

基于堆叠稀疏自编码器的风机故障预警系统研究

苏振禄

泉州海洋职业学院 海洋工程学院 福建石狮 362700

DOI:10.12238/ems.v7i12.16467

[摘要] 针对风机在运行过程中, 现有深度学习模型无法准确及时的监测风机运行状态, 结合堆叠自编码器思想, 提出一种基于堆叠稀疏自编码的风机故障预警方法。首先采集风力发电机在正常运行条件下的状态变量作为堆叠稀疏自编码器的输入, 以此建立针对风力发电机系统正常运行状态的估计模型。依据正常与异常运行状态间的数据结构差异, 当系统进入异常状态, 模型预测的输出与实际观测值之间的偏差增大。应用滑动窗口算法, 计算出用于监测风力发电机系统状态的关键指标, 基于高斯分布的预警阈值设置方法, 实现对潜在故障的及时预警。利用实际 SCADA 系统历史运行数据进行仿真, 结果表明, 相比于其他方法, 堆叠稀疏自编码模型具有更好的估计性能, 且能够在故障发生前及时发出预警信息, 对运行人员具有一定的指导意义。

[关键词] 风机; 故障预警; 状态估计; 堆叠稀疏自编码器; 滑动窗口

引言

风力发电机因其复杂的内部结构和特殊的工作环境, 导致其在运行期间较易出现故障^[1]。这些故障一旦发生, 在电机内部会引起发电效率的变化, 并且在严重的情况下可能会损害执行器与传感器。面临重大故障时, 可能需要进行停机维修和零部件替换, 从而引发显著的维修成本。通常, 这些故障是逐渐发展而成, 从微小变化累积到质的变化。如果能在此过程中实施有效的状态监控^[2], 并在故障早期提供预警, 以防止由于设备故障而造成的更大经济损失。因此, 监测风力发电机的运行状态成为一项关键任务。

风场在风机日常运行阶段, 通过大量的传感器监测风机各部件的参数变化趋势, 因此风机会产生大量的历史数据, 现场工作人员通过监视这些数据来判断风机的运行状况, 只能通过运行人员的经验, 无法有效的利用海量的历史数据, 通过对历史数据进行分析, 可实现对设备状况的评估, 给予现场维护指导建议。在大数据挖掘和状态预测方面的研究中^[3], 神经网络一直具有较好的效果, 其中无监督学习代表一自编码器算法, 不需要关注输入输出间的相互影响, 只需将多维输入数据进行训练, 便可使输出尽可能重构输入, 利用这一特性, 可以对设备的多状态进行估计。通过无监督学习等深度学习方法提取风力发电机参数的特征分布。依据运行参数, 构建一个反映系统实际工作状态的模型, 进而获得对风力发电机运行参数的监测和预测值。然后, 利用模型预测值与实际观测值之间的重构误差来实现风力发电机系统的故障预警。

因此, 本文提出一种基于堆叠稀疏自编码器的风力发电机异常状态监测方法。利用堆叠稀疏自编码器强大的特征提取能力构建风机运行状态估计模型。结合滑动窗口算法与高斯分布的预警阈值设置方法, 实现对潜在故障的及时预警。这种方法不仅增强了故障诊断的准确性, 而且通过早期故障预警, 有助于避免系统故障对风电发电效率的负面影响。

1 深度学习模型

1.1 自编码器

自编码器 (Auto-Encoder, AE)^[4] 是一种用于无监督学习的神经网络, 其核心目的在于通过最小化输入与输出之间的误差来提取隐层特征并学习参数。具体而言, 自编码器的训练过程旨在调整网络参数, 以便输出可以尽量复原输入数据, 从而实现两者之间的重构误差最小化。该网络架构通常包含编码阶段, 用于将输入数据转换为一个更低维的表示, 以及解码阶段, 该阶段旨在用低维表示重构原始数据, 如图 1 所示。

