

TSEncoder: 基于 SAVMD 和多源数据融合的故障分类^①



季龙炳, 周 宇, 钱 巨

(南京航空航天大学 计算机科学与技术学院, 南京 211106)

通信作者: 周 宇, E-mail: zhouyu@nuaa.edu.cn

摘 要: 针对实际运行机械设备信号易受噪声干扰导致故障特征难以准确提取, 以及设备单一位置信息无法全面反映运行状态的问题, 本研究提出了一种改进的信号自适应分解与多源数据融合的时空故障分类方法. 首先, 提出了一种改进的信号自适应分解算法 SAVMD (signal adaptive variational mode decomposition), 并构建加权峭度稀疏度指标 WKS (weighted kurtosis sparsity) 筛选出富含特征信息的 IMF (intrinsic mode function) 分量, 以实现信号重构. 其次, 将不同位置传感器的多源数据进行融合, 并以周期性采样得到的数据集作为模型的输入. 最后, 构建了一个时空故障分类模型来处理多源数据, 通过改进的稀疏自注意力机制降低噪声干扰, 并利用双编码器机制实现对时间步长和空间通道信息的有效处理. 在 3 个公开的机械设备故障数据集上进行实验, 平均准确率分别达到了 99.1%、98.5% 和 99.4%. 与其他故障分类方法相比表现更好, 具有良好的自适应性和鲁棒性, 为机械设备的故障诊断提供了一种可行的方法.

关键词: 信号自适应分解; 多源数据融合; 时空故障分类模型; 故障分类

引用格式: 季龙炳, 周宇, 钱巨. TSEncoder: 基于 SAVMD 和多源数据融合的故障分类. 计算机系统应用, 2025, 34(1): 223–235. <http://www.c-s-a.org.cn/1003-3254/9750.html>

TSEncoder: Fault Classification Based on SAVMD and Multi-source Data Fusion

Ji Long-Bing, ZHOU Yu, QIAN Ju

(College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China)

Abstract: Aiming at the problems that mechanical equipment signals in actual operation are susceptible to noise interference, making it difficult to accurately extract fault features, and that the information from a single position of the equipment cannot fully reflect operational status, this study proposes an improved spatio-temporal fault classification method of signal adaptive decomposition and multi-source data fusion. Firstly, an improved signal adaptive decomposition algorithm named signal adaptive variational mode decomposition (SAVMD) is proposed, and a weighted kurtosis sparsity index named weighted kurtosis sparsity (WKS) is constructed to filter out intrinsic mode function (IMF) components rich in feature information for signal reconstruction. Secondly, multi-source data from different position sensors are fused, and the data set obtained by periodic sampling is used as the input of the model. Finally, a spatio-temporal fault classification model is built to process multi-source data, which reduces noise interference through an improved sparse self-attention mechanism and effectively processes time step and spatial channel information by using a dual-encoder mechanism. Experiments on three public mechanical equipment fault datasets achieve average accuracy rates of 99.1%, 98.5%, and 99.4% respectively. Compared with other fault classification methods, it has better performance, good adaptability and robustness, and provides a feasible method for fault diagnosis of mechanical equipment.

^① 基金项目: 国防基础科研项目 (JCKY2022605C006)

收稿时间: 2024-06-18; 修改时间: 2024-07-10, 2024-08-01; 采用时间: 2024-08-13; csa 在线出版时间: 2024-11-15

CNKI 网络首发时间: 2024-11-18

Key words: signal adaptive decomposition; multi-source data fusion; spatiotemporal fault classification model; fault classification

机械装备作为国防工业的关键基础部件,广泛应用于航空航天、电力、汽车工业等关乎国民经济和国防建设发展的领域.在机械设备运行过程中,由于疲劳、磨损和过载等因素,机械设备的故障概率大幅度上升,因此进行故障诊断对降低设备的维护成本以及提升生产效率具有重要意义.随着智能制造和工业物联网的逐步发展,大量的低成本加速度传感器和声学传感器逐渐部署到现代工业制造领域,获得大量不同位置的工业设备运行状态参数.单一传感器无法全面表示机械设备的运行状态而且容易受到外部环境的影响,多源数据融合可以得到运转设备的物理、空间等属性.因此,有必要根据加速度传感器收集的多维度振动数据来实现旋转设备的故障诊断.目前在故障诊断领域主要分为基于信号处理的方法和数据驱动的深度学习方法.

滚动轴承作为旋转机械设备中的主要组件如图1所示,其是否失效直接影响机械设备运行状态和健康状况.滚动轴承故障诊断常采用振动分析法^[1],而采集到的振动信号往往为多个分量的混合,且振动信号具有非平稳、非线性特征,这些振动信号含有丰富的故障特征,可用于故障识别.然而混合在其中的噪声使得提取有用的特征信号变得困难,如何从这类信号中提取有效特征是对机械系统进行故障诊断和早期故障征兆预警的关键.



图1 滚动轴承故障

目前在信号特征提取方面有了较为广泛的研究,对于非平稳振动信号,其频谱随时间变化,所以单一频域或时域表示无法反映所有特征.在这种情况下,时域

分析用于振动信号的分析^[2]就是一种非常有效的方法,以及短时傅里叶变换(STFT)、离散小波变换(DWT)、经验模态分解(VMD)都有被提出并且应用.但是基于信号处理的方法进行故障诊断存在各种噪声干扰故障信号的提取和识别,同时特征提取也非常困难,某些特征非常微弱难以从复杂的背景中提取,需要复杂的算法和大量的计算资源.

近年来基于数据驱动的深度学习方法^[3]成为一个新的研究方向,解决了无法从海量数据中提取特征的难题.基于深度学习的故障诊断近年来取得了显著进展,通过卷积神经网络(CNN)、长短时记忆网络(LSTM)等模型,有效提取和分析故障信号特征.深度学习与传统信号处理方法结合,提供了端到端的解决方案.同时一些基于并行神经网络^[4]、Transformer^[5]、动态辅助模型^[6]、迁移学习^[7]和半监督对抗神经网络^[8]等被广泛运用.上述基于深度卷积神经网络的智能故障分类方法取得了不错的效果,但是在实际的工程应用中发现,信号里包含的大量的噪声导致神经网络无法有效提取故障特征,并且对于信号的处理人工干预度较高.目前大量的故障诊断研究基本只考虑单传感器数据,未充分利用到机械设备多位置的特征信息,而且简单的数据组合或者是拼接也会导致分类准确度不足.

针对以上问题,本文设计了一种SAVMD算法自适应决定信号分解数量和惩罚因子来减少人工干预度,同时构建指标WKS来完成筛选重构,能够选择性地保留那些富含故障信息的分量,并有效降低噪声;接着对处理后的单通道数据进行融合拼接,充分考虑通道间的相关性,基于滑动窗口采样得到富含故障特征的多维数据;最后构建时空故障分类模型TSEncoder,基于稀疏注意力机制减少无关信号噪声的影响,使用双编码器机制分别处理时间步长和通道信息,同时添加一个门控单元合并双编码器特征,最后输出故障分类结果.

为了评估本文所提出的一种故障诊断方法的有效性,在3个公开的实验台数据集(美国CWRU数据集,东南大学实验台数据集,渥太华大学实验台数据集)上进行实验,结果表明本方法的平均准确度分别到达了99.1%、98.5%和99.4%,比传统方法有了较大的提升.

本文的贡献在如下几个方面.

- (1) 本文提出了噪声情况下的信号处理方法, 基于 SAVMD 算法完成信号的自适应分解, 构建 WKS 指标筛选保留富含故障信息的分量.
- (2) 本文构建了一种适用于多源数据的时空故障诊断模型, 利用多源数据通道间的相关性, 融合了双编码器和稀疏自注意力机制.
- (3) 本文在 3 个真实的实验台数据集上的实验结果表明, SAVMD-TSEncoder 的表现优于传统数据处理方法和深度学习模型, 并具有更强的鲁棒性.

1 相关研究

当前已经有许多方法专注于机械设备的故障诊断研究, 主要分为基于信号处理的故障诊断和基于深度学习的故障诊断两类.

1.1 基于信号分解的方法

目前, 在机械设备的故障诊断中由于存在大量噪声干扰的问题, 导致故障特征难以被准确提取出来. 针对此问题, 国内外很多学者提出了很多信号分解方法, 例如 Huang 等人^[9]提出了 EMD 方法, 但是在应用中存在模态混叠和端点效应问题. Wu 等人^[10]提出的集合经验模态分解 EEMD 抑制的 EMD 方法中存在的问题, 之后 VMD 分解方法被提出用来解决多分量非平稳信号的分解问题, 但是模态个数和惩罚因子需要认为经验设置. 王琇峰等人^[11]根据惩罚因子和模态分量中心

频率间的关系和能量角度来选择最优参数组合. 最终通过不同的分解方式将信号分解成不同的模式, 通过分析信号的量纲来判断故障类型.

