



(12) 发明专利申请

(10) 申请公布号 CN 107025153 A

(43) 申请公布日 2017. 08. 08

(21) 申请号 201610065609. 6

(22) 申请日 2016. 01. 29

(71) 申请人 阿里巴巴集团控股有限公司

地址 英属开曼群岛大开曼资本大厦一座四  
层 847 号邮箱

(72) 发明人 丁永明 周俊 崔卿 瞿神全

(74) 专利代理机构 北京博浩百睿知识产权代理  
有限责任公司 11134

代理人 宋子良

(51) Int. Cl.

G06F 11/22(2006. 01)

G06F 11/30(2006. 01)

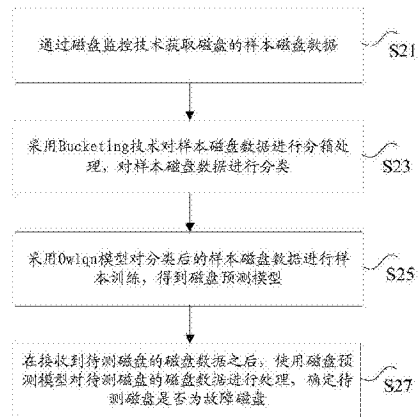
权利要求书2页 说明书12页 附图4页

(54) 发明名称

磁盘的故障预测方法和装置

(57) 摘要

本发明公开了一种磁盘的故障预测方法和装置。其中,该方法包括:通过磁盘监控技术获取磁盘的样本磁盘数据,其中,样本磁盘数据包括多个维度上的样本数据;采用 Bucketing 技术对样本磁盘数据进行分箱处理,对样本磁盘数据进行分类;采用 Owlqn 模型对分类后的样本磁盘数据进行样本训练,得到磁盘预测模型;在接收到待测磁盘的磁盘数据之后,使用磁盘预测模型对待测磁盘的磁盘数据进行处理,确定待测磁盘是否为故障磁盘。本发明解决了现有技术的硬盘故障预测系统中一些容易致使硬盘故障的因素不能被采集或量化导致的预测结果不准确的技术问题。



1. 一种磁盘的故障预测方法,其特征在于,包括:

通过磁盘监控技术获取磁盘的样本磁盘数据,其中,所述样本磁盘数据包括多个维度上的样本数据;

采用Bucketing技术对所述样本磁盘数据进行分箱处理,对所述样本磁盘数据进行分类;

采用Ow1qn模型对所述分类后的样本磁盘数据进行样本训练,得到磁盘预测模型;

在接收到待测磁盘的磁盘数据之后,使用所述磁盘预测模型对所述待测磁盘的磁盘数据进行处理,确定所述待测磁盘是否为故障磁盘。

2. 根据权利要求1所述的方法,其特征在于,所述样本磁盘数据为SMART磁盘数据,其中,所述样本磁盘数据至少包括如下四个维度上的样本数据:原始值、标准值、最差值和累积值。

3. 根据权利要求2所述的方法,其特征在于,在通过磁盘监控技术获取磁盘的样本磁盘数据之后,所述方法还包括:

对每个维度上的样本数据进行如下任意一种或多种运算:差分运算、平方运算和分布求和运算,使得任意一个维度上的样本数据被扩展出新的维度上的样本数据。

4. 根据权利要求1至3中任意一项所述的方法,其特征在于,采用Bucketing技术对所述样本磁盘数据进行分箱处理,对所述样本磁盘数据进行分类,包括:

确定预先划分的每个分箱的取值范围以及每个分箱对应的ID值;

通过将每个维度上的样本数据离散化至对应的分箱来对所述样本磁盘数据进行分类,得到所述每个维度上的样本数据所对应的ID值。

5. 根据权利要求4所述的方法,其特征在于,采用Ow1qn模型对所述分类后的样本磁盘数据进行样本训练,得到磁盘预测模型,包括:

所述Ow1qn模型对所述每个维度上的样本数据所对应的ID值进行训练,得到每个维度上的样本数据的权重值;

根据所述每个维度上的样本数据及对应的权重值,确定所述磁盘预测模型,其中,所述磁盘预测模型包括所述每个维度上的样本数据的预测结果。

6. 根据权利要求5所述的方法,其特征在于,所述每个维度上的样本数据的预测结果为所述样本磁盘数据进行分类后得到的预测值。

7. 根据权利要求6所述的方法,其特征在于,在接收到待测磁盘的磁盘数据之后,使用所述磁盘预测模型对所述待测磁盘的磁盘数据进行处理,确定所述待测磁盘是否为故障磁盘,包括:

接收到所述待测磁盘的磁盘数据之后,将所述待测磁盘的磁盘数据离散化至对应的分箱,得到所述待测磁盘的磁盘数据所对应的ID值;

根据所述待测磁盘的磁盘数据所对应的ID值确定所述待测磁盘的磁盘数据的权重值;

根据所述待测磁盘的磁盘数据的权重值从所述磁盘预测模型中确定所述待测磁盘是否为故障磁盘。

8. 一种磁盘的故障预测装置,其特征在于,包括:

获取模块,用于通过磁盘监控技术获取磁盘的样本磁盘数据,其中,所述样本磁盘数据包括多个维度上的样本数据;

分类模块,用于采用Bucketing技术对所述样本磁盘数据进行分箱处理,对所述样本磁盘数据进行分类;

训练模块,用于采用Owlqn模型对所述分类后的样本磁盘数据进行样本训练,得到磁盘预测模型;

确定模块,用于在接收到待测磁盘的磁盘数据之后,使用所述磁盘预测模型对所述待测磁盘的磁盘数据进行处理,确定所述待测磁盘是否为故障磁盘。

9.根据权利要求8所述的装置,其特征在于,所述样本磁盘数据为SMART磁盘数据,其中,所述样本磁盘数据至少包括如下四个维度上的样本数据:原始值、标准值、最差值和累积值。

10.根据权利要求9所述的装置,其特征在于,所述装置还包括:

运算模块,用于对每个维度上的样本数据进行如下任意一种或多种运算:差分运算、平方运算和分布求和运算,使得任意一个维度上的样本数据被扩展出新的维度上的样本数据。

11.根据权利要求8至10中任意一项所述的装置,其特征在于,所述分类模块包括:

第一确定子模块,用于确定预先划分的每个分箱的取值范围以及每个分箱对应的ID值;

分类子模块,用于通过将每个维度上的样本数据离散化至对应的分箱来对所述样本磁盘数据进行分类,得到所述每个维度上的样本数据所对应的ID值。

12.根据权利要求11所述的装置,其特征在于,所述训练模块包括:

训练子模块,用于所述Owlqn模型对所述每个维度上的样本数据所对应的ID值进行训练,得到每个维度上的样本数据的权重值;

第二确定子模块,用于根据所述每个维度上的样本数据及对应的权重值,确定所述磁盘预测模型,其中,所述磁盘预测模型包括所述每个维度上的样本数据的预测结果。

13.根据权利要求12所述的装置,其特征在于,所述每个维度上的样本数据的预测结果为所述样本磁盘数据进行分类后得到的预测值。

14.根据权利要求13所述的装置,其特征在于,所述确定模块还包括:

离散模块,用于接收到所述待测磁盘的磁盘数据之后,将所述待测磁盘的磁盘数据离散化至对应的分箱,得到所述待测磁盘的磁盘数据所对应的ID值;

第三确定子模块,用于根据所述待测磁盘的磁盘数据所对应的ID值确定所述待测磁盘的磁盘数据的权重值;

