



(12) 发明专利申请

(10) 申请公布号 CN 102129397 A

(43) 申请公布日 2011. 07. 20

(21) 申请号 201010611191. 7

(22) 申请日 2010. 12. 29

(71) 申请人 深圳市永达电子股份有限公司

地址 518000 广东省深圳市南山区高新技术南区深圳国际创新研究院研发大楼
D301-309 室

(72) 发明人 王明文 戚建淮

(74) 专利代理机构 北京英特普罗知识产权代理有限公司 11015

代理人 齐永红

(51) Int. Cl.

G06F 11/00 (2006. 01)

G06N 3/08 (2006. 01)

G11B 19/04 (2006. 01)

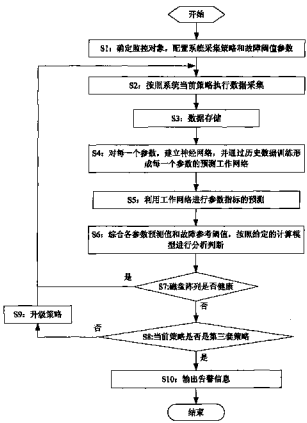
权利要求书 2 页 说明书 8 页 附图 2 页

(54) 发明名称

一种自适应磁盘阵列故障预测方法及系统

(57) 摘要

本发明公开了一种自适应磁盘阵列故障预测方法及系统,所述方法基于反向传播神经网络,通过神经网络具有的自学习机制,通过对能够表征磁盘阵列健康状态的参数的历史数据的搜集,通过逼近的方法总结出能够表征该参数随时间变化的规律,从而对该参数值进行预测。本方法共设置了第一、第二、第三由低到高的三套参数采集策略,参数个数逐渐递增而时间粒度逐渐递减,并提出了一种根据预测参数值和参数故障阈值的计算磁盘阵列健康状况的通用计算模型,从而能够有效预测磁盘阵列故障情况。自适应磁盘阵列故障预测系统包括信息采集模块、自适应配置模块、数据存储模块、自适应指标预测模块、参考故障阈值模块、分析决策模块、用户管理模块和用户界面模块。



1. 一种自适应磁盘阵列故障预测方法,其特征在于:包括如下主要步骤:

按照粗参数粒度、粗时间粒度策略采集一段时间内的磁盘阵列的健康和性能指标参数并建立每个指标参数的历史数据库;

每个指标参数的历史数据库建立后,对每一个新采集到的指标参数,根据其历史数据通过人工神经网络方法训练一参数预测网络;并利用训练好的参数预测网络对磁盘阵列的健康和性能指标参数进行初次预测;

将初次预测的结果与一定参数粒度、一定时间粒度所对应的一参考阈值进行比较,按照给定的计算模型进行计算和分析,如果判断出磁盘阵列未来一段时间内健康状态良好,则可以继续按照该参数粒度、时间粒度策略实施预测;如果判断出未来一段时间内磁盘阵列健康状态将要开始恶化,则增加检测参数的粒度并缩短预测故障的时间粒度,以得到磁盘阵列系统更精确的故障预测信息;

再对增加检测参数并缩短预测故障的时间粒度后的每一个新采集到的指标参数,根据其历史数据通过人工神经网络方法重新训练参数预测网络,并利用训练好的新的参数预测网络对磁盘阵列的健康和性能指标参数进行预测,得到预测结果;

将再次的预测结果与对应增加的检测参数粒度以及缩短的时间粒度的第二参考阈值进行比较,按照给定的计算模型进行计算和分析,如果判断出磁盘阵列未来一段时间内健康状态良好,则可以保持当前的检测参数和时间粒度实施预测;反之,如果判断出磁盘阵列系统即将发生故障时,则产生报警信息,通知相关人员或者外部系统进行处理。

2. 根据权利要求1所述的自适应磁盘阵列故障预测方法,其特征在于:

前述增加检测参数粒度并缩短时间粒度的第二次预测可以有多次,如果在多次预测中的任意一次判断出磁盘阵列未来一段时间内健康状态良好,则可以保持当前的检测参数和时间粒度实施预测;如果判断出未来一段时间内磁盘阵列健康状态将要进一步恶化,则进一步增加检测参数至细参数粒度并缩短预测故障的时间粒度至细时间粒度,以得到磁盘阵列系统更精确的健康信息;直至判断出磁盘阵列系统即将发生故障时,则产生报警信息,通知相关人员或者外部系统进行处理。

3. 根据权利要求1所述的自适应磁盘阵列故障预测方法,其特征在于:所述的磁盘阵列系统健康和性能参数,至少包括磁盘数据读取速度、磁盘数据写入速度、磁盘数据读写速度、错误读取率、启动/停止次数、重新分配扇区数、旋转重试次数、磁盘校准重试次数、ULTRA DMA 奇偶校验错误率、多区域错误率等。

4. 根据权利要求1所述的自适应磁盘阵列故障预测方法,其特征在于:所述的检测参数粒度和检测时间粒度,可以由系统管理人员通过系统进行人工配置得到。

5. 一种自适应磁盘阵列故障预测系统,与外部系统磁盘阵列系统和告警系统分别相连,故障预测系统定期采集磁盘阵列的健康和性能参数,通过系统内部的参数预测和综合分析,在预测出磁盘阵列故障时向外部告警系统发出告警信号,由外部告警系统进行进一步的处理;该磁盘阵列故障预测系统包括:

