



(12) 发明专利

(10) 授权公告号 CN 112737799 B

(45) 授权公告日 2023. 03. 28

(21) 申请号 201910975285.3

H04L 41/0681 (2022.01)

(22) 申请日 2019.10.14

H04L 41/0677 (2022.01)

(65) 同一申请的已公布的文献号

H04L 43/067 (2022.01)

申请公布号 CN 112737799 A

H04W 24/04 (2009.01)

H04L 41/0654 (2022.01)

(43) 申请公布日 2021.04.30

(56) 对比文件

(73) 专利权人 中国移动通信有限公司研究院

CN 109358602 A, 2019.02.19

地址 100053 北京市西城区宣武门西大街
32号

CN 109992484 A, 2019.07.09

专利权人 中国移动通信集团有限公司

CN 108737147 A, 2018.11.02

CN 110245168 A, 2019.09.17

(72) 发明人 李光宇 邓娟 周岩 吕永霖

CN 105844291 A, 2016.08.10

US 2015096043 A1, 2015.04.02

(74) 专利代理机构 北京派特恩知识产权代理有
限公司 11270

审查员 张琦

专利代理师 高洁 张颖玲

(51) Int. Cl.

H04L 41/06 (2022.01)

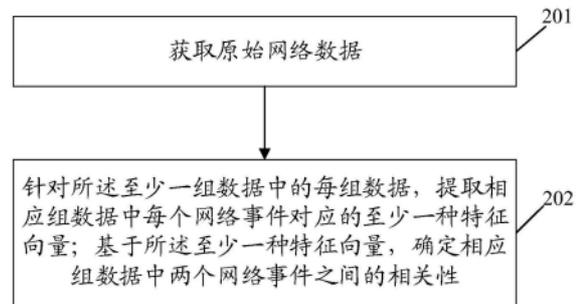
权利要求书2页 说明书12页 附图5页

(54) 发明名称

一种数据处理方法、装置及存储介质

(57) 摘要

本发明公开了一种数据处理方法、装置及存储介质。其中,方法包括:获取原始网络数据;对所述原始网络数据中的网络事件进行组合,得到至少一组数据;所述至少一组数据中的每组数据包含两个网络事件;针对所述至少一组数据中的每组数据,提取相应组数据中每个网络事件对应的至少一种特征向量;基于所述至少一种特征向量,确定相应组数据中两个网络事件之间的相关性;满足预设条件的用于在相应两个网络事件发生异常时进行告警。



1. 一种数据处理方法,其特征在于,所述方法包括:
 - 获取原始网络数据;
 - 从所述原始网络数据中选取与至少一个滑动窗口对应的至少两个网络事件;
 - 将与所述至少一个滑动窗口中每个滑动窗口对应的网络事件组成数据集合,得到至少一个数据集合;
 - 从所述至少一个数据集合中选取包含两个网络事件的数据集合,将选取的数据集合作为第一数据集合;
 - 统计所述第一数据集合中两个网络事件在第二数据集合中出现的总次数;所述第二数据集合为所述至少一个数据集合中除所述第一数据集合外的其它数据集合;
 - 当统计的总次数大于次数阈值时,将相应第一数据集合中事件发生频次满足预设条件的两个网络事件进行组合,得到至少一组数据;所述至少一组数据中的每组数据包含两个网络事件;
 - 针对所述至少一组数据中的每组数据,提取相应组数据中每个网络事件对应的至少一种特征向量;
 - 基于所述至少一种特征向量,确定相应组数据中两个网络事件之间的相关性;满足预设条件的相关性用于在相应两个网络事件发生异常时进行告警。
2. 根据权利要求1所述的方法,其特征在于,所述提取相应组数据中每个网络事件对应的至少一种特征向量,包括:
 - 针对相应组数据中每个网络事件,确定待提取特征向量的至少一种类别;
 - 针对所述至少一种类别中每个类别,当相应组数据中每个网络事件的数据中包含与相应类别对应的第一数据时,提取所述第一数据的特征向量,得到至少一种特征向量。
3. 根据权利要求2所述的方法,其特征在于,所述方法还包括:
 - 当相应组数据中每个历史网络告警事件的数据中未包含与相应类别对应的第一数据时,基于相应组数据中每个历史网络告警事件的数据包含的第二数据,得到与相应类别对应的第一数据;
 - 提取所述第一数据的特征向量,得到至少一种特征向量。
4. 根据权利要求1所述的方法,其特征在于,所述基于所述至少一种特征向量,确定相应组数据中两个网络事件之间的相关性,包括:
 - 按照特征向量的类别,对相应组数据中两个网络事件分别对应的至少一种特征向量进行排序,得到排序结果;
 - 对所述排序结果中类别相同的两个特征向量进行相关性计算,得到至少一个相关性结果;
 - 将所述至少一个相关性结果中每个相关性结果和对应的预设权重进行加权计算,得到相应组数据中两个网络事件之间的相关性。
5. 根据权利要求4所述的方法,其特征在于,所述对所述排序结果中类别相同的两个特征向量进行相关性计算,包括:
 - 针对所述排序结果中类别相同的两个特征向量,确定与所述类别匹配的相关性计算方式;
 - 基于确定的相关性计算方式,对所述排序结果中类别相同的两个特征向量进行相关性

计算。

6. 根据权利要求1所述的方法,其特征在于,所述方法还包括:

当相应两个网络事件发生异常时,若对应的相关性大于相关性门限,则对相应两个网络事件进行告警。

7. 根据权利要求2所述的方法,其特征在于,所述针对相应组数据中每个网络事件,确定待提取特征向量的至少一种类别,包括:

接收第一操作;

基于所述第一操作,确定待提取特征向量的至少一种类别。

8. 一种数据处理装置,其特征在于,所述装置包括:

获取单元,用于获取原始网络数据;

组合单元,用于从所述原始网络数据中选取与至少一个滑动窗口对应的至少两个网络事件;将与所述至少一个滑动窗口中每个滑动窗口对应的网络事件组成数据集合,得到至少一个数据集合;从所述至少一个数据集合中选取包含两个网络事件的数据集合,将选取的数据集合作为第一数据集合;统计所述第一数据集合中两个网络事件在第二数据集合中出现的总次数;所述第二数据集合为所述至少一个数据集合中除所述第一数据集合外的其它数据集合;当统计的总次数大于次数阈值时,将相应第一数据集合中事件发生频次满足预设条件的两个网络事件进行组合,得到至少一组数据;所述至少一组数据中的每组数据包含两个网络事件;

确定单元,用于针对所述至少一组数据中的每组数据,提取相应组数据中每个网络事件对应的至少一种特征向量;基于所述至少一种特征向量,确定相应组数据中两个网络事件之间的相关性;满足预设条件的用于在相应两个网络事件发生异常时进行告警。

9. 一种数据处理装置,其特征在于,所述装置包括:

通信接口,用于获取原始网络数据;

处理器,用于从所述原始网络数据中选取与至少一个滑动窗口对应的至少两个网络事件;将与所述至少一个滑动窗口中每个滑动窗口对应的网络事件组成数据集合,得到至少一个数据集合;从所述至少一个数据集合中选取包含两个网络事件的数据集合,将选取的数据集合作为第一数据集合;统计所述第一数据集合中两个网络事件在第二数据集合中出现的总次数;所述第二数据集合为所述至少一个数据集合中除所述第一数据集合外的其它数据集合;当统计的总次数大于次数阈值时,将相应第一数据集合中事件发生频次满足预设条件的两个网络事件进行组合,得到至少一组数据;所述至少一组数据中的每组数据包含两个网络事件;以及针对所述至少一组数据中的每组数据,提取相应组数据中每个网络事件对应的至少一种特征向量;基于所述至少一种特征向量,确定相应组数据中两个网络事件之间的相关性;满足预设条件的用于在相应两个网络事件发生异常时进行告警。

