

控制与决策

Control and Decision

基于时序上采样卷积神经网络的风机叶片结冰检测

姜娜, 严蜜, 李柠

引用本文:

姜娜, 严蜜, 李柠. 基于时序上采样卷积神经网络的风机叶片结冰检测[J]. *控制与决策*, 2022, 37(8): 2017–2025.

在线阅读 View online: <https://doi.org/10.13195/j.kzyjc.2020.1657>

您可能感兴趣的其他文章

Articles you may be interested in

[基于W-DenseNet的减压阀不平衡样本故障诊断模型](#)

W-DenseNet-based fault diagnosis model of pressure-reducing valve with unbalanced samples

控制与决策. 2022, 37(6): 1513–1520 <https://doi.org/10.13195/j.kzyjc.2020.1708>

[自适应感受野网络的行人重识别](#)

Adaptive receptive network for person re-identification

控制与决策. 2022, 37(1): 119–126 <https://doi.org/10.13195/j.kzyjc.2020.0505>

[基于深度时序特征迁移的轴承剩余寿命预测方法](#)

Remaining useful life prediction of bearing based on deep temporal feature transfer

控制与决策. 2021, 36(7): 1699–1706 <https://doi.org/10.13195/j.kzyjc.2019.1809>

[基于聚类簇结构特性的自适应综合采样法在入侵检测中的应用](#)

Toward intrusion detection via cluster structure-based adaptive synthetic sampling approach

控制与决策. 2021, 36(8): 1920–1928 <https://doi.org/10.13195/j.kzyjc.2019.1672>

[嵌入重采样技术的C4.5决策树集成分类算法的临床医学预测](#)

Clinical prediction of C4.5 decision tree classification algorithm with embedded resampling technique

控制与决策. 2021, 36(6): 1342–1350 <https://doi.org/10.13195/j.kzyjc.2019.1247>

基于时序上采样卷积神经网络的风机叶片结冰检测

姜娜, 严蜜, 李柠[†]

(上海交通大学自动化系, 上海 200240)

摘要: 风电机组叶片结冰检测对风电机组的安全性、可靠性以及经济性具有非常重要的现实意义. 针对风电机组运行观测数据的非平衡和单点无时序性问题, 提出一种基于过采样和时序上采样卷积神经网络的风机叶片结冰检测方法. 首先, 采用数据自适应综合过采样算法对原始非平衡数据集进行重采样, 实现对非平衡数据集的均衡; 然后, 提出并构建一种时序上采样卷积神经网络模型, 将原始单点向量型数据进行重构并上采样为二维网格型数据, 同时将其自动映射成为稀疏的特征表示, 实现准确的风机叶片结冰检测功能; 最后, 将所提出方法在真实风场采集的数据集上进行验证. 实验结果表明, 所提出的风机叶片结冰检测方法在数据集非平衡且采集条件有限(单点无时序性数据)的情况下, 具有一定的有效性、稳定性和可行性.

关键词: 叶片结冰检测; 非平衡数据; 向量型数据; 时序上采样; 一维卷积神经网络; 深度学习

中图分类号: TP206.3

文献标志码: A

DOI: 10.13195/j.kzyjc.2020.1657

引用格式: 姜娜, 严蜜, 李柠. 基于时序上采样卷积神经网络的风机叶片结冰检测 [J]. 控制与决策, 2022, 37(8): 2017-2025.

Icing detection of wind turbine blade based on the time-dimensional upsampling convolutional neural network

JIANG Na, YAN Mi, LI Ning[†]

(Department of Automation, Shanghai Jiao Tong University, Shanghai 200240, China)

Abstract: The icing detection of the wind turbine blade has very important practical significance for the safety, reliability and economy of wind turbines. Aimed at the problem of imbalanced and single-time-point non-sequentiality of wind turbine operating observation data, a method is proposed based on the oversampling and the time-dimensional upsampling convolutional neural network model. Firstly, the adaptive synthetic algorithm is applied to original dataset to achieve the balance of the imbalanced dataset. Then, a time-dimensional upsampling convolutional neural network model is proposed and constructed. On one hand, the model can reconstruct and upsample the original single-time-point vector data into the two-dimensional grid data. On the other hand, it can automatically map the data into a sparse feature representation, to achieve an accurate icing detection of the wind turbine blade. Finally, the method is verified on a dataset collected from a real wind farm, and the experimental results show that the proposed icing detection method of the wind turbine blade is effective, stable, and feasible when the dataset is imbalanced and limited.

Keywords: icing detection of blade; imbalanced data; vector data; temporal upsampling; one-dimensional convolutional neural network; deep learning

0 引言

随着化石能源存储减少, 当今社会的环境问题日益严峻, 风能作为清洁无污染、可持续发展的新能源受到了广泛关注. 由于地理因素限制, 我国的风电场主要集中在海拔高、湿度大、温度低的北部沿海地区和山区. 这些地区寒冷潮湿, 冬季时会导致风电机组叶片大量覆冰. 结冰对机组出力和安全运行影响非常大, 不仅对整个风电机组的输出功率造成损失, 还

极易引发叶片失速、机械故障、结冰脱落等安全隐患^[1]. 因此, 研究有效、可靠、经济且实施方便的风电机组叶片结冰检测方法具有非常重要的现实意义.

传统的风电机组结冰检测方法主要分为两大类: 直接检测方法和间接检测方法^[1-2]. 直接检测方法通常通过测量冰在风电机组表面沉积而导致的特性变化检测叶片是否结冰^[2], 例如电容法^[3]、超声波法^[4]、热红外辐射法^[5]、固有频率法^[6]、高光谱成像法^[7]

收稿日期: 2020-11-30; 录用日期: 2021-06-04.

基金项目: 国家自然科学基金项目(61773260).

[†]通讯作者. E-mail: ning_li@sjtu.edu.cn.

