

DOI:10.20033/j.1003-7241.(2026)01-0086-05

基于随机森林算法的IT设备系统运行状态自动监测

李宇新^{1,2}, 李建伟¹, 赵凤坤^{1,2}, 王 璨^{1,2}, 牟 宣^{1,2}

(1. 中石油煤层气有限责任公司, 北京 100028; 2. 中联煤层气国家工程研究中心有限责任公司, 北京 100095)

摘要: 针对IT设备系统的运行中受环境外部噪声影响, 导致监测结果的准确度低的问题。提出基于随机森林算法的IT设备系统运行状态自动监测方法。通过分析IT设备系统运行信号的外部噪声方差, 构建相应的概率密度函数, 结合小波变换处理去除运行信号的外部噪声。利用傅里叶变换的处理解析去噪信号的幅值频谱。构建随机森林算法的决策树结构, 通过算法的自主学习过程得到运行状态的自动监测结果。测试结果表明, 所提方法的PR曲线更接近图像的右上角, 证明所提方法能提高监测结果准确度, 具备更好的应用性能水平, 应用价值更高。

关键词: IT设备; 系统运行状态; 随机森林算法; 自动监测; 运行状态监测; 设备监测

中图分类号: TP242

文献标志码: A

文章编号: 1003-7241(2026)01-0086-05

Automatic monitoring of IT equipment system operation status based on random forest algorithm

LI Yuxin^{1,2}, LI Jianwei¹, ZHAO Fengkun^{1,2}, WANG Can^{1,2}, MOU Xuan^{1,2}

(1. PetroChina Coalbed Methane Company Ltd., Beijing 100028, China;

2. China Union Coalbed Methane National Engineer Research Center Company Ltd., Beijing 100095, China)

Abstract: Addressing the issue of low accuracy in monitoring results due to the influence of external noise on the operation of IT equipment systems, a method for automatic monitoring of IT equipment system operation status based on the random forest algorithm is proposed. By analyzing the external noise variance of IT equipment system operation signals, corresponding probability density functions are constructed. Combined with wavelet transform processing, the external noise in the operation signals is removed. Fourier transform processing is then utilized to analyze the amplitude spectrum of the denoised signals. The decision tree structure of the random forest algorithm is built, and through the algorithm's self-learning process, automatic monitoring results of the operational status are obtained. Test results indicate that the PR curve of the proposed method is closer to the upper right corner of the graph, proving that this method can improve the accuracy of monitoring results and possess a higher level of application performance, making it more valuable for practical applications.

Keywords: IT equipment; system operation status; random forest algorithm; automatic monitoring; operation status monitoring; equipment monitoring

IT设备系统在当今信息化社会中发挥着越来越重要的作用。而相应的运行状态自动监测技术不仅可以提高设备的运行效率, 还有助于确保系统的稳定性和安全性。通过实时监控设备状态, 及时发现并解决问题, 从而保障系统的顺畅运行。此外, 自动监测技术还有助于减少人力投入, 提高运维效率, 降低运维成本。因此, IT设备系统运行状态自动监测具有重要的现实意义和广泛的应用前景。在这样的情况下, 国内外很多专家学者提出了设备运行状态的监测方法。

文献[1]提出了一种通过联合自注意机制和残差网络对消费电子中的智能传感器进行自动监控的方法。利

用残差特征提取和自我注意机制来挖掘传感器故障的显著表征, 并使用长短期记忆网络融合时域上下文信息。智能传感器在复杂环境中运行时, 容易受到温度、湿度、电磁干扰等因素的影响, 从而影响了监测结果的准确度。文献[2]提出了基于IEC61850的设备运行状态监测方法。以IEC61850为标准, 对运行数据进行分层处理, 构建设备监测的对象模型。解析设备的通信接口及通信协议, 实现设备运行状态的监测。但IEC61850标准主要通过通信服务接口来定义模型架构, 可能需要根据具体需求进行模型扩展, 增加了监测过程实现的复杂性。文献[3]提出了基于宽频振动传感器的设备运行状态监测方法。利用宽频振

收稿日期: 2024-06-21; 录用日期: 2024-07-08

基金项目: 国家自然科学基金项目(61202494)

作者简介: 李宇新(1983—), 男, 硕士, 高级工程师, 研究方向: 云计算与网络工程。

引用本文: 李宇新, 李建伟, 赵凤坤, 等. 基于随机森林算法的IT设备系统运行状态自动监测[J]. 自动化技术与应用, 2026, 45(1): 86-90. (LI Yuxin, LI Jianwei, ZHAO Fengkun, et al. Automatic monitoring of IT equipment system operation status based on random forest algorithm[J]. Techniques of Automation and Applications, 2026, 45(1): 86-90.)

