

基于小波神经网络的齿轮箱故障诊断研究

汪鲁才, 彭滔, 张颖

WANG Lu-cai, PENG Tao, ZHANG Ying

湖南师范大学 工学院, 长沙 410081

Polytechnic College, Hunan Normal University, Changsha 410081, China

E-mail: pengtao20005@126.com

WANG Lu-cai, PENG Tao, ZHANG Ying. Gearbox fault diagnosis based on wavelet neural network. *Computer Engineering and Applications*, 2007, 43(28): 203-205.

Abstract: This paper focuses on the system structure and algorithms of Wavelet Neural Network. For various features of gear vibrating signals in frequency domain, feature vectors are extracted as inputs of the Wavelet Neural Network which are capable of mapping the feature vectors to the corresponding fault modes. Based on this algorithm, a gear fault diagnosis model has been designed. Simulation results indicate this model, compared with the conventional BP neural network model, can remarkably reduce the training time. It is feasible for mechanical fault diagnosis.

Key words: wavelet analysis; neural network; fault diagnosis

摘要: 论述了小波神经网络的系统结构及算法, 并根据齿轮振动信号的频域变化特征, 提取特征向量作为输入, 利用小波神经网络建立特征向量与故障模式之间的映射关系, 建立了基于该算法的齿轮故障诊断模型。仿真结果表明: 与传统的 BP 神经网络相比, 该模型显著缩短了训练时间。该小波神经网络进行机械故障诊断是有效的。

关键词: 小波分析; 神经网络; 故障诊断

文章编号: 1002-8331(2007)28-0203-03 **文献标识码:** A **中图分类号:** TP18

齿轮及齿轮箱作为机械设备中一种必不可少的连接和传递动力的通用零部件, 在现代工业设备中得到了广泛的应用。在齿轮工作过程中, 齿轮、轴承和轴都会产生振动信号, 当零件发生故障时, 振动信号产生不同形式的变化, 利用振动信号携带的运行状态信息进行故障模式的识别是机械故障诊断中常用的方法, 其信号处理的主要数学工具是傅立叶变换^[1]。小波分析是 20 世纪 80 年代在傅立叶分析的基础上发展起来的一种强有力的数学工具, 作为一种时频域分析方法, 小波分析比傅立叶分析有着突出的优点。人工神经网络具有强大的学习能力、多输入并行处理能力、非线性逼近和容错能力以及自适应能力, 能出色完成模式识别任务。而信号分析和模式识别正是设备故障诊断所依赖的两个关键技术。因此, 研究如何将二者完美地结合起来^[2], 应用于旋转机械的故障诊断, 具有广泛的应用价值。

1 小波变换原理

小波分析的基本思想是用一族函数去表示或逼近一信号或函数。所谓小波, 就是指具有零均值、在时域和频域内能量局部化的函数, 其波形表现为两端衰减为零的小波形。设基小波函数 $\psi(t)$, 基小波经尺度伸缩和时间平移便得到一个子波族, 其形式为^[3]:

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}} \psi\left(\frac{t-b}{a}\right) \quad (1)$$

式中: a 为尺度参数; b 为时间位置参数。 a 用来调整子波覆盖的频率范围, b 用来调整子波的时域位置, 系数 $1/\sqrt{a}$ 用来实现子波能量的归一化。对一个时变信号 $x(t)$, 其连续小波变化定义为:

$$W(a,b) = \frac{1}{\sqrt{a}} \int_c \psi_{a,b}^*(t) x(t) dt \quad (2)$$

式中 $\psi_{a,b}^*(t)$ 为 $\psi_{a,b}(t)$ 的复共轭; c 为复参数; 尺度参数 a 和时间位置参数 b 在取值范围内连续变化。尺度参数 a 不仅影响时频窗口在频率轴上的位置, 也影响窗口的形状 (但窗口面积固定)。这样小波变换对不同的频率在时域上的取样步长是可调节的, 即在低频时小波变换的时间分辨率较低, 而频率分辨率较高; 在高频时小波变换的时间分辨率较高, 而频率分辨率较低, 这正符合低频信号变化缓慢而高频信号变换迅速的特点。这便是它优于经典的傅立叶变换和短时傅立叶变换的地方。

