



(12) 发明专利

(10) 授权公告号 CN 109388548 B

(45) 授权公告日 2020.12.22

(21) 申请号 201811145255.1

(22) 申请日 2018.09.29

(65) 同一申请的已公布的文献号
申请公布号 CN 109388548 A

(43) 申请公布日 2019.02.26

(73) 专利权人 京东数字科技控股有限公司
地址 100176 北京市大兴区经济技术开发区
区科创十一街18号C座2层221室

(72) 发明人 聂健 祝浩 张雯 曲洪涛

(74) 专利代理机构 北京英赛嘉华知识产权代理
有限责任公司 11204

代理人 王达佐 马晓亚

(51) Int. Cl.
G06F 11/34 (2006.01)

(56) 对比文件

CN 107368718 A, 2017.11.21

CN 106649527 A, 2017.05.10

CN 107229557 A, 2017.10.03

CN 108595448 A, 2018.09.28

CN 108550052 A, 2018.09.18

US 20140095425 A1, 2014.04.03

审查员 张镭

权利要求书2页 说明书10页 附图4页

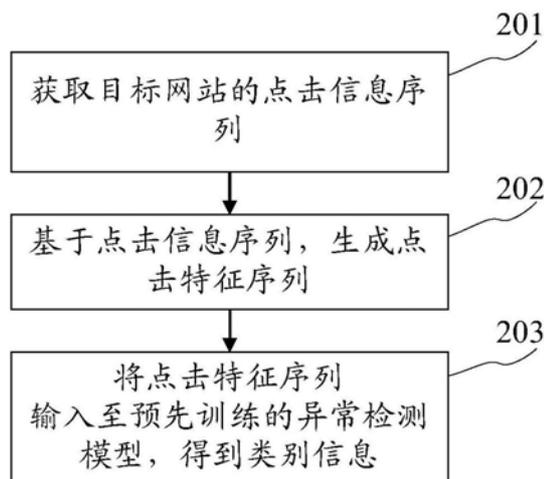
(54) 发明名称

用于生成信息的方法和装置

(57) 摘要

本申请实施例公开了用于生成信息的方法和装置。该方法的一具体实施方式包括：获取目标网站的点击信息序列；基于该点击信息序列，生成点击特征序列；将该点击特征序列输入至预先训练的异常检测模型中，得到类别信息，其中，异常检测模型用于表征点击特征序列与类别信息之间的对应关系，类别信息用于表征点击行为是否异常。该实施方式实现了对使用程序进行交互操作的识别。该实施方式实现了对用户在目标网站的异常操作行为的检测。

200



1. 一种用于生成信息的方法,包括:

获取目标网站的点击信息序列;

基于所述点击信息序列,生成点击特征序列;

将所述点击特征序列输入至预先训练的异常检测模型中,得到类别信息,其中,所述异常检测模型用于表征点击特征序列与类别信息之间的对应关系,所述类别信息用于表征点击行为是否异常;

其中,所述基于所述点击信息序列,生成点击特征序列包括:

对于所述点击信息序列中的点击信息,根据预设的历史数据关系图,生成与该点击信息对应的子图和该点击信息的属性信息,其中,所述历史数据关系图中的节点用于表征数据实体,所述历史数据关系图中的节点具有基于统计信息而得到的属性信息,所述历史数据关系图中的边用于表征用户行为,所述历史数据关系图提供查询接口,以方便按条件进行查询得到该关系图的子图;

基于所生成的属性信息,生成所述点击特征序列。

2. 根据权利要求1所述的方法,其中,所述基于所生成的点击信息的属性信息,生成所述点击特征序列,包括:

对于所述点击信息序列中的点击信息,提取以下至少一项:该点击信息所对应的页面的停留时间,该点击信息所对应的页面的内容类别;

基于所生成的属性信息和所提取的信息,生成所述点击特征序列。

3. 根据权利要求1所述的方法,其中,所述异常检测模型通过以下步骤训练得到:

获取训练样本集合,其中,训练样本包括样本点击特征序列和与样本点击特征序列对应的样本类别信息;

将所述训练样本集合中的训练样本的样本点击特征序列作为输入,将与输入的样本点击特征序列对应的样本类别信息作为期望输出,训练得到所述异常检测模型。

4. 根据权利要求1-3之一所述的方法,其中,所述方法还包括:

响应于所得到的类别信息是表征点击行为异常的信息,发送提示信息。

5. 一种用于生成信息的装置,包括:

获取单元,被配置成获取目标网站的点击信息序列;

生成单元,被配置成基于所述点击信息序列,生成点击特征序列;

识别单元,被配置成将所述点击特征序列输入至预先训练的异常检测模型中,得到类别信息,其中,所述异常检测模型用于表征点击特征序列与类别信息之间的对应关系,所述类别信息用于表征点击行为是否异常;

其中,所述生成单元包括:

属性信息生成模块,被配置成对于所述点击信息序列中的点击信息,根据预设的历史数据关系图,生成与该点击信息对应的子图和该点击信息的属性信息,其中,所述历史数据关系图中的节点用于表征数据实体,所述历史数据关系图中的节点具有基于统计信息而得到的属性信息,所述历史数据关系图中的边用于表征用户行为,所述历史数据关系图提供查询接口,以方便按条件进行查询得到该关系图的子图;

特征序列生成模块,被配置成基于所生成的属性信息,生成所述点击特征序列。

6. 根据权利要求5所述的装置,其中,所述特征序列生成模块包括:

提取子模块,被配置成对于所述点击信息序列中的点击信息,提取以下至少一项:该点击信息所对应的页面的停留时间,该点击信息所对应的页面的内容类别;

生成子模块,被配置成基于所生成的属性信息和所提取的信息,生成所述点击特征序列。

7. 根据权利要求5所述的装置,其中,所述异常检测模型通过以下步骤训练得到:

获取训练样本集合,其中,训练样本包括样本点击特征序列和与样本点击特征序列对应的样本类别信息;

将所述训练样本集合中的训练样本的样本点击特征序列作为输入,将与输入的样本点击特征序列对应的样本类别信息作为期望输出,训练得到所述异常检测模型。

8. 根据权利要求5-7之一所述的装置,其中,所述装置还包括:

发送单元,被配置成响应于所得到的类别信息是表征点击行为异常的信息,发送提示信息。

9. 一种服务器,包括:

一个或多个处理器;

存储装置,其上存储有一个或多个程序;

