

一种用于轴承故障诊断的 二维卷积神经网络优化方法

肖雄, 王健翔*, 张勇军, 郭强, 宗胜悦

(北京科技大学工程技术研究院, 北京市 海淀区 100083)

A Two-dimensional Convolutional Neural Network Optimization Method for Bearing Fault Diagnosis

XIAO Xiong, WANG Jianxiang*, ZHANG Yongjun, GUO Qiang, ZONG Shengyue

(Institute of Engineering Technology, University of Science and Technology Beijing, Haidian District, Beijing 100083, China)

ABSTRACT: Intelligent bearing fault diagnosis is a hot research field of mechanical big data condition monitoring. The traditional data-driven fault diagnosis method is extremely time-consuming and requires high expert experience for signal extraction-based feature extraction. In order to eliminate the pre-defined effects of parameters and improve the recognition rate while improving the feature extraction, this paper proposed a two-dimensional convolutional neural network optimization method for bearing fault diagnosis based on the research of one-dimensional convolution neural network fault diagnosis method. This method introduced a new data preprocessing method, which converted the original time domain signal data into a two-dimensional gray image to extract the transformed image features, eliminated the influence of manual features, and collects the experimental errors before verifying the classification. The data set adds noise reduction processing and optimizes the parameter adaptive learning rate for the convolutional neural network gradient descent algorithm. The simulation and experimental results show that the proposed two-dimensional optimized convolutional neural network fault diagnosis method is based on the signal-picture conversion format under 64×64 . The AMSGrad algorithm can improve the accuracy of the fault prediction model to 98%, and the training speed is faster and higher. Classification accuracy and anti-noise performance can achieve a recognition accuracy of less than 5% loss in the actual speed range.

KEY WORDS: deep learning; fault diagnosis; signal conversion; convolutional neural network

摘要: 轴承智能故障诊断是机械大数据状态监测的热门研究领域, 传统的数据驱动故障诊断方法对基于信号处理的特征提取环节极为耗时, 并且对专家经验要求高, 为消除其带来的参数预定义影响, 实现快速特征提取的同时提高识别率, 在研究一维卷积神经网络故障诊断方法的基础上, 提出一种用于轴承故障诊断的二维卷积神经网络优化方法。该方法引入了一种新的数据预处理方式, 将原始时域信号数据转换成二维灰度图像来提取转换后的图像特征, 消除手工特征的影响; 同时, 在验证分类前对实验采集故障数据集添加了降噪处理, 并对卷积神经网络梯度下降算法进行参数自适应学习率优化。仿真与实验结果表明, 所提出的二维优化卷积神经网络故障诊断方法在选取 64×64 的信号-图片转换格式下, AMSGrad 算法能将故障预测模型的准确度提升至 98%, 训练速度更快, 同时具有更高的分类准确性和抗噪性能, 使其在实际转速范围内能达到损失小于 5% 的识别准确率。

关键词: 深度学习; 故障诊断; 信号转换; 卷积神经网络

0 引言

轴承是旋转机械中最关键的零件之一, 近年来电机轴承智能故障诊断一直是研究的热点^[1], 新出现故障的早期检测对于当前的复杂系统十分重要, 即可节省时间和成本, 又利于采取必要措施来避免危险情况。目前, 故障诊断的方法可以分为基于模型^[2-4]、基于信号^[5-6]、基于知识和混合-主动方法^[7], 其中基于知识的方法也被称为数据驱动方法, 需要大量的历史数据来建立无先验已知模型或信号模式的系统故障模型, 非常适用于难以建立精确模型或提取信号特征的复杂系统。随着智能制造水平的提高, 企业数据的收集速率与数量级均得到了广泛

基金项目: 国家自然科学基金项目(61873025); 中国博士后科学基金面上资助项目(2018M631340); 教育部中央高校基础科研费项目(FRF-TP-17-045A1)。

Project Supported by National Natural Science Foundation of China (61873025); China Postdoctoral Science Foundation Funded Project (2018M631340); Fundamental Research Funds for the Central Universities (FRF-TP-17-045A1).

提升, 机电装备的状态监测迈入“大数据”时代^[8], 充分利用大量的生产数据^[9-10], 为数据驱动故障诊断方法提供了新的机会。

处理数据驱动故障诊断中数据的主流方法之一便是机器学习, 将支持向量机(support vector machine, SVM)应用于故障诊断的最初尝试是在 20 世纪 90 年代后期^[11], 随着人工智能技术的发展, SVM 的应用也在不断更新。文献[12]利用核主元分析法(kernel principal component analysis, KPCA)与 SVM 对断路器进行故障诊断, 通过分离正常数据样本空间与故障数据样本空间以提高诊断模型的抗干扰性能; 文献[13]成功利用局部图嵌入加权罚支持向量机(local graph embedding weighted-penalty SVM, LGEWP-SVM)对模拟电路进行故障诊断; 文献[14]利用局部均值分解, 针对不同工况下出现的多种变化情况将采集到的分合闸振动信号自适应分解, 实现断路器工作模式的识别。另一种较完善的主流数据驱动故障诊断方法则是人工神经网络(artificial neural network, ANN), 但由于存在易陷入局部最小值、收敛速度慢等问题, 之后研究人员不断地改进人工神经网络以应用于故障诊断中。文献[15]将经过改进的神经网络应用于电机定子的故障诊断, 并结合捕食搜索策略优化的遗传算法来选择神经网络训练的初始权值和阈值, 提高了反向传播(back propagation, BP)神经网络训练的速度和故障识别准确率; 相对于传统神经网络的缺陷, 文献[16]通过引入概率神经网络(probabilistic neural network, PNN)对高压断路器进行故障诊断, 提高了高压断路器的故障诊断精确率。然而, 目前所用的机器学习与普通的人工神经网络方法不能产生原始数据的判别特征, 并且与信号特征提取过程相结合。而特征提取过程本身是一项复杂的工作, 此过程对最终结果有很大的影响, 给诊断系统的实时性与设计成本均带来一定负担。

随着机器学习的发展, 深度学习(deep learning, DL)已成为克服上述缺陷的有效方法。深度学习可自动学习原始数据的抽象表示特征, 这可避免工程师设计的手工特征, 并且与传统的机器学习方法相比均取得了良好的效果。最常见的数据类型是时域信号, 处理一维信号的多种深度学习方法已经在实时电机故障诊断得到应用。文献[17]基于局部线性嵌入定义一个新的综合特征指标来量化滚动轴承性能, 利用遗传算法(genetic algorithm, GA)对深度置信网络的关键参数进行优化, 然而遗传算法需要

