

Rolling Bearing Fault Monitoring Research Based on the BP Neural Network

Shujuan Liao, Pan Fu, Erqing Zhang

Mechanical Engineering School, Southwest Jiaotong University, Chengdu
Email: amymissyouchina@sina.com

Received: Apr. 16th, 2014; revised: May 20th, 2014; accepted: May 30th, 2014

Copyright © 2014 by authors and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

The rolling bearing is one of the most common parts in machinery and equipment. Its normal running is related to the safety of the whole machine. So the fault diagnosis of rolling bearing is of great significance. In this paper, the platform system to collect normal bearing and the fault bearing vibration signals is established. According to the principle of feature selection, the time domain and frequency domain features are extracted and the eigenvalues are normalized. BP neural network has been trained based on the bearing characteristics of known state. Bearing of unknown state is identified by using the established network, and it is proved that this method has got a very good effect.

Keywords

Rolling Bearing, Fault Monitoring, Feature Extraction, BP Neural Network

基于BP神经网络的滚动轴承故障监测研究

廖术娟, 傅攀, 张尔卿

西南交通大学机械工程学院, 成都
Email: amymissyouchina@sina.com

收稿日期: 2014年4月16日; 修回日期: 2014年5月20日; 录用日期: 2014年5月30日

摘要

滚动轴承是机械设备中最常用的零件之一，它能否正常运行关系到整台机器的安全，所以对滚动轴承进行故障诊断具有重大的意义。本文搭建状态监测系统平台采集正常轴承和故障轴承的振动信号，根据特征选取原则，提取时域和频域特征并将特征值归一化。根据已知状态的轴承特征值训练BP神经网络，之后利用已建立的网络识别未知状态的轴承，经过试验验证，该方法得到了非常好的监测效果。

关键词

滚动轴承，故障监测，特征提取，BP神经网络

1. 引言

轴承是一种精密的机械元件，也是各种机械设备中举足轻重的基础件。滚动轴承作为机械设备中应用最广泛的标准部件之一，其本身精度高，对使用环境要求也相对较高，而且也是最易损坏并导致事故发生的部件[1]。轴承监测是保证产品尤其是高精度产品质量既能符合国家标准和企业内控标准，又可以很好地满足用户需求的手段，不但重要而且非常必要[2]。据统计，在使用滚动轴承的旋转机械设备中，约有 30% 的机械故障是与滚动轴承损伤有关的[3] [4]。所以，对滚动轴承的状态监测和故障诊断迫在眉睫。轴承的故障诊断可采用温度诊断技术、油样分析技术、油膜电阻诊断技术、声发射诊断技术和振动诊断技术等，而从适用、实用、有效的观点看，振动诊断法是目前最有效的滚动轴承状态监测与故障诊断方法[5]-[8]。本文基于滚动轴承的振动信号数据，通过建立识别神经网络对轴承的运行状态进行监测及故障诊断。

2. 实验系统及信号

一个完整的滚动轴承故障诊断系统应该包括数据采集和故障诊断两部分。其中数据采集部分用于采集和存储滚动轴承的振动信号；而故障诊断部分用于连续或者间断地处理振动信号，实现故障诊断，并将诊断结果显示出来。在本实验中，主要采用了压电式加速度传感器来测量振动信号，将传感器安装在合适的测点位置；由于传感器测得的振动信号比较微弱，因此传感器测得的信号需接入电荷放大器进行放大；放大器输出的信号是模拟信号，计算机无法识别，因此振动信号还需 A/D 卡进行模数转换，得到计算机可以识别和处理的数字信号，本文所涉及实验中采用的是 PCI-1711 研华数据采集卡。

3. 振动信号数据的特征提取

在设计好滚动轴承状态的监测装置后，基于所采用的装置采集滚动轴承的振动信号，再利用 MATLAB 对采集到的信号进行数据处理和分析。首先通过零均值化进行数据预处理，结果如图 1 所示。

