

基于大数据分析的热工设备运行状态评估模型构建

邵新玉

新疆天富能源股份有限公司天河热电分公司 新疆维吾尔自治区 石河子 832000

【摘要】：在智慧电厂的建设与运行过程中，热工设备作为核心生产设备，其运行状态直接影响电厂的发电效率、能源消耗和安全生产。本文针对智慧电厂热工设备运行状态评估的需求，结合大数据分析技术的优势，深入研究热工设备运行数据的特点与规律。通过采集多源异构的热工设备运行数据，运用数据预处理、特征提取等技术，构建基于大数据分析的热工设备运行状态评估模型。该模型能够实现对热工设备运行状态的准确评估与预测，及时发现潜在故障隐患，为热工设备的优化运行、预防性维护提供科学依据，提高智慧电厂的生产效率与安全性。

【关键词】：大数据分析；热工设备；运行状态评估；模型构建；智慧电厂

DOI:10.12417/2811-0722.26.06.003

引言

传统的热工设备运行状态评估主要依赖人工巡检和简单的仪表监测，存在评估效率低、准确性差、无法及时发现潜在故障等问题。随着物联网、传感器等技术在电厂中的广泛应用，热工设备产生了海量的运行数据，涵盖设备的温度、压力、流量、振动等多维度信息。这些数据蕴含着设备运行状态的关键特征，但传统的数据分析方法难以对其进行有效处理和挖掘。大数据分析技术具有强大的数据处理、分析和预测能力，能够从海量数据中提取有价值的信息，为热工设备运行状态评估提供了新的技术途径。构建基于大数据分析的热工设备运行状态评估模型，对于保障热工设备安全稳定运行、提升智慧电厂的智能化管理水平具有重要意义。

1 热工设备运行状态评估现状与问题分析

1.1 传统评估方法概述

传统的热工设备运行状态评估方法主要包括人工巡检法、定期检测法和简单指标监测法。人工巡检法是由工作人员定期对设备进行现场检查，通过目视、听声、触摸等方式，判断设备是否存在异常；定期检测法按照固定的时间周期，对设备进行全面的性能检测和故障排查；简单指标监测法则是通过监测设备的少数关键指标，如温度、压力等，与设定的阈值进行比较，判断设备运行状态。

1.2 现存问题分析

传统评估方法存在诸多不足。首先，人工巡检法依赖巡检人员的经验和责任心，主观性强，容易出现漏检、误判等情况，且巡检效率低，难以满足大规模热工设备的评估需求；其次，定期检测法检测周期固定，无法及时发现设备在检测周期内出现的突发故障，存在安全隐患；最后，简单指标监测法仅关注少数关键指标，忽略了设备运行数据之间的关联关系，无法全面反映设备的运行状态，对于潜在的复杂故障难以有效识别。

表1 评估方法

评估方法	优点	缺点
------	----	----

人工巡检法	操作简单，可直观发现部分明显问题	主观性强，效率低，易漏检误判
定期检测法	检测全面，能发现潜在问题	检测周期固定，无法及时响应突发故障
简单指标监测法	指标明确，判断方式简单	无法反映数据关联，难以识别复杂故障

2 基于大数据分析的热工设备运行状态评估模型构建

2.1 数据采集与预处理

2.1.1 数据采集

热工设备运行数据呈现多源性与异构性特征，其来源广泛且数据格式多样。在实时监测数据采集方面，借助物联网（IoT）技术，部署各类传感器构建感知网络。例如，在锅炉设备上安装高精度温度传感器，可实现对炉膛温度、蒸汽温度等关键指标每秒1次的高频采集；压力传感器则用于监测汽包压力、管道压力，采集频率同样适配设备运行动态变化需求。这些传感器通过工业现场总线或无线通信技术，将数据实时传输至统一的数据采集平台。

设备运行日志数据蕴含设备运行过程中的关键事件与参数变化信息，通常存储于设备控制系统或专用日志服务器中。通过开发数据接口程序，利用数据库连接技术（如 ODBC、JDBC），可实现对日志数据的定期抽取，例如每小时同步一次日志信息，涵盖设备启停时间、参数调整记录、报警事件等内容。

2.1.2 数据预处理

采集到的原始数据往往存在缺失、噪声和异常值等问题，严重影响后续分析的准确性，因此需进行系统的数据预处理。对于缺失数据，依据数据缺失比例和分布特点选择合适的插补方法。当缺失比例较低（如小于5%）且数据具有平稳性时，采用均值插补法，以该变量的历史均值填充缺失值；若数据存在明显的趋势性，则运用回归插补法，通过建立与其他相关变量的回归模型预测缺失值。例如，在处理某段时间内缺失的蒸

