



(12)发明专利申请

(10)申请公布号 CN 109919335 A

(43)申请公布日 2019.06.21

(21)申请号 201910179477.3

(22)申请日 2019.03.11

(71)申请人 西安电子科技大学

地址 710071 陕西省西安市太白南路2号

(72)发明人 杨冲昊

(74)专利代理机构 北京华仲龙腾专利代理事务
所(普通合伙) 11548

代理人 李静

(51)Int.Cl.

G06Q 10/00(2012.01)

G06Q 10/04(2012.01)

G06N 3/04(2006.01)

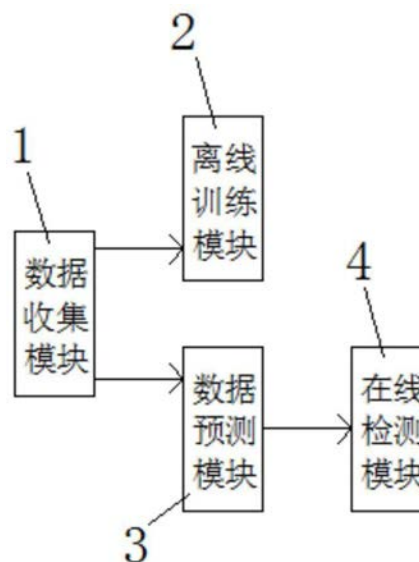
权利要求书1页 说明书4页 附图1页

(54)发明名称

基于深度学习的磁盘故障预测系统

(57)摘要

本发明公开了基于深度学习的磁盘故障预测系统,涉及磁盘故障预测技术领域,包括数据收集模块和数据预测模块,所述数据收集模块的输出端电性连接有离线训练模块,所述数据预测模块电性连接于数据收集模块的输出端,且数据预测模块的输出端电性连接有在线检测模块。本发明的有益效果是:该基于深度学习的磁盘故障预测系统,通过数据收集模块、离线训练模块、数据预测模块及在线检测模块的共同作用,便于替代传统的阈值算法中的当某一属性超过预定义的值时,触发SMART标记预警,从而达到提前预估磁盘故障的作用,有效避免其对磁盘的故障预测检测率下降的现象。



1. 基于深度学习的磁盘故障预测系统, 包括数据收集模块(1) 和数据预测模块(3), 其特征在于: 所述数据收集模块(1) 的输出端电性连接有离线训练模块(2), 所述数据预测模块(3) 电性连接于数据收集模块(1) 的输出端, 且数据预测模块(3) 的输出端电性连接有在线检测模块(4)。

2. 根据权利要求1所述的基于深度学习的磁盘故障预测系统, 其特征在于: 所述数据收集模块(1) 的终端设备为磁盘。

3. 根据权利要求1所述的基于深度学习的磁盘故障预测系统, 其特征在于: 所述离线训练模块(2) 由卷积神经网络对数据进行训练。

4. 根据权利要求1所述的基于深度学习的磁盘故障预测系统, 其特征在于: 所述数据预测模块(3) 通过循环神经网络训练的数据得出数据预测模型。

5. 根据权利要求1所述的基于深度学习的磁盘故障预测系统, 其特征在于: 所述在线检测模块(4) 与数据预测模块(3) 之间为单向电性连接。

6. 根据权利要求1所述的基于深度学习的磁盘故障预测系统, 其特征在于: 所述在线检测模块(4) 通过检测器来对磁盘故障做出判断。

基于深度学习的磁盘故障预测系统

技术领域

[0001] 本发明涉及磁盘故障预测技术领域,具体为基于深度学习的磁盘故障预测系统。

背景技术

[0002] 随着科学技术的发展及社会时代的进步,数据的海量存储问题给数据中心带来了新的挑战,虽然如今SSD的存在给存储带来新的方式,但是由于成本等各种原因,数据中心仍主要由硬盘存储占主要地位,而随着数据量的增大,存储容量和磁盘数量的增加。

[0003] 现有的磁盘在使用过程中,通过利用阈值算法来监控其的工作状态,但当某一属性超过预定义的值时,将会触发SMART标记预警,从而达到提前预估磁盘故障的作用,然而使得其对磁盘的故障预测检测率下降,则磁盘发生故障不易检测,从而导致数据丢失,影响人力财力。

