

基于动态集群的风电机组异常状态检测方法

于华楠,李靖雨,王鹤,李石强,边竞

(东北电力大学 现代电力系统仿真控制与绿色电能新技术教育部重点实验室,吉林 吉林 132012)

摘要:针对风电机组异常状态的检测问题,提出了考虑相似机组运行状态的风电机组异常检测方法。基于滑动时窗和K-means聚类算法对风电机组运行数据进行分析,提出了风电机组动态集群方法,进而建立了考虑时空相关性的风电机组集群。提出基于自适应权重与Levy飞行策略的北方苍鹰优化(WLNGO)算法;利用五折交叉验证(5CV)改进WLNGO算法,得到WLNGO-5CV算法,并利用该算法对核极限学习机(KELM)的超参数进行优化,进一步提出WLNGO-5CV-KELM回归模型。结合滑动时窗对相似机组预测残差进行统计分析得到实时预警阈值,消除了工况等因素对风电机组的影响,能够对目标风电机组进行可靠的异常检测。通过对中国东北某风电场的实际数据进行仿真分析,验证了所提方法的有效性和准确性。

关键词:风电机组;WLNGO-5CV-KELM回归模型;时空相关性;动态集群;异常状态监测;数据采集与监控系统

中图分类号:TM315

文献标志码:A

DOI:10.16081/j.epae.202412038

0 引言

近年来,随着风电的规模化开发和利用,风力发电已经跻身全国能源战略的关键组成部分^[1-3]。然而,风电机组工作环境恶劣,在运行过程中可能会面临各种异常情况^[4]。因此,在风电机组发生故障前,准确地对其工作状态进行监测,采取预防性维护措施^[5]可以降低故障的发生率,进而保障机组健康稳定地运行。目前,风电机组数据主要通过数据采集与监控(supervisory control and data acquisition, SCADA)系统获取^[6-7],然而SCADA系统采集的数据中存在大量异常数据,对于分析运行数据的真实规律以及特征参数相关性的挖掘与应用造成严重影响。因此,有大量的学者对采集到的数据进行预处理^[8-9],并基于处理后的数据建立风电机组状态模型。

风电机组历史故障状态信息较少,导致正常和异常样本不平衡,可通过挖掘SCADA历史数据中正常状态和故障状态信息,进而建立故障辨识模型^[10],但此类方法需要大量的故障数据,往往难以满足条件。目前常用的方法是从SCADA历史数据中提取正常状态运行数据,以建立风电机组状态预测模型^[11-12]。为充分挖掘参数间的关系,文献[13]提出了一种长短期记忆(long short-term memory, LSTM)、模糊综合和特征迁移学习相结合的风电机组运行状态预测方法,该方法可以充分学习风电

组的特征,提高了状态预测的准确性。文献[14-16]利用LSTM分别与堆叠降噪自动编码器、自动编码器、卷积神经网络相结合建立观测参数与目标参数之间的逻辑关系,但是此类方法未能考虑风电机组运行过程中受自然环境影响而导致训练数据出现异常波动的情况,因此具有一定的局限性。文献[17]采用遗传算法(genetic algorithm, GA)对反向传播神经网络(back propagation neural network, BPNN)模型进行优化并建立预测模型,虽然考虑了风速波动对预测结果的影响,但在实际运行中,风电机组的运行状态受多种因素影响,仅通过对风速进行划分未必能够提高风电机组状态检测的准确性。

尽管上述方法可以实现风电机组运行状态的预测,但由于采用的算法单一,提升模型精度和泛化能力存在一定难度。考虑到各种概率估计模型的优点,文献[18]提出了一种基于组合概率估计模型的异常检测方法,通过采用假设检验(hypothesis testing, HT)方法,实现了风电机组早期的故障预警,虽然组合概率估计模型的检测结果优于独立概率估计模型,但此类方法未能考虑相似机组运行状态对预测结果的影响;文献[19-20]建立回归模型时虽然考虑了其他机组预测模型的辨识结果,但仅采用平均绝对误差对本机组历史样本和其他机组近期样本建立的预测模型进行选择,只能筛选出其他机组中预测精度较高的模型,并未考虑到不同地理环境、天气状况和不同时刻下风电机组的相似性,其效果未必最佳。

为了较好地实现风电机组的异常状态检测,本文提出了一种基于动态集群的风电机组异常状态检测方法。首先,采用局部离群因子(local outlier factor, LOF)和最大互信息系数(maximal information

收稿日期:2024-03-22;修回日期:2024-10-13

在线出版日期:2025-01-03

基金项目:吉林省科技发展计划项目(20220203163SF)

Project supported by the Science and Technology Development Plan Project of Jilin Province(20220203163SF)

coefficient, MIC)对风电机组状态参数中的数据进行预处理;其次,考虑到不同时段下运行环境及风电机组状态的差异,提出了动态集群方法,通过使用特定尺度的窗口进行滑动截取数据,采用K-means聚类算法对各时段数据集进行聚类,从而对风电场内运行状态相似的风电机组进行动态划分;然后,通过在北方苍鹰优化(northern goshawk optimization, NGO)算法的第1个狩猎阶段引入自适应权重,第2个狩猎阶段引入Levy飞行策略对北方苍鹰位置进行更新,提出基于自适应权重与Levy飞行策略的北方苍鹰优化(adaptive weight and Levy flight based northern goshawk optimization, WLNGO)算法,并利用五折交叉验证(5-fold cross validation, 5CV)改进WLNGO算法以实现核极限学习机(kernel extreme learning machine, KELM)超参数的优化,从而进一步提出WLNGO-5CV-KELM回归模型;最后,运用统计学与滑动窗口相结合的方法计算相似机组的动态阈值作为目标机组的检测阈值,同时采用熵权法将多个相似机组阈值进行融合,实现对风电机组异常状态的检测。

1 数据预处理及风电机组动态集群划分

1.1 风电机组状态参数的选择与处理

在风电机组运行状态检测中,输入变量直接影响模型的准确性,因此选择合适的参数进行风电机组运行状态预测至关重要。SCADA系统记录了自然环境参数、设备运行参数等状态信息,其主要组件和状态参数分别如附录A图A1和表A1所示^[20]。本文主要针对表A1中的参数研究风电机组异常状态检测。

在选择状态参数作为预测模型的输入时,需要考虑以下2个因素。

1)在选择特征参数进行建模时,需要考虑不同组件间存在的关联性。

由于MIC不依赖于数据的分布,可以有效地检测到线性和非线性相关性,能够避免同一类型状态参数的信息冗余。鉴于状态参数之间的关联性,本文采用MIC来衡量变量之间相关性。

首先,按照式(1)计算变量 M 、 N 之间的互信息。

$$I(M; N) = \iint p(m, m') \log_2 \frac{p(m, m')}{p(m)p(m')} dm dm' \quad (1)$$

式中: $I(M; N)$ 为变量 M 、 N 之间的互信息; $p(m, m')$ 为联合概率密度函数; $p(m)$ 和 $p(m')$ 为边缘密度函数。

然后,对所求最大互信息值进行归一化处理得到MIC,如式(2)所示。

$$C_{MI}(M; N) = \max_{ab < B} \frac{I(M; N)}{\log_2 \min(a, b)} \quad (2)$$

式中: $C_{MI}(M; N)$ 为变量 M 、 N 之间的MIC值; a 、 b 分别为横、纵轴方向上划分的格子数量,本质上代表网格分布; B 为划分二维平面网格的最大单元数,为数据量的0.6次方。

2)在数据采集、转换、传输、存储过程中,SCADA系统中会不可避免地产生异常数据,因此需要对数据样本进行预处理。

为保证模型能够学习正常行为的特征和模式,在构建预测模型前利用LOF算法对SCADA系统中采集到的数据进行清洗,以确保其代表机组正常运行状态的数据。LOF算法基于数据点与其邻近点之间的密度关系来检测异常值,其计算公式为:

$$d_{reach, k}(p, s) = \max(d(p, s), d_k(p)) \quad (3)$$

式中: $d_{reach, k}(p, s)$ 为点 s 到点 p 的第 k 可达距离; $d(p, s)$ 为点 p 与点 s 之间的距离; $d_k(p)$ 为点 p 与其第 k 个最近邻点之间的距离。

