



(12)发明专利

(10)授权公告号 CN 106055452 B

(45)授权公告日 2019.06.14

(21)申请号 201610355129.3

(22)申请日 2016.05.25

(65)同一申请的已公布的文献号
申请公布号 CN 106055452 A

(43)申请公布日 2016.10.26

(73)专利权人 北京百度网讯科技有限公司
地址 100085 北京市海淀区上地十街10号
百度大厦2层

(72)发明人 董辉 宋磊 侯翔宇 孟伟彬

(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 11/30(2006.01)
H04L 12/24(2006.01)

(56)对比文件

CN 104408100 A,2015.03.11,
CN 103379136 A,2013.10.30,
CN 101320348 A,2008.12.10,
US 2007130099 A1,2007.06.07,
US 2005203934 A1,2005.09.15,
张珠玉等.基于网格的TCP网络日志二次聚
类算法.《暨南大学学报(自然科学版)》.2011,
唐球等.基于差分压缩的大规模日志压缩系
统.《通信学报》.2015,

审查员 史雪飞

权利要求书3页 说明书8页 附图6页

(54)发明名称

创建交换机日志模板的方法和装置

(57)摘要

本申请公开了创建交换机日志模板的方法和装置。所述方法的一具体实施方式包括：获取一个型号的交换机的原始日志；获取所述原始日志中的消息类型和详细消息，如果获取不到消息类型，则创建聚类标签作为消息类型；对所述详细消息进行分词，得到关键词；根据所述关键词的词频对所述关键词重新排序，将词频高的关键词排在前面；根据所述重新排序的关键词创建多叉树，每个消息类型作为树的根节点，所述重新排序的关键词作为树的节点，每个消息类型对应一个多叉树；深度优先遍历所述多叉树，根据所述多叉树中的路径创建所述型号交换机的每个消息类型所对应的日志模板。该实施方式创建交换机日志模板，使用该模板对日志进行压缩。



1. 一种创建交换机日志模板的方法,其特征在于,所述方法包括:
 - 获取一个型号的交换机的原始日志;
 - 获取所述原始日志中的消息类型和详细消息,如果获取不到消息类型,则创建聚类标签作为消息类型;
 - 对所述详细消息进行分词,得到关键词;
 - 根据所述关键词的词频对所述关键词重新排序,将词频高的关键词排在前面;
 - 根据所述重新排序的关键词创建多叉树,每个消息类型作为树的根节点,所述重新排序的关键词作为树的节点,每个消息类型对应一个多叉树;
 - 深度优先遍历所述多叉树,根据所述多叉树中的路径创建所述型号交换机的每个消息类型所对应的日志模板。
2. 根据权利要求1所述的创建交换机日志模板的方法,其特征在于,所述方法还包括:
 - 获取一个型号的交换机的新增日志;
 - 获取所述新增日志中的消息类型和详细消息,如果获取不到消息类型,则创建聚类标签作为消息类型;
 - 使用所述日志模板过滤所述新增日志;
 - 将所述日志模板无法匹配的新增日志中的详细消息进行分词,得到关键词;
 - 根据所述关键词的词频对所述关键词重新排序,将词频高的关键词排在前面;
 - 根据所述重新排序的关键词创建多叉树,每个消息类型作为树的根节点,所述重新排序的关键词作为树的节点,每个消息类型对应一个多叉树;
 - 深度优先遍历所述多叉树,根据所述多叉树中的路径创建所述型号交换机的所述新增日志中的每个消息类型所对应的新增日志模板。
3. 根据权利要求1或2所述的创建交换机日志模板的方法,其特征在于,如果所述多叉树中一个节点的子节点超过节点阈值数目,则删除所述节点的所有子节点,所述节点作为最后一个子节点。
4. 根据权利要求1或2所述的创建交换机日志模板的方法,其特征在于,所述创建聚类标签作为消息类型,包括:
 - 将每一条消息类型未知的日志按照语义分成了五种类别并分别赋予权重值,所述五种类别包括:只有数字或数字与符号,数字、字母和符号,符号和字母,只有字母,只有符号;
 - 提取所述日志中五类语义的频度,将所述日志转化为一个固定五个长度的词频向量;
 - 计算所述词频向量与已知的消息类型集合的相似度,得到一组相似度结果,如果最大相似度大于等于预设的相似度阈值,则将所述日志归到对应的消息类型;如果最大相似度小于预设的相似度阈值,则将所述词频向量作为一个新的消息类型。
5. 根据权利要求1或2所述的创建交换机日志模板的方法,其特征在于,当一个日志模板是另外一个日志模板的子集时,通过对节点打标签的形式来标识一个节点是否是一条路径的结束。
6. 根据权利要求5所述的创建交换机日志模板的方法,其特征在于,将每个消息类型中的日志模板按照树的深度排序,当一个日志模板是另外一个日志模板的子集时,优先采用树的深度较大的日志模板进行匹配。
7. 一种创建交换机日志模板的装置,其特征在于,所述装置包括:

获取单元,配置用于获取一个型号的交换机的原始日志;

解析单元,配置用于获取所述原始日志中的消息类型和详细消息,如果获取不到消息类型,则创建聚类标签作为消息类型;

处理单元,配置用于对所述详细消息进行分词,得到关键词;根据所述关键词的词频对所述关键词重新排序,将词频高的关键词排在前面;

创建单元,配置用于根据所述重新排序的关键词创建多叉树,每个消息类型作为树的根节点,所述重新排序的关键词作为树的节点,每个消息类型对应一个多叉树;深度优先遍历所述多叉树,根据所述多叉树中的路径创建所述型号交换机的每个消息类型所对应的日志模板。

8. 根据权利要求7所述的创建交换机日志模板的装置,其特征在于,所述装置还配置用于:

获取一个型号的交换机的新增日志;

获取所述新增日志中的消息类型和详细消息,如果获取不到消息类型,则创建聚类标签作为消息类型;

使用所述日志模板过滤所述新增日志;

将所述日志模板无法匹配的新增日志中的详细消息进行分词,得到关键词;

根据所述关键词的词频对所述关键词重新排序,将词频高的关键词排在前面;

根据所述重新排序的关键词创建多叉树,每个消息类型作为树的根节点,所述重新排序的关键词作为树的节点,每个消息类型对应一个多叉树;

