



(12) 发明专利

(10) 授权公告号 CN 114500318 B

(45) 授权公告日 2024. 02. 06

(21) 申请号 202210127498.2

H04L 43/16 (2022.01)

(22) 申请日 2022.02.11

H04L 43/50 (2022.01)

H04L 41/0631 (2022.01)

(65) 同一申请的已公布的文献号

申请公布号 CN 114500318 A

(56) 对比文件

(43) 申请公布日 2022.05.13

CN 103684835 A, 2014.03.26

CN 107291533 A, 2017.10.24

(73) 专利权人 中国工商银行股份有限公司

CN 112615883 A, 2021.04.06

CN 112787939 A, 2021.05.11

地址 100140 北京市西城区复兴门内大街  
55号

CN 113347249 A, 2021.09.03

CN 113918525 A, 2022.01.11

(72) 发明人 曹璘 王亚军 陈万红 邱枫

WO 2021022710 A1, 2021.02.11

(74) 专利代理机构 中科专利商标代理有限责任  
公司 11021

审查员 罗璨

专利代理师 张博

(51) Int. Cl.

H04L 43/08 (2022.01)

H04L 43/10 (2022.01)

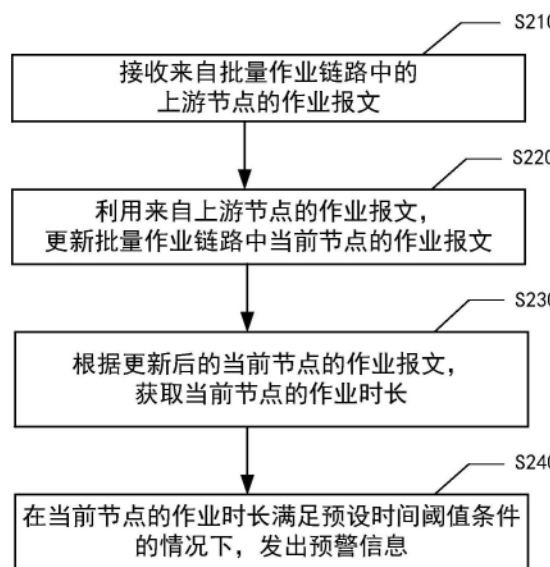
权利要求书2页 说明书12页 附图8页

(54) 发明名称

一种批量作业监控方法及装置、设备及介质

(57) 摘要

本公开提供了一种批量作业监控方法,应用于金融、计算机技术领域,该方法包括:接收来自批量作业链路中的上游节点的作业报文;利用来自上游节点的作业报文,更新批量作业链路中当前节点的作业报文,其中,当前节点在批量作业链路中位于上游节点的下游;根据更新后的当前节点的作业报文,获取当前节点的作业时长;在当前节点的作业时长满足预设时间阈值条件的情况下,发出预警信息。本公开还提供了一种批量作业监控装置、设备、存储介质和程序产品。



1. 一种批量作业监控方法,包括:

接收来自批量作业链路中的上游节点的作业报文,其中,所述作业报文包括累计作业时长;

利用来自所述上游节点的作业报文,更新所述批量作业链路中当前节点的作业报文,其中,所述当前节点在所述批量作业链路中位于所述上游节点的下游;

其中,所述利用来自所述上游节点的作业报文,更新所述批量作业链路中当前节点的作业报文包括:

更新所述当前节点的累计作业时长,包括:

通过所述当前节点的历史日志文件获取历史作业处理时长;

通过局部加权线性回归算法处理所述历史作业处理时长和所述当前节点的作业文件大小,确定所述当前节点的处理时长;

通过所述当前节点获取所述批量作业链路中各个上游节点的累计作业时长并获得上游节点累计作业时长的最大值;

通过所述当前节点的处理时长和所述上游节点累计作业时长的最大值进行运算,更新所述当前节点的累计作业时长;

根据更新后的所述当前节点的作业报文,获取所述当前节点的作业时长;

在所述当前节点的作业时长满足预设时间阈值条件的情况下,发出预警信息。

2. 根据权利要求1所述的方法,其中,所述作业报文还包括批量作业工程名称、作业节点名称、作业文件大小和报错节点名称;

其中,所述作业文件大小表示作业节点待处理的文件大小;

其中,所述累计作业时长表示所述批量作业链路累计的作业时间;

其中,所述报错节点名称用于记录所述批量作业链路中发生作业中断的节点名称。

3. 根据权利要求2所述的方法,其中,所述利用来自所述上游节点的作业报文,更新所述批量作业链路中当前节点的作业报文还包括以下至少之一:更新所述当前节点的作业文件大小以及更新所述当前节点的报错节点名称。

4. 根据权利要求3所述的方法,其中,所述更新所述当前节点的作业文件大小包括:

通过所述当前节点获取所述批量作业链路中各个上游节点的作业文件大小并进行累加,获得作业文件大小累加值;

通过所述当前节点的历史日志文件获取转发作业文件大小,其中,所述历史日志文件用于记录所述当前节点的历史作业信息,所述历史日志文件包括历史作业文件总数、历史作业处理时长以及发送到下游节点的转发作业文件大小;

利用局部加权线性回归算法处理所述作业文件大小累加值和所述转发作业文件大小,更新所述当前节点的作业文件大小。

5. 根据权利要求3所述的方法,其中,所述更新所述当前节点的报错节点名称包括:

利用所述当前节点实时检测所述批量作业链路中的所述上游节点的作业状态;

在所述上游节点发生作业中断的情况下,将所述当前节点的报错节点名称更新为发生作业中断的上游节点的名称;

通过所述当前节点利用作业报文将报错节点名称在批量作业链路中向下游节点传播;

在所述批量作业链路中发生作业中断的节点恢复的情况下,将所述当前节点的作业报

文中的报错节点名称重置为空。

6. 根据权利要求2所述的方法,其中,所述在所述当前节点的作业时长满足预设时间阈值条件的情况下,发出预警信息,包括:

在所述批量作业链路中上游节点的累计作业时长的最大值超过所述当前节点的预设开始作业时间阈值的情况下,通过所述当前节点发出第一预警信息;

在所述当前节点的作业处理时长超过所述当前节点预设的作业处理时间阈值的情况下,通过所述当前节点发出第二预警信息。

7. 根据权利要求1所述的方法,还包括:

将所述更新后的当前节点的作业报文发送到所述批量作业链路中的下游节点。

8. 一种用于实现如权利要求1-7中任一项所述的批量作业监控方法的批量作业监控装置,包括:

接收模块,用于接收来自批量作业链路中的上游节点的作业报文,其中,所述作业报文包括累计作业时长;

更新模块,用于利用来自所述上游节点的作业报文,更新所述批量作业链路中当前节点的作业报文,其中,所述当前节点在所述批量作业链路中位于所述上游节点的下游;

其中,所述利用来自所述上游节点的作业报文,更新所述批量作业链路中当前节点的作业报文包括:

更新所述当前节点的累计作业时长,包括:

通过所述当前节点的历史日志文件获取历史作业处理时长;

通过局部加权线性回归算法处理所述历史作业处理时长和所述当前节点的作业文件大小,确定所述当前节点的处理时长;

通过所述当前节点获取所述批量作业链路中各个上游节点的累计作业时长并获得上游节点累计作业时长的最大值;

通过所述当前节点的处理时长和所述上游节点累计作业时长的最大值进行运算,更新所述当前节点的累计作业时长;

获取模块,用于根据更新后的所述当前节点的作业报文,获取所述当前节点的作业时长;

预警模块,用于在所述当前节点的作业时长满足预设时间阈值条件的情况下,发出预警信息。

9. 一种电子设备,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,

其中,当所述一个或多个程序被所述一个或多个处理器执行时,使得所述一个或多个处理器执行根据权利要求1~7中任一项所述的方法。

10. 一种计算机可读存储介质,其上存储有可执行指令,该指令被处理器执行时使处理器执行根据权利要求1~7中任一项所述的方法。

## 一种批量作业监控方法及装置、设备及介质

### 技术领域

[0001] 本公开涉及金融、计算机技术领域,具体涉及一种批量作业监控方法、装置、设备、介质和程序产品。

### 背景技术

[0002] 随着银行等金融行业信息化程度的不断提高,用户数量的不断增加,后台服务器产生的数据量大幅增长。此外,随着银行等信息科技部门后台架构的不断升级调整,应用系统间集成和信息交互的复杂度不断提升,大量的应用系统之间通过批量作业完成数据的交互同步,这对批量作业的性能处理需求提出了更高的要求。批量作业在传输中,因为网络中断,传输文件大小,批量节点服务器性能处理能力,批量文件格式等原因,会造成批量作业的延迟,甚至中断,影响应用系统间的批量文件传输和信息交互。目前的技术方案中,作业节点无法感知批量作业链路中上下游节点作业状态,影响了批量作业效率,并推高了运维成本。

### 发明内容

[0003] 鉴于上述问题,本公开提供了一种基于报文的批量作业监控方法、装置、设备、介质和程序产品。

[0004] 本公开的第一个方面,提供了一种批量作业监控方法,包括:

[0005] 接收来自批量作业链路中的上游节点的作业报文;