1.2 基于深度学习的方法

卷积神经网络凭借其强大的特征提取能力和非线性表征能力成为其中的研究热点, 例如尹文哲等人^[12]提出的一种基于 CNN-SVM 的轴承故障诊断方法完成了对滚动轴承的端到端故障诊断, 王琦等人^[13]基于一维 CNN 迁移学习解决了实际工程中数据分布不同且难以标记的问题, 实现了对滚动轴承的诊断, 李可等人^[14]等人对传统的 VMD 算法参数优化后再进行模型的训练得到了较好的分类结果, 杨瑞恒等人^[15]利用 Res2Net 网络完成轴承的故障诊断, 邢芷恺等人^[16]利用多维度数据进行融合后再利用 CNN 进行训练, Jing 等人^[17]在模型训练前将多维传感器信号进行拼接, 利用 1D-CNN 完成故障诊断, 朱丹宸等人^[18]也是将数据在模型输入前完成融合, 利用 CNN 完成轴承的智能诊断. 上述基于深度卷积神经网络的智能故障分类方法都取得了不错的效果.

2 SAVMD 信号分解算法

VMD 在处理上间歇性和含有噪声的信号具有更好的分解能力, 但是需要确定分解参数. 因此我们设计了一种自适应信号分解算法 SAVMD, 同时构建 WKS 指标筛选保留富含故障信息的分量, 具体流程如图 2 所示.

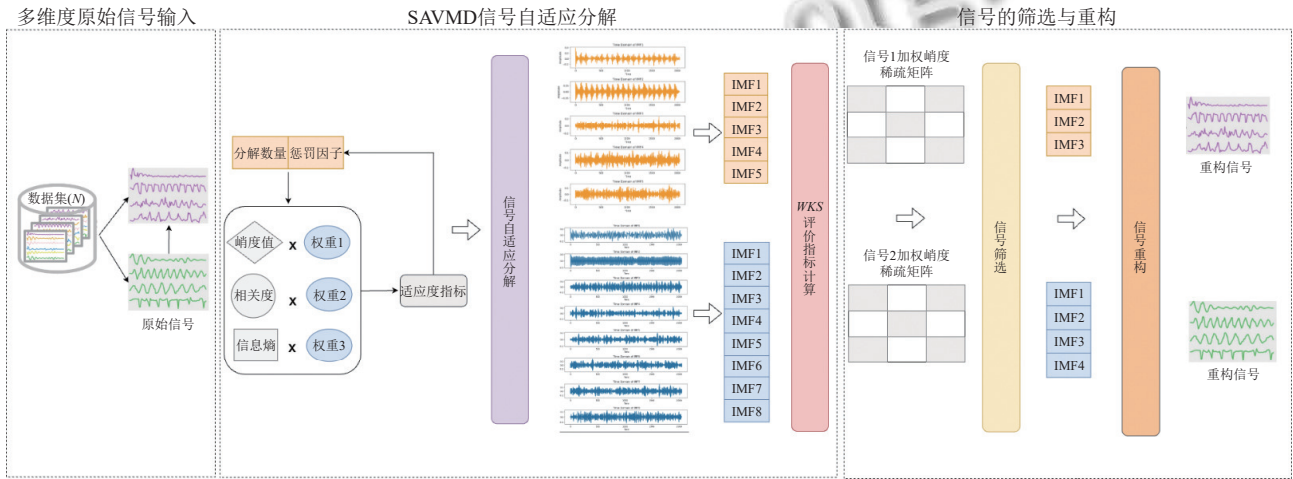


图 2 信号处理流程图

2.1 VMD 变分模态分解

变分模态分解 (variational mode decomposition) VMD 是 2014 年提出的一种非递归信号处理方法, 通

过将时间序列数据分解为一系列具有有限带宽的本征模态函数 (IMF), 迭代搜寻变分模态的优解, VMD 可以自适应更新各 IMF 的最优中心频率和带宽.

假设多成分信号 $f(t)$ 是由 K 个有限带宽的模式分量 $u_k(t)$ 组成, 每个 IMF 分量的中心频率为 $w_k(t)$, 约束条件是模态和等于输入信号^[19-21], 相应的变分约束模型为:

$$\begin{cases} \min \left\{ \sum_k \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) \times u_k(t) \right] e \right\| \right\} \\ \text{s.t. } \sum_k u_k = f \end{cases} \quad (1)$$

其中, $\{u_k\} = \{u_1, \dots, u_k\}$ 代表分解后的 IMF 分量, $\{w_k\} = \{w_1, \dots, w_k\}$ 代表各个组成部分的中心频率。

为了找到约束变分问题的最优解, 首先引入拉格朗日乘子 $\lambda(t)$ 和二阶惩罚因子 α , 将变分约束问题转化为无约束变分问题。其中二阶惩罚因子 α 可以保证高斯噪声环境下信号重构的准确性。拉格朗日乘子 $\lambda(t)$ 可以保证约束条件的严格性。求解该问题可以将约束优化问题等效为一个无约束优化问题, 扩展的拉格朗日表达式如下:

$$\begin{aligned} L(\{u_k\}, \{w_k\}, \lambda) = & \alpha \sum_k \left\| \partial_t \left[\partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) \times u_k(t) \right] e^{-jw_k t} \right] \right\|_2^2 \\ & + \left\| f(t) - \sum_k u_k(t) \right\|_2^2 + \left\langle \lambda(t), f(t) - \sum_k u_k(t) \right\rangle \end{aligned} \quad (2)$$

求解该约束问题可通过交替方向乘子法 (ADMM) 连续更新各分量及其中心频率, 最终得到无约束模型的鞍点, 即原问题的最佳解, 其思想简述为固定另外两个变量, 更新其中一个变量, 如下:

$$\begin{cases} u_k^{n+1} = \argmin_{u_k} L(\{u_{i \neq k}^{n+1}\}, \{w_k^n\}, \lambda^n) \\ w_k^{n+1} = \argmin_{w_k} L(\{u_k^n\}, \{w_{i \neq k}^{n+1}\}, \lambda^n) \\ \lambda^{n+1} = \lambda^n + \rho \left(f(t) - \sum_k u_k^{n+1}(t) \right) \end{cases} \quad (3)$$

(a) 求解模态分量 u_k^{n+1} :

$$\begin{aligned} u_k^{n+1} = & \argmin_{u_k} \alpha \sum_k \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) \times u_k(t) \right] e^{-jw_k t} \right\|_2^2 \\ & + \left\| f(t) - \sum_k u_k(t) \right\|_2^2 + \left\langle \lambda(t), f(t) - \sum_k u_k(t) \right\rangle \\ = & \argmin_{u_k} \alpha \sum_k \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) \times u_k(t) \right] e^{-jw_k t} \right\|_2^2 \\ & + \left\| f(t) - \sum_k u_k(t) + \frac{\lambda(t)}{2} \right\|_2^2 \end{aligned} \quad (4)$$

由于傅里叶变换的等效性, 因此利用 Parseval 定理在频域内求解:

$$\begin{aligned} \hat{u}_k^{n+1} = & \argmin_{u_k} \int_0^\infty 4\alpha(w - w_k)^2 |u_k(w)|^2 \\ & + 2 \left| f(w) - \sum_i u_i(w) + \frac{\lambda(w)}{2} \right|^2 dw \end{aligned} \quad (5)$$

对上述求导可知:

$$\hat{u}_k^{n+1}(w) = \frac{\hat{f}(w) - \sum_{i \neq k} \hat{u}_i(w) + \hat{\lambda}(w)/2}{1 + 2\alpha(w - w_k)^2} \quad (6)$$

(b) 求解中心频率 w_k^{n+1} :

$$w_k^{n+1} = \argmin_{w_k} \sum_k \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) \times u_k(t) \right] e^{-jw_k t} \right\|_2^2 \quad (7)$$

同理可知:

$$w_k^{n+1} = \frac{\int_0^\infty w |u_k(w)| dw}{\int_0^\infty |u_k(w)| dw} \quad (8)$$