3.1. 滚动轴承时域特征值提取

在时域内分别抽取如下七种特征值：均方根值、峰值、峰值因子、峭度因子、脉冲因子、裕度系数和波形因子。各特征参数的计算公式及物理意义如下：

- 1) 均方根值：其与轴承表面波纹等缺陷引起的无规则振动波形的异常有较好的相关性；

$$RMS = \sqrt{\frac{1}{N} \sum_{i=1}^N (X_i - \bar{X})^2} \quad (1)$$

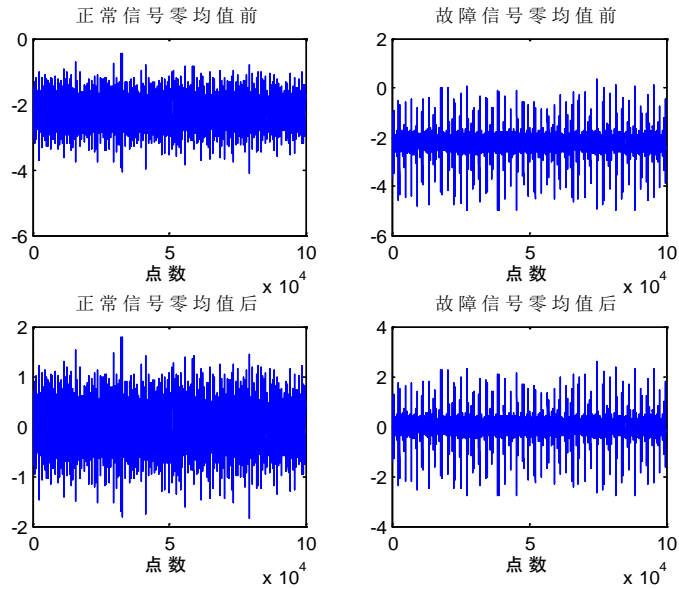


Figure 1. Normal and fault bearing signals zero mean contrast figure
图 1. 正常信号和故障信号零均值前后对比图

2) 峰值：峰值反映了信号偏离平均值的程度；

$$C_k = \frac{1}{N} \sum_{i=1}^N \frac{(x_i - \bar{x})^4}{\sigma^4} - 3 \quad (2)$$

3) 峰值因子：适用于点蚀类故障的诊断；

$$C = X_{peak} / X_{rms} \quad (3)$$

4) 峭度系数：表示轴承振动信号振幅的规则性，当过大时则意味着有故障出现；

$$K_v = \frac{\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^4}{\left(\frac{1}{N} \sum_{i=1}^N x_i^2 \right)^2} \quad (4)$$

5) 脉冲因子：反映了振动脉冲的频率；

$$I = X_{peak} / \bar{x} \quad (5)$$

6) 裕度系数：反映信号冲击程度的一个指标，对轴承的冲击故障较为敏感；

$$clf = \frac{|x_{\max}|}{\left(\frac{1}{N} \sum_{i=1}^N \sqrt{|x_i|} \right)^2} \quad (6)$$

7) 波形因子：可用来判定滚动轴承的损伤类型。

$$S = X_{rms} / \bar{x} \quad (7)$$

动力设备在各种负荷状态下的运行状态叫做工况。对于滚动轴承，在相同的负荷状态下，不同的转速也属于不同的工况，所以在对信号的处理和分析中，应在相同的转速下，提取特征值。在本实验中，正常轴承和故障轴承在 $z_s = 20 \text{ rpm}$, $z_s = 25 \text{ rpm}$, $z_s = 30 \text{ rpm}$ 三种工况下采集振动信号，每种工况下又分成

3 组数据, 得 $2 \times 3 \times 3 = 18$ 组数据。以滚动轴承转速为 20 rpm(20 转/每分钟)时的状态数据为例说明, 时域特征结果见表 1。

正常轴承和故障轴承 $z_s = 20$ 时分别提取 3 组特征, 根据表 1 得到时域特征值曲线对比图如图 2, 由图可知在相同的工况下, 时域特征值的重复性好, 不同工况下, 峰值、峭度因子、峰值因子、脉冲因子和裕度因子特征值在幅值上差别很大, 尤其是峭度因子和裕度因子在正常状态下和故障状态下区别非常明显。

