(19)国家知识产权局

(12)发明专利

(10)授权公告号 CN 111625625 B
(45)授权公告日 2024.12.17

(21)申请号 202010460929.8
(22)申请日 2020.05.27
(65)同一申请的已公布的文献号
申请公布号 CN 111625625 A
(43)申请公布日 2020.09.04
(73)专利权人 腾讯科技(深圳)有限公司
地址 518000 广东省深圳市南山区高新区科技中一路腾讯大厦35层
(72)发明人 许景禧
(74)专利代理机构 深圳市深佳知识产权代理事务所(普通合伙) 44285
专利代理师 常忠良
(51)Int.Cl.
G06F 16/33(2019.01)
G06F 11/34(2006.01)
G06F 16/35(2019.01)
(56)对比文件
CN 110210512 A,2019.09.06
CN 105653427 A,2016.06.08
CN 109634818 A,2019.04.16
审查员 蔡智勇

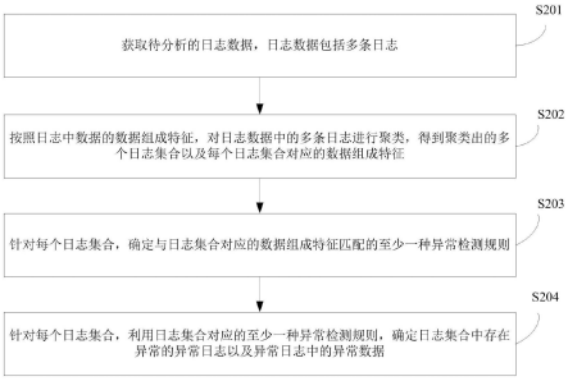
权利要求书3页 说明书17页 附图7页

(54)发明名称

确定异常日志的方法、装置、计算机设备和存储介质

(57)摘要

本申请公开了一种确定异常日志的方法、装置、计算机设备和存储介质,该方法包括:获取待分析的日志数据,日志数据包括多条日志;按照日志中数据的数据组成特征,对日志数据中的多条日志进行聚类,得到聚类出的多个日志集合以及每个日志集合对应的数据组成特征,日志集合包括至少一条日志;针对每个日志集合,确定与日志集合对应的数据组成特征匹配的至少一种异常检测规则;针对每个日志集合,利用日志集合对应的至少一种异常检测规则,确定日志集合中存在异常的异常日志以及异常日志中的异常数据。本申请的方案可以降低从大量日志数据中确定出异常日志的复杂度。



CN 111625625 B

1.一种确定异常日志的方法,其特征在于,包括:

获取待分析的日志数据,所述日志数据包括多条日志;

按照日志中包含的功能标识和字符特征,对所述日志数据中的多条日志进行聚类,得到聚类出的多个日志集合以及每个所述日志集合对应的数据组成特征,所述日志集合包括至少一条日志,日志中的功能标识用于表征日志中记录的运行状态信息所归属的功能模式的类型,所述数据组成特征表征日志中包含的数据项的个数和种类,以及每种数据项对应的数值,所述字符特征包括日志中包含的实数的个数、实数之外的字符的数量以及实数之外的字符的具体内容中的一种或者几种;

确定设定的检测模式,所述检测模式为人工检测模式和自动检测模式中的一种;

在设定的检测模式为自动检测模式的情况下,针对每个所述日志集合,基于所述日志集合对应的数据组成特征,确定所述日志集合中的日志所对应的所述功能模式,按照数据组成特征与异常检测规则的对应关系查找所述功能模式匹配的至少一种异常检测规则,所述异常检测规则包括日志中不同种数据项所需满足的条件;

针对每个所述日志集合,利用所述日志集合对应的至少一种异常检测规则,确定所述日志集合中存在异常的异常日志以及所述异常日志中的异常数据;

在设定的检测模式为人工检测模式的情况下,显示日志展现界面,所述日志展现界面显示有所述日志数据中的多条日志;

响应于所述日志展现界面检测到日志选择操作,确定所述日志选择操作所选择的目标日志;

确定所述目标日志所属的目标日志集合;

显示所述目标日志集合中各条日志的数据。

2.根据权利要求1所述的方法,其特征在于,还包括:

在所述日志集合中存在异常日志的情况下,确定所述日志集合中各异常日志中的异常数据所归属的至少一种目标数据项,所述目标数据项属于所述日志集合的各日志中具有的至少一种数据项;

针对所述日志集合对应的每种目标数据项,构建所述目标数据项的异常标识曲线图,所述异常标识曲线图包括所述日志集合的各日志中所述目标数据项的数值对应的数值变化曲线图,且在所述数值变化曲线图中标示有目标数据项中异常数据的信息;

显示出所述日志集合中各目标数据项的异常标识曲线图。

3.根据权利要求1所述方法,其特征在于,还包括:

针对每个日志集合,从存储的日志样本数据对应的多个日志样本集合中,确定与所述日志集合对应的数据组成特征匹配的目标日志样本集合,所述日志样本数据包括多个无异常数据的日志样本,且,所述多个日志样本集合为基于日志样本中数据的数据组成特征对所述日志样本数据的多个日志样本聚类得到的;

针对每个日志集合,确定所述日志集合在所述日志数据中的第一日志出现频率以及所述日志集合对应的目标日志样本集合在所述日志样本数据中的第二日志出现频率,并在基于所述第一日志出现频率和第二日志出现频率确定出所述日志集合中日志的数量存在数量增多异常的情况下,将所述日志集合确定为存在异常风险的风险日志集合,其中,所述第一日志出现频率为所述日志集合中包含的日志的第一数量与所述日志数据中包含的日志

的第一总数量的比值,所述第二日志出现频率为所述日志集合对应的目标日志样本集合中包含日志的第二数量与所述日志样本数据中包含的日志的第二总数量之间的比值;

输出针对所述风险日志集合的提示信息。

4.根据权利要求3所述的方法,其特征在于,所述输出针对所述风险日志集合的提示信息,包括:

输出针对所述风险日志集合的规则补充提示,所述规则补充提示用于提示用户针对所述风险日志集合对应的数据组成特征增设异常检测规则。

5.根据权利要求1所述的方法,其特征在于,在所述显示所述目标日志集合中各条日志的数据之前,还包括:

分别确定所述目标日志集合对应的各种数据项各自的数值变化曲线,目标日志集合对应的数据项为目标日志集合的日志中包含的数据项,每种数据项的数值变化曲线为基于所述目标日志集合的各条日志中该数据项的数值,构建出的数值变化曲线;

所述显示所述目标日志集合中各条日志的数据,包括:

显示曲线图展现界面,所述曲线图展现界面展现有所述目标日志集合对应的各种数据项的数值变化曲线。

6.一种确定异常日志的装置,其特征在于,包括:

日志获取单元,用于获取待分析的日志数据,所述日志数据包括多条日志;

日志聚类单元,用于按照日志中数据的数据组成特征,对所述日志数据中的多条日志进行聚类,得到聚类出的多个日志集合以及每个所述日志集合对应的数据组成特征,所述日志集合包括至少一条日志;

模式确定单元,用于确定设定的检测模式,所述检测模式为人工检测模式和自动检测模式中的一种;

规则匹配单元,用于针对每个所述日志集合,基于所述日志集合对应的数据组成特征,确定所述日志集合中的日志所对应的功能模式,按照数据组成特征与异常检测规则的对应关系查找所述功能模式匹配的至少一种异常检测规则,所述异常检测规则包括日志中不同种数据项所需满足的条件;

异常确定单元,用于针对每个所述日志集合,利用所述日志集合对应的至少一种异常检测规则,确定所述日志集合中存在异常的异常日志以及所述异常日志中的异常数据;

所述日志聚类单元,包括:

日志聚类子单元,用于按照日志中包含的功能标识和字符特征,对日志数据中的多条日志进行聚类,其中,日志中的功能标识用于表征日志中记录的运行状态信息所归属的功能模式的类型,日志的数据组成特征表征日志中包含的数据项的个数和种类,以及每种数据项对应的数值,所述字符特征包括日志中包含的实数的个数、实数之外的字符的数量以及实数之外的字符的具体内容中的一种或者几种;

日志展现单元,用于在设定的检测模式为人工检测模式的情况下,显示日志展现界面,所述日志展现界面显示有所述日志数据中的多条日志;

日志选择单元,用于响应于所述日志展现界面检测到日志选择操作,确定所述日志选择操作所选择的目标日志;

集合确定单元,用于确定所述目标日志所属的目标日志集合;

数据显示单元,用于显示所述目标日志集合中各条日志的数据。

7.一种计算机设备,其特征在于,所述计算机设备包括:处理器和存储器;

所述处理器,用于调用并执行所述存储器中存储的程序;

所述存储器用于存储所述程序,所述程序至少用于实现如上权利要求1至5任一项所述的确定异常日志的方法。

8.一种存储介质,其特征在于,所述存储介质中存储有计算机可执行指令,所述计算机可执行指令被处理器加载并执行时,实现如上权利要求1至5任一项所述的确定异常日志的方法。

确定异常日志的方法、装置、计算机设备和存储介质

技术领域

[0001] 本申请涉及数据处理技术领域,尤其涉及一种确定异常日志的方法、装置、计算机设备和存储介质。

背景技术

[0002] 在软件产品(如应用程序,或者,包含多个应用程序的系统平台等)运行过程中,均会记录表征其运行状况的日志。通过对软件产品的日志进行分析可以及时发现软件产品存在的问题。

[0003] 其中,软件产品中一般都会包含实现不同功能模式的程序模块,由于不同程序模块的差异性,不同程序模块所产生的日志形式会有所不同,因此,软件产品所产生的日志数据不仅数据量大,而且日志数据也较为复杂多样。同时,由于软件产品中不同程序模块的异常表现形式不同,因此,针对不同程序模块的日志数据的异常分析方式也会有所不同。基于此,为了从软件产品的日志数据中定位出表征软件产品存在异常的异常日志,需要人工对日志数据中各条日志逐条分析,导致确定异常日志的复杂度较高。

发明内容

[0004] 有鉴于此,本申请提供了一种确定异常日志的方法、装置、计算机设备和存储介质,以降低从大量日志数据中确定出异常日志的复杂度。

[0005] 为实现上述目的,一方面,本申请提供了一种确定异常日志的方法,包括:

[0006] 获取待分析的日志数据,所述日志数据包括多条日志;

[0007] 按照日志中数据的数据组成特征,对所述日志数据中的多条日志进行聚类,得到聚类出的多个日志集合以及每个所述日志集合对应的数据组成特征,所述日志集合包括至少一条日志;

[0008] 针对每个所述日志集合,确定与所述日志集合对应的数据组成特征匹配的至少一种异常检测规则;

[0009] 针对每个所述日志集合,利用所述日志集合对应的至少一种异常检测规则,确定所述日志集合中存在异常的异常日志以及所述异常日志中的异常数据。

[0010] 在一种可能的情况中,所述按照日志中数据的数据组成特征,对所述日志数据中的多条日志进行聚类,包括:

[0011] 按照日志中包含的功能标识和字符特征,对所述日志数据中的多条日志进行聚类,其中,日志中的功能标识用于表征日志中记录的运行状态信息所归属的功能模式的类型。

[0012] 在又一种可能的情况中,还包括:

[0013] 在所述日志集合中存在异常日志的情况下,确定所述日志集合中各异常日志中的异常数据所归属的至少一种目标数据项,所述目标数据项属于所述日志集合的各日志中具有的至少一种数据项;