第四确定子模块,用于根据所述待测磁盘的磁盘数据的权重值从所述磁盘预测模型中确定所述待测磁盘是否为故障磁盘。

## 磁盘的故障预测方法和装置

### 技术领域

[0001] 本发明涉及磁盘领域,具体而言,涉及一种磁盘的故障预测方法和装置。

### 背景技术

[0002] 目前,硬盘是存储数据的主要介质,硬盘一旦出故障,便会造成巨大的数据损失。因此如何保证硬盘的稳定性能非常重要。在通常状态下,硬盘在24小时中出错的概率在是万分之一左右,当一台服务器具有十块硬盘时,服务器硬盘出错的概率就会上升到千分之一,而随着当前网站等业务的发展,服务器需要使用的硬盘会越来越多,多块硬盘同时出错的概率也会提升。

[0003] 通常情况下,数据存储通常会有多个备份,如mysql主备库,GFS文件默认3个备份。在大量数据存储平台上,如果多个硬盘同时出故障,那么这些硬盘上存储着同一个文件的备份的概率就会很高,即如果多块硬盘同时出现故障,就会导致一些文件的丢失,对于一些线上的服务,大都依赖于服务器中存储的海量数据,如果硬盘出故障,就会导致上述在线服务异常,甚至暂停使用。

[0004] 由于上述原因,需要具有预测硬盘是否会出错的系统需要有一套系统能提前告诉我们哪些硬盘会出错,数据可能丢失导致硬盘故障的原因有很多,最常见的有以下几种:外部振动、温度和湿度、电器元件损坏、声音和灰尘,在上述因素中,有些因素能够被采集到,比如温度和湿度、一些元器件数据,但是更多的数据无法被采集和量化,因此便会导致预测结果不准确。

[0005] 针对现有技术的硬盘故障预测系统中一些容易致使硬盘故障的因素不能被采集或量化导致的预测结果不准确的问题,目前尚未提出有效的解决方案。

### 发明内容

[0006] 本发明实施例提供了一种磁盘的故障预测方法和装置,以至少解决现有技术的硬盘故障预测系统中一些容易致使硬盘故障的因素不能被采集或量化导致的预测结果不准确的技术问题。

[0007] 根据本发明实施例的一个方面,提供了一种磁盘的故障预测方法,包括:通过磁盘监控技术获取磁盘的样本磁盘数据,其中,样本磁盘数据包括多个维度上的样本数据;采用Bucketing技术对样本磁盘数据进行分箱处理,对样本磁盘数据进行分类;采用Owlqn模型对分类后的样本磁盘数据进行样本训练,得到磁盘预测模型;在接收到待测磁盘的磁盘数据之后,使用磁盘预测模型对待测磁盘的磁盘数据进行处理,确定待测磁盘是否为故障磁盘。

[0008] 根据本发明实施例的另一方面,还提供了一种磁盘的故障预测装置,包括:获取模块,用于通过磁盘监控技术获取磁盘的样本磁盘数据,其中,样本磁盘数据包括多个维度上的样本数据;分类模块,用于采用Bucketing技术对样本磁盘数据进行分箱处理,对样本磁盘数据进行分类;训练模块,用于采用Owlqn模型对分类后的样本磁盘数据进行样本训练,

得到磁盘预测模型；确定模块，用于在接收到待测磁盘的磁盘数据之后，使用磁盘预测模型对待测磁盘的磁盘数据进行处理，确定待测磁盘是否为故障磁盘。

[0009] 在本发明实施例中，采用通过磁盘监控技术获取磁盘的样本磁盘数据，其中，样本磁盘数据包括多个维度上的样本数据；采用Bucketing技术对样本磁盘数据进行分箱处理，对样本磁盘数据进行分类；采用Owlqn模型对分类后的样本磁盘数据进行样本训练，得到磁盘预测模型的方式，通过在接收到待测磁盘的磁盘数据之后，使用磁盘预测模型对待测磁盘的磁盘数据进行处理，达到了确定待测磁盘是否为故障磁盘的目的，从而实现了预测磁盘故障的技术效果，进而解决了现有技术的硬盘故障预测系统中一些容易致使硬盘故障的因素不能被采集或量化导致的预测结果不准确的技术问题。

## 附图说明

[0010] 此处所说明的附图用来提供对本发明的进一步理解，构成本申请的一部分，本发明的示意性实施例及其说明用于解释本发明，并不构成对本发明的不当限定。在附图中：

[0011] 图1是根据本发明实施例的一种磁盘的故障预测方法的计算机终端的硬件结构框图；

[0012] 图2是根据本发明实施例一的磁盘的故障预测方法的流程图；

[0013] 图3是根据本发明实施例的一种可选的磁盘的故障预测方法的流程图；

[0014] 图4是根据本发明实施例的一种磁盘的故障预测装置的结构示意图；

[0015] 图5是根据本发明实施例的一种可选的磁盘的故障预测装置的结构示意图；

[0016] 图6是根据本发明实施例的一种可选的磁盘的故障预测装置的结构示意图；

[0017] 图7是根据本发明实施例的一种可选的磁盘的故障预测装置的结构示意图；

[0018] 图8是根据本发明实施例的一种可选的磁盘的故障预测装置的结构示意图；以及

[0019] 图9是根据本发明实施例的一种计算机终端的结构框图。

## 具体实施方式

[0020] 为了使本技术领域的人员更好地理解本发明方案，下面将结合本发明实施例中的附图，对本发明实施例中的技术方案进行清楚、完整地描述，显然，所描述的实施例仅仅是本发明一部分的实施例，而不是全部的实施例。基于本发明中的实施例，本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其他实施例，都应当属于本发明保护的范围。

[0021] 需要说明的是，本发明的说明书和权利要求书及上述附图中的术语“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。应该理解这样使用的数据在适当情况下可以互换，以便这里描述的本发明的实施例能够以除了在这里图示或描述的那些以外的顺序实施。此外，术语“包括”和“具有”以及他们的任何变形，意图在于覆盖不排他的包含，例如，包含了一系列步骤或单元的过程、方法、系统、产品或设备不必限于清楚地列出的那些步骤或单元，而是可包括没有清楚地列出的或对于这些过程、方法、产品或设备固有的其它步骤或单元。

[0022] 实施例1

[0023] 根据本发明实施例，提供了一种磁盘的故障预测方法实施例，需要说明的是，在附

图的流程图示出的步骤可以在诸如一组计算机可执行指令的计算机系统中执行,并且,虽然在流程图中示出了逻辑顺序,但是在某些情况下,可以以不同于此处的顺序执行所示出或描述的步骤。

[0024] 本申请实施例一所提供的方法实施例可以在移动终端、计算机终端或者类似的运算装置中执行。以运行在计算机终端上为例,图1是根据本发明实施例的一种磁盘的故障预测方法的计算机终端的硬件结构框图。如图1所示,计算机终端10可以包括一个或多个(图中仅示出一个)处理器102(处理器102可以包括但不限于微处理器MCU或可编程逻辑器件FPGA等的处理装置)、用于存储数据的存储器104、以及用于通信功能的传输模块106。本领域普通技术人员可以理解,图1所示的结构仅为示意,其并不对上述电子装置的结构造成限定。例如,计算机终端10还可包括比图1中所示更多或者更少的组件,或者具有与图1所示不同的配置。

