

# 基于 DCRNN 的风机 SCADA 数据故障诊断

唐宏芬, 张真真, 吴春, 曹庆才

中国大唐集团新能源科学技术研究院有限公司, 北京 西城, 100052

## Fault Diagnosis of Wind Turbines SCADA Data Based on DCRNN

Tang Hong-fen, Zhang Zhenzhen, Wu Chun

China Datang Corporation Renewable Energy Science and Technology Research Institute, Xicheng District, Beijing

**摘要:** 为准确诊断风电机组的故障类型, 充分利用计算资源, 自动识别风电机组的不同健康状态, 提出了一种新的深度卷积循环神经网络(DCRNN)。首先, 该方法采用一维卷积神经网络(1D-CNN), 网络的每一层都使用不同大小的卷积核来挖掘 SCADA 数据不同尺度的空间特征; 其次, 利用 GRU 对提取的空间特征进一步捕捉数据的时间相关性, 将空间特征提取能力和时间特征提取能力结合在一起。在一个真实风场的 SCADA 数据集上验证了该方法的优越性, 并与支持向量机(SVM)、随机森林(Random Forest)、CNN 和 GRU 方法进行了比较。

**关键词:** 深度卷积循环神经网络(DCRNN); 风电机组; 故障诊断; 深度学习

**ABSTRACT:** In order to accurately diagnose the fault types of wind turbines, make full use of computing resources, and automatically identify the different health states of wind turbines, a new deep Convolutional Recurrent Neural Network (DCRNN) is proposed. First, the method uses a one-dimensional Convolutional Neural Network (1D-CNN), each layer of the network uses different sizes of convolution kernels to mine the spatial features of different scales of SCADA data; secondly, the use of GRU to further extract the spatial features capture the temporal correlation of the data, and combine the spatial feature extraction capabilities with the temporal feature extraction capabilities. The superiority of this method is verified on a SCADA data set of a real wind field, and compared with Support Vector Machine (SVM), Random Frest (Random Forest), CNN and GRU methods.

**KEY WORD:** Deep Convolutional Recurrent Neural Network (DCRNN); wind turbine; fault diagnosis; deep learning

## 0 引言

风力发电已经成为了世界上主流的发电方式之一<sup>[1]</sup>, 因其规模的不断扩大, 风力发电机组的装机量也不断增加, 但是, 随着风电装机量的逐年增加, 故障的数量也会随之增长<sup>[2]</sup>。风力发电机组的维护和运行成本已经变得非常高<sup>[3]</sup>。为了降低运维成本, 提高风力发电机组的性能, 迫切的需要准确、高效的故障诊断方法, 以便及时发现故障<sup>[4]</sup>。许多学者正在研究更好的风机故障诊断方法, 大致可分为两类: 基于模型的诊断方法<sup>[5-7]</sup>和数据驱动的诊断方法<sup>[2],[8],[9]</sup>。前者基于严格的数学推导和动态过程, 而实际的风机监控数据采集(SCADA)数据中含有大量的非线性因素。因此, 基于模型的故障诊断方法不能很好地挖掘风机 SCADA 数据中包含的故障信息, 非线性因素严重, 限制了其在实际工程应用

中的性能。数据驱动的故障诊断方法不需要严格的数学推导, 就能够提取出风机 SCADA 数据中包含的特征。

近年来, 数据驱动的故障诊断方法得到了广泛的关注, 相关工作取得了较大进展。Nassim<sup>[10]</sup>等人提出了机器学习中的支持向量机方法用于风机故障诊断。Zhang<sup>[11]</sup>等设计了一种利用随机森林和 XGboost 的风机故障诊断方法的数据驱动故障状态检测方法。传统的故障诊断方法通常需要特征提取、特征选择和特征分类<sup>[12-13]</sup>。为了充分利用对时间序列数据的长依赖关系, Lei<sup>[14]</sup>等提出了一种利用长短时记忆网络进行风机故障诊断的方法, 实现了端到端的深度学习模型; Xia<sup>[15]</sup>等人利用多传感器和卷积神经网络进行旋转机械故障诊断, Kong<sup>[16]</sup>等利用 CNN 和门控循环单元提取风机数据的时空特征, 进