信息采集模块:定期采集磁盘阵列的健康和性能参数信息,通过数据的规整转化,装载进统一的信息对象数据结构,方便后续处理;

自适应配置模块:根据目前所执行的采集策略以及对磁盘阵列健康状态的预测结果,自适应的决定是否更换所执行的采集策略;

数据存储模块 :负责将系统内各种数据存入相应的数据库中 ;

自适应指标预测模块 :根据系统所确定的采集策略,利用最近一段时间的历史数据自适应的构建针对不同参数指标的多个故障预测神经网络 ;其中首先建立神经网络,然后通过统计一段时间内系统的历史健康和性能指标参数,输入神经网络以训练神经网络的权值,达到训练结束条件后,就得到了针对某一指标的训练好的神经网络 ;对于训练好的神经网络,输入实时采集的前段时间的参数后,输出得到下一时间单位的参数预测值 ;

参考故障阈值模块 :读取磁盘阵列系统健康和性能参数的阈值 ;

分析决策模块 :对自适应指标预测模块输出的参数预测值进行分析,根据所监控磁盘阵列的型号以及相关参数故障阈值信息,基于系统的计算模型进行综合的分析计算,得出所监控磁盘阵列系统的健康状态,并根据系统所采用的监控策略决策下一步的操作 ;

用户管理模块 :负责对系统各级用户的增加、删除、授权以及用户的认证,只有经过授权的用户合法用户才能查阅所监控的磁盘阵列对象健康状态信息 ;以及,

用户界面模块 :是用户与系统进行交互的界面,系统管理人员通过系统展示模块进行系统配置和系统管理 ;系统展示模块通过各种报表、折线图、拓扑图等形式展示系统当前及未来一段时间内全局的健康状态信息及磁盘阵列健康状态的发展趋势等。

一种自适应磁盘阵列故障预测方法及系统

技术领域

[0001] 本发明涉及计算机网络管理技术领域,具体涉及一种磁盘阵列故障预测系统的实现方法和装置。

背景技术

[0002] 随着大型计算机信息系统的不断投入使用,海量数据存储系统获得了越来越广泛的应用,其中最重要的是磁盘阵列系统(RAID, Redundant Array of Independent Disks)。RAID 最初是由加利福尼亚大学伯克利分校的 D. A. Patterson 教授在其论文中提出来的,作为一种数据保存手段,它的作用是提供专用服务器中接入多个磁盘(专指硬盘)时,以磁盘阵列方式组成一个超大容量、响应速度快、可靠性高的存储子系统。

[0003] 通过数据分块技术和冗余机制,磁盘阵列系统具备了一定的数据可靠性保障,但这种保障机制是有限的:具体到某一业务系统的磁盘阵列,往往使用质量一致性相近的相同型号和批次的磁盘,在长期使用之后,当一块磁盘出现故障时,其它磁盘出现故障的可能性也就很大,如果短时间内多块磁盘连续发生故障,则依靠磁盘阵列系统自身的可靠性机制无法恢复数据。

[0004] 由于磁盘阵列上往往存储着非常重要的业务数据,一旦出现不可恢复的灾难性故障,将会造成难以估量的损失。目前磁盘阵列运行在线监控系统,往往只能对已经发生的磁盘阵列故障进行检测和告警,更进一步,人们希望能够对磁盘阵列系统可能发生的故障进行预测。申请号 200510088506.3 的中国专利“磁盘装置的故障预测方法及使用该方法的磁盘装置”以及申请号 200710004243.2 的中国专利“预测存储设备故障的装置、系统和方法”分别对此进行了阐述。在传统的磁盘故障预测方法中,主要是通过实时采集磁盘系统的故障监控指标,并通过与预设的预测故障指标阈值进行比较,当实时采集的指标超过预测故障指标阈值时,通过一定的分析计算后进行判别处理。传统的磁盘阵列故障预测方法主要存在如下问题:

[0005] (1) 参数采集方式缺乏自适应性。往往只是对固定的几个故障预测参数进行采集和计算,所以只能针对特定的磁盘阵列系统进行信息采集和故障预测。

[0006] 对于采集参数多采集周期短的系统,往往影响系统性能。而对于采集参数少的系统,故障预测的准确性又受到影响。

[0007] 缺乏对故障告警时间的提前预测。往往是在检测到故障参数超过阈值时发出告警,而对于何时故障参数将会超过阈值并没有任何的预测或者预测精度差,而当故障参数超过阈值时很可能离磁盘出现实际故障的时间已经很短,不利于备品备件的提前准备。

发明内容

[0008] 本发明针对上述问题,提出一种自适应磁盘阵列故障预测系统,可以根据磁盘阵列的历史运行数据推断磁盘阵列所处的健康状况,进而根据磁盘阵列的健康状况自适应的调节故障检测参数的数量和检测周期,达到对磁盘阵列故障预测的自适应粒度调节,从而