深度优先遍历所述多叉树,根据所述多叉树中的路径创建所述型号交换机的所述新增日志中的每个消息类型所对应的新增日志模板。

9. 根据权利要求7或8所述的创建交换机日志模板的装置,其特征在于,所述创建单元还配置用于:

如果所述多叉树中一个节点的子节点超过节点阈值数目,则删除所述节点的所有子节点,所述节点作为最后一个子节点。

10. 根据权利要求7或8所述的创建交换机日志模板的装置,其特征在于,所述创建聚类标签作为消息类型,包括:

将每一条消息类型未知的日志按照语义分成了五种类别并分别赋予权重值,所述五种类别包括:只有数字或数字与符号,数字、字母和符号,符号和字母,只有字母,只有符号;

提取所述日志中五类语义的频度,将所述日志转化为一个固定五个长度的词频向量;

计算所述词频向量与已知的消息类型集合的相似度,得到一组相似度结果,如果最大相似度大于等于预设的相似度阈值,则将所述日志归到对应的消息类型;如果最大相似度小于预设的相似度阈值,则将所述词频向量作为一个新的消息类型。

11. 根据权利要求7或8所述的创建交换机日志模板的装置,其特征在于,所述创建单元还配置用于:

当一个日志模板是另外一个日志模板的子集时,通过对节点打标签的形式来标识一个节点是否是一条路径的结束。

12. 根据权利要求11所述的创建交换机日志模板的装置,其特征在于,所述创建单元还配置用于:

将每个消息类型中的模板按照树的深度排序,当一个日志模板是另外一个日志模板的子集时,优先采用树的深度较大的模板进行匹配。

创建交换机日志模板的方法和装置

技术领域

[0001] 本申请涉及计算机技术领域,具体涉及互联网技术领域,尤其涉及创建交换机日志模板的方法和装置。

背景技术

[0002] 交换机日志是交换机设备在存活周期内产生的重要数据,通过交换机日志可以获取交换机的状态,包括端口的抖动、协议的抖动、板卡故障、电源故障等,而传统的设备监控系统都是基于交换机日志,通过特定的规则匹配日志,从而发现故障点并进行报警。

[0003] 随着互联网数据的爆炸性增长以及业务的增多,很多互联网企业对自建网络的投入也越来越大,IDC(Internet Data Center,互联网数据中心)中涉及到的网络设备也越来越多,尤其是交换机设备。

[0004] 一个典型的案例是在企业自建IDC中,存在大规模的交换机设备,这些设备分别来自不同的厂商、存在多种型号,同时分布在多个IDC中。过多的设备会产生大规模的交换机日志,对于报警规则的挖掘以及排障都带来了困难,需要一种自动化手段对日志进行压缩。

发明内容

[0005] 本申请的目的在于提出一种创建交换机日志模板的方法和装置,来解决以上背景技术部分提到的技术问题。

[0006] 第一方面,本申请提供了创建交换机日志模板的方法,所述方法包括:获取一个型号的交换机的原始日志;获取所述原始日志中的消息类型和详细消息,如果获取不到消息类型,则创建聚类标签作为消息类型;对所述详细消息进行分词,得到关键词;根据所述关键词的词频对所述关键词重新排序,将词频高的关键词排在前面;根据所述重新排序的关键词创建多叉树,每个消息类型作为树的根节点,所述重新排序的关键词作为树的节点,每个消息类型对应一个多叉树;深度优先遍历所述多叉树,根据所述多叉树中的路径创建所述型号交换机的每个消息类型所对应的日志模板。

[0007] 在一些实施例中,所述方法还包括:获取一个型号交换机的新增日志;获取所述新增日志中的消息类型和详细消息,如果获取不到消息类型,则创建聚类标签作为消息类型;使用所述日志模板过滤所述新增日志;将所述日志模板无法匹配的新增日志中的详细消息进行分词,得到关键词;根据所述关键词的词频对所述关键词重新排序,将词频高的关键词排在前面;根据所述重新排序的关键词创建多叉树,每个消息类型作为树的根节点,所述重新排序的关键词作为树的节点,每个消息类型对应一个多叉树;深度优先遍历所述多叉树,根据所述多叉树中的路径创建所述型号交换机的每个消息类型所对应的新增日志模板。

[0008] 在一些实施例中,如果所述多叉树中一个节点的子节点超过节点阈值数目,则删除所述节点的所有子节点,所述节点作为最后一个子节点。

[0009] 在一些实施例中,所述创建聚类标签作为消息类型,包括:将每一条消息类型未知

的日志按照语义分成了五种类别并分别赋予权重值,所述五种类别包括:只有数字或数字与符号,数字、字母和符号,符号和字母,只有字母,只有符号;提取所述日志中五类语义的频度,将所述日志转化为一个固定五个长度的词频向量;计算所述词频向量与已知的消息类型集合的相似度,得到一组相似度结果,如果最大相似度大于等于预设的相似度阈值,则将所述日志归到对应的消息类型;如果最大相似度小于预设的相似度阈值,则将所述词频向量作为一个新的消息类型。

[0010] 在一些实施例中,当一个日志模板是另外一个日志模板的子集时,通过对节点打标签的形式来标识一个节点是否是一条路径的结束。

[0011] 在一些实施例中,将每个消息类型中的日志模板按照树的深度排序,当一个日志模板是另外一个日志模板的子集时,优先采用树的深度较大的日志模板进行匹配。

[0012] 第二方面,本申请提供了一种创建交换机日志模板的装置,其特征在于,所述装置包括:获取单元,配置用于获取一个型号的交换机的原始日志;解析单元,配置用于获取所述原始日志中的消息类型和详细消息,如果获取不到消息类型,则创建聚类标签作为消息类型;处理单元,配置用于对所述详细消息进行分词,得到关键词;根据所述关键词的词频对所述关键词重新排序,将词频高的关键词排在前面;创建单元,配置用于根据所述重新排序的关键词创建多叉树,每个消息类型作为树的根节点,所述重新排序的关键词作为树的节点,每个消息类型对应一个多叉树;深度优先遍历所述多叉树,根据所述多叉树中的路径创建所述型号交换机的每个消息类型所对应的日志模板。

