

摘要

滚动轴承是机械设备中最常用的部件，它的运行状态直接影响到整台机器的性能。据统计，在旋转机械的现场故障中，滚动轴承损伤而引起的故障约占 30%。对滚动轴承进行精确、可靠的故障诊断具有极大的现实意义，本文提出了一种基于改进 BP 算法优化神经网络的滚动轴承故障诊断方法。

首先，系统地阐述了滚动轴承的振动机理及其几种典型故障的振动特征。根据故障的不同引起不同形式振动的机理，建立了多种单一损伤类故障振动的数学模型，为相应的故障诊断提供了坚实的理论基础。

其次，研究了小波变换在滚动轴承故障诊断中信号去噪方面与信号奇异性检测两方面的应用。滚动轴承振动信号容易受到随机噪声的污染，如何去噪成为滚动轴承故障诊断的关键问题之一。针对传统的消噪方法可能会将信号中一些能量小的有用信号当作噪声消除，提出了一种改进的小波消噪方法，能更好地提高信噪比，为进一步故障诊断决策提供了可靠的依据。小波奇异性方法克服了其他方法对滚动体故障不敏感的缺陷，可有效地用于滚动轴承的状态检测和故障诊断。

最后，针对 BP 算法存在收敛速度、局部最小问题和训练瘫痪等缺点，采用了优化原理对其进行了改进。同时利用小波包分解与重构提取滚动轴承的动态信号的能量作为故障特征参数，将其作为 BP 网络的输入，并用共轭梯度法对网络进行训练，然后由 BP 神经网络进行故障诊断。通过对实测滚动轴承振动信号进行仿真分析，实验结果表明采用该方法与其它方法相比，收敛速度快，诊断精度高，可以准确地实现滚动轴承的故障诊断。

关键词：小波变换，消噪，奇异性检测，滚动轴承，BP 神经网络，故障诊断

ABSTRACT

Rolling bearing is the common component in machinery. Its running state can influence the performance of the whole machine directly. According to statistics, in the rotating machinery fault at the scene, 30% of faults are caused by the rolling bearing. It is of great realistic significance to diagnosis the fault of rolling bearing accurately and reliably. This paper proposed a new fault diagnosis method of rolling bearing using improved BP neural network.

Firstly, systematically expounded the vibration of rolling bearings and the vibration characte of their typical failure. According to the different failure caused by different forms of vibration, bulding vibration signal model with single defect on each component and analysis fault charactera of the vibration signal. And then, the use of bearing vibration signal to diagnose rolling bearing is feasible.

Secondly, study and analyze the wavelet transform at different aspects of fault diagnosis of rolling bearing applications: denoising; singularity detection. Rolling element bearing vibration signal is vulnerable to random noise pollution, how to denoising becomes one of the key issues about rolling element bearing fault diagnosis. The traditional denoising methods may treated the useful signals which have small energy as noise, and eliminated it. Then this paper expressed the wavelet denoising method. In the area of denoising, the wavelet denoising improved the signal noise ratio perfectly and provided a reliable basis of fault diagnosis decision-making. In addition, wavelet singularity detection can overcome shortcomings that the other methods are not sensitive to detect the rolling faults. It can be effectively used for rolling bearing state detection and fault diagnosis.

Finally, aimed at shortcomings in BP algorithm, such as convergence rate, the smallest partial paralysis and so on, a new rolling fault diagnosis method was produced. Wavelet transformation was used to extract feature of dynamic vibration information, and then input to BP neural network to train the network with conjugate gradient algorithm. Moreover, applying the BP neural network to identify the failure of the rolling bearings. By analyzed on the rolling bearing vibration signals, the experiment results showed that,

compared to other methods, the new method is faster to converge and has a higher precision in fault diagnosis applications, and can accurately realize the rolling fault diagnosis.

KEY WORDS: wavelet transform, denoising, singularity detection, rolling bearing, BP neural network, fault diagnosis.

原创性声明

本人声明，所呈交的学位论文是本人在导师指导下进行的研究工作及取得的研究成果。尽我所知，除了论文中特别加以标注和致谢的地方外，论文中不包含其他人已经发表或撰写过的研究成果，也不包含为获得中南大学或其他单位的学位或证书而使用过的材料。与我共同工作的同志对本研究所作的贡献均已在论文中作了明确的说明。

作者签名：黄一祥 日期：2009年5月15日

学位论文授权使用授权书

本人了解中南大学有关保留、使用学位论文的规定，即：学校有权保留学位论文并根据国家或湖南省有关部门规定送交学位论文，允许学位论文被查阅和借阅；学校可以公布学位论文的全部或部分内容，可以采用复印、缩印或其它手段保存学位论文。同时授权中国科学技术信息研究所将本学位论文收录到《中国学位论文全文数据库》，并通过网络向社会公众提供信息服务。

作者签名：黄一祥 导师签名：申昭杰 日期：2009年5月15日

第一章 绪论

1.1 课题研究的背景及意义

1.1.1 研究背景

当今的时代被称为信息时代，信息科学技术的快速发展和广泛渗透已经成为现今社会的一个重要的时代特征^[1]。随着工业技术的迅速发展，现代机械设备正朝着大型化，复杂化和高度自动化方向发展，由此使用设备的功能越来越多，性能指标越来越高，组成和结构越来越复杂，价格越来越贵，对设备的管理与维修的要求也相对提高。诚然这些设备对国民经济的影响也越来越大，许多故障若不能及时发现并加以预防，一旦发生事故，则将导致整个生产过程无法正常进行，甚至造成重大经济损失和人员伤亡。

滚动轴承是机械设备中最常用的部件且工作条件最为恶劣，并起着关键作用，它的运行状态正常与否往往直接影响整台机器的性能（包括精度、可靠性及寿命等），对生产过程、产品质量等方面也有直接影响。据统计，在旋转机械的现场故障中，由于滚动轴承损伤而引起的故障约占 30%，例如电机故障中 80%表现为电机轴承故障。由于轴承特定的使用环境造成其寿命的随机性比较大，即有的轴承大大超过设计寿命而亦然完好地工作，有的轴承则还未达到设计寿命就出现各种各样的故障。由于轴承的这个特点，因此在实际应用中如何准确的监测并判别滚动轴承的工作状态一直都是国内外机械故障诊断研究中的重点之一。

1.1.2 研究的意义

设备故障诊断是识别机器或机组运行状态的科学，它研究的是机器或机组运行状态的变化在诊断信息中的反映，其研究内容包括对设备运行现状的识别诊断、对其运行过程的监测以及对其运行发展趋势的预测三个方面。

滚动轴承故障诊断的目的是保证轴承在一定的工作环境和一定的工作期间可靠有效的运行，以保证整个机械的工作精度。及时地识别滚动轴承的状态，这不仅可以从防止机械工作精度下降，减少或杜绝事故发生，而且可以最大限度的发挥轴承的工作潜力，减少经济开支。

综上所述，对滚动轴承故障诊断技术进行探索和研究，对于保证机组长期、可靠、高效运行，从而保证企业的安全生产与跳高经济效益具有重大的意义与实际应用价值。

1.2 故障诊断方法及其发展现状与趋势

1.2.1 故障诊断方法概述

故障诊断技术是随着机械工程、测试技术、信号处理、计算机应用技术、人工智能等众多理论基础的发展而发展起来的新兴综合性科学技术。经过几十年的研究和应用,其理论和方法日益丰富和完善,产生了许多故障诊断方法。从故障发生的部位看,可以分为仪表故障(通常称为传感器故障),执行器故障和元部件故障;根据故障性质,可以分为突变故障和缓变故障;从建模角度出发,又可以分为乘性故障和加性故障。目前故障检测与诊断方法主要可以分为基于解析模型的方法、基于信号处理的方法和基于知识的方法^[2]。

基于解析模型的故障诊断方法应用在线系统辨识技术来实时地为系统建立数学模型,当系统中存在故障时,系统的输入输出关系就会改变,这些变化就会反映在数学模型中,因而通过观测系统数学模型参数变化,便能判定系统是否存在故障。基于解析模型的故障诊断方法能够深入研究系统本质的动态性质并进行实时诊断,但系统模型通常难以获得,且由于建模误差、扰动及噪声的存在,使得鲁棒性问题日益突出。

基于信号处理的方法不需要对象的准确模型,是一种传统的故障诊断技术,适应性强。当可以得到被控过程的输入输出信号,但很难建立被控对象的解析数学模型时,可采用该方法。

上述的基于系统模型的故障诊断方法要求得到较为精确的系统模型,但在实际工程中往往难以建立系统的模型或模型不精确,这就限制了这种故障诊断方法在实际中的使用。基于知识的方法同基于信号的故障诊断方法类似,同样不需要对象的精确数学模型,常见的有专家系统方法、模糊诊断方法、遗传算法诊断方法、人工神经网络法以及基于诊断对象的模型知识的诊断方法。

1.2.2 滚动轴承故障诊断技术的发展现状及趋势

(一) 滚动轴承故障诊断技术的发展现状

滚动轴承的故障诊断在国外大概开始于20世纪60年代^[3]。在其后的几十年时间里,随着科学技术的不断发展,各种方法和技巧不断产生、发展和完善,应用领域不断扩大,监视与诊断的有效性不断提高。

我国是从1983年开始,在熟悉苏联维修体制的基础上,不断地研究美、日、德、瑞典等国的各种维修体制,从中知道状态监测与诊断维修能改善管理,具有较大的优点,方才开始接触这方面的问题。化工、电力、石化行业等很多行业也投入了较大的资金对重点设备进行监测,特别加强了大型旋转机械的振动监测和压力容器等设备的安全性能检验,并开发了相关设备在线监测和故障诊断系统。

振动理论^[4]适用于各种类型各种工况的轴承，可以有效地诊断出早期微小故障，且可实时测定、处理简单直观和诊断结果可靠等优点，所以目前的滚动轴承的故障诊断方法大都以振动理论为基础，通过对振动信号的分析来诊断轴承的故障。James E. Berry 表明^[5]，在过去的时间中，在滚动轴承振动信号分析方面有很多研究。文章^[5]表明，频谱内容不是振动幅值，而是揭示轴承状态的关键。作者用一种表格程序总结如何决定何时需要更换轴承，给出了一个单独的标准。共振解调技术是近年来发展起来的对滚动轴承的点蚀、剥落、擦伤和裂纹等所谓的元件工作表面损伤类故障行之有效的办法。王志刚^[6]等人针对低速重载轴承特点，提出了适合于低速重载轴承故障诊断的共振解调方法。

K.F.Martin 等^[7]详细地说明了应用包络技术进行故障检测和诊断，并描述了一种通过对故障诊断和正常轴承的频谱进行归一化的技术，这种方法对检测故障特征频率很敏感。另外国内关于利用包络分析法诊断轴承故障的文献中很多是利用 Hilbert 变换实现包络解调。武和雷等^[8]人采用能量算子对故障轴承的高频共振信号进行包络解调，提取故障特征，从而达到诊断故障的目的。

传统的滚动轴承故障诊断方法有频域分析和时域分析方法等均假定振动信号为平稳信号，但现场实际滚动轴承故障轴承振动信号具有非平稳的特性。为了克服传统检测方法的缺陷，W.Y.Wang 和 M.J.Harrap 将时频分析方法应用于滚动轴承的损伤检测^[9]。通过观察滚动轴承振动信号的时频分布情况，来判断滚动轴承是否有故障。其研究表明，在几种时频分析方法中，伪 Winger 分布用于滚动轴承的局部损伤如外圈或内圈剥落的检测具有相对较好的效果。但这些方法都存在计算量大、实时性差的特点。常用的时频分析方法有短时傅立叶变换、Wiger-Vile 分布和小波分析等。陈刚等^[10]将振动信号进行小波分解与重构，然后对其中某粗略信号进行频谱分析，其频谱特征可以揭示故障的有无及其发证部位。通过滚动轴承内环点蚀、外环划痕、滚动体划痕三种典型故障的实验，验证了该方法在滚动轴承故障诊断方面的可行性和有效性。

目前，美、英、日等国都相继开发了以微机为中心的滚动轴承工况监测和诊断系统。滚动轴承诊断系统主要有俄罗斯的 VAST 公司开发的滚动轴承自动诊断系统 DREAM, Bently 公司的 REBAM 系统。国内有航空航天部 608 所的唐德尧教授等人于 1984 年开发成功的 JK86411 自动试验诊断系统；还有南京航空航天大学振动所赵淳生教授等人开发的 NMS 系列轴承故障诊断系统。M.Subrahmanyam 等采用两种神经网络方法，采用有监督的误差反向传播（EBP）技术和无监督的自适应共振理论（ART2）训练的多层前向神经网络。从一个正常轴承和两种不同类型的故障轴承采集加速度信号，通过处理获得统计特征参数，作为网络的输入，输出代表轴承的状态。训练的结果用于识别轴承的状态。

目前国内故障诊断技术的研究主要集中在以下几个方面：

(1) 传感器的研究：传感器技术是反映设备状态参数的仪表技术。国内先后开发了电涡流传感器、速度传感器、加速度传感器和温度传感器等，最近开发的传感技术有光导纤维、激光、声发射等。

(2) 信号分析与处理技术的研究：从传统的谱分析、时序分析和时域分析，到引入一些更新的信号分析手段，如短时傅里叶分析、Winger 谱分析和小波变换等，弥补了传统分析方法的不足。

(3) 人工智能与专家系统的研究：这方面的研究已成为诊断技术的发展主流，目前已有“人工机械故障诊断专家系统”，但这一技术在工程方面的研究尚未达到人们所期待的水平。

(4) 诊断系统的开发与研究：从单机巡检与诊断到上下位机的主从结构，直至以网络为基础的分布式结构，系统的结构越来越复杂，实时性越来越高。

(5) 专门化与便携式诊断仪器和设备的研制与开发。

(二) 滚动轴承故障诊断技术的发展趋势

随着信号检测技术、计算机技术、数字信号处理技术、人工智能技术的迅速发展^[11-12]，基于分形理论和混沌的故障诊断技术、极大熵谱法、神经网络技术、小波理论和光谱分析等新的理论和方法开始应用于滚动轴承的诊断研究，为复杂的滚动轴承的故障诊断提供了更可靠、更准确的信息依据。今后的发展方向可体现在以下几个方面^[13]：

(1) 时域分析和频谱分析在轴承故障诊断中的应用日益完善。

(2) 对于轴承故障诊断的理论和方法进一步深入研究，并且各种研究成果将会逐步应用到实际生产中。

(3) 故障诊断智能系统进一步深入研究，多种轴承故障分析方法结合，如小波神经网络、模糊识别与小波分析相结合等新分析方法应用智能专家系统，提高诊断的效率和准确率。

(4) 随着计算机和网络技术的发展，远程故障诊断将是一个重要的方向。

1.3 轴承故障诊断的流程

滚动轴承的故障诊断要先识别其运行的状态，再研究运行状态在故障诊断中的信息反映。主要包括对轴承运行状态的监测、故障预报和故障监视等以下五个环节。整个轴承故障诊断的过程如图 1-1 所示。

(1) 信号采集：根据滚动轴承的工况，通过传感器检测滚动轴承的振动、压力、温度等能够反映滚动轴承工作情况或状态信号。

(2) 特征提取：从状态信号中，利用信号处理器提取特征信号，从而反映轴承工作状态。

(3) 状态识别：根据提取的特征信息，通过一定的故障识别方法识别轴承的状

态，判断轴承的运行状态是否工作正常。

(4) 故障分析：当轴承出现故障时，根据故障的征兆来进一步分析故障的类型、部位、性质、产生原因及发展趋势等。

(5) 决策干预：根据轴承状态及发展趋势，做出调整、控制、自诊治或维修等决策，对轴承的使用进行干预。

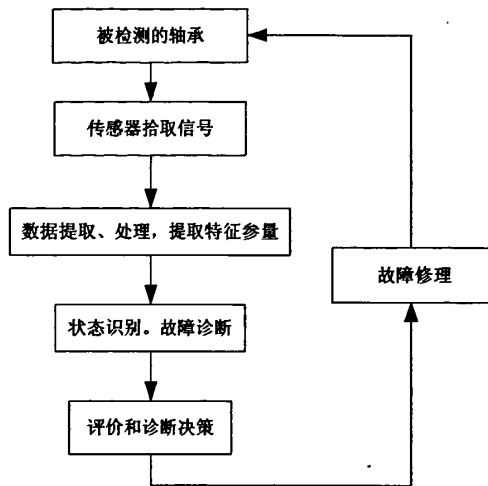


图 1-1 故障诊断过程流程图

1.4 小波理论和神经网络在故障诊断中的应用

1.4.1 小波理论在故障诊断中的应用

小波变换是 80 年代后期发展起来的应用数学分支，最初由法国学 Daubechies 和 Mallat 引入信号处理研究领域，是一种时间-尺度分析方法，具有多分辨率分析的特点。它在 20 世纪 90 年代被广泛的应用于信号处理领域，因其能够把信号特性分配到各个不同尺度的小波变换系数上，则可根据小波变换系数的分析与处理，就可以对信号进行压缩、奇异性检测以及去噪等方面^[14-15]。

故障信号的特征提取^[16]是当前轴承故障诊断中的瓶颈问题，它直接影响到故障诊断的准确性和可靠性。滚动轴承早期故障信号信号比较微弱，往往容易在采集和传输的过程中被其它随机噪声污染。如何把微弱的有用信号同噪声分开，提高信号的信噪比，为故障特征频率的提取提供可靠的依据，成为滚动轴承故障诊断中的重要任务之一。文献[17]对近些年在机械诊断领域内小波技术的研究及其在滚动轴承故障诊断等方面应用中取得的进展进行了详细系统地介绍。

传统的消噪方法主要有信号滤波、自适应消噪、时域平均等^[18-19]，这些方法难以将滚动轴承微弱的故障信号从背景信号中分离，也可能会将信号中一些能量小的有用信号当作噪声消除，有的实施起来也有一定的难度。而小波变换是一个频率和

时间的局部变换,具有多分辨分析的特点,通过每层高频分解的系数进行阈值处理后重构,可有效地达到消除噪声的目的。小波去噪的方法有很多种^[20]:(1)小波变换模极大值法;(2)非线性小波变换阈值法;(3)平移不变量小波去噪法;(4)原子分解的基追踪去噪法等。比如文献[21]简单介绍了一种小波去噪的原理,并加以应用,结果表明小波去噪能够有效地保留信号中的尖峰和突变部分,最大限度地反映了原信号本身的特性。另外国内外有不少学者把小波理论用于故障诊断领域的成功实例^[22-25],其中文献[22]应用小波分析以提高故障检测灵敏度。

1.4.2 神经网络理论在故障诊断中的应用

神经网络用于故障诊断始于 20 世纪 80 年代末期。最初故障诊断采用的神经网络是反向传播(BP-Back Propagation)神经网络,后来人们又采用了多种神经网络结构模型,如径向基函数(RBF-Radial Basis Function)神经网络、Hopfield 神经网络、ART 神经网络等。由于 ANN 是对人脑或自然的神经网络若干基本特性的抽象和模拟,是一种非线性的动力学系统。

它具有大规模的并行处理和分布式消息存储能力、自组织性既很强的学习、联想、容错、抗干扰等能力、具有良好的自适应性。许多学者、科技工作者从事这一新领域的研究,大量的研究成果如雨后春笋般涌现出来。国内外有不少用于故障诊断领域的成功实例,比如文献[26-33]:其中文献[33]分析了前向型神经网络动力系统,并提出新算法来克服传统 BP 算法学习速度慢、容易陷入局部最小等缺点。改进后的算法用于轴承故障诊断中,结果表明该方法能有效地缩短网络在训练中滞留于局部极小区域的时间,大大地提高了网络的学习速度。

1.5 主要的研究内容及论文结构安排

本文对滚动轴承的故障诊断过程中遇到的相关问题进行了详细的分析和研究,其主题在于如何有效地将小波理论应用于滚动轴承振动信号的故障诊断中,继而准确地判别滚动轴承是否存在故障及其故障的类型。围绕此主题,各个研究的章节内容安排如下:

第一章:介绍了课题的研究背景及其意义,综述了故障诊断方法的发展现状及其趋势,研究了小波理论和神经网络在故障诊断系统中的应用及其发展过程,从而提出了本文的研究内容。

第二章:分析了滚动轴承的振动机理及其典型故障的振动特征,为后面的研究奠定了坚实的理论基础。

第三章:详细地阐述了小波分析的基本理论,继而实验研究了小波变换在滚动轴承故障诊断中不同方面的应用:信号去噪方面和信号奇异性检测。

第四章:简单介绍了人工神经网络的基本原理,并对 BP 神经网络的工作原理进

行了系统的阐述。重点讨论研究了 BP 算法的不足以及各种优化算法，并通过实验对比分析研究了各种优化算法的优缺点，并从中得出本课题提出的优化算法用于神经网络中是可行有效的，其优化效果比其他优化算法更优越。

第五章：简单介绍了 BP 神经网络的设计以及实验数据的来源，然后结合小波变换和改进的 BP 神经网络提出了一种改进的滚动轴承故障诊断方法。首先利用小波变换对滚动轴承的动态信号进行精细的划分，提取滚动轴承动态信号的故障特征信息；然后利用改进的 BF 神经网络进行滚动轴承故障分类与诊断。这种方法既克服了传统 BP 的收敛速度、局部最小问题和训练瘫痪等缺点，极大地提高了诊断精度，实验仿真结果表明该方法的优越性和有效性。

第六章：对所进行的研究工作进行系统的总结和展望。

第二章 滚动轴承的振动理论

2.1 滚动轴承的振动机理

滚动轴承的典型结构一般由内圈、外圈、滚动体和支持架等四部分组成。滚动轴承在工作时，一般是外圈与轴承或机壳相连接，固定或相对固定，内圈与机械的传动轴相连接随轴转动。在机械运转时，由于轴承本身的结构特点、加工装配误差及运行过程中出现的故障等内部因素，以及传动轴上其他零部件的运动和力的作用等外部因素，当轴以一定的速度并在一定载荷下运转时对轴承和轴承座或外壳组成的振动系统产生激励，使该系统振动，其振动产生的机理如图 2-1 所示。