等. 间接检测方法主要是基于对导致结冰的天气条件变量或者能够直接反映结冰影响的变量进行采集记录, 并将新的测量值与旧的测量值进行比较, 从而确定风机叶片表面上的积冰量^[2]. 间接检测技术包括实际功率与额定功率之间的测量差异、双风速仪之间的测量差异、环境温度和露点的测量、液态水含量与中值液滴直径的测量差异、噪音检测、叶片共振频率变化检测等方法^[1]. 传统的风机叶片检测方法通常需要外置传感器或结冰检测系统, 不仅增加了风电机组的机械复杂性, 而且需要额外的安装和维护成本. 同时, 传感器的退化和故障将严重损害信号的准确性, 并降低结冰检测系统的可靠性. 因此, 在实际应用中价值较低. 开发具有及时、高可靠性且低成本的结冰检测方法, 开始受到风场和研究人员的关注.

风机数据采集和监控 (supervisory control and data acquisition, SCADA) 系统作为风场的基本组成部分, 能够提供丰富的有关风电机组运行状态的外界环境、电气和机械等信息. 由于 SCADA 系统访问的便捷性和存储数据的大量性, 基于 SCADA 数据的故障诊断方法^[8-13] 成为风电机组故障诊断领域重点研究问题之一, 基于 SCADA 数据的风机叶片结冰检测也成为当今的研究热点. 基于 SCADA 数据的方法通常采用机器学习算法, 将叶片结冰检测问题转化为分类或聚类问题, 从而对风机叶片是否结冰的概率进行预测或者进行异常检测. 目前, 很多机器学习算法应用于基于 SCADA 数据的风机叶片结冰检测中. 如 Zhou 等^[14] 采用支持向量机 (support vector machine, SVM) 算法对风机叶片结冰进行检测; Peng 等^[15] 通过动态主成分分析提取关键特征, 并结合埃尔曼人工神经网络 (elman artificial neural network, EANN) 对风机叶片结冰故障进行预测; Zhang 等^[16] 将机理方法与数据分析相结合, 提取对风机结冰敏感的特征, 基于随机森林 (random forest, RF) 模型对风机叶片结冰进行检测; 邱瑞等^[17] 基于领域知识和包装法对风机 SCADA 数据进行特征选择, 应用极端梯度提升模型 (extreme gradient boosting, XGBoost) 对风机叶片结冰进行早期预测. 为了节省上述方法进行特征选择和提取工作所需的人力与时间, 同时降低对专家经验的需求, 深度学习 (deep learning, DL) 模型^[18-21] 作为一种可以自动提取特征的表示学习方法, 开始广泛应用于基于数据的风机叶片结冰检测中. 如 Chen 等^[18] 考虑风机的多个运行阶段并基于深度神经网络 (deep neural network, DNN) 学习深层特征表示, 以监测风力发电机的健康状况; Ma 等^[19] 根据结冰与风机性能的关系

模型进行特征选择, 提出了一种基于深度置信网络 (deep belief network, DBN) 的叶片结冰预测方法; Liu 等^[20] 采用深度自动编码器 (deep autoencoder, DAE) 网络从 SCADA 数据中自动学习多级故障特征, 并利用集成技术建立集成结冰检测模型; Chen 等^[21] 利用决策树进行特征筛选, 并结合机理模型与卷积神经网络 (convolutional neural networks, CNN) 构建风机叶片结冰预测模型.

然而, 在实际应用中仍然存在一些问题需要解决. 由于存储技术的限制和风机结冰的不确定性, SCADA 系统存储的数据集往往存在类别非平衡、数据缺乏时序性等问题. 非平衡是指由于风机叶片结冰的不确定性, SCADA 数据集内的风机叶片结冰观测数据量远远少于正常运行观测数据量的情况. 数据集的类别非平衡会导致训练模型更加关注风机正常运行数据而忽视结冰状态数据, 从而使模型对于风机结冰的检测漏报率非常高. 同时, SCADA 数据集中的样本为多变量的单点一维向量型数据, 该数据仅记录了风机在结冰检测时刻的状态, 缺乏时序性. 但实际上, 风机结冰的过程是一个随时间变化的过程, 风机在某时刻是否结冰与其当前时刻的状态有关, 且能够从之前时刻的状态信息中反应出来, 因此, 数据时序性的缺失也会降低模型的检测精度.

针对上述两个问题, 本文提出一种基于时序上采样卷积神经网络的风机叶片结冰检测方法. 针对数据类别非平衡问题, 采用自适应综合过采样方法对风机叶片结冰类样本进行过采样, 均衡原非平衡风机数据集. 针对多变量单点一维向量型数据样本, 在时间维度上进行上采样, 将其升维成二维网格型数据, 以扩展风机数据在时间维度上有价值的相关特征信息, 并基于一维卷积神经网络对其局部特征进行自动提取和风机结冰检测. 实验结果表明, 所提出的风机叶片结冰检测方法在数据集非平衡且采集条件有限 (单点无时序性数据) 的情况下, 具有一定的有效性、稳定性和可行性.

1 风机叶片结冰检测方法

1.1 数据清洗和预处理

1.1.1 数据缺失值和异常值处理

在实际的风机 SCADA 系统中, 有时由于信息处理错误、传感器工作异常、人工检修等问题, 系统采集的数据存在缺失、异常等问题, 因此, 所提出方法首先对数据进行缺失值和异常值处理. 对于数据缺失值的情况, 本文所用数据集缺失值样本占总体的比例较小, 因此采用同类均值插补的方法进行填充. 对

于异常值数据,采用 3σ 模型进行检测和处理^[22]. 假设样本点 x_i 有 f_I 个变量,定义 f_I 个变量的均值为 $\bar{x}_i = \frac{1}{f_I} \sum_{d=1}^{f_I} x_{i,d}$,对应的标准差为

$$\sigma_i = \sqrt{\frac{1}{f_I} \sum_{d=1}^{f_I} (x_{i,d} - \bar{x}_i)^2}$$

若 $x_{i,d}$ 的值在 $(\bar{x}_i - \sigma_i, \bar{x}_i + \sigma_i)$ 内,则认为 $x_{i,d}$ 的值为正常值,否则视其为异常值,并基于下式对 $x_{i,d}$ 进行修正:

$$x_{i,d} = \begin{cases} x_{i,d} - \sigma_i, & x_{i,d} < \bar{x}_i - \sigma_i; \\ x_{i,d} + \sigma_i, & x_{i,d} \geq \bar{x}_i + \sigma_i. \end{cases} \quad (1)$$

