

文章编号: 0258-0926(2021)03-0160-06; doi:10.13832/j.jnpe.2021.03.0160

基于局部离群因子和神经网络模型的设备状态 在线监测方法研究

沈江飞¹, 李怀洲², 黄立军¹, 毛晓明¹, 张 圣¹

1. 苏州热工研究院有限公司, 江苏苏州, 215004; 2. 福建宁德核电有限公司, 福建宁德, 355209

摘要: 核电厂重大设备状态在线监测是保障核电厂安全和经济运行的重要技术, 针对传统阈值监测的固有缺陷, 提出一种基于局部离群因子 (LOF) 和神经网络模型的设备状态在线监测方法。此方法属于多参数动态阈值监测方法, 首先分析监测对象的故障模式和故障现象, 选择一组可覆盖故障现象的传感器测点; 根据设备运行特点采集足够长的历史运行数据, 筛除异常数据; 计算历史运行数据的 LOF, 以历史运行数据为输入、LOF 为输出, 建立并训练得到神经网络模型; 最后基于神经网络模型和传感器测点实时数据计算设备健康指数, 监控当前设备健康状态。将本文的监测方法用于循环水泵泵体健康状态的监测, 并采集了一段时间的正常数据和异常数据以验证其监测效果, 验证结果表明, 本文提出的监测方法可以提前 10 d 进行预警, 降低误报率, 大幅提升监控效能。

关键词: 核电设备; 智能监测; 局部离群因子 (LOF); 神经网络

中图分类号: TL113 **文献标志码:** A

Study on Online Monitoring of Equipment Condition Based on Local Outlier Factor and Artificial Neural Networks Model

Shen Jiangfei¹, Li Huaizhou², Huang Lijun¹, Mao Xiaoming¹, Zhang Sheng¹

1. Suzhou Nuclear Power Research Institute, Suzhou, Jiangsu, 215004, China; 2. Fujian Ningde Nuclear Power Co., Ltd., Ningde, Fujian, 355209, China

Abstract: The centralized online monitoring technology plays the most important role in nuclear power plants for the safety of major equipments and economic operation. In order to solve the false alarm and alarm failure problems in the traditional online monitoring, a new artificial intelligence monitoring method based on the local outlier factor and artificial neural networks model is put forward in this paper. This method is one of the multiple parameter dynamic threshold detection method. Firstly, a group of monitoring parameters of equipment is selected by analyzing the failure modes and failure phenomena of equipment. Secondly, enough data of this group of parameters needs to be collected and the abnormal data needs to be screened out. Thirdly, all the selected data is used to calculate the local outlier factor, and then the neural network model will be established by inputting the selected data and the local outlier factor. Finally, the neural network model can be used to assess the equipment health index with the real-time data of equipment parameters as input, and the health index represents the real-time health of equipment. In this paper, this method is used to develop a monitoring model of circulating water pump. In order to verify the validity of the model, enough monitoring data of healthy equipment and malfunction equipment are used to verify the monitoring results. The results show that the method can provide a pre-alarm for the early failure of the equipment with low false alarm rate, greatly improving the monitoring efficiency.

Key words: Nuclear power equipment, Intelligent monitoring, Local outlier factor (LOF), Artificial neural networks

收稿日期: 2020-03-30; 修回日期: 2020-11-26

作者简介: 沈江飞 (1991—), 男, 硕士研究生, 目前主要从事核电厂设备状态监测技术研究, E-mail: shenjiangfei@cgnpc.com.cn

0 引言

设备持续良好运行是核电厂稳定运行的基本保障，为保障设备运行可靠，核电厂设备工程师以及运行人员需要对设备运行状态进行实时监测，根据设备运行状态实施相应维护措施。基于设备运行状态的维护措施可在延长设备寿命的同时，保障设备安全可靠运行，显著提升设备可靠性和经济性，已经成为未来核电厂设备维护的主要策略方向^[1]。

