

CENTURY
21

高等学校教材

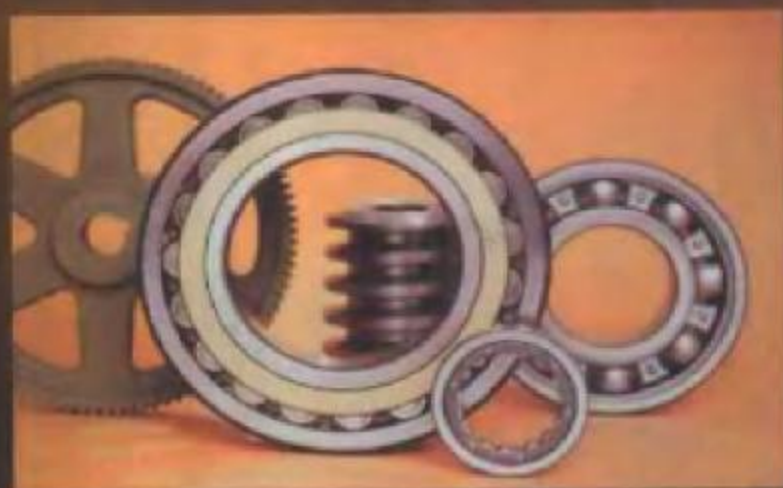
Textbook for Higher Education

机械设备故障 诊断技术及应用

王江萍 主编



JIXIESHEBEIGUZHANGZHEN DUANJISHUJIYINGYONG



□责任编辑 / 傅高明

□封面设计 / 小 果

ISBN 7-5612-1076-0



9 787561 210765 >

ISBN 7-5612-1076-0/TH · 63

定价：20.00元

344

978-7-100-14111-1

1

机械设备故障诊断技术及应用

主编 王江萍



A0973954

西北工业大学出版社

【内容提要】 本书全面系统地论述了机械设备故障诊断原理和技术。全书分两大部分共 9 章, 第一部分 1~4 章, 叙述了机械设备故障诊断技术的基本概念、机械物理信号分析基础、建模方法及时序模型在机械设备故障诊断中的应用、设备状态识别方法等诊断原理和常用技术; 第二部分 5~9 章以应用技术为主, 主要介绍了旋转机械、往复机械、主要传动零部件等的故障诊断技术, 以及无损检测技术等。

本书可作为高等院校机械工程及相关专业的本科、研究生教学用书, 也可作为机械设计、制造及维护管理方面工程技术人员的参考书。

图书在版编目(CIP)数据

机械设备故障诊断技术及应用/王江萍主编. —西安:
西北工业大学出版社, 2001. 8
ISBN 7-5612-1076-0

I. 机... II. 王... III. 机械设备-故障诊断 IV. TH17

中国版本图书馆 CIP 数据核字 (2001) 第 036779 号

出版发行: 西北工业大学出版社

通信地址: 西安市友谊西路 127 号 邮编: 710072 电话: 029-8493844

网 址: <http://www.nwpup.com>

E-mail: fxb@nwpup.com

印刷者: 西安电子科技大学印刷厂

开 本: 787 mm×1 092 mm 1/16

印 张: 15.5

字 数: 373 千字

版 次: 2001 年 8 月第 1 版 2001 年 8 月第 1 次印刷

印 数: 01~3 000 册

定 价: 20.00 元

前 言

机械设备故障诊断技术是 20 世纪 70~80 年代迅速发展起来的一门多学科交叉和融合的新型技术,现代微电子技术、检测技术、信息技术、计算机技术等为这一新型技术提供了先进的、多样化的技术手段,使之不断地完善和充实。故障诊断技术为保障设备安全运行、对设备故障的发展作出早期预报、对出现故障的原因作出判断并提出对策建议、为实现从现行的“定期维修”向更合理的“预测维修”的转变提供了条件。

近年来国内陆续出版了一批设备故障诊断方面的专著和译著,多是以故障诊断丛书的形式,分列专题出版的。这些著作及译著所论述的角度各有侧重,内容都比较丰富,并各有自己的特色。

本书的内容比较完整,既阐明故障诊断的原理和方法,又介绍故障诊断系统的实际组成和各种诊断方法的具体实现及其应用,反映故障诊断技术的最新发展,如模糊集理论、人工神经网络、小波变换等在故障诊断中的应用在书中也作了较为详细的讨论。本书所提供的数据和图表,多是作者在实验室及工程实践中的研究结果,而有些采用参考文献中的结果,主要是为了说明方法原理和基本规律。在进行实际诊断中,需根据具体的诊断对象和要求,采用对应的检测数据。

本书的第 1~6 章由王江萍编写,第 7~9 章由徐爱荣编写。全书由王江萍负责统稿,彭勇教授对全书进行了审阅,并提出了宝贵的意见。

本书的编写得到了西安石油学院的立项资助,在编写过程中得到了学院教务处和机械工程系有关老师的支持和帮助,在此致以诚恳的感谢。

由于编者的水平有限,加之机械故障诊断是一项新发展的科学技术,还没有形成较为完整的体系,许多原理方法尚在研究之中,因此,书中难免有错误和不妥之处,恳请读者批评指正。

编 者

2000 年 11 月

目 录

第 1 章 绪论	1
1.1 机械设备故障诊断技术的意义、目的和内容.....	1
1.2 设备故障的信息获取和检测方法	3
1.3 机械设备故障诊断方法的分类	6
1.4 机械设备故障诊断技术的发展概况	8
习题与思考题	10
第 2 章 机械故障诊断中的信号分析与处理	11
2.1 机械物理信号分析的基础知识.....	11
2.2 检测信号的时域分析方法.....	16
2.3 检测信号的频域分析方法.....	30
2.4 倒频谱分析方法.....	41
2.5 小波分析.....	47
习题与思考题	55
第 3 章 机械故障诊断的时序模型分析方法	56
3.1 时间序列模型结构特征.....	56
3.2 自回归模型的参数、阶次的确定	58
3.3 自回归谱的概念和应用.....	64
3.4 设备状态变化趋势性及预测.....	69
习题与思考题	74
第 4 章 状态识别与判决方法原理	76
4.1 概述.....	76
4.2 主分量分析.....	77
4.3 贝叶斯决策理论方法.....	82
4.4 距离函数分类法.....	89
* 4.5 故障诊断的模糊模式识别方法.....	93
* 4.6 神经网络及其在故障诊断中的应用	107

习题与思考题	121
第 5 章 旋转机械的振动监测与故障诊断	123
5.1 概述	123
5.2 旋转机械振动监测参数与分析	126
5.3 旋转机械典型故障的机理和特征	133
5.4 旋转机械振动故障诊断示例	145
习题与思考题	148
第 6 章 往复机械的监测与诊断	149
6.1 往复机械的动力性能监测	149
6.2 往复机械故障诊断的油样分析法	151
6.3 瞬时转速检测法	160
6.4 往复机械故障的振动诊断法	162
6.5 柴油机供油系统故障诊断压力波形分析法	168
习题与思考题	169
第 7 章 典型传动部件的故障及其诊断方法	170
7.1 滚动轴承异常的基本形式	170
7.2 滚动轴承的振动类型及其故障特征	171
7.3 滚动轴承故障诊断方法	176
7.4 齿轮异常的基本形式及振动特点	182
7.5 齿轮故障诊断方法	192
习题与思考题	198
第 8 章 机械制造过程工况监测与故障诊断	199
8.1 概述	199
8.2 切削过程刀具磨损与破损的监测与诊断	200
8.3 金属切削颤振的在线监测	205
8.4 切削状态的在线识别	208
8.5 磨削过程表面烧伤的识别	211
8.6 FMS 和 CIMS 的监测与诊断系统概述	214
习题与思考题	217
第 9 章 无损检测技术	218
9.1 光学检测	218
9.2 渗透与磁粉检测	220
9.3 涡流检测	223
9.4 射线检测	226

9.5 超声波检测	229
9.6 声发射检测	234
习题与思考题	237
参考文献	238

绪 论

1.1 机械设备故障诊断技术的意义、目的和内容

随着科学技术与生产的高度发展,各学科相互渗透、相互交叉、相互促进,形成了设备诊断技术这一生命力旺盛的新兴学科。它是一种了解和掌握设备在使用过程中的状态,确定其整体或局部是正常或异常,早期发现故障及其原因,并能预报故障发展趋势的技术。通俗地说,它是一种给设备“看病”的技术,这里所说的“设备”是指机械设备和某些电气设备。

机械设备故障诊断技术日益获得重视与发展的原因是,随着科学技术与生产的发展,机械设备工作强度不断增大,生产效率、自动化程度越来越高,同时设备更加复杂,各部分的关联愈加密切,往往某处微小故障就爆发链锁反应,导致整个设备乃至与设备有关的环境遭受灾难性的毁坏。这不仅造成巨大的经济损失,而且会危及人身安全,后果极为严重。例如,1973年美国三里岛核电站堆芯损坏事故;1985年美国航天飞机“挑战者号”的坠毁;1984年印度博帕尔市农药厂异氰酸甲酯毒气外泄事故;1986年前苏联切尔诺贝利核电站泄漏事故;1986年欧洲莱茵河瑞士化学工业污染事故等等,都是设备故障造成的震惊世界的恶性事故。国内,1982年江苏某化肥厂合成氨压缩机组发生强烈振动,三次停机,损失达千万元以上;1995年山西某电厂一台20万千瓦汽轮发电机组毁坏,直接损失达千万元。据统计,重要设备因事故停机造成的损失极为严重:一个乙烯球罐停产一天,损失产值500万元,利润200万元;一台大型化纤设备停产1小时,损失产值80万元。英国有人在1984年发表论文认为,对大型汽轮发电机组进行振动监视,获利与投资之比为17:1,在英国西南地区,每台发电机组如减少2.5%的事故与检修损失,每年获利可达5.5亿英镑。这表明采用设备诊断技术,保证设备可靠而有效地运行是极为重要的。

设备诊断技术日益获得重视与发展的另一个重要原因是能改革维修体制,大量节省维修费用。当前,国内外对机械设备主要采用计划维修。在许多场合下,这是非常不合理的,不该修的修了,不仅费时花钱,甚至降低设备工作性能;该修的又没修,不仅降低设备寿命,而且导致事故。英国曾对2000

家工厂调查,结果表明采用诊断技术后,每年设备维修费可节约3亿英镑。日本有资料指出,采用诊断技术后,每年设备维修费减少20%~50%,故障停机减少75%。我国的大型钢厂,每年设备维修费达2亿元至3亿元。某港口有5台15吨带斗机,原规定每台每年小修停机45天,实行状态监测与维护后5台每年共停机45天,多获产值160万元。某机械施工单位拥有工程机械232台,采用状态维修后,维修材料费降低30%,维修工作量降低47%。有资料表明,大量生产的发动机,其修理劳动量是制造劳动量的5~10倍。日本航空公司很早就对JT3D喷气发动机采用监控技术,使大修寿命从1200小时提高到12000小时以上,直至取消大修。大量正反事例都表明:对机械设备,特别是关键设备实行工况监控与故障诊断的必要性与迫切性。

在机械制造领域中,如柔性制造系统(FMS)、计算机集成制造系统(CIMS)等,故障诊断技术也具有相同的重要性。然而在机械制造业中,大量的单件、小批量生产,在传统的生产环境中,一般机床设备操作与质量控制主要靠人进行,这时故障诊断技术的地位就没有前述连续生产系统显得那么重要。但对于某些关键机床设备,因故障存在而导致加工质量降低,使整个机器产品质量不能保证,这时故障诊断技术也不容忽视。

从设备诊断技术的起源与发展来看,设备诊断技术的目的应是“保证可靠地、高效地发挥设备应有的功能”。这包含三点:一是保证设备无故障,工作可靠;二是保证物尽其用,设备要发挥其最大的效益;三是保证设备在将有故障或已有故障时,能及时诊断出来,正确地加以维修,以减少维修时间,提高维修质量,节约维修费用,应使重要的设备能按其状态进行维修(即视情维修或预知维修),改革目前按时维修的体制。应指出,设备诊断技术应为设备维修服务,可视为设备维修技术的内容,但它决不仅限于为设备维修服务,正如前两点所示,它还应保证设备能处于最佳的运行状态,这意味着它还应为设备的设计、制造与运行服务。例如,它应能保证动力设备具有良好的抗振、消振、减振能力,具有良好的出力能力等。还应指出,所谓故障是指设备丧失其规定的功能。显然,故障不等于失效,更不等于损坏,失效与损坏是严重的故障。

设备诊断技术的最根本的任务是通过测取设备的信息来识别设备的状态,因为只有识别了设备的有关状态,才有可能达到设备诊断的目的。概括起来,正如对人体诊断一样,一是预防与保健,二是看病与处置。对于设备的诊断,一是防患于未然,早期诊断;二是诊断故障,采取措施。具体讲,设备诊断技术应包括以下5方面内容。

1. 正确选择与测取设备有关状态的特征信号

显然,所测取的信号应该包含设备有关状态的信息。例如,诊断起重桁架有无裂纹决不能靠测取桁架各点温度来判定,因温度信号中不包含裂纹有无的信息,而测取桁架的振动信号则可达到目的,因为振动信号中包含了结构有无裂纹的信息。这种信号可称为特征信号。

2. 正确地从特征信号中提取设备有关状态的有用信息

一般来讲,从特征信号来直接判明设备状态的有关情况,查明故障的有无,是比较难的。例如,一般难于从结构的振动信号直接判明结构有无裂纹,还需要根据振动理论、信号分析理论、控制理论等提供的理论与方法,加上试验研究,对特征信号加以处理,提取有用的信息(称为征兆),才有可能判明设备的有关状态。例如,理论分析与试验研究表明,从振动信号中计算出的固有频率这一征兆固然可用,但对结构有无裂纹产生并不敏感,而计算出的频率特性(或称频响函数)却存在着十分敏感的频带,因此,以频率特性作为征兆则更为合适。所谓敏感的强弱,是指征兆的相对变化与有关状态相对变化之比的大小。这在概念上是成立的,但在实际上就不一定计算或测试得出来。

征兆,可以是结构的物理参数(如质量、刚度等)、结构的模态参数(如固有频率、模态阻尼、模型等),可以是设备的工作特征(如耗油率、工作转速、功率等),可以是信号的统计特性(如均值、方差、自功率谱等),也可以是由信号中得出的其他特征量(如自回归模型参数等)。

3. 根据征兆正确地进行设备的状态诊断

一般来讲,还不能直接采用征兆来进行设备的故障诊断、识别设备的状态。这时,可以采用多种的模式识别理论与方法,对征兆加以处理,构成判别准则,进行状态的识别与分类。例如,对发动机的正常状态、阀撞击状态与连杆撞击状态,在测取振动信号,采用时序方法加以处理,建立自回归模型,将自回归参数与残差方差作为征兆后,可用此征兆构成 Kullback - Liebler 信息距离这一判别准则,来识别发动机所处的状态。显然,状态诊断这一步是设备诊断重点之所在。当然,这决不表明设备诊断的成败只取决于状态诊断这一步,特征信号与征兆的获取正确与否,应该是能否进行正确的状态诊断的前提。

由此应指出,征兆既用于由外表现象推断内部状态,此时可称为症候;又用于由现在现象推断未来状态,此时可称为预兆。状态诊断既包括诊断设备是否将发生什么故障,此即早期诊断,也包括诊断设备已发生什么故障,此即故障诊断。

4. 根据征兆与状态正确地进行设备的状态分析

当状态为有故障时,则应采用有关方法进一步分析故障位置、类型、性质、原因与趋势等。例如,故障树分析是分析故障原因的一种有效方法,当然,故障的原因往往是次一级的故障;如轴承烧坏是故障,其原因是输油管不输油,不输油是因油管堵塞,后者是因滤油器失效等等,这些原因就可称为第二、三、四级故障。正因为故障的原因可能是次级故障,从而有关的状态诊断方法也可用于状态分析。当状态为无故障时,则可用 Kalman 滤波、时序模型等方法进一步分析状态趋势,预计未来情况。

5. 根据状态分析正确地作出决策

干预设备及其工作进程,以保证设备可靠、高效地发挥其应有功能,达到设备诊断的目的。所谓干预包括人为干预和自动干预,即包括调整、修理、控制、自诊断等等。

应当指出,实际上往往不能直接识别设备的状态,因此事先要建立同状态一一对应的基准模式,由征兆所作出的判别准则,此时是同基准模式相联系来对状态进行识别与分类的。显然,将上述设备诊断内容加以概括,可得到图 1-1 所示的设备诊断过程框图。

所谓简易诊断主要指状态诊断这步只判明设备有无故障,但不作故障分析;而所谓精密诊断则包括状态诊断与状态分析这两步,还可能要判明多种故障的情况。

一般所谓工况监视实际上是状态监视,故障是设备的异常状态;因此,工况监视是故障诊断(亦即设备诊断)的基础。设备诊断过程可以说是设备的工况监控、分析与干预过程。

1.2 设备故障的信息获取和检测方法

1.2.1 设备故障信息的获取方法

对设备故障进行诊断,首先应获取有关信息。信息是提供人们判断或识别状态的重要依据,是指某些事实和资料的集成。信号是信息的载体,因而设备故障诊断技术在一定意义上是

属于信息技术的范畴。充分地检测足够量的、能反映系统状态的信号对诊断来说是至关重要的,一个好的诊断系统首先应该能正确地、全面地获取监测和诊断所必须的全部信息。下面介绍信息获取的几种方法。

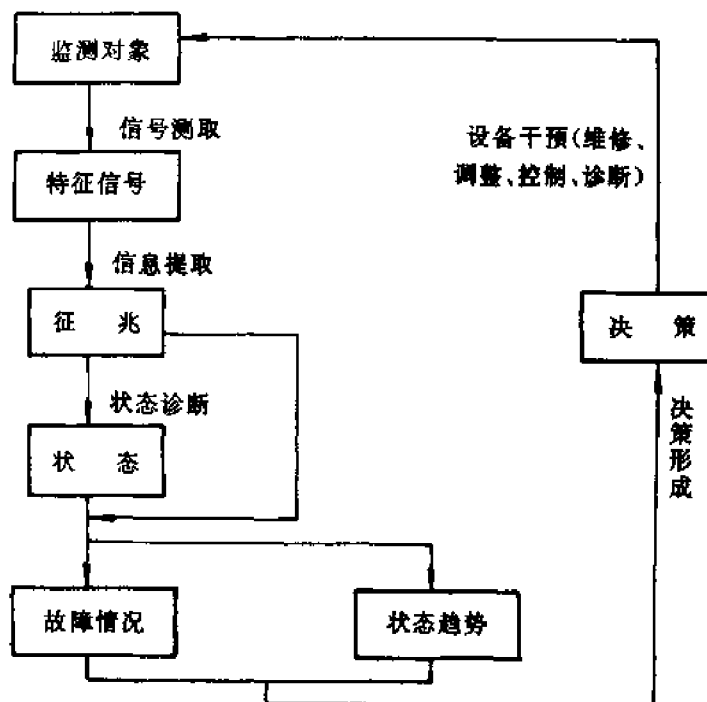


图 1-1 设备诊断过程框图

1. 直接观测法

应用这种方法对机器状态作出判断主要靠人的经验和感官,且限于能观测到的或接触到的机器零件。这种方法可以获得第一手资料,更多的是用于静止的设备。在观测中有时使用了一些辅助的工具和仪器,如倾听机器内部声音的听棒,检查零件内孔有无表面缺陷的光学窥镜,探查零件表面有无裂纹的磁性涂料及着色渗透剂等,来扩大和延伸人的观测能力。

2. 参数测定法

根据设备运行的各种参数的变化来获取故障信息是广泛应用的一种方法。因为机器运行时各部件的运动必然会有各种信息,这些信息参数可以是温度、压力、振动或噪声等,它们都能反映机器的工作状态,如根据轴瓦下部油压变化可以了解转子对中情况,分析油液中金属碎屑情况可以了解轴瓦磨损程度等。在运转的设备中,振动是最重要的信息来源,在振动信号中包含了各种丰富的故障信息,任何机器在运转时工作状态发生了变化,必然会从振动信号中反映出来。对旋转机械来说,目前在国内外应用最普遍的方法是利用振动信号对机器状态进行判别。从测试手段来看,利用振动信号进行测试也最方便、实用,要利用振动信号对故障进行判别,首先应从振动信号中提取有用的特征信号,即利用信号处理技术对振动信号进行处理。目前应用最广泛的处理方法是进行频谱分析,即从振动信号中的频率成分和分布情况来判断故障。

其他如噪声、温度、压力、变形等参数也是故障信息的重要来源。

3. 磨损残余物的测定

机器零件,例如轴承、齿轮、活塞环、缸套等在运行过程中的磨损残渣可以在润滑油中找到。测定的方法分为三种:一种是直接检查残渣,以及测定油膜间隙内电容或电感的变化、润滑油混浊度的变化等方法以迅速获得零件失效的信息。第二种是残渣的收集,例如采用磁性探头、特殊的过滤器等收集齿轮、滚动轴承等工作表面疲劳引起的大块剥落颗粒。第三种方法是油样分析,可以确定机器中什么零件在磨损。

测定机器零部件如轴承、齿轮、活塞环等的磨损残渣在润滑油中的含量,也是一种有效的获取故障信息的方法。根据磨损残渣在润滑油中含量及颗粒分布可以掌握零件磨损情况,并可预防机器故障的发生。

4. 设备性能指标的测定

设备性能包括整机及零部件性能。通过测量机器性能及输入、输出量的变化信息来判断机器的工作状态也是一种重要方法。

例如,柴油机耗油量与功率的变化,机床加工零件精度的变化,风机效率的变化等均包含着故障信息。

对机器零部件性能的测定,主要反映在强度方面,这对预测机器设备的可靠性,预报设备破坏性故障具有重要意义。

1.2.2 设备故障的检测方法

由于机器运行的状态千差万别,因而出现的故障也多种多样,采用的检测方法各不相同。本节将对具体的各种故障应采用的方法及其应用范围作一介绍。

1. 振动和噪声的故障检测

机器在运行的过程中振动是诊断的重要信息,它反映了机器的状态。根据检测的振动和噪声信号,可采用以下方法对机器状态进行诊断:

(1) 振动法:对机器主要部位的振动值如位移、速度、加速度、转速及相位值等进行测定,与标准值进行比较,据此可以宏观地对机器的运行状况进行评定,这是最常用的方法。

(2) 特征分析法:对测得的上述振动量在时域、频域、时-频域进行特征分析,用以确定机器各种故障的内容和性质。

(3) 模态分析与参数识别法:利用测得的振动参数对机器零部件的模态参数进行识别,以确定故障的原因和部位。

(4) 冲击能量与冲击脉冲测定法:利用共振解调技术以测定滚动轴承的故障。

(5) 声学法:对机器噪声的测量可以了解机器运行情况并寻找故障源。

2. 材料裂纹及缺陷损伤的故障检测

材料裂纹包括应力腐蚀裂纹及疲劳裂纹,一般可采用下述方法进行检测:

(1) 超声波探伤法:该方法成本低,可测厚度大,速度快,对人体无害,主要用来检测平面型缺陷。

(2) 射线探伤法:主要采用 X 和 γ 射线。该方法主要用于展示体积型缺陷,适用于一切材料,测量成本较高,对人体有一定损害,使用时应注意。

(3) 渗透探伤法:主要有荧光渗透与着色渗透两种。该方法操作简单,成本低,应用范围

广,可直观显示,但仅适用于有表面缺陷的损伤类型。

(4) 磁粉探伤法:该法使用简便,较渗透探伤更灵敏,能探测近表面的缺陷,但仅适用于铁磁性材料。

(5) 涡流探伤法:这种方法对封闭在材料表面下的缺陷有较高的检测灵敏度,它属于电学测量方法,容易实现自动化和计算机处理。

(6) 激光全息检测法:它是 20 世纪 60 年代发展起来的一种技术,可检测各种蜂窝结构、叠层结构、高压容器等。

(7) 微波检测技术:它也是近几十年来发展起来的一种新技术,对非金属的贯穿能力远大于超声波方法,其特点是快速、简便,是一种非接触式的无损检测。

(8) 声发射技术:它主要对大型构件结构的完整性进行监测和评价,对缺陷的增长可实行动态、实时监测,且检测灵敏度高,目前在压力容器,核电站重点设备及放射性物质泄漏,输送管道焊接部位缺陷等方面的检测获得了广泛的应用。

3. 设备零部件材料的磨损及腐蚀故障检测

这类故障除采用上述无损检测中的超声探伤法外尚可用下列方法:

(1) 光纤内窥技术:它是利用特制的光纤内窥探测器直接观测到材料表面磨损及腐蚀情况。

(2) 油液分析技术:油液分析技术可分为两大类:一类是油液本身的物理、化学性能分析,另一类是对油液污染程度的分析。具体的方法有光谱分析法与铁谱分析法。

4. 温度、压力、流量变化引起的故障检测

机器设备系统的有些故障往往反映在一些工艺参数,如温度、压力、流量的变化中,在温度测量中除常规使用的装在机器上的热电阻、热电偶等接触式测温仪外,目前在一些特殊场合使用的非接触式测温方法有红外测温仪和红外热像仪,它们都是依靠物体的热辐射进行测量。

1.3 机械设备故障诊断方法的分类

机器设备有各种类型,其工作条件又各不相同,故对不同机器的故障往往需要采用不同的方法来诊断。对机器进行故障诊断的方法可以按如下几种方式进行分类。

1.3.1 按诊断的目的和要求分类

1. 功能诊断和运行诊断

功能诊断是针对新安装或刚维修后的机器或机组,需要检查它们的运行工况和功能是否正常,并且按检查的结果对机器或机组进行调整。而运行诊断是针对正常工作的机器或机组,需要监视其故障的发生和发展。

2. 定期诊断和连续监控

定期诊断是每隔一定时间,例如 1 个月或数月对工作状态下的机器进行常规检查。连续监控则是采用仪表和计算机信息处理系统对机器运行状态进行不间断地监视或控制。两种诊断方式的采用,取决于设备的关键程度、设备事故影响的严重程度、运行过程中性能下降的快慢,以及设备故障发生和发展的可预测性,如表 1-1 所示。

表 1-1 两种诊断方法采用的条件

性能下降速度	故障不可预测	故障可预测
快	连续监控	定期更换
慢	定期诊断	定期诊断

3. 直接诊断和间接诊断

直接诊断是直接确定关键部件的状态,如主轴承间隙、齿轮齿面磨损、燃气轮机叶片的裂纹以及在腐蚀环境下管道的壁厚等。直接诊断往往受到机器结构和工作条件的限制而无法实现,这时,就不得不采用间接诊断。

所谓间接诊断是通过二次诊断信息来间接判断机器中关键零部件的状态变化。多数二次诊断信息属于综合信息,例如前述的用润滑油温升来反映主轴承的运行状态,因此,在间接诊断中出现误诊和漏检两种情况的可能性都会增大。

4. 常规工况下诊断和特殊工况下诊断

多数诊断在机器正常工作条件下就能进行的,只有在个别情况下才需要创造特殊的工作条件来拾取信息。例如,动力机组的起动和停车过程,需要跨过转子扭转、弯曲的几个临界转速。利用起动和停车过程的振动信号作出的瀑布图,常包含着许多在常规诊断中所得不到的诊断信息。

5. 在线诊断与离线诊断

在线诊断是指对于大型、重要的设备为了保证其安全和可靠运行需要对所监测的信号自动、连续、定时的进行采集与分析,对出现的故障及时作出诊断;离线诊断是通过磁带记录仪或数据采集器将现场的信号记录并储存起来,再在实验室进行回放分析,对于一般中小型设备往往采用离线诊断方式。

1.3.2 按诊断对象分类

- (1) 旋转机械诊断技术:如汽轮发电机组、燃气轮机组、水轮机组、风机及泵离心等。
- (2) 往复机械诊断技术:包括内燃机、往复式压缩机及泵等。
- (3) 工程结构诊断技术:如海洋平台、金属结构、框架、桥梁、容器等。
- (4) 运载器和装置诊断技术:如飞机、火箭、航天器、舰艇、火车、汽车、坦克、火炮、装甲车等。
- (5) 通讯系统诊断技术:如雷达、电子工程等。
- (6) 工艺流程诊断技术:主要是生产流程、传送装置及冶金压延等设备。

1.3.3 按信息提取方式分类

信号是信息的载体,设备出现故障时所表现出的征兆是通过检测信息,即信号来体现的,所以可以按找特征信号与征兆之间的关系对诊断方法进行分类。

1. 函数分析法

特征信号与征兆之间存在定量的函数关,可用数学分析方法,例如状态空间分析,由特征信号求出征兆。

2. 统计分析法

可用数理统计方法由特征信号求出征兆。统计分析法又可为分非参数模型统计法即传统的信号处理方法和参数模型统计法两种。它根据信号的采样数据,首先建立差分方程形式的参数模型,再用模型的参数或用模型计算出信号统计特性、结构固有的特性或其他特性作为征兆。这种方法之一就是时序模型方法,其最大优点之一是没有传统信号处理方法中对数据加“窗”的致命弱点,其次几乎能将蕴含在采样数据中全部信号凝聚在少数几个模型参数之中。

1.3.4- 按照状态诊断方式分类

设备的各种运行状态均会通过其相应的各种征兆表现出来,所以按照征兆与状态之间的关系可将诊断方法分为以下几种。

1. 对比诊断法

目前应用最广,应事先通过统计归纳、试验研究、分析计算,确定同各有关状态一一对应的征兆(即基准模式或标准档案),然后将获得的征兆同基准模式对比,即可确定设备的状态。

2. 函数诊断法

在征兆与状态之间如存在定量的函数关系,则在获得征兆后即可用相应的函数关系计算出状态。

3. 逻辑诊断法

在征兆与状态间如存在逻辑关系时,则在获得征兆后即可用相应物理或数理逻辑关系推理判明有关状态。如润滑油检测法、激光全息法等均属物理逻辑法,决策布尔函数法则属数理逻辑法。

4. 统计诊断法

一般模式识别理论中的统计模式法,它用于征兆与状态之间存在统计关系时。

5. 模糊诊断法

它是一种较新的诊断方法,其特点有二:第一,它采用多因素诊断,因为一种状态可在不同程度地引起多种征兆,而一种征兆又可在不同程度上反映多种状态;第二,它模仿人利用模糊逻辑而精确识别事物这一特性。这样,它根据所获得的征兆,列出征兆隶属度模糊向量,再根据以实践为基础所得到的模糊矩阵,利用模糊数学方法,计算出状态隶属度模糊向量,最后根据此向量中各元素的大小确定有关状态的情况。

6. 智能诊断法

人工智能的目的是使计算机去做原来只有人才能做的智能工作,包括推理、理解、规划、决策、抽象、学习等功能。专家系统是实现人工智能的重要形式,目前已广泛用于诊断、解释、设计、规划、决策等各个领域。现在国内外已发展了一系列用于设备故障诊断的专家系统,获得了良好的效果。

专家系统由知识库、推理机以及工作存储空间(包括数据库)组成。实际的专家系统还应有知识获取模块、知识库管理维护模块、结束模块、显示模块以及人机界面等。

1.4 机械设备故障诊断技术的发展概况

设备的故障诊断,实际上自有工业生产以来就已存在,早期人们依据对设备的触摸,对声

音、振动等状态特征的感受,凭借工匠的经验,可以判断某些故障的存在,并提出修复的措施,例如有经验的工人常利用听棒来判断旋转机械轴承及转子的状态。但是故障诊断技术作为一门学科,则是 20 世纪 60 年代以后才发展起来的。

对设备故障诊断技术的发展情况,已有不少文献进行了回顾和综述。最早开展故障诊断技术研究的大概是美国,美国 1961 年开始执行阿波罗计划后出现了一系列设备故障,促使 1967 年在美国宇航局(NASA) 倡导下,由美国海军研究室(ONR) 主持美国机械故障预防小组(MFPG),积极从事故障诊断技术的研究和开发。1971 年 MFPG 划归美国国家标准局(NSB) 领导,成为一个官方领导的组织,下设故障机理研究、检测、诊断和预测技术、可靠性设计和材料耐久性评价四个小组,平均每年召开两次会议。美国机械工程师学会(ASME) 领导下的锅炉压力容器监测中心(NBBI) 对锅炉压力容器和管道等设备的诊断技术作了大量的研究,制订了一系列有关设备设计、制造、试验和故障诊断及预防的标准规程,研究推行了设备的声发射(Acoustic Emission) 诊断技术。其他如 Johns Mitchel 公司的超低温水磁和空压机监测技术, SPIRE 公司的用于军用机械的轴与轴承诊断技术, TEDECO 公司的润滑油分析诊断技术等都在国际上具有特色。在航空运输方面,美国在可靠性维修管理的基础上,大规模地对飞机进行状态监测,发展了应用计算机控制的飞行器数据系统(AIDS),利用大量飞行中的信息来分析飞机各部位的故障原因并能发出消除故障的命令。这些技术已普遍用于波音 747 这一类巨型客机,大大提高了飞行的安全性。据统计,世界班机的每亿旅客公里的死亡率已从 20 世纪 60 年代的 0.6 左右下降到 20 世纪 70 年代的 0.2 左右。在旋转机械故障诊断方面,首推美国西屋公司,从 1976 年开始研制,到 1990 年已发展成网络化的汽轮发电机组智能化故障诊断专家系统,其三套人工智能诊断软件(汽轮机 TurbinAID,发电机 GenAID,水化学 ChemAID) 共有诊断规则近 1 万条,已对西屋公司所产机组的安全运行发挥了巨大的作用,取得了很大的经济效益。还有以 Bentley Nevada 公司的 DDM 系统和 ADRE 系统为代表的多种机组在线监测诊断系统等。

英国在 20 世纪 60 年代末 70 年代初,以 R. A. Collacott 为首的英国机械保健中心(U. K. Mechanical Health Monitoring Center) 开始诊断技术的开发研究。1982 年曼彻斯特大学成立了沃福森工业维修公司(WIMU), Michael Neale and Associte 公司等几家公司,担任政府的顾问、协调和教育工作,开展了咨询、制定规划、合同研究、业务诊断、研制诊断仪器、研制监测装置、开发信号处理技术、教育培训、故障分析、应力分析等业务活动。在核发电方面,英国原子能机构(UKAEA) 下设一个系统可靠性服务站(SRS) 从事诊断技术的研究,包括利用噪声分析对炉体进行监测,以及对锅炉、压力容器、管道的无损检测等,起到了英国故障数据中心的作用。在钢铁和电力工业方面英国也有相应机构提供诊断技术服务。

设备诊断技术在欧洲其他一些国家也有很大进展,它们在广度上虽不大,但都在某一方面具有特色或占领先地位,如瑞典的 SPM 轴承监测技术,挪威的船舶诊断技术,丹麦的振动和声发射技术,等等。

如果说美国在航空、核工业以及军事部门中诊断技术占有领先地位,那么日本在某些民用工业,如钢铁、化工、铁路等部门发展得很快,占有某种优势。他们密切注视世界性动向,积极引进消化最新技术,努力发展自己的诊断技术,研制自己的诊断仪器。例如 1970 年英国提出了设备综合工程学后,日本设备工程师协会紧接着在 1971 年开始发展自己的 TPM(全员生产维修),并每年向欧美派遣“设备综合工程学调查团”,了解诊断技术的开发研究工作,经过 6 年于

1976年基本达到实用阶段。日本机械维修学会、计测自动控制学会、电气学会、机械学会也相继设立了自己的专门研究机构。国立研究机构中,机构技术研究所和船舶技术研究所重点研究机械基础件的诊断技术;东京大学、东京工业大学、京都大学、早稻田大学等高等学校着重基础性理论研究。其他民办企业,如三菱重工、川崎重工、日立制作所、东京芝浦电气等以企业内部工作为中心开展应用水平较高的实用项目,例如三菱重工的白木万博在旋转机械故障诊断方面开展了系统的工作,他所研制的“机械保健系统”在汽轮发电机组故障监测和诊断方面已起到了有效的作用。

我国于1983年由原国家经委发布了《国营工业交通设备管理试行条例》,1987年国务院正式颁布了《全民所有制工业交通企业设备管理条例》,规定:“企业应当积极采用先进的设备管理方法和维修技术,采用以设备状态监测为基础的设备维修方法”,其后冶金、机械、核工业等部门还分别提出了具体实施要求,使我国故障诊断技术的研究和应用在全国普遍展开。自1985年以来,由中国设备管理协会设备诊断技术委员会、中国振动工程学会机械故障诊断分会和中国机械工程学会设备维修分会分别组织的全国性故障诊断学术会议已先后召开十余次,极大地推动了我国故障诊断技术的发展。现在全国已有很多单位开展设备诊断技术的研究工作,全国各行业都很重视在关键设备上装备故障诊断系统,特别是智能化的故障诊断专家系统,其中突出的有电力系统、石化系统、冶金系统以及高科技产业中的核动力电站、航空部门和载人航天工程等。工作比较集中的是大型旋转机械故障诊断系统,已经开发了20种以上的机组故障诊断系统和十余种可用来做现场简易故障诊断的便携式现场数据采集器。一些高等院校已培养了大量以设备故障诊断技术为选题的硕士研究生和博士研究生。我国的故障诊断事业正在蓬勃发展,将在我国经济建设中发挥越来越大的作用。

习题与思考题

- 1-1 机械设备状态监测与故障诊断的主要内容是什么?
- 1-2 机械设备状态监测与故障的常用识别方法有哪些?
- 1-3 简述获取诊断信息的方法。
- 1-4 设备状态识别方法有哪些不确定性?会产生哪几类错误?

机械故障诊断中的信号分析与处理

2.1 机械物理信号分析的基础知识

信号或动态数据的处理与分析是机械设备故障诊断的前提和基础。本章所说的信号是指测量信号,它是对系统的某些物理量,如位移、速度、加速度、应力、应变等进行观测获得的数据,其共同特点是随时间而变化,故将其称为动态信号,它们代表了系统的运行状态和特征。对信号分析与处理的目的是改变信号的形式,提取有用的信息,以便对所研究的机械运行状态作出估计和辨别。这里的“时间”是泛指概念,有时可以是空间坐标或时空坐标。

2.1.1 测量信号分类

按照信号随时间变化的规律不同可按图 2-1 对信号进行分类。

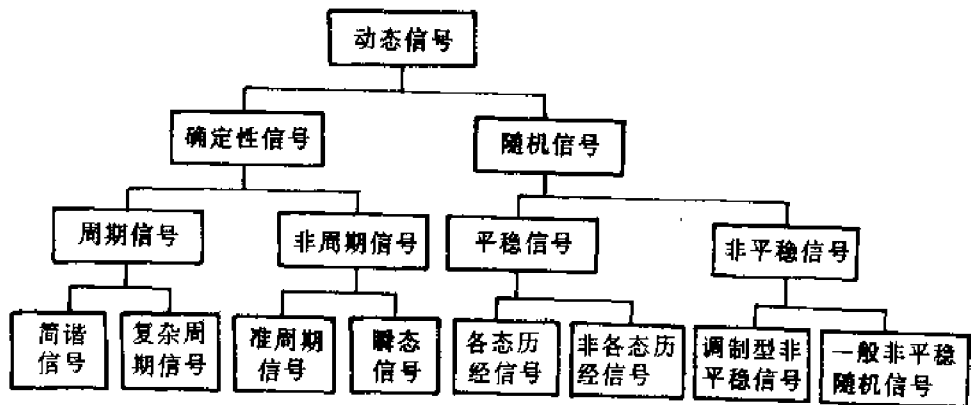


图 2-1 信号分类

1. 确定性信号

如果描述系统状况的状态变量可以用确定的时间函数来表述,则称这样的物理过程是确定性的,而描述它们的测量数据就是确定性信号,如图 2-2 所示。

周期信号包括简谐信号和复杂周期信号。表述简谐信号的基本物理量

是频率、振幅和初相位；复杂周期信号可借助傅里叶级数展成一系列离散的简谐分量之和，其中任意两个分量的频率比都是有理数。

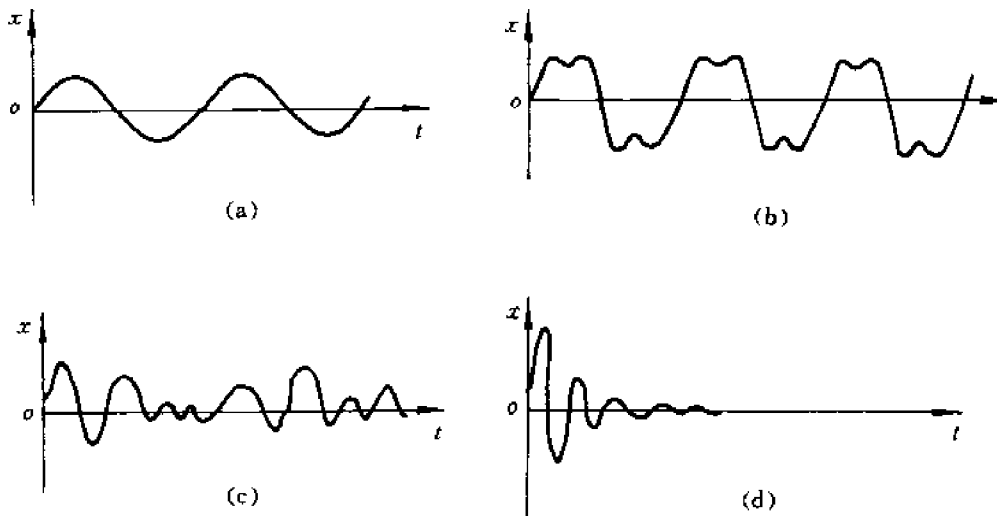


图 2-2 确定性信号

非周期信号包括准周期信号和瞬态信号。准周期信号也是由一些不同离散频率的简谐信号合成的信号，但它不具有周期性，组成它的简谐分量中总有一个分量与另一个分量的频率比为无理数；瞬态信号的时间函数为各种脉冲函数或衰减函数，如有阻尼自由振动的历程就是瞬态信号。瞬态信号可借助傅里叶变换而得到确定的连续频谱函数。

2. 随机信号

如果描述系统状况的状态变量不能用确切的时间函数来表述，无法确定状态变量在某时刻的确切数值，其物理过程具有不可重复性和不可预知性时，则称这样的物理过程是随机的，而描述它们的测量数据就是随机信号，在数学上称为随机过程。对随机过程的研究通常转化为对随机变量的研究，可借助概率论和随机过程理论来进行。

在工程实践中，通常是在相同的条件下，对某台设备（或同一型号的设备）进行大量的重复试验所得的试验数据进行统计分析，来研究其规律性。图 2-3 示出其随机试验各次观测所得的时间历程，这些时间历程的集合总体就表达了该随机过程，记为

$$X(t) = \{x_1(t) \quad x_2(t) \quad \cdots \quad x_N(t) \quad \cdots\} \quad (2.1)$$

其中的时间历程称为样本函数。随机过程的随机性是通过各个样本函数之间的区别以及这种区别的不可预测性体现出来的，随机过程在某时刻 t_i 的取值 $x_1(t_i), x_2(t_i), \cdots, x_N(t_i), \cdots$ 为一随机变量。

如果随机过程 $X(t)$ 各样本函数不同时刻取值的随机变量的统计特性（如均值、均方值、概率密度等）分别相等，即统计特性与统计时间无关，则称 $X(t)$ 为平稳随机过程；反之，称为非平稳随机过程。对平稳随机过程，若用任一样本函数得到的时间统计特性与随机过程 $X(t)$ 所有样本统计特性（集合统计特性）相等时，这样的随机过程为各态历经平稳随机过程。正常工作的机械系统，表征其过程的随机信号是平稳和弱平稳的；对于过渡状态下的机械系统，其信号往往是非平稳的。三种类型的随机信号是这样区分的：如果在时间的某一时刻，信号的平均值、方差和高阶矩（包括峭度、偏斜度指标等）都保持不变，则信号是平稳的；如果不考虑高阶

矩,只考虑平均值和方差,而这两个参数保持不变,则称信号是弱平稳的。很明显,弱平稳的条件要比平稳的要求放宽多了。本书中对随机信号的讨论仅限于各态历经平稳随机过程的范围,且测量样本长度是有限的。

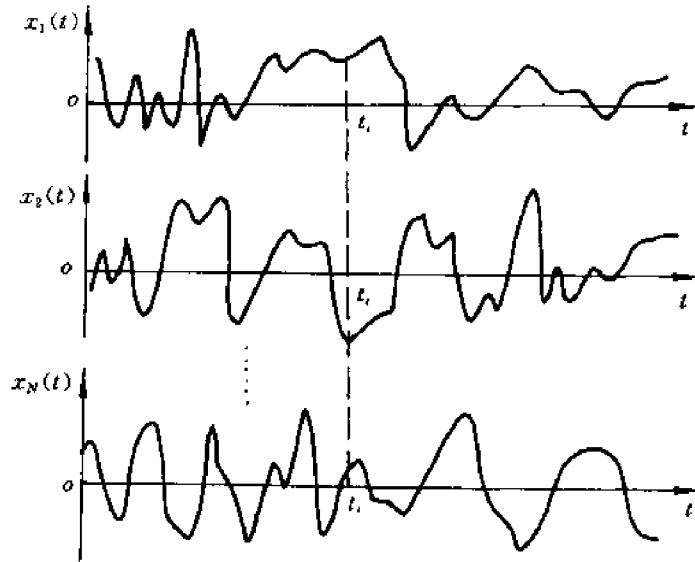


图 2-3 随机过程的样本函数

2.1.2 随机过程的数字特征

随机过程的概率分布函数可以完全刻划随机过程全部统计特性,但在实际问题中要确定随机过程的分布函数并加以分析往往比较困难,甚至是不可能的。从实际应用上看,在很多情况下,只要知道随机过程的某些数字特征就可以了,这些数字特征能从不同角度反映随机过程的统计特性,且运算简单、测量方便。下面简要讨论随机过程的一些基本数字特征。

1. 数学期望(均值函数)

设 $X(t)$ 是一随机过程,其数学期望定义为

$$\mu_x(t) = E[X(t)] = \int_{-\infty}^{\infty} xp(x,t)dx \quad (2.2)$$

式中 $p(x,t)$ 是 $X(t)$ 的概率密度函数。 $E[X(t)]$ 是随机过程 $X(t)$ 的所有样本函数 $x_j(t)$ ($j = 1, 2, \dots$) 在各个时刻 t 的函数值的平均,可认为是随机过程在各个时刻的摆动中心。

2. 均方值

定义随机过程 $X(t)$ 的二阶原点矩

$$\phi_x^2 = E[X^2(t)] = \int_{-\infty}^{\infty} x^2 p(x,t)dx \quad (2.3)$$

式中 $\phi_x^2(t)$ 为 $X(t)$ 的均方值。均方值反映了过程的能量特征,其正平方根值称为均方根值。

3. 方差(均方差值)

定义随机过程 $X(t)$ 的二阶中心矩

$$\sigma_x^2(t) = D[X(t)] = E\{[X(t) - \mu_x(t)]^2\} = \int_{-\infty}^{\infty} [x(t) - \mu_x(t)]^2 p(x,t)dx \quad (2.4)$$

式中 $\sigma_x^2(t)$ 为 $X(t)$ 的方差。方差的正平方根 $\sigma_x(t)$ 称为 $X(t)$ 的标准差,它表示随机过程 $X(t)$

在时刻 t 对于均值 $\mu_x(t)$ 的偏离程度,是数据分散度的测度,在信号分析中代表了信号电平的大小。

如果把均值作为描述随机过程的静态分量,那么标准差就是描述随机过程的动态分量。利用(2.2),(2.3)和(2.4)三式,容易推得

$$\sigma_x^2 = \psi_x^2(t) - \mu_x^2(t) \quad (2.5)$$

4. 相关函数

设 $X(t_1)$ 和 $X(t_2)$ 是随机过程 $X(t)$ 在任意两个时刻 t_1 和 t_2 时的状态, $p(x_1, x_2, t_1, t_2)$ 是相应的二维联合概率密度函数,定义二阶原点混合矩

$$R_x(t_1, t_2) = E[X(t_1)X(t_2)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1 x_2 p(x_1, x_2, t_1, t_2) dx_1 dx_2 \quad (2.6)$$

式中 $R_x(t_1, t_2)$ 为随机过程 $X(t)$ 的自相关函数,简称相关函数。它描述了随机过程 $X(t)$ 自身在两个不同时刻状态之间的线性依从关系和相似程度。

类似地,定义 $X(t_1)$ 和 $X(t_2)$ 的二阶中心混合矩

$$C_x(t_1, t_2) = E\{[X(t_1) - \mu_x(t_1)][X(t_2) - \mu_x(t_2)]\} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [x_1 - \mu_x(t_1)] \cdot [x_2 - \mu_x(t_2)] p(x_1, x_2, t_1, t_2) dx_1 dx_2 \quad (2.7)$$

式中 $C_x(t_1, t_2)$ 为随机过程 $X(t)$ 的自协方差函数,简称协方差函数,它就是已中心化的自相关函数。

引入无量纲的标准化相关系数函数

$$\rho_x(t_1, t_2) = \frac{C_x(t_1, t_2)}{\sigma_x(t_1)\sigma_x(t_2)} \quad (2.8)$$

式中 $\sigma_x(t_i)$ ($i = 1, 2$) 为标准偏差, $\rho_x(t_1, t_2)$ 又简称为自相关系数,且 $-1 \leq \rho_x(t_1, t_2) \leq 1$ 。

当 $\rho_x(t_1, t_2) = \pm 1$ 时,称过程的两个状态 $X(t_1)$ 与 $X(t_2)$ 是完全线性相关的;当 $\rho_x(t_1, t_2) = 0$ 时,称过程是完全不相关的;而 $-1 < \rho_x(t_1, t_2) < 1$ 时,称过程是部分相关的。

在工程上,随机过程的均值和自相关函数是描述随机过程的最重要的统计特性,这是因为,对某些随机过程,例如高斯随机过程,如果已知一阶和二阶统计特性,就能完全决定该过程的全部概率结构;从试验数据的统计计算来看,一阶和二阶统计量近似表达了过程的主要特性,且易于获得并易于进行计算分析。

推广到两个或两个以上的随机过程,可获得描述多个随机过程之间相互依赖关系的互相关函数,包括两个随机过程 $X(t)$ 和 $Y(t)$ 的互相函数 $R_{xy}(t_1, t_2)$, 互协方差函数 $C_{xy}(t_1, t_2)$, 互相关系数函数 $\rho_{xy}(t_1, t_2)$ 。

5. 高阶矩函数

(1) 偏斜度:定义随机过程 $X(t)$ 的三阶中心矩

$$\alpha_3(t) = E\{[X(t) - \mu_x(t)]^3\} = \int_{-\infty}^{\infty} [x(t) - \mu_x(t)]^3 p(x, t) dx \quad (2.9)$$

式中 $\alpha_3(t)$ 为 $X(t)$ 的偏斜度。

(2) 峭度:定义随机过程 $X(t)$ 的四阶原点矩

$$\alpha_4(t) = E[X^4(t)] = \int_{-\infty}^{\infty} x^4 p(x, t) dx \quad (2.10)$$

式中 $\alpha_4(t)$ 为 $X(t)$ 的峭度。

偏斜度和峭度用于将随机过程 $X(t)$ 的概率分布密度和正态分布曲线进行比较,定量地确定偏离正态分布的程度。

2.1.3 测量数据分析与处理方法的分类

测量数据是定量地描述客观事物的物理量。对被测对象监测获取的测量信号是很复杂的,它既包括与所研究的对象存在着直接和间接关系的有用信息,又含有与研究对象无关的干扰,如观测误差和环境噪声等。对测量数据进行分析与处理的目的是去伪存真、去粗取精、由表及里、由此及彼的加工过程,以便最大限度地抑制或消除噪声,突出或提取有用信息,提高信噪比,找出客观事物本身的内在规律和客观事物之间的相互关系。测量数据处理的方法主要有如下几类。

1. 按任务分

(1) 预处理:包括数据准备、编辑和检验,以及剔除奇异点、零均值化、消除趋势项等。

(2) 二次处理:如数据压缩和变换等。由于信号的频率结构更能反映事物的本质,因此快速傅里叶变换(FFT)得到了广泛的应用。

(3) 最终处理:获得最终有用的信息,并进行显示、记录和打印。

2. 按方式分

(1) 在线处理:在进行在线实时状态监测与故障诊断时,常常要求信号的产生和分析的结果几乎同时完成,既不舍弃观测信号,又不使数据“积压”。为此,就必须使信号处理与分析的时间小于或等于相应信号变化所需的采集时间,以便使信号分析的速度能赶上输入信号的变化,具有良好的同时性。

对于 FFT 分析仪,处理时间通常指计算 N 点的快速傅里叶变换所需要的运算时间。这个时间标志着分析仪的计算速度、性能指标、规模大小和应用场合。一台 FFT 分析仪能实时分析的最高频率为 $f_{\max} = N/(k \cdot t_A)$,其中 N 为采样点数, t_A 为处理 N 点所需的时间, k 为采样频率与被分析信号最高频率 f_{\max} 的比值,一般取 $k = 2.56$ 。例如 CF-920FFT 分析仪能在 10 ms 内处理 1 024 点谱分析,则 FFT 实时处理最高频率 $f_{\max} = 40$ kHz,为中速 FFT 分析仪。

(2) 离线处理:一般是分析那些处理时间不受限制的信号。它是首先把现场信号记录下来(一般采用磁带记录仪或数据存储记录仪等),然后再重放信号,对记录下来的数据进行事后分析加工。

3. 按手段分

(1) 模拟式分析:就是对时间和幅值都连续的时间历程信号采用各种模拟式仪器来进行的分析,其主要器件是模拟式滤波器和各种记录仪。

(2) 数字式分析:就是对时间历程信号经过时间上离散化和幅值上量化后的数字信号进行的分析,其主要内容是频谱分析和数字滤波。

数字信号处理的基本组成如图 2-4 所示。

数字式信号处理具有高精度(14 位字长可达 10^{-4} 的精度),高性能指标(分辨率高,动态范围大,线性相位特性好等),高速度,高灵活性,高可靠性,以及便于集成化、小型化、自动化等特点,在现代故障诊断系统中一般采用数字式信号处理。

本章主要介绍数字信号处理的简要内容,以便为机械设备的状态监测和故障诊断提供必要的基础知识。

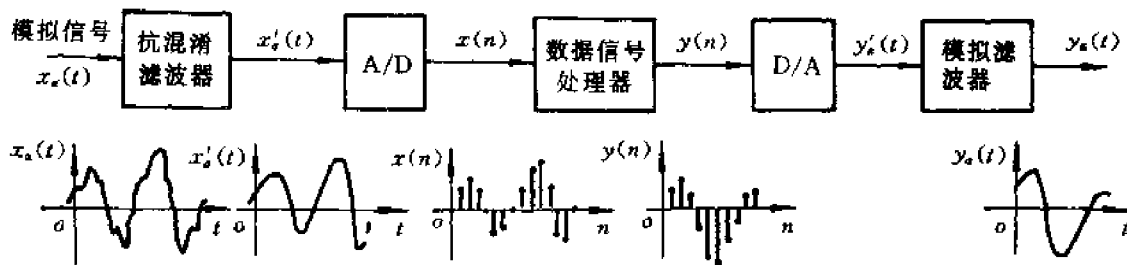


图 2-4 数字信号处理的基本组成

2.2 检测信号的时域分析方法

2.2.1 波形分析及动态指标

在机械故障诊断中,振动信号是最常用的检测信号,直接对振动时域信号的时间历程进行分析和评估是状态监测和故障诊断最简单和最直接的方法,特别是当信号中含有简谐信号、周期信号或短脉冲信号时更为有效。直接观察时域波形可以看出周期、谐波、脉冲,利用波形分析可直接识别共振现象和拍频现象。当然这种分析对比较典型的信号或特别明显的信号以及较有经验的人员才比较适用。此外,还可利用各种动态指标进行诊断。

1. 时域故障诊断的概率分析法

对于各态历经的平稳随机过程可用其时间历程的概率分布来描述。图 2-5 示出某一信号的时间历程及其概率密度函数 $p(x)$, $p(x)$ 可由下列关系式计算:

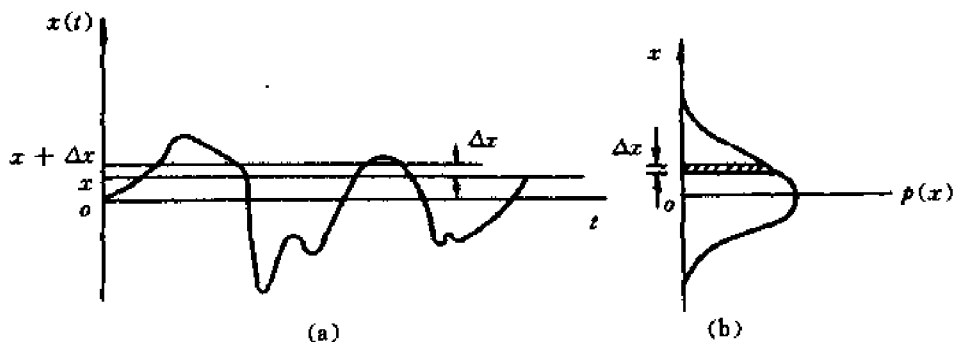


图 2-5 信号及其概率密度函数

$$p(x) = \lim_{\Delta x \rightarrow 0} \frac{P[x < x(t) < x + \Delta x]}{\Delta x} = \lim_{\Delta x \rightarrow 0} \frac{1}{\Delta x} \left[\lim_{T \rightarrow 0} \frac{T_x}{T} \right] \quad (2.11)$$

式中 $P[x < x(t) < x + \Delta x]$ 表示瞬时值落在增量 $(x, x + \Delta x)$ 范围内可能出现的概率, T_x 是在总的观测时间 T 中信号 $x(t)$ 位于 $(x, x + \Delta x)$ 区间内的所有时间之和。图 2-5(b) 表示 $p(x) - x$ 关系,此时横坐标为幅值 x ,故而有将信号的概率密度函数分析称为幅值域分析。利用信号的概率密度分布可对机械设备进行故障诊断,图 2-6 为一高速滚动轴承工作时振动加速度幅值的概率密度函数 $p(x)$ 图,其中实线为正常轴承,虚线为某故障轴承的 $p(x)$ 图。由于磨损、腐蚀、压痕等使振幅增大,谐波增多,反映到 $p(x)$ 图上使其变峭,两旁展宽。

2. 故障诊断的动态指标

(1) 峰值 x_p : 指信号可能出现的最大瞬时值 $\max x(t)$, 它是信号强度的一种描述, 有时人们也用峰峰值 x_{p-p} (即 $\max x(t) - \min x(t)$) 这个指标表示信号的变化范围。在测试时, 对需测信号的峰值事先应有足够的估计, 以便调整测量仪器的范围。

(2) 均值 μ_x 和绝对平均值 $\mu_{|x|}$: 均值是指信号幅值的算术平均值, 可用下式来定义

$$\mu_x = \frac{1}{T} \int_0^T x(t) dt \quad (2.12)$$

式中 T 是观察或测量时间。对周期信号, T 就是信号本身的重复循环周期, 对于均值相等的信号, 其随时间的变化规律并非完全相同, 因此均值只是反映信号中心趋势的一个标志。

绝对平均值的定义为

$$\mu_{|x|} = \frac{1}{T} \int_0^T |x(t)| dt \quad (2.13)$$

在工程中表示了信号经整流后的直流分量。

假如信号 $x(t)$ 的离散值为 $x_i (i = 1, 2, \dots, N)$, 则可得到均值和绝对平均值的一致估计分别为

$$\hat{\mu}_x = \frac{1}{N} \sum_{i=1}^N x_i \quad (2.14)$$

$$\hat{\mu}_{|x|} = \frac{1}{N} \sum_{i=1}^N |x_i| \quad (2.15)$$

(3) 均方值和均方根植: 均方值的定义为

$$\psi_x^2 = \frac{1}{T} \int_0^T x^2(t) dt$$

或

$$\hat{\psi}_x^2 = \frac{1}{N} \sum_{i=1}^N x_i^2 \quad (2.16)$$

均方根值的定义为

$$\psi_x = \left[\frac{1}{T} \int_0^T x^2(t) dt \right]^{\frac{1}{2}}$$

或

$$\hat{\psi}_x = \left[\frac{1}{N} \sum_{i=1}^N x_i^2 \right]^{\frac{1}{2}} \quad (2.17)$$

均方值和均方根植都是描述动态信号强度的指标。幅值的平方具有能量的含义, 因此均方值表示了单位时间内的平均功率。尽管并非所有信号的均方值都有功率量纲, 但在信号分析中仍形象地称为信号功率。而信号的均方根植由于有幅值的量纲, 在工程中又称为有效值。

此外还有方根幅值指标, 定义为

$$x_r = \left[\frac{1}{T} \int_0^T |x(t)|^{\frac{1}{2}} dt \right]^2$$

或

$$\hat{x}_r = \left[\frac{1}{N} \sum_{i=1}^N |x_i|^{\frac{1}{2}} \right]^2 \quad (2.18)$$

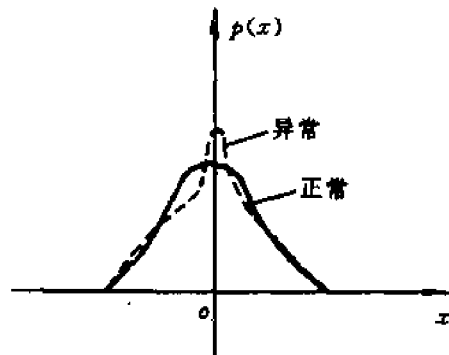


图 2-6 滚动轴承振动信号的概率密度

(4) 方差:方差的定义为

$$\sigma_x^2 = \frac{1}{T} \int_0^T [x(t) - \mu_x]^2 dt$$

或
$$\hat{\sigma}_x^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \hat{\mu}_x)^2 \quad (2.19)$$

方差描述信号偏离中心趋势的波动强度,因此它是信号的波动分量。方差的正平方根称为标准差 σ_x ,当信号的均值为零时,标准差就是有效值。

以上介绍的四种指标均属有量纲的动态指标,通常被简易仪器的测量所采用,并以此来判断设备是否出现异常及其严重程度。但这些指标的数值与测量仪器的频率范围有关,不同频率范围的仪器所测的值不完全相同,频率范围宽,阻频带衰减特性缓慢的仪器读数偏高。

(5) 偏斜度和峭度:两者的数值可以如下确定

偏斜度指标
$$a_3 = \frac{1}{T} \int_0^T [x(t) - \mu_x]^3 dt$$

或
$$\hat{a}_3 = \frac{1}{N} \sum_{i=1}^N (x_i - \hat{\mu}_x)^3 \quad (2.20)$$

峭度指标
$$a_4 = \frac{1}{T} \int_0^T x^4(t) dt$$

或
$$\hat{a}_4 = \frac{1}{N} \sum_{i=1}^N x_i^4 \quad (2.21)$$

偏斜度和峭度用于和正态分布曲线进行比较,两者分别反映信号概率分布的中心不对称程度和概率密度函数峰顶的凸平度。比值 $R_3 = a_3/\sigma_x^3$ 称为偏度,此处 σ_x 为标准差。偏度也是概率密度函数不对称性程度的度量。图 2-7 列举了 $R_3 < 0$, $R_3 = 0$ 和 $R_3 > 0$ 三种不同情况下概率密度函数。

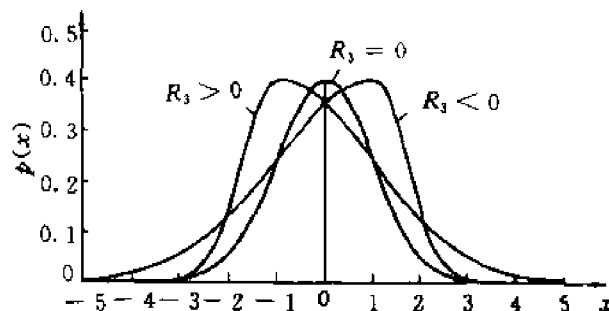


图 2-7 偏度与概率密度函数的对应关系

比值 $R_3 = a_3/\sigma_x^3$ 称为陡度,是概率密度分布陡峭程度的度量。对于正态分布来说,其偏度等于零,陡度等于 1;对于一般的实际信号来说,偏度接近于零,陡度基本等于 1。陡度 R_4 存在三种情况,它们分别对应图 2-8 所示的概率分布函数。峭度对信号中的冲击特性较敏感。它可以用作滚动轴承故障诊断中。如轴承圈出现裂纹,滚动元件或滚珠轴承边缘剥裂等在时域波形中都可能引起相当大的脉冲,用峭度作为故障诊断特征量是很有效的,但用于滑动轴承的故障诊断就不灵敏了。

在旋转机械振动监测和故障诊断中,对波形复杂的振动信号,常采用其峰峰值,即最大峰值与其相邻的最低谷值之间的幅值作为振动大小的特征量,称为振动的“通频幅值”。

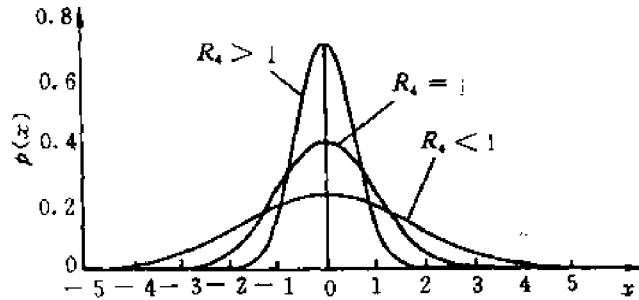


图 2-8 陡度与概率密度函数的对应关系

利用系统中某些特征点振动响应的均方根值作为故障诊断的判断依据是最简单、最常用的一种方法。均方根值诊断法多适用作稳态振动的情况，当机器振动不平稳，振动响应随时间变化时，可用振幅—时间图诊断法。该方法在研究系统的过渡过程（开机和停机）中是有效的，根据曲线的变化可判断系统的状态和故障。

当时间信号中包含的信息不是来自一个零件或部件，而是属于多个元件时，例如在多级齿轮的振动信号中往往包含有来自高速齿轮、低速齿轮以及轴承等部件的信息，在这种情况下，可利用下列的一些无量纲动态指标进行故障诊断或趋势分析。

$$\begin{array}{ll}
 \text{波形指标} & K = \frac{\psi_x}{\mu_x} \\
 \text{峰值指标} & G = \frac{x_p}{\psi_x} \\
 \text{脉冲指标} & I = \frac{x_p}{\mu_x} \\
 \text{裕度指标} & L = \frac{x_p}{x_r}
 \end{array} \quad (2.22)$$

对于这些无量纲指标的基本要求是：

(1) 对故障和缺陷足够敏感。

(2) 对信号的幅值和频率变化不敏感，即与机器运行的工况无关，只依赖于幅值分布的形状。当机器连续运行后质量下降时，例如机器中运动副的游隙增加，齿轮或滚动轴承的撞击增加，相应的振动信号中的冲击脉冲增多，幅值分布的形状也随之作缓慢的变化。分析结果证实，波形指标 K 和峰值指标 G 对于冲击脉冲的多少和幅值分布形状的变化不够敏感，而裕度指标 L 和脉冲指标 I 则能够识别上述的变化，从而可以在机器的振动、噪声信号诊断中加以应用。

在选择上述各动态指标时，按其诊断能力由大到小顺序排列，大体上为峭度 → 裕度指标 → 脉冲指标 → 峰值指标 → 波形指标。

图 2-9 所示是一轴承外圈在工作到 21 小时出现损伤以后，峭度和峰值指标的变化趋势。由图可见，当轴承正常工作时，两者都接近于 3，当出现损伤时，峭度的变化趋势非常明显，其值可达 13，这是因为信号中脉冲成分比较明显的缘故。而峰值指标比起峭度变化来就不够明显。

2.2.2 时域同步平均法

时域同步平均法是在混有噪声干扰的信号中提取周期性分量的有效方法，也称相干检波法。

当随机信号中包含有确定性的周期信号时，如果截取信号的采样时间等于周期性信号的周期 T ，将所截得的信号叠加平均，就能将该周期信号从随机信号、非周期信号以及与指定周

期 T 不一致的其他周期信号中分离出来,而保留指定的周期分量及其高频谐波分量,提高欲研究周期信号的信噪比。即使该周期信号较弱也可分离出来,这是谱分析法所不及的,这就是时域同步平均法的基本思路。如果事先不知道周期信号的周期,可通过相关分析来确定信号的周期。对于旋转机械,截取的周期应和机器运行的转动周期同步起来。例如转一圈采一帧(或整转几圈采一帧),如此循环采集若干帧信号进行平均,故该方法称时域同步平均法。

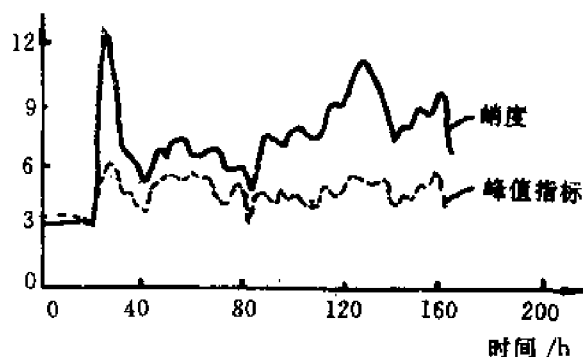


图 2-9 轴承外圈损伤时峭度和峰值指标的比较

如果一信号 $x(t)$ 由周期信号 $y(t)$ 和白噪声 $n(t)$ 组成: $x(t) = y(t) + n(t)$, 我们以 $y(t)$ 的周期去截取信号 $x(t)$, 共截得 N 段, 然后将各段对应点相加, 由于白噪声的不相关性, 可得到

$$x(t_i) = Ny(t_i) + \sqrt{N}n(t_i) \quad (2.23)$$

再对 $x(t_i)$ 平均, 便得到输出信号 $y_1(t_i)$

$$y_1(t_i) = y(t_i) + \frac{n(t_i)}{\sqrt{N}} \quad (2.24)$$

此时输出的白噪声是原来输入信号 $x(t)$ 中的白噪声的 $1/\sqrt{N}$, 因此信噪比将提高。

图 2-10 所示是截取不同的段数 N , 进行时域同步平均的效果。由图可见, 虽然原信号 ($N=1$) 的信噪比 SNR 很低 ($\text{SNR} = 0.5$), 但经过多段平均后, 信噪比大大提高。由图可见, 当 $N=256$ 时, 可以得到几乎接近于理想的正弦信号。而原始信号中的正弦分量, 几乎完全被其他信号和随机噪声所淹没。

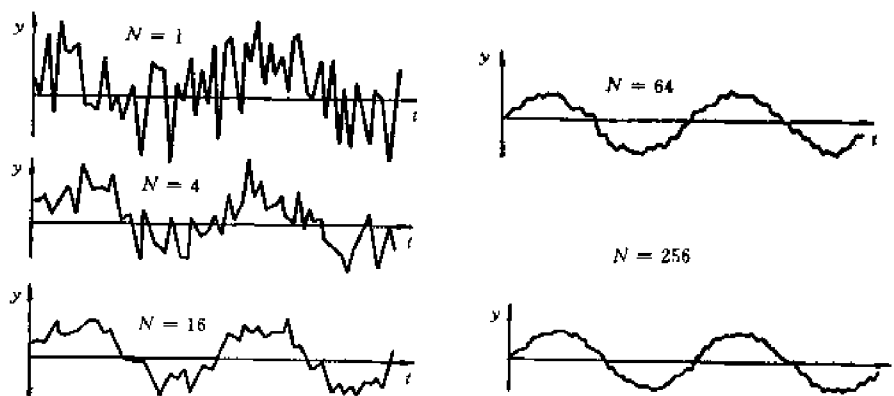


图 2-10 用时域同步平均法提高信噪比

时域同步平均的基本原理如图 2-11 所示, 系统需摄取一个输入信号和一个时标信号。例如在齿轮故障诊断时, 时标可以将某一齿轮轴的一整转定为脉冲周期 T , 乘以一定的传动比

后,化成指定的周期 T' ,输入信号即可以此周期分段采样再迭加平均,并经平滑化输出。因此,时域同步平均法和谱分析不同,后者只需摄取一个输入信号,而前者除振动信号外,还要摄取时标信号。其次,时域同步平均和谱分析方法的差异还在于:谱分析提供了各个频带内的功率,其大小主要取决于该频带内能量最大的振源,谱分析不能略去任何输入信号分量,因而,待检齿轮的信号可能完全淹没在噪声之中;而时域同步平均法可以消除与给定周期无关的全部信号分量,因此可以在噪声环境下工作。

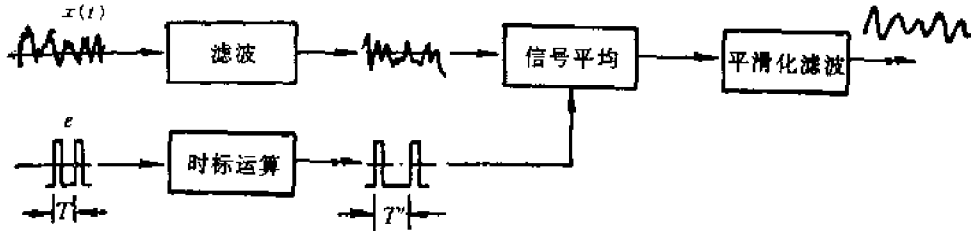


图 2-11 时域同步平均的原理

对于回转机械或往复机械运行中所产生的带有周期重复性的机械物理信号 $x(t)$,以 T_s 为时间间隔进行离散采样,得到数值 $x(nT_s), n = 0, 1, 2, 3, \dots$ 。采用时域同步平均的方法提取周期分量 $y(t)$,可以看作是用鸡冠滤波器(comb filter)进行滤波的过程。设 N 为迭加、平均的周期总数, r 为迭加循环数, M 为一个周期中的采样数目, T_s 为采样间隔, $x(nT_s)$ 为滤波器的输入,则滤波器的输出

$$y(nT_s) = \frac{1}{N} \sum_{r=0}^{N-1} x(n - rM)T_s, n = (n-1)M, (N-1)M + 1, \dots, NM - 1 \quad (2.25)$$

对式(2.25)进行 z 变换,设 $y(nT_s), x(nT_s)$ 的 z 变换为 $Y(z), X(z)$,有

$$Y(z) = \frac{X(z)}{N} \sum_{r=0}^{N-1} z^{-rM} = \frac{X(z)}{N} \cdot \frac{1 - z^{-MN}}{1 - z^{-M}} \quad (2.26)$$

滤波器的传递函数 $H(z)$

$$H(z) = \frac{Y(z)}{X(z)} = \frac{1}{N} \cdot \frac{1 - z^{-MN}}{1 - z^{-M}} \quad (2.27)$$

令 $z = e^{j\omega T_s}$,且有周期 $T = MT_s = 1/f_0 = 2\pi/\omega_0$,传递函数 $H(z)$ 的幅频特性为

$$\begin{aligned} |H(\omega)| &= \frac{1}{N} \left| \frac{1 - e^{-jMN\omega T_s}}{1 - e^{-jM\omega T_s}} \right| = \frac{1}{N} \left| \frac{1 - e^{-j2\pi N\omega/\omega_0}}{1 - e^{-j2\pi\omega/\omega_0}} \right| = \\ &= \frac{1}{N} \left| \frac{e^{-j\pi N\omega/\omega_0} (e^{j\pi N\omega/\omega_0} - e^{-j\pi N\omega/\omega_0})}{e^{-j\pi\omega/\omega_0} (e^{j\pi\omega/\omega_0} - e^{-j\pi\omega/\omega_0})} \right| = \\ &= \frac{1}{N} \left| \frac{(e^{j\pi N\omega/\omega_0} - e^{-j\pi N\omega/\omega_0})/2j}{(e^{j\pi\omega/\omega_0} - e^{-j\pi\omega/\omega_0})/2j} \right| = \frac{1}{N} \left| \frac{\sin(\pi N\omega/\omega_0)}{\sin(\pi\omega/\omega_0)} \right| \end{aligned} \quad (2.28)$$

式中, ω 为信号的角频率, ω_0 为基频。传递函数 $H(z)$ 的相频特性

$$\begin{aligned} \varphi(\omega) &= \arg H(\omega) = \arg \frac{1}{N} \left(\frac{1 - e^{-j2\pi N\omega/\omega_0}}{1 - e^{-j2\pi\omega/\omega_0}} \right) = \\ &= \arg \frac{e^{-j\pi N\omega/\omega_0} \cdot \sin(\pi N\omega/\omega_0)}{N e^{-j\pi\omega/\omega_0} \cdot \sin(\pi\omega/\omega_0)} = \arg \frac{e^{-j\pi N\omega/\omega_0}}{e^{-j\pi\omega/\omega_0}} = \pi(N-1) \frac{\omega}{\omega_0} \end{aligned} \quad (2.29)$$

图 2-12(a) 是时域同步平均过程当 $N = 4$ 的示意图,图 2-12(b) 是由(2.28)式得出的

鸡冠滤波器的幅频特性曲线。

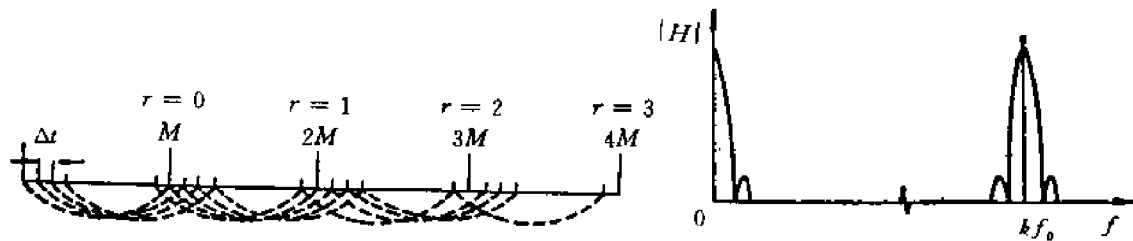


图 2-12 由机械信号中提取周期分量

(a) 当 $N = 4$ 时的时域同步平均过程

(b) 鸡冠型滤波器的频响曲线

当 $\omega/\omega_0 = f/f_0 = k$ 时,也就是当频率 f 是 f_0 的整数倍, $k = 0, 1, 2, \dots$, 为正整数,即周期 T 是各谐波的公共周期。对于这些周期性的谐波分量,鸡冠滤波器传递函数的增益 $|H(\omega)|$ 在 $\omega = k\omega_0$ 的值由罗必塔法则求得

$$|H(\omega)|_{\omega=k\omega_0} = \frac{1}{N} \left| \frac{\sin(\pi N\omega/\omega_0)}{\sin(\pi\omega/\omega_0)} \right|_{\omega=k\omega_0} = \frac{1}{N} \left| \frac{N\cos\pi Nk}{\cos\pi k} \right| = 1 \quad (2.30)$$

可见时域同步平均过程等价于具有中心频率 $\omega = k\omega_0$, 也就是 $f = kf_0$ 的鸡冠滤波器。图 2-13 是滤波器关于频率 ω/ω_0 的幅频特性 $|H(\omega)|$ 曲线的形状。由 (2.28) 式可知滤波器的极点数等于 $N-1$ (当 N 为偶数) 或 N (当 N 为奇数), 滤波器的零点数等于 N (当 N 为偶数) 或 $N-1$ (当 N 为奇数)。图 2-13 是 $N = 4$ 和 $N = 8$ 的 $|H(\omega)|$ 曲线形状。

当 $\omega/\omega_0 = k$, $k = 0, 1, 2, \dots$ 时的幅频特性 $|H(\omega)|$ 值称为滤波器主瓣峰值, 此时 $|H(\omega)| = 1$ 。滤波器有一个主瓣和若干个旁瓣, 第一个旁瓣的峰值等于 0.212, 第二个旁瓣的峰值等于 0.128, 第三个旁瓣的峰值等于 0.091, 第四个等于 0.058 等等。

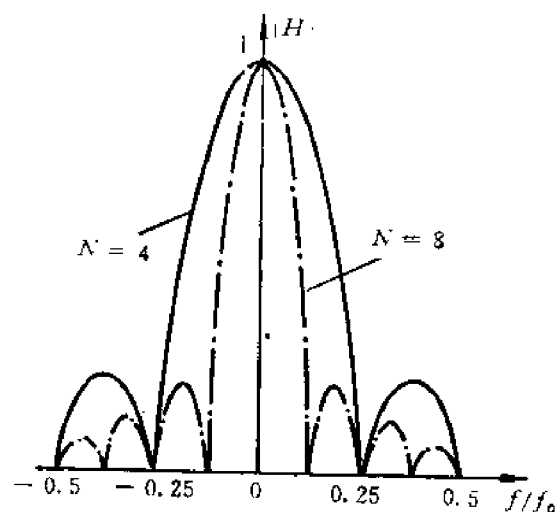


图 2-13 鸡冠滤波器幅值响应 $|H|$ 曲线形状

鸡冠滤波器在频域里抑制了白噪声 $a(t)$, 使通过滤波器后的白噪声的等价噪声带宽 (ENB) 随着 N 的增大而减小。我们知道, 输出功率谱密度 $S_y(f)$ 与输入功率谱密度 $S_x(f)$ 之比等于滤波器传递函数 $H(f)$ 模的平方

$$\frac{S_y(f)}{S_x(f)} = |H(f)|^2 \quad (2.31)$$

通过滤波器后噪声能量的改变, 等价噪声带宽

$$\text{ENB} = \int_{-f_N}^{f_N} \frac{S_y(f)}{S_x(f)} df = \int_{-f_N}^{f_N} |H(f)|^2 df \quad (2.32)$$

由于时域同步平均时, 截取信号段的周期为 T , 即频率为 f_0 , 所以 $f_N = f_0/2$, $f/f_0 = \omega/\omega_c$ 的范围为 $[-0.5, 0.5]$, 并令 $u = \omega/\omega_0$, 将(2.28)式代入(2.32)式积分

$$\begin{aligned} \text{ENB} &= \int_{-0.5}^{0.5} \left| \frac{1}{N} \frac{\sin \pi N u}{\sin \pi u} \right|^2 du = \frac{1}{N^2 \pi} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \left(\frac{\sin N x}{\sin x} \right)^2 dx = \\ &= \frac{1}{N^2 \pi} N \pi = \frac{1}{N} \quad (x = \pi u) \end{aligned} \quad (2.33)$$

通过时域同步平均, 从方差意义上讲, 使平均后的信噪比缩小了 N 倍, 相当于输出的噪声能量是输入噪声能量的 $1/N$ 。因此鸡冠滤波器抑制了白噪声, 提高了信噪比。

我们选择了基频 f_0 ($f_0 = 1/MT$), 就是将滤波器的中心频率放在 $f_k = k/MT$ ($k = 0, 1, 2, \dots$) 上, 如图 2-12(b) 所示, 这样对于基频的谐波 $k f_0$ 就能有效地提取出来。

图 2-14 所示是正常齿轮和带缺陷齿轮一整转的时域同步平均信号, 由图可见, 这些缺陷在时域同步平均信号中极易加以区别。

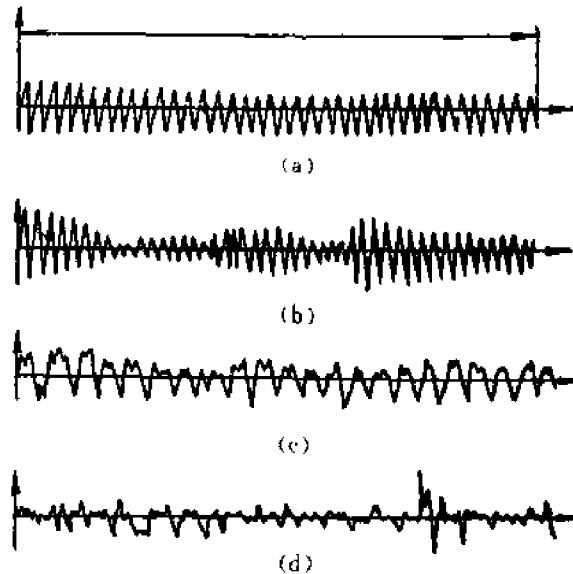


图 2-14 用时域平均法识别齿轮缺陷

(a) 正常齿轮 (b) 齿向安装不平行 (c) 齿面严重磨损 (d) 齿面有剥落

2.2.3 相关函数诊断法

相关函数及其分析是振动信号在时延域上的描述, 在系统的振源识别和故障诊断中有着广泛的应用。

1. 相关分析的基本概念

观测图 2-15 所示 4 个信号, $x_2(t), y_1(t), y_2(t)$ 两两之间都很相似, 但 $x_1(t)$ 与 $x_2(t), y_1(t), y_2(t)$ 中的任何一个都不相似。如果要问 $x_2(t)$ 和 $y_1(t), y_2(t)$ 中的哪个更相似, 用目视

观察的方法直接进行比较很难得出结论,所以,这种直接观测方法虽然比较明显,但比较粗糙且没有数值定量,因此我们用波形相似性来定量比较、分析信号之间的相似程度。

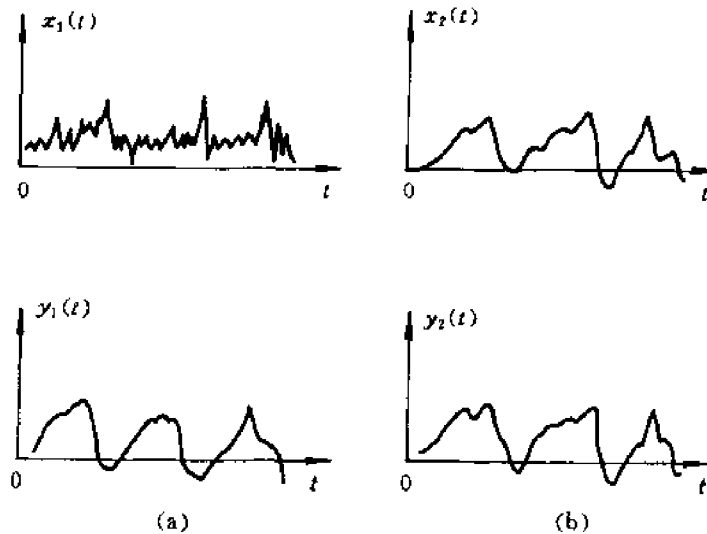


图 2-15 四种图线比较其相似性

设两个信号 $x(t)$ 和 $y(t)$,如图 2-16 所示。其离散值分别为 x_1, x_2, \dots, x_N 和 y_1, y_2, \dots, y_N ,则两者的统计均方差为

$$\beta = \frac{1}{N} \sum_{i=1}^N (x_i - y_i)^2 \quad (2.34)$$

此式的意义表示两个信号的相似的程度,也就是说 β 的数值大表示两个信号差别大而不相似, β 的数值小表示两个信号差别不大而相似。将上式展开可得

$$\beta = \frac{1}{N} \sum_{i=1}^N x_i^2 + \frac{1}{N} \sum_{i=1}^N y_i^2 - \frac{2}{N} \sum_{i=1}^N x_i y_i \quad (2.35)$$

式(2.35)的前两项表示信号的均方值,即表示总能量,在各态历经平稳随机过程中随机信号的总能量为常数。所以两个信号的相似程度完全取决于第三项的大小,记为

$$R_{xy} = \frac{1}{y} \sum_{i=1}^N x_i y_i \quad (2.36)$$

当 R_{xy} 的数值大,则 β 就小,其意义表示两个信号相似性较好,反之则相似性差。

我们也可以分析两个信号或同一信号在不同时刻的相似性,如图 2-17 所示。如果它们是各态历经和平稳的,那么相似性就与时间的起点选择无关,仅仅与分析的时间间隔有关,设时间间隔为 τ ,那么信号在不同时刻的相似性可用下列 τ 的函数来描述

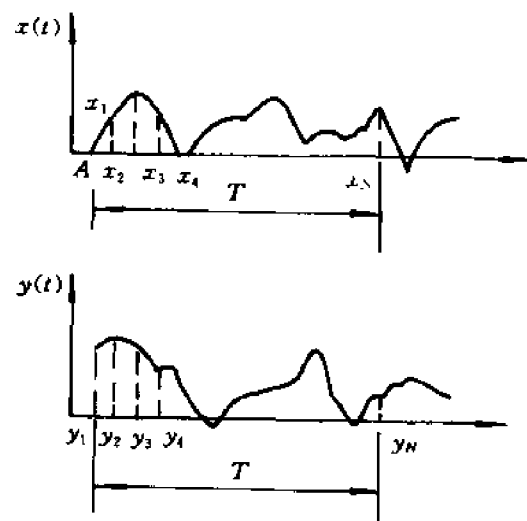


图 2-16 两个波形的离散值

两个信号比较

$$R_x(\tau) = \frac{1}{N} \sum_{i=1}^N x_i x_{i+\tau} \quad (2.37)$$

$$R_{xy}(\tau) = \frac{1}{N} \sum_{i=1}^N x_i y_{i+\tau} \quad (2.38)$$

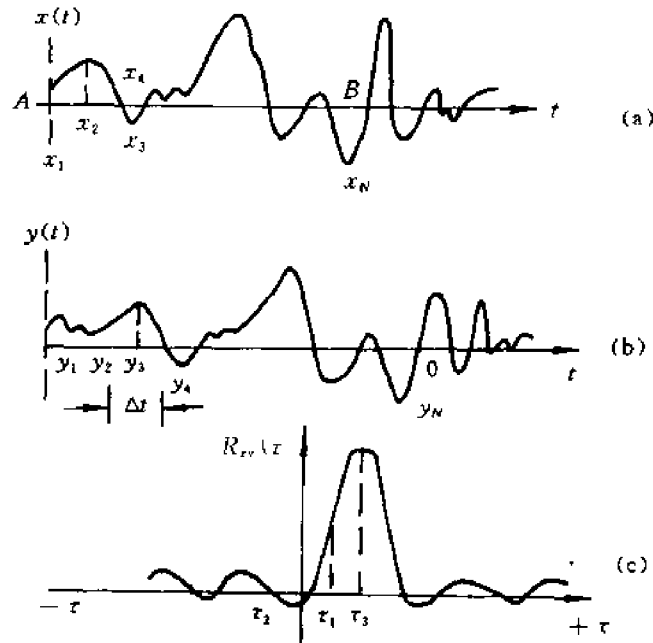


图 2-17 计算相关函数方法

2. 相关函数的定义和性质

由上述讨论可知, $R_{xy}(\tau)$ 值可以定量地衡量两个信号之间的相似程度, $R_{xy}(\tau)$ 不仅与两个信号本身的特点有关, 还与两个信号之间的相对移动值 τ 有关。由此可见, $R_{xy}(\tau)$ 的物理意义是它全面地描述了两个信号之间的相似性, 所以称 $R_{xy}(\tau)$ 为相关函数。

(1) 自相关函数的定义和性质: 信号 $x(t)$ 的自相关函数是描述信号在一个时刻的取值和另一个时刻取值之间的相似关系, 可用下式定义

$$R_x(\tau) = \frac{1}{T} \int_0^T x(t)x(t+\tau)dt$$

或

$$R_x(k) = \frac{1}{N} \sum_{i=1}^N x_i x_{i+k} \quad (2.39)$$

式中, T, N 为信号观测时间; τ, k 为时间间隔。

自相关函数 $R_x(\tau)$ 除了描述信号在不同时刻的相似性外, 它还从另一方面反映信号幅值变化剧烈的程度。如果时间间隔 τ 很小时信号幅值之间的差异就很大, 则信号的变化很剧烈, 自相关函数 $R_x(\tau)$ 值就小; 反之, 即使时间间隔 τ 很大时信号幅值仍很接近, 则信号的变化比较缓慢, $R_x(\tau)$ 值就较大。

自相关函数具有以下主要性质:

- ① 自相关函数 $R_x(\tau)$ 是偶函数, 即 $R_x(\tau) = R_x(-\tau)$;
- ② 当 $\tau = 0$ 时, 自相关函数 $R_x(0)$ 等于信号的均方值, 即 $R_x(0) = \psi_x^2$;

- ③ 当 $\tau \neq 0$ 时, 自相关函数 $R_x(\tau)$ 的值总是小于 $R_x(0)$, 即 $R_x(\tau) < R_x(0)$;
 ④ 当时间间隔 τ 足够大, 自相关函数 $R_x(\tau)$ 接近于信号的均值函数的平方, 即

$$\lim_{\tau \rightarrow \infty} R_x(\tau) = R_x(\infty) = \mu_x^2$$

- ⑤ 周期信号的自相关函数仍是周期信号, 两者周期相同, 但不反映相位信息。

以上性质的证明读者可自己完成。

例 2-1 求 $x(t) = A\sin(\omega t + \theta)$ 的自相关函数, 其中 A 和 ω 为常数, 而 θ 为在 $0 \sim 2\pi$ 范围内均匀分布的随机变量。

【解】
$$R_x(\tau) = \frac{1}{T} \int_0^T A\sin(\omega t + \theta) A\sin[\omega(t + \tau) + \theta] dt$$

式中 $T = 2\pi/\omega$ 为正弦函数的周期, 令 $\omega t + \theta = \varphi$, 则 $dt = d\varphi/\omega$

于是
$$R_x(\tau) = \frac{A^2}{2\pi} \int_0^{2\pi} \sin\varphi \sin(\varphi + \omega\tau) d\varphi = \frac{A^2}{2} \cos\omega\tau$$

可见正弦函数的自相关函数是一个同频率的余弦函数, 在 $\tau = 0$ 时具有最大值, 但原信号中的相位信息消失了。

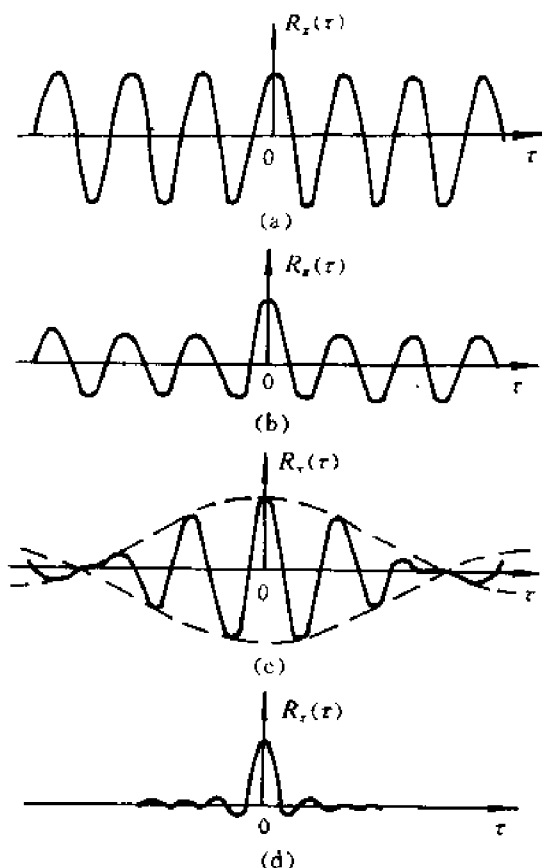


图 2-18 常见的时间波形的自相关函数

- (a) 正弦波 (b) 正弦波加随机噪声
 (c) 窄带随机噪声 (d) 宽带随机噪声

图 2-18 为几种典型信号的自相关函数(设信号的均值为零), 其中图 2-18(a) 为相位随机正弦信号的自相关函数。图 2-18(c) 为窄带噪声的自相关函数, 其表达式 $R_x(\tau) =$

$kB\sin(2\pi B\tau)/2\pi B\tau$, k 为系数, B 为信号的带宽。当 B 增加时, 波形变窄; 当 $B \rightarrow \infty$ 时, $R_x(\tau)$ 成了 $\delta(\tau)$ 即宽带噪声的自相关函数, 如图 2-18(d) 所示。图 2-18(b) 为正弦信号叠加随机噪声的自相关函数, 由图可知, 它是图 2-18(a) 和图 2-18(c) 叠加而成。可以看到, 当 τ 较大时, 随机噪声部分的自相关函数已衰减掉, 剩下周期信号的自相关函数。利用这个性质, 可以用较大的时延 τ 计算信号的自相关函数, 以抑制噪声的影响, 从而将周期性成分检测出来, 这是自相关函数的重要应用。

例如, 用噪声诊断机器设备状态时, 正常状态的机器噪声是大量的、无序的、大小接近的随机冲击的结果, 所以具有较宽而均匀的频谱。当机器运行状态不正常时, 在随机噪声中将出现有规则的、周期性的脉冲, 其大小要比随机冲击大的多。如当机器中轴承磨损而使间隙增大时, 轴与轴承间就会有撞击的现象; 如果滚动轴承的滚道出现剥落、齿轮传动中某个齿面严重磨损等情况出现时, 在随机噪声中都会出现周期信号。因此, 用噪声诊断机器设备故障时首先要在噪声中查出隐藏的周期分量, 特别是在故障发生初期, 周期信号不明显、直观难以发现的时候, 可以采用自相关分析方法, 依靠 $R_x(\tau)$ 的幅值和波动的频率查出机器缺陷之所在。

图 2-19 所示是拖拉机变速箱噪声的自相关曲线。图 2-19(a) 是正常状态下的变速箱噪声自相关函数, 当 $\tau = 0$ 时, $R_x(\tau)$ 有一峰值, 随着 τ 的增大, 自相关函数迅速趋近于零, 这说明变速箱噪声是随机噪声。在图 2-19(b) 中, 变速箱的随机噪声中夹杂有周期振动, 当 τ 值增大时, 自相关函数并不向横坐标衰减, 这种情况标志着运行状态不正常。

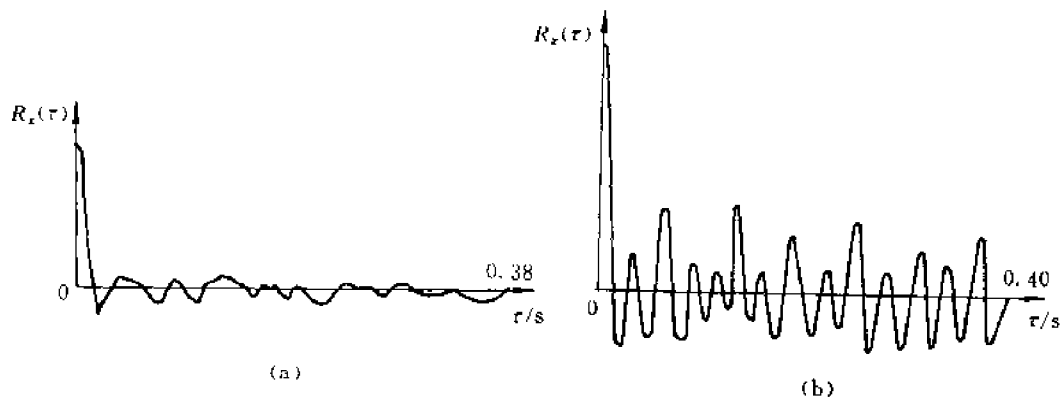


图 2-19 拖拉机变速箱噪声的自相关函数
(a) 正常状态噪声自相关函数 (b) 异常状态噪声自相关函数

(2) 互相关函数的定义和性质: 互相关函数是描述两个信号之间的相似关系, 可定义为

$$R_{xy}(\tau) = \frac{1}{T} \int_0^T x(t)y(t+\tau)dt$$

或

$$R_{xy}(k) = \frac{1}{N} \sum_{i=1}^N x_i y_{i+k} \quad (2.40)$$

互相关函数有以下一些基本性质:

① 互相关函数的峰值不一定在 $\tau = 0$, 峰值点偏离原点的距离表示两信号取得最大相关程度的时移 τ , 如图 2-20(a) 所示。

② $R_{xy}(\tau) = R_{yx}(-\tau)$, 互相关函数是一非奇非偶的实函数, 具有反对称性, 当 $x(t)$ 和 $y(t)$ 取值互换时, 则互相关函数的图形以纵坐标成镜像对称, 如图 2-20(b) 所示。

$$\textcircled{3} \lim_{\tau \rightarrow \infty} R_{xy}(\tau) = R_{xy}(0) = \mu_x \mu_y$$

④ 周期信号的互相关函数也是同频率的周期信号,且保留了原两信号的相位差信息。

从互相关函数的性质可见,互相关函数比自相关函数含有更多的信号信息,因此应用也更广,如测量滞后时间及信号的平均传播速度、确定信号的传递通道、检测被噪声污染了的信号等等。

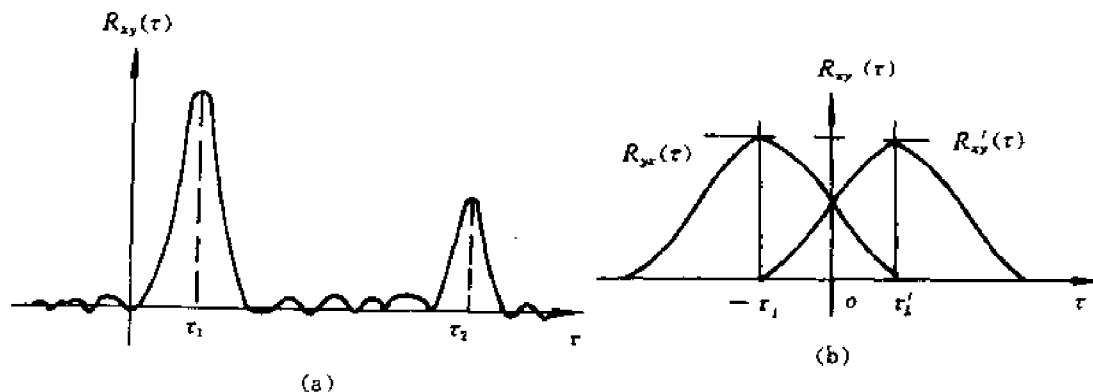


图 2-20 互相关函数的图形

(3) 相关系数:由于信号是有物理单位的函数,因此自相关函数也是一个有单位的函数。不同信号的自相关程度很难相互比较,所以实际处理时用相关系数来表达。自相关系数为

$$\rho_x(\tau) = \frac{R_x(\tau)}{R_x(0)} \quad |\rho_x(\tau)| \leq 1 \quad (2.41)$$

类似互相关系数为

$$\rho_{xy}(\tau) = \frac{R_{xy}(\tau)}{\sqrt{\sigma_x^2 \sigma_y^2}} \quad |\rho_{xy}(\tau)| \leq 1 \quad (2.42)$$

相关系数是一个无量纲的物理量,因不必知道测试系统的标定系数和其物理单位,故在信号分析中得到广泛的应用。

3. 相关分析的应用举例

相关分析为解决设备状态监测和故障诊断中的技术问题提供了不少信息,因此有着广泛的应用,下面举几个例子。

例 2-2 利用互相关函数测量滞后时间,确定深埋在地下的输油管漏损位置。

【解】 图 2-21 上漏损处 K 视为向两侧传播的声源,在管道上方分别放置传感器 1 和 2,因为传感器的位置与漏处位置间的距离不等,则漏油的音响传至传感器就有时差,将两传感器测得的音响信号 $x_1(t)$ 和 $x_2(t)$ 进行互相关分析,在互相关函数图上得出 $\tau = \tau_m$ 处有最大值,则 τ_m 就是上述时间差。由 τ_m 值即可定出漏损处位置

$$S = \frac{1}{2} v \tau_m$$

式中, S 为两传感器中点至漏损处的距离; v 为音响通过管道传播速度。

例 2-3 图 2-22 给出了用互相关函数诊断汽车驾驶员座椅上的振动源。座椅上的振动信号为 $y(t)$,前轮轴梁和后轮轴梁上的振动信号分别为 $x(t)$ 和 $z(t)$ 。分别求 $R_{xy}(\tau)$ 与 $R_{zy}(\tau)$ 。从图上可看出座椅的振动主要是由于前轮振动而引起。

例 2-4 图 2-23 给出了某机器中的 6306 轴承在不同状态下的振动加速度信号的自相关

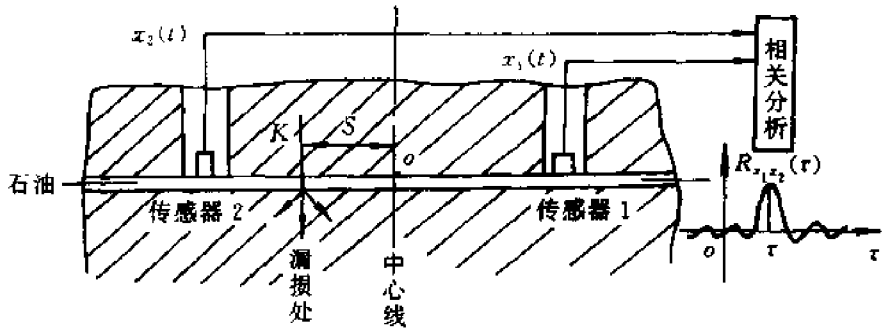


图 2-21 地下输油管道漏损位置的检测

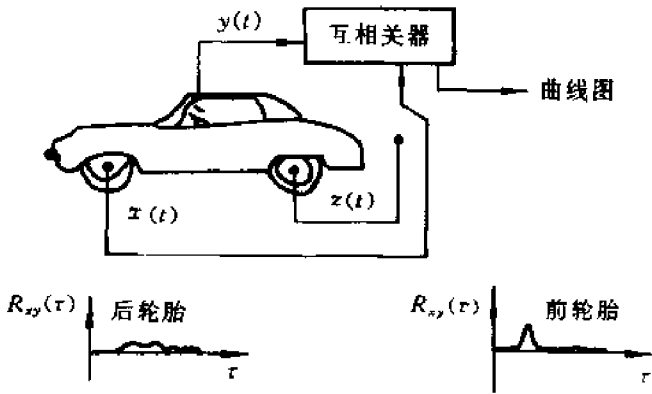


图 2-22 汽车振动信号的互相关关系

函数。图(c)为正常轴承的自相关函数图形接近于宽带随机噪声的自相关函数。在图(a)及图(b)上分别因外、内圈滚道上有疵点而在间隔为 14 ms 和 11 ms 处出现了峰值。

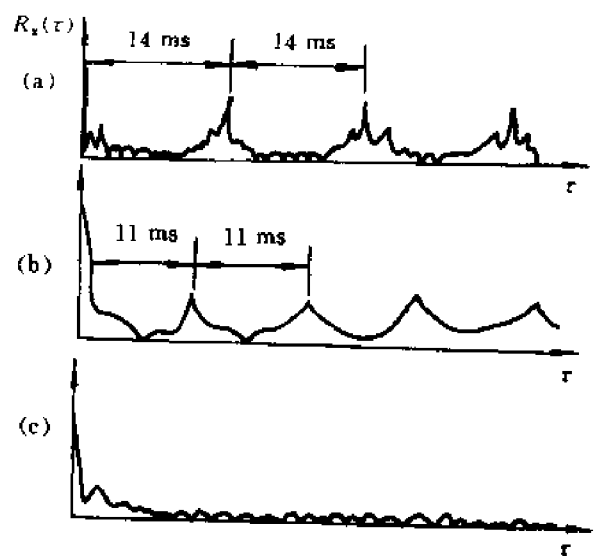


图 2-23 轴承振动信号的自相关函数图

(a) 外圈滚道上有疵点 (b) 内圈滚道上有疵点 (c) 正常轴承

2.3 检测信号的频域分析方法

2.3.1 傅里叶级数及幅值谱

让我们首先观察图 2-24 中所示的周期信号。按照傅里叶分析的原理,这一信号可以分解为许多谐波分量之和:

$$x(t) = \frac{a_0}{2} + \sum_{n=1}^{\infty} (a_n \cos n\omega_0 t + b_n \sin n\omega_0 t) \quad (2.43)$$

式中基频 $\omega_0 = 2\pi/T$, 系数 $a_n (n=0, 1, 2, \dots)$ 和 $b_n (n=1, 2, \dots)$ 由下式确定

$$a_n = \frac{2}{T} \int_0^T x(t) \cos n\omega_0 t dt$$

$$b_n = \frac{2}{T} \int_0^T x(t) \sin n\omega_0 t dt$$

将这些谐波分量投影到幅值—频率坐标平面上,就可以得到许多离散的分量,反映了不同频率分量的幅值谱。由此可见,时域分析和频域分析二者是反映了同一信号的不同侧面,它们是互相补充的。

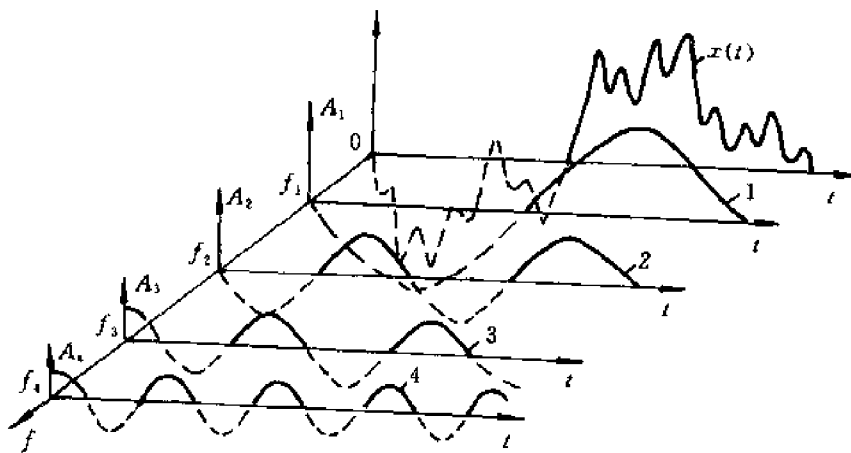


图 2-24 机械信号的时域和频域分析

1— $A_1 \cos(2\pi f_1 t + \varphi_1)$; 2— $A_2 \cos(2\pi f_2 t + \varphi_2)$; 3— $A_3 \cos(3\pi f_3 t + \varphi_3)$; 4— $A_4 \cos(2\pi f_4 t + \varphi_4)$

各频率分量的幅值和相位分别为

$$A_n = \sqrt{a_n^2 + b_n^2} \quad (2.44)$$

及

$$\varphi_n = \arctan \frac{b_n}{a_n} \quad (2.45)$$

傅里叶级数亦可以表示成如下的复数形式

$$x(t) = \sum_{n=-\infty}^{\infty} C_n e^{jn\omega_0 t} \quad (2.46)$$

及

$$C_n = \frac{1}{2} (a_n - jb_n) = \frac{1}{T} \int_0^T x(t) e^{-jn\omega_0 t} dt \quad (2.47)$$

$$\varphi_n = \arctan \frac{\text{Im}[C_n]}{\text{Re}[C_n]}$$

称 $A_n - \omega$, $|C_n| - \omega$ 关系为幅值谱, $\varphi_n - \omega$ 关系为相位谱, $A_n^2 - \omega$, $|C_n|^2 - \omega$ 关系为功率谱。

2.3.2 傅里叶变换

假设时域函数 $x(t)$ 是非周期的, 并且在实数域上满足绝对可积条件。若将 $x(t)$ 看作周期函数在 $T \rightarrow \infty$ 时的极限, 就可以将傅里叶级数的定义推广到更一般的函数, 即傅里叶积分

$$X(f) = F[x(t)] = \int_{-\infty}^{\infty} x(t) e^{-j2\pi ft} dt \quad (2.48)$$

由于 $x(t)$ 绝对可积, 上述积分一定存在。称为 $x(t)$ 的傅里叶变换(或谱)。反之, 若谱 $X(f)$ 已知, 则可由下式求 $X(f)$ 的傅里叶反变换

$$x(t) = F^{-1}[X(f)] = \int_{-\infty}^{\infty} X(f) e^{j2\pi ft} dt \quad (2.49)$$

由式(2.48)和(2.49)定义的变换对称为傅里叶变换对。对其分析可知, 与周期信号类似, 非周期信号也可以分解成许多不同频率分量的叠加, 所不同的是, 由于非周期信号的周期 $T \rightarrow \infty$, 基频 $f_0 \rightarrow df$, 所以它包含了从零到无限大的所有频率分量。各频率分量的幅值为 $X(f)df$, 这是无穷小量, 所以频谱不能再用幅值表示, 而必须用密度函数描述, 称 $|X(f)| - \omega$ 关系为幅值谱密度, $|X(f)|^2 - \omega$ 关系为功率谱密度, $\varphi_n = \arctan \frac{\text{Im}[X_n]}{\text{Re}[X_n]} - \omega$ 关系为相位谱密度。

2.3.3 离散傅里叶变换

1. 采样与混叠

将时域信号转换到频域中去进行分析, 最普遍的方法是利用数字计算机或数据处理机, 通过快速傅里叶变换进行的。为此, 首先要对连续信号进行采样, 使之成为离散信号。

采样过程是通过采样脉冲序列 $\delta_T(t)$ 与连续时间信号 $x(t)$ 相乘来完成的。如图 2-25(b) 所示的采样脉冲序列

$$\delta_T(t) = \sum_{n=-\infty}^{\infty} \delta(t - nT_s) \quad (2.50)$$

采样信号 $x_s(t) = x(t)\delta_T(t)$ (2.51)

如果 $F[x(t)] = X(\omega)$, $F[\delta_T(t)] = \Delta(\omega)$

则根据卷积定理, 有

$$X_s(\omega) = \frac{1}{2\pi} X(\omega) * \Delta(\omega) \quad (2.52)$$

又, 采样脉冲序列是一个周期函数, 所以序列的傅里叶变换为

$$\Delta(\omega) = 2\pi \sum_{n=-\infty}^{\infty} C_n \delta(\omega - n\omega_s) \quad (2.53)$$

式中, C_n 是 $\delta_T(t)$ 的傅里叶系数, $C_n = \frac{1}{T_s} \int_{-\frac{T_s}{2}}^{\frac{T_s}{2}} \delta_T(t) e^{jn\omega_s t} dt = \frac{1}{T_s}$

所以 $X_s(\omega) = \sum_{n=-\infty}^{\infty} C_n X(\omega - n\omega_s) = \frac{1}{T_s} \sum_{n=-\infty}^{\infty} X(\omega - n\omega_s)$ (2.54)

此式表明,一个连续信号经过采样以后,它的频谱将沿着频率轴每隔一个采样频率 $\omega_s(2\pi/T_s)$,重复出现一次,即频谱产生了周期延拓,其幅值被傅里叶系数 C_n 所加权,因为 $C_n = 1/T_s$,所以频谱形状不变。但由于频域信号的周期延拓,使频谱图形易发生混叠效应,如图 2-25(c) 所示。离散信号在频率区间 $[-\frac{\omega_s}{2}, \frac{\omega_s}{2}]$ 内的频谱既包含此区间内连续信号的频谱,又有其他频段的频谱混叠进来,随着 $|\omega|$ 增大,混叠进来的频段增多。那么怎样才能不产生这种现象呢?

