DOI: 10.3901/JME.2019.07.081

独立自适应学习率优化深度信念网络在轴承故障诊断中的应用研究*

沈长青 汤盛浩 江星星 石娟娟 王 俊 朱忠奎
(苏州大学轨道交通学院 苏州 215131)

摘要: 对机械装备轴承等关键对象的健康状态监测正在步入大数据、智能化时代。传统的轴承故障诊断方法大多数依靠人工提取特征, 这需要依赖于复杂的信号处理方法以及丰富的专业经验积累。深度学习方法作为一种可以学习数据深层次特征的新的机器学习方法, 将其引入机械故障诊断领域, 并对其运行效率、故障识别精度进行提升, 将进一步提高基于深度学习方法的故障诊断领域的实用性。提出一种基于 Nesterov 动量法的独立自适应学习率优化的深度信念网络, 引入 Nesterov 动量法代替传统动量法预测参数下降的位置, 控制参数达到最优点的速度, 避免了传统动量法引起的错过最优点问题; 利用独立自适应学习率在梯度更新时自适应选择下降步长, 加快模型训练, 提高模型的泛化能力。试验结果表明, 在诊断精度上, 相比支持向量机和标准深度信念网络, 提出的方法对不同载荷工况下轴承故障识别均获得了最高的精度; 在运行效率上, 相比现有一些优化算法, 该优化模型能够稳定有效的加快模型训练速度, 提升深度信念网络的泛化能力, 有效地实现轴承故障诊断。

关键词: 机械装备; 故障诊断; Nesterov 动量法; 独立自适应学习率; 深度信念网络

中图分类号: TH17

Bearings Fault Diagnosis Based on Improved Deep Belief Network by Self-individual Adaptive Learning Rate

SHEN Changqing TANG Shenghao JIANG Xingxing SHI Juanjuan
WANG Jun ZHU Zhongkui

(School of Rail Transportation, Soochow University, Suzhou 215131)

Abstract: The health monitoring of key component of mechanical equipment such as bearings is entering the era of big data and intelligence. The traditional method of bearing fault diagnosis depends on manual extraction, which relies on complex signal processing methods and abundant professional experience. As a new machine learning method which can learn deep features from data, the deep learning method is introduced into the field of mechanical fault diagnosis, and the operation efficiency and fault recognition accuracy of mechanical fault diagnosis are improved, which will further improve the practicability of deep learning method in fault diagnosis field. A novel self-adaptive deep belief network with Nesterov momentum (NM-based ADDBN) is proposed. Nesterov momentum is used to replace the traditional momentum method to predict the next position of the parameters, and control parameters' speed towards the optimal values, avoiding missing the optimal values caused by the traditional momentum method. Self-individual adaptive learning rate is used to adaptively select decrease step length for the gradient update, and to speed up the model training and improve the generalization ability of the model. Experimental results show that compared with support vector machine and standard depth belief network, the proposed method obtain the best precision performance for bearing fault recognition under different load conditions. For operational efficiency, the optimization model can effectively and steadily speed up the model training, improve generalization ability of deep belief network, and realize the rolling bearing fault diagnosis effectively compared with current optimization methods.

Key words: mechanical equipment; fault diagnosis; Nesterov momentum; self-individual adaptive learning rate; deep belief network

* 国家自然科学基金资助项目(51875375, 51875376)。20180611 收到初稿,
20181109 收到修改稿

0 前言

滚动轴承作为旋转机械中的关键部件,在航空航天、电力、冶金以及机械等部门中广泛使用。旋转机械的工作状态与滚动轴承有密切关系,据统计,在使用滚动轴承的旋转机械设备中,约 30% 的机械故障是与轴承损伤有关联的。滚动轴承一旦发生意外故障,正常生产进程以及产品质量就会受影响并因此产生巨大的经济损失,严重者甚至危及人身安全,导致重大的灾难性事故。因此,为了提高滚动轴承的可靠性,保证机械设备安全运行,发展有效的故障诊断手段非常重要。