更加高效更加可靠的保护磁盘数据。

[0009] 为了实现上述目的,本发明所描述的自适应磁盘阵列故障预测方法,包括如下主要步骤:按照粗参数粒度、粗时间粒度策略采集一段时间内的磁盘阵列的健康和性能指标参数并建立每个指标参数的历史数据库;每个指标参数的历史数据库建立后,对每一个新采集到的指标参数,根据其历史数据通过人工神经网络方法训练一参数预测网络;并利用训练好的参数预测网络对磁盘阵列的健康和性能指标参数进行初次预测;将初次预测的结果与一定参数粒度、一定时间粒度所对应的一参考阈值进行比较,按照给定的计算模型进行计算和分析,如果判断出磁盘阵列未来一段时间内健康状态良好,则可以继续按照该参数粒度、时间粒度策略实施预测;如果判断出未来一段时间内磁盘阵列健康状态将要开始恶化,则增加检测参数的粒度并缩短预测故障的时间粒度,以得到磁盘阵列系统更精确的故障预测信息;再对增加检测参数并缩短预测故障的时间粒度后的每一个新采集到的指标参数,根据其历史数据通过人工神经网络方法重新训练参数预测网络,并利用训练好的新的参数预测网络对磁盘阵列的健康和性能指标参数进行预测,得到预测结果;

[0010] 将再次的预测结果与对应增加的检测参数粒度以及缩短的时间粒度的第二参考阈值进行比较,按照给定的计算模型进行计算和分析,如果判断出磁盘阵列未来一段时间内健康状态良好,则可以保持当前的检测参数和时间粒度实施预测;反之,如果判断出磁盘阵列系统即将发生故障时,则产生报警信息,通知相关人员或者外部系统进行处理。

[0011] 前述增加检测参数粒度并缩短时间粒度的第二次预测可以有几次,如果在多次预测中的任意一次判断出磁盘阵列未来一段时间内健康状态良好,则可以保持当前的检测参数和时间粒度实施预测;如果判断出未来一段时间内磁盘阵列健康状态将要进一步恶化,则进一步增加检测参数至细参数粒度并缩短预测故障的时间粒度至细时间粒度,以得到磁盘阵列系统更精确的健康信息;直至判断出磁盘阵列系统即将发生故障时,则产生报警信息,通知相关人员或者外部系统进行处理。

[0012] 本发明所述的磁盘阵列系统健康和性能参数,至少包括磁盘数据读取速度、磁盘数据写入速度、磁盘数据读写速度、错误读取率、启动/停止次数、重新分配扇区数、旋转重试次数、磁盘校准重试次数、ULTRA DMA 奇偶校验错误率、多区域错误率等。

[0013] 本发明的方法所述的检测参数粒度和检测时间粒度,可以由系统管理人员通过系统进行人工配置得到。一般情况下初始的参数粒度可选 3~5 个主要的检测参数,中间参数粒度增加为 6~9 个检测参数,最终参数粒度增加为 10~15 个检测参数;初始的时间粒度可选为 1 个月,中间的时间粒度选为 1 周,最终的时间粒度可选为 1 天。实际的策略配置应根据系统的实际情况灵活掌握。当然,本领域一般技术人员应当理解,这些示例性数值并非意在限定本发明的保护范围。

[0014] 本发明所述的自适应磁盘阵列故障预测系统,与外部系统磁盘阵列系统和告警系统分别相连。故障预测系统定期采集磁盘阵列的健康和性能参数,通过系统内部的参数预测和综合分析,在预测出磁盘阵列故障时向外部告警系统发出告警信号,由外部告警系统进行进一步的处理。磁盘阵列故障预测系统自身包含有下述功能模块:

[0015] 信息采集模块:定期采集磁盘阵列的健康和性能参数信息,通过数据的规整转化,装载进统一的信息对象数据结构,方便后续处理。

[0016] 自适应配置模块:根据目前所执行的采集策略以及对磁盘阵列健康状态的预测结

果,自适应的决定是否更换所执行的采集策略。具体的说,即是否由第一套策略上升为第二套策略,或者是否由第二套策略上升为第三套策略。

[0017] 数据存储模块:负责将系统内各种数据存储进相应的数据库中,主要的数据包括系统定时采集的磁盘阵列健康和性能参数,承担参数预测工作的人工神经网络的各种网络参数,以及各种健康和性能参数的预测故障参考阈值等。

[0018] 自适应指标预测模块:根据系统所确定的采集策略,利用最近一段时间的历史数据自适应的构建针对不同参数指标的多个故障预测神经网络。首先建立神经网络,然后通过统计一段时间内系统的历史健康和性能指标参数,输入神经网络以训练神经网络的权值,达到训练结束条件后,就得到了针对某一指标的训练好的神经网络。对于训练好的神经网络,输入实时采集的前段时间的参数后,输出得到下一时间单位的参数预测值。

[0019] 参考故障阈值模块:读取磁盘阵列系统健康和性能参数的阈值,系统管理员可以通过手工重新设置各参数的阈值。

[0020] 分析决策模块:对自适应指标预测模块输出的参数预测值进行分析,根据所监控磁盘阵列的型号以及相关参数故障阈值信息,基于系统的计算模型进行综合的分析计算,得出所监控磁盘阵列系统的健康状态,并根据系统所采用的监控策略决策下一步的操作。