我们考查一个余弦信号,其幅值为 1,频率为 ω ,采样间隔为 T_s ,则第 k 个样本值为 $\cos(k\omega \cdot T_s)$ 。

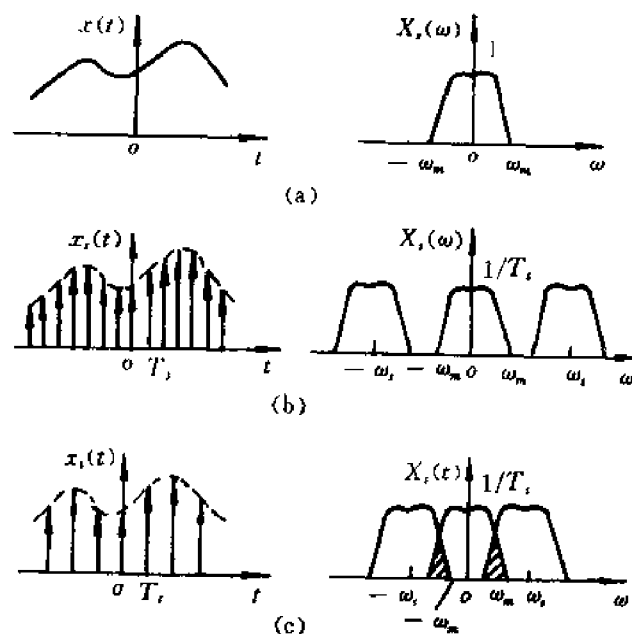


图 2-25 信号的采样频混现象

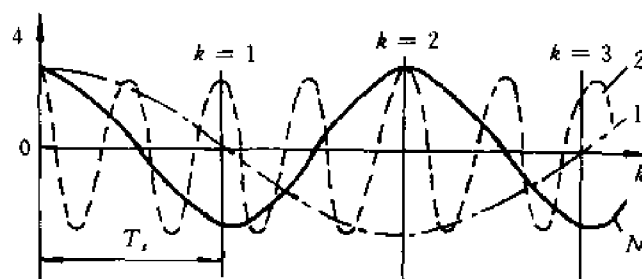


图 2-26 褶迭频率

$$1-\omega_1 < \pi/T_s, 2-\omega_2 > \pi/T_s, N-\omega_N = \pi/T_s$$

保持采样的时间间隔等于 T_s 不变,改变频率 ω 值,观察采样的情况。很明显当频率由 ω_1 增大到 ω_2 时,波形波动将越来越激烈,当 ω 增加到 $\omega_N = \pi/T_s$ 时,就是采样后能够把信号波动的频率保留下来的最迅速的波动。当 ω 进一步增大到 $\omega_2 > \pi/T_s$ 时,采样后原来的波动已完全消失,如图 2-26 中的虚线波形所示。反之,当 ω 下降到 $\omega_1 < \omega_N$ 时,采样后原来的波动仍能保持,

如图 2-26 中的点划线所示。

假定频率 ω 在 π/T_s 和 $2\pi/T_s$ 之间, 则取 $\omega' = 2\pi/T_s - \omega$,

$$x_k = \cos(k\omega \cdot T_s) = \cos[(2\pi/T_s - \omega)k T_s] = \cos(k\omega'/T_s)$$

由此可见, 这时频率 ω 和 ω' 无法区分, 互相产生混淆的现象。换言之, 在 0 到 π/T_s 区间以外的频率分量, 将褶迭到这一区间中来, 如图 2-26 所示, 这样称 $\omega_N = \pi/T_s$ 或 $f_N = 1/2T_s$ 为褶迭频率或 Nyquist 频率。

设 f_{\max} 是欲分析信号的最高频率, 则在选择采样间隔 T_s 时保证

$$T_s \leq \frac{1}{2f_{\max}} \quad (2.55)$$

就不会发生混叠现象, 这就是所谓的采样定理, 它是数字信号采集和处理中一个很重要的概念。

2. 快速傅里叶变换(FFT)

离散信号的时域与频域转换是依靠离散傅里叶变换(DFT)来实现的。设时域中的离散信号为 $x(n)$, $n = 0, 1, \dots, N-1$, 其频域变换为 $X(k)$, 则有

$$X(k) = \sum_{n=0}^{N-1} x(n)e^{-j2\pi nk/N}, k = 0, 1, \dots, N-1 \quad (2.56)$$

和
$$x(n) = \frac{1}{N} \sum_{k=0}^{N-1} X(k)e^{j2\pi nk/N}, n = 0, 1, \dots, N-1 \quad (2.57)$$

常用的离散傅里叶变换的算法是快速傅里叶变换 FFT, 它的特点是大大节约计算时间。例如, 当数据长度 $N = 1024$ 时, 计算时间可节约 100 倍以上。快速傅里叶变换 FFT 的基本原理如下:

设有一信号, 其长度为 N

$$x(1), x(2), \dots, x(N-1)$$

计算其离散傅里叶变换。首先将上述信号分解为两个信号: $g(n)$ 是 $x(n)$ 中的偶样本(假定 N 是偶数),

$$g(n) = x(2n), n = 0, 1, \dots, N/2 - 1$$

而 $q(n)$ 是 $x(n)$ 的奇样本,

$$q(n) = x(2n+1), n = 0, 1, \dots, N/2 - 1$$

假定信号 $g(n)$ 的离散傅里叶变换为 $G(k)$, 它是一个 $N/2$ 个点的变换

$$G(k) = \sum_{n=0}^{N/2-1} g(n)e^{-j2\pi nk/(N/2)}, k = 0, 1, \dots, N/2 - 1 \quad (2.58)$$

这里信号的长度变为 $N/2$, 并且, 在式(2.58)中, 如果用 $k + N/2$ 代替 k , 则式(2.58)的计算式不会改变(因为在指数中增加了 $2\pi j$ 的整数倍, 其值不会变化)。这样, $k = N/2, N/2 + 1, \dots, N - 1$ 时, $G(k)$ 是由

$$G(k + N/2) = G(k) \quad (2.59)$$

所定义的。

用同样的方法可以计算 $Q(k)$

$$Q(k) = \sum_{n=0}^{N/2-1} q(n)e^{-j2\pi nk/(N/2)}, k = 0, 1, \dots, N/2 - 1 \quad (2.60)$$

k 值同上面一样, 可以延伸到 $k = N/2, N/2 + 1, \dots, N - 1$

$$Q(k + N/2) = Q(k) \quad (2.61)$$

式(2.58)与式(2.60)共需计算 $2(N/2)^2$ 次乘法运算以得到

$$G(k) \quad k = 0, 1, \dots, N-1$$

$$Q(k) \quad k = 0, 1, \dots, N-1$$

下面可以看到 $G(k), Q(k)$ 组合起来可以求出 $X(k)$, 考虑

$$G(k) + e^{-j2\pi k/N} Q(k) = \sum_{n=0}^{N/2-1} g(n) e^{-j2\pi nk/(N/2)} + e^{-j2\pi nk/N} \sum_{n=0}^{N/2-1} q(n) e^{-j2\pi nk/(N/2)}$$

$$k = 0, 1, \dots, N-1 \quad (2.62)$$

将 $g(n) = x(2n), q(n) = x(2n+1)$ 代入式(2.62), 并将指数因子移入右端的求和符号内

$$\sum_{n=0}^{N/2-1} x(2n) e^{-j2\pi(2n)k/N} + \sum_{n=0}^{N/2-1} x(2n+1) e^{-j2\pi(2n+1)k/N} = X(k) \quad (2.63)$$

这就是说两个子变换 $G(k), Q(k)$ 可以合并得到原来的变换 $X(k)$, 这个用两个 $N/2$ 点变换组合起来, 来计算一个 N 点变换的过程, 称为数据并合过程, 并合过程需要增加附加的 N 个乘法运算, 即总共要进行 $2(N/2)^2 + N$ 次乘法运算, 这比起直接计算一个 N 点变换来要节省 $N^2 - (N^2/2 + N) = N^2/2 - N$ 次乘法运算。

$G(k)$ 和 $Q(k)$ 两个变换还可以再分别分成奇部和偶部, 经过并合过程求出 $G(k)$ 和 $Q(k)$, 这样 $G(k)$ 和 $Q(k)$ 分别需要 $[2(N^2/4) + N/2]$ 次乘法运算。为得到 $X(k)$, 就需要 $4(N^2/4) + 2N$ 次运算。如果 N 是 2 的指数, 则这一分解过程可以一直继续 $\log_2 N$ 次, 直到最后, 当 $N = 1$ 时就不需要乘法运算。

上式中乘法运算次数的 N^2 将会消失, 但每次分解为奇部和偶部时, 又将引入新的 N 次乘法运算。由于我们分解了 $\log_2 N$ 次, 故需 $N \log_2 N$ 次乘法运算。这样新的乘法运算次数仅为原来乘法运算次数的 $(N \log_2 N)/N^2 = (N \log_2 N)/N$ 。

以上是一种快速傅里叶变换的基本原理。目前, 专用的数据处理机和通用数字计算机中, 多数都配有快速傅里叶变换的标准程序。

3. 截断与泄漏

在进行数字信号处理时, 只能对有限长的离散信号进行时域、频域变换, 即必须把时域信号截断。为了将信号截取一段长度, 这时需要在时域中乘以窗函数(见图 2-27(d)), 因而引起信息损失, 使窗外的信息损失掉。设要求的样本长度为 N , 窗函数的长度为 T_0 , 则采样的数目 $N = T_0/T_s$ 。在频域中, 这个称为矩形窗函数的权函数与离散信号的乘积, 仍然变换成为频域的卷积, 如图 2-27(e) 所示。由于矩形窗函数的频域变换是 $(\sin x)/x$, 它和原来信号卷积的结果, 引起频域信号的皱纹, 能量将会从原来的频率上泄漏到两边频带, 造成频谱谱峰模糊, 甚至移位, 并使原来真正的频带稍有变宽。在极端情况下, 来自强频率分量的旁瓣可能淹没邻近单元的弱频率分量的主瓣, 这种效应在信号分析中称为强信号旁瓣抑制了弱信号主瓣。为了减少这种泄漏现象, 人们寻找了多种窗函数, 如哈宁窗函数、余弦窗函数等等。

时域信号经过以上的采样、加窗(见图 2-27(a)~(d))后, 所得到的谱(见图 2-27(e))是连续谱, 是无法计算的, 需要将谱乘以频域中的采样函数进行采样, 才能得到离散谱(见图 2-27(g))。设我们要求谱中一个谱峰间隔中采样 N 个脉冲: $f_s/f_0 = f_s \times T_0 = T_0/T_s = N$, 这样, 在时域中就成了一个限时信号(见图 2-27(e))和间距为 T_0 的脉冲二者的卷积。卷积的结果, 产生一个周期信号, 其周期为 T_0 。通过以上的步骤, 最后才得到离散傅里叶变换(见图 2-

27(g), 频域), 如果从上述周期信号中各取出其基本周期来, 就得到两个有限长的序列 $x(n)$ 和 $X(k)$ (见图 2-27(i)), 它们形成对偶的离散傅里叶变换关系式(2.56) 和式(2.57)。

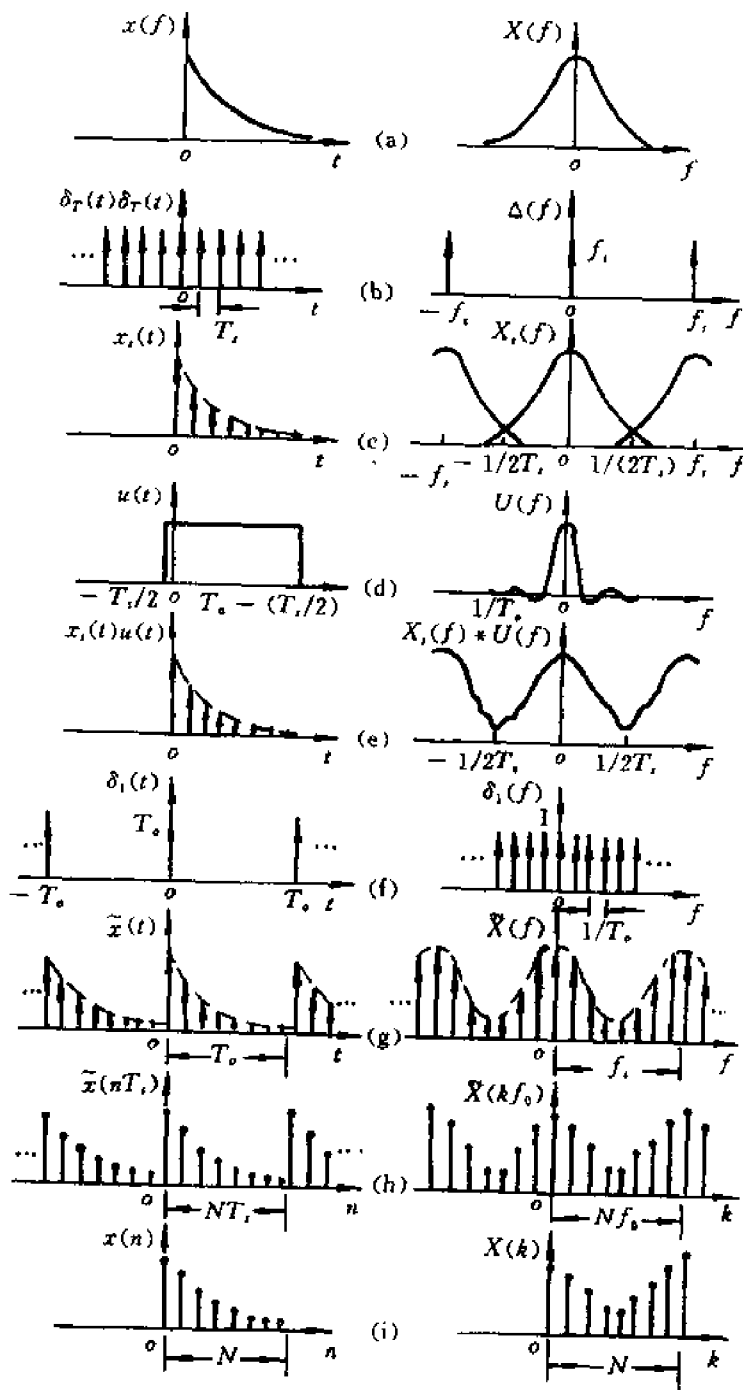


图 2-27 离散傅里叶变换的图解分析

但是, 图 2-27 所讨论的离散傅里叶变换中还忽略了这样的问题: 用自相关函数经 FFT 转换为功率谱密度函数时, 谱密度约计量的方差不会由于样本数目的增多而趋向于零。所得到的约计量将绕真值剧烈地摆动。为了使谱平滑化, 需要采用二次窗函数, 使相关函数在时域中乘以二次窗函数, 达到使谱平滑化的目的, 例如

$$\omega(t) = 0.54 + 0.46\cos\pi t$$

这是一种改进的吐克窗函数(Turkey Window)。采用二次窗函数的效果如图 2-28 所示。

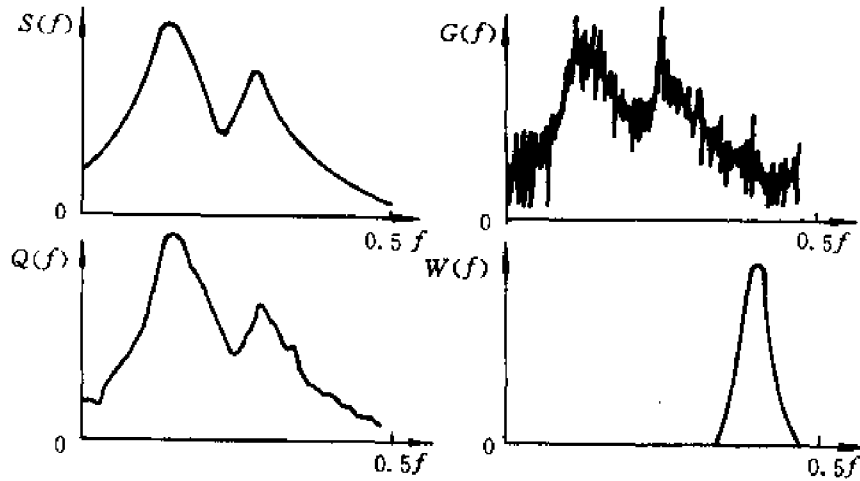


图 2-28 二次窗函数的作用

$S(f)$ —实际功率谱; $G(f)$ —周期图; $Q(f)$ —采用二次窗函数光滑化后的谱; $W(f)$ —二次窗函数

4. 细化谱分析

FFT 分析方法是在 $f_s/2$ 采样频率范围内对 $N/2$ 个采样点数进行变换,谱线间隔($f_0 = f_s/N$)决定了频率分辨能力,即 f_0 越小,谱图的分辨率越高, f_0 较大时,将由于栅栏效应而丢掉有用信息。当采样频率 f_s 选定时, f_0 值决定于采样点数 N ,可见,若要提高频率分辨率,又要求上限频率($f_s/2$)不变,则需要增加时窗长度 T_0 ,即增加采样点数,这样计算工作就要增大,特别是对专用信号处理机,一旦制成,其可处理的最大点数即已固定,并不是可随意增大的,所以既要不损失上限频率,又要增大分辨率是很困难的。故此,就引出了窄带谱的细化快速傅里叶变换分析。

窄带谱的频率细化,犹如频谱的局部放大,如图 2-29 所示,能使某些感兴趣的重点频区得到较高的分辨率。频率细化方法有很多种,下面介绍可选频带的频率细化分析方法。可选频带频率细化分析方法,又称为复调制细化分析方法,是基于复调制的高分辨率的傅里叶分析方法,简称为 ZOOM-FFT 方法。其基本思想是利用频移,将时域样本改造,使相应频谱原点移到感兴趣的中心频率处,再重新采样作 FFT,即可得到更高的频率分辨率,其运算过程如图 2-30 所示。图中:

(1) 时域信号 $x(t)$,其频谱为 $X(f)$,经抗频混滤波,滤波器截止频率为 $f_c \leq f_s/2$, f_s 为采样频率;

(2) 模拟信号经过转换,其采样序列的周期频谱为 $X_0(f)$,频率间隔为 f_0 ;

(3) 复调制,根据傅里叶变换的频移定理,在时域乘以频移因子 $e^{-j2\pi F_0 t}$,在频域有 F_0 的频移, F_0 是欲细化频段的中心,经 F_0 频移后, F_0 已成为新频谱的零点;

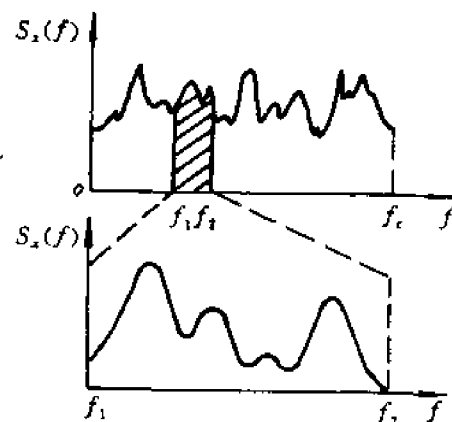


图 2-29 窄带谱的频率细化

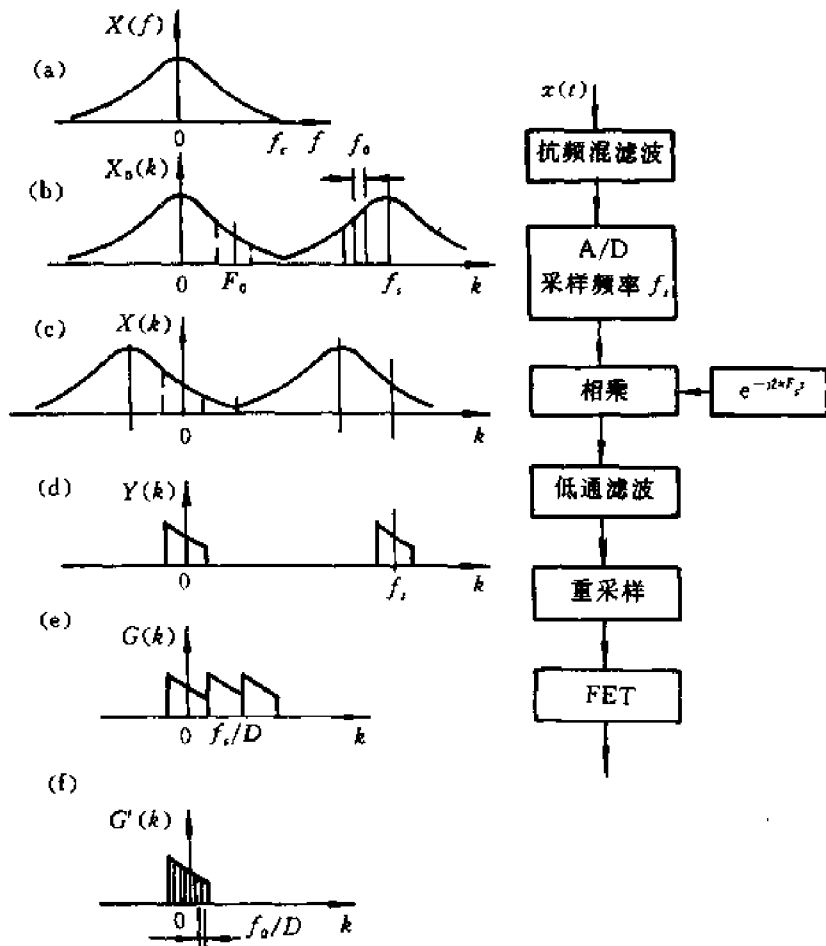


图 2-30 ZOOM-FFT 原理图

(4) 用低通数字滤波器, 将观测频带以外的高频成分滤除, 以防止采样频率降低后引起无用频带对有用频带成分的混叠;

(5) 重采样, 采样的周期为 $T_s \cdot D$, D 是细化倍数, 在频域则按 f_s/D 作周期化;

(6) FFT 处理, 对时域序列作仍为 N 点的 FFT 处理, 在频域得到 N 点谱线, 由于时域重采样时采样频率降低了 D 倍, 而采样点仍保持不变, 这就使总时间窗长增长了 D 倍, 所以使频率分辨率提高了 D 倍, 使 FFT 后在所截取的频段内得到了细化 D 倍的频谱。

ZOOM-FFT 主要适用于: 包含大量谐波的信号, 例如齿轮箱故障诊断用的振动信号; 改进弱阻尼、密模度频响函数的分析结果, 以便获得准确的谱峰及对应的频率值等场合。

2.3.4 随机信号的功率谱分析

根据巴塞伐定理, 同一个信号在时域内所包含的总功率, 应该等于频域中所包含的总功率。这样, 由式(2.43), 有

$$P = \frac{1}{T} \int_0^T x^2(t) dt = \frac{1}{T} \int_0^T \left[\frac{a_0}{2} + \sum_{n=1}^N a_n \cos n\omega_0 t + b_n \sin n\omega_0 t \right]^2 dt \quad (2.64)$$

当 $m \neq n$ 时, 下面一些关系成立

$$\int_0^T a_m a_n \cos m\omega_0 t \cos n\omega_0 t dt = 0$$

$$\int_0^T b_m b_n \sin m\omega_0 t \sin n\omega_0 t dt = 0$$

$$\int_0^T a_m a_n \cos m\omega_0 t \sin n\omega_0 t dt = 0$$

$$\int_0^T \frac{a_0}{2} \sin n\omega_0 t dt = 0$$

这时可以导出

$$P = \frac{1}{T} \int_0^T x^2(t) dt = \frac{a_0^2}{4} + \sum_{n=1}^N \frac{1}{2} (a_n^2 + b_n^2) = \frac{a_0^2}{4} + \sum_{n=1}^N \frac{A_n^2}{2} \quad (2.65)$$

其中

$$A_n^2 = a_n^2 + b_n^2$$

而 A_n^2 就是第 n 个谐波分量所包含的功率 P_n 。这样由各个谐波分量的离散功率所形成的谱称为功率谱,如图 2-31(a) 所示。图 2-31(b) 是将各个中心频率 $\omega_1, \omega_2, \dots, \omega_n$ 处的功率用矩形面积表示,将带宽 $\Delta\omega_1, \Delta\omega_2, \dots, \Delta\omega_n$ 除以面积,则纵坐标成为功率谱密度 $S_x(\omega)$ 。当 $\Delta\omega_i \rightarrow 0$ 时,功率谱密度成为连续的曲线,如图 2-31 所示。

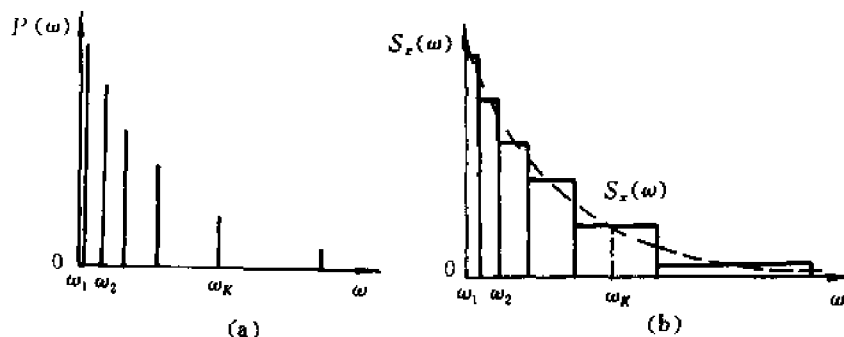


图 2-31 离散功率谱与功率谱密度曲线

(a) 离散功率谱 (b) 功率谱密度曲线

1. 自功率谱(自谱)密度函数

如同相关函数在时域分析一样,功率谱密度是频域分析的一种重要工具。

在时域中相关函数有自相关函数和互相关函数之分,在频域中功率谱密度也有自谱密度 $S_x(\omega)$ 和互谱密度 $S_{xy}(\omega)$ 之分。根据维纳-辛钦定理,自相关函数和自谱密度是一傅里叶变换对,即

$$S_x(\omega) = \int_{-\infty}^{\infty} R_x(\tau) e^{-j\omega\tau} d\tau \quad (2.66)$$

$$R_x(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} S_x(\omega) e^{j\omega\tau} d\omega \quad (2.67)$$

因为 $S_x(\omega)$ 与 $R_x(\tau)$ 都是实偶函数,可以用余弦函数代替指数函数

$$S_x(\omega) = \int_{-\infty}^{\infty} R_x(\tau) \cos\omega\tau d\tau \quad (2.68)$$

$$R_x(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} S(\omega) \cos\omega\tau d\omega \quad (2.69)$$

设 $x(t)$ 的傅里叶变换是 $X(\omega)$,由于时域和频域内同一信号的功率应该相等,因此

$$P = \frac{1}{T} \int_{-\infty}^{\infty} x^2(t) dt = \frac{1}{2\pi T} \int_{-\infty}^{\infty} |X(\omega)|^2 d\omega = \frac{1}{2\pi} \int_{-\infty}^{\infty} S_x(\omega) d\omega$$

有 $S_x(\omega) = |X(\omega)|^2 / T$ (2.70)

反映了幅值谱 $X(\omega)$ 和功率谱密度 $S_x(\omega)$ 之间的关系。由此可见,要得到一个时域信号 $x(t)$ 的功率谱密度 $S_x(\omega)$ 可以通过两个途径,一个途径是先求出自相关函数 $R_x(\tau)$,再利用式(2.66),(2.67)所示的傅里叶变换;另一个途径是先求出 $x(t)$ 的傅里叶变换幅值谱 $X(\omega)$,再由式(2.70)求出其功率谱密度。前一种方法叫相关图法,而后一种则称为周期图法。由于自谱密度不包含相位信息,所以时域信号不能由它的自功率谱密度惟一地确定,只能由它的幅值谱进行复原。

自谱密度是用得最多、最普遍的一种频域分析方法。在机械故障诊断中有着广泛的用途。图2-32所示是滚动轴承在加载运转时的振动信号,经过处理后所得到的功率谱密度。试验是在专用的轴承试验台上进行的,用来代替常规的人工验收精密滚动轴承。

由图2-32可以看到,试验台接收和处理的信号频带相当宽,从低频一直到80 kHz高频。这样,滚动轴承在正常运行下各个滚动元件、套圈、保持器所激发的高频共振信号可以全部接收和处理,为此,采用了一种特殊的高频共振技术和压电陶瓷传感器。从图上可以看到,在0到5.1 kHz之间有三个清晰的谱峰。在12 kHz,24~30 kHz处,以及68~72 kHz处均有附加的谱峰。由此可见,在功率谱图上,信息是十分丰富的,需要选择可靠和灵敏的频带作为诊断频带,并且和滚动轴承在运动中的缺陷联系起来。

采用这种方法验收精密滚动轴承,试验的结果与熟练技师人工验收的效果不相上下,有些轴承缺陷,人工验收不能发现,试验台却可以及时发现并提出警告。由于严格了验收规范,就减少了大修中更换轴承的次数,以及运行中发生事故的可能性。

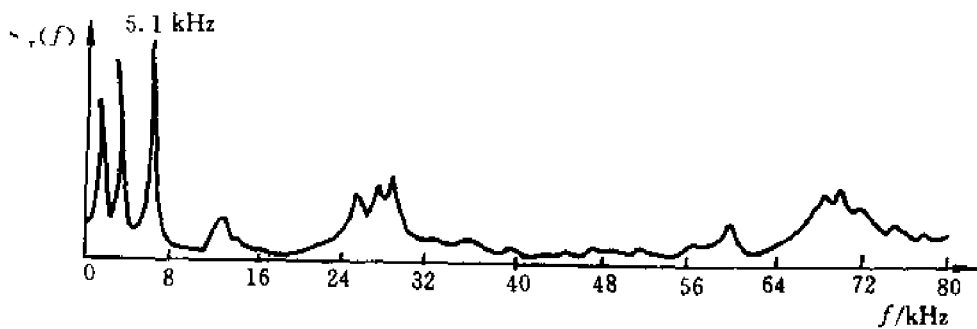


图2-32 滚动轴承振动信号的频域分析

图2-33(a)所示是拖拉机发动机的噪声功率谱密度,发动机转速为1300 r/min,处于满载运行的情况。图中四条曲线相当于活塞与缸套间隙为 $h = 0.5, 0.3, 0.2$ 和 0.1 mm的情况。图2-33(b)所示是在各个共振峰区域内信号功率的变化;在0.71~0.79 kHz频带内,虽然总的趋势与其他谱峰相同,但间隙变化的影响很小;在9 kHz的谱峰处,其功率与活塞、缸套间的间隙无关;而在1.6与3.2 kHz的谱峰处,功率受间隙的影响最大。

图2-34是发动机连杆轴承间隙变化时,安装在汽缸头上的压电加速度计所接收的振动信号频谱也随之变化。试验还证明,图中1.2 kHz处谱峰的功率随着连杆轴承间隙的增大而呈抛物线关系增加。

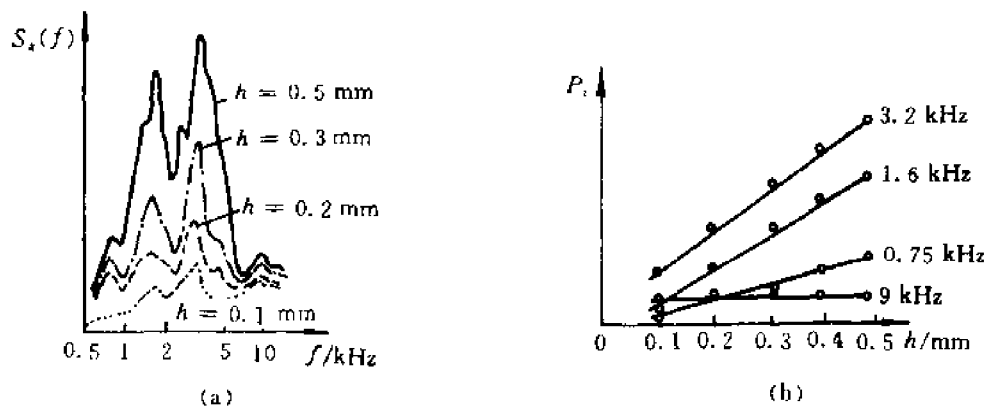


图 2-33 拖拉机发动机在不同的活塞缸套间隙下的噪声的测定

(a) 功率谱; (b) 峰值处的功率

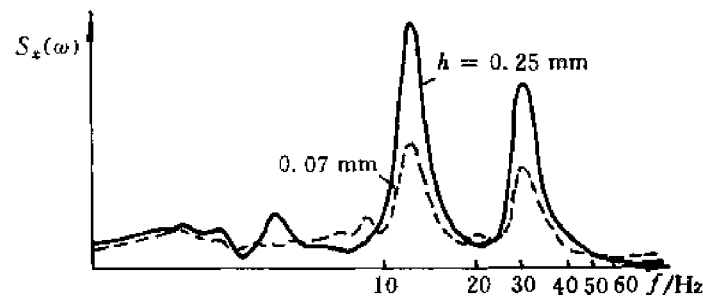


图 2-34 发动机在连杆轴承间隙变化时的振动谱

2. 互功率谱(互谱) 密度函数

如同自相关函数与自谱密度一样,互相关函数和互谱密度呈一对傅里叶变换对,其基本关系为

$$S_{xy}(\omega) = \int_{-\infty}^{\infty} R_{xy}(\tau) e^{-i\omega\tau} d\tau \quad (2.71)$$

由于 $R_{xy}(\tau)$ 不是偶函数,所以 $S_{xy}(\omega)$ 是复函数,具有如下特性:

- (1) $S_{xy}(\omega) = S_{yx}^*(\omega)$, 即 $S_{xy}(\omega)$ 和 $S_{yx}(\omega)$ 互为共轭函数;
- (2) 互谱密度与自谱密度之间存在有不等式 $|S_{xy}(\omega)| \leq S_x(\omega)S_y(\omega)$;
- (3) 互谱密度函数的标准化形式,称为凝聚函数

$$r_{xy}^2(\omega) = |S_{xy}(\omega)|^2 / [S_x(\omega)S_y(\omega)] \quad 0 \leq r_{xy}^2(\omega) \leq 1$$

式中 $S_x(\omega)$ 和 $S_y(\omega)$ 是信号 $x(t)$ 和 $y(t)$ 的自谱密度。在特定的某个圆频率 ω 下,如果 $r_{xy}^2(\omega) = 0$,则两个信号在此频率下是不相干的;若对所有的 ω ,总有 $r_{xy}^2(\omega) = 0$,则两个信号是完全不相干的;若对所有的 ω ,总有 $r_{xy}^2(\omega) = 1$,则称两个信号是完全相干的。必须注意,为得到正确的凝聚函数, $S_x(\omega)$, $S_y(\omega)$ 和 $S_{xy}(\omega)$ 必须在完全相同的条件(滤波器带宽、记录长度等)下求出,否则,求出的凝聚函数将会出现很大的误差。

互谱密度不像自谱密度那样具有明显的物理意义,引入这个概念主要是为了描述两个随机信号的相关性,在实际应用中,常常利用互谱密度确定系统的传递函数。

2.4 倒频谱分析方法

如果一实测信号 $y(t)$ 是由两个分量 $x(t)$ 和 $s(t)$ 叠加形成的, 即 $y(t) = x(t) + s(t)$, 则当两个分量的能量分别集中在不同的频率里时, 可用频域分析中的线性滤波或功率谱分析。当所要提取的分量以一定的形状作周期性重复, 而另一个分量是随时间变化的噪声时, 可用时域分析中的信号平均方法或相关分析。这些方法都可有效地处理线性迭加信号。

在工程实测的振动或声响信号不是振源信号本身, 而是振源或声响信号 $x(t)$ 经过传递系统到测点的输出信号 $y(t)$, 传递系统动态特性是由脉冲响应 $h(t)$ 描述的, 则振源或声响信号 $x(t)$ 与输出信号 $y(t)$ 有如下关系

$$y(t) = x(t) * h(t) \quad (2.72)$$

即输出 $y(t)$ 是输入 $x(t)$ 与脉冲响应 $h(t)$ 的卷积, 这时用处理线性迭加信号的方法就不够了, 而倒频谱能很好地处理这类问题。

2.4.1 倒频谱时频域转换的物理意义

功率倒频谱是 Bogert, Healy, Tukey 等人 1962 年提出来的。倒频谱可将输入信号与传递函数区分开来, 便于识别。当机械故障信号的频谱图出现难以识别的多族调制边频时, 应用倒频谱分析, 还可以分解和识别故障频率, 分析和诊断产生故障的原因。所以自倒频谱产生以来, 它已在回声、语音分析、地震、机械故障诊断、噪声分析等方面得到广泛的应用。

功率倒频谱的表达式为

$$C_p(\tau) = \{F^{-1}[\lg |X(f)|^2]\}^2 = \{F^{-1}[\lg S_x(f)]\}^2 \quad (2.73)$$

式中, $X(f)$ 与 $S_x(\omega)$ 分别为信号 $x(t)$ 的傅里叶变换与自功率谱密度函数。工程上常用上式的平方根, 即

$$C_x(\tau) = F^{-1}[\lg S_x(f)] \quad (2.74)$$

称为幅值倒频谱。如果将 $C_x(\tau)$ 与信号 $x(t)$ 的自相关函数 $R_x(\tau) = F^{-1}[S_x(f)]$ 进行比较就可以看到: 幅值倒频谱与自相关函数有类似之处, 所不同的是, 自相关函数是直接从自功率谱求傅里叶逆变换, 而幅值倒频谱则是对自功率谱的对数求傅里叶逆变换。

功率倒频谱或幅值倒频谱中的自变量 τ , 称为倒频率, 它具有与信号 $x(t)$ 及其自相关函数 $R_x(\tau)$ 中的自变量有相同的时间量纲。 τ 值大者, 称为高倒频率, 表示频谱图上的快速波动和密集谱频。与此相反, τ 值小者, 称为低倒频率, 表示频谱图上的缓慢波动和疏散谱频。在某些场合使用倒频谱而不用自相关函数, 是因为倒频谱在功率谱的对数转换时, 给幅值较小的分量有较高的加权, 其作用是既可帮助判别谱的周期性, 又能精确地测出频率间隔; 此外, 在某些情况下, 倒频谱之所以优于自相关函数, 还由于自相关函数检测回波峰值时, 与频谱形状的关系十分密切, 经过回波之后实际上已不可能加以检测; 而功率谱的对数对这种回波的影响是不敏感的。所以, 在自相关函数无法分解的场合, 倒频谱对频谱形状的不敏感性, 使它获得了许多应用。

2.4.2 倒频谱的基本原理

对功率谱作倒频谱变换, 其根本原因是在倒频谱上可以较容易地识别信号的组成分量, 便

于提取其中所关心的信号成分。例如一个系统的脉冲响应函数是 $h(t)$ ，输入为 $x(t)$ ，那么输出信号 $y(t)$ 等于 $x(t)$ 和 $h(t)$ 的卷积，倒频谱的作用就是将卷积变成简单的迭加。

对式(2.72)两边取傅里叶变换，根据卷积定理，时域中的卷积转换成频域中的相乘

$$Y(\omega) = X(\omega) \cdot H(\omega) \quad (2.75)$$

将式(2.75)取幅值平方，便得到功率谱的关系式

$$S_y(\omega) = S_x(\omega) \cdot |H(\omega)|^2 \quad (2.76)$$

两边取对数， $\lg S_y(\omega) = \lg S_x(\omega) + \lg |H(\omega)|^2$

由于傅里叶变换的线性性质，这个相加关系保留在倒频谱中

$$F^{-1}\{\lg S_y(\omega)\} = F^{-1}\{\lg S_x(\omega)\} + F^{-1}\{\lg |H(\omega)|^2\}$$

即 $C_y(\tau) = C_x(\tau) + C_h(\tau) \quad (2.77)$

式(2.77)含意是：如果输入信号 $x(t)$ 或系统的脉冲响应 $h(t)$ 中有一个已知，就可以从输出信号 $y(t)$ 的倒频谱 $C_y(\tau)$ 中除去，得到另一分量的倒频谱，例如 $C_h(\tau)$ ，对其进行傅里叶变换可得到 $\lg |H(\omega)|^2$ ，再进行指数运算，便得到传递函数的幅值 $|H(\omega)|$ 了。

图3-35为输入、输出和系统影响的对数功率谱及其倒频谱图。从图中可以看出：功率谱由两部分组成，其一是 $\lg S_x(f)$ ，是输入信号的谱，有明显的周期性，频率间隔为 Δf ；其二则是缓慢变化的中线，是系统的影响 $\lg |H(\omega)|^2$ ，二者合成为输出信号的谱 $\lg S_y(\omega)$ 。倒频谱有两个明显的波峰，高倒频率 τ_2 ($\tau_2 = 1/\Delta f$) 表示了输入信号的特征；低倒频率 τ_1 表示了系统的影响。显然，在倒频谱中，输入与系统响应是一种可分离的叠加性谱图，这为分解或判定其中任一分量提供了先决条件。

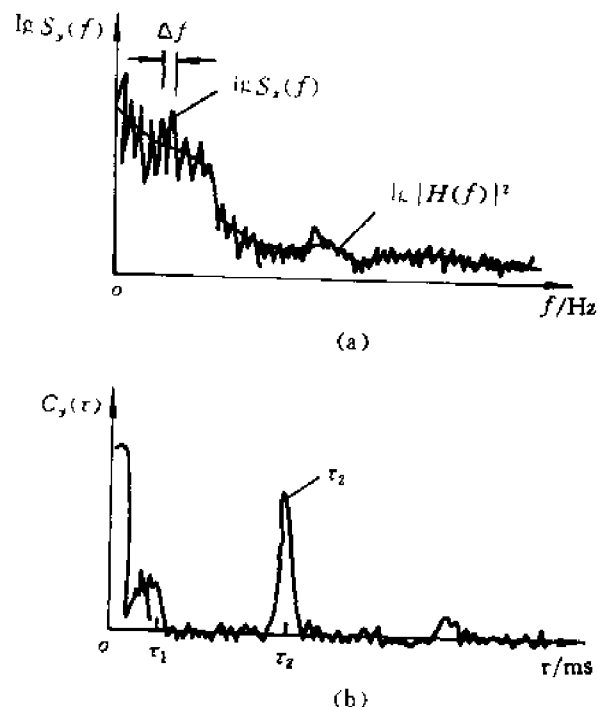


图 2-35 输入、输出和系统响应的功率及倒频谱
(a) 对数功率谱 (b) 倒频谱

利用倒频谱对信息进行分解的基本步骤如图 2-36 所示。

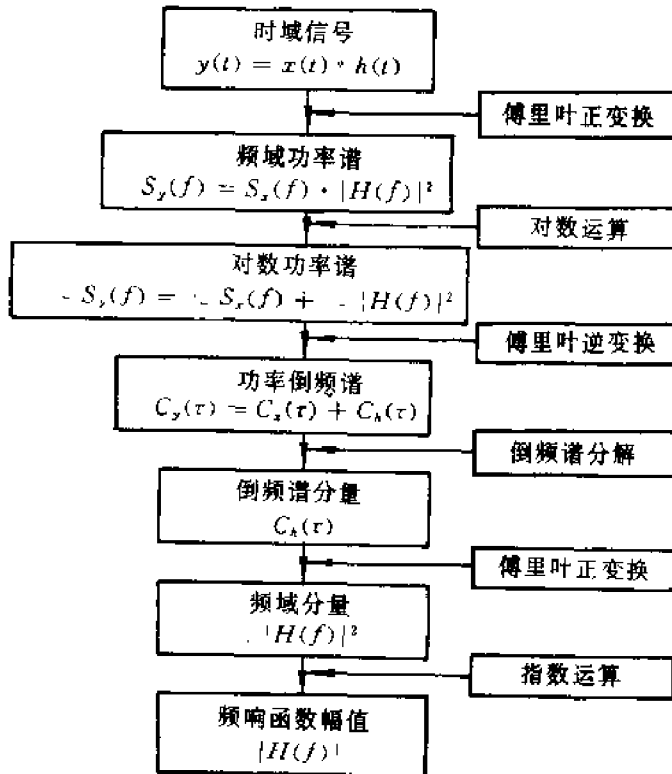


图 2-36 利用倒频谱对信号进行分析的基本步骤

2.4.3 倒频谱的应用 —— 回声的分析和剔除

由理想的平坦反射表面所产生的回声与原始信号混合,结果可用图 2-37 表示。图中 $x(t)$ 是原始信号; $y(t)$ 是掺有回声的混合(输出)信号;系数 α 表示回声能量的衰减, α 值范围为 $0 < \alpha < 1$; τ_0 表示回声的延迟时间。由原始信号 $x(t)$ 所产生的回声可表示为 $\alpha x(t - \tau_0)$ 。这样,我们实际记录下来的混合信号 $y(t)$ 是

$$y(t) = x(t) + \alpha x(t - \tau_0)$$

利用 α 函数的性质改写 $y(t)$

$$y(t) = x(t) + \alpha x(t - \tau_0) = x(t) * [\delta(t) + \alpha \delta(t - \tau_0)] \quad (2.78)$$

显然,具有回声的混合信号 $y(t)$ 可用原始信号 $x(t)$ 和一对脉冲函数的卷积来表示,如图 2-38 所示。脉冲函数之一在时间原点上,其强度(面积)等于 1,另一个在回声延迟时间 τ_0 上,其强度小于 1,相当于回声的衰减。若能设法将 $\delta(t) + \alpha \delta(t - \tau_0)$ 除去,就可以得到真实信号 $x(t)$ 及其真实功率谱 $S_x(f)$ 了。

利用倒频谱对其进行分析,对式(2.78)两边取傅里叶变换得

$$F[y(t)] = F[x(t)] * F[\delta(t) + \alpha \delta(t - \tau_0)]$$

$$Y(f) = X(f) (1 + \alpha e^{-j2\pi f \tau_0})$$

功率谱的关系式

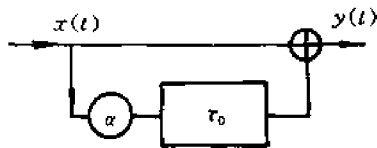


图 2-37 回声与原始信号混合

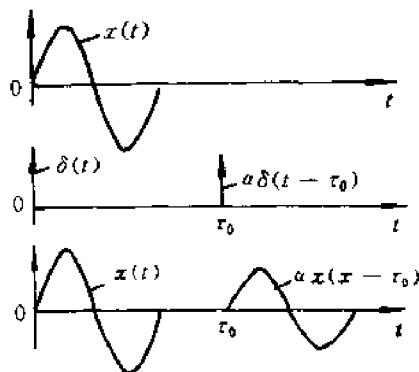


图 2-38 具有回声的混合信号的卷积表示

$$S_y(f) = S_x(f) |1 + \alpha e^{-j2\pi f\tau_0}|^2 = S_x(f) (1 + \alpha e^{-j2\pi f\tau_0})(1 + \alpha e^{j2\pi f\tau_0}) \quad (2.79)$$

对式(2.79)两边取对数

$$\lg S_y(f) = \lg S_x(f) + \lg(1 + \alpha e^{-j2\pi f\tau_0}) + \lg(1 + \alpha e^{j2\pi f\tau_0})$$

因为 $|\alpha e^{\pm j2\pi f\tau_0}| < 1$, $\lg(1 + \alpha e^{\pm j2\pi f\tau_0})$ 可展为幂级数, 所以有

$$\begin{aligned} \lg S_y(f) = \lg S_x(f) + \alpha e^{-j2\pi f\tau_0} - \frac{\alpha^2}{2} e^{-j2\pi f2\tau_0} + \frac{\alpha^3}{2} e^{-j2\pi f3\tau_0} - \dots + \\ \alpha e^{j2\pi f\tau_0} - \frac{\alpha^2}{2} e^{j2\pi f2\tau_0} + \frac{\alpha^3}{3} e^{j2\pi f3\tau_0} - \dots \end{aligned} \quad (2.80)$$

对式(2.80)两边取傅里叶逆变换, 利用公式 $F^{-1}[e^{\pm j2\pi f\tau_0}] = \delta(\tau \pm \tau_0)$ 便得到倒频谱 $C_y(\tau)$ 的表达式

$$\begin{aligned} C_y(\tau) = C_x(\tau) + \alpha \delta(\tau - \tau_0) - \frac{\alpha^2}{2} \delta(\tau - 2\tau_0) + \frac{\alpha^3}{3} \delta(\tau - 3\tau_0) - \dots \\ + \alpha \delta(\tau + \tau_0) - \frac{\alpha^2}{2} \delta(\tau + 2\tau_0) + \frac{\alpha^3}{3} \delta(\tau + 3\tau_0) - \dots \end{aligned} \quad (2.81)$$

式中 $C_y(\tau) = F^{-1}[\lg S_y(f)]$ $C_x(\tau) = F^{-1}[\lg S_x(f)]$

由上面的分析知道, 回声在倒频谱中形成了一系列的 δ 脉冲函数, 这些脉冲处在相当于时间轴 τ (倒频率) 上已知的位置, 这些位置可由计算回声延迟时间 τ_0 得到。在倒频谱上位于 $\tau_0, 2\tau_0, 3\tau_0, \dots$ 的地方, 可看到有幅值递减的脉冲峰。

还可证明, 在有回声混合信号 $y(t)$ 的功率谱 $S_y(f)$ 中存在周期成分, 由式(2.79)有

$$\begin{aligned} S_y(f) = S_x(f) |1 + \alpha e^{-j2\pi f\tau_0}|^2 = S_x(f) |1 + \alpha \cos 2\pi f\tau_0 - j\alpha \sin 2\pi f\tau_0|^2 = \\ S_x(f)(1 + \alpha^2) + 2\alpha S_x(f) \cos 2\pi f\tau_0 \end{aligned}$$

显然, 由于 $2\alpha S_x(f) \cos 2\pi f\tau_0$ 的存在, 随着 f 的变化, 使输出功率谱中出现了周期分量。周期分量的频率周期 $\Delta f = 1/\tau_0$ 。

图 2-39 表明信号中有回声, 在功率谱 $S_y(f)$ 中出现了频率周期 Δf 的周期成分, 而在倒频谱中, 在 $\tau_0, 2\tau_0, \dots$ 等时刻出现了幅值递减的峰值。如果从倒频谱减去这些脉冲峰值, 则关于回声的信息就被删去了。对剔除了回声脉冲峰值的倒频谱取傅里叶正变换, 再取指数函数, 便重新得到相当于去掉回声信号的真实功率谱了。

图 2-40 是一个用白噪声作为声音信号,加有一个反射面来产生回声,再用倒频谱进行分析的过程说明。图 2-40(a) 表示混有回声的白噪声信号的平均功率谱,可以清楚地看到功率谱的周期结构,其频率间隔为 $\Delta f(\text{Hz})$, $1/\Delta f$ 恰好等于回声的延迟时间 $\tau_0(\text{s})$,这一延迟时间取决于回声反射经过的路程和声音在空气中的传播速度。图 2-40(b) 是图 2-40(a) 的倒频谱,这里功率谱中的周期变成倒频谱中脉冲峰值的间隔 τ_0 ,在倒频率 τ 为 $\tau_0, 2\tau_0, \dots$ 的地方有幅值递减的脉冲峰。图 2-40(c) 是编辑了的倒频谱,用鸡冠滤波器令倒频谱中回声延迟时间间隔 τ_0 及其整数倍的位置上的脉冲峰值为零,这样就除去了回声的影响。图 2-40(d) 是对编辑了的倒频谱取傅里叶正变换后得到的功率谱。图 2-40(e) 是没有任何反射面得到的无回声信号的功率谱。图 2-40(d) 与图 2-40(a) 相比较可见,两谱图形差别很大,但图 2-40(d) 与图 2-40(e) 却很相近,说明了用倒频谱分析并采用鸡冠滤波器获得的剔除回声影响的功率谱,是令人满意的。

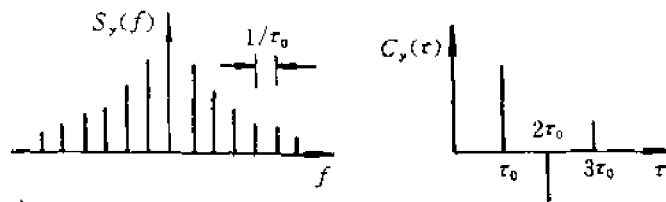


图 2-39 当信号中有回声时功率谱和倒频谱上的特征

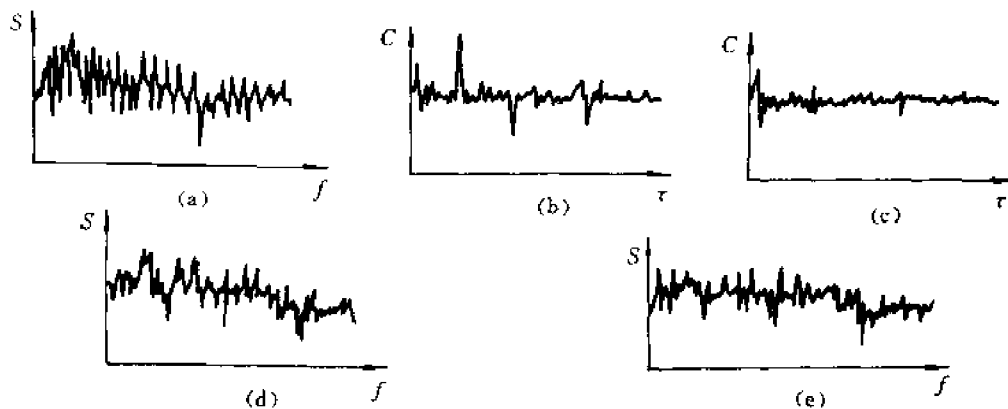


图 2-40 从功率谱上剔除回声的影响

(a) 平均功率谱; (b) 由(a)得到的倒频谱; (c) 编辑了的(即剔除了回声脉冲的)倒频谱;
(d) 由(c)得到的功率谱; (e) 无回声的平均功率谱

通过倒频率处理去掉回声的功率谱有以下特点:

- (1) 谱图上虚假的谱峰减少;
- (2) 噪声信号的主要频率成分突出;
- (3) 功率谱的谱峰高度能够比较真实地反映各频率分量在量值上的比例关系。

求得剔除回声影响的功率谱的全过程,用框图表示如图 2-41 所示。

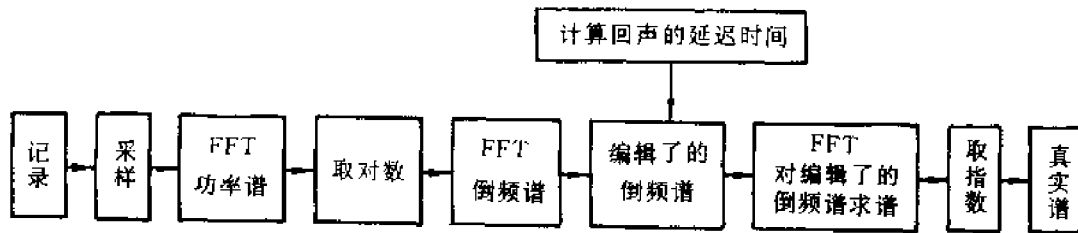


图 2-41 从功率谱上易除回声的过程框图

2.4.4 计算实例

在车间里对一台国产车床床头箱的噪声,用精密声级计和 B&K 磁带记录仪进行测试记录。车床在室内的空间位置如图 2-42 所示,测试过程中没有任何隔声和消声装置,因此床头箱的噪声通过箱壁散发出来后,由空气传播到声级计,同时还被地面、左墙、后墙和天花板等方面反射后再传播到声级计。这样由声级计接收到的信号中就掺杂了多方面反射进来的回声。将记录下来原始信号进行处理,得到的功率谱如图 2-43 所示。图 2-45 是图 2-43 的倒频谱,其中四个虚线谱峰为消除回声的影响而删掉的谱峰。在此例中,只考虑地面、后墙、左墙和天花板等主要的四个方面回声的影响,其他方面如前墙、右墙则因距车床较远,不予考虑。

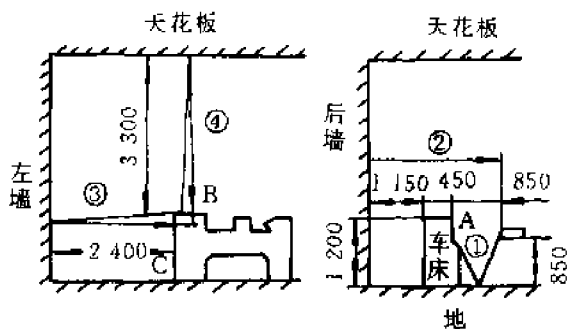


图 2-42 车床位置图

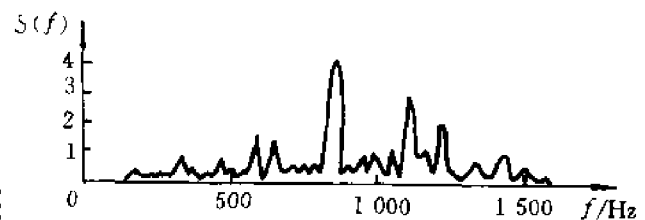


图 2-43 原始功率谱

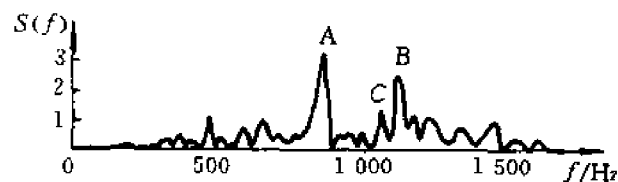


图 2-44 删除回声功率谱

如图 2-42 所示,由床头箱 A 面散射出来的噪声经过地而反射再到声级计(路线)与直接传播到声级计的路程差约为 1 m,我们近似取声音在空气中传播速度为 330 m/s,这样路线噪声的回声延迟时间 τ_1 为

$$\tau_1 \approx 1\text{m}/(330\text{m/s}) \approx 0.003 \text{ s} = 3 \text{ ms}$$

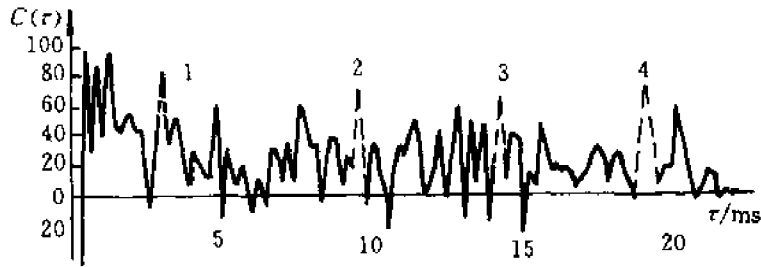


图 2-45 功率倒频谱

同理, A 面散射出来的噪声经过后墙反射后传到声级计(路线)与直接传到声级计的路程差约为 3.2 m, 回声延迟的时间 τ_2 为

$$\tau_2 \approx 3.2 \text{ m}/(330\text{m/s}) \approx 0.0096\text{s} = 9.6 \text{ ms}$$

C 面散发出来的噪声经过左墙反射后传到声级计(路线)与直接到声级计的路程差约 4.75 m, 回声延迟时间 τ_3 为

$$\tau_3 \approx 4.75 \text{ m}/(330\text{m/s}) \approx 0.0144\text{s} = 14.4 \text{ ms}$$

B 面散发出来的噪声经过天花板反射传到声级计(路线)与直接传到声级计的路程差约为 6.4 m, 回声的延迟时间 τ_4 为

$$\tau_4 \approx 6.4 \text{ m}/(330\text{m/s}) \approx 0.0194\text{s} = 19.4 \text{ ms}$$

对照图 2-45, 可以看到, 在倒频率(时间)轴上位于 3 ms, 9.6 ms, 14.4 ms 和 19.4 ms 处, 有明显的脉冲峰值, 可将这 4 个脉冲删除掉(图 2-45 中虚线表示删除掉的谱峰)。在计算机运算处理时, 即可将这 4 个峰值的数值冲零。由上面理论推证中知道, 对于位于 $k\tau_i$ ($i=1, 2, 3, 4$; $k=2, 3, \dots$) 处的脉冲峰值衰减很快, 比 τ_i ($i=1, 2, 3, 4$) 处的峰值要小得多, 作近似处理, 我们没有将它们删除。这样便得到了编辑了的倒频谱, 对其取傅里叶正变换和指数运算, 便回到了频域, 得到如图 2-44 所示的功率谱, 这就是比较真实的床头箱噪声功率谱。

对照图 2-43 和图 2-44, 可发现原始功率谱图形脉动大, 虚假的谱峰多, 图 2-44 中 A, B, C 三个谱峰较突出, 表明它们所对应的频率分量在噪声(2 500 Hz 以下)中占明显的优势, 通过实际分析, 频率 A 是机床电机(国产, 7 500 W, 1 500 r/min)的机壳共鸣声, 这一频率分量已为电机生产厂所做的电机噪声分析所证实。频率 B 是齿轮 z_1 和 z_2 两轮的一次啮合频率。频率 C 是齿轮 z_1 和 z_3 两轮的二次啮合频率。对噪声的主要分量能够心中有数, 就便于采取措施。

2.5 小波分析

2.5.1 从傅里叶变换到小波变换

如前所述, 傅里叶变换可以将时域信号变换到频域中的谱。就振动分析来说, 各频段的谱分量可以告诉我们信号的各个组成部分, 表征着信号的不同来源和不同特征。FFT 算法和现代谱理论的发展使得信号谱估计可以在很短的时间内完成, 从而实现对观测信号的实时分析。频谱估计现已成为故障诊断领域中十分重要的特征分析工具。

傅里叶变换的不足之处在于它只适用于稳态信号分析, 而非稳态信号在工程领域中是广泛存在的, 例如变速机械的振动等。加窗傅里叶变换是为了适应非稳态信号分析发展起来的一

种改进方法,时域信号 $x(t)$ 的加窗傅里叶变换如下

$$X(\omega, \tau) = \int_R x(t) w(t - \tau) e^{-j\omega t} dt \quad (2.82)$$

其中 $w(t - \tau)$ 为窗函数, τ 为可变参数, 变动 τ 可控制窗函数沿时间轴平移, 以实现信号 $x(t)$ 的按时逐段分析。由于式(2.82)中窗函数的大小和形状是固定的, 因此难以适应信号频率高低不同的局部化要求; 而且, 如果在信号中有短时(相对于窗长)、高频成分, 如在故障监测中, 对突变信号的分析和谐估计, 这种变换也不是非常有效的; 实际应用中, 要求对低频信号采用宽时窗, 高频信号采用窄时窗, 以提高谱线分辨率。小波分析发展了加窗傅里叶变换的局部化思想, 采用时窗宽度可调的小波函数替代式(2.82)中的窗函数, 它的窗宽随频率增高而缩小, 符合高频信号的分辨率较高的要求。

限于篇幅, 本节仅对小波变换的构造、部分性质及其在故障诊断中应用有关的理论作以简解。

2.5.2 小波函数及积分小波变换

1. 小波函数

概括地说, 对于函数 $\psi(t) \in L^2(R)$, 若满足 $\int_{-\infty}^{\infty} \psi(t) dt = 0$ 时, 称为小波函数或基小波, 它通过平移和缩放产生的一个函数族 $\psi_{b,s}(t)$

$$\psi_{b,s}(t) = \frac{1}{|s|^{1/2}} \psi\left(\frac{t-b}{s}\right) \quad (2.83)$$

称为由小波基 $\psi(t)$ 生成的依赖参数 b, s 的分析小波或连续小波, 其中 b, s 分别为平滑和伸缩因子, 统称为尺度因子。

用这一可变宽度的函数作变换基, 即可得到不是单一, 而是一系列不同分辨率的变换, 即小波变换, 它的主要特点是具有用多重分辨率来刻画信号局部特征的能力, 从而它很适合于探测正常信号中夹带的瞬态反常现象并展示其成分, 这在旋转机械、往复机械的状态监测及早期故障诊断中具有重要的意义。

2. 小波变换

信号 $x(t) \in L^2(R)$ 在尺度 s 上的小波变换定义为

$$W_\omega x(b, s) = \langle x(t), \psi_{b,s}(t) \rangle = \frac{1}{\sqrt{|s|}} \int_{-\infty}^{+\infty} x(t) \psi^*\left(\frac{t-b}{s}\right) dt \quad (2.84)$$

式中 $\psi^*[(t-b)/s]$ 是 $\psi[(t-b)/s]$ 的共轭函数, 符号 $\langle x, \psi \rangle$ 表示二信号的内积。如果小波函数 $\psi(t) \in L^2(R)$ 当其傅里叶变换 $\Psi(\omega)$ 满足容许性条件

$$C_\psi = \int_{-\infty}^{\infty} \frac{|\Psi(\omega)|^2}{|\omega|} d\omega < +\infty \quad (2.85)$$

时, 小波变换是可逆的, 且具有以下重构公式(小波反变换)

$$x(t) = \frac{1}{C_\psi} \iint_{\phi_{R_s}} \{W_\omega x(b, s)\} \left\{ \frac{1}{\sqrt{|s|}} \psi\left(\frac{t-b}{s}\right) \right\} \frac{ds db}{s^2} \quad (2.86)$$

容许性条件成立的必要条件为

$$\Psi(\omega) \Big|_{\omega=0} = \frac{1}{2\pi} \int_{-\infty}^{\infty} \psi_{b,s}(t) e^{j\omega t} dt \Big|_{\omega=0} = \frac{1}{2\pi} \int_{-\infty}^{\infty} \psi_{b,s}(t) dt = 0 \quad (2.87)$$

式(2.87)表明 $\psi_{b,s}(t)$ 必为衰减的振荡波形,即 $\psi_{b,s}(t)$ 必须具有小的波形,这就是 $\psi_{b,s}(t)$ 被称为“小波”的原因。具体地说,任何形如式(2.84)并满足式(2.85)所示容许性条件的正交函数族均可用来构成小波函数。当然,实际应用中还需从生成方便、可以形成有效的数值算法等多方面加以考虑。小波函数生成一直是该领域重要的研究方向之一,有关内容详见小波分析的著作。

函数的小波变换可理解为对其进行带通滤波,即将信号分解到一系列带宽和中心频率不同的频率通道的过程,图 2-46 是 $\text{sinc}x$ 函数及其小波分解。由图可见, $\text{sinc}x$ 函数被分解成很多频率通道,频率通道中心起始及终止值分别为 $2^{-4}\omega_0$ 和 $2^8\omega_0$ (表示带通滤波器的中心频率),各频道中心频率按对数尺度线性增加,每个频道频率变化范围比较小(波形频率接近通道中心频率)。

小波变换从信号中所提取出的成分主要由小波和其傅里叶变换在时域和频域的波形确定。图 2-47 是一类典型小波函数当取不同值时的波形,当 s 减少时, $\psi_{b,s}(t)$ 的局部性增强,而 $\Psi_{b,s}(\omega)$ 的局部性下降;当 s 增大时, $\psi_{b,s}(t)$ 的局部性下降,而 $\Psi_{b,s}(\omega)$ 的局部性增强,由此可见其波动性及带通性。

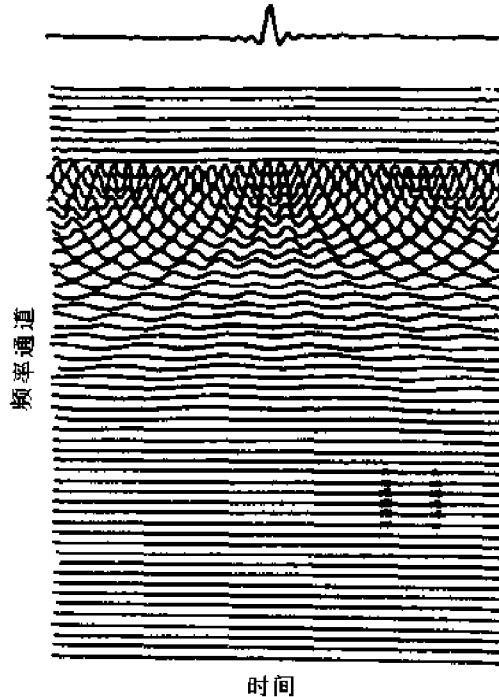


图 2-46 SINC 函数及其小波分解

3. 小波变换的特点

(1) 时-频窗口可调性: 设窗口小波函数的中心与半径分别为 t^* 和 Δ_ψ , 则函数 $\psi[(t-b)/s]$ 是中心在 $b+st^*$ 且半径为 $s\Delta_\psi$ 的一个窗函数, 因此, 小波变换 $W_\psi x(b, s)$ 给出了信号具有一个时间窗 $[b-st^*-s\Delta_\psi, b+st^*+s\Delta_\psi]$ 的局部化信息。这个窗对于小 s 的值变窄而对大的 s 值变宽。

下面考虑小波函数的傅里叶变换

$$\frac{1}{2\pi} \Psi_{b,s}(\omega) = \frac{|s|^{-1.2}}{2\pi} \int_{-\infty}^{\infty} e^{i\omega t} \psi\left(\frac{t-b}{s}\right) dt = \frac{s|s|^{-1.2}}{2\pi} e^{i\omega b} \Psi(\omega) \quad (2.88)$$

并假设窗函数的中心与半径分别为 ω^* 和 Δ_Ψ ; 其次, 如果设: $\eta(\omega) = \Psi(\omega + \omega^*)$, 那么 η 也是一

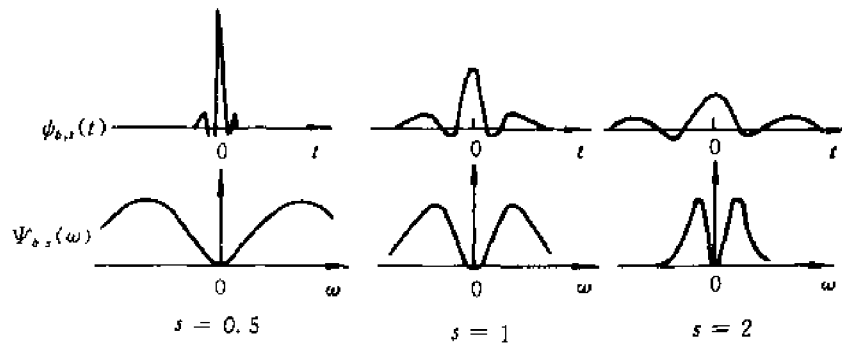


图 2-47 小波函数图形

个中心在原点且半径为 Δ_{ψ} 的窗函数,由公式(2.88),并根据 Parseval 恒等式,小波变换成为

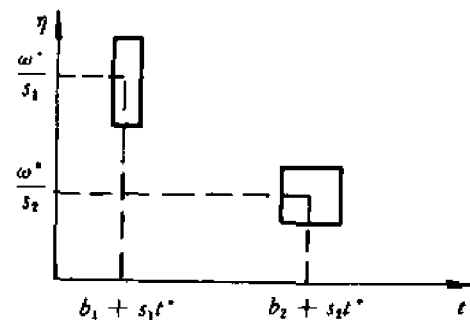
$$W_{\psi}x(b,s) = \frac{|s|^{-1/2}}{2\pi} \int_{-\infty}^{\infty} X(\omega) e^{i\omega b} \eta^* [s(\omega - \omega^*/s)] d\omega \quad (2.89)$$

公式(2.89)中的表示说明,除了倍数 $|s|^{-1/2}/2\pi$ 和一个线性相位 $e^{-i\omega b}$ 之外,同样的量 $W_{\psi}x(b,s)$ 又给出了具有一个“频率窗” $[\omega^*/s - \Delta_{\psi}/s, \omega^*/s + \Delta_{\psi}/s]$ 的信号 $x(t)$ 频谱 $X(\omega)$ 的局部信息。所以矩形时间-频率窗(图 2-48)

$$[b + s^* - s\Delta_{\psi}, b + s^* + s\Delta_{\psi}] \times [\omega^*/s - \Delta_{\psi}/s, \omega^*/s + \Delta_{\psi}/s]$$

的大小、位置,确定了小波变换所提取出的信号成分,小波变换 $W_{\psi}x(b,s)$ 的数值,表示的就是位于这个窗口内的信号能量大小。

窗口形状随尺度 s 而变化,但窗口面积保持不变,所以,对于检测高频现象(即小的 $s > 0$),窗会自动变窄,这意味着在高频带将有较好的时间分辨率;而对于检测低频特性(即大的 $s > 0$),窗会自动变宽,这意味着在低频带将有越来越高的频率分辨率。通过改变尺度因子 s 和平滑因子 b 的数值,调节窗口的形状、位置,合理地选取小波变换在时域和频域的分辨力,提取信号中位于感兴趣的频带和时段内的信号成分,从而实现可调窗口的时、频局部分析。

图 2-48 时间-频率窗($s_1 < s_2$)

(2) 小波变换是一个线性运算;因为小波变换是信号与小波之间的一个内积,而且一个矢量函数的连续小波变换是一个矢量,这个矢量的分量是不同的连续小波变换。

(3) 小波变换满足能量守恒定理;这意味着当信号施以小波变换时信息没有损失。

(4) 与加窗傅里叶变换相同,它是冗余的,由于是连续变化的,一个分析窗与另一个分析窗绝大部分内容是重叠的,即其相关性很强。

2.5.3 离散小波变换

正如傅里叶变换可分为积分傅里叶变换和离散傅里叶变换一样,小波变换也包含积分小波变换和离散小波变换,它们分别应用于连续信号和数字信号的分析。离散小波变换是将数字信号分解成一族小波函数的叠加,这样的分解使人们可以分析信号在特定的时、频窗范围内的

细节。当然,为了实现信号的小波分解,首先必须找到一个小波函数族。

如前所述,变动式(2.84)中的参数 s 和 b 可以生成小波函数族。 s 的变动使函数拉伸或压缩,形成不同“级”的小波; b 的变动使函数平移,形成不同“位”的小波。对于数字信号分析,是通过对小波的伸缩因子 s 和平移因子 b 的采样而离散化的,最常用的是二进离散,即参数 s 按二进规则(2^k)取值, $b = m2^{-k}$ ($k, m \in Z$) 等间隔取值。经过这种离散化后的小波和相应的小波变换成为二进小波变换,所以信号 $x(t)$ 在尺度 $s = 2^k$ 上的二进小波变换为

$$W_{\psi}x(2^{-k}m, 2^k) = \langle x(t), \psi_{2^k}(t) \rangle = 2^{-k/2} \int_{-\infty}^{\infty} x(t) \psi^* \left(\frac{t - 2^{-k}m}{2^k} \right) dt \quad (2.90)$$

小波变换 $W_{\psi}x(b, s)$ 给出了 $x(t)$ 变化位置 ($b + s^*$, t^* 为窗函数中心), 速率(s) 和量 ($W_{\psi}x(b, s)$ 的值) 的度量,而二进小波变换 $W_{\psi}x(2^{-k}m, 2^k)$ 给出在第 k 个倍频程(或频带)的局部谱信息。

假设信号 $x(t) \in L^2(R)$ 可用 $x_N \in V_N$ 来逼近,其中的子空间 $V_j = V_{j-1} \oplus W_{j-1}$ 形成 $L^2(R)$ 的子空间的一个嵌套序列,闭子空间 W_{j-1} 是 V_{j-1} 在 V_j 中的正交补子空间,它们的并在空间中是稠密的,它们的交是零空间 $\{0\}$ 。在这个意义上讲, x_N 具有惟一的分解

$$x_N = x_{N-1} + g_{N-1}$$

式中 $x_{N-1} \in V_{N-1}$ 和 $g_{N-1} \in W_{N-1}$;继续这个过程,则有

$$x_N = \dots + x_{N-M} + g_{N-M} + \dots + g_{N-2} + g_{N-1} \quad (2.91)$$

式中,对于任何 $j, x_j \in V_j, g_j \in W_j$,公式(2.91)中的惟一“分解”称为小波分解。上述关系中,闭子空间序列 $\{V_j\}_{j \in Z}$ 是 $L^2(R)$ 的一个多分辨分析或逼近。由此可得数字信号 $x(t)$ 的二进小波分解的数学表达式如下

$$\begin{aligned} x(t) &= a_0 \varphi(t) + \sum_{k, m \in Z} a_{k, m} \psi(2^k t - m) = \\ &= a_0 \varphi(t) + a_{1,0} \psi(2t) + a_{1,1} \psi(2t - 1) + \\ &= a_{2,0} \psi(4t) + a_{2,1} \psi(4t - 1) + a_{2,2} \psi(4t - 3) + \\ &= a_{3,0} \psi(8t) + \dots + a_{3,7} \psi(8t - 7) + \dots + a_{k,1} \psi(2^k t - m) + \dots \end{aligned} \quad (2.92)$$

式中 $a_0 \varphi(t)$ 表示 $x(t)$ 的直流分量,零级小波只有 $\varphi(t)$ 一项,二级小波由 $\psi(2t)$ 与 $\psi(2t - 1)$ 两个移位小波叠加组成,依此类推, k 级小波由 $2k$ 个移位小波 $\psi(2^k t - m)$ ($m = 0, 1, \dots, 2k - 1$) 叠加组成。

在小波分解表达式(2.92)中,每级小波实际上代表着不同倍频程段内的信号成分,所有频段正好不相交地布满整个频率轴,因此小波分解可以实现频域局部分析。另一方面,由于各级小波为多个移位小波加权和,各移位小波系数又反映了相应频段的信号在各时间段上的信息,即同时实现了时域局部分析。要实现式(2.92)所示的小波分解,关键问题是确定其中各小波分量的系数。如果所采用的小波函数族满足正交性条件,那么理论上可按下式确定各小波系数

$$a_{k, m} = \int x(t) \psi(2^k t - m) dt \quad (2.93)$$

但由于小波函数通常比较复杂甚至不具有解析表达式,实际上积分表达式(2.93)只是从理论上反映了小波系数、小波函数和信号 $x(t)$ 三者之间的关系,计算出小波系数还必须采用其他可行的方法。就目前的研究水平而言,最成功的算法是 Mallat 算法,该算法利用小波的正交性导出各系数矩阵的正交关系,从高级到低级逐级滤去信号中的各级小波。为叙述简便,假设数字信号 $x(t)$ 有 8 个采样点,其小波分解式中包含零级、一级和二级小波,分别记为 $x_0, x_1,$

x_2 , 其中 x_1 只含 $a_0\psi(t)$ 一项, x_1 由两个移位小波 $a_{1,0}\psi(2t)$ 和 $a_{1,1}\psi(2t-1)$ 叠加而成, x_2 由四个移位小波 $a_{2,0}\psi(4t)$, $a_{2,1}\psi(4t-1)$, $a_{2,2}\psi(4t-2)$ 和 $a_{2,3}\psi(4t-3)$ 叠加而成, 加上常数项 $x_0 = a_0\psi(t)$, 分解式中共有 8 项, 与信号采样点相同。若同级小波作为一个整体, 参照公式(2.91), (2.92), 则有 8 个采样点的数字信号 $x(t)$ 的分解式可写成如下形式

$$x_N = x_4 + x_0 + x_1 + x_2 \quad (2.94)$$

我们采用图 2-49 并结合小波分解式(2.94) 来介绍 Mallat 算法的主要思想。

$x(t) = x_4 + x_0 + x_1 + x_2$			
$Lx = x_4 + x_0 + x_1$		$Hx; x_2(a_{2,0}, a_{2,1}, a_{2,2}, a_{2,3})$	
$LLx; x_4 + x_0$	$HLLx; x_1(a_{1,0}, a_{1,1})$		
$LLLx; x_4(a_4)$	$HLLLx; x_0(a_0)$		

图 2-49 Mallat 算法示意图

Mallat 算法不直接计算积分表达式, 而是利用小波函数族的正交性, 从高级到低级滤出信号中的各级小波。以上述含 8 个采样点的数字信号 $x(t)$ 为例, Mallat 算法的第一步是从中滤出二级小波 x_2 , 同时确定二级小波中各移位小波的系数 $a_{2,0}, a_{2,1}, a_{2,2}$ 和 $a_{2,3}$, 并将信号分解成 x_2 和 $x_4 + x_0 + x_1$ 的叠加。这一过程相当于一低通滤波器, 对应于二级小波的高频信号 Hx 被分离出来, 而低频信号 Lx (零、一级小波及常数项) 全部保留, 如图 2-49 所示。算法的第二步是从 Lx 中再滤出一级小波并确定一级小波系数, 即 $a_{1,0}$ 和 $a_{1,1}$ 。如此进行下去, 直至滤出各级小波并确定所有的系数, 小波分解也就完成了。Mallat 算法概念清楚、计算简便, 其在小波分析中的地位, 相当于傅里叶分析中 FFT 算法。但要完整地介绍该算法需要较大的篇幅, 具体内容参见有关小波分析的著作。

小波分析技术的进一步介绍涉及许多较深的数学概念, 一些与工程上应用有关的内容如多元小波、小波包分解、多分辨分析与小波基理论等请读者参阅有关文献。

2.5.4 基于小波分析的故障诊断

如前所述, 傅里叶分析的理论基础是待分析信号的平稳性。对于非平稳信号, 傅里叶分析可能给出虚假的结果, 从而导致故障的误诊断。对设备故障诊断问题来说, 由于以下原因, 使傅里叶分析的应用受到限制。

(1) 由于机器转速不稳、负荷变化以及机器故障等原因产生的冲击、摩擦导致非平稳振动信号的产生。

(2) 由于机器各零部件的结构不同, 致使振动信号所包含的不同零部件的故障频率分布在不同的频道范围内。特别是当机器隐藏有某一零部件的早期微弱缺陷时, 它的缺陷信息被其他零部件的振动信号和随机噪声所淹没。

对于这类问题, 小波分析方法具有无可比拟的优点。由于小波分解技术能够将任何信号(平稳或非平稳) 分解到一个由小波伸缩而成的基函数族上, 信息量完整无缺, 在通频范围内得到分布在不同频道内的分解序列, 在时域和频域均具有局部化的分析功能。因此, 可根据故障诊断的需要选取包含所需零部件故障信息的频道序列, 进行深层信息处理以查找机器故障源。

1. 故障信号的分解

对于往复机械, 振动信号的频谱在通频带上均有大量的能量分布, 这样仅用频谱分析方法, 由于其频率分辨率低, 且对不同的频率成分在时域上的分辨率是不变的, 因而难以对故障类型作出判定。而小波分析对不同的频率成分, 在时域上的分辨率是可调的, 高频者小, 低频者

大,它能够将信号分解成多尺度成分,并对于大小不同的尺度成分采用相应的时域与频域步长,从而能够不断地聚焦到信号的任意微小细节,具有比傅里叶变换更强的特征提取功能。

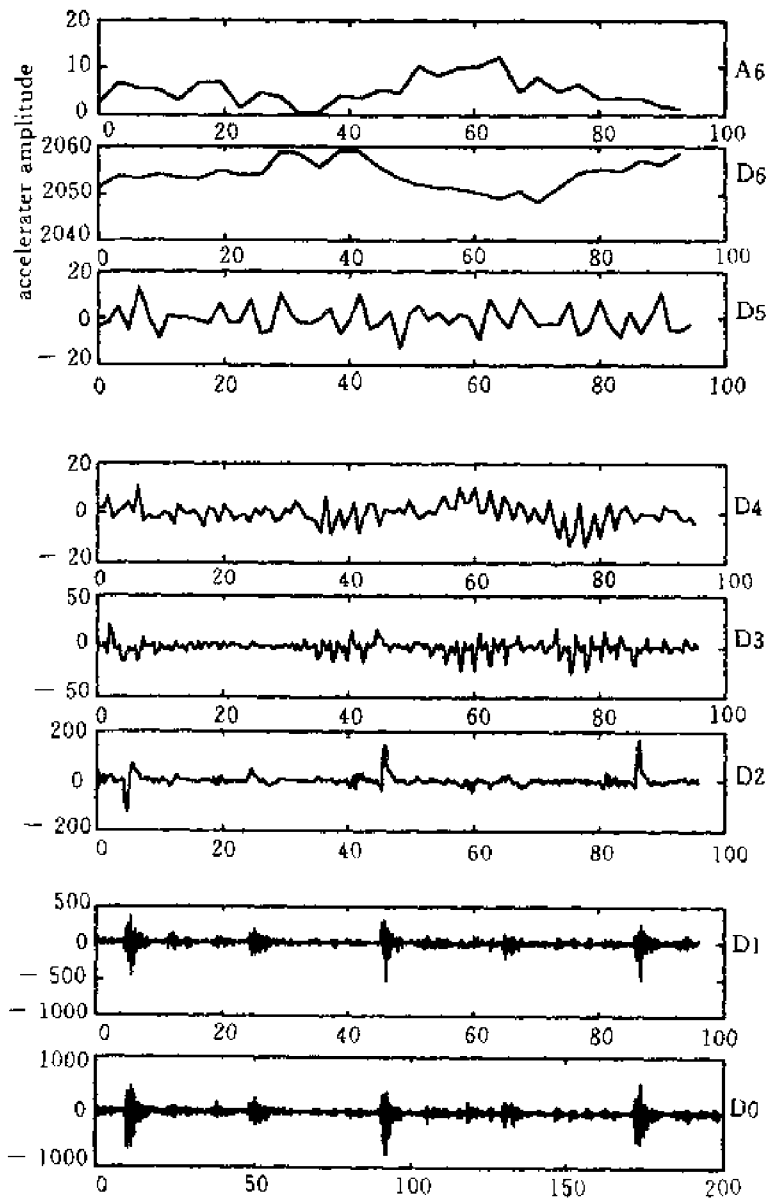


图 2-50 空气压缩机气阀振动信号的小波分解

图 2-50 为空气压缩机气阀振动信号的二进小波分解表示,在分解结果中,各个小波序列 D1, D2, ..., D6 和 A6 分别包含了原始信号的从高到低的各频率段信息,这里分析频率上限为 15 kHz,所以 A6 位于 0 ~ 234.375 Hz 频带, D6 位于 234.375 ~ 468.75 Hz 频带, D5 位于 468.75 ~ 937.5 Hz 频带, D4 位于 937.5 ~ 1875 Hz 频带, D3 位于 1.875 ~ 3.75 kHz 频带, D2 位于 3.75 ~ 7.5 kHz 频带, D1 位于 7.5 ~ 15 kHz 频带。在全频带内正交分解的结果,信息量既无冗余,也不疏漏,而且信号分解和重构可有针对性地选择有关频带信息和剔除噪声干扰,从中提取故障特征,以便进行故障类型的识别。

2. 局部异常信号的检测

在设备探伤或诊断中,常常只对设备异常区域所引起的信号局部变化感兴趣,如果将这些

局部异常区分离出来,就可对设备故障的大小和位置作出分析。

图2-51是一段在钢丝绳无损检测中获得的信号和小波分析图形,信号波形中的跳变部分为由断绳所引起的局部异常成分。

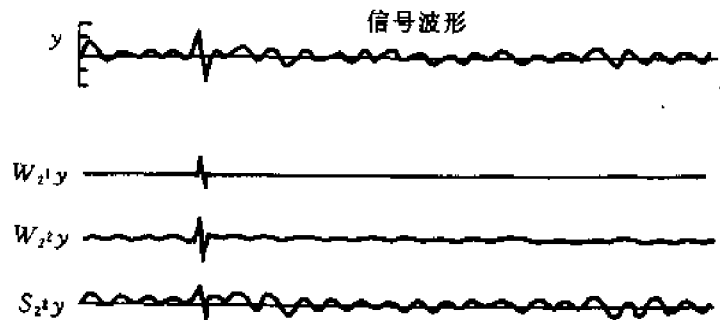


图2-51 钢丝绳检测信号和小波分析图解

突变信号的峰点在对应的同一位置的所有尺度上都产生相应的最大值点,因此通过小波变换后,根据各尺度上最大值点的位置和数值就可以确定出信号异常的程度。特别是当信号中含有较强高频噪声时,时域波形往往很混乱,不易准确判断,对信号进行小波变换后,噪声一般只干扰低阶小波系数,这时可由高阶小波中的最大值点位置来判定峰点位置。

3. 干扰信号的剔除

实测中的信号往往受到多种因素的干扰,如常见的高频噪声和低频扰动等。通过小波变换将信号分解为位于不同频段和时段内的成分,若干扰信号与有用信号位于不同的频带内,则只要将干扰信号所对应的那一阶小波系数置零,然后按小波反演公式对信号进行重构,即可达到消除干扰的目的。图2-52是反映柴油机燃油系统故障的高压油管中的压力波形,为了便于分析,采用小波变换与重构消去了信号中的高频干扰。

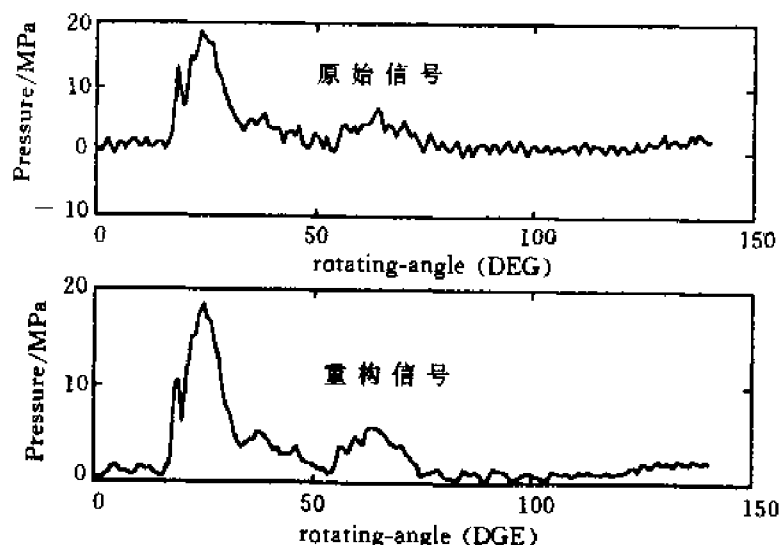


图2-52 柴油机高压油管的压力波形与小波变换波形

习题与思考题

- 2-1 新、旧机器的振动噪声信号的概率分布密度函数有什么区别?自相关函数有什么不同?
- 2-2 机械故障诊断中常用的幅值指标有哪些?无量纲指标有哪些?对无量纲指标有什么要求?
- 2-3 什么是自相关函数?自相关函数有什么性质?在机械设备状态监测中有什么用途?
- 2-4 什么是互相关函数?互相关函数有什么性质?在机械设备状态监测中有什么用途?
- 2-5 设一信号 $x(t)$, 它是由两个频率、相角均不相等的余弦函数叠加而成, 其数学表达式为 $x(t) = A_1 \cos(\omega_1 t + \varphi_1) + A_2 \cos(\omega_2 t + \varphi_2)$ 求该信号的自相关函数。
- 2-6 一管道发生泄漏, 现分别在相距 1 500 mA, B 两处用传感器测量, 测得信号经互相关处理的相关函数峰值在 $\tau = 1.25$ s, 已知管道中声速为 300 m/s, 求泄漏处的位置。
- 2-7 已知信号的自相关函数 $R_x(\tau) = A \cos(\omega\tau)$, 求出该信号的均方值 ψ_x^2 和均方根值 ψ_x 。
- 2-8 已知一个随机信号 $x(t)$ 的自功率谱密度函数为 $S_x(f)$, 将 $x(t)$ 输入到传递函数为的系统, 其传递函数为 $H(s) = 1/(\tau s + 1)$, 试求该系统的输出信号 $y(t)$ 的自功率谱密度函数为 $S_y(f)$, 及输入、输出两函数的互功率谱密度函数 $S_{xy}(f)$ 。
- 2-9 对三个余弦函数 $x_1(t) = \cos 4\pi t$, $x_2(t) = \cos 12\pi t$, $x_3(t) = \cos 20\pi t$ 进行理想采样, 采样频率 $f_s = 6$ Hz。求三个采样输出序列, 比较这三个结果, 画出 $x_1(t)$, $x_2(t)$, $x_3(t)$ 的波形及采样点位置, 并解释频率混叠现象。
- 2-10 已知一有限长序列 $x(n) = \begin{cases} 1 & 2 \leq n \leq 6 \\ 0 & n = 0, 1, 7, 8, 9 \end{cases}$
- (1) 用直接 DFT 方法求 $\text{DFT}[x(n)] = X(k)$, 再由所得结果求 $\text{IDFT}[X(k)] = x(n)$;
- (2) 用 FFT 方法求 $X(k)$, 再以所得结果求 $\text{IFFT}[X(k)] = x(n)$ 。
- 2-11 分析说明小波变换的基本原理, 比较其与傅里叶变换的相同点与不同之处。
- 2-12 定性说明如何进行时域同步平均?
- 2-14 试对比时域同步平均与功率谱分析的异同, 时域同步平均与自相关函数在提取周期分量时的原理和结果。

第 3 章

机械故障诊断的时序模型分析方法

前述已知,由于动态过程十分复杂,人们很难从观测数据直接分析系统的变化规律,虽然数学模型不等同于客观系统,但它能对系统作最本质的描述。如果说某种模型能正确地描述系统,意味着它能正确地表示系统的内在规律,动态过程状态的变化,反映在其数学模型的结构、参数和特征函数的变化。因此,建立数学模型的目的是便于人们更方便地应用数学工具,分析和认识客观过程的特征,尤其是人们所关心的那些特征,以便更深刻、更集中地了解过程的规律,判断设备状态的变化趋势及状态的属性。除此之外,模型通常还可以用于了解过程的规律和发展趋势进行预报和控制。因此研究动态系统时序模型的方法,将会给状态监测与故障诊断带来便利。

3.1 时间序列模型结构特征

3.1.1 机械设备运行过程数据序列的特点

机械设备在运行过程中的各种运行参数,以及所产生的振动、噪声、温升等一类信号量,都可以看作是一个时间过程,它即为所观测动态系统的输出。可以将传感器拾取的、连续变化的参数经过模/数(A/D)转换,得到一个离散的时间序列: $\{x_k\}$, ($k = 1, 2, \dots, N$),这一时间序列通常具有以下特点:

- (1) 由于动态过程是随机过程,所以时间序列是随时间而随机地变化的序列,它一般是平稳或可近似认为是平稳的随机离散信号;
- (2) 系统的输入,即产生这一随机时间序列的原因无法确知;
- (3) 由于机械系统相互耦合,十分复杂,加大了对时间序列分析的难度。

一般来说,通过分析,把复杂系统抽象为简单的物理模型只能作一般规律分析,很难用于对实际机器状态的监测与诊断。在这种场合下,时间序列模型(简称为时序模型)则具有无可比拟的优势。

3.1.2 时序模型的概念

为了对时间序列进行数学描述,研究时间序列的变化规律,需要建立数

学模型,这种模型通称为时序模型。时序模型方法在目前已经广泛应用于生物医疗、地球物理、语言识别、机械振动、噪声工程等各个领域,具有对一个物理过程进行预报、识别诊断、在线控制等多种用途。机器诊断的模型方法,就是在机器运行过程中,首先选定恰当的诊断参数,然后建立其时序模型,通过对时序模型的相应判据进行分析,以诊断机器状态的变化。一般情况下,模型方法可以比较可靠地回答机器属于正常或异常运行状态的问题,而不能准确地回答为什么的问题。但是,在相当多的场合,能够回答前一个问题,已经是十分难能可贵的了,因为这对于事故预防,已经发挥了积极的作用。

平稳时间序列 $\{y_k\}(k=1,2,\dots,N)$,一般地说, $E y_k = \mu_y \neq 0$ 。为方便起见,令 $x_k = y_k - \mu_y$,显然 $E x_k = \mu_x = 0$,于是得序列 $\{x_k\}(k=1,2,\dots,N)$,仍然是平稳时间序列。均值为零的平稳时间序列可表示为下面三种形式中的一种。

1. 自回归模型 AR(m)

任何一个时刻 k 上的数值 x_k 可表示为过去 m 个时刻上数值的线性组合加上 k 时刻的白噪声,即可表示为

$$x_k = \phi_1 x_{k-1} + \phi_2 x_{k-2} + \dots + \phi_m x_{k-m} + a_k \quad (3.1)$$

式中 $\{a_k\}(k=1,2,\dots)$ 是白噪声,满足 $E a_k = 0, D a_k = \sigma_a^2 (0 < \sigma_a^2 < \infty), E a_k a_i = 0 (k \neq i)$,即离散白噪声是互不相关的、均值为零且方差相同的随机变量序列。常数 m 叫做阶次,常数系数 $\phi_i (i=1,2,\dots,m)$ 称为自回归系数,且 $m > 0, \phi_m \neq 0$ 。可以表示为线性差分方程(3.1)形式的平稳序列 $\{x_k\}(k=1,2,\dots,N)$,称为具有自回归模型, m 阶自回归模型记为AR(m),它的含义是:系统在 k 时刻的输出 x_k 可用此系统在 k 时刻前的 m 个输出 $x_{k-1}, x_{k-2}, \dots, x_{k-m}$ 与 k 时刻的白噪声的线性组合表示。

2. 滑动平均模型 MA(n)

x_k 可表示为白噪声 $\{a_k\}$ 在 k 时刻和 k 时刻以前 n 个时刻上数值 $a_{k-1}, a_{k-2}, \dots, a_{k-n}$ 的加权和,或者说滑动和的形式,即可表示为

$$x_k = a_k - \theta_1 a_{k-1} - \theta_2 a_{k-2} - \dots - \theta_n a_{k-n} \quad (3.2)$$

式中常数 n 叫做阶次,常数系数 $\theta_i (i=1,2,\dots,n)$ 称为滑动平均系数,且 $n > 0, \theta_n \neq 0$ 。可以表示为线性差分方程(3.2)形式的平稳序列 $\{x_k\}(k=1,2,\dots,N)$,称为具有滑动平均模型, n 阶滑动平均模型记为MA(n),它的含义是:系统在 k 时刻的输出 x_k 可看成此系统在 k 时刻与 k 时刻前 n 个互相独立的白噪声输入的线性和。

3. 自回归滑动平均模型 ARMA(m, n)

可表示为线性差分方程形式

$$x_k - \phi_1 x_{k-1} - \phi_2 x_{k-2} - \dots - \phi_m x_{k-m} = a_k - \theta_1 a_{k-1} - \theta_2 a_{k-2} - \dots - \theta_n a_{k-n} \quad (3.3)$$

