

基于一维卷积神经网络的风机故障预测

梁 川

(华润江中制药集团有限责任公司,江西 南昌 330096)

摘 要:风电场风机叶片结冰会造成风电机组的效率降低,严重时会导致叶片断裂,严重威胁风电场的正常生产运行。文章提出了一种基于一维卷积神经网络的风机叶片结冰预测方法,能够根据风机运行的 SCADA 数据对叶片结冰的早期过程进行精确预测,并采用某风电场的 2 台风机电数据进行模型验证和对比试验,并提出了针对不平衡数据集处理的加权精确率的评价指标。实验结果表明文章所提方法能有效预测风机的早期故障,对设备作出预测性维护,减少经济损失。

关键词:卷积神经网络;叶片结冰;故障预测

中图分类号:TH165

文献标志码:A

文章编号:2095-2945(2020)32-0132-04

Abstract: Ice accumulation on wind turbine blades of wind farms will reduce the efficiency of wind turbines. In severe cases, the blades will break, which seriously threatens the normal production and operation of wind farms. This paper proposes a wind turbine blade icing prediction method based on a one-dimensional convolutional neural network, which can accurately predict the early process of blade icing based on the SCADA data of the wind turbine operation. Using 2 wind turbine data of a wind farm, this paper conducts model verification and comparative experiments, and then puts forward evaluation indexes of weighted accuracy rate for imbalanced data set processing. The experimental results show that the method proposed in this paper can effectively predict the early failure of the fan, make predictive maintenance on the equipment, and reduce economic losses.

Keywords: convolutional neural network; blade icing; fault prediction

1 概述

随着经济的发展,人们对于能源的依赖逐步加强,化石燃料能源对生态环境负面影响很大,因此,可再生能源的发展备受关注。其中,风能便是常用可再生能源之一^[1]。但将风能转化为电能的风机时常会受到低温潮湿环境影响,叶片容易发生结冰现象,从而影响风机的发电性能和安全性^[2]。因此,对风机叶片结冰预测研究具有较高的应用价值。

故障预测与健康管理^[3](PHM)是指利用传感器采集系统的数据信息,借助于信息技术、人工智能推理算法来监控,在系统发生故障之前对其故障进行预测,并结合现有的资源信息提供一系列的维护保障建议或决策,它是一种集故障检测、隔离、健康预测与评估及维护决策于一身的综合技术。在整个过程中,故障的预测的精确性和准确性对整个系统起的关键作用,具有很大的潜在研究价值。Mostafa B 等人^[4]提出了一种基于时域和频域(FFT 相角和 PSD)统计分析、主成分分析(PCA)和自适应神经模糊推理系统(ANFIS)的智能故障预测方法。Amruthnath N 等人^[5]提出了一种基于模型聚类的无监督学习故障类别预测,避免了对历史数据的需求。

国内外对于风机叶片结冰的故障预测的研究尚处于起步阶段,由于风机叶片本身结构复杂,因此一般的结冰状态监测主要是在风机叶片上加装叶片传感器。Simani 等^[6]提出了数据驱动方法对风机的早期故障进行检测,主要运用了

模糊系统和神经网络来描述测量和故障之间的强非线性关系。Hongyuan Xiao 等^[7]采用人工蜂群(ABC)和 EMD 分解方法对支持向量机(SVM)的模型进行优化,提高了对风机故障诊断和预测的精度。张明德等^[8]提出了一种基于多尺度卷积策略 CNN 的滚动轴承故障诊断方法,其将表征滚动轴承故障的一维振动信号输入模型,然后利用卷积层中的多尺度卷积结构对原始信号进行卷积运算,最后通过池化层进行特征信息的提炼和简化,从而实现滚动轴承故障诊断。

本文以国内某风机场的 2 台风机电数据分别进行模型训练和测试,采用了一种基于一维卷积神经网络的模型结构对风机叶片结冰进行预测,有效提高对风机结冰故障的预测准确率。提出了针对不平衡数据集处理的加权精确率评价指标,提升了结果的可靠性。