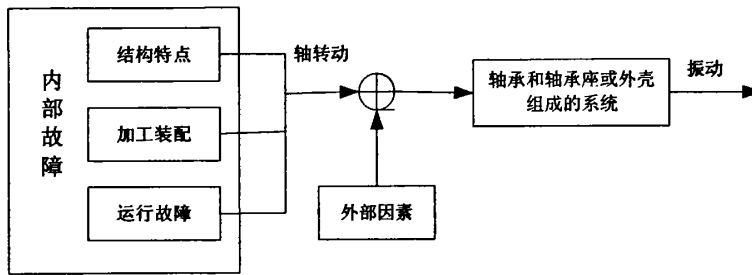


图 2-1 滚动轴承振动产生机理

显然在实际诊断中，通过布置在轴承座或外壳适当位置的传感器拾取的振动信号是上述各种内部和外部激励源施加于轴承以及轴承座和外壳组成的振动系统的综合振动。如果不考虑轴承加工和装配误差，在这一综合振动中，运行故障内部因素所引起的振动信号为振动因素。那么如何从综合振动中把运行故障引起的振动信号提取出来，从而有效地诊断出轴承的运行故障时轴承工矿监测的关键问题^[34]。

2.2 滚动轴承振动类型分析

在工作过程中，滚动轴承的振动主要可分为一下三类：（1）与轴承弹性有关的振动，轴承是弹性变形体，承受载荷时，由于承载滚动体的不断变换使得轴承在运行时发生弹性振动；（2）与轴承加工有关的振动，在加工轴承各元件中不可避免地会出现误差，比如擦衡、表面波纹、装配误差等都会引起轴承的振动；（3）与轴承使用有关的振动，轴承在实际工况运行中，由于发生疲劳剥落、断裂、破碎等故障而产生振动，其振动反映了轴承的损伤情况，该类振动是我们最感兴趣的，也是故

障监测中的主要分析对象。第一类振动与轴承是否异常无关，但它决定了振动系统的传递特性，后两类振动则反映了轴承的表面状况。以下详细地对滚动轴承的各种振动进行研究与分析。

2.2.1 滚动轴承的固有振动

滚动轴承在运行工作时，滚动体与内圈或外圈之间可能产生冲击，而诱发轴承元件的振动。当振动频率与轴承元件固有频率相等时，这一受迫振动将加剧。其固有频率仅仅取决于元件本身的材料，形状和质量，与转轴转速无关^[35]。

钢球滚动体的固有频率的计算公式为

$$f_{bm} = 4.8 \times 10^4 / r \quad (2-1)$$

式(2-1)中 r 为钢球的半径(mm)

轴承套圈径向弯曲振动的固有频率计算公式为

$$f_n = 9.4 \times 10^5 \frac{hn(n^2 - 1)}{b^2 \sqrt{n^2 + 1}} \quad (2-2)$$

式(2-2)中 n 为固有频率的阶数($n-1$)， $n=2,3,4,\dots$ ， b 套圈宽度(mm)， h 套圈厚度(mm)。

轴承元件固有频率可能在 20-60 KHz 范围内，通常将此段频率作为诊断频带。其原因为轴承元件的缺陷或不规则表面在运行中引起的脉冲冲击能激发轴承某个元件产生固有频率振动，从而放大了在此频带中的故障信号，提高了故障信号的信噪比。

2.2.2 滚动轴承发生若干故障时的振动特征

轴承系统在运行过程中，由于各种原因而产生的振动信号十分复杂。引起轴承振动的激励是多方面的，比如轴承元件的制造误差和装配误差、轴承疲劳点蚀、剥落、磨损等。不同的原因，对轴承系统的激励是不同的。下面对各种不同的激励进行详细的分析与研究^[36]。

(1) 精加工波纹引起的振动

制造轴承时，如在滚道上或滚动体上的精加工表面留有波纹，当凸起数达到某一量值时，就会产生如表 2-1 的振动，表中的 z 为滚动体个数， f_s 为轴旋转频率， $n=1,2,3,\dots$ 。

由表 2-1 可知波纹度一般将引起比滚动体在滚道上的通过频率高很多的高频振动、噪声及轴心的振摆，其结果不会引起轴承径向振动，在一定的条件下还会引起轴向振动。

(2) 轴承元件本身构造引起的振动

滚动轴承在承受一定载荷时，由于不同位置承载滚动体数目不一样，因而承载刚度有所变化，引起轴心起伏波动^[35]。这时的振动称为滚动体传输振动，它由滚动

体的公转产生，其主要频率成分为 zf_b ， f_b 为滚动体频率。

当滚动体直径不一致，一个滚动体直径比其它滚动体大时，轴心随着滚动体的频率 f_b 而变动；另外 f_b 因轴向的刚性不同而产生差别，发生包括 f_b 以及 $nf_b \pm f_s$ （其中 $n=1,2,\dots$ ）两种频率成分的振动，其振动频率一般在一千赫兹以下，没有像固有振动那样的频率成分。

对于旋转轴弯曲或轴承与轴撞歪的情形时，将发生具有 $zf_b \pm f_s$ 频率成分的振动。

表 2-1 精加工波纹引起的振动

故障原因	特征频率	波纹的凸起数
外圈波纹	nf_o	$nz \pm 1$
内圈波纹	$nf_i \pm f_s$	$nz \pm 1$
滚动体波纹	$2nf_b \pm f_o$	$2n$

(3) 轴承刚性非线性引起的振动

滚动轴承是靠滚到与滚动体的弹性接触来承受载荷的，具有弹簧的性质。当轴承出现润滑状态不良时，就会出现非线性弹簧的特性而产生振动，其频率有轴的旋转频率 f_s 、高次谐波 $2f_s$ ， $3f_s$ ， \dots 、分数谐波 $1/2f_s$ ， $1/3f_s$ ， \dots 。但该类振动一般发生在深槽球轴承上，在自动调心轴承和滚子轴承不常发生。

(4) 轴承不同轴引起的振动

当轴承不对中或轴承装配不良时都会发生如表 2-2 所示的低频振动。

表 2-2 轴承不同轴引起振动的振动特征

故障原因	特征频率	备注
a) 两个轴承不对中		
b) 轴承架内表面划伤或进入异物	$1/2f_s$	发生左栏中频率的振动，球
c) 轴承支架的装配部分松动		轴承容易发生振动
d) 轴承本身装配不良		
a) 内圈面的圆度差	$2f_s$	发生左栏中频率的振动，
b) 轴径圆度差		球轴承容易发生振动

(5) 轴承损伤而引起的振动

a) 当轴承偏心时，轴的中心（内圈中心）便以外圈中心为中心作振摆。这时的振动频率为 nf_s ， f_s 为旋转频率，其中 $n=1,2,\dots$ 。

b) 轴承磨损时，即轴承滚道、滚动体、保持架等由于机械原因或润滑杂质引起表面磨损，它是一种渐变性故障，磨损后振动水平（幅值）明显高于正常轴承。

c) 外环有点蚀故障时, 则产生的振动频率是外环一点和(单一)滚动体相接触的基本频率 f_s 再乘以滚动体个数 z 的乘积 zf_s 及其高次谐波 nzf_s 。由于损伤处的位置和承载方向之间的位置关系是一定的, 所以与振幅调制无关。

d) 内圈有点蚀时, 即当滚动轴承内圈的某个部分存在剥落、压痕、裂纹、损伤等故障时, 便会发生含有振动频率为 znf_i 、 $znf_i \pm f_s$ 等多种频率的振动。

e) 滚动体有点蚀故障时, 发生的振动与内圈点蚀一样由于存在径向间隙, 随点蚀与滚动体发生冲击的位置而产生的振动的振幅发生变化。其振动频率有 $2nf_b \pm f_s$ 以及 $2nf_b$ 。

2.2.3 轴承故障特征频率的计算

当滚动轴承出现外圈、内圈、滚动体出现点蚀等表面损伤类故障时, 会产生一定(特征)频率的宽带冲击信号, 引起轴承的减幅振荡。此类振荡发生的频率称为故障特征频率, 它由轴的转速、轴承几何尺寸以及损伤点的位置来确定。下面给出根据轴承各元件的相对运动关系来推出轴承故障特征频率的计算公式如下:

z 个滚动体与外圈上某一故障点接触的频率为

$$f_o = \frac{z}{2} \left(1 - \frac{d}{D} \cos \alpha\right) f_s \quad (2-3)$$

z 个滚动体与内圈上某一故障点接触的频率为

$$f_i = \frac{z}{2} \left(1 + \frac{d}{D} \cos \alpha\right) f_s \quad (2-4)$$

滚动体上某一故障点与外圈或内圈接触的频率为

$$f_b = \frac{D}{2d} \left[1 - \left(\frac{d}{D}\right)^2 \cos^2 \alpha\right] f_s \quad (2-5)$$

式(2-3)、(2-4)、(2-5)中内圈的旋转频率为 f_s , 轴承节径为 D (mm), 滚动体直径为 d (mm), 接触角为 α , 滚动个数为 z , f_c 为滚动体的公转频率, 即保持器的转动频率 f_b 。

f_o 、 f_i 、 f_b 分别称为外圈、内圈和滚动体的通过频率。当上述得到的某一个固定点是局部损伤点(比如点蚀点、剥落点、烧伤点等)时, f_o 、 f_i 、 f_b 分别称为局部损伤点撞击滚动轴承元件的频率, 所以又分别称为外圈、内圈和滚动体的故障特征频率。根据此特征频率可判断轴承发生故障的位置, 但需要强调的是随着故障程度的增加, 经常还会出现一些以特征频率为主频, 以轴频为差值的调制边频现象, 而当外圈、内圈和滚动体同时出现故障, 即多重故障时, 由特征频率来诊断故障相对来说是较为困难的。

2.3 轴承局部表面损伤类故障理论模型

滚动轴承在运行过程中出现的故障, 按其振动信号的特征不同, 可分为两大类:

(1) 表面损伤类故障, 包括点蚀、剥落、插伤等; (2) 磨损类故障。在正常使用的情况下, 滚动轴承工作表面磨损故障经历时间较长, 是一种渐变型故障, 不会马上引起轴承破坏, 其危害程度远小于表面损伤类故障, 因此重点研究表面损伤类故障。下面对详细地研究介绍一下轴承局部表面损伤类故障的理论模型。

2.3.1 外圈上有单一故障

假定传感器放置于轴承外圈上, 当轴承运转时, 传感器与故障点相对位置保持不变, 所以故障点引起的脉冲力的大小和方向也不变。并假定脉冲是理想脉冲, 及脉冲作用时间为零。

如果 $t=0$ 时刻刚好有一滚动体和故障点相接处而产生脉冲, 则外圈上局部故障点在轴承运转过程中所产生的系列脉冲力可表示为

$$\Delta_o(t) = \sum_{k=-\infty}^{+\infty} d_o(t - kT_o) \quad (2-6)$$

式 (2-6) 中 d_o 为脉冲力强度, $\delta(t)$ 单位脉冲函数, $T_o = 1/f_o$ 为脉冲时间间隔, f_o 为外圈故障特征频率, k 为正整数。

对于轴承来说在其运转过程中, 由于外圈故障点所引起的故障脉冲以频率 f_o 不断出现, 是比引起一系列的高频衰减振动, 因为阻尼通常比较大, 所以衰减较快, 所需时间与脉冲间隔周期相比要小得多, 因此各个衰减振动基本上是相对独立的, 一串脉冲力就有一串相互分离的高频衰减振动。对其进行幅值包络得到的是一系列指数衰减脉冲, 其包络波形及其频谱图可参见文献[37], 可表示为

$$e(t) = \begin{cases} e^{-t/T}, & t > 0 \\ 0, & t \leq 0 \end{cases} \quad (2-7)$$

则高频衰减振动的包络解调信号表示为

$$\psi_o(t) = \sum_{k=-\infty}^{+\infty} d_o \delta(t - kT_o) * e(t) \quad (2-8)$$

2.3.2 内圈上有单一故障

同外圈相比, 内圈上有故障点时情况较为复杂, 主要是因为大型回转轴承运转时, 故障点相对固定不动, 而传感器和载荷随着外圈的转动而变化, 因为故障点有时处于载荷区内, 有时处于载荷区外。当故障点位于载荷区内, 它与滚动体接触产生脉冲力, 且脉冲力的大小和仿效与故障点所处的位置有关; 而当故障点处于载荷区外时, 如果不考虑运动惯性力的作用, 产生脉冲力可忽略不计。由此可知内圈故障引起的脉冲力大小和方向受载荷的分布和故障点位置的影响。

假定不考虑载荷分布和故障点的位置, 则故障点与滚动体接触一次就产生一个

脉冲力，且产生的脉冲力大小一致，这样由内圈故障引起的一系列脉冲力可表示为

$$\Delta_i(t) = \sum_{k=-\infty}^{+\infty} d_i(t - kT_i) \quad (2-9)$$

式(2-9)中 d_i 为脉冲力强度， $\delta(t)$ 单位脉冲函数， $T_i = 1/f_i$ 为脉冲时间间隔， f_i 为内圈故障特征频率， k 为正整数。

由文献[38]可得轴承受径向载荷时的载荷分布表达式为

$$q(\varphi) = q_{\max} \left[1 - \frac{1}{2\varepsilon} (1 - \cos \varphi) \right]^n \quad (2-10)$$

其中对于球轴承， $n=1.5$ ；对于滚动轴承， $n=1.1$ 。载荷分布角可表示为

$$|\varphi| \leq \varphi_{\max} = \cos^{-1} \frac{C_d}{2\delta_{\max} + C_d} \quad (2-11)$$

载荷分布系数可表示为

$$\varepsilon = \frac{1}{2} \left(1 - \frac{C_d}{2\delta_{\max} + C_d} \right) \quad (2-12)$$

式(2-11)中 C_d 为径向游隙， δ_{\max} 最大接触变形。无径向游隙时， $C_d = 0$ ， $\varepsilon = 0.5$ ， $\varphi_{\max} = 90^\circ$ ；有径向游隙时， $C_d > 0$ ， $\varepsilon < 0.5$ ， $\varphi_{\max} < 90^\circ$ 。

假定内圈损伤点与滚动体接触时产生的脉冲力产生的脉冲力方向与传感器轴线方向成 φ 角，所以在传感器轴线方向的投影为脉冲力的大小成与 $\cos \varphi$ 之积，如果用 $p(\varphi)$ 表示脉冲力作用方向的影响，则有

$$p(\varphi) = \cos \varphi \quad (2-13)$$

假定脉冲力的大小与其作用位置处的载荷分布密度成正比，令 $\varphi = 2\pi f_s t$ (f_s 为轴频)，则作用于传感器轴线方向的脉冲力大小为

$$f_i(t) = \Delta_i(t) q(2\pi f_s t) p(2\pi f_s t) \quad (2-14)$$

则载荷分布和脉冲作用方向影响的相位一致，于是令

$$C_i(t) = q(2\pi f_s t) p(2\pi f_s t) \quad (2-15)$$

则

$$f_i(t) = \Delta_i(t) C_i(t) \quad (2-16)$$

单一内圈故障点引起的振动包络信号为

$$\varphi_i(t) = A_i [\Delta_i(t) C_i(t)] * e(t) \quad (2-17)$$

式(2-17)中 A_i 为力与振动之间的转换系数。

根据时域卷积和频率乘积之间的对应关系，其对应的频谱为

$$\Phi_i(f) = A_i [\Delta_i(f) * C_i(f)] E(f) \quad (2-18)$$

2.3.3 滚动体上有单一故障

当滚动体故障点与外圈接触时产生的脉冲力直接传播至外圈，而当滚动体故障

点与内圈接触时产生的脉冲力要经过滚动体和外圈的界面传播后才作用于外圈，由于在滚动体内及通过界面传播时能量损失较大，所以这个脉冲力的强度肯定比产生时要小得多。在不考虑载荷分布情况和脉冲力作用方向的影响时，有滚动体上一个故障点所引起的系列脉冲力可表示为

$$\Delta_b(t) = \Delta_{bo}(t) + \Delta_{bi}(t) \quad (2-19)$$

其中

$$\Delta_{bo}(t) = \sum_{k=-\infty}^{+\infty} d_{bo} \delta(t - kT_b) \quad (2-20)$$

$$\Delta_{bi}(t) = \sum_{k=-\infty}^{+\infty} d_{bi} \delta(t - kT_b - \frac{1}{2}T_b) \quad (2-21)$$

式(2-20)和式(2-21)中 $d_{bo} > d_{bi}$ ， d_{bo} 为滚动体上故障点与外圈接触产生的脉冲力强度， d_{bi} 为滚动体上故障点与内圈接触产生的脉冲力强度， $T_b = 1/f_b$ 为滚动体自转周期。

参照前面分析内圈故障点的方法，用函数 $C_b(t)$ 表示载荷分布和脉冲力作用方向影响，可得

$$C_b(t) = q(2\pi f_c t) p(2\pi f_c t) \quad (2-22)$$

式(2-22)中 f_c 为滚动体公转频率。

滚动体故障点引起的振动包络信号为

$$\varphi_b(t) = \varphi_{bo}(t) + \varphi_{bi}(t) \quad (2-23)$$

其中

$$\varphi_{bo}(t) = A_b [\Delta_{bo}(t) C_b(t)] * e(t) \quad (2-24)$$

$$\varphi_{bi}(t) = A_b [\Delta_{bi}(t) C_b(t)] * e(t) \quad (2-25)$$

同样根据时域和频域中卷积和乘积之间的对应关系可得到包络频谱为

$$\Phi_b(f) = \Phi_{bo}(f) + \Phi_{bi}(f) \quad (2-26)$$

其中

$$\Phi_{bo}(f) = A_b [\Delta_{bo}(f) * C_b(f)] E(f) \quad (2-27)$$

$$\Phi_{bi}(f) = A_b [\Delta_{bi}(f) * C_b(f)] E(f) \quad (2-28)$$

两个脉冲信号的相位差为

$$\Delta\varphi_b = -m\pi \quad (2-29)$$

其中 $m = 0, 1, 2, \dots$ ；对应于 $f = 0, f_b, 2f_b, \dots$ 。

再由矢量加法可得

$$|\Phi_b(f)| = [\Phi_{bo}(f)^2 + \Phi_{bi}(f)^2 + 2|\Phi_{bo}(f)| \cdot |\Phi_{bi}(f)| \cdot \cos \Delta\varphi_b]^{1/2} \quad (2-30)$$

将(2-29)代入可得

$$\Phi_b(f) = \begin{cases} |\Phi_{bo}(f)| + |\Phi_{bi}(f)|, m = 0, 2, 4, \dots \\ |\Phi_{bo}(f)| - |\Phi_{bi}(f)|, m = 1, 3, 5, \dots \end{cases} \quad (2-31)$$

由此可以知道, 尽管滚动体上的故障点自转一周时分别与外圈和内圈接触一次, 产生两个脉冲力。但由于与内圈接触产生的脉冲力经传递后大大衰减, 作用到传感器上比外圈接触产生的脉冲力要小得多, 因此滚动体的故障特征频率应为 f_b 而不是 $2f_b$ 。

2.3.4 轴承元件有多个故障

用前面的分析理论可用来分析滚动轴承元件表面有两个或两个以上故障点的包络谱, 由于本文重点讨论单一故障, 所以多点故障的分析的过程及方法可参考文献[37]。

滚动轴承元件存在多个故障点时解调谱是每个故障点引起的幅值谱的矢量和, 得到的频谱线不像单个故障点时那么有规律, 在以特征频率各阶倍频为中心的谱线分布中, 不一定是故障特征频率倍频幅值最大。并从文献[38]中得知利用矢量叠加原理还可以分析滚动轴承元件有三个或三个以上故障点时的包络幅值谱, 但前提是假设传感器安装在载荷最大的地方, 如果与实际测试中有出入的话, 只需要考虑信号的不同传递途径, 加入相应的传递函数就可以作相应的分析。

2.4 滚动轴承时域的统计特征

假设采集到的轴承振动信号表示为 $\{x_i\}$ ($i=1, 2, \dots, N$, N 为采样点数), 则信号的均方根值的表达式为

$$X_{RMS} = \sqrt{\frac{1}{N} \sum_{i=1}^N x_i^2} \quad (2-32)$$

把振动信号 $\{x_i\}$ 分成 n 段, 利用某一种峰值计数法从信号中找出 n 个峰值 $\{x_{pj}\}$ ($j=1, 2, \dots, n$), 则振动信号峰值的表达式为

$$X_{PEAK} = \frac{1}{n} \sum_{j=1}^n x_{pj} \quad (2-33)$$

峰值因子是表示波形是否有冲击的指标。假如传感器或放大器的灵敏度有所变动, 也不会出现测量误差, 因为峰值因子不受振动信号绝对水平所左右, 其表达式为

$$C = X_{PEAK} / X_{RMS} \quad (2-34)$$

裕度因子其表达式为

$$L = X_{PEAK} / X_r \quad (2-35)$$

式 (2-35) 中

$$X_r = \left[\frac{1}{N} \sum_{i=1}^N \sqrt{|x_i|} \right]^2 \quad (2-36)$$

波形因子为

$$S = X_{RMS} / X_{av} \quad (2-37)$$

式(2-37)中

$$X_{av} = \frac{1}{N} \sum_{i=1}^N |x_i| \quad (2-38)$$

脉冲因子为

$$I = X_{PEAK} / X_{av} \quad (2-39)$$

峭度因子是概率密度分布峭度程度的度量，表示振动波形中是否有冲击或尖峭程度的值，其表达式为

$$K = \frac{N \sum_{i=1}^N x_i^4}{X_{RMS}^4} \quad (2-40)$$

正常轴承的振动信号近似服从正态分布，所以其峭度值约为 3^[39]。当轴承发生故障时，其峭度值将大于 3，根据这个也可以判断轴承是否发生了故障^[40]。

轴承振动信号的时域参数可分为两大类：有量纲参数和无量纲参数。有量纲参数指标依赖历史数据并且对载荷和转速等变化比较敏感，而无量纲参数指标判断轴承故障上不受轴承型号、转速和载荷等因素的影响，无需考虑相对标准值或与以前的数据进行比较，另外它不受信号绝对水平的影响，所以即使测点同以往的位置略有不同，对参数计算结果不会产生明显的影响。由此选择峰值因子、脉冲因子、裕度因子、波形因子和峭度因子等五个参数组成轴承运行状态的特征参数。