将点 p 的第 k 可达距离邻域 Q 内所有的点到点 p 的第 k 可达距离的平均值的倒数定义为点 p 的局部可达密度,其计算公式为:

$$d_{lr, k}(p) = \frac{|N_k(p)|}{\sum_{s \in N_k(p)} d_{reach, k}(p, s)} \quad (4)$$

式中: $d_{lr, k}(p)$ 为点 p 的局部可达密度; $N_k(p)$ 为点 p 的第 k 可达距离内所有点的集合; $|N_k(p)|$ 为点 p 的第 k 可达距离内所有点的数量。

LOF异常分数 $L_{OF, k}(p)$ 的计算公式为:

$$L_{OF, k}(p) = \frac{\sum_{s \in N_k(p)} \frac{d_{lr, k}(s)}{d_{lr, k}(p)}}{|N_k(p)|} \quad (5)$$

若 $L_{OF, k}(p) > 1$,则表明点 p 为异常点。

经过上述分析与处理,选择清洗后并且与待检测目标变量相关性较强的状态参数用于模型训练。

1.2 风电机组动态集群方法

由于相似机组和目标机组在正常运行过程中存在相似的运行环境和状态,通过将这些相似性考虑在内,可以更好地捕捉数据中的不确定性,从而提高异常检测的可靠性。因此,相似风电机组的选择将直接影响检测结果。本文将相似机组动态集群分为数据预处理与特征提取、基于时间尺度的动态集群与划分这2个阶段。

在相似机组选择过程中,若仅考虑单一变量进行相似性评估,则无法提供充分的信息来描述机组的全面特性。因此,使用多个变量和多个因素来评估相似性,其评估结果更加可靠。首先,通过SCADA系统获取风电机组的各项运行数据,基于MIC计算变量间的相关性,通过比较平均风速、功率、温度等相关参数指标,选择与待检测目标变量相

关性较高的变量评估风电机组之间运行状态的相似性。

考虑到风电场内的风电机组在不同时刻的运行状态可能存在差异,本文通过对风电机组进行动态集群实现不同时刻相似机组的划分,最大限度地得到相似机组。首先,通过时间尺度动态提取数据,如附录A图A2所示,将分析时段的数据通过尺度为 q 的窗口进行滑动截取,采用 K -means聚类算法对各时段数据集进行聚类,其计算公式见附录A式(A1)、(A2),遍历全时间尺度,得到实时风电机组集群,进而获取运行状态相似的风电机组。

由于 K -means聚类算法在初始聚类中心的选取上具有随机性,为减少这种随机性对聚类结果的影响,本文在选择初始聚类中心时采用了一种系统性策略,通过构建相异度矩阵,确定初始聚类中心,其计算公式见附录A式(A3)~(A5)。此外, K -means聚类算法通常需要人为设定聚类数目,若子集群划分过多,则可能会出现目标风电机组自成一类的情况,导致集群失去整体性;若子集群划分过少,则可能出现不同子集群内风电机组数量差异巨大的情况,导致集群失去平衡。本文通过计算原始数据中的戴维斯-博尔丁指数(Davies-Bouldin index, DBI)^[21]来衡量 K -means聚类算法的最佳聚类数目,最优聚类解决方案具有最小的DBI值。DBI的计算公式为:

$$D_{BI} = \frac{1}{k_s} \sum_{i=1}^{k_s} \max_{j \neq i} \{D_{i,j}\} \quad (6)$$

式中: k_s 为聚类数; D_{BI} 为DBI值; $D_{i,j}$ 为第 i 个和第 j 个簇的簇内到簇间距离比。

2 风电机组状态预测模型的构建

2.1 风电机组状态参数预测模型

2.1.1 KELM

KELM是一种改进的极限学习机(extreme learning machine, ELM),通过引入核函数和正则化函数解决ELM随机的隐藏层节点和输出层权值导致的预测波动,提升模型的拟合能力和预测精度^[22],其目标函数 $f(\mathbf{x})$ 可表示为:

$$f(\mathbf{x}) = \begin{pmatrix} K(\mathbf{x}_1, \mathbf{x}_1) \\ \vdots \\ K(\mathbf{x}_1, \mathbf{x}_n) \end{pmatrix} \left(\frac{I}{C} + \mathbf{\Omega}_{ELM} \right)^{-1} \mathbf{W} \quad (7)$$

式中: $K(\cdot, \cdot)$ 为核函数; \mathbf{x} 为输入向量; C 为正则化系数; I 为单位矩阵; $\mathbf{\Omega}_{ELM}$ 为KELM取代ELM中随机矩阵 $\mathbf{H}\mathbf{H}^T$ 的核函数,其计算过程见附录A式(A6)、(A7); \mathbf{W} 为输出期望。

2.1.2 WLNGO-5CV算法优化KELM

尽管KELM通过引入核函数和正则化系数解决了ELM的输入权值随机初始化问题,获得了稳定的

预测输出,但其预测性能受正则化系数 C 和核参数 σ 的影响较大。为解决这一问题,本节采用WLNGO算法确定KELM的最佳正则化系数 C 和核参数 σ 。

NGO算法的狩猎策略可以分为2个阶段,针对各阶段分别建立数学模型,具体如下。

在北方苍鹰的第1个狩猎阶段,它会随机选定1只猎物然后迅速发动攻击。在此阶段,NGO算法行为的数学表达式为:

$$u_{i,j}^{new,p1} = \begin{cases} u_{i,j} + r(p_{i,j} - Eu_{i,j}) & F_{Pi} < F_i \\ u_{i,j} + r(u_{i,j} - p_{i,j}) & F_{Pi} \geq F_i \end{cases} \quad (8)$$

式中: F_{Pi} 为第 i 只北方苍鹰 P_i 的猎物位置的目标函数值,即适应度值; F_i 为第 i 只北方苍鹰的目标函数值; $u_{i,j}^{new,p1}$ 为第 i 只北方苍鹰在第1个狩猎阶段中第 j 维度的新状态; $u_{i,j}$ 为第 i 只北方苍鹰在位置 j 的状态; r 为属于 $[0,1]$ 的随机数; E 取值为1或2。

在第2个狩猎阶段,北方苍鹰袭击猎物后,猎物会试图逃跑。假设此猎物于半径 R_h 的攻击位置,则第2个狩猎阶段的数学表达式为:

$$\begin{cases} R_h = 0.02 \left(1 - \frac{A}{\tau} \right) \\ u_{i,j}^{new,p2} = u_{i,j} + R_h(2r - 1)u_{i,j} \end{cases} \quad (9)$$

式中: A 为NGO算法第2个狩猎阶段的当前迭代次数; τ 为NGO算法的最大迭代次数; $u_{i,j}^{new,p2}$ 为第 i 只北方苍鹰在第2个狩猎阶段中第 j 维度的新状态。

NGO算法易陷入局部最优,因此为了提高算法的性能,本文对该算法进行改进并提出了WLNGO算法。在NGO算法中的第1个狩猎阶段,通过调整自适应权值可以控制个体在搜索空间中的分布,从而减少盲目跟风并增加搜索多样性。本文采用一种呈指数变化的自适应权重方法:在算法的早期阶段,使用较大的权重实现强大的全局搜索性能,以确保搜索范围;随着迭代次数的增加,当接近最优解时,权重值呈指数递减,从而显著提高了算法的局部搜索能力。自适应权重 w 的计算公式为:

$$w = e^{-(10t/T)^2} \quad (10)$$

式中: t 为WLNGO算法第1个狩猎阶段的当前迭代次数; T 为WLNGO算法的最大迭代次数。

权重 w 的引入使NGO算法在第1个狩猎阶段的位置更新公式由式(8)变为:

$$u_{i,j}^{new,p1} = \begin{cases} u_{i,j} + rw(p_{i,j} - Eu_{i,j}) & F_{Pi} < F_i \\ u_{i,j} + rw(u_{i,j} - p_{i,j}) & F_{Pi} \geq F_i \end{cases} \quad (11)$$

