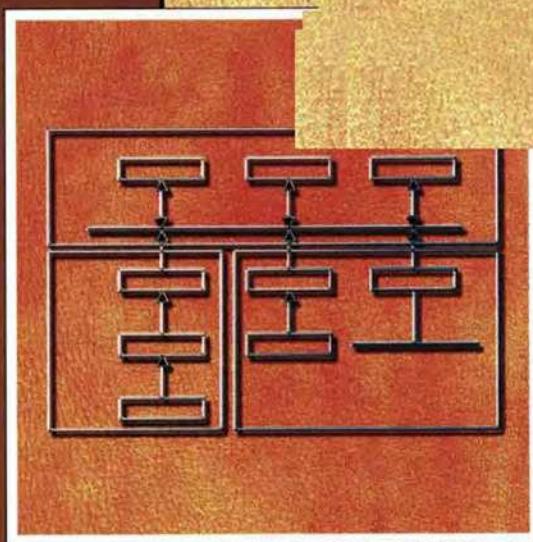




研究生系列教材

智能传感器 系统(第二版)



刘君华 主编
汤晓君 张勇 编著
李世维 胡红利 编著

西安电子科技大学出版社
<http://www.xduph.com>

研究生系列教材

- 通信网的安全——理论与技术 王育民等(教育部推荐教材)
 自适应信号处理 沈福民
 数字信号处理——时域离散随机信号处理 丁玉美等
 随机信号处理 陆光华等
 高性能数字信号处理与高速实时信号处理 苏涛等
 实时信号处理系统设计 苏涛等
 多传感器数据融合及其应用 杨万海
 时频测控技术 周渭等
 数字视觉视频技术 李玉山
 计算机图形学 穆柏青等
 微波电路CAA与CAD 王家礼
 电磁波时域有限差分方法(第二版) 葛德彪等(教育部推荐教材)
 电磁场有限元方法 金建铭
 微电子器件可靠性 史保华等
 数值分析 宋国乡等
 数值泛函与小波理论 冯象初等
 应用泛函分析原理 李广民等
 应用统计学教程(上、下册) 赵玮等(教育部推荐教材)
 近代解析应用数学基础 王长清
 随机过程 毛用才等 随机过程 张卓奎等
 组合数学 马光思 组合数学(第二版) 姜建国等
 非线性光学 石顺祥等(教育部推荐教材)
 光的电磁理论——光波的传播与控制 石顺祥等
 红外物理 张建奇等
 现代光学 刘继芳
 光电探测原理 安毓英等
 预测与决策 李华等
 科技英语写作教程 秦获辉
 英语听力教程 秦获辉
 科技英语阅读高级教程 秦获辉
 自然辩证法概论教程 顾民等
 科学社会主义理论与实践 汪向阳等
 高等教育管理导论 柏昌利
 非线性控制理论及应用 贺显曜等
 工程电动力学(修订版) 王一平
 移动地理信息系统开发技术 李斌兵
 现代密码学 杨晓元
 智能传感器系统(第二版) 刘君华

ISBN 978-7-5606-2380-1



9 787560 623801 >

定价：41.00元



研究生系列教材

智能传感器系统

(第二版)

刘君华 主编

汤晓君 张 勇 李世维 胡红利 编著

西安电子科技大学出版社

内 容 简 介

本书在第一版的基础上对智能传感器系统及相应智能化技术进行了与时俱进的全面阐述，重点突出三个方面：一是概述智能传感器系统硬件的两种实现形式，即传感器（经典的或现代的）经不同集成度调理电路芯片与 CPU 相结合的虚拟仪器形式以及与 MPU 相结合的微计算机/微处理器形式；二是全面概述了基本智能化功能软件模块的实现技术，并依次介绍了多种经典和新兴的信息处理技术作为智能化技术工具的原理与方法，同时又注意介绍其软件模块在两种形式智能传感器系统中的实现；三是介绍新型的模糊与无线网络智能传感器系统。

本书内容丰富、新颖，反映了该领域前沿的最新技术，既具先进性又具实用性；文字简洁、深入浅出，避免了深奥的理论阐述与生涩的数学推导，但不失说理性，便于自学。本书可作为高等院校研究生教材，也可作为进行科学的研究、工程实验的科研工作者及工程技术人员的参考书。

图书在版编目(CIP)数据

智能传感器系统/刘君华等著. —2 版. —西安：西安电子科技大学出版社，2010.5

(研究生系列教材)

ISBN 978 - 7 - 5606 - 2380 - 1

I. 智… II. 刘… III. 智能控制—传感器—研究生—教材 IV. TP212.6

中国版本图书馆 CIP 数据核字(2010)第 003366 号

策 划 云立实

责任编辑 张晓燕 云立实

出版发行 西安电子科技大学出版社(西安市太白南路 2 号)

电 话 (029)88242885 88201467 邮 编 710071

网 址 www.xduph.com 电子邮箱 xdupfxb001@163.com

经 销 新华书店

印刷单位 陕西天意印务有限责任公司

版 次 2010 年 5 月第 2 版 2010 年 5 月第 4 次印刷

开 本 787 毫米×1092 毫米 1/16 印张 25.5

字 数 601 千字

印 数 1~3000 册

定 价 41.00 元

ISBN 978 - 7 - 5606 - 2380 - 1 / TP · 1197

XDUP 2672002 - 1

* * * 如有印装问题可调换 * * *

本社图书封面为激光防伪覆膜，谨防盗版。



自 1999 年本书第一版问世以来，智能传感器系统以及与之伴随的智能传感器技术均以前所未有的速度蓬勃发展。因此，著者对智能传感器一书再版的内容做了大幅度调整。

作为实现智能传感器系统重要基石的集成化技术，更是发展迅猛异常，不但已有与不同种类传感器适配的、不同集成度的调理电路芯片商品，更有含微处理器的单片全系统集成的器件芯片商品大量流行于市。本版书中割舍了与这些集成调理电路、典型集成电路元件制造工艺有关的集成技术，割舍了现代传感器技术中与制作典型微结构以及微机械工艺主要技术有关的内容，这部分内容其他相关专著会比我们阐述得更为清楚。我们重点介绍这些不同集成度、适配不同类型传感器的集成调理电路以及单片全系统集成器件芯片的技术性能、指标与其使用方法，为读者构建自己的智能传感器系统提供一种新的思路与新的途径。

智能传感器/变送器仍然是现场控制总线(FCS)系统中的主要角色。其发展势头与发展规模随着自动化技术的发展更加波澜壮阔。现场总线控制系统是智能传感器最大的工业实用舞台，在智能传感器实用化进程中具有首屈一指的重要作用。对于现场总线控制系统中的智能传感器/变送器，本版书割舍了在使用安装中涉及的有关具体技术问题的介绍；相应的智能化功能，如刻度转换与非线性自校正、自校零与自校准、温度补偿、PID 控制等，均放在有关章节中统一介绍；其智能化通信功能，由于与现场总线流行的各种通信协议密切相关，相关的技术培训班可以更好地解决使用中所需具备的具体技术问题，故本版书中不再介绍。书中增加了在网络化与通信功能发展中，重要性日益凸现的无线网络智能传感器系统。

赋予传感器系统以“智能”的智能传感器技术，随着信息处理技术的飞速发展而快速地发展着。智能传感器工作者辛勤而有成效的工作，使那些经典的和新兴的信息处理技术相继不断地被挖掘、移植与引入到智能传感器系统这一新的领域中来，成为实现智能化功能新的有效技术手段。本书力求反映这种形势。书中涉及的每一种信息处理技术都已有相应的专著论述，本书无力仿效它们去讲述理论本身，而是本着拿来为我所用的宗旨，着重介绍应用这些理论实现智能化功能的基本原理与方法，为使用智能化技术奠定基础。

本书的特色表现在每章之后的示例中。全书的示例都以提高传感器稳定性、消除交叉敏感的影响这一智能化功能作为目标，而且均针对压阻式压力传感器来介绍不同信息处理手段。所需的信息处理知识都在相应章节基础知识部分提供，无需借助其他专著，其目的在于介绍它们作为智能化技术工具入门的基本思路，为读者举一反三、扩大应用的广度与深度留以广阔的空间和无限的余地。传感器的种类数不胜数，新的信息处理技术也将层出不穷，期望广大读者善于捕捉与挖掘。

全书共分 13 章。第 1 章为概述；第 2 章为智能传感器系统中的经典传感技术基础；第 3 章为智能传感器系统的组建与集成调理电路芯片介绍；第 4 章为基本智能化功能与其软

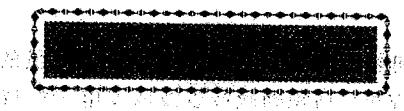
件实现；第 5~11 章分别介绍多元回归分析法、神经网络技术、支持向量机技术、粒子群优化算法、主成分分析技术、小波变换、线性相位滤波器与自适应滤波器等及其在智能传感器系统中的应用；第 12 章为智能模糊传感器系统；第 13 章为无线网络智能传感器系统。智能传感器系统硬件模块集成实现的方法在第 3 章介绍。本书重点介绍了传感器经不同集成度调理芯片与 CPU 相结合的虚拟仪器形式实现和与 MPU 相结合的微计算机/微处理器实现两种方式。顺应发展形势，本次再版更加突出了智能传感器智能化功能的实现技术，第 4~12 章列举了多种基于信号分析与处理技术的智能化功能软件模块设计方法，这些软件模块可以在 LabVIEW 或/和 MATLAB 软件平台上实现，同时也注意介绍将它们移植到内存容量小的微计算机/微处理器式智能传感器系统中复现的思路与方法。这些软件模块都是经过实践检验可实现的，它们都是我们课题组不同历史时期科研成果在教学中的反映。

本书力求内容的基础性与先进性相结合，基础理论与智能化功能相结合，尤其是智能化功能的原理学习与工程可实现性相结合。在文字方面力求简明易懂，深入浅出；在内容方面力求避免深邃的理论阐述与生涩的数学推导，但仍不失清晰的说理性。

本书的编写要感谢西安交通大学的研究生郝惠敏、李玉军、肖丹、王颖、唐建文、文会飞、朱凌建、王影花、李小辉、刘斐、罗政元等。他们多次仔细阅读底稿，制作并反复修改书中的示例。

由于作者水平有限，若有不当之处，恳请批评指正。

编著者
2010 年 1 月



智能传感器技术是一门正在蓬勃发展的现代传感器技术，是涉及微机械与微电子技术、计算机技术、信号处理技术、电路与系统、传感技术、神经网络技术、模糊理论等多种学科的综合性技术。

在智能传感器发展进程中，由于对其“智能”含义的理解不断深化，各个时期的学者给予智能传感器的定义也随着历史的推移而演变。本书第1章将详细介绍智能传感器的发展概貌。如：

“把传统的传感器与微处理器集成在一块芯片上的传感器称为智能传感器”。这种说法强调的是在工艺上将传感器与微处理器二者紧密结合。关于这方面的内容主要反映在本书第3、4章中。

进而有“所谓智能传感器就是一种带有微处理器兼有检测和信息处理功能的传感器”的说法。这种提法突破了传感器与微处理器的结合必须在工艺上集成在一块芯片上的框框，着重于二者赋予智能的结合可以使传感器系统的功能由以往只起“信息检测”作用扩展到兼而具有“信息处理”功能。而按照传统概念，“信息处理”功能只有“仪器”才可以具有。这部分内容在本书第5、6、9章中详述。

再而有如H.Schodel、E.Beniot等人认为“一个真正意义上的智能传感器，必须具备学习、推理、感知、通信以及管理等功能。”这种观点更强调了智能化的功能。这种功能相当于一个具备知识与经验丰富的专家的能力。然而，知识的最大特点是它所具有的模糊性，所以模糊传感器是一种新型智能传感器。有关智能传感器的通信功能与模糊传感器的内容在本书的第7、8章详述。

无论怎样，智能传感器是在传统传感器技术基础上发展起来的，它与传统传感器技术仍然有着密不可分的关系（我们将在第2章详述）。但是，它绝不是传统观念中的传感器。

为了反映智能传感器的全貌，本书共分9章：第1章，概述；第2章，智能传感器系统中的经典传感器技术基础；第3章，不同集成度智能传感器系统举例；第4章，智能传感器系统的集成方法与技术；第5、6章，智能传感器系统智能化功能的实现Ⅰ、Ⅱ；第7章，通信功能与总线接口；第8章，模糊传感器；第9章，人工神经网络在智能传感器系统中的应用。

为了便于广大读者更清楚地理解书中所涉及的内容，书中力求避免深奥而又生涩的数学推导，而着重于基本要领的阐述。同时也进行必要的简明易懂的推证，以增强说理性。并列举有大量实例使其理论联系实际。书中特别注意知识层次的连贯性与系统性，在介绍智能传感器系统时有机地结合了所需的基础知识（包括经典传感器技术），以期帮助不同领域和不同知识背景的读者，在学习时不会感到大的跳跃，而能步入发展智能传感器事业的行列。

此外，朱长纯教授参加了本书大纲编写，并和博士生李炳乾参加了第4章的编写工作；

博士生常炳国参加了第6、7、8、9章的编写工作；博士生刘月明参加了第3章的编写工作。在此表示感谢。

本书编写过程中承蒙陕西省计量测试研究所赵继文研究员的大力支持，并对全书提出了许多有益意见；赵维安研究员、马积勋教授以及西安电子科技大学出版社总编李荣才教授和云立实编辑也给予了热情帮助和积极支持，叶德福编审对本书做了认真细致的审读加工，在此一并表示衷心感谢。

本书应西安电子科技大学出版社之约，结合教学和科研实践而编写。由于编者水平有限，难免有不当与错误之处，恳请读者指正，不吝赐教。

编著者

1998年5月

于西安交通大学

目 录

第1章 概述	1
1.1 传感器技术发展的重要性	1
1.2 智能传感器发展的历史背景	2
1.3 智能传感器的功能与特点	5
1.3.1 智能传感器的功能	5
1.3.2 智能传感器的特点	5
1.4 智能传感器概念与传感器系统	6
1.5 智能传感器实现的途径	7
1.5.1 非集成化实现	7
1.5.2 集成化实现	8
1.5.3 混合实现	10
1.5.4 智能传感器的几种形式	11
1.5.5 改善传感器系统性能的多传感器智能化技术	11
1.6 现场总线智能变送器/传感器经典实例简介	12
1.6.1 现场总线控制系统(FCS)中的传感器与仪表	12
1.6.2 经典实例简介	14
参考文献	20
第2章 智能传感器系统中的经典传感技术基础	22
2.1 传感器系统的基本特性与技术指标	22
2.1.1 静态特性与静态技术指标	23
2.1.2 动态特性与动态技术指标	27
2.2 几种变换器工作原理	33
2.2.1 基于压阻效应的电阻变换器	33
2.2.2 基于电容效应的电容变换器	35
2.2.3 基于固有频率变化效应的谐振式变换器	37
2.3 集成化压力传感器与加速度传感器	38
2.3.1 压阻式压力传感器	38
2.3.2 电容式压力传感器	44
2.3.3 谐振式压力传感器	46
2.3.4 加速度传感器	49
2.4 提高传感器性能的技术途径	56
2.4.1 合理选择结构、参数与工艺	56
2.4.2 基于差动对称结构的差动技术	57
2.4.3 补偿	66
2.4.4 多信号测量法	70
参考文献	73
第3章 智能传感器系统的组建与集成调理电路芯片介绍	74
3.1 智能传感器系统的基本组成形式	74

3.1.1	传感器	74
3.1.2	信号调理	75
3.1.3	数据采集与转换	75
3.1.4	计算机及其 I/O 接口设备	75
3.2	基于虚拟仪器平台实现数据采集与显示功能	76
3.2.1	数据采集卡(DAQ)的基本性能指标	76
3.2.2	数据采集卡的安装	77
3.2.3	I/O 接口设备 Lab PCI 6024E 数据采集卡简介	77
3.2.4	实现数据采集卡软件驱动前的参数设置	78
3.2.5	[示例 3-1]基于 DAQ 与 PC 实现虚拟仪器式的数据采集与显示	78
3.3	电阻电桥式传感器的单片集成调理电路——MAX1450 芯片	80
3.3.1	MAX1450 芯片的引脚功能与结构框图	80
3.3.2	MAX1450 的基本功能与补偿校准功能	82
3.4	适配压阻式传感器的单片集成调理电路——MAX1460 芯片	86
3.4.1	结构框图	86
3.4.2	功能与原理简介	86
3.5	适配变压器式传感器的单片集成调理电路——AD698 芯片简介	88
3.5.1	LVDT 变压器式传感器简介	89
3.5.2	AD698 的结构框图与工作原理	89
3.5.3	主要参数设置与使用方法	91
3.6	适配电容式传感器的集成调理电路——CAV414 芯片	94
3.6.1	CAV414 的结构框图	94
3.6.2	工作原理与引脚连接	94
3.6.3	[示例 3-2]基于 CAV414 的电容式变送器的构成	96
3.7	数据采集系统——单片集成接口芯片 ADuC812	96
3.7.1	简介	96
3.7.2	结构框图与组成模块的功能	97
3.7.3	主要参数特点	100
3.7.4	应用举例——温度的测量与显示系统	100
3.8	温度传感器系统——全系统单片集成芯片 MAX6625	102
3.8.1	MAX6625 结构框图及引脚排列	102
3.8.2	主要功能与技术指标	104
3.8.3	工作原理	104
3.8.4	MAX6625 的操作与使用	105
3.8.5	MAX6625 器件的编程举例	106
3.9	XTR101 4~20mA 回路变送器芯片	109
3.9.1	简介	109
3.9.2	结构框图及引脚功能	109
3.9.3	主要参数	110
3.9.4	应用方法举例	110
3.10	适配桥路式传感器的信号调理电路——ZMD31050 芯片	111
3.10.1	简介	111
3.10.2	功能框图与引脚排列	112

3.10.3 主要参数	113
3.10.4 应用举例	113
3.11 双轴加速度传感器系统——全系统单片集成芯片 ADXL202	114
3.11.1 简介	114
3.11.2 结构框图与功能	114
3.11.3 主要参数特点	116
3.11.4 使用说明及典型应用举例	116
参考文献	118
第4章 基本智能化功能与其软件实现	120
4.1 改善线性度及智能化非线性刻度转换功能	120
4.1.1 查表法	121
4.1.2 曲线拟合法	123
4.1.3 [示例 4-1]与铂电阻配用的智能化刻度转换模块的设计(曲线拟合法)	124
4.2 改善静态性能提高系统精度及智能化自校零、自校准功能	126
4.2.1 两基准法	127
4.2.2 多基准法	128
4.3 改善稳定性抑制交叉敏感及智能化多传感器数据融合功能	130
4.3.1 单传感器系统	130
4.3.2 交叉敏感与传感器系统的稳定性	131
4.3.3 多传感器技术改善传感器系统性能的基本方法	132
4.4 改善动态性能扩展频带及智能化频率自补偿功能	134
4.4.1 数字滤波器的数学基础——Z 变换简介	135
4.4.2 扩展频带的数字滤波法	142
4.4.3 扩展频带的频域校正法	145
4.4.4 [示例 4-2]采用数字滤波法将测温传感器(一阶系统)频带扩展 $A \geq 10$ 倍	146
4.5 提高信噪比与分辨力及智能化信号提取与消噪功能	147
4.5.1 数字滤波技术	147
4.5.2 相关	159
4.5.3 频域谱分析法	167
4.5.4 [示例 4-3]交流电桥调幅波解调器中滤波器参数的确定	168
4.5.5 [示例 4-4]设计一个巴特沃斯低通数字滤波器	170
4.5.6 [示例 4-5]设计一个巴特沃斯高通数字滤波器 $H_s(z)$	172
4.5.7 [示例 4-6]二阶带通数字滤波器在 LabVIEW 平台上的实现、 检验及向 MPU 移植与复现	173
4.6 增强自我管理与自适应能力及智能化控制功能	178
4.6.1 模拟 PID 控制器的传递函数	178
4.6.2 数字 PID 控制器脉冲传递函数	179
参考文献	181
第5章 多元回归分析法及其在智能传感器系统中的应用	182
5.1 多元回归分析法与定常系数多元回归方程	182
5.1.1 二传感器数据融合——二元回归分析法	182
5.1.2 三传感器数据融合——三元回归分析法	185
5.2 回归分析法与可变系数回归方程	187

5.2.1 工作原理	187
5.2.2 回归方程可变系数 $A_0(T) \sim A_5(T)$ 的确定	187
5.3 应用举例	190
5.4 示例	191
5.4.1 [示例 5-1] 基于回归分析模型法降低一个干扰量影响的智能化软件模块设计	191
5.4.2 [示例 5-2] 基于回归分析法模型降低两个干扰量影响的智能化软件模块设计	195
参考文献	200
第 6 章 神经网络技术与其在智能传感器系统中的应用	201
6.1 概述	201
6.2 神经网络基础知识	202
6.2.1 神经网络结构	202
6.2.2 神经元模型	203
6.2.3 神经元作用函数	203
6.2.4 BP 神经网络	206
6.2.5 径向基(RBF)神经网络	209
6.3 示例	211
6.3.1 [示例 6-1] 基于神经网络模型法温度自补偿智能化模块的设计	211
6.3.2 [示例 6-2] 基于 RBF 神经网络法抗两个干扰量影响的智能化软件模块的设计	219
参考文献	224
第 7 章 支持向量机技术在智能传感器系统中的应用	225
7.1 统计学习理论与支持向量机的基础知识	225
7.1.1 统计学习理论	225
7.1.2 支持向量机	230
7.2 支持向量机的训练、检验与测量	240
7.2.1 训练样本及检验样本的制备	240
7.2.2 支持向量机的训练	240
7.2.3 支持向量机的检验	241
7.2.4 测量	241
7.2.5 移植	241
7.3 基于 SVM 方法的三传感器数据融合原理	241
7.3.1 三传感器数据融合的智能传感器系统的组成	241
7.3.2 [示例 7-1] 降低两个干扰量影响的 SVM 智能化软件模块的设计	243
7.3.3 [示例 7-2] 使用 SVC(支持向量分类)对两组分混合气体进行定性识别	249
参考文献	255
第 8 章 粒子群优化算法与其在智能传感器系统中的应用	257
8.1 粒子群优化算法发展与应用概况	257
8.1.1 群智能	257
8.1.2 群智能的主要算法	258
8.1.3 群智能算法的特点	258
8.2 粒子群优化算法的基础知识	258
8.2.1 基本粒子群优化算法	259
8.2.2 标准粒子群优化算法	260

8.2.3 粒子群优化算法流程	261
8.3 粒子群优化算法的发展与改进简介	262
8.4 应用示例	263
8.4.1 待优化的参数与待改善的传感器	263
8.4.2 [示例 8-1]遍历优化 LS-SVM 模型参数	266
8.4.3 [示例 8-2]采用粒子群优化算法优化一个参数	269
8.4.4 [示例 8-3]采用粒子群优化算法优化两个参数	273
参考文献	277
第 9 章 主成分分析及其在智能传感器系统中的应用	278
9.1 主成分分析法	278
9.1.1 二维空间中的 PCA	278
9.1.2 PCA 算法	280
9.2 PCA 算法在消除传感器漂移中的应用	281
9.2.1 PCA 算法实现传感器故障检测的思想	281
9.2.2 [示例 9-1]建立传感器阵列、获取关联数据与漂移数据	283
9.2.3 [示例 9-2]传感器发生漂移的识别	286
参考文献	290
第 10 章 小波分析及其在智能传感器系统中的应用	291
10.1 小波分析基础	291
10.1.1 小波分析与短时 Fourier 变换	291
10.1.2 离散小波	293
10.1.3 小波级数	294
10.1.4 多分辨分析	294
10.1.5 小波包分析	297
10.2 Matlab 工具箱中小波分析函数	298
10.2.1 小波分析函数	298
10.2.2 小波包函数	305
10.3 [示例 10-1]小波数字滤波的实现	310
参考文献	314
第 11 章 线性相位滤波器与自适应滤波器	315
11.1 线性相位滤波器	315
11.1.1 线性相位与线性相位滤波器	315
11.1.2 线性相位有限冲击响应滤波器的数学模型	317
11.1.3 线性相位 FIR 滤波器的窗口设计	321
11.1.4 [示例 11-1]用窗口法设计一个线性相位低通 FIR 滤波器	328
11.1.5 [示例 11-2]线性相位 FIR 滤波器在 Matlab 中的实现	328
11.2 自适应滤波器	333
11.2.1 自适应滤波器的结构	333
11.2.2 自适应滤波理论与算法	333
11.2.3 Matlab 中的自适应滤波函数	337
11.2.4 [示例 11-3]自适应滤波在系统辨识中的应用	339
参考文献	340

第 12 章 模糊智能传感器系统	342
12.1 模糊集合理论概述	342
12.1.1 模糊集合的定义及其表示方法	343
12.1.2 隶属函数的确定方法及常用形式	344
12.1.3 模糊集合的基本运算	348
12.1.4 模糊关系的定义及合成	348
12.1.5 语言变量与模糊推理	349
12.2 模糊传感器系统	351
12.2.1 测量结果“符号化表示”的概念	351
12.2.2 模糊传感器的基本概念和功能	352
12.2.3 模糊传感器的结构	354
12.2.4 模糊传感器语言描述的产生方法	355
12.2.5 模糊传感器对测量环境的适应性	358
12.2.6 模糊传感器隶属函数的训练算法	359
12.3 示例	362
12.3.1 [示例 12-1]模糊温度测试仪	362
12.3.2 [示例 12-2]模糊温度控制仪	366
参考文献	371
第 13 章 无线传感器网络	372
13.1 概述	372
13.1.1 无线传感器网络研究与应用状况	372
13.1.2 无线传感器网络通信协议	373
13.1.3 无线传感器网络与 Internet 的互联	374
13.1.4 实现远程监测的无线传感器网络系统的典型结构	374
13.2 无线传感器网络与 Internet 的互联	375
13.2.1 基于 LabVIEW 虚拟仪器的网络化方法	375
13.2.2 [示例 13-1]设计“远程无线传感器网络监测平台”虚拟仪器程序	376
13.2.3 [示例 13-2]将本地“远程无线传感器网络监测平台”VI 制作成为网络化 VI	378
13.3 无线传感器网络	383
13.3.1 无线传感器网络中的传感器节点	383
13.3.2 无线传感器网络中的汇聚节点	384
13.3.3 [示例 13-3]工业监测用无线传感器网络的硬件芯片与引脚连接	384
13.3.4 [示例 13-4]工业监测用无线传感器网络的软件设计流程举例	390
参考文献	395

第1章 概 述

本章内容

- 传感器技术发展的重要性
- 智能传感器发展的历史背景
- 智能传感器的功能与特点
- 智能传感器概念与传感器系统
- 智能传感器实现的途径
- 现场总线智能变送器/传感器经典实例简介

通过对传感器技术发展的重要性与历史背景的概述和回顾，可以看到发展智能传感器是历史的必然。同时本章全面概述了智能传感器的功能、特点、概念与实现的途径。

1.1 传感器技术发展的重要性

传感器技术是当今世界令人瞩目、迅猛发展的一种高新技术，是当代科学技术发展和一个国家综合实力的重要标志。

它的重要性体现在以下两方面，这也是传感器迅猛发展的推动力。

(1) 近几十年来传感器技术总是一个时代技术发展进步的瓶颈。

传感器、执行器、计算机是组成近代自动化系统的三大环节。在工业自动化技术发展时代，传感器是瓶颈。因此，在微处理器开始普及的 20 世纪 80 年代，掀起了“传感器热”，日本把传感器列为 20 世纪 80 年代 10 大技术之首，美国把传感器技术列为 20 世纪 90 年代 22 项关键技术之一，甚至有“谁掌握了传感器，谁就能掌握一个时代”的说法。

当今，我们正处于信息化时代，传感器技术是信息获取科学与技术的核心技术，信息获取科学与技术又是构成信息技术的三大支柱之一，是信息的源头和基础。但是传感器目前仍是信息技术发展的瓶颈，信息获取技术(传感/检测技术)大大落后于信息处理技术(计算机技术)与信息传输技术(通信、网络技术)，所以传感器仍然居于推动科学技术进步发展的关键与基础的地位，是众多科学家、技术工作者攻坚的热点。

(2) 传感器具有广阔的市场和社会需求。

传感器的销售值反映一个国家科技发达与社会进步的程度。20 世纪 80 年代，日本、西欧市场传感器销售值年增长率为 30%~40%，英国传感器销售额 1990 年比 1980 年增长 24 倍。近十几年来发达国家传感器的产量及市场需求年增长率平均在 10% 以上。这是因为它是：

① 工业生产的倍增器。通常一部高档轿车需要 200~300 个传感器，一架飞机需要

3600 个传感器，一个发电站需要近万个传感器，一个钢厂需要 2 万个传感器……正是由于以传感器为前端的测量仪器系统保证了庞大的发电厂、化工厂、钢铁厂等的稳定生产、产品的质量和效率，从而才有巨额的产值、效益与市场的倍增。而这些带传感器的仪器仪表系统仅占企业固定资产 10%~15%，因此它们对于工业生产具有“四两拨千斤”的拉动作用。

由于传感器的使用，使生产工艺过程的控制和产品性能的检测有了保证，所以它是提高产品竞争力的强有力的手段，是获得经济效益的有效途径。据有关资料，全美电站的相关数据表明，如果主汽流量精度改善 1%，电站的燃烧成本(热效率)将会改善 1%，每年可节约 3 亿美元；若传感器及其测量仪表可利用率提高 1%，则每年可节约 30 亿美元；美国的电站采用了先进的传感器和控制技术后，使全美经济每年获益达 110 亿美元之多。

② 科学研究的“先行官”。伟大的科学家门捷列夫说过：“没有测量就没有科学”。诺贝尔奖设立至今，众多得奖科学家都是借助于先进仪器的诞生才获得重大的科学发现。例如，正是因为有了改进的仪器仪表，才使人类基因的测试提前 3 年完成。人类进一步对地下、海洋空间、宇宙星球的探索都离不开传感器及其仪器仪表系统，而且还需研制在这些极端条件下可正常工作的传感器。

③ 军事上的“战斗力”。现代武器装备几乎都配备了相关的检测传感器及其控制仪器仪表。

④ 确认证据的“物化法官”。在产品质检、环境污染监测、违禁药物检测、指纹识别、假钞与金属利器识别等方面，传感器及其仪器仪表是确认证据的科学依据。

⑤ 安全的屏障。煤矿瓦斯监测与安全预警需要高灵敏度、高稳定性、快速响应的气体传感器；气体液体传输，如我国的“西气东输”管道，需要高精度的油气计量传感器；国家领空、领海、边境安全等都需要大面积、广域传感检测系统。

⑥ 传染病预警、环境保护、可持续发展的战略保障。SARS、禽流感等突发性流行病的快速诊断需要具有快速分析能力的生物传感器；大气环境监测(SO_2 、 NO_x 、 CO_2 、微尘等)，水质监测(COD、BOD 等)，以及电磁辐射、噪声、室内有害气体(甲醛等)等的监测无不需要相应的传感器及其仪器仪表系统。

所以社会需要它！海陆空、吃穿住，传感器无处不在，也必须无处不在。

1.2 智能传感器发展的历史背景

作为获取信息工具的传感器，位于信息系统的最前端。其特性的好坏、输出信息的可靠性对整个系统质量至关重要。因此，传感器的性能必须适应系统使用的要求。

回顾自动化系统对传感器的要求，对了解智能传感器提出的背景是很有益处的。

自动化系统对传感器最基本的要求又是最急切的要求是：降低现行传感器的价格性能比。

我们知道，每种现代自动化过程，都包括有如图 1-1 所示的三种主要功能块：执行器、计算机(或微处理器)及传感器。传感器时时检测“对象”的状态及其相应的物理参量，并及时馈送给计算机；计算机相当于人的大脑，经过运算、分析、判断，根据“对象”状态偏