目前，国外广泛采用基于高级模式识别技术（APR）的智能监测方法对核电厂设备进行在线监测，通过对设备正常运行状态下的历史数据进行训练建模，监测并识别相似的运行模式，再在相似运行模式下判断设备参数是否异常，从而大幅提高监测灵敏度^[2]。国内如丁锋^[3]、王凯^[4]等利用自回归模型对时间序列等相关参数的监测进行研究和验证，但是该方法只适用于与时间具有明确关系的参数监测；李锡文^[5]、余良武^[6]、董玉亮^[7]等基于信息融合、分析特征值、高斯混合模型等技术，计算特定设备的故障特征参数，监控设备的运行状态，此方法建立的模型只能用于特定设备，且需要单独寻找每个设备的特征函数。强文渊^[8]等通过滑动窗截取数据段并对每个数据段进行聚类分析，识别每一段监测数据的分布特征，由于需要对每一时刻的数据进行筛选和聚类计算，计算量大，且不适用于工况变化复杂的设备。丁显^[9]等通过模糊神经网络模型计算目标参数与实测值的偏差变化规律，从而实现对风电设备状态的监测，此法需要找到关键特征参数，普适性较低。上述研究针对特定设备的状态智能预警做了极有意义的研究，但是上述方法的普适性相对不足，难以被大规模应用且监测目标特征的物理意义难以解释，更为关键的是，上述方法都难以给出设备劣化程度的量化指标。

基于上述原因，本文以核电厂设备状态在线监测方法为研究对象，研究建立一种基于局部离群因子（LOF）和神经网络模型的设备智能在线状态监测模型，以满足核电厂大量系统和设备的健康状态监测智能化需求，提前发现和预警重要设备的早期异常状况。

1 局部离群因子模型监测方法

1.1 局部离群因子（LOF）

状态监测过程中离群点一般是指异常运行数

据，其产生可能来自于设备运行环境变化、设备零部件自然老化磨损或异常损伤、数值测量误差和传输误差等。由于核电设备运行相对稳定，可靠性高，因此故障数据较少，而正常运行状态下的数据经过长年积累，已经几乎覆盖所有可能的运行工况，因而可以以正常运行状态下的数据为参考，通过实时监测数据的离群程度来判断设备运行状态是否异常。当设备运行状态偏离正常运行模式时，监测模型发出预警，提醒工程师进一步检查设备并明确报警原因，以根据检查结果维修设备或优化监测模型。

定义 $d(p,o)$ 为数据对象 p 与对象 o 之间的欧式距离， o 为数据集 D 中除 p 以外的集合 $D \setminus \{p\}$ 中的某一数据；对任意的超参数 k ，可定义数据对象 p 的第 k 距离为 $[k \cdot \text{dist}(p)]$ ，满足以下 2 项条件时有 $k \cdot \text{dist}(p) = d(p,o)$ ^[10]：

(1) 至少存在 k 个数据对象 $o' \in D \setminus \{p\}$ 满足 $d(p,o') \leq d(p,o)$ 。

(2) 至少存在 $k-1$ 个数据对象 $o' \in D \setminus \{p\}$ 满足 $d(p,o') < d(p,o)$ 。

将数据对象 p 的第 k 距离 $k \cdot \text{dist}(p)$ 内所有对象的集合称为 p 的第 k 距离邻域，记作 $N_k(p)$ ，其表达式为：

$$N_k(p) = \{q | d(p,q) \leq k \cdot \text{dist}(p)\} \quad (1)$$

对任意的 k ，定义数据对象 p 相对数据对象 o 的可达距离为 $\text{reach} \cdot \text{dist}(p,o)$ ，其值为“对象 o 的第 k 距离”与“对象 p 与对象 o 的欧式距离”中的较大值，表达式为：

$$\text{reach} \cdot \text{dist}(p,o) = \max\{k \cdot \text{dist}(o), d(p,o)\} \quad (2)$$

根据式 (1) 和式 (2) 定义的可达距离和第 k 距离邻域，可定义数据对象 p 的局部可达密度 $\text{lrd}_k(p)$ 为数据对象 p 到其邻域 $N_k(p)$ 内数据对象的可达距离平均值的倒数，可表示为：

$$\text{lrd}_k(p) = \left[\frac{\sum_{o \in N_k(p)} \text{reach} \cdot \text{dist}_k(p,o)}{|N_k(p)|} \right]^{-1} \quad (3)$$