[0014] 针对所述日志集合对应的每种目标数据项,构建所述目标数据项的异常标识曲线图,所述异常标识曲线图包括所述日志集合的各日志中所述目标数据项的数值对应的数值变化曲线图,且在所述数值变化曲线图中标示有目标数据项中异常数据的信息;

[0015] 显示出所述日志集合中各目标数据项的异常标识曲线图。

[0016] 在又一种可能的情况中,还包括:

[0017] 针对每个日志集合,从存储的日志样本数据对应的多个日志样本集合中,确定与所述日志集合对应的数据组成特征匹配的目标日志样本集合,所述日志样本数据包括多个无异常数据的日志样本,且,所述多个日志样本集合为基于日志样本中数据的数据组成特征对所述日志样本数据的多个日志样本聚类得到的;

[0018] 针对每个日志集合,确定所述日志集合在所述日志数据中的第一日志出现频率以及所述日志集合对应的目标日志样本集合在所述日志样本数据中的第二日志出现频率,并在基于所述第一日志出现频率和第二日志出现频率确定出所述日志集合中日志的数量存在数量增多异常的情况下,将所述日志集合确定为存在异常风险的风险日志集合,其中,所述第一日志出现频率为所述日志集合中包含的日志的第一数量与所述日志数据中包含的日志的第一总数量的比值,所述第二日志出现频率为所述日志集合对应的目标日志样本集合中包含日志的第二数量与所述日志样本数据中包含的日志的第二总数量之间的比值;

[0019] 输出针对所述风险日志集合的提示信息。

[0020] 在又一种可能的情况中,在所述针对每个所述日志集合,确定与所述日志集合对应的数据组成特征匹配的至少一种异常检测规则之前,还包括:

[0021] 确定设定的检测模式,所述检测模式为人工检测模式和自动检测模式中的一种;

[0022] 所述针对每个所述日志集合,确定与所述日志集合对应的数据组成特征匹配的至少一种异常检测规则,包括:

[0023] 在设定的检测模式为自动检测模式的情况下,针对每个所述日志集合,确定与所述日志集合对应的数据组成特征匹配的至少一种异常检测规则;

[0024] 所述方法还包括:

[0025] 在设定的检测模式为人工检测模式的情况下,显示日志展现界面,所述日志展现界面显示有所述日志数据中的多条日志;

[0026] 响应于所述日志展现界面检测到日志选择操作,确定所述日志选择操作所选择的目标日志;

[0027] 确定所述目标日志所属的目标日志集合;

[0028] 显示所述目标日志集合中各条日志的数据。

[0029] 又一方面,本申请还提供了一种确定异常日志的装置,包括:

[0030] 日志获取单元,用于获取待分析的日志数据,所述日志数据包括多条日志;

[0031] 日志聚类单元,用于按照日志中数据的数据组成特征,对所述日志数据中的多条日志进行聚类,得到聚类出的多个日志集合以及每个所述日志集合对应的数据组成特征,所述日志集合包括至少一条日志;

[0032] 规则匹配单元,用于针对每个所述日志集合,确定与所述日志集合对应的数据组成特征匹配的至少一种异常检测规则;

[0033] 异常确定单元,用于针对每个所述日志集合,利用所述日志集合对应的至少一种

异常检测规则,确定所述日志集合中存在异常的异常日志以及所述异常日志中的异常数据。

[0034] 又一方面,本申请还提供了一种计算机设备,所述计算机设备包括:处理器和存储器;

[0035] 所述处理器,用于调用并执行所述存储器中存储的程序;

[0036] 所述存储器用于存储所述程序,所述程序至少用于实现如任一项所述的确定异常日志的方法。

[0037] 又一方面,本申请还提供了一种存储介质,所述存储介质中存储有计算机可执行指令,所述计算机可执行指令被处理器加载并执行时,实现如上任一项所述的确定异常日志的方法。

[0038] 由以上内容可知,本申请中,按照日志的数据组成特征,将待分析的日志数据中的多条日志进行聚类,由于日志中数据的数据组成特征可以反映出日志是针对软件产品的哪项功能的运行状态所生成的,因此,基于数据组成特征可以将记录同一功能的运行状态的日志聚类到同一个日志集合;而且,通过日志集合对应的数据组成特征可以反映出该日志集合中的日志所针对的功能,因此,可以将相应功能对应的异常检测规则(即日志集合的数据组成特征匹配的异常检测规则)确定为该日志集合中日志所适用的异常检测规则,从而能够确定日志集合中的日志所适用的异常检测规则,进而可以直接利用相应的异常检测规则对日志集合中的日志进行异常检测,避免了人工逐条分析日志异常所导致的复杂度,也就降低了确定异常日志的复杂度。

附图说明

[0039] 为了更清楚地说明本申请实施例中的技术方案,下面将对实施例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申请的实施例,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,还可以根据提供的附图获得其他的附图。

[0040] 图1示出了本申请的一种确定异常日志的方法所适用的一种场景的组成架构示意图;

[0041] 图2示出了本申请提供的确定异常日志的方法的一种流程示意图;

[0042] 图3示出了本申请针对存在异常日志的日志集合输出异常标识曲线图的一种流程示意图;

[0043] 图4示出了本申请中数据项的数值变化曲线图的一种示意图;

[0044] 图5示出了本申请中数据项的异常标识曲线图的一种示意图;

[0045] 图6示出了本申请中包含异常报告和异常标识曲线图的异常显示界面的一种示意图;

[0046] 图7示出了本申请一种确定异常日志的方法的又一种流程示意图;

[0047] 图8示出了本申请中异常汇总报表的一种示意图;

[0048] 图9示出了本申请一种确定异常日志的方法的又一种流程示意图;

[0049] 图10示出了本申请一种确定异常日志的装置的一种组成结构示意图;

[0050] 图11示出了本申请中一种计算机设备的一种组成架构示意图。

具体实施方式

[0051] 本申请的方案适用于从大量日志数据中确定出异常日志,以降低确定异常日志的复杂度。在本申请中提到的日志可以为应用等软件产品运行过程中所产生的运行日志,通过对日志进行异常分析可以及时发现软件产品中存在的性能差以及故障等问题。

[0052] 本申请的方案可以应用于个人计算机、服务器或者多个服务器构成的服务器系统,为了提高确定异常日志的效率,本申请还可以应用于云平台或者其他计算系统。

[0053] 为了便于理解,以本申请的方案应用于云平台这一场景为例说明。如图1所示,其示出了本申请所适用的一种场景的组成架构示意图。

[0054] 由图1可以看出,该场景包括:云平台10,云平台可以包括多个云服务器101。

[0055] 该场景还可以包括多个应用客户端20,应用客户端中可以运行有待监测的应用。相应的,应用客户端在应用运行过程中,会产生与应用运行相关的日志,如,应用运行中内存以及处理器等的使用情况等以及应用的一些数据响应情况等等。

[0056] 如,假设需要基于游戏应用的日志分析游戏应用是否存在异常,那么,应用客户端可以为游戏客户端,游戏客户端会生成并存储游戏运行过程中的相关日志。

[0057] 其中,应用客户端20可以将应用运行中产生的日志发送给云平台10。

[0058] 相应的,云平台可以获得不同应用客户端对应的日志数据,得到多份日志数据,并依次对每份日志数据进行异常分析。

[0059] 其中,云平台也成为云计算平台,其基于云技术构建出的网络平台。其中,云技术(Cloud technology)是指在广域网或局域网内将硬件、软件、网络等系列资源统一起来,实现数据的计算、储存、处理和共享的一种托管技术。

[0060] 云技术(Cloud technology)是基于云计算商业模式应用的网络技术、信息技术、整合技术、管理平台技术、应用技术等的总称,可以组成资源池,按需所用,灵活便利。技术网络系统的后台服务需要大量的计算、存储资源,如图像存储以及编码等等。伴随着互联网行业的高度发展和应用,将来每个物品都有可能存在自己的识别标志,都需要传输到后台系统进行逻辑处理,不同程度级别的数据将会分开处理,各类行业数据皆需要强大的系统后盾支撑,只能通过云计算来实现。

[0061] 其中,云计算(cloud computing)是一种计算模式,它将计算任务分布在大量计算机构成的资源池上,使各种应用系统能够根据需要获取计算力、存储空间和信息服务。提供资源的网络被称为“云”。“云”中的资源在使用者看来是可以无限扩展的,并且可以随时获取,按需使用,随时扩展,按使用付费。

[0062] 作为云计算的基础能力提供商,会建立云计算资源池(简称云平台,一般称为IaaS(Infrastructure as a Service,基础设施即服务)平台,在资源池中部署多种类型的虚拟资源,供外部客户选择使用。云计算资源池中主要包括:计算设备(为虚拟化机器,包含操作系统)、存储设备、网络设备。

[0063] 在图1的场景中,通过在云平台的云服务器可以完成日志数据的收集、分析以及异常日志的确定等处理。

[0064] 可以理解的是,图1是以云平台直接从应用客户端获得应用客户端中生成的日志数据为例,在实际应用中,云平台还可以通过其他设备或者通过其他网络途径获得日志数据,对此不加限制。

[0065] 另外,图1是以获得应用客户端的日志数据为例说明,在实际应用中,如果需要对某个系统或者网络平台等进行分析,则云平台还可以获得系统或者网络平台等运行相关的日志数据。

[0066] 需要说明的是,图1是为了便于理解,以本申请应用于云平台为例说明,当利用单个计算机设备、服务器或者服务器集群分析日志数据时,只需要将图1所示场景中的云平台替换为相应的设备或者服务器集群等即可,其场景类似,在此不再赘述。

[0067] 下面结合流程图对本申请的方案进行介绍。

[0068] 如图2所示,其示出了本申请一种确定异常日志的方法的一种流程示意图,本实施例的方法可以应用于前面提到的计算机设备、服务器或者云平台等等。本实施例的方法可以包括:

[0069] S201,获取待分析的日志数据,日志数据包括多条日志。

[0070] 如,待分析的日志数据可以为软件产品的运行日志,日志数据中每条日志都可以记录软件产品运行过程中的运行状态数据等。如,某个客户端中应用的运行日志,也可以是某个系统或者网络平台的运行日志等等。

[0071] 可以理解的是,对于应用或者软件系统等等,不同时刻可能会产生不同的日志,因此,每份日志数据实际上是包含了分别在多个不同时刻生成的多条日志。如,日志数据可以为一份日志数据文本,该日志数据文本中可以包含多行日志,该多行日志可以看成是多个时刻生成的日志序列。

[0072] 可选的,为了区分日志数据中各条日志,每条日志可以对应有日志标识,该日志标识可以用于唯一标识该日志数据中的一条日志。如,日志标识可以为日志序号。例如,日志数据由多行日志组成的情况下,每行日志对应一个行号,该行号就是日志的标识。

[0073] 需要说明的是,在实际应用中,可能会同时获得多份来自不同设备端的日志数据,但是对于每份日志数据的分析过程均相同,均可以采用与本申请中后续步骤进行异常日志的确定。

[0074] S202,按照日志中数据的数据组成特征,对日志数据中的多条日志进行聚类,得到聚类出的多个日志集合以及每个日志集合对应的数据组成特征。

[0075] 其中,日志的数据组成特征表示日志中所包含的数据内容所具有的特征。如,日志的数据组成特征可以表征日志中包含的数据项的个数和种类,以及每种数据项对应的数值的具体形式等等。

