(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 109145030 A
(43)申请公布日 2019.01.04

(21)申请号 201810668596.0

(22)申请日 2018.06.26

(71)申请人 阿里巴巴集团控股有限公司
地址 英属开曼群岛大开曼

(72)发明人 金璐 应杭

(74)专利代理机构 北京三友知识产权代理有限公司 11127
代理人 李辉

(51)Int.Cl.
G06F 16/2458(2019.01)
G06K 9/62(2006.01)

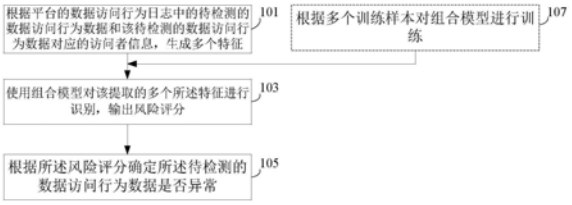
权利要求书2页 说明书7页 附图4页

(54)发明名称

一种异常数据访问的检测方法和装置

(57)摘要

本说明书实施例提供一种异常数据访问的检测方法和装置，该方法包括：根据平台的数据访问行为日志中的待检测的数据访问行为数据和该待检测的数据访问行为数据对应的访问者信息，生成多个特征；使用组合模型对所述多个特征进行识别，输出风险评分，其中，所述组合模型中包括至少一个无监督分类模型和至少一个半监督分类模型；以及根据所述风险评分确定所述待检测的数据访问行为数据是否异常。



CN 109145030 A

1.一种异常数据访问的检测方法,包括:

根据平台的数据访问行为日志中的待检测的数据访问行为数据和该待检测的数据访问行为数据对应的访问者信息,生成多个特征;

使用组合模型对所述多个特征进行识别,输出风险评分,其中,所述组合模型中包括至少一个无监督分类模型和至少一个半监督分类模型;以及

根据所述风险评分确定所述待检测的数据访问行为数据是否异常。

2.如权利要求1所述的检测方法,其中,所述多个特征包括:

访问者风险特征、平台风险特征、访问者访问行为特征和被访问数据的风险特征。

3.如权利要求1所述的检测方法,其中,所述方法还包括:

根据多个训练样本对所述组合模型进行训练,

其中,在对多个所述特征进行识别时,使用训练完成后的所述组合模型进行所述识别。

4.如权利要求3所述的检测方法,其中,根据训练样本对所述组合模型进行训练,包括:

根据平台的数据访问行为日志中的作为训练样本的数据访问行为数据和该作为训练样本的数据访问行为数据对应的访问者信息,生成多个训练用特征的组合,其中,部分训练样本被标记为异常;以及

根据该多个训练用特征的组合,对所述组合模型进行训练。

5.如权利要求4所述的检测方法,其中,根据训练样本对所述组合模型进行训练,还包括:

根据被标记为异常的训练样本在所述多个训练样本中的数量占比,调整所述组合模型中所述无监督分类模型的识别结果的权重和所述半监督分类模型的识别结果的权重。

6.一种异常数据访问的检测装置,包括:

第一生成单元,其根据平台的数据访问行为日志中的待检测的数据访问行为数据和该待检测的数据访问行为数据对应的访问者信息,生成多个特征;

识别单元,其使用组合模型对所述多个特征进行识别,输出风险评分,其中,所述组合模型中包括至少一个无监督分类模型和至少一个半监督分类模型;以及

判断单元,其根据所述风险评分确定所述待检测的数据访问行为数据是否异常。

7.如权利要求6所述的检测装置,其中,所述多个特征包括:

访问者风险特征、平台风险特征、访问者访问行为特征和被访问数据的风险特征。

8.如权利要求6所述的检测装置,其中,所述装置还包括:

训练单元,其根据多个训练样本对所述组合模型进行训练,

其中,所述识别单元在对多个所述特征进行识别时,使用训练完成后的所述组合模型进行所述识别。

9.如权利要求8所述的检测装置,其中,所述训练单元包括:

第二生成单元,其根据平台的数据访问行为日志中的作为训练样本的数据访问行为数据和该作为训练样本的数据访问行为数据对应的访问者信息,生成多个训练用特征的组合,其中,部分训练样本被标记为异常;以及