[0013] 在一些实施例中,所述装置还配置用于:获取一个型号的交换机的新增日志;获取所述新增日志中的消息类型和详细消息,如果获取不到消息类型,则创建聚类标签作为消息类型;使用所述日志模板过滤所述新增日志;将所述日志模板无法匹配的新增日志中的详细消息进行分词,得到关键词;根据所述关键词的词频对所述关键词重新排序,将词频高的关键词排在前面;根据所述重新排序的关键词创建多叉树,每个消息类型作为树的根节点,所述重新排序的关键词作为树的节点,每个消息类型对应一个多叉树;深度优先遍历所述多叉树,根据所述多叉树中的路径创建所述型号交换机的每个消息类型所对应的新增日志模板。

[0014] 在一些实施例中,所述创建单元还配置用于:如果所述多叉树中一个节点的子节点超过节点阈值数目,则删除所述节点的所有子节点,所述节点作为最后一个子节点。

[0015] 在一些实施例中,所述创建聚类标签作为消息类型,包括:将每一条消息类型未知的日志按照语义分成了五种类别并分别赋予权重值,所述五种类别包括:只有数字或数字与符号,数字、字母和符号,符号和字母,只有字母,只有符号;提取所述日志中五类语义的频度,将所述日志转化为一个固定五个长度的词频向量;计算所述词频向量与已知的消息类型集合的相似度,得到一组相似度结果,如果最大相似度大于等于预设的相似度阈值,则将所述日志归到对应的消息类型;如果最大相似度小于预设的相似度阈值,则将所述词频向量作为一个新的消息类型。

[0016] 在一些实施例中,所述创建单元还配置用于:当一个日志模板是另外一个日志模板的子集时,通过对节点打标签的形式来标识一个节点是否是一条路径的结束。

[0017] 在一些实施例中,所述创建单元还配置用于:将每个消息类型中的模板按照树的深度排序,当一个日志模板是另外一个日志模板的子集时,优先采用树的深度较大的模板

进行匹配。

[0018] 本申请提供的创建交换机日志模板的方法和装置,通过获取所述原始日志中的消息类型和详细消息,根据详细消息中的关键词创建多叉树,根据多叉树中的路径创建交换机的每个消息类型所对应的日志模板,以采用增量迭代训练方式对交换机日志进行压缩。

附图说明

[0019] 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:

[0020] 图1是本申请可以应用于其中的示例性系统架构图;

[0021] 图2是根据本申请的创建交换机日志模板的方法的一个实施例的流程图;

[0022] 图3是根据本申请的创建交换机日志模板的方法的词频向量的示意图;

[0023] 图4a、4b和4c是根据本申请的创建交换机日志模板的方法的一个应用场景的示意图;

[0024] 图5是根据本申请的创建交换机日志模板的方法的又一个实施例的流程图;

[0025] 图6是根据本申请的创建交换机日志模板的装置的一个实施例的结构示意图;

[0026] 图7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

[0027] 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。

[0028] 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。

[0029] 图1示出了可以应用本申请的创建交换机日志模板的方法或创建交换机日志模板的装置的实施例的示例性系统架构100。

[0030] 如图1所示,系统架构100可以包括交换机101、102、103,网络104和服务器105。网络104用以在交换机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。

[0031] 交换机101、102、103的日志通过网络104传输到服务器105。交换机101、102、103上可以安装有各种日志采集工具的客户端,例如安装有可以将其内部的日志信息传输到远程的日志服务器的rsyslog等。

[0032] 交换机101、102、103可以是接入交换机的任意两个网络节点提供独享的电信号通路的网络设备,包括但不限于以太网交换机、快速以太网交换机、千兆以太网交换机、FDDI交换机、ATM交换机和令牌环交换机等。

[0033] 服务器105可以是提供各种服务的服务器,例如采集交换机101、102、103的日志,对采集到的日志进行结构化处理并创建日志模板的服务器。

[0034] 需要说明的是,本申请实施例所提供的创建交换机日志模板的方法一般由服务器105执行,相应地,创建交换机日志模板的装置一般设置于服务器105中。

[0035] 应该理解,图1中的交换机、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交换机、网络和服务器的。

[0036] 继续参考图2,示出了根据本申请的创建交换机日志模板的方法的一个实施例的流程200。所述的创建交换机日志模板的方法,包括以下步骤:

[0037] 步骤201,获取一个型号的交换机的原始日志。

[0038] 在本实施例中,创建交换机日志模板的方法运行于其上的电子设备(例如图1所示的服务器)可以通过有线连接方式或者无线连接方式从交换机获取原始日志。

[0039] 步骤202,获取原始日志中的消息类型和详细消息,如果获取不到消息类型,则创建聚类标签作为消息类型。

[0040] 在本实施例中,获取原始日志中的消息类型和详细消息,如果获取不到消息类型,则创建聚类标签作为消息类型。其中,所述详细消息需要去除变量。

[0041] 在本实施例的一些可选的实现方式中,所述创建聚类标签作为消息类型,包括:将每一条消息类型未知的日志按照语义分成了五种类别并分别赋予权重值,所述五种类别包括:只有数字或数字与符号,数字、字母和符号,符号和字母,只有字母,只有符号,如表1中所示;提取所述日志中五类语义的频度,将所述日志转化为一个固定五个长度的词频向量,如图3所示;计算所述词频向量与已知的消息类型集合的相似度,得到一组相似度结果,如果最大相似度大于等于预设的相似度阈值,则将所述日志归到对应的消息类型;如果最大相似度小于预设的相似度阈值,则将所述词频向量作为一个新的消息类型。

[0042]

类别	定义	举例	权重
1	只有数字或数字与符号	1, 3/6, 192.168.1.1	0.1
2	数字、字母和符号	Vlan-interface1203, vty0, Ten-GigabitEthernet1/0/4	0.2
3	符号和字母	IS-IS, udp-port, aaa.cfg, line-protocol	0.6
4	只有字母	linkdown, state, interface	0.8
5	只有符号	<, >, =, :, @	1

[0043] 表1

[0044] 步骤203,对详细消息进行分词,得到关键词。

[0045] 在本实施例中,采用Luene进行分词,得到关键词。

[0046] 步骤204,根据关键词的词频对关键词重新排序,将词频高的关键词排在前面。

[0047] 在本实施例中,根据关键词的词频对关键词重新排序,将词频高的关键词排在前面。

[0048] 步骤205,根据重新排序的关键词创建多叉树。

[0049] 在本实施例中,根据步骤204中重新排序的关键词创建多叉树。

[0050] 在本实施例的一些可选的实现方式中,如果所述多叉树中一个节点的子节点超过节点阈值数目,则删除所述节点的所有子节点,所述节点作为最后一个子节点。例如,如果一个节点的子节点超过10个(经验值),则砍掉该节点的所有子节点,该节点作为最后一个