很多参数, 如交叉率和变异率, 并且这些参数的选择严重影响解的品质, 而目前这些参数的选择大部分是依靠经验, 加之未能及时利用网络的反馈信息, 故算法的搜索速度比较慢, 对于解更精确的需求时尤为明显。文献[18]使用 3 层稀疏自编码器来提取原始数据的特征, 并应用最大平均差异项来最小化训练数据和测试数据之间的差异惩罚; 文献[19]利用堆叠降噪自编码器对输入的信号进行数据重构, 得到更鲁棒性的特征表示, 易于进行故障分类, 然而自编码器在生成模型过程中会出现乱码与噪声, 对故障分类的结果会有一定的影响。文献[20]利用一维卷积神经网络将电机故障检测的特征提取和分类阶段融合为一个整体的学习过程。这几种深度学习方法已经应用于故障诊断, 这些端对端的方法省去了人为特征的提取, 但都是一维时间序列信号, 特征提取容易出现特征丢失现象, 然而目前主流的二维卷积神经网络结构并不直接适用于一维振动信号, 导致需加深常用的一维卷积神经网络深度来获取较大的感受野, 从而抑制过拟合, 这从一定程度上增加了设计的难度。本文引入了一种新的数据预处理方法, 将原始时域信号数据转换成二维灰度图像, 没有任何预定义的参数, 这可尽可能地消除专家经验, 并可使用在图像识别应用广泛的卷积神经网络, 如 LeNet-5。

本文首先介绍传统一维信号的故障诊断方法, 然后详细阐述基于二维信号的故障诊断优化方法, 包括信号-图像数据预处理、卷积神经网络结构设计; 梯度下降算法动态自适应学习率优化, 验证改进后模型的抗噪声能力; 最后采集实验数据以验证其有效性。

1 基于一维信号的卷积神经网络故障诊断方法

在传统的驱动数据故障诊断方法中, 数据预处理方法至关重要, 因为数据驱动方法不能直接处理原始信号。数据预处理方法的主要功能之一是从大量的历史数据中提取原始信号的特征。但是, 特征提取依赖于技术人员的经验与技术, 提取正确的特征是一项复杂的工作, 特征值的准确性影响到最终的分类结果。由于原始的电机振动信号是一维的时序信号, 常用的预处理方法包括主成分分析(principal component analysis, PCA)、偏最小二乘法(partial least square, PLS)和独立分量相关(independent component analysis, ICA), 这些方法常用于 SVM、

ANN 诊断方法的预处理; 针对于卷积神经网络诊断方法的预处理包括归一化处理、数据集增强技术。

由于处理对象是一维时域信号, 在卷积神经网络的选取上, 选择较为简单的一维卷积神经网络, 该卷积神经网络常用于语音识别领域(voice recognition)。卷积神经网络是多级神经网络, 包含滤波级与分类级。其中, 滤波级用来提取输入信号的特征, 分类级对学习到的特征进行分类, 两级网络参数是共同训练得到的。滤波级包含卷积层、池化层与激活层 3 个基本单元, 而分类级由全连接层组成。常用的一维卷积神经网络结构包含 2 个卷积层、2 个池化层、1 个全连接隐含层及 1 个 Softmax 层。所选取的一维卷积神经网络如表 1 所示。

表 1 一维卷积神经网络结构参数

Tab. 1 1-D convolutional neural network structure parameters

编号	网络层	卷积核大小	卷积核数目
C1	卷积层 1	5×1	32
S2	最大池化层 1	2×1	32
C3	卷积层 2	5×1	64
S4	最大池化层 2	2×1	64
L5	输出层	120	1
L6	Softmax	10	1

目前, 常用的二维卷积神经网络结构均含有堆叠式的 3×3 的卷积核。这样既可加深网络深度, 也可实现以较少的参数获取较大的感受野, 从而抑制过拟合。然而, 对于一维振动信号, 2 层 3×1 卷积的结构, 以 6 个权值为代价, 仅仅获取了 5×1 的感受野, 反而将上述优势变成了劣势, 因此视觉领域的网络结构不适用于轴承故障诊断领域。故本文通过对一维振动信号进行图像可视化处理, 可利用卷积神经网络在图像识别的强大特点与优势。

2 基于二维信号的卷积神经网络故障诊断方法

在上述方法中, 由于一维时间序列信号存在数据量不足的情况, 在训练过程中容易出现过拟合的现象; 同时常用的特征提取方法容易出现特征丢失的现象, 本文结合信号-图像处理技术提出一种二维卷积神经网络优化方法以最大可能地保留原信号的特征。

本文提出的数据预处理方法是将原始的一维时域信号转换为二维的灰度图, 在大量图片数据集的基础上, 可使用在图像处理领域广泛应用的二维卷积神经网络 LeNet-5。

2.1 信号-图像数据预处理

灰度图的每个像素点包含 0~255 整数值, 表示为图像每个像素点的灰度值, 信号-图像数据预处理方法如图 1 所示, 原为一维时间序列的离散信号, 其中信号中包含 i 个信号点 $p(i)$, 信号点最大值与最小值分别为 $p_{\max}(i)$ 、 $p_{\min}(i)$ 。首先, 将信号点 $p(i)$ 进行归一化处理, 经过离散化处理、取整并转换为灰度值的信号为 $p^m(i)$, 其中函数 f 是将归一化的离散信号进行取整。

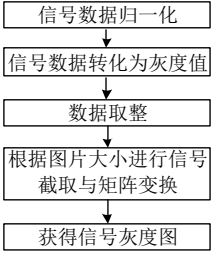


图 1 信号-图像转换原理图

Fig. 1 Signal-to-image conversion schematic

$$p^m(i) = f\left\{\frac{p(i) - p_{\min}(i)}{p_{\max}(i) - p_{\min}(i)}\right\} \times 255 \tag{1}$$