10. 一种数据处理装置,包括存储器、处理器及存储在存储器上并可在处理器上运行的计算机程序,其特征在于,所述处理器执行所述程序时实现权利要求1至7任一项所述方法的步骤。

11. 一种计算机存储介质,其上存储有计算机指令,其特征在于,所述指令被处理器执行时实现权利要求1至7任一项所述方法的步骤。

一种数据处理方法、装置及存储介质

技术领域

[0001] 本发明涉及无线网络领域,尤其涉及一种数据处理方法、装置及存储介质。

背景技术

[0002] 随着移动网络的快速发展,网络结构越来越复杂,网络中的硬件设备和软件可能会产生大量的网络事件,如告警事件。实际应用时,可以从产生的网络事件中确定具有关联关系的网络事件,利用具有关联关系的网络事件进行故障分析、网络告警等等。目前,针对事件发生时间间隔较短的网络事件,通过计算频繁度,判断网络事件之间是否关联。

[0003] 上述方式中基于频繁度确定具有关联的网络事件,由于网络事件发生具有随机性,因此基于频繁度计算网络事件之间的关联度,可能导致将没有关联关系的网络事件判定为具有关联关系,计算精度较差。

发明内容

[0004] 基于此,本发明实施例期望提供一种数据处理方法、装置及存储介质。

[0005] 本发明实施例的技术方案是这样实现的:

[0006] 本发明实施例提供一种数据处理方法,所述方法包括:

[0007] 获取原始网络数据;

[0008] 对所述原始网络数据中的网络事件进行组合,得到至少一组数据;所述至少一组数据中的每组数据包含两个网络事件;

[0009] 针对所述至少一组数据中的每组数据,提取相应组数据中每个网络事件对应的至少一种特征向量;

[0010] 基于所述至少一种特征向量,确定相应组数据中两个网络事件之间的相关性;满足预设条件的关联性用于在相应两个网络事件发生异常时进行告警。

[0011] 上述方案中,所述提取相应组数据中每个网络事件对应的至少一种特征向量,包括:

[0012] 针对相应组数据中每个网络事件,确定待提取特征向量的至少一种类别;

[0013] 针对所述至少一种类别中每个类别,当相应组数据中每个网络事件的数据中包含与相应类别对应的第一数据时,提取所述第一数据的特征向量,得到至少一种特征向量。

[0014] 上述方案中,所述方法还包括:

[0015] 当相应组数据中每个历史网络告警事件的数据中未包含与相应类别对应的第一数据时,基于相应组数据中每个历史网络告警事件的数据包含的第二数据,得到与相应类别对应的第一数据;

[0016] 提取所述第一数据的特征向量,得到至少一种特征向量。

[0017] 上述方案中,所述对所述原始网络数据中的网络事件进行组合,得到至少一组数据,包括:

[0018] 从所述原始网络数据中选取与至少一个滑动窗口对应的至少两个网络事件;

[0019] 从选取的至少两个网络事件中选取事件发生频次满足预设条件的两个历史网络告警事件进行组合,得到至少一组数据。

[0020] 上述方案中,所述从选取的至少两个网络事件中选取事件发生频次满足预设条件的两个网络事件进行组合,包括:

[0021] 将与所述至少一个滑动窗口中每个滑动窗口对应的网络事件组成数据集合,得到至少一个数据集合;

[0022] 从所述至少一个数据集合中选取包含两个网络事件的数据集合,将选取的数据集合作为第一数据集合;

[0023] 统计所述第一数据集合中两个网络事件在第二数据集合中出现的总次数;所述第二数据集合为所述至少一个数据集合中除所述第一数据集合外的其它数据集合;

[0024] 当统计的总次数大于次数阈值时,将相应第一数据集合中事件发生频次满足预设条件的两个网络事件进行组合。

[0025] 上述方案中,所述对所述原始网络数据中的网络事件进行组合,得到至少一组数据,包括:

[0026] 将所述原始网络数据中任意两个网络事件进行组合,得到至少一组数据。

[0027] 上述方案中,所述基于所述至少一种特征向量,确定相应组数据中两个网络事件之间的相关性,包括:

[0028] 按照特征向量的类别,对相应组数据中两个网络事件分别对应的至少一种特征向量进行排序,得到排序结果;

[0029] 对所述排序结果中类别相同的两个特征向量进行相关性计算,得到至少一个相关性结果;

[0030] 将所述至少一个相关性结果中每个相关性结果和对应的预设权重进行加权计算,得到相应组数据中两个网络事件之间的相关性。

[0031] 上述方案中,所述对所述排序结果中类别相同的两个特征向量进行相关性计算,包括:

[0032] 针对所述排序结果中类别相同的两个特征向量,确定与所述类别匹配的相关性计算方式;

[0033] 基于确定的相关性计算方式,对所述排序结果中类别相同的两个特征向量进行相关性计算。

[0034] 上述方案中,所述方法还包括:

[0035] 当相应两个网络事件发生异常时,若对应的相关性大于相关性门限,则对相应两个网络事件进行告警。

[0036] 上述方案中,所述针对相应组数据中每个网络事件,确定待提取特征向量的至少一种类别,包括:

[0037] 接收第一操作;

[0038] 基于所述第一操作,确定待提取特征向量的至少一种类别。

[0039] 本发明实施例提供一种数据处理装置,所述装置包括:

[0040] 获取单元,用于获取原始网络数据;

[0041] 组合单元,用于对所述原始网络数据中的网络事件进行组合,得到至少一组数据;

所述至少一组数据中的每组数据包含两个网络事件；

[0042] 确定单元,用于针对所述至少一组数据中的每组数据,提取相应组数据中每个网络事件对应的至少一种特征向量;基于所述至少一种特征向量,确定相应组数据中两个网络事件之间的相关性;满足预设条件的关联性用于在相应两个网络事件发生异常时进行告警。

[0043] 本发明实施例提供一种数据处理装置,所述装置包括:

[0044] 通信接口,用于获取原始网络数据;

[0045] 处理器,用于对所述原始网络数据中的网络事件进行组合,得到至少一组数据;所述至少一组数据中的每组数据包含两个网络事件;以及针对所述至少一组数据中的每组数据,提取相应组数据中每个网络事件对应的至少一种特征向量;基于所述至少一种特征向量,确定相应组数据中两个网络事件之间的相关性;满足预设条件的关联性用于在相应两个网络事件发生异常时进行告警。

[0046] 本发明实施例提供一种数据处理装置,包括存储器、处理器及存储在存储器上并可在处理器上运行的计算机程序,所述处理器执行所述程序时实现上述任一所述方法的步骤。

[0047] 本发明实施例提供一种计算机存储介质,其上存储有计算机指令,所述指令被处理器执行时实现上述任一所述方法的步骤。

[0048] 本发明实施例提供的数据处理方法、装置及存储介质,获取原始网络数据;对所述原始网络数据中的网络事件进行组合,得到至少一组数据;所述至少一组数据中的每组数据包含两个网络事件;针对所述至少一组数据中的每组数据,提取相应组数据中每个网络事件对应的至少一种特征向量;基于所述至少一种特征向量,确定相应组数据中两个网络事件之间的相关性;满足预设条件的关联性用于在相应两个网络事件发生异常时进行告警。采用本发明实施例的技术方案,对原始网络数据进行分组,提取每组数据中两个网络事件的特征向量,计算每组数据中两个网络事件对应的特征向量的相关性,并根据计算的相关性结果判定两个网络事件之间的关联关系,从而使关联分析结果更准确。

