

文章编号:1672-6197(2014)06-0029-06

一种基于神经网络的网络设备故障预测系统

侯晓凯¹, 李师谦¹, 王杰琼², 胡 彬¹, 邓 晶¹

(1. 中国人民解放军 72671 部队, 山东 济南 250022; 2. 山东省中医药大学 药学院, 山东 济南 250355)

摘 要: 引入基于状态的维修技术, 构建了基于多状态的在网运行设备故障预测模型. 根据故障的严重性将预警等级划分为四层, 对于不同的预警级别, 分别构建神经网络. 该模型解决了设备故障预测精度不高的难题, 提升了基于多状态的故障预测能力.

关键词: 神经网络; 多状态; 网络设备; 故障预测

中图分类号: T393

文献标志码: A

DOI:10.13367/j.cnki.sdgc.2014.06.007

Neural network fault prediction system in network equipment

HOU Xiao-kai¹, LI Shi-qian¹, WANG Jie-qiong², HU Bin¹, DENG Jing¹

(1. 72671 Troops, The Chinese People's Liberation Army, Ji'nan 250022, China;

2. School of Medicine, Shandong University of Traditional Chinese Medicine, Ji'nan 250355, China)

Abstract: We introduced the maintenance technology based on condition, and constructed network equipment fault prediction model based on the multi-state. The warning level is divided into four layers according to the seriousness of the fault, and neural network are constructed for different warning levels. This model solves the problem that equipment fault prediction accuracy is not high, and enhances the ability of multi-state fault prediction.

Key words: neural network; multi-state; network equipment; fault prediction

为了克服传统维修方式的不足, 随着状态监测和故障诊断技术的不断进步, 上个世纪末逐渐发展起来一种新的维修方式——基于状态的维修(CBM). 该维修方式综合运用各种技术手段获取设备的运行状态信息, 并运用数据分析与维修决策技术对设备状态进行实时或者周期性的评价, 最终做出科学化的维修决策. 实现了通过状态监测预测即将发生的故障, 制订合理的维修决策.

1 基于状态的维修

基于状态的维修过程示意图如图 1 所示.

基于状态的维修主要根据获得的设备状态信息预测设备(或部件)的剩余寿命, 以一定的优化准则为目标对设备做出维修决策, 即判断设备是否需要

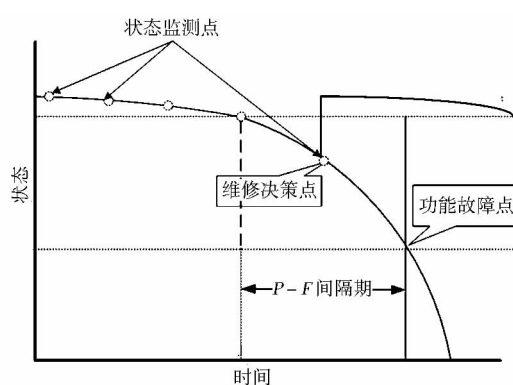


图 1 基于状态维修示意图

进行预防性维修, 若需要, 何时进行维修最合适. 这种维修方式的维修间隔期是不固定的, 其最大的特点是根据每个设备具体的状态, 在设备故障发生前进行维修. 因此, 基于状态的维修能克服传统维修方

收稿日期: 2014-04-28

作者简介: 侯晓凯, 男, hou_xk@126.com

式的不足,实现精确维修.对于设备,基于状态的维修可以降低维修保障费用、提高设备完好率和任务成功率^[1];通过减少备件、保障设备、维修人力等保障资源需求,降低维修保障费用^[2];通过减少维修,特别是计划外维修次数,缩短维修时间,提高战备完好率;通过状态监测,减少任务过程中故障引起的风险,提高任务成功率.

2 故障预测的前提条件

对于一些故障,其性能劣化是一个逐渐劣化的过程,即存在“潜在故障—功能故障”间隔,这个过程也就是 $P-F$ 间隔期,如图 2 所示.“ P ”点是潜在故障点,是能够发现故障的点,在此之前,故障是没有任何征兆的,当经历“ P ”点之后,如果还不采取维修措施,它通常会以较快的速度退化到功能故障点“ F ”,即设备已不可用.“ $P-F$ 间隔期”的存在是复杂设备开展 CBM 的前提条件,即 CBM 仅适用于存在“ $P-F$ 间隔期”的故障.