根据卷积神经网络的输入要求, 需要将转换为灰度值的离散信号进行长度截取, 信号截取长度为 N , 截取后的长度为 N^2 , 通常 N 的取值为 32、64 和 128 等。截取后 n 个离散信号为 $p_n(i)$, n 的范围为 $\left[1, f\left(\frac{i_{\max}}{N^2}\right)\right]$, f 为取整函数, 其中:

$$P_n(i) = p^m(i, i + N^2 - 1) \tag{2}$$

经过截取后的离散信号仍为 $1 \times N^2$ 的一维矩阵, 此时需要将其转换为 $N \times N$ 的二维矩阵, 最终经过转换的离散信号, 即表示灰度图的矩阵为

$$\begin{bmatrix} P(i) & \cdots & P(i + N - 1) \\ \vdots & & \\ P(i + N^2 - N) & \cdots & P(i + N^2 - 1) \end{bmatrix}$$

这种数据处理方法的优点是, 它提供了一种保留原始信号的二维特征的方法。更重要的是, 这种数据预处理方法可以在没有任何预定义参数的情况下进行计算, 并且尽可能减少技术人员的经验。

2.2 卷积神经网络 LeNet-5

LeNet-5 是一种经典的卷积神经网络(convolutional neural network, CNN), 其具有深度较浅、训练速度快、准确率能够得到一定保证的优点, 在工业界中得到了有效的应用。共包含 7 层结构, 其中包括 3 个卷积层、2 个最大池化层、1 个全连接层和 1 个输出层。在激活函数的选择上, 传

统的激活函数 Sigmoid 容易出现梯度消失、训练时间长的问题, 本文选择了线性整流函数(rectified linear unit, ReLU)激活函数 R_{eLU} , 有

$$R_{\text{eLU}}(x) = \begin{cases} x, & x > 0 \\ 0, & x \leq 0 \end{cases} \quad (3)$$

LeNET-5 卷积神经网络具体的结构如表 2 所示。

表 2 LeNet-5 结构参数
Tab. 2 LeNet-5 structure parameters

编号	网络层	卷积核大小	卷积核数目
C1	卷积层 1	5×5	6
S2	最大池化层 1	2×2	6
C3	卷积层 2	5×5	16
S4	最大池化层 2	2×2	16
C5	卷积层 3	5×5	16
F6	全连接层	120	1
L7	输出层	84	1

2.3 Adam 梯度下降优化算法

卷积神经网络训练过程中, 模型训练的精度影响最终测试分类的准确率, 由于参数及超参数多, 如果超参数选择不好, 使用随机梯度下降法(stochastic gradient descent, SGD)训练时往往陷入局部最优点。同时, 对于每一个参数都用相同的学习率进行更新, 但是在实际应用中各参数的重要性是不同的。本文采用自适应矩估计(adaptive moment estimation, Adam)优化算法对于不同的参数要动态地采取不同的学习率, 让目标函数更快地收敛, 加快训练速度, 避免陷入局部最优点。

虽然 Adam 是常见的、有效的梯度下降算法, 但是仍存在一些缺点: 自适应算法类似于过学习的效果, 生成的模型面对整体分布时是过拟合的; 同时, 相比于固定学习率的 SGD 算法, Adam 的学习率可能会发生剧烈变化, 这样会导致 Adam 可能不收敛, Adam 可能错失全局最优解。

2.4 AMSGrad 梯度下降优化算法

针对 2.3 节 Adam 算法的不足, 文献[21]提出了一种改进的 Adam 算法, 即 AMSGrad 算法。

在 Adam 算法中学习率可能是负的, 所以本文引入了一种替代方法, 把超参数设置为随着时间变化而变化, 从而保证学习率始终是个非负数。

2.5 卷积神经网络 DnCNN

由于信号可能存在一定程度的噪声干扰, 需要对经过信号-图像转换后的验证集图片进行降噪处理, 本文采用经过改进的 CNN 即前馈降噪卷积神经网络(deep CNN, DNCNN)的原因如下:

1) 卷积神经网络可有效提高利用图像特征的容量和灵活性;

2) CNN 训练正则和学习方法有相当大的提升;

3) 并行计算可提高运行速度。

DnCNN 采用残差学习策略, 在卷积神经网络的隐含层中隐式地移除干净图片(clean image)。即输入为带噪的观测图片(noisy image), 输出为移除干净图片的残差(噪声)图片(residual image)。其中包括 3 部分:

1) Conv+ReLU。第 1 层, 输入为 $N \times N \times C$ 的图像, 经过 64 个 $3 \times 3 \times 3$ 的卷积核卷积, 输出为 64 张 $N \times N$ 的特征图, 即 $N \times N \times 64$ 的图像。

2) Conv+BN+ReLU。第 $(2 \sim (d-1))$ 层, 有 64 个 $3 \times 3 \times 64$ 的卷积核, 所以这些层的输入输出都为 $N \times N \times 64$ 的图像。在卷积和激活函数间添加批量归一化。

3) Conv。通过 C 个 $3 \times 3 \times 64$ 的卷积核, 重建 c 维图像, 作为输出。

其中: Conv 表示卷积层; ReLU 表示激活函数, 即线性整流函数; BN (batch normalization) 表示批次标准化, 可加速收敛。在经过信号-图像处理, 通过加深卷积神经网络 DnCNN 的处理, 得到较为纯净的去噪图片作为模型的验证样本。

2.6 故障诊断总体流程

整体流程包括对故障数据集与现场信号的信号-图像转换、模型训练与模型验证诊断, 流程如图 2 所示。

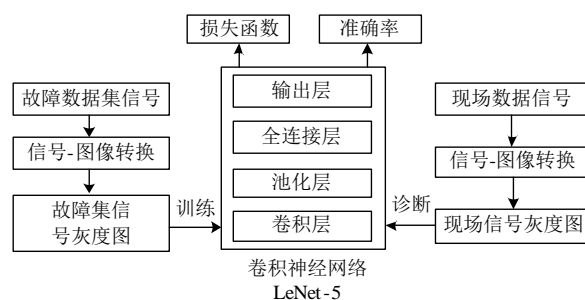


图 2 轴承故障训练-诊断流程图

Fig. 2 Bearing fault training-diagnosis process

故障诊断模型训练阶段将故障集信号通过信号-图像转换制作成故障集图像, 并作为训练数据集, 通过卷积神经网络 LeNet-5 训练出故障诊断模型, 通过模型的代价函数判断模型的鲁棒性, 并验证诊断准确率; 故障诊断模型验证阶段将采集验证信号通过信号-图像转换制作成验证集图像, 并作为验证数据集, 通过卷积神经网络 DnCNN 对含有

