

DOI: 10.19650/j.cnki.cjsi.J2412941

基于两阶段多维数据生成与实时健康指数的 风机齿轮箱故障预警*

马同旭¹, 刘帅^{2,3}, 刘卫亮^{2,3}, 刘长良^{2,3}

(1. 华北电力大学 控制与计算机工程学院 保定 071003; 2. 华北电力大学 保定市综合能源系统状态检测与优化调控
重点实验室 保定 071003; 3. 华北电力大学 河北省发电过程仿真与优化控制技术创新中心 保定 071003)

摘要:针对风电机组关键部件维护时准确率与虚警率难以平衡的问题,提出一种基于两阶段多维数据生成与实时健康指数的风电机组齿轮箱故障预警方法。首先,为反映风速对机组运行状态影响,提出一种基于历史风速的高斯混合隐马尔科夫模型对风速进行短期预测;然后,为提高预警准确度,提出一种基于两阶段多维数据生成的实时动态阈值设定方法,依据预测风速序列和生成器形成当前时刻油温阈值区间;最后,综合齿轮箱油温实际值和健康状态判别器输出,确定风电机组齿轮箱是否处于异常状态。实际数据的仿真结果表明,所提方法可有效降低虚警率,提前17 h预警风电机组齿轮箱潜在故障。

关键词:风电机组;数据采集与监控系统;生成对抗网络;隐马尔科夫模型;风速预测;故障预警

中图分类号: TH17 **文献标识码:** A **国家标准学科分类代码:** 460.99

Fault warning of wind turbine gearbox based on two-stage multidimensional data generation and real-time health index

Ma Tongxu¹, Liu Shuai^{2,3}, Liu Weiliang^{2,3}, Liu Changliang^{2,3}

(1. School of Control and Computer Engineering, North China Electric Power University, Baoding 071003, China;
2. Baoding key Laboratory of State Detection and Optimization Regulation for Integrated Energy System, North China Electric Power
University, Baoding 071003, China; 3. Hebei Technology Innovation Center of Simulation & Optimized Control for
Power Generation, North China Electric Power University, Baoding 071003, China)

Abstract: Aiming at the problem that the accuracy rate and false alarm rate are difficult to balance during the maintenance of key components of wind turbines, a wind turbine gearbox fault early warning method based on two-stage multi-dimensional data generation and real-time health index is proposed. To account for the impact of wind speed on the unit's operational state, a Gaussian Mixture Hidden Markov Model leveraging historical wind speed data is proposed to forecast short-term wind speed. Next, to enhance early warning accuracy, a real-time dynamic threshold-setting method utilizing a two-stage multi-dimensional data generation process is introduced. Based on the predicted wind speed sequence and generator data, the threshold interval for the current oil temperature is established. Finally, the actual gearbox oil temperature values and the output of the health status discriminator are integrated to assess whether the wind turbine gearbox is in an abnormal state. Simulation results using real-world data demonstrate that the proposed method significantly reduces the false alarm rate and provides a potential fault warning for the wind turbine gearbox up to 17 hours in advance.

Keywords: wind turbine; supervisory control and data acquisition system; generating adversarial network; hidden Markov model; wind speed prediction; fault warning

0 引言

随着风电装机总量的逐年增加,风能资源分布不平衡,风电机组运行环境普遍恶劣,风电机组的故障率较高^[1]、运维成本占比较大。且如今大量风电机组投运多年,进入故障高发期。为了减少运维成本,提高风电机组的运行可靠性,优化风电场的运维策略,有必要对风电机组进行故障预警,提前获知风电机组潜在的故障情况,减少经济损失^[2]。

目前,投运多年的风电机组绝大多数为双馈式异步风电机组,齿轮箱是其重要的易损部件。齿轮箱油温与可预测的缓变性故障的发生有强关联性,专家学者对基于油温预测的残差分析技术做出了大量研究。文献[3]基于卷积自编码和双向长短时神经网络的特征选择特性和非线性特性构建了油温预测模型,对预测值和残差进行分析,设置自适应阈值监测异常状态的趋势变化。文献[4]融合了数据采集与监控系统(supervisory control and data acquisition system, SCADA)数据与状态监测系统特征参数,建立了包含齿轮箱油温的主传动系统多参数预测模型,采用多残差融合方法对主传动系统状态进行评估。文献[5]提出一种蛭螂算法优化门控循环网络的油温预测模型,随机森林袋外估计算法提取特征变量,蛭螂优化算法优化模型超参数。文献[6]通过构建 Stacking 异源集成模型,充分发挥了不同算法的优势,预测齿轮箱油温时具有更高的拟合优度和更小的预测误差。上述文献虽然可以对齿轮箱油温进行准确地预测,但预警策略单一,且残差分析技术易受噪声的干扰,误报警次数较多。

生成对抗网络(generative adversarial network, GAN)是 1 种无监督学习方法,可以从未标记的数据中学习分布的模式和特征,并生成与训练数据相似的高质量样本,同时即使外部噪声的干扰,训练后的判别器输出具有很好的稳定性。文献[7]基于 GAN 对健康数据进行学习,训练出可区分机组健康状态的判别器模型,基于判别器结果对整机状态进行监测。文献[8]设计了改进的带有传输层的自编码网络作为生成器,以消除模糊数据对故障预警的影响,并引入最小二乘法学习健康数据的特征分布,但其采用固定阈值和时间长度进行预警,仍有改进空间。文献[9]通过深度卷积 GAN 模型中判别器的输出,自动创建健康状态监测的阈值,计算样本差异分析的指标,为风电机组轴承故障严重程度提供了定量指标。文献[10]将长短时记忆网络作为 GAN 框架的基本网络结构,学习时序数据的时间相关性,基于生成数据的重构误差和判别器输出,提出异常检测指标,基于多个公开数据集的测试结果表明:所提的 GAN 框架可以实现基于时

序数据的异常监测。

囿于大多数基于 GAN 的文献仅使用判别器的判别结果作为状态监测的唯一指标,未充分挖掘生成器生成多维数据的能力,无法对具体部件或变量进行故障预警的现状,本文提出一种基于两阶段多维数据生成和实时健康指数的方法,采用隐马尔科夫模型对风速进行短期预测;根据预测风速,基于最小二乘条件生成对抗网络(least squares conditional generative adversarial networks, LSCGAN)模型生成对应预测风速下的多维时间序列数据;最后,综合齿轮箱油温实际值和判别器输出,计算齿轮箱状态健康指数,进而确定风电机组齿轮箱是否处于异常状态,可以在实现精确预警风电机组齿轮箱故障的同时,有效降低虚警率。