网络在网运行设备的性能也是一个逐渐劣化的过程,在经过一个时间点(潜在故障点“ P ”)后会产生告警信息和日志信息,性能开始劣化,如不及时进行维修,很快会导致设备故障(功能故障点“ F ”),因此,通过建立故障预测标准准确判断设备的“ P ”点,跟进设备的运行态势,在“ F ”点之前进行预防性的维修是很有必要的.

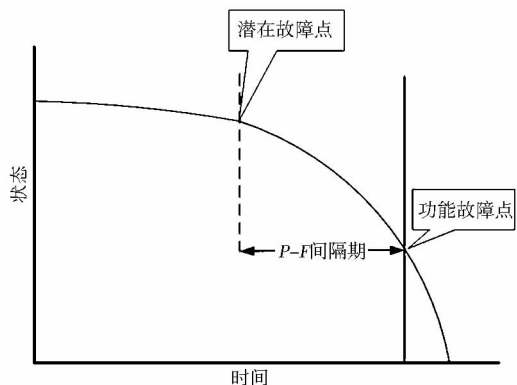


图 2 $P-F$ 间隔期

3 基于状态的故障预测模型

基于状态的故障预测模型分为设备的状态信息、提取故障特征、故障分级和故障预测 4 个模块,如图 3 所示.其中设备的状态信息为后续模块提供

了基本信息,通过对其分析提取出与设备运行状态紧密相关的故障特征,故障分级根据故障的严重性划分为蓝、黄、橙、红 4 个级别,这样,通过判断设备运行状态处于何种级别,一方面可以更直观的显示设备的运行状态,比较清楚的展现设备的劣化趋势;另一方面也可以针对不同的级别分别进行故障预测,使得故障预测更准确.故障预测就是通过数学建模,预测出即将发生的故障特征,从而判断设备运行状态的未来走势.

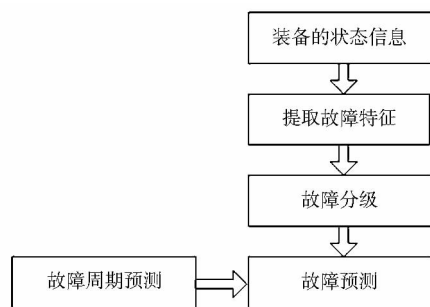


图 3 基于状态的故障预测模型

4 故障特征量的提取

对设备进行基于状态的故障预测,其实质是利用预测模型对特征量进行预测,再通过预测的特征量的值来预测未来一段时间设备的健康状态,所以故障特征量的选择显得尤为重要,要选择对设备运行状态影响较大的特征量,而对于在线运行的计算机网络设备来说,可以通过 SNMP 和 SYSLOG 获取其大量运行信息,比如:生产厂商、路由表、设备的运行时间、板卡的温度、CPU 利用率、内存利用率、SYSLOG 日志、端口流量、端口丢包率等,通常认为设备的运行时间、板卡的温度、CPU 利用率、内存利用率和 SYSLOG 日志信息对设备的运行趋势影响较大,因此,选择它们作为预测模型的特征量,并且从概率的角度出发,做出以下 6 点合理假设:

- (1) 设备的运行时间越趋近或超过故障周期,认为发生故障的可能性越大.
- (2) 板卡的温度超过标准温度(夏季 $23 \pm 3^\circ\text{C}$, 冬季 $20 \pm 3^\circ\text{C}$)越高,认为发生故障的可能性越大.
- (3) CPU 利用率越高,认为发生故障的可能性越大.
- (4) 内存利用率越高,认为发生故障的可能性越大.

(5) SYSLOG 级别越低,认为发生故障的可能性越大;从前面的介绍可知:对应于 0 至 7 个级别,0 级别说明是最严重的消息,设备即将不可用;1 级别说明要立即关注系统的状态,发生故障的可能性次之,依次类推,6、7 级别发生故障的可能性非常低,只是一些报告性、调试性信息,可以通过 SYSLOG 采集模块过滤掉。

(6) SYSLOG 日志信息的刷新速度,认为刷新速度越快,出现故障的可能性越大。

5 故障分级

一方面为更直观的显示设备的运行状态、运行趋势,另一方面可以针对不同的级别,分别分析故障特征量对预测结果的影响,提高预测的准确性。参照突发气象灾害用 4 种颜色即“蓝、黄、橙、红”来表示其严重性,将设备故障的级别也用这四种颜色表示。那么,从设备运行状态来说,“蓝色”预示着设备已经历了潜在故障点,进入“黄色”反映出故障的严重性上升了,“橙色”较“黄色”进一步上升,“红色”则预示着即将到达功能故障点;从趋势上来说,从“蓝色”到“红色”反映出故障的严重性在逐渐上升。反映到 P - F 间隔图上如图 4 所示。

