

基于动态卷积的混合模型性能评估方法

迟 骋, 李向阳, 张志利*, 汪 潇

(火箭军工程大学, 陕西 西安 710025)

摘 要: 围绕大型复杂设备的安全、准确的性能评估急需, 针对现有基于机器学习的性能评估方法难以捕捉时序数据的动态特性, 且难以同时处理非线性和线性数据问题, 提出了一种基于动态卷积的混合模型性能评估方法。该方法首先将动态数据输入动态卷积模块, 从而更有效地进行特征提取, 并保留数据的动态特性。然后, 利用门控 Transformer 能够捕捉更复杂的长期依赖关系, 而自回归模型能够更好地捕捉数据线性特征的特点, 采用门控 Transformer 和自回归模型自适应加权的方法, 以更好地适应不同使用场景和领域的的数据。在 C-MAPSS 数据集的验证结果表明: 预测误差与模型效率均有一定改进, 尤其在 FD002 数据集中, 对于均方根误差和 Score 评分函数, 该方法比其他效果最好的方法分别提升了 5.92、116。

关键词: 动态卷积; 门控 Transformer; 自回归模型; 混合模型; 非线性数据; 线性数据

中图分类号: TJ76

文献标志码: A

文章编号: 2097-2741(2024)05-053-07

Performance Evaluation Method of Hybrid Model Based on Dynamic Convolution

CHI Cheng, LI Xiangyang, ZHANG Zhili*, WANG Xiao

(Rocket Force University of Engineering, Xi'an 710025, Shaanxi)

ABSTRACT: There is an urgent need for the safe and accurate performance evaluation of large and complex equipment. Considering the incapacity of existing performance evaluation methods based on machine learning in capturing the dynamic characteristics of time series data and handling both nonlinear and linear data simultaneously, a hybrid model performance evaluation method based on dynamic convolution was proposed. Firstly, the proposed method input dynamic data into a dynamic convolution module for more effective feature extraction, while preserving the dynamic characteristics of the data. Then, the adaptive weighting method of gated transformer and autoregressive model is adopted to better suit data from different usage scenarios and fields. For the gated transformer can capture more complex long-term dependencies, and the autoregressive model can better capture the linear characteristics of the data. The proposed method was verified in the C-MAPSS dataset, the results showed that the prediction error and model efficiency were both improved to a certain extent. In particular, in the FD002 dataset, the proposed method improved the root mean square error and Score scoring function by 5.29 and 116 respectively compared to other methods with the best performance.

收稿日期: 2024-07-03

第一作者: 迟骋 (1996—), 男, 硕士研究生, 主要从事复杂设备性能评估方法研究。E-mail: azir928@foxmail.com

***通信作者:** 张志利 (1966—), 男, 教授, 主要从事智能感知与导航技术研究。E-mail: 157918018@qq.com

引用格式: 迟骋, 李向阳, 张志利, 等. 基于动态卷积的混合模型性能评估方法[J]. 火箭军工程大学学报, 2024, 38 (5): 53-58, 80.

KEYWORDS: dynamic convolution; gated transformer; autoregressive model; hybrid model; nonlinear data; linear data

近年来,随着现代工业技术的发展,大型复杂设备的性能评估已成为一项关键技术。设备性能评估通过在设备运行过程中不同时间段提供性能状态值^[1],帮助设备管理人员分析和判断设备的当前运行状态。在此基础上,可以预测设备的性能状态,确定早期故障,预测性能状态的趋势^[2]以及设备的剩余寿命^[3]。这些关于设备退化过程的信息能够预测潜在的性能下降和故障发生,从而指导用户及时做出维修决策^[4],避免故障和经济损失,减少设备停机时间和维修成本,最终实现设备可靠性和经济效益的平衡。因此,近年来人们对设备性能评估的关注度不断上升^[5]。