$\text{lrd}_k(p)$ 衡量了数据对象 p 在其附近集合内的稀疏程度， $\text{lrd}_k(p)$ 越大则表明分布较为稠密，反之则表明其分布较为稀疏。通过 $\text{lrd}_k(p)$ 可以定义 LOF：

$$\text{LOF}_k(p) = \frac{\sum_{o \in N_k(p)} \text{lrd}_k(o)}{|N_k(p)|} \quad (4)$$

根据式(4)可知, $LOF_k(p)$ 可以表示第 k 距离邻域的数据对象 p 与邻域对象的密度差异指标。 $LOF_k(p)$ 的大小取决于局部环境,其可以给每个数据点对象都分配 1 个离散度指标 LOF。大量研究表明,若数据点的 LOF 值远远大于 1,表示 p 点的密度与数据的整体密度差异较大,则认为 p 点为离群点。假设 LOF 接近于 1,表示 p 点与数据的整体密度差异较小,因此可认为 p 点为正常点。

1.2 监测过程

受环境、负载变化、运行操作等影响,设备运行状态常常会发生不规律的变化,设备传感器测点数据的分布区域和分布密度较为离散。因此在判断异常点时,需要结合其局部邻域内点的分布情况进行判定。若其局部邻域内的点都很密集,则认为此数据点为正常数据点;若其距离最近的点距离都较远,则认为此数据点为异常点。

LOF 的大小可直观且量化反映样本点与其周围临近点的密度对比,非常适用于设备状态监测中异常数据点的判断。若以设备正常稳态运行作为筛选条件,对设备运行的历史监测数据进行筛选,并将筛选后的数据作为基础数据集,在每一新的时刻将新采集的监测数据加入基础数据集组成一个新数据集,并重新计算新数据集中新采集监测数据对象的 LOF 值。若 LOF 值过大,则表明当前时刻设备运行状态偏离历史正常运行状态。

由于基础数据集通常是指近 1~3 a 的数据,数据采集间隔不高于 1 h,至少有上万个数据点。而新监测数据点只有 1 个,但是按照上述计算模型,由于新监测数据的加入组成新数据集,每次监测设备状态都需要将新数据集整体作为输入,因此近 99.99%的数据点被反复计算。以 900 MW 的核电机组为例,需用到的监测模型约 200 个,每个模型的监测频率不低于 5 min/次,若采用 LOF 模型直接监测设备运行状态,计算量极为庞大且浪费计算资源,尤其不适用于核电厂大量设备集中监测的情况。因而需要引入神经网络模型,通过训练、回归分析的方式预测新监测数据点的 LOF 的近似值,避免直接计算 LOF 造成大量重复计算。

2 神经网络拟合模型监测方法

设备运行状态的评估往往需要采集 5~20 维度甚至上百维度的特征参数进行评估,由于机械

设备运行状态复杂,受干扰因素多,数据点的 LOF 与数据点各特征参数之间的关系是一个复杂的多维度非线性关系,因此为预测数据点的 LOF 的近似值,需采用具备复杂非线性拟合能力的神经网络模型。

人工神经网络是模拟生物神经网络完成信息处理的一种数学模型,具有很强的非线性拟合能力,不需要获得关系方程即可训练得到输入-输出的映射关系,在函数逼近、信号处理等方面被广泛应用。BP 神经网络是依据误差逆传播算法训练的具有多层前馈的网络,其权值与阈值通过神经网络不断进行反向传递,最终使得误差的平方和在神经网络中达到最小值,即在进行数据拟合时具有精度高、误差小等优点。

本文选用单隐层 BP 神经网络,以各监测特征参数的基础数据集作为输入,以 LOF 作为输出,拟合 LOF 与各监测特征参数的关系模型,训练神经网络模型。当设备状态监测需求产生时,只需输入新监测数据点并计算神经网络模型的预测结果,即可得到当前数据点的 LOF 的近似值。若计算结果不符合预期,也只需要对基础数据进行更新补充,并再次训练神经网络模型即可。