在NGO算法中的第2个狩猎阶段引入Levy飞行策略对北方苍鹰位置进行更新,避免种群陷入局部最优,扩大搜索能力,具体如式(12)所示。

$$L_{Levy}(\beta) = \frac{\mu}{|v|^{\frac{1}{\beta}}} \quad (12)$$

式中: $L_{Levy}(\beta)$ 为 Levy 飞行的步长, $\beta=1.5$; μ, v 服从正态分布, 即 $\mu \sim N(0, \sigma_\mu^2), v \sim N(0, \sigma_v^2), \sigma_\mu$ 和 σ_v 的计算公式见附录 A 式(A8)。

由此可知, NGO 算法中第 2 个狩猎阶段的位置更新公式由式(9)变为:

$$u_{i,j}^{new,p2} = u_{i,j} + R_h(2r-1)u_{i,j} + \alpha L_{Levy}(\beta) \quad (13)$$

式中: α 为步长缩放因子。

虽然 KELM 在建模时通过调整超参数来平衡拟合及泛化性能,但是在模型开发过程中,依赖单一验证集来评估模型性能可能会导致模型在训练数据上表现良好,但在未见过的测试数据上性能不佳,即“过拟合”。因此,为了避免由于数据随机性造成的验证偏差,本文采用 5CV,将训练样本随机均匀分成 5 份互不相交且数量大致相同的子集,分别记为 Fold1—Fold5。然后,依次选择其中 4 份作为训练集,剩余的 1 份作为验证集,进行 5 次独立的模型训练和验证。最后,将 5 次验证结果的平均值作为最终的交叉验证误差。

WLNGO-5CV-KELM 算法的整体流程框架如图 1 所示。

本文将适应度函数 $F_{fitnessl} (l=1, 2, \dots, 5)$ 设计为测试集均方误差(mean-square error, MSE)的最小值,如式(14)所示。

$$F_{fitnessl} = \operatorname{argmin}(M_{SE}) \quad (14)$$

式中: M_{SE} 的计算公式见附录 A 式(A9)。

WLNGO-5CV-KELM 算法的实现流程如下。

步骤 1: 选取风电机组实测数据,根据式(1)—(5)对数据进行预处理并定量分析得到与风电机组

运行状态密切相关的特征参数。

步骤 2: 针对 NGO 算法,在其第 1 个狩猎阶段根据式(10)、(11)引入自适应权重,在其第 2 个狩猎阶段根据式(12)、(13)引入 Levy 飞行策略,得到 WLNGO 算法,从而避免了 NGO 算法陷入局部最优。

步骤 3: 使用 WLNGO 算法生成一系列解(特征向量),然后在原始训练样本集上使用这些解更新训练样本。利用 5CV 训练 KELM 模型,计算每个解的适应度值。借助 WLNGO 算法的优化机制,迭代搜索最优解。

步骤 4: 获取最佳解(最佳特征向量),根据最佳解重新提取训练集和测试集。利用最优超参数重新训练模型,并对测试集进行预测,以评估模型的预测效果。

本文采用均方根误差(root mean square error, RMSE)、平均绝对误差(mean absolute error, MAE)和判定系数 R^2 ,作为衡量模型预测效果可靠性的评价指标,计算公式见附录 A 式(A10)—(A12)。

3 风电机组状态参数异常辨识

3.1 动态残差阈值的划分

风电机组在正常运行时,重构误差在一定范围内波动,而当系统发生异常时,重构误差值有可能超出正常的控制范围,导致重构误差的趋势发生明显变化,越过预设的阈值。因此,通过监测重构误差的趋势是否超过阈值,可以判定风电机组是否发生异常。残差 R_{ec} 的计算公式为:

$$R_{ec} = \hat{y}_c - y_c \quad (15)$$

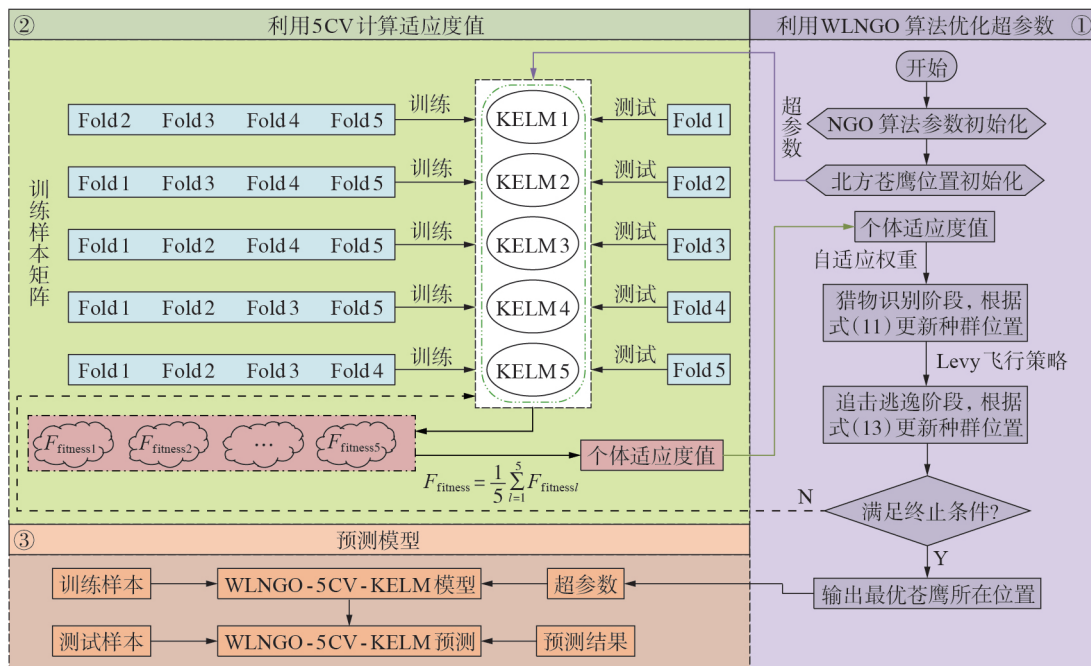


图 1 WLNGO-5CV-KELM 算法的整体流程框架

Fig.1 Overall process framework of WLNGO-5CV-KELM algorithm

式中: c 为时间序列中的序列编号; y_c 为实测值; \hat{y}_c 为预测值。

由于风电机组运行状态受工况影响,在不同风速区间下各部件的运行温度有较大差异,正常运行的风电机组在不同时间段可能会面临不同的风况。即使风电机组正常运行,其运行状态也可能在一定范围内随机变化,并不能表示系统出现异常。在这种情况下,仅通过目标机组正常历史状态下的预测残差来分析风电机组未来运行状态的效果不佳。因此,本文采用滑动时窗对残差进行统计分析,其置信度为 $1-\alpha$ 的置信区间上限为:

$$U_{cd} = \bar{R}_{cc} + \frac{S}{\sqrt{q}} t_{\alpha/2}(q-1) \quad (16)$$

式中: U_{cd} 为第 c 个时间点的第 d 台机组的计算阈值; \bar{R}_{cc} 为窗口中残差均值; S 为窗口中残差值的标准差; $t_{\alpha/2}$ 为 t 分布的 $\alpha/2$ 分位点。