1 算法基础

1.1 隐马尔科夫模型

马尔科夫链是一个在状态空间中从一个状态转换到另一个状态的随机过程^[11]。这一过程必须具有“无记忆”特性:即由当前状态决定下一状态的概率分布,当前状态之前的事件均与下一状态无关。基于马尔科夫链发展而来的隐马尔科夫模型(hidden Markov model, HMM),被用来研究一系列的隐藏状态。该模型提出基于马尔科夫链生成一组不可见的状态序列,并从该不可见的状态序列中生成一组可见的状态序列,以上是两个随机的序列生成过程^[12]。其中基于马尔科夫链所产生的不可见序列称为状态序列,用状态转移概率分布来表征各状态之间的相互联系;由状态序列生成的可见序列被称为观测序列,用观测概率分布来描述状态和观测数据的关系。故隐马尔可夫模型由观测概率分布、初始概率分布及状态转移概率分布所确定的^[13]。初始状态概率向量 π_i 表示在 $t = 1$ 时处于状态 q_i 的概率,其表示为:

$$\pi_i = p(i_1 = q_i) \quad (1)$$

A 是状态转移概率矩阵,表示系统在 t 时刻处在状态 q_i 的条件下,在 $t+1$ 时刻转换到状态 q_j 的概率,其表示为:

$$A = [a_{ij}]_{N \times N}, a_{ij} = P(i_{t+1} = q_j | i_t = q_i) \quad (2)$$

B 是观测概率矩阵,表示 t 时刻处在状态 q_j 的条件下,产生观测值 v_k 的概率,其表示为:

$$B = [b_j(k)]_{N \times M}, b_j(k) = P(o_t = v_k | i_t = q_j) \quad (3)$$

π 、 A 可以确定状态序列, B 可以确定观测序列。 π 与 A 决定了不可见的马尔科夫链,产生一组不可见的状态, B 决定如何由隐藏状态产生可见观测值,并与隐状态序列共同决定观测序列。

1.2 条件生成对抗网络

CGAN (conditional generative adversarial network,

CGAN)是1种生成对抗网络的变体,用于生成具有特定条件的样本,在生成器和判别器之间引入附加条件信息,以便生成更具控制性的输出。在CGAN中,生成器接收条件向量,它可以是任何形式的条件信息,例如类别标签、文本描述等,生成器利用这个条件向量生成与之相关的样本。判别器负责将生成器生成的样本与真实样本进行区分,并提供反馈给生成器以进行改进。通过对抗训练,CGAN能够生成与条件匹配的高质量样本,提供了更好的控制能力和样本多样性。

CGAN 优化目标函数 $U(D,G)$ 的表达式为:

$$\min_G \max_D U(D,G) = E_{x \sim p_x(x)} [\log D(x|c)] + E_{z \sim p_z(z)} [\log (1 - D(G(z|c)))] \quad (4)$$

其中, E 为均值; $x \sim p_x(x)$ 表明 x 服从真实数据分布; c 为条件信息; $z \sim p_z(z)$ 表明 z 服从随机数据分布。

2 故障预警方法

2.1 基于高斯混合隐马尔科夫模型的风速预测

本文选择历史风速数据作为隐马尔科夫模型的数据集。考虑到风速是一个连续随机变量,故本文采用高斯混合隐马尔科夫模型 (Gaussian mixture hidden Markov model, GMHMM) 来描述风速随机分布的特性^[14]。高斯混合分布是1种可以对复杂密度函数进行拟合的多个高斯分布的线性叠加。通过对多个高斯分布进行叠加,并调节其均值、协方差矩阵和系数,进而实现对任意连续概率密度函数的准确近似。观测概率矩阵 B 的参数由高斯混合分布重建,其重新估计的公式如下:

$$b_j(O_t) = \sum_{i=1}^I c_{ji} b_{ji}(O_t) = \sum_{i=1}^I c_{ji} N(O_t, \mu_{ji}, \Sigma_{ji}) \quad (5)$$

预测风速时采用模式识别:求出GMHMM中当前时刻对应的对数似然值,然后遍历历史数据的似然值,找到与当前时刻数据似然值最接近的历史状态,进而对风速进行预测。对风速进行准确预测是保证风电机组安全稳定、高效发电的关键举措^[15]。为提高风速预测的准确度,本文采用基于多个历史状态的加权平均预测法^[16]。其公式如下:

$$F_{0+1} = F_0 + \sum_{i=1}^d \omega_i (F_{t+1} - F_t) \quad (6)$$

式中: F_0 表示预测时刻上一时刻的风速; F_{0+1} 表示预测时刻的风速; $F_{t+1}-F_t$ 表示历史风速差值; ω_i 表示权重,权重表达式为:

$$\omega_i = \frac{1/|L_i - L_0|}{\sum_{j=1}^d 1/|L_j - L_0|}, \sum_{i=1}^d \omega_i = 1 \quad (7)$$

2.2 基于风速预测和 LSCGAN 的齿轮箱故障预警框架

1) 基于风速预测的改进 LSCGAN 模型

为了满足生成不同工况下多维时序数据的需求、在最小二乘 GAN (least square GAN, LSGAN) 的基础上添加风速作为条件信息,其使用最小二乘损失代替交叉熵损失,解决了传统 GAN 生成质量不高以及训练过程不稳定这两个缺陷。

本文在 LSCGAN 的基础上,移除了生成器输入中的随机噪声,仅将风速作为条件信息输入生成器,这样可有效降低模型训练和应用过程中,噪声对生成数据精度的影响。从 SCADA 数据库获取风电机组健康状态的历史数据,进行数据清洗和特征选择。将风速数据作为条件信息输入生成器,获得生成数据,并将其与真实数据分别输入判别器,得到判别结果。通过交替训练生成器和判别器,直至生成数据和真实数据的判别结果均收敛至 0.5,完成模型训练。流程图如图 1 所示。

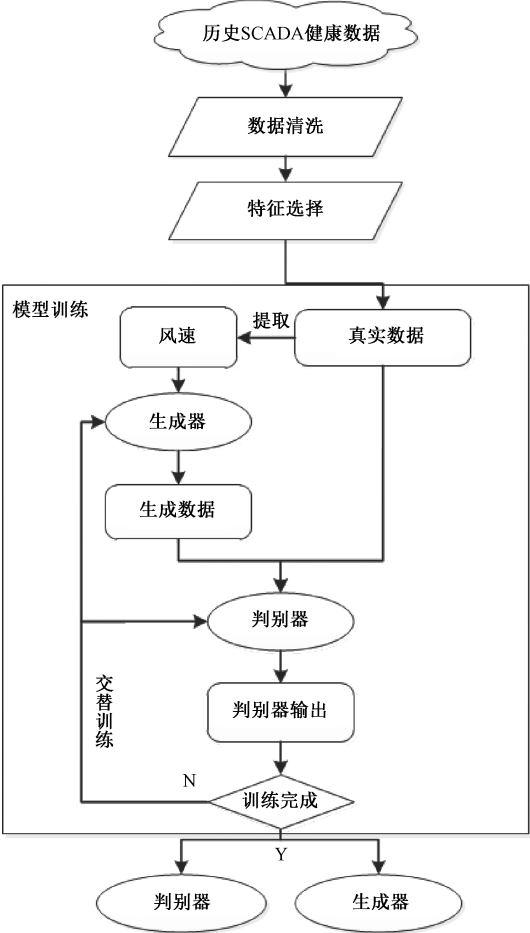


图 1 LSCGAN 训练流程

Fig. 1 LSCGAN training flow chart

2) 两阶段多维数据生成策略

首先,对 SCADA 系统实时数据进行预处理,剔除异常数据,并进行关键特征选择,构建出适用于模型的综合

特征向量^[17]。第 1 阶段:将实时风速数据整合入 GMHMM 的历史数据库中,以优化模型。GMHMM 通过识别与预测时刻最匹配的历史状态,采用加权平均法,实现未来短期风速的准确预测。第 2 阶段:将得到的预测风速输入到训练后的生成器模型中,输出包含齿轮箱油温在内的多维时序数据预测结果。整体流程如图 2 所示。