子节点。这样做的目的是防止模板膨胀,因为针对每一个消息类型,一般不会超过10个有效的状态。例如,以登录日志为例,处理用户名其它部分关键词一致,如果不进行子节点数量的限制,就会导致过多的模板表示同一含义。

[0051] 步骤206,深度优先遍历多叉树,根据多叉树中的路径创建该型号交换机的每个消息类型所对应的日志模板。

[0052] 在本实施例中,深度优先遍历步骤205创建的多叉树,根据多叉树中的路径创建该型号交换机的每个消息类型所对应的日志模板。

[0053] 在本实施例的一些可选的实现方式中,当一个日志模板是另外一个日志模板的子集时,通过对节点打标签的形式来标识一个节点是否是一条路径的结束(该节点可能不是最长路径的结束)。

[0054] 在本实施例的一些可选的实现方式中,当一个日志模板是另外一个日志模板的子集时,优先采用树的深度较大的日志模板进行匹配。在节点不是最长路径的结束的情况下,用路径最长的匹配。

[0055] 继续参见图4a-4c,图4a-4c是根据本实施例的创建交换机日志模板的方法的应用场景的一个示意图。在图4a-4c的应用场景中,在图4a中,图左侧为经过按照词频排序后的日志,图右侧为构造出的多叉树。在图4b中,节点down的子节点数目多于10个,因此删除down的子节点,并将down作为最后一个子节点。在图4c中,节点up是一条路径的结束,对其打上标签,但它不是最长路径,第一模板的路径比第二模板的路径长。采用模板进行日志匹配时优先采用第一模板进行匹配。

[0056] 本申请的上述实施例提供的方法通过对已经消息类型的日志进行分词处理得到关键词后创建多叉树,创建了交换机日志模板,可用于增量迭代训练方式压缩日志,日志压缩比可达到2000:1。

[0057] 进一步参考图5,其示出了创建交换机日志模板的方法的又一个实施例的流程500。该创建交换机日志模板的方法的流程500,包括以下步骤:

[0058] 步骤501,获取一个型号的交换机的新增日志。

[0059] 在本实施例中,该步骤与步骤201基本相同,区别在于该步骤获取的是新增的日志。

[0060] 步骤502,获取新增日志中的消息类型和详细消息,如果获取不到消息类型,则创建聚类标签作为消息类型。

[0061] 步骤502与步骤202基本相同,在此不再赘述。

[0062] 步骤503,使用日志模板过滤新增日志。

[0063] 在本实施例中,使用在步骤206中创建的日志模板过滤新增日志,得到原日志模板无法匹配的日志进行增量训练。

[0064] 步骤504,将日志模板无法匹配的新增日志中的详细消息进行分词,得到关键词。

[0065] 步骤504与步骤203基本相同,在此不再赘述。

[0066] 步骤505,根据关键词的词频对关键词重新排序,将词频高的关键词排在前面。

[0067] 步骤506,根据重新排序的关键词创建多叉树。

[0068] 步骤507,深度优先遍历多叉树,根据多叉树中的路径创建该型号交换机的每个消息类型所对应的新增日志模板。

[0069] 步骤505-507与步骤204-206基本相同,在此不再赘述。

[0070] 从图5中可以看出,与图2对应的实施例相比,本实施例中的创建交换机日志模板的方法的流程500突出了对新增日志进行过滤的步骤。由此,本实施例描述的方案可以采用增量迭代训练的方式对日志进行压缩。

[0071] 进一步参考图6,作为对上述各图所示方法的实现,本申请提供了一种创建交换机日志模板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。

[0072] 如图6所示,本实施例所述的创建交换机日志模板的装置600包括:获取单元601、解析单元602、处理单元603和创建单元604。其中,获取单元601配置用于获取一个型号的交换机的原始日志;解析单元602配置用于获取所述原始日志中的消息类型和详细消息,如果获取不到消息类型,则创建聚类标签作为消息类型;处理单元603配置用于对所述详细消息进行分词,得到关键词;根据所述关键词的词频对所述关键词重新排序,将词频高的关键词排在前面;创建单元604配置用于根据所述重新排序的关键词创建多叉树,每个消息类型作为树的根节点,所述重新排序的关键词作为树的节点,每个消息类型对应一个多叉树;深度优先遍历所述多叉树,根据所述多叉树中的路径创建所述型号交换机的每个消息类型所对应的日志模板。

[0073] 在本实施例的一些可选的实现方式中,该创建交换机日志模板的装置600还配置用于:获取一个型号的交换机的新增日志;获取所述新增日志中的消息类型和详细消息,如果获取不到消息类型,则创建聚类标签作为消息类型;使用所述日志模板过滤所述新增日志;将所述日志模板无法匹配的新增日志中的详细消息进行分词,得到关键词;根据所述关键词的词频对所述关键词重新排序,将词频高的关键词排在前面;根据所述重新排序的关键词创建多叉树,每个消息类型作为树的根节点,所述重新排序的关键词作为树的节点,每个消息类型对应一个多叉树;深度优先遍历所述多叉树,根据所述多叉树中的路径创建所述型号交换机的每个消息类型所对应的新增日志模板。

[0074] 在本实施例的一些可选的实现方式中,创建单元604还配置用于:如果所述多叉树中一个节点的子节点超过节点阈值数目,则删除所述节点的所有子节点,所述节点作为最后一个子节点。

[0075] 在本实施例的一些可选的实现方式中,创建聚类标签作为消息类型,包括:将每一条消息类型未知的日志按照语义分成了五种类别并分别赋予权重值,所述五种类别包括:只有数字或数字与符号,数字、字母和符号,符号和字母,只有字母,只有符号;提取所述日志中五类语义的频度,将所述日志转化为一个固定五个长度的词频向量;计算所述词频向量与已知的消息类型集合的相似度,得到一组相似度结果,如果最大相似度大于等于预设的相似度阈值,则将所述日志归到对应的消息类型;如果最大相似度小于预设的相似度阈值,则将所述词频向量作为一个新的消息类型。