离设定值的方向与程度，对执行器下达修正动作的命令；执行器相当于人的手脚，按大脑的命令对“对象”进行操作。如此反复不止，以使“对象”在允许的误差范围内维持在所设定的状态。

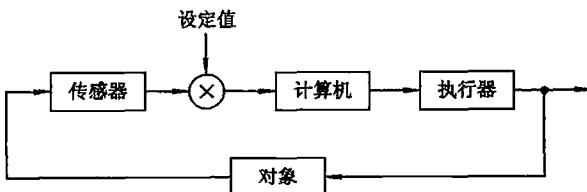


图 1-1 自动化(控制)系统框图

透视人类自动化的历史进程，我们看到：

首先，人们发展了执行器：通过制造人造工具——执行器，如水磨，以扩大机械动力；待蒸汽机出现，导致了“工业革命”。

然后，计算机出现，通过扩展计算机程序的“智能”，产生了“信息革命”。

现在，人们正在通过应用传感器来扩展“感觉”，去获得信息数据，以便校正自动化过程中的偏差，并能根据各种情况的变化作出实时、正确的处理。

在控制环路中，只有当三个功能块都经济耐用时，才可能实现工业现场的许多工作不是由人工去做，而是由自动化机械或机器人来做，甚至家务工作也可如此。这种标志人类生活主要变化的时代就叫“自动化时代”。我们正处于“自动化时代”的发展期，但是传感器技术发展的落后阻碍了自动化的进程。

从图 1-2 中我们可以看到，若以 1970 年的价格性能比为 1，则 1990 年的价格性能比是：执行器（以电动机为例）为 $1/10$ ，计算机为 $1/1000$ ，传感器为 $1/3$ 。其中，以计算机的价格性能比下降幅度最为惊人。这是由于半导体集成电路工艺的迅速发展，使大规模集成电路芯片制作成本大幅度降低。相对而言，近 30 年来计算机性能价格比的提高更是遥遥领先，而传感器的价格性能仍居高不下，与其它两个功能块的发展形势极不适应。

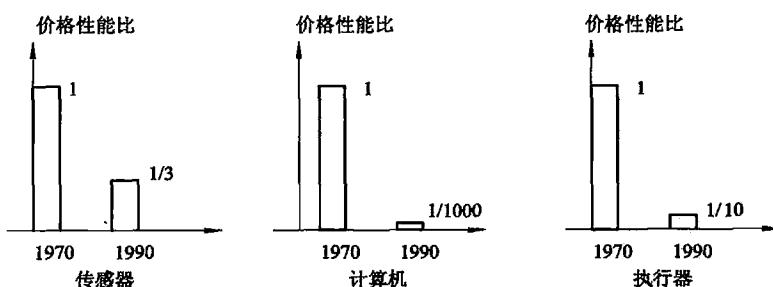


图 1-2 传感器、计算机及执行器的价格性能比

然而，传统的传感器技术已达到其技术极限。它的价格性能比不可能再有大的下降。它在以下几方面存在严重不足：

- 因结构尺寸大，故时间(频率)响应特性差；
- 输入输出特性存在非线性，且随时间而漂移；

- 参数易受环境条件变化的影响而漂移；
- 信噪比低，易受噪声干扰；
- 存在交叉灵敏度，选择性、分辨率不高。

以上不足是传统传感器性能不稳定、可靠性差、精度低的主要原因。它的“手工艺品”式的制作过程，制作材料的多品种、高性能的要求，是成本价格居高不下的原因。

自动化系统对传感器的进一步的、急切的要求是：增加品种，减小体积和重量。

随着自动化领域不断扩展，需要测量的参量日益增加，而且一些特殊领域进而需要传感器小型化和轻量化。如在线监测，除了温度、压力、流量等热工参量外，还迫切需要监测机械振动参量以及化学成分与物理成分，在线监测电站每一个燃烧器的煤粉量和一次风量，等等，不可胜数。特别是自动控制系统的飞速发展，对传感器进一步提出了数字化、智能化、标准化的紧迫需求。

生产过程自动化在经历了 20 世纪 50 年代和 60 年代的集中控制，70 年代的分散型控制系统(DCS)之后，为适应多点多参数日益复杂的大型控制系统的需要，80 年代以来出现了基于现场总线的开放型控制系统(FCS)。它是对分散型控制系统(DCS)的继承、完善和进一步发展，是继 DCS 之后自动化领域的又一次重大变革。

现场总线是连接测控系统中各智能装置(包括智能传感器)的双向数字通信网络。其主要特点分述如下。

1. 传输数字信号

用数字信号取代原来的 4~20 mA 标准模拟信号，提高可靠性和抗干扰能力。这就要求传感器由可输出 4~20 mA 标准信号的变送器改变为带数字总线接口并输出数字信号。所有现场传感器，通过数字总线接口都方便地挂接在一条环形现场总线上。这样可以大大削减现场与控制室(高/上位计算机)之间一对一的连接导线，节约初期安装费用，大大简化整个系统的布线和设计。这种节约对一个大型、多点测量系统是很有意义的，譬如：

一个电站	需要 5000 台传感器及其仪表；
一个钢铁厂	需要 20 000 台传感器及其仪表；
大型石油化工厂	需要 6000 台传感器及其仪表；
大型发电机组	需要 3000 台传感器及其仪表；
一部汽车	需要 30~100 台传感器；
一架飞机	需要 3600 台传感器；
：	

2. 标准化

总线采用统一标准，使系统具有开放性。不同厂家的产品，在硬件、软件、通信规程、连接方式等方面互相兼容、互换联用，既方便用户使用，又易于安装维修。不少大公司都推出了自己的现场总线标准。国际化的统一标准的工作正在加紧进行中。

3. 智能化

采用智能与控制职能分散下放到现场装置的原则，现场总线网络的每一节点处安装的现场仪表应是“智能”型的，即安装的传感器应是“智能传感器”。在这种控制系统中，智能型现场装置是整个控制管理系统的主体。这种基于现场总线的控制系统，要求必须使用智

能传感器，而不是一般传统的传感器。

现场总线中的智能传感器，通常称为智能变送器。这种智能传感器带有标准数字总线接口，能够自己管理自己，它将所检测到的信号经过变换处理后，以数字量形式通过现场总线与高/上位计算机进行信息通信与传递，有的也同时兼有 4~20 mA 标准模拟信号输出。

所以，智能传感器是应现代自动化系统发展的需要而提出来的，同时也是传感技术克服自身落后向前发展的必然趋势。

1.3 智能传感器的功能与特点

1.3.1 智能传感器的功能

目前还没有关于传感器智能化功能的明确定义，一般来说；可以从以下几方面概括其功能：

- (1) 在自我完善能力方面，
 - ① 具有改善静态性能，提高静态测量精度的自校正、自校零、自校准功能；
 - ② 具有提高系统响应速度，改善动态特性的智能化频率自补偿功能；
 - ③ 具有抑制交叉敏感，提高系统稳定性的多信息融合功能。
- (2) 在自我管理与自适应能力方面，
 - ① 具有自检验、自诊断、自寻故障、自恢复功能；
 - ② 具有判断、决策、自动量程切换与控制功能。
- (3) 在自我辨识与运算处理能力方面，
 - ① 具有从噪声中辨识微弱信号与消噪的功能；
 - ② 具有多维空间的图像辨识与模式识别功能；
 - ③ 具有数据自动采集、存储、记忆与信息处理功能。
- (4) 在交互信息能力方面，具有双向通信、标准化数字输出以及拟人类语言符号等多种输出功能。

1.3.2 智能传感器的特点

与传统传感器相比，智能传感器的特点简述如下。

1. 精度高

智能传感器有多项功能来保证它的高精度，如通过自动校零去除零点，与标准参考基准实时对比以自动进行整体系统标定，自动进行整体系统的非线性等系统误差的校正，通过对采集的大量数据进行统计处理以消除偶然误差的影响等，从而保证了智能传感器的高精度。

2. 高可靠性与高稳定性

智能传感器能自动补偿因工作条件与环境参数发生变化而引起的系统特性的漂移，如温度变化而产生的零点和灵敏度的漂移；在被测参数变化后能自动改换量程；能实时自动

进行系统的自我检验，分析、判断所采集到的数据的合理性，并给出异常情况的应急处理（报警或故障提示）。因此，有多项功能保证了智能传感器的高可靠性与高稳定性。

3. 高信噪比与高分辨力

由于智能传感器具有数据存储、记忆与信息处理功能，通过软件进行数字滤波、相关分析等处理，可以去除输入数据中的噪声，将有用信号提取出来；通过数据融合、神经网络技术，可以消除多参数状态下交叉灵敏度的影响，从而保证在多参数状态下对特定参数测量的分辨能力，故智能传感器具有高信噪比与高分辨力。

4. 强自适应性

智能传感器具有判断、分析与处理功能，它能根据系统工作情况决策各部分的供电情况、与高/上位计算机的数据传送速率，使系统工作在最优低功耗状态和优化传送效率。

5. 较低的价格性能比

智能传感器所具有的上述高性能，不是像传统传感器技术那样通过追求传感器本身的完善、对传感器的各个环节进行精心设计与调试、进行“手工艺品”式的精雕细琢来获得的，而是通过与微处理器/微计算机相结合，采用廉价的集成电路工艺和芯片以及强大的软件来实现的，所以具有较低的价格性能比。

由此可见，智能化设计是传感器传统设计中的一次革命，是世界传感器的发展趋势。作为商品，在 20 世纪 80 年代初期有美国霍尼韦尔公司的压阻式 ST - 3000 型压力(差)智能变送器，后有美国 SMAR 公司生产的 LD302 系列电容式智能压力(差)变送器，美国罗斯蒙特公司生产的电容式智能压力(差)变送器系列，日本横河电气株式会社生产的谐振式 EJA 型智能压力(差)变送器。此外，世界各国正在利用计算机和智能技术研究、开发各种其他类型的智能传感器/变送器。

1.4 智能传感器概念与传感器系统

传感器本身是一个系统，随着科学技术的发展，这个系统的组成与研究内容也在不断更新。人们提出“传感器系统”，是因为当前世界传感技术发展的重要趋势就是传感器系统的发展。所谓传感器系统，简单地讲就是传感器、计算机和通信技术的结合，而智能传感器系统与微传感器系统是其中的两个主要研究方向。前者着重点在如何赋予传感器系统以“智能”，后者以实现微小结构为主要目标。

智能传感器系统是一门现代综合技术，是当今世界正在迅速发展的高新技术，至今还没有形成规范化的定义。早期，人们简单、机械地强调在工艺上将传感器与微处理器两者紧密结合，认为“传感器的敏感元件及其信号调理电路与微处理器集成在一块芯片上就是智能传感器”。

随着以传感器系统发展为特征的传感技术的发展，人们逐渐发现将传感器与微处理器集成在一块芯片上构成智能传感器，在实际中并不总是必需的，而且也不经济；重要的是传感器（通过信号调理电路）与微处理器/微型计算机赋以智能的结合。若没有赋予足够的“智能”的结合，只能说是“传感器微机化”，还不能说是智能传感器。于是进而认为“所谓智

能传感器，就是一种带有微处理器兼有检测信息和信息处理功能的传感器”、“传感器(通过信号调理电路)与微处理器赋予智能的结合，兼有信息检测与信息处理功能的传感器就是智能传感器”。这些提法突破了传感器与微处理器结合必须在工艺上集成在一块芯片上的框框，而着重于两者赋予智能的结合可以使传感器系统的功能由以往只起“信息检测”作用扩展到兼而具有“信息处理”功能。而传统观念认为，仪器系统是执行信息处理任务的，即将有用信息从含有噪声的输入信号中提取出来并予以显示的装置。也就是说，“信息处理”功能仅属于“仪器”所有。目前，把一个大的仪器系统与敏感元件采用微机械加工与集成电路微电子工艺压缩后，共同制作在一块芯片上，封装在一个小外壳里构成的智能传感器系统的全集成化工艺已经实现，并有商品化器件出售。

另一方面，工业现场总线控制系统中的传感器/变送器，都是带微处理器的智能传感器/变送器。它们是形体较大的装置，既不是仅有获取信息功能的那种传统传感器，也不是只有信息处理功能的那种传统仪器。因此，传统的传感器与仪器的那种“壁垒森严”的、“不可逾越”的界线正在消失。智能传感器系统是既有获取信息功能，又有信息处理功能的传感器系统。

H. Schodel, E. Beniot 等人更进一步强调了智能化功能，认为“一个真正意义上的智能传感器，必须具备学习、推理、感知、通信以及管理等功能”。这种功能相当于一个具备知识与经验丰富的专家的能力。然而，知识的最大特点是它所具有的模糊性。20世纪80年代末，L. Foulloy 在他的论文《An Ultrasornic Fuzzy Sensor》中提出了模糊传感器的概念，他认为“模糊传感器是一种能够在线实现符号处理的智能传感器”；D. Stipanieer 也认为“模糊传感器是一种智能测量设备”。当然，这种智能传感器也不一定是全集成化的。

关于智能传感器的中、英文称谓，目前也尚未统一。John Brignell 和 Nell White 认为“Intelligent Sensor”是英国人对智能传感器的称谓，而“Smart Sensor”是美国人对智能传感器的俗称。而 Johan H. Huijsing 在《Integrated Smart Sensor》一文中按集成化程度的不同，将智能传感器分称为“Smart Sensor”和“Integrated Smart Sensor”。“Smart Sensor”的中文译名有“灵巧传感器”、“智能传感器”。本书采用智能传感器系统(Intelligent Sensor System)的称谓，简称为智能传感器(Intelligent Sensor)，并且认为“传感器与微处理器赋予智能的结合，兼有信息检测与信息处理功能的传感器就是智能传感器(系统)”。模糊传感器也是一种智能传感器(系统)，将传感器与微处理器集成在一块芯片上是构成智能传感器(系统)方式的一种。

1.5 智能传感器实现的途径

目前传感技术的发展是沿着三条途径实现智能传感器的。

1.5.1 非集成化实现

非集成化智能传感器是将传统的经典传感器(采用非集成化工艺制作的传感器，仅具有获取信号的功能)、信号调理电路、带数字总线接口的微处理器组合为一整体而构成的一个智能传感器系统。其框图如图 1-3 所示。

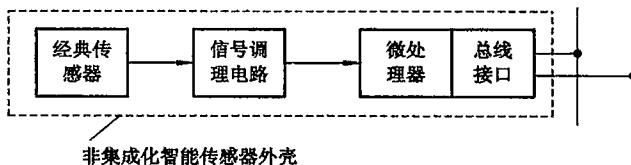


图 1-3 非集成化智能传感器框图

图 1-3 中的信号调理电路是用来调理传感器输出的信号的，即将传感器输出信号进行放大并转换为数字信号后送入微处理器，再由微处理器通过数字总线接口挂接在现场数字总线上。这是一种实现智能传感器系统的最快途径与方式。例如美国罗斯蒙特公司、SMAR 公司生产的电容式智能压力(差)变送器系列产品，就是在原有传统非集成化电容式变送器基础上附加一块带数字总线接口的微处理器插板后组装而成的，并开发配备通信、控制、自校正、自补偿、自诊断等智能化软件，从而实现智能传感器。

这种非集成化智能传感器是在现场总线控制系统发展形势的推动下迅速发展起来的。因为这种控制系统要求挂接的传感器/变送器必须是智能型的，对于自动化仪表生产厂家来说，原有的一整套生产工艺设备基本不变。因此，对于这些厂家而言，非集成化实现是一种建立智能传感器系统最经济、最快捷的途径与方式。

广大的科研工作者、工程技术人员为了提高测量精度，提高系统的性能，也都针对各自的传统传感器采用各种智能化技术，构建了自己的智能传感器系统。因此，智能传感器系统非集成化实现不仅在自动化仪表，而且在更广阔的领域都在蓬勃地发展着。

1.5.2 集成化实现

集成化实现智能传感器系统，是建立在大规模集成电路工艺及现代传感器技术两大技术基础之上的。

1. 大规模集成电路工艺技术

利用大规模集成电路工艺技术将由硅材料制作的敏感元件、信号调理电路、微处理单元集成在一块芯片上构成智能传感器系统。

然而，要在一块芯片上实现智能传感器系统存在着许多困难的、棘手的问题。例如：

- 哪一种敏感元件比较容易采用标准的集成电路工艺来制作？
- 选用何种信号调理电路，使需要外接的元件如精密电阻、电容、晶振等最少？
- 由于制作了敏感元件后留下的芯片面积有限，因而需要寻求占用面积最小的模/数转换器型式，如 $\Delta - \Sigma$ 等。
- 由于芯片面积有限制，以及制作敏感元件与数字电路的优化工艺不兼容，因而微处理器系统及可编程只读存储器的规模、复杂性与完善性受到很大限制。
- 对功耗与自热、电磁耦合带来的相互影响，在一块芯片内应如何消除？

此外，还有其它许多问题，这里不一一列举。

2. 现代传感器技术

1) 现代传感器技术的技术特征

与传统经典传感器制作工艺完全不同的现代传感器技术是于 20 世纪 70 年代开始发展

起来的。它以既有优良电性能、又有极好的机械性能的硅材料为基础，采用微米($1\text{ }\mu\text{m}\sim 1\text{ mm}$)级的微机械加工技术(包括硅的各向异性刻蚀技术，干湿法刻蚀技术，控制腐蚀深度的自停止技术，形成空腔、梁等可动三维结构的牺牲层技术，将分离件整合的键合技术等)代替制作经典传感器的车、铣、刨、磨、焊等宏观加工工艺，国外也称其为专用集成微型传感技术(ASIM)。由现代传感器技术制作出的敏感元件也是 MEMS(Micro-Electronic Mechanical System)技术的开端。由现代传感器技术制作的传感器通常称为集成传感器或固态传感器。已作为工业产品的集成传感器有 20 世纪 70 年代美国霍尼尔公司生产的硅压阻式传感器，90 年代初日本横河株式会社生产的硅谐振式传感器，90 年代末美国罗斯蒙特及日本 FUJI 公司生产的硅电容式压力(差)传感器，以及美国摩托罗拉公司生产的硅加速度传感器等。

2) 集成/固态传感器的特点

(1) 微型化。微型压力传感器已经可以小到放在注射针头内送进血管测量血液流动情况，装在飞机或发动机叶片表面用以测量气体的流速和压力。美国最近研究成功的微型加速度计可以使火箭或飞船的制导系统的重量从几公斤下降至几克。

(2) 结构一体化。压阻式压力(差)传感器是最早实现一体化结构的。传统的做法是先分别由宏观机械加工金属圆膜片与圆柱状环，然后把二者粘贴形成周边固支结构的“金属杯”，再在圆膜片上粘贴电阻变换器(应变片)而构成压力(差)传感器，这就不可避免地存在蠕变、迟滞、非线性特性。采用微机械加工和集成化工艺，不仅“硅杯”一次整体成型，而且电阻变换器与硅杯是完全一体化的，进而可在硅杯非受力区制作调理电路、微处理器单元，甚至微执行器，从而实现不同程度乃至整个系统的一体化。

(3) 精度高。比起分体结构，传感器结构本身一体化后，迟滞、重复性指标将大大改善，时间漂移大大减小，精度提高。后续的信号调理电路与敏感元件一体化后可以大大减小由引线长度带来的寄生参量的影响，这对电容式传感器更有特别重要的意义。

(4) 多功能。微米级敏感元件结构的实现特别有利于在同一硅片上制作不同功能的多个传感器。例如，压阻式压差传感器是采用微机械加工技术最先实用化的集成传感器，但是它受温度与静压影响，总精度只能达到 0.1%，致力于改善它的温度性能花费了 20 余年时间却无重大进展。美国霍尼韦尔公司发展了多功能敏感元件，率先于 20 世纪 80 年代初期研制成功的 ST - 3000 型智能变送器，就是在一块硅片上制作了感受压力、压差及温度三个参量的，具有三种功能(可测静压、压差、温度)的敏感元件结构的传感器，不仅增加了传感器的功能，而且又通过采用数据融合技术消除了温度与静压的影响，提高了传感器的稳定性与精度(详细讨论见第 4 章)。

(5) 阵列式。微米技术已经可以在一平方厘米大小的硅芯片上制作含有几千个压力传感器的阵列。譬如丰田中央研究所半导体研究室用微机械加工技术制作的集成化应变计式面阵触觉传感器，在 $8\text{ mm}\times 8\text{ mm}$ 的硅片上制作了 $1024(32\times 32)$ 个敏感触点(桥)，基片四周还制作了信号处理电路，其元件总数约 16 000 个。

敏感元件构成阵列后，配合相应图像处理软件，可以实现图形成像且构成多维图像传感器。这时的智能传感器就达到了它的最高级形式。

敏感元件组成阵列后，通过计算机/微处理器解耦运算、模式识别、神经网络技术的应用，有利于消除传感器的时变误差和交叉灵敏度的不利影响，可提高传感器的可靠性、稳

定性与分辨能力。

传感器的集成化实现是传感器的发展方向，它又是传感器向微型化、阵列化、多功能化、智能化方向发展的基础。随着微电子技术的飞速发展，大规模集成电路工艺技术日臻完善，MEMS 技术、微纳米技术、现代传感器技术协同发展，现已有不同集成度的电路芯片及传感器系统芯片商品面市(将在第 3 章中介绍)。

由于在一块芯片上实现智能传感器全系统，并不总是希望的，也并不总是必须的，所以，一种更为可行的混合实现智能化的方式迅速得到发展。

1.5.3 混合实现

根据需要与可能，将系统各个环节，如敏感单元、信号调理电路、微处理器单元、数字总线接口，以不同的组合方式集成在两块或三块芯片上，并装在一个外壳里，实现混合集成。可采用如图 1-4 所示的几种方式来实现。

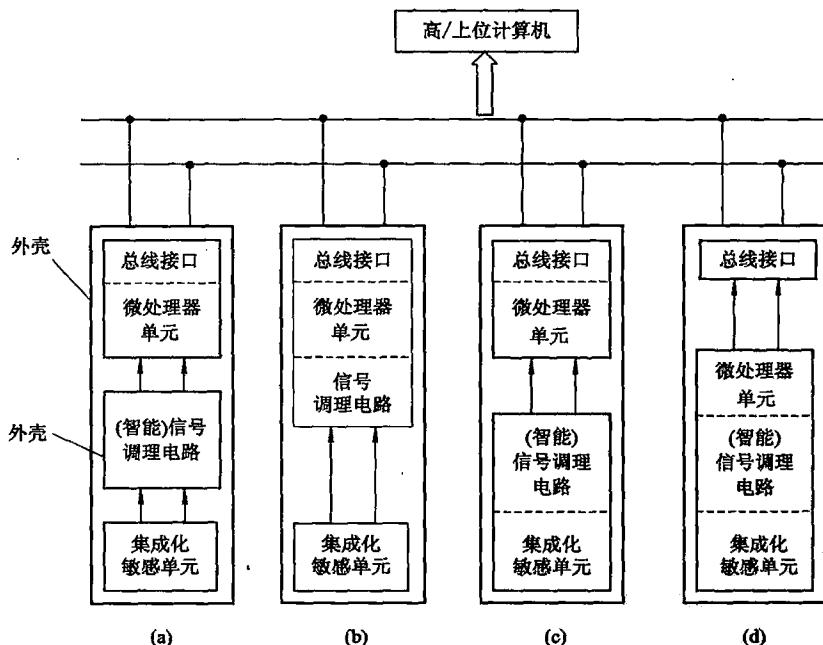


图 1-4 在一个封装中可能的混合集成实现方式

集成化敏感单元包括(对结构型传感器)弹性敏感元件及变换器。

信号调理电路包括多路开关、仪用放大器、基准、模/数转换器(ADC)等。

微处理器单元包括数字存储器(EPROM、ROM、RAM)，I/O 接口，微处理器，数/模转换器(DAC)等。

在图 1-4(a)中，三块集成化芯片封装在一个外壳里。

在图 1-4(b)、(c)、(d)中，两块集成化芯片封装在一个外壳里。

图 1-4(a)、(c)中的(智能)信号调理电路具有部分智能化功能，如自校零、自动进行

温度补偿，这是因为这种电路带有零点校正电路和温度补偿电路，它们常不与微处理单元封装在一起而单独出售。图(a)、(b)中的集成化敏感单元也可以代之以片外外接传感器。

实现智能传感器系统可利用的各种集成化芯片器件及硬件商品将在第3章中介绍。

1.5.4 智能传感器的几种形式

若按具有的智能化程度来分类，智能传感器有三种存在形式。

1. 初级形式

初级形式就是组成环节中没有微处理器单元，只有敏感单元与(智能)信号调理电路，二者被封装在一个外壳里。这是智能传感器系统最早出现的商品化形式，是最广泛使用的形式，被称为“初级智能传感器”。从功能来讲，它只具有比较简单的自动校零、非线性的自校正、温度自动补偿功能。这些简单的智能化功能是由硬件电路来实现的。通常称这种硬件电路为智能调理电路。

2. 中级形式/自立形式

中级形式是指在组成环节中除敏感单元与信号调理电路外，必须含有微处理器单元，即一个完整的传感器系统全部封装在一个外壳里，如现场总线中使用的各种型号的智能传感/变送器。其中的传感器可以是集成化的也可以是经典的，它具有1.3.1节所列的完善的智能化功能，这些智能化功能主要是由强大的软件来实现的。

3. 高级形式

高级形式的智能传感器集成度进一步提高，敏感单元实现多维阵列化，同时配备了更强大的信息处理软件，从而具有更高级的智能化功能。这时的传感器系统不仅具有1.3.1节所述的完善的智能化功能，而且还具有更高级的传感器阵列信息融合功能，或具有成像与图像处理等功能。

显然，对于集成化智能传感器系统而言，集成化程度越高，其智能化程度也就越可能达到更高的水平。

1.5.5 改善传感器系统性能的多传感器智能化技术

当前，多传感器智能化技术迅速发展，已成为改善传感器系统性能最有效的手段。多传感器智能化技术包括两大方面：

其一，将多个传感器与计算机(或微处理器)组建成智能化传感器系统，其深刻内涵是提高某点(单点)位置处某一个参量(单参量) x_1 的测量准确度，而不是一般意义的多点或多参量测量系统。

其二，将多个传感器获得的多个信息的数据进行融合处理，实现某种改善传感器性能的智能化功能，在抑制交叉敏感、改善传感器稳定性的同时，线性度也得到改善。

详细介绍请见第4章的4.1和4.3节以及第5、6、7、8章。

综上所述，我们可以看到，智能传感器系统是一门涉及多种学科的综合技术，是当前正在发展中的高新技术。因此，作为一个设计智能传感器系统的工程师，除必须掌握经典的、现代的传感器技术外，还必须具有信号分析与处理、计算机软件设计、通信与接口、电路与系统等多种学科方面的基础知识。当然，智能传感器系统的发展也需要有多种学科的

工程师并肩合作、共同努力。

1.6 现场总线智能变送器/传感器经典实例简介

1.6.1 现场总线控制系统(FCS)中的传感器与仪表

随着工业生产的发展，需要的测控点和测控参数越来越多，使得一个自动控制系统日益庞大而复杂，原有的分散型控制系统(DCS, Distributed Control System)已不能适应需要。20世纪80年代以来，开放型控制系统开始发展，即现场总线控制系统(FCS, Fieldbus Control System)，这是继DCS之后自动化领域的又一次重大变革。我们现在正处于用新一代的现场总线控制系统取代传统的分散控制系统(DCS)的转变时代。这是工业过程控制仪表发展的必然趋势。

1. 现场总线结构

现场总线控制系统的典型结构如图1-5所示。图中现场总线的节点是现场设备或现场仪表，如传感器/变送器、调节器、调节阀、步进电机、记录仪、条形阅读器等等，但不是传统的单功能的现场仪表，而是具有综合功能的智能仪表。

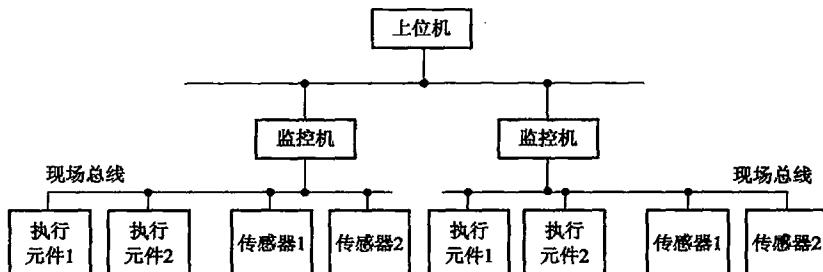


图1-5 现场总线控制系统的典型结构

图中为数众多的智能设备——智能执行元件与智能传感器/变送器位于现场生产、运行区域，处于FCS控制系统的“末梢”节点上。智能传感器/变送器通过总线接口挂接在环形现场总线上，通过现场总线实现现场智能化设备之间的数字通信以及现场设备和上级控制中心(上位机)之间的信息传递。

现场总线网络协议规定了现场应用进程之间的相互可操作性、通信方式、层次化通信服务功能的划分、信息流向及传递规则。这样，可以实现不同厂家生产的种类繁多的产品在硬件、软件、通信规程、连接方式诸方面互相兼容，互换联用。因此，一个智能化设备若想方便地安装、挂接到现场总线上，实现与其它设备及上位机的正常通信，必须遵守统一的网络协议。由于工业现场的固有特点，所以现场总线互联网与商用互联网既有一定的共性，又有很大的不同。这两者的差异主要反映在网络的通信协议中。

至今，现场总线仍然处于群雄争霸的状态，尚无一个最完善的、最权威的、一致公认的标准协议。有关现场总线网络通信协议请参阅有关书籍、手册。

2. 现场总线设备为什么需要“智能”

如果传感器与仪表都是没有智能的，只会采集数据，对采得的数据毫无“感觉”，甚至是否进行数据采集也要由中心控制室下达命令，势必中心控制室的中心计算机就不得不时刻关注每台传感器与仪表（对电站有5000台，对钢铁厂有2万台，对石油化工厂有6000台……）的详细状况，譬如传感器及其仪表的温度补偿情况、自校准情况，工作是否正常，数据是否可靠，等等，再根据所知情况与数据进行分析、判断后作出决策，之后对某个执行器发出控制命令，这就相当于5000、6000或2万人的军团，每个士兵的动作都只由一个总司令来指挥一样。这样，中心计算机负担过重，难以适应现代工业化大生产日益复杂的要求。解决此问题的办法是“分散”或“分布”智能，也就是给现场的传感器/变送器、执行器等设备配备“大脑”——微计算机/微处理器。这样，传统的传感器/变送器与微计算机/微处理器相结合并赋予智能而成为智能传感器/变送器。例如，智能流量变送器由微型计算机中的非线性校正、温度补偿软件等功能模块可以获得排除干扰噪声后的瞬时流量值，经累加运算软件功能模块可获得累计流量值。而且还有控制功能软件智能化模块，可将获得的瞬时流量值与设定值相比较后对执行器——调节阀经现场总线下达控制驱动指令。

决策模式既可以是PID控制模式，也可以是模糊控制模式或其他模式，依控制软件功能模块的类型而定，同时还可以采用调节阀的阀门特性的自校验和自诊断软件功能模块等。这样，我们可以清楚看到：在这种现场总线控制系统的现场设备中，已经组成了对单一量的自行测量、自行数据处理、自行分析判断与决策的控制系统。也就是说，FCS放弃了DCS的控制站，把DCS控制站的功能块分散地分配给现场仪表，从而构成虚拟控制站。这样，许多控制功能从控制室移至现场仪表，大量的过程检测与控制的信息就地采集、就地处理、就地使用，在新的技术基础上实施就地控制。现场智能传感器/变送器将调控了的对象状态参数（如流量）通报给控制室的上位计算机。上位机主要对其进行总体监督、协调、优化控制与管理，实现了彻底的分散控制。譬如，处于监控中心的上位计算机，根据全局优化分析，确定某位置压力值的设定值，该位置处的测控系统就根据设定的压力值进行控制。若上位机下达新的压力设定值，则该位置处的测控系统按新的设定值进行自动调控。在这个局域的分散控制系统中的现场传感器/变送器及执行器都是智能型的，并带有标准数字总线接口，之间的信息反馈与指令均通过现场总线进行。