2 基于一维卷积神经网络的风机故障预测算法

卷积神经网络(Convolutional Neural Network, CNN)是一种深度学习架构^[9]。其独特的多卷积层结构使得模型具有学习序列数据更深层次特征的能力,提高模型的鲁棒性。卷积运算是 CNN 的核心操作,其本质就是对数据进行特征提取,其通过不同的卷积核对输入数据进行卷积运算,从而获取数据的全局化特征,再经过池化操作对提取到的特征进行下采样,在降低运算量的同时也可以一定程度抑制模型的过拟合。卷积神经网络主要涉及三个基本概念^[8]:局部感知、参数共享和池化。在卷积层中,每一个卷积核都会连接到输入数据的一个小区域上,然后通过时间步

作者简介:梁川(1967-),男,硕士,高级工程师,研究方向:智能制造。

的方式在输入矩阵上进行不断扫描,这个过程就是局部感知。卷积核通过对扫描区域不断进行内积运算,从而实现特征映射。其数学模型表述如下:

$$X_{t:t+k-1} = x_t \oplus x_{t+1} \oplus \cdots \oplus x_{t+k-1}$$

$$z_t = f(W^T X_{t:t+k-1} + b)$$

$$P_t = \max(z_t)$$

其中, $X_{t:t+k-1}$ 表示经过时间步处理的时间序列, k 表示卷积核的大小; W^T 为共享的权重参数, b 为共享偏置, 共享权重和共享偏置的机制使得模型的训练参数大大减少, 也使得卷积神经网络学到的模式具有了平移不变性, 从而使得卷积神经网络在训练模型时可以高效利用数据, 提高模型的训练速度和泛化能力。 $f(\cdot)$ 为激活函数, 激活函数为模型引入了非线性因素, 使得模型的表达能力大大增强。 $\max(\cdot)$ 函数为最大池化函数, 其可以在保留卷积层特征的同时降低模型训练的数据量。

卷积神经网络最初是被设计出来处理图像数据, 由于每次都是直接在图像矩阵上用同样的卷积核像素点进行卷积操作, 所以与传统的全连接神经网络相比, 卷积神经网络能够大幅降低模型训练所需要的参数, 从而提高模型的训练速度。一维卷积神经网络^[9]是卷积神经网络的演变版, 可以很好地应用于传感器数据的时间序列分析, 与专门用来处理时间序列数据的循环神经网络相比, 一维卷积神经网络具有结构更加简单, 处理速度更快的优点。图 1 为一维卷积神经网络的工作原理, 其利用一维卷积核在时间序列上进行滑动, 由于其每次处理的数据为一个时间周期而不是某一时刻的数据, 所以它能够提取到序列数据在时间维度上体现出来的前后变化关系, 从而学习到更多的数据信息, 进而提高模型的精度以及泛化能力。

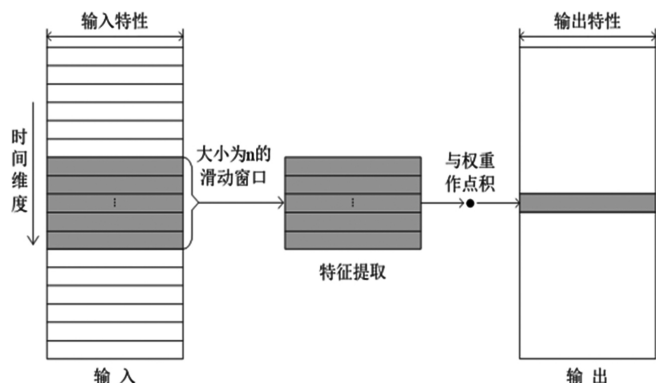


图 1 一维卷积神经网络的工作原理

3 实验

本文采用的数据集来自 2017 年国内首届工业大数据创新竞赛, 该数据集包含两台风机近两个月的 SCADA 系统数据及对应的正常和故障状态的时间段, 数据集共有 26 个特征维度, 共包含 584382 条样本。