[0006] 利用来自上游节点的作业报文,更新批量作业链路中当前节点的作业报文,其中,当前节点在批量作业链路中位于上游节点的下游;

[0007] 根据更新后的当前节点的作业报文,获取当前节点的作业时长;

[0008] 在当前节点的作业时长满足预设时间阈值条件的情况下,发出预警信息。

[0009] 根据本公开的实施例,上述作业报文包括批量作业工程名称、作业节点名称、作业文件大小、累计作业时长和报错节点名称;

[0010] 其中,作业文件大小表示作业节点待处理的文件大小;

[0011] 其中,累计作业时长表示批量作业链路累计的作业时间;

[0012] 其中,报错节点名称用于记录批量作业链路中发生作业中断的节点名称。

[0013] 根据本公开的实施例,上述利用来自上游节点的作业报文,更新批量作业链路中当前节点的作业报文包括以下至少之一:更新当前节点的作业文件大小、更新当前节点的累计作业时长以及更新当前节点的报错节点名称。

[0014] 根据本公开的实施例,上述更新当前节点的作业文件大小包括:

[0015] 通过当前节点获取批量作业链路中各个上游节点的作业文件大小并进行累加,获得作业文件大小累加值;

[0016] 通过当前节点的历史日志文件获取转发作业文件大小,其中,历史日志文件用于记录当前节点的历史作业信息,历史日志文件包括历史作业文件总数、历史作业处理时长

以及发送到下游节点的转发作业文件大小；

[0017] 利用局部加权线性回归算法处理作业文件大小累加值和转发作业文件大小,更新当前节点的作业文件大小。

[0018] 根据本公开的实施例,上述更新当前节点的累计作业时长包括:

[0019] 通过当前节点的历史日志文件获取历史作业处理时长;

[0020] 通过局部加权线性回归算法处理历史作业处理时长和当前节点的作业文件大小,确定当前作业节点的处理时长;

[0021] 通过当前节点获取批量作业链路中各个上游节点的累计作业时长并获得上游节点累计作业时长的最大值;

[0022] 通过当前作业节点的处理时长和上游节点累计作业时长的最大值进行运算,更新当前节点的累计作业时长。

[0023] 根据本公开的实施例,上述更新当前节点的报错节点名称包括:

[0024] 利用当前节点实时检测批量作业链路中的上游节点的作业状态;

[0025] 在上游节点发生作业中断的情况下,将当前节点的报错节点名称更新为发生作业中断的上游节点的名称;

[0026] 通过当前节点利用作业报文将报错节点名称在批量作业链路中向下游节点传播;

[0027] 在批量作业链路中发生作业中断的节点恢复的情况下,将当前作业节点的作业报文中的报错节点名称重置为空。

[0028] 根据本公开的实施例,上述在当前节点的作业时长满足预设时间阈值条件的情况下,发出预警信息包括:

[0029] 在批量作业链路中上游节点的累计作业时长的最大值超过当前节点的预设开始作业时间阈值的情况下,通过当前节点发出第一预警信息;

[0030] 在当前节点的作业处理时长超过当前节点预设的作业处理时间阈值的情况下,通过当前节点发出第二预警信息。

[0031] 根据本公开的实施例,上述批量作业监控方法还包括:

[0032] 将更新后的当前节点的作业报文发送到批量作业链路中的下游节点。

[0033] 本公开的第二个方面,提供了一种批量作业监控装置,包括:

[0034] 接收模块,用于接收来自批量作业链路中的上游节点的作业报文;

[0035] 更新模块,用于利用来自上游节点的作业报文,更新批量作业链路中当前节点的作业报文,其中,当前节点在批量作业链路中位于上游节点的下游;

[0036] 获取模块,用于根据更新后的当前节点的作业报文,获取当前节点的作业时长;

[0037] 预警模块,用于在当前节点的作业时长满足预设时间阈值条件的情况下,发出预警信息。

[0038] 本公开的第三个方面,提供了一种电子设备,包括:

[0039] 一个或多个处理器;

[0040] 存储装置,用于存储一个或多个程序,

[0041] 其中,当一个或多个程序被一个或多个处理器执行时,使得一个或多个处理器执行上述批量作业监控方法。

[0042] 本公开的第四个方面,提供了一种计算机可读存储介质,其上存储有可执行指令,

该指令被处理器执行时使处理器执行上述批量作业监控方法。

[0043] 本公开的第五个方面,提供了一种计算机程序产品,包括计算机程序,计算机程序被处理器执行时实现上述批量作业监控方法。

[0044] 本公开所提供的基于报文的批量作业监控方法、批量作业监控装置、设备、介质和计算机程序产品,通过报文在批量作业链路中传输上下游节点的作业状态信息,提高了批量作业的效率;同时优化报文里的相关字段,可以进行提前预警,方便了运维人员进行批量作业管理。

## 附图说明

[0045] 通过以下参照附图对本公开实施例的描述,本公开的上述内容以及其他目的、特征和优点将更为清楚,在附图中:

[0046] 图1示意性示出了根据本公开实施例的批量作业监控方法的应用场景图;

[0047] 图2示意性示出了根据本公开实施例的批量作业监控方法的流程图;

[0048] 图3示意性示出了根据本公开实施例的更新当前节点的作业文件大小的流程图;

[0049] 图4示意性示出了根据本公开实施的更新当前节点的累计作业时长的流程图;

[0050] 图5示意性示出了根据本公开实施例的当前节点历史作业时长的示意图;

[0051] 图6示意性示出了根据本公开实施例的权重系数 $k_2=1$ 的拟合图;

[0052] 图7示意性示出了根据本公开实施例的权重系数 $k_2=0.01$ 的拟合图;

[0053] 图8示意性示出了根据本公开实施例的权重系数 $k_2=0.003$ 的拟合图;

[0054] 图9示意性示出了根据本公开实施例的更新当前节点的报错节点名称的流程图;

[0055] 图10示意性示出了根据本公开实施例的报错节点在批量作业链路中传递的过程图;

[0056] 图11示意性示出了根据本公开实施例的基于作业报文的批量作业流程图;

[0057] 图12示意性示出了根据本公开实施例的一种批量作业监控装置的结构框图;

[0058] 图13示意性示出了根据本公开实施例的适于实现批量作业监控方法的电子设备的方框图。

## 具体实施方式

[0059] 以下,将参照附图来描述本公开的实施例。但是应该理解,这些描述只是示例性的,而并非要限制本公开的范围。在下面的详细描述中,为便于解释,阐述了许多具体的细节以提供对本公开实施例的全面理解。然而,明显地,一个或多个实施例在没有这些具体细节的情况下也可以被实施。此外,在以下说明中,省略了对公知结构和技术的描述,以避免不必要地混淆本公开的概念。

[0060] 在此使用的术语仅仅是为了描述具体实施例,而并非意在限制本公开。在此使用的术语“包括”、“包含”等表明了所述特征、步骤、操作和/或部件的存在,但是并不排除存在或添加一个或多个其他特征、步骤、操作或部件。

[0061] 在此使用的所有术语(包括技术和科学术语)具有本领域技术人员通常所理解的含义,除非另外定义。应注意,这里使用的术语应解释为具有与本说明书的上下文相一致的含义,而不应以理想化或过于刻板的方式来解释。

[0062] 在使用类似于“A、B和C等中至少一个”这样的表述的情况下,一般来说应该按照本领域技术人员通常理解该表述的含义来予以解释(例如,“具有A、B和C中至少一个的系统”应包括但不限于单独具有A、单独具有B、单独具有C、具有A和B、具有A和C、具有B和C、和/或具有A、B、C的系统等)。

[0063] 批量作业的主要特点是:批量文件数量多、批量文件一般比较大、时效性较低、涉及多个系统应用的串行处理等。批量链路中某一个上游节点的批量作业处理时效过长,或者发生中断,往往会影响到下游所有批量链路节点的处理时效。目前的技术方案,主要根据生产经验值,在本节点配置上游批量文件到达时间阈值,和本节点批量处理时效阈值。存在的问题:批量链路中的节点只能感知本节点批量文件当前的实际处理情况,不能感知上游发生的批量异常。当上游批量文件到达时间超过了阈值,或者本节点处理时效超过阈值,本节点批量作业才会报警,通知运维人员处理,这给本节点,尤其是下游节点处理批量作业,带来了较高的时间处理成本。

[0064] 有鉴于此,本公开提供了一种批量作业监控方法、一种批量作业监控装置、设备、存储介质以及程序产品。上述批量作业监控方法,通过有效利用批量作业链路中上游节点的作业状态信息以及当前节点的历史日志文件,可以有效对当前节点的作业时长和作业文件大小进行预估,从而达到提前预警、节约运维成本的目的。

[0065] 在本公开的技术方案中,所涉及的用户个人信息的获取,存储和应用等,均符合相关法律法规的规定,采取了必要保密措施,且不违背公序良俗。

[0066] 本公开的实施例提供了一种批量作业监控方法,可以应用在包括但不限于金融领域。

[0067] 图1示意性示出了根据本公开实施例的批量作业监控方法的应用场景图。

[0068] 如图1所示,根据该实施例的应用场景100可以包括终端设备101、102、103、网络104以及服务器105、106、107。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。

