

DOI:10.3969/j.issn.1001-4551.2021.06.016

# 基于动态贝叶斯网络的智能工厂 设备健康评估方法研究\*

高柯柯, 于重重\*, 晏 臻

(北京工商大学 人工智能学院, 北京 100048)

**摘要:**针对传统的机械工厂设备健康评估方法存在设备重组困难、主要故障提取无特点以及设备间的交互不明显问题,提出了一种基于动态贝叶斯网络的设备健康评估方法。利用智能工厂生产线上实时采集的数据,对工厂数据进行了特征提取,以每小时产量作为评估工厂健康程度的指标;利用随机森林算法提取了影响该工厂生产线产量的主要故障类型,构建了动态贝叶斯网络,并进行了设备健康评估。研究表明:动态贝叶斯网络有着出众的学习能力,当输入样本足够多时,设备健康等级的预估由初始的 38.4% 上升到了 64.1%,先验概率会越来越趋近于真实分布,更有利于分析设备间的交互关系及生产线设备的健康状况;该方法可以对设备进行有效的评估,达到有效维护设备的目的。

**关键词:**随机森林算法;动态贝叶斯网络;健康评估方法;故障类型

中图分类号:TH186;TP181

文献标识码:A

文章编号:1001-4551(2021)06-0768-06

## Health assessment for equipment in intelligent factory based on the dynamic Bayesian

GAO Ke-ke, YU Chong-chong, YAN Zhen

(College of artificial intelligence, Beijing Technology & Business University,  
Beijing 100048, China)

**Abstract:** In order to address the problem that there were difficulties in equipment reorganization, featureless extraction of major faults and insignificant interaction between devices in the health assessment methods of traditional equipments in factories, a dynamic Bayesian network-based equipment health assessment method was proposed. Using the real-time data collected from the production line of the intelligent factory, the factory data was characteristically extracted and the hourly output was used as an indicator to assess the health condition of the factory; and the hourly output was used as an index to evaluate the health of the factory; the random forest algorithm was used to extract the main fault types that affected the output of the factory's production line, and a dynamic Bayesian network was constructed for equipment health assessment. The results demonstrate that dynamic Bayesian networks have a superior ability to learn, and that when the input sample is large enough, the health level of equipment is estimated to rise from 38.4% to 64.1%, the prior probability gets closer to the true distribution, which is more conducive to analyzing the interactions between equipment and the health of the production line equipment. It can provide an effective way to evaluate the equipment and achieve effective maintenance of the equipment.

**Key words:** random forest algorithm; dynamic Bayesian network (DBN); health assessment; fault types

收稿日期:2020-10-18

基金项目:北京市自然科学基金资助项目(4202015)

作者简介:高柯柯(1995-),女,甘肃天水人,硕士研究生,主要从事时间序列数据分析方面的研究。E-mail:18811356398@163.com

通信联系人:于重重,女,博士,教授。E-mail:yucc@btbu.edu.cn

## 0 引言

随着工业 4.0 和智能制造技术的兴起,智能工厂的建设逐渐被各国企业所重视<sup>[1,2]</sup>。在这种趋势下,工业设备日益向复杂化和智能化的方向发展,设备的安全问题越来越受到企业的重视。同时,由于工业设备系统运行工况繁杂多变、运行环境相对恶劣,更增加了系统性能退化失效的速率和发生故障的概率。更为重要的是,工厂生产线上的设备一旦由于失效引发故障,将会对工厂经济造成重大的损失。因此,在大数据背景的时代下,利用工厂实时采集的海量数据,在设备性能退化的早期,及时且有效地对设备进行健康状态评估,合理制定维护计划和维修方案变得极为重要。

