

基于 KECA-PLS 的风电机组健康状态评估

马 亮¹, 任 超^{2,3}, 刘月文^{2,3}, 郝增孝²

(1. 大唐(赤峰)新能源有限公司, 内蒙古 赤峰 024000;

2. 内蒙古工业大学, 呼和浩特 010051;

3. 内蒙古自治区高等学校 智慧能源技术与装备工程研究中心, 呼和浩特 010080)

摘要: 针对风电机组整机性能退化趋势的及时准确预测需求, 提出了一种结合核熵主成分分析与偏最小二乘法(KECA-PLS)的健康评估算法; 该研究采用前置局部异常因子算法(LOF)对数据预处理, 并通过高斯混合模型(GMM)对风电场数据集进行分类聚类, 分类标签指导后续分析; 利用核熵成分分析(KECA)对数据实施降维处理并提取关键特征, 结合 SPE 统计量监控风电机组运行状态; 考虑到 SCADA 数据的非平稳与非线性特性, 将 PLS 算法融入核熵框架以进行故障预测, 依据预测残差动态设定报警阈值, 确保故障的早期警示; 算法应用模糊评判机制并生成风电机组退化状态的雷达图, 直观展示了故障演进态势; 该算法在某风电场的实际运营数据上的实施验证, 证明了其有效评估风电机组即时健康状况的能力, 以及清晰展示故障演变过程的可视化效果。

关键词: 偏最小二乘法(PLS); 状态监测; 健康评估; 风电机组; 可视化

A kernel Entropy PLS Algorithm and its Application for Wind Turbine Performance Degradation Prognostic

MA Liang¹, REN Chao^{2,3}, LIU Yuewen^{2,3}, HAO Zengxiao²

(1. Datang New Energy Co. Ltd., Chifeng 024000, China;

2. Inner Mongolia university of technology, Hohhot 010051, China;

3. Smart Energy Technology and Equipment Engineering Research Center of Higher Education Institutions in Inner Mongolia Autonomous Region, Hohhot 010080, China)

Abstract: In response to the demand for timely and accurate prediction of the performance degradation trends of wind turbines, a health assessment algorithm combining Kernel Entropy Principal Component Analysis (KECA) and Partial Least Squares (PLS) is proposed. This study uses the Local Outlier Factor (LOF) algorithm for data preprocessing and applies Gaussian Mixture Model (GMM) to classify and cluster the wind farm dataset, where the classification labels guide subsequent analysis. Kernel Entropy Component Analysis (KECA) is utilized to perform dimensionality reduction and extract key features from the data, and the SPE statistic is used to monitor the operational status of the wind turbine. Considering the non-stationary and nonlinear characteristics of SCADA data, the PLS algorithm is integrated into the Kernel Entropy framework for fault prediction, with dynamic alarm thresholds set based on prediction residuals to ensure early fault warnings. The algorithm employs a fuzzy judgment mechanism and generates a radar chart to visually display the wind turbine degradation state, providing an intuitive view of fault evolution. The implementation and validation of this algorithm on actual operational data from a wind farm demonstrated its effectiveness in assessing the real-time health status of wind turbines and its capability in clearly visualizing the fault evolution process.

Keywords: partial least squares (PLS); condition monitoring; health assessment; wind turbines; visualize

收稿日期: 2024-03-20; 修回日期: 2024-05-09。

基金项目: 国家自然科学基金(62363029); 内蒙古自治区高等学校科学研究项目(NJZY22365); 企业研发项目(CDTH20230066684)。

作者简介: 马 亮(1986-), 男, 大学本科, 高级工程师。

刘月文(1982-), 女, 副教授。

引用格式: 马 亮, 任 超, 刘月文, 等. 基于 KECA-PLS 的风电机组健康状态评估[J]. 计算机测量与控制, 2025, 33(5): 351-360.

0 引言

随着能源危机的加剧,风能作为一种关键的绿色可再生能源正逐渐在世界能源结构中占据重要地位。风电机组通常在恶劣的环境中运行,面临着剧烈的工况波动和强烈的机械应力。由于机组的运行状态和负荷承受着随机性和不平稳性,因此研究提升风电机组的可利用率,降低运行维护成本以提高风场经济效益变得至关重要。

目前已经在机组的关键部件状态监测与故障诊断方面取得了广泛研究的成果^[1-2]。在此基础上,进一步研究关于风电机组的故障预测与健康管(PHM, prognostic and health management),开发健康状态诊断与评估系统,实现对风电机组全生命周期的健康管理,已经成为当前研究的热点。PHM 利用各种算法和智能模型监控、预测和管理健康状态,以实现视情维护或预测维护。关于风电机组整机性能退化的预测与评估是风机 PHM 研究的关键内容。

预测并评估风机的健康状态需要考虑反映风机状态的各类参数,这些参数主要来自 SCADA (supervisory control and data acquisition) 系统提供的运行与状态数据以及独立增设的传感器测量数据等。通过提取风场海量数据中的特征信息,即可获知机组的运行状态。目前,一些风场已经在原有 SCADA 系统的基础上集成了独立的状态监测系统 (CMS, condition monitoring system),主要关注传动链振动信号和油液信号的监测。虽然 CMS 信号频率快,但由于数据获取困难、缺乏完备性,并且无法涵盖全部故障信息,因此近年来许多学者开始利用 SCADA 参数进行相关研究^[3-5]。利用 SCADA 数据进行风机性能退化趋势的研究首先可以从风功率数据出发。由于风速的随机变化影响发电机轴所受负载,负载与转子和发电机间转矩成比例,从而影响风机的故障率^[6]。因此,风功率曲线的变化能间接揭示发电机组的运行状态,反映风机性能退化趋势。

基于风功率分析风电机组健康状态的方法包括数据驱动法、模型驱动法和混合方法^[7]。数据驱动法可细分为基于信号、统计学习和机器学习知识驱动^[8],以及基于数据驱动模型的分析法等^[9-10],其基本思想有:(1)“趋势分析”,例如利用风速和转矩的散点图进行回归分析,根据时间序列曲线变化趋势定义评价指标;(2)“聚类”,对正常和异常观测值自动分类,由测试值到聚类中心的“距离”判断异常发生,例如 GMM (Gaussian mixed model) 和 SOM (self-organization mapping net) 等;(3)“残差法”基于非线性回归分析方法对响应变量建立预测模型描述风机的正常运行状态,由预测值与实测值间的“残差”判断风机异常与失效行为的发生,例如支持向量机^[11]、相关性分析^[12]、

fuzzy sys-tem^[13]、非线性回归^[14]等。风电机组运行状态预测模型的建立后,可以进一步建立机组健康状态评估模型,例如基于劣化度模型评价^[15]、门控循环网络^[16]等方法融合机组性能预测结果,以确定健康状态等级。然而,传统的预测方法往往存在过度依赖数据和完备性的问题。为了减少预测误差,采用了复杂算法,但同时引入了新的不利因素,即模型对数据样本的依赖较大,从而导致评价结果的稳定性不足。为了克服这些问题,本文提出了一种新的风电机组健康状态评估算法,结合了核熵主成分分析和偏最小二乘 (KECA-PLS)。该方法首先通过核熵主成分分析进行降维,提取特征信息,实现对故障的监测。随后,在核熵空间引入偏最小二乘法 (PLS) 用于故障预测,具体预测对象包括齿轮箱油温、齿轮箱主轴温度和齿轮箱润滑油入口压力。通过考虑高阶信息,该方法能够提高预测性能。最后,通过对预测模型生成的多个残差进行融合,并借助模糊评判法,实现对健康状态的评估。通过建立健康状态雷达图的可视化方式呈现给运行人员,使其能够更直观地了解风电机组的健康状况。这一整合方法旨在克服传统方法的局限性,为风电机组的运行和维护提供更可靠的支持。