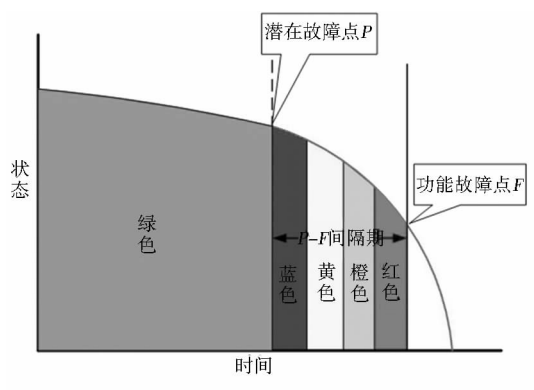


图 4 故障分级

而故障分级则是根据设备的故障特征的阈值进行划分的,如果故障特征的指标达到“蓝”、“黄”、“橙”、“红”级别,则判断设备的状态进入了相应的级别,且故障特征的阈值是可以修正的。这里,通过对以往获得的大量历史样本进行统计分析,认为“蓝”、“黄”、“橙”、“红”4 个级别对应的阈值见表 1 至表 4。

表 1 蓝色预警的阈值

类 别	标 准
SYSLOG 级别	3 级
SYSLOG 刷新速度/条·h ⁻¹	5<每小时的刷新速度<10
板卡温度/℃	40<板卡温度<50
CPU 利用率/%	70<CPU 利用率<75
内存利用率/%	70<内存利用率<75
设备的运行时间	故障维修周期前 1 年
备注	故障周期为通过预测获得

表 2 黄色预警的阈值

类 别	标 准
SYSLOG	4 级
SYSLOG 刷新速度/条·h ⁻¹	10<刷新速度<15
板卡温度/℃	50<板卡温度<60
CPU 利用率/%	75<CPU 利用率<80
内存利用率/%	75<内存利用率<80
设备的运行时间	故障维修周期前 6 个月
备注	故障维修周期为通过预测获得

表 3 橙色预警的阈值

类 别	标 准
SYSLOG 级别	5 级
SYSLOG 刷新速度/条·h ⁻¹	30<刷新速度<40
板卡温度/℃	60<板卡温度<70
CPU 利用率/%	80CPU 利用率<85
内存利用率/%	80<内存利用率<85
设备的运行时间	故障维修周期前 3 个月
备注	故障维修周期为通过预测获得

表 4 红色预警的阈值

类 别	标 准
SYSLOG 级别	6 级
SYSLOG 刷新速度/条·h ⁻¹	40<刷新速度
板卡温度/℃	70<板卡温度
CPU 利用率/%	85<CPU 利用率
内存利用率/%	85<内存利用率
设备的运行时间	故障维修周期前 1 个月或已超过维修周期
备注	故障维修周期为通过预测获得

6 故障预测

故障预测就是通过对历史和当前的故障特征值进行分析,预测出未来的故障特征值,判断设备的预警级别.经过实践证明,当故障发生时,以上 6 种故障特征中 SYSLOG 级别和 SYSLOG 刷新速度并不会因时间的推移而发生变化,可以直接用来判断设备所处的预警级别.而板卡温度、CPU 利用率和内存利用率之间关系密切,比如 CPU 和内存的高利用率势必带来板卡温度的升高,即未来设备的板卡温度、CPU 利用率和内存利用率与当前及历史上设备的板卡温度、CPU 利用率和内存利用率之间存在一种非线性的关系,而 BP 网络恰能以任意的精度逼近任何非线性函数,且 MATLAB 神经网络工具箱提供了对 BP 神经网络的仿真.因此采用 BP 神经网络理论来进行故障预测的.

6.1 反向传播神经网络结构设计

6.1.1 网络隐层节点数目选择

在设计 BP 神经网络时,一般先考虑设计一个隐层,当一个隐层的隐节点数仍不能改善网络性能时,才考虑再增加一个隐层^[3].这里,我们选择单隐层 BP 网络进行设计仿真.而设置多少个隐节点则通过试凑法来确定,确定隐节点数的经验公式^[4]:

$$m = \sqrt{n+l} + \alpha \quad (1)$$

式中: m 为隐层节点数, n 为输入层节点数, l 为输出层节点数, α 为 1~10 之间的常数.