汽流量数据时,可基于蒸汽压力、温度等相关变量构建线性回归模型,计算并填充缺失值。

针对噪声数据,采用中值滤波、高斯滤波等算法进行平滑处理。中值滤波适用于去除脉冲噪声,通过将数据点的取值替换为其邻域内的中值,可有效消除随机干扰;高斯滤波则基于高斯函数对数据进行加权平均,能够在保留数据趋势的同时,抑制高斯噪声。以设备振动信号处理为例,经高斯滤波后,信号中的高频噪声被有效过滤,真实振动特征得以凸显。

异常值的识别采用基于统计学的 3σ 原则和孤立森林算法相结合的方式。 3σ 原则依据数据的均值和标准差,将超出均值 ± 3 倍标准差范围的数据判定为异常;孤立森林算法通过构建二叉树对数据点进行孤立,孤立路径短的数据点被识别为异常值。对于识别出的异常值,若为测量误差导致,则进行修正;若为真实异常情况,则保留并标记,以供后续深入分析。

2.2 特征提取与选择

2.2.1 特征提取

在时域分析中,通过计算数据序列的统计量来获取特征。对于设备振动信号和电流信号,均值反映了信号的平均水平,可用于判断设备运行的基准状态;方差则衡量了数据的离散程度,方差越大,表明信号波动越剧烈,可能暗示设备运行不稳定。例如,当汽轮机轴承振动信号的方差突然增大时,可能预示着轴承磨损加剧。此外,峰值、峭度、偏度等统计特征也具有重要意义。峰值可反映设备在运行过程中受到的瞬时冲击;峭度对信号中的异常冲击敏感,常用于检测设备的早期故障;偏度则描述了数据分布的不对称性,能够辅助判断设备运行的异常趋势。

时频分析结合了时域和频域分析的优点,能够同时展示信号在时间和频率上的变化特征。常见的时频分析方法如短时傅里叶变换(STFT)、小波变换等。短时傅里叶变换通过加窗函数对信号进行分段傅里叶变换,可分析信号的时变频率特性;小波变换则具有良好的局部化分析能力,能够自适应地调整分析尺度,对瞬态信号和非平稳信号具有更好的分析效果。在热工设备故障诊断中,时频分析可用于捕捉故障发生瞬间的信号特征变化。例如,当锅炉燃烧器出现点火故障时,其火焰强度信号通过小波变换后,在特定时间-频率区域会出现异常的能量分布,通过对这些特征的分析,可快速定位故障发生的时间和原因。

2.2.2 特征选择

经过特征提取后,得到的特征集合往往维度较高,包含大量冗余和不相关特征,这些特征不仅会增加模型训练的计算量,还可能干扰模型对关键信息的学习,降低评估准确性。因此,需要运用特征选择方法对特征进行筛选,保留最具代表性的特征子集。

主成分分析(PCA)是一种常用的线性降维方法,其核心思想是通过正交变换将原始特征转换为一组线性无关的新特征,即主成分。这些主成分按照方差从大到小排列,方差越大表示该主成分包含的原始数据信息越多。在热工设备运行状态评估中,PCA可将多个相关性较高的特征(如设备不同位置的温度、压力特征)转换为少数几个综合特征,这些综合特征能够保留原始特征的大部分信息。例如,对于锅炉的多个温度监测点数据,通过PCA处理后,可得到2-3个主成分,这些主成分既减少了数据维度,又能反映锅炉整体的温度分布状态。在实际应用中,通常根据累计方差贡献率来确定保留的主成分数量,一般当累计方差贡献率达到85%-95%时,即可认为保留的主成分能够较好地代表原始数据。

互信息方法基于信息论原理,通过计算特征与设备运行状态标签之间的互信息值来衡量特征的重要性。互信息值越大,表明该特征与设备运行状态的相关性越强,对评估模型的贡献越大。在热工设备特征选择中,可计算每个特征与设备故障类型(正常、轻微故障、严重故障等)之间的互信息,设定阈值去除互信息值较低的特征。例如,在对汽轮机运行状态评估时,计算振动频率、蒸汽流量、润滑油压力等特征与故障类型的互信息,发现某些次要的润滑油参数与故障类型的互信息值极小,可将其从特征集合中剔除。此外,互信息方法还可用于检测特征之间的冗余性,若两个特征之间的互信息值过高,说明它们携带的信息存在大量重复,可保留其中一个特征,进一步优化特征集合。

2.3 评估模型构建

2.3.1 模型选择

根据热工设备运行状态评估的特点和需求,选择合适的评估模型。常用的模型包括支持向量机(SVM)、随机森林、神经网络等。支持向量机适用于小样本、非线性数据的分类和回归问题;随机森林具有较好的抗噪声能力和泛化能力;神经网络能够处理复杂的非线性关系,具有强大的自学习和自适应能力。在实际应用中,可通过对比实验,选择性能最优的模型。