模型训练单元,其根据该多个训练用特征的组合,对所述组合模型进行训练。

10.如权利要求9所述的检测装置,其中,

所述模型训练单元还根据被标记为异常的训练样本在所述多个训练样本中的数量占

比，调整所述组合模型中所述无监督分类模型的识别结果的权重和所述半监督分类模型的识别结果的权重。

一种异常数据访问的检测方法和装置

技术领域

[0001] 本说明书涉及电子信息领域，尤其涉及一种异常数据访问的检测方法和装置。

背景技术

[0002] 对公司各类信息系统的数据访问进行安全管控，例如实现员工的异常数据访问行为检测，可以实现敏感信息防泄漏，保障数据安全。该异常数据访问行为例如包括批量数据下载、批量文件下载、批量数据查询等等。

[0003] 现有的员工异常数据访问行为的检测算法可分为两类：

[0004] 1)基于规则引擎的异常检测算法：提取特征参数，判断是否超过预先设定的阈值；

[0005] 2)基于有监督机器学习的算法：提取数据访问过程中的行为特征，然后利用大量的正负样本对员工行为进行训练，确定模型参数，然后基于训练好的模型判断员工行为是否属于异常。

[0006] 应该注意，上面对技术背景的介绍只是为了方便对本说明书的技术方案进行清楚、完整的说明，并方便本领域技术人员的理解而阐述的。不能仅仅因为这些方案在本说明书的背景技术部分进行了阐述而认为上述技术方案为本领域技术人员所公知。

发明内容

[0007] 本说明书的发明人发现，上述现有的员工异常数据访问行为检测算法都存在各自的缺陷，例如：基于规则引擎的异常检测算法只能实现粗层次的行为识别，无法自适应的调整规则和风险阈值；基于有监督机器学习的算法极大的依赖训练样本中已确定的异常行为的样本量，存在冷启动的问题，此外，该方法缺乏对未知的异常行为的识别能力。其中，训练样本中已确定的异常行为通常称为黑样本。

[0008] 本说明书实施例提供一种异常数据访问的检测方法及装置，使用包括至少一个无监督分类模型和至少一个半监督分类模型的组合模型对多个特征进行识别，从而检测异常的数据访问，由此，适用于多种场景下，访问者在数据访问过程中的异常行为检测，且无需大量该类异常行为的黑样本数据对模型进行训练。

[0009] 为了实现上述目的，本说明书提供一种异常数据访问的检测方法，包括：

[0010] 根据平台的数据访问行为日志中的待检测的数据访问行为数据和该待检测的数据访问行为数据对应的访问者信息，生成多个特征；使用组合模型对所述多个特征进行识别，输出风险评分，其中，所述组合模型中包括至少一个无监督分类模型和至少一个半监督分类模型；以及根据所述风险评分确定所述待检测的数据访问行为数据是否异常。

[0011] 本说明书还提供一种异常数据访问的检测装置，包括：

[0012] 第一生成单元，其根据平台的数据访问行为日志中的待检测的数据访问行为数据和该待检测的数据访问行为数据对应的访问者信息，生成多个特征；识别单元，其使用组合模型对所述多个特征进行识别，输出风险评分，其中，所述组合模型中包括至少一个无监督分类模型和至少一个半监督分类模型；以及判断单元，其根据所述风险评分确定所述待检

测的数据访问行为数据是否异常。

[0013] 本说明书的有益效果在于:适用于多种场景下,访问者在数据访问过程中的异常行为检测,且无需大量该类异常行为的黑样本数据对模型进行训练。

[0014] 参照后文的说明和附图,详细公开了本发明的特定实施方式,指明了本发明的原理可以被采用的方式。应该理解,本发明的实施方式在范围上并不因而受到限制。在所附权利要求的精神和条款的范围内,本发明的实施方式包括许多改变、修改和等同。

[0015] 针对一种实施方式描述和/或示出的特征可以以相同或类似的方式在一个或更多个其它实施方式中使用,与其它实施方式中的特征相组合,或替代其它实施方式中的特征。

[0016] 应该强调,术语"包括/包含"在本文使用时指特征、整件、步骤或组件的存在,但并不排除一个或更多个其它特征、整件、步骤或组件的存在或附加。

附图说明

[0017] 为了更清楚地说明本说明书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的前提下,还可以根据这些附图获得其他的附图。

[0018] 图1是本说明书实施例1的异常数据访问的检测的方法的一个示意图;

[0019] 图2是本说明书实施例1的组合模型的一个示意图;

[0020] 图3是本说明书实施例1的对组合模型进行训练的方法的一个示意图;

[0021] 图4是使用本实施例的检测方法检测异常数据访问一个实例的示意图;

[0022] 图5是本说明书实施例2的异常数据访问的检测装置的一个示意图;

[0023] 图6是本说明书实施例2的训练单元的一个示意图;

