

引文:江新星,吴杰,薛一冰,等.基于 SOM-BP 级联神经网络的电驱离心泵健康状态识别方法[J].油气储运,2025,44(3):350-359.

JIANG Xinxing, WU Jie, XUE Yibing, et al. Health status recognition method of electrically driven centrifugal pump based on cascade-structured SOM-BP neural networks[J]. Oil & Gas Storage and Transportation, 2025, 44(3): 350-359.

# 基于 SOM-BP 级联神经网络的电驱离心泵健康状态识别方法

江新星<sup>1</sup> 吴杰<sup>2</sup> 薛一冰<sup>1</sup> 彭世亮<sup>3</sup>

1. 国家管网集团储运技术发展有限公司压缩机组维检修分公司; 2. 国家管网集团湖南分公司;  
3. 中国石油大学(北京)机械与储运工程学院

**摘要:**【目的】离心泵预测性维护是提升设备可靠性与运行效率的核心技术之一,在该过程中,对离心泵设备的健康状态识别是关键环节。然而,传统的健康状态识别方法多依赖于机器学习技术,高度依赖足量标记数据,难以直观清晰地表征监测数据与健康状态之间的对应关系,使其在实际复杂工况中的应用效果受限,亟需开发更精准、高效且适应性更强的健康状态识别方法。【方法】提出一种基于自组织映射(Self-Organization Map, SOM)神经网络与 BP(Back Propagation)神经网络级联的电驱离心泵健康状态识别方法。首先采用 SOM 神经网络方法对离心泵全生命周期振动数据进行预处理,提取时域、频域及时频域的多种统计特征与熵特征,从而全面表征设备的运行状态;其次,采用主成分分析法(Principal Component Analysis, PCA)对已提取的轴承振动信号特征进行降维与融合,有效减少冗余信息和计算复杂度,优化输入参数的模式,提升建模效率;最后,综合 SOM 神经网络与 BP 神经网络的优点,建立了基于 SOM-BP 级联神经网络的电驱离心泵健康状态识别模型。【结果】以某电驱离心泵的健康状态监测数据集为算例,对比了 SOM-BP 模型与常见的机器学习方法(随机森林模型、XG-boost 模型)识别电驱离心泵健康状态的准确率,以  $R^2$ 、MSE、RMSE 为模型评价指标,结果表明:基于 SOM-BP 级联神经网络模型的  $R^2$  值、MSE 值、RMSE 值分别为 0.901、 $0.8 \times 10^{-6} \text{ m}^2/\text{s}^4$ 、 $9.12 \times 10^{-4} \text{ m/s}^2$ ,显著优于传统的机器学习方法,展现出良好的鲁棒性与适应性。【结论】基于 SOM-BP 级联神经网络计算方法不仅提升了离心泵健康状态识别的精度,还可为离心泵故障诊断与剩余寿命预测提供数据支撑,同时为其他旋转机械的健康状态管理与诊断提供了新思路。(图 5,表 2,参 25)

**关键词:** 自组织映射; BP 神经网络; 电驱离心泵; 健康状态识别

中图分类号: TE832

文献标识码: A

文章编号: 1000-8241(2025)03-0350-10

DOI: 10.6047/j.issn.1000-8241.2025.03.011

## Health status recognition method of electrically driven centrifugal pump based on cascade-structured SOM-BP neural networks

JIANG Xinxing<sup>1</sup>, WU Jie<sup>2</sup>, XUE Yibing<sup>1</sup>, PENG Shiliang<sup>3</sup>

1. Compressor Maintenance and Repair Company, PipeChina Pipeline Technology Development Co., Ltd.; 2. Hunan Branch of PipeChina;  
3. College of Mechanical and Transportation Engineering, China University of Petroleum (Beijing)

**Abstract:** [Objective] Predictive maintenance is recognized as one of the core approaches to improving the reliability and operational efficiency of centrifugal pumps. A crucial component of implementing this strategy lies in the health status recognition of these pumps. However, traditional methods for health status recognition primarily depend on machine learning techniques that require a substantial amount of labeled data. This dependency often leads to their limitations in effectively and clearly representing relationships between monitoring data and health status, thus constraining their application effect in real-world conditions, which are typically complex. Therefore, it is urgently needed to develop health status recognition methods that are more accurate, efficient, and adaptable. [Methods] This paper presents a health status recognition method for electrically driven centrifugal pumps, utilizing a cascade configuration of the Self-Organizing Map (SOM) neural network and the Back Propagation (BP) neural network. First, the SOM neural network was employed to preprocess the life-cycle vibration data of centrifugal pumps, focusing on extracting various statistical characteristics and entropy features in the time, frequency, and

time-frequency domains, to fully characterize their operating states. Next, the extracted bearing vibration signal features were subjected to dimensionality reduction and integration using Principal Component Analysis (PCA). This facilitates the reduction of redundant information and computational complexity, the optimization of the pattern of input parameters, and the enhancement of modeling efficiency. Finally, a health status recognition model was established for electrically driven centrifugal pumps based on cascaded SOM-BP neural networks, leveraging the advantages of both neural networks. **[Results]** Based on the health status monitoring dataset of an electrically driven centrifugal pump, a comparison was conducted between the SOM-BP model and common machine learning techniques, such as random forest and XGBoost models, to evaluate their accuracy in health status recognition of the pump, using  $R^2$ , MSE, and RMSE as evaluation metrics. The results indicated that the  $R^2$ , MSE, and RMSE values generated by the proposed model were 0.901,  $0.8 \times 10^{-6} \text{ m}^2/\text{s}^4$ , and  $9.12 \times 10^{-4} \text{ m/s}^2$ , respectively. These results were significantly superior to those obtained from traditional machine learning methods, demonstrating enhanced robustness and adaptability. **[Conclusion]** The proposed method has demonstrated effectiveness in enhancing the accuracy of health status recognition for centrifugal pumps. Additionally, it provides valuable data support for fault diagnosis and residual life prediction of these pumps. The results of this study present a novel perspective on health status management and diagnosis that can be applied to other rotating machinery. (5 Figures, 2 Tables, 25 References)

**Key words:** Self-Organizing Map (SOM), Back Propagation (BP) neural network, electrically driven centrifugal pump, health status recognition

随着工业技术的快速发展, 离心泵作为石油化工、能源输送等行业的关键工业设备, 其运行状态直接关系到整个生产系统的安全性与可靠性。如何准确识别离心泵的健康状态, 及时发现潜在故障, 提升设备运行效率与可靠性, 成为现代工业维护的核心课题<sup>[1-2]</sup>。目前工业领域普遍采用定时维护策略, 尽管这种方式能够在一定程度上减少故障发生的风险, 但其固定的维护周期往往无法针对设备的实际运行状态进行动态调整, 导致过度维护或维护不足, 致使浪费资源的同时还难以及时发现潜在隐患<sup>[3]</sup>。此外, 定时维护过程中产生的大量监测数据缺乏有效的管理与分析手段, 使设备健康状态难以量化, 进一步影响了维护决策的科学性与效率。

