

【航空电子电气工程】

基于 HPRIME-TCN-BiGRU-Attention 的 滚动轴承故障诊断

姚政豪¹, 胡明茂^{1,2}, 巩加玉¹, 龚青山^{1,2}, 姜宇¹, 王旭¹

(1. 湖北汽车工业学院 汽车智能制造学院, 湖北 十堰 442000;

2. 湖北汽车工业学院 汽车动力传动与电子控制湖北省重点实验室, 湖北 十堰 442000)

摘要:针对滚动轴承运行过程中不同故障类型的振动信号异常现象以及故障诊断模型超参数调优问题,提出了一种结合混合并行霜冰优化算法(HPRIME)、时间卷积网络(TCN)与注意力(Attention)机制的双向门控循环单元(BiGRU)的轴承故障诊断方法,以提高故障诊断的准确度。将通过参数优化后的变分模态分解(VMD)算法提取的滚动轴承的振动信号作为样本特征,采用 TCN 将样本特征转化为数据矩阵,并输入 BiGRU-Attention 中,引入改进的 HPRIME,分别对 VMD 和 BiGRU-Attention 的超参数进行优化,以此构建 HPRIME-TCN-BiGRU-Attention 故障诊断模型。选用凯斯西储大学公开的滚动轴承驱动端振动数据验证模型的有效性,结果表明,本方法在三种变负载与两种数据不平衡工况下的平均准确度分别为 98.85% 与 98.86%,较原始 GRU 模型准确性提升了 8.08%。此外,结合东南大学的轴承数据,验证了所提出方法具有良好的鲁棒性和泛化能力。

关键词:滚动轴承;故障诊断;霜冰优化算法;双向门控循环单元;时间卷积网络

中图分类号: TH 133.33; TH 17 **文献标识码:** A DOI:10.13486/j.issn.2097-4973.2025.06.004

在航空航天领域,滚动轴承作为关键组件,被广泛应用于飞机发动机、起落架、直升机传动系统等设备。这些设备通常在高速、重载及振动等复杂工况下运行,滚动轴承的可靠性直接影响设备的安全性和任务完成能力。在高速、重载与振动等严苛环境下,滚动轴承容易出现滚动体与滚道的润滑不良、异常磨损与摩擦旋转等故障或潜在故障^[1-2],进而导致传动系统损坏,严重影响整机的服役性能。因此,研究滚动轴承故障诊断技术,对保障设备稳定运行及降低停机经济损失具有重大理论意义与实际价值^[3]。多年来,已出现多种诊断方法,如油液分析、振动信号分析、腐蚀监测等^[4]。其中,振动信号分析因其可以在不停机或不拆卸机器的情况下识别故障,而备受青睐^[5]。上述故障诊断方法可通过数据监测与诊断取得显著效果,但处理海量数据时效率低^[6]。为此,引入机器学习充分挖掘数据中的隐含非线性关系已成为该领域的热点。

目前,机器学习已在轴承故障诊断领域取得显著成效。夏逸飞等提出了一种基于多尺度知识蒸馏与增量学习的轴承诊断方案,该方案能够有效缓解灾难性遗忘并保持良好的诊断效果^[7]。在考虑环境噪声及变转速工况的背景下,张璐莹等提出了一种时间卷积网络(Temporal Convolutional Networks, TCN)和

收稿日期:2025-04-10

基金项目:国家自然科学基金项目(523755084)

第一作者简介:姚政豪(2000—),男,福建莆田人,硕士研究生,主要从事工业大数据应用研究。

E-mail:2023111106@huat.edu.cn

通信作者简介:胡明茂(1980—),男,湖北十堰人,教授,硕士生导师,博士,主要从事制造信息系统研究。

E-mail:hu@huat.edu.cn

双向门控循环单元(Bi-directional Gated Recirculation Unit, BiGRU)结合的轴承故障诊断模型,该模型通过注意力(Attention)机制抓取数据关键信息,提高了预测准确率^[8]。在滚动轴承现有诊断研究中,机器学习虽然都取得了较好的诊断效果,但仍需要依赖大量的实验数据来调整模型参数。为解决这一问题,陈志刚等提出了使用改进的麻雀优化算法(Sparrow Search Algorithm, SSA)来优化核极限学习机(Kernel Extreme Learning Machine, KELM)的核心参数以建立模型,且同时保持了较高的准确率^[9]。在轴承振动数据预处理方法研究中,杨森等提出运用改进的 SSA 选取变分模态分解(Variational Mode Decomposition, VMD)参数,结合极限学习机(Extreme Learning Machine, ELM)构建轴承故障诊断模型,该模型引入优化算法来调整 VMD 的超参数,以此来增强其特征提取能力,从而提升模型效果^[10]。

优化后的 TCN 在数据处理方面的效果优于卷积神经网络(Convolutional Neural Network, CNN)和长短期记忆(Long Short-Term Memory, LSTM)神经网络。Ding 等提出了结合 TCN、软阈值算法和自注意(Self-attention)机制的旋转机械故障诊断方法,该方法的识别准确率超过了传统 TCN 模型、CNN 和 LSTM^[11]。Zhang 等将 VMD、TCN、BiGRU 以及 SSA 相融合,进行短期风电预测研究,验证了使用优化算法选取模型超参数的可靠性和实用性^[12]。TCN 具有长效记忆、可并行执行卷积、灵活感受野和稳定梯度等优点。但在预测时,模型超参数难以取优,需经多次调试得到相对较优的结果,这增加了模型的使用成本。针对 TCN-BiGRU-Attention 预测效果的偶然性和参数需多次调整问题,本文提出了一种结合混合并行霜冰优化算法(Hybrid Parallel Rime-ice Optimization Algorithm, HPRIME)的 HPRIME-TCN-BiGRU-Attention 滚动轴承故障诊断方法,并通过滚动轴承驱动端振动数据,对模型的准确率进行了分析。