作为数据驱动方法的典型代表,机器学习在设备性能评估领域尤为突出^[6],已被广泛应用于多个领域的设备性能评估^[7],许多深度学习模型已被用于机械设备的性能评估和预测。Wang 等^[8]提出了一种新的深度预测框架——多尺度卷积注意力网络 (Multiscale Convolutional Attention Network, MSCAN),用于机器的性能评估。Jiang 等^[9]提出了一种基于双注意力的精确阶段划分卷积神经网络用于滚动轴承性能预测,其中包括两部分,即首次预测时间的确定和性能的估计。Qin 等^[10]提出了一种新的退化趋势约束的变分自编码器来构建具有明显退化趋势的性能指标向量。该方法首先构建轴承寿命周期数据集和齿轮疲劳数据集的性能指标向量,然后利用基于宏观-微观注意力的长短期记忆 (Macroscopic-microscopic-attention-based Long Short Term Memory, MMAL-STL) 方法对所构建的性能指标进行预测。然而,基于机器学习的性能评估方法面临以下几个挑战:在航空航天、海洋水资源等高度复杂的领域,设备具有独特性、数量少、复杂性高的特点,数据的多样性会增大数据特征提取的难度;Transformer 模型计算复杂度高、序列长度限制、并行化处理有局限性和对大规模数据的需求。为了降低多传感器数据的维度并检测异常,通常采用两步方法,其缺陷之一是两个基线模型的联合性能容易陷入局部最优,因为两个模型是分别训练的。

针对上述问题,本文提出了一种基于动态卷积的混合模型性能评估方法。首先将动态样本输

入动态卷积网络,然后将其分别输入自回归模型和非线性存储网络。所提出的方法对线性模型和非线性模型分别赋予自适应权重,并将两个网络的输出输入到单层感知机。

本文的创新点如下。

1) 动态模块通过动态生成适应每个动态数据的卷积核,避免了动态数据差异过大导致的特征冗余,从而更有效地进行特征提取,并能保留数据的动态特性。

2) 该方法引入了门控 Transformer 网络 (Gated Transformer Networks, GTN),通过门控机制增强模型能力的变体,结合了长短期记忆网络 (Long Short-Term Memory, LSTM) 和门控循环单元 (Gated Recurrent Unit, GRU) 的优势,能够有效捕捉长距离依赖并控制信息的遗忘与更新。相比传统 Transformer,GTN 在多变量时间序列分类任务中表现更好,特别适合处理时间序列数据。

3) 通过结合线性预测模型 (自回归模型) 和非线性预测模型,所提出模型可以更好地适应不同的使用场景和领域。例如,自回归模型在短期建模中表现良好,而门控 Transformer 能够捕捉更复杂的长期依赖关系。

1 本文方法

本文方法的架构见图 1,为了避免动态数据差异过大导致的特征冗余,首先将动态样本输入动态卷积网络,然后分别输入到自回归模型和非线性存储网络。线性模型和非线性模型分别被赋予自适应权重,将两个网络输入单层感知机。

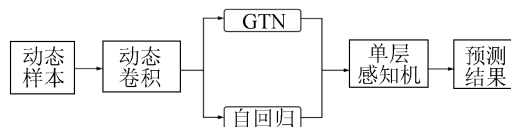


图 1 本文评估方法总体架构

Fig. 1 Overall architecture of evaluation method proposed in this article

1.1 动态卷积

在设计动态卷积时,要设计一个能够有效处理时序数据尺度变化的模块。因此,生成具有不

同膨胀率的动态卷积核, 可以采用重新缩放特征^[11]或几个卷积核的加权等方法, 这比文献^[12]更适合处理数据的尺度问题。

动态卷积网络的结构如图2所示, 动态卷积核 (Dynamic Convolution Kernel, DCK) 模块有两个分支用于输入。下支路将输入的通道减少 m 个, 通道缩减由 1×1 卷积层 f_{θ_1} (即图2中的 “conv1”) 操作, θ_1 为其参数。上分支的输出为 $F_i = f_{\theta_1}(X_i) \in \mathbb{R}^{h \times w \times c}$ 。其中 w 为数据的宽度, h 为数据的高度。