当所述一个或多个程序被所述一个或多个处理器执行,使得所述一个或多个处理器实现如权利要求1-4中任一所述的方法。

10. 一种计算机可读介质,其上存储有计算机程序,其中,该程序被处理器执行时实现如权利要求1-4中任一所述的方法。

用于生成信息的方法和装置

技术领域

[0001] 本申请实施例涉及计算机技术领域,具体涉及用于生成信息的方法和装置。

背景技术

[0002] 随着网络购物的快速普及,防范盗卡、盗账户进行购物变现的技术也变得愈来愈重要。相关的技术主要有两类:一是依靠专家经验总结一系列风险规则来识别风险操作;二是通过人工选择用户购物特征,通过机器学习方法建立预测模型,对用户的操作进行风险分类。

发明内容

[0003] 本申请实施例提出了用于识别交互操作的方法和装置。

[0004] 第一方面,本申请实施例提供了一种用于识别交互操作的方法,该方法包括:获取目标网站的点击信息序列;基于点击信息序列,生成点击特征序列;将点击特征序列输入至预先训练的异常检测模型中,得到类别信息,其中,异常检测模型用于表征点击特征序列与类别信息之间的对应关系,类别信息用于表征点击行为是否异常。

[0005] 在一些实施例中,基于点击信息序列,生成点击特征序列包括:对于点击信息序列中的点击信息,根据预设的历史数据关系图,生成与该点击信息对应的子图和该点击信息的属性信息;基于所生成的属性信息,生成点击特征序列。

[0006] 在一些实施例中,基于所生成的点击信息的属性信息,生成点击特征序列,包括:对于点击信息序列中的点击信息,提取以下至少一项:该点击信息所对应的页面的停留时间,该点击信息所对应的页面的内容类别;基于所生成的属性信息和所提取的信息,生成点击特征序列。

[0007] 在一些实施例中,异常检测模型通过以下步骤训练得到:获取训练样本集合,其中,训练样本包括样本点击特征序列和与样本点击特征序列对应的样本类别信息;将训练样本集合中的训练样本的样本点击特征序列作为输入,将与输入的样本点击特征序列对应的样本类别信息作为期望输出,训练得到异常检测模型。

[0008] 在一些实施例中,上述方法还包括:响应于所得到的类别信息是表征点击行为异常的信息,发送提示信息。

[0009] 第二方面,本申请实施例提供了一种用于生成信息的装置,该装置包括:获取单元,被配置成获取目标网站的点击信息序列;生成单元,被配置成基于点击信息序列,生成点击特征序列;识别单元,被配置成将点击特征序列输入至预先训练的异常检测模型中,得到类别信息,其中,异常检测模型用于表征点击特征序列与类别信息之间的对应关系,类别信息用于表征点击行为是否异常。

[0010] 在一些实施例中,生成单元包括:属性信息生成模块,被配置成对于点击信息序列中的点击信息,根据预设的历史数据关系图,生成与该点击信息对应的子图和该点击信息的属性信息;特征序列生成模块,被配置成基于所生成的属性信息,生成点击特征序列。

[0011] 在一些实施例中,特征序列生成模块包括:提取子模块,被配置成对于点击信息序列中的点击信息,提取以下至少一项:该点击信息所对应的页面的停留时间,该点击信息所对应的页面的内容类别;生成子模块,被配置成基于所生成的属性信息和所提取的信息,生成点击特征序列。

[0012] 在一些实施例中,异常检测模型通过以下步骤训练得到:获取训练样本集合,其中,训练样本包括样本点击特征序列和与样本点击特征序列对应的样本类别信息;将训练样本集合中的训练样本的样本点击特征序列作为输入,将与输入的样本点击特征序列对应的样本类别信息作为期望输出,训练得到异常检测模型。

[0013] 在一些实施例中,上述装置还包括:发送单元,被配置成响应于所得到的类别信息是表征点击行为异常的信息,发送提示信息。

[0014] 第三方面,本申请实施例提供了一种服务器,该服务器包括:一个或多个处理器;存储装置,其上存储有一个或多个程序;当一个或多个程序被一个或多个处理器执行,使得一个或多个处理器实现如第一方面中任一实现方式描述的方法。

[0015] 第四方面,本申请实施例提供了一种计算机可读介质,其上存储有计算机程序,该程序被处理器执行时实现如第一方面中任一实现方式描述的方法。

[0016] 本申请实施例提供的用于生成信息的方法和装置,首先获取目标网站的点击信息序列。然后,基于点击信息序列,生成点击特征序列。接下来,将点击特征序列输入至预先训练的异常检测模型中,得到类别信息。所得到的类别信息可以用于表征点击行为是否异常。从而通过对点击信息序列的分析实现了对用户异常操作行为的检测。

附图说明

[0017] 通过阅读参照以下附图所作的对非限制性实施例所作的详细描述,本申请的其它特征、目的和优点将会变得更明显:

[0018] 图1是本申请的一个实施例可以应用于其中的示例性系统架构图;

[0019] 图2是根据本申请的用于生成信息的方法的一个实施例的流程图;

[0020] 图3是根据本申请实施例的用于生成信息的方法的一个应用场景的示意图;

[0021] 图4是根据本申请的用于生成信息的方法的又一个实施例的流程图;

[0022] 图5是根据本申请的用于生成信息的装置的一个实施例的结构示意图;

[0023] 图6是适于用来实现本申请实施例的服务器的计算机系统的结构示意图。

具体实施方式

[0024] 下面结合附图和实施例对本申请作进一步的详细说明。可以理解的是,此处所描述的具体实施例仅仅用于解释相关发明,而非对该发明的限定。另外还需要说明的是,为了便于描述,附图中仅示出了与有关发明相关的部分。

[0025] 需要说明的是,在不冲突的情况下,本申请中的实施例及实施例中的特征可以相互组合。下面将参考附图并结合实施例来详细说明本申请。

[0026] 图1示出了可以应用本申请的用于生成信息的方法或用于生成信息的装置的示例性架构100。

[0027] 如图1所示,系统架构100可以包括终端设备101、102、103,网络104和服务器105。

网络104用以在终端设备101、102、103和服务器105之间提供通信链路的介质。网络104可以包括各种连接类型,例如有线、无线通信链路或者光纤电缆等等。

[0028] 终端设备101、102、103通过网络104与服务器105交互,以接收或发送消息等。终端设备101、102、103上可以安装有各种通讯客户端应用,例如网页浏览器应用、购物类应用、搜索类应用、即时通信工具、邮箱客户端、社交平台软件、阅读类应用等。

[0029] 终端设备101、102、103可以是硬件,也可以是软件。当终端设备101、102、103为硬件时,可以是具有显示屏并且支持网页浏览的各种电子设备,包括但不限于智能手机、平板电脑、电子书阅读器、膝上型便携计算机和台式计算机等等。当终端设备101、102、103为软件时,可以安装在上述所列举的电子设备中。其可以实现成多个软件或软件模块(例如用来提供分布式服务),也可以实现成单个软件或软件模块。在此不做具体限定。

