



(12)发明专利申请

(10)申请公布号 CN 110704772 A

(43)申请公布日 2020.01.17

(21)申请号 201810654762.1

(22)申请日 2018.06.22

(71)申请人 北京京东尚科信息技术有限公司  
地址 100195 北京市海淀区杏石口路65号  
西杉创意园四区11号楼东段1-4层西  
段1-4层

申请人 北京京东世纪贸易有限公司

(72)发明人 黄栌桥 吴萌

(74)专利代理机构 北京律智知识产权代理有限  
公司 11438

代理人 阚梓瑄 王卫忠

(51)Int.Cl.

G06F 16/958(2019.01)

G06F 11/30(2006.01)

权利要求书2页 说明书12页 附图8页

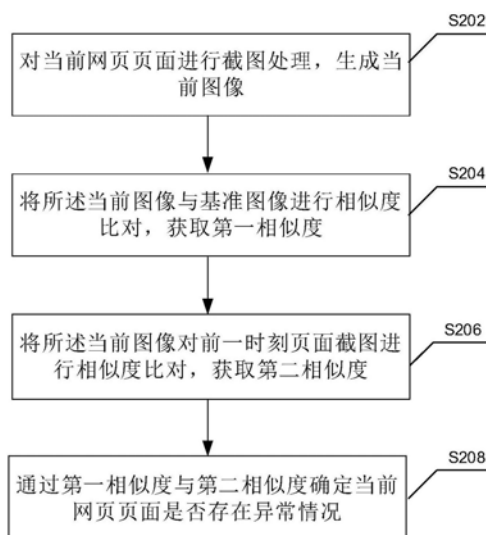
(54)发明名称

页面异常监控方法、系统、装置、电子设备及  
计算机可读介质

(57)摘要

本公开涉及一种页面异常监控方法、系统、  
装置、电子设备及计算机可读介质。涉及计算机  
信息处理领域,该方法包括:对当前网页页面进  
行截图处理,生成当前图像;将所述当前图像与  
基准图像进行相似度比对,获取第一相似度;将  
所述当前图像对前一时刻页面截图进行相似度  
比对,获取第二相似度;以及通过第一相似度与  
第二相似度确定当前网页页面是否存在异常情  
况。本公开涉及的页面异常监控方法、系统、装  
置、电子设备及计算机可读介质,能够自动探测  
网页页面的异常情况,以便及时修复网页缺陷。

20



1. 一种页面异常监控方法,其特征在于,包括:  
对当前网页页面进行截图处理,生成当前图像;  
将所述当前图像与基准图像进行相似度比对,获取第一相似度;  
将所述当前图像对前一时刻页面截图进行相似度比对,获取第二相似度;以及  
通过第一相似度与第二相似度确定当前网页页面是否存在异常情况。
2. 如权利要求1所述的方法,其特征在于,对当前网页页面进行截图处理,生成当前图像包括:  
对当前网页页面进行截图处理,生成第一图像;以及  
将所述第一图像进行灰化处理,生成所述当前图像。
3. 如权利要求1所述的方法,其特征在于,对当前网页页面进行截图处理,生成当前图像还包括:  
以像素为单位对所述当前图像进行切分处理,获取预定个切分图像单元。
4. 如权利要求3所述的方法,其特征在于,将所述当前图像与基准图像进行相似度比对,获取第一相似度包括:  
将所述当前图像中切分图像单元分别与所述基准图像中的切分图像单元进行相似度对比,获取第一相似度。
5. 如权利要求4所述的方法,其特征在于,将每一个切分图像单元分别与所述基准图像中的切分图像单元进行相似度对比,获取第一相似度包括:  
将当前图像中切分图像单元与所述基准图像中的切分图像单元相似单元数记为M;  
将当前图像中切分图像单元与所述基准图像中的切分图像单元不相似单元数记为N;  
以及  
通过相似度公式确定所述当前图像与所述基准图像的所述第一相似度。
6. 如权利要求5所述的方法,其特征在于,相似度公式包括:  
$$S=M/N*100\%;$$
  
其中,S为相似度,M为相似单元数;N为不相似单元数。
7. 如权利要求1所述的方法,其特征在于,通过第一相似度与第二相似度确定当前网页页面是否存在异常情况包括:  
通过第一相似度与第二相似度生成第一相似度曲线与第二相似度曲线;以及  
通过第一相似度曲线与第二相似度曲线确定当前网页页面是否存在异常情况。
8. 如权利要求1所述的方法,其特征在于,通过第一相似度曲线与第二相似度曲线确定当前网页页面是否存在异常情况包括:  
当所述第一相似度曲线与第二相似度曲线满足第一特征时,确定所述当前网页页面为正常状态;以及  
当所述第一相似度曲线与第二相似度曲线满足第二特征时,确定当前网页页面为正常状态。
9. 如权利要求1所述的方法,其特征在于,通过第一相似度曲线与第二相似度曲线确定当前网页页面是否存在异常情况包括:  
当所述第一相似度曲线与第二相似度曲线满足第三特征时,计算获取相似度方差;  
当所述相似度方差小于预定阈值时,确定所述当前网页页面为正常状态;以及

当所述相似度方差大于预定阈值时,确定当前网页页面存在异常情况。

10. 如权利要求1所述的方法,其特征在于,还包括:

对用户行为进行分析,以生成页面异常信息;

根据页面异常信息,对当前网页页面进行截图处理,生成当前图像。

11. 一种页面异常监控系统,其特征在于,包括:

截图服务器,用于对当前网页页面进行截图处理,生成当前图像;

爬虫服务器,用于周期性获取所述当前图像,并将所述当前图像与基准图像进行相似度比对,获取第一相似度;将所述当前图像对前一时刻页面截图进行相似度比如,获取第二相似度;以及

网络服务器,用于通过第一相似度与第二相似度确定当前网页页面是否存在异常情况。

12. 一种页面异常监控装置,其特征在于,包括:

截图模块,用于对当前网页页面进行截图处理,生成当前图像;

第一对比模块,用于将所述当前图像与基准图像进行相似度比对,获取第一相似度;

第二对比模块,用于将所述当前图像对前一时刻页面截图进行相似度比如,获取第二相似度;以及

判断模块,用于通过第一相似度与第二相似度确定当前网页页面是否存在异常情况。

13. 一种电子设备,其特征在于,包括:

一个或多个处理器;

存储装置,用于存储一个或多个程序;

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1-10中任一所述的方法。

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

## 页面异常监控方法、系统、装置、电子设备及计算机可读介质

### 技术领域

[0001] 本公开涉及计算机信息处理领域,具体而言,涉及一种页面异常监控方法、系统、装置、电子设备及计算机可读介质。

### 背景技术

[0002] 在网站服务类应用中,用户是通过不同页面提供的辅助功能来完成与网站服务系统之间的信息交互,进而达到用户访问网站的最终目的。以购物类的网站为例,用户从挑选商品到最终付款,会浏览到大量不同的页面,可例如为:店铺首页、活动页面、商品详情页、购物车页面,以及付款页面等等。

[0003] 在一般情况下,如果某个页面的数据或者渲染发生错误,研发人员在该错误发生的第一时间是不能够得知该错误的,一般情况下,这个错误需要等到业务部门发现之后,再通知到研发部门。业务部门通知到研发部门页面异常的时间与页面异常的发生时间的周期可长可短。如何能够让页面的研发人员及时快速的得知页面的异常情况,从而快速的对页面进行修复,是一个现有技术中的难题。

[0004] 因此,需要一种新的页面异常监控方法、系统、装置、电子设备及计算机可读介质。

[0005] 在所述背景技术部分公开的上述信息仅用于加强对本公开的背景的理解,因此它可以包括不构成对本领域普通技术人员已知的现有技术的信息。

### 发明内容

[0006] 有鉴于此,本公开提供一种页面异常监控方法、系统、装置、电子设备及计算机可读介质,能够自动探测网页页面的异常情况,以便及时修复网页缺陷。

[0007] 本公开的其他特性和优点将通过下面的详细描述变得显然,或部分地通过本公开的实践而习得。

[0008] 根据本公开的一方面,提出一种页面异常监控方法,该方法包括:对当前网页页面进行截图处理,生成当前图像;将所述当前图像与基准图像进行相似度比对,获取第一相似度;将所述当前图像对前一时刻页面截图进行相似度比如,获取第二相似度;以及通过第一相似度与第二相似度确定当前网页页面是否存在异常情况。

[0009] 在本公开的一种示例性实施例中,对当前网页页面进行截图处理,生成当前图像包括:对当前网页页面进行截图处理,生成第一图像;以及将所述第一图像进行灰化处理,生成所述当前图像。