下分支为内核生成器分支。核生成器 f_{θ_2} 生成 3×3 的不同扩张率的卷积核, 生成的核记为 $f_{\theta_2}(X_i)$ 。使用不同的膨胀率来捕获不同尺度的对象, 因为不同域的对象通常具有不同的尺度 (见图2)。

$$\hat{O}_i = f_{\theta_2}(X_i) \otimes F_i = f_{\theta_2}(X_i) \otimes f_{\theta_1}(X_i)$$

式中: 符号 $\otimes$ 指深度卷积; $\hat{O}_i$ 为输出特征。注意到动态卷积核 $f_{\theta_2}(X_i)$ 、 $f_{\theta_1}(X_i)$ 依赖于输入实例 X_i , 因此有适应样本的多样性。

在实现中, 一个内核生成器包括1个全局平均池化层、1个通道缩减层和1个内核生成层。平均池化层通过压缩空间维度来整合空间上下文。核生成层, 即图2中的 “conv2”, 增加了特征图1的元素, 其输出被重新整形为生成的核。

1.2 门控Transformer网络

门控Transformer网络 (Gated Transformer Networks, GTN) 用于多变量时间序列分类。GTN架构包括3部分。

1) 嵌入层 (Embedding)

将原始Transformer中的词嵌入层替换为全连接层, 并添加 $\tanh$ 激活函数。并且添加位置编码 (Positional Encoding), 以利用时间步之间的顺序相关性。

2) 双塔Transformer (Two-tower Transformer)

时间步编码器 (Step-wise Encoder): 使用带掩码的自注意力机制, 计算时间步之间的注意力权重。通过多头自注意力层和位置全连接前馈层进行特征提取, 保持残差连接和层归一化。

通道编码器 (Channel-wise Encoder): 计算不同通道之间的注意力权重, 不添加位置编码, 通过掩码机制捕捉通道之间的相关性。

3) 门控机制 (Gating)

合并时间步编码器和通道编码器的输出。每个编码器输出后添加全连接层和非线性激活层, 得到向量 C 和 S 。将 C 和 S 拼接, 通过线性投影层得到 h , 经过 softmax 函数计算门控权重 g_1 和 g_2 。每个门控权重分别与对应塔的输出相乘, 得到最终合并的特征向量。

这种架构有效地结合了时间步间和通道间的相关性, 提高了多变量时间序列分类的性能。具体架构如图3所示。

1.3 线性回归

自回归 (Autoregressive, AR) 模型是一种线性模型, 它利用观测值与若干滞后观测值之间的依赖关系进行预测。理论上, 非线性递归网络比AR模型更具表现力, 功能更强大。然而, 在特定的实际数据集中, 这种方法并不总是有效。事实上, AR模型在短期预测建模中也能产生良好的效果。因此, 本文将AR模型与非线性记忆网络部分并行结合。

AR模型的公式为

$$\hat{z}_h = c \sum_{i=1}^{h-1} \lambda_{h-i} + \sum_{i=1}^{h-1} \lambda_{h-i} * z_{h-i}$$

式中: $\lambda_1, \lambda_2, \dots, \lambda_{h-1}$ 为AR模型的权重; c 为常数; $\hat{z}_h$ 为过去时间值 $[z_1, z_2, \dots, z_{h-1}]$ 的预测值。