近年来, 学者们提出多种离心泵健康状态识别方法, 涵盖传统的振动分析、频谱分析、健康因子评估及基于机器学习的智能算法等<sup>[4-8]</sup>。早期的振动分析、频谱分析是通过对振动信号时域、频域特征提取, 初步建立了故障诊断体系, 但在适应早期退化状态与多源数据融合方面存在不足<sup>[9]</sup>。健康因子评估方法采用层次分析法、T-S(Takagi-Sugeno)模糊故障树等技术综合评估设备状态<sup>[10]</sup>, 虽在一定程度上提供了预测依据, 但对复杂故障的敏感性仍不足。近年来, 基于机器学习的模型(卷积神经网络、长短期记忆网络等)展现了自动特征提取的潜力, 但其对标记数据的高依赖性与计算复杂度限制了其在实际工况下的广泛应用<sup>[5, 11]</sup>。

自组织映射(Self Organization Map, SOM)神经网络与 BP(Back Propagation)神经网络因其各自的优势逐渐被引入设备健康状态识别领域。SOM 神经网络在特征聚类与降维方面表现优越<sup>[12]</sup>, 能够从高维数据中提取关键特征, 但该算法无监督学习的特点, 因此无法直接完成分类; BP 神经网络具备强大的监督分类能力<sup>[13-15]</sup>, 能有效实现输入与输出的非线性映射, 但易陷入局部最优, 且训练效率较低。将 SOM 神经网络与 BP 神经网络结合起来的 SOM-BP 级联神经网络模型近年来在光伏逆变器、锂电池健康状态评估等领域应用取得良好效果<sup>[16-19]</sup>, 通过融合 SOM 的特征提取能力与 BP 的分类能力, 可显著提升诊断的准确性与鲁棒性。为此, 提出一种基于 SOM-BP 级联神经网络的电驱离心泵健康状态识别方法, 可高效处理离心泵的多种复杂实际工况与多源监测数据, 以期为中心泵的故障诊断与剩余寿命预测提供支持, 同时为其他旋转机械设备的健康状态管理与诊断提供新思路。

## 1 电驱离心泵健康状态识别

### 1.1 离心泵振动信号时间序列分析

为提高电驱离心泵运行的安全性, 最大限度地减少停机时间, 确保高效完成生产任务, 研究提出采用预测性维护方法进行离心泵健康状态识别<sup>[20]</sup>, 因此, 预计组件与系统的当前退化状态并预测未来的演变趋势

尤为必要。近年来,研究提出预测设备剩余使用寿命(设备将根据规格继续执行其功能的时间)的方法,即在故障诊断中引入健康状态指标,测量设备退化的程度。根据设备及其退化机制的不同,健康状态指标可以是直接测量的信号。通常情况下,健康状态指标的提取可通过预处理原始数据以降低噪声来实现,也可使用时域的统计指标(均值、标准差等)以及频域或时频域的其他指标来提取退化趋势。

为此,提出离心泵健康状态评价准则,以量化转动设备的退化情况。假设从一组设备中收集到 $N$ 条运行故障轨迹数据,其中有 $S$ 个信号可以直接或间接地与设备的运行状态、退化状态相关,其测量值是可用的。健康状态指标应该具有3个关键属性:趋势不可逆性、趋势性及趋势显著性。由于设备退化过程不可逆,这将导致设备的退化具有非单调趋势;同时设备退化过程中健康状态指标的演变应具有相同趋势,可用相同的函数形式来描述;最后,设备在故障发生时健康状态指标的特征都应相同。综上,为判别离心泵泵体退化状况,提出用于评估电驱离心泵健康状态的3个指标:趋势不可逆性、趋势性、趋势显著性。

趋势不可逆性指标 $F_1$ 的具体表达式<sup>[21]</sup>为:

$$F_1 = \frac{1}{N} \left| \sum_{i=1}^N M_i \right| \quad (1)$$

$$M_i = \frac{n_i^+}{n_i - 1} - \frac{n_i^-}{n_i - 1}, \quad i = 1, 2, \dots, N \quad (2)$$

式中: $M_i$ 为电驱离心泵单次运行到故障轨迹的单调性; $N$ 为电驱离心泵所有故障轨迹的总次数; $n_i$ 为电驱离

心泵第 $i$ 次运行到故障轨迹中的观测总数; $n_i^+$ 、 $n_i^-$ 分别为以正、负一阶导数为特征的观测数。

同时提出一个基于运行到失效轨迹中正一阶导数与正二阶导数分数对比的趋势指数。所提出的指数已被证明极易受到噪声的影响<sup>[22]</sup>,因此将趋势性指标 $F_2$ 定义为线性相关系数的最小值,其表达式为:

$$F_2 = (|f_{ij}|), \quad i = 1, 2, \dots, N; j = 1, 2, \dots, N \quad (3)$$

式中: $f_{ij}$ 为第 $i$ 次与第 $j$ 次运行至故障轨迹之间的线性相关系数。

最后,趋势显著性指标 $F_3$ <sup>[22]</sup>定义为可用运行到故障轨迹失效时健康状态指标值的标准差除以轨迹开始与结束时健康状态指标值的平均变化,将得到的值进行指数加权,以给出所需的0到1比例的度量。 $F_3$ 的具体表达式为:

$$F_3 = \frac{\text{std}(H_{\text{fail}}) - \text{mean}(H_{\text{start}} - H_{\text{fail}})}{|H_{\text{start}} - H_{\text{fail}}|} \quad (4)$$

式中: $H_{\text{start}}$ 、 $H_{\text{fail}}$ 分别为运行到故障轨迹开始、结束时的健康状态指标值。

健康状态评价准则指标集 $F$ 的表达式为:

$$F = [F_1, F_2, F_3] \quad (5)$$

相比电驱离心设备的工作参数评价方法<sup>[1]</sup>,评价准则指标集 $F$ 可从趋势不可逆性、趋势性、趋势显著性3个方面综合判别离心泵泵体退化状况,并利用基于时间序列数据分析技术<sup>[8,23]</sup>的9项能够精细化反映离心泵转动轴承系统健康表征数据发生劣化的数据特征值(表1),定性判定电驱离心设备在异常工况下指标体系的变化规律。