1.1.2 数据自适应综合过采样

针对风电数据存在的非平衡问题,本文采用自适应综合过采样(adaptive synthetic, ADASYN)^[23]方法对风机叶片结冰故障类(少数类)样本进行过采样,改善风电不平衡数据集,从而得到均衡分布的风电数据集. 基于ADASYN算法对非平衡风电数据集过采样步骤如下.

step 1: 计算非平衡率(imbalanced ratio, IR)为

$$IR = m_s/m_l. \quad (2)$$

其中: m_s 为少数类样本数量, m_l 为多数类样本数量, $IR \in (0, 1)$.

step 2: 计算需要合成的少数类样本数量 G 为

$$G = m_l - m_s. \quad (3)$$

step 3: 基于欧氏距离计算得到每个少数类样本 x_i 的 k 近邻,则其 k 近邻中多数类样本占比 r_i 为

$$r_i = \Delta_i/k, \quad i = 1, 2, \dots, m_s. \quad (4)$$

其中: Δ_i 为少数类样本 x_i 的 k 近邻中属于多数类样本的近邻数量, $r_i \in [0, 1]$.

step 4: 计算每个少数类样本 x_i 所需合成的样本数量占比 \hat{r}_i 为

$$\hat{r}_i = r_i / \sum_{i=1}^{m_s} r_i, \quad (5)$$

故 $\sum_{i=1}^{m_s} \hat{r}_i = 1$.

step 5: 对每个少数类样本 x_i 计算其所需合成的样本数量 g_i 为

$$g_i = \hat{r}_i \times G. \quad (6)$$

step 6: 从每个待合成的少数类样本 x_i 周围的 k 近邻中,随机选择1个少数类样本 x_{zi} ,基于下式合成新样本 x_{si} :

$$x_{si} = x_i + \lambda(x_{zi} - x_i), \quad (7)$$

其中 $\lambda \in [0, 1]$. 重复合成新样本,直至满足step 5中所需的合成数目 g_i 为止.

1.1.3 数据特征标准化

风电数据为多变量数据,各变量的量纲和数值量级具有不一致性. 为了消除各变量间的差异,使其对之后检测模型的训练具有相同的影响程度,从而加快模型权重参数的收敛,需要对数据进行特征标准化. 本文采用标准差标准化(zero-mean normalization, z-score)方法,基于下式对过采样后的数据集进行特征标准化:

$$x'_{i,d} = \frac{x_{i,d} - \mu_d}{\sigma_d}. \quad (8)$$

其中: $x_{i,d}$ 为样本 x_i 变量 d 的取值, $x'_{i,d}$ 为 $x_{i,d}$ 经过特征标准化后的值, μ_d 和 σ_d 分别为变量 d 的均值和标准差.

1.2 基于时间维度上采样卷积神经网络的风机叶片结冰检测模型

针对SCADA数据集类别不均衡、数据缺乏时序性的问题,提出基于时序上采样卷积神经网络(time-dimensional upsampling convolutional neural network, TDUS-CNN)的风机叶片结冰检测模型,扩充风机数据在时间维度上有价值的相关信息并实现自动特征提取和结冰检测的目的. 所提出模型主要由两部分构成:时间维度上采样网络和一维卷积神经网络. TDUS-CNN模型架构如图1所示. 其中:时间维度上采样网络结构由输入层、重构层以及时间维度上采样层组成,一维卷积神经网络结构由卷积层、最大池化层、全局平均池化层以及分类器层组成.

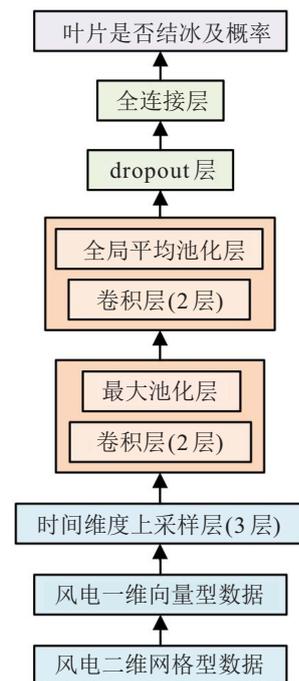


图1 基于TDUS-CNN的风机叶片结冰检测模型架构

1.2.1 基于时间维度的单点风机数据上采样

本文对单点风机数据在时间维度上进行上采样,将风机一维向量型数据升维成二维网格型数据,从

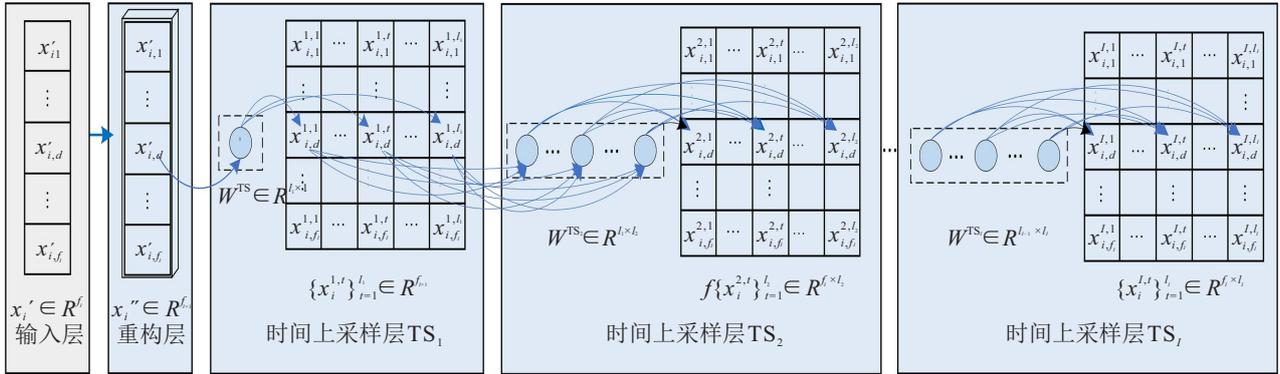


图2 时间维度单点风机数据上采样网络结构

重构层将输入层输入的风机一维向量型数据 $X_i^f \in R^{f_I}$ 重构为二维网格型数据 $X_i'' \in R^{f_I \times 1}$, 即

$$X_i^f = [x_1, x_2, \dots, x_{f_I}], \quad (9)$$

$$X_i'' = [x_1, x_2, \dots, x_{f_I}]^T, \quad (10)$$

其中 f_I 为变量总数. 为描述方便, 后续内容使用 X_i 替代.

