

# 基于平衡分布自适应迁移学习的多风电机组运行状态监测方法

张雅洁<sup>1,2</sup>, 王 罗<sup>3</sup>, 刘宇璐<sup>1,2</sup>, 乐 波<sup>3</sup>, 韩 爽<sup>1,2</sup>, 苏 营<sup>3</sup>, 刘永前<sup>1,2</sup>

(1. 新能源电力系统国家重点实验室(华北电力大学), 北京 102206; 2. 华北电力大学 新能源学院, 北京 102206; 3. 中国长江三峡集团有限公司, 北京 100038)

**摘要:** 风电机组状态的准确监测对风电机组安全稳定运行和经济效益提升至关重要。但是,受不同风电机组运行数据分布差异的影响,现有状态监测方法在多风电机组应用场景下存在精度和效率难以兼顾的问题,而平衡分布自适应迁移学习(BDA)可以拉近数据距离,同化数据分布。因此,文章提出了一种基于BDA的多风电机组状态监测方法。首先,基于Copula熵的互信息法挖掘风电机组运行状态关键影响参量;然后,构建基于门控循环单元模型(GRU)和序贯概率比检验(SPRT)方法的单风电机组状态监测模型;最后,构建基于BDA的多风电机组运行数据分布同化模型,并用于多风电机组运行状态监测。算例结果表明,所提方法可以有效节省建模成本和计算成本,能够在保障多风电机组运行状态监测精度的前提下,显著提升监测效率。

**关键词:** 风电机组; 状态监测; 平衡分布自适应迁移学习; 序贯概率比检验; 门控循环单元

**中图分类号:** TK81 **文献标志码:** A **文章编号:** 1671-5292(2024)08-1068-06

## 0 引言

由于风电机组运行环境恶劣、承受载荷大、风况多变、运行存在不确定性,导致风电机组运行情况复杂多变,设备故障频繁,运维问题不断<sup>[1]</sup>。因此,对风电机组状态进行准确监测有助于运维人员更好地判断风电机组的健康衰退趋势,提高风力发电机组的可靠性,避免故障造成的停机和事后维修损失,提高风电场运营收益。

风电机组的状态监测和故障诊断方法主要有物理模型和数据驱动两类。风电机组结构复杂,建立纯物理模型难度大。而数据驱动的方法使用数据采集与监控数据(SCADA)训练状态监测模型,不需要风电机组先验知识,是目前的研究热点。文献[2]基于高斯过程判断风电机组是否处于异常状态,并通过主成分分析进一步确定导致机组出现异常的关键部件,但在选择输入时依赖专家经验,没有考虑输入参量之间的相关性。文献[3]提出了一种主带平滑策略,该方法依据数据筛选后保留网格的数目和相对位置,对网格边缘符合风电机组出力特性的数据进行补充,使保留的数据主带更为平滑,提升了功率曲线的建模

精度。文献[4]结合支持向量机和果蝇优化算法回归风电机组参量,计算劣化度进行状态划分。但上述文献研究基于传统的机器学习,泛化能力差,而深度学习的特征挖掘能力强,如文献[5]构建了基于卷积神经网络和门控循环网络的风电机组多向时空状态监测模型,通过挖掘数据高维时空特征从而提高模型精度。文献[6]利用基于最大互信息系数挖掘风电机组参量,提高模型的精度。文献[7]结合注意力机制和自编码器提高模型的鲁棒性,监测风电机组的状态。

现有研究注重提升单台风电机组状态监测精度,但由于不同风电机组所处位置具有不同的风能资源条件,数据分布存在明显差异,须针对自身数据重复建模提高监测精度,但会显著增加人工建模和计算成本,难以兼顾多风电机组状态的监测精度和效率。

针对上述问题,本文提出了基于平衡分布自适应迁移学习(BDA)的多风电机组状态监测方法,同化不同风电机组的数据分布,平衡多风电机组状态监测的精度和效率。首先,基于Copula熵的互信息法挖掘风电机组功率的关键影响参量;然

收稿日期: 2022-10-27。

基金项目: 中国长江三峡集团有限公司企业科技项目(212103368)。

通信作者: 韩 爽(1974-),女,博士,教授,博士生导师,研究方向为风电场功率回归、风电场规划、设计及运行。

E-mail: hanshuang1008@sina.com

后,以关键影响参量为输入,以风电机组有功功率为输出,构建基于门控循环单元模型(GRU)和序贯概率比检验(SPRT)方法的单风电机组状态监测模型;最后,基于 BDA 同化多台风电机组的运行数据分布,结合 GRU 模型和 SPRT 方法构建多风电机组状态监测模型,并以国内某风电场真实数据作为算例,验证所提方法的有效性和正确性。