而监测风机的状态。Jiang<sup>[17]</sup>等人提出了一种利用时间信息去噪自编码器的风机故障检测方法。

通常的 CNN<sup>[18]</sup>模型，其每一层在大小上的卷积核都是均匀的，不能同时提取出不同尺度的特征<sup>[19]</sup>。本文基于目前广泛使用的 CNN 模型，提出了一种基于数据驱动端到端学习模型，能够自动提取出多尺度特征，提高诊断分析性能。与传统方法不同的是<sup>[9]</sup>，该方法不需要手动提取和选择特征，减少了决策过程中的主观因素。

## 1 SCADA 数据

风电机组 SCADA(Supervisory Control And Data Acquisition)系统通过安装在机组关键零部件上的传感器，以特定频率进行风机运行状态参数的采集、存储和远程传输。因为不同风电机组安装的 SCADA 系统来自不同厂商，所以其采样频率也不相同，采样周期一般为几秒钟至数分钟不等。同时，监测采集到的参数量也不同，一般为几十个到上百个不等，主要分为连续量与离散量两大类。现有风电机组 SCADA 系统主要采用单一参数设定阈值或趋势分析对关键零部件进行健康监测，容易出现故障误报，无法实现有效的健康监测。

风电 SCADA 系统监测的数据之间存在着复杂的关联关系，各参数之间潜在复杂的关联关系蕴含着风电机组的运行规律，正常运行状态下各参数之间的关联关系是稳定的，一旦风电机组比如叶片发生故障，各参数之间的稳定关系会被打破，并发生异常变化。

## 2 卷积神经网络

CNN 是神经网络的一种变体，由卷积层、激活函数层和池化层组成。

每个卷积层由几个卷积单元组成。卷积运算的目的是从原始数据中提取不同层次的特征。第一个卷积层可能只能提取一些低级特征。卷积层越多，提取的特征越复杂。与其他网络相比，CNN 利用稀疏连接，使核小于输入，并在相邻层的神经元之间加强局部连接模式。这样可以更有效地描述单元间复杂的相互作用，也可以降低过拟合风险。CNN 中的每个核(或权重矩阵)在整个感受野中使用，但只学习一次，而不是为每个位置学习一组单独的权重。因此，CNN 是一种非常有效的方法，它在整个输入中应用同一个局部区域的线性变换来描述变换。

假设一个卷积层的结构如下：

(1) 输入特征映射组  $X \in R^{M \times N \times D}$  为三维张量，其中每个切片矩阵  $X^d \in R^{M \times N}$  为一个输入特征映射， $1 \leq d \leq D$ ；

(2) 输出特征映射组  $Y \in R^{M' \times N' \times P}$  为三维张量，其中每个切片矩阵  $Y^p \in R^{M' \times N'}$  为一个输出特征映射， $1 \leq p \leq P$ ；

(3) 卷积核  $W \in R^{m \times n \times D \times P}$  为四维张量，其中每个切片矩阵  $W^{p,d} \in R^{m \times n}$  为一个二维卷积核， $1 \leq d \leq D, 1 \leq p \leq P$ ；

用卷积核  $W^{p,1}, W^{p,2}, \dots, W^{p,D}$  分别对输入特征映射  $X^1, X^2, \dots, X^D$  进行卷积，然后将卷积结果相加，并加上一个偏置  $b$  得到卷积层的净输入  $Z^p$ ，再经过非线性激活函数后得到特征映射  $Y^p$ 。

$$Z^p = W^p \otimes X + b^p = \sum_{d=1}^D W^{p,d} \otimes X^d + b^p$$

$$Y^p = f(Z^p)$$

其中， $W^p \in R^{m \times n \times D}$  为三维卷积核， $f(\cdot)$  为线性激活函数，一般用 ReLU 函数。

若希望卷积层输出  $P$  个特征映射，重复计算  $P$  次，得到  $P$  个输出特征映射  $Y^1, Y^2, \dots, Y^P$ 。整个计算流程图 1 所示。