对于第 TS_j 层的时间上采样层, 根据下式对每个特征的时间序列进行全连接操作:

$$\{x_{i,d}^{j,t}\}_{t=1}^{l_j} = \{x_{i,d}^{j-1,t}\}_{t=1}^{l_{j-1}} W^j + b^j. \quad (11)$$

其中: l_j 和 l_{j-1} 分别为 $\{x_{i,d}^{j,t}\}_{t=1}^{l_j}$ 和 $\{x_{i,d}^{j-1,t}\}_{t=1}^{l_{j-1}}$ 的序列长度; $W^j \in R^{l_{j-1} \times l_j}$ 和 $b^j \in R^{1 \times l_j}$ 分别为第 TS_j 层上采样层的权重和偏置; 第 TS_{j-1} 层时间上采样层在第

而扩展风机数据在时间维度上有价值的相关特征信息. 该上采样网络结构如图2所示, 主要包括输入层、重构层、时间上采样层.

d 个特征上的输出序列为 $\{x_{i,d}^{j-1,t}\}_{t=1}^{l_{j-1}} \in R^{1 \times l_{j-1}}$, 通过第 TS_j 层上采样得到输出序列 $\{x_{i,d}^{j,t}\}_{t=1}^{l_j} \in R^{1 \times l_j}$.

1.2.2 一维卷积神经网络

一维卷积神经网络 (one-dimensional convolutional neural network, 1D-CNN) 模型具有局部连接、权值共享、平移不变等特性, 这些特性能够使卷积网络在特定局部区域内进行稀疏交互, 缩减网络的参数量, 并在保持原始输入数据结构信息的前提下提取重要的风机叶片结冰特征信息, 从而提高模型的训练效率和准确性. 因此, 本文采用 1D-CNN 模型对升维后的风电网格型数据进行自动特征提取和叶片结冰检测. 1D-CNN 结构如图3所示, 主要包括输入层、卷积层、最大值池化层、全局平均值池化层以及分类器层.

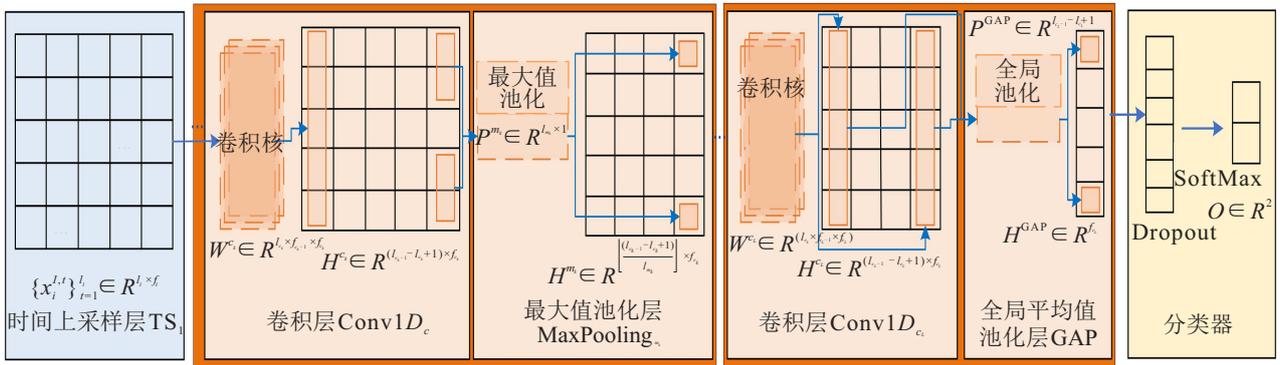


图3 1D-CNN结构

输入层将通过时序上采样处理后的网格型数据 $\{x_i^t\}_{t=1}^{l_I} \in R^{f_I \times l_I}$ 进行转置, 然后输入给后续的卷积层 $Conv1D_1$, 其中 l_I 为时间序列长度.

卷积层通过一系列的卷积核对前一层输入的二维网格型数据进行扫描和运算, 组合得到高维特

征. 第 c_k 层卷积层的前向传播为

$$s_i^{c_k}[j_c] = H^{c_k-1}[j_c : j_c + l_{c_k} - 1] \cdot w_i^{c_k} + b_i^{c_k}, \quad (12)$$

$$h_i^{c_k}[j_c] = \text{ReLU}(s_i^{c_k}[j_c]). \quad (13)$$

其中: H^{c_k-1} 为第 c_k 层卷积层的输入, $H^{c_k-1} \in$

$R^{l_{c_k-1} \times f_{c_k-1}}$; $w_i^{c_k}$ 和 $b_i^{c_k}$ 分别为第 c_k 层卷积层的第 i 个卷积核的权重和偏移量, $w_i^{c_k} \in R^{f_{c_k-1} \times l_{c_k}}$; $H^{c_k-1}[j_c : j_c + l_{c_k} - 1]$ 为 H^{c_k-1} 中第 j_c 到第 $(j_c + l_{c_k} - 1)$ 个行向量组成的矩阵, $H^{c_k-1}[j_c : j_c + l_{c_k} - 1] \in R^{f_{c_k-1} \times l_{c_k}}$; $h_i^{c_k}$ 为第 c_k 层卷积层的第 i 个卷积核经过卷积和 ReLU 处理后得到的特征向量; $h_i^{c_k}[j_c]$ 为 $h_i^{c_k}$ 的第 j_c 个元素, 且 $j_c \in [1, 2, \dots, l_{c_k-1} - l_{c_k} + 1]$.