[0076] 在本实施例的一些可选的实现方式中,创建单元604还配置用于:当一个日志模板是另外一个日志模板的子集时,通过对节点打标签的形式来标识一个节点是否是一条路径的结束。

[0077] 在本实施例的一些可选的实现方式中,创建单元604还配置用于:将每个消息类型中的模板按照树的深度排序,当一个日志模板是另外一个日志模板的子集时,优先采用树

的深度较大的模板进行匹配。

[0078] 下面参考图7,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700的结构示意图。

[0079] 如图7所示,计算机系统700包括中央处理单元603(CPU)701,其可以根据存储在只读存储器(ROM)702中的程序或者从存储部分708加载到随机访问存储器(RAM)703中的程序而执行各种适当的动作和处理。在RAM 703中,还存储有系统700操作所需的各种程序和数据。CPU 701、ROM 702以及RAM 703通过总线704彼此相连。输入/输出(I/O)接口705也连接至总线704。

[0080] 以下部件连接至I/O接口705:包括键盘、鼠标等的输入部分706;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707;包括硬盘等的存储部分708;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709。通信部分709经由诸如因特网的网络执行通信处理。驱动器710也根据需要连接至I/O接口705。可拆卸介质711,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710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708。

[0081] 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,其包括有形地包含在机器可读介质上的计算机程序,所述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709从网络上被下载和安装,和/或从可拆卸介质711被安装。在该计算机程序被中央处理单元603(CPU)701执行时,执行本申请的方法中限定的上述功能。

[0082] 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,所述模块、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。

[0083] 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、解析单元、处理单元和创建单元。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“获取一个型号的交换机的原始日志的单元”。

[0084] 作为另一方面,本申请还提供了一种非易失性计算机存储介质,该非易失性计算机存储介质可以是上述实施例中所述装置中所包含的非易失性计算机存储介质;也可以是单独存在,未装配入终端中的非易失性计算机存储介质。上述非易失性计算机存储介质存储有一个或者多个程序,当所述一个或者多个程序被一个设备执行时,使得所述设备:获取一个型号的交换机的原始日志;获取所述原始日志中的消息类型和详细消息,如果获取不

到消息类型,则创建聚类标签作为消息类型;对所述详细消息进行分词,得到关键词;根据所述关键词的词频对所述关键词重新排序,将词频高的关键词排在前面;根据所述重新排序的关键词创建多叉树,每个消息类型作为树的根节点,所述重新排序的关键词作为树的节点,每个消息类型对应一个多叉树;深度优先遍历所述多叉树,根据所述多叉树中的路径创建所述型号交换机的每个消息类型所对应的日志模板。

[0085] 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所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

