

PSO 优化深度神经网络诊断齿轮早期点蚀故障

李嘉琳¹, 何巍华^{1,2}, 曲永志³
(1. 东北大学 机械能与自动化学院, 辽宁 沈阳 110819; 2. 伊利诺伊大学芝加哥分校 机械及工业工程系, 伊利诺伊州 芝加哥 60607; 3. 武汉理工大学 机电工程学院, 湖北 武汉 430070)

摘 要: 基于数据驱动方法诊断齿轮故障时一般会用傅里叶变换等进行特征提取,特征提取方法的选取对诊断结果影响很大. 提出应用深度神经网络来诊断齿轮早期点蚀故障,直接以采集的振动信号作为网络输入,可以避免特征提取环节产生误差. 此外,应用粒子群算法优化深度神经网络,使训练过程更稳定、诊断率更高. 在分析结果时应用主成分分析法对网络输出进行降维. 用实验采集的数据训练并测试网络,诊断正确率能达到 90% 之上,证明所提出的方法是合理、可用的.

关 键 词: 齿轮; 早期点蚀; 粒子群算法; 深度神经网络; 主成分分析

中图分类号: TH 132.41 **文献标志码:** A **文章编号:** 1005-3026(2019)07-0974-06

Diagnosis of Gear Early Pitting Faults Using PSO Optimized Deep Neural Network

LI Jia-lin¹, HE David^{1,2}, QU Yong-zhi³
(1. School of Mechanical Engineering & Automation, Northeastern University, Shenyang 110819, China; 2. Department of Mechanical & Industrial Engineering, University of Illinois at Chicago, Chicago 60607, USA; 3. School of Mechanical and Electronic Engineering, Wuhan University of Technology, Wuhan 430070, China. Corresponding author: HE David, E-mail: heweihua@me.neu.edu.cn)

Abstract: When gear faults were diagnosed based on the data-driven method, feature extraction was generally performed by Fourier transform, etc. The feature extraction method used has a great influence on the diagnosis results. Therefore, deep neural network (DNN) was proposed to diagnose early gear pitting faults and the vibration signals are directly used as the network inputs to avoid errors caused by feature extraction. In addition, the particle swarm optimization (PSO) algorithm was applied to optimize the DNN for obtaining a more stable training process and better diagnosis results. Principal component analysis (PCA) algorithm was used to reduce the dimensions of the DNN outputs. The data collected from the experiment was used to train and test the DNN. The fault diagnostic accuracy can reach over 90%, which proves that the proposed method is reasonably effective.

Key words: gear; early pitting; PSO algorithm; deep neural network (DNN); principal component analysis (PCA)

齿轮是现代机械设备中最常见的传动零件,广泛应用于飞机、汽车、机床等现代机械设备,所以对其的故障监测十分必要. 齿轮的故障包括齿轮断齿、齿轮点蚀、齿面划痕、齿面磨损等,其中齿轮点蚀故障大约占齿轮全部故障的 31%^[1].

齿轮点蚀故障的诊断方法有很多,可以先对时域信号进行处理,之后根据经验分析是否发生故障. 文献[2]分别应用了傅里叶变换 (Fourier transform, FT) 和小波变换 (wave transform, WT) 来处理振动信号,并分析处理结果判断齿轮的状况. 文献[3]应用了双频谱分析法处理振动信号,并对比健康和故障齿轮的频谱差异. 文献[4]通

过信号的静态特征如均方差和峭度等区别齿轮状况,并对信号进行了包络分析.文献[5]应用稀疏全变分模型实现振动信号冲击调整分离并进行包络分析诊断齿轮故障.以上方法均是信号处理后依靠经验人为判断是否发生故障.此外还有应用特征提取和智能算法来进行故障诊断的,文献[6]分别应用了支持向量机(SVM)和人工神经网络(artificial neural network, ANN)来诊断轴承故障,从时域信号中提取平均值、均方差和峭度等 9 个统计量作为 ANN 的输入,并应用遗传算法(GA)优化 ANN.文献[7]从时域信号提取了 24 个统计量并应用 ANN 来诊断齿轮-轴承的混合故障.文献[8]应用希尔伯特-黄变换提取 6 个特征参数作为自组织映射(self-organization mapping, SOM)神经网络的输入进行齿轮故障诊断.文献[9]结合混沌优化策略和人工蜂群搜索算法提出了引力搜索径向基神经网络诊断机车齿轮箱故障.文献[10]应用连续的 WT 处理声发射信号,之后应用 ANN 来进行故障诊断.