近年来,基于机器学习的方法已经成功运用于轴承故障监测与诊断,例如人工神经网络(Artificial neural network, ANN)^[1]、模糊推断(Fuzzy inference, FI)^[2]和支持向量机(Support vector machines, SVMs)^[3]。上述基于传统机器学习的滚动轴承故障诊断技术首先通过复杂的信息处理技术对采集到的信号进行人工提取特征,然后使用机器学习方法对这些提取到的特征进行分类,进而识别机械健康状态,这需要相应的数学基础和机械故障诊断知识。此外不同的特征具有不同的表达意义,很难人工选取统一的、适用于不同模型的特征,为了获得更好的识别精度就需要花费更多的时间去人工提取适合的特征。深度学习(Deep learning)^[4-6]作为机器学习研究中的一个新领域,近年来被运用于轴承故障领域,包括深度神经网络(Deep neural networks, DNNs)^[7]、卷积神经网络(Convolutional neural networks, CNNs)^[8]和深度信念网络(Deep belief networks, DBNs)^[9]。深度学习通过组合多层非线性学习层模仿人脑的机制解释数据,从中自动提取出数据的深层次特征,结合模式识别进行故障分类,无需复杂的人工提取特征和丰富的故障诊断知识。作为深度学习神经网络的经典模型之一,DBN 通过贪婪逐层无监督学习自动提取数据内在特征,避免人工提取特征带来的复杂性和不确定性,提高分类精度。李巍华等^[10]利用 DBN 直接从轴承振动原始信号提取更加抽象的高层表示,实现轴承故障的分类识别。SHAO 等^[11]运用 DBN 自动从振动信号提取更深层次的特征,实现对感应电动机工作状态的监测。机械装备故障诊断呈现大数据的趋势,针对高维或大数据量的样本,常规深度学习模型在运行效率上难以得到有效保证。为解决此类问题,一些优化算法可被用来加快 DBN 训练速度并提高分类效果,例如自适应次梯度法(Adaptive subgradient method, ASM)^[12]和自适应

时刻估计方法(Adaptive moment estimation, Adam)^[13]已被用于图像识别的深度学习优化,但在机械装备数据深度学习中应用还很少。另一方面,优化方法选择不当可能存在梯度消失现象,反而降低故障诊断精度。

本文利用深度信念网络在自动学习特征上的优势,将深度信念网络用于滚动轴承故障状态识别。针对上述深度学习在机械设备故障诊断领域中面临的挑战,将 Nesterov 动量法与独立自适应学习率算法相结合,提出 IADLR-NM 优化算法,旨在解决遇到维度高、结构复杂或大数据量的样本时深度学习模型的学习速度变慢和学习率难以选择的问题。在此基础上,提出基于 Nesterov 动量法的自适应深度信念网络 NM-based ADDBN,直接从频域信号中提取数据深层特征,实现对滚动轴承的故障诊断,该模型通过比较前后两次梯度大小预测参数下降位置,进而对梯度下降进行正或负补偿,保证参数不会越过最优点;利用独立自适应学习率自适应选择梯度下降步长,加快模型学习速度,有效提高模型的泛化能力。

1 深度信念网络

深度信念网络是一种概率生成模型,它是由多个受限玻尔兹曼机(Restricted boltzmann machines, RBMs)堆叠而成。通过对大量无标签数据进行无监督学习,可以自动提取信号的深层特征,实现对原始数据的降维,通过结合 Softmax 回归模型就可以实现对轴承的故障分类。

1.1 受限玻尔兹曼机

受限玻尔兹曼机由一层可视层 v 和一层隐藏层 h 组成,分别包含 n 个可视神经元 $v=\{v_1, v_2, \dots, v_n\}$ 和 m 个隐藏神经元 $h=\{h_1, h_2, \dots, h_m\}$ 。

如图 1 所示,在可视层与隐藏层层内不存在神经元连接,可视层与隐藏层通过权重 W 连接,神经元只有激活和未激活两种状态,分别用 0 和 1 表示。那么,对于一组确定的神经元状态 (v, h) ,RBM 的能量函数可以定义为

$$E(v, h; \theta) = -\sum_{i=1}^m \sum_{j=1}^n w_{ij} v_i h_j - \sum_{i=1}^m a_i v_i - \sum_{j=1}^n b_j h_j \quad (1)$$

式中, $\theta=\{W, a, b\}$, v_i 表示可视层第 i 个神经元的状态, h_j 表示隐藏层第 j 个神经元的状态, a_i 表示可视层第 i 个神经元的偏置, b_j 表示隐藏层第 j 个神经元的偏置, w_{ij} 表示可视层第 i 个神经元与隐藏层第 j 个神经元的连接权重。