[0076] 可以理解的是,由于软件产品一般会具有多种不同的功能模式,如软件产品包括不同功能模块,不同功能模块实现不同功能模式。例如,游戏应用中可能会包含数据编解码功能和图像渲染功能,游戏应用会在分别执行这两项功能模式的过程,会分别得到这两种功能模式下各自产生的日志。

[0077] 相应的,软件产品记录的日志可以包含针对不同功能模式的运行状况的日志,而不同功能模式对应的日志中所包含的数据内容以及具体形式会有较大差别。但是针对同一功能模式的日志中数据内容以及具体形式相同或者近乎相同(即功能模式数据组成特征相同或者相似),因此,可以基于数据组成特征可以确定日志数据中哪些日志是针对同一功能模式生成的日志。基于此,可以基于数据组成特征,将针对同一功能模式的日志聚类到一个日志集合中。

[0078] 其中,每个日志集合中可以包括至少一条日志。可以理解的是,具有相同的数据组成特征的日志会被聚类到同一日志集合,因此,日志集合内各条日志普遍适用的数据组成特征就是该日志集合对应的数据组成特征。类似的,日志集合对应的数据组成特征反映出的日志集合内的日志所针对的功能模式。

[0079] 可以理解的是,由于同一日志集合内包含具有相同数据组成特征的日志,因此,一条日志只能属于某一个日志集合,而不同日志集合内的日志也不存在重叠。

[0080] 可以理解的是,日志的数据组成特征可以通过多种维度来反映。作为一种可能的情况,日志的数据组成特征可以日志中包含的功能标识和字符特征。其中,日志中的功能标识用于表征日志中记录的运行状态信息所归属的功能模式的类型,如,功能标识可以为功能模式的名称等。字符特征可以包括:日志中包含的实数的个数、实数之外的字符的数量以及实数之外的字符的具体内容等特征中的一种或者几种。相应的,可以按照日志中包含的功能标识和字符特征,对日志数据中的多条日志进行聚类。

[0081] 例如,以游戏应用的日志数据为例,在游戏应用中会涉及到渲染这一功能模式,那么渲染这一功能模式的日志中会记录有:表征渲染模式的文件名称、分辨率以及渲染数据的传输速率。相应的日志可以为:**.render;分辨率:1920*180;传输速率:25bps,那么该日志中包含的功能标识为“**.render”,且包含有三个实数,分别为1980、180和25,实数之外的字符至少包括“分辨率”和“传输速率”。相应的,如果其他日志也具有该功能标识“**.render”,具有“分辨率”和“传输速率”这两个字符串,且具有三个实数,则可以认为这些日志同样是针对渲染模式生成的日志,从而将这些日志聚类到一个日志集合。

[0082] S203,针对每个日志集合,确定与日志集合对应的数据组成特征匹配的至少一种异常检测规则。

[0083] 其中,与日志集合对应的数据组成特征匹配的异常检测规则实际上就是适用于该日志集合中日志对应的功能模式的异常检测规则。

[0084] 如,可以预先针对每种数据组成特征配置至少一种异常检测规则,相应的,可以按照数据组成特征与异常检测规则的对应关系,确定适合对日志集合中各条日志进行异常检测的只收一种异常检测规则。

[0085] 又如,还可以预先配置不同功能模式各自适用的至少一种异常检测规则。相应的,针对每个日志集合,可以基于日志集合对应的数据组成特征,确定该日志集合中的日志所对应的功能模式,然后再查找该功能模式适合的至少一种异常检测规则。

[0086] 其中,异常检测规则可以包括日志中不同种数据项所需满足的条件。其中,日志中的数据项也可以称为参数项,不同数据项表征日志中记录的功能模式的一种状态指标。比如,针对日志中的数据项可以包括:CPU使用率,以及内存占用率等等,相应的,异常检测规则可以包括,CPU使用率大于设定CPU使用率,内存占用率大于设定占用率等等。

[0087] 可以理解的是,异常检测规则可以根据不同的软件产品,并结合实际需要设定,对此不加限制。

[0088] S204,针对每个日志集合,利用日志集合对应的至少一种异常检测规则,确定日志集合中存在异常的异常日志以及异常日志中的异常数据。

[0089] 可以理解的是,如果日志集合中的日志与异常检测规则中异常条件相符,则说明日志属于异常日志。其中,日志属于异常日志说明日志中至少一个数据项的数值存在异常,

因此,基于异常检测规则可以定位异常日志中出现异常的异常数据。如,异常数据可以为异常日志中存在异常的数据项以及存在异常的数据项的取值。

[0090] 在一种可能的情况下,为了便于对缓存数据以及对数据进行异常检测,针对每个日志集合,还可以提取该日志集合中日志中具有的至少一种数据项,将提取出的各数据项分别作为待构建的数据表中的字段,并基于该日志集合中每条日志中该至少一种数据项各自数值,构建数据表。

[0091] 其中,日志集合中各条日志所具有的数据项相同,因此,可以提取日志集合中任意一条日志所具有的所有数据项。如日志集合中每条日志都包含内存占用率和CPU使用率这两个数据项,则会提取这两个数据项作为数据表的两个字段。

[0092] 其中,基于日志集合构建出的数据表中每条记录对应该日志集合中的一条日志,且不同日志对应数据表中的不同记录。如,日志集合中的日志具有在日志数据中唯一的行号,那么数据表中可以通过日志对应的行号标识日志对应的记录。

[0093] 举例说明,假设日志集合中包含日志行号为11和13的两条日志,且从日志集合的日志中提取出的字段分别为CPU使用率和内存占用率,那么基于日志集合构建的数据表可以具有CPU使用率和内存占用率,且,在该数据表中的一条记录为日志行号为11的日志,在该条记录中记录有CPU使用率对应的数值以及内存使用率对应的数值。相应的,数据表中还有日志行号为13的记录。

[0094] 在通过数据表缓存日志集合中各条日志的数据后,在确定出日志集合适用的至少一种异常检测规则之后,可以基于至少一种异常检测规则对该数据表中各条记录进行异常检测,以检测出存在字段的取值存在异常的异常记录。相应的,该异常记录对应的日志为异常日志,且异常记录中存在异常的字段的数据为该异常日志中的异常数据。

[0095] 可见,在本申请实施例中,可以按照日志的数据组成特征,将待分析的日志数据中的多条日志进行聚类,由于日志中数据的数据组成特征可以反映出日志是针对软件产品的哪项功能的运行状态所生成的,因此,基于数据组成特征可以将记录同一功能的运行状态的日志聚类到同一个日志集合;而且,通过日志集合对应的数据组成特征可以反映出该日志集合中的日志所针对的功能,因此,可以将相应功能对应的异常检测规则(即日志集合的数据组成特征匹配的异常检测规则)确定为该日志集合中日志所适用的异常检测规则,从而使得确定日志集合中的日志所适用的异常检测规则提供了可能,进而可以直接利用相应的异常检测规则对日志集合中的日志进行异常检测,避免了人工逐条分析日志异常所导致的复杂度,也就降低了确定异常日志的复杂度。

[0096] 可以理解的是,在步骤S204确定出日志集合中的异常日志之后,为了能够便于用户直观的了解到哪些日志中的哪些数据项的数值存在异常,本申请还可以针对日志集合输出能够反映异常日志中存在异常的数据项对应的曲线图。如,参见图3,其示出了本申请中针对存在异常日志的日志集合输出异常标识曲线图的一种流程示意图,本实施例的流程可以包括:

[0097] S301,针对每个日志集合,在该日志集合中存在异常日志的情况下,确定日志集合中各异常日志中的异常数据所归属的至少一种目标数据项。

[0098] 其中,针对任意一个异常日志,该异常日志可能会有至少一个数据项的数据存在异常,而存在数据异常的数据项及其数值就是异常数据,因此,每项异常数据都对应了一个

数据项。为了便于区分,将异常日志中存在数据异常的数据项称为目标数据项。

[0099] 举例说明,日志集合中存在异常的异常日志包括日志1和日志2,日志1中数据项a1的数据存在异常,而日志2中数据项a2的数据存在异常。那么对于日志1而言,其异常数据归属的目标数据项为数据项a1;而日志2中异常数据归属的目标数据项为数据项a2。

[0100] 其中,日志集合中异常日志可以有一个或者多个。在日志集合中异常日志有多个时,不同异常日志中存在数据异常的目标数据项的种类和个数均有所差别,本申请会统计出所有存在数据异常的目标数据项,从而得到至少一个目标数据项。如上面例子,虽然日志1中数据项a1存在数据异常,数据项a1属于存在异常的目标数据项;虽然日志1中数据项a2不存在数据异常,但是日志2中该数据项a2的数据存在异常,因此,数据项a2也属于存在异常的目标数据项。

[0101] S302,针对日志集合对应的每种目标数据项,构建目标数据项的异常标识曲线图。

[0102] 其中,目标数据项的异常标识曲线图包括日志集合的各日志中目标数据项的数值的数值变化曲线图,且在该数值变化曲线图中标示有目标数据项中的异常数据的信息。

[0103] 可以理解的是,每个目标数据项对应了一条数值变化曲线图,目标数据项的数值变化曲线图是基于日志集合中各个日志内该目标数据项的数值构建的,在该数值变化曲线图中可以呈现出日志集合中每个日志的该目标数据项的数值。

[0104] 如,目标数据项的数值变化曲线图中的坐标系可以包含相互垂直的两条坐标轴,一条坐标轴中每个值标识日志集合中一条日志,另一条坐标轴的不同值表示目标数据项的数值,相应的,可以依次在该坐标系中标注出日志集合中各个日志的目标数据项的取值,从而得到反映日志集合中各个日志的该目标数据项的数值变化的数值变化曲线图。

[0105] 为了便于理解,可以参见图4,图4示出了数据项的数值变化曲线图的一种示意图。在图4中以日志集合的各个日志中包含的数据项为CPU使用率的情况为例,并在图4中示出了CPU使用率的数值变化曲线图。

[0106] 在图4中横坐标表示各个日志的标识号,如日志集合包含的各个日志在日志数据中的行号。因此,图4的数值变化曲线中每个点对应的横坐标都对应了一个日志的行号。如图4,日志集合中至少包括行号为12、行号为20的日志和行号为25的日志等等,当然,对于该日志集合中包含的其他行号的日志无法一一列出。

[0107] 同时,图4的纵坐标标识CPU使用率的具体取值。相应的,根据日志集合中各条日志中记录的CPU使用率,可以在该坐标系中该日志对应位置处的纵坐标上标出相应的数值。比如,行号为12的日志中CPU使用率为40%,则该数值变化曲线图中对应横轴为20这一位置处的CPU使用率为60%,具体如图4的数值变化曲线图中坐标(20,60)这一点所示。

[0108] 可以理解的是,为了能够使得用户可以基于目标数据项的数值变化曲线图可以直观看出哪些日志的该目标数据项的数值存在异常,本申请可以在目标数据项的数值变化曲线图中标示有目标数据项中存在异常的异常数据的信息,从而得到异常标识曲线图。如,在目标数据项的数值变化曲线图中标出存在异常的坐标点,或者是,该目标数据项存在异常数据的日志所在的横坐标。

[0109] 例如,参见图5,其示出了本申请一种异常标识曲线图的一种示意图。

[0110] 图5是在图5示出的CPU使用率的数值变化曲线图的基础上得到的异常标识曲线。对比图4和图5可以看出,在图5中针对CPU使用率存在异常的日志,会在异常标识曲线图的