1 算法原理

1.1 变分模态分解算法

VMD 是一种用于信号分解的方法,通过将复杂信号分解为多个固有模态函数简化轴承振动信号处理^[13]。在处理信号时,通过迭代求解变分模态模型的最优解,来确定各模态分量,使各模态分量总和等于原始轴承振动信号,并且各分量之间的带宽最小,其表达式为

$$\min_{\{u_k\}, \{\omega_k\}} \left\{ \sum_k \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \right\}, \text{ s. t. } \sum_k u_k = f.$$

式中: $\{u_k\}$ 和 $\{\omega_k\}$ 分别表示所有和对应中心频率的集合; f 表示轴承原始振动信号; k 表示模态数, $k = 1, 2, 3, \dots, n$; $\delta(t)$ 表示狄拉克分布函数; $\left(\delta(t) + \frac{j}{\pi t} \right) u_k(t)$ 表示经过希尔伯特变换后 $u_k(t)$ 的频谱; ∂_t 表示梯度运算。

为便于求解,将约束变分问题转化为非约束变分问题,引入拉格朗日乘子 $\lambda(t)$, 得到增广拉格朗日公式

$$L(\{u_k\}, \{\omega_k\}, \lambda) = \alpha \sum_k \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 + \langle \lambda(t), f(t) - \sum_k u_k(t) \rangle + \left\| f(t) - \sum_k u_k(t) \right\|_2^2.$$

式中: α 表示惩罚因子, λ 表示拉格朗日算子。公式使用交替方向乘子法,转化为增广拉格朗日函数的鞍点问题。VMD 的实现步骤如下:

步骤 1 初始化 $\{\hat{u}_k^1\}$ 、 $\{\lambda^1\}$ 、 $\{\omega_k^1\}$ 和 n 的值,输入合适的模态数 k 和惩罚因子 α 。

步骤 2 对 \hat{u}_k^1 、 λ^1 、 ω_k^1 进行更新迭代。

$$\omega_k^{n+1} = \frac{\int_0^\infty \omega |\hat{u}_k(\omega)|^2 d\omega}{\int_0^\infty |\hat{u}_k(\omega)|^2 d\omega}, \hat{u}_k^{n+1} = \frac{\hat{f}(\omega) - \sum_{i \neq k} \hat{u}_i(\omega) + \frac{\hat{\lambda}(\omega)}{2}}{1 + 2\alpha (\omega - \omega_k)^2}, \hat{\lambda}^{n+1}(\omega) \leftarrow \hat{\lambda}^n(\omega) + \tau(\hat{f}(\omega) - \sum_k \hat{u}_k^{n+1}(\omega)).$$

步骤 3 判断迭代停止条件为

$$\sum_k \frac{\|\hat{u}_k^{n+1} - \hat{u}_k^n\|_2^2}{\|\hat{u}_k^n\|_2^2} < \epsilon。$$

式中： ϵ 表示判断精度，若满足 $\epsilon > 0$ ，则输出 k 个模态分量，否则返回步骤 2。

1.2 时间卷积网络算法

TCN 是一种专门为分析时间序列数据而设计的卷积神经网络^[14]，它由因果卷积、扩展卷积和残差连接三个关键部分组成。因果卷积确保网络产生与输入相同长度的输出，以及防止未来信息泄漏到过去，其与标准卷积神经网络的不同之处在于因果卷积采用单向结构，确保模型输入和输出的大小相同；扩展卷积通过在输入端进行间隔采样，以更好地解决感知层受限的问题。扩展卷积可以用

$$F(T) = (X_d * f)(T) = \sum_{i=0}^{n-1} f(i) \times x_{T-d \cdot i}$$

表示，输入 $\vec{X} = (x_0, x_1, \dots, x_t)$ 和滤波器 $f: \{0, 1, 2, \dots, n-1\} \rightarrow \mathbf{R}$ 。式中： d 表示扩展因子， n 表示滤波器的大小。

为了解决隐藏层过多可能导致梯度消失等问题，TCN 引入了残差块，残差块运算为

$$o = \text{Activation}(x + F(x))。$$

式中： $x \in \mathbf{R}^n$ ， $F(\cdot)$ 表示扩展卷积， $\text{Activation}(\cdot)$ 表示激活函数。

1.3 BiGRU-Attention 神经网络

BiGRU 模型采用一种具有前向传播和后向传播的双向递归神经网络组成的结构^[15]。BiGRU 同时考虑前后特征的变化规律，相关计算公式为

$$\begin{cases} \vec{h}_t = \text{GRU}(x_t, \vec{h}_{t-1}), \\ \overleftarrow{h}_t = \text{GRU}(x_t, \overleftarrow{h}_{t-1}), \\ h_t = \mathbf{W}_1 \vec{h}_t + \mathbf{W}_2 \overleftarrow{h}_t + b_t。 \end{cases}$$

式中： \vec{h}_t 和 \overleftarrow{h}_t 分别表示前向和后向的隐藏状态， \mathbf{W}_1 和 \mathbf{W}_2 分别表示对应的权重矩阵， h_t 为 t 时刻 BiGRU 层的最终输出状态， b_t 是 t 时刻的偏置项。

为了能够长时间保留轴承故障信息并突出对预测结果影响较大的特征值，同时忽略无关信息，采用注意力机制来进一步处理 BiGRU 的输出信息，计算得到注意力权重矩阵和加权后的特征，具体表达式为

$$u_i = \tanh(\mathbf{W}_s h_i + b_s), \sum_{i=1}^t \frac{\exp(u_i^T u_s)}{\sum_{i=1}^t \exp(u_i^T u_s)} = 1, v = \sum_{i=1}^t a_i h_i, s = \text{softmax}(\mathbf{W}_v v + b_v)。$$