图 1 受限玻尔兹曼机结构

可视层与隐藏层神经元的联合概率分布定义为

$$P(v, h; \theta) = \frac{1}{Z(\theta)} \exp(-E(v, h; \theta)) \quad (2)$$

式中， $Z(\theta)$ 为归一化因子，定义为

$$Z(\theta) = \sum_v \sum_h \exp(-E(v, h; \theta)) \quad (3)$$

由于可视层与隐藏层层内不存在神经元连接，可视层神经元和隐藏层神经元的条件概率都是条件独立的，有

$$P(h | v; \theta) = \frac{P(v, h; \theta)}{P(v; \theta)} = \prod_j P(h_j | v; \theta) \quad (4)$$

$$P(v | h; \theta) = \frac{P(v, h; \theta)}{P(h; \theta)} = \prod_i P(v_i | h; \theta) \quad (5)$$

由于神经元只存在两种状态，进一步得到激活概率。

$$P(h_j = 1 | v; \theta) = \frac{\sum_{h_{k \neq j}} p(h_j = 1, h_{k \neq j}, v; \theta)}{\sum_h p(h, v; \theta)} = \sigma \left(\sum_{i=1}^n w_{ij} v_i + b_j \right) \quad (6)$$

$$p(v_i = 1 | h; \theta) = \frac{\sum_{v_{k \neq i}} p(v_i = 1, v_{k \neq i}, h; \theta)}{\sum_v p(v, h; \theta)} = \sigma \left(\sum_{j=1}^m w_{ij} h_j + a_i \right) \quad (7)$$

式中， σ 表示 sigmoid 函数，定义为

$$\sigma = 1 / (1 + e^{-x}) \quad (8)$$

为了最大程度的拟合输入数据，使用随机梯度上升法最大化 RBM 的对数似然函数，获取最优参数 θ^*

$$\theta^* = \arg \max_{\theta} L(\theta) = \arg \max_{\theta} \sum_v \ln P(v | \theta) \quad (9)$$

则对数似然函数关于 θ 各个参数的偏导数为

$$\begin{aligned} \frac{\partial \ln p(v; \theta)}{\partial w_{ij}} &= \langle v_i h_j \rangle_{\text{data}} - \langle v_i h_j \rangle_{\text{model}} \\ \frac{\partial \ln p(v; \theta)}{\partial a} &= \langle v_i \rangle_{\text{data}} - \langle v_i \rangle_{\text{model}} \\ \frac{\partial \ln p(v; \theta)}{\partial b} &= \langle h_j \rangle_{\text{data}} - \langle h_j \rangle_{\text{model}} \end{aligned} \quad (10)$$

式中， $\langle \cdot \rangle_{\text{data}}$ 表示训练数据已知时，隐藏层的概率分布，该项比较容易计算， $\langle \cdot \rangle_{\text{model}}$ 表示可视层与隐藏层的联合分布，由于 $Z(\theta)$ 的存在，联合分布计算困难。因此，HINTON 等提出使用对比散度 Contrastive divergence (CD) 算法估计 $\langle \cdot \rangle_{\text{model}}$ 这一项的近似值。在 CD 算法中，先利用式(6)和 Gibbs 抽样计算所有隐藏层神经元的二值状态，然后使用式(7)和 Gibbs 抽样确定可视层神经元状态，进而对下一步可视层产生了一个重构(reconstruction)，利用得到的可视层重构数据再次计算隐藏层神经元的二值状态，重复上述步骤，当抽样步数 k 足够大时，就可以得到 RBM 模型所定义的联合分布，即 $CD-k$ 。

下面为 RBM 基于 $CD-k$ 算法的主要步骤。

算法： k 步对比散度

输入：训练数据 S ，RBM 模型可视层神经元个数为 m ，隐藏层神经元个数为 n

输出：近似梯度 Δw_{ij} 、 Δa_i 和 Δb_j $\Delta w_{ij} = \Delta a_i = \Delta b_j = 0$

For all the $s \in S$ do

$v^{(0)} \leftarrow s$

For $t = 0, 1, \dots, k-1$ do

For $j = 1, 2, \dots, n$ do sample $h_j^{(t)} \sim P(h_j | v^{(t)})$

For $i = 1, 2, \dots, m$ do sample

$v_i^{(t+1)} \sim P(v_i | h^{(t)})$

For $i = 1, 2, \dots, m, j = 1, 2, \dots, n$ do

$\Delta w_{ij} \leftarrow \Delta w_{ij} + \eta_w (P(h_j = 1 | v^{(0)}) \cdot v_i^{(0)} -$