健康评估是一种通过对系统的状态信息进行收集和学习,分析健康状态,洞悉系统故障,提取规避风险,从而保证系统运行可靠性的一项技术。GASPERIN M 等<sup>[3]</sup>从包络谱中提取出了齿轮啮合频率的功率密度,并且据此对齿轮的健康状态进行了评估。JAVED K 等<sup>[4]</sup>利用模糊聚类的方法,对系统的健康状态进行了自动评估。YAN J 等<sup>[5]</sup>利用 logistic 回归函数,对电梯门运动系统的健康状态进行了评估。WANG C 等<sup>[6]</sup>提出了一种结合有效特征提取方法和模糊 c 均值聚类算法的深度森林算法,实现了对轨道航天器当前健康状态的评估。余道明等<sup>[7]</sup>提出了一种集新型深度自编码和最小量化误差相结合的,应用于滚动轴承全寿命周期的健康评估方法。王村松等<sup>[8]</sup>采用一种多变量深度森林分类器,实现了设备健康状态的离线训练和在线评估。林涛等<sup>[9]</sup>采用改进栈式自编码器的方法,实现了对风电机组发电机的健康评估。

尽管以上的建模评估方法已经取得了较好的理论成果,但在实际应用的场景中,由于场景设置的复杂性不同,设备运行状况也受到许多其他因素的影响。考虑到运行环境的复杂性,前述的建模方式难度倍增,这就使得分析不同工况下设备的运行状态,找出主要影响设备运行的因素,进而对设备进行健康评估变得尤为重要。

本文主要对某压缩机工厂的设备进行健康评估。由于该工厂生产线上各设备协同运行共同完成生产,中间任何一个环节发生故障,都会影响后续的生产,进而影响整个流程的产出;且同一生产线上各设备之间存在着关联性,各设备的运行状况实时都在发生变化。

动态贝叶斯网络可以实现实时故障诊断,随着毗邻时间步骤将不同变量联系起来,本文采用动态贝叶斯网络对设备的健康状态进行评估。

## 1 基于随机森林算法的主要故障提取

本文的实验数据来源于某压缩机工厂的相关数据,数据来自该工厂 2017 年 9 月 3 日到 2018 年 3 月 30 日共半年的数据,主要包括了监测数据、故障数据和产量数据。

监测数据主要是指生产过程中设备上传感器采集上传的实时数据,不同的设备有着不同的状态参数,主要包括电压、电流、转速、焊接时间等,采集间隔固定,基本都在 5 min 左右进行一次采集。

监测数据如表 1 所示。

表 1 监测数据

参数	单位
电压	V
电流	A
初期速度	mm/min
收弧速度	mm/min
焊接时间	s
搭接速度	s
升降动作时间	s

故障数据是指再生产过程中监测出的设备故障类型。数据主要包括故障码以及故障发生时间,故障码为 0 的字段表示正常,非 0 字段表示故障,不同的故障码对应着相应的故障类型。故障数据实时监控采集,一旦出现故障,设备会上报故障码到工业系统。当故障完成修复,则会上传故障码 0,表示设备已回复正常工作状态。

不同设备有着不同的故障类型,故障数据如表 2 所示。

表 2 故障数据

故障设备	故障码	日期
设备 1	108	2017/9/3 13:45:14
设备 1	0	2017/9/3 13:50:54
设备 5	128	2017/9/3 13:52:49
设备 5	0	2017/9/3 13:55:24

产量数据指生产线最后的成品产出数量。产量数据再生产过程中实时采集,每当有 1 件成品产出,累计产数加 1,同时记录此刻的时间。

产量数据如表 3 所示。

表 3 产量数据

累计产量	时间
212 912	2017/9/2 12:05:13
212 913	2017/9/2 12:05:22
212 914	2017/9/2 12:05:39

### 1.1 特征构造

该工厂所采集的数据,除了监测数据有着固定的采集频率,故障数据和产量数据均没有固定的采集频率。(1)产量数据。产量数据是每当新增一个产出,上传一次数据,每次生产一个产出所耗用的时间称之为节拍。节拍数据在工况正常的时候通常维持在 9 s 左右;(2)故障数据。故障数据则是每当发现一个故障或者异常的时候会上传相应的故障码,因此,故障数据的采集没有一个较为规律的周期。结合上述情况,需要对数据进行划分处理。