附图说明

[0049] 图1为相关技术中确定两个网络事件的关联度的实现流程示意图;

[0050] 图2为本发明实施例数据处理方法的实现流程示意图;

[0051] 图3为本发明实施例数据处理方法的具体实现过程示意图;

[0052] 图4a为本发明实施例网络事件A的时间分布示意图一;

[0053] 图4b为本发明实施例网络事件B的时间分布示意图一;

[0054] 图5a为本发明实施例网络事件A的时间分布示意图二;

[0055] 图5b为本发明实施例网络事件B的时间分布示意图二;

[0056] 图6为本发明实施例数据处理装置的组成结构示意图一;

[0057] 图7为本发明实施例数据处理装置的组成结构示意图二。

具体实施方式

[0058] 为了能够更加详尽地了解本发明实施例的特点与技术内容,下面结合附图对本发

明实施例的实现进行详细阐述,所附附图仅供参考说明之用,并非用来限定本发明。

[0059] 相关技术中,随着移动网络的快速发展,无线网络结构越来越复杂,无线网络中的硬件设备和软件可能会产生大量的网络事件,如告警事件。实际应用时,可以从产生的网络事件中确定具有关联关系的网络事件,利用具有关联关系的网络事件进行故障分析、网络告警等等。目前,没有统一的方法来分析网络事件之间的关联性,通常可以通过滑动窗口对网络事件进行时间分割,再对时间间隔满足条件的网络事件计算频繁度,判断网络事件之间是否具有关联关系。图1是相关技术中确定两个网络事件的关联度的实现流程示意图,如图1所示,包括:

[0060] 步骤1:收集无线网络中已经发生的网络事件对应的原始网络数据,所述原始网络数据至少包含告警名称、告警发生时间、告警恢复时间、告警发生位置等。

[0061] 步骤2:对网络事件按照预设规则进行排序,例如,按照网络事件的发生时间进行排序,或者按照网络事件的发生位置进行排序。

[0062] 步骤3:将排序后的网络事件划分为发生时间间隔满足预设条件的网络事件集合。例如,利用长度为n的滑动窗口,从按事件发生时间排序的网络事件中过滤出事件发生时间间隔相差不超过n秒的所有网络事件,假设产生4个网络事件集合:[A,B,C],[A,C],[B,D],[B,C,D]。

[0063] 步骤4:针对网络事件集合中每个子集合,统计每个子集合中所有网络事件在其它子集合中出现的频繁度,即针对每个网络集合中数量大于1的子集合,分别统计该子集合中所有网络事件在其它子集合中是否出现、在所有集合中的出现次数。假设网络事件集合包含的子集有:[A,B],[A,C],[B,C],[B,D],[C,D],[A,B,C],[B,C,D]。

[0064] 步骤5:筛选出频繁度满足预设条件的子集合。例如,以50%为频繁度门限,筛选的子集合为[A,C],[B,C],[B,D],其中,网络事件A和C之间具有关联关系,网络事件B和C之间具有关联关系,网络事件B和D之间具有关联关系。

[0065] 上述方式中基于频繁度计算网络事件之间的关联度,至少包含以下缺陷:一、精确度较差:由于网络事件发生具有随机性,可能导致将没有关联关系的网络事件判定为具有关联关系,计算精度较差。二、可能遗漏关联关系:由于不同厂家的网络设备的告警上报方式不同,加上网络延时、网络结构复杂等因素,网络事件的发生时间可能存在偏差,时间间隔较短的网络事件之间未必具有关联关系,而时间间隔较大的网络事件之间可能具有关联关系。如此,仅通过网络事件的发生时间和网络事件发生的频繁度判断网络事件之间是否具有关联关系,使得至少两类网络事件可能会被忽略,一类是事件发生时间相隔较大的网络事件,另一类是网络事件发生频繁度不满足条件的网络事件。

[0066] 基于此,本发明实施例中,获取原始网络数据;对所述原始网络数据中的网络事件进行组合,得到至少一组数据;所述至少一组数据中的每组数据包含两个网络事件;针对所述至少一组数据中的每组数据,提取相应组数据中每个网络事件对应的至少一种特征向量;基于所述至少一种特征向量,确定相应组数据中两个网络事件之间的相关性;满足预设条件的相关性用于在相应两个网络事件发生异常时进行告警。

[0067] 图2是本发明实施例数据处理方法的实现流程示意图;如图2所示,所述方法包括:

[0068] 步骤201:获取原始网络数据;对所述原始网络数据中的网络事件进行组合,得到至少一组数据;所述至少一组数据中的每组数据包含两个网络事件;

[0069] 这里,所述原始网络数据可以是指无线网络中的设备或系统发生故障进行告警时产生的网络事件对应的数据,如告警名称、告警发生时间、告警恢复时间、告警发生位置等。

[0070] 实际应用时,考虑到事件发生时间相隔较短且事件发生频次较多的两个网络事件之间可能具有关联关系,因此,可以从原始网络数据中选取事件发生时间相隔较短且事件发生频次较多的两个网络事件进行组合,作为待分析数据。

[0071] 基于此,在一实施例中,所述对所述原始网络数据中的网络事件进行组合,包括:从所述原始网络数据中选取与至少一个滑动窗口对应的至少两个网络事件;从选取的至少两个网络事件中选取事件发生频次满足预设条件的两个历史网络告警事件进行组合,得到至少一组数据。

[0072] 这里,所述滑动窗口可以是根据所述网络事件的发生时间确定的时间窗口,也可以是根据所述网络事件的发生位置确定的距离窗口。

[0073] 这里,当所述滑动窗口为时间窗口时,可以对所述原始网络数据中的网络事件按照事件的发生时间进行排序,得到排序后的事件序列;从所述事件序列中选取与至少一个时间窗口对应的至少两个网络事件。每个时间窗口可以表征发生时间间隔满足预设条件的至少两个网络事件。

[0074] 这里,当所述滑动窗口为距离窗口时,可以对所述原始网络数据中的网络事件按照事件的发生时位置进行排序,得到排序后的事件序列;从所述事件序列中选取与至少一个距离窗口对应的至少两个网络事件。每个距离窗口可以表征发生位置间隔满足预设条件的至少两个网络事件。

[0075] 进一步地,在一实施例中,所述从选取的至少两个网络事件中选取事件发生频次满足预设条件的两个网络事件进行组合,包括:将与所述至少一个滑动窗口中每个滑动窗口对应的网络事件组成数据集合,得到至少一个数据集合;从所述至少一个数据集合中选取包含两个网络事件的数据集合,将选取的数据集合作为第一数据集合;统计所述第一数据集合中两个网络告警事件在第二数据集合中出现的总次数;所述第二数据集合为所述至少一个数据集合中除所述第一数据集合外的其它数据集合;当统计的总次数大于次数阈值时,将相应第一数据集合中事件发生频次满足预设条件的两个网络事件进行组合。

[0076] 实际应用时,考虑到不同厂家的网络设备的告警上报方式不同、网络延时、网络结构复杂等因素,网络事件的发生时间可能存在偏差,时间间隔较短且发生频次较多的网络事件之间未必具有关联关系,而时间间隔较大的网络事件之间可能具有关联关系。如此,可以从原始网络数据中选取任意两个网络事件进行组合,作为待分析数据。

[0077] 基于此,在一实施例中,所述对所述原始网络数据中的网络事件进行组合,得到至少一组数据,包括:将所述原始网络数据中任意两个网络事件进行组合,得到至少一组数据。