2.5 本章小结

本章介绍了滚动轴承的振动机理，详细地研究分析了滚动轴承各个不同振动类型及其故障特征频率的计算。

重点研究了轴承的各个原件表面局部单一损伤类故障的理论模型，此外还简单介绍了滚动轴承的时域统计特征，为后面的研究工作奠定了必要的坚实的理论基础。

第三章 小波分析的理论与应用研究

3.1 傅立叶变换和短时傅立叶变换

设为系统的输入, $f(t) \in L^1(R)$, 即 $\int_{-\infty}^{\infty} |f(t)| dt < \infty$, 则函数 $f(t)$ 的连续傅立叶变换为

$$\hat{f}(\omega) = \int_{-\infty}^{\infty} e^{-i\omega t} f(t) dt \quad (3-1)$$

$\hat{f}(\omega)$ 傅立叶逆变换定义为

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \hat{f}(\omega) e^{i\omega t} d\omega \quad (3-2)$$

计算傅立叶变换, 需用数值积分, 即取 $f(t)$ 在 R 上的离散点上的值来计算这个积分。但实际应用中希望用计算机进行信号的频谱分析及其其他方面的处理工作, 所以要求信号在时域和频域上是离散的, 且为有限长。离散时间傅立叶变换 (Discrete Fourier Transform, DFT) 定义。

给定实的或复的离散时间序列 f_0, f_1, \dots, f_{N-1} , 设该序列绝对可积, 即满足

$\sum_{n=0}^{N-1} |f_n| < \infty$, 则序列 $\{f_n\}$ 的离散傅立叶变换为

$$X(k) = F(f_n) = \sum_{n=0}^{N-1} f_n e^{-i\frac{2\pi}{N}kn} \quad k = 0, 1, \dots, N-1 \quad (3-3)$$

序列 $\{X(k)\}$ 的离散傅立叶逆变换 (IDFT) 为

$$f_n = \frac{1}{N} \sum_{k=0}^{N-1} X(k) e^{-i\frac{2\pi}{N}kn} \quad n = 0, 1, \dots, N-1 \quad (3-4)$$

其中, k 是对频率域的离散化, n 是对时间域的离散化, 而且它们都是以 N 点为周期的。离散傅立叶变换序列 $\{X(k)\}$ 是以 2π 为周期, 且具有共轭对称性。

当 $f(t)$ 是实轴上以 2π 为周期的函数时, 即 $f(t) \in L^2(0, 2\pi)$, 则 $f(t)$ 可以表示成傅立叶级数形式, 即

$$f(t) = \sum_{n=-\infty}^{\infty} C_n e^{-in t} \quad (3-5)$$

为了研究信号在局部范围内的频域特征, D.Gabor 于 1946 年提出了 Gabor 变换。

其基本思想为：取 Gauss 函数 $g_a(t) = \frac{1}{2\sqrt{\pi a}} e^{-t^2/(4a)}$ 作为窗口函数，用 $g_a(t-b)$ 同待分析函数 $f(t)$ 相乘，然后在进行傅立叶变换。

$$G_b^a f(\omega) = \int_{-\infty}^{\infty} f(t) g_a(t-b) e^{-i\omega t} dt \quad (3-6)$$

Gabor 变换只是 STFT 的一种特例，STFT 的窗口函数不局限于 Gauss 函数。由式 (3-6) 可以得知，Gabor 变换把一维时域信号映射到二维时频域，即由 $t \rightarrow (\omega, b)$ ，精确分解了 $f(t)$ 的傅立叶变换，以便给出它的局部谱信息。这种局部信息的提供只是一种近似方法，它的精度取决于窗函数的选择。窗函数越宽，时频的局部性越差，而频域的局部性越好，反之亦然。

若函数 $g(t) \in L^2(R)$ ， $\hat{g}(\omega) \in L^2(R)$ ，则 $g(t)$ 被称为窗口函数，它的时窗与频窗中心定义为

$$\begin{cases} t_0 = \int_{\mathbb{R}} t |g(t)|^2 dt / \|g(t)\|_2^2 \\ \omega_0 = \int_{\mathbb{R}} \omega |\hat{g}(\omega)|^2 d\omega / \|\hat{g}(\omega)\|_2^2 \end{cases} \quad (3-7)$$

其时窗和频窗半径分别为

$$\begin{cases} \Delta_g = \left[\int_{\mathbb{R}} (t-t_0)^2 |g(t)|^2 dt / \|g(t)\|_2^2 \right]^{1/2} \\ \Delta_{\hat{g}} = \left[\int_{\mathbb{R}} (\omega-\omega_0)^2 |\hat{g}(\omega)|^2 d\omega / \|\hat{g}(\omega)\|_2^2 \right]^{1/2} \end{cases} \quad (3-8)$$

从概率的观点上理解，窗函数 $g(t)$ 在时域和频域上的能量集中在 $[t_0 - \Delta_g, t_0 + \Delta_g] \times [\omega_0 - \Delta_{\hat{g}}, \omega_0 + \Delta_{\hat{g}}]$ 范围内， Δ_g 和 $\Delta_{\hat{g}}$ 的大小表征着窗函数在时域和频域上局部化性质的好坏。但根据 Heisenberg 测不准原理，可知 $\Delta_g \times \Delta_{\hat{g}} \geq 1/2$ 。加窗 Fourier 变换的和是固定不变的，即构成“刚性”窗。当分析一个信号时，低频时时窗宽一些，希望高频时时窗窄一些，显然 STFT 不满足要求。

3.2 小波变换

对于平稳信号来说，Fourier 变换及各种基于 Fourier 变换的信号分析方法是有效的；对于非平稳信号，小波变换则是一种有效的信号处理方法，它是信号的时间-尺度分析方法，具有多辩分析的特点，且在时频域都具有表征信号局部特征的能力，是一种窗口大小固定不变但其形状可改变的时频局部化分析方法。即在低频部分具有较高的频率分辨率和较低的时间分辨率，在高频部分具有较高的时间分辨率和较低的频率分辨率^[41-43]。

设 $\psi(t)$ 为一平方可积函数，即 $\psi(t) \in L^2(R)$ ，若其傅立叶变换 $\hat{\psi}(\omega)$ 满足条件

$$C_\psi = \int_R \frac{|\hat{\psi}(\omega)|^2}{|\omega|} d\omega < \infty \quad (3-9)$$

则称 $\psi(t)$ 为一个基本小波或小波母函数，我们称式 (3-9) 为小波函数的容许条件。

将小波母函数 $\psi(t)$ 进行伸缩和平移，就可以产生一个函数族 $\psi_{a,b}(t)$

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}} \left(\frac{t-b}{a} \right) \quad a, b \in R; a > 0 \quad (3-10)$$

称为由小波基 $\psi(t)$ 生成的依赖参数 a, b 的分析小波或连续小波，其中 a 为伸缩因子， b 为平移因子。

小波母函数 $\psi(t)$ 的时频窗为 $[t_\psi - \Delta_\psi, t_\psi + \Delta_\psi] \times [\omega_\psi - \Delta_\omega, \omega_\psi + \Delta_\omega]$ ，则 $\psi_{a,b}(t)$ 的时频窗为 $[b + at_\psi - a\Delta_\psi, b + at_\psi + a\Delta_\psi] \times [\omega_\psi/a - \Delta_\omega/a, \omega_\psi/a + \Delta_\omega/a]$ 。可见对于低频信号，时间窗变宽而频率窗变窄；对于高频信号，时间窗变窄而频率窗变宽。因为频率与每单位时间的周期数成正比，所以取一个窄的时间窗可以更精确地确定高频现象，而取一个宽的时间窗可以更充分地分析低频特性^[44]。小波变换正好满足时频分析中自适应可调窗的要求，克服了加窗 Fourier 变换刚性时频窗的不足，所以小波变换被称为“数学显微镜”，可以聚焦到信号的任意局部细节。

3.2.1 连续小波变换

将任意 $L^2(R)$ 空间中的函数 $f(t)$ 在小波基下展开，称为这种展开为函数 $f(t)$ 的连续小波变换，其表达式为

$$W_f(a,b) = \langle f(t), \psi_{a,b}(t) \rangle = \frac{1}{\sqrt{a}} \int_R f(t) \overline{\psi\left(\frac{t-b}{a}\right)} dt \quad (3-11)$$

其中 $\overline{\psi}$ 为 ψ 的共轭函数。如果小波函数 $\psi(t) \in L^2(R)$ ，当其 Fourier 变换 $\hat{\psi}(\omega)$ 满足容许条件式 (3-9) 时，小波变换存在着逆变换，且具有以下小波反变换

$$f(t) = \frac{1}{C_\psi} \int_0^\infty \frac{da}{a^2} \int_{-\infty}^\infty W_f(a,b) \psi_{a,b}(t) db \quad (3-12)$$

容许性条件式 (3-9) 成立的必要条件为

$$\hat{\psi}(\omega) = \frac{1}{2\pi} \int_{-\infty}^\infty \psi_{a,b}(t) e^{-i\omega t} dt \Big|_{\omega=0} = \frac{1}{2\pi} \int_{-\infty}^\infty \psi_{a,b}(t) dt = 0 \quad (3-13)$$

式 (3-13) 表明 $\psi_{a,b}(t)$ 必须为衰减的振荡波形（能量有限），即 $\psi_{a,b}(t)$ 必须具有小的波形，这便是其称之为“小波”的原因^[45]。

函数 $f(t)$ 经过小波变换后，把一个一元函数 $f(t)$ 映射为伸缩因子为 a 和平移因子为 b 的二元函数，如果固定尺度 a ，则小波变换反映的是信号 $f(t)$ 在某一固定频带内随时间 t 的演变情况；如果令尺度 a 变换，则小波变换反映的是信号 $f(t)$ 在各个频带

内随时间 t 的变化情况。因此只要选全一个适当的小波函数,就有可能观察到信号的局部频率特性以及这种局部频率特性随时间的演变情况。在振动信号的处理中,它可由粗到细逐步给出振动信号在不同尺度下的波形。

信号处理的关键是用无冗余的信息完整地表达原始信号^[46],但是连续小波变换系数有很大的冗余量,造成计算的极不方便与低效率,因为信号完全有可能只需要其中的一个子集来重构。为此,通过加强小波基的选取条件,通过对伸缩因子 a 和平移因子 b 不同离散化产生了二进小波、R 小波等,为信号有效分解和重构提供了广阔的空间。

3.2.2 离散小波变换

令 $a = a_0^j$, $b = ka_0^j b_0$, 其中 $j, k \in Z$, $a_0 > 1$, 则相应的离散小波函数为

$$\psi_{j,k}(t) = a_0^{-j/2} \psi(a_0^{-j}t - kb_0) \quad (3-14)$$

相应的离散小波变换为

$$W_f(a_0^j, ka_0^j b_0) = \langle f(t), \psi_{j,k}(t) \rangle = \int_{-\infty}^{\infty} f(t) \overline{\psi_{j,k}(t)} dt \quad (3-15)$$

取 $a_0 = 2$, 即当 $a = 2^j$, $b = k2^j b_0$ 时, 若 $\psi_{b_0,j,k} = 2^{-j/2} \psi(2^{-j}t - kb_0)$ ($j, k \in Z$) 满足

稳定性条件 $A \leq \sum_{-\infty}^{\infty} |\hat{\psi}(2^{-j}\omega)|^2 \leq B$, $0 < A \leq B < \infty$, 则生成 $L^2(R)$ 的小波框架 $\{\psi_{b_0,j,k}\}$ 。

当 $\{\psi_{b_0,j,k}\}$ 线性无关时, 则构成 $L^2(R)$ 的 Riesz 基。当抽样速率 $b_0 = 1$ 时, Riesz 基为

$\{\psi_{j,k}(t) = 2^{-j/2} \psi(2^{-j}t - k)\}$, 只需 a, b 二进离散点集上的分析结果也能实现对原信号

的完全重构, 此时的 $\psi(t)$ 成为 R 函数。

当 R 函数的对偶存在且满足双正交性质时, 该 R 函数成为 R 小波, 满足进一步的条件还可构成正交和半正交小波。由此可得知, 从基小波到 R 小波, 小波基构成条件逐渐苛刻, 但小波变换结果的信息冗余度在不断降低^[47]。

3.2.3 多分辨分析和 Mallat 算法

(一) 多分辨分析

为了构造上述提到的 R 小波, Mallat 等^[48]提出了多分辨分析, 统一了小波函数的构造, 并在此基础上提出了 Mallat 塔式算法。

定义函数 $\phi(t) \in L^2(R)$ 为尺度函数, 若其整数平移系列 $\phi_k(t) = \phi(t - k)$ 满足

$$\langle \phi_k(t), \phi_{k'}(t) \rangle = \delta_{kk'} \quad (3-16)$$

其中

$$\delta_{kk'} = \begin{cases} 1, & k = k' \\ 0, & k \neq k' \end{cases} \quad (3-17)$$

定义 $\phi_k(t)$ 由在 $L^2(\mathbb{R})$ 空间张成的闭子空间为 V_0 , 称为零尺度空间

$$V_0 = \overline{\text{span}_k \{\phi_k(t)\}} \quad k \in \mathbb{Z} \quad (3-18)$$

对于任意 $f(t) \in V_0$, 有

$$f(t) = \sum_k a_k \phi_k(t) \quad (3-19)$$

假设 $\phi(t)$ 尺度函数在平移的同时又进行了尺度的伸缩, 则可得到一个尺度和位移均可变化的函数集合, 即

$$\phi_{j,k}(t) = 2^{-j/2} \phi(2^{-j}t - k) = \phi_k(2^{-j}t) \quad (3-20)$$

则称每一个固定尺度 j 上的平移系列 $\phi_k(2^{-j}t)$ 所张成的空间 V_j 为尺度 j 的尺度空间, 即

$$V_j = \overline{\text{span}_k \{\phi_k(2^{-j}t)\}} \quad k \in \mathbb{Z} \quad (3-21)$$

对于任意 $f(t) \in V_j$, 有

$$f(t) = \sum_k a_k \phi_k(2^{-j}t) \quad (3-22)$$

由上可知, 尺度函数 $\phi(t)$ 在不同尺度下其平移系列张成了一系列的尺度空间 $\{V_j\}_{j \in \mathbb{Z}}$ 。由式 (3-20) 可知, 随着尺度 j 的增大, 函数 $\phi_{j,k}(t)$ 的定义域变大, 实际的平移间隔也变大, 则它的线性组合不能表示 (小于该尺度) 的细微变化, 因此其张成的尺度空间只能包括大尺度的缓变信号。相反, 随着尺度 j 的减小, 函数 $\phi_{j,k}(t)$ 的定义域变小, 实际的平移间隔也变小, 则它的线性组合能表示函数的更细微的变化, 因此其张成的尺度空间所包含的函数增多, 尺度空间增大。

多分辨分析是指满足下述性质的一系列闭子空间 $\{V_j\}$, $j \in \mathbb{Z}$:

(a) 一致单调性: $\dots V_2 \subset V_1 \subset V_0 \subset V_{-1} \subset \dots$

(b) 渐近完全性: $\bigcap_{j \in \mathbb{Z}} V_j = \{0\}$; $\bigcup_{j \in \mathbb{Z}} V_j = L^2(\mathbb{R})$

(c) 伸缩规则性: $f(t) \in V_j \Leftrightarrow f(2^j t) \in V_0 \quad j \in \mathbb{Z}$

(d) 平移不变性: $f(t) \in V_0 \Leftrightarrow f(t-k) \in V_0 \quad k \in \mathbb{Z}$

(e) 正交基存在性: 存在 $\phi(t) \in V_0$, 使得 $\{\phi(t-n)\}_{n \in \mathbb{Z}}$ 是 V_0 正交基, 即

$$V_0 = \overline{\text{span}_n \{\phi(t-n)\}} \quad \int_{\mathbb{R}} \phi(t-n)\phi(t-m)dt = \delta_{m,n} \quad (3-23)$$

其中正交基存在性条件可放宽为 Riesz 基存在性。若 $\{\phi(t-k)\}$ 为空间 V_0 的正交

基, 则 $\{\phi_{j,k}(t)\}_{k \in \mathbb{Z}}$ 必为子空间 V_j 的标准正交基。

由多分辨率分析的定义可知, 所有的闭子空间 $\{V_j\}$ 都是由统一尺度函数 $\phi(t)$ 伸缩后的平移系列张成的尺度空间, 它们具有包含关系, 但是不具有正交性, 因此它们的基 $\{\phi_{j,k}(t)\}_{k \in \mathbb{Z}}$ 在不同尺度间不具有正交性, 即 $\{\phi_{j,k}(t)\}_{k \in \mathbb{Z}}$ 不能作为空间 $L^2(\mathbb{R})$ 的正交基。

为了寻找一组空间 $L^2(\mathbb{R})$ 的正交基, 定义尺度空间 $\{V_j\}_{j \in \mathbb{Z}}$ 的补空间如下:

设 W_j 为 V_j 在 V_{j-1} 中的补空间, 即

$$V_{j-1} = V_j \oplus W_j, \quad W_j \perp V_j, \quad j \in \mathbb{Z} \quad (3-24)$$

显然任意子空间 W_n 与 W_m 是相互正交的, 并且 $W_n \perp W_m$, 当 $n \neq m$ 且 $n, m \in \mathbb{Z}$ 时, 由多分辨率分析定义知, $L^2(\mathbb{R})$ 可以表示为

$$L^2(\mathbb{R}) = \bigoplus_{j \in \mathbb{Z}} W_j \quad (3-25)$$

因此 $\{W_j\}_{j \in \mathbb{Z}}$ 构成了 $L^2(\mathbb{R})$ 的一系列正交子空间, 并且由式 (3-24) 得

$$W_j = V_{j-1} - V_j \quad (3-26)$$

若 $f(t) \in W_0$, 则 $f(2^{-j}t) \in V_{j-1} - V_j$, 即

$$f(t) \in W_0 \Leftrightarrow f(2^{-j}t) \in W_j \quad (3-27)$$

若设 $\{\psi_{0,k}(t)\}_{k \in \mathbb{Z}}$ 为空间 W_0 的一组正交, 由式 (3-27) 对所有尺度 $j \in \mathbb{Z}$,

$\{\psi_{j,k}(t) = 2^{-j/2} \psi(2^{-j}t - k)\}_{j,k \in \mathbb{Z}}$ 必为空间 W_j 的正交基。 $\psi_{j,k}(t)$ 的整个集合 $\{\psi_{j,k}(t)\}_{j,k \in \mathbb{Z}}$ 必然构成了空间 $L^2(\mathbb{R})$ 的一组正交基。 $\psi_{j,k}(t)$ 正是由同一母函数伸缩平移得到的正交小波基, 称 $\psi(t)$ 为小波函数, 相应地称 W_j 是尺度 j 的小波空间。

对于任意函数 $f(t) \in V_0$, 可以将其分解为细节部分 W_1 和大尺度逼近部分 V_1 , 然后将大尺度逼近部分 V_1 进一步分解。如此重复就可得到任意尺度上的逼近部分和细节部分, 这就是多分辨率的框架。下面分析函数如何向尺度空间和小波空间投影的问题。

设 $f_s^j(t)$ 为函数 $f(t)$ 向尺度空间 V_j 投影后得到的尺度 j 下的概貌信号, 则

$$f_s^j(t) = \sum_k c_{j,k} \phi_{j,k}(t) \quad (3-28)$$

其中 $c_{j,k} = \langle f(t), \phi_{j,k}(t) \rangle$, 称为尺度展开系数。

若将函数 $f(t)$ 向不同尺度的小波空间 W_j 投影, 则可得到不同尺度下的细节信号 $f_d^j(t)$, 其表达式为

$$f_d^j(t) = \sum_k d_{j,k} \psi_{j,k}(t) \quad (3-29)$$

其中 $d_{j,k} = \langle f(t), \psi_{j,k}(t) \rangle$, 称为小波展开系数。

若将 $f(t) \in L^2(R)$ 按以下空间组合展开, 即

$$L^2(R) = \sum_{j=-\infty}^J W_j \oplus V_j \quad (3-30)$$

其中 J 为任意设定的尺度, 则

$$f(t) = \sum_{j=-\infty}^J \sum_{k=-\infty}^{\infty} d_{j,k} \psi_{j,k}(t) + \sum_{k=-\infty}^{\infty} c_{J,k} \phi_{J,k}(t) \quad (3-31)$$

当 $J \rightarrow \infty$ 时, 式 (3-30) 变为

$$f(t) = \sum_{j=-\infty}^J \sum_{k=-\infty}^{\infty} d_{j,k} \psi_{j,k}(t) \quad (3-32)$$

式 (3-28) 和式 (3-29) 成为离散正交小波变换的分解公式, 式 (3-31) 和式 (3-32) 称为离散正交小波变换综合公式。

由多分辨率分析得知, $\phi(t)$ 、 $\psi(t)$ 分别为尺度空间 V_0 及小波空间 W_0 的一个标准正交基, 又由于 $V_0 \subset V_{-1}$, $W_0 \subset V_{-1}$, 所以 $\phi(t)$ 、 $\psi(t)$ 也必然属于空间 V_{-1} , 也即 $\phi(t)$ 、 $\psi(t)$ 可用 V_{-1} 空间的正交基 $\phi_{-1,n}(t)$ 线性展开为

$$\phi(t) = \sum_n h_0(n) \phi_{-1,n}(t) = \sqrt{2} \sum_n h_0(n) \phi(2t-n) \quad n \in Z \quad (3-33)$$

$$\psi(t) = \sum_n h_1(n) \phi_{-1,n}(t) = \sqrt{2} \sum_n h_1(n) \phi(2t-n) \quad n \in Z \quad (3-34)$$