表1 能够精细化反映离心泵转动轴承系统健康表征数据发生劣化的数据特征值表  
Table 1 Eigenvalues for accurate representation of the health characterization data degradation for centrifugal pump bearing system

| 特征值    | 含义                         | 作用                           |
|--------|----------------------------|------------------------------|
| 转频     | 计算时间序列数据的频率分量              | 通过傅里叶变换或小波变换等频域分析方法提取转频      |
| 特征频率   | 识别时间序列数据中的主要频率成分           | 使用傅里叶变换或小波变换等方法,通过分析频谱确定特征频率 |
| 倍频     | 检测时间序列数据中的倍频成分,即频率的整数倍     | 通过寻找频谱峰值的倍数来提取倍频             |
| 有效值    | 计算时间序列数据的RMS值              | 表示信号的整体能量                    |
| 峭度值    | 计算时间序列数据的峭度                | 描述信号的尖锐程度或峰值的陡峭程度            |
| 脉冲值    | 检测时间序列数据中的脉冲信号,即短时突发的高幅值信号 | 捕捉设备运行中的突发性事件与高幅值冲击          |
| 叶轮通过频率 | 识别时间序列数据中叶轮通过的频率成分         | 通过频谱分析或滤波方法来提取该特征            |
| 裕度因子   | 计算时间序列数据的裕度因子              | 描述信号的动态范围与幅值分布情况             |
| 绝对均值   | 计算时间序列数据的绝对均值              | 衡量信号的整体幅值水平                  |



## 1.2 离心泵健康状态指标构建

主成分分析(Principal Component Analysis, PCA)法是一种广泛应用的数学统计方法,主要用于数据降维,同时能够保留数据集中最显著的变异特征。通过 PCA 计算,原始数据中的变量被转换为一组线性不相关的变量,即主成分,这些主成分根据其自身对原始数据集方差的解释能力进行排序。PCA 计算可在减少数据维度的同时最大限度地降低信息损失,为此,采用 PCA 对已提取的轴承振动信号特征进行降维与融合,从而有效减少冗余信息与计算复杂度,优化输入参数的模式,提升建模效率。

首先,对  $n$  维输入数据进行线性变换,找取数据特征中心,标准化后的数据矩阵为  $\bar{X}$ ,计算得到协方差矩阵为  $P$ ,其表达式为:

$$P = \text{cov}(\bar{X}) = \frac{1}{n-1} \bar{X} \bar{X}^T \quad (6)$$

然后,计算协方差矩阵  $P$  的特征值  $\lambda_j$  以及特征向量  $\alpha_j$ ,其表达式为:

$$P = EDE^T \quad (7)$$

$$E = (\alpha_1, \alpha_2, \dots, \alpha_j) \quad (8)$$

式中:  $D$  为按特征值大小降序排列的对角矩阵;  $E$  为特征值  $\lambda_j$  对应特征向量  $\alpha_j$  的集合。

通过线性变换即可求得各主成分向量  $Y_1, Y_2, \dots, Y_n$ ,从而得到  $n$  个互不相关的主成分,每个主成分数据都是  $n$  维输入数据的线性组合,每个维度的主成分都属于时间序列数据,其表达式为:

$$\begin{cases} Y_1 = \beta_1^1 X_1^* + \dots + \beta_j^1 X_j^* + \dots + \beta_n^1 X_n^* \\ \vdots \\ Y_i = \beta_i^1 X_1^* + \dots + \beta_j^i X_j^* + \dots + \beta_n^i X_n^* \\ \vdots \\ Y_n = \beta_n^1 X_1^* + \dots + \beta_j^n X_j^* + \dots + \beta_n^n X_n^* \end{cases} \quad (9)$$

式中:  $Y_i$  为第  $i$  个主成分数据(输出数据);  $X_j^*$  为第  $j$  维标准化时间序列数据(输入数据);  $\beta_j^i$  为对应于第  $i$  个主成分、第  $j$  维输入数据的线性组合系数。

通过计算各主成分的方差,可以对其进行排序,形成第 1 主成分、第 2 主成分等。方差越大的主成分包含的数据有效特征信息越丰富,越能够更好地表征轴承的退化趋势。因此,采用 PCA 计算可将原始的多维特征数据融合为单维或二维数据,同时保留原始特征中的关键信息。

## 1.3 SOM-BP 级联神经网络算法

### 1.3.1 SOM 神经网络

SOM 神经网络是由 Kohonen<sup>[24]</sup>提出的一种无监督学习型网络,是典型的自组织神经网络之一。与有监督神经网络相比,SOM 神经网络更接近于人脑的认知过程。在未知信息类型的前提下,该网络通过学习样本间的关系与模式,自适应地调整网络的连接权值,从而实现对本样本的分类。与其他竞争型神经网络不同,SOM 神经网络由输入层与竞争层组成,不仅能够对输入向量进行分类,还能在竞争层直观地呈现输入向量的分布情况,其显著特点是竞争层中神经元之间具有相互联系,分类过程由多个神经元协作完成,从而有效克服了单一神经元决定分类结果的局限性。SOM 神经网络的学习过程是无监督的,无需为样本数据设置标签。在接收到样本数据后,会将数据划分为不同区域,每个区域内的神经元通过相互作用自适应地对输入做出响应,从而完成输入向量的分类。

SOM 神经网络(图 1)的训练过程相对直观。当输入向量被随机输入到网络时,竞争层会计算该输入向量与每个神经元之间的欧式距离。通过对比所有神经元的欧式距离,确定当前输入向量的获胜神经元。随后,更新获胜神经元及其邻域内神经元的连接权值,而邻域外神经元的连接权值保持不变。经过多次迭代,输入向量与神经元之间的欧式距离逐渐减小,网络最终在满足“达到最大迭代次数或误差小于阈值”终止条件时收敛。最终,与同一获胜神经元相对应的输入向量被划分为同一类。此时,训练过程完成,输入向量

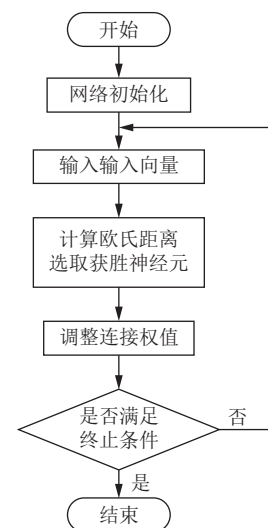


图 1 SOM 神经网络算法流程图  
Fig. 1 Flow chart of SOM neural network algorithm

与竞争层神经元之间的关系已确定。当新的测试向量输入时, SOM 神经网络会将其归类到最接近的神经元对应的类别中。

SOM 神经网络的优点在于无需样本标签, 不依赖于大量的训练样本, 同时其结构简单, 算法易于实现, 但其分类效果会受到竞争层节点数量与拓扑结构的影响。此外, 其局限性主要体现在以下几个方面: ①在训练之前, 需要预先确定竞争层神经元的数量与拓扑结构。这两个超参数在训练过程中无法调整, 目前只能通过大量实验来确定, 缺乏理论指导。②在训练过程中, 由于初始化阶段某些神经元与所有样本的距离过远, 导致这些神经元可能始终无法获胜或参与优胜邻域, 致使权值未能更新。同时, 部分神经元可能因被过度使用而影响网络的整体分类效果。③如果在训练过程中需要添加新的类别, 则必须重新训练整个网络。④虽然 SOM 神经网络不需要大量的训练样本, 但在样本数量过少的情况下, 分类效果可能会受到限制。⑤如果算法随机初始化的连接权值选择不当, 可能会导致网络收敛时间延长, 甚至无法收敛。