由于产量数据、故障数据都是实时采集的,时间精

表 4 特征表

日期	产量	设备 1_108	设备 1_128	设备 1_190	设备 1_192	设备 1_电压	设备 1_转速
2017090300	352	0	0	78	0	normal	normal
2017090301	371	0	0	247	0	normal	normal
2017090302	338	3 145	0	0	0	normal	normal
2017090303	349	3 600	0	0	0	normal	normal
2017090304	349	3 600	0	0	0	normal	normal

### 1.2 相关性分析

在动态贝叶斯网络中,需要将特征转化为节点构造出图模型。而上述得到的特征表维度较多,且特征表具有明显的稀疏性。为了减少动态贝叶斯网络的节点个数,减少计算消耗,提高动态贝叶斯网络实时评估的效率,同时剔除由于异常或报警而上传的故障类型,本文通过分析不同故障与产量之间的相关性,找出影响产量的主要故障。

首先笔者统计出频率发生较高的故障类型及其故障的修复时长,对重点故障进行初步的筛选。同时,由于设备的工作情况直接或者间接影响到生产线的产量,不同故障的故障时间对产量有着不同程度的影响,可以采用随机森林算法,将故障对产量的影响程度进行排序,从而提取出重点故障类别来。

笔者对产品的产量进行统计分析,将每小时的产量进行分为 A、B、C、D、E、F 6 个等级。其中,每小时产量 >400 为 A 等,每小时产量  $\in (350,400]$  为 B 等,产量  $\in (300,350]$  为 C 等,产量  $\in (250,300]$  为 D 等,产量  $\in (200,250]$  为 E 等,每小时产量 <200 为 F 等。

最后笔者得到产量的分布图,如图 1 所示。

### 1.3 随机森林提取主要故障

随机森林是决策树的延伸方法之一,是一种由多颗决策树构成的集成学习算法<sup>[10]</sup>。由于决策树存在易发生过拟合的现象<sup>[11]</sup>,为了改善该缺点,随机森林的预测通过组合多个弱分类器,最终结果通过投票或取均值,使得整体模型的结果具有较高的精确度和泛

确到秒,同时采集间隔又是不固定的,为了更加直观地发现故障和产量之间的联系,本文对故障数据和产量数据按小时进行切分。监测数据基本有着固定的采样间隔,采样频率为 5 min。故本文选择将监测数据进行二值处理,即分为正常和不正常两种类型。

结合上述特征,笔者对某机械工厂生产线本体 A 设备组上的各个设备在每个小时段上进行统计,可以构造出特征表。其中,设备 1\_108 表示设备 1 发生故障 108 的时长,以此类推。该表一共包括日期、产量、故障时长(以 s 为单位)、监测参数状态,共计 122 维。

部分特征表的数据截图如表 4 所示。

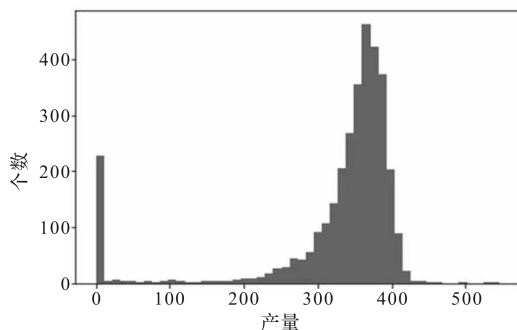


图 1 每小时产量分布

化性能<sup>[12]</sup>。

笔者将每小时不同故障的发生次数和故障恢复时间作为特征输入到随机森林,产量等级作为分类标签;对模型的分类结果的准确率进行排序,排名靠前的特征即为影响产量的重要特征。