[0069] 用户可以使用终端设备101、102、103通过网络104与服务器105、106、107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(仅为示例)。

[0070] 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。

[0071] 服务器105、106、107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所浏览的网站提供支持的后台管理服务器(仅为示例)。后台管理服务器可以对接收到的用户请求等数据进行分析等处理,并将处理结果(例如根据用户请求获取或生成的网页、信息、或数据等)反馈给终端设备。

[0072] 需要说明的是,本公开实施例所提供的基于报文的批量作业监控方法一般可以由服务器105、106、107执行。相应地,本公开实施例所提供的基于报文的批量作业监控系统一般可以设置于服务器105、106、107中。本公开实施例所提供的基于报文的批量作业监控方法也可以由不同于服务器105、106、107且能够与终端设备101、102、103和/或服务器105通

信的服务器或服务器集群执行。相应地,本公开实施例所提供的基于报文的批量作业监控系统也可以设置于不同于服务器105、106、107且能够与终端设备101、102、103和/或服务器105、106、107通信的服务器或服务器集群中。

[0073] 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、网络和服务器的。

[0074] 以下将基于图1描述的场景,通过图2~图11对本公开实施例的基于报文的批量作业监控方法进行详细描述。

[0075] 图2示意性示出了根据本公开实施例的批量作业监控方法的流程图。

[0076] 如图2所示,该实施例的基于报文的批量作业监控方法包括操作S210~操作S240。

[0077] 在操作S210,接收来自批量作业链路中的上游节点的作业报文。

[0078] 上述作业报文用于记录批量作业链路中各个作业节点的作业状态,通过作业报文可以有效的在批量作业链路中传递作业信息,能够提高批量作业的效率。

[0079] 在操作S220,利用来自上游节点的作业报文,更新批量作业链路中当前节点的作业报文,其中,当前节点在批量作业链路中位于上游节点的下游。

[0080] 批量作业链路中的当前节点可以根据上游节点的作业报文,及时更新当前节点的作业报文。

[0081] 在操作S230,根据更新后的当前节点的作业报文,获取当前节点的作业时长。

[0082] 作业时长例如可以包括当前节点处理作业文件的处理时长以及上游各个节点的累计作业时长的最大值。

[0083] 在操作S240,在当前节点的作业时长满足预设时间阈值条件的情况下,发出预警信息。

[0084] 通过与预设时间阈值进行比较,在当前节点的作业时长超过预设时间阈值的时候,向相关运维人员发出预警信息。

[0085] 上述实施例公开的批量作业监控方法,通过报文传在批量作业链上传递作业节点的作业状态信息,可以有效的对批量作业链中的上下游节点进行监控,大大降低了时间处理成本。

[0086] 根据本公开的实施例,上述作业报文包括批量作业工程名称、作业节点名称、作业文件大小、累计作业时长和报错节点名称;

[0087] 其中,作业文件大小表示作业节点待处理的文件大小;

[0088] 其中,累计作业时长表示批量作业链路累计的作业时间;

[0089] 其中,报错节点名称用于记录批量作业链路中发生作业中断的节点名称。

[0090] 下面结合表1对上述作业报文作进一步的说明。

[0091] 如表1所示,表1示意性的示出了批量作业链路中作业报文的格式以及作业报文中各个字段的含义。

[0092] 表1作业报文



	. batchName	..	batchNode	..	batchPacketSize	..	accumulateTime	..	errorBatchNode	.
[0093]	. 批量作业工程名称	..	作业节点名称表示当前作业节点的名称	..	作业文件大小表示当前作业节点处理的批量文件大小	..	截止当前作业节点, 预估链路累计的处理时间	..	上游批量报错节点名称	.

[0094] 其中, batchName表示批量作业工程名称, 即链路中需要添加监控的批量工程名称; batchNode表示作业节点的名称, 即向下游发送报文的节点名称, 如Node[i], Node[j]; batchPacketSize表示作业节点的作业文件大小, 即需要传向下游的批量文件总大小, 其中当前节点实时预估本节点需要传向下游的批量文件总大小; accumulateTime表示截止到当前节点批量作业链路中作业的累计时长, 在当前节点有多个上游节点的情况下, 该字段取所以上游字段的最大值; errorBatchNode表示报错节点名称, 即上游某个节点发生批量中断, 向下游传送节点名称, 下游节点依次传送此节点名; 该上游节点中断恢复后, 此字段更新为空。

[0095] 上述batchPacketSize字段在当前节点读取到上游的batchPacketSize字段的数值发生变化, 当前节点发往下游的batchPacketSize的值随历史经验值变化。batchPacketSize的值不可小于当前节点当前实际文件总大小。

[0096] 上述accumulateTime字段在当前节点接收上游各个节点心跳报文(即本公开的作业报文), 读取各个节点报文的batchPacketSize字段并累加, 根据历史经验值, 预估当前节点需要处理批量文件的时效processTime(即本公开所述的处理时长), 与接收到的上游accumulateTime的值(一般为各个上游节点的最大值)相加, 写入当前节点发往下游的accumulateTime字段, 其中, 当前节点读取到上游的batchPacketSize字段的数值发生变化, 当前节点预估处理批量文件时长随历史经验值变化, 发往下游的accumulateTime的值相应变化, accumulateTime的值不可小于本节点当前实际时间; 若当前节点批量作业发生中断, 接管重做, 则使用重做时间加上当前节点预估批量处理时间processTime, 作为accumulateTime发往下游节点。

[0097] 根据本公开的实施例, 上述利用来自上游节点的作业报文, 更新批量作业链路中当前节点的作业报文包括以下至少之一: 更新当前节点的作业文件大小、更新当前节点的累计作业时长以及更新当前节点的报错节点名称。

[0098] 通过本公开提供的作业报文传递批量作业链路中的信息, 并通过上述作业报文的更新, 能够及时有效的传递作业信息和作业节点的状态, 从而提高作业和运维效率。

[0099] 图3示意性示出了根据本公开实施例的更新当前节点的作业文件大小的流程图。

[0100] 如图3所示, 更新当前节点的作业文件大小包括操作S310~操作S330。

[0101] 在操作S310, 通过当前节点获取批量作业链路中各个上游节点的作业文件大小并

进行累加,获得作业文件大小累加值。

[0102] 当前节点通过接收到的作业报文(或者心跳报文)读取上游节点的batchPacketSize字段,并累加,预估本节点需要往下游传送的批量文件大小,写入发往下游节点的作业报文的batchPacketSize字段。

[0103] 在操作320,通过当前节点的历史日志文件获取转发作业文件大小,其中,历史日志文件用于记录当前节点的历史作业信息,历史日志文件包括历史作业文件总数、历史作业处理时长以及发送到下游节点的转发作业文件大小。

[0104] 在预估发往下游节点的作业报文的batchPacketSize字段时,当前节点可以利用本节点的作业历史日志文件来估计转发到下游的作业文件大小。

[0105] 在操作S330,利用局部加权线性回归算法处理作业文件大小累加值和转发作业文件大小,更新当前节点的作业文件大小。

[0106] 其中,局部加权线性回归算法由公式(1)和公式(2)表示:

$$[0107] \quad \hat{w}_1 = (X_1^T W_1 X_1)^{-1} X_1 W_1 y_1 \quad (1), \quad W_1(i, j) = \exp\left(\frac{x_1(i) - x_1(j)}{-2k_1^2}\right) \quad (2),$$

[0108] 其中, $X_1$ 为特征矩阵,包含样本 $x_1$ 的矩阵形式, $y_1$ 为目标变量向量, $W_1$ 表示高斯核对应的权重,是一个对角元素的权重矩阵,权重系数 $k_1$ 为用户指定,根据学习结果选择相对最优值。

[0109] 本公开实施例提供的上述作业文件大小更新方法,能够充分利用当前节点的历史日志文件,从而实现对作业文件大小更加精准的预估。

[0110] 表2是根据本公开实施例的历史日志文件表,下面结合表2对历史日志文件作进一步说明。

[0111] 表2历史日志文件表

totalSize	processTime	NodeX1	NodeX2	NodeX3	...
导入文件总大小	本节点批量处理时长	传向下游节点 NodeX1 的文件大小	传向下游节点 NodeX2 的文件大小	传向下游节点 NodeX3 的文件大小	

[0113] 如表2所示,当前节点维护历史日志文件表;其中,totalSize表示当前节点的历史作业文件总数,读取上游批量作业执行完毕的心跳报文作为有效输入,以上游Node[i], Node[j], Node[k]等各节点批量导入文件的总大小为参数,记录本节点批量作业执行时长,和需要传向下游的各个节点批量文件大小,作为历史记录值。定期使用历史记录表的数据做机器学习,使用例如局部加权线性回归(Locally Weighted Linear Regression, LWLR)的算法,对历史日志文件表中的相关字段的值进行拟合。当历史经验覆盖不到的数值,可以使用默认先验数值。