由于特征提取困难,本文提出应用深度神经网络(deep neural network, DNN)来诊断早期齿轮点蚀故障,不进行特征提取操作,直接将采集的振动信号做 DNN 的输入.

1 算法实现

1.1 DNN 简介

DNN 的正向传播主要依靠众多神经元的计算来完成,其结构如图 1 所示,工作过程可以用式(1)、式(2)来表示:

$$u_k = \sum_{i=1}^n w_{ki} x_i, \quad (1)$$

$$y_k = f(u_k - b_k). \quad (2)$$

其中: x_i 表示第 i 个输入; w_{ki} 表示与第 i 输入量相连的权值; u_k 表示所有输入的加权和; b_k 为神经元阈值; $f(\cdot)$ 为激活函数; y_k 为神经网络的输出.

激活函数的种类有很多,如 sigmoid, tanh 及 ReLU. 本文应用的是 ReLU 激活函数,如式(5)所示.

$$f_{\text{sig}} = 1 / (e^{-z} + 1), \quad (3)$$

$$f_{\text{tanh}} = (e^z - e^{-z}) / (e^z + e^{-z}), \quad (4)$$

$$f_{\text{ReLU}} = \max(0, z). \quad (5)$$

式(6)为 ReLU 激活函数的导函数,可以发现输入小于 0 的部分都会被置零,这使得网络有较好的稀疏性.深度学习的目的就是从中众多复杂、密集的特征中提取‘关键因子’,所以要解开这些

缠绕的复杂关系就要求网络有良好的稀疏特性,而且还能有效避免梯度消失和梯度爆炸.

$$\frac{d}{dz} f_{\text{ReLU}} = \begin{cases} 1, & z > 0; \\ 0, & z \leq 0. \end{cases} \quad (6)$$

网络的输出层使用 softmax 函数作分类器,式(7)为第 i 个神经的输出.

$$f_{\text{softmax}} = e^i / \sum_j e^j. \quad (7)$$

其中: i 为所求神经元编号; j 为输出层神经元数.

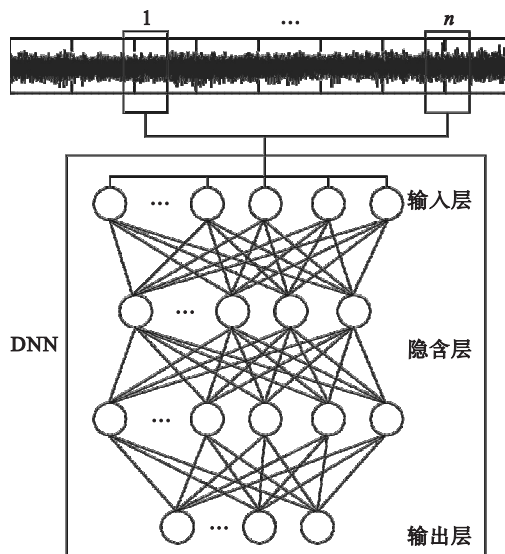


图 1 DNN 结构

Fig. 1 Structure of DNN

DNN 的反向传播过程即为网络的误差反向传播及自我修正的过程,式(8)为均方差误差函数,式(9)为交叉熵误差函数:

$$E_{\text{MSE}} = 1/2 (\mathbf{O} - \mathbf{Y})^2, \quad (8)$$

$$E_{\text{cross-entropy}} = -(\mathbf{O} \lg \mathbf{Y} + (1 - \mathbf{O}) \lg (1 - \mathbf{Y})). \quad (9)$$

其中: \mathbf{O} 为期望输出; \mathbf{Y} 为实际输出.