噪声的信号进行降噪处理,通过训练出的故障诊断模型得出故障类别归属权重,判断故障类型,并通过整体验证集数据得出故障诊断模型的预测准确率。

3 模型训练与实验验证

3.1 时域信号处理结果

本文使用的训练故障数据集为凯斯西储大学(CWRU)提供的电机轴承数据。在这个数据集中,有 4 种不同的转速状态,分别为 1730、1750、1772、1790r/m,每种转速状态中有滚动故障(BA)、内圈故障(IF)和外圈故障(OF)3 种故障类型,每种故障类型有 3 种不同的损伤大小,总共有 15 个故障状况和 1 个正常状况(NO)。损伤尺寸为 0.18、0.36、0.54 和 0.72mm;其中,外圈故障位置角度包括 0°、90°和 180°。

选取相同运转速度条件下收集从动端振动信号,作为电机轴承故障集,由于在信号-图像转换中不同图片大小对故障的识别精度与效率存在一定的影响,本文选取不同大小的转换格式,经过转换后的图片大小与数量如表 3 所示。

表 3 信号-图像转换结果
Tab. 3 Signal-image conversion result

图片大小	像素点	图片数量		
		状态	速度	总计
16×16	256	475	7600	30400
32×32	1024	120	1920	7680
64×64	4096	30	480	1920
128×128	16384	8	128	512

该方法通过图片的标签来表明故障归属,最大限度地保留了原信号的特征,同时具有一定的样本容量,保证了卷积神经网络的训练效果,防止过拟合。转换后的样本图片如图 3 所示,选取一种转速状态下,图片大小为 64×64 时,信号共存在 16 种状态,每个状态选取一张图片。

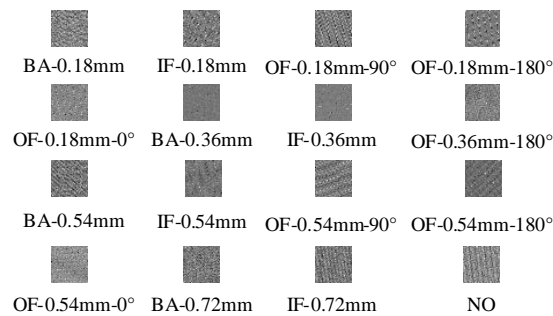
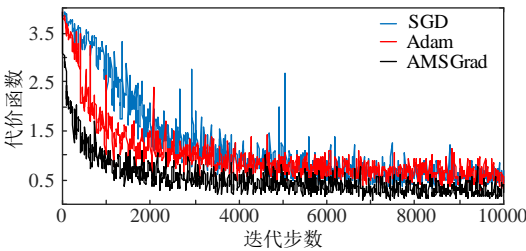


图 3 轴承故障信号-图像转换图

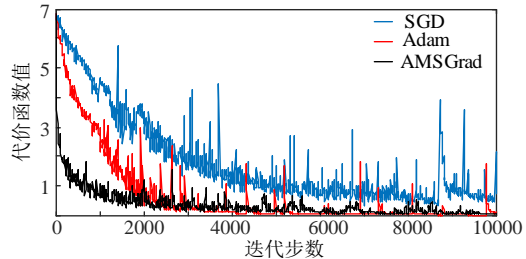
Fig. 3 Bearing fault signal-image conversion

3.2 模型训练与验证

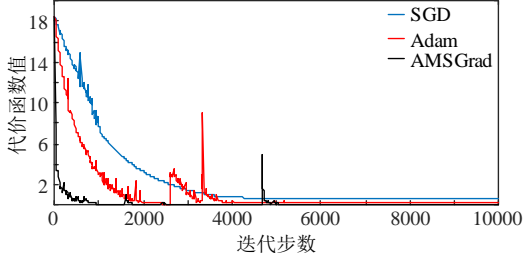
本文选用基于 Openstack 的 Windows Server 2016 虚拟机作为训练卷积神经网络的平台,服务器配置为 16 核、16G 内存、200G 硬盘空间,卷积神经网络模型的程序由 python3.6.5 编写,并使用谷歌开源框架 TensorFlow。在训练过程中,迭代步数为 10000 次,每批处理的样本个数 batch_size 为 50,在选用 16×16、32×32、64×64 和 128×128 图片大小情况下使用 SGD 算法、Adam 算法与 AMSGrad 算法训练模型的代价函数下降趋势如图 4 所示。代价函数即所有样本误差的平均,代价函数值越小,所训练的模型拟合性越好,3 种梯度下降算法的代价函数最终值比较如表 4 所示。



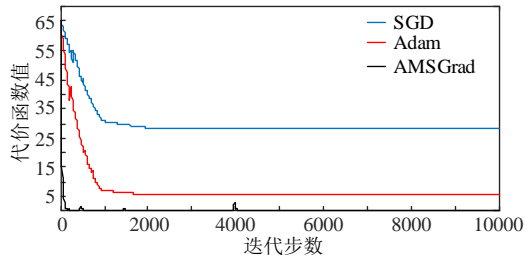
(a) 图片大小 16×16



(b) 图片大小 32×32



(c) 图片大小 64×64



(d) 图片大小 128×128

图 4 3 种梯度下降算法的代价函数值对比图

Fig. 4 Comparison of cost functions of three gradient descent algorithms

表 4 3 种梯度下降算法的代价函数

Tab. 4 Cost functions of three gradient descent algorithms

图片大小	代价函数最终值			训练速度/ (s/批次)
	SGD 算法	Adam 算法	AMSGrad 算法	
16×16	0.50	0.50	0.24	0.04
32×32	0.47	0.04	0.04	0.26
64×64	0.52	0.15	0.05	0.47
128×128	28.45	5.62	0.01	1.10

以代价函数最终值为指标，同一种算法下，训练样本图片格式由 16×16—128×128 时，SGD 算法的代价函数最终值基本在 0.5 上下，只有当为 128×128 时，最终值过大，为 28.45，而且所有格式下均存在多次较大的震荡；当训练样本图片格式一致时，Adam 与 SGD 算法相比，其代价函数的最终值偏小；相较 SGD 与 Adam 算法，AMSGrad 算法的代价函数最终值趋于更小，在 32×32、64×64 格式下分别是 0.04、0.05；经过对比可知，AMSGrad 算法能使训练过程中代价函数值最小，更小的代价函数值意味着模型具有更好的准确性与鲁棒性。