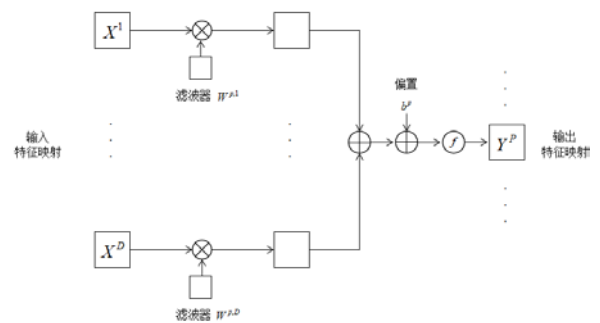


图 1 卷积层从输入特征映射组到输出特征映射组的计算示例

Fig.1 Example of calculating input feature mapping to output feature mapping using convolution layer

激活函数层可以增强判定函数和整个神经网络的非线性特性，而本身并不会改变卷积层。相比于其他激活函数，ReLU 得到了广泛的应用，因为它可以显著提高神经网络的训练速度，而不会对模型的泛化准确度造成显著影响。其定义如下：

$$f(x) = \max(0, x)$$

池化层是 CNN 的另一个重要部分。其作用是进行特征选择，降低特征数量，从而减少网络参数

数量,这也避免了过拟合[32]。直观地看,这种机制是有效的,因为特征的精确位置远不如其相对位置重要。

假设池化层的输入特征映射组为  $X \in R^{M \times N \times D}$ , 对于其中每一个特征映射  $X^d$ , 将其划分为很多区域  $R_{m,n}^d, 1 \leq m \leq M', 1 \leq n \leq N'$ , 这些区域可以重叠,也可以不重叠。常用的池化函数有两种,分别是最大池化(Maximum Pooling)和平均池化(Mean Pooling)。最大池化是取一个区域内所有神经元的最大值,其公式如下:

$$Y_{m,n}^d = \max_{i \in R_{m,n}^d} x_i$$

其中,  $x_i$  为区域  $R_{m,n}^d$  内每个神经元的激活值。平均池化是取区域内所有神经元的平均值,其公式如下:

$$Y_{m,n}^d = \frac{1}{|R_{m,n}^d|} \sum_{i \in R_{m,n}^d} x_i$$

对每一个输入特征映射  $X^d$  的  $M' \times N'$  个区域进行子采样,得到池化层的输入特征映射  $Y^d = \{Y_{m,n}^d\}, 1 \leq m \leq M', 1 \leq n \leq N'$ 。

### 3 用于风机故障诊断的 DCRNN 模型

用于风机故障诊断的 DCRNN 模型框架如图 1 所示。DCRNN 的输入是原始的多变量 SCADA 时间序列,模型的输出是故障状态标签。我们提出的 DCRNN 以端到端方式依次提取隐藏在原始 SCADA 数据中的时空相关特征。如图 2 所示,本文提出的框架主要分为三个阶段:空间特征提取,时间特征提取和最终分类。具体来说,我们首先建立一个 CNN 模型,探索 SCADA 数据中多个传感器变量之间的空间关系,然后利用 GRU 对提取的空间特征进一步捕捉时间相关性。最后,将学习到的时空特征输入 softmax 层,进行多类故障分类任务。所提框架中每个阶段的更多细节描述如下。