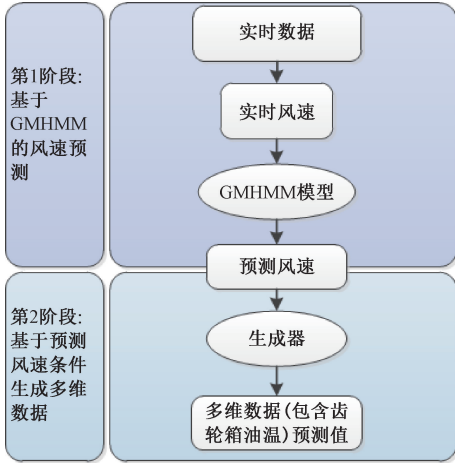


图 2 两阶段多维数据生成框架

Fig. 2 Two-stage multidimensional data generation framework diagram

3) 基于健康指数的故障预警策略

本文提出 1 种状态健康指数,它综合了生成器生成的齿轮箱油温预测值和判别器的判别结果,可以直观地表明齿轮箱的状态,整体预警策略图如图 3 所示。

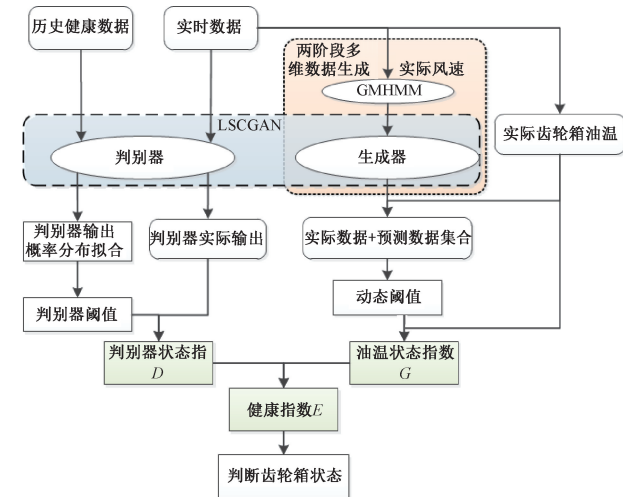


图 3 健康指数预警策略

Fig. 3 Health index early warning strategy diagram

基于两阶段多维数据生成法,设置齿轮箱油温的动态阈值。对未来一段时间内的风速进行预测,将预测风速作

为生成器的输入,得到健康状态油温预测值,通过设置合适的滑窗^[18],根据实际数据-预测数据组成的数据集,基于 3σ 准则确定当前时刻齿轮箱油温的上、下限阈值。

基于健康状态下的判别器结果,设置判别器实时输出的静态阈值。将健康数据输入训练好的判别器模型中,可得判别器输出散点图,通过绘制判别器输出直方图,并对输出直方图进行拟合,基于置信区间原理确定判别器输出的静态阈值。

综合以上 2 种预警策略,本文设计 1 种健康状态指数。首先,定义齿轮箱油温的状态指数 T ,可以根据齿轮箱油温实际值和所求的上下限阈值来计算。其公式如下:

$$T = \begin{cases} 1 - \frac{\left| T_{\text{actual}} - \frac{T_{\text{upper}} + T_{\text{lower}}}{2} \right|}{T_{\text{upper}} - T_{\text{lower}}}, & T_{\text{lower}} \leq T_{\text{actual}} \leq T_{\text{upper}} \\ 0, & T_{\text{actual}} < T_{\text{lower}} \text{ or } T_{\text{actual}} > T_{\text{upper}} \end{cases} \quad (8)$$

式中: T_{actual} 为齿轮箱油温实际值; T_{upper} 为动态上限阈值; T_{lower} 为动态下限阈值。

其次,定义判别器输出的状态指数 D ,可以根据判别器的输出值和阈值来计算。其公式如下:

$$D = \begin{cases} 1 - \frac{1 - D_{\text{output}}}{10(1 - D_{\text{lower}})}, & D_{\text{output}} \geq D_{\text{lower}} \\ 0, & D_{\text{output}} < D_{\text{lower}} \end{cases} \quad (9)$$

式中: D_{output} 为判别器输出; D_{lower} 为判别器下限阈值。

最后,定义综合指数 E ,用来表明齿轮箱的状态:

$$E = \lambda D + (1 - \lambda) T \quad (10)$$

式中: λ 为权重系数。

当健康状态指数 E 超过阈值时,判断风电机组齿轮箱出现故障,否则说明监测的状态参数虽有波动但各参数之间特征分布没有改变,并非处于异常状态,继续监测。

3 实例分析

3.1 基于 GMHMM 的风速预测

1) 数据选择

选取某风电机组 2017 年 8 月 11 日至 9 月 11 日的分钟级风速数据作为模型的数据集,其中 80% 数据作为训练集,剩余 20% 数据作为测试集。

2) 隐状态数目的确定

在构建 GMHMM 模型的过程中,随着训练样本的增加以及模型的参数的增加,模型的准确性也在不断地提高。但是,随着模型参数的增加,模型的复杂性也会越来越高,这就造成了过拟合的问题。因此,在模型复杂性和

准确性方面寻求一个理想的平衡点是非常必要的。贝叶斯信息准则(Bayesian information criterion,BIC)是1种模型选择准则,用于比较不同模型的拟合优度。它结合了模型的最大似然估计和模型的复杂度,以帮助确定最合适的模型。其表达式如下:

$$BIC = k\ln(n) - 2\ln(L)$$

(11)

在不同的 k 值下,对 BIC 进行了计算,较低的值说明在此参数下的模型与已有的数据集非常适合,BIC 在不同 k 值下的值如表 1 所示。

表 1 BIC 在不同隐状态数 k 下的数值

Table 1 BIC values under different number of hidden states		
隐状态数 k	似然值 LL	BIC
2	-33 005	66 027
3	-31 581	63 187
4	-31 195	62 424
5	-31 152	62 346
6	-30 916	61 882
7	-30 937	61 933

根据 BIC 计算公式,BIC 值存在极小值点,而后随着 k 的增大,BIC 值不断增大,由表 1 可知当设定模型的隐状态数目为 6 时,BIC 的数值最小,模型最优。

3) 风速预测

基于 GMHMM 进行风速预测主要依靠其“模式识别”的特性。本文基于分钟级历史风速求取某一时刻及其未来一段时间共 100 min 的风速预测值,通过绘制风速预测曲线和计算相应的模型评价指标:均方根误差(root mean square error, RMSE)、平均绝对误差(mean absolute error, MAE)和平均绝对百分比误差(mean absolute percentage error, MAPE),来获取合适的参数,并获得良好的风速预测效果^[19]。

对基于多个历史状态加权预测法,通过实例分析不同个数的历史状态下的预测精度,进而决定出最优的历史状态数。本文选取历史状态数 N 分别为 1、5、10、15、20、25、30 的加权法预测未来 100 min 的风速变化,其中 $N=25$ 时的预测风速曲线图如图 4 所示。