[0025] 存储器104可用于存储应用软件的软件程序以及模块,如本发明实施例中的磁盘的故障预测方法对应的程序指令/模块,处理器102通过运行存储在存储器104内的软件程序以及模块,从而执行各种功能应用以及数据处理,即实现上述的磁盘的故障预测方法。存储器104可包括高速随机存储器,还可包括非易失性存储器,如一个或者多个磁性存储装置、闪存、或者其他非易失性固态存储器。在一些实例中,存储器104可进一步包括相对于处理器102远程设置的存储器,这些远程存储器可以通过网络连接至计算机终端10。上述网络的实例包括但不限于互联网、企业内部网、局域网、移动通信网及其组合。

[0026] 传输装置106用于经由一个网络接收或者发送数据。上述的网络具体实例可包括计算机终端10的通信供应商提供的无线网络。在一个实例中,传输装置106包括一个网络适配器(Network Interface Controller, NIC),其可通过基站与其他网络设备相连从而可与互联网进行通讯。在一个实例中,传输装置106可以为射频(Radio Frequency, RF)模块,其用于通过无线方式与互联网进行通讯。

[0027] 在上述运行环境下,本申请提供了如图2所示的磁盘的故障预测方法。图2是根据本发明实施例一的磁盘的故障预测方法的流程图。

[0028] 步骤S21,通过磁盘监控技术获取磁盘的样本磁盘数据,其中,样本磁盘数据包括多个维度上的样本数据。

[0029] 在上述步骤中,磁盘监控技术用于监视并记录磁盘状态,

[0030] 在一种可选的实施例中,样本磁盘数据可以是样本磁盘的数据吞吐性能、马达启动时间、寻道错误率等。

[0031] 此处需要说明的是,在使用现有技术(例如S.M.A.R.T,自我监测、分析及报告技术)对磁盘进行监测时,能够得到多维度的体现磁盘状态的数据,以及根据监测得到的数据对磁盘是否故障,或是否会在未来的较短时间内发生故障做出分析,这样的分析在磁盘监测技术监测的数据的基础上进行的,然而磁盘的状态还能通过其他数据量体现,这些数据量可能是不能被检测或不能被量化的数据量,因此本申请建立了磁盘预测模型,使用磁盘预测模型对磁盘的故障状态进行分析,其中,磁盘预测模型由Owlqn模型对样本磁盘数据进行样本训练得到。上述实施例的方案通过对磁盘样本数据的样本训练建立了预测故障的模型,使得向磁盘监控系统输入待测磁盘样本数据后,能够根据预测故障的模型,对待测磁盘进行故障状态分析,避免了在分析磁盘故障时,采用对单一或固定的多个样本数据进行分

析,导致的不可统计或不可量化的磁盘数据对磁盘故障预测结果的影响。

[0032] 步骤S23,采用Bucketing技术对样本磁盘数据进行分箱处理,对样本磁盘数据进行分类。

[0033] 在上述步骤中,在对样本磁盘数据进行分箱处理时能够采用多种分箱方法达到平滑数据的目的,其中,对样本磁盘数据进行分箱的方法包括按照箱内数据的平均值平滑数据、按照箱内数据的中间值平滑数据以及按照箱内数据的边界值平滑数据。

[0034] 在一种可选的实施例中,可以先将样本磁盘数据集中的多个样本数据分至多个分箱中,在此示例中,将样本磁盘数据分至5个分箱中,将样本磁盘数据分至不同的分箱中时可将样本磁盘数据按照升序排列,然后计算每个分箱中的数据量,在将样本磁盘数据按照每个分箱中应该有的数据量分至5个分箱,然后对每个分箱中的数据进行处理,在此实施例中采用按照箱内数据的平均值平滑数据的方法进行处理,即计算得到每个分箱中数据的平均值,然后该分箱内所有数据均变为该平均值。

[0035] 此处需要说明当时,对样本磁盘数据进行分箱处理用于将每个分箱中的数据进行平滑处理,由于每个分箱中的数据都较为相近,因此分箱处理在达到稳定平滑数据的基础上,并不会影响下一步骤中对样本磁盘数据进行训练的结果。

[0036] 此处还需要说明的是,对样本磁盘数据进行分箱处理的方法包括上述实施例中的任意一种分箱方法,且不限于此,任何能够达到平滑或稳定数据目的的方法都可用于对样本磁盘数据的分箱处理。

[0037] 步骤S25,采用Ow1qn模型对分类后的样本磁盘数据进行样本训练,得到磁盘预测模型。

[0038] 在上述步骤中,对样本磁盘数据进行训练可以为将处理后的样本磁盘数据输入至Ow1qn模型,其中,上述样本磁盘数据为预先知晓真实值的样本,样本的真实值可以是1或者0,用于表示样本为正样本或负样本,正样本用于表示该样本为故障磁盘,负样本表示该样本为正常磁盘。

[0039] 在一种可选的实施例中,每个输入的样本磁盘数据都能够从Ow1qn模型中得到相应的输出值,在得到样本磁盘数据集中每个样本相应的输出值之后,获取所有正样本的输出值,构成正样本输出值区间,同样获取所有负样本的输出值,构成负样本的输出区间,由此得到磁盘预测模型。

[0040] 步骤S27,在接收到待测磁盘的磁盘数据之后,使用磁盘预测模型对待测磁盘的磁盘数据进行处理,确定待测磁盘是否为故障磁盘。

[0041] 需要进一步说明的是,在使用Ow1qn模型对样本磁盘数据进行样本训练时,使用的是对样本磁盘数据进行分类后的样本数据,并且对分类后的样本数据进行了分箱处理,使得分类后的每个类别中的样本数据离散化,从而能够对样本磁盘的样本数据进行训练。

[0042] 在一种可选的实施例中,上述样本磁盘数据可以包括:底层数据读取错误率、启动/停止计数、重映射扇区数、通电时间累计、主轴起旋重试次数、磁盘校准重试次数、磁盘通电次数、温度以及写错误率,可以根据磁盘历史故障情况获取样本磁盘数据。例如,可以按照正负样本比例为1:5的比例进行样本获取,其中,正样本为存在故障的磁盘,负样本为不存在故障的磁盘。

[0043] 此处需要说明的是,在通过磁盘监控技术获取磁盘的样本磁盘数据时,由于预测

磁盘故障的各个机构使用的磁盘并不一定相同,且由于各个机构不同温湿度等环境因素对磁盘的影响,使得不同机构的磁盘的好坏比例并不相同,为了使样本磁盘数据的训练提供更可靠的样本磁盘数据,还可以根据机构的实际上磁盘损坏情况进行获取样本磁盘数据。

[0044] 由此,解决了现有技术的硬盘故障预测系统中一些容易致使硬盘故障的因素不能被采集或量化导致的预测结果不准确的技术问题

[0045] 根据本申请上述实施例,在一种优选的方案中,样本磁盘数据为SMART磁盘数据,其中,样本磁盘数据至少包括如下四个维度上的样本数据:原始值、标准值、最差值和累积值。

[0046] 上述原始值为磁盘运行时的当前参数;上述标准值为正常磁盘运行时各项参数的数值;上述最差值为磁盘运行时,磁盘的各项检测参数曾出现过与正常值偏差最大的非正常值;上述累计值为磁盘的各项检测参数从磁盘使用至当前时刻的累计结果。

[0047] 在一种可选的实施例中,磁盘的各项参数可以是对磁盘的各项属性进行描述的信息,可以包括错误读取率、加电次数、重新分配扇区数、旋转重试次数、磁盘校准重试次数以及奇偶校验错误率中的一项或多项,也可以包括磁盘的其他属性信息。