(其中 $m > 0, n > 0, \phi_m \neq 0, \theta_n \neq 0$)的平稳序列 $\{x_k\}(k=1,2,\dots,N)$,称为具有自回归滑动平均模型,记为ARMA(m, n)。常数 m, n 叫做自回归滑动平均模型的阶次。系数 ϕ, θ 的含义与上述相同。自回归滑动平均模型的含义是:在时刻 k 的输出 x_k 是系统在 k 时刻前的 m 个输出 $x_{k-1}, x_{k-2}, \dots, x_{k-m}$ 与由 $k-n$ 到 k 时刻中 $n+1$ 个互相独立的白噪声输入的线性和。

在自回归滑动平均模型式(3.3)中取 $m > 0, n = 0$,变成自回归模型式(3.1);如果取 $m = 0, n > 0$,式(3.3)又变成滑动平均模型式(3.2),因而,自回归滑动平均模型是较一般的模型,自回归模型和滑动平均模型是它的特殊形式。

一个自回归模型可以逼近ARMA和MA模型。对式(3.3)的等式两端进行 z 变换,将 a_k 看

作是一个系统的输入,即白噪声输入,将 x_k 看作是系统的输出,引用前述关于自谱密度和传递函数之间的关系: $S_x = |H(z)|^2 S_a$, 其中

$$H(z) = \frac{1 - \sum_{i=1}^n \theta_i z^{-i}}{1 - \sum_{i=1}^m \phi_i z^{-i}} = \frac{\Theta(z)}{\Phi(z)} \quad (3.4)$$

或
$$H(z) = 1 / \frac{\Phi(z)}{\Theta(z)} = \frac{1}{\Phi'(z)}$$

则在式(3.4)中,用连除法将分子的多项式除以分母的多项式,可以得到一个具有无穷多项的多项式作为分母,而分子为1.这就说明,一个自回归滑动平均模型可以用一个高阶的自回归模型逼近;同样滑动平均模型也可以用一个高阶的自回归模型逼近,其逼近的程度取决于所取的自回归模型的阶次。

一般情况下,由于自回归模型中的系数 ϕ_i 可应用线性最小二乘方原理进行确定,计算简单且速度快,特别有利于在线诊断.因此,目前生产实践中,低阶自回归模型仍然是最常用、最普通的一种时序模型。

3.2 自回归模型的参数、阶次的确定

3.2.1 AR模型参数的最小二乘方估计

对于AR(m)模型

$$x_k = \phi_1 x_{k-1} + \phi_2 x_{k-2} + \cdots + \phi_m x_{k-m} + a_k$$

$$a_k \sim \text{NID}(0, \sigma_a^2)$$

所谓参数估计,是根据已知的观测数据 $\{x_k\} (k=1, 2, \dots, N)$, 按某一方法估计出 $\phi_i (i=1, 2, \dots, m)$ 和 σ_a^2 这 $m+1$ 个参数.对模型的残差 $\{a_k\} (k=1, 2, \dots)$, 由上式有

$$a_k = x_k - \phi_1 x_{k-1} - \phi_2 x_{k-2} - \cdots - \phi_m x_{k-m}$$

根据的 σ_a^2 含义,对模型而言, σ_a^2 是模型残差序列 $\{a_k\}$ 的方差,故有

$$\sigma_a^2 = E[a_k^2] = \frac{1}{N} \sum_{k=m+1}^N (x_k - \phi_1 x_{k-1} - \phi_2 x_{k-2} - \cdots - \phi_m x_{k-m})^2 \quad (3.5)$$

可见,一旦估计出 $\phi_i (i=1, 2, \dots, m)$, 即可按上式估计 σ_a^2 . 所以,通常所指的参数估计,即是指估计出 $\phi_i (i=1, 2, \dots, m)$ 这 m 个参数。

1. 几个基本命题

- (1) 当 $k \neq j$ 时, $E a_k a_j = 0$, 即在不同时刻, a_k 是相互独立的, a_k 与 a_{k-1}, a_{k-2}, \dots 均不相关。
- (2) a_k 的分布是正态的, 即 $a_k \sim \text{NID}(0, \sigma_a^2)$ 。
- (3) 当 $j > 0$ 时, $E x_{k-j} a_k = 0$, 即 a_k 与 x_{k-1}, x_{k-2}, \dots 均不相关, 这从基本命题(1)即可看出。

2. 样本自相关函数

平稳序列 $\{x_k\} (k=1, 2, \dots, N)$, 因为 $E x_k = 0$, 所以自相关函数和自协方差函数相同, 为

$$r_k = E(x_j x_{j+k}) = \frac{1}{N} \sum_{j=1}^{N-k} x_j x_{j+k}, \quad k=0, 1, 2, \dots, K (K < N) \quad (3.6)$$

$k=0, r_0$ 就等于序列的方差 σ_x^2 , 定义样本自相关系数为

$$\rho_k = r_k/r_0, k = 0, 1, 2, \dots, K (K < N) \quad (3.7)$$

3. AR 模型参数的最小二乘方估计

假设由平稳序列 $\{x_k\} (k = 1, 2, \dots, N)$ 建立的自回归模型为

$$x_k = \phi_1 x_{k-1} + \phi_2 x_{k-2} + \dots + \phi_m x_{k-m} + a_k$$

按最小二乘方原理确定自回归系数, 即选 $\phi_i (i = 1, 2, \dots, m)$ 使均方偏差达到最小。

$$\begin{aligned} \delta = \sigma_a^2 &= E(x_k - \sum_{j=1}^m \phi_j x_{k-j})^2 = \\ &= E(x_k - \phi_1 x_{k-1} - \phi_2 x_{k-2} - \dots - \phi_m x_{k-m})^2 \end{aligned}$$

这是多元函数求极值问题, 为此, 让

$$\frac{\partial \delta}{\partial \phi_1} = E[2(x_k - \phi_1 x_{k-1} - \phi_2 x_{k-2} - \dots - \phi_m x_{k-m})(-x_{k-1})] = 0$$

$$\frac{\partial \delta}{\partial \phi_2} = E[2(x_k - \phi_1 x_{k-1} - \phi_2 x_{k-2} - \dots - \phi_m x_{k-m})(-x_{k-2})] = 0$$

.....

$$\frac{\partial \delta}{\partial \phi_m} = E[2(x_k - \phi_1 x_{k-1} - \phi_2 x_{k-2} - \dots - \phi_m x_{k-m})(-x_{k-m})] = 0$$

化简得

$$\left. \begin{aligned} r_0 \phi_1 + r_1 \phi_2 + r_2 \phi_3 + \dots + r_{m-1} \phi_m &= r_1 \\ r_1 \phi_1 + r_0 \phi_2 + r_3 \phi_3 + \dots + r_{m-2} \phi_m &= r_2 \\ &\dots\dots\dots \\ r_{m-1} \phi_1 + r_{m-2} \phi_2 + r_{m-3} \phi_3 + \dots + r_0 \phi_m &= r_m \end{aligned} \right\} \quad (3.8)$$

各个等式的两边除以 r_0 , 得

$$\left. \begin{aligned} \rho_0 \phi_1 + \rho_1 \phi_2 + \rho_2 \phi_3 + \dots + \rho_{m-1} \phi_m &= \rho_1 \\ \rho_1 \phi_1 + \rho_0 \phi_2 + \rho_3 \phi_3 + \dots + \rho_{m-2} \phi_m &= \rho_2 \\ &\dots\dots\dots \\ \rho_{m-1} \phi_1 + \rho_{m-2} \phi_2 + \rho_{m-3} \phi_3 + \dots + \rho_0 \phi_m &= \rho_m \end{aligned} \right\} \quad (3.9)$$

写成矩阵的形式有

$$\begin{bmatrix} 1 & \rho_1 & \rho_2 & \dots & \rho_{m-1} \\ \rho_1 & 1 & \rho_1 & \dots & \rho_{m-2} \\ & & \dots\dots\dots & & \\ \rho_{m-1} & \rho_{m-2} & \dots & & 1 \end{bmatrix} \begin{bmatrix} \phi_1 \\ \phi_2 \\ \dots \\ \phi_m \end{bmatrix} = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \dots \\ \rho_m \end{bmatrix} \quad (3.10)$$

或有

$$P_m \varphi = \rho$$

式中 $P_m = \begin{bmatrix} 1 & \rho_1 & \rho_2 & \dots & \rho_{m-1} \\ \rho_1 & 1 & \rho_1 & \dots & \rho_{m-2} \\ & & \dots\dots\dots & & \\ \rho_{m-1} & \rho_{m-2} & \dots & & 1 \end{bmatrix}$ 为自相关矩阵, $\varphi = [\phi_1 \phi_2 \dots \phi_m]^T$ 为参数矩

阵, $\rho = [\rho_1 \rho_2 \dots \rho_m]^T$ 为自相关系数矩阵。

由于 $\rho_i (i = 1, 2, \dots, m)$ 是可以直接由 $\{x_k\}$ 约计的, 只要在式(3.10)中解出 $\phi_i (i = 1, 2, \dots, m)$ 即可

$$\varphi = P_m^{-1} p \quad (3.11)$$

式(3.11)称为 Yule-Walker 方程。由于自相关矩阵 P_m 中主对角线为 1, 主对角线两侧诸元素两两相等, 即此矩阵主对角线两侧的元素对称于主对角线, 因此属于 Toeplitz 矩阵。在 Toeplitz 矩阵求逆的运算中, 可以采用 Levinson 循环递推法, 将计算机的运算过程加以简化。

3.2.2 AR 模型阶次确定

自回归模型的阶次 m 可以用经验方法加以确定。设 N 为序列 $\{x_i\}$ 中 x 的个数, 则一般可取:

当 $N = 20 \sim 50$ 时, $m = N/2$

当 $N = 50 \sim 100$ 时, $m = N/3 \sim N/2$

当 $N = 100 \sim 200$ 时, $m = (2N)/\ln(2N)$

比较科学的定阶方法是用 AIC 指标定阶, AIC 值与模型的残差和阶次 m 有关

$$AIC = N \ln(\sigma_a^2) + 2m \quad (3.12)$$

它的物理概念是: 提高模型拟合的阶次, 则残差 σ_a^2 将减小, 而阶次 m 将增大。这样 AIC 值将有一个极小值, 对应于此极小值的模型阶次可认为是最佳的模型阶次。

3.2.3 机械故障时序模型诊断法

1. 根据 AR 模型参数 φ_i 进行诊断

时间序列中蕴含了大量的系统状态信息, 通过时序模型的建立, 将其凝聚成为少数几个模型参数, 也就是说, 模型参数浓缩了系统状态的信息, 所以可以直接依据模型参数对系统状态进行识别和诊断。

利用 φ_1 在 VDF 车床上进行颤振识别试验。测取尾架顶尖处的振动加速度信号, 采样间隔 $T_s = 0.5 \mu s$, 采样数据长度 $N = 128$, 得到参考时序 $\{x_k\}_R$ 。为研究颤振从无到有的发展过程及其特征, 考虑到颤振一般在 $1 \sim 3 s$ 间隔内产生(这与具体的切削条件有关), 因此, 在远离颤振发生时, 每隔 $3.6 s$ 采样一次, 建模一次。而在临近颤振发生时, 每隔 $0.9 s$ 采样一次, 建模一次。图 3-1 示出了在颤振从无到有这一发展过程中参考 AR_R 模型参数 ϕ_{1k} 的变化规律, 图中横坐标为时间坐标, 以 $3.6 s$ 为一个单位, 则一个单位对应于采样一次, 建模一次的 ϕ_{1k} 数值。由图可见, 在远离颤振以前的 4 次采样间隔的时间 ($3.6 s \times 4 = 14.4 s$) 内, ϕ_{1k} 变化平坦; 在 4 次采样后, 颤振即将发生, ϕ_{1k} 急剧增大, 然后维持较大的值。因此, 实用中可直接将图 3-1 作为参考模式, 对实际切削过程的振动加速度信号进行监测, 并按前述相同的步骤不断得到待检时序 $\{x_i\}_T$, 并不断建立待检模型 AR_T , 一旦发现 AR_T 的模型参数 ϕ_{1T} 具有如图 3-1 所示的急剧增大趋势时, 则发出警报信号, 采取控制措施。

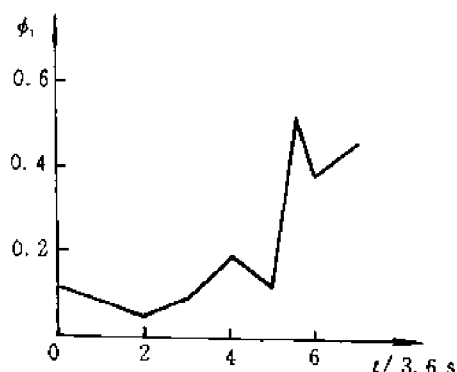


图 3-1 切削颤振识别

图 3-2 所示是 12 个时域信号, 分为正常与异常状态两类。为了区分这两类信号, 首先对这

12个信号建立AR(3)模型,然后如图所示,以模型的第一个和第二个系数 ϕ_{31} 和 ϕ_{32} 为坐标作图,将正常信号与异常信号区分开来。一般来说,采用这样的方法识别准确性不会太高,因为正常信号与各种异常信号建模时的最佳阶次不一定完全相同。在需要建立高阶自回归模型时,由自回归系数建立的是多维空间,不能直观地、迅速地识别各类信号。基于这些原因,这一方法应用较少。

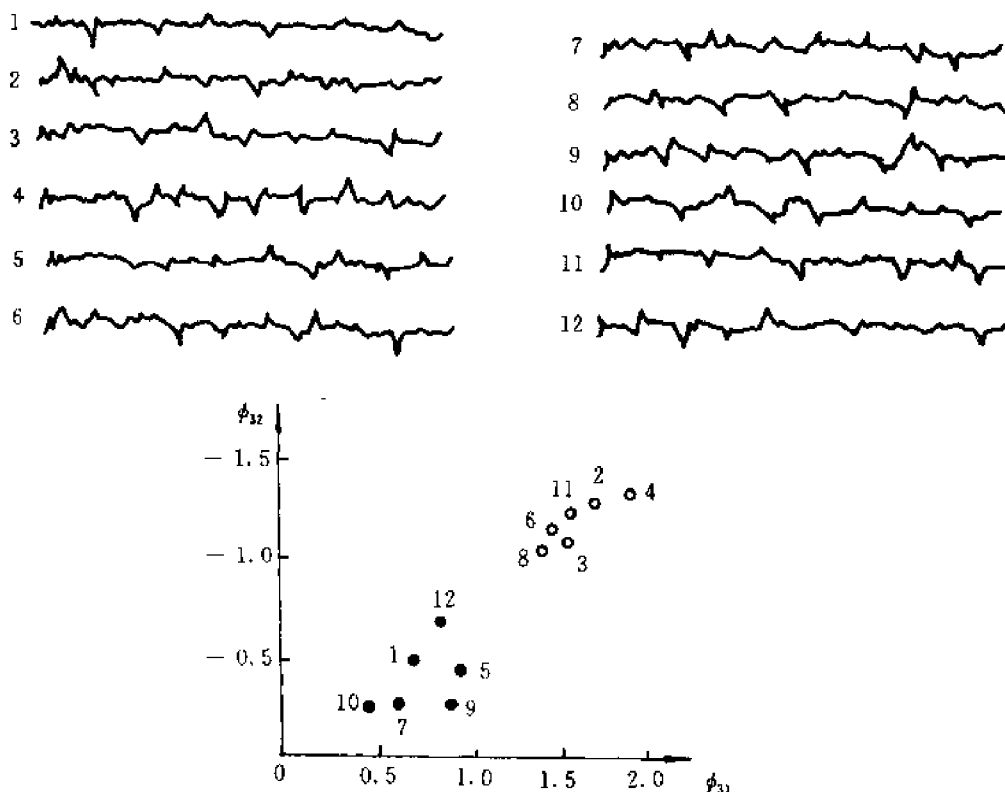


图 3-2 利用自回归系数进行识别(●-正常,○-异常)

2. 根据模型的残差方差 σ_a^2 进行诊断

在时序方法中, σ_a^2 具有不同的含义,对系统而言, σ_a^2 是系统输入白噪声 $\{a_k\}$ 的方差;对模型而言, σ_a^2 是模型的残差序列 $\{a_k\}$ 的方差。对AR模型, σ_a^2 由式(3.5)确定,当模型参数确定后, σ_a^2 亦可由模型直接算出

$$\sigma_a^2 = \frac{1}{N} \sum_{k=m+1}^N (x_k - \sum_{i=1}^m \phi_i x_{k-i} + \sum_{j=1}^n \theta_j a_{k-j})^2 \quad (3.13)$$

将 σ_a^2 直接用于故障诊断的思想是:在某一参考状态(一般是正常工况)下,取得这一参考状态的参考时序 $\{x_k\}_R$,并建立参考模型ARMA_R或AR_R,按式(3.13)或式(3.5)算得参考模型的残差方差 σ_{aR}^2 , σ_{aR}^2 即是该参考状态的特征量。当用于故障诊断时,在待检状态下,取得待检时序 $\{x_k\}_T$,再将待检时序 $\{x_k\}_T$ 代入参考模型ARMA_R或AR_R,由式(3.13)或式(3.5)算得相对于参考模型ARMA_R或AR_R的残差方差 σ_{aT}^2 ,即在式(3.13)或式(3.5)中,模型参数 ϕ, θ 仍为参考模型的参数,但其中的各 $\{x_k\}$ 值用 $\{x_k\}_T$ 代入。显然,如果待检状态与参考状态属于同一状态,则待检时序 $\{x_k\}_T$ 应满足参考模型,从而 σ_{aT}^2 与 σ_{aR}^2 应差别不大;反之,如果待检状态与参考状态不属于同一状态,则 $\{x_k\}_T$ 就不会满足参考模型,从而 σ_{aT}^2 与 σ_{aR}^2 差别甚大。例如,若取

参考状态为工况正常状态,在工况监测中,当 σ_{rT}^2 与 σ_{rk}^2 差别不大时,则认为工况正常,一旦 σ_{rT}^2 与 σ_{rk}^2 相差较大时,应认为工况异常。

下面采用残差方差来判断电机转子质量偏心是否超过给定的界限。在确定参考模型的试验中,首先使电机在正常状态下运行,测取电机振动加速度信号,并建立 ARMA_N(2,1)模型

$$x_k = 1.96x_{k-1} - 0.93x_{k-2} + a_k + 0.693a_{k-1}$$

然后,在不同的偏心载荷(见表3-1)下,对持续0.5s的电机振动加速度信号采样100个点得到 $\{x_k\}$ ($k=1,2,\dots,100$),再将 $\{x_k\}$ 代入上述参考模型,计算出 $\{x_k\}$ 相对于上述模型的残差方差

$$\sigma_e^2 = \frac{1}{N} \sum_{k=3}^N (x_k - 1.96x_{k-1} + 0.93x_{k-2} - 0.693a_{k-1})^2$$

作出 σ_e^2 的点图,如图3-3所示。图中 M 是电机在正常状态下运转时,多次试验得到的残差方差 σ_e^2 的平均值,置信限为 $M+3\sigma$, σ 为上述 σ_e^2 值的均方差,横坐标是按偏心质量的大小依次对电机编号。

表3-1 电机转子不同的质量偏心量

状态	质量/g	力/N	符号
A	正常		+
B	9.1	0.8896	×
C	27.2	4.0043	◇
D	45.4	6.6723	△
E	90.7	8.8964	×

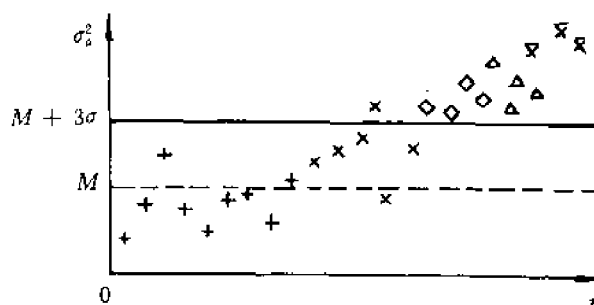


图3-3 诊断电机回转质量的偏心状态

由图可见,偏心质量越大, σ_e^2 也越大。此图作为诊断的参考模式,具体诊断时,在待检电机运行时,测取其振动加速度信号,离散采样100个数据得到待检时序 $\{x_k\}_T$ ($k=1,2,\dots,100$),将 $\{x_k\}_T$ 代入上述参考模型算得待检残差 σ_{rT}^2 ,即可根据 σ_{rT}^2 落在图3-3中的位置判断出待检电机的偏心质量大小。

3. 利用 AIC 指标进行状态识别

AIC 准则是检验 AR 模型适应性的一条极为重要的准则。当作为判别函数使用时,其思路为:若对某一状态的信号建立 AR 模型,记为模型 C,算出相应的 AIC 值,记为 AIC_C,再将该信号截成两段,对这两段信号分别建立两个模型 A 与 B,算出相应 AIC_A 和 AIC_B。对同一状态的信号,由于模型 C 使用的数据多于模型 A, B 的数据 N_A, N_B ,则应有: AIC_C < AIC_A + AIC_B,这就意味着,前后两次采样是来源于同一总体,在前后两段运行时间内系统状态没有变化。反之,

如果当上述不等式不成立时,意味着前后两次抽样源于不同的总体,系统的状态在前后两段时间内已经产生了变化。由此得到由 AIC 指标构成的判据为

$$\begin{cases} AIC_C < AIC_A + AIC_B & \text{系统状态无变化} \\ AIC_C > AIC_A + AIC_B & \text{系统状态有变化} \end{cases}$$

具体运用 AIC 指标比较时,由于采样数目 N_A, N_B 和 N_C 不等,应经过标准化,即可对上述判据中的值分别除以其数据个数,得到判据形式为

$$\begin{cases} \frac{AIC_C}{N_C} < \frac{1}{2} \left(\frac{AIC_A}{N_A} + \frac{AIC_B}{N_B} \right) & \text{系统状态无变化} \\ \frac{AIC_C}{N_C} > \frac{1}{2} \left(\frac{AIC_A}{N_A} + \frac{AIC_B}{N_B} \right) & \text{系统状态有变化} \end{cases} \quad (3.14)$$

国外有人曾经对脑电图经一定时间间隔采样后建立 AR 模型,识别人在熟睡和醒后的两种状态。每次采样数为 400 个 ($N = 400$),如图 3-4 所示。当采样数累计到 1 200 个以前时,用 AIC 指标判断,有

$$\frac{AIC_C(13)}{1\ 200} < \frac{1}{2} \left[\frac{AIC_A(13)}{800} + \frac{AIC_B(9)}{400} \right]$$

说明三次采样可以采用同一个 AR 模型,表明在该段时间内,人体状态无变化。当采样数累计到 1 200 ~ 1 600 个时,有

$$\frac{AIC_C(14)}{1\ 600 - 1\ 200} > \frac{1}{2} \left[\frac{AIC_A(13)}{1\ 200} + \frac{AIC_B(13)}{800} \right]$$

用 AIC 指标检验认为不能与前面三组数据合并成一个 AR 模型,表明在此期间,人体状态发生了变化。又当累计数在 1 600 ~ 2 800 期间时,有

$$\frac{AIC_C(12)}{2\ 800 - 1\ 600} < \frac{1}{2} \left[\frac{AIC_A(11)}{2\ 400 - 1\ 600} + \frac{AIC_B(13)}{2\ 000 - 1\ 600} \right]$$

表明在此时间内,人体在新状态下无变化。由此可见,这种模型合并和分离的情况,完全客观地反映了人在熟睡和醒后两种不同状态的变化过程。采样数累计在 0 ~ 1200 期间时,人体处于某一种状态;采样数累计在 1 200 ~ 2 800 期间时,人体处于另一种状态;采样数累计在 1200 ~ 1600 期间,是两个状态的转换过程。这一结果由图 3-4 中相应各段数据的功率谱图得到了证实;前三个谱图的形状和谱峰均相当接近,而与后三个谱图明显不同,中间的谱图又不同与此两者。这个例子有力地说明了诊断方法的客观性和有效性。

由于 AR 模型建模比较简单,计算机运行时间短,对系统状态的识别又有一定的准确性,因此适宜于在在线监控中采用。

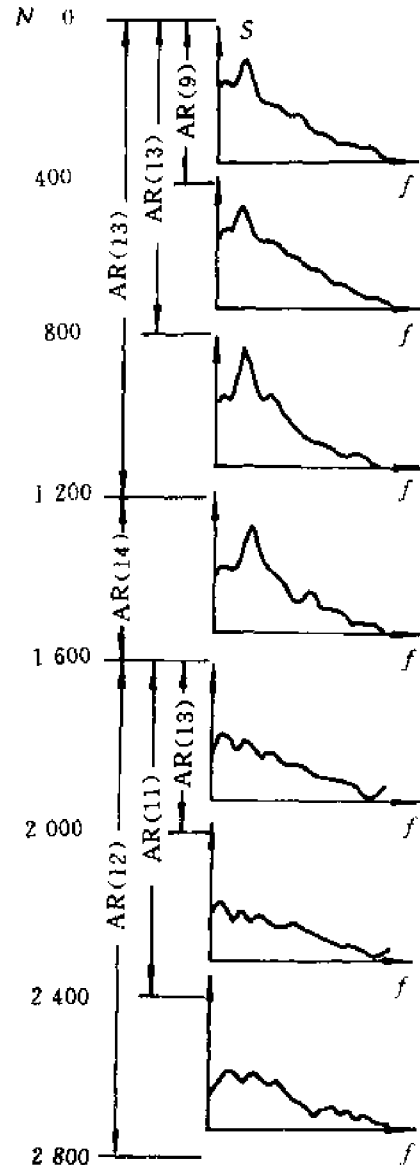


图 3-4 用 AIC 指标识别状态的变化

除了上述直接根据时序模型的个别参数或个别特征进行故障诊断外,还可将模型参数构

成模式向量,应用模式识别的方法对机械设备运行状态作出分类,详细内容将在第4章中讨论。

3.3 自回归谱的概念和应用

自回归谱是自回归时序模型经过频域变换得到的一种功率谱密度函数。自回归谱反映了一个时间序列在频域中的组成情况。因此,它是机械设备故障诊断中极为有效的工具。

3.3.1 自回归谱的概念

假定我们已经采用 Yule-Walker 方程获得了式(3.1)所示的自回归模型,并且应用 AIC 准则确定了模型的最佳阶次。这时,可以对模型作 z 变换,确定在白噪声 a_k 输入下,输出为 x_k 时的系统传递函数

$$H(z) = \frac{X(z)}{A(z)} = \frac{1}{1 - \phi_1 z^{-1} - \phi_2 z^{-2} - \dots - \phi_m z^{-m}}$$

根据系统输入、输出的自功率谱与传递函数的关系,将 $z = e^{j2\pi f T_s}$ 代入,有

$$S_x(f) = |H(e^{j2\pi f T_s})|^2 S_a(f)$$

式中, T_s 为采样间隔; $S_a(f)$ 为输入白噪声的功率谱密度, $S_a(f) = \sigma_a^2 T_s$ 。这样可以得到时间序列 $\{x_k\}$ 的自回归谱

$$S_x(f) = \frac{\sigma_a^2 T_s}{|1 - \sum_{k=1}^m \phi_k e^{-j2\pi k f T_s}|^2} \quad (3.15)$$

对于一阶自回归模型 $x_k = \phi_1 x_{k-1} + a_k$, 利用式(3.15)可以求出其自回归谱

$$S_x(f) = \frac{\sigma_a^2}{1 + \phi_1^2 - 2\phi_1 \cos 2\pi f}$$

这里,采样间隔 T_s 取为 1, 相当于采样频率 $f_s = 0.5$ 或 $\omega_s = \pi$ 。当 $\omega = 2\pi f$ 自 0 到 π 变化时,分母将单调增或单调减,视 ϕ_1 值的正负而定。

当 ϕ_1 为正时, $\omega = 0$, $S_x(f) = \max$

当 ϕ_1 为负时, $\omega = \pi$, $S_x(f) = \max$

由此可见,一阶自回归模型在谱图上形不成谱峰,如图 3-5(a) 所示。

进一步观察一个带噪声的正弦波 $x_k = A \sin k\omega_0 T_s$, 确定其自回归模型和功率谱密度函数

$$\begin{aligned} \text{因为 } x_{k-1} &= \frac{2A \sin(k-1)\omega_0 T_s \cdot \cos \omega_0 T_s}{2 \cos \omega_0 T_s} = \\ &= \frac{A \sin k\omega_0 T_s + A \sin(k-2)\omega_0 T_s}{2 \cos \omega_0 T_s} \end{aligned}$$

因此,相应的自回归模型是 AR(2)

$$x_k - (2 \cos \omega_0 T_s) x_{k-1} + x_{k-2} = a_k$$

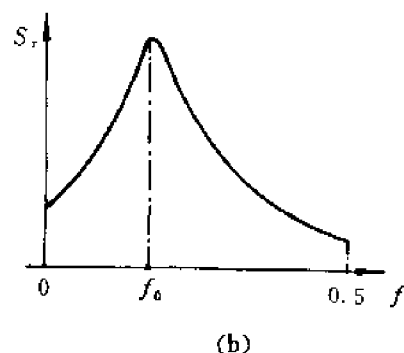
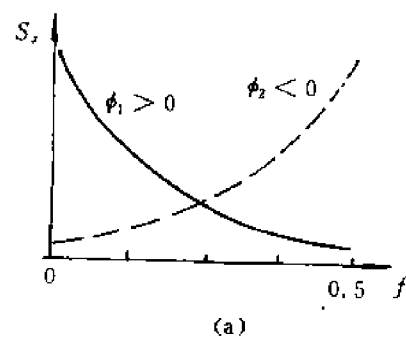


图 3-5 自回归模型的谱
(a) 一阶自回归模型谱
(b) 二阶自回归模型谱

取采样间隔 $T_s = 1$, 噪声的方差为 σ_n^2 , 经过频域变换, 可以得到 x 的功率谱密度函数

$$S_x(f) = \frac{\sigma_n^2}{4(\cos^2 \omega_0 + \cos^2 \omega) - 8\cos \omega_0 \cos \omega}$$

可以看到, 当 $\omega \rightarrow \omega_0$ 时, S_x 将会出现一个谱峰, 如图 3-5(b) 所示。

自回归谱的基本优点是:

- (1) 谱峰尖锐, 频率定位准确、清晰;
- (2) 当两个谱峰的位置十分邻近时具有很强的分辨力;
- (3) 对周期性较强的序列不要求严格按周期采样;
- (4) 在保证获得足够信息的前提下, 可以大大减少采样数目;
- (5) 整个分析工作可以在微型计算机上进行。

由于自回归谱具有上述的一系列优点, 特别是能够提供比较准确的频域信息, 对于复杂的机器运行信号, 通过自回归谱分析, 可以找出各个频率分量及其在信号中的比重, 因此宜于在故障诊断中应用。

3.3.2 自回归谱的应用

1. 电动机运转噪声的自回归谱分析

电动机在运转过程中产生的振动和噪声直接反映了电动机的工作状态。利用振动和噪声信号对电动机进行监控和故障诊断已经在国内外的一些化肥厂、轧钢厂等连续生产过程中得到采用。造成电动机噪声的原因是很多的, 主要有通风噪声、电磁噪声、轴承噪声和其他部件的机械振动声, 它们有各自的频率特性。因此, 准确地确定电动机噪声的各个频率成分和相应的幅值, 有助于对电动机运行中的故障和结构工艺上的缺陷进行诊断。

图 3-6 是精密声级计测量 Y200L-6 型电动机噪声的光线示波器记录曲线。图中(a)是测点位于电动机前上端时的记录, (b) 位于电动机左前端时的记录。由于光线示波器振子频响特性的限制, 高频分量没有保留下来。图 3-7 是对应于图 3-6(a) 记录曲线的自回归谱, 图 3-8 是对应于图 3-6(b) 中记录曲线的自回归谱。谱上标明了各个谱峰处的频率, 比较图 3-7 和图 3-8 所示的两个自回归谱, 可以看到, 对应于各个谱峰处的频率基本相同, 甚至十分接近。从谱峰的高度来看前测点上所测出的电磁噪声、机壳共振声均较左前测点所测出的结果为大。

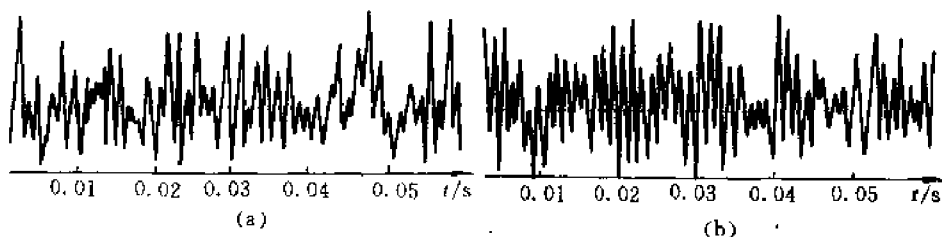


图 3-6 Y200L-6 型电动机噪声记录
(a) 正前测点 (b) 左前测点

电动机的通风噪声分为共鸣与涡流声两类。前者是由于风扇的叶片、风道、散热肋板设计不合理或者叶片与导风装置配合不善所引起, 其频率是 $f = nmz/60$ Hz, 式中 m 为风扇的叶片数目, z 为谐波次数, 一般为一、二次谐波最大, n 为电机转速。涡流声又称气体紊流声, 与叶片

的形状、尺寸、风道结构有关,其特点是频率范围在 100~3 000 Hz 的宽频带白噪声。在图 3-7 与图 3-8 所示的两个自回归谱上都没有找到通风噪声,可能的原因是风扇位于电机的后部,而上述两个测点都在电机前部的缘故。表 3-2 是对这一电动机噪声的诊断结论。

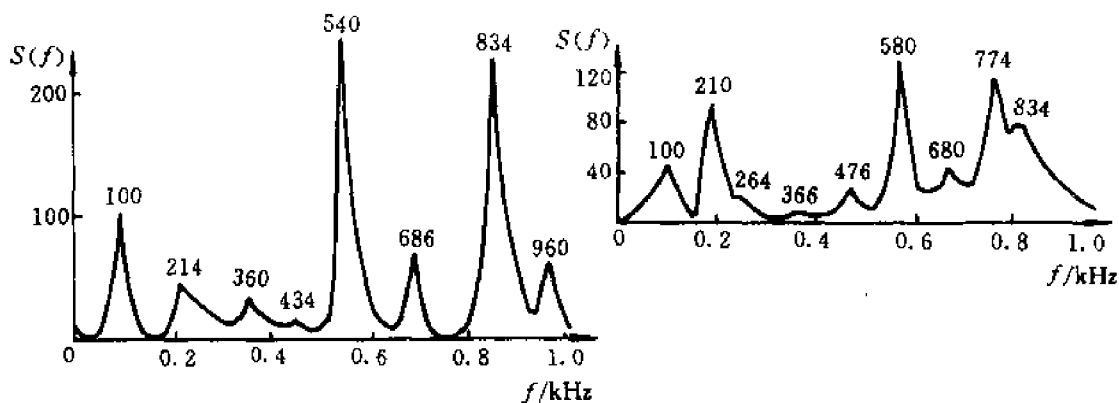


图 3-7 电动机噪声的自回归谱(在正前测点)

图 3-8 电动机噪声的自回归谱(在左前侧点)

表 3-2 Y200L-6 型电动机噪声源的诊断

正前测点(见图 3-7)		左前测点(见图 3-8)		噪声源诊断
f/kHz	$S(f)$	f/kHz	$S(f)$	
100	96.9	100	48.2	2 倍电源频率,磁极径向拉力脉动噪声
214	48.7	210	90.2	4 倍电源频率,磁极径向拉力脉动噪声
		264	21.0	后轴承轴向窜动噪声
434	17.1	476	26.3	前轴承噪声
540	242.7	580	127.7	机壳共振声
686	79.5	680	45.4	机壳共振声
		774	110.6	磁噪声
834	214.4	834	75.7	磁噪声
960	47.3			磁噪声

注:表中未列的频率分量原因不明。

2. 磨削振动的自回归谱分析

在一台装有静压主轴承的外圆磨床上测量砂轮重新修正后直到磨钝为止各个阶段中工件的振动状态。所用的量仪是涡流式测振仪,工件材料分别为 45, 55 和 T8 钢,砂轮采用粒度为 46 的中软氧化铝砂轮。图 3-9 是一组典型的记录曲线。仔细观察图中各个振动信号的变化,可以看到拍的现象从用新修的砂轮磨削时起就已经存在(见图 3-9(a)),所不同的是,在砂轮整个磨钝过程中振幅不断加大,而频率却不断降低。这种拍的现象在机器振动中经常遇到,在线性振动中可以理解为两个振幅相近、频率相近的正弦振动合成。这种形成拍的频率分量的特点是它们的频率十分接近,只有谱的分辨率十分高时,才能将它们分解出来。图 3-10 是用自回归谱

方法分离出的、频率定位准确的三个正弦分量,这在一般的数据处理机上是无法获得的。

图 3-10 所示为砂轮快要磨钝时的自回归谱。图上三个相邻谱峰的频率和相对振幅如表 3-3 所示。由于拍的形成还可以用非线性振动来说明,例如,一个以频率 $f_+ = (f_1 + f_2)/2$ 振动的物体放置在另一个以频率 $f_- = (f_1 - f_2)/2$ 振动的支架上时,也可以得到拍的现象。表 3-3 中两个 f_- 的频率 25.6Hz 和 48.4Hz,十分接近砂轮的回转频率 20.5Hz 和 50Hz 的电源频率,值得进一步加以探查。

表 3-3 磨削振动中拍的频率合成

频率 f , Hz	振幅 S /cm	$(f_1 + f_2)/2$	$(f_2 + f_3)/2$	$(f_2 - f_1)/2$	$(f_3 - f_2)/2$
$f_1 = 582.0$	0.88	630.3		48.4	
$f_2 = 678.7$	1.56	630.3	704.2	48.4	25.6
$f_3 = 729.8$	1.33		704.2		25.6

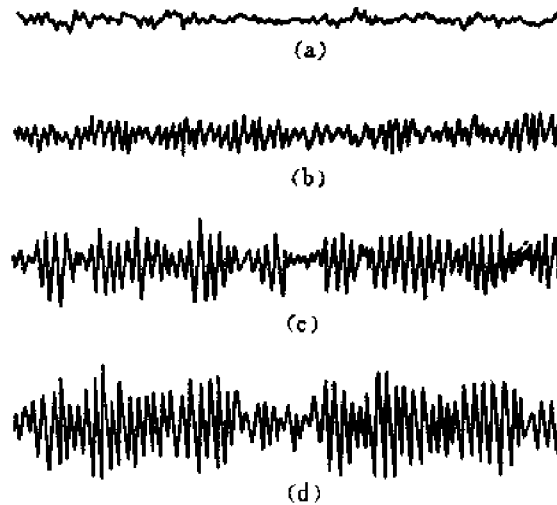


图 3-9 磨削振动的记录曲线

3. 刀具磨损状态判别的 AR 谱分析

为了研究刀具磨损状态所造成的影响,在车床上进行刀具磨损试验,测取切削过程的切削力信号,得到参考时序 $\{x_k\}_R$ 。对这一时间序列进行 FFT 变换所得频谱毛刺较多,谱峰不明显,不易进行分析。采用自回归谱分析法,建立参考时序模型 AR_R ,采用 AIC 准则检验,模型阶数 $m = 5 \sim 9$,根据 AR_R 算出 AR 谱 $S_r(f)_R$ 。图 3-11 示出了刀具磨损量分别为 0, 0.29, 0.57, 0.86 时的 AR 谱 $S_r(f)_R$,图 3-12 示出了同一刀具持续工作 40 min 期间的四个阶段的 AR 谱 $S_r(f)_R$ 。从图中可看出,每一 AR_R 模型对应有两个谱峰,第一谱峰在 265 ~ 313 Hz 之间,随着磨损量和工作时间的增加,谱峰频率向高频方向移动,谱峰幅值也逐渐增高,磨损量越大,谱峰幅值增高越快;第二谱峰在 605 ~ 615 Hz 之间,随着磨损量和工作时间的增加,谱峰频率基本不变,但谱峰幅值则逐渐减小。总的趋势是,随着磨损量和工作时间的增加,第一主峰增长,第二主峰下降,其物理意义为:切削系统的能量由高频向低频转移。根据上述 $S_r(f)_R$ 的情况,可有三种诊断方案:

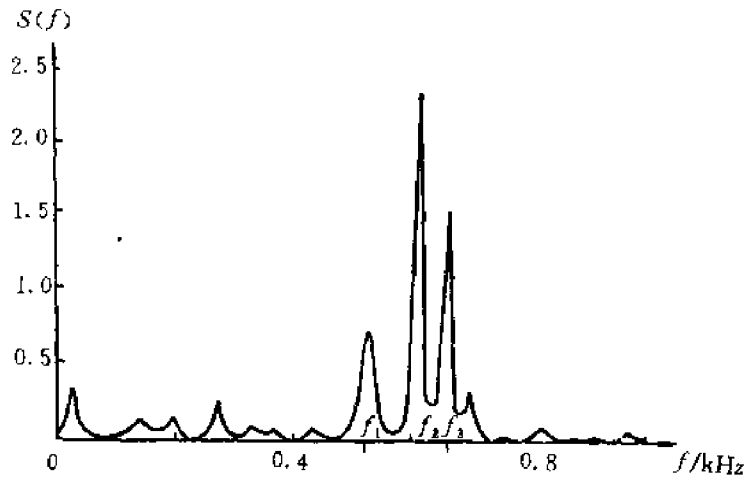


图 3-10 磨削振动的自回归谱

第一,直接采用 265 ~ 313 Hz 附近的第一主峰幅值 S_{1R} 作为判据。实际监测与诊断时,将待检 AR 谱 $S_x(f)_T$ 图中的第一主峰幅值 S_{1T} 与参考谱 $S_x(f)_R$ 图的第一主峰幅值 S_{1R} 进行比较,可以随时了解切削过程中的刀具磨损量大小,当 S_{1T} 超过刀具最大磨损量所对应的 S_{1Rmax} 时,控制系统即可发出换刀信号,实现了刀具磨损的在线监控。

第二,直接采用 605 ~ 656 Hz 附近的第二主峰幅值 S_{2K} 作为判据。监测方法同上,所不同的是由于 S_{2T} 随磨损量的增大而减小,所以,当 S_{2T} 小于刀具最大磨损量所对应的 S_{2Rmin} 时,就发出换刀信号。

第三,采用谱峰幅值比 S_{1K}/S_{2K} 作为判据,监测方法仍同上。

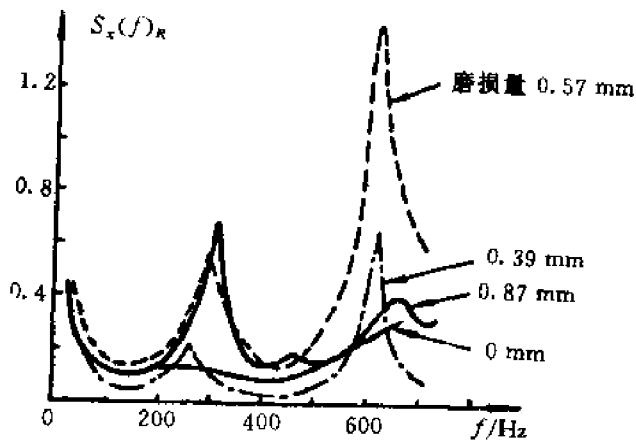


图 3-11 不同磨损量时的 AR 谱

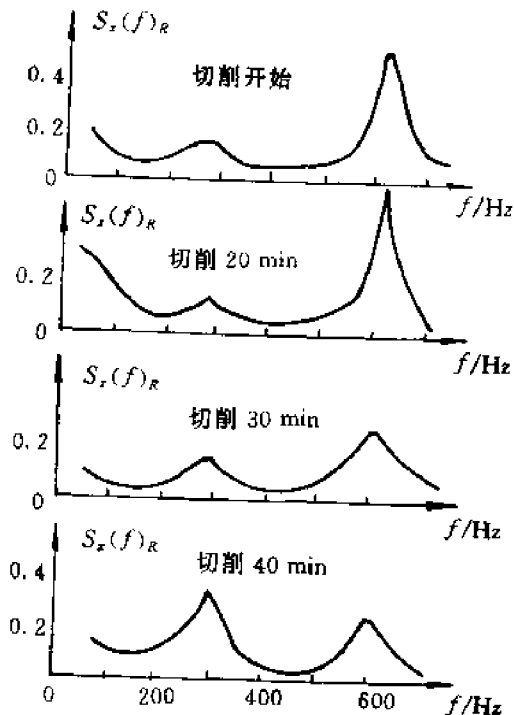


图 3-12 刀具在不同工作阶段时的 AR 谱

3.4 设备状态变化趋势性及预测

3.4.1 设备状态的趋势分析与预测技术的特点

1. 设备状态发展的规律性

很多事物的发展,差不多都与它过去的状态有关系,观察过去的状态可知它的现在,观察过去和现在的状态也会预知未来。这是因为事物的发展,都带有一定的延续性或称为“惯性”。例如某机组轴承温度在上午8时为40℃,9时为41℃,10时为43℃,11时为46℃。如果孤立地观察,可以说轴承温度都在合格范围以内,但是从所观测数据与时间的关系来看,则可说明该轴承温度在上升。以时间为横坐标,以所测数据为纵坐标绘出曲线,即可按曲线延伸来预测发展趋势。根据图3-13,由曲线延伸可以预测到13时以后轴承温度将达到50℃。如果轴承温度继续升高,很可能导致质的变化,造成轴承损坏,甚至带来更大的危害。

通过上述分析可知,设备状态预测的主要依据是设备故障的发展规律与趋势,其具有以下特点:

(1) 设备缺陷和异常的延续性:设备故障往往来自元件、部件本身或它们之间连接部位的缺陷。一般情况下缺陷既以形成就会出现缺陷的状态特征,不经修复或处理,缺陷不仅不会自行消失,而且还可能发展,发展到一定程度就会突然形成故障。掌握各种缺陷向故障发展的规律,将有助于提供有使用价值的状态预测。

(2) 设备故障的相关性:设备缺陷的发展规律,常和设备内部结构、部件之间的关联以及设备的运行条件密切相关,如高压电器局部放电逐渐增大常和内部绝缘劣化有关系,高温金属蠕变速度增大和运行温度有关系。查明故障机理,即可依据故障因、果的函数关系,建立预测模型。

(3) 设备故障的相似性:设备缺陷及其扩展变化,有时具有共同的特征。在同型设备或近似类型设备上,则有更大的近似性。搜集同型设备发生的缺陷及其扩展变化的情况,再加上被预测设备的具体运行条件,将为建立预测模型提供有利的借鉴。

2. 预测技术的特点

(1) 科学性:预测技术是按科学方法和程序建立模型,是在掌握主观和客观诸因素和各种因素相互关系的基础上,对事物发展趋势作出的判断,所以是科学的预测。

(2) 近似性:预测是对未来发展的科学估计,是对事物发展规律的认识与推演,它只是预测大略的趋向,不可能预测某台机组运行到某月、某日,甚至某时一定会损坏。预测与实际情况的发展常会有小的偏差,在数值上可能与实际略有出入,所以预测具有近似性。

(3) 局限性:由于外部各种因素变化的影响,以及人们对未来认识的局限性,预测结果常常是在指定的范围内有效,也就是预测结果不能表达事物发展的全过程。因此,预测具有一定的局限性。

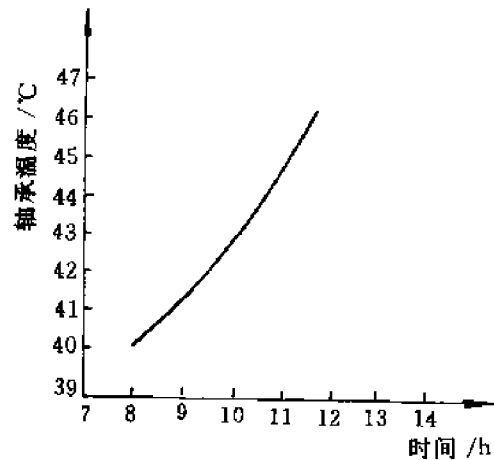


图 3-13 轴承温度发展趋势

在设备状态监测与故障诊断中,使用的预测方法较多,本节仅对时间序列预测法作一讨论。

3.4.2 时间序列预测法

时间序列通常是按小时、日、周、月、年观测事物的变化。在某些场合不是按时间观察统计,而是按温度、电流等观察统计的数据,习惯仍使用时间序列这一术语。

时间序列用于预测的基本思想是认为历史将延续到未来,即一种事物过去随时间而变化的趋势,也是今后该事物随时间而变化的趋势,预测的方法就是时间序列的外推。时间序列预测技术是通过预测目标本身时间序列的处理来研究其变化趋势的。这一变化趋势往往包含有:

(1) 长期趋势分量:它反映了事物的主要变化趋势,对于作长期的较粗略的预测是很有用的。

(2) 季节变动分量或周期变动分量:它是由事物某些局部特性引起的,对作短期预测有实际意义。

(3) 随机性变动分量:它是指由于各种事前无法预料的因素而引起的对时间序列宏观上的影响,它使测量结果产生一定的分散。

1. 确定性时间序列预测技术

确定性时间序列预测的一般步骤是:首先求出基本的发展趋势,分析可能存在的波动,再通过对随机变动的分析,确定一个合理的预测区间,然后进行预测。下面介绍几种常用的确定性时间序列预测技术。

(1) 滑动平均法:滑动平均法认为未来的状态与近期的状态有关,而远期的状态并不重要。所以该方法是不断引入新数据来修改平均值,以消除变动的偶然因素的影响,得出事物发展的主导趋势。其数学模型为

$$M_k = \frac{x_{k-1} + x_{k-2} + \cdots + x_{k-n}}{n} \quad (k > n) \quad (3.16)$$

式中, M_k 为预测值(观测序列平均值); x_k 为观测序列实际值; n 为预测资料期(滑动平均包含的观测值的个数)。

例如,设近期三个月的实测数据为 x_1, x_2, x_3 , 则预测第四个月的数据为 $M_4 = \frac{x_1 + x_2 + x_3}{3}$; 同法类推,第五个月的数据预测值为 $M_5 = \frac{x_2 + x_3 + x_4}{3}$ 。

滑动平均法预测误差的大小取决于滑动平均所包含的观测值个数 n , n 值越大,对实际值的修正作用越强,预测线越平滑,灵敏度也就越差,其结果只能反映预测事件的发展方向和趋势;反之, n 值越小,预测线接近实际值,灵敏度越高。所以 n 值应根据要求的预测精度和 x 值变化大小而选取。如果要求预测值比较准确, n 值应取小一些,可在 3~5 之间;反之,如果想得到事物变化的大致趋势, n 值可取的大一些,可在 10~30 之间。滑动平均法只适合作近期预测。

(2) 加权滑动平均预测法:在滑动平均法中,每个数据在平均中的作用是等同的,不能反映距预测期越近的数据对预测值影响越大的情况,所以把简单滑动平均法修改为加权滑动平均预测法。根据距离预测期的远近,分别赋予各个观测数据一个不同的权数,近期数据对于预测值的影响较大,其权数大些,远期数据的影响相对较小,其权数可小些。其数学模型为

$$M_k = w_{k-1}x_{k-1} + w_{k-2}x_{k-2} + \dots + w_{k-n}x_{k-n} \quad (k > n) \quad (3.17)$$

式中, M 为预测值(观测序列平均值); x 为观测序列实际值; w 为数据的权数, $w_{k-1} > w_{k-2} > \dots$; $\sum w = 1, 0 < w < 1$ 。

加权滑动平均预测法可以比较好地反映实际值的变化情况, 预测误差比滑动平均法小。

(3) 指数平滑法: 这种方法主要是强调近期数据对预测值的影响, 可以任意选择近期数据的权数, 但也不忽略远期数据的作用。所以指数平滑法是以近期的实际值和近期的预测值为依据, 经过修正后得出预测值, 不需要存储很多的历史观测数据, 它实质上也是一种加权平均法, 不过它的权数是由近期实际值和近期预测值的误差来确定的, 而且它在整个时间序列中是有规律排列的, 其数学模型为

$$M_k = M_{k-1} + \alpha(x_{k-1} - M_{k-1}) \quad (3.18)$$

式中, M_k 为第 k 期预测值; M_{k-1} 为第 $k-1$ 期预测值; x_{k-1} 为第 $k-1$ 期实际观测值; α 为平滑系数或权数 ($0 < \alpha < 1$)。

用式(3.18)计算的预测值, 其大小主要取决于近期的预测误差 ($x_{k-1} - M_{k-1}$), 以及平滑系数 α 。 α 的取值将直接影响预测精度, 当时间序列的波动较大, 不具备长期的稳定趋势时, 应取较大的 α 值, 一般取 $0.5 \sim 0.9$; 当时间序列变动缓慢, 或虽有不规则的起伏, 但长期趋势较平稳时, 取较小的 α 值, 一般取 $0.01 \sim 0.3$ 。

2. 平稳随机时间序列的预报原理

(1) 一般原理: 平稳随机时间序列预测技术不同于确定性时间序列预测技术, 它是把时间序列作为随机过程来研究的。由于考虑了时间序列的随机特征和统计特征, 所以能比确定性时间序列预测法提供更多的信息。在设备诊断技术中常用模型研究系统特性和工作状态, 预测设备状态变化的趋势。

假设 $\{x_k$ 为平稳时间序列, 以 x_k 表示 $\{x_k\}$ 在 k 时刻及其以前的观测值 $\{x_i\} (i = 1, \dots, k)$ 的记录, 若根据观测序列 x_k 对 x_{k+l} 作出某种最优意义下的估计, 则称该估计值 \hat{x}_{k+l} 为 k 时刻时间序列的 l 步预报, 记为 $x_k(l)$ 。怎样计算估计值 \hat{x}_{k+l} 认为是最优预报结果呢? 这里采用最小方差线性估计原则。用 x_1, x_2, \dots, x_k 对 x_{k+l} 作最小方差线性预报, 取

$$x_k(l) = \hat{x}_{k+l} = c_0 + \sum_{j=1}^k c_j x_j \quad (3.19)$$

式中 c_0, c_1, \dots, c_k 是常数, 选择 c_0, c_1, \dots, c_k , 使得平均平方误差达到最小, 亦即

$$e_k^2(l) = E(x_{k+l} - \hat{x}_{k+l})^2 = E(x_{k+l} - c_0 - \sum_{j=1}^k c_j x_j)^2 = \min$$

称 $e_k(l) = x_{k+l} - \hat{x}_{k+l}$ 为 l 步预报误差。或写为

$$x_{k+l} = e_k(l) + \hat{x}_{k+l} \quad (3.20)$$

也就是说, x_{k+l} 由预报值和预报误差两部分组成, 预报误差是不可预报的部分, 它包含了新的信息。

后面要讲的预报方法, 建立在下述基本引理基础上。若已经观测到平稳时间序列 x_1, x_2, \dots, x_k 的数值, 则

- 1) 将来第 $k+l$ 个时刻的白噪声估计值为 0, 即 $\hat{a}_{k+l} = 0$;
 - 2) 现在或过去第 j 个时刻平稳时间序列估计值为其观测值, 即 $\hat{x}_j = x_j, (1 \leq j \leq k)$ 。
- (2) 时间序列的 ARMA 模型预报方法: 对于如下的时间序列 ARMA(m, n) 模型其中 $\theta_0 =$

$\mu_x(1 - \varphi_1 - \dots - \varphi_m)$ 。考虑 x_{k+l} ($l = 1, 2, \dots$) 的最小方差线性预报(假设 $m > n$)，当 $l \leq n$ 时

$$x_k(l) = \hat{x}_{k+l} = \theta_0 + \sum_{i=1}^m \varphi_i \hat{x}_{k+l-i} - \sum_{j=1}^n \theta_j \hat{a}_{k+l-j} \quad (3.21)$$

当 $l > n$ 时

$$x_k(l) = \hat{x}_{k+l} = \theta_0 + \sum_{i=1}^m \phi_i \hat{x}_{k+l-i} \quad (3.22)$$

由式(3.21)和(3.22)可得时间序列的 ARMA(m, n) 预报具有如下特点:

1) 当 $l \leq n$ 时, 预报公式(3.21)中包含白噪声项 $a_k, a_{k-1}, a_{k-2}, \dots$ 白噪声序列的值是无法直接观测到的。若 ARMA 模型满足可逆性条件, 则由序列 $\{x_k\}$ 的观测值应用模型方程(3.20)可迭代计算白噪声序列的值, 但这样的计算是十分繁琐的。

2) 当 $l > n$ 时, 预报算式(3.22)中不含白噪声序列的值, 但是它包含前一步预报值, 且各步预报值之间具有式 3.22 所示的递推关系。显然, 最初的 n 步预报必然涉及白噪声序列值的递推计算。

3) AR 模型和 MA 模型预报均为 ARMA 模型预报的特例。若在模型方程和相应的预报公式中分别令 $\varphi_i = 0$ ($i = 1, \dots, m$) 或 $\theta_j = 0$ ($j = 1, \dots, n$), 则得到 AR 序列和 MA 序列的预报公式。可见, AR 模型的预报不涉及白噪声序列的值, 计算简便。

4) 由于 ARMA 模型的预报计算比较复杂, 因此工程中更多地采用 AR 模型作为预报模型。

(3) 时间序列的 AR 模型预报方法: 一个平稳时间序列的 AR 模型为

$$x_k = \theta_0 + \sum_{i=1}^m \phi_i x_{k-i} + a_k$$

已经观测到 x_1, x_2, \dots, x_k ($k > m$) 的数值, 在上式中取 $k = k+l$, 并在等式两边取估计值, 得到

$$x_k(l) = \hat{x}_{k+l} = \theta_0 + \sum_{i=1}^m \phi_i \hat{x}_{k+l-i} + \hat{a}_{k+l}$$

有基本引理, 得

$$x_k(l) = \hat{x}_{k+l} = \theta_0 + \sum_{i=1}^m \phi_i \hat{x}_{k+l-i} \quad (3.23)$$

式中 $\theta_0 = \mu_x(1 - \phi_1 - \dots - \phi_m)$ 。在预报公式(3.23)中分别取 $l = 1, 2, \dots$, 可分别得到一步、二步、... 预报值, 即

$$\begin{aligned} \text{取 } l = 1, \quad x_k(1) &= \hat{x}_{k+1} = \theta_0 + \phi_1 x_k + \phi_2 x_{k-1} + \dots + \phi_m x_{k+1-m} \\ \text{取 } l = 2, \quad x_k(2) &= \hat{x}_{k+2} = \theta_0 + \phi_1 \hat{x}_{k+1} + \phi_2 \hat{x}_k + \dots + \phi_m \hat{x}_{k+2-m} \\ \text{取 } l = 3, \quad x_k(3) &= \hat{x}_{k+3} = \theta_0 + \phi_1 \hat{x}_{k+2} + \phi_2 \hat{x}_{k+1} + \dots + \phi_m \hat{x}_{k+3-m} \\ &\dots \dots \dots \end{aligned}$$

需要指出, 在计算二步预报值时要用到一步预报值, 在计算三步预报值时要用到一步、二步预报值, 等等。

现在介绍计算一步预报误差范围的方法。由式(3.20), 有

$$e_k(1) = x_{k+1} - \hat{x}_{k+1} = a_{k+1}$$

即 k 时刻一步预报误差等于第 $k+1$ 时刻的白噪声的数值。一般情况下, 用

$$Ee_k^2(1) = Ea_{k+1}^2 = \sigma_a^2$$

刻划一步预报的精度。对正态平稳时间序列 $\{x_k\}$, 一步预报误差 $e_k(1)$ 服从正态分布, 所以

$$P\{|e_k(1)| < 2\sigma_e\} \approx 0.95$$

式中, $\sigma_e = \sqrt{\sigma_e^2}$ 。因而, 一步预报误差绝对值不超过 $2\sqrt{\sigma_e^2}$ 的概率约为 95%, 即置信概率为 0.95 的一步预报绝对误差的范围为 $2\sqrt{\sigma_e^2}$ 。用它可以判断一步预报效果的好坏。

例 某条河流上的一个水文站从 1915 年到 1973 年记录的每年最大径流量见表 3-1 y_k 栏, 共 59 个数据。要求建立 AR 模型, 并预报今后三年的年最大径流量。

表 3-4 水文记录数据

k	y_k	x_k	k	y_k	x_k	k	y_k	x_k	k	y_k	x_k	k	y_k	x_k
1	15 600	6 931	13	8 640	-29	25	9 130	461	37	10 700	2 031	49	2 340	-6 329
2	8 960	291	14	6 380	-2 289	26	7 480	-1 189	38	6 190	-2 479	50	11 100	2 431
3	10 400	1 731	15	6 810	-1 859	27	6 980	-1 689	39	9 610	940	51	5 090	-3 579
4	10 600	1 931	16	8 820	151	28	9 650	981	40	7 580	-1 089	52	10 900	2 231
5	20 800	2 131	17	14 400	5 731	29	7 260	-1 409	41	9 990	1 321	53	6 490	-2 179
6	9 880	1 211	18	7 440	-1 229	30	8 750	81	42	6 150	-2 519	54	12 600	3 931
7	9 850	1 181	19	7 240	-1 429	31	9 900	1 231	43	8 250	-419	55	6 640	-2 029
8	10 900	2 231	20	6 430	-2 239	32	7 310	-1 359	44	6 030	-2 639	56	7 430	-1 239
9	8 810	141	21	11 000	-2 331	33	9 040	371	45	8 980	311	57	6 760	-1 909
10	9 960	1 291	22	7 340	-1 329	34	7 310	-1 359	46	6 180	-2 489	58	10 000	1 331
11	12 200	3 531	23	9 260	591	35	8 850	181	47	9 630	961	59	9 300	631
12	7 510	1 195	24	5 390	-3 379	36	7 840	-829	48	4 990	821			

【解】 1) 先计算均值

$$\mu = \frac{1}{59}(15\,600 + 8\,960 + \dots + 10\,000 + 9\,300) = 8\,669$$

令 $x_k = y_k - 8669$, x_k 的数据见表 3-4 中的 x_k 栏。

2) 根据 x_k 的数据, 建立二阶自回归模型 AR(2)。利用式(3.6)和(3.7)计算模型参数估计所需的自相关函数

$$\hat{\gamma}_0 = \frac{1}{59}(6\,931^2 + 291^2 + 1\,731^2 + \dots + 1\,331^2 + 631^2) = 5\,020\,385$$

$$\hat{\gamma}_1 = \frac{1}{59}(6\,931 \times 291 + 291 \times 1\,731 + \dots + 1\,331 \times 631) = -1\,156\,994$$

$$\hat{\gamma}_2 = \frac{1}{59}(6\,931 \times 1\,713 + 291 \times 1\,931 + 1\,731 \times 2\,131 + \dots + (-1\,909) \times 631) = 1\,470\,118$$

所以

$$\hat{\rho}_1 = \frac{\hat{\gamma}_1}{\hat{\gamma}_0} = -0.23, \quad \hat{\rho}_2 = \frac{\hat{\gamma}_2}{\hat{\gamma}_0} = 0.29$$

利用 Yule-Walker 方程(3.11)估计模型参数

$$\begin{bmatrix} \hat{\phi}_1 \\ \hat{\phi}_2 \end{bmatrix} = \begin{bmatrix} 1 & \hat{\rho}_1 \\ \hat{\rho}_1 & 1 \end{bmatrix}^{-1} \begin{bmatrix} \hat{\rho}_1 \\ \hat{\rho}_2 \end{bmatrix}$$

所以

$$\hat{\phi}_1 = \frac{\hat{\rho}_1(1-\hat{\rho}_2)}{1-\hat{\rho}_1^2} = \frac{(-0.23)(1-0.29)}{1-(-0.23)} = -0.172$$

$$\hat{\phi}_2 = \frac{\hat{\rho}_2 - \hat{\rho}_1^2}{1-\hat{\rho}_1^2} = \frac{0.29 - (-0.23)^2}{1-(-0.23)} = 0.253$$

得到关于 x_k 的线性模型

$$x_k + 0.172x_{k-1} - 0.253x_{k-2} = a_k$$

将 $x_k = y_k - 8669$ 代入上式,得到关于 y_k 的线性模型

$$(y_k - 8669) + 0.172(y_{k-1} - 8669) - 0.253(y_{k-2} - 8669) = a_k$$

化简得

$$y_k = 7966 - 0.172y_{k-1} + 0.253y_{k-2} + a_k$$

3) 由上式 AR(2) 的模型方程,得到预报公式为

$$\hat{y}_{k+i} = 7966 - 0.172\hat{y}_{k+i-1} + 0.253\hat{y}_{k+i-2}$$

利用 $y_{73\text{年}} = 9300, y_{72\text{年}} = 10000$,得到

$$\begin{aligned}\hat{y}_{74\text{年}} &= 7966 - 0.172z_{73\text{年}} + 0.253z_{72\text{年}} = \\ &7966 - 0.172 \times 9300 + 0.253 \times 10000 = 8896\end{aligned}$$

又

$$\begin{aligned}\hat{y}_{75\text{年}} &= 7966 - 0.172\hat{y}_{74\text{年}} + 0.253z_{73\text{年}} = \\ &7966 - 0.172 \times 8896 + 0.253 \times 9300 = 8789\end{aligned}$$

又

$$\begin{aligned}\hat{y}_{76\text{年}} &= 7966 - 0.172\hat{y}_{75\text{年}} + 0.253\hat{y}_{74\text{年}} = \\ &7966 - 0.172 \times 8789 + 0.253 \times 8896 = 8705\end{aligned}$$

于是得到了 74,75,76 年年最大径流量的预报值。

4) 讨论一步预报误差

$$\begin{aligned}\sigma_a^2 &= E(x_{k+1} - \phi_1 x_k - \phi_2 x_{k-1})^2 = \hat{y}_0 - \phi_1 \hat{y}_1 - \phi_2 \hat{y}_2 = \\ &5020385 - 0.172 \times 1156994 - 0.253 \times 1470118 = 4453433\end{aligned}$$

因而 $e_k(1) = a_{k+1} = 2\sqrt{\sigma_a^2} = 4221$,一步预报误差范围是 4221,预报效果不好。

习题与思考题

3-1 平稳序列 $\{x_k\}$ 的样本自相关系数如下 ($\mu_y = 0.03, r_0 = 3.34$):

k	1	2	3	4	5
ρ_k	-0.800	0.670	-0.518	-0.390	-0.310

假定模型识别为 AR(1),试求 y_k 的模型方程和 σ_a^2 的值。

3-2 平稳序列 $\{x_k\}$ 的样本自相关系数如下 ($\mu_y = 0.09, r_0 = 1.15$):

k	1	2	3	4	5
ρ_k	0.427	0.475	0.169	0.253	0.126

假定模型识别为 AR(2),试求 y_k 的模型方程和 σ_a^2 的值。

- 3-3 自回归过程 $x_k = 0.8x_{k-1} + a_k$ $a_k \sim \text{NID}(2,4)$,
求:(1) x_k 的均值;(2) x_k 的方差;(3) $\tau=3$ 时的自相关系数 ρ_τ ;(4) 功率谱密度函数 $S_x(f)$ 。
- 3-4 时域故障识别如何用 AIC 指标进行?
- 3-5 自回归谱用于机械故障诊断有什么优点?
- 3-6 试设计应用自回归谱分析电动机产生噪声原因的检测、分析系统。
- 3-7 根据 AR 模型参数 $\varphi_i (i=1,2,\dots,m)$ 进行故障诊断应如何实施?其特点是什么?
- 3-8 如何根据模型残差方差 σ_a^2 进行故障诊断?
- 3-9 平稳序列 $\{y_k\}$ 的线性模型为 $y_k = 0.05 - 0.8y_{k-1} + a_k$, 而 $\sigma_a^2 = 1.2$ 。已知观测值 $y_{100} = 3.2$, 试用递推法求预报值 $y_{100}(1), y_{100}(2), y_{100}(3)$, 并求置信概率为 95% 的一步预报绝对误差的范围(假定正态平稳序列)。
- 3-10 平稳序列 $\{y_k\}$ 的线性模型为 $y_k = -0.34 + a_k + 0.62a_{k-1}$, 而 $\sigma_a^2 = 0.96$, 利用观测值 y_1, y_2, \dots, y_{50} , 算得 $a_{50} = 1.26$, 试用递推法求预报值 $y_{50}(1), y_{50}(2), y_{50}(3)$ 。

第 4 章

状态识别与判决方法原理

4.1 概 述

状态监测与故障诊断技术的方法很多,其实质是模式识别与分类问题,即根据机械设备的运行信息来识别机械设备的有关状态。从模式识别的技术角度看,状态识别与分类是由模式空间经过特征空间到类型空间的演化过程,如图 4-1 所示。将机械设备作为被监测对象,它们可用适当选择的、足够的函数来描述,或者说它们在物理上是可测量的,这些可测数据的维数一般来说是无限多的。

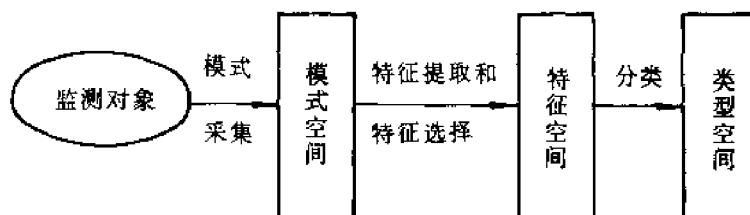


图 4-1 模式空间、特征空间和类型空间的转换

在可测数据的集合中,适当选择一些能反映机械设备运行状态的测量参数,这些测量数据构成了观察样本,对样本分别进行观测数据的综合就构成模式,所有的观察样本数据则构成模式空间,显然,模式空间的维数与选择的样本和测量方法有关,也与特定的应用有关,一般说来是很大的,但都是一个有限值。在模式空间里,每个样本都是一个点,点的位置由该模式在各维上数据来确定,由可测数据集合到模式空间的过程称为模式采集。

模式空间的维数虽多,但有些并不能揭示样本的实质,对模式空间里的各坐标元素进行综合分析,获取最能揭示样本属性的观测量作为主要特征,这些主要特征就构成特征空间,显然特征的维数大大压缩了,由模式空间到特征空间所需的综合分析,往往包含适当的变换和选择,称之为特征提取与选择。

由某些知识和经验可以确定分类准则,称之为判决规则。根据适当的判决规则,把特征空间里的样本区分成不同的类型,从而把特征空间塑成了类型空间,类型空间里不同的类型之间的分界面,常称为决策面,类型空间的维数与类型的数目相等,一般情况是小于特征空间的维数,由特征空间到类型空间所需要的操作是分类判决。

从被监测对象可测数据空间,通过模式空间、特征空间到类型空间,经历了模式采集、特征提取和选择,以及分类判决等完整的模式识别过程,可以用图 4-1 形象地表达出来。

在故障诊断中,机械设备的不同状态是不同的模式类,所以机械设备状态监测与故障诊断是一个典型的模式识别系统的设计和实现过程。识别系统设计又称为训练过程或学习过程,是指用一定数量的样本进行分类器的设计,把所研究系统的状态分为若干类,在故障诊断中就是确定系统的各参考模式类;识别系统的实现又称为识别过程,是指利用所设计的分类器对待识别样本(或称待检测状态)进行分类决策,判断待检状态应属于哪一类。所以故障诊断中的模式识别的本质工作就是:如何通过对机械设备外部征兆的监测,取得特征参数的正确信息进行分析和识别。因此研究适合于机械设备状态识别特点与要求的识别理论是解决问题的关键,本章介绍几种常用识别与判决方法的原理及其应用。

4.2 主分量分析

4.2.1 基本概念

主分量分析是用较少数量的特征对样本进行描述以达到降低空间维数,又尽可能多地保持原有分类信息的方法,它既可在时域识别中使用,又可在频域识别中使用。假定有一特征向量 X 由两个分量 x_1 和 x_2 组成,相应的有 N 个试验点

$$\begin{aligned} x_1 &= x_{11}, x_{12}, \dots, x_{1N} \\ x_2 &= x_{21}, x_{22}, \dots, x_{2N} \end{aligned}$$

图 4-2 所示。现在需要寻找一个新的坐标系 D_1, D_2 , 使全部样本点投影到新的坐标 D_1 上的分量弥散为最大,即方差为最大。这样在 D_1 方向上就保存了原来样本最多的信息量,亦即有可能用一个分量来代替原来的两个分量。由此可见,主分量分析实质上是作一线性变换,使原来的坐标系旋转到主分量方向

$$\begin{aligned} y_1 &= a_{11}x_1 + a_{12}x_2 \\ y_2 &= a_{21}x_1 + a_{22}x_2 \end{aligned}$$

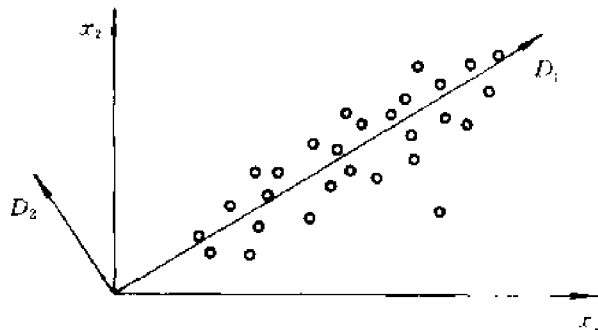


图 4-2 将坐标系旋转到主分量方向

现设机械设备的某一状态由 n 维模式向量 $X = \{x_i\} (i = 1, 2, \dots, n)$ 描述,它对应于原始空间中的一点,且存在一正交函数集 $A = \{A_j(i), i, j = 1, 2, \dots, n\}$,使得

$$Y = \sum_{j=1}^n x_j A_j = AX \quad (4.1)$$

式中 $Y = (y_1, y_2, \dots, y_n)^T$, 求 Y 的转置矩阵,有

$$Y^T = X^T A^T \quad (4.2)$$

将式(4.1)与(4.2)相乘并取数学期望

$$\begin{aligned} Y^T &= X^T A^T \\ E[YY^T] &= AE[XX^T]A^T \\ C_y &= AC_x A^T \end{aligned} \quad (4.3)$$

其中 C_x, C_y 分别为 X 和 Y 的协方差矩阵,适当选择变换矩阵 A ,可以使各分量 y_i 间相互独立, $y_1, y_2, \dots, y_j, \dots, y_n$ 两两之间的协方差为零,而使 C_y 成为对角矩阵,即

$$C_y = \text{diag}(\sigma_1^2, \sigma_2^2, \dots, \sigma_n^2) = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n) \quad (4.4)$$

消除了原有向量 X 的各分量之间的相关性,实现了将 X 变换成 Y 的主分量分析,每个 y_i 称为一个主分量,它是综合原信号 X 的性质而形成的一种具有代表某种故障信息的新特征。

由于 C_x 是实对称矩阵, C_y 是对角阵,变换矩阵 A 是正交矩阵,所以关系式 $A^T = A^{-1}$ 成立,即变换矩阵的转置与其逆矩阵相等,这时有

$$C_x = A^T C_y A \quad (4.5)$$

所以在式(4.4)中, $\lambda_1, \lambda_2, \dots, \lambda_n$ 为 C_x 的特征值,且 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$;变换矩阵 A 的列向量 A_i 是 C_x 的特征向量, $A = (A_1, A_2, \dots, A_n)$ 。

根据协方差的概念, λ_i 等于 Y 中的第 i 个分量的方差,从而 $\sum_{i=1}^n \lambda_i$ 则代表了整体的方差。因此,可按特征值计算比值 $\lambda_i / \sum_{i=1}^n \lambda_i$,它反映了 Y 中的第 i 个分量对整体方差的贡献率,贡献率越大,该分量就越重要,一般从 $\lambda_1, \lambda_2, \dots, \lambda_n$ 中选取前 m ($m < n$) 个,当

$$\eta = \sum_{i=1}^m \lambda_i / \sum_{i=1}^n \lambda_i > 85\%$$

可以只选取这 m 个分量组成新特征向量 $Y = (y_1, y_2, \dots, y_m)^T$,取代原始模式向量 X ,向量 Y 所对应的空间是向量 X 所对应空间的子空间,由此可见,主分量分析实质上是作一线性变换,使原来的坐标旋转到主分量方向,得到一个子空间,其坐标是 D_1, D_2, \dots, D_m ($m < n$),当原始故障样本 X 投影到此子空间的坐标上以后,其投影分量的方差为最大,亦即 A_i 使 Y_i 的方差达最大,这样在 A_1 方向上就保存了原来故障向量最多的信息量, A_2 方向的信息量次之,等等。从信息论的角度看,也就是说 X 的各样本投影到此子空间后投影分量的平均信息量,即熵最大。

进行主分量分析的步骤是:

- (1) 将原始样本数据对其均值标准化;
- (2) 计算协方差矩阵 C_x ;
- (3) 求 C_x 的特征值 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$;
- (4) 求出 λ_i 对应的特征向量 A_i ;
- (5) 将 A_i 标准化,使 $|A_i| = 1$;

(6) 在 n 个主分量方向中选取 $m < n$ 个主分量方向,这时,经过特征抽取后保存下来的信息量为 $\eta = \sum_{i=1}^m \lambda_i / \sum_{i=1}^n \lambda_i$; 如果 η 足够大,则说明经过主分量分析和特征抽取后仍然保存了足够的信息。

例如,由两个特征 x_1 和 x_2 组成的特征向量 X 具有协方差矩阵

$$C_x = \begin{bmatrix} 604.4 & 561.6 \\ 561.6 & 592.5 \end{bmatrix}$$

时,可求出其特征值 $\lambda_1 = 1160.139, \lambda_2 = 36.875$;相应的特征向量为 $A_1 = (0.710, 0.703)$; $A_2 = (-0.703, 0.710)$;由 A_1 主分量方向保存的信息量为

$$\eta = \frac{\lambda_1}{\lambda_1 + \lambda_2} = \frac{1160.139}{1160.139 + 36.875} = 97\%$$

这样在 D_1 方向上就保存了原样本最多的信息量,而与之垂直的 D_2 只具有次要作用,所以选择 $Y = y_1$ 作为主特征向量.经过正交变换后的新协方差矩阵为

$$C_y = \begin{bmatrix} 1160.139 & 0 \\ 0 & 36.875 \end{bmatrix}$$

相应的新主分量为

$$Y = \begin{bmatrix} 0.710 & 0.703 \\ -0.703 & 0.710 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

即

$$y_1 = 0.710x_1 + 0.703x_2$$

$$y_2 = -0.710x_1 + 0.710x_2$$

4.2.2 主分量分析应用实例

1. 特征压缩与故障诊断

作为主分量分析在机械故障诊断中应用的一个实例,是对 4135 柴油机故障的识别.柴油机以 750 r/min 的速度运行,燃爆压力为 25×10^5 Pa,模拟了 4 种状态:正常运行的(A),活塞环破坏的(B),连杆轴座间隙超限的(C),汽缸套移位(D).办法是将振动加速度传感器放置在柴油机外壳靠近活塞上止点处,记录每种运动状态下的振动加速度信号,然后对其进行频域转换,求出功率谱如图 4-3 所示.这些频谱很相似,难以区别.现在 0 到 6 Hz 频带中找出与 15 个特征频率相对应的功率谱,从而将每个功率谱用一个 15 维的特征向量表示,通过主分量分析,将原来 $n = 15$ 个特征向量所组成的特征空间变换到 $m = 2$ 和 $m = 3$ 个主分量所组成的特征空间,当 $m = 2$ 时,贡献率 $\eta = 96.3\%$,当 $m = 3$ 时,贡献率 $\eta = 96.3\%$,几乎保留了原特征空间的全部信息.图 4-4 给出了用两个主分量对上述 4 类柴油机运动状态进行识别的结果,这样检查了 40 个样本记录,没有出现误诊断.

2. 利用主分量分析将故障信息分离

主分量分析与机械故障信息的分离有着内在的联系.它能从多故障信号并存的原始机械信号中提取一种故障或多种完全故障信息.

设原始信号中包含多种故障为 $F_i (i = 1; d)$,各故障之间的相关性为 $R_{i,j} (i, j = 1; d)$;且

$$0 \leq R_{i,j} < 1 \quad i \neq j$$

$$R_{i,j} = 1 \quad i = j$$

如果经主分量分析后的两组短数据信号 S_1, S_2 中包含有同一故障信息,则有

$$S_1 \cap S_2 = F_{12} \neq \Phi$$

$$R(S_1, S_2) \neq 0$$

其中 Φ 表示空集.现所获主分量为 $\{y_i, i = 1; d\}$,每一 y_i 与其对应的故障集 F_{y_i} 之间存在一定的关系

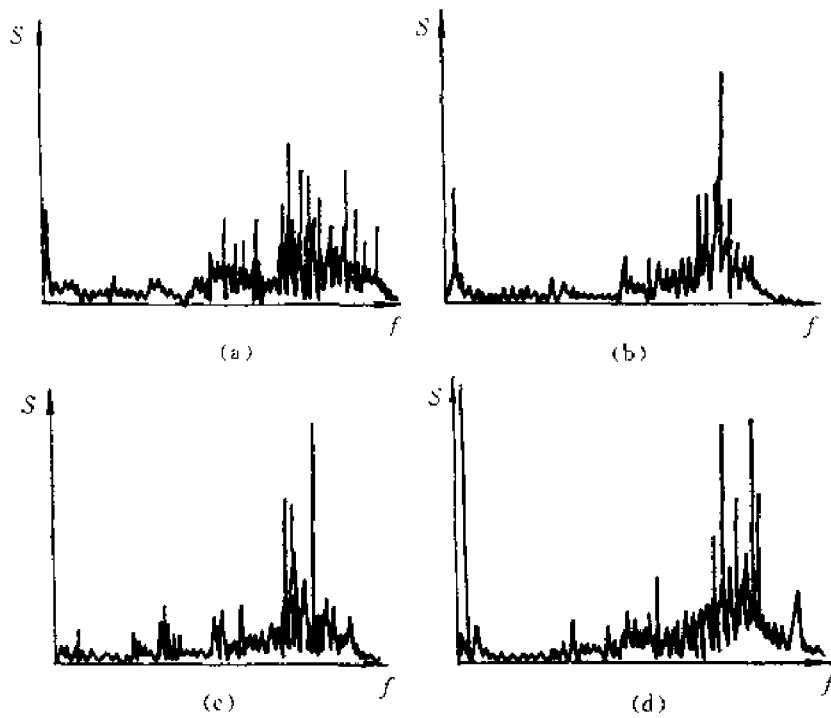


图 4-3 柴油机功率谱密度

(a) 正常 (b) 活塞环破坏 (c) 连杆轴座间隙超限 (d) 汽缸套移位

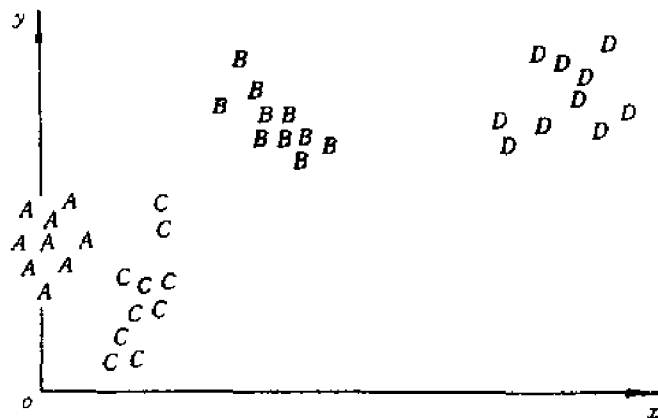


图 4-4 柴油机四种状态的两个主分量的分析

$$y_i = f(F_{y_i})$$

F_{y_i} 为一集合, 中间包含有某些故障, 由于 y_i 之间互不相关, 有

$$R(y_i, y_j) = 0 \quad i \neq j$$

即

$$R[f(F_{x_i}), f(F_{y_j})] = 0$$

亦即

$$F_{x_i} \cap F_{y_j} = \Phi$$

由此可见 y_i, y_j 中包含着不同的故障信息, 同一故障信息不可能分布在不同的主分量中。换句话说, 采用主分量分析将原故障信号中的线性相关, 变为线性独立, 原故障信号中各种故障信

息的共同作用、相互影响,随之得到了分离。

例如对齿轮箱中一对齿轮齿面磨损和轮齿产生裂纹的故障信号进行故障分离与诊断。图 4-5 为齿轮箱振动加速度信号的 FFT 谱图,图 4-6(a) 和图 4-6(b) 为一阶及二阶主分量的自回归谱(AR 谱)估计。

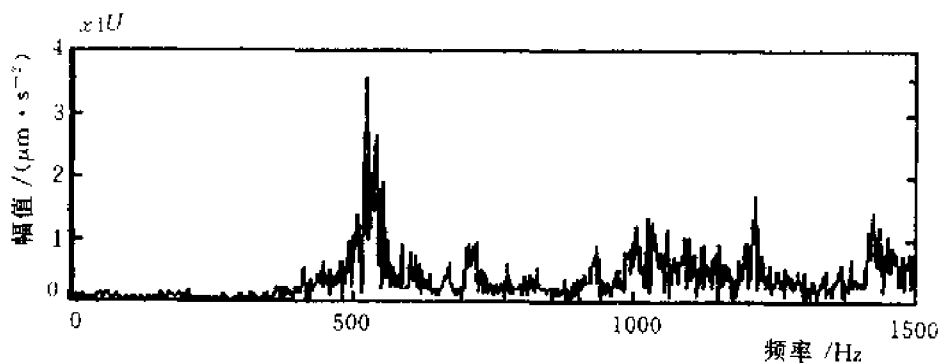
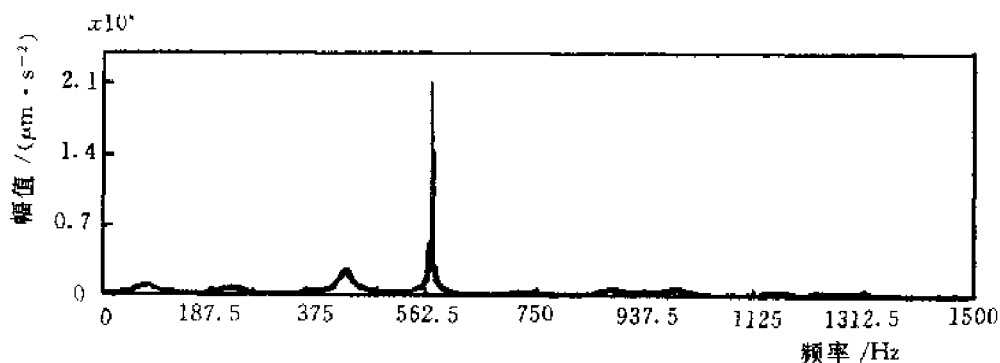
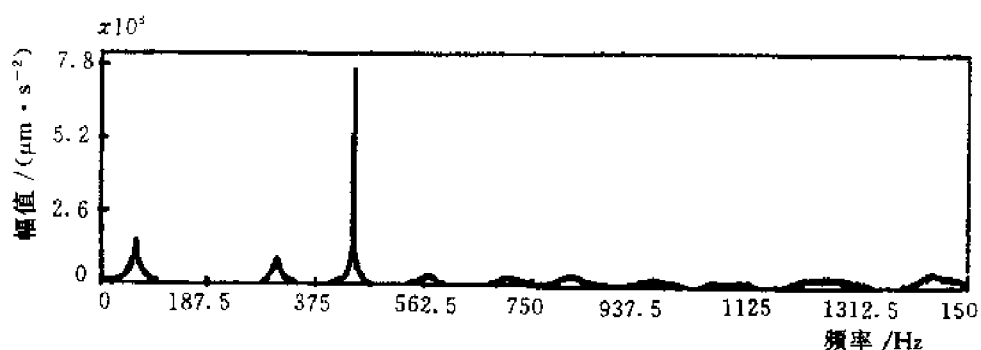


图 4-5 齿轮振动信号 FFT 谱



(a)



(b)

图 4-6 自回归谱(AR 谱)

(a) 一阶主分量 AR 谱 (b) 二阶主分量 AR 谱

从图 4-5 看到,原始振动信号的 FFT 谱很复杂,它是齿轮轴的转动频率及其谐频、齿轮的啮合频率及其谐频,以及齿轮自身各阶固有频率的某种组合振动引起的。由于频谱边频带的数量大且相互重叠,用此频谱图是很难识别出各特征频率来的。

采用主分量—AR 谱分析,首先应用主分量分析对原始振动加速度信号进行故障分离,根据累计方差贡献率 $(\sigma_1^2 + \sigma_2^2 + \dots + \sigma_m^2) / (\sigma_1^2 + \sigma_2^2 + \dots + \sigma_n^2)$,即信息量足够大原则,在 n 个主分量方向中选取 $m < n$ 个主分量方向,将原始向量 X 分别在 m 个主分量方向上投影,每个 σ_i^2 反映了原始向量 X 在相应主分量方向上投影的能量大小,而后再对各主分量方向上的信号投影分量分别作 AR 谱分析,从中找出振动频率的组成成分。

基于上述思想,通过对时域振动信号进行主分量分析,提取特征值后,得到一阶主分量 AR 谱如图 4-6(a) 所示,从图中看到,能量集中于单一峰值上,其对应频率约为 570 Hz,与故障齿轮的啮合频率(齿轮齿数为 46,转速为 750 r/min)基本相符,说明该齿轮齿面严重损伤。图 4-6(b) 的二阶主分量 AR 谱中,450 Hz 处的最高谱峰,由经验公式计算和类比,判断为齿轮啮合时的固有频率,它是由故障齿轮引起的脉冲激发所致。另外,由于齿轮等间隔两轮齿产生裂纹,在故障齿异常啮合时发生较大的冲击振动,冲击的重复频率为齿轮轴的转动频率的 2 倍,图中 50 Hz 处的频谱谱线是冲击频率的低阶谐波,从而证实了齿轮存在的缺陷。

4.3 贝叶斯决策理论方法

4.3.1 贝叶斯公式及应用

贝叶斯决策理论是统计模式识别方法中的一个基本方法,用这个方法进行分类时要求:各别类总体的概率分布是已知的;要决策分类的类别数是一定的。贝叶斯分类法是以概率密度函数为基础来描述工况状态的变化。

假设机械设备运行状态有 n 种特征观测量 x_1, x_2, \dots, x_n ,这些特征所有可能的取值范围构成了 n 维特征空间,称 $X = [x_1, x_2, \dots, x_n]^T$ 为 n 维特征向量,符号 T 表示转置。要识别分类问题有 m 个类别,各别类状态用 $w_i (i = 1, 2, \dots, m)$ 来表示,故状态空间可写成, $\Omega = (w_1, w_2, \dots, w_m)$,对应于各别类 w_i 出现的先验概率 $P(w_i)$ 及类条件概率 $P(X/w_i)$ 是已知的,如图 4-7 所示,则有

$$P(w_i/X) = \frac{P(X/w_i)P(w_i)}{\sum_{j=1}^m P(X/w_j)P(w_j)} \quad (4.6)$$

得到的条件概率 $P(w_i/X)$ 称为状态的后验概率,即状态的模式向量 X 属于状态空间类 w_i 的概率。这就是著名的贝叶斯公式,可见贝叶斯公式实质上是通过观测 X 把状态的先验概率 $P(w_i)$ 转化为状态的后验概率 $P(X/w_i)$,如图 4-8 所示。

下面举例说明贝叶斯公式的应用。

例 4-1 设定一个故障为 w ,一个征兆为 x 。其他所有故障为 \bar{w} ,其他所有征兆为 \bar{x} 。在征兆 x 发生的情况下,假设征兆必须是由故障 w 引起的,则根据式(4.6)可知,征兆 x 存在时故障 w 发生的概率为

$$P(w/x) = \frac{P(x/w)P(w)}{P(x/w)P(w) + P(x/\bar{w})P(\bar{w})}$$

式中, $P(w)$ 为故障 w 发生的先验概率;

$P(x/w)$ 为故障 w 发生引起征兆 x 发生的概率;

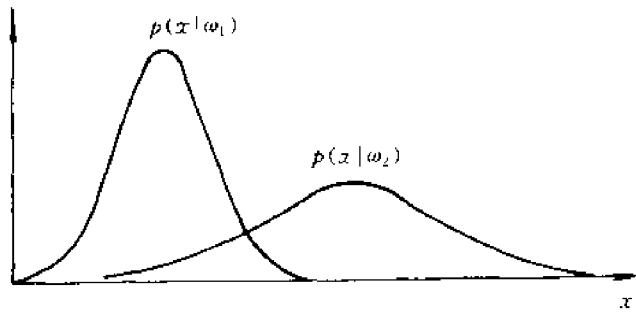


图 4-7 类别条件概率密度函数

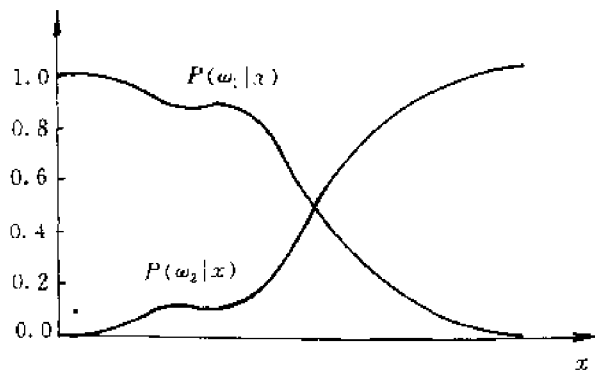


图 4-8 状态的后验概率

$P(\bar{w})$ 为其他故障 \bar{w} 发生的先验概率;

$P(x/\bar{w})$ 为其他故障 \bar{w} 发生引起征兆 x 发生的概率。

由上式可知:即使在 $P(x/w)$ 高, $P(x/\bar{w})$ 低的情况下,若 $P(w)$ 小,则 $P(w/x)$ 仍较低,如 $P(x/w) = 0.95, P(x/\bar{w}) = 0.1, P(w) = 0.01, P(\bar{w}) = 0.15$, 则

$$P(w/x) = P(w/x) = \frac{0.95 \times 0.01}{0.95 \times 0.01 + 0.1 \times 0.15} = 0.39$$

该例说明,即使故障 w 引起征兆 x 出现的可能性很大,且其他故障的 \bar{w} 引起征兆 x 出现的可能性小,如果故障 w 发生的可能性小,而其他故障发生的可能性相对较大,则在出现征兆 x 的情况下,故障 w 发生的可能性仍较小。

可见,这一分析结论与人的直觉认识(在故障 w 发生引起征兆 x 的可能性大,而其他故障 \bar{w} 引起征兆 x 可能性小的情况下,若征兆 x 出现,则存在故障 w 的可能性大)是不同的,究其原因,在于各种故障发生的概率不同,一般来说,被诊断对象的各种故障发生的可能性不是一成不变的,因此,可以作出如下结论:诊断过程中必须考虑被诊断对象的运行的历史状况,以期准确地获取各种故障发生的概率变化情况,并为各种故障确定准确的先验概率。

4.3.2 基于最小错误率的贝叶斯决策规则

在模式分类问题中,人们往往希望尽量减少分类的错误,从这样的要求出发,利用概率论中的贝叶斯公式,就能得出使错误率为最小的分类规则,称之为基于最小错误率的贝叶斯

决策。

以两类(机器工况正常 w_1 和异常 w_2) 识别问题为例,基于最小错误率的贝叶斯决策规则为:如果 $P(w_1/X) > P(w_2/X)$ 则把 X 归类于正常状态 w_1 ,反之, $P(w_1/X) < P(w_2/X)$,则把 X 归类于异常状态 w_2 ,上面的规则可简写为

$$\text{如果 } P(w_i/X) = \max_{j=1,2} P(w_j/X) \quad \text{则 } X \subseteq w_i \quad (4.7)$$

利用贝叶斯公式(4.6)还可以得到以下几种贝叶斯决策规则的等价形式

$$(1) \text{ 如果 } P(X/w_i)P(w_i) = \max_{j=1,2} P(X/w_j)P(w_j), \quad \text{则 } X \subseteq w_i \quad (4.8)$$

$$(2) \text{ 如果 } L(x) = \frac{p(X/w_1)}{p(X/w_2)} > \text{ 或 } < \frac{P(w_2)}{P(w_1)} \quad \text{则 } X \subseteq w_1 \quad \text{或} \quad X \subseteq w_2 \quad (4.9)$$

(3) 对式(4.9)取对数;并冠以负号,则有

$$\text{如果 } h(X) = -\ln[L(X)] = -\ln p(X/w_1) + \ln p(X/w_2) < \text{ 或 } > \ln \frac{P(w_2)}{P(w_1)}$$

$$\text{则 } X \subseteq w_1 \quad \text{或} \quad X \subseteq w_2 \quad (4.10)$$

式(4.8)是利用贝叶斯公式代入(4.7)消去共同的分母而得出的,(4.9)中的 $L(X)$ 在统计学中称为似然比,而 $P(w_2)/P(w_1)$ 称为似然比阈值, $p(X/w_i)$ 是状态特征观测值 X 的类条件概率密度。

例 4-2 假设某机械设备正常状态 w_1 和异常状态 w_2 两类的先验概率分别为 $P(w_1) = 0.9$ 和 $P(w_2) = 0.1$,现有一待诊断状态,其特征观测值为 X 。从类条件概率密度分布曲线上查得 $P(X/w_1) = 0.2$, $P(X/w_2) = 0.4$,试对该状态 X 进行分类。

【解】 利用贝叶斯公式,分别计算出 w_1 及 w_2 的后验概率

$$P(w_1/X) = \frac{P(X/w_1)P(w_1)}{\sum_{j=1}^2 P(X/w_j)P(w_j)} = \frac{0.2 \times 0.9}{0.2 \times 0.9 + 0.4 \times 0.1} = 0.88$$

$$P(w_2/X) = 1 - P(w_1/X) = 0.182$$

根据贝叶斯决策规则式(4.7),有 $P(w_1/X) = 0.88 > P(w_2/X) = 0.182$ 所以决策结果把 X 归类于正常状态。

从这个例子可见,决策结果取决于实际观测到的类条件概率密度 $P(X/w_i)$ 和先验概率 $P(w_i)$ 。该例中,由于正常状态 w_1 的先验概率较异常状态 w_2 的先验概率大好几倍,使先验概率在作出决策中起了主导作用。

贝叶斯决策分类可能发生误诊断,一为“谎报”,即将正常状态 w_1 误判为异常状态 w_2 ,另一种是“漏检”,就是把异常状态 w_2 误判为正常状态 w_1 ,误诊概率是谎报和漏检两概率之和。误诊概率或称为错误率用 $P(e)$ 表示,其定义为

$$P(e) = \int_{-\infty}^{\infty} P(e, X) dX = \int_{-\infty}^{\infty} P(e/X) p(X) dX \quad (4.11)$$

式中 $\int_{-\infty}^{\infty} \cdot dX$ 表示在整个 n 维特征空间积分,可见错误率是指平均错误率。

决策规则(4.7)实际上是对每个 X 都使 $P(e/X)$ 取小者,这就使式(4.11)的积分也必须达到最小,即使平均错误率 $P(e)$ 达到最小,这就证明了最小错误率贝叶斯决策规划确实使错误最小。

对两类别问题,从式(4.7)的决策规则可知,如果 $P(w_2/X) > P(w_1/X)$,则决策应为 w_2 ;

显然在作出决策 w_2 时, x 的条件错误概率为 $P(w_1/X)$, 反之, 则为 $P(w_2/X)$, 可表示为

$$P(e) = \begin{cases} P(w_1/X), & \text{当 } P(w_2/X) > P(w_1/X) \\ P(w_2/X), & \text{当 } P(w_1/X) > P(w_2/X) \end{cases} \quad (4.12)$$

如图 4-9 所示, 令 t 为两类的分界面, 且特征向量 x 是一维时, t 将 x 轴分为两个区域 Ω_{w_1} 和 Ω_{w_2} , Ω_{w_1} 为 $(-\infty, t)$, Ω_{w_2} 为 (t, ∞) , 则有

$$P(e) = \int_{-\infty}^t P(w_2/X) p(X) dX + \int_t^{\infty} P(w_1/X) p(X) dX = \\ \int_{-\infty}^t p(w_2/X) P(X) dX + \int_t^{\infty} p(w_1/X) P(X) dX$$

也可写为

$$P(e) = P(X \subseteq \Omega_{w_1}, w_2) + P(X \subseteq \Omega_{w_2}, w_1) = \\ P(X \subseteq \Omega_{w_1}/w_2)P(w_2) + P(X \subseteq \Omega_{w_2}/w_1)P(w_1) = \\ P(w_2) \int_{\Omega_{w_1}} p(X/w_2) dX + P(w_1) \int_{\Omega_{w_2}} p(X/w_1) dX = \\ P(w_2)P_2(e) + P(w_1)P_1(e) \quad (4.13)$$

图 4-9 表示了式(4.13)的几何意义, 斜线面积为 $P(w_2)P_2(e)$, 纹线面积为 $P(w_1)P_1(e)$, 两者之和为 $P(e)$ 。

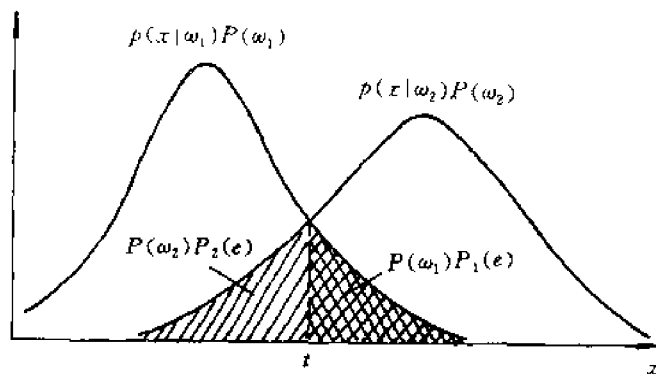


图 4-9 最小错误率

以上讨论很容易推广到 n 维特征空间的情况。在多类决策过程中, 把特征空间划分为 $\Omega_{w_1}, \Omega_{w_2}, \dots, \Omega_{w_m}$ 个区域, 其相应的最小错误率贝叶斯决策规则为

$$\text{如果 } P(w_i/X) = \max_{j=1, \dots, m} P(w_j/X) \quad \text{则 } X \subseteq w_i \quad (4.14)$$

其等价形式为

$$\text{如果 } P(X/w_i)P(w_i) = \max_{j=1, \dots, m} P(X/w_j)P(w_j) \quad \text{则 } X \subseteq w_i \quad (4.15)$$

多类别决策过程中, 要把特征空间分割成 $\Omega_{w_1}, \Omega_{w_2}, \dots, \Omega_{w_m}$ 个区域, 可能错分的情况很多, 平均错误概率 $P(e)$ 计算如下

$$P(e) = 1 - P(c) = 1 - \sum_{j=1}^m P(X \subseteq \Omega_{w_j}/w_j)P(w_j) = \\ 1 - \sum_{j=1}^m \int P(X/w_j)P(w_j) dX \quad (4.16)$$

式中, $P(c)$ 为决策分类平均正确分类概率。

4.3.3 基于最小风险的贝叶斯决策规则

1. 最小风险贝叶斯决策

在实际诊断问题中,经常需要考虑比错误率更广泛的概念,这就是风险问题,或决策损失。在误诊中,谎报与漏检两种不同的错误判断所造成损失的严重程度不同,将正常状态误判为异常状态,会造成停机检查和维修浪费,而将异常状态误判为正常状态,就可能造成机毁人亡。显然漏检的损失比谎报的损失要严重的多。最小风险贝叶斯决策正是考虑各种错误造成损失不同而提出的一种决策规则。

假设类型空间 Ω_w 有 m 个类别

$$\Omega_w = \{w_1, w_2, \dots, w_m\}$$

观测量 \mathbf{X} 是 n 维随机向量

$$\mathbf{X} = \{x_1, x_2, \dots, x_n\}^T$$

其中 x_1, x_2, \dots, x_n 为一维随机变量

决策空间 A 由 l 个决策, $A = \{a_1, a_2, \dots, a_l\}$

这里 l 与 m 不同是由于除了对 m 个类别有 m 种不同的决策外,还允许采取其他决策,如采取“拒绝”的决策,这时就有 $l = m + 1$, 损失函数(条件风险)为 $\lambda(a_i, w_j)$; $i = 1, 2, \dots, l$; $j = 1, 2, \dots, m$, λ 表示本应属 w_j 类的样本决策到 a_i 类时所带来的损失。

对于给定的 \mathbf{X} , 在采取决策 a_i 情况下的条件期望损失或称为条件风险 $R(a_i/\mathbf{X})$ 为

$$R(a_i/\mathbf{X}) = E[\lambda(a_i, w_j)] = \sum_{j=1}^m \lambda(a_i, w_j) P(w_j/\mathbf{X}) \quad i = 1, 2, \dots, l \quad (4.17)$$

条件风险 $R(a_i/\mathbf{X})$ 表示本应属于 w_j 类的样本判到 a_i 的各种判别的 $\lambda(a_i, w_j)$ 与对应后验概率的加权和,它反映了对某一 \mathbf{X} 的取值采取决策 a_i 所带来的风险。

在考虑误判带来的损失时,希望损失最小,如果在采取每一个决策时,都使其条件风险最小,则对所有 \mathbf{X} 作出决策时,其期望风险也必然最小,这样的决策就是最小风险贝叶斯决策,其规则为

$$\text{如果 } R(a_k/\mathbf{X}) = \min_{i=1, \dots, l} R(a_i/\mathbf{X}), \quad \text{则 } a = a_k \quad (4.18)$$

决策过程可按下列步骤进行:

(1) 在已知 $P(w_j)$, $P(\mathbf{X}/w_i)$, 及给出待识别 \mathbf{X} 的情况下,根据贝叶斯公式计算后验概率

$$P(w_j/\mathbf{X}) = \frac{P(\mathbf{X}/w_j)P(w_j)}{\sum_{i=1}^m P(\mathbf{X}/w_i)P(w_i)} \quad j = 1, 2, \dots, m$$

(2) 利用后验概率及根据所研究的具体问题,分析错误决策造成损失的严重程度而确定的损失函数 $\lambda(a_i, w_j)$,按式(4.17)计算采取 a_i 的条件风险 $R(a_i/\mathbf{X})$;

(3) 从 l 个 $R(a_i/\mathbf{X})$, ($i = 1, 2, \dots, l$), 按照决策规则(4.18)找出使条件风险最小的决策 a_k , 则 a_k 就是最小风险贝叶斯决策。

例 4-3 在例 4-2 条件的基础上,按最小风险贝叶斯决策规则进行分类。已知, $P(w_1) = 0.9$, $P(w_2) = 0.1$, $P(\mathbf{X}/w_1) = 0.2$, $P(\mathbf{X}/w_2) = 0.4$, 且 $\lambda(a_1, w_1) = 0$, $\lambda(a_1, w_2) = 6$, $\lambda(a_2,$

$w_1) = 1, \lambda(a_2, w_2) = 0。$

【解】 由例 4-2 的计算结果可知后验概率为 $P(w_1/X) = 0.818, P(w_2/X) = 0.182$ 按式(4.17) 计算的条件风险

$$R(a_1/X) = \sum_{j=1}^m \lambda(a_1, w_j) P(w_j/X) = \lambda(a_1, w_2) P(w_2/X) = 1.092$$

$$R(a_2/X) = \lambda(a_2, w_1) P(w_1/X) = 0.818$$

由于 $R(a_1/X) > R(a_2/X)$, 即决策为 w_2 的条件风险小于决策为 w_1 的条件风险, 因此我们采取决策 w_2 , 即判断待识别状态为异常状态 w_2 。

本例与例 4.2 相比, 分类结果正好相反, 这是因为这里影响决策结果的因素又多了一个“损失”, 且两类错误决策所造成的损失相差很悬殊, 因此“损失”起了主导作用。

2. 最小风险贝叶斯决策与最小错误率决策的关系

根据贝叶斯公式, 考虑 $P(X)$ 为公共项, 条件风险 $R(a_i/X)$ 可描述为

$$R(a_i/X) = \sum_{j=1}^m \lambda(a_i, w_j) P(w_j) P(X/w_j) \quad (4.19)$$

若损失函数 $\lambda(a_i, w_j) = \begin{cases} 0 & i = j \\ 1 & i \neq j \end{cases}$

即正确决策时, 无损失; 错误决策时, 损失一样都为 1, 这时

$$R(a_i/X) = \sum_{j=1}^m \lambda(a_i, w_j) P(w_j/X) = \sum_{\substack{j=1 \\ j \neq i}}^m P(w_j/X) \quad (4.20)$$

$\sum_{\substack{j=1 \\ j \neq i}}^m P(w_j/X)$ 表示对 X 采取决策 w_i 的条件错误概率, 所以, 在 $\lambda(a_i, w_j)$ 取值为 0 或 1 时, 使

$$R(a_i/X) = \min_{i=1, \dots, m} R(a_i/X)$$

的最小风险贝叶斯决策就等价于

$$\sum_{\substack{j=1 \\ j \neq i}}^m P(w_j/X) = \min_{i=1, \dots, m} \sum_{\substack{j=1 \\ j \neq i}}^m P(w_j/X)$$

的最小错误率贝叶斯决策。

显然, 此风险最小决策等于最小错误率决策, 最小错误率决策是最小风险决策的一个特例。

4.3.4 最小最大决策规则

在实际机械故障诊断中, 会遇到决策要处理的各种类别先验概率 $P(w_i)$ 未知, 或是变化的, 如果按某个固定的 $P(w_i)$ 条件下进行决策往往得不到最小错误或最小风险, 这里所要介绍的最小最大决策就是在考虑 $P(w_i)$ 变化的情况, 如何使最大可能的风险最小, 即在最差的情况下争取最好的结果。不失一般性, 讨论仍以两类问题为例。

设损失函数 λ_{11} 为 $X \in w_1$ 时决策为 $X \in w_1$ 的损失; λ_{21} 为 $X \in w_1$ 时决策为 $X \in w_2$ 的损失;

λ_{22} 为 $X \in w_2$ 时决策为 $X \in w_2$ 的损失; λ_{12} 为 $X \in w_2$ 时决策为 $X \in w_1$ 的损失。

通常作出错误决策总是比作出正确决策所带来得损失要大, 即

$$\lambda_{21} > \lambda_{11} \quad \text{及} \quad \lambda_{12} > \lambda_{22}$$

再假定类别 w_1 的决策域为 Ω_1 , 类别 w_2 的决策域为 Ω_2 , 而 $\Omega_1 + \Omega_2 = \Omega$, Ω 为整个特征空间, 即决策是把整个特征空间分割成不相交的二个区域和。

总风险可按式(4.17) 得出

$$\begin{aligned} R &= \int_{\Omega_1} R(a_1/X) p(X) dX + \int_{\Omega_2} R(a_2/X) p(X) dX = \\ &\int_{\Omega_1} [\lambda_{11} P(w_1) p(X/w_1) + \lambda_{12} P(w_2) p(X/w_2)] dX + \\ &\int_{\Omega_2} [\lambda_{21} P(w_1) p(X/w_1) + \lambda_{22} P(w_2) p(X/w_2)] dX \end{aligned} \quad (4.21)$$

考虑

$$\begin{aligned} P(w_1) + P(w_2) &= 1 \\ \int_{\Omega_2} p(X/w_1) dX &= 1 - \int_{\Omega_1} p(X/w_1) dX \end{aligned}$$

则式(4.21) 可写为

$$\begin{aligned} R &= \lambda_{12} + (\lambda_{12} - \lambda_{22}) \int_{\Omega_1} p(X/w_2) dX + \\ &P(w_1) \{ (\lambda_{11} - \lambda_{22}) (\lambda_{21} - \lambda_{11}) \int_{\Omega_2} p(X/w_1) dx - (\lambda_{12} - \lambda_{22}) \int_{\Omega_1} p(X/w_2) dX \} = \\ &a + bP(w_1) \end{aligned} \quad (4.22)$$

式中: $a = \lambda_{22} + (\lambda_{12} - \lambda_{22}) \int_{\Omega_1} p(X/w_2) dX$,

$$b = \{ (\lambda_{11} - \lambda_{22}) + (\lambda_{21} - \lambda_{11}) \int_{\Omega_2} p(X/w_1) dX - (\lambda_{12} - \lambda_{22}) \int_{\Omega_1} p(X/w_2) dX \}$$

式(4.22) 表明, 如果决策域 Ω_1 和 Ω_2 已确定, 则 a, b 为常数, 总风险 R 就是 $P(w_1)$ 的一个线性函数。 $R_{\max} = a + b$, 最大风险出现在 $P(w_1) = 1$ 或 $P(w_2) = 1$, $P(w_1)$ 在 $(0, 1)$ 内取值如何使最大可能的风险为最小呢?