自动控制系统的飞速发展，对智能型传感器/变送器的需求，成为智能传感器产生、发展的强大社会推动力。目前，用于现场总线控制系统中的、具有智能的智能传感器/变送器也称为现场总线仪表。

3. 现场总线仪表的主要特点

与DCS系统中的现场仪表相比，FCS系统现场仪表的主要特点是：

(1) 具有多种基本的智能化功能，可用于改善静态、动态特性，提高精度和稳定性。如量程设定和零点调整，刻度转换可以多种单位表示被测量，非线性自校正、频率补偿、温度补偿、自检等。

(2) 具有控制与基本参数存储功能，可实现自我管理并提高自适应能力。大量过程检测与控制信息实行现场采集、现场处理、现场使用，现场实时控制，使过程控制基本分散到现场。现场设备都是具有微处理器的，不论是智能传感器/变送器还是执行元件的微处

理器中均可装入 PID(或其他模式)控制模块以实现现场实时控制功能。控制模块通常装入智能传感器/变送器的微处理器中。

在这些现场总线智能设备的微处理器中，不仅存储着各种智能化功能软件模块，还存储了自身的基本信息，如出厂日期批号、技术指标及基本参数，以方便查询与系统维护。

(3) 开放性与互换性。最高理想情况是采用统一的国际标准通信协议，在未能实现最高理想的情况下，可采用同一种通信协议，不同厂家的产品在硬件、软件、通信规程、连接方式等方面可相互兼容、互换使用。这对用户的安装、使用、操作、维修和产品的扩展都十分有利。

(4) 带有数字总线接口，实现通信功能。所有的智能化现场设备，包括变送器、执行器等都通过接口挂接在总线上。现场设备之间，如流量变送器与调节阀之间要执行闭环控制功能需要的信息传递，同时它们与管理中心的上位机之间的通信均是通过各自的接口经现场总线传输进行的。现场总线采用双绞线、光缆或无线方式，目前主要以双绞线为主。也就是说，上位机与所有现场设备的连接只有两根导线，这两根线不仅可以承担现场设备所需的供电，而且承担了它们之间全数字化、双向串行通信。用数字信号取代模拟信号可以提高抗干扰能力，延长信息传输距离，而且大量削减了现场与控制室之间的导线安装费用。而在 DCS 系统中，每一台现场设备就要有一对导线与控制室相连。当然，这种多站的通信必须遵守统一的通信规范和标准。

1.6.2 经典实例简介

现场总线智能传感器/变送器的开发研究多姿多彩。有的生产厂家提供 OEM 集成/固态传感器芯片器件；有的公司专为 OEM 芯片器件配备含有微处理器及通信接口电路系统的印刷电路板，并把印刷电路板与 OEM 芯片器件组装成一个智能传感器；有的按某种现场总线标准对现有用于工业自动化中的传统传感器/变送器加装微处理器与数字总线接口使之成为智能化传感器。追溯本溯源主要有三种类型：基于集成/固态压力(差)传感器的美国霍尼韦尔公司的压阻式传感器，与日本横河电机株式会社的谐振式传感器以及基于传统传感器的美国罗斯蒙特与南美 SMAR 公司的差动电容式传感器，它们在智能传感器/变送器的历史进程以及当今的市场份额中都占据着重要地位。

1. ST - 3000 系列智能变送器

美国霍尼韦尔(Honeywell)公司于 1983 年率先推出了压阻式智能压力变送器 ST - 3000。ST - 3000 型智能变送器是世界上第一台智能变送器，其对传统的现场仪表而言是一次深刻的变革，它开创了现场仪表的新纪元，为工业自动化及其系统应用向更高层次发展奠定了基础。

1) ST - 3000 智能变送器的结构及工作原理

图 1 - 6 为 ST - 3000 智能变送器的结构框图。它由两部分组成：一部分为传感器芯片及调理电路；另一部分为微处理器及存储器。传感器芯片上有由集成工艺制作的三个传感器：压力/压差(ΔP)、静压(P)和温度(T)传感器。其中压力/压差传感器和静压传感器均接成全桥差动电路形式(具体工作原理详见 2.3.1 与 2.4.2 节)。三个传感器的静态标定数据，即表征输入输出特性的标定数据均存入 PROM 中。整个变送器的测压原理为：待测压力首先作用在传感器芯片的硅膜片上，引起传感器的电阻值的相应变化，此阻值的变化由

形成于传感器芯片上的惠斯登电桥检出，桥路输出信号经调理电路进行调理，并由 A/D 转换器转换成数字信号，送入微处理器。与此同时，在此传感器芯片上形成的两个辅助传感器（温度传感器和静压传感器）检测出表体温度和过程静压。辅助传感器的输出也被转换成数字信号并送至微处理器。在微处理器内，来自三个传感器的数字信号进行运算处理，转换成一个对应的 4~20 mA 的模拟输出信号或数字输出信号。

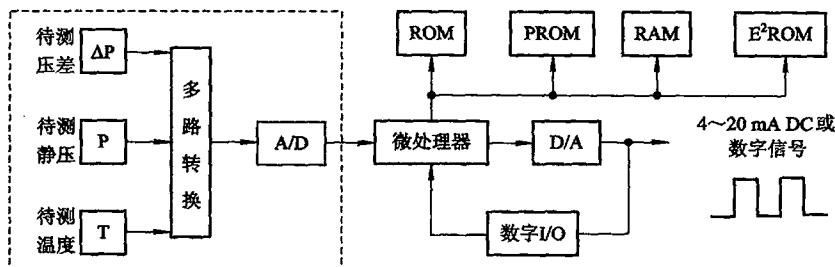


图 1-6 ST-3000 智能变送器的结构框图

2) ST-3000 智能变送器的性能特点

ST-3000 智能变送器的主要性能特点如下：

(1) 宽量程比。ST-3000 的量程比通常可达 100:1，最大可达 400:1(一般传感器仅为 10:1)，这样，当被测压力发生显著变化时，只需调整量程，而不必更换或增加变送器，故一台 ST-3000 变送器可覆盖多台变送器的量程。这是因为 PROM 中的输入输出特性是在宽量程范围内标定的，也反映了传感器本身硅膜片的大量程性能。

(2) 高精度和高稳定性。由于每台 ST-3000 出厂前都在与工作现场相似的环境进行了实验标定，其内存储器 PROM 中存储着一套完整的用于消除交叉灵敏度的特征参数和三个传感器数据的融合计算方程式，从而可以消除环境温度及静压变化对被测压力/压差的影响，保证高精度和高稳定性。ST-3000 模拟输出时的精度达量程的±0.075%，其数字输出(DE)方式时的精度可达量程的±0.062%或读数的±0.125%。消除交叉敏感的数据融合算法详见 5.1.2 节。

(3) 双向通信能力。由于 ST-3000 所具有的双向数字通信能力，使其能方便地用于现场总线测控系统中，符合现代自动化测控系统的要求，且通过与现场通信器(SFC)的远距离通信，可以很容易地实现工作现场与中央控制室之间所进行的参数设定、调整和作业。

(4) 完善的自诊断功能。自诊断功能通过手持现场通信器 SFC 实现，将 SFC 与 ST-3000 连接通信，由 SFC 发出自诊断命令，可对 ST-3000 的通信线路、过程回路和变送器不断进行检测，将检测的结果以简明的语言在 SFC 上显示出来。如在 SFC 上显示“STATUS CHECK=OK”，则表示变送器和 SFC 工作正常；否则，就会把发生的故障显示出来，使用户知道问题所在及如何去修正。这种诊断可远程进行，这样一方面可使操作员不必处于恶劣的工作环境现场，又可大大减少维护时间，既方便了操作又可降低维修成本。

(5) 宽域温度及静压补偿。ST-3000 的温度使用范围可达 -40~110°C，静压可达 0~210 kgf/cm²，且在这么宽的使用范围内可使温度和静压得到补偿。这是因为 ST-3000 具有内部温度与静压传感器，由三个传感器数据进行融合给出最终输出值。

2. 3051 型智能压力变送器

3051 型智能变送器是美国罗斯蒙特公司(现为费希尔-罗斯蒙特公司, FISHER - ROSEMOUNT)在保持 20 世纪 80 年代 1151 型差动电容式变送器优良传统的基础上坚持连续改进获得的新发展,于上世纪 90 年代初推出的现场总线智能仪表系列之一,并于 1998 年开始向我国市场供货。

1) 3051 型智能压力变送器的特点

归纳起来,3051 型智能压力变送器具有如下特点:

(1) 3051 型智能压力变送器测量性能优越,可用于所有压力、液位与流量测量场合,测量精度为±0.075%,其稳定性可做到工作 5 年不需要调整传感器零点漂移。其优异的总体性能为压力变送器树立了一个新的标准,即将温度影响和静压影响考虑在内的综合指标作为精度的指标。

(2) 3051 型智能压力变送器具有多种输出协议:标准输出为 4~20 mA 并基于 HART 协议的数字信号。

(3) 3051 型智能压力变送器的性能和功能不断改进,但具有向前、向后的兼容性。

2) 3051 型智能压力变送器结构

3051 型智能压力变送器由传感和数据处理两部分组成,其原理结构如图 1-7 所示,外形如图 1-8 所示。

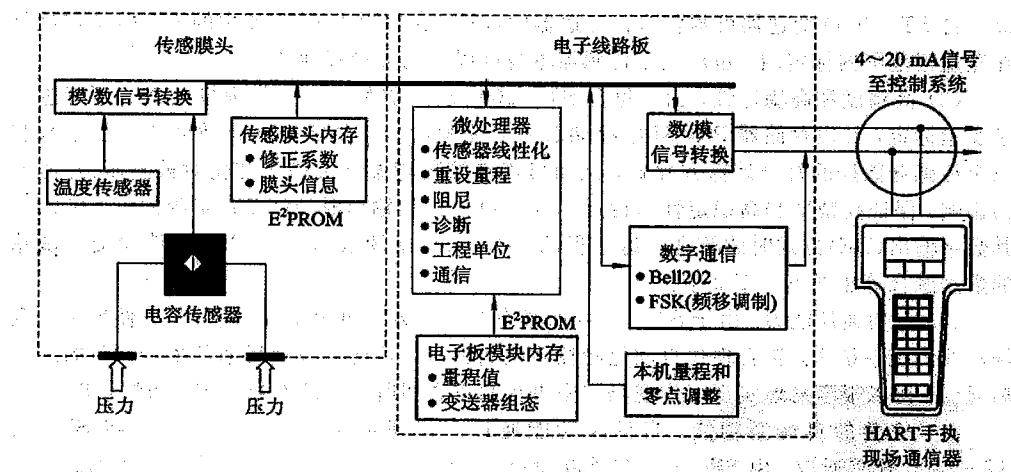


图 1-7 3051 智能压力变送器结构框图

图中各部分的组成和功能阐述如下:

(1) 传感部分。传感部分由测量压力的差动电容式传感器、用于对压力传感器进行温度补偿的温度传感器、模/数转换器以及保存传感器修正系数、传感器膜片参数的片外传感膜头存储器组成。有关差动电容式传感器的工作原理详见 2.3.2 节; 温度补偿原理详见 4.3.3 节。

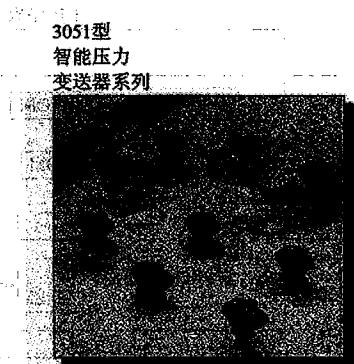


图 1-8 3051 智能变送器外形图

(2) 数据处理部分。数据处理部分由微处理器、片外存储器、数/模转换器、数字通信以及本机量程和零点调整几部分组成。

微处理器要完成如下功能：

- 修正传感器特性；
- 设置量程；
- 设置阻尼系数；
- 对传感器进行故障自诊断；
- 设定工程单位；
- 确定智能传感器与上位机的通信接口与通信格式。

片外存储器用来存放传感器量程以及智能变送器组态参数等。变送器的工作参数包括：

- 零点与量程设定值；
- 线性或平方根输出；
- 阻尼；
- 工程单位选择。

片外存储器将有关信息数据输入变送器，以便对变送器进行识别与物理描述。

数字通信模块遵循 HART 协议，被调制的频移键控信号叠加在 4~20 mA 模拟信号上，通过现场总线实现智能变送器与执行器之间以及智能变送器与上位机之间的通信。

3. 现场总线压力变送器 LD302

LD302 是南美 SMAR 公司于 1995 年推出的第一代现场总线仪表的系列之一。它已在我国安庆石化总厂的腈纶厂投入使用。

1) 硬件构成框图

LD302 是一种测控压差、绝对压力、表压、液位和流量等工业过程参数的现场总线变送器，也称现场仪表。其硬件组成框图如图 1-9 所示。它由传感器组件板、主电路板和显示板组成。



控制在一定范围内,这种方法是以提高成本为代价的。

智能传感器另辟蹊径,通过与微处理器的结合,实现了零位漂移和灵敏度漂移的自动校正。通常采用的自动校准方法有以下几种。

2.2.1 实现自校准功能的方法一

该方法实现的原理框图如图 2.5 所示,校准过程不含传感器。工作过程是:标准发生器首先产生标准电压 U_R 和零点标准值,然后微处理器控制多路转换器分时段选通,以分别选通标准电压 U_R 和零点标准值,并记录下这两种情况下的调理电路输出,从而最终消除零点漂移和灵敏度漂移对传感器性能的影响。

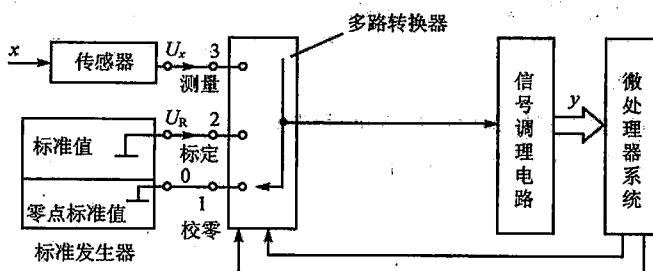


图 2.5 实现自校准功能方法一的原理框图

具体说来,校准过程实质上是三步测量法。

第一步:输入信号是零点标准值,输出值为 $y_0 = a_0$;

第二步:输入信号是标准电压 U_R ,输出值为 y_R ;

第三步:输入信号为传感器输出信号 U_x ,输出值为 y_x 。

通过第一步和第二步可以分别校准调理电路部分的零位和灵敏度,在此基础上可以得到调理电路输出结果为 y_x 时,传感器的真实输出信号 U_x ,之后再反推至传感器的真实输入信号 x ,这样就消除了零点和灵敏度漂移对测量结果的影响。

对于宽量程多档多增益系统,应当对每档增益值都进行自校,但这样需要产生多个标准值,会增加校准成本,因此可以设计“斜率比动态校准法”进行自校,可以用一个标准值对多个增益进行实时标定。原理框图如图 2.6 所示。

图 2.6 中,当标准发生器产生三角波信号 U_R 时,调理电路的输出 y_R 为梯形波,增益值不同则梯形波的斜率不同,因此根据输入与输出信号的斜率比即可确定出调理电路的增益。将阈值电压比较器的上、下限比较电压分别设为 4.5 V 和 -4.5 V,输入信号在 -4.5 V 上升至 4.5 V 时,输出信号为高电平,否则为低电平。微处理器分别记录下 U_R 和 y_R 的上升时间,则被校调理电路的增益 a_1 为 $a_1 = (t_{R2} - t_{R1}) / (t_{y2} - t_{y1})$, t_{R1} 、 t_{R2} 、 t_{y1} 和 t_{y2} 可参见图 2.6。

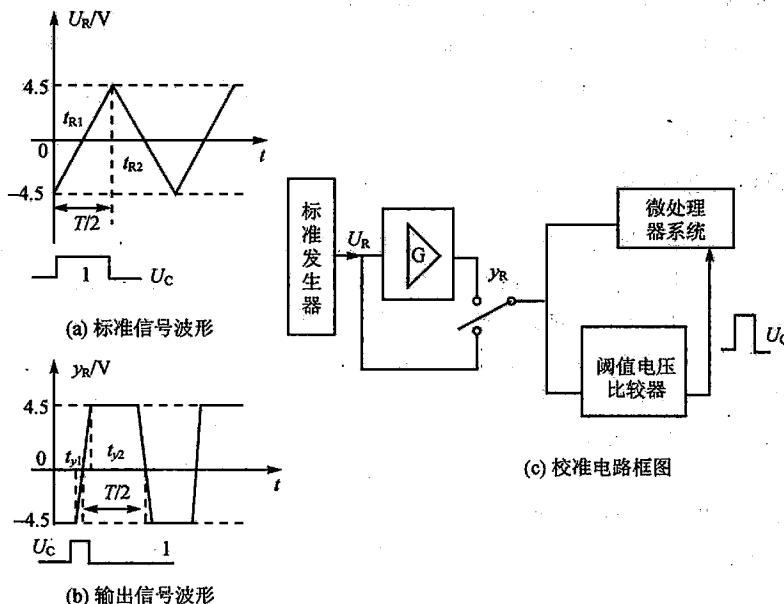


图 2.6 斜率比动态校准法原理及框图

2.2.2 实现自校准功能的方法二

方法二与方法一的区别，在于方法二能够实时自校准包含传感器在内的整个传感器系统。原理如图 2.7 所示。

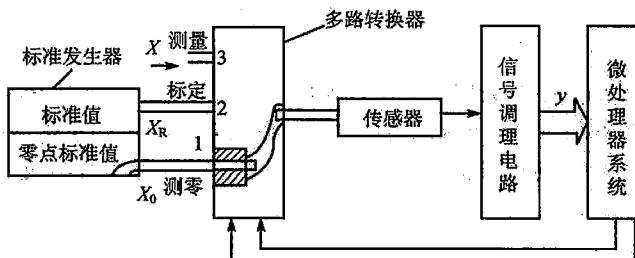


图 2.7 实现自校准功能方法二的原理框图

图 2.7 中，标准值 X_R 、零点标准值 X_0 与传感器输入的被测目标参数 X 的属性相同。假定校正的是压力传感器，则标准值和零点标准值 X_R 、 X_0 须是压力信号，而多路转换器是可传输流体介质的气动多路开关。执行的仍然是三步测量法。整个传感器系统的精度由标准发生器产



生的标准值的精度决定,校准过程中要求被校系统的各环节如传感器、放大器等,在三步测量所需时间内保持稳定。在三步测量所需时间间隔之前和之后产生的零点、灵敏度漂移都不会引起测量误差,故这种实时在线自校准功能,可以采用低精度的传感器、放大器等元器件而获取高精度的测量结果。

2.2.3 实现自校准功能的方法三

前面两种方法均要求被校系统的特性为线性,从而只需要两个标准值即可完善地标定出系统的零点和增益。那么对于非线性的系统特性如何处理呢?原理上讲可通过增加标定点数的方法来实现,但为了增加标定的实时性,点数又不宜增加过多。因此,通常采用施加三个标准值的标定方法(三点标定法),具体标定过程如下:

① 依次输入三个标准值得到对应输出,表示为 (x_{R1}, y_{R1}) 、 (x_{R2}, y_{R2}) 和 (x_{R3}, y_{R3}) 。

② 列出非线性自校准曲线:

$$x = C_0 + C_1 y + C_2 y^2 \quad (2.5)$$

并依据①中得到的测量结果,应用最小二乘法求取式(2.5)中的系数。

③ 求出系数后,智能传感器系统即转入测量状态,根据传感器的输出结果反推至传感器输入端,即可得到校准后的真实传感器输入信号。

这种方法要求传感器系统在标定与测量期间的输入-输出特性保持不变。实验证明,对于一个零漂、温漂在100℃的温度变化范围内,总误差达±1%的压力传感器系统,采用满量程精度为±0.02%的标准压力值进行实时三点校准,整个系统的短时精度可以达到优于±0.1%。但是这种方法要求提供至少三个标准值和一套外围设备,这不是任何一个待测目标参量都能做到的事情。相对来说,目前最为成熟的是智能压力传感器系统。

2.3 自补偿

传感器在实际运行过程中,会因多种误差因素的影响而导致性能下降,因此误差补偿技术的应用势在必行。特别是时域中的温度误差补偿,以及频域中工作频带的扩展,应用非常广泛。下面对这两种误差补偿技术进行简要介绍,其基本思想也可作为其他干扰因素误差补偿的借鉴。

2.3.1 温度补偿

对于非温度传感器而言,温度是传感器系统中最主要的干扰量,在经典传感器中主要采用结构对称方式来消除其影响。而随着微处理器技术的快速发展,在经典传感器与其相结合而产生的智能传感器系统中,则通常采用监测补偿法,即通过对干扰量的监测,再经过相应的软件处理来达到误差补偿的目的。



以压阻式压力传感器为例,因其敏感元件等基本部分由半导体材料制成,故工作特性易受温度影响,因此对其进行温度误差补偿具有典型性和重要的工程应用价值。

1. 温度信号的获取

一般来说,温度的测量需要放置测温元件,但对于压阻式压力传感器而言,可以通过“一桥二测”技术,即通过同一个电桥,实现温度和传感器输出信号的同时测量。图2.8给出了采用恒流源供电的压阻式压力传感器的典型结构图。

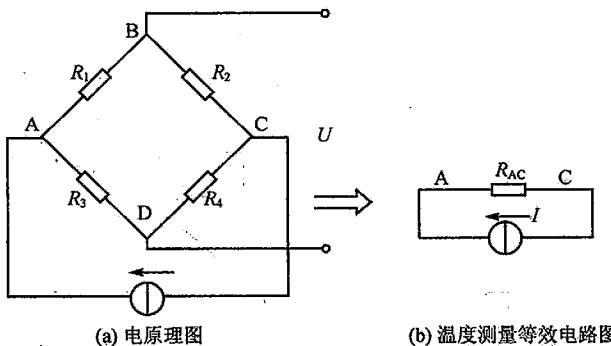


图 2.8 压阻式压力传感器

当被测压力和干扰温度同时作用时,各桥臂的阻值表达式为

$$R_1 = R_3 = R + \Delta R + \Delta R_T \quad (2.6)$$

$$R_2 = R_4 = R - \Delta R + \Delta R_T \quad (2.7)$$

进一步得等效电阻 R_{AC} 为

$$R_{AC} = R + \Delta R_T \quad (2.8)$$

因此

$$U_{AC} = IR + I\Delta R_T \quad (2.9)$$

式中: I 为恒流源电流值; R 为压阻式传感器初始值; ΔR_T 为温度改变所引起的桥臂电阻变化。

通过上述推导可以清楚地看到,A、C 两点的电压是随 ΔR_T 而变化的,是温度的函数,因此监测其数值就得到了与温度变化有固定函数关系的电压信号,之后按照函数关系解算即可得到温度信号。

而监测 B、D 两点的电压,得到的信号当中既包含真实压力变化引起的输出信号变化,也包含温度干扰引起的输出信号变化。下面的关键问题就是把温度引起的干扰信号分离出来,而干扰信号包含零点漂移和灵敏度漂移两个部分。

2. 零点和灵敏度温度漂移的补偿

零点漂移补偿的前提是传感器的特性具有重复性。补偿的基本思想与一般仪器消除零点的思想完全相同。也就是说,假定传感器的工作温度为 T ,则应在传感器输出值 U 中减去该



工作温度下对应的零点电压 $U_0(T_i)$ 。可见,补偿的关键是先测出传感器的零点漂移特性,并保存至内存中。大多数传感器的零点漂移特性呈现出严重的非线性,如图 2.9 所示。

因此,由温度 T_i 求取该温度下的零点电压 $U_0(T_i)$,实际上相当于非线性校正中的线性化处理问题。

对于压阻式压力传感器,在输入压力保持不变的情况下,其输出信号将随温度升高而下降,如图 2.10 所示。图中 $T > T_1$ 。

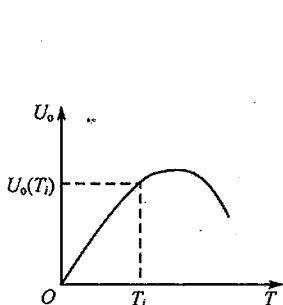


图 2.9 传感器的零点漂移特性

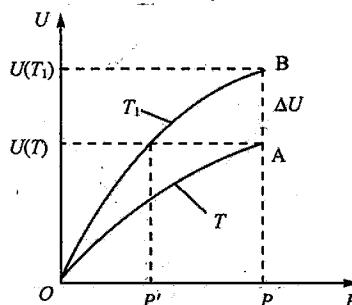


图 2.10 压阻式压力传感器的灵敏度温度漂移特性

因此,若不考虑温度变化对灵敏度的影响,将会产生明显的测量误差。常用的补偿方法有两种:一种是在压阻式压力传感器的温度变化范围内,分成多组测量不同温度下的特性,然后,根据实际工作温度插值获取所应施加的补偿电压;另一种是非线性拟合的方法,拟合传感器输出与温度间的非线性关系,嵌入微处理器中进行补偿。

2.3.2 频率补偿

频率补偿的实质是拓展智能传感器系统的带宽,以改善系统的动态性能。目前主要采用两种方法:数字滤波法和频域校正法。

1. 数字滤波法

数字滤波法的补偿思想是:给当前传感器系统(传递函数为 $W(s)$)附加一个传递函数为 $H(s)$ 的环节,于是新系统的总传递函数($I(s) = H(s) \cdot W(s)$),可以满足动态性能要求。补偿过程的示意图如图 2.11 所示。

以一阶系统为例,其传递函数和频率特性分别为

$$W(s) = \frac{1}{1 + \tau s} \quad (2.10)$$

$$W(j\omega) = \frac{1}{1 + j\omega\tau} \quad (2.11)$$

若想将其频带扩展 K 倍,即转折角频率为

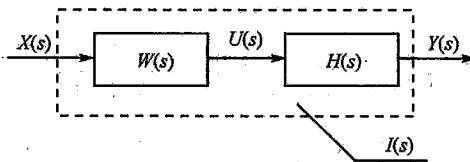


图 2.11 数字滤波法补偿示意图

$$w'_\tau = Kw_\tau \quad (2.12)$$

这等价于将其时间常数减小 $1/K$ 。具体的实现方法读者可自行查阅信号分析处理方面的书籍，这里就不再赘述。图 2.12 给出了一个测温传感器系统性能改进的实例。

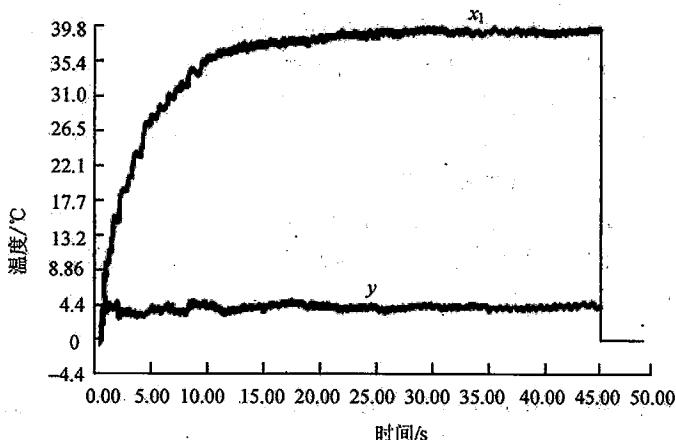


图 2.12 测温传感器系统阶跃响应改善实例图

从图 2.12 中可以看出，经数字滤波后，测温系统的阶跃响应 y 的上升沿明显改善，且由于数字滤波是编程实现的，故调整灵活方便。

2. 频域校正法

图 2.13 给出了系统动态特性频域校正的过程。与数字滤波一样，频域校正的前提是已知系统的传递函数。

频域校正分为采样($y(t) \rightarrow y(n)$)、快速傅里叶变换(FFT)、复数除法运算($Y(m)/W(m)$)及快速傅里叶反变换(IFFT)四步。其核心思想是：通过频域校正对畸变的 $y(t)$ 进行处理，得到能够真实反映输入信号 $x(t)$ 的频谱 $Z(m)$ ，然后进行傅里叶反变换以求取输入信号的真值，从而达到消除误差的目的。

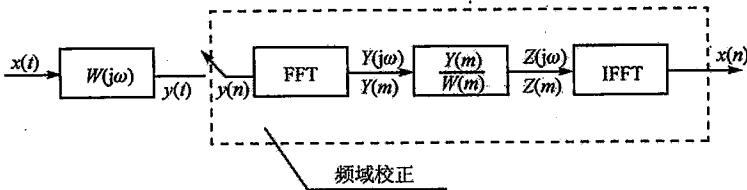


图 2.13 系统动态特性频域校正过程示意图

2.4 增益的自适应控制

关于智能传感器系统的增益设置,要在系统自身数据容量与被测量范围、系统的精度与信噪比、系统的灵敏度与分辨率等诸多因素之间折中选择来确定。若增益过小,数据字的信息容量就会浪费,信噪比可能很低,测量误差大而不能满足要求;反之,若增益过大,信息也会因系统内的数据字信息容量不够而损失掉。所以增益设置必须仔细权衡,根据具体情况折中确定。下面通过一个例子说明增益选择的基本规则。

例 2.1 考虑一个增益可程控放大器跟随一个 8 位 A/D 转换器组成的数据采集系统,若要求由 A/D 转换器量化噪声产生的相对误差不大于 0.5 %,试确定量程切换的原则。

解:已知 A/D 转换器的量化值 q 为

$$q = \frac{U_H}{2^b}$$

式中: U_H 为 A/D 转换器满刻度输出时对应的电压值; b 为 A/D 转换器的位数。

尽管输入是连续变化的模拟量,但 A/D 转换器只能输出离散值,根据 A/D 转换的基本工作原理,A/D 转换过程中必然产生量化误差,通常情况下可以设其为随机变量——噪声。在最坏情况下,最大量化误差一般为

$$\epsilon_m = \frac{1}{2}q$$

根据题意要求,量化噪声产生的相对误差不大于 0.5 %,即

$$\delta = \frac{\epsilon_m}{U_i} \leqslant 0.5 \%$$

则输入电压

$$U_i \geqslant \frac{\epsilon_m}{\delta} = \frac{0.5q}{0.5\%} = 100q$$

可见,输入电压的最小值不得小于 100 个量化值;小于这个阈值,则应通过微控制器发出指令,调整程控放大器增益。而若输入电压的最大值大于 255 个量化值,也会出现问题,因为大于 255 之后的电压,只能恒定输出 255。因此,也应适当设定输入电压的上限值,超过这个

上限时,程控放大器的增益应适当减小。

因此,可得量程切换的原则如下:

- $U_i < 100q$, 增益增大;
- $U_i > 250q$, 增益减小。

而增益的控制则通过微控制器发出相应指令执行,这样就完成了一个增益自适应控制的全过程。

由于增益自适应控制的情况千变万化,没有一个统一的原则,所以应当根据实际情况进行分析处理。但是通过上述例子可以清楚地看出,增益自适应控制的出发点是固定增益电路出现了难以避免的不足,而增益自适应控制的优点也正是弥补了这个不足。最后应当强调的是增益自适应控制是伴随着微控制器、大规模集成电路技术的发展应运而生的,这是实现增益自适应控制的硬件平台。