### 1.3.2 BP 神经网络基本原理

BP 神经网络是由 Rumelhart 等<sup>[25]</sup>于 1986 年提出的一种基于误差反向传播算法训练的多层前馈网络, 是典型的监督学习神经网络之一。常见的 BP 神经网络一般分为输入层、隐含层、输出层 3 层, 其学习过程包括两个阶段: ①信息的正向传播, 将输入层的训练样本经过传递函数处理后传递至输出层, 由输出层生成结果并计算误差。②训练误差的反向传播, 将训练误差反馈至各个神经元, 按照误差大小分配至网络中的每个神经元, 各神经元依据误差调整其连接权值与阈值。当训练次数达到预设值或训练误差满足目标要求时, 网络训练完成并输出结果。

BP 神经网络算法(图 2)的训练过程也相对清晰。当训练样本输入到网络后, 信息首先从输入层逐层传递至隐含层与输出层, 并在输出层产生预测结果。随后, 将该预测结果与实际目标值进行对比, 计算得到误差。该误差通过反向传播算法逐层返回到各个神经元, 并用以调整网络各层之间的连接权值与阈值, 使网络的输出逐渐接近真实目标值。经过多次迭代, 训练误差不断减小, 网络最终在满足“达到最大迭代次数或误差小于阈值”终止条件时收敛, 此时网络完成对输入与输出关系的学习。当新的测试样本输入时,

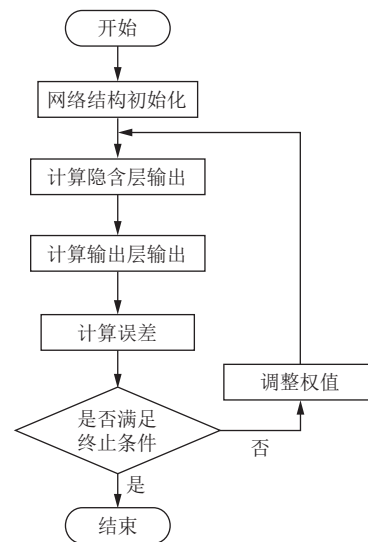


图 2 BP 神经网络算法流程图  
Fig. 2 Flow chart of BP neural network algorithm

BP 神经网络便会利用已训练好的权值与阈值, 对该样本进行前向传播并输出预测结果, 从而实现分类、回归等任务。

大部分学者研究神经网络模型都以 BP 神经网络为基础, 作为严谨的数学理论支撑。然而, BP 神经网络也存在以下 4 点局限性: ①在训练过程中, BP 神经网络通过逐步调整连接权值与阈值来改善性能, 但这种优化往往朝着局部改善的方向进行, 容易陷入局部极小值。②学习率的设定没有明确的理论依据, 通常需要通过实验结果进行比较来选择合适的参数。此外, 隐含层神经元的数量没有固定的设置方法, 只能在经验公式的基础上通过多次尝试进行调整。③如果训练样本过于冗杂、不具有代表性, 或者样本数量不足, 都会对网络的性能产生负面影响。④由于算法的初始权值与阈值是随机生成的, 每次训练的结果可能会有所差异。

### 1.3.3 SOM-BP 级联神经网络

将 SOM 神经网络的输出作为 BP 神经网络的输入, 即为 SOM-BP 级联神经网络<sup>[19]</sup>, 其基本思想是将无监督学习网络与有监督学习网络串联, 构建一个新的混合网络。通过 SOM 神经网络的泛化能力, 可以有效克服容差因素对分类的影响, 因此无需大量的训练数据; 结合 BP 神经网络的有监督学习, 则能够完成输入到输出的非线性映射。

SOM-BP 级联神经网络<sup>[16]</sup>(图 3)包括输入层、竞争层、隐含层、输出层 4 层, 其中  $R$ 、 $S_1$ 、 $S_2$ 、 $S_3$  分别为输入层、竞争层、隐含层、输出层的节点数;  $f_0$ 、 $f_1$ 、

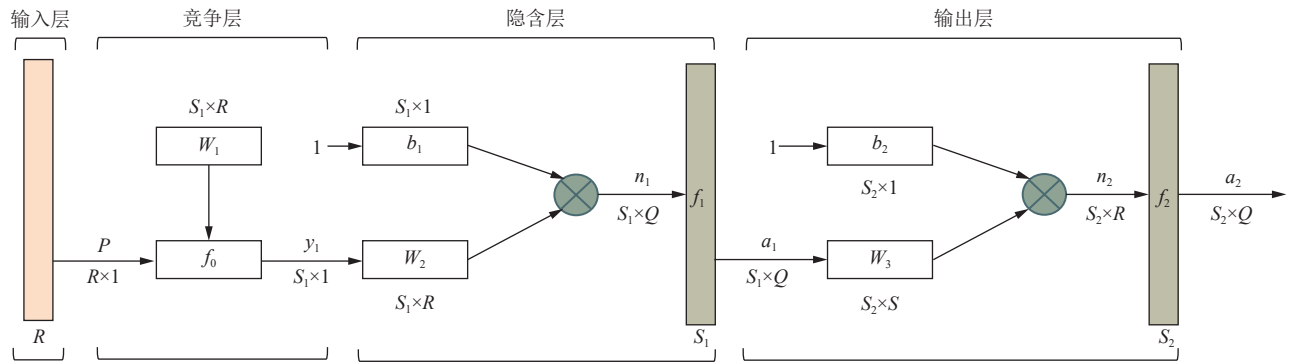


图3 SOM-BP 级联神经网络结构图

Fig. 3 Cascade structure of SOM-BP neural networks

$f_1$  分别为竞争层、隐含层、输出层的传递函数;  $W_1$ 、 $W_2$ 、 $W_3$  分别为竞争层、隐含层、输出层的连接权值;  $b_1$ 、 $b_2$  分别为隐含层、输出层的阈值;  $y_1$  为竞争层神经元的输出;  $a_1$ 、 $a_2$  分别为隐含层、输出层神经元的加权输入;  $n_1$ 、 $n_2$  分别为隐含层、输出层神经元的数量;  $Q$  为神经元输出的维度。