[0010] 在本公开的一种示例性实施例中,对当前网页页面进行截图处理,生成当前图像还包括:以像素为单位对所述当前图像进行切分处理,获取预定个切分图像单元。

[0011] 在本公开的一种示例性实施例中,将所述当前图像与基准图像进行相似度比对,获取第一相似度包括:将所述当前图像中切分图像单元分别与所述基准图像中的切分图像单元进行相似度对比,获取第一相似度。

[0012] 在本公开的一种示例性实施例中,将每一个切分图像单元分别与所述基准图像中

的切分图像单元进行相似度对比,获取第一相似度包括:将当前图像中切分图像单元与所述基准图像中的切分图像单元相似单元数记为M;将当前图像中切分图像单元与所述基准图像中的切分图像单元不相似单元数记为N;以及通过相似度公式确定所述当前图像与所述基准图像的所述第一相似度。

[0013] 在本公开的一种示例性实施例中,相似度公式包括:

[0014]  $S=M/N*100\%$ ;

[0015] 其中,S为相似度,M为相似单元数;N为不相似单元数。

[0016] 在本公开的一种示例性实施例中,通过第一相似度与第二相似度确定当前网页页面是否存在异常情况包括:通过第一相似度与第二相似度生成第一相似度曲线与第二相似度曲线;以及通过第一相似度曲线与第二相似度曲线确定当前网页页面是否存在异常情况。

[0017] 在本公开的一种示例性实施例中,通过第一相似度曲线与第二相似度曲线确定当前网页页面是否存在异常情况包括:当所述第一相似度曲线与第二相似度曲线满足第一特征时,确定所述当前网页页面为正常状态;以及当所述第一相似度曲线与第二相似度曲线满足第二特征时,确定当前网页页面为正常状态。

[0018] 在本公开的一种示例性实施例中,通过第一相似度曲线与第二相似度曲线确定当前网页页面是否存在异常情况包括:当所述第一相似度曲线与第二相似度曲线满足第三特征时,计算获取相似度方差;当所述相似度方差小于预定阈值时,确定所述当前网页页面为正常状态;以及当所述相似度方差大于预定阈值时,确定当前网页页面存在异常情况。

[0019] 在本公开的一种示例性实施例中,对用户行为进行分析,,以生成页面异常信息;根据页面异常信息,对当前网页页面进行截图处理,生成当前图像。

[0020] 根据本公开的一方面,提出一种页面异常监控系统,该系统包括:截图服务器,用于对当前网页页面进行截图处理,生成当前图像;爬虫服务器,用于周期性获取所述当前图像,并将所述当前图像与基准图像进行相似度比对,获取第一相似度;将所述当前图像对前一时刻页面截图进行相似度比如,获取第二相似度;以及网络服务器,用于通过第一相似度与第二相似度确定当前网页页面是否存在异常情况。

[0021] 根据本公开的一方面,提出一种页面异常监控装置,该装置包括:截图模块,用于对当前网页页面进行截图处理,生成当前图像;第一对比模块,用于将所述当前图像与基准图像进行相似度比对,获取第一相似度;第二对比模块,用于将所述当前图像对前一时刻页面截图进行相似度比如,获取第二相似度;以及判断模块,用于通过第一相似度与第二相似度确定当前网页页面是否存在异常情况。

[0022] 根据本公开的一方面,提出一种电子设备,该电子设备包括:一个或多个处理器;存储装置,用于存储一个或多个程序;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上文的方法。

[0023] 根据本公开的一方面,提出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上文中的方法。

[0024] 根据本公开的页面异常监控方法、系统、装置、电子设备及计算机可读介质,能够自动探测网页页面的异常情况,以便及时修复网页缺陷。

[0025] 应当理解的是,以上的一般描述和后文的细节描述仅是示例性的,并不能限制本

公开。

### 附图说明

[0026] 通过参照附图详细描述其示例实施例,本公开的上述和其它目标、特征及优点将变得更加显而易见。下面描述的附图仅仅是本公开的一些实施例,对于本领域的普通技术人员来讲,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。

[0027] 图1是根据一示例性实施例示出的一种页面异常监控方法及装置的系统框图。

[0028] 图2是根据一示例性实施例示出的一种页面异常监控方法的流程图。

[0029] 图3是根据另一示例性实施例示出的一种页面异常监控方法的流程图。

[0030] 图4是根据另一示例性实施例示出的一种页面异常监控方法中的统计曲线示意图。

[0031] 图5是根据另一示例性实施例示出的一种页面异常监控方法中的统计曲线示意图。

[0032] 图6是根据另一示例性实施例示出的一种页面异常监控方法中的统计曲线示意图。

[0033] 图7是根据另一示例性实施例示出的一种页面异常监控方法中的统计曲线示意图。

[0034] 图8是根据一示例性实施例示出的一种页面异常监控系统的示意图。

[0035] 图9是根据一示例性实施例示出的一种页面异常监控系统的框图。

[0036] 图10是根据另一示例性实施例示出的一种页面异常监控装置的框图。

[0037] 图11是根据一示例性实施例示出的一种电子设备的框图。

[0038] 图12是根据一示例性实施例示出一种计算机可读存储介质示意图。

### 具体实施方式

[0039] 现在将参考附图更全面地描述示例实施例。然而,示例实施例能够以多种形式实施,且不应被理解为限于在此阐述的实施例;相反,提供这些实施例使得本公开将全面和完整,并将示例实施例的构思全面地传达给本领域的技术人员。在图中相同的附图标记表示相同或类似的部分,因而将省略对它们的重复描述。

[0040] 此外,所描述的特征、结构或特性可以以任何合适的方式结合在一个或更多实施例中。在下面的描述中,提供许多具体细节从而给出对本公开的实施例的充分理解。然而,本领域技术人员将意识到,可以实践本公开的技术方案而没有特定细节中的一个或更多,或者可以采用其它的方法、组元、装置、步骤等。在其它情况下,不详细示出或描述公知方法、装置、实现或者操作以避免模糊本公开的各方面。

[0041] 附图中所示的方框图仅仅是功能实体,不一定必须与物理上独立的实体相对应。即,可以采用软件形式来实现这些功能实体,或在一个或多个硬件模块或集成电路中实现这些功能实体,或在不同网络和/或处理器装置和/或微控制器装置中实现这些功能实体。

[0042] 附图中所示的流程图仅是示例性说明,不是必须包括所有的内容和操作/步骤,也不是必须按所描述的顺序执行。例如,有的操作/步骤还可以分解,而有的操作/步骤可以合并或部分合并,因此实际执行的顺序有可能根据实际情况改变。

[0043] 应理解,虽然本文中可能使用术语第一、第二、第三等来描述各种组件,但这些组件不应受这些术语限制。这些术语乃用以区分一组件与另一组件。因此,下文论述的第一组件可称为第二组件而不偏离本公开概念的教导。如本文中所使用,术语“及/或”包括相关联的列出项目中的任一个及一或多者的所有组合。

[0044] 本领域技术人员可以理解,附图只是示例实施例的示意图,附图中的模块或流程并不一定是实施本公开所必须的,因此不能用于限制本公开的保护范围。

[0045] 图1是根据一示例性实施例示出的一种页面异常监控方法及装置的系统框图。

[0046] 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。

[0047] 用户可以使用终端设备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,例如购物类应用、网页浏览器应用、搜索类应用、即时通信工具、邮箱客户端、社交平台软件等。

[0048] 终端设备101、102、103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、膝上型便携计算机和台式计算机等等。

[0049] 服务器105可以是提供各种服务的服务器,例如对用户利用终端设备101、102、103所浏览的购物类网站提供页面支持的后台管理服务器。后台管理服务器可以对接收到的产品信息查询请求等数据进行分析等处理,并将处理结果反馈给终端设备。

[0050] 服务器105可例如对当前网页页面进行截图处理,生成当前图像;服务器105可例如将所述当前图像与基准图像进行相似度比对,获取第一相似度;服务器105可例如将所述当前图像对前一时刻页面截图进行相似度比如,获取第二相似度;服务器105可例如通过第一相似度与第二相似度确定当前网页页面是否存在异常情况。

[0051] 服务器105可以是一个实体的服务器,还可例如为多个服务器组成,服务器105可例如由截图服务器,爬虫服务器,以及网络服务器组成。

[0052] 其中截图服务器,可用于对当前网页页面进行截图处理,生成当前图像;爬虫服务器,可用于周期性获取所述当前图像,并将所述当前图像与基准图像进行相似度比对,获取第一相似度;将所述当前图像对前一时刻页面截图进行相似度比如,获取第二相似度;以及网络服务器,可用于通过第一相似度与第二相似度确定当前网页页面是否存在异常情况。

[0053] 需要说明的是,本公开实施例所提供的页面异常监控方法可以由服务器105执行,相应地,页面异常监控装置可以设置于服务器105中。而提供给用户进行商品浏览的网页端与进行页面截图的监控端一般位于终端设备101、102、103中。