[0024] 图7是本说明书实施例3的电子设备的一个构成示意图。

具体实施方式

[0025] 实施例1

[0026] 本说明书实施例1提供一种异常数据访问的检测方法。图1是该方法的一个示意图,如图1所示,该方法包括:

[0027] 步骤101、根据平台的数据访问行为日志中的待检测的数据访问行为数据和该待检测的数据访问行为数据对应的访问者信息,生成多个特征;

[0028] 步骤103、使用组合模型对该提取的多个所述特征进行识别,输出风险评分,其中,所述组合模型中包括至少一个无监督分类模型和至少一个半监督分类模型;

[0029] 步骤105、根据所述风险评分确定所述待检测的数据访问行为数据是否异常。

[0030] 本实施例的检测方法适用于多种场景下,访问者在数据访问过程中的异常行为检测,且无需大量该类异常行为的黑样本数据对模型进行训练。

[0031] 在本实施例中,平台例如可以是企业的内部数据平台。访问者可以是企业中等能够访问该企业的内部数据平台的人员,例如企业的员工等,员工是企业的内部数据平台的特殊的访问者,员工的访问权限和/或访问的痕迹记录都不同于一般的访问者,例如:员工能够被允许访问企业内部数据平台上敏感等级较高的数据;和/或员工对企业内部数据平

台的访问的痕迹记录更为详细，并且该痕迹记录能够保留更长的时间。对于企业员工之外的访问者，通常仅能允许访问企业内部数据平台上敏感等级较低的数据甚至没有权限访问企业的内部数据平台，并且，访问的痕迹记录一般比较简单，保留时间也较短。

[0032] 在本实施例的步骤101中，可以通过访问者信息数据挖掘员工自身风险特征，如员工中的待离职或外包属性人员一般具有较高风险。通过访问者在各类平台的数据访问行为日志挖掘出该平台风险特征、该平台中被访问的数据风险特征和访问者的访问行为特征等。通过生成多个特征，能够全面刻画员工的数据访问行为异常程度和风险程度，提高检测的准确性。

[0033] 在本实施例的步骤101中，根据平台的数据访问行为日志中的待检测的数据访问行为数据和该待检测的数据访问行为数据对应的访问者信息，生成上述的多个特征，其中，上述的多个特征可以是特征向量的形式。在本实施例中，生成上述多个特征所用的具体算法可以参考现有技术。

[0034] 在本实施例的步骤103中，使用组合模型对步骤101中生成的多个特征进行识别，输出风险评分。其中，该组合模型中包括至少一个无监督分类模型和至少一个半监督分类模型。

[0035] 图2是本实施例1的组合模型200的一个示意图。如图2所示，该组合模型200中可以包括2个无监督分类模型和1个半监督分类模型，其中，该2个无监督分类模型可以是孤立森林(Isolation Forest)模型201和一类支持向量机(one-class SVM)模型202，该1个半监督分类模型可以是正例和无标记学习(Positive and unlabeled learning)模型203。此外，本实施例可以不限于此，无监督分类模型和半监督分类模型可以是具有其它数量，并且可以是其他种类。

[0036] 在本实施例中，步骤103所使用的组合模型可以是经过训练的组合模型。对该组合模型进行训练的方法见后述的说明。

[0037] 在本实施例的步骤103中，组合模型对多个特征进行识别，可以输出风险评分，例如，组合模型中的各模型可以分别对多个特征进行并行识别，得到各自的识别结果，并将各自的识别结果按照一定的权重进行加权，得到该风险评分进行输出。

[0038] 在本实施例的步骤105中，该风险评分可以与一个阈值进行比较，如果风险评分等于或高于该阈值，判断为该待检测的数据访问行为数据异常，如果风险评分低于该阈值，判断为该待检测的数据访问行为数据正常。

[0039] 此外，当该待检测的数据访问行为数据被判断为异常时，可以对该数据访问行为数据标记为异常，从而将其作为黑样本，该黑样本可以被输入到训练样本集，用于对组合模型进行进一步的训练，从而进一步提高该组合模型的识别准确性。

[0040] 在本实施例中，如图1所示，该方法还包括：

[0041] 步骤107、根据多个训练样本对组合模型进行训练。

[0042] 其中，步骤107训练完成的该组合模型可以被用于步骤103中，对步骤101生成的多个特征进行识别。

[0043] 此外，在本实施例中，可以根据一定的周期或条件来进行步骤107的训练，而不需要每次在执行步骤103之前都通过步骤107进行一次训练。

[0044] 图3是对组合模型进行训练的方法的一个示意图，如图3所示，进行训练的方法可

以包括：

[0045] 步骤301、根据平台的数据访问行为日志中的作为训练样本的数据访问行为数据和该作为训练样本的数据访问行为数据对应的访问者信息，生成多个训练用特征的组合，其中，部分训练样本被标记为异常；以及