在 SOM-BP 级联神经网络的训练过程中, 首先通过输入量  $x$ 、权重矩阵  $W$  进行归一化处理, 得到  $P_q$ 、 $W_j$ , 然后计算两者的欧氏距离来确定最接近的获胜神经元, 欧氏距离  $d$  的计算表达式为:

$$d = \min \|P_q - W_j\| = \min \sqrt{2(1 - W_j^T P_q)} \quad (10)$$

式中:  $P_q$  为第  $q$  个输入样本的特征向量;  $W_j$  为第  $j$  个神经元的权重向量。

欧氏距离最小的神经元胜出, 获胜神经元可调整其权值,  $W_j$  通过学习率逐渐接近最佳值, 直到误差减

小至 0 或达到设定值, 训练过程完成。

将 SOM 训练得到的神经元结果作为 BP 神经网络的输入, 训练样本为  $(v_q, t_q)$ 。训练过程包括计算当前接收到的输入向量  $v_q$ 、 $W_j$  的输出, 得到相应的神经网络输出  $a_q^2$ , 其表达式为:

$$e = \frac{1}{2} \sum_{q=1}^M (t_q - a_q^2)^2 \quad (11)$$

式中:  $e$  为神经网络输出与目标输出之间的误差;  $M$  为训练样本的总数;  $t_q$  为第  $q$  个样本的目标输出;  $a_q^2$  为第  $q$  个样本的 BP 神经网络输出, 即网络根据输入进行计算后的输出值。

根据误差反馈调整神经元的连接权重与偏置, 直到训练误差达到预设阈值或达到最大训练轮数为止。

基于 SOM-BP 级联神经网络的离心泵健康状态识别(图 4, 其中  $t_1$ 、 $T_1$  分别为第 1 阶段 SOM 神经

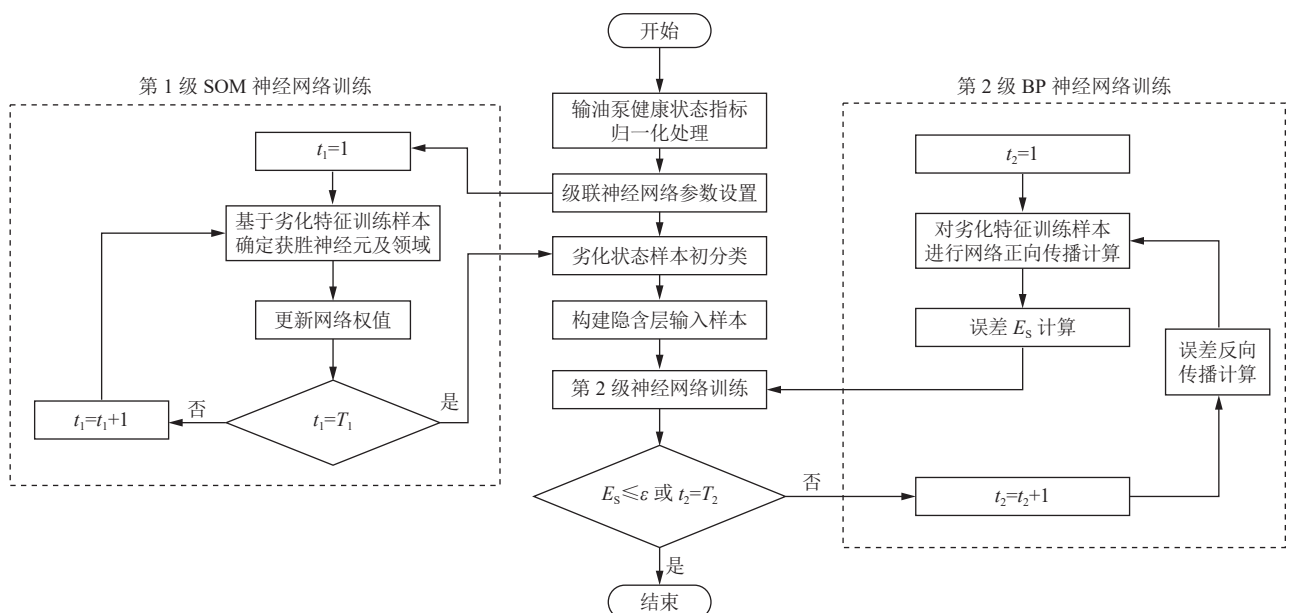


图4 基于 SOM-BP 级联神经网络的离心泵健康状态识别流程图

Fig. 4 Flow chart of health status recognition for centrifugal pump based on cascaded SOM-BP neural networks

网络训练中的迭代次数、最大迭代次数;  $t_2$ 、 $T_2$  分别为第2阶段BP神经网络训练中的迭代次数、最大迭代次数;  $E_s$  为第2阶段BP神经网络的误差;  $\varepsilon$  为误差阈值)的具体算法步骤为: ①第1阶段SOM神经网络训练, 使样本在无监督的情况下进行特征提取与初步分类。先将输入的健康状态指标数据进行归一化处理, 后设置SOM网络的参数, 并将劣化状态样本进行初步分类, 再根据样本数据更新SOM网络的权值。此阶段通过计算欧氏距离来确定获胜神经元, 并进行相应的权值调整, 直到训练次数达到预定的最大次数  $T_1$  或满足误差要求时训练完成。②第2阶段BP神经网络训练, 将第1阶段SOM神经网络训练的结果作为BP神经网络的输入样本, 用于准确识别离心泵的健康状态。BP神经网络先进行正向传播计算, 后根据输出与目标值计算误差, 并通过反向传播更新权值, 逐步优化网络, 直到误差  $E_s$  小于设定的阈值  $\varepsilon$  或达到最大训练次数  $T_2$  时训练完成。

#### 1.3.4 误差评价指标

预测模型的性能往往需要误差评价指标来对其进行评价, 选取决定系数  $R^2$ 、均方误差(Mean Squared Error, MSE)、均方根误差(Root Mean Square Error, RMSE)作为模型的评价指标。 $R^2$  用于评估神经网络模型的拟合程度, 取值范围为  $[0, 1]$ , 当  $R^2=1$  时, 表示模型完美地拟合了数据, 即预测值与实际观测值完全一致; 当  $R^2=0$  时, 表示模型没有捕捉到任何有用的信息, 模型的预测效果与简单地将所有观测值设为平均值的效果相同, 且  $R^2$  越大, 说明模型对数据的拟合程度越好, 预测效果越准确。MSE、RMSE 可衡量模型预测结果与真实结果之间的差异, MSE 值越小, 表示模型预测效果越好, 模型的泛化能力越强; RMSE 值越小, 说明模型的预测越准确。

## 2 实例应用

首先自行搭建单级电驱离心泵故障模拟试验平台, 主要由电机、单级悬臂式离心泵、蓄水罐、电控箱、试验台底、管道等附件组成。故障模拟试验工作介质为循环水, 所用离心泵额定转速为 1 485 r/min。在离心泵机组壳体上共布置了 4 个测点, 用于测量振动加速度。共进行正常运行、转子不平衡故障、轴承