$P(h_j = 1 | v^{(k)}) \cdot v_i^{(k)})$

$\Delta a_i \leftarrow \Delta a_i + \eta_a (v_i^{(0)} - v_i^{(k)})$

$\Delta b_j \leftarrow \Delta b_j + \eta_b (P(h_j = 1 | v^{(0)}) - P(h_j = 1 | v^{(k)}))$

HINTON 指出当使用训练样本初始化 $v^{(0)}$ ，仅需要很少的抽样步数就可以得到足够好的近似，一般取 $k=1$ 。此时，参数 $\theta = \{W, a, b\}$ 的更新准则变成

$$\Delta w_{ij} \leftarrow \Delta w_{ij} + \eta_w (\langle v_i h_j \rangle^0 - \langle v_i h_j \rangle^1)$$

$$\Delta a_i \leftarrow \Delta a_i + \eta_a (\langle v_i \rangle^0 - \langle v_i \rangle^1)$$

$$\Delta b_j \leftarrow \Delta b_j + \eta_b (\langle h_j \rangle^0 - \langle h_j \rangle^1) \quad (11)$$

式中， η_w 、 η_a 和 η_b 分别表示连接权重 W 、可视层

偏差值 a 和隐藏层偏差值 b 的学习率, 其值范围在 0 到 1 之间。

1.2 深度信念网络模型

深度信念网络实质上是由若干个 RBM 层层堆叠而成的, 其顶层提取出的数据深层特征连接 Softmax 回归模型用于分类。图 2 所示为三层 RBM 层叠而成的 DBN 模型。首先, 可视层 v 与第一层隐藏层 h_1 组成第一个 RBM 模型; 其次, 第一层隐藏层 h_1 与第二层隐藏层 h_2 组成第二个 RBM 模型; 最后, 第二层隐藏层 h_2 与第三层隐藏层 h_3 组成第三个 RBM 模型。

图 2 深度信念网络结构

DBN 模型训练过程主要分成两个步骤: 预训练和全局微调。在预训练过程中, 使用贪婪逐层无监督算法分别训练每一个 RBM 模型, 具体方法为: 将原始数据输入 DBN 的可视层中训练第一个 RBM 模型, 固定训练好的 RBM 模型参数, 使用其隐藏层神经元状态作为第二个 RBM 模型的输入, 重复上述步骤训练完所有的 RBM 模型。将顶层 RBM 训练后得到隐藏层神经元状态作为 Softmax 回归模型的输入, 结合数据标签有监督的训练 Softmax 模型。在全局微调阶段, 使用有标签数据结合反向传播算法(Back propagation, BP)优化 DBN 模型参数, 提高分类效果。

2 深度信念网络优化

2.1 Nesterov 动量法

随机梯度下降法(Stochastic gradient descent, SGD)常用于似然函数的梯度下降, 然而当面对一个方向的坡度比其他方向都要陡峭的沟壑时, SGD 常常不会沿着沟壑方向加速下降到局部最优值, 只是沿着该方向反复振荡, 逐渐向最优值靠拢, 导致 RBM 参数训练缓慢, 如图 3 所示。

图 3 SGD 沿沟壑梯度下降示意图

动量法(Momentum)^[14]能够有效加快 SGD 通过沟壑时梯度下降的速度并减缓振荡, Momentum 通过将上一次更新的梯度乘以一个动量因子 γ , 再加上此时的梯度作为这次更新的梯度, 让每一次的参数更新方向不仅仅取决于当前参数的梯度, 还受到上一次参数更新方向的影响。

$$\begin{aligned} v_t &= \gamma v_{t-1} + \eta \nabla_{\theta} J(\theta_{t-1}) \\ \theta_t &= \theta_{t-1} - v_t \end{aligned} \quad (12)$$

式中, γ 表示动量因子, 它表示对上一次更新方向的衰减权重, 通常设置为 0.9。若两次梯度的方向一致, 则会加速在这一方向的梯度下降; 若两次梯度方向相反, 则会抑制高曲率方向上的振荡, 如图 4 所示。

图 4 使用 Momentum 后 SGD 沿沟壑梯度下降示意图

然而, Momentum 的梯度下降是盲目的, 它不能判断参数将要下降的位置, 只会一直加速下降, 很容易在坡度在下降转为上升时错过最优解。Nesterov 动量法(Nesterov momentum, NM)^[15]能够有效解决这个问题, 它通过计算 $J(\theta - \gamma v_{t-1})$ 的梯度大概预测下一次参数下降的位置, 而后再做出修正, 帮助参数进行有效的下降。

$$\begin{aligned} v_t &= \gamma v_{t-1} + \eta \nabla_{\theta} J(\theta_{t-1} - \gamma v_{t-1}) \\ \theta_t &= \theta_{t-1} - v_t \end{aligned} \quad (13)$$