随机森林算法模型结构图如图 2 所示。

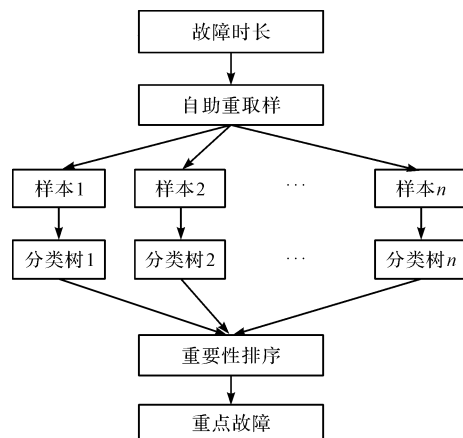


图 2 随机森林模型

笔者最终选定排序在前 10 的故障作为影响产量的主要故障,依次是设备:1(108,128,190),设备 2(103,106,113,114),设备 3(121),设备 4(134),设备 5(126);括号内为设备的故障码。

具体故障及其解释如表 5 所示。

表 5 主要故障及其说明

设备	故障码	故障说明
设备 1	108	1ST 安全光栅
设备 1	128	90°翻转检测工件异常
设备 1	190	气缸超时
设备 2	103	上料机通讯异常
设备 2	106	入口托盘升降机入料超时
设备 2	113	本体输送线热继保护
设备 2	114	设备非常停止
设备 3	121	清渣台数到
设备 4	134	维修门打开报警
设备 5	126	上料安全门异常

## 2 基于动态贝叶斯的设备健康评估

### 2.1 动态贝叶斯理论

动态贝叶斯网络(DBN)是一种以概率论和图论为基础,将时间维融入到贝叶斯网络(BN),以表示时间关联性,最终形成了能够处理时序信息的动态概率推理模型<sup>[13]</sup>。

和贝叶斯网络不同的是,动态贝叶斯网络在贝叶斯网络的基础上考虑到了时间因素,因此更适合对非静态的时序数据进行分析<sup>[14,15]</sup>。动态贝叶斯网络能够反映出贝叶斯网络随着时间变化的动态特性,从而进一步可发现各变量之间的因果关系;同时,当动态贝叶斯网络搭建完成以后,除了能够给出变量之间的关系,还能够在线输入真实案例对网络模型的条件概率表(conditional probability table, CPT)进行推理和更新。因此,在故障健康评估过程中,分析人员通过动态贝叶斯网络能够直观地、实时地找出系统各环节之间的因果关系,从而有效地制定设备的健康评估方案。

参照贝叶斯网络,动态贝叶斯网络同样采用概率积分来表述其不确定性,并且对于概率积分,最基本的规则就是采用关联实践的概率,即:

$$P(A, B) = P(A/B) \times P(B) \quad (1)$$

式中: $P(A, B)$ —事件  $A$  和  $B$  发生的实际概率; $P(B)$ —事件  $B$  的绝对概率; $P(A/B)$ —给定事件  $B$  发生的情况下事件  $A$  发生的条件概率。

$P(A/B)$ 可以由式(1)变换为下式:

$$P(A/B) = \frac{P(A, B)}{P(B)} \quad (2)$$

而当涉及多源融合信息时,贝叶斯网络的公式如下式所示:

$$P(A_i/B) = \frac{P(B/A_i)P(A_i)}{\sum_j P(B/A_j)P(A_j)} \quad (3)$$

### 2.2 数据预处理

动态贝叶斯网络同贝叶斯网络相同,节点类型通常为有限的离散值,而不同时段故障时长是一个真实采集的实数。考虑到贝叶斯网络的实时计算性能,各节点可取的离散值个数不宜过大的原因,笔者将不同时段故障时长划分为高、低、无 3 档。当故障时长超过了不同时段发生该故障的时间的平均值,则用“high”表示。同理,故障时长低于在不同时段发生该故障的时间的平均值,则用“low”表示,而在该时段没有发生故障则用“none”表示。不同故障中“high”、“low”、“none”的先验概率则取值为训练样本中各档的比值。