### 1 基本思路

本文所提基于 BDA 的多风电机组运行状态监测主要分为 3 个关键模块,基本思路如下:

①基于 Copula 熵互信息法的风电机组关键影响参量挖掘,挖掘风电机组有功功率的关键影响参量,作为后续 GRU 模型的输入;

②基于 GRU 模型和 SPRT 方法的单风电机组状态监测方法,基于 GRU 构建风电机组正常行为模型,并结合 SPRT 判断单机组状态,为多风电机组状态监测提供数据和模型支撑;

③基于 BDA 的多风电机组状态监测方法,同化源域和目标域数据分布,并用于多风电机组状态监测。

多风电机组运行状态监测流程如图 1 所示。



图 1 风电机组状态监测流程图

Fig.1 Condition monitoring flow chart of wind turbine

## 2 基于 Copula 熵互信息法的风电机组状态关键影响参量挖掘

### 2.1 基于 Copula 熵的互信息法

为消除冗余参量,提高后续正常行为模型的

精度和训练效率,采用基于 Copula 熵的互信息法进行特征提取,准确评估参量之间的相关度。

对参量  $X, Y$  进行归一化,获得参量  $U, V$ ,采用核密度估计法确定参量  $U, V$  的边缘概率分布  $u, v$ 。计算参量  $U, V$  的高斯-Copula 函数。根据 Copula 概率密度函数  $c(u, v)$  替换互信息的联合概率密度,计算参量互信息值  $I, I$  越大,相关度越高。

### 2.2 关键影响参量挖掘流程

挖掘风电机组功率关键影响参量,消除冗余参量,提高后续正常行为模型的效率。

采集风电机组 SCADA 运行数据,采用四分位法清洗风电机组 SCADA 运行数据<sup>[8]</sup>,计算 SCADA 参量与风电机组功率的互信息值,确定风电机组运行状态关键影响参量。

## 3 基于 GRU 模型和 SPRT 方法的单风电机组状态监测方法

### 3.1 GRU

长短期记忆网络(LSTM)内部结构复杂导致计算资源浪费,因此 GRU 在 LSTM 的基础上用重置门  $R_t$  和更新门  $Z_t$  改进了结构、精简了门控单元数量<sup>[9]</sup>,提高了训练效率,能有效降低过拟合的风险。GRU 具体结构如图 2 所示。

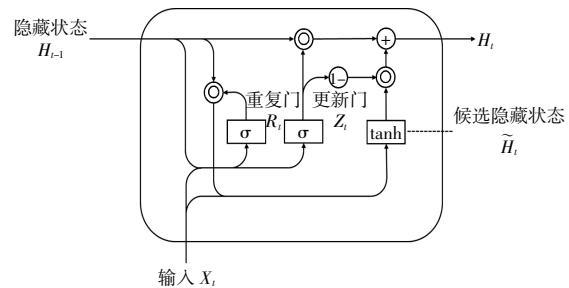


图 2 GRU 神经网络

Fig.2 GRU neural network

$R_t$  决定是否重置信息积累,局部是否发生显著变化,有助于捕捉时间序列里短期的依赖关系, $Z_t$  决定了是否对过去隐藏状态进行更新,能挖掘时间序列里长期的依赖关系。

### 3.2 SPRT

SPRT 是一种统计决策方法<sup>[10]</sup>。为快速准确判断风电机组功率残差的变化,本文采用了 SPRT 监测风电机组的运行状态。

已知随机参量序列  $X=[x_1, x_2, x_3, \dots, x_n]$ ,提出两点假设:假设  $H_0$ ,风电机组正常运行;假设  $H_1$ ,

风电机组亚健康运行。同时,构成序贯概率比检验。

误报警概率  $\alpha$  和漏报警概率  $\beta$  的对应阈值分别为  $A, B$ 。若似然比满足  $\lambda_n < A$ , 则停止检验, 接受假设  $H_0$ ; 若似然比满足  $\lambda_n > B$ , 则停止检验, 拒绝假设  $H_0$ , 接受假设  $H_1$ ; 如似然比满足  $A \leq \lambda_n \leq B$ , 则无法判定, 取值继续进行检验, 直至满足上述条件, 做出最后判断。