为了更好地理解 Nesterov 动量法, 推导 Nesterov 动量法的等效形式, 探讨它与动量法之间的联系。根据式(13), 计算 $\theta_t - \gamma v_t$

$$\begin{aligned} \theta_t - \gamma v_t &= \theta_{t-1} - v_t - \gamma v_t = \\ &= \theta_{t-1} - (\gamma + 1)[\gamma v_{t-1} + \eta \nabla_{\theta} J(\theta_{t-1} - \gamma v_{t-1})] = \\ &= \theta_{t-1} - \gamma v_{t-1} - \gamma^2 v_{t-1} - (\gamma + 1)\eta \nabla_{\theta} J(\theta_{t-1} - \gamma v_{t-1}) \end{aligned} \quad (14)$$

令

$$\hat{\theta}_t = \theta_t - \gamma v_t$$

$$\begin{aligned}\hat{\theta}_{t-1} &= \theta_{t-1} - \gamma v_{t-1} \\ \hat{v}_t &= \gamma^2 v_{t-1} + (\gamma + 1) \eta \nabla_{\theta} J(\hat{\theta}_{t-1})\end{aligned}\quad (15)$$

根据式(14)和式(15)得到

$$\hat{\theta}_t = \hat{\theta}_{t-1} - \hat{v}_t \quad (16)$$

展开 $\hat{v}_t$ 得到

$$\begin{aligned}\hat{v}_t &= (\gamma + 1) \eta \nabla_{\theta} J(\hat{\theta}_{t-1}) + \gamma^2 v_{t-1} = \\ &(\gamma + 1) \eta \nabla_{\theta} J(\hat{\theta}_{t-1}) + \gamma^2 (\gamma v_{t-2} + \eta \nabla_{\theta} J(\theta_{t-2} - \gamma v_{t-2})) = \\ &(\gamma + 1) \eta \nabla_{\theta} J(\hat{\theta}_{t-1}) + \gamma^2 \eta \nabla_{\theta} J(\hat{\theta}_{t-2}) + \\ &\gamma^3 \eta \nabla_{\theta} J(\hat{\theta}_{t-3}) + \gamma^4 \eta \nabla_{\theta} J(\hat{\theta}_{t-4}) + \dots\end{aligned}\quad (17)$$

同理得到 $\hat{v}_{t-1}$

$$\begin{aligned}\hat{v}_{t-1} &= (\gamma + 1) \eta \nabla_{\theta} J(\hat{\theta}_{t-2}) + \gamma^2 \eta \nabla_{\theta} J(\hat{\theta}_{t-3}) + \\ &\gamma^3 \eta \nabla_{\theta} J(\hat{\theta}_{t-4}) + \dots\end{aligned}\quad (18)$$

$$\begin{aligned}\hat{v}_t - \gamma \hat{v}_{t-1} &= (\gamma + 1) \eta \nabla_{\theta} J(\hat{\theta}_{t-1}) - \gamma \eta \nabla_{\theta} J(\hat{\theta}_{t-2}) = \\ &\eta [\nabla_{\theta} J(\hat{\theta}_{t-1}) + \gamma [\nabla_{\theta} J(\hat{\theta}_{t-1}) - \nabla_{\theta} J(\hat{\theta}_{t-2})]]\end{aligned}\quad (19)$$

根据式(16)、(19)得到 Nesterov 动量法的等效形式

$$\begin{aligned}v_t &= \gamma v_{t-1} + \eta \nabla_{\theta} J(\theta_{t-1}) + \gamma \eta [\nabla_{\theta} J(\theta_{t-1}) - \nabla_{\theta} J(\theta_{t-2})] \\ \theta_t &= \theta_{t-1} - v_t\end{aligned}\quad (20)$$