最后对得到的频率分量进行收敛判断, 不断迭代更新。

2.2 SAVMD 信号自适应分解

和传统的 EMD 相比, VMD 在处理间歇性以及含有噪声的信号上具有更好的分解能力。在 VMD 参数设置合理的情况下, 分量的模式混合现象可以得到更好地抑制。然而 VMD 也有不足, 它的两个重要参数, 用于确定分解数量的 k 和用于带宽调整的惩罚因子 α , 需要凭经验手动确定, 会影响分解结果^[22]。在 α 值固定的情况下, 如果 k 太大, 信号会被过度分解这导致同一频段被分解为多模态。如果 k 太小, 信号分解不足, 不同频段的中心频率出现在相同的模式。另一种情况, 当 k 值固定时, 会导致 α 过大模式之间的带宽非常小, 导致有用特征的损失信息。如果 α 值太小, 各模式的带宽变宽, 导致更多的噪音被保留。因此, 进行参数的自适应选择来避免人工手动确定就显得十分重要。

为了解决这个问题, 我们设计了一种新颖的信号分解方法 SAVMD 来减少人工干预度, 不需要凭借先验知识手动确认分解数量和惩罚因子的值, 通过对峭度值、相关度和信息熵 3 者加权值的融合来自定义适应度函数, 以此衡量信号分解质量。同时构建阈值标准来确定 k 值和 α 的值, 这里设置的阈值是 5%, 即下一次信

号的适应度提升值低于 5%, 最终实验证明通过此方法的信号分解质量是优于中心值确认法的. 完整的 SAVMD 分解过程如算法 1 所示.

算法 1. SAVMD 信号自适应分解算法

输入: $u_k^1, w_k^1, n \leftarrow 0$, 模态分解数量 $k \leftarrow 1$, 惩罚因子初始值 $\alpha \leftarrow 50$.

输出: u_k^n, w_k^n , 模态分解数量 k , 惩罚因子 α .

Step 1. **while** $n \leftarrow n+1$ **do**

Step 2. 计算 $Q(k)$:

$$Q(k) = \frac{QualityIndex(k) - QualityIndex(k-1)}{QualityIndex(k-1)} \quad (9)$$

Step 3. **while** $Q(k) < 5\%$

Step 4. 更新 $\hat{u}_k$ 的值, 对于 $w \geq 0$ 的情况:

$$\hat{u}_k^{n+1}(w) = \frac{\hat{f}(w) - \sum_{i \neq k} \hat{u}_i(w) + \hat{\lambda}(w)/2}{1 + 2\alpha(w - w_k)^2} \quad (10)$$

Step 5. 更新 w_k 的值:

$$\lambda^{n+1}(w) \leftarrow \lambda^n(w) + \beta(b(w) - \sum_k u_k^{n+1}(w)) \quad (11)$$

Step 6. 计算分信号峭度值、相关系数、信息熵

Step 7. 计算指标权重 $QualityIndex$ 的值:

$$QualityIndex \leftarrow q_1 \times kurtosis + q_2 \times correlation - q_3 \times entropy \quad (12)$$

Step 8. $k \leftarrow k+1, \alpha \leftarrow \alpha+50$

Step 9. 更新 λ 的值, 对于 $w \geq 0$:

$$\lambda^{n+1}(w) \leftarrow \lambda^n(w) + \beta(b(w) - \sum_k u_k^{n+1}(w)) \quad (13)$$

Step 10. **end while**

Step 11. 收敛性判断:

Step 12. **if** $\frac{\sum_n \|u_N^{k+1} - u_N^k\|_2^2}{\|u\|_2^2} \leq \epsilon$ **then**

Step 13. **break**

Step 14. **end if**

Step 15. **end while**

首先需要计算当前的信号分解质量 $QualityIndex(k)$ 相对于上一次的分解质量 $QualityIndex(k-1)$ 的增幅, 当这个指标 $Q(k)$ 大于 5% 才会进入到循环体; 第 4 行和第 5 行用于更新分解模态信号分量和中心频率; 第 6 行计算分解信号的峭度值、相关度和信息熵; 第 7 行通过每个值以及相应的权重计算得到指标权重 $QualityIndex$ 来作为分解的适应度函数; 第 8 行同时进行 k 值和惩罚因子的更新; 第 9 行更新拉格朗日乘子的值; 第 10 行若满足循环条件就回到第 4 行; 第 12 行进行分解信号收敛性判断, 判断是否收敛, 最终输出得到最优信号分解数量 k 和惩罚因子 α .

2.3 信号的筛选与重构

原始信号经过上述的 SAVMD 自适应信号分解后得到一系列的本征模态分量 IMF, 但是这些 IMF 分量中存在这一些过分解的情况或者是包含虚假信息, 若是

不对这些分量进行筛选, 会影响到我们重构信号的质量, 对故障分类的准确度造成影响. 筛选的目的是要去掉那些包含噪声的 IMF, 而保留那些包含有用信息的 IMF.

目前的主要 IMF 筛选方法包括根据频谱直接筛选的方法 (需要频率信息: 对于这种筛选方法, 研究者需要清楚地知道他们所需要的信号频段). 目前常用的方法包括根据相关度的筛选方法、根据峭度的筛选方法以及一些自适应筛选方法等. 但是信号的振动性质不能由这些单一的指标进行说明, 因此本文构造了加权峭度稀疏值指标 (weighted kurtosis sparsity, WKS), 通过自定义的新型评价指标来筛选 IMF 分量. WKS 定义如下:

$$\left\{ \begin{array}{l} S(i) = \frac{\sqrt{\frac{1}{N} \sum_{t=1}^N u_i^2(t)}}{\frac{1}{N} \sum_{t=1}^N |u_i(t)|} \\ C(i) = \frac{1}{N-1} \sum_{t=1}^N \left(\frac{u_i - \bar{u}_i}{s} \right)^4 - 3 \\ W_i = S(i) \cdot C(i) \\ I_{wks} = W(i) - \frac{1}{K} \sum_{i=1}^K W(i) \end{array} \right. \quad (14)$$

其中, $S(i)$ 为第 i 个模态分量 $u_i(t)$ 的稀疏度; N 为模态分量长度; K 为模态分量个数; $C(i)$ 为第 i 个模态分量 $u_i(t)$ 的峭度值; W_i 为模态分量稀疏值和峭度值的点积结果, 最后对每一个模态分量的加权峭度稀疏值和原始 W_i 同时做一个差值运算, 得到 I_{wks} . 当通过上述公式计算得到 I_{wks} 的值大于 0 时候, 我们认为此分量保留了较多的特征值信息, 意味着该分量包含的信息量或者是重要性超过了期望值, 因此选择该模态分量作为我们信号重构的值.

具体操作是, 对于通过 SAVMD 分解得到的每一个 IMF 计算其稀疏度, 接着使用峭度值计算公式得到每一个 IMF 的峭度值, 代表着信号富含有效特征的程度. 接着将稀疏值和信号峭度值做点积运算得到 w_i , 最后和期望算子做一个差值运算得到最后的 I_{wks} . 当该值大于 0 时候, 选择该分量作为我们信号重构的值.

3 多源数据融合的时空故障诊断模型

首先, 针对数据利用不充分的缺陷, 将不同轴承位

置的振动数据融合成多通道数据, 基于故障规律和滑动窗口进行数据集采样. 接着构建了采用双编码器机制的故障诊断模型 TSEncoder, 利用稀疏自注意力来减少信号无关噪声的影响, 最后基于门控单元融合特征, 输出故障分类结果.

3.1 多源数据融合和数据采样

受到在图像分类任务中使用 RGB 三色通道来全面描述图像特征的思想的启发, 轴承的故障特征也可以用 m 个位置通道来表示. 因此, 将不同位置 $X_t^1, \dots, X_t^m$ (大小为 N) 的数据建模为 m 个单通道数据 (大小 $N \times 1$), 然后将 m 个单通道数据合并得到多通道融合数据 (大小 $N \times m$). 机械设备上不同的传感器都是用来监测不同属性的信息, 我们充分考虑通道之间的相关性. 因此可以直接将这些单源数据拼接成多维度数据来代表同一位置的多态性, 即将单通道转化为 m 通道, 充分利用通道间隐藏的异常特征, 具体如图 3 所示.