内圈故障、轴承外圈故障、轴承滚动体故障、轴承松动故障、碰摩故障 7 组试验, 前 6 组试验在转子转速分别为 743 r/min、1 485 r/min 工况下进行, 第 7 组试验仅在转子转速为 1 485 r/min 工况下进行。每组试验 4 个测点分批次进行测量, 加速度传感器的采样率为 32 768 Hz, 其中每个测点收集 400 组样本, 波形长度为 1 024, 由此得到的每个文件内部形式为一个  $400 \times 1\,024$  的矩阵。

采用无迹卡尔曼滤波对收集到的单级电驱离心泵原始故障监测信号进行滤波预处理, 减缓原始信号中存在的噪声信息。然后, 对原始故障监测信号进行归一化处理, 消除数据的大小对模型效果的影响。最后, 将 80% 的数据作为训练集, 10% 的数据作为验证集, 10% 的数据作为测试集, 开展后续的模型对比实验。基于 Pytorch 深度学习框架搭建模型, 实验环境为 RTX GeoForce 2080ti、Pytorch2.0.1、Python3.9。所有模型的超参数都在验证集上进行选优, 以保证模型达到最优的效果。为验证所提出的 SOM-BP 级联神经网络模型在离心泵健康状态识别中的性能和准确性, 对比随机森林、XG-boost 及 SOM-BP 级联神经网络模型在相同数据集上的表现, 并通过  $R^2$ 、MSE、RMSE 评价指标进行模型性能的综合评估(表 2)。可见, SOM-BP 级联神经网络模型的  $R^2$  最接近 1, 且其 MSE、RMSE 为 3 种模型中最小。

表 2 离心泵健康状态识别 3 种模型预测振动加速度评价结果对比表

Table 2 Comparison of evaluation results for vibration acceleration predictions of centrifugal pump across three health status recognition models

| 模型            | $R^2$ | MSE/( $\text{m}^2 \cdot \text{s}^{-4}$ ) | RMSE/( $\text{m} \cdot \text{s}^{-2}$ ) |
|---------------|-------|------------------------------------------|-----------------------------------------|
| 随机森林          | 0.758 | $7.1 \times 10^{-6}$                     | $2.6189 \times 10^{-3}$                 |
| XG-boost      | 0.831 | $4.5 \times 10^{-6}$                     | $1.5792 \times 10^{-3}$                 |
| SOM-BP 级联神经网络 | 0.901 | $0.8 \times 10^{-6}$                     | $9.1200 \times 10^{-4}$                 |

根据 3 种模型对离心泵健康状态识别预测结果热力图(图 5)可知: 随机森林模型的热力图显示了该模型能够较为清晰地识别离心泵健康状态的变化, 特别是在较为明显的故障模式上, 模型预测结果与实际监测数据具有较高的一致性。然而, 在一些复杂或微弱的故障信号上, 随机森林模型的预测准确度有所下降。热力图中的一些“过渡区域”表明, 随机森林模型在处理较为复杂的时空依赖性时, 可能存在一定的预测误差, 尤其是在数据变化较为复杂的区域, 预测误



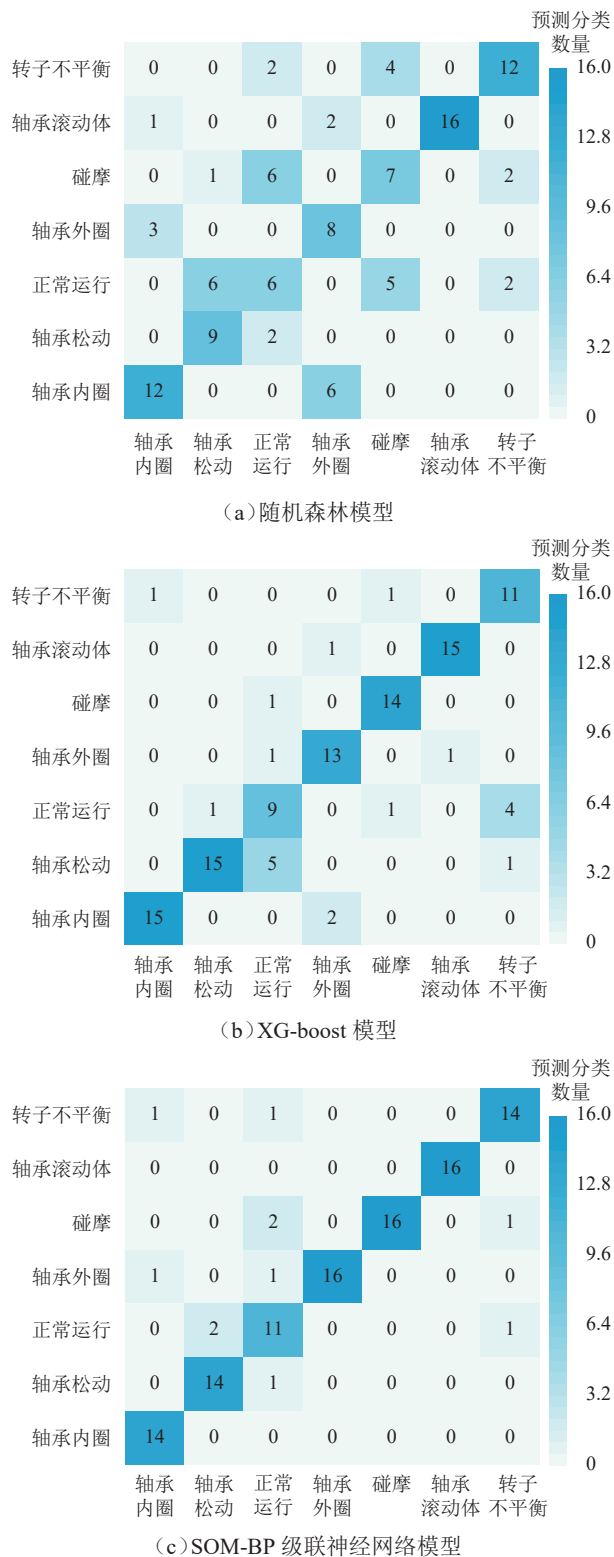


图 5 3 种模型对离心泵健康状态识别预测结果热力图  
Fig. 5 Heat maps of prediction results for health status recognition of centrifugal pump by three models

差有所增大(图 5a)。XG-Boost 模型通过优化的梯度提升算法,在大多数情况下能很好地拟合数据,并准确预测离心泵健康状态。相比随机森林模型,XG-Boost 模型在训练与预测时表现出更强的稳定性及较

高的精度。然而,当数据具有较强的时序依赖关系与复杂的空间结构时,XG-Boost 模型的性能受到一定限制,未能充分捕捉到电驱离心泵健康状态中的长期依赖性(图 5b)。SOM-BP 模型表现出最优异的预测精度,不仅能够准确捕捉到电驱离心泵健康状态的变化趋势,而且对复杂的时空特征具有较强的适应性与高效性,因此在多个评价指标上超越了随机森林与 XG-Boost 算法。SOM-BP 模型中的 SOM 算法能够有效地将高维数据映射到低维空间,发现数据中的潜在结构,从而提高模型的泛化能力。BP 算法则在此基础上进行精细化的训练,进一步提高了模型的预测精度(图 5c)。综上,3 种模型都能识别离心泵健康状态,但 SOM-BP 级联神经网络模型预测的准确度更高,在电驱离心泵健康状态识别方面具有显著优势,能够为实际应用提供更加可靠的预测结果。