[0054] 图2是根据一示例性实施例示出的一种页面异常监控方法的流程图。页面异常监控方法20至少包括步骤S202至S208。

[0055] 如图2所示,在S202中,对当前网页页面进行截图处理,生成当前图像。可例如,对当前网页页面进行截图处理,生成第一图像;以及将所述第一图像进行灰化处理,生成所述当前图像。

[0056] 其中,灰化是指将图片中一个像素的R、G、B三个通道的值转化为一个亮度值来表示。由于很容易就能做到对Web页面每次截图的坐标和小大完全一致,所以计算两张截图的相似度时不用考虑位移。而且Web页面截图与真实的照片不同,不受光照条件的影响,所以

在对图片进行灰化后,单独利用截图中像素点的灰度值来计算相似度,就可以满足本发明描述的场景。

[0057] 在一个实施例中,对当前网页页面进行截图处理,生成当前图像还包括:以像素为单位对所述当前图像进行切分处理,获取预定个切分图像单元。对当前页面进行截图,并保存为A1。先对A1进行灰化处理,然后按照横向X个像素和纵向Y个像素为单位,把图片进行切分。可例如一张分辨率为PX\*PY的图片,按10(px)\*10(px)为最小单元进行切分,共切分出N个单元,每个单元中有100个像素。

[0058] 在S204中,将所述当前图像与基准图像进行相似度比对,获取第一相似度。可例如,将所述当前图像中切分图像单元分别与所述基准图像中的切分图像单元进行相似度对比,获取第一相似度。

[0059] 在一个实施例中,可例如对同一页面在不同时间进行两次截图,并保存为A1和A2。其中,A1代表当前图像,A2代表基准图像。先对A1和A2进行灰化处理,然后按照横向X个像素和纵向Y个像素为单位,分别把图片进行切分,然后依次对切分后的单元进行对比。一张分辨率为PX\*PY的图片,按10(px)\*10(px)为最小单元进行切分,共切分出N个单元,每个单元中有100个像素。

[0060] 其中基准图像可例如为任一时刻截取到的当前网页图像,还可例如为每天固定时间周期截图到的当前网页的图像,本申请不以此为限。

[0061] 在一个实施例中,将每一个切分图像单元分别与所述基准图像中的切分图像单元进行相似度对比,获取第一相似度包括:将当前图像中切分图像单元与所述基准图像中的切分图像单元相似单元数记为M;将当前图像中切分图像单元与所述基准图像中的切分图像单元不相似单元数记为N;以及通过相似度公式确定所述当前图像与所述基准图像的所述第一相似度。

[0062] 可例如,如果A1图像中的某一个单元U1和A2图像中对应的单元U2中灰度值不相同的像素超过一个预设阈值,则标记为U1和U2不相似,所有相似的单元数量为M,不相似单元数记为N,则A1与A2的页面相似度为:

[0063]  $S=M/N*100\%$ 。

[0064] 在S206中,将所述当前图像对前一时刻页面截图进行相似度比较,获取第二相似度。可例如,将所述当前图像中切分图像单元分别与所述前一时刻页面截图中的切分图像单元进行相似度对比,获取第二相似度。

[0065] 在一个实施例中,可例如对同一页面在连续两个时间周期分别进行截图,并保存为A1和A3。其中,A1代表当前图像,A3代表前一时刻页面截图。先对A1和A3进行灰化处理,然后按照横向X个像素和纵向Y个像素为单位,分别把图片进行切分,然后依次对切分后的单元进行对比。一张分辨率为PX\*PY的图片,按10(px)\*10(px)为最小单元进行切分,共切分出N个单元,每个单元中有100个像素。

[0066] 可例如,如果A1图像中的某一个单元U1和A3图像中对应的单元U3中灰度值不相同的像素超过一个预设阈值,则标记为U1和U3不相似,所有相似的单元数量为M,不相似单元数记为N,则A1与A3的页面相似度为:

[0067]  $S=M/N*100\%$ 。

[0068] 在S208中,通过第一相似度与第二相似度确定当前网页页面是否存在异常情况。



可例如,通过第一相似度与第二相似度生成第一相似度曲线与第二相似度曲线;以及通过第一相似度曲线与第二相似度曲线确定当前网页页面是否存在异常情况。

[0069] 在一个实施例中,当所述第一相似度曲线与第二相似度曲线满足第一特征时,确定所述当前网页页面为正常状态。

[0070] 在一个实施例中,当所述第一相似度曲线与第二相似度曲线满足第二特征时,确定当前网页页面为正常状态。

[0071] 在一个实施例中,当所述第一相似度曲线与第二相似度曲线满足第三特征时,计算获取相似度方差;当所述相似度方差小于预定阈值时,确定所述当前网页页面为正常状态。

[0072] 在一个实施例中,当所述第一相似度曲线与第二相似度曲线满足第三特征时,计算获取相似度方差;当所述相似度方差大于预定阈值时,确定当前网页页面存在异常情况。

[0073] 根据本公开的页面异常监控方法,通过对比不同时刻的页面截图的相似度数据,进而判断当前页面是否存在异常情况的方式,能够自动探测网页页面的异常情况,以便及时修复网页缺陷。

[0074] 应清楚地理解,本公开描述了如何形成和使用特定示例,但本公开的原理不限于这些示例的任何细节。相反,基于本公开公开的内容的教导,这些原理能够应用于许多其它实施例。

[0075] 基于上文介绍的相似度计算公式,对一个固定页面建立相似度统计数据的思路如下:截取被跟踪页面(记为P)的一张截图作为基线图片(B0),之后按固定频率对P进行截图,计算所截图片跟B0的相似度(记为 $S_{current}$ ),记录所有 $S_{current}$ 的值,并根据所记录的 $S_{current}$ 计算页面P的波动范围(如图3所示),如果下次截图的 $S_{current}$ 超过此范围,则可初步认为认为该页面P出现异常。

[0076] 但是,受限于现有技术和页面包含动态内容原因,除非完全静止的页面(即页面没有动态变化的部分),否则相似度很难达到100%准确。所以页面相似度一般会有个波动,本实施例中,通过对页面进行灰化将这种波动降到最低。但是,真实环境中页面的相似度曲线可能呈现更为复杂的情况,这里对不同情况进行分类并分别加以阐述。

[0077] 做如下的参数定义:当前页面截图与基准图片的相似度记为 $S_1$ ;当前页面截图与前一次截图的相似度记为 $S_2$ 。

[0078] 在一个实施例中,页面基本不随时间变化(静态页面,或者页面上只有很小一部分会动态变化)。例如“用户个人信息”页面。除了用户余额、消息条数、购物车数量会动态变化以外,其余部分基本不随时间变化。此类页面的相似度统计曲线如图4所示:由于变化的部分并不随时间呈现周期性的变化,例如用户余额,还和用户的购买行为有关。可以利用技术手段在截图的时候把这部分区域排除掉,让相似度曲线 $S_1$ 和 $S_2$ 简化为类直线。

[0079] 在一个实施例中,页面固定部分有周期性的变化(搜索结果页、带BI功能的页面。)

[0080] 例如在购物网站首页输入“苹果”搜索得到的结果页,页面一侧的商品推荐可能每次的内容都不一样,但搜索结果的具体内容在一段时间内都是固定的。此类页面的 $S_1$ 和 $S_2$ 相似度曲线如图5所示:图5中 $S_1$ 的相似度曲线由于搜索后台物品的相关性数值发生变化,导致搜索结果出现改变,相似度从80%左右降到30%左右,然而由于搜索结果的变化是固定在一个平面范围的,所以在第一次变化后,搜索结果即使再次发生变化,相似度也将稳定

在30%左右。如果这段时间页面发生渲染异常,并且异常的内容就位于搜索结果变化的平面范围中,通过S1相似度统计是感知不到的。

[0081] 以引入了S2相似度曲线,从图5可以看到,S2曲线的相似度变化只在搜索结果变化后的第一次截图计算中出现,因为S2是当前截图与前次截图的对比,所以后续又会恢复到之前的相似度水平。此类页面可以基于S2曲线,将相似度波动范围固定在[30,31]和[80,81]这两个区间,甚至剔除掉变化时的相似度值,只保留[80,81]这个区间。

[0082] 上文提到的两类页面都可以通过建立相似度统计曲线来探测页面的异常情况,并借助技术手段剔除干扰元素,达到简化曲线的目的。但这种方法只适用于静态页面或者变化比较规律的页面,下面介绍相似度曲线不符合以上两种实施例中的相似度曲线时的判断场景。