式(10)~式(12)比较了 sigmoid 为传递函数时,交叉熵损失函数和均方差损失函数的反向传播效果,可以发现交叉熵作为损失函数时,网络修正更迅速.

$$\frac{\partial E_{\text{MSE}}}{\partial \mathbf{w}} = (\mathbf{O} - \mathbf{Y}) f'_{\text{sig}}(\mathbf{w} \mathbf{z} + \mathbf{b}) \mathbf{z}, \quad (10)$$

$$\frac{\partial E_{\text{cross-entropy}}}{\partial \mathbf{w}} = (\mathbf{Y} - \mathbf{O}) \mathbf{z}, \quad (11)$$

$$\frac{\Delta \mathbf{w}_{\text{MSE}}}{\Delta \mathbf{w}_{\text{cross-entropy}}} = \frac{|\mathbf{O} - \mathbf{Y}| \cdot |f'_{\text{sig}}(\mathbf{w} \mathbf{z} + \mathbf{b})| \cdot |\mathbf{z}|}{|\mathbf{Y} - \mathbf{O}| \cdot |\mathbf{z}|} \leq 0.25. \quad (12)$$

为了避免过拟合情况出现,反向修正过程中应用 L2 正则化提高网络的泛化能力,式(13)~式(17)所示对误差函数进行修正.

$$E = E_0 + \frac{\lambda}{2n} \sum_w w^2, \tag{13}$$

$$\frac{\partial E}{\partial w} = \frac{\partial E_0}{\partial w} + \frac{\lambda}{n} w, \tag{14}$$

$$\frac{\partial E}{\partial b} = \frac{\partial E_0}{\partial b}, \tag{15}$$

$$w_{\text{new}} = w - \eta \left(\frac{\partial E_0}{\partial w} + \frac{\lambda}{n} w \right) = \left(1 - \frac{\eta \lambda}{n} \right) w - \eta \frac{\partial E_0}{\partial w}, \tag{16}$$

$$1 - \frac{\eta \lambda}{n} < 1. \tag{17}$$

其中: E_0 为原误差函数; E 为修正后误差函数; η 为神经网络的学习率; λ 为正则系数; n 训练样本量; w 为权值向量; b 为神经元阈值向量.

1.2 PSO 算法

粒子群优化 (particle swarm optimization, PSO) 算法被广泛地应用在生活中的各个领域. 其基本原理可以描述为: p 维空间中, 存在着一个种群规模为 n 的粒子集. 种群规模为粒子的个数, 即鸟群中鸟的个数. 其中第 i 个粒子的速度和位置可以用 $v_i = (v_{i1}, v_{i2}, \cdots, v_{ip})$ 和 $x_i = (x_{i1}, x_{i2}, \cdots, x_{ip})$ 来表示, 每个粒子都存在一个适应度 f , 适应度的大小相当于每个鸟与食物之间的距离, 根据粒子的适应度可以对群体的个体极值 P_b 和群体极值 G_b 进行更新, 之后可以利用个体极值和种群极值对粒子的速度和位置进行计算, 计算公式为

$$v_{ij}(t+1) = w_g v_{ij}(t) + c_1 \varepsilon_1 (P_{bj}(t) - x_{ij}(t)) + c_2 \varepsilon_2 (G_{bj} - x_{ij}(t)), \tag{18}$$

$$x_{ij}(t+1) = x_{ij}(t) + v_{ij}(t+1). \tag{19}$$

其中: i 表示第 i 个粒子; j 表示第 j 维空间; c_1, c_2 为学习因子, c_1 为粒子的自身部分, 表示自身对寻优的认识和影响, c_2 为社会部分, 表示粒子受种群影响; t 为迭代次数; $\varepsilon_1, \varepsilon_2$ 为均匀分布在 $0 \sim 1$ 之间的随机数; w_g 为惯性权重, 表示受上次速度的影响.

1.3 PCA 算法

主成分分析 (principal component analysis, PCA) 是一种常用的数据算法. 当数据维度较高时, 用其降低数据维度以便进行数据分析, 常用于图片压缩、因子分析等. 本文应用 PCA 处理 DNN 网络的输出, 更直观地反映出 DNN 的诊断效果.