最大值池化层基于下式, 通过最大化值池化操作提取最显著的特征:

$$p_i^{m_k}[j_p] = \max(h_i^{c_k}[j_p l_{m_k} - l_{m_k} + 1 : j_p l_{m_k}]), \quad (14)$$

$$l_{m_k-1} = l_{c_k-1} - l_{c_k} + 1. \quad (15)$$

其中: $h_i^{c_k}[j_p l_{m_k} - l_{m_k} + 1 : j_p l_{m_k}]$ ($j_p \in [1, 2, \dots, \lfloor l_{m_k-1} / h_{p_{k+1}} \rfloor$) 为 $h_i^{c_k}$ 的第 $j_p l_{m_k} - l_{m_k} + 1$ 到第 $j_p l_{m_k}$ 个元素组成的列向量; l_{m_k} 为最大值池化窗口的高; l_{m_k-1} 为 $h_i^{c_k}$ 的高; $p_i^{m_k}$ 为第 m_k 层最大值池化层第 i 个特征的池化向量, $p_i^{m_k} \in R^{\lfloor l_{m_k-1} / l_{m_k} \rfloor}$; $p_i^{m_k}[j_p]$ 为 $p_i^{m_k}$ 的第 j_p 个元素.

全局平均池化层对每个特征序列求平均值, 将二维网格数据转化为一维向量数据 $H^{GAP} \in R^{f_{c_L}, f_{c_L}}$ 为最后一层卷积层的卷积核数目.

分类器层包括 Dropout 层和 SoftMax 回归层. Dropout 层随机为网络中的神经元赋值零权重, 减小网络对数据微小变化响应的敏感度, 提高对不可见数据处理的泛化能力. SoftMax 层通过 SoftMax 回归操作输出该样本是否发生叶片结冰故障的概率.

1.3 风机叶片结冰检测算法流程

综上所述, 所提出的基于时序上采样卷积神经网络的风机叶片结冰检测算法流程分为离线和在线两部分.

离线部分主要基于历史数据构建和训练风机叶片结冰检测 TDUS-CNN 模型. 首先, 根据均值插补和 3σ 原则对原始训练集进行缺失值和异常值检测处理; 然后, 根据式 (2) 判断数据集类别分布是否平衡, 若非平衡则采用 ADASYN 算法对少数类别样本进行过采样, 均衡原非平衡数据集; 接着, 基于平衡的数据集进行特征标准化处理, 并保存该特征标准化模型; 最后, 构建所提出的风机结冰检测 TDUS-CNN 模型, 并基于特征标准化后的数据集对其进行训练, 保存性能最优的模型.

在线部分主要基于在线的数据和离线构建的模型风机叶片是否结冰进行检测. 首先, 对在线数据进行缺失值和异常值检测处理; 然后, 基于离线部分保存的特征标准化模型对数据进行归一化处理; 最后, 将预处理后的数据输入给离线保存的风机结冰检测

模型, 以便对该时刻风机叶片是否结冰进行检测.

2 实验及结果分析

2.1 数据描述

为了验证所提出的基于时序上采样卷积神经网络的风机叶片结冰检测方法, 本文采用 WAIC 2020 工业互联网算法大赛提供的云南某高原风场的结冰标记数据集进行实验. 如表 1 所示, 该数据集主要包括 38 个风机状态监测变量以及样本是否结冰的标签.

表 1 WAIC 2020 风电机组叶片结冰检测数据集变量

序号	符号	描述
1	CabinInsideTemp	机舱控制柜内温度
2	CabinOutTemp	机舱控制柜外温度
3	ConverCoolWaterTemp	变流器冷却水温度
4	GBCoolWaterTemp	齿轮箱冷却水温度
5	GearBoxOilTemp	齿轮箱油温
6	GebBTempB	发电机轴承温度 B
7	GenAirTempIn	发电机冷却进口温度
8	GenAirTempOut	发电机冷却出口温度
9	GenBTempA	发电机轴承温度 A
10	GenStatorUT	发电机 U 相定子温度
11	GenStatorVT	发电机 V 相定子温度
12	GenStatorWT	发电机 W 相定子温度
13	GeneratorSpeed	发电机转速
14	GridCospi	电网功率因数
15	GridFQ	电网频率
16	GridIA	电网 A 相电流
17	GridIB	电网 B 相电流
18	GridIC	电网 C 相电流
19	GridPower	机组有功功率
20	GridRePower	机组无功功率
21	GridULA	电网 A 相电压
22	GridULB	电网 B 相电压
23	GridULC	电网 C 相电压
24	HSpdShaftTemp	高速轴温度
25	HydraulicSystemPressure	液压系统压力
26	LSpdShaftTemp	低速轴温度
27	NaceVibEffec	机舱振动
28	NaceVibX	机舱 X 方向振动
29	NaceVibY	机舱 Y 方向振动
30	NacelleCabinetTemp	机舱柜温度
31	PitchAngle	变桨角度
32	PitchAngle1	叶片 1 变桨角度
33	PitchAngle2	叶片 2 变桨角度
34	PitchAngle3	叶片 3 变桨角度
35	RotorSpeed	风轮转速
36	TowerBaseCabinetTemp	风电机组塔底控制柜温度
37	TwistPos	偏航扭缆角度
38	YawPosition	偏航位置
39	label	0 为不结冰, 1 为结冰

该数据集的组成和样本类别分布信息如表2所示,主要包括训练集和测试集两部分:训练集为多台风电机组2017年12月~2018年3月期间的历史运行数据,样本总数为9650,其中结冰样本个数为657,样本类别非平衡程度为6.8%;测试集为多台风电机组2018年12月~2019年3月期间的历史运行数据,样本总数为3369,其中结冰样本个数为226,样本类别非平衡程度为6.7%。

表2 WAIC 2020 风电机组叶片结冰检测数据集信息

数据集	样本总数	正样本数	负样本数	非平衡程度/%
训练集	9650	657	8993	6.8
测试集	3369	226	3143	6.7

2.2 评价指标

本文采用召回率 R 和 AUC 对模型作为评估指标. 假设 TP、FP、TN、FN 分别为检测正确的正样本个数、检测错误的负样本个数、检测正确的负样本个数以及检测错误的负样本个数,则召回率 R 为检测正确的结冰状态次数占实际结冰状态次数的比例

$$R = TPR = \frac{TP}{TP + FN} \quad (16)$$

AUC 为 ROC 曲线下的面积,ROC 是以负正类率 (false positive rate, FPR) 特异度为横坐标,召回率为纵坐标的曲线. 负正类率为

$$FPR = \frac{FP}{FP + TN} \quad (17)$$

在风机叶片结冰检测模型训练阶段,本文采用五折交叉验证对模型进行评估,然后选择评估结果最优的模型在测试集上进行验证.

