

基于深度张量投影网络的机械故障诊断方法研究

黄文静, 李志农, 王发麟, 陈亮亮, 龙盛蓉

(南昌航空大学无损检测技术教育部重点实验室, 江西南昌 330063)

摘要: 针对现有的、基于深度卷积神经网络的故障诊断方法利用池化层对高阶输入张量降维时容易破坏张量数据, 造成数据信息丢失, 以及网络结构相对复杂的不足, 构造了一种深度张量投影网络。该网络利用张量投影层代替传统卷积神经网络中的池化层, 在对输入的高阶张量数据进行降维时, 不会对张量数据造成破坏, 避免了特征信息的丢失, 提高了模型对故障的识别准确率; 并且张量投影层是一种维度可变的降维层, 可以简化网络结构。在此基础上, 结合高阶谱和深度张量投影网络各自的优点, 提出了基于深度张量投影网络的机械故障诊断方法。在提出的方法中, 利用高阶谱提取故障信号特征, 将得到的高阶张量谱图输入到构建的深度张量投影网络模型中进行高阶张量降维和识别。提出的方法成功应用到齿轮箱故障诊断中。实验结果表明, 所提方法能够更好地保留原始故障信息, 有效识别不同类型的故障, 准确率优于传统深度卷积神经网络故障诊断方法。

关键词: 故障诊断; 深度张量投影网络; 高阶谱

中图分类号: TH165⁺.3; TP183 **文献标志码:** A **文章编号:** 1004-4523(2024)04-0657-10

DOI: 10.16385/j.cnki.issn.1004-4523.2024.04.012

引言

深度卷积神经网络作为深度学习的经典算法之一, 被广泛应用于各个领域。深度卷积神经网络是由提取特征的卷积层和处理特征的池化层交叠构成的多层神经网络。其中, 池化层作为卷积神经网络中最重要的组成部分, 具有降低数据维度、扩大网络视野和增加网络非线性能力等特点^[1], 池化方式有最大池化和平均池化两种方式。基于深度卷积神经网络的优势, 学者们将深度卷积神经网络广泛应用于机械故障诊断中, 通过对振动信号的时域、频域或者时、频域进行预处理, 将处理后的信号作为深度卷积神经网络的输入, 从而实现了对机械关键部件的故障诊断。例如, 唐波等^[2]利用短时傅里叶变换得到滚动轴承振动时间序列的二维时频谱, 然后通过卷积神经网络自适应提取时频谱中不同故障的特征, 从而实现了故障诊断。Wang 等^[3]提出了一种基于深度卷积神经网络的综合状态监测框架, 提高了深度卷积神经网络模型的分类精度和计算效率。Xin 等^[4]提出了一种将时域、频域和时频数据组织为混合多对象的方法, 并通过深度卷积神经网络诊断故障, 该方法弥补了单一类型数据无法揭示振动信号中丰富的幅度和频率信息的不足。卞景艺等^[5]针

对机电装备故障诊断需要大量专家经验、故障特征识别困难的问题, 在一维深度卷积神经网络基础上进行了改进, 构建了多尺度一维深度卷积神经网络, 提出了基于该网络的故障诊断方法。李长文等^[6]针对传统的采煤机摇臂齿轮故障诊断方法不能自主提取特征, 导致齿轮故障诊断精度和效率不佳等问题, 构建了基于深度残差网络的采煤机摇臂齿轮故障诊断模型。牛乃平等^[7]构建的作用于时域振动信号的一维卷积神经网络故障诊断模型实现了对采煤机摇臂齿轮的故障诊断。胡葛庆等^[8]利用经验模态分解提取振动信号特征, 再利用深度卷积神经网络进行特征融合, 实现了对行星齿轮箱的故障诊断。Liu 等^[9]通过小波包分解将齿轮振动信号转换为二维时频表示, 并将其输入到深度卷积神经网络模型中进行故障分类。Cheng 等^[10]针对不同测量位置的损伤敏感程度或损伤数据完整度的问题, 提出了一种基于振动信号深度卷积神经网络方法, 并应用于早期故障检测中。Heydarzadeh 等^[11]提出了一种深度卷积神经网络齿轮故障诊断模型, 并将离散小波变换得到的初始特征作为该网络模型的输入。Wu 等^[12]提出了以一维卷积神经网络模型对定轴齿轮箱和行星齿轮箱进行故障诊断的方法, 其准确率高于传统的故障诊断方法。Xin 等^[13]提出了一种基于时

收稿日期: 2022-07-26; 修订日期: 2022-09-28

基金项目: 国家自然科学基金资助项目(52075236); 江西省自然科学基金重点项目资助项目(20212ACB202005)。

频表示和深度卷积神经网络的可靠、有效的旋转机械故障诊断方法。Wang等^[14]结合改进的变分模式和深度卷积神经网络,并将其应用到轴承故障诊断中。Li等^[15]针对CNN架构的可解释性问题,提出了一种新颖的小波驱动深度神经网络方法,该方法利用连续小波卷积层替代标准CNN的第一个卷积层。Zhou等^[16]提出了一种部分可解释的CNN模型,用于燃气轮机的故障诊断,通过优化输入测量参数序列来提高典型CNN的性能。Grezmak等^[17]在逐层相关传播的基础上开发了一种可解释的深度卷积神经网络方法,并将此方法用于齿轮箱的故障诊断。然而,上述基于深度卷积神经网络的故障诊断方法在结构上都是采用了“卷积-池化”交替堆叠的方式来构建故障诊断模型,并且池化层在将数据维度减小时,容易破坏张量数据,造成特征信息的丢失,导致模型泛化能力差。因此,有必要寻找新的深度学习模型来弥补现有深度卷积神经网络中存在的不足。

基于此,本文利用张量投影层(TensorProjection Layer, TPL)^[18]代替传统卷积神经网络的池化层,提出了一种基于深度张量投影网络(Deep TensorProject Networks, DTPN)的故障诊断模型,它

在对张量数据降维时,不会造成故障特征信息的丢失,而且拥有较高的准确率。张量投影层是维度可变的降维层,可以简化网络结构。将所提方法应用到故障诊断中,并与传统的卷积神经网络故障诊断方法进行对比,通过实验进一步验证了所提模型的有效性。

1 深度张量投影网络模型的构造

在传统的卷积神经网络中,池化层通常用于卷积层之后,具有降维和压缩特征等特点。而本文利用张量投影层代替池化层构造的深度张量投影网络,不仅能对上一层的卷积层输出的张量数据进行降维,还不会破坏张量结构,从而保留了数据原始特征信息,提高了机械故障诊断的准确率;而且张量投影层是一种可选择降维层,它可以根据任务的需求和数据的特性来调整降维策略,进而简化网络结构。