[0021] 用户管理模块:负责对系统各级用户的增加、删除、授权以及用户的认证,只有经过授权的用户合法用户才能查阅所监控的磁盘阵列对象健康状态信息。用户的角色至少包括系统管理员和普通用户两级,用户需提交的信息包括但不限于姓名、部门、职位、电子邮箱、移动电话、办公电话等。

[0022] 用户界面模块:是用户与系统进行交互的界面,系统管理人员通过系统展示模块进行系统配置和系统管理。系统展示模块通过各种报表、折线图、拓扑图等形式展示系统当前及未来一段时间内全局的健康状态信息及磁盘阵列健康状态的发展趋势等。

[0023] 采用上述技术方案,本发明有益的技术效果在于:通过建立针对磁盘阵列系统的故障预测系统,可以根据磁盘阵列的健康和性能参数的历史数据推断其未来的参数指标,从而对磁盘阵列系统可能发生的故障进行提前预测。由于采用了一系列自适应的技术,既能够保证系统的性能,又能较好的保证系统的故障预测精度,从而有效的保护磁盘阵列系统上所存储的数据。

附图说明

[0024] 下面将结合附图及实施例对本发明作进一步说明,附图中:

[0025] 图 1 是本发明自适应磁盘阵列故障预测方法及系统实施例的结构框图;

[0026] 图 2 是本发明自适应磁盘阵列故障预测方法及系统中用于参数预测的人工神经网络的网络拓扑图;

[0027] 图 3 是本发明自适应磁盘阵列故障预测方法及系统中故障预测方法的流程图。

具体实施方式

[0028] 如图 1 所示,是本发明自适应磁盘阵列故障预测系统实施例的结构框图。根据该框图,本发明一种具体实施方式中的自适应磁盘阵列故障预测系统 12 需要通过互联网络与磁盘阵列系统 11 以及外部告警系统 13 进行连接。磁盘阵列故障预测系统 12 实现磁盘

阵列故障的分析和预测功能,该系统包括信息采集模块 121、自适应配置模块 122、数据存储模块 123、自适应指标预测模块 124、参考故障阈值模块 125、分析决策模块 126、用户管理模块 127 以及用户界面模块 128。

[0029] 信息采集模块 121 提供对于磁盘阵列系统运行期间健康和性能参数的采集功能。信息采集模块 121 接收来自自适应配置模块 122 的配置策略信息,并按照设定的配置策略对磁盘阵列系统 11 进行信息采集,采集的信息经规格化后交由数据存储模块 123 进行处理。

[0030] 自适应配置模块 122 的功能是确定磁盘阵列系统的信息采集策略。自适应配置模块 122 接收来自用户界面模块 128 的用户输入配置信息,以及来自分析决策模块 126 的磁盘阵列预测健康状态信息,判断是维持本级策略还是升级到更高一级策略,并依此向信息采集模块 121 发出控制指令。

[0031] 数据存储模块 123 用于存储系统内的重要数据。数据存储模块接收来自信息采集模块 121 提供的磁盘阵列健康和性能参数信息,接收来自参考故障阈值模块 125 提供的磁盘阵列及与其故障相关的健康和性能参数阈值信息,接收来自用户管理模块 127 提供的用户、用户组及权限信息,以及来自用户界面模块 128 的其它系统信息并存储在数据库中,供自适应指标预测模块 124、分析决策模块 126 以及用户界面模块 128 使用。

[0032] 自适应指标预测模块 124 用于构建针对各参数的基于人工神经网络的预测系统并进行参数值的预测。自适应指标预测模块 124 接收来自数据存储模块 123 的神经网络配置信息以及磁盘阵列参数历史数据用于训练针对该参数的故障预测网络,接收实时采集的数据对未来的参数值进行预测。自适应指标预测模块 124 输出预测结果给分析决策模块 126 进行磁盘阵列故障预测,同时将预测结果输出给数据存储模块 123 进行存储。

[0033] 参考故障阈值模块 125 接收来自用户界面模块 128 的故障阈值查询和配置信息,并输出给数据存储模块 123 进行存储。参考故障阈值模块 125 输出故障参数参考阈值给分析决策模块 126 进行计算分析。

[0034] 分析决策模块 126 提供对于磁盘阵列未来健康状态信息的计算和分析,并决定是否产生告警信息。分析决策模块 126 接收来自自适应指标预测模块 124 的参数预测信息和参考故障阈值模块 125 的参数故障阈值信息,按照给定的计算模型进行计算分析后,给出在当前策略下,磁盘阵列系统在未来一段时间内运行健康状态的信息。分析决策模块 126 输出分析结果到自适应配置模块 122,在预测出磁盘阵列的故障后,输出告警信息到用户界面模块 128,同时输出告警信息到外部告警系统 13。

[0035] 用户管理模块 127 提供对于用户和用户组的增加、删除、认证、授权和访问控制等功能。用户管理单元 127 接收用户的操作指令,并通过用户界面单元 128 展示相应的操作结果。

