

[风电]

# 基于SCADA数据驱动的海上风电机组运行状态诊断方法

文/张智伟  
上海海湾新能风力发电有限公司

传统海上风电机组运行状态诊断方法直接对异常数据进行判定，未对海上风电机组运行数据进行预处理，造成传统方法诊断精准度较低。提出基于SCADA数据驱动的海上风电机组运行状态诊断方法。实验结果表明该研究方法相比于传统方法诊断精准度更高，具有更好的诊断效果。

风力发电是目前世界上增长最快、具有竞争力的可再生能源。中国海上风力发电储量约为7.5亿kW，是陆地的三倍。随着海洋风能开发饱和度的不断提高，海洋风电已成为风能发展的新方向。由于台风、气流和空气照明等海上风电机组所处的封闭环境，在运行过程中不可避免地会出现运行障碍。如果能够及时准确地监测和诊断风电机组的运行状态，风电机组的故障将得到纠正。这对于海上风电机组的运行和维护非常重要，可以降低总成本和大规模风电机组损害。鉴于恶劣的环境和对风电机组故障的敏感性，风电机组的运行和维护成本仍然很高，严重限制了风电行业的发展。在我国，由于海上风能发展缓慢，可靠性测试的机组数据不足。海上风电场的SCADA数据驱动可以收集机组的实时运行数据，但尚未实现有效的机组运行状态诊断。因此，应该寻求一种有效的方法来诊断海上风电机组的运行状态。一旦机组离开了保证期，进入了运行和维护期，安装更多的风电机组，就会对风电机组运行状态和监测造成影响，增加运行和维护成本。为确保风电机组在其整个生命周期中的安全性和稳定性，SCADA通常用作风力发电机组运行状态监测数据驱动。它可以了解风力发电机的采集、监测、自动控制和信号，对反映和分析风力发电机的运行

状态具有非常重要的作用。然而，大多数SCADA数据驱动在使用过度的边界警报制度。其中，BP神经网络算法、向量机制方法和非线性故障诊断方法是建立训练矩阵所必需的，由于只有当故障发生并恶化到一定程度时才会发出提醒，因此不能发出故障提醒，也不能完全反映机组子数据驱动或组件的状态。为了监测风电机组的状态并获得故障信号，有必要在SCADA监测数据的基础上建立风电机组状态诊断和故障信号算法。目前，基于SCADA监测数据诊断风电机组运行状态的主要方法包括BP神经网络算法、向量机支持、非线性状态诊断方法和最小二乘安装法（多项式回归法）等。通过选择训练数据，预测风电机组活动和训练数据的选择对人为因素有显著影响。训练过程不需要最小二乘安装法（多项式回归法），通过比较和分析正常位置和指定位置之间的距离，确定是否符合监测数据，并对呼吸机或部件的位置进行快速诊断。然而，目前使用最低二次设备法（多项式回归法）的SCADA分析和研究是基于海上风电机组，而不是对海上风电机组的研究。本文提出基于SCADA数据驱动的海上风电机组运行状态诊断方法，分析了海上汽轮发电机SCADA数据的监测，以诊断海上发电机组运行状态。

## 基于SCADA数据驱动的海上风电机组运行状态诊断方法设计

### 1. 海上风电机组运行数据预处理

风电机组的运行条件复杂,可能会受到不同的因素影响,如风速变化、电子网络限制和湍流等。这些情况可能导致重大减值数据。因此,有必要验证筛选数据,以确保计算结果的准确性,将比较区间 $N$ 内所有有效数据置进行划分。

所选风速根据IEC61400-12风速分类标准应分为0.5 m/s的风速带,可以划分为

$$N = \frac{V_{\max} - V_{\min}}{0.5} \quad (1)$$

式中, $N$ 是分段间隔的数量; $V_{\min}$ 是使用时的最小风速值; $V_{\max}$ 是使用时的最大风速值。

将SCADA数据驱动中要建模的两组数据根据风速划分为风速区间。计算每个风速范围的预期数据监测值公式为

$$x_k = N \sum p_i x_i \quad (2)$$

式中, $p_i = (n_i/n) \times 100\%$ ; $n_i$ 为 $i$ 个单元之间的监控数据的数量; $n$ 是第 $k$ 个单元之间数据总数; $x$ 是第 $i$ 个单元间的相关 $x_i$ 值。

当监测数据包含函数关系 $y=(a_x)$ 时,使用监控数据 $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ 通风电机的时间段的第 $k$ 个多项式进行拟合,以便满足分布点

$$y_i = Na_0 + a_1 x_i + a_2 x_i^2 + \dots + a_k x_i^k \quad (3)$$

式中, $a_k$ 为具有多项式模型系数, $y_i$ 为具有原始数据 $y_i$ 设备的值。根据多项式回归方法, $y_i$ 设备的值与 $\sigma$ 数据的初始数量之间的故障率最低,即

$$\frac{\sigma(\sum [y_i - \hat{y}_i])}{\sigma a_i} x_k = 0 \quad (4)$$

令 $R^2 = \sum [y_i - \hat{y}_i]$ ,代入式(3)中可得

$$R^2 = \sum y_i^2 - 2 \sum (a_0 + a_1 x_i + a_2 x_i^2 + \dots + a_k x_i^k) \quad (5)$$