6.1.2 传递函数的选择

BP 属于多层网络,其神经元常用的传递函数包括 log-sigmoid 型函数 logsig,tan-sigmoid 函数 tan-sig,以及线形函数 purelin.

需要指出,sigmoid 型传递函数的曲线形状是 S 型的^[5],log-sigmoid、tan-sigmoid 型函数都是如此.如果 BP 网络的输出层采用 sigmoid 型函数,输出就被限制在 $[-1, 1]$ 范围之内,而如果采用线性函数 purelin 作为输出层的传递函数,那么输出就可以取任何值.由于网络的输出向量值已经被限制在 $[-1, 1]$ 中,因此输出层神经元传递函数我们选用 log-sigmoid 函数.

而中间层神经元传递函数的选择,我们依据 BP 网络设计的一般原则设定为 S 型正切函数 tan-sigmoid.

6.1.3 神经网络的训练

网络设计完成后,要应用历史样本对其进行训

练.训练时对所有样本正向进行一轮并反向修改权值一次称为一次训练.在训练过程中要反复使用样本集数据.

6.2 训练样本的确定及分级

6.2.1 训练样本的确定

BP 神经网络首先要利用历史训练样本对网络进行训练,该过程是 BP 神经网络预测的基础.

当前校园网网络历史训练样本的主要来源为长期在网运行的 H3C 智能网管系统所采集的大量数据.经统计整理,在 2010 年至 2013 年 10 月份之间,某校园网路由交换设备共发生 103 次设备故障,其中设备或(板卡)出现温度过高、CPU 和内存利用率较高征兆的占 77 次,比例达到 76.9%.其他导致故障的主要原因为:误操作、电源系统故障等,而这些故障信息大多可以通过 SYSLOG 获取.可见,温度、CPU 利用率和内存利用率是预测故障的一个重要切入点,而 SYSLOG 信息是一个极为重要的补充.

表 5 为通过 H3C 智能网管系统获取的 2011 年某校园网路由器 SIP 501 母板故障前一段时间内的运行状态信息,该板卡 4 月 11 日突然无法工作,导致对上阻断.该表摘取了 3 月 13 日达到“蓝色”告警之后的数据,也就是说通过本系统在 3 月 13 日就可以判断设备进入潜在故障点,并可预测出未来板件的运行趋势,从而采取一定措施避免阻断的发生.表 5 中, X_1 和 Y_1 、 X_2 和 Y_2 、 X_3 和 Y_3 分别代表上一天和下一天板卡的温度值(单位:℃)、CPU 利用率(单位:%)和内存利用率(单位:%).

表 5 2011 年某路由器故障前设备运行状态信息

日期	输入			输出			级别
	X_1	X_2	X_3	Y_1	Y_2	Y_3	
3 月 13 日	40	60	58	40	60	58	蓝
3 月 14 日	40	60	68	40	61	58	蓝
3 月 15 日	40	61	58	41	61	59	蓝
3 月 16 日	41	62	59	42	63	60	蓝
3 月 17 日	42	63	60	42	63	60	蓝
3 月 18 日	42	63	60	43	64	61	蓝
3 月 19 日	43	64	61	43	65	62	蓝
3 月 20 日	43	65	62	44	66	62	蓝
3 月 21 日	44	66	62	44	68	63	蓝
3 月 22 日	44	68	63	45	69	64	蓝
3 月 23 日	45	69	64	45	70	65	蓝
3 月 24 日	45	70	65	47	71	67	蓝

续表

日期	输入			输出			级别
	X_1	X_2	X_3	Y_1	Y_2	Y_3	
3月25日	47	71	67	48	71	67	蓝
3月26日	48	72	67	49	73	69	蓝
3月27日	49	73	69	50	75	69	蓝
3月28日	50	75	69	51	75	70	黄
3月29日	51	75	70	52	75	70	黄
3月30日	52	75	70	53	75	70	黄
3月31日	53	75	70	54	76	70	黄
4月1日	54	76	70	55	76	70	黄
4月2日	55	76	70	56	77	70	黄
4月3日	56	77	70	57	78	72	黄
4月3日	57	78	72	57	78	72	黄
4月3日	57	78	72	58	78	72	黄
4月4日	58	78	72	59	78	72	黄
4月5日	59	78	72	60	80	75	黄
4月6日	60	80	75	62	82	76	橙
4月7日	62	82	76	64	83	76	橙
4月8日	64	83	76	65	84	77	橙
4月9日	65	84	77	70	85	80	橙
4月10日	70	85	80	75	90	90	红
4月11日	75	90	90	75	90	90	红

6.2.2 训练样本的分级

获得的训练样本按照“蓝、黄、橙、红”4个级别进行划分,对不同的级别分别用神经网络进行预测.