2 小波神经网络模型及算法实现

2.1 模型结构

由于小波函数构造的多样性和复杂性, 使得小波与神经网络

基金项目: 湖南省自然科学基金(the Natural Science Foundation of Hunan Province of China under Grant No.04JJ40042)。

作者简介: 汪鲁才(1969-), 男, 教授, 主要研究方向: 模式识别与智能系统, 图像处理, 卫星图像信息融合和语音信号处理技术; 彭滔(1982-), 女, 硕士研究生, 主要研究方向: 人工智能。

网络的结合有很多种形式。按小波与神经网络结合的紧密程度可分为两种：一种是松散型结合，即小波分析作为神经网络的前置处理手段。用小波分析对信号进行预处理，即以小波空间作为模式识别的特征空间，通过将小波基与信号的内积进行加权和来实现信号的特征提取，然后将提取的特征向量送入神经网络处理。另一种是紧致型结合，指用小波函数或尺度函数作为激励函数，形成神经元，直接融合。本文采用了后一种方式，选用图 1 所示的小波神经网络拓扑结构。

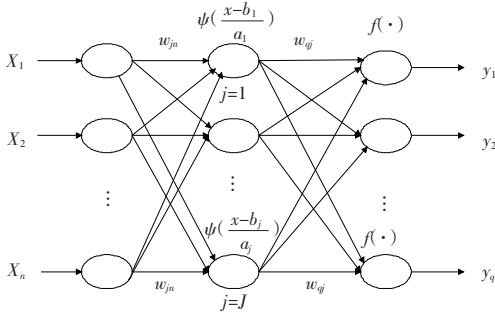


图 1 小波神经网络结构图

网络的输入层有 n 个神经元，输出层有 q 个神经元，中间隐含层有 j 个神经元，网络中输入模式向量 $X_i=[X_1^p, X_2^p, \dots, X_n^p]$ ，实际输出向量 $Y_i=[Y_1^p, Y_2^p, \dots, Y_n^p]$ ， w_{jn} 为输入层和隐含层之间的连接权值， w_{qj} 为隐含层和输出层之间的连接权值，隐含层和输出层节点的输入是前一层节点的输出的加权和，每个节点的激励程度由它的激活函数来决定， a_j 为第 j 个隐层节点的伸缩系数， b_j 为第 j 个隐层节点的平移系数。

2.2 基小波的选择

对于小波网络中小波函数的选择，目前还没有统一的理论或方法来确定选用哪种类型的小波函数或尺度函数适合作为实际不同网络中神经元的激励函数。但是，小波变换的小波系数为如何选择小波基提供依据。另外，还要根据信号处理的目的来决定尺度的大小。因此根据经验和实际的不同情况本文选取 Mexican Hat (mexh) 小波^[4]

$$\psi(x) = \left(\frac{2}{\sqrt{3}}\pi^{-\frac{1}{4}}\right)(1-x^2)e^{-\frac{x^2}{2}} \quad (3)$$

作为网络的神经元响应函数。这里需要说明的是，虽然 Mexican Hat 小波函数不具有正交性，但因这一小波网络仅仅是用于故障诊断(分类)，不是用于严格的逼近信号或函数(即重构)，小波的正交性并不是很重要。

2.3 算法实现

小波网络和三层 BP 网络很相似，所以小波网络的权值的调整规则可以借用 BP 算法^[5]。BP 算法的实质是梯度下降法。其基本思想是以神经网络期望的输出与网络的实际输出的误差平方和为学习目标函数，根据使其最小化的原则来调整网络的权值。为此定义如下的误差函数：

$$E(W) = \frac{1}{2} \| Y(k) - \hat{Y}(W, k) \|^2 \quad (4)$$

上式中 k 为离散时间变量， W 为网络所有权值组成的向量， $Y(k)$ 为期望的网络输出向量， $\hat{Y}(W, k)$ 为网络的实际输出向量， $\|\cdot\|$ 表示向量的 Euclidean 范数。梯度下降法就是沿着