3.2. 滚动轴承频域特征值提取

对信号进行时域分析时, 有时一些信号的时域参数相同, 但并不能说明信号就完全相同。因为信号不仅随时间变化, 还与频率、相位等信息有关, 这就需要进一步分析信号的频率结构, 并在频域中对信号进行描述。动态信号从时间域变换到频域主要通过傅立叶级数和傅立叶变换实现, 其中周期信号靠傅立叶级数, 非周期信号靠傅立叶变换。

Table 1. Rolling bearing vibration signal time domain features

表 1. 滚动轴承振动信号时域特征值

z _s = 20 时的滚动轴承振动信号时域特征值							
	均方根值(jfgz)	峰值(fz)	峭度因子(qdyz)	峰值因子(fzyz)	脉冲因子(mcyz)	裕度因子(ydyz)	波形因子(bxyz)
正常 信号	0.3251	1.7468	4.1545	5.3733	7.0149	8.4392	1.3055
	0.3285	1.7375	4.1002	5.2900	6.8938	8.2873	1.3032
	0.3189	1.6557	4.0881	5.1917	6.7713	8.1472	1.3043
故障 信号	0.3815	2.7333	13.1025	7.1650	13.0510	18.7229	1.8215
	0.3914	2.7435	11.8474	7.0086	12.5505	17.9573	1.7907
	0.3828	2.7038	12.2842	7.0633	12.6333	17.9816	1.7886

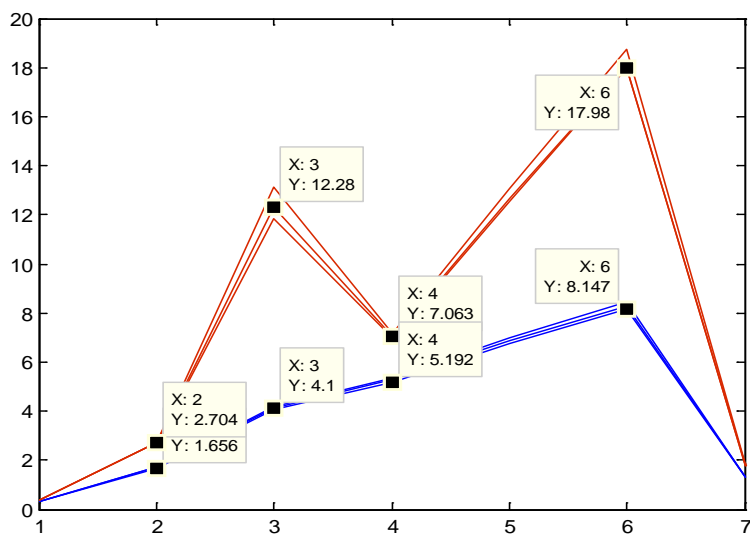


Figure 2. Normal and fault bearing signals time domain features contrast figure ($z_s = 20$ rpm)

图 2. $z_s = 20$ 时正常信号和故障信号的时域特征值对比图(红色——故障信号、蓝色——正常信号)

在本实验中，利用傅里叶变换提取频域特征。傅里叶变换(FFT)是一种频域分析方法，当信号被分解成各次谐波以后，我们就可以从频域来分析处理问题。信号的频域特性是信号的内在本质，而信号的时域波形只是信号的外在形式。显然，从本质上分析处理问题将会更深入，得到的分析结果也将会更全面、更具有优越性。

在频域中，首先提取如下三种特种值：重心频率(zxpl)，频率方差(plfc)，均方频率(jfpl)。

根据表 2 得正常轴承和故障轴承振动信号频域特征对比图如图 3，从图中可以明显看出故障轴承信号的频域特征较正常信号相反，在幅值上均小于正常轴承信号的特征值。

由图 2、图 3 可以得出：当转速为 20 rpm 时，正常轴承信号和故障轴承信号所提取的时、频域特征值对比非常明显。为了验证选取的特征值是否具有普适性，分别在转速为 20 rpm、25 rpm 和 30 rpm 状态下采集信号，并分别求取这三种状态下信号的频率特征，其对比结果如图 4 所示，正常轴承在不同的转速下频域特征值有很好的重复性，故障轴承在不同的转速下频率特征值表现出幅值差异很大，无规律。