2.5 传感器系统的自检

自检是智能传感器自动开始或人为触发开始执行的自我检验过程。其作用是对智能传感器系统的软硬件功能进行检测,给出检测结果以判断传感器的性能,有助于提高智能传感器系统的可靠性。

自检通常有三种实现方式。

1. 开机自检

每当工作电源接通后进行,在随后的测试工作中不再进行。这种自检一般用于检查显示装置、ROM、RAM 和总线,有时也用于对插件进行检查。

2. 周期性自检

周期性自检是在传感器工作过程中,定期检查关键部位的工作状态,以确保传感器始终处于最优工作状态。周期性自检完全是自动的,在测控间歇期进行,不干扰传感器的正常工作。除非检测到故障,周期性自检并不为操作者所察觉。

3. 键控自检

键控自检是需要人工干预的检测手段,用于那些不能在正常运行操作中进行的自检项目,可通过操作面板上的“自检按键”,由操作人员干预进行。这种自检方式简单方便,可以灵活进行,有助于补充周期性自检所不能完成的工作。

智能传感器内部的微处理器,具有强大的逻辑判断能力和运行功能,通过灵活的编程即可方便地完成各项自检项目。

2.5.1 ROM 自检

ROM 是存放智能传感器工作软件的存储器,因此存储内容的可靠性至关重要。对 ROM

的自检通常采用“校验和法”，即在 ROM 中保留一个单元，在该单元中写入校验字，使 ROM 中的每一列具有奇数(或偶数)个“1”，从而使 ROM 的校验和为全“1”(或全“0”)，这称做 ROM 的奇(或偶)校验。如表 2.1 所列，前面 7 个单元写入的是程序代码，最后一个单元写入奇校验字 11001011，使 8×8 ROM 的校验和全为“1”。

表 2.1 ROM 自检原理

ROM 地址	ROM 中内容	备注
0	11111110	
1	11000110	
2	10101010	
3	01111110	
4	01010101	
5	10001101	
6	00000000	
7	11001011	校验字
	11111111	校验和

2.5.2 特殊功能寄存器自检

智能传感器中有必要检测关键性的标志寄存器和特殊功能寄存器的有效性。测试标志寄存器时，可以将相应标志位置“1”，然后执行一条条件转移指令，如果根据当前的标志位能够正确跳转，说明标志位正常；否则需要开展进一步检查。对于可读/写寄存器的检查，可通过先写入、再读出的方法来判断寄存器的正确性。通常情况下，为确保测试正确性，可以先后进行两次写入和读出过程，第一次全部写入“1”，第二次全部写入“0”，如果测试正确，则可转入下面的测试程序，否则给出错误信息，停机检查。

2.5.3 RAM 自检

静态 RAM 的自检分两种情况：当 RAM 为空白时，自检的目的是判断其能否正确写入和读出。通常采用的方法是写入“AA”，然后逐字节读出，检查是否存在问题；而后再写入“55”以执行同样的过程，若没有问题则自检通过。若 RAM 中已写入数据，则采用“异或”的方法，先读出 RAM 单元中的数据，存入寄存器，将其求反后与原单元内容作“异或”运算，若结果全“1”，则表明单元工作正常。

对于设计有动态 RAM 的智能传感器，应注意自检时必须对动态 RAM 进行定时刷新，因此首先要对定时器部件进行检查。



2.5.4 总线自检

微控制器的总线通过缓冲器与 I/O 接口器件和各种插件相连。总线自检的目的就是检查经过缓冲器的总线(外部总线)传递信息是否正确。由于总线没有记忆能力,因此必须设置相应的寄存器,以分别保持地址总线和数据总线上的信息,如图 2.14 所示。

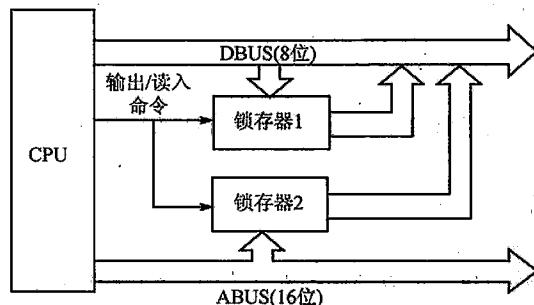


图 2.14 总线自检原理图

总线自检时,CPU 首先对相应锁存器执行一条输出指令,使地址总线和数据总线上的信息保存在锁存器中,然后再对锁存器进行读入操作,即将锁存器中的地址和数据总线信息重新读入 CPU,并和原来的输出信息进行比较,即可判断出外部总线是否存在故障。需要注意的是,对外部总线进行自检时,应分别检查每一条信号线。如对 16 位地址总线来说,需采用 32 步独立测试来完成,前 16 步逐次置地址信号线为“1”,后 16 步逐次置“0”;如果测试中某条地址线总是保持“1”(或“0”),或地址线相互短路,则利用该方法可方便地检出。

2.5.5 A/D 和 D/A 自检

A/D 和 D/A 转换器分别在智能传感器输入、输出通道中占有重要地位,其转换精度直接影响系统控制运算精度,因此必须对其进行周期性自检。

1. A/D 转换器自检

通常采取的方法是,将智能传感器内的参考/基准电源接在 A/D 转换器的输入端,然后启动采集程序,并将采样结果与机内存入的预定值进行比较;若误差在允许范围内,则 A/D 转换器工作正常。

2. D/A 转换器自检

D/A 转换器的自检通常需要 A/D 转换器的配合。自检的方法是由微控制器产生扫描电压信号(如锯齿波),将该信号接入 D/A 转换器输出端,再经 A/D 转换器采回,判断采样结果与机内预定值之间的误差是否在允许范围内。可见,D/A 转换器自检必须在 A/D 转换器自检后进行,以消除 A/D 转换器可能带来的误差影响。



2.5.6 I/O 接口电路自检

不同智能传感器的 I/O 接口电路不同。因此，测试方法差异性较大，需要具体问题具体分析，不能千篇一律。例如，若智能传感器系统带有键盘和显示器，通常的检测方法是读取相应的标志寄存器来判断其状态正确与否。而有些 I/O 接口电路专门设计了自检电路，这个时候只需正常启动其自检电路即可。

2.5.7 插件自检

功能较强的智能传感器常常有多个插件，因此对插件插接状况的检查非常重要。插件自检包括两个内容：

- ① 检查插件是否已经插入；
- ② 检查插件是否正常工作。

前者的检查方法比较简单。可以利用插件与 CPU 之间的应答信号出现与否来判断。而后的实现比较麻烦，常采用模拟方法进行，且要引入故障诊断学的有关内容才能解决，因此这里不再详述。

2.5.8 显示面板自检

显示面板的自检一般有两种方式。一种方式是令面板上显示器的所有字段发光，然后再令所有字段不发光，以此来检查显示器。检查结果正常后，按下仪器任意按键，即可进入正常工作状态。另一种自检方式是显示某些特征字符，几秒钟后自动消失，进入其他初始状态或正常工作状态。

综上所述，智能传感器自检项目繁多，就不再一一列举了。自检过程中，当检查到系统出现某种故障时，一般借助系统本身的显示装置给出显示代码，并发出声光报警信号，以便于操作人员根据显示信息查找故障类型，判断故障发生部位。

智能传感器设计中，自检是必需的，自检项目越多，使用和维修越方便，但是相应的用于自检的软、硬件结构也越复杂。因此设计人员应根据仪器的工作特性和用户要求，确定自检项目，并尽可能利用被检对象本身提供的信号和现有资源实现自检，使系统自检工作能够简单方便地进行。

2.6 自诊断

众所周知，科学技术的进步使得许多航天器应运而生，诸如火箭、导弹等飞行器在控制系统中都应用了大量传感器，以监测系统运行过程中的参数。如果这些传感器在使用过程中发生故障，包括硬故障（传感器损坏）和软故障（传感器性能变差），都可能导致整个系统运行瘫



痪。因此,当某个传感器发生故障后,希望能够及时进行检测并且进行故障隔离。这项工作得到了越来越广泛的重视。

目前为止,广泛应用的传感器故障诊断方法主要有三大类:硬件冗余方法、解析冗余方法和人工神经网络方法。每种方法都有其优势与不足,下面简介硬件冗余方法和解析冗余方法的优缺点。

2.6.1 硬件冗余方法

硬件冗余方法是最早采用的诊断方法,其核心思想就是对容易失效的传感器设置一定的备份,然后通过表决器的方法进行管理。也就是说,硬件是由两个、三个、或者四个完全一样的,且是测量相同被测量的设备构成。对冗余设备输出量进行相互比较,可以验证整个系统输出的一致性。

一般而言,双重冗余配置只能判断有无传感器故障,不能分离故障;三重冗余系统则可以判断故障的有无,以及进行故障的分离。硬件冗余方法的优点是不需要被控对象的数学模型,且鲁棒性很强;缺点是设备复杂,体积和质量很大,成本较高。对于一些重要系统,如火箭、航天飞机等,利用硬件冗余备份方法是现实可行的;而对于大多数情况,特别是需要测量很多量的情况下,硬件冗余方法的不足就凸显出来了。

2.6.2 解析冗余方法

解析冗余方法的实质就是建立被测对象(含传感器)的动态模型,通过比较模型输出和实际输出之间的差异来判断传感器是否发生故障。其原理如图 2.15 所示。

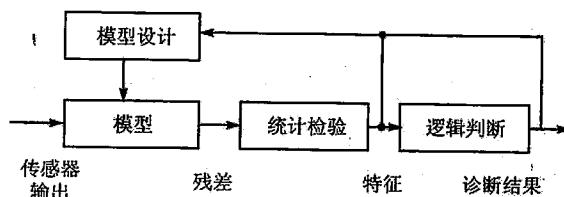


图 2.15 解析冗余方法原理框图

从图 2.15 中可以看出,解析冗余方法的大致步骤如下:

① 模型设计。根据被控对象的特性、传感器的类型、故障类型和系统要求等,建立相应的被控对象的数学模型。

② 设计与传感器故障相关的残差。在相同控制量的作用下,传感器输出信号和由模型所得值之差,即为残差。在没有传感器故障时,残差应为零。当有传感器故障时,残差不再为零,其中包含了传感器故障信号。



③ 进行统计检验和逻辑分析,以诊断某些类型的传感器故障。

根据数学模型产生方法的不同,有不同类型的解析冗余方法。目前主要有:观测器组方法、故障检测滤波器方法、一致性空间方法、状态和参数辨识方法及基于知识的方法等。总的来说,用解析冗余方法进行传感器故障诊断,能够定位故障来源,也就是说能够确定哪个传感器发生了故障,并可估计故障大小和严重程度。同时,解析冗余方法还不需要增加硬件设备,因此成本较低。

但是,这种方法也存在不足。当系统参数存在不确定性,以及系统参数随时间变化而变化时,或者系统中有未知的输入干扰,都会给诊断结果带来不利影响。因此必须要求方法具有鲁棒性,也就是说,对应的传感器故障诊断和检验算法必须对系统参数时变、未知输入干扰等干扰因素具有抑制能力。另外,这种方法也必须知道被控对象的精确数学模型,因此当系统存在高度非线性,而且难以得到系统的数学模型时,这种方法就无能为力了。最后要说明的就是,这种方法能够进行传感器的故障诊断,但不能恢复故障传感器的信号。

对于解析冗余方法的具体工作过程,感兴趣的读者可以自行查阅容错控制方面的书籍加以了解。

2.7 噪声抑制与弱信号检测

传感器获取的信号中常常夹杂噪声及各种干扰信号。作为智能传感器系统,不仅应具有获取信息的功能,还应具有信息处理的功能,因此噪声抑制和微弱信号检测是智能传感器设计过程中必须正视的问题。

2.7.1 噪声抑制技术

广义上讲,噪声是扣除被测信号真实值之后的各个测量值。通常情况下,常把可以减小或消除的外部扰动称为干扰,也就是说,干扰是非被测信号或非测量系统引起的,是外界影响造成;而把由于材料或器件的物理原因引起的扰动称为噪声。

理论上讲,干扰是属于理想上可排除的噪声。不少干扰源发出的干扰是有规律的,有些是周期性的,有些只是瞬时的,因此可以通过接地、屏蔽、滤波等手段加以消除或削弱。而来自于被测对象、传感器乃至整个测控系统内部的噪声,往往由大量的短尖脉冲组成,其幅度和相位都是随机的,这类噪声不可能完全消除,只能设法减小。常用的噪声抑制技术包括滤波和相关等,是依据噪声和有用信号频谱的分布特点来选取的。当噪声频谱和信号频谱不重合时,可以考虑利用滤波技术滤除噪声;而当信号和噪声频谱有重叠时,则要考虑利用相关技术等方法来消除噪声。

1. 滤 波

滤波可以分为模拟滤波和数字滤波两种。模拟滤波器需要在原始信号后面增加新的元器



件；若是有源滤波器，还要增加新的电源，因此就带来了新的干扰源。而数字滤波器只需对信号进行采集，然后送入微控制器/微计算机中进行处理，不需要增加额外的硬件设备，且数字滤波还可以实现很低频率信号的滤波。所以，数字滤波技术近年来得到了广泛的重视和应用。不过也应看到，模拟滤波器实时性好的特点也是数字滤波器不具备的。两者相比，性能各有千秋，应根据设计要求灵活选择。

对于数字滤波器，从实现方法上来讲，有 IIR (Infinite Impulse Response) 滤波器和 FIR (Finite Impulse Response) 滤波器之分。FIR 滤波器可以对给定的频率特性直接进行设计，而 IIR 滤波器常用的方法是利用成熟模拟滤波器的设计方法来进行设计。由于 FIR 系统只有零点，所以要获取好的衰减特性，阶数必须取得高。但它也有自己的特点，一是系统总是稳定的，二是易实现线性相位，三是允许设计多通带滤波器。后两项都是 IIR 滤波器难以实现的，而 IIR 滤波器的优势则在于较低的阶数即可获取较好的衰减性能。此外，同步加算平均滤波、移动平均滤波、中位值滤波等滤波方法也在噪声抑制中得到了有效应用。下面以同步加算平均滤波为例来说明滤波对噪声抑制的作用。

同步加算平均滤波的工作过程如下：

① 首先将记录信号 $x(t)$ 的长度分为 m 段，每段长度相同，得到 m 个样本函数，而 $x(t)$ 是有用周期信号 $s(t)$ 与噪声信号 $N(t)$ 之和，即

$$x(t) = s(t) + N(t) \quad (2.13)$$

② 假定 $N(t)$ 是随机噪声，则均值为零，故噪声的强度用均方值衡量，每个样本的均方值为 σ^2 ， m 个样本的总体均方值之和为 $m\sigma^2$ ，均方根值为 $\sqrt{m}\sigma$ 。而对于周期信号 $s(t)$ ， m 个样本的总体均值为 $ms(t)$ ，所以样本平均后的信噪比为

$$\left(\frac{s}{N}\right)_m = \frac{ms(t)}{\sqrt{m}\sigma} = \sqrt{m} \left(\frac{s}{N}\right) \quad (2.14)$$

式中： $\left(\frac{s}{N}\right)_m$ 是样本平均后的信噪比， $\left(\frac{s}{N}\right)$ 是样本平均前的信噪比。

③ 可见，采用同步相加平均后信噪比提高了 \sqrt{m} 倍。图 2.16 给出了整个过程示意图，(a) 为信号分段示意图，(b) 为同步加算平均滤波后的效果。

2. 相 关

相关分析是随机信号处理过程中的常用技术。按照不同的应用情况又分为自相关函数和互相关函数两种。

自相关函数是描述同一随机过程相隔 τ 的两个时间点上取值的相关程度；而互相关函数则是描述两个不同的随机过程在相隔 τ 的两个时间点上取值的相关程度。下面就举例说明相关分析在噪声抑制中的应用。

例 2.2 利用相关估计去除噪声中的工频干扰。

设输入信号为 $s(t)$ ，其含有微弱待测有用信号 $x(t)$ 和工频干扰信号 $x_i(t)$ 。工频干扰信号

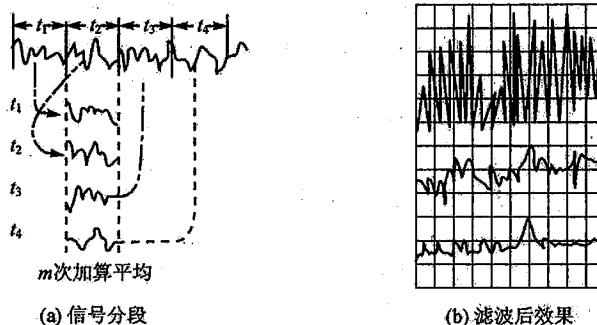


图 2.16 同步加算平均滤波示意图

$x_i(t)$ 可表示为

$$x_i(t) = C \sin(\omega_i t + \varphi)$$

将其展开可得

$$x_i(t) = C \sin \varphi \cos \omega_i t + C \cos \varphi \sin \omega_i t = A \cos \omega_i t + B \sin \omega_i t$$

可见,如果求得 A 、 B 及 ω_i ,则工频干扰信号就完全确定下来,将测得信号 $s(t)$ 减去工频干扰 $x_i(t)$,即可得到有用信号 $x(t)$ 。那么如何求出 A 、 B 及 ω_i 呢? 步骤如下:

① 对信号 $s(t)$ 进行采样。假定采样周期为 T_0 ,采样时间为 $t_p = NT_0$,采样时间为工频周期的整数倍,则可得采样序列为

$$s(nT_0) = x(nT_0) + \hat{A} \cos \hat{\omega}_i n T_0 + \hat{B} \sin \hat{\omega}_i n T_0 \quad (n = 0, 1, \dots, N-1)$$

为简单起见,令 $T_0=1$,则上式变为

$$s(n) = x(n) + \hat{A} \cos \hat{\omega}_i n + \hat{B} \sin \hat{\omega}_i n \quad (n = 0, 1, \dots, N-1)$$

式中: \hat{A} 、 \hat{B} 、 $\hat{\omega}_i$ 分别是 A 、 B 、 ω_i 的估计值。

② 将采样序列与 $\cos \hat{\omega}_i n$ 进行互相关,并求延迟 $\tau=0$ 时的互相关函数值 $\hat{R}_A(0)$ 。

根据互相关函数定义可得

$$\hat{R}_A(\tau) = \frac{1}{t_p} \int_0^{t_p} s(t) \cos \omega_i (t + \tau) dt$$

因 t_p 为有限值,上式只是 $R_A(\tau)$ 的估计值,且由于 $s(t)$ 和 $\cos \hat{\omega}_i t$ 经过采样了,所以上式积分运算可以变为求和运算。在 $\tau=0$ 时:

$$\hat{R}_A(0) = \frac{1}{N} \sum_{n=0}^{N-1} [x(n) + \hat{A} \cos \hat{\omega}_i n + \hat{B} \sin \hat{\omega}_i n] \cdot \cos \hat{\omega}_i n$$

因 $x(n)$ 与 $\cos \hat{\omega}_i n$ 独立无关,以及 $\sin \hat{\omega}_i n$ 与 $\cos \hat{\omega}_i n$ 三角函数的正交性,可进一步得

$$\hat{R}_A(0) = \frac{1}{N} \sum_{n=0}^{N-1} \hat{A} \cos^2 \hat{\omega}_i n = \frac{\hat{A}}{2N} \sum_{n=0}^{N-1} (1 + 2 \cos \hat{\omega}_i n) = \frac{\hat{A}}{2}$$



于是

$$\hat{A} = 2\hat{R}_A(0) = \frac{2}{N} \sum_{n=0}^{N-1} s(n) \cos \hat{\omega}_i n$$

③ 将采样序列与 $\sin \hat{\omega}_i n$ 进行互相关，并求延迟 $\tau=0$ 时的互相关函数值 $\hat{R}_B(0)$ 。

完全参照和②相同的步骤，可得

$$\hat{B} = 2\hat{R}_B(0) = \frac{2}{N} \sum_{n=0}^{N-1} s(n) \sin \hat{\omega}_i n$$

④ 计算工频干扰实时的角频率 $\hat{\omega}_i$ 。采用搜索试探法进行，按照②和③中的计算方法计算 \hat{A} 和 \hat{B} 时，应首先设定 $\hat{\omega}_i$ 的初值，并以一定步长改变 $\hat{\omega}_i$ 的设定值，以求得不同 $\hat{\omega}_i$ 下的 \hat{A} 和 \hat{B} ，再计算两者的模：

$$\hat{C} = \sqrt{\hat{A}^2 + \hat{B}^2}$$

\hat{C} 是 $\hat{\omega}_i$ 的单峰函数，故其关系曲线的峰值处对应的 $\hat{\omega}_i$ 即为工频干扰的实时角频率。

⑤ 有用信号的分离。至此，工频干扰信号的 \hat{A} 、 \hat{B} 、 $\hat{\omega}_i$ 全部求出，工频干扰信号可由下式求出，即

$$x_i(n) = \hat{A} \cos \hat{\omega}_i n + \hat{B} \sin \hat{\omega}_i n$$

待测有用信号为

$$x(n) = s(n) - \hat{A} \cos \hat{\omega}_i n - \hat{B} \sin \hat{\omega}_i n$$

可见，灵活运用相关分析技术，即从采样信号中分离出了工频干扰信号，达到了噪声抑制的目的。

2.7.2 弱信号检测技术

弱信号检测的前提是要区分弱信号和噪声的不同。前面已经提到，噪声没有重复性的特点，是随机出现的，不同时刻的噪声之间是不相关的；而弱信号则是有用信号，信号间是有关联的。

弱信号检测的方法多种多样，下面举几种常用的检测方法加以说明。

1. 窄带滤波法

这个方法是利用信号的功率谱密度较窄，而噪声的功率谱密度相对很宽的特点，使用一个窄带通滤波器，将有用信号的功率提取出来。由于窄带通滤波器只让噪声功率的很小一部分通过，而滤掉了大部分的噪声功率，所以输出信噪比能得到很大的提高。

例如，对 $1/f$ 噪声来说，当其通过一个传输系数为 K_v 、带宽 $B = f_2 - f_1$ 的系统后，输出噪声为

$$E_{\text{now}}^2 = \int_{f_1}^{f_2} K_v^2 K \frac{1}{f} df = K_v^2 K \ln \left(1 + \frac{B}{f_1} \right) \quad (2.15)$$

式中: E_{now} 为输出噪声电压。

可见, 带宽越窄, 输出噪声电压的均方值就越小, 对噪声的抑制能力就越强。

也可以通过有限正弦信号及白噪声的功率谱密度曲线来说明窄带滤波法是如何提高信噪比的, 如图 2.17 所示。

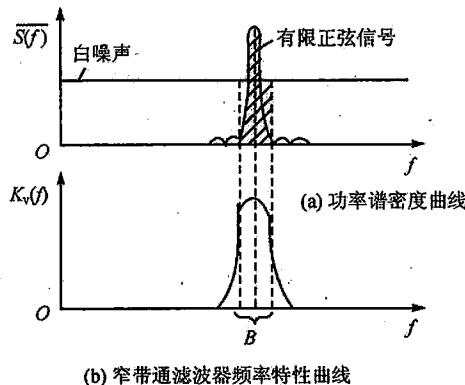


图 2.17 窄带滤波法图例

由图 2.17 看出, 使用了窄带通滤波器后, 则

$$\text{输出信噪比} = \frac{\text{信号主峰下的面积}}{\text{画斜线的矩形面积}} > 1$$

如果 B 选得很窄, 则输出信噪比还能更大一些。

窄带通滤波器的实现方式很多, 常见的有双 T 选频、LC 调谐、晶体窄带滤波器、金属振子机械滤波器等。其中双 T 选频和金属振子机械滤波器可以做到相对带宽 (B/f_0) 千分之几 (f_0 为中心频率), 晶体窄带滤波器可以做到万分之几, LC 调谐的相对带宽则比它们都要大。这些滤波器的带宽相比其他方法, 带宽还是太宽, 因此这种方法不能检测深埋在噪声中的信号, 通常它只用在对噪声特性要求不很高的场合中。

最后需要说明, 窄带滤波法不仅适用于周期性正弦信号波形的复现, 而且也能用来检测单次信号是否存在。需要说明的是, 对于单次信号而言, 由于通过滤波器后, 信号的一小部分频率分量被滤掉了, 所以这种方法不能复现单次信号的波形, 而只能用来提取信号存在与否的信息。

2. 锁定接收法

锁定接收法的原理电路如图 2.18 所示。其中 $U_1(t)$ 为输入信号, $U_2(t)$ 为参考信号, 这两个信号同时输给乘法器进行运算, 然后再经过积分器, 最后得到输出信号 $U_o(t)$ 。

若 $U_1(t)$ 为正弦信号电压, $U_{s1}(t) = U_{s1} \sin(\omega_1 t + \phi_1)$, 而 $U_2(t) = U_2 \sin(\omega_2 t + \phi_2)$, 亦为正

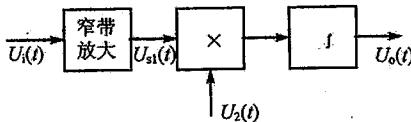


图 2.18 锁定接收原理电路图

弦电压，并且 $\omega_1 = \omega_2$ ，则锁定放大器的输出为

$$U_{so}(t) = \frac{K_v}{2} U_{sl} U_2 \cos(\phi_1 - \phi_2) \quad (2.16)$$

式中： K_v 为系统的传输增益。

由于输入信号的相角 ϕ_1 及参考信号的相角 ϕ_2 均为常数，因而式(2.16)为恒定值，也就是说锁定放大器得到了直流输出信号。

如果 $U_{nl}(t)$ 为窄带白噪声（或称窄带高斯噪声）， $U_{nl}(t) = \rho(t) \sin[\omega_1 t + \phi(t)]$ ，其中的幅度 $\rho(t)$ 和相角 $\phi(t)$ 均随时间作不规则变化， $U_2(t) = U_2 \sin(\omega_2 t + \phi_2)$ ，并且 $\omega_1 = \omega_2$ ，则锁定放大器输出为

$$U_{no}(t) = \frac{1}{2T} U_2 \int_0^T \rho(t) \{ \cos[\phi(t) - \phi_2] - \cos[2\omega_1 t + (\phi(t) + \phi_2)] \} dt \quad (2.17)$$

当积分时间 $T \rightarrow \infty$ 时， $U_{no}(t) = 0$ ，这表明当积分时间很大时，锁定放大器对噪声的抑制能力很强。在实际中，由于 T 不可能做得很大，或者有时为了制作方便，积分器是用低通滤波器来代替时，这时锁定放大器输出的噪声不为零，而是在零附近起伏变化。

衡量锁定放大器噪声特性的指标也是用信噪比改善来表征的，即

$$SNIR = \Delta f_{ni} / \Delta f_c \quad (2.18)$$

式中： Δf_{ni} 为输入白噪声的带宽， Δf_c 为锁定放大器的等效噪声带宽，主要由放大器中的积分器来决定。

锁定放大器是 20 世纪 60 年代才出现的精密测量仪器，几十年来通过不断的研制和改进，已逐渐成为多用途分析仪器，广泛用于光学、声学、振动、磁学、热学、超导、微波、半导体等到各个方面。目前，国外正着力于进一步扩大锁定放大器的功能；同时从电路结构上研制新的锁定放大器。如为了解决方波调制时产生的奇次谐波干扰，设计了脉冲载波调制锁定放大器；为了进一步增强抑制噪声的能力，设计了外差式锁定放大器等。

3. 取样积分法

取样积分法是利用周期性信号的重复特性，在每个周期内对信号的一部分取样一次，然后经过积分器取出平均值，因而各个周期内取样平均信号的总体便展现了待测信号的真实波形。与此同时，由于取样点的重复及积分器的抑噪作用，大大提高了信噪比。

图 2.19 为取样积分器的原理电路图，其中 $U_1(t)$ 为输入周期性重复信号， $U_2(t)$ 为与 $U_1(t)$ 周期同步的参考信号。受 $U_2(t)$ 控制的时基信号驱动取样脉冲的产生，使得在取样脉冲期间取样门



被接通,于是输入信号通过取样门进入积分器进行运算,最后得到输出信号 $U_o(t)$ 。

为了具体说明取样积分的原理,绘出图 2.20 所示波形。图 2.20 中最上面的图给出的是周期为 T 的输入信号 $U_1(t)$ (只画出了一个波形),下面画的依次是各个周期内取样脉冲的波形。

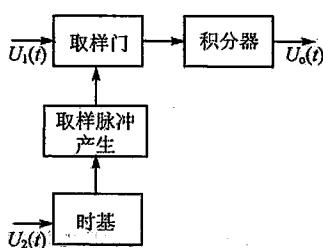


图 2.19 取样积分器原理电路图

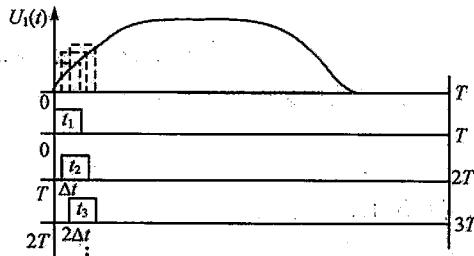


图 2.20 输入周期性信号与取样脉冲的时间关系图

设各取样脉冲坐标原点分别为 $0, T, 2T, \dots$, 各取样脉冲中点距原点时间为 t_1, t_2, t_3, \dots , 且 t_2 脉冲相对于原点 T 延迟 Δt 时间, t_3 脉冲相对于原点 $2T$ 延迟 $2\Delta t$ 时间等。因而, 相对于起始原点 0 来说, 各次脉冲的中点位置分别为时间 $t_1, t_2 + T = t_1 + \Delta t + T, t_3 + 2T = t_1 + 2\Delta t + 2T, \dots$ 。设 n_s 是扫描完 $U_1(t)$ 一个波形的所有取样脉冲数目, 显然有 $n_s \Delta t = T$, 在每一个脉冲期间对输入信号取样积分一次, 各次所得的输出信号分别为

$$\left. \begin{aligned} U_o(t_1) &= \frac{1}{RC} \int_{t_1 - (T_p/2)}^{t_1 + (T_p/2)} K_v U_1(t) dt \\ U_o(t_2) &= \frac{1}{RC} \int_{t_2 + T - (T_p/2)}^{t_2 + T + (T_p/2)} K_v U_1(t) dt \\ &\vdots \\ U_o(t_s) &= \frac{1}{RC} \int_{t_s + (s-1)T - (T_p/2)}^{t_s + (s-1)T + (T_p/2)} K_v U_1(t) dt \end{aligned} \right\} \quad (2.19)$$

式中: RC 为积分器的时间常数, T_p 为取样脉冲的宽度, K_v 为系统的电压传输系数。式(2.19)中的各式均为常数值, 且分别与各个时刻信号波形的幅度成一定的比例关系; 也就是说每取样积分一次就得到了一个与输入信号相关联的直流输出电压, 取样积分 n_s 次则得到 n_s 个直流输出电压, 这些直流输出电压按时间的编序就复现了输入信号 $U_1(t)$ 的波形。由于一个脉冲

只在一个周期内出现一次, 所以 $U_1(t)$ 的一个波形要显现出来, 需要花费 $T_s = n_s T$ 的时间(又称扫描时间)。

那么采用取样积分的方法为什么能够提高信噪比呢? 由图 2.20 可见, 因为 $T_p > \Delta t$, 那么取样脉冲扫过 T_p 时间是需要一定数量的, 设这个数目为 n_p , 即 $n_p \Delta t = T_p$ 。于是在 T_p 内每 Δt 长度上的信号波形将要重复取样 n_p 次(见图 2.21, 其中 $n_p = 5$)。其抑制噪声的思想可参见图 2.16 的同步加算平均滤波示意图, 输