2.3.2 模型训练与优化

(1) 数据划分

首先将预处理后的热工设备运行数据划分为训练集和测试集。常见的数据划分比例为7:3或8:2,即70%-80%的数据用于模型训练,20%-30%的数据用于模型测试。此外,为充分利用数据并更准确地评估模型性能,还可采用K折交叉验证方法。将数据集划分为K个互不相交的子集,每次训练时选择其中K-1个子集作为训练集,剩余1个子集作为测试集,重复K次训练和测试过程,最终将K次测试结果的平均值作为模型性能的评估指标。这种方法能够有效避免因数据划分方式不同导致的评估偏差,更全面地反映模型的泛化

能力。

(2) 模型训练

利用训练集数据对选定的模型进行训练。在训练过程中，不同模型需要调整的参数各不相同。对于支持向量机，关键参数包括核函数类型及其参数（如 RBF 核的 γ 值）、惩罚因子 C 等。 γ 值控制着数据映射到高维空间后的分布情况， γ 值越大，模型对训练数据的拟合能力越强，但也容易导致过拟合； C 值用于平衡模型的训练误差和泛化能力， C 值越大，模型对训练数据的拟合要求越高。

(3) 模型优化

为防止模型过拟合，提高泛化能力，需采用多种优化策略。正则化是常用的方法之一，对于神经网络，L1 和 L2 正则化通过在损失函数中添加正则化项，约束模型参数的大小，避免模型过于复杂；对于支持向量机，惩罚因子 C 也起到类似的正则化作用。此外，还可采用 Dropout 技术（常用于神经网络），在训练过程中以一定概率随机丢弃神经元，减少神经元之间的相互依赖，增强模型的泛化能力。在训练过程中，还需关注模型的训练误差和验证误差变化趋势，当训练误差持续下降而验证误差开始上升时，表明模型可能出现过拟合，此时应及时调整参数或停止训练。

3 案例分析

3.1 案例背景

选取某智慧电厂的锅炉设备作为研究对象，该锅炉设备在运行过程中产生了大量的温度、压力、流量、振动等数据。采用传统评估方法对锅炉设备进行运行状态评估时，存在故障发现不及时、维护成本高等问题。为提高锅炉设备的运行安全性和经济性，构建基于大数据分析的运行状态评估模型。

3.2 模型应用与实施

3.2.1 数据采集与处理

在锅炉设备上安装各类传感器，采集设备运行过程中的实

时数据，并从设备管理系统中获取历史运行数据和维护记录数据。对采集到的数据进行预处理，包括缺失值填充、噪声去除和标准化处理。

3.2.2 特征提取与选择

运用时域分析和频域分析方法，从预处理后的数据中提取温度变化率、压力波动幅度、振动频谱能量等 50 个特征。采用主成分分析方法，对特征进行筛选，最终保留 20 个最具代表性的特征。

3.2.3 模型构建与训练

选择神经网络作为评估模型，构建包含输入层、隐藏层和输出层的三层神经网络结构。输入层节点数为 20，对应提取的 20 个特征；隐藏层节点数通过试凑法确定为 30；输出层节点数为 3，分别表示设备正常运行、轻微故障和严重故障三种状态。利用训练集对神经网络进行训练，经过 1000 次迭代训练后，模型的损失函数收敛。

3.3 结果分析

将测试集数据输入训练好的模型进行评估，结果显示模型的准确率达到 92%，召回率达到 90%。与传统评估方法相比，基于大数据分析的评估模型能够更准确地判断锅炉设备的运行状态，提前发现潜在故障隐患。例如，在一次评估中，模型预测锅炉设备存在轻微故障，经现场检查发现设备的某一阀门存在轻微泄漏，及时进行维修后，避免了故障的进一步扩大。

4 结论

综上所述，本文构建了基于大数据分析的热工设备运行状态评估模型，通过数据采集与预处理、特征提取与选择、评估模型构建等步骤，实现了对热工设备运行状态的准确评估。案例分析表明，该模型相比传统评估方法具有更高的准确性和有效性，能够为热工设备的优化运行和预防性维护提供科学依据。在未来的研究中，可进一步优化模型算法，结合更多的设备数据和外部因素，提高模型的预测能力和泛化能力，为智慧电厂的发展提供更有力的技术支持。

参考文献：

- [1] 薛艳龙,裴玉隆.提高热工计量设备运行可靠性的技术措施[J].机械管理开发,2023,38(12):228-230.
- [2] 樊筱.提高热工计量设备运行可靠性的技术措施[J].内蒙古石油化工,2020,46(01):97-98.
- [3] 王静,孙开元.热工仪表与自动控制对火电设备机组节煤降耗的影响分析[J].冶金与材料,2019,39(01):155-156.
- [4] 罗云.电厂热工自动化系统检修常见问题分析及处理[J].科技与创新,2016,(03):147-148.
- [5] 罗云.电厂热工自动化系统检修常见问题分析及处理[J].科技与创新,2016,(03):147-148.