[0048] 在一种可选的实施例中,可以采用HDTune、CrystalDiskInfo等软件获取样本磁盘数据。

[0049] 根据本申请上述实施例,在一种优选的方案中,步骤S21在通过磁盘监控技术获取磁盘的样本磁盘数据之后,上述方法还包括:

[0050] 步骤S211,对每个维度上的样本数据进行如下任意一种或多种运算:差分运算、平方运算和分布求和运算,使得任意一个维度上的样本数据被扩展出新的维度上的样本数据。

[0051] 在一种可选的实施例中,对样本数据中的原始值进行差分运算、平方运算和分布求和运算,从而能得到原始值的差分值、原始值的平方值以及原始值的分布求和值,因此在知晓样本磁盘数据的原始值的基础上,得到另外四个维度的样本磁盘数据;同样可以将样本数据中的标准值、最差值以及累积值分别作上述运算,得到更多维度的样本磁盘数据。

[0052] 需要说明的是,对样本磁盘数据进行多种运算得到更多维度的样本磁盘数据,能够提高对样本磁盘数据的利用率以及对样本磁盘数据进行训练时,样本磁盘数据的敏感度,从而提高故障预测模型的准确度。

[0053] 根据本申请上述实施例,在一种优选的方案中,步骤S23采用Bucketing技术对样本磁盘数据进行分箱处理,对样本磁盘数据进行分类,包括:

[0054] 步骤S231,确定预先划分的每个分箱的取值范围以及每个分箱对应的ID值。

[0055] 在上述步骤中,划分的每个分箱的取值范围的目的在于确定与样本磁盘数据集合中的数据对应的分箱,即样本磁盘数据所属的范围对应的分箱即为该样本磁盘数据所属的分箱。确定每个分箱的ID值用于区分不同的分箱。

[0056] 步骤S233,通过将每个维度上的样本数据离散化至对应的分箱来对样本磁盘数据进行分类,得到每个维度上的样本数据所对应的ID值。

[0057] 在一种可选的实施例中,将样本磁盘数据分配至不同的分箱后,以分箱的ID号对分配至该分箱的数据进行替换,即将每一维度上的样本磁盘数据都被替换为该样本磁盘数据对应的分箱ID值,使得原始的样本磁盘数据的每一个维度上的数据都被替换为如上的整



数值。

[0058] 在另一种可选的实施例中,例如在设置5个取值范围不同的分箱,且每个分箱的ID值分别为1,2,3,4,5的情况下,每个分箱中都包含不同的数据,当样本磁盘数据A落入分箱1和分箱3的取值范围内时,样本磁盘数据A的ID值可以为10100。按照上述实施例中的方案,使得每个维度上的样本数据都能够得到与之对应的ID值。

[0059] 根据本申请上述实施例,在一种优选的方案中,步骤S25,采用0w1qn模型对分类后的样本磁盘数据进行样本训练,得到磁盘预测模型,包括:

[0060] 步骤S251,0w1qn模型对每个维度上的样本数据所对应的ID值进行训练,得到每个维度上的样本数据的权重值。

[0061] 在上述步骤中,每个维度上的样本数据的权重值为该样本为“1”的概率,即为该样本为正样本的概率。

[0062] 在一种可选的是实例中,将待测磁盘数据表示为 $(\bar{x}_i, \bar{y}_i)$ ,其中, $\bar{x}_i = x_i^1, \dots, x_i^n$ , $y_i$ 为0或1,0w1qn模型获取用于训练的样本数据后,输出每个磁盘特征数据的权重值,即每个磁盘特征数据为故障磁盘数据的概率。权重值可以通过如下公式计算得到:权重值

$\tilde{y}_i = f(\bar{x}_i) = w_0 + \sum_{k=1}^n w_k \cdot x_i^k$ , $i$ 用于表示第*i*个样本, $n$ 用于表示*n*个维度, $k$ 表示1至*n*之间任意

一个维度, $w_k$ 用于表示*k*维度上的权重值,其中, $w_0$ 为截距,需要注意的是,输出的权重值需要

满足条件: $J = \|y_i - l(\tilde{y}_i)\| + \sum_{i=0}^n (w_i)$ 能取得最小值, $J$ 为最优化目标函数。

[0063] 步骤S253,根据每个维度上的样本数据及对应的权重值,确定磁盘预测模型,其中,磁盘预测模型包括每个维度上的样本数据的预测结果。

[0064] 在一种可选的实施例中,得到待测磁盘的磁盘数据后,计算待测磁盘的预测值,其中,

计算待测磁盘的预测值可以根据如下公式进行计算: $l(\tilde{y}_i) = \frac{1}{1 + e^{-\tilde{y}_i}}$ 。上述预测值即为训

练样本磁盘数据得到的预测结果,由于样本磁盘是否为故障磁盘为已知量,因此,在得到预测结果后,将正样本磁盘的预测结果和负样本磁盘的预测结果进行区分,得到故障磁盘的预测值的取值范围和正常磁盘的预测值的取值范围。

[0065] 在一种可选的实施例中,向0w1qn模型输入样本数据对应的ID值,并向0w1qn模型输入ID值对应的样本磁盘的故障状态,使0w1qn模型记忆ID值以及ID值对应的磁盘故障状态,再向0w1qn模型重复输入ID值,验证0w1qn模型是否能够输出ID值对应的故障状态。

[0066] 根据本申请上述实施例,在一种优选的方案中,每个维度上的样本数据的预测结果为样本磁盘数据进行分类后得到的预测值。

[0067] 根据本申请上述实施例,在一种优选的方案中,步骤S27在接收到待测磁盘的磁盘数据之后,使用磁盘预测模型对待测磁盘的磁盘数据进行处理,确定待测磁盘是否为故障磁盘,包括:

[0068] 步骤S271,接收到待测磁盘的磁盘数据之后,将待测磁盘的磁盘数据离散化至对应的分箱,得到待测磁盘的磁盘数据所对应的ID值。

[0069] 在上述步骤中,将待测磁盘的磁盘数据离散化至对应的分箱,得到待测磁盘的磁

盘数据所对应的ID值,可以采用上实施例中的步骤S231至步骤S233中提出的方案实施。

[0070] 步骤S273,根据待测磁盘的磁盘数据所对应的ID值确定待测磁盘的磁盘数据的权重值。

[0071] 在一种可选的是实例中,将待测磁盘数据表示为 $(\bar{x}_i, \bar{y}_i)$ ,其中, $\bar{x}_i = x_i^1, \dots, x_i^n$ ,  $y_i$ 为0或1,owlqn模型获取用于训练的样本数据后,输出每个磁盘特征数据的权重值,即每个磁盘特征数据为故障磁盘数据的概率。权重值可以通过如下公式计算得到:权重值

$$\bar{y}_i = f(\bar{x}_i) = w_0 + \sum_{k=1}^n w_k \cdot x_i^k。需要注意的是,输出的权重值需要满足  $J = \|y_i - I(\bar{y}_i)\| + \sum_{i=0}^n (w_i)$ 。$$