2.3 实验过程和结果分析

将训练集的非平衡数据在数据过采样阶段进行均衡,得到分布平衡的训练集,其中正样本数为8993,负样本数为8893,样本总数为17786,重采样后样本均值由 2.3091×10^{-16} 变为 0.1972,方差由 1.0000 变为 1.3500. 原始数据、特征标准化后的数据以及通过 ADASYN 算法过采样和特征标准化后的数据分布如图4所示. 其中:图4(a)、图4(c)和图4(e)为选取特征16和特征19所得的二维散点图,图4(b)、图4(d)和图4(f)为通过 TSNE 降维可视化方法得到的二维散点图. 由图4(b)和图4(d)可见,通过特征标准化后的数据被缩放到一定范围内,不同类数据相较于原数据的可分性也被增强. 此外,由图4可见,原始数据不同类别的样本分布是极度非平衡的,结冰状态的数据明显少于正常状态的数据. 通过 ADASYN 过采样后,结冰状态的数据样本和正常状态下的数据样本分布基本得到了均衡.

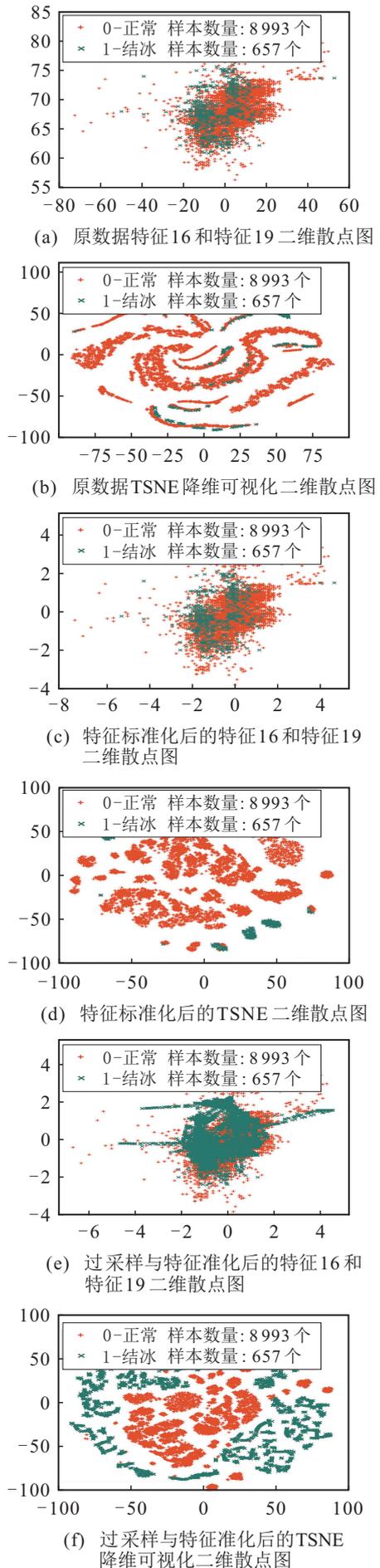


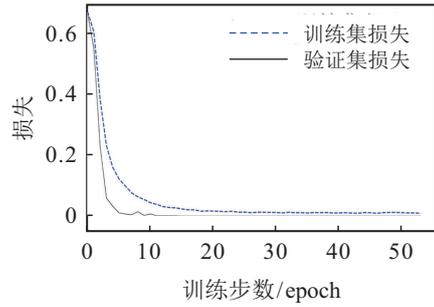
图4 特征二维散点图

基于均衡的数据集对所提出的风机叶片结冰检测模型进行训练. 模型的训练过程主要包括超参数的选择和模型参数的训练两部分. 模型的超参数主要包括: 时间上采样层 TSUP 的层数及其相应层神经元的个数、卷积层的层数及其相应卷积核的个数、卷积核的高度、最大值池化层的层数和窗口的高度、全连接层神经元的个数. 本文基于网格搜索法对模型结构超参数进行确定, 最终得到较优的网络结构的超参数如表 3 所示.

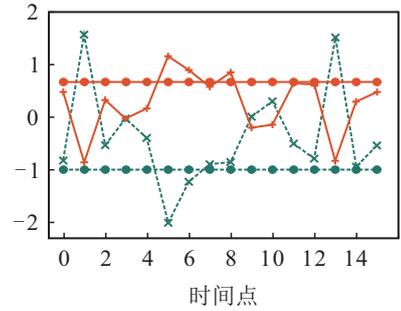
表 3 TDUS-CNN 风机叶片结冰检测模型结构参数

超参数类别	参数
输入层大小	38
重构层输出大小	1 × 38
时间上采样层层数	3
时间上采样层 TSUP_1 神经元数	4
时间上采样层 TSUP_2 神经元数	8
时间上采样层 TSUP_3 神经元数	16
卷积层层数	6
卷积层高度	4
卷积层 Conv1D_1 卷积核数, 卷积层 Conv1D_2 卷积核数	32, 32
卷积层 Conv1D_3 卷积核数, 卷积层 Conv1D_4 卷积核数	16, 16
卷积层 Conv1D_5 卷积核数, 卷积层 Conv1D_6 卷积核数	8, 8
最大值池化层层数	2
最大值池化层窗口高度	2
全局平均池化层输出大小	8
全连接层神经元数	8
分类器神经元数	2