[0083] 在一个实施例中,页面非固定部分非周期性变化(商家首页,频道首页)。例如“XX购物”首页,大部分内容都是动态内容,而且页面不同部分的变化频率也不同(轮播大图相比底部抢购的变化频率低),这就导致整体页面每次截图的相似度变化幅度不一样。并且动态部分的变化频率不可预期,这也导致整体页面的相似度变化频率无规律可循(如图6所示)。总的来说,此类页面无法通过S1和S2曲线建立统计规律和探测。

[0084] 虽然无法通过相似度曲线来探测页面异常,但相似度的波动幅度依然是具有参考价值,因为一般来说变动越频繁的页面,出错的概率越大,此类页面就需要重点关注。本申请中引入相似度方差的概念来描述这种变动,即:

[0085] 相似度方差  $D(S) = \frac{\sum_{k=1}^n (S_k - S_{ave})^2}{n}$  ;

[0086] 其中,D(S)为方差值, $S_k$ 为当前相似度, $S_{ave}$ 为平均相似度,n为截图数量。

[0087] 结合上文提到的三种类型的页面相似度曲线,如果计算D(S2),类型三的面明显会比类型一、二的方差大。这样能够通过方差来量化页面的变化幅度。当方差数大于预定阈值时,确定当前网页页面出现异常。

[0088] 借助于上文中对不同情况的描述,下面详细介绍页面异常判断的方法。

[0089] 图7是根据另一示例性实施例示出的一种页面异常监控方法的流程图。页面异常监控方法70是对页面异常监控方法20中,步骤S208“通过第一相似度与第二相似度确定当前网页页面是否存在异常情况”的具体描述。

[0090] 如图7所示,在S702中,获取第一相似度与第二相似度。

[0091] 在S704中,通过第一相似度与第二相似度生成第一相似度曲线与第二相似度曲线。

[0092] 在S706中,当所述第一相似度曲线与第二相似度曲线满足第一特征时,确定所述当前网页页面为正常状态。第一特征可例如为如4所述的曲线特征。

[0093] 在S708中,当所述第一相似度曲线与第二相似度曲线满足第二特征时,确定当前网页页面为正常状态。第二特征可例如为如5所述的曲线特征。

[0094] 在S710中,当所述第一相似度曲线与第二相似度曲线满足第三特征时,计算获取相似度方差。第三特征可例如为如6所述的曲线特征。

[0095] 在S712中,当所述相似度方差小于预定阈值时,确定所述当前网页页面为正常状态。

- [0096] 在S714中,当所述相似度方差大于预定阈值时,确定当前网页页面存在异常情况。
- [0097] 根据本公开的页面异常监控方法,不仅能够对于静态页面和页面变化较规律的页面进行异常情况监控,还能够通过方差判断的方式,判断含有较多动态内容的网页页面。即通过相似度统计主动探测页面异常,能够极大缩短页面异常从发生到被发现的时间。
- [0098] 图8是根据一示例性实施例示出的一种页面异常监控系统的示意图。在一个实施例中,还可例如通过用户行为分析,生成页面异常信息;根据页面异常信息,对当前网页页面进行截图处理,生成当前图像。还可例如,在根据相似度曲线判断当前网页异常后,通过监控用户的指定操作,来进一步确认网页的异常情况。本申请不以此为限。
- [0099] 其中,用户行为分析是指通过在被监控页面注入脚本,当用户浏览该页面并完成某些特定操作时(比如点击、刷新),触发特定请求发送到服务器,再由服务器针对不同请求完成对应的数据分析。
- [0100] 如图8所示,异常监控系统800可例如由相似度分析系统802和用户行为分析系统804构成。
- [0101] 如图8所示,其中,相似度分析系统802包括:截图服务器8022,爬虫服务器8024,以及网络服务器8026。其中,对当前网页页面进行截图处理,生成当前图像。
- [0102] 爬虫服务器8022用于周期性获取所述当前图像,并将所述当前图像与基准图像进行相似度比对,获取第一相似度;将所述当前图像对前一时刻页面截图进行相似度比如,获取第二相似度。
- [0103] 网络服务器8024用于通过第一相似度与第二相似度确定当前网页页面是否存在异常情况。
- [0104] 数据库8026用户储存相似度数据。
- [0105] 如图8所示,其中,相似度分析系统804包括:代理服务器8041,负载均衡服务器8043,分析服务器8045,数据结构服务器8047,以及分布式数据库8049。
- [0106] 其中,代理服务器8041用于监控用户指定操作,生成第一请求,所述指定操作包括页面刷新操作;并用于对所述第一请求进行过滤处理,以剔除无效请求;
- [0107] 负载均衡服务器8043用于解析所述第一请求获取用户信息及访问信息;负载均衡服务器8043中包括多个逻辑集合单元60431,用于第一请求的处理和计算。
- [0108] 分析服务器8045用于根据所述用户信息与访问信息确定所述用户本次指定操作与上次指定操作的时间间隔;以及当所述时间间隔小于预定阈值时,生成页面异常信息。
- [0109] 数据结构服务器8047用于储存所述用户信息及访问信息;
- [0110] 分布式数据库8049用于储存页面异常信息。
- [0111] 还可例如包括:消息队列服务器8040,用于为存储过滤后的有效请求的中间服务器。
- [0112] 相似度分析系统802中,爬虫服务器8022(Crawler Server)负责周期性爬取被监控页面的截图,计算并生成相似度值。网络服务器8024(Web Server)负责相似度统计的可视化、相似度数据的管理、爬取策略的管理,并提供相似度数据接口供分析服务器8045(Analyze worker)调用,以便生成异常监控数据。
- [0113] 用户分析系统804中分析服务器8045(Analyze worker)的循环周期可例如为1分钟,所以当页面异常发生时,可以借由用户刷新行为及时的探测问题页面。当然,如果页面

发生异常,但没有用户访问该页面,那自然不会触发用户行为分析,这种情况下对页面异常的探测时间就会延长为相似度统计的周期时间。如果相似度统计频率为一天,那该页面异常得等到下次相似度统计时才能被发现。实际上这种情况很少发生,因为用户访问频率如此低的页面也不具备监控价值。

[0114] 本发明页面异常监控方法,通过将页面异常这种确定性问题转化为概率问题,通过相似度统计和用户行为分析两个方面来量化异常的概率,为最终的页面异常判断提供数据。

[0115] 本发明页面异常监控方法,通过概率来求解一个确定性问题:即通过相似度统计和用户行为分析来主动探测页面异常,并综合两方面的数据为最终判断页面异常提供依据。极大缩短页面异常从发生到被发现的时间。

[0116] 本领域技术人员可以理解实现上述实施例的全部或部分步骤被实现为由CPU执行的计算机程序。在该计算机程序被CPU执行时,执行本公开提供的上述方法所限定的上述功能。所述的程序可以存储于一种计算机可读存储介质中,该存储介质可以是只读存储器,磁盘或光盘等。

[0117] 此外,需要注意的是,上述附图仅是根据本公开示例性实施例的方法所包括的处理的示意性说明,而不是限制目的。易于理解,上述附图所示的处理并不表明或限制这些处理的时间顺序。另外,也易于理解,这些处理可以是例如在多个模块中同步或异步执行的。

[0118] 下述为本公开装置实施例,可以用于执行本公开方法实施例。对于本公开装置实施例中未披露的细节,请参照本公开方法实施例。

[0119] 图9是根据一示例性实施例示出的一种页面异常监控系统的框图。页面异常监控系统90包括:截图服务器902,爬虫服务器904,以及网络服务器906。

[0120] 截图服务器902用于对当前网页页面进行截图处理,生成当前图像。

[0121] 爬虫服务器904用于周期性获取所述当前图像,并将所述当前图像与基准图像进行相似度比对,获取第一相似度;将所述当前图像对前一时刻页面截图进行相似度比如,获取第二相似度。

[0122] 网络服务器906用于通过第一相似度与第二相似度确定当前网页页面是否存在异常情况。

[0123] 根据本公开的页面异常监控系统,通过对比不同时刻的页面截图的相似度数据,进而判断当前页面是否存在异常情况的方式,能够自动探测网页页面的异常情况,以便及时修复网页缺陷。

[0124] 图10是根据另一示例性实施例示出的一种页面异常监控装置的框图。页面异常监控装置1000包括:截图模块1002,第一对比模块1004,第二对比模块1006,判断模块1008。

[0125] 截图模块1002用于对当前网页页面进行截图处理,生成当前图像;

[0126] 第一对比模块1004用于将所述当前图像与基准图像进行相似度比对,获取第一相似度;

[0127] 第二对比模块1006用于将所述当前图像对前一时刻页面截图进行相似度比如,获取第二相似度;