1.4 霜冰优化算法

霜冰优化算法 (Rime-ice Optimization Algorithm, RIME) 是一种基于自然现象的元启发式优化算法^[16]，该算法的原理是模仿自然中霜冰的生成过程。RIME 包括霜冰初始化、软霜搜索策略、硬霜穿刺机制和正向贪婪选择机制四个阶段。RIME 的随机初始化， $\mathbf{R}_{ij}^{\text{init}}$ 为初始化的霜冰粒子群体，其公式为

$$\mathbf{R}_{ij}^{\text{init}} = [S_1, S_2, \dots, S_i]^T, S_i = [x_{i1} x_{i2} \dots x_{ij}]。$$

软霜搜索策略作为算法核心的一部分，利用霜冰粒子的高度随机性和广泛覆盖性，使算法在迭代中能够覆盖整个搜索空间。

$$R_{ij}^{\text{new}} = R_{\text{best},j} + r_1 \times \cos \theta \times \beta \times (h \times (Ub_{ij} - Lb_{ij}) + Lb_{ij}), r_2 < E。$$

式中： R_{ij}^{new} 表示更新后粒子的新位置，是种群中最优代理的第 j 个粒子； Lb_{ij} 和 Ub_{ij} 分别表示第 j 个变量的下界和上界；随机参数 r_1 用于控制粒子移动的方向和步长； θ 表示角度， $\cos \theta$ 控制粒子运动的方向，与 β 结合，其变化随迭代次数而变化； β 表示环境因子，模拟外部环境影响，其值随迭代次数阶梯式减小，用于调整步长并确保算法收敛； h 表示黏附度， $h \in (0, 1)$ 用于随机缩放由上下界的搜索区间；随机参数 r_2 用于

控制粒子位置是否更新。

$$\theta = \pi \times \frac{t}{10 \times T}, \beta = 1 - \lceil \frac{\omega \times t}{T} \rceil / \omega。$$

附着系数 E 直接影响粒子是否更新其位置。随着迭代次数的增加, E 的值逐渐增大, 这意味着粒子之间的凝结概率也随之增加, 有助于算法更有效地探索搜索空间。附着系数公式为 $E = \sqrt{(t/T)}$ 。

由于生长方向相同, 每个霜冰代理可以轻易地交叉穿越, 这一现象称为霜冰穿刺。受此现象启发, 该机制可用于更新代理之间的算法, 使算法粒子能够交换, 从而提高算法跳出局部最优解的能力。替换粒子的公式为 $R_{ij}^{\text{new}} = R_{\text{best}, j}, r_3 < F^{\text{normr}}(S_i)$ 。正向贪婪选择机制确保了种群在每次迭代中都朝着更优的方向进化, 从而提高了算法的整体性能和问题解决的执行效率。

2 改进霜冰优化算法

RIME 存在过早收敛和易陷入局部最优的问题, 针对 RIME 的缺陷, 本文提出了一种 HPRIME。该算法通过与自适应差分进化算法并行计算, 利用贪婪选择机制存储的记忆参数不断挖掘最优解, 最终找到全局最优解。具体的改进策略如下。

2.1 动态软霜搜索策略

动态软霜搜索策略作为原算法的核心部分, 保留了高度随机性和广泛覆盖性, 为防止该算法过早陷入局部最优, 取消了更新判断机制, 引入动态变化的附着系数 E , 将原本的迭代次数函数改成自适应差异 δ 来反应个体在全局和局部中搜索的状态, 以当前最优适应度 f_{best} 和前一代最优适应度 $f_{\text{best}_{\text{prev}}}$ 的差值绝对值加上防止 δ 为 0 的常数 δ , 来帮助算法更有效地探索搜索空间, 优化后该算法更新公式为

$$R_{ij}^{\text{new}} = (N_{\text{rand}} - 0.5) \times 2 \times \cos(\frac{\pi t}{T/10}) \times E, E = e^{-\delta}, \delta = |f_{\text{best}} - f_{\text{best}_{\text{prev}}}| + \delta。$$

式中: $N_{\text{rand}} \in [0, 1]$, 用于调整粒子的搜索方向和幅度; t 表示当前迭代次数; T 表示最大迭代次数; E 直接影响粒子是否更新其位置。

2.2 自适应差分变异算子

将硬霜穿刺策略替换成自适应的差分变异算子, 目的是为降低陷入局部最优的风险。采用著名的 DE/current-to-best/1 变异算子^[17]取代原有的位置更新策略, 从而提高算法跳出局部最优的能力。替换粒子更新公式为

$$\begin{cases} uCR = (1 - c) \times uCR + c \times \text{mean}(CR_s), \\ uF = (1 - c) \times uF + c \times \frac{\sum F_s^2}{\sum F_s}。 \end{cases}$$

式中: CR 表示自适应更新交叉概率, F 表示缩放因子, 它们的初始值均设为 0.5; uCR 表示交叉概率均值; uF 表示缩放因子均值; c 表示学习因子, $c \in [1, 0]$ 控制新的 CR 和 F ; CR_s 是交叉成功时的交叉概率; F_s 表示变异操作成功的缩放因子。变异操作公式为

$$\vec{V}_i = \vec{R}_i + F \times (R_{p\text{best}} - \vec{R}_i) + F \times (R_{r_1} - R_{r_2})。$$