[0036] 用户界面模块 128 提供用户与磁盘阵列故障预测系统之间的交互界面,并提供磁盘阵列故障预测状态的集中展示界面。用户界面模块 128 接收来自用户的命令输入,并输出给用户管理模块 127、参考故障阈值模块 125、自适应指标预测模块 124、自适应配置模块 122 以及数据存储模块 123 以用于系统的配置。用户界面模块 128 接收来自数据存储模块 123 的数据信息,根据用户的查询请求,生成相应的统计报表和图形展示给用户。

[0037] 本发明所描述的自适应磁盘阵列故障预测方法,包括如下主要步骤:

[0038] 按照粗参数粒度、粗时间粒度策略采集磁盘阵列的健康和性能指标参数并作为历史数据进行存储；

[0039] 对每一个采集到的指标参数，根据其历史数据通过人工神经网络方法训练一参数预测网络；利用训练好的参数预测网络对磁盘阵列的健康和性能指标参数进行预测；

[0040] 将预测结果与粗参数粒度、粗时间粒度所对应的第一参考阈值进行比较，按照给定的计算模型进行计算和分析，如果判断出磁盘阵列未来一段时间内健康状态良好，则可以继续按照粗参数粒度、粗时间粒度策略实施预测；反之，如果判断出未来一段时间内磁盘阵列健康状态将要开始恶化，则增加检测参数至中参数粒度并缩短预测故障的时间粒度至中时间粒度，以得到磁盘阵列系统更精确的故障预测信息；

[0041] 对增加检测参数并缩短预测故障的时间粒度后的每一个新采集到的指标参数，根据其历史数据通过人工神经网络方法重新训练参数预测网络；利用训练好的新的参数预测网络对磁盘阵列的健康和性能指标参数进行预测；

[0042] 将预测结果与对应中参数粒度以及中时间粒度的第二参考阈值进行比较，按照给定的计算模型进行计算和分析，如果判断出磁盘阵列未来一段时间内健康状态良好，则可以保持当前的检测参数和时间粒度实施预测；反之，如果判断出未来一段时间内磁盘阵列健康状态将要进一步恶化，则进一步增加检测参数至细参数粒度并缩短预测故障的时间粒度至细时间粒度，以得到磁盘阵列系统更精确的健康信息；

[0043] 对每一个增加检测参数并缩短预测故障的时间粒度后采集到的指标参数，根据其历史数据通过人工神经网络方法训练参数预测网络；利用训练好的参数预测网络对磁盘阵列的健康和性能指标参数进行预测；

[0044] 将预测结果与第三参考阈值进行比较，按照给定的计算模型分析计算磁盘阵列系统所处的健康状态。当细粒度的检测和故障预测显示磁盘阵列系统即将发生故障时，则产生报警信息，通知相关人员或者外部系统进行处理。

[0045] 本发明所述的磁盘阵列系统健康和性能参数，至少包括磁盘数据读取速度、磁盘数据写入速度、磁盘数据读写速度、错误读取率、启动 / 停止次数、重新分配扇区数、旋转重试次数、磁盘校准重试次数、ULTRA DMA 奇偶校验错误率、多区域错误率等。

[0046] 在本发明所述的磁盘阵列故障预测实现方法中，所指的给定的计算模型其具体计算方法如下，这里设故障预测系统采集的磁盘阵列参数为 $A_i (i = 1, \dots, n)$ ，对于参数 $A_i (i = 1, \dots, n)$ ，其对应的阈值分别为 $T_i (i = 1, \dots, n)$ ，实际检测值为 $V_i (i = 1, \dots, n)$ ；

[0047] 对于参数 $A_i (i = 1, \dots, n)$ ，计算参数距离比率 $D_i = \text{abs}((V_i - T_i) / T_i)$ ，其中 $\text{abs}()$ 表示取绝对值函数。距离比率表示磁盘参数实际测量值距离参数故障预测阈值的远近，越小则距离越近；

[0048] 计算包含所有检测参数信息的总距离比率 $D = (D_1 + \dots + D_n) / n$ 。

[0049] 比较总距离比率 D 与参考总距离比率阈值 T ，(T 为按照策略人为设定的值)，如果 $D > T$ ，则磁盘阵列处于健康状态，如果 $D \leq T$ ，则启动下一套策略或发出故障预警信号。对于不同的操作策略， T 的取值不相同，一般情况下，对于初始的参数粒度和时间粒度，可以取 $T = 0.2$ ，对于中间的参数粒度和时间粒度，可以取 $T = 0.1$ ，对于最终的参数粒度和时间粒度，可以取 $T = 0.05$ 。

[0050] 在本发明所述的磁盘阵列故障预测方法中，所涉及到利用人工神经网络的方法进

行故障预测网络的构建。基本思想是通过神经网络 BP 学习算法（误差反向传播算法）来刻画磁盘阵列健康性能参数的变化趋势，从而对参数作出预测，最后再根据综合各种参数的预测值，对磁盘阵列可能的故障进行预测。BP 学习算法是一种有监督的学习过程，它是根据给定的（输入，输出）样本对来进行学习，并通过调整网络连接权来体现学习的效果。就整个神经网络来说，它有两种状态：

[0051] 第一是学习阶段：先将学习样本对的输入加在网络的输入端，沿着前向（即输入层→输出层）在各层神经元按输入和激励函数的方式产生输出。然后将输出层神经元的实际输出值和期望输出值之差逆向（即输出层→输入层）传播到各层神经元，并根据误差的大小和符号相应地调整各连接权值。此过程一直进行到神经网络权连接方式能够给定输入样本条件下以一定精度产生给定输出结果为止，即认为学习阶段结束；