应用映射矩阵 V 将 n 维特征向量 x 映射到 k ($k < n$) 维空间内, 如式 (20) 所示. 映射矩阵 V 的求解见式 (21), 式 (22), 先求解矩阵 Ψ , 再求解 Ψ 的协方差矩阵 C , 最后求解协方差矩阵 C 的特征

向量和特征值, 之后由前 k 个较大特征值对应的特征向量组成的矩阵即为映射矩阵 V .

$$Y = V^T x, \tag{20}$$

$$\Psi = x_{ij} - \frac{1}{s} \sum_{i=1}^s x_i, \tag{21}$$

$$C = \frac{1}{s} \sum_{i=1}^s \Psi \Psi^T. \tag{22}$$

其中: x 为特征向量; Y 为降维矩阵; s 为样本数; Ψ 为特征矩阵中元素每列平均值后的新矩阵; C 为 Ψ 的协方差矩阵.

2 实验设计及数据采集

实验齿轮箱如图 2 所示, 由两个额定功率 45 kW 的西门子电机驱动, 电机 1 为驱动电机, 电机 2 为受载电机, 并配有润滑和冷却系统, 三向振动传感器安装在从动轮轴承座上, 可采集 X, Y, Z 三个方向的振动数据, 采样频率为 10 240 Hz. 齿轮相关参数见表 1.

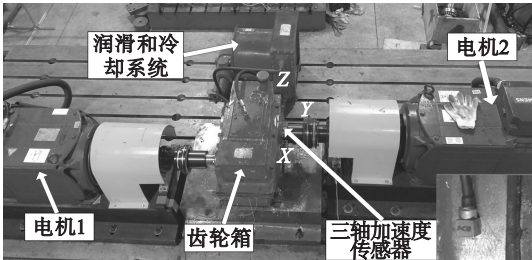


图 2 齿轮箱实验台
Fig. 2 Gearbox test rig

表 1 齿轮相关参数
Table 1 Gear parameters

参数	主动轮	从动轮
齿数/个	40	72
齿轮模数	3	3
基圆直径/mm	112.763	202.974
节圆直径/mm	120	216
压力角/(°)	20	20
齿距/mm	9.424 78	9.424 78
螺距/mm	8.466 7	8.466 7
顶隙系数	0.25	0.25
齿顶高/mm	4.5	3.588
齿根高/mm	2.25	3.162
实际中心距/mm	170	170

图 3 为 6 种不同程度齿轮点蚀故障, C1 为一对健康齿轮组 (没在图中显示); C2 为从动轮的齿 1 齿面 10% 点蚀; C3 为从动轮的齿 1 齿面 30% 点蚀; C4 为从动轮的齿 1 齿面 50% 点蚀; C5

为从动轮的齿 1 齿面 50% 点蚀 + 齿 72 齿面 10% 点蚀;C6 为从动轮的齿 1 齿面 50% 点蚀 + 齿 72 齿面 10% 点蚀 + 齿 2 齿面 10% 点蚀;C7 为从动轮的齿 1 齿面 50% 点蚀 + 齿 72 齿面 30% 点蚀 + 齿 2 齿面 10% 点蚀. 故障是用钻头在从动轮齿上人为制造的,为了模拟齿轮早期点蚀情况,人为制造的故障较小,且制造故障时仅拆除齿轮箱防护罩不对齿轮进行拆装,这样可以避免发生装备误差.

实验包含了 7 种不同类型的齿轮振动数据采集,运行工况为转速 100 r/min,转矩 50 N·m,采集了稳定运行后 12 s 内的振动数据,每组齿轮进行 5 次实验. 图 4 为传感器采集的 7 组齿轮 3 个

方向的振动信号,不同齿轮类型的振动波形有明显的区别. 对比图 4a, 4b, 4c 可以发现 X, Y 轴方向的振动幅值相仿,Z 轴的振动信号幅值最大.