式中: \vec{V}_i 表示个体 i 的变异向量; \vec{R}_i 表示当前个体 i 的位置向量; $R_{p\text{best}}$ 表示从当前种群中最优的 $p\%$ 中随机选择的个体, p 设置为 0.05; R_{r_1} 和 R_{r_2} 分别表示当前种群和冰箱种群中随机选择的两个个体。交叉操作公式为

$$\vec{U}_i^{(j)} = \begin{cases} \vec{V}_i^{(j)}, r \leq CR \text{ 或 } j = j_{\text{rand}}, \\ \vec{R}_i^{(j)}, \text{其他}。 \end{cases}$$

式中: $\vec{U}_i^{(j)}$ 表示个体 i 在第 j 个维度上的实验向量; $\vec{V}_i^{(j)}$ 表示个体 i 在第 j 个维度上的变异向量; $\vec{R}_i^{(j)}$ 表示

个体 i 在第 j 维度上的原位置向量; j_{rand} 表示在所有维度中随机选择的一个维度, 保证至少一个维度发生变异; r 表示作为随机数与 CR 进行比较, $r \in [0, 1]$ 。

2.3 改进贪婪选择机制

RIME 的正向贪婪选择策略仅考虑保留最优解, 而改进贪婪选择机制在保留最优解的基础上, 将所对应解、成功交叉和变异参数进行储存, 提高算法的整体性能的同时为下次迭代提供方案:

$$\text{if } f(U_i) < f(R_i), \text{ then } R_i = U_i; \text{ if } f(R_i^{\text{new}}) < f(R_{\text{worst}}), \text{ then } A = A \cup \{R_i^{\text{new}}\}.$$

式中: 如果试验个体 U_i 的适应度 $f(U_i)$ 优于当前个体 R_i 的适应度 $f(R_i)$, 则更新个体 R_i , 即将 U_i 替换为 R_i ; 当 R_i^{new} (更新后的个体) 比种群 A 中最差的个体 R_{worst} 更优时, 将 R_i^{new} 加入种群, 达到最大迭代次数时输出最优解。HPRIME 流程见图 1。

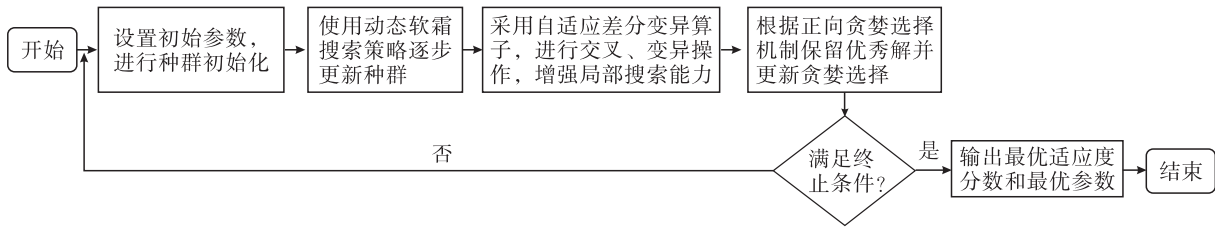


图 1 HPRIME 流程图

3 滚动轴承故障诊断模型

3.1 故障诊断流程

基于 HPRIME-VMD-TCN-BiGRU-Attention 的轴承故障诊断模型流程如图 2 所示。首先, 将滚动轴承的正常振动和故障振动信号输入 VMD 进行分解, 获得一组 IMF 分量, 并计算每个 IMF 分量的组合熵(排列熵/互信息熵), HPRIME 根据总 IMF 分量最小为适应度优化 k 和 a 两个参数。使用优化后参数对 VMD 进行构建。随后, 根据最小组合熵, 采用 VMD 提取 9 个指标特征(均值、方差、峭度等)后划分训练集和测试集。接着, 将提取特征数据输入 HPRIME-TCN-BiGRU-Attention 模型中对学习率、神经元个数、注意力机制键值和正则化参数进行优化。最后, 采用 HPRIME 输出的最优超参数训练 TCN-BiGRU-Attention 故障诊断模型输出滚动轴承故障诊断结果。

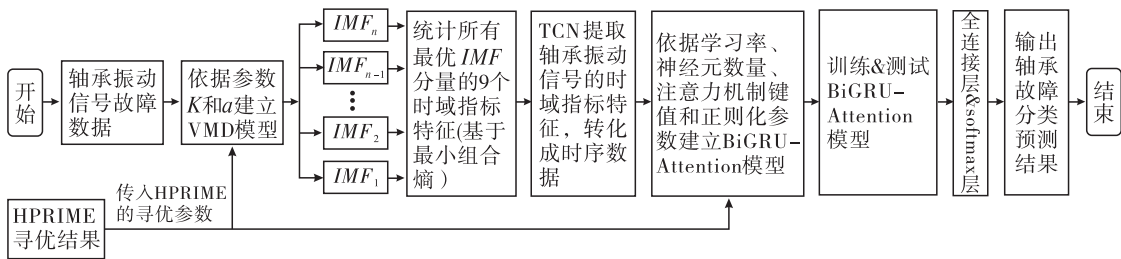


图 2 轴承故障诊断流程图

3.2 模型评价指标

模型评价采用准确率、宏观 F1 分数为评价指标, 公式为

$$P_{\text{Accuracy}} = \frac{TP + TN}{TP + TN + FP + FN}, F_1 = \frac{1}{N} \sum_{i=1}^N \frac{2 \times TP_i}{2 \times TP_i + FP_i + FN_i}.$$

