引；

[0094] 拦截模块，用于根据由用户正常标签产生推送信息的概率，对用户接收的推送信息进行拦截；

[0095] 拦截信息分析模块，用于对于拦截的推送信息，根据由该用户检测标签产生该推送信息的概率进行筛选，如果由该用户检测标签产生该推送信息的概率高于给定的阈值，则将该用户检测标签加入到疑似泄漏标签集合；

[0096] 渠道检索模块，用于根据疑似泄漏标签集合，搜索渠道索引，得到对应的疑似泄漏渠道ID列表；

[0097] 输出模块，用于检测该推送信息是否来源于所搜索到的渠道，如果是，删除对应的渠道，将剩下的渠道作为疑似泄漏渠道输出。

[0098] 容易理解的是，本实施例的装置可以应用于应用系统的后台服务器，其中拦截模块可以集成在用户终端，在用户终端侧进行拦截，该拦截模块可以采用第三方的客户端如安全卫士，或专门的客户端来进行拦截。

[0099] 本实施例检测标签添加模块在用户拥有的正常标签基础上为用户添加检测标签时，新添加的检测标签与用户现有标签同时出现的概率低于设定的第一阈值。即新生成的检测标签与用户标签集中现有标签均不相似，具有差异性，同时出现的概率低，从而不会相互发生影响。

[0100] 本实施例渠道关联模块在根据用户标签数据集为给定的渠道赋予检测标签时，执行如下操作：

[0101] 对于给定的渠道,根据其历史行为计算其可信度;

[0102] 以该渠道的渠道ID作为变量,从设定的HASH函数集中选取一个HASH函数;

[0103] 基于渠道可信度抽样用户群;

[0104] 对抽样得到的用户群中每一个用户,以用户ID作为变量,根据抽取得到的HASH函数从该用户的检测标签中选出该渠道对应的检测标签;

[0105] 建立[用户ID、检测标签]到渠道ID的渠道索引。

[0106] 本实施例拦截模块在根据由用户正常标签产生推送信息的概率,对用户接收的推送信息进行拦截时,执行如下操作:

[0107] 如果推送信息由正常标签产生的概率低于设定的第二阈值,则进行拦截,否则向用户展示该推送信息。

[0108] 本实施例检测标签添加模块还用于根据用户正常标签的变化更新用户检测标签的步骤,具体执行如下步骤:

[0109] 根据新的正常标签与现有检测标签同时出现概率,删除与用户新的正常标签同时出现概率高的检测标签;

[0110] 重新为用户添加新的检测标签,新添加的检测标签与用户现有标签同时出现的概率低于第一阈值。

[0111] 本实施例渠道关联模块还用于从渠道索引中除去被删除检测标签相关项。从而在用户产生新的正常标签时,及时对用户标签集进行更新。

[0112] 以上实施例仅用以说明本发明的技术方案而非对其进行限制,在不背离本发明精神及其实质的情况下,熟悉本领域的技术人员当可根据本发明作出各种相应的改变和变形,但这些相应的改变和变形都应属于本发明所附的权利要求的保护范围。