从代价函数下降趋势上看，对比可见，除个别节点外，训练样本的目标函数值稳步下降，在 3000 步之前，相比于 SGD、Adam 算法，AMSGrad 算法使代价函数更快地下降、震荡更小，分别在 2100、1200、700、200 步就能使代价函数值趋于稳定。

从训练速度看，模型的训练速度分别为 0.04、0.26、0.47 和 1.1s/批次。通过对比可知：图片格式大小越大，模型训练速度越慢。

SGD 算法、Adam 算法与 AMSGrad 算法训练模型的准确率如图 5 所示，3 种梯度下降算法的准确率比较如表 5 所示。

以准确率最终值为指标，同一种算法下，训练样本图片格式由 16×16—128×128 时，图片格式大小越大，准确率越高，但是增至 64×64 后，受样本数量与硬件设备原因制约准确率上升空间有限，并且是以牺牲速度为代价；当训练样本图片格式一致时，相较 SGD 与 Adam 算法，AMSGrad 算法的准确率最终值更高，趋于 99%，而且 AMSGrad 算法使准确率更快地上升，震荡更小。

从转换图片大小来看，小图片大数据量与大图片小数据量都不是最好的选择，前者由于样本个体特征信息不完全带来的学习过程损失，导致训练效果不佳的现象；在图片格式为 16×16 的情况下，训练速度为 0.04s/批次，但识别精度为 86%。后者由

于图片过大导致模型训练速度慢，影响了诊断前期的模型建立速度；在图片格式为 128×128 的情况下，训练速度为 1.1s/批次，识别精度为 99%。因此，在实际的模型训练中，选取合适的图片大小即 64×64 时，训练速度为 0.47s/批次，识别精度为 98%，可以兼顾训练速度与训练精度。

同时，考虑经典的一维信号故障诊断方法，其

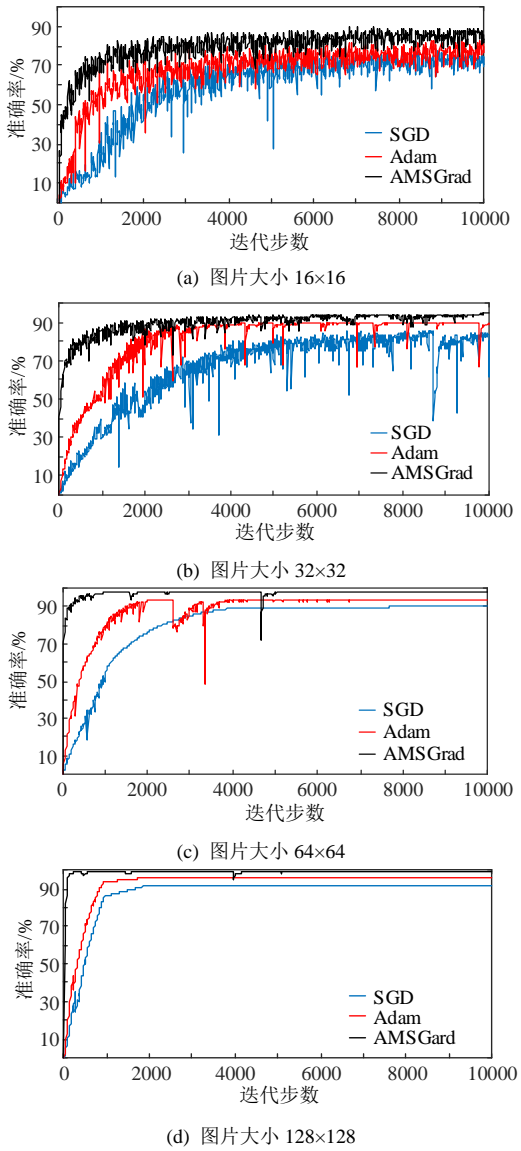


图 5 3 种梯度下降算法的准确率对比图
Fig. 5 Comparison of accuracy of three gradient descent algorithms

表 5 3 种梯度下降算法的准确率

Tab. 5 Accuracy of three gradient descent algorithms

图片格式	准确率/%			训练次数/ (s/批次)
	SGD	Adam	AMSGrad	
16×16	70	72	86	0.04
32×32	81	89	94	0.26
64×64	90	92	98	0.47
128×128	90	95	99	1.10

准确率如表 6 所示。在广泛应用的经典一维信号故障诊断中, SVM 与 BP 神经网络的训练过程需要进行特征向量的选取, 由于需要专家经验的人工选取, 最终的准确率并不高; 一维卷积神经网络用于故障诊断是一种端到端的诊断方法, 虽然不需要特征向量的选取, 但由于一维信号的样本长度、样本数量的局限性, 导致一维卷积神经网络的感受野不够大, 训练过程中存在过拟合的情况, 导致最终的准确率降低, 从而使卷积神经网络大规模样本处理能力的优势无法发挥。

表 6 经典诊断方法的准确率
Tab. 6 Accuracy of classical diagnostic methods

诊断方法	准确率/%
SVM+特征选取	85
BP 神经网络	86
一维卷积神经网络(CNN)	96

综上分析, 综合考虑代价函数、准确率最终值, 及训练速度与训练效果的平衡, 在训练模型时选取图片大小为 64×64 的训练集, 利用二维卷积神经网络进行训练, 训练使用的梯度下降法选用 AMSGrad 算法。

3.3 噪声干扰验证

实际电机的运行信号包含一定的外界干扰噪声, 为验证所训练模型对含有噪声信号的验证准确率, 将数据集中的故障信号添加白噪声作为验证所用的验证集。

在白噪声的添加上, 本文对故障信号分别添加 0~10dB 的白噪声, 并对信号同样进行 16×16、32×32、64×64、128×128 图片大小规格的图片转化, 首先对未进行图像去噪的验证图片进行验证, 分类结果的准确率对比如表 7 所示。

表 7 未降噪下不同噪声情况下的模型识别准确率
Tab. 7 Model recognition accuracy under different noise conditions without noise reduction

图片大小	准确率/%					
	0dB 白 噪声	2dB 白 噪声	4dB 白 噪声	6dB 白 噪声	8dB 白 噪声	10dB 白 噪声
16×16	86.07	78.22	76.91	74.32	73.12	70.15
32×32	94.46	88.45	85.28	83.99	82.01	80.01
64×64	98.55	93.86	92.15	91.22	90.15	90.01
128×128	99.10	94.11	94.00	92.47	91.25	90.99