构造深度张量投影网络时,首先利用卷积层对张量数据进行特征提取,然后经过TPL进行张量降维,这样层层堆叠后,再通过铺平层对数据进行一维化,接着进入全连接层,最后由分类器进行分类输出,其结构图如图1所示。

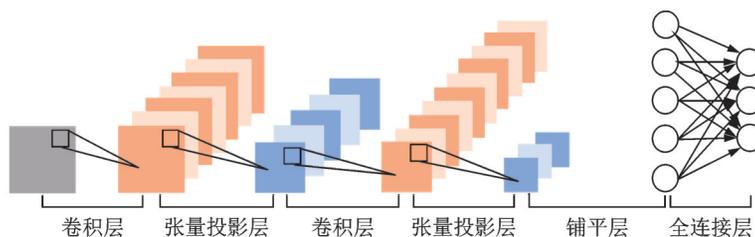


图1 深度张量投影网络结构图

Fig. 1 Structure diagram of Deep TensorProjection Networks

1.1 前向传播

假设深度张量投影网络的输入是三阶张量 $\{X_i\}_{i=1}^n$,其通过卷积层后输出大小为 $p_1 \times p_2 \times p_3$ 的 $\{X_i^l\}_{i=1}^n$,如下式所示:

$$X_i^l = f(\text{conv}(X_i, w^l) + b^l) \quad (1)$$

式中 w^l 表示权重; b^l 表示偏置; i 表示张量数; l 表示网络层数($l=1, 2, \dots, l_0$)。卷积层输出的 X_i 通过张量投影层后输出大小为 $q_1 \times q_2 \times q_3$ 的 $\{Z_i^l\}_{i=1}^n$ ^[18],并且 $q \leq p$,即:

$$Z^l = X^l \times_1 U_1^T \times_2 U_2^T \times_3 U_3^T \quad (2)$$

式中 \times_k 代表 k 模积; U_1, U_2, U_3 为大小为 $p_k \times q_k$ 的正交矩阵,并且有 $U_i^T U_i = I_{q_i}$ ($i=1, 2, 3$), I_q 为单位矩阵。

神经网络通过梯度下降来训练参数,但如果只是简单地运用算法对式(2)中的参数 $\{U_k\}_{k=1}^3$ 进行训练更新,那么 $\{U_k\}_{k=1}^3$ 将不会满足正交。因此,引入大小为 $p_k \times q_k$ 的矩阵 $\{S_k\}_{k=1}^3$ 。 U_k 与 S_k 的关系为:

$$U_k = S_k (S_k^T S_k + \epsilon_k^2 I_{q_k})^{-\frac{1}{2}}; k=1, 2, 3 \quad (3)$$

式中 $\{\epsilon_k^2\}_{k=1}^3$ 为正常数。

然后,经过最后一层张量投影层输出 $\{Z_i\}_{i=1}^n$,再进入铺平层得到输出 F_i ,即:

$$F_i = \text{Flatten}(Z_i) \quad (4)$$

接着,铺平层得到的输出 F_i 再进入到全连接层得到 FC_i ,即:

$$FC_i = f(w F_i + b) \quad (5)$$

最后, FC_i 进入到Softmax分类器中,预测模型

输出,即:

$$\hat{y} = \text{Softmax}(FC_i) \quad (6)$$

1.2 反向传播

在上述的正向传播后,需要通过反向传播对参数进行更新优化。在模型的反向传播中,运用的是交叉熵损失函数,如下式所示:

$$\text{loss} = -[y_i \lg(\hat{y}_i) + (1 - y_i) \lg(1 - \hat{y}_i)] \quad (7)$$

式中 y_i 为真实样本标签; \hat{y}_i 为预测样本标签。

设模型所有的权重参数为 w , 则模型总损失函数表达式为:

$$L(w, b) = \sum \text{loss}(w, b) = -\frac{1}{n} [y_i \lg(\hat{y}_i) + (1 - y_i) \lg(1 - \hat{y}_i)] \quad (8)$$

要使模型识别的准确率达到最优效果,寻找最优参数,即:

$$w := w + \eta \frac{\partial L}{\partial w} \quad (9)$$

式中 $:=$ 为更新运算符; η 为学习率,这里需要通过计算 $\frac{\partial L}{\partial w}$ 找到最小损失函数 L , 则 $\frac{\partial L}{\partial w}$ 的计算公式为:

$$\frac{\partial L}{\partial \text{vec}(w'_i)} = \sum_{i=1}^n \frac{\partial L}{\partial \text{vec}(Z'_i)} \frac{\partial \text{vec}(Z'_i)}{\partial \text{vec}(w'_i)} \quad (10)$$

式中 $\{w'_h\}_{h=1}^{H_l}$ 为第 l 层对应的训练权重参数, $h=1, 2, \dots, H_l$, H_l 为第 l 层的总权重个数。在式(10)中,最重要的是计算 $\frac{\partial L}{\partial \text{vec}(Z'_i)}$, 这也是反向传播的关键。网络第 l 层的输入是第 $l-1$ 层的输出,即:

$$Z_i^{l-1} = X_i^l \quad (11)$$

因此可得^[18]:

$$\frac{\partial L}{\partial \text{vec}(Z_i^{l-1})} = \frac{\partial L}{\partial \text{vec}(X_i^l)} = \frac{\partial L}{\partial \text{vec}(Z_i^l)} \frac{\partial \text{vec}(Z_i^l)}{\partial \text{vec}(X_i^l)} \quad (12)$$

最终,从网络的最后一层 l_0 到第一层依次计算出 $\left\{ \frac{\partial L}{\partial \text{vec}(Z_i^{l-1})} \right\}_l$ 。

下面详细介绍张量投影层的反向传播过程。

首先,令:

$$M_k = S_k^T S_k + \epsilon_k^2 I_{q_k} \quad (13)$$

$$G_k = (S_k^T S_k + \epsilon_k^2 I_{q_k})^{-\frac{1}{2}} = M_k^{-\frac{1}{2}} \quad (14)$$

则得到:

$$U_k = W_k G_k \quad (15)$$

由式(12)得到:

$$\frac{\partial L}{\partial \text{vec}(X_i)} = \frac{\partial L}{\partial \text{vec}(Z_i)} \frac{\partial \text{vec}(Z_i)}{\partial \text{vec}(X_i)} = \frac{\partial L}{\partial \text{vec}(Z_i)} \cdot (U_3^T \otimes U_2^T \otimes U_1^T) \quad (16)$$

通过链式法则求 $\frac{\partial L}{\partial \text{vec}(S_k)}$, 如下式所示:

$$\frac{\partial L}{\partial \text{vec}(S_k)} = \sum_{i=1}^n \frac{\partial L}{\partial \text{vec}(Z_i)} \frac{\partial \text{vec}(Z_i)}{\partial \text{vec}(U_k)} \frac{\partial \text{vec}(U_k)}{\partial \text{vec}(S_k)} \quad (17)$$