1 风电机组健康状态预测与评估方法

本文研究了风电机组齿轮箱的状态监测与故障预测。图 1 展示了整个算法的总体流程框图。该算法分为 4 个主要阶段:(1) 数据预处理和工况划分;(2) 基于 KECA 的风电机组状态监测;(3) 在 KECA 投影空间中进行基于 PLS 的故障预测;(4) 利用故障预测结果融合残差进行健康评估。

1.1 数据预处理

数据预处理旨在提高数据质量,通过清理、筛选、剔除和转换等操作处理冗余数据。这包括参数选择、对非结构化数据进行结构化模型表示,以及与多元多尺度数据的融合,以有效提升数据的质量。

1.1.1 异常点剔除

数据中的异常值包括离群点和重复点,通常是由于测量误差、记录缺失或其他异常原因导致的。为了处理这些异常值,首先观察风速与功率的时间序列曲线,以识别并剔除重复点。对于离群点,由于数据集缺乏标签变量,本文采用了无监督的侦测方法,从特征空间的距离、概率和密度三个方面对离群点进行界定,通过这三种方法的综合评估,可以有效地识别和定位数据中的异常点。在本文中,从特征空间的密度角度出发,采用基于密度的局部异常值思想的 LOF (local outlier factor) 方法来识别离群点。

LOF 算法的基本描述如下:K-近邻 (k-distance) 就是在距离数据点 p 最近的几个点中,第 k 个最近的点与

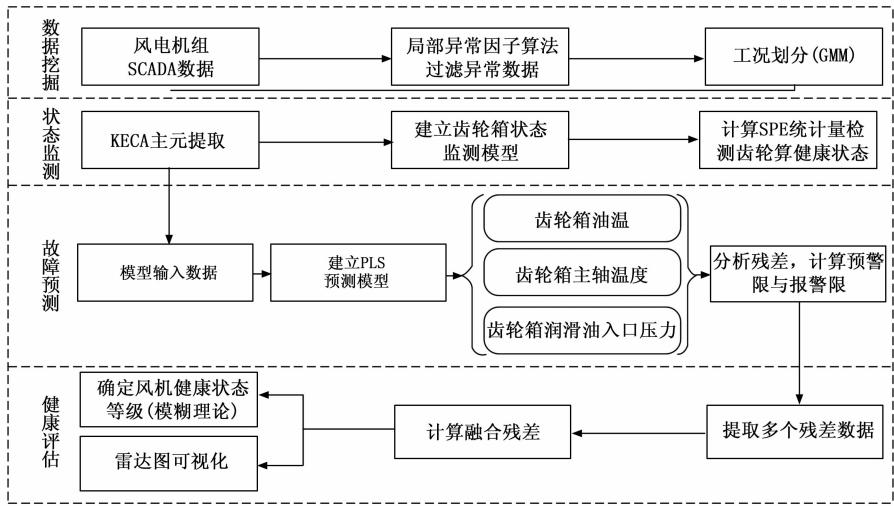


图 1 算法总体框图

点 p 之间的距离记为 $k\text{-distance}(p)$ 。局部异常因子算法利用“局部可达密度”来表示某个样本的异常程度, 当某个样本点的局部密度越大, 则越有可能是异常点。

1.1.2 工况划分

使用高斯混合建模 (GMM) 方法进行工况的自适应划分, 以提取准线性段的风功率数据。

GMM 方法的核心思想基于统计分布的聚类, 即假设样本数据包含若干“自然小类”, 每个小类所包含的观测值均来自于某个特定的统计分布。通过估计这些分布的参数, GMM 能够将复杂的数据集划分为多个子群体, 每个子群体呈现出相似的统计特性, 从而实现对风功率数据的自适应划分, 特别是在准线性段, 能够更好地提取数据中的潜在规律。观测的整体被视为来自多个统计分布的有限混合分布的随机样本。因此, 将非高斯分布特征的数据分解为多个高斯分布函数的线性组合:

$$H(x) = \sum_{i=1}^n p_i N(\mu_i, \delta_i) \tag{1}$$

其中: p_i 为第 i 个成分的先验概率, 满足: $p_i \geq 0$, $\sum_{i=1}^n p_i = 1$; $N(\mu_i, \delta_i)$ 是第 i 个成分的高斯分布函数。

使用期望最大化 (EM, Expectation-Maximization) 聚类算法, 在已知样本数据但未知其所属类别的情况下, 算法通过迭代过程估计数据的潜在概率分布, 并利用这些估计值来更新聚类的参数, 最终实现对数据集的聚类。确定成分参数包括密度函数的均值 $\hat{\mu}$ 和方差 $\hat{\delta}$:

$$LL(\mu, \delta | X, z) = \log \prod_{i=1}^n N(x_i, z_i | \mu, \delta) = \sum_z p \sum_i \log N(x_i, z_i | \mu, \delta) \tag{2}$$

通过建立有限高斯混合模型对风功率数据进行处理, 通过各成分的风速变量的均值 (类质心) μ_i 、方差 δ_i 和占比 (混合概率) p_i , 来截取具有准线性关系的风速与输出功率的数据段。

1.2 基于 KECA 的风电机组状态监测

为了解决处理非线性结构数据时的限制, Scholkopf 等人提出了核主成分分析 (KPCA), 该方法通过使用核函数将数据映射到高维空间, 从而实现非线性降维。在高维空间中, 原本在低维空间中线性不可分的数据能够变得更易于分离。之后, Robert Jenssen^[17] 等人提出了核熵成分分析 (KECA)。该方法采用核技术实现数据由从低维到高维空间的映射, 以便更好地处理非线性关系的数据。本文利用 KECA 方法进行白化处理, 相较于传统的 PCA 和 KPCA 算法, 能够在风电机组状态监测问题上表现出更为灵活和适应非线性数据的特点。通过引入信息熵以更有效地揭示数据集中的结构信息, KECA 在捕捉数据特征方面显示出显著的优势。通过对比分析, 清晰地呈现了 KECA 相对于 PCA 和 KPCA 的优越性, 进一步验证了 KECA 在风电机组状态监测中的应用潜力。

基于 KECA 白化处理, 在核映射空间建立 ECA 模型, 并计算相应的 SPE 监测统计量。这一步骤是基于对主元模型的训练和对待检测数据的映射, 通过分析监测统计量的变化来识别系统中的异常情况。KECA 的引入为风电机组状态监测提供了一种更为灵活且适应非线性关系的有效工具, 有望在实际应用中展现出更好的性能。本文使用的检测统计量是平方预测误差 (Squared Prediction Error, SPE)。SPE 统计量用于监测过程或系统的行为是否偏离了正常运行状态, 通过计算实际观测值与预测值之间的误差的平方和, 来评估当前状态与期望状态的差异。当 SPE 值超过预定的控制限时, 表明系统可能出现异常或故障, 需进一步调查和处理。SPE 统计量的计算如下:

$$SPE_i = E_i E_i^T = t_i (I - P_R P_R^T) t_i^T \tag{3}$$

式中, t_i 为输入向量 x 在特征空间中第 i 个核主元, P_R 为核空间提取特征向量。

SPE 统计量的控制限计算公式如下:

$$\text{SPE}_v = \theta_1 \left[\frac{c_v \sqrt{2\theta_2 h^2}}{\theta_1} + 1 + \frac{\theta_2 h(h-1)}{\theta_1^2} \right]^{1/h} \quad (4)$$