### 3.3 单风电机组状态监测模型构建

基于 GRU 模型和 SPRT 方法的单风电机组状态监测模型构建步骤如下:

①挖掘关键影响参量, 对风电机组 SCADA 数据进行预处理;

②构建基于 GRU 的风电机组正常行为模型, 获得回归功率;

③计算回归功率与实际功率的残差值, 基于 SPRT 比标记单点结果;

④设置连续  $t$  个点的似然比状态为  $w$ , 状态阈值为  $w_1$ , 当  $w > w_1$  时, 判断风电机组亚健康运行; 否则, 判断风电机组正常运行。计算正常运行状态数据的抽样状态分布, 可设  $w_1$  为 0.7。

$$w = \frac{1}{t} (x_1 + \dots + x_i + \dots + x_t) \quad (1)$$

式中:  $x_i$  为单点似然比判断结果, 其中正常为 0, 异常为 1 ( $i=1, 2, 3, \dots, t$ )。

正常运行状态数据的抽样分布如图 3 所示。

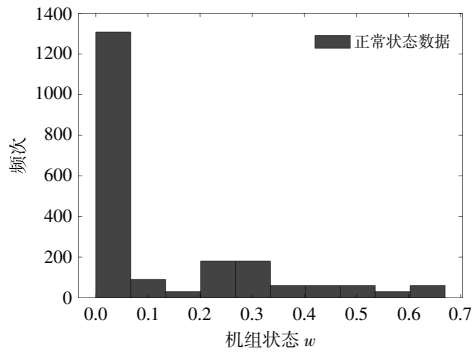


图 3 正常状态数据的抽样分布

Fig.3 Sampling distribution diagram of normal state data

## 4 基于 BDA 的多风电机组状态监测方法

### 4.1 BDA 算法

BDA 可根据实际数据分布的情况, 通过平衡因子  $\mu$  自适应地调整分布适配过程中边缘分布和条件分布的权重, 拉近二者的概率分布, 从而在不重复建模的情况下提高模型的精度<sup>[11]</sup>。

①应用 TCA 缩小数据边缘分布, 计算源域和目标域最大均值差异矩阵(MMD)和中心矩阵, 选择一个核函数映射优化 MMD, 拉近数据的概率分布。

②适配源域和目标域条件分布。用类条件概率  $P(x_i|y_i)$  近似  $P(y_i|x_i)$  来逼近条件概率, 并通过迭代提高精度。

③用  $\mu$  动态调整边缘分布和条件分布的权重, 其计算式为

$$DISTANCE(D_s, D_t) \approx (1-\mu)DIS[P(x_s), P(x_t)] + \mu DIS[P(y_s|x_s), P(y_t|x_t)] \quad (2)$$

式中:  $D_s, D_t$  分别为源域数据和目标域数据;  $P(x_s), P(x_t)$  为二者的边缘分布;  $(y_s|x_s), P(y_t|x_t)$  为二者的条件分布;  $\mu \in [0, 1]$  为平衡因子。

不同数据分布的源域和目标域如图 4 所示。图中, 圆形和正方形分别代表数据集中的两类数据。当目标域类型如图 4(b)所示时, 源域和目标域本身存在较大的差异性,  $\mu$  应接近于 0; 当目标域类型如图 4(c)所示时, 源域和目标域的边缘分布相似, 应接近于 1。

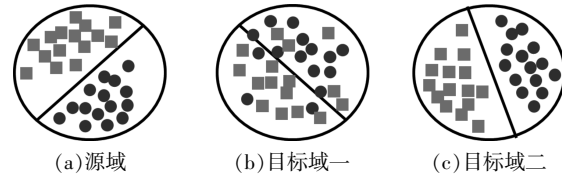


图 4 不同数据分布的源域和目标域

Fig.4 Source domain and target domain of different data distribution

### 4.2 多风电机组状态监测模型构建

同一风电场的不同风电机组因所处位置不同, 具有不同的风资源条件, 数据分布存在显著差异。若想提高风电机组状态监测模型的准确度, 须针对机组自身数据调整模型, 而人工建模成本高。开展多风电机组运行状态监测既可以提高监测模型的精度, 又节省了重复计算的资源, 解决了多风电机组精度和效率不可兼顾的问题。因此, 本文提出了基于 BDA 的多风电机组状态监测模型。