$\partial R / \partial P(w_1) = 0$, 如果 $b = 0$, 则 $R = a$, R 与 $P(w_1)$ 无关, 即最大可能风险达到最小值。另一种合理的解释是既然 $P(w_1)$ 为变化或未知的, R_{\max} 只能在 $P(w_1) = 0$ 或 $P(w_1) = 1$ 达到。这时要求 R 取最小, 只有遍历所有 $P(w_1)$ 时找 R 的最大值中的最小者。由于 $R \sim P(w_1)$ 的线性关系, 只有 $P(w_1) = 0$ 或 $P(w_1) = 1$ 和 $P(w_1)$ 为 $0 \sim 1$ 之间的任意值时, R 均相等, 才能满足这一情况, 因此 R 只能为一水平线, 也就是 $P(w_1)$ 的系数为 0, 即

$$(\lambda_{11} - \lambda_{22}) + (\lambda_{21} - \lambda_{11}) \int_{\Omega_2} p(X/w_1) dX - (\lambda_{12} - \lambda_{22}) \int_{\Omega_1} p(X/w_2) dX = 0 \quad (4.23)$$

则总风险为

$$R = \lambda_{22} + (\lambda_{12} - \lambda_{22}) \int_{\Omega_1} p(X/w_2) dX = a \quad (4.24)$$

所以不管 $P(w_1)$ 作什么变化, 总风险 R 都不在变化, 而等于 a 。这时就使最大风险最小。因此, 最大最小决策的任务就是寻找使贝叶斯风险为最大时的决策域 Ω_1 和 Ω_2 , 它对应于(4.23) 积分方程的解。在求出使贝叶斯风险为最大时的决策域 Ω_1 和 Ω_2 以及对应于 $\partial R / \partial P(w_1) = 0$ 时的 $P_i^*(w_1)$, 最大最小决策规则就完全与 4.3.3 节中的决策规则相似。

4.4 距离函数分类法

为能将模式集划分成不同的类别,可以定义一种相似性的测度来度量特征空间同一类样本间的类似性和不同样本间的差异性。两个特征向量之间的距离是它们相似性的一种很好度量,也就是说,同类模式具有聚类性,不同类状态的模式有各自的聚类域和聚类中心,因此,可将待检模式与参考模式间的距离作为判别函数,判别待检状态的属性。

4.4.1 空间距离函数

1. 欧氏(Eudidean)距离

设状态存在 L 个类别 ω_j , 其参考模式向量 $Y^j = (y_1^j, y_2^j, \dots, y_n^j)$, $j = 1, 2, \dots, L$; 待检模式向量 $X = (x_1, x_2, \dots, x_n)^T$, 则欧氏距离定义为

$$d_E^2(X, Y^j) = \sum_{i=1}^n (x_i - y_i^j)^2 = (X - Y^j)^T (X - Y^j) \quad j = 1, 2, \dots, L \quad (4.25)$$

为消除模式特征分量的量纲对分类结果的影响,须对特征数据进行归一化处理,如

$$x_i = \frac{x_i - x_{\min}}{x_i - x_{\max}}$$

式中, x_{\max} 和 x_{\min} 分别为特征参数的最大值和最小值。

考虑到特征矢量中的诸分量对分类起的作用不同,可采用加权方法,构造加权欧氏距离

$$d_w^2(X, Y^j) = (X - Y^j)^T W (X - Y^j) \quad j = 1, 2, \dots, L \quad (4.26)$$

式中, W 为权系数矩阵。取不同的权矩阵 W , 也就引出不同的距离判别函数。

对于状态有 L 个类别 ω_j , $j = 1, 2, \dots, L$, 欧氏距离判据形式为

如果 $d_w^2(X, Y^j) = \min_{j=1, 2, \dots, L} d_w^2(X, Y^j)$ 则 $X \in \omega_j$

上式表明: 两点距离越近, 其相似性越大, 则可认为属于同一个群聚域, 或属于同一类别。

2. 马氏(Mahalanobis)距离

这是加权欧氏距离中用得较多的一种, 其形式为

$$d_M^2(X, Y^j) = (X - Y^j)^T C^{-1} (X - Y^j) \quad j = 1, 2, \dots, L \quad (4.27)$$

式中, C^{-1} 为 X 与 Y^j 的互协方差矩阵的逆矩阵, 即

$$C^{-1} = (X, Y^j)^T)^{-1}$$

或

$$C^{-1} = [(X - \mu_x)(Y^j - \mu_y)^T]^{-1}$$

式中 μ_x, μ_y 分别为 X 和 Y^j 的均值。

马氏距离的优点是排除了特征参数之间的相互影响。

3. 一般形式的明氏(Minkowski)度量

$$d_M(X, Y^j) = \left[\sum_{i=1}^n (x_i - y_i^j)^s \right]^{1/s} \quad (4.28)$$

式中, s 为正整数。显然, 当 $s = 2$, 即为欧氏距离, $s = 1$ 时, $d(X, Y^j) = \sum_{i=1}^n |x_i - y_i^j|$, 称为“街坊距离”。

上述空间距离函数都是在 R^n 中度量的, 要求待检模式阶数与所有参考模型的阶数均

相等。

4. 空间距离判别的应用

现以时间序列模型参数作为特征而得到残差偏移的距离函数为例。

设自回归 AR 模型的矩阵形式

$$\mathbf{X}\Phi = \mathbf{A}$$

式中, \mathbf{X} 为时序样本矩阵; Φ 为自回归系数矢量; \mathbf{A} 为残差矢量, 可得残差平方和

$$\mathbf{S} = \mathbf{A}^T \mathbf{A} = \Phi^T \mathbf{X}^T \mathbf{X} \Phi = \Phi^T \mathbf{R} \Phi$$

式中, $\mathbf{R} = \mathbf{X}^T \mathbf{X}$ 为样本序列的自相关函数。

设待检模型残差 $\mathbf{A}_T = \mathbf{X}_T \Phi_T$, 将其代入参考模型 $\mathbf{X}_R \Phi_R = \mathbf{A}_R$ 中, 得残差 $\mathbf{A}_{RT} = \mathbf{X}_T \Phi_R$, 定义 $\mathbf{A}_{RT} - \mathbf{A}_T$ 为残差偏移距离, 它表示待检模型与参考模型之间的接近程度, 于是有

$$\mathbf{A}_{RT} - \mathbf{A}_T = \mathbf{X}_T \Phi_R - \mathbf{X}_T \Phi_T = \mathbf{X}_T (\Phi_R - \Phi_T)$$

定义残差偏移距离

$$d_A^2 = (\mathbf{A}_{RT} - \mathbf{A}_T)^T (\mathbf{A}_{RT} - \mathbf{A}_T) = (\Phi_R - \Phi_T)^T \mathbf{X}_T^T \mathbf{X}_T (\Phi_R - \Phi_T) = (\Phi_R - \Phi_T)^T \mathbf{R}_T (\Phi_R - \Phi_T)$$

式中, $\mathbf{R}_T = \mathbf{X}_T^T \mathbf{X}_T$ 为待检序列的自相关函数。

从距离的意义来讲, 残差距离实质是以自相关函数矩阵为权矩阵的欧氏距离。

4.4.2 相似性指标

相似性指标是在作聚类分析时判断两个特征矢量点是否属于同一类的统计量, 待检状态应归入相似性指标最大的状态类别。

1. 角度相似性指标

定义两模式向量 \mathbf{X} 与 \mathbf{Y} 之间夹角的余弦为角度相似性指标。即

$$S(\mathbf{X}, \mathbf{Y}) = \frac{\sum_{i=1}^n x_i y_i}{\sqrt{\sum_{i=1}^n x_i^2 \sum_{i=1}^n y_i^2}} \quad (4.29)$$

或 $S(\mathbf{X}, \mathbf{Y}) = \mathbf{X}^T \mathbf{Y} / (\|\mathbf{X}\| \cdot \|\mathbf{Y}\|)$

式中 $\|\mathbf{X}\|$ 和 $\|\mathbf{Y}\|$ 分别为模式向量 \mathbf{X} 和 \mathbf{Y} 的模。 $S(\mathbf{X}, \mathbf{Y})$ 有一个重要的性质即坐标放缩、旋转对其值无影响, 当 $S(\mathbf{X}, \mathbf{Y}) = 1$, 即 \mathbf{X} 和 \mathbf{Y} 间的夹角为零时, 相似性达到最大。

2. 相关系数

$$S_R(\mathbf{X}, \mathbf{Y}) = \frac{\sum_{i=1}^n (x_i - \mu_x)(y_i - \mu_y)}{\sqrt{\sum_{i=1}^n (x_i - \mu_x)^2 \sum_{i=1}^n (y_i - \mu_y)^2}} \quad (4.30)$$

式中, μ_x, μ_y 分别为 \mathbf{X} 和 \mathbf{Y} 的均值, 相关系数越大, 表示相似性越强。

4.4.3 信息距离判别法

信息距离函数是由信息论中有关信息量的计算导出的, 它们是用来作为两个概率分布之间距离的度量, 这种距离又称“伪距离”。两个概率分布之间的“伪距离”越小, 它们之间的近似程度越高。

1. Kullback 信息量距离

Kullback 信息量距离是比较两个概率密度函数 p_1, p_2 之间的差异, 它是 p_1 与 p_2 间的互熵, 其表达式为

$$I(p_1, p_2) = \int p_1(x) \ln \frac{p_1(x)}{p_2(x)} dx$$

$$I(p_2, p_1) = \int p_2(x) \ln \frac{p_2(x)}{p_1(x)} dx \quad (4.31)$$

式(4.31)度量了两个概率密度函数的交叠程度,为它们之间距离的度量。当两类概率密度相同,即 $p_1(x)$ 等于 $p_2(x)$, Kullback 信息量(即互熵)等于零,当两类概率密度完全不交叠时,互熵取最大值。一般情况下,互熵为非负,因此 $I(p_1, p_2)$ 和 $I(p_2, p_1)$ 提供了两类之间的平均可分性信息。

在状态识别中,采用 Kullback 信息量度量两个状态之间的信息距离,可以证明,当参考状态序列 Y 与待检状态序列 X 服从正态分布时, Kullback 信息量为

$$I(p_R, p_T) = \ln \frac{\sigma_T}{\sigma_R} + \frac{1}{2\sigma_T^2} [\sigma_R^2 + (Y - X)^T R_T (Y - X)] - \frac{1}{2}$$

$$I(p_T, p_R) = \ln \frac{\sigma_R}{\sigma_T} + \frac{1}{2\sigma_R^2} [\sigma_T^2 + (Y - X)^T R_R (Y - X)] - \frac{1}{2} \quad (4.32)$$

式中, σ_R^2, σ_T^2 分别为参考状态序列 Y 与待检状态序列 X 的方差, R_T, R_R 为两状态序列的自相关函数。显然,当待检状态与参考状态相同,即 $X = Y, \sigma_R^2 = \sigma_T^2$, 则

$$I(p_R, p_T) = I(p_T, p_R) = 0$$

一般情况下,由于 $I(p_R, p_T) \neq I(p_T, p_R)$, 这表明将待检状态从参考状态区分出来的信息量不等于参考状态从待检状态中区分出来的信息量,可见,作为“距离函数”这是不理想的。

2. J 散度距离

由于 $I(p_1, p_2)$ 与 $I(p_2, p_1)$ 无对称性,在同一情况下,取值各不相同,定义 J 散度为两个状态总的平均信息,等于两类平均可分信息之和,即

$$J = I(p_1, p_2) + I(p_2, p_1) \quad (4.33)$$

当设备工况相同时, $X = Y, \sigma_R^2 = \sigma_T^2$, 有 $J = 0$, J 越小,两类模式的状态越接近。

3. Itakura 信息距离

时序模型的残差和残差的方差含有丰富的信息,当时序 $\{x_k\}$ 通过 AR 模型后,就意味着 $\{x_k\}$ 通过一个 AR 滤波器而凝聚成模型的残差 $\{a_k\}$ 作为输出,故可定义 Itakura 信息距离函数为

$$d_I(p_{RT}, p_T) = \frac{\sigma_{RT}^2}{\sigma_T^2} - 1 \quad (4.34)$$

式中 σ_T^2 和 σ_{RT}^2 分别是待检时序 $\{x\}_T$ 通过自身滤波器 AR_T 和参考滤波器 AR_R 后的残差, p_T 和 p_{RT} 分别是输出 $\{a_k\}_T$ 和 $\{a_k\}_{RT}$ 的概率密度函数,显然,当 $\sigma_T^2 = \sigma_{RT}^2$ 时,则 $d_I(p_{RT}, p_T) = 0$, 即待检状态属于该参考状态。

4. Kullback - Leiber 信息量

Kullback - Leiber 信息量(简称 K - L 信息量)是前述 Kullback 信息距离的特例,其定义为

$$d_{KL}(p_{RT}, p_T) = \ln \frac{\sigma_R^2}{\sigma_T^2} + \frac{\sigma_{RT}^2}{\sigma_R^2} - 1 \quad (4.35)$$

$d_{KL}(p_{RT}, p_R)$ 是比较残差序列 $\{a_k\}_{RT}$ 与 $\{a_k\}_R$ 的概率密度函数 p_{RT} 和 p_R 之间的差异。

假定有 c 个参考时间序列 $\{x^i\}, i = 1, 2, \dots, c$, 来源于 c 个不同的机器状态,由此 c 个参考序

列建立的 c 个时序模型为

$$x_k^1 = \phi_1^1 x_{k-1}^1 + \cdots + \phi_p^1 x_{k-p}^1 + a_k^1 \quad \text{残差的方差为 } \sigma_1^2$$

$$x_k^2 = \phi_1^2 x_{k-1}^2 + \cdots + \phi_q^2 x_{k-q}^2 + a_k^2 \quad \text{残差的方差为 } \sigma_2^2$$

.....

$$x_k^c = \phi_1^c x_{k-1}^c + \cdots + \phi_r^c x_{k-r}^c + a_k^c \quad \text{残差的方差为 } \sigma_c^2$$

现有一待检状态 $\{x\}_T$ 的时序模型为 $x_k^T = \phi_1^T x_{k-1}^T + \cdots + \phi_s^T x_{k-s}^T + a_k^T$ 残差的方差为 σ_T^2

可根据 K-L 信息量 $d_{KL}(p_{mT}, p_m) = \ln \frac{\sigma_m^2}{\sigma_T^2} + \frac{\sigma_{mT}^2}{\sigma_m^2} - 1$ 之值为最小, 将 $\{x\}_T$ 归类到 $\{r^m\}$ 类型中,

$$\text{式中} \quad \sigma_{mT}^2 = \sum_{k=u+1}^N [x_k^T - (\phi_1^m x_{k-1}^T + \cdots + \phi_u^m x_{k-u}^T)]^2$$

即将待识别序列 $\{x\}_T$ 代入参考模型 $\{x^m\}$ 中后所得到的残差。当 $\sigma_T^2 = \sigma_m^2 = \sigma_{mT}^2$, 即待识别序列本身所建立的模型的残差与将之代入参考模型后所得到的残差, 以及参考模型本身的残差三者相等时, K-L 信息量为零。

例 4-4 根据发动机振动信号建立的三个参考模型分别对应机器的正常运行 A, 连杆撞击 B 和阀撞击 C 三种运行状态。求出待检发动机运行状态 T 的 K-L 信息量如下表。

表 4-1 用 K-L 信息量对信号分类

	A	B	C	T
A	0.000	0.176	0.212	0.182
B	0.087	0.000	0.020	0.026
C	0.084	0.019	0.000	0.005
T	0.079	0.025	0.005	0.000

由表可见, 待检发动机运行状态是属于阀撞击 C 这一运行状态。

5. Bhattacharyya 信息距离

对于一个由 N 个样本组成的平稳时间序列, 其与参考序列之间的 Bhattacharyya 距离可如下求出

$$d_b(p_R, p_T) = -\ln \int \sqrt{p_R(x) \cdot p_T(x)} dx \quad (4.36)$$

式中 $p_R(x)$, $p_T(x)$ 分别为参考时序 $\{x\}_R$ 和待检时序 $\{x\}_T$ 的概率密度函数。当两者均为正态分布时, 可推得

$$d_b(p_R, p_T) = \frac{1}{2} \ln \left(\frac{\det C}{\sqrt{\det C_R \det C_T}} \right) \quad (4.37)$$

式中 C_R, C_T 为参考时序与待检时序的自协方差矩阵, $C = (C_R + C_T)/2$ 。

Bhattacharyya 信息距离的特点是直接根据时间序列的统计特性进行判别, 而不需对时序拟合 AR 模型或 ARMA 模型。

4.4.4 应用距离函数分类时应注意的问题

应用距离函数对机械设备运行状态进行分类, 首先要找到能表达各种工况的特征矢量作为训练样本, 求得在各种状态下模式点的聚类中心, 将与之对应的特征矢量作为参考模式; 再计算待检样本与聚类中心的距离, 按最近邻准则确定其状态属性。对于两类问题, 这种方法十分有效, 多类问题, 由于决策函数复杂, 实时性差, 在生产中应用存在困难。

距离函数使用还存在其他困难,如要已知状态类别的模式集,这往往难以得到。另外,运行状态处于异常时,其工况样本的聚类性很差,所求得的聚类中心不一定能代表该类状态的属性。

4.5 故障诊断的模糊模式识别方法

前几节讨论的故障模式识别诊断方法,是建立在诊断对象的故障及故障原因是明确、清晰和肯定的,即模式是明确、清晰和肯定的基础之上的。而在很多情况下,设备故障的信息环境基本上是一个模糊环境,模糊性的存在一方面是由于设备运行状态的划分设有确切的含义,在量上没有明确的界限,造成状态分类亦此亦彼的性态,这些性态的类属是不清晰的。另一面由于人们对机械设备症状的观察本身也是不明确的,主观的成分较高,导致对同一台设备的评价得到不确切的结论。

解决上述问题有效的方法是应用模糊数学,将模糊现象与因素间关系用数学表达方式描述并进行运算,分析设备故障诊断中各个环节中所遇到的各种模糊信息,对它们进行科学的、定量的处理与解释,这就是模糊诊断技术。

4.5.1 隶属概念与隶属函数

在经典集合论中,论域 U 中的任意一个元素 u 与集合 A 的关系只有 $u \in A$ 或 $u \notin A$ 两种情况,二者必居且仅居其一,用函数表示为

$$x_A(u) = \begin{cases} 1 & \text{当 } u \in A \\ 0 & \text{当 } u \notin A \end{cases} \quad (4.38)$$

式中, x_A 称为集合 A 的特征函数。 x_A 在 u 处的值 $x_A(u)$ 称为 u 对 A 的隶属度, $x_A(u) = 1$ 时,表示 u 属于且绝对隶属于 A ; $x_A(u) = 0$ 时,表示 u 不属于且绝对不属于 A 。

模糊集合论则把 u 对 A 的隶属度从 0,1 二值逻辑推广到可取 $[0,1]$ 闭区间中任意值的连续逻辑,此时的特征函数 x_A 称为隶属函数,用 $\mu(x)$ 表示,它表征着所论及的特征 X 以多大程度隶属于状态空间 $\Omega = (w_1, w_2, \dots, w_m)$ 中哪一个子集合 w_j ,用 $\mu_w(x)$ 表示,且满足 $0 \leq \mu_w(x) \leq 1$ 。

在故障诊断中,可以将故障发生时出现的状态空间 Ω 作为论域,而把引起故障征兆 $X = (x_i) (i = 1, 2, \dots, 2^n)$ 的故障原因(即某中运行状态)看成论域的模糊子集 $w_j (j = 1, 2, \dots, m)$ 。所以诊断问题就是确定 X 的某个元素 x , 以多大程度隶属于哪个模糊子集的问题。隶属函数在模糊数学中占有重要地位,它是把模糊性进行数值化描述,使事物的不确定性在形式上用数学方法进行计算。因此,进行模糊状态识别的首要问题是研究如何确定隶属函数。

确定隶属函数常用的方法有以下几种:

- (1) 通过模糊统计来确定;
- (2) 采用二元对比排序法;
- (3) 借用常见的模糊分布来确定;
- (4) 利用动态信号处理的结果,经过适当转换得到隶属函数;
- (5) 通过神经网络模型来学习和获取隶属函数;
- (6) 通过主观识别、个人经验、人为评分等获取隶属函数。

下面仅对方法(3)和(4)进行讨论。

1. 常用的模糊分布

表4-2列出了几种常用的隶属函数。其中(偏小型)戒上型分布适用于 x 很小的隶属函数。由于隶属函数分布在第一象限,所以论域均取正值;(偏大型)戒下型分布适用于 x 较大时的隶属函数,由于隶属函数分布在第一象限,所以论域亦取正值。

表 4-2 常用隶属函数图表

类型	隶属函数名称	隶属函数图形	隶属函数表达式
偏小型 (戒上型)	降半 Γ 形分布		$\mu(x) = \begin{cases} 1 & x \leq a \\ e^{-k(x-a)} & x > a, k > 0 \end{cases}$
	降半正态形分布		$\mu(x) = \begin{cases} 1 & x \leq a \\ e^{-k(x-a)^2} & x > a, k > 0 \end{cases}$
	降半哥西形分布		$\mu(x) = \begin{cases} 1 & x \leq a \\ \frac{1}{1 + a(x-a)^\beta} & x > a, k > 0, \beta > 0 \end{cases}$
	降半凹(凸)分布		$\mu(x) = \begin{cases} 1 - ax^k & 0 \leq x \leq \frac{1}{\sqrt{a}} \\ 0 & \frac{1}{\sqrt{a}} < x \end{cases}$
	降半梯形分布		$\mu(x) = \begin{cases} 1 & 0 \leq x \leq a_1 \\ \frac{a_2 - x}{a_2 - a_1} & a_1 \leq x \leq a_2 \\ 0 & a_2 < x \end{cases}$
	降半岭形分布		$\mu(x) = \begin{cases} 1 & 0 \leq x \leq a_1 \\ \frac{1}{2} - \frac{1}{2} \sin \frac{\pi}{a_2 - a_1} (x - \frac{a_1 + a_2}{2}) & a_1 < x < a_2 \\ 0 & a_2 < x \end{cases}$
	偏大型 (戒下型)	升半 Γ 形分布	
升半正态形分布			$\mu(x) = \begin{cases} 0 & x \leq a \\ 1 - e^{-k(x-a)^2} & x > a, k > 0 \end{cases}$

续 表

类型	隶属函数名称	隶属函数图形	隶属函数表达式
偏大型 (戒下型)	升半哥西形分布		$\mu(x) = \begin{cases} 1 & x \leq a \\ \frac{1}{1 + a(x-a)^{-\beta}} & x > a, k > 0, \beta > 0 \end{cases}$
	升半凹(凸)分布		$\mu(x) = \begin{cases} 0 & x \leq a \leq a \\ a(x-a)^k & a < x < a + \frac{1}{\sqrt{a}} \\ 1 & a + \frac{1}{\sqrt{a}} \leq x \end{cases}$
	升半梯形分布		$\mu(x) = \begin{cases} 0 & 0 \leq x \leq a_1 \\ \frac{x-a_1}{a_2-a_1} & a_1 \leq x \leq a_2 \\ 1 & a_2 < x \end{cases}$
	升半岭形分布		$\mu(x) = \begin{cases} 0 & 0 \leq x \leq a_1 \\ \frac{1}{2} - \frac{1}{2} \sin \frac{\pi}{a_2-a_1} (x - \frac{a_1+a_2}{2}) & a_1 < x \leq a_2 \\ 1, & a_2 < x \end{cases}$
中间型 (对称型)	矩形分布		$\mu(x) = \begin{cases} 0 & 0 \leq x \leq a-b \\ 1 & a-b < x \leq a+b \\ 0 & a+b < x \end{cases}$
	尖Γ形分布		$\mu(x) = \begin{cases} e^{k(x-a)} & x \leq a \\ e^{-k(x-a)} & x > a, k > 0 \end{cases}$
	正态形分布		$\mu(x) = e^{-k(x-a)^2} \quad k > 0$
	哥西形分布		$\mu(x) = \frac{1}{1 + a(x-a)^\beta} \quad a > 0, \beta \text{ 为正偶数}$

续 表

类型	隶属函数名称	隶属函数图形	隶属函数表达式
中间型 (对称型)	梯形分布		$\mu(x) = \begin{cases} 0 & 0 \leq x \leq a - a_2 \\ \frac{a_2 + x - a}{a_2 - a_1} & a - a_2 < x < a - a_1 \\ 1 & a - a_1 \leq x \leq a + a_1 \\ \frac{a_2 - x + a}{a_2 - a_1} & a + a_1 < x < a + a_2 \\ 0 & a + a_2 \leq x \end{cases}$
	岭形分布		$\mu(x) = \begin{cases} 0 & x \leq -a_2 \\ \frac{1}{2} + \frac{1}{2} \sin \frac{\pi}{a_2 - a_1} (x - \frac{a_1 + a_2}{2}) & -a_2 < x \leq a_2 \\ 1 & -a_1 < x \leq a_1 \\ \frac{1}{2} - \frac{1}{2} \sin \frac{\pi}{a_2 - a_1} (x - \frac{a_1 + a_2}{2}) & a_1 < x \leq a_2 \\ 0 & a_2 < x \end{cases}$

2. 动态信号处理方法

对机械系统运行过程中的动态信号进行采集、分析处理来识别机械系统所处的状态或具有的状态,存在着随机性和模糊性,随机性是由因果关系不确定造成的,概率统计和随机过程方法对此进行了长期、深入的研究,动态信号处理就采用这一理论和方法;模糊性是指事物在质和量上没有明确的界限,即衡量尺度不清楚,模糊数学为解决此类问题提供了手段,将动态信号处理与模糊数学方法结合起来,是解决状态监测和故障诊断中随机性和模糊性这种不确定问题的适宜途径。

(1) 自相关系数法:当研究线性动态信号 $x(t)$ 中周期分量(或噪声)的多少时,就进入了模糊领域。设 $X = [x_i(t), i = 1, 2, \dots, n]$ 为表征论域中某一状态 w_j 的特征向量,则 w_j 表示 $x_i(t)$ 中周期分量大小的模糊子集,隶属函数(表示 $x_i(t)$ 属于 w_j 的程度)为

$$\mu_w(x_i) = \frac{E[\max R_x(\tau)] - (\bar{x}_i)^2}{\sigma_{x_i}^2} \quad (4.39)$$

式中, $R_x(\tau)$, \bar{x}_i , $\sigma_{x_i}^2$ 分别是 $x_i(t)$ 的自相关函数、均值和方差; $E[\max R_x(\tau)]$ 是当 $0 < \tau_0 \leq \tau$ 时自相关函数的极大值的平均值; τ_0 是一足够大的正数。这样 $\mu_w(x_i) \in [0, 1]$ 。

由于周期函数的自相关函数仍是同周期的周期函数,噪声的自相关函数因噪声的不相关性,当时滞 τ 足够大时($\tau > \tau_0$) 趋于零,所以当 $0 < \tau_0 < \tau$, 式(4.39)表示的隶属函数就反映了信号 $x(t)$ 中周期分量多少的程度。

(2) 凝聚函数法:当研究系统的输出信号 $y(t)$ 在多大程度上受到输入信号 $x(t)$ (或 $x_i(t)$, $i = 1, 2, \dots, n$) 的影响,或信号 $y(t)$ 与信号 $x(t)$ 在频域中的相关程度时,就使问题进入了模糊空间。设论域中的模糊子集 w 表示在某频带里信号 $x(t)$ 与信号 $y(t)$ 的相干程度,则属于 w 的隶属函数可用该频带内的凝聚函数的最大值 $\max \gamma_{xy}^2(f)$ 表示

$$\mu_w(x) = \max \frac{|S_{xy}(f)|^2}{S_x(f)S_y(f)} = \max \gamma_{xy}^2(f) \quad (4.40)$$

式中, $S_{xy}(f)$ 是 $x(t)$ 与 $y(t)$ 的互谱密度; $S_x(f), S_y(f)$ 分别是 $x(t), y(t)$ 的自谱密度; $0 \leq \gamma_{xy}^2(f) \leq 1, \mu_w(x) \in [0, 1]$ 。

若信号处理所得到的参数其范围不在 $[0, 1]$ 区间, 则可经过适当转换, 用 $[0, 1]$ 之间的值表示隶属函数。

(3) 峭度指标法: 时域信号 $x(t)$ 的峭度指标 R_k 峭度定义为

$$R_k = \int_{-\infty}^{\infty} x^4 p(x) dx / \left[\int_{-\infty}^{\infty} x^2 p(x) dx \right]^2 \quad (4.41)$$

式中, x 代表信号 $x(t), p(t)$ 是 $x(t)$ 的概率密度函数。

R_k 值可灵敏地反映信号 $x(t)$ 的变化情况。根据所研究对象的实际情况, 分析其状态的 R_k 值变化规律, 可构造一定的模糊分布来刻画其隶属函数。表 4-3 是齿轮疲劳试验过程中不同阶段时的振动信号的 R_k 值, 表 4-4 是铣刀在加工过程中随刀具磨损量增大时声发射信号的 R_k 值。

表 4-3 齿轮疲劳试验 R_k 值

阶段	峭度指标 R_k
初期	3.63
中期	8.18
后期	26.6

表 4-4 铣刀磨损声发射 R_k 值

铣刀磨损量	峭度指标 R_k
0.05	0.585
0.10	0.856
0.20	1.013
0.30	1.731

虽然表 4-3 与表 4-4 中的 R_k 值不同, 但都随状态发展变化而使 R_k 值增大, 表明各状态属于“故障严重”(即齿轮疲劳损坏及铣刀磨损严重)这一模糊子集 w 的程度随 R_k 值增大, 还可见随状态发展 R_k 值增加的幅度呈非线性增大, 故可用(偏大型)戒下型的升半 Γ 分布来描述其隶属函数, 其表达式为

$$\mu_w(R_k) = \begin{cases} 0 & \text{当 } R_k \leq a \\ 1 - e^{-k(R_k - a)} & \text{当 } R_k > a, \text{ 其中 } k > 0 \end{cases} \quad (4.42)$$

例 4-5 对表 4-3 齿轮疲劳试验, 若令 $R_k = 3.0$ 时, $\mu_w(R_k) = 0$, 可求得 $a = 3.0$, 取 $R_k = 26.60$ 时, 令齿轮疲劳损坏故障严重性隶属度为 0.95, 可得 $k = 0.138$, 则隶属函数可写为

$$\mu_w(R_k) = \begin{cases} 0 & \text{当 } R_k \leq 3.0 \\ 1 - e^{-0.138(R_k - 3.0)} & \text{当 } R_k > 3.0 \end{cases}$$

同理, 对表 4-4 铣刀磨损故障来说, 可得 $a = 0.5$, 并设 $R_k = 1.731$ 时, 磨损故障严重性隶属度为 0.90, 可得, 则隶属函数为

$$\mu_w(R_k) = \begin{cases} 0 & \text{当 } R_k \leq 0.5 \\ 1 - e^{-1.87(R_k - 0.5)} & \text{当 } R_k > 0.5 \end{cases}$$

可见, 不同的工况或不同的故障状态, 它们的故障严重性程度在信号处理方法中给出了不同的

峭度指标值 R_i , 相互之间不便于比较。若用隶属函数表示, 则故障严重这一模糊概念可以用同一尺度去衡量, 即用 $[0, 1]$ 之间的隶属度值进行衡量, 不同工况状态之间也便于相互比较。

(4) 谱距离指标(J 散度)法: 在对动态信号分析和识别时, 会遇到这样的问题: n 个机械状态信号 $x_1(t), x_2(t), \dots, x_n(t)$ 其对应的功率谱分别为 $S_1(f), S_2(f), \dots, S_n(f)$, 现有一待识别信号 $y(t)$, 那么 $y(t)$ 应归属于上述 n 个状态中的哪一个呢? 这个问题可用 $y(t)$ 功率谱 $S_y(f)$ 与 n 个谱之间的散度 $J(y, x_i) (i = 1, 2, \dots, n)$ 来分析。当 $J(y, x_i)$ 为最小时, 则判定所处的状态与 $x_i(t)$ 所处的状态属于同一类, J 散度公式为

$$J(y, x_i) = \frac{1}{2m} \sum_{k=1}^m \left[\frac{S_y(k)}{S_i(k)} + \frac{S_i(k)}{S_y(k)} \right] - 1 \quad i = 1, 2, \dots, n \quad (4.43)$$

显然, $J(y, x_i) = J(x_i, y) \geq 0$, 当 $S_y(f) = S_i(f)$, 即对同一个谱, $J(y, x_i) = 0$ 。

分类问题是一个模糊问题, 在一般模式识别问题中所有的距离函数均是值域在 $[0, +\infty]$ 上的实函数, 不同状态和类别的距离值其差别很大, 精确研究类别之间的相似程度时, 用它往往难以表达清楚, 使问题模糊化, 若论域中的模糊集合 w 定义为 $y(t)$ 与 $\{x_i(t), i = 1, 2, \dots, n\}$ 的相似程度, 由 J 散度指标可知, 当其取最小值时, 表明二状态的相似程度最大, 因此可用偏小型函数来表示这种相似程度, 即相似测度。相似测度是模糊测度的一种, 可表示模糊集合的隶属度, 功率谱(或幅值谱) J 散度属于 w 的隶属函数, 可用降半哥西分布表示为

$$\mu_w(J) = \begin{cases} 1 & J \leq 0 \\ \frac{1}{1+aJ} & J > 0, a > 0, \text{且为两常数} \end{cases} \quad (4.44)$$

由于 $J \in [0, +\infty]$, 故上式可写成

$$\mu_w(J) = \frac{1}{1+aJ} \quad J \geq 0, a > 0 \quad (4.45)$$

这样, 使距离值转化成 $[0, 1]$ 区间中的值, 有利于相互之间进行比较。

(5) 状态参量法: 不同时期机械所处状态的动态信号的某些参量 u , 如谱峰值、均方差、频率的幅值、偏心等, 其值增大表示状态向故障严重方向发展, 那么参量 u 属于严重异常状态模糊子集 w 的隶属函数可定义为(升半梯形分布)

$$\mu_w(u) = \begin{cases} 0 & u \leq u_{\min} \\ \frac{u - u_{\min}}{u_{\max} - u_{\min}} & u_{\min} < u \leq u_{\max} \\ 1 & u > u_{\max} \end{cases} \quad (4.46)$$

式中: u_{\min} 和 u_{\max} 是根据工程要求设定的下限和上限值。

有了隶属函数的描述, 借助模糊诊断原理, 可以对机械设备的运行动态信号进行识别、分类、评判、推理与决策, 从而能够全面正确地对机械设备作状态监测和故障诊断。

4.5.2 基于识别算法的模糊识别

1. 模糊识别算法与归属原则

模糊识别的识别算法可叙述为: 假设有 n 个模糊集 w_1, w_2, \dots, w_n (代表 n 种类型), 当一个识别算法作用于诊断对象 u 时, 就产生隶属度 $\mu_{w_i}(u) (i = 1, 2, \dots, n)$ 表示诊断对象 u 属于集合 w_i 的程度, 如果一个识别算法的清晰描述已经给出, 这个算法称为明确的; 如果算法没有清晰描述, 这种算法称为不明确的。人们通常是通过不明确的算法直接对诊断对象 u 进行识别, 而

模式识别则是将一个不明确的算法转换为明确的算法,从对诊断对象本身进行识别转化为对它的模式进行识别。

识别算法的工作原则上分三步进行:

(1) 特征抽取:从诊断对象 x 中提取与识别有关的诸特征,并测出 x 在各特征上的具体数据,将 x 转化为模式 $x = \{x_1, x_2, \dots, x_m\}$,称为诊断对象的特征模式。这步是基础,特征抽取是否得当,将直接影响识别的结果。

(2) 建立隶属函数:建立一个明确算法以产生隶属函数 $\mu_{w_i} (i = 1, 2, \dots, n)$, x 属于模糊集合 w_i 的隶属度 $\mu_{w_i}(x)$ 依赖于 x_1, x_2, \dots, x_m 。隶属函数的确定还没有一般、普遍的原则,应用中的许多公式还带有主观性和经验性的成分。

(3) 识别判决。按某种归属原则对 x 进行判决,指出它应归属哪一类型。

识别判决一般分直接法与间接法两种,直接法又称个体识别方法,它是按最大隶属原则来进行决策的。间接法又称群体识别方法,它是按最小距离原则或按择近原则来归类的。

① 最大隶属原则。设 Ω 为全体被识别对象构成的论域, w_1, w_2, \dots, w_n 是 Ω 的 n 个模糊子集,现对一个确定的对象 $x_0 \in \Omega$ 进行识别。由于模式 w_1, w_2, \dots, w_n 是模糊的,而确定的对象 x_0 是清晰的,因此要用最大隶属原则进行归类。

若 w_1, w_2, \dots, w_n 中每一个 $w_i (i = 1, 2, \dots, n)$ 的隶属函数已确定,则对任一个元素 $x_0 \in \Omega$ 均可按下述隶属原则确定其归属,假若 x_0 满足

$$\mu_{w_i}(x_0) = \max[\mu_{w_1}(x_0), \mu_{w_2}(x_0), \dots, \mu_{w_n}(x_0)] \quad (4.17)$$

可认为 x_0 隶属于 w_i ,即待识别故障 x_0 应属于模糊子集 w_i 。

② 择近原则。设论域 Ω 上有 n 个模糊子集: w_1, w_2, \dots, w_n ,而被识别的对象是模糊的,它也是论域 Ω 上的模糊子集 B ,这时就要考虑 B 与每个 $w_i (i = 1, 2, \dots, n)$ 的贴近程度 (B, w_i) 。 B 和哪一个 w_i 最贴近就认为它属于哪一类,这就要用到择近原则。

若 w_i 和 B 满足

$$(B, w_i) = \max\{(B, w_1), (B, w_2), \dots, (B, w_n)\} \quad (4.18)$$

则认为 B 和 w_i 最贴近。

2. 模糊性度量

在应用最大隶属原则和择近原则进行模糊识别时,为衡量两个模糊集合之间的贴近程度,经常用到模糊数学中的模糊性变量概念建立模式识别关系。

(1) 模糊度:设论域 Ω 上任一个模糊子集 w ,为量度其模糊性大小,定义

$$D: w \rightarrow [0, 1] \quad (4.19)$$

为 w 的模糊度 $D(w)$,它应满足以下三种情况:

① 当且仅当 $\mu_{w_i}(x_i)$ 只取 0 或 1 时, $D(w) = 0$, $x_i \in \Omega$, $\mu_{w_i}(x_i)$ 是 x_i 对 w 的隶属度。模糊度取值为普通集合中的 0 时,说明普通集合并不模糊。

② 当 $\mu_{w_i}(x_i) = 0.5$ 时, $D(w)$ 应取最大值,即 $D(w) = 1$,也就是说,当隶属度为 0.5 时最为模糊。

③ 对任意 $x \in \Omega$,设 Ω 上有两个模糊子集 A 和 B ,若 $\mu_A(x) \geq \mu_B(x) \geq 0.5$,或 $\mu_A(x) \leq \mu_B(x) \leq 0.5$,则 $D(B) \geq D(A)$,就是说,越靠近 0.5 就越模糊,反之,离 0.5 越远就越清晰。最模糊时模糊度为 1;最清晰时模糊度为 0,在这种情况下,应属于普通集合。

(2) 模糊熵:设一系统有 n 个状态 w_1, w_2, \dots, w_n ,其各自的概率为 $P(w_1), P(w_2), \dots$,

$P(w_n)$, 则系统熵定义为

$$H[P(w_1), P(w_2), \dots, P(w_n)] = - \sum_{i=1}^n P(w_i) \ln[P(w_i)] \quad (4.50)$$

模糊事件 A 的熵定义为

$$H(A) = - \sum_{i=1}^n \mu_w(A) P(w_i/A) \ln[P(w_i/A)] \quad (4.51)$$

式中, $P(w_i/A)$ ($i = 1, 2, \dots, n$) 为模糊事件 A 属于状态模糊子集 w_i 的概率。由上式可知模糊熵就是一个模糊事件 A 的模糊度。

(3) 距离: 常用距离来度量模糊集。由特征向量 $X = \{x_1, x_2, \dots, x_n\}$ 所描述的模糊集 A 与 B 间的距离 $d(A, B)$ 应满足: 非负性、对称性及传递性, 几种常用的距离见表 4-5。

表 4-5 几种常用的距离

距离类型	数学表达式
海明距离	$d_1(A, B) = \frac{1}{n} \sum_{i=1}^n \mu_A(x_i) - \mu_B(x_i) $
欧几里德距离	$d_2(A, B) = \frac{1}{\sqrt{n}} \sqrt{\sum_{i=1}^n (\mu_A(x_i) - \mu_B(x_i))^2}$
闵可夫斯基距离	$d_3(A, B) = \left[\frac{1}{n} \sum_{i=1}^n \mu_A(x_i) - \mu_B(x_i) ^p \right]^{1/p}, p > 0$
极小极大化距离	$d_4(A, B) = \frac{\sum_{i=1}^n \mu_A(x_i) - \mu_B(x_i) }{\sum_{i=1}^n (\mu_A(x_i) + \mu_B(x_i))}$

例 4-6 设 $X = \{x_1, x_2, x_3, x_4, x_5, x_6\}$, 各特征对模糊集 A, B 的隶属度表述为

$$A = \frac{0.6}{x_1} + \frac{0.8}{x_2} + \frac{1}{x_3} + \frac{0.8}{x_4} + \frac{0.6}{x_5} + \frac{0.4}{x_6}, B = \frac{0.4}{x_1} + \frac{0.6}{x_2} + \frac{0.3}{x_3} + \frac{1}{x_4} + \frac{0.8}{x_5} + \frac{0.6}{x_6}$$

求各种距离的大小。

【解】 $d_1(A, B) = \frac{1}{6} \sum_{i=1}^6 |\mu_A(x_i) - \mu_B(x_i)| \approx 0.28$

$$d_2(A, B) = \frac{1}{\sqrt{6}} \sqrt{\sum_{i=1}^6 (\mu_A(x_i) - \mu_B(x_i))^2} \approx 0.34$$

取 $p = 1/3$, 则有 $d_3(A, B) = \left[\frac{1}{6} \sum_{i=1}^6 |\mu_A(x_i) - \mu_B(x_i)|^{1/3} \right]^3 \approx 0.16$

$$d_4(A, B) = \frac{\sum_{i=1}^6 |\mu_A(x_i) - \mu_B(x_i)|}{\sum_{i=1}^6 (\mu_A(x_i) + \mu_B(x_i))} \approx 0.22$$

(4) 贴适度: 用距离来度量模糊度, 在离散情况下需要把很多项相加求和, 当论域元素很多时, 求和较困难。为此, 引入贴适度的概念。

设 A, B 是由特征向量 $X = \{x_1, x_2, \dots, x_n\}$ 所描述的两个模糊集, 则几种在实际中常用的

贴近度见表 4-6。

表 4-6 几种常用的贴近度

距离贴近度类型	数学表达式
海明贴近度	$\sigma_1(A, B) = 1 - d_1(A, B) = 1 - \frac{1}{n} \sum_{i=1}^n \mu_A(x_i) - \mu_B(x_i) $
欧几里德贴近度	$\sigma_2(A, B) = 1 - d_2(A, B) = 1 - \frac{1}{\sqrt{n}} \sqrt{\sum_{i=1}^n (\mu_A(x_i) - \mu_B(x_i))^2}$
闵可夫斯基贴近度	$\sigma_s(A, B) = 1 - [d_s(A, B)]^s = 1 - \frac{1}{n} \sum_{i=1}^n \mu_A(x_i) - \mu_B(x_i) ^s$
极小极大化贴近度	$\sigma_4(A, B) = 1 - d_4(A, B) = 1 - \frac{\sum_{i=1}^n \mu_A(x_i) - \mu_B(x_i) }{\sum_{i=1}^n (\mu_A(x_i) + \mu_B(x_i))}$

4.5.3 故障诊断的模糊综合评判方法

综合评判是多目标决策问题的一个数学模型。故障诊断的模糊综合评判就是应用模糊变换原理和最大隶属度原则,根据各故障原因与故障征兆之间的不同程度的因果关系,在综合考虑所有征兆的基础上,来诊断设备发生故障的可能原因。这里,首先是对各个征兆进行单独评判,然后再对所有征兆进行综合评判。

1. 征兆集

在设备故障诊断中,对设备的每一种故障,根据设备的各种资料和维修经验可统计出该种故障发生时所可能表现出的各种征兆。设共有 n 种不同的征兆,则各种不同的征兆构成的集合就是征兆集,它可表示为 $X = \{x_1, x_2, \dots, x_n\}$ 。

例如,对汽轮机转子“轴系不对中”故障而言,其征兆可能有某轴承处轴颈垂直方向的“通频振幅 x_1 ”,各阶谐波的“一阶振幅 x_2 ”、“二阶振幅 x_3 ”等。

x_i 对 X 的关系是普通集合关系,因此,征兆集是一个普通集合。模糊综合评判就是在综合考虑所有征兆的基础上,评判设备发生异常状态的可能原因。

2. 故障(状态原因)集

在设备故障诊断中,根据设备的各种资料和实际的经验可统计出各种故障状态。设共有 m 种故障状态,则所有可能故障构成的集合可表示为 $\Omega = \{\omega_1, \omega_2, \dots, \omega_m\}$ 。这些故障状态具有不同程度的模糊性。

例如,汽轮机转子主要故障有:“初始不平衡 ω_1 ”,“转子部件脱落 ω_2 ”,“转子暂时热弯曲 ω_3 ”,“汽封碰磨 ω_4 ”,“轴向碰撞 ω_5 ”等。

3. 单故障模糊评判

首先对故障集 Ω 中的一个故障 $\omega_i (i = 1, 2, \dots, m)$ 作单故障模糊评判,确定被评判对象对

征兆集元素 $x_j (j = 1, 2, \dots, n)$ 的隶属度(可能性程度) r_{ij} , 这样就得出第 i 个故障 ω_i 的单故障模糊集

$$r_i = (r_{i1}, r_{i2}, \dots, r_{in})$$

它是征兆集 X 上的模糊子集, 这样 m 个故障的评价集就构造出一个总的评价模糊矩阵 R

$$R = \begin{bmatrix} r_{11} & r_{12} & \dots & r_{1n} \\ r_{21} & r_{22} & \dots & r_{2n} \\ \dots & \dots & \dots & \dots \\ r_{m1} & r_{m2} & \dots & r_{mn} \end{bmatrix} \quad (4.52)$$

R 即是故障论域 U 到征兆论域 X 的一个模糊关系, $\mu_R(\omega_i, x_j) = r_{ij}$ 表示 ω_i 和 x_j 之间隶属关系的程度, 即评判对象按 ω_i 评判时取 x_j 的亲疏程度。

4. 权重集(重要程度系数)

单故障评判是比较容易办到的, 多故障的综合评判就比较困难了。因为, 一方面, 对于被评判的诊断对象, 从不同的故障着眼可以得到绝然不同的结论; 另一方面, 在诸故障 $\omega_i (i = 1, 2, \dots, m)$ 之间, 有些故障在总评价中的影响程度可能大些, 而另一些故障在总评价中的影响程度可能要小些, 说明各个故障原因在总评价中所起的作用的重要程度是一个模糊择优问题。用 $a_i (i = 1, 2, \dots, m)$ 表示各原因 ω_i 在总评价中重要程度的权数, 因此各权数组成的集合为故障论域 U 上的模糊子集 A , 记作

$$A = \frac{a_1}{\omega_1} + \frac{a_2}{\omega_2} + \dots + \frac{a_m}{\omega_m} \quad (4.53)$$

或者

$$A = (a_1, a_2, \dots, a_m)$$

式中 $a_i (0 \leq a_i \leq 1)$ 为 ω_i 对 A 的隶属度。它是单故障 ω_i 在总评价中的影响程度大小的度量, 在一定程度上也代表单故障 ω_i 评定等级的能力。注意: a_i 可能是一种调整系数或者限制系数, 也可能是普通权系数, A 称为 Ω 的因素重要程度模糊子集, a_i 称为因素 ω_i 的重要程度系数。可依实际情况主观确定或用隶属函数法确定。

5. 模糊综合评判

当模糊向量 A 和模糊关系矩阵 R 为已知时, 作模糊变换来进行模糊综合评判。

$$B = A \cdot R = (a_1, a_2, \dots, a_m) \cdot \begin{bmatrix} r_{11} & r_{12} & \dots & r_{1n} \\ r_{21} & r_{22} & \dots & r_{2n} \\ \dots & \dots & \dots & \dots \\ r_{m1} & r_{m2} & \dots & r_{mn} \end{bmatrix} = (b_1, b_2, \dots, b_n) \quad (4.54)$$

B 中的各元素 b_j 是在广义模糊合成运算下得出的运算结果, 权重集 A 和模糊关系矩阵 R 的合成, 一般用综合评判模型 $M(\dot{*}, \ddagger)$ 表示, 其中 $\dot{*}$ 为广义模糊“与”运算, \ddagger 为广义模糊“或”运算。

B 称为征兆集 X 上的模糊子集, $b_j (j = 1, 2, \dots, n)$ 为征兆 x_j 对综合评判所得模糊子集 B 的隶属度。如果要选择一个决策, 则可按照最大隶属度原则选择最大的 b_j 所对应的征兆 x_j 作为综合评判的结果。

在广义模糊合成运算下综合评判模型, 即式(4.54)的意义在于 $r_{ij} (i = 1, 2, \dots, m, j = 1,$

2, …, n) 为单独考虑故障原因 ω_i 的影响时诊断对象对征兆 x_j 的隶属度; 而通过广义模糊“与”运算 ($a \overset{+}{*} r_{ij}$) 所得的结果 (记为 r_{ij}^*), 就是在全面综合考虑各种故障的影响时诊断对象对征兆 x_j 的隶属度, 也就是在考虑故障 ω_i 在总评判中的影响程度 a_i 时对隶属度 r_{ij} 所进行的调整或限制。最后通过广义模糊“或”运算对各个调整 (或限制) 后的隶属度 r_{ij}^* 进行综合处理, 即可得出合理的综合评价结果。

式(4.54)所示的模糊变换 R (单故障评判矩阵), 可以看作是从故障论域 Ω 到征兆论域 X 的一个模糊变换器, 也就是说每输入一个模糊向量 A 就可输出一个相应的综合评判结果 B 。

6. 模糊综合评判的 5 种具体模型

就理论上而言, 上述的广义模糊合成运算有无穷多种, 但在故障诊断的实际应用中, 经常采用的具体模型有以下 5 种。

模型 1: $M(\wedge, \vee)$, 即用 \wedge 代替 $\overset{+}{*}$, \vee 代替 $\overset{+}{\vee}$, 有

$$b_j = \bigvee_{i=1}^m (a_i \wedge r_{ij}) \quad j = 1, 2, \dots, n \quad (4.55)$$

式中 \wedge, \vee 分别为取小(min) 和取大(max) 运算, 即

$$b_j = \max[\min(a_1, r_{1j}), \min(a_2, r_{2j}), \dots, \min(a_m, r_{mj})]$$

在此模型中, 单故障 ω_i 的评价对征兆 x_j 的隶属度 r_{ij} 被调整为

$$r_{ij}^* = a_i \wedge r_{ij} = \min(a_i, r_{ij}) \quad j = 1, 2, \dots, n$$

这清楚地表明, a_i 是在考虑多故障时调整后的隶属度 r_{ij} 的上限, 换句话说, 我们在考虑多故障时, 诊断对象对各征兆 x_j ($j = 1, 2, \dots, n$) 的隶属度都不能大于 a_i 。因此, a_i 是在考虑多故障时 r_{ij} 的调整系数。

用 \vee 代替 $\overset{+}{*}$ 的意义是: 在决定 b_j 时, 对每个征兆 x_j 而言, 只考虑调整后的隶属度 r_{ij}^* 最大的起主要作用的那个故障, 而忽略了其他故障的影响。可见, 模型 $M(\wedge, \vee)$ 是一种“主故障决定型”的综合评判。它的优点是计算方便, 缺点是运算太粗糙, 诊断中往往丢掉有价值的信息, 以至所得诊断结果常常不太令人满意。

模型 2: $M(\cdot, \vee)$, 即用 \cdot 代替 $\overset{+}{*}$, \vee 代替 $\overset{+}{\vee}$, 于是

$$b_j = \bigvee_{i=1}^m (a_i \cdot r_{ij}) \quad j = 1, 2, \dots, n \quad (4.56)$$

式中“ \cdot ”为普通实数乘法, “ \vee ”为取大(max) 运算。

此模型与模型 $M(\wedge, \vee)$ 的意义很相近, 其区别仅在于 $M(\cdot, \vee)$ 以 $r_{ij}^* = a_i r_{ij}$ 代替了 $M(\wedge, \vee)$ 的 $r_{ij}^* = a_i \wedge r_{ij}$, 也就是说, 用对 r_{ij} 乘一小于 1 的系数来代替给 r_{ij}^* 规定一个上限。此模型中, 因为也是用 \vee 代替 $\overset{+}{*}$, 所以模型 $M(\cdot, \vee)$ 也是一种“主故障突出型”的综合评判。

模型 3: $M(\wedge, \oplus)$, 即用 \wedge 代替 $\overset{+}{*}$, \oplus (有界算子) 代替 $\overset{+}{\vee}$, 于是

$$b_j = \oplus \sum_{i=1}^m a_i \wedge r_{ij} \quad j = 1, 2, \dots, n \quad (4.57)$$

这里“ \wedge ”为取小(min) 运算, $a \oplus \beta = \min(1, a + \beta)$, $\sum_{i=1}^m$ 为对 m 个数在 \oplus 运算下求和, 即

$$b_j = \min\left[1, \sum_{i=1}^m \min(a_i, r_{ij})\right]$$

由上式可以看出,与模型 $M(\wedge, \vee)$ 中一样,在模型 $M(\wedge, \oplus)$ 中也是对 r_{ij} 的规定上限 a_i , 给以 r_{ij} 的调整, 即有 $r_{ij}^* = a_i \wedge r_{ij}$, 其区别在于, 该模型是对各 r_{ij}^* 作有上界相加以求 b_j 。因此, a_i 也是在考虑多故障时 r_{ij} 的调整系数。形式上这个模型是一种对每一种征兆 x_j 都同时对应各种故障的综合评判。

该模型在取小运算 $(a_i \wedge r_{ij})$ 时, 仍会丢失大量有价值的信息, 以至所得诊断结果常常不太令人满意。当 a_i 和 r_{ij} 取值较大时, 相应的 b_j 值均可能等于上限 1; 当 a_i 取值较小时, 相应的 b_j 值均可能等于各 a_i 之和, 这样就会得不到有意义的诊断结果。

模型 4: $M(\cdot, \oplus)$, 即用 \cdot 代替 $*$, \oplus (有界算子) 代替 $*$, 有

$$b_j = \oplus \sum_{i=1}^m a_i \cdot r_{ij} \quad j = 1, 2, \dots, n \quad (4.58)$$

式中“ \cdot ”为普通实数乘法, $\alpha \oplus \beta = \min(1, \alpha + \beta)$, $\sum_{i=1}^m$ 为对 m 个数在 \oplus 运算下求和, 即

$$b_j = \min[1, \sum_{i=1}^m a_i r_{ij}]$$

此模型是在模型 $M(\cdot, \vee)$ 的基础上改进而成的。模型 $M(\cdot, \oplus)$ 在决定 b_j 时, 是用对调整后的 $r_{ij}^* = a_i r_{ij}$ 取上界和来代替模型 $M(\cdot, \vee)$ 中对 $r_{ij}^* = a_i r_{ij}$ 取最大。该模型有下列重要特点:

(1) 在确定诊断对象对征兆 x_j 的隶属度 b_j 时, 综合考虑了所有故障 $\omega_i (i = 1, 2, \dots, m)$ 的影响, 而不是像模型 $M(\cdot, \vee)$ 那样只考虑对 b_j 影响程度最大的那个故障。

(2) 由于同时考虑到所有故障的影响, 所以各 a_i 的大小具有刻画各故障 ω_i 的重要性程度的权数的意义, 因此, a_i 应满足要求: $\sum_{i=1}^m a_i = 1$ 。

所以模型 $M(\cdot, \oplus)$ 是一种“加权平均型”的综合评判, 在此模型中, 模糊向量 $A = (a_1, a_2, \dots, a_m)$ 具有权向量的意义。

应该指出, 由于 $\sum_{i=1}^m a_i r_{ij} \leq 1$, 运算 \oplus 实际上已蜕化为普通实数加法“+”。因此模型 $M(\cdot, \oplus)$ 可改变成为模型 $M(\cdot, +)$, 即用 \cdot 代替 $*$, $+$ 代替 $*$, 这就是普通矩阵乘法, 所以有

$$b_j = \sum_{i=1}^m a_i \cdot r_{ij} \quad j = 1, 2, \dots, n \quad (4.59)$$

式中 \cdot 和 $+$ 分别为普通实数的乘法和加法, 权系数 a_i 的和满足条件: $\sum_{i=1}^m a_i = 1$ 。

模型 5: $M(\text{乘幂}, \wedge)$, 即用普通乘幂代替 $*$, \wedge 代替 $*$ 。

$$b_j = \bigwedge_{i=1}^m r_{ij}^{\cdot} \quad j = 1, 2, \dots, n \quad (4.60)$$

在此模型中, 考虑多故障时, 对故障论域 Ω 到征兆论域 X 的模糊关系矩阵 R 中元素 r_{ij} 的调整为 $r_{ij}^* = r_{ij}^{\cdot}$, 该模型最大的特点是在各调整值 r_{ij}^* 中取其最小者作为 b_j , 这说明在这一模型中 r_{ij} 和 b_j 不再是各自相应的隶属度, 而是某种评判指标。这里规定评判指标的最小者为最佳者。

以上 5 种模型, 均称为综合评判初始模型。

4.5.4 模糊综合评判故障诊断应用实例

1. 柴油机系统故障的模糊诊断

这里以柴油机系统故障分析与诊断为例介绍模糊综合评判故障诊断的应用。柴油机是一个极为复杂的系统,由曲柄连杆机构、配气机构、燃油系统、电器系统以及润滑、冷却、增压等部分组成。在柴油机工作过程中,各子系统及各零部件均会发生故障,而且故障的各种征兆往往不易发现或难以完全发现,故障现象和故障原因之间的对应关系也是模糊的。我们应用模糊综合评判对其进行分析 and 诊断。

(1) 确定征兆集 X : 以柴油机系统可检测信息组成柴油机系统的故障征兆集为

$$X = \{\text{振动, 压力, 温度}\} = \{x_1, x_2, x_3\}$$

(2) 建立故障集 Ω : 人为地将柴油机系统按各子系统划分, 所以柴油机系统的故障集为

$$\Omega = \{\text{曲柄连杆机构故障, 配气机构故障, 燃油系统故障, 其他系统故障}\} = \{\omega_1, \omega_2, \omega_3, \omega_4\}$$

另外对柴油机系统定义一个评价集为

$$G = \{\text{正常, 不太正常, 不正常}\} = \{g_1, g_2, g_3\}$$

(3) 建立模糊关系矩阵 R 及权重集 A : 设第 i 个征兆的单故障原因模糊评判集为

$$r_i = \frac{r_{i1}}{g_1} + \frac{r_{i2}}{g_2} + \frac{r_{i3}}{g_3}$$

或简单地表示为

$$r_i = \{r_{i1}, r_{i2}, r_{i3}\}$$

式中 r_{ij} 为第 i ($i = 1, 2, 3$) 个征兆对评价集中第 j ($j = 1, 2, 3$) 个元素的隶属度。将各单故障原因模糊评判集的隶属度为行组成模糊综合评判矩阵

$$R = \begin{Bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{Bmatrix}$$

R 为征兆集 X 到评价集的模糊关系矩阵。

又设对各个征兆 x_i ($i = 1, 2, 3$) 赋予的权重为

$$A_i = \frac{a_{i1}}{x_1} + \frac{a_{i2}}{x_2} + \frac{a_{i3}}{x_3}$$

式中 a_{ij} ($i = 1, 2, 3$) 为第 j 个征兆对故障集 Ω 中第 i ($i = 1, 2, 3$) 个元素的权数。总的权重矩阵为

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \\ a_{41} & a_{42} & a_{43} \end{bmatrix}$$

(4) 诊断结果描述: 对柴油机系统故障集 Ω 的综合评判结果为

$$\Omega = A \cdot R = \begin{bmatrix} \omega_{11} & \omega_{12} & \omega_{13} \\ \omega_{21} & \omega_{22} & \omega_{23} \\ \omega_{31} & \omega_{32} & \omega_{33} \\ \omega_{41} & \omega_{42} & \omega_{43} \end{bmatrix}$$

式中 $\omega_k = \bigvee_{j=1}^3 a_{ij} \bigvee_{j=1}^3 r_{jk} \quad i = 1, 2, 3, 4; k = 1, 2, 3$

将 Ω 中最后一列元素和第一列元素对应之差,按大小排列,如 $\omega_{13} - \omega_{11} \geq \omega_{23} - \omega_{21} \geq \omega_{33} - \omega_{31} \geq \omega_{43} - \omega_{41}$,则故障按隶属度大小排列为: $\omega_1, \omega_2, \omega_3, \omega_4$,所以故障最有可能发生在 ω_1 。

现对一台六缸柴油机进行故障诊断,试验中人为地减少某一缸的供油量,采集振动、压力、温度等信息,其中振动信息的评判是通过自相关值来进行的。

将采集到的征兆信息代入各单故障原因对于评价集的隶属函数,得到模糊综合评判矩阵为

$$R' = \begin{bmatrix} 0.16 & 0.28 & 0.42 \\ 0.97 & 0.58 & 0.24 \\ 0.91 & 0.46 & 0.02 \end{bmatrix}$$

对其作归一化处理,即用其各行元素之和,除遍原来的每个元素值,得归一化后的单故障原因模糊综合评判矩阵为

$$R = \begin{bmatrix} 0.186 & 0.326 & 0.488 \\ 0.542 & 0.324 & 0.134 \\ 0.655 & 0.331 & 0.014 \end{bmatrix}$$

据经验,各故障对征兆的全重分配为

$$A = \begin{bmatrix} 0.45 & 0.23 & 0.32 \\ 0.65 & 0.00 & 0.35 \\ 0.80 & 0.00 & 0.20 \\ 0.10 & 0.45 & 0.45 \end{bmatrix}$$

则模糊综合评判为

$$\Omega = AR = \begin{bmatrix} 0.45 & 0.23 & 0.32 \\ 0.65 & 0.00 & 0.35 \\ 0.80 & 0.00 & 0.20 \\ 0.10 & 0.45 & 0.45 \end{bmatrix} \begin{bmatrix} 0.186 & 0.326 & 0.488 \\ 0.542 & 0.324 & 0.134 \\ 0.655 & 0.331 & 0.014 \end{bmatrix} = \begin{bmatrix} 0.32 & 0.326 & 0.45 \\ 0.35 & 0.331 & 0.448 \\ 0.20 & 0.326 & 0.488 \\ 0.45 & 0.331 & 0.134 \end{bmatrix}$$

按第三列减第一列后,得到

$$\omega_{31} - \omega_{33} = 0.288 > \omega_{13} - \omega_{11} = 0.13 > \omega_{23} - \omega_{21} = 0.098 > \omega_{43} - \omega_{41} = 0.316$$

所以故障发生在 ω_3 ,即燃油系统故障,诊断结果正确。

2.5 种模型的应用效果

通过对工程机械液压系统故障分析与诊断,介绍上述 5 种模型的应用效果。

(1) 确定征兆集 X :任何液压机械设备,其液压系统的故障征兆大致可归结为:压力不足 x_1 ,流量不足 x_2 ,温度高 x_3 ,操纵无反应 x_4 ,系统发生振动 x_5 ,噪声增大 x_6 和漏油加剧 x_7 等。因此,液压系统的故障征兆集为 $X = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7\}$ 。

(2) 建立故障集 Ω :找出各种可能的故障为:液压泵故障 ω_1 ,油马达故障 ω_2 ,液压缸压障

ω_3 , 压力阀故障 ω_4 , 流量阀故障 ω_5 , 方向阀故障 ω_6 , 管系故障 ω_7 , 液压油故障 ω_8 , 滤油器故障 ω_9 和其他故障 ω_{10} 。因此, 故障集为 $\Omega = \{ \omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6, \omega_7, \omega_8, \omega_9, \omega_{10} \}$ 。

(3) 诊断结果: 故障论域 Ω 到征兆论域 X 的评判模糊矩阵为

$$R = \begin{bmatrix} 0.139 & 0.192 & 0.010 & 0.104 & 0.139 & 0.143 & 0.083 \\ 0.069 & 0.038 & 0.010 & 0.069 & 0.069 & 0.107 & 0.167 \\ 0.069 & 0.038 & 0.010 & 0.034 & 0.034 & 0.071 & 0.208 \\ 0.345 & 0.077 & 0.260 & 0.276 & 0.069 & 0.036 & 0.083 \\ 0.034 & 0.115 & 0.240 & 0.269 & 0.034 & 0.071 & 0.125 \\ 0.069 & 0.077 & 0.080 & 0.010 & 0.139 & 0.143 & 0.250 \\ 0.207 & 0.054 & 0.320 & 0.034 & 0.276 & 0.214 & 0.010 \\ 0.034 & 0.038 & 0.010 & 0.010 & 0.139 & 0.107 & 0.010 \\ 0.010 & 0.038 & 0.040 & 0.172 & 0.010 & 0.010 & 0.010 \end{bmatrix}$$

故障重要程度模糊子集(权重集)为

$$A = (0.188, 0.070, 0.064, 0.155, 0.128, 0.096, 0.107, 0.176, 0.037, 0.032)$$

用 5 种模型计算故障诊断的评判结果如表 4-7 所示。

表 4-7 5 种模糊综合评判模型的故障诊断评判结果

模型	$A \cdot R = B = (b_1, b_2, b_3, b_4, b_5, b_6, b_7)$						
	b_1	b_2	b_3	b_4	b_5	b_6	b_7
模型 1M(\wedge, \vee)	0.176	0.128	0.176	0.155	0.176	0.176	0.107
模型 2M(\cdot, \vee)	0.053	0.030	0.056	0.043	0.049	0.038	0.027
模型 3M(\wedge, \oplus)	0.763	0.694	0.707	0.672	0.658	0.725	0.616
模型 4M(\cdot, \oplus)	0.132	0.101	0.137	0.132	0.120	0.103	0.099
模型 5M(乘幂, \wedge)	0.649	0.598	0.581	0.551	0.661	0.597	0.445

4.6 神经网络及其在故障诊断中的应用

人工神经网络是一门高度综合的交叉学科, 它有研究和发展涉及神经生理科学、数理科学、信息科学和计算机科学等众多学科领域。

人工神经网络是模仿生物脑结构和功能的一种信息处理系统, 虽然目前的模仿还处于低水平, 但已显示出一些与生物脑类似的特点: ① 大规模并行结构与信息的分布式存储和并行处理, 克服了传统的智能诊断系统出现的无穷递归、组合爆炸及匹配冲突问题, 它特别适用于快速处理大量的并行信息。② 具有良好的自适应性, 系统在知识表示和组织、诊断求解策略与实施等方面可根据生存环境自适应、自组织达到自我完善。③ 具有较强的学习、记忆、联想、识别功能, 系统可根据环境提供的大量信息, 自动进行联想、书记及聚类等方面的自组织学习, 也可在导师的指导下学习特定的任务, 从而达到自我完善。④ 具有很强的容错性, 当外界输入到神经网络中的信息存在某些局部错误时, 不会影响到整个系统的输出性能。神经网络已经在信号处理、模式识别、目标跟踪、机器人控制、专家系统、系统辨识等众多领域显示出其极大的应

用价值,作为一种新的模式识别技术或一种知识处理方法,神经网络在机械故障诊断中显示了其极大的应用潜力。

神经网络在故障诊断领域的应用主要集中在三个方面:一是从模式识别角度应用神经网络作为分类器进行故障诊断;二是从预测角度应用神经网络作为动态预测模型进行状态预测;三是从知识角度建立基于神经网络的诊断专家系统。本节着重讨论第一方面的问题。

4.6.1 神经网络基础

1. 生物神经元

神经元是生物体中参与信息处理的各种神经细胞的总称,其形状和大小是多种多样的,就结构而言,存在有共性,图 4-10 所示为神经元的基本结构,它主要由三部分组成:① 细胞体,它是神经元的主体,含有细胞核、细胞质和细胞膜。② 树突,它是神经元的输入端,接受其他神经元的传递信号,其形状类似树状分枝。③ 轴突,它是由细胞体伸出的一条最长的突起,用来传出细胞体产生的输出电信号,其端部的众多神经末梢为信号的输出端子,每一条神经末梢可以与其他神经元形成功能性接触,该接触部分称为突触。

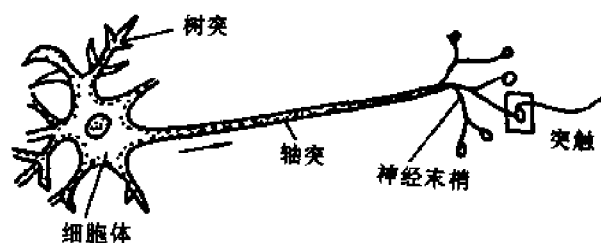


图 4-10 神经元解剖结构

神经元具有兴奋和抑制两种工作状态。当传入的神经冲动,使细胞膜电位升高到阈值时,细胞进入兴奋状态,产生神经冲动,由轴突输出。相反,若传入的神经冲动,使细胞膜电位下降到低于阈值时,细胞进入抑制状态,没有神经冲动输出。

2. 人工神经元

人工神经元模型是生物神经元的抽象与模拟,它是神经网络的基本处理单元,其模型结构如图 4-11 所示,它是一个多输入单输出的非线性阈值元件,假定 x_1, x_2, \dots, x_n 表示某一神经元的 n 个输入, w_{ij} 表示第 j 个神经元与第 i 个神经元的突触连接强度,其值称为权值, A_i 表示第 i 个神经元的输入总和,相应于生物神经细胞的膜电位,称为激活函数, y_i 表示第 i 个神经元的输出, h_i 表示神经元的阈值,则人工神经元的输出可描述为

$$y_i = f(A_i)$$

$$A_i = \sum w_{ij} x_j - h_i \quad (4.61)$$

式中 $f(A)$ 是表示神经元输入-输出关系的函数,称为作用函数或传递函数。常用的传递函数可归结为三种形式:阈值型、S型和伪线性型,如图 4-12 所示,这样,就有三类基本的神经元模型。

(1) 阈值型神经元:阈值型神经元是二值型神经元,其输出状态值为 1 或 0,分别代表神经元的兴奋和抑制状态。用数学表达式表示为

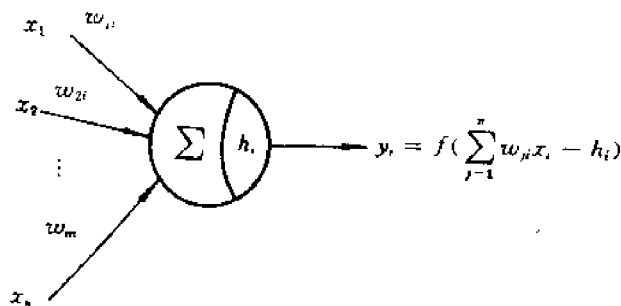


图 4-11 形式神经元模型

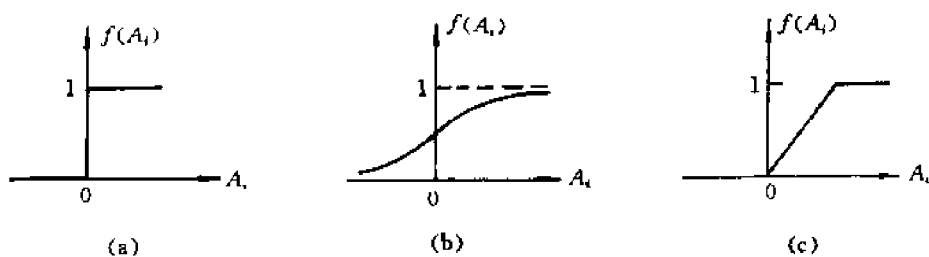


图 4-12 常用的作用函数形式

(a) 阈值型; (b) S型; (c) 伪线性型

$$y_i = f(A_i) = \begin{cases} 1 & A_i \geq 0 \\ 0 & A_i < 0 \end{cases} \quad (4.62)$$

对阈值型神经元, 权值 w_{ji} 可在 $(-1, +1)$ 区间连续取值, 取负表示抑制两神经元间的连接强度, 正值表示加强。

当神经元输出为 -1 或 1 时, $f(A_i)$ 为 sgn 函数。

$$y_i = f(A_i) = \text{sgn}(A_i) = \begin{cases} 1 & A_i \geq 0 \\ -1 & A_i < 0 \end{cases} \quad (4.63)$$

(2) S型神经元。神经元输出值是在 $(0, 1)$ 或 $(-1, 1)$ 内连续取值, 输入-输出特性多采用反指数、对数或双曲正切等 S 状曲线表示, 如

$$y_i = f(A_i) = \frac{1}{1 + e^{\beta A_i}} \quad \beta > 0$$

或

$$y_i = f(A_i) = \frac{1}{2} \left[1 + \text{th} \frac{A_i}{A_0} \right] \quad (4.64)$$

图 4-12(b) 所示的是上式的 $f(A_i)$ 在 β 取不同值时的曲线, 显然当 β 趋于无穷时, S 状曲线趋近于阶跃函数。通常情况下 β 取值为 1。

有时在网络中还采用下列简单的非线性函数。

$$y_i = f(A_i) = \frac{A_i}{1 + |A_i|} \quad (4.65)$$

S 型传递函数反映了神经元的非线性输出特性。

(3) 伪线性型神经元: 神经元的输入-输出特性满足一定的工镜线性关系, 其输出可表示为

$$y_i = f(A_i) = \begin{cases} 0 & A_i \leq 0 \\ cA_i & 0 < A_i \leq A_c \\ 1 & A_i > A_c \end{cases} \quad (4.66)$$

式中, c, A_c 均表示常量。

3. 神经网络的拓扑结构

神经网络是由大量的神经元相互连接而成的网络, 根据连接方式不同, 神经网络主要可分成两大类: 没有反馈的前向网络和反馈网络。

(1) 前向网络: 前向网络由输入层、中间层(或叫隐层)和输出层组成, 如图 4-13(a) 所示, 中间层(隐层)可有若干层, 每一层的神经元只接受前一层神经元的输出。网络的输入模式经过各层的顺次处理后得到输出层输出。

(2) 反馈网络: 反馈网络与前向网络不同, 网络中任意两个神经元都可能连接, 图 4-13(b) 所示, 因此输入信号要在神经元之间反复往返传递, 从某一初态开始, 经过若干次变化, 渐渐趋于某一稳定状态或进入周期振荡或混沌等状态。

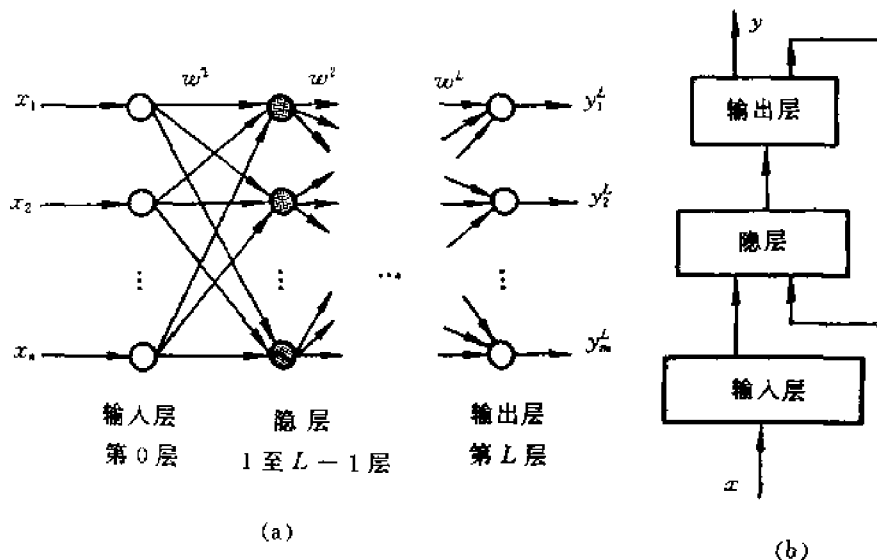


图 4-13 神经网络的拓扑结构

(a) 前向网络; (b) 反馈网络

4. 神经网络的学习规则

人工神经网络最有价值的特性就是它的自适应功能, 这种自适应功能是通过所谓的学习或训练实现的。任何一个神经网络模型要实现某种功能的操作, 必须先对它进行训练, 即使得它学会所要完成的任务, 并把这些学得的知识记忆(存储)在网络的权重中。

人工神经网络的学习规则可分为如下几种:

(1) 相关规则: 仅依赖于连接间的激活水平改变权重, 如贺普(Hebb)规则及其各种修正形式等。

(2) 纠错规则: 依赖于输出节点的外部节点的反馈改变网络权重, 如感知器学习规则、 δ 规则以及广义 δ 规则等。

(3) 竞争学习规则:类似于聚类分析算法,学习表示为自适应于输入空间的事件分布,如矢量化(LOQ)算法、SOM算法、以及ART训练算法都利用了竞争学习规则。

(4) 随机学习规则:利用随机过程、概率统计和能量函数的关系来调节连接权,如模拟退火算法;此外基于生物进化规则的基因遗传(GA)算法在某种程度上也可视为一类随机学习算法。

尽管神经网络的学习规则多种多样,但它们一般都可归结为以下两类:

(1) 有指导学习:不但需要学习用的输入事例(也称训练样本,通常为一矢量),同时还要求与之对应的表示所需期望输出的目标矢量。进行学习时,首先计算一个输入矢量的网络输出,然后同相应的目标输出比较,比较结果的误差用来按规定的算法改变加权。如上述纠错规则以及随机学习规则就是典型的有指导学习。

(2) 无指导学习:不要求有目标矢量,网络通过自身的“经历”来学会某种功能,在学习时,关键不在于网络实际输出是否与外部的期望输出相一致,而在于调整权重以反映学习样本的分布,因此整个训练过程实质是抽取训练样本集的统计特性,如上述纠错规则和竞争学习规则。

值得指出的是在工程实践中,有指导学习和无指导学习并不是相互冲突的,目前已经出现了一些融合有指导学习和无指导学习的训练算法。如在应用有指导学习训练一个网络后,再利用一些后期的无指导学习来使得网络自适应于环境的变化。

4.6.2 神经网络基本模型

目前已提出的神经网络模型大约有上百种。在信号分析和模式识别领域,应用最多的是以下4种基本模型和它们的改进型,即多层感知器, Hopfield神经网络, 自组织神经网络和概率神经网络。

1. 多层感知器与BP算法

多层感知器是一种典型的有指导学习的前馈神经网络,图4-14示出了一种典型的结构。隐层和输入层中任一神经元的输入等于与之相邻的低一层中各神经元输出的加权和。以简单的两类分类问题为例,其基本原理是:将输入的样本通过一个简单权向量集 W 的超平面,分离成两种模式类型如“A”类或“B”类。基本过程是:在提供网络的正确输出样本下,按照误差校正学习规则去寻找合适的权集 $W = (w_0, w_1, w_2, \dots, w_n)$,以使如果输入样本属于“A”类,感知器输出为“A”;如果输入样本属于“B”类,则输出为“B”类,如图4-15所示。

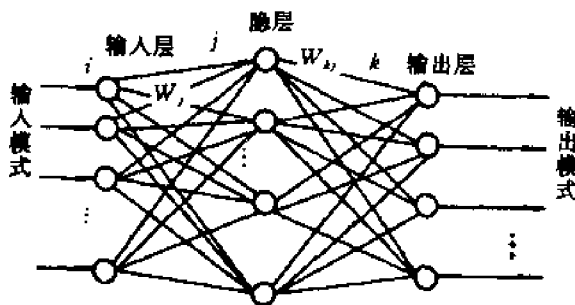


图4-14 多层感知器

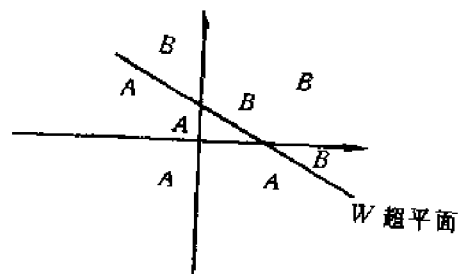


图4-15 感知器分类

感知器学习方程可以表示为

$$W^{new} = W^{old} + (Y - Y')X \quad (4.67)$$

式中, W 为权向量矩阵; Y 为正确的输出向量矩阵(训练期间提供); Y' 为感知器输出向量矩阵; X 为输入向量矩阵。

由式(4.67)可以看出, 如果感知器在输出中产生一个误差向量 $(Y - Y')$, 则需重新校正 W 超平面(权向量)去消除这个误差, 如果输出误差 $(Y - Y') = 0$, 则感知器输出正确。

BP 算法, 即误差反向传播算法实现了多层感知器学习的设想。当给定网络一个输入模式时, 它由输入层单元传到隐层单元, 经隐层单元处理后, 再送到输出层单元, 由输出层单元处理后产生一个输出模式。在正向传播阶段, 每一层神经元的状态只影响下一层神经元的状态。如果输出响应与期望模式有误差, 不满足要求, 那么就转入误差后向传播, 将误差值沿原来的连接通路逐层返回传送并修正各层连接权值, 使误差信号达到最小。对于给定的一组训练模式, 不断用一个训练模式训练网络, 重复前向传播和误差反向传播过程, 当各个训练模式都满足要求时, 我们就说网络已学习好了。下面详细讨论采用 BP 算法的学习过程。

由图 4-14 可知, 输入层中任一神经元的输出为输入模式分量的加权和。其余各层中, 设某一层中任一神经元 j 输入为 net_j , 输出为 y_j , 与这一层相邻的低一层中任一神经元 i 的输出为 y_i 。则有

$$\begin{aligned} net_j &= \sum_i w_{ji} y_i \\ y_j &= f(net_j) \end{aligned} \quad (4.68)$$

式中, w_{ji} 为神经元 j 与神经元 i 之间的连接权; $f(\cdot)$ 为神经元的输出函数, 我们取它为 S 形函数, 即

$$y_j = f(net_j) = \frac{1}{1 + e^{-(net_j + h_j)/\theta_0}} \quad (4.69)$$

式中, h_j 为神经元 j 的阈值, 它影响输出函数水平方向的位置; θ_0 用来修改输出函数的形状。

设输出层第 k 个神经元的实际输出为 y_k , 输入为 net_k , 与输出层相邻的隐层中第 j 个神经元的输出为 y_j , y_k 和 net_k 分别为

$$net_k = \sum_j w_{kj} y_j \quad (4.70)$$

$$y_k = f(net_k) \quad (4.71)$$

对于一个输入模式 X_p , 若输出层中第 k 个神经元的期望输出为 O_{pk} , 实际输出为 y_{pk} , 则输出层的输出方差为

$$E_p = \frac{1}{2} \sum_k (O_{pk} - y_{pk})^2 \quad (4.72)$$

若输入 N 个模式, 则网络的系统均方差为

$$E = \frac{1}{2N} \sum_p \sum_k (O_{pk} - y_{pk})^2 = \frac{1}{N} \sum_p E_p \quad (4.73)$$

权值 w_{kj} 的修改应使 E (或 E_p) 最小。因此, w_{kj} 应沿 E_p 的负梯度方向变化。也就是说, 当输入 X_p 时, w_{kj} 的修正量 $\Delta_p w_{kj}$ 应与 $(-\partial E_p / \partial w_{kj})$ 成正比, 即

$$\Delta_p w_{kj} = -\eta \frac{\partial E_p}{\partial w_{kj}} \quad (4.47)$$

$(-\partial E_p / \partial w_{kj})$ 又可以写成

$$-\frac{\partial E_p}{\partial w_{kj}} = -\frac{\partial E_p}{\partial net_k} \frac{\partial net_k}{\partial w_{kj}} \quad (4.75)$$

由式(4.70)得

$$\frac{\partial net_k}{\partial w_{kj}} = \frac{\partial}{\partial w_{kj}} \sum_j w_{kj} \cdot y_{ji} = y_{ji} \quad (4.76)$$

令 $\delta_{jk} = -\partial E_p / \partial net_k$, 由式(4.71)和式(4.72)得

$$\delta_{jk} = -\frac{\partial E_p}{\partial y_{jk}} \cdot \frac{\partial y_{jk}}{\partial net_k} = (O_{jk} - y_{jk}) f'(net_k) \quad (4.77)$$

由式(4.69)和式(4.70)得

$$f'(net_k) = \frac{\partial}{\partial net_k} \left(\frac{1}{1 + e^{-(net_k + \theta_k)/\theta_0}} \right) = y_{jk} (1 - y_{jk}) \quad (4.78)$$

因此

$$\delta_{jk} = (O_{jk} - y_{jk}) y_{jk} (1 - y_{jk}) \quad (4.79)$$

$$\Delta_p w_{kj} = \eta \delta_{jk} y_{ji} \quad (4.80)$$

对于与输出层相邻的隐层中神经元 j 和比该隐层低一层中的神经元 i , 权值 w_{ji} 的修正量仍应为

$$\Delta_p w_{ji} = -\eta \frac{\partial E_p}{\partial w_{ji}} = -\eta \frac{\partial E_p}{\partial net_j} \frac{\partial net_j}{\partial w_{ji}} = -\eta \frac{\partial E_p}{\partial net_j} = \eta \delta_{ji} y_{ji} \quad (4.81)$$

$$\delta_{ji} = -\frac{\partial E_p}{\partial net_j} = -\frac{\partial E_p}{\partial y_{ji}} \frac{\partial y_{ji}}{\partial net_j} = -\frac{\partial E_p}{\partial y_{ji}} f'(net_j) = -\frac{\partial E_p}{\partial y_{ji}} y_{ji} (1 - y_{ji}) \quad (4.82)$$

式中 $(-\partial E_p / \partial y_{ji})$ 不能直接计算, 可以根据其他已知量计算。具体算法为

$$\begin{aligned} -\frac{\partial E_p}{\partial y_{ji}} &= -\sum_k \frac{\partial E_p}{\partial net_k} \frac{\partial net_k}{\partial y_{ji}} = \sum_k \left(-\frac{\partial E_p}{\partial net_k} \right) \cdot \frac{\partial}{\partial y_{ji}} \sum_m w_{km} y_{jm} \\ &= \sum_k \left(-\frac{\partial E_p}{\partial net_k} \right) w_{kj} = \sum_k \delta_{jk} w_{kj} \end{aligned}$$

因此得到

$$\delta_{ji} = y_{ji} (1 - y_{ji}) \sum_k \delta_{jk} w_{kj} \quad (4.84)$$

$$\Delta_p w_{ji} = \eta \delta_{ji} y_{ji} \quad (4.85)$$

如式(4.79)和式(4.84)所示, 输出层中神经元的输出误差反向传播到前面各层, 以各层之间的权值进行修正。

BP 算法的具体步骤如下:

第一步: 权值和神经元阈值初始化。给所有权值和阈值赋以在 $[0, 1]$ 上分布的随机数。

第二步: 输入样本模式, 指定输出层各神经元的期望输出值 O_1, O_2, \dots, O_M 。

$$O_j = \begin{cases} 1 & X \text{ 属于第 } j \text{ 类} \\ 0 & X \text{ 不属于第 } j \text{ 类} \end{cases} \quad j = 1, 2, \dots, M$$

第三步: 依次计算每层神经元的实际输出, 直到计算出输出层各神经元的实际输出 $(y_1, y_2, y_3, \dots, y_M)$ 。各神经元的实际输出根据式(4.69)计算。

第四步: 修正每个权值。从输出层开始, 逐步向后递推, 直到第一隐层。递推公式为

$$w_{ji}(t+1) = w_{ji}(t) + \eta \delta_{ji} y_{ji}$$

式中, $w_{ji}(t)$ 是 t 时刻从神经元 i (输入层或隐层神经元) 到上一层神经元 j (隐层或输出层神经

元)的连接权。 y_i 是神经元*i*在*t*时刻的实际输出。 η 是步长调整因子, $0 < \eta < 1$ 。如果神经元*j*是输出层一个神经元,则

$$\delta_j = y_j(1 - y_j)(O_j - y_j)$$

如果神经元*j*是一个隐层神经元,则

$$\delta_j = y_j(1 - y_j) \sum_k \delta_k w_{kj}$$

式中, y_j 是神经元*j*在*t*时刻的实际输出; k 是神经元*j*的高一层神经元的编号。如果权值按下面方式修正,收敛会更快,且权值会平滑的变化,即

$$w_{ji}(t+1) = w_{ji}(t) + \eta \delta_j y_i + \alpha [w_{ji}(t) - w_{ji}(t+1)]$$

式中, α 是平滑因子, $0 < \alpha < 1$ 。若把神经元的阈值当成一个权值,相应的输入模式增加一个分量 1,则阈值可以用调整权值的方法调整。

第五步:转到第二步。如此循环,直到权值稳定为止。

BP 算法是一个很有用的算法,因此受到广泛的重视。但也存在一些问题,主要是:存在局部极小值问题;算法收敛速度慢;选取隐单元的数目尚无一般的指导原则;新加入的学习样本会影响已学习样本的学习结果。对隐层的数目对分类性能的问题,李普曼作了简单的论证,其结果如图 4-16 所示。可以证明,包含两个隐层的多层感知器能形成任意复杂的判决界面。即使同类模式分布在模式空间几个不连通的区域中,这种多层网络也能进行正确的判决。一般说来,隐层越多,网络的学习能力越强。鲁姆尔哈特等人发现,若隐单元的数目以指数规律增加,则学习异或问题的速度线性增加。另外一些人发现,对于其他一些问题随隐层数目的增加,学习速度会减小。


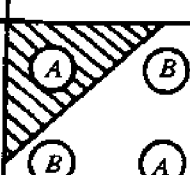

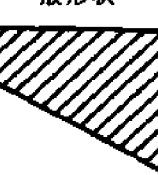
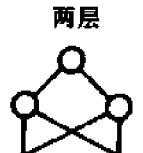
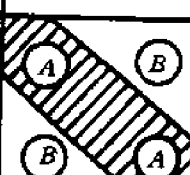
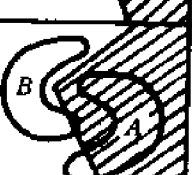
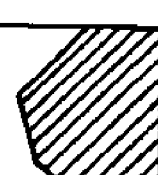
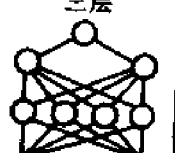
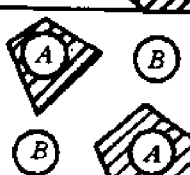


网络结构	分类 分类区域	异或问题 区域类型	网状分类区域	分类区域 一般形状
单层 	由超平面划分的 两个半平面 区域			
两层 	突状开区域 或闭区域			
三层 	任意形状			

图 4-16 单层感知器和多层感知器形成的判决区

2. Hopfield 神经网络

人具有很强的模式识别能力,其主要原因之一就是人具有联想记忆能力。具体地说,人不仅能识别记忆中的一个完整的模式,而且能根据记忆中模式的部分信息进行正确的识别和分类。例如,人能够认出有某种程度缺损或模糊的字符等等。可见人的识别能力具有很强的容错性。所以许多研究人员对人的联想记忆特征进行了长期不懈的研究,提出了一些模拟人脑联想记忆功能的神经网络模型。其中最重要的并对未来神经网络的发展产生了重大影响的是 Hopfield 提出的反馈形式的神经网络模型。在该模型中 Hopfield 引入了网络能量函数来描述网络系统的稳定性;记忆的样本模式以向量形式分布存储于神经元之间的连接权上,并使其对应网络能量函数的局部极小值;每个神经元的输入输出特性为一有界的非线性函数(S型函数);各神经元以随机等概率导步方式进行计算。这种网络由初始状态(初始输出模式向量)向稳定状态演化的过程就是寻找记忆的过程,即一个不完整的模式向完整的模式演化的过程。这种记忆方式为内容寻址记忆,与实际神经系统的记忆方式十分相似。

(1) 离散时间 Hopfield 神经网络:由图 4-17 可知,离散时间 Hopfield 神经网络的各神经元相互连接,神经网络中的每个神经元的输出通过连接权与其余各神经元的输入端连接。输入模式矢量的各分量及神经元的输出值取 +1 和 0(或 -1),它们分别代表神经元的激活或抑制状态。神经元的个数与输入模式矢量的维数相同。样本模式记忆在神经元之间的连接权上。

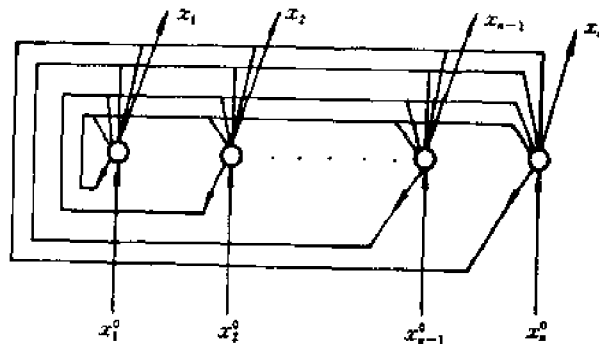


图 4-17 离散时间 Hopfield 神经网络

若有 M 类模式,并设 $X_s = (x_{s1}, x_{s2}, \dots, x_{sn})$ 是第 s 类的样本,为了存储 M 个类样本,要求存储网络的稳定状态集为 $\{X_s\}, s = 1, 2, \dots, M$ 。神经元 i 和神经元 j 之间的权值 w_{ij} 为

$$w_{ij} = \sum_{s=1}^M x_{si} x_{sj} \quad (4.86)$$

若神经元 i 膜电位为 u_i ,输出为 x_i (称为神经元 i 的状态),则 u_i 和 x_i 关系为

$$x_i = f(u_i) = f\left(\sum_{j=1}^n w_{ij} x_j\right) \quad (4.87)$$

式中,函数 $f(\cdot)$ 定义为符号函数,即

$$f(u_i) = \begin{cases} +1 & u_i > 0 \\ -1 & u_i < 0 \end{cases}$$

若输入一个未知模式 X_0 ,网络的初始状态 X 由 X_0 决定,即令 $x_i = X_{0i}, i = 1, 2, \dots, n$ 。那么根据以上算法,网络从初始状态开始逐步演化,最终趋向于一个稳定状态,即网络最终输出一个

与未知模式最相似的样本模式。为了说明这一点,我们在这里定义网络的能量函数为

$$E = -\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n w_{ij} x_j x_i \quad (4.88)$$

由式(4.88)可知, E 随 x_i 的变化而变化。由某一个神经元状态的变化 Δx_i 引起的 E 的变化量为

$$\Delta E = -\frac{1}{2} \left(\sum_{j=1}^n w_{ij} x_j \right) \Delta x_i \quad (4.89)$$

式中, $w_{ij} = w_{ji}$, $w_{ii} = 0$ 。由式(4.89)可知,当式(4.89)中的 $\sum(\cdot)$ 项为正时, Δx_i 也为正值;当 $\sum(\cdot)$ 项为负时, Δx_i 也为负值。这就是说,当 x_i 随时间变化时, E 的变化量总是小于零。因为 E 是有界的,所以算法最终使网络达到一个不随时间变化的稳定状态。

Hopfield 网络算法的具体步骤如下:

第一步:给神经元的连接权赋值,即存储样本模式。设 $X_s = (x_{s1}, x_{s2}, \dots, x_{sn})^T$ 是第 s 个模式类的样本模式,则神经元 i 和神经元 j 之间的连接权为

$$w_{ij} = \sum_{s=1}^M x_{si} x_{sj} \quad i, j = 1, 2, \dots, n; \quad i \neq j$$

第二步:输入未知模式 $X_0 = (x_{01}, x_{02}, \dots, x_{0n})^T$,用 x_{0i} 设置神经元 i 的初始状态。若 $x_i(t)$ 表示神经元 i 在 t 时刻的状态(输出),则 $x_i(t)$ 的初始值为

$$x_i(t) = x_{0i} \quad i = 1, 2, \dots, n$$

第三步:用迭代算法计算 $x_i(t+1)$,直到算法收敛。 $x_i(t+1)$ 根据式(4.87)计算,即

$$x_i(t+1) = f \left[\sum_{j=1}^n w_{ij} x_j(t) \right] \quad i = 1, 2, \dots, n$$

计算进行到神经元的输出不随进一步的迭代而变化时算法收敛。此时神经元的输出即为与未知模式匹配最好的样本模式。

第四步:转到第二步,输入新模式。

(2) 连续时间 Hopfield 神经网络:对于连续时间情况,Hopfield 提出了图 4-18 所示的神经网络模型,这同样由 n 个神经元互连而成,但神经元的输出不在是离散值 0(或-1)和 1,而可以在某一区间连续变化。

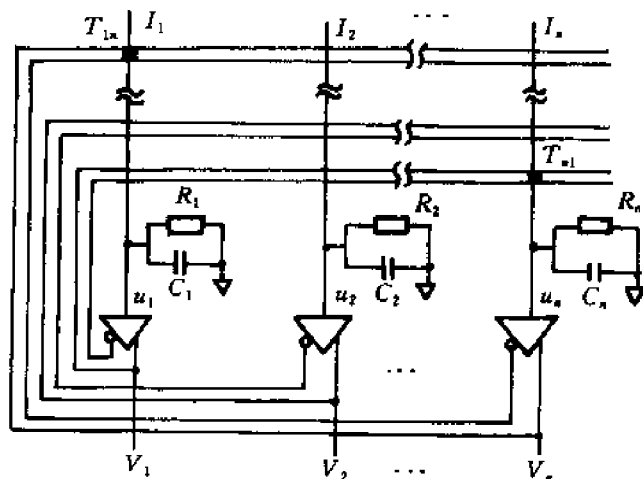


图 4-18 连续时间 Hopfield 神经网络

在连续时间模型中,神经元可以看成由电阻、电容和运算放大器组成。电阻 R_i 和电容 C_i 并联,模拟生物神经元输出的时间常数。神经元 i 和神经元 j 之间的连接权 w_{ij} 可以看成神经元 j 的输出到神经元 i 的输入之间的互电导,它模拟神经元之间的突触特性。运算放大器模拟生物神经元的非线性特性。放大器的“+”输出端和“-”输出端分别模拟生物神经元的兴奋和抑制特性。 $w_{ij} > 0$ (对应于兴奋性突触) 时,神经元 i 的输入和神经元 j 的“+”输出端相接; $w_{ij} < 0$ (对应于抑制性突触) 时,神经元 i 的输入和神经元 j 的“-”输出端相接。每个神经元的输出通过连接权与其余神经元的输入连接。外界输入 I_i 用来建立一般的兴奋电平。若神经元 i 的输入为 u_i , 输出为 V_i , 且神经网络有 n 个神经元互连, 则神经元状态变量的动态变化可用下面的非线性和微分方程描述为

$$\begin{cases} C_i \frac{du_i}{dt} = \sum_{j=1}^n w_{ij} V_j - \frac{u_i}{R_i} + I_i \\ V_i = f(u_i), \text{ 成 } u_i = f^{-1}(V_i) \end{cases} \quad i = 1, 2, \dots, n \quad (4.90)$$

这里取 $f(u_i)$ 为 S 形函数。

$$f(u_i) = \frac{1}{1 + e^{-(u_i + h_i)/\theta_0}}$$

式中, h_i 为神经元的 i 阈值, θ_0 用来改变 $f(u_i)$ 曲线渐近“0”和“1”的陡峭或缓慢程度。

$f(u_i)$ 连续时间网络模型的系统能量函数定义为

$$E = \frac{1}{2} \sum_i \sum_j w_{ij} V_i V_j - \sum_i V_i I_i + \sum_i \frac{1}{R_i} \int_0^{V_i} f^{-1}(V) dV \quad (4.91)$$

下面根据 E 的定义讨论网络的稳定性。要使网络随时间最终趋于一个稳定状态, E 随时间的增加必须逐渐减小, 即必须 $dE/dt < 0$ 。为此, 我们求 dE/dt , 并假定 $w_{ij} = w_{ji}$, $C_i > 0$ 。根据式 (4.90) 和式 (4.91) 得到 dE/dt 为

$$\begin{aligned} \frac{dE}{dt} &= \sum_i \frac{dE}{dV_i} \frac{dV_i}{dt} = \sum_i \frac{dV_i}{dt} \left(- \sum_j w_{ij} V_j - I_i + \frac{u_i}{R_i} \right) = \\ &= \sum_i \frac{dV_i}{dt} C_i \frac{du_i}{dt} = \sum_i C_i \frac{du_i}{dV_i} \left(\frac{dV_i}{dt} \right)^2 = \sum_i C_i f^{-1}(V_i) \left(\frac{dV_i}{dt} \right)^2 \end{aligned}$$

由于 $C_i > 0$, 且 $f^{-1}(V_i)$ 单调递增, 因此有 $dE/dt \leq 0$, 即从任意初态开始, 随着时间的增加, 网络的状态轨道总是朝着能量减小的方向运动, 网络最终达到稳定平衡点, 即 E 的极小值点。

$$E = - \frac{1}{2} \sum_i \sum_j w_{ij} V_i V_j - \sum_i V_i I_i \quad (4.93)$$

根据式 (4.93), 要记忆 M 个类样本 $X_i = (x_{i1}, x_{i2}, \dots, x_{in})$, 神经元 i 和神经元 j 之间的权值可以根据下式构成

$$w_{ij} = \sum_k x_{ki} x_{kj}, \quad i = 1, 2, \dots, n \quad (4.94)$$

每个 X_i 是网络的一个渐近平衡点, 也对应 E 的一个局部极小点。当输入模式是记忆中一个样本的某种程度的变形时, 网络随时间演化, 量终达到一个稳定状态, 输出一个与输入模式最相似的样本。

Hopfield 神经网络模型的局限性主要有两个方面。其一, 网络能够记忆和正确回顾的样本数是相当有限的。如果记忆的样本数太多, 网络可能收敛于一个不同于所有记忆中样本的伪模式。若网络用于模式类的样本数小于网络中神经元数 (或模式矢量的分量数) 的 0.15 倍时, 收

敏于伪样本的情况才不会发生。其次,如果记忆中的某一样本的某些分量与别的记忆样本的对应分量相同时,这个记忆样本可能是一个不稳定的平衡点。

3. 自组织神经网络

多层感知器的学习和分类是以已知一定的先验知识为条件的,即网络权值的高速是在监督情况下进行的。而在实际应用中,有时并不能提供所需的先验知识,这就需要网络具有能够自学习的功能。Kohonen 提出的自组织特征映射图就是一种具有自学习功能的神经网络,这种网络是基于生理学和脑科学研究成果提出来的。脑神经科学研究表明:传递感觉的神经元排列是按某种规律有序进行的,这种排列往往反映所感受的外部刺激的某些物理特征。例如在听觉系统中,神经细胞的纤维是按照其最敏感的频率分布而排列的。为此,Kohonen 认为,神经网络在接受外界输入时,将会分成不同的区域,不同的区域对不同的模式具有不同的响应特征,即不同的神经元以最佳的方式响应不同性质的信号激励,从而形成一种拓扑意义上的有序图。这种有序图也称为特征图,它实际上是一种非线性映射关系,它将信号空间中各模式的拓扑关系几乎不变地反映在这张图上,即各神经元的输出响应上。由于这种映射是通过无监督的自适应过程来完成的,所以也称为自组织特征图。

如图 4-19 所示,在这种网络中,输出节点与邻域其它节点广泛相连,并互相激励。输入节点和输出节点之间通过强度 $w_{ji}(t)$ 相连接,通过某种规则,不断地调整 $w_{ji}(t)$,使得在稳定时,每一邻域的所有节点对某种输入具有类似的输出,并且这种聚类的概率分布与输入模式的概率分布相接近。

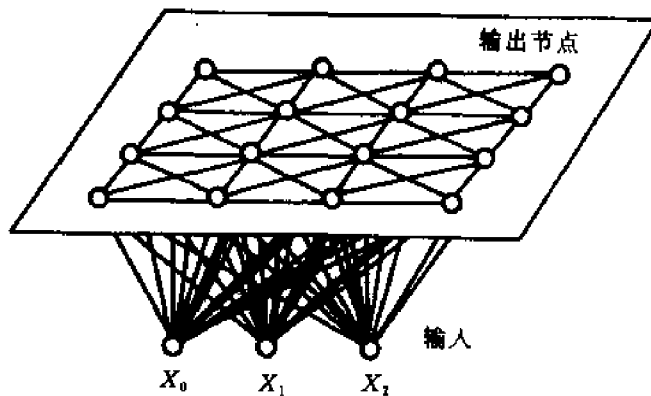


图 4-19 自组织神经网络

完成自组织特征映射的算法较多。一种常用的自组织算法步骤如下:

- (1) 权值初始化并选定邻域的大小;
- (2) 输入模式;
- (3) 计算空间距离

$$d_j = \sum_{i=1}^{N-1} [x_i(t) - w_{ji}(t)]^2 \quad (4.95)$$

式中, $x_i(t)$ 是 t 时刻 i 节点的输入; $w_{ji}(t)$ 是输入节点 i 与输出节点 j 的连接强度; N 为输入节点的数目;

- (4) 选择节点 j^* , 它满足 $d_{j^*} = \min d_j$;
- (5) 按下式改变 j^* 和其邻域节点的连接强度

$$w_{ji}(t+1) = w_{ji}(t) + \eta(t)[x_i(t) - w_{ji}(t)]$$

式中, $\eta(t)$ 称为衰减因子, 取值范围为 $0 < \eta(t) < 1$;

(6) 返回到(2) 直到满足 $[x_i(t) - w_{ji}(t)] < \epsilon$ (ϵ 为给定的误差)。

通过这种无指导的学习, 稳定后的网络输出就对输入模式生成自然的特征映射, 从而达到自动聚类的目的。

4. 概率神经网络

概率神经网络与统计信号处理的许多概念有着紧密的联系。当这种网络用于检测和模式分类时, 可以得到贝叶斯最优结果。图 4-20 为一概率神经网络的示意图, 它通常由 4 层组成。第一层为输入层, 每个神经元均为单输入单输出, 其传递函数也为线性的, 这一层的作用只是将输入信号用分布的方式来表示。第二层称之为模式层, 它与输入层之间通过连接权值 w_{ji} 相连接, 该层第 i 个神经元的输入 $Z_i = X^T \cdot W_i$, 其中 X 为输入列矢量, $W_i = [w_{i1}, w_{i2}(t), \dots, w_{ip}(t)]^T$ 。模式层神经元的传递函数不再是通常的 S 函数, 而为 $f(Z_i) = \exp[(Z_i - 1)/\sigma^2]$, 模式层结构如图 4-21 所示。第三层称之为累加层, 它具有线性求和的功能, 这层的神经元数目与欲分的模式数目相同(图 4-20 中对应着两模式)。第四层输出层具有判决功能, 它的神经元输出为离散值 1 和 -1(或 0), 分别代表着输入模式的类别。

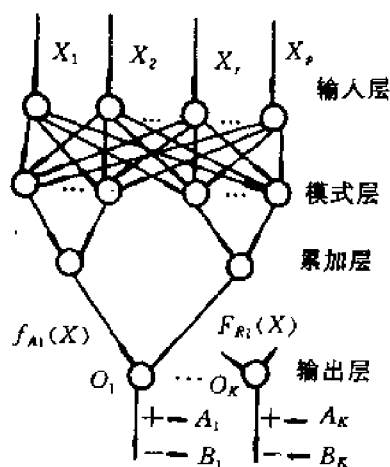


图 4-20 概率神经网络

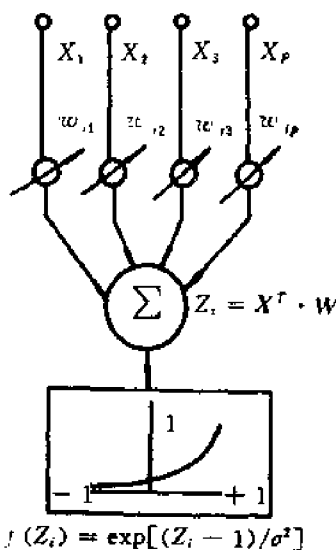


图 4-21 模式层第 i 个神经元

概率神经网络具有如下特点:

- (1) 训练容易, 收敛速度快, 从而非常适用于实时处理;
- (2) 可以完成任意的非线性变换, 所形成的判决曲面与贝叶斯最优准则下的曲面相接近;
- (3) 具有很强的容错性;
- (4) 模式层的传递函数可以选用各种用来估计概率密度的核函数, 并且分类结果对核函数的形式不敏感;
- (5) 各层神经元的数目比较固定, 因而易于硬件实现。

这种网络已经较广泛地应用于非线性滤波、模式分类、联想记忆和概率密度估计当中。但

是它的权值学习算法及有关的理论有待于更进一步的研究。

4.6.3 神经网络在机械故障诊断中的应用

多层感知器与 BP 算法在工程中应用很广,应用中选择网络结构参数非常重要。输入层节点数一般根据输入的特征多少来定,输出层节点数选取有两种方式:一种是根据输出的分类数来定,一种是将输出按二进制编码;隐层节点数选取还不够清楚,依靠经验和试验来选取。在我们所举例的神经网络学习与诊断系统中,构造的三层感知器分类器的输出数与故障类别数相等,隐层节点数等于输入层节点数,输入层节点数为待识别状态信号的特征数。

表 4-8 是一组旋转机械故障的训练示例,表内的值表示各训练示例的特征值大小,其取值区间为 $[0,1]$,如在不平衡训练示例中, $0 \sim 1/4$ 倍频的振动幅值的当量值为 0; $1/4 \sim 3/4$ 倍频的振动幅值的当量值为 0; $3/4 \sim 1$ 倍频的振动幅值的当量值还为 0;1 倍频的振动幅值的当量值为 0.9;2 倍频的振动幅值的当量值为 0.1;等等,其余类推。

表 4-8 旋转机械的故障训练示例

	$0 \sim \frac{1}{4}$ 倍频	$\frac{1}{4} \sim \frac{3}{4}$ 倍频	$\frac{3}{4} \sim 1$ 倍频	1 倍频	2 倍频	3 倍频	高次偶频	低次偶频
不平衡	0	0	0	0.9	0.1	0	0	0
油膜涡动	0	0.6	0	0.3	0.1	0	0	0
不对中	0	0	0	0.6	0.4	0	0	0

将这些故障训练示例输入到一个具有 8 个输入层节点,8 个中间层节点和 3 个输出层节点的网络中,经过 1 200 次迭代,形成了一个网络,该网络的记忆效果如表 4-9 所示,经过 12 000 次迭代所形成网络的记忆效果如表 4-10 所示。

表 4-9 1 200 次迭代所形成网络的记忆效果

	不平衡	油膜涡动	不对中
不平衡鼓掌	0.94	0.00	0.06
油膜涡动鼓掌	0.00	0.96	0.04
不对中鼓掌	0.06	0.04	0.90

表 4-10 12 000 次迭代所形成网络的记忆效果

	不平衡	油膜涡动	不对中
不平衡鼓掌	0.98	0.00	0.02
油膜涡动鼓掌	0.00	0.96	0.04
不对中鼓掌	0.06	0.02	0.92

表 4-9 中第一行表示,当输入一组不平衡故障时,得出该故障的置信度为 0.94,而其他故障几乎为 0;第二行表示,当输入一组油膜涡动故障时,得出该故障的置信度为 0.96,而其他故障几乎为 0;第三行表示,当输入一组不对中故障时,得出该故障的置信度为 0.90,而其他故障几乎为 0。表 4-10 的结果有所改进,其值已趋于稳定。通过比较表 4-9 与表 4-10,可看出训练中迭代次数越多,所得到的网络越能够更好地联想出训练示例。但训练次数也不宜过长,只要

满足精度要求,训练次应尽可能少,以减少训练时间。

习题与思考题

4-1 如果只知道各种故障类的先验概率,最小错误率贝叶斯决策规则应如何表示?