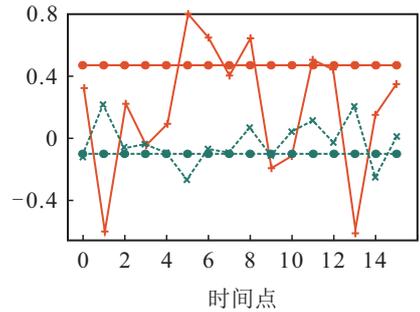
对于模型参数的训练, 本文以交叉熵为损失函数, 采用 Adam 优化算法对其进行训练, 从而得到较优的模型参数, 训练过程中的训练 loss 和验证 loss 收敛曲线如图 5(a) 所示. 训练好后的模型对原始数据上采样后的部分特征曲线如图 5(b) 和图 5(c) 所示. 由图 5(a) ~ 图 5(c) 可见, 不同类别的特征曲线趋势是不同的, 相同类别特征之间的曲线趋势有一定的相似性. 同时, 上采样的过程在原数据的基础上增加了一些波动, 从而在数值和趋势两方面扩充了网络可以学习的信息, 用于后续的特征提取和风机叶片结冰检测. 由 TDUS-CNN 模型自动提取的特征具有一定的稀疏性, 正常的数据通过 TDUS-CNN 模型所提取的特征 3 和特征 5 全部被置零, 相反, 结冰的数在除特征 3 和特征 5 之外的其他特征被全部置 0. 选取特征 3 和特征 6 对数据样本进行二维散点可视化, 得到自动映射成为稀疏的特征表示结果如图 5(d) 所示, 可见, 两种类别的数据通过 TDUS-CNN 模型后已经能够很好地区分. 经过测试, 所提出基于 TDUS-CNN 的风机结冰检测模型的召回率 R 和 AUC 分别为 0.9469 和 0.9639.



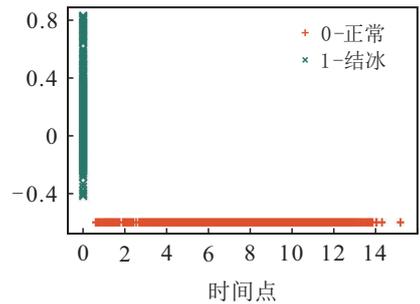
(a) 较优模型训练过程损失变化



(b) 上采样后特征 1 曲线图



(c) 上采样后特征 2 曲线图



(d) TDUS-CNN 模型提取特征二维散点图

图 5 实验结果

为了验证所提出模型的性能, 将其与 SVM 模型、最邻近节点 (k -nearest neighbor, KNN) 模型、XGBoost 模型、多层感知机 (multi-layer perception, MLP) 模型和堆栈自编码器 (stacked autoencoder, SAE) 进行对比. 分别采用原始训练集和过采样后的平衡训练集对 6 种模型通过五折交叉验证的方法进行训练与评估, 得到基于 6 种方法性能较优的训练模型, 并对其

在相同的测试集上进一步进行对比评估。

各模型基于原始不平衡数据集和均衡过后的平衡数据集进行训练得到的实验结果如表4和表5所示。评估指标主要包括五折交叉验证训练得到的AUC均值和标准差、较优模型在测试集上进行测试所得的AUC和 R 。由表4和表5可见,相较于其他模型,所提出基于TDUS-CNN风机叶片结冰检测模型取得的训练AUC均值、测试AUC和 R 最高,性能最好。基于原始不平衡数据集和均衡过后的平衡数据集训练得到的模型性能,从训练AUC的均值看,TDUS-CNN模型比其他模型平均高出1.05%和1.87%;从测试AUC看,TDUS-CNN模型比其他模型平均高出11.84%和17.56%;从测试 R 而言,TDUS-CNN模型比其他模型平均高出31.68%和17.56%;从训练AUC的标准差而言,TDUS-CNN模型的标准差也相对较小,表现较为稳定,鲁棒性较高。此外,基于均衡过后的平衡数据集训练所得到的模型性能,显然优于基于原始不平衡的数据集训练所得到的模型性能。

表4 与其他算法对比分析(原始不平衡数据集)

模型	训练集 AUC	测试集 AUC	测试集 R
SVM	0.998 5±0.003 0	0.791 7	0.584 1
KNN	0.989 8±0.008 6	0.756 2	0.513 3
XGBoost	0.963 1±0.072 4	0.786 2	0.575 2
MLP	0.996 1±0.003 3	0.891 0	0.818 6
SAE	0.997 6±0.002 0	0.867 2	0.747 8
TDUS-CNN	0.999 2±0.001 4	0.911 8	0.827 4

表5 与其他算法对比分析(均衡过后的平衡数据集)

模型	训练集 AUC	测试集 AUC	测试集 R
SVM	0.999 9±0.000 1	0.870 4	0.743 4
KNN	0.998 4±0.000 7	0.926 2	0.867 3
XGBoost	0.975 8±0.009 4	0.831 7	0.672 6
MLP	0.960 0±0.011 0	0.911 2	0.876 1
SAE	0.974 6±0.076 1	0.944 6	0.924 8
TDUS-CNN	0.999 9±0.000 1	0.963 9	0.946 9

实验结果表明,相较于其他模型,所提出的TDUS-CNN风机叶片结冰检测模型,由于采用上采样方法从数值和趋势两方面补充了网络可以学习的信息,同时基于该信息采用1D-CNN模型进行局部特征提取,将稠密的原始数据特征映射成为稀疏的特征,从而降低了后续模型的复杂度,提高了模型的检测精度。此外,由于在数据预处理部分采用ADASYN算法对原始非平衡的数据集进行均衡,克服了由于数据类别非平衡给模型带来的少数类别分类精度低的问题。

3 结论

本文针对风电数据非平衡和单点无时序性问题,提出一种基于过采样和时序上采样卷积神经网络的方法,用于数据驱动的风机叶片结冰检测。首先,采用ADASYN算法对原始非平衡的数据集进行过采样,人工合成少数类样本,对原数据集进行均衡;其次,构建并训练TDUS-CNN模型,一方面对原始单点一维向量型数据进行重构升维,并在时序维度进行上采样,以扩展风机数据在时间维度上有价值的相关特征信息,另一方面采用1D-CNN模型对上采样后的数据自动进行局部特征提取,将原始稠密的数据特征映射成为稀疏的特征表示。实验结果表明,所提出的风机叶片结冰检测方法在数据集非平衡且采集条件有限(单点无时序性数据)的情况下优于其他方法,能够表现出有效且稳定的性能,较为准确地识别出风机叶片的结冰状态。

在后续研究中,将把时间戳属性作为一个特征,构建具有时序性的非定长时间序列数据,并用于风机叶片结冰提前预测的研究中。

参考文献(References)

