

基于重构二维数组的 CNN 齿轮系统故障诊断方法

陈 琳¹, 马芸婷¹, 郭利宏¹, 魏 彬¹, 王文祺²

- (1. 内蒙古包钢钢联股份有限公司轨梁轧钢厂, 内蒙古 包头 014010;
2. 内蒙古包钢钢联股份有限公司稀土钢冷轧板材厂, 内蒙古 包头 014010)

摘 要: 面对海量的齿轮系统振动信号数据, 常规信号分析方法已难以精准提取故障特征, 也无法完成高效的故障分类工作。针对这一问题, 文章提出一种重构一维样本数据为二维数据的方法, 并将重构之后的二维数据输入到 CNN 网络中进行故障提取与故障分类, 解决端到端的故障诊断问题, 最终取得了 99.7% 的高准确率, 实验结果表明重构二维数据并输入到 CNN 网络的方法适用于齿轮系统的故障诊断。

关键词: 故障诊断; 数据重构; 深度学习; 卷积神经网络

中图分类号: TH132.41

文献标识码: B

文章编号: 1009-5438(2026)03-0070-06

Fault Diagnosis Method of CNN for Gear System Based on Reconstruction of Two-dimensional Array

Chen Lin¹, Ma Yunting¹, Guo Lihong¹, Wei Bin¹, Wang Wenqi²

- (1. Rail and Beam Rolling Plant of Inner Mongolia Baotou Steel Union Co., Ltd.,
Baotou 014010, Inner Mongolia Autonomous Region, China;
2. Cold Rolled Rare Earth Steel Plate Plant of Inner Mongolia Baotou Steel Union Co., Ltd.,
Baotou 014010, Inner Mongolia Autonomous Region, China)

Abstract: In the face of a bulk of vibration signal data of gear system, fault features have been difficult to be accurately extracted and efficient fault classification could not be completed with conventional signal analysis methods. In the paper, a method of reconstruction from one-dimensional sample data to two-dimensional data is proposed aiming at the problems and the reconstructed two-dimensional data are input into the convolutional neural network (CNN) for fault feature extraction and fault classification to solve the problem of end-to-end fault diagnosis. Eventually, the high accuracy rate of 99.7% is achieved. The experimental results showed that the method of reconstructing two-dimensional data and inputting them into the CNN was suitable for fault diagnosis of gear system.

Key words: fault diagnosis; data reconstruction; deep learning; CNN

在多个工业制造域以及国防交通设施中, 发挥着关键性作用的齿轮系统其旋转机械特性表现得尤为显著。齿轮系统运行状态监测技术与故障辨识方

法的重要性已不容忽视。多重激励源存在于结构复杂的齿轮系统中, 导致所产生振动信号呈现出非平稳特性与显著的非线性特征。振动信号分析成为故

障监测诊断研究的切入点。

企业长期存在着对于齿轮系统故障监测与诊断技术的投入明显不足的现象。在多数情况下,专家经验积累以及温度参数等常规监测手段仍是多数情况下故障诊断工作的主要方法。在这些传统方法中滞后性缺陷表现明显,有效的故障预警功能难以实现。学者们逐渐关注能够反映系统冲击特征的振动信号^[1-5]。基于振动监测方法的特征提取与故障诊断研究正在成为新的趋势。

频谱分析基础上的实时状态监测技术,构成了当前主流技术方案。当面临多故障并发、频率耦合及重要频率模糊等复杂工况时,当前主流技术方案往往无法解决问题。人工参与特征提取环节,采用传统信号处理方法,如小波包变换(WPT)^[6]、经验模态分解(EMD)^[7]以及提升小波变换(LWT)^[8]等的必要环节。虽然取得了一定成效,但实时性难以保证。人工神经网络(ANN)^[9]、支持向量机(SVM)^[10]等模式识别方法,仅适用于特定数据集,人工提取的特征参数导致诊断方法的普适性受限^[11]。

齿轮系统的连续工作、测点布置密度的提升以及采样频率的不断提高使齿轮系统从投运至报废的全生命周期数据采集形成了海量数据集^[12]。步入“大数据”时代的是齿轮系统振动监测领域,如何实现从海量振动信号中提取各类故障特征,并快速准确完成自动化分类识别成为亟待解决的关键问题,深度学习技术为此提供了新的解决途径^[13]。