可以看到, 相比于动量法, Nesterov 动量法在更新梯度时增加了一项 $\gamma \eta [\nabla_{\theta} J(\theta_{t-1}) - \nabla_{\theta} J(\theta_{t-2})]$, 如果当前计算得到的梯度大于前一次的梯度, 参数在加速下降, 模型认为下一次梯度下降将进一步加速, 对梯度下降进行正补偿; 若后一次梯度相对前一次梯度来说下降幅度变小, 参数在减缓下降速度, 模型认为下一次梯度下降将进一步减速, 对梯度下降进行负补偿, 保证参数始终以合适的速度下降至最优点以避免错过最优值。从数学的角度来说, $\gamma \eta [\nabla_{\theta} J(\theta_{t-1}) - \nabla_{\theta} J(\theta_{t-2})]$ 可以认为是目标函数的近似二阶导数, Nesterov 动量法在梯度更新时考虑了目标函数的近似二阶导数。

2.2 独立自适应学习率

在训练过程中, 学习率是一个非常重要的参数。学习率太大会引起参数 θ 的解在最优值附近振荡发散, 导致训练误差增大, 模型分类效果不理想。一个小的学习率固然可以避免上述这些问题, 同时它也会导致收敛速度慢, 需要花费更多的时间进行训练。为了获得理想的分类效果和训练速度, 一种独立自适应学习率(Self-individual adaptive learning rate, IADLR)^[16]被用来训练 RBM 模型。对于每一个连接权重 w_{ij} 更新时, 有

$$h_{ij}(k) = \begin{cases} h_{ij}(k-1) + \alpha & (\text{grad}_{ij}^k \text{grad}_{ij}^{k-1}) > 0 \\ h_{ij}(k-1) \times (1 - \alpha) & (\text{grad}_{ij}^k \text{grad}_{ij}^{k-1}) < 0 \end{cases} \quad (21)$$

$$\Delta w_{ij}^k = \eta \cdot h_{ij}^k \cdot \nabla_{\theta} J(\theta_{t-1} - \gamma v_{t-1}) \quad (22)$$

式中, grad_{ij}^k 表示第 k 次训练后权重 w_{ij} 的梯度, $h_{ij}(0) = 1$ 。若当前权重梯度方向与上一次梯度一致, 则 $h_{ij}(t)$ 增加 α ; 若当前权重梯度方向与上一次梯度相反, 则 $h_{ij}(t)$ 缩小 $1 - \alpha$ 倍。参数 α 取值应该较小, 例如 0.1, 这样才能保证当学习率连续增大时, 学习率不会增长过快, 模型越过最优解, 或学习率连续减少时, 学习率不会快速衰减, 导致梯度无法下降。为了避免梯度消失问题, $h_{ij}(t)$ 范围应该限制在 0.01~100。

2.3 基于独立自适应学习率和 Nesterov 动量法的深度信念网络

在经典深度信念网络模型的基础上, 结合独立自适应学习率和 Nesterov 动量法, 建立基于 Nesterov 动量法的自适应深度信念网络(NM-based ADDBN)。图 5 所示为该深度信念网络训练流程图。

图 5 深度信念网络训练流程图

利用贪婪逐层无监督算法和无标签数据训练每

一个 RBM 完成深度信念网络的初始化, 然后使用深度信念网络学习到的数据深度特征训练 Softmax 回归模型, 最后采用 BP 算法和共轭梯度法完成整个深度信念网络的微调。

3 试验验证

3.1 数据描述

为验证基于 Nesterov 动量法的自适应深度信念网络对轴承故障诊断的效果, 利用轴承故障试验平台采集各种故障下的轴承振动信号。如图 6 所示, 该试验平台包括驱动电动机、正常轴承、测试轴承和加载装置等。加速度传感器放置在测试轴承 12 点方向。测试轴承型号为 6205-2RS SKF, 通过线切割在轴承表面设置宽度为 0.2 mm 的单点故障, 故障类型包括内圈 (Outer race, OR) 故障、滚动体 (Ball) 故障和外圈故障 (Inner race, IR)。电动机转速为 961 r/min, 分别在载荷 0 kN 和载荷 3 kN 两个工况下采集振动信号, 信号采样频率为 10 kHz, 四种状态的振动信号如图 7、8 所示。

图 6 滚动轴承振动数据采集试验平台

图 7 0 kN 载荷下轴承不同状态的振动信号

图 8 3 kN 载荷下轴承不同状态的振动信号

选取 1 248 个采样点为一个时域样本长度, 则频域样本长度为 624 个采样点, 每种状态的样本个数为 400 个。试验中, 随机选取每类样本总数的 3/4 作为训练数据, 剩下的 1/4 作为测试样本。对不同载荷下每类状态样本集标记标签用于 Softmax 回归模型的分类, 对不同载荷下同一类状态样本集使用相同的标签, 试验样本集详细信息如表 1 所示。