在 MATLAB^[6] 中用 X_1 、 X_2 、 X_3 代表输入序列, Y_1 、 Y_2 、 Y_3 代表输出序列,按照蓝、黄、橙、红4个级别区分如下:

“蓝色”级别对应的输入输出样本为

$$X_1=[40,40,40,41,42,42,43,43,44,44,45,45,47,48,49]$$

$$X_2=[60,60,61,62,63,63,64,65,66,68,69,70,71,72,73]$$

$$X_3=[58,58,58,59,60,60,61,62,62,63,64,65,67,67,69]$$

$$Y_1=[40,40,41,42,42,43,43,44,44,45,45,47,48,49,50]$$

$$Y_2=[60,61,62,63,63,64,65,66,68,69,70,71,72,73,75]$$

$$Y_3=[58,58,59,60,60,61,62,62,63,64,65,67,67,69,69]$$

“黄色”级别对应的输入输出样本为

$$X_1=[50,51,52,53,54,55,56,57,57,58,59]$$

$$X_2=[75,75,75,75,76,76,77,78,78,78,78]$$

$$X_3=[69,70,70,70,70,70,70,72,72,72,72]$$

$$Y_1=[51,52,53,54,55,56,57,57,58,59,60]$$

$$Y_2=[75,75,75,76,76,77,78,78,78,78,80]$$

$$Y_3=[70,70,70,70,70,70,72,72,72,72,75]$$

“橙色”级别对应的输入输出样本为

$$X_1=[60,62,64,65]$$

$$X_2=[80,82,83,84]$$

$$X_3=[75,76,77,77]$$

$$Y_1=[62,64,65,70]$$

$$Y_2=[82,83,84,85]$$

$$Y_3=[76,77,77,80]$$

“红色”级别对应的输入输出样本为

$$X_1=[70,75]$$

$$X_2=[85,90]$$

$$X_3=[80,90]$$

$$Y_1=[75,75]$$

$$Y_2=[90,90]$$

$$Y_3=[90,90]$$

分别对 X_1 、 X_2 、 X_3 、 Y_1 、 Y_2 、 Y_3 进行归一化到 $[-1,1]$ 之间,采用的表达式如下:

$$V=\frac{V-V_{\min}}{V_{\max}-V_{\min}} \quad (2)$$

在 MATLAB 中,用 *premnmx* 命令实现:

$$V=\text{premnmx}(V) \quad (3)$$

6.3 网络的创建与实现

对于不同的预警级别,分别构建神经网络.不同级别的神经网络结构区别在隐含层的节点数目上,根据经验公式:

$$4<\text{隐含层节点数}<14 \quad (4)$$

对于不同的预警级别,遍历所有的隐含层节点数,寻找最优的目标.

for int i=4:1:14

$X_s=[X_1;X_2;X_3]'$;

$Y_s=[Y_1;Y_2;Y_3]'$;

Net=newff([-1 1;-1 1;-1 1],[i,3],{'tansig','logsig'},'trainlm');

Net.trainParam.show=100;

Net.trainParam.epoch=2000;

Net.trainParam.goal=0.0001;

[net,tr]=train(Net,Xs',Ys');

训练过程中,“蓝色”和“黄色”告警级别对应的网络误差的变化情形如图 5,图 6 所示.

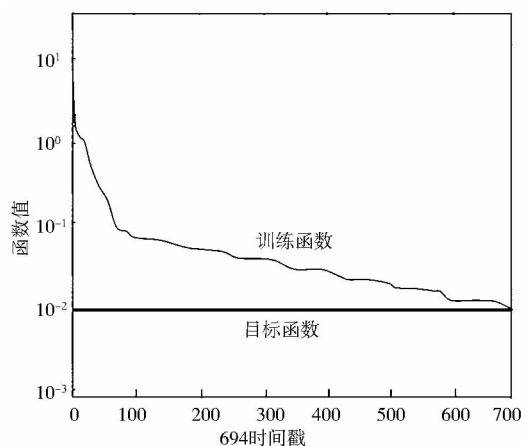


图5 BP 仿真误差曲线(“蓝色”告警级别)