[0114] 图4示意性示出了根据本公开实施的更新当前节点的累计作业时长的流程图。

[0115] 如图4所示,当前节点更新作业报文中累积作业时长的方法包括操作S410~操作S440。

[0116] 在操作S410,通过当前节点的历史日志文件获取历史作业处理时长。

[0117] 在操作S420,通过局部加权线性回归算法处理历史作业处理时长和当前节点的作业文件大小,确定当前作业节点的处理时长。

[0118] 其中,上述局部加权线性回归算法由公式(3)和公式(4)确定,

$$[0119] \quad \hat{w}_2 = (X_2^T W_2 X_2)^{-1} X_2 W_2 y_2 \quad (3), \quad W_2(i, j) = \exp\left(\frac{x_2(i) - x_2(j)}{-2k_2^2}\right) \quad (4),$$

[0120] 其中, $X_2$ 是特征矩阵, $y_2$ 是目标变量向量, $W_2$ 是高斯核,是一个对角元素的权重矩阵,权重系数 $k_2$ 为用户指定,根据学习结果选择相对最优值。

[0121] 在操作S430,通过当前节点获取批量作业链路中各个上游节点的累计作业时长并获得上游节点累计作业时长的最大值。

[0122] 在上游节点有多个时,当前节点选择累计作业时长最大的那个上游节点作为更新自己累计作业时长的初始值。

[0123] 在操作S440,通过当前作业节点的处理时长和上游节点累计作业时长的最大值进行运算,更新当前节点的累计作业时长。

[0124] 下面结合图5~图8对上述当前节点更新作业报文中累计作业时长字段的流程作进一步详细说明。

[0125] 图5示意性示出了根据本公开实施例的当前节点历史作业时长的示意图。

[0126] 如图5所示,X轴表示导入的历史作业文件总数,Y轴表示历史作业时长,从图5可以看出历史作业时长和历史作业文件总数存在一定线性关系,可以利用局部加权线性回归算法对二者之间的关系进行拟合。

[0127] 图6示意性示出了根据本公开实施例的权重系数 $k_2=1$ 的拟合图。

[0128] 如图6所示,X轴表示导入的作业文件总数,Y轴表示作业时长,由图6可见当 $k_2=1$ 时,不能准确拟合样本输入。

[0129] 图7示意性示出了根据本公开实施例的权重系数 $k_2=0.01$ 的拟合图。

[0130] 如图7所示,X轴表示导入的作业文件总数,Y轴表示作业时长,由图7可见当 $k_2=0.01$ 时,可以较为能准确拟合样本输入。

[0131] 图8示意性示出了根据本公开实施例的权重系数 $k_2=0.003$ 的拟合图。

[0132] 如图8所示,X轴表示导入的作业文件总数,Y轴表示作业时长,由图8可见当 $k_2=0.003$ 时,存在过度拟合的现象,拟合曲线在使用中不能适应实际输入变量的变化。

[0133] 通过拟合图6~拟合图8与实际作业时长图5的对比,当 $k_2=0.01$ 能够准确拟合作业文件大小和作业时长之间的关系。

[0134] 图9示意性示出了根据本公开实施例的更新当前节点的报错节点名称的流程图。

[0135] 如图9所示,当前节点更新报错节点名称的方法包括操作S910~操作940。

[0136] 在操作S910,利用当前节点实时检测批量作业链路中的上游节点的作业状态。

[0137] 在操作S920,在上游节点发生作业中断的情况下,将当前节点的报错节点名称更新为发生作业中断的上游节点的名称。

[0138] 在操作S930,通过当前节点利用作业报文将报错节点名称在批量作业链路中向下

游节点传播。

[0139] 在操作S940,在批量作业链路中发生作业中断的节点恢复的情况下,将当前作业节点的作业报文中的报错节点名称重置为空。

[0140] 通过对当前节点的作业报文中报错节点名称的更新,使得批量作业链路中出现作业故障的节点能够及时被其下游节点感知,从而能够提升预警效率。

[0141] 图10示意性示出了根据本公开实施例的报错节点在批量作业链路中传递的过程图。

[0142] 如图10所示,在批量作业链路中,节点中Node[i]作为上游配置的心跳报文(即本公开的作业报文)发往下游节点Node[i+1],下游节点Node[i+1]监测报文。若上游Node[i]发生批量中断,则记录到Node[i]发往Node[i+1]中的心跳报文,指明Node[i]发生批量中断。Node[i+1]向下游Node[i+2]中发送心跳报文中,指明Node[i]发生批量中断。Node[i+2]接收到信息,通过心跳报文向下游告知Node[i]处发生批量中断。同理,批量中断恢复后,上游节点发送的心跳报文中对应的中断内容清空,下游根据上游报文中的内容同步更新本节点中断报警。

[0143] 根据本公开的实施例,在当前节点的作业时长满足预设时间阈值条件的情况下,发出预警信息包括:在批量作业链路中上游节点的累计作业时长的最大值超过当前节点的预设开始作业时间阈值的情况下,通过当前节点发出第一预警信息;在当前节点的作业处理时长超过当前节点预设的作业处理时间阈值的情况下,通过当前节点发出第二预警信息。

[0144] 以Node[i+1]作为当前节点,对上述预警信息的发出过程作进一步说明。当前节点设置有文件到齐后批量开始时间阈值,若当前节点的某一个上游Node[j]心跳报文(即本公开的作业报文)传来的预估链路累计处理时间accumulateTime超过了当前节点的开始时间阈值,则当前节点预警批量开始时间过晚,供当前节点应用系统管理员评估影响。

[0145] 若当前节点的某一个上游Node[k]心跳报文传来的预估批量文件大小batchPacketSize,叠加Node[i],Node[j]传来的batchPacketSize,得到当前节点批量作业文件总大小(即本公开的作业文件大小),经过历史经验评估,超过了当前节点的执行时长阈值,则当前节点预警批量执行时间过长,供当前节点应用系统管理员评估影响。

[0146] 根据本公开的实施例,上述批量作业监控方法还包括:将更新后的当前节点的作业报文发送到批量作业链路中的下游节点。

[0147] 本公开实施例公开的上述批量作业监控方法,通过作业报文将批量作业链路中上下游节点的作业状态进行高效地传递,实现了上下游节点作业状态的实时共享,能够提高批量作业的效率,同时能够节约运维时间和运维成本。

[0148] 图11示意性示出了根据本公开实施例的基于作业报文的批量作业流程图,下面结合图11对上述批量作业监控方法作进一步详细说明。

[0149] 如图11所示,当前节点Node[i+1]接收多个上游节点(如Node[i]、Node[j]以及Node[k])的作业报文,读取各个上游节点的发送到当前节点的预估批处理文件的大小,同时读取各个上游节点的批处理时间结束时间,即作业报文中累计作业时长。当各个上游节点的作业时长超过了当前节点开始作业的时间阈值时,向运维人员发出预警信息,即作业时间开始太晚,请求运维人员及时进行作业调度。

[0150] 当前节点在读取上游各个节点的作业文件大小,利用LWLR进行当前节点作业文件大小预估和作业处理时长预估。

[0151] 在当前节点预估的作业处理时长超过预设的时间阈值时,向运维人员发出预警信息,以方便运维人员进行处理。

[0152] 将更新的当前节点的作业报文发送到下游各个节点。

[0153] 图12示意性示出了根据本公开实施例的一种批量作业监控装置的结构框图。

[0154] 如图12所示,该实施例的批量作业监控装置1200包括接收模块1210、更新模块1220、获取模块1230以及预警模块1240。

[0155] 接收模块1210,用于接收批量作业链路中上游节点的作业报文;

[0156] 更新模块1220,用于利用上游节点的作业报文,更新批量作业链路中当前节点的作业报文;

[0157] 获取模块1230,用于根据更新后的当前节点的作业报文,获取当前节点的作业时间;

[0158] 预警模块1240,用于在当前节点的作业时间超过预设时间阈值的情况下,发出预警信息。

[0159] 上述批量作业的监控装置,能够对批量作业的链路节点,配置从上游到下游的健康检查,通过批量节点间报文的传递交互,实现对批量作业风险的提前预警。该装置可以为下游应用系统的管理员提前预警批量文件的传输情况。为下游提早处理批量作业提供了丰富的时间。因此,预期整体运营成本可以得到有效降低。

[0160] 根据本公开的实施例,接收模块1210、更新模块1220、获取模块1230以及预警模块1240中的任意多个模块可以合并在一个模块中实现,或者其中的任意一个模块可以被拆分成多个模块。或者,这些模块中的一个或多个模块的至少部分功能可以与其他模块的至少部分功能相结合,并在一个模块中实现。根据本公开的实施例接收模块1210、更新模块1220、获取模块1230以及预警模块1240中的至少一个可以至少被部分地实现为硬件电路,例如现场可编程门阵列(FPGA)、可编程逻辑阵列(PLA)、片上系统、基板上的系统、封装上的系统、专用集成电路(ASIC),或可以通过对电路进行集成或封装的任何其他的合理方式等硬件或固件来实现,或以软件、硬件以及固件三种实现方式中任意一种或以其中任意几种的适当组合来实现。或者,接收模块1210、更新模块1220、获取模块1230以及预警模块1240中的至少一个可以至少被部分地实现为计算机程序模块,当该计算机程序模块被运行时,可以执行相应的功能。