①基于 Copula 熵的互信息法挖掘关键影响参量。

②构建单台风电机组状态监测模型, 将关键印象参量数据和状态监测结果作为源域数据, 另一台风电机组数据作为目标域数据; 为消除不同机组间数据分布差异, 采用 BDA 同化数据分布。

③输入迁移后的 2 号风电机组数据, 构建基

于GRU 的风电机组正常行为模型。

④基于 SPRT 分析功率残差, 结合时间状态阈值监测多风电机组状态。

多风电机组状态监测流程如图 5 所示。

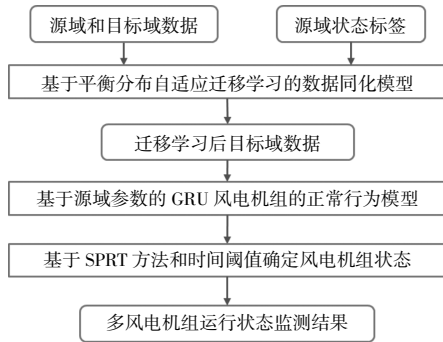


图 5 多风电机组状态监测流程  
Fig.5 Condition monitoring flow chart of multiple wind turbine

## 5 算例分析

### 5.1 数据基础

选用国内某风电场的 SCADA 数据及其故障申诉表进行实验验证, 数据的采样周期为 1 min, 数据的时间长度为 1 个月。

### 5.2 结果分析

采集 28 号风电机组 2018 年 4 月的 SCADA 数据作为样本集。SCADA 数据中参量种类繁多, 冗余参量降低了网络的精度和计算速度, 因此采用基于 Copula 熵的互信息法筛选回归模型的输入参量, 挖掘与风电机组功率高度相关的参量, 计算 SCADA 中所有参量与功率的互信息系数(表 1)。

表 1 不同参量与功率的互信息系数

Table 1 Mutual information coefficient of different variables and power

参量	$I$
主轴转速	0.491 375 931
发电机转速	0.468 855 146
风速	0.455 729 784
桨距角	0.399 833 351
齿轮箱驱动端轴承温度	0.349 464 318
齿轮箱非驱动端轴承温度	0.341 317 418
齿轮箱油池温度	0.271 456 461
发电机定子 U 相线圈温度	0.261 365 209
发电机定子 V 相线圈温度	0.262 293 845
发电机定子 W 相线圈温度	0.257 678 783
主轴承外圈温度	0.233 509 172
浆电机温度	0.232 989 881
齿轮箱分配器位置油压	0.231 990 587
齿轮箱油泵吸油口油压	0.211 139 204

互信息系数越大表示此参数和功率变化相关度越高, 选取互信息系数大的主轴转速、发电机转速、风速、桨距角、齿轮箱驱动端轴承温度、齿轮箱非驱动端轴承温度这 6 个参量为关键影响参量进行后续的风电机组状态监测。

以风电机组正常状态下的关键影响参量为输入, 以有功功率为输出, 训练基于 GRU 的风电机组正常行为模型。以 6:2:2 的比例划分训练集、验证集和测试集, 设置模型优化器为 Amad 优化器, 隐藏层节点数为 64。

利用 SPRT 分析有功功率回归值和实际值的残差, 根据风电机组正常行为模型输出的前 200 个测试样本计算  $H_0, H_1$  的均值  $\mu_0$  和标准差  $\sigma_0$ , 具体参数设置如表 2 所示。

表 2 假设  $H_0, H_1$  参数  
Table 2 Assume  $H_0, H_1$  parameter

假设内容	$\mu$	$\sigma$
风电机组正常运行	0.010 7	0.015 1
风电机组亚健康运行	$2\mu_0$	$2\sigma_0$

图 6 为 SPRT 的功率残差分析曲线。

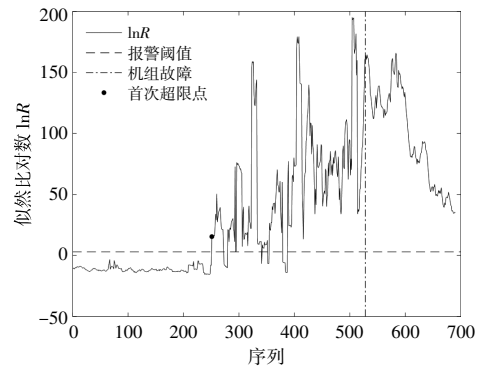


图 6 SPRT 的功率残差分析结果

Fig.6 Power residual analysis results of SPRT

由图 6 可知, SPRT 方法对功率残差的异常变化的判断灵敏, 在第 241 点, 风电机组首次超过报警阈值, 此后似然比对数不断增大且长期超过报警上限。

图 7 为风电机组状态判断结果。

由图 7 可知, 自 241 点起, 风电机组状态值大于阈值, 且 421 点后一直超限, 判断为亚健康状态。对照故障表信息可知, 该模型提前 284 点判断出风电机组的异常, 验证了所提模型对风电机组状态监测的有效性。