[0072] 步骤S275,根据待测磁盘的磁盘数据的权重值从磁盘预测模型中确定待测磁盘是否为故障磁盘。

[0073] 在一种可选的实施例中,得到待测磁盘的磁盘数据后,计算待测磁盘的预测值,其中,

计算待测磁盘的预测值可以根据如下公式进行计算: $I(\bar{y}_i) = \frac{1}{1 + e^{-\bar{y}_i}}$ 。得到待测磁盘的预测值后,将待测磁盘的预测值与训练样本磁盘数据得到的正样本的取值范围和负样本的取值范围进行比对,若待测磁盘的预测值落入正样本的取值范围,则可以认为该待测磁盘为故障磁盘,若待测磁盘的预测值落入负样本的取值范围,则可以认为该待测磁盘为正常磁盘。

[0074] 需要说明的是,对于前述的各方法实施例,为了简单描述,故将其都表述为一系列的动作组合,但是本领域技术人员应该知悉,本发明并不受所描述的动作顺序的限制,因为依据本发明,某些步骤可以采用其他顺序或者同时进行。其次,本领域技术人员也应该知悉,说明书中所描述的实施例均属于优选实施例,所涉及的动作和模块并不一定是本发明所必须的。

[0075] 通过以上的实施方式的描述,本领域的技术人员可以清楚地了解到根据上述实施例的方法可借助软件加必需的通用硬件平台的方式来实现,当然也可以通过硬件,但很多情况下前者是更佳的实施方式。基于这样的理解,本发明的技术方案本质上或者说对现有技术做出贡献的部分可以以软件产品的形式体现出来,该计算机软件产品存储在一个存储介质(如ROM/RAM、磁碟、光盘)中,包括若干指令用以使得一台终端设备(可以是手机,计算机,服务器,或者网络设备等)执行本发明各个实施例所述的方法。

[0076] 如图3所示,提供了一种磁盘的故障预测方法,该方法可以包括如下步骤S31至步骤S37:

[0077] S31,获取样本磁盘的样本数据。

[0078] 在上述步骤中,样本磁盘的样本数据可以是SMART磁盘数据。具体的,在上述步骤中,可以通过HDTune、CrystalDiskInfo等软件获取样本磁盘数据。

[0079] S32,对样本数据进行差分运算。

[0080] 具体的,在上述步骤中,差分运算指磁盘在某一时刻的特征数据与过该磁盘在24小时之前的特征数据做差运算得到的值。

[0081] S33,对差分运算得到的结果进行分布求和和/或平方运算。

[0082] 上述步骤对每个维度上的样本数据进行如下任意一种或多种运算:差分运算、平方运算和分布求和运算,使得任意一个维度上的样本数据被扩展出新的维度上的样本数

据。

[0083] S34,得到训练和预测使用的数据。

[0084] S35,采用分箱进行离散化。

[0085] 上述步骤划分的每个分箱的取值范围的目的在于确定与样本磁盘数据集合中的数据对应的分箱,即样本磁盘数据所属的范围对应的分箱即为该样本磁盘数据所属的分箱。确定每个分箱的ID值用于区分不同的分箱,并对每个分箱中的数据进行离散化处理。

[0086] S36,通过Owlqn模型进行训练。

[0087] 在上述步骤中,通过Owlqn模型对样本磁盘的样本数据进行训练得到磁盘预测模型。

[0088] S37,得到磁盘的预测结果。

[0089] 在上述步骤中,使用上述步骤构建的磁盘预测模型对待测磁盘进行预测,得到预测值后,与模型中的预测取值范围进行比对,得到待测磁盘的预测结果。

[0090] 实施例2

[0091] 根据本发明实施例,还提供了一种用于实施上述磁盘的故障预测方法的装置,如图4所示,该装置包括:获取模块40、分类模块42、训练模块44和确定模块46。

[0092] 其中,获取模块40,用于通过磁盘监控技术获取磁盘的样本磁盘数据,其中,样本磁盘数据包括多个维度上的样本数据;分类模块42,用于采用Bucketing技术对样本磁盘数据进行分箱处理,对样本磁盘数据进行分类;训练模块44,用于采用Owlqn模型对分类后的样本磁盘数据进行样本训练,得到磁盘预测模型;确定模块46,用于在接收到待测磁盘的磁盘数据之后,使用磁盘预测模型对待测磁盘的磁盘数据进行处理,确定待测磁盘是否为故障磁盘。

[0093] 此处需要说明的是,上述获取模块40、分类模块42、训练模块44和确定模块46对应于实施例一种的步骤S21至步骤S27所实现的实例和应用场景相同,但不限于上述实施例一所公开的内容。需要说明的是,上述模块作为装置的一部分可以运行在实施例一提供的计算机终端10中。

[0094] 根据本申请上述实施例,在一种优选的方案中,样本磁盘数据为SMART磁盘数据,其中,样本磁盘数据至少包括如下四个维度上的样本数据:原始值、标准值、最差值和累积值。

[0095] 根据本申请上述实施例,在一种优选的方案中,结合图5所示,上述装置还包括:

[0096] 运算模块50,用于对每个维度上的样本数据进行如下任意一种或多种运算:差分运算、平方运算和分布求和运算,使得任意一个维度上的样本数据被扩展出新的维度上的样本数据。

[0097] 此处需要说明的是,上述获取模块50对应于实施例一种的步骤S211所实现的实例和应用场景相同,但不限于上述实施例一所公开的内容。需要说明的是,上述模块作为装置的一部分可以运行在实施例一提供的计算机终端10中。

[0098] 根据本申请上述实施例,在一种优选的方案中,结合图6所示,上述分类模块42包括:

[0099] 第一确定子模块60,用于确定预先划分的每个分箱的取值范围以及每个分箱对应的ID值;分类子模块62,用于通过将每个维度上的样本数据离散化至对应的分箱来对样本

磁盘数据进行分类,得到每个维度上的样本数据所对应的ID值。

[0100] 此处需要说明的是,上述第一确定子模块60和分类子模块62对应于实施例一种的步骤S231和步骤S233所实现的实例和应用场景相同,但不限于上述实施例一所公开的内容。需要说明的是,上述模块作为装置的一部分可以运行在实施例一提供的计算机终端10中。

[0101] 根据本申请上述实施例,在一种优选的方案中,结合图7所示,上述训练模块44包括:

[0102] 训练子模块70,用于Owlqn模型对每个维度上的样本数据所对应的ID值进行训练,得到每个维度上的样本数据的权重值;第二确定子模块72,用于根据每个维度上的样本数据及对应的权重值,确定磁盘预测模型,其中,磁盘预测模型包括每个维度上的样本数据的预测结果。

[0103] 此处需要说明的是,上述训练子模块70和第二确定子模块72对应于实施例一种的步骤S251和步骤S253所实现的实例和应用场景相同,但不限于上述实施例一所公开的内容。需要说明的是,上述模块作为装置的一部分可以运行在实施例一提供的计算机终端10中。

[0104] 根据本申请上述实施例,在一种优选的方案中,每个维度上的样本数据的预测结果为样本磁盘数据进行分类后得到的预测值。

[0105] 根据本申请上述实施例,在一种优选的方案中,结合图8所示,上述确定模块46还包括:

[0106] 离散模块80,用于接收到待测磁盘的磁盘数据之后,将待测磁盘的磁盘数据离散化至对应的分箱,得到待测磁盘的磁盘数据所对应的ID值;第三确定子模块82,用于根据待测磁盘的磁盘数据所对应的ID值确定待测磁盘的磁盘数据的权重值;第四确定子模块84,用于根据待测磁盘的磁盘数据的权重值从磁盘预测模型中确定待测磁盘是否为故障磁盘。