得益于计算设备的革新,其在各个领域展现出强大应用潜力,作为机器学习重要分支的深度学习技术也在故障诊断领域凸显出来。近年来,卷积神经网络(CNN)通过优化调整卷积核参数实现对原始图像分类识别的方法,使得CNN具有自动特征学习能力的这一特性在多个领域得以成功应用^[14-18]。

已有学者尝试将CNN引入齿轮系统故障诊断研究:采用振动信号特征统计图作为网络输入实现故障分类^[19]。也有一些学者通过多小波变换和时频变换将处理后的二维图像数据输入CNN进行诊断^[20],但这些方法均需借助其他信号处理手段进行数据转换,未能充分发挥深度学习直接从原始信号中自动学习特征的优势。

本文提出一种基于重构二维数组,图像化处理齿轮传动系统故障辨识的框架。此框架的关键步骤是将振动传感信号自动重组为二维图像数据,直接

输入至卷积神经网络架构中,完成端到端诊断任务。该框架在诊断效能上,显著优于传统特征选择结合模式识别的方法。同步展开的还有针对卷积核尺寸、网络层级组合等核心参数的系统性探究工作。

1 卷积神经网络

Yann Le Cun教授1989年在纽约大学首次提出了CNN结构,如图1所示。本质上可归类为前馈式神经网络的特殊变型。核心机制在于设计多个具备特征提取功能的滤波器:原始输入数据通过这些滤波器实施逐层卷积运算与池化操作,数据深层隐藏的拓扑特征被逐步抽离出来。在网络深度不断增加的过程中,所获取的特征呈现出更加抽象化的趋势——最终形成对输入数据平移、旋转及尺度变化保持稳定的表征系统。该网络的两大核心特征在于稀疏连接特性与权重共享机制。前者通过建立层间非全连接的拓扑关联,能够有效减少训练参数量;后者则显著抑制算法过拟合现象值得关注。子采样技术的引入使得数据本身的统计特性得到充分利用应当注意。典型的CNN处理流程表现为交替进行的卷积与子采样操作序列:接近输出层时转换为常规的多层神经网络结构,进行最终处理尤为关键。区别于其他神经网络的显著特点,体现在可直接接收二维特征图或原始数据图,作为输入而无需预先进行矢量化转换。

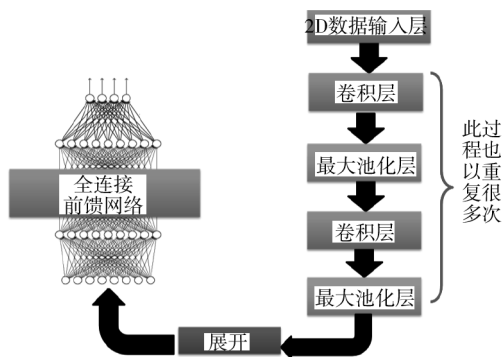


图1 CNN结构图

实现卷积层的过程中,前一层的特征图会与当前层的卷积核完成数学卷积运算——配合激活函数的非线性变换,共同生成新的输出特征图。多层级的特征提取过程,通过这种循环迭代的方式持续进行,其数学模型可表述为公式(1):

$$x_i^l = f(\sum_{i \in M_j} x_i^{l-1} \times k_{ij}^l + b_j^l) \quad (1)$$

式中: M_j 为输入特征图; l 为第 l 层网络; k 为卷积核; b 为网络偏置; x_j^l 为 l 层的输出; x_j^{l-1} 为 l 层的输入。

尺度压缩映射乃子采样层主要功能体现所在, 上层数据图经由其处理可获得显著效果: 数据维度降低现象明显发生, 尺度不变性特征表示得以形成。模型过拟合现象的防止作用亦被该层同时具备。这些优势具有不可忽视性。该层神经元计算方法可表述如下:

$$x_j^l = f(\beta_j^l \text{down}(x_j^{l-1}) + b_j^l) \quad (2)$$

式中: $\text{down}(\cdot)$ 为子采样函数; β 为网络乘性偏置。

卷积神经网络输出层的处理方式中普遍存在下列现象: 前一层级联特征图通过全连接线性变换实现转换操作。Softmax 分类器在该环节的应用频率较高实例可见, 多类别分类问题的解决有效性正是该方法优势所在。模型数学表达形式如式(3)所示:

$$O = f(b_o + \omega_o f_v) \quad (3)$$

式中: f_v 为特征向量; b_o 为偏差向量; ω_o 为权值矩阵。

卷积神经网络显著的特性之一是局部特征的提取能力化表现。通过层级递进的特征组合与抽象化处理过程, 高层表征得以形成。该机制在设备故障诊断任务与模式识别领域效果较优。

2 CNN 网络模型组成

本研究采用的 CNN 架构方案由以下组件构成: 输入层单元、两组交替排列的卷积-池化层组合体、全连接层模块以及分类器单元。

2.1 卷积层

作为核心构件的卷积层承担着特征提取的功能。特定卷积核对应每一个特征映射图, 理论上可具有任意的尺寸, 实际应用中, 多采用小型正方形形态。常见尺寸包括 3×3 、 5×5 和 7×7 三种规格。前一层输入数据经由权重参数实施卷积运算后, 生成的特征输出将作为后续层的输入源数据。与传统滤波器需要人工设定参数的方式不同, CNN 中的权值矩阵与偏置项, 通过反向传播算法完成训练流程实现。

2.2 池化层

池化层的主要功能是空间维度的大幅度压缩, 该操作可使权重参数量下降 75%, 计算资源消耗随之降低明显体现。其次池化层还可以抑制过拟合, 使得训练集与测试集间的准确度差异呈现缩小趋势。通常在卷积层之后会配置池化层, 但其并非必然存在的连接, 其具体设置需依据实验情况确定选

择。最大池化法在各种池化操作中性能特征值最优, 表现优于平均池化和 L-2 范式池化。基于此因素下, 本研究采用 2×2 窗口尺寸配合步长 2 的最大池化策略方案设计, 针对输入数据的每个深度切片独立执行空间尺寸变换, 同时保持深度不变性特征, 仅剩余四分之一激活信息。

输入的数据尺寸 B_i, C_i, D_i , 池化尺寸为 $F \times F$, 步长为 S 的池化操作, 可以计算得出的输出尺寸分别为:

$$B_{i+1} = \frac{B_i - F}{S} + 1 \quad (4)$$

$$C_{i+1} = \frac{C_i - F}{S} + 1 \quad (5)$$

$$D_{i+1} = D_i \quad (6)$$

2.3 全连接层

经过两层卷积-池化操作模块的串联处理, 全连接网络结构单元被后续接入。特征向量与类别标记间的关联性建立机制构成了该层级的核心功能实现。过拟合现象在全连接特性作用下凸显出来, 结合 Dropout 的随机梯度下降优化方法, 能显著降低该问题的出现频度。权值更新过程采用的数学表述形式如下:

$$v_{i+1} = 0.9 \times v_i - 0.01 \times \beta \times w_i - \beta \times \left(\frac{\partial L}{\partial w} \Big|_{D_i} w_i \right) \quad (7)$$

$$w_{i+1} = w_i + v_{i+1} \quad (8)$$

式中: i 为迭代次数; β 为学习率; D_i 代表训练样本集; v 代表动量参数。

2.4 分类器

本研究模型架构中 Softmax 回归层作为最终分类单元被配置使用。全连接层处理后产生的特征向量 V_i 作为输入数据流向该层。经过 Softmax 函数运算处理后的数学表达呈现如下形式:

$$\text{softmax}(v_i) = \frac{e^{V_i}}{\sum_j e_j^{V_i}} \quad (9)$$

从公式(9)可以发现, 经过 Softmax 标准化处理的神经网络输出向量, 已经转变为具备概率分布特征的数学对象。实际测试数据证实, 该向量能够有效反映输入样本属于不同类别的可能性大小。本质上完成了从特征空间到概率空间的非线性映射转换过程。实例表明, 整个神经网络系统形成了计算输入样本类别概率的封闭式运算体系。