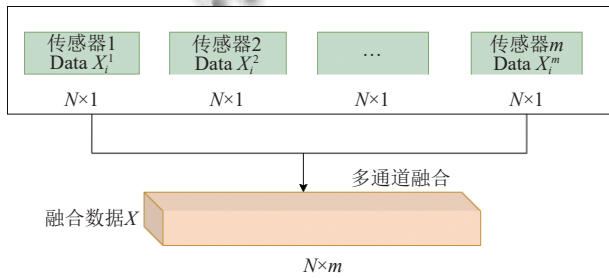


图3 多通道数据融合

根据采样频率 f 和旋转周期 T , 们可以用式 (15) 计算出一个旋转周期内的采样点数. 由于轴承的振动信号是通过旋转产生的, 所以当轴承异常时, 通过旋转产生的振动信号会出现异常波动. 因此轴承的异常信号也具有周期特征, 且周期与马达的旋转周期一致, 因此用式 (18) 定义每个样本的长度 L .

$$\begin{cases} T = \frac{RPM}{60} \\ n = \frac{f}{T} \\ L = \alpha \times n, \alpha \in N \end{cases} \quad (15)$$

考虑到数据集采样点总数小且固定, 在故障分类中, 需要使用数据增强技术来缓解训练样本数据不足的问题, 滑动窗口是最常用的方法. 因此从原始振动信号中, 根据样本长度 L 和滑动步长 S 对样本进行采样, 样本长度设置为轴承在一个或多个旋转周期中采集的采样点.

3.2 融合多维信号的时空 TSEncoder 模型

为了对上述得到的多维故障信号进行分类, 我们基于 Transformer^[23-30]提出了一种网络结构 TSEncoder, 利用稀疏化自注意力机制减少无关信号噪声对于分类结果的影响, 在时间和空间两个维度对数据进行分类. 传统的 Transformer 是由解码器和编码器堆叠而成并且对序列生成和预测任务有相应的位置嵌入, 对于多维信号分类的问题, 我们修改 Transformer 来适应我们的下游任务, 主要包括稀疏化自注意力机制、双编码器机制和门控单元, 具体如图 4 所示.

3.2.1 稀疏自注意力机制

传统的 Self-attention 机制是 Transformer 的重要组成部分, 通过查询向量矩阵 Q 、键向量矩阵 K 、值向量矩阵 V 去计算注意力, 但是其存在的问题是在计算权重时会受到噪声的干扰, 导致注意力集中在无关的元素上, 稀疏自注意力机制^[31-35]通过引入一个掩码或者是限制条件使得注意力集中在有意义的元素上. 我们通常认为稀疏性表现在不同的 query 对于 key 的关注度不同, 并将 query 划分为活跃值和惰性值, 筛选出惰性值完成查询向量矩阵的稀疏化操作. 对原来的 Q, K 矩阵筛选出最活跃的 q_i, k_j 组成 $\bar{Q}, \bar{K}$, 代替原来的 Q, K , 同时时间复杂度的数量级也得到了下降, 最后得到稀疏化机制的自注意力公式如下所示.

$$A(Q, K, V) = \text{Softmax}\left(\frac{\bar{Q} \bar{K}^T}{\sqrt{d}}\right) V \quad (16)$$

3.2.2 双编码器机制

我们得到的多维度信号数据具有多个通道, 其中每个通道是一个单变量的时间序列. 我们通常认为不同的通道、不同的传感器、不同的数据类型之间存在着隐含的相关性, 因此捕获步长 (时间) 和通道 (空间) 的信息是我们进行多维信号故障分类任务的关键. 我们提出了一种双编码器机制, 每个编码器通过注意力机制和掩码设计捕获了时间步长和通道信息, 同时关注时间和空间的信息.

(1) 时间步长编码器: 为了对时间信息进行编码, 在多头稀疏化自注意力层中, 利用缩放点积注意力公式筛选得到时间步长上最值得关注的注意力矩阵. 和传统的 Transformer 结构一样, 使用基于位置的全连接前馈层堆叠在每个多头稀疏化自注意力层上来增强特征提取. 两个子层周围的残余连接都保存了直接信息

和梯度信息, 然后进行层归一化。

(2) 空间通道编码器: 和上述过程类似, 通道信息编码器对不同通道之间的所有时间步长内的信息利用稀

疏自注意力机制筛选得到计算注意力权重矩阵. 由于多维度信号数据力的通道位置信息没有相对或绝对相关性, 类似于切换通道顺序一样, 时间序列应该没有变化。

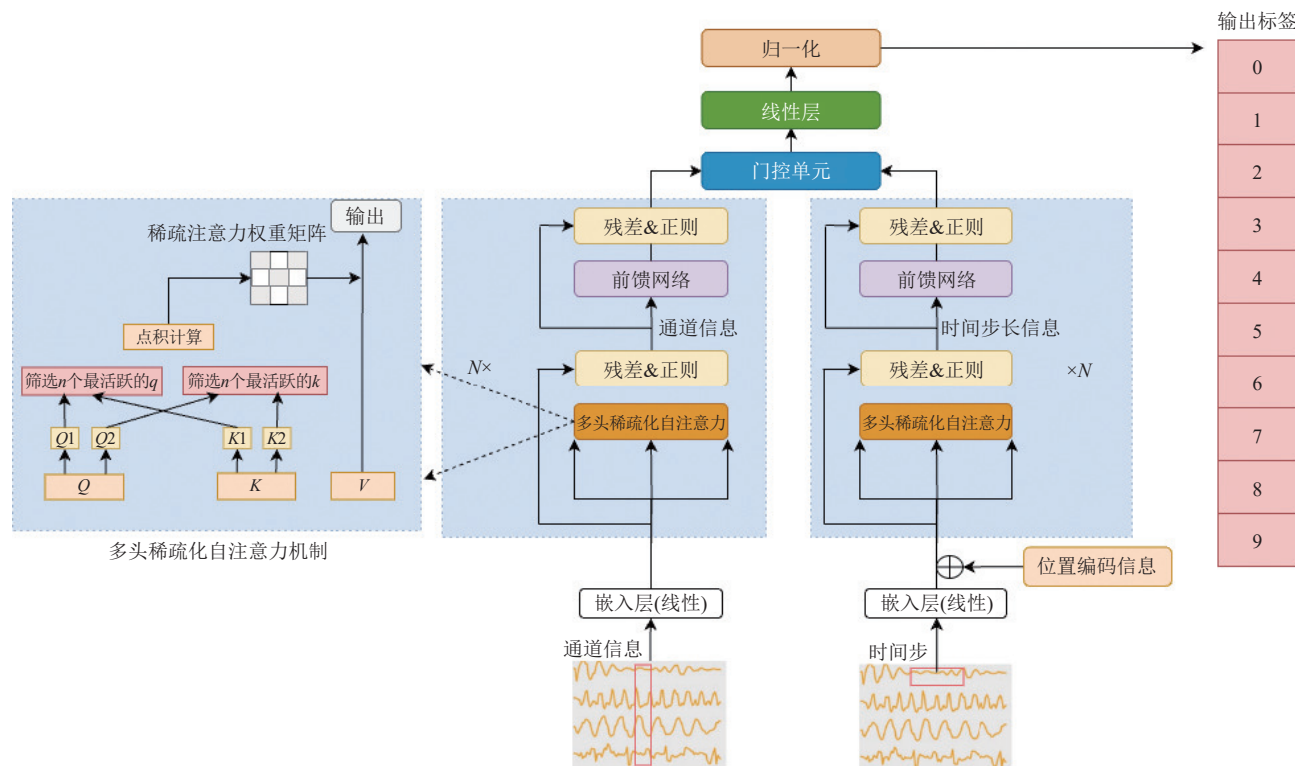


图4 TSEncoder 模型

3.2.3 门控单元

对于不同的数据集, 不同的注意力机制有好有坏, 为合并双编码器产生的时间步长信息和空间通道信息, 一种最简单的方式是直接对这两种信息进行相加。

但是我们设计了一种门控机制来学习每个编码器的权重, 当得到每个编码器的输出之后, 这里包含一个带有非线性激活的全连接层 C 和 S 。

将它们通过一个线性的映射进行拼接得到 h 。经过归一化操作之后, 对门控单元的权重计算得到 g_1 和 g_2 , 之后每个权重被添加到编码器输出计算得到最后的特征向量, 具体如式 (17) 所示。

$$\begin{cases} h = W \cdot \text{Concat}(C, S) + b \\ g_1, g_2 = \text{Softmax}(h) \\ y = \text{Concat}(C \cdot g_1, S \cdot g_2) \end{cases} \quad (17)$$