在不同的工况下($z_s = 20$ 、 $z_s = 25$ 和 $z_s = 30$)，所得到的滚动轴承信号频域特征是不同的，虽然在 $z_s = 20$ 时故障轴承频域幅值的差异非常明显。当 $z_s = 25$ 和 $z_s = 30$ 时，情况发生了改变，不仅幅值差异很大，而且变化趋势也有所不同。说明了不同的工况，提取的频域特征不具有重复性和变化趋势一致性，根据

Table 2. Rolling bearing vibration signal frequency domain features
表 2. 滚动轴承振动信号频域特征值一

z _s = 20 时滚动轴承振动信号频域特征一			
	重心频率	频率方差	均方频率
正常 信号	0.6537	0.6667	1.0940
	0.6494	0.6657	1.0874
	0.6647	0.6790	1.1208
故障 信号	0.4706	0.3148	0.5362
	0.4756	0.3194	0.5456
	0.4893	0.3332	0.5726

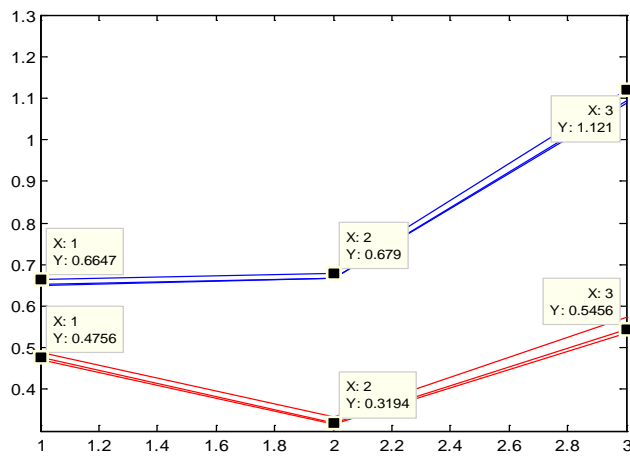


Figure 3. Normal and fault bearing signals frequency domain features contrast figure ($z_s = 20$ rpm)

图 3. $z_s = 20$ 时正常信号和故障信号的频域特征对比图(红色——故障信号、蓝色——正常信号)

特征值选取原则，重心频率、频率方差、均方频率都不可以作为模式识别的特征量，重新选取频域特征。

为了得到的特征值在相同工况下有很好的重复性且变化趋势具有一致性，不同工况下又有很好的差异性，本实验中，对三种工况采集到的数据分别分段提取 FFT 谱的分段最大值均值特征(将采集信号的 FFT 谱按能量特征分成 9 段)，得到的特征值如表 3 所示。

正常状态和故障状态，分别各提取 $z_s = 20$, $z_s = 25$, $z_s = 30$ 转速下的一组特征，根据表 3 中分段 FFT 频谱最大值均值得到的正常状态和故障状态特征值曲线对比图如图 5，从图中可以看到在不同转速下信号有较好的重复性，而且正常信号和故障信号之间的 D3、D7、D8 又具有比较明显的差异性。

由图 5 可知，相同工况下选取的九个特征值中，根据有效特征值的选取原则：1) 同状态信号重复性好；2) 不同状态信号差异性好，处于横坐标为 3、7、8 上的特征值可以选作模式识别的特征值。

对时域和频域特征值的比较筛选，得到八组特征值用于模式识别的特征值，整理数据如表 4。

4. 模式识别

4.1. 特征值归一化

当所有样本的输入信号都为正值时，与第一隐含层神经元相连的权值只能同时增加或减小，从而导致学习速度很慢。为了避免出现这种情况，加快网络学习速度，需要对输入信号进行归一化，使得所有

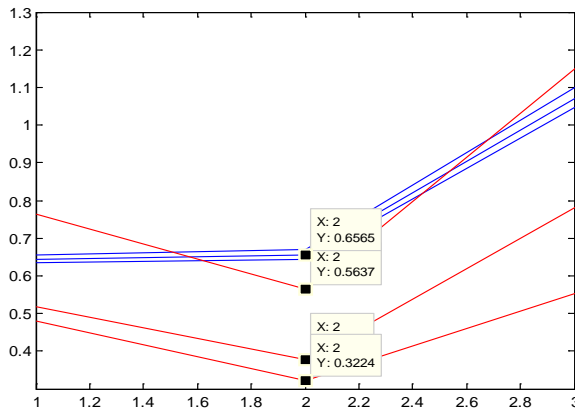


Figure 4. Normal and fault bearing signals frequency domain features contrast figure at three different speeds
图 4. 三种不同转速下正常信号和故障信号的频域特征对比图(红色——故障信号、蓝色——正常信号)