[0161] 图13示意性示出了根据本公开实施例的适于实现批量作业监控方法的电子设备的方框图。

[0162] 如图13所示,根据本公开实施例的电子设备1300包括处理器1301,其可以根据存储在只读存储器(ROM)1302中的程序或者从存储部分1308加载到随机访问存储器(RAM)1303中的程序而执行各种适当的动作和处理。处理器1301例如可以包括通用微处理器(例如CPU)、指令集处理器和/或相关芯片组和/或专用微处理器(例如,专用集成电路(ASIC))等等。处理器1301还可以包括用于缓存用途的板载存储器。处理器1301可以包括用于执行根据本公开实施例的方法流程的不同动作的单一处理单元或者是多个处理单元。

[0163] 在RAM 1303中,存储有电子设备1300操作所需的各种程序和数据。处理器1301、

ROM 1302以及RAM 1303通过总线1304彼此相连。处理器1301通过执行ROM 1302和/或RAM 1303中的程序来执行根据本公开实施例的方法流程的各种操作。需要注意,所述程序也可以存储在除ROM 1302和RAM 1303以外的一个或多个存储器中。处理器1301也可以通过执行存储在所述一个或多个存储器中的程序来执行根据本公开实施例的方法流程的各种操作。

[0164] 根据本公开的实施例,电子设备1300还可以包括输入/输出(I/O)接口1305,输入/输出(I/O)接口1305也连接至总线1304。电子设备1300还可以包括连接至I/O接口1305的以下部件中的一项或多项:包括键盘、鼠标等的输入部分1306;包括诸如阴极射线管(CRT)、液晶显示器(LCD)等以及扬声器等的输出部分1307;包括硬盘等的存储部分1308;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1309。通信部分1309经由诸如因特网的网络执行通信处理。驱动器1310也根据需要连接至I/O接口1305。可拆卸介质1311,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1310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1308。

[0165] 本公开还提供了一种计算机可读存储介质,该计算机可读存储介质可以是上述实施例中描述的设备/装置/系统中所包含的;也可以是单独存在,而未装配入该设备/装置/系统中。上述计算机可读存储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执行时,实现根据本公开实施例的方法。

[0166] 根据本公开的实施例,计算机可读存储介质可以是非易失性的计算机可读存储介质,例如可以包括但不限于: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)、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。在本公开中,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。例如,根据本公开的实施例,计算机可读存储介质可以包括上文描述的ROM 1302和/或RAM 1303和/或ROM 1302和RAM 1303以外的一个或多个存储器。

[0167] 本公开的实施例还包括一种计算机程序产品,其包括计算机程序,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。当计算机程序产品在计算机系统中运行时,该程序代码用于使计算机系统实现本公开实施例所提供的物品推荐方法。

[0168] 在该计算机程序被处理器1301执行时执行本公开实施例的系统/装置中限定的上述功能。根据本公开的实施例,上文描述的系统、装置、模块、单元等可以通过计算机程序模块来实现。

[0169] 在一种实施例中,该计算机程序可以依托于光存储器件、磁存储器件等有形存储介质。在另一种实施例中,该计算机程序也可以在网络介质上以信号的形式进行传输、分发,并通过通信部分1309被下载和安装,和/或从可拆卸介质1311被安装。该计算机程序包含的程序代码可以用任何适当的网络介质传输,包括但不限于:无线、有线等等,或者上述的任意合适的组合。

[0170] 在这样的实施例中,该计算机程序可以通过通信部分1309从网络上被下载和安装,和/或从可拆卸介质1311被安装。在该计算机程序被处理器1301执行时,执行本公开实施例的系统中限定的上述功能。根据本公开的实施例,上文描述的系统、设备、装置、模块、单元等可以通过计算机程序模块来实现。

[0171] 根据本公开的实施例,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

行本公开实施例提供的计算机程序的程序代码,具体地,可以利用高级过程和/或面向对象的编程语言、和/或汇编/机器语言来实施这些计算程序。程序设计语言包括但不限于诸如 Java, C++, python, “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、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)。

[0172] 附图中的流程图和框图,图示了按照本公开各种实施例的系统、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,上述模块、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,框图或流程图中的每个方框、以及框图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。

[0173] 本领域技术人员可以理解,本公开的各个实施例和/或权利要求中记载的特征可以进行多种组合和/或结合,即使这样的组合或结合没有明确记载于本公开中。特别地,在不脱离本公开精神和教导的情况下,本公开的各个实施例和/或权利要求中记载的特征可以进行多种组合和/或结合。所有这些组合和/或结合均落入本公开的范围。

[0174] 以上对本公开的实施例进行了描述。但是,这些实施例仅仅是为了说明的目的,而并非为了限制本公开的范围。尽管在以上分别描述了各实施例,但是这并不意味着各个实施例中的措施不能有利地结合使用。本公开的范围由所附权利要求及其等同物限定。不脱离本公开的范围,本领域技术人员可以做出多种替代和修改,这些替代和修改都应落在本公开的范围之内。