[0128] 判断模块1008用于通过第一相似度与第二相似度确定当前网页页面是否存在异常情况。

[0129] 根据本公开的页面异常监控装置,通过对比不同时刻的页面截图的相似度数据,进而判断当前页面是否存在异常情况的方式,能够自动探测网页页面的异常情况,以便及时修复网页缺陷。

[0130] 图11是根据一示例性实施例示出的一种电子设备的框图。

[0131] 下面参照图11来描述根据本公开的这种实施方式的电子设备200。图11显示的电子设备200仅仅是一个示例,不应对本公开实施例的功能和使用范围带来任何限制。

[0132] 如图11所示,电子设备200以通用计算设备的形式表现。电子设备200的组件可以包括但不限于:至少一个处理单元210、至少一个存储单元220、连接不同系统组件(包括存储单元220和处理单元210)的总线230、显示单元240等。

[0133] 其中,所述存储单元存储有程序代码,所述程序代码可以被所述处理单元210执行,使得所述处理单元210执行本说明书上述电子处方流转处理方法部分中描述的根据本公开各种示例性实施方式的步骤。例如,所述处理单元210可以执行如图2,图7中所示的步骤。

[0134] 所述存储单元220可以包括易失性存储单元形式的可读介质,例如随机存取存储单元(RAM)2201和/或高速缓存存储单元2202,还可以进一步包括只读存储单元(ROM)2203。

[0135] 所述存储单元220还可以包括具有一组(至少一个)程序模块2205的程序/实用工具2204,这样的程序模块2205包括但不限于:操作系统、一个或者多个应用程序、其它程序模块以及程序数据,这些示例中的每一个或某种组合中可能包括网络环境的实现。

[0136] 总线230可以为表示几类总线结构中的一种或多种,包括存储单元总线或者存储单元控制器、外围总线、图形加速端口、处理单元或者使用多种总线结构中的任意总线结构的局域总线。

[0137] 电子设备200也可以与一个或多个外部设备300(例如键盘、指向设备、蓝牙设备等)通信,还可与一个或者多个使得用户能与该电子设备200交互的设备通信,和/或与使得该电子设备200能与一个或多个其它计算设备进行通信的任何设备(例如路由器、调制解调器等等)通信。这种通信可以通过输入/输出(I/O)接口250进行。并且,电子设备200还可以通过网络适配器260与一个或者多个网络(例如局域网(LAN),广域网(WAN)和/或公共网络,例如因特网)通信。网络适配器260可以通过总线230与电子设备200的其它模块通信。应当明白,尽管图中未示出,可以结合电子设备200使用其它硬件和/或软件模块,包括但不限于:微代码、设备驱动器、冗余处理单元、外部磁盘驱动阵列、RAID系统、磁带驱动器以及数据备份存储系统等。

[0138] 通过以上的实施方式的描述,本领域的技术人员易于理解,这里描述的示例实施方式可以通过软件实现,也可以通过软件结合必要的硬件的方式来实现。因此,根据本公开实施方式的技术方案可以以软件产品的形式体现出来,该软件产品可以存储在一个非易失性存储介质(可以是CD-ROM,U盘,移动硬盘等)中或网络上,包括若干指令以使得一台计算设备(可以是个人计算机、服务器、或者网络设备等)执行根据本公开实施方式的上述方法。

[0139] 图12示意性示出本公开示例性实施例中一种计算机可读存储介质示意图。

[0140] 参考图12所示,描述了根据本公开的实施方式的用于实现上述方法的程序产品400,其可以采用便携式紧凑盘只读存储器(CD-ROM)并包括程序代码,并可以在终端设备,例如个人电脑上运行。然而,本公开的程序产品不限于此,在本文件中,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。

[0141] 所述程序产品可以采用一个或多个可读介质的任意组合。可读介质可以是可读信号介质或者可读存储介质。可读存储介质例如可以为但不限于电、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。可读存储介质的更具体的例子(非穷举的列表)包括:具有一个或多个导线的电连接、便携式盘、硬盘、随机存取存储器(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。

[0142] 所述计算机可读存储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,其中承载了可读程序代码。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。可读存储介质还可以是可读存储介质以外的任何可读介质,该可读介质可以发送、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。可读存储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、有线、光缆、RF等等,或者上述的任意合适的组合。

[0143] 可以以一种或多种程序设计语言的任意组合来编写用于执行本公开操作的程序代码,所述程序设计语言包括面向对象的程序设计语言—诸如Java、C++等,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设备上执行、部分地在用户设备上执行、作为一个独立的软件包执行、部分在用户计算设备上部分在远程计算设备上执行、或者完全在远程计算设备或服务器上执行。在涉及远程计算设备的情形中,远程计算设备可以通过任意种类的网络,包括局域网(LAN)或广域网(WAN),连接到用户计算设备,或者,可以连接到外部计算设备(例如利用因特网服务提供商来通过因特网连接)。

[0144] 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一个该设备执行时,使得该计算机可读介质实现如下功能:对当前网页页面进行截图处理,生成当前图像;将所述当前图像与基准图像进行相似度比对,获取第一相似度;将所述当前图像对前一时刻页面截图进行相似度比如,获取第二相似度;以及通过第一相似度与第二相似度确定当前网页页面是否存在异常情况。

[0145] 本领域技术人员可以理解上述各模块可以按照实施例的描述分布于装置中,也可以进行相应变化唯一不同于本实施例的一个或多个装置中。上述实施例的模块可以合并为一个模块,也可以进一步拆分成多个子模块。

[0146] 通过以上的实施例的描述,本领域的技术人员易于理解,这里描述的示例实施例可以通过软件实现,也可以通过软件结合必要的硬件的方式来实现。因此,根据本公开实施例的技术方案可以以软件产品的形式体现出来,该软件产品可以存储在一个非易失性存储介质(可以是CD-ROM,U盘,移动硬盘等)中或网络上,包括若干指令以使得一台计算设备(可以是个人计算机、服务器、移动终端、或者网络设备等)执行根据本公开实施例的方法。

[0147] 以上具体地示出和描述了本公开的示例性实施例。应可理解的是,本公开不限于这里描述的详细结构、设置方式或实现方法;相反,本公开意图涵盖包含在所附权利要求的精神和范围内的各种修改和等效设置。

[0148] 此外,本说明书说明书附图所示出的结构、比例、大小等,均仅用以配合说明书所

公开的内容,以供本领域技术人员了解与阅读,并非用以限定本公开可实施的限定条件,故不具技术上的实质意义,任何结构的修饰、比例关系的改变或大小的调整,在不影响本公开所能产生的技术效果及所能实现的目的下,均应仍落在本公开所公开的技术内容得能涵盖的范围内。同时,本说明书中所引用的如“上”、“第一”、“第二”及“一”等的用语,也仅为便于叙述的明了,而非用以限定本公开可实施的范围,其相对关系的改变或调整,在无实质变更技术内容下,当也视为本公开可实施的范畴。