[0107] 此处需要说明的是,上述离散模块80、第三确定子模块82和第四确定子模块84对应于实施例一种的步骤S271和步骤S275所实现的实例和应用场景相同,但不限于上述实施例一所公开的内容。需要说明的是,上述模块作为装置的一部分可以运行在实施例一提供的计算机终端10中。

[0108] 实施例3

[0109] 本发明的实施例可以提供一种计算机终端,该计算机终端可以是计算机终端群中的任意一个计算机终端设备。可选地,在本实施例中,上述计算机终端也可以替换为移动终端等终端设备。

[0110] 可选地,在本实施例中,上述计算机终端可以位于计算机网络的多个网络设备中的至少一个网络设备。

[0111] 在本实施例中,上述计算机终端可以执行磁盘的故障预测方法中以下步骤的程序代码:通过磁盘监控技术获取磁盘的样本磁盘数据,其中,样本磁盘数据包括多个维度上的样本数据;采用Bucketing技术对样本磁盘数据进行分箱处理,对样本磁盘数据进行分类;采用Owlqn模型对分类后的样本磁盘数据进行样本训练,得到磁盘预测模型;在接收到待测磁盘的磁盘数据之后,使用磁盘预测模型对待测磁盘的磁盘数据进行处理,确定待测磁盘是否为故障磁盘。

[0112] 可选地,图9是根据本发明实施例的一种计算机终端的结构框图。如图9所示,该计算机终端A可以包括:一个或多个(图中仅示出一个)处理器91、存储器93、以及传输装置95。

[0113] 其中,存储器可用于存储软件程序以及模块,如本发明实施例中的磁盘的故障预测方法和装置对应的程序指令/模块,处理器通过运行存储在存储器内的软件程序以及模块,从而执行各种功能应用以及数据处理,即实现上述的磁盘的故障预测方法。存储器可包括高速随机存储器,还可以包括非易失性存储器,如一个或者多个磁性存储装置、闪存、或者其他非易失性固态存储器。在一些实例中,存储器可进一步包括相对于处理器远程设置的存储器,这些远程存储器可以通过网络连接至终端A。上述网络的实例包括但不限于互联网、企业内部网、局域网、移动通信网及其组合。

[0114] 处理器可以通过传输装置调用存储器存储的信息及应用程序,以执行下述步骤:通过磁盘监控技术获取磁盘的样本磁盘数据,其中,样本磁盘数据包括多个维度上的样本数据;采用Bucketing技术对样本磁盘数据进行分箱处理,对样本磁盘数据进行分类;采用Owlqn模型对分类后的样本磁盘数据进行样本训练,得到磁盘预测模型;在接收到待测磁盘的磁盘数据之后,使用磁盘预测模型对待测磁盘的磁盘数据进行处理,确定待测磁盘是否为故障磁盘。

[0115] 可选的,上述处理器还可以执行如下步骤的程序代码:样本磁盘数据为SMART磁盘数据,其中,样本磁盘数据至少包括如下四个维度上的样本数据:原始值、标准值、最差值和累积值。

[0116] 可选的,上述处理器还可以执行如下步骤的程序代码:对每个维度上的样本数据进行如下任意一种或多种运算:差分运算、平方运算和分布求和运算,使得任意一个维度上的样本数据被扩展出新的维度上的样本数据。

[0117] 可选的,上述处理器还可以执行如下步骤的程序代码:确定预先划分的每个分箱的取值范围以及每个分箱对应的ID值;通过将每个维度上的样本数据离散化至对应的分箱来对样本磁盘数据进行分类,得到每个维度上的样本数据所对应的ID值。

[0118] 可选的,上述处理器还可以执行如下步骤的程序代码:Owlqn模型对每个维度上的样本数据所对应的ID值进行训练,得到每个维度上的样本数据的权重值;根据每个维度上的样本数据及对应的权重值,确定磁盘预测模型,其中,磁盘预测模型包括每个维度上的样本数据的预测结果。

[0119] 可选的,上述处理器还可以执行如下步骤的程序代码:每个维度上的样本数据的预测结果为样本磁盘数据进行分类后得到的预测值。

[0120] 可选的,上述处理器还可以执行如下步骤的程序代码:接收到待测磁盘的磁盘数据之后,将待测磁盘的磁盘数据离散化至对应的分箱,得到待测磁盘的磁盘数据所对应的ID值;根据待测磁盘的磁盘数据所对应的ID值确定待测磁盘的磁盘数据的权重值;根据待测磁盘的磁盘数据的权重值从磁盘预测模型中确定待测磁盘是否为故障磁盘。

[0121] 在本发明实施例中,采用通过磁盘监控技术获取磁盘的样本磁盘数据,其中,样本磁盘数据包括多个维度上的样本数据;采用Bucketing技术对样本磁盘数据进行分箱处理,对样本磁盘数据进行分类;采用Owlqn模型对分类后的样本磁盘数据进行样本训练,得到磁盘预测模型的方式,通过在接收到待测磁盘的磁盘数据之后,使用磁盘预测模型对待测磁盘的磁盘数据进行处理,达到了确定待测磁盘是否为故障磁盘的目的,从而实现了预测磁

盘故障的技术效果,进而解决了现有技术的硬盘故障预测系统中一些容易致使硬盘故障的因素不能被采集或量化导致的预测结果不准确的技术问题。

[0122] 本领域普通技术人员可以理解,图9所示的结构仅为示意,计算机终端也可以是智能手机(如Android手机、iOS手机等)、平板电脑、掌上电脑以及移动互联网设备(Mobile Internet Devices, MID)、PAD等终端设备。图9其并不对上述电子装置的结构造成限定。例如,计算机终端A还可包括比图9中所示更多或者更少的组件(如网络接口、显示装置等),或者具有与图9所示不同的配置。

[0123] 本领域普通技术人员可以理解上述实施例的各种方法中的全部或部分步骤是可以通程序来指令终端设备相关的硬件来完成,该程序可以存储于一计算机可读存储介质中,存储介质可以包括:闪存盘、只读存储器(Read-Only Memory, ROM)、随机存取器(Random Access Memory, RAM)、磁盘或光盘等。

[0124] 实施例4

[0125] 本发明的实施例还提供了一种存储介质。可选地,在本实施例中,上述存储介质可以用于保存上述实施例一所提供的一种磁盘的故障预测方法所执行的程序代码。

[0126] 可选地,在本实施例中,上述存储介质可以位于计算机网络中计算机终端群中的任意一个计算机终端中,或者位于移动终端群中的任意一个移动终端中。

[0127] 可选地,在本实施例中,存储介质被设置为存储用于执行以下步骤的程序代码:通过磁盘监控技术获取磁盘的样本磁盘数据,其中,样本磁盘数据包括多个维度上的样本数据;采用Bucketing技术对样本磁盘数据进行分箱处理,对样本磁盘数据进行分类;采用Owlqn模型对分类后的样本磁盘数据进行样本训练,得到磁盘预测模型;在接收到待测磁盘的磁盘数据之后,使用磁盘预测模型对待测磁盘的磁盘数据进行处理,确定待测磁盘是否为故障磁盘。

[0128] 可选地,上述存储介质还被设置为存储用于执行以下步骤的程序代码:样本磁盘数据为SMART磁盘数据,其中,样本磁盘数据至少包括如下四个维度上的样本数据:原始值、标准值、最差值和累积值。

[0129] 可选地,上述存储介质还被设置为存储用于执行以下步骤的程序代码:对每个维度上的样本数据进行如下任意一种或多种运算:差分运算、平方运算和分布求和运算,使得任意一个维度上的样本数据被扩展出新的维度上的样本数据。