4 实验分析

4.1 信号分析

实验信号 $x(t)$ 是由某一周期性脉冲信号 $s(t)$ 以及一

个信噪比为 5 dB 的高斯白噪声 $x_n(t)$ 组成, 信号的持续时间为 1 s, 信号的采样频率 $f_s=2048$ Hz, 采样点个数 2048, 信号故障频率为 20 Hz, 载波信号的频率 $f_1=180$ Hz, 信号的基本信息构成如下:

$$\begin{cases} s(t) = 0.8 \exp(-100t_0) \cos(2\pi f_1 t) \\ x(t) = s(t) + x_n(t) \end{cases} \quad (18)$$

实验信号的时域表示和频率表示如图 5 所示, 分别是原始信号, 加噪信号和噪声信号. 加入的白噪声引入了很多宽频率对原始信号 $s(t)$ 的干扰; 从图 6 看出, 信号故障频率为 20 Hz, 其中的 40 Hz, 60 Hz 特征频率是包络谱的主要组成部分. 从图 7 我们可以发现加入的噪声对这些信号的幅值有明显的影响, 在原始信号 $s(t)$ 的包络谱中, 特征频率分量 20 Hz 的幅值为 205, 但是加入噪声之后相同特征频率的幅值变成了 165, 下降了 20.5% 左右, 另外两组 40 Hz 和 60 Hz 对应的包络谱幅值也有相应的下降。

4.1.1 信号自适应分解与重构

利用上述的 SAVMD 方法用来搜寻 $[k, \alpha]$ 的最佳组

合, 具体算法流程如第 2.3 节, 得到的最优结果是 $k = 8$, $\alpha = 500$.
最终分解得到的 8 个 IMF 具有不同的中心频率,

分别是 140 Hz, 180 Hz, 230 Hz, 360 Hz, 480 Hz, 640 Hz, 780 Hz, 940 Hz; 如此看出之前 VMD 模态分解存在模式混合的问题在这里得到了提升.

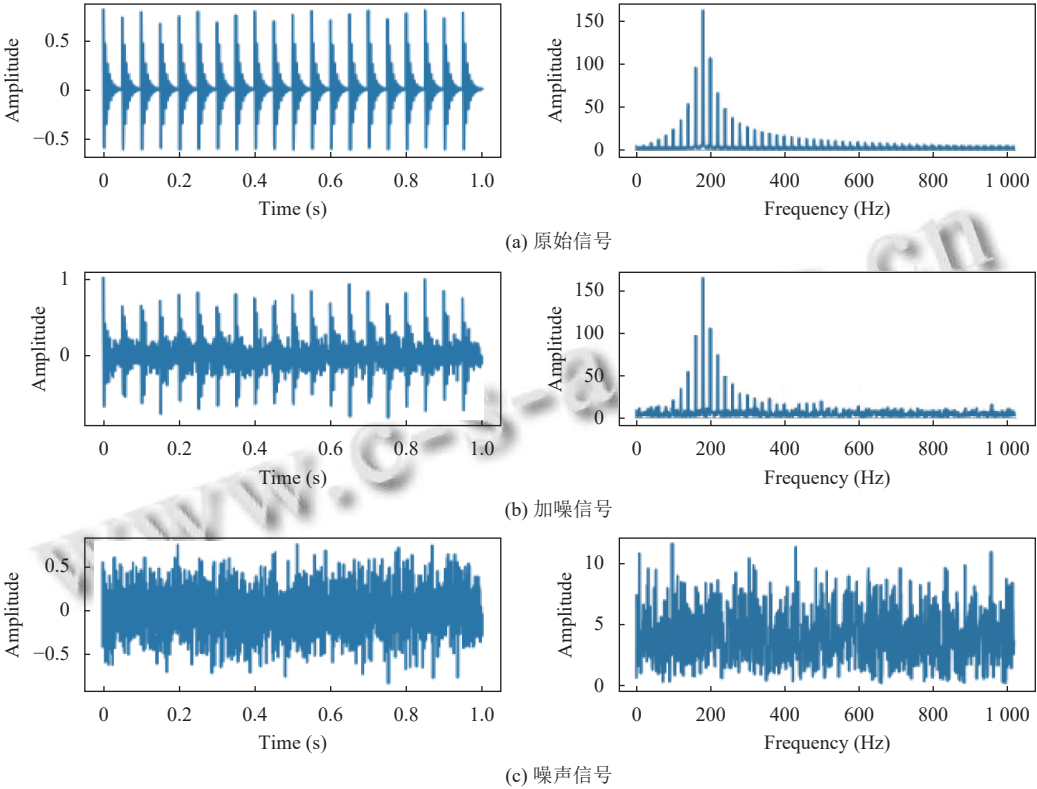


图 5 实验信号时域和频域

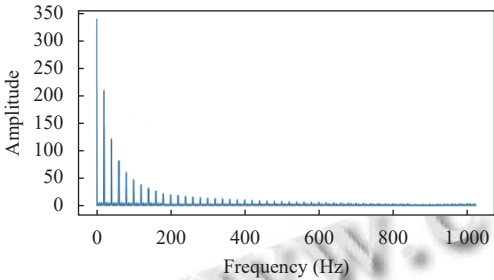


图 6 原始信号包络谱

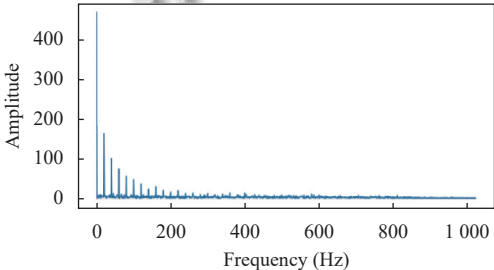


图 7 加噪信号包络谱

同时根据第 2.3 节所描述的, 我们计算 8 个 IMF 分量的 WKS 值如表 1 所示, 选择 WKS 值高于 0 的

IMF 分量进行信号的重建, 因此我们选择 IMF1, IMF2, IMF3, IMF4, IMF5, 重构后的信号包络谱如图 8 所示.

表 1 IMF 分量								
指标	IMF1	IMF2	IMF3	IMF4	IMF5	IMF6	IMF7	IMF8
峭度值	0.74	-0.63	-0.33	1.66	1.50	-0.17	0.22	0.03
稀疏值	1.34	1.19	1.23	1.36	1.31	1.24	1.27	1.25
WKS	0.018	0.024	0.037	0.042	0.138	-0.007	-0.052	-0.022
有效IMF	√	√	√	√	√	—	—	—

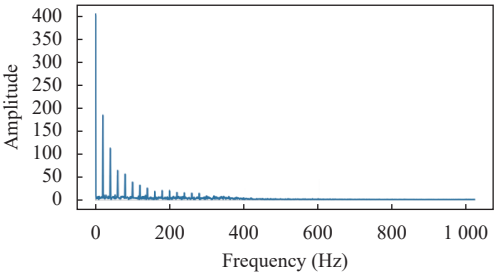


图 8 SAVMD 重构包络谱

为了更好地进行对比, 普通 VMD 方法也被运用于分解故障信号 $x(t)$. 通过中心频率观测法^[24]来确定 k 值, 在 VMD 中使用 $k = 4$, $\alpha = 2000$, 分解信号按照相

同的方法计算信号的 WKS 的值, 选择相应的值进行信号的重构, 重建后的信号包络谱如图 9 表示, 可以发现相同的特征频率下信号的幅值不及 SAVMD 下的分解效果, 20 Hz 下的幅值为 180, 下降了 2.7%, 40 Hz 下的幅值为 101, 下降了 14.4%, 60 Hz 下的幅值为 55, 下降了 26.6%. 原始信号 $s(t)$ 、加噪信号 $x(t)$ 、SAVMD 分解信号和 VMD 分解信号后的包络幅值变化变化参照表 2.

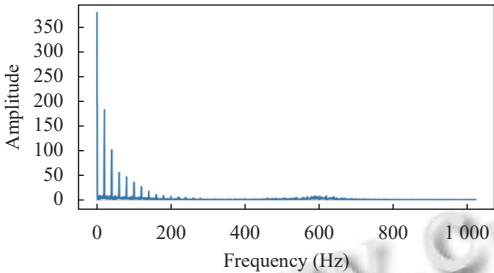


图 9 VMD 重构包络谱

表 2 包络幅值变化

包络幅值 (Hz)	$s(t)$	$x(t)$	SAVMD	VMD
20	210	175	185	180
40	125	100	118	101
60	80	70	75	55

4.1.2 重构信号性能评价

为了定量比较图 8 和图 9 中的重建信号的质量, 我们计算了性能指标——峭度、SNR、RMSE 和能量值, 如表 3 所示.