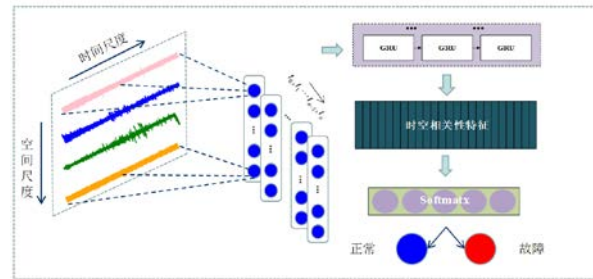


图 2 用于风机故障诊断所提的 DCRNN 模型架构  
Fig. 2 The proposed DCRNN model structure

#### 3.1 基于 CNN 模块的空间多尺度特征学习

CNN 以其强大的特征提取能力,广泛应用到了各个领域,本文从空间特征提取角度,利用 CNN 端到端地提取 SCADA 数据中不同变量间的关系。首先设计单个卷积神经网络模块,结构图如图 2 所示,每个卷积神经网络模块包括一层卷积层、一层池化下采样层、一层 Dropout 层和 BatchNormalization 层,本文设计的 CNN 网络模型由三层卷积神经网络模块组成,卷积核大小分别为  $k_1, k_2, k_3$ ,从空间尺度提取 SCADA 输入数据中不同变量间的空间特征,所得到的特征学列仍然具备时序关系,从而可以继续输入到下一个时序特征提取模块。

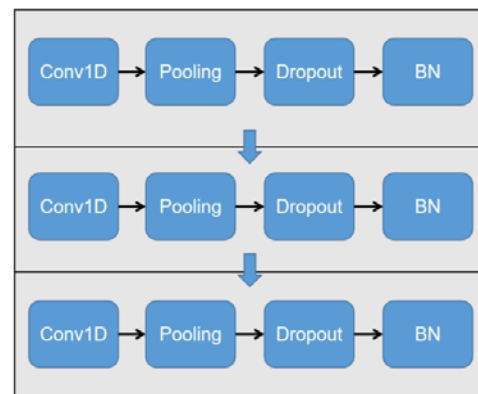


图 3 CNN 空间特征提取模块  
Fig.3 CNN for spatial features extract

#### 3.2 基于 GRU 的时间特征学习

风机故障往往不是突然发生的,而是由于监测数据发生了一些长期的缓慢而微妙的变化。CNN 模块仅挖掘了 SCADA 数据的空间特征,并没有考虑数据的时间相关性。我们利用 GRU 进一步挖掘数据的时间相关性,丰富故障特征,为多类故障的分类做准备。

### 3.2.1 循环神经网络

循环神经网络(RNN)是一种用于处理序列数据的神经网络。然而,标准 RNN 在训练过程中存在梯度消失和梯度爆炸的问题,这将影响最终的训练效果。此外,标准的 RNN 不能记住长期的数据,并且会丢失相关的早期特征,即“记忆衰退”。

### 3.2.2 GRU

门控循环单元 (GRU) 是一种流行的 RNN 体系结构,图4所示为 GRU 的原理图。与传统的 RNN 相比,隐藏单元被门控单元所取代,门控单元可以调节单元内部的信息流,而不需要单独的记忆单元。

GRU 层的计算描述如下:

$$z_t = \sigma_g (W_z x_t + U_z h_{t-1} + b_z)$$

$$r_t = \sigma_g (W_r x_t + U_r h_{t-1} + b_r)$$

$$\tilde{h}_t = \sigma_h (W_h x_t + U_t (r_t \otimes h_{t-1}) + b_h)$$

$$h_t = (1 - z_t) \otimes h_{t-1} + z_t \otimes \tilde{h}_t$$

其中,可学习的参数包括  $W \in R^{d \times k}$ ,  $V \in R^{d \times d}$ ,  $b \in R^d$ ,  $\otimes$  表示元素积,  $d$  和  $k$  表示输入和隐藏向量维数。 $x_t$  和  $h_t$  是输入和输出向量,  $z_t$  和  $r_t$  是更新门向量和重置门向量,  $\sigma_g$  和  $\sigma_h$  是激活函数。