发明内容

[0004] 本发明的目的在于提供基于深度学习的磁盘故障预测系统,解决了上述背景技术中提出的磁盘在使用过程中,通过利用阈值算法来监控其的工作状态,但当某一属性超过预定义的值时,将会触发SMART标记预警,从而达到提前预估磁盘故障的作用,然而使得其对磁盘的故障预测检测率下降,则磁盘发生故障不易检测,从而导致数据丢失,影响人力财力的问题。

[0005] 为实现上述目的,本发明提供如下技术方案:基于深度学习的磁盘故障预测系统,包括数据收集模块和数据预测模块,所述数据收集模块的输出端电性连接有离线训练模块,所述数据预测模块电性连接于数据收集模块的输出端,且数据预测模块的输出端电性连接有在线检测模块。

[0006] 可选的,所述数据收集模块的终端设备为磁盘。

[0007] 可选的,所述离线训练模块由卷积神经网络对数据进行训练。

[0008] 可选的,所述数据预测模块通过循环神经网络训练的数据得出数据预测模型。

[0009] 可选的,所述在线检测模块与数据预测模块之间为单向电性连接。

[0010] 可选的,所述在线检测模块通过检测器来对磁盘故障做出判断。

[0011] 本发明提供了基于深度学习的磁盘故障预测系统,具备以下有益效果:

[0012] 该基于深度学习的磁盘故障预测系统,通过数据收集模块、离线训练模块、数据预测模块及在线检测模块的共同作用,便于替代传统的阈值算法中的当某一属性超过预定义的值时,触发SMART标记预警,从而达到提前预估磁盘故障的作用,有效避免其对磁盘的故障预测检测率下降的现象,同时,通过数据收集模块的作用,能够根据磁盘SMART技术导出过去一周带有已标记故障、非故障的磁盘数据,且在离线训练模块的作用下,能够将所有标记磁盘数据指标输入到卷积神经网络中进行训练,得到一个故障检测模型,而通过数据预测模块的作用,能够对单个磁盘过去一段时间的各项指标数据输入到循环神经网络中,得到一个数据预测模型输出为规定时间内该磁盘的各项指标的预测值,且在在线检测模块的

作用下,能够将数据预测模型的输出作为输入,将预测出的数据输入到卷积神经网络故障检测器,从而判断该磁盘在未来是否发生故障。

附图说明

[0013] 图1为本发明基于深度学习的磁盘故障预测系统的流程结构示意图;

[0014] 图2为本发明基于深度学习的磁盘故障预测系统的ROC曲线建模结构示意图。

[0015] 图中:1、数据收集模块;2、离线训练模块;3、数据预测模块;4、在线检测模块。

具体实施方式

[0016] 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实施例,而不是全部的实施例。

[0017] 在本发明的描述中,除非另有说明,“多个”的含义是两个或两个以上;术语“上”、“下”、“左”、“右”、“内”、“外”、“前端”、“后端”、“头部”、“尾部”等指示的方位或位置关系为基于附图所示的方位或位置关系,仅是为了便于描述本发明和简化描述,而不是指示或暗示所指的装置或元件必须具有特定的方位、以特定的方位构造和操作,因此不能理解为对本发明的限制。此外,术语“第一”、“第二”、“第三”等仅用于描述目的,而不能理解为指示或暗示相对重要性。

[0018] 在本发明的描述中,需要说明的是,除非另有明确的规定和限定,术语“相连”、“连接”应做广义理解,例如,可以是固定连接,也可以是可拆卸连接,或一体地连接;可以是机械连接,也可以是电连接;可以是直接相连,也可以通过中间媒介间接相连。对于本领域的普通技术人员而言,可以根据具体情况理解上述术语在本发明中的具体含义。

[0019] 请参阅图1,本发明提供一种技术方案:基于深度学习的磁盘故障预测系统,包括数据收集模块1、离线训练模块2、数据预测模块3和在线检测模块4,数据收集模块1的输出端电性连接有离线训练模块2,数据预测模块3电性连接于数据收集模块1的输出端,且数据预测模块3的输出端电性连接有在线检测模块4。

[0020] 实施例1

[0021] 请参阅图2,事件预测的目标是在将来尽可能准确地预测磁盘故障,给定二元分类器和输入,一般有四种结果:假设输入为正类。如果分类器预测为正,则结果为真阳性。如果分类器预测为负,则结果为假阴性。如果输入为负且结果为负,那么它是真正负数,如果输入为负结果为正,则为误报;