$E(W)$ 的负梯度方向不断修正 W 的值，直至使 $E(W)$ 达到最小值。用数学公式可以表示为：

$$W(k+1) = W(k) + \eta(k) \frac{\partial E(W)}{\partial W} \Big|_{W=W(k)} \quad (5)$$

其中 $\eta(k)$ 是控制权值调整速度的变量，通常与计算的步长有关。

如图 1 的小波神经网络模型可以得到：

$$y_q^p(t) = f \left(\sum_{j=1}^N w_{qj} \psi \left(\frac{\sum_{n=1}^N w_{jn} x_n^p + b_{1j}}{a_j} \right) - b_{2j} \right) + b_{2q} \quad (6)$$

其中 b_{1j} 为第 j 个隐层节点的阈值， b_{2q} 为第 q 个输出层节点的阈值， p 是输入样本的模式个数。为了方便推导用大写的 X 表示输入，小写的 x 表示输出。则式(6)重写如下：

$$y_q^p(t) = f(X_j^p) = f \left(\sum_{j=1}^J w_{qj} x_j^p + b_{2q} \right) \quad (7)$$

选取 Mexican Hat 小波作为网络的神经元响应函数，则

$$\psi'(x) = \left(\frac{2}{\sqrt{3}}\pi^{-\frac{1}{4}} \right) e^{-\frac{x^2}{2}} e^{-\frac{x^2}{2}} + \frac{x^2}{2} \psi(x) \quad (8)$$

按照梯度下降法的思想，定义误差函数：

$$E_{\text{总}} = \sum_{p=1}^P \sum_{q=1}^Q (y_q^p - \bar{y}_q^p)^2 \quad (9)$$

则分别计算如下：

$$(1) \Delta w_{qj} = -\eta \frac{\partial E}{\partial w_{qj}} = -\eta \frac{\partial E}{\partial X_q^p} \times \frac{\partial X_q^p}{\partial w_{qj}} \quad (10)$$

$$\text{令 } \delta_q = -\frac{\partial E}{\partial X_q^p} = -(x_q^p - \bar{x}_q^p) x_q^p (1 - x_q^p) \quad (11)$$

则 $\Delta w_{qj} = \eta \delta_q x_j^p$ 。

$$(2) \Delta w_{jn} = -\eta \frac{\partial E}{\partial w_{jn}} = -\eta \frac{\partial E}{\partial X_j^p} \times \frac{\partial X_j^p}{\partial w_{jn}} \quad (12)$$

$$\text{令 } \delta_j = -\frac{\partial E}{\partial X_j^p} = \sum_q \frac{\partial E}{\partial X_q^p} \frac{\partial X_q^p}{\partial X_j^p} \psi' \left(\frac{X_j^p - b_j}{a_j} \right) \frac{1}{a_j} - \sum_q \delta_q w_{qj} \psi' \left(\frac{X_j^p - b_j}{a_j} \right) \frac{1}{a_j} \quad (13)$$

则 $\Delta w_{jn} = \eta \delta_j x_n^p$ 。

$$(3) \Delta b_j = -\eta \frac{\partial E}{\partial b_j} = -\eta \frac{\partial E}{\partial X_j^p} \frac{\partial X_j^p}{\partial b_j} \text{, 由式(13)可得:}$$

$$\Delta b_j = -\eta \sum_q \delta_q w_{qj} \psi' \left(\frac{X_j^p - b_j}{a_j} \right) \frac{-1}{a_j} = \eta \delta_j \quad (14)$$

$$(4) \Delta a_j = -\eta \frac{\partial E}{\partial a_j} = -\eta \frac{\partial E}{\partial X_j^p} \frac{\partial X_j^p}{\partial a_j} \text{, 由式(13)(14)可得:}$$

$$\Delta a_j = -\eta \sum_q \delta_q w_{qj} \psi' \left(\frac{X_j^p - b_j}{a_j} \right) \left(\frac{-1}{a_j} \right) \left(\frac{X_j^p - b_j}{a_j} \right) = -\eta \delta_j \frac{X_j^p - b_j}{a_j} \quad (15)$$