由于 LOF 的计算采用了欧式距离,因而需要对各个监测数据进行归一化处理,即:

$$x'_i = \frac{\max(x) - x_i}{\max(x) - \min(x)} \quad (5)$$

式中, x 为原始数据; x' 为归一化后的数据;下标 i 为数据点序号。

基于 LOF 和神经网络模型的多参数智能监测模型流程如图 1 所示。为减少瞬时随机参数变化造成误报警的情况,定义健康指数(HI)为设备当前设备状态与正常运行状态的偏离程度,HI 的计算可表示为:

$$HI_t = (1 - \lambda)HI_{t-1} + \lambda LOF_t \quad (6)$$

式中, λ 为平滑系数;下标 t 为时刻。

HI 的计算是采用指数加权移动平均统计的方法,兼顾历史观察值和当前观察值之间的关系,从而减小误报概率,HI 值越大表明设备当前运行参数偏离最优状态越严重。

3 建模与案例验证

循环水泵是核电厂的重要冷源设备,在机组常规运行期间,要求其长期稳定可靠运行,因而需要对其运行状态进行长期监测,以避免影响机

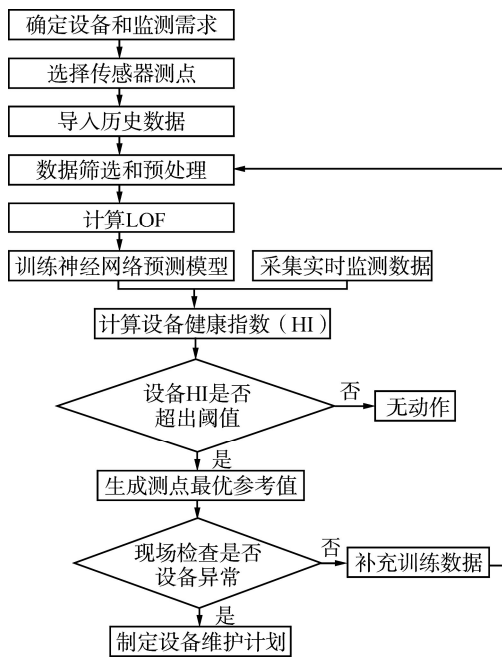


图1 基于LOF和神经网络模型的在线参数智能监测流程

Fig. 1 Intelligent Online Monitoring Process Based on LOF and Neural Network Model

组发电和运行稳定性。循环水泵主要包括电机、齿轮箱和泵体3个部分，本文选择对泵体运行状态进行监测，主要通过监测轴承温度情况来监测轴承、叶轮的损伤，监测点的选择如表1所示。

表1 测点选择
Tab. 1 Selected Parameters

序号	测点名称	监测参数
1	CRF001MI	泵电机电流
2	CRF102MT	泵上推力轴承温度
3	CRF104MT	泵上径向轴承温度
4	CRF142MT	泵下径向轴承温度
5	SRI001MT	轴承润滑油冷源温度
6	CRF502MT	CRF冷却水温度
7	CGR015MT	轴承润滑油温度

CRF—循环水系统；SRI—常规岛闭路冷却水系统；CGR—循环水泵润滑系统

为考虑环境因素和机组状态变化对设备运行状态的影响，选取某核电厂一整年的监测点数据（2018年8月至2019年8月）作为训练数据，在训练之前筛除故障阶段和设备维护过程的数据、监测范围外数据以及因传感器造成的突变数据。由于LOF模型本身具备一定异常数据筛选能力，因而对训练数据的筛选要求不高，取 $k=20$ 计算LOF值，计算获得LOF的分布规律，见图2。