最后,得到多项式回归系数矩阵为

$$A = (R^2 x_k)^{-1} R^2 x_k \quad (6)$$

SCADA数据之间的关系决定了政治回归阶数 $k$ 的顺序。鉴于风速与有效容量之间的三次关系,本文将 $k$ 阶定为3阶。

### 2. 异常数据判定

预处理后的数据在风电机正常位置的系数定义为 $b_i$ ,在等待位置的系数向量定义为 $a_i$ 。通过引入差值 $C$ ,将显示模型的两个系数之间的距离为

$$C = A \frac{\int \sum (a_i - b_i) x^i dx}{x_{\max} - x_{\min}} \quad (7)$$

式中, $x_{\max}$ 和 $x_{\min}$ 是 $x$ 的最大值和最小值。

不同的 $C$ 值可用于确定风电机组中的差异是否处于不同的时期。 $C$ 值越高,两状态之间的差异就越大,表明等待情况与正常情况不同。这对应于正态分布 $3\sigma$ 。如果差值 $C$ 的值达到所选择的阈值,则可以确定呼吸机子数据驱动(部件)不起作用。以计算期望值 $\mu$ 标准偏差 $\sigma$ ,极限值选择标准 $\sigma$ 根据规则,可以选择正常的机器操作作为样本数据,概率很低,因此可以将 $C$ 值作为判断正态分布随机变量中异常数据的判断阈值。

### 3. 基于SCADA数据驱动实现海上风电机组运行状态诊断

本文提出了一种SCADA数据驱动的海上风电机组状态监测方法(以BCDM为代表)。通过上述计算的异常数据判断阈值 $C$ 取代其最初的独立假设,实现风电机组状态的有效判别。

SCADA数据管理器是一种用于解决状态识别问题的统计方法。整合SCADA数据的管理,风电场的SCADA数据驱动,通常以指定的时间间隔收集不同风电机组组件的状态变量和运行环境参数,监测风电机组的变量,令

$$X = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & & \vdots \\ x_{T1} & x_{T2} & \dots & x_{Tp} \end{bmatrix} \quad X \in R^{T \times R} \quad (8)$$

就本文而言,每个监测变量具有 $T$ 个观测值, $p$ 的监测变量为 $X$ 。风电机组运行状态寄存器中的位置对应于每个矩阵行中的参数,并且对应的级别标签是 $T \times R$  ( $R=1,2$ )。视为仅处于正常( $C1$ )和 $k$ 故障( $C2$ )的风电机组。库提取功能用于突出不同因素之间的相关性,如果在任何时候给出新的 $x$ 观测值,则可以使用贝叶斯公式进行计算,从而实现风电机组的状态判别。公式为

$$P(C_k | x) = X \frac{p(x | C_k) p(C_k)}{\sum p(x | C_k) p(C_k)} \quad (9)$$

式中, $p(C_k)$ 是 $C_k$ 的第一概率,则 $p(x | C_k)$ 是相关其次的概率,并且 $p(x | C_k)$ 为条件概率,考虑密度函数分母是两个风能单元的标量,假设 $k=1$ ,  $2p(x | C_k)p(C_k)$ 相同,则可以从式(9)中获得

$$P(C_k|x) \propto p(x|C_k)p(C_k) \tag{10}$$

$p(C_k)$ 的概率可以通过将数量除以样本数据的总数来获得；属于给定类别的样本数据。为了简化 $p(x|C_k)$ 的计算，通常假设属性的变量彼此无关，但事实上变量之间可能存在一些相关性。因此，本文使用基于SCADA数据驱动来诊断变量之间的相关性。