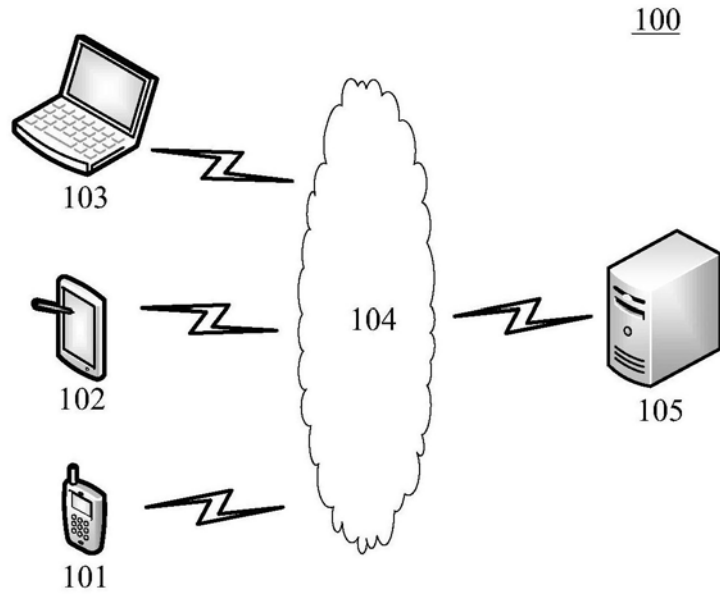


图1

20

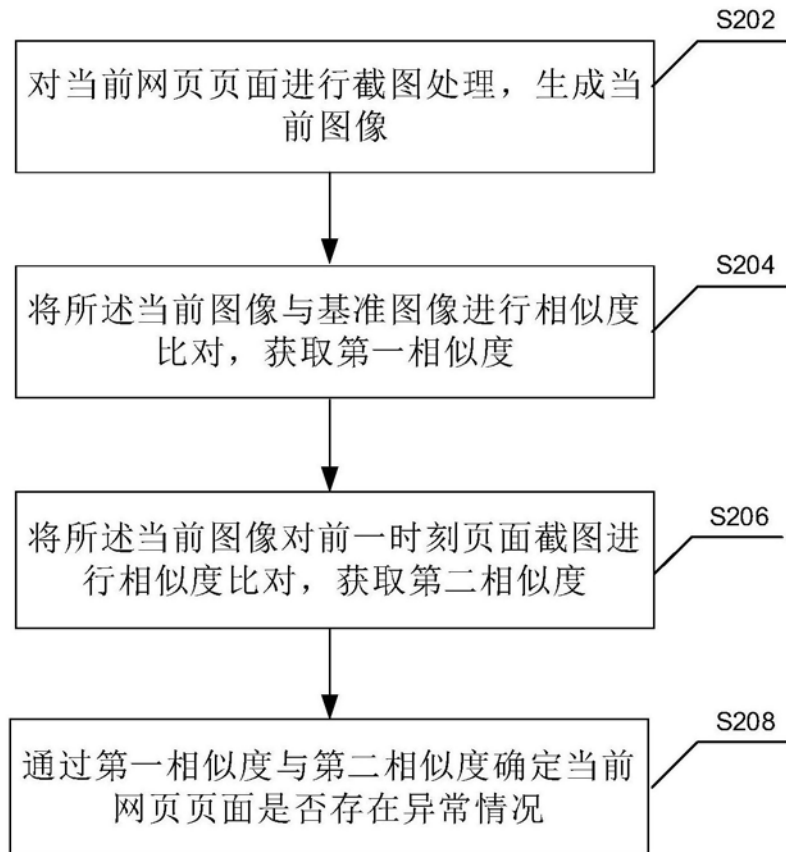


图2



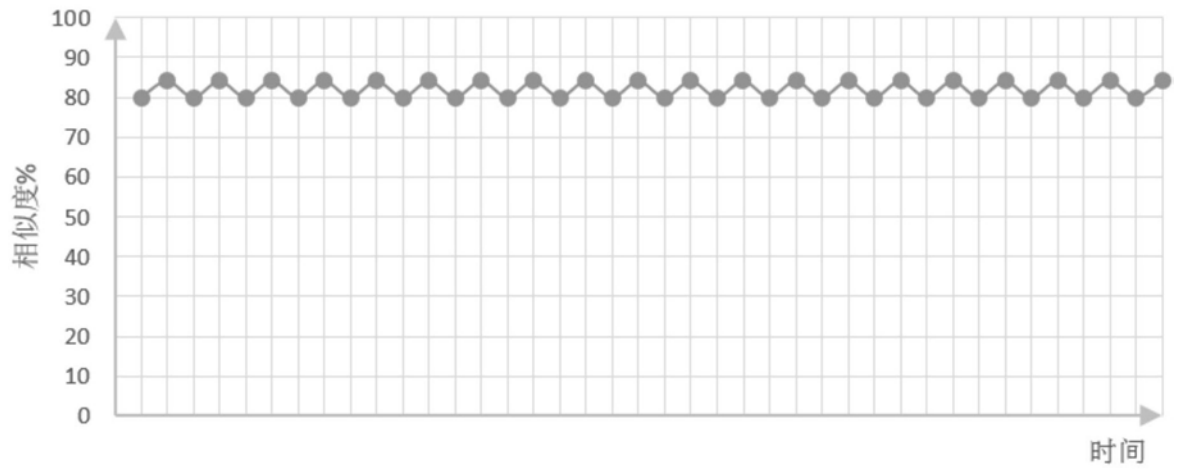


图3

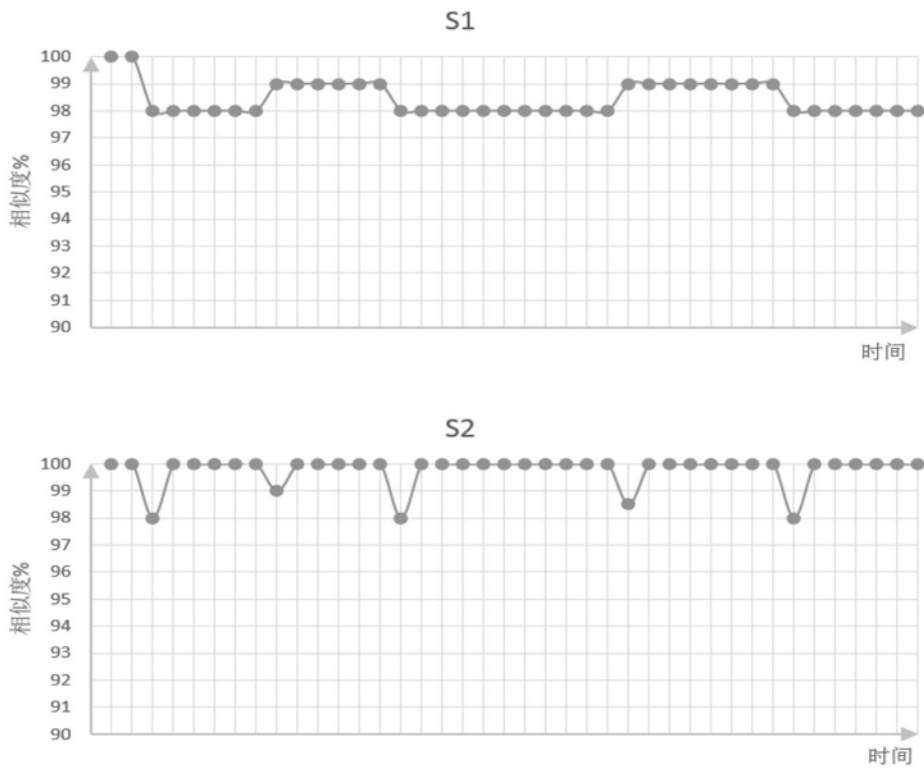