[0046] 步骤303、根据该多个训练用特征的组合，对所述组合模型进行训练。

[0047] 在本实施例中，步骤301与步骤101类似，区别在于针对的对象不同，步骤101针对待检测的数据访问行为数据及其对应的访问者的信息生成特征，而步骤301针对样本集中作为每个样本的数据访问行为数据及其对应的访问者的信息生成特征。

[0048] 在本实施例的步骤303中，可以根据该多个训练用特征的组合，对该组合模型中的各无监督分类模型和半监督分类模型进行训练，例如，对各无监督分类模型和半监督分类模型中网络各节点的权值进行训练。其中，对各无监督分类模型和半监督分类模型进行训练的具体方法可以参考现有技术。

[0049] 如图3所示，对组合模型进行训练的方法还可以包括：

[0050] 步骤305、对各无监督分类模型和各半监督分类模型的各自的识别结果的权重进行训练。

[0051] 例如，可以根据被标记为异常的训练样本在用于训练的多个训练样本中的数量占比，调整该组合模型中各无监督分类模型的识别结果的权重和各半监督分类模型的识别结果的权重。由此，能根据各类场景下已标注的黑样本量自动进行组合模型中各类模型的权重寻优，使得组合模型的稳定性和鲁棒性更强。

[0052] 图4是使用本实施例的检测方法检测异常数据访问一个实例的示意图。如图4所示，该实例的可以包括如下的步骤：

[0053] 步骤401、对应于上述步骤101和步骤301，在该步骤401中，对员工各类平台数据访问行为日志和员工信息进行收集和处理，该处理例如可以是调整数据或信息格式等处理；

[0054] 步骤403、根据步骤401的处理结果，生成多个特征；

[0055] 步骤405、基于训练样本生成的特征A被用于对组合模型进行训练；

[0056] 步骤407、基于待检测的数据访问行为数据生成的特征B被输入到训练完成的组合模型中，用于对该待检测的数据访问行为数据进行检测，在该步骤407中，组合模型输出风险评分；

[0057] 步骤409、根据组合模型输出的风险评分和预设的风险阈值，判断该待检测的数据访问行为数据是否异常。

[0058] 根据本实施例，在组合模型中将无监督或半监督分类模型进行组合，并进行并行计算。在训练过程中不需要事先标注样本，可有效的避免冷启动问题；并且，组合模型可识别出多种异常点，有效识别新的攻击手段。

[0059] 实施例2

[0060] 本实施提供一种异常数据访问的检测装置，与实施例1的方法对应。

[0061] 图5是异常数据访问的检测装置的一个示意图，如图5所示，该装置500包括：第一生成单元501，识别单元502和判断单元503。

[0062] 其中，第一生成单元501根据平台的数据访问行为日志中的待检测的数据访问行为数据和该待检测的数据访问行为数据对应的访问者信息，生成多个特征；识别单元502使

用组合模型对所述多个特征进行识别，输出风险评分，其中，所述组合模型中包括至少一个无监督分类模型和至少一个半监督分类模型；判断单元503根据所述风险评分确定所述待检测的数据访问行为数据是否异常。

[0063] 如图5所示，装置500还包括：

[0064] 训练单元504，其根据多个训练样本对所述组合模型进行训练。

[0065] 其中，识别单元502在对多个所述特征进行识别时，使用训练单元504训练完成后的所述组合模型进行所述识别。

[0066] 图6是训练单元的一个示意图，如图6所示，该训练单元504包括：

[0067] 第二生成单元601，其根据平台的数据访问行为日志中的作为训练样本的数据访问行为数据和该作为训练样本的数据访问行为数据对应的访问者信息，生成多个训练用特征的组合，其中，部分训练样本被标记为异常；模型训练单元602，其根据该多个训练用特征的组合，对所述组合模型进行训练。

[0068] 在本实施例中，模型训练单元602还根据被标记为异常的训练样本在所述多个训练样本中的数量占比，调整所述组合模型中所述无监督分类模型的识别结果的权重和所述半监督分类模型的识别结果的权重。

[0069] 本实施例2中关于各单元的说明，可以参考实施例1中对于各步骤的说明。

[0070] 根据本实施例，在组合模型中将无监督或半监督分类模型进行组合，并进行并行计算。在训练过程中不需要事先标注样本，可有效的避免冷启动问题；并且，组合模型可识别出多种异常点，有效识别新的攻击手段。