- 4-2 (1) 对 c 类情况推广最小错误率贝叶斯决策规则;
 (2) 指出此时使错误率最小等价于后验概率最大,即

$$P(\omega_j/X) > P(\omega_i/X) \quad \text{对一切 } j \neq i \text{ 成立。}$$

4-3 对两类问题,证明最小风险贝叶斯决策规则可表示为

$$\text{若 } \frac{p(X/\omega_1)}{p(X/\omega_2)} > (\text{或 } <) \frac{(\lambda_{12} - \lambda_{22})p(\omega_2)}{(\lambda_{21} - \lambda_{11})p(\omega_1)} \quad \text{则 } X \in \begin{cases} \omega_1 \\ \omega_2 \end{cases}$$

4-4 谎报与漏检有什么不同?在设备状态监测中哪一种失误应控制得多一些?

4-5 主分量分析的原理是什么?步骤有哪些?

4-6 若有下列两类样本集:

$$X_1^1 = (0,0,0)^T \quad X_1^2 = (0,0,1)^T$$

$$X_2^1 = (1,0,0)^T \quad X_2^2 = (0,1,0)^T$$

$$X_3^1 = (1,0,1)^T \quad X_3^2 = (0,1,1)^T$$

$$X_4^1 = (1,1,0)^T \quad X_4^2 = (1,1,1)^T$$

用主分量分析分别把特征空间维数降到 $d=2$ 和 $d=1$,并用图画出样本在特征空间中的位置。

4-7 设备状态识别中,常用的空间距离函数有哪些?其特点如何?

4-8 设备状态识别中,常用的信息距离判别函数有哪些?其特点如何?

4-9 应用距离函数对机械设备状态进行分类识别时应注意哪些问题?

4-10 设论域 $\Omega = \{a, b, c, d, e, f\}$ 上 5 类模式

$$A_1 = \frac{0.8}{a} + \frac{0.3}{b} + \frac{0.2}{c} + \frac{0}{d} + \frac{0.5}{e} + \frac{0.1}{f}$$

$$A_2 = \frac{0.7}{a} + \frac{1}{b} + \frac{0.3}{c} + \frac{0}{d} + \frac{0.3}{e} + \frac{0.9}{f}$$

$$A_3 = \frac{0.2}{a} + \frac{1}{b} + \frac{0.8}{c} + \frac{0.4}{d} + \frac{0.5}{e} + \frac{0.1}{f}$$

$$A_4 = \frac{0.8}{a} + \frac{0}{b} + \frac{0.4}{c} + \frac{0.2}{d} + \frac{0.7}{e} + \frac{0}{f}$$

$$A_5 = \frac{0.5}{a} + \frac{0.3}{b} + \frac{0.6}{c} + \frac{0.6}{d} + \frac{0}{e} + \frac{0.4}{f}$$

今有样本

$$B = \frac{0.7}{a} + \frac{0.4}{b} + \frac{0.6}{c} + \frac{0.1}{d} + \frac{0.2}{e} + \frac{0.8}{f}$$

问应将 B 划归哪一类。

4-11 试举出利用最大隶属原则和择近原则进行故障模糊识别的例子。

4-12 对某产品质量作综合评判,考虑从 4 种因素来评价产品, $X = \{x_1, x_2, x_3, x_4\}$

将产品质量分为 4 等 $U = \{u_1, u_2, u_3, u_4\}$, 设单因素决断为模糊映射 $\tilde{f}: X \rightarrow \tilde{F}(U)$

$$\tilde{f}(x_1) = (0.3, 0.6, 0.1, 0); \quad \tilde{f}(x_2) = (0, 0.2, 0.5, 0.3)$$

$$\tilde{f}(x_3) = (0.5, 0.3, 0.1, 0.1) \quad \tilde{f}(x_4) = (0.1, 0.3, 0.2, 0.4)$$

设有两种因素全重分配

$$A_1 = (0.5, 0.2, 0.2, 1); \quad A_2 = (0.2, 0.4, 0.1, 0.3)$$

试评价此产品按两种全重分配情况下, 分别相对地属于哪级产品。

- 4-13 试举一模糊综合评判的实际机械故障诊断的例子(由建立评判空间到综合评判结束)。
- 4-14 多层感知器网络中只有一层连接权可调节, 那么, 采用多层感知器的原因何在?
- 4-15 试述 BP 网络学习过程中, 输出误差后向传播及权值调节的原理。
- 4-16 神经网络是如何用于机械故障诊断的? 试从本章应用举例中给以总结。

旋转机械的振动监测与故障诊断

5.1 概 述

5.1.1 振动的几种类型

旋转机械指汽轮机、燃气轮机、发电机、电动机、离心压缩机、水轮机、航空发动机等机械设备,它的主要构成部件有转子、支承转子的轴承、定子或机器壳体、联轴器等等。转速范围一般为几千 r/min 至几十万 r/min,这类机组通常称为高速旋转机械,当然也有低速的水轮机机组,其转速只有 100r/min。由于旋转机械的结构及其零部件的加工和安装方面的缺陷,使机器在运行时引起振动,其振动类型可分为横向振动、轴向振动和扭转振动三类。其中过大的横向振动往往是机器破坏的主要原因,所以成了振动监测的主要对象,也是对机组状态进行诊断的主要依据。

转轴组件在其运行转速低于其一阶横向自振频率时,称为刚性转子,例如,常见的电动机转子。但也有运行转速高于本身的一阶横向自振频率的转子,例如燃气轮机转子、压缩机转子等,这类转子称为柔性转子。

旋转机械中的振动可分为两种形式:其一是同步振动,又称强迫振动,主要由转子的不平衡、联轴器的不对中、安装不良等原因造成。其振动的频率为转子的回转频率及其倍频;振动的振幅,在转子的临界转速前,随着转速的增加而增大,超过临界转速,则随转速 n 的增加而减小,在临界转速处,有一共振峰值。对转子来说,其激振原因是由于转子的不平衡(转子的质心与回转轴线偏离)而在旋转时产生的离心惯性力,该力的大小取决于转子的不平衡程度。因此,对转子进行动平衡处理,减小转子质心与回转轴线的偏距,即可减轻这种振动。在设计上,使转子在远离临界转速处运行,以避免共振的发生。

另一种振动的形式是亚同步振动,又称自激振动,其振动频率低于转子的回转频率。由于这种差异而在转子和定子中产生交变应力,并且,这种振动常常在某个转速下(大于临界转速)突然发生,因而对旋转机械具有极大的危害性。亚同步振动对环境条件的变化是十分敏感的,机器的微小差别,

其稳定性可能具有极大的差异。

5.1.2 转子的临界转速

旋转机械在启停升降速过程中,往往在某个(或某几个)转速下出现振动急剧增大的现象,有时甚至在工作转速下振动也比较强烈。其振动原因往往是由于转子系统处于临界转速附近产生共振。

在无阻尼的情况下,转子的临界转速等于其横向固有频率,因此转子的临界转速个数与转子的自由度相等。对实际转子来说,理论上有无穷多个临界转速,但由于转子的转速限制,往往只能遇见数个临界转速。

在有阻尼的情况下,转子的临界转速略高于其横向固有频率。

根据转子的工作转速 n 与其第一阶临界转速 $n_{c,r,1}$ 间的关系,可划分:

$n < 0.5n_{c,r,1}$	刚性转子
$0.5n_{c,r,1} \leq n < 0.7n_{c,r,1}$	准刚性转子
$n \geq 0.7n_{c,r,1}$	柔性转子

刚性转子与柔性转子两者的动力学特性有很大不同,这对于动平衡来说十分重要。

5.1.3 旋转机械振动评定标准

目前最常采用的是通频振幅来衡量机械运行状态的,根据所使用传感器的种类分为:

(1) 轴承振动评定:这可以利用接触式传感器(例如磁电式振动速度传感器或压电式振动加速度传感器)放置在轴承座上进行测量。

(2) 轴振动值评定:这可利用非接触式传感器(例如涡流式传感器)测量轴相对于机壳的振动值或轴的绝对振动值。

评定参数可用振动位移峰—峰值和振动烈度(即均方根值,它代表了振动能量的大小)来表示。

1. 以轴承振动位移峰—峰值作评定标准

表 5-1 为水电部 1959 年颁发的《电力工业技术管理法规》中关于汽轮发电机组轴承的振动标准。要求机组垂直、水平和轴向均满足该标准。

表 5-1 水电部汽轮机组振动标准(双振峰)

转速 r/min	标准 / μm		
	优	良	合格
1 500	30	50	70
3 000	20	30	50

表 5-2 为机械部关于《离心鼓风机和压缩机技术条件》中规定的轴承振动标准。

表 5-2 离心鼓风机和压缩机振动标准

振动标准 μm	转速 / ($\text{r} \cdot \text{min}^{-1}$)			
	$\leq 3\,000$	$\leq 6\,500$	$\leq 10\,000$	$> 10\,000 \sim 16\,000$
主轴承	50	≤ 40	≤ 30	≤ 20
齿轮轴承		≤ 40	≤ 40	≤ 30

表 5-3 为国际电工委员会 IEC 推荐的汽轮机振动标准。

表 5-3 IEC 汽轮机振动标准

振动标准 μm	转速 / (r · min ⁻¹)				
	≤ 1 000	1 500	3 000	3 600	≥ 6 000
轴承上	75	50	25	21	12
轴上(靠近轴承)	150	100	50	44	20

由上述三表可以看出,转速低,允许的振动值大;转速高,允许的振动值小。这是因为对于同样的振动值,高速机组比低速机组更易出现故障。同时还必须强调,上述表中的振幅均为双振幅,即峰-峰值,而有些国家、公司采用峰值标准,相应允许振动值要减少 1/2,这点应给予重视,以免造成不必要的损失。

2. 以轴承振动烈度作为评定标准

在制定上述振动标准时,假设:

- (1) 机组振动为单一频率的正弦波振动;
- (2) 轴承振动和转子振动基本上有一固定的比值,因此可利用轴承振动代表转子振动;
- (3) 轴承座在垂直、水平方向上的刚度基本上相等,即认为是各向同性的。

实际证明上述假设与事实不尽相符,所测得的振动多数是由数种频率的振动合成的;轴承组水平刚度明显低于垂直刚度;转子振动和轴承座振动的比值,可以是 2 ~ 50 倍,它和轴承型式、间隙、轴承座刚度、油膜特性等有关,且同类机组亦不尽相同,因此,为了较全面的反映机组的振动情况,必须制定其他的振动标准。

国际标准化组织 ISO3945 给出了用振动烈度评定功率大于 300 kW,转速为 600 ~ 12 000 r/min 的大型原动机和其他具有旋转质量的大型机器,例如电动机和发电机、蒸汽轮机和燃气轮机、涡轮压缩机、涡轮泵和风扇等振动特性的国际标准,如表 5-4 所示。

表中的支承情况分为刚性和柔性两种,刚性支承是指机器—支架系统的第一阶固有频率高于主激振频率。

表 5-4 ISO3945 振动标准

轴承振动烈度 v_{rms} , (mm · s ⁻¹)	支承分类	
	刚性	柔性
0.46	好	好
0.71		
1.17	良	良
1.8		
2.8		
4.6	及格	及格
7.1		
11.2	不可用	不可用
18.0		
28.0		
45.0		

振动烈度计算公式为

$$v_{rms} = \sqrt{\frac{1}{T} \int_0^T v^2(t) dt} = \sqrt{\frac{1}{n} (v_1^2 + v_2^2 + \dots + v_n^2)} =$$

$$\sqrt{\frac{1}{n}(A_1^2\omega_1^2 + A_2^2\omega_2^2 + \cdots + A_n^2\omega_n^2)} \quad (5.1)$$

式中, $\omega_1, \omega_2, \dots, \omega_n$ 为非简谐振动的各个角频率; v_1, v_2, \dots, v_n 为相应角频率下的振动速度值; A_1, A_2, \dots, A_n 为相应角频率下的振动位移峰值。

3. 以轴振动的位移峰—峰值作为评定标准

在轴承座上测量机组的振动较为方便, 然而机组转子的振动要通过油膜传到轴承座, 所测得的振动幅值受油膜刚度和轴承刚度的影响, 因此直接测量转轴的振动值以判断转子的振动特性将能更确切地反映振动的实质。

表 5-3 所列出的 IEC 推荐的振动标准中包含有轴振动的标准。

美国石油学会给出了功率不超过 1 000 kW 的中小型涡轮机械轴振动的振动标准 API617, 其振动许可值为

$$A = 25.4 \sqrt{\frac{12\,000}{n}} \quad (5.2)$$

式中, A 为振动许可值(双振幅), 单位为 μm ; n 为机器的转速, 单位为 r/min 。

以上振动标准不能机械地套用, 还应结合机组的振动趋势综合考虑, 如长期振动较小的机组或测点, 当其振动值增加但还仍未超过振动标准时, 这也是故障征兆, 应给予足够的重视, 以防产生不良后果。

5.2 旋转机械振动监测参数与分析

5.2.1 监测参数

对于一台机器, 有许多物理量可以测量, 为了达到故障诊断目的, 应该选择那些量作为监测参数呢? 由于机器的振动情况直接反映了机器运行状态的优劣, 机器的许多故障都以振动形式反映出来, 振动为故障诊断提供了重要信息, 因此振动是故障诊断必须监测的参数之一。此外, 与之相关的过程参数、工艺参数等也是故障诊断的有用参数。

监测参数可分为动态参数和静态参数两种。

1. 动态参数

(1) 振幅: 它表示振动的严重程度, 可用位移、速度或加速度表示。常用的振动传感器的使用频率范围, 如图 5-1 所示。

(2) 振动烈度: 近年来国际上已统一使用振动烈度作为描述机器振动状态的特征量。其表达式见式(5.1)。

(3) 相位: 它对于确定旋转机械的动态特性、故障特性及转子的动平衡等具有重要意义。

2. 静态参数

(1) 轴心位置: 指在稳定情况下, 轴承中心相对于转轴轴颈中心的位置。在正常工况下, 转轴在油压、阻尼作用下在一定的位置上浮动。在异常情况下, 由于偏心太大, 会发生轴承磨损的故障。

(2) 轴向位置: 是机器转子上止推环相对于止推轴承的位置, 当轴向位置过小时, 易造成动静摩擦, 产生不良后果。

(3) 差胀:指旋转机械中转子与静子之间轴向间隙的变化值。它对机组安全启动具有十分重要的意义。

(4) 对中度:指轴系转子之间的连接对中程度,它与各轴承之间的相对位置有关,不对中故障是旋转机械的常见故障之一。

(5) 温度:轴瓦温度反映轴承运行情况。

(6) 润滑油压:反映滑动轴承油膜的建立情况。

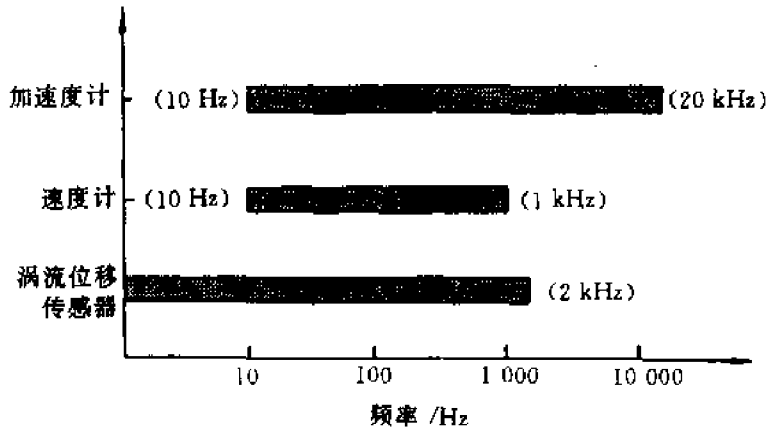


图 5-1 常用振动传感器的测量范围

5.2.2 旋转机械故障振动分析

1. 常态频域分析

(1) 振动频率:线性系统中振动的频率应等于激振力的频率,而激振力又是零部件故障产生的,因此,测量了转轴组件的振动频率,就可以找到激励源。但是,当转轴组件对中不良、松动或过载时,系统出现非线性刚度,振动频率中会包含有激振频率的高次谐波。图 5-2 所示是一台在管道共振区附近运行的汽轮机,由于内部质量不平衡造成过载,从而迫使轴承在非线性区域工作。图上出现激振频率的高次谐波。

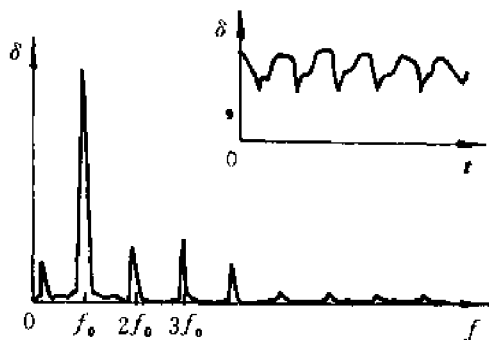


图 5-2 汽轮机由于不平衡引起的振动

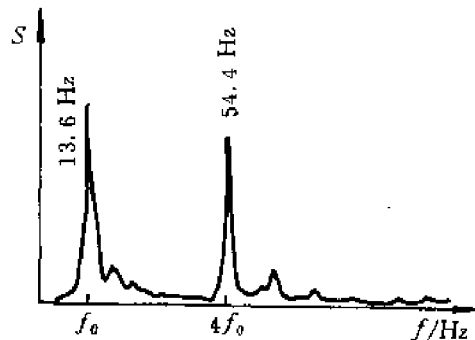


图 5-3 一台大型水泵的振动频谱

(2) 多重频率:如果振动传感器测量的信号是两个信号迭加,则此两个信号的频率可以在谱图上区分出来。图 5-3 是一台转速为 800 r/min 的大型水泵的振动频谱。由于水泵的叶轮上

有四个叶片,每个叶片上的水量不等(不平稳)而产生基频信号,频率 54.4 Hz(4×13.6) 是叶片与水撞击形成的。

(3) 脉冲激发:当机器受到冲击载荷时,机器就会按其固有频率进行振动,如图 5-4(a) 所示,功率谱图上将有一谱峰位于机器的固有频率处。如果机器中零件的缺陷比较严重,则此固有频率还被缺陷的重复频率 f_1 所调制而产生边频,如图 5-4(b) 所示。这方面典型的例子是齿轮传动,将在第 7 章讨论。

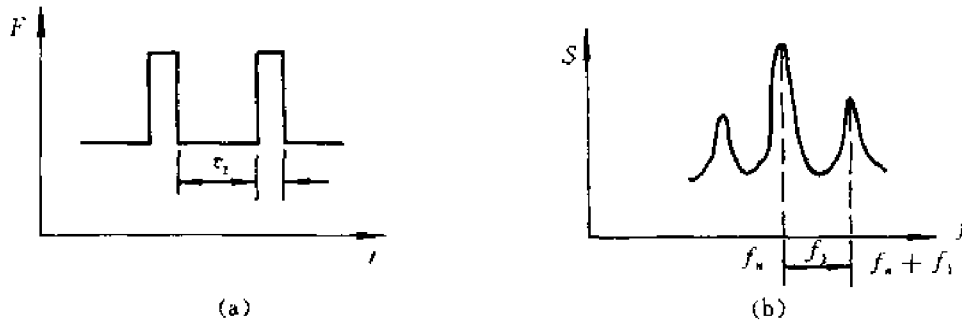


图 5-4 机械系统对脉冲激发的响应

(a) 脉冲激发 (b) 振动谱图

(4) 拍:拍的现象在旋转机械振动中经常出现,当两种振动频率相近且幅值相等时,叠加起来就会产生拍

$$x = a\sin\omega_1 t + a\sin\omega_2 t = 2a\cos\left(\frac{\omega_1 - \omega_2}{2}\right)\sin\left(\frac{\omega_1 + \omega_2}{2}\right)$$

如图 5-5 所示,时域信号的包络,相当于频率等于 $0.5(\omega_1 - \omega_2)$ 的缓慢波动。在功率谱上,可以分辨出两个频率十分接近,高度又接近相等的谱峰。

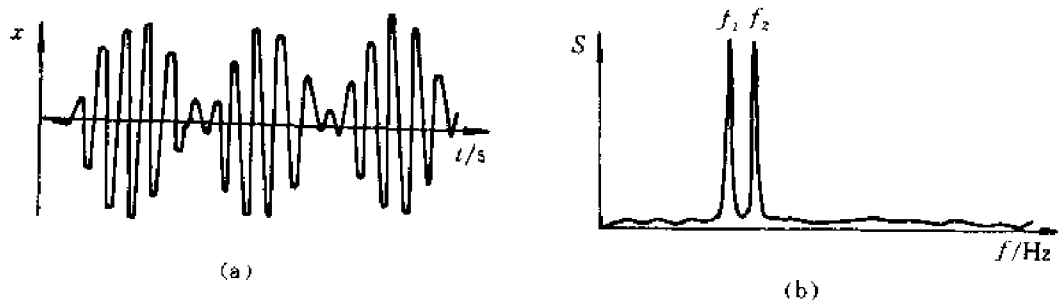


图 5-5 拍及其在频域中的分解

(a) 时域曲线 (b) 功率谱

(5) 频率和差规律:它是指旋转机械功率谱图上各个谱峰的中心频率等于两个频率的和或差值,例如,等于 $(\omega_2 - \omega_1)$, $(\omega_1 + \omega_2)$, $(\omega_1 + 2\omega_2)$, $(2\omega_2 - \omega_1)$, $(2\omega_1 + \omega_2)$, $(2\omega_1 - \omega_2)$ 等等。这种现象是当时域信号形成拍并且单边削平时(见图 5-6)经频域变换得到的。

图 5-7 说明了这种时域信号的由来。旋转机械中诸如失衡、齿轮和滚动轴承的缺陷、油膜涡动、封存的流体,以及轴颈与油封的摩擦等激振原因促使转子以一定的频率振动,当机器存在对中不良、松动和刚度的非线性时,转子的振动传到定子上就形成了单边削平的现象。单边削平的拍在功率谱上出现的频率分量的多少,取决于削平的深度。

最早用来分析频率和差规律的例子是燃气轮机的振动,燃气轮机的转子存在两个激振原因,一是失衡,激振频率等于转子的转动频率 f_1 ; 另一个激振原因是转子中残存的流体,其激振频率 f_2 等于 $0.883f_1$ 。在两个振动的共同作用下,转子振动便形成拍。图 5-8 解释了当转子在轴承间隙中偏心安装时如何使定子的输出波形出现削平的现象。设 $y_0(t)$ 为转子的输入波形, $y(t)$ 为定子的输出波形,当转子轴颈与定子轴承接触并一起振动时,相当于

$$\text{当 } y_0(t) \geq y_1 \text{ 时} \quad y(t) = y_0(t)$$

而当转子与定子脱开后,即

$$\text{当 } y_0(t) < y_1 \text{ 时} \quad y(t) = y_1$$

式中 y_1 为干涉距离。图 5-9 是这台燃气轮机外壳振动的功率谱图。

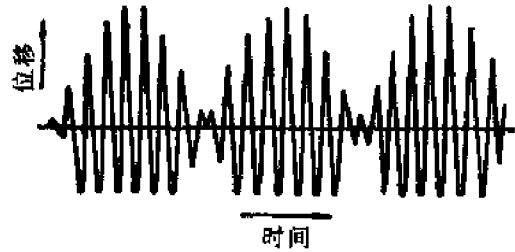


图 5-6 拍经单边削平后的时域信息

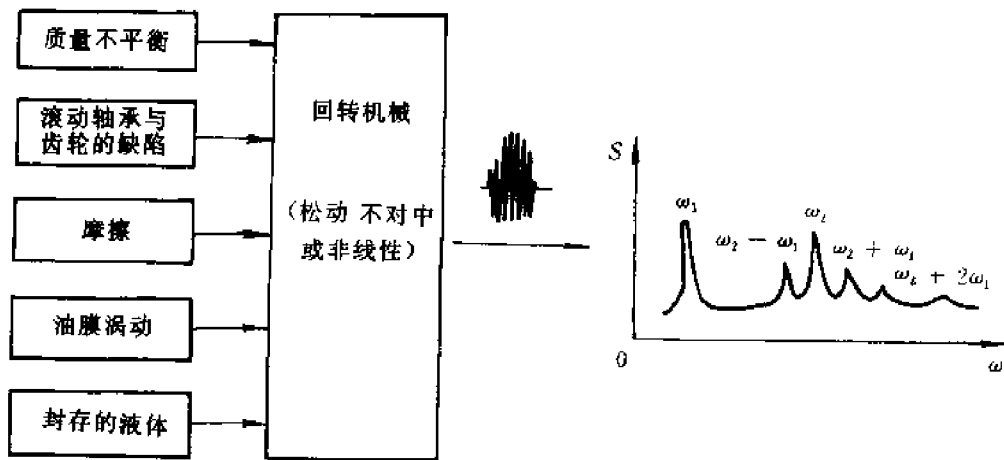


图 5-7 利用频率的和差规律诊断回转机械的故障

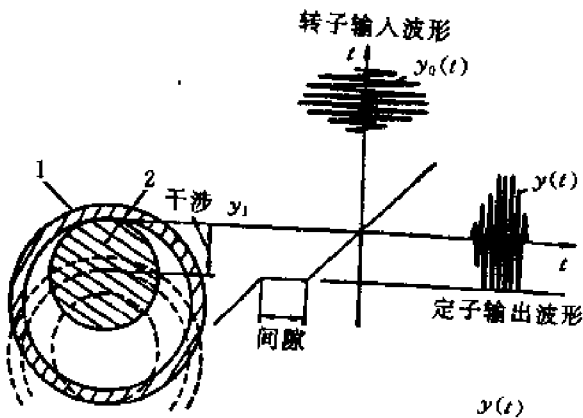


图 5-8 拍削平现象的一种解释图

1-定子轴承,2-转子轴颈;

实线-轴承与轴颈一起振动;虚线-轴承与轴颈分开

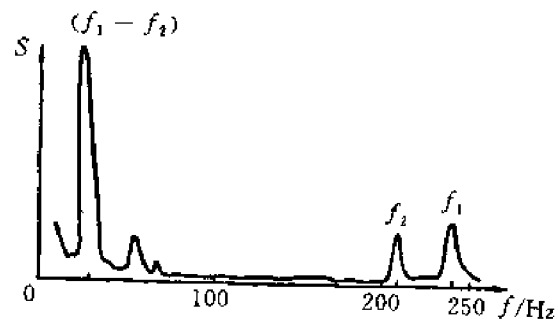


图 5-9 燃气轮机振动谱的频率和差规律

作为频率和差规律的另一个例子是一台 225 kW 电动机的振动频谱,当承受轴向载荷的滚动轴承产生缺陷时,谱图上出现一系列和频和差频。随着滚道上缺陷的不断增大,由于单边削平的深度增加,和频和差频的数目也随之增加。图 5-10 所示是内圈滚道圆周上,有整圈剥落的情况下出现的一群谱峰。这些谱峰所在的中心频率都是钢球通过内滚道的频率($f_i = 148$ Hz)与转轴组件回转频率($f_1 = 29.6$ Hz)之和或差。这两种频率是分别由于转轴组件质量失衡和内滚道缺陷所引起的。

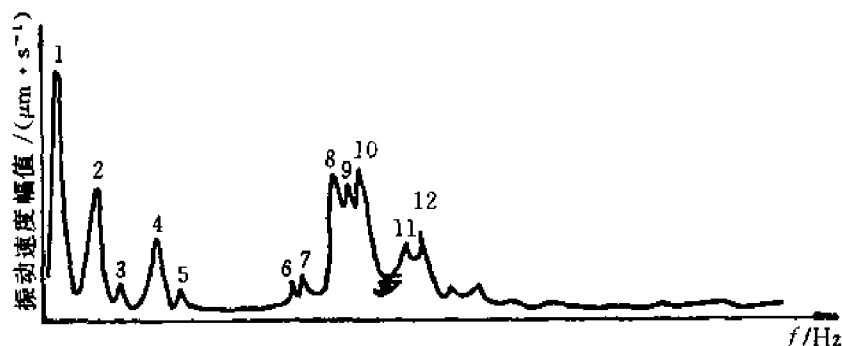


图 5-10 一台 225 kW 电动机水平方向振动谱

轴承 SKF6313; f_1 — 转轴组件的转速, 29.6 Hz; f_o — 钢球通过外圈滚道频率, 92 Hz;

f_i — 钢球通过内圈滚道频率, 148 Hz

1 — 32 Hz(f_1); 2 — 112 Hz($f_i - f_1$); 3 — 184 Hz($2f_o$); 4 — 256 Hz($2f_i - f_1$); 5 — 320 Hz($2f_1 + f_i$);

6 — 568 Hz($4f_1 - f_i$); 7 — 500 Hz($4f_o$); 8 — 688 Hz($5f_i - 2f_1$); 9 — 712 Hz($5f_1 - f_i$);

10 — 744 Hz($5f_o$); 11 — 856 Hz($6f_i - f_1$); 12 — 888 Hz($6f_1$)

上面所讨论的 5 种情况都是以频域分析为基础,亦即是以激振频率来查找激振原因的,但频域分析必须和时域识别密切结合,有些特征能够在时域信息中表现出来,但却不能在频谱图中明显地反映出来。如图 5-11 所示,图 5-11(a) 表示一台汽轮发电机组中电机振动的时域曲线,图 5-11(c) 表示由于轴弯曲或不对中产生的另一种振动的时域曲线,即单边削平波形。图 5-11(b), (d) 分别为图 5-11(a), (c) 的频域谱图。这两种情况下,在谱图上都是有一个基频分量和一些高度逐渐降低的高次谐波分量,但在时域信号中却表现了很大的差异。在图 5-11(a) 中仅仅是基频分量和二次谐波分量的叠加,而在图 5-11(c) 中组件可能存在非线性振动。由此可见,频域和时域分析结合是十分重要的。

(6) 轴心轨迹: 转子在轴承中高速旋转时不只围绕自身中心旋转,还环绕某一中心作涡动运动。产生涡动运动的原因可能是转子不平衡、对中不良、动静摩擦等,这种涡动运动的轨迹称为轴心轨迹。轴心轨迹的获取一般采用两个互成 90° 安置的涡流式位移传感器,在各自的方向上测量转轴组件相对机座的振动。图 5-12 所示是一台燃气轮机排气轴的轴线运动轨迹。图 5-12(a) 是水平和垂直两个方向上轴颈的振动频谱,由谱图上可以看到两个方向上的振动都包含有回转频率的高次谐波,并且振动的水平分量比垂直分量大。图 5-12(b) 是两个方向的合成,即轴心轨迹的图示。通过分析轴心轨迹的运动方向和转轴的旋转方向,可以确定转轴的进动方向(正进动或逆进动)。图 5-12(c) 是两个方向振动的时域信号。

2. 暂态频域分析

将机械系统的启停过程称为暂态过程。由于转轴组件从启动、升速到达额定转速的过程经

历了全部各种转速,在各个转速下的振动状态可以用来对临界转速、固有频率、阻尼系数各个参数辨识之用,因此启动和停车过程包含了丰富的信息,是常规运行状态下所无法获得的。对暂态过程进行频域分析常采用以下3种图形分析法。

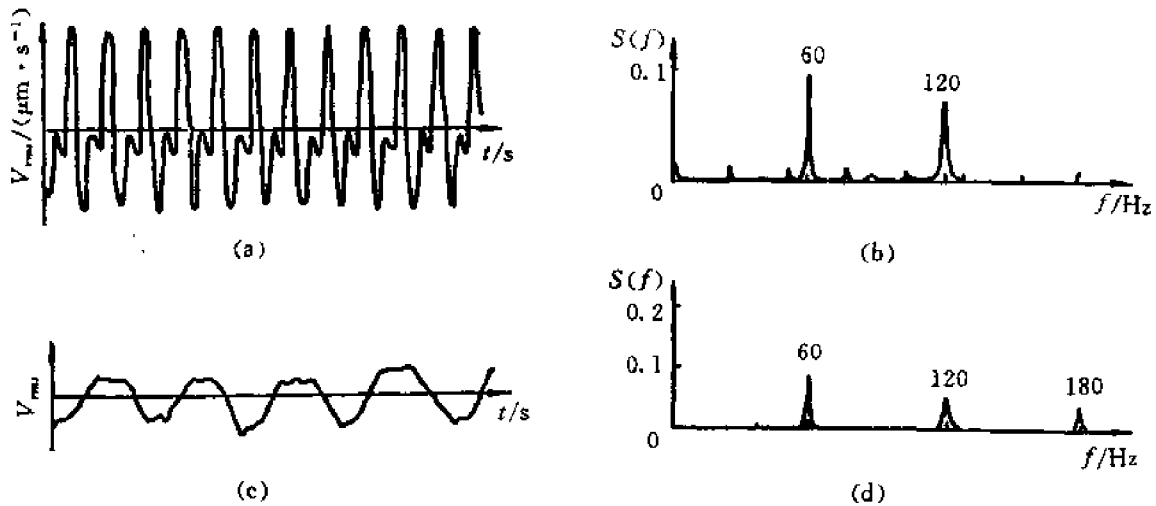


图 5-11 转轴组件不对中引起的振动

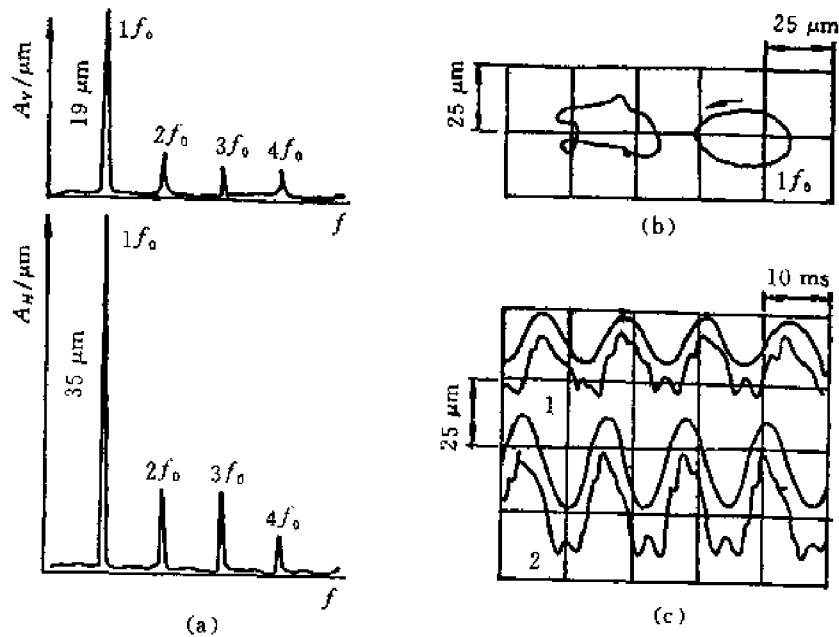


图 5-12 燃气轮机转子运转时的轴心轨迹

A_v -振幅垂直分量(μm); A_H -振幅水平分量(μm); 1-垂直分量; 2-水平分量

(1) 波德(Bode)图:波德图是机器振幅与频率、相位与频率的关系曲线,如图 5-13 所示。图中的横坐标是转轴组件的转速,一个纵坐标是振动的幅值,另一纵坐标是振动的相位。波德图上的幅值是将振动信号经过同步跟踪数字式向量滤波器过滤得到的,只包含有与转速相同的一个基频分量,其他高次谐波均已滤除。由于基频分量主要是由转子失衡引起的,因此波德图有时也称为失衡响应图。由图 5-13 所示的波德图上可以得到如下的信息:

一是转子系统在各个转速下的振幅和相位;

二是通过振幅峰值和相位偏移时的转速,可以判断共振频率(或临界转速值)为3 750 r/min;

三是在190°处有一个12.7 μm大小的慢滚动向量;

四是反映系统阻尼的放大系数约等于4(共振峰处的幅值100 μm除以共振后的幅值25 μm);

五是测量共振时振幅峰值的高度可以确定系统的阻尼。

(2) 极坐标图: 极坐标图又称为奈魁斯特(Nyquist)图,它是把上述幅频特性曲线和相频特性曲线综合在极坐标上表示出来,即它是在转轴组件启动过程中,当转速增加时,将不同转速下的幅值和相位作在极坐标平面上所连成的曲线,如图5-14所示。图上各点的极半径表示振幅值,角度表示相位。由图5-14所示的奈魁斯特图上,可以得到如下的信息:

- 一是转轴自零转速到运行转速整个范围内对失衡力的响应;
- 二是共振频率(或临界转速)为3 750 r/min;
- 三是在165°处有一个12.7 μm大小的慢滚动向量;
- 四是在共振时有180°相位偏移;
- 五是共振过程中相位角滞后(与转轴转动方向相反移动);
- 六是在1 500 r/min和3 000 r/min处结构有亚共振区。

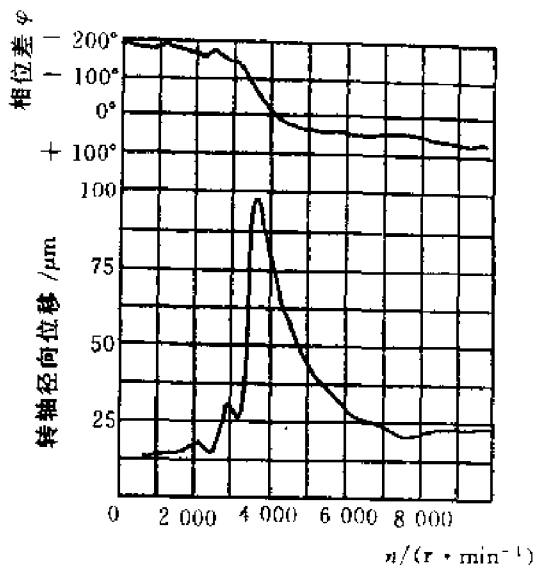


图5-13 转轴组件暂态过程的波德图

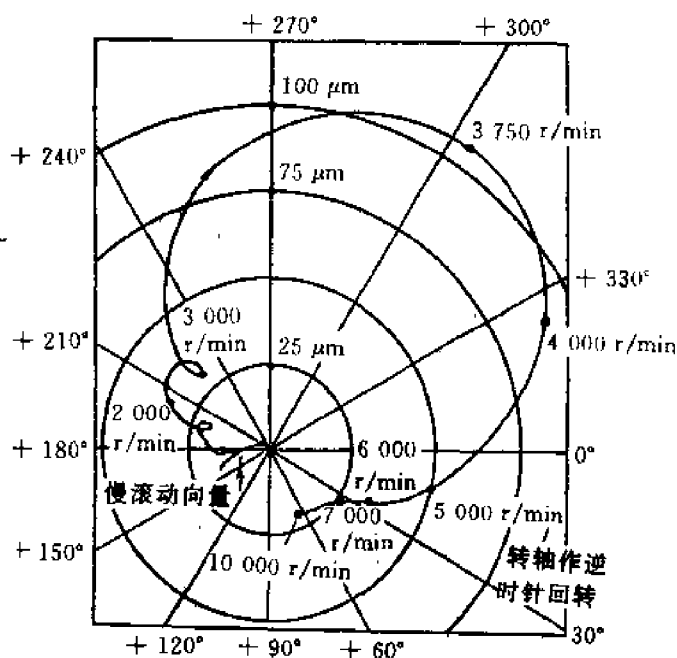


图5-14 转轴组件暂态过程的奈魁斯特图
图上箭头表示极坐标速度增长的方向

比较两种图示的方法,一般来说,奈魁斯特图比较直观地反映出轴心的暂态位置,使用也比较方便。例如,在共振速度区能清楚地表示出转子的响应和 180° 相位翻转;由子轴弯曲产生的慢滚动向量可以简单地用图解法从任意向量中减去;可以从内回路中找到结构共振区以及它们的相位关系。因此,奈魁斯特图是较为常用的方法。

(3) 瀑布图:除了上述两种图示方法外,还有一种所谓瀑布图示法,如图 5-15 和 5-16 所示,也是描述暂态过程的工具。它实质上是在启动或停车过程中,在不同转速下振动的功率谱图的迭置,因此纵坐标是机器的转速和幅值,转速自零到额定转速,横坐标是频率。图 5-15 是在不同转速下作出的多个功率谱迭置而成的三维瀑布图。

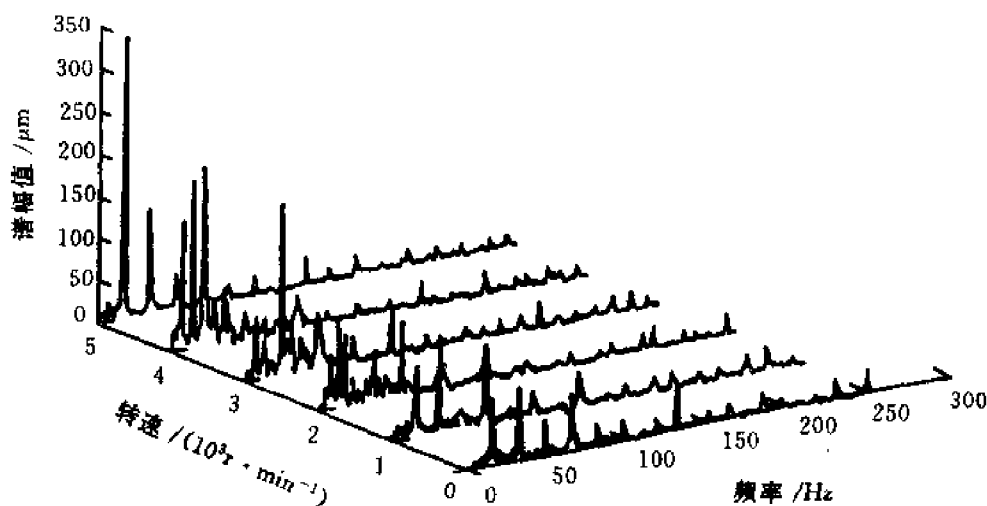


图 5-15 由机器在不同转速下的功率谱迭置成的瀑布图

由瀑布图(见图 5-16)可以看到回转频率 $n(r/min)$ 及其各次谐波下谱峰的高度,从而判断机器的临界转速、阻尼比和振源。有时在瀑布图上还会看到频率稳定,不随转速改变的谱峰,一般这类谱峰可能是转子上零件的自振频率(如叶片)。在图 5-16 中,临界转速约在 $4\,000\text{ r/min}$ 附近;由于高次谐波分量很小,主要是回转频率处的谱峰,可以判断转子的失衡比较严重。此外,在瀑布图上还有一个频率为 60 Hz 的谱峰,位置稳定,不随机器转速的升高而改变,这主要是电磁脉动的影响。

3. 趋势分析

趋势分析是把所测得的特征数据值和预报值按一定的时间顺序排列起来进行分析。这些特征数据可以是通频振动、 $1\times$ 振幅、 $2\times$ 振幅、 $0.5\times$ 振幅、轴心位置等,时间顺序可以按前后各次采样,按小时采样,按天采样等,趋势分析在故障诊断中起着重要的作用。图 5-17 为 $1\times$ 振动趋势示意图。

5.3 旋转机械典型故障的机理和特征

5.3.1 转子不平衡

不平衡是旋转机械最常见的故障。引起转子不平衡的原因有:结构设计不合理,制造和安

装误差,材质不均匀,受热不均匀,运行中转子的腐蚀、磨损、结垢、零部件的松动和脱落等。转子不平衡故障包括:转子质量不平衡、转子初始弯曲、转子热态不平衡、转子部件脱落、转子部件结垢、联轴器不平衡等,不同原因引起的转子不平衡故障规律相近,但也各有特点。

1. 转子质量不平衡

所有不平衡都可归结为转子的质量偏心,为此,首先分析带有偏心质量的转子的振动情况。转子的质量不平衡所产生的离心力始终作用在转子上,它相对于转子是静止的,其振动频率就是转子的转速频率,也称为工频(即工作频率),在频谱分析时,首先要找的就是工频分量。转子不平衡的振动特征有:

(1) 刚性转子:由于转子质量不平衡产生的离心力 $F = mr\omega^2$ (r 为偏心质量) 与转速的平方成正比,即振幅与转速的平方成正比,振动频率与转子转速一致,所以在轴承座测得的振动随转速增大而增大,但不一定与转速平方成正比,这是由于轴承与转子之间的非线性所致。图 5-18 为一台机床主轴振幅随转速变化的曲线。

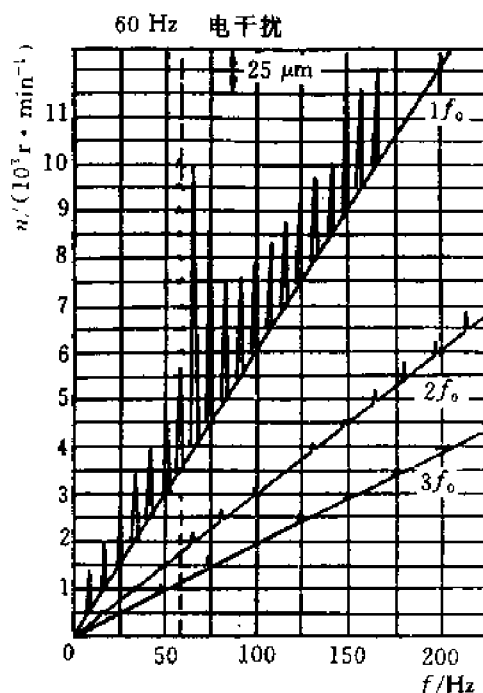


图 5-16 瀑布图分析

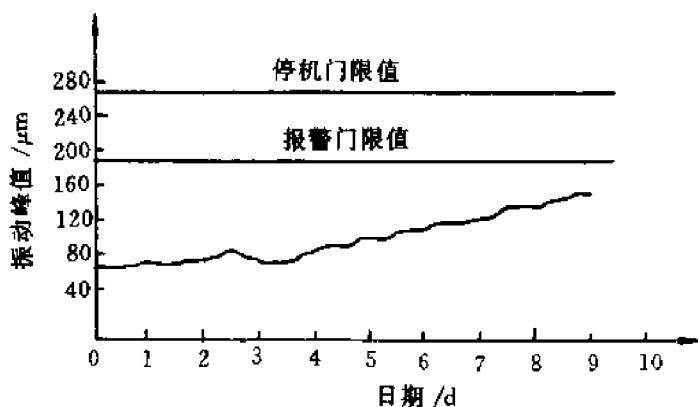


图 5-17 通频振动峰—峰值趋势图

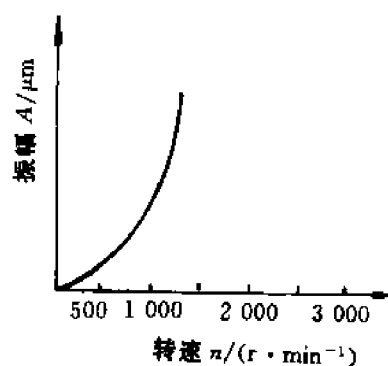


图 5-18 刚性转子不平衡的主要特征

(2) 柔性转子。转子稳态振动是一个与转速同频的强迫振动,振动幅值随转速按振动理论中的共振曲线规律变化,转速频率的高次谐波幅值很低,在临界转速处达到最大值,图 5-19(a) 所示;时域波形接近于一个正弦函数,图 5-19(b) 所示。因此转子不平衡故障的突出表现为一倍频振动幅值大。

表示转子各横截面中心位移的复数向量相角是不同的,因此轴线弯曲成空间曲线,并以转子转速绕 oz 轴转动,如图 5-20 所示。

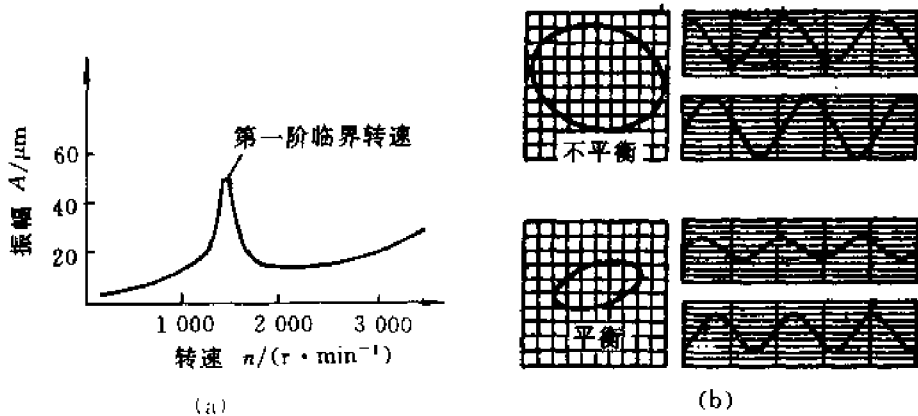


图 5-19 柔性转子不平衡的主要特征

实际上,由于轴承在不同的方向上刚度不相等,油膜阻尼的非线性以及转子的非线性等因素的影响,使轴承在不同方向上的振动大小并不一样,通常是水平方向刚度较小,振动幅值较大,使轴心轨迹成为椭圆形,并且会出现较小的高次谐波,使整个频谱呈所谓的" 枞树形",如图 5-21 所示。

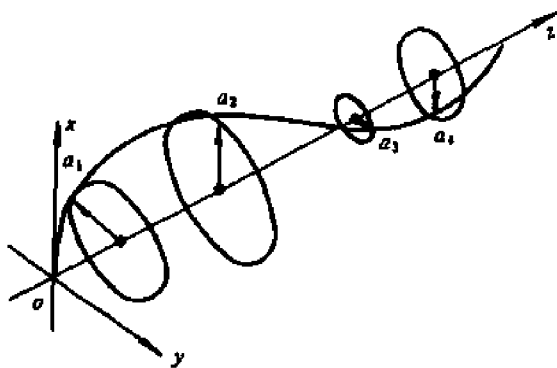


图 5-20 转子轴线形状示意图

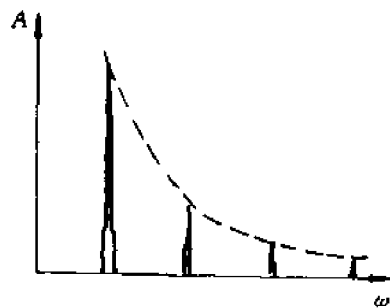


图 5-21 转子不平衡故障谱图

2. 转子初始弯曲

人们习惯上将转子的初始弯曲与质量初始不平衡同等看待,实际上是有区别。所谓质量不平衡是指各横截面的质心连线与其几何中心连线存在偏差,而转子弯曲是指各横截面的几何中心连线与旋转轴线不重合,二者都会使转子产生偏心质量,从而使转子产生不平衡振动。

初弯转子具有与质量不平衡转子相似的振动特征,所不同的是初弯转子在转速较低时振动较明显,趋于初弯值,可通过检查转子各部位的径向跳动量予以判断。在汽轮发电机组中,通常是在盘车时和盘车后测量晃动度的大小来判断转子是否存在初始弯曲。

3. 转子热态不平衡

在机组的启动和停机过程中,由于热交换速度的差异,使转子横截面产生不均匀的温度分布,使转子发生瞬时热弯曲,产生较大的不平衡。热弯曲引起的振动一般与负荷有关,改变负荷,振动相应地发生变化,但在时间上较负荷的变化滞后。随着盘车或机组的稳态运行,整机温

度趋于均匀,振动会逐渐减小。

4. 转子部件脱落

运行中的转子部件突然脱落也会引起转子不平衡,使转子振幅突然发生变化,严重影响机组的正常运行。为了防止脱落部件在惯性力作用下飞出使机体发生二次事故,必要时应及时停机检修。

可以将部件脱落失衡现象看作对工作状态的转子的瞬时阶跃响应。由于瞬态响应最终要衰减为零,因此,部件脱落的主要特征是振动会突然发生变化而后趋于稳定,振动的幅值一般会有较明显的增大。

5. 转子部件结垢

如果工件的质量不合格,随着时间的推移,将在转子的动叶和静叶表面产生尘垢,使转子原有的平衡遭到破坏,振动增大。由于结垢需要相当长的时间,所以振动是随着年月逐渐增大的。并且由于通流条件变差,轴向推力增加,轴向位移增大,机组级间压力逐渐增大,效率逐渐下降。

6. 联轴器不平衡

由于制造、安装的偏差或者动平衡时未考虑联轴器的影响,可能使联轴器产生不平衡。联轴器不平衡具有与质量不平衡相似的振动特征,通常是联轴器两端轴承的振动较大,相位基本相同。

以上各种不平衡故障的振动特征与质量不平衡基本相同。

5.3.2 转子不对中

转子不对中通常是指相邻两转子的轴心线与轴承中心线的倾斜或偏移程度。转子不对中可分为联轴器不对中和轴承不对中,联轴器不对中又可分为平行不对中、偏角不对中和平行偏角不对中三种情况。

1. 联轴器不对中

(1) 平行不对中:当转子轴线之间存在径向位移时,联轴器的中间齿套与半联轴器组成移动副,不能相对转动,但中间齿套却与半联轴器产生滑动而作平面圆周运动,即中间齿套的中心是沿着以径向位移 y 为直径作圆周运动,如图 5-22 所示。设 A 为主动转子的轴心投影, B 为从动转子的轴心投影, K 为中间齿套的轴心, AK 为中间齿套与主动轴的连线, BK 为中间齿套与从动轴的连线, AK 垂直 BK , 如图 5-23 所示, 设 AB 长为 D , K 点坐标为 $K(x, y)$, 取 θ 为自变量, 则有

$$\begin{aligned}x &= D\sin\theta\cos\theta = \frac{1}{2}D\sin(2\theta) \\y &= D\cos\theta\cos\theta - \frac{1}{2}D = \frac{1}{2}D\cos(2\theta)\end{aligned}\quad (5.3)$$

对 θ 求导, 得

$$dx = D\cos(2\theta)d\theta, \quad dy = -D\sin(2\theta)d\theta$$

K 点的线速度为

$$V_k = \sqrt{(dx/dt)^2 + (dy/dt)^2} = Dd\theta/dt \quad (5.4)$$

由于中间齿套平面运动的角速度 $(d\theta/dt)$ 等于转轴的角速度, 即 $d\theta/dt = \omega$, 所以 K 点绕圆

周中心运动的角速度 ω_k 为

$$\omega_k = 2V_k/D = 2\omega \quad (5.5)$$

式中, V_k 为点 K 的线速度。由式(5.5)可知, K 点的转动为转子角速度的两倍, 因此当转子高速运转时, 就会产生很大的离心力, 激励转子产生径向振动, 其振动频率为转子工频的两倍。此外由于不对中而引起的振动有时还包含有大量的谐波分量, 但其关键的是一个很大的 2 倍频分量。



图 5-22 联轴器平行不对中

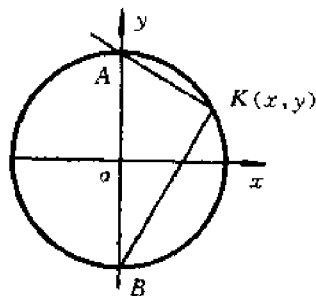


图 5-23 联轴器齿套运动分析

(2) 偏角不对中: 当转子轴线之间存在偏角位移时, 如图 5-24 所示, 从动转子与主动转子的角速度是不同的。从动转子的角速度为

$$\omega_2 = \omega_1 \cos\alpha / (1 - \sin^2\alpha \cos^2\varphi_1) \quad (5.6)$$

式中, ω_1, ω_2 分别为主动转子和从动转子的角速度, α 为从动转子的偏斜角, φ_1 为主动转子的转角。

从动转子每转动一周其转速变化两次, 如图 5-25 所示, 变化范围为

$$\omega_1 \cos\alpha \leq \omega_2 \leq \omega_1 / \cos\alpha \quad (5.7)$$

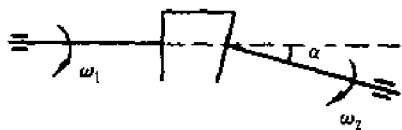


图 5-24 联轴器偏角不对中图

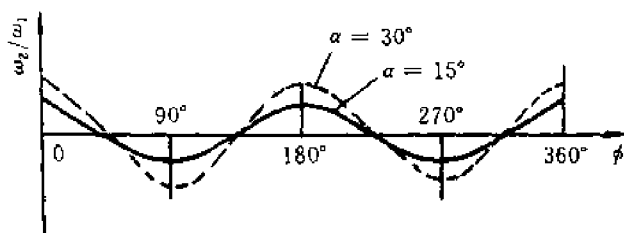


图 5-25 转速比的变化曲线

偏角不对中使联轴器附加一个弯矩, 弯矩的作用是力图减小两轴中心线的偏角。轴旋转一周, 弯矩作用方向交变一次, 因此, 偏角不对中增加了转子的轴向力, 使转子在轴向产生工频振动。

(3) 平行偏角不对中: 实际上, 各转子轴线之间往往既有径向位移又有偏角位移, 因此当转子运转时, 就有一个 2 倍频的附加径向力作用于靠近联轴器的轴承上, 有一个同频的附加轴向力作用于止推轴承上, 从而激励转子发生径向和轴向振动。

2. 轴承不对中

轴承不对中实际上反映的是轴承座标高和左右位置的偏差。由于结构上的原因, 轴承在水

平方向和垂直方向上具有不同的刚度和阻尼,不对中的存在加大了这种差别。虽然油膜既有弹性又有阻尼,能够在一定程度上弥补不对中的影响,但当不对中过大时,会使轴承的工作条件改变,使转子产生附加的力和力矩,甚至使转子失稳和产生碰摩。

轴承不对中使轴颈中心的平衡位置发生变化,使轴系的载荷重新分配。负荷大的轴承油膜呈现非线性,在一定条件下出现高次谐波振动,负荷较轻的轴承易引起油膜涡动进而导致油膜振荡。支承负荷的变化还使轴系的临界转速和振型发生改变。

3. 不对中故障的特征

(1) 转子径向振动出现 2 倍频,以 1 倍频和 2 倍频分量为主,不对中越严重,2 倍频所占比例越大;1 倍频和 2 倍频分量的比例通常可以用来判断问题的严重程度;

(2) 相邻两轴承的油膜压力反方向变化,一个油膜压力变大,另一个则变小;

(3) 典型的轴心轨迹为香蕉形,正进动;

(4) 联轴器不对中时轴向振动较大,振动频率为 1 倍频,振动幅值和相位稳定;

(5) 轴承不对中时径向振动较大,有可能出现高次谐波,振动不稳定;

(6) 振动对负荷变化敏感。当负荷改变时,由联轴器传递的扭矩立即发生改变,振动值随负荷的增大而增高;如果联轴器不对中,则转子的振动状态也立即发生变化。由于温度分布的变化,轴承座的热膨胀不均匀而引起轴承不对中,使转子的振动也要发生变化。但由于热传导的惯性,振动的变化在时间上要比负荷的改变滞后一段时间。

5.3.3 转子碰摩

随着机组参数的不断提高,动静间隙的不断缩小,以及运行过程中不平衡、不对中、热弯曲等的影响,经常发生转子碰摩故障。在国产 20 万千瓦汽轮发电机组中,已有多台因动静碰摩而造成转子弯曲的严重事故。根据摩擦部位不同,碰摩分两种情况,转子外缘与静止件接触而引起的摩擦,称为径向碰摩;转子在轴向与静止件接触而引起的摩擦,称为轴向碰摩。从不同的角度,摩擦还可分为局部摩擦和全周摩擦;早期、中期和晚期碰摩等。

1. 摩擦振动对转子的影响

转子碰摩是一复杂的过程,从机理上分析,摩擦振动对转子有以下 4 方面的影响:

(1) 直接影响:转子运动可分为自转和进动两种形式。摩擦对自转的影响在于附加了一个力矩,因此,在转子原有力矩不变的条件下有可能使转子转速发生波动。至于进动,由于摩擦力的干预可能使正进动转化为反进动,特别是全周摩擦,常常产生所谓的“干摩擦”现象,从而引起自激振动,影响转子正常运行,甚至损坏机组。

(2) 间接影响:摩擦的作用使动静部件相互抵触,相当于增加了转子的支承条件,增大了系统的刚度,改变了转子的临界转速及振型,且这种附加支承是不稳定的,从而可能引起不稳定振动及非线性振动。

(3) 冲击影响:局部碰摩除了摩擦作用外还会产生冲击作用,其直观效应是给转子施加了一个瞬态激振力,激发转子以固有频率作自由振动。虽然自由振动是衰减的,但由于碰摩在每个旋转周期内都产生冲击激励作用,在一定条件下有可能使转子振动成为叠加自由振动的复杂振动。

(4) 热变形:摩擦引起的热变形可能引起转子弯曲,加大偏心量,使振动增大。

转子碰摩的定量分析比较困难,一般来说,转子与静止件发生摩擦时,转子受到静止附加

作用力,它是非线性的和时变的,因此使转子产生非线性振动,在频谱图上表现出频谱成分丰富,不仅有工频,还有高次和低次谐波分量。当摩擦加剧时,这些谐波分量的增长很快。典型的碰摩故障的波形和频谱如图 5-26 所示。

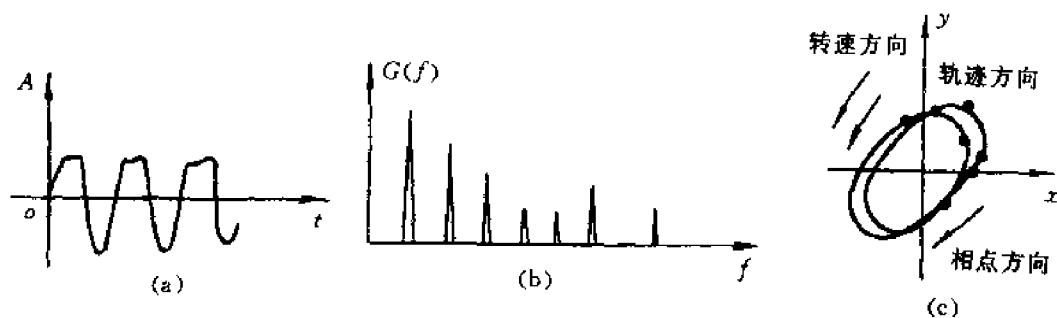


图 5-26 转子碰摩的波形和频谱图
(a) 波形 (b) 频谱 (c) 轴心轨迹

转子径向碰摩主要影响转子的径向振动,对转子的轴向振动影响较小。但当转子发生轴向碰摩时,除了对径向振动产生影响外,由于轴向力的存在,使轴向位移和轴向振动增大,有时还会使级间压力发生变化,造成机组效率的下降。

2. 转子碰摩故障的特征

- (1) 转子失稳前频谱丰富,波形畸变,轴心轨迹不规则变化,正进动;
- (2) 转子失稳后波形严重畸变或削波,轴心轨迹发散,反进动;
- (3) 轻微摩擦时同频幅值波动,轴心轨迹带有小圆环;
- (4) 碰摩严重时,各频率成分幅值迅速增大;
- (5) 系统的刚度增加,临界转速区展宽,各阶振动的相位发生变化;
- (6) 工作转速下发生的轻微摩擦振动,其振幅随时间缓慢变化,相位逆转动方向旋转。

5.3.4 油膜振荡

1. 半速涡动分析

当轴颈在轴瓦中转动时,在轴颈与轴瓦之间的间隙中形成油膜,油膜的流体动压力使轴颈具有承载能力。当油膜的承载力与外载荷平衡时,轴颈处于平衡位置;当转轴受到某种外来扰动时,轴颈中心就会在静平衡位置附近发生涡动,其振动频率约为转子回转频率的一半,因而,常称为半速涡动或半频涡动。半速涡动是一种自激振动,涡动幅值保持在一稳定值,一般幅值较小,但半速涡动可能演变为发散情况,是属于不稳定振动。

假设油在轴承中无端泄,轴瓦表面的油膜流动速度为零,而轴颈表面的油膜流动速度与转速为 ω 的轴颈表面线速度相同。因此在层流的假设下,油膜沿径向的速度分布如图 5-27 所示,在连心线上 AB 截面流入油楔的流量 $r\omega b(c+e)/2$ 与在 CD 处流出的流量 $r\omega b(c-e)/2$ 之差应等于因轴心涡动引起收敛楔隙内流体容积的增加率,即

$$\frac{1}{2}r\omega b(c+e) - \frac{1}{2}r\omega b(c-e) = 2rbe\omega_s \quad (5.8)$$

由此得

$$\omega_s = \frac{1}{2}\omega$$

式中, r 为轴颈半径; b 为轴承宽度; c 为轴承间隙; e 为轴心偏心距; ω 为轴颈转动角速度; ω_w 为轴颈涡动角速度。

这就是所谓半速涡动的含义。实际上, 由于轴承端泄等因素的影响, 一般涡动频率略小于转速的一半, 约为转速的 $0.42 \sim 0.48$ 倍。

2. 油膜振荡现象

转轴的转动在失稳转速以前是平稳的, 当达到失稳转速后即发生半速涡动。随着转速升高, 涡动角速度也将随之增加, 但总保持着约等于转动速度之半的比例关系, 半速涡动一般并不剧烈。当转轴转速升到比第一阶临界转速的 2 倍稍高以后, 由于这时半速涡动的涡动速度与转轴的第一阶临界转速相重合即产生共振, 表现为强烈的振动现象, 称为油膜振荡。油膜振荡一旦发生之后, 就将始终保持约等于转子一阶临界转速的涡动频率, 而不再随转速的升高而升高, 在图上出现一个“平台”, 如图 5-28 所示。

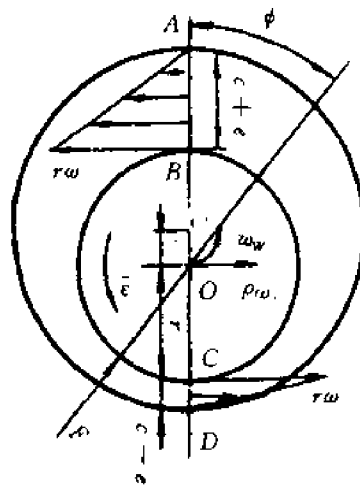


图 5-27 半速涡动的原理

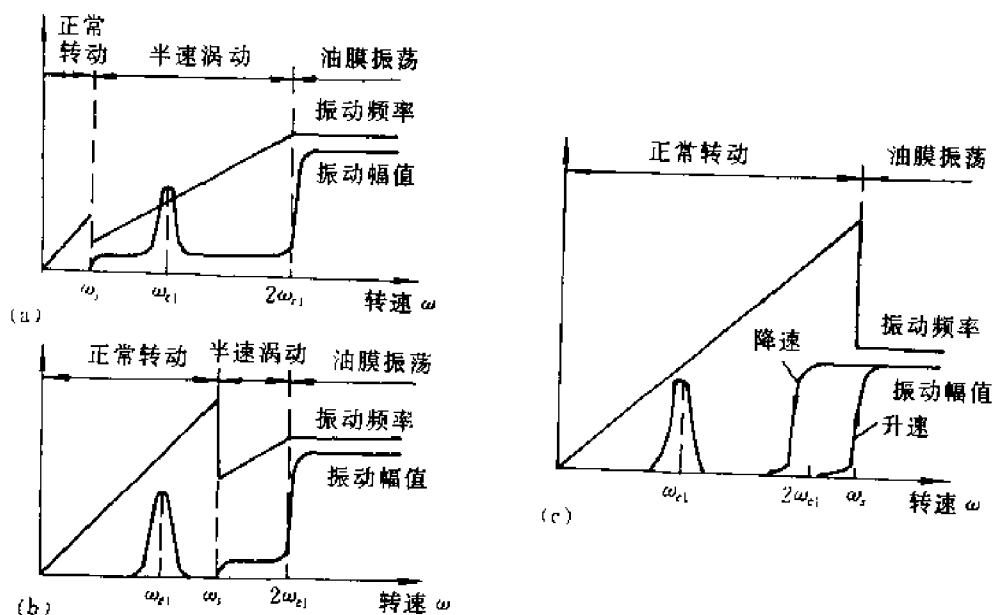


图 5-28 油膜振荡的转速特性

图 5-28 表示油膜振荡的转速特性, 分三种情况, 图中均表明了随转速 ω 变化的正常转动、半速涡动和油膜振荡的三个阶段, 其中一条曲线表示振动频率的变化, 一条曲线表示振动幅值的变化。图 5-28(a) 表示失稳转速在一阶临界转速之前; 图 5-28(b) 表示失稳转速在一阶临界转速之后, 这两种情形的油膜振荡都在稍高于二倍临界转速的某一转速时发生; 图 5-28(c) 表示失稳转速在 2 倍临界转速之后, 转速在稍高于 2 倍临界转速时, 转轴并没有失稳, 直到比 2 倍临界转速高出较多时, 转轴才失稳; 而降速时油膜振荡消失的转速要比升速时发生油膜振荡的转速低, 表现出油膜振荡的一种“惯性”现象。

2. 油膜振荡故障的特征

- (1) 油膜振荡总是发生在转速高于转子系统一阶临界转速的 2 倍以上；
- (2) 油膜振荡的频率接近转子的一阶临界转速，即转速再升高，其频率基本不变；
- (3) 油膜振荡时，转子的挠曲呈一阶振型；
- (4) 油膜振荡时，振动的波形发生畸变，在工频的基波上叠加了低频成分，有时低频分量占主导地位，低频振动的幅值，轴承座振动可达 $40\ \mu\text{m}$ 以上，轴振动可达 $100 \sim 150\ \mu\text{m}$ 以上，且振幅不稳，轴心轨迹发散

(5) 油膜振荡时，转子涡动方向与转子转动方向相同，轴心轨迹呈花瓣形，正进动，如图 5-29 所示；

(6) 油膜振荡的发生和消失具有突然性，并具有惯性效应，即升速时产生振荡的转速比降速时振荡消失的转速要大；

(7) 油膜振荡剧烈时，随着油膜的破坏，振荡停止，油膜恢复后，振荡再次发生，这样持续下去，轴颈与轴承不断碰摩，产生撞击声，轴瓦内油膜压力有较大波动；

(8) 油膜振荡对转速和油温的变化较敏感，一般当机组发生油膜振荡时，随着转速的增加，振动不下降，随着转速的降低，振动也不立即消失，称为滞后现象；提高进油温度，振动一般有所降低；

(9) 轴承载荷越小或偏心率 $\epsilon = e/c$ 越小，越易发生油膜振荡。

4. 防止油膜振荡的措施

为了预防和消除油膜振荡，可以根据转子系统的实际情况采取以下若干措施：

(1) 消除油膜振荡的诱发因素：① 改善转子的平衡状态，限制振幅放大因子；② 消除转子不对中故障，限制低次谐波分量；③ 消除动静间隙不均匀，限制非线性激振力。

(2) 改变轴承参数：① 提高轴承比压，减小轴承宽度，抬高轴承标高，在下瓦中部开环形槽等，但不超过轴承允许的最大承载能力；② 降低润滑油粘度，将粘度较高的油换成粘度较低的油，提高进油温度，以不发生干摩擦、油质劣化为限；

(3) 选择合适的轴承形式：根据轴承类型和结构尺寸的不同，每种轴承有其稳定工作的范围。一般认为各种轴承的稳定性从优到劣可按下列次序排列：可倾瓦、偏置三油叶、对称三油叶、椭圆、三油楔、圆柱轴承。

(4) 增加转子系统刚度，提高转子系统的临界转速。转子固有频率越高，发生油膜振荡的失稳转速也越高，系统失稳转速应在工作转速的 125% 以上。

5.3.5 其他常见典型故障

1. 转轴裂纹

导致转轴裂纹最重要的原因是高周期疲劳、低周期疲劳、蠕变和应力腐蚀开裂，此外也与转子工作环境中含有腐蚀性化学物质等有关，而大的扭转和径向载荷，加上复杂的转子运动，造成了恶劣的机械应力状态，最终也将导致轴裂纹的产生。

裂纹在转子旋转的动态应力下，始终处于“开”和“闭”的周期变化过程中。定性表示裂纹转轴的挠度变化如图 5-30 所示。

裂纹轴响应中除 $1X$ 分量外，还有 $2X, 3X, 5X$ 等阶谐波分量，利用转子升速通过 $\omega_1/2$ ，

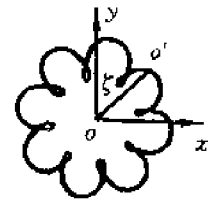


图 5-29 油膜振荡时的涡动轨迹

$\omega_1/3$ 转速时相应的 2 倍频、3 倍频成分被共振放大的所谓超谐波共振现象,也可监测轴裂纹。图 5-31 为某裂纹转子的升速共振频谱图,从图中可以看出它包含有 $\omega_1/2, \omega_1/3$ 临界转速分量。一般在低于临界转速运行时,所观测到的高阶成分较明显,而在高于临界转速状态下运行时,高阶成分不明显。

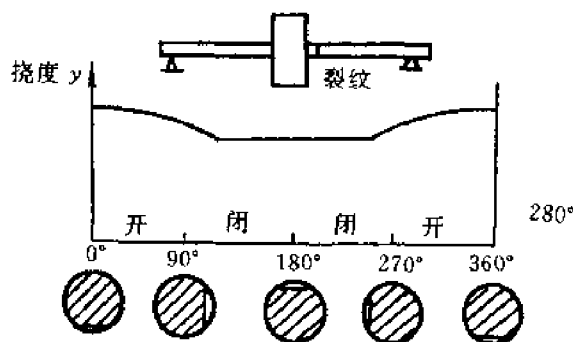


图 5-30 裂纹转轴的挠度变化

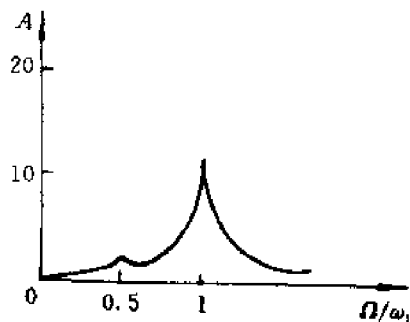


图 5-31 某裂纹转子升速共振频谱图

此外,裂纹转子的动平衡会遇到反复无常的变化,这是由于裂纹转子的非线性特性。

裂纹转子的监测和诊断方法有:

(1) 稳态响应法:对裂纹转子的监测和诊断要着眼于各阶谐波分量幅值 $1X, 2X$ 和 $3X$ 的大小以及随时间的变化。 $1X, 2X$ 和 $3X$ 分量幅值随时间稳定增长的趋势表明转子可能存在裂纹。

(2) 滑停法:此法将机组从工作转速滑降至零转速,在降速过程中测量振动响应并进行谱分析,若转子产生裂纹或裂纹有进一步的扩展,则在转速过临界及 $1/2, 1/3$ 临界转速时,振动响应将有明显的改变。

(3) 温度瞬间法:此法原理是快速降低蒸汽温度,使转子表面产生拉伸的热应力,如果有裂纹存在,拉应力将使裂纹张开,使转子振动瞬间增大。通过快速降温或快速升温的办法可以发现转子是否有裂纹。

裂纹故障的特征为:

- (1) 各阶临界转速较正常时要小,尤其在裂纹严重时;
- (2) 由于裂纹造成刚度变化且不对称,转子的共振转速扩展为一个区;
- (3) 裂纹转子轴系在强迫响应时,一阶分量的分散度比无裂纹时大;
- (4) 转速超过临界转速后,一般各高阶谐波振幅较未超过时小;
- (5) 恒定转速下,各阶谐波幅值 $1X, 2X$ 和 $3X$ 及其相位不稳定,且尤以 $2X$ 突出;
- (6) 裂纹引起刚度不对称,使转子动平衡发生困难,往往多次试重也达不到所要求的平衡精度。

2. 旋转失速

旋转失速是压缩机最常见的一种不稳定现象。当压缩机流量减少时,由于冲角增大,叶栅背面将发生流体分离,流道将部分或全部被堵塞。这样失速区会以某速度向叶栅运动的反方向传播。实验表明,失速区传播的相对速度低于叶栅转动的绝对速度。因此观察到的失速区沿转子的转动方向移动,故称分离区,这种相对叶栅的旋转运动为旋转失速。旋转失速使压缩机中

的流动情况恶化,压比下降,流量及压力随时间波动。在一定转速下,当入口流量减少到某一值 Q_{min} 时,机组会产生强烈的旋转失速,强烈的旋转失速会进一步引起整个压缩机组系统的一种危险性更大的不稳定的气动现象,即喘振。此外,旋转失速时压缩机叶片受到一种周期性的激振力,如旋转失速的频率与叶片的固有频率相吻合,则将引起强烈振动,使叶片疲劳损坏造成事故。

旋转失速故障的识别特征:

- (1) 旋转失速发生在压气机上;
- (2) 振动幅值随出口压力的增加而增加;
- (3) 振动发生在流量减小时,且随着流量的减小而增大;
- (4) 振动频率与工频之比为小于1的常值;
- (5) 转子的轴向振动对转速和流量十分敏感;
- (6) 一般排气端的振动较大;
- (7) 排气压力有波动现象;
- (8) 机组的压比有所下降,严重时压比突降。

3. 喘振

旋转失速严重时可以导致喘振,但二者并不是一回事。喘振除了与压缩机内部的气体流动情况有关之外,还同与之相连的管道网络系统的工作特性有密切的联系。

压缩机总是和管网联合工作的,为了保证一定的流量通过管网,必须维持一定压力,用来克服管网的阻力。机组正常工作时的出口压力是与管网阻力相平衡的。但当压缩机的流量减少到某一值 Q_{min} 时,出口压力会很快下降,然而由于惯性作用,管网中的压力并不马上降低,于是,管网中的气体压力反而大于压缩机的出口压力,因此,管网中的气体就倒流回压缩机,一直到管网中的压力下降到低于压缩机出口压力为止。这时,压缩机又开始向管网供气,压缩机的流量增大,恢复到正常的工作状态。但当管网中的压力又回到原来的压力时,压缩机的流量又减少,系统中的流体又倒流。如此周而复始产生了气体强烈的低频脉动现象——喘振。

喘振故障的识别特征:

- (1) 诊断对象为压气机组或其他带长导管、容器的流体动力机械;
- (2) 振动发生时,机组的入口流量小于相应转速下的最小流量;
- (3) 振动的频率一般在 $0 \sim 10$ Hz之内;
- (4) 机组及与之相连的管道都发生强烈振动;
- (5) 有倒流现象;
- (6) 出口压力(压力表)呈大幅度的波动;
- (7) 机组的功率(表指针)呈周期性的变化;
- (8) 振动前有失速现象;
- (9) 振动时有周期性的吼叫声;
- (10) 机组的工作点在喘振区(或附近)。

4. 迷宫密封的气流激振

气体在迷宫中流动是一种复杂的三维流动。当转子因挠曲、偏磨、安装偏心或旋转产生涡动运动时,密封腔内周向的间隙不均匀,即使密封腔内入口处的压力周向分布是均匀的,在该腔的出口处却形成了不均匀的周向压力分布,形成了一个作用于转子上的合力,此力在与转子

偏心位移相垂直方向上的切向分力相互作用,就将激励转子作进一步的涡动,成为转子一个不稳定的激励力,可能导致转子失稳。失稳时的频率因不同的气体状态及迷宫几何形状而不相同。

迷宫密封中的流体力激振所引起的机器振动频率,往往表现为低于工作转速的亚异步振动。许多机器的振动还与机组的负荷与转速有关,在操作时存在一个与转速、负荷等因素密切相关的“阈值”,当机器运行到这个值时,只要很小的转速或负荷的变化,就可能导致机器强烈振动,使原来运行稳定的转子运行不稳定,或是机器在低负荷下运行稳定,在高负荷下运行不稳定。

迷宫密封气流涡动故障特征:

- (1) 涡动频率一般为 0.6 ~ 0.9 倍工频;
- (2) 轴心轨迹呈椭圆形,正进动;
- (3) 强振时有可能激发转子的一阶自振频率,表现为自激振动;
- (4) 转速存在一个“阈值”,在其值附近可导致强烈振动;
- (5) 负荷也存在一个“阈值”,在其值附近可导致强烈振动;
- (6) 强振时的主频为转子的一阶固有频率,频带较宽;
- (7) 振动的再现性强;
- (8) 一般在转子不平衡、不对中、偏心时易发生。

5. 不均匀气流涡动

汽轮机、燃气轮机、压气机等转子都有叶片,除离心压气机外,气(汽)体在叶轮周围是轴向流动的,气流对叶片产生周向力。如果转子没有弯曲,则叶轮与固定内腔的径向间隙沿周向是相同的,因此气流沿周向是均匀分布的,它对叶轮各叶片的周向力相等,所有这些力的合力是一个推动或阻碍叶轮转动的力偶。如果轴发生了弯曲,则叶轮偏向内腔的一侧,径向间隙沿周向是不均匀分布的。图 5-32 表示汽轮机气流驱使叶轮转动,这时,气流加于叶轮上的周向力在间隙大的一边小于间隙小的一边,即 $F_{n1} > F_{n2}$ 。各叶片所受周向力的总和除了力偶外,还有与轮心 O' 的位移垂直的力: $F_t = F_{n1} - F_{n2}$,这个力使转子产生涡动,涡动的方向与转子运转的方向一致。涡动的频率约为 0.6 ~ 0.9 倍的转速。随着转速提高涡动频率接近系统的固有频率,且气流压力足够大时,这会发生振荡。这一失稳机理同油膜失稳是类似的。

不均匀气流涡动故障的主要识别特征:

- (1) 振动频率为 0.6 ~ 0.9 倍工频;
- (2) 转子有偏心弯曲造成的间隙不均;
- (3) 振动对气流压力、流量的改变非常敏感;
- (4) 负荷存在一个“阈值”,在其值附近可导致剧烈振动;
- (5) 在一个由多个转子组成的轴系中,气流涡动常发生在气流压力高的转子上,如在汽轮发电机组中,蒸汽振荡主要发生于高压转子。

6. 转子内壁吸附液体

在某些空心转子中,有时可能在转子内壁的局部吸附了油或水汽等冷凝后的液体,汽轮机大轴中心孔进油也属于这种情况。当转轴有弯曲变形时,这种液体的离心力也会使转子失稳,如图 5-33 所示。当转轴弯曲时,液体沿轴心位移 OO' 的方向被甩向转子的内壁,但此液体并不是停留在 OO' 的延长线上,而是因粘性被内壁粘带至延长线一侧,液体的重心位点与 OO' 的夹

角为 φ 的直线上。设液体的离心力为 F_c ，它可以分解为 F_{c1} 和 F_{c2} 两个分力， F_{c1} 与位移 OO' 垂直，促使转子运动失稳。经过分析可知，失稳角速度 ω ，高于临界角速度 ω_c ，但小于它的两倍，即 $\omega_c < \omega < 2\omega_c$ ，或者说，转子失稳角速度与转速之比为 $0.5 \sim 1.0$ ，涡动是正向的，其频率等于临界转速。

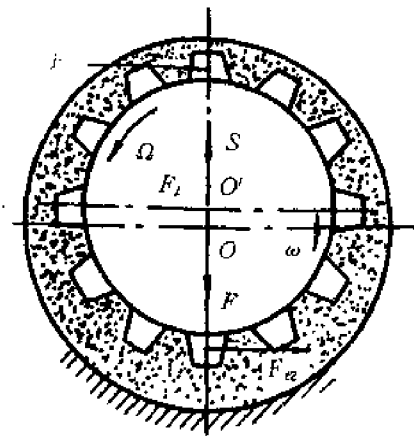


图 5-32 蒸汽不均转子受力图

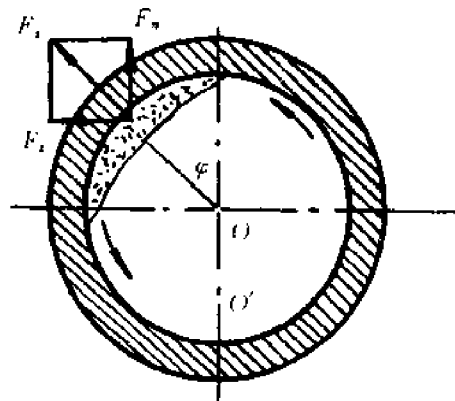


图 5-33 转子内壁吸附液体受力图

5.4 旋转机械振动故障诊断示例

5.4.1 H 型离心压缩机转子-定子碰摩振动故障

1. 机组简况

某厂环氧乙烷装置离心式空气压缩机由高、低压两缸组成。低压缸为 H 型三轴式压缩机，其高速末端又驱动一高压缸，其结构如图 5-34 所示，有关主要性能参数如表 5-6 所示。

表 5-6 离心压缩机主要性能参数

	低压缸	高压缸
转速 / (r · min ⁻¹)	8 383	13 987
功率分配 / kW	4 000	2 000

2. 机组启动时振动故障

正常情况下 LP_3 轴承处转子振动最大可达 $41 \sim 45 \mu\text{m}$ ，检修后再次启动机组时，振动达 $325 \mu\text{m}$ 以上，居高不下，无法投入运行。

3. 振动数据采集及信号处理

对振动信号的监测分析结果表明，振动状态有突变，即有跳跃现象发生，转子振动呈现两种状态：

(1) 不稳定同频振动状态。如图 5-35 所示，由频谱图可见工频振动占主导，有很小的 $1/2$ 分频成分，随着振动增大，轨迹图变扁变大直至振动剧烈时扁圆形发散。

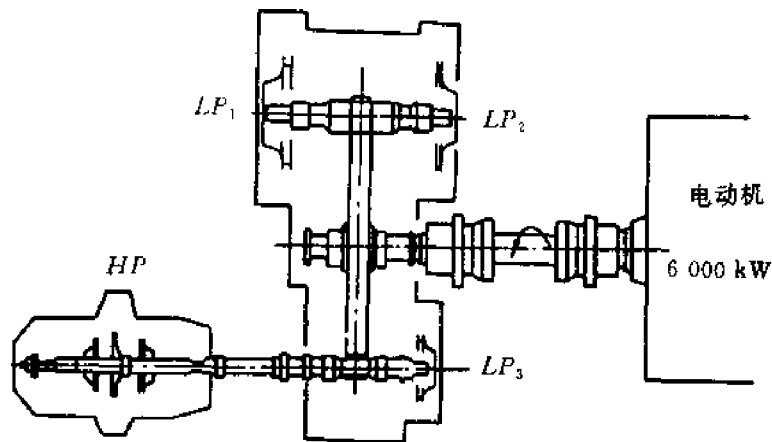


图 5-34 H 型离心压缩机结构简图

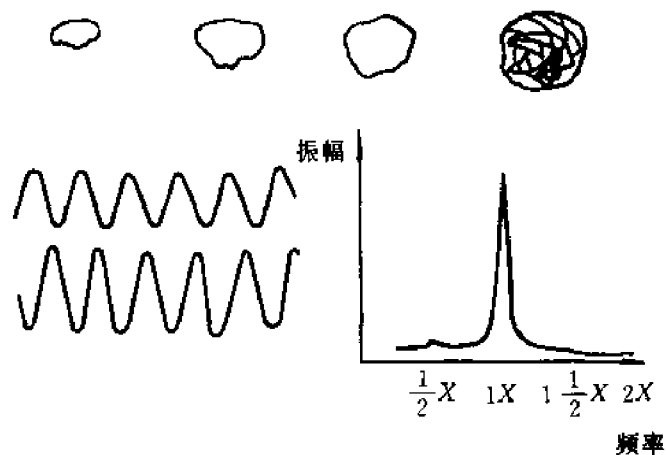


图 5-35 同频振动轨迹、波形及频谱图

(2) $1/2$ 分频振动状态。如图 5-36 所示，由频谱图可见 $(1/2)X$ 频率成分占主导，工频成分很小。波形畸变，波峰时大时小。振动小时轨迹为双圆套在一起，随振动加大逐渐分开，呈“8”字形，轨迹明显变长，即一个方向振动大，另一方向较小。

4. 故障诊断

本例突出征兆是分频振动。产生分频振动的候选故障集主要有 4 种可能的故障：支承系统、碰摩、流体动力和分频谐振，其中第 4 种可以排除。另外，H 型压缩机无隔板，不会发生因隔板倾斜产生流体动力激振，因此流体动力原因亦可排除。余下有支承系统、碰摩两种可能。但从相位特征看，振动从高振幅陡降时相位变化 720° ，即从反进动变为正进动，说明高振幅时为反进动，所以工频振动应是碰摩，从而可初推断 $(1/2)X$ 分频振动也是碰摩产生的。

解体检查验证说明轴承间隙过大是引起转子—定子碰摩的主要原因，经减小顶部间隙、调整对中及增大润滑油粘度后，故障彻底消除。

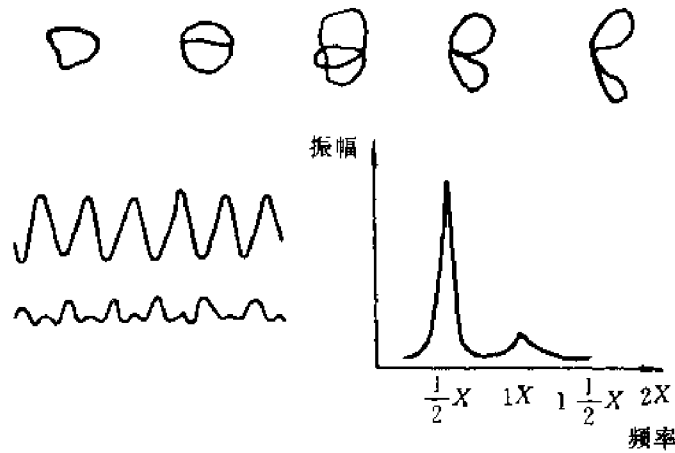


图 5-36 分频振动轨迹、波形及频谱图

5.4.2 某船用汽轮机振动故障的诊断

1. 机组结构及故障概况

某船用机组是由平行布置的高、低压汽轮机及与其相联的两级转动减速齿轮箱所组成,如图 5-37 所示,其高压汽轮机工作转速为 6 900 r/min。当高压汽轮机转速在 5 200 r/min 以下时,机组运行正常,各测点振动都较小。高低压缸前、后轴承的振动幅值均小于 $10 \mu\text{m}$,试车中当转速升至 6 200 r/min 左右时,高压缸后轴承突然起振,振幅(单峰值)达 $40 \mu\text{m}$,当转速升至 6 900 r/min 时,振幅达到 $60 \mu\text{m}$,大大超过允许振动值。

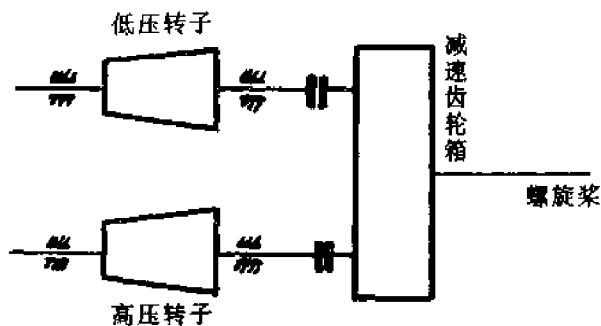


图 5-37 机组结构示意图

2. 现场测试分析

- (1) 振动具有明显的突发性,起振前振动很小($3 \mu\text{m}$),起振后振动很大($40 \mu\text{m}$);
- (2) 起振后随着转速升高振幅值也升高;
- (3) 起振转速与进气方式有关,在部分进气时一般为 $5\ 860 \sim 6\ 180 \text{ r/min}$,而在全周进气时,起振转速推迟,直到 $6\ 780 \sim 6\ 900 \text{ r/min}$ 左右才起振;
- (4) 频谱分析,振动为低频振动,如图 5-38 所示,振动主导频率约为 40 Hz。

3. 故障诊断

经反复分析,并进行转子临界转速及失稳转速的计算,排除多种故障原因后,分析该机组

轴承的长径比为 0.69, 比压为 5.79 MPa, 比压偏低。经理论计算, 其失稳转速为 2 121 r/min, 可见该轴承的稳定性较差。因此断定该机组的振动是由于油膜涡动而发展为油膜振荡。消除振动最有效的措施是采用减小长径比, 提高轴承比压(改变间隙和油温都作过试验, 效果不大)。改变前后的参数变化如表 5-7 所示。

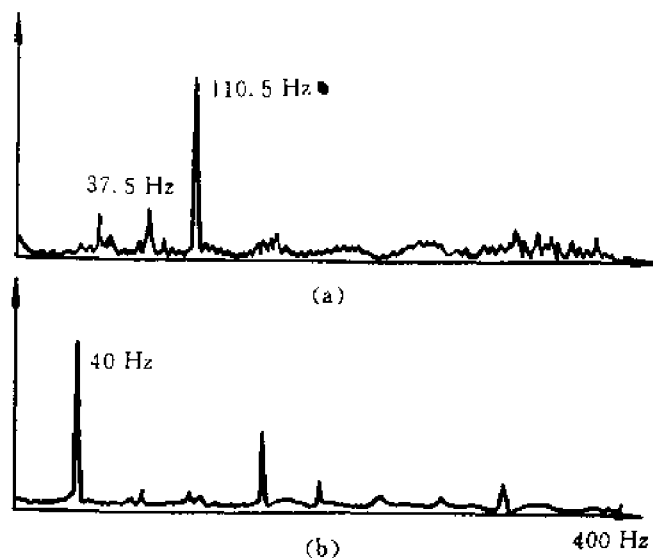


图 5-38 振动频谱图

对高压后轴承进行了上述处理后, 进行验证性试验, 其结果令人满意, 达到了预期效果, 高压后轴承的振幅从原来的 60 μm 降至 2.5 μm , 几乎降低 24 倍, 高压转子运行十分稳定。

表 5-7 处理前后轴承参数

	长径比 L/D	比压 /MPa	轴瓦长 /mm	振幅 / μm
改变前	0.693	5.79	104	60
改变后	0.5	8.04	75	2.5

习题与思考题

- 5-1 旋转机械振动评定方法及其表述。
- 5-2 拍的形成、振动信号单边削平在频率结构上有什么规律和特点?
- 5-3 进行暂态频域分析的手段有哪些?从暂态分析中可以得到什么信息?
- 5-4 转子振动监测方法有哪些?各有何特点?
- 5-5 转子的主要故障有哪些?各有何特点?
- 5-6 产生转子不平衡的原因是什么?转子不平衡产生的振动有什么特点?
- 5-7 转子不对中的类型有哪些?引起转子不对中的原因是什么?
- 5-8 油膜振荡产生的原因是什么?其表现特征如何?
- 5-9 若一转子的旋转转速从零升至 17 000 r/min, 其一阶临界转速为 8 000 r/min, 试问半速涡动频率等于多少?发生油膜振荡时的频率又是多少?
- 5-10 什么是旋转失速?什么是喘振?两者有何异同?