其中展开系数 $h_0(n)$, $h_1(n)$ 为

$$h_0(n) = \langle \phi(t), \phi_{-1,n}(t) \rangle \quad h_1(n) = \langle \psi(t), \phi_{-1,n}(t) \rangle \quad (3-35)$$

式 (3-33) 和式 (3-34) 描述的是相邻二尺度空间基函数之间的关系, 二式称为二尺度方程。需注意的是, 二尺度关系存在于任意相邻尺度 j 和 $j-1$ 之间, 即

$$\phi_{j,0}(t) = \sum_n h_0(n) \phi_{j-1,n}(t) \quad n \in Z \quad (3-36)$$

$$\psi_{j,0}(t) = \sum_n h_1(n) \phi_{j-1,n}(t) \quad n \in Z \quad (3-37)$$

由二尺度方程可知, 滤波器组 $h_0(n)$ 和 $h_1(n)$ 系数描述了两尺度空间函数之间的内在联系, 并且唯一对应于 $\phi(t)$ 和 $\psi(t)$ 。以下几点是 $h_0(n)$ 和 $h_1(n)$ 一些重要性质。

(a) $h_0(n)$ 和 $h_1(n)$ 的总和分别为

$$\sum_n h_0(n) = 2 \quad n \in Z \quad (3-38)$$

$$\sum_n h_1(n) = 0 \quad n \in Z \quad (3-39)$$

(b) 频域初值

$$H_0(\omega = 0) = 1 \quad H_1(\omega = 0) = 0 \quad (3-40)$$

(c) 递推关系

$$\hat{\phi}(\omega) = \prod_{j=1}^{\infty} H_0(2^{-j}\omega) \quad (3-41)$$

$$\hat{\psi}(\omega) = H_1\left(\frac{\omega}{2}\right) \prod_{j=2}^{\infty} H_0(2^{-j}\omega) \quad (3-42)$$

(d) 正交性

由同一尺度间尺度函数的平移正交性，可推得

$$\langle h_0(n-2k), h_0(n-2l) \rangle = \delta(k-l) \quad (3-43)$$

式 (3-43) 说明 $h_0(n)$ 满足偶次移位正交性

由同一尺度间小波函数的平移正交性，可推得

$$\langle h_1(n-2k), h_1(n-2l) \rangle = \delta(k-l) \quad (3-44)$$

式 (3-44) 说明 $h_1(n)$ 也满足偶次移位正交性

并且，由同一尺度间的尺度函数和小波函数的正交条件，可推得

$$\langle h_0(n-2k), h_1(n-2l) \rangle = 0 \quad (3-45)$$

式 (3-45) 说明滤波器 $h_0(n)$ 和 $h_1(n)$ 之间也是偶次移位正交的。

(e) 滤波器 $H_0(\omega)$, $H_1(\omega)$ 的特性

$$|H_0(\omega)|^2 + |H_0(\omega + \pi)|^2 = 1 \quad (3-46)$$

$$|H_1(\omega)|^2 + |H_1(\omega + \pi)|^2 = 1 \quad (3-47)$$

关于多分辨分析的直观理解，在这里以一个三层的分解进行说明，其小波分解树如图 3-1 所示。从图 3-1 中可以明显的看出，多分辨分析只是对低频部分进行进一步分解，使得低频部分的分辨率越来越高，而高频部分不予以考虑。图 3-1 中 S 表示信号，A 表示信号分解后的低频部分，D 表示信号分解后的高频部分，末尾的序号表示小波分解的层数，分解具有的关系为： $S=A_3+D_3+D_2+D_1$ 。注明一下，如果要进行进一步的分解，则可以把低频部分 A3 分解成低频部分 A4 和高频部分 D4，以下再分解以此类推。

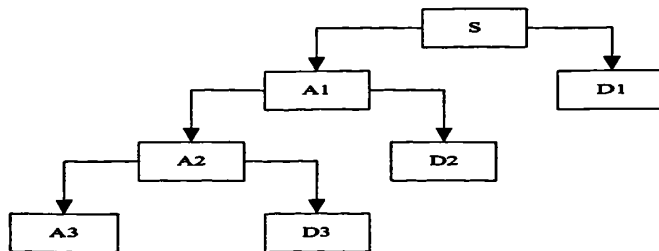


图 3-1 三层多分辨率分析树

有上述可知,对系统的输出信号之所以采用多分辨率分析,是因为信号的整个变换趋势(轮廓)和细节(突变)可以分别表示在不同分辨层次上,轮廓信息可以在低分辨率下提取,细节信息可以在较高分辨率下表现。

(二) Mallat 算法

将二尺度方程对时间进行伸缩和平移,则

$$\begin{aligned}\phi(2^{-j}t-k) &= \sqrt{2} \sum_n h_0(n) \phi(2^{-j+1}t-2k-n) & n \in Z \\ &= \sqrt{2} \sum_n h_0(n) \phi(2^{-j+1}t-2k-n)\end{aligned}\quad (3-48)$$

$$\begin{aligned}\psi(2^{-j}t-k) &= \sqrt{2} \sum_n h_1(n) \phi(2^{-j+1}t-2k-n) \\ &= \sqrt{2} \sum_n h_1(n) \phi(2^{-j+1}t-2k-n)\end{aligned}\quad (3-49)$$

令 $m=2k+n$, 则

$$\phi(2^{-j}t-k) = \sqrt{2} \sum_m h_0(m-2k) \phi(2^{-j+1}t-m) \quad (3-50)$$

$$\psi(2^{-j}t-k) = \sqrt{2} \sum_m h_1(m-2k) \phi(2^{-j+1}t-m) \quad (3-51)$$

由多分辨率分析,定义

$$V_{j-1} = \overline{\text{span}\{2^{(-j+1)/2} \phi(2^{-j+1}t-k)\}} \quad (3-52)$$

那么任意 $f(t) \in V_{j-1}$ 在 V_{j-1} 空间的展开式为

$$f(t) = \sum_k c_{j-1,k} 2^{(-j+1)/2} \phi(2^{-j+1}t-k) \quad (3-53)$$

将 $f(t)$ 分解一次(即分别投影到 V_j 和 W_j 空间),则有

$$f(t) = \sum_k c_{j,k} 2^{-j/2} \phi(2^{-j}t-k) + \sum_k d_{j,k} 2^{-k/2} \psi(2^{-j}t-k) \quad (3-54)$$

式(3-53)中 $c_{j,k}$ 和 $d_{j,k}$ 为尺度 j 上的展开系数,且

$$c_{j,k} = \langle f(t), \phi_{j,k}(t) \rangle = \int_{\mathbb{R}} f(t) 2^{-j/2} \overline{\phi(2^{-j}t-k)} dt \quad (3-55)$$

$$d_{j,k} = \langle f(t), \psi_{j,k}(t) \rangle = \int_{\mathbb{R}} f(t) 2^{-j/2} \overline{\psi(2^{-j}t-k)} dt \quad (3-56)$$

将式(3-50)代入式(3-53),得

$$c_{j,k} = \sum_m h_0(m-2k) \int_{\mathbb{R}} f(t) 2^{(-j+1)/2} \overline{\phi(2^{-j+1}t-m)} dt \quad (3-57)$$

由于

$$\int_{\mathbb{R}} f(t) 2^{(-j+1)/2} \overline{\phi(2^{-j+1}t-m)} dt = \langle f(t), \phi_{j-1,m}(t) \rangle = c_{j-1,m} \quad (3-58)$$

则式 (3-57) 变为

$$c_{j,k} = \sum_m h_0(m-2k)c_{j-1,m} \quad m \in Z \quad (3-59)$$

同理可得知

$$d_{j,k} = \sum_m h_1(m-2k)c_{j-1,m} \quad (3-60)$$

有上述可知, j 尺度空间的尺度系数 $c_{j,k}$ 和小波系数 $d_{j,k}$ 可有 $j-1$ 尺度空间的尺度系数 $c_{j-1,k}$ 经滤波器系数 $h_0(n)$ 和 $h_1(n)$ 进行加权求和得到。实际中的滤波器 $h_0(n)$ 和 $h_1(n)$ 长度都是有限的, 如紧支集正交小波、Haar 小波等; 或近似有限长的, 如样条小波等, 因此使分解运算非常简单。

将 V_j 空间尺度系数 $c_{j,k}$ 进一步分解下去, 可分别得到空间 V_{j+1} 、 W_{j+1} 的尺度系数 $c_{j+1,k}$ 和小波系数 $d_{j+1,k}$

$$c_{j+1,k} = \sum_m h_0(m-2k)c_{j,m} \quad (3-61)$$

$$d_{j+1,k} = \sum_m h_1(m-2k)c_{j,m} \quad (3-62)$$

同样将尺度空间 V_{j+1} 继续分解下去, 可得任意尺度空间 V_j 。式 (3-61) 和式 (3-62) 给出了一种小波分解的快速算法, 这就是著名的 Mallat 塔式分解算法。

系数重构的快速算法如下

设函数 $f(t) \in V_{j-1}$, 则有式 (3-53)

$$f(t) = \sum_k c_{j-1,k} 2^{(-j+1)/2} \phi(2^{-j+1}t - k) \quad k \in Z$$

或表示成尺度分解一次的形式 (3-54)

$$f(t) = \sum_k c_{j,k} 2^{-j/2} \phi(2^{-j}t - k) + \sum_k d_{j,k} 2^{-j/2} \psi(2^{-j}t - k)$$

将二尺度方程代入上式, 得

$$f(t) = \sum_k c_{j,k} \sum_n h_0(n) 2^{(-j+1)/2} \phi(2^{-j+1}t - 2k - n) + \sum_k d_{j,k} \sum_n h_1(n) 2^{(-j+1)/2} \psi(2^{-j+1}t - 2k - n)$$

上式两边同时对 $\phi_{j-1,m}(t)$ 进行内积, 得

$$c_{j-1,m} = \sum_k c_{j,k} h_0(m-2k) + \sum_k d_{j,k} h_1(m-2k) \quad (3-63)$$

式 (3-63) 则为 Mallat 塔式重构算法。

3.2.4 小波包算法

令正交小波基的滤波器分别为 $h_n = h_0(n)$ 和 $g_n = h_1(n)$, 并将尺度函数 $\phi(t)$ 记为 $u_0(t)$, 小波函数 $\psi(t)$ 改记为 $u_1(t)$, 则原来关于 $\phi(t)$ 和 $\psi(t)$ 的二尺度方程变为

$$\begin{cases} u_0(t) = \sqrt{2} \sum_k h_k u_0(2t-k) \\ u_1(t) = \sqrt{2} \sum_k g_k u_0(2t-k) \end{cases} \quad (3-64)$$

由公式

$$\begin{cases} u_{2n}(t) = \sqrt{2} \sum_{k \in Z} h_k u_n(2t-k) \\ u_{2n+1}(t) = \sqrt{2} \sum_{k \in Z} g_k u_n(2t-k) \end{cases} \quad (3-65)$$

定义的函数的集合 $\{u_n(t), n=0,1,2,\dots\}$ 为由 $u_0(t) = \phi(t)$ 所确定的小波包。由此，小波包 $\{u_n(t)\}$ 是包括尺度函数 $u_0(t)$ 和小波函数 $u_1(t)$ 在内的一个具有一定联系的函数的集合。

设 $\{u_n(t), n=0,1,2,\dots\}$ 是由标准正交化的尺度函数生成的，则小波包具有下列性质

a) 对于任意的 $n=0,1,2,\dots$ ，有

$$\langle u_n(\cdot, -j), u_n(\cdot, -k) \rangle = \delta_{jk} \quad j, k \in Z \quad (3-66)$$

b) 对于任意的 $n=0,1,2,\dots$ ，有

$$\langle u_{2n}(\cdot, -j), u_{2n+1}(\cdot, -k) \rangle = 0 \quad j, k \in Z \quad (3-67)$$

令

$$U_n = \overline{\text{span}\{u_n(t-k), k \in Z\}} \quad (3-68)$$

且定义空间

$$U_j^n = \overline{\text{span}\{2^{-j/2} u_n(2^{-j}t-k), k \in Z\}} \quad n=0,1,2,\dots \quad (3-69)$$

则空间 U_j^n 为由函数 $u_n(t)$ 在 j 尺度下的整数平移之线性组合所生成的子空间在 $L^2(R)$ 中的闭包，因此 $u_n(t)$ 在 j 尺度下的整数平移系列 $\{2^{-j/2} u_n(2^{-j}t-k), k \in Z\}$ 为空间 U_j^n 的一组正交基。

由于 $u_0(t) = \phi(t)$ ， $u_1(t) = \psi(t)$ ，且由正交小波基的理论知道，由 $\phi(t)$ 及 $\psi(t)$ 的伸缩平移生成的 $L^2(R)$ 的子空间为

$$\begin{cases} V_j = \overline{\text{span}\{2^{-j/2} \phi(2^{-j}t-k), k \in Z\}} \\ W_j = \overline{\text{span}\{2^{-j/2} \psi(2^{-j}t-k), k \in Z\}} \end{cases} \quad (3-70)$$

则由空间 U_j^n 的定义知

$$\begin{cases} V_j = U_j^0 \\ W_j = U_j^1 \end{cases} \quad j \in Z \quad (3-71)$$

又已知 V_j, W_j 具有性质

$$V_{j-1} = V_j \oplus W_j, \quad V_j \perp W_j, \quad j \in Z \quad (3-72)$$

因此，式 (3-72) 可以写为

$$U_{j-1}^0 = U_j^0 \oplus U_j^1, \quad U_j^0 \perp U_j^1, \quad j \in Z \quad (3-73)$$

现在把这一性质推广到小波包，则有

$$U_{j-1}^n = U_j^{2n} \oplus U_j^{2n+1}, U_j^{2n} \perp U_j^{2n+1}, j \in Z \quad (3-74)$$

小波空间 W_j 的二进细致分解为

$$W_j = U_{j+k}^{2^k} \oplus U_{j+k}^{2^k+1} \oplus \dots \oplus U_{j+k}^{2^k+1-1} \quad (3-75)$$

对取定的 $m = 0, 1, 2, \dots, 2^k - 1, k \in Z, j = 1, 2, \dots$ 函数系

$$\{2^{-(j+k)/2} u_{2^{j+k}m}(2^{-(j+k)}t-l)\}_{l \in Z} \quad (3-76)$$

是空间 $U_{j+k}^{2^k+m}$ 的标准正交基。小波包对 W_j 的分解相当于 W_j 的第 j 个频带被分割成 2^k 个子频带。

小波包的分解及重构的算法如下：

设 $f_j^n(t) \in U_j^n$ ，则 $f_j^n(t)$ 可表示为

$$f_j^n(t) = \sum_{l \in Z} d_l^{j,n} 2^{-j/2} u(2^{-j}t-l) \quad (3-77)$$

$f_j^n(t)$ 分解为 $f_{j+1}^{2n}(t)$ 与 $f_{j+1}^{2n+1}(t)$ ，小波包分解算法为

$$\begin{cases} d_l^{j+1,2n} = \sum_k h_{k-2l} d_k^{j,n} \\ d_l^{j+1,2n+1} = \sum_k g_{k-2l} d_k^{j,n} \end{cases} \quad (3-78)$$

小波包分解实际上是通过一组低、高通组合的共轭正交滤波器不断将信号分割到不同的频带上，滤波器组每作用一次，信号长度减少一半。

$f_{j+1}^{2n}(t)$ 与 $f_{j+1}^{2n+1}(t)$ 重构 $f_j^n(t)$ 的重构算法为

$$d_l^{j,n} = \sum_k (h_{l-2k} d_k^{j+1,2n} + g_{l-2k} d_k^{j+1,2n+1}) \quad (3-79)$$

以上分析介绍的小波包分析能够为信号提供一种更加精确的分析方法，它将频带多层次划分，对多分辨分析没有细分的高频部分进一步分解。它的直观理解就是：一个三层的小波包分析树如图 3-2 所示，图 3-2 中 S 表示信号，A 表示低频部分，D 表示高频部分，末尾的序号表示分解的层数。分解具有的关系为

$$S = AA3 + DAA3 + ADA3 + DDA3 + AAD3 + DAD3 + ADD3 + DDD3 \quad (3-80)$$

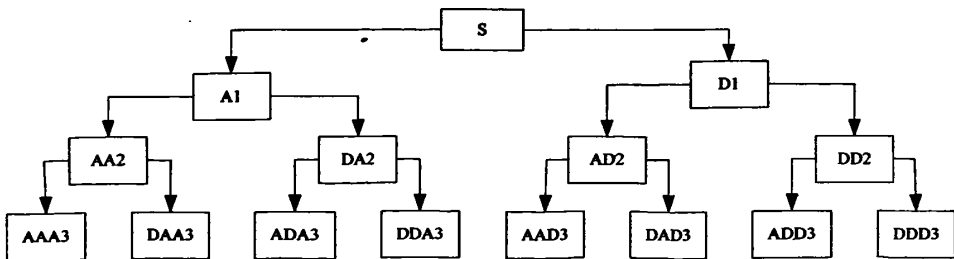


图 3-2 三层小波包分解树

3.3 小波变换在轴承故障诊断中的应用及实验研究

3.3.1 信号去噪

滚动轴承早期故障信号信号比较微弱，往往容易在采集和传输的过程中被其它随机噪声污染。如何把微弱的有用信号同噪声分开，提高信号的信噪比，为故障特征频率的提取提供可靠的依据，成为滚动轴承故障诊断中的重要任务之一。

传统的消噪方法主要有信号滤波、自适应消噪、时域平均等^[49]，这些方法难以将滚动轴承微弱的故障信号从背景信号中分离，也可能会将信号中一些能量小的有用信号当作噪声消除，有的实施起来也有一定的难度。而小波变换是一个频率和时间的局部变换，具有多分辨分析的特点，通过每层高频分解的系数进行阈值处理后重构，可有效地达到消除噪声的目的。在应用小波消噪时，关键是阈值门限的适当设置。通过对仿真信号和滚动轴承内圈故障振动信号的分析说明了该方法的实用性和有效性，并具有 Fourier 变化无法比拟的优越性。

1988 年 S.Mallat 提出了多分辨分析的概念及 mallat 算法。Mallat 算法又称塔式算法，其分解算法在本章 3.2.3 中有详细的介绍。下面简单介绍一下小波消噪的基本原理及相关的实验研究。

(一) 小波消噪的基本原理

假定含噪声信号的模型可以表示成如下形式^[50]：

$$y(t) = x(t) + \sigma * n(t), t = 0, 1, \dots, N-1 \quad (3-81)$$

其中， $y(t)$ 为含噪信号， $x(t)$ 为有用信号， $n(t)$ 为噪声信号。

在实际工程中，有用信号通常表现为低频信号或一些比较平稳的信号，而噪声信号则通常表现为高频信号。在这里假定滚动轴承故障诊断信号中的噪声为高斯白噪声，噪声级为 1。

基于小波变换的多分辨分析的 Mallat 算法，将信号分解为频率小于 2^j 的低频部分 $L(t)$ 和频率介于 2^j 和 2^{j+1} 之间的高频部分 $H(t)$ 。可采用以门限阈值等形式对小波系数进行处理，然后再由 Mallat 重构算法对信号进行重构即可以达到消噪的目的。对信号小波消噪^[51]的目的就是运用小波方法从被噪声污染的信号 $y(t)$ 中提取出有用信号 $x(t)$ 。

(二) 小波消噪的实验研究

(1) 小波消噪和 FFT 消噪的比较

本文选择了 Matlab 软件中的仿真信号 Blocks，该信号中含有若干不连续点和奇异点，因此拿它作为原始信号，信号长度为 $N=2000$ ，见图 3-3 (a) 所示。小波消噪方法中采用选择具有正交、对称性的小波基 sym7，对信号做三层分解，经多次实验表明，选用 heursure 软阈值规则可更有效的进行消噪处理。另外还采用傅立叶分析对信号进行消噪处理。

在原始信号中叠加高斯白噪声，信噪比为 7，见图 3-3 (b) 所示。分别用上述两种方法对信号进行消噪处理，实验结果分别见图 3-4 (a) 和图 3-4 (b) 所示。

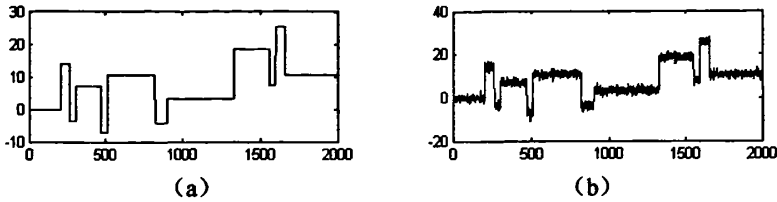


图 3-3 信号的时域图

(a) 原始信号 (b) 含噪信号

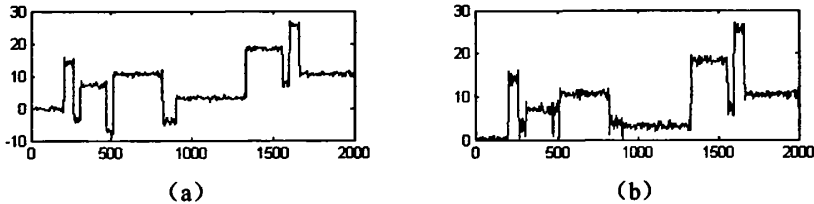


图 3-4 信号去噪后的时域图

(a) 小波去噪后信号 (b) 用傅立叶分析去噪信号

实验结果表明，用小波消噪可以很好地保存有用信号中的尖峰和突变部分，而用 FFT 分析进行滤波消噪，它不能将有用信号的高频部分和由噪声引起的高频干扰加以有效地区分。因此小波消噪有着傅立叶分析不可比拟的优点。