参考常规的 BP 算法引入动量系数 α ，网络的各参数按照下面的公式进行调整：

$$w_{qj}(t+1) = w_{qj}(t) - \eta \frac{\partial E}{\partial w_{qj}} + \alpha \Delta w_{qj}$$

$$w_{jn}(t+1) = w_{jn}(t) - \eta \frac{\partial E}{\partial w_{jn}} + \alpha \Delta w_{jn} \quad (16)$$

$$a_j(t+1) = a_j(t) - \eta \frac{\partial E}{\partial a_j} + \alpha \Delta a_j$$

$$b_j(t+1) = b_j(t) - \eta \frac{\partial E}{\partial b_j} + \alpha \Delta b_j$$

另外还要对网络的隐层和输出层的阈值进行调整:

$$b1_j(t+1) = b1_j(t) - \eta \frac{\partial E}{\partial b1_j} + \alpha \Delta b1_j$$

$$b2_q(t+1) = b2_q(t) - \eta \frac{\partial E}{\partial b2_q} + \alpha \Delta b2_q \quad (17)$$

3 故障诊断模型的建立及训练

由于齿轮及齿轮箱本身结构复杂, 工作环境恶劣等原因, 在工作过程中容易受到损害和出现故障。齿轮失效又是诱发机器故障的重要因素。据统计, 传动机械中 80% 的故障是由齿轮引起的, 旋转机械中齿轮故障占其故障的 10% 左右^[6]。为了说明问题, 本文只研究齿轮故障的诊断。对于齿轮的故障, 这里选取了频域中的几个特征量。频域中齿轮故障比较明显的是在啮合频率处的边缘带上。所以, 在频域特征信号的提取中选取了在 2、4、6 档时, 在 1、2、3 轴的边频带族 $f_s \pm n f_c$ 处的幅值 $A_{i,j1}$ 、 $A_{i,j2}$ 、 $A_{i,j3}$, 其中 f_s 表示齿轮的啮合频率, f_c 是轴的转频, $n=1, 2, 3$, $i=2, 4, 6$ 表示档位, $j=1, 2, 3$ 表示轴的序号, 在 2 轴和 3 轴上有两对齿轮啮合, 所以用 1、2 分别表示两个啮合频率。这样一来, 网络的输入就是一个 15 维的向量。这些数据具有不同的单位和量级, 所以在输入神经网络之前应该首先进行归一化处理。输出模式为: F1 无故障(1 0 0); F2 齿根裂纹(0 1 0); F3 断齿(0 0 1)。经归一化处理后的学习样本如表 1 所示(篇幅所限, 只列出其中 3 组数据)^[7]。

4 仿真结果与分析

根据上面的理论和推导, 本文构造了一个输入节点 N 为 15(对应 15 个特征参数)和输出节点 Q 为 3(对应 3 种故障状态)的三层小波神经网络。隐层传递函数取 Mexican Hat 小波基函数。隐层节点数的确定是根据输入和输出节点, 再根据公式 $P=(N+Q)^{0.5}+L$ 得到, 其中 L 为 0 到 10 间的常数, 可得隐含层神经元节点数 P 取 4~14 之间的数。针对故障样本数量, 以及经验公式在实际网络训练中调试选择, 最终最佳的隐层节点数

为 13。附加动量项的引入使得网络在修正其权重和阈值时, 不仅考虑误差在梯度上的作用, 而且考虑在误差曲面上的变换趋势的影响。它减小了学习过程的震荡趋势, 从而改善了收敛性, 找到更优的解。动量因子一般取值为 0.9 左右, 本文取值 0.923。选好了最佳的隐层节点数和优化的动量因子后, 在大量的仿真试验中选用不同的学习率 η 来训练网络。结果表明大的学习率可能会导致系统的不稳定, 但是小的学习速率将可能导致较长的训练时间, 收敛很慢。反复比较, 本文选取优化的学习率 $\eta=0.01$ 。