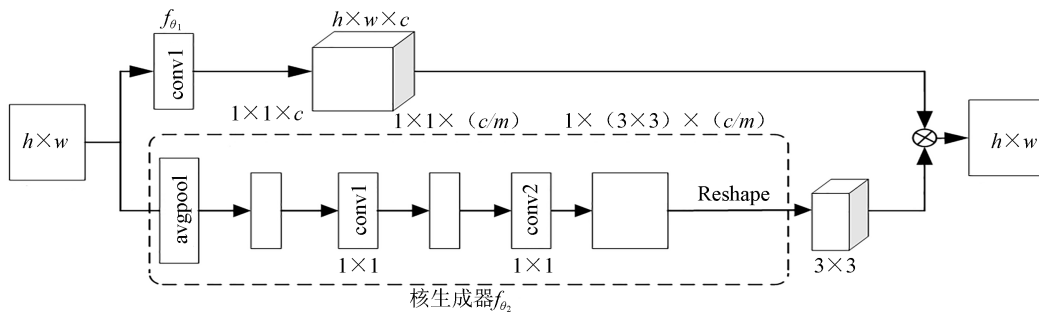


图2 动态卷积网络结构

Fig. 2 Structure of dynamic convolution networks

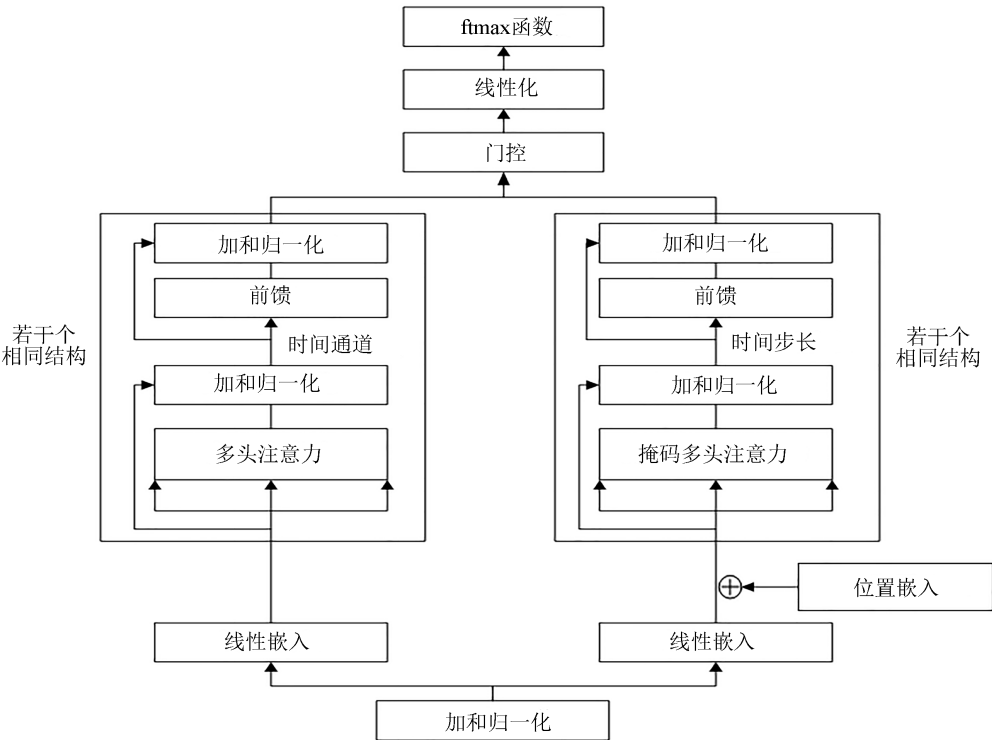


图 3 门控Transformer网络
Fig. 3 Gated transformer network

最后，采用单层感知机来对GTN和线性回归加权求和。在输出层，通过计算预测模型输出与真实值 z_h 之间的差异来获得预测误差。

2 实验与结果分析

2.1 C-MAPSS 数据集

C-MAPSS是目前最流行的性能预测模拟数据集，由美国航空航天局（National Aeronautics and Space Administration, NASA）卓越故障预测研究中心提供，用于飞机涡轮引擎的性能评估。数据来源于在MATLAB和Simulink环境下构建的涡扇发动机引擎退化仿真模型。通过模拟低压涡轮、高压涡轮、低压压气机、高压压气机和风扇这5个主要的发动机组件的退化过程，获取发动机在不同工况和故障模式下，由不同初始条件的引擎从正常运行直到发生故障的全过程数据。