[0130] 可选地,上述存储介质还被设置为存储用于执行以下步骤的程序代码:确定预先划分的每个分箱的取值范围以及每个分箱对应的ID值;通过将每个维度上的样本数据离散化至对应的分箱来对样本磁盘数据进行分类,得到每个维度上的样本数据所对应的ID值。

[0131] 可选地,上述存储介质还被设置为存储用于执行以下步骤的程序代码:Owlqn模型对每个维度上的样本数据所对应的ID值进行训练,得到每个维度上的样本数据的权重值;根据每个维度上的样本数据及对应的权重值,确定磁盘预测模型,其中,磁盘预测模型包括每个维度上的样本数据的预测结果。

[0132] 可选地,上述存储介质还被设置为存储用于执行以下步骤的程序代码:每个维度上的样本数据的预测结果为样本磁盘数据进行分类后得到的预测值。

[0133] 可选地,上述存储介质还被设置为存储用于执行以下步骤的程序代码:接收到待测磁盘的磁盘数据之后,将待测磁盘的磁盘数据离散化至对应的分箱,得到待测磁盘的磁

盘数据所对应的ID值;根据待测磁盘的磁盘数据所对应的ID值确定待测磁盘的磁盘数据的权重值;根据待测磁盘的磁盘数据的权重值从磁盘预测模型中确定待测磁盘是否为故障磁盘。

[0134] 上述本发明实施例序号仅仅为了描述,不代表实施例的优劣。

[0135] 在本发明的上述实施例中,对各个实施例的描述都各有侧重,某个实施例中沒有详述的部分,可以参见其他实施例的相关描述。

[0136] 在本申请所提供的几个实施例中,应该理解到,所揭露的技术内容,可通过其它的方式实现。其中,以上所描述的装置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,例如多个单元或组件可以结合或者可以集成到另一个系统,或一些特征可以忽略,或不执行。另一点,所显示或讨论的相互之间的耦合或直接耦合或通信连接可以是通过一些接口,单元或模块的间接耦合或通信连接,可以是电性或其它的形式。

[0137] 所述作为分离部件说明的单元可以是或者也可以不是物理上分开的,作为单元显示的部件可以是或者也可以不是物理单元,即可以位于一个地方,或者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者全部单元来实现本实施例方案的目的。

[0138] 另外,在本发明各个实施例中的各功能单元可以集成在一个处理单元中,也可以是各个单元单独物理存在,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,也可以采用软件功能单元的形式实现。

[0139] 所述集成的单元如果以软件功能单元的形式实现并作为独立的产品销售或使用,可以存储在一个计算机可读取存储介质中。基于这样的理解,本发明的技术方案本质上或者说对现有技术做出贡献的部分或者该技术方案的全部或部分可以以软件产品的形式体现出来,该计算机软件产品存储在一个存储介质中,包括若干指令用以使得一台计算机设备(可为个人计算机、服务器或者网络设备等)执行本发明各个实施例所述方法的全部或部分步骤。而前述的存储介质包括:U盘、只读存储器(ROM,Read-Only Memory)、随机存取存储器(RAM,Random Access Memory)、移动硬盘、磁碟或者光盘等各种可以存储程序代码的介质。