横坐标上标出异常日志对应的坐标位置501,如图5中横坐标上的白色圆点所示。为了便于直观看到各个异常日志的横坐标位置,在图5中仅仅以CPU使用率存在异常的日志有两个,分别为行号为20和行号为1001的日志为例,相应的,这两个异常日志的横坐标分别为20和1001。

[0111] 可以理解的是,在本申请实施例中,可以在确定出日志集合中的异常日志之后,再针对日志集合中各个异常日志对应的存在异常的目标数据项,分别构建目标数据项的数值变化曲线图。还可以是,在聚类得到日志集合之后,便针对日志集合中各个日志涉及到的每个数据项分别构建数值变化曲线图,相应的,在确定出日志集合中的异常日志之后,可以针对每个存在异常数据的目标数据项,直接从已构建的各个数据项的数值变化曲线图中,确定目标数据项的数值变化曲线图。

[0112] 其中,针对任意一个日志集合的每个数据项,构建该数据项的数值变化曲线图的具体方式可以有多种可能:

[0113] 如,在一种可能的实现方式中,可以提取日志集合中各条日志中该数据项的取值,并基于各条日志中该数据项的取值,构建该数据项的数值变化曲线图。

[0114] 又如,在又一种可能的实现方式中,在针对每个日志集合分别构建出该日志集合对应的数据表的情况下,可以直接基于数据表中该数据项对应字段在各个记录中的取值,构建该数据项的数值变化曲线图。在该种情况中,数据表中每个字段对应的列就是该数据项中各个日志在该数据项中的取值,从而可以直接基于各条记录在该列中的取值,生成该数据项的数值变化曲线图。

[0115] S303,显示出日志集合中各目标数据项的异常标识曲线图。

[0116] 在一种可能的实现方式中,针对每个日志集合,在完成该日志集合的异常日志分析之后,可以输出针对该日志集合中各目标数据项的异常标识曲线图。在该种情况中,针对每个日志集合,在显示日志集合的日志中目标数据项的异常标识曲线图时,还可以标识出该目标日志集合对应的数据组合特征,以便用户直观了解到是针对软件产品的哪个功能模式中该目标数据项的异常标识曲线图,并基于该异常标识曲线图了解到该功能模式对应的日志中哪些日志存在异常。

[0117] 在又一种可能的实现方式中,本申请可以在分析完所有日志集合后,统一输出存在异常日志的各个日志集合中各目标数据项的异常标识曲线图。

[0118] 可以理解的是,在实际应用中,在输出日志集合中目标数据项的异常标识曲线图的同时,本申请还可以针对日志集合的异常报告,异常报告中可以指示有日志集合中存在异常的异常日志以及异常日志中存在异常的具体信息等。

[0119] 如,可以在分析完各个日志集合之后,可以输出异常显示界面。在异常显示界面中显示各个日志集合的异常报告,并在该异常显示界面中呈现每个日志集合的各个目标数据项的异常标识曲线图。

[0120] 如图6所示,其示出了本申请中异常显示界面的一种示意图。

[0121] 由图6可以看出在异常显示界面中可以显示出针对日志集合中的异常日志所存在的各种具体异常原因等信息的异常报告601,同时,在异常报告的下方还可以呈现出该日志集合中存在异常的各目标数据项各自的异常标识曲线602,从而使得用户可以直观了解到具体异常原因,并能够基于异常标识曲线确定存在异常的日志。

[0122] 可以理解的是,在实际应用中,如果在正常情况下,针对某种功能模式生成的日志一般数量较少,那么如果检测到日志数据该种功能模式的日志的数量较多,则说明该种功能模式可能存在异常。因此,为了能够及时发现某种功能模式存在异常,本申请还可以基于日志集合中日志出现的频率是否存在异常,来分析日志集合是否属于存在异常的风险日志集合,进而及时基于风险日志集合确定出存在异常风险的功能模式。

[0123] 其中,检测日志集合是否为风险日志集合的操作可以是与前面实施例中针对日志集合进行异常日志检测的操作为两种并行确定异常日志的方式,当然,也可以是在前面针对日志集合进行异常日志检测的基础上,分析日志集合是否属于风险日志集合。

[0124] 为了便于理解,下面以一种情况为例介绍,如参见图7,其示出了本申请一种确定异常日志的方法的又一种实现流程。本实施例的流程可以包括:

[0125] S701,获取待分析的一份日志数据,该日志数据包括多条日志。

[0126] S702,按照日志中数据的数据组成特征,对日志数据中的多条日志进行聚类,得到聚类出的多个日志集合以及每个日志集合对应的数据组成特征。

[0127] S703,针对每个日志集合,将该日志集合中各日志具有的至少一种数据项作为待构建的数据表中的字段,并基于该日志集合中每条日志中该至少一种数据项各自数值,构建数据表。

[0128] S704,针对每个日志集合,确定与日志集合对应的数据组成特征匹配的至少一种异常检测规则,并基于该至少一种异常检测规则,确定日志集合的数据表中存在异常的异常记录以及异常记录中的异常数据。

[0129] 其中,由于日志集合对应的数据表中每条记录对应该日志集合中的一条日志,因此,异常记录对应的日志就是异常日志,相应的,异常记录中的异常数据确定为异常记录对应的异常日志中的异常数据。

[0130] 以上步骤S701到S704具体可以参见前面实施例的相关介绍,在此不再赘述。

[0131] 需要说明的是,在本申请是以聚类出日志集合之后,通过针对每个日志集合生成数据表,通过数据表缓存各个日志集合中各条日志的数据为例说明。但是如果不生成数据表,而利用异常检测规则直接对日志集合中各条日志进行异常检测也同样适用于本实施例。

[0132] S705,输出异常显示界面。

[0133] 其中,在该异常显示界面中可以显示出各个日志集合对应的异常报告,异常报告中可以提示出日志集合中存在异常的异常日志的原因以及异常情况等;同时,异常显示界面还可以呈现日志集合中存在异常的目标数据项的异常标识曲线。

[0134] 例如,用户可以根据点击异常报告中指示出的存在异常的异常日志,则可以在异常显示界面中呈现出该异常日志的目标数据项在该异常日志所在的日志集合中对应的异常曲线标识图,或者是,呈现出该异常日志所在的日志集合中各个目标数据项的异常标识曲线图等。

[0135] 当然,此处以异常显示界面的一种情况为例,对于前面提到的异常显示界面的其他情况也同样适用于本实施例。

[0136] S706,针对每个日志集合,从存储的日志样本数据对应的多个日志样本集合中,确定与该日志集合对应的数据组成特征匹配的目标日志样本集合。

[0137] 其中,日志样本数据为在软件产品不存在运行异常的情况下,获得到的日志数据。相应的,日志样本数据中包括多个无异常数据的日志样本,且日志样本数据中多个日志样本被聚类到多个日志样本集合。

[0138] 如,可以预先基于日志样本中数据的数据组成特征对日志样本数据中的多个日志样本进行聚类,得到多个日志样本集合以及每个日志样本集合对应的数据组成特征。在此基础上,针对聚类出的日志集合,可以基于日志集合对应的数据组成特征,查询对应相同数据组成特征的日志样本集合。

[0139] 为了便于区分,将与日志集合的数据组成特征匹配的日志样本集合称为目标日志样本集合。

[0140] S707,针对每个日志集合,确定该日志集合在日志数据中的第一日志出现频率以及该日志集合对应的目标日志样本集合在日志样本数据中的第二日志出现频率,如果基于该第一日志出现频率和第二日志出现频率确定出该日志集合中日志的数量存在数量增多异常的情况下,将该日志集合确定为存在异常风险的风险日志集合。

[0141] 其中,第一日志出现频率为日志集合中包含的日志的第一数量与日志数据中包含的日志的第一总数量的比值。如,日志集合中包含有日志的条数为50条,而步骤S701获得到的日志数据中一共有1000条日志,那么该日志集合对应的第一日志出现频率为1/20。

[0142] 相应的,第二日志出现频率为日志集合对应的目标日志样本集合中包含日志的第二数量与该日志样本数据中包含的日志的第二总数量之间的比值。

[0143] 其中,基于该第一日志出现频率和第二日志出现频率确定出该日志集合中日志的数量存在数量增多异常的方式可以有多种。如,针对一个日志集合,如果第一日志出现频率与第二日志出现频率的比值大于设定阈值,则确定日志集合中的日志的数量存在数量增多异常的情况。又如,针对一个日志集合,如果该第一日志出现频率与第二日志出现频率的差值大于设定值,则确定该日志集合中的日志的数量存在数量增多异常的情况。

[0144] 可以理解的是,针对一个日志集合,该日志集合对应的目标日志样本集合的第二日志出现频率表示该日志集合中日志所对应的功能模式在正常情况下所能日志的出现频率,因此,如果日志集合对应的第一日志出现频率远远大于第二日志出现频率,则说明针对该功能模式出现了较多的日志,这与正常情况不同,从而说明该功能模式存在异常的可能性增大,因此,将该功能模式对应的日志集合确定为风险日志集合,有利于用户及时发现该功能模式所可能存在的风险。

[0145] S708,输出针对风险日志集合的提示信息。

[0146] 其中,针对风险日志集合的提示信息用于提示用户该风险日志集合的日志所针对的功能模式存在异常风险。如果存在多个风险日志集合,则可以针对该多个风险日志集合输出相应的提示信息。

[0147] 如,提示信息可以包括风险日志集合对应的功能模式以及该功能模式存在风险的提示信息。

[0148] 又如,作为以一种可选方式,提示信息可以包括针对该风险日志集合的规则补充提示,该规则补充提示用于提示用户针对该风险日志集合对应的数据组成特征增设异常检测规则。可以理解的是,如果日志集合属于风险日志集合,则说明该日志集合对应的功能模式存在异常风险,那么为了及时发现该种异常风险可以提示用户更为有针对性和详细的部

署这对该功能模式的异常检测规则,以便更为及时和可靠的基于功能模式对应的日志发现该功能模式所可能存在的异常。

[0149] 相应的,在获得用户针对该风险日志集合对应的数据组合特征(或者功能模式)设定的至少一种异常检测规则之后,可以存储该数据组合特征(或者功能模式)对应的至少一种异常检测规则。

[0150] 需要说明的是,在本实施例中在聚类出各个日志集合并生成相应的数据表之后,是以并行执行基于日志集合对应的异常检测规则对日志集合进行异常检测以及检测日志集合是否为风险日志集合的操作为例,即,步骤S704-S705与步骤S706到S708并行执行。但是可以理解的是,在实际应用中,也可以是在对分别基于各个日志集合的异常检测规则,确定出各个日志集合中的异常日志之后,再执行步骤S706到S708;或者是,在确定出各个日志集合中的异常日志之后,针对不存在异常日志的日志集合可以再执行步骤S706到S708,对此不加限制。

[0151] 可以理解的是,在本申请实施例中,可能会同时获得多份日志数据,例如,从不同应用客户端分别获得日志数据,从而得到来源自不同应用客户端的多份日志数据。在获得多份日志数据的情况下,可以依次按照本申请如上任意一个实施例的方式依次处理各份日志数据即可。

[0152] 特别的,在获得多份日志数据,并分析出每份日志中各个日志集合内存在的异常日志之后,还可以针对该多份日志数据生成异常汇总报表。其中,该异常汇总报表展现出每份日志数据存在的异常情况,其中,每份日志数据中存在的异常情况是该日志数据聚类出的各个日志集合中的异常日志所对应的异常情况的汇总。