往复机械的监测与诊断

往复机械种类很多,有往复式压缩机、内燃机(柴油机及汽油机)、往复泵等,其应用范围十分广泛,因此,和旋转机械一样,对往复机械进行状态监测与故障诊断具有十分重要意义。

往复机械的故障主要有两种,一种是结构性的故障,另一种是性能方面的故障。结构性故障是指零件磨损、裂纹、装配不当、动静部件间的碰磨、油路堵塞等;而性能方面故障表现在机器性能指标达不到要求,如功率不足、油耗量大、转速波动较大等。显然,结构性故障会反映在机器的性能中,通过性能的评定,也可反映结构性故障的存在和其严重程度。

对于往复机械的监测和诊断,一般应用振动诊断法,通常还须辅以其他的检测方法和手段,如温度监测,油样的光谱、铁谱分析及性能参数的测定等。

表 6-1 是有关工业用柴油机 410 次停机故障的分类统计。这些统计数字反映了往复机械故障的一个侧面。

表 6-1 柴油机故障统计

故障分类	故障发生率 / (%)	故障分类	故障发生率 / (%)
喷油设备及供油系统故障	27.0	调速器齿轮故障	3.9
漏水故障	17.3	燃油泄漏	3.5
气门及气门座故障	11.9	漏气	3.2
轴承故障	7.0	机座故障	0.9
活塞组件故障	6.6	曲轴故障	0.2
漏油及润滑系统故障	5.2	其他故障	5.0
涡轮增压器故障	4.4		
齿轮及传动装置故障	3.9	总计	100.0

6.1 往复机械的动力性能监测

对往复机械性能进行监测可以了解其工作情况并有助于判断其故障。对往复机械中的内燃机来说具体的监测内容有以下几点。

1. 汽缸的压力检测

监测汽缸压力可以检查燃油燃烧是否处于良好状态,内燃机燃油的耗量与燃烧状态存在密切关系。汽缸中的燃烧过程与燃油的品质,预热温度及喷油系统工作状态有关,因此监测汽缸压力具有十分重要意义。

以曲轴转角作为时间横坐标,以汽缸压力为纵坐标的示功图包含有丰富的信息,我们可以利用示功图上压力膨胀线上某一固定曲轴转角时的压力值来获得有关燃烧好坏的信息,如过后燃烧或燃烧阻滞方面信息。根据示功图上最高压缩压力及压力升高率,通过计算可以得到内燃机功率值,燃烧放热规律等性能指标,还可以判断气阀、活塞环与缸套的工作状态及其气密性。利用各缸平均指示压力值可以校检各缸间的功率是否平衡,等等。

对各汽缸测得的压力值进行互相分析也有助于内燃机的工况监测与故障诊断。试验表明,对于调整良好的内燃机,各缸压力信号之间的互相关系数峰值应近似为1,峰值之间所对应的时延即为两缸之间的发火间隔角。对于同一汽缸,利用不同样本的压力变化值进行互相分析也可对汽缸工作过程有无异常进行检查。如各缸工作过程不均衡,则互相关曲线将出现异常。

2. 温度信号的检测

温度是柴油机工作时最基本的热工参数。它对柴油机的综合性能影响很大,是既方便检测又可进行有效诊断的参数,温度信号包括柴油机的水温、油温、排气温度及轴承温度等。

3. 启动性能的检测

启动性能是柴油机的重要指标,影响启动性能的主要因素是启动动力源性能,启动机性能,汽缸压缩性能,供油系统、进气系统以及环境温度等。为此应对动力源(蓄电池)的性能、汽缸压缩性能、供油压力波、进气压力波、启动瞬时转速等进行检测,据此进行单项或多项相关性诊断。

4. 动力性能的检测

动力性能是柴油机的核心,影响动力性能的因素主要有汽缸的气密性、供油量、喷油提前角、进气量等,可以利用内燃机瞬时转速变化对气密性和动力性进行监测和诊断。详见后一节。

5. 增压系统的检测

目前,大部分柴油机(除高速小型柴油机外)都装有废气涡轮增压器,增压器工作好坏直接影响主机工作的状态,因此,为了监测涡轮与压气机的效率,需要测量涡轮进出口压力、温度及其转速,对于压气机也要测量相应的温度与压力参数。

对中间冷却器中增压空气压降的监测可以了解冷却器的污染程度,对空气过滤器前后压降的监测可以确定过滤器有否堵塞现象。

6. 进排气系统的检测

进排气系统监测的内容有:进排气压力,各缸支管排气温度,总管排气温度及排气的成分等,各缸排气温度的变化将反映进排气阀及喷油器工作状态,瞬时排气温度的测定能监测喷油器的故障。

对于往复式空压机性能监测参数,除各点温度、压力外,尚应监测排气量及汽缸的漏气量,以了解气阀,活塞等的工作状态。

另外,往复机械中燃油品质恶化,排气阀泄漏,活塞环磨损,冷却器失效等故障也会对其性能参数发生影响,反之,若往复机械性能发生变化也可判别其故障的存在,因此,利用往复机械性能的监测来判别其故障是一种方便、实用的诊断方法。

6.2 往复机械故障诊断的油样分析法

机械设备故障诊断油样分析是20世纪70年代开始发展起来的新的监测技术,由于该技术的独特作用,目前已被愈来愈多的部门所采用。整个油样分析工作分为采样、检测、诊断、预测和处理五个步骤进行。从润滑油中采样,必须采集能反映当前机器中各个零部件运行状态的油样,具有代表性;检测是指对油样进行分析,测定油样中磨损残渣的数量和粒度分布,初步回答机器的磨损状态,是正常磨损,还是异常磨损;当机器属于异常磨损状态时,需要进一步进行诊断,即确定磨损零件和磨损的类型(例如,磨料磨损、疲劳剥落等);所谓预测,是指预测处于异常磨损状态的机器零件的剩余寿命和今后的磨损类型。根据所预测的磨损零件、磨损类型和剩余寿命即可对机器进行处理:确定维修的方式、维修的时间以及确定需要更换的零部件等。

对往复机械来说主要是采用光谱分析和铁谱分析方法,若两者结合,相互补充,可获得更佳效果,由于受条件限制,目前我国油样分析是以铁谱分析方法为主。

6.2.1 油样光谱分析法

油样光谱分析法是指用原子吸收或原子发射光谱分析润滑油中金属的成分和含量,判断磨损的零件和磨损的严重程度的方法。这种方法对有色金属比较适用。

物质的原子是由原子核和在一定轨道上绕其旋转的核外电子组成。正常状态下,电子的负电核与原子核的正电核是相等的,即电子沿固定轨道绕原子核旋转产生的离心力与原子核对它的吸引力相等,原子处于稳定状态,原子具有的能量也一定。当外来能量加到原子上时,核外电子吸收能量,便从较低能级跃迁到高能级的轨道上去,此时,原子的能量状态是极不稳定的,电子会自动地从较高能级跃回到低能级的轨道,同时以发射光子的形式把它吸收的能量再辐射放出,即发出光。各种不同的元素的原子辐射的光都有各自一定的波长,各种波长的光就组成了光谱。

利用棱镜或光栅的色散作用,将某一光源辐射的光束中代表不同元素的不同波长的光分解开来,并通过感光板记录下来,再利用测光器测出谱线的强度,根据光谱线的强度和元素浓度的特定比例关系,就可求得元素的浓度。元素的浓度愈高,则受激发后产生的辐射线的强度也愈高。从而得到元素光谱的定量分析值,即元素的浓度,用ppm表示。

因为光谱线的强度与光谱的激发条件、激发光源的强度、曝光的时间、以及感光板的性能等诸因素有关,所以一般是测量被分析元素的谱线与一个内标元素的对照谱线的相对强度。内标元素一般是人为外加的,即在分析的油样中和标准样品中,同时加入一定的某一元素为内标。

油样光谱分析法可分为以下几种:

(1) 发射光谱分析法:利用气体火焰、交流电弧或直流电弧等离子体,以及电火花等方法激发油样,以获得发射光谱,从而测得谱线的相对强度,求得分析元素的含量。

(2) 直读发射光谱分析法:将发射光谱通过光电转换仪(如光电倍增管)直接转换为电流,通过测量仪器再转换为元素的浓度,这种方法称为直读发射光谱。一次油样可分析全部所含元素,分析速度快、操作简便。

(3) 原子吸收光谱分析法:在光源和分光镜中间加入透明的物质,把某种单色光的成分吸

收掉,产生原子吸收光谱。被分析的元素都有特定的波长,通过测量原子蒸汽对辐射光的吸收,从而测定油样中的元素的浓度。虽然原子吸收光谱对周围环境的干扰影响较小,分析精度及可靠性较高,但是一种阴极灯只能分析一种元素,操作比较麻烦,分析速度慢。

原子吸收或原子发射光谱仪均有三个主要部分:激发源、光学系统和信号处理及读出系统。如图 6-1 所示。

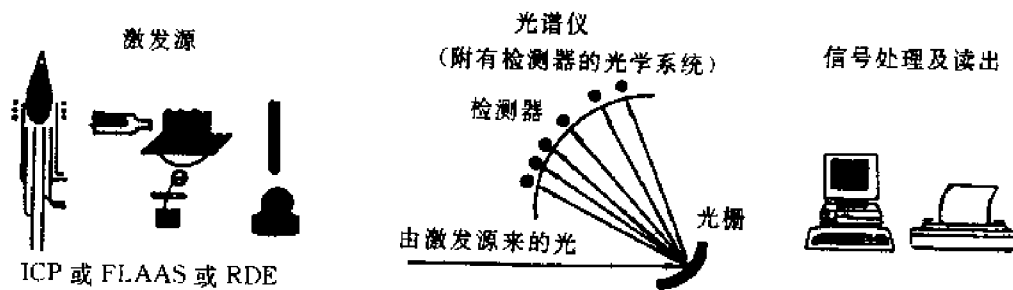


图 6-1 光谱仪系统主要部件

机械设备状态监测油液分析所采用的光谱仪多为原子发射光谱仪(AES),其激发源多为转盘电极型(RDE)或电感耦合等离子体型(ICP)。两者不同之处在于采用不同的方式激发油样,如图 6-2 所示。在 RDE-AES 技术中,有一个旋转的电极,将油样不断地送入到它与一个固定棒状碳电极所形成的间隙中,然后用高压电弧闪激油样,使油样中的各个元素均发出光或辐射能。在 ICP-AES 技术中,激发技术是由惰性气体氩气产生无电极等离子体,氩气不断地通过一匝或三匝射频圈内的等离子体炬管,该线圈与射频交流电发生器相连接,氩气的作用犹如变压器的次级线圈,因此,氩气可以加热到极高的温度。油样被吸入炬管中心,并进入等离子体,油样中的元素被激发,发出辐射能。

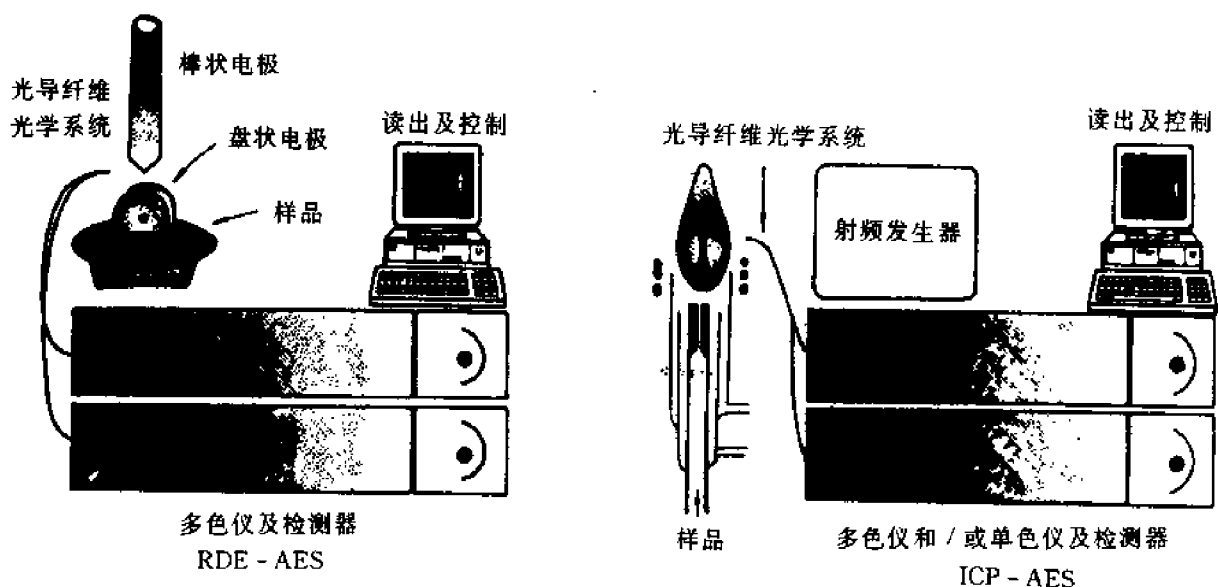


图 6-2 RDE-AES 及 ICP-AES 光谱仪系统简图

RDE-AES及ICP-AES光谱仪工作原理如图6-3所示:用透镜或光导纤维使激发源的辐射能聚焦到光学系统上,通过光学系统的光,照射到一个凹面光栅上,光栅使光色散为因元素而异的各个波长的谱线。用光电倍增管来检测辐射能,并将其转换为倍增的电信号,一个光电倍增管对应某一元素的特定波长光线。电子处理转换器将电信号转换成待测元素的浓度值,仪器屏幕上显示检测结果,最后由打印机输出结果。

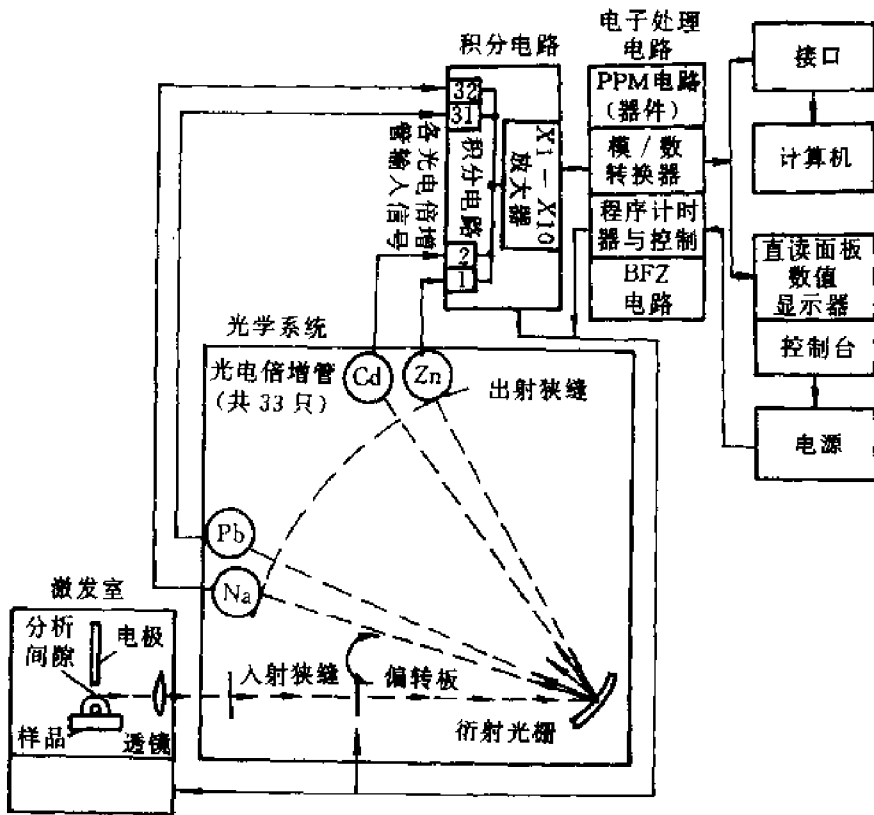


图6-3 电子发射光谱仪工作原理

光谱分析给出的元素成分及其含量 ppm 值,可以提供下列信息:

(1) 磨粒元素的成分及其含量:根据设备运行副零件的材料构成,可以判断磨粒产生的可能部位。

(2) 添加剂元素及污染物元素的成分及含量:根据润滑油的性能要求,可以判断润滑油的劣化变质程度。

(3) 磨粒的增长率:单位时间主要磨粒元素含量 ppm 值可表示磨粒的增长速度,即

$$\text{磨粒的增长率} = \text{磨粒的增加量 ppm 值} / \text{取样间隔时间 (ppm/h)}$$

根据磨粒的增长率可以判断摩擦副的磨损趋势及其严重程度。

例如,柴油机主轴瓦及连杆轴瓦的材料为钢背网状铝锡合金,这种合金是以锡—铝共晶软化相的形式存在的。通过油样光谱分析可知,润滑油中微量的锡和铝的存在,来自主轴瓦和连杆轴瓦的磨损。其机理是由于运转初期润滑油供应瞬时中断,摩擦副油膜破裂以致直接接触摩擦,出现局部高温使低熔点的共晶锡液珠自合金中析出。镁的存在,是来自球墨铸铁曲轴轴颈的磨损,铜和锌的存在则是来自连杆小头锡青铜衬套的磨损,等等。这样,油样光谱分析方法

不但可以定性地判断磨损的零件,而且可以从润滑油中金属成分含量的多少,定量地判断出零件磨损的程度。

油样采集间隔时间应由不同设备的运行特点而定,往往最合适的取样间隔时间是在积累了长期经验之后而确定的。例如,铁路内燃机车油样光谱分析取样间隔时间约为机车行走 5 000 km 左右为宜。通常,可根据工作环境,设备的质量状况,管理水平,以及设备维修情况等因素综合而定。工作环境好,维修质量较高,设备技术状态良好时,取样间隔时间可以长一些;反之,取样间隔时间要短一些。

列表、记录并追踪被监测设备油样增长率及其相应的时间或运行距离,绘制相应的设备磨损趋势曲线,对于设备运行过程中,需要换油或补充新油的情况下,绘制磨损趋势线图时,应作相应的修正,以反映设备磨损过程的真正磨损趋势,图 6-4 为磨粒元素浓度值与时间或距离的关系曲线。显然,一个成功的机械设备状态监测工作,油样光谱分析结果都应该认真建立自身的数据资料档案或数据库。

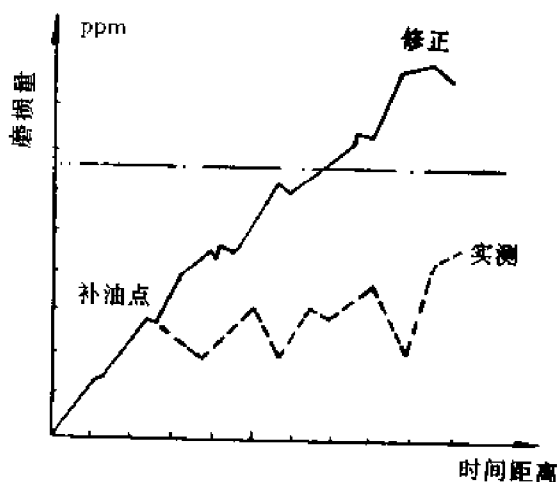


图 6-4 磨粒元素浓度值与时间或距离的关系曲线

油样光谱分析法采用标准的光谱分析仪。这种仪器在生产上使用比较方便,但其价格比较昂贵,从采样到取得分析结果有较长的滞后时间。此外,由于方法本身的限制,不能给出磨损残渣的形貌细节,所分析的残渣的大小一般只能小于 $10\ \mu\text{m}$ 。

6.2.2 油样铁谱分析法

油样铁谱分析法是目前使用最广泛、最有发展前途的润滑油样分析方法。它的基本原理是将油样按一定的严格操作步骤稀释在玻璃试管或玻璃片上,使之通过一个强磁场。在强磁场的作用下,不同大小的残渣所能通过的距离不同,用光学或电子显微镜观察、分析这些磨损残渣的形貌、数量、粒度、和成分,根据油样中残渣沉淀的情况判断出机械设备的运行工况、关键零部件磨损的程度。

铁谱分析中使用的基本仪器是铁谱仪。铁谱仪主要有直读式、分析式和旋转式等,它们均为离线测量分析仪器。目前已研制出能在润滑系统中分离测量磨粒的在线铁谱仪。这里我们主要介绍直读式和分析式两种铁谱仪。

1. 直读式铁谱仪

直读式铁谱仪的主要特点是可以比较迅速而方便地测出磨损特性指标,对设备的磨损状态给出定量分析,特别适用于设备的状态监测,目前在工厂、港口、船舶等处得到广泛应用。

直读式铁谱仪的工作原理如图 6-5 所示。

稀释后的油样装入高处的试管中,经毛细管,油样被虹吸向下流入沉积管,高梯度强磁场装置位于沉积管下方,当油样缓慢流入沉积管时,油样中的铁磁磨粒在磁场作用下,有序地沉积排列在沉积管下部,如图 6-6 所示。磨粒在沉积管中沉降的速度取决于本身的尺寸、形状、密度和磁化率,以及润滑油的粘度、密度和磁化率等许多因素。当其他因素固定后,磨粒的沉降速度与其尺寸的平方成正比,同时还与磨粒进入磁场后离管底的高度有关。从图 6-6 中可以看到,在入口处沉淀有大颗粒和一部分小颗粒残渣,尺寸约在 $5 \sim 10 \mu\text{m}$,在距入口 5 mm 处沉淀有部分小颗粒残渣,尺寸约在 $1 \sim 2 \mu\text{m}$ 。

图 6-5 中两道光束分别穿过上述大小磨粒沉积区,并被位于对侧的光电元件所接受,光电元件所接受到的光束强弱,表示磨粒沉积的数量,即反映磨粒的浓度。设左侧光束测出的光密度值为 D_L ,代表大于 $5 \mu\text{m}$ 的“大磨粒”的相对数量,右侧光束测出的光密度值为 D_S ,代表 $1 \sim 2 \mu\text{m}$ 的“小磨粒”的相对数量。为了铁谱定量分析方便,通常规定大磨粒的读数 D_L 和小磨粒的读数 D_S 之和表示油样的磨粒数量,代表磨损程度;大磨粒的读数 D_L 和小磨粒的读数 D_S 之差表示油样磨粒的尺寸分布,又称为磨损严重度;而大小磨粒数之和 $D_L + D_S$ 与大小磨粒数之差 $D_L - D_S$ 的乘积表示油样磨损严重指数。即

磨粒数量(磨损度): $D_L + D_S$

磨粒尺寸分布(磨损严重度): $D_L - D_S$

磨损严重度指数 I : $(D_L + D_S)(D_L - D_S)$ 或 $D_L^2 - D_S^2$

显然,磨损严重度指数 I 既与总磨损量有关,又与磨损的严重程度有关,所以 I 不但可以反映磨损状况的变化过程,又可以反映磨损状况的严重程度,是铁谱技术定量参数中的重要指标之一。在设备状态监测中,还可归纳如下两个参数:

磨粒浓度: $[(D_L + D_S) / \text{油样量}] 100\%$

大磨粒百分数: $[(D_L - D_S) / (D_L + D_S)] 100\%$

记录绘制磨损程度指标曲线的变化,可以有效地监测设备的早期磨损,长期监测某一特定设备的经验积累,就可以定出磨损程度指标的基准线,从而提高状态监测的有效性。图 6-7 所示为季度取样的趋势图,图中 D_1 表示正常磨损状态下的磨损度, D_2 表示异常磨损状态下的磨损度。在正常磨损状态下, D_1 可能超过 D_2 ,而当磨损趋于严重时, D_2/D_1 比值可能变得很大。

2. 分析式铁谱仪及分析方法

直读式铁谱仪只能提供有关残渣数量和大小的信息,为了进一步确定残渣的形态和成分,

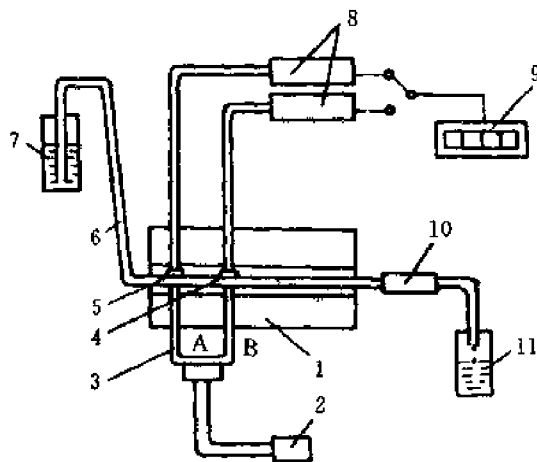


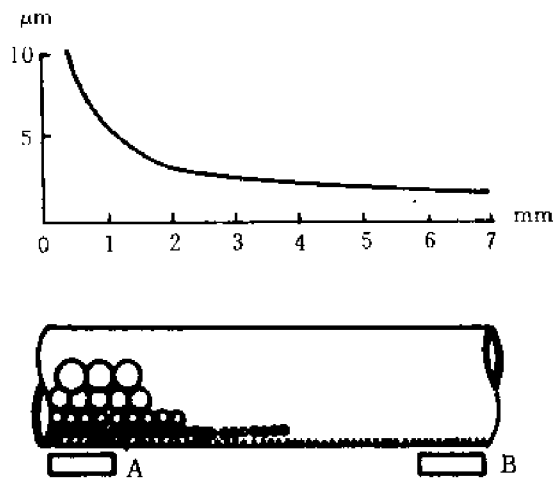
图 6-5 直读式铁谱仪结构简图

磁场装置; 2—光源; 3—光通道; 4—沉积管;

5—光电传感器; 6—毛细管; 7—油样; 8—信息转换器;

9—数字显示器; 10—虹吸泵; 11—废油;

A—第一束光 B—第二束光



A—前传感器;B—后传感器

图 6-6 沉积管内的磨粒排列

A,B—光电元件

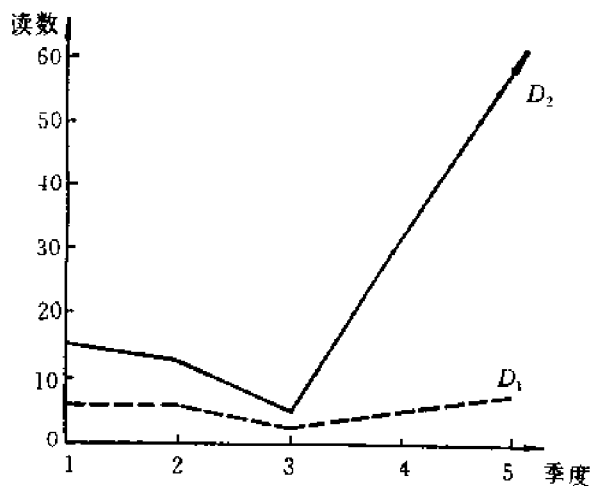


图 6-7 直读式铁谱仪的磨损度趋势

需要采用分析式铁谱仪。分析式铁谱仪是最先研制出来的铁谱技术仪器,由铁谱仪和铁谱显微镜两部分组成。图 6-8 为分析式铁谱仪工作原理图。

按一定要求从设备润滑系统中取得的油样,经稀释、加热后约 2 ml 的待检油样放入玻璃管中,稳定低速率的微量泵输送油样到放置在强磁场装置上方、且成一定倾斜角($1^{\circ} \sim 3^{\circ}$)的玻璃基片上(亦称铁谱基片)。油样由上端以约 15 m/h 的流速流过高梯度的强磁场区,从基片下端流入回油管,然后排入贮油杯中。油样中的磨粒在高梯度强磁场的作用下,按一定的规律排列沉积在基片上,经四氟乙烯溶剂冲洗去除底片上的残油,待固定剂全部挥发干后,垂直向上取下铁谱片,然后用显微镜对残渣磨粒进行观察,根据磨粒的形态可以确定磨损的类型,还可根据磨粒沉积的位置和形态区别出有色金属残渣。例如,沉积部位偏下的大颗粒残渣,其长轴方向与磁力线方向成较大的角度,说明其磁敏感性较低;残渣表面的孔洞和变形褶皱也说明它们比较软,等等。图 6-9 为铁谱片形状和磨粒尺寸分布。实验分析表明,“大磨粒”(≥ 5 μm)一般沉积在距出口端约 50~56 mm 的入口区,“小磨粒”(1~2 μm)一般沉积在距出口端约 50 mm 处,而 < 1 μm 的“细小磨粒”通常沉积在距出口端约 30 mm 以下的区域。

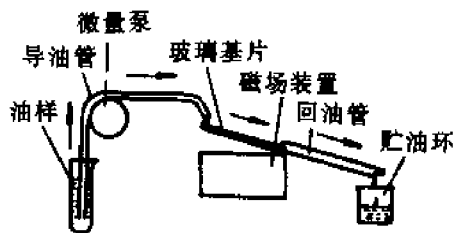
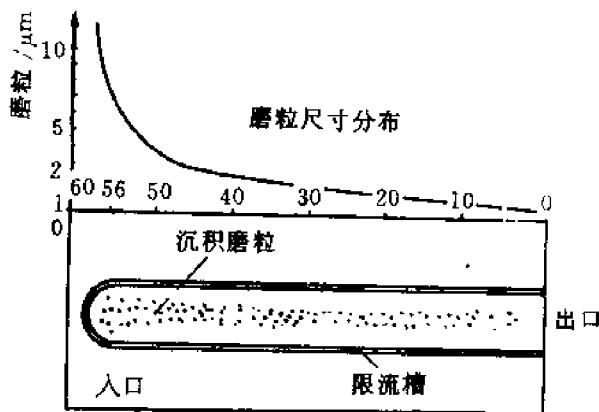


图 6-8 分析式铁谱仪工作原理图



6-9 铁谱片和磨粒尺寸分布

磨损研究表明, 润滑工况下, 相对运动的两表面的磨损状态与磨损过程中产生的磨粒数量、磨粒的尺寸及其分布密切相关。非正常的磨损均会导致磨粒浓度的变化, 严重磨损总是伴随着较大磨粒的数量增加。所以, 测量、记录油样磨粒的浓度变化、尺寸分布变化及其趋势就可以相对定量的诊断和监测设备的磨损状况。图 6-10 为一般金属表面磨损过程与磨粒尺寸及磨粒数量的关系。

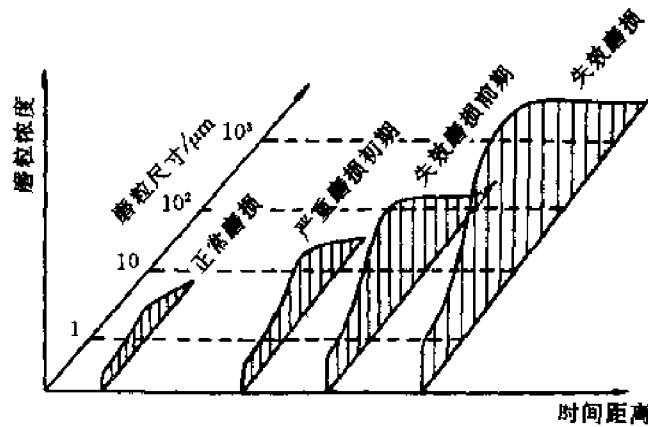


图 6-10 磨损过程与磨粒尺寸

分析式铁谱分析技术的各磨损程度定量指标与直读式铁谱仪的分析指标基本相同, 主要有:

磨粒数量(磨损度): $D_L + D_S$

磨粒尺寸分布(磨损严重度): $D_L - D_S$

磨损严重度指数 I : $D_L^2 - D_S^2$

由于分析式铁谱仪谱片制作的误差、光密度读数器对被测点的圈点误差以及定标误差等, 上述定量分析值准确性较难控制, 许多铁谱技术工作者取直读式铁谱读数和光谱分析值作为磨损量的指标, 分析式铁谱分析技术以观察、分析磨粒的形状、大小、表面形貌等为主。

3. 磨损类型及磨粒相关性

现将不同磨损状态下形成的磨粒在显微镜下的形态描述如下:

(1) 正常滑动磨损磨粒: 摩擦磨损或正常滑动磨损磨粒是指机器正常滑动磨损产生的磨粒。在正常的磨合期内, 磨损表面上形成了特殊的一层, 对钢而言, 这发生在厚度大约在 $1 \mu\text{m}$ 以下的薄层内, 该层称为剪切混合层。正常滑动磨损磨粒是从剪切混合层中剥落下来的薄碎片, 其典型的主要尺寸范围为 $0.5 \sim 15 \mu\text{m}$ 或更小, 它们具有光滑的表面和 $0.15 \sim 1 \mu\text{m}$ 范围的厚度。其主要尺寸与厚度之比从大颗粒的约 $10:1$ 变化到 $0.5 \mu\text{m}$ 颗粒的约 $3:1$ 。

正常滑动磨损磨粒是剪切混合层部分剥落的结果。若剪切混合层未完全剥落, 润滑系统中, 象沙粒这样的过量污染物可能使磨损的发生率增加一个数量级以上, 虽然不大可能发生灾难性的事故, 但这种系统能够迅速遭到磨损。在这种情况下, 虽然大部分磨屑可能是正常滑动磨损的颗粒, 即将发生的故障还是可以用磨粒数量明显增加来预报。磨屑实际数量上的增加取

决于污染的类型和数量。易于发生这种问题的零部件是那些具有大致相同硬度的配合面,例如,柴油机的汽缸壁和活塞环的运动配合面。对这类油样的磨粒分析将揭示出污染颗粒及磨屑。

(2) 切削磨损磨粒:它是由一个摩擦表面切入另一个摩擦表面形成的,或者是由润滑油中夹杂的砂粒、其他部件的磨损残渣切削较软的摩擦表面形成的,其形状如带状切屑,宽度为 $2 \sim 5 \mu\text{m}$,长度为 $25 \sim 100 \mu\text{m}$,偶尔由于磨损表面有硬夹杂物存在,可能会产生大约 $5 \mu\text{m}$ 长、 $0.25 \mu\text{m}$ 宽的小颗粒。

切削磨损颗粒是不正常磨损,应当仔细地检测它们的存在及数量。若系统中的大多数切削磨损颗粒达到了大约几微米长、几分之一微米宽,应当怀疑是否有污染物存在;若系统显示出大的切削磨损磨粒($50 \mu\text{m}$ 长)的数量在不断增加,预示着部件将可能发生失效。

(3) 滚动疲劳磨损磨粒:滚动轴承疲劳磨损产生的颗粒,像滚柱端部磨损、滚道和涉及到滑动的任何其他接触产生的磨粒。已经发现有性质不同的颗粒类型与滚动轴承疲劳有关,即疲劳剥落颗粒、球状颗粒和层状颗粒。

疲劳剥落颗粒呈片状,其主要尺寸与厚度之比大约为 $10:1$,它们具有光滑的表面和任意不规则的周边,最大尺寸达 $100 \mu\text{m}$ 。当产生宏观剥落而发生疲劳时,颗粒的尺寸可能继续增大,按照大于 $10 \mu\text{m}$ 的颗粒的数量是否增加,可以判断出初期出现的不正常磨损。某些设备中产生的磨屑数量可能不大,但这些磨屑会引起设备功能上的严重损失。

球状颗粒是在轴承疲劳裂纹中产生的,若发生这种情况,那么球状颗粒的出现可以作为即将发生故障而需要采取措施的一种预报,因为往往在轴承产生任何实际剥落之前就能检测到这种球状颗粒。迄今为止,在所有检测的工业系统中,滚动轴承疲劳剥落都是发生在直径为 $1 \sim 5 \mu\text{m}$ 球状钢颗粒大量产生以后。据估计,在失效过程中,轴承要产生数百万颗球状颗粒。

层状颗粒非常薄,游离金属颗粒主要尺寸为 $20 \sim 50 \mu\text{m}$,厚度约为 $1 \sim 2 \mu\text{m}$,主要尺寸与厚度之比一般为 $30:1$ 。层状颗粒可能在轴承的整个使用期内产生,但是在疲劳剥落开始发生以后,这种颗粒产生的数量将会增加,所以,若它们的数量增加,而且发现有不明根源的严重磨损发生,便表明滚动轴承工作中出现了故障。同样地,若随着球状颗粒的大量产生,层状颗粒的数量也增加,这表明滚动轴承存在着疲劳裂纹了,这些裂纹将导致材料的剥落。

(4) 滚动疲劳兼滑动疲劳磨粒:主要是由齿轮节圆上的材料疲劳剥落形成的。产生的颗粒形状不规则,主要尺寸与厚度之比在 $4:1 \sim 10:1$ 之间变化。当齿轮的载荷过大、速度过高时,齿面上也会出现凹凸不平,表面粗糙的擦伤。一旦发生擦伤,它通常影响到齿轮上的每一个齿,并产生大量的磨屑。齿面存在的拉伸应力会引起疲劳裂纹,而且裂纹向轮齿深处扩散,直至剥落形成块状磨粒。

(5) 严重滑动磨损磨粒:它是在摩擦面的载荷过高或速度过高的情况下由于剪切混合层不稳定而形成的。磨粒呈大颗粒剥落,引起磨损率增加。若因载荷或速度的原因而使磨损表面上的应力进一步增加,整个表面将遭到破坏,磨损率会达到灾难性的数值。

严重滑动磨损磨粒的尺寸范围在 $20 \mu\text{m}$ 以上,厚度在 $2 \mu\text{m}$ 以上,由于滑动的结果,其中有些磨粒带有表面擦伤,并且具有锐利的直边。在这种磨损形式下,当磨损进一步加剧时,磨粒上的擦伤和直的刃边会变得更加突出。

6.2.3 油样分析诊断技术应用

油样分析技术在机械设备状态监测与故障诊断中的应用实例较多,这里仅以柴油机状态监测为例进行说明。

采用柴油机作为发电动力设备的发电厂中,监测柴油机的运转状态,防止突发事件的发生,加强早期预防柴油机关键运动件的故障,指导设备现代化润滑管理是十分重要的。油样分析技术是柴油机状态监测最有效的方法之一。

1. 腐蚀度

用以监测润滑油腐蚀性以及柴油机主要零件的腐蚀程度。通常借助于分析铁谱技术,将腐蚀磨损的颗粒沉积在铁谱片上,检测其颗粒的数量及形貌。借助于光谱技术,可以检测主要零件的腐蚀程度,例如检测 Fe, Cu, Pb 等主要金属元素的浓度及其变化。借助于直读铁谱技术可以检测出大磨粒数与小磨粒数的相对比例。

2. 氧化度

用以监测润滑油被氧化的程度,利用红外光谱测定润滑油中氧、氧化物的变化趋势以及添加剂衰减的趋势。

3. 污染度

用以监测润滑油被金属或非金属杂质污染的程度。在许多情况下,引起故障的原因是外来污染物对润滑油的污染,例如灰尘、泥沙、积碳颗粒以及油变质形成的化合物、集合物等。通常污染度可分为 10 级:1~3 级有少量污染物,4~7 级有一定数量污染物,应注意,大于 7 级应采取措

施。

润滑油中金属磨粒浓度的监测,大多数借助于光谱技术测量,用发射光谱仪可以监测柴油机 9 个最有价值的元素浓度值。

Fe, Cr, Mo 反映主要钢铁零件的磨损和腐蚀;

Cu, Pb, Sn 反映轴承的磨损、腐蚀;

Si 反映环境的污染程度;

Al 反映活塞裙部或 AlSn 合金轴承的磨损;

V 反映燃烧产物或燃油的污染程度。

由于发射光谱仪一般只能最有效地检测润滑油中金属颗粒小于 $10\ \mu\text{m}$ 的悬浮于油中的磨粒,所以发射光谱测出的金属元素浓度变化范围,可以很好地反映零件的腐蚀和摩擦副表面正常磨损的变化趋势。如果有异常磨损现象,发射光谱仪测出的元素浓度的突然变化,可以给出很明显的反映。

柴油机出现的异常磨损,如擦伤、刮伤、磨粒磨损、粘着磨损以及疲劳磨损等,产生的磨粒其尺寸往往都大于 $10\ \mu\text{m}$,严重磨损时可产生几百微米的磨粒,所以监测预防柴油机的烧瓦、拉缸等严重故障,还应借助于铁谱分析技术。直读式铁谱仪可以较快地给出大磨粒相对值 D_L 和小磨粒相对值 D_S ,尤其是 D_L 值对于分析严重磨损更有意义。 $D_L + D_S$ 代表了磨粒的总相对浓度,所以对于柴油机状态监测时,应记录追踪 $D_L + D_S$ 的变化趋势;同时,在分析严重磨损原因时,还应参考磨粒的形貌显微观察,从而作出比较可靠的结论。

图 6-11 为某发电厂柴油机高磨损时检测的铁谱及光谱参数值,表 6-2 为某发电厂柴油机轴承烧瓦故障前后,润滑油油样铁谱分析、光谱分析的检测参数值。

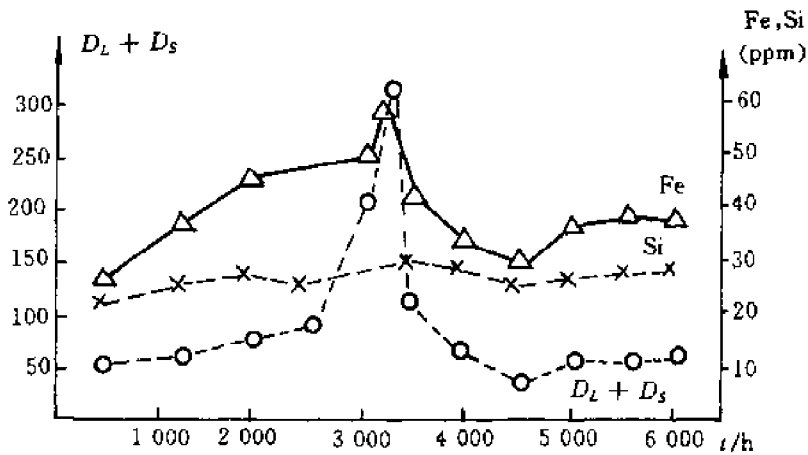


图 6-11 某柴油机高磨损时检测的铁谱及光谱参数值

表 6-2 某柴油机轴承失效油样分析结果

油样编号	粘度 40℃ mm ² /s	污染度 等级	光谱分析			铁谱分析		
			Fe/ppm	Cu/ppm	Pb/ppm	D_L	D_S	$D_L + D_S$
1	169	4.0	18.3	< 2.0	3.7	6.5	5.0	11.5
2	162	5.5	19.0	< 2.0	4.5	11.2	8.2	19.4
3	158	6.5	17.7	3.0	1.8	23.9	8.5	32.4
4	155		16.9	2.2	2.8	39.8	19.8	59.6

6.3 瞬时转速检测法

在往复机械的监测与故障诊断中,无论是内燃机的输出轴瞬时转速,还是压缩机的输入轴转速,都综合反映了机器的工作状态和工作质量,并包含有机器各部件工作状态的大量信息。例如,柴油机调速齿杆位置固定后,转速的平稳意味着柴油机工况的稳定性,转速波动则表示柴油机工况发生了变化。柴油机的瞬时转速可定义为:飞轮齿圈上两个相邻齿间隔内的平均转速,即工作循环内每 $360/z$ 度曲轴转角的平均转速(z 为柴油机飞轮齿数)。尽管这种定义不够严格,但在实际工程应用中,其精度能够满足要求,而且可直接利用飞轮齿圈等分工作循环,使测量方便易行。具体测试时可用磁电或涡流传感器测量飞轮齿圈任意两轮齿之间的时间间隔,用来作为瞬时转速的测量;另用一传感器安装在齿轮轴上的转轮处,作为上止点检测器,以此来确定相位。

由于内燃机缸内气体压力的波动及往复惯性力的变化,其输出的扭矩是波动的,在阻力矩一定情况下,曲轴的瞬时转速也是波动的,曲轴的瞬时转速在一定程度上反映了机器工作状态和工作质量。

6.3.1 汽缸动力性能的诊断

内燃机工作正常时,各缸的动力性能基本一致,柴油机运转平稳,测取的瞬时转速波形规则、波形峰谷分布均匀,有很好的周期性;当某缸工作不正常时,特别是机器发生故障时,如某缸供油量不正常或供油提前角不当等故障出现时,动力的一致性遭到破坏,该缸的动力性能将出现差异,这种差异将在其瞬时转速波形图中表示出来。图6-12为某柴油机瞬时转速波形图,上图表示各缸动力性能正常,下图表示某缸有断油的故障,其波形表现凌乱,不规则。

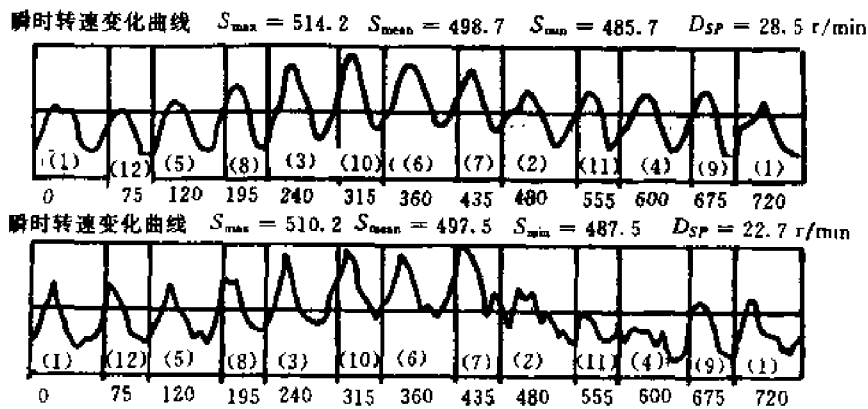


图 6-12 柴油机瞬时转速波形图

上图 — 各缸动力性正常,下图 — 某缸断油故障

柴油机各缸动力性能的差异,将会影响其运转的平稳性,当某缸动力性能下降,如供油不足时,其爆发压力会低于正常值,在瞬时转速波形上表现为其上升沿变化较缓,速度峰值变低;反之,当某缸动力性能高于正常值时,如供油过量,汽缸爆发压力高于正常值,其波形表现为上升沿变较陡,速度峰值高于正常峰值,因此,根据瞬时波形的变化可以方便地诊断柴油机各缸的动力性能。

瞬时转速的波动是由于各缸的顺序工作、负载及阻力等因素所致,当某缸处于压缩上止点时,瞬时转速降至最低点,爆发冲程至下止点时,瞬时转速上升到最高点,该转速差直接反映了汽缸的工作性能,对于柴油机爆发工作缸,其爆发瞬间对曲轴作的功远大于其余缸合力功。因此,通过比较各缸爆发时瞬时转速的变化可以诊断各缸的动力性能。

6.3.2 汽缸气密性的诊断

柴油机汽缸的气密性反映了汽缸磨损、装配等状况,与其动力性能及经济性能有密切关系,气密性不良将引起输出功率下降,经济性变坏。因此,柴油机工作一段时间后,应该检测汽缸的气密性。采用测量起动倒拖时瞬时转速的方法来诊断汽缸气密性较其他方法更简便、实用,因为起动倒拖瞬时转速波形与柴油机汽缸压缩压力有很好的对应关系。应用时,断油起动柴油机,到转速基本稳定后测取瞬时转速,根据瞬时转速波动情况判断各缸的压缩性,从而检测各缸的气密性。

图6-13是实测的起动瞬时转速波形图,图中带括号的数字是按柴油机点火顺序排列的缸号, S_{max} 为测量工作循环内转速最大值, S_{min} 为转速最小值, S_{mean} 为测量工作循环内平均转速,

D_{sp} 为最大转差 ($S_{max} - S_{min}$)。由图 6-13 上图可知各缸瞬时转速起伏均匀, 比较规则, 其原因是柴油机各缸压缩性能良好, 工作循环内各缸压缩阻力基本一致, 对转速的影响基本相同, 因此测得的转速波形比较均匀, 说明各缸气密性良好。图 6-13 下图表示第 2 缸有漏气、气密性较差时的瞬时转速波形, 直观上看, 上、下两图有很大差别, 波形不在均匀, 在漏气缸位上形成一个凸起, 其原因是: 漏气缸的压缩阻力比其他缸小, 倒拖力矩不变时, 在漏气缸压缩上止点处转速不但下降, 反而上升, 从而在转速波形上形成一个凸起, 此时的平均转速为 $S_{mean} = 248.1$, 最大转差为 $D_{sp} = 14.9 \text{ r/min}$, 都比正常状态时测得的值 ($S_{mean} = 241.8, D_{sp} = 9.5 \text{ r/min}$) 增大了。实践表明, 不同缸的漏气及漏气程度不同, 都能从瞬时转速波形图上表示出来。

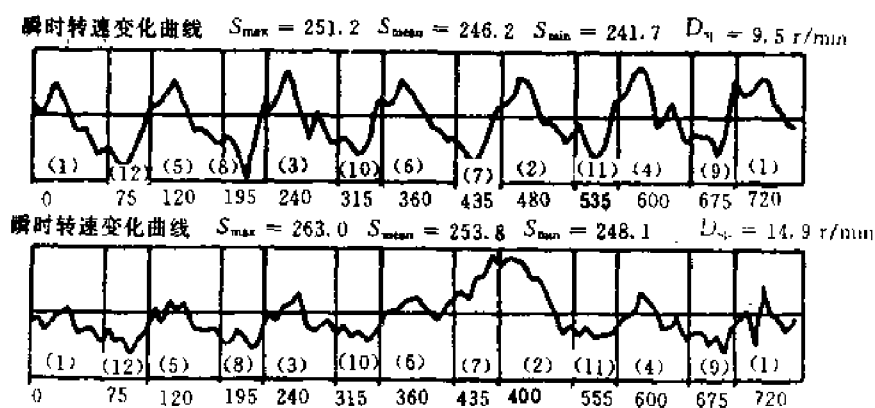


图 6-13 某柴油机两种瞬时转速比较

上图—各缸气密性良好; 下图—第 2 缸气密性较差

6.4 往复机械故障的振动诊断法

6.4.1 往复机械的脉动性

往复机械的特点是运动件多, 而且复杂, 在其工作时引起振动的激励源很多, 振动的主要形式是脉动性。往复机械的运动机构主要是曲柄连杆机构, 活塞(或柱塞) 往复运动引起的振动、气体(液体) 的脉动、各部件之间的周期性撞击都会使机体产生周期性脉动。为了说明这一点, 下面讨论一下周期性脉冲信号。设周期信号的周期为 T , 单个脉冲作用时间是 ζ , 如图 6-14(a) 所示, 此周期脉冲的频谱如图 6-14(b) 所示。

这种脉冲频谱特征在实测的振动信号中得到很好的体现。图 6-15 所示的往复式五柱塞注水泵泵头水平方向振动信号就是典型的冲击信号, 冲击源主要是柱塞往复运动惯性力通过连杆、曲轴产生的周期性激励, 进、排液阀以一定的频率撞击阀座所激励的信号等的综合响应。它的频谱谱峰分布于两个频带内, 一群谱峰分布在的 $0 \sim 152 \text{ Hz}$ 低频带内, 一群分布于 $260 \sim 450 \text{ Hz}$ 的中高频带内, 相邻谱峰频率相差 6.17 Hz , 为柱塞泵的曲轴回转频率(转速为 370 r/min), 频谱呈转频的倍频分布, 和理论上的分析接近。同样, 往复式压缩机也有相似的脉动性, 只是频带的分布有所不同而已。所要说明的是: 理论分析只考虑一个脉冲, 实际频谱是许

多冲击信号在所测点的叠加信号的谱,各信号相位不同,传到测点的时间也不同,因此各信号在测点叠加要抵消掉一部分,所以叠加的结果未必使振动加强,在某些频率上的能量会变得很小,谱图上高频区(260~500 Hz)谱峰的高低错落正说明了这一点。正常信号的脉动特征在机器出现故障时会有所改变,其表现形式是谱图的能量分布及峰值的变化。

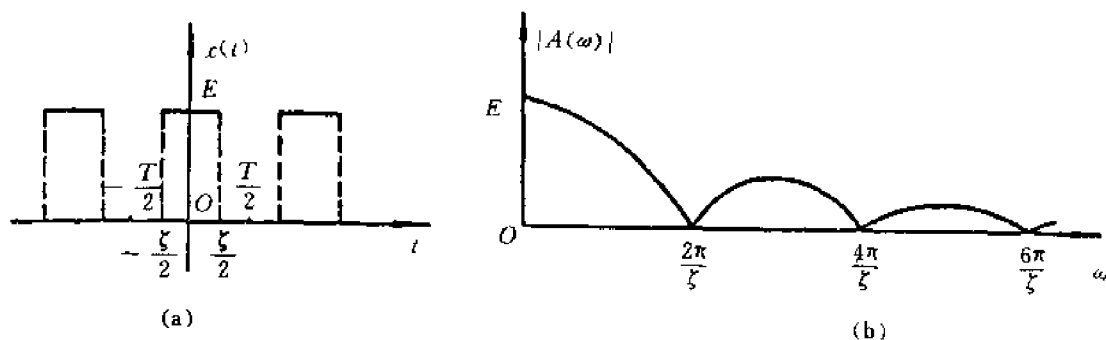


图 6-14 周期性脉冲信号及其频谱

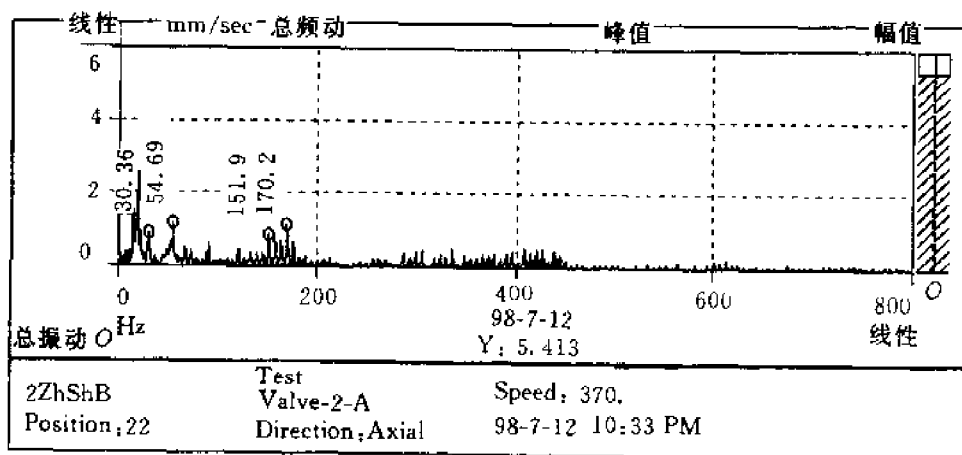


图 6-15 五柱塞水泵泵头水平方向振动频谱

6.4.2 振动诊断法

在往复机械中振动法的应用不如旋转机械那样广泛和有效,这是因为往复机械转速低,要求传感器有良好的低频特性,因而在传感器选用方面有一定限制。此外,由于往复机械结构复杂,运动件多,工作时振动激励源多,对不同零部件,这些激励源的作用是不同的,因而利用振动信号进行分析困难较多,但近年由于振动分析技术的发展,已日益得到更多的应用。

1. 传递函数法

传递函数是用来描述系统的动态特性,缸套与机身的传递函数可以通过振动实验求得,目前研究较多的是利用发动机缸盖系统的动态特性诊断汽缸内的故障。

发动机缸盖系统是一个复杂的机械系统,主要表现在它本身的结构复杂和承受多种激励。在发动机工作时,缸盖系统承受汽缸内气体压力、气门落座瞬时冲击力、活塞不平衡往复惯性力和曲轴不平衡回转惯性力以及随机激励等。从整个发动机结构来看,气体压力、气门落座冲

击力使缸盖产生相对机身的振动;而不平衡惯性力、沿活塞连杆向下传递的气体压力则通过机身传递到缸盖上,使机身和缸盖一起振动。图 6-16 是发动机缸盖振动响应的时域波形,从图中看到:对缸盖振动影响较大的是气体压力、气阀落座冲击和排气阀开启,进气阀关闭和机身振动对缸盖振动影响较小。对于这样一个系统的振动特征,可以简化成一个多输入单输出线性系统。由发动机的工作特性及图 6-16 可知:缸内气体压力(燃烧压力)、排气阀落座冲击和开启以及进气阀关闭所产生的响应各自按照一定规律作用于缸盖,根据缸盖表面测得的振动响应推断各个激励性质,从而对故障进行判断。

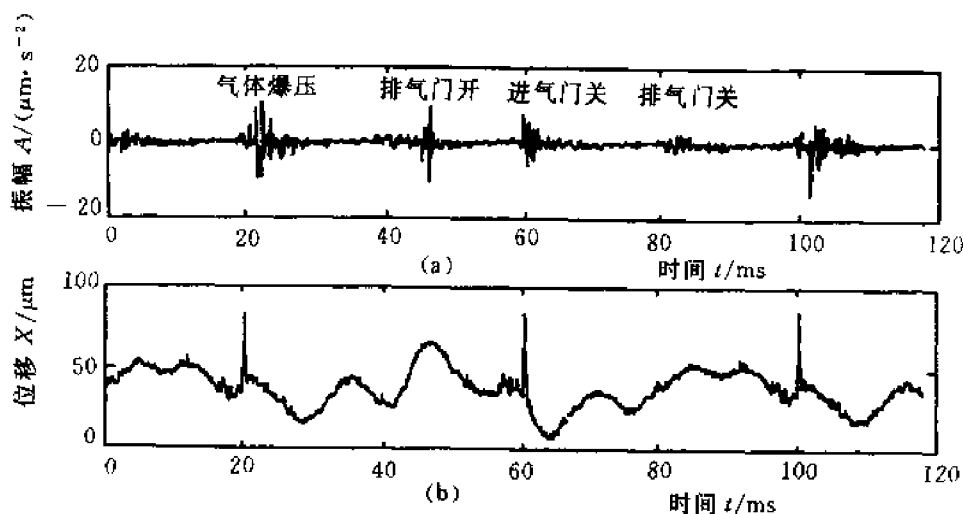


图 6-16 发动机缸盖振动信号

2. 能量谱法

应用振动信息频域分析中的响应能量谱也可对故障进行诊断,振动信号能量谱计算公式为

$$E = \sum_{i=1}^N S_i \Delta f \quad (6.1)$$

式中, S_i 为功率谱密度值, Δf 为频率分辨率。用式(6.1)可以计算出某个频带内的能量 E , 或整个频带内的能量 E 。因为当发动机某部件发生故障时,其能量谱会发生变化,将实测的能量谱值与正常工作状态下的参考谱值进行比较,即可判别汽缸活塞组的工作状态如何。一般来说,故障状态下的总能量值要比正常状态下的大得多。还可将能量谱法发展成为功率谱的数理统计法,测量正常工作状态发动机机体的振动信号,计算的功率谱建立标准谱;再测取实际工作状态下信号的功率谱,将二者相比较,采用如下公式

$$C = \frac{\sum_{i=1}^N w_i S_i w_i S_s}{\sqrt{\sum_{i=1}^N (w_i S_i)^2 \sum_{i=1}^N (w_i S_s)^2}} \quad (6.2)$$

式中, S_i 为实际测得的功率谱, S_s 为标准谱, w_i 为加权系数。因为在实测的功率谱中,对故障敏感的特征频率有一定的带宽,这样,对不同的谱值取不同的权系数,有助于提高对故障的敏感

程度。系数 C 的值在 $0 \sim 1$ 之间, C 的值越小, 说明实测功率谱与标准谱的差距越大, 故障也就越严重。如果建立故障时的功率谱参考值, 用实测谱与其比较时, C 的值越大, 说明实际工况的故障程度越严重。

3. 时域特征量法

利用时域信号中的特征量来判断柴油机故障也是十分有效的方法, 常用的时域特征量有 Kullback-Leiber 信息距离指标, Bhattacharyya 距离指标等, 具体计算这些指标的方法可参阅第 4 章。

除以上几种方法外, 其他如评定缸体表面振动加速度总振级方法, 也是在实际中经常应用的方法, 综合运用上述各种方法可以有效地确定汽缸—活塞组的各种故障。

6.4.3 振动诊断法应用举例

1. 柴油机拉缸时的故障判别

汽缸—活塞组为柴油机工作的动力部分, 它的故障将会导致柴油机不能正常工作, 甚至会使柴油机损坏, 造成巨大的经济损失, 拉缸是汽缸—活塞组十分严重的故障, 所谓拉缸, 是指汽缸套表面与活塞表面间相互作用而造成的严重表面损伤。

造成柴油机拉缸的原因有多种, 如果汽缸与活塞之间的间隙不正常就会导致故障的发生, 间隙过大时, 燃烧气体在活塞环和汽缸壁之间有泄漏, 有可能使活塞环因位置偏斜而粘牢咬死在环槽内; 如果活塞和汽缸套之间的间隙过小, 活塞会在汽缸套中咬死, 产生卡瓦、拉缸等故障。其他如活塞与汽缸套之间润滑不良, 活塞环断裂、活塞销装配过紧都会引起拉缸故障。

当间隙过小发生拉缸时, 在缸体表面测得功率谱密度图中高频成分 ($> 3 \text{ kHz}$) 明显增加, 如图 6-17 所示。这与正常工作情况下不同的特征说明了此时活塞作用为宽频带激励, 反映到缸体振动上是能量分布带宽增加, 同时, 总振级测量值明显小于基准值。据此, 可以判别拉缸已经发生。

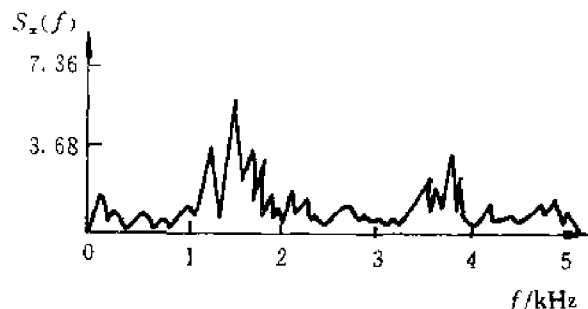


图 6-17 拉缸时表面振动功率谱密度

2. 汽缸活塞磨损故障的判别

可以利用缸体表面振动加速度总振级进行判别, 若正常工作状态下各测点的振动加速度总振级为 L , 实测各点的振动总振级为 L_a , 比较这两值的倍数, 可以确定汽缸磨损状态并确定

磨损极限。

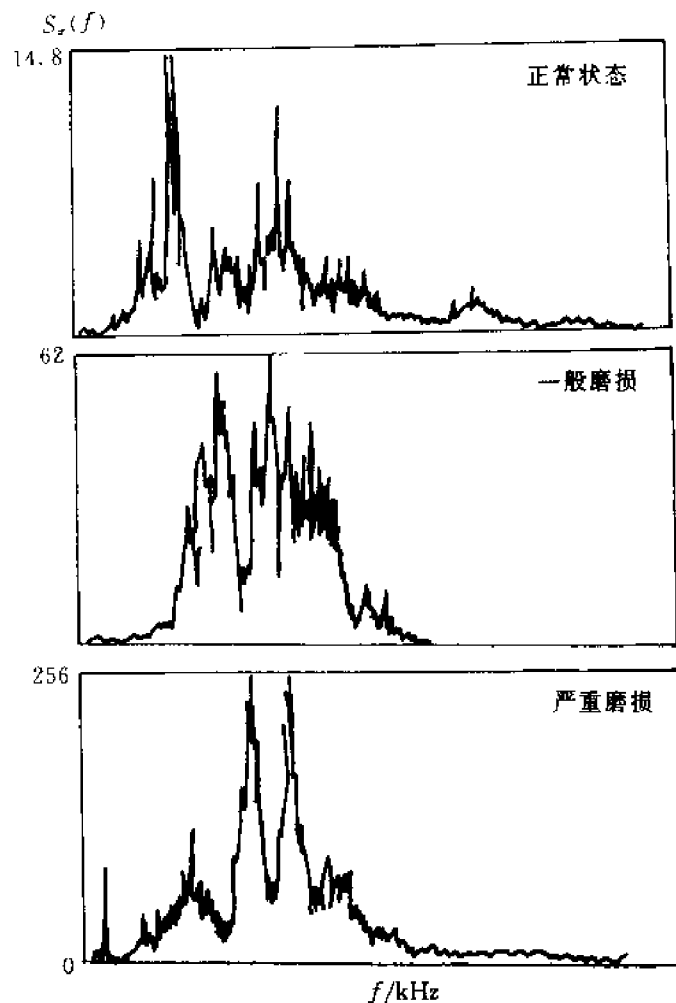


图 6-18 车用发动机在不同间隙状态下机身表面振动功率谱

发动机汽缸磨损可以通过活塞—汽缸套之间的间隙反映出来,图 6-18 是 X4105CQ 型车用发动机在不同间隙状态下,机身表面振动响应的功率谱。从图中可知,功率谱峰值随汽缸磨损量的增加而加大,达到极限磨损后,峰值急剧增加。表 6-2 是 1 缸处在不同间隙情况下机身的振动加速度总振级,从中可见,磨损量增加,总振级呈上升趋势。

表 6-2 机身振动加速度总振级

汽缸套磨损量 /mm	0	0.12	0.20	0.60
总振级 $g/(9.8m \cdot s^{-2})$	25	80	150	210

3. 气阀漏气故障

气阀间隙变大将导致气阀漏气,影响发动机的性能,而气阀不严漏气也是柴油机的常见故障。

图 6-19 是一车用发动机在不同气门间隙时测得的振动响应时域波形,测点是在进气门底座附近。由图可见,气门间隙过大时,气门激励引起的缸体表面瞬态响应比较明显,且随着气门间隙增大而出现超前的特征,响应加速度的振幅随气门间隙增大而增大。

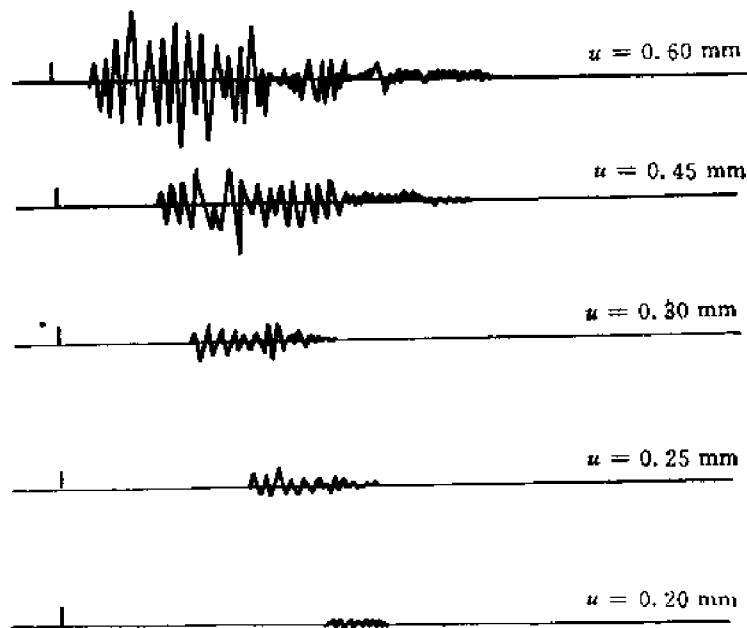


图 6-19 不同间隙时表面测点响应时域波形

响应时域波形均方根值反映信号能量的大小,因气门激励响应为瞬态响应,其中包括有限的能量,故可用均方根值来描述时域响应的特征,表 6-3 给出了进气门落座表面测点振动响应的均方根值。从表中可知,随气门间隙的增大,均方根值也增大。

表 6-3 进气门落座表面测点振动均方根值

间隙 /mm	0.60	0.45	0.30	0.25	0.20
动均方根值 / $(m \cdot s^{-2})$	5.383	4.234	1.161	0.889	0.633

根据实际测试结果,可以看出,故障波形作用在同一转速、同一测点上,其幅值随气阀间隙增大而增大,两缸气阀间隙异常时,其波形相互独立,各自在相应的相位上出现故障波形,因此,利用时域波形分析方法可以判断发动机气阀漏气的故障及其部位和程度。

还可在不同试验条件下测取进排气门开启和底座冲击时的振动响应信号并进行频谱分析以获得信号的频谱特征。谱的总能量随间隙的增大而增大(也随转速升高而增大),其特点是谱的能量均集中在某一频率附近,此频率为进气门测点的气门落座特征频率,气门开启和底座激励在表面的测点产生的振动响应,都有各自的特征频率,谱的能量主要集中在特征频率附近的频带内,而特征频带内的能量随气门间隙及发动机转速增大而增大。

从发动机缸盖系统的响应分析可知,气体爆炸是一个低频($f < 500 \text{ Hz}$)的激振力,如果气门漏气在气体爆炸作用力上将增加一个准“白噪声”作用力,缸盖响应的高频部分能量将增加。这是因为从狭缝喷流的声学特性的研究表明,漏气的声学信号相当于一个频率范围很宽的准“白噪声”信号。因此,根据测得的高频信号特征可以判别气门漏气情况。

在实际测试时应注意传感器安装的位置,因为漏气能量相对其他激励能量较小,气门漏气反映在缸盖上的振动特征,易被其他干扰信号所淹没。另外,应防止高频信号被滤掉,必须采用

高通滤波器。

除了以上从时域及频域能量谱对气门漏气故障进行诊断外,也可应用 AR 谱进行识别,测试表明,AR 谱突出了各测点响应信号的特征频率,AR 谱各种能量指标随气门间隙增大而增大,利用这些指标法可以判断气门间隙的变化情况。

以上对振动法在往复机械的诊断中的应用作了系统的介绍,以活塞—汽缸拉缸、磨损和气门漏气三种故障作为实例进行了分析。往复机械的故障种类很多,除上述几种外,尚有曲轴连杆组件中各种部件的故障,有配气机构方面的故障、供油系统的故障等,这里不作介绍。

6.5 柴油机供油系统故障诊断压力波形分析法

主要由燃油泵、出油阀、高压油管和喷油器组成的供油系统,是柴油机的一个重要组成部分,它直接影响燃烧过程和柴油机的工作性能,据统计资料表明(见表 6-1)柴油机的故障 30% 以上是发生在供油系统。

供油系统工作不正常的结果是直接降低功率和热效率,功率下降后,必须增加供油量以满足增大功率的需要,热效率降低则使废热大量增加,增加的废热往往成为引发一些重大故障的根源,如可引起活塞过热,排气门烧蚀,润滑油结焦和水温、油温不正常升高等。再比如喷油器的喷射能力改变或喷油器针阀运动受阻,这都会影响喷油雾化质量,造成因燃烧不良引起的故障。这类故障刚发生时并无明显的异常现象,但缸内结焦,润滑油损坏以及影响各缸载荷分配等影响运转可靠性的因素在不断发展,由于严重结焦,活塞环局部粘结,排气门的密封性也被破坏,进一步恶化了燃烧过程,使柴油机运转不正常。

由此可见,监测和诊断供油系统的工作状态,对保证柴油机可靠安全的运行和工作寿命,以及增加动力经济性甚为重要。

传统的诊断方法是依靠生产人员的经验,根据柴油机发生故障时的外部征象来判断的,这种经验法只有在故障表现的比较严重时才容易判断,而不能在早期发现故障,所以它是一种比较被动的方法。另外,运用柴油机表面振动信号的频谱形状对供油系统的某些故障进行诊断也有研究,但由于频率范围很广,干扰故障特征的提取,给准确判断故障类型造成很大困难。

供油系统在不解体诊断情况下,监测雾化油的质量(平均油滴直径、靠近喷孔的油束锥角等)是相当困难的或是不可能的,但获取供油系统的燃油压力波形相对容易些,以燃油压力分析为基础,可以完成对供油系统的状态监测,判断供油系统的技术状况、典型的故障和异常喷射等。图 6-20(a),(b) 分别为 4135 柴油机正常工况下与喷油压力过高时燃油压力波形,与正常喷射压力波形相比,喷油压力过高时的波形结构不但出现二次喷射,三次压力峰值也较大,而且供油提前角与喷油提前角均有所减小。另外,在实验中,还先后人为设置了多种供油系统的典型故障,对出油阀密封锥面磨损、喷油器针阀磨损、喷油器针阀卡紧,喷油压力改变、供油提前角变化、供油量变化及高压油管接头处漏油等十种故障状态进行了监测和诊断,从获取的压力波形中看到,不同的故障类型,压力波形的结构、供油提前角、喷油提前角、最大喷射压力等均有较大变化。所以根据油压波形的结构特点能够分辨供油系统工作正常与否,同时也为进一步进行智能诊断提供了必要的实验依据。

另外从图中还可以看出,不管是哪种运行状态,喷油泵端与喷油器端的油压波形差别较大,这是由于在大幅度变化的压力作用下,燃油存在可压缩性,而且高压油管富有弹性,使高压

系统形成一个弹性系统,燃油在高压系统中的流动也就产生弹性振动。在供油过程中,当出油阀开启时,高压油管中喷油泵端燃油产生的压力波向喷油器端传播,如果不足以升起针阀,则压力波全部被反射,向喷油泵端传播,与该处新产生的压力波叠加起来,又被反射,向喷油器端传播。当压力传播使喷油器端燃油压力升高到大于针阀开启压力时,针阀即打开,喷油开始,此时传至喷油器端的压力波仍有部分被反射回去。所以在整个供油过程中,压力波往复传播多次反射,高压油管中的压力也就随时间和地点而变。由于上述的压力波动现象存在,使实际喷油过程与柱塞的供油过程很不一致。

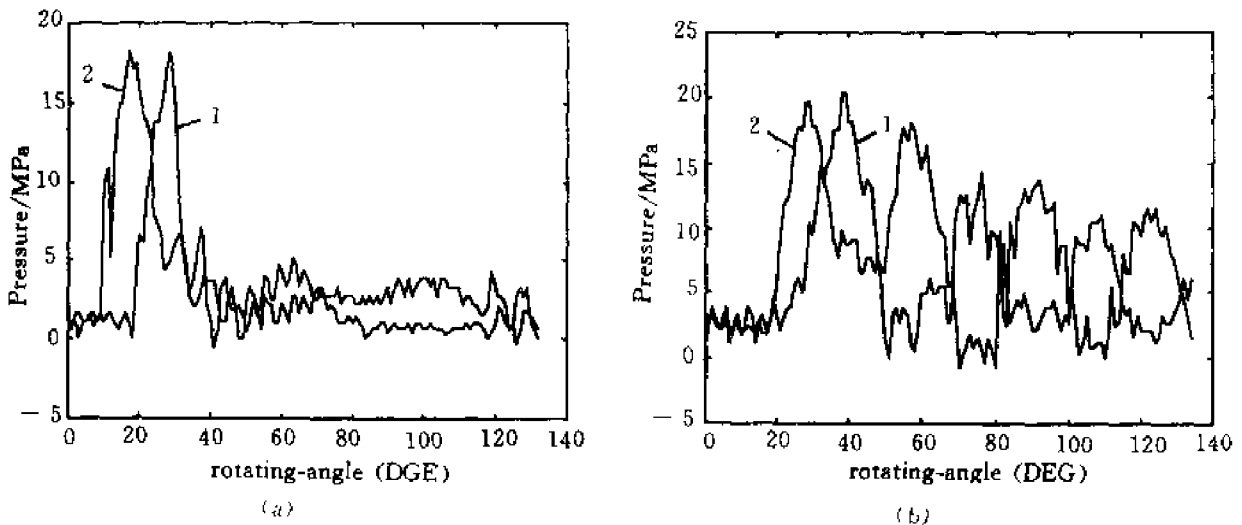


图 6-20 燃油压力波动分析

(a) 供油系统正常状态燃油压力波形 (b) 喷油压力过高时燃油压力波形

1—喷油器入口压力波形,2—出油阀出口压力波形,40°—上止点位置

习题与思考题

- 6-1 对往复机械进行性能检测的目的是什么?
- 6-2 油样光谱分析方法的原理是什么?
- 6-3 试述原子发射光谱仪的组成及其工作原理。
- 6-4 在机械故障诊断中,光谱分析可提供哪些信息?
- 6-5 常用油样铁谱分析仪器有哪些?其原理如何?
- 6-6 油样铁谱分析的诊断指标有哪些?各自的含义是什么?
- 6-7 如何应用油样分析方法进行故障诊断?
- 6-8 往复机械振动特征与旋转机械有何不同?原因何在?
- 6-9 用振动信号作为监测参数对往复机械进行故障诊断的方法有哪些?如何应用?
- 6-10 往复机械与旋转机械相比,故障诊断的难点何在?

第7章

典型传动部件的故障及其诊断方法

在机械设备中,机器的动力传递要靠传动部件来实现,传动部件的状态如何直接影响着整机的功能。因此,对其运行状态进行监测和故障诊断具有十分重要的意义。本章将主要对滚动轴承和齿轮装置的故障及诊断方法进行讨论。

7.1 滚动轴承异常的基本形式

滚动轴承是机械设备中最常见和最易损坏的部件之一。在使用过程中的异常形式主要有以下几种:

1. 疲劳剥落

这是滚动轴承常见的一种异常形式。在滚动轴承中,滚道和滚动体表面既承受载荷,又相对滚动。由于交变载荷的作用,首先在表面一定深度处形成裂纹,继而扩展到使表层形成剥落坑,最后发展到大片剥落。这种疲劳剥落现象造成了运行时的冲击载荷,使振动和噪声加剧。造成疲劳剥落的主要原因是疲劳应力。

2. 磨损

这是滚动轴承另一种常见的异常形式。轴承滚道、滚动体、保持架、座孔或安装轴承的轴颈,由于机械原因及杂质异物的侵入引起表面磨损。磨粒的存在是轴承磨损的基本原因,润滑不良会使磨损加剧。磨损导致轴承游隙增大,表面粗糙,降低机器运行精度,增大振动和噪声。

3. 塑性变形

轴承因受到过大的冲击载荷、静载荷、落入硬质异物等在滚道表面上形成凹痕或划痕,而且一旦有了压痕,压痕引起的冲击载荷会进一步使邻近表面剥落,这样,载荷的累积作用或短时超载就有可能引起轴承塑性变形。

4. 腐蚀

润滑油、水或空气水分引起表面锈蚀(化学腐蚀)、轴承内部有较大电流通过造成的电腐蚀、以及轴承套圈在座孔中或轴颈上微小相对运动造成的微振腐蚀等造成了轴承零件表面的腐蚀。

5. 断裂

载荷过大或疲劳常引起轴承零件破裂。热处理、装配引起的残余应力,

运行时的热应力过大也会引起轴承零件的裂纹或破裂。

6. 胶合

胶合指滚道和滚动体表面由于受热而局部融合在一起的现象。常发生在润滑不良、高速、重载、高温、起动加速度过大等情况下。由于摩擦发热，轴承零件可以在极短时间内达到很高的温度，导致表面烧伤，或某处表面上的金属粘附到另一表面上。

7.2 滚动轴承的振动类型及其故障特征

在工作过程中，滚动轴承的振动通常分为两类：其一为与轴承的弹性有关的振动，其二为与轴承滚动表面的状况（波纹、伤痕等）有关的振动。前者与轴承的异常状态无关，而后者反映了轴承的损伤情况。

滚动轴承在运转时，滚动体在内、外圈之间滚动。如果滚动表面损伤，滚动体在损伤表面转动时，便产生一种交变的激振力。由于滚动表面的损伤形状是无规则的，所以激振力产生的振动，将是由多种频率成份组成的随机振动。从轴承滚动表面状况产生振动的机理可以看出，轴承滚动表面损伤的形态和轴的旋转速度，决定了激振力的频率；轴承和外壳，决定了振动系统的传递特性。因此，振动系统的最终振动频率，由上述二者共同决定。也就是说，轴承异常所引起的振动频率，由轴的旋转速度、损伤部分的形态及轴承与外壳振动系统的传递特性所决定。

通常，轴的旋转速度越高，损伤越严重，其振动的频率就越高；轴承的尺寸越小，其固有振动频率越高。因此，轴承异常所产生的振动，对所有的轴承都没有一个共同的特定频率；即使对一个特定的轴承，当产生异常时，也不会只发生单一频率的振动。

为了加深对滚动轴承振动特性的认识，下面对其振动情况作进一步分析。

7.2.1 滚动轴承的固有振动频率

滚动轴承在工作时，滚动体与内环或外环之间可能产生冲击面引起轴承各元件的固有振动。各轴承元件的固有频率与轴承的外形、材料和质量有关，而与轴的转速无关。钢球的固有频率为

$$f_m = \frac{0.424}{r} \sqrt{\frac{E}{2\rho}} \quad (7.1)$$

式中， r 为钢球的半径(m)， ρ 为材料密度(kg/m³)， E 为钢球的弹性模量(N/m²)。

当滚动轴承为钢材时，其内外环的固有频率计算公式为

$$f_{(\omega)n} = 9.40 \times 10^5 \times R^2 \sqrt{\frac{EI}{M}} \times \frac{n(n^2 - 1)}{\sqrt{n^2 + 1}} \quad (7.2)$$

式中， R 为圆环回转轴到中性轴的半径(m)， n 为节点数， I 为圆环截面绕中性轴的惯性矩(m⁴)， M 为圆环单位长度内的质量(kg/m)。

需要注意的是，轴承元件的固有频率值，要受安装状态的影响，以上仅为轴承元件在自由状态下的计算公式。一般情况下，滚动轴承的固有振动频率通常可达数千赫到数十千赫，是非常高的振动频率。

7.2.2 滚动轴承的缺陷特征频率

图 7-1 为滚动轴承元件运动分析简图。图 7-1(a) 为外环固定而内环随轴转动的情况；图

7-1(b) 为内环固定而外环旋转的情况。

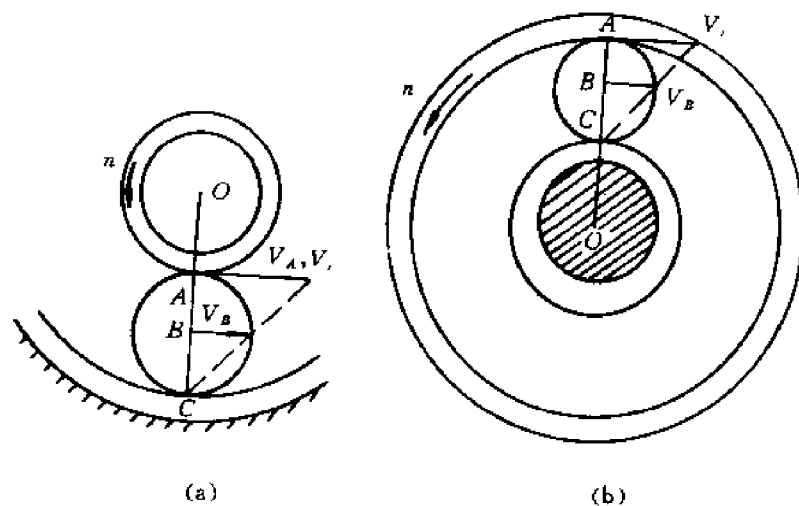


图 7-1 滚动轴承元件运动分析

为便于推导轴(轴承)旋转时运动元件缺陷的特征频率,作如下假设:

- (1) 滚动体与滚道之间无滑动接触;
- (2) 每个滚动体直径相同,且均匀分布在内外滚道之间;
- (3) 径向、轴向受载荷时各部分无变形。

1. 不受轴向力时轴承缺陷特征频率

(1) 外环固定内环随轴转动时单个滚动体(或保持架)相对于外环的旋转频率:从图 7-1(a)可知,内环滚道的切线速度为

$$V_i = \pi D_i f_r = \pi (D_m - d) f_r \quad (7.3)$$

式中, f_r 为轴的旋转频率, d 为滚动体的直径, D_i 为内环滚道的直径, D_m 为轴承滚道节径,即内外滚道的平均值。

因为滚动体滚而不滑,所以滚动体与内环滚道接触点 A 的速度为

$$V_A = V_i$$

又因外环固定,所以滚动体与接触点 C 的速度为

$$V_C = 0$$

而滚动体中心 B 的速度(即保持架的速度)为

$$V_B = \frac{1}{2} V_A = \frac{\pi}{2} (D_m - d) f_r \quad (7.4)$$

单个滚动体(或保持架)相对于外环的旋转频率为

$$f_{B_e} = \frac{V_B}{l_m} = \frac{\frac{\pi}{2} (D_m - d) f_r}{\pi D_m} = \frac{1}{2} \left(1 - \frac{d}{D_m}\right) f_r \quad (7.5)$$

其中 l_m 为滚道节圆周长。

(2) 内环固定外环旋转时,单个滚动体(或保持架)相对于内环的旋转频率:若外环的旋转频率仍为 f_r ,则保持架相对内环的切向速度从图 7-1(b)可知为

$$V_B = \frac{1}{2}V_A = \frac{\pi}{2}(D_m + d)f_r \quad (7.6)$$

单个滚动体(或保持架)相对于内环的旋转频率为

$$f_{B_i} = \frac{V_B}{l_m} = \frac{\frac{\pi}{2}(D_m + d)f_r}{\pi D_m} = \frac{1}{2}\left(1 + \frac{d}{D_m}\right)f_r \quad (7.7)$$

(3) 内外环均转动:若内外环相对转动频率仍为 f_r ,则当内外环同向旋转时两者相对转动频率等于内外环转动频率之差;反向旋转时为两频率之和。

(4) 轴承内外环有一缺陷时的特征频率:如果内环滚道上某一处有缺陷时,则 Z 个滚动体滚过该缺陷时的频率为

$$f_i = f_{B_i}Z = \frac{1}{2}\left(1 + \frac{d}{D_m}\right)f_r Z \quad (7.8)$$

如果外环道上某一处有缺陷时,则 Z 个滚动体滚过该缺陷时的频率为

$$f_o = f_{B_o}Z = \frac{1}{2}\left(1 - \frac{d}{D_m}\right)f_r Z \quad (7.9)$$

(5) 单个滚动体某处有缺陷时的特征频率:单个滚动体相对于外环的转动频率为

$$f_n = f_{B_o} \frac{\pi(D_m + d)}{\pi d} = \frac{1}{2}\left(1 - \frac{d^2}{D_m^2}\right)f_r \frac{D_m}{d} \quad (7.10)$$

单个滚动体相对于内环的旋转频率为

$$f_r = f_{B_i} \frac{\pi(D_m - d)}{\pi d} = \frac{1}{2}\left(1 - \frac{d^2}{D_m^2}\right)f_r \frac{D_m}{d} \quad (7.11)$$

可见 $f_n = f_r$ 。

如果单个有缺陷滚动体每自转一周只冲击内环滚道(或外环滚道)一次,则其频率为

$$f_{Ri} = f_{Ri} = f_n = \frac{1}{2}\left(1 - \frac{d^2}{D_m^2}\right)f_r \frac{D_m}{d} \quad (7.12)$$

如果该滚动体每自转一周冲击内外滚道各一次,则频率加倍,即

$$f_{Ri} = \left(1 - \frac{d^2}{D_m^2}\right)f_r \frac{D_m}{d} \quad (7.13)$$

如果滚动体是滚珠,其运转中不但有公转和自转,还会发生摇摆,因此,滚珠表面缺陷对滚道有时产生冲击,有时无冲击,从而会产生断续性故障信号。

(6) 保持架与内环或外环某处发生碰撞的频率:

$$\text{保持架碰外环的频率为} \quad f_{B_o} = \frac{1}{2}\left(1 - \frac{d}{D_m}\right)f_r \quad (7.14)$$

$$\text{保持架碰内环的频率为} \quad f_{B_i} = \frac{1}{2}\left(1 + \frac{d}{D_m}\right)f_r \quad (7.15)$$

2. 承受轴向力时轴承缺陷特征频率

由于滚珠轴承具有相当大的间隙,在承受轴向力时就会形成图7-2所示的状态:轴承内外环轴向相互错开,滚珠与滚道的接触点由 A, B 点移到 C, E 点。

此时,虽然轴承的节径不变,但内滚道的工作直径变大,外滚道的工作直径变小,就是说滚珠的工作直径由 d 变为 $d \cos \alpha$ 。在不受轴向力的轴承缺陷特征频率的计算公式中,只与滚珠直径和轴承节径两个物理量有关,因此,只须将不受轴向力时轴承缺陷特征频率计算公式中的滚

珠直径 d 用 $d \cos \alpha$ 代替, 便可得到受轴向力时轴承缺陷特征频率。这样, 受轴向力时轴承缺陷特征频率就具有形式为

$$\text{内滚道缺陷} \quad f_i = \frac{1}{2} \left(1 + \frac{d}{D_m} \cos \alpha \right) f_r Z \quad (7.16)$$

$$\text{外滚道缺陷} \quad f_o = \frac{1}{2} \left(1 - \frac{d}{D_m} \cos \alpha \right) f_r Z \quad (7.17)$$

$$\text{滚珠缺陷} \quad f_{rs} = \frac{1}{2} \left(1 - \frac{d^2}{D_m^2} \cos^2 \alpha \right) f_r \cdot \frac{D_m}{d} \quad (7.18)$$

$$\text{保持架碰外环} \quad f_{bs} = \frac{1}{2} \left(1 - \frac{d}{D_m} \cos \alpha \right) f_r \quad (7.19)$$

$$\text{保持架碰内环} \quad f_{bi} = \frac{1}{2} \left(1 + \frac{d}{D_m} \cos \alpha \right) f_r \quad (7.20)$$

式中, α 为轴承的压力角。

例 7-1 某 204 滚珠轴承, 节圆直径 $D_m = 33.5 \text{ mm}$, 滚珠直径 $d = 7.938 \text{ mm}$, 滚珠数 $Z = 8$, 轴转速 $n = 3000 \text{ r/min}$, 压力角 $\alpha = 0^\circ$ 。如果每次只出现一处缺陷, 试计算缺陷的重复频率。

【解】

$$\text{内滚道缺陷} \quad f_i = \frac{1}{2} \left(1 + \frac{d}{D_m} \cos \alpha \right) f_r Z = \frac{1}{2} \times 50 \left(1 + \frac{7.938}{33.5} \right) \times 8 = 247.39 \text{ Hz}$$

$$\text{外滚道缺陷} \quad f_o = \frac{1}{2} \left(1 - \frac{d}{D_m} \cos \alpha \right) f_r Z = \frac{1}{2} \times 50 \left(1 - \frac{7.938}{33.5} \right) \times 8 = 152.61 \text{ Hz}$$

$$\text{滚珠缺陷} \quad f_{rs} = \frac{1}{2} \left(1 + \frac{d^2}{D_m^2} \cos^2 \alpha \right) f_r \cdot \frac{D_m}{d} = \frac{1}{2} \times 50 \left(1 - \frac{7.938^2}{33.5^2} \right) \times \frac{33.5}{7.938} = 99.58 \text{ Hz}$$

$$\text{保持架碰外环} \quad f_{bs} = \frac{1}{2} \left(1 - \frac{d}{D_m} \cos \alpha \right) f_r = \frac{1}{2} \times \left(1 - \frac{7.938}{33.5} \right) = 19.08 \text{ Hz}$$

$$\text{保持架碰内环} \quad f_{bi} = \frac{1}{2} \left(1 + \frac{d}{D_m} \cos \alpha \right) f_r = \frac{1}{2} \times \left(1 + \frac{7.938}{33.5} \right) = 30.92 \text{ Hz}$$

7.2.3 滚动轴承的振动及其故障特征

1. 滚动轴承的时域波形特征

(1) 正常情况下, 滚动轴承的振动时域波形如图 7-3 所示; 从图中可以看出, 其波形有两个特点: 一是无冲击, 二是变化慢。

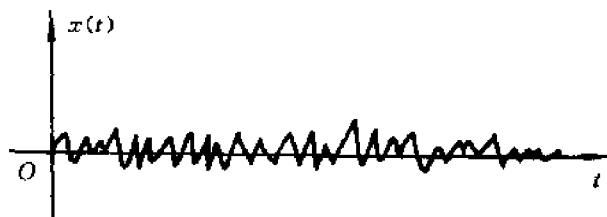


图 7-3 正常轴承的振动波形

(2) 轴承元件发生异常时, 就会产生冲击脉冲振动; 冲击脉冲周期为基阶故障特征频率的倒数。冲击脉冲宽度在 μs 数量级, 它将激起系统或结构的高频响应(固有振动), 响应水平取决

于系统或结构的固有频率及阻尼的大小。图7-4为轴承损伤引起的振动响应及其频率。每组图中,上图为损伤引起的冲击脉冲,中间图形为冲击脉冲产生的振动,下图反映了损伤的特征频率。

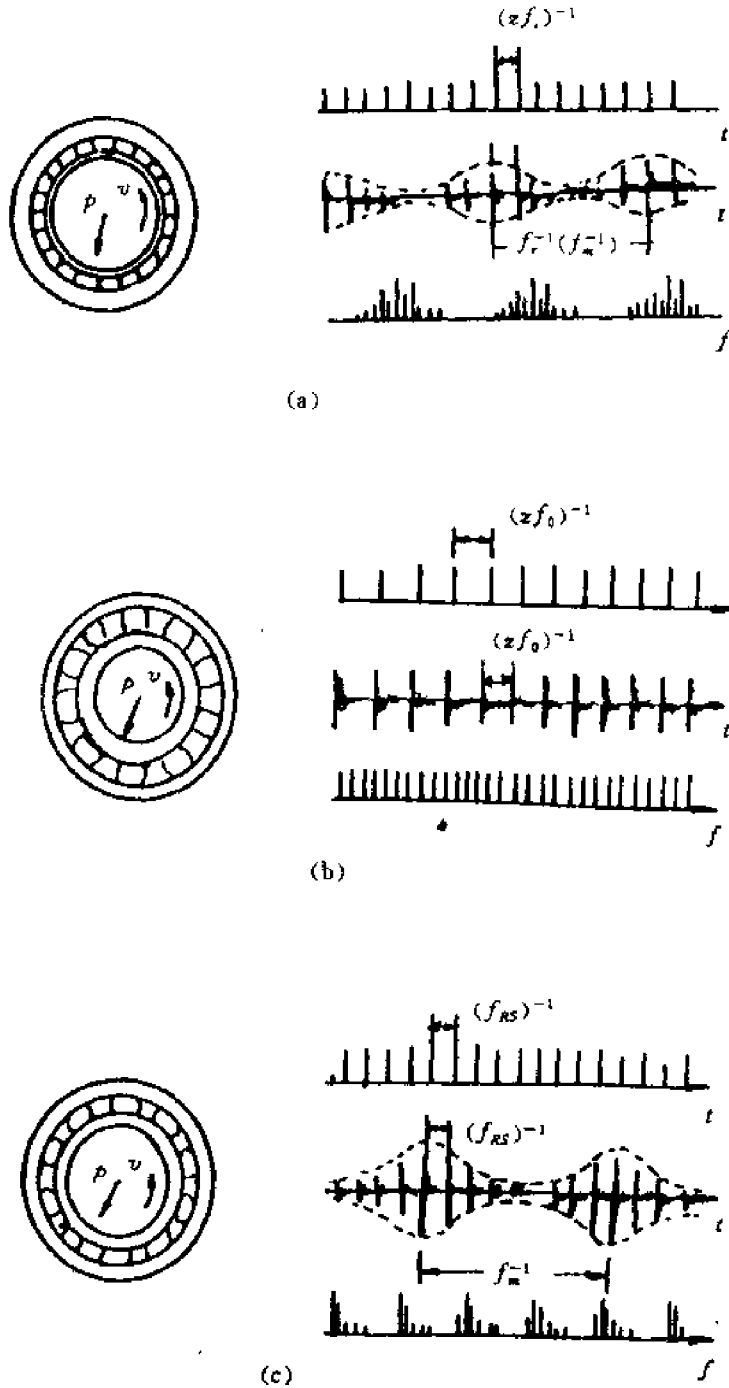


图7-4 滚动轴承损伤引起的振动波形
(a) 内环损伤 (b) 外环损伤 (c) 滚动体损伤

对于内环,由于轴承径向间隙的存在,根据损伤和滚动体发生冲击的位置,响应振幅会受到调制,其调制波的频率为轴的旋转频率 f_r 和振动体的公转频率 f_m ,其中最主要的还是通过

频率对固有频率的脉冲调制,如图 7-4(a) 所示。

对于外环,由于损伤处的位置与承载方向之间的位置关系是一定的,故无振幅调制现象,如图 7-4(b) 所示。

对于滚动体,它与内环情况相当,其调制波的频率为 f_m ,如图 7-4(c) 所示。

(3) 轴承偏心引起的振动:当轴承内环严重磨损或开裂时,轴的中心(即内环中心)便以外环中心为中心作振动,振动的频率成份为 nf_r ($n = 1, 2, \dots$)。

(4) 滚动体的非线性伴生振动:滚动轴承靠滚道与滚动体的弹性接触来承受载荷,因此具有“弹簧”的性质。这个“弹簧”的刚性很大,当润滑状态不良时,就会出现非线性弹簧性质的振动。轴向非线性伴生振动频率为轴的旋转频率 f_r ,分数谐波 $\frac{1}{2}f_r, \frac{1}{3}f_r, \dots$ 及其高次谐波 $2f_r, 3f_r, \dots$;而径向非线性伴生振动是 Zf_r 的各次谐波及 f_r 的分数谐波成分。

(5) 不同轴引起的振动:当两个轴承不对中,轴承装配不良等都会引起低频振动。

2. 幅值域中的概率密度特征

图 7-5 为滚动轴承正常时和发生剥落损伤时轴承振动信号的幅值概率密度分布。从图中可以看出,轴承发生剥落时,幅值分布的幅度广,这是由于存在剥落的冲击振动。这样,从概率密度分布的形状,就可以进行异常诊断。

显然,根据上述轴承的各种振动特征,不但有可能判别运转中的轴承是否已出现故障,而且可进一步判断故障的类型及故障发生的元件。

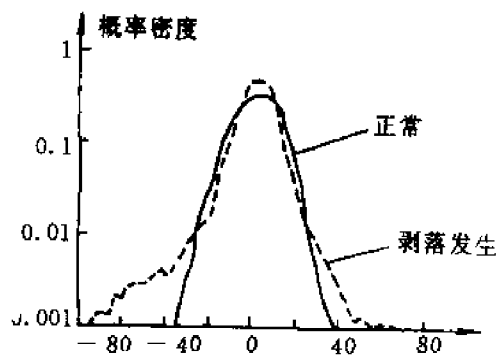


图 7-5 轴承振动的概率密度分布

7.3 滚动轴承故障诊断方法

目前,用于滚动轴承监测和诊断的方法很多,本节主要讨论利用振动信号对其进行监测的方法。并对利用光纤监测技术和接触电阻诊断法对滚动轴承进行故障诊断的方法作以简单介绍。

7.3.1 振动诊断法

滚动轴承在工作过程中会产生各种各样的异常和损伤,多数故障都会使轴承的振动加剧,这样,振动信号就成为诊断轴承故障的主要信息。采用振动诊断法主要有以下优点:

- (1) 可以检测出各种类型轴承的异常现象;
- (2) 在故障初期就可发现异常,并可在旋转中测定;
- (3) 由于振动信号发自轴承本身,所以不需特别的信号源;
- (4) 信号检测和处理比较简单。

在滚动轴承的振动诊断中,较常用的诊断方法有如下几种。

1. 有效值和峰值判断法

滚动轴承振动的瞬时值随着时间在不断地进行变化,作为表现这种振动变化大小的方法

广泛地使用有效值。有效值是振动幅值的均方根值。

由于有效值是对时间的平均,所以对具有表面皱裂无规则振动波形的异常,其测定值的变动小,可给出恰当的评价。但是,对表面剥落或伤痕等具有瞬变冲击振动的异常是不适用的。这是由于冲击波峰的振幅大,但持续时间短,如作时间平均,则有无峰值的差异几乎表现不出来。对于这种形态的异常,可用峰值进行判断。

峰值是在某个时间内振幅的最大值。它对瞬时现象也可得出正确的指示值,特别对初期阶段轴承表面剥落,非常容易由峰值的变化检测出来。它对滚动体对保持架的冲击及突发性外界干扰、或灰尘等原因引起的瞬时振动比较敏感,所以比起有效值来,测定值的变化可能很大。

2. 峰值指标法

峰值指标是指峰值与有效值的比。如上所述,由表面剥落或伤痕引起的瞬时冲击振动,峰值比有效值的反映灵敏,使用峰值指标正是利用峰值的该性质。一般来讲,正常轴承振动的峰值指标约为 5,当轴承发生伤痕时,峰值指标有时会达到 10。所以用该方法也较容易对滚动轴承的异常作出判断。

该方法的最大特点是:由于峰值指标的值不受轴承尺寸、转速及负荷的影响,所以正常异常的判断可非常单纯地进行;此外,峰值指标不受振动信号的绝对水平所左右,所以传感器或放大器的灵敏度即使变动,也不会出现测定误差。这种方法对表面皱裂或磨损之类的异常,却几乎没有诊断能力。

3. 振幅概率密度分析法

前面,已利用概率密度分布对正常和有疲劳剥落的轴承进行了区分,但这种方法还需要定量化,因此,为把概率密度分布的幅度用量表示出来,用式(7.21)求出概率密度分布的陡度 R_k ,把异常的程度数量化,然后根据 R_k 的大小判断轴承异常情况。

$$R_k = \frac{\int_{-\infty}^{\infty} x^4 p(x) dx}{\sigma_x^4} \quad (7.21)$$

式中, x 为瞬时幅值; $p(x)$ 为概率密度函数; σ_x 为标准偏差。

一般来讲,对正常轴承, R_k 的大小约为 3;当剥落发生时, R_k 将变大。 R_k 与峰值指标类似,因其与轴承转速、尺寸、负荷等条件无关,因此使用起来比较单纯,对轴承好坏的判定非常简单。其缺点是对轴承表面皱裂、磨损等异常缺乏检测能力,它主要适用于轴承表面有伤痕的情况。

4. 时序模型参数分析法

时序模型参数分析法,是一种把轴承振动信号采样值看作一个时间序列,并建立数学模型,然后利用模型的参数对轴承故障进行诊断的一种方法。

常用的模型为自回归模型 $AR(m)$,假设根据观测值 $x_k (k = 1, 2, \dots, N)$ 建立的模型为

$$x_k = \varphi_1 x_{k-1} + \varphi_2 x_{k-2} + \dots + \varphi_m x_{k-m} + a_k \quad (7.22)$$

$$\sigma_a^2 = \frac{1}{N} \sum_{k=m+1}^N a_k^2 \quad (7.23)$$

自回归模型参数 $\varphi_j (j = 1, 2, \dots, m)$ 表征了被测系统的某些特性,如动态特性、信号的频率结构模式、信号的能量大小等。在建立模型时, φ_j 是通过残差方差 σ_a^2 最小面获得的。

如果轴承出现故障或工况发生变化,系统状态相应改变,那么表征系统特性的模型参数 φ_i 也会随之变化。这样,用来真实拟合描述系统的差分方程的阶次 m 也将相应发生变化。如果系统状态的改变不足以引起模型阶次的变化,则用原来的 φ_i 值来计算 a_k ,但 σ_e^2 值将会增大。可以看出,模型阶次 m 和残差方差 σ_e^2 集中地代表了系统的特性。

为了对不同工作条件下的滚动轴承故障进行有效地诊断,应把信号强度变化这一因素加以排除。为此,引入归一化残差方差指标。

$$\text{NRSS} = \frac{\sigma_e^2}{\sigma_x^2} \quad (7.24)$$

式中, σ_x^2 为观测数据 x_k 的方差。

弄清滚动轴承在正常和各种异常时 m 及NRSS的变化规律,就能对轴承的状态作出诊断。

5. 冲击脉冲法

当两个不平的表面相互撞击时,就会产生冲击波,即冲击脉冲。这个冲击脉冲的强弱反映了撞击的猛烈程度。基于这个原理,通过检测轴承滚动体与滚道的撞击程度,可以了解轴承的工作状态。

如果滚动轴承的某些元件有损伤,轴承工作时,这些零件在接触过程中就会发生机械冲击,产生冲击脉冲力的幅度变化极大。通过加速度传感器可以测得此冲击引起的高频衰减振动波形,从而可对滚动轴承的故障作出判断。

振动加速度的振幅大小是与异常程度成比例的,因此可以利用冲击波形的最大值 x_p ,或冲击波形的绝对平均值 $\mu_{|x|}$ 进行异常判断。当转速较低时(300 r/min以下),平均值很小,据此进行异常与否判断则很困难,因此用最大值进行诊断。有时也用 $x_p/\mu_{|x|}$ 来判断异常, $x_p/\mu_{|x|}$ 大表示轴承有损伤, $x_p/\mu_{|x|}$ 小则表示发生了润滑不良或磨损异常。

6. 包络法

包络分析法是利用包络检波和对包络谱的分析,根据包络谱峰来识别故障。

事实表明,当滚动轴承元件产生缺陷而在运行中引起脉动时,不但会引起轴承外圈及传感器本身产生高频固有振动,且此高频振动的幅值还会受到上述脉动激发力的调制。

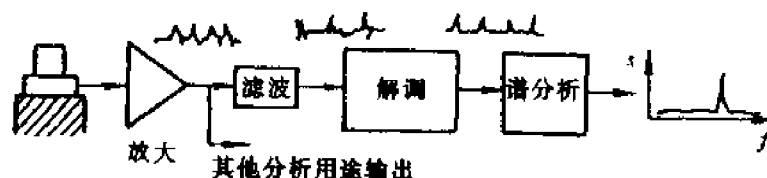


图 7-6 包络法原理框图

在包络法中(见图 7-6),将上述经调制的高频分量拾取,经放大、滤波后送入解调器,即可得到原来的低频脉动信号,再经谱分析即可获得功率谱。

包络法不仅可根据某种高频固有振动的出现与否,判断轴承是否异常,而且可根据包络信号的频率成分识别出产生故障的元件(如内圈、外圈,滚动体)。

包络法把与故障有关的信号从高频调制信号中解调出来,从而避免与其他低频干扰的混

滑,故有很高的诊断可靠性和灵敏度,是目前最常用、最有效地诊断滚动轴承故障的方法之一。

7. 高通绝对值频率分析法

将加速度计测得的振动加速度信号经电荷放大器后,再经过 1 kHz 的高通滤波器,只抽出其高频成分,然后将滤波后的波形作绝对值处理,再对经绝对值处理后的波形进行频率分析,即可判明各种故障原因。