(2) 小波消噪阈值的选取

用小波对信号进行消噪过程中，都要用阈值进行小波分解系数的量化处理，其重要的环节就是阈值的选择和阈值的量化^[52]。小波消噪阈值的选取有以下四种规则：

(a) 基于 Stein 无偏似然估计(SURE)阈值估计(Rigrsure)；(b) 长度对数阈值(Sqtwolog)；(c) 启发式 SURE 阈值(Heursure)；(d) 极小极大方差阈值(Minimaxi)。

(a) 基于 Stein 无偏似然估计(SURE)阈值估计(Rigrsure)

无偏似然估计是一种基于 Stein 的自适应阈值选择。对一个给定的阈值 t ，得到它的似然估计，再将非似然 t 最小化，就得到了所选的阈值，它是一种软件阈值估计。设信号 $x(k)$ 为一个离散事件序列， $k=1,2,\dots,n$ ，令信号 $y(k)$ 为 $|x(k)|$ 的升序序列，再令

$$y_1(k) = y^2(k) \quad (3-82)$$

则阈值 thr_1 的计算公式为

$$y_2(k) = \sum_{i=1}^k y_1(i) \quad (3-83)$$

$$r(k) = \frac{n-2k+y_2(k)+(n-k)y_1(k)}{n} \quad (3-84)$$

$$thr_1 = \sqrt{\min(r)} \quad (3-85)$$

(b) 长度对数阈值 (Sqrtwolog)

固定阈值 thr_2 的计算公式为

$$thr_2 = \sqrt{2 \log(n)} \quad (3-86)$$

其中, n 为信号 $x(k)$ 的长度。

(c) 启发式 SURE 阈值 (Heursure)

启发式阈值原则是无偏似然估计原则和固定阈值原则的折中。如果信噪比很小, 按无偏似然估计原则处理信号; 如果噪声较大, 则采用固定阈值原则处理信号。若信号 $x(k)$ 的长度为 n , 令

$$eta = \frac{\|x\|^2 - n}{n} \quad (3-87)$$

$$crit = \frac{[\log(n)/\log 2]^{1.5}}{\sqrt{n}} \quad (3-88)$$

则阈值 thr_3 按如下计算

$$thr_3 = \begin{cases} thr_2, eta < crit \\ \min(thr_1, thr_2), 其他 \end{cases} \quad (3-89)$$

(d) 极小极大方差阈值 (Minimaxi)

极值阈值选择采用极大极小原理选择阈值, 它产生一个最小均方误差的极值。若信号 $x(k)$ 的长度为 n , 按极值阈值原则的阈值 thr_4 计算公式为

$$thr_4 = \begin{cases} 0, n < 32 \\ 0.3936 + \frac{0.1829 \log(n)}{\log 2}, n > 32 \end{cases} \quad (3-90)$$

为了得出不同阈值选择规则消噪结果的优劣, 继续采用上述仿真信号以及方法进行研究。实验结果如图 3-5 所示, 可以得出以上几种方法都能达到消噪较满意的效果, 经多次实验比较得出, heursure 软阈值消噪优越于其他三种方法。

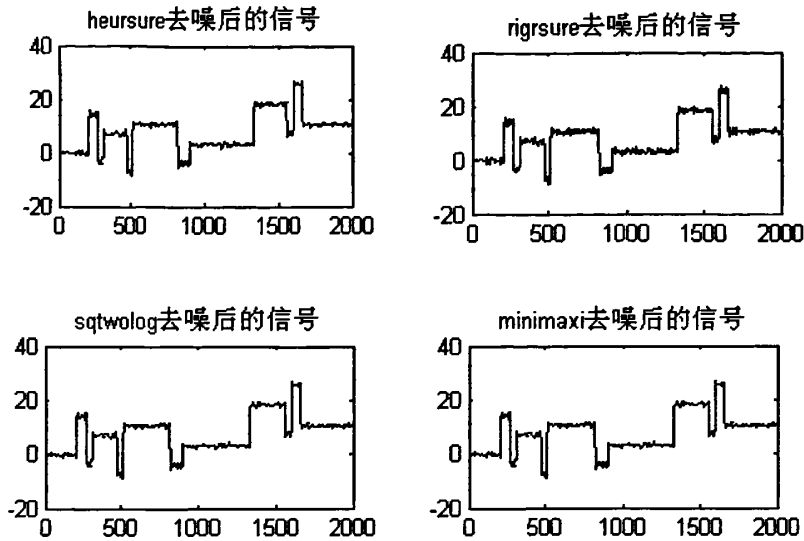


图 3-5 不同阈值选取的实验结果

(三) 小波消噪在滚动轴承故障诊断中的应用

采集带有内圈故障轴承（型号 6205-2RS JEM SKF）振动信号，试验装置和试验数据均来自美国 Case Western Reserve University 电气工程实验室^[53]。试验装置如图 3-6 所示，试验中通过对功率计的控制获得期望的扭矩载荷，振动加速传感器垂直固定在感应电机输出轴支撑轴承上方的机壳上进行数据采集。1.5kW 的三相感应电机通过自校准联轴节与一个功率计和一个扭矩传感器相连，最后驱动风机进行运转。

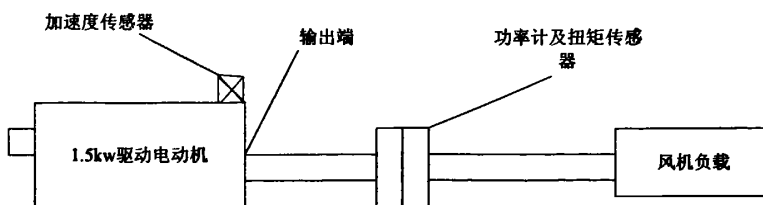
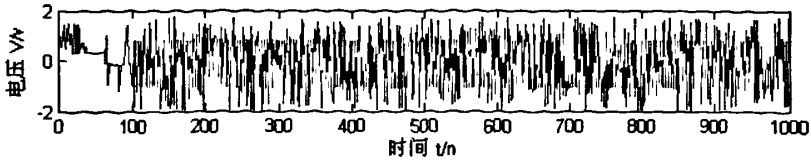
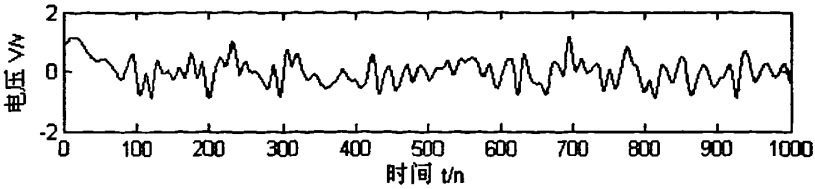


图 3-6 滚动轴承实验台示意图

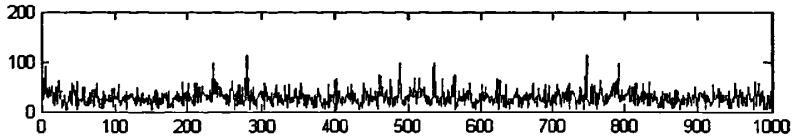
模拟的运行状态下滚动轴承内圈故障运行状态，采样频率为 12kHz，采样点数为 1024，其时域波形如图 3-7 (a) 所示。在图 3-7 (c) 原始振动信号的频谱图中，信号中存在峰值冲击，并由大量的噪声成分。若直接对其进行谱分析，不能准确判定故障的特征频率，也就不能判定故障的确切位置。



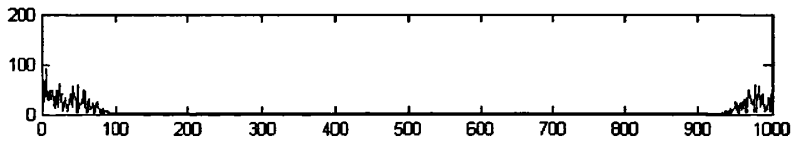
(a) 内圈故障原始信号时域波形



(b) 小波去噪后内圈故障信号时域波形



(c) 原始信号的谱图



(d) 消噪后信号的谱图

图 3-7 实验结果

经过多次实验对比研究,选择 sym8 小波对信号进行三层分解,选用 heursure 软阈值进行阈值处理,能更好地得到消噪后的信号,如图 3-7 (b) 所示。从图 3-7 (d) 中消噪后的信号谱图很容易发现频谱的峰值,并与实际理论相吻合。

研究和论述了小波分解和重构的算法,并给出了小波消噪的理论,在此基础上提出了小波消噪的滚动轴承故障诊断方法。滚动轴承振动信号是一种非平稳信号,存在着冲击成分,并容易受到随机噪声的污染。小波变换能有效地保留信号中的冲击成分,最大限度真实地反映原始信号本身的特性。通过实验证明小波消噪能更好地提高信噪比,为进一步故障诊断决策提供了可靠的依据。

3.3.2 信号奇异性检测

(一) 奇异性理论介绍^[54-56]

信号中的奇异点及不规则的突变部分经常带有比较重要的信息，它是信号重要的特征之一。信号的奇异性是指信号在某处有间断或某阶倒数不连续。反过来说无限次可导的函数是光滑的或没有奇异性。数学上常采用 Lipschitz 指数来度量奇异点的奇异性，Lipschitz 指数也称为奇异性指数。

若 n 是一非负整数， $n < \alpha \leq n+1$ ，如果存在着两个常数 A 和 $h_0 > 0$ ，即次多项式 $P_n(h)$ ，使得对任意的 $h \leq h_0$ ，均有

$$|f(t_0 + h) - P_n(t_0 - h)| \leq A|h|^\alpha \quad (3-91)$$

则 $f(t)$ 称为在 t_0 处的 Lipschitz α 指数。 $f(t)$ 在 t_0 处的 Lipschitz α 指数刻画了函数在该点的正则性。Lipschitz α 指数表征了信号在该点的奇异性大小，Lipschitz α 指数越大，函数越光滑；Lipschitz α 指数越小，该点的奇异性越大。函数在一点连续、可微，则在该点的 Lipschitz α 指数为 1；若 $f(t)$ 在 t_0 处的 Lipschitz α 下于 1，则称函数在 t_0 点是奇异性；一个在 t_0 点不连续但有界的函数，该点的 Lipschitz α 指数为 0。

信号和噪声具有不同的奇异性。奇异性不同，其小波变换模的极大值在不同尺度下的传播特性不同。小波变换能够通过多尺度分析提取信号的奇异点，基本原理是当信号在奇异点附近的 Lipschitz 指数 $\alpha > 0$ 时，其小波变换的模极大值随尺度的增大而增大；当 $\alpha < 0$ 时，则随尺度的增大而减小。而噪声对应的 Lipschitz 指数 α 远小于 0，而信号边沿对应的 Lipschitz 指数 α 大于或等于 0。因此利用小波变换可以区分噪声和信号边沿，信号突变点的小波变换域常对应于小波变换系数的极值点或过零点，并且信号的奇异性的大小同小波变换系数的极值随尺度的变换规律相对应。

在尺度 2^j 下，若 $\frac{d}{dx} W_2 f(x)$ 在 x_0 有一过零点，称 x_0 点是小波变换的局部极值点。

若在 x_0 的某一个邻域 δx_0 内，对一切 $x \in \delta x_0$ ，存在：

$$|W_2 f(x)| \leq |W_2 f(x_0)| \quad (3-92)$$

称 x_0 点是在尺度下的小波变换的模极大值点。

如果信号 $f(t)$ 在时间 t_0 点出现异常，那么在各个尺度上 t_0 点附近的 $|W_2 f(x)|$ 都会产生一个局部极大值点，且随着尺度的减小，这些模的局部极大值点收敛于 t_0 点。因此利用小波变换模极大值点检测出信号 $f(t)$ 的所有突变点，并且可根据小波变换模与尺度的关系而求出 Lipschitz 指数 α ，就是求出该点奇异性的大小。小波变换的模极大值都是出现在信号有突变的地方，并且突变点处高频成分较多，所有函数的奇异点可以从其小波变换的高频部分的模极大值检测出来。由奇异性的大小来判断机组出现异常的激烈程度。

采用以下方法来计算 Lipschitz 指数 α ：

设在第 j 层（尺度相对应为 2^j ）小波变换上的某一模极大值点 t_0 处的模极大值为 α_j ，即

$$\alpha_j = |W_j f(t_0)| \quad (3-93)$$

模极大值点处在各个尺度上的模极大值构成一序列 $\{\alpha_j | j=1 \sim J\}$, 在 j 较小时有如下的近似:

$$\alpha_j = |W_j f(t_0)| \approx A 2^{j\alpha} \quad (3-94)$$

两边去对数, 得

$$j\alpha = \log_2 \alpha_j - \log_2 A \quad (3-95)$$

从而有

$$\alpha = \log_2 \alpha_{j+1} - \log_2 \alpha_j \quad (3-96)$$

不同尺度下求出的 α 会略有不同, 将其平均即可得到 Lipschitz 指数 α 的估算值。

在轴承振动信号的处理中, 一般要检测的奇异性如间断与尖点其 Lipschitz 指数小于 1, 因而使用具有一阶消失矩的小波就可以了。

(二) 小波奇异性检测的实验研究

实验所用的振动信号数据来自不同状态的同一种轴承, 包括了正常及外圈故障、内圈故障、滚动体故障等四种情形。用上述的方法来计算振动信号总奇异点的奇异性。实验所用的某种轴承的参数如表 3-1 所示。

表 3-1 轴承参数

滚珠个数 z	接触角 α	轴承节径	滚动体直径	转速 f_s (Hz)
		D (mm)	d (mm)	
8	0	33.5	7.938	20

由第二章的式 (2-3)、式 (2-4)、式 (2-5) 得知, 外圈的故障特征频率为 153Hz, 内圈的故障特征频率 247Hz, 滚动体的故障特征频率 100Hz。由小波变换得到的四种状态下轴承振动信号奇异点如表 3-2 所示。

滚动轴承不同元件的局部故障都会引起信号中的冲击和瞬变, 但它们由 Lipschitz 指数刻划的奇异性却是不同的。且由表 3-2 可知, 在不同的状态下, 轴承振动信号奇异点的 Lipschitz 指数的取值范围有所不同。其中滚动体局部故障引发的冲击的奇异性最为严重, 又内圈局部故障引发的奇异性次之, 而由外圈局部故障引发的冲击奇异程度最轻。由此可知, 振动信号的奇异点的 Lipschitz 指数可以作为指示滚动轴承状态并进行故障诊断的一项指标。

表 3-2 滚动轴承振动信号的奇异性

轴承的状态	Lipschitz 指数	平均值
正常	0.544~0.786	0.651
外圈故障	0.346~0.407	0.374
内圈故障	0.273~0.360	0.325
滚动体故障	0.235~0.279	0.261

3.4 本章小结

本章首先介绍了 Fourier 变换和 Gabor 变换的基本原理，并指出其在实际应用中的局限性，比如频率混叠和栅栏效应等，一般适用分析平稳、缓变信号。

详细地讨论了连续小波变换、离散小波变换、Mallat 塔式算法、多分辨率分析及其小波包算法等小波理论的主要内容。

最后研究了小波变换在滚动轴承故障诊断两个方面的应用：(1) 信号去噪方面；(2) 小波奇异性方法克服了其他方法对滚动体故障不敏感的缺陷，可有效地用于滚动轴承的状态检测和故障诊断。滚动轴承的滚动表面出现局部故障时，其振动信号中包含了大量的冲击成分，对这些冲击成分进行分析有助于对滚动轴承的故障进行诊断，而数学上用 Lipschitz 指数来衡量信号突变点奇异性的一个定量指标，它由信号的小波变换求得。并且文中给出了 Lipschitz 指数的一种简便计算方法，避免了传统算法过大的计算量；由 Lipschitz 指数刻画的奇异性可以作为一项反映滚动轴承运行状态和故障类型的特征指标。研究表明滚动轴承不同元件的局部故障都会引起信号中的冲击和瞬变，但它们由 Lipschitz 指数刻划的奇异性却是不同的。

第四章 BP 神经网络算法的优化研究

4.1 人工神经网络的简介^[57-61]

人工神经网络(Artificial Neural Networks, ANNs)是一种旨在模仿人脑及其功能的信息处理系统。当前, 人工神经网络技术已被广泛应用于信号处理、自动控制、电子通信、故障诊断、金融预测与管理等各个领域。它具有以下几个特点:

(1) 良好的容错性和联想记忆功能

人工神经网络模型的信息储存是分布式的或全息式的, 模型各个部分所储存的信息互相支持, 相互补充, 从而赋予模型较强的容错性和联想记忆能力, 使其不会因为部分神经元的损坏而严重影响其总体功能, 也不会因为输入信号受到一定程度的噪声污染而严重歪曲其输出。所谓联想记忆就是从数据到数据的一种影射, 它的知识存储不仅是分布式的, 也是联想式的。

(2) 高度并行性

人工神经网络不但结构上是并行的, 它的处理顺序也是并行和同时的, 在同一层内的处理单元都是同时操作的。这种并行模式赋予模型高速的信息处理能力和对输入变化的快速响应能力, 解决了在线检测的问题。

(3) 自适应性

神经网络的自适应性包括以下四个方面: 学习性、自组织能力、推理能力以及可训练性^[58]。推理能力(即泛化能力)是网络对以前未曾遇见过的输入做出反应的能力, 泛化本身具备进一步学习和自调节的能力。

(4) 高度的非线性

工程界普遍面临的问题是如何实现输入空间到输出空间的非线性映射, 寻求输入到输出之间的非线性关系模型。然而对大部分无模型的非线性系统, 神经网络能很好的模拟。

神经网络是由众多简单的神经元连接而成, 它有三种基本层次结构: 前向网络、反馈网络和自组织竞争型网络, 其典型例子分别是误差反向传播网络(BP 网络)、hopfield 网络(由美国加州工学院物理学家 Hopfield JJ 提出)和 Kohonen 网络(由芬兰学者 Kohonen 提出)。由于 BP 神经网络是一种典型的神经网络且发展最为成熟, 所以在实际应用中绝大部分的模型都采用 BP 网络或它的变化形式, 它可以逼近任意的非线性映射关系^[59-61]。

4.2 BP 神经网络的工作原理

神经网络不需要预先给出关于模式的先验知识和判别模型，能够通过自组织和自学习机制自动形成所要求的决策域，有很强的模式识别能力，因此神经网络在设备的故障诊断中得到了广泛的应用。在人工神经网络的实际应用中，绝大部分都是采用 BP 网络或它的变化形式。BP 神经网络由输入层、隐层、输出层三层构成，它的拓扑结构如图 4-1 所示。

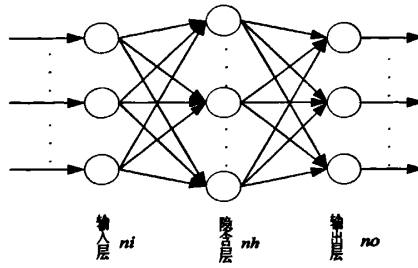


图 4-1 三层 BP 神经网络

输入层节点数 ni ，隐层节点数 nh ，输出层节点数 no ，样本数 np 。输入和期望输出矢量为 \bar{x}_p ， \bar{y}_p ， $p=1,2,\dots,np$ 。其中 $\bar{x}_p = \{x_{pi}\}_{i=1,2,\dots,ni}$ ， $\bar{y}_p = \{d_{pk}\}_{k=1,2,\dots,no}$ ， $j=1,2,\dots,nh$ 。 $w_{j,i}$ ， θ_j ， $w_{k,j}$ ， θ_k 分别为隐层和输出层的权值和阈值。目标误差为 ε ，各层神经元激活函数采用 S 型函数，即

$$f(x) = \frac{1}{1 + \exp(-x)} \quad (4-1)$$

对 p 个输入模式的误差计算公式为

$$E_p = \frac{1}{2} \sum_{k=1}^{no} (d_{pk} - y_{2_{pk}})^2 \quad (4-2)$$

式中 d_{pk} 为 BP 网络的期望输出， $y_{2_{pk}}$ 为网络的实际输出。

全部样本的均方误差为

$$E = \frac{1}{2 \cdot np} \sum_{p=1}^{np} \sum_{k=1}^{no} (d_{pk} - y_{2_{pk}})^2 \quad (4-3)$$

输入层输出计算公式为

$$y_{0_i} = x_i, \quad i=1,2,\dots,ni \quad (4-4)$$

隐含层输出计算公式为

$$\begin{cases} net_j = \sum_{i=1}^{ni} w_{ji} y0_i + \theta_j \\ y1_j = f(net_j) \end{cases} \quad j=1,2,\dots,nh \quad (4-5)$$

输出层输出计算公式为

$$\begin{cases} net_k = \sum_{i=1}^{nh} w_{ki} y1_i + \theta_k \\ y2_k = f(net_k) \end{cases} \quad k=1,2,\dots,no \quad (4-6)$$

由式(4-6)可得到实际输出值后,当实际输出值与期望值之间的误差大于可接受范围时就需要校正。其指导思想是对网络的权值 W 、阈值 θ 修正,使误差函数沿着负梯度方向下降,直到网络输出误差精度达到目标精度要求,学习结束。其过程中对于隐层和输出层的调整,每次调整权值增量 Δw_{kj} 满足

$$\Delta w_{kj} = -\eta \frac{\partial E}{\partial w_{kj}} = -\eta \frac{\partial E}{\partial net_k} \cdot \frac{\partial net_k}{\partial w_{kj}} = \eta \delta_k y1_j \quad (4-7)$$

式中

$$\delta_k = -\frac{\partial E}{\partial net_k} = -\frac{\partial E}{\partial y2_k} \cdot \frac{\partial y2_k}{\partial net_k} = (d_k - y2_k) \cdot f'(net_k) \quad (4-8)$$