[0153] 如,参见图8,其示出了本申请中异常汇总报表的一种示意图。

[0154] 在图8的异常汇总报表中各个字段代表了可能存在的各种异常情况。如,异常情况包括:高CPU使用率、低内存以及发送/接收码率高等等情况。

[0155] 在图8的异常汇总报表中每一行表示一份日志数据对应的异常汇总情况,其中,日志数据可以通过相应的文件名称等标识,为了便于描述和较为直观,在图8中采用日志数据1、日志数据2等等来表示不同的日志。

[0156] 其中,如果该日志数据中存在某个字段表示的异常情况,则该日志数据所在行对应该字段的取值为1,否则为0。例如,日志数据1中存在高CPU使用率,但是不存在低内存的异常情况,则异常汇总报表中日志数据1对应“高内存使用率”这一字段的取值为1,而对应“低内存”这一字段的取值为0。

[0157] 可以理解的是,在完成多份日志数据的分析并生成该异常汇总报表之后,可以直接显示该异常汇总报表;也可以是,在检测到用户输入展现汇总报表的指令后,显示该异常汇总报表。

[0158] 可以理解的是,在以上实施例中均以计算机或者云平台等自动完成对日志数据进行异常日志检测为例说明。但是可以理解的是,在实际应用中,通过对日志数据中各条日志进行聚类,也可以实现辅助用户更为有效和高效确定异常日志。

[0159] 具体的,在聚类得到多个日志集合之后,如果确定出需要进入人工检测模式,则可以显示日志展现界面,该日志展现界面显示有日志数据中的多条日志。在此基础上,响应于该日志展现界面检测到日志选择操作,确定该日志选择操作所选择的目标日志,并确定该

目标日志所属的目标日志集合,从而显示该目标日志集合中各条日志的数据。

[0160] 由于通过聚类将属于同一功能模式的日志聚类到同一日志集合,那么在用户需要查看某条日志时,将与该日志属于同一功能模式的日志的数据展现出来,以便用户同时查看到该功能模式下的多条日志,从而更为便捷和高效的分析该功能模式是否存在异常。

[0161] 其中,显示该目标日志集合中各条日志的数据的具体方式可以有多种,下面以一种情况为例说明。如,参见图9,其示出了本申请一种确定异常日志的方法又一种流程示意图,本实施例的方法可以包括:

[0162] S901,获取待分析的一份日志数据,该日志数据包括多条日志。

[0163] S902,按照日志中数据的数据组成特征,对日志数据中的多条日志进行聚类,得到聚类出的多个日志集合以及每个日志集合对应的数据组成特征。

[0164] S903,针对每个日志集合,将该日志集合中各日志具有的至少一种数据项作为待构建的数据表中的字段,并基于该日志集合中每条日志中该至少一种数据项各自数值,构建数据表。

[0165] 作为一种可选方式,构建日志集合的数据表之后,还可以针对数据表中每个字段表示的数据项,生成该数据项的数值变化曲线图。其中,生成数据项的数值变化曲线的过程可以参见前面的相关介绍,在此不再赘述。在此基础上,后续需要展现某个数据项的数值变化曲线图时,则可以直接显示该数据项的数值变化曲线图;或者是,在需要生成该数据项的常标识曲线图时,可以直接在该数据项的数值变化曲线图上进行构建。

[0166] S904,确定设定的检测模式是否为自动检测模式,如果是,则执行步骤S905;如果否,则执行步骤S909;

[0167] 其中,在该步骤S904可以确定设定的检测模式,该检测模式为人工检测模式和自动检测模式中的一种,如果设定的检测模式为自动检测模型,则执行步骤S905到S908的操作;否则,则执行步骤S909,以辅助人工检测。

[0168] 可以理解的是,该检测模式可以是在聚类出各个日志集合之后,提示用户输入或者选择的检测模式;也可以是在获得日志数据之后,显示模式展现界面,用户可以在该显示模式展现界面中选择或者输入用户希望的检测模式,相应的,响应于该显示模式展现界面中的模式选择或者输入操作,确定设定的检测模式。

[0169] S905,针对每个日志集合,确定与日志集合对应的数据组成特征匹配的至少一种异常检测规则,并基于该至少一种异常检测规则,确定日志集合的数据表中存在异常的异常记录以及异常记录中的异常数据。

[0170] S906,针对每个日志集合,在该日志集合中存在异常日志的情况下,确定日志集合中各异常日志中的异常数据所归属的至少一种目标数据项。

[0171] S907,针对日志集合对应的每种目标数据项,构建目标数据项的异常标识曲线图。

[0172] 其中,异常标识曲线图为日志集合的各日志中目标数据项的数值的数值变化曲线图,且在数值变化曲线图中标示有目标数据项中的异常数据的信息。

[0173] 如,可以针对日志集合的该目标数据项,先构建该目标数据项的数值变化曲线图,或者是,在S903之后已构建出目标数据项的数值变化曲线图,则直接获得该目标数据项的数值变化曲线图。然后,基于该目标数据项的数值变化曲线图中构建出异常标识曲线图。

[0174] S908,显示出日志集合中各目标数据项的异常标识曲线图。

[0175] S909,在设定的检测模式为人工检测模式的情况下,显示日志展现界面。

[0176] 其中,该日志展现界面显示有日志数据中的多条日志。

[0177] 其中,在日志展现界面中还可以将日志的标识采用设定显示方式显示,该设定显示方式可以包括字体加粗或者亮度增强等方式,以突出该日志的标识,使得用户直观了解到可以点击该日志,以查看该日志相关信息。

[0178] 可选的,在该日志展现界面中显示的日志中还可以标示出该日志中各个数据项的取值,以便于用户直观的看到日志中涉及到的各个数据项的数据。如,可以采用区别于日志中其他数据的不同颜色或者不同亮度来显示日志中数据项的数据等。

[0179] S910,响应于日志展现界面检测到日志选择操作,确定日志选择操作所选择的目标日志。

[0180] 其中,日志选择操作可以为日志展现界面中的日志被点击,或者是其他触发选择日志的操作等。

[0181] 为了便于区分,将日志选择操作所触发选择的日志称为目标日志。

[0182] 如,在日志展现界面通过特定显示方式显示日志的标识时,那么可以响应于日志标识被点击的操作,将该被点击的日志标识对应的日志确定为目标日志。

[0183] S911,确定目标日志所属的目标日志集合,并分别确定该目标日志集合对应的各种数据项各自的数值变化曲线。

[0184] 其中,目标日志所被聚类到的日志集合就是该目标日志所属的目标日志集合。

[0185] 其中,目标日志集合对应的数据项为目标日志集合的日志中包含的数据项,每种数据项的数值变化曲线为基于所述目标日志集合中各条日志中该数据项的数值构建出的数值变化曲线。

[0186] 可以理解的是,如果聚类出日志集合之后,仅仅是生成了日志集合的数据表,而未基于数据表,为日志集合涉及到的各个数据项分别生成数值变化曲线,则该步骤S911可以为基于目标日志集合对应的数据表,分别构建目标日志集合涉及到的各个数据项各自对应的数值变化曲线。构建数据项的数值变化曲线的过程可以参见前面的相关介绍。

[0187] 如果在进入人工检测模式之前已经生成了该目标日志集合中各个数据项对应的数值变化曲线,则可以直接查询并获取相应的数值变化曲线即可。

[0188] S912,显示曲线图展现界面,该曲线图展现界面展现有目标日志集合对应的各种数据项的数值变化曲线。

[0189] 其中,曲线图展现界面展现出的每个数据项的数值变化曲线均可以参见图4所示,在此不再赘述。

[0190] 在本实施例中,如果设定的检测模式为人工检测模式,那么本申请可以显示日志展现界面,在此基础上,如果用户点击日志展现界面中的某条日志,则会确定该日志被聚类到的日志集合,并显示出日志集合的各日志所涉及到的数据项对应的数值变化曲线,使得用户可以直观看到该日志对应的功能模式下记录的每个数据项在不同日志中的取值变化情况,从而有利于用户更为高效的发现异常数据,从而高效定位异常日志。

[0191] 对应本申请的一种确定异常日志的方法,本申请还提供了一种确定异常日志的装置。如图10所示,其示出了本申请一种确定异常日志的装置一个实施例的组成结构示意图,本实施例的装置可以包括:

[0192] 日志获取单元1001,用于获取待分析的日志数据,该日志数据包括多条日志;

[0193] 日志聚类单元1002,用于按照日志中数据的数据组成特征,对该日志数据中的多条日志进行聚类,得到聚类出的多个日志集合以及每个日志集合对应的数据组成特征,日志集合包括至少一条日志;

[0194] 规则匹配单元1003,用于针对每个日志集合,确定与该日志集合对应的数据组成特征匹配的至少一种异常检测规则;

[0195] 异常确定单元1004,用于针对每个日志集合,利用该日志集合对应的至少一种异常检测规则,确定该日志集合中存在异常的异常日志以及异常日志中的异常数据。

[0196] 在一种可能的实现方式中,该日志聚类单元,包括:

[0197] 日志聚类子单元,用于按照日志中包含的功能标识和字符特征,对日志数据中的多条日志进行聚类,其中,日志中的功能标识用于表征日志中记录的运行状态信息所归属的功能模式的类型。

[0198] 在又一种可能的实现方式中,该装置还可以包括:

[0199] 数据项确定单元,用于在日志集合中存在异常日志的情况下,确定日志集合中各异常日志中的异常数据所归属的至少一种目标数据项,该目标数据项属于日志集合的各日志中具有的至少一种数据项;

[0200] 异常曲线构建单元,用于针对日志集合对应的每种目标数据项,构建该目标数据项的异常标识曲线图,该异常标识曲线图包括日志集合的各日志中该目标数据项的数值对应的数值变化曲线图,且在该数值变化曲线图中标示有目标数据项中异常数据的信息;

[0201] 曲线显示单元,用于显示出日志集合中各目标数据项的异常标识曲线图。

[0202] 在又一种可能的实现方式中,本申请实施例的装置还可以包括:

[0203] 样本集匹配单元,用于针对每个日志集合,从存储的日志样本数据对应的多个日志样本集合中,确定与日志集合对应的数据组成特征匹配的目标日志样本集合,日志样本数据包括多个无异常数据的日志样本,且,多个日志样本集合为基于日志样本中数据的数据组成特征对日志样本数据的多个日志样本聚类得到的;

[0204] 风险日志集确定单元,用于针对每个日志集合,确定日志集合在日志数据中的第一日志出现频率以及日志集合对应的目标日志样本集合在日志样本数据中的第二日志出现频率,并在基于第一日志出现频率和第二日志出现频率确定出日志集合中日志的数量存在数量增多异常的情况下,将日志集合确定为存在异常风险的风险日志集合,其中,第一日志出现频率为日志集合中包含的日志的第一数量与日志数据中包含的日志的第一总数量的比值,第二日志出现频率为日志集合对应的目标日志样本集合中包含日志的第二数量与日志样本数据中包含的日志的第二总数量之间的比值;

[0205] 风险提示单元,用于输出针对风险日志集合的提示信息。

[0206] 可选的,该风险提示单元,包括:

[0207] 规则补充提示单元,用于输出针对该风险日志集合的规则补充提示,该规则补充提示用于提示用户针对风险日志集合对应的数据组成特征增设异常检测规则。

[0208] 在又一种可能的实现方式中,该装置还可以包括:

[0209] 模式确定单元,用于在规则匹配单元确定与日志集合对应的数据组成特征匹配的至少一种异常检测规则之前,确定设定的检测模式,该检测模式为人工检测模式和自动检