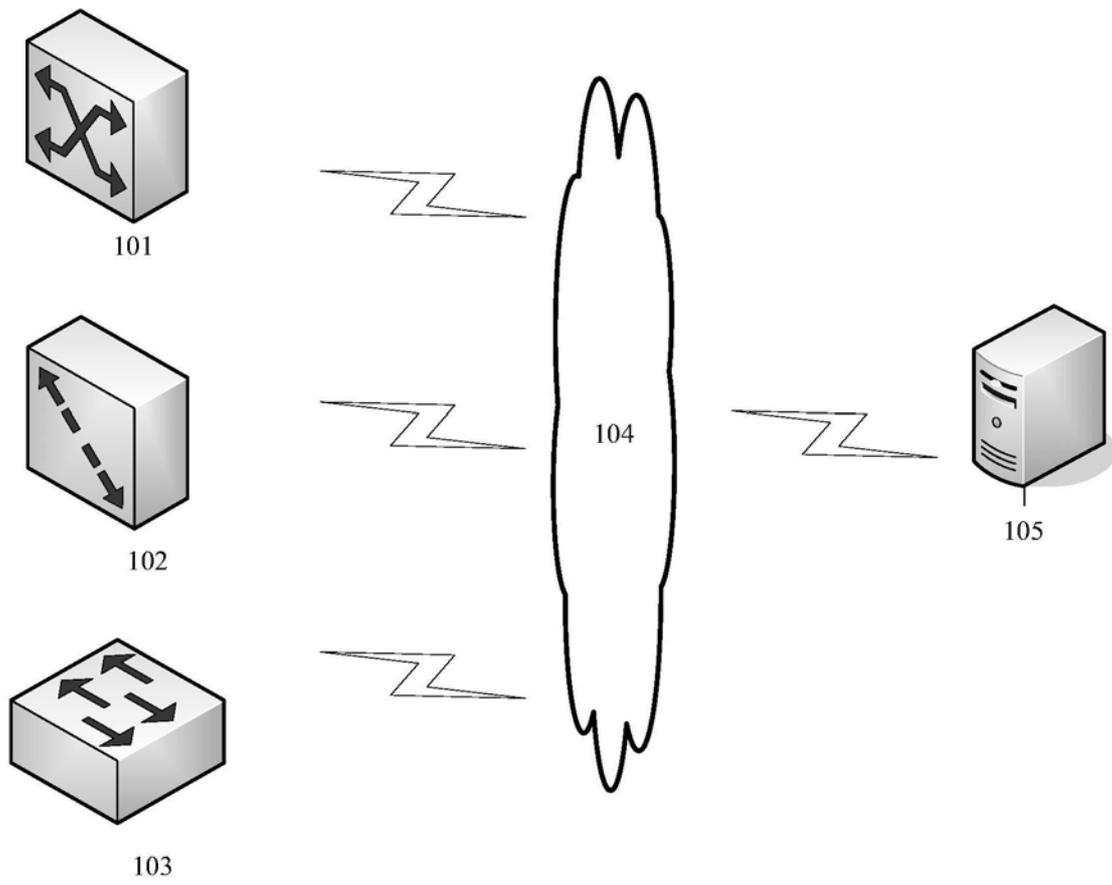


图1

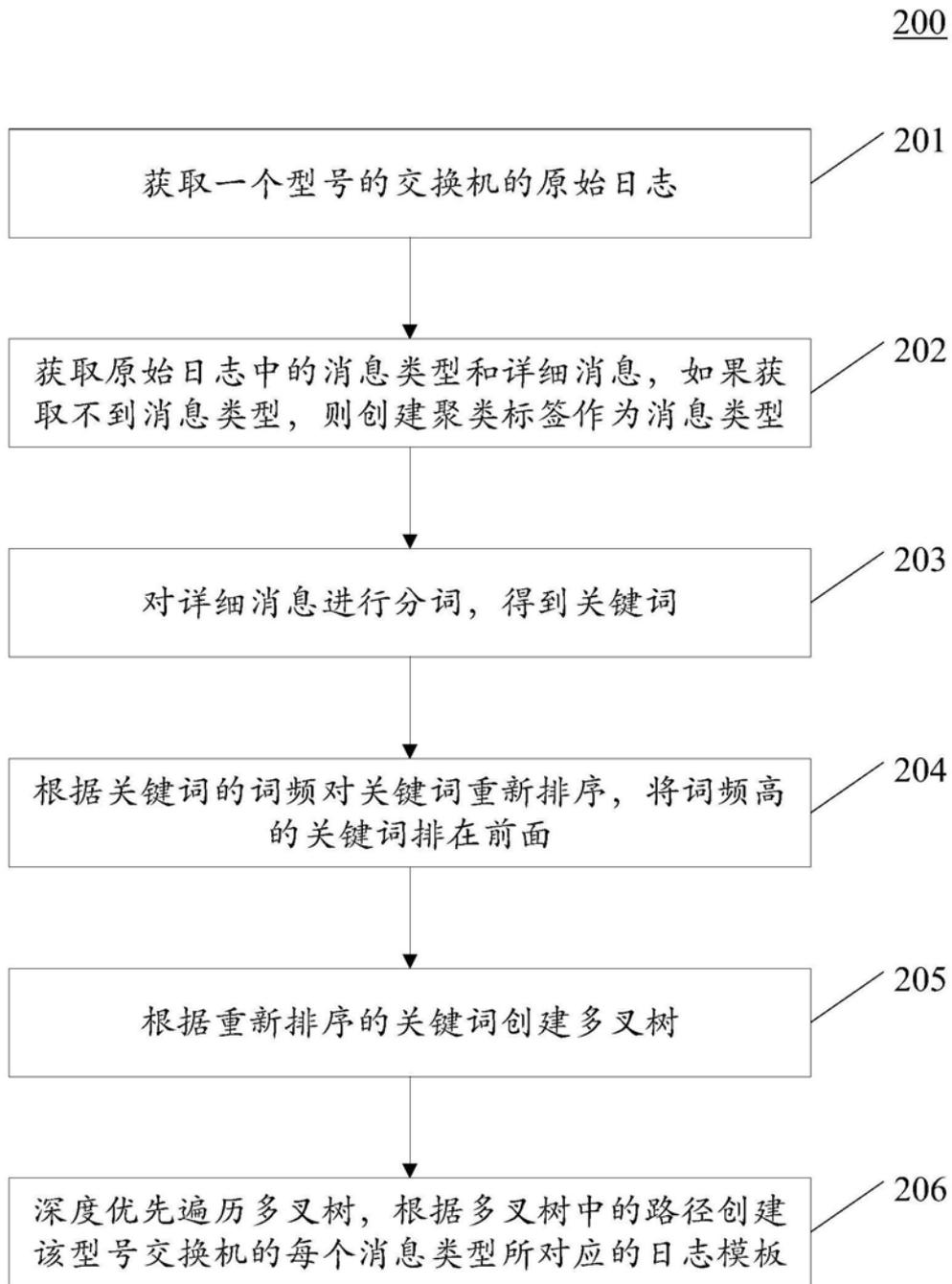


图2

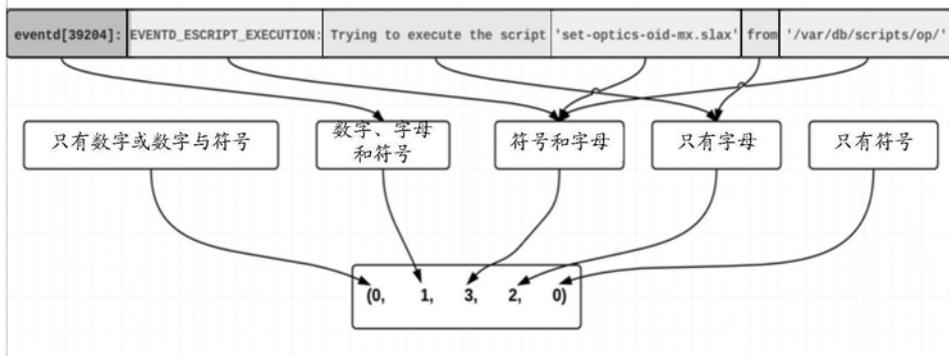


图3