[0078] 这里,可以将所述原始网络数据中任意两个类型不同的网络事件进行组合,得到至少一组数据。

[0079] 这里,从原始网络数据中选取任意两个网络事件进行组合,如此,能够避免遗漏可能具有关联关系的网络事件的发生。

[0080] 步骤202:针对所述至少一组数据中的每组数据,提取相应组数据中每个网络事件对应的至少一种特征向量;基于所述至少一种特征向量,确定相应组数据中两个网络事件

之间的相关性。

[0081] 这里,满足预设条件的相关性用于在相应两个网络事件发生异常时进行告警。

[0082] 实际应用时,不同类型的网络事件确定的特征向量的类别可以不同,比如,针对类型为a的网络事件A和B,需要确定三种类别的特征向量,才能够确定出网络事件A和B之间是否具有关联关系。而针对类型为b的网络事件C和D,需要确定五种类别的特征向量,才能够确定出网络事件C和D之间是否具有关联关系。

[0083] 基于此,在一实施例中,所述提取相应组数据中每个网络事件对应的至少一种特征向量,包括:针对相应组数据中每个网络事件,确定待提取特征向量的至少一种类别;针对所述至少一种类别中每个类别,当相应组数据中每个网络事件的数据中包含与相应类别对应的第一数据时,提取所述第一数据的特征向量,得到至少一种特征向量。

[0084] 这里,所述第一数据可以是网络事件对应的告警发生时间、告警发生位置、告警恢复时间中之一的数据。

[0085] 举例来说,当类别表征网络事件的发生时间时,假设网络事件的发生时间为2019年9月22日,则所述第一数据可以是2019年9月22日对应的数据;当类别表征网络事件的发生位置时,假设网络事件的发生位置为位于a大厦的网络设备A,则所述第一数据可以是a大厦中网络设备a的地理位置对应的数据。

[0086] 进一步地,在一实施例中,所述方法还包括:当相应组数据中每个历史网络告警事件的数据中未包含与相应类别对应的第一数据时,基于相应组数据中每个历史网络告警事件的数据包含的第二数据,得到与相应类别对应的第一数据;提取所述第一数据的特征向量,得到至少一种特征向量。

[0087] 这里,所述第一数据可以是指网络事件对应的告警发生频次等数据。

[0088] 这里,所述第二数据可以是指所述相应组数据中包含的告警发生时间、告警发生位置、告警恢复时间中之一的数据。

[0089] 实际应用时,可以针对每种类型的网络事件预先设置待提取特征向量的类别,也可以借助用户的经验,确定每个网络事件对应的特征向量的类别。

[0090] 进一步地,在一实施例中,所述针对相应组数据中每个网络事件,确定待提取特征向量的至少一种类别,包括:接收第一操作;基于所述第一操作,确定待提取特征向量的至少一种类别。

[0091] 这里,所述第一操作可以是指用户的输入操作;所述输入操作用于指示待特征向量的至少一种类别。

[0092] 这里,所述第一操作对应的内容可以是待提取特征向量的类别对应的编号,也可以是待提取特性向量的类别。

[0093] 这里,所述类别可以表征网络事件的发生时间、发生位置、发生频次等等。

[0094] 实际应用时,针对相应组数据中两个网络事件,对于提取的类别相同的特征向量,可以使用同一种相关性计算方式,计算两个特征向量之间的相关性结果。

[0095] 基于此,在一实施例中,所述基于所述至少一种特征向量,确定相应组数据中两个网络事件之间的相关性,包括:按照特征向量的类别,对相应组数据中两个网络事件分别对应的至少一种特征向量进行排序,得到排序结果;对所述排序结果中类别相同的两个特征向量进行相关性计算,得到至少一个相关性结果;将所述至少一个相关性结果中每个相关

性结果和对应的预设权重进行加权计算,得到相应组数据中两个网络事件之间的相关性。

[0096] 这里,所述两个特征向量的类别相同可以是指两个特征向量的取值范围相同,也可以是指两个特征向量的数据格式相同。

[0097] 进一步地,在一实施例中,所述对所述排序结果中类别相同的两个特征向量进行相关性计算,包括:针对所述排序结果中类别相同的两个特征向量,确定与所述类别匹配的相关性计算方式;基于确定的相关性计算方式,对所述排序结果中类别相同的两个特征向量进行相关性计算。

[0098] 这里,所述相关性计算方式可以包括互相关、卷积、皮尔森系数等计算方式。

[0099] 进一步地,在一实施例中,所述方法还包括:当相应两个网络事件发生异常时,若对应的相关性大于相关性门限,则对相应两个网络事件进行告警。

[0100] 这里,本发明实施例中,能够将相关性计算技术应用到无线网络事件的关联分析中,通过计算特征向量的相关性,可以分析两个网络事件之间相互影响的程度,也可以预测两个网络事件的变化趋势。

[0101] 这里,本发明实施例中,提出一种新的判定网络事件关联的方式即基于相关性分析网络事件之间的关联性,不仅关注网络事件的发生时间间隔,而且更关注网络事件的特征向量的时间分布、位置分布等,从而使关联分析结果更加准确。

[0102] 这里,本发明实施例中,提出一种在判断频繁度基础上,结合网络事件的时间、空间特征,例如发生时间、恢复时间、产生频繁度、产生位置等,计算不同网络事件之间上述特征向量的相关性,并根据计算的相关性结果判定网络事件之间的关联关系,从而能够使计算结果更准确,能够更精确地发现可能相关的网络事件。

[0103] 这里,本发明实施例中,将原始网络数据中任意两个网络事件进行组合,可以更全面的发现可能具有关联关系的网络事件组合。

[0104] 采用本发明实施例提供的技术方案,对原始网络数据进行分组,提取每组数据中两个网络事件的特征向量,计算每组数据中两个网络事件对应的特征向量的相关性,并根据计算的相关性结果判定两个网络事件之间的关联关系,从而使关联分析结果更准确。

[0105] 下面以具体实施例详细说明本发明实施例数据处理方法的实现原理。

[0106] 图3是本发明实施例数据处理方法的具体实现过程示意图,如图3所示,包括:

[0107] 步骤1:收集无线网络中已经发生的网络事件对应的原始网络数据,所述原始网络数据至少包含告警名称、告警发生时间、告警恢复时间、告警发生位置等。

[0108] 步骤2:对网络事件按照预设规则进行排序,例如,按照网络事件的发生时间进行排序,或者按照网络事件的发生位置进行排序。

[0109] 步骤3:将排序后的网络事件划分为发生时间间隔满足预设条件的网络事件集合。例如,利用长度为n的滑动窗口,从按事件发生时间排序的网络事件中过滤出事件发生时间间隔相差在n秒内的所有网络事件,假设产生4个网络事件集合:[A,B,C],[A,C],[B,D],[B,C,D]。

[0110] 步骤4:针对网络事件集合中每个子集合,统计每个子集合中所有网络事件在其它子集合中出现的频繁度,即针对每个网络集合中数量大于1的子集合,分别统计该子集合中所有网络事件在其它个子集合中是否出现、在所有集合中的出现次数。假设网络事件集合包含的子集有:[A,B],[A,C],[B,C],[B,D],[C,D],[A,B,C],[B,C,D]。

[0111] 步骤5:筛选出频繁度满足预设条件的子集合,作为待计算相关性的组合数据。例如,以50%为频繁度门限,筛选的子集合为[A,C],[B,C],[B,D]。

[0112] 步骤6:以筛选的子集合为[A,C]为例,假设需要提取n个特征向量,则为n个特征向量分配对应的n个权重。

[0113] 这里,n个特征向量可以用 f_1, f_2, \dots, f_n 表示,分别表征网络事件的告警发生时间、告警恢复时间、告警频繁度、告警发生位置等。n个权重可以用 w_1, w_2, \dots, w_n 表示,且满足 $w_1 + w_2 + \dots + w_n = 1$ 。