式中: P_{Accuracy} 和 F_1 分别表示准确率和宏观 F1 分数, 准确率衡量模型整体正确分类率, 包括正类和负类。宏观 F1 分数衡量各类别的分类表现是否均衡; TP 和 TN 分别表示真正例和真负例; FP 和 FN 分别表示假正例和假负例; N 表示类别总数; TP_i 、 FP_i 和 FN_i 分别表示第 i 类的真正例数量、假正例数量以及假负例数量。

4 故障诊断实验验证

4.1 实验数据

实验选用美国凯斯西储大学公开的滚动轴承驱动端振动数据^[18]来验证该模型的有效性。实验涵盖了四种不同的轴承状态,分别是 0 负载、1 负载、2 负载条件下损坏尺寸为 0.021 mm 的外圈故障信号、内圈故障信号、滚动体故障信号与正常信号。每个数据类别下分别选择了 1 200 个样本数据,每个样本数据的长度为 2 048 个采样点,经特征提取后得到 120 个样本数据。数据样本组成参见表 1。

表 1 样本数据

标签类别	负载/hp	损坏位置	损坏尺寸/mm	样本数量
1	0/1/2	正常	—	120
2	0/1/2	内圈	0.021	120
3	0/1/2	外圈	0.021	120
4	0/1/2	滚动体	0.021	

为展示故障特征提取方法的效果,运用 t-SNE 方法进行数据可视化(图 3)。通过这一方法发现原始振动数据存在难以区分各个故障状态的显著问题。在不同的特征提取方式中,VMD 结合样本熵的特征向量表现较差,难以区分不同的状态。而使用 VMD 结合排列熵的特征向量虽然在一定程度上区分正常状态与滚动体故障,但在其他故障状态之间却存在明显的重叠现象。相比之下,VMD 结合组合熵的特征向量构建方法显著提升了故障状态的区分能力,能够清晰地区分出四种不同的故障状态。由此可见,VMD+组合熵(排列熵)的特征向量构建方法展现出了最佳的效果,能有效识别和区分不同的故障状况,为后续的故障诊断工作提供了依据。

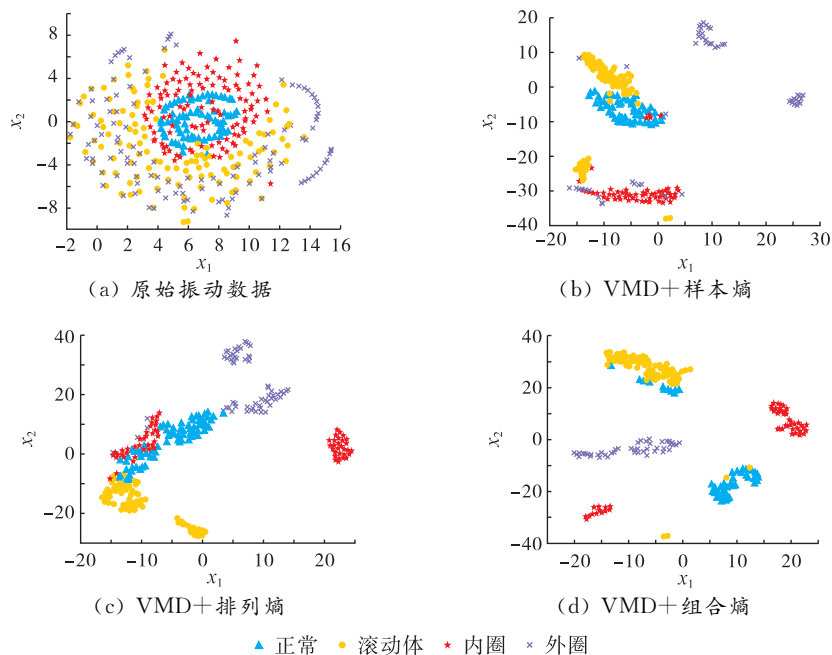


图 3 不同提取方法的 t-SNE 可视化

4.2 特征提取

首先,通过判断各轴承故障类型的 IMF_s 最小组合熵来选择最优的 IMF ,并计算 9 个时域特征;然后,将计算结果作为预处理数据输入 TCN 进行时序特征提取;最后,再将提取的数据输入 BiGRU-Attention 中,

通过 HPRIME 对模型的预测错误率作为适应度目标函数值,实现超参数调优。

进行 VMD 分解时,惩罚因子 α 和模态分解层数 K 的参数较为敏感,对其分解效果有很大影响。以任一驱动轴承故障振动信号为例,使用 HPRIME 得到 VMD 最优参数,迭代次数为 30, $K \in (3, 10)$ 取整数, $\alpha \in (100, 2\ 500)$ 取整数。收敛曲线如图 4 所示,适应度最小值出现在第四次迭代,其中 HPRIME 寻优收敛效果最好,与改进前的 RIME 相比收敛速度和精度都得到了提升^[19]。

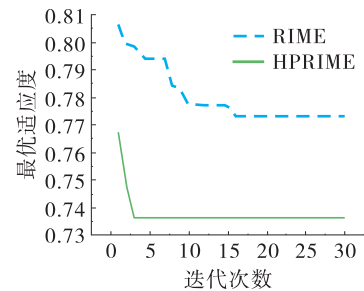


图 4 VMD 参数寻优收敛曲线

在信号处理中,通过 VMD 提取信号的内模态函数,可以清晰隔离信号中的各模态组件,便于分析其特性。选择具有最小组合熵的 IMF,有利于确保信号成分的高纯度和解释性^[20],因为这些 IMF 更规整,所提取特征更直接反映信号的本质属性,而非噪声或复杂性。尽管高熵值可能指示故障,但选用低熵值 IMF 有助于突出故障引起的规律变化。利用最优 IMF 的指标进行故障诊断,这些指标捕捉信号的统计和形态特性,有助于识别各类机械故障^[21]。