[0071] 实施例3

[0072] 本说明书实施例3提供一种电子设备，所述电子设备包括：如实施例2所述的使用分类器进行分类的装置。

[0073] 图7是本说明书实施例3的电子设备的一个构成示意图。如图7所示，电子设备700可以包括：中央处理器(CPU)701和存储器702；存储器702耦合到中央处理器801。其中该存储器702可存储各种数据；此外，还存储用于执行异常数据访问的检测的方法的程序，并且在中央处理器801的控制下执行该程序。

[0074] 在一个实施方式中，异常数据访问的检测装置的功能可以被集成到中央处理器701中。

[0075] 其中，中央处理器701可以被配置为：

[0076] 根据平台的数据访问行为日志中的待检测的数据访问行为数据和该待检测的数据访问行为数据对应的访问者信息，生成多个特征；使用组合模型对所述多个特征进行识别，输出风险评分，其中，所述组合模型中包括至少一个无监督分类模型和至少一个半监督分类模型；以及根据所述风险评分确定所述待检测的数据访问行为数据是否异常。

[0077] 在本实施例中，所述多个特征包括：访问者风险特征、平台风险特征、访问者访问行为特征和被访问数据的风险特征。

[0078] 在本实施例中，中央处理器701还可以被配置为：根据多个训练样本对所述组合模型进行训练，其中，在对多个所述特征进行识别时，使用训练完成后的所述组合模型进行所述识别。

[0079] 在本实施例中，中央处理器701还可以被配置为：根据平台的数据访问行为日志中

的作为训练样本的数据访问行为数据和该作为训练样本的数据访问行为数据对应的访问者信息，生成多个训练用特征的组合，其中，部分训练样本被标记为异常；以及根据该多个训练用特征的组合，对所述组合模型进行训练。

[0080] 在本实施例中，中央处理器701还可以被配置为：根据被标记为异常的训练样本在所述多个训练样本中的数量占比，调整所述组合模型中所述无监督分类模型的识别结果的权重和所述半监督分类模型的识别结果的权重。

[0081] 此外，如图7所示，电子设备700还可以包括：输入输出单元703和显示单元704等；其中，上述部件的功能与现有技术类似，此处不再赘述。值得注意的是，电子设备700也并不是必须要包括图7中所示的所有部件；此外，电子设备700还可以包括图7中没有示出的部件，可以参考现有技术。

[0082] 本说明书实施例还提供一种计算机可读程序，其中当在异常数据访问的检测装置或电子设备中执行所述程序时，所述程序使得异常数据访问的检测装置或电子设备执行实施例1所述的异常数据访问的检测方法。

[0083] 本说明书实施例还提供一种存储有计算机可读程序的存储介质，其中，所述存储介质存储上述计算机可读程序，所述计算机可读程序使得异常数据访问的检测装置或电子设备执行实施例1所述的异常数据访问的检测方法。

[0084] 结合本发明实施例描述的异常数据访问的检测装置可直接体现为硬件、由处理器执行的软件功能单元或二者组合。例如，图5和6中所示的功能框图中的一个或多个和/或功能框图的一个或多个组合，既可以对应于计算机程序流程的各个软件功能单元，亦可以对应于各个硬件模块。这些软件功能单元，可以分别对应于实施例1所示的各个步骤。这些硬件模块例如可利用现场可编程门阵列(FPGA)将这些软件功能单元固化而实现。本说明书实施例方法所述的功能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时，可以存储在一个计算设备可读取存储介质中。基于这样的理解，本说明书实施例对现有技术做出贡献的部分或者该技术方案的部分可以以软件产品的形式体现出来，该软件产品存储在一个存储介质中，包括若干指令用以使得一台计算设备(可以是个人计算机，服务器，移动计算设备或者网络设备等)执行本说明书各个实施例所述方法的全部或部分步骤。而前述的存储介质包括：U盘、移动硬盘、只读存储器(ROM，Read-Only Memory)、随机存取存储器(RAM，Random Access Memory)、磁碟或者光盘等各种可以存储程序代码的介质。

[0085] 针对图5和6描述的功能框图中的一个或多个和/或功能框图的一个或多个组合，可以实现为用于执行本说明书所描述功能的通用处理器、数字信号处理器(DSP)、专用集成电路(ASIC)、现场可编程门阵列(FPGA)或其它可编程逻辑器件、分立门或晶体管逻辑器件、分立硬件组件、或者其任意适当组合。针对图5和6描述的功能框图中的一个或多个和/或功能框图的一个或多个组合，还可以实现为计算设备的组合，例如，DSP和微处理器的组合、多个微处理器、与DSP通信结合的一个或多个微处理器或者任何其它这种配置。