### 3 结论

1) 基于 SOM-BP 级联神经网络的电驱离心泵健康状态识别方法能够有效地结合无监督学习与监督学习的优势,实现高维特征数据的降维与精确识别。相比传统的随机森林、XG-Boost 模型,该方法在离心泵健康状态识别的准确性上表现优越,其  $R^2$  值接近 1, MSE 值、RMSE 值均最低,表明 SOM-BP 级联神经网络模型在处理复杂时空特征与捕捉设备退化趋势方面具有更高的适应性与鲁棒性,可为电驱离心泵的智能化健康管理与预测性维护提供有效支持。同时,该方法通过 PCA 算法优化了数据特征,提高了模型训练效率与特征融合效果,为工业设备故障诊断领域提供了新的技术思路,具有广泛的实际应用潜力。

2) SOM 神经网络的训练依赖于预设的竞争层神经元数量与拓扑结构,这些超参数无法在训练过程中动态调整,且缺乏理论指导,依赖大量实验进行选择,导致训练效率与效果可能受限。虽然采用 PCA 降维提高了计算效率,但在处理大规模高维数据时,计算开销仍然较大,训练时间较长。此外, SOM 神经网络对噪声数据较为敏感,且初始化过程可能导致收敛慢或无法收敛。因此,未来的研究可以考虑引入自动化超参数优化技术,如遗传算法或贝叶斯优化,来动态调整超参数;同时,探索混合模型优化,使用深度学习、循环神经网络等技术,以提高对复杂故障模式的识别



能力。此外,应增加模型对异常数据的鲁棒性,并优化模型以适应实时在线监测,从而进一步提升该方法在工业设备健康管理中的应用潜力。

### 参考文献:

- [1] 谢小青,孙卫平.预测性维护:离心泵维护方式的未来[J].水泵技术,2021(1):1-5.
- XIE X Q, SUN W P. Predictive maintenance: the future of centrifugal pump maintenance[J]. Pump Technology, 2021(1): 1-5.
- [2] 刘志伟,刘锐,徐劲松,李毅,周黎明.复杂系统故障预测与健康管理的(PHM)技术研究[J].计算机测量与控制,2010,18(12):2687-2689,2751. DOI: [10.16526/j.cnki.11-4762/tp.2010.12.001](https://doi.org/10.16526/j.cnki.11-4762/tp.2010.12.001).
- LIU Z W, LIU R, XU J S, LI Y, ZHOU L M. Research of complex system's prognostic and health management[J]. Computer Measurement & Control, 2010, 18(12): 2687-2689, 2751.
- [3] 彭宇,刘大同.数据驱动故障预测和健康管理综述[J].仪器仪表学报,2014,35(3):481-495. DOI: [10.19650/j.cnki.cjsi.2014.03.001](https://doi.org/10.19650/j.cnki.cjsi.2014.03.001).
- PENG Y, LIU D T. Data-driven prognostics and health management: a review of recent advances[J]. Chinese Journal of Scientific Instrument, 2014, 35(3): 481-495.
- [4] 刘强.基于数字孪生技术的离心泵叶轮机械故障诊断方法研究与应用[D].兰州:兰州理工大学,2023.
- LIU Q. Research and application of fault diagnosis method for centrifugal pump impeller based on digital twinning technology[D]. Lanzhou: Lanzhou University of Technology, 2023.
- [5] 王前江.基于深度学习的离心泵故障诊断方法研究[D].郑州:郑州大学,2021.
- WANG Q J. Research on fault diagnosis method of centrifugal pump based on deep learning[D]. Zhengzhou: Zhengzhou University, 2021.
- [6] 高帆,李洪元,吴帆.基于频谱分析的离心泵健康状态监测及故障诊断[J].自动化仪表,2019,40(7):24-28. DOI: [10.16086/j.cnki.issn1000-0380.2019030330](https://doi.org/10.16086/j.cnki.issn1000-0380.2019030330).
- GAO F, LI H Y, WU F. Health status monitoring and fault diagnosis based on spectrum analysis for centrifugal pump[J]. Process Automation Instrumentation, 2019, 40(7): 24-28.
- [7] 乔佳伟,田慕琴.基于AHP-TOPSIS综合评价法的离心泵健康状态评估[J].工矿自动化,2022,48(9):69-76. DOI: [10.13272/j.issn.1671-251x.17984](https://doi.org/10.13272/j.issn.1671-251x.17984).
- QIAO J W, TIAN M Q. Health condition assessment of centrifugal pump based on AHP-TOPSIS comprehensive evaluation method[J]. Journal of Mine Automation, 2022, 48(9): 69-76.
- [8] 王前江,陈磊,徐向阳,岳树旭.基于KECA的离心泵健康状态评价方法[J].组合机床与自动化加工技术,2021(7):181-184. DOI: [10.13462/j.cnki.mmtamt.2021.07.042](https://doi.org/10.13462/j.cnki.mmtamt.2021.07.042).
- WANG Q J, CHEN L, XU X Y, YUE SHUAI X. Evaluation method of centrifugal pump health state based on KECA[J]. Modular Machine Tool & Automatic Manufacturing Technique, 2021(7): 181-184.
- [9] 侯继洁.基于故障树和剩余使用寿命的设备健康状态评估技术研究[D].北京:北京化工大学,2024.
- HOU J J. Research on equipment health assessment techniques based on fault trees and remaining useful life[D]. Beijing: Beijing University of Chemical Technology, 2024.
- [10] 王语嘉,李金波,侯继洁,马波.基于T-S模糊故障树和剩余寿命的离心泵健康评估方法[J].机电工程,2024,41(12):2272-2281. DOI: [10.3969/j.issn.1001-4551.2024.12.015](https://doi.org/10.3969/j.issn.1001-4551.2024.12.015).
- WANG Y J, LI J B, HOU J J, MA B. Health assessment method of centrifugal pump based on T-S fuzzy fault tree and remaining useful life[J]. Journal of Mechanical & Electrical Engineering, 2024, 41(12): 2272-2281.
- [11] 江璐鑫.基于机器学习的输油管道系统典型运行工况识别方法研究[D].北京:中国石油大学(北京),2023.
- JIANG L X. Research on the recognition method of typical operation condition of oil pipelines based on machine learning[D]. Beijing: China University of Petroleum (Beijing), 2023.
- [12] 崔恒志,韦磊,马洲俊,张海强,朱红,李建岐.基于电气距离和SOM-BP的配电网拓扑识别研究[J].自动化技术与应用,2021,40(9):146-151. DOI: [10.3969/j.issn.1003-7241.2021.09.033](https://doi.org/10.3969/j.issn.1003-7241.2021.09.033).
- CUI H Z, WEI L, MA Z J, ZHANG H Q, ZHU H, LI J Q. Research on distribution network topology recognition based on electrical distance and SOM-BP[J]. Techniques of Automation and Applications, 2021, 40(9): 146-151.
- [13] 廖睿.基于GA-BP神经网络的城市供水管网健康状态评估[D].蚌埠:安徽财经大学,2021.
- LIAO R. Health assessment of urban water supply network based on GA-BP neural network[D]. Bengbu: Anhui University of Finance and Economics, 2021.
- [14] 吴青峰,杨艺涛,刘立群,胡秀芳,薄利明,杨杰豹.基于GA-SA-BP神经网络的锂电池健康状态估算方法[J].电力系统保护与控制,2024,52(19):74-84. DOI: [10.19783/j.cnki.pspc.240248](https://doi.org/10.19783/j.cnki.pspc.240248).