4.3 实验结果分析

4.3.1 实验参数设置

将提取特征的数据划分为 7 : 3 用于训练和测试。对比 RIME 和 HPRIME 在 TCN-BiGRU-Attention 模型的超参数选取的效果,HPRIME 表现出最佳的寻优能力,寻优参数结果学习率为 0.001 352 22,隐藏层神经元个数为 55,自注意力机制键值为 16.933,正则化参数为 $1.121\ 4 \times 10^{-4}$ 。

4.3.2 变负载工况实验

使用 Matlab 2024b 构建了模型,针对驱动端轴承的故障和正常振动信号进行了负载工况信真实验。为了进行故障诊断的对比研究,选取 PSO-LSSVM、GWO-TCN-BiLSTM-Attention、RIME-TCN-BiGRU-Attention 作为对照组,为确保公平,均采用 HPRIME-VMD 预处理的时域特征指标作为输入。对比实验表明,在诊断精度方面,HPRIME-TCN-BiGRU-Attention 模型的表现最佳。为避免实验的偶然性,取 10 次实验的准确率均值和标准差作为参考值,实验结果如表 2 所示。

表 2 各模型诊断准确度与标准差对比结果

故障诊断模型	0 负载		1 负载		2 负载	
	准确率/%	标准差	准确率/%	标准差	准确率/%	标准差
SSA-SVM	94.04	2.72	95.53	3.04	77.78	10.84
PSO-LSSVM	92.65	0.53	96.05	1.54	88.01	5.65
GWO-TCN-BiLSTM-Attention	95.87	2.16	97.91	2.16	93.26	1.95
RIME-TCN-BiGRU-Attention	96.67	1.17	98.61	0.55	94.11	1.42
HPRIME-TCN-BiGRU-Attention	99.08	0.77	99.52	0.46	97.97	1.24

HPRIME-TCN-BiGRU-Attention 模型在不同负载条件下表现出卓越的诊断性能和结果稳定性。在 0 负载、1 负载和 2 负载条件下,该模型的诊断准确率分别达到 99.08%、99.52% 和 97.97%,标准差均控制在 1.5 以内;其平均准确率经过计算达到 98.85%,相较基准模型(原始 GRU 模型)的平均准确率提高了 10.93%,表现出良好的诊断性能。传统模型(SSA-SVM、PSO-LSSVM)在高负载条件下的准确率显著下降,结果波动较大,反映出其对复杂环境的适应性较差。深度学习模型通过 TCN、BiGRU 和 Attention 的结合,有效提升了故障特征提取和分类性能。特别是 HPRIME-TCN-BiGRU-Attention 模型在应对负载变化和信号复杂性方面展现出更强的鲁棒性与泛化能力。这得益于该模型的结构优势,其中 Attention

提高了模型对轴承复杂信号的识别能力;TCN 增强了模型对时间序列的理解和处理能力,从而提高了故障检测的准确性;BiGRU 同时考虑了时间序列的前向和后向信息,使得模型在轴承故障识别时更加全面。HPRIME 优化了 BiGRU 在超参数选择上的不足,适当的学习率可以加快模型的收敛速度,避免过高的学习率导致模型波动,或过低的学习率导致收敛缓慢。

4.3.3 数据不平衡工况实验

在实际情况中,机械设备长期处于正常运行状态,故障数据通常相对稀少,正常样本数量的丰富使得模型在训练时对这些样本的学习更加充分,从而对故障数据表现出更高的敏感性。而数量较少的故障样本则很容易被忽视或出现过拟合的情况,进而影响模型的整体性能。这种数据不均衡现象显著减弱了模型在实际应用中的分类能力,因此,必须采取有效的措施来应对这一挑战。为模拟数据不平衡的真实场景,本文选取实验数据中的 0 负载工况为例,按 3:1 和 5:1 的正常数据与故障数据划分数据集 A 和 B,采用五折交叉验证的方法减小单次划分数据集带来的偶然影响,使模型性能评价更稳定,实验结果见表 3。

表 3 各模型数据不平衡工况实验结果

故障诊断模型	数据集 A		数据集 B	
	准确率/%	宏观 F1	准确率/%	宏观 F1
SSA-SVM	94.44	93.62	84.21	65.21
PSO-LSSVM	91.67	90.25	92.98	88.68
GWO-TCN-BiLSTM-Attention	94.52	93.45	96.49	95.16
RIME-TCN-BiGRU-Attention	96.38	95.46	96.84	96.21
HPRIME-TCN-BiGRU-Attention	98.61	98.57	99.12	98.87

实验结果表明,在数据不平衡的场景下,不同故障诊断模型的性能存在显著差异。传统模型(SSA-SVM 和 PSO-LSSVM)在数据不平衡加剧时,性能显著下降,尤其是宏观 F1 分数在 B 数据集上明显降低。这反映出它们对少数类样本的学习能力不足。而基于 TCN 和 Attention 的深度学习模型表现更为稳定,尤其是 HPRIME-TCN-BiGRU-Attention 模型在 A、B 数据集上均取得了最优表现,准确率分别达到 98.61% 和 99.12%,其平均准确率达到 98.86%,与基准模型相比提升了 10.68%。在少数故障样本的情况下,该模型仍较为准确地捕捉到故障特征,从而识别出大多数故障样本,拥有较高的宏观 F1 分数,展现出良好的学习能力和鲁棒性。因此,HPRIME-TCN-BiGRU-Attention 模型能够有效应对数据不平衡问题,在复杂故障诊断任务中具备显著优势。

4.3.4 消融实验

为验证每个模块的有效性,将 HPRIME-TCN-BiGRU-Attention 逐步拆解,分析各个模块对总体模型故障诊断准确度的影响^[22]。采用 0 负载状况作为实验验证数据,详细对比结果如图 5 所示。