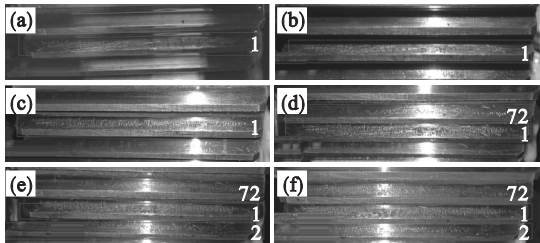


图 3 6 种齿轮点蚀故障
Fig. 3 Six gear pitting faults
(a)—C2;(b)—C3;(c)—C4;(d)—C5;
(e)—C6;(f)—C7.

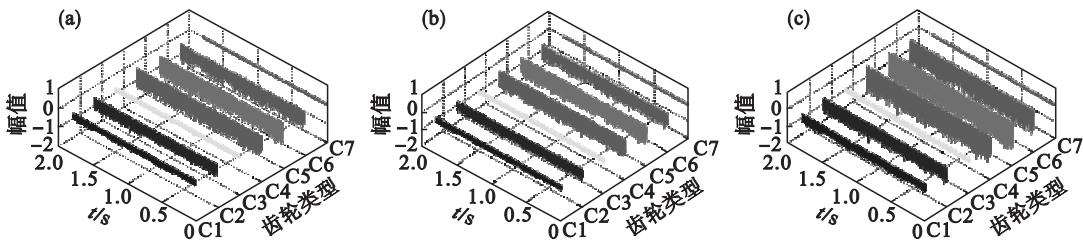


图 4 齿轮箱 3 个方向振动信号
Fig. 4 3-D vibration signals of the gearbox
(a)—X 轴;(b)—Y 轴;(c)—Z 轴.

3 DNN 训练与测试

通过上一节可知 Z 轴的振动信号幅值最大,所以用其训练 DNN 的效果会更好. 7 种齿轮类型,每种进行 5 次数据采集,共 35 组信号. 其中驱动齿轮转速 100 r/min,则从动齿轮转速为 55.6 r/min,大约 1 r/s. 所以从动齿轮旋转一周传感器采集的数据点大概为 10 240 个. 之后将每组信号进行分段,为确保每段数据中都包含故障信息,每段的数据点个数稍大于 10 240,本文选择截取后的每段包含 11 000 个数据点,所以现有样本量为 11 000 × 770.

3.1 DNN 训练

每种齿轮类型分别进行了 5 次数据采集,应用前 4 次采集的数据训练,第 5 次采集数据测试. 训练输入样本为 11 000 × 616,测试输入样本为 11 000 × 154. DNN 网络输入层神经元 11 000 个,输出层神经元 7 个,包含 3 个隐含层. DNN 根据 1.1 节介绍的不断地迭代寻找最优网络参数.

图 5 为有无 PSO 优化的 DNN 训练曲线,观察图 5a 中的局部放大图可知,没有 PSO 优化的网络起始误差略小于 6,之后有一个将近 20 的误

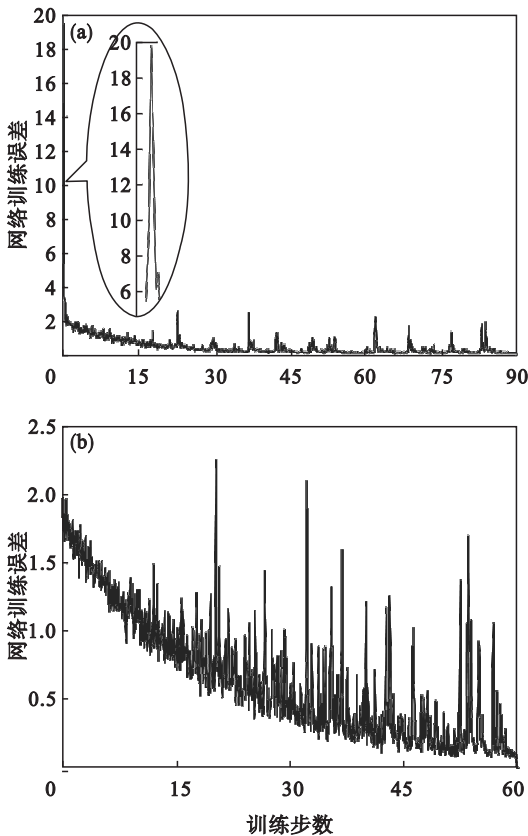


图 5 网络训练误差曲线
Fig. 5 Network training error curve
(a)—DNN 训练;(b)—PSO 优化 DNN 训练.