[0086] 本说明书中各个实施例采用递进的方式描述，每个实施例重点说明的都是与其它实施例的不同之处，各个实施例之间相同或相似部分互相参见即可。

[0087] 对所公开的实施例的上述说明，使本领域专业技术人员能够实现或使用本说明书。对这些实施例的多种修改对本领域的专业技术人员来说将是显而易见的，本文中所定义的一般原理可以在不脱离本说明书的精神或范围的情况下，在其它实施例中实现。因此，

本说明书将不会被限制于本文所示的这些实施例,而是要符合与本文所公开的原理和新颖特点相一致的最宽的范围。

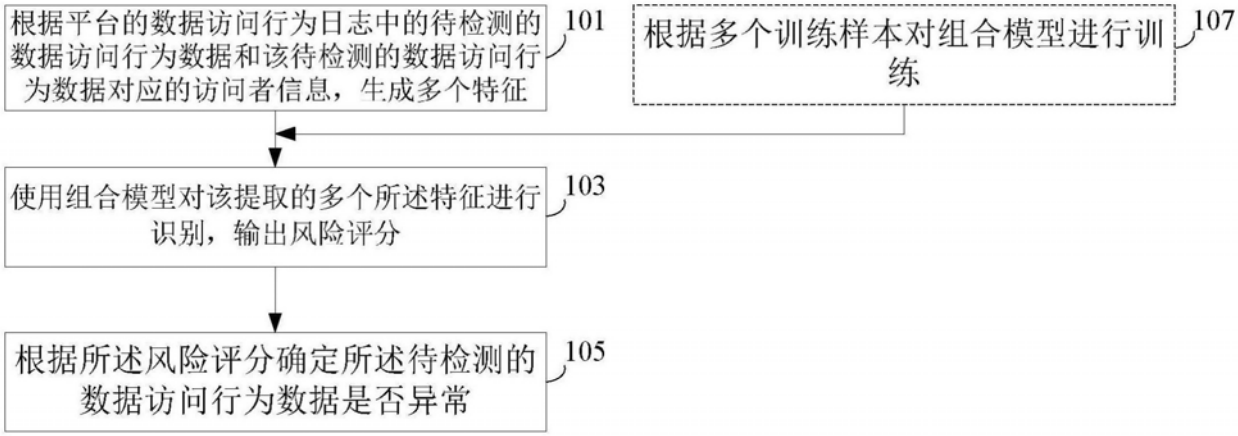

根据平台的数据访问行为日志中的待检测的数据访问行为数据和该待检测的数据访问行为数据对应的访问者信息，生成多个特征
101
根据多个训练样本对组合模型进行训练
107
使用组合模型对该提取的多个所述特征进行识别，输出风险评分
103
根据所述风险评分确定所述待检测的数据访问行为数据是否异常
105