3.1 数据预处理

数据集中包含了正常数据、故障数据以及无效数据, 分别将其对应的样本标签标记为 0, 1, 2。其中, 对于无效数

据, 在实际的模型训练中将其删除。如图 2 所示, 在风机的 26 个特征维度中, 由于风机三个叶片的相同属性值表现出强线性关系, 可通过取这些相同属性的平均值来代替原有的多个特征。经处理后, 数据集剩余 553715×20 组数据。

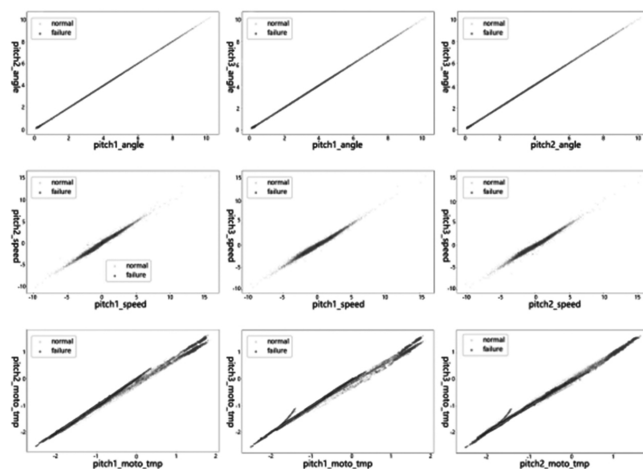


图 2 特征相关性示意图

为了将单纯的分类问题转换成故障预测问题, 需要对风机数据集进行重新分割, 但该数据集本身存在数据采样周期不一致的问题, 需要对数据进行重采样, 将数据处理成等时间间隔的标准时序数据。通过计算, 该数据集采样的平均周期维持在 7.5s 左右, 考虑到神经网络对数据量的要求比较高, 本文采用周期为 10s 对原数据集进行重新采样, 把数据分割成等时间间隔的时序数据。通过重采样处理好的数据为一个时间步对应一个标签, 但是这种数据无法被用作故障预测, 故障预测问题需要用设备历史数据来推断设备未来的状态, 所以需要重新对数据进行重新分割。如图 3 所示, 经过多次尝试, 本文最终选择每次取 20 条数据作为一个时间窗口, 将其作为历史数据, 再取对应时间窗口后面的第 10 条数据的标签值作为设备未来状态的标签, 最后以 10 为步长在数据集上进行滑动, 将数据集重新分割成符合故障预测的标准数据集。经数据分割处理后, 数据集被分为 55369 组大小为 (20, 20) 的新样本。

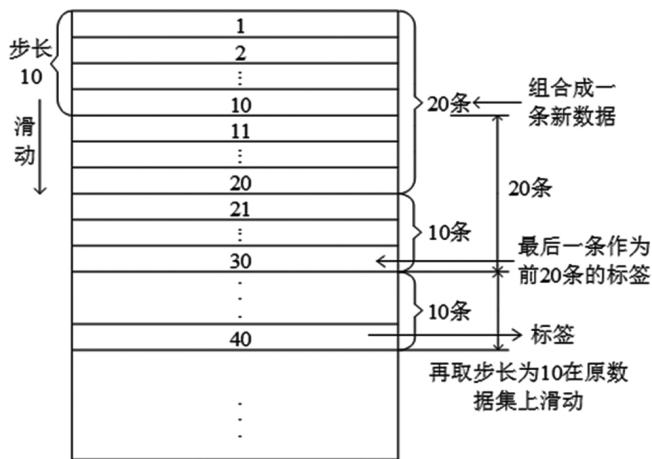


图 3 数据集分割示意图

3.2 数据集划分及模型训练

在处理完成的新数据集中取最后的 15369 条样本作

为模型最终的测试集,其中故障样本971条,正常样本14398条。将前面的40000条样本按7:3的比例划分为训练集与验证集,将数据输入所提模型进行训练,模型结构如图4所示。