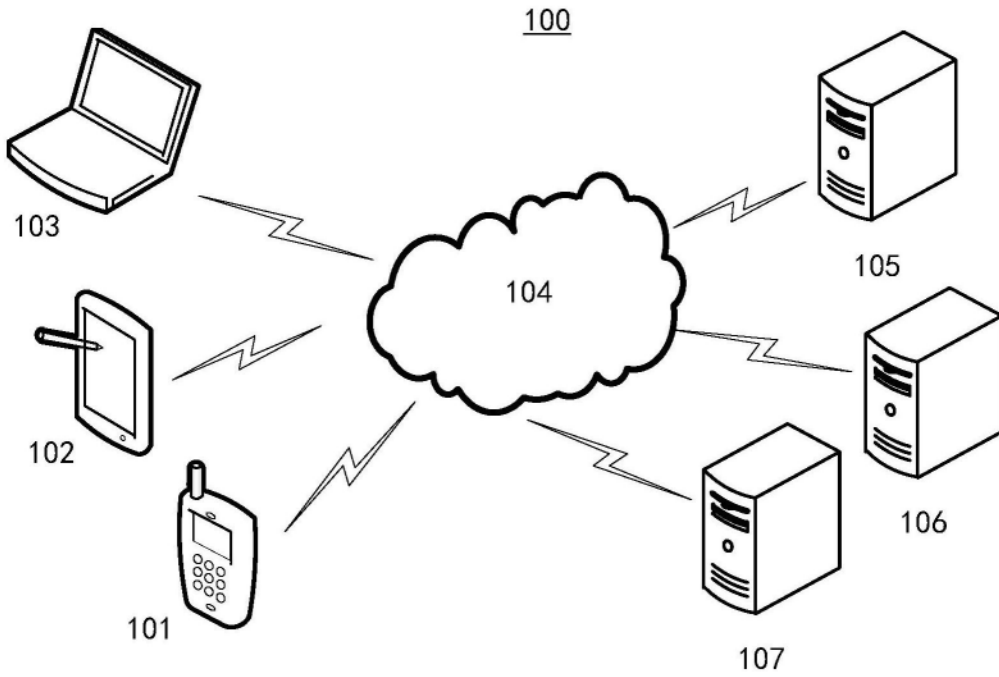


图1

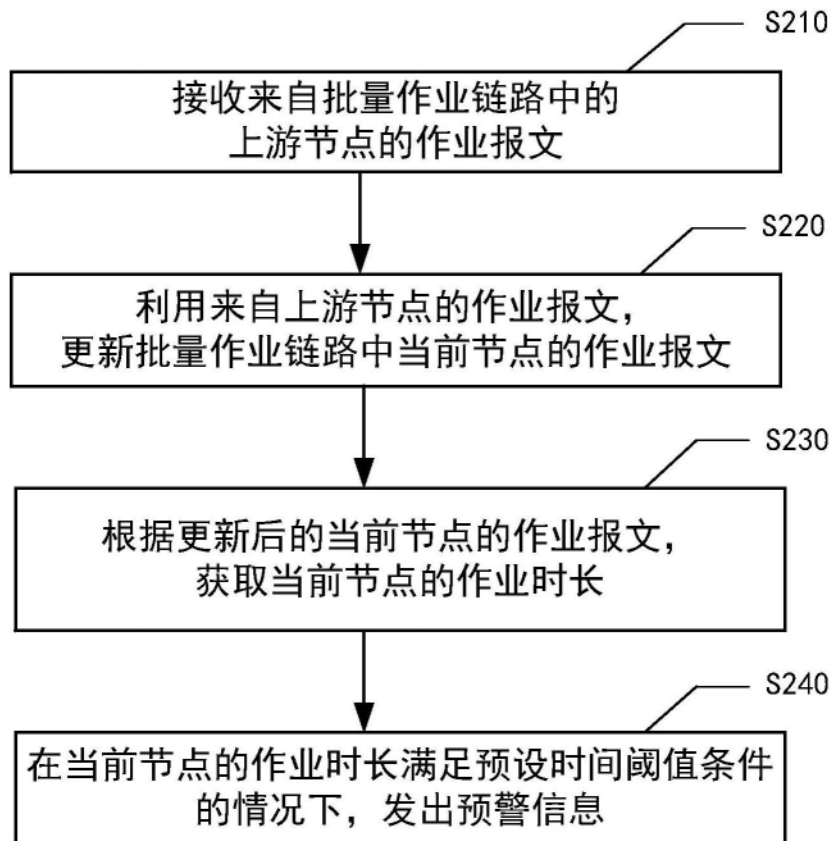


图2



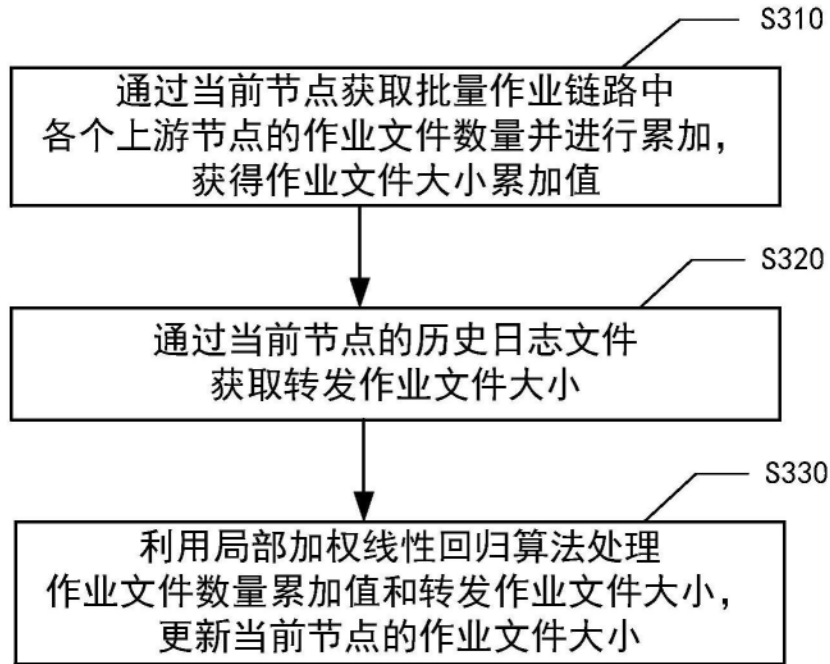


图3

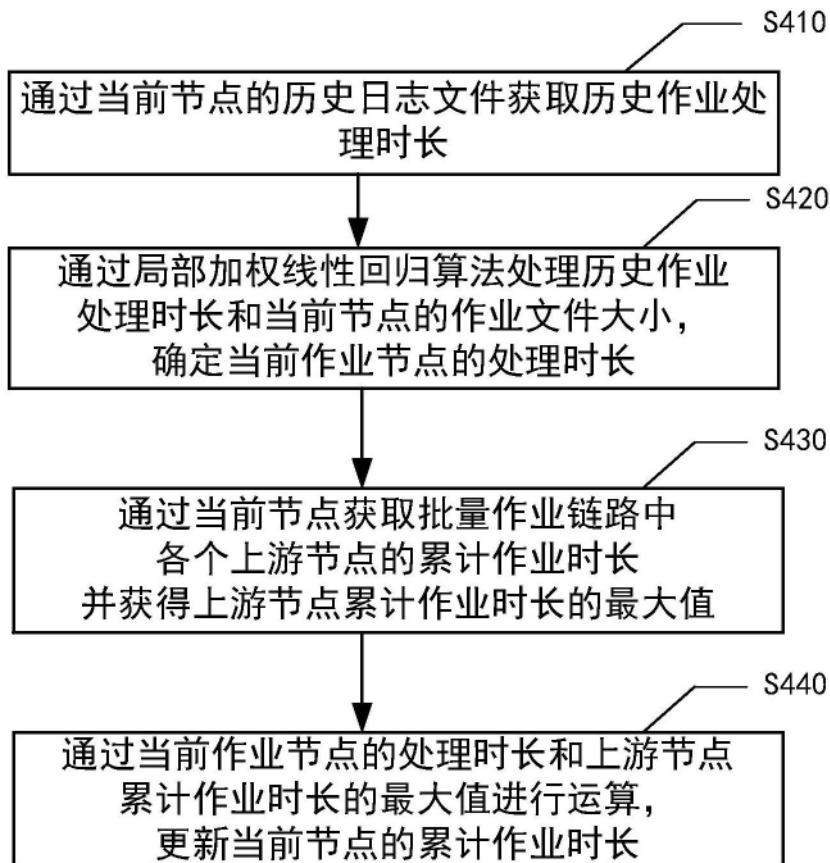


图4