将 Z_i, X_i 分别以 k 模积的第 1 模(轴或维)展开,得:

$$Z_{i(1)} = U_1^T X_{i(1)} (U_3 \otimes U_2) = I_{q_1} U_1^T X_{i(1)} (U_3 \otimes U_2) \quad (18)$$

$X_{i(1)}, Z_{i(1)}$ 分别表示 X_i, Z_i 各自的第 1 模展开。将两边向量化,可得:

$$\begin{aligned} \text{vec}(Z_{i(1)}) &= \text{vec}(I_{q_1} U_1^T X_{i(1)} (U_3 \otimes U_2)) = \\ &= \left\{ [(U_3^T \otimes U_2^T) X_{i(1)}^T] \otimes I_{q_1} \right\} \text{vec}(U_1^T) = \\ &= \left\{ [(U_3^T \otimes U_2^T) X_{i(1)}^T] \otimes I_{q_1} \right\} K_{p_1, q_1} \text{vec}(U_1) \end{aligned} \quad (19)$$

式中 K_{p_1, q_1} 为交换矩阵。根据以上推导,可得:

$$\frac{\partial \text{vec}(Z_{i(1)})}{\partial (U_1)} = \left\{ [(U_3^T \otimes U_2^T) X_{i(1)}^T] \otimes I_{q_1} \right\} K_{p_1, q_1} \quad (20)$$

引入置换矩阵 $P_{q_1, q_2, q_3}^{(1)}$, 然后通过链式法则可得:

$$\begin{aligned} \frac{\partial \text{vec}(Z_i)}{\partial (U_1)} &= \frac{\partial \text{vec}(Z_i)}{\partial \text{vec}(Z_{i(1)})} \frac{\partial \text{vec}(Z_{i(1)})}{\partial (U_1)} = \\ &= P_{q_1, q_2, q_3}^{(1)} \left\{ [(U_3^T \otimes U_2^T) X_{i(1)}^T] \otimes I_{q_1} \right\} K_{p_1, q_1} \end{aligned} \quad (21)$$

同理:

$$\frac{\partial \text{vec}(Z_i)}{\partial (U_2)} = \frac{\partial \text{vec}(Z_i)}{\partial \text{vec}(Z_{i(2)})} \frac{\partial \text{vec}(Z_{i(2)})}{\partial (U_2)} =$$

$$P_{q_1, q_2, q_3}^{(2)} \left\{ [(U_3^T \otimes U_1^T) X_{i(2)}^T] \otimes I_{q_2} \right\} K_{p_2, q_2} \quad (22)$$

$$\frac{\partial \text{vec}(Z_i)}{\partial (U_3)} = \frac{\partial \text{vec}(Z_i)}{\partial \text{vec}(Z_{i(3)})} \frac{\partial \text{vec}(Z_{i(3)})}{\partial (U_3)} =$$

$$P_{q_1, q_2, q_3}^{(3)} \left\{ [(U_2^T \otimes U_1^T) X_{i(3)}^T] \otimes I_{q_3} \right\} K_{p_3, q_3} \quad (23)$$

接着令 $S_k \rightarrow S_k + \Delta S_k$, 即:

$$U_k + \Delta U_k = (S_k + \Delta S_k)(G_k + \Delta G_k) \quad (24)$$

忽略 $\Delta S_k \Delta G_k$, 可得:

$$\Delta U_k = \Delta S_k G_k + S_k \Delta G_k = I_{p_k} \Delta S_k G_k + S_k \Delta G_k \quad (25)$$

对上式两边进行向量化后,进行变换求偏导,得:

$$\frac{\partial \text{vec}(U_k)}{\partial \text{vec}(S_k)} = G_k^T \otimes I_{p_k} + (I_{q_k} \otimes S_k) \frac{\partial \text{vec}(G_k)}{\partial \text{vec}(S_k)} \quad (26)$$

并且有:

$$\frac{\partial \text{vec}(G_k)}{\partial \text{vec}(S_k)} = \frac{\partial \text{vec}(G_k)}{\partial \text{vec}(M_k)} \frac{\partial \text{vec}(M_k)}{\partial \text{vec}(S_k)} \quad (27)$$

由式(14)可知 G 与 M 之间的关系, 并且 M 是一个 $q \times q$ 的对称正定矩阵。在 M 中引入一个 ΔM , 即:

$$\Delta G = \lim_{\epsilon \rightarrow 0} \frac{(M + \epsilon \Delta M)^{-\frac{1}{2}} - M^{-\frac{1}{2}}}{\epsilon} \quad (28)$$

将式(28)变形为:

$$\Delta G = M^{-\frac{1}{4}} V \left\{ \lim_{\epsilon \rightarrow 0} \frac{(I_q + \epsilon \mathbf{A})^{-\frac{1}{2}} - I_q}{\epsilon} \right\} V^T M^{-\frac{1}{4}} = \frac{1}{2} M^{-\frac{1}{4}} V \mathbf{A} V^T M^{-\frac{1}{4}} = \frac{1}{2} M^{-\frac{3}{4}} \Delta M M^{-\frac{3}{4}} \quad (29)$$

式中 \mathbf{A} 为矩阵 M 的特征值构成的对角矩阵; V 为正交矩阵, $V V^T = I_q$, 对式(29)两边向量化、求偏导, 即:

$$\frac{\partial \text{vec}(G_k)}{\partial \text{vec}(M_k)} = -\frac{1}{2} \left(M_k^{-\frac{3}{4}} \otimes M_k^{-\frac{3}{4}} \right) \quad (30)$$

再令 $S_k \rightarrow S_k + \Delta S_k$, 即:

$$M_k + \Delta M_k = (S_k + \Delta S_k)^T (S_k + \Delta S_k) \quad (31)$$

忽略 $\Delta S_k^T S_k$, 则得到:

$$\Delta M_k = S_k^T \Delta S_k + (\Delta S_k)^T S_k = S_k^T \Delta S_k I_{q_k} + I_{q_k} (\Delta S_k)^T S_k \quad (32)$$

通过对式(32)两边向量化, 求偏导, 得^[18]:

$$\frac{\partial \text{vec}(M_k)}{\partial \text{vec}(S_k)} = K_{q_k \times q_k} (I_{q_k} \otimes S_k^T) + I_{q_k} \otimes S_k^T \quad (33)$$