[0140] 以上所述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。

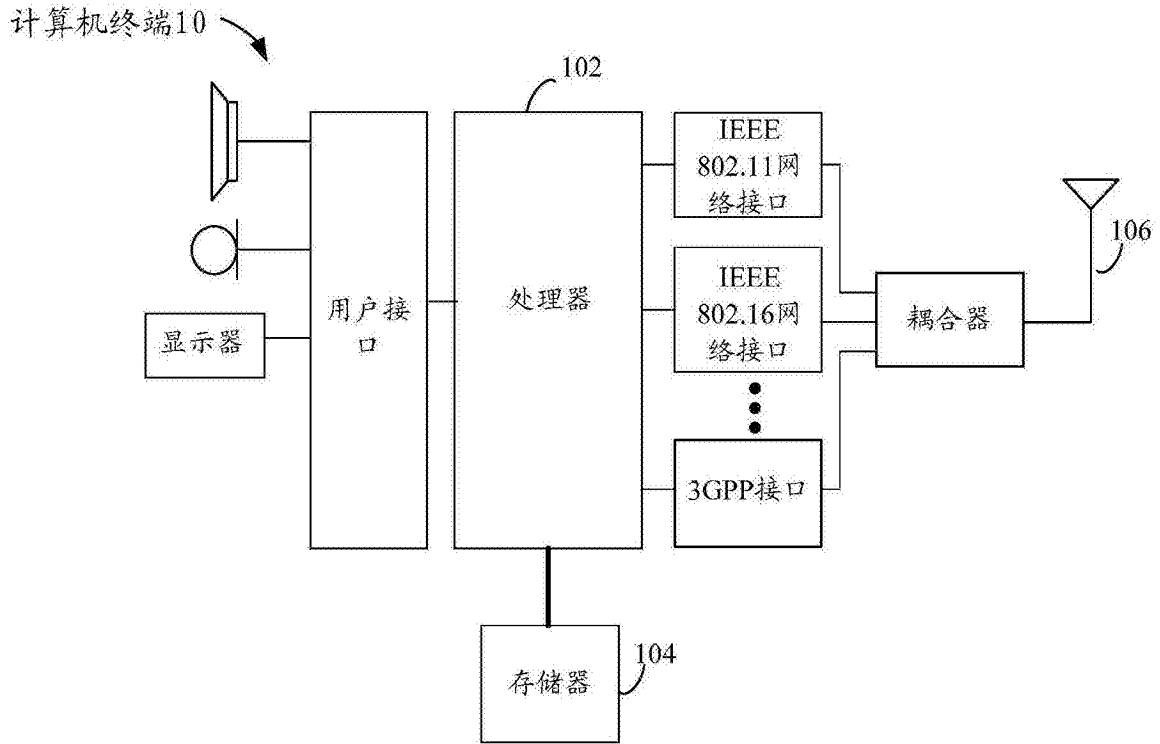


图1

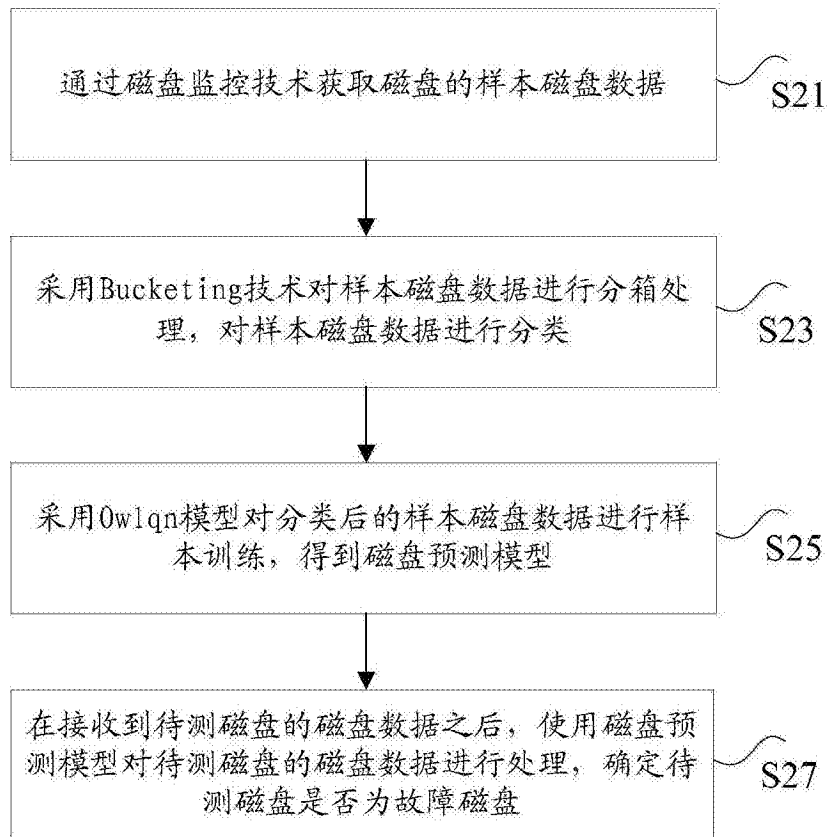


图2



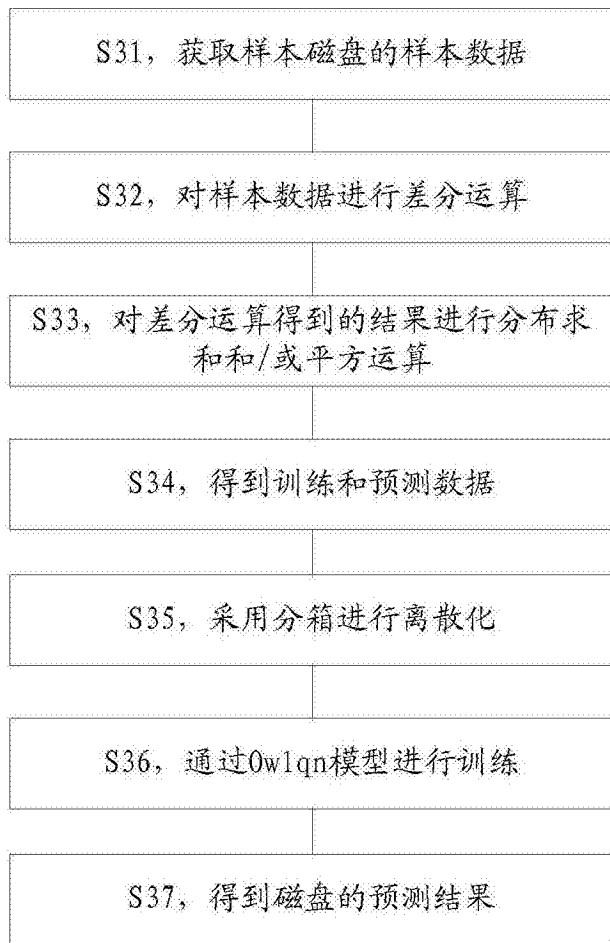


图3



图4

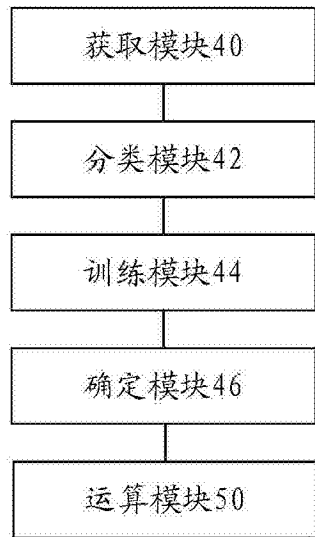


图5



图6

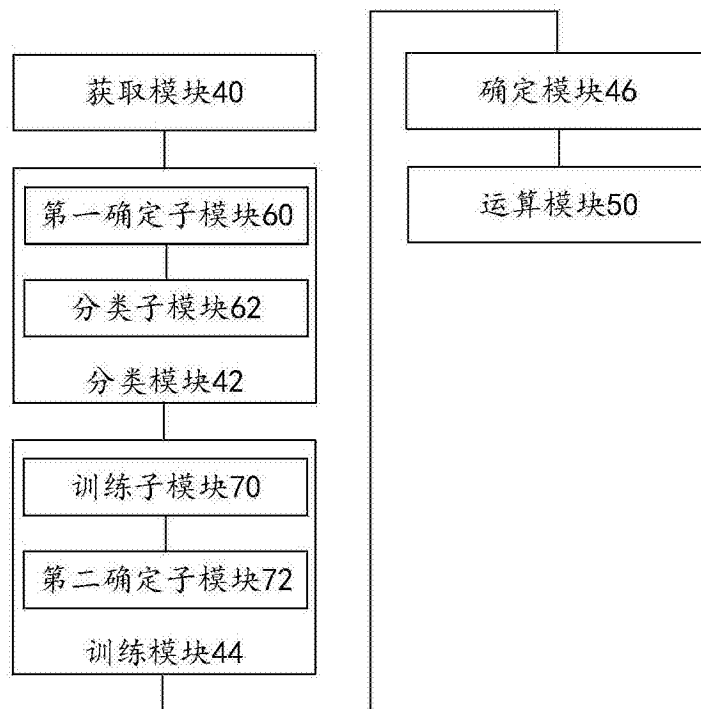


图7

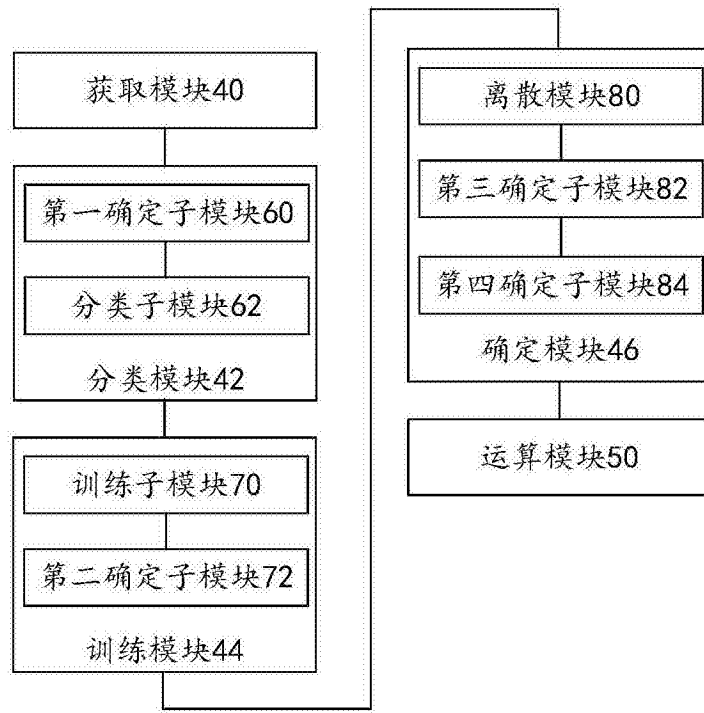


图8

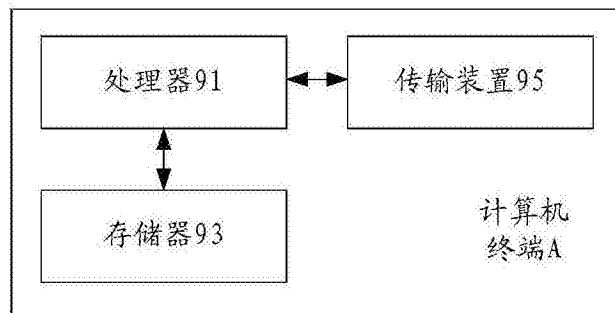


图9