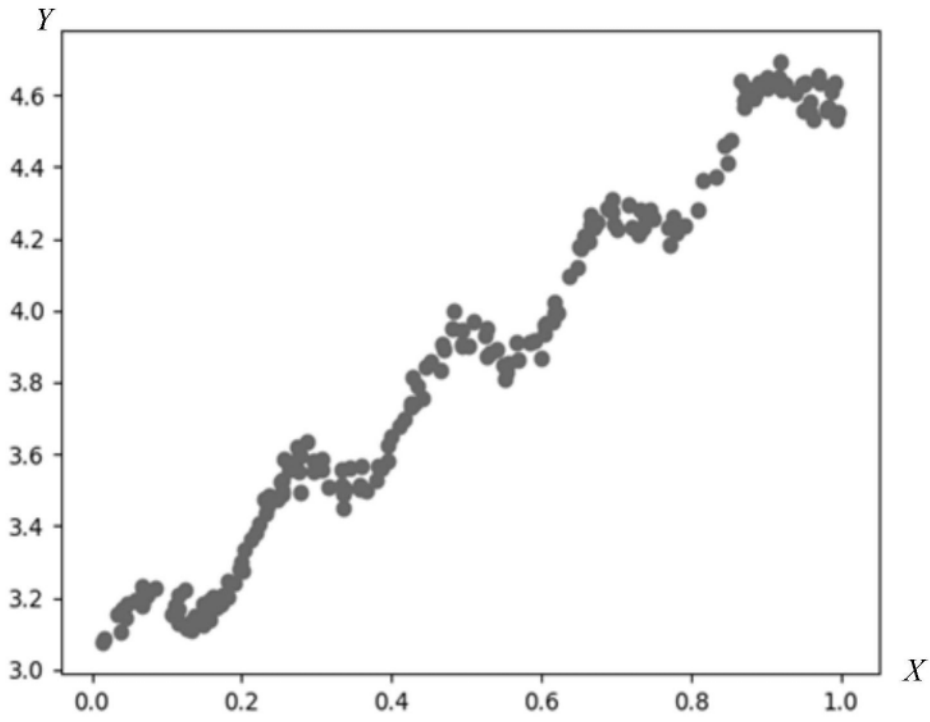


图5

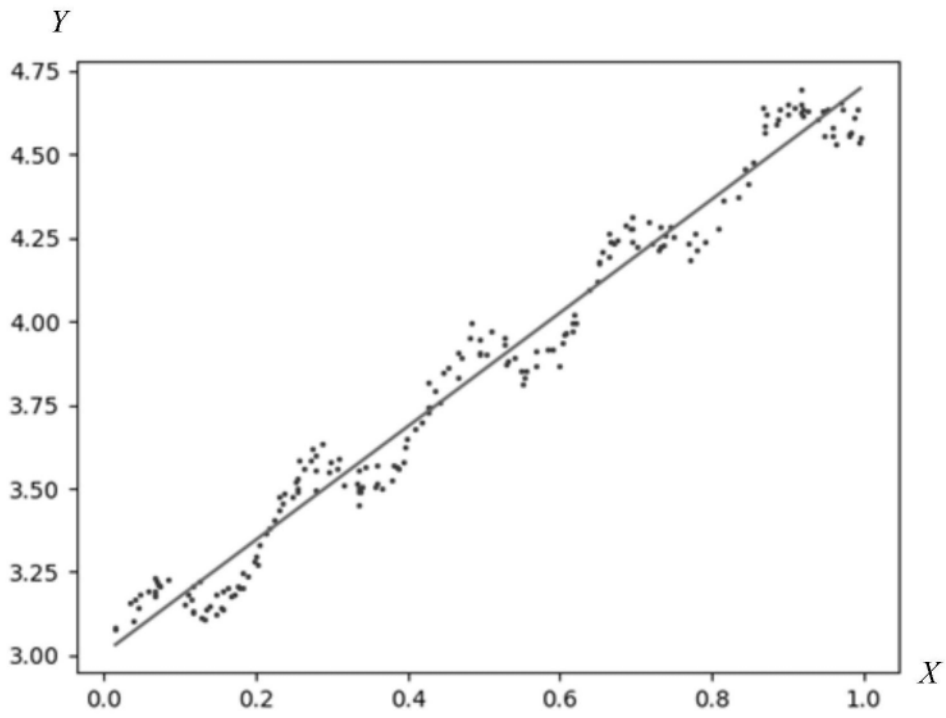


图6

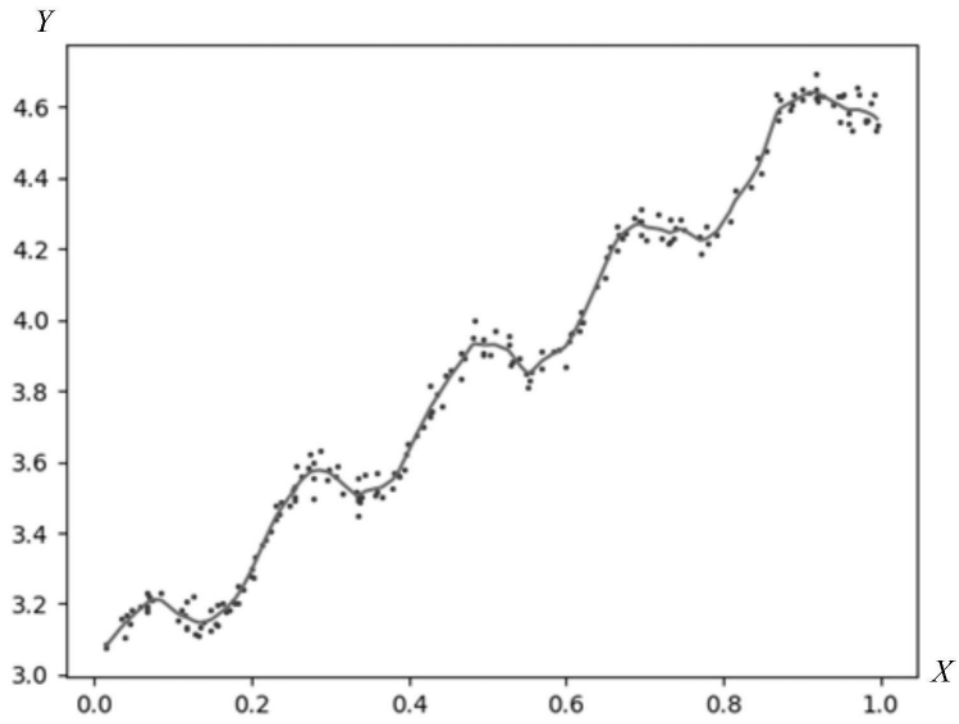


图7

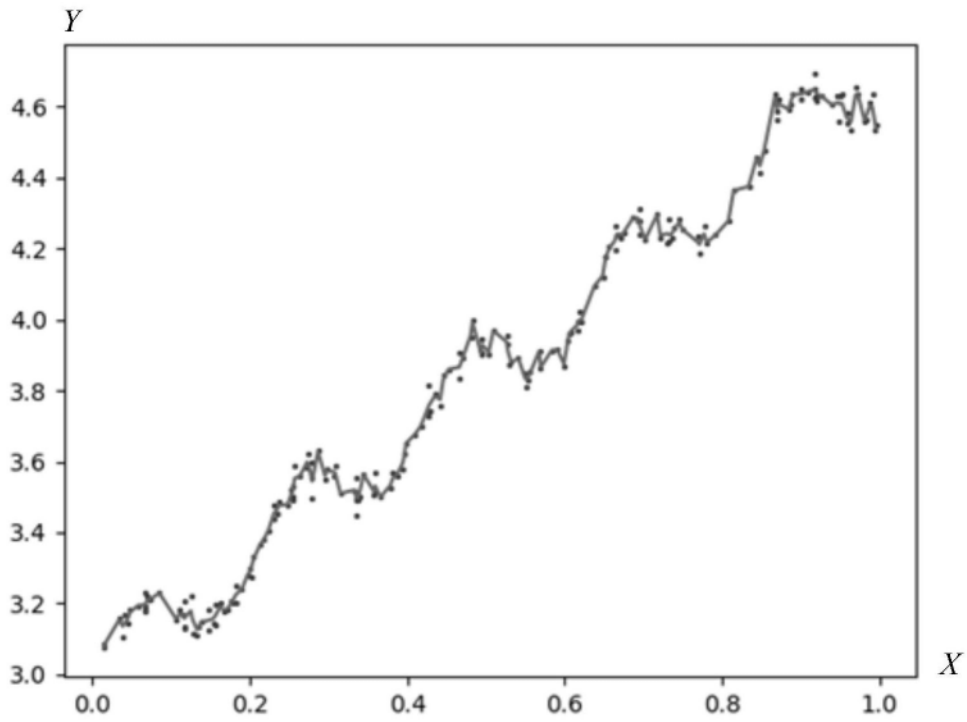


图8

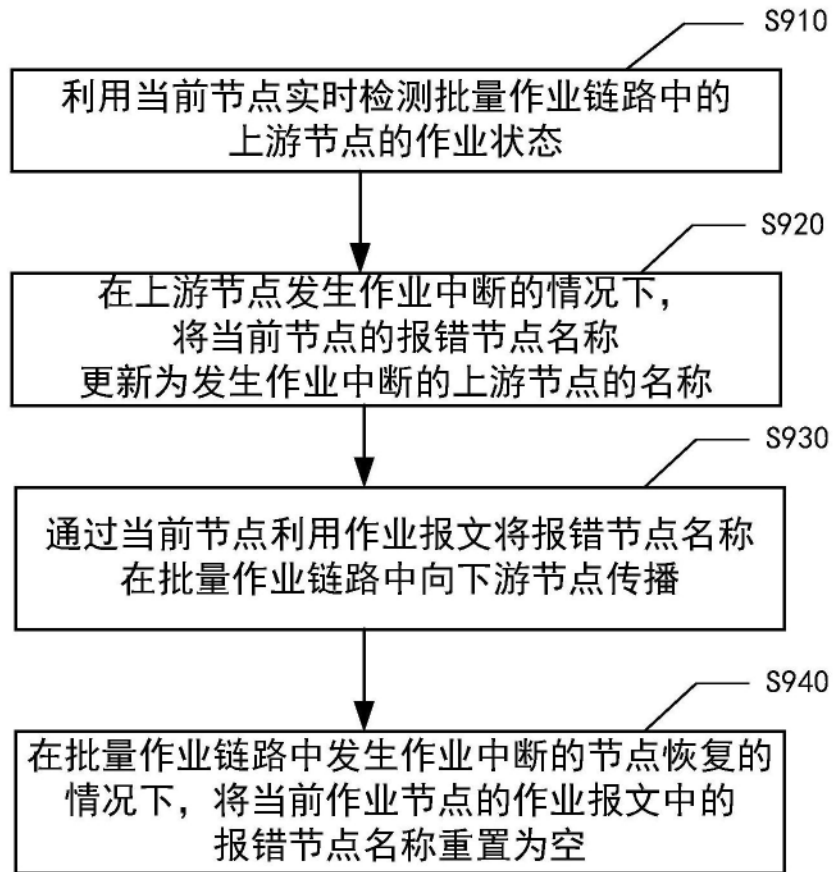