动传感器采集设备运行的振动信号,结合设备的局部放电特征,修正信号的超声时延。并提取信号的时频特征,实现设备运行状态的监测。虽然宽频振动传感器可以覆盖较宽的频率范围,能够获取丰富的设备振动信息,但这也使得数据处理的复杂性有所增加,可能导致监测结果的准确度下降。文献[4]提出基于数据挖掘的电气设备状态自动检测方法。通过建立自适应核函数提取状态特征,分割成时间序列,分析测量距离值与序列匹配之间的映射关系,建立自动检测模型,采用蒙特卡洛方法预测设备寿命周期内的故障概率,如果数据预处理步骤不完善,可能会对自动检测模型的准确性和可靠性产生影响。这些方法虽然各具优势,但也面临外部噪声干扰、标准实施复杂性以及数据处理难度等挑战,需进一步研发和优化以提高监测结果的准确性和可靠性。

考虑到上述文献所提出的设备运行状态监测方法在实际应用中均存在监测准确度不高的问题,难以满足 IT 设备在应用中的状态管理监测需求,为了缓解这一问题,本研究结合随机森林算法,提出了一种 IT 设备系统运行状态自动监测方法。该方法通过解析 IT 设备系统的运行信号,通过随机森林算法的自主学习过程,得到准确度较高的自动监测结果。

1 IT 设备系统运行状态自动监测

1.1 IT 设备系统运行信号预处理

本研究通过解析 IT 设备系统在运行过程中所产生的信号来实现相应的运行状态的监测。考虑到 IT 设备系统由于设备之间可能存在复杂的电磁干扰,使得采集的设备运行信号含有大量的外部噪声,本研究对运行信号进行了预处理。

定义设备系统运行信号所服从的高斯分布表达式为^[5]

$$f(x) = \frac{\alpha}{2\sigma x(\alpha^{-1})} e^{-\frac{|x-\mu|}{\sigma\alpha}} \quad (1)$$

式中, x 表示采集的 IT 设备系统运行信号, $f(x)$ 表示该信号的高斯分布, α 表示特征指数, σ 表示尺度参数, e 表示自然常数, μ 表示位置参数。

根据这一定义,分析运行信号的外部噪声方差,计算公式为

$$\eta(x) = f(x) \left(\frac{2n}{N \cdot l} \right)^{\frac{1}{2}} \quad (2)$$

式中, $\eta(x)$ 表示运行信号 x 的外部噪声方差, N 表示信号长度, n 表示信号区段, l 表示信号区段长度, λ 表示信号噪声阈值。

根据运行信号的区段在高斯分布中将其划分为 n 个单高斯分布的形式^[6]。在此基础上,为每个单高斯分布的运行信号添加相应的隐变量,从而计算信号区段的高斯混合分布总方差为

$$\hat{\eta}_n(x) = \sum_{i=1}^N \eta(x) \sigma^2 \quad (3)$$

式中, $\hat{\eta}_n(x)$ 表示运行信号 x 在划分为 n 个区段后的高斯混合分布总方差。

基于式(3)的计算,构建运行信号外部噪声分布的概率密度函数,其表达式为^[7-8]

$$G(x) = \frac{\hat{\eta}_n(x)}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}} \quad (4)$$

式中, $G(x)$ 表示 IT 设备运行信号的外部噪声分布概率密度函数, X 表示添加的隐变量。

将解析出的信号噪声部分与非噪声部分分离,从而实现信号的去噪。本研究引入小波去噪方法完成这一步骤。构建小波去噪的变换函数表达式为^[9]

$$\psi(t) = a^{-\frac{x}{2}} \int \left(\frac{t-ka}{a_0^x} \right)^b dt \quad (5)$$