[0114] 这里,分配权重之后,依次判断网络事件的数据中是否包含与n个特征向量的类别对应的数据,若网络事件的数据中未包含与n个特征向量的类别对应的数据,通过网络事件的已有数据计算出与n个特征向量的类别对应的数据。

[0115] 例如,网络事件的数据中包含告警发生时间、告警恢复时间、告警发生位置,但未包含告警发生频次,则可以根据告警发生时间及告警发生位置等数据进行换算得到告警发生频次对应的数据。

[0116] 这里,以筛选的子集合为[A,C]为例,假设提取网络事件A的n个特征向量为: $f_{1A}, f_{2A}, \dots, f_{nA}$,提取网络事件C的n个特征向量为: $f_{1C}, f_{2C}, \dots, f_{nC}$ 。

[0117] 步骤7:根据特征向量的数据类型及特点,选取对应的相关性计算方法,如互相关、卷积、皮尔森系数等,并计算出各特征向量之间的相关性结果。

[0118] 这里,各特征向量之间的相关性结果可以用 c_1, c_2, \dots, c_n 表示。

[0119] 以筛选的子集合为[A,C]为例,首先,针对特征向量 f_{1A}, f_{1C} ,根据特征向量的数值范围或类型等选取对应的相关性计算方法,如皮尔森系数。然后,判断特征向量 f_{1A} 和 f_{1C} 的长度是否相等:若长度相等,直接计算 f_{1A} 和 f_{1C} 的皮尔森系数,得出相关性结果 c_1 ;若长度不等,例如 f_{1A} 长度为10, f_{1C} 长度为20,则分别计算 f_{1A} 与 f_{1C} 的第1-10个元素、2-11个元素、...、第11-20个元素的皮尔森系数,然后取均值作为相关性结果 c_1 ;依次计算特征向量 $f_2 \sim f_n$,重复上述步骤,得到各个相关性结果 $c_2 \sim c_n$ 。最后,根据各特征向量分配的权重 $w_1 \sim w_n$ 及相关性结果,加权算出网络事件组合[A,C]的相关性结果 C_{ac} 。

[0120] 步骤8:根据两个网络事件对应的相关性结果,判断两个网络事件之间的关联关系。

[0121] 以筛选的子集合为[A,C]为例,若两个网络事件对应的相关性结果高于相关性门限T,则网络事件A和C之间确实具有关联关系。

[0122] 举例来说,图4a为网络事件A的时间分布示意图,图4b为网络事件B的时间分布示意图,网络事件B的发生次数相对较少,样本点数较少。从图4a和4b可看出,网络事件B的发生时间与网络事件A的发生时间的间隔较远,如果使用滑动窗口分析,则无法发现网络事件A与网络事件B之间的任何关联关系,但从网络事件的发生时间变化规律进行分析,则可以发现网络事件B的时间分布规律与网络事件A的某一部分(如虚线框所示)的时间分布规律十分相似,也就是说,通过计算网络事件A和网络事件B的特征向量之间的相关性,可以判断出A与B存在关联。

[0123] 举例来说,图5a为网络事件A的时间分布示意图,图5b为网络事件B的时间分布示意图。从图5a和5b可看出,网络事件B的发生时间与网络事件A的发生时间的间隔较近,如果使用滑动窗口分析,则确定网络事件A与网络事件B之间具有关联关系,但通过计算网络事

件A和网络事件B的特征向量之间的相关性,可以判断出A与B不存在关联。

[0124] 这里,分析网络事件的关联关系时仅考虑事件发生时间间隔,无法全面、客观地评估网络事件之间的关联关系,还需要加入网络事件的特征分布的相关性信息。

[0125] 为实现本发明实施例数据处理方法,本发明实施例还提供一种数据处理装置,图6为本发明实施例数据处理装置的组成结构示意图;如图6所示,所述装置包括:

[0126] 获取单元61,用于获取原始网络数据;

[0127] 组合单元62,用于对所述原始网络数据中的网络事件进行组合,得到至少一组数据;所述至少一组数据中的每组数据包含两个网络事件;

[0128] 确定单元63,用于针对所述至少一组数据中的每组数据,提取相应组数据中每个网络事件对应的至少一种特征向量;基于所述至少一种特征向量,确定相应组数据中两个网络事件之间的相关性;满足预设条件的相关性用于在相应两个网络事件发生异常时进行告警。

[0129] 在一实施例中,确定单元63,具体用于:针对相应组数据中每个网络事件,确定待提取特征向量的至少一种类别;针对所述至少一种类别中每个类别,当相应组数据中每个网络事件的数据中包含与相应类别对应的第一数据时,提取所述第一数据的特征向量,得到至少一种特征向量。

[0130] 在一实施例中,所述确定单元63,具体用于:接收第一操作;基于所述第一操作,确定待提取特征向量的至少一种类别。

[0131] 在一实施例中,所述装置还包括:

[0132] 提取单元,用于当相应组数据中每个历史网络告警事件的数据中未包含与相应类别对应的第一数据时,基于相应组数据中每个网络告警事件的数据包含的第二数据,得到与相应类别对应的第一数据;提取所述第一数据的特征向量,得到至少一种特征向量。

[0133] 在一实施例中,所述组合单元62,具体用于:从所述原始网络数据中选取与至少一个滑动窗口对应的至少两个网络事件;从选取的至少两个网络事件中选取事件发生频次满足预设条件的两个网络告警事件进行组合,得到至少一组数据。

[0134] 在一实施例中,所述组合单元62,具体用于:将与所述至少一个滑动窗口中每个滑动窗口对应的网络事件组成数据集合,得到至少一个数据集合;从所述至少一个数据集合中选取包含两个网络事件的数据集合,将选取的数据集合作为第一数据集合;

[0135] 统计所述第一数据集合中两个历史网络告警事件在第二数据集合中出现的总次数;所述第二数据集合为所述至少一个数据集合中除所述第一数据集合外的其它数据集合;

[0136] 当统计的总次数大于次数阈值时,将相应第一数据集合中事件发生频次满足预设条件的两个网络事件进行组合。

[0137] 在一实施例中,所述组合单元62,具体用于:将所述原始网络数据中任意两个网络事件进行组合,得到至少一组数据。

[0138] 在一实施例中,所述确定单元63,具体用于:按照特征向量的类别,对相应组数据中两个网络事件分别对应的至少一种特征向量进行排序,得到排序结果;对所述排序结果中类别相同的两个特征向量进行相关性计算,得到至少一个相关性结果;将所述至少一个相关性结果中每个相关性结果和对应的预设权重进行加权计算,得到相应组数据中两个网

络事件之间的相关性。

[0139] 在一实施例中,所述确定单元63,具体用于:针对所述排序结果中类别相同的两个特征向量,确定与所述类别匹配的相关性计算方式;基于确定的相关性计算方式,对所述排序结果中类别相同的两个特征向量进行相关性计算。

[0140] 在一实施例中,所述装置还包括:

[0141] 告警单元,用于当相应两个网络事件发生异常时,若对应的相关性大于相关性门限,则对相应两个网络事件进行告警。

[0142] 实际应用时,所述获取单元61由所述数据处理装置中处理器结合通信接口实现;所述组合单元62、确定单元63、提取单元、告警单元可由所述数据处理装置中的处理器实现。