[0030] 服务器105可以是提供各种服务的服务器,例如为终端设备101、102、103上显示网页提供支持的后台服务器。后台服务器可以对获取的点击信息序列进行分析,生成点击特征序列,进而确定点击信息序列所对应的点击行为是否异常。

[0031] 需要说明的是,服务器可以是硬件,也可以是软件。当服务器为硬件时,可以实现成多个服务器组成的分布式服务器集群,也可以实现成单个服务器。当服务器为软件时,可以实现成多个软件或软件模块(例如用来提供分布式服务),也可以实现成单个软件或软件模块。在此不做具体限定。

[0032] 需要说明的是,本申请实施例所提供的用于生成信息的方法一般由服务器105执行,相应地,用于生成信息的装置一般设置于服务器105中。

[0033] 应该理解,图1中的终端设备、网络和服务器的数目仅仅是示意性的。根据实现需要,可以具有任意数目的终端设备、网络和服务器。

[0034] 继续参考图2,示出了根据本申请的用于生成信息的方法的一个实施例的流程200。该用于生成信息的方法包括以下步骤:

[0035] 步骤201,获取目标网站的点击信息序列。

[0036] 在本实施例中,目标网站可以是根据实际的应用需求,预先指定的任意网站,例如可以是某电商网站。目标网站也可以是根据规则而定的网站,例如可以是一段时期内访问量最高的网站。

[0037] 在本实施例中,点击信息可以包括但不限于以下至少一项:用于记录用户访问网站时的浏览、点击等行为的数据,所访问的网页上的内容信息。其中,记录用户访问网站时的浏览行为的数据例如可以是用户打开和关闭网页的时刻。记录用户访问网站时的点击行为的数据例如可以是用户点击链接打开新网页的时刻,也可以是用户在网络购物过程中点击加入收藏的商品信息,还可以是用户点击的URL(Uniform Resource Locator,统一资源定位符)。所访问的网页上的内容信息可以包括网页上所展示的文字或图片信息,例如可以是网页上所展示的产品信息(例如:产品所属类别、产品名称、报价、商家名称),也可以是交易金额。点击信息序列可以是用户端浏览器与网站服务器间的一次会话(session)过程中所产生的点击信息按照时间顺序排列而成的。作为示例,上述点击信息可以是打开页面和关闭页面的时刻,也可以是用户端所采用的浏览器版本、IP(Internet Protocol,网际协议)地址、域名。

[0038] 在本实施例中,用于生成信息的方法的执行主体(如图1所示的服务器105)可以通

通过各种方法获取目标网站的点击信息序列。作为示例,上述执行主体可以从目标网站的网站服务器所记录的网站记录中获取上述点击信息序列。作为又一示例,上述执行主体可以通过在目标网站的页面上嵌入自定义的JS (Javascript,一种脚本语言) 代码来获取上述点击信息序列。

[0039] 步骤202,基于点击信息序列,生成点击特征序列。

[0040] 在本实施例中,上述执行主体可以基于从步骤201获取的点击信息序列,通过各种方式生成点击特征序列。其中,上述点击特征序列中的点击特征可以用于表征与网站服务器进行交互的客户端所对应的用户的相关信息。上述相关信息可以包括但不限于以下至少一项:客户端与网站服务器进行交互的情况,用户的信息,用户历史浏览信息,用户所在群体的信息。作为示例,上述客户端与网站服务器进行交互的情况可以包括但不限于以下至少一项:页面停留时间,页面内容所属类别,页面总访问时长。用户的信息可以包括但不限于以下至少一项:用户注册时长,性别,年龄,所在地区。用户历史浏览信息可以包括但不限于以下至少一项:最近7天是否登录,平均浏览时间,经常购买的商品类别。用户所在群体可以是依据用户不同方面的信息进行分类得到的。用户所在群体的信息可以用于表征与用户属于同一类别的其他用户的相关信息。例如,用户A浏览了与用户B浏览过的页面相同的页面,用户A所在群体的信息可以包括用户B浏览过的其他页面的信息。

[0041] 具体地,对于点击信息序列中的每个点击信息,上述执行主体可以从该点击信息中提取信息,基于所提取的信息,生成该点击信息所对应的点击特征。然后,将所生成的点击特征组合成上述点击特征序列。

[0042] 作为示例,点击信息序列可以包括2个点击信息。其中,点击信息a可以包括在 T_1 时刻打开页面X,页面X上展示了各种手机。点击信息b可以包括在 T_2 时刻关闭页面X。由此,所生成的点击特征可以是页面X的停留时间 T_2-T_1 和页面X内容所属类别(例如可以是数码产品)。因而,所生成的点击特征序列可以是 $\{T_2-T_1, \text{数码产品}\}$ 。

[0043] 步骤203,将点击特征序列输入至预先训练的异常检测模型中,得到类别信息。

[0044] 在本实施例中,上述执行主体可以将通过步骤202生成的点击特征序列输入至预先训练的异常检测模型,得到类别信息。其中,异常检测模型用于表征点击特征序列与类别信息之间的对应关系,类别信息用于表征点击行为是否异常。类别信息可以是各种形式的信息,比如:数字、文字、符号等等。作为示例,可以用“1”来表示点击行为异常,用“0”来表示点击行为正常。通常,上述点击行为指的是上述点击特征序列所对应的点击行为。

[0045] 需要说明的是,作为示例,异常检测模型可以是技术人员基于对大量的点击特征序列和类别信息的统计而预先制定的、存储有多个点击特征序列与类别信息的对应关系的对应关系表。也可以是技术人员基于对大量数据的统计而预先设置的规则。上述规则例如可以是:如果点击特征序列中特定页面的停留时间小于预设的时间阈值,则上述点击特征序列所对应的类别信息为表征点击行为异常的信息。上述特定页面例如可以是在购买价格高于预设的价格阈值的商品的订单确认页面之前浏览的针对该商品进行介绍的页面。

[0046] 在本实施例的一些可选的实现方式中,上述执行主体可以将通过步骤202生成的点击特征序列中的点击特征序列依次输入至预先训练的异常检测模型,得到与点击特征序列中的点击特征序列数目对应的类别信息。然后,可以根据所得到的类别信息,确定上述点击特征序列的类别信息。确定方式在此不作限制。作为示例,可以将占比超过半数的类别信

息作为点击特征序列的类别信息。作为又一示例,可以为每个所得到的类别信息附以不同的权重,根据计算出的结果确定类别信息。

[0047] 在本实施例的一些可选的实现方式中,异常检测模型可以是利用机器学习方法,通过大量的训练样本对各种人工神经网络训练得到的模型。作为示例,上述异常检测模型可以通过以下步骤训练得到:

[0048] 第一步,获取初始异常检测模型。

[0049] 在本实施例中,训练步骤的执行主体可以首先获取初始异常检测模型。其中,初始异常检测模型可以是现有的各种神经网络。例如,RNN (Recurrent Neural Network,循环神经网络),LSTM (Long Short-Term Memory,长短期记忆网络)等。

[0050] 第二步,获取训练样本集合。每个训练样本可以包括样本点击特征序列和与样本点击特征序列对应的样本类别信息。实践中,训练样本可以通过多种方式得到。作为示例,可以由技术人员对历史数据进行如前述步骤202的加工,将正常点击行为所对应的点击特征序列和表征点击行为正常的类别信息关联存储;将异常点击行为所对应的点击特征序列和表征点击行为异常的类别信息关联存储,最终得到训练样本。通过大量的数据形成大量的训练样本,进而组成训练样本集合。

[0051] 第三步,利用机器学习的方法,将训练样本集合中的训练样本中的样本点击特征序列作为初始异常检测模型的输入,将与输入的样本点击特征序列对应的样本类别信息作为期望输出,训练得到上述异常检测模型。

[0052] 具体地,上述训练步骤的执行主体可以将训练样本集合中的训练样本的样本点击特征序列输入初始异常检测模型,得到该训练样本的类别信息。然后,可以利用预设的损失函数计算所得到的类别信息与该训练样本的样本类别信息之间的差异程度。接下来,可以,基于计算所得的差异程度,调整初始异常检测模型的网络参数,并在满足预设的训练结束条件的情况下,结束训练。最后,将训练得到的初始异常检测模型确定为预先训练的异常检测模型。

[0053] 需要说明的是,上述预设的训练结束条件可以包括但不限于以下至少一项:训练时间超过预设时长;训练次数超过预设次数;计算所得的差异程度小于预设的差异阈值。

[0054] 还需要说明的是,基于所生成的类别信息与该训练样本的样本类别信息之间的差异程度,可以采用各种方式调整初始异常检测模型的结构参数。例如,可以采用BP (Back Propagation,反向传播) 算法或者SGD (Stochastic Gradient Descent,随机梯度下降) 算法来调整初始异常检测模型的网络参数。

[0055] 值得注意的是,上述训练步骤的执行主体可以与用于生成信息的方法的执行主体相同或者不同。如果相同,则上述训练步骤的执行主体可以在训练得到异常检测模型后将训练好的异常检测模型的结构信息和网络参数存储在本地。如果不同,则上述训练步骤的执行主体可以在训练得到异常检测模型后将训练好的异常检测模型的结构信息和网络参数发送给用于生成信息的方法的执行主体。

[0056] 在本实施例的一些可选的实现方式中,上述用于生成信息的方法的执行主体响应于所得到的类别信息是表征点击行为异常的信息,还可以发送提示信息。其中,提示信息可以是各种形式,例如文字、数字。

[0057] 继续参见图3,图3是根据本申请实施例的用于生成信息的方法的应用场景的一个

示意图。在图3的应用场景300中,用户可以通过终端设备301浏览网页并进行点击操作。作为示例,用户通过终端设备301浏览某电商网站,打开页面1和页面2。其中,页面1是单反相机的商品详情介绍页面,售价25,000RMB。页面2是通过点击操作购买页面1所展示的单反相机后跳转到的订单确认页面。与终端设备通信连接的后台网页服务器302可以获取目标网站AA的点击信息序列303。其中, t_1 可以是打开页面1的时刻, t_2 可以是跳转到页面2的时刻, x 可以是“单反相机”, y 可以是“25,000RMB”。接下来,基于上述点击信息序列303,可以生成点击特征序列304。其中, a 可以是页面1的停留时间 t_2-t_1 ,例如可以是2s; b 可以是页面2的类型“订单确认”; c 可以是交易金额“25,000RMB”。然后,服务器302可以将上述点击特征序列304输入至预先训练的异常检测模型。根据预设的规则:若点击特征序列中的订单确认金额大于20,000RMB,且相应商品详情介绍页面的停留时间小于3s,则该点击特征序列所对应的点击行为异常。服务器302可以得到表征上述点击行为异常的类别信息305,例如是“黑”。可选地,若服务器302得到表征上述点击行为异常的类别信息,可以向终端设备301发送提示信息,例如“操作异常”。

[0058] 本申请的上述实施例提供的方法,首先通过获取目标网站的点击信息序列。然后,基于上述点击信息序列,生成点击特征序列。最后,将上述点击特征序列输入至预先训练的异常检测模型中,得到类别信息。从而可以根据目标网站的点击信息,实现在目标网站上对异常操作的识别,从而可以对诸如异常购物等行为进行拦截。

[0059] 进一步参考图4,其示出了用于生成信息的方法的又一个实施例的流程400。该用于生成信息的方法的流程400,包括以下步骤:

[0060] 步骤401,获取目标网站的点击信息序列。

[0061] 步骤402,对于点击信息序列中的点击信息,根据预设的历史数据关系图,生成与该点击信息对应的子图和该点击信息的属性信息。

[0062] 在本实施例中,预设的历史数据关系图可以是基于大量的历史数据而构建的关系图。历史数据可以包括但不限于以下至少一项:用户信息、订单信息(例如,收货人、地址、手机号等)、商品信息、商家信息、商品收藏信息。关系图可以用于表征历史数据中数据之间的关联关系。

[0063] 在本实施例中,历史数据关系图可以通过如下方法构建:首先,利用信息抽取技术,将大量的、结构化的网站日志记录抽取转化为半结构化的关系图;然后,利用图数据库进行保存;最后,提供查询接口,以方便按条件进行查询得到该关系图的子图。其中,子图可以由图中与某节点关联紧密的节点集合形成的,其可以表示该节点的局部统计属性。

[0064] 需要说明的是,上述节点还可以具有基于统计信息而得到的属性信息。属性信息可以表示由点击信息所提取出来的节点的个体特征。属性信息例如可以是表征历史行为是否存在异常的标签,也可以是与该节点关联最紧密的节点信息。作为示例,若用户不存在异常行为历史,且经常购买数码产品。那么,该用户的属性信息可以是标签为“白”和数码产品。

[0065] 具体地,可以将历史数据中的用户信息、商家信息、商品信息、手机号、IP地址信息等数据看成实体(对应于图中的节点),将浏览、购买等行为看成关系(对应于图中的边),将传统的结构化的网站日志记录转化为实体间具有关联关系的关系图。作为示例,用户在某商家购买商品,即可以构建起用户节点、商家节点、商品节点、收货手机/地址等节点间的关