使用 MAE、MAPE 和 RMSE 对预测值的准确度进行分析,结果如表 2 所示。

随着历史状态数 N 的增加,风速预测效果越好,但 MAE、MAPE 和 RMSE 之间也越来越接近。基于 20、25、30 个历史状态数的各指标差值较小,这 3 种不同状态个数方法的预测准确度几乎没有差别。综合考虑了模型的复杂性和准确性,基于 25 个历史状态数的加权法能更好地解决模型预测问题。综合图 4 和表 2 可以看出:基于

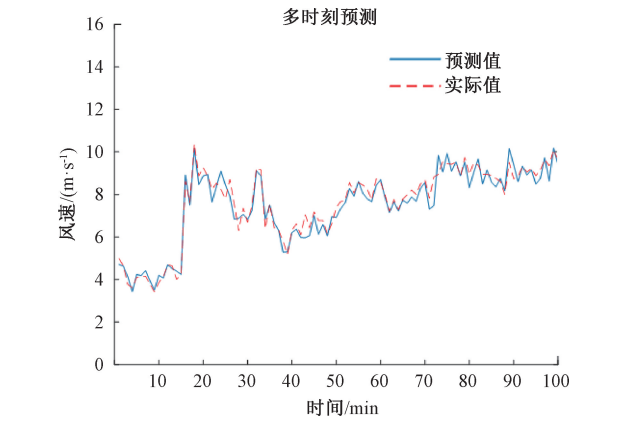


图 4 $N=25$ 时加权预测风速曲线
Fig. 4 Weighted predicted wind speed curve when $N=25$

表 2 不同历史状态数下的风速预测的 MAE、MAPE 和 RMSE

Table 2 MAE、MAPE and RMSE of wind speed prediction under different historical states			
历史状态数 N	MAE	MAPE	RMSE
1	1.410 5	0.194 9	1.779 8
5	0.929 5	0.128 4	1.256 2
10	0.822 1	0.112 8	1.085 6
15	0.721 8	0.099 5	0.986 3
20	0.687 4	0.093 8	0.933 0
25	0.620 9	0.082 9	0.869 1
30	0.655 1	0.087 4	0.934 0

多个历史状态数的加权预测在同样的情况下表现较好,但超过一定时间段后,预测值与实际值的误差逐渐增大,因此高斯混合隐马尔科夫模型预测方法仅适用于短期的风速预测研究。

3.2 基于改进生成对抗网络的故障预警

1) 数据预处理与特征选择

在生产过程中,由于传感器精度、外界干扰等因素,存在异常数据,需做数据预处理方可建立准确的预警模型^[20]。本文选取某风电场的某个风电机组 2017 年 8 月 11 日至 9 月 11 日的分钟级 SCADA 数据作为模型数据。首先去除数据样本中由于传感器故障、数据传输错误等产生的缺失值^[21];其次,基于风速-功率之间的关系,剔除掉在风速不等于 0 且机组没有出力的时刻或时段内的单位机组数据,得到风电机组在正常状态运行时段的数据^[22],如图 5 所示。数据清洗后,留下 23 646 个数据作为训练模型的数据集。

对清洗后的 SCADA 数据进行归一化处理。本文采

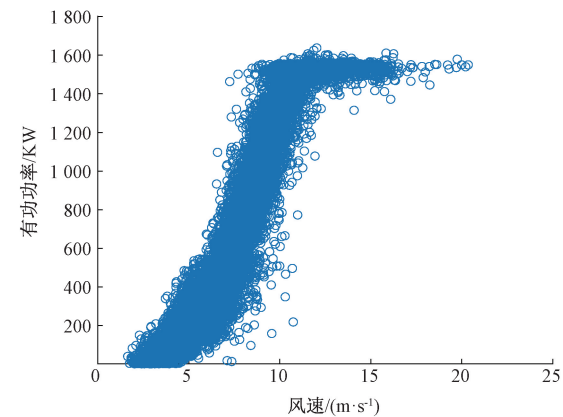


图 5 数据预处理后风速-功率

Fig. 5 Wind speed-power diagram after data preprocessing

用具有快速收敛特性、良好的解释性和对数据分布无特别要求的斯皮尔曼相关性分析方法,来提取能直接或者间接体现风机齿轮箱油温特性的特征参数集^[23],最终选取 L1 侧相电流、发电机有功功率、发电机定子 U 相线圈温度、风速、主轴转速、齿轮箱高速轴驱动端轴承温度、齿轮箱高速轴非驱动端轴承温度以及齿轮箱油温 8 个变量作为判别器的输入特征参数,如表 3 所示。

表 3 所确定的特征参数的 Spearman 相关性系数

Table 3 The Spearman correlation coefficient of the determined characteristic parameters

特征参数	Spearman
L1 侧相电流	0.760 0
发电机有功功率	0.744 1
发电机定子 U 相线圈温度	0.803 2
风速	0.728 9
主轴转速	0.718 3
齿轮箱高速轴驱动端轴承温度	0.943 0
齿轮箱高速轴非驱动端轴承温度	0.957 7
齿轮箱油温	1

2) 基于 LSCGAN 的模型训练

经过数据预处理和特征选择的 SCADA 数据作为训练集。提取风速作为生成器输入,生成其他 7 维数据,然后将生成的 7 维数据再与条件信息合并,与真实数据一起送入判别器进行判别,交叉循环进行训练。训练时利用 Adam 优化计算模型的参数,生成器和判别器的 Adam 优化器损失函数定义为最小二乘损失函数,优化器的初始学习率为 0.000 1。训练过程中通过设置软标签和 Two Timescale Update Rule 的方法来提高训练过程的稳定性。

当生成数据与真实数据的判别结果均收敛至 0.5 附近时,训练完成。从上面描述的 GAN 算法原理可以知道,风速数据作为条件信息输入生成器可生成齿轮箱油温预测曲线,如图 6 所示,图中虚线为实际值,实线为预测值,横线部分表示该时间段内存在数据缺失;将健康数据输入判别器模型,判别器输出如图 7 所示,其理论输出在 0.5 附近波动。

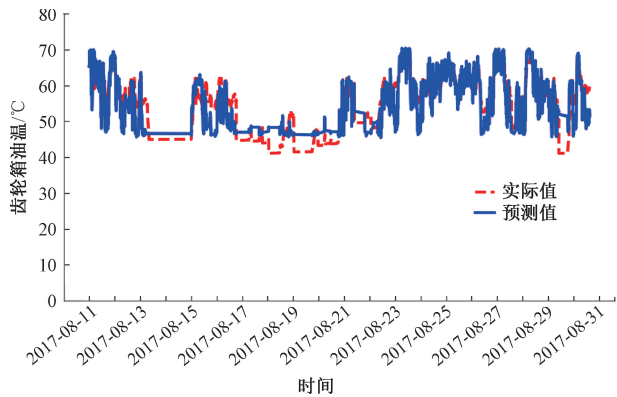


图 6 齿轮箱油温模型预测值和实际值曲线

Fig. 6 Gearbox oil temperature model predicted value and actual value curve

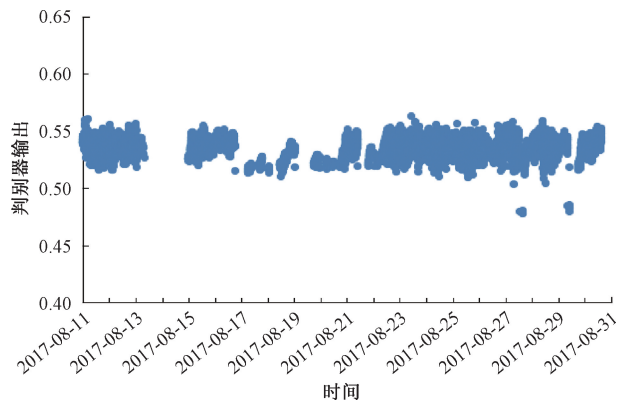


图 7 GAN 判别器对健康运行状态数据的判别结果散点

Fig. 7 The scatter plot of the discriminant results of the GAN discriminator for healthy running state data