式中, t 表示信号的采集时间, $\psi(t)$ 表示相应时间下,运行信号的小波变换函数, a 表示平移因子, k 表示小波系数, b 表示尺度因子。

利用上述的小波变换函数将分析的外部噪声与运行信号分离,实现运行信号的去噪过程。

通过上述步骤,完成 IT 设备系统运行信号的预处理步骤。

1.2 解析信号幅值频谱

对去除了外部噪声后的 IT 设备系统运行信号再次进行分析处理。本研究结合傅里叶变换处理,将去噪信号的幅值参数转换为幅值频谱图的形式。

定义去噪后 IT 设备系统运行信号的表达式为

$$Y(t) = \sum_{i=1}^N x(m) W_v^p \quad (6)$$

式中, $Y(t)$ 表示去噪信号, m 表示旋转因子, v 表示信号频域, p 表示信号时域, W 表示 v 为基的指数函数的 p 次幂。

根据 IT 设备系统运行信号的周期性和对称性特点,在式(6)的基础上,为去噪信号添加加权系数,以此表征信号的幅值特征值^[10-12]。其中,信号幅值的表达式为

$$F(Y) = Y(t) (2r-1)^e e^{-\frac{Y^2\pi}{m}} \quad (7)$$

式中, F 表示信号幅值, r 表示加权系数。

根据信号中含有的时域特征与频域特征,将信号幅值参数转换为幅值频谱图的形式。构建这一转换过程中应用的傅里叶变换表达式为

$$p(Y) = \int_{-\infty}^{+\infty} \frac{Y(t)}{F(Y)} e^{-i2\pi mt} \quad (8)$$

经过傅里叶变换的处理,得到设备系统运行信号的时域-频域关系,由此生成相应的幅值频谱图。

通过上述步骤,完成 IT 设备系统运行信号幅值频谱的解析过程。

1.3 运行状态自动分析

为了实现运行状态的自动监测,本研究引入了随机森林算法。随机森林算法是一种集成技术,它利用多棵决策

树的集合来提升预测的精确度并防止过拟合。该算法通过集体决策的方式,即“投票”策略,来增强预测效果。随机森林算法擅长处理高维数据,无需预先选择特征,能够实现特征重要性的有效评估。利用该算法对解析的信号幅值频谱特征进行自主学习,从而实现自动监测过程。

随机森林算法结合多棵决策树,每棵树基于自动采样数据构建,并在每个节点随机选取特征子集以确定最佳分割。构建随机森林算法的决策树结构,如图1所示。

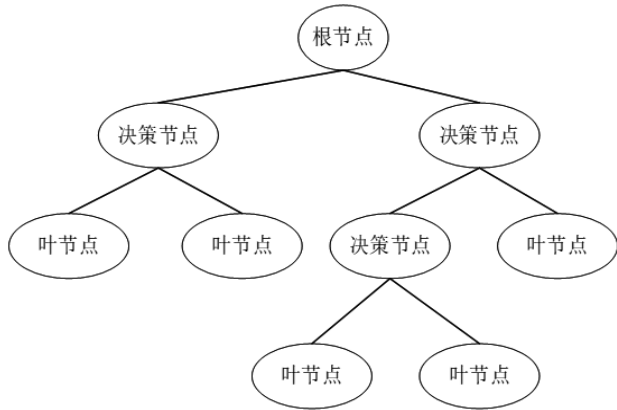


图1 随机森林算法决策树结构

Fig.1 Decision tree structure of random forest algorithm

如图1所示,本研究采用随机森林算法中的 CART 形式的决策树结构。在每个节点上,算法从所有特征中随机选择一部分特征,然后根据基尼值选择最佳分割特征。定义该结构下的基尼指数为^[13]

$$g(D, B) = \sum_{c=1}^c \frac{D^c}{D} s(D^c) \quad (9)$$

式中, g 表示基尼指数, D 表示样本集, B 表示样本特征, c 表示样本取值, s 表示样本基尼值。

在式(9)中,样本基尼值 s 的计算为

$$s(D) = \sum_y^M P_y (1 - P_y) \quad (10)$$

式中, y 表示训练样本, P_y 表示样本占比, M 表示样本类别总数。

根据上述计算定义随机森林算法的决策树结构参数。将解析的 IT 设备系统运行信号的幅值频谱作为训练样本代入到该算法中进行自主学习,这一过程如图2所示^[14-15]。