本文使用C-MAPSS数据集对所提出的半监督模型进行评估。该数据集包括FD001、FD002、FD003和FD004子数据集。每个子数据集都有1个训练集和1个测试集。这些子数据集由引擎编号、运行周期、操作条件和21个传感器测量数据组成。数据集的详细信息见表1。

表 1 C-MAPSS 数据集信息
Tab. 1 C-MAPSS data set information

子数据集	操作条件	故障模式	训练集周期数	测试集周期数
FD001	1	1	100	100
FD002	6	1	260	259
FD003	1	2	100	100
FD004	6	2	249	248

2.2 数据预处理

2.2.1 数据标准化

由于不同传感器之间存在差异，所获得的监测数据具有不同的量纲。通过对数据进行标准化处理，可以减少因量纲不同带来的影响，提高模型对性能评估的精确度。本文选用z-score标准化方法，使处理后的数据符合均值为0、标准差为1的标准正态分布。其具体计算公式为

$$x_i' = \frac{x_i - \mu_i}{\sigma_i}$$

式中： x_i' 为标准化后的数据； x_i 为要标准化的数据； μ_i 为对应数据集中第*i*个特征的均值； σ_i 为对应的标准差。

2.2.2 指数平滑

在性能评估预测时，传感器获取的数据通常

包含大量噪声, 这些噪声可能会影响最终的预测结果, 因此, 需要对数据进行降噪处理。常用的降噪方法包括滑动平均法、加权滑动平均法和指数平滑法等。滑动平均法对所有传感器监测数据赋予相同的权重, 但在实际应用中, 不同传感器的数据重要性各不相同, 因此需要对其进行加权处理。加权滑动平均法为不同传感器的监测数据赋予不同的权重。指数平滑法结合了滑动平均法和加权滑动平均法的优点, 同时考虑了当前数据和历史数据的影响, 其原理是任一时刻的指数平滑值是当前观察值与前一时刻指数平滑值的加权平均。具体计算公式为

$$P_t = \alpha x_t + (1 - \alpha) P_{t-1} \quad (1)$$

$$P_{t-1} = \alpha x_{t-1} + (1 - \alpha) P_{t-2} \quad (2)$$

$$P_t = \alpha x_t + (1 - \alpha) (\alpha x_{t-1} + (1 - \alpha) P_{t-2}) \quad (3)$$

式中: α 为平滑系数, 取值范围为 $[0, 1]$; P_t 为在时间步 t 上经过平滑后的值; x_t 为时间步 t 上的实际数据; P_{t-1} 为在前一个时间步 $t-1$ 上经过平滑后的值。

从式 (3) 可以看出, P_t 是 P_{t-2} 的 $(1 - \alpha)^2$ 倍, 而在式 (1) 中 P_t 是 P_{t-1} 的 $(1 - \alpha)$ 倍。由此可见, 指数平滑法在对数据进行权重分配时, 会给当前数据赋予较大的权重, 而给历史数据赋予较小的权重, 并且呈指数关系衰减。由于具有这种指数衰减特性, 因而与其他去噪方法相比, 采用该方法进行平滑处理后得到的值更接近当前时刻的观测值, 在降噪的同时减小了信息的损失。同时, 通过指数平滑处理可使传感器的总体趋势变得更加明显, 有利于后续的模型预测。因此, 为了过滤噪声, 使模型能够更好地捕捉传感器信号的变化趋势, 本文对传感器数据进行指数平滑处理。

2.3 评估标准

为了评估所提出方法的预测性能, 本文采用均方根误差 (RMSE) 和 NASA 提出的 Score 评分函数 S 作为模型的评价指标:

$$E_{\text{RMSE}} = \sqrt{\frac{1}{n} \sum_{i=1}^n (d_i)^2} \quad (4)$$

$$S = \begin{cases} \sum_{i=1}^n \exp(-\frac{d_i}{13}) - 1, & d_i < 0 \\ \sum_{i=1}^n \exp(\frac{d_i}{10}) - 1, & d_i \geq 0 \end{cases} \quad (5)$$