图1

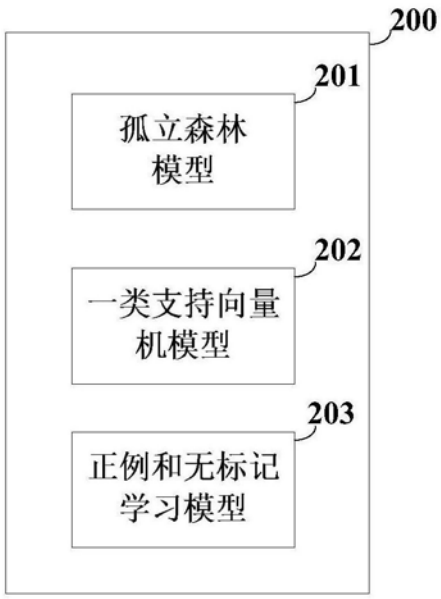

200
201
孤立森林模型
202
一类支持向量机模型
203
正例和无标记学习模型

图2

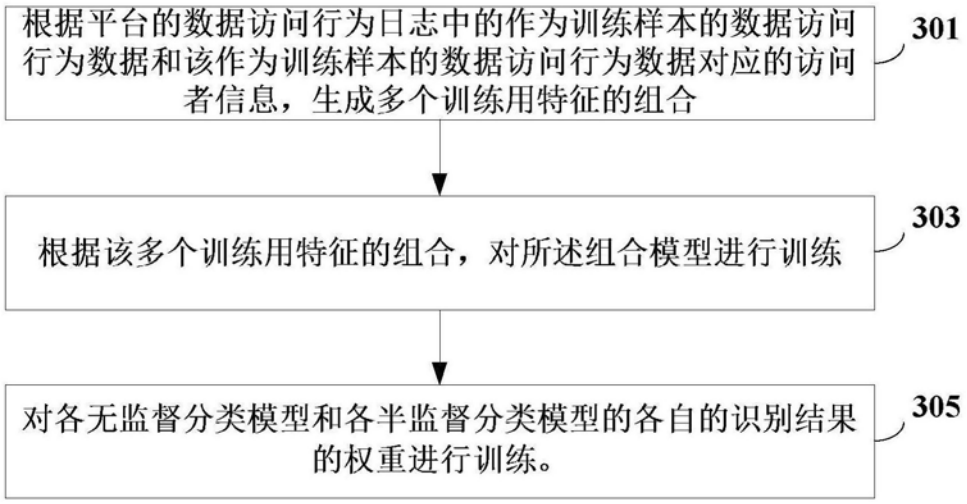

根据平台的数据访问行为日志中的作为训练样本的数据访问行为数据和该作为训练样本的数据访问行为数据对应的访问者信息，生成多个训练用特征的组合
301
根据该多个训练用特征的组合，对所述组合模型进行训练
303
对各无监督分类模型和各半监督分类模型的各自的识别结果的权重进行训练。
305