表 3 信号重构性能指标

指标	$s(t)$	$x(t)$	SAVMD	VMD
峭度值	4.7195	2.4998	2.5570	2.4849
信噪比	—	3.0440	6.8681	4.0251
RMSE	—	0.1031	0.0686	0.0897
Energy	66.329	87.294	65.287	78.133

从表 3 看出, 原始信号 $s(t)$ 的峭度值为 4.7195, 加入噪声的信号 $x(t)$ 的峭度值是 2.4998, 表示噪声覆盖了故障并使得其特征不太明显. 使用基于 SAVMD- WKS 的方法进行信号的分解和重构, 同时去除部分噪声, 信号的峭度值增加到 2.5570, 高于信号 $x(t)$. 而使用普通的 VMD 方法我们得到的峭度值相较于 $x(t)$ 没有得到优化.

对信噪比 SNR 进行比较, 故障信号 $x(t)$ 的 SNR 为 3.0440 dB, 由于基于 SAVMD- WKS 的方法消除了部分噪声, 重建信号的 SNR 增加到了 6.8681 dB, 而基于 VMD 的方法仅仅增加 SNR 到了 4.0251 dB, 小于前者.

对 RMSE 进行比较, 较小的均方根误差值代表较少的扩散信号, 带有噪声的信号 $x(t)$ 的 RMSE 的值为 0.1031; 经过 SAVMD- WKS 消除噪声之后, RMSE 下降至 0.0686; 然而经过 VMD 方法之后 RMSE 值下降至 0.0897, 仍然高于前者.

对信号能量值进行比较, 值越小代表着信号越规则. 原始信号 $s(t)$ 的能量值为 66.329, 由于噪声的存在, 信号 $x(t)$ 的能量值增加到了 87.294. 经过 SAVMD 消除噪声以及基于 WKS 的信号重构, 能量值减少到了 65.287, 低于 VMD 分解的方式为 78.133.

4.2 数据集描述

为了说明实验方法的有效性, 我们选用了 3 种故障信号数据集, 分别是美国凯斯西储大学轴承数据集 (CWRU)、东南大学机械数据集 (SUFD) 以及渥太华大学轴承数据集. 这里使用的 CWRU 数据集 (图 10) 是在 12 kHz 采样频率下采集了两个传感器位置的数据, 也就是 2 个通道数据, 其故障形成是采用电火花技术加工而成, 分为转子、内圈和外圈故障, 其尺寸大小为 0.007 mm、0.014 mm 和 0.021 mm, 因此共分为 10 种故障类型, 并且是以 mat 格式保存的.

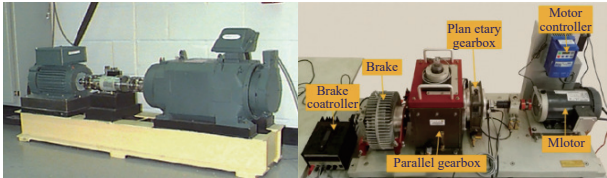


图 10 故障诊断实验台

故障类型分别为: NS、IF7、OF7、BF7、IF14、OF14、BF14、IF21、OF21 和 BF21, 其中 NS 代表正常工作状态, OF 代表外圈故障, BF 代表转子故障, IF 代表内圈故障, 具体如表 4 所示.

表 4 故障类型

标签值	故障类型	损伤程度
0	正常 (NA)	0
1	内圈故障	0.007
2	转子故障	0.014
3	外圈故障	0.021
4	内圈故障	0.007
5	转子故障	0.014
6	外圈故障	0.021
7	内圈故障	0.007
8	转子故障	0.014
9	内圈故障	0.021

东南大学的实验数据集主要分为轴承数据集和齿轮箱数据集,出于统一性考虑我们这里只使用轴承数据集,其中主要包括5种类型,包括转子故障、内圈故障、外圈故障、复合故障和正常状态.轴承数据集一共包括8个通道数据,具体如下:第1列电机振动信号;第2,3,4列分别对应 x, y, z 这3个方向的振动信号;第5列对应着电机扭矩;第6,7,8列分别对应着减速器 x, y, z 这3个方向的振动信号.

渥太华大学轴承数据集包含了不同故障类型的振动数据,实验台包含5种故障类型,包括正常状态、转子故障、内圈故障、外圈故障和组合故障(内圈、外圈和转子).每个故障类型中含有Channel_1和Channel_2的通道信号,它们分别对应着滚动轴承的振动信号(Channel_1)和转速信号(Channel_2).信号的采样频率为200 kHz,采样时间为10 s.

4.3 故障分类流程

首先采用SAVMD得到最优的模态数量 k 和惩罚因子 α ,实现对数据信号的自适应分解;接着采取自定义的WKS指标对分解得到的 k 个IMF分量进行筛选,除去虚假IMF,同时达到对信号的降噪处理,对有效的IMF分量进行信号重构.然后,对于多通道数据集采取融合处理以及周期性采样处理,将重构后的信号分为训练集、测试集和验证集;最后将处理得到的多通道数据输入到时空故障诊断模型中,通过门控单元处理最后得到分类结果.以美国凯斯西储大学CWRU数据集进行说明,使用SAVMD进行信号分解,得到每一种故障类型的最优参数 $[k, \alpha]$ 选择,如表5所示;计算信号每一个IMF分量的WKS值,筛选出值大于0的IMF分量.

表5 最优参数选择

序号	故障类型	$[k, \alpha]$
0	正常(NA)	[4, 200]
1	内圈故障007	[6, 300]
2	转子故障007	[6, 300]
3	外圈故障007	[7, 350]
4	内圈故障014	[8, 400]
5	转子故障014	[8, 400]
6	外圈故障014	[7, 350]
7	内圈故障021	[8, 400]
8	转子故障021	[7, 350]
9	内圈故障021	[8, 400]

根据上述筛选得到的IMF分量进行信号重构,对于每种故障类型的2维度信号进行融合,进行周期性

采样的方法确定每个样本点个数为420,滑动步长为210,以0.75:0.15:0.15的比例将重构信号分为训练集、验证集和测试集,维度分别为(5267, 420, 2), (1117, 420, 2), (1117, 420, 2).采用双编码机制的Transformer和门控单元进行训练.

神经网络中的参数众多,不同的参数选取对最后结果的影响是不同的,一些参数影响训练的速度,一些参数影响分类的准确度,这里的时空故障诊断模型核心是两个编码器、一个Gated门控单元和一个Softmax层,训练参数经过反复调试后具体设置如表6所示.

表6 模型参数

超参	取值
d_model	512
d_hidden	1024
d_input	420
d_channel	2
d_output	10
q, v	8, 8
h	8
n	8
Dropout	0.2
Epoch	100
Batch size	3
Lr	1E-4

根据上述的分类流程,最后得到在CWRU数据集下的训练集和测试集的准确度以及Loss值变化如图11和图12所示,发现能够快速收敛并且稳定性较高.

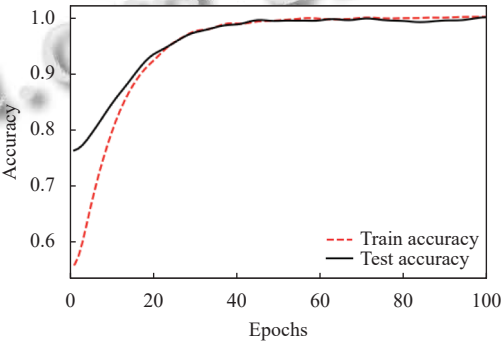


图11 训练集和测试集平均准确度

图13为测试结果的混淆矩阵,测试集数量为1117,每一个故障类别的数量都为110个左右,且每一种故障类别几乎只有一个标签发生了偏移,本方法的平均准确度约为99%.

4.4 对比实验

为了更好地说明所提故障分类方法的有效性,下面进行了多个对比实验.