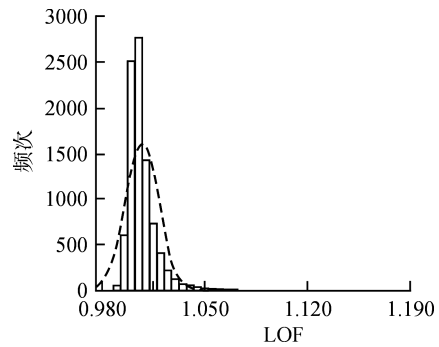


图2 LOF分布规律

Fig. 2 LOF Coefficient Distribution

使用正态分布曲线拟合LOF的分布规律，所得拟合结果的均值为1.007，标准差为0.01132。按照正态分布双边检验规则可计算得到置信度为99.99%时LOF的置信区间为[0.961, 1.052]。因此在正常情况下LOF值小于0.9或大于1.2为绝对小概率事件，因此可以选择0.9和1.2作为参数异常判断标准的上下限值。

以训练数据和计算得到的LOF值为输入，构建神经网络预测模型，隐层神经单元数量为10，训练函数采用LM(Levenberg-Marquardt)算法，采用Matlab神经网络工具箱建立神经网络模型。基于训练得到的神经网络模型计算2019年9月的设备监测数据的LOF值，选取平滑系数 $\lambda=0.2$ 代入式(6)，计算获得设备的HI指数变化趋势，如图3所示。

由图3可知，9月期间设备的HI指数趋势平稳，并维持在1.0附近，远低于设定的上下限，即在设备运行正常时，难以触发监测模型误报警。这与现场反馈的9月期间设备运行状态正常的情况一致。

据现场反馈，在10月1日至11月6日期间，

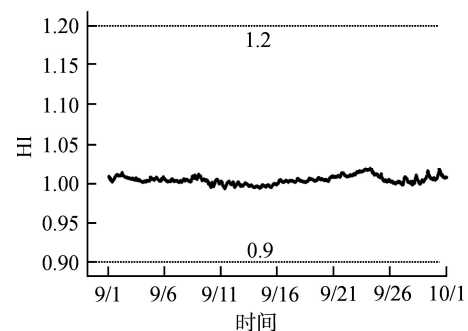


图3 9月正常运行期间设备HI健康指数变化趋势

Fig. 3 Health Index(HI) during Normal Operation Period in September
9/1—9月1日；其余类同

由于该循环水泵上方球面滚子推力轴承的铜保持架出现断裂,并于11月6日导致设备异常停机,期间泵组轴承温度出现小幅波动。采用LOF和神经网络方法建立和训练得到的监测模型对此设备在上述运行期间的数据进行验证,数据采样频率为5 min/次,设备健康指数HI的监测结果如图4所示。

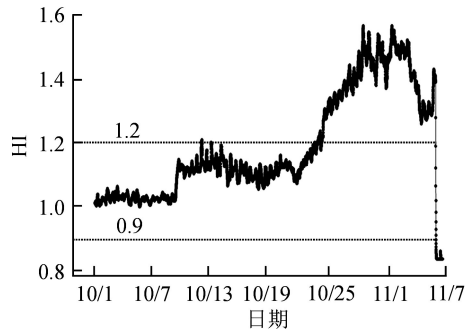


图4 10月故障运行期间设备HI健康指数
Fig. 4 Health Index (HI) during Abnormal Operation Period in October

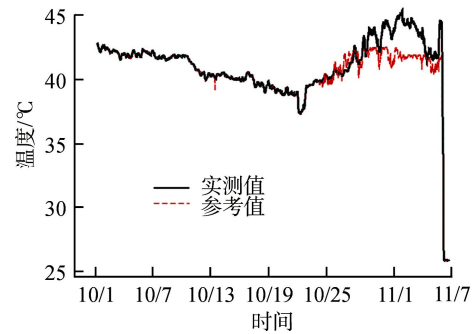
由图4可知,10月9日后,设备的HI出现明显上涨但尚处于限值内,经研究发现此期间工作人员对循环水泵的下轴承进行了补脂工作,因此导致CRF142MT测点温度大幅上涨,数据点偏离正常运行数据,从而引起HI指数增大。10月22日后,设备的HI再次出现快速上涨,10月25日超出设定的预警值1.2,并于11月1日左右达到最大值(约1.57),随后在11月5日电厂运行巡检发现设备有严重异响,11月6日泵组停机。