相比于未融合原始 GRU 神经网络模型,TCN-GRU、TCN-BiGRU 和 TCN-BiGRU-Attention 模型的准确率分别提高了 1.52%、3.90% 和 5.28%。可以看出,三种算法对 GRU 神经网络有正向促进作用。相较于仅添加单个 TCN 算法,融合模型能够更好地捕捉时间序列数据的前后文信息。因此,不同模型在加入新的机制和优化策略后,轴承故障诊断准确率均有所提升。其中,HPRIME-TCN-BiGRU-Attention 表现最为优异,准确率达到 99.08%,比原始 GRU 模型提升了 8.08%。这表明 HPRIME-TCN-BiGRU-Attention 能够在故障诊断任务中提供更高的准确性和可靠性。

4.3.5 模型泛化能力评估

为了评估本文所提模型对其他数据的学习能力,采用东南大学 0 负载下的驱动端数据^[23]共四种故障(滚子故障、内圈故障等)进行实验,为尽量避免其他因素的干扰,实验结果取 10 次重复实验平均准确率。

为更加直观地描述诊断效果,绘制某次分类结果混淆矩阵,如图 6 所示。由图 6 可知,本文模型仅对滚子故障、外圈故障这两种故障的少量样本分类错误,对其他故障均分类准确。该结果表明,本文模型平均准确率较高,具有良好的泛化能力。

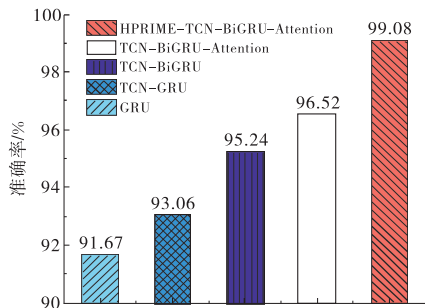


图 5 消融实验测试集实验结果

滚子故障	59				1	98.3%	1.7%
复合故障		60				100.0%	
健康状态			60			100.0%	
内圈故障				60		100.0%	
外圈故障	2				58	96.7%	3.3%
	96.7%	100.0%	100.0%	100.0%	98.3%		
	3.3%				1.7%		
	滚子故障	复合故障	健康状态	内圈故障	外圈故障		
	预测类						

图 6 东南大学轴承数据分类结果

5 结论

本文提出了一种改进优化算法的故障诊断模型,旨在实现轴承故障诊断模型超参数的自动选择,以提高滚动轴承故障诊断的准确性。实验结果表明,该方法能够有效调优轴承故障诊断模型的超参数,准确率较高,并能够高效诊断轴承故障。该研究对解决轴承故障诊断结果不准确与诊断模型参数优化困难等问题具有重要意义。

(1) 采用动态软霜搜索策略增强了算法初期探索的能力,使用自适应差分进化策略与改进的 RIME 并行寻优模型的超参数。采用贪婪选择机制,既保留当前最优解,又储存当前最优解的交叉和变异因子,提高了算法的适应能力和后期精确搜索能力。

(2) 选用滚动轴承在不同工况下的故障数据,主要包含三种变负载工况和两种数据不平衡工况。对比实验表明,利用 HPRIME 对 VMD 和 TCN-BiGRU-Attention 故障诊断模型的平均准确率分别为 98.85%和 98.86%,与基准模型相比,分别提升了 10.93%和 10.68%,具备较为良好的泛化性和鲁棒性。

(3) HPRIME 能够自适应调整模型参数,在大数据时代具有实际意义。消融实验验证结果表明,相比原始 GRU 模型提升了 8.08%;相对于其他模型,HPRIME-TCN-BiGRU-Attention 的诊断效果最好。为进一步验证该方法的泛化能力,采用东南大学轴承数据进行验证,结果表明,所提方法在迁移能力上有效,准确率超过 98%。因此,该方法在轴承故障诊断上具有一定的可行性,具备较高的工程应用价值。但该方法缺乏实际应用案例,今后的研究将着重于混合优化算法故障诊断模型的实际场景应用,为轴承数据驱动的故障诊断方法提供方法和思路。

参考文献:

[1] AHERWAR A,KHALID M S. Vibration analysis techniques for gearbox diagnostic;a review[J]. International journal of advanced engineering technology,2012,3(2):4-12.

[2] DE SÁ SÓ MARTINS D H C,VIANA D P,DE LIMA A A,et al. Diagnostic and severity analysis of combined failures composed by imbalance and misalignment in rotating machines [J]. The international journal of advanced manufacturing technology,2021,114(9):3077-3092.

[3] ALSHORMAN O,ALKAHATNI F,MASADEH M,et al. Sounds and acoustic emission-based early fault diagnosis of induction motor;a review study[J]. Advances in mechanical engineering,2021,13(2):1-19.