系图。

[0066] 在本实施例中,对于点击信息序列中的点击信息,上述执行主体可以首先从该点击信息中提取与该点击信息对应的页面相关信息。页面相关信息可以包括但不限于用户信息、商品信息和店铺信息。然后,上述执行主体可以上述页面相关信息作为查询条件,在预设的历史数据关系图中进行查询,得到一个子图。由此,点击信息序列中的每个点击信息都可以对应一个查询到的子图,并且子图中包含了所对应的点击信息中的数据及数据之间的关联关系。

[0067] 还需要说明的是,由于所得到的子图是在预设的历史数据关系图中查询得到的,子图中也可以包含所对应的点击信息中无法直接得到的信息。具体地,可以利用传播算法得到子图中节点的属性信息。由于点击信息序列中的每个点击信息对应一个子图,子图中的节点对应点击信息中的数据,因而可以将节点的属性信息作为点击信息的属性信息。其中,传播算法可以是能够用于图上的各种算法,例如LPA(Label Propagation Algorithm, 标签传播算法),AP(Affinity Propagation, 近邻传播)算法。作为示例,点击信息中的用户信息对应于图中的用户节点。可以利用上述算法通过预设的历史数据关系图中与之相连的节点的“黑/白”标签,计算出该用户节点的标签为“黑/白”的概率值。

[0068] 在本实施例中,点击信息序列中的每一次点击可以看作用户与网页内容建立一种关系。作为示例,用户作为买家通过客户端浏览电商网站,按照某一关键词搜索商品的过程就是建立了该用户与此类商品的关联关系。

[0069] 步骤403,对于点击信息序列中的点击信息,提取以下至少一项:该点击信息所对应的页面的停留时间,该点击信息所对应的页面的内容类别。

[0070] 在本实施例中,页面的内容类别可以是根据页面上的信息内容所确定的类别。作为示例,电商网站通常可以根据商品的种类来确定页面的内容类别,例如生活用品、服饰、图书音像、数码产品等。对于点击信息序列中的每条点击信息,上述执行主体可以从该点击信息中提取该点击信息所对应的页面的停留时间和该点击信息所对应的页面的内容类别。其中,对于页面的停留时间的解释和说明与前述实施例中的步骤202中的描述一致,此处不再赘述。

[0071] 在本实施例的一些可选的实现方式中,还可以从点击信息序列中提取全局特征信息。其中,全局特征信息可以包括但不限于以下至少一项:交易总金额、页面访问总时长、用户近期购物频率、常用地址、经常访问的时间段。

[0072] 步骤404,基于所生成的属性信息和所提取的信息,生成点击特征序列。

[0073] 在本实施例中,基于步骤402所生成的属性信息和步骤403所提取的信息,上述执行主体可以生成点击特征序列。

[0074] 需要说明的是,上述基于属性信息和所提取的信息,生成点击特征序列的方式与前述实施例中的步骤202中的描述相类似,只是点击特征序列中的元素有所改变,此处不再赘述。

[0075] 步骤405,将点击特征序列输入至预先训练的异常检测模型中,得到类别信息。

[0076] 上述步骤401和步骤405分别与前述实施例中的步骤201和203一致,上文针对步骤201和步骤203的描述也适用于步骤401和步骤405,此处不再赘述。

[0077] 从图4中可以看出,与图2对应的实施例相比,本实施例中的用于生成信息的方法

的流程400体现了对于点击信息序列中的点击信息,根据预设的历史数据关系图,生成与该点击信息对应的子图和该点击信息的属性信息的步骤。由此,本实施例描述的方案可以利用预设的历史数据关系图生成点击特征序列,从而使得所生成的点击特征序列更具有个体行为习惯特征,进而有助于提升异常检测的准确度。

[0078] 进一步参考图5,作为对上述各图所示方法的实现,本申请提供了用于生成信息的装置的一个实施例,该装置实施例与图2所示的方法实施例相对应,该装置具体可以应用于各种电子设备中。

[0079] 如图5所示,本实施例提供的用于生成信息的装置500包括获取单元501、生成单元502和识别单元503。其中,获取单元501,被配置成获取目标网站的点击信息序列。生成单元502,被配置成基于点击信息序列,生成点击特征序列。识别单元503,被配置成将点击特征序列输入至预先训练的异常检测模型中,得到类别信息,其中,异常检测模型用于表征点击特征序列与类别信息之间的对应关系,类别信息用于表征点击行为是否异常。

[0080] 在本实施例中,用于生成信息的装置500中:获取单元501、生成单元502和识别单元503的具体处理及其所带来的技术效果可分别参考图2对应实施例中的步骤201、步骤202和步骤203的相关说明,在此不再赘述。

[0081] 在本实施例的一些可选的实现方式中,上述生成单元502可以包括:属性信息生成模块(图中未示出)和特征序列生成模块(图中未示出)。其中,属性信息生成模块,被配置成对于点击信息序列中的点击信息,根据预设的历史数据关系图,生成与该点击信息对应的子图和该点击信息的属性信息。特征序列生成模块,被配置成基于所生成的属性信息,生成点击特征序列。

[0082] 在本实施例的一些可选的实现方式中,特征序列生成模块可以包括:提取子模块(图中未示出)和生成子模块(图中未示出)。其中,提取子模块,被配置成对于点击信息序列中的点击信息,提取以下至少一项:该点击信息所对应的页面的停留时间,该点击信息所对应的页面的内容类别。生成子模块,被配置成基于所生成的属性信息和所提取的信息,生成点击特征序列。

[0083] 在本实施例的一些可选的实现方式中,异常检测模型可以通过以下步骤训练得到:获取训练样本集合,其中,训练样本包括样本点击特征序列和与样本点击特征序列对应的样本类别信息;将训练样本集合中的训练样本的样本点击特征序列作为输入,将与输入的样本点击特征序列对应的样本类别信息作为期望输出,训练得到异常检测模型。

[0084] 在本实施例的一些可选的实现方式中,用于生成信息的装置500还包括:发送单元(图中未示出),被配置成响应于所得到的类别信息是表征点击行为异常的信息,发送提示信息。

[0085] 本申请的上述实施例提供的装置,首先通过获取单元501,获取目标网站的点击信息序列。然后,生成单元502基于点击信息序列,生成点击特征序列。接下来,识别单元503将点击特征序列输入至预先训练的异常检测模型中,得到类别信息。所得到的类别信息可以用于表征点击行为是否异常。从而通过对点击信息序列的分析实现了对用户异常操作行为的检测。

[0086] 下面参考图6,其示出了适于用来实现本申请实施例的服务器的计算机系统600的结构示意图。图6示出的服务器仅仅是一个示例,不应对本申请实施例的功能和使用范围带