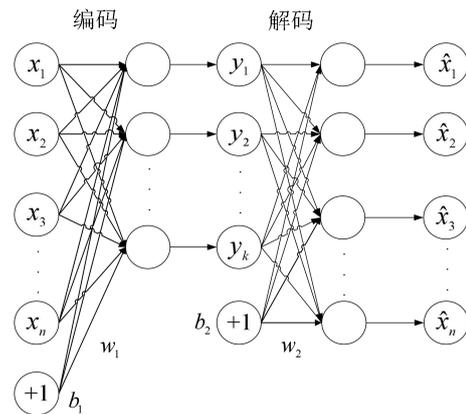


图 1 自编码器结构

在模型的训练过程中, 涉及到编码和解码两个主要阶段。初始阶段是编码过程, 其中输入数据 x 被传递通过隐层网络, 实现从原始数据到编码表示的映射, 产生编码结果 y , 随后, 在解码过程中, 编码得到的 y 被进一步处理以生成最终输出 \hat{x} , 该输出旨在重构原始输入 x 。

$$y = f(w_1 x + b_1) \quad (1)$$

$$\hat{x} = f(w_2 y + b_2) \quad (2)$$

式中, $f(\cdot)$ 表示激活函数, 通常为 Sigmoid 函数, w_1 、 w_2 、 b_1 、 b_2 分别代表两种变换过程中的权重和偏置参数。

1.2 堆叠稀疏自编码器

稀疏自编码器 (Sparse Auto-Encoder, SAE)^[5] 是一种通过引入稀疏正则化项到损失函数中来实现数据特征的稀疏化提取的机制。这种方法主要目的是为了减少自编码器对训练数据的过度拟合, 通过限制隐藏层中激活单元的数量来达到这一点。稀疏自编码器强制少数隐藏层神经元激活, 从而在大量数据中提取有意义的特征, 这有助于提高模型的泛化能力并保持数据特征的紧凑表示。通过应用 Sigmoid 函数确定隐藏层单元的输出, 得到的值接近 1 表示单元处于激活状态, 而值接近 0 则表示单元处于抑制状态。 $a_j(x)$ 表示隐含层第 j 个单元的激活水平, 则该单元的平均激活水平可以表示为:

$$\hat{\rho}_j = \frac{1}{k} \sum_{i=1}^k a_j[x(i)] \quad (3)$$

设定一个接近于零的阈值 $\hat{\rho}_j$ 以促使大多数隐藏层单元保持在抑制状态。 ρ 这个阈值称为稀疏性参数, 它在损失函

数中引入了一个稀疏惩罚项，该惩罚项用于量化两个随机变量（平均值为 ρ 和 $\hat{\rho}_j$ ）之间的差异。稀疏性惩罚项可以通过 Kullback-Leibler (KL) 散度来表达，从而形成惩罚项的具体数学形式。该方法通过最小化损失函数中的 KL 散度来保持稀疏性，从而使得大多数隐含层单元保持非激活状态，有助于防止过拟合并提高模型的泛化能力。

$$PN = \sum_{j=1}^{s_2} KL\left(\rho \parallel \hat{\rho}_j\right) \quad (4)$$

$$KL\left(\rho \parallel \hat{\rho}_j\right) = \rho \ln \frac{\rho}{\hat{\rho}_j} + (1-\rho) \ln \frac{1-\rho}{1-\hat{\rho}_j} \quad (5)$$

式中， s_2 表示隐藏层中单元的数量；当实际激活率 $\hat{\rho}_j$ 接近稀疏性参数 ρ 时，KL 散度趋近于零。这反映了模型的实际激活率与设定的稀疏目标之间的接近程度。损失函数可以定义为：

$$J(W, b) = \frac{1}{k} \sum_{i=1}^k \left\| x(i) - \hat{x}(i) \right\|^2 + \frac{\lambda}{2} \sum_{l=1}^{n_l} \sum_{i=1}^{s_l} \sum_{j=1}^{s_{l+1}} [W_{ij}(l)]^2 \quad (6)$$