差波峰,接着误差下降至 2 左右开始较平稳地下降. 观察有 PSO 优化的 DNN 训练误差图 5b, 误差曲线就是从 2 开始下降的,并没有像图 5a 中开始出现较大的误差波动,而且应用 PSO 优化的 DNN 训练所用步数明显较少.

3.2 DNN 测试及结果讨论

训练好的 DNN 网络要用测试数据来测试其诊断效果. 输出层应用了 softmax 分类器,7 个神经元对应 7 种齿轮状况,所以根据输出向量判定齿轮状况方式为输出值最大的神经元对应的齿轮状况即为当前齿轮的状况,表 2 所示为 7 种故障类型测试准确率,图 6 为有无 PSO 优化的不同齿轮类型的误诊率.

表 2 7 种齿轮故障诊断准确率

Table 2 Test accuracy of 7 gear faults

齿轮类型	DNN	PSO-DNN
C1	1	1
C2	0.863 6	1
C3	1	1
C4	0.818 2	1
C5	1	0.954 5
C6	0.50	0.545 5
C7	1	1
平均准确率	0.883 1	0.928 6

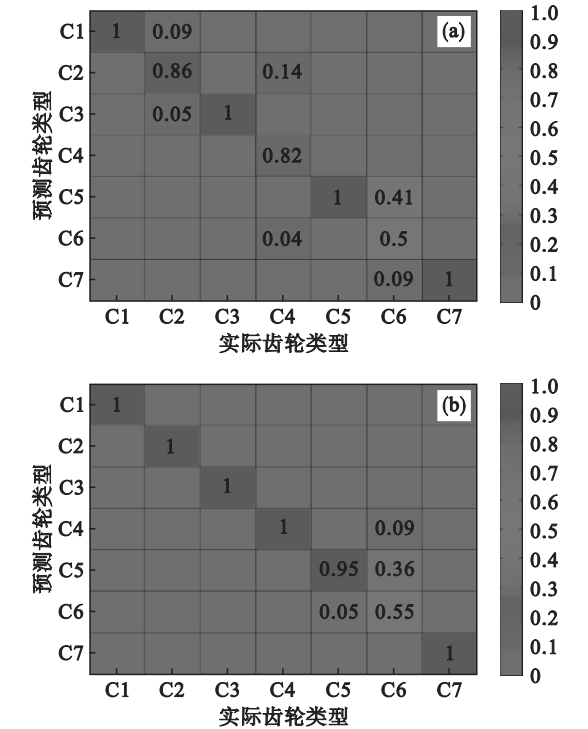


图 6 不同方法诊断 7 种齿轮类型误诊率
Fig. 6 Misdiagnosis rate of different methods
(a)—DNN; (b)—PSO 优化 DNN.

从图 6 中可以看出,误诊一般都发生在相近齿轮类型之间,如图 6a 中 DNN 诊断网络将部分 C2 误诊成 C1 和 C3. 结合表 2 所示的齿轮类型的说明,可以发现 C2 ~ C6 六种齿轮的点蚀程度逐渐增加,实验设定的点蚀程度较小而且相邻齿轮类型的故障情况十分相似,所以就会出现相邻齿轮类型误诊的情况. 总体来说,要分辨的齿轮类型越相近、故障程度越小,发生误诊的可能性就越大,对比图 6a,图 6b 可以发现经过 PSO 优化 DNN 误诊的情况已经明显改善.