Table 3. Rolling bearing vibration signal frequency domain features
表 3. 滚动轴承信号频域特征值二

		z _s = 20 滚动轴承信号频域特征二								
		D1	D2	D3	D4	D5	D6	D7	D8	D9
正常信号		1.5590	0.6126	0.0787	0.0165	0.0583	0.0226	0.9472	0.3522	0.0337
		1.7083	0.6035	0.0905	0.0163	0.0660	0.0219	0.8763	0.3489	0.0339
		1.7254	0.6615	0.1033	0.0162	0.0720	0.0221	1.0205	0.3451	0.0338
故障信号		1.9691	0.6751	0.8933	0.0403	0.1415	0.0594	0.2354	0.1220	0.0258
		2.1748	0.6519	1.1732	0.0415	0.1528	0.0591	0.3739	0.0963	0.0259
		2.2267	0.6232	0.8153	0.0431	0.1407	0.0605	0.2087	0.1144	0.0268

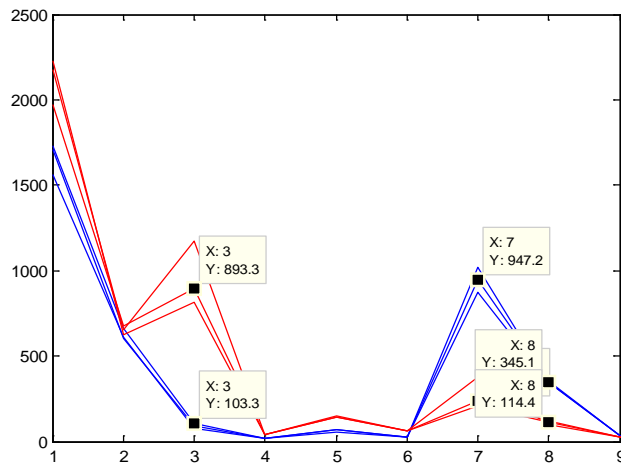


Figure 5. Normal and fault bearing signals frequency domain features contrast figure ($z_s = 20$ rpm)
图 5. $z_s = 20$ 时正常信号和故障信号的频域特征对比图(红色——故障信号、蓝色——正常信号)

Table 4. Rolling bearing vibration signal features
表 4. 信号有效特征值对比

	峰值(fz)	峭度因子(qdyz)	峰值因子(fzyz)	脉冲因子(mcyz)	裕度因子(ydyz)	D3	D7	D8
正常信号	1.7468	4.1545	5.3733	7.0149	8.4392	0.0787	0.9472	0.3522
	1.7375	4.1002	5.2900	6.8938	8.2873	0.0905	0.8763	0.3489
	1.6557	4.0881	5.1917	6.7713	8.1472	0.1033	1.0205	0.3451
故障信号	2.7333	13.1025	7.1650	13.0510	18.7229	0.8933	0.2354	0.1220
	2.7435	11.8474	7.0086	12.5505	17.9573	1.1732	0.3739	0.0963
	2.7038	12.2842	7.0633	12.6333	17.9816	0.8153	0.2087	0.1144

样本的输入信号其均值接近于 0 或与其均方差相比很小。归一化就是要把数据通过某种算法处理后，限制在一定范围内。首先归一化是为了后面数据处理的方便，其次是保证程序运行时收敛加快。归一化的具体作用是归纳统一样本的统计分布性。

神经网络是以样本在事件中的统计分布几率来进行训练(概率计算)和预测的，归一化是统一在(0, 1)之间的统计概率分布，所以在本实验中，对提取的轴承信号特征值采用线性函数转换，其表达式为： $y = (x - \text{MinValue}) / (\text{MaxValue} - \text{MinValue})$ ，其中 x 、 y 分别为转换前、后的值， MaxValue 、 MinValue 分别为样本的最大值和最小值。归一化后的特征值数据如表 5 所示。