综上, 张量投影层的反向传播的参数更新过程就完成了。

整个 DTPN 模型通过正向传播与反向传播实现模型参数更新, 使得参数达到最优, 模型诊断的准确性也达到最佳。

2 基于 DTPN 的故障诊断模型流程

在构造的 DTPN 的基础上, 本文提出了基于 DTPN 的故障诊断方法, 该方法的诊断过程如图 2 所示, 主要包括信号样本划分、高阶张量谱图的构造、DTPN 设计与训练三部分。具体步骤如下:

(1) 信号的采集与样本划分: 将采集到的振动信

号按等窗口长度进行划分, 即得到信号样本; 使用重叠切片的方式, 对数据样本进行增强。

(2) 高阶张量谱图的构造: 传统的信号分析方法很难捕捉到频率随时间的变化, 即使事先采用消噪预处理, 也可能出现因消噪方法的参数或者基函数选择不当而难以达到较好的消噪效果, 而这种选择又往往缺乏依据, 在消噪的过程中, 如果只是简单地基于功率谱分析, 相位耦合信息不能被提取, 不可避免地造成有用信息的丢失。因此, 为解决采集的振动信号中含有噪声信号的问题, 本文引入高阶谱的信号分析方式对振动信号进行特征提取。高阶谱^[19-21]是随机过程高阶累积量的多维傅里叶变换, 本文运用的是双谱, 也称为三阶谱分析方法, 该分析方法同时包含了幅度和相位信息, 能够在不同的工作条件下, 有效地探测给定信号中非线性相互作用, 并实现相位信息的保留及高斯噪声的消除。

将得到的信号样本进行高阶谱分析, 先将一个信号样本分成 P 个片段, 然后计算它的离散系数和三阶自相关系数 β , 最后得到信号样本的高阶谱表达式为^[21]:

$$B(\omega_1, \omega_2) = \frac{1}{P} \sum_{p=1}^P \beta_p(\omega_1, \omega_2) \quad (34)$$

其中, $\omega_1 = \frac{2\pi f_s}{N_0} \lambda_1$; $\omega_2 = \frac{2\pi f_s}{N_0} \lambda_2$; f_s 为采样步数; N_0 为每个信号片段的长度; λ 为特征值。

从式(34)可以看出, 高阶谱是一个信号的二维重定向谱, 它含有两个周期为 2π 的独立频率分量。传感器采集到的振动信号通过高阶谱分析得到一个二维张量的高阶谱图, 从而得到一个 $(\omega_1, \omega_2, \text{通道数})$ 三阶张量。

(3) DTPN 设计与训练: 通过交替堆叠卷积层与张量投影层构造 DTPN, 并设置卷积层中卷积核的大小、步长和张量投影层中的张量维数。将数据集的训练集放入建立好的 DTPN 模型中进行训练, 并通过上述的反向传播算法更新优化模型参数, 使得模型效果达到最优。

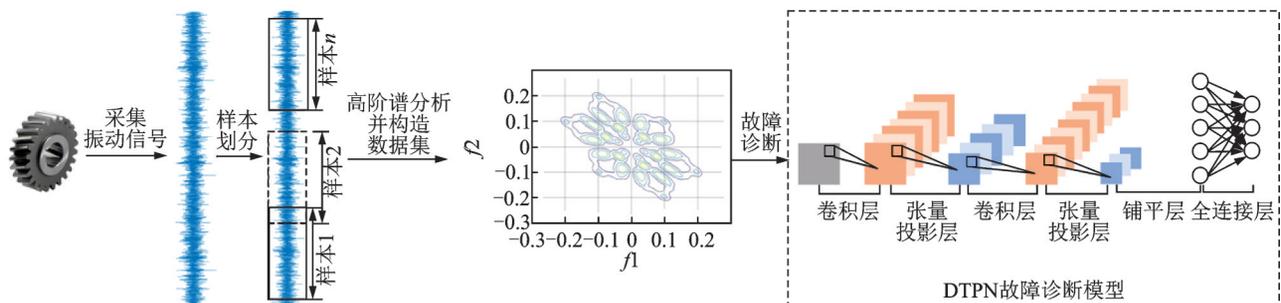


图 2 DTPN 故障诊断流程图

Fig. 2 DTPN fault diagnosis flow chart

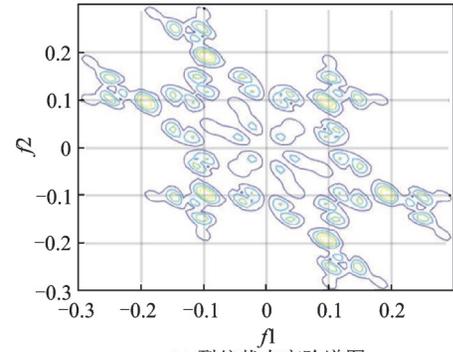
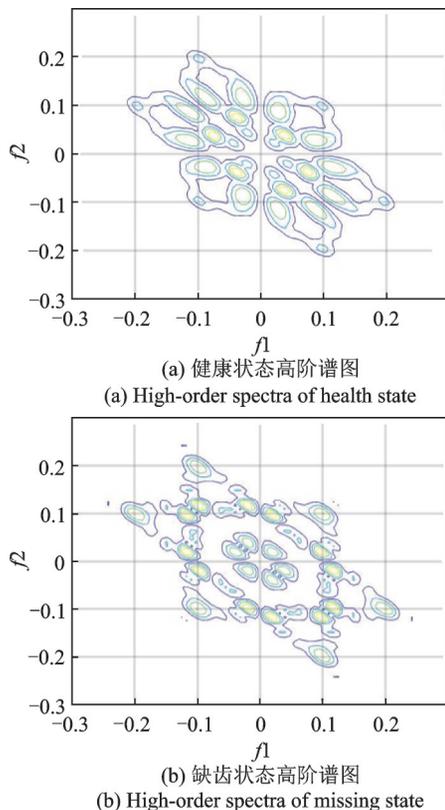
本文提出的方法将故障振动信号转换为高阶谱形式,不仅能去除信号的高斯噪声,还能保留它的相位信息,更全面地显示出了信号的特征。然后利用DTPN网络对高阶张量谱数据进行特征提取与分类,此方法不会对张量数据结构造成破坏导致信息丢失,更好地保留了故障的特征信息。

3 实验验证

3.1 实验数据描述

本文使用的实验数据来自于康涅狄格大学提供的齿轮故障数据集^[22]。该数据集的采样频率为20 kHz,对输入轴上的小齿轮引入了九种不同的齿轮状态,包括健康、缺齿、裂纹、剥落以及五种不同程度的削屑尖端。在原始数据集中,每类有104个样本,每个样本3600个点。本文采用了此数据集的健康、缺齿、裂纹三种齿轮状态,每类状态400个样本,每个样本1024个样本点,并采用等长窗口对齿轮振动信号进行重叠切片,从而增强数据的可靠性。