其中: $\theta_i = \sum_{j=p+1}^n \lambda_j^i, i=1,2,3; h=1-\frac{2\theta_1\theta_3}{\theta_2^2}$; C_γ 为

标准正态分布检验水平为 γ 的临界值。若 $\text{SPE} < \text{SPE}_\gamma$, 则表示输入向量的 SPE 统计正常。

本研究提出了一种基于 KECA 算法的风电机组状态监测方法。该方法的实现过程主要包括离线训练和在线监测两个关键步骤, 具体流程如图 2 所示。

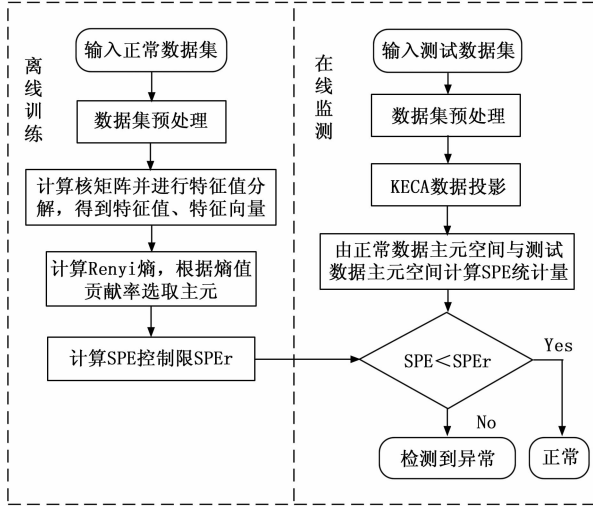


图 2 基于 KECA 算法的齿轮箱状态监测流程图

1.3 基于 PLS 的核熵空间预测模型

基于 KECA 的风电机组状态监测仅能判断某个特定时段内是否发生故障, 但不能评估风电机组的整体健康状况。KPLS 算法在风电机组的预测过程中, 当出现缓慢且不易察觉的故障时, 可能无法进行有效监测, 可能导致大量误报警情况。这是因为微小故障信息通常会被噪声掩盖, 且往往蕴含在高阶统计量中。而 KPLS 算法仅能够提取二阶统计量信息, 未能有效地揭示二阶以上的高阶统计量所包含的信息熵。为解决这一问题, 本文在 KECA 投影空间中引入 PLS 进行故障预测。该方法对于微小缓变故障更为敏感, 有助于提升对风电机组状态变化的察觉。

1.3.1 基于核熵空间的 PLS 算法

在核熵空间中应用 PLS 算法的过程通过非线性映射将输入数据映射到一个新的高维空间, 以便更好地捕捉数据的内在结构。在这一高维空间中, 数据的非线性特性被转化为线性关系, 从而使得传统的 PLS 算法能够更有效地提取数据中的潜在模式和信息。

信息熵是用来度量系统不确定性的指标。在这里, 选用 Renyi 熵进行排序, 以更好地理解 and 利用数据的不确定性特征。Renyi 熵的定义如下所示:

$$H(p) = -\log \int p^2(x) dx \quad (5)$$

其中: $p(x)$ 是数据 D 的概率密度函数。

由于式 (5) 为单调函数, 所以该式可以表示为如下公式:

$$V(p) = \int p^2(x) dx \quad (6)$$

用 Parzen 窗密度估计对式 (6) 进行估计得到下式:

$$\hat{p}(x) = \frac{1}{N} \sum_{i=1}^N W_{\sigma}(x, x_i) \quad (7)$$

将式 (7) 代入式 (6) 可以得到:

$$\hat{V}(p) = \int \hat{p}^2(x) dx = \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N \int W_{\sigma}(x, x_i) W_{\sigma}(x, x_j) dx = \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N \int W_{\sqrt{2}\sigma}(x, x_i) \quad (8)$$

高斯函数的卷积积分依然是高斯函数, 则对式 (8) 进行符号简化可以得到 $W_{\sqrt{2}\sigma}(x_i, x_j)$ 。选用高斯函数作为核函数, $W_{\sqrt{2}\sigma}(x_i, x_j)$ 则可以表示如下式:

$$\hat{V}(p) = \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N k(x_i, x_j) = \frac{1}{N^2} \mathbf{I}^T \mathbf{K} \mathbf{I} \quad (9)$$

其中: $\mathbf{I}$ 是 $N \times 1$ 全 1 向量, $\mathbf{K}$ 是 $N \times N$ 的核矩阵。

对核矩阵 $\mathbf{K}$ 进行特征分解 $\mathbf{K} = \boldsymbol{\Phi}^T \boldsymbol{\Phi} = \mathbf{E} \mathbf{D} \mathbf{E}^T$, 其中 $\mathbf{D} = \text{diag}(\lambda_1, \dots, \lambda_N)$, $\mathbf{E}$ 是 $\mathbf{D}$ 对应的特征向量阵 $\mathbf{E} = (\mathbf{e}_1, \dots, \mathbf{e}_N)$ 。由此得下式 [19]:

$$\hat{V}(p) = \frac{1}{N^2} (\sqrt{\lambda_i} \mathbf{e}_i^T \mathbf{1})^2 \quad (10)$$

在 PLS 算法中的广义特征向量是 $\mathbf{X}^T \mathbf{Y} \mathbf{Y}^T \mathbf{X}$, 而 $\boldsymbol{\omega}$ (权值向量) 是广义特征矩阵的最大的特征值对应的特征向量:

$$\mathbf{X}^T \mathbf{Y} \mathbf{Y}^T \mathbf{X} \boldsymbol{\omega} = \lambda \boldsymbol{\omega} \quad (11)$$

$\mathbf{X}$ 的得分向量 $\mathbf{t}$ 的计算公式如下:

$$\mathbf{t} = \mathbf{X} \boldsymbol{\omega} \quad (12)$$

但是, 在核特征空间中 $\boldsymbol{\omega}$ 和 $\mathbf{t}$ 并不能直接得到, 需要对非线性迭代偏最小二乘法 (NIPALS) 算法进行核熵化处理。由公式 (11) 和公式 (12) 可得到下式:

$$\mathbf{X} \mathbf{X}^T \mathbf{Y} \mathbf{Y}^T \mathbf{X} \boldsymbol{\omega} = \lambda \mathbf{X} \boldsymbol{\omega} \quad (13)$$

$$\text{即: } \mathbf{X} \mathbf{X}^T \mathbf{Y} \mathbf{Y}^T \mathbf{t} = \lambda \mathbf{t} \quad (14)$$

$$\mathbf{K} \mathbf{Y} \mathbf{Y}^T \mathbf{t} = \lambda \mathbf{t} \quad (15)$$

在本文算法中可令: $\mathbf{Z} = \mathbf{X} \mathbf{X}^T \mathbf{Y} \mathbf{Y}^T$ 由公式 (10) 计算相应 Renyi 熵, 选择 Renyi 熵估计中贡献量最大的特征值以及其相应的特征向量, 该特征向量即为 $\mathbf{X}$ 的得分矩阵 $\mathbf{t}$ 。之后, 根据 $\mathbf{t}$ 的值计算预测变量的得分矩阵 $\mathbf{u}$ 。

1.3.2 基于核熵空间 PLS 的残差预测

当齿轮箱出现故障迹象时, 残差特性将出现异常波动, 通过监测残差异常变化趋势可判别齿轮箱是否存在故障。这种监测方法允许在齿轮箱发生潜在故障时快速做出反应。通过实时跟踪残差特性的变化, 可以提前察