预测结果与真实标签之间的差异程度, 通过特定设计的损失函数完成量化评估。

3 实验信号分析

3.1 数据采集

为验证本文所提方法的可行性,将直齿圆柱齿轮的五种故障状态作为验证对象。在动力传动故障模拟实验台上进行实验,实验台结构示意图如图 2 所示。应用加速度传感器,共计 3 个,将其分别布置于垂直向、径向与轴向三个方位上,并采集振动信号数据。采样频率设定为 5 060 Hz,持续采样 1 s,获得 5 060 个数据。常见的齿轮状态为断齿、正常、磨损、齿根裂纹和缺齿。对这五种状态分别采样,每种情况采 300 组,5 种情况共获得 1 500 组数据,构成 1 500 行 $\times$ 5 060 列的数据矩阵。

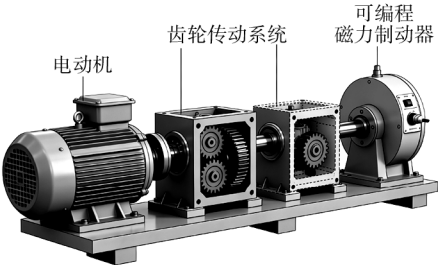


图 2 传动故障模拟实验台结构图

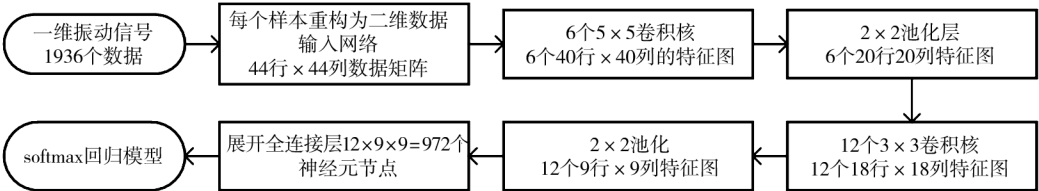


图 3 CNN 结构图

3.4 CNN 参数设计

本文的 CNN 模型由 6 层与一个分类器组成,其中有一个输入层,两组交替出现的卷积层和池化层,以及全连接层和一个分类器。

3.4.1 第一组卷积层设置

采用 5 行 $\times$ 5 列规格的卷积核进行特征提取操作,其输出的感受野尺寸与之相同。共计部署 6 个该规格的卷积核组,用于从重构后的图像样本中捕获 6 种不同特征类型。生成 6 张特征图谱作为输出结果,这些图谱将作为后续池化层的输入数据。

3.4.2 第一组池化层设置

选择最大池化方法实施降采样处理,设定 2 行 $\times$ 2 列的采样窗口尺寸配合步长参数为 2 个的

3.2 数据重构

为保障数据样本的多样性与完整性要求,实施了一维数据向二维数据的转换操作。通过实验过程,采集到五类故障模式的数据各 300 组,总计 1 500 组实验数据。基于主成分分析原理,各组均选取贡献率较高的前 1 936 个特征点进行重构,形成 44 行 $\times$ 44 列规格的二维图像矩阵。最终获得的数据集维度表现为 44 行 $\times$ 44 列 $\times$ 1 500 组的三维结构形式。

训练集与测试集的划分,遵循常规比例原则。从总体数据中,随机抽取 1 125 组作为训练样本集(约占 75%),其维度构成为 44 行 $\times$ 44 列 $\times$ 1 125 组;剩余的 375 组数据(约占 25%)构成测试样本集,维度表现为 44 行 $\times$ 44 列 $\times$ 375 组。完成重组后的数据集被导入 CNN 网络架构实施特征提取与分类运算任务。

3.3 卷积神经网络构建

依据重构后的实验数据,规定如下实验方案: CNN 的各层所得特征图个数和特征图尺寸和全连接层的神经元节点数如图 3 所示。

操作方式。对 6 张输入特征图谱分别执行独立处理,在维持网络深度不变的前提下完成尺寸调整。最终输出 6 张 20 行 $\times$ 20 列规格的特征图谱。

3.4.3 第二组卷积层设置

使用 3 行 $\times$ 3 列规格的卷积核开展二次特征提取工作,其感受野范围相应调整为 3 行 $\times$ 3 列尺度。配置 12 个该类型卷积核实现更精细的特征捕获,从样本数据中析出 12 类差异化特征。经过运算产生 12 张 18 行 $\times$ 18 列尺度的特征图谱并输送至后续池化处理环节。