随着时间的推移,对每个窗口中的残差进行统计分析,得到风电机组的动态阈值。最后,利用熵权法将各相似机组的阈值进行融合。

首先,对阈值进行标准化处理,具体公式见附录A式(A13);然后,按照式(17)计算每个标准化阈值的概率 p_{cd} 。

$$p_{cd} = \frac{z_{cd}}{\sum_{c=1}^n z_{cd}} \quad (17)$$

式中: z_{cd} 为标准化处理后的阈值; n 为样本序列中的样本数量。

每个相似机组的熵 E_d 为:

$$E_d = -k' \sum_{c=1}^n p_{cd} \ln p_{cd} \quad (18)$$

式中: $k' = 1/\ln n$,为归一化常数。

根据熵值计算第 d 台机组的权重 w_d ,如式(19)所示。

$$w_d = \frac{1 - E_d}{b_g - \sum_{d=1}^{b_g} E_d} \quad (19)$$

式中: b_g 为相似机组的个数。

使用权重对各相似机组的阈值进行加权,得到融合后的阈值 $U(t)$,如式(20)所示。

$$U(t) = \sum_{d=1}^{b_g} w_d U_{cd} \quad (20)$$

上述方法能够有效考虑数据的不确定性和变化性,更好地适应风电机组运行状态的变化。

3.2 异常状态检测流程

本文提出的异常状态检测流程如图2所示,主要包括以下步骤。

1)利用MIC和LOF算法对SCADA系统采集到的数据进行筛选和清洗。

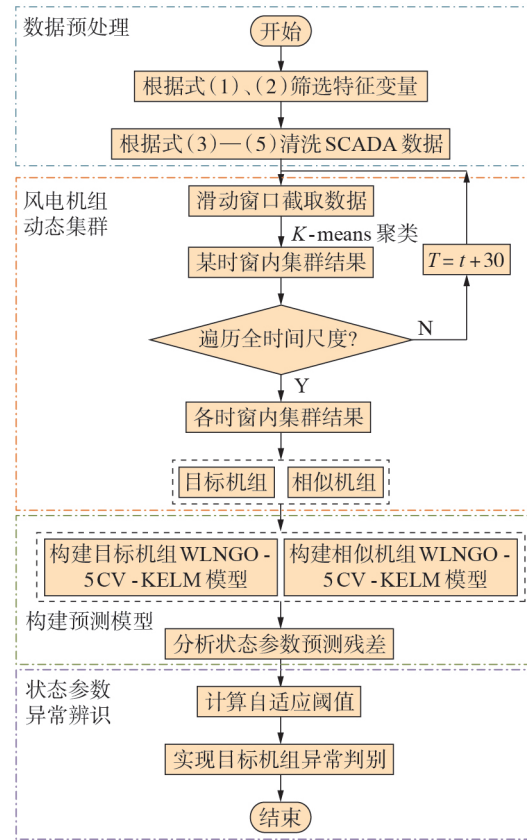


图2 异常状态检测流程图

Fig.2 Flowchart of abnormal state detection

2)在动态时间尺度下,通过滑动时窗提取数据并运用K-means聚类算法进行集群分析,以识别每个时窗中相似的机组。最后,通过获取每个时间窗口中相似机组的交集,得到整个时间范围内的相似机组。

3)利用本文所提WLNGO-5CV-KELM预测模型分别构建目标机组和相似机组的预测模型。

4)通过滑动时窗和统计学相结合的方式,计算相似机组预测残差的动态阈值。借助熵权法对相似机组的阈值进行加权融合,最终实现对目标机组异常状态的检测。

4 仿真实验与分析

为了验证本文所提风电机组性能评估方法的有效性和准确性,以东北某风电场实际风电机组运行数据为例进行动态集群的风电机组异常检测,然后对评估结果进行分析,并与传统评估方法所得结果进行对比。

4.1 风电机组数据预处理与相似机组动态集群

2022年8月16日22:35,SCADA系统记录了1号风电机组(目标机组)发电机输出端轴承温度过热报警。为了分析发电机输出端轴承温度参数在警报前的变化,以故障发生前1.5个月为分析时段。

为了对目标变量进行诊断,需要对SCADA系统采集到的13类状态特征进行筛选。图3展示了根据互信息值大小筛选得到的一些常见状态特征与发电机输出端轴承温度之间的相关性的结果。由图可见,特征7—10,即发电机定子绕组U温度、发电机定子绕组V温度、发电机定子绕组W温度、发电机前端轴承温度与发电机输出端轴承温度相关性较强。

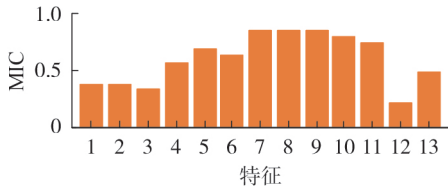


图3 状态特征与发电机输出端轴承温度之间的相关性柱状图

Fig.3 Histogram of correlation between state features and generator output end bearing temperature

考虑到环境对预测结果的影响,本文将风速作为模型输入参数之一,最终确定5个状态特征作为预测模型的输入变量,如附录A表A2所示。在数据采集过程中会产生异常数据,本文通过LOF算法对筛选出的变量进行数据清理,并利用箱线图对结果进行可视化,如附录A图A3所示。

为了验证不同输入变量对预测结果的影响,本文分别仅将风速作为输入变量和将多个变量(包括风速)作为输入变量对模型进行训练,所得模型的预测结果分别如附录A图A4(a)、(b)所示。由图可见,将多个变量(包括风速)作为输入变量对模型进行训练后,预测结果与发电机输出端轴承的实测值更加贴合,这说明训练模型时使用多个变量具有更强的灵敏度。从附录A表A3可以看出,利用包含风速在内的多个变量训练的模型在各项评估指标上的表现优于仅利用风速作为变量训练的模型。

为了筛选出与目标机组运行状态相似的风电机组,利用2022年7月1日—2022年7月31日期间风电场中风电机组的发电机定子绕组温度、发电机前端轴承温度和风速作为K-means聚类模型的输入特征,从附录A图A5可以看出本时段的最佳聚类数目为4,其聚类集群如附录A表A4所示。在此阶段,3、7、9、15号风电机组和目标机组在同一集群内。利用滑动时窗,得到不同时刻下的聚类结果;然后,遍历全时间尺度,得到实时风电机组集群;最后,对每个时间尺度下的相似机组取交集,得到3、7号风电机组为与目标机组运行状态相似的风电机组。

利用本文提出的WLNGO-5CV-KELM回归模型预测3、7号风电机组的发电机输出端轴承温度,结果如附录A图A6所示。由图可见,基于正常历史数据进行聚类得到的相似风电机组在后续的预测过程

中,预测值和实测值的变化趋势基本一致,表明相似机组均处于正常运行状态。因此,选择3、7号风电机组作为相似机组。

4.2 发电机输出端轴承温度预测模型

为验证本文所提方法的优越性,以风电机组发电机输出端轴承温度的测试结果为例,对比分析不同模型对预测精度的影响。WLNGO-5CV-KELM、NGO-KELM^[23]、KELM^[24]模型的测试结果如图4所示。由图可见,本文所提模型的预测结果与实测值最接近,而NGO-KELM、KELM模型的预测结果相对实测值存在一定的偏移。

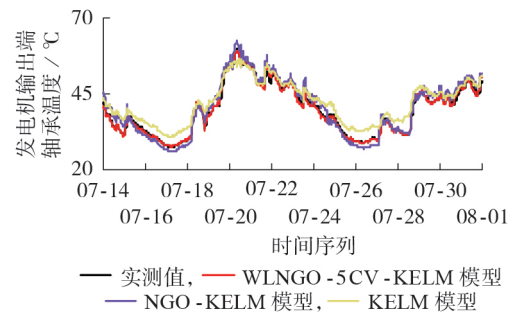
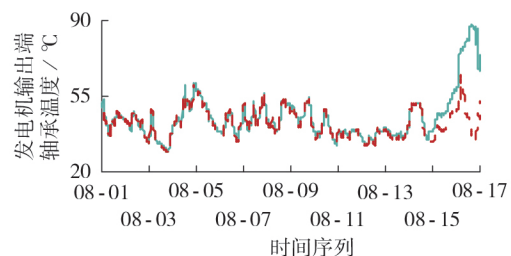


图4 不同模型的对比

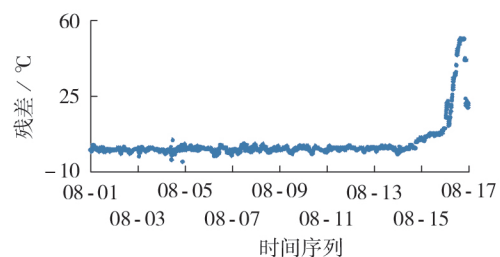
Fig.4 Comparison among different models

计算WLNGO-5CV-KELM、NGO-KELM、KELM模型的MAE、RMSE、 R^2 值,结果如附录A表A5所示。由表可见,本文方法的评价指标均优于其他方法,因此本文所提模型可有效提高预测精度。