数据样本按照上述方式划分好后,对其进行高阶谱分析。将每个振动信号样本以高阶张量谱图的形式输出,齿轮每类状态的高阶谱图如图3所示。其中 f_1 和 f_2 分别表示高阶谱的两个独立频率。得到的高阶谱图的大小为 128×128 ,然后将其转换为灰度图像。本文使用的振动信号数据都是来自于同



(c) 裂纹状态高阶谱图
(c) High-order spectra of crack state

图3 齿轮每类状态的高阶谱图

Fig. 3 High-order spectra of each type of gear state

一个通道,最终生成的三阶张量大小为 $128 \times 128 \times 1$ 。

综合以上阐述,构造的数据集共有1200个样本,每类样本的大小是 $400 \times 128 \times 128 \times 1$ 。因为模型是监督学习,所以需要将每类齿轮状态样本打上不同的标签,如表1所示。

表1 实验数据样本

Tab. 1 Experimental data samples

故障类型	样本维度	样本个数	标签
健康	$128 \times 128 \times 1$	400	0
缺齿	$128 \times 128 \times 1$	400	1
裂纹	$128 \times 128 \times 1$	400	2

将构造好的数据样本集打乱,并按8:2的比例分成训练集和测试集,其中训练集样本960个,测试集样本240个。

3.2 齿轮故障诊断结果分析

为验证深度张量投影网络对齿轮故障诊断效果的影响,将提出的方法与传统的深度卷积神经网络进行了对比研究。将上述构造好的张量数据集分别放入DTPN模型和DCNN模型中进行训练和测试,其中DTPN模型和DCNN模型的参数分别如表2和3所示。

表2 DTPN模型各层参数

Tab. 2 Parameters of each layer of DTPN model

结构	输出维度	参数	激活函数
卷积层1	$(n, 128, 128, 16)$	160	ReLU
张量投影层1	$(n, 64, 64, 16)$	16384	—
卷积层2	$(n, 64, 64, 32)$	4640	ReLU
张量投影层2	$(n, 32, 32, 32)$	4096	—
卷积层3	$(n, 32, 32, 64)$	18496	ReLU
张量投影层3	$(n, 16, 16, 64)$	1024	—
铺平层	$(n, 8912)$	—	—
全连接层1	$(n, 160)$	2621600	ReLU
全连接层2	$(n, 3)$	483	Softmax

表3 DCNN模型各层参数

结构	输出维度	参数	激活函数
卷积层1	(n, 128, 128, 16)	160	ReLU
池化层1	(n, 64, 64, 16)	—	—
卷积层2	(n, 64, 64, 32)	4640	ReLU
池化层2	(n, 32, 32, 32)	—	—
卷积层3	(n, 32, 32, 64)	18496	ReLU
池化层3	(n, 16, 16, 64)	—	—
铺平层	(n, 8912)	—	—
全连接层1	(n, 160)	2621600	ReLU
全连接层2	(n, 3)	483	Softmax

在DCNN模型中,池化层采用的是大小为(2, 2)的平均池化层,两个模型中所有卷积层的卷积核大小都为(3, 3),且步长均为1。由于DTPN模型是用张量投影层取代传统卷积神经网络的池化层,每个张量投影层和相对应的平均池化层的输出维度要求保持一致,以保证对比的有效性。两个模型的学习率 η 为0.001,迭代次数设置为150次。分别使用两个模型对输入张量数据分类识别,其结果如图4和5所示。

由图4可知,随着迭代次数的增加,DTPN故障诊断模型的训练集和测试集的准确率逐渐提高,最终两条曲线重合,都达到了100%,同时,训练集与测试集的loss值稳定下降,逐渐收敛于0,迭代过程

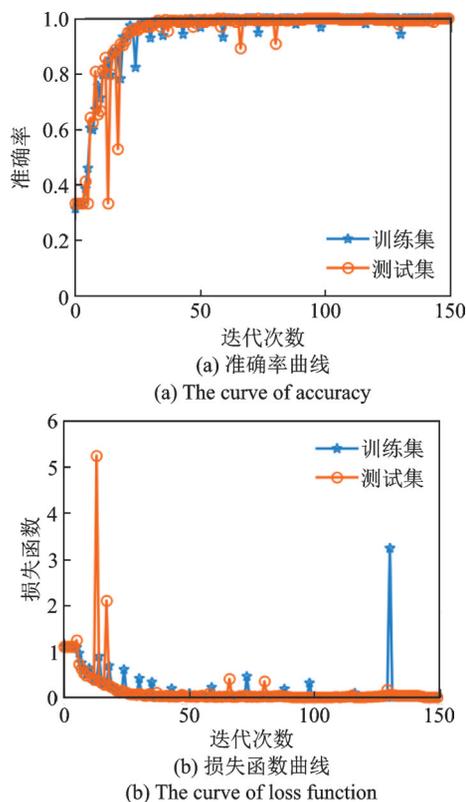


图4 DTPN模型诊断结果曲线
Fig. 4 Curve of diagnosis result of DTPN model

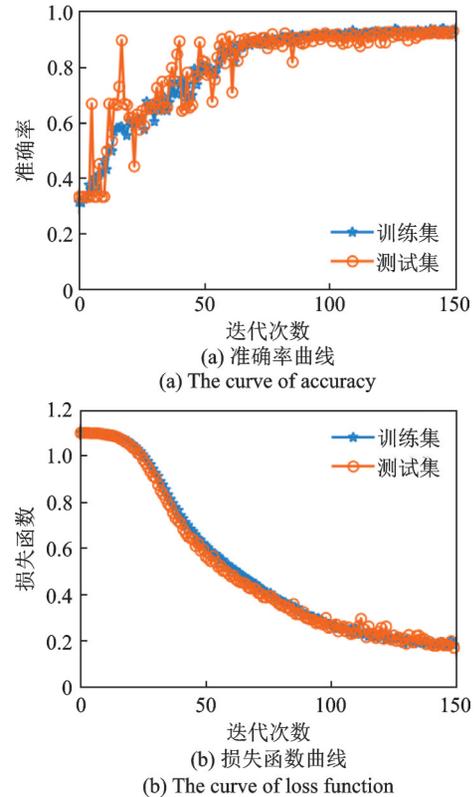


图5 DCNN模型诊断结果曲线
Fig. 5 Curve of diagnosis result of DCNN model

中未出现过拟合现象。从图5中看出,虽然DCNN故障诊断模型的训练集和测试集的准确率随着迭代次数的增加也在逐步提高,并且逐渐重合,但是最终的准确率只有93%左右,而且迭代过程中模型的准确率波动较大,不及DTPN模型稳定。同时,DCNN模型的损失函数曲线也在慢慢下降并有收敛的趋势,但最终的损失值仍然有0.2。综上,DTPN模型的识别准确率要高于DCNN模型,且诊断效果优于DCNN模型。