来任何限制。

[0087] 如图6所示,计算机系统600包括中央处理单元(CPU)601,其可以根据存储在只读存储器(ROM)602中的程序或者从存储部分608加载到随机访问存储器(RAM)603中的程序而执行各种适当的动作和处理。在RAM 603中,还存储有系统600操作所需的各种程序和数据。CPU 601、ROM 602以及RAM 603通过总线604彼此相连。输入/输出(I/O)接口605也连接至总线604。

[0088] 以下部件连接至I/O接口605:包括键盘、鼠标等的输入部分606;包括诸如阴极射线管(CRT)、液晶显示器(LCD)等的输出部分607;包括硬盘等的存储部分608;以及包括诸如LAN卡、调制解调器等的网络接口卡的通信部分609。通信部分609经由诸如因特网的网络执行通信处理。驱动器610也根据需要连接至I/O接口605。可拆卸介质611,诸如磁盘、光盘、磁光盘、半导体存储器等等,根据需要安装在驱动器610上,以便于从其上读出的计算机程序根据需要被安装入存储部分608。

[0089] 特别地,根据本公开的实施例,上文参考流程图描述的过程可以被实现为计算机软件程序。例如,本公开的实施例包括一种计算机程序产品,其包括承载在计算机可读介质上的计算机程序,该计算机程序包含用于执行流程图所示的方法的程序代码。在这样的实施例中,该计算机程序可以通过通信部分609从网络上被下载和安装,和/或从可拆卸介质611被安装。在该计算机程序被中央处理单元(CPU)601执行时,执行本申请的方法中限定的上述功能。

[0090] 需要说明的是,本申请的计算机可读介质可以是计算机可读信号介质或者计算机可读存储介质或者是上述两者的任意组合。计算机可读存储介质例如可以是一—但不限于——电、磁、光、电磁、红外线、或半导体的系统、装置或器件,或者任意以上的组合。计算机可读存储介质的更具体的例子可以包括但不限于:具有一个或多个导线的电连接、便携式计算机磁盘、硬盘、随机访问存储器(RAM)、只读存储器(ROM)、可擦式可编程只读存储器(EPROM或闪存)、光纤、便携式紧凑磁盘只读存储器(CD-ROM)、光存储器件、磁存储器件、或者上述的任意合适的组合。在本申请中,计算机可读存储介质可以是任何包含或存储程序的有形介质,该程序可以被指令执行系统、装置或者器件使用或者与其结合使用。而在本申请中,计算机可读的信号介质可以包括在基带中或者作为载波一部分传播的数据信号,其中承载了计算机可读的程序代码。这种传播的数据信号可以采用多种形式,包括但不限于电磁信号、光信号或上述的任意合适的组合。计算机可读的信号介质还可以是计算机可读存储介质以外的任何计算机可读介质,该计算机可读介质可以发送、传播或者传输用于由指令执行系统、装置或者器件使用或者与其结合使用的程序。计算机可读介质上包含的程序代码可以用任何适当的介质传输,包括但不限于:无线、电线、光缆、RF等等,或者上述的任意合适的组合。

[0091] 可以以一种或多种程序设计语言或其组合来编写用于执行本申请的操作的计算机程序代码,所述程序设计语言包括面向目标的设计语言——诸如Java、Smalltalk、C++、还包括常规的过程式程序设计语言——诸如“C”语言或类似的设计语言。程序代码可以完全地在用户计算机上执行、部分地在用户计算机上执行、作为一个独立的软件包执行、部分在用户计算机上部分在远程计算机上执行、或者完全在远程计算机或服务器上执行。在涉及远程计算机的情形中,远程计算机可以通过任意种类的网络——包括局域网(LAN)

或广域网 (WAN) 一连接到用户计算机, 或者, 可以连接到外部计算机 (例如利用因特网服务提供商来通过因特网连接)。

[0092] 附图中的流程图和框图, 图示了按照本申请各种实施例的系统、方法和计算机程序产品的可能实现的体系架构、功能和操作。在这点上, 流程图或框图中的每个方框可以代表一个模块、程序段、或代码的一部分, 该模块、程序段、或代码的一部分包含一个或多个用于实现规定的逻辑功能的可执行指令。也应当注意, 在有些作为替换的实现中, 方框中所标注的功能也可以以不同于附图中所标注的顺序发生。例如, 两个接连地表示的方框实际上可以基本并行地执行, 它们有时也可以按相反的顺序执行, 这依所涉及的功能而定。也要注意的, 框图和/或流程图中的每个方框、以及框图和/或流程图中的方框的组合, 可以用执行规定的功能或操作的专用的基于硬件的系统来实现, 或者可以用专用硬件与计算机指令的组合来实现。

[0093] 描述于本申请实施例中所涉及到的单元可以通过软件的方式实现, 也可以通过硬件的方式来实现。所描述的单元也可以设置在处理器中, 例如, 可以描述为: 一种处理器, 包括获取单元、生成单元、识别单元。其中, 这些单元的名称在某种情况下并不构成对该单元本身的限定, 例如, 获取单元还可以被描述为“获取目标网站的点击信息序列的单元”。

[0094] 作为另一方面, 本申请还提供了一种计算机可读介质, 该计算机可读介质可以是上述实施例中描述的服务器中所包含的; 也可以是单独存在, 而未装配入该服务器中。上述计算机可读介质承载有一个或者多个程序, 当上述一个或者多个程序被该服务器执行时, 使得该服务器: 获取目标网站的点击信息序列; 基于点击信息序列, 生成点击特征序列; 将点击特征序列输入至预先训练的异常检测模型中, 得到类别信息, 其中, 异常检测模型用于表征点击特征序列与类别信息之间的对应关系, 类别信息用于表征点击行为是否异常。

[0095] 以上描述仅为本申请的较佳实施例以及对所运用技术原理的说明。本领域技术人员应当理解, 本申请中所涉及的发明范围, 并不限于上述技术特征的特定组合而成的技术方案, 同时也应涵盖在不脱离上述发明构思的情况下, 由上述技术特征或其等同特征进行任意组合而形成的其它技术方案。例如上述特征与本申请中公开的 (但不限于) 具有类似功能的技术特征进行互相替换而形成的技术方案。