利用本文提出的WLNGO-5CV-KELM回归模型预测下一时刻发电机输出端轴承温度,目标机组预测结果如图5所示。由图可见,发电机输出端轴承



(a) 目标机组预测结果



(b) 目标机组预测残差

— 实测值, - - - 预测值, · 残差

图5 目标机组预测结果

Fig.5 Prediction results of target wind turbine

温度在2022年8月15日08:25开始出现偏离正常温度的趋势,在2022年8月16日08:45预测残差开始明显增大。由于风电机组在不同时刻下受风速影响,因此预测残差的突然增大不能反映发电机输出端温度的异常状态,传统的阈值诊断方法可能导致误判。

4.3 基于动态集群相似机组的异常状态检测

基于相似机组的异常状态检测结果如图6(a)所示。由图可见,2022年8月15日10:05检测到异常温度,说明风电机组有开始劣化的趋势,并且在风电机组运行临近故障时也有连续越限情况发生,说明距离故障时刻越近预警越频繁,而日志实际报警时间为2022年8月16日22:35。本文所提基于相似机组的动态阈值法的预警时刻比风电机组实际故障时刻提前了36 h30 min。基于目标机组自身数据划分动态阈值的方法的异常检测结果如图6(b)所示。由图可见,发电机输出端轴承温度在2022年8月16日02:10出现越限情况,检测到异常状态的时刻比本文所提方法晚了16 h5 min。

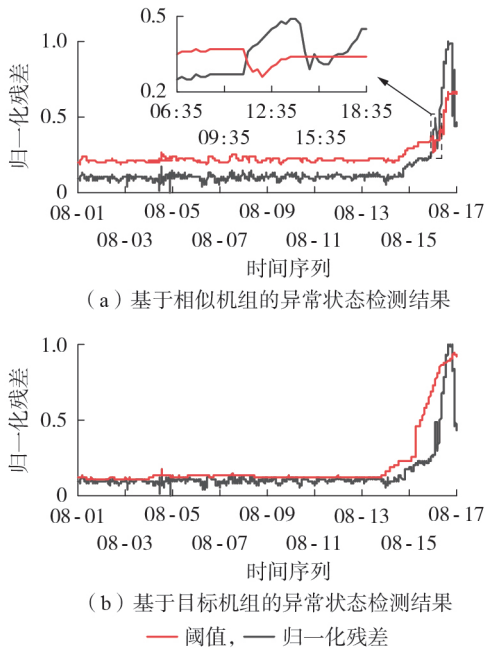


图6 异常状态检测结果

Fig.6 Abnormal state detection results

4.4 异常状态检测算法对比

为了验证本文所提基于动态集群风电机组异常状态检测方法的有效性,使用 3σ 准则^[25]进行阈值划分,如附录A图A7所示。由图可见,利用 3σ 准则计算得到报警阈值 $U(t)=0.6$,故发电机输出端轴承温度在2022年8月15日17:35出现越限。但是风电机组的运行状态可能受到多种因素的影响,这些因素存在复杂的相关性,固定阈值往往难以考虑这种复杂性,导致检测性能下降。

最后,采用基于分位数回归的区间估计模型来检测风电机组的异常状态,结果如附录A图A8所示。由图可见:该方法能够在系统故障报警前检测到温度异常;在2022年8月16日01:15检测到发电机输出端轴承温度超出预测区间,判断机组状态异常;但由于该模型性能上的差异,使得真实值在2022年8月8日00:06超出阈值区间范围,出现误判。因此,风电机组异常状态判断结果除了与判别方法有关,在一定程度上还与预测模型的性能存在一定的关系。

综合上述的多组实验结果可知,本文提出的基于动态集群的风电机组异常状态检测方法具有更多的优势,能够更早地检测到风电机组的异常状态,降低误判情况的发生概率,为维修人员提供充足的抢修时间。

5 结论

本文通过对SCADA系统采集到的数据进行分析,建立了风电机组状态辨识模型,对于检测过程中存在的可靠性问题,利用运行状态相似的风电机组预测结果残差分布对目标机组异常状态进行检测,进一步提高异常状态检测的准确性,并基于东北某风电场的实际数据进行仿真验证和分析。本文所得结论如下。

1) 本文通过对KELM的超参数进行优化,提出了WLNGO-5CV-KELM预测模型,有效提高了模型的预测精度。

2) 受环境与工况因素影响,仅考虑单一变量在整个时间段对风电机组进行聚类,其聚类结果划分不精确。通过滑动时窗对风电机组的多个相关变量进行动态集群,能够实现更加精细的风机集群划分。

3) 综合考虑风电机组运行环境和自身参数,通过动态集群获取运行状态相似的风电机组,并对其预测残差进行阈值划分,作为目标机组的判别阈值,提高了风电机组异常状态检测的可靠性。

4) 基于实际运行数据对本文所提方法的有效性进行仿真验证,结果表明:考虑机组间的时空相关性进行动态集群划分,并利用相似机组计算动态阈值,可以有效提高风电机组异常状态的检测能力。

附录见本刊网络版(<http://www.epae.cn>)。

参考文献:

- [1] 刘璐洁,李传文,符杨,等. 考虑台风影响的海上风电机组部件剩余寿命预测方法[J]. 电力自动化设备,2024,44(6):35-42. LIU Lujie, LI Chuanwen, FU Yang, et al. Residual life prediction of offshore wind turbine components considering typhoon effects[J]. Electric Power Automation Equipment, 2024, 44(6): 35-42.
- [2] 唐贵基,朱星皓,王晓龙,等. 基于VEITD和OSMHD的风电机组轴承损伤识别[J]. 电力自动化设备,2023,43(6):101-107.