笔者将表 4 中提出的特征表进行调整,得到和贝叶斯网络相匹配的样本格式,如表 6 所示。

表 6 样本数据

IDnum	yield	equipme- nt_1_108	equipme- nt_1_128	equipme- nt_1_190	equipme- nt_1_H	equipme- nt_3_S
1	B	none	none	none	healthy	normal
2	C	none	none	none	healthy	normal
3	B	low	low	none	healthy	normal
4	A	none	none	none	healthy	normal
5	C	none	none	low	healthy	normal

表 6 中,“IDnum”为样本的标号,“yield”表示产量等级,共分为 A、B、C、D、E、F 6 个等级。命名规则如下:首个字符表示设备名称,“S”表示状态,“H”表示健康程度。如“equipment\_5\_121”表示设备 5 发生故障 121 的时长,共分为“high”、“low”、“none”3 个等级。“equipment\_5\_H”表示设备 5 的健康情况,分为“healthy”和“unhealthy”两个等级。

同样地,“equipment\_5\_S”表示设备 5 的监控数据状态,分为“normal”和“unnormal”两个等级,其余以此类推。

### 2.3 实验与结果分析

经笔者初步构建的动态贝叶斯网络图,如图 3 所示。

笔者将已经处理好的数据集输入动态贝叶斯网络中,当贝叶斯网络对样本数据进行学习后,可以发现,原先设定的先验概率由于对样本的学习而发生了改变,如图 4 所示。

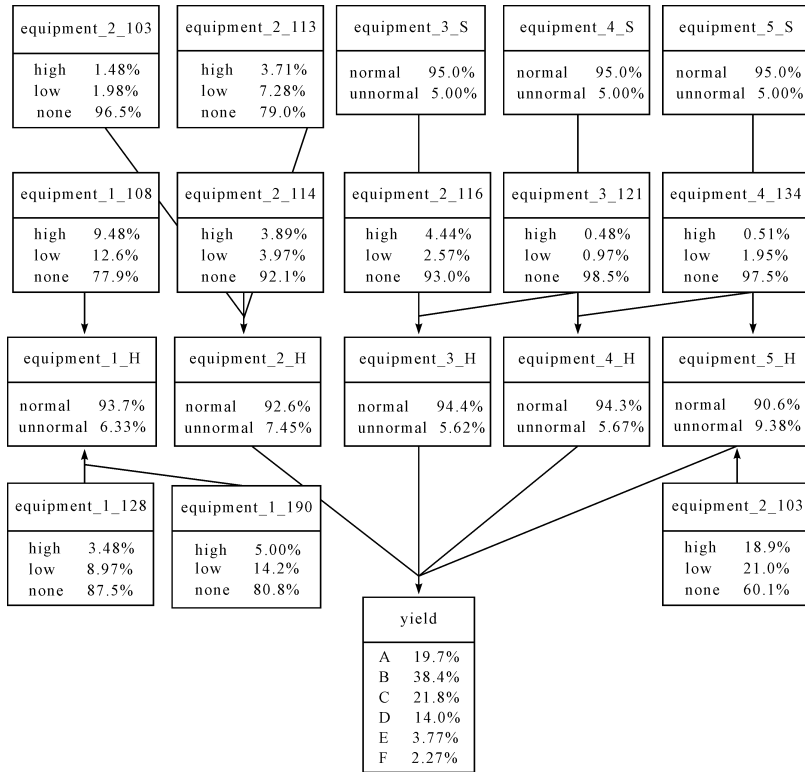


图 3 动态贝叶斯网络结构图

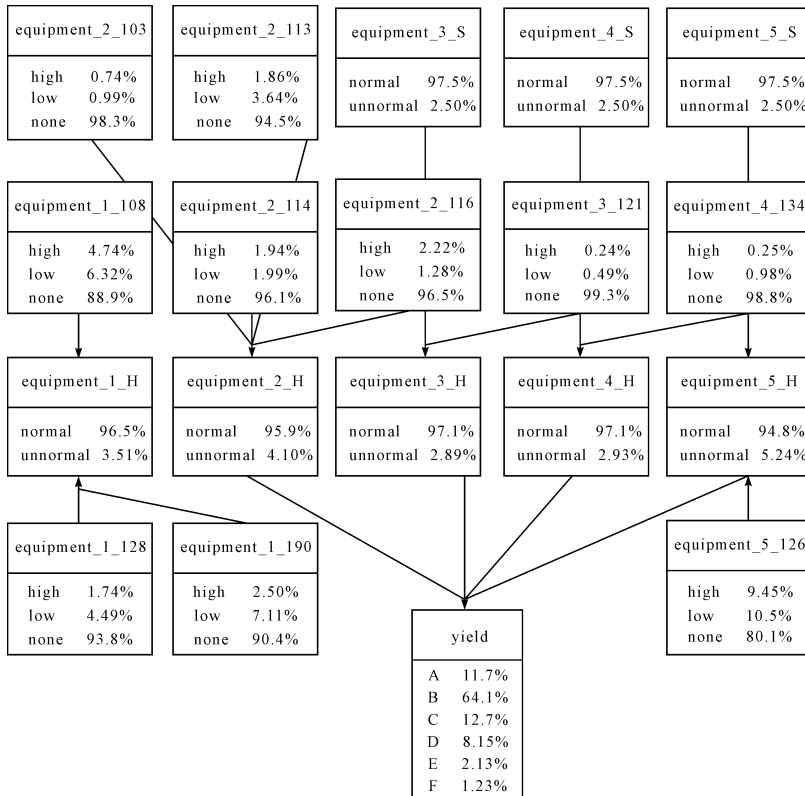


图 4 学习后的动态贝叶斯网络图