绘制健康数据的判别器输出分布直方图,并利用统计学中的常用分布对所绘直方图进行拟合,如下图 8(a)~(c)所示。经对比发现:韦伯分布(Weibull distribution)能够很好地拟合判别器判别结果的分布。此外,已有研究证明韦伯分布可以更好地刻画实际风速的分布,并且由于风机运行状态与风速之间存在强相关性,本文利用韦伯分布来拟合判别器输出的概率分布,并采用最大似然方法进行参数估计,从而获得最佳的韦伯分布参数: $k=84.35$, $\lambda=0.539$ 。由韦伯分布的累积概率密度分布

函数求出相应的阈值 T , 设定允许的异常值比率 x , 使区间 $[0, T]$ 中所含数据的比率是 x , 这里取 $x = 0.3\%$, 得到 $T = 0.497$, 其具体数据分布如图 8(c) 所示。

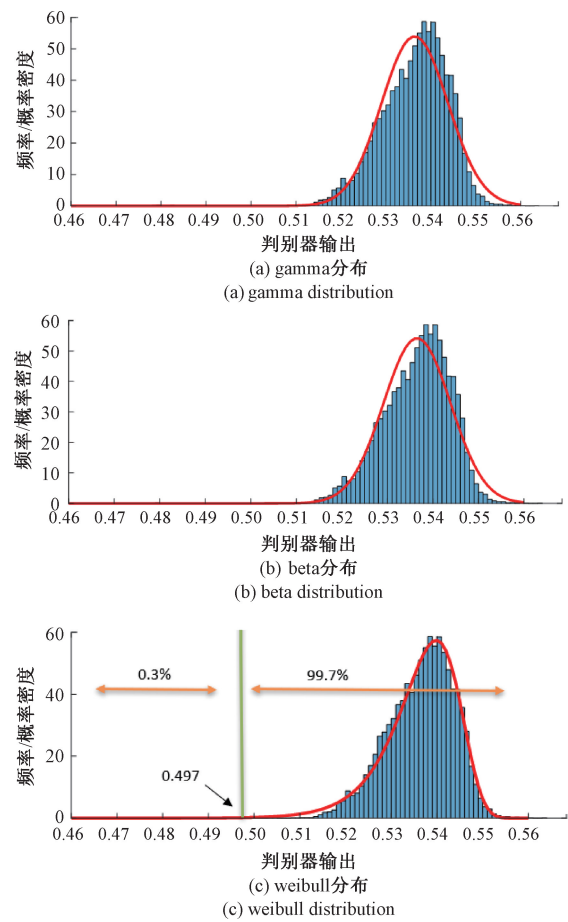


图 8 判别器输出直方图及不同分布的拟合曲线
Fig. 8 Discriminator output histograms and fitting curves of different distributions

3) 健康状态实例分析

基于训练完成的 GMHMM 模型和 LSCGAN 模型, 对风电机组 2017 年 11 月 11 日之后一个月内的齿轮箱健康运行数据进行分析。设置滑动窗宽度为 100, 依据两阶段法生成齿轮箱油温预测值, 形成实际油温-预测油温数据集, 基于 3σ 准则确定当前时刻的阈值上限和下限。如图 9 所示, 实线为实际值, 虚线为动态的阈值上限, 点划线为动态的阈值下限。将数据输入判别器, 可得实际数据的判别结果如图 10 所示。

图 9 和 10 表明: 风电机组在健康运行状态下, 齿轮箱油温实际值虽然出现几次极短时间的超阈值状况, 但对应的判别器输出并未超阈值, 这表明即使齿轮箱油温有所波动, 但对应多维数据的特征分布并未变化, 从图 11 中也可看出健康指数 E 并未超出阈值, 因此判定齿

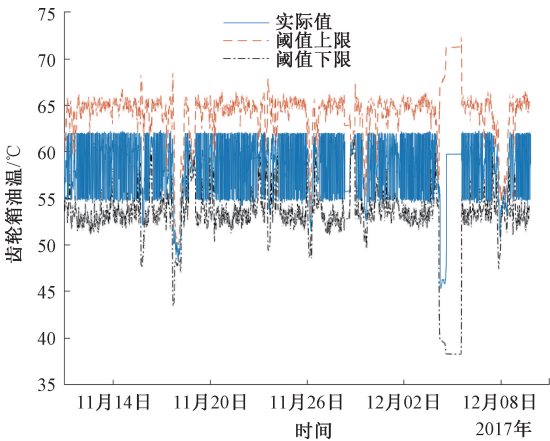


图 9 齿轮箱油温实际值和动态阈值曲线
Fig. 9 Gearbox oil temperature actual value and dynamic threshold curve

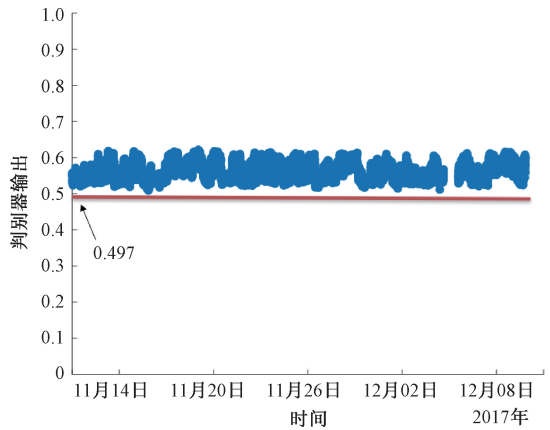


图 10 齿轮箱状态判别结果散点
Fig. 10 Gearbox state discriminant results scatter plot

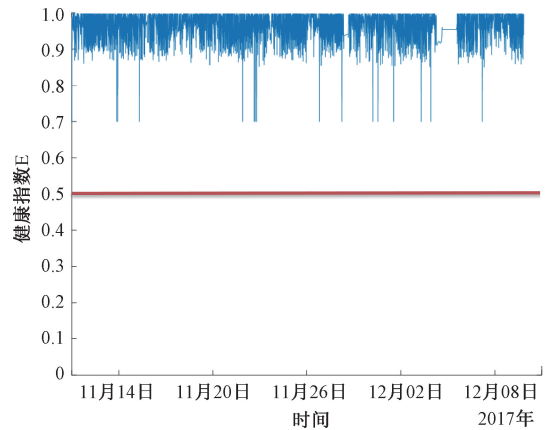


图 11 齿轮箱状态健康指数
Fig. 11 Gearbox state health index diagram

轮箱油温是受外界噪声干扰产生的波动, 而非自身出现了故障, 齿轮箱处于健康运行状态。

4) 故障状态实例分析

使用已训练好的多维生成器模型和状态判别模型,对 2018 年 1 月 18 日之后的齿轮箱故障数据进行监测。如图 12 所示实线为齿轮箱油温的真实值,虚线为动态的阈值上限,点划线为动态的阈值下限。将真实数据输入训练完成的判别器可得真实数据的判别结果如图 13 所示。虽然齿轮箱油温在 1 月 22 日前出现频繁的短时间超阈值情况,但对应时刻下判别器的输出结果并未超阈值,判定为外部扰动导致的油温波动,而非风电机组齿轮箱处于异常状态,齿轮箱正常工作。