[0052] 第二是工作阶段：当待测样本输入到已学习好的神经网络输入端时，根据类似输入产生类似输出的原则，神经网络按内插或外延的方式在输出端生成所求的解答来。

[0053] 如图 2 所示，神经网络纵向分层，将神经网络第 k 层第 i 元素的输入和记为 I_i^k ，输出记为 O_i^k ，将第 $k-1$ 层 i 元素向第 k 层 j 元素的连接权值记为 $w_{i,j}^{k-1,k}$ ，设神经元激励函数为 f ，这里 f 可选择常用的神经元激励函数，如 S 形函数。则神经元激励函数根据神经元的输入计算产生神经元的输出，即有

$$[0054] \quad O_i^k = f(I_i^k)$$

[0055] 其中

$$[0056] \quad I_j^k = \sum_i w_{i,j}^{k-1,k} O_i^{k-1}$$

[0057] 误差函数定义为输出层各元素误差的平方和，即

$$[0058] \quad r = \frac{1}{2} \sum_j (O_j^m(W, X) - y_j)^2$$

[0059] 式中， $O_j^m(W, X)$ 为与权矢量 W 和输入矢量 X 有关的第 m 层（即输出层）第 j 元素的实际输出值， y_j 为该元素的期望输出；根据梯度下降的原理，我们希望各层神经元相应权值的改变量正比于误差函数对该权值的负导数，即

$$[0060] \quad \Delta w_{i,j}^{k-1,k} = -\eta \frac{\partial r}{\partial w_{i,j}^{k-1,k}} = -\eta \frac{\partial r}{\partial I_j^k} \frac{\partial I_j^k}{\partial w_{i,j}^{k-1,k}}$$

$$[0061] \quad = -\eta \frac{\partial r}{\partial I_j^k} O_i^{k-1} = -\eta d_j^k O_i^{k-1}$$

[0062] 式中， η 为常数，根据经验指定。

[0063] 对于磁盘阵列故障预测网络，其具体的学习算法如下：

[0064] 首先根据应用需求，建立网络拓扑结构，即三层网络，多输入单输出的网络；

[0065] 再将网络各权值 $w_{i,j}$ 赋予小的非零随机实数值，设定学习率 η 和惯性系数 α 。

[0066] 再输入 p 个（输入，输出）样本对中的第 i ($i = 1, \dots, p$) 个样本对。注：对样本对而言，输入为根据输入节点数所取的历史数据的健康参数平均值，输出为下一时间段的实际健康参数平均值。如对于 5 输入 1 输出的网络，对于以天为时间粒度的情况，如果已经拥有之前从 1 号到 31 号总共 31 天的历史数据，则第一个样本对为（1 号数据，2 号数据，3

号数据,4 号数据,5 号数据),6 号数据),第二个样本对为((2 号数据,3 号数据,4 号数据,5 号数据,6 号数据),7 号数据),以此类推,第二十六个样本对为((26 号数据,27 号数据,28 号数据,29 号数据,30 号数据),31 号数据)。

[0067] 按公式 $O_j^k = f(I_j^k) = f(\sum_i w_{i,j}^{k-1,k} O_i^{k-1})$ 计算出针对第 i 个样本对的网络各层元素的实际输出值。

[0068] 并按下式调整各连接权值 $w_{i,j}^{k-1,k}(t+1) = \alpha w_{i,j}^{k-1,k}(t) - \eta d_j^k O_i^{k-1}$, 其中当 $k = m$, $d_j^m = (O_j^m - y_j) f'(I_j^m)$; 当 $k < m$, $d_j^k = (\sum_l w_{j,l}^{k,k+1} d_l^{k+1}) f'(I_j^k)$ 。

[0069] 最后输入第 $i+1$ 组样本对,……,循环利用 p 个样本对,直至 $w_{i,j}$ 趋于稳定不变为止。

[0070] 其具体的工作算法如下:

[0071] 在网络输入层输入前 n 个时刻的磁盘阵列某一健康性能指标实际测量值 $\lambda_1, \lambda_2, \dots, \lambda_n$ 。

[0072] 按公式 $O_j^k = f(I_j^k) = f(\sum_i w_{i,j}^{k-1,k} O_i^{k-1})$ 计算出网络各层元素的实际输出值。输出层的输出值即为对 $n+1$ 时刻性能指标的预测值 λ_{n+1} 。

[0073] 将输出的预测值重组为输入队列计算后续时刻的预测值(比如将 $\lambda_2, \dots, \lambda_n, \lambda_{n+1}$ 作为输入队列,即可预测出 $n+2$ 时刻的预测值 λ_{n+2}),以此类推,即可预测出后面一定时间段内的性能值。