为便于分析和诊断,当HI指数超出设定的限值时,即设备发生异常时,自动计算各个测点数据的参考值以用于快速发现当前设备各测点数据的异常情况和偏离程度,参考值的生成方法可以选择最靠近当前监测数据的前 n 个对象的均值。若取 $n=10$,各测点的计算参考值和实测值的变化趋势如图5所示。

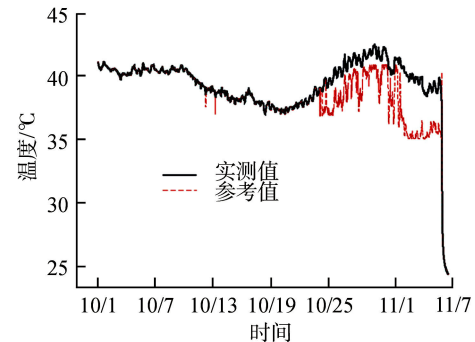
由图5可知,通过比较实测值和参考值曲线,在10月22日左右循环水泵上轴承CRF102MT和CRF104MT测点处的实测值就开始偏离参考值,并持续上涨,这是因为CRF102MT和CRF104MT测点处的温度过高,导致相应的冷却系统温度测点数据的参考值偏高,引起CGR015M测点处润滑油温度、SRI001MT测点处冷却水温度和CRF502MT测点处海水温度的参考值相比实测值偏高。设备监测测点的实测值偏离正常运行状态

下的训练数据,HI指数快速上涨并超出上限值。

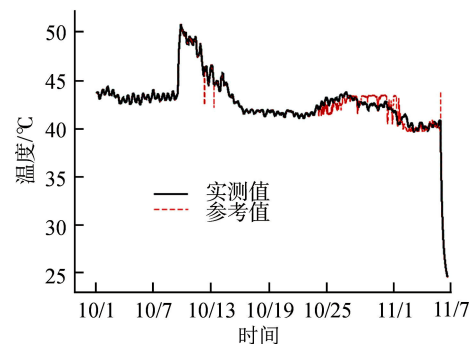
监测对象为低转速设备,没有安装振动传感器,然而从轴承出现异常直至损坏停机过程中,轴承温度一直远低于电厂分散控制系统(DCS)报警值,因而无法通过传统DCS在线监测手段发