3.4.4 第二组池化层参数设置

保持 2 行 $\times$ 2 列采样窗口与步长参数为 2 的设置方案。对 12 张输入特征图实施独立降采样操作,

网络深度维度未发生改变,而空间维度产生变化。最终获得 12 张 9 行×9 列规格的特征图谱输出结果。

3.4.5 全连接层参数设置

经由前述两组卷积-池化层的多级特征提取与降维处理后,将生成的 12 张特征图展开为 972 个数的一维向量输入全连接层。进行模式学习分类器分类,分类器采用 Softmax 回归模型完成最终类别判定其具体计算过程:权重矩阵化参数与偏置项数值分别予以表示,构成 CNN 架构最终输出表征量。

$$O = \begin{bmatrix} P(y = 1 | x; W_1, b_1) \\ P(y = 2 | x; W_2, b_2) \\ P(y = 3 | x; W_3, b_3) \\ P(y = 4 | x; W_4, b_4) \\ P(y = 5 | x; W_5, b_5) \end{bmatrix}$$
$$= \frac{1}{\sum_{j=1}^k \exp(W_j x + b_j)} \begin{bmatrix} \exp(W_1 x + b_1) \\ \exp(W_2 x + b_2) \\ \exp(W_3 x + b_3) \end{bmatrix} \quad (10)$$

式中:W 为权重矩阵;b 为偏置值;代表卷积神经网络的最终结果。

3.5 结果分析

结果分析性讨论经重构处理的样本数据集规模达 1 500 组×44 行×44 列,随机抽取 1 125 组样本数据作为训练集构建用,剩余 375 组则作为测试集使用。测试集数据输入后,每个样本将输出五类状态概率分布值,选取概率峰值所对应状态作为预测标签值。预测标签与真实标签一致性检验中,匹配成功则判定为预测正确,失配则归为预测错误。测试集总体错误样本占比计算后,所得百分比即为网络模型最终准确率指标值。迭代次数固定为 100 次,学习率参数设定为 0.1 s。权重更新后每次从测试集中随机选取 15 组数据进行验证性测试。网络系统总运行时间达 300 s,其中模型训练过程耗时 265 s,共完成 100 次迭代操作,单次迭代平均时耗 2.65 s。数据预处理阶段包括导入与重构操作共计消耗 35 s 时间资源。分类准确率达到 99.7%,表明了端到端故障诊断范式得以有效实现。

3.6 模型对比

本方法与传统故障诊断方法差异性特征表现,对原始 1 500 组×1 936 个规模数据集实施 EMD 算法与 ELMD 算法双通道信号分解处理实验。EMD 算法分解产生 7 个本征模态分量,ELMD 算法分解

生成 7 个乘积函数成分。上述特征分量分别作为输入特征导入浅层 BP 神经网络架构中,训练阶段采用 1 125 组样本数据,测试阶段使用 375 组样本数据进行分类精度验证,完整流程从数据采集开始至故障分类结果输出终止(见表 1)。

表 1 运行结果对比表

网络	训练样本 /个	测试样本 /个	正确识别 /个	识别精度 /%
2D-CNN	1 125	375	374	99.7
ELMD-BP	1 125	375	354	94.4
EMD-BP	1 125	375	350	93.3

由表 1 可以得出:当采用相同数据集进行实验时,经过重构处理的二维数组输入 CNN 网络后,相较于常规故障诊断手段展现出了显著优越的性能。

4 结论

本研究构建了一种创新性的齿轮系统复合诊断框架,该框架通过将一维振动信号转换为二维数组结构,并输入至 CNN 网络实现特征自动学习与分类诊断。并应用传动故障模拟实验台获得五种故障状态的振动数据,输入网络进行模型验证。