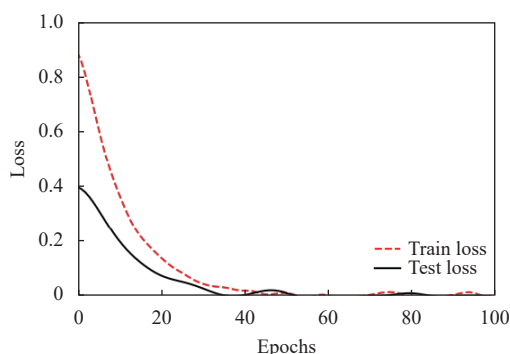


图 12 训练集和测试集损失

(1) 对比 TSEncoder 和其他故障分类方法

分别在 CWRU 数据集 (10 种故障类型, 2 通道数据)、SUFD 数据集 (5 种故障类型, 8 通道数据) 以及 WTHD 数据集 (5 种故障类型, 2 通道数据) 上进行实验。其中, 训练集、验证集、测试集的样本数量是按照 0.75:0.15:0.15 的比例进行划分的。分别在一种改进注意力机制的 CRRA-CNN^[10]、改进注意力机制的 SetNet-

ResNet^[10]、多层卷积网络 MCNN-LSTM^[36]、双向稀疏 Transformer (BST)^[37]以及本文的 TSEncoder 时空模型上使用两种信号分解方法进行对比实验。在 3 种数据集上使用不同方法的效果如图 14 所示, 平均准确度见表 7。

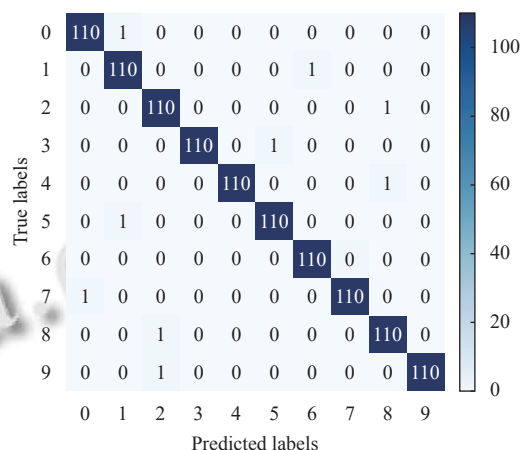


图 13 测试集混淆矩阵

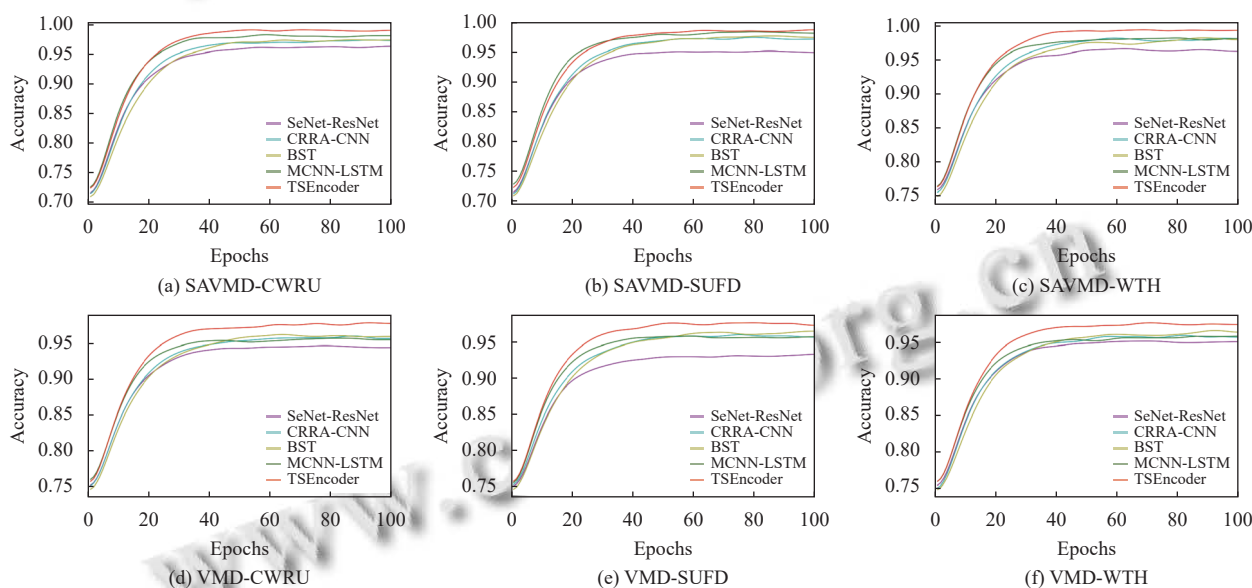


图 14 3 种测试集上不同方法的平均准确度

(2) 对比单维度数据和多维度数据

为了说明融合多位置传感器数据进行故障分类的有效性, 我们在 3 个数据集上分别进行单维度数据和多维度数据的提取处理。对于单维度数据, 随机选取 3 个数据集某一通道的数据进行实验。由表 8 可以发现, 在 5 种故障分类模型 M1: SeNet-ResNet, M2: CRAA-CNN, M3: BST, M4: MCNN-LSTM, M5: TSEncoder 下, 基于多源数据融合的平均分类准确度总是高于单源数

据, 同时在本文所提出的方法下分类精度表现最优。

4.5 消融实验

为了说明本文所提出的 SAVMD 的时空故障分类方法各个组成部分对于 TSEncoder 模型整体性能包括分类、鲁棒性等效果的提升, 设计了对应的消融实验, 实验结果见表 9。对于采用信号自适应分解方法和 TSEncoder 模型中的双编码器门控机制进行单独的实验以验证其实验效果。设计 2 个变体进行比较: 1) M1:

使用不做处理的原始信号直接进行故障分类. 2) M2: 没有使用稀疏自注意力机制进行权重计算, 处理双编码器的输出没有使用门控机制进行权重的计算而是直接进行拼接. 3) M3 是本文中采取的方法. 所有的消融实验均在 3 个数据集上进行, 参数设置和训练过程保持一致.

表 7 不同方法在 3 种数据集的平均准确度 (%)

分类方法	CWRU	SUFD	WTHD
VMD-SeNet-ResNet ^[10]	94.5	93.2	94.8
SAVMD-SeNet-ResNet	96.2	95.0	96.4
VMD-CRAA-CNN ^[10]	95.7	95.1	95.2
SAVMD-CRAA-CNN	98.2	97.3	97.9
VMD-BST ^[37]	96.1	95.6	96.4
SAVMD-BST	98.7	97.4	97.8
VMD-MCNN-LSTM ^[36]	95.6	96.9	95.8
SAVMD-MCNN-LSTM	98.2	98.1	98.0
VMD-TSEncoder	97.5	97.8	97.7
SAVMD-TSEncoder	99.1	98.5	99.4

表 8 不同维度下测试集的平均准确度 (%)

数据维度	M1	M2	M3	M4	M5
CWRU-1	93.1	95.6	95.3	94.1	98.0
CWRU-2	96.2	98.2	98.7	98.2	99.1
SUFD-1	92.3	95.8	94.3	95.4	97.2
SUFD-8	95.0	97.3	97.4	98.1	98.5
WTHD-1	94.5	92.6	95.5	95.9	98.3
WTHD-2	96.4	97.9	97.8	98.0	99.4

表 9 消融实验 (%)

模型	CWRU	SUFD	WTHD
M1	94.2	93.7	95.1
M2	95.6	94.5	95.4
M3	99.1	98.5	99.4

5 结束语

本文提出了一种改进信号自适应分解和多源数据融合的故障分类方法, 在 3 种公开数据集上验证了其可行性. 由于信号中存在的强噪声掩盖了故障特征, 我们通过基于 SAVMD-WKS 的信号自适应分解和重构完成降噪的同时尽可能保留了和故障相关的特征. 此外, 多传感器的数据包含了更多和设备状态相关的信息, 我们提出一种基于多源数据融合的时空故障诊断模型完成时间和空间信息的融合, 相较于传统的分类模型准确度得到了提升.

目前由于在信号的自适应分解中需要花费较长时间, 方法的实时性较差, 这是未来的研究目标之一. 此外将本方法应用到迁移学习中去解决变工况问题和有

限样本情况也是一个重要的研究目标.