小波神经网络的结构确定以后, 利用故障征兆域、故障原因之间的关系, 抽取故障样本作为网络的学习对象, 对其进行训练。其学习的过程中, 在选择了适当的初始阈值后, 需经过多次反复调节连接权, 才能最终得到符合要求的小波。其最终输出最大误差如图 2 所示。从仿真的结果来看, 在得到合适的网络初始参数的情况下, 经过 37 次的训练, 网络就收敛于样本, 此时其最大误差不超过 0.01(在程序中给定)。为了说明问题, 同时用 BP 网络做了大量的仿真试验。采用与小波网络相类似的网络参数, 在最大误差要求为 0.01 时, 需经过 2 866 步才能达到收敛, 如图 3。对 BP 算法采取了一系列的优化措施, 使其收敛速度大大改进, 但是改进的 BP 仍然需要 147 步才能收敛, 如图 4。经过大量的试验可以看到 BP 网络收敛通常需要的步数大约在 110~3 000 步之间, 而小波神经网络通常 30~50 步之间即可达到满足要求的最大误差。很明显, 相同条件下, 小波神经网络具有较快的收敛速度, 这主要是由于小波函数具有快速衰减性, 小波网络属于局部逼近网络, 而通常的 BP 网络是一种全局网络。

5 结论

小波变换具有时频局部特性和变焦特性, 而神经网络又有自学习、自适应、鲁棒性、容错性和推广能力等特性, 如何把两者的优势结合起来, 一直是人们关注的问题。本文针对齿轮常见故障的振动信号建立征兆表, 作为小波网络的输入, 训练网络, 得到输出, 证明小波网络用于故障诊断的有效性。并将小波网络与 BP 网络做比较, 表明该算法极大地提高了运算速度,

表 1 齿轮箱各特征参数学习样本

序号	特征样本												状态			
1	0.228 6	0.129 2	0.072 0	0.159 2	0.133 5	0.073 3	0.115 9	0.094 0	0.052 2	0.134 5	0.009 0	0.126 0	0.361 9	0.069 0	0.182 8	F1
2	0.260 3	0.171 5	0.070 2	0.271 1	0.149 1	0.133 0	0.096 8	0.191 1	0.254 5	0.087 1	0.006 0	0.179 3	0.100 2	0.078 9	0.090 9	F2
3	0.172 9	0.234 7	0.182 9	0.181 1	0.292 2	0.065 5	0.077 4	0.227 3	0.205 6	0.092 5	0.007 8	0.185 2	0.350 1	0.168 0	0.266 8	F3

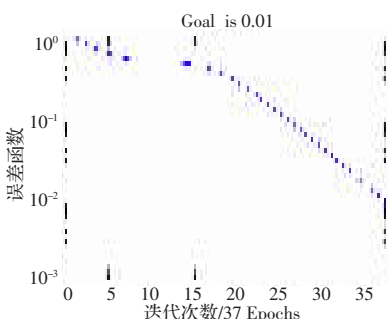


图 2 小波神经网络算法误差-训练次数曲线

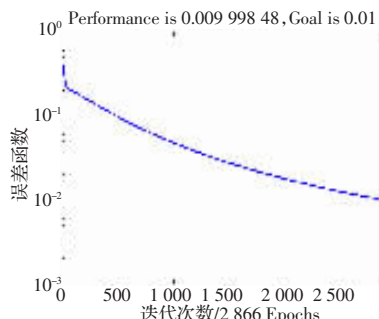


图 3 常规 BP 算法误差-训练次数曲线

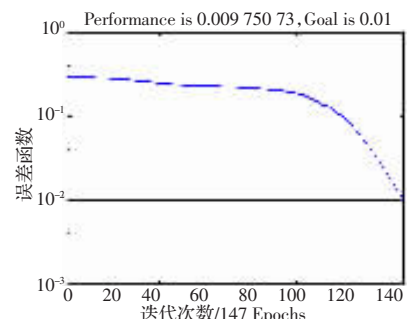


图 4 改进 BP 算法误差-训练次数曲线