[0143] 需要说明的是:上述实施例提供的数据处理装置在进行参数谱提取时,仅以上述各程序模块的划分进行举例说明,实际应用中,可以根据需要而将上述处理分配由不同的程序模块完成,即将装置的内部结构划分成不同的程序模块,以完成以上描述的全部或者部分处理。另外,上述实施例提供的数据处理装置与数据处理方法实施例属于同一构思,其具体实现过程详见方法实施例,这里不再赘述。

[0144] 本发明实施例还提供了一种数据处理装置,如图7所示,该装置70包括:通信接口71、处理器72、存储器73;其中,

[0145] 通信接口71,能够与其它设备进行信息交互;

[0146] 处理器72,与所述通信接口71连接,用于运行计算机程序时,执行上述智能设备侧一个或多个技术方案提供的方法。而所述计算机程序存储在存储器73上。

[0147] 当然,实际应用时,数据处理装置70中的各个组件通过总线系统74耦合在一起。可理解,总线系统74用于实现这些组件之间的连接通信。总线系统74除包括数据总线之外,还包括电源总线、控制总线和状态信号总线。但是为了清楚说明起见,在图7中将各种总线都标为总线系统74。

[0148] 本申请实施例中的存储器73用于存储各种类型的数据以支持控制装置70的操作。这些数据的示例包括:用于在数据处理装置70上操作的任何计算机程序。

[0149] 上述本申请实施例揭示的方法可以应用于所述处理器72中,或者由所述处理器72实现。所述处理器72可能是一种集成电路芯片,具有信号的处理能力。在实现过程中,上述方法的各步骤可以通过所述处理器72中的硬件的集成逻辑电路或者软件形式的指令完成。上述的所述处理器72可以是通用处理器、数字数据处理器(DSP, Digital Signal Processor),或者其他可编程逻辑器件、分立门或者晶体管逻辑器件、分立硬件组件等。所述处理器72可以实现或者执行本申请实施例中的公开的各方法、步骤及逻辑框图。通用处理器可以是微处理器或者任何常规的处理器等。结合本申请实施例所公开的方法的步骤,可以直接体现为硬件译码处理器执行完成,或者用译码处理器中的硬件及软件模块组合执行完成。软件模块可以位于存储介质中,该存储介质位于存储器73,所述处理器72读取存储器73中的信息,结合其硬件完成前述方法的步骤。

[0150] 在示例性实施例中,数据处理装置70可以被一个或多个应用专用集成电路(ASIC, Application Specific Integrated Circuit)、DSP、可编程逻辑器件(PLD, Programmable Logic Device)、复杂可编程逻辑器件(CPLD, Complex Programmable Logic Device)、现场

可编程门阵列 (FPGA, Field-Programmable Gate Array)、通用处理器、控制器、微控制器 (MCU, Micro Controller Unit)、微处理器 (Microprocessor)、或者其他电子元件实现,用于执行前述方法。

[0151] 可以理解,本申请实施例的存储器73可以是易失性存储器或者非易失性存储器,也可包括易失性和非易失性存储器两者。其中,非易失性存储器可以是只读存储器 (ROM, Read Only Memory)、可编程只读存储器 (PROM, Programmable Read-Only Memory)、可擦除可编程只读存储器 (EPROM, Erasable Programmable Read-Only Memory)、电可擦除可编程只读存储器 (EEPROM, Electrically Erasable Programmable Read-Only Memory)、磁性随机存取存储器 (FRAM, ferromagnetic random access memory)、快闪存储器 (Flash Memory)、磁表面存储器、光盘、或只读光盘 (CD-ROM, Compact Disc Read-Only Memory);磁表面存储器可以是磁盘存储器或磁带存储器。易失性存储器可以是随机存取存储器 (RAM, Random Access Memory),其用作外部高速缓存。通过示例性但不是限制性说明,许多形式的RAM可用,例如静态随机存取存储器 (SRAM, Static Random Access Memory)、同步静态随机存取存储器 (SSRAM, Synchronous Static Random Access Memory)、动态随机存取存储器 (DRAM, Dynamic Random Access Memory)、同步动态随机存取存储器 (SDRAM, Synchronous Dynamic Random Access Memory)、双倍数据速率同步动态随机存取存储器 (DDRSDRAM, Double Data Rate Synchronous Dynamic Random Access Memory)、增强型同步动态随机存取存储器 (ESDRAM, Enhanced Synchronous Dynamic Random Access Memory)、同步连接动态随机存取存储器 (SLDRAM, SyncLink Dynamic Random Access Memory)、直接内存总线随机存取存储器 (DRRAM, Direct Rambus Random Access Memory)。本申请实施例描述的存储器旨在包括但不限于这些和任意其它适合类型的存储器。

[0152] 需要说明的是:“第一”、“第二”等是用于区别类似的对象,而不必用于描述特定的顺序或先后次序。

[0153] 在本申请所提供的几个实施例中,应该理解到,所揭露的设备和方法,可以通过其它的方式实现。以上所描述的设备实施例仅仅是示意性的,例如,所述单元的划分,仅仅为一种逻辑功能划分,实际实现时可以有另外的划分方式,如:多个单元或组件可以结合,或可以集成到另一个系统,或一些特征可以忽略,或不执行。另外,所显示或讨论的各组成部分相互之间的耦合、或直接耦合、或通信连接可以通过一些接口,设备或单元的间接耦合或通信连接,可以是电性的、机械的或其它形式的。上述作为分离部件说明的单元可以是、或也可以不是物理上分开的,作为单元显示的部件可以是、或也可以不是物理单元,即可以位于一个地方,也可以分布到多个网络单元上;可以根据实际的需要选择其中的部分或全部单元来实现本实施例方案的目的。另外,在本发明各实施例中的各功能单元可以全部集成在一个处理模块中,也可以是各单元分别单独作为一个单元,也可以两个或两个以上单元集成在一个单元中;上述集成的单元既可以采用硬件的形式实现,也可以采用硬件加软件功能单元的形式实现。本领域普通技术人员可以理解:实现上述方法实施例的全部或部分步骤可以通过程序指令相关的硬件来完成,前述的程序可以存储于一计算机可读取存储介质中,该程序在执行时,执行包括上述方法实施例的步骤;而前述的存储介质包括:移动存储设备、ROM、RAM、磁碟或者光盘等各种可以存储程序代码的介质。

[0154] 本申请所提供的几个方法实施例中所揭露的方法,在不冲突的情况下可以任意组

合,得到新的方法实施例。

[0155] 本申请所提供的几个产品实施例中所揭露的特征,在不冲突的情况下可以任意组合,得到新的产品实施例。

[0156] 本申请所提供的几个方法或设备实施例中所揭露的特征,在不冲突的情况下可以任意组合,得到新的方法实施例或设备实施例。

[0157] 以上所述,仅为本发明的具体实施方式,但本发明的保护范围并不局限于此,任何熟悉本技术领域的技术人员在本发明揭露的技术范围内,可轻易想到变化或替换,都应涵盖在本发明的保护范围之内。因此,本发明的保护范围应以所述权利要求的保护范围为准。