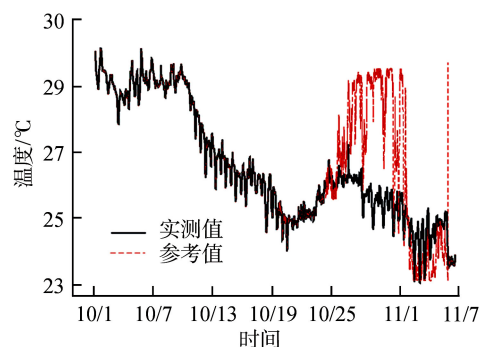
a CRF102MT 测点



b CRF104MT 测点



c CRF142MT 测点



d SRI001MT 测点

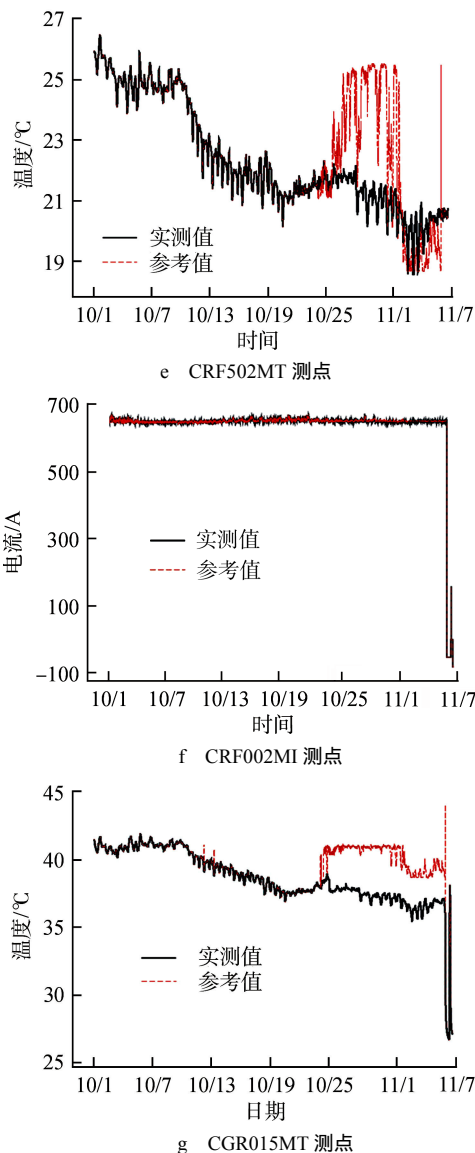


图5 设备参数和参考值趋势图

Fig. 5 Trend Chart of Equipment Parameters and Reference Value

现设备异常。相比运行巡检发现设备异音的时间，通过本文提出的监测模型可以提前 10 d 发出预警，从而为提前制定干预维护措施及应急方案争取了宝贵机会和时间，避免紧急停机而造成损失。

4 结论

针对核电厂设备状态在线智能监测技术难点，采用基于 LOF 和神经网络模型的设备状态在线监测技术，可有效监测到设备故障早期异常，提

前发现潜在问题并进行预警，主要技术特点如下：

(1) 采用 LOF 和 HI 作为评价设备参数状态的指标，通过离群程度表征设备故障劣化程度，直观显示设备健康状态劣化的趋势和程度。

(2) 采用 LOF 与神经网络模型相结合的方式建模，利用神经网络模型拟合 LOF 与设备传感器监测参数之间的关系，计算量小，节省计算资源，非常适用于核电厂大量设备的集约化监测。

(3) 监测模型对训练数据在设备运行工况覆盖的全面程度上要求较高，当训练数据较少时，容易出现误报警，且计算 LOF 的超参数 k 的取值对计算结果影响较大，需要根据各设备和参数情况进行调试选择。

参考文献：

- [1] AHMAD R, KAMARUDDIN S. An overview of time-based and condition-based maintenance in industrial application[J]. Computers & Industrial Engineering, 2012, 63(1): 135-149.
- [2] MALEY S. Continuous online monitoring guidebook. volume 5: COLM Technology infrastructure application [R]. Palma: EPRI, 2019: 1-70.
- [3] 丁锋. 复杂机械系统监测诊断中的时间序列模式识别技术[J]. 控制与检测, 2005(10): 45-47.
- [4] 王凯, 胡怡婧, 徐青. 基于时间序列 ARMA 模式识别的基坑监测预警优化研究[J]. 工业安全与环保, 2018, 7(44): 29-32.
- [5] 李锡文, 杨明金, 杜润生, 等. 模糊模式识别在铣刀磨损监测中的应用[J]. 机械科学与技术, 2007, 9(26): 1113-1118.
- [6] 余良武, 刘东风, 房友龙, 等. 油液监测故障模式识别理论研究现状[J]. 润滑与密封, 2018, 2(23): 115-118.
- [7] 董玉亮, 李亚琼, 曹海斌, 等. 基于运行工况辨识的风电机组健康状态实时评价方法[J]. 中国电机工程学报, 2013, 4(33): 88-95.
- [8] 强文渊, 韦家富, 刘友波, 等. 基于滑动时窗 K-Means 聚类的 CVT 早期故障诊断[J]. 电力电容器与无功补偿, 2019, 6(40): 89-95.
- [9] 丁显, 徐进, 滕伟, 等. 基于 RF 和 ANFIS 算法的直驱风电机组故障预警[J]. 噪声与振动控制, 2018, 2(38): 209-214.
- [10] 辛丽玲. 基于密度差异的离群点检测研究[D]. 北京: 北京交通大学, 2015.

(责任编辑：周 茂)