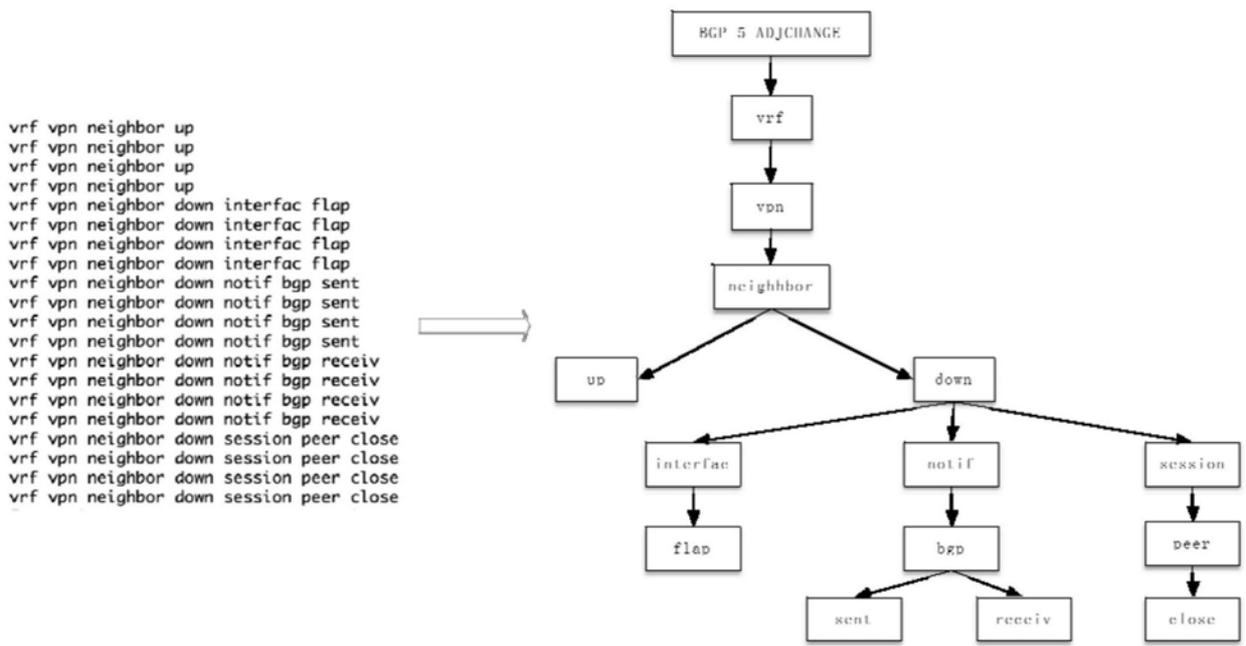


图4a

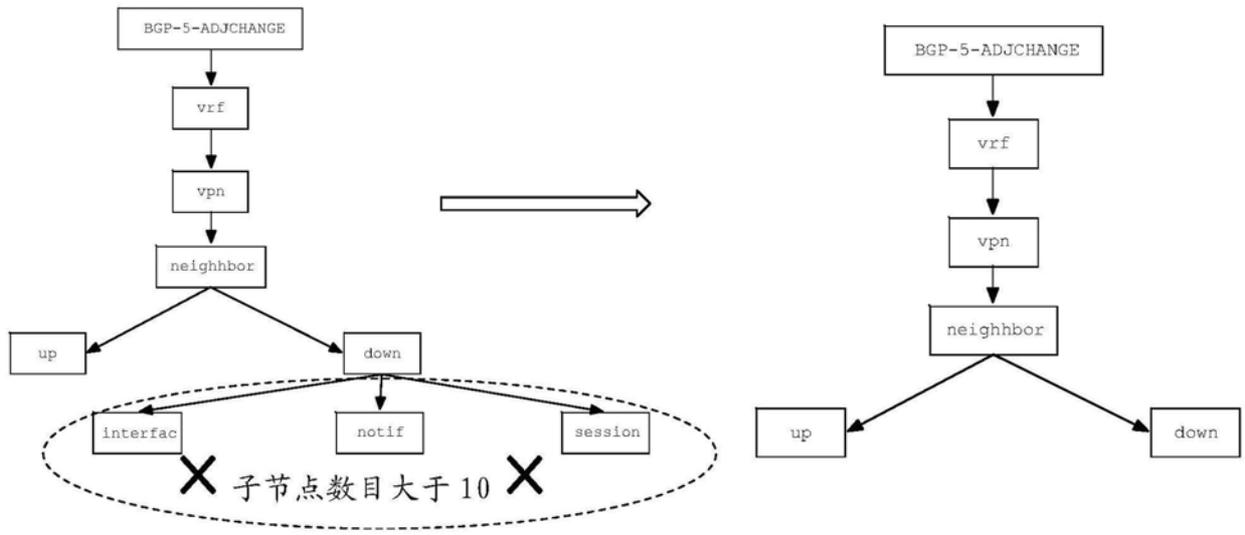


图4b

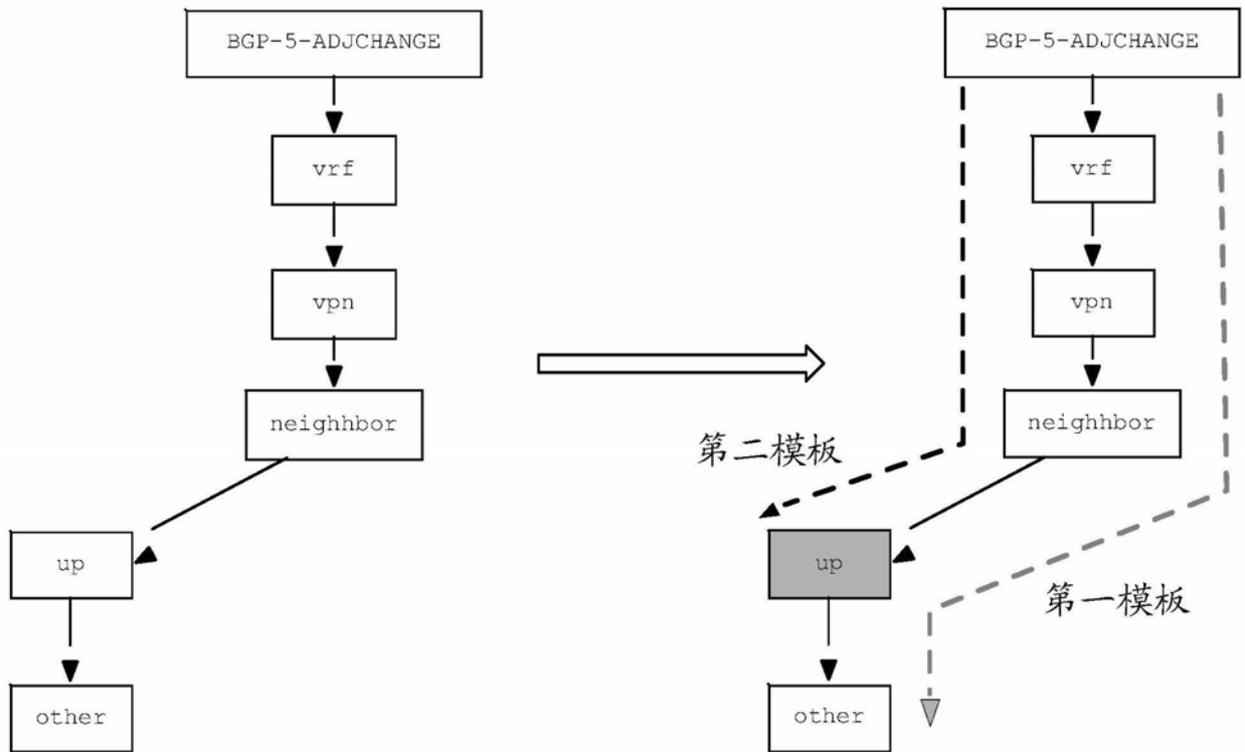


图4c