基于如图2所示的算法学习过程,得到对 IT 设备系统运行信号的学习结果。在此基础上,为算法的学习过程添加状态特征的关联,这一步骤如式(11)所示。

$$S(y) = \operatorname{argmax}_y \sum_{c=1}^{D^c} I(h(x)) \quad (11)$$

式中, S 表示运行状态特征的关联系数, $h(x)$ 表示 IT 设备系统的运行状态特征, I 表示线性函数。

随机森林通过多数投票的方式来决定最终的分类结果。通过算法对状态特征的自主关联,实现对 IT 设备系统运行状态的自动分析,从而得到相应的自动监测结果。

通过上述步骤,完成基于随机森林算法的 IT 设备系统运行状态自动监测方法的设计。

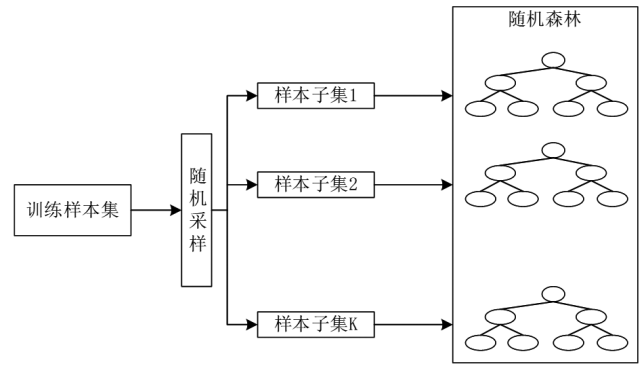


图2 随机森林算法示意图

Fig.2 Schematic diagram of random forest algorithm

2 监测分析

2.1 IT 设备系统

以某地区的数据中心所使用的 IT 设备系统为依托开展 IT 设备系统运行状态监测的实践测试实验。该数据中心的 IT 设备系统包含 20 台主要节点,其连接拓扑结构如图3所示。

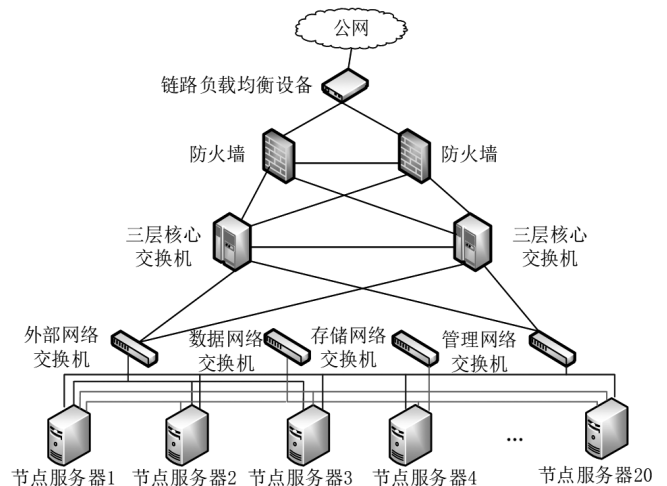


图3 IT 设备系统拓扑结构

Fig.3 IT equipment system topology structure

如图3所示,本次实验所依托的数据中心 IT 设备系统为六层拓扑结构。上述结构中,网络交换机的设备参数如表1所示。

表1 网络交换机设备参数

Tab.1 Network switch equipment parameters

参数	参数设置
RJ-45 端口数量	8 个
端口最大传输速率	1 000 Mb/s
1 000 Mb/s	100 Gb/s
光纤端口	SFP
传输距离	100 km
光口	1 000 M 单模双纤 LC

3 层核心交换机的设备参数如表2所示。

表2 3层核心交换机设备参数

Tab. 2 Parameters of layer 3 core switch equipment

参数	参数设置
风道	左进右出
主控板槽位	2
交换网板槽位	2(主控集成)
业务板槽位	8
交换容量	502.5 Tbp/s
包转发率	155 550 Mpp/s
生成树	MSTP 协议