由于学习的样本数据中,所有的监控节点均为“normal”,可以看到:学习后的检测数据节点也从图 3 中的 95% 的可能性为“normal”上升到图 4 中的 97.5%;而故障数据也随着输入样本中相应故障数据

为“none”,对“none”置信度有了一定程度的提升。

同时,由以上分析可得知:本文每小时产量数据分布集中在 B 等级,产量节点随着输入数据为 B 等级,对 B 的预估也由初始的 38.4% 上升到了 64.1%。

综合以上分析可以看出:动态贝叶斯网络有着出众的学习能力,当输入的样本足够多,先验概率会越来越趋近于真实分布。

通过实时上传的数据,动态贝叶斯健康评估网络可以实时在线学习。动态贝叶斯网络根据上一时间点的推理结果与当前时间点的证据信息更新节点概率,从而实现实时的健康评估。

笔者对训练好的动态贝叶斯网络进行验证,对未来5个时刻的产量进行了预测,得到的预测结果如表7所示。

表7 动态贝叶斯网络实验结果

时间 点	真实 产值	预测(P[A,B,C,D,E,F])
1	B	[15.8%,28.6%,21.7%,14.5%,12.3%,7.1%]
2	B	[14.2%,32.6%,19.7%,16.5%,8.3%,8.5%]
3	A	[33.7%,27.6%,20.4%,7.5%,5.3%,5.5%]
4	B	[15.8%,31.2%,22.5%,19.7%,8.7%,2.1%]
5	C	[12.8%,21.9%,32.1%,17.4%,10.9%,4.9%]

从表7的实验结果可以看出:在5个时刻中,动态贝叶斯网络评估最大概率对应的等级都与真实的健康等级相同,该结果验证了动态贝叶斯网络在实时健康评估上的有效性。

### 3 结束语

随着工业设备智能化的快速发展,制造企业对设备的可靠性分析不再停留在设备的故障与否,而是更多地考虑到设备未来长期的健康状况。

本文提出了一种基于动态贝叶斯网络的设备健康评估方法,首先利用生产线上实时采集的数据对工厂数据进行了特征提取,再利用随机森林算法提取出了主要故障类型,构建了动态贝叶斯网络,最后进行了设备健康评估。