研究对象的风电机组数据上用SCADA数据记录的风速、容量、温度和湍流四个参数。基于SCADA数据函数的风电机组状态诊断过程如图1所示。

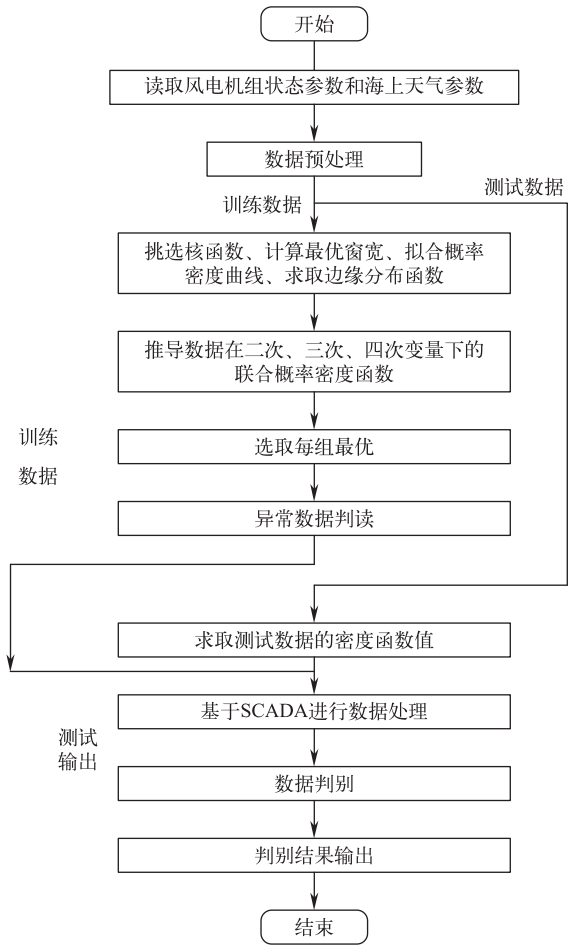


图1 基于SCADA数据函数的风能机组状态诊断过程

第一步：对风速、容量、温度和湍流等变量进行归一化，以防止不同维度的影响，减少异常数据干扰。此外，计算训练数据频率，计算Kendall水平的其他相关系数，探索数据之间的相关性特征。

第二步：选择均匀性和准确性好的高斯核函数，计算最佳窗口宽度，对应每个可变密度曲线，诊断非参数核密度，计算训练数据极限值。

第三步：使用每个变量极限值的最大概率诊断方法来计算四个参数的组合。如果风电机组处于正常状态，则将计算并使用相关的五个柯伊伯函数来优化柯伊伯功能。

第四步：计算实验数据密度函数和分布函数的值，与SCADA数据驱动程序联合识别风电机组状态。将判别方法的准确性与当地的实际操作条件登记处进行比较，比较四个判别参数组合的判别结果，并进行输出，至此完成基于SCADA数据驱动的海上风电机组运行状态诊断。

实验论证

为了验证基于SCADA数据驱动的海上风电机组运行状态诊断方法的有效性，进行了对比实验，设置传统方法1和传统方法2为对照组，对三种方法的诊断精准度进行对比。

1. 实验准备

本文将选择一台定义为5 MW的风力发电机组，对福建省某海域进行分析。机组基本信息见表1。

表1 机组基本信息

| 名称 | 额定功率/kW | 切入风速度/(m/s) | 额定风速/(m/s) | 切出风速/(m/s) | 发电机额定转速/(r/min) |
|----|---------|-------------|------------|------------|-----------------|
| 参数 | 5 000   | 3.5         | 11.8       | 25         | 1212            |

风电机组开始使用测试发电网进行并联发电。选取SCADA数据作为正常数据，用作待判和待分析的预期数据，使用三种方法分别对其进行实验，并记录实验结果见表2。

2. 对比实验

表2 三种方法结果对比

| 故障部位 | 故障存在数/个 | 传统方法1识别故障数/个 | 传统方法2识别故障数/个 | 本文方法识别故障数/个 |
|------|---------|--------------|--------------|-------------|
| 相间电路 | 10      | 4            | 6            | 10          |
| 定子绕组 | 6       | 4            | 5            | 6           |

本文设计的基于SCADA数据驱动的海上风电机组运行状态诊断方法相比于传统方法最终得出的诊断误差最小。传统方法1和传统方法2的诊断误差显著。实验结果表明本文研究方法相比于传统方法诊断精准度更高，具有更好的诊断效果，更具实用性。

## 结束语

通过分析影响海上风力发电机组运行的因素，对海上风力发电机运行的指标进行数据预处理和异常数据判定，基于SCADA数据驱动实现海上风电机组运行状态诊断，诊断结果也接近实际运行情况，具有较强的可操作性，可为海上风电机的维护提供参考依据。EA

(收稿日期：2023-08-23)

## 参考文献

- [1] 马佩强, 李斌, 白琴琴, 等. 地面自动气象观测设备运行状态信息检测技术研究与应用[J]. 中状态高新科技, 2023(8): 99-100+107.
- [2] 韩伟, 张志刚. 基于监控设备运行大数据的智能变电站运行状态分析方法研究[J]. 现代工业经济和信息化, 2023, 13(3): 46-47+50.
- [3] 梁月放. 基于物联网嵌入式单片机的电气设备运行状态远程监控[J]. 廊坊师范学院学报(自然科学版), 2023, 23(1): 39-44.
- [4] 刘忱辰, 周银莹, 秦地茂. 基于特征数据聚类的紫外线消毒数据驱动运行状态实时监测方法[J]. 中状态医疗设备, 2023, 38(1): 65-70+76.
- [5] 王利欣, 李波, 胡成军, 等. 基于物联网远程集中监控数据驱动的煤矿掘进设备运行状态研究[J]. 能源与环保, 2022, 44(11): 187-192.
- [6] 李朝瑞, 郭龙, 戚永青, 等. 基于RBF的电网大型设备运行状态实时监测方法[J]. 电子设计工程, 2022, 30(22): 108-111+116.
- [7] 王硕, 谢远森, 张群安, 等. Impala快速查询技术在配电网设备运行状态检测中的应用[J]. 自动化技术与应用, 2022, 41(8): 122-125.
- [8] 姬建富, 刘宏伟, 侯云飞, 等. 基于温度传感器的变压器冷却风机运行状态在线监控方法[J]. 工业仪表与自动化装置, 2022(4): 80-84.