测模式中的一种；

[0210] 规则匹配单元具体为，用于在设定的检测模式为自动检测模式的情况下，针对每个日志集合，确定与该日志集合对应的数据组成特征匹配的至少一种异常检测规则；

[0211] 该装置还可以包括：

[0212] 日志展现单元，用于在设定的检测模式为人工检测模式的情况下，显示日志展现界面，该日志展现界面显示有该日志数据中的多条日志；

[0213] 日志选择单元，用于响应于该日志展现界面检测到日志选择操作，确定该日志选择操作所选择的目标日志；

[0214] 集合确定单元，用于确定目标日志所属的目标日志集合；

[0215] 数据显示单元，用于显示该目标日志集合中各条日志的数据。

[0216] 可选的，该装置还可以包括：

[0217] 曲线确定单元，用于在数据显示单元显示目标日志集合中各条日志的数据之前，分别确定目标日志集合对应的各种数据项各自的数值变化曲线，目标日志集合对应的数据项为目标日志集合的日志中包含的数据项，每种数据项的数值变化曲线为基于目标日志集合的各条日志中该数据项的数值，构建出的数值变化曲线；

[0218] 数据显示单元具体为，用于显示曲线图展现界面，所述曲线图展现界面展现有所述目标日志集合对应的各种数据项的数值变化曲线。

[0219] 又一方面，本申请还提供了一种计算机设备，该计算机设备可以为个人计算机、服务器或者云平台中的节点等等。如图11，其示出了本申请提供的计算机设备的一种组成架构示意图。在图11中，该计算机设备1100可以包括：处理器1101和存储器1102。

[0220] 可选的，该计算机设备还可以包括：通信接口1103、输入单元1104和显示器1105和通信总线1106。

[0221] 其中，处理器1101、存储器1102、通信接口1103、输入单元1104和显示器1105均通过通信总线1106完成相互间的通信。

[0222] 在本申请实施例中，该处理器1101，可以为中央处理器，特定应用集成电路等。

[0223] 该处理器可以调用存储器1102中存储的程序，具体的，处理器可以执行以上实施例中云计算机设备侧所执行的操作。

[0224] 存储器1102中用于存放一个或者一个以上程序，程序可以包括程序代码，所述程序代码包括计算机操作指令，在本申请实施例中，该存储器中至少存储有用于实现以上任意一个实施例中的确定异常日志的方法的程序。

[0225] 在一种可能的实现方式中，该存储器1102可包括存储程序区和存储数据区，其中，存储程序区可存储操作系统、以上所提到的程序，以及图像播放等功能所需的应用程序等；存储数据区可存储根据计算机设备的使用过程中所创建的数据。

[0226] 该通信接口1103可以为通信模块的接口。

[0227] 本申请还可以包括输入单元1104，该输入单元可以包括触摸感应单元、键盘等等。

[0228] 该显示器1105包括显示面板，如触摸显示面板等。

[0229] 当然，图11所示的计算机设备结构并不构成对本申请实施例中计算机设备的限定，在实际应用中计算机设备可以包括比图11所示的更多或更少的部件，或者组合某些部件。

[0230] 另一方面,本申请还提供了一种存储介质,该存储介质中存储有计算机可执行指令,所述计算机可执行指令被处理器加载并执行时,实现如上任意一个实施例中的确定异常日志的方法。

[0231] 需要说明的是,本说明书中的各个实施例均采用递进的方式描述,每个实施例重点说明的都是与其他实施例的不同之处,各个实施例之间相同相似的部分互相参见即可,且不同实施例可以相互结合。对于装置类实施例而言,由于其与方法实施例基本相似,所以描述的比较简单,相关之处参见方法实施例的部分说明即可。

[0232] 最后,还需要说明的是,在本文中,诸如第一和第二等之类的关系术语仅仅用来将一个实体或者操作与另一个实体或操作区分开来,而不一定要求或者暗示这些实体或操作之间存在任何这种实际的关系或者顺序。

[0233] 对所公开的实施例的上述说明,使本领域技术人员能够实现或使用本发明。对这些实施例的多种修改对本领域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发明的精神或范围的情况下,在其它实施例中实现。因此,本发明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。

[0234] 以上仅是本发明的优选实施方式,应当指出,对于本技术领域的普通技术人员来说,在不脱离本发明原理的前提下,还可以做出若干改进和润饰,这些改进和润饰也应视为本发明的保护范围。

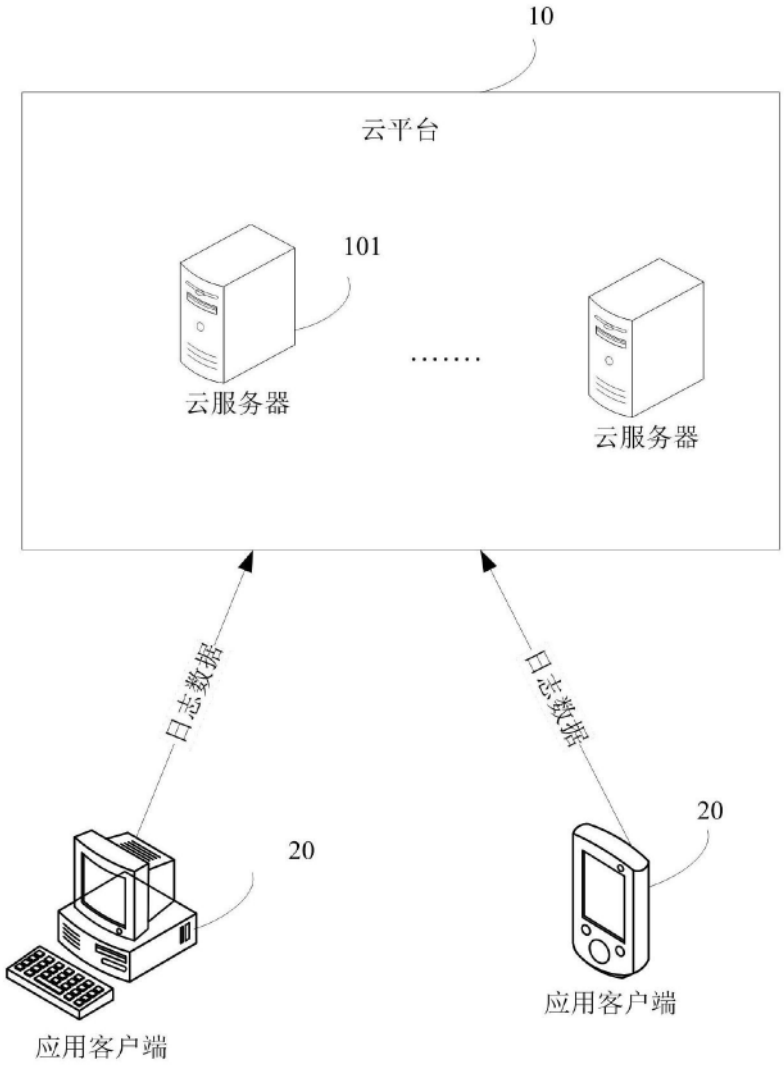


图1

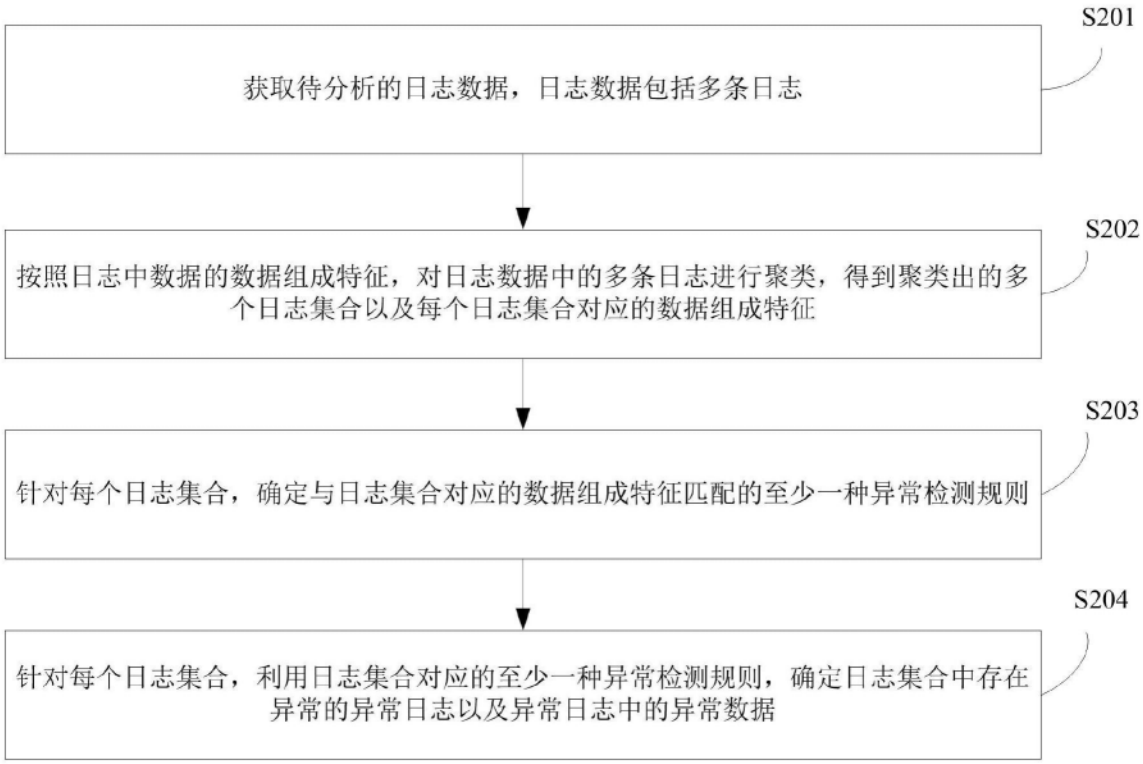

S201
获取待分析的日志数据，日志数据包括多条日志
S202
按照日志中数据的数据组成特征，对日志数据中的多条日志进行聚类，得到聚类出的多个日志集合以及每个日志集合对应的数据组成特征
S203
针对每个日志集合，确定与日志集合对应的数据组成特征匹配的至少一种异常检测规则
S204
针对每个日志集合，利用日志集合对应的至少一种异常检测规则，确定日志集合中存在异常的异常日志以及异常日志中的异常数据