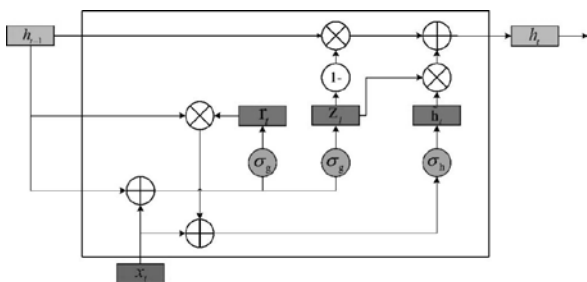


图4 GRU 结构图

Fig.4 The structure of GRU

GRU 可以避免标准 RNN 中的问题。更新门向量和重置门向量可以决定输入中的哪些信息应该被记住,哪些应该被忘记,并且针对不同的特征学习自适应权重。另外,与 LSTM 相比,GRU 保持近似的性能,并且参数更少,可以抑制过拟合。图5为使用 GRU 提取 SCADA 数据的时间相关性示意图。

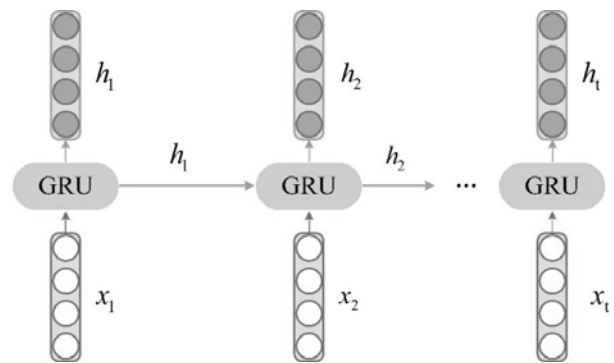


图5 时间特征提取原理图

Fig.5 Schematic diagram of temporal feature extraction

图5展示了基于 DCRNN 模型的状态监测流程。它主要包括两个阶段:离线训练阶段和在线诊断阶段。考虑到 SCADA 数据中不同传感器变量的值量纲不同,因此进行数据归一化处理,将多变量数据转换到相同的范围,使每个变量对模型的贡献相等:

$$K' = \frac{K - K_{min}}{K_{max} - K_{min}}$$

其中  $K$  和  $K'$  分别为原始数据和归一化数据。 $K_{min}$  和  $K_{max}$  分别是每个传感器变量数据  $K$  的最小值和最大值。

具体步骤描述如下:

1) 离线训练阶段:这一阶段的目的是利用来自风电场的大量带标记的 SCADA 数据,包括无故障数据和来自多个故障状态的故障数据,训练基于 DCRNN 方法的多类故障分类器。首先对大量历史数据进行数据归一化预处理,然后作为训练集进行 DCRNN 模型训练。经过训练后,最后根据验证精度保存最优模型,并用于在线诊断。

2) 在线诊断阶段:新的在线监测测量数据样本首先经过相同的数据预处理步骤,然后被送入在离线训练阶段训练良好的 DCRNN 模型中。该模型将自动进行时空特征提取和计算,最后给出故障分类结果,即故障标签。

## 4 实验

### 4.1 数据说明

选用某实际风场中的多维 SCADA 时间序列数据<sup>[20]</sup>,具有10min的采样时间间隔。数据集共有40000个样本,为了训练和测试模型,把数据划分成如表1所示的训练集、验证集、测试集。各个数据文件均大约有400个样本左右,一个文件就是一个标

签。若标签为1,则表明在最近一周内出现了风电机组叶片开裂故障,若为0,则表明风电机组没有发生叶片开裂故障即处于正常状态。

表1 数据集划分  
Table 1 Data set division

数据集	时间间隔	样本数
训练集	10min	25000
验证集	10min	75000
测试集	10min	75000

风电机组的 SCADA 系统实时采集蕴含机组状态信息的工况参数、环境参数、控制参数和状态参数。是实时掌握各关键子系统健康运行状态的重要数据。本文所用 SCADA 数据集共包括与风电机组叶片有联系的75维特征,表2给出了其中的一部分特征。