觉到异常情况, 从而及时采取维护和修复措施, 有助于降低潜在故障带来的风险和损失。然而, 仅依赖于单一参数的残差可能不足以充分反映齿轮箱的整体健康状况。因此, 通过对多个残差进行综合分析, 可以提高系统的稳定性和可靠性。为此, 本文提出了融合残差概念, 将其定义为 β :

$$\beta = \begin{cases} 0, & \forall \lambda_i < L_i \\ \frac{\sum_{i=1}^n \alpha_i \lambda_i}{n}, & \exists \lambda_i > L_i \\ \max(\lambda_i), & \forall \lambda_i > L_i \end{cases} \quad (16)$$

其中: λ_i 为第 i 个残差 ($i=1, 2, \dots, n$), L_i 为第 i 个残差的阈值。 α_i 为参数影响齿轮箱健康状态的权重因子, 通过对风电机组早期运行状态的观察以及结合经验推断进行分析, 综合推断出这些权重因子的具体数值。

1.4 基于模糊综合评判法的齿轮箱健康状态评估

在进行风电机组齿轮箱健康状态评估时, 依据健康状态分级原则, 将机组状态参数异常的评估结果划分为 5 个不同等级, 具体包括: $L = \{l_1, l_2, l_3, l_4, l_5\} = \{\text{健康}, \text{良好}, \text{注意}, \text{恶化}, \text{疾病}\}$ 。本文中, 依据半梯形和三角形组合构建隶属度分布函数。半梯形和三角形的组合形式能够更贴近机组在退化过程中的真实状态变化, 反映了系统性能随时间的渐进变化。通过这种隶属度函数的应用, 期望更准确地描述机组性能的演化趋势, 从而为后续的分析 and 决策提供更可靠的基础。通过隶属度函数描述健康状态的不同等级如下:

$$l_1 = \begin{cases} 1, & \beta < a_1 \\ \frac{a_2 - \beta}{a_2 - a_1}, & a_1 \leq \beta < a_2 \\ 0, & \beta \geq a_2 \end{cases} \quad (17)$$

$$l_2 = \begin{cases} 1, & \beta < a_1 \\ \frac{\beta - a_1}{a_3 - a_1}, & a_1 \leq \beta < a_3 \\ \frac{a_4 - \beta}{a_4 - a_3}, & a_3 \leq \beta < a_4 \\ 0, & \beta \geq a_4 \end{cases} \quad (18)$$

$$l_3 = \begin{cases} 1, & 0 \leq \beta < a_3 \\ \frac{\beta - a_3}{a_4 - a_3}, & a_3 \leq \beta < a_4 \\ \frac{a_5 - \beta}{a_5 - a_4}, & a_4 \leq \beta < a_5 \\ 0, & \beta \geq a_5 \end{cases} \quad (19)$$

$$l_4 = \begin{cases} 1, & 0 \leq \beta < a_4 \\ \frac{\beta - a_4}{a_5 - a_4}, & a_4 \leq \beta < a_5 \\ \frac{a_6 - \beta}{a_6 - a_5}, & a_5 \leq \beta < a_6 \\ 0, & \beta \geq a_6 \end{cases} \quad (20)$$

$$l_5 = \begin{cases} 0, & 0 \leq \beta < a_5 \\ \frac{\beta - a_5}{a_6 - a_5}, & a_5 \leq \beta < a_6 \\ 1, & \beta \geq a_6 \end{cases} \quad (21)$$

根据早期风电机组实际运行数据, 经过多次实验, 选取 $a_1 = 1.5$ 、 $a_2 = 4.86$ 、 $a_3 = 8.54$ 、 $a_4 = 11.56$ 、 $a_5 = 14.58$ 、 $a_6 = 16$, 以此为基础进行模糊综合评判, 从而尽可能准确地分析出风电机组的健康状态。

2 案例应用与算法分析

为验证本文提出方法的有效性和适用性, 选取某风电场的 2 MW 风电机组 (型号 UP-96-2000) 进行实验。数据来源于该机组在故障前近两个月 (2 月 21 日至 4 月 16 日) 的 SCADA 数据集, 采样周期为 5 秒。该机组的主要参数如下: 切入风速为 3 m/s, 额定风速为 10 m/s, 切出风速的 10 分钟和 3 秒均值分别为 25 m/s 和 35 m/s; 叶轮转速范围为 9.88~19.5 RPM, 额定转速为 17.28 RPM; 发电机的额定转速为 1 780 RPM。该机组于 4 月 16 日因主轴高温故障停机, 故障源为齿轮箱故障。

2.1 数据预处理

对数据进行预处理时, 采用局部异常因子算法 (LOF, local outlier factor)。预处理后的功率数据结果如图 3 所示, 其中横轴表示风速, 纵轴表示输出功率。从原始数据的观察可以发现, 存在一些堆积点和离群点。例如, 在输出功率为 1 000 kW 时, 有一小部分数据分布在风速为 15~25 m/s 之间。这些数据被识别为异常数据, 并通过 LOF (局部离群因子) 方法处理后, 异常数据的突出部分得到了有效滤除, 从而验证了该异常数据检测方法的有效性。此步骤有助于提升数据质量, 去除了可能对后续分析和建模产生干扰的异常值。通过这一数据清洗过程, 能够确保后续分析和建模工作的可靠性与准确性, 为模型的稳健性和可信度提供保障。

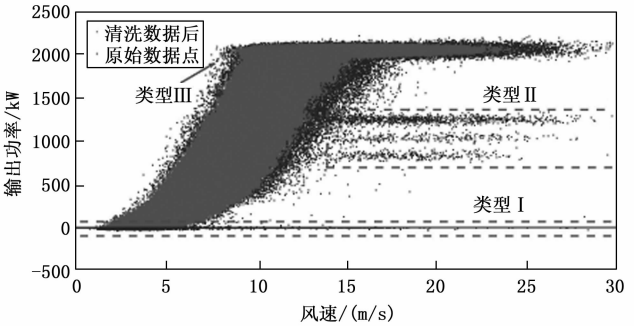


图 3 数据预处理前后对比

2.2 GMM 分类数据

基于 GMM 方法对 SCADA 数据进行聚类, 设定聚类数为 3。首先, 选择风力功率数据作为输入数据, 接着根据风力数据的分类结果将其扩展到所有数据类别。在拟合高斯混合模型后, 得到以下结果: 第一类的混合

比例为 0.395 318, 均值为 5.375 820; 第二类的混合比例为 0.389 258, 均值为 9.196 038; 第三类的混合比例为 0.215 425, 均值为 14.237 322。接下来, 利用高斯混合模型对数据进行聚类处理, 最终获得了每个数据向量对应的标签。

根据图 4 可得, 可以明显看到高斯混合模型的分类结果呈现为三个清晰的类别。在这种情况下, 将根据时间和其他特征信息对聚类结果进行分类, 可以观察到每个类别都具有相应的簇心。