[0022] 准确度:真阳性数量/假阳性数量;

[0023] 精确度:真阳性数量/真阳性和假阳性的总数;

[0024] 召回率:真阳性数量/阳性总数量;

[0025] $TPR = \text{正确预测故障的磁盘数量} / \text{实际故障的磁盘总数}$;

[0026] $FPR = \text{错误预测故障的磁盘数量} / \text{未故障的磁盘总数}$;

[0027] 通常,FPR与TPR曲线称为“接收器操作特性”(ROC)曲线,对于任何预测均匀分布在正样本和负样本中的随机预测算法,真阳性率等于假阳性率。因此,ROC中的对角线表示真正无偏的随机算法的性能,如果ROC高于对角线,则算法不是真正随机的并且具有一些预测能力。对于所有FPR,理想算法将具有 $TPR = 1, 0 \leq FPR \leq 1$,对于任何现实技术,TPR随着FPR

的增加而增加,最初TPR随着FPR的增加而急剧增加,然后随着FPR的增加,TPR的增长速度减慢;

[0028] 预测算法的性能通常以“ROC曲线下面积”(AUC)为特征,对于任何真正无偏的随机算法,AUC=0.5,另一方面,对于理想算法,AUC=1,我们根据经验将ROC曲线建模。

[0029] 实施例2

[0030] 导出数据有:磁盘是否故障标签(正负1,+1表示磁盘正常,-1表示磁盘故障)、原始读取错误率,旋转时间,重新分配的扇区数,错误率,开机时间,报告的不可纠正的错误,温度,回复的硬件,当前待定扇区数,重新分配的扇区数等

[0031]

```
00001, -1, 0.451613, 0.259259, -1.000000, 0.317073, 0.894737, 0.878788, 0.898990, -0.657143, -0.139785, 1.000000, 0.935255, -1.000000
00001, -1, 0.483871, 0.259259, -1.000000, 0.317073, 0.894737, 0.878788, 0.898990, -0.657143, -0.118280, 1.000000, 0.935255, -1.000000
00001, -1, 0.000000, 0.259259, -1.000000, 0.317073, 0.894737, 0.878788, 0.898990, -0.657143, -0.118280, 1.000000, 0.935255, -1.000000
00001, -1, 0.096774, 0.259259, -1.000000, 0.317073, 0.894737, 0.878788, 0.898990, -0.714286, -0.139785, 1.000000, 0.935255, -1.000000
00001, -1, 0.193548, 0.259259, -1.000000, 0.317073, 0.894737, 0.878788, 0.898990, -0.657143, -0.139785, 1.000000, 0.935255, -1.000000
00001, -1, 0.290323, 0.259259, -1.000000, 0.317073, 0.894737, 0.878788, 0.898990, -0.657143, -0.118280, 1.000000, 0.935255, -1.000000
00001, -1, 0.290323, 0.259259, -1.000000, 0.317073, 0.894737, 0.878788, 0.898990, -0.657143, -0.139785, 1.000000, 0.935255, -1.000000
00001, -1, 0.290323, 0.259259, -1.000000, 0.317073, 0.894737, 0.878788, 0.898990, -0.657143, -0.139785, 1.000000, 0.935255, -1.000000
00001, -1, 0.322581, 0.259259, -1.000000, 0.317073, 0.894737, 0.878788, 0.898990, -0.657143, -0.139785, 1.000000, 0.935255, -1.000000
00001, -1, 0.354839, 0.259259, -1.000000, 0.317073, 0.894737, 0.878788, 0.898990, -0.657143, -0.139785, 1.000000, 0.935255, -1.000000
00001, -1, 0.387097, 0.259259, -1.000000, 0.317073, 0.894737, 0.878788, 0.898990, -0.714286, -0.139785, 1.000000, 0.935255, -1.000000
00001, -1, 0.387097, 0.259259, -1.000000, 0.317073, 0.894737, 0.878788, 0.898990, -0.714286, -0.139785, 1.000000, 0.935255, -1.000000
00001, -1, 0.387097, 0.259259, -1.000000, 0.317073, 0.894737, 0.878788, 0.898990, -0.657143, -0.139785, 1.000000, 0.935255, -1.000000
00001, -1, 0.387097, 0.259259, -1.000000, 0.317073, 0.894737, 0.878788, 0.898990, -0.657143, -0.139785, 1.000000, 0.935255, -1.000000
00001, -1, 0.419355, 0.259259, -1.000000, 0.317073, 0.894737, 0.878788, 0.898990, -0.657143, -0.139785, 1.000000, 0.935255, -1.000000
00001, -1, 0.419355, 0.259259, -1.000000, 0.317073, 0.894737, 0.878788, 0.898990, -0.657143, -0.161290, 1.000000, 0.935255, -1.000000
00001, -1, 0.451613, 0.259259, -1.000000, 0.317073, 0.894737, 0.878788, 0.898990, -0.657143, -0.139785, 1.000000, 0.935255, -1.000000
00001, -1, 0.451613, 0.259259, -1.000000, 0.317073, 0.894737, 0.878788, 0.898990, -0.657143, -0.139785, 1.000000, 0.935255, -1.000000
```