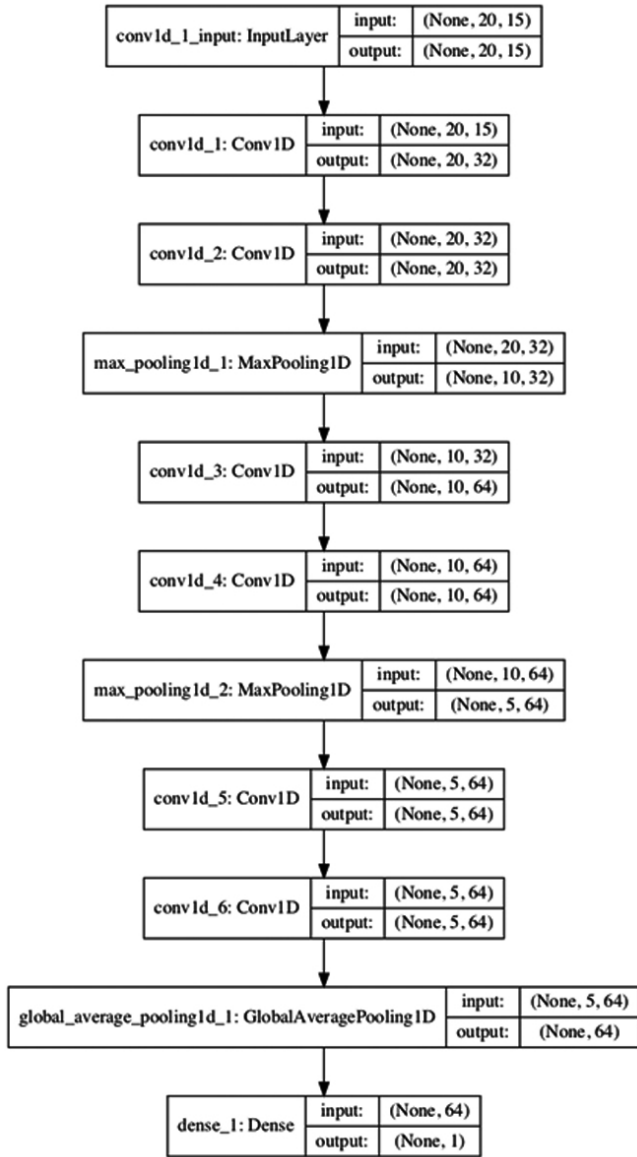


图4 模型结构

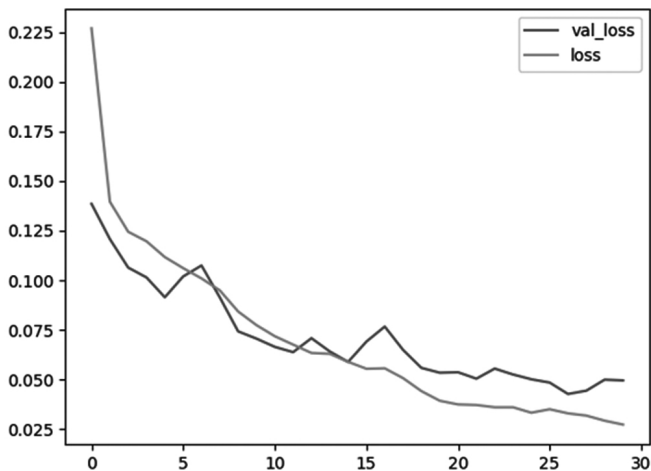


图5 模型损失曲线

图5和图6分别展示了模型训练过程中训练集和验证集上的损失和准确率的变化曲线,其中横坐标为模型训练次数。从图中可以看出,无论是训练集还是验证集,损失值整体呈下降趋势,精度呈上升趋势,模型拟合情况良好,没有出现过拟合情况。后期的精度虽然出现了一定的下降,但仍具有上升趋势,通过增加训练次数可能让模型具有更好的表现效果。