输出层采用 softmax 分类器,输出值在 0 ~ 1 之间,判定齿轮状况采用的是输出值最大神经元对应的故障类型. 所以当输出值为 0.5 或 0.99 时齿轮类型判定结果相同,但两种结果对应的网络判别效果有明显差别. 为了更好地判断网络的诊断效果,应用主成分分析法对输出矩阵进行降维,提取前 3 个主成分绘制 3 维图,如图 7 所示. 图中包含 7 种图案对应 7 种齿轮类型,每种图案有 22 个点,共 154 个点对应 154 组测试数据. 相同图案的点越集中,不同图案的点越分散表明网络判别效果越好,还可以根据点的混杂情况判断不同齿轮类型间的误判情况.

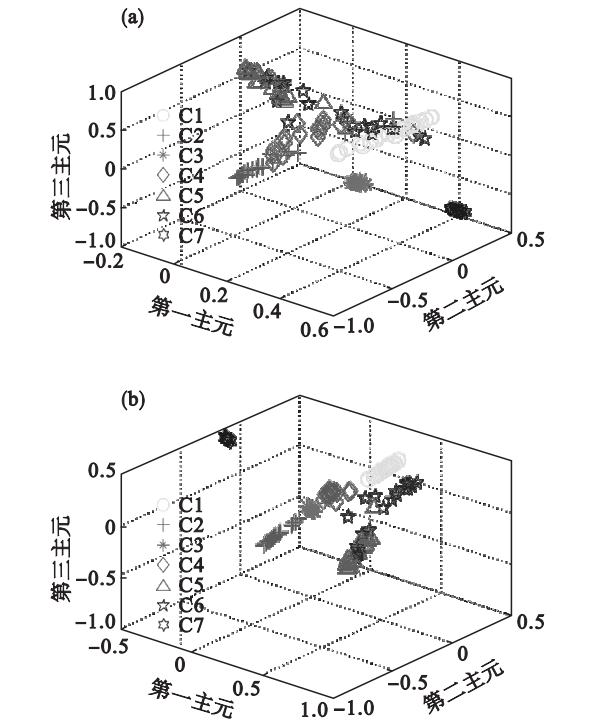


图 7 主成分分析处理输出矩阵的结果
Fig. 7 Results of PCA process output vectors
(a)—DNN; (b)—PSO 优化 DNN.

对比图 7a 和图 7b 可以发现,有 PSO 优化的 DNN 网络诊断效果明显要好. 图 7a 中 C2, C4, C5 和 C6 混杂在一起,之间肯定存在误判情况. 而图

7b 中仅 C6 的部分混杂在 C5 中,误判率较低。

仅仅一次训练和测试并不能完全体现 DNN 网络的诊断效果,图 8 所示为三种方法连续训练 20 次的准确率曲线。其中 PSO 优化后的 DNN 网络诊断效果最好,其次是 DNN 方法,最差的是 FFT-ANN 方法。

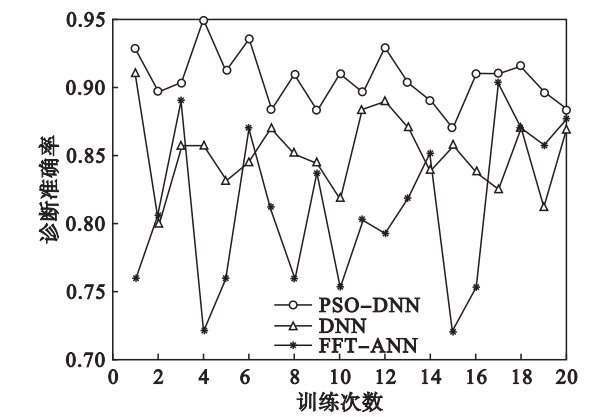


图 8 20 次训练的诊断准确率曲线
Fig. 8 Diagnostic accuracy curves of 20 times' training

全部样本要分成两部分,一部分用于 DNN 训练,剩余的用于测试,训练和测试占总样本比例对诊断效果也有影响。图 9 为训练数据占总样本量比例对 DNN 诊断准确率的影响,可以发现,随着训练样本的增加,DNN 的诊断准确率也随之增加,其中有 PSO 优化的 DNN 网络效果明显要好。