图2

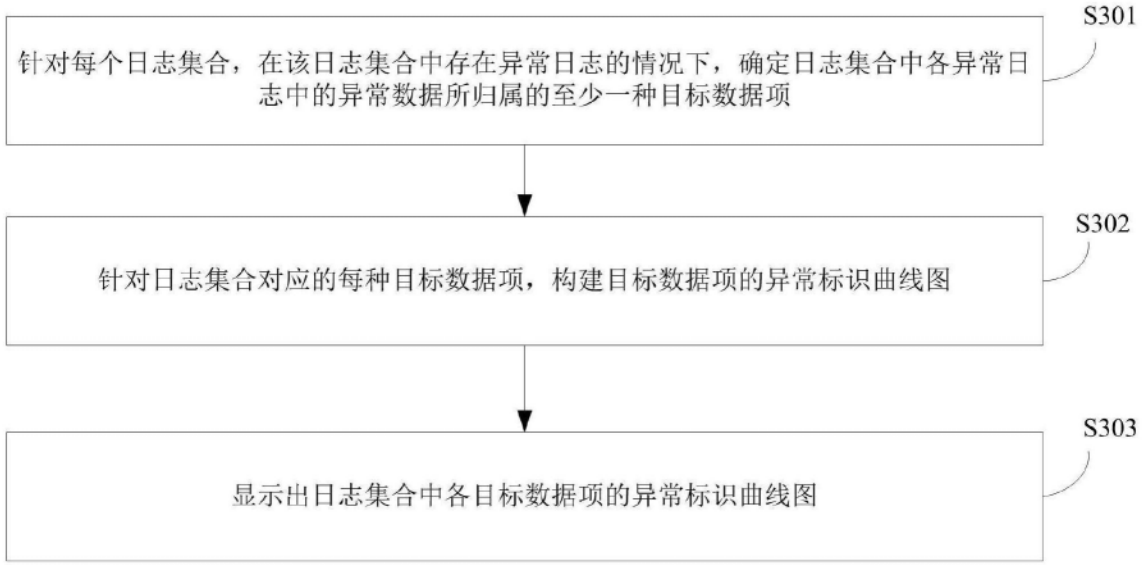

S301
针对每个日志集合，在该日志集合中存在异常日志的情况下，确定日志集合中各异常日志中的异常数据所归属的至少一种目标数据项
S302
针对日志集合对应的每种目标数据项，构建目标数据项的异常标识曲线图
S303
显示出日志集合中各目标数据项的异常标识曲线图

图3

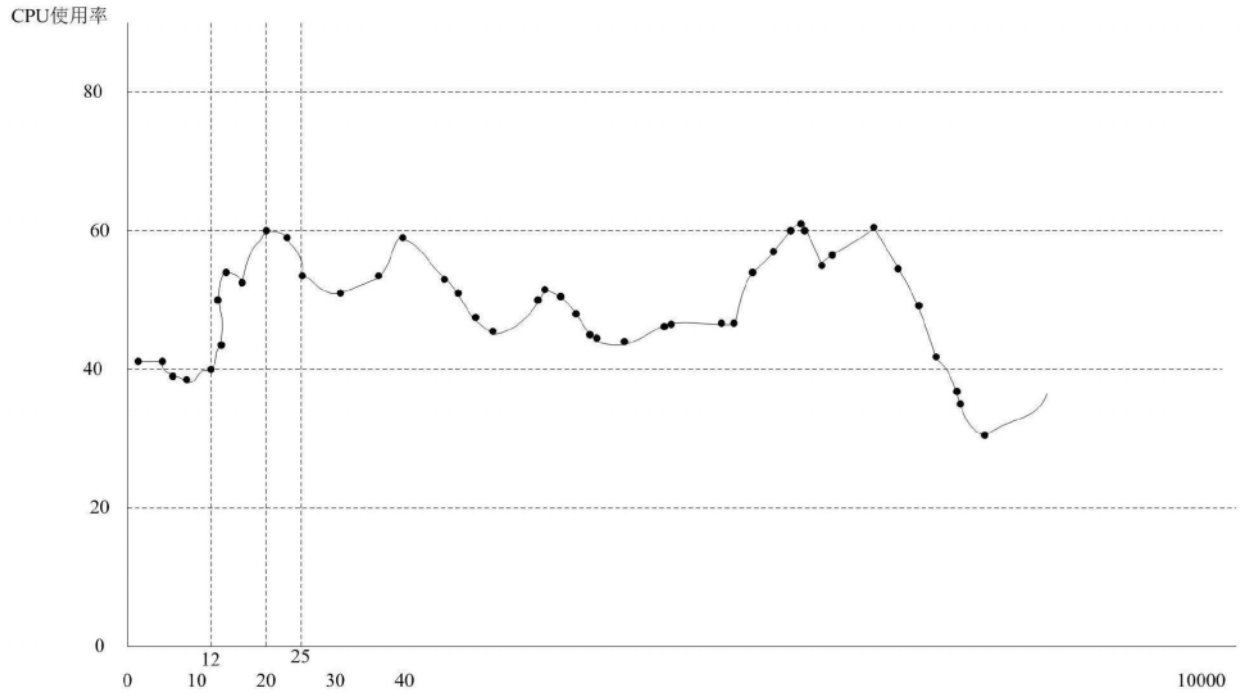

CPU使用率
80
60
40
20
0
12
25
0
10
20
30
40
10000

图4

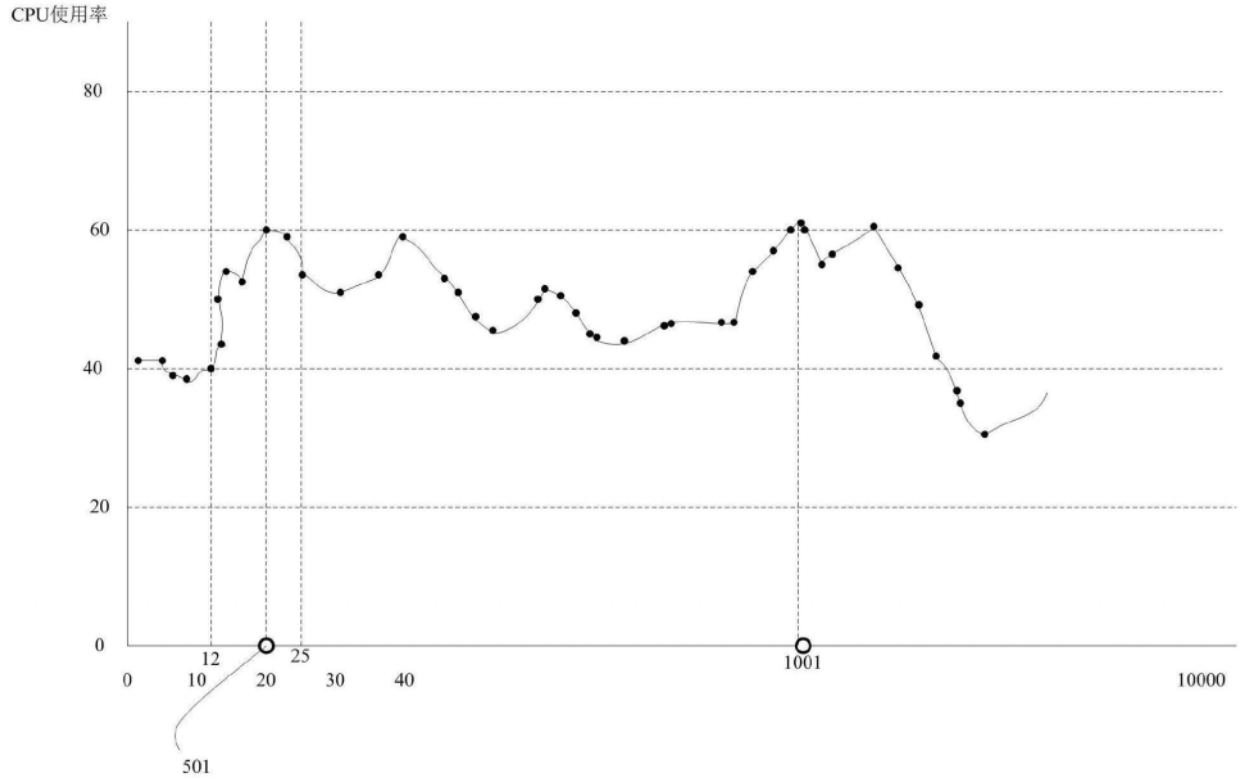

CPU使用率
80
60
40
20
0
12
25
1001
0
10
20
30
40
10000
501

图5

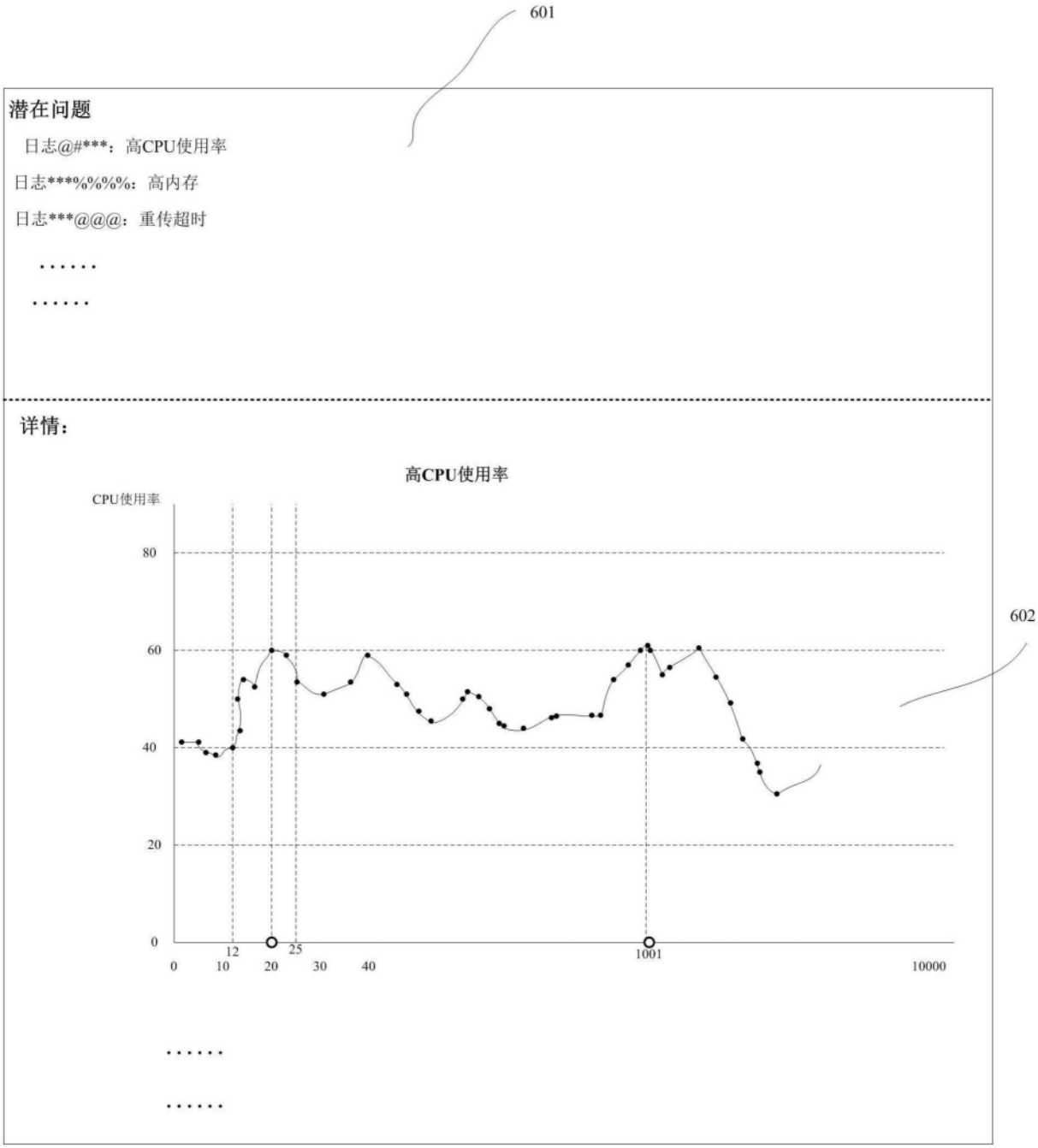

601
潜在问题
日志@#***：高CPU使用率
日志***%%%%：高内存
日志***@@@：重传超时
.....
.....
详情：
高CPU使用率
CPU使用率
80
60
40
20
0
12
25
1001
0
10
20
30
40
10000
602
.....
.....