对于 28 号风电机组, 本文构建的基于 GRU

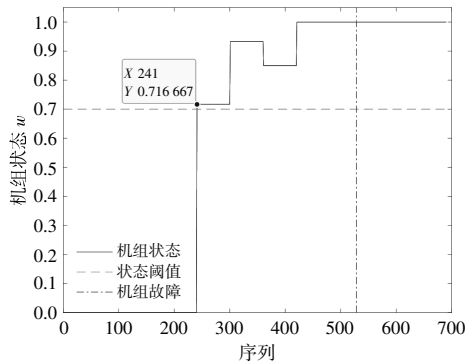
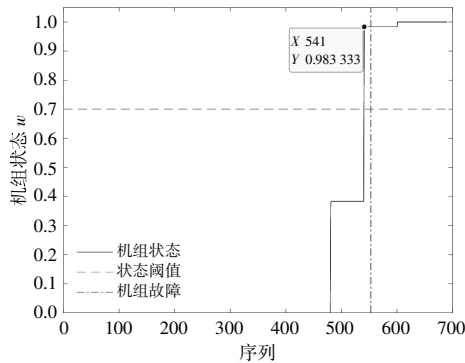


图7 风电机组状态判断结果

Fig.7 Judgment results of wind turbine condition

模型和 SPRT 方法的单风电机组状态监测模型, 以其 SCADA 数据和模型判断的结果作为源域数



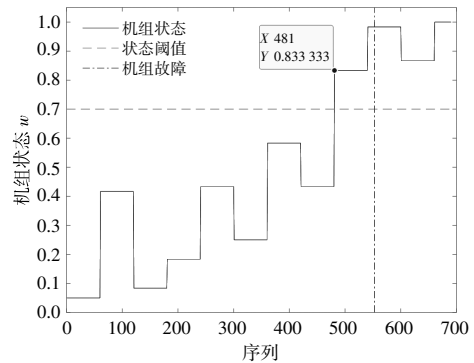
(a) 未迁移数据

据,以同一风电场的 30 号风电机组数据作为目标域数据。分别用边缘分布自适应(TCA)、联合分布适配(JDA)和 BDA 3 种迁移学习方法同化数据分布,并将迁移前后的结果输入根据 28 号风电机组训练的风电机组状态监测模型,所得精度对比结果如表 3 所示。结合时间状态阈值判断 30 号风电机组状态监测结果如图 8 所示。

表 3 3 种迁移学习方法结果对比

Table 3 The comparison results of three transfer learning methods

迁移学习	RMSE(迁移前)	RMSE(迁移后)	提升 %
TCA	13.13	8.700	4.430
JDA	13.13	6.892	6.238
BDA	13.13	5.684	7.446



(b) 迁移后数据

图 8 BDA 前后机组状态判断结果

Fig.8 Wind turbine condition judgment results before and after BDA

由表 3 可知:TCA 仅考虑拉近数据的边缘分布,精度提升幅度较低;JDA 考虑了数据的边缘分布和条件分布,精度比 TCA 提高了 1.808%;BDA 根据数据本身的性质去调整条件分布和边缘分布,模型精度比迁移前提高了 7.446%,比 TCA 提高了 3.016%,比 JDA 提高了 1.208%,说明在 GRU 模型不重复建模的情况下 BDA 显著提高了模型的精度。

由图 8 可知,经 BDA 处理数据后,亚健康状态的判断比迁移学习前提前了 60 个点。结合表 3 可知,迁移学习后模型提前判断出风电机组异常,说明迁移学习的过程没有影响 SCADA 数据的质量,迁移后的模型准确判断了风电机组的运行状态,兼顾了多风电机组运行状态监测模型的精度和效率,验证了所提模型的有效性和优越性。

## 6 结论

针对多风电机组状态监测方法精度和效率难

以兼顾的问题,本文提出了基于 BDA 学习的多风电机组运行状态监测方法,得到以下结论。

①基于 GRU 模型和 SPRT 方法的单风电机组状态监测方法能消除冗余参量的影响,可提取参量之间时序关系,提高了正常行为模型的精度,可准确判断风电机组的运行状态。