[0158] 需要说明的是:本发明实施例所记载的技术方案之间,在不冲突的情况下,可以任意组合。

[0159] 以上所述,仅为本发明的较佳实施例而已,并非用于限定本发明的保护范围。

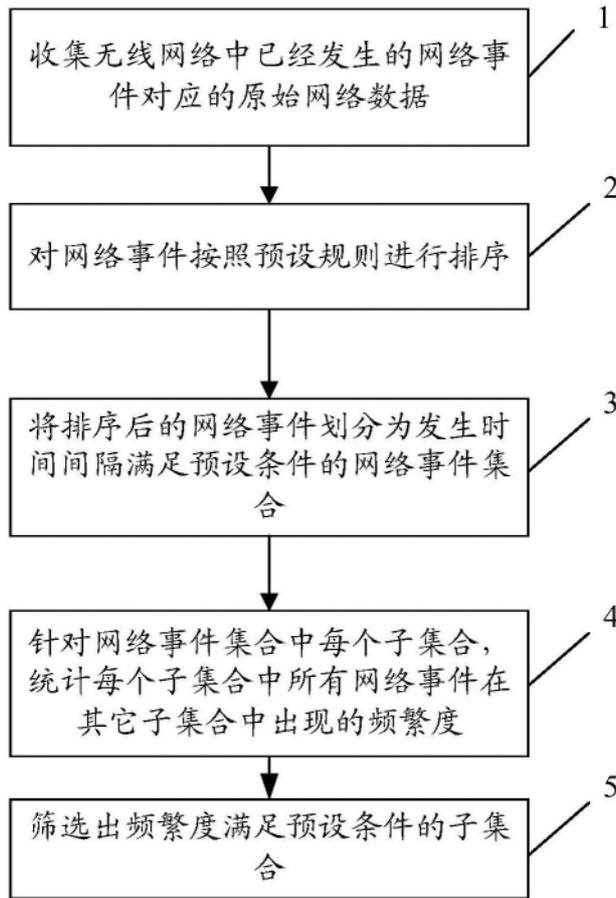


图1

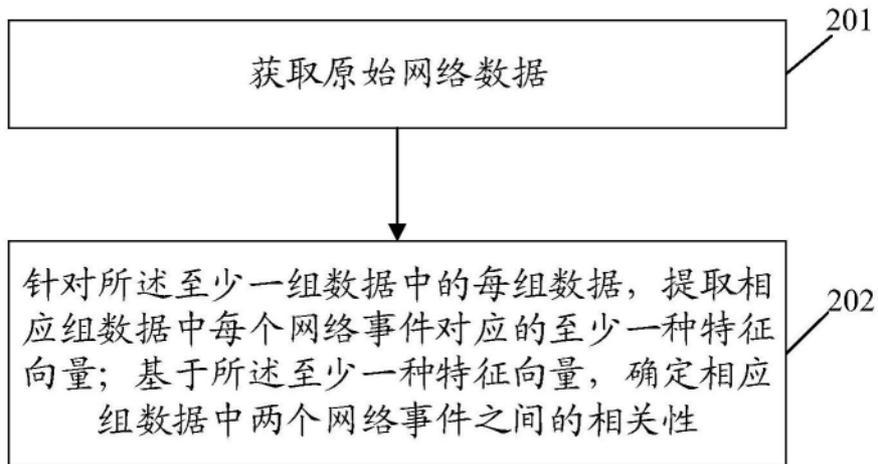


图2

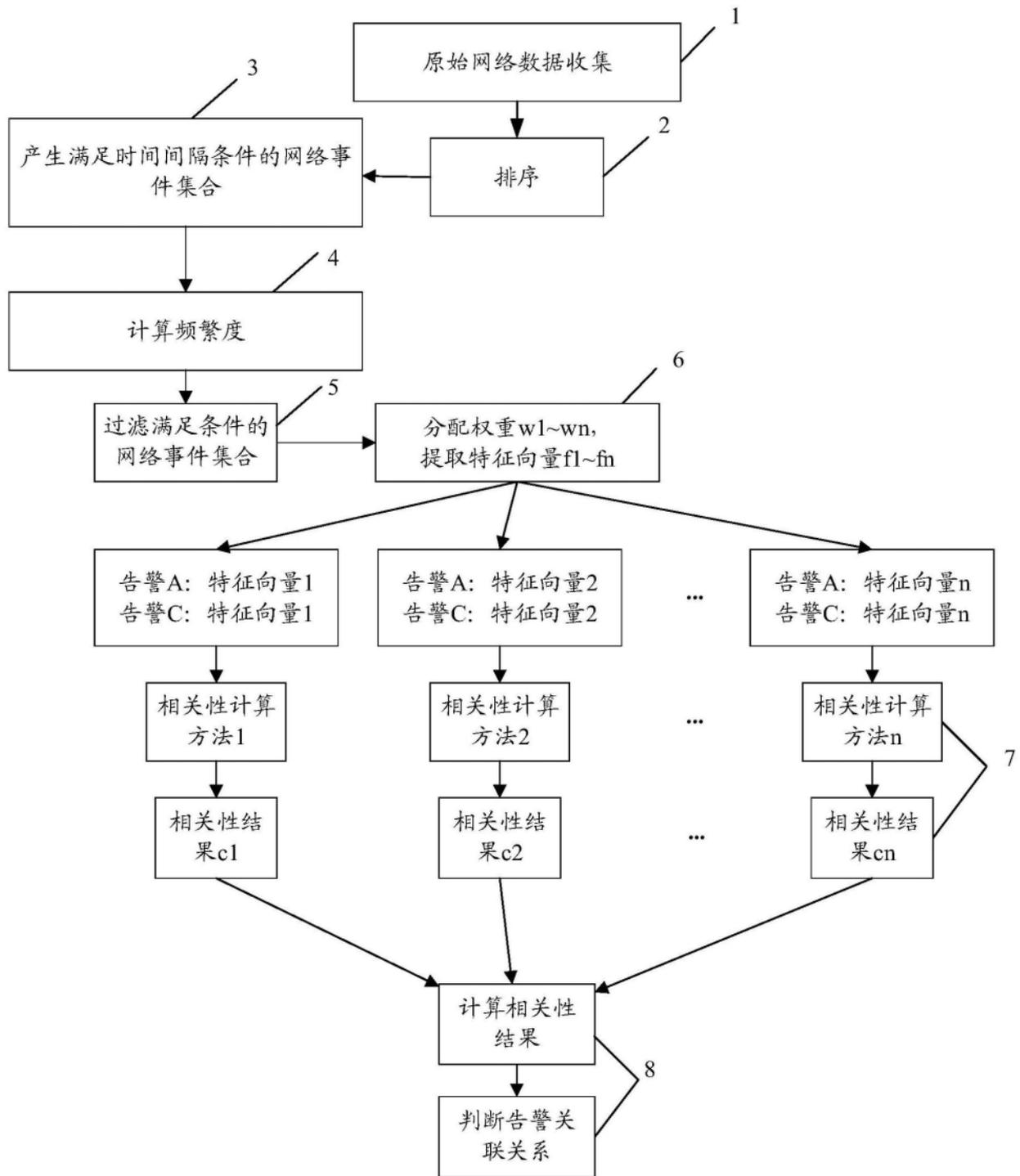


图3

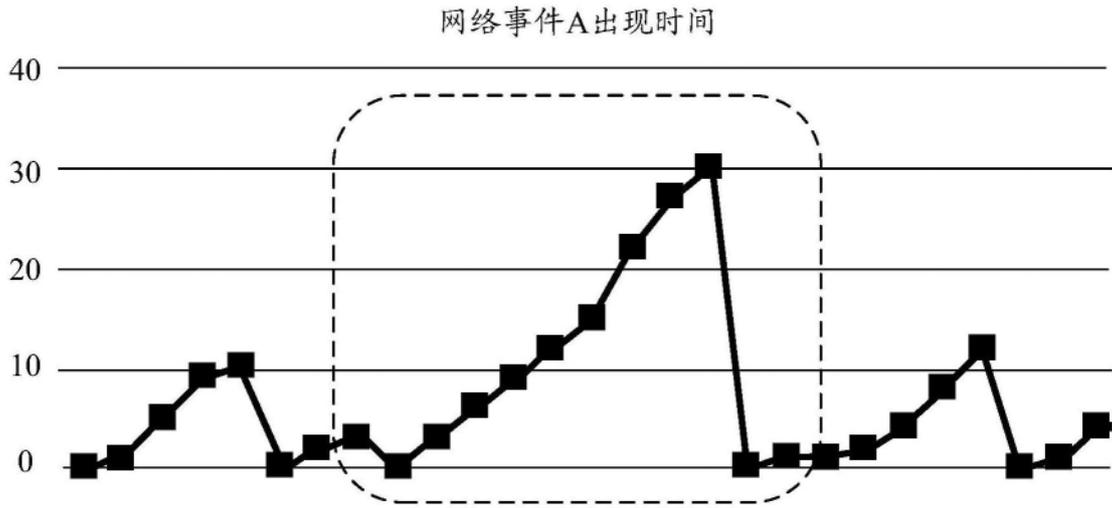