100

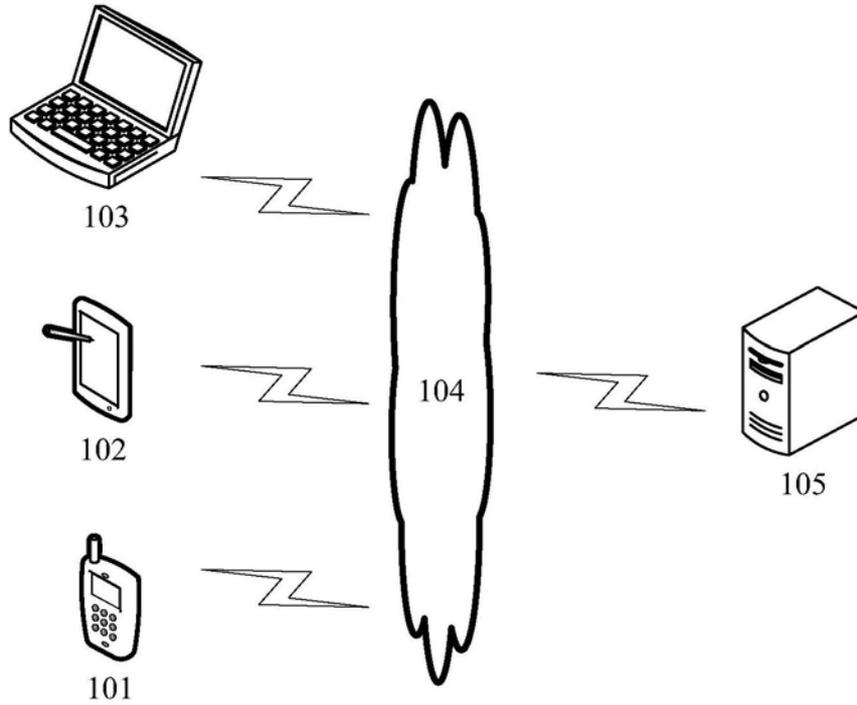


图1

200

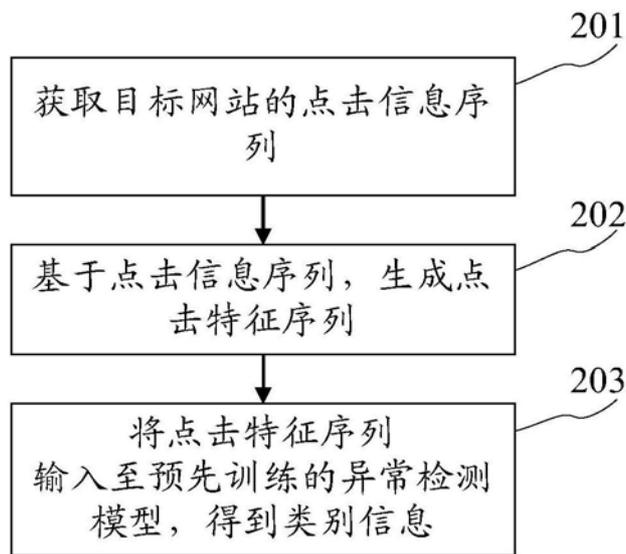


图2

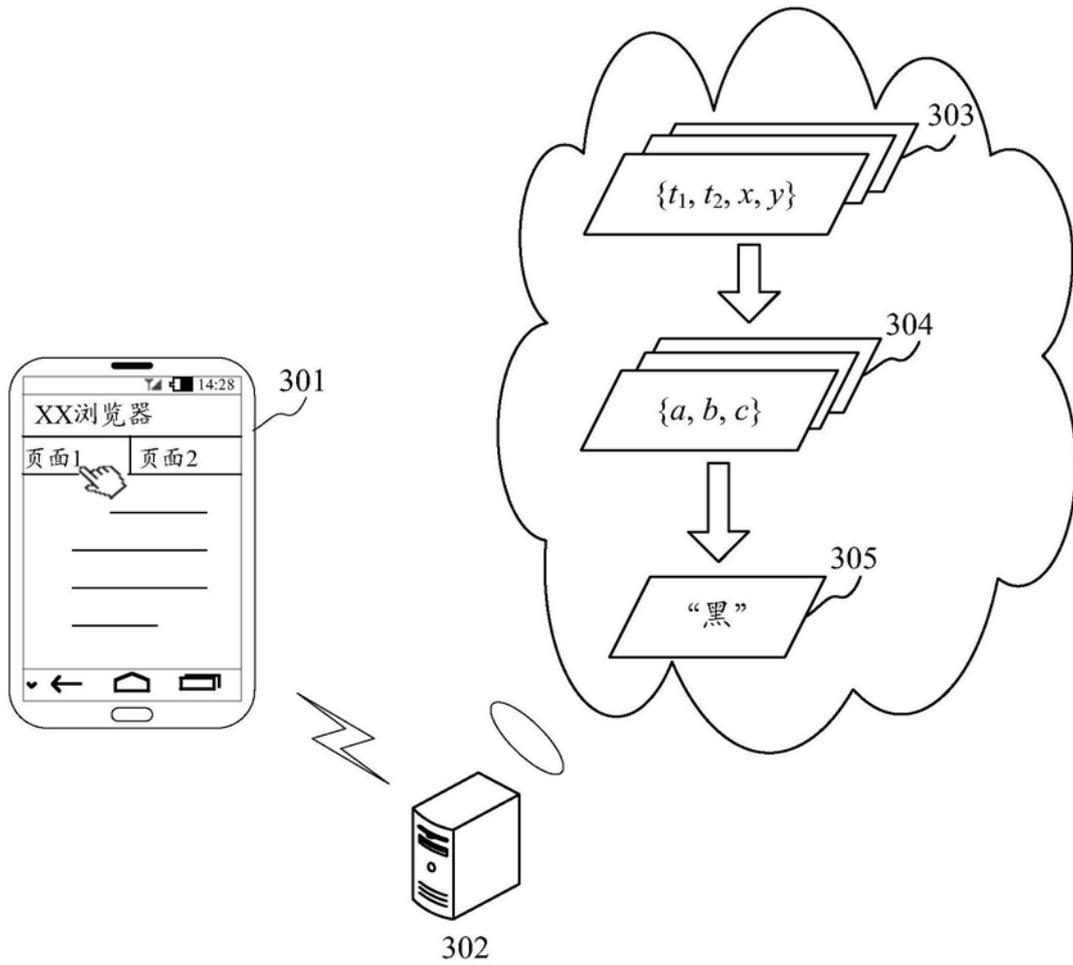