链路负载均衡设备的主要参数如表3所示。

表3 链路负载均衡设备参数

Tab. 3 Parameters of link load balancing equipment

参数	参数设置
传输速率	300 Mbps
支持协议	IEEE 802.11b/g/n
速度	300 m/s
线网络支持频率	2.4 GHz
串口	TTL+RS485
工作功耗	300 mA@12V DC
主芯片	32位处理器

以上述所示的设备参数为标准数值分析IT设备系统的运行状态。

在该IT设备系统环境下,开展本次运行状态的监测实验。

2.2 运行状态自动监测

在上述的IT设备系统环境下,应用本文所设计的方法对其运行状态进行自动监测,分析其应用可行性。

采集IT设备系统在一段时间内的运行信号,对本文方法中的随机森林算法进行训练,结果如图4所示。

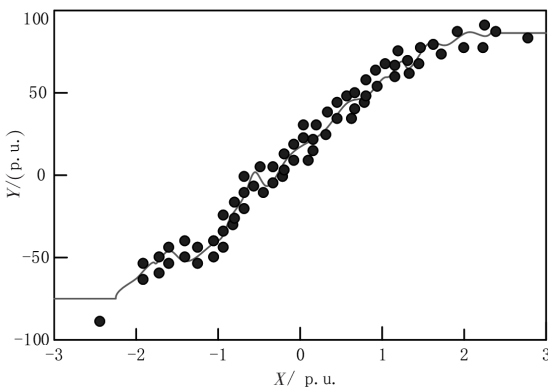


图4 随机森林算法回归训练

Fig. 4 Regression training of random forest algorithm

如图4所示,在本次测试中,随机森林算法对IT设备系统运行信号的学习未出现过拟合现象。根据这一训练结果,设置本文方法中的随机森林算法参数,如表4所示。

完成随机森林算法的训练后,应用本文方法对IT设备系统的运行状态进行监测。

经由本文方法对IT设备系统运行信号进行处理后,得到结果如图5所示。

表4 随机森林算法参数设置

Tab. 4 Parameter settings for random forest algorithm

参数项	参数设置
决策树数量	180
决策树最大深度	50
内部节点分裂最小样本数	2
叶节点最小样本数	15
最大特征数	$\log_2 50$
采样方式	有放回采样
并行处理	5
随机种子值	2
权重	0.05
最大叶节点数	500

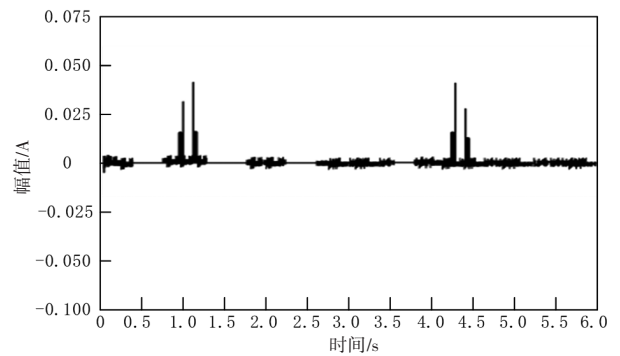


图5 运行信号处理结果

Fig. 5 Signal processing results during operation

如图5所示,经由本文方法处理后,去除了原始信号中的外部噪声干扰。根据去噪处理后的信号,将其分解转换为幅值频谱的形式,结果图像如图6所示。

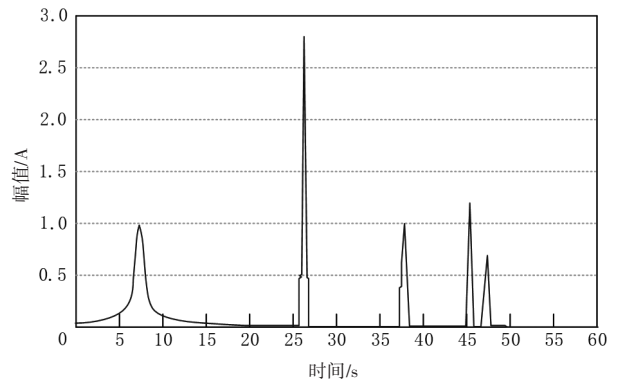


图6 运行信号幅值频谱图

Fig. 6 Frequency spectrum of operating signal amplitude

分析图6可知,在本次监测时段内,IT设备系统运行信号的各个频率分量呈现直流衰减混合叠加的状态。通过随机森林算法对各个频率分量幅值特征进行自动分析,得到相应的IT设备系统运行状态。经分析,本文方法监测该时段内IT设备系统处于启动至稳定运行的过渡状态。与实际的运行状态对比可知,本文方法的监测结果准确,说明其具备实际应用的可行性。