[0074] 如图 2 所示,是图 1 中自适应指标预测模块 124 中用于预测的人工神经网络的网络拓扑图实施例。用于故障预测的 BP 神经网络结构为一个三层网络结构。输入层取 10 个神经元,其输入表示磁盘阵列按离散时间序列(1,2, ..., 10)采集到的表征磁盘阵列健康和性能状况的某参数,中间层取 4 个神经元,输出层取 1 个神经元,其输出表示在未来时间的磁盘阵列参数,即在粗时间粒度条件下通过前 10 个月的历史数据预测未来 1 个月的参数值,在中时间粒度条件下通过前 10 周的历史数据预测未来 1 周的值,在细时间粒度条件下通过前 10 天的历史数据预测未来 1 天的参数值。

[0075] 如图 3 所示,是本发明自适应磁盘阵列故障预测方法及系统的故障预测方法的主要流程图。该方法包括以下步骤:

[0076] 步骤 S1:用户通过用户界面模块 128 设置磁盘阵列故障预测系统 12 的监控对象,并配置系统所需要的各种系统参数,包括每一种采集策略所包含的参数个数、参数名称及预测时间粒度,每一种参数的预测故障阈值,人工神经网络的配置参数等。

[0077] 步骤 S2:信息采集模块 121 按照系统的既定策略,定期采集表征磁盘阵列系统 11 的健康和性能的参数信息,并进行规格化处理。

[0078] 步骤 S3:数据存储模块 123 对采集的信息进行存储,存储到数据库中供后续处理。

[0079] 步骤 S4:自适应指标预测模块 124 针对每一个参数,建立神经网络,并通过历史数据训练形成每一个参数的预测工作网络。

[0080] 步骤 S5:自适应指标预测模块 124 利用最近的历史数据,通过人工神经网络的自动计算对下一时间粒度单位的参数值进行预测。

[0081] 步骤 S6:分析决策模块 126 综合各参数预测值和故障参考阈值,按照给定的计算

模型进行分析判断。

[0082] 步骤 S7 :分析决策模块 126 判断磁盘阵列在当前策略下的预测值是否正常? 如果正常,则转步骤 S2 继续进行监控;否则转下一步。

[0083] 步骤 S8 :分析决策模块 126 判断当前策略是否是第三套策略(细粒度策略)? 如果是,则转步骤 S10;否则转下一步。

[0084] 步骤 S9 :分析决策模块 126 通知自适应配置模块 122 升级当前策略,即如果当前策略为第一套策略,则升级到第二套策略;如果当前策略为第二套策略,则升级到第三套策略。升级后转步骤 S2。

[0085] 步骤 S10 :分析决策模块 126 向用户界面模块 128 和外部告警系统 13 输出磁盘阵列预测故障告警信息。数据采集和预测处理过程完毕。

[0086] 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,可轻易想到的变化或替换,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该以权利要求的保护范围为准。