由此可见,经过重构处理的二维数组数据,在卷积神经网络中产生了相似的处理效果。成熟的 2D-CNN 网络架构,应用于一维振动信号分析领域展现出了令人满意的适用性。

训练完成的卷积神经网络模型,能够直接接收原始振动信号作为输入源。预处理环节、特征提取阶段以及故障分类步骤被网络自主完成这一事实,表明端到端的快速诊断目标可实现。齿轮系统故障可应用此模式,此模式下不仅可以获得较高诊断准确率,同时保持了较短的计算耗时。

参 考 文 献

[1] 王子顺,汪韵朗,赵东辉,等. 齿轮箱行星轮系故障诊断与原因分析[J]. 机械传动,2024,48(9):128-138.

[2] 李俊逸,尧远,刘明浩. 基于振动信号最优特征提取算法的风力机齿轮箱 SVM 故障诊断[J]. 太阳能学报,2024,45(7):626-633.

[3] 李宗强,戴永成. 变速箱故障诊断方法及其在维修中的应用[J]. 汽车知识,2024,24(6):149-151.

- [4] 车畅,李明辉,马晨佩,等.包装机深沟球轴承多模态智能诊断模型与自适应参数优化研究[J].包装工程,2025,46(19):247-257.
- [5] 许永靖,尹建超,陈祥,等.基于太阳轮局部故障啮合周期特性的故障诊断指标[J].天津理工大学学报,2025,41(2):30-36.
- [6] 黄辉,张冠勇.小波分析在齿轮箱故障诊断中的应用研究[J].建筑机械,2024(10):165-168.
- [7] 周继忠,谢云.基于 EMD 分解和奇异值差分谱的重型柴油车内燃机接地信号检测[J].内燃机与配件,2025(19):64-66.
- [8] 李彩霞. LWT-FLP 去噪算法在 FBG 传感器信号处理中的应用[J].石河子科技,2023(1):62-64.
- [9] 苏莫涵.基于多源传感器的滚动轴承故障诊断[J].信息记录材料,2024,25(12):73-76.
- [10] 施庚伟,柯海森,朱勃旬,等.基于改进 SVM 和加权多特征融合的萃取分液异常检测方法[J].现代电子技术,2025,48(21):105-112.
- [11] Surajet. Khonjun, Rapeepan Pitakaso, Thanat-kij. Srichok, et al. Advancing biomedical waste classification through a hybrid ensemble of deep Learning, reinforcement Learning, and differential evolution algorithms [J]. Waste Management, 2026, 209: 115210-115210.
- [12] 刘泽,鲁杰,张松,等.基于时频分析与卷积神经网络的滚动轴承故障诊断[J].机械管理开发, 2025, 40(10): 82-85.
- [13] 王凡.融合 CNN 和 Transformer 的高分辨率遥感影像滑坡检测方法[J].遥感信息, 2025, 40(5): 158-165.
- [14] 许鑫辉,杜强,柯丽,等.基于卷积自编码器模型的脑血流阻抗图像重建方法研究[J].医疗卫生装备, 2025, 46(10): 1-8.
- [15] 程阳,黄锦娜,朱家兵,等.融合 U-Net 和 Transformer 的医学图像分割算法[J].现代信息科技, 2025, 9(19): 62-69.
- [16] 吴胜利,郑子润,邢文婷,等.基于多信息融合的 INFO-VMD-CNN 的齿轮箱故障诊断方法[J].振动与冲击, 2025, 44(13): 309-316.
- [17] 张发振,张清华,秦宾宾,等.基于 CNN-Transformer 网络融合时频域的滚动轴承剩余使用寿命预测[J].机床与液压, 2025, 53(14): 7-14.
- [18] 闫国华,胡以怀.以信号图像表示的基于 CNN 的永磁同步电机故障诊断方法[J].船舶工程, 2025, 47(1): 114-123.
- [19] 赵师兵,张志明.基于时域信号特征和卷积神经网络的模拟电路故障诊断算法[J].计算机应用, 2022, 42(S2): 320-326.
- [20] JING L, ZHAO M, LI P, et al. A convolutional neural network based feature learning and fault diagnosis method for the monitoring of gearbox [J]. Measurement, 2021, 111: 1-10.