- TANG Guiji, ZHU Xinghao, WANG Xiaolong, et al. Wind turbine bearing damage identification based on VEITD and OSMHD[J]. Electric Power Automation Equipment, 2023, 43(6): 101-107.
- [3] 阮益闽, 宗启航, 姚伟, 等. 计及典型控制的风电场调频能力量化评估及影响因素分析[J]. 电力系统自动化, 2024, 48(8): 42-52.
- RUAN Yimin, ZONG Qihang, YAO Wei, et al. Quantitative assessment and analysis of influencing factors on frequency regulation capability of wind farms considering typical control[J]. Automation of Electric Power Systems, 2024, 48(8): 42-52.
- [4] BEBARS A, ELADL A A, ABDULSALAM G M, et al. Internal electrical fault detection techniques in DFIG-based wind turbines: a review[J]. Protection and Control of Modern Power Systems, 2022, 7(2): 1-22.
- [5] BANGALORE P, TJERNBERG L B. An artificial neural network approach for early fault detection of gearbox bearings[J]. IEEE Transactions on Smart Grid, 2015, 6(2): 980-987.
- [6] 金晓航, 泮恒拓, 许壮伟, 等. 基于SCADA数据和单分类简化核极限学习机的风电机组发电机状态监测[J]. 计算机集成制造系统, 2022, 28(8): 2408-2418.
- JIN Xiaohang, PAN Hengtuo, XU Zhuangwei, et al. Condition monitoring of wind turbine generators using SCADA data and OC-RKELM[J]. Computer Integrated Manufacturing Systems, 2022, 28(8): 2408-2418.
- [7] RENSTRÖM N, BANGALORE P, HIGHCOCK E. System-wide anomaly detection in wind turbines using deep autoencoders[J]. Renewable Energy, 2020, 157: 647-659.
- [8] 马良玉, 程善珍. 基于支持向量数据描述和XGBoost的风电机组异常工况预警研究[J]. 电工技术学报, 2022, 37(13): 3241-3249.
- MA Liangyu, CHENG Shanzhen. Abnormal state early warning of wind turbine generator based on support vector data description and XGBoost[J]. Transactions of China Electrotechnical Society, 2022, 37(13): 3241-3249.
- [9] 吴永斌, 张建忠, 袁正舫, 等. 风电场风功率异常数据识别与清洗研究综述[J]. 电网技术, 2023, 47(6): 2367-2380.
- WU Yongbin, ZHANG Jianzhong, YUAN Zhengxi, et al. Review on identification and cleaning of abnormal wind power data for wind farms[J]. Power System Technology, 2023, 47(6): 2367-2380.
- [10] 王桂兰, 赵洪山, 米增强. XGBoost算法在风机主轴轴承故障预测中的应用[J]. 电力自动化设备, 2019, 39(1): 73-77, 83.
- WANG Guilian, ZHAO Hongshan, MI Zengqiang. Application of XGBoost algorithm in prediction of wind motor main bearing fault[J]. Electric Power Automation Equipment, 2019, 39(1): 73-77, 83.
- [11] 魏书荣, 张鑫, 符杨, 等. 基于GRA-LSTM-Stacking模型的海上双馈风力发电机早期故障预警与诊断[J]. 中国电机工程学报, 2021, 41(7): 2373-2383.
- WEI Shurong, ZHANG Xin, FU Yang, et al. Early fault warning and diagnosis of offshore wind DFIG based on GRA-LSTM-Stacking model[J]. Proceedings of the CSEE, 2021, 41(7): 2373-2383.
- [12] 陈俊生, 李剑, 陈伟根, 等. 采用滑动窗口及多重加噪比堆栈降噪自编码的风电机组状态异常检测方法[J]. 电工技术学报, 2020, 35(2): 346-358.
- CHEN Junsheng, LI Jian, CHEN Weigen, et al. A method for detecting anomaly conditions of wind turbines using stacked denoising autoencoders with sliding window and multiple noise ratios[J]. Transactions of China Electrotechnical Society, 2020, 35(2): 346-358.
- [13] ZHU Yongchao, ZHU Caichao, TAN Jianjun, et al. Anomaly detection and condition monitoring of wind turbine gearbox based on LSTM-FS and transfer learning[J]. Renewable Energy, 2022, 189: 90-103.
- [14] ZHANG Chen, HU Di, YANG Tao. Anomaly detection and diagnosis for wind turbines using long short-term memory-based stacked denoising autoencoders and XGBoost[J]. Reliability Engineering & System Safety, 2022, 222: 108445.
- [15] CHEN Hansi, LIU Hang, CHU Xuening, et al. Anomaly detection and critical SCADA parameters identification for wind turbines based on LSTM-AE neural network[J]. Renewable Energy, 2021, 172: 829-840.
- [16] ZHANG Fanghong, ZHU Yuze, ZHANG Chuanjiang, et al. Abnormality detection method for wind turbine bearings based on CNN-LSTM[J]. Energies, 2023, 16(7): 3291.
- [17] 李泽宇, 郭创新, 朱承治. 采用GA-BPNN与TLS模型的风电机组异常辨识方法[J]. 电力系统自动化, 2020, 44(9): 95-102.
- LI Zeyu, GUO Chuangxin, ZHU Chengzhi. Anomaly identification method of wind turbine based on genetic algorithm-back propagation neural network and t-location scale model[J]. Automation of Electric Power Systems, 2020, 44(9): 95-102.
- [18] ZENG Xiangjun, YANG Ming, FENG Chen, et al. A generalized wind turbine anomaly detection method based on combined probability estimation model[J]. Journal of Modern Power Systems and Clean Energy, 2023, 11(4): 1136-1148.
- [19] 孙鹏, 李剑, 寇晓涛, 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.
- SUN Peng, LI Jian, KOU Xiaoshi, et al. Wind turbine status parameter anomaly detection based on prediction models and fuzzy theory[J]. Electric Power Automation Equipment, 2017, 37(8): 90-98.
- [20] 曾祥军, 冯琛, 杨明, 等. 考虑运行状态相似性的风电机组数据异常检测方法[J]. 电力系统自动化, 2022, 46(11): 170-180.
- ZENG Xiangjun, FENG Chen, YANG Ming, et al. Data anomaly detection method for wind turbines considering operation state similarity[J]. Automation of Electric Power Systems, 2022, 46(11): 170-180.
- [21] YANG Jingxian, ZHANG Shuai, XIANG Yue, et al. LSTM auto-encoder based representative scenario generation method for hybrid hydro-PV power system[J]. IET Generation Transmission & Distribution, 2020, 14(24): 5935-5943.
- [22] 叶永盛, 徐燕龙, 李阳, 等. 基于相似周和模态分解的融合模型电动汽车负荷预测[J/OL]. 电网技术. (2024-05-17)[2024-07-22]. <https://doi.org/10.13335/j.1000-3673.pst.2024.0225>.
- [23] LI Zhailong, SUN Benchao, ZHU Wenming, et al. Thermal error modeling of motorized spindle and application of miniature radiator in motorized spindle[J]. The International Journal of Advanced Manufacturing Technology, 2024, 131(3): 1107-1118.
- [24] 宋佳, 石若凌, 郭小红, 等. 基于核极限学习机的飞行器故障诊断方法[J]. 清华大学学报(自然科学版), 2020, 60(10): 795-803.
- SONG Jia, SHI Ruoling, GUO Xiaohong, et al. KELM based diagnostics for air vehicle faults[J]. Journal of Tsinghua University(Science and Technology), 2020, 60(10): 795-803.
- [25] WANG Shaoko, ZHANG Zhaoyan, WANG Penguang, et al. Failure warning of gearbox for wind turbine based on 3σ -median criterion and NSET[J]. Energy Reports, 2021, 7: 1182-1197.

作者简介:

于华楠(1981—),女,副教授,博士,研究方向为电力系统信号分析与检测、压缩感知在电力系统中的应用等(E-mail: yhn810117@163.com);

李靖雨(1999—),女,硕士研究生,研究方向为风电机组异常检测(E-mail: 1193994933@qq.com);

王鹤(1983—),男,教授,博士,研究方向为智能电网技术、新能源发电并网技术、继电保护技术等(E-mail: wanghe_nedu@163.com)。

(编辑 任思思)

(下转第94页 continued on page 94)

Dual-end traveling wave fault location method for deep-sea offshore wind-integrating system via MMC-HVDC using instantaneous energy spectrum of wavefront of backward traveling wave

LIU Le¹, CHEN Xuming¹, KANG Xiaoning¹, MA Xiaowei^{1,2}, LI Shichuang¹,
ZHAO Boyang¹, LI Xinying¹, LIU Xin²

(1. School of Electrical Engineering, Xi'an Jiaotong University, Xi'an 710049, China;
2. Northwest Branch of State Grid Corporation of China, Xi'an 710048, China)

Abstract: The accuracy and reliability of existing traveling wave fault location methods are significantly affected by factors such as the protection sampling frequency, strong noise interference, short fault distance and high transition resistance. To address these issues, a dual-end traveling wave fault location method based on wavelet adaptive threshold denoising (AWTD) and variational mode decomposition (VMD) combined with Hilbert transform is proposed. The AWTD algorithm is utilized to preprocess the fault traveling wave data for noise reduction. Then, the VMD algorithm is applied to extract high-frequency intrinsic mode functions (IMFs) containing fault distance information. The fifth-level IMF is processed using the Hilbert transform to obtain its instantaneous energy spectrum. By utilizing the maximum value of the instantaneous energy spectrum, the traveling wave arrival times at both ends are achieved, thus obtaining precise timing for the arrival of traveling waves at relay points, so the fault distance can be predicted by combining the traveling wave velocities at both ends of the line. Extensive tests are carried out by dual-end and three-end typical deep-sea wind power integrated grid models established in the PSCAD/EMTDC and RTDS simulation platforms, the results demonstrate that the proposed fault location method is not affected by fault resistance or fault type, and can achieve accurate fault localization under different sampling frequencies, near-end faults, strong noise interference and real-time simulation environments. Therefore, the proposed method has certain engineering application value.