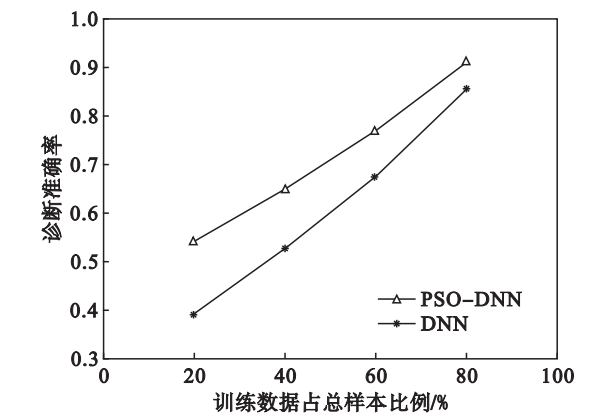


图 9 训练样本量对诊断准确率的影响
Fig. 9 Effect of training sample size on the diagnostic accuracy

4 结 论

本文去掉特征提取环节,直接应用实验采集的振动信号和 DNN 来诊断早期齿轮点蚀故障,

并应用 PSO 算法对 DNN 进行优化。结果表明: PSO 优化 DNN 不仅能增加训练稳定性而且能提高诊断效果,故障诊断率达到 90% 之上。PSO-DNN 的诊断效果和诊断稳定性要好于 DNN,传统的 ANN 诊断效果最差且多次诊断结果波动较大。训练样本占总样本比例越大,训练后的网络诊断效果越好。

参考文献:

[1] 李耀华,王星州. 直升机齿轮箱故障诊断方法[J]. 机械科学与技术,2018,44(3):192-197.
(Li Yao-hua, Wang Xing-zhou. A method for the fault diagnosis of helicopter gearbox [J]. *Mechanical Science and Technology for Aerospace Engineering*, 2018, 44(3):192-197.)

[2] Vernekar K, Kumar H, Gangadharan K V. Gear fault detection using vibration analysis and continuous wavelet transform [J]. *Procedia Materials Science*, 2014, 5(2):1846-1852.

[3] Liu Y B, Yan K G, Zhou Y B, et al. Bispectral analysis for diagnosis of gear pitting fault [C]// World Congress on Intelligent Control and Automation. New York: IEEE, 2008: 6847-6851.

[4] Elasha F, Mba D, Ruizcarcel C. Pitting detection in worm gearboxes with vibration analysis [J]. *Engineering Failure Analysis*, 2014, 42:366-376.

[5] 陈伟,王友仁,孙灿飞,等. 行星齿轮箱耦合调制振动信号分离与故障诊断[J]. 航空动力学报,2018,33(5):1112-1120.
(Chen Wei, Wang You-ren, Sun Can-fei, et al. Coupled modulated vibration signal component separation and fault diagnosis of planetary gearbox [J]. *Journal of Aerospace Power*, 2018, 33(5):1112-1120.)

[6] Samanta B, Al-Balushi K R, Al-Araimi S A. Artificial neural networks and support vector machines with genetic algorithm for bearing fault detection [J]. *Engineering Applications of Artificial Intelligence*, 2003, 16(7/8):657-665.

[7] Dhamande L S, Chaudhari M B. Detection of combined gear-bearing fault in single stage spur gear box using artificial neural network [J]. *Procedia Engineering*, 2016, 144:759-766.

[8] Cheng G, Cheng Y L, Shen L H, et al. Gear fault identification based on Hilbert-Huang transform and SOM neural network [J]. *Measurement*, 2013, 46(3):1137-1146.

[9] 卫晓娟,丁旺才,李宁洲,等. 基于引力搜索 RBF 神经网络的机车齿轮箱故障诊断[J]. 铁道学报,2016,38(2):19-26.
(Wei Xiao-juan, Ding Wang-cai, Li Ning-zhou, et al. Faulty diagnosis of locomotive gearbox based on gravitational search RBF neural network [J]. *Journal of the China Railway Society*, 2016, 38(2):19-26.)

[10] Wu J D, Chan J J. Faulted gear identification of a rotating machinery based on wavelet transform and artificial neural network [J]. *Expert Systems with Applications*, 2009, 36(5):8862-8875.