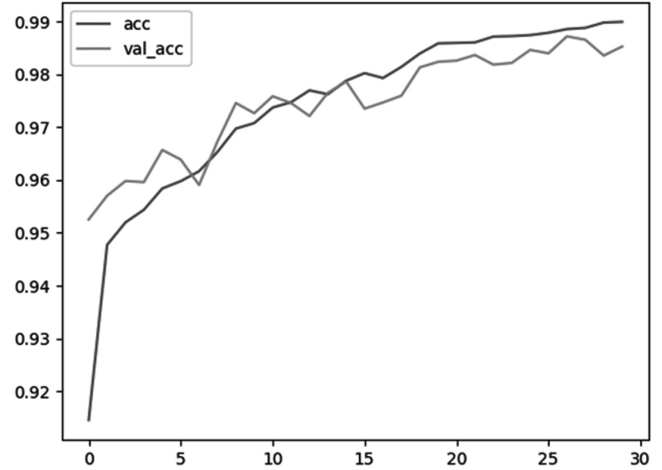


图6 模型准确率曲线

4 实验结果分析

在测试集上对模型进行测试,结果如图5所示,其中上方部分为真实值,下方部分为预测值。

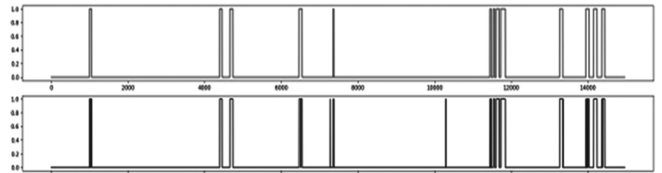


图7 模型预测结果

为判定此算法的性能,本文采用了准确率(Accuracy)、精确率(Precision)、召回率(Recall)、特异度(Specificity)和加权准确率(Weighted-Accuracy)。各个指标的定义如下:

$$ACC = \frac{TP + TN}{TP + TN + FP + FN}$$

$$PRE = \frac{TP}{TP + FP}$$

$$REC = \frac{TP}{TP + FN}$$

$$SPE = \frac{TN}{TN + FP}$$

$$WA = \frac{TP}{TP + FN} * \frac{TN + FP}{TP + FP + FN + TN} + \frac{TN}{TN + FP} * \frac{TP + FN}{TP + FP + FN + TN}$$

其中,TP指正类判定为正类,FP指负类判定为正类,FN指正类判定为负类,TN指负类判定为负类。准确率是指分类模型所有判断正确的结果占总观测值的比重;精确率是指被分为正类的示例中实际为正类的比例;召回率是指所有正类样本中被预测正确的概率;特异度是值在所有负类样本中被预测正确的概率;加权准确率是综合考虑正类和负类的样本数量,使用加权的方法来解决样本不平衡所带来的评价不严谨的问题。

表 1 本文所提方法的风机故障预测效果

性能指标	数值
准确率	96.92%
精确率	99.46%
召回率	97.23%
特异度	92.18%
加权正确率	92.46%

表 1 展示了本文所提方法在各个评价指标下的故障预测效果。

5 结论

现有风电机组结冰预测方式的准确率和精确率还有一定欠缺,本文提出的基于一维卷积神经网络的风机结冰故障预测方法,对准确率和精确率有了些许提升,该方法在风机预测时提前预测给出结冰预警,有效避免因风机结冰而产生的不必要损失,有效运用了风机的数据,对风机结冰故障预测的研究有实际意义。

使用加权精确率的评价指标可以综合正类的准确率和负类的准确率,进一步降低原始数据不平衡而导致的单个精确率的可参考性。

参考文献:

- [1]薛阳,王琳,王舒,等.一种结合 CNN 和 GRU 网络的超短期风电预测模型[J].可再生能源,2019(3):456-462.
- [2]张延伍.可再生能源政策对发电行业的影响分析[D].华北电力大学,2019.