表 1 轴承试验样本集描述

轴承状态	训练样本	测试样本	标签
正常(0, 3 kN)	300	100	1
内圈故障(0, 3 kN)	300	100	2
滚动体故障(0, 3 kN)	300	100	3
外圈故障(0, 3 kN)	300	100	4

3.2 诊断结果与对比分析

为了减少随机因素的影响, 重复进行 10 次试验, 取其平均结果作为最终识别结果。本试验中, NM-ADDBN 模型包括一层可视层、两层隐藏层和一层分类层, 其中可视层神经元数目为 624 个, 第一层隐藏层的神经元数目为 500, 第二层隐藏层神经元数目为 200, 分类层类别数为 4, 初始化各层连接权重 $W \sim N(0, 0.01)$, 在预训练过程中, 第一个 RBM 的自适应学习率初始化为 0.096, 第二个 RBM 的自适应学习率初始化为 0.016, 添加 L2 权重衰减项正则化梯度, 惩罚大的权重项, 衰减系数为 0.000 1。设置 IADLR-NM 优化算法参数 $\alpha = 0.1$ 、 $\gamma = 0.9$ 。在微调过程中, 每次迭代完成后输入测试样本, 计算测试误差和精度, 观察模型的泛化能力。

试验结果如表 2 所示, 可以看到, 对于不同载荷工况下的振动信号, NM-based ADDBN 都很好地

识别轴承的故障状态，十次试验结果都为 100%。为了更好地说明提出方法的有效性，使用与试验所用 NM-based ADDBN 结构相同的标准深度信念网络(Standard DBN)以及支持向量机(SVM)对试验数据进行故障识别。试验结果也示于图 8 中，可以看到，标准 DBN 对 0 kN 载荷工况下的信号识别较好，平均识别率为 99.93%，而对于 3 kN 载荷工况下的信号，标准 DBN 的识别精度不能保证，平均测试精度为 91.05%。对于 SVM，两种载荷工况下的振动信号的测试精度整体偏低，不能很好地识别出轴承健康状态，其中更高载荷工况的信号取得了更差的精度。

表 2 振动信号识别结果(%)

试验 次数	SVM		Standard DBN		NM-ADDBN	
	载荷	载荷	载荷	载荷	载荷	载荷
	0 kN	3 kN	0 kN	3 kN	0 kN	3 kN
1	50	42.65	100	79.5	100	100
2	50	42.65	100	98.75	100	100
3	50	42.65	99.75	99.25	100	100
4	50	42.65	100	75.5	100	100
5	50	42.65	100	97.5	100	100
6	50	42.65	99.75	99.75	100	100
7	50	42.65	100	88.25	100	100
8	50	42.65	100	97.25	100	100
9	50	42.65	100	78	100	100
10	50	42.65	99.75	96.75	100	100
平均	50	42.65	99.93	91.05	100	100

为了更好地验证 Nesterov 动量法和独立自适应学习率对深度信念网络模型泛化能力的提升，使用 Mometum, Adadelata^[17]和 Adam 三种方法与本文提出的方法进行比较。计算十次试验结果的平均值作为最终结果，如表 3 所示。可以看到，对于不同载荷工况下的振动信号，IADLR-NM 和 Adadelata 获得了最好的测试精度 100%，而 Adam 获得了最低的测试精度 80.45%和 67.95%。图 9 所示为四种算法的测试误差，对于 0 kN 载荷工况下的信号，IADLR-NM 得到的测试误差在前期迭代中相比于 Adadelata 较慢，在第十二次迭代之后，IADLR-NM 取得更小的误差。对于 3 kN 载荷工况下的信号，IADLR-NM

表 3 四种算法的平均分类结果(%)

方法	载荷 0 kN		载荷 3 kN	
	训练	测试	训练	测试
Momentum	99.75	99.93	93.69	91.05
Adam	81.79	80.45	67.96	67.95
Adadelata	100	100	100	100
IADLR-NM	100	100	100	100