对于非输出层同样根据上面的推导方法可求得

$$\Delta w_{ji} = \eta \delta_j y0_i \quad (4-9)$$

其中

$$\delta_i = \sum_{k=1}^{no} \delta_k w_{ki} f'(net_i) \quad (4-10)$$

式(4-9)中 η 为学习系数。

为使网络的输出误差趋于极小值,对于网络的输入的每一组训练模式,一般要经过很多次的循环记忆训练,才能使网络记住这一模式。当每次循环记忆训练结束后,都要进行学习结果的判别。判别的目的是主要是检查输出误差是否已经小到允许的程度。如果小到允许的程度,就可以结束整个学习过程,否则还要进行循环训练。BP算法的工作流程图如图4-2所示。

4.3 BP算法的不足及其优化算法的研究

4.3.1 BP算法的不足

BP算法尽管有很大的应用潜力,但由于采用梯度下降法,仍存在很多自身的限制和不足。主要体现在局部最小问题、收敛速度和训练瘫痪问题等方面。

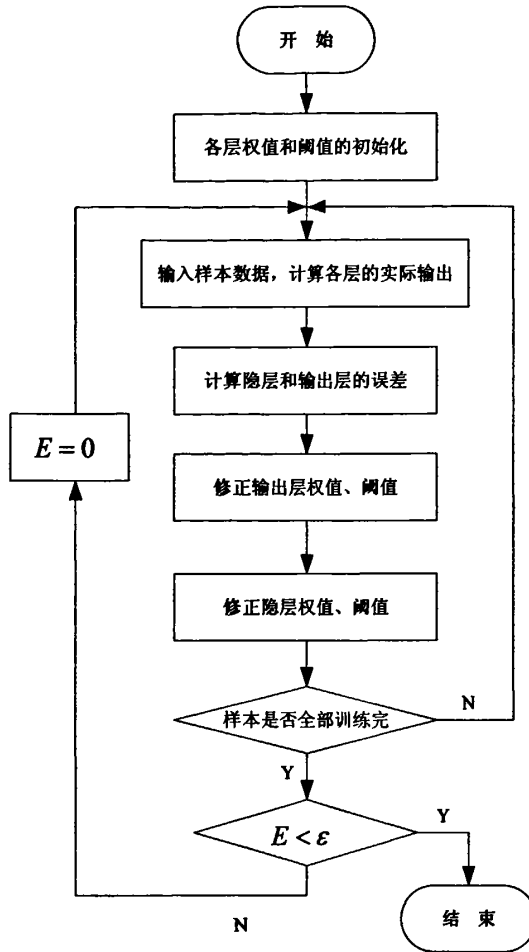


图 4-2 BP 算法的工作流程图

(1) 容易陷入局部最小问题: BP 学习算法用梯度下降法调整网络权值, 对于凸状的误差函数来说, 它有唯一的最小值。它可使网络权值收敛到一个解, 但不能保证所有为误差超平面的全局最小解, 但可能是一个局部最小解。且对实际问题中经常遇到的非凸状误差函数却不能取得最佳结果, 甚至在网络训练后也不知道误差函数是否取得全局最小。

(2) 网络学习收敛问题: 洛曼哈脱等人曾用偏微分方程作了证明, 当网络的权值的调整步长为无限小时, 收敛才有效。这意味着要用无限长的收敛时间。也就是说没有证明 BP 法在有限步内收敛。实际上在处理一些复杂的问题上, BP 算法可能要几个小时甚至更长时间的训练, 这主要是由于学习速率太小所造成。可采用变化的学习速率或自适应的学习速率来加以改进。

(3) 网络训练瘫痪问题: 在有些情况下, 网络训练过程中权值的修改实际上停顿了。一旦进入这种状态, 它的训练时间按数量级增长, 网络处于瘫痪状态。

4.3.2 传统的改进算法

(1) 统计算法

随机选取权值和阈值, 对于样本集合计算最后的输出, 训练过程中计算误差的平方和, 用最小随机量调整它, 如果误差平方和减少则保留改变后的权值和阈值, 否则恢复原权值和阈值。这种改进的好处不采用学习规则, 而是比较误差函数。但是计算速度较慢, 迭代次数比增加。

(2) Sigmoid 函数的改进

神经网络的输出域为 $[0, 1]$, 在实际应用中会遇到输出域在 $[-1, 1]$, 这种情况可用改进的 Sigmoid 函数 $f(x) = 2/[1 + \exp(-x)] - 1$, 这样扩大了输出范围。

(3) 加入动量项

在训练样本时, 在权值和阈值的修正算法中, 各加入一个相应的动量项

$$\Delta w(k+1) = -\eta \frac{\partial E}{\partial w} + \alpha w(k) \quad (4-11)$$

其中 α 为动量因子, 通常取 0 到 1 之间的某一值, 一般取 0.95 左右。增加动量项可以起到平滑梯度方向的剧烈变化, 增加算法的稳定性, 这样可以加速收敛和防止振荡。

附加动量项的实质^[61]是将最后一次权值变化的影响, 通过一个动量因子来传递。当动量因子取值为零时, 权值的变化则是根据梯度下降法产生; 当动量因子取值为 1 时, 新的权值变化则是设置为最后一次权值的变化, 而依梯度法产生的变化部分则被忽略掉。当增加了动量项后, 促使权值的调节向着误差曲面底部的平均方向变化, 当网络权值进入误差曲面底部的平坦区时, 梯度将变得很小, 从而防止了 $\Delta w(k+1) = 0$ 的出现, 有助于使网络从误差曲面的局部极小值中跳出来。

如果初始误差点的斜率下降方向与通向最小值的方向背道而驰, 则附加动量法失效。因此其缺点是它对训练的初始值有要求, 必须使其值在误差曲线上的位置所出误差下降方向与误差最小值的运动方向一致。

(4) 自适应调整学习率

通常调节学习速率的原则^[62]是: 检查权值的修正值是否真正降低了误差函数, 如果确实如此, 则说明选取的学习速率值小, 可以对其适当的增加一个量; 若不是则减小学习速率的值。通常 BP 算法在学习计算中, 学习率 η 和动量系数 α 是由经验选定, 对学习率来说, η 越大, 权变化越大, 收敛快, 但 η 越大又使误差增加, 从而使迭代次数加大, 并可能引起振荡。所以 η 应在不引起振荡的情况下, 尽可能取大的 η 。这就是说在训练开始, 可以给学习率 η 较大的值, 发现振荡时减少 η , 学习速率不应固定不变。1989 年和 1990 年 R.Salomon 用一种简单的进化策略来调节学习速率。

(5) 弹性反向传播算法

弹性反向传播算法是根据局部梯度信息实现权重的直接修改,对于每个权引入它的各自的更新值 $\Delta_j^{(t)}$,它独自确定权更新值的大小。它基于在误差函数 E 上的局部梯度信息,按照以下的学习规则更新:

$$\Delta_j^{(t)} = \begin{cases} \eta^+ \times \Delta_j^{(t-1)}, (\frac{\partial E^{(t-1)}}{\partial w_{ij}} \times \frac{\partial E^{(t)}}{\partial w_{ij}} > 0) \\ \eta^- \times \Delta_j^{(t-1)}, (\frac{\partial E^{(t-1)}}{\partial w_{ij}} \times \frac{\partial E^{(t)}}{\partial w_{ij}} < 0) \\ \Delta_j^{(t-1)}, (\text{其他}) \end{cases} \quad (4-12)$$

式(4-12)中 $\frac{\partial E^{(t)}}{\partial w_{ij}}$ 表示在模式集的所有模式(批学习)上求和的梯度信息, (t)

表示时刻 t 或第 t 次学习, E 表示平方和误差,其中 $0 < \eta^- < 1 < \eta^+$ 。

在每个时刻,如果目标函数的梯度改变它的符号,它表示最后的更新值太大,更新值 $\Delta_j^{(t)}$ 应由权更新值减小倍数因子 η^- 得到减少;如果目标函数的梯度保持它的符号,更新值应由权更新值增大倍数因子 η^+ 得到增大。为了减少自由地可调参数的数目,增大倍数因子 η^+ 和减小倍数因子 η^- 被设置到固定值: $\eta^+ = 1.2$, $\eta^- = 0.5$ 。

权更新遵循的原则:如果导数是正(增加误差),这个权由它的更新值减小;如果导数是负,更新值增加。

已知

$$\text{sign}(x) = \begin{cases} 1, (x > 0) \\ -1, (x < 0) \\ 0, (\text{其他}) \end{cases} \quad (4-13)$$

所以

$$\Delta W^{(t)} = \text{sign}\left(\frac{\partial E^{(t)}}{\partial W}\right) \times \Delta_j^{(t)} \times \frac{\partial E^{(t-1)}}{\partial W} \quad (4-14)$$

当学习开始时,所有的更新值 $\Delta_j^{(0)}$ 被设置为初始值 Δ_0 ,因为它直接确定了前面权步的大小,它应该按照权自身的初值进行选择。为了使权不至于变得太大,设置最大权更新值 Δ_{\max} ,默认上界设置为 $\Delta_{\max} = 50.0$ 。要说明的是,阈值的更新也遵循上述相同的规则。

目前国内外在优化BP算法方面作了大量的研究,比如Bo Yang等^[63]用一种改进的遗传算法来优化BP神经网络的初始权值,仿真结果表明:这种方法可以降低BP神经网络陷入局部极小的可能性,并且可以有效的改善收敛速度。Yang Huizhong等^[64]采用带有自适应动量项的变步长BP算法,可以降低BP神经网络陷入局部极小的

可能性, 并且可以有效的改善收敛速度。

4.3.3 基于优化原理的网络训练算法

节 4.3.2 中提到的第 (3)、(4) 方法实际上并没有跳出所谓 Heb 学习原则的圈子, 其每一步的搜索方向还是负梯度方向。然而神经网络的实质是一类优化问题, 因此节 4.2 中所描述的问题可用数值优化方法, 可如下描述:

考虑三层网络, 如图 4-1 所示, 它由输入层、隐层和输出层构成。输入层节点数 ni , 隐层节点数 nh , 输出层节点数 no , 样本数 np 。输入和期望输出矢量为 $\bar{x}_p, \bar{y}_p, p=1, 2, \dots, np$ 。其中 $\bar{x}_p = \{x_{pi}\}, i=1, 2, \dots, ni, \bar{y}_p = \{d_{pk}\}, k=1, 2, \dots, no$ 。 $w_{j,i}, \theta_j, w_{k,j}, \theta_k$ 分别为隐层和输出层的权值和阈值。将训练过程转化为在已建立的网络系统下的无约束的非线性最优化问题, 即

$$\min\{f(\bar{z})\} \quad \bar{z} \in R^n \quad (4-15)$$

问题的目标函数为

$$f(\bar{z}) = \frac{1}{2np} \sum_{p=1}^{np} \|\bar{y}_p - \bar{o}_p\|^2 = \frac{1}{2np} \sum_{p=1}^{np} \sum_{k=1}^{no} (y_{pk} - o_{pk})^2 \quad (4-16)$$

式 \bar{z} 为变量, 它表示为权值和阈值。变量的总维数为 $n = (ni + 1)nh + (nh + 1)no$, 权值和阈值的变量 \bar{z} 的表示关系满足

$$z(nx) : nx = (j-1)(ni+1) + i, i=1, 2, \dots, ni, j=1, 2, \dots, nh : \text{表示 } w_{j,i}$$

$$nx = j \cdot (ni+1), j=1, 2, \dots, nh : \text{表示 } \theta_j$$

$$\text{设 } nnc = (ni+1)nh$$

$$z(nx) : nx = nnc + (k-1)(nh+1) + j, j=1, 2, \dots, nh, k=1, 2, \dots, no : \text{表示 } w_{k,j}$$

$$nx = nnc + k(nh+1), k=1, 2, \dots, no : \text{表示 } \theta_k$$

则

$$\bar{z} = [z(1), z(2), \dots, z(n)]^T \quad (4-17)$$

负梯度为

$$\bar{s} = -\nabla f = \left[\frac{\partial f}{\partial z(1)}, \frac{\partial f}{\partial z(2)}, \dots, \frac{\partial f}{\partial z(n)} \right]^T \quad (4-18)$$

对于任意给定的 $\bar{z}^{(0)}$, 采用一种优化方法可以确定变量 \bar{z}^* , 使得获得极小值 $f(\bar{z}^*)$, 它可能是局部最小值或全局最小点。

本文采用的是神经网络共轭梯度法来优化训练 BP 网络。神经网络共轭梯度法^[65]是 1952 年由赫斯坦内斯和斯蒂费尔提出, 后于 1964 年由费彻和里夫斯进一步研究和完善 (简称 FR 法), 它是求解无约束最优化问题的一种有效的最优化方法。共轭梯度法既能在一定程度上克服 BP 算法收敛速度、局部最小问题和训练瘫痪等不

足, 程序简单容易实现, 特别适用于目标函数具有二次型性质的问题。共轭梯度法的实质, 是对最速下降法进行修正, 使搜索方向变成共轭方向 (将负梯度方向旋转一个角度), 即每一步的搜索方向都要对该步的负梯度方向进行修正。共轭梯度法的基本思想, 是把共轭性与最速下降法 (最优梯度法) 结合起来, 利用已知处的梯度构成一组共轭方向并沿这组方向进行搜索, 求出目标函数的极小值点。

对于 n 维非线性目标函数 $f(\bar{z})$ 搜索过程如下: 从随机初始点 $\bar{z}^{(0)}$ 开始第一次沿着负梯度方向 $\bar{s}^{(0)} = -\bar{g}^{(0)} = -\nabla f(\bar{z}^{(0)})$, 得到新起点 $\bar{z}^{(1)} = \bar{z}^{(0)} + \bar{\alpha}^{(0)}\bar{s}^{(0)}$, 计算新梯度点

$\bar{g}^{(1)} = \nabla f(\bar{z}^{(1)})$, 由梯度性质相邻梯度值正交即 $[\bar{g}^{(0)}]^T \bar{g}^{(1)} = 0$, 得到 $[\bar{s}^{(0)}]^T \bar{g}^{(1)} = 0$, 取

$$\bar{s}^{(1)} = -\bar{g}^{(1)} + \beta^{(0)}\bar{s}^{(0)}, \quad \text{由 } [\bar{s}^{(1)}]^T [\bar{g}^{(1)} - \bar{g}^{(0)}] = 0 \text{ 可得 } \beta^{(0)} = \frac{\|\bar{g}^{(1)}\|^2}{\|\bar{g}^{(0)}\|^2}, \quad \bar{s}^{(k)} \text{ 方向可由}$$

$$\bar{s}^{(k)} = -\bar{g}^{(k)} + \beta^{(k)}\bar{s}^{(k-1)}, \quad \text{其中 } \beta^{(k)} = \frac{\|\bar{g}^{(k+1)}\|^2}{\|\bar{g}^{(k)}\|^2}。 \text{ 经过 } N \text{ 次计算后, 要重新开始。结束}$$

条件为 $\|\nabla f(\bar{z}^{(k)})\| < \varepsilon$ 。该优化算法的流程图如图 4-3 所示。

文献[66]采用了两种算法事例计算得出了一个分析结果, 如图 4-4 所示, 我们可以看到两种算法对目标函数的计算次数关系。从图上我们可以看到共轭梯度法的计算次数远少于 BP 算法和梯度法。

4.4 实验研究分析

建立滚动轴承运行状态的 BP 神经网络分类器^[67], 采用传统的改进 BP 算法和本课题的新优化算法对相同结构的网络进行训练。其中通过对某轴承运行状态的振动信号进行处理, 提取特征参量组成的网络学习样本, 其部分样本数据如表 4-1 所示, 其中第二章提到的峰值因子、脉冲因子、裕度因子、波形因子和峭度因子等五个特征参数, 作为神经网络的输入。部分网络测试样本如表 4-2 所示。

输入层采用全通型, 隐含层采用 Sigmoid 函数为激活函数, 输出层采用线性函数, 性能函数即误差函数采用均方误差函数。网络训练参数: 学习率 0.07; 训练目标误差 0.01; 最大训练次数设置为 1000。经过反复实验对比后, BP 网络结构最终选定为 N (5-13-4)。用改进方法的实验输出结果见表 4-3。

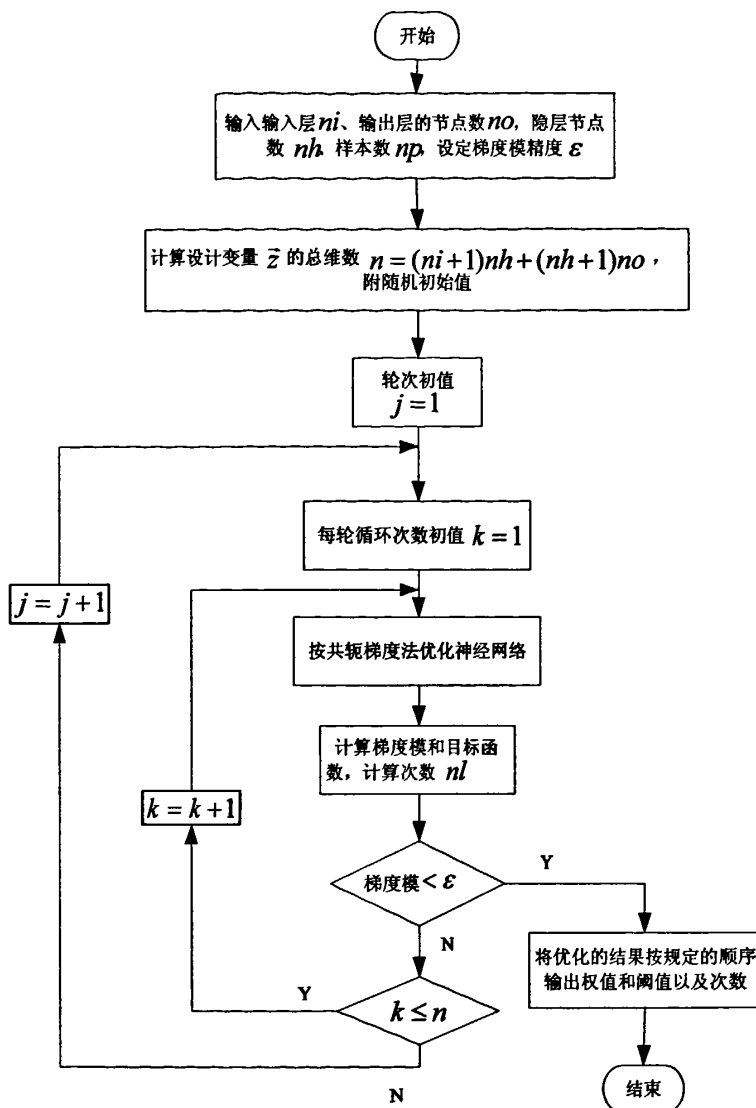


图 4-3 优化算法的流程图

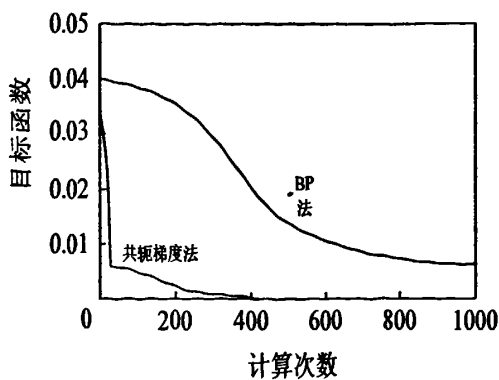


图 4.4 两种算法对目标函数的计算次数关系

表 4-1 部分网络学习样本

轴承的状态	峰值因子	脉冲因子	裕度因子	波形因子	峭度因子
正常	4.19	5.55	6.51	1.22	3.04
	4.20	5.67	6.49	1.23	3.08
	4.08	5.80	6.41	1.25	3.16
	4.32	5.64	6.81	1.29	3.20
外圈故障	6.58	7.67	9.48	1.36	3.64
	6.50	7.54	9.61	1.29	3.57
	6.45	7.54	9.61	1.29	3.52
	6.02	7.60	9.52	1.34	3.56
内圈故障	8.75	10.8	16.1	1.51	4.38
	5.68	10.6	16.2	1.45	4.41
	8.56	10.5	15.8	1.42	4.35
	8.61	11.0	15.2	1.44	4.38
滚动体故障	9.81	16.1	19.7	1.63	9.65
	9.58	15.4	20.1	1.59	9.38
	9.69	15.8	20.4	1.59	9.78
	9.75	15.8	20.0	1.58	9.89

表 4-2 部分网络测试样本

轴承的状态	峰值因子	脉冲因子	裕度因子	波形因子	峭度因子
正常	4.18	5.45	6.41	1.12	2.95
	4.22	5.67	6.51	1.22	3.06
外圈故障	6.24	7.63	9.58	1.39	3.66
	6.48	7.67	9.48	1.41	3.70
内圈故障	8.71	10.6	16.1	1.52	4.37
	8.65	10.7	15.8	1.46	4.38
滚动体故障	9.81	16.1	20.8	1.68	9.69
	9.79	15.7	20.4	1.69	9.66

通过分析比较表 4-3 中的网络的实际输出和网络的期望输出,可以发现它们比较接近,输出结果比较的理想,实验结果表明采用优化过的 BP 神经网络来识别滚动轴承的运行状态是可行有效的,并说明了利用滚动轴承振动信号提取的无量纲参数组成的特征向量可以表征滚动轴承的状态。

表 4-3 神经网络的实验输出

轴承的状态	实际输出	期望输出
正常	(0.0021, 0.0148, 0.0034, 0.0164)	(0, 0, 0, 0)
	(0.0093, 0.0044, 0.0594, 0.0321)	(0, 0, 0, 0)
外圈 故障	(0.0005, 0.9702, 0.0203, 0.0416)	(0, 1, 0, 0)
	(0.0013, 0.9851, 0.0068, 0.0088)	(0, 1, 0, 0)
内圈 故障	(0.0029, 0.0386, 0.9384, 0.0678)	(0, 0, 1, 0)
	(0.0012, 0.0005, 0.09124, 0.0925)	(0, 0, 1, 0)
滚动体 故障	(0.0062, 0.0031, 0.0633, 0.09985)	(0, 0, 0, 1)
	(0.0053, 0.0061, 0.0060, 0.9772)	(0, 0, 0, 1)