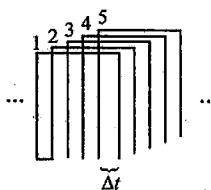


图 2.21 重复取样示意图

出信噪比与 n_p 成正比例关系, n_p 愈大, 则信噪比改善愈明显。另一方面, 当噪声通过积分器后也会被大大衰减。双管齐下, 取样积分法即可得到良好的信噪抑制比。

取样积分器是从 20 世纪 60 年代由实验室研制开始的, 到 70 年代才逐渐成为商品。它在光学、声学、电子学、磁学、无损检验、生物物理等许多方面都有重要的应用。

其他的微弱信号检测方法还有双路消噪法、同步累积法、相关分析法等, 每种方法都有其应用的特点和局限性, 在选取前一定要综合考虑各方面的要求后再作决定。

2.8 多传感器信息融合

现实世界的多样性决定了采用单一的传感器已不能全面地感知和认识自然界, 于是多传感器及其数据融合技术应运而生。根据 JDL (Joint Directors of Laboratories Data Fusion Working Group, 1986 年建立) 的定义, 多传感器数据融合 (multi-sensor data fusion) 是一种针对单一传感器或多传感器数据或信息的处理技术, 通过数据关联、相关和组合等方式以获得对被测环境或对象的更加精确的定位、身份识别及对当前态势和威胁的全面而及时的评估。

多传感器信息融合就像人脑综合处理信息一样, 可以充分利用多传感器资源, 把多传感器在空间或时间上的冗余或互补信息依据某种准则进行组合, 以获得对被测对象的一致性解释或描述。

和传统的单传感器技术相比, 多传感器数据融合技术具有许多优点, 下面列举的是一些有代表性的方面。

① 采用多传感器数据融合可以提高检测的可信度。例如采用多个雷达系统可以使得对同一目标的检测更可信。

② 降低不确定度。例如采用雷达和红外传感器对目标进行定位, 雷达通常对距离比较敏感, 但方向性不好; 而红外传感器则刚好相反, 具备较好的方向性, 但对距离测量的不确定性较大。将两者相结合, 可以使得对目标的定位更精确。

③ 改善信噪比, 提高测量精度。例如人们通常用到的对同一被测量进行多次测量, 然后取平均的方法。

④ 增强系统的鲁棒性。采用多传感器技术, 当某个传感器不工作, 失效的时候, 其他的传感器还能提供相应的信息。例如用于汽车定位的 GPS 系统, 由于受地形、高楼、隧道、桥梁等的影响, 可能得不到需要的定位信息; 如果和汽车其他常规惯性导航仪表如里程表、加速度计等联合起来, 就可以解决此类问题。

⑤ 增加对被检测量的时间和空间的覆盖程度。

⑥ 降低成本。例如采用多个普通传感器可以取得和单个高可靠性传感器相同的效果, 但成本却可以大大降低。

多传感器数据融合技术在军事和非军事领域均得到了广泛应用。在军事领域, 主要包括

以下几个方面：

① 目标自动识别。如现代空战中用到的目标识别系统就是一个典型的多传感器系统，它包括地面雷达系统、空中预警雷达系统、机载雷达系统和卫星系统等，相互之间通过数据链传递信息。地面雷达系统功率可以很大，探测距离远，但容易受云层等的影响；空中预警雷达可以做到移动探测，大大地延伸了防御和攻击距离。这两种机载雷达都采用相控阵雷达，可同时跟踪近 20 个目标，并对当前最具威胁的多个目标提出示警，引导战机同时对 4~5 个目标进行攻击。

② 自动导航。如各种技术导弹的制导，往往也采用多传感器技术，目前使用较多的有激光制导、电视制导、红外制导等，可以分为主动制导和被动制导。又如远程战略轰炸机和战斗机用到的自动导航和巡航技术等。在最近发生的几次局部战争、海湾战争和伊拉克战争中，第一波攻击都是在夜间由各种战机发起对地攻击，没有良好的导航系统是不可想象的。

③ 战场监视。打击效果检测以期作出战果评估。

④ 遥感遥测、无人侦察、卫星侦察及自动威胁识别系统。

在非军事领域，多传感器数据融合技术的应用主要包括以下几个方面：

① 工业过程监测和维护。在工业过程领域，多传感器数据融合已被广泛应用于系统故障识别和定位，以及以此为依据的报警、维护等。例如在核反应堆中就用到此类技术。

② 机器人。为了能使机器人充分了解自己所处的环境，机器人上安装有多种传感器（如 CCD 摄像头、超声传感器、红外传感器等），多传感器数据融合技术使得机器人能够作为一个整体，自由、灵活、协调地运动，同时识别目标，区分障碍物并完成相应任务。

③ 医疗诊断。多传感器数据融合技术在医学领域也得到了广泛应用。如将采用 CT (Computerized Tomography)、核磁共振成像、PET (Positron Emission Tomography) 和光学成像等不同技术获得的图像进行融合，可以对肿瘤等病症进行识别和定位。

④ 环境监测。例如每天都接触到的天气预报，实际上是对卫星云图、气流、温度、压力以及历史数据等多种传感器信息进行融合后作出的决策推理。

2.8.1 多传感器数据融合框架

多传感器数据融合中将传感器看作是一种获取被感知环境在给定时刻信息的装置，可定义为一个具有两个自变量的函数，一个自变量为被感知环境；另一个自变量为时间。用数学公式表达为

$$S(E, t) = \{V(t), e(t)\} \quad (2.20)$$

式中： E 、 t 为自变量； V 为映射结果； e 为不确定度。

从多传感器数据融合的角度来看，传感器相互之间的工作方式主要分为三种，即互补方式、竞争方式和协同方式。在互补方式下，各个传感器工作相互独立，互不依赖，各传感器提供的信息相互之间形成互补关系，没有重合部分，通过将各传感器信息组合可以得到对被测对象



更加全面的认识。图 2.22 是由 R. Brooks 和 S. Iyengar 描述的关于雷达的例子。由于受观测角度的影响,每个雷达系统都只能观测到整个区域 1/4 部分的范围,将 4 个雷达系统提供的信息综合起来,就可以得到对整个区域的观测信息。在竞争模式下,各传感器提供对同一信息的不同测量结果,图 2.23 仍然是关于雷达探测的例子,图中采用 4 个雷达系统对同一区域进行观测,因而增强了系统的可靠性和鲁棒性,即当某些雷达系统不能正常工作时仍能得到被观测区域的信息。在协同工作模式下,各传感器独立工作,通过将其各自获取的信息进行融合,从中提取出由任意单一传感器均无法获得的信息。例如采用 2 个 CCD 摄像头从不同的角度对某一物体进行观测,任何一个 CCD 传感器都可以获得关于被测物体的深度的信息;如果来自 2 个 CCD 传感器的数据是匹配的,根据 2 个传感器观测角度之间的关系,就可以获得被测物体的三维信息,而单独采用 2 个 CCD 传感器中的任何一个都不可能获得被测物体的三维信息。又如可以采用多个温度传感器来测量某一温度场中不同点的温度,在获得各传感器数据的基础上可以根据各传感器之间的位置关系推算出该温度场的温度梯度,而温度梯度是任何一个温度传感器都无法直接提供的。

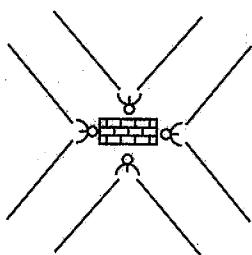


图 2.22 互补式传感器工作示例图

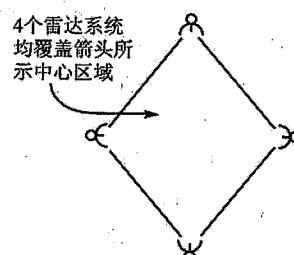


图 2.23 竞争式传感器工作示例

在实际使用中,传感器的各种工作模式并不是互相排斥的,一个多传感器系统中不同的传感器可以工作在多种方式下。对于不同工作方式下的传感器,其数据融合也略有不同。互补工作方式下的传感器,其数据融合相对较为简单,一般只需简单地将它们结合起来获得对被测对象更完整的信息即可。对于协同工作方式下的传感器,其数据融合需构造全新的抽象传感器以获得由各传感器实体无法获得的信息。对于竞争工作方式下的传感器,其数据融合涉及冗余信息的处理,融合过程相对较为复杂。

下面介绍多传感器数据融合中经常用到的“可靠抽象传感器”(reliable abstract sensor)的概念。为了说清楚这种传感器的概念,先对有关的三种传感器,即实际(具体)传感器、抽象传感器和可靠抽象传感器作一比较。

实际传感器可以获取待测物理量的信息。

抽象传感器 A 是一分段连续函数,函数来自在窄密集区间 $[l_A, \mu_A]$ ($l_A < \mu_A$) 上物理量的效应。 l_A, μ_A 为实数,分别是抽象传感器 A 区间的上、下限。直观上,这一区间表示的是取自

所采集到的物理量值,它可能受到物理过程不确定度、系统延迟以及处理器处理策略等的影响。一个抽象传感器如果没有包含物理变量的准确值或者区间太宽,就可能出现故障,这个故障可能是由于实际传感器有故障、环境噪声、系统连接故障等因素引起的。

应用抽象传感器,以实际传感器作为输入,采用相应的故障容许算法,便得到可靠抽象传感器。它始终包含了物理量在某时间间隔的信息、区间,所得结果基本是准确的。

下面给出相关术语的正式描述(可视作定义):

① 一个抽象传感器的宽度定义为 $\mu_A - l_A$ 。宽度决定了抽象传感器的精度,宽度越窄,精度越高。然而宽度常常是确定的。

② 一个抽象传感器被称为是“正确抽象传感器”,表示其区间内包含了物理量的实际值,区间不超过 Γ 。参数 Γ 由设计人员指定。任两个正确抽象传感器 A 和 B 必定是相交有重叠(重合)的,因为它们均含有被测物理量的值。特定情况下,这样的两个传感器允许 $\mu_A < l_B$ 。

③ 如果有 $(\mu_A - l_A) < (\mu_B - l_B)$, 则表示正确传感器 A 的精度高于正确传感器 B。

④ 如果 $l_A \leq l_B$ 和 $\mu_A \geq \mu_B$, 则区间 A 完全重叠区间 B。

⑤ 给定 n 个独立的抽象传感器,假设在进行测量的时间间隔内有不多于 f 个抽象传感器有故障。这个假设通常被理解为:一旦有一个传感器出现并持续保持故障,则有故障的传感器不会多于 f 个。需要注意的是:这里的故障传感器指的是抽象传感器,不是实际物理传感器。确认抽象传感器为故障传感器有两种情况:一是物理传感器获得的值不准确;二是抽象传感器的区间大于参数 Γ 的指定值。可以认为:任一 $n-f$ 个或更多区间的交叠可能包含真值,而少于 $n-f$ 个区间的交叠则不包含真值。

构建一个可靠抽象传感器,可以确保包含被测量的真值,其宽度至少为 Γ 。构建的方法有两种:

① 构建为有限宽度的单一正确区间。这个区间可以取所有可能区间的“并”而得到。

② 是 $n-f$ 个最小数目的区间集或者包含有物理量真值的更多区间的交集。

显然对于①,通过取区间的“并”得到正确区间,但正确区间并不唯一,所确定的正确区间内的点不一定属于其他可能的正确区间。若某个抽象传感器被认为是正确区间,则为了提高精度,可靠抽象传感器内某些区间就可能被丢弃,但在对物理量值进行采集的时刻,所认定的正确区间获得的信息可能偏偏不可用,因此,采取方法②显得更合理一些。

将单一正确区间定义为“生成区间”,可靠抽象传感器的宽度即为生成区间的宽度。

构建可靠抽象传感器的基本原则由 K. Marzullo 提出。他提出了一种所谓的“故障容许传感器的平均算法”。这种算法采用“区间图”的形式,对 n 个区间集,进行确定故障容许的算法,用于故障传感器的检测,并当故障传感器数目限定为 f 个时,来限定可靠抽象传感器的宽度。K. Marzullo 经理论证明得出:当 $f < n/2$ 时,生成区间的宽度与一个抽象传感器一样窄,甚至还要窄;而当 $f > n/3$ 时,生成区间的宽度与一个正确传感器相同。由于 $f \geq n/2$ 时,两个正确传感器之间的距离可能是任意大小,因此可靠抽象传感器的宽度不能被限定,从而均假设

$f < n/2$ 。算法的详情可参看相关文献。

可靠抽象传感器是为了解决多传感器系统中的传感器失效问题,即一个多传感器构成的测量系统,在最多允许多个传感器失效的情况下仍能得到正确的测量结果。K. Marzullo 提出,对于 N 个传感器构成的系统,能够获得可靠数据的前提是允许出错的传感器数目 $< N/2D$,其中 D 是测量数据的维数。

为了更好地理解什么是可靠抽象传感器,下面举例说明。用三个温度计来测量某点的温度,各自给出的测量值和不确定度都满足温度计的精度要求,分别为 $(15 \pm 2)^\circ\text{C}$ 、 $(16 \pm 2)^\circ\text{C}$ 和 $(17 \pm 2)^\circ\text{C}$,则采用 Marzullo 规则,可能的正确温度范围为 $14 \sim 18^\circ\text{C}$ 。具体只需将各温度传感器数据范围列出(见表 2.2),从表中可以找到有两个或两个以上温度传感器数据重合的部分就是可靠抽象传感器的输出(对于 N 为 3, D 为 1, 允许出错的传感器个数为 1),表中“ \times ”部分表示各传感器温度范围,“○”部分为可靠抽象传感器输出温度范围。

表 2.2 可靠抽象传感器示例

温度/ $^\circ\text{C}$ 类 别	12	13	14	15	16	17	18	19	20
传感器 1		\times	\times	\times	\times	\times			
传感器 2			\times	\times	\times	\times	\times		
传感器 3				\times	\times	\times	\times	\times	
可靠抽象传感器				○	○	○	○	○	

2.8.2 多传感器数据融合结构

在多传感器系统中,各种传感器的数据具有不同特征,可能是实时的或非实时的、模糊的或确定的、互相支持的或互补的,也可能是互相矛盾的或竞争的。多传感器数据融合与经典的信号处理方法也存在本质的区别,数据融合所处理的多传感器具有更复杂的形式。根据信息处理的不同层次,传感器数据融合可分为数据级、特征级和决策级。

在数据级融合中主要是将来自不同传感器的数据直接组合后得到统一的输出数据,例如 2.8.1 节讲过的抽象传感器的例子就属于这一类。数据级融合的主要目的是为了获得对被观测对象的统一的数据描述,用到的关键技术包括数据转化、相关和关联等。特征融合主要是为了获得关于被测对象的统一特征描述,可以根据各个传感器的数据直接融合出特征,也可以根据各个传感器数据分别提取出特征,然后再融合。决策融合是多传感器数据融合的最终目标;同样,也可以由特征融合后得到统一决策;也可以根据单一传感器作出的决策最后再融合得到统一的决策。

图 2.24 给出了多传感器数据融合的基本构架。根据系统输入/输出的特点,多传感器数



据融合又可以分为以下几种类型。

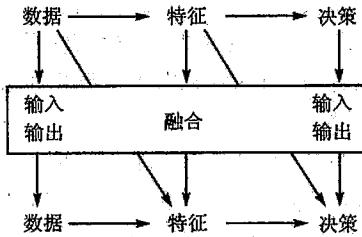


图 2.24 多传感器数据融合的层次

① Data In-Data Out, 即输入为数据, 输出也为数据。其主要采用传统的信号处理技术等来完成相应工作。

② Data In-Feature Out, 即输入为数据, 融合输出的结果为特征。一个典型的例子就是人对物体深度的感知, 即可以从双眼看到的物体图像中融合提取出物体深度信息。

③ Feature In-Feature Out, 即输入为特征, 融合输出的结果也为特征。

④ Feature In-Decision Out, 即输入为特征, 融合输出的结果为决策。

⑤ Decision In-Decision Out, 即输入为决策, 融合输出的结果也为决策。

根据数据融合发生的地点, 多传感器数据融合系统可以分为集中式(centralized)和分布式(decentralized 或 distributed)。集中式融合的特点是存在一个融合中心, 它收集来自所有传感器子系统的数据、特征或决策并完成融合计算。与集中式融合相对应, 分布式融合中没有明显的融合中心, 各传感器系统都可以看作一个融合中心, 它们通常构成一个网络, 通过通信获得其他传感器的数据, 并不同程度地完成融合计算。

根据数据融合中各传感器连接的方式, 多传感器数据融合系统又可分为串行、并行和串并行混合式。串联型多传感器数据融合是指先将两个传感器数据进行一次融合, 再把融合的结果与下一个传感器数据进行融合, 依次进行下去直至所有的传感器数据都融合完为止。串联融合时, 每个传感器既具有接收数据、处理数据的功能, 又具有信息融合的功能, 各传感器的处理同前一级传感器输出的信息有很大关系, 最后一个传感器综合了所有前级传感器输出的信息, 得到的输出将作为串联融合系统的结论。因此, 串联融合时, 前级传感器的输出对后级传感器输出的影响比较大。并联型多传感器数据融合是指所有传感器输出数据都同时输入给数据融合中心, 传感器之间没有影响, 融合中心对各种类型的数据按适当的方法进行综合处理, 最后输出结果。并联融合时, 各传感器的输出之间不会相互影响。串并联混合型多传感器数据融合是指串联和并联两种形式的综合, 可以先串联后并联, 也可以先并联后串联。串行融合和并行融合的结构如图 2.25 所示。

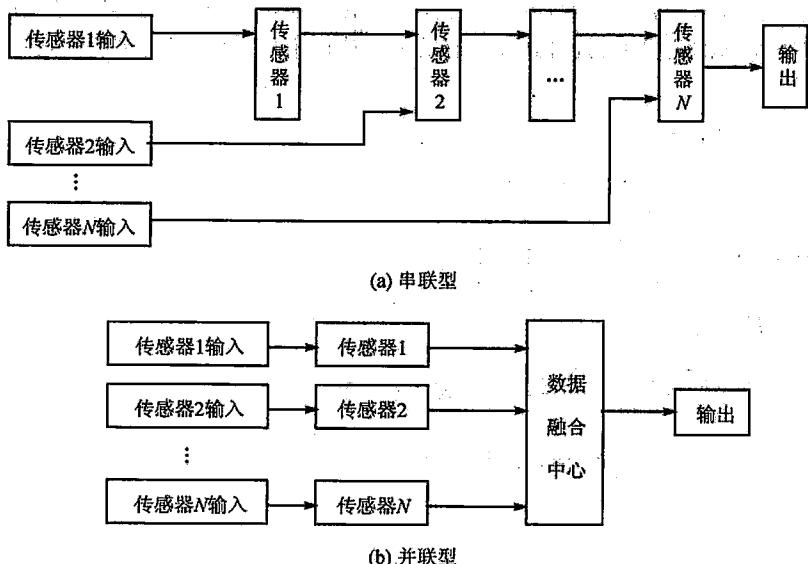


图 2.25 串行融合和并行融合结构

2.8.3 多传感器数据融合算法

多传感器数据融合虽然未形成完整的理论体系,但不少应用领域根据各自的具体应用背景,已经形成了许多成熟并且有效的融合方法。下面对目前几种使用较多的融合算法加以介绍。

1. 基于卡尔曼滤波的多传感器数据融合方法

1959年,美国太空署(NASA)开始研究载人登月计划。当时有两个非常棘手的问题:一是飞船的中途制导和导航;二是液体燃料助推器在大扰动条件下的自动驾驶。导航问题中主要需要解决对太空船运动状态的估计。其测量信息主要来自三个子系统:飞船设备的惯性测量装置、天文观测仪和地面的测轨系统。匈牙利人R. E. Kalman提出了后来以其命名的卡尔曼滤波器,成功地将其应用于解决导航问题。

为了说明卡尔曼滤波在多传感器数据融合中的应用,先从一个简单的例子开始。假设对某一物理量进行测量,如果测量了 N 次,且假设每次测量使用同一传感器,各次测量互相独立且其精度相同,则可以通过求这 N 次测量的平均值来获得对被测未知量的一个更准确的估计。用 X_1, X_2, \dots, X_N 来表示 N 次测量值,它们均可以看作是随机变量,且服从方差为 σ^2 的正态分布,则对被测量的估计可写为

$$\hat{Y}_N = (X_1 + X_2 + \dots + X_N)/N \quad (2.21)$$

从多传感器数据融合的角度,可以看作是对单传感器在不同时间获得的数据的一种融合。如果每得到一个新的测量值,就将其和已有的测量值做平均,就可以得到和上述过程相对应的递推公式。

第一步: $\hat{Y}_1 = X_1, \sigma^2 = \sigma^2$

$$\begin{cases} \hat{Y}_2 = \frac{(X_1 + X_2)}{2} = \hat{Y}_1 + K_2(X_2 - \hat{Y}_1) \\ \sigma_2^2 = \frac{1}{2}\sigma^2 = \frac{1}{2}\sigma_1^2 = (1 - K_2)\sigma_1^2 \\ K_2 = \frac{1}{2} \end{cases}$$

$$\begin{cases} \hat{Y}_3 = \frac{(X_1 + X_2 + X_3)}{3} = \hat{Y}_2 + K_3(X_3 - \hat{Y}_2) \\ \sigma_3^2 = \frac{1}{3}\sigma^2 = \frac{2}{3}\sigma_2^2 = (1 - K_3)\sigma_2^2 \\ K_3 = \frac{1}{3} \end{cases}$$

⋮

$$\begin{cases} \hat{Y}_n = \frac{(X_1 + X_2 + \dots + X_n)}{n} = \hat{Y}_{n-1} + K_n(X_n - \hat{Y}_{n-1}) \\ \sigma_n^2 = \frac{1}{n}\sigma^2 = \frac{n-1}{n}\sigma_{n-1}^2 = (1 - K_n)\sigma_{n-1}^2 \\ K_n = \frac{1}{n} \end{cases}$$

对上述递推过程可作如下解释,即对当前被测对象的估计可以看作是由两部分构成:一部分是根据上一步的估计得到的对当前被测对象的预测;另一部分是放大的“新息”,即当前观测值和预测值的差,放大的“新息”可以看作是对当前预测值的修正。

下面来看多传感器的情况,为了注重原理,仍然采用最简单的例子。假设用两个传感器来测量某个物理量 x (比如用激光和尺子来测量长度),它们分别给出的测量值为 x_1 和 x_2 ,两者满足如下的正态分布:

$$p(x_i) = \frac{1}{\sigma_i \sqrt{2\pi}} e^{-\frac{(x_i - \bar{x}_i)^2}{2\sigma_i^2}} \quad (i = 1, 2) \quad (2.22)$$

式中: \bar{x}_1 和 \bar{x}_2 分别为均值, σ_1^2 和 σ_2^2 分别为方差。

关于如何考虑两个传感器的测量结果,实际上就是一个多传感器数据融合的问题。传统数据处理方法已经给出了这个问题的答案,即加权平均。由于两个传感器给出的测量结果方差不一样,可以利用它们使得加权后的结果方差最小。如用 \hat{x} 表示加权平均的结果,则有

$$\hat{x} = \omega x_1 + (1 - \omega)x_2 \quad (2.23)$$



由于两次测量相互独立,且都满足正态分布,所以加权后的方差满足

$$\hat{\sigma}^2 = \omega^2 \sigma_1^2 + (1 - \omega)^2 \sigma_2^2 \quad (2.24)$$

为了使得加权后的方差最小,权重 ω 的选择应满足 $\partial\hat{\sigma}/\partial\omega=0$ 。进一步可推导出:

$$\left. \begin{array}{l} \omega = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} \\ \hat{x} = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} x_1 + \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} x_2 \\ \hat{\sigma}^2 = \frac{\sigma_1^2 \sigma_2^2}{\sigma_1^2 + \sigma_2^2} \end{array} \right\} \quad (2.25)$$

对式(2.25)的物理意义可以作出如下解释,即加权平均的过程实际上是根据它们各自的精度将两次测量值重新组合后得到对被观测量的估计,精度高(方差小)的取较大的权重,精度低(方差大)的取较小的权重,而且加权后的方差比任何一次测量值的方差都小。通过加权平均,对被观测量的估计精度可以得到改善,不确定度降低。

对上述加权平均过程同样可以用递推形式来表示。假设两个传感器测量系统提供的测量值是依次得来的,则

第一步: $\hat{x}_1 = x_1, \hat{\sigma}_1 = \sigma_1$

第二步: $\left. \begin{array}{l} \hat{x}_2 = x_1 + K_2(x_2 - \hat{x}_1) \\ \hat{\sigma}_2^2 = (1 - K_2)\hat{\sigma}_1^2 \\ K_2 = \hat{\sigma}_1^2 / (\hat{\sigma}_1^2 + \sigma_2^2) \end{array} \right.$

从两个例子的递推形式可以看出,不论是前面第一个关于单传感器多数据算术平均的例子,还是多个传感器数据加权平均的例子,其本质都是相同的。

虽然本节一开始就提到了卡尔曼滤波,但到目前为止并没有给出其数学表示,下面就具体给出其数学表示。考虑下面两式给出的随机线性离散系统模型:

$$\mathbf{X}_k = \Phi_{k|k-1} \mathbf{X}_{k-1} + \Gamma_k \mathbf{W}_k \quad (2.26)$$

$$\mathbf{Z}_k = \mathbf{H}_k \mathbf{X}_k + \mathbf{V}_k \quad (2.27)$$

式中: k 表示当前时刻; \mathbf{X}_k 为系统的 n 维状态向量; \mathbf{Z}_k 为系统的 m 维观测向量; \mathbf{W}_k 为系统的 p 维随机向量; \mathbf{V}_k 为系统的 m 维观测噪声向量; $\Phi_{k|k-1}$ 为 $k-1$ 时刻到 k 时刻系统的 $n \times n$ 维状态转移矩阵; Γ_k 为 $n \times p$ 维干扰输入矩阵; \mathbf{H}_k 为 $m \times n$ 维观测矩阵; \mathbf{W}_k 和 \mathbf{V}_k 都是均值为 0 的高斯噪声,其方差矩阵分别为 \mathbf{Q}_k 和 \mathbf{R}_k ,即

$$\mathbf{E}(\mathbf{W}_k) = \mathbf{0}, \quad \mathbf{E}(\mathbf{W}_k \mathbf{W}_k^T) = \mathbf{Q}_k \quad (2.28)$$

$$\mathbf{E}(\mathbf{V}_k) = \mathbf{0}, \quad \mathbf{E}(\mathbf{V}_k \mathbf{V}_k^T) = \mathbf{R}_k \quad (2.29)$$

根据状态向量和观测向量在时间上存在的不同对应关系,可以把估计问题分为滤波、预测和平滑。用 $\hat{\mathbf{X}}_{kj}$ 表示根据 j 时刻及 j 以前时刻的观测值对 k 时刻状态 \mathbf{X}_k 作出某种估计,则可以写出递推形式的卡尔曼滤波方程:

$$\hat{X}_{k|k-1} = \Phi_{k|k-1} \hat{X}_{k-1} \quad (2.30)$$

$$\varepsilon_k = Z_k - H_k \hat{X}_{k|k-1} \quad (2.31)$$

$$\hat{X}_k = \hat{X}_{k|k-1} + K_k \varepsilon_k \quad (2.32)$$

$$K_k = P_{k|k-1} H_k^T (H_k P_{k|k-1} H_k^T + R_k)^{-1} \quad (2.33)$$

$$P_k = (I - K_k H_k) P_{k|k-1} \quad (2.34)$$

$$P_{k|k-1} = \Phi_{k|k-1} P_{k-1} \Phi_{k|k-1}^T + \Gamma_{k|k-1} Q_k \Gamma_{k|k-1}^T \quad (2.35)$$

式中： I 为单位矩阵； ε_k 为“新息”； K_k 为增益矩阵； $P_{k|k-1}$ 和 P_k 分别称为预测方差矩阵和估计方差矩阵； $P_{k|k-1}$ 和 P_k 的定义为

$$P_{k|k-1} = E[(X_k - \hat{X}_{k|k-1})(X_k - \hat{X}_{k|k-1})^T] \quad (2.36)$$

$$P_k = E[(X_k - \hat{X}_k)(X_k - \hat{X}_k)^T] \quad (2.37)$$

鉴于篇幅所限，这里不再对卡尔曼滤波的导出过程进行详细解释，读者可参阅相关文献与书籍，其基本思想是增益矩阵的选取应该使得估计方差最小。图 2.26 描述了卡尔曼滤波的原理和过程，有助于加深理解。

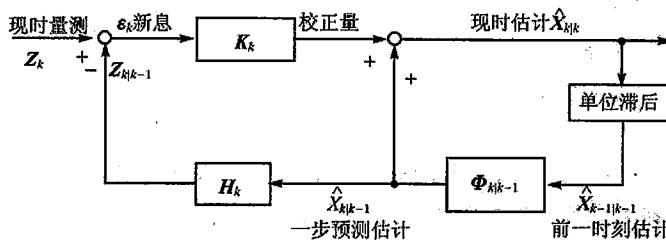


图 2.26 卡尔曼滤波原理示意图

卡尔曼滤波的基本思路是用当前量测值与上一时刻的预测估计值的偏差（称为“新息”）乘以一定的权重 K_k 来不断修正下一状态的估计。从式(2.32)可以看出：权重大表示对观测值的依赖增大，对先前状态的估计依赖相应减小；权重小表示对观测值的依赖减小，对先前状态的估计依赖增大。而权重大小的选择是通过使估计误差 P_k 达到最小来实现的。卡尔曼滤波正是不断通过调整对量测和估计的依赖程度，快速而有效地实现对被估计状态的最佳估计。

卡尔曼滤波和多传感器的数据融合有何关系呢？本节一开始所举的两个例子实际上就是卡尔曼滤波的特例（对应于状态转移矩阵 $\Phi_{k|k-1}$ 、观测矩阵 H_k 和干扰输入矩阵 Γ_k 均为 1 的情况），它们可以看作利用卡尔曼滤波对单一传感器在不同时刻多数据源的融合以及对多传感器数据的融合，融合的结果提高了精度，降低了不确定度。

2. 基于贝叶斯决策的多传感器数据融合方法

贝叶斯决策的基础是贝叶斯统计学。下面就贝叶斯统计学中用到的基本概念和公式加以



简单介绍。

设事件 A 和 B 均为事件域 E 中发生的事件,且事件 B 发生的概率 $P(B) > 0$,则有

$$P(A | B) = \frac{P(A \cap B)}{P(B)} \quad (2.38)$$

式中: $P(A|B)$ 为事件 B 发生条件下事件 A 发生的概率; $P(A \cap B)$ 为事件 A 和 B 同时发生的概率。

式(2.38)为传统概率论中的条件概率公式,可以进一步推出:

$$P(A | B) = \frac{P(B | A)P(A)}{P(B)} \quad (2.39)$$

式(2.39)是贝叶斯统计理论最基本的公式。进一步推广,设事件域 E 由 B_1, B_2, \dots, B_k 等 k 个非关联的事件构成, A 为 E 中的某一事件,则

$$P(B_1 \cup B_2 \cup \dots \cup B_k) = 1 \quad (2.40)$$

$$A = (A \cap B_1) \cup \dots \cup (A \cap B_k) \quad (2.41)$$

$$P(A) = \sum_k P(A \cap B_k) = \sum_k P(A | B_k)P(B_k) \quad (2.42)$$

$$P(B_k | A) = \frac{P(A \cap B_k)}{P(A)} = \frac{P(A | B_k)P(B_k)}{\sum_k P(A | B_k)P(B_k)} \quad (2.43)$$