该方法具有以下特点:

(1)反映出了采集数据与产量之间的隐层关系;

(2)提取出了主要故障类型并对生产线设备进行了健康评估,解决了传统工厂主要设备故障类型提取无特点以及设备间的交互不明显问题;

(3)对设备的健康评估取得了较为理想的效果,验证了动态贝叶斯网络在实际工程应用中的有效性和结果的可解释性。

在后续的研究工作中,需要继续优化动态贝叶斯网络,使得该模型能更广泛地应用于实际工程之中。

### 参考文献(References):

- [1] 焦洪硕,鲁建厦.智能工厂及其关键技术研究现状综述[J].机电工程,2018,35(12):1249-1258.
- [2] 陈国金,姜周曙,苏少辉,等.基于工业4.0的智能工厂实验系统的搭建及应用[J].现代教育技术,2017,27(7):121-126.
- [3] GAPERIN M, JURII D, BOKOSKI P, et al. Model-based prognostics of gear health using stochastic dynamical models[J]. *Mechanical Systems & Signal Processing*, 2011, 25(2):537-548.
- [4] JAVED K, GOURIVEAU R, ZERHOUNI N. A new multi-variate approach for prognostics based on extreme learning machine and fuzzy clustering[J]. *IEEE Transactions on Cybernetics*, 2015, 45(12):2626-2639.
- [5] YAN J, KOC M, LEE J. A prognostic algorithm for machine performance assessment and its application[J]. *Production Planning and Control*, 2004, 15(8):796-801.
- [6] WANG C, LU N, CHENG Y, et al. A telemetry data based diagnostic health monitoring strategy for in-orbit spacecrafts with component degradation[J]. *Advances in Mechanical Engineering*, 2019, 11(4):1-14.
- [7] 余道明,贾民平,张苑.一种新型深度自编码网络的滚动轴承健康评估方法[J].东南大学学报:自然科学版,2018,48(5):801-806.
- [8] 王村松,陆宁云,程月华,等.基于无标签、不均衡、初值不确定数据的设备健康评估方法[J].控制与决策,2020,35(11):2687-2695.
- [9] 林涛,赵成林,刘航鹏,等.基于改进栈式自编码器的风电机组发电机健康评估[J].计算机工程与科学,2020,42(3):517-522.
- [10] 李兵,韩睿,何怡刚,等.改进随机森林算法在电机轴承故障诊断中的应用[J].中国电机工程学报,2020,40(4):1310-1319.
- [11] VIJAYKUMAR V, MICHELLE C, SINA S, et al. Quantifying and characterizing tonic thermal pain across subjects from EEG data using random forest models[J]. *IEEE Transactions on Bio-medical Engineering*, 2017, 64(12):2988-2996.
- [12] 吴潇雨,和敬涵,张沛,等.基于灰色投影改进随机森林算法的电力系统短期负荷预测[J].电力系统自动化,2015,39(12):50-55.
- [13] 李明,洪梅.基于动态贝叶斯网络的海战场自然环境评估与预测[J].军事运筹与系统工程,2018,32(1):37-43.
- [14] 杨静,邹梅,黄微.基于动态贝叶斯网络的网络舆情危机等级预测模型[J].情报科学,2019,37(5):92-97.
- [15] 柯赞.基于动态贝叶斯网络的舆情预测模型研究[J].统计与决策,2016(20):26-28.

[编辑:李辉]

#### 本文引用格式:

高柯柯,于重重,晏臻.基于动态贝叶斯网络的智能工厂设备健康评估方法研究[J].机电工程,2021,38(6):768-773.

GAO Ke-ke, YU Chong-chong, YAN Zhen. Health assessment for equipment in intelligent factory based on the dynamic Bayesian[J]. *Journal of Mechanical & Electrical Engineering*, 2021, 38(6):768-773.

《机电工程》杂志: <http://www.meem.com.cn>