通过 DnCNN 对增加噪声的图片进行降噪处理后作为验证集, 并对相应图片格式的模型进行分类验证, 分类的准确率对比如表 8 所示。

表 8 降噪处理下不同噪声情况下的模型识别准确率
Tab. 8 Model recognition accuracy under different noise conditions under noise reduction

图片大小	准确率/%					
	0dB 白 噪声	2dB 白 噪声	4dB 白 噪声	6dB 白 噪声	8dB 白 噪声	10dB 白 噪声
16×16	86.07	85.34	85.11	84.55	84.01	83.66
32×32	94.46	93.66	92.78	92.56	92.32	92.29
64×64	98.55	98.14	97.95	97.54	97.20	96.99
128×128	99.10	98.21	97.67	97.15	97.02	96.99

从噪声干扰方面看, 经过训练的模型对未经降噪处理的验证信号识别率较低, 在较大噪声的情况下甚至出现了损失大于 15% 的情况。在降噪处理方面, 经过训练的模型对经过降噪处理验证集的识别率保持了较高的识别率, 在较大噪声的情况下, 针对不同的图片大小仍保持损失小于 4% 的识别准确率, 可证明利用 DnCNN 对含有噪声的信号进行处理仍具有较高的识别率。

3.4 实际信号验证

为验证训练模型的预测精度, 通过现场扭矩传感器采集转矩信号作为实验数据集, 对数据集进行分割, 每一种工况类型的训练集与验证集的比例为 9:1, 现场采集设备与电机参数如图 6 所示。图中, 传动装置可通过与异步电机转动轴的接触与分离实现轴承转速达到指定转速, 扭矩传感器记录在装载不同轴承下的扭矩。

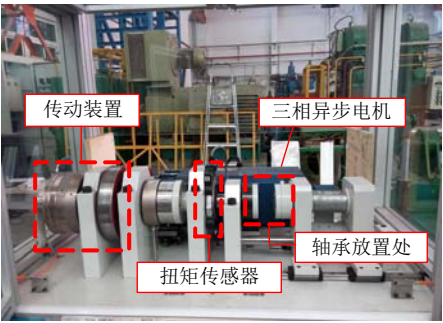


图 6 运行设备图

Fig. 6 Operation equipment

实验中共采集了 4 组轴承在一定转速范围下的震动数据, 其中选取滚动磨损 0.25±0.03mm (BA)、内圈故障 0.23±0.03mm (IF)、外圈故障 1–0.50±0.05mm–150° (OF)、外圈故障 2–0.30±0.05mm–90° (OF)。转速均在 1680~1840r/m 范围内, 每间隔 10r/m 选取一个转速值作为转速设定, 在每个转速设定值下每个轴承情况采集 100 组信号, 并通过本文的数据处理方法将其转化为二维灰度图, 将其作为测试数据集, 选取图片大小为 64×64 的转换格式。并在

进行信号-图片可视化转化后利用 DnCNN 对图片进行降噪处理,形成去噪验证集;并将验证集内的数据进行模型分类验证。

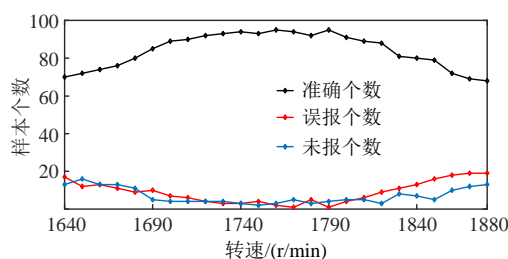


图 7 AMSGrad 梯度下降算法
滚动磨损信号预测情况对比图

Fig. 7 Comparison of recognition effects of AMSGrad
gradient descent algorithms

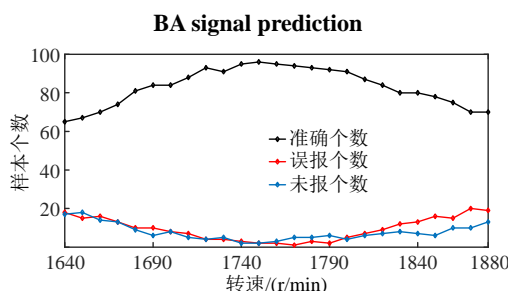


图 8 AMSGrad 梯度下降算法
内圈故障信号预测情况对比图

Fig. 8 Comparison of recognition effects of AMSGrad
gradient descent algorithms IF signal prediction

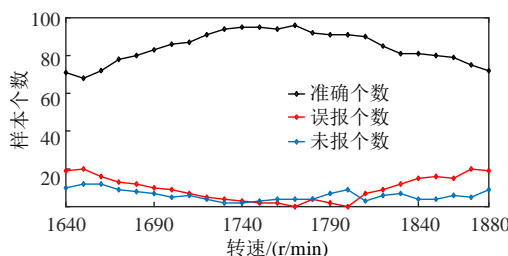


图 9 AMSGrad 梯度下降算法
外圈故障信号 1 预测情况对比图

Fig. 9 Comparison of recognition effects of AMSGrad
gradient descent algorithms

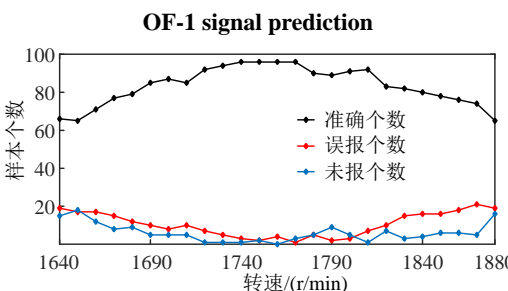


图 10 AMSGrad 梯度下降算法
外圈故障信号 2 预测情况对比图

Fig. 10 Comparison of recognition effects o
f AMSGrad gradient descent algorithms
OF-2 signal prediction

64×64、128×128 图片大小规格的故障信号的分类结果如图 7—10 所示,其中:“误报”表示模型将验证数据分类为其他故障类型;“未报”表示模型将验证数据分类为正常运行类型。