下面举例说明上述公式的具体应用及贝叶斯统计学中的有关概念。假设某一公司在其下属的 5 个分厂都生产同一种产品,分销商向该公司成批量进货,为了保证进货的质量,分销商将在每次进货的时候进行抽检。其具体做法是从该批货物中任意抽取 3 个样品,若其中有 1 个是次品,则认为整批货物都是次品,从而实行退货处理。现假设 5 个分厂中第 2 个分厂的次品率为 5%,其余 4 个分厂的次品率均为 2%。假设分销商检查到某一批货物为次品,问该批产品来自第 2 分厂的概率是多少?用 r 表示分厂编号,取值可从 1 取到 5。用 $P[F]$ 表示某批货物来自第 r 个分厂的概率,在没有其他已知条件的情况下,假设某批货物来自各个分厂的概率相等,即 $P[(F=r)] = 1/5$,这个概率通常也被称为先验概率。用条件概率 $P[D|F]$ 表示来自第 r 分厂的产品的次品率,这个概率通常也被称为似然函数。根据已知条件有

$$P[D | (F=r)] = 0.02 \quad (r = 1, 3, 4, 5); \quad P[D | (F=2)] = 0.05$$

现在需要求条件概率 $P[(F=2)|D]$,通常被称为后验概率。根据贝叶斯公式,后验概率为

$$P[(F=2) | D] = \frac{P[(F=2) \cap D]}{P(D)} = \frac{P[D | (F=2)]P[(F=2)]}{P(D)} = 0.37$$

通过上述例子,有助于掌握贝叶斯统计学中的几个基本概念,即先验概率、似然函数和后验概率。贝叶斯决策则建立在贝叶斯统计学的基础上,鉴于篇幅,此处只简单介绍下述三种基本方法。

① 最大后验概率决策。对应前面讲述的关于产品检验的例子,在满足条件 $P[(F=2)|$

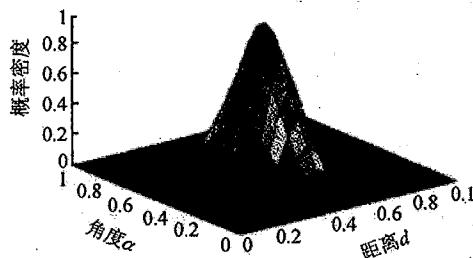


$D] > P[(F \neq 2)]$ 的前提下,作出的决策为 $F=2$,即结论是该批产品来自第 2 分厂。

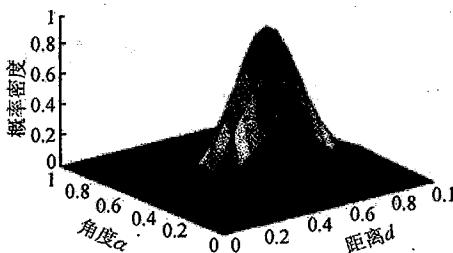
② 极大似然决策。对应前面讲述的关于产品检验的例子,在满足条件 $P[D|(F=2)] > P[D|(F \neq 2)]$ 的前提下,作出的决策为 $F=2$ 。

③ Neyman-Pearson 决策。在满足条件 $\frac{P[D|(F=2)]}{P[D|(F \neq 2)]} > t$ (t 为域值) 时,作出决策为 $F=2$ 。

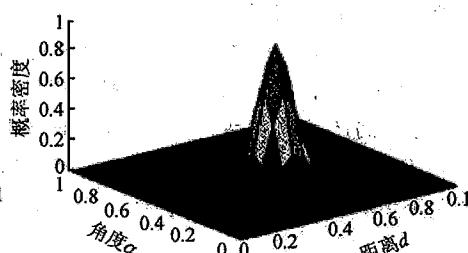
前面简单介绍了贝叶斯统计和贝叶斯决策最基本的知识,它们都可以用在多传感器数据融合当中。假设在机器人上安装两种传感器来对机器人行进前方障碍物进行检测和定位,其中一种为雷达,另一种为红外传感器。由于两种传感器自身的特点,雷达对于被测物体的距离比较敏感,但方向性不好。红外传感器则刚好相反,其对角度比较敏感,对距离则不太敏感。假设两种传感器检测结果的概率密度函数已知,分别如图 2.27(a) 和(b) 所示,则根据贝叶斯公式可对二者进行融合,得到在两种传感器观测数据下关于被测物体角度和距离的概率密度函数,如图 2.27(c) 所示。



(a) 红外传感器的概率密度



(b) 雷达的概率密度



(c) 数据融合后的概率密度

图 2.27 采用雷达和红外传感器的机器人障碍物定位的融合计算过程

从图 2.27 中可以看出,数据融合后的结果其角度和距离的不确定度均要优于任何一种传感器的结果,充分说明了贝叶斯决策在传感器数据融合中的作用。

2.8.4 多传感器数据融合实例

如前所述,多传感器数据融合技术自20世纪70年代出现后,逐步在航空航天、遥感遥测、机械制造、医学等领域得到了广泛应用。为加深读者对多传感器数据融合技术尤其是有关融合算法的理解,下面以多传感器数据融合技术在多机器人系统定位中的利用为例加以说明。

在移动机器人定位相关研究中,主要包括移动机器人自身的定位(即在一个事先设定或是完全未知的环境中如何确定机器人自身的位置)和目标的定位(包括机器人行进路径上障碍物的定位以便于机器人选择合适的路径从而避开障碍物,也包括机器人行进目标位置的定位),在此基础上机器人可以进一步判断自身与周围物体的相对关系。下面关于多传感器数据融合技术在多机器人系统定位中应用的相关例子主要是根据Carnegie Mellon大学机器人研究所的Ashley W. Stroupe等人的研究成果。

首先针对图2.28所示的简单观测情况。1、2两个机器人分别从水平和垂直方向对某一目标物体进行观测。为简化问题,假设观测数据呈高斯分布,且其不确定度随机器人到目标位置距离的增加而加大。它们均满足如下的高斯分布:

$$p(\mathbf{X}_i) = \frac{1}{2\pi \sqrt{|\mathbf{C}_i|}} e^{-\frac{1}{2}(\mathbf{x}_i - \bar{\mathbf{x}}_i)^T \cdot \mathbf{C}_i^{-1} \cdot (\mathbf{x}_i - \bar{\mathbf{x}}_i)} \quad (i=1,2) \quad (2.44)$$

式中:下标*i*代表机器人编号; \mathbf{X}_i 为机器人对目标位置二维坐标的观测值; $\bar{\mathbf{x}}_i$ 为其均值; \mathbf{C}_i 为观测数据的协方差矩阵,可以表示为 $\mathbf{C}_i = \begin{bmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{bmatrix}$ 。

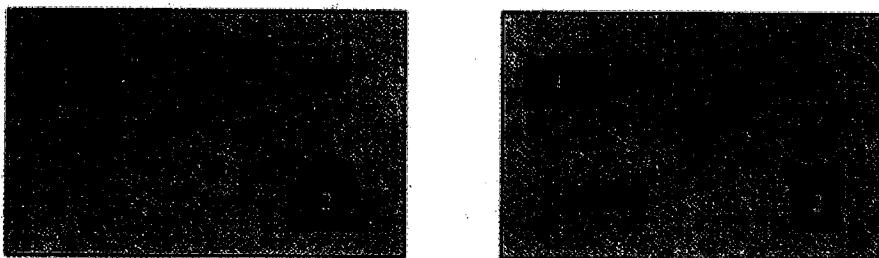


图2.28 简单观测情况

对于图2.28所示情况,取 $\bar{\mathbf{x}}_1 = \begin{bmatrix} 10 \\ 10 \end{bmatrix}$, $\bar{\mathbf{x}}_2 = \begin{bmatrix} 12 \\ 10 \end{bmatrix}$, $\mathbf{C}_1 = \begin{bmatrix} 25 & 0 \\ 0 & 4 \end{bmatrix}$, $\mathbf{C}_2 = \begin{bmatrix} 1 & 0 \\ 0 & 4.84 \end{bmatrix}$,对 \mathbf{X}_i 的融合可以采用二维卡尔曼滤波算法

$$\hat{\mathbf{X}} = \mathbf{X}_1 + \mathbf{C}_1(\mathbf{C}_1 + \mathbf{C}_2)^{-1}(\mathbf{X}_2 - \mathbf{X}_1)$$

图2.29给出了融合前后各机器人观测数据的概率分布和融合后的概率分布,从中可以看出融合后不确定度得到了很大改善。

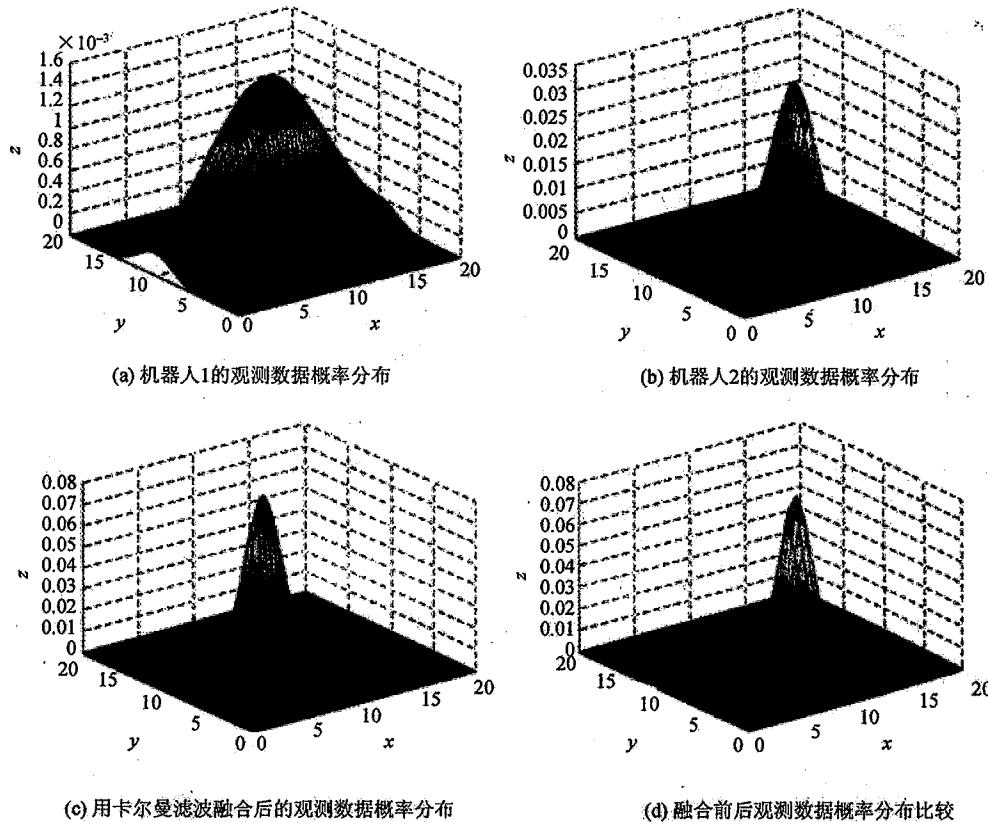


图 2.29 融合前和融合后各机器人观测数据的概率分布

对于更一般的情况,当传感器观测方向和水平方向或垂直方向不重合时,在数据融合时需要进行必要的坐标变换。详细步骤可以参考 Ashley W. Stroupe 等人的论文,在此不再加以详细介绍。

2.9 模糊技术及其应用

众所周知,人类思维的重要特征之一就是能对模糊事件进行识别和判断。人对传感器信号的处理,采用的是一种高级的模糊算法,从而使能够利用低精度的传感信号,以及低速度、低精度的运算,作出许多准确而有效的判断。

传统的传感器测量是一种数值测量,它将被测量映射到实数集合中,以数值符号的形式来

描述被测量状态,因此称之为数值传感器。它一方面具有精度高、无冗余的优点;另一方面又存在测量结果不易理解,数值存储量大和涉及人类自身行为以及某些高层逻辑信息难以描述的问题。

1965年美国加利福尼亚大学控制论专家L. A. Zadeh发表了题为*Fuzzy Sets*的论文,提出了模糊集合理论,为处理客观世界中存在的模糊性问题提供了有利工具。随着科学技术的迅猛发展,模糊理论和方法日臻完善,已广泛应用于自然科学和社会科学的各个领域。模糊逻辑在传感器测量领域的应用和模糊传感器的出现是测量学科迅速发展的结果,不仅拓宽了经典测量学科,而且使测量科学向人类的自然语言理解方面迈出了重要的一步。

基于模糊逻辑技术的模糊传感器的研究最早见于20世纪80年代末,但到目前为止,尚没有严格、公认的定义。一般认为,模糊传感器是在经典传感器数值测量的基础上,经过模糊推理与知识集成,以自然语言符号描述的形式输出测量结果。因此,信息的符号表示与符号信息系统的研究是模糊传感器研究的基石。下面就先从模糊数学的基本概念讲起,最后给出模糊传感器的基本结构及其应用实例。

2.9.1 模糊数学基础

1. 模糊集合

精确集合(或经典集合)的特点是具有精确的边界,例如,实数大于6的精确集合可以表示为

$$A = \{X \mid X > 6\} \quad (2.45)$$

式(2.45)中的边界没有任意性,大于6属于集合A,否则就不属于A。这里引入隶属函数的概念,定义为 $\mu_A(X)$,则

$$\mu_A(X) = \begin{cases} 1, & X \in A \\ 0, & X \notin A \end{cases} \quad (2.46)$$

精确集合适用于不同的应用领域,且是数学和计算机科学的重要工具,但其不能反映人的思维方法和概念。人们表达某一事物或对象时,往往是抽象的和不精确的。例如,说某人“个儿高”或“个儿矮”(“距离长”或“距离短”)时,不能说身高1.80 m的人“高”,1.79 m的人就“矮”(5 m“长”,4.9 m就“短”)。所以“高”与“矮”之间的界限是模糊的、平滑的,因此需要一种模糊集合来描述,这种集合没有精确的边界,从“属于一个集合”到“不属于一个集合”是逐渐过渡的。因此,模糊集合的概念得以提出。

如果X是对象x的集合,则将X的模糊集合A定义为有序对的集合,即

$$A = \{(x, \mu_A(x)) \mid x \in X\} \quad (2.47)$$

式中:X称为论域, $\mu_A(x)$ 称为模糊集A的隶属函数(或简写为MF)。

隶属函数可把X中每一个元素映射成0与1之间的隶属度(或隶属值)。显然,模糊集合是精确集合的推广,表现在精确集合的隶属函数的值只允许有1或0;而模糊集的隶属函数可

以在 0 与 1 之间连续变化。模糊集合的构造取决于确立合适的论域和适当的隶属函数。隶属函数的指定是主观的。对同一概念，不同的人所指定的隶属函数会有巨大的差别。这个主观性来源于个人感受和表达抽象概念上的差异，但是与随机性无关，因为它是客观现象。

当论域 X 为连续空间时，人们往往把 X 划分为几个模糊集合，它们的隶属函数大体上以一致的方式覆盖论域 X ，这些模糊集合通常具有一个与日常形容语句相符的名字，如“大”、“中”、“小”等，把它们称做语言值或语言标识，则论域 X 称做语言变量。

语言变量是以多元组 $(x, T(x), X, G, M)$ 为特征的，其中： x 是变量的名称； $T(x)$ 是 x 术语的集合，即 x 的语言值名称的集合； G 是产生 x 值名称的句法规则； M 是与各值含义有关的语法规则。

例如：用热敏电阻作为敏感元件检测某热源的温度，若被测温度范围为 $0 \sim 120^{\circ}\text{C}$ ，是可能温度的集合，则模糊集合 B = “温度 57.5°C 左右”可表示为

$$B = \{x, \mu_B(x) \mid x \in X\}$$

式中

$$\mu_B(x) = \frac{1}{1 + \left(\frac{x - 57.5}{10}\right)^4}$$

该隶属函数可表示在图 2.30 中。

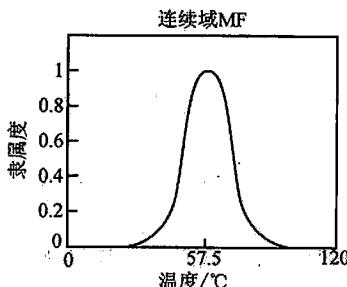


图 2.30 模糊集合“温度 57.5°C 左右”

这时它的术语集合 t (温度)可以是：

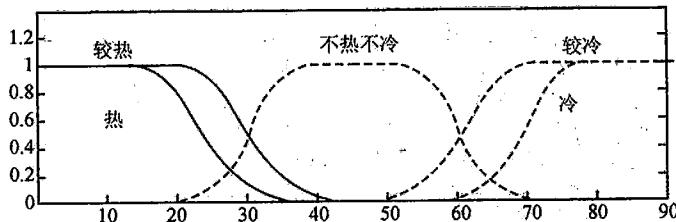
$$t(\text{温度}) = \{\text{热、较热、不热不冷、较冷、冷}\}$$

式中： t (温度)中每一项由论域 $\{0, 120\}$ 中一个模糊集合来表征，如图 2.31 所示。图中表示了几种典型的隶属函数。

句法规则是术语集 t (温度)语言值产生的方法；语法规则定义了术语集中每一个语言值的隶属函数。

2. 模糊集合的运算与隶属函数的参数化

与精确集合的运算相对应，模糊集合也有并、交、补等相似的基本运算。

图 2.31 术语集 t (温度)典型的隶属函数

① 包含(或子集): 仅且仅当所有 $x, \mu_A(x) \leq \mu_B(x)$, 则模糊集合 A 被包含在模糊集合 B 中。记作: $A \subseteq B \Leftrightarrow \mu_A(x) \leq \mu_B(x)$ 。

② 并(析取): 两个模糊集合 A 和 B 的“并”为模糊集合 C , 写成 $C = A \cup B$, 其隶属函数与 A 和 B 的关系为

$$\mu_C(x) = \max(\mu_A(x), \mu_B(x)) = \mu_A(x) \vee \mu_B(x) \quad (2.48)$$

③ 交(合取): 两个模糊集合 A 和 B 的“交”是模糊集合 C , 写成 $C = A \cap B$, 其隶属函数与 A 和 B 的关系为

$$\mu_C(x) = \min(\mu_A(x), \mu_B(x)) = \mu_A(x) \wedge \mu_B(x) \quad (2.49)$$

④ 补(负): 模糊集合 A 的补用 \bar{A} ($-A$ 、非 A) 表示, 定义为

$$\mu_{\bar{A}}(x) = 1 - \mu_A(x) \quad (2.50)$$

图 2.32 给出了交、并、补三种运算的运算过程。图(a)说明两个模糊集合 A 和 B ; 图(b)是 A 的补; 图(c)是 A 和 B 的并; 图(d)是 A 和 B 的交。

实际应用中常用的一维隶属函数有三角形、梯形、高斯型和钟形等, 如图 2.33 所示, 它们都可以用一组参数来描述。其中高斯型和钟形隶属函数的平滑性和概念精确, 在模糊集合的运算中应用日益广泛。

一般化的钟形隶属函数为

$$\text{bell}(x; a, b, c) = \frac{1}{1 + \left| \frac{x-c}{a} \right|^{2b}} \quad (2.51)$$

适当地选择参数集 $\{a, b, c\}$ 可以获得所需的钟形隶属函数。调节 c 和 a , 可以改变隶属函数的中心和宽度; 用 b 来控制隶属函数交叉点上的斜率。图 2.34 表示了参数改变对隶属函数的影响。

有时需要利用具有两个输入的隶属函数, 每个输入处于不同的论域, 这种隶属函数称为二维隶属函数, 而普通的隶属函数称为一维隶属函数。二维隶属函数通常有两种类型: 复合式和非复合式。如果一个二维隶属函数可以表示成两个一维隶属函数的解析式, 则它是复合式的, 否则为非复合式的。复合式的二维隶属函数可以看成是由连接符 AND 或 OR 连接起来的两

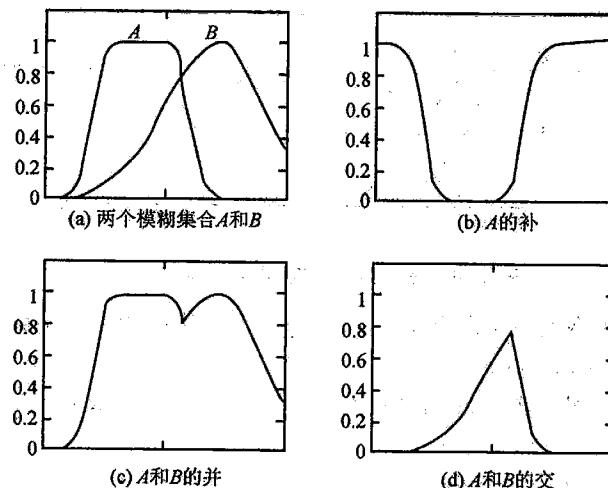


图 2.32 模糊集合的运算

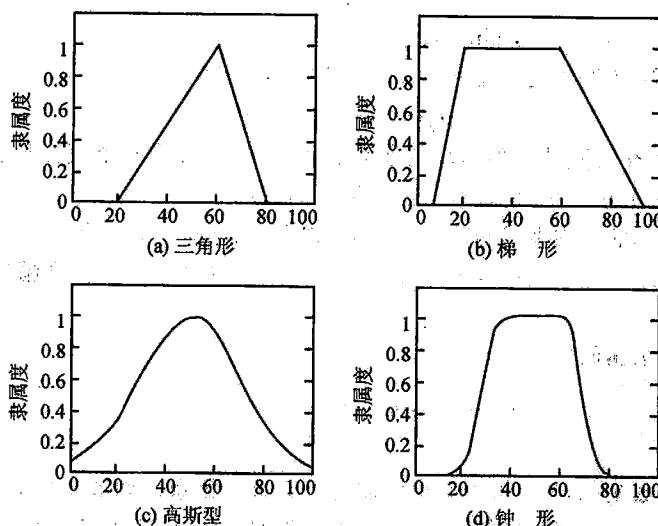


图 2.33 四种参数化隶属函数

个语句。根据这个条件,二维隶属函数定义为其两个组成部分 AND 或 OR 的集结。对模糊集合经典的 AND 和 OR 运算是 \max 和 \min 。

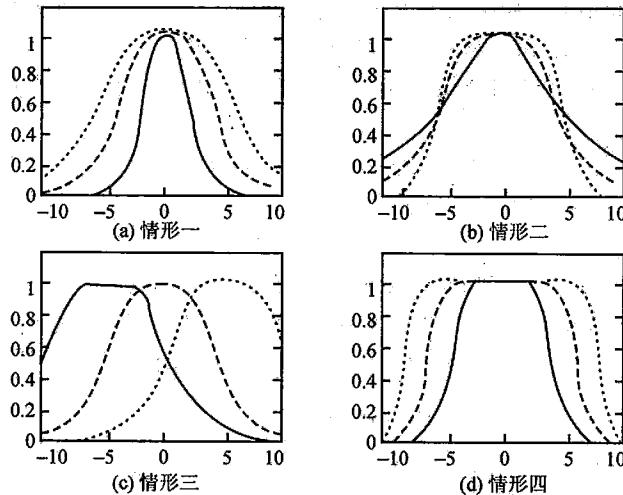


图 2.34 钟形隶属函数参数改变对其形状的影响

3. 模糊关系

精确关系表示两个或两个以上的集合元素之间的关联、交互或互联是否存在；而模糊关系则表示两个以上集合元素之间关联、交互或互联存在或不存在的程度。二元模糊关系是 $X \times Y$ 中的模糊集合，它把 $X \times Y$ 中的元素映射成0和1之间的隶属度。单元模糊关系是具有一定维隶属函数的模糊集合；而二维模糊关系是具有二维隶属函数的模糊集合。模糊关系的应用包括模糊控制和决策。

令 X 和 Y 是两个论域，则模糊关系 $R(X, Y)$ 是 $X \times Y$ 空间中的模糊集合，可表示为

$$R(X, Y) = \{((x, y), \mu_R(x, y)) \mid (x, y) \in X \times Y\} \quad (2.52)$$

式中： \times 为直接积(笛卡儿积)算符。该式称做 $X \times Y$ 中的二元模糊关系。 $\mu_R(x, y)$ 实际上就是一个二维的隶属函数。

例如，如果有 $X=Y=R^+$ (正实轴)， R ="y比x大得多"，则模糊关系的隶属函数 R 可以定义为

$$\mu_R(x, y) = \begin{cases} (y - x)/(x + y + 2), & y > x \\ 0, & y \leqslant x \end{cases}$$

假如 $X=(3, 4, 5)$, $Y=(3, 4, 5, 6, 7)$ ，则可将模糊关系 $R(X, Y)$ 方便地表示成关系矩阵：

$$R = \begin{bmatrix} 0 & 0.111 & 0.200 & 0.273 & 0.333 \\ 0 & 0 & 0.0091 & 0.167 & 0.231 \\ 0 & 0 & 0 & 0.077 & 0.143 \end{bmatrix}$$

矩阵中第 i 行 j 列的元素等于 X 的第 i 元素和 Y 的第 j 元素之间的隶属度。

在不同乘积空间中的模糊关系可以通过复合运算联合起来。模糊关系的复合运算有很多种,但是最常用的是 max-min 和 max-乘积两种复合运算。

设 $R(X, Y)$ 和 $S(Y, Z)$ 分别为定义在 $X \times Y$ 和 $Y \times Z$ 空间中的两个模糊关系,则其复合运算为

$$P(X, Z) = R(X, Y) \cdot S(Y, Z) \quad (2.53)$$

或简写为

$$P = R \cdot S \quad (2.54)$$

现在举例说明 max-min 和 max-乘积两种复合运算的方法。

假定: R = “ x 与 y 有关”, S = “ y 与 z 有关”分别是定义在 $X \times Y$ 和 $Y \times Z$ 中的两个模糊关系,式中 $X = \{1, 2, 3\}$, $Y = \{\alpha, \beta, \gamma, \delta\}$, $Z = \{a, b\}$, R 和 S 可以表示成以下的关系矩阵:

$$R = \begin{bmatrix} 0.1 & 0.3 & 0.5 & 0.7 \\ 0.4 & 0.2 & 0.8 & 0.9 \\ 0.6 & 0.8 & 0.3 & 0.2 \end{bmatrix}, \quad S = \begin{bmatrix} 0.9 & 0.1 \\ 0.2 & 0.3 \\ 0.5 & 0.6 \\ 0.7 & 0.2 \end{bmatrix}$$

根据 R 和 S ,求出 $R \cdot S$,即推导“ x 与 z 有关系”的模糊关系。为了简单起见,假定只对 $2 \in X$ 和 $a \in Z$ 之间的关系程度感兴趣。

若采用 max-min 复合,则

$$\begin{aligned} \mu_{R \cdot S}(2, a) &= \max(0.4 \wedge 0.9, 0.2 \wedge 0.2, 0.8 \wedge 0.5, 0.9 \wedge 0.7) = \\ &= \max(0.4, 0.2, 0.5, 0.7) = 0.7 \end{aligned}$$

若采用 max-乘积复合,则

$$\begin{aligned} \mu_{R \cdot S}(2, a) &= \max(0.4 \wedge 0.9, 0.2 \wedge 0.2, 0.8 \wedge 0.5, 0.9 \wedge 0.7) = \\ &= \max(0.36, 0.04, 0.40, 0.63) = 0.63 \end{aligned}$$

图 2.35 表示两个模糊关系的复合,图中 X 中的元素 2 和 Z 中的元素 a 是通过连接两个元素的四条可能路径(实线)而建立的。2 与 a 的关系程度是四条路径强度的最大者,而每条路径的强度是其组成连接强度的最小者(或乘积)。

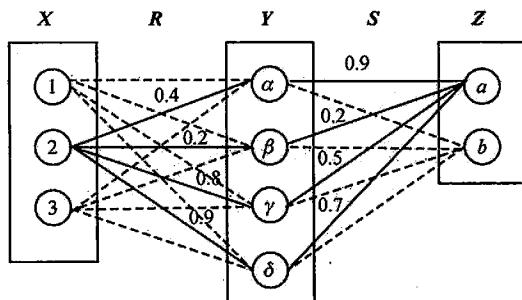


图 2.35 模糊关系的复合



2.9.2 模糊规则与模糊推理

在模糊推理中,经常遇到模糊 if-then 规则,它们有以下的形式:

$$\text{if } x \text{ 是 } A, \text{ then } y \text{ 是 } B. \quad (2.55)$$

式中: A 和 B 是语言值,分别由 X 和 Y 中模糊集合确定,“ x 是 A ”通常叫做前提或前件;“ y 是 B ”叫做结论或后件。

式(2.55)可简化为 $A \rightarrow B$ 。实质上,该表达式描述了变量 x 和 y 之间的关系。因此可以把 if-then 规则定义为乘积空间中的二元模糊关系。一般来说,可以用两种方法来解释模糊隐含或模糊规则 $A \rightarrow B$ 。

① $A \rightarrow B$ 解释为 A 与 B 相关,那么可以得到四种不同的模糊关系:

$$R_m = A \times B = \int_{X \times Y} \mu_A(x) \wedge \mu_B(y)/(x, y) \quad \text{或} \quad \mu_{A \rightarrow B}(x, y) = \min \{\mu_A(x), \mu_B(y)\} \quad (2.56)$$

$$R_p = A \times B = \int_{X \times Y} \mu_A(x) \mu_B(y)/(x, y) \quad \text{或} \quad \mu_{A \rightarrow B}(x, y) = \mu_A(x) * \mu_B(y) \quad (2.57)$$

$$R_{bp} = A \times B = \int_{X \times Y} \mu_A(x) \otimes \mu_B(y)/(x, y) = \int_{A \times B} 0 \vee (\mu_A(x) + \mu_B(y) - 1)/(x, y) \quad \text{或} \quad \mu_{A \rightarrow B}(x, y) = \max \{0, \mu_A(x) + \mu_B(y) - 1\} \quad (2.58)$$

$$R_{dp} = A \times B = \int_{X \times Y} \mu_A(x) \otimes \mu_B(y)/(x, y)$$

$$\text{或} \quad \mu_{A \rightarrow B}(x, y) = \mu_A(x) \otimes \mu_B(y) = \begin{cases} \mu_A(x), & \text{如 } \mu_B(y) = 1 \\ \mu_B(y), & \text{如 } \mu_A(x) = 1 \\ 0, & \text{其他} \end{cases} \quad (2.59)$$

② $A \rightarrow B$ 解释成“ A 传递给 B ”,那么又可得到四种不同的模糊隐含函数:

$$R_a = \neg A \cup B = \int_{X \times Y} 1 \wedge (1 - \mu_A(x) + \mu_B(y))/(x, y) \quad \text{或} \quad \mu_{A \rightarrow B}(x, y) = \min (1, 1 - \mu_A(x) + \mu_B(y)) \quad (2.60)$$

$$R_{nm} = \neg A \cup B = \int_{X \times Y} (1 - \mu_A(x)) \vee (\mu_A(x) + \mu_B(y))/(x, y)$$

$$\text{或} \quad \mu_{A \rightarrow B}(x, y) = \max \{(1 - \mu_A(x)), \min [\mu_A(x), \mu_B(y)]\} \quad (2.61)$$

$$R_s = \neg A \cup B = \int_{X \times Y} (1 - \mu_A(x)) \vee (\mu_B(y))/(x, y)$$

$$\text{或} \quad \mu_{A \rightarrow B}(x, y) = \max (1 - \mu_A(x), \mu_B(y)) \quad (2.62)$$

$$R_d = \neg A \cup B = \int_{X \times Y} (\mu_A(x)) \tilde{\wedge} (\mu_B(y))/(x, y)$$

$$\text{或} \quad \mu_{A \rightarrow B}(x, y) = \mu_A(x) \tilde{\wedge} \mu_B(y) = \begin{cases} 1, & \mu_A(x) \leq \mu_B(y) \\ \mu_B(y)/\mu_A(x), & \mu_A(x) > \mu_B(y) \end{cases} \quad (2.63)$$