Key words: deep-sea offshore winds; traveling wave based fault location; wavelet adaptive threshold denoising; variational mode decomposition; Hilbert transform; instantaneous energy spectrum

(上接第71页 continued from page 71)

Abnormal state detection method based on dynamic clustering of wind turbines

YU Huanan, LI Jingyu, WANG He, LI Shiqiang, BIAN Jing

(Key Laboratory of Modern Power System Simulation and Control & Renewable Energy Technology Ministry of Education,
Northeast Electric Power University, Jilin 132012, China)

Abstract: Aiming at the problem of abnormal state detection of wind turbine (WT), a method of abnormal state detection of WT considering the operating state of similar WTs is proposed. Based on sliding time-window and *K*-means clustering algorithm, the WT operation data are analyzed, the dynamic clustering method of WTs is proposed, and then the WT clustering considering spatio-temporal correlation is established. The WLNGO (adaptive weight and Levy flight based northern goshawk optimization) algorithm is proposed. The 5-fold cross validation (5CV) is used to improve WLNGO algorithm, and the WLNGO-5CV algorithm is proposed, which is used to optimize the hyperparameters of the kernel extreme learning machine (KELM), and the WLNGO-5CV-KELM regression model is further proposed. Then, the real-time warning threshold is obtained by combining the sliding time-window with the statistical analysis of the prediction residuals of similar WTs, which eliminates the influence of working conditions and other factors on WTs, and is able to reliably detect the anomalies of target WT. The effectiveness and accuracy of the proposed method are verified by simulation analysis of actual data from a wind farm in Northeast China.

Key words: wind turbines; WLNGO-5CV-KELM regression model; spatio-temporal correlation; dynamic clustering; abnormal state detection; supervisory control and data acquisition system

附录 A

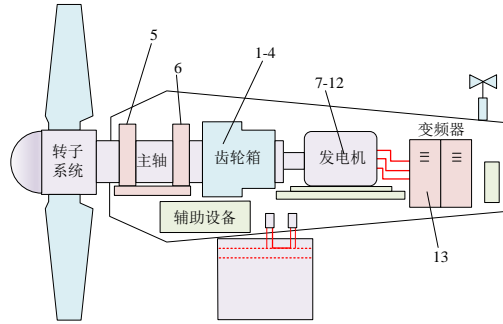


图 A1 风电机组的主要组件

Fig.A1 Main components of wind turbine

表 A1 风电机组状态参数

Table A1 State parameters of wind turbine

| 编号 | 机组状态参数 | 变量符号 | 单位 |
|----|--------------|----------|--------------------|
| 1 | 齿轮箱输入轴温度 | T_i | $^{\circ}\text{C}$ |
| 2 | 齿轮箱输出轴温度 | T_o | $^{\circ}\text{C}$ |
| 3 | 齿轮箱油温 | T_1 | $^{\circ}\text{C}$ |
| 4 | 齿轮箱冷却水温度 | T_2 | $^{\circ}\text{C}$ |
| 5 | 主轴承温度（转子侧） | T_3 | $^{\circ}\text{C}$ |
| 6 | 主轴承温度（齿轮箱侧） | T_4 | $^{\circ}\text{C}$ |
| 7 | 发电机绕组温度（U 相） | T_{u1} | $^{\circ}\text{C}$ |
| 8 | 发电机绕组温度（V 相） | T_{v1} | $^{\circ}\text{C}$ |
| 9 | 发电机绕组温度（W 相） | T_{w1} | $^{\circ}\text{C}$ |
| 10 | 发电机前端轴承温度 | T_a | $^{\circ}\text{C}$ |
| 11 | 发电机冷却风温度 | T_{ai} | $^{\circ}\text{C}$ |
| 12 | 机舱电柜温度 | T_5 | $^{\circ}\text{C}$ |
| 13 | 输出功率 | P | kW |

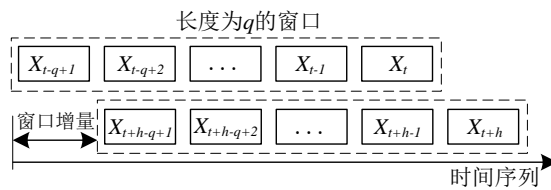


图 A2 滑动窗口示意图

Fig.A2 Schematic of sliding window

图中： $[X_t, \dots, X_{t+q+1}]$ 为在 t 时刻的窗口； $[X_{t+h+q+1}, \dots, X_{t+h}]$ 为 t 时刻的下一窗口； h 为窗口增量。

表 A2 输入变量

Table A2 Input variables

| 参数名称 | 变量符号 | 单位 |
|--------------|----------|------------------------------|
| 发电机绕组温度（U 相） | T_{u1} | $^{\circ}\text{C}$ |
| 发电机绕组温度（V 相） | T_{v1} | $^{\circ}\text{C}$ |
| 发电机绕组温度（W 相） | T_{w1} | $^{\circ}\text{C}$ |
| 发电机前端轴承温度 | T_a | $^{\circ}\text{C}$ |
| 风速 | V_0 | $\text{m}\cdot\text{s}^{-1}$ |

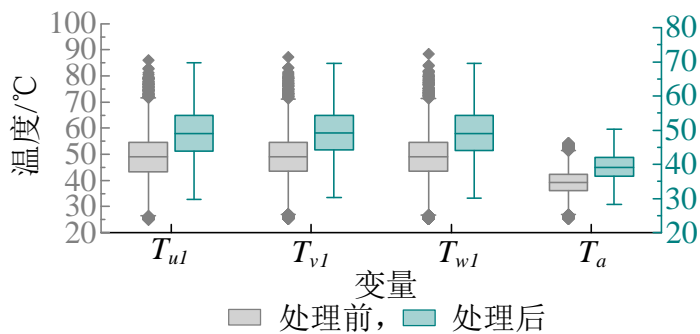
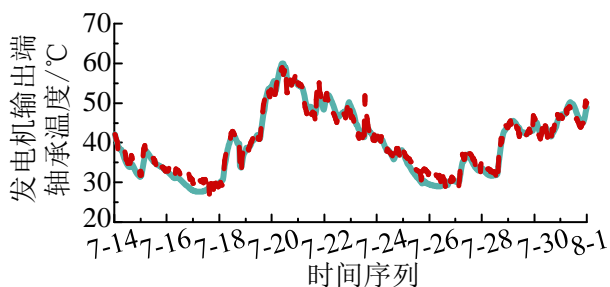
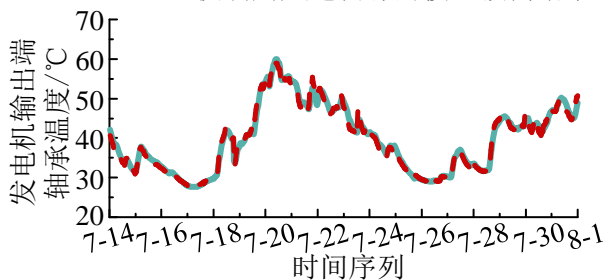


图 A3 数据清洗前后变量箱线图

Fig.A3 Box diagram of variables before and after data cleaning



(a) 仅利用风速训练的模型预测结果



(b) 利用包含风速在内的多个变量的预测结果

— 实测值, - - 预测值

图 A4 基于验证数据的发电机输出端轴承温度预测结果

Fig.A4 Prediction results of generator output end bearing temperature based on validation data

表 A3 不同输入变量训练模型对比

Table A3 Comparison between training models with different input variables

| 方法 | MAE | RMSE | R ² |
|-----------|------|------|----------------|
| 仅利用风速训练模型 | 0.49 | 0.64 | 0.89 |
| 利用多变量训练模型 | 0.17 | 0.22 | 0.98 |

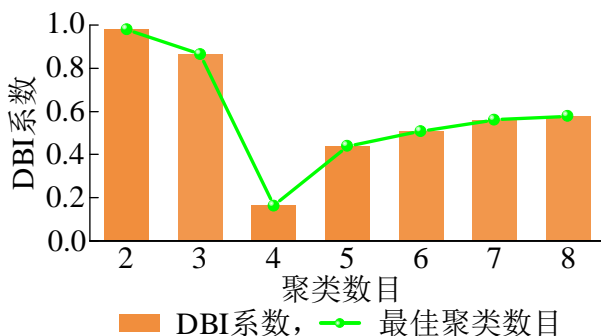
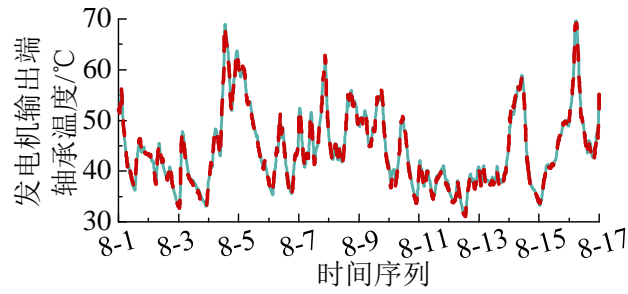