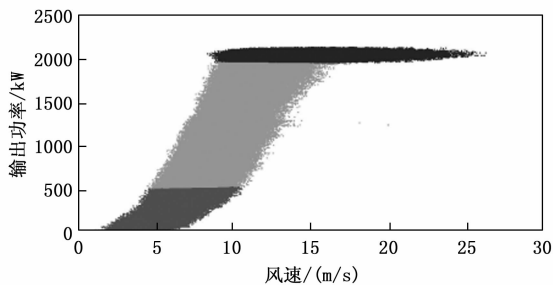


图 4 分类结果可视化

2.3 基于核熵主元分析的监测结果比较

为更好地验证本文方法的有效性, 与传统的 PCA、KPCA 方法进行了对比试验。如图 5 所示, 展示了正常工况下, 三种监测方法对该风电机组齿轮箱状态的监测结果。虚线表示 99% 的控制限, 实线则为正常运行下的 SPE 统计量监控图。由图可见, 所有方法均存在误报警点。然而, 三种方法中, PCA 方法的效果最差, 这可能是由于风机齿轮箱数据具有较强的非线性特征, 而经过非线性变换后的主元效果相对更为优秀, 能够更好地捕捉数据中的复杂模式和变化。此外, 相对于 KPCA, 本文方法表现出更加优越的监测性能, 分析原因可知, KECA 通过 Renyi 二次熵的大小来选择主元, 能

够更有效地揭示数据的结构信息。这种对数据结构的敏感性使得 KECA 在处理风电机组状态监测问题上更具优势, 提高了监测模型的可靠性和适用性。

将待检测样本输入监测模型后, 实际情况表明, 4 月 16 日该风电机组因主轴高温故障导致停机, 故障部件为齿轮箱。针对齿轮箱的状态监测, 采用了 PCA、KPCA 和 KECA 三种监测模型, 并将结果呈现在图 6 中。尽管所有图中都能观察到误报警点, 但在 PCA-SPE 图中, 这些误报警点的 SPE 值却反而高于故障点的 SPE 值, 频繁的误报警会降低监测系统的可信度和可靠性。这种现象表明, PCA 模型在处理复杂的非线性数据时存在一定的局限性, 可能无法有效区分正常工况与异常工况之间的微小差异。相比之下, KPCA 和 KECA 模型能够更好地捕捉数据的非线性特征, 减少误报警点的数量, 并提高了监测系统在故障检测中的敏感性和准确性。尤其是在 3 月 30 日至 4 月 10 日期间产生的异常数据点, 更增加了对误报警点和故障点明确定位的难度。

相比之下, KPCA-SPE 和 KECA-SPE 图中很少出现这种情况, 表明 PCA 在应对非线性数据时存在一定的局限性, 尤其是在区分异常情况的敏感性上存在缺陷。KPCA-SPE 和 KECA-SPE 方法都能有效检测到故障状态, 其中 KECA-SPE 的监测效果更为突出。更为显著的是, KECA-SPE 比其他两种方法更早地发现了故障状态, 并且总体上误报警点的数量更少。这进一步突显了在处理非线性数据时, 选择适当的监测模型至关重要, 以确保能够准确识别和区分故障点与误报警点, 从而提高监测系统的准确性和可靠性。

2.4 基于核熵空间的 PLS 残差预测

借助上一节计算得到的实验训练样本, 构建了一个预测模型, 通过对目标参数进行计算, 得到相应的预测

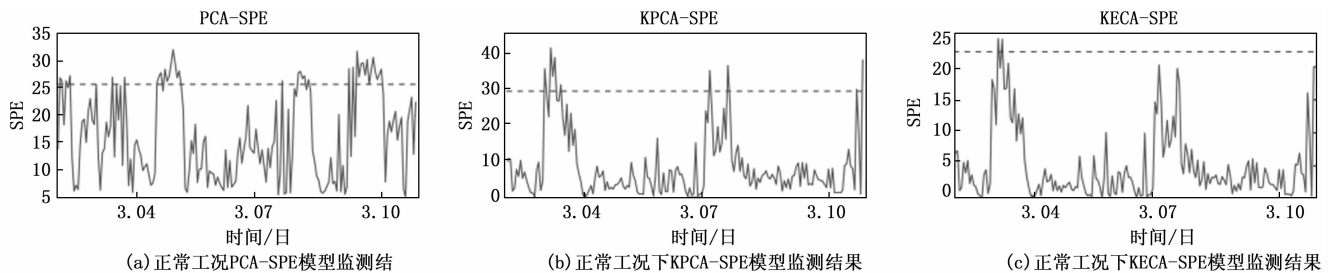


图 5 正常工况下不同监测模型效果对比

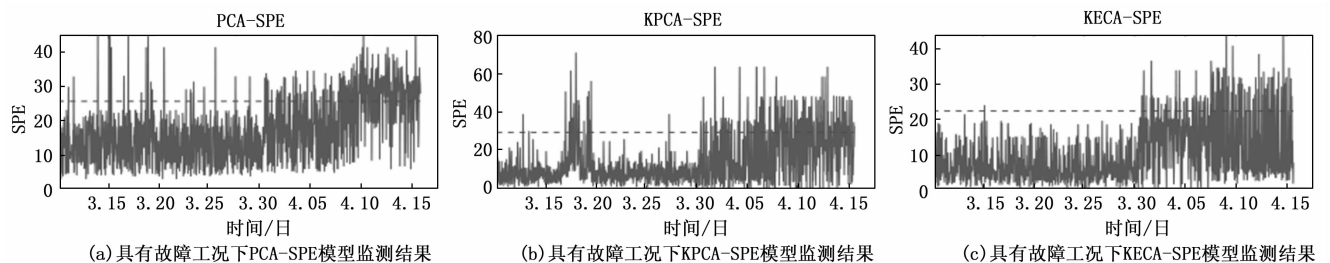


图 6 具有故障工况下 3 种不同模型的监测结果图

输出。之后, 将预测结果与实际目标参数进行比较, 并计算模型的精确度, 以验证本文提出的预测模型的有效性。为此, 采用了两个常用的评价指标: 平均绝对误差 (MAE, mean absolute error) 和平均相对误差 (MRE, mean relative error), 来评估预测模型的性能。图 7 展示了本文方法的预测结果, 如图 8 所示, 与传统的 KPLS 方法进行了比较, 以更好地体现算法的优越性。

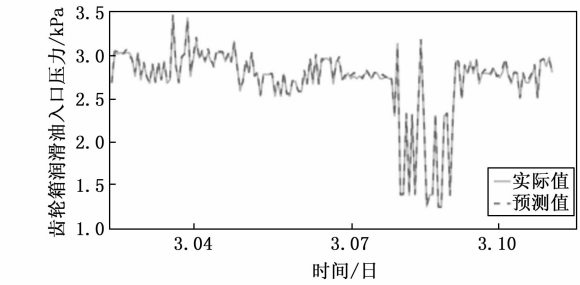


图 7 齿本文方法对入口压力的预测效果

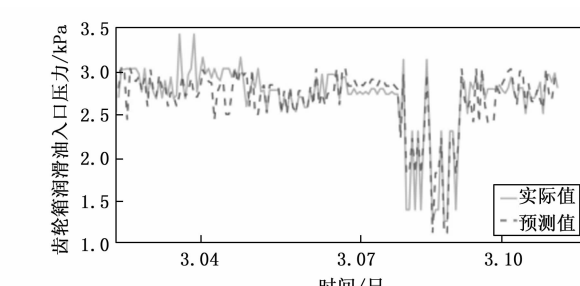


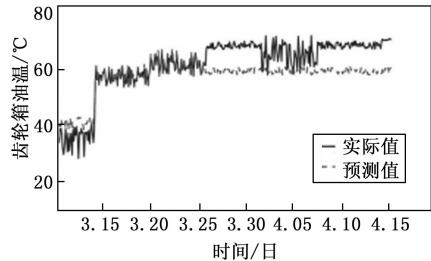
图 8 齿轮箱润滑油入口压力的 KPLS 预测效果

表 1 为 KPLS 模型与本文方法的误差对比。对比表 1 可发现, 本文所用模型预测精度更高。

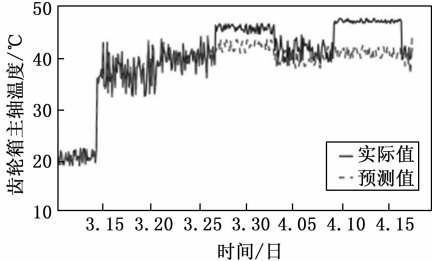
表 1 误差指标表

目标参数	模型	MAE	MRE
齿轮箱油温	KPLS	3.362 0	0.072 5
	本文方法	0.412 8	0.007 8
齿轮箱主轴温度	KPLS	3.627 5	0.131 0
	本文方法	0.272 7	0.008 2
齿轮箱润滑	KPLS	0.146 2	0.059 4
油入口压力	本文方法	0.022 0	0.008 1