表2 部分特征  
Table 2 Partial features

编号	参数	编号	参数
1	轮毂转速	6	变频器出口压力
2	轮毂角度	7	发电机空气温度 1
3	叶片 1 角度	8	发电机空气温度 2
4	叶片 2 角度	9	机舱控制柜温度
5	叶片 3 角度	10	液压制动压力

#### 4.2 对比分析

为了评估 DCRNN 方法的性能,采用传统的浅层机器学习方法以及几种深度学习方法进行对比,主要有支持向量机(SVM)、随机森林(Random Forest)、CNN、GRU 等。

对比方法设置如下:

RF: 500个估计量,8个叶节点的随机森林分类器。

SVM: 在{1, 5, 10, 50, 100, 200}和{0.0001, 0.0005, 0.001, 0.01, 0.02}的范围内,通过5次交叉验证,确定最优参数 C 和 r。核函数为径向基函数。

CNN: 一维卷积神经网络,能够挖掘 SCADA 数据中的空间特征。

GRU: 循环神经网络的变体。能够挖掘 SCADA 数据中固有的时序特征。

所有深度学习模型的模型参数均为:选择批量大小32;选择优化器 Adam;设定学习率为0.001;最大迭代次数为1000次。

在本文中,风机故障诊断的目标是对健康状态进行分类。为了评估对比模型的性能,计算总体和每一类的评估指标,准确率、召回率和 f1-score 的相关计算公式如表3所示。

表3 评价指标及其计算公式  
Table 3 Evaluation metrics and calculation formulas

评价指标	计算公式
准确率	$\frac{tp}{tp + fp}$
召回率	$\frac{tp}{tp + fn}$
f1-score	$\frac{2(\text{精度召回率})}{\text{精度召回率}}$

其中  $tp$ 、 $tn$ 、 $fp$  和  $fn$  分别为第  $i$  类中的真阳性、真阴性、假阳性和假阴性。

#### 4.3 对比结果

几种对比方法的实验准确率对比如图8所示,从中可以看出本文所提的 DCRNN 网络模型的准确率最高,为92.3%,高于其他对比模型。

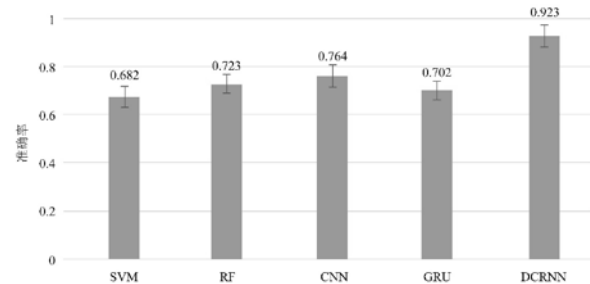


图6 不同方法的分类准确率

Fig.6 Classification accuracy of different methods

浅层机器学习方法网络较浅,只能挖掘数据的浅层特征,所以其效果不佳;单独的 CNN 只能挖掘 SCADA 数据的空间特征,单独的 GRU 只关注 SCADA 数据的时间相关性;而本文所提的 DCRNN 模型既考虑了不同传感器之间的空间相关性,又考虑了同一传感器的时间关联性,所以 DCRNN 拥有最佳的分类性能。

#### 5 结论

针对多变量 SCADA 数据在空间上和时间上的特征难以挖掘的问题,本文提出了一种新的 DCRNN 模型,用于风电机组的故障诊断。模型的核心思想是结合3层卷积神经网络和 GRU 方法,将空间特征学习和时间特征学习能力整合到一个统一的框架中。采用离线训练模型、在线故障诊断的策