在此表达式中，第二项代表基于模型输出与目标输出的误差项， λ 为正则化参数； n_l 代表神经网络的层数。通过这种方式，总损失函数结合了预测误差、稀疏性约束和权重衰减，为模型提供了一个综合的优化目标。总的损失函数为：

$$J_{sparse}(W, b) = J(W, b) + \beta PN \quad (7)$$

在该公式中， β 表示稀疏性惩罚项参数，用于调控稀疏惩罚在总损失函数中的影响。为了优化这一损失函数并找到最佳的权重矩阵 W 和偏置向量 b ，采用拟牛顿方法。这种优化算法适用于寻找多变量函数的局部最小值，且不需要计算二阶导数。通过迭代更新 W 和 b ，拟牛顿方法能够有效地最小化总损失函数，进而使得稀疏自编码器能够达到优秀的隐藏层表征能力，确保了自编码器可以学习到数据中的重要特征，从而获得更好的数据表示。

$$W_{ij}(l) = W_{ij}(l) - \alpha \frac{\partial}{\partial W_{ij}(l)} J_{sparse}(W, b) \quad (8)$$

$$b_i(l) = b_i(l) - \alpha \frac{\partial}{\partial b_i(l)} J_{sparse}(W, b) \quad (9)$$

式中： α 为学习率。

堆叠稀疏自编码器^[6]是通过将多个稀疏自编码器 (SAE) 顺序堆叠而成的深度网络结构。在这一体系中，第一个 SAE 的隐藏层输出被用作第二个 SAE 的输入。通过这种方式，每个 SAE 的隐藏层输出依次成为下一个 SAE 的输入。这个过程通过多层的前向传播连续进行，从而实现数据特征的逐层降维。在完成前向传播后，通过构建反向传播 (BP) 神经网络来调整和微调整个网络的权重和偏置参数。这个微调过程旨在优化网络的整体性能，从而增强模型的泛化能力，并确保模型能够有效地学习并表示数据的复杂结构。通过这种分层训练和微调的方法，堆叠稀疏自编码器能够有效地捕获数据的深层特征，提高模型对新数据的预测和泛化能力。

1.3 基于堆叠稀疏自编码器的状态重构模型

图 2 展示了基于堆叠稀疏自编码器的网络模型架构，该模型包括输入层、多个稀疏自编码器组成的编码层、一个 BP 神经网络以及重构的输出层。在这种结构中，输入层负责接收风力发电机各项运行参数的历史数据，并对这些数据进行预处理操作，如去除异常值和进行标准化处理，以便于后续分析。

编码层由多个稀疏自编码器网络层组成，负责对输入数据进行压缩和降维处理。这一过程旨在有效提取数据中的时

间序列特征，减少数据的维度同时保留其关键信息。经过编码层处理后的降维数据，随后被用作 BP 神经网络的输入，进行监督式学习和网络参数的微调。通过应用反向传播算法 (BP 算法)，模型能够对其参数进行优化，从而增强其对新数据处理的泛化能力。整体而言，该网络模型通过其多层结构实现了对风力发电机运行参数的深度学习和特征提取，致力于提升预测性能和模型的泛化能力。

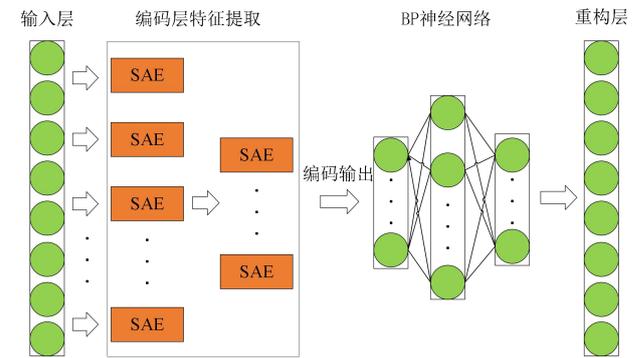


图 2 堆叠稀疏自编码器网络模型结构

2 风机运行状态异常预警方法

2.1 模型变量选取

在风机 SCADA 系统中记录了风速、功率、发电机转速、转子温度、轴承温度等等近百个参数。本文基于风电机组运行参数的内在关系对风电机组的异常运行工况进行预警，结合风电机组运行机理的深入理解和历史数据分析^[7]，选取齿箱油温、电机轴承温度、发电机转速、有功功率、风速、机舱温度进行建模，旨在确保模型能够准确地捕捉到可能导致性能下降或故障的关键指标。详细列表可参见表 1。