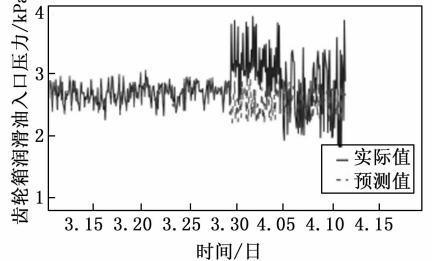
如图 9 所示, 展示了齿轮箱油温、齿轮箱主轴温度和齿轮箱润滑油入口压力的预测结果。从图中可以观察到, 在 3 月 26 日前后, 该风电机组的齿轮箱油温和主



(a) 齿轮箱油温本文方法预测结果



(b) 齿轮箱主轴温度本文方法预测结果



(c) 齿轮箱润滑油入口压力本文方法预测结果

图 9 核熵 PLS 对 3 月 12 日~4 月 16 日数据的预测结果

轴温度出现了高温异常现象。而齿轮箱润滑油入口压力则在 4 月 1 日前后才出现了高压异常, 显示出一定的滞后性。为了进一步分析异常的原因及其时间关系, 计算了残差值, 结果如图 10 所示。

图 10 展示了通过本文方法预测模型得到的目标参数残差图, 采用核密度估计 (KDE) 计算残差的阈值上限, 并将其作为高温 (高压) 报警限。当残差值超出设定的阈值时, 表示齿轮箱状态出现异常, 触发高温 (高压) 报警。从图中可以观察到, 齿轮箱主轴温度在 3 月 26 日之后已经超出了高温报警线, 且此后主轴温度持续处于高温异常状态, 表明存在潜在的故障风险。同时, 齿轮箱油温在 3 月 26 日之后也持续超出了高温报警线, 进一步验证了齿轮箱在该时间段出现的异常状况。通过这种残差监控方法, 能够及时捕捉到齿轮箱的异常行为, 并为故障预警提供早期提示。而在齿轮箱润滑油入口压力的残差图中, 残差波动出现在 4 月 1 日之后, 相较于齿轮箱油温和主轴温度的异常, 呈现出滞后现象。

通过齿轮箱主轴温度和油温两个残差的趋势图, 可以判断该风电机组齿轮箱部件可能由于轴承摩擦产生高温故障。具体而言, 轴承的过度摩擦会导致温度升高, 进而通过热传递效应导致齿轮箱内部油温的升高。这一过程形成了温度异常的连锁反应, 使得齿轮箱部件的健康状态逐步恶化, 最终可能引发齿轮箱的故障。此外, 高温故障还可能引起润滑油入口压力的变化, 但该变化较为缓慢。从而可以得出结论, 在齿轮箱出现异常时, 通过对目标参数进行预测与监测, 可以较早地预见到故障的发生。然而, 不同目标参数的异常发生时间存在差异, 最长可达 5 天左右。如果仅依赖单一参数的监测, 可能导致预测结果的片面性和误差。为解决这一问题, 本文采用了多参数融合残差的方法, 能够更全面地捕捉系统的整体状态, 从而提高了齿轮箱健康状况评估的准确性和可靠性。

对比图 9 和图 10 可以看出, 经过多参数融合后, 残差对异常情况的敏感度显著提高。这使得在正常状态下的异常判断更加准确和精细。通过融合多个目标参数的残差信息, 能够更全面地反映系统的整体状态, 进而提升对异常情况的早期检测能力, 从而实现更精确的故

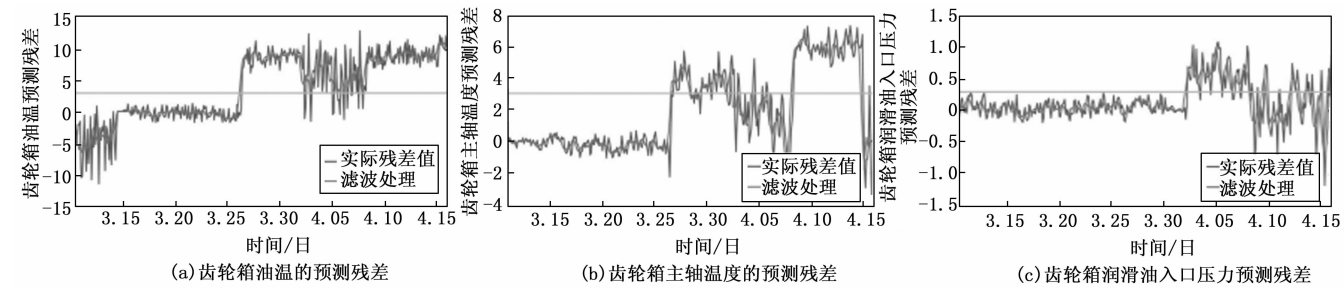


图 10 预测残差

障预警。

根据式 (16) 计算三个目标残差的融合残差 β ，本文通过试凑法选用权重因子 $\alpha_1 = \alpha_2 = 0.4$ ， $\alpha_3 = 0.2$ ，结果如图 11 所示。

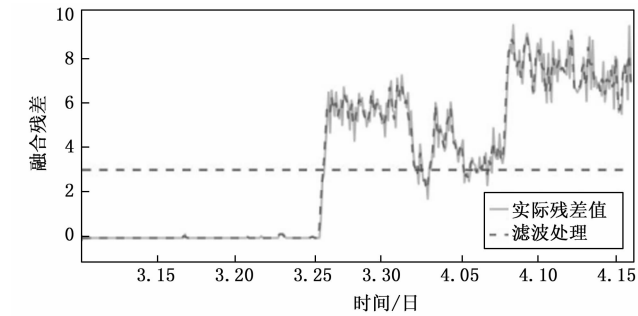


图 11 融合残差

将融合残差 β 代入式 (19) ~ (23)，并设定置信水平为 0.9 后，对隶属度进行基本的置信度分配，从而评估了风电机组齿轮箱的健康状态，具体结果见图 11。由图中可以看出，该齿轮箱健康状态经历了 4 个阶段：阶段一（3 月 10 日—3 月 21 日左右）齿轮箱呈现为“健康”状态，机组正常运行，没有明显的异常迹象；阶段二（3 月 22 日—3 月 26 日左右）齿轮箱的健康状态由“良好”过渡至“注意”，这提示需要加强维护工