图3

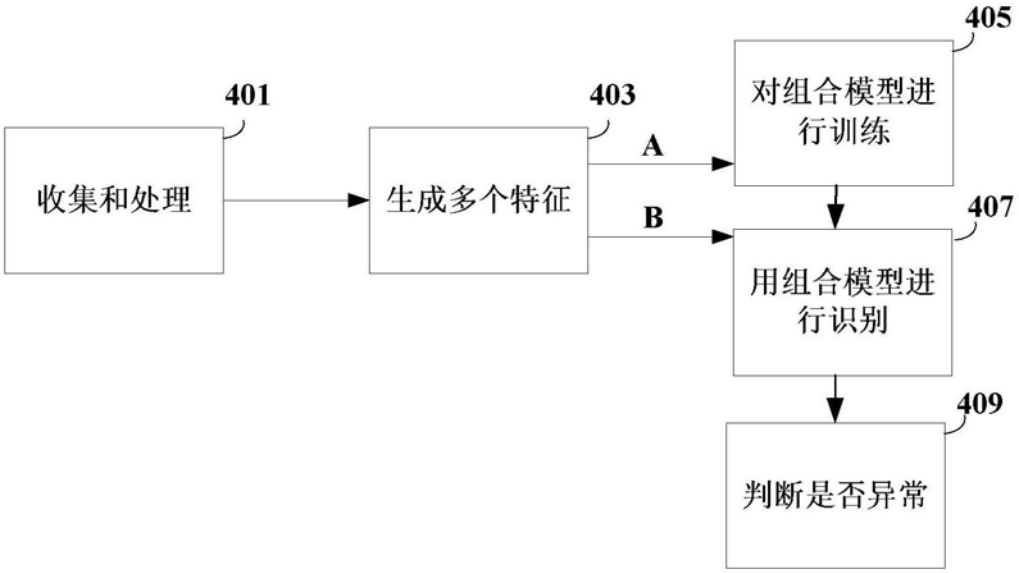

401
收集和处理
403
生成多个特征
A
B
405
对组合模型进行训练
407
用组合模型进行识别
409
判断是否异常

图4

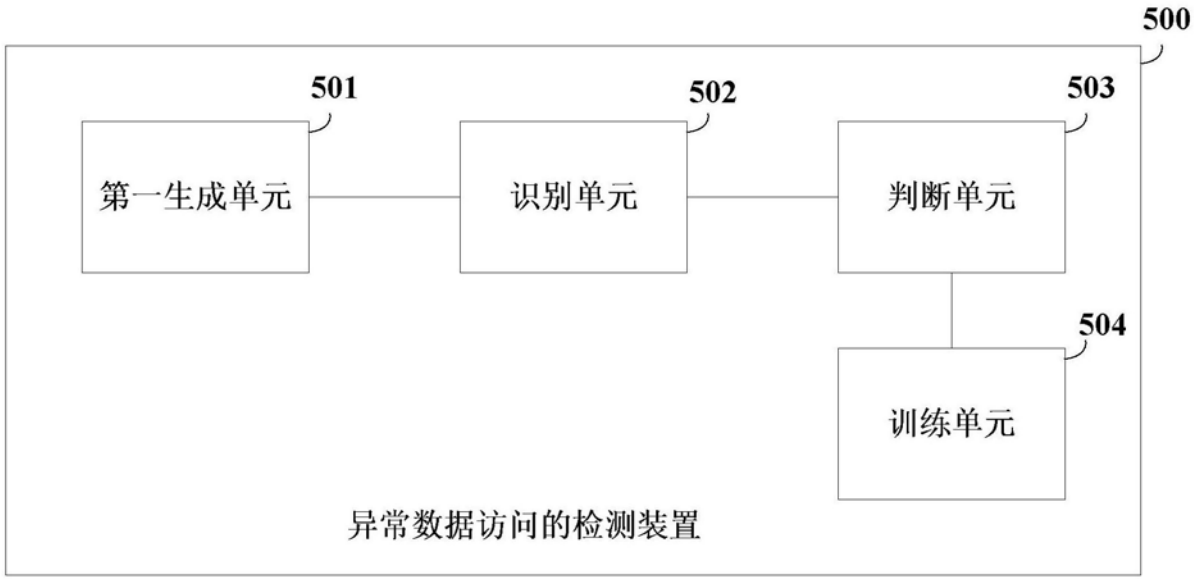

500
501
502
503
第一生成单元
识别单元
判断单元
504
训练单元
异常数据访问的检测装置

图5

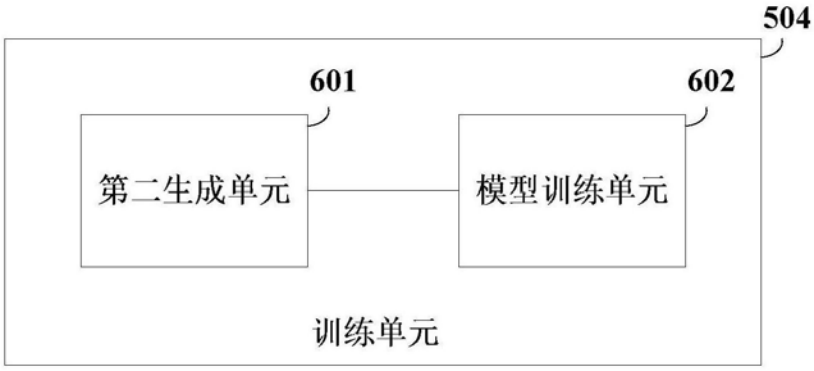

504
601
602
第二生成单元
模型训练单元
训练单元

图6

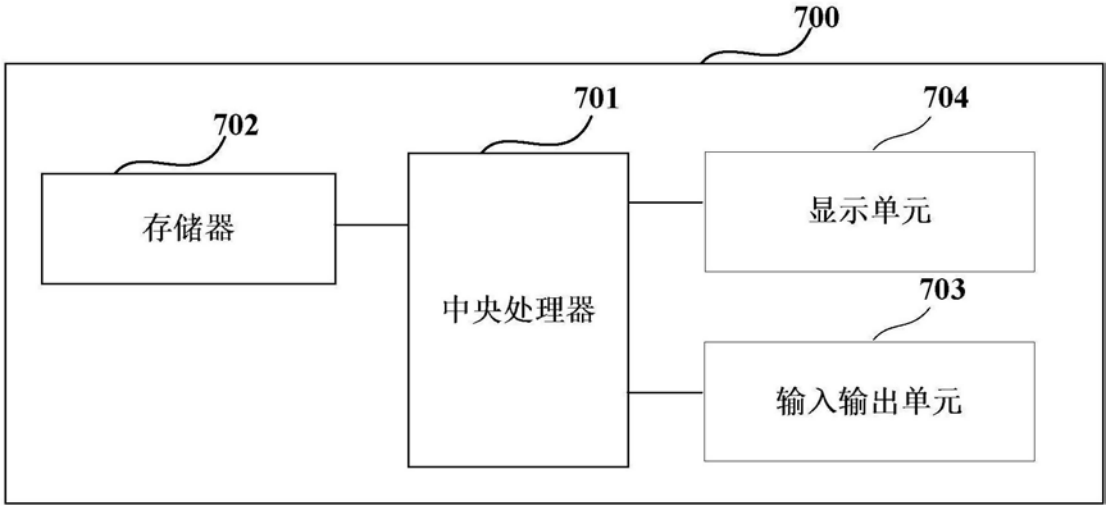


图7