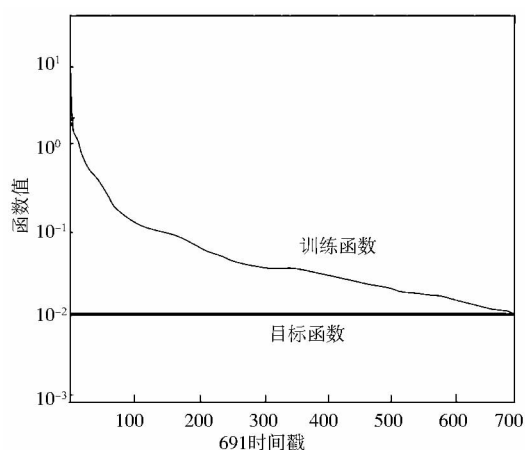


图6 BP 仿真误差曲线(“黄色”告警级别)

本系统投入运行后,2012 年 10 月该路由器再次出现板卡温度过高,CPU 利用率和内存利用率过高的现象,经过该系统预测值与实际值相比基本吻合.如图 7、图 8、图 9 所示:

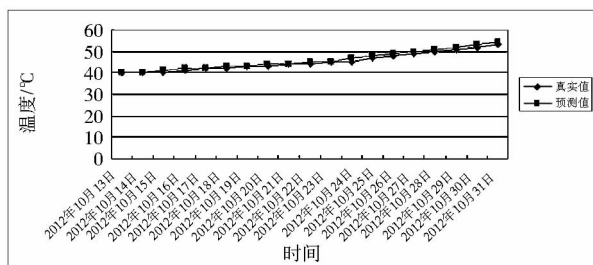


图7 故障预测误差曲线(温度)

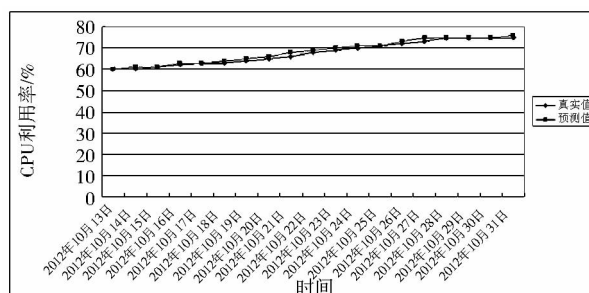


图8 故障预测误差曲线(CPU 利用率)

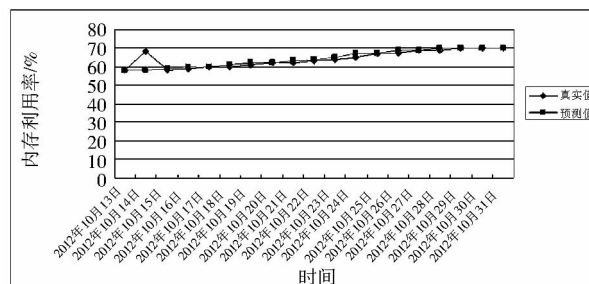


图9 故障预测误差曲线(内存利用率)

通过以上分析可知,当设备处于“蓝色”和“黄色”告警级别时,神经网络计算所得的预测值与实际值基本吻合,当进入“橙色”和“红色”告警级别时,由于设备(或板件)的劣化速度加剧,训练样本较少,已无法正确预测.因此,对于在网运行设备的故障预测最好在“蓝色”和“黄色”告警级别就做好维修工作.

参考文献:

- [1] 李春. 故障预测与健康管理(PHM)技术介绍[J]. 中国高新技术企业, 2008, 15(8): 43-44.
- [2] 刘志伟. 复杂系统故障预测与健康管理(PHM)技术研究[J]. 计算机测量与控制, 2010, 18(12): 2687-2689.
- [3] 徐萍. 基于小波分析和神经网络的 BFI 预测研究[D]. 大连: 大连海事大学, 2005.
- [4] 洪卫国. 人工神经网络在太阳能热水器市场预测中的应用[D]. 上海: 上海交通大学, 2013.
- [5] 孙妍姑. 基于 BP 神经网络的图像识别技术研究[J]. 淮南师范学院学报, 2010, 63(5): 22-23.
- [6] 廖芳瑜. 基于 BP 神经网络分析的在线矿浆品位分析仪设计[J]. 企业科技与发展, 2012(7): 19-20.

(编辑: 姚佳良)