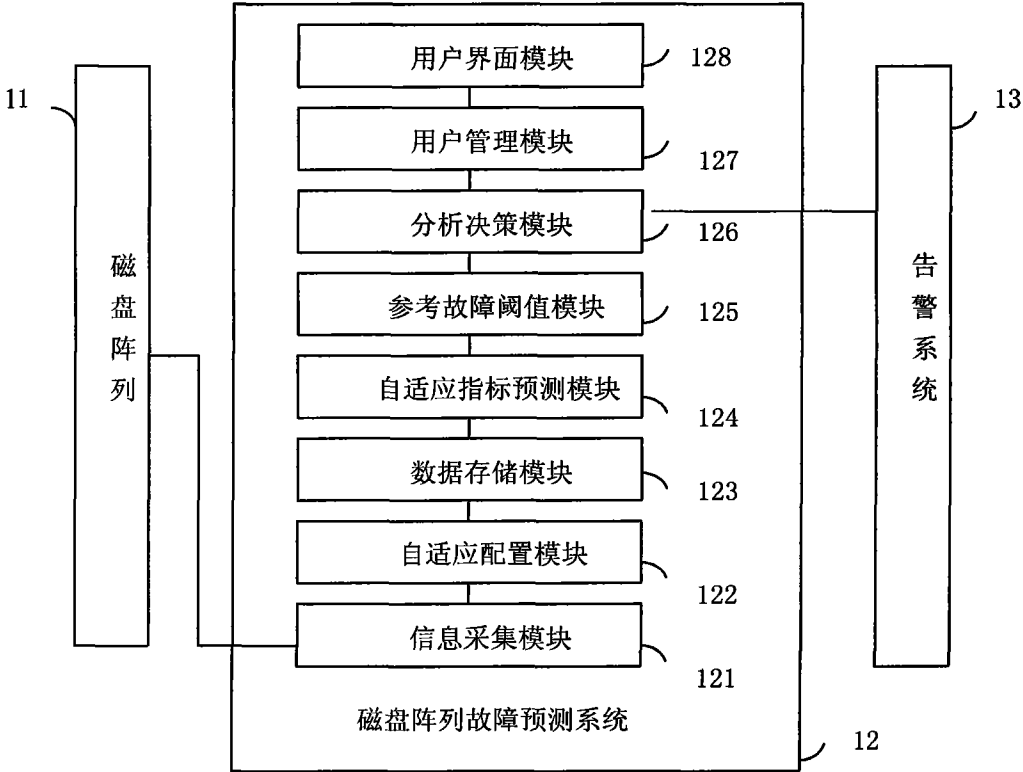


图 1

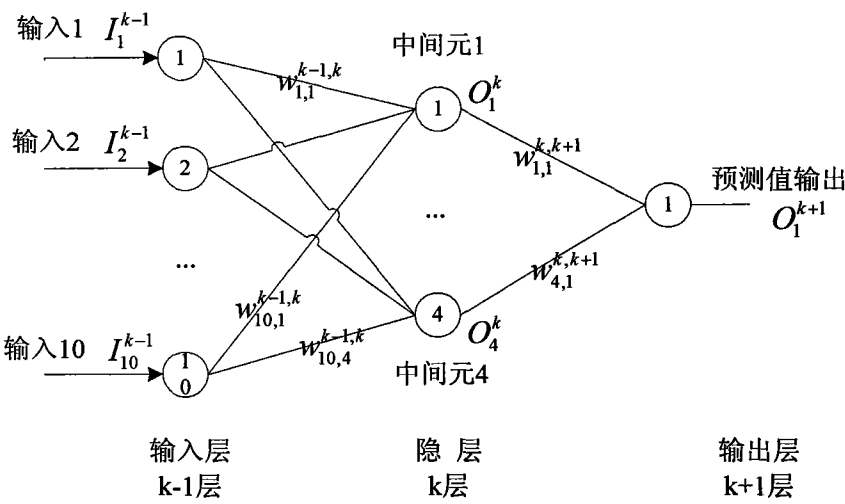


图 2

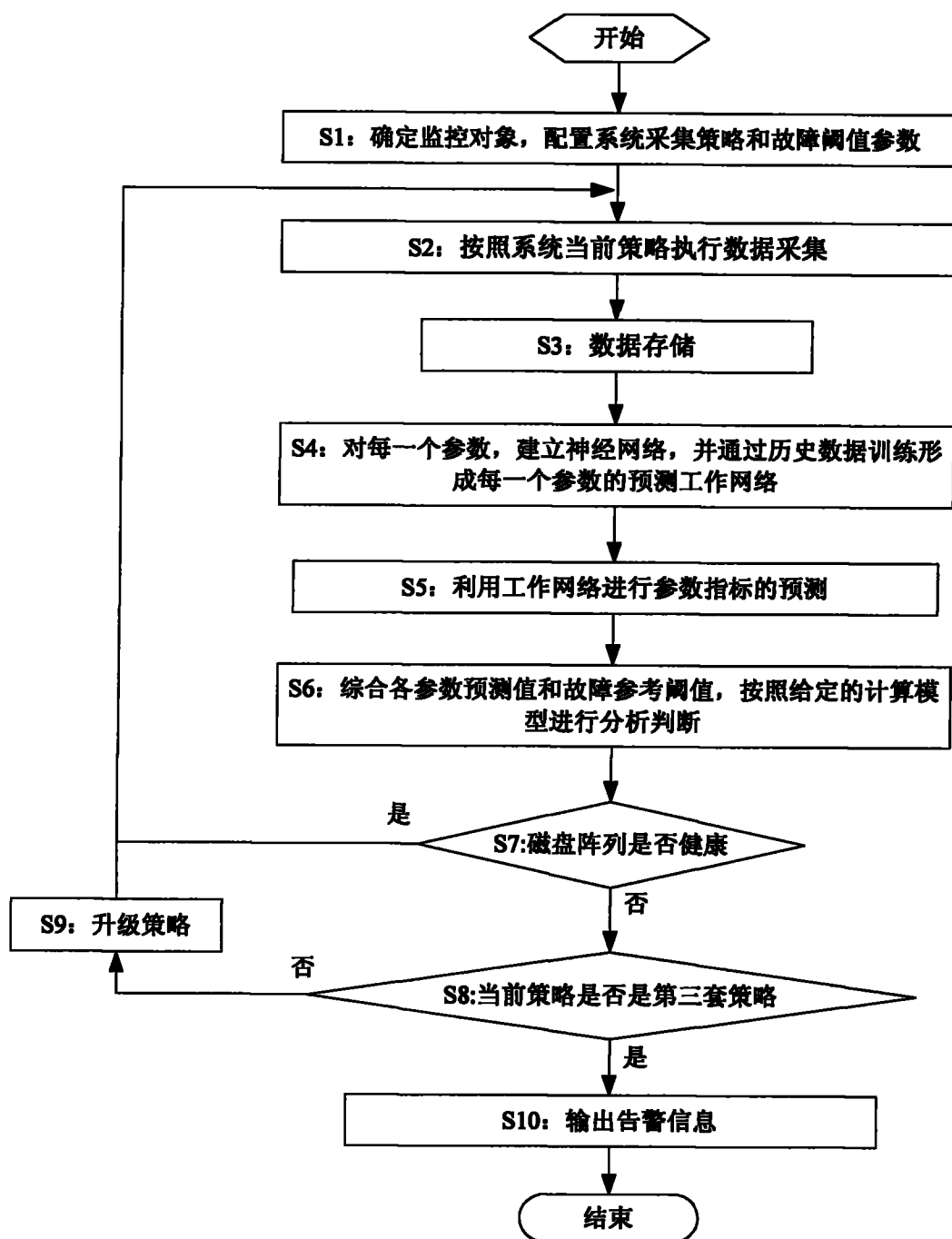


图 3