2.3 性能分析

为进一步验证本文方法在实践应用中的监测性能,分别采用文献[2]、文献[3]以及文献[4]所提出的设备运行状态监测方法为对比方法,设计IT设备系统运行状态监

测的对比实验。

经测试,得到不同方法的监测结果如图7所示。通过将各方法的监测结果绘制为PR曲线,得到如图7所示的对比结果。由该结果可知,与其他方法相比,本文方法的PR曲线更接近图像的右上角,这说明本文方法的监测结果能够在保证较高召回率的同时保持较高的准确度。

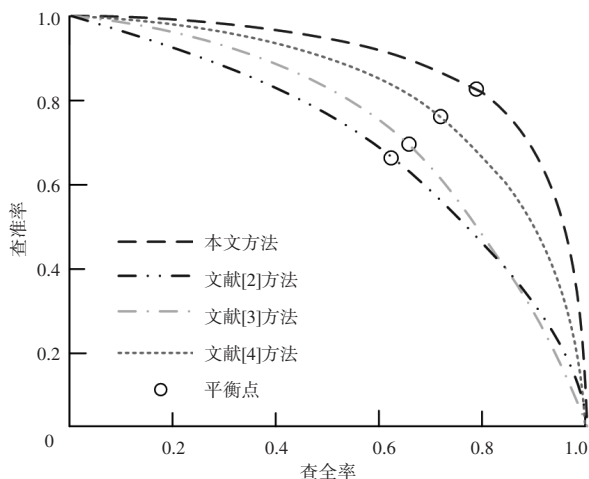


图7 不同方法监测结果对比

Fig. 7 Comparison of monitoring results using different methods

综上所述,本文方法在IT设备系统领域的实际应用中,能够得到较高准确度的运行状态监测结果,具备更好的应用性能水平,应用价值更高。

3 结论

本文面向IT设备系统在实际应用中的需求,提出了基于随机森林算法的IT设备系统运行状态自动监测方法。尽管该方法在实践测试中的许多方面表现出了优异的性能,监测准确度较高,但仍然存在一些不足之处。其中的不足在于该方法在处理高维信号时可能由于维度的增加,直接导致计算复杂度和内存需求的急剧上升,从而

遭遇性能瓶颈。因此,在后续的研究工作中,还需针对IT设备系统运行信号的预处理步骤进行深入研究,确保每个高维信号均能实现降维过程,以保证监测方法对运行信号的全解析,从而进一步提高状态监测的准确度。