图9

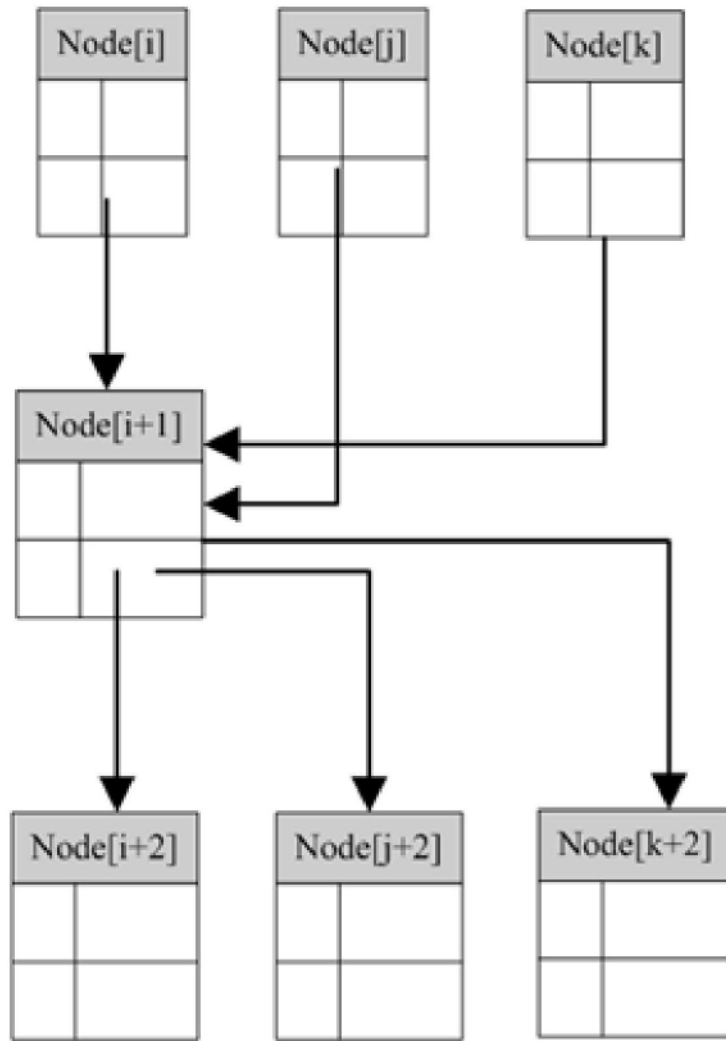


图10

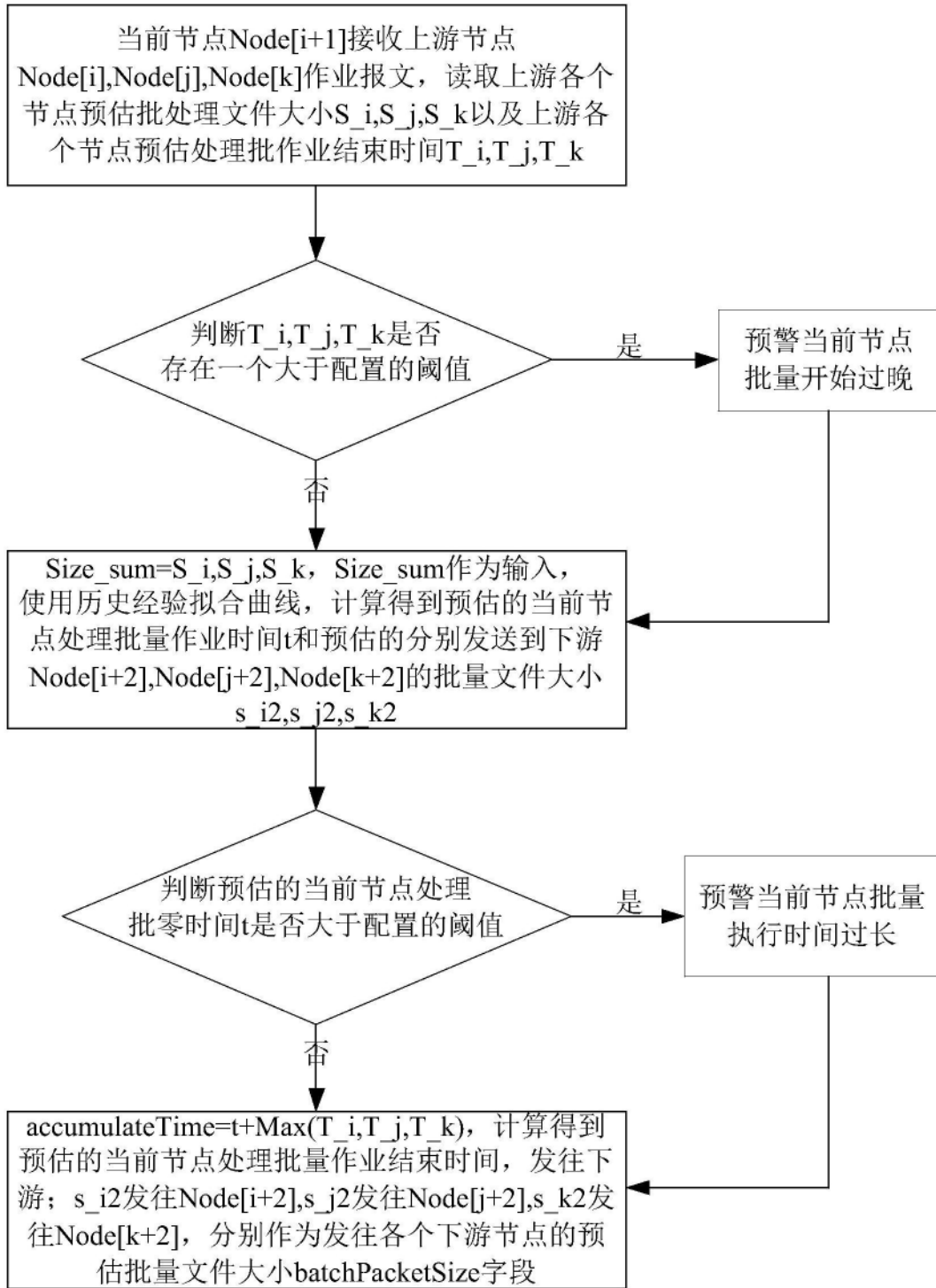


图11

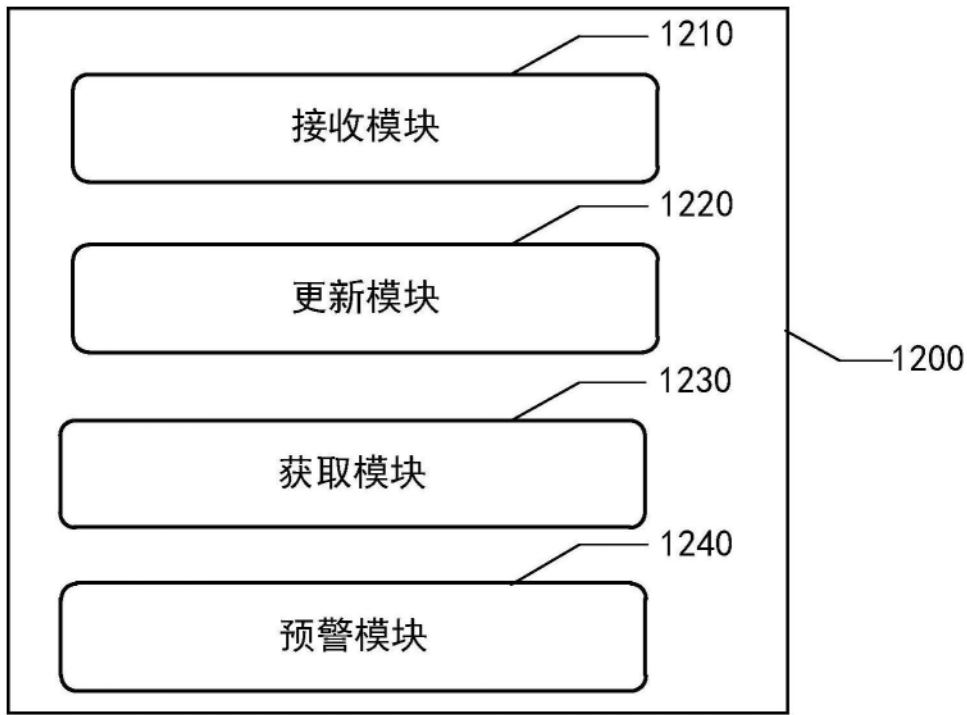


图12

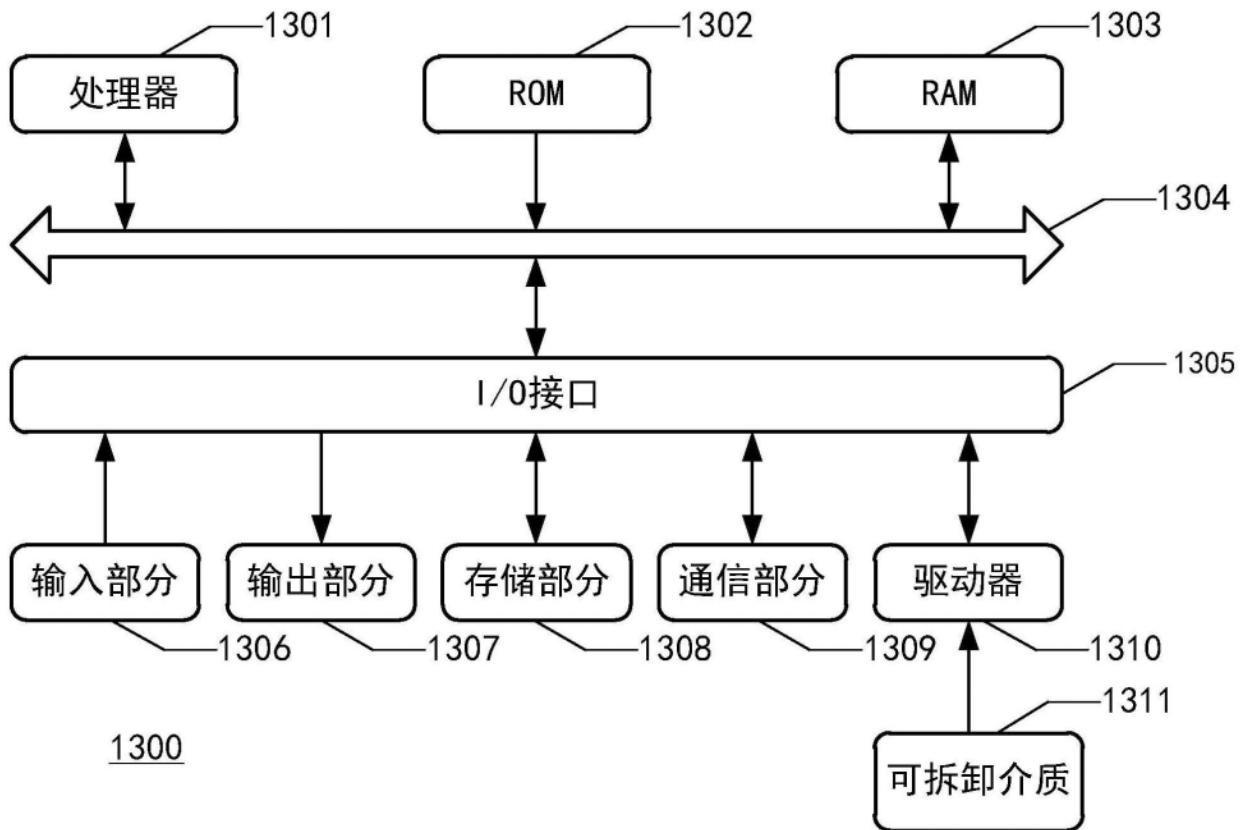


图13