图4

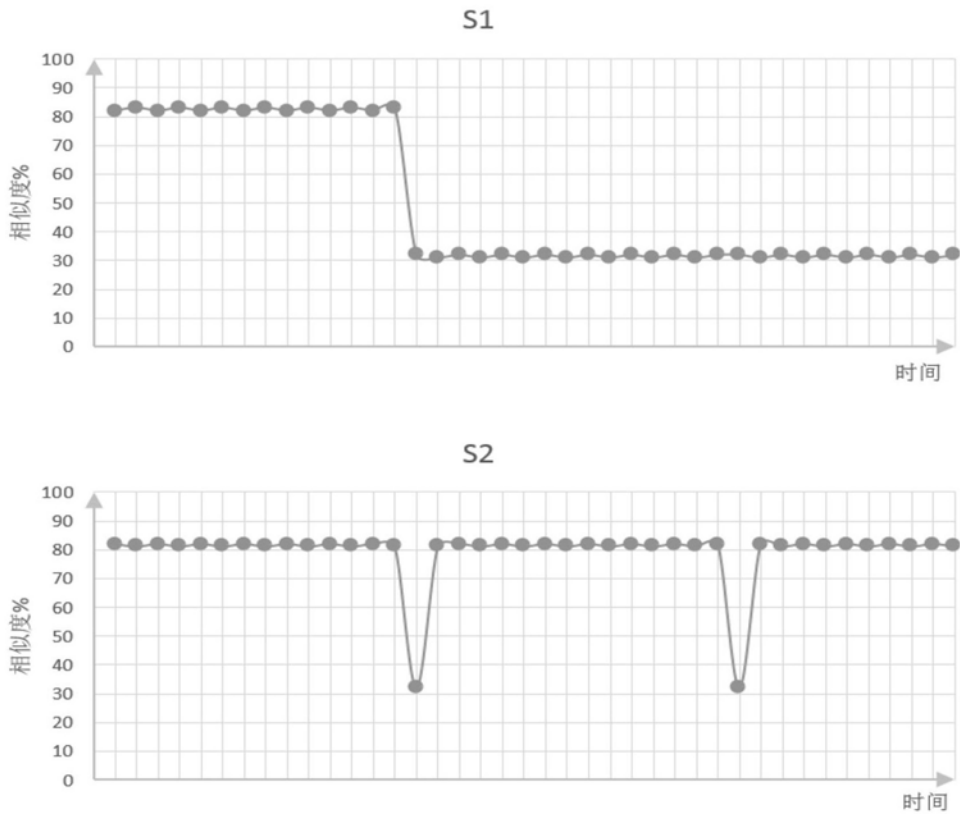


图5

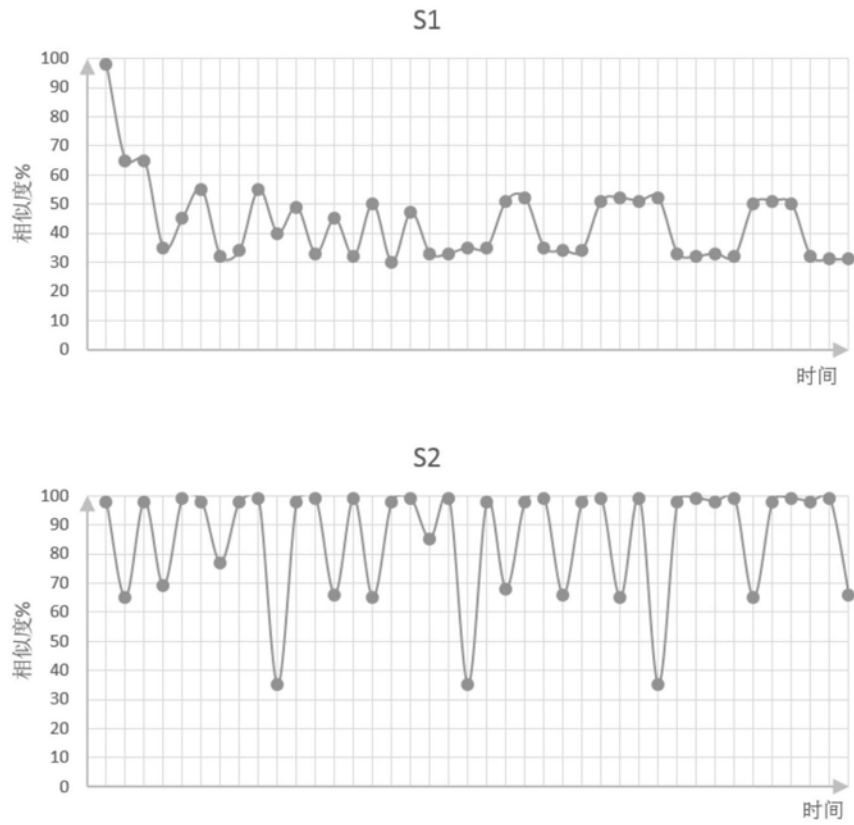


图6

70

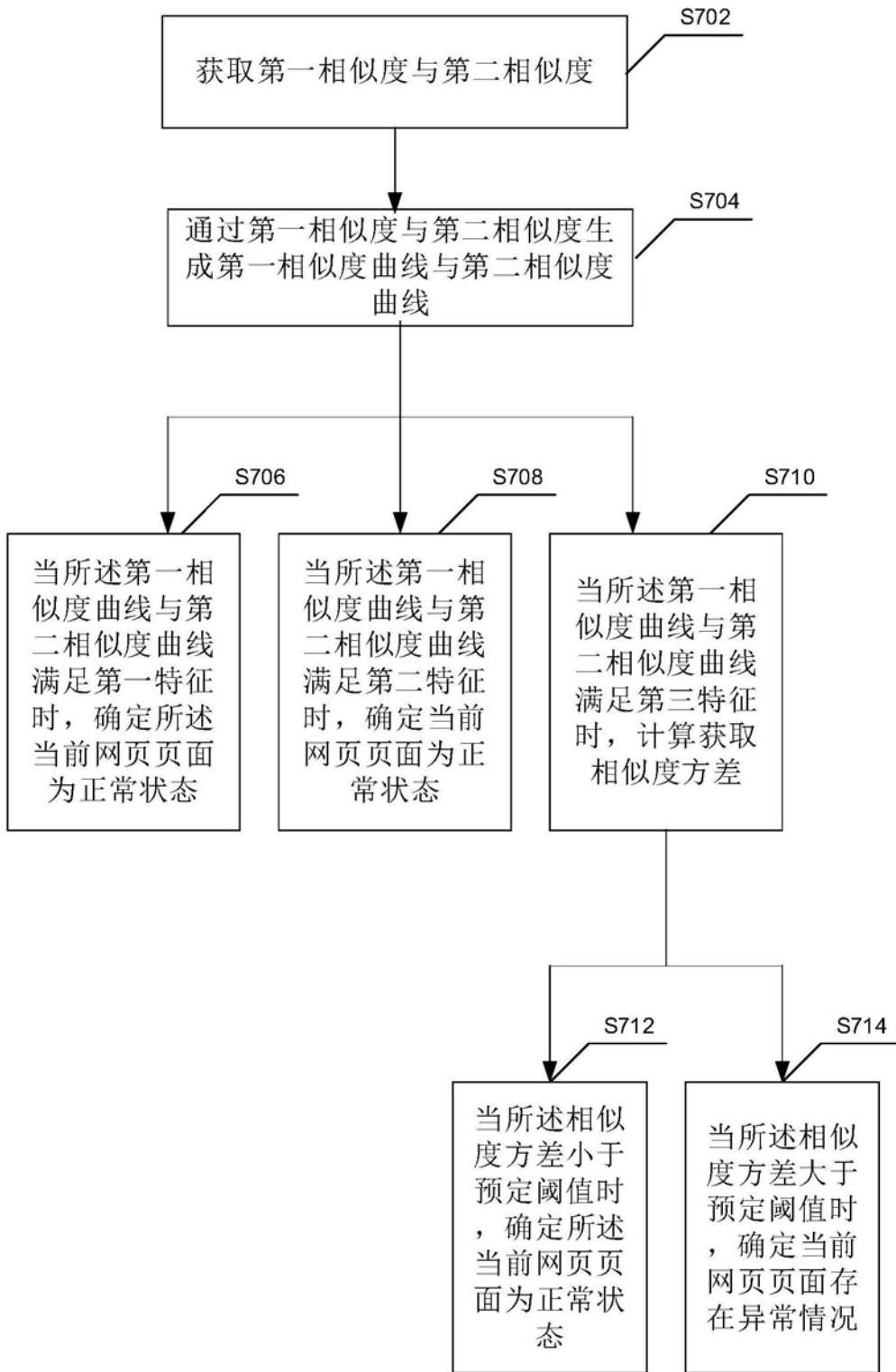


图7

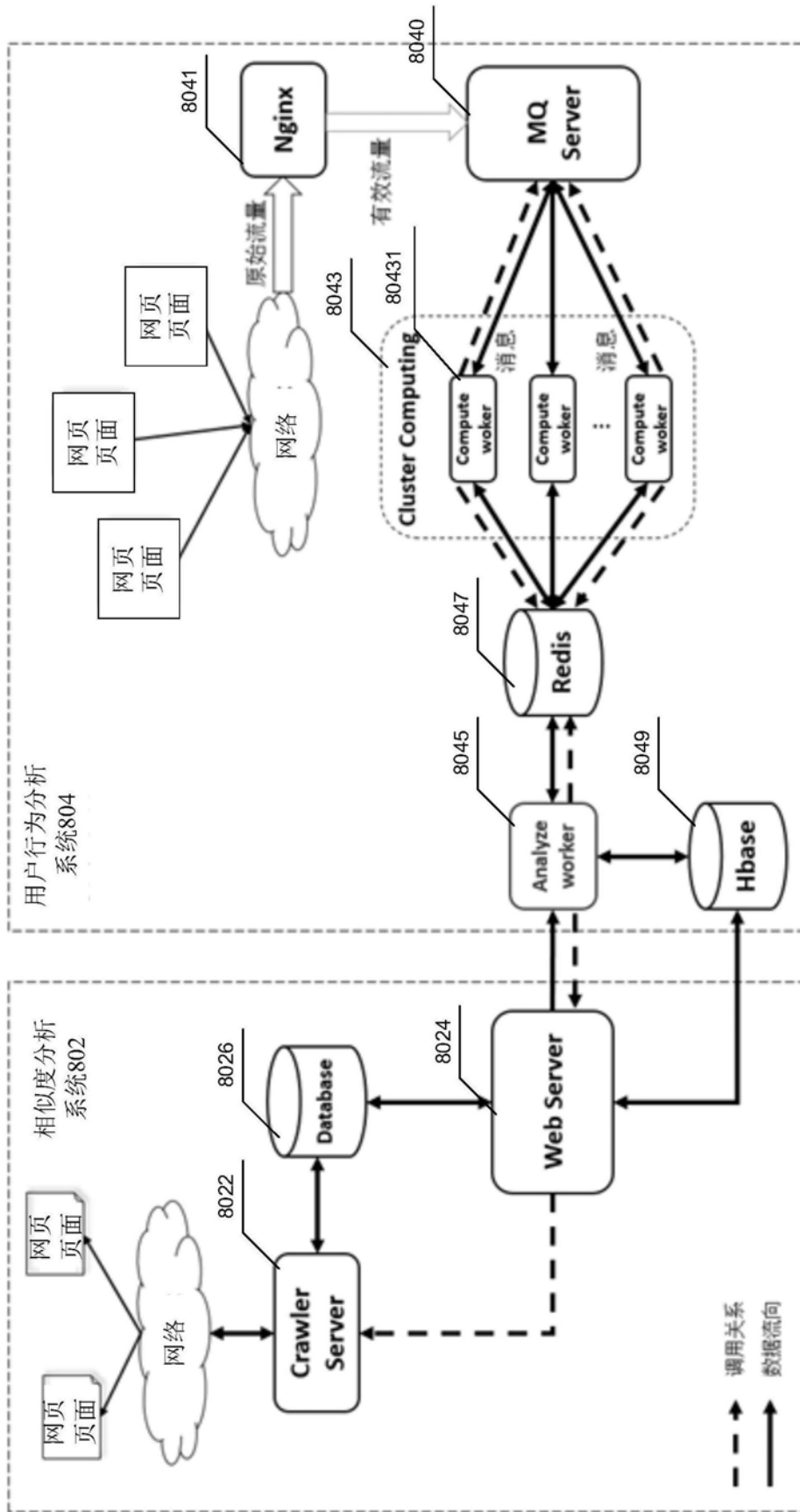