参考文献

- [1] ZHONG S, ALI R. Joint self-attention mechanism and residual network for automated monitoring of intelligent sensor in consumer electronics [J]. IEEE Transactions on Consumer Electronics, 2024, 7(1): 1302-1309.
- [2] 钟化兰, 黄扬海. 基于IEC61850的智能变电站设备运行状态数据实时监测方法[J]. 计算机应用与软件, 2024, 41(5): 96-100, 137.
- [3] 相晨萌, 高树国, 刘宏亮, 等. 基于宽频振动传感器的电力设备状态监测方法可行性分析[J]. 高压电器, 2024, 60(4): 83-91.
- [4] 王森. 机械设备电气故障自动检测系统优化设计[J]. 自动化技术与应用, 2024, 43(4): 134-137, 168.
- [5] 和征, 李忠鹏, 杨小红. 基于数字孪生合K-邻算法的车间设备运行状态预测研究[J]. 制造技术与机床, 2024(3): 193-199.
- [6] 张军六, 刘珊, 李瑞. 数字孪生技术下输变电设备运行状态智能感知[J]. 计算机仿真, 2023, 40(12): 123-127, 144.
- [7] 施胜丹, 黄金军, 赵文强, 等. 基于Hankel矩阵与图结构的GIS设备运行状态识别方法[J]. 噪声与振动控制, 2022, 42(5): 159-164, 233.
- [8] 王雪, 唐正聪, 刘兴杰. 电压暂降随机影响下考虑低压脱扣器保护动作机制的电压敏感设备运行状态评估[J]. 电工技术学报, 2022, 37(15): 3794-3804.
- [9] 周鹤, 黄建军. 基于大数据分析的光通信系统关键设备状态识别[J]. 激光杂志, 2023, 44(12): 167-172.
- [10] 牛硕丰, 李涛, 等. GIS开断设备典型机械故障模拟及运行状态智能诊断研究[J]. 高压电器, 2023, 59(12): 36-47.
- [11] 杜海红, 陈黎明, 王冬冬, 等. 基于Mask R-CNN的电力关键设备载波运行状态检测[J]. 电子设计工程, 2024, 32(5): 100-103, 108.
- [12] 杨青, 吴松丽. 基于大数据挖掘的船舶通信系统关键设备状态分析[J]. 舰船科学技术, 2023, 45(5): 136-139.
- [13] 刘雯, 李超. 基于改进神经网络的电压监测仪故障自动诊断方法[J]. 电子设计工程, 2023, 31(10): 167-171.
- [14] 郭春玲. 基于随机森林算法的英语自动翻译设备运行态势监测方法[J]. 自动化与仪器仪表, 2023(1): 178-183.
- [15] 秦潘昊, 陈威宇, 胡秦然, 等. 新型电力系统设备状态监测与故障诊断传感芯片关键技术与展望[J]. 电力系统自动化, 2024, 48(6): 83-95.

(上接第82页)

- [3] 郝素华, 张志勇. 燃煤电厂SO₃脱除、测试技术研究进展[J]. 资源节约与环保, 2022(12): 5-10.
- [4] 韩军赞, 张洁, 闫威卓, 等. 燃煤电厂和钢铁厂烧结烟气有色烟羽排放特征与管控建议[J]. 环境工程, 2024, 42(2): 144-151.
- [5] 张军营, 崔向峰, 王志康, 等. 煤燃烧非常规污染物排放控制[J]. 洁净煤技术, 2023, 29(10): 1-16.
- [6] 杨林, 徐有和, 张伟. 基于数据挖掘的发电厂烟气循环排放自动监测系统[J]. 自动化技术与应用, 2024, 43(3): 57-61.
- [7] 赵西, 温崇荣, 李金金. 控制冷凝法对烟气中三氧化硫冷凝分离效率研究[J]. 石油与天然气化工, 2023, 52(6): 23-29, 38.
- [8] 梁玉兰, 周川, 郑芳, 等. 燃煤烟气治理不同阶段SO₃检测方法的对比研究[J]. 现代化工, 2023, 43(9): 227-230.
- [9] 陈永强. SO₄²⁻分析方法在SO₃测试中的应用研究[J]. 能源与环境, 2023, (4): 125-127.

- [10] 李秉辉. 基于可调谐二极管激光吸收光谱技术的SO₃气体测量参数优化[J]. 应用激光, 2023, 43(12): 133-138.
- [11] 黄思林. 基于中红外激光吸收光谱的CH₄浓度检测仪研究[J]. 电子器件, 2022, 45(4): 1004-1008.
- [12] 李俊, 范斌斌, 曾庆杰, 等. 基于8.309 μm QCL的硫化氢/甲烷开路式检测方法研究[J]. 光学精密工程, 2024, 32(4): 467-477.
- [13] 王婷, 韩志洋, 俞跃, 等. 基于TDLAS技术气体浓度测量的温度修正方法[J]. 激光与红外, 2024, 54(2): 208-213.
- [14] 柳山虎, 吴勇, 梁立振, 等. 基于TDLAS的气体检测技术对比与分析[J]. 计量与测试技术, 2024, 51(4): 35-39.
- [15] LI JD, DING YJ, LI Z, et al. Simultaneous measurements of SO₂ and SO₃ in the heterogeneous conversions of SO₂ using QCL absorption spectroscopy[J]. Applied Physics B, 2022(128): 61.