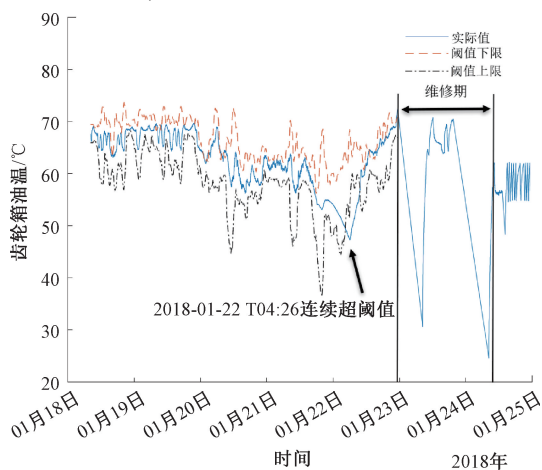


图 12 齿轮箱油温实际值和动态阈值曲线

Fig. 12 Actual gearbox oil temperature and dynamic threshold curve

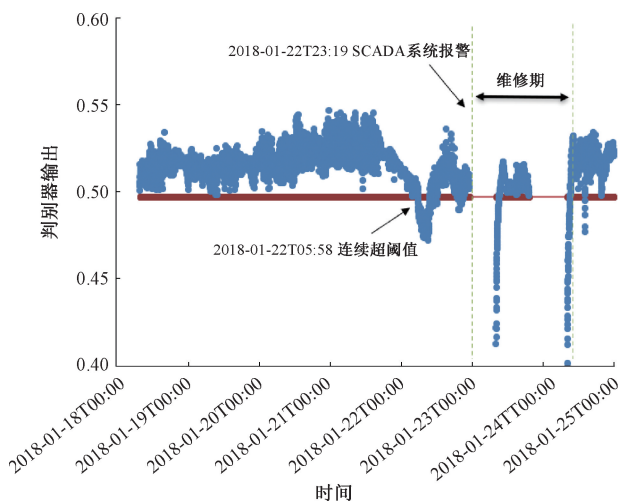


图 13 齿轮箱状态判别结果散点

Fig. 13 Gearbox state discriminant results scatter plot

为了直观地表明齿轮箱的状态,本文计算所得的健康指数曲线图如图 14 所示。从图 14 中可以直观地看出,在“2018-01-22”之前齿轮箱的健康指数 E 虽然有些波动,但并未超过阈值;仅在“2018-01-22T05:58”时齿轮箱油温实际值和判别器的输出结果同时超出阈值,齿轮

箱的健康指数 E 超出阈值,此时判定齿轮箱处于异常运行状态,而 SCADA 系统在“2018-01-22T23:19”时才发出报警。由此可知,本文所提方法对齿轮箱有着良好的预警效果,且相较于 SCADA 系统提早约 17 h 检测出了风电机组齿轮箱异常的状况。

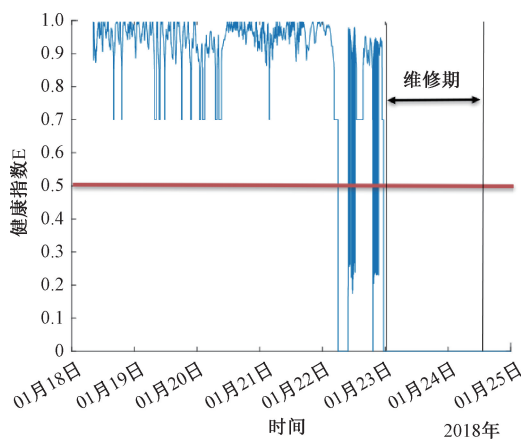


图 14 齿轮箱状态健康指数

Fig. 14 Gearbox state health index diagram

3.3 生成多维数据质量分析

为了验证两阶段生成的多维数据的质量,本文采用 XGBoost 模型对生成的多维数据进行验证分析,XGBoost 是一种高效的集成学习算法,能够捕捉到数据中的复杂非线性关系,学习多维时序数据的特征分布,这对于预测齿轮箱油温这类具有复杂特征和动态变化的数据非常有用。对同一风电机组健康运行月份的 SCADA 数据进行数据预处理。依据两阶段法,基于处理后的风速数据进行多维数据生成,生成数据与实际数据的各个特征向量对比图如图 15 所示。

使用真实的多维数据训练 XGBoost 模型。确保模型充分训练,使用训练好的 XGBoost 模型对生成的多维数据进行预测,得到预测油温。将预测油温与真实油温进行对比,并绘制折线图,直观地比较预测油温与真实油温的趋势和分布,如图 16 所示,图中实线为实际值,虚线为预测值。

从图 15 和 16 可以看出:两阶段法生成的多维数据的各特征向量与实际数据相比,各点数据误差较小,变化趋势相同;由生成的多维数据产生的预测油温曲线与真实曲线拟合度较高,且与真实数据的变化趋势一致,由此可以得出采用两阶段法生成的多维数据的质量较好,可以用于后续的使用和分析。

3.4 故障预警时间对比

风电机组故障预警的越早,及时发现潜在问题,可以更早采取措施避免或延缓故障发生,从而减少设备损坏,延长风电机组的使用寿命。分别利用卷积神经网络

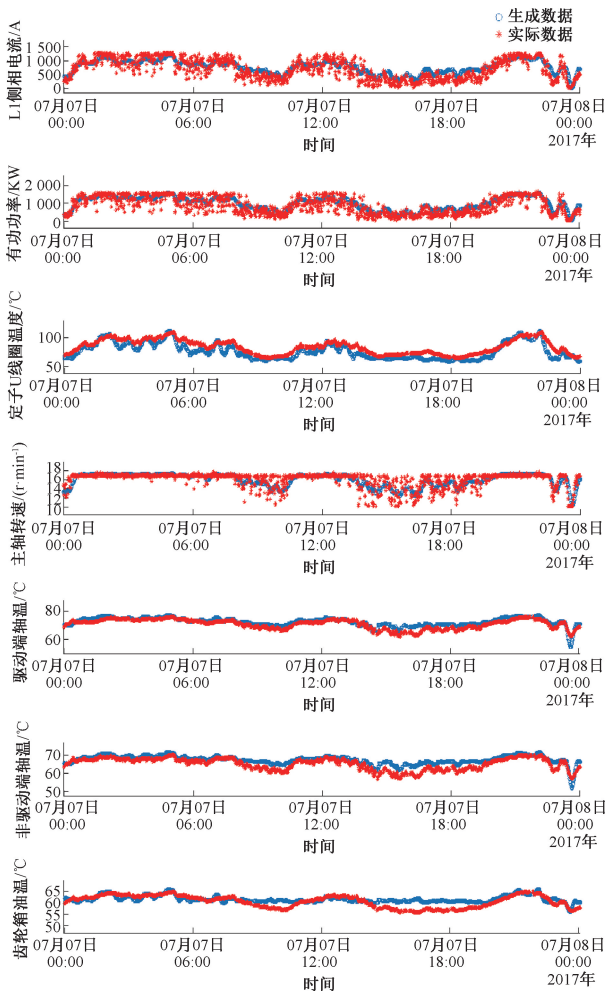


图 15 两阶段法生成数据与真实数据对比曲线

Fig. 15 The two-stage method generates a comparison curve between the generated data and the actual data