由图 7—10 可见,在训练集给出的转速范围内,各类故障信号均能达到不错的故障识别效果。以滚动磨损信号的分类结果为例,在训练集给出的转速 1730~1790r/min 内,滚动磨损信号的识别准确率达到 94%以上,误报、未报个数在 5 个以下,达到了不错的故障识别效果。而在训练集给定的转速之外,即当转速小于 1730r/min、或者转速大于 1790r/min 时,故障的识别率下降,与给定转速范围相差越大,识别准确个数逐渐减少,误报、未报个数逐渐增多。在转速范围为 1700~1800r/min 内的识别准确率达到 90%以上,误报、未报个数在 10 个以内;在转速范围为 1660~1830r/min 内的识别准确率达到 80%以上,误报、未报个数在 15 个以内;在转速范围为 1640~1880r/min 内的识别准确率达到 75%以上,误报、未报个数在 25 个以内。

4 结论

本文提出一种用于轴承故障诊断的二维卷积神经网络优化方法,通过采用信号-图像方法对样本信号与验证信号进行处理,再结合 AMSGrad 算法进行训练,并对比了 3 种梯度下降算法的训练效果与分类准确率,得出了以下结论:

1) 信号-图像预处理方法处理的故障信号能更多地保留原信号内的特征,减少了对经验的依赖,相比于传统的数据与处理方式能节省出更多的处理时间,提高了故障分类的精度,同时减少了外部噪声带来的诊断干扰;

2) 所提出的优化方法能更加精确、快速地训练卷积神经网络模型,减少了由于代价函数震荡带来的预测精度下降,能更快地达到更好的训练效果,并在实际验证样本训练集给定范围内达到很高的识别准确率。

参考文献

- [1] 刘长良,武英杰,甄成刚.基于变分模态分解和模糊 C 均值聚类的滚动轴承故障诊断[J].中国电机工程学报,2015,35(13):3358-3365.
Liu Changliang, Ying Wujie, Zhen Chenggang. Rolling bearing fault diagnosis based on variational mode decomposition and fuzzy C means clustering [J]. Proceedings of the CSEE, 2015, 35(13): 3358-3365(in

- Chinese).
- [2] Gonzalez-Prieto I, Duran M J, Rios-Garcia N, et al. Open-switch fault detection in five-phase induction motor drives using model predictive control[J]. IEEE Transactions on Industrial Electronics, 2018, 65(4): 3045-3055.
- [3] 杨明, 柴娜, 李广. 等. 基于电机驱动系统的齿轮故障诊断方法对比研究[J]. 电工技术学报, 2016, 31(19): 132-140.
- Yang Ming, Chai Na, Li Guang. et al. A comparative study of gear fault diagnosis methods based on the motor drive system[J]. Transactions of China Electrotechnical Society, 2016, 31(19): 132-140(in Chinese).
- [4] 王臻, 李承, 张舜钦, 等. 基于改进矩阵束滤波与检测的异步电动机故障诊断新方法[J]. 中国电机工程学报, 2015, 35(12): 3969-3977.
- Wang Zhen, Li Cheng, Zhang Shunqin, et al. A new filter and detection approach of fault in asynchronous motor based on improved matrix pencil[J]. Proceedings of the CSEE, 2015, 35(12): 3969-3977(in Chinese).
- [5] 魏书荣, 李正茂, 符杨, 等. 计及电流估计差的海上双馈电机定子绕组匝间短路故障诊断[J]. 中国电机工程学报, 2018, 38(13): 213-219.
- Wei Shurong, Li Zhengmao, Fu Yang, et al. Inter-turn faults diagnosis on stator windings of offshore wind DFIGs considering current estimated difference [J]. Proceedings of the CSEE, 2018, 38(13): 213-219(in Chinese).
- [6] 袁莉芬, 孙业胜, 何怡刚, 等. 基于小波包优选的模拟电路故障特征提取方法[J]. 电工技术学报, 2018, 33(11): 158-165.
- Yuan Lifen, Sun Yesheng, He Yigang, et al. Fault feature extraction method for analog circuit based on preferred wavelet packet[J]. Transactions of China Electrotechnical Society, 2018, 33(11): 158-165(in Chinese).
- [7] Gao Z W, Cecati C, Ding Seven X. A survey of fault diagnosis and fault-tolerant techniques—Part I: Fault diagnosis with model-based and signal-based approaches[J]. IEEE Transactions on Industrial Electronics, 2015, 62(6): 3757-3767.
- [8] Wang D, Peter W T. Prognostics of slurry pumps based on a moving-wear degradation index and a general sequential Monte Carlo method[J]. Mechanical Systems and Signal Processing, 2015, 56-57: 213-229.
- [9] 周国亮, 朱永利, 王桂兰. 实时大数据处理技术在状态监测领域中的应用[J]. 电工技术学报, 2014, 29(S1): 432-437.
- Zhou Guoliang, Zhu Yongli, Wang Guilan. Real-time big data processing technology application in the field of state monitoring[J]. Transactions of China Electrotechnical Society, 2014, 29(S1): 432-437(in Chinese).
- [10] 朱永利, 李莉, 宋亚奇, 等. ODPS 平台下的电力设备监测大数据存储与并行处理方法[J]. 电工技术学报, 2017, 32(9): 199-210.
- Zhu Yongli, Li Li, Song Yaqi, et al. Storage and parallel processing of big data of power equipment condition monitoring on ODPS platform[J]. Transactions of China Electrotechnical Society, 2017, 32(9): 199-210(in Chinese).
- [11] AYDMJ T, Duin R P W. Pump failure determination using support vector data description[C]//Proceedings of the 19th ACM Conference on Information and Knowledge Management, Toronto, 2010.
- [12] 梅飞, 梅军, 郑建勇, 等. 基于 KPCA-SVM 的断路器故障稳健诊断方法[J]. 电工技术学报, 2014, 29(S1): 50-58.
- Mei Fei, Mei Jun, Zheng Jianyong, et al. The robust fault diagnostic method for circuit breaker based on KPCA and SVM[J]. Transactions of China Electrotechnical Society, 2014, 29(S1): 50-58(in Chinese).
- [13] 廖剑, 史俊贤, 周绍磊, 等. 基于局部图嵌入加权罚 SVM 的模拟电路故障诊断方法[J]. 电工技术学报, 2016, 31(4): 28-35.
- Liao Jian, Shi Junxian, Zhou Shaolei, et al. Analog circuit fault diagnosis based on local graph embedding weighted-penalty SVM[J]. Transactions of China Electrotechnical Society, 2016, 31(4): 28-35(in Chinese).
- [14] 孙曙光, 张强, 杜太行, 等. 基于振动信号的低压万能式断路器分合闸故障程度评估方法的研究[J]. 中国电机工程学报, 2017, 37(18): 5473-5482.
- Sun Shuguang, Zhang Qiang, Du Taihang, et al. Study of evaluation method for low voltage conventional circuit breaker switching fault degree based on vibration signal[J]. Proceedings of the CSEE, 2017, 37(18): 5473-5482(in Chinese).
- [15] 史丽萍, 汤家升, 王攀攀, 等. 采用最优小波树和改进 BP 神经网络的感应电动机定子故障诊断[J]. 电工技术学报, 2015, 30(24): 38-45.
- Shi Liping, Tang Jiasheng, Wang Panpan, et al. Stator fault diagnosis of induction motors using the optimal wavelet tree and improved BP neural network [J]. Transactions of China Electrotechnical Society, 2015, 30(24): 38-45(in Chinese).
- [16] 杨凌霄, 朱亚丽. 基于概率神经网络的高压断路器故障