图3

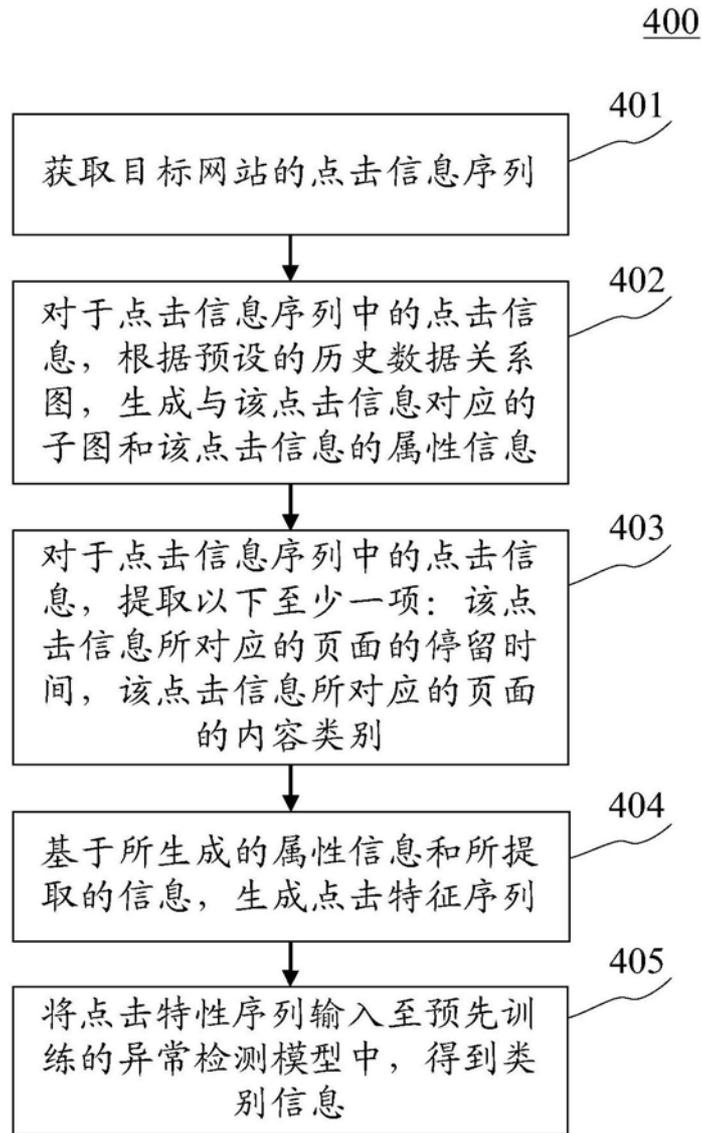


图4

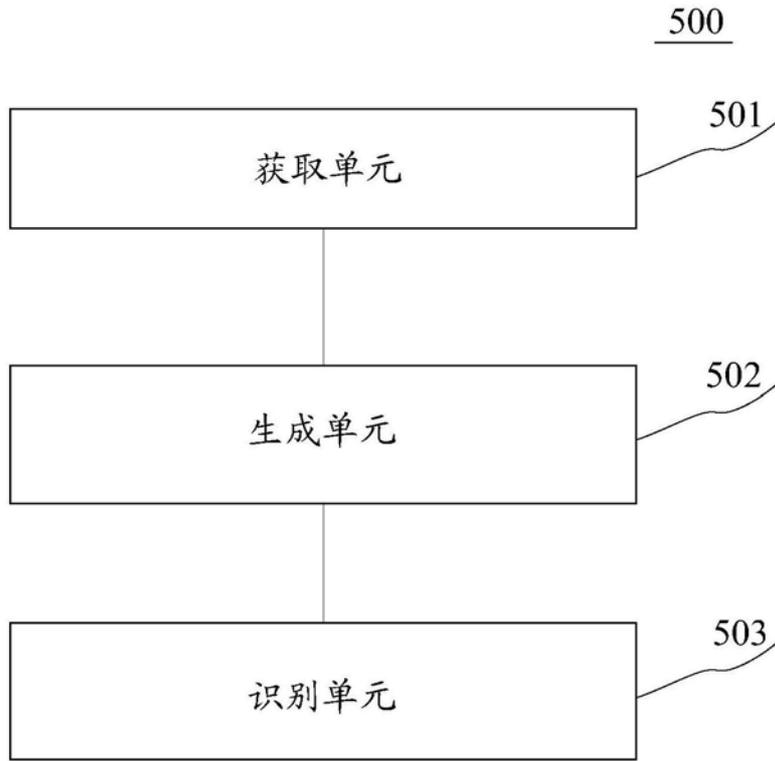


图5

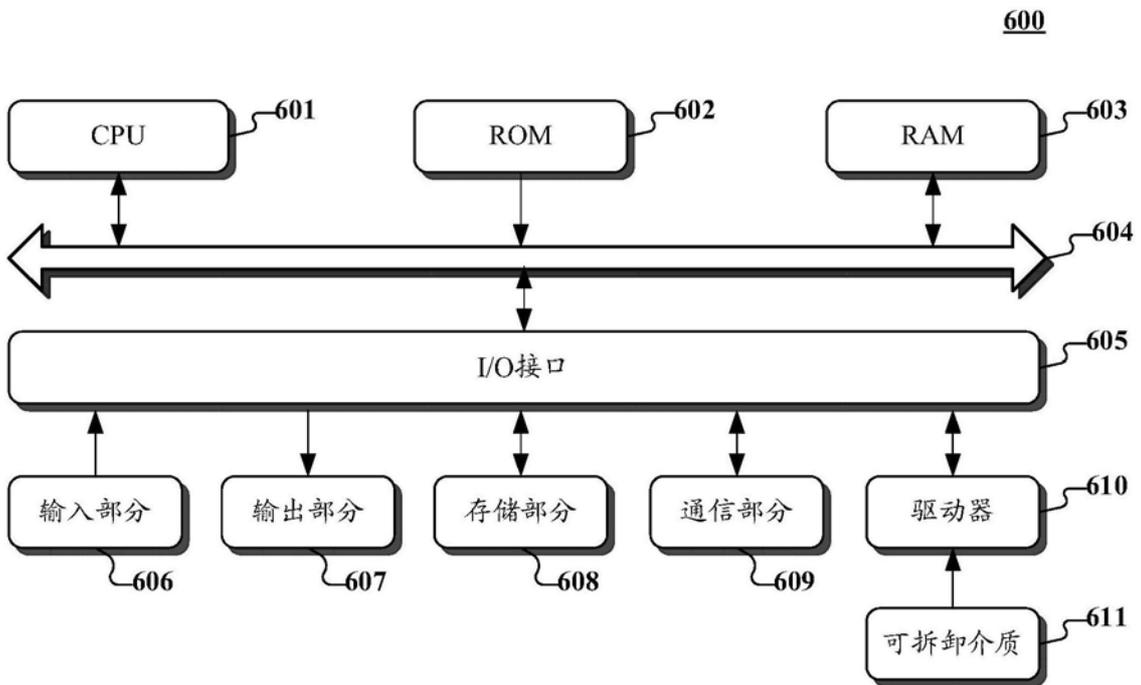


图6