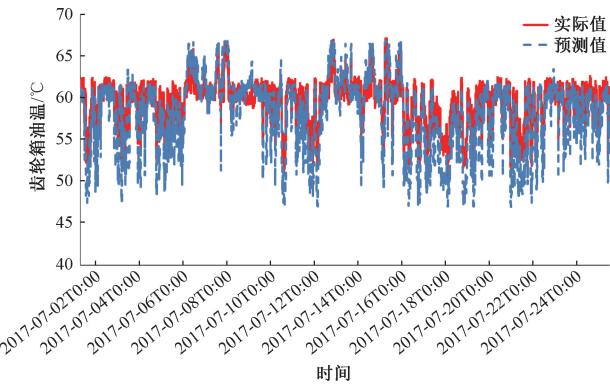


图 16 XGBoost 模型预测值与真实值对比曲线

Fig. 16 The comparison curve between the predicted value and the real value of the XGBoost model

term memory, LSTM)^[25] 和门控循环单元 (gate recurrent unit, GRU)^[26] 这几种常用的预测时序数据的方法构建油温预测模型,对同一风电机组同一故障案例进行故障预警,通过预警时间进行对比,如表 4 所示。从表 4 可以看出:本文所提方法相较于原来使用的 SCADA 系统报警,提前了约 17 h 发出齿轮箱油温异常预警,比常见的深度学习模型提前了约 1~3 h。

表 4 不同方法故障预警时间对比表

Table 4 Comparison of fault warning times across different methods

预警方法	预警时间	提前预警小时数/h
SCADA 系统	2018-01-22T23:19	—
CNN	2018-01-22T07:12	16.12
DNN	2018-01-22T07:54	15.42
LSTM	2018-01-22T09:04	14.25
GRU	2018-01-22T07:58	15.35
本文	2018-01-22T05:58	17.35

3.5 虚警率分析

在风电场景下,目前普遍存在健康数据和故障数据的样本量严重不平衡的问题。虚警率是一个重要的性能指标,通常被定义为虚警样本数与总正常样本数的比值。在实际的风电机组运行过程中,每次虚警都会给风电企业带来很大的经济损失。因此,降低虚警率可以减少系统不必要的负荷,有效提高风电机组运维的稳定性和可靠性。同样分别利用 CNN、DNN、LSTM、GRU 这几种常用的预测时序数据的方法构建模型,对同一风电机组某一健康运行月份进行故障预警,通过评价指标进行对比,如表 5 所示。

表 5 不同方法健康月份预测结果对比

Table 5 Comparison table of different methods of health month prediction result

预警方法	健康月份超阈值样本数	虚警率/%
本文	1	0.005
CNN	166	0.83
DNN	184	0.92
LSTM	241	1.205
GRU	194	0.97

本文所提方法在鲁棒性上有较好的性能,其在健康运行月份中超阈值的样本数目只有 1 个,比其他几种方法都要小得多,即虚警率最小,能有效减少健康运行状态下对数据样本的误判。

(convolutional neural network, CNN)、深度神经网络 (deep neural network, DNN)^[24]、长短时记忆网络 (long short-

4 结 论

现有的齿轮箱故障预警方法主要使用残差类技术,通过实际数据与预测数据的残差来判断是否存在故障,但其容易受到噪声干扰,导致在风电机组正常运行时产生误报警。本文提出了 1 种两阶段多维数据生成与实时健康指数的风电机组齿轮箱故障预警方法,其特点在于:

1) 提出基于 GMHMM 的加权平均风速预测法,通过 MAE、MAPE、RMSE 这 3 项指标的对比,基于 25 个历史状态的加权平均预测法的预测精度较好;

2) 相较于 LSGAN 模型,本文添加风速作为条件信息,并移除了随机噪声,通过交叉训练,可以生成不同工况下的高质量多维时序数据。

3) 提出了 1 种实时健康指数,综合了基于两阶段数据生成的动态阈值及判别器的输出,可有效降低实际应用过程中的虚警率。

本文使用现场数据,对基于两阶段法生成的多维数据的质量进行了验证分析,结果表明生成的多维数据的特征分布与真实分布一致,数据质量较好。同时,实际故障案例仿真分析结果表明:本文提出的方法能提前约 17 h 检测出齿轮箱的异常状态,且相较于仅分析齿轮箱油温的残差进行故障预警的方法,本文方法可以有效降低故障预警的虚警率。

参考文献

- [1] 吕游,封烁,郑茜,等. 基于 LeNet5like 的迁移学习风电机组叶片覆冰故障诊断研究[J]. 仪器仪表学报, 2024,45(3):128-143.
LYU Y, FENG SH, ZHENG X, et al. Research on fault diagnosis of wind turbine icing characteristics based on LeNet5like transfer learning[J]. Chinese Journal of Scientific Instrument, 2024,45(3):128-143.
- [2] 刘杰,曹静,赵昕. 基于 OOB-GWO-SVR 的风电机组齿轮箱故障预警[J]. 电子测量与仪器学报, 2022, 36(12):97-105.
LIU J, CAO J, ZHAO X. Wind turbine gearbox fault warning based on OOB-GWO-SVR[J]. Journal of Electronic Measurement and Instrumentation, 2022, 36(12):97-105.
- [3] 向玲,朱浩伟,丁显,等. 基于 CAE 与 BiLSTM 结合的风电机组齿轮箱故障预警方法研究[J]. 动力工程学报, 2022,42(6):514-521.
XIANG L, ZHU H W, DING X, et al. Research on fault early warning method of wind turbine gearbox based on CAE and BiLSTM[J]. Journal of Power Engineering, 2022,42(6):514-521.
- [4] CHWN H G, CHEN J, DAI J CH, et al. Early fault

warning method of wind turbine main transmission system based on SCADA and CMS data[J]. Machines, 2022, 10(11):1018-1018.