参考文献

- 1 胡智勇, 胡杰鑫, 谢里阳, 等. 滚动轴承振动信号处理方法综述. 中国工程机械学报, 2016, 14(6): 525–531.
- 2 Chen XW, Feng ZP. Time-frequency analysis of torsional vibration signals in resonance region for planetary gearbox fault diagnosis under variable speed conditions. IEEE Access, 2017, 5: 21918–21926. [doi: [10.1109/ACCESS.2017.2763172](https://doi.org/10.1109/ACCESS.2017.2763172)]
- 3 Zhu ZQ, Lei YB, Qi GQ, *et al.* A review of the application of deep learning in intelligent fault diagnosis of rotating machinery. Measurement, 2023, 206: 112346. [doi: [10.1016/j.measurement.2022.112346](https://doi.org/10.1016/j.measurement.2022.112346)]
- 4 Guo JY, Yang YL, Li H, *et al.* A parallel deep neural network for intelligent fault diagnosis of drilling pumps. Engineering Applications of Artificial Intelligence, 2024, 133: 108071. [doi: [10.1016/j.engappai.2024.108071](https://doi.org/10.1016/j.engappai.2024.108071)]
- 5 Xiao YM, Shao HD, Wang J, *et al.* Bayesian variational Transformer: A generalizable model for rotating machinery fault diagnosis. Mechanical Systems and Signal Processing, 2024, 207: 110936. [doi: [10.1016/j.ymsp.2023.110936](https://doi.org/10.1016/j.ymsp.2023.110936)]
- 6 Wang CX, Zhang YX, Zhao ZB, *et al.* Dynamic model-assisted transferable network for liquid rocket engine fault diagnosis using limited fault samples. Reliability Engineering & System Safety, 2024, 243: 109837.
- 7 Misbah I, Lee CKM, Keung KL. Fault diagnosis in rotating machines based on transfer learning: Literature review. Knowledge-based Systems, 2024, 283: 111158. [doi: [10.1016/j.knosys.2023.111158](https://doi.org/10.1016/j.knosys.2023.111158)]
- 8 Han T, Xie WZ, Pei ZY. Semi-supervised adversarial discriminative learning approach for intelligent fault diagnosis of wind turbine. Information Sciences, 2023, 648: 119496. [doi: [10.1016/j.ins.2023.119496](https://doi.org/10.1016/j.ins.2023.119496)]
- 9 Huang NE, Shen Z, Long SR, *et al.* The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. Proceedings of the Royal Society Series A: Mathematical, Physical and Engineering Sciences, 1998, 454(1971): 903–995. [doi: [10.1098/rspa.1998.0193](https://doi.org/10.1098/rspa.1998.0193)]
- 10 Wu ZH, Huang NE. Ensemble empirical mode decomposition: A noise-assisted data analysis method. Advances in Adaptive Data Analysis, 2009, 1(1): 1–41. [doi: [10.1142/S1793536909000047](https://doi.org/10.1142/S1793536909000047)]
- 11 王琇峰, 文俊, 倪泽行, 等. 基于 AVMD 和二阶 FWEO 的滚动轴承故障诊断. 振动、测试与诊断, 2023, 43(2): 246–253.

- 12 尹文哲, 夏虹, 彭彬森, 等. 基于 CNN-SVM 的核电厂轴承故障诊断方法. 哈尔滨工程大学学报, 2023, 44(3): 410–417. [doi: [10.11990/jheu.202106040](https://doi.org/10.11990/jheu.202106040)]
- 13 王琦, 邓林峰, 赵荣珍. 基于一维 CNN 迁移学习的滚动轴承故障诊断. 振动、测试与诊断, 2023, 43(1): 24–30.
- 14 李可, 牛园园, 宿磊, 等. 参数优化 VMD 的滚动轴承故障诊断方法. 振动工程学报, 2023, 36(1): 280–287.
- 15 杨瑞恒. 基于多尺度特征融合的轴承故障诊断方法研究 [硕士学位论文]. 贵阳: 贵州大学, 2022.
- 16 邢芷恺, 刘永葆, 王强, 等. 多源数据融合与改进注意力机制的轴承智能故障诊断. 推进技术, 2023, 44(5): 2204017.
- 17 Jing LY, Wang TY, Zhao M, *et al.* An adaptive multi-sensor data fusion method based on deep convolutional neural networks for fault diagnosis of planetary gearbox. *Sensors*, 2017, 17(2): 414. [doi: [10.3390/s17020414](https://doi.org/10.3390/s17020414)]
- 18 朱丹宸, 张永祥, 潘洋洋, 等. 基于多传感器信号和卷积神经网络的滚动轴承故障诊断. 振动与冲击, 2020, 39(4): 172–178.
- 19 Dragomiretskiy K, Zosso D. Variational mode decomposition. *IEEE Transactions on Signal Processing*, 2014, 62(3): 531–544. [doi: [10.1109/TSP.2013.2288675](https://doi.org/10.1109/TSP.2013.2288675)]
- 20 Ma TW, Zhang XF, Jiang H, *et al.* Early fault diagnosis of shaft crack based on double optimization maximum correlated kurtosis deconvolution and variational mode decomposition. *IEEE Access*, 2021, 9: 14971–14982. [doi: [10.1109/ACCESS.2021.3052852](https://doi.org/10.1109/ACCESS.2021.3052852)]
- 21 Liu H, Xiang JW. A strategy using variational mode decomposition, l-kurtosis and minimum entropy deconvolution to detect mechanical faults. *IEEE Access*, 2019, 7: 70564–70573. [doi: [10.1109/ACCESS.2019.2920064](https://doi.org/10.1109/ACCESS.2019.2920064)]
- 22 雷伟, 廖光忠, 裴浪. 基于改进 DenseNet 模型的滚动轴承故障诊断. 计算机技术与发展, 2024, 34(3): 207–213. [doi: [10.3969/j.issn.1673-629X.2024.03.030](https://doi.org/10.3969/j.issn.1673-629X.2024.03.030)]
- 23 任利强, 贾舒宜, 王海鹏, 等. 基于深度学习的时间序列分类研究综述. 电子与信息学报, 2024, 46(8): 3094–3116. [doi: [10.11999/JEIT231222](https://doi.org/10.11999/JEIT231222)]
- 24 Shao SY, McAleer S, Yan RQ, *et al.* Highly accurate machine fault diagnosis using deep transfer learning. *IEEE Transactions on Industrial Informatics*, 2019, 15(4): 2446–2455. [doi: [10.1109/TII.2018.2864759](https://doi.org/10.1109/TII.2018.2864759)]
- 25 Zhou H, Zhang S, Peng J, *et al.* Informer: Beyond efficient Transformer for long sequence time-series forecasting. *Proceedings of the 38th AAAI Conference on Artificial Intelligence*. Vancouver: AAAI Press, 2021. 11106–11115.
- 26 Dempster A, Petitjean F, Webb GI. ROCKET: Exceptionally fast and accurate time series classification using random convolutional kernels. *Data Mining and Knowledge Discovery*, 2020, 34(5): 1454–1495. [doi: [10.1007/s10618-020-00701-z](https://doi.org/10.1007/s10618-020-00701-z)]
- 27 Karim F, Majumdar S, Darabi H, *et al.* Multivariate LSTM-FCNs for time series classification. *Neural Networks*, 2019, 116: 237–245. [doi: [10.1016/j.neunet.2019.04.014](https://doi.org/10.1016/j.neunet.2019.04.014)]
- 28 Zhang XC, Gao YF, Lin J, *et al.* TapNet: Multivariate time series classification with attentional prototypical network. *Proceedings of the 38th AAAI Conference on Artificial Intelligence*. Vancouver: AAAI Press, 2020. 6845–6852.
- 29 Wang SN, Li BZ, Khabisa M, *et al.* Linformer: Self-attention with linear complexity. *arXiv:2006.04768*, 2020.
- 30 Kitaev N, Levskaya A. Reformer: The efficient Transformer. *Proceedings of the 8th International Conference on Learning Representations*. Addis Ababa: ICLR, 2020. 1–12.
- 31 Child R, Gray S, Radford A, *et al.* Generating long sequences with sparse Transformers. *arXiv:1904.10509*, 2019.
- 32 Yang H, Desell T. Robust augmentation for multivariate time series classification. *arXiv:2201.11739*, 2022.
- 33 Tsai Y H H, Bai S, Yamada M, *et al.* Transformer dissection: An unified understanding for Transformer’s attention via the lens of kernel. *Proceedings of the 2019 Conference on Empirical Methods in Natural Language*. Hong Kong: Association for Computational Linguistics, 2019. 4344–4353.
- 34 Dempster A, Schmidt DF, Webb GI. MiniRocket: A very fast (almost) deterministic transform for time series classification. *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*. Singapore: ACM, 2021. 248–257.
- 35 Tay Y, Bahri D, Metzler D, *et al.* Synthesizer: Rethinking self-attention for Transformer models. *Proceedings of the 38th International Conference on Machine Learning*. PMLR, 2021. 10183–10192.
- 36 Chen XH, Zhang BK, Gao D. Bearing fault diagnosis base on multi-scale CNN and LSTM model. *Journal of Intelligent Manufacturing*, 2021, 32(4): 971–987. [doi: [10.1007/s10845-020-01600-2](https://doi.org/10.1007/s10845-020-01600-2)]
- 37 王慧强, 陈楚皓, 吕宏武, 等. 基于双向稀疏 Transformer 的多变量时序分类模型. 小型微型计算机系统, 2024, 45(3): 555–561.

(校对责编: 王欣欣)