表 1 模型所选参数

参数	单位	符号
齿箱油温	°C	x_1
电机轴承温度	°C	x_2
发电机转速	r/min	x_3
有功功率	kW	x_4
风速	m/s	x_5
机舱温度	°C	x_6

2.2 数据预处理

为了削弱不同参数数据间的量纲差异，并加快模型的收敛速度，在建模阶段采取了归一化处理，如公式 (10) 所示：

$$x^* = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (10)$$

其中， x 和 x^* 分别表示各参数原始数据和归一化后的数据； x_{\max} 与 x_{\min} 为最大值和最小值。归一化处理确保了所有参数在同一尺度上进行比较和计算，减少了不同量纲之间可能引起的偏差，从而有助于提高算法的稳定性和效率。

2.3 预警方法

利用上述模型，能够计算出各参数正常运行时的预期值。通过正常以及故障状态下的预测值与实际值之间的差异不同，识别设备的具体状况，根据此种关联，构造合适的预警阈值就可实现异常状态预警。

使用适当宽度的滑动窗口算法可以有效地捕捉统计特性的持续变化，同时减少随机波动的干扰，从而增强状态监测的准确性和可靠性。

基于监测变量的估计残差，研究采用平均绝对误差作为监测指标，以构建滑动窗口监测模型。该模型的核心在于通过分析估计残差来评估系统的性能，如公式 (11) 所示：

$$\varepsilon = x_i - \hat{x}_i \quad (11)$$

式中, x_i 表示实际观测值中的第 i 个变量; \hat{x}_i 代表该变量的预测估计值。

为了进一步分析系统性能, 对残差序列实施相同的滑动窗口技术, 以此计算得出监测指标。该监测指标通过整合特定时间窗口内的残差数据, 提供了对系统状态变化的连续性评估。这种方法不仅增强了对系统行为的洞察, 而且通过持续监控残差变化, 有助于及时识别和预防潜在的系统异常或故障。求取监测指标 R_c 如下:

$$R_c = \frac{1}{m} \sum_{t=1}^m |\varepsilon(t)| \quad (12)$$

式中, $\varepsilon(t)$ 表示估计残差。

基于统计学中心极限定理可知, 模型估计值与实际值的差值样本的平均值始终围绕在总体均值周围, 并且呈高斯分布。故预警阈值可通过高斯分布理论的阈值上限公式^[8]求取:

$$U_{th} = \mu + t_{1-\frac{\alpha}{2}} \frac{S}{\sqrt{N}} k \quad (13)$$

式中, μ 为监测指标的均值; N 为样本数; S 为监测指标对应的标准差; $t_{1-\frac{\alpha}{2}}$ 为随置信度取值的不同而变化的小概率率值; 系数 k 一般为 2。

通过分析风力发电机系统的监测指标及其变化趋势, 可以有效地对系统的运行参数进行跟踪和监视。通过这些参数指标, 可以精确地定位到可能的故障原因, 并且能够实施早期故障诊断。此方法不仅增强了对系统运行状态的了解, 而且通过早期识别异常或偏差, 有助于采取预防措施, 避免潜在的故障发展成更严重的问题。