图8

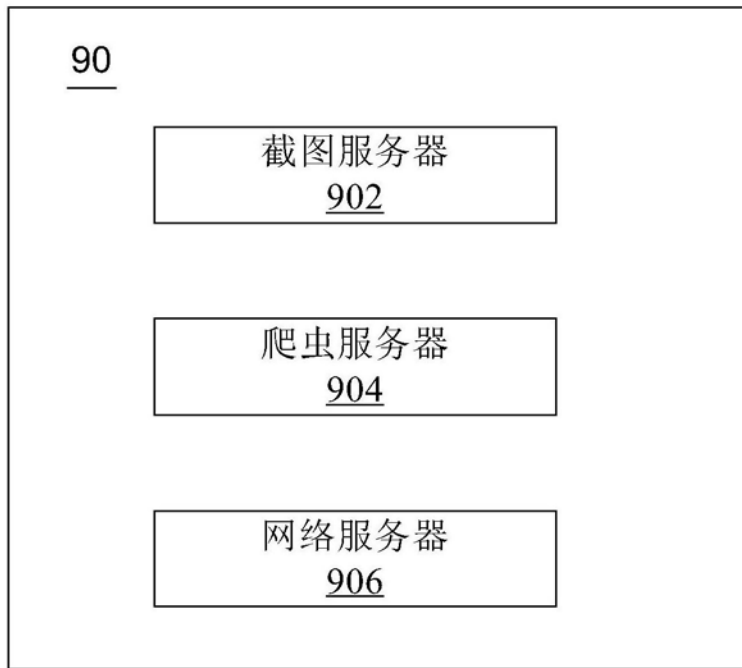


图9

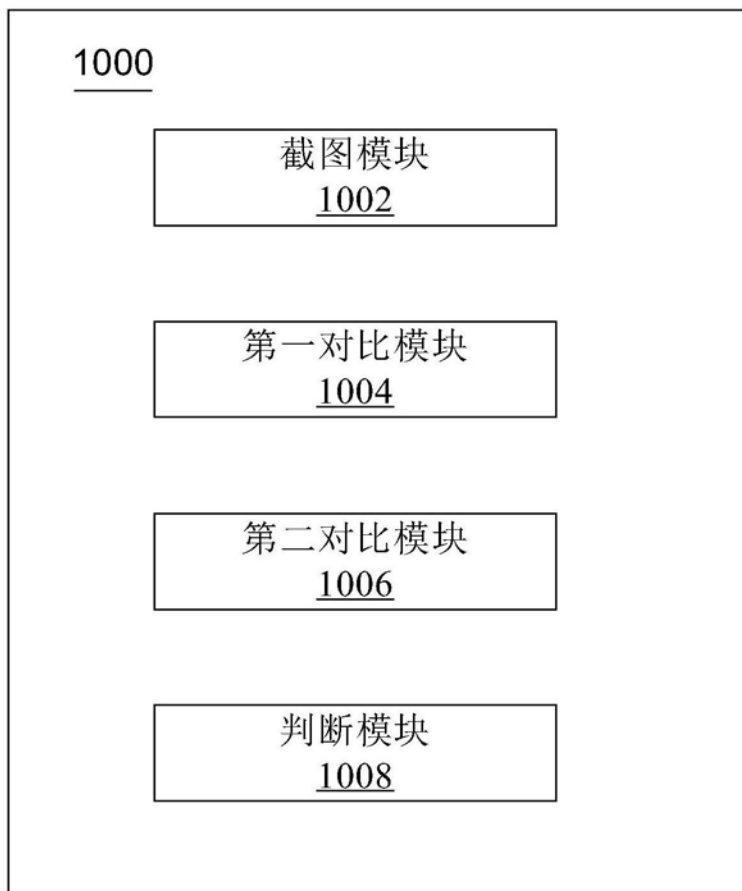


图10

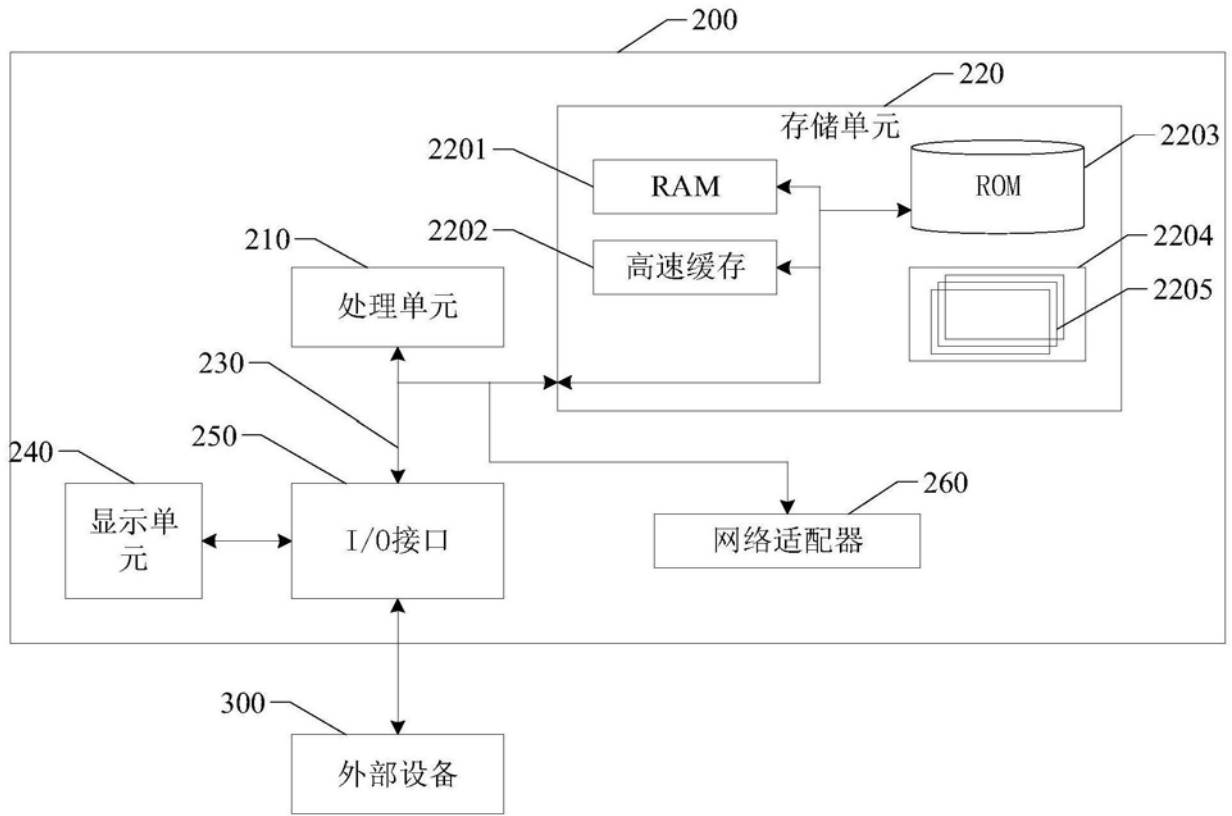


图11

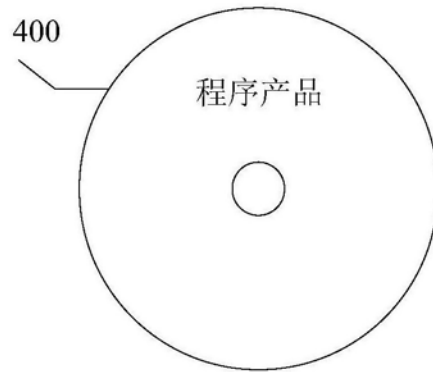


图12