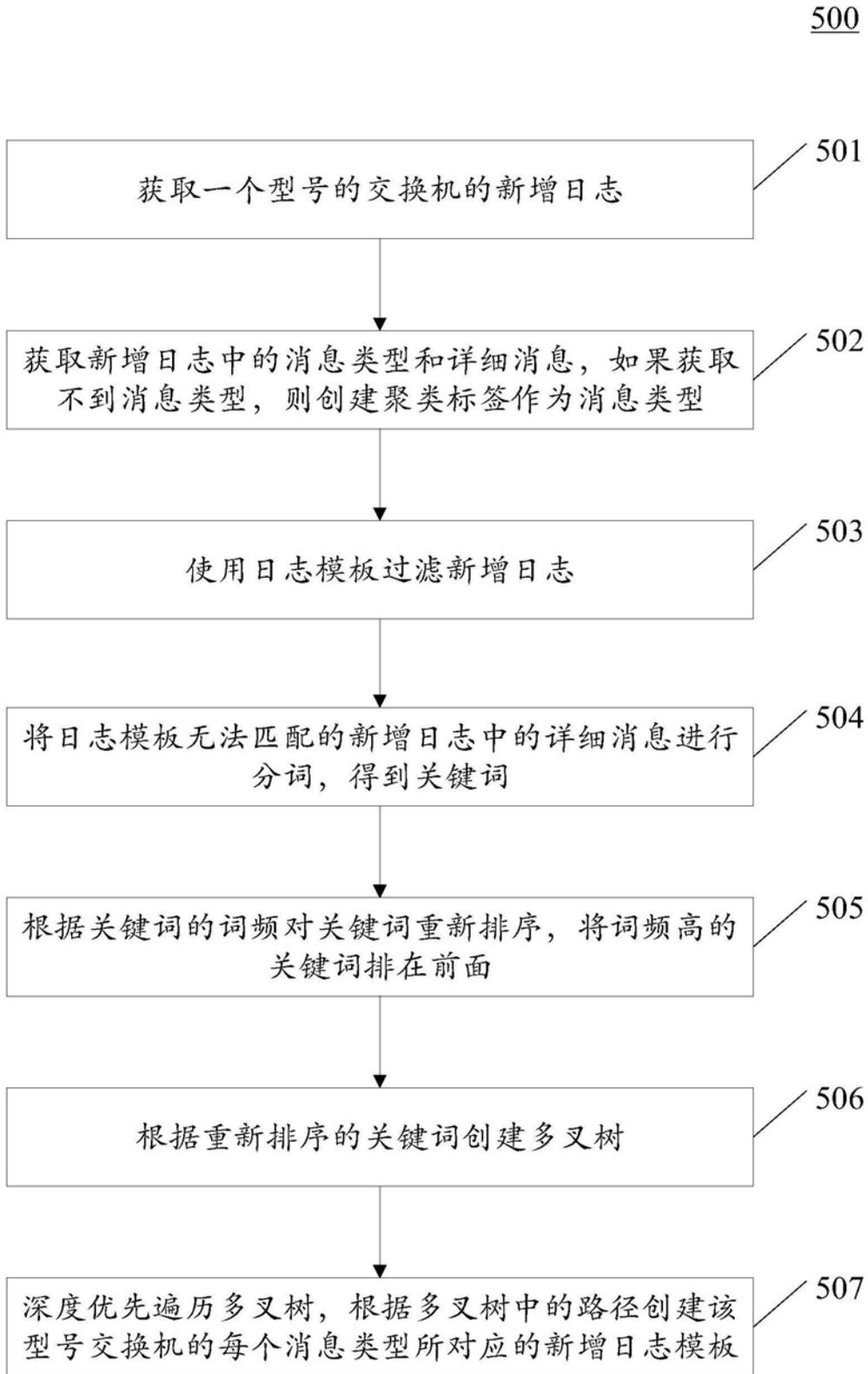


图5

600

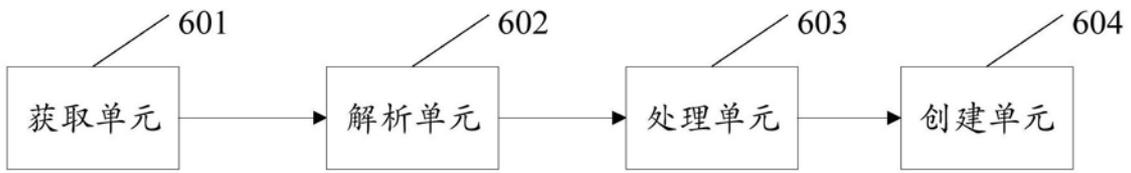


图6

700

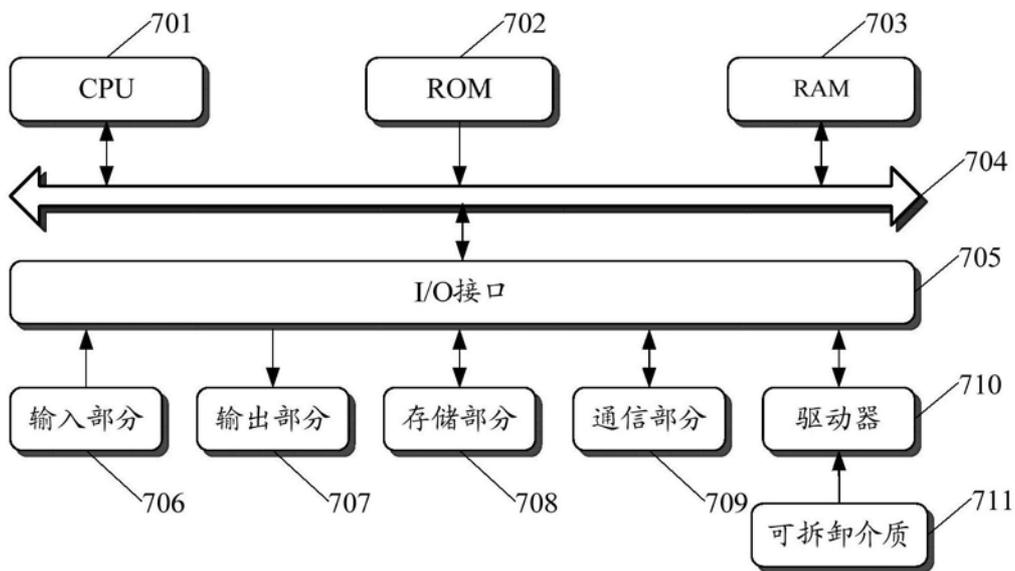


图7