2.4 风机运行状态异常预警流程

基于堆叠稀疏自编码器的风机异常状态预警方法步骤如下: 预警方法分为两个核心环节: 离线分析和在线监测。

(1) 离线分析阶段: 在这个阶段, 首先对风力发电机在历史正常运行期间收集的数据进行预处理, 这包括数据的清洗、标准化等步骤, 随后将这些数据分割为训练集和测试集。使用训练集数据, 构建堆叠稀疏自编码器网络, 用于估计系统状态, 并从中提取出风力发电机在正常运行条件下的特征信息。此后, 利用测试集数据对构建的模型进行验证, 以确保模型的准确性和可靠性。在这个过程中, 还会根据正常运行数据计算出预警阈值。

(2) 在线监测阶段: 在实际运行中, 系统会实时收集风力发电机的运行数据。这些实时数据经过预处理后, 被输入到已经训练好的堆叠稀疏自编码器网络中, 以得到各运行参数的估计值。然后, 通过构建的模型生成差值序列, 并应用滑动窗口方法实时计算状态监测指标。一旦这些监测指标超过预先设定的预警阈值, 系统就会立即触发异常状态的预警信号, 以便及时进行故障诊断或采取必要的干预措施。

3 实例分析

选取某风电机组 SCADA 系统采集的 2014 年 1 月至 12 月的历史运行数据进行分析, 其中 80% 用于模型训练, 20% 用于测试, 以评估模型的性能。为了验证模型在实时监测中的有效性, 选取电机轴承温度过高与齿箱油温过高两组故障数据作为在线监测。上述数据采样周期均为 10min。

3.1 风机状态估计模型建立

表 3-1 展示了模型的输入和输出向量构成。模型输入向量为 $\mathbf{x} = \{x_1, x_2, x_3, x_4, x_5, x_6\}$, 输出向量为 $\hat{\mathbf{x}} = \{\hat{x}_1, \hat{x}_2, \hat{x}_3, \hat{x}_4, \hat{x}_5, \hat{x}_6\}$, 对应输入的估计输出。模型

训练是通过训练集完成的, 旨在优化堆叠稀疏自编码器网络的参数配置, 这些参数的详细信息可参见表 2。

表 2 模型结构参数

项目	数值
第一层稀疏自编码器隐含层个数	20
第二层稀疏自编码器隐含层个数	20
迭代次数	400
稀疏性参数	0.1
权重衰减系数	0.001
稀疏惩罚项系数	4
初始学习率	0.01

在确定模型参数后, 进行堆叠稀疏自编码器网络的训练过程。测试集中效果如图 3 所示, 用于评估模型的准确性和可靠性。

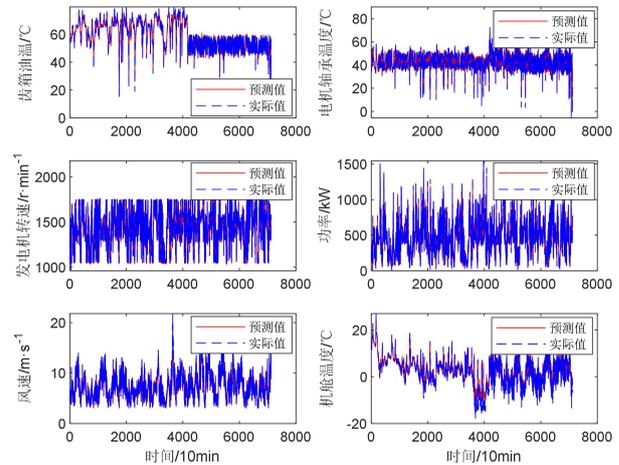


图 3 测试集模型验证结果

图 3 为在风力发电机正常运行期间, 模型对各运行参数的估计效果较好。模型的重构能力通过平均绝对百分比误差 (MAPE) 进行评估, 该指标反映了模型预测值与实际值之间的相对误差, 其计算公式如下所示。

$$\delta_i = \frac{1}{N} \sum_{j=1}^N \frac{|x_{ij} - \hat{x}_{ij}|}{x_{ij}} \times 100\% \quad (14)$$

式中, δ_i 为参数 x_i 的实际值与重构值的 MAPE 误差值。

各参数测试集的 MAPE 分别为 0.3008、0.3285、0.0791、0.6441、1.6415、9.5823, 平均 MAPE 值为 2.0961, 从中可以看出 MAPE 值较小, 模型具有较好的预测性能。