[4] KUMAR S,LOKESHA M,KUMAR K,et al. Vibration based fault diagnosis techniques for rotating

- mechanical components; review paper[J]. IOP conference series; materials science and engineering, 2018,376:012109.
- [5] MOHD GHAZALI M H,RAHIMAN W. Vibration analysis for machine monitoring and diagnosis: a systematic review[J]. Shock and vibration,2021,2021(1):9469318.
- [6] SAUCEDO-DORANTES J J,DELGADO-PRIETO M,ORTEGA-REDONDO J A,et al. Multiple-fault detection methodology based on vibration and current analysis applied to bearings in induction motors and gearboxes on the kinematic chain[J]. Shock and vibration,2016,2016(1):5467643.
- [7] 夏逸飞,皋军,邵星,等. 基于多尺度知识蒸馏与增量学习的滚动轴承故障诊断方法[J]. 振动与冲击, 2024,43(12):276-285.
- [8] 张璐莹,侯立群. 基于注意力时间卷积网络和双向门控循环单元的轴承故障诊断[J]. 电力科学与工程, 2023,39(6):62-70.
- [9] 陈志刚,杨远鹏,王衍学,等. 基于 KPCA-CGSSA-KELM 的滚动轴承故障诊断[J/OL]. 轴承(2024-06-13) [2024-07-08]. <https://kns-cnki-net.wvwn.huat.edu.cn/kcms/detail/41.1148.TH.20240612.1906.002.html>.
- [10] 杨森,王恒迪,崔永存,等. 基于改进 AFSA 的参数优化 VMD 和 ELM 的轴承故障诊断[J]. 组合机床与自动化加工技术,2023(4):67-70.
- [11] DING L,LI Q. Fault diagnosis of rotating machinery using novel self-attention mechanism TCN with soft thresholding method[J]. Measurement science and technology,2024,35(4):047001.
- [12] ZHANG Y,ZHANG L,SUN D,et al. Short-term wind power forecasting based on VMD and a hybrid SSA-TCN-BiGRU network[J]. Applied sciences,2023,13(17):9888.
- [13] DRAGOMIRETSKIY K,ZOSSO D. Variational mode decomposition[J]. IEEE transactions on signal processing,2013,62(3):531-544.
- [14] LEA C,FLYNN M D,VIDAL R,et al. Temporal convolutional networks for action segmentation and detection[C]//2014 IEEE congress on evolutionary computation. Donostia:IEEE,2017:156-165.
- [15] 徐鹏,皋军,邵星. 基于 AMCNN-BiGRU 的滚动轴承故障诊断方法研究[J]. 振动与冲击,2023,42(18):71-80.
- [16] SU H,ZHAO D,HEIDARI A A,et al. RIME:a physics-based optimization[J]. Neurocomputing, 2023,532:183-214.
- [17] TANABE R,FUKUNAGA A. Success-history based parameter adaptation for differential evolution [C]//2013 IEEE congress on evolutionary computation. Cancun:IEEE,2013:71-78.
- [18] 西储大学电机滚动轴承振动信号数据集[EB/OL]. [2022-12-16]. <https://csegroups.case.edu/bearingdatacenter/-home>.
- [19] 郝涛,胡明橙,王莉静. 基于 COT-NGO-VMD 与 LSTM 的变转速滚动轴承故障诊断[J]. 组合机床与自动化加工技术,2023(12):188-192.
- [20] DU W,WANG Z,GONG X,et al. Optimum IMF's selection based envelope analysis of bearing fault diagnosis in plunger pump[J]. Shock and vibration,2016,2016(1):1248626.
- [21] YANG J,HUANG D,ZHOU D,et al. Optimal IMF selection and unknown fault feature extraction for rolling bearings with different defect modes[J]. Measurement,2020,157:107660.
- [22] 师国东,胡明茂,宫爱红,等. 基于 XGBoost-MSIWOA-LSTM 的车辆油耗优化预测模型[J]. 计算机集成制造系统,2025,31(9):3467-3484.
- [23] SHAO S,MCALEER S,YAN R,et al. Highly accurate machine fault diagnosis using deep transfer

learning[J]. IEEE transactions on industrial informatics, 2018, 15(4): 2446-2455.

Fault diagnosis method for rolling bearings based on HPRIME-TCN-BiGRU-Attention

YAO Zhenghao¹, HU Mingmao^{1,2}, GONG Jiayu¹, GONG Qingshan^{1,2}, JIANG Yu¹, WANG Xu¹

(1. School of Automotive Intelligent Manufacturing,

Hubei University of Automotive Technology, Shiyan 442000, China;

2. Hubei Provincial Key Laboratory of Automobile Power Transmission

and Electronic Control, Hubei University of Automotive Technology, Shiyan 442000, China)

Abstract: To address the abnormal vibration signals of rolling bearings under various fault types during operation and the hyperparameter tuning issues of fault diagnosis models, a bearing fault diagnosis method combining the hybrid parallel Rime-ice optimization algorithm (HPRIME), temporal convolutional network (TCN), and bidirectional gated recurrent unit (BiGRU) with Attention mechanism is proposed to improve diagnostic accuracy. The vibration signals of rolling bearings extracted by the optimized variational mode decomposition (VMD) algorithm are used as sample features. TCN transforms these features into a data matrix, which is then input into the BiGRU-Attention model. An improved HPRIME is introduced to optimize the hyperparameters of both VMD and BiGRU-Attention, thereby constructing the HPRIME-TCN-BiGRU-Attention fault diagnosis model. The effectiveness of the model is validated using publicly available rolling bearing vibration data from Case Western Reserve University. The results show that the proposed method achieves average accuracy of 98.85% and 98.86% under three variable load conditions and two data imbalance scenarios, respectively, representing an 8.08% improvement over the original GRU model. Additionally, the method's robustness and generalization capabilities are verified using bearing data from Southeast University.

Keywords: rolling bearings; fault diagnosis; RIME optimization; bidirectional gated recurrent unit; temporal convolutional network

(责任编辑:王新亮)

引用格式 姚政豪,胡明茂,巩加玉,等. 基于 HPRIME-TCN-BiGRU-Attention 的滚动轴承故障诊断[J]. 山东航空学院学报, 2025, 42(6): 26-36.

YAO Z H, HU M M, GONG J Y, et al. Fault diagnosis method for rolling bearings based on HPRIME-TCN-BiGRU-Attention[J]. Journal of Shandong University of Aeronautics, 2025, 42(6): 26-36.