- [5] 付景,林永君,刘帅,等. 基于 OOB-DBO-GRU 的风电机组齿轮箱故障预警[J/OL]. 中国测试, 1-7[2024-08-01].
FU J, LIN Y J, LIU SH, et al. Fault warning of wind turbine gearbox based on OOB-DBO-GRU[J/OL]. China Test, 1-7[2024-08-01].
- [6] 马良玉,耿妍竹,梁书源,等. 基于 Stacking 多模型融合的风电机组齿轮箱油池温度异常预警[J]. 中国电机工程学报, 2023,43(S1):242-251.
MA L Y, GENG Y ZH, LIANG SH Y, et al. Temperature anomaly warning of wind turbine gearbox oil pool based on Stacking multi-model fusion[J]. Chinese Journal of Electrical Engineering, 2023,43(S1):242-251.
- [7] 金晓航,许壮伟,孙毅,等. 基于生成对抗网络的风电机组在线状态监测[J]. 仪器仪表学报, 2020,41(4):68-76.
JIN X H, XU ZH W, SUN Y, et al. Online condition monitoring of wind turbines based on generative adversarial networks[J]. Chinese Journal of Scientific Instrument, 2020,41(4):68-76.
- [8] WANG AN Q, QIAN ZH, PEI Y, et al. A de-ambiguous condition monitoring scheme for wind turbines using least squares generative adversarial networks[J]. Renewable Energy, 2022, 185:267-279.
- [9] CHEN P, LI Y, WANG K SH, et al. A threshold self-setting condition monitoring scheme for wind turbine generator bearings based on deep convolutional generative adversarial networks[J]. Measurement, 2021,167:108234.
- [10] LI D, CHEN D CH, JIN B H, et al. MAD-GAN: Multivariate anomaly detection for time series data with generative adversarial networks[J]. Spring, Charm, 2019:11730.
- [11] 江俊. 基于 HMM 修正视角的期货价格预测[D]. 浙江:浙江财经大学, 2023.
JIANG J. Futures price forecasting based on HMM correction perspective[D]. Zhejiang: Zhejiang University of Finance and Economics, 2023.
- [12] 王晨. 基于隐马尔科夫模型的股票价格指数预测[D]. 山东:山东大学, 2018.
WANG CH. Prediction of stock price index based on hidden Markov model[D]. Shandong: Shandong University, 2018.
- [13] 王越,孙若莹. 基于高斯混合隐马尔科夫模型的螺纹钢价格波动预测[J]. 冶金经济与管理, 2022(2):36-40.
WANG Y, SUN R Y. Prediction of price fluctuation of rebar based on Gaussian mixture hidden Markov

- model[J]. Metallurgical Economy and Management, 2022(2):36-40.
- [14] 黄宇,张冰哲,庞慧珍,等. 基于混合 Copula 优化算法的风速预测方法研究[J]. 太阳能学报, 2022, 43(10):192-201.
HUANG Y, ZHANG B ZH, PANG H ZH, et al. Research on wind speed prediction method based on mixed Copula optimization algorithm[J]. Journal of Solar Energy, 2022,43(10):192-201.
- [15] 王逸文,王维莉,刘贤超,等. 融合两阶段分解与 iJaya-ELM 的短期风速预测模型[J]. 电子测量与仪器学报, 2023,37(7):186-195.
WANG Y W, WANG W L, LIU X CH, et al. A short-term wind speed prediction model combining two-stage decomposition and iJaya-ELM[J]. Journal of Electronic Measurement and Instrumentation, 2023,37(7):186-195.
- [16] 尤晴. 基于 PSO 优化 LSSVM 的短期风速预测研究[D]. 哈尔滨:哈尔滨工业大学, 2022.
YOU Q. Research on short-term wind speed prediction based on PSO optimized LSSVM[D]. Harbin: Harbin Institute of Technology, 2022.
- [17] DAO P B. Condition monitoring and fault diagnosis of wind turbines based on structural break detection in SCADA data[J]. Renewable Energy, 2022, 185:641-654.
- [18] LIU Y R, WU Z D, WANG X L. Research on fault diagnosis of wind turbine based on SCADA data[J]. IEEE Access, 2020, 8:185557-185569.
- [19] 王旭光,张可,李潇,等. 基于变量相关注意力机制的短期风速预测[J]. 太阳能学报, 2023, 44(8):467-476.
WANG X G, ZHANG K, LI X, et al. Short-term wind speed prediction based on variable-related attention mechanism[J]. Journal of Solar Energy, 2023,44(8):467-476.
- [20] 徐健,刘长良,王梓齐,等. 基于缺失数据填补的风电齿轮箱状态监测研究[J]. 仪器仪表学报, 2022, 43(9):88-97.
XU J, LIU CH L, WANG Z Q, et al. Research on condition monitoring of wind turbine gearbox based on missing data filling[J]. Chinese Journal of Scientific Instrument, 2022,43(9):88-97.
- [21] 金晓航,许壮伟,孙毅,等. 基于 SCADA 数据分析和稀疏自编码神经网络的风电机组在线运行状态监测[J]. 太阳能学报, 2021, 42(6):321-328.
JIN X H, XU ZH W, SUN Y, et al. Online operation condition monitoring of wind turbines based on SCADA data analysis and sparse autoencoder neural network[J]. Journal of Solar Energy, 2021,42(6):321-328.
- [22] ZHANG ZH Y, WANG SH K, WANG P G, et al. Research on fault early warning of wind turbine based on IPSO-DBN[J]. Energies, 2022, 15(23):9072.
- [23] 贾科,杨哲,魏超,等. 基于斯皮尔曼等级相关系数的新能源送出线路纵联保护[J]. 电力系统自动化, 2020,44(15):103-111.
JIA K, YANG ZH, WEI CH, et al. Pilot protection based on spearman rank correlation coefficient for transmission line connected to renewable energy source[J]. Automation of Electric Power Systems, 2020,44(15):103-111.
- [24] XIANG L, YANG X, HU A J, et al. Condition monitoring and anomaly detection of wind turbine based on cascaded and bidirectional deep learning networks[J]. Applied Energy, 2022,305:117925.
- [25] 向玲,王朋鹤,李京蓄. 基于 CNN-LSTM 的风电机组异常状态检测[J]. 振动与冲击, 2021,40(22):11-17.
XIANG L, WANG P H, LI J X. Abnormal state detection of wind turbines based on CNN-LSTM[J]. Journal of Vibration and Shock, 2021,40(22):11-17.
- [26] 马永光,冯勇升. 基于 HICEEMDAN-PCA-GRU 的风电机组齿轮箱故障预警方法研究[J]. 太阳能学报, 2023,44(4):67-73.
MA Y G, FENG Y SH. Research on fault early warning method of wind turbine gearbox based on HICEEMDAN-PCA-GRU[J]. Journal of Solar Energy, 2023,44(4):67-73.

作者简介



马同旭, 2022 年于华北电力大学获得学士学位, 现为华北电力大学硕士研究生, 主要研究方向为风电机组故障预警。

E-mail: matongxu9@163.com

Ma Tongxu received his B. Sc. degree from North China Electric Power University in 2022. Now he is a postgraduate student of North China Electric Power University. The main research direction is wind turbine fault warning.



刘帅(通信作者), 2013 年于华北电力大学获得学士学位, 2019 年于华北电力大学获得博士学位, 现为华北电力大学副教授, 主要研究方向为数据驱动建模、控制系统优化、风电机组故障预警。

E-mail: liushuai@ncepu.edu.cn

Liu Shuai (Corresponding author) obtained B. Sc. degree from North China Electric Power University in 2013 and Ph. D. degree from North China Electric Power University in 2019. Now he is an associate professor of North China Electric Power University. The main research directions are data-driven modeling, control system optimization, and wind turbine fault warning.