[0032] 表1

[0033] 第1列:表示其序列号的磁盘索引;

[0034] 第2列:磁盘的类标签,对于故障的磁盘为-1,对于良好的磁盘为+1;

[0035] 第3列:SMART ID#1的值,原始读取错误率;

[0036] 第4列:SMART ID#3的值,旋转时间;

[0037] 第5列:SMART ID#5的值,重新分配的扇区数;

[0038] 第6列:SMART ID#7的值,寻求错误率;

[0039] 第7列:SMART ID#9的值,开机时间;

[0040] 第8列:SMART ID#187的值,报告的不可纠正的错误;

[0041] 第9列:SMART ID#189,High Fly Writes的值;

[0042] 第10列:SMART ID#194的值,温度摄氏度;

[0043] 第11列:SMART ID#195的值,恢复的硬件ECC;

[0044] 第12列:SMART ID#197的值,当前待定扇区计数;

[0045] 第13列:SMART ID#5的RAW_VALUE,重新分配的扇区数;

[0046] 第14列:SMART ID#197的RAW_VALUE,当前待定扇区计数。

[0047] 综上所述,该基于深度学习的磁盘故障预测系统,使用时,首先通过根据磁盘SMART技术导出过去一周带有已标记故障、非故障的磁盘数据,再将所有标记磁盘数据指标输入到卷积神经网络中进行训练,得到一个故障检测模型,随后,将单个磁盘过去一段时间的各项指标数据输入到循环神经网络中,得到一个数据预测模型输出为规定时间内该磁盘的各项指标的预测值,最后,通过将数据预测模型的输出作为输入,将预测出的数据输入到卷积神经网络故障检测器,从而判断该磁盘在未来是否发生故障。

[0048] 以上所述,仅为本发明较佳的具体实施方式,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,根据本发明的技术方案及其发明构思加以等同替换或改变,都应涵盖在本发明的保护范围之内。

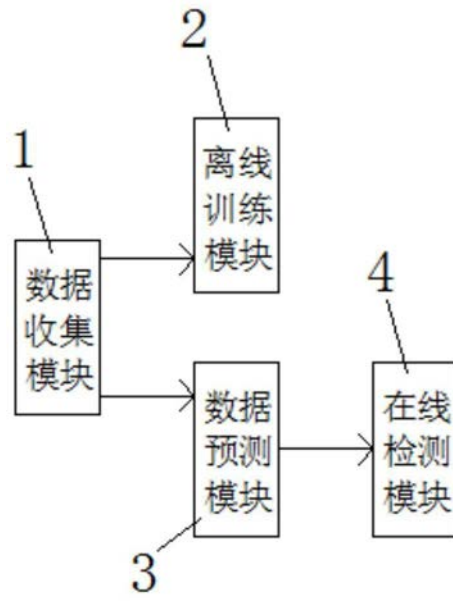


图1

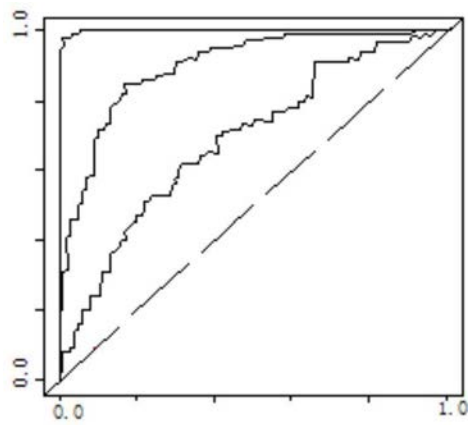


图2