[3]曾声奎,Michael G. Pecht, 吴际.故障预测与健康管理(PHM)技术的现状与发展[J].航空学报,2005,26(5):626-632.

[4]Mostafa B, Hossein J, Ebrahim E. Intelligent Prediction of Fault Severity of Tractor's Gearbox by Time-domain and Frequency-domain(FFT phase angle and PSD) Statistics Analysis and ANFIS [J]. Journal of Mechanical Engineering, 2018,47(1): 51.

[5]Amruthnath N, Gupta T. Fault class prediction in unsupervised learning using model-based clustering approach [C]// International Conference on Information & Computer Technologies. IEEE, 2018.

[6]Simani S, Farsoni S, Castaldi P. DATA-DRIVEN TECHNIQUES FOR THE FAULT DIAGNOSIS OF A WIND TURBINE BENCHMARK[J]. International Journal of Applied Mathematics and Computer ence, 2018,28(2):247-268.

[7]Hongyuan Xiao, Yuejia Sun, Yibo Jin. Optimization about Fault Prediction and Diagnosis of Wind Turbine Based on Support Vector Machine [C]// International Conference on Measuring Technology & Mechatronics Automation. IEEE Computer Society, 2018.

[8]张明德,卢建华,马婧华.基于多尺度卷积策略 CNN 的滚动轴承故障诊断[J].重庆理工大学学报(自然科学),2020(06):102-110.

[9]周飞燕,金林鹏,董军.卷积神经网络研究综述[J].计算机学报,2017(6):1229-1251.

[10]谢远东,雷文平,韩捷.基于全矢 1D-CNN 的轴承故障诊断研究[J].设备管理与维修,2019(15):21-24.

(上接 131 页)

溶液平均浓度。 $u_{r(0)}$ 的自由度 $v_{r(0)}=mn-2=13$ 。

3.3 合成不确定度的评定

3.3.1 标准不确定度分量一览表

表 3 标准不确定度分量一览表

名称	不确定度来源	标准不确定度值
$u_{r(a)}$	测量重复性	2.11×10^{-3}
$u_{r(b)}$	标准溶液及其配制过程	3.7×10^{-3}
$u_{r(c)}$	测试仪器	2.0×10^{-2}
$u_{r(d)}$	标准曲线绘制	1.2×10^{-2}

3.3.2 合成标准不确定度 u 的计算

$$u = \sqrt{u_{r(a)}^2 + u_{r(b)}^2 + u_{r(c)}^2 + u_{r(d)}^2}$$

$$= \sqrt{(2.11 \times 10^{-3})^2 + (3.7 \times 10^{-3})^2 + (2.0 \times 10^{-2})^2 + (1.2 \times 10^{-2})^2}$$

$$= 0.024$$

3.4 扩展不确定度的评定

置信水平为 0.95 时,取 $k=2$,计算得水中硅含量的扩

展不确定度为 $U=K \times u=2 \times 0.024=0.048$ 。

4 测量不确定度的报告

用电感耦合等离子体发射光谱法测定水和废水中硅含量的测量不确定度结果为 $(2.95 \pm 0.048) \text{mg/L}$, $k=2$, $P=95\%$ 。

5 结果讨论

结合表 3 中各个标准不确定度分量的大小可知,采用 ICP-MS 测定水中硅含量的实验过程中,其不确定度影响因素主要来自于测试仪器误差、标准工作曲线拟合以及标准工作液的稀释配制,因此,在日常检验工作中,相关机构应注重大型仪器的保养维护,使用精度较高的量具和有证标准物质试剂,同时检验人员还需严格按照标准规范操作,以尽可能减少测量误差,此外,对于特殊样品还可以增加平行测定次数,以便得到更准确的测定结果。

参考文献:

- [1]化学分析中不确定度的评估指南 CNAS-GL006:2019[S].
- [2]水质 32 种元素的测定 电感耦合等离子体发射光谱法 HJ766-2015[S].
- [3]生活饮用水标准检验方法金属指标 GB/T5750.6-2006[S].