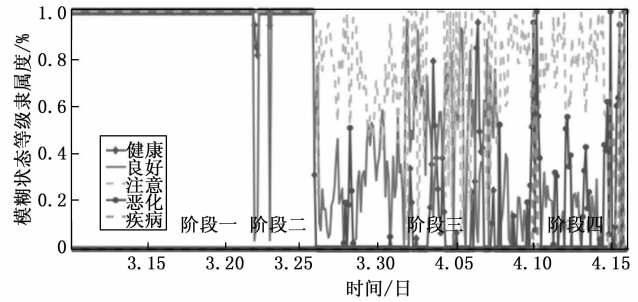


图 12 健康评估结果图

作以确保设备的正常运行；阶段三（3 月 25 日到 4 月 10 日）齿轮箱的健康状态明显退化，多半处于“恶化”状态，暗示可能发生了异常情况，需要更加深入的关注和检修；阶段四（4 月 10 日—4 月 16 日）齿轮箱的恶化趋势更加明显，到 4 月 15 日左右已经达到“疾病”状态，表明存在着明显故障。整个过程从“注意”到“疾病”状态历时 15 天，与 KECA-SPE 状态趋势图的观察结果一致。由于过渡速度极快，推测关键或重要部件/子系统可能发生异常，建议制定紧急运维计划，进行详细的故障排查。

在图 13 中选择 5 个日期，分别对应风电机组的五种状态：健康、良好、注意、恶化、疾病。这些日期与

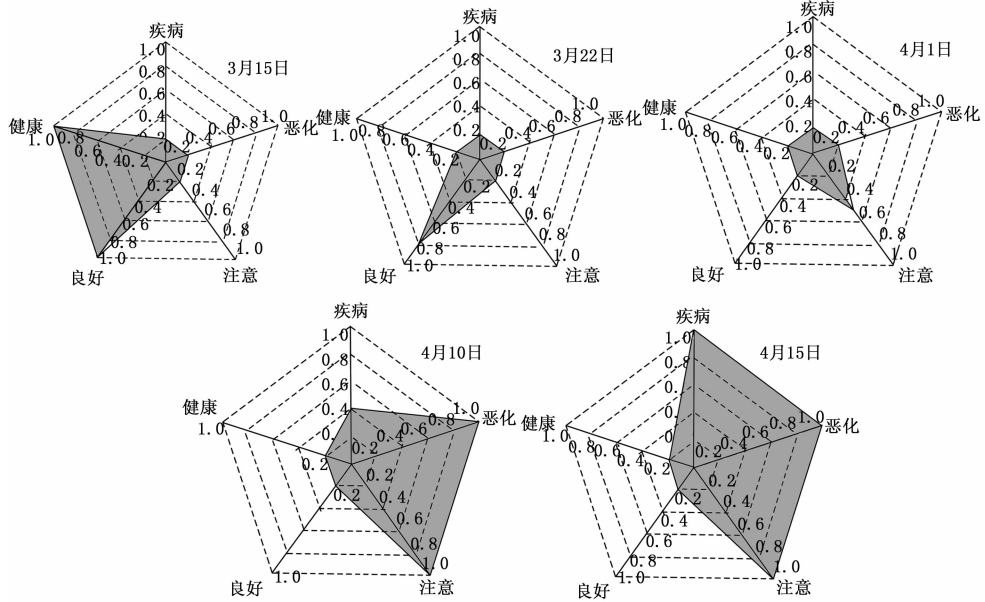


图 13 不同日期下风机健康指标日期雷达图

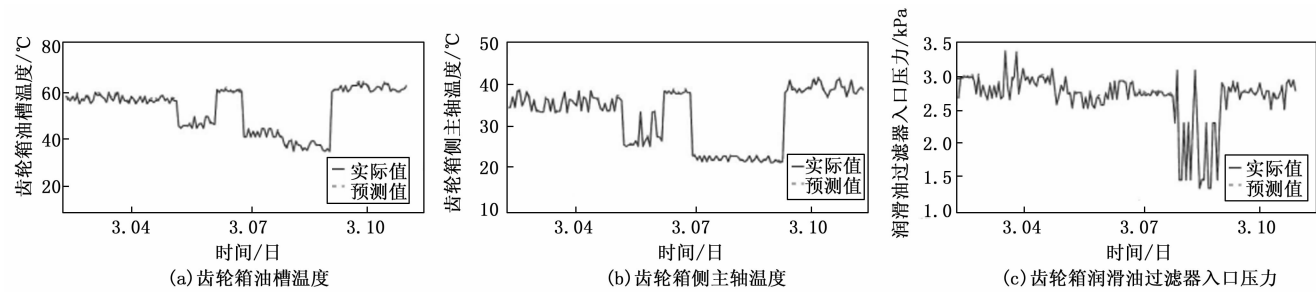


图 14 KEPLS 模型预测效果图

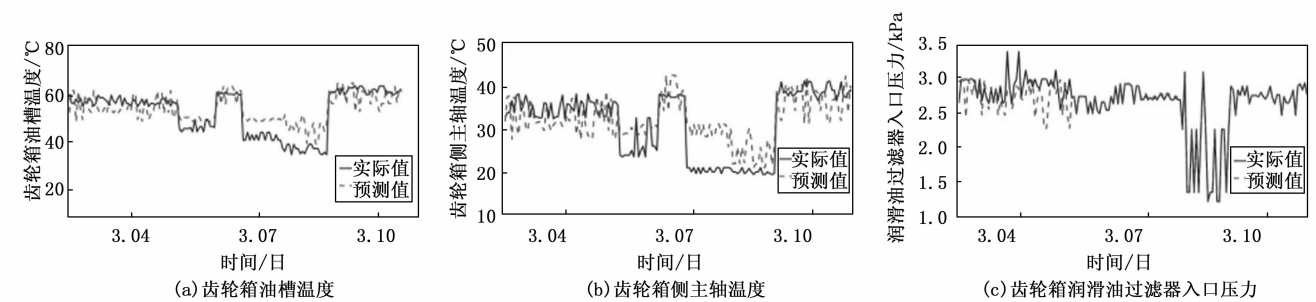


图 15 PLS 模型预测效果图

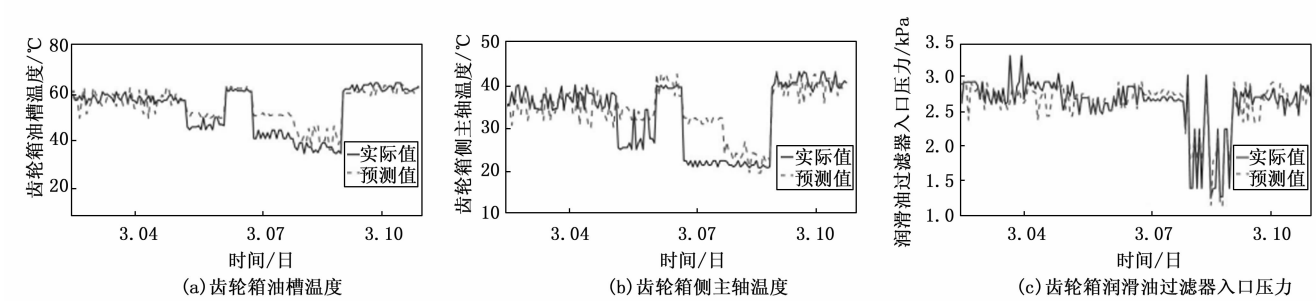


图 16 KPLS 模型预测效果图

图 12 中的健康评估结果一致, 两图的比对表明风机健康预测结果完全一致。

2.5 对比经典算法

KEPLS 模型的预测结果如图 14 所示。为了更好地体现该方法的有效性和优越性, 进行了 PLS 和 KPLS 模型的对比实验, 结果如图 15 和图 16 所示。