图4a

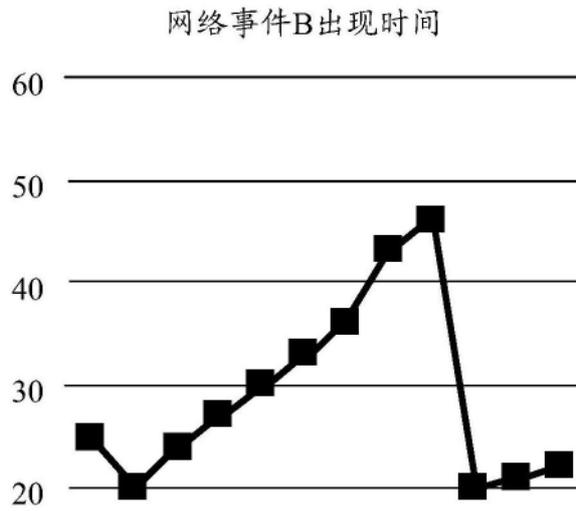


图4b

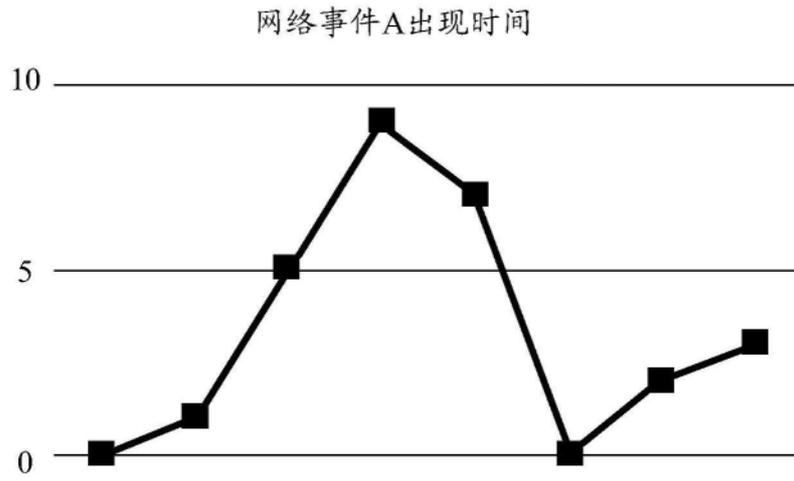


图5a

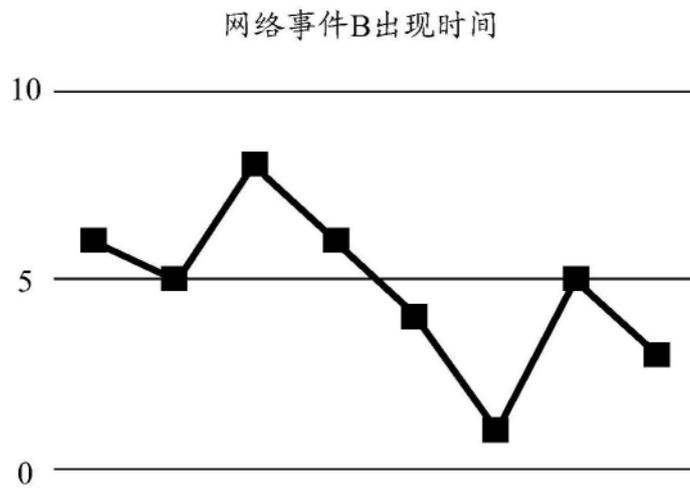


图5b

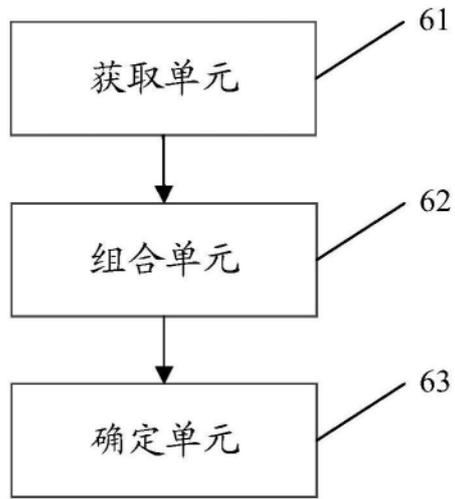


图6

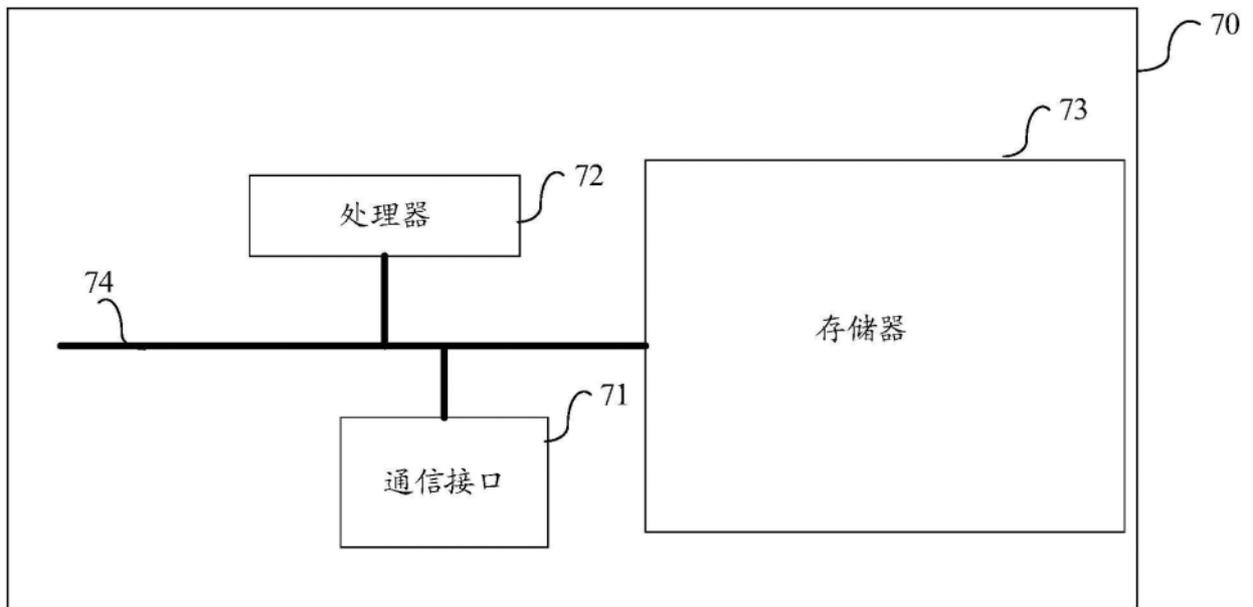


图7