卷积层的隐式特征可以反映DTPN故障诊断模型的特征提取能力。为了展示其特征提取能力,本文运用t-SNE算法对卷积层输出的高维特征张量进行降维。降维后的各卷积层特征散点图如图6所示。由于特征散点图的横纵坐标无实际意义,这里分别用P1和P2表示。从图6(a)可以看出,经过第一层卷积层后,由于齿轮的健康状态和缺齿故障关联性比较强,两种齿轮状态样本混合在一起,而裂纹故障已经呈现出聚类的状态,只有个别点存在误判的现象。通过一个张量投影层后,从图6(b)第二层卷积层的特征散点图可以看出,这时的裂纹故障已经完全分离聚类,并且健康状态和缺齿故障也在逐渐分离。从图6(c)看出,再次经过张量投影层后,齿轮的健康状态和缺齿故障已经完全分离聚类。

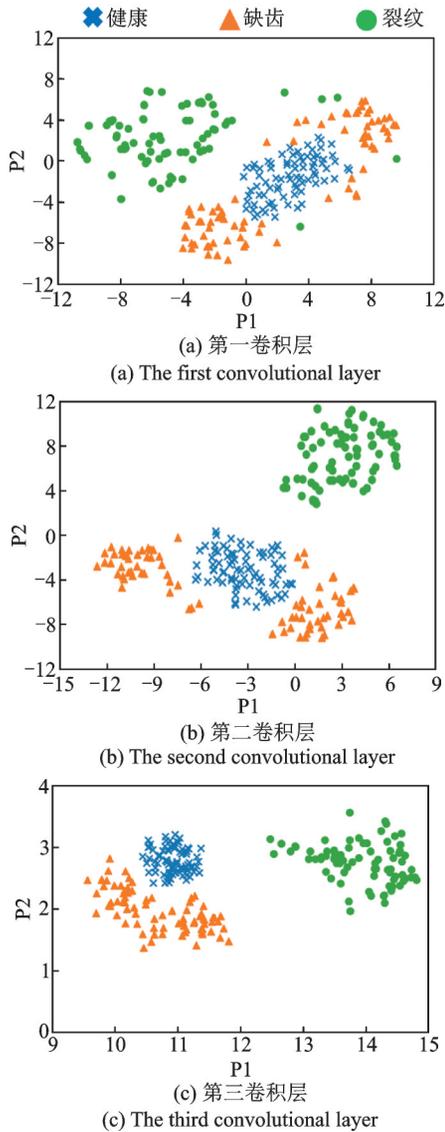


图 6 DTPN 特征散点图

Fig. 6 The feature scatter plot of DTPN model by t-SNE

同样,对 DCNN 故障诊断的各个卷积层也做了 t-SNE 可视化分析,如图 7 所示。从图 7(a)可以看出,经过第一卷积层后,与 DTPN 模型的第一卷积层的特征散点图类似,裂纹故障已经呈现出聚类的状态,但是齿轮健康状态和缺齿故障还未分离,出现重叠。经过第一次平均池化层后,从图 7(b)所示的第二卷积层特征散点图看出,虽然裂纹故障已经完全聚类,但是另外两种齿轮状态还是混合的状态,还没有出现分离聚类的趋势,但此时的 DTPN 模型的这两种齿轮状态已经开始各自聚类。从图 7(c)第三卷积层的特征散点图看出,经过第二次平均池化层后,齿轮的健康状态和缺齿故障开始慢慢分离聚类,但有些部分仍然混合在一起。从图 6 和 7 的两种模型的特征散点图可以看出,经过张量投影层后的卷积层的特征提取能力要比经过平均池化层的强,进一步说明了 DTPN 故障诊断模型

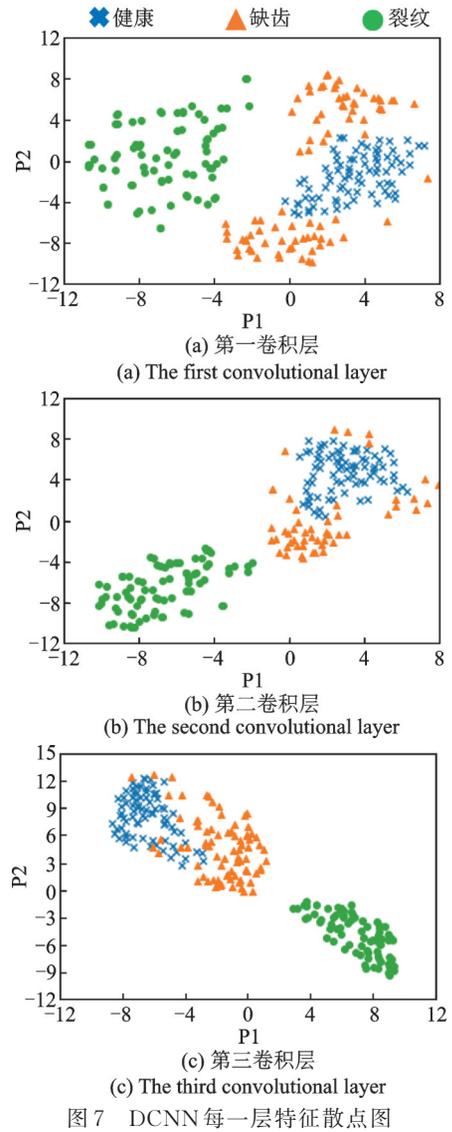


图 7 DCNN 每一层特征散点图

Fig. 7 The feature scatter plot of DCNN model by t-SNE

比 DCNN 故障诊断模型更好地保留了原始故障特征信息。如果只是单纯地通过准确率描述模型的好坏,无法看出具体预测类别与真实类别的关系,因此通过混淆矩阵来进一步分析 DTPN 故障诊断模型与 DCNN 故障诊断模型对不同类别数据的预测情况,如图 8 所示,横坐标表示预测标签,纵坐标表示真实标签。

从图 8(a)看出, DTPN 故障诊断模型的识别准确率为 100%, 而图 8(b)中, DCNN 故障诊断模型对健康和缺齿状态存在错误分类, 识别准确率只有 93% 左右, 低于 DTPN 模型。这更能说明 DTPN 故障诊断模型优于 DCNN 故障诊断模型。

由于深度张量投影网络可以通过修改张量投影层的输出维数来简化网络结构, 所以将张量投影层的输出维数减小, 然后把上述齿轮状态的高阶谱张量数据输入修改后的 DTPN 模型中进行实验结构分析, 并与修改前的 DTPN 模型进行对比, 其参数变化和识别结果如表 4 和 5 所示。