式 (4)、(5) 中: d_i 为预测值与实际值的误差; n 为数据集的样本总数。RMSE 和 Score 分数越低,

表明预测效果越好。

2.4 性能评估实验结果

将所提出的方法在 C-MAPSS 数据集的 4 个子集上进行了实验, 并将得到的预测结果与其他文献中基于相同数据集的模型结果进行比较。对比方法有: LSTM、双向长短期记忆网络 (BiLSTM)、深度卷积神经网络 (Deep Convolutional Neural Network, DCNN)、注意力门控卷积神经网络 (Attention-Gated Convolutional Neural Network, AGCNN)、分层深度神经网络 (Hierarchical Deep Neural Network, HDNN)。

为了捕捉数据的动态特性, 通过滑窗对数据集进行重构。由于模型预测性能存在一定的随机性, 故进行了多次实验, 计算了 RMSE 和 Score 的平均值, 结果见表 2。

由表 2 可以看出: LSTM 模型的表现中规中矩, 但在 FD002 和 FD004 上的 RMSE 较高, 分别为 24.49 和 28.17, 但得分较高; BiLSTM 模型在所有数据集上的 RMSE 稍有改善, 尤其在 FD001 和 FD003 上的 RMSE 分别为 13.65 和 12.74, 但得分依然较高; 相比之下, DCNN 模型在 FD001 和 FD003 上的 RMSE 分别为 11.96 和 12.46, 同时得分也较低, 显示出其更高的精度和稳定性; AGCNN 模型同样表现不俗, 在 FD001 和 FD002 上的 RMSE 分别为 12.42 和 19.43, 在特定数据集上的表现优异; 本文模型的表现最为均衡, 在所有数据集上的 RMSE 和得分表现良好, 特别是在 FD001 和 FD003 上的得分分别为 221.00 和 297.01, 凸显其在整体优化和精度提升方面的显著优势。

综上所述, DCNN 和本文模型在各数据集上的综合表现最为优越, DCNN 在 RMSE 上略胜一筹, 而本文模型在得分上表现最为出色, 特别是在 FD001 和 FD003 数据集上展现出显著优势, 表明在实际应用中, 本文模型能够实现更高的精度和更好的综合性能。

2.5 消融研究

为了研究所提出模型中各成分的贡献, 在 C-MAPSS 测试数据上进行了消融研究, 得到的结果见表 3。其中: R-GTN 表示移除 GTN 模块的结果; R-DCK 表示移除动态卷积的结果。

由表 3 可以看出在模型的结构方面: 当移除 GTN 模块后, RMSE 和 Score 分数均增大, 尤其是在具有复杂工况的 FD004 数据集上最为明显, 可见 GTN 部分对模型的影响较大, 证明了 GTN

表 2 采用各模型得到的结果
Tab. 2 Results obtained by each model

模型	RMSE				Score			
	FD001	FD002	FD003	FD004	FD001	FD002	FD003	FD004
LSTM	16.14	24.49	16.18	28.17	338.00	4 450.00	852.00	5 550.00
BiLSTM	13.65	23.18	12.74	24.86	295.00	4 130.00	317.00	5 430.00
DCNN	11.96	20.34	12.46	22.43	229.00	2 730.00	535.00	3 370.00
AGCNN	12.42	19.43	13.39	21.50	225.51	1 492.00	227.09	3 392.00
本文方法	13.41	14.14	11.42	15.16	221.00	1 376.00	297.01	3 422.01

注：加粗字体表示各模型在 4 个子数据集上取得的最佳结果。

表 3 消融实验的 RMSE 和 Score 结果
Tab. 3 RMSE and Score results of ablation experiment

模型	RMSE				Score			
	FD001	FD002	FD003	FD004	FD001	FD002	FD003	FD004
R-GTN	15.34	25.13	15.67	27.63	357	4 572	878	5 798
R-DCK	14.81	24.57	13.54	23.77	237	4 571	362	5 852