- 诊断[J]. 电力系统保护与控制, 2015, 43(10): 62-67.
- Yang Lingxiao, Zhu Yali. High voltage circuit breaker fault diagnosis of probabilistic neural network[J]. Power System Protection and Control, 2015, 43(10): 62-67(in Chinese).
- [17] Shao H, Jiang H, Zhang X, et al. Rolling bearing fault diagnosis using an optimization deep belief network [J]. Measurement Science and Technology, 2015, 26(11): 11502.
- [18] Wen L, Gao L, Li X Y. A new deep transfer learning based on sparse auto-encoder for fault diagnosis[J]. IEEE Transactions on Systems Man and Cybernetics: Systems, 2017(99): 1-9.
- [19] 候文擎, 叶鸣, 李巍华. 基于改进堆叠降噪自编码的滚动轴承故障分类[J]. 机械工程学报, 2018, 54(7): 87-96.
- Hou Wenqing, Ye Ming, Li Weihua. Rolling element bearing fault classification using improved stacked de-noising auto-encoders[J]. Journal of Mechanical Engineering, 2018, 54(7): 87-96(in Chinese).
- [20] Ince T, Kiranyaz S, Eren L, et al. Real-time motor fault detection by 1-d convolutional neural networks[J]. IEEE Transactions on Industrial Electronics, 2016, 63(11): 7067-7075.
- [21] Reddi S J, Kale S, Kumar S. On the convergence of adam and beyond[C]//International Conference on Learning Representations (ICLR), Vancouver, Canada, 2018.



肖雄

收稿日期: 2018-10-09。

作者简介:

肖雄(1989), 男, 博士, 研究方向为变频调速系统中的非线性控制算法研究与应用、工业过程中故障诊断研究, docxxiao@ustb.edu.cn;

*通信作者: 王健翔(1993), 男, 硕士研究生, 研究方向为新型电机的运行理论及控制, wangjianxiang1025@126.com。

(责任编辑 李婧妍)

A Two-dimensional Convolutional Neural Network Optimization Method for Bearing Fault Diagnosis

XIAO Xiong, WANG Jianxiang*, ZHANG Yongjun, GUO Qiang, ZONG Shengyue

(Institute of Engineering Technology, University of Science and Technology Beijing)

KEY WORDS: deep learning; fault diagnosis; signal conversion; convolutional neural network

Bearing intelligent fault diagnosis is a hot research field of mechanical big data condition monitoring. The traditional data-driven fault diagnosis method is extremely time-consuming and requires high expert experience for signal extraction-based feature extraction, in order to eliminate the pre-defined effects of parameters. In order to improve the recognition rate while improving the feature extraction, this paper proposes a two-dimensional convolutional neural network optimization method for bearing fault diagnosis based on the research of one-dimensional convolution neural network fault diagnosis method. This method introduces a new data preprocessing method, which converts the original time domain signal data into a two-dimensional gray image to extract the transformed image features, eliminating the influence of manual features.

Each pixel of the grayscale image contains an integer value of 0~255, which is represented as the gray value of each pixel of the image. The original signal $p(i)$ is a discrete signal of a one-dimensional time series, wherein the signal includes some signal points, and the signal point is first performed. Normalized processing, is discretized, rounded, and converted to grayscale values $p^m(i)$. After truncating, the discrete signal is $p_n(i)$, and the range is $[1, f(\frac{i_{\max}}{N^2})]$, f for the rounding function. The results of signal-image conversion are shown in Fig. 1.

$$p^m(i) = f\left\{\frac{p(i) - p_{\min}(i)}{p_{\max}(i) - p_{\min}(i)}\right\} \times 255 \quad (1)$$

$$P_n(i) = p^m(i, i + N^2 - 1) \quad (2)$$

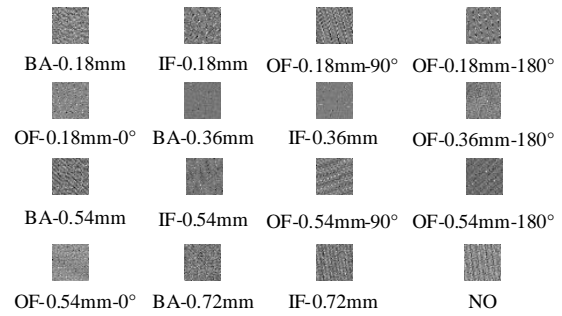


Fig. 1 Fault signal signal-image conversion

The fault diagnosis accuracy and training speed in the case of four kinds of picture size conversion formats and three gradient descent algorithms are shown in Tab.1. According to the data in the table, 64×64 is the best size of the pictures.

Tab. 1 Accuracy of three gradient descent algorithms

Size of picture	Accuracy/%			Training speed/ (s/batch)
	SGD	Adam	AMSGrad	
16×16	70	72	86	0.04
32×32	81	89	94	0.26
64×64	90	92	98	0.47
128×128	90	95	99	1.10

In terms of noise interference, the training model has a low recognition rate for the verification signal without noise reduction, and in the picture noise reduction processing, the DnCNN processing contains a noise signal with a high recognition rate. In the range of speeds given by the training set, all kinds of fault signals can achieve a good fault recognition effect. In addition to the given rotational speed of the training set, the recognition rate of the fault decreases, and the difference from the given rotational speed range is larger, the number of accurate identifications is gradually reduced, and the number of false positives and unreported numbers are gradually increased.