图 7-7 为高通绝对值频率分析法的测试分析原理框图。图 7-7(c) 给出了振动波形绝对值处理结果。

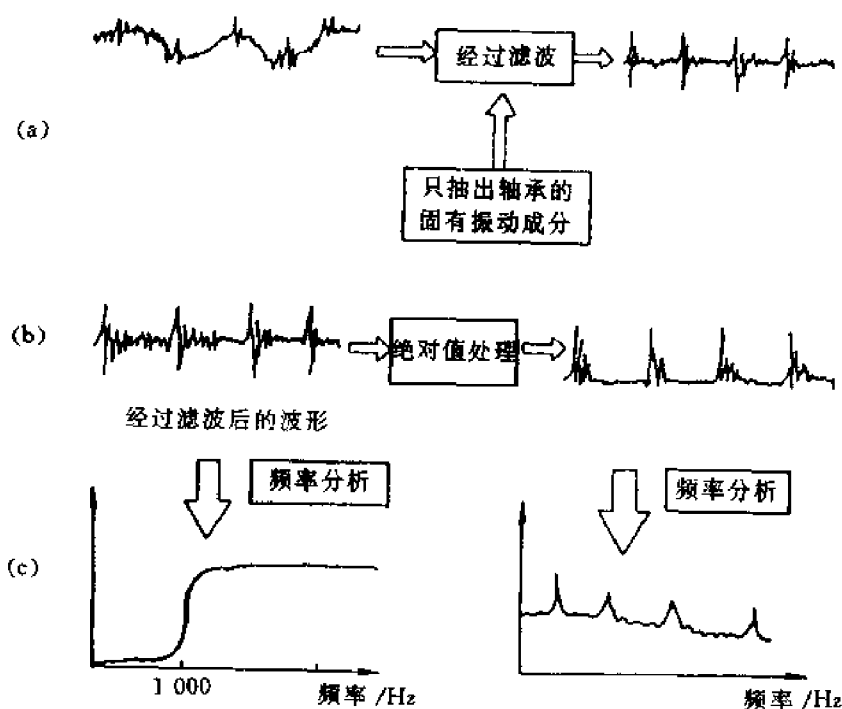


图 7-7 高频绝对值频率分析法的测试分析原理图

7.3.2 滚动轴承其他监测方法

1. 光纤监测技术

(1) 基本原理:前面所讨论的一些振动监测方法,通常是在轴承座上安装传感器,即用传感器测量轴承盖的振动信号。这样所检测的信号中完全接收了外界干扰,轴承的故障信号可能会因为较弱而被淹没。这里所讨论的光纤监测技术,则直接从轴承套圈的表面提取信号,其基本原理如图 7-8(a) 所示。

用光导纤维束制成的位移传感器包含有发送光纤束和接收光纤束,光线由发送光纤束经过传感器端面与轴承套圈表面的间隙反射回来,再由接收光纤束接收,经过光电元件转换为电压输出。间隙量 d 改变时,导光锥照射在轴承表面的面积也随之改变。传感器输出电压—间隙量特性曲线如图 7-8(b) 所示。

在图 7-8(b) 中,特性曲线开始有一段线性区,这是由于导光锥照射在轴承表面的面积越来越大,接收光纤束所接收的照度不断增大,直到达到峰值为止。此后,当间隙量进一步增大时,接收光纤所接收的照度与间隙量的平方成反比,其输出电压逐渐下降。

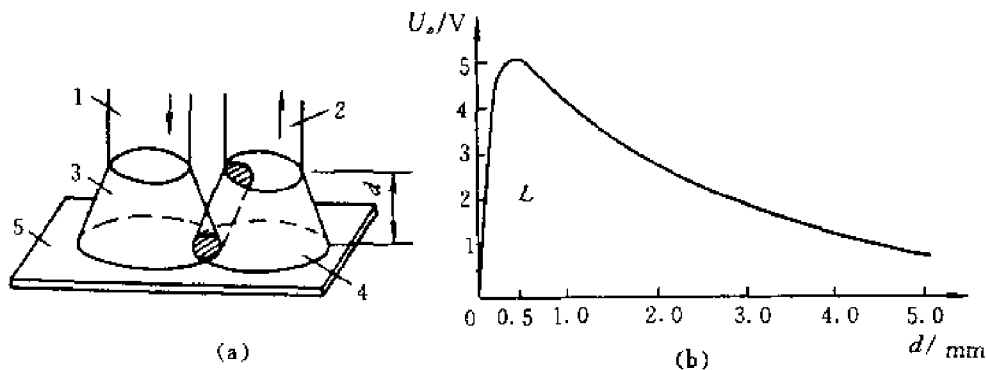


图 7-8 光纤式位移传感器的原理及特性曲线

1 发送光纤束;2 接收光纤束;3—导光锥;4—反光锥;5—轴承表面;L—前侧线性区

图 7-9 是发送光纤束和接收光纤束在传感器横截面中的布置方式。整个传感器由 600 根直径为 $60\ \mu\text{m}$ 的光导纤维组成,分为发送光束和接收光束。图 7-9(a) 中为两种纤维在横截面内为随机分布,图 7-9(b) 为相间分布,图 7-9(c) 则为圆环状分布。在这三种分布形式中,圆环状分布最为常用,等间隔分布最为灵敏,但制造起来最困难。

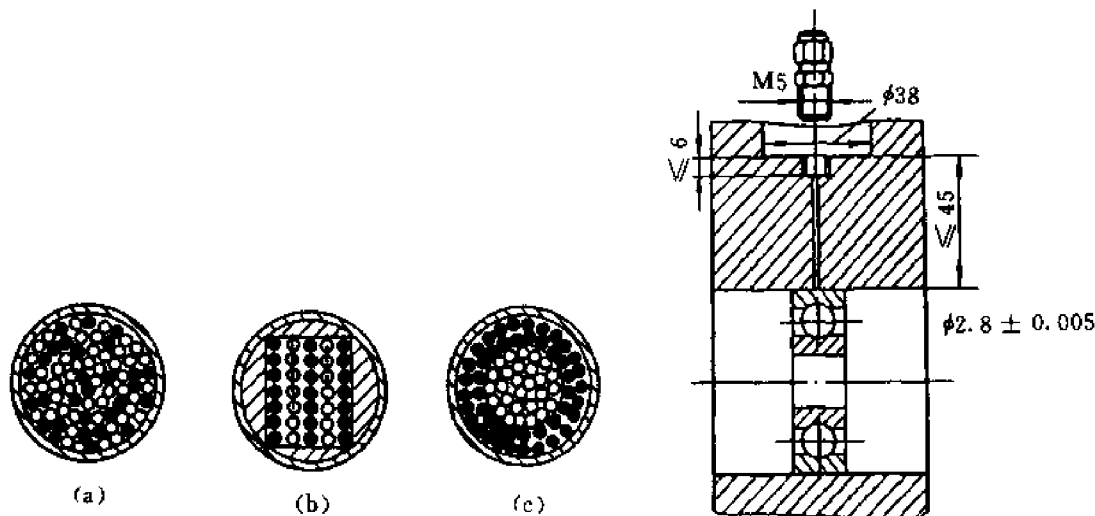


图 7-9 光纤传感器的横截面形式

●—发送光纤;○—接收光纤

图 7-10 光纤传感器的安装方法

由上所述可知,采用光纤传感器,可减小或消除振动传递通道的影响,从而提高信噪比。可以直接反映滚动轴承的制造质量,轴承工作表面磨损的程度,轴承的载荷,润滑和间隙的情况以及进行现场动平衡。另外,光导纤维位移传感器还具有高灵敏度(可达 $50\ \text{mV}/\mu\text{m}$ 的输出电压)、外形细长、便于安装的优点。图 7-10 所示是这种光纤传感器在轴承振动监测中的安装实例。

(2) 诊断指标:常采用的诊断指标包括有效值、峰值有效值比、轴承速率比。下面将分述这些指标的含义。

有效值 ψ_r :轴承由于其零件的制造缺陷,如表面粗糙度、波纹度和圆度误差形成不规则的

轮廓,运行时就会产生振动。这一振动由光纤传感器接收后,即可得图7-11所示的 ψ_x 脉动波形。图7-11(a)为一个接近理想的高精度电动机轴承形成的波形,其套圈的弹性变形接近简谐图形,其波数等于通过测点的钢球数目;图7-11(b)为精度级最低的轴承形成的复杂波形,这种轴承不但表面粗糙度大,几何形状误差大,而且钢球直径也有明显的不同。因此,可用光纤传感器直接在所使用的机器上测量轴承的质量,它是一种简单而有效的测试方法。

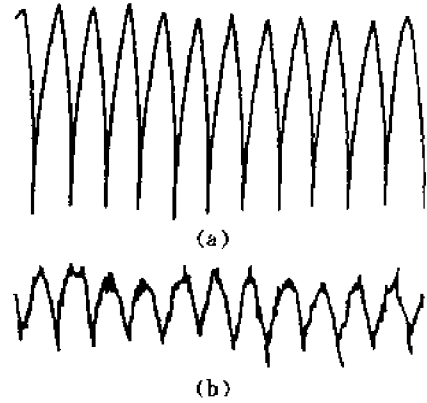


图7-11 均方根幅值的变化反映轴承制造质量的不同

峰值有效值比 x_P/ψ_x : 对于经过一段时间运行的滚动轴承,其工作表面会由于磨损而变得粗糙。虽然此时轴承表面粗糙状况也可用上述有效值指标来反映。但是,当轴承零件上有局部的剥落、凹坑一类缺陷时,有效值就无法反映出来。这时通过峰值有效值比则可以反映出来。一般来说,当 $x_P/\psi_x > 1.5$ 时,认为轴承零件上有局部缺陷产生。

轴承速率比 BSR: 定义为钢球通过频率与轴的回转频率之比。

BSR 值取决于轴承的载荷和间隙的大小以及轴承的润滑状况,图7-12为 BSR 值与轴向载荷的关系。图中的阴影部分是轴承正常工作时的 BSR 值。当 BSR 值偏高时,则可能是载荷过高、润滑不良或者轴承间隙过大;当 BSR 值偏低时,则可能是载荷不足、润滑过多(例如润滑油脂涂敷过多)或者轴承间隙过小。BSR 值可以说是机器中轴承运行性能的直接指标。

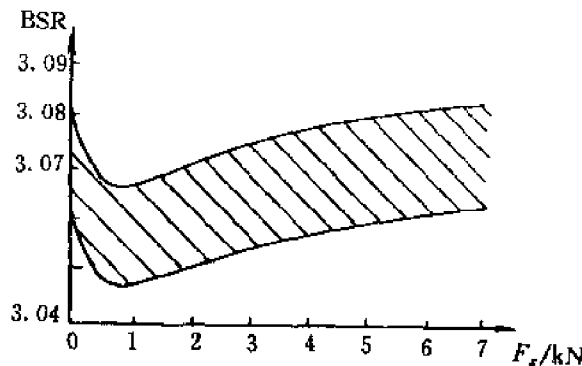


图7-12 滚动轴承的 BSR 值

由于载荷是由钢球与滚道传递的,当钢球通过滚道上的监测点时,滚道将以钢球的接触点为中心产生弹性变形区,这样,光纤传感器可以直接测量这一变形,从而确定钢球的通过频率。而轴的回转频率则需另外安装一位移传感器作为时标加以确定。

2. 接触电阻法

接触电阻法所依据的基本原理和振动测量完全不同,它是与振动监测法相互补充的一种监测技术。

轴承在运转过程中,滚道面与滚动体之间便会形成油膜,这样在内外圈之间就有很大的电阻,正常轴承,其油膜厚度至少是表面粗糙度的四倍,由于润滑剂是有机碳氢化合物,轴承内外圈间的平均电阻很高,一般在 $1 \sim 1 \times 10^6 \Omega$ 之间变化(见图7-13),而当轴承零件出现剥落、腐

蚀、裂纹或磨损时，油膜破坏，接触电阻下降。根据这一性质，可对轴承故障进行诊断。

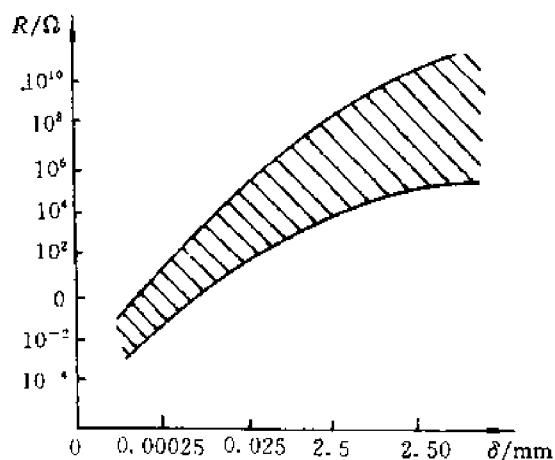


图 7-13 油膜厚度与接触电阻的关系

用接触电阻法监测轴承的故障时，需要在轴承上施加一微小的直流电压，大小约为 100 mV，测量轴承接触表面间的接触电阻（见图 7-14）。依靠图 7-14 所示的仪器和线路，可以求出在不同轴承缺陷下的接触电阻谱。

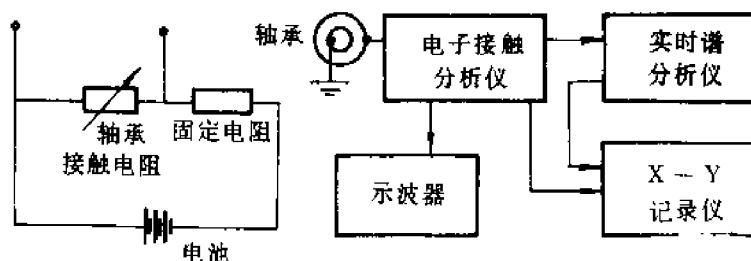


图 7-14 接触电阻法的原理

显然，振动监测和接触电阻监测二者对于不同的轴承缺陷敏感的程度是不一样的。振动监测法对剥落、凹坑比较敏感，而接触电阻法对磨损、腐蚀这一类缺陷比较敏感。两者是互相补充的。因此用两种方法测试具有不同内部缺陷的滚动轴承，所得的结果很可能导致相反的结论。

当前，滚动轴承运行的故障诊断和监测技术已发展的较完善。测振，测温，测磨损，测润滑剂各种方法都有应用。例如测温技术，可以采用温度计、热电偶、双金属片、易熔金属以及液晶和温敏涂料等各种手段。润滑剂检查可用目测油标、润滑油成分化验、流量监测以及铁谱、磁塞等多种方式进行。此外，在关键的部位，还可以用放射性元素测量磨损，等等。这里就不一一详细讨论了，其中有的方法和一般的测试技术没有很大的区别，可以参阅有关的文献。

7.4 齿轮异常的基本形式及振动特点

齿轮传动是机械设备中最常见的传动方式，现代机械对齿轮传动的要求日益提高，既要求

齿轮能在高速、重载、特殊介质等恶劣环境条件下工作,又要求齿轮装置具有高平稳性、高可靠性和结构紧凑等良好的工作性能,由此使得齿轮发生故障的因素愈来愈多,而齿轮异常又是诱发机器故障的重要因素。因此,齿轮故障诊断技术的应用研究是非常重要的课题。

7.4.1 齿轮异常的基本形式

齿轮由于制造、操作、维护以及齿轮材料、热处理、运行状态等因素不同,产生异常的形式也不同,常见的齿轮异常有以下几种形式。

1. 齿面磨损

润滑油不足或油质不清洁将造成齿面剧烈的磨粒磨损,使齿廓显著改变,侧隙加大,以至由于齿厚过度减薄导致断齿。一般情况下,只有在润滑油料中夹杂有直径在 $30\ \mu\text{m}$ 以上的磨粒时,齿轮才会在运行中引起齿面磨粒磨损。

2. 齿面胶合和擦伤

重载和高速的齿轮传动,齿面工作区温度很高,如润滑条件不好,齿面间的油膜破裂,一个齿面的金属会熔焊在与之啮合的另一个齿面上,在齿面上形成垂直于节线的划痕胶合。新齿轮未经跑合时,常在某一局部产生这种现象,使齿轮擦伤。

3. 齿面接触疲劳

齿轮在实际啮合过程中,既有相对滚动,又有相对滑动,而且相对滑动的摩擦力在节点两侧的方向相反,从而产生脉动载荷。这两种力的作用使齿轮表面层深处产生脉动循环变化的剪应力,当这种剪应力超过齿轮材料的剪切疲劳极限时,接触表面将产生疲劳裂纹,继而裂纹扩展,最终使齿面剥落小块金属,在齿面上形成小坑,称其为点蚀。当“点蚀”扩大连成一片时,形成齿面上金属块剥落。此外,材质不均或局部擦伤,也易在某一齿上首先出现接触疲劳,产生剥落。

4. 弯曲疲劳与断齿

在运行过程中承受载荷的轮齿,如同悬臂梁,其根部受到脉冲循环的弯曲应力作用最大,当这种周期性应力超过齿轮材料的弯曲疲劳极限时,会在根部产生裂纹,并逐步扩展,当剩余部分无法承受外载荷时就会发生断齿现象。齿轮由于工作中严重的冲击、偏载以及材质不均也都可能引起断齿。断齿和点蚀是齿轮故障的主要故障模式。

凶轮异常还可分为局部的和分布的。前者集中于某个或几个齿上,后者分布在齿轮各轮齿上。

7.4.2 齿轮振动及其特点

1. 齿轮的振动频率

(1) 齿轮振动的类型:齿轮在运行过程中产生的振动是比较复杂的,由于齿轮所受的激励不同,从而使齿轮产生的振动类型也不同。下面分别介绍各种类型振动产生的原因及其特征:

第一,齿轮啮合过程中由于周节误差、齿形误差或均匀磨损等都会使齿与齿之间发生撞击,撞击的频率就是它的啮合频率。齿轮在此周期撞击力的激励下产生了以啮合频率为振动频率的强迫振动,频率范围一般在几百到几千赫内。

第二,由于齿轮啮合过程中轮齿发生弹性变形,使刚刚进入啮合的轮齿发生撞击,因而产生沿着啮合线方向作用的脉动力,于是也会产生以啮合频率为频率的振动。对于齿廓为渐开线的齿轮,在节点附近为单齿啮合,而在节点两侧为双齿啮合,故其刚度是非简谐的周期函数,所产生的强迫振动与上述第一种情况不同,不仅有以啮合频率为频率的基频振动,而且还有啮合频率的高次谐波振动。

第三,齿与齿之间的摩擦在一定条件下会诱发自激振动,主要与齿面加工质量及润滑条件

有关,自激振动的频率接近齿轮的固有频率。

第四,齿与齿之间撞击是一种瞬态激励,它使齿轮产生衰减自由振动,振动频率就是齿轮的固有频率,通常固有频率在 $1 \sim 10$ kHz 内。

第五,齿轮、轴、轴承等元件由于不同心、不对称、材料不均匀等会产生偏心、不平衡,其离心惯性力使齿轮轴系统产生强迫振动,振动的频率等于轴的转动频率(一般在 100 Hz 以内)及其谐频。

第六,由于齿面的局部损伤而产生的激励,其相应的强迫振动频率等于损伤的齿数乘以轴的转动频率。

综上所述,齿轮的振动频率基本上可归纳为三类:即轴的转动频率及其谐频,齿轮的啮合频率及其谐频,齿轮自身的各阶固有频率。而齿轮的实际振动往往是上述各类振动的某种组合,图 7-15 是齿轮几种常见工作状态下振动的时间历程曲线和相应的幅频谱图。从图中看到,不同状态下其时域和频域图形均有明显的区别。但这些振动曲线都是经过低通滤波后得到的,它只显示出其中频率较低的转动频率和啮合频率及它们的谐频,滤去了高频的自由衰减振动。实际上,齿轮的自由振动经由轴、轴承传到齿轮箱体时,高频冲击振动已衰减,犹如通过一个机械低通滤波器,因此在轴承座等处测得的振动信号,一般只包含转动频率与啮合频率及其谐频。

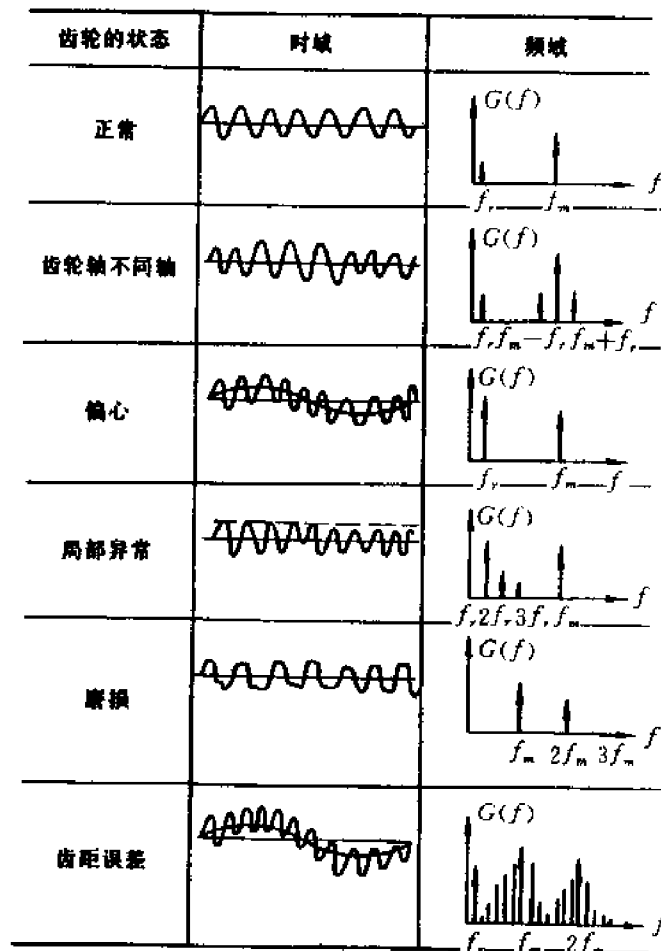


图 7-15 各种状态齿轮的振动(低频部分)

f_r —转动频率 f_m —啮合频率

齿轮箱各不同部件故障的振动特征,如表 7-1 所示。

表 7-1 齿轮箱故障的振动特征简表

部件	失效类型	振动频率	振幅特征	振动方向	其他
转子	失衡	f_r	随 f_r 增大 $f_r = f_s$ 时有峰值	径向	受悬臂式载荷时有轴向振动
轴	弯曲	$f_r, 2f_r$ 及 nf_r	随 f_r 增大	径向最大	
	截面扁平	$2f_r$	同上	径向	
联轴器	对中不良	$f_r, 2f_r$ 及 nf_r	变化不定	轴向较大	齿轮联轴节的振动特征基本上与齿轮相同,但 $f_r = f_s$ 时有峰值
	配合松	$f_r/n, f_r$ 或 nf_r	同上	径向	
	不平衡	f_r	同上	径向	
齿轮	齿面损伤	损伤齿数 $\times f_r$	随 f_r 增大	径向	磨损严重时出现高阶振动, f_r 的振动能量明显增大
	断齿	断齿数 $\times f_r \cdot f_s$	同上	径向	
滚动轴承	内圈剥落	$0.5nZ(1 + \frac{d}{D} \cos \alpha) f_r$	变化不定	径向	轴承的高频振动(10 ~ 60 kHz)不易传给其他部件
	外圈剥落	$0.5nZ(1 - \frac{d}{D} \cos \alpha) f_r$	同上	径向	
	钢球剥落	$n \frac{d}{D} [1 + (\frac{d}{D})^2 \cos^2 \alpha] f_r$	同上	径向	
滑动轴承	润滑不良	f_r	同上	径向	
	油膜涡动	$(0.42 - 0.48) f_r$	突变	径向	
	油膜振荡	f_s	同上	径向	
基础	翘曲(不平)	f_r, f_r 及 nf_r	随 f_r 增大	轴向较大	
	刚性不好	f_r	随 f_r 增大而减小	径向	
		f_r —轴的转动频率 f_s —轴的临界转动频率 f_s —齿轮的固有频率 a —轴承的压力角 n —自然数 1, 2, 3...		Z —轴承钢球数 d —轴承钢球直径 D —轴承平均直径	

(2) 齿轮各类振动频率的计算:

齿轮及轴的转动频率 f_r

$$f_r = \frac{N}{60} \text{ Hz} \quad (7.25)$$

式中, N 为齿轮及轴的转速 (r/min)。

齿轮的啮合频率 f_m ; 对于定轴转动的齿轮有

$$f_m = Z_i \frac{N_i}{60} \quad (7.26)$$

式中, Z_i 为第 i 个齿轮的齿数, N_i 为第 i 个齿轮的转速 (r/min)。

由式(7.26)可知一对啮合齿轮的啮合频率是相同的。

对于有固定齿圈的行星轮系有

$$f_m = Z_r(N_r \pm N_c)/60 \quad (7.27)$$

式中, Z_r 为齿轮的齿数; N_r 为该齿轮的转速(r/min); N_c 为转臂的转速(r/min), 当 N_r 与 N_c 转向相反时 N_c 取正号, 否则就取负号。

齿轮以啮合频率为振动频率振动的特点:

一是振动频率随齿轮的转速变化;
二是由于非线性的影响, 往往有啮合频率的高阶谐波振动;

三是随着转速升高, 振动能量增大, 噪声增强;

四是当啮合频率接近或等于齿轮的固有频率时, 齿轮发生共振形成强烈振动。

例 7-2 图 7-16 为某车床主轴齿轮箱, 当主轴转速为 1 000 r/min 时的传动简图。各传动轴的转动频率与各齿轮的啮合频率列于表 7-2 中, 可以看出这种频率的数值相差一个数量级或更多。

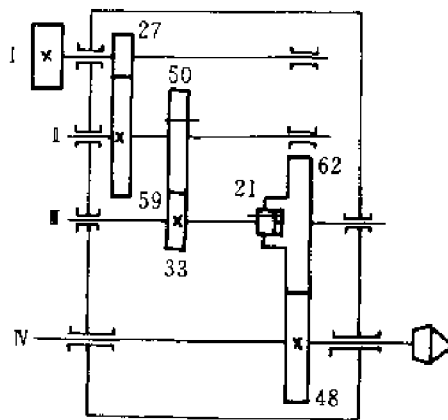


图 7-16 车床主轴齿轮箱简图

表 7-2 齿轮箱各轴的转动频率与齿轮的啮合频率

轴号	轴的转速 / (r · min ⁻¹)	轴转动频率 / Hz	啮合齿轮副	齿轮啮合频率 / Hz
I	1120	18.7	27/50	504
II	513	8.5	50/33	427
III	777	12.9	21/21	272
IV	1 000	16.7	62.48	802
计算公式	N_i	$f_r = \frac{N_i}{60}$	Z_1 / Z_2	$f_m = Z_1 f_n = Z_2 f_r$

齿轮的固有频率, 单个齿轮固有频率的计算及测定方法: 单个齿轮的固有频率是指齿轮轴向振动的固有频率。把齿轮近似地看作是周边自由、中间固定的圆板时, 其轴向振动的固有频率可近似地用下式计算

$$f_n = \frac{a_{NS}}{2\pi R^2} \sqrt{\frac{Et^3}{12(1-\mu^2)\rho_A}} \quad (7.28)$$

式中, R 为齿轮的分度圆半径; E 为齿轮材料的弹性模量; t 为齿轮厚度; μ 为齿轮材料的泊松比; ρ_A 为齿轮的单位面积质量; a_{NS} 为齿轮的振型常数(下标 N 为振型的径向节线数, S 为振型的节圆数, 如图 7-17 所示。

式(7.28)是将齿轮简化为圆板得到的, 用此式计算出来的固有频率有一定误差, 准确的数值需用实验测定。稳态激振法可同时测出齿轮的振型和相应的固有频率, 其方法如下: 将齿轮装到垂直固定的芯轴上, 在水平的齿轮表面撒上细砂粒, 沿轴向激振齿轮, 缓慢地提高激振频率。当齿轮发生共振时,

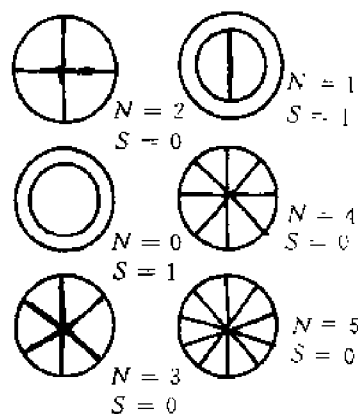


图 7-17 齿轮的几种振型

激振频率就是齿轮的固有频率,而齿轮上细砂料这时就逐渐移动到节圆和径向节线附近,形成明显的振型。齿轮的几种振型如图 7-17。任何一个齿轮都具有多种振型,而每种振型有各自的固有频率,因此每个齿轮都有多个固有频率。

例 7-3 某钢质齿轮其模数 $m = 5 \text{ mm}$, 齿数 $Z = 39$, 齿厚 $t = 10 \text{ mm}$ 。表 7-3 为振型常数 a_{NS} 的数值。同时用计算法与实验法求得其固有频率列于表 7-4 中。

表 7-3 振型常数 a_{NS} 的数值

$\begin{matrix} N \\ S \end{matrix}$	0	1	2	3	4	5	6
0	—	—	5.251	12.23	21.62	33.06	43.56
1	9.076	20.52	35.24	52.91	73.10	95.26	121.00
2	38.52	59.86	83.91	111.30	142.80	175	210.30
3	78.8	119.00	154.00				

表 7-4 某齿轮的固有频率

齿轮的振型		齿轮的固有频率 / Hz	
节线数 N	节圆数 S	计算值	实测值
2	0	1 218	1 160
0	1	2 106	2 160
3	0	2 837	2 760
1	1	4 761	4 520
4	0	5 016	4 640
5	0	7 670	6 800
6	0	10 106	9 080

从表 7-4 中可以看出,计算值都偏离实测值,其原因就在于用计算法时将齿轮简化为无内孔的圆板。显然,若齿轮的内孔相当大时,式(7.28)就不再适用。

单组齿啮合时齿轮的固有频率:齿轮啮合时的固有频率实际是指齿轮轴系扭转振动的固有频率。振动系统由啮合齿轮和传动轴组成,齿轮轴系的扭振包括轴的扭振和轮齿的弹性振动。轴的刚度系数约为 $(0.5 \sim 2) \times 10^6 \text{ kN/m}^2$,而齿轮轮齿的刚度系数约为 $(2 \sim 10) \times 10^7 \text{ kN/m}^2$,两者相差一个数量级。因此在分析轴的弹性变形引起的齿轮扭振时,可以认为轮齿是刚性的;而在分析由于轮齿弹性变形产生的载荷所引起的齿轮扭振时,就不考虑转动轴及其他零件的惯性载荷,认为转矩是一个常量。显然,与齿轮损伤信息有关的是轮齿弹性引起的振动,此时振动系统可认为由齿轮体(质量块)与轮齿(弹性体)构成,而齿的刚度周期性变化、齿面损伤及扭矩的变化等为齿轮扭振提供了激励。

直齿圆柱齿轮扭振的固有频率可用下式进行近似计算

$$f_n = \frac{1}{2\pi} \sqrt{\frac{K_t}{M_t}} \quad (7.29)$$

式中, K_t 为齿轮副的等效刚度系数; M_t 为齿轮副的等效质量。

$$\frac{1}{K} = \frac{1}{K_G} + \frac{1}{K_P} \quad (7.30)$$

式中, K_G 为大齿轮的刚度系数; K_P 为小齿轮的刚度系数。

K_c 与 K_p 的数值由图 7-18 根据齿轮的齿数 Z 和修正系数 x 来确定。用式(7.30) 求出 K 后,再由图 7-19 根据啮合(重叠)系数和 K 值求出齿轮副的等效刚度系数 K_e 。

$$\frac{1}{M_e} = \frac{1}{M_G} + \frac{1}{M_P} \quad (7.31)$$

式中, M_e 为齿轮副的等效质量; M_G 为大齿轮的等效质量; M_P 为小齿轮的等效质量。

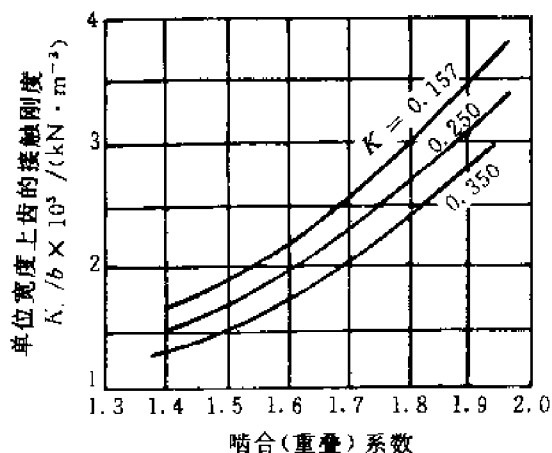


图 7-18 齿轮的刚度变化曲线

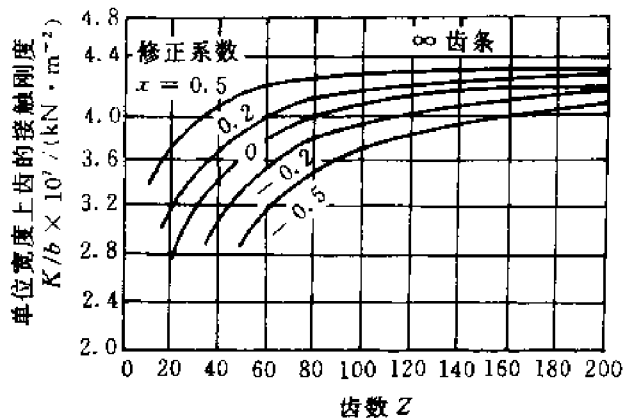


图 7-19 齿轮副的等效刚度变化曲线

M_G 与 M_P 的数值计算公式与齿轮的形状有关,其中最常见两种齿轮的计算公式为齿轮的形状如图 7-20(a) 时

$$M_i = (1.82mZ_i)^2 b_i \times 10^{-6} \text{ kg}$$

齿轮的形状如图 7-20(b) 时

$$M_i = M_w + M_a$$

$$M_w = (1.82mZ_i)^2 b_i \times 10^{-6} \text{ kg}$$

其中

$$M_a = \frac{\pi\rho}{4} \times \frac{(b_a - b_i)^2}{r_g^2} r_a^2 \text{ kg}$$

式中, i 为 G, P ; Z 为齿数; ρ 为材料的密度; m 为齿轮的模数。

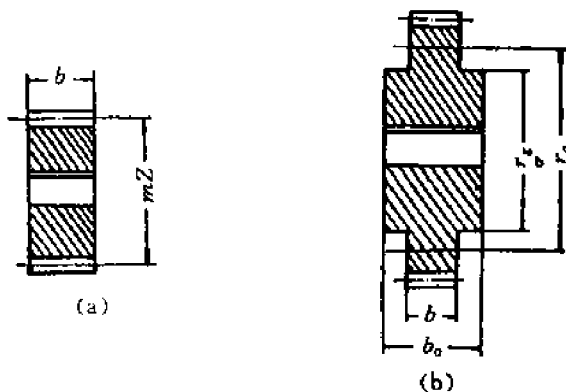


图 7-20 两种常见齿轮的示意图

2. 齿轮振动频谱的特点

(1) 齿轮振动的边频带谱: 齿轮在运行中无论齿轮发生异常与否, 齿的啮合都会发生冲击啮合振动, 其振动波形表现出振幅受到调制的特点, 既调幅又调频。图 7-21 为某齿轮箱的振动频谱。从图中可以看出齿轮振动的频谱图是非常复杂的, 除了有明显表示啮合频率的谱线 ① 和啮合频率的高阶谐频的谱线 ②、③、④ 外, 还有许多按一定规律分布的小谱线, 这就是在齿轮振动的频谱图中常见的边频带谱。它是由于几种动载同时作用在齿轮上, 使其同时产生几种振动, 它们相互叠加产生调制的结果。通常载频为啮合频率及其高阶谐频或其他高频成分, 而轴的转动频率及其高阶谐频则为调制频率。齿轮发生故障时, 啮合频率的振动成分或轴的转动频率及谐频的振动成分随之发生显著的变化。因此, 如果能从齿轮振动的边频带谱中分析出载频或调制信号的频率, 对于分析齿轮的故障将十分有用。

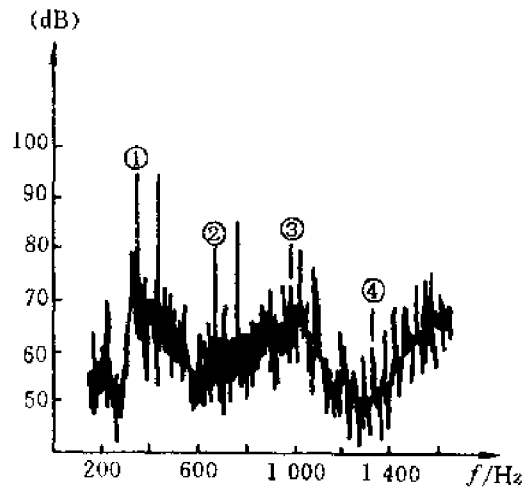


图 7-21 齿轮箱的振动频谱

(2) 振动信号的调制原理: 齿轮振动信号既有幅值调制又有频率调制, 这两种调制在频谱图中均表现为在啮合频率及其谐频的两侧各有一簇边频带, 各边频带的间隔即是调制信号的频率。下面以幅值调制为例简述产生边频带的原理。

幅值调制就是载频的时域信号幅值受到调制信号的调制。若载频是啮合频率 f_m , 其时域信号用 $x(t)$ 表示, 调制信号为余弦函数 $\cos 2\pi f_0 t$ 或正弦函数 $\sin 2\pi f_0 t$ (f_0 即是调制信号的频率) 则幅值调制信号为 $x(t)\cos 2\pi f_0 t$ 或 $x(t)\sin 2\pi f_0 t$ 。根据傅里叶变换的性质, 我们有

$$F[x(t)e^{j2\pi f_0 t}] = F[x(t)(\cos 2\pi f_0 t + j\sin 2\pi f_0 t)] = X(f_m - f_0) \quad (7.32)$$

$$F[x(t)e^{-j2\pi f_0 t}] = F[x(t)(\cos 2\pi f_0 t - j\sin 2\pi f_0 t)] = X(f_m + f_0) \quad (7.33)$$

$$\left. \begin{aligned} \text{式(7.32) + (7.33) 得 } F[x(t)\cos 2\pi f_0 t] &= \frac{1}{2}[X(f_m - f_0) + X(f_m + f_0)] \\ \text{式(7.32) - (7.33) 得 } F[x(t)\sin 2\pi f_0 t] &= \frac{1}{2j}[X(f_m - f_0) - X(f_m + f_0)] \end{aligned} \right\} \quad (7.34)$$

式(7.34)表明: 时域中的幅值调制(相乘)在频域中反映为频移, 即原载频的谱线 f_m 的位置移动了 f_0 , 而且由调前的一条谱线 f_m 变为两条谱线($f_m - f_0, f_m + f_0$), 分布在原谱线 f_m 的两侧, 形成边频带, 如图 7-22 所示。

齿轮振动中的调制信号一般不是正弦信号或余弦信号, 而是以一定周期变化的脉冲信号。从振动分析中得知, 脉冲信号可以看成是一系列简谐信号的组合。齿轮局部缺陷激发的是窄脉冲, 其频谱在较宽的频率范围内具有相等而较小的幅值; 若齿轮的缺陷是连续的, 则激发的脉冲较宽, 频谱的频带窄且幅值衰减较快。因此, 这两种缺陷的边频带有明显的差别: 前者的边频带范围宽而幅值小且变化平缓; 而后者的边频带窄, 集中于载频谱线附近, 幅值较大而衰减较快。这样, 就可根据这两种故障的时域波形或频域边频带加以区别, 如图 7-23 所示。

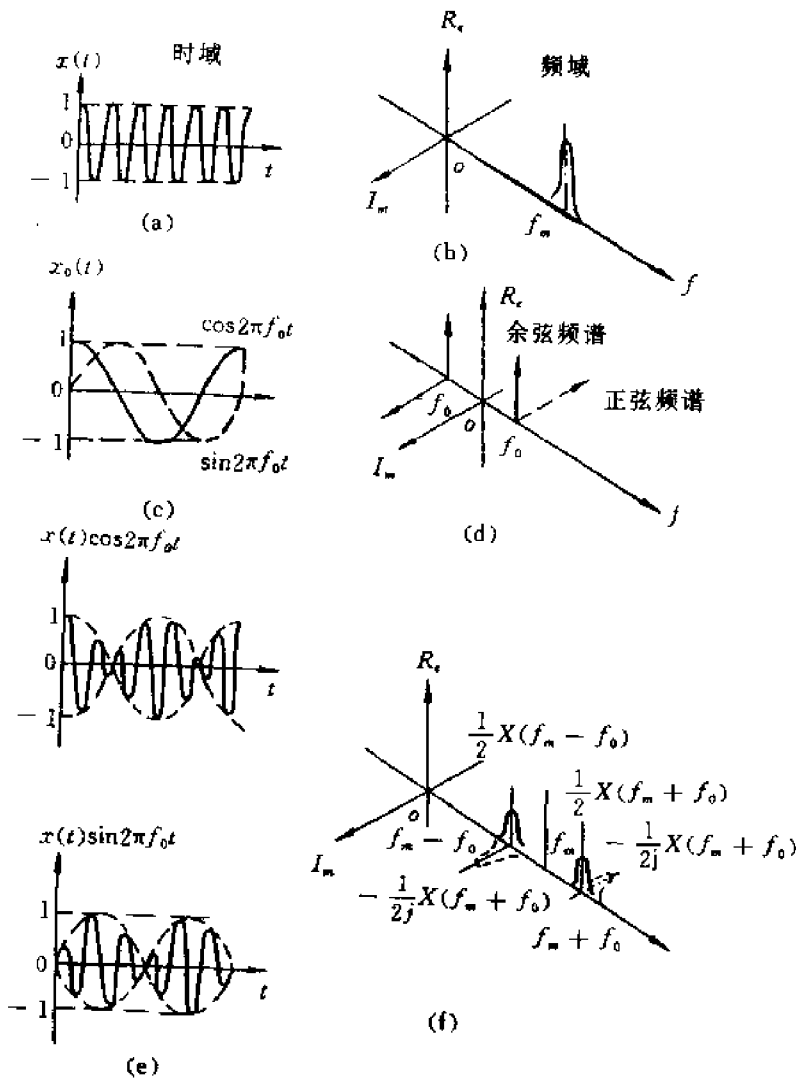


图 7-22 边频带谱的形成原理

(a) 载波 (b) 载波频谱 (c) 调制波 (d) 调制波频谱 (e) 幅值调制信号 (f) 幅值调制信号频谱(边频带谱)

频率调制就是载频信号受到调制信号的调制后,变成变频信号,如图 7-24 所示。

齿轮由于齿间误差或载荷发生周期性变化而产生振动时,不但有幅值调制,同时也有频率调制。频率调制波的频谱也是在载频(如 f_m) 谱线的两侧产生等间隔的边频带,边频带的间隔 f_0 就是调制信号(往往是与齿轮故障有关的信号)的频率。由于幅值调制与频率调制的边频都相同,所以在频谱图中这两种调制的边频带是重叠的。对于同一频率的边频带谱线,如果两者的相位相同,则它们的幅值相加;两者的相位相反时,则它们的幅值相减。所以载频谱线两侧的边频带的分布

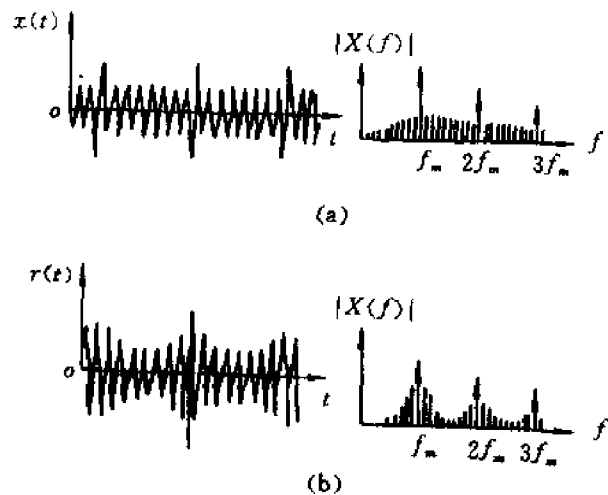


图 7-23 两种缺陷在时域和频域上的差别

(a) 局部缺陷 (b) 连续缺陷

一般是不对称、不规则的,如图 7-25 所示。

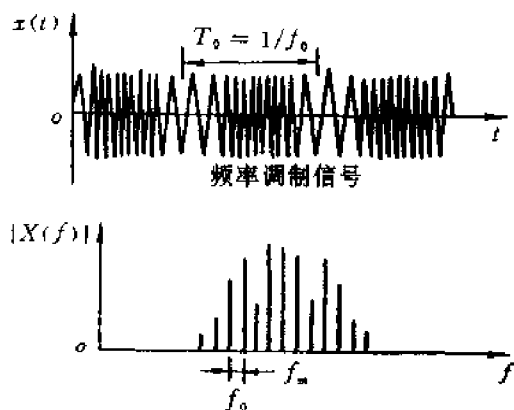


图 7-24 频率调制信号及其频谱

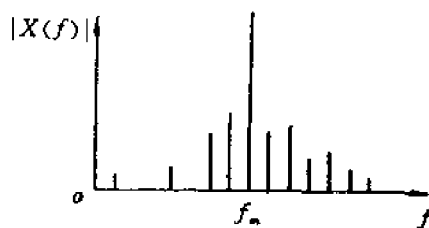


图 7-25 两种调制的综合频谱

齿轮振动时,由于各种因素的影响往往既有以啮合频率为基频的振动,又有它的高阶谱频分量;而轴的转动频率也常有高阶谱频分量。齿轮振动的各调制边频可用下式表示

$$f_{\text{边}} = Pf_m \pm Mf_{r1} \pm Nf_{r2} \quad (7.35)$$

式中, f_m 为齿轮副的啮合频率; f_{r1}, f_{r2} 为主动齿轮、被动齿轮的转动频率; $P = 1, 2, 3, \dots$ 为啮合频率各阶谐波的序数; $M, N = 1, 2, 3, \dots$ 为分别为主、被动齿轮转动频率的各阶谐波的序数。

由式(7.35)可以看出,齿轮振动的边频带分布非常复杂,倘若齿轮箱中同时有几对齿轮啮合的话,几对齿轮振动的边频带重叠在一起,其频谱图就更加复杂了,在这种情况下就很难直接从频谱图中识别出各特征频率来。

(3) 齿轮振动频谱图的组成成分:齿轮振动频谱图的谱线一般有下列几种:

一是齿轮的转动频率及其低阶谐波。它主要是由于转轴对中不良、轴变形、零部件松动等原因引起,使齿轮在运转过程中产生附加脉冲。图 7-26(a) 为有附加脉冲时的振动信号,其曲线不对称于零线,而被周期(轴系失衡产生振动信号)信号调制的齿轮振动信号如图 7-26(c) 所示,其振动曲线对称于零线,显然由时域曲线就很容易地将两者区分开来。

二是齿轮的啮合频率及其谐波、边频带。如前所述,这些振动成分是由齿形误差、齿面损伤等原因引起的。

三是齿轮副的各阶固有频率。它是由于齿轮啮合时齿间撞击(往往是由故障所致)而引起的齿轮自由衰减振动。从振动波形曲线上看,它是衰减曲线,在振动频谱图中,它们位于高频区,幅值较小,易被噪声信号淹没。

四是齿轮加工机床分度齿轮的啮合频率及其谐波(称为隐含成分)。其谱线往往在啮合频率的附近。这是由于当齿轮加工

机床分度齿轮的误差较大时,影响到被加工齿轮的齿形精度而引起的振动。此种谱线并非所有齿轮的振动频谱图中都是明显的,因为即使加工机床的分度误差较大,但当齿轮经过一段时间

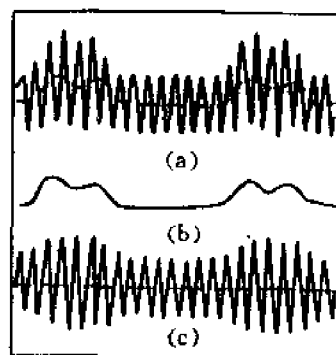


图 7-26 附加脉冲的特征

(a) 被附加脉冲调制的振动信号;

(b) 附加脉冲;

(c) 被周期信号调制的振动信号

的运转后,其分度误差的影响会由于齿面磨损而减少,其对应谱线的幅值也就随之变小。不过可利用此种谱线了解齿轮磨合、磨损的程度。

7.5 齿轮故障分析方法

7.5.1 频谱分析法

1. 频率细化分析技术

如上所述,齿轮的振动频谱图包含着丰富的信息,不同的齿轮故障具有不同的振动特征,其相应的谱线发生特定的变化。齿轮各种工作状态的振动频谱如图 7-27 所示。

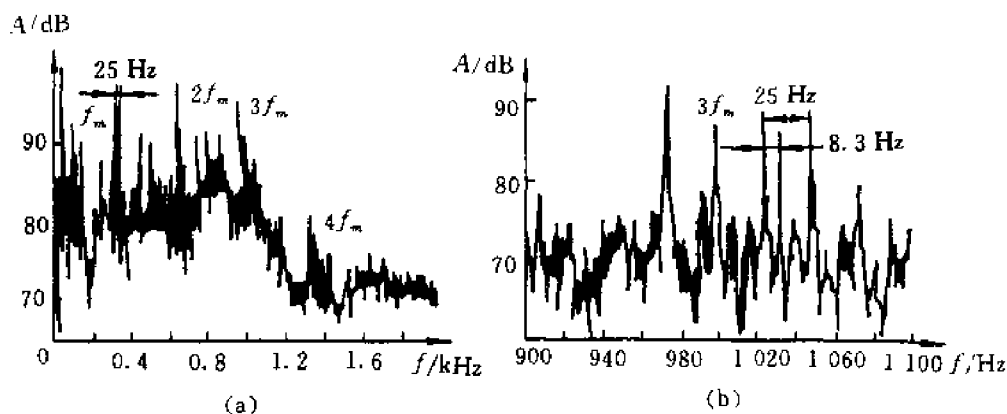


图 7-27 齿轮振动信号的频谱分析

(a) 普通频谱图 (b) 细化频谱图

对信号进行频谱分析时必须有足够高的频率分辨率。当边频带的间隔(故障频率)小于频率分辨率时,就分析不出齿轮的故障,此时可采用频率细化分析技术以提高分辨率。以某齿轮变速箱的频谱图(见图 7-27(a))为例,从图中可以看出,在所分析的 0~2 kHz 频率范围内,有 1~4 阶的啮合频率的谱线,还可较清晰地看出有间隔为 25 Hz 的边频带,而在两边频带间似乎还有其他的谱线,但限于频率分辨率已不能清晰分辨。为此,对其中 900~1100 Hz 的频段进行细化分析,其细化频谱如图 7-28(b)。由此图中可清晰地看出边频带的真实结构,两边频带的间隔为 8.3 Hz,它是由于转动频率为 8.3 Hz 的小齿轮轴不平衡引起的振动分量对啮合频率调制的结果。本例表明,用振动频谱的边频带进行齿轮的故障诊断时必须要有足够的频率分辨率,否则会造成误诊或漏诊,影响诊断结果的准确性。

2. 倒频谱分析

对于同时有数对齿轮啮合的齿轮箱振动频谱图,由于每对齿轮啮合都将产生边带谱,几个边带谱交叉分布在一起,仅进行频率细化分析是不行的,还需要进一步做倒频谱分析。倒频谱将原来谱上成族的边带谱线简化为单根谱线,使监测者便于观察,而齿轮有故障时的振动频谱具有的边带谱恰是具有等间隔(故障频率)的结构,利用倒频谱这个优点,可以检测出功率谱中肉眼难以辨识的周期性信号。

图 7-28 说明了倒频谱具有检测周期性的能力。图 7-28(a) 是齿轮箱振动信号的频谱, 频率范围为 $0 \sim 20$ kHz。其中包含啮合频率(4.3 kHz)的三次谐波, 谱上没有分解出边频带。图 7-28(b) 是 2000 线细分功率谱, 频率范围为 $3.5 \sim 13.5$ kHz。谱中包含前三次啮合谐波, 但不包含两根轴回转频率的低次谐波。再将图 7-28(b) 中 $7.5 \sim 9.5$ kHz 的频带用高分辨率的信号分析仪进行细化处理, 可以得到图 7-28(c) 所示的谱图。在这一谱图中已经可以看到由轴转速形成的边频带。图 7-28(d) 是图 7-28(b) 功率谱的倒频谱。倒频谱上清楚地表明了对应于两根轴回转频率(85 Hz 与 50 Hz)的两个分量 A_1 (11.8 ms) 和 B_1 (20.0 ms), 而在高分辨率谱图 7-28(c) 中却难以分辨出来。

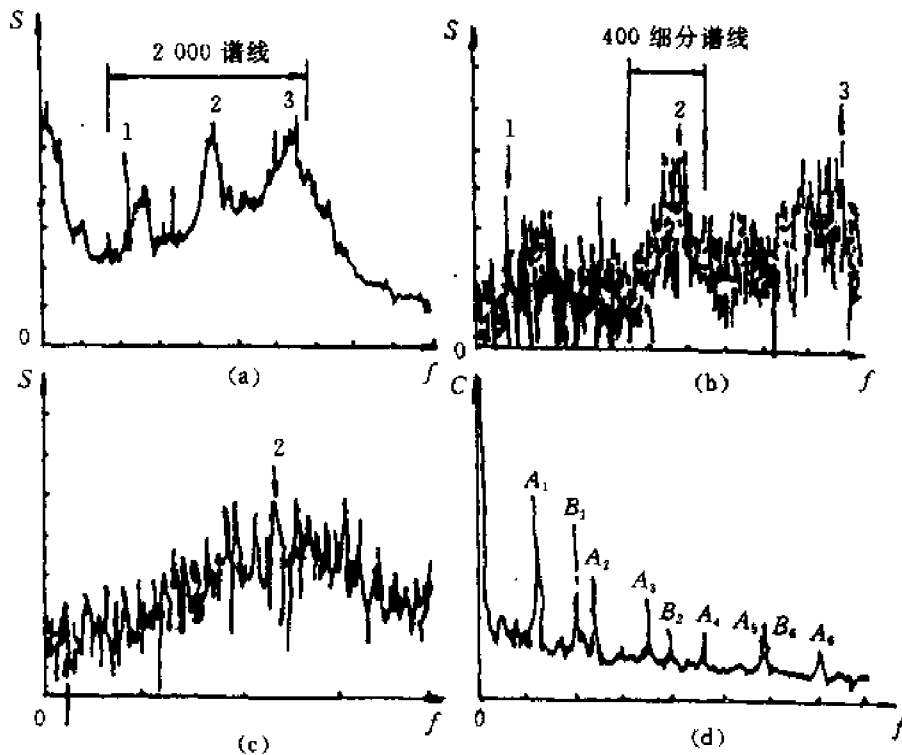


图 7-28 齿轮箱振信号边频带的倒频谱分析

1—啮合频率; 2,3—高次频率; A_1, A_2, \dots —周期 11.8 ms 谐波; B_1, B_2, \dots —周期 20 ms 谐波

图 7-29 说明倒频谱用于诊断的第二个优点, 即能精确地辨识频谱中的周期特性。图中所示是正常和异常状态下卡车变速箱一档齿轮啮合时振动的功率谱和倒频谱。正常状态的功率谱无明显周期性, 而从异常状态的功率谱中可看出有大量间距约为 10 Hz 的边频, 相应的倒频率为 95.9 ms (10.4 Hz)。在倒频谱图上能清晰的看到两个倒频率, 除 95.9 ms 的倒频率外, 还有一系列对应于输入轴转速的倒频谱谐波 (28.1 ms 或 35.6 Hz)。因为输出轴的回转频率为 5.4 Hz, 所以, 最初怀疑调制频率是输出轴的二次谐波, 但这样, 调制频率应该是 10.8 Hz 而不是 10.4 Hz。最后找到空转不受载荷的二档齿轮是调制源, 其回转频率准确地等于 10.4 Hz。由此说明了倒频谱辨识周期性的精确度。

由以上分析可看出, 倒频谱分析对于齿轮故障诊断是一种有效的方法。当频率分辨率不够时, 将频率细化分析与倒频谱分析结合起来, 则可得到满意的结果。

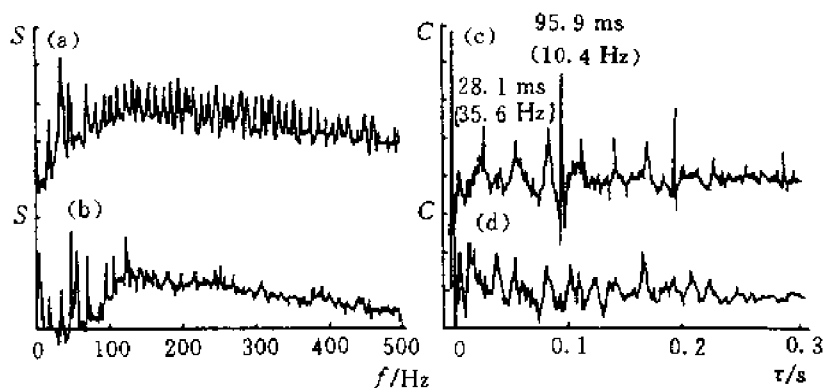


图 7-29 卡车齿轮箱正常与异常时的功率谱和倒频谱
(a),(c) 异常; (b),(d) 正常

3. 瀑布图分析法

除了倒频谱分析方法外,前述瀑布图分析方法也可以在齿轮箱故障诊断中应用。改变齿轮箱输入轴的转速并作出相应的振动功率谱,就可以得到瀑布图。在瀑布图上,有些谱峰的位置随输入轴转速的变化而偏移。一般只要增加 20% 转速即可发现这种偏移,它们一般属于齿轮强迫振动的频率。相反,有些谱峰的位置始终不变,不随输入轴转速的变化而改变,这种谱峰就属于共振所引起,这种共振可能是齿轮传动系统共振,也可能是箱体的。它们可以通过增加阻尼或改变设计来解决。

7.5.2 时域诊断

应用时域同步平均法可从复杂的振动信号中分离出与参考脉冲频率相等的最低周期成分以及它的各阶谐波成分。此法应用于齿轮箱的故障诊断时,可从总的振动信号中提取出感兴趣的那对啮合齿轮的振动信号,而把其他部件的振动信号及噪声成分都一概除去,从而大大地提高了信号的信噪比。由于滚动轴承的内环,外环或滚动体有损伤时,其振动信号与轴的转动频率不同步,因此,时域同步平均法也可将齿轮箱中齿轮故障引起的振动与轴承故障引起的振动区分开来。如果想要得到另一对啮合齿轮的振动信号,则只需使参考脉冲的频率等于其齿轮轴的转速即可。

根据时域同步平均法得到的时域信号曲线,可直观地分析出齿轮的某些故障,如齿面剥落、断齿等,也可对时间平均后得到的时域信号进一步作频谱分析。

图 7-30 是用时域同步平均法对不同状态下的齿轮检测时所得的信号。图 7-30(a) 是正常齿轮的时域同步平均信号,信号是由均匀的啮合频率分量组成,没有明显的高次谐波,整个信号长度相当于齿轮一转的时间。图 7-30(b) 是齿轮安装错位的情况。信号的啮合频率分量受到幅值调制,调制信号的频率比较低,主要是齿轮转速及其倍频。图 7-30(c) 是齿轮齿面严重磨损的情况。啮合频率分量出现较大的高次谐波分量,但由图中可见,磨损仍然是均匀磨损。图

7-30(d)的情况不同于前三种,在齿轮一转的信号中,有突跳现象,这种情况是在个别齿断裂时出现的。

图7-31所示是对一个齿轮链用时域同步平均法提取信号的情况。这一图示说明了时标提取的位置不同时,得到的信号也不同。图中的大齿轮 Z_1 上齿面有局部的剥落,当时标信号由位置 A 的传感器拾取时,由于大齿轮依次与两个小齿轮 Z_2, Z_4 啮合,经过的转角为 61.2° ,在时域同步平均信号 A 上有两个脉冲,其时间间隔相当于大齿轮转过 61.2° 所需的时间。相反,时标信号由位置 B 的传感器拾取时,大齿轮每转过一转时,大齿轮上有剥落的齿面与小齿轮只啮合一次,产生一次撞击,在时域同步平均信号 B 上只有一个脉冲。

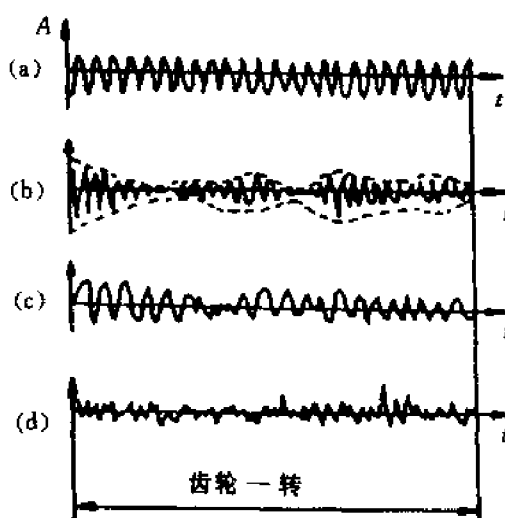


图7-30 齿轮在各种状态下的时域同步平均信号

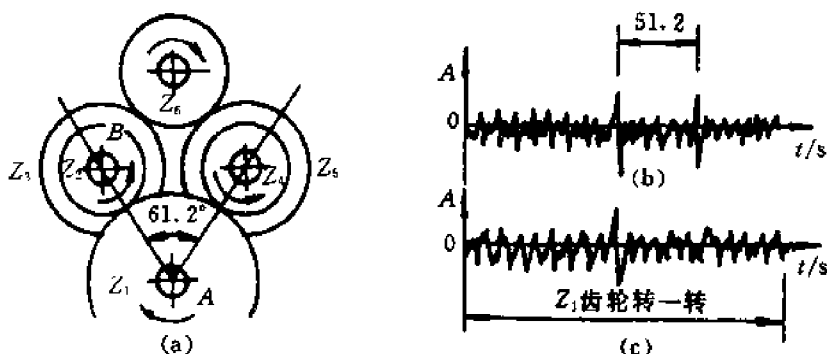


图7-31 齿轮链的时域同步平均信号受时标周期的影响

$$Z_1 = 51, Z_2 = Z_4 = 16, Z_3 = Z_5 = 31, Z_6 = 18$$

(a) 齿轮链信号测试; (b) 在 A 处接受的时标信号; (c) 在 B 处接受的时标信号

图7-32所示是使用时域同步平均法对光栅测量信号处理后的效果。两个光栅分别装在汽车齿轮箱的输入与输出轴上,在不同的输出扭矩(见图7-32(a))和不同的输入轴转速(见图7-32(b))下记录了传动误差曲线,然后将此传动误差作时域同步平均处理,得到齿轮箱中第三个齿轮的传动误差。

图7-32(a)所示的一组曲线,是齿轮箱输出轴转速保持在 48 r/min 而输出扭矩由零增大到 $150 \text{ N}\cdot\text{m}$ 时的传动误差。当载荷小时有脱啮现象存在,载荷逐渐增加时,开始时单齿啮合误差减小,而后又逐渐增大。这可以解释为轮齿弯曲和输入轴悬臂弯曲共同作用的结果。后一种弯曲效应开始时对齿轮的啮合螺旋线有所改善,而当进一步增大时却造成相反方向的错位。图7-32(b)所示是保持输出扭矩不变,逐渐增大转速时的情况。在 290 r/min 附近齿轮箱内部有共振现象,此时单齿啮合误差显著增大。

比较频域方法和时域方法,可以看到,频谱分析只需用加速度传感器拾取一个信号,而时

域分析则除了加速度传感器外,还需要一个时标信号。频域方法不能略去输入信息中的任何分量,因此待检的齿轮信号可能淹没在噪声中,而时域方法则能够有效地消除与时标周期无关的分量。频域方法依靠边频带来反映个别齿的缺陷,而时域方法可以很直观地查出个别齿的节距误差、剥落与断裂现象。这些都是时域方法不可忽视的优点。

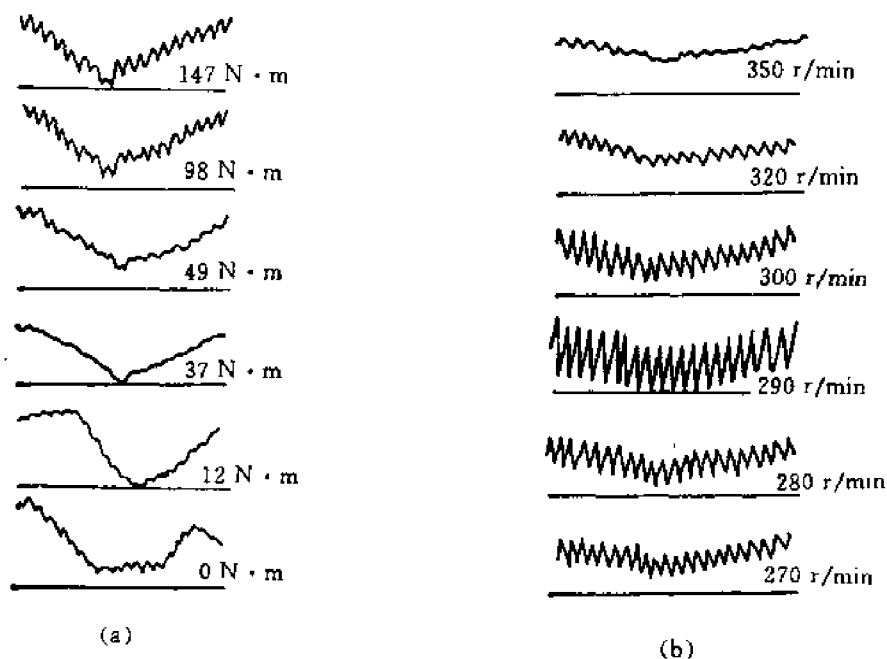


图 7-32 汽车齿轮箱输入轴上单一齿轮的传动误差
(a) 输出轴转速恒定时;(b) 输出扭矩恒定时

7.5.3 齿轮的精密诊断

齿轮箱故障的精密诊断,不仅要判断其运行状态是否异常及发生异常的部位,还要求判断异常的类型和异常的程度。齿轮的精密诊断是以频率分析为基础的。齿轮振动的频率范围很宽(从几赫到几十千赫),用同一个频谱图来表示它所包含的所有频率含量显然是不恰当和有困难的,因为频率范围与频率分辨率之间是有矛盾的。考虑到各种齿轮失效类型的特征频率分布,一般可分为3个阶段:① 0~100 Hz:各轴的转动频率;② 100~1 000 Hz:齿轮的啮合频率;③ 1 000~10 000 Hz:齿轮的固有频率。应根据需要选择分析的频段,一般分析的频率在0~1 kHz(低频)时,即可从频谱图中得到轴的转动频率及其谐频的谱线与啮合频率及其谐频的谱线,以及它们的边频带,从中可分析出故障的特征频率来。若谱线分布很复杂,很密集时,可进一步进行细化频谱分析以及倒频谱分析。几种典型齿轮状态低频(滤去固有频率成分)的时域波形和频谱示于图7-15,从图中可知,若同时分析时域波形与频谱,齿轮失效的类型是容易确定的。

以齿轮失效型式中最常见的轮齿磨损为例,当齿轮所有的轮齿均匀磨损而使齿隙增大时,在啮合过程中产生的冲击振动(衰减自由振动)的振幅和其他振动成分相比是相当大的,而且各次冲击振动的振幅差不多都相等,如图7-33(a)所示,冲击振动的频率(近似等于固有频率)为1kHz以上的高频。而冲击的重复频率就是啮合频率,所以啮合频率振动的成分也相应增

大,并随着磨损程度的加剧,齿轮刚度的非线性愈加明显,啮合频率的高次谐波成分随之增大,振动波形(低频)由正弦波变为如图 7-33(b) 的波形。

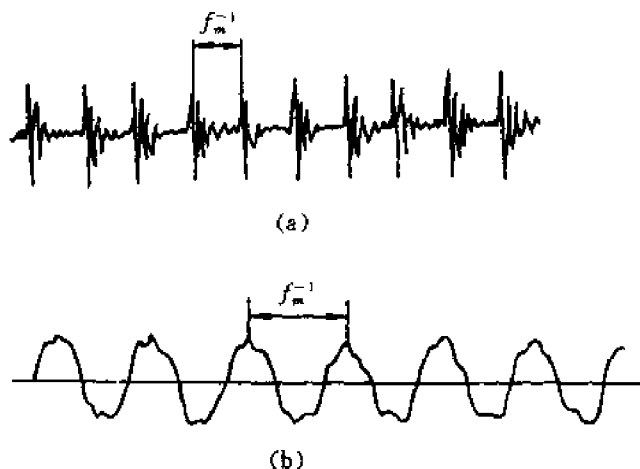


图 7-33 齿轮磨损后引起的振动(高频部分)

(a) 高频部分 (b) 低频部分

而当齿轮仅是个别轮齿有严重磨损时,其振动波形如图 7-34 所示。因为只有在异常啮合时才发生较大的冲击振动,冲击的重复频率即是轴的转动频率 f_r ,所以它的频谱图中轴的转动频率及其高阶谐波谱线相应较大。



图 7-34 局部有异常的齿轮发生的振动

如果需确定齿轮上故障齿的位置,可用时域同步平均法。图 7-35,因为异常齿啮合时,冲击振动的振幅要比其他齿的大,所以曲线上幅值最大的峰值位置即是异常齿的位置。

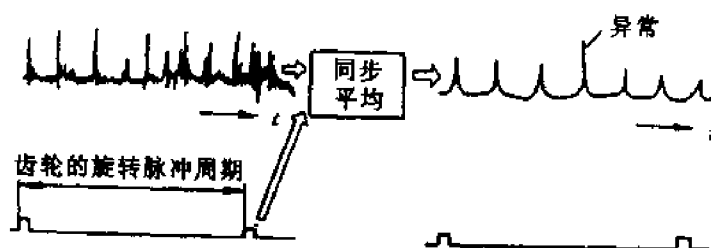


图 7-35 时域同步平均分析

习题与思考题

- 7-1 滚动轴承常发生那些异常现象?引起各种异常的原因是什么?
- 7-2 推导滚动轴承缺陷特征频率的计算公式。
- 7-3 简述滚动轴承内环、外环及滚动体有损伤时,引起振动的时频域特点。
- 7-4 损伤轴承与正常轴承幅值域中的概率密度相比有什么不同?
- 7-5 在滚动轴承振动故障诊断方法中,有效值判断法、峰值指标法各适于检测什么性质的故障?轴承发生损伤时,峰值指标如何变化?
- 7-6 冲击脉冲法进行轴承振动故障诊断时,使用哪些指标对轴承正常与异常情况进行判断?
- 7-7 试说明滚动轴承的包络分析诊断法的优点。
- 7-8 阐述用光纤监测法和接触电阻法进行滚动轴承故障诊断的诊断原理及说明各自的优点。
- 7-9 齿轮异常的基本形式有哪几种?
- 7-10 试分析齿轮运行中可能产生的振动形式及其频率特点。
- 7-11 在齿轮箱的振动频谱图中,如何识别调制现象?
- 7-12 为什么说倒频谱分析对于齿轮故障诊断是一种有效的方法?
- 7-13 如何确定齿轮上损伤轮齿的位置?
- 7-14 齿轮的轮齿均匀磨损、个别齿严重磨损时,其频域特性有什么不同?
- 7-15 齿轮振动信号的幅值调制和频率调制是怎样产生的?它们有什么共同点?

机械制造过程工况监测与故障诊断

8.1 概 述

8.1.1 关于机械制造过程工况监测与故障诊断

机械制造过程的工况监视与故障诊断是设备故障诊断学的一个分支,它是新发展的一个学科领域。它比旋转机械运行状态的监测与诊断滞后一个时期,因为在传统的生产环境中,以人为主体,机械制造过程的工况状态,人们可以很容易进行判断,有故障也很容易排除,对故障诊断技术的要求就不那么迫切,研究基础相对较薄弱。直至 20 世纪 80 年代初,随着柔性制造系统(FMS)和计算机集成制造系统(CIMS)的发展,在这种生产环境下,主要靠计算机对过程进行控制,大大地减少了人对过程的直接干预,若对过程状态不能自动识别,对故障不能自动排除,系统就无稳定性可言,也就不能使之形成生产力。它有待解决的问题更多,工况监测比故障诊断的层次更多、系统更大,除设备本身的故障之外,还包括一切与生产工艺过程有关的状态和故障。因此,在这种形势下,其工况监测与故障诊断技术越来越受到人们关注,它成为实现机械制造过程自动化、少人化或无人化的重要技术保证。本章针对机械制造过程的特点来进一步说明工况监测与故障诊断的技术原理及其应用。

8.1.2 机械制造过程工况监测与故障诊断的特点

机械制造系统的动态特性不同于机器设备运行系统,它有以下特点:

1. 离散性与断续性

就制造系统而言,信息的主要形式是离散的,如零件尺寸、加工精度以及各种经济与技术数据等等。就加工过程而言,在一次走刀中的切削加工可以是连续的(车、钻、磨等),也可以是断续的(如铣);从一个零件的制造过程而言,工序与工序是两个相互独立的过程,而对加工质量来说,工序与工序又是相关的。

2. 缓变性与突变性

在固定的加工条件下,一台机床的动态特性是缓变的,如机床的温升、零件磨损、应力的分布等都是缓变过程;但象刀具损坏、折断等往往是在瞬间出现的,属于突变性故障。

3. 随机性与趋向性

由于机械加工过程中的随机因素干扰大,因此机械加工过程中各种物理量的变化,如切削力、切削温度、刀具磨损与刀具寿命和切削条件的关系往往是含有趋向性的随机过程。

4. 模糊性

在现象与因素关系上,大部分呈模糊性,即一部分因果关系是透明的,而另一部分是黑色的,属于灰色系统。在状态分析中,需要用到各种建模方法,没有适用于各种情况的通用数学模型。在状态分类中,可分性是基本的,但类别间往往无确定的边界,客观上也存在模糊区,使得状态分类困难。

8.1.3 机械制造过程工况监测与故障诊断的方法与策略

任何一种制造系统,它的最终目的是在保证产品质量前提下,降低制造成本,提高生产率。因此,机械制造过程的监测与诊断的目的,就是为了保证达到上述目标,在线识别过程状态,通过参数调节,提供稳定的生产条件,分析过程中的问题。从监测诊断内容上看,有质量监控和过程监控,二者相互有关,但侧重面不同。前者是以产品质量控制为中心,后者是以保证加工过程稳定为目标。在传统的生产环境中,由于以人为主体,过程的工况状态,人们可以很容易进行判断,故应当以产品质量为主要目的;而在自动化生产环境中主要是靠计算机控制生产过程,这时,产品质量和过程监控都是十分突出的问题。

鉴于加工过程的随机性等特点,以下将以机械加工过程的动态特性的识别为主,结合实例进一步介绍监测诊断方法原理的应用及机械加工状态识别中一些特殊问题和解决方法。

8.2 切削过程刀具磨损与破损的监测与诊断

8.2.1 监视信号选择与实验分析系统

振动信号是一种信息载体,其突出优点是频响范围宽,对切削过程中的异常现象反映敏感,受环境条件限制较少,检测装置比较简单,安装灵活,调整方便,在生产条件下容易实现。选择刀杆垂直方向振动加速度作为原始特征信号,用加速度传感器拾取。检测和分析系统如图 8-1 所示。

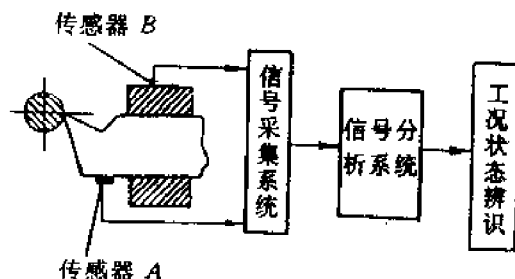


图 8-1 刀具磨损测试示意图

8.2.2 刀具磨损过程的频谱特征

切削过程中的刀具磨损使切削力发生变化,势必影响刀杆系统振动参数的变化。图8-2和图8-3是在刀具前角 $\gamma_0 = 12^\circ$,后角 $\alpha_0 = 7^\circ$,刃倾角 $\lambda_s = 0^\circ$;工件材料为45钢;切削用量是吃刀量 $a_p = 2\text{ mm}$,进给量 $y = 0.301\text{ mm/r}$,切削速度 $v = 125\text{ m/min}$ 试验条件下所得的低频功率谱图和特征变化规律图。当机床空转时,机床系统的振动频率为417.9 Hz(见图8-2(a))。当刀具一接触工作,由于刀具与工件之间产生摩擦,并且刀具与工件连成一体,和刀具没有与工件接触时比较,显然机械系统振动模态已经发生了变化,主峰频率位置开始向低频移动,旁瓣开始发展(见图8-2(b)),随着进给量的增大,旁瓣位置继续向低频方向移动(见图8-2(c))。图8-3(a,b)表明了低频段频谱变化规律,由图分析可知,它的特点是随着刀具磨损量的增加,主峰幅值开始增加较快,然后趋向平缓,而主峰频率位置则由高频向低频方向移动,且出现新谱峰。这种变化反映了刀具磨损的产生和发展,通过切削力激发起刀具—工件—机床加工系统的各段振动成分,从而导致系统振动模态参数发生变化;多谱峰的出现使得信号能量分散,因此,谱峰幅值增加缓慢;而刀具工件接触区条件的恶化引起的切削阻尼的增加,则使主峰频率位置降低。和图8-3(c)所示的实际刀具磨损规律比较,规律基本一致,但界限值并不突出。

图8-4为高频段($f > 1000\text{ Hz}$)功率谱图变化,图8-5表示频谱参数随时间的变化规律。可以看出,其变化规律与低频段基本一致,但规律性更为突出,边界明确。可以证明:刀杆的振动固有频率是在高频段(本例试验条件下约为3000~4000 Hz),因此,可以认为高频段频谱特性的变化,是刀具磨损通过切削力激发刀杆振动模态参数变化所造成的,且同前刀面与切削的接触长度和后刀面与工件表面的摩擦面长度的变化有关。而低频段是由于刀具磨损通过工件激发加工系统振动模态参数变化所造成的,机械系统的强迫振动一般

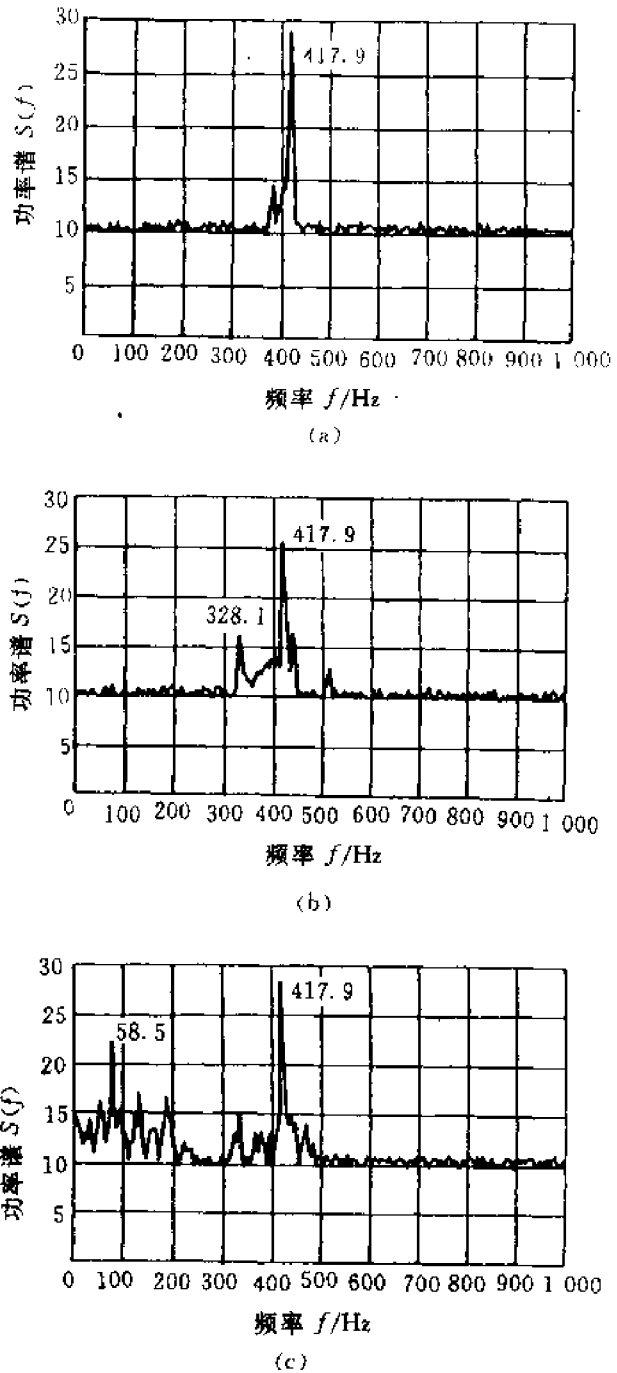


图8-2 低频段功率谱变化

都在 1 000 Hz 以内,因此,高频谱可以有效地隔离或削弱加工系统的谐波成分,而主要与刀具磨损的变化有关。谱峰幅值和频率的变化趋势与刀具磨损过程的规律性也更为一致,尤其是从正常磨损阶段向急剧磨损阶段过渡或是刀具即将发生破损时,频谱特性的变化相当显著。

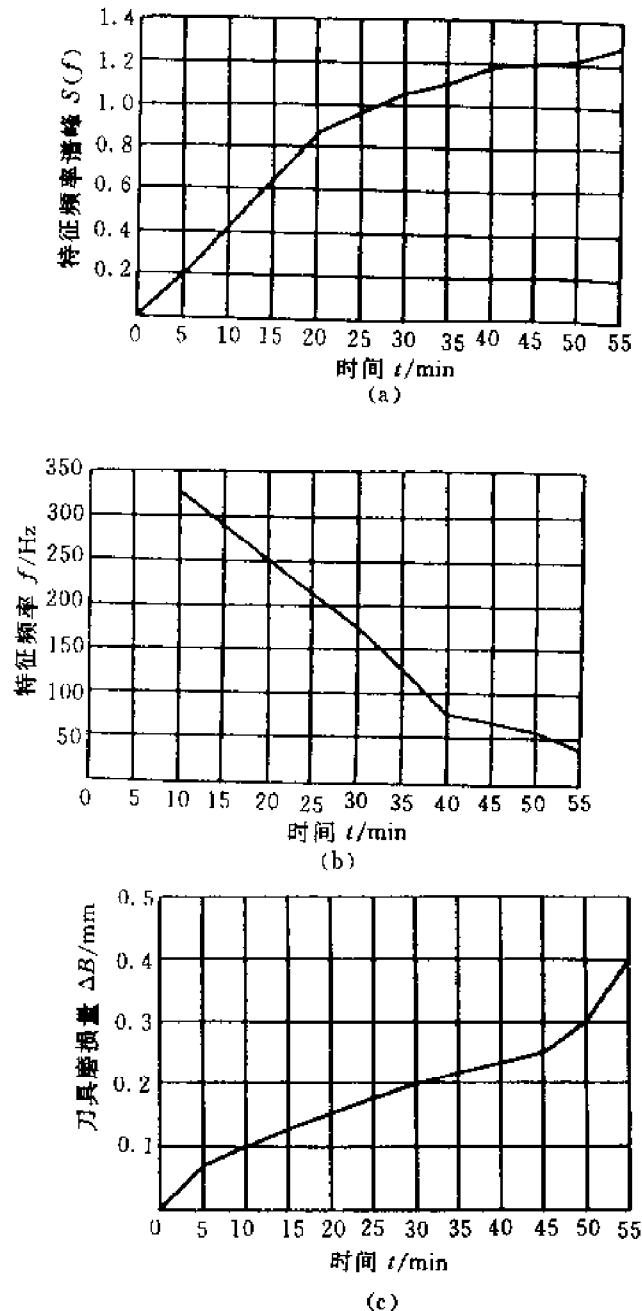


图 8-3 低频段特征频率与谱峰的变化和刀具实际磨损规律的比较

(a) 谱峰幅值随时间的变化 (b) 谱峰频率随时间的变化 (c) 刀具磨损随时间的变化

应该指出,振动加速度信号的频谱特性与切削条件及刀具—工件—机床加工系统的振

动模态有关,如果切削条件和加工系统不同,谱图特征将有区别。上述试验结果说明:从监测诊断的目的出发,要求选择对工况状态最敏感的特征量,对于切削过程刀具磨损与破损的状态识别,它不是在低频,而是在高频。

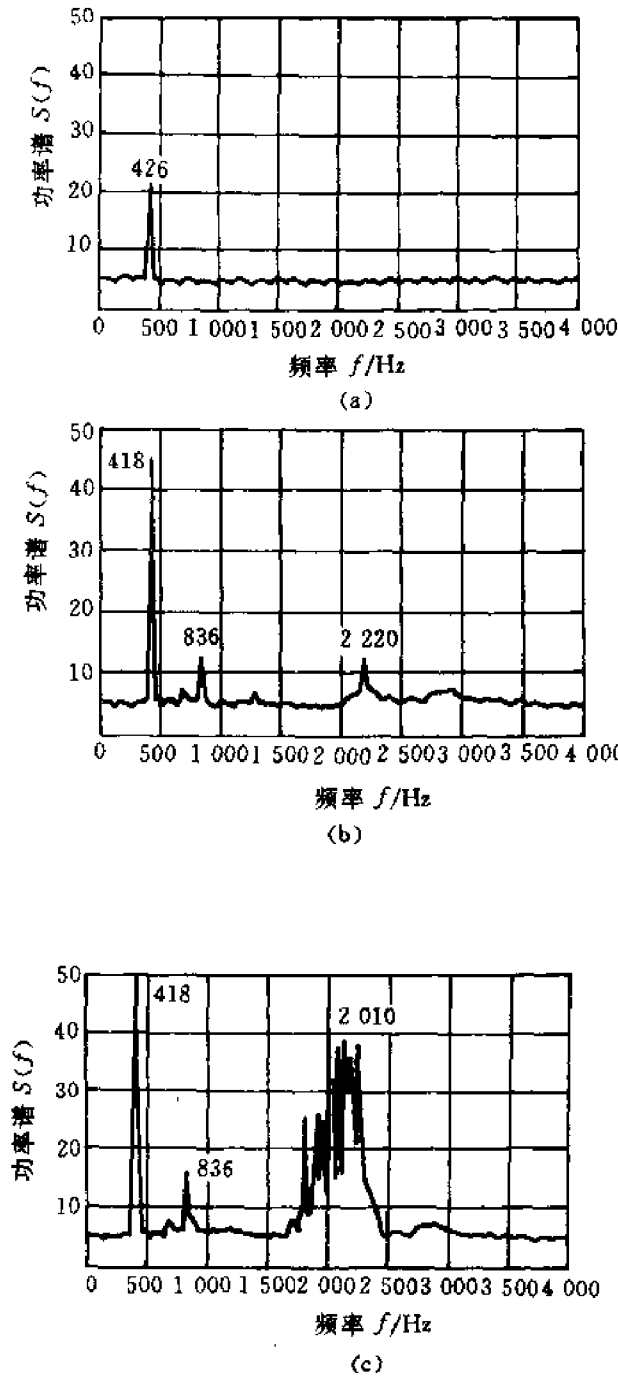


图 8-4 高频段 ($f > 1000$ Hz)

(a) 空转 (b) $V_B = 0.07$ mm (c) $V_B = 0.35$ mm

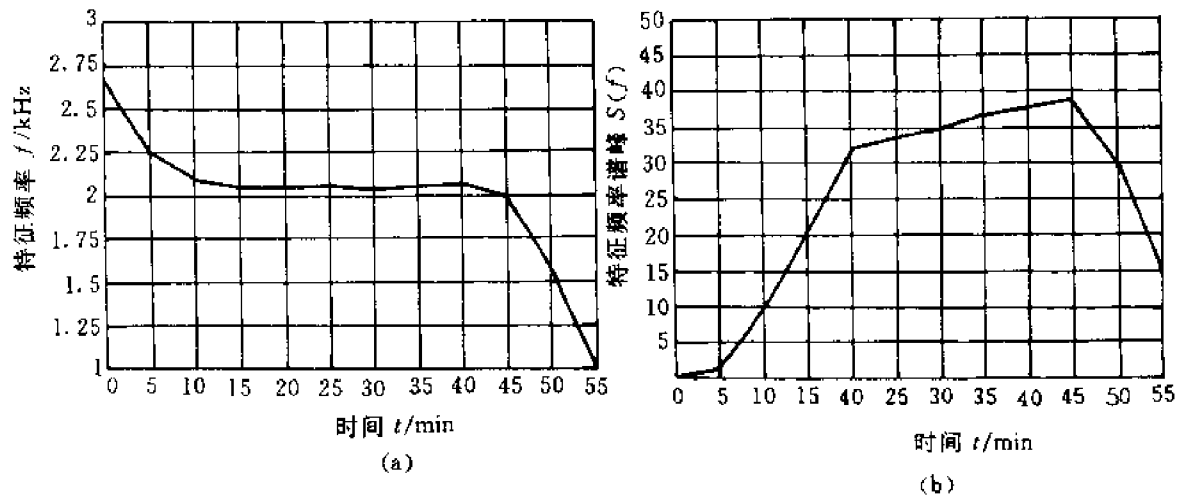


图 8-5 高频段频谱参数变化规律

(a) 特征频率位置随时间的变化 (b) 特征频率的谱幅值随时间的变化

8.2.3 刀具磨损过程的时序模型分析

前述已知,时序模型的结构、参数、残差和自协方差函数都能表达动态过程的特征。图 8-6 和图 8-7 分别表示了自回归模型参数和残差方差与刀具磨损过程之间的关系。显然,它们都能反映刀具磨损的变化规律,当刀具进入急剧磨损区时,刀具与工件接触状态的变化使数据间相关性增大(刀具如崩刃则相关性变小),故自回归模型参数急剧增大(刀具如崩刃则急剧减小);又因切削条件的恶化使随机成分增加,从而使残差平方和也增大。

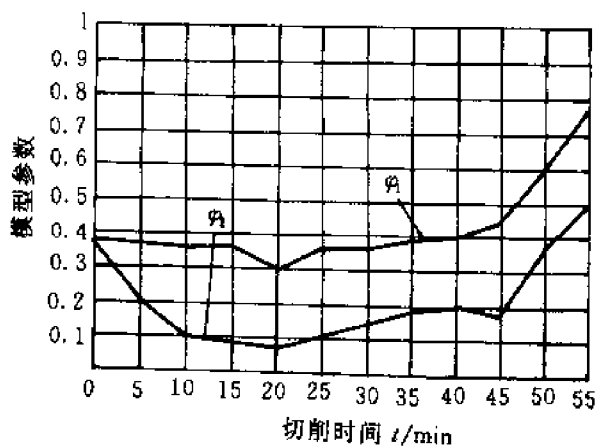


图 8-6 AR 模型参数随时间的变化

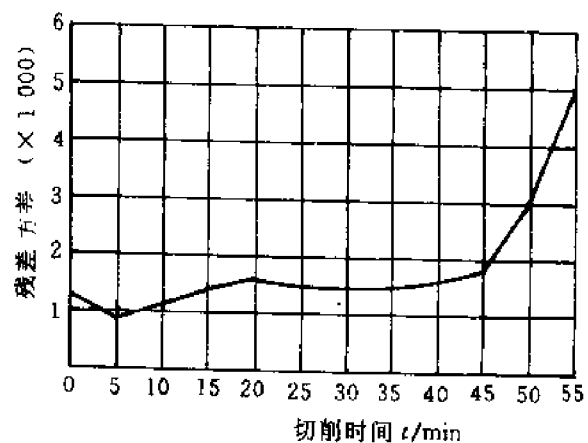


图 8-7 残差方差随时间的变化

8.2.4 刀具磨损过程的统计特征分析

很多时域统计特征量均与刀具磨损状态相关,其中特别是二阶矩统计特性对刀具磨损状态变化的反映最为敏感。图 8-8 和图 8-9 分别是一步自相关函数和样本方差在刀具磨损过程中的变化曲线,这种变化反映了由于刀具磨损导致的信号能量的变化,与频谱特性所表达的规

律性是一致的。但应该注意的是,这些图形是间断取样得到的,由于刀具磨损过程的波动性较大,在连续切削过程中,上述特征量的变化也呈现出波动性,但其变化趋势是一致的,尤其是刀具从正常磨损向急剧磨损过渡阶段的变化以及发生崩刃时,这种变化特征是相当显著的。

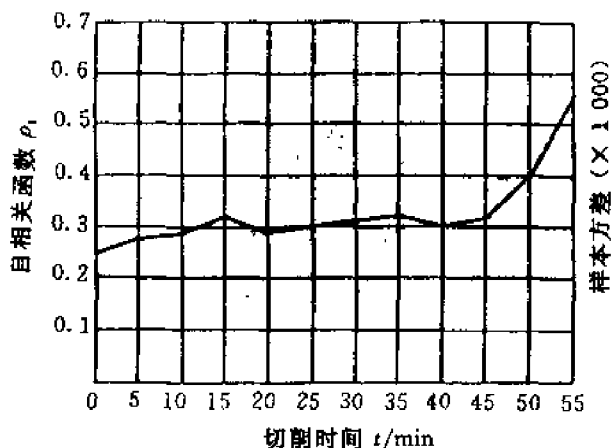


图 8-8 一步自相关函数随时间的变化

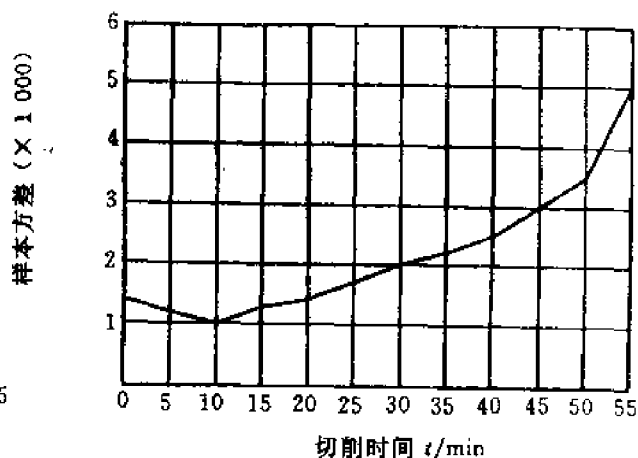


图 8-9 样本方差随时间的变化

由前述时域、频域或统计特性分析可以看出,在刀具正常磨损状态(含初期磨损)和异常磨损状态(包括急剧磨损和破损)之间,有关特征量的变化趋势图上均有一个与刀具磨损过程变化特性相对应的转折点(见图 8-5~图 8-9),初看上去,这个转折点可以用来作为状态分类的门限值,但实际上此转折点是随机的,它是一个过渡区;即使是相同的切削条件下,由于工件和刀具材料的不均匀性,以及加工系统的外界环境的随机干扰,对于不同的实现,其转折点也是不确定的。显然,若用某一过程的转折点作为门限值去决定另一过程的磨损状态,误判率是很大的。因此,还需应用这些特征量构成模式空间,在此基础上去研究状态识别的方法。

8.3 金属切削颤振的在线监测

颤振是金属切削过程中的重要问题,它涉及加工过程的稳定性和加工质量,不能等颤振发生才控制,人们感兴趣的是颤振预兆的在线识别。对切削颤振预兆的识别也就是对切削颤振进行早期诊断,属于两类属性的模式分类问题。

8.3.1 特征信号的选择

颤振是一种振动现象,图 8-10 表示切削过程从平稳到失稳的过渡过程中振动加速度的时域信号,显然,无颤振与已颤振的时域信号存在明显的差别,后者的振动周期明显,振幅增长很大。考虑到传感器安装方便和信号传递可靠,以及尾架受切削过程中各种影响较小,选择尾架顶尖的垂直方向振动加速度信号作为特征信号是合适的。

8.3.2 切削颤振的统计特征分析

实验表明:与加工系统有关的频率变化主要在 1 000 Hz 以下,颤振频率亦在此范围内,故

通常以 1 000 Hz 作为截止频率, 对时域信号进行低通滤波后再进行采样计算。

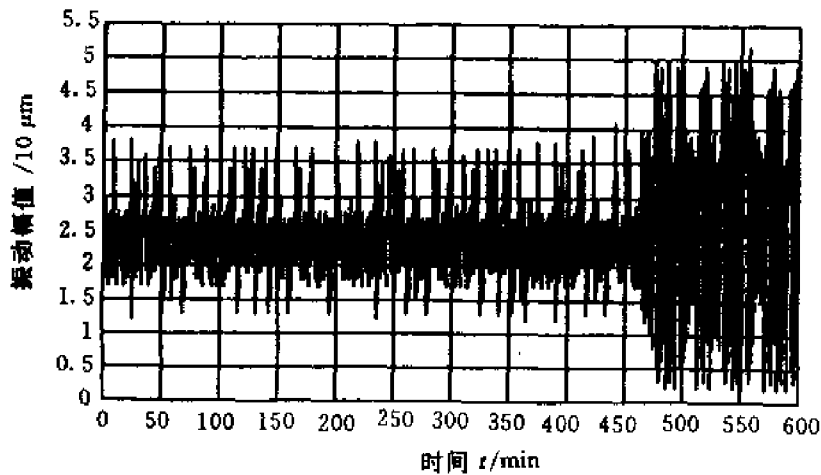


图 8-10 振动时域信号

随着颤振的孕育、形成和发展, 振动逐渐加强, 振动能量逐渐加大, 表现为振动信号的幅值增大, 如图 8-10 所示。前述已知, 随机信号可用方差来描述其强度, 由图 8-11(a) 可知, 在颤振即将发生的瞬间, 信号方差迅速增大, 一旦颤振发生后, 信号方差有所下降, 这表明幅度趋于稳定, 而方差值仍远大于无颤振时的方差。

图 8-11(b) 表示时域信号的一步自相关函数, 它反映了随着颤振的产生和发展, 相邻采样数据的相关性呈现出增强的趋势, 但其规律性不如方差显著。

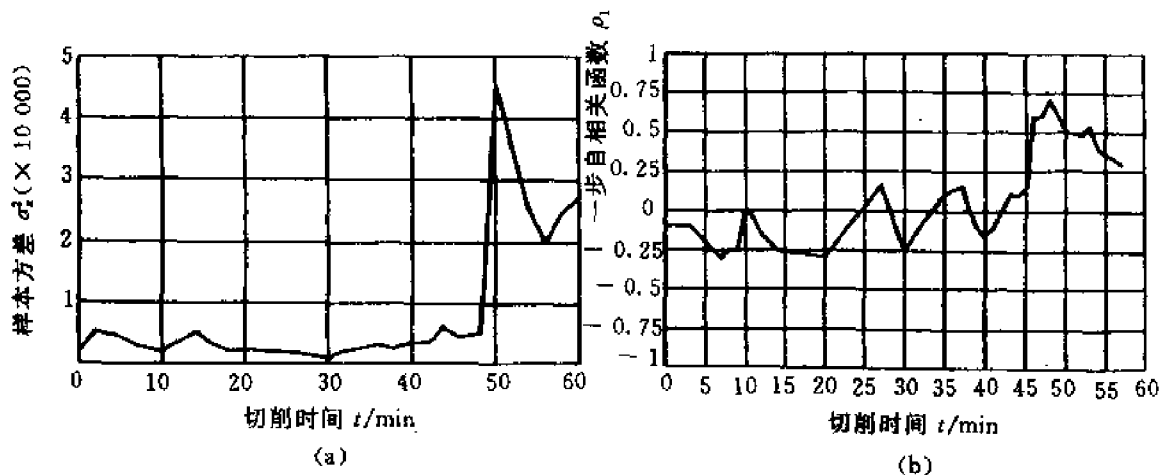


图 8-11 样本方差和一步自相关函数随时间的变化

(a) 样本方差 (b) 一步自相关函数

8.3.3 颤振的频域特征分析

频域中颤振预兆是极为明显的, 图 8-12(a, b, c) 表示加工时, 颤振发生前后, 振动加速度信号的功率谱的变化。在颤振尚未发生前的平稳切削过程中, 产生的振动较小, 振幅较低, 功率

谱的主频带位于较高频段(约为 300 Hz 左右),如图 8-12(a) 所示,它这主要是反映机械加工系统的振动。而当颤振经孕育到形成时,其能量开始从小变大,切削系统振动加强,在 300Hz 以下,如图 8-12(b) 中 200 Hz 左右出现新谱峰。随着时间的延续,主峰频率逐渐向低频方向移动直至颤振频率附近,此例颤振频率为 200 Hz 左右,这时切削系统失稳。

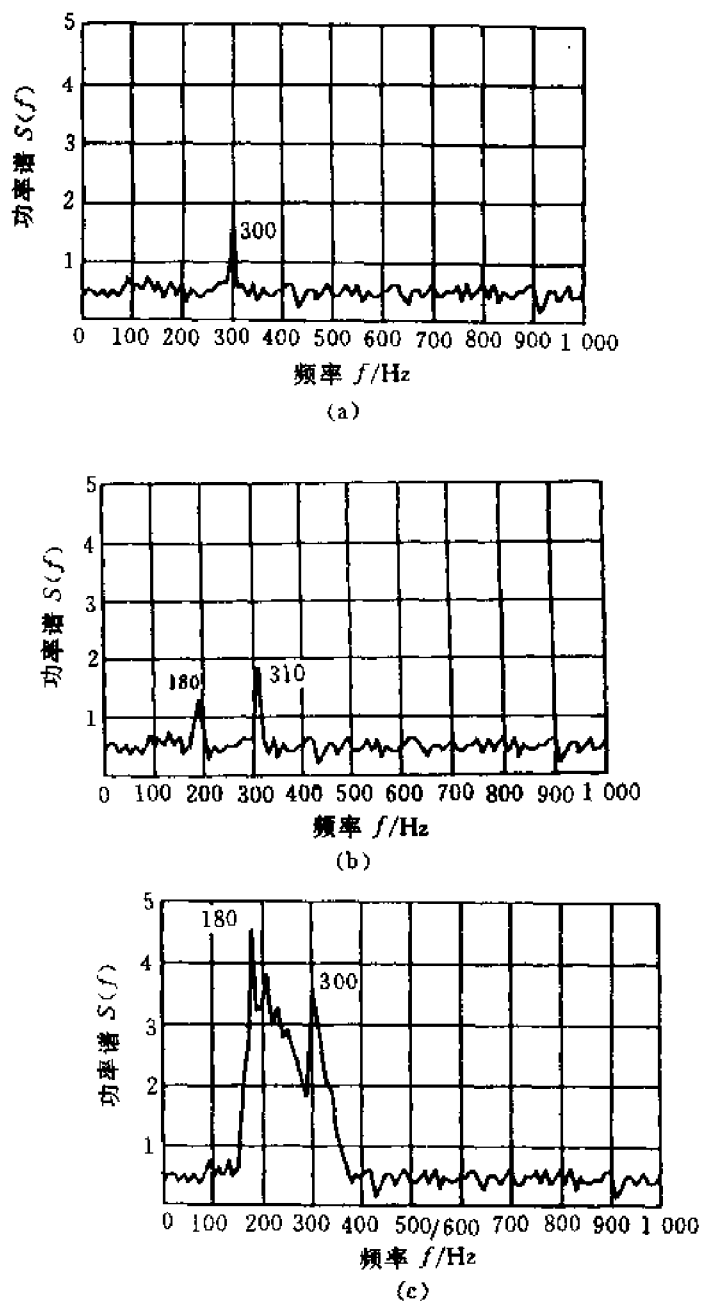


图 8-12 车削时颤振发生与发展

(a) 颤振未发生 (b) 颤振已发生 (c) 颤振剧烈

综上所述,在切削过程从稳定向不稳定的过渡过程发展中,振动加速度信息将有两个方面的变化:在时域方面,信号能量逐渐增大,可由信号方差 σ_x^2 加以定量描述,即当颤振将发生时,

σ^2 将增大;在频域方面,信号的主频带将从高频段向低频段移动,可由信号的一步自相关函数 ρ_1 加以定量描述,即当颤振将发生时, ρ_1 将增大。因此,信号的能量和主频带是描述切削过程状态变化的两个有效的特征量。

8.3.4 切削颤振的在线控制

根据机床颤振的非线性理论,机床颤振的振幅对于切削用量有极其敏锐的依赖关系。切削颤振发生后,只要适当改变切削速度或增大进给量,即可有效地抑制颤振。因此,在切削过程中,计算机一旦识别出颤振即将发生的预兆,即可迅速启动控制系统,调整切削用量,从而把颤振消灭在孕育阶段。

图 8-13(a) 和(b) 分别是在颤振稳定后,突然降低主轴转速和增大进给量时加速度信号变化的时域波形图,可见调整后,振幅立即迅速下降,颤振消失。这说明适当增加切削负荷,有时反而有利于切削稳定性。图 8-13(c) 是颤振发生后不采取任何控制措施的信号波形图,这时振幅不会减小,振动将持续下去,直到刀具损坏为止。

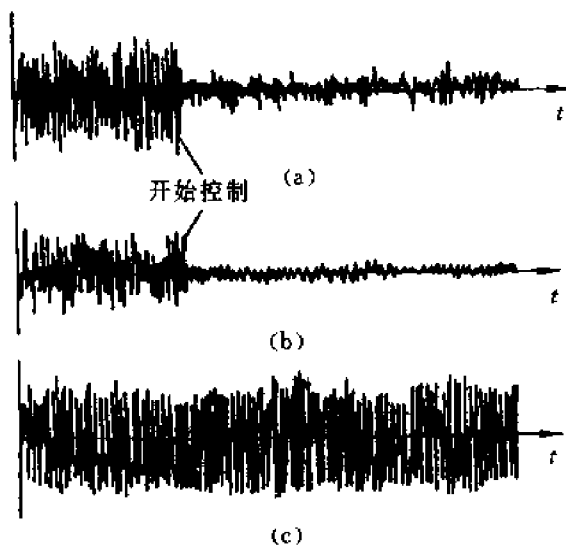


图 8-13 机床颤振的抑制

8.4 切削状态的在线识别

8.4.1 概述

由于切削过程的复杂性和断削的随机性,要在任何条件下都能得到正常的断削,是十分困难的,即使在切削初期断削正常,在切削过程中由于削形的易变性,切削形态仍有可能改变。不良削形将有可能造成切削过程的故障,在传统的加工环境中,靠人工控制与排除,而在自动化生产环境中,靠计算机控制,切削缠绕在刀具上,往往导致加工系统不能正常运行。因此,随着自动化生产的发展,切削状态的识别与控制,引起了人们的很大关注。