- [1] Wei K X, Yang Y, Zuo H Y, et al. A review on ice detection technology and ice elimination technology for wind turbine[J]. Wind Energy, 2020, 23(3): 433-457.
- [2] Madi E, Pope K, Huang W M, et al. A review of integrating ice detection and mitigation for wind turbine blades[J]. Renewable and Sustainable Energy Reviews, 2019, 103: 269-281.
- [3] Gerardi J J, Dahl P R, Hickman G A. Piezoelectric sensor[P]. United States: 5191791, 1993-03-09.
- [4] Zhang Z H, Zhou W S, Li H. Icing estimation on wind turbine blade by the interface temperature using distributed fiber optic sensors[J]. Structural Control and Health Monitoring, 2020, 27(6): e2534.
- [5] Gómez M C Q, García M F P, Sánchez T J M. Ice detection using thermal infrared radiometry on wind turbine blades[J]. Measurement, 2016, 93: 157-163.
- [6] Gantasala S, Luneno J C, Aidanpaa J O. Detection of ice mass based on the natural frequencies of wind turbine blade[J]. Wind Energy Science Discussions, 2016: 1-17.
- [7] Rizk P, Al Saleh N, Younes R, et al. Hyperspectral imaging applied for the detection of wind turbine blade damage and icing[J]. Remote Sensing Applications: Society and Environment, 2020, 18: 100291.
- [8] 任浩, 屈剑锋, 柴毅, 等. 深度学习在故障诊断领域中的研究现状与挑战[J]. 控制与决策, 2017, 32(8): 1345-1358.
(Ren H, Qu J F, Chai Y, et al. Deep learning for fault

- diagnosis: The state of the art and challenge[J]. Control and Decision, 2017, 32(8): 1345-1358.)
- [9] 张静, 李柠, 李少远, 等. 基于数据的风电机组发电机健康状况评估[J]. 信息与控制, 2018, 47(6): 694-701.
(Zhang J, Li N, Li S Y, et al. Health assessment of wind-turbine generator based on data[J]. Information and Control, 2018, 47(6): 694-701.)
- [10] 曲建岭, 余路, 袁涛, 等. 基于卷积神经网络的层级化智能故障诊断算法[J]. 控制与决策, 2019, 34(12): 2619-2626.
(Qu J L, Yu L, Yuan T, et al. A hierarchical intelligent fault diagnosis algorithm based on convolutional neural network[J]. Control and Decision, 2019, 34(12): 2619-2626.)
- [11] 王村松, 陆宁云, 程月华, 等. 基于无标签、不均衡、初值不确定数据的设备健康评估方法[J]. 控制与决策, 2020, 35(11): 2687-2695.
(Wang C S, Lu N Y, Cheng Y H, et al. Equipment health risk assessment based on unlabeled, unbalanced data under uncertain initial condition[J]. Control and Decision, 2020, 35(11): 2687-2695.)
- [12] Jiang N, Hu X Z, Li N. Graphical temporal semi-supervised deep learning-based principal fault localization in wind turbine systems[J]. Proceedings of the Institution of Mechanical Engineers — Part I: Journal of Systems and Control Engineering, 2020, 234(1): 095965181990103.
- [13] Jiang N, Li N. A wind turbine frequent principal fault detection and localization approach with imbalanced data using an improved synthetic oversampling technique[J]. International Journal of Electrical Power & Energy Systems, 2021, 126: 106595.
- [14] Zhou G F, Tan W, Zhang D. Ice detection for wind turbine blades based on PSO-SVM method[J]. Journal of Physics: Conference Series, 2018, 1087: 022036.
- [15] Peng C, Tang Z, Chen Q, et al. Short-term prediction of generator blade ice fault based on multi-AN[C]. The 7th International Conference on Advanced Cloud and Big Data. Suzhou, 2019: 291-296.
- [16] Zhang L, Liu K, Wang Y, et al. Ice detection model of wind turbine blades based on random forest classifier[J]. Energies, 2018, 11(10): 2548.
- [17] 邸瑞, 陈正鸣, 吕嘉. 基于SMOTE的XGBoost算法在风机叶片结冰预测中的应用[J]. 信息技术, 2019, 43(12): 81-85.
(Di R, Chen Z M, Lyu J. Application of XGBoost based on SMOTE in prediction of wind turbine blades icing[J]. Information Technology, 2019, 43(12): 81-85.)
- [18] Chen L T, Xu G H, Liang L, et al. Learning deep representation for blades icing fault detection of wind turbines[C]. IEEE International Conference on Prognostics and Health Management. Seattle, 2018: 1-8.
- [19] Ma J Q, Ma L X, Tian X C. Wind turbine blade icing detection based on deep belief network[C]. The 4th International Conference on Mechanical, Control and Computer Engineering. Hohhot, 2019: 226-263.
- [20] Liu Y, Cheng H, Kong X G, et al. Intelligent wind turbine blade icing detection using supervisory control and data acquisition data and ensemble deep learning[J]. Energy Science & Engineering, 2019, 7(6): 2633-2645.
- [21] Chen X X, Lei D, Xu G Q. Prediction of icing fault of wind turbine blades based on deep learning[C]. IEEE 2nd International Conference on Automation, Electronics and Electrical Engineering. Shenyang, 2019: 295-299.
- [22] 常永亮. 基于现场数据的大型风电机组故障诊断研究[D]. 北京: 华北电力大学, 2016.
(Chang Y L. Fault diagnosis research on wind turbine based on the running data in fxiel[D]. Beijing: North China Electric Power University, 2016.)
- [23] He H B, Bai Y, Gracia E A, et al. ADASYN: Adaptive synthetic sampling approach for imbalanced learning[C]. IEEE International Joint Conference on Neural Networks. Hong Kong, 2008: 1322-1328.

作者简介

姜娜(1991—), 女, 博士生, 从事基于数据的风机系统健康运维的研究, E-mail: firecy@sjtu.edu.cn;

严蜜(1995—), 女, 博士生, 从事基于度量学习的不平衡分类任务、工业流程的故障诊断的研究, E-mail: yanmi1995@sjtu.edu.cn;

李柠(1974—), 女, 教授, 博士生导师, 从事分布式优化与控制、工业大数据分析、故障预测与智能运维、自主系统主动感知与智能决策等研究, E-mail: ning_li@sjtu.edu.cn.

(责任编辑: 魏冰)