3.2 预警实例

选取齿箱油温过高与电机轴承温度过高两组故障数据, 用作故障验证集。验证集涵盖了对这些数据进行归一化处理, 使用 3.1 节训练的模型进行监测分析, 采用滑动窗口法与高斯分布理论计算状态监测指标, 设置窗口宽度 m 为 60, 即每个窗口覆盖了 600 分钟的数据, 取置信度 $1-\alpha$ 为 0.99。这种方法有助于定量分析和识别数据中的异常状态, 从而验证模型对于故障检测的有效性。

图 4 为齿箱油温过高故障数据的估计效果, 可以看到模型在机组正常运行时具有较好的估计效果, 第 215 个样本点时, 齿箱油温太高, 重构能力下降, 此时误差变大, 综合图 5 的齿箱油温状态监测曲线, 在第 159 个样本点发出预警, 提前 560min 发出异常状态预警, 相较于传统的 3-sigma 准则, 可以及时发出预警信息。

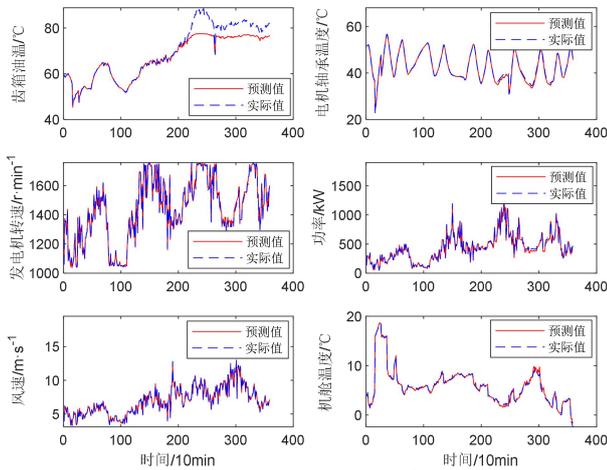


图4 验证集1的估计效果

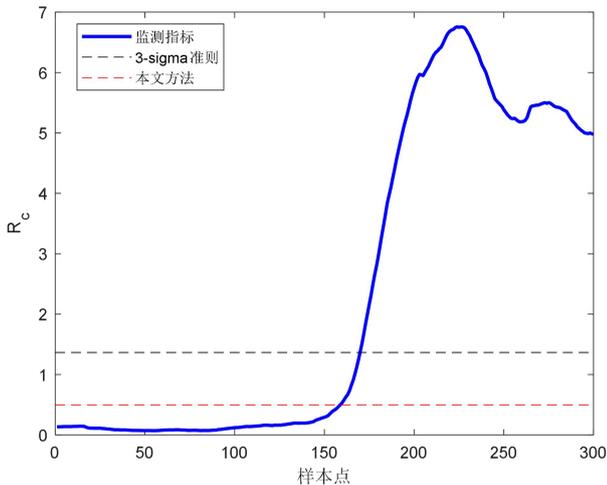


图5 齿箱油温状态监测曲线

图6为电机轴承温度故障数据的估计效果,第183个样本点时,电机轴承温度超限,重构能力下降,此时误差变大,综合图7的电机轴承温度监测曲线,在第125个样本点发出预警,提前580min发出异常状态预警。

当某个变量监测曲线在发生故障后明显偏离正常运行状态时,可以通过该变量判断可能的故障原因,实现对故障的初步诊断。预警模型可以实现对多个运行参数的状态进行监测,现场人员可通过模型发出的预警信息并结合各运行参数的状态监测曲线综合判断所发生的故障,实现早期故障诊断。