②基于 BDA 的多风电机组状态监测方法能够避免多风电机组重复建模而导致的人工建模和计算成本增加的问题,兼顾多风电机组运行状态的监测精度和效率。基于 BDA 的多风电机组运行状态监测方法的精度比迁移前提升了 7.446%,明显优于基于 TCA 和 JDA 的监测方法。

## 参考文献:

[1] 刘永前,周家慷,阎洁,等.基于多任务学习的风速实时回归方法[J].可再生能源,2021,39(4):481-487.  
 [2] Chen N, Yu R, Chen Y, et al. Hierarchical method for

- wind turbine prognosis using SCADA data [J]. *IET Renewable Power Generation*, 2017, 11(4): 403–410.
- [3] 李航涛. 风电机组功率曲线绘制与应用研究[D]. 北京: 华北电力大学, 2017.
- [4] 林涛, 张丽, 蔡睿琪, 等. 基于改进果蝇算法优化支持向量机的风电机组性能评估[J]. *可再生能源*, 2019, 37(1): 132–137.
- [5] Xiang L, Yang X, Hu A, et al. Condition monitoring and anomaly detection of wind turbine based on cascaded and bidirectional deep learning networks [J]. *Applied Energy*, 2022, 305: 117925.
- [6] 李林晏, 韩爽, 张雅洁, 等. 基于迁移成分分析的多风电机组运行状态识别方法 [J]. *分布式能源*, 2022, 7(1): 12–19.
- [7] Jia X, Han Y, Li Y, et al. Condition monitoring and performance forecasting of wind turbines based on denoising autoencoder and novel convolutional neural networks[J]. *Energy Reports*, 2021(7): 6354–6365.
- [8] Han S, Qiao Y, Yan J, et al. Mid-to-long term wind and photovoltaic power generation prediction based on Copula function and long short term memory network[J]. *Applied Energy*, 2019, 239: 181–191.
- [9] Qiao Y, Han S, Zhang Y, et al. A multivariable wind turbine power curve modeling method considering segment control differences and short-time self-dependence[J]. *Renewable Energy*, 2024, 222: 119894.
- [10] Wald A. *Sequential Analysis* [M]. Chicago: Courier Corporation, 2004.
- [11] Wang J, Chen Y, Hao S, et al. Balanced distribution adaptation for transfer learning [A]. 2017 IEEE International Conference on Data Mining (ICDM) [C]. New Orleans: IEEE, 2017. 1129–1134.

## Condition monitoring for multiple wind turbines based on balanced distribution adaptive transfer learning

Zhang Yajie<sup>1,2</sup>, Wang Luo<sup>3</sup>, Liu Yulu<sup>1,2</sup>, Yue Bo<sup>3</sup>, Han Shuang<sup>1,2</sup>, Su Ying<sup>3</sup>, Liu Yongqian<sup>1,2</sup>  
 (1.State Key Laboratory of Alternate Electrical Power System with Renewable Energy Sources (North China Electric Power University), Beijing 102206, China; 2.School of New Energy, North China Electric Power University, Beijing 102206, China; 3.China Three Gorges Corporation, Beijing 100038, China)

**Abstract:** Accurate condition monitoring of wind turbines is crucial to the safe and stable operation of wind turbines and the improvement of economic benefits. However, affected by the divergence in the distribution of operating data of different wind turbines, the existing condition monitoring methods have the problem of difficulty in taking into account the accuracy and efficiency in the application scenario of multiple wind turbines. BDA can shorten the data distance and reduce the data distribution divergence. Therefore, this paper propose a multi-wind turbine condition monitoring method based on balanced distribution adaptive transfer learning. Firstly, the mutual information method based on Copula entropy is used to mine the key influencing parameters of the wind turbine condition; then, a wind turbine condition monitoring model is established based on the GRU model and SPRT method; wind turbine operation data distribution assimilation model based on BDA is constructed, and used for multi-wind turbine condition monitoring. Results show that the proposed method can effectively save the modeling cost and calculation cost, and can significantly improve the monitoring efficiency on the premise of ensuring the monitoring accuracy of the operating state of multiple wind turbines.

**Keywords:** wind turbine; condition monitor; balanced distribution adaptive transfer learning; sequential probability ratio test; gated recurrent unit