略,在实际风场的 SCADA 数据集上进行实验评估,与多种机器学习故障诊断方法和多种深度学习故障诊断方法对比,证明了该方法的有效性与可靠性。

## 参 考 文 献

- [1] Tong W. Wind power generation and wind turbine design[M]. WIT press, 2010.
- [2] Yu D, Chen Z M, Xiahou K S, et al. A radically data-driven method for fault detection and diagnosis in wind turbines[J]. International Journal of Electrical Power & Energy Systems, 2018, 99: 577-584.
- [3] Qiao W, Lu D. A survey on wind turbine condition monitoring and fault diagnosis—Part I: Components and subsystems[J]. IEEE Transactions on Industrial Electronics, 2015, 62(10): 6536-6545.
- [4] Wu X, Jiang G, Wang X, et al. A multi-level-denoising autoencoder approach for wind turbine fault detection[J]. IEEE Access, 2019, 7: 59376-59387.
- [5] Isermann R. Model-based fault-detection and diagnosis—status and applications[J]. Annual Reviews in control, 2005, 29(1): 71-85.
- [6] Isermann R, Balle P. Trends in the application of model-based fault detection and diagnosis of technical processes[J]. Control engineering practice, 1997, 5(5): 709-719.
- [7] Simani S, Fantuzzi C, Patton R J. Model-based fault diagnosis techniques[M]//Model-based Fault Diagnosis in Dynamic Systems Using Identification Techniques. Springer, London, 2003: 19-60.
- [8] Yin S, Wang G, Karimi H R. Data-driven design of robust fault detection system for wind turbines[J]. Mechatronics, 2014, 24(4): 298-306.
- [9] Pashazadeh V, Salmasi F R, Araabi B N. Data driven sensor and actuator fault detection and isolation in wind turbine using classifier fusion[J]. Renewable Energy, 2018, 116: 99-106.
- [10] Laouti N, Sheibat-Othman N, Othman S. Support vector machines for fault detection in wind turbines[J]. IFAC Proceedings Volumes, 2011, 44(1): 7067-7072.
- [11] Zhang D, Qian L, Mao B, et al. A data-driven design for fault detection of wind turbines using random forests and XGboost[J]. IEEE Access, 2018, 6: 21020-21031.
- [12] Ruiz M, Mujica L E, Alferez S, et al. Wind turbine fault detection and classification by means of image texture analysis[J]. Mechanical Systems and Signal Processing, 2018, 107: 149-167.
- [13] Gong W, Chen H, Zhang Z, et al. A novel deep learning method for intelligent fault diagnosis of rotating machinery based on improved CNN-SVM and multichannel data fusion[J]. Sensors, 2019, 19(7): 1693.
- [14] Lei J, Liu C, Jiang D. Fault diagnosis of wind turbine based on Long Short-term memory networks[J]. Renewable energy, 2019, 133: 422-432.
- [15] Xia M, Li T, Xu L, et al. Fault diagnosis for rotating machinery using multiple sensors and convolutional neural networks[J]. IEEE/ASME Transactions on Mechatronics, 2017, 23(1): 101-110.
- [16] Kong Z, Tang B, Deng L, et al. Condition monitoring of wind turbines based on spatio-temporal fusion of SCADA data by convolutional neural networks and gated recurrent units[J]. Renewable Energy, 2020, 146: 760-768.
- [17] Jiang G, Xie P, He H, et al. Wind turbine fault detection using a denoising autoencoder with temporal information[J]. IEEE/Asme transactions on mechatronics, 2017, 23(1): 89-100.
- [18] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[J]. Advances in neural information processing systems, 2012, 25: 1097-1105.
- [19] Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2015: 1-9.
- [20] 风电叶片开裂故障预警 DataFountain 大数据竞赛 . <https://www.datafountain.cn/competitions/302>

### 作者简介:

唐宏芬(1977-),女,广西南宁,博士,高级工程师,研究方向为风电机组控制策略优化、风电机组健康状况诊断与故障预警

张真真(1991-),女,山东枣庄,硕士,工程师,研究方向为风机故障预警和风电数据分析

吴春(1973-),男,河北宣化,硕士,高级工程师,研究方向为风机震动、继电保护

曹庆才(1991-),男,山东临沂,硕士,工程师,研究方向为风电机组设计、智慧电站