4.2. BP 神经网络

BP(Back Propagation)神经网络，通过误差反向传播来不断调整网络的权值和阈值的学习过程，由信息的正向传播和误差的反向传播两个过程组成。输入层各神经元负责接收来自外界的输入信息，并传递给中间层各神经元；中间层是内部信息处理层，负责信息变换，根据信息变换能力的需求，中间层可以设计为单隐层或者多隐层结构；最后一个隐层传递到输出层各神经元的的信息，经进一步处理后，完成一次学习的正向传播处理过程，由输出层向外界输出信息处理结果。当实际输出与期望输出不符时，进入误差的反向传播阶段。误差通过输出层，按误差梯度下降的方式修正各层权值，向隐层、输入层逐层反

Table 5. Rolling bearing vibration signal normalized features
表 5. 归一化后的特征值

		峰值	峭度因子	脉冲因子	裕度系数	峰值因子	D3	D7	D8
正常	1	0.0837	0.0074	0.0920	0.0388	0.0276	0	0.9097	1.0000
	2	0.0752	0.0013	0.0498	0.0195	0.0132	0.0108	0.8224	0.9871
	3	0	0	0	0	0	0.0225	1.0000	0.9723
故障	1	0.9906	1.0000	1.0000	1.0000	1.0000	0.7443	0.0329	0.1004
	2	1.0000	0.8608	0.9207	0.9203	0.9276	1.0000	0.2035	0
	3	0.9635	0.9092	0.9485	0.9335	0.9299	0.6730	0	0.0707

传。周而复始的信息正向传播和误差反向传播过程，是各层权值不断调整的过程，也是神经网络学习训练的过程，此过程一直进行到网络输出的误差减少到可以接受的程度，或者预先设定的学习次数为止。

4.3. BP 神经网络的建立

BP 网络是由一个输入层，一个或多个隐层以及一个输出层组成，上下层之间实现全连接，而每层神经元之间无连接，输入的神经元可以根据要求解的问题和数据表示方式确定。在故障诊断中，输入层神经元个数一般等于每个样本中包含的特征值的个数，输出层的维数可根据设计者的要求确定。在正向传播进程中，输入信息从输入层经隐层加权处理传向输出层，经功能函数运算后得到的输出值与期望值进行比较，若有误差，则误差反向传播，沿原先的连接通道返回，通过逐层修改各层的权重系数，减小误差。随着这种误差逆向传播修正的不断进行，网络对输入模式响应的正确率也不断上升。

其中隐层的设计：隐层的神经元个数选择是一个较为复杂的问题，往往需要设计者多次试验来确定，因而不存在一个理想的解析式来表示。隐层单元的个数与问题的要求、输入/输出单元的数目都有着直接关系。隐层单元个数太多会导致学习时间过长、误差不一定最佳，也会导致容错性差、不能识别以前没有见到的样本等问题，因此一定存在一个最佳的隐层单元数。在本实验中，主要参照以下三个公式确定最佳隐层单元数：

$$n_2 = 2n_1 + 1 \tag{8}$$

其中 n_1 是输入层神经元数， n_2 是隐层神经元数；

$$n_1 = \sqrt{n + m} + a \tag{9}$$

其中， m 是输出神经元数， n 是输入神经元数， a 是 [1, 10] 之间的常数；

$$n_1 = \log_2 n \tag{10}$$

其中 n 是输入神经元数。

三层 BP 神经网络的网络结构图如图 6。

4.4. 神经网络训练及结果

神经网络训练程序如下：

```
net=newff(minmax(x),[21,2],{'tansig','logsig'},'traingdx');
net.trainParam.epochs = 300;
net.trainParam.lr = 0.1;
net.trainParam.goal = 0.0001;
```