优化后的模型得到的测试误差收敛最快并且稳定，获得的误差小。在相同数量的样本下，NM-ADDBN 能够提取出数据更具代表性的深层特征，提高了网络的泛化能力。

(a) 载荷为 0 kN

(b) 载荷为 3 kN

图 9 四种优化算法的测试结果

4 结论

基于 Nesterov 动量法的自适应深度信念网络可以自动学习到数据的深层特征，有效地实现对轴承的故障诊断。

(1) 深度信念网络可以直接从频率信号中提取出数据的深层特征，能够有效地应用于故障分类，无需复杂的信号处理方法和故障诊断知识。

(2) 提出的基于 Nesterov 动量法的自适应深度信念网络相比传统的 SVM 浅层网络，具有更强的分类能力，能更有效地实现轴承的故障诊断。

(3) Nesterov 动量法和独立自适应学习率可以有效地提升深度信念网络的泛化能力，使得提取出的数据深层特征更具有代表性，有效地实现对滚动轴承的故障诊断。

参 考 文 献

- [1] BANGALORE P, TJEMBERG L B. An artificial neural network approach for early fault detection of gearbox bearings[J]. IEEE Transactions on Smart Grid, 2017, 6(2): 980-987.
- [2] MORETO M, CIESLAK D A G. Fuzzy inference system to automatic fault classification in power plants[J]. IEEE Latin America Transactions, 2016, 14(2): 746-751.
- [3] SHEN C, WANG D, KONG F, et al. Recognition of rolling bearing fault patterns and sizes based on two-layer support vector regression machines[J]. Smart Structures & Systems, 2014, 13(3): 453-471.
- [4] HINTON G E, SALAKHUTDIONOV R R. Reducing the dimensionality of data with neural networks[J]. Science, 2006, 313(5786): 504-507.
- [5] BENGIO Y. Learning deep architectures for AI[J]. Foundations & Trends® in Machine Learning, 2009, 2(1): 1-55.
- [6] 雷亚国, 贾峰, 周昕, 等. 基于深度学习理论的机械装备大数据健康监测方法[J]. 机械工程学报, 2015, 51(21): 49-56.
- LEI yaguo, JIA feng, ZHOU xin, et al. A deep learning-based method for machinery health monitoring with big data [J]. Journal of Mechanical Engineering, 2015, 51(21): 49-56.
- [7] JIA F, LEI Y, LIN J, et al. Deep neural networks: A promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data[J]. Mechanical Systems & Signal Processing, 2016, 72-73: 303-315.
- [8] GUO X, CHEN L, SHEN C. Hierarchical adaptive deep convolution neural network and its application to bearing fault diagnosis[J]. Measurement, 2016, 93: 490-502.
- [9] SHAO H, JIANG H, WANG F, et al. Rolling bearing fault diagnosis using adaptive deep belief network with dual-tree complex wavelet packet[J]. Isa Transactions, 2017: 187-201.
- [10] 李巍华, 单外平, 曾雪琼. 基于深度信念网络的轴承故障分类识别[J]. 振动工程学报, 2016, 29(2): 340-347.
- LI Weihua, SHAN Waiping, ZENG Xueqiong. Bearing fault identification based on deep belief network[J]. Journal of Vibration Engineering, 2016, 29(2): 340-347.
- [11] SHAO S, SUN W, WANG P, et al. Learning features from vibration signals for induction motor fault diagnosis[C]// International Symposium on Flexible Automation. IEEE, 2016: 71-76.
- [12] DUCHI J, HAZAN E, SINGER Y. Adaptive subgradient methods for online learning and stochastic optimization[J]. Journal of Machine Learning Research, 2011, 12(7): 2121-2159.
- [13] KINGMA D, BA J. Adam: A method for stochastic optimization [J]. Computer Science, 2015, 1: 1-15.
- [14] TAN W, ZHAO C, WU H. Intelligent alerting for fruit-melon lesion image based on momentum deep learning[J]. Multimedia Tools & Applications, 2016, 75(24): 1-21.
- [15] BOTEV A, LEVER G, BARBER D. Nesterov's accelerated gradient and momentum as approximations to regularised update descent[C]// International Joint Conference on Neural Networks. IEEE, 2017: 1.
- [16] TIELEMAN T, HINTON G. Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude[J]. COURSE: Neural Networks for Machine Learning, 2012, 4(2): 26-31.
- [17] ZEILER M D. ADADELTA: An adaptive learning rate method[J]. arXiv Preprint arXiv: 2012, 1: 1212.5701.

作者简介: 沈长青, 男, 1987 年出生, 博士, 副教授, 硕士研究生导师。主要研究方向为机械状态信号处理、监测与智能诊断。

E-mail: cqshen@suda.edu.cn

朱忠奎(通信作者), 男, 1974 年出生, 博士, 教授, 博士研究生导师。主要研究方向为车辆系统动力学与控制、机械故障诊断。

E-mail: zhuzhongkui@suda.edu.cn