目前监测控制切削形状的方法有摄像法和利用切削特征的识别法,前者可以监视切削形状,但不能直接控制,必须经过数值转换变成数字量才好控制,而后者既可监测又可控制。本节

以特征分析和状态识别为主,说明切削形状的变化规律。

8.4.2 信号采集及预处理

实验结果表明:由于切削在折断过程中撞击刀杆的方向主要为走刀方向,因此刀杆走刀方向的加速度信号与切削状态的关系最为密切。针对切削过程干扰噪声太大,切削折断所产生振动微弱的特点,需对检测信号进行带通滤波、放大处理,以提高信噪比。为了剔除背景噪声的干扰和避免非削形改变(如切削用量的变化,工件材料不均匀性引起的硬质点的出现等)所引起的信号能量的变化,需要对采样后的数字信号剔除奇异点和进行归一化处理,数据归一化后再进行零均值处理。

8.4.3 切削的折断频率 f_c 的计算方法

在切削过程中,由于短切削的折断具有周期性,使得作用在刀具上的力相应发生周期性的变化,于是在刀具上产生了形成长切削时所没有的规律性的强迫振动,据此,研究切削的折断频率与功率谱图上频率的关系,便有可能获得有关削形的信息。

切削的折断频率 f_c 的计算方法是

$$f_c = \frac{v \times 1\,000}{l} = 1\,000 \frac{v}{l} \quad (8.1)$$

式中, f_c 为切削的折断频率(Hz); v 为切削速度(m/s); l 为形成切削前的切削层长度(mm)。

切削层长度 l 用重量法计算

$$l = \frac{M_c}{\gamma a_p \rho} \quad (8.2)$$

将式(8.2)代入式(8.1),则得

$$f_c = \frac{\gamma a_p \rho}{M_c} v \quad (8.3)$$

式中, M_c 为切削质量(g); γ 为进给量(mm/r); a_p 为吃刀量(mm); ρ 为材料密度(kg/m^3),碳素钢为 $7\,850 \text{ kg}/\text{m}^3 = 7.85 \text{ g}/\text{cm}^3 = 7.85 \times 10^{-3} \text{ g}/\text{mm}^3$,因为 M_c 的单位是(g), a_p 和 γ 的单位是 mm,故计算时用 $\rho = 7.85 \times 10^{-3} \text{ g}/\text{mm}^3$ 代入式(8.3),使量纲一致。

在较宽的切削用量范围内($a_p = 0.5 \sim 3 \text{ mm}$, $\gamma = 0.1 \text{ mm}/\text{r}$, $v = 40 \sim 202 \text{ m}/\text{min}$) 进行断削试验发现,切削的折断频率 f_c 均低于 200Hz,且大都在 10 ~ 100Hz 范围内。

8.4.4 切削折断状态的频域特征分析

图 8-14 是在不同切削条件下刀杆振动加速度信号的功率谱图。由图可见,无论在何种切削条件下,只要出现折断削,谱图上即有一个与切削折断频率相对应的谱峰出现, f_c 提高时,该谱峰频率亦随之提高。说明此谱峰频率与 f_c 有较好的一致性,因此,用功率谱的特征频率可表征切削是否折断,如图 8-15(a) 所示。当出现连续削时,该谱峰消失,谱图的低频段较为平坦如图 8-15(b) 所示。

顺便指出的是因为采用了低通滤波器,图 8-14 上没有机床加工系统的振动频率(300 Hz 左右),图 8-15 标明了 0 ~ 500 Hz 间的频率分量,加工系统的振动频率在 250 Hz 附近。由于切削折断所施加于加工系统的强迫振动,使得机械加工系统的频率左移,并出现许多新的小谱峰。

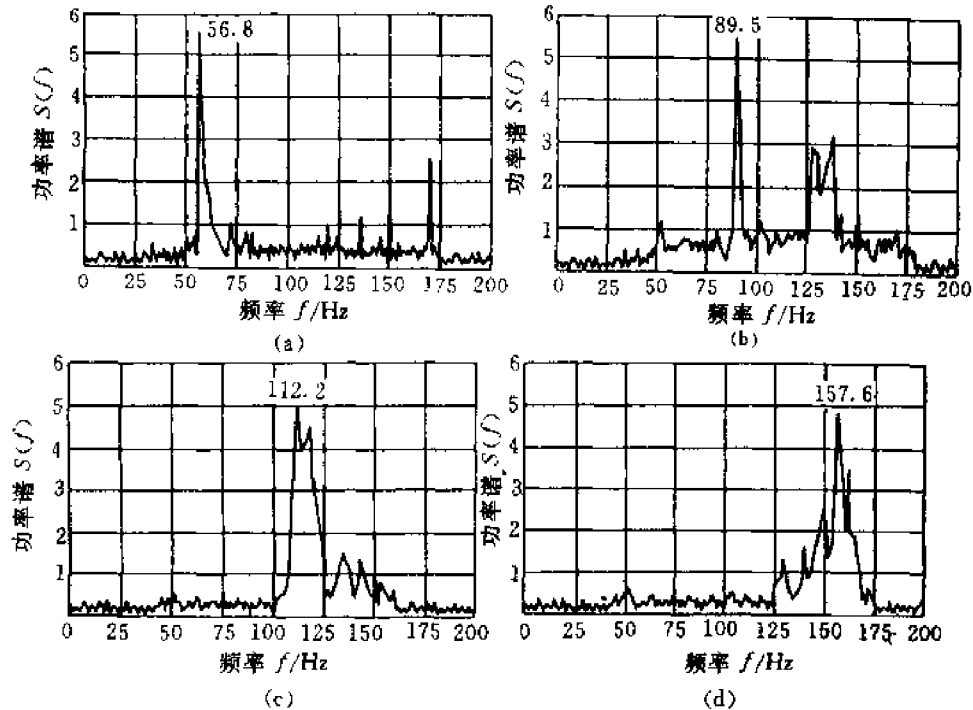
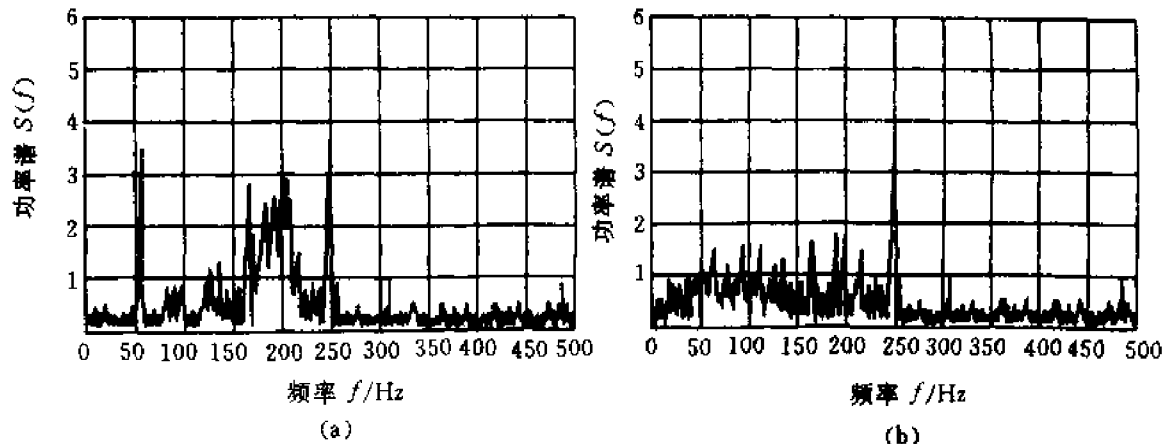


图 8-14 弧形削的功率谱图

- (a) $v = 53.8 \text{ m/min}$ (弧形削, $f_c = 47 \text{ Hz}$) (b) $v = 84.7 \text{ m/min}$ (弧形削, $f_c = 75 \text{ Hz}$)
 (c) $v = 107.6 \text{ m/min}$ (弧形削, $f_c = 84 \text{ Hz}$) (d) $v = 151.3 \text{ m/min}$ (弧形削, $f_c = 145 \text{ Hz}$)

图 8-15 切削的折断频率 f_b 在加工系统频率域的位置

为了定量描述连续削和折断削在频域中的差异,进一步提取频域的数字特征量,研究表明:敏感频段内信号功率谱密度函数的二阶中心矩比较理想,其定义为

$$M_2 = \frac{1}{f_{\max}} \sum_{i=1}^{f_{\max}} [S_i(f) - \bar{S}(f)]^2 \quad (8.4)$$

式中, f_{\max} 为敏感频段的频率上限; $S(f)$ 为功率谱密度函数; $\bar{S}(f)$ 为敏感频段内功率谱密度函数值的均值。在同一类切削条件下,折断削的 M_2 值均高于连续削的值,这是由于 M_2 表征了功

率谱密度函数偏离其均值的平均程度。形成短削时,低频段出现谱峰,谱图上谱峰增多,使得折断削的功率谱密度函数曲线波动较大,因此使 M_2 值增大;而形成连续切削时,低频区较为平坦(见图 8-15(b)),故 M_2 值较小。

8.4.5 切削状态的统计特性

试验结果表明:在同一类切削条件下,折断切削的样本方差 σ_x^2 较连续切削要大,而一步自相关函数 ρ_1 也高于连续切削的相应值,这说明了当切削状态由连续切削变为折断切削时,由于出现了与切削折断频率位置相对应的谱峰,使得敏感频段内的能量增大,且其他低频成分也增加,这符合两种切削状态的形成规律。若将 σ_x^2 和 ρ_1 结合起来组成新的特征量

$$r_p = \rho_1 \sigma_x^2 \quad (8.5)$$

便有更强的变化趋势,因为

$$\rho_1 = \frac{r_1}{r_0}$$

而 σ_x^2 是方差,等于 r_0 ,故 $r_p = r_1$,此即样本数据的一步自协方差函数。

8.5 磨削过程表面层烧伤的识别

8.5.1 概述

磨削烧伤是机械制造过程中的重要课题之一,它的实质是磨削表面层金属组织在磨削热作用下发生变化,磨削碳钢及其合金钢时,工件表面产生过回火层,在极薄层内形成残余拉应力,显微硬度降低,回火马氏体呈不均匀分布,变质层范围约 0.01 ~ 0.02 mm 左右,其稳定性差,在外载作用下,由于残余应力重新分布,容易形成微观裂纹,在交变应力作用下,容易使微观裂纹发展成为宏观裂纹,引起突发性故障。几十年来,人们对此作了大量的工作。在工程中解决这一问题的办法,一是在实验室研究的基础上确定典型零件的磨削工艺,用工艺条件保证磨削质量;二是抽样检查(酸洗、金相检查),但是都难以避免误检而将不合格的零件误认为合格品装入到机器中去,使机器具有潜在发生突发性故障的危险,应用工况监测与故障诊断的目的就是把烧伤的零件在磨削过程中识别出来,它是磨削质量控制的重要措施之一。

8.5.2 特征信号的选择与测量

由于磨削烧伤是一种相变,它的直接影响因素是磨削温度,其他因素(如磨削力等)也有影响,但不是主要的,因此选用磨削温度作为特征信号是毋庸置疑的。当然如果能测量工件与砂轮接触面的温度,是最理想的,然而不论哪种切削加工,都被砂轮与工件所遮盖。传统的测量磨削温度的方法是在工件或砂轮中埋置热电偶,但在实际生产中,可以说没有哪种工件可以允许在其中或砂轮中埋置热电偶,所以这种方法只适用于实验室,不适用于生产条件。因为,在实验室所得到的样本和生产条件下所得到的样本不属于同一母体,作为研究磨削温度的变化规律是可以的,但不能用作生产条件下状态监测的判别量。因此,用磨削热辐射流温度作为特征信号来监测磨削表面层的烧伤,其道理可从图 8-16 所示的磨削区磨削热的形成与分布来理解,磨削过程产生的热向工件、砂轮、周围介质中传播。

由图可知,磨削热源的形成可从以下几点说明:

(1) 磨粒磨削点的温度 θ_{dot} : 它是磨粒与工件之间的摩擦、耕犁、切削等作用所产生的温度的总和, 温度最高, 瞬时温度可达 $1\ 000^{\circ}\text{C}$ 以上, 以点热原形式存在, 磨削就在这里形成, 因此也是磨削热的热源。

(2) 砂轮磨削区的温度 θ_w : 它是砂轮与工件接触区内所有磨粒点的所有热能产生的温度 θ_{dot} 的总和, 虽然磨削区温度梯度很大, 但由于磨粒数量多, 磨削区温度仍然很高, 通常磨削碳钢高达 $700 \sim 800^{\circ}\text{C}$, 它传播到工件表面的温度也有 $500 \sim 600^{\circ}\text{C}$, 这是影响磨削表面烧伤的直接原因。

(3) 工件中的平均温度 θ_x : 它是指磨削热传到工作中的热能 Q_x 引起的整个工件的平均温升。通常只有几十度, 它对工件几何尺寸精度有影响, 但不能作为判别磨削烧伤的依据。

由此看来, 由磨粒点的热源产生的磨削区温度是产生烧伤的直接因素, 但由于被砂轮与工件自身所遮盖, 在生产条件下难以测量。由于磨削热辐射流的温度虽然不等于磨削区的温度 θ_w , 但取决于磨削区的温度, 采用磨削辐射热 Q_r 作为信号源, 用红外探测器拾取信号, 可检测性好。红外探测器所拾取信号经采样, A/D 转换, 输入数据处理及状态识别系统。磨削时, 磨削热辐射流以锥形方式放射出来, 并形成火花流, 在接触面的切线方向密度最大, 愈远离切线方向愈稀疏(见图 8-17(a))。

8.5.3 磨削热辐射流温度分布特征

磨削热辐射流温度分布特征如图 8-17(b) 所示, 磨削刚离开砂轮—工件接触区 a 点到 b 点 ($ab \approx 5 \sim 10\ \text{mm}$), 由于尚未吸收空气中足够的氧, 氧化还原过程尚未充分形成, 磨削还没有形成明亮的火花, 至 b 点后, 上述氧化、还原过程逐渐充分, 爆炸形成了可目见的火花流, 随着距 a 点距离 s 增加, 温度逐渐升高 (bc 段) 至 M 点时, 温度最高, 由于冷却的作用, 在 M 点以后, 温度又逐渐降低, 最后火花熄灭, 火花流形成过程中, 色泽是变化的, 大致是暗红—红—黄, 在点 M 及其附近呈橙黄色, 温度最高, 测量时探头应安装在 b 点附近, 探头所测得的热辐射流温度不是个别磨削温度, 而是同时通过热敏面积的所有磨削(包括脱落磨粒)所形成温度的统计平均值。显然, 火花愈密集, 意味着在时刻 t 通过磨削愈多, 从统计意义上讲, 就是样本多, 其均值愈接近真值, 即测量结果愈准确。反之, 若火花分布疏散, 随机成分大, 就愈不准确。通过

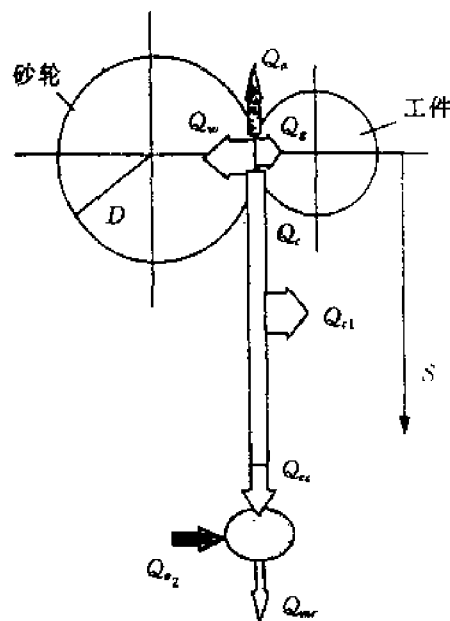


图 8-16 磨削区与磨削热的生成与传布

- Q_w - 磨削区热量进入砂轮部分;
- Q_x - 磨削区热量进入工件部分;
- Q_c - 被空气及冷却液带走的磨削区热量;
- Q_r - 被磨削带走的总热量;
- Q_{c1} - 被空气及冷却液带走的磨削流热量;
- Q_{ox} - 产生氧化的磨削流热量;
- Q_{ox2} - 切削与空气氧化产生的热量;
- $Q_{m,r}$ - 检测的磨削流的辐射热

热敏磨削热辐射流温度的统计平均值是指时刻 t 而言, 当时间 t 改变, 即磨削过程连续进行, 测量值 $\{x_t\}$ 是随 t 而改变的, 它与砂轮表面状态、周围介质(如冷却液)和磨削表面的 FeO 爆炸的随机性有关。因此, 测量值 $\{x_t\}$ 是一个连续的随机过程, 它表达了磨削热辐射流温度的动态变化, 可以利用现代各种信号处理及系统辨识理论与方法研究它的动态特性及状态属性。

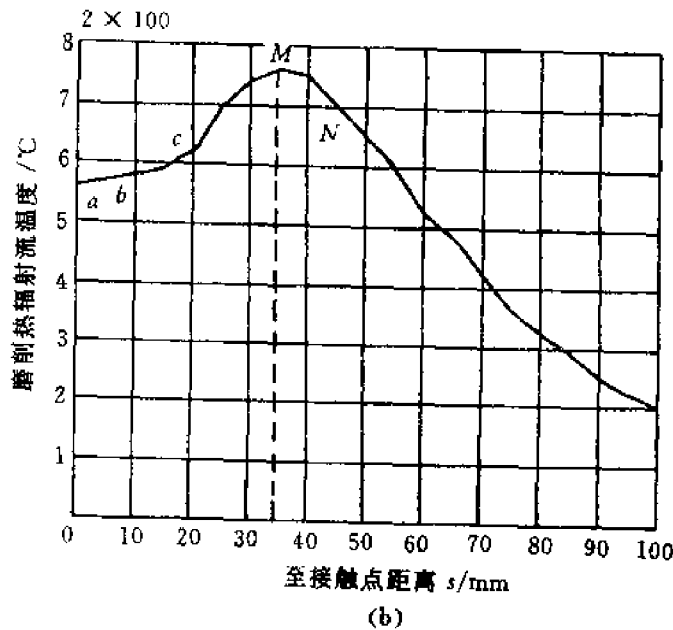
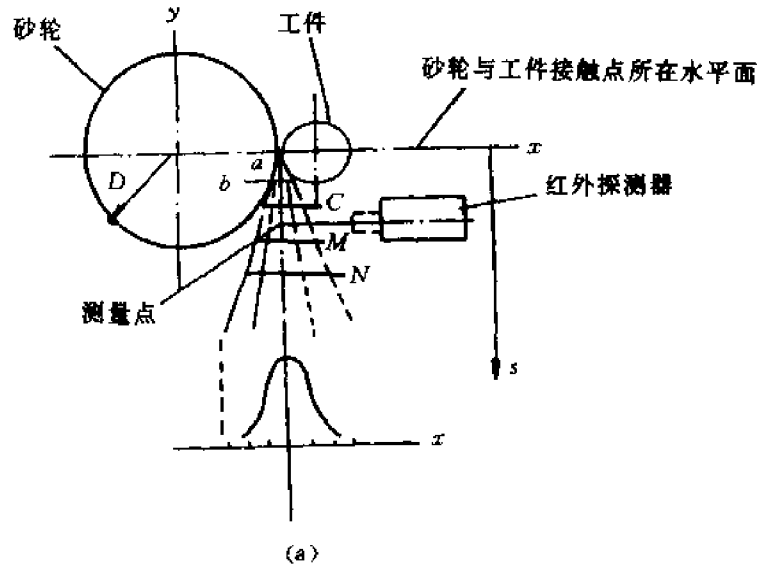


图 8-17 磨削辐射热与火花流温度分布

8.5.4 磨削热辐射流温度的统计特性

1. 磨削火花温度的自相关分析

设 $\{x_i, i = 1, 2, \dots, N\}$ 为磨削火花温度的时间序列, 并假设其均值 $\mu_x = 0$ 。由前述自相关函数的定义

$$R_x(k) = E[x_i, x_{i+k}] = \frac{1}{N} \sum_{i=1}^N x_i x_{i+k} \quad (8.6)$$

自相关系数为

$$\rho(k) = \frac{R_x(k)}{R_x(0)} \quad (8.7)$$

当 $k=0$, $R_x(0)$ 就是信号方差, 显然 $0 < \rho_1(k) < 1$ 。 ρ_1 表示一步自相关函数, ρ_1 大表示信号相关性大, 新修正砂轮的磨削力及磨削温度都比较小, 这时 ρ_1 主要受磨粒几何形状及其分布的随机性所影响, 故未烧伤信号的 ρ_1 较大。烧伤时砂轮堵塞、磨粒钝化都是随机的, 使得磨削火花温度随机变化大, 噪声成分增加, 故 ρ_1 较小。可见磨削火花温度系统不仅是连续随机过程, 而且是相关性强的随机过程。

2. 概率密度函数分析

随机信号分析的一个重要方面是概率密度函数, 它提供了观测值分布的统计特性, 同一类随机过程的概率分布应基本一致。图 8-18 表示磨削烧伤和未烧伤两类状态的概率密度分布, 为了比较, 将标准正态分布曲线也同时画出。由图可知, 两种状态都近似接近正态分布, 两类状态在概率密度分布曲线上的差异主要表现在分布曲线的“高”和“矮”。

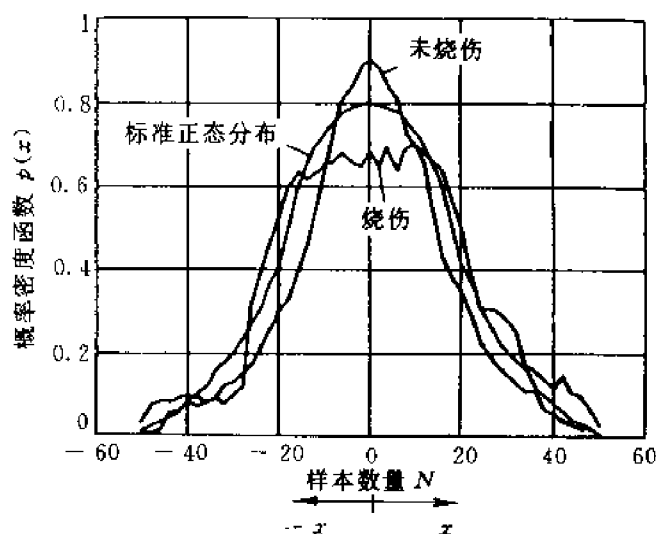


图 8-18 两类磨削状态的概率密度函数

8.5.5 磨削热辐射流温度的频域特性

磨削热辐射流的功率谱如图 8-19 所示。烧伤和未烧伤两种状态的谱峰主要在 200 Hz 以内。烧伤时频率位置向高频方向移动, 并在 $f=0 \sim 100$ Hz 和 $f=500 \sim 600$ Hz 处出现新谱峰。磨削火花温度的频谱特性与砂轮的表面状态有关, 烧伤的砂轮表面堵塞, 磨粒钝化, 能量增加, 但能量分散, 由于它们都是随机的, 故频率成分增加。

8.6 FMS 和 CIMS 监测与诊断系统概述

在计算机集成制造系统(CIMS)中, 车间、制造单元、机床设备和运输设备等都由计算机控制, 组成一个多级计算机控制网络。在 CIMS 和柔性制造系统(FMS)中, 对设备和生产过程的检测、诊断和控制是必不可少的, 诊断控制系统是系统网络的组成部分。

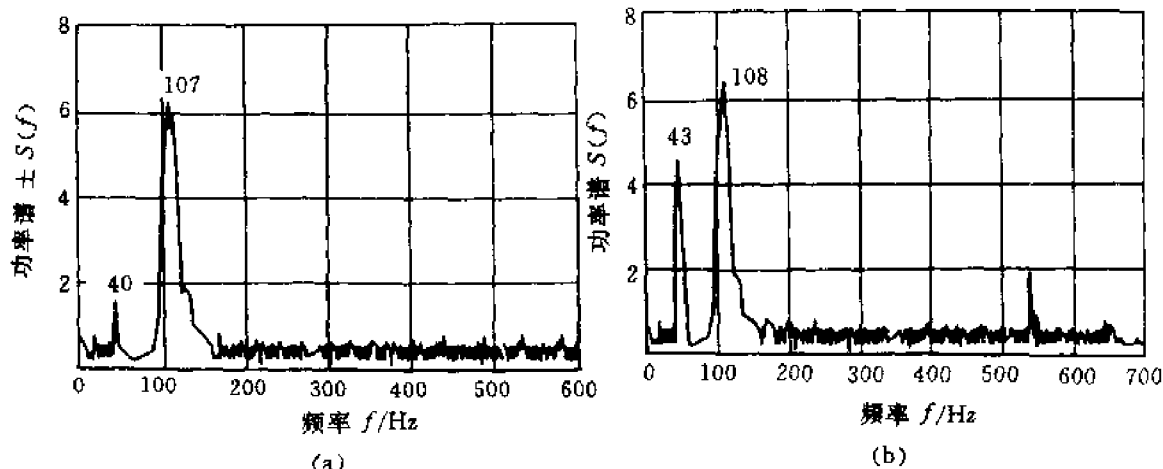


图 8-19 烧伤与未烧伤时磨削火花流温度的功率谱
(a) 未烧伤 (b) 烧伤

FIMS 中的设备应保证一定的无故障时间,如表 8-1 所示。

表 8-1 FIMS 中设备的无故障时间

设备名称	无故障时间 /h
柔性加工单元	650
柔性装配单元	650
柔性锻压单元	650
数控系统	6 000
计算技术设备	6 000
输送小库	6 000
工业机器人	2 500
检测和适应设备	2 500
电气、液压和气动自动系统	5 500

CIMS 的诊断监控系统一般可分为(车间)系统级、单元级和加工中心级三个层次。系统级的诊断监控系统主要是对各子系统间的通讯、协调和集成管理。加工(或装配)单元一般是一个完整的子系统。单元级诊断监测系统的原理的示例见图 8-20。诊断监测系统对单元内的各加工中心、小车、刀库等部分出现的误差、错误以及系统参数的异常变化,能迅速诊断并进一步作出相应的调整补偿决策和控制指令,保证单元的正常工作状态。诊断监控系统的结构示意图如图 8-21 所示。

加工中心级的诊断监控系统可相对独立工作,它也是单元级监控系统的一部分,如图 8-22 所示。一般由刀具状态诊断、设备状态诊断和加工工件质量诊断三部分组成。能够监测机床主轴功率和切削力;诊断机床主要部分的工作状态;监测刀具的位置与损伤、监测工件位置;监

控油温、油压、电压等加工环境参数；实现系统不同层次间的通讯等。总体监控系统技术流程如图 8-23 所示。

目前，国内外都在 FMS 和 CIMS 对生产过程的监测、诊断和控制方面积极开展工作。主要的研究内容有：诊断监控流程、人工智能(AI)技术应用、诊断决策方法、各种信号的采集分析方法与手段(如新型传感器、计算机视觉系统)等。存在的主要问题是传感器落后，数据处理实时性不够，信息储量中只有少部分信息被利用，监测装置只适用单一目标决策，方法简单，判别函数不能自动生成，软硬件和计算机数字控制(CNC)机床接口能力差等等。

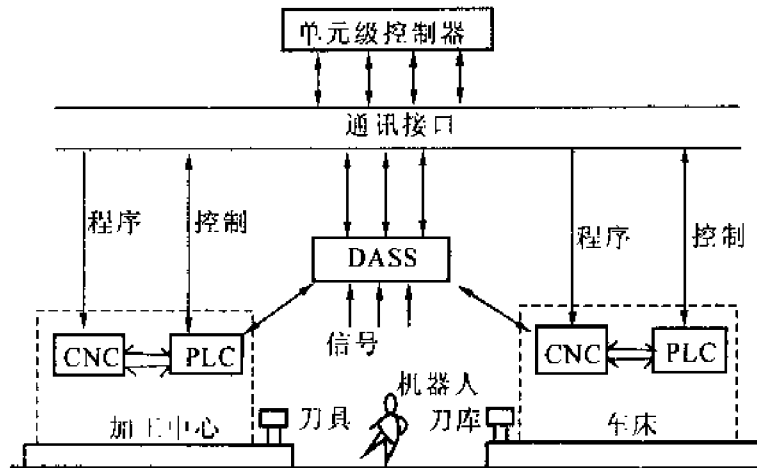


图 8-20 容错加工单元及监控系统
DASS—数据采集与处理系统

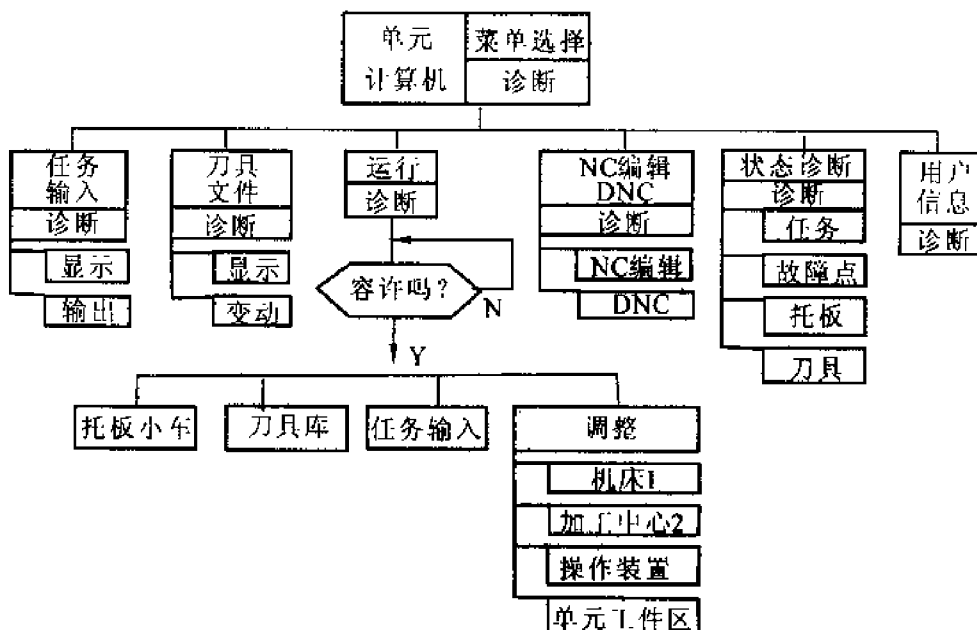


图 8-21 单元级监控决策系统

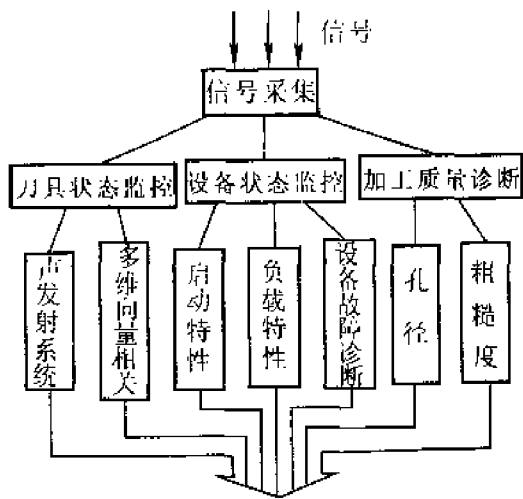


图 8-22 加工中心监控系统

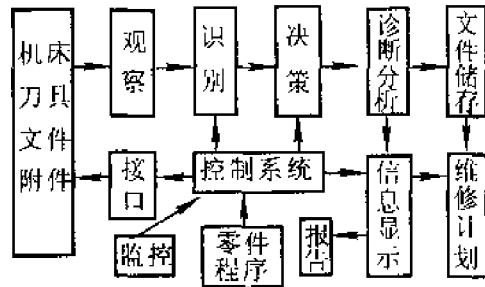


图 8-23 总体监控系统技术流程图

在 FMS 和 CIMS 的环境中,监测诊断系统要求具有如下功能:

- (1) 能进行多种、多通道信号的检测和大批量的信号处理,并具有智能接口;
- (2) 具有现代信息处理和知识处理的软硬件,能自动生成判别函数判别复杂状态,可进行状态特征分析、推理诊断和预报;
- (3) 设备与计算机联机,实现在线诊断监测;
- (4) 具有较大的柔性,并与 FMS 的柔性有机的结合起来等;
- (5) 具有自诊断功能,能诊断系统本身的故障。

总之,由单一目标向多功能、柔性方向发展,是监测诊断技术今后的重要任务之一,它不仅是自动化生产环境的需要,对促进当前机械制造工业的发展也具有重要意义。

习题与思考题

- 8-1 试分析切削过程中刀具磨损过程的特征频率随刀具磨损量变化的变化规律。
- 8-2 刀具磨损过程时序模型的模型参数和残差方差是如何变化的?
- 8-3 金属切削颤振的振动信号随着颤振的发生与发展其特征量的变化有什么特点?如何进行切削颤振的在线控制?
- 8-4 分析切削折断频率与功率谱特征频率的关系以及其在加工系统频域中的位置。
- 8-5 磨削过程中磨削热辐射流温度的分布特征是什么?
- 8-6 如何用磨削火花温度的自相关系数判断磨削砂轮的状态?磨削砂轮的状态会对磨削火花温度的频域特性产生什么影响?
- 8-7 论述多功能、柔性监测诊断系统研究的意义。

第 9 章

无损检测技术

无损检测技术是指对机器零部件和结构件(金属和非金属)进行非破坏性检测,以期发现表面和内部缺陷的一项专门诊断监测技术。

在现代机械故障诊断中,无损检测技术是探求故障根源和机理,预测隐患的重要手段。可以认为,无损检测是实现预防维修的先决条件之一。在工业生产中,无损检测技术对于保证产品质量、降低产品成本、提高安全可靠性以及改进制造工艺等都具有重要意义。

无损检测的方法很多,有人按照不同的原理方法和检测及信息处理方法,详细地统计了已经应用和正在研究的各种无损检测方法,总共达 70 余种。在工业生产中,目前应用最广泛的无损检测方法主要是光学检测法、渗透检测法、磁粉检测法、涡流检测法、射线检测法、超声波检测法和声发射检测法等。

为了对常用各种检测方法的检测能力作一大致介绍,这里引用《国内外设备诊断仪器实用信息手册》中的资料如表 9-1 所示。

表 9-1 表面裂纹的极限检测能力

检测方法	裂纹尺寸/mm			注释
	宽	长	深	
目视法	0.1	2	—	受表面状况和光照的影响
渗透法	10^{-2}	1	0.1	与表面状况有关
磁粉法	10^{-2}	1	0.2	与磁化方式有关,限于铁磁材料
涡流法	10^{-2}	1	0.2	与激励频率有关,适于导体
超声波法	10^{-2}	2	1	与反射情况有关
X射线法	0.3	5	0.3	垂直于表面、与厚度有关
声发射法	10^{-3}	10^{-3}	10^{-3}	仅在裂纹扩展时才有信号

9.1 光学检测

与其他方法相比,光学检测虽然显得陈旧,但有时却能发现一些其他方法难以发现的缺陷,只要探头能接近到离被测表面 24 in 的距离(约等于 610 mm),并且所成角度不小于 30° ,光学检测即可正常进行。

这种方法在更多的情况下要借助于简单放大镜和内窥镜进行观察。其

他光学辅助检测手段包括显微镜、潜望镜,有时还要利用摄影及电视技术。为了使缺陷的成像更为清晰,在必要时采用化学和电磁学方法(后面将述及)。下面着重介绍内窥镜。

内窥镜是为腔室和管道进行直接光学检测而设计的,它最早用于步枪枪膛和大炮炮膛的检测。内窥镜是一种单目镜光学系统,由一个物镜(有时带一个棱镜)、中继镜和一个目镜组成(见图9-1)。其工作装置具有各种不同的形式以适应不同的工作需要,如物距大小、目标大小、所在部位深度和反射率、缺陷尺寸和检测人口相对位置等。针对待测表面的不同方位和可能遇到的各种障碍,设置了不同观察角度的工作部件。直角(侧向)检测系统能够检测那些接近的角落,而后倾内窥镜能够对内部具有凸肩的管道进行精确的检测。

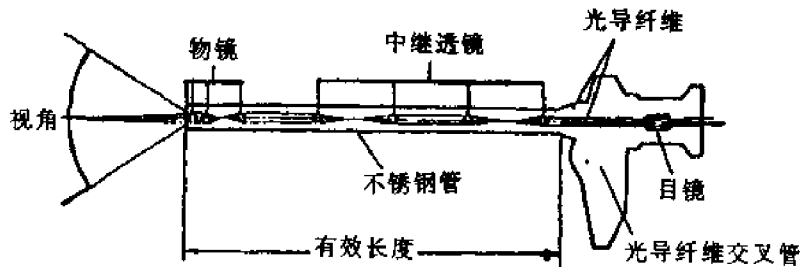


图9-1 硬管窥镜

内窥镜光源(如:紫外线光源、氙气光源和冷光光源)设在单独的一个装置中,其光线由分布在硬管内壁的成束光导纤维(每根直径约 $30\mu\text{m}$)送到前端,射向观察目标。

在小型管道内窥镜中,中继镜系统由一根单独的柱透镜所代替,被测物像在其中传播。

在挠性内窥镜中,以图像传输光纤束代替硬管式的中继镜组和柱透镜,更适合观察现场的实际需要。光纤束的每一根纤维都很纤细(直径 $7\sim 10\mu\text{m}$)柔软,入射的光线由于内反射作用而能沿其方向曲线传递。为了强化这种内反射作用,避免相邻纤维间互相干扰,在每根纤维外壁涂敷反光材料。

在图像传输光纤束中,为了保持图像不失真,这些光纤的相对位置是经过精心排列的,作用纤维越细其图像分辨力越高。

在光纤内窥镜中照明光线由光导纤维及连接器引向视场(见图9-2)。以内窥镜为主体,并与专用光源、照相机、摄影机和电视系统可组成一整套内孔观察记录和重放系统,大大便利了故障检测工作。

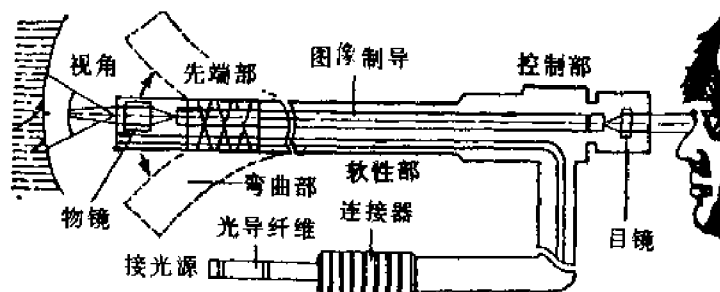


图9-2 光纤内窥镜

上述几种内窥镜已形成系列商品。

9.2 渗透与磁粉检测

9.2.1 渗透检测

渗透检测是把受检验零件表面清洗干净待干燥后,敷以专用的渗透液,由于表面细微开口缺陷的毛细作用将渗透液吸入其中,然后把零件表面残存的渗透液清洗掉,再涂敷显像剂把缺陷中的渗透液吸出,从而显现缺陷图像。

这种检测方法简单有效,已经有数十年的历史。它的局限性在于只能检验表面开口的缺陷,对于藏在表面以下的内部缺陷无能为力,而且它不适合检查多孔性材料或多孔性表面缺陷,因其缺陷图像很难辨认清楚。

1. 渗透检测过程

渗透检测的基本流程可概括为图 9-3 所示的四个阶段。

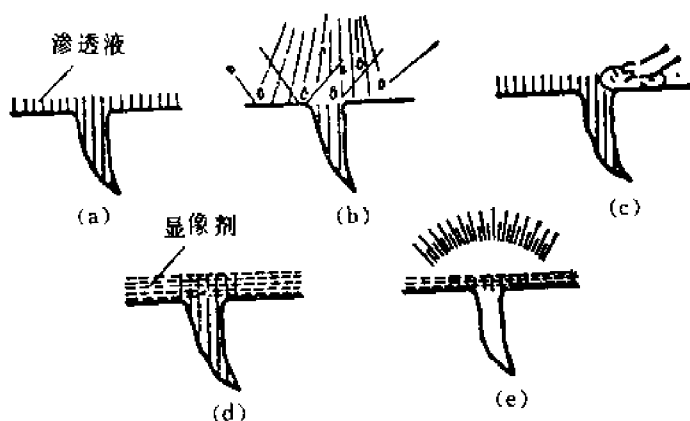


图 9-3 渗透检验的基本流程

(a) 渗透 (b) 水清洗 (c) 溶剂清洗 (d) 显像 (e) 观察

(1) 渗透过程:把被检验的零件表面处理干净之后,对于小零件,可将其浸渍于渗透液中,对于大零件,用喷雾器或刷子把渗透液涂在零件表面上,让荧光渗透液或着色渗透液与零件接触,如果零件有裂纹、孔隙和其他开裂处,渗透液就渗入其中。

(2) 清洗过程:待渗透液充分渗透到缺陷内之后,用清洗剂或水把零件表面的渗透液洗掉。由于清洗剂不能渗入到留有渗透液的裂纹和孔隙中,故裂纹和孔隙中的渗透液得以保留下来。

(3) 显像过程:清洗过的零件经干燥后,把显像剂(白色粉末)喷洒到零件表面上,残留在裂纹和孔隙中的渗透液就会被显像剂吸出到表面上,形成放大的黄绿色荧光或者红色的显示迹痕。

(4) 观察过程:用荧光渗透液的显示迹痕在紫外线的照射下能发出明亮的荧光,用着色渗透液的显示迹痕在自然光线下呈红色,因此很容易识别,用肉眼观察就可以发现很细微的缺陷。迹痕有三种:连续线表示裂纹,点线表示狭窄的裂纹,分散的点痕表示孔隙。

在操作过程中,要特别注意,为了使渗透液尽可能多的渗入缺陷中并且防止对渗透液的污染以致降低灵敏度,零件表面必须清除干净。其次,在清洗零件表面残存渗透液时,既要使残液清除干净,又要防止吸入缺陷中的渗透液流失,影响检验灵敏度。

2. 渗透检测的应用

在工业生产中,液体渗透检测常被用于工艺条件试验,成品质量检验和设备的局部检查等。它可以用于检验非多孔性的黑色和有色金属材料以及非金属材料,用来显示下列各种缺陷:

- (1) 铸件表面的裂纹、缩孔、疏松和气孔等;
- (2) 锻件、轧制件和冲压件表面的裂纹、分层和折叠等;
- (3) 焊接件表面的裂纹、融合不良、气孔等;
- (4) 塑料、陶瓷、玻璃等非金属材料 and 器件的表面裂纹等缺陷;
- (5) 金属材料的磨削裂纹、疲劳裂纹、应力腐蚀裂纹、热处理淬火裂纹等;
- (6) 各种金属、非金属容器泄漏的检查;
- (7) 设备检修时局部检查。

液体渗透检测法不适于检查多孔性材料或多孔性表面缺陷,因为缺陷显示的图像难以判断。

渗透检测能测出的最小尺寸,是由检测剂的性能、检测方法、检测操作的好坏和试件表面的粗糙度等因素决定的,一般约为深0.02 mm,宽0.001 mm。另外,在荧光渗透检测时,若使用荧光辉度高的渗透液,在检测的同时把交变应力加在试件上,可进一步提高检测灵敏度。

9.2.2 磁粉检测

磁粉检测是应用较早较广的一种无损检测手段,主要用于检测铁磁性材料(零件)表面和近表面的各种缺陷,对裂纹、夹层和未焊透等缺陷极为敏感。

这种检测的基本原理是当铁磁材料磁化时,如其表面或近表面存在缺陷,则在相应部位必然发生漏磁场,用适当的仪器或显示物质表明这种漏磁场的存在,即可达到检测的目的。最常用的方法有磁粉法、检测线圈法、磁场测定法和磁带记录法,其中以磁粉法应用最广泛。

磁粉法,采用的是三氧化二铁,四氧化三铁等强磁性粉末,工作时被漏磁场所吸引从而显示缺陷的位置和形状。为增强显示效果,磁粉中掺入荧光剂,在紫外光线照射下发出黄绿色荧光。检测线圈法是使线圈沿被测工件表面移动,遇到缺陷处的漏磁场时,由于感应电流的产生而能显示缺陷的所在。磁场测定法则是利用磁电转换元件(如霍尔元件、磁敏二极管等)或差接线圈来感知缺陷漏磁场的存在,可以获得较高的灵敏度和工效,往往用于成批钢材的检验。磁带记录法就是在大型焊制钢管检验磁化时,把具有高剩磁磁粉的磁带贴在焊缝上,如有缺陷漏磁场就会记录在磁带上,磁带上的检测信息经处理即可在荧光屏上显示,以供判读。

影响磁粉检测灵敏度的因素很多,其中主要因素如下:

(1) 零件本身的磁化特性和形状:磁化不充分,零件表面的细小缺陷不能清晰显示;磁化过强时则可能出现伪像。在检测时,将零件磁化到饱和或近饱和状态,使零件达到充分磁化,可以得到较高的灵敏度。

(2) 零件表面粗糙度:零件表面的粗糙度数值越小,清洁程度越好,检测灵敏度就越高;反之,灵敏度下降。

(3) 零件的尺寸、形状和缺陷的方位：零件表面的缺陷越大越深，磁化时，产生的漏磁场越强，检测灵敏度越高。埋藏在表面下的缺陷，距表面越远，则灵敏度越低。

(4) 被检测零件的磁场强度和方向：当裂纹方向与磁感应线垂直时，可获得最大的检测灵敏度，而当裂纹方向与磁感应线相一致时，缺陷可能完全不被显示。

(5) 使用的磁粉剂：磁粉的磁性越好粒度越细，检测灵敏度越高，磁粉粗大易于沉淀且难以被漏磁场吸附。另外，磁粉的颜色与被检工件应有足够的对比度，例如对黑皮工件和内孔检查应采用荧光磁粉进行检测。

(6) 操作使用方法：工件表面的镀层或涂层对检测灵敏度有一定影响。磁粉检验工序一般在涂镀工序之前进行。

磁化的时间不宜过长，每次磁化要通电 3 ~ 5 次，每次通电的理论磁化时间为 1/200 s，实际操作可通电 1 ~ 2 s。

零件的各种磁化方法如图 9-4 所示，纵向磁场可用细旋线圈或磁轭产生（见图 9-4(a)，(b)），在零件的内部和外部形成纵向磁力线，它能把零件上的横向缺陷检测出来，图 9-4(b) 中的零件缺陷是 A 而不是 B。由纵向通过零件的电流可直接获得环形磁场，零件上纵向裂纹与环形磁力线正交，这些裂纹把磁路切断，因而能被检测出来，如图 9-4(c)，(d)，(e) 所示。图 9-4(e) 说明被测零件的间接磁化，图 9-4(d) 为直接磁化。

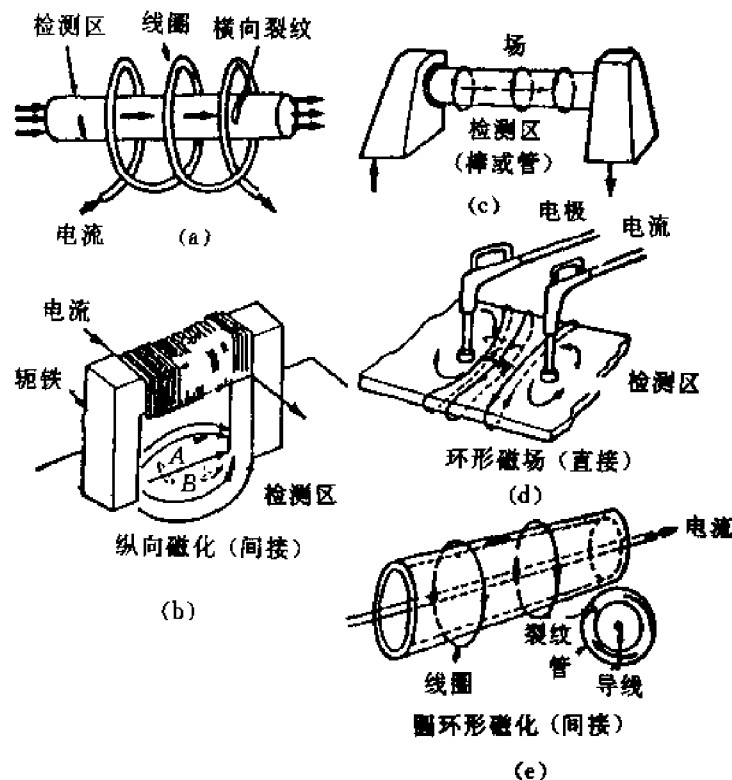


图 9-4 零件的各种磁化方法

在磁化电流方面，交流、直流、整流交流和脉冲直流四种类型都能应用。由于交流电产生的磁场集中在金属表面，所以最适合于检查表面裂纹，而直流电用于检查金属表面下隐藏裂纹最为适宜，其他两种电流用于检查金属表面下隐藏的缺陷也是很有效的。通常情况下，用较低的

电压源提供电流,而这个电流量应能满足检测的需要,但如果电流太大会使磁粉紧密堆积,缺陷不能清楚地显示出来。通过反复试验总结经验就能确定各种情况下的最佳磁化电流。我国及其他工业国家对磁化电流均有标准规定。

磁粉检测优点在于:①能直观地显示出缺陷位置、形状、走向和沿表面方向的尺寸,并能大致确定缺陷的性质;②具有很高的检测灵敏度;③几乎不受工件大小和几何形状的限制,通过采用多种磁化方法,能检测到零件的各个部位。但它局限于:①只能用于检查铁磁性材料,不能用于检查铝、镁、钛、铜及合金和奥氏体不锈钢等非铁磁性材料;②只能用来检查材料或工件的表面及近表面缺陷,而不能检查埋藏较深的内部缺陷,可检测的近表面缺陷埋藏深度一般不超过1~2 mm。

9.3 涡流检测

9.3.1 涡流检测基本原理

涡流检测是以电磁感应为基础的,它的基本原理是:通以交变电流的检测线圈靠近导电试件时,由于线圈交变磁场的作用,试件中会感生出涡流,涡流又产生使检测线圈阻抗发生变化的反作用磁场;由于试件表面或近表面缺陷的存在,会使涡流的大小,分布和流动形式等发生畸变,相应地涡流产生的反作用磁场也发生变化,通过检出由于反作用磁场的变化而引起的检测线圈阻抗的变化,便可检测出试件中的缺陷。

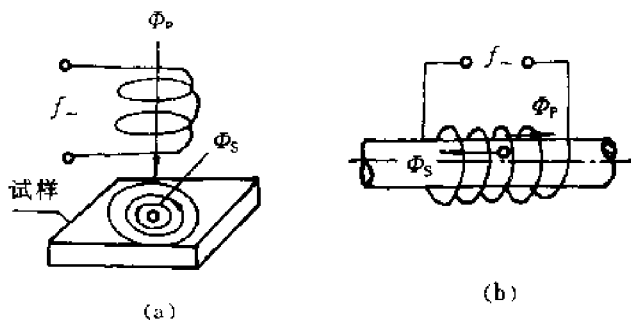


图 9-5 探测线圈与试件放置图

(a) 探头线圈放置于试件上 (b) 试件放置于穿过式线圈内

Φ_0 —线圈在未放置试件时的初级磁场; Φ_s —试件中由涡流产生的次级磁场

如图 9-5,在线圈中通以交变电流,就会产生交变磁场 Φ_p 。若将试件(导体)放在线圈磁场附近,或放在线圈中,试件在线圈产生的交变磁场作用下,就会在其表面感应出旋涡状的电流,称为涡流。涡流又产生一交变反磁场 Φ_s 。根据楞次定律, Φ_s 的方向与原有激励磁场 Φ_p 的方向相反。 Φ_p 与 Φ_s 两个交变磁场叠加形成一个合成磁场,使线圈内磁场发生了变化。因而流经线圈的电流 I 也跟着变化。如果加于线圈两端的电压 V 恒定,则

$$I = \frac{V}{Z} \quad (9.1)$$

式中, I 为流过线圈的电流(A); Z 为线圈阻抗(Ω)。

由公式(9.1)可知电流 I 随线圈阻抗 Z 的变化而变化,而

$$Z = R + j2\pi fL \quad (9.2)$$

式中, f 为线圈中交变电流频率(Hz); R 为线圈直流电阻(Ω); L 为线圈电感(H)。

如在线圈中没有放置试件,它在空气中的阻抗为 Z_0 ,以图 9-6 中 P_0 点表示。

$$Z_0 = R_0 + j\omega L_0 \quad (9.3)$$

当线圈放置在试件附近时其阻抗为 Z_1 ,以 P_1 点表示。

$$Z_1 = R_1 + j\omega L_1 \quad (9.4)$$

图 9-6 为试验线圈阻抗平面图。线圈阻抗由 P_0 点变至 P_1 点,和空载线圈的磁场 Φ_p 与涡流产生的反磁场 Φ_s 叠加之后的变化相一致。

磁场的改变导致了试验线圈阻抗改变,涡流磁场的大小与试件导电率 σ 、试件直径 d 、磁导率 μ 以及试件中的缺陷(裂纹或气孔等)有关,由此可见,涡流磁场直接反映出材料内部性能的信息,只要测量出线圈阻抗的变化,就可以测量出材料有关信息。但涡流检测线圈测出的阻抗变化是各种信息的综合。若需要测出材料内部某一特定信息(如裂纹)时就必须依靠线圈的设计以及仪器的合理组成,抑制掉不需要的干扰信息,突出所需要检测的信息。一般是将检测头线圈接收到的信号变成电信号输入到涡流仪中,进行不同的信号处理,在示波器或记录仪上显示出来,以表示材料中是否有缺陷。如试件表面有裂纹会阻碍涡流流过或使它流过的途径发生曲扭,最终影响涡流磁场。使用检测线圈便可把这些变化情况检测出来。

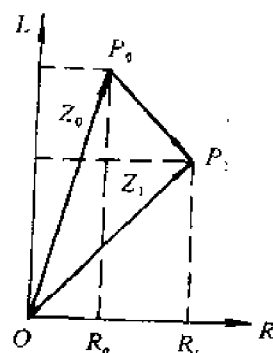


图 9-6 试验线圈阻抗图

缺陷对线圈阻抗的影响可看作是电导率和几何尺寸两个参数影响的综合结果,因此,它的效应应该介于电导率效应和直径效应之间。

在实际试件中,缺陷的状况是各种各样的,对于缺陷位置、深度和形状等综合影响所产生的缺陷效应无法进行理论计算。通常都是通过试验,取得在各种材料中不同形状、尺寸和位置的缺陷、在不同检测频率下的结果,制成参考图表,以便为实际检测提供依据。

9.3.2 涡流检测的特点

1. 涡流检测的主要优点

(1) 涡流检测适用范围广。涡流检测应用最多的是检测,特别适合导电材料表面(或近表面)检测。灵敏度高,可自动显示、报警、标记、记录,并常用于材质鉴别(即材料分选)、电导率测定、膜厚测定、尺寸测定等,对少数金属材料如石墨也可用涡流法进行检测。

(2) 涡流检测使用电磁场信号,因电磁场传播不需要任何介质,所以涡流检测不用耦合剂,探头可以不接触零件。因此,可以实现高速度、高效率自动检测和其他自动检测。目前对管材、棒材、丝材成批生产中涡流检测速度已高达 2 500 m/min 以上。因此,可用于现代工业高速生产线上,采用微机处理等控制手段,检测效率高。

(3) 涡流检测使用的是电磁信号,电磁场传播不受材料温度变化(居里点)的限制,因此可用于高温检测。而且探头可设计成多种多样,以适应特殊场合的要求。例如可对复杂型面的汽轮机叶片裂纹检测,进行内孔表面裂纹检测,对细小的钨丝、薄皮管材也可用涡流法检测其缺陷。

(4) 涡流检测还可根据显示器或记录器的指示,估算出缺陷的位置和大小。有的还可记录

成像。检测结果可以保存备查。

2. 涡流检测的缺点

(1) 由于涡流表面的趋肤效应,距表面较深的缺陷难以检测出来。

(2) 影响涡流因素多,如检测缺陷时其指示往往受材质变化和传送装置振动等干扰因素影响,必须采用(各种电子电路装置)信息处理方法将干扰信号抑制掉,才能显示出反映缺陷的信号。

(3) 要准确判断缺陷的种类、形状和大小是困难的,须作模拟试验或作标准试块予以对比,因此要求检测人员具有一定水平的专业知识和实践经验。

(4) 涡流对形状复杂的零件存在边界效应。检测时较困难,所以一般复杂零件很少采用涡流检测法。

9.3.3 涡流检测仪及检测图像

涡流检测仪主要由三部分组成:探头、二次仪表、工件夹和传送装置。当探头(检测线圈)通以交变电流时便产生激励磁场,使试件产生涡流,此涡流又产生反磁场。由于涡流磁场中包含有各种缺陷信息,探头可将它们收集并转变为电信号,经过二次仪表处理之后在荧光屏上显示。示波器图像是通过测量探头和导体之间空隙(空气层、绝缘涂料层等)的变化而给出的,如图9-7所示。图中,横坐标表示电阻,纵坐标是感抗,从原点 o 到任一点的阻抗向量对应着一个特定含意的数值,此值对应着不同的空气层或覆盖层涂料厚度等。通常调整示波器可以产生一个向右下方的运动迹线。图9-7的曲线是通过改变线圈到某导体表面的距离而获得的。不同导体,能产生不同的曲线,如图9-8所示。

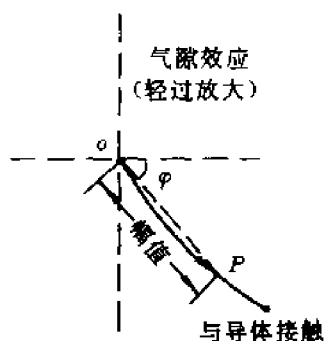


图9-7 线圈阻抗随气隙变化而变化的规律

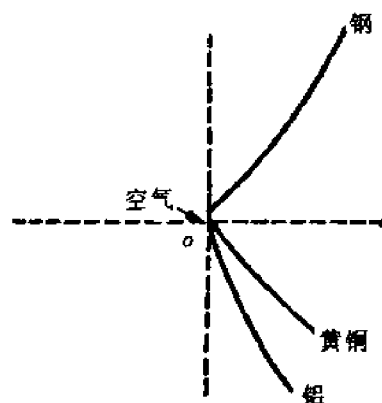


图9-8 线圈靠近不同金属表面时的阻抗变化

涡流检测技术的应用基础是惠斯登电桥原理(见图9-9(a))。电桥的每个臂上都有两个阻抗元件。开始时先将两臂调平衡,使表针指零。当指针偏离零位时,就说明作为探头的那个阻抗元件打破了平衡。

绝对分析和差动分析两种方式都可用在涡流检测技术中。每种方式都使用了两个阻抗均为 Z_0 的线圈。

绝对分析,是将一个线圈置于探头内,移动到试件的上方或穿过试件进行探查,而另一个线圈或者跟电子设备相接,或者与邻近的一个无缺陷的而形状材质完全相同的工件相联,作为参考、补偿系统(见图9-9(b))。在这种情况下,第一个线圈相对于第二个线圈的阻抗变化(Z_1

对 Z_0) 就是探头所测实况的度量。这种方法称作绝对式涡流检测分析。它非常适合检测因振动、高温气腐蚀等而逐渐产生的缺陷。但是,它对于一些与缺陷无关的因素,如沿管长方向温度差、传导率或磁导率的微量变化过于敏感。

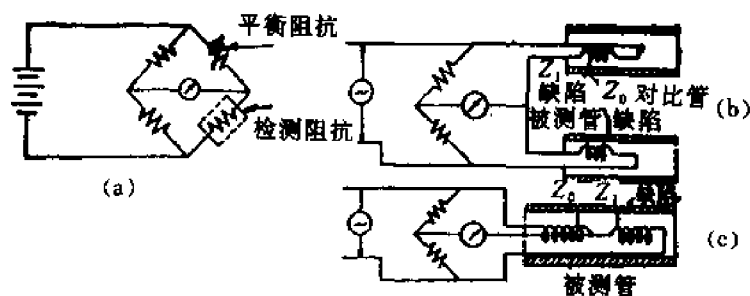


图 9-9 涡流检测电路原理

(a) 惠斯登电桥 (b) 绝对方法 (c) 差动方法

差动涡流检测克服了上述过高的敏感性。在差动检测中,旋向相反的两个线圈放在同一个探头内,如图 9-9(c),二者产生相背的磁场,结果趋于互相抵消,而当其中任一线圈遇到试件裂纹时,就有信号输出。例如,一个差动探头在一个表面有微小裂纹的管材中移动,当探头接近裂纹区时,前面的那个线圈将首先感应面产生不平衡磁场。两线圈完全通过缺陷时,将描绘出图 9-10(b) 所示的连续环形曲线,下弯和回返段相继描绘出来。当缺陷处在两线圈正中间时,曲线恰好回到原点。当继续移动探头时,产生相反变化的磁场从而描绘出上半个环形,构成一个“8”字形曲线。通过分析这种图像,可探明缺陷的位置和大小。

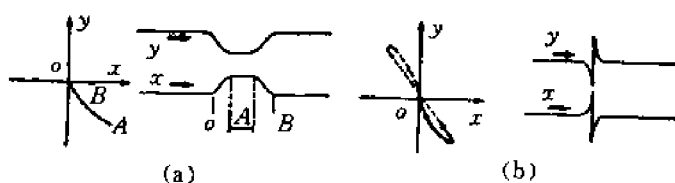


图 9-10 涡流检测 CRT 显示曲线

(a) 绝对式 (b) 差动式

在绝对涡流检测中幅值大小分析有其局限性,因为信号的弦长受缺陷的宽度、形状和方位的影响,而这些因素对于薄壁零件而言不像缺陷深度那么重要。这样一来,对每一构件如果需要确切地检测出其缺陷的特征,就要事先确定出一些校正的标准以便表示出各种遇到的缺陷类型、几何形状和尺寸。相比之下,相位角的变化仅依赖于缺陷的深度(见图 9-11)。由于缺陷深度是关键,是昂贵装置抗破坏性能的最重要因素,所以对于复杂情况检测往往就首先采用相位分析。

9.4 射线检测

前述各种无损检测法用于检测表面或近表面的缺陷。但更多的深藏缺陷还需要穿透性更强的方法和能反映各元件内部结构或状态变化的装置来检测。能量的吸收或发射是这些方法

的共同特点。

射线检测是应用最广泛的整体检测技术。工业上广泛应用的射线检测有 X 射线检测和 γ 射线检测。X 射线和 γ 射线均是电磁波，以 $3 \times 10^8 \text{ m/s}$ 的恒定速度在空间传播。X 射线和 γ 射线的粒子分别称为 X 光子和 γ 光子。光子是不带电的，具有一定的能量。射线检测中所说的射线能量，指的是 X 光子和 γ 光子的能量。X 射线和 γ 射线具有以下性质：

- (1) 不可见，依直线传播；
- (2) 它本身不带电，不受电场和磁场的影响；
- (3) 具有很强的穿透力，能穿透可见光不能穿透的物质，其中包括肌肉、骨骼、黑纸和金属等；
- (4) 具有反射、干涉、衍射、折射等现象，但这些现象与可见光有区别；
- (5) 能使物质产生光电子和返跳电子及诱发荧光 X 射线，以及引起散射现象；
- (6) 能被物质吸收，强度减弱；
- (7) 能使某些物质发生光化学作用，使照相胶片感光；也能使某些物质发生荧光作用；
- (8) 能使物质电离；
- (9) 能引起生物效应，杀伤生物细胞。

9.4.1 X 射线检测

X 射线对各种物质有不同的穿透能力，其衰减率是由所给试件的厚度、密度、原子位置(序数)决定的，因此，辐射强度能够揭示试件的异常密度(过大或过小)。用胶片或荧光屏来反映射束量的多少，为人们提供了物质内部结构状况的图片(见图 9-12)。

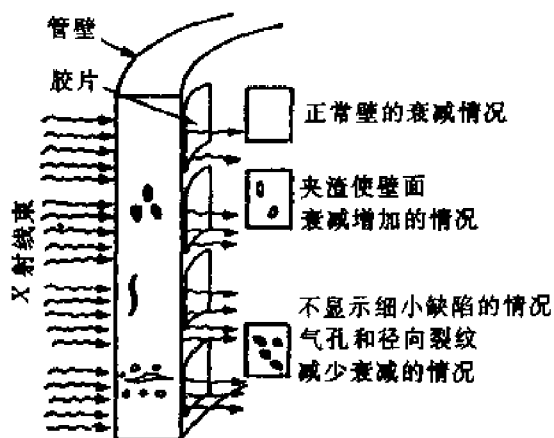


图 9-12 X 射线检测试件缺陷示意图

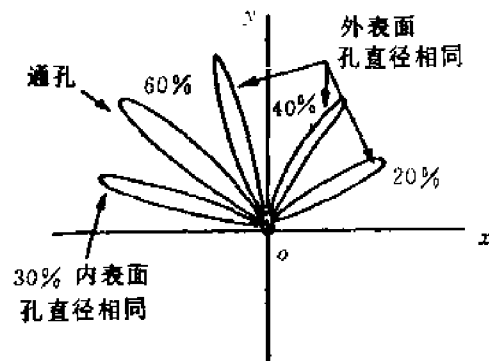


图 9-11 相位解与缺陷深度

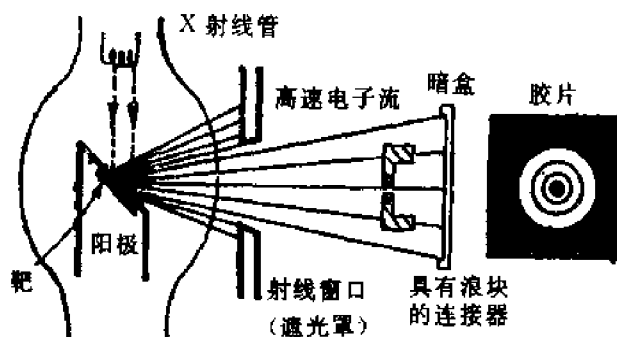


图 9-13 便携式 X 射线机原理图

X 射线在特制的真空管中产生，真空管包括一个加热灯丝和一个靶(见图 9-13)，靶一般由钨制成，它嵌在阳极中。当真空管两端施加高电压时，灯丝释放高能电子束，电子束受驱使，

射向靶子,当电子束在阳极靶上猝然停止时,阳极靶便发射 X 射线。管电压越高,X 射线的能量越大,穿透力越强。阴极灯丝的电流越大,X 射线强度也越强。X 射线谱随管电压和管电流的变化曲线如图 9-14 所示。这两种情况都会增加 X 射线分辨内部缺陷的能力。

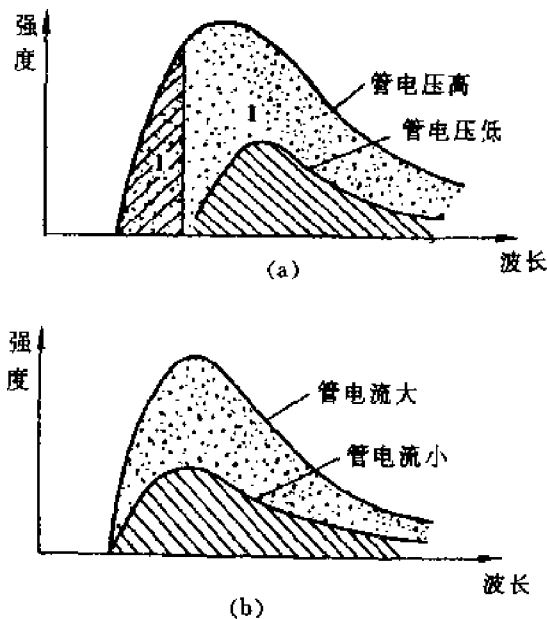


图 9-14 X 射线谱随管电压和管电流的变化
(a) 管电流一定 (b) 管电压一定

检测厚壁的区域或密度大的材料(如钢、黄铜)需要较高能量,约 2×10^5 eV 以上,而对铝件或其他轻金属件, $1 \times 10^5 \sim 1.5 \times 10^5$ eV 的能量就足够了。不管施加多高电压,几何条件对射线照片清晰度的影响是相同的。其要点如下:

- (1) 在靶上的聚集点要小(相对于缺陷的尺寸);
- (2) 将 X 射线管与检测区的距离拉大些,不致于因射线束散射而使其强度过分减弱;
- (3) 使胶片与被测件极为接近;
- (4) 为了对缺陷进行准确分析,透照射线束方向应尽可能与试件表面垂直;
- (5) 胶片应与最主要表面平行。因为缺陷部位常常不能预测,所以需要在同方向上多次曝光。

9.4.2 γ 射线检测

一般说来,X 射线源体积较大而难以放入大多数部件中。 γ 和中子辐射源之所以得到应用,正在于其体积小、射线强。 γ 射线是由放射性元素的原子核衰变后发射出来的。有些放射性元素,如镭,能自然衰变产生 γ 射线。其他元素则通过人工激发(通过辐射或裂变)产生 γ 射线,例如,钴 60 和铯 137 就是如此,它们正在取得广泛的应用。

γ 射线源的对比度很高,而且 γ 射线源体积很小,可以放在许多小工件内部,射线可以从里向外射,提高了透照图像的清晰度。能量为 1.2M eV 的钴 60 γ 射线源可与能量相同、但结构庞

大的 X 射线源相媲美,可对核蒸汽发生器管支持盘中微小裂纹进行检测。

中子源在两方面不同于 X 和 γ 射线源,它在核相互作用或裂变中产生,且具有粒子性而不是波动性。它亦可用于射线照相检测,特点是:中子的吸收(率)并非其途径材料原子密度的直接函数。所以,某些轻元素对中子吸收作用比许多重元素大。例如,氢元素、锂元素对中子的吸收作用明显高于铜元素、铁元素。所以,当对含有轻元素杂质的重金属厚壁零件拍射线图片时,可获得高反差的影像。

各种射线源和设备的功能非常广泛,这种技术已处于无损检测的前列。仅为工件厚度 2% 的裂纹可被迅速探出。严格控制各临界参数,可使灵敏度达到 0.5% ~ 1%。应用这种方法,可检测出结构缺陷和装配中的差错。它在铸件、锻件、管道的质量检查方面有广泛应用。射线检测用于核反应堆元件焊缝的检查已成为常规工序。它是在建造整个核反应装置后的维护业务中全面检查其焊缝质量的基础,也是检查现场安装质量、维修质量和寿命预测的依据。

9.5 超声波检测

9.5.1 检测原理

由于可闻声波频率不高($16 \sim 20 \times 10^4$ Hz),波长太大,用于机器零件检测时可能在缺陷周围发生绕射而不能检测出损伤。超声波之所以适于检测,是因为它处于极高的频率之中,频率在 $2 \times 10^4 \sim 10^{10}$ Hz 范围的声波称为超声波,实际用于检测的超声波频率在 400 kHz ~ 25 MHz 之间,其中用得最多的是 1 ~ 5 MHz,波长甚短,对普通缺陷不会发生绕射漏检问题。它是根据介质点的振动方向和波动传播方向不同的关系来区分的,超声波可以分为纵波、横波、表面波、板波等。

持续时间无穷的波为连续波,脉冲波则是持续时间(通常是微秒数量级)有限的波。超声波检测中绝大多数使用脉冲波。把超声波脉冲从探头射入被检物,如果其内部有缺陷,则一部分入射的超声波在缺陷处被反射,利用探头能接收信号的性能,可以不必损坏被检物面检出缺陷的部位及其大小。这种检测方法称为超声波检测。

在超声波检测法中,目前应用最多的是脉冲反射法。这种检测方法的理论基础是回声原理。图 9-15 为脉冲反射法的原理示意图。把脉冲振荡器发生的电压加到探头的压电晶片上,晶片因在厚度方向产生伸缩而发生机械振动,发出超声波脉冲。如果被检物是铁和铝的话,超声波就以 600 m/s 的固定速度在内部传播,如果被检物有裂纹等缺陷,超声波脉冲的一部分就从缺陷处反射回到探头的晶片(叫做缺陷回波),不碰到缺陷的超声波脉冲则在被检物底面

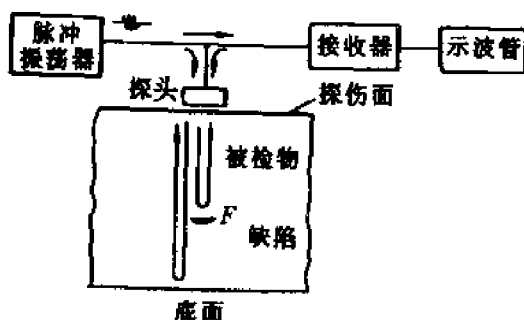


图 9-15 脉冲反射法的原理

反射回来(叫做底面回波),因此,缺陷处反射的超声波先回到晶片,底面反射回来的超声波后回到晶片。回到晶片的超声波使晶片发生振动,在晶片的两电极间就会产生频率与超声波相

等,强度与超声波成正比的高频电压,高频电压通过接收器进入示波管。通常在超声波检测中使用一个探头,它既做发射又做接收。这样,进入接收器内的高频电压包括两部分:一部分是振荡器发生的高频电压,另一部分是回波转换的高频电压。因此,当在示波管横坐标上以脉冲振荡器的起振时间为基点,把辉点向右移动时,在示波管上可以得到如图9-16所示的波形图。在这个波形图上,就可以判断有无缺陷,缺陷的部位及其大小。缺陷的部位 L_F 可以根据比例求出,缺陷的大小可根据回波高度 h_F 来估计。当缺陷很大时,可移动探头,按显示缺陷的延伸尺寸求出。

为了使超声波很好的传入被检物,探头与被检物表面之间应涂上耦合剂,被检物表面光滑时,可使用机油、合成浆糊和水作耦合剂;表面粗糙时,可使用甘油或者水玻璃作耦合剂。

超声波检测的优点主要是:设备轻巧,操作方便,成本低,灵敏度高,检测速度快,对人体无害。既适合在制造厂生产线上成批检查,也可以用于野外作业对机器零部件检查。B型、C型扫描检测仪还可以显示缺陷的形状,直观可靠。但一般用的A型扫描仪器只能以脉冲波形显示缺陷,判读困难,需要经过专门训练才能掌握好。

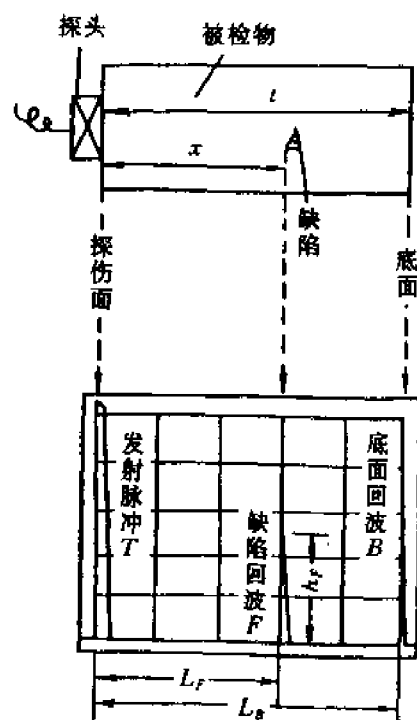


图9-16 超声波探伤图形

9.5.2 超声波探头

主要由压电晶片、外壳等组成。按其工作角度可分为直探头、斜探头、可变角探头以及聚焦探头等。

图9-17(a)所示的直探头又称平探头,可以同时发、收纵波。它由压电晶片、保护膜、阻尼吸收块、外壳和电缆接插电极等组成。保护膜的作用是避免晶片磨损。保护膜厚度最好为超声波在保护膜材料中传播的半波长的整数倍,这时声波穿透率最大。阻尼吸收快能把晶片周围

(除向工件发射面之外)的声能吸收掉,使电脉冲停止后晶片很快停止振动。

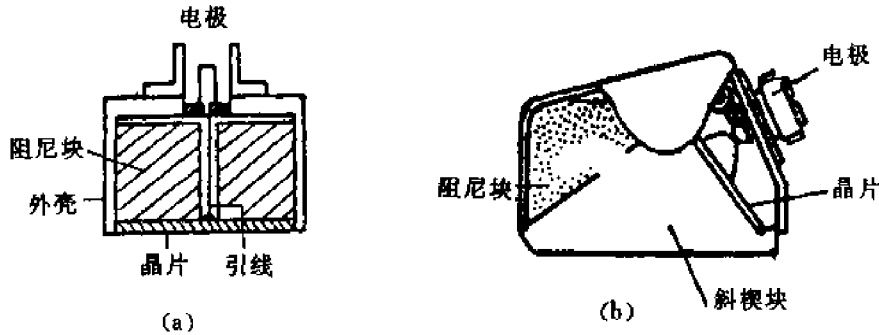


图 9-17 探头结构
(a) 直探头 (b) 斜探头

斜探头又称横波探头,可以同时发、收横波。它由吸收块及斜楔块等组成,如图 9-17(b)所示。自晶片发出来的纵波经斜楔块与工件接触进入工件,折射后变成横波。而由楔块底面反射回去的超声波则被吸收块吸收掉。斜楔块上和晶片接触的表面与和工作接触表面之间的夹角成为探头角度,即超声波入射到工件中的入射角。此角一般取值在第一与第二临界值之间,以保证去掉纵波,保留横波而且避免产生表面波。

可变角探头是指入射角可以按需选择变化的探头,其具体结构又有多种不同形式。

聚焦探头,特点是指向性好,其结构是在平探头晶片的辐射面上安装有机玻璃或环氧树脂制成的凹球面声透镜,它可以使晶片发出的超声波束聚焦在一定距离处。

9.5.3 超声波检测仪

1. 电路

超声波检测仪主要电路构成如图 9-18 所示。由同步电路产生周期性同步脉冲信号,分两路进行同步触发:一路触发时间扫描电路产生锯齿波电压,使示波屏产生一水平扫描线;另一路触发高频脉冲发射电路,使激发一种幅度大、时间短的脉冲作用在晶片上以激发超声波。发射电路即高频脉冲发生器产生超声波电振荡,加大晶片上激发超声波,经耦合媒介射入工件,在遇到缺陷或工件界面时,即反射回到探头晶片转变成电信号输入到接收放大器。放大器把发射脉冲和反射脉冲进行放大、检波、抑制后加于示波管垂直偏转板上以显示始波、底波和伤波等。

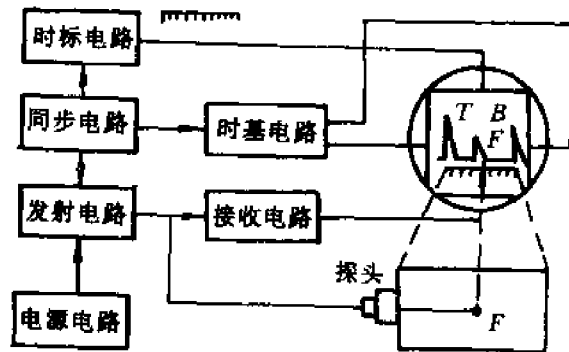


图 9-18 超声波检测仪框图

2. 检测仪显示缺陷方式

(1) A 型显示(A 扫描): A 型扫描显示使用最广泛,它可以在荧光屏上以脉冲高度来显示

缺陷大小,根据脉冲位置(水平扫描线上)来判断缺陷深度和位置。它的使用范围广,可用于纵波、横波检测。设备较简单、轻便。但它检测是以波形高低位置为依据,不直观,难于准确判断。较易对缺陷定性判断,而定量判断需要经验,较困难。

(2)B型显示(B扫描):B扫描的图像,可以显示工件在截面上缺陷的断面形状,即缺陷在此截面上的范围、深度、大小。B型仪器原理如图9-19。

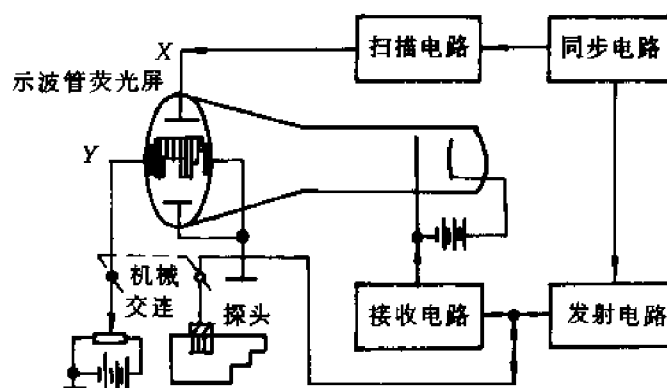


图 9-19 B型显示检测仪

B型显示使用越来越多,为了有利于检测自动化和不使探头磨损,常采用浸液法工作方式。它和A型扫描仪器结构基本相同,不同之处是探头在工件表面移动和图示电位计可变接点是连动的。此电位计上加有直流电压,正的可变端与示波管垂直偏转板Y相接。因此探头沿工件移动时,垂直偏转板上电压随着改变。这就使水平扫描线在荧光屏上随着垂直移动,因此可以把沿探头路线所切割的工件断面情况显示出来。

荧光屏X坐标代表工件深度,纵坐标代表探头在工件上的位置。当扫描电路产生的锯齿波电压加在示波管水平偏转板X上时,对水平扫描线进行亮度控制,使荧光屏上扫描线看不出或隐约可见,即使示波管不发射电子或发射电子很少。当超声波在工件底部或缺陷处反射回来时,通过放大器放大后加到示波管的阴极(反射波为负)或栅极(反射波为正)上。使阴极发射电子或使电子密度增大,在荧光屏上显示出亮点,亮点在X坐标上的位置代表深度。因此由荧光屏上显示出来的断面图形,便可看出工件被探头扫描断面上有否缺陷和显示出工件轮廓。

(3)C型显示(C扫描):C型显示可以显示出工件内部的缺陷平面图形,其原理如图9-20所示。

由两个电位计(分别代表X轴和Y轴)的直流电压来控制示波管水平和垂直偏转极。使示波管荧光屏的横坐标和纵坐标分别代表工件表面的横坐标和纵坐标。两个电位计和探头连动。当探头在工件表面移动时,偏转电压相应改变,使荧光屏上光点的坐标位置与探头在工件表面的坐标位置相对应。

C型检测仪接通电源工作时,同步电路向两处输出触发信号,一个输到发射电路,使它产生高频振

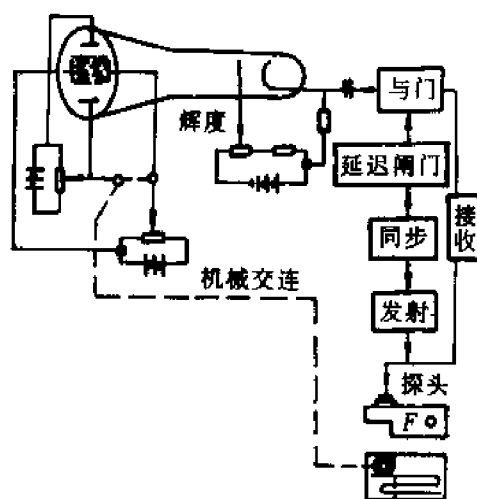


图 9-20 C型显示检测仪

荡脉冲,加到探头晶片上产生超声波,另一个同时加到深度电路,使深度电路工作。此深度电路是由延迟电路及闸门电路组成。触发信号经过延迟电路延迟一段时间后使闸门电路工作,由闸门电路产生一可调范围的矩形电压波,加到与门上。当接收电路接收到始波 T_0 、伤波 F 及底波 B 时,经放大后加到与门电路上。当闸门矩形波来到时,与门电路才有输出。调整延迟电路延迟时间及闸门时间,使只有缺陷波才能通过与门。与门电路输出的缺陷波加到示管上,使荧光屏发出亮点。这样探头在工件上扫描时便能把缺陷在平面上的图形显示在荧光屏上。

3. 检测仪的使用

A 型显示在实际检测时,除了采用前述的脉冲反射法外,还经常采用透射法和共振法。

透射检验用在试件的前后表面均可接近的场合。它除了需用两个独立的转换发射器和接收器之外,与上述技术相同。裂纹、气孔等可使声束发生散射,故而减少穿透的振动幅度,据此可以测得材料缺陷的数目,但不能给出其位置信息。这种工序用在薄壁元件的检测上,而对其采用脉冲反射法是不够精确的。这是因为工件的厚度太小,振波聚焦成束又需要一定行程,故不能精确测量裂纹位置。

共振技术可以单面测试壁厚。任何一块材料在声波沿其厚度方向传播中都有自己固有的共振频率,此频率取决于在其中传播的声速和壁厚。在实际测试中加在试件上的声频率是在一个连续范围内变化的(扫频),当壁厚等于共振频率半波长的整数倍时,试件就会发生共振,此时,转换器上能量增加,同时在 CRT 上显示波峰,通过事先校验的 CRT 刻度即可读出壁厚值。

上述装置能用在细晶粒的铸件和锻件上。有内部缺陷的锻件,假设在材料加工的碾轧垂直方向进行检测,则最容易,因为内部缺陷多与之垂直(见图 9-21)。因此,直通射束测试最适于检验轧制板材,特别是叠层材料。然而对近表面层的检验能力不强而且耗费时间较多。斜向波束测试克服了这些限制。它也用于重型盘、导管、管道等,带来巨大收益的应用是涡轮机转子孔洞附近裂纹的检测和尺寸的估计。在平板形或圆筒形零件中超声波以适当角度传播时可在表面间重复多次反射达到较远距离(见图 9-22)。脉冲反射效应将显示沿途任何地方遇到的缺陷,经过适当校准其位置可以读出,定位测试中横、纵波都可以使用,但是上述两者则不能同时使用。横波通常适用于这类检测,在对管道进行检测时,壁厚不能超过管子直径的 20%,而另一方面横波以一标准角度($45^\circ, 60^\circ, 70^\circ$)之一传播,将不会发现内壁或接近内壁的缺陷(见图 9-23)。

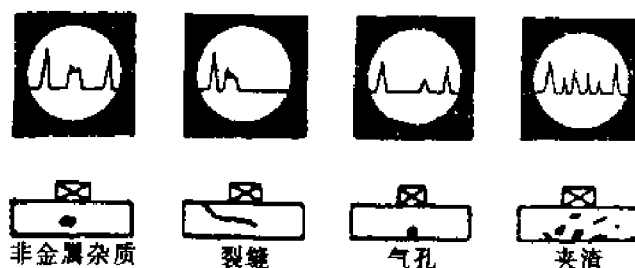


图 9-21 反射信号与缺陷形式及尺寸的关系

另一试验技术是把转换器和试件浸入水中,它可克服通常情况下难以提供优质振动耦合的困难。因为水是一种很好的耦合剂,故不需直接接触,这种方法效率高,最适于规则剖面的内部扫描,而不必改变转换器和楔形体结构,这种方法尤其适于粗糙工件。如把大型试件浸放水

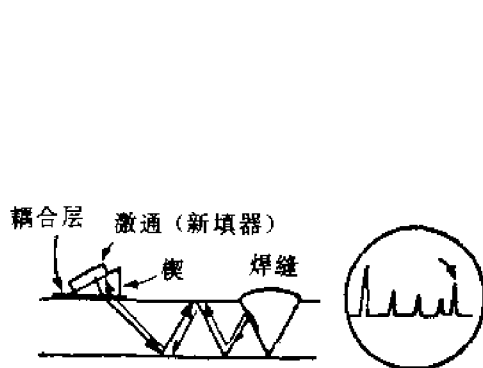


图 9-22 内表层的长距离检测

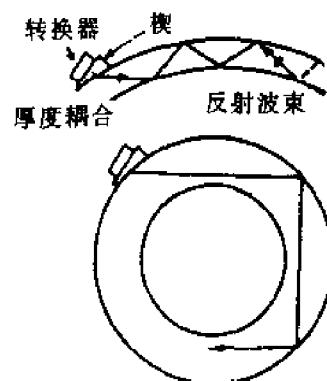


图 9-23 环形焊件的角度超声束检测

中有困难,等效方法是探头晶体界面上施加泡沫水流。

超声波检测仪的检测灵敏度要用标准试块进行调整,把标准试块人工缺陷的反射波高度调整到适当的高度,或者把被检物无缺陷部位底面的反射波调整到适当的高度。两种方法各有所长,一般检测钢板和焊缝时使用前者,检测铸锻件时则两者都用。

超声波检测对于平面状的缺陷,不管其厚度如何薄,只要超声波是垂直射向它时,就可以取得很高的反射波。但另一方面,对球形缺陷,假如缺陷不是相当大,或者不是较密集的话,就不能得到足够的缺陷反射波。因此,对钢板的层迭、分层和裂纹的检测率是很高的,而对单个气孔的检测分辨率则很低。

9.6 声发射检测

9.6.1 声发射与检测

金属材料由于内部晶格的位错,晶界滑移,或者由于内部裂纹的发生和发展,均要以弹性波的形式释放出应变能,这种现象称为声发射。各种材料声发射的频率很宽,从次声频、声频到超声频。所以,声发射也称作应力波发射。声发射是一种常见的物理现象,如果释放的应变能足够大,就产生听得见的声音,如锡片受力弯曲时就可听见劈啪声,这就是锡受力产生孪生变形的声发射。但多数金属材料受力发生塑性变形或断裂时的声发射信号很微弱,人耳难以查觉,需要借助电子仪器放大处理后才能被检测出来。

单个声发射的持续时间短、频带宽。其高频成分穿过物体时迅速衰减,而低频成分又容易和机械噪声相混不易分辨。因此,通常声发射检测都选择在某一频率段内进行,例如,金属材料声发射研究选 $10^5 \sim 10^6$ Hz 频段。

声发射检测与多数无损检测的区别表现在两方面:其一,多数无损检测方法是射线穿透检测,被检零件处于静止,被动状态;而声发射检测是动态无损检测,只有被检零件受到一定载荷,有开放形裂纹发生和发展的前提下才会有声发射可以接受,因而声发射检测可以实时地反映缺陷的动态信息,实现状况监测和险情报警;其二,多数无损检测方法是射线按一定途径穿透试件,而声发射是试件本身发射的弹性波,由传感器加以接收。因此,接收到的信号,其幅值、

相位、频率不能直接表征声发射源发出的信号。

声发射探头工作时不必移动、操作简便、灵敏度高、不受材料种类的限制。

声发射检测的技术关键之一是排除背景噪声的干扰。

声发射源发出的声信号以弹性波的形式向四周传播,到达材料表面即转变为表面波,经过耦合剂到达传感器的可以是表面波、横波和纵波及其混合形式,由传感器把它们转变为电信号,通过声发射仪接受并处理,最后显示数据和图形。

9.6.2 声发射信号的特征和计量

声发射信号可分为连续型和突发型两类。

声发射信号的典型波形如图 9-24 所示, V_p 为最大输入输出幅度, t_r 为上升时间, V_i 是门檻电压(常取为反射波或其他形式波幅高)。传感器每振荡一次输出的一个脉冲称为一个振铃。把各脉冲振铃顶峰相连形成包络线,包络线的每一主要起伏(V_i 值以上)称为一个事件。事件的宽度以 t_e 表示, t_i 为事件时间间隔, $t_e + t_i$ 为事件持续时间。

目前所采用的表征参数都是通过对仪器输出波形的处理而得到的。这些参数主要有声发射事件、振铃计数和总计数、幅度及幅度分布、能量及能量分布、有效电压值、频率、波形和声源位置坐标等。声发射波形的能量定义为将声发射信号的幅度平方,然后进行包络检波,求出包络检波后的包络线所围的面积,即可用以衡量信号所含能量的大小。

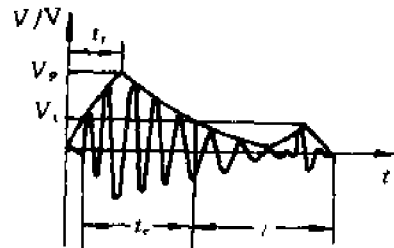


图 9-24 声发射信号的波形

声发射源的定位是依据传感器接收到的声发射信号来测定物体内部产生声发射的裂纹位置,定位的根据是信号到达各个传感器的时间差和先后顺序。

为了在固体材料表面某一范围测量出缺陷的位置,可以将几个压电换能器按一定的几何关系放置在固定点上,组成换能器阵列以测定声发射源发射的声波传播到各个换能器的相对时差。将这些相对时差代入满足该阵列几何关系的一组方程求解,便可以得到缺陷的位置坐标。在实践中,为了推得求解声源位置的计算方程式并简化计算,换能器通常是按特定规则的几何图形布置的。声发射源定位方法有直线定位法、归一化正方阵定位法、平面正方形定位法、平面三角形定位法等。

归一化正方阵定位法是一种将声源位置坐标按换能器位置坐标归一化的定位方法。如图 9-25 所示,将四个换能器置于直角坐标系中的位置 $(1,1)$, $(-1,1)$, $(-1,-1)$, $(1,-1)$, 声发射源 $P(x,y)$ 的声波到达换能器 1 的传输时间为 t_1 , 而换能器 2,3,4 相对于换能器 1 的时差为 $\Delta t_2, \Delta t_3, \Delta t_4$, 那么, $P(x,y)$ 应该位于分别以换能器 1,2,3,4 的位置为圆心,以 $vt_1, v(t_1 + \Delta t_2), v(t_1 + \Delta t_3), v(t_1 + \Delta t_4)$ 为半径的 4 个圆的交点上。4 个圆的交点只有一个,所以方程组只能有惟一解。

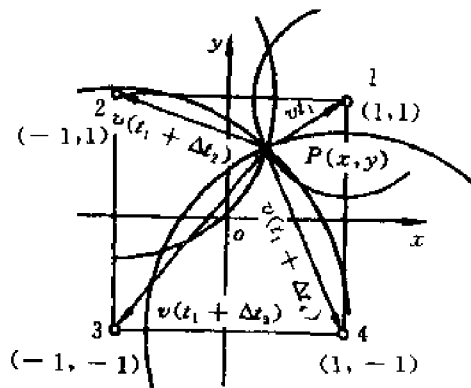


图 9-25 归一化正方阵定位方法

平面正方形定位法是将换能器置于直角坐标系中正方形四个顶点的定位方法；平面正三角形定位法是将三个换能器布置在正三角形的顶点，再将另一换能器置于正三角形的内心，并取为直角坐标系的原点。

关于上述各种定位方法的数学模型和计算方法，已有成熟的处理软件，使用十分方便，这里不再介绍。

由于声波在介质内部传播的过程中因内摩擦而衰减，以致声发射信号上升到最大幅度的时间拖长，从而，影响利用时间差进行声源定位的精度。

9.6.3 声发射检测仪器

声发射检测原理如图9-26所示。来自试件的声波被传感器接收，转变为电信号，经过前置放大器放大，滤波后提高信噪比，主放大器将信号进一步放大。信号处理器包括事件形成、方波整形，然后提供给事件计数；还包括振铃形成、方波整形，然后提供给振铃计数。计数器以数字显示或经数-模转换进而驱动X-Y记录仪绘出图形。

按接收声发射信号的通道数目，声发射仪可分为单通道、双通道和多通道三类。

声发射传感器多数是由锆钛酸铅(PZT-5)压电陶瓷片制成的压电纵波传感器，这类传感器通常是不加阻尼衬垫的。差动传感器可以消除电气噪声的干扰，它是由一块压电陶瓷片对半切开后制成的。

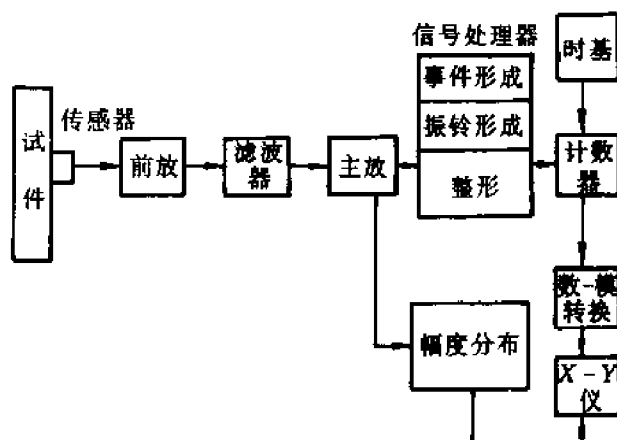


图 9-26 声发射仪工作原理

9.6.4 声发射检测的应用

在工业生产中，声发射检测主要应用在以下几方面：

(1) 金属塑性变形的声发射分析：金属屈服时的变形、晶粒的不均匀变形、某些合金受力超过屈服极限而产生的加工硬化，都伴随着声发射现象。

(2) 评价表面渗层的脆性：如钢材表面渗碳、渗氮、渗硼、碳氮共渗和离子氮化等工艺，用声发射与力学性能试验相结合的方法评价渗层质量要比传统的维氏硬度压痕试验更有效。

(3) 断裂韧性的声发射分析：最早的应用是检测裂纹失稳扩展产生的突发声发射，以此时的相应载荷定为开裂点。

(4) 检测疲劳裂纹扩展:可以早期发现疲劳裂纹,与断裂力学计算相结合可以预测零件寿命。用声发射检测疲劳裂纹有连续监测法和间歇监测法,后者适用于背景噪声高的情况。

此外,控制焊接质量、评价压力容器安全性、泄漏监测、运转机械的状况监测,以及内部放电监测等都是声发射检验的典型应用场合。

目前,声发射技术在结构完整性的检测方面已获得十分广泛的应用。对于运行状态下构件缺陷的发生和发展进行在线监测,声发射方法已经成为不可缺少的手段。构件裂纹的发生和发展,压力容器水压试验的指示,氢脆和应力腐蚀裂纹,中子辐射脆化,周期性超载和应变老化,焊接质量的监测以及声图像分析等。

习题与思考题

- 9-1 无损检测常用那些方法?试比较各检测方法的优缺点。
- 9-2 渗透检测包括哪几个过程?该检测方法的特点是什么?
- 9-3 试叙述磁粉检测的基本原理。磁粉法和检测线圈法有什么相同和不同之处?磁化试件时应考虑什么因素?
- 9-4 涡流检测的基本原理是什么?该方法依据检测线圈何种物理量的变化进行检测分析的?
- 9-5 差动涡流检测法与绝对涡流检测法相比有什么优点?
- 9-6 在工业生产中,采用射线检测法检测时,为什么常用 X 射线检测和 γ 射线检测?
- 9-7 γ 射线检测与 X 射线检测相比,有何特点?
- 9-8 实际用于超声波检测的声波频率在什么范围内?为什么?超声波检测中目前最常用的是何种方法?该方法是如何进行检测的?
- 9-9 超声波检测仪有哪几种显示缺陷的方式?
- 9-10 声发射检测与多数无损检测有什么区别?是否可在有背景噪声下进行声发射检测?
- 9-11 声发射检测采用哪些表征参数?
- 9-12 说明声发射检测中归一化正方阵定位法。

参考文献

- 1 屈梁生,何正嘉. 机械故障诊断学. 上海:上海科技出版社,1986
- 2 黄文虎等. 设备故障诊断原理、技术及应用. 北京:科学出版社,1997
- 3 徐敏. 我国设备故障诊断技术发展简史与未来. 全国设备诊断技术学术会议论文集—95(武汉).7—15
- 4 雷继尧,何世德. 机械故障诊断基础知识. 西安:西安交通大学出版社,1989
- 5 袁海德. 冲击振动计量与测试. 武汉:湖北科学技术出版社,1985
- 6 卢文祥,杜润生. 工程测试与信息处理. 武汉:华中理工大学出版社,1992
- 7 刘金环,任玉田. 机械工程测试技术. 北京:北京理工大学出版社,1990
- 8 张安华. 机电设备状态监测与故障诊断技术. 西安:西北工业大学出版社,1995
- 9 邝朴生等. 现代机器故障诊断学. 北京:农业出版社,1991
- 10 汤和,徐滨宽. 机械设备的计算机辅助诊断. 天津:天津大学出版社,1992
- 11 李常禧. 电力设备诊断技术概论. 北京:水利电力出版社,1994
- 12 邝朴生,蒋文科等. 设备诊断工程. 北京:中国农业科技出版社,1997
- 13 钟秉林,黄任. 机械故障诊断学. 北京:机械工业出版社,1997
- 14 何岭松等. 小波分析及其在设备故障诊断中的应用. 武汉:华中理工大学学报, Vol. 21, No. 1, 82—87, 1993(2)
- 15 耿中行. 小波分析方法及其在机械状态监测信号处理中的应用. [学位论文]. 西安:西安交通大学, 1993
- 16 程正兴译. 小波分析导论. 西安:西安交通大学出版社,1995
- 17 Mallat, S. , A Theory of Multiresolution Signal Decomposition; the Wavelet Representation. IEEE Trans. Pattern Anal. Machine Intell. 11, 674—693, 1989
- 18 徐佩霞,孙功宪. 小波分析与应用实例. 合肥:中国科学技术大学出版社,1996
- 19 王江萍,王鸿飞. 基于小波多分辨分析的往复机械故障特征提取与识别. 西安:西安石油学院学报 Vol. 13, 30—32, 1998(1)
- 20 王鸿飞,王江萍. 小波分析在柴油机监测信号处理中的应用. 西安:机械科学与技术, 440—441, 1998(3)
- 21 汪金鑫. 随机过程. 西安:西安交通大学出版社,1987
- 22 杨叔子,吴雅. 机械故障诊断的时序方法. 西安:西安交通大学出版社,1989
- 23 沈清,汤霖. 模式识别导论. 长沙:国防科技大学出版社,1991
- 24 边肇祺等. 模式识别. 北京:清华大学出版社,1987
- 25 温熙森等. 模式识别与状态监控. 长沙:国防科技大学出版社,1997
- 26 吴今培. 模糊诊断理论及其应用. 北京:科学出版社,1995
- 27 郭桂蓉. 模糊模式识别. 长沙:国防科技大学出版社,1993
- 28 萧筱南. 实用模糊数学. 北京:亚洲出版社,1993
- 29 李孝安等. 神经网络与神经计算机导论. 西安:西北工业大学出版社,1994
- 30 程相君等. 神经网络原理及其应用. 北京:国防工业出版社,1995
- 31 赵振宇等. 模糊理论和神经网络的基础与应用. 北京:清华大学出版社, 南宁:广西科学技术出版社, 1996
- 32 刘雄等. 转子监测和诊断系统. 西安:西安交通大学出版社,1991
- 33 张正松等. 旋转机械振动监测及故障诊断. 北京:机械工业出版社,1991
- 34 王江萍. 机械故障信号主分量分析的最大熵谱分析. 西安:机械科学与技术, 980—982, 1998(6)

- 35 王江萍等. 柴油机故障诊断技术的现状与展望. 西安:机械科学与技术,878—882,1997(5)
- 36 杨建国,周铁尘. 内燃机振动监测与故障诊断. 大连:大连海运学院出版社,1994
- 37 王江萍等. 柴油机运行状态监测信息获取试验研究. 西安:西安石油学院学报 Vol. 13,57—61,1998(3)
- 38 李柱国. 油液分析诊断技术. 上海:上海科学技术文献出版社,1997
- 39 严新平. 油液监测技术应用软件的发展与展望. 全国设备诊断技术学术会议论文集—93(石家庄),508—512
- 40 严勇年等. 应用铁谱和光谱技术监测柴油机磨损状况. 荆门:石油机械 Vol. 20, No. 4, 37—40, 1992
- 41 杨其明,周长春. 内燃机车柴油机光谱铁谱监测系统. 第四届全国机械设备故障诊断学术会议论文集(1994 杭州),15—21
- 42 C. BA. 安邦贤. 根据示功图诊断柴油机. 国外内燃机车 93(11),51
- 43 郝志勇等. 内燃机气缸压力振动识别研究. 天津:内燃机学报 Vol. 10, No. 1, 43—48, 1994
- 44 杨建国等. 车用发动机缸内部件状态监测与故障诊断. 武汉:武汉水运工程学院学报, Vol. 17, No. 1, 66—71, 1993
- 45 谭达明等. 柴油机工作过程故障振动诊断的基础研究. 天津:内燃机学报 Vol. 10, No. 4, 341—346, 1992
- 46 刘忠文等. 故障状态系数及其应用. 武汉:武汉水运工程学院学报 Vol. 12, No. 4, 59—63, 1988
- 47 杨建国,周铁尘. 利用振动信号诊断柴油机主要运动部件的磨损状态. 第四届全国机械设备故障诊断学术会议论文集(1994 杭州),548—553
- 47 杨建国,周铁尘. 用表面振动信号诊断 135 型柴油机活塞—气缸套的磨损状态. 第四届全国机械设备故障诊断学术会议论文集(1994 杭州),532—537
- 48 黄宜谅等. 内燃机瞬时转速的研究. 天津:内燃机学报 Vol. 10, No. 2, 55—160, 1992
- 49 陈祥初等. 柴油机瞬时转速检测及其在故障诊断中的应用. 第四届全国机械设备故障诊断学术会议论文集(1994 杭州),577—581
- 50 沈仲贤,邓昆. 用飞轮瞬时转速诊断缸内工作过程. 第四届全国机械设备故障诊断学术会议论文集(1994 杭州),582—585
- 51 裴峻峰,杨其俊. 机械故障诊断技术. 东营:石油大学出版社,1997, 4
- 52 能源部西安热工研究所主编. 热工技术手册, Vol. 6, 电厂金属. 北京:水利电力出版社,1989, 11
- 53 《机械(动力)师手册》编委会编著. 机械(动力)师手册. 北京:机械工业出版社,1996, 12
- 54 黄长艺,严普强主编. 机械工程测试技术基础. 北京:机械工业出版社,1995
- 55 王江萍等. 五柱塞泵机组振动试验测试与振动特性分析. 兰州:石油矿场机械,11—22,2001(1)
- 56 王江萍等. 五柱塞泵运行状态振动测试与诊断. 第一届全国诊断工程技术学术会议文集(1998 武汉),552—555