- WU Q F, YANG Y T, LIU L Q, HU X F, BO L M, YANG J B. Lithium battery state of health estimation method based on a GA-SA-BP neural network[J]. Power System Protection and Control, 2024, 52(19): 74–84.
- [15] 徐元中, 曹翰林, 吴铁洲. 基于 SA-BP 神经网络算法的电池 SOH 预测[J]. 电源技术, 2020, 44(3): 341–345. DOI: [10.3969/j.issn.1002-087X.2020.03.010](https://doi.org/10.3969/j.issn.1002-087X.2020.03.010).
- XU Y Z, CAO H L, WU T Z. Estimation of SOH for battery based on SA-BP neural network[J]. Chinese Journal of Power Sources, 2020, 44(3): 341–345.
- [16] 张书婷, 朱珠. 基于 SOM-BP 神经网络的光伏逆变器软故障诊断研究[J]. 中国新技术新产品, 2024(13): 22–24. DOI: [10.3969/j.issn.1673-9957.2024.13.008](https://doi.org/10.3969/j.issn.1673-9957.2024.13.008).
- ZHANG S T, ZHU Z. Research on soft fault diagnosis of photovoltaic inverters based on SOM-BP neural network[J]. New Technologies and New Products of China, 2024(13): 22–24.
- [17] 黄玮. 基于 FA 优化 SOM-BP 神经网络的混合动力汽车故障诊断[D]. 合肥: 合肥工业大学, 2020.
- HUANG W. Fault diagnosis of hybrid vehicles based on FA optimized SOM-BP neural network[D]. Hefei: Hefei University of Technology, 2020.
- [18] 张振坤. 基于 SOM-BP 云神经网络的计算机网络对抗效能评估[J]. 舰船电子对抗, 2021, 44(2): 61–63, 87. DOI: [10.16426/j.cnki.jcdzdk.2021.02.013](https://doi.org/10.16426/j.cnki.jcdzdk.2021.02.013).
- ZHANG Z K. Effectiveness evaluation of computer network countermeasure based on SOM-BP cloud neural network[J]. Shipboard Electronic Countermeasure, 2021, 44(2): 61–63, 87.
- [19] 高利民, 许庆阳, 李锋, 杨吉, 孟景辉, 杨树忠. 基于 SOM-BP 混合神经网络的道岔设备退化状态研究[J]. 中国铁道科学, 2020, 41(3): 50–58. DOI: [10.3969/j.issn.1001-4632.2020.03.06](https://doi.org/10.3969/j.issn.1001-4632.2020.03.06).
- GAO L M, XU Q Y, LI F, YANG J, MENG J H, YANG S Z. Research on degradation state of turnout equipment based on SOM-BP hybrid neural network[J]. China Railway Science, 2020, 41(3): 50–58.
- [20] 陆宁云, 陈闯, 姜斌, 邢尹. 复杂系统维护策略最新研究进展: 从视情维护到预测性维护[J]. 自动化学报, 2021, 47(1): 1–17. DOI: [10.16383/j.aas.c200227](https://doi.org/10.16383/j.aas.c200227).
- LU N Y, CHEN C, JIANG B, XING Y. Latest progress on maintenance strategy of complex system: from condition-based maintenance to predictive maintenance[J]. Acta Automatica Sinica, 2021, 47(1): 1–17.
- [21] 刘军, 安柏任, 张维博, 甘乾煜. 大型风力发电机组健康状态评价综述[J]. 电力系统保护与控制, 2023, 51(1): 176–187. DOI: [10.19783/j.cnki.pspc.220373](https://doi.org/10.19783/j.cnki.pspc.220373).
- LIU J, AN B R, ZHANG W B, GAN Q Y. Review of health status evaluation of large wind turbines[J]. Power System Protection and Control, 2023, 51(1): 176–187.
- [22] 董玉亮, 李亚琼, 曹海斌, 何成兵, 顾煜炯. 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, 33(11): 88–95, 15. DOI: [10.13334/j.0258-8013.pcsee.2013.11.016](https://doi.org/10.13334/j.0258-8013.pcsee.2013.11.016).
- DONG Y L, LI Y Q, CAO H B, HE C B, GU Y J. Real-time health condition evaluation on wind turbines based on operational condition recognition[J]. Proceedings of the CSEE, 2013, 33(11): 88–95, 15.
- [23] WANG Q F, LIU X J, WEI B K, CHEN W W. Online incipient fault detection method based on improved  $\ell_1$  trend filtering and support vector data description[J]. IEEE Access, 2021, 9: 30043–30059. DOI: [10.1109/ACCESS.2021.3058907](https://doi.org/10.1109/ACCESS.2021.3058907).
- [24] KOHONEN T. The self-organizing map[J]. Proceedings of the IEEE, 1990, 78(9): 1464–1480. DOI: [10.1109/5.58325](https://doi.org/10.1109/5.58325).
- [25] RUMELHART D E, HINTON G E, WILLIAMS R J. Learning representations by back-propagating errors[J]. Nature, 1986, 323(6088): 533–536. DOI: [10.1038/323533a0](https://doi.org/10.1038/323533a0).
- (编辑: 张雪琴)
- 
- 基金项目:** 国家自然科学基金青年项目“智能化设备健康状态评估与故障诊断方法研究”, 52401412。
- 作者简介:** 江新星, 男, 1989 年生, 工程师, 2014 年硕士毕业于中国石油大学(北京)安全工程专业, 现主要从事压缩机、泵机组故障诊断等相关研究。地址: 河北省廊坊市广阳经济技术开发区畅祥道 6 号, 065000。电话: 15652755658。Email: [532554117@qq.com](mailto:532554117@qq.com)
- Received: 2024-12-12
  - Revised: 2025-01-12
  - Online: 2025-02-12