4.5 本章小结

本章简单介绍了人工神经网络的基本原理, 并对 BP 神经网络的工作原理进行了系统的阐述。重点讨论研究了 BP 算法的不足以及各种优化算法, 并通过实验分析研究了各种优化算法, 并从由实验结果得出本课题提出的优化算法用于神经网络中是可行有效地, 其优化效果比其他优化算法更优越。

第五章 基于小波包的滚动轴承优化智能故障诊断

5.1 实验数据来源

试验装置的合理设计和适宜的获取数据的方法,在试验装置运行时正确记录数据,保证数据能够真实准确地反映装置运行的状态,这有利于后面的分析和研究。试验系统的组成如图 5-1 所示,主要包括轴承、加速度传感器、数据采集系统和轴承诊断系统等四个方面。所使用的轴承包括了正常和有故障的轴承,试验系统运行时所获取的数据代表相应的轴承的状态;加速度传感器的输出与被测物体的振动的加速度成正比,与速度传感器或位移传感器相比,加速度传感器频率响应范围更宽,能够获得更多的频率成分信号,且其承受负载能力更强,机械性能更好;数据采集系统是将由加速度传感器测得的模拟信号进行 A/D 转换,然后存入计算机中,数据文件为 Matlab 文件格式。

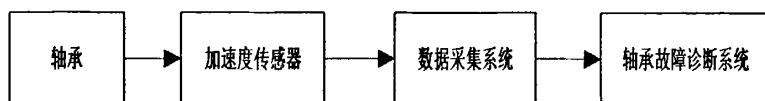


图 5-1 试验系统框图

机械故障振动诊断方法受到广泛地关注,开展了大量的研究工作,但多将注意力集中于新好处理方法的研究及其在此基础上的故障特征的提取、故障征兆判断等问题上,而忽略了由参数选择和多传感器位置配置所决定的信息源故障信息含量等问题。如果信息源不含有所需的最低限度的故障信息,则随后无论采用多么先进的处理方法,比如近代谱分析方法、模糊逻辑、时频变换、人工神经网络以及各种辨识和判断方法,也难于得到所期望的结果^[69]。

所以采用的实验数据来源为:采集带有内圈故障轴承(型号 6205-2RS JEM SKF)振动信号,试验装置和试验数据均来自美国 Case Western Reserve University 电气工程实验室^[51-53]。试验装置如图 5-2 所示,试验中通过对功率计的控制获得期望的扭矩载荷,振动加速传感器垂直固定在感应电机输出轴支撑轴承上方的机壳上进行数据采集。1.5kW 的三相感应电机通过自校准联轴节与一个功率计和一个扭矩传感器相连,最后驱动风机进行运转。

由于客观条件的限制,很难在短时间内收集到大量的故障轴承用于研究,故在研究滚动轴承故障的过程中采用故障模拟的试验研究方法。故障模拟是在一定条件

下人为地在轴承上制造某些故障来模拟现实中的一些故障形式，然后通过分析研究，从而判断和验证故障诊断方法的可行性。本文研究的轴承故障均由人为地划上电弧产生。

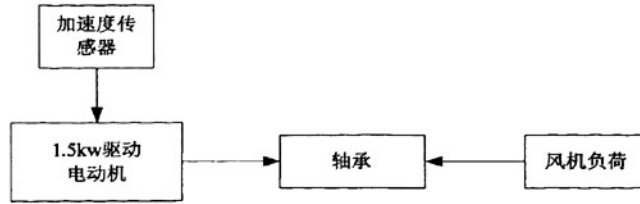


图 5-2 滚动轴承实验台示意图

另外传感器的安装位置十分重要，它直接关系到测采集数据的有效性，即系统采集的数据是否正确。本实验中传感器的位置被安装在电机驱动端的 12 点钟的位置。

5.2 基于小波包的滚动轴承特征提取

当用一个含有丰富频率成分的信号作为输入对系统进行激励时，由于系统故障对各频率成分的抑制和增强作用发生改变，通常它会明显地对某些频率成分起着抑制作用，而对另外一些频率成分起着增强作用。因此其输出与正常系统输出相比，相同频带内信号的能量会有着较大的差别，它使某些频带内信号能量减小，而使另外一些频带内信号能量增大。因此在各频率成分信号的能量中，包含着丰富的故障信号，某种或某几种频率成分能量的改变即代表了一种故障情况^[50]。

基于小波包变换得到的能量特征向量，可以作为故障特征向量进行故障诊断，同时，这种故障诊断方法实施故障特征提取和诊断时，无需被诊断系统的数学模型，就可以迅速地进行故障诊断检测，并能准确地进行故障定位^[70]。

由于小波包分解具有多分辨率的特点，利用它将轴承振动信号进行三层小波包，以获取轴承振动信号在不同频带内的能量，其分解结构如图 5-3 所示。

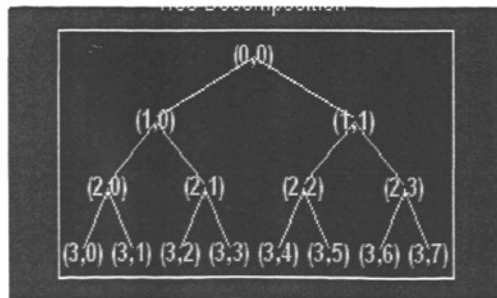


图 5-3 轴承振动信号小波包分解结构图

图 5-3 中(0,0)表示原始信号， $(i,0)$ 表示小波包分解的第 i 层的低频系数 X_{i0} ， (i,j)

表示小波包分解的第*i*层的高频系数 $X_{ij}(i=1,2,3;j=1,2,\dots,7)$ 。小波包分解提取滚动轴承特征的具体算法及步骤如下:

(1) 对滚动轴承的振动信号进行3层小波包分解。分别提取第3层从低频到高频8个频率成份的信号特征: 低频系数 X_{30} , 高频系数 $X_{31}, X_{32}, X_{33}, X_{34}, X_{35}, X_{36}, X_{37}$ 。

(2) 对小波包分解系数重构, 提取各频带范围的信号。设 S_{3j} 是 X_{3j} 的重构信号, 则总信号*S*表示为

$$S = S_{30} + S_{31} + S_{32} + S_{33} + S_{34} + S_{35} + S_{36} + S_{37} \quad (5-1)$$

将最高频率归一化为1, 则8个频率成分所代表的频率范围如表5-1所示

表 5-1 各频率成分所代表的频率范围

重构信号	S_{30}	S_{31}	S_{32}	S_{33}
频率范围	0~0.125	0.125~0.250	0.250~0.375	0.375~0.500
重构信号	S_{34}	S_{35}	S_{36}	S_{37}
频率范围	0.500~0.625	0.625~0.750	0.750~0.875	0.875~1.000

(3) 求各频带信号的总能量。设 $S_{3j}(j=0,1,\dots,7)$ 对应的能量为 $E_{3j}(j=0,1,\dots,7)$, 则有

$$E_{3j} = \int |S_{3j}(t)|^2 dt = \sum_k^n |x_{jk}|^2 \quad (5-2)$$

其中 $x_{jk}(j=0,1,\dots,7,k=1,2,\dots,n)$ 表示重构信号 S_{3j} 的离散点的幅值。

(4) 构造特征向量。由于系统出现故障时, 会对各频带内的信号的能量有较大的影响, 因此, 以能量为元素构造一个特征向量*T*:

$$T = [E_{30}, E_{31}, E_{32}, E_{33}, E_{34}, E_{35}, E_{36}, E_{37}] \quad (5-3)$$

当能量较大时, $E_{3j}(j=0,1,\dots,7)$ 通常是一个较大的数值, 给数据分析带来一些不方便, 为此, 可选归一化后的向量*T*为特征向量, 令

$$E = \sqrt{\sum_{j=0}^7 |E_{3j}|^2} \quad (5-4)$$

则

$$T' = [E_{30}/E, E_{31}/E, E_{32}/E, E_{33}/E, E_{34}/E, E_{35}/E, E_{36}/E, E_{37}/E] \quad (5-5)$$

小波包技术故障特征提取用于滚动轴承优化智能故障诊断的流程如图5-4所示。

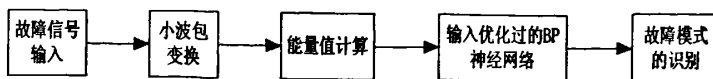


图 5-4 优化智能故障诊断流程图

5.3 诊断模型的建立及设计

BP 网络结构的设计包括输入层和输出层节点数目的确定、隐含层数目和隐含层节点数的确定等。

5.3.1 输入、输出层节点数的确定

输入与输出层的节点数数目是由问题的类别和应用要求来决定的：输入层的节点数目取决于数据源的维数，输出层的节点数取决于要解决问题的模式大小。

该研究的是滚动轴承的故障识别，为了使建立的神经网络模型能正确的映射出不同情况下的滚轴轴承的实际运行状态，采用小波包分解将输入信号分成 8 个频段提取轴承的特征信号，将其作为神经网络的输入，因此输入层节点数目为 8 个。

研究中考虑的模式类型包括正常和 3 种故障，为了便于直观的判断出故障的种类，及时方便地对结果进行处理，输出层的节点数的决定采用的方法是：对应多少种类故障就设定多少个输出。对于处理的故障种类不是很多的情况，这种方法直观、简单。因此输出层的节点数目为 4 个。

5.3.2 隐含层层数和隐含层节点数的选择

(1) 隐含层层数和隐含层节点数的选择

只有当隐含层层数和每个隐含层节点数给定时，神经网络系统才能变成第四章所描述的对于给定样本集合条件下确定权值和阈值的目标函数的优化问题。对于一个给定的样本集合，建立一个能描述其样本特征的多层神经网络。其网络的输入层节点数和输出层节点数是设定的，它反映结构的特性。那么隐含层层数和每个隐含层节点数取多少才算合适，是许多学者研究的问题。

隐含层节点数选择过少，网络获得信息过少，难以成功学习到事物的本质信息；隐含层神经元过多又容易使网络学习过饱和，最终整个网络容错性不好。但目前人工神经网络尚处在学科发展的不成熟期，其理论还没有完善到能提供一套可赖以遵循的设计方法。对于多层神经网络来说，首先要确定选用几层隐含层。Robert Hecht-Nielso 证明了当各节点具有不同的门限时，对于任何在闭区间内一个连续函数都可以用一个隐含层的网络来逼近，因而一个三层的基于 BP 算法的神经网络可以完成任意 n 维到 m 维的映射^[68]。此外有理论的分析也表明当各节点均采用 S 型函数时，一个隐含层就足以实现任意判决分类问题，两个隐含层则足以表示输入图形的任意输出函数。但经验表明对小型网络的边界判决问题，两层隐含层网络并不比单隐含层网络更优越，所以目前认为二进制分类或判决边界问题，一个隐含层就足够了。但是如果要求输出是输入函数的任意连续函数，那就要用两个隐含层或者采用不同的激活函数。

尽管理论分析证明,单隐含层网络能映射一切连续函数,但对于实际问题隐含层数的选择主要还是取决于问题的性质及复杂程度。研究表明,本课题用一个隐含层即能满足要求。此外对于多层前馈网络而言,隐含层神经元数目的确定也是一个重点。当前的开发设计还侧重于实验、试探多种模型方案,在实验中改进,直到选取一个满意的方案为止。同样当前隐含层神经元的选择也没有成熟的理论作为指导。目前常用的隐含层选择方法有如下几种:

$$a) \quad nh = \log 2^m \quad (5-6)$$

$$b) \quad nh = \text{sqt}(ni * no) \quad (5-7)$$

$$c) \quad nh = \text{sqt}(ni + no) + a \quad (5-8)$$

$$d) \quad nh = \sqrt{ni + no} + a \quad (5-9)$$

其中 ni 为输入层节点数, nh 为隐层节点数, no 为输出层节点数 no , 常数 $a=1\sim 10$ 。还有的文献介绍了通过对选择不同节点的网络进行蓄念,直到全值不再发生变化使网络稳定为止,然后依据测试误差的最小,来确定网络的隐层节点数。但该方法比较的费时间。根据上述的几种办法,本题采用式(5-4)方法来选择隐含层节点数数目。

(2) 初始权值的选择

由于网络系统是非线性的,初始权值强烈影响最终的解,若初始权值过高,可能会使部分乃至全部神经元的净注入较大,从而运行在 S 型转移函数斜率平缓的饱和区,致使权值调整量很小,网络出与麻痹状态,直接影响网络选练时间。因此一般希望经过初始权值后的每个神经元的输出值都接近于零,这样可以保证每个神经元的权值都能够在它们的 S 型激活函数变化最大之处进行调节。网络的初始值通常是 (-1, 1) 之间的随机数。为了防止上述现象的发生,威得罗等人提出了一种选定初始权值的策略。

5.4 仿真实验研究及其结果分析

由上述可知,采用 3 层 BP 神经网络,输入层为 8 个节点,隐层节点数的选取可根据经验或通过多次训练学习后,考虑网络的学习次数和识别率总和比较后选定。根据式(5-9)选择隐层节点数,即为

$$nh = \sqrt{ni + no} + a \quad (5-10)$$

nh 为隐节点数, ni 为输入层节点数, no 为输出节点数, 常数 $a=1\sim 10$ 。

根据轴承运行状态进行编码,即输出层为 4 个节点,对应滚动轴承 4 种状态,其期望输出为:

正常	(0, 0, 0, 0)
外圈故障	(0, 1, 0, 0)
内圈故障	(0, 0, 1, 0)

滚动体故障 (0, 0, 0, 1)

综上所述可得, BP 神经网络的结构为 $8 \times 13 \times 4$ 的三层网络结构, 具体结构如图 5-5 所示。

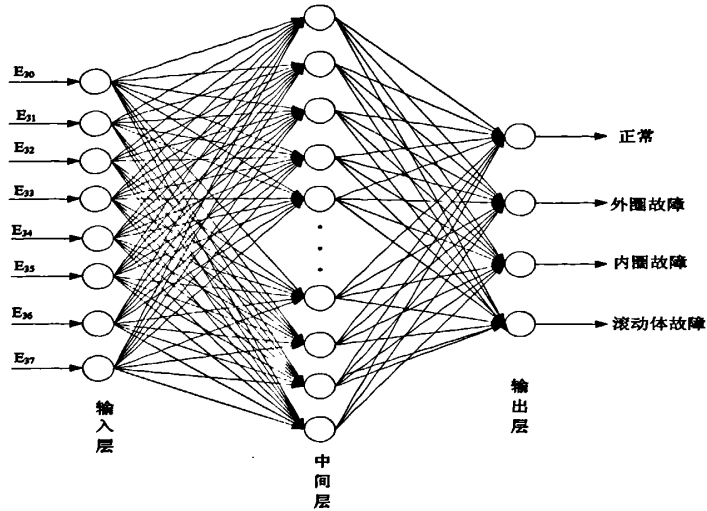


图 5-5 BP 神经网络结构图

表 5-2 和表 5-3 分别给出了部分训练样本数据和部分测试样本数据, 以检验各种不同算法相同结构的网络性能。

表 5-2 标准部分训练样本

故障类型	E ₃₀	E ₃₁	E ₃₂	E ₃₃	E ₃₄	E ₃₅	E ₃₆	E ₃₇
正常	-0.8736	-0.8952	1.0000	-0.3840	0.6034	0.0863	-0.1051	-1.0000
	-0.3365	-1.0000	0.1936	-0.7691	1.0000	-1.0000	0.3845	-0.7045
	-1.0000	0.0181	0.3709	-0.9223	-0.0771	0.6049	0.3967	-0.5998
外圈故障	0.0975	-0.3959	-0.9342	-0.4205	-0.0166	0.3271	-0.1199	0.3020
	-0.9687	0.8186	-0.3704	-0.4768	-0.1870	0.0132	-0.1699	0.4397
	-0.2648	0.0871	-0.0992	-0.3758	-0.9204	1.0000	0.0506	-0.1356
内圈故障	0.1782	0.8044	-0.4259	-1.0000	-0.7729	0.0998	-0.6963	1.0000
	0.2862	0.3316	-0.9154	-0.5633	-1.0000	-0.4889	1.0000	0.0564
	0.0049	-0.4019	-0.3936	-0.0571	-0.1050	-0.4717	-0.1174	0.4151
滚动体故障	-0.5109	0.4827	-0.4416	-0.0169	-0.7799	-0.2485	0.2991	0.3937
	-0.1004	-0.4947	-0.4345	0.5462	-0.5795	-0.1216	-0.3861	0.7902
	0.2491	0.9838	-0.6065	-0.6623	-0.5646	0.3597	-0.4360	-0.0630

表 5-3 部分测试样本数据

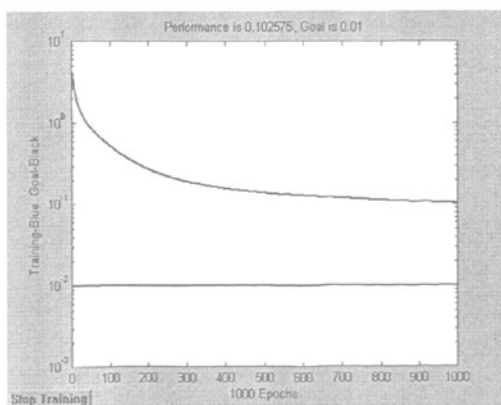
故障类型	E ₃₀	E ₃₁	E ₃₂	E ₃₃	E ₃₄	E ₃₅	E ₃₆	E ₃₇
正常	-0.4021	-0.3702	-0.3729	1.0000	0.3815	0.7108	-1.0000	-0.8224
外圈故障	1.0000	-0.6780	-0.7309	-0.5284	0.0434	-0.6012	0.2129	-0.3122
内圈故障	-0.3456	-0.2769	-1.0000	-0.6142	0.4132	-0.2647	0.9454	-0.7009
滚动体故障	-0.3797	1.0000	-0.2556	-0.3624	0.0695	0.1710	-0.5445	-0.5678

在离线情况下每类滚动轴承各取得样本，利用小波包分解提取各个样本的能量特征向量，并分别用传统的改进 BP 算法和共轭梯度法对神经网络进行学习；然后对滚动轴承的检验样本数据进行了故障诊断，结果表明新的方法比传统的 BP 算法诊断正确率高，诊断速度快，能够准确地实现轴承的故障诊断。

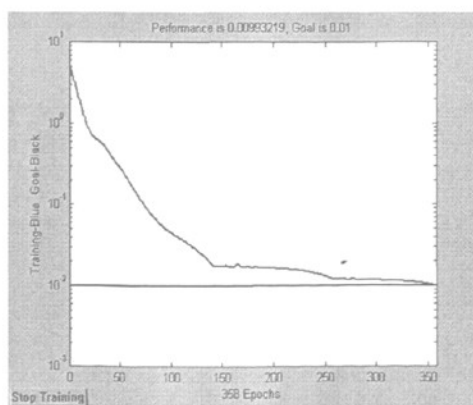
为了分析新方法的优越性和有效性，在相同的网络结构下，对同一组非训练样本数据分别用不同的算法和本课题采用的改进方法用于轴承的故障诊断中。在选择不同算法时，网络使用相同的传递函数，为对数 Sigmoid 传递函数，即

$$f(x) = \frac{1}{1 + \exp(-x)} \quad (5-11)$$

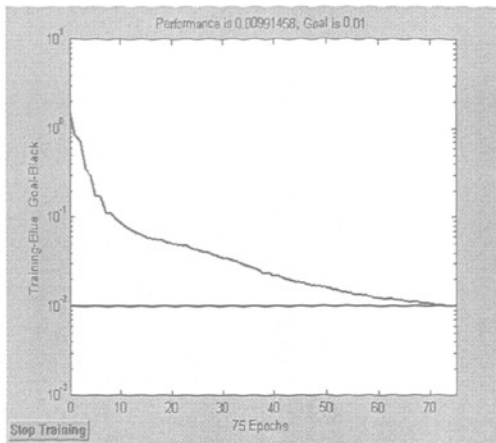
性能函数即误差函数采用均方误差函数。各种算法的共同网络训练参数：学习率 0.07；训练目标误差 0.001；最大训练次数设置为 1000。每种故障模式输入 20 组，前 10 组数据作为学习样本数据，后 10 组数据作为检验样本数据，以检验经过训练后 BP 网络对故障信号的识别能力。经过反复实验对比后，BP 网络结构最终选定为 N(8-13-4)。各种算法训练网络的误差性能曲线如图 5-6 所示。用本课题采用的改进方法的实验结果见表 5-4。