表 2 显示了 PLS、KPLS 模型和 KEPLS 模型的误差比较。

表 2 误差指标表			
目标参数	模型	MAE	MRE
齿轮箱油槽温度	PLS	4.937 1	0.103 5
	KPLS	3.361 4	0.072 5
	KEPLS	0.431 7	0.008 2
齿轮箱侧主轴温度	PLS	4.660 3	0.160 5
	KPLS	3.626 8	0.131 0
	KEPLS	0.278 6	0.008 4
齿轮箱润滑油过滤器压力	PLS	0.066 9	0.022 5
	KPLS	0.146 2	0.059 4
	KEPLS	0.023 2	0.008 6

表 2 中的误差指标相比, 本文所用模型的预测精度更高。

3 结束语

本文研究了一种基于核熵 PLS (KEPLS) 的风电机组退化趋势预测与运行状态评估算法。该方法能够实现风电机组齿轮箱的状态监测、过程多变量数据的预测, 并通过融合预测残差综合评估风电机组齿轮箱的健康状态。首先, 建立 KEPLS 预测模型并计算目标参数的残差; 然后, 定义融合残差计算方法, 并借助模糊综合评判对风电机组齿轮箱的健康状况进行评估; 最后, 将 KEPLS 模型与 PLS 模型和 KPLS 模型进行了比较, 结果表明, KEPLS 模型具有较好的拟合效果和较高的故障灵敏性, 验证了该算法在风电机组健康评估中的有效性和优越性。

参考文献:

[1] 龙 寰, 杨 婷, 徐劭辉, 等. 基于数据驱动的风电机组状态监测与故障诊断技术综述 [J]. 电力系统自动化, 2023, 47 (23): 55-69.

- [2] 王中行, 周元贵, 张学广. 基于数据驱动的风电机组状态监测与故障诊断技术综述 [J]. 东北电力大学学报, 2024, 44 (1): 42-51.
- [3] ZHAO Y, LI D, DONG A, et al. Fault prediction and diagnosis of wind turbine generators using SCADA data [J]. Energies, 2017, 10 (8): 1210.
- [4] TAUTZ-WEINERT J, WATSON S J. Using SCADA data for wind turbine condition monitoring-a review [J]. IET Renewable Power Generation, 2017, 11 (4): 382-394.
- [5] CAO M, QIU Y, FENG Y, et al. Study of wind turbine fault diagnosis based on unscented Kalman filter and SCADA data [J]. Energies, 2016, 9 (10): 847.
- [6] 马良玉, 程东炎, 梁书源, 等. 基于 Light GBM-VIFMIC-SFS 的风电机组故障诊断输入特征选择方法 [J]. 热力发电, 2024, 53 (1): 154-164.
- [7] LETURIONDO U, SALGADO O, CIANI L, et al. Architecture for hybrid modelling and its application to diagnosis and prognosis with missing data [J]. Measurement, 2017, 108 (1): 152-162.
- [8] HAECKELL M W, ROLFES R, KANE M B, et al. Three-tier modular structural health monitoring framework using environmental and operational condition clustering for data normalization: Validation on an operational wind turbine system [J]. Proceedings of the IEEE, 2016, 104 (8): 1632-1646.
- [9] BOGOEVSKA S, SPIRIDONAKOS M, CHATZI E, et al. A data-driven diagnostic framework for wind turbine structures; A holistic approach [J]. Sensors, 2017, 17 (4): 720.
- [10] DU M, YI J, MAZIDI P, et al. A parameter selection method for wind turbine health management through SCADA data [J]. Energies, 2017, 10 (2): 253.
- [11] LEAHY K, HU R L, KONSTANTAKOPOULOS I C, et al. Diagnosing wind turbine faults using machine learning techniques applied to operational data [C] //2016 IEEE international conference on prognostics and health management (icphm). IEEE, 2016: 1-8.
- [12] 申云乔, 江政儒, 叶新红. 大数据技术在风电机组运行状态评估及故障诊断中的应用分析 [J]. 电工技术, 2023 (20): 38-40.
- [13] 孙 鹏, 李 剑, 寇晓适, 等. 采用预测模型与模糊理论的风电机组状态参数异常辨识方法 [J]. 电力自动化设备, 2017, 37 (8): 90-98.
- [14] 李雄威, 郭晓雅, 李庚达, 等. 一种基于非线性偏最小二乘的风电机组齿轮箱状态监测方法 [J]. 可再生能源, 2022, 40 (10): 1346-1351.
- [15] 金晓航, 秦治伟, 郭远晶, 等. 基于改进劣化度模型的风电机组日常运行状态评估 [J]. 太阳能学报, 2023, 44 (1): 239-246.
- [16] 李 斌, 宋 威, 赵 凯, 等. 基于 mRMR-GRU 的风电机组运行状态评估 [J]. 可再生能源, 2023, 41 (7): 906-911.
- [17] JENSSEN R. Kernel entropy component analysis [J]. IEEE transactions on pattern analysis and machine intelligence, 2009, 32 (5): 847-860.
- [18] 吕 岩. 一种低成本高可靠的星载计算机存储器容错方法 [J]. 空间控制技术与应用, 2020, 46 (3): 66-71.
- [19] 虞业砾, 施敏华, 郑倩云. FPGA 软件重构验证方法研究 [J]. 电子设计工程, 2020, 28 (21): 6-10.
- [20] 邱 伟. 基于 DSP 和 FPGA 系统的动态重构设计 [J]. 电子技术与软件工程, 2014 (3): 3-12.
- [21] 杨鹏飞, 刘 波, 党佳乐, 等. 面向条件受限环境的动态可重构异构计算平台 [J]. 空间控制技术与应用, 2020, 46 (3): 7-15.
- [22] 王 超, 韩笑冬, 王 睿, 等. 支持网络互连的可重构卫星平台关键技术研究 [J]. 通信学报, 2017, 38 (A01): 5-15.
- [23] 段文杰, 王大轶, 刘成瑞. 一种闭环冗余系统可重构控制方法 [J]. 空间控制技术与应用, 2014, 40 (6): 21-25, 47-55.
- [24] LIU B, CAI L. Monte carlo reliability model for single-event transient on combinational circuits [J]. IEEE Transactions on Nuclear Science, 2017, 64 (12): 2933-2937.
- [25] 蒲卫华, 谢成清, 姜文志, 等. COTS 星载计算机容错设计及可靠性研究 [J]. 空间控制技术与应用, 2020, 46 (3): 72-77.
- [26] 季月明, 张振力, 吴志远, 等. 一种针对 COTS 产品的板级总剂量试验方法 [J]. 航天器环境工程, 2020, 37 (3): 299-302.
- [27] 赵 键, 杨 芳. 中国高分辨率敏捷小卫星的技术创新及应用实践 [J]. 航天器工程, 2021, 30 (6): 23-30.
- [28] Digitalglobe. WorldView-4 data sheet [EB/OL]. [2021-10-12]. <http://gbdxdocs.digitalglobe.com/docs/worldview-4>.
- [29] 张 兵. 智能遥感卫星系统 [J]. 遥感学报, 2011, 15 (3): 415-431.
- [30] 李德仁, 王 密, 沈 欣, 等. 从对地观测卫星到对地观测脑 [J]. 武汉大学学报 (信息科版), 2017, 42 (2): 143-149.
- [31] YANG C T, CHEN S T, CHENG W H, et al. A heterogeneous cloud storage platform with uniform data distribution by software-defined storage technologies [J]. IEEE Access, 2019, 7: 1.
- [32] 宫经刚, 宁 宇, 吕 楠. 美国高轨天基态势感知技术发展及启示 [J]. 空间控制技术与应用, 2021, 47 (1): 1-7.