图 A5 机组聚类数目

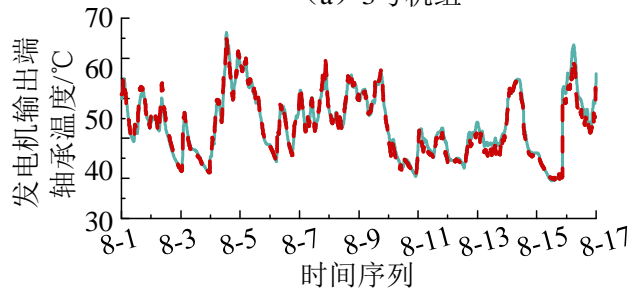
Fig.A5 Cluster number of wind turbines

表 A4 相似机组的选择
Table A4 Selection of similar wind turbines

| 序号 | 集群 | 机组编号 |
|----|------|----------------|
| 1 | 集群 1 | 1、3、7、9、15 |
| 2 | 集群 2 | 2、5、11、12、17 |
| 3 | 集群 3 | 8、4、6、14、19 |
| 4 | 集群 4 | 10、13、16、18、20 |



(a) 3号机组



(b) 7号机组

— 实测值, - - 预测值

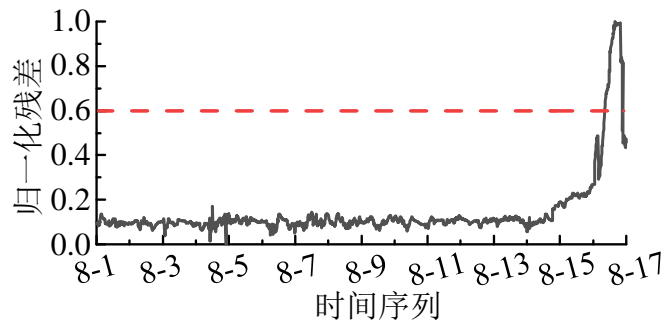
图 A6 相似机组预测结果

Fig.A6 Prediction results of similar wind turbines

表 A5 不同模型的精度对比

Table A5 Comparison of accuracy among different models

| 模型 | MAE | RMSE | R^2 |
|----------------|------|------|-------|
| WLNGO-5CV-KELM | 0.17 | 0.22 | 0.98 |
| NGO-KELM | 0.51 | 0.62 | 0.82 |
| KELM | 0.77 | 0.81 | 0.79 |



— 归一化残差, - - 文献[25]

图 A7 异常状态预警图

Fig.A7 Abnormal status warning chart

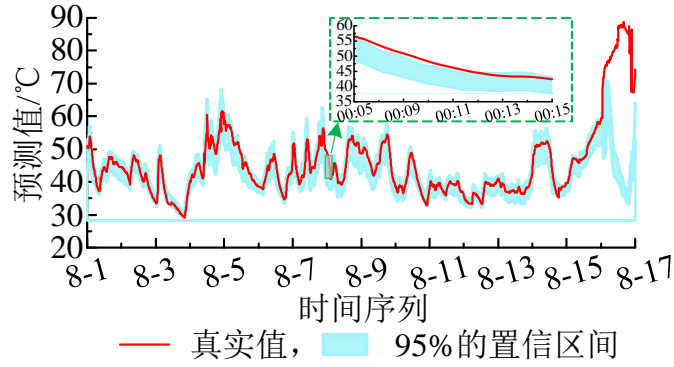


图 A8 分位数回归区间估计图

Fig.A8 Graph of interquartile regression interval estimation

K -means 聚类算法的计算公式为:

$$d(l, C_i) = \sqrt{\sum_{j=1}^m (l_j - C_{ij})^2} \quad (A1)$$

$$a_j = \frac{1}{|C_i|} \sum_{l \in C_i} l \quad (A2)$$

式中: l 为数据对象; C_i 为第 i 个聚类中心; m 为数据对象的维度; l_j 、 C_{ij} 分别为 l 、 C_i 的第 j 个属性值; a_j 为更新后的聚类中心。

给定数据集 $X_{\gamma \times \lambda} = \{x_1, x_2, \dots, x_\gamma\}$, 其中, $x_i = [x_{i1}, x_{i2}, \dots, x_{i\lambda}]$, γ 为数据集样本总数, λ 为数据集样本维度。以综合相似性距离定义 2 个样本间的相异性, 记为 $\text{dis}(x_i, x_j)$, 样本间相异度矩阵 M_{dis} 为:

$$M_{\text{dis}} = \begin{bmatrix} 0 & \text{dis}(x_1, x_2) & \cdots & \text{dis}(x_1, x_\gamma) \\ \text{dis}(x_2, x_1) & 0 & \cdots & \text{dis}(x_2, x_\gamma) \\ \vdots & \vdots & \ddots & \vdots \\ \text{dis}(x_\gamma, x_1) & \text{dis}(x_\gamma, x_2) & \cdots & 0 \end{bmatrix} \quad (A3)$$

均值相异性是数据点 x_i 与数据集中每个对象的距离平均值, 记为 $A_{\text{dis}}(x_i)$, 如式 (A4) 所示。

$$A_{\text{dis}}(x_i) = \frac{1}{\gamma} \sum_{j=1}^{\gamma} \text{dis}(x_i, x_j) \quad (A4)$$

数据集的总体相异性与全体数据的分布有关, 体现了数据集的稀疏程度, 数据集的总体相异性定义如式 (A5) 所示。

$$T_{\text{dis}} = \frac{1}{\gamma^2} \sum_{i=1}^{\gamma} \sum_{j=1}^{\gamma} \text{dis}(x_i, x_j) \quad (A5)$$

结合式 (A3) — (A5), 确定聚类中心方法的步骤为: 首先, 将初始相异度矩阵中均值相异性最大的点选为第 1 个初始聚类中心点; 然后, 去除已选初始点后, 在新的相异度矩阵中选择均值相异性最大的点进行判断; 最后, 如果该点与已选初始点的相异性大于总体相异性, 则将其作为第 2 个聚类中心; 如果相异性小于总体相异性, 则选择均值相异性第二大的样本进行判断, 直至选出所有初始点。

核函数计算式为:

$$\Omega_{\text{EIM}} = \mathbf{H}\mathbf{H}^T = h(\mathbf{x}_i)h(\mathbf{x}_j) = K(\mathbf{x}_i, \mathbf{x}_j) \quad (A6)$$

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\sigma^2}\right) \quad (A7)$$

式中: \mathbf{x}_i 、 \mathbf{x}_j 为输入向量; σ 为核函数 $K(\mathbf{x}_i, \mathbf{x}_j)$ 的参数因子。

σ_μ 和 σ_v 的计算式为:

$$\begin{cases} \sigma_\mu = \left\{ \frac{\Gamma(1+\beta) \sin(\pi\beta/2)}{\Gamma[(1+\beta)/2] \beta 2^{(\beta-1)/2}} \right\}^{1/\beta} \\ \sigma_v = 1 \end{cases} \quad (\text{A8})$$

式中: Γ 为 Gamma 函数。

均方误差的计算公式为:

$$M_{SE} = \frac{1}{\xi} \sum_{i=1}^{\xi} (\hat{y}_i - y_i)^2 \quad (\text{A9})$$

式中: i 为采样点; ξ 为采样点个数。 M_{SE} 误差越小表明预测的数据与原始数据重合度越高。

均方根误差的计算公式为:

$$R_{MSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (\text{A10})$$

平均绝对误差的计算公式为:

$$M_{AE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (\text{A11})$$

判定系数的计算公式见式 (A12), 其值越接近 1, 表示模型的预测效果越好。

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2} \quad (\text{A12})$$

式中: \bar{y}_i 为实测值的平均值。

阈值标准化处理的公式为:

$$z_{cd} = \frac{U_{cd} - \min(U_d)}{\max(U_d) - \min(U_d)} \quad (\text{A13})$$

式中: z_{ij} 为标准化处理后的阈值; $\min(U_j)$ 和 $\max(U_j)$ 分别为第 j 个相似机组的最小值和最大值。