可有效捕捉更复杂的长期依赖关系，为后期预测提供良好的特征表达；当移除动态卷积模块后，RMSE 和 Score 分数也均增大，但增大幅度小于 R-GTN，证明了 DCK 可更有效地进行特征提取，并能保留数据的动态特性。

3 结论

本文提出了一种基于动态卷积的混合模型性能评估方法，利用动态卷积模块学习时序数据中的复杂模式，旨在克服现有机器学习方法在处理时序数据动态特性方面的局限性，增强了模型对数据的理解能力。该方法引入门控 Transformer 和自回归模型方法，通过自适应加权使这两种模型能够在不同的使用场景和数据领域中发挥各自的优势，提供更加灵活和精确的性能评估。实验结果表明：基于动态卷积的混合模型性能评估方法在预测误差方面表现显著。

在实际应用中，该方法非常适用于机电控制系统。首先，从机电控制系统采集数据，并输入本文提出的模型进行分析，通过模型评估可以判断机电系统的性能状态。若检测到性能下降或异常，系统会根据分析结果进行调整，如调节控制参数、执行维护操作或发出警报通知。这样可以有效监控和优化机电控制系统运行，确保其始终

处于最佳工作状态，提高系统的可靠性和效率。

参考文献：

- [1] ZHAO J R, YUAN M, CUI J, et al. A novel hierarchical training architecture for Siamese Neural Network based fault diagnosis method under small sample[J]. Measurement, 2023, 215: 112851.
- [2] HAN Y Y, XU M M, SUN X Q, et al. Gear health monitoring and RUL prediction based on MSB analysis[J]. IEEE Sensors Journal, 2022, 22 (5): 4400-4409.
- [3] WANG X K, YANG L T, CAO E Y, et al. A tensor-based t-SVD-LSTM remaining useful life prediction model for industrial intelligence[J]. IEEE Transactions on Industrial Informatics, 2022, 99: 1-12.
- [4] MEI S J, YUAN M, CUI J, et al. Machinery condition monitoring in the era of industry 4.0: A relative degree of contribution feature selection and deep residual network combined approach[J]. Computers and Industrial Engineering, 2022, 168: 108129.
- [5] CHA Y J, WANG Z L. Unsupervised novelty detection based structural damage localization using a density peaks-based fast clustering algorithm[J]. Structural Health Monitoring, 2018, 17(2): 313-324.
- [6] WANG X K, REN L, YUAN R X, et al.

- QTT-DLSTM: A cloud-edge-aided distributed LSTM for cyber-physical-social big data[J]. IEEE Transactions on Neural Networks and Learning Systems, 2022, 34(10): 7286-7298.
- [7] CHA Y J, CHOI W, BÜYÜKÖZTÜRK O. Deep learning-based crack damage detection using convolutional neural networks[J]. Computer-Aided Civil and Infrastructure Engineering, 2017, 32(5): 361-378.
- [8] WANG B, LEI Y G, LI N P, et al. Multiscale convolutional attention network for predicting remaining useful life of machinery[J]. IEEE Transactions on Industrial Electronics, 2021, 68(8): 7496-7504.
- [9] JIANG F, DING K, HE G L, et al. Dual-attention-based multiscale convolutional neural network with stage division for remaining useful life prediction of rolling bearings[J]. IEEE Transactions on Instrumentation and Measurement, 2022, 71: 3525410.
- [10] QIN Y, ZHOU J H, CHEN D L. Unsupervised health indicator construction by a novel degradation-trend-constrained variational autoencoder and its applications[J]. IEEE / ASME Transactions on Mechatronics, 2022, 27(3): 1447-1456.
- [11] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ, USA: IEEE, 2018: 7132-7141.
- [12] LI Y S, YUAN L, CHEN Y P, et al. Dynamic transfer for multi-source domain adaptation[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ, USA: IEEE, 2021: 10993-11002.