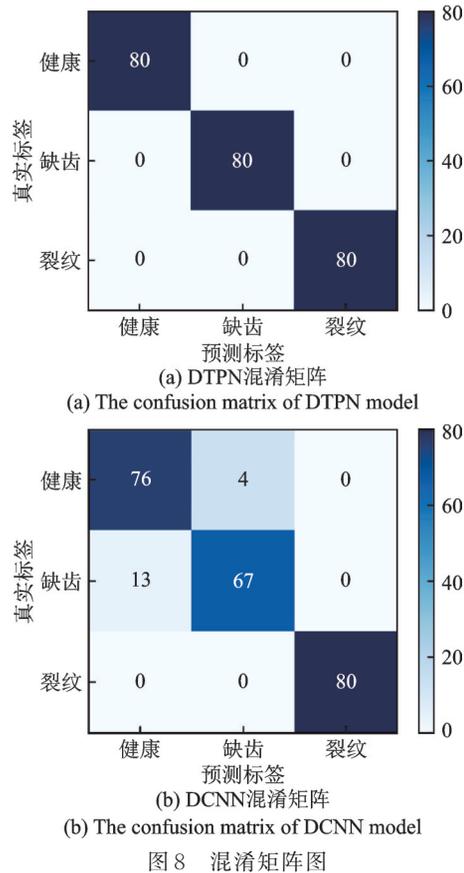


Fig. 8 Confusion matrix diagram of each model

表4 改变TPL维度后的DTPN模型各层参数

Tab. 4 The parameters of each layer of the DTPN model after changing the dimension of the TPL

结构	输出维度	参数	准确率
张量投影层1	$(n, 60, 60, 16)$	15360	100%
张量投影层2	$(n, 26, 26, 32)$	3120	
张量投影层3	$(n, 9, 9, 64)$	468	

表5 改变TPL维度前的DTPN模型各层参数

Tab. 5 The parameters of each layer of the DTPN model before changing the dimension of the TPL

结构	输出维度	参数	准确率
张量投影层1	$(n, 64, 64, 16)$	15360	100%
张量投影层2	$(n, 32, 32, 32)$	4096	
张量投影层3	$(n, 16, 16, 64)$	1024	

通过对比表4和5发现,将每个张量投影层的输出张量维度减小,模型参数也减少的同时,准确率不变。训练集和测试集的故障诊断结果曲线如图9所示,图中,训练集和测试集的准确率曲线都在逐渐上升至100%并保持不变。同时,训练集和测试集的损失函数曲线也在稳定下降,并逐渐收敛至0。

由于DTPN可以在保证故障识别准确率的前提下简化网络结构,而DCNN一般需要增加网络层数

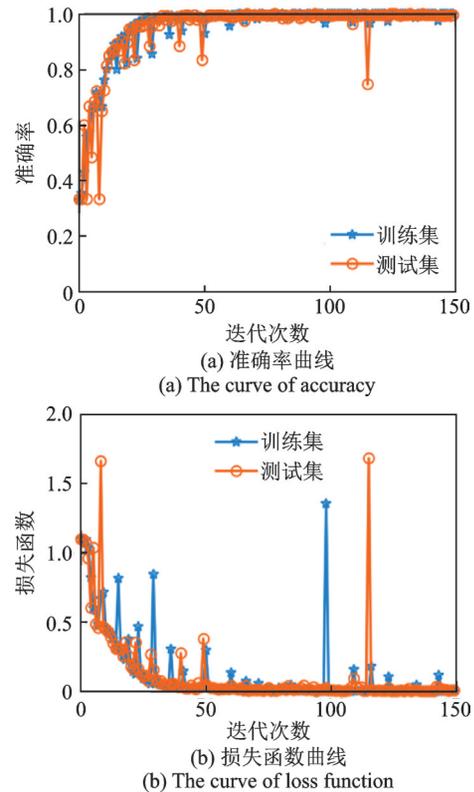


图9 修改TPL层维度后的DTPN诊断结果曲线

Fig. 9 DTPN diagnosis result curve after modifying the dimension of TPL layer

以提高准确率。对此,对比具有两层张量投影层的DTPN模型与分别具有两层、三层和四层池化层的DCNN模型,并保证经过最后一层张量投影层和池化层的输出维度一致,得到不同模型的准确率如表6所示。由表6可知,虽然增加池化层数提高了DCNN故障诊断模型准确率,但网络的复杂度也在随之提高,并且最终准确率不及DTPN故障诊断模型。

表6 不同模型准确率对比

Tab. 6 Comparison of accuracy rates of different models

模型	输出维度	准确率
两层张量投影层的DTPN	$(n, 8, 8, 32)$	100%
两层池化层的DCNN	$(n, 8, 8, 32)$	91%
三层池化层的DCNN	$(n, 8, 8, 32)$	92%
四层池化层的DCNN	$(n, 8, 8, 32)$	94%

4 结论

本文针对传统深度卷积神经网络的池化层容易丢失数据信息等问题,提出了一种深度张量投影网络的故障诊断模型,并对齿轮进行故障诊断。

(1)本文利用张量投影层取代卷积神经网络中的池化层,构建了深度张量投影网络模型。张量投影层不但可以对高阶张量数据进行降维,而且还能

保证不破坏张量数据结构,避免造成数据特征信息丢失。

(2)针对齿轮振动信号的非线性、非平稳和复杂等特性,通过运用高阶谱分析振动信号,在去除信号的高斯噪声的同时,还保留了其相位信息,捕获了非线性系统的相位耦合信息。

(3)结合高阶谱分析,将所提模型用于齿轮的故障诊断中,证明了DTPN故障诊断模型的有效性。并与DCNN故障诊断模型进行了对比。结果表明,DTPN故障诊断模型更好地保留了原始故障特征信息,泛化能力也优于DCNN故障诊断模型。

(4)所提出的张量投影层是一种维度可变的降维层,除了能够简化网络结构、降低网络的复杂度之外,还能保证模型的准确性和泛化能力。

参考文献:

- [1] Boureau Y L, Ponce J, Lecun Y. A Theoretical Analysis of Feature Pooling in Visual Recognition[C]// ICML 2010—Proceedings, 27th International Conference on Machine Learning, 2010:111-118.
- [2] 唐波,陈慎慎.基于深度卷积神经网络的轴承故障诊断方法[J].电子测量与仪器学报,2020,34(3):88-93.
Tang Bo, Chen Shenshen. Method of bearing fault diagnosis based on deep convolutional neural network[J]. Journal of Electronic Measurement and Instrumentation, 2020, 34(3): 88-93.
- [3] Wang J J, Ma Y L, Huang Z G, et al. Performance analysis and enhancement of deep convolutional neural network[J]. Business & Information Systems Engineering, 2019, 61(3): 311-326.
- [4] Xin Y, Li S M, Wang J R, et al. Intelligent fault diagnosis method for rotating machinery based on vibration signal analysis and hybrid multi-object deep CNN[J]. IET Science, Measurement & Technology, 2020, 14(4): 407-415.
- [5] 卞景艺,刘秀丽,徐小力,等.基于多尺度深度卷积神经网络的故障诊断方法[J].振动与冲击,2021,40(18):204-211.
Bian Jingyi, Liu Xiuli, Xu Xiaoli, et al. Fault diagnosis method based on a multi-scale deep convolutional neural network[J]. Journal of Vibration and Shock, 2021, 40(18): 204-211.
- [6] 李长文,程泽银,张小刚,等.基于深度残差网络的采煤机摇臂齿轮故障诊断[J].工矿自动化,2021,47(3):71-78.
Li Changwen, Cheng Zeyin, Zhang Xiaogang, et al. Fault diagnosis of shearer rocker gear based on deep residual network[J]. Journal of Mine Automation, 2021, 47(3): 71-78.
- [7] 牛乃平,张小刚,丁华.基于1D-CNN的采煤机摇臂齿轮故障诊断[J].煤矿机械,2020,41(11):148-150.
Niu Naiping, Zhang Xiaogang, Ding Hua. Fault diagnosis of shearer rocker arm gear based on 1D-CNN[J]. Coal Mine Machinery, 2020, 41(11): 148-150.
- [8] 胡葛庆,陈徽鹏,程哲,等.基于经验模态分解和深度卷积神经网络的行星齿轮箱故障诊断方法[J].机械工程学报,2019,55(7):9-18.
Hu Niaoqing, Chen Huipeng, Cheng Zhe, et al. Fault diagnosis for planetary gearbox based on EMD and deep convolutional neural networks[J]. Journal of Mechanical Engineering, 2019, 55(7): 9-18.
- [9] Liu C Y, Meerten Y, Declercq K, et al. Vibration-based gear continuous generating grinding fault classification and interpretation with deep convolutional neural network[J]. Journal of Manufacturing Processes, 2022, 79: 688-704.
- [10] Cheng Z, Hu N, Chen J, et al. Fault detection of planetary gearboxes based on deep convolutional neural network[C]// 2019 Prognostics and System Health Management Conference (PHM-Qingdao). 2019: 1-5.
- [11] Heydarzadeh M, Hedayati S, Nourani M, et al. Gear fault diagnosis using discrete wavelet transform and deep neural networks[C]// Proceedings of the IECON 2016-42nd Annual Conference of the IEEE Industrial Electronics Society. New York:IEEE.2016: 1494-1500.
- [12] Wu C Z, Jiang P C, Ding C, et al. Intelligent fault diagnosis of rotating machinery based on one-dimensional convolutional neural network[J]. Computers in Industry, 2019, 108: 53-61.
- [13] Xin Y, Li S M, Cheng C, et al. An intelligent fault diagnosis method of rotating machinery based on deep neural networks and time-frequency analysis[J]. Journal of Vibroengineering, 2018, 20(6): 2321-2335.
- [14] Wang R, Xu L, Liu F K. Bearing fault diagnosis based on improved VMD and DCNN[J]. Journal of Vibroengineering, 2020, 22(5): 1055-1068.
- [15] Li T F, Zhao Z B, Sun C, et al. WaveletKernelNet: an interpretable deep neural network for industrial intelligent diagnosis[J]. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2021, 52(4): 2302-2312.
- [16] Zhou D J, Yao Q B, Wu H, et al. Fault diagnosis of gas turbine based on partly interpretable convolutional neural networks[J]. Energy, 2020, 200: 117467.
- [17] Grezmaek J, Wang P, Sun C, et al. Explainable convolutional neural network for gearbox fault diagnosis[J]. Procedia CIRP, 2019, 80: 476-481.

- [18] Morimoto T, Huang S Y. TensorProjection Layer: a tensor-based dimensionality reduction method in CNN [J]. arXiv: 2004.04454.2020.
- [19] Jiang L L, Liu Y L, Li X J, et al. Using bispectral distribution as a feature for rotating machinery fault diagnosis[J]. Measurement, 2011, 44(7): 1284-1292.
- [20] Liu Y B, Yan K G, Zhou Y B, et al. Bispectral analysis for diagnosis of gear pitting fault [C] //2008 7th World Congress on Intelligent Control and Automation. 2008: 6847-6851.
- [21] Hasan M J, Sohaib M, Kim J M. A multitask-aided transfer learning-based diagnostic framework for bearings under inconsistent working conditions[J]. Sensors, 2020, 20(24): 7205.
- [22] Cao P, Zhang S L, Tang J. Preprocessing-free gear fault diagnosis using small datasets with deep convolutional neural network-based transfer learning[J]. IEEE Access, 2018, 6: 26241-26253.

Mechanical fault diagnosis method based on Deep TensorProjection Networks

HUANG Wen-jing, LI Zhi-nong, WANG Fa-lin, CHEN Liang-liang, LONG Sheng-rong

(Key Laboratory of Nondestructive Testing of Ministry of Education, Nanchang Hangkong University, Nanchang 330063, China)

Abstract: The shortcomings of fault diagnosis methods based on deep convolutional neural networks is that, tensor data is easily destroyed when reducing the dimension of high-order input tensors by pooling layers, which results in a loss of data information, and the relatively complex network structure. Therefore, a Deep TensorProjection Networks method is constructed via replacing the pooling layer in the traditional CNN by a TensorProjection Layer. The TensorProjection Layer reduces the dimensionality of input high-order tensor data without causing damage to the data, thus avoiding the impact of the loss of feature information, and greatly improving the recognition accuracy of the model. The dimensionality of the TensorProjection Layer used for dimensionality reduction is variable, thus simplifying the networks structures. Based on this, combined with the respective advantages of high-order spectrum and deep TensorProjection networks, a mechanical fault diagnosis method based on deep TensorProjection networks is proposed. In the proposed method, the feature of fault signal is extracted by high-order tensor spectrum, which is input into the constructed model for reducing high-order tensor dimensionality and identifying faults. The proposed method is applied to diagnose gearbox faults. Experimental results show that the proposed method can better retain the original fault information and effectively recognize the different types of faults. And the accuracy is better than traditional deep convolutional neural network fault diagnosis methods.

Key words: fault diagnosis; Deep TensorProjection Networks; high-order spectrum

作者简介: 黄文静(1998—),女,硕士研究生。E-mail: 1341065370@qq.com。

通讯作者: 李志农(1966—),男,教授,博士生导师。E-mail: lizhinong@tsinghua.org.cn。