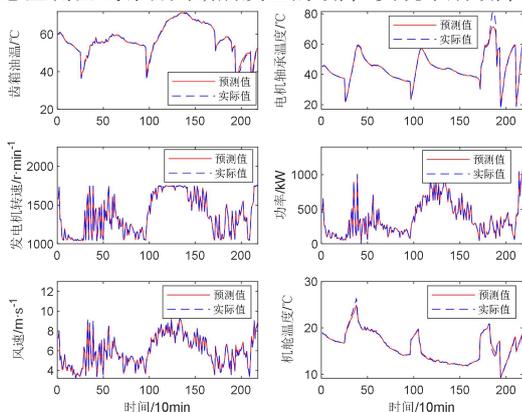


图6 验证集2的估计效果

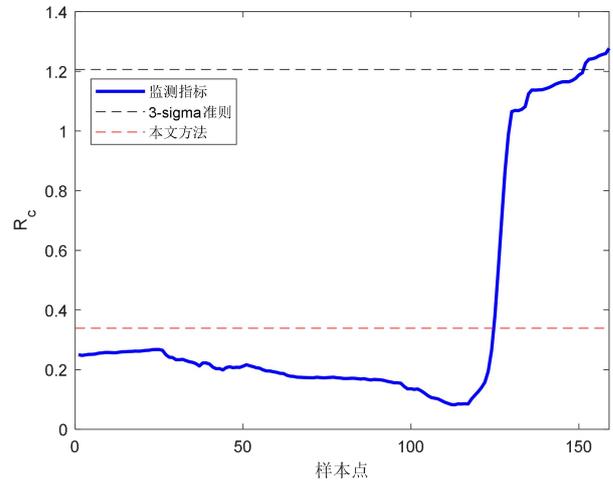


图6 电机轴承温度状态监测曲线

4 总结

- (1) 风力发电机运行工况多变,基于堆叠稀疏自编码器的状态估计模型,充分提取时间序列的内部特征,模型的输出可以很好的跟踪实际值。
- (2) 建立基于滑动窗口法的故障预警模型,以单个运行参数监测为例进行研究,实现对故障的精准定位,达到早期故障诊断的目的。
- (3) 该方法对维持风机安全稳定运行具有重要意义,所提故障预警方法还可以应用于其他机械、化工等多变量、多工况工业过程的异常状态识别。

【参考文献】

[1] 崔楷, 许宜菲, 李雪松, 等. 基于广义回归神经网络的风电机组性能预测模型及状态预警[J]. 科学技术与工程, 2020, 20 (32): 13220-13228.

[2] 赵洪山, 刘辉海, 刘宏杨, 等. 基于堆叠自编码神经网络的风电机组发电机状态监测与故障诊断[J]. 电力系统自动化, 2018, 42 (11): 102-108.

[3] 尹诗, 侯国莲, 胡晓东, 等. 基于 AC-GAN 数据重构的风电机组主轴轴承温度监测方法[J]. 智能系统学报, 2021, 16 (06): 1106-1116.

[4] 袁非牛, 章琳, 史劲亨, 等. 自编码神经网络理论及应用综述[J]. 计算机学报, 2019, 42 (01): 203-230.

[5] OLSHAUSEN B A, FIELD D J. Emergence of simple-cell receptive field properties by learning a sparse code for natural images[J]. Nature, 1996, 381 (6583): 607- 609.

[6] 张智恒, 周凤星, 严保康, 等. 基于改进堆叠稀疏降噪自编码器的轴承故障诊断[J]. 轴承, 2021, (02): 35-41.

[7] 马良玉, 梁书源, 程东炎, 等. 基于 QM-DBSCAN 与 BiLSTM 的风电机组异常工况预警研究[J]. 计量学报, 2024, 45 (09): 1384-1393.

[8] 任建亭, 汤宝平, 雍彬, 等. 基于深度变分自编码神经网络融合 SCADA 数据的风电齿轮箱故障预警[J]. 太阳能学报, 2021, 42 (04): 403-408.

作者简介: 苏振禄 (1992-), 男, 学士, 中级, 研究方向为机电设备故障诊断与维护研究, E-mail: 504310179@qq.com.

课题名称: 泉州海洋职业学院 2025 年校级课题 (QH202550)