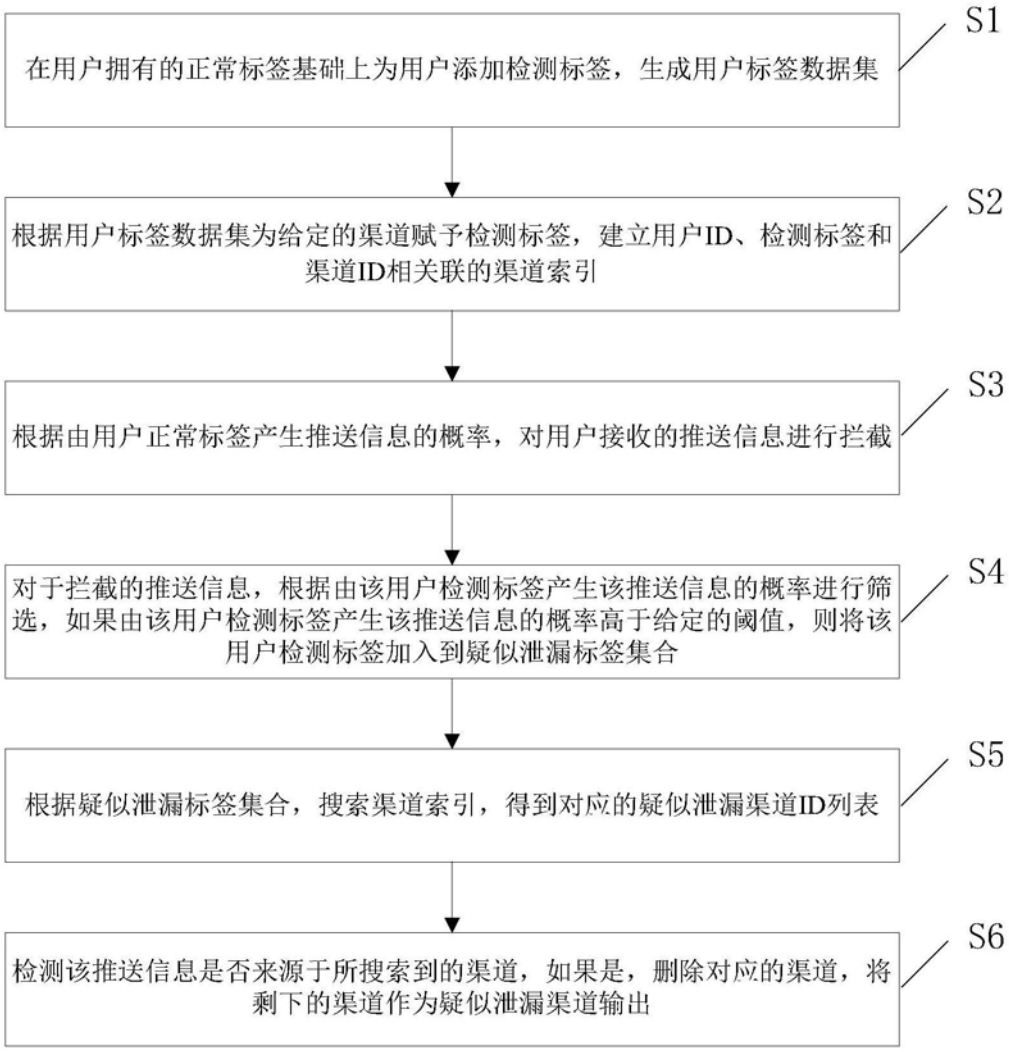

在用户拥有的正常标签基础上为用户添加检测标签，生成用户标签数据集
S1
根据用户标签数据集为给定的渠道赋予检测标签，建立用户ID、检测标签和渠道ID相关联的渠道索引
S2
根据由用户正常标签产生推送信息的概率，对用户接收的推送信息进行拦截
S3
对于拦截的推送信息，根据由该用户检测标签产生该推送信息的概率进行筛选，如果由该用户检测标签产生该推送信息的概率高于给定的阈值，则将该用户检测标签加入到疑似泄漏标签集合
S4
根据疑似泄漏标签集合，搜索渠道索引，得到对应的疑似泄漏渠道ID列表
S5
检测该推送信息是否来源于所搜索到的渠道，如果是，删除对应的渠道，将剩下的渠道作为疑似泄漏渠道输出
S6

图1

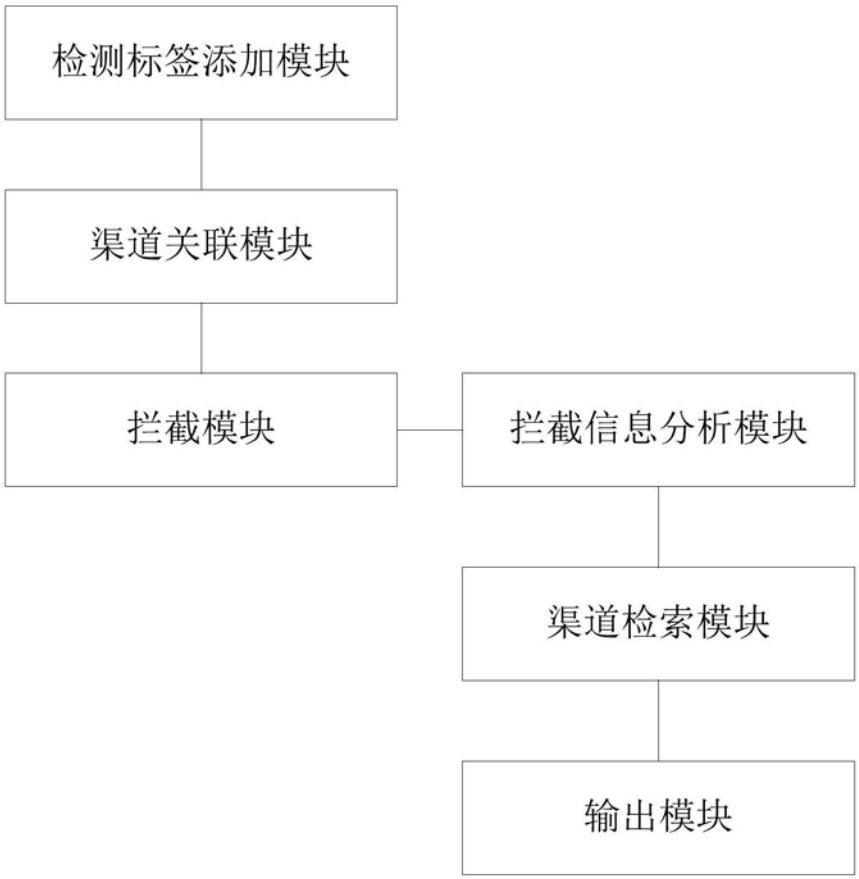


图2