图6

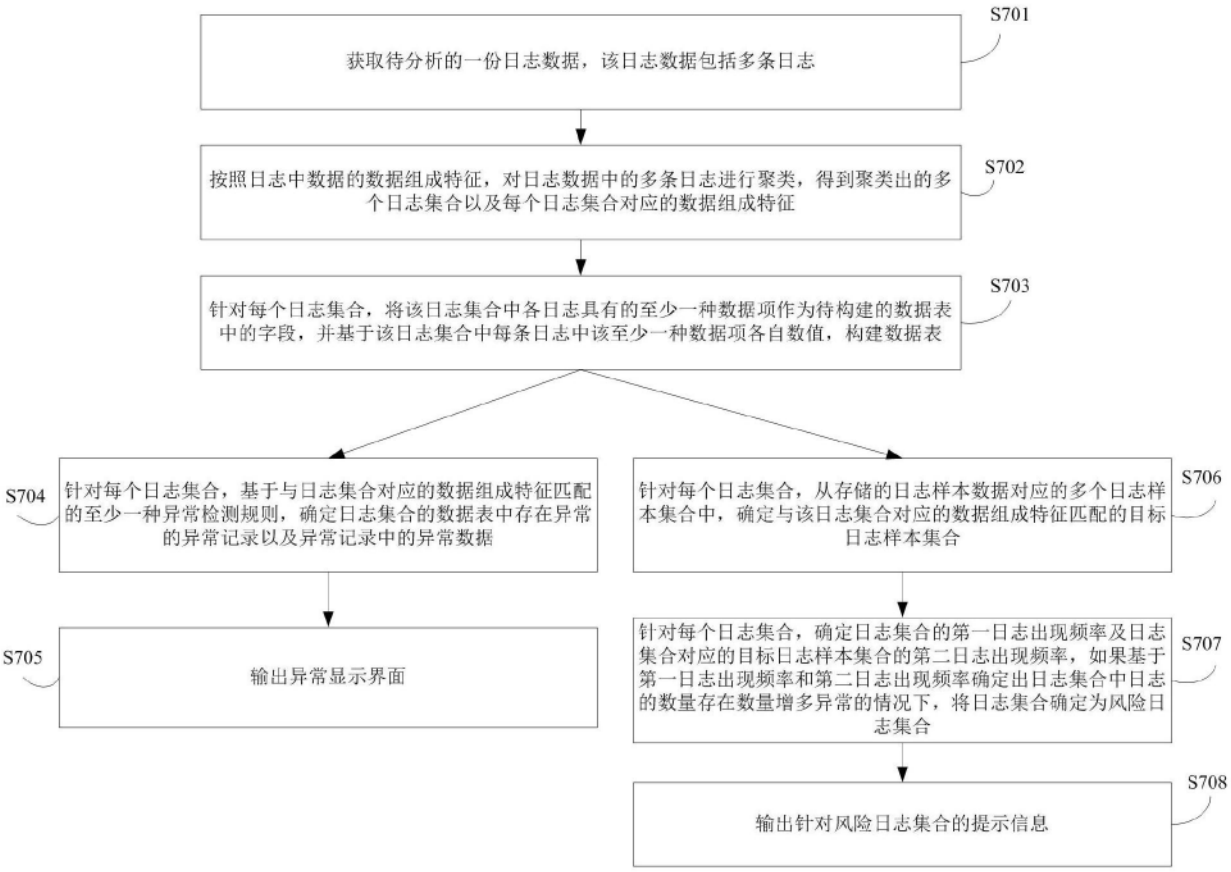


图7

日志数据	高CPU使用率	低内存	……	发送/接收码率高
日志数据1	1	0	…….	1
日志数据2	0	1	…….	1
…….	…….	…….	…….	…….
日志数据n	0	1	…….	0

图8

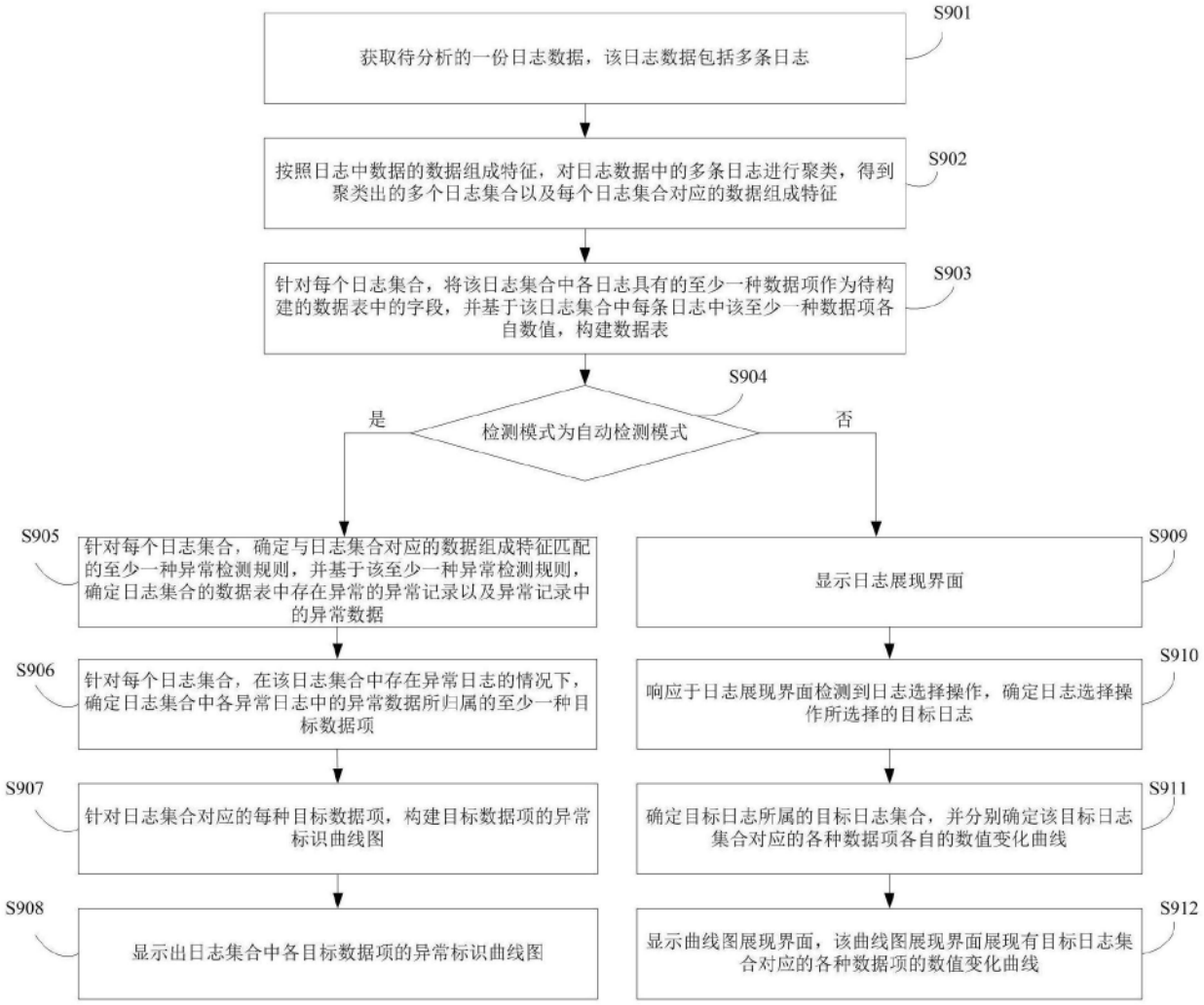

S901
获取待分析的一份日志数据，该日志数据包括多条日志
S902
按照日志中数据的数据组成特征，对日志数据中的多条日志进行聚类，得到聚类出的多个日志集合以及每个日志集合对应的数据组成特征
S903
针对每个日志集合，将该日志集合中各日志具有的至少一种数据项作为待构建的数据表中的字段，并基于该日志集合中每条日志中该至少一种数据项各自数值，构建数据表
S904
检测模式为自动检测模式
是
否
S905
针对每个日志集合，确定与日志集合对应的数据组成特征匹配的至少一种异常检测规则，并基于该至少一种异常检测规则，确定日志集合的数据表中存在异常的异常记录以及异常记录中的异常数据
S906
针对每个日志集合，在该日志集合中存在异常日志的情况下，确定日志集合中各异常日志中的异常数据所归属的至少一种目标数据项
S907
针对日志集合对应的每种目标数据项，构建目标数据项的异常标识曲线图
S908
显示出日志集合中各目标数据项的异常标识曲线图
S909
显示日志展现界面
S910
响应于日志展现界面检测到日志选择操作，确定日志选择操作所选择的目标日志
S911
确定目标日志所属的目标日志集合，并分别确定该目标日志集合对应的各种数据项各自的数值变化曲线
S912
显示曲线图展现界面，该曲线图展现界面展现有目标日志集合对应的各种数据项的数值变化曲线

图9


1001
日志获取单元
1002
日志聚类单元
1003
规则匹配单元
1004
异常确定单元

图10

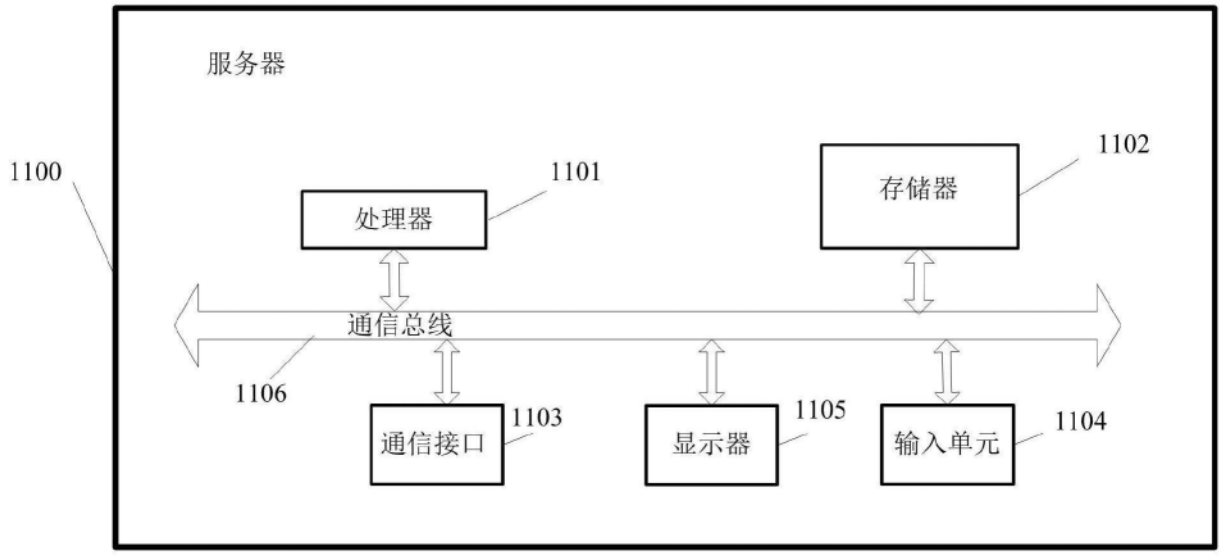

服务器
1100
处理器
1101
存储器
1102
通信总线
1106
通信接口
1103
显示器
1105
输入单元
1104

图11