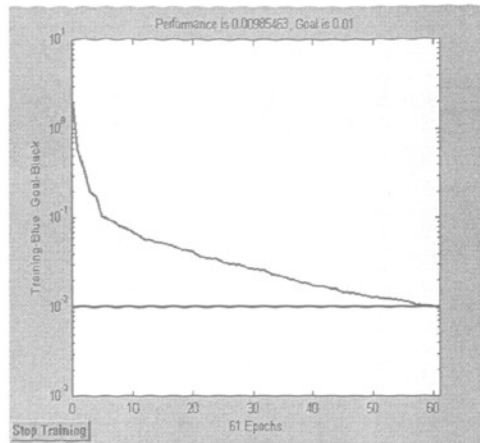
(a) 采用传统的 BP 算法



(b) 采用自适应学习率算法



(c) 采用有弹回的 BP 算法

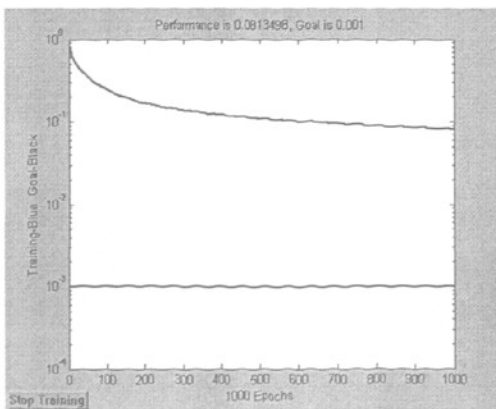


(d) 采用共轭梯度法

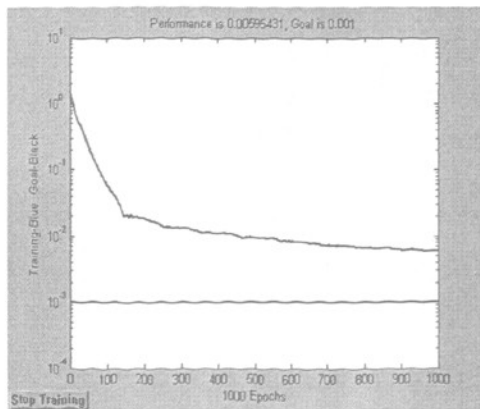
图 5-6 各种改进算法的误差曲线和训练次数

下面分析当模式识别的精度要求提高时，采用不同的算法对网络进行训练，对故障类别进行识别的结果又是怎样？当网络参数改为：训练目标误差 0.001；最大训练次数设置为 1000 时，各种不同算法的误差曲线和训练次数如图 5-7 所示。

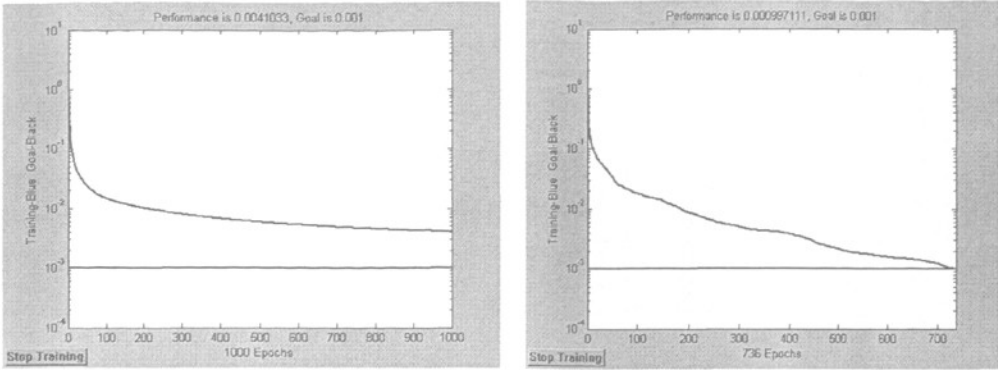
通过分析比较图 5-6 和图 5-7 可知，当网络目标误差为 0.01 和网络目标误差为 0.001 时，在限定最大训练次数（1000）内，普通 BP 网络都未能达到训练目标；相对普通 BP 网络来说，自适应修改学习率算法和有弹回 BP 算法在网络目标误差为 0.01 时都能用较少的训练次数达到了所要求的训练目标，两者方法明显地改善了 BP 神经网络的训练结果。需要说明一点的是采用有弹回 BP 算法，训练次数明显少于自适应修改学习速率算法，改进的效果优越于自适应修改学习速率。但当网络目标误差改为 0.001 时这两种算法都未能在限定最大训练次数（1000）内达到训练目标。反之无论当网络目标误差为 0.01 或网络目标误差为 0.001 时，在限定最大训练次数（1000）内，共轭梯度算法都能达到训练目标。结果表明共轭梯度算法在训练网络时收敛速度很快，优化网络的效果更优越于自适应修改学习率算法和有弹回 BP 算法。



(a) 采用传统的 BP 算法



(b) 采用自适应学习率算法



(c) 采用有弹回的 BP 算法

(d) 采用共轭梯度法

图 5-7 各种改进算法的误差曲线和训练次数

表 5-4 实验结果数据

	正常	外圈故障	内圈故障	滚动体故障
实验输出	(0.0109, 0.0228, -0.0166, 0.0078)	(0.0193, 0.9604, 0.0073, 0.0018)	(-0.0123, -0.0186, 1.0113, 0.0247)	(0.0600, 0.0807, 0.0468, 1.0042)
期望输出	(0, 0, 0, 0)	(0, 1, 0, 0)	(0, 0, 1, 0)	(0, 0, 0, 1)

根据实验结果表 5-4，可分析得出以下结论：

(1) 误差在期望的范围之内，实际输出值与期望输出值基本吻合。实验结果表明，该方法用于滚动轴承的故障诊断中具有一定的优越性和有效性。

(2) 根据小波包分解的频率划分性质，准确地提取了滚动轴承故障的特征，可保证不丢失故障的特征信息；然后将提取滚动轴承的故障特征向量作为神经网络的输入，减小了神经网络的规模与计算量，提高了诊断的准确性。

(3) 只要选择足够多的实际故障样本训练 BP 神经网络，网络则具有较好的容错性和稳定；另外通过理论分析和实验数据结果表明将神经网络用于滚动轴承的离线故障诊断是可行有效的。

5.5 本章小结

本章首先简单介绍了 BP 神经网络的设计以及实验数据的来源，然后结合小波变换和改进的 BP 神经网络提出了一种新改进的滚动轴承故障诊断方法。首先利用小波变换对滚动轴承的动态信号进行精细的划分，提取滚动轴承动态信号的故障特征信息；然后利用改进的 BP 神经网络进行滚动轴承故障分类与诊断。这种方法既克服了传统 BP 的收敛速度、局部最小问题和训练瘫痪等缺点，极大地提高了诊断精度，实验仿真结果表明该方法的优越性和有效性。

第六章 结论与展望

6.1 结论

本文主要研究设计了用于滚动轴承故障诊断的智能系统。在参考国内外有关轴承故障诊断技术的相关文献资料的基础上,对如何正确及时地判别轴承的运行状态等问题进行了较深入的系统研究,取得的主要研究成果如下:

(1) 对小波变换的去噪理论和奇异性理论在滚动轴承状态监测与故障诊断中的应用进行了研究。

实验研究表明:小波变换具有多分辨分析的特点,通过每层高频分解的系数进行阈值处理后再重构,可有效地达到消除噪声的目的;小波奇异性理论克服了其他方法对滚动体故障不敏感的缺陷,可有效地用于滚动轴承的状态检测和故障诊断。

(2) 重点讨论研究了 BP 算法的不足以及各种优化算法,并通过实验对比分析研究了各种优化算法的优缺点,并从中得出本课题提出的优化算法用于神经网络中是可行并有效的,其优化效果比其他优化算法更优越。

(3) 详细系统地分析了滚动轴承的振动机理,故障类型以及特性,采用对滚动轴承的振动信号进行时域和时频域提取滚动轴承正常、内圈故障、外圈故障和滚动体故障的特征,综合信号处理、故障诊断以及神经网络等多方面的知识,以仿真实验为基础,提出了一种基于改进 BP 神经网络的滚动轴承智能故障诊断技术。

其基本思想是:采用滚动轴承的实测数据,用小波包理论提取滚动轴承的振动特征,作为 BP 神经网络的输入,从而建立滚动轴承智能诊断系统模型。实验研究结果表明,本文设计的轴承智能诊断系统能正确判断滚动轴承的运行状态,达到及时故障识别及判断的目的,取得了比较满意的效果。

6.2 后续工作及展望

由于受条件、时间和精力限制,本次课题所研究的故障智能诊断系统以及对其中神经网络优化改进技术都远没有达到完美无缺的地步,欢迎老师和同学指出其中的不足和提出有建设性的建议。下一步的研究工作着重在以下方面改进:

(1) 在小波分析中,最优小波包基的选择对信号处理乃至故障诊断都有着莫大的关系,因此今后将进一步研究如何自适应地选择最优小波包基。

(2) 本文只对轴承正常、内圈故障、外圈故障、滚动体故障等四种运行状态下的振动信号进行了诊断,提出的改进故障智能诊断用于多种故障并发时的诊断识别

尚需要进一步的研究。

(3) 仿真和实验验证表明故障诊断正确率还没有达到 100%，如何优化神经网络的学习样本设计还有待于进一步研究。

参考文献

- [1] 张润林. 旋转机械故障机理与诊断技术[M]. 北京: 机械工业出版社, 2002
- [2] 周东华, 王桂增. 故障诊断技术: 综述[J]. 化工自动化与仪表, 1998, 25(1):58~62
- [3] 梅宏斌. 滚动轴承振动监测与诊断理论·方法·系统[M]. 北京: 机械工业出版社, 1995
- [4] 陈向东, 赵登峰, 王国强, 许纯新. 基于神经网络的滚动轴承故障监测[J]. 轴承, 2003(2):23~26
- [5] James E. Berry. How to Track Rolling Element Bearing Health with Vibration Signature Analysis [J]. Sound and Vibration, 1991:24~35
- [6] 王志刚, 李友荣, 朱瑞荪, 吕勇. 一种用于低速重载轴承故障诊断的共振解调法[J]. 煤矿机械, 2002, (8):74~76
- [7] K.F. Martin, P. Thorpe: Normalised spectra in monitor of rolling bearing elements[J]. Wear, Vol. 210:277~285
- [8] 武和雷, 朱善安, 林瑞仲, 郑剑锋. 基于能量算子解调法的滚动轴承故障诊断[J]. 农业机械学报, 2003, 34(1):118~120
- [9] W.Y. Wang, M.J. Harrap. Condition monitoring of rolling element bearings by using cone kernel time-frequency distribution. In Proc. SPIE Conf. On Measure, Technol, and Intellinstrum., 1993, 2101(1):290~29
- [10] 陈刚, 廖明夫. 基于小波分析的滚动轴承故障诊断研究[J]. 科学技术与工程. 2007, 7(12):2810~2814
- [11] Fu, Z.; Brown, D.J; Haynes, B.P. A new method of non-stationary signal analysis for control motor bearing fault diagnosis [J]. Intelligent Signal Processing, 2003, IEEE International Symposium on 4-6 Sept. 2003:99~104
- [12] Li.B.; Chow, M.Y.; Tipsuwan, Y.; Hung, J.C. Neural-network-based motor rolling bearing fault diagnosis[C]. Industrial Electronics, IEEE Transactions on Volume 47, Issue5, Oct. 2000: 1060~1069
- [13] 仇学青, 张鑫. 滚动轴承故障诊断研究的国内现状和发展方向[J]. 煤矿机械, 2007, 28(6): 6~8
- [14] Jan W. Kantelhardt, H.E Duardo, Roman, Martin Greiner. Discrete Wavelet Approach to Multifractality. Physica, A, 1995,(220):219~238
- [15] R R Coifman, Y Meyer, M.V Wickerhauser. Wavelet Analysis and Signal Processing. In Wavelets and Their Applications, Boston: Jones and Barlett, 1992:153~178.
- [16] 关毅林, 崔岩. 列车滚动轴承故障特征获取和诊断[J]. 现代机械, 2003, (5): 69~72

- [17] 何正嘉等.小波技术在机械监测诊断领域的应用现状与进展[J].西安交通大学学报, 2001, (5): 540~547
- [18] 张永学.提升小波变换在信号去噪中的应用[J].石油仪器.2007, 21(1):80~82.
- [19] 陆爽, 田野.滚动轴承微弱故障信号检测的研究[J].组合机床与自动化加工技术, 2004, (8):3~5.
- [20] 郭计云, 王福明.基于小波变换的信号去噪方法研究[J].科技情报开发与经济, 2007, 17(6): 202~203
- [21] 刘兴旺, 李训铭等.小波变换在脉象信号消噪处理中的应用[J].计算机与现代化, 2007(1): 16~18
- [22] Lindh T, Ahola J, Kamarainen J K, et al. Bearing damage detection based on statistical discrimination of stator current. Symposium on Diagnostics for Electric Machines, Power Electronics and Drives, Atlanta (USA), 2003, 1:177~181
- [23] Abbaszadeh K, Milimonfared J, Haji M, Toliyat H A. Broken bar detection in induction motor via wavelet transformation. Proceedings of the 27th Annual Conference of the IEEE Industrial Electronics Society, Denver (USA), 2001, 1:95~99
- [24] Eren L, Devaney M J. Bearing Damage Detection Via Wavelet Packet Decomposition of the Technology Conference. Stator Current. IEEE Instrumentation and Measurement. Anchorage, AK, USA, 2002
- [25] Shinji Ustsumi, Zen-ichiro Kawasaki, Kenji Matsu-ura, Masatake Kawada. Use of wavelet transform and fuzzy system theory to distinguish wear Particles in lubricating oil for bearing diagnosis[J]. Electrical engineering in Japan, 2001, 134(1):36~44
- [26] Yuan H Y, Chen G J. A Method of Fault diagnosis in nonlinear analog circuit based on neural network[C]. Proceeding of the 7of International Conference on Electornic Measurement Instruments, 2005, 8: 68~70
- [27] Zhi-Yong Luo, Zhong-Ke Shi. Wavelet Neural Network Method for Fault Diagnosis of Push-pull Circuits[J]. Proceedings of the 4th International Conference on Machine Learning and Cybernetics, 2005: 3327~3332
- [28] 孙璐, 唐广群, 张德政.神经网络技术在破碎机故障诊断中的应用[J].矿冶, 2007, 16(3): 85~90
- [29] Peng Lu, Da-Ping Xu, Yi-Bin Liu. Method of Fault Diagnosis on Multilayer BP Wavelet Networks and Its Applications [J]. Proceedings of the 3rd International Conference on Machine Learning and Cybernetics, 2004: 6296~6299
- [30] 胡婧, 杨曙年.基于 BP 神经网络的滚动轴承故障诊断研究[J].机械与电子, 2006(4): 9~11

- [31] Li Er guo, Yu Jin shou. In Input-Training Neural Network-Nonlinear Principal Componet Analysis Approach for Fault Diagnosis [J]. Proceedings of the 4th world Congress on Intelligent Control and Automation, 2004, 4 (4): 2733~2759
- [32] Yaagoub N, Al-Nassar, Mohsin Siddiqui, Ahmed Z, Al-Garni. Artificial neural networks in vibration control of rotor-bearing systems [J]. Simulation Practice and Theory, 2001, 7(8): 729~740
- [33] 王太勇, 商同等. 一种改进 BP 算法在滚动轴承故障诊断中的应用 [J]. 中国机械工程, 2001, 12(10): 1179~1181
- [34] Eren L, Devaney M J. Bearing Damage Detection Via Wavelet Packet Decomposition of the Technology Conference. Stator Current. IEEE Instrumentation and Measurement Anchorage, AK, USA, 2002
- [35] 璞良贵, 纪名刚. 机械设计 [M]. 北京: 高等教育出版社, 2000
- [36] 李辉. 滚动轴承和齿轮振动信号分析与故障诊断方法 [D]. 西北工业大学硕士学位论文, 2001: 13~17
- [37] 陈鹏. 基于神经网络的滚动轴承故障诊断方法研究与应用 [D]. 四川大学硕士学位论文, 2006: 19~20
- [38] 陈凯. 异步电动机轴承故障检测方法的研究 [D]. 华北电力大学硕士学位论文, 2006: 14~15
- [39] P.D.MeFadden&M.M.TOOZHY, Application of Synchronous Averaging to Vibration Monitoring of Rolling Bearings. Mechanical Systems and Signal Proecessing, 2000, 14(6): 891~906.
- [40] 王卓, 田振华, 赵丁选. 滚动轴承的振动监测与故障诊断系统 [J]. 轴承, 2002, 3: 22~27
- [41] Albert Boggess, Francis J. Narcowich.. A First Course in Wavelets with Fourier Analysis [M]. 北京: 电子工业出版社, 2002
- [42] C. Karanikas. The Hausdorff Dimension of Verv Weak Self-similar Fractals Described by the Haar Wavelet System. Chaos, Solitons and Fractals, 2000, (11): 275~280
- [43] R Zbigniew, Struzik. Wavelet Methods in (Financial) Time-series Processing. Physica A, 2001, (296): 307~319
- [44] 杨建国. 小波分析及其工程应用 [M]. 北京: 机械工业出版社, 2005
- [45] Jaideva C. Goswami Andrew K. Chan 著, 许天周等译. 小波分析: 理论、算法及其应用 [M]. 北京: 国防工业出版社, 2007
- [46] 陈刚, 廖明夫. 基于小波分析的滚动轴承故障诊断研究 [J]. 科学技术与工程, 2007, 7(12): 2810~2814

- [47] 孙建奎.小波分析及其应用[M].北京:机械工业出版社,2005
- [48] Mallat S. Theory for multi-resolution signal decomposition: The wavelet representation [J]. IEEE Transactions Pattern Analysis and Machine Intelligence, 1989, 11(7):674~693.
- [49] 张永学.提升小波变换在信号去噪中的应用[J].石油仪器.2007, 21(1):80~82.
- [50] 胡昌华,张军波,夏军,张伟.基于 MATLAB 的系统分析和设计—小波分析[M].西安电子科技大学出版社,1999:217~225
- [51] 曾庆虎,邱静,刘冠军,谭晓栋.小波相关特征尺度熵在滚动轴承故障诊断中的应用[J].国防科技大学学报,2007,29(6):102~105.
- [52] GG Yen and Kuo-Chung Lin. Wavelet Packet Feature Extraction for Vibration Monitoring. IEEE Transactions on Industrial Electronics, 2000, 650~666.
- [53] <http://www.eecs.case.edu/laboratory/bearing>
- [54] 唐英,孙巧.滚动轴承振动信号的小波奇异性故障检测研究[J].振动工程学报,2002,15(1):111~113
- [55] 蔡立良,周春华,杨小强,史长根.基于小波变换的信号奇异性检测在滚动轴承故障诊断中的应用研究[J].机床与液压,2006,(7):242~243
- [56] 陈恩利,于雪梅,申永军.奇异性检测在滚动轴承外圈损伤故障诊断中的应用[J].轴承,2006,(5):28~30
- [57] 刘乐平,林凤涛.基于小波包特征向量与神经网络的滚动轴承故障诊断[J].轴承,2008,(4):46~48
- [58] Simon Haykin. Neural Networks: A Comprehensive Foundation [M]. New Jersey, USA: Prentice Hall PTR, 1999.
- [59] Frederic M.Ham, Ivica Kostanic. Principles of neurocomputing for science& engineering [M].Beijing: China Machine Press, 2003
- [60] Martin T.Hagan, Howard B. Demuth, Mark Beale. Neural Network Design [M].Beijing: China Machine Press, 2002
- [61] 从爽,神经网络、模糊系统及其在运动控制中的应用[M].合肥:中国科学技术大学出版社,2001:19~23
- [62] Rumelhart D., McClelland J. Parallel Distributed Processing, Explorations in the Microstructure of Cognition. Cambridge: Bradford Books, MIT Press. pp. Vol-Vo2. 1986
- [63] Bo Yang, Xiao-Hong Su, Ya-Dong Wang; Machine Learning and Cybernetics,2002. Proceedings. 2002 International Conference on, Volume:1, 4-5, ov. 2002, Pages: 64~68 vol. 1
- [64] Yang Hui zhong, Li Dan j ing, Tao Zhenl in, Zhang Suzhen.Variable-step BP training

- algorithm with an adaptive momentum term; Intelligent Control and Automation, 2002. Proceedings of the 4th World Congress on, Volume:3, 10-14 June 2002, Pages:1961~1965 vol.3
- [65] Chang Wing-Fai; Mak Man-Wai; A conjugate gradient learning algorithm for recurrent neural networks; Neurocomputing Volume:24, Issue: 1-3, February, 1999, pp. 173~189
- [66] 侯祥林, 虞刚, 虞和济. 神经网络共轭梯度优化算法在激光加工定位中的应用[J]. 机械工程学报, 2002, 38(11):143~146.
- [67] 周开利、康耀红. 神经网络模型及其 MATLAB 仿真程序设计[M]. 北京: 清华大学出版社, 2005:70~85
- [68] 阎平凡. 人工神经网络与模拟进化计算[M]. 北京: 清华大学出版社, 2005
- [69] Yaagoub N, Al-Nassar, Mohsin Siddigui, Ahmed Z, Al-Garni. Artificial neural networks in vibration control of rotor-bearing systems[J]. Simulation Practice and Theory, 2001, 7(8):729~740
- [70] GG Yen and Kuo-Chung Lin. Wavelet Packet Feature Extraction for Vibration Monitoring. IEEE Transactions on Industrial Electronics, 2000:650~666.

致 谢

在论文完成之际，谨向我的导师申群太教授致以最衷心的感谢。感谢他多年来给予我学习和生活上的关怀和帮助，不断地给我指明方向，让我能抓住课题的关键，少走弯路顺利地完成毕业设计的工作。导师渊博的学识、敏锐的学术洞察力、严谨的治学态度、忘我的工作精神、朴实谦和的作风和孜孜不倦的教诲时刻深深地感染了我，将使我终生受益匪浅。

在此感谢刘鹏飞师兄、赵莉师姐、陈秋凤、全晓明、胡凯、谭境、罗锋、陈颀、何君霄等，没有他们的悉心指导和督促，研究的进度不会这么快。另外三年来实验室师兄姐妹给予了无私的帮助和大力地支持，在此表示我最衷心的感谢。

感谢我的同学们，是他们给了我各方面的帮助和关心，纯真的同学之情值得我永远珍惜和留念。

最后，深深地感谢我的父母和亲人，是他们无私的关怀、鼓励和支持，促使我奋发进取，持之以恒，并顺利完成学业，谨以此文作为献给他们最好的礼物。

谨此再次衷心的感谢所有关心和帮助我的人。

作者：黄一样
2009年3月
于中南大学

攻读硕士学位期间的主要成果

发表的学术论文:

[1] 黄一样, 申群太.小波消噪在滚动轴承故障诊断的应用研究.仪器与仪表用户, 已录用.