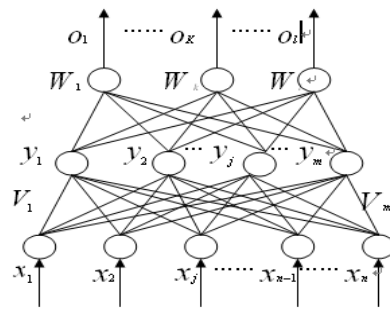
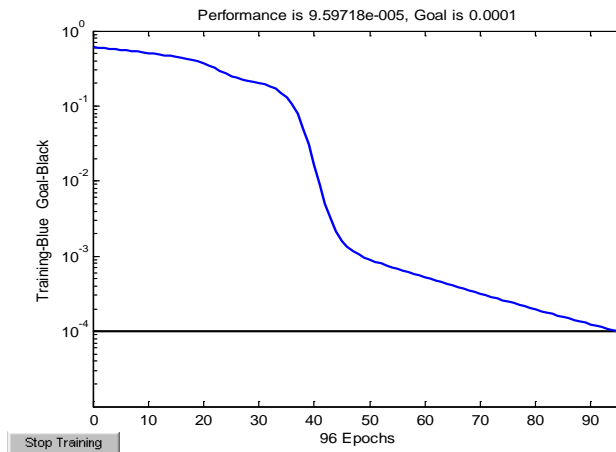



Figure 6. BP network structure of three layers
图 6. 三层 BP 神经网络的网络结构示意图

```
net=train(net,x,t);
y=sim(net,x);
error=y-t
res(i)=norm(error);
程序训练得到结果为:
```

```
error =
-0.0166   -0.0186   -0.0157    0.0069    0.0118    0.0076
0.0031    0.0030    0.0034   -0.0010   -0.0026   -0.0011
```



可见网络训练收敛速度快，稳定性好。使用验证样本对已经训练好的网路进行仿真，比较仿真结果与目标结果，并通过比较计算训练样本、验证样本平均误差评价此 BP 神经网络的有效性。而通过识别结果得知，所建立的 BP 网络识别结果准确，达到了预期的目标。

5. 总结

从对滚动轴承故障监测的实际出发，以及滚动轴承状态监测的重要性的发展前景，设计选择监测装置，并通过选用的监测装置采集正常轴承和故障轴承的诊断信号数据，对获得的数据提取时域特征和频域特征，通过正常信号和故障信号的特征值对比图，根据特征值的选取原则，提取有效特征值，如果不加选择地选取过多故障特征作为神经网络的输入，不仅不能提高系统正确诊断率，反而会降低诊断率。以神经网络理论为基础，结合滚动轴承故障诊断的特点，设计出适合于滚动轴承故障诊断的 BP 神经网络。

络，将时域特征和频域特征进行不同组合作为 BP 神经网络的输入，对 BP 神经网络进行训练及验证，运用训练网络对未知轴承状态进行监测，实验证明该方法能够对轴承运行状态进行有效的实时检测和诊断。

参考文献 (References)

- [1] 胡倩 (2006) 基于 BP 神经网络的滚动轴承缺陷诊断研究. 硕士学位论文, 华中科技大学, 武汉.
- [2] 李萌 (2008) 旋转机械轴承故障的特征提取与模式识别的方法研究. 博士学位论文, 吉林大学, 长春.
- [3] 张俊 (2009) 基于小波分析和神经网络的滚动轴承故障诊断. 硕士学位论文, 北京交通大学, 北京.
- [4] 刘华胜 (2007) 基于 EMD 的滚动轴承故障诊断方法研究. 硕士学位论文, 大连理工大学, 大连.
- [5] 王美波 (2008) 基于声学方法的滚动轴承故障信号分析方法研究. 硕士学位论文, 大庆石油学院, 大庆.
- [6] 潘紫微 (2009) 基于小波包分析和高阶模糊神经网络的滚动轴承故障诊断. *煤矿机械*, **12**, 238-240.
- [7] Kurfess, T.R., Billington, S. and Liang, S.Y. (2006) Advanced diagnostic and prognostic techniques for rolling element bearings. *Condition Monitoring and Control for Intelligent Manufacturing*, 137-165.
- [8] Wang, D., Miao, Q., Fan, X.F. and Huang, H.-Z. (2009) Rolling element bearing fault detection using an improved combination of Hilbert and wavelet transforms. *Journal of Mechanical Science and Technology*, **23**, 3292-3301.