而模糊推理或模糊近似推理可以分如下四步进行：

① 计算兼容度。把已知事实与模糊规则的前提进行比较，求出相对于每一前提隶属函数的兼容度。

② 求激励强度。用模糊和(AND)或(OR)算子，把相对于前提隶属函数的兼容度结合起来，形成激励强度，它说明规则前提部分满足的程度。

③ 求定性(演绎)结果隶属函数。把激励强度施加于规则的结果隶属函数，以产生一个定性结果隶属函数(该隶属函数代表了激励强度如何传递并用在模糊隐含语句中)。

④ 求总输出隶属函数。把所有定性结果隶属函数集结，获得总的输出隶属函数。

2.9.3 多级映射原理

采用自然语言描述被测量，通常只能使用较少的概念表示(例如用九级语言变量表示温度：非常非常热、非常热、热、有点热、不冷不热、有点冷、冷、非常冷、非常非常冷)。与数值测量无限可分比较，用自然语言描述的细致程度不够。如温度测量范围为 $0 \sim 1000^\circ\text{C}$ ，采用九级概念划分表示，则可分辨的区间温度范围是 $0 \sim 125^\circ\text{C}$ ；此外，在有些情况下，传感器需要同时能够实现符号输出和进行非线性校正后直读的数值测量输出，因此需要采用多级映射原理去描述被测量。

这里介绍针对多级映射的信息传输、变换过程。图 2.36 给出了多级映射原理的模糊传感器信息传输链示意图。由图 2.36 可知，被测温度 t 经温度敏感元件转换成电量 E_Q (这里 E_Q 可以是电压、电流、电阻或其他电学量)。 E_Q 经第一次映射变成第一级符号表示结果 t_{S1} 。如果第一级映射描述细致度不够，可以第二次映射，得到第二级映射的符号表示结果 t_{S2} 。如果需要，可以进行第三级映射，得到第三级映射符号表示结果 t_{S3} 。重复上述过程，直到得到合适的符号表示结果为止。

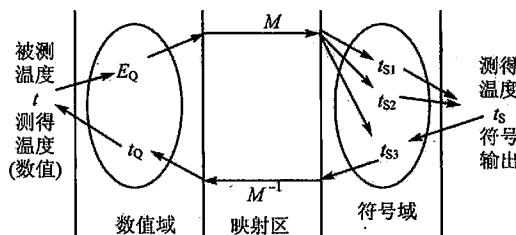


图 2.36 模糊温度传感器多级映射信息传输链



假如某个模糊传感器采用三级映射表示结果,那么符号测量结果 t_s 可以表示为

$$t_s = \{t_{s1}, t_{s2}, t_{s3}\} \quad (2.64)$$

为了校正传感器的非线性,并获得数值测量结果,可将模糊传感器语言变量的隶属函数上的特征测量点与非线性校正的数据校正点统一起来,然后将测得的被测温度符号输出 t_s 进行 M^{-1} 逆映射,可以得到经过非线性校正后的数值测量结果 t_q 。

图 2.37 给出了采用多级映射原理进行传感器非线性校正和符号过程示意图。如果图 2.37 中的数据源自 K 型热电偶,测温范围为 $-200 \sim 1000^\circ\text{C}$,则被测温度为 380°C 时,对应的热电偶输出热电势 E_q 为 15.552 mV 。

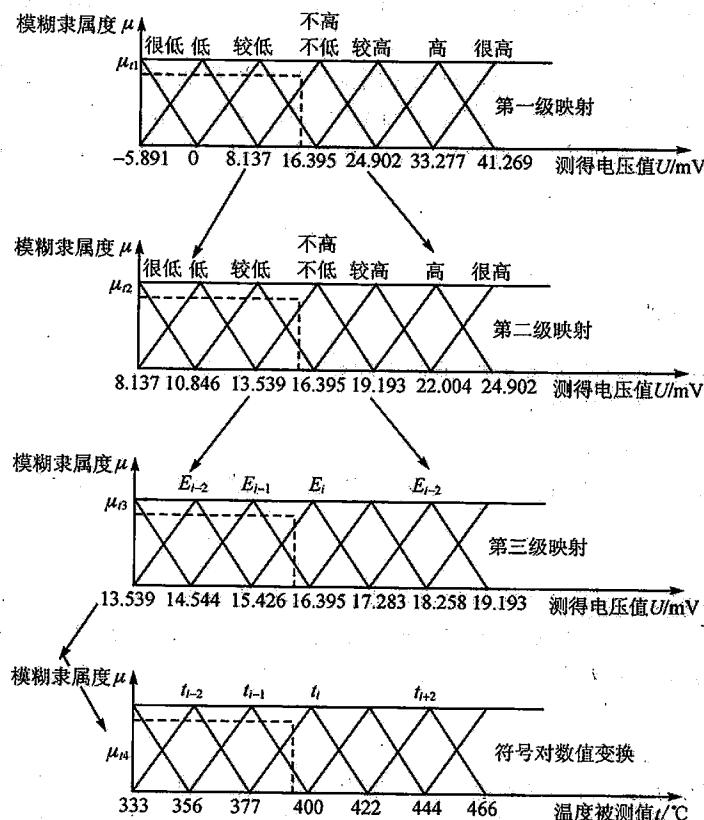


图 2.37 多级映射在模糊温度传感器中应用的示意图

可见,第一级符号表示用七个语言变量。不失一般性,语言变量采用很高、高、较高、不高不低、较低、低、很低表示。由于用语言变量描述测量结果的分辨率是非端点语言变量覆盖范

围的 1/2,这样,第一级映射每个语言变量符号表示的分辨范围是 200 ℃。显然用一级映射符号表示描述细致度不够。为此,采用多级映射方法。第二级映射是将第一级映射中的一个语言变量(非端点处变量)再划分为七个子语言变量(见图 2.37 中的第二级映射表示图)。这样第二级映射子语言变量温度测量分辨范围是 33.3 ℃。如果采用第三级映射,则可以得到 22.1 ℃的子子语言变量温度范围。对 K 型热电偶,测量范围为 -200~1 000 ℃时,当采用三级映射时,可使符号测量结果表示的分辨率达到 11.05 ℃。

若用三级映射可得到合适的描述,最后则要进行论域变换,实现符号对数值的变换。论域变换的基本方法如下:

- ① 以测得的电量为变量来表述隶属函数曲线上的特征测量点,用该点温度校正数据代替。
- ② 将测得电量所属的语言变量按对应原则转换成以测得温度表征的语言变量。
- ③ 对任一给定 E_Q 测得电量,在两个不同论域具有相同隶属函数数值。
- ④ 采用插值法(或其他方法)得到被测温度值测量结果。测得温度值可采用下式计算,即

$$t_Q = (1 - \mu_i)(t_i - t_{i-1}) + t_{i+1} \quad (2.65)$$

$$\mu_i = \mu_{i3} = 1 - \frac{E_Q - E_{i-1}}{E_i - E_{i-1}} \quad (2.66)$$

2.9.4 模糊传感器

模糊传感器是以数值测量为基础,能产生和处理与其相关的符号信息,实现被测对象信息自然语言符号化表示的智能传感器。

1. 模糊传感器的基本功能

模糊传感器作为一种智能传感器,应该具有智能传感器的基本功能,即学习、推理联想、感知和通信功能。

(1) 学习功能

模糊传感器特殊和重要的功能是学习功能。人类知识集成的实现,测量结果的高级逻辑表达等都是通过学习功能完成的。能够根据测量任务的要求学习相关知识是模糊传感器与传统传感器的重要差别。模糊传感器的学习功能是通过有导师学习算法和无导师自学习算法实现的。

(2) 推理联想功能

模糊传感器可分为一维传感器和多维传感器。当一维传感器接收外界刺激时,可以通过训练时的记忆联想得到符号化测量结果。多维传感器接受多个外部测量刺激时,可通过人类知识的集成进行推理,实现时空信息整合与多传感器信息融合,以及复合概念的符号化表示等。模糊传感器的推理联想功能需要通过推理机构和知识库来实现。

(3) 感知功能

模糊传感器与一般传感器一样可以感知由传感元件确定的被测量,但根本区别在于其不



仅可以输出数值量,且能够输出语言符号量。因此模糊传感器必须具有数值/符号转换器。

(4) 通信功能

传感器通常作为一个大系统中的子系统进行工作,因此模糊传感器应该能够与上级系统进行信息交换,所以通信功能也是模糊传感器的基本功能。

2. 模糊传感器的基本结构

(1) 逻辑结构与物理结构

图 2.38 是模糊传感器的简化逻辑结构框图。所谓模糊传感器的逻辑结构,就是其在逻辑上要完成的功能。一般来讲,模糊传感器逻辑上可以分为转换部分和符号处理与通信部分。从功能上看,有信号调理与转换层、数值/符号转换层、符号处理层、有指导学习层和通信层。这些功能有机地集成在一起,完成数值/符号转换功能。

(2) 多维模糊传感器结构

图 2.39 是多维模糊传感器硬件结构框图。由图可知,由敏感元件、信息调理以及 A/D 组成的基础测量单元完成传统的传感测量任务。由数值预处理、数值/符号转换器、概念合成器、数据库、知识库构成的符号生成与处理单元实现模糊传感器核心工作——数值/符号转换。单一被测量的一维情况只是多维模糊传感器的一个特殊情况。

数值符号的模糊化必须有知识库中专家知识的支持,知识库中存放的知识主要有符号量及其隶属函数、合成概念的推理合成规则、被测对象的背景知识以及测量系统的有关情况等,具体体现在论域上语言符号概念隶属函数的产生上。知识库中经验隶属函数可以通过模糊统计法、选择比较法等方法产生;而对于不同的测量对象具体实现时的输入信息模糊化过程,则可以在知识库中经验的指导下,通过语义关系自动产生,也可以在导师指导下通过学习和训练来产生并修正隶属函数曲线,这种方法通过调整符号量的隶属函数,可使模糊传感器适合不同的测量目的。

通过训练若干点,模糊传感器可以自动产生一个概念序列放在数据库中,当有数值测量结果送入时,按最大隶属度原则选择一个符号量输出,即实现了数值/符号转换。

学习系统主要是为了调整概念而设计的,当测量系统用于不同的测量目的时,或不同用户有不同要求时,用户通过学习系统来调整符号量的隶属关系,达到自己的测量目的,调整好的符号量放入知识库中。

推理器主要用来实现复合概念的生成。由于复合概念是建立在经验知识基础上的,故测量结果不能通过公式计算,必须利用知识库中的经验知识通过模糊推理来实现。为实现有导师学习,还必须具有输入设备,用户通过它实现对传感器的控制和调整概念。

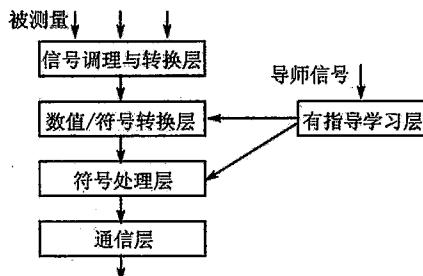


图 2.38 模糊传感器的简化逻辑结构框图

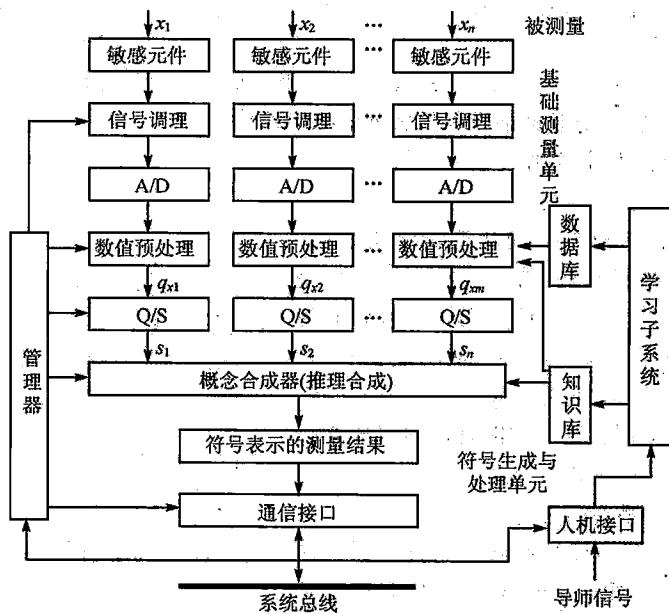


图 2.39 多维模糊传感器硬件结构框图

通信接口实现模糊传感器与上级系统之间的信息交换,把测量结果(数值量与符号量)输出到系统总线,并从系统总线上接收上级系统的命令。人机接口是模糊传感器与操作者进行信息交流的通道。

管理器主要是对测量系统自身进行管理,接收上级系统的命令,开启/关闭测量系统,调节放大器的放大倍数,根据上级系统的要求决定输出量的类型(数值量或者语言符号量)等。

3. 有导师学习结构的实现

具有学习功能使得模糊传感器的智能化水平提高。图 2.40 是具有学习功能的模糊传感器示意图。可以看出,有导师学习法的基本原理是基于比较导师和传感器对同一被测量 x 的定性描述的差别进行学习的。

对同一被测量 x ,如果导师的语言描述符号为 $l(x)$,模糊传感器结构的描述为 $l'(x)$,则 $l(x)$ 与 $l'(x)$ 进行比较,结果如下:

如果 $l(x) \geq l'(x)$,则 $e=+$,那么 $\mu=\text{增加}$;

如果 $l(x) \leq l'(x)$,则 $e=-$,那么 $\mu=\text{减少}$;

如果 $l(x)=l'(x)$,则 $e=0$,那么 $\mu=\text{保持}$ 。

e 是误差, μ 为控制量,被控量为概念的隶属函数,控制行为是“增加”、“减少”和“保持”。“增加”是指隶属曲线向数值小的方向平移或扩展,“减少”是指隶属函数曲线向数值大的方向

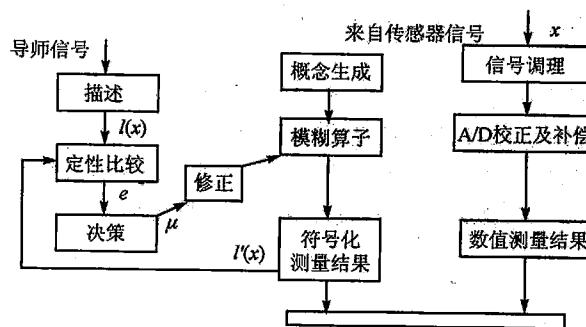


图 2.40 具有学习功能的模糊传感器示意图

平移或扩展，“保持”是指隶属函数保持不变。

基于上述有导师学习的基本原理，实现模糊传感器有导师学习功能的结构，关键在于导师信号的获取。

导师信号可以分为经验知识信号和直观感觉信号两类。所谓经验知识信号是指在某些测量任务下，经过长期实践经验总结，导师对被测量的描述符号隶属度为 1 的相应数值已经可以事先确定；而直观感觉信号是指导师对被测量状态的直接描述符号，其间不经过数值描述符号这一过程。可以采用人机接口的方式将上述这两种导师信号输入至模糊传感器内部，以便进行学习。

导师指导信号的获取步骤如下：

- ① 确定测量范围的上下限；
- ② 选择论域上描述被测符号的个数；
- ③ 选择被训练概念；
- ④ 输入该被训练概念对应隶属度为 1 的特征测量点；
- ⑤ 输入该被训练概念对应的关键点（例如，隶属函数为 0.5）；
- ⑥ 通过相应训练算法产生概念的隶属函数。

模糊传感器的概念生成能否产生适合测量目的的准确语言符号量，关系到测量的准确程度。与模糊控制类比，它相当于模糊控制中的模糊化，但很多方面又有所不同，因此对其转换基础和方法的研究有着重要的理论价值和实际意义。

2.9.5 模糊传感器应用实例

下面通过模糊温度传感器的例子来说明模糊传感器的具体工作过程。该模糊传感器以热敏电阻为敏感元件，以单片机为核心构成硬件平台。

(1) 硬件结构

模糊温度传感器由热敏元件、信号调理单元、A/D 转换器及典型的单片机系统组成,如图 2.41 所示。

图 2.41 中,CPU 对应选定的微处理器, RAM 和 EPROM 分别为单片机系统的数据和程序存储器。单片机系统采用 8279 作为键盘、显示器和导师指导信号接口。模糊温度传感器是以传统数值测量为基础的,而输出则有数值法和符号两种。由于符号输出具有较大的冗余,换句话说,可以采用较低测量准确度的器件,为此敏感元件采用热敏电阻。图 2.42 是模糊温度传感器接口电路图。

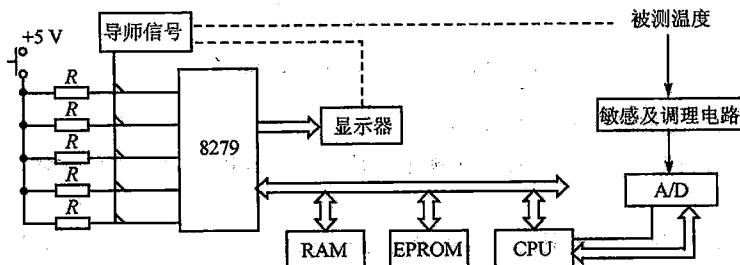


图 2.41 模糊温度传感器原理图

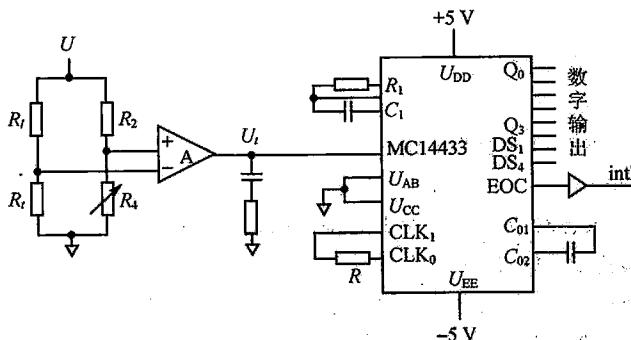


图 2.42 模糊温度传感器接口电路图

(2) 数值/符号变换原理

模糊温度传感器被测温度范围是 $0 \sim 120^{\circ}\text{C}$,按照很热、热、较热、不冷不热、较冷、冷和很冷七个语言概念划分基础概念。第一、二级子概念也按照线性划分原理进行划分。

在传统的数值测量已经完成,即在 A/D 转换结束后进行传感器的非线性校正。

假定被测温度为 57.5°C ,按照线性划分七个语言概念, $U_{57.5^{\circ}\text{C}}$ 落在不热不冷、较冷两个语



言概念的交集。为提高语言变量描述细节的程度,采用多级映射进行第一级和第二级子概念的映射。由图 2.43 可知,第一级子概念映射 $U_{57.5}^{\circ}\text{C}$ 落在不高不低、较低两个语言概念的交集,而在第二级子概念映射 $U_{57.5}^{\circ}\text{C}$ 落在较低、低两个语言概念的交集。用符号表示,有

$$U_{57.5}^{\circ}\text{C} = \{\text{不热不冷 / 较冷, 不高不低 / 较低, 较低 / 低}\}$$

如果利用最大隶属函数判别准则,还可以表示为

$$U_{57.5}^{\circ}\text{C} = \{\text{不热不冷, 不高不低, 较低}\}$$

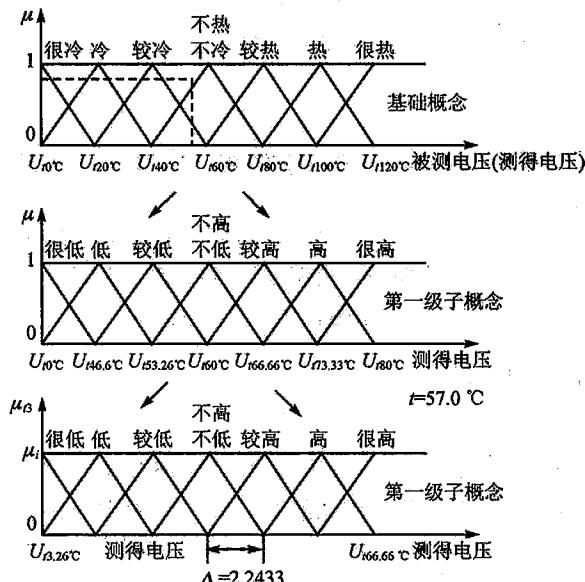


图 2.43 数值/符号转换原理示意图

(3) 有导师学习

事实上,用自然语言表示的语言概念很少是与被测物理量成线性关系的。因此应将人类的知识与经验集成到模糊温度传感器,采用有导师指导学习方法完成。众所周知,游泳池内的水温“热”的含义与炼钢炉中温度“热”的含义显然是不同的。

这里采用三角隶属函数的简单情况实现实现有导师学习。

首先,按照线性划分的方法对被测量整个范围进行划分,得到被测量与符号表示有线性关系的表示结果。

然后,改变被测量状态,同时将模糊传感器切换到训练状态。这时,如果显示的结果与导师的结果不一致,则要通过调整来选择出此温度下对应的符号结果。图 2.44 是一个三角形隶属函数学习后的函数曲线。

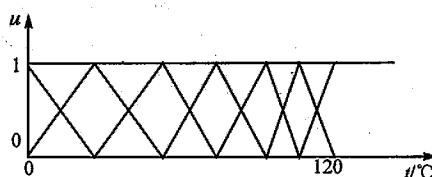


图 2.44 三角形隶属函数学习后的函数曲线

总的来说,模糊温度传感器的工作过程可以这样描述:首先建立不同语言值下的模糊隶属函数,根据情况可以将语言值细分为多级,具体分几级则根据需求而定;建立完隶属函数后,则根据实际温度传感器的输出电压信号进行多级映射,给出符号输出表达结果。如果需要数值输出表示结果,则要进行逆映射,在前述隶属函数的基础上,反推出温度高低。可以看出,整个工作过程中最关键的环节就是隶属函数的建立,这也是影响模糊传感器性能的核心环节。

2.10 人工神经网络技术及其应用

神经网络技术的研究源于 20 世纪 40 年代,期间经历了几次高潮和低谷;最近的一次研究高潮源于 20 世纪 80 年代,由于在神经网络模型、学习算法和应用等方面都取得了令人鼓舞的众多成果,因此,其在智能传感器中的应用也得到了广泛关注。

2.10.1 神经网络的基本知识

1. 人工神经网络模型

Kononen 曾给出神经网络的定义。他认为,神经网络是由一些简单的(通常为自适应的)单元及其层次组织的大规模并行连接构造的网络,并按照生物神经系统的同样方式处理真实世界的客观事物,从而反映了人脑功能的若干特征;但它并非神经系统的真实描写,而是它的简化、抽象和模拟。

人工神经网络是由人工建立的以有向图为拓扑结构的动态系统,它通过对连续或断续的输入作状态响应,并进行信息的处理,或者说将大量的基本神经元,通过一定的拓扑结构组织在一起,构成群体并行分布式处理的计算结构。人的思维有逻辑性和直观性两种不同的基本方式。逻辑性的思维是指根据逻辑规则进行推理的过程,它先将信息化成概念,并用符号表示;然后,根据符号运算的方式,按串行模式进行逻辑推理。这一过程可以写成串行的指令,由计算机来执行。而直观性思维是将分布式存储的信息综合起来,并在瞬间产生想法或解决问题的办法。这种思维方式与逻辑性思维的不同之处有两点:

- ① 信息是通过神经元上的兴奋模式以分布式存储在网络上;
- ② 信息处理是通过神经元之间同时相互作用的动态过程来完成的。

人工神经网络就是模拟人思维的第二种方式,即直观性思维,这是一个非线性动力学系统,其特色在于信息的分布式存储和并行协同处理。

神经元是神经网络中的基本单元。图 2.45 给出了一个简单的人工神经元模型。



假设 x_1, x_2, \dots, x_n 表示某(第 i 个)神经元的 n 个输入; W_{ij} 表示第 j 个神经元与第 i 个神经元的连结强度, 其值称为权值; A_i 表示第 i 个神经元的输入总和, 称为激活函数; y_i 表示第 i 个神经元的输出; Q_i 表示第 i 个神经元的阈值。因此, 神经元的输出可以描述为

$$y_i = f(A_i) \quad (2.67)$$

$$A_i = \sum_{j=1}^n W_{ij} x_j - Q_i \quad (2.68)$$

式中: $f(A_i)$ 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。常用的作用函数有三种: 阈值型、S型和分段线性型(伪线性型)。

(1) 阈值型神经元

阈值型神经元是一种最简单的神经元, 由美国心理学家 Mc. Culloch 和数学家 Pitts 共同提出, 因此, 通常称为 M-P 模型。

M-P 模型神经元是二值型神经元, 其输出状态取值为 1 或 0, 分别代表神经元的兴奋状态和抑制状态。其数学表达式为

$$y_i = f(A_i) = \begin{cases} 1 & (A_i > 0) \\ 0 & (A_i \leq 0) \end{cases} \quad (2.69)$$

对于 M-P 模型神经元, 权值 W_{ij} 可在 $(-1, 1)$ 区间内连续取值。取负值表示抑制两神经元间的连接强度, 取正值表示加强。

(2) S型神经元

这是常用的一种连续型神经元模型, 输出值是在某一范围内连续取值的。输入-输出特性多采用指数函数表示, 用数学公式表示为

$$y_i = f(A_i) = \frac{1}{1 + e^{A_i}} \quad (2.70)$$

S型作用函数反映了神经元的非线性输入-输出特性。

(3) 分段线性型

神经元的输入-输出特性满足一定的区间线性关系, 其输出可以表示为

$$y_i = \begin{cases} 0 & (A_i \leq 0) \\ CA_i & (0 \leq A_i \leq A_C) \\ 1 & (A_C < A_i) \end{cases} \quad (2.71)$$

式中: C, A_i 表示常量。

如果将大量功能简单的基本神经元通过一定的拓扑结构组织起来, 构成群体并行分布式的计算结构, 那么这种结构就是人工神经网络结构。

根据神经元之间连接的拓扑结构上的不同, 可将神经网络分为两大类: 分层网络、相互连

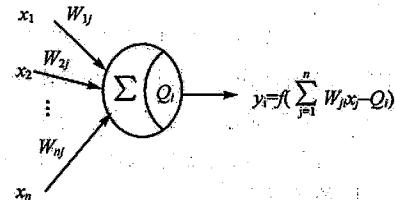


图 2.45 简单的人工神经元模型

$$U_1(P_m) + U_2(P_m) + U_3(P_m) + U_4(P_m) = 196 \text{ mV}$$

将上述数据代入式(2-138)中, 可得总特性 $P-U$ 的线性度 δ_L 为

$$\delta_L = \frac{\sum_{i=1}^4 \delta_i U_i(P_m)}{\sum_{i=1}^4 U_i(P_m)} = 2.3 \times 10^{-4}$$

由计算可见, 非线性大大改善, 非线性误差减小近两个数量级。

实现上述非线性补偿的典型结构为一具有双岛结构的方形硅膜片, 如图 2-44 所示。两对压敏电阻 R_{A1} 、 R_{A2} 与 R_{B1} 、 R_{B2} 分置于边槽和中央沟槽表面。当硅膜片受压后应力高度集中于压敏电阻所在的沟槽表面, 同一沟槽内应力 σ_{AX} 、 σ_{BX} 变化平缓, 如图 2-44(c)所示。边槽和中央沟槽内应力符号相反, 两对压敏电阻 R_A 、 R_B 分别受横向拉伸应力和横向压缩应力。若两对压敏电阻接成电桥则具有如图 2-43 所示的非线性特性。其中曲线 1、2 是压敏电阻 R_{A1} 、 R_{A2} 的特性, 曲线 3、4 则是压敏电阻 R_{B1} 、 R_{B2} 的特性。因为桥臂电阻 R_{A1} 、 R_{A2} 、 R_{B1} 、 R_{B2} 对桥路输出的贡献 $U_{A1}(P)$ 、 $U_{A2}(P)$ 、 $U_{B1}(P)$ 、 $U_{B2}(P)$ 与应力 $\sigma_{A1}=\sigma_{A2}=\sigma_{AX}$, $\sigma_{B1}=\sigma_{B2}=\sigma_{BX}$ 成正比, 即

$$\frac{U_A(P)}{U_B(P)} = -\frac{\sigma_{AX}}{\sigma_{BX}}$$

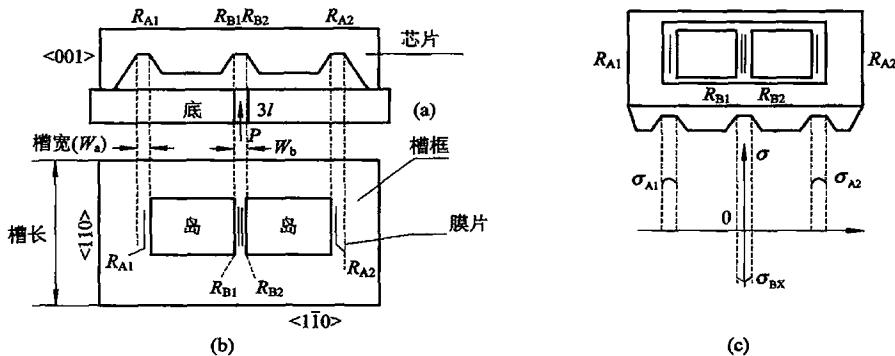


图 2-44 具有非线性内补偿作用的压力传感器的结构示意图

(a) 横断面; (b) 顶视图; (c) 槽中的应力

而沟槽应力 σ_{AX} 、 σ_{BX} 之比可以通过调整沟槽宽度 W_A 、 W_B 之比 n 来调节, 从而实现式(2-138)中的分子项为零, 即

$$\sum_{i=1}^4 \delta_i U_i(P_m) = 2\delta_A U_A(P) + 2\delta_B U_B(P) = 0$$

则有

$$\frac{U_A(P)}{U_B(P)} = -\frac{\delta_B}{\delta_A}$$

达到上述总的非线性误差 $\delta=0$ 的最佳补偿条件是 $n=W_A/W_B=2.48$ 。我国产品 PT14 系列集成压力传感器中就采用这种双岛结构硅膜片, 其非线性、迟滞均达到 10^{-4} 数量级。

这种补偿法也算作是一种结构补偿法。前面提到的机械结构对称、电结构对称(惠斯

顿电桥)、使被测物理量即目标参数产生差动信号以及其它干扰参量产生共模信号等, 都属于结构补偿范围。

2. 拼凑补偿法

1) 温度自补偿

几乎所有的未经补偿的初级经典传感器都受工作温度的影响。如果这种传感器不是测量温度的, 那么它对温度的敏感程度就是交叉灵敏度。这样, 温度的变化除会引起零位漂移外, 还会产生增益的变化, 从而使该传感器对其基本目标测量参量的灵敏度产生温度漂移。所以对传感器进行温度补偿, 一直是传感器工作者关注的重点。压阻式传感器具有许多优点, 但它的主要缺点是易受温度的影响, 因此围绕对它进行温度补偿做了大量工作。其中一些补偿网络、补偿电路已与敏感元件一起集成实现了一体化, 构成了具有温度自动补偿功能的集成化的初级智能传感器系统。拼凑的方式不胜枚举, 只需寻找某个电路元件也敏感温度, 其作用与传感器的温度漂移量方向相反, 而数值要拼凑得相等就能起抵消补偿作用。这里我们仅就利用电源进行补偿为例说明“拼凑”补偿的基本思想。当然“拼凑”补偿法也在很多其它类型传感器中广泛应用。

由公式(2-113)、(2-114)已给出电压源供电、恒流源供电时, 电桥输出电压 U 与桥臂阻值的关系, 而桥臂阻值是由被测物理量 P ——目标参量决定的。所以电桥输出电压 U 是目标参量 P 的函数。我们将式(2-113)、式(2-114)重写如下:

电压源供电:

$$U = \frac{R_1 R_4 - R_2 R_3}{(R_1 + R_2)(R_3 + R_4)} E = Ef(P)$$

恒流源供电:

$$U = \frac{R_1 R_4 - R_2 R_3}{R_1 + R_2 + R_3 + R_4} I = I\Phi(P)$$

如果输出 U 还受非目标参量——温度变化 ΔT 的影响, 则上两式为

电压源供电:

$$U = Ef(P)[1 + \alpha \Delta T + \beta \Delta T^2 + \dots] \quad (2-139)$$

恒流源供电:

$$U = I\Phi(P)[1 + \alpha' \Delta T + \beta' \Delta T^2 + \dots] \quad (2-140)$$

由上两式可见, 当目标参量恒定不变时, $P=$ 常量, 电桥输出 U 随温度变化量 ΔT 而变, 其温度交叉灵敏度为 α (电压源供电)、 α' (恒流源供电)。设想两式中的电源 E 、 I 也将随温度的变化而变化, 但其温度灵敏度 α_E 、 α_I 与 α 、 α' 数值相等、符号相反, 那么电桥输出 U 的表达式则为

电压源供电:

$$\begin{aligned} U &= f(P)E[1 + \alpha_E \Delta T + \beta_E \Delta T^2 + \dots][1 + \alpha \Delta T + \beta \Delta T^2 + \dots] \\ &\approx f(P)E[1 + \alpha_E \Delta T][1 + \alpha \Delta T] \\ &\approx f(P)E[1 + (\alpha_E + \alpha)\Delta T] \end{aligned}$$

当 $\alpha_E = -\alpha$ 时, 有

$$U \approx f(P)E \quad (2-141)$$

式(2-141)表明, 在 $\alpha_E = -\alpha$ 条件下, 电桥输出 U 不是温度的函数。

恒流源供电：

$$\begin{aligned} U &= \Phi(P) I [1 + \alpha_1 \Delta T + \beta_1 \Delta T^2 + \dots] [1 + \alpha' \Delta T + \beta' \Delta T^2 + \dots] \\ &\approx \Phi(P) I [1 + \alpha_1 \Delta T] [1 + \alpha' \Delta T] \\ &\approx \Phi(P) I [1 + (\alpha_1 + \alpha') \Delta T] \end{aligned}$$

当 $\alpha_1 = -\alpha'$ 时，有

$$U \approx \Phi(P) I \quad (2-142)$$

上式表明，在 $\alpha_1 = -\alpha'$ 条件下，电桥输出 U 不是温度的函数。

因此，问题归结为制作一个电压源 E 或恒流源 I ，使它们与未补偿前电桥输出具有相反的温度特性，从而使传感器的电桥输出不受温度影响，在温度变化时输出 U 保持恒定。当然完全“拼凑”好是相当困难的，但是大大减小温度交叉灵敏度还是可能的。

2) 频率补偿

一个未补偿的初级经典传感器的工作频带往往较窄，使用范围受到限制，因此希望扩展它的工作频带。频带的扩展可以采用具有拼凑补偿思想的硬件电路来实现，也可采用软件编程来实现，有关内容将在第 4 章中详述。

2.4.4 多信号测量法

为了消除系统误差，提高测量精度，多信号测量法是一种有效的可行方法，常用的是二信号法与三信号法。

1. 二信号法

被测量 U_x 与参考基准 U_R 两个信号，经过相同的路径由相同的系统测量，如图 2-45 所示。若该系统的增益为 A ，系统的误差源为 E_o ，则

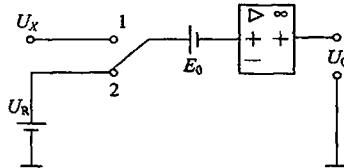


图 2-45 二信号测量法

测量 U_x 时输出

$$U_{o1} = (U_x + E_o)A$$

测量 U_R 时输出

$$U_{o2} = (U_R + E_o)A$$

求其比

$$\frac{U_{o1}}{U_{o2}} = \frac{(U_x + E_o)A}{(U_R + E_o)A}$$

推导 U_x 的表达式为

$$U_x = \frac{U_{o1}}{U_{o2}} U_R + E_o \left(\frac{U_{o1}}{U_{o2}} - 1 \right)$$

实际上我们由比值计算 U_x 的公式是

$$U_x = \frac{U_{01}}{U_{02}} U_R \quad (2-143)$$

故通过比值求 U_x 的绝对误差是

$$\Delta U_x = E_0 \left(\frac{U_{01}}{U_{02}} - 1 \right) \quad (2-144)$$

通过比值计算 U_x 的相对误差是

$$\frac{\Delta U_x}{U_x} = \frac{E_0}{U_x} \left(\frac{U_{01}}{U_{02}} - 1 \right) \quad (2-145)$$

如果不采用二信号求比值的方法来求 U_x ，而是直接用该系统输出值 U_{01} 来求 U_x ，则其相对误差为

$$\frac{\Delta U_{01}}{U_{01}} = \frac{AE_0}{U_{01}} \approx \frac{E_0}{U_x} \quad (2-146)$$

比较式(2-145)、式(2-146)可见，二信号比值法求 U_x 的相对误差是直接测量 U_x 的相对误差的 $\left(\frac{U_{01}}{U_{02}} - 1\right)$ 倍，所采用的参考信号 U_R 越接近被测量 U_x 时，输出值 U_{01} 也越近 U_{02} ，则相对误差越小，而且系统增益 A 变化产生的误差也可消除。

2. 三信号法

如图 2-46 所示。 R_x 为待测电阻，实际上它可以是用于温度测量的铂电阻，由已知电流 I_s 通过其压降 $I_s R_x$ 来求 R_x 。若测量系统增益为 A ，系统误差源为 E_0 ，再引入两个标准电阻 R_1 、 R_2 ，则

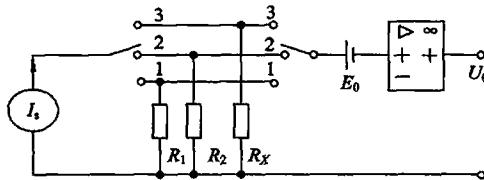


图 2-46 三信号测量法

R_x 上压降的测量值

$$U_{03} = (I_s R_x + E_0) A$$

R_2 上压降的测量值

$$U_{02} = (I_s R_2 + E_0) A$$

R_1 上压降的测量值

$$U_{01} = (I_s R_1 + E_0) A$$

单次测量值中均含有系统误差 E_0 。现在为消除 E_0 的影响作下列运算：

$$N = \frac{U_{03} - U_{01}}{U_{02} - U_{01}} = \frac{I_s R_x - I_s R_1}{I_s R_2 - I_s R_1} = \frac{R_x - R_1}{R_2 - R_1} \quad (2-147)$$

式(2-147)表明，差值之比 N 中不再含有误差 E_0 ，进而可推得求 R_x 的表达式如下：

$$R_x = N(R_2 - R_1) + R_1 \quad (2-148)$$

上式表明， R_x 由精密标准电阻 R_1 、 R_2 及不含系统误差 E_0 的比值 N 决定。三信号法的优点是：

- (1) 只要在三个信号量的测量期间内，系统误差 E_0 保持不变，则 E_0 可完全被消除。
- (2) 只要在三个信号量的测量期间内，增益 A 不变，则可消除由增益 A 变化引入的误差。
- (3) 只要在对三个信号量的测量期间内，恒流源供电电流 I_s 保持不变，则可消除恒流源波动引入的误差。

由此可知，采用三信号法可以用低精度的放大器与恒流源获得高精度的测量结果。

3. 三步测量法

基于三信号法，可以进一步引伸出一种行之有效的三步测量法。若被测量为 U_x ，一个标准量为 U_R ，另一个标准量选为零，如图 2-47 所示，则可按下述三步进行测量：

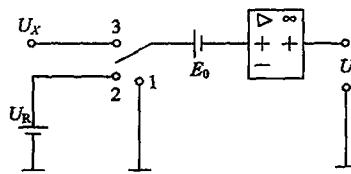


图 2-47 三步测量法

第一步：测零点。系统输入为零时之输出：

$$U_{01} = E_0 A$$

第二步：标定。系统输入为标准量 U_R 时之输出：

$$U_{02} = (U_R + E_0) A$$

第三步：测量。系统输入为被测目标参量 U_x 时之输出：

$$U_{03} = (U_x + E_0) A$$

进行相减、相除运算求比值 N ：

$$N = \frac{U_{03} - U_{01}}{U_{02} - U_{01}}$$

将上述三步测量值 U_{01} 、 U_{02} 、 U_{03} 代入上式，得

$$N = \frac{(U_x + E_0) A - E_0 A}{(U_R + E_0) A - E_0 A} \approx \frac{U_x}{U_R} \quad (2-149)$$

于是可得 U_x 的表达式如下：

$$U_x = N \cdot U_R = \frac{U_{03} - U_{01}}{U_{02} - U_{01}} U_R \quad (2-150)$$

根据式(2-149)和式(2-150)可见，被测目标参量 U_x 由比值 N 与标准量 U_R 决定。而比值 N 中已消除了系统误差 E_0 ，同时也不存在增益 A ，因此三步测量法的第一步实质为测量零点；第二步是以标准量 U_R 为标准对全系统总增益作实时标定；第三步是测量。三步测量法的优点是：

- (1) 只要在三步测量时间内，系统增益未变化从而可为常量时，在三步测量期间之外增益变化不引入误差。
- (2) 系统误差 E_0 只要在三步测量时间内保持不变，则所引入的误差可以完全消除。
- (3) 为保证测量精度需要有足够精度的标准信号 U_R 。

这种三步测量法目前主要用于对经典传感器的后续环节的总体校验与标定，大大降低了对后续环节稳定性的要求，这些参数只需在三步测量的时间内暂时保持不变就行，测量精度主要依靠参考标准信号 U_R 来保证。

如果参考标准信号不是由基准电压、标准电阻、标准电容等电气量来提供的，而是由与被测目标量同类属性的标准来提供的，如标准压力发生器、标准恒温器等，那么三步测量法就用来进行包括传感器在内的传感器系统的整体自校与检验。这样就可实现采用低精度的传感器及其低精度的后续系统，获得高精度的测量结果，而成为高精度传感器系统。

对于与微型计算机相结合的智能传感器系统，三步测量法的优越性发挥得更加淋漓尽致，微型计算机的数据采集速度比人工操作快很多，因此对传感器及其后续系统的要求可以更加降低，只需在极短的三步测量时间内($<60\text{ ms}$)稳定不变即可。

参 考 文 献

- [1] 刘君华, 等. 传感器技术及应用实例[M]. 北京: 电子工业出版社, 2008
- [2] 刘沁. 新型工业用硅电容差压传感器、变送器的研制[C]. 第九届全国敏感元件与传感器学术会议论文集. 西安: 全国敏感元件与传感器学术团体联合组织委员会, 2005: 15-21
- [3] 刘君华. 现代测试技术与测试系统设计[M]. 西安: 西安交通大学出版社, 1999
- [4] 宋宗炎. 现代传感器技术[J]. 测控技术, 1996, 15(4): 5
- [5] 鲍敏杭. 微电子传感器的新进展[C]. 第四届全国敏感元件与传感器学术会议论文集. 上海: 全国敏感元件与传感器学术团体联合组织委员会, 1995: 117
- [6] 俞朴, 等. 试论提高传感器性能的技术途径[C]. 第二届全国敏感元件与传感器学术会议论文集. 上海: 全国敏感元件与传感器学术团体联合组织委员会, 1991: 857
- [7] 杨明, 刘君华. 实现精密测温的脉冲调宽与三步替代法[J]. 电子测量与仪器学报, 1991, 5(1): 43-48
- [8] EJA 智能变送器选型样本. 横河川仪有限公司
- [9] ST-3000 型智能变送器使用说明. 美国霍尼韦尔公司
- [10] 3051 型智能压力变送器系列. 费希尔-罗斯蒙特资讯集. FISHER - ROSEMOUNT 中国公司, 1998(11)

第3章 智能传感器系统的组建与集成调理电路芯片介绍

本章内容

- 智能传感器系统的基本组成形式
- 基于虚拟仪器平台实现数据采集与显示功能
- 各种不同集成度调理电路芯片简介

随着微机械加工技术和大规模集成电路工艺技术的迅猛发展，沿着集成化、非集成化、混合集成化三条途径，智能传感器系统获得了飞速发展。目前，已经有单片集成及各种不同集成度商品化芯片以及功能极高的 I/O 接口设备——数据采集卡可供构建智能传感器系统使用。

在智能传感器系统三条实现途径基础上，本章介绍构建智能传感器系统的两种形式：一是传感器与个人计算机(Personal Computer, PC)相结合的虚拟仪器形式；二是传感器与微处理器 MPU 相结合的智能仪器形式。不论哪种结合形式都需要调理电路。因此，本章还介绍了大量不同集成度调理电路芯片供读者构建智能传感器系统参考。

3.1 智能传感器系统的基本组成形式

智能传感器系统主要由传感器、调理电路、数据采集与转换、计算机及其 I/O 接口设备四大部分组成，如图 3-1 所示。

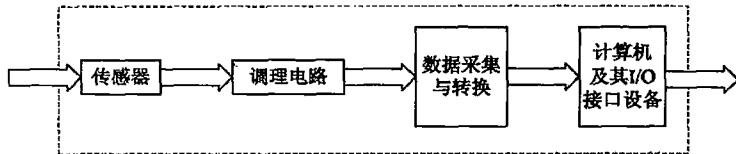


图 3-1 智能传感器系统的基本组成

系统各组成部分的功能分述如下。

3.1.1 传感器

传感器完成信号的获得，它将规定的被测参量按一定规律转换成相应的可用输出信号。被测参量可以是各种非电参量，也可以是电气参量。如电力输电线路高电压电网，可通过电压互感器将高电压变为 100 V 电压，通过电流互感器将电网大电流变为 5 A 后，仍需采用电压、电流传感器或变送器将 100 V 电压及 5 A 电流分别转换成 5 V 低电压，送入

A/D转换器。在单片集成芯片中，A/D与微计算机集成在同一芯片中；在非集成化实现方式中，A/D往往在数据采集卡中。

3.1.2 信号调理

来自传感器的输出信号通常是含有干扰噪声的微弱信号。因此，后面配接的信号调理电路的基本作用有三个：其一是放大，将信号放大到与数据采集卡(板)中的A/D转换器相适配；其二是预滤波，抑制干扰噪声信号的高频分量，将频带压缩以降低采样频率，避免产生混淆，如果信号调理电路输出的是规范化的标准信号，即4~20 mA的电流信号，则称这种信号调理电路与传感器的组合为变送器；其三是转换，将传感器输出的电参量，如电容C、电感L或M、电阻R的改变量，转换为电压或频率量，即 C/u 、 L/u 、 M/u 、 R/u 转换等。此外，根据需要还可进行信号的隔离与变换等。

3.1.3 数据采集与转换

数据采集部分由采样/保持(S/H)与多路切换开关(MUX)组成，实现对多传感器多点多通道输入信号的分时或并行采样。时间连续信号 $x(t)$ 经过采样后变为离散时间序列 $x(n)$ ， $n=0, 1, 2, \dots$ 。

数据转换部分为A/D、D/A转换器或V/F转换器。A/D(模/数)转换器将信号的采样值转换为幅值离散化的数字量，或由V/F(伏/频)转换器转换为脉冲频率；D/A(数/模)转换器输出模拟控制信号。在以PC与传感器接合的非集成化实现方式中，A/D、D/A、MUX以及可编程放大器PAG集中放在一块DAQ数据采集卡(板)中，并将DAQ插入PC相应的空槽中。

3.1.4 计算机及其I/O接口设备

计算机是神经中枢，它使整个测量系统成为一个智能化的有机整体，在软件导引下按预定的程序自动进行信号采集与存储，自动进行数据的运算分析与处理，指令以适当的形式输出，显示或记录测量结果。根据采用的计算机类型，传感器系统可分为以下两种形式。

(1) 智能仪器式智能传感器系统：以微型计算机或微处理器(Microprocessor)为核心，于20世纪80年代初开始应用，是测量技术与计算机最初结合的形式。如美国霍尼韦尔公司的ST-3000型和美国罗斯蒙特公司的3051型智能压力变送器，将微处理器所在的数据处理主板放到压力传感器腔内，将传感器与计算机赋予智能，形成了智能传感器/变送器，打破了传感器与仪器的界限。

(2) 虚拟/集成仪器式智能传感器系统：以个人计算机(Personal Computer)为核心，充分利用PC的运算与分析处理功能和显示功能，打破了计算机与仪器的界限。由于其有更强大的运算与信号分析处理和显示功能，所以与以微处理器为核心构成的仪器相比，虚拟/集成仪器式智能传感器系统有更强大的智能，实现起来更容易、更快捷。

图3-1所示的智能传感器系统有多种集成实现方式，如不同组成环节的集成，有配接电阻型、电容型、电感型不同类型传感器的信号调理芯片，有数据采集、转换与微处理器的集成芯片，还有全系统的单片集成，或采用PC与其I/O接口设备DAQ(数据采集卡)硬件集成。

3.2 基于虚拟仪器平台实现数据采集与显示功能

在 DAQ 与 PC 的硬件集成平台上先构建一个虚拟仪器——“数据采集与显示仪”，再与任一传感器及其调理电路相结合，这是虚拟/集成仪器式智能传感器系统最基本的形式，也是实现智能传感器系统最便捷的方式。其中，数据采集卡(DAQ)作为 PC 的一种外围输入/输出(I/O)设备，主要执行来自传感器及其调理电路的信号的采集、放大、模/数(A/D)与数/模(D/A)转换任务；PC 则对采集输入的数字量数据进行进一步的信号分析与处理，其最基本的任务是对采集输入的数字量数据进行刻度转换与显示。

3.2.1 数据采集卡(DAQ)的基本性能指标

数据采集卡的主要性能指标概述如下。

1. 模拟信号输入部分

- (1) 模拟输入通道数：该参数表明数据采集卡所能够采集到的最多的信号路数。
- (2) 信号的输入方式：一般数据采集卡设备的可供选择的待采集信号输入方式有以下 4 种。

- 单端输入：信号的一个端子接地。
- 差动输入：信号两端均浮地。
- 单极性：信号幅值范围为 $[0, A]$ ， A 为信号最大幅值。
- 双极性：信号幅值范围为 $[-A, A]$ 。

使用者可根据实际情况进行选择。

- (3) 模拟信号的输入范围(量程)：一般根据信号输入极性而定，如单极性输入的典型值为 $0 \sim 10$ V，双极性输入的典型值为 $-5 \sim 5$ V。
- (4) 放大器增益。
- (5) 模拟输入阻抗：采集卡固有参数，一般不由用户设置。

2. A/D 转换部分

1) 采样速率

采样速率是指在单位时间内数据采集卡对模拟信号的采集次数，是数据采集卡的重要技术指标。为了使采样后输出的离散时间序列信号能无失真地复现原输入信号，由采样定理可知采样频率 f_s 至少应为输入信号最高有效频率 f_{\max} 的 2 倍，否则会出现频率混淆误差。实际系统中，为了保证数据采集精度，一般有下列关系：

$$f_s = (7 \sim 10)f_{\max} \times N \quad (3-1)$$

式中， N 为多通道数据采集系统的通道数。

2) 位数 b

位数是指 A/D 转换器输出二进制数的位数。如图 3-2 所示的 8 位 A/D，当输入电压由 $U=0$ 增至满量程值 $U=U_H$ 时，一个 8 位($b=8$)A/D 的数字输出由 8 个“0”变为 8 个“1”，共计变化 2^8 个状态，故 A/D 转换器产生一个最低有效位数字量输出改变量，相应的

输入量 $U_{\min} = 1\text{LSB} = q$ 可由下式计算：

$$1\text{LSB} = q = \frac{U_H}{2^8 - 1} \quad (3-2)$$

式中， q 为量化单位； U_H 为满量程输入电压，通常等于 A/D 转换器的电源电压。

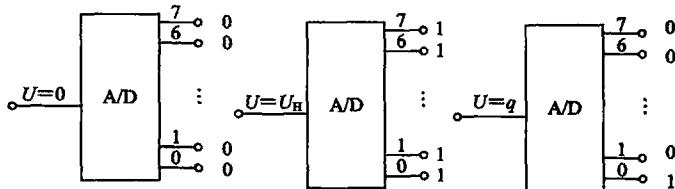


图 3-2 8 位 A/D 的输入与输出

3) 分辨率与分辨力

分辨率与分辨力指数据采集卡可分辨的输入信号最小变化量。分辨率一般以 A/D 转换器输入的二进制位数 b 或 BCD 码位数表示；分辨力为 1LSB(最低有效位数)。

3. D/A 模数转换部分

(1) 分辨率：当输入数字发生单位数码变化，即 1LSB 时，所对应输出模拟量的变化量即为分辨率，通常用 D/A 转换器的转换位数 b 表示。

(2) 标称满量程：相当于数字量标称值 2^b 的模拟输出量。

(3) 响应时间：数字量变化后，输出模拟量稳定到相应数值范围内(1LSB/2)所经历的时间。

以上为数据采集卡的主要性能指标。对一些功能丰富的数据采集卡，还有定时/计数等其他功能，相应地还有其他相关指标，这里不再赘述。

3.2.2 数据采集卡的安装

数据采集卡通常都是插卡式结构，即将 DAQ 卡插入 PC 相应的标准总线扩展插槽内，与 PCI 总线或 ISA 总线相连，就可在 PC 的控制下完成数据采集、模拟信号输出等功能。

3.2.3 I/O 接口设备 Lab-PCI-6024E 数据采集卡简介

Lab-PCI-6024E 是 NI 公司 E 系列多功能数据采集卡之一，是一种性能优良的低价位的适合 PC 及其兼容机的采集卡。它能够完成信号采集转换(A/D)、数字信号的模拟输出(D/A)及定时/计数等功能。它集 12 位 A/D 转换器、12 位 D/A 转换器、16 路单端接地的模拟输入通道、8 位或 24 位并行输入输出线(5 V/TTL)及两路 24 位定时器与计数器为一体，支持 DMA 方式和双缓冲区模式，保证了实时信号不间断采集与存储，具体功能如下。

1. 模拟信号输入部分

(1) 输入通道：16 路单端接地的模拟输入通道。

(2) 信号的输入方式：差动方式、单端有参考地及单端无参考地方式。

(3) 增益：有 0.5、1.0、10.0 和 100.0 四种增益可选。

(4) 输入电压范围: $\pm 50 \text{ mV}$, $\pm 500 \text{ mV}$, $\pm 5 \text{ V}$, $\pm 10 \text{ V}$, 单、双极性各有 4 挡可选。

2. A/D 转换部分

(1) 采样速率: 最高采样率为 200 kHz。

(2) 位数: 12 位。

3. D/A 转换部分

(1) 分辨率: 12 位。

(2) 标称满量程: $[-10 \text{ V}, 10 \text{ V}]$ 。

(3) 响应时间: $10 \mu\text{s}$ 。

3.2.4 实现数据采集卡软件驱动前的参数设置

对于 NI 公司生产的 I/O 接口设备, 如 DAQ 卡、GPIB、VISA、VXI、IMAQ 等, 进行数据采集前的参数设置是在 NI 公司提供的“Measurement & Automation”软件窗口中进行的。

1. 调用“Measurement & Automation”

“Measurement & Automation”图标在 Windows 桌面上, 它是在安装完 LabWindows/CVI 或 LabVIEW 后自动产生的。

调用方法: 鼠标双击“Measurement & Automation”图标。

2. 寻找所用 NI 公司数据采集卡的名称列表项

执行路径: [Devices and Interfaces] → [Traditional] → [NI - DAQDevices]
在“Measurement & Automation”浏览窗口中应出现所用数据采集卡名称的列表项。

3. 参数设置

出现所用数据采集卡名称的列表项后, 双击 [Properties] 选项即可进入参数设置用户界面。参数设置有 6 项: System、AQ、AO、Accessory、OPC 和 Remote Access, 可依据所用的具体数据采集卡的性能参数逐项进行设置。

3.2.5 [示例 3-1] 基于 DAQ 与 PC 实现虚拟仪器式的数据采集与显示

如果将数据采集与显示这一基本功能用虚拟仪器(VI)形式实现, 即利用数据采集卡 DAQ 与 PC 硬件先设计实现一个“信号采集与显示仪”, 那么与不同传感器及其调理电路相配合就可以构建测量不同参量的智能传感器系统。

1. 信号采集与显示仪的功能

该信号采集与显示仪可以采集和显示一路信号, 参数如下。

频率范围: $0 \sim 10 \text{ kHz}$ 。

幅值范围: $0.1 \sim 6 \text{ V}$ 。

采用 Lab-PCI-6024E 卡, 信号由 0 通道输入, 由 LabVIEW 软件驱动。

2. 设计提示

在设计 VI 之前, 需在“Measurement & Automation”驱动软件下, 进行相关参数的设置。设置完毕后, 进入 LabVIEW 进行前面板和流程图的设计。

1) 参数设置

因为只需数据采集与输入，故只需设置 AI 项。

AI 设置项：因为输入信号幅值为 0.1~6 V，故选模拟信号输入方式为有参考地单端输入，量程为 0~10 V。

2) 前面板控件

(1) 一个输出显示型控件，显示采集信号的波形。由 [Controls]→[Graph]→[Waveform Graph] 路径调用 Graph 控件。

(2) 两个输入控制型数字控件，设置采样频率和采样点数。采样频率选择 100 kHz，因为被测信号的最高频率为 10 kHz。

3) 流程图设计

双击 AI Acquire Waveform.vi 图标，即可进行流程图设计。参数设置如下。

(1) 调入路径：[Functions]→[NI Measurements]→[Data Acquisition]→[Analog Input]→[AI Acquire Waveform vi]。

(2) AI Acquire Waveform.vi 图标及端口如图 3-3 所示。

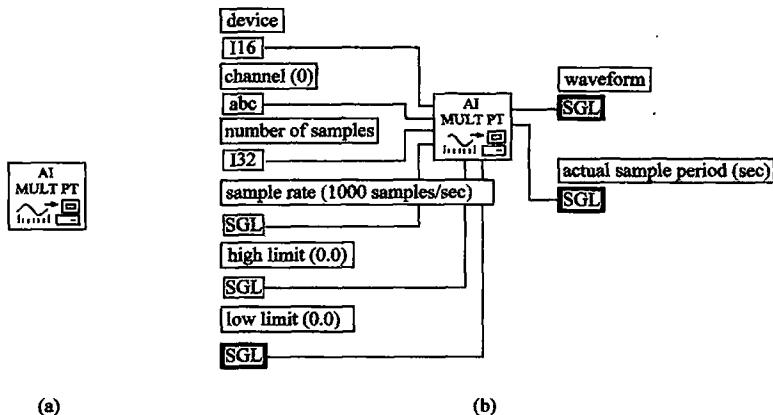


图 3-3 AI Acquire Waveform.vi 图标与其端口

(a) AI Acquire Waveform.vi 图标；(b) AI Acquire Waveform.vi 图标与其端口

输入端口：

- device——DAQ 的设备号，已在 Measurement & Automation 中设定，本例为 1。
- channel——待采集的模拟信号输入通道号，本例由 0 通道输入。
- number of samples——采样点数，本例由前面板设定。
- sample rate——采集速率，本例由前面板设定。
- high limit (0.0)——设置模拟信号的最大输入值，已在 Measurement & Automation 设置软件中设定。
- low limit (0.0)——设置模拟信号的最小输入值，已在 Measurement & Automation 设置软件中设定。

输出端口：

- waveform——存放采集所得数据的一维数组，它是前面板 Graph 控件的数据输

入端。

- actual sample period: 实际采样周期，这个数据和本例没有关系。

设计好的前面板与流程图如图 3-4 所示。

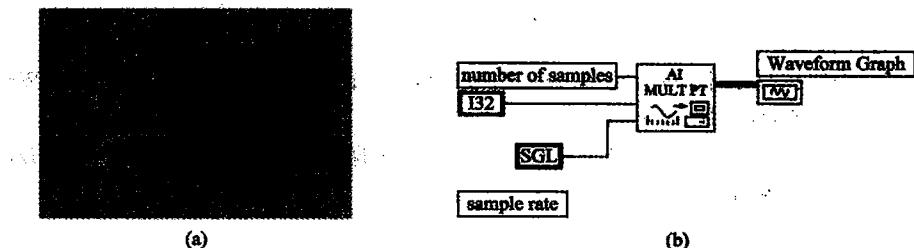


图 3-4 连续信号的采集与显示演示仪

(a) 前面板; (b) 流程图

3. 运行检验

将幅值在 0.1~6 V、频率在 0~10 kHz 范围内的正弦信号或其他波形的连续信号接入采集卡的模拟 0 通道，然后运行程序，在面板的图形控件上应显示出该信号的波形。

3.3 电阻电桥式传感器的单片集成调理电路 ——MAX1450 芯片

MAX1450 是美国 Maxim 公司针对压阻式压力传感器的应用而设计的信号调理集成芯片，可用于加速度计、应变检测和其它低电平桥式电阻型传感器。

MAX1450 芯片外接元件较少，只需简单的设置和调整，便可对传感器的失调、满量程输出、失调温度系数、满量程输出温度系数及满量程输出非线性进行校准和补偿，使信号输出精度提高到 1%（相对于传感器的线性误差），简化了传感器输出信号调理电路的结构。

3.3.1 MAX1450 芯片的引脚功能与结构框图

1. 引脚图与功能简介

MAX1450 采用 20 引脚 SSOP 封装，单 5 V 电源供电，芯片引脚如图 3-5 所示。

2. 引脚功能简介

引脚 1、20(INP、INM)：传感器输出信号正、负接入端，输入阻抗为 $1 M\Omega$ 。

引脚 13、6、7(A2、A1、A0)：可编程放大器增益设置数字量输入端，连接至 V_{DD} 为高电平，连接至 V_{SS} 或悬空为低电平。

引脚 8(OFFTC)：失调温度补偿电压 V_{OFFTC} 输入调节端。

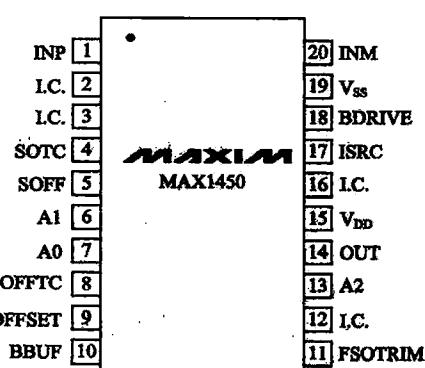


图 3-5 MAX1450 引脚图

引脚 9(OFFSET)：失调电压 V_{OFFSET} 输入调节端。

引脚 4、5(SOTC、SOFF)：分别为失调温度补偿电压 V_{OFFTC} 符号位、失调电压 V_{OFFSET} 符号位输入端。在输出 $V_{OUT} = V_{PGA} \pm V_{OFFTC} \pm V_{OFFSET}$ 中，SOTC、SOFF 连接到 V_{DD} 时相应项取正，连接到 V_{SS} 或悬空时取负。

引脚 11(FSOTRIM)：传感器桥的激励电流设置输入端。

引脚 17(ISRC)：电流源参考端，该端与 V_{SS} 间可连接参考电阻。

引脚 18(BDRIVE)：芯片对传感器的激励电流输出端。

引脚 10(BBUF)：BDRIVE 端的桥压缓冲输出。

引脚 14(OUT)：信号调理芯片的电压输出端。

引脚 15、19(V_{DD} 、 V_{SS})：工作电源正、负接人端。

引脚 2、3、12(I.C.)：内部连接端，使用时悬空。

3. 结构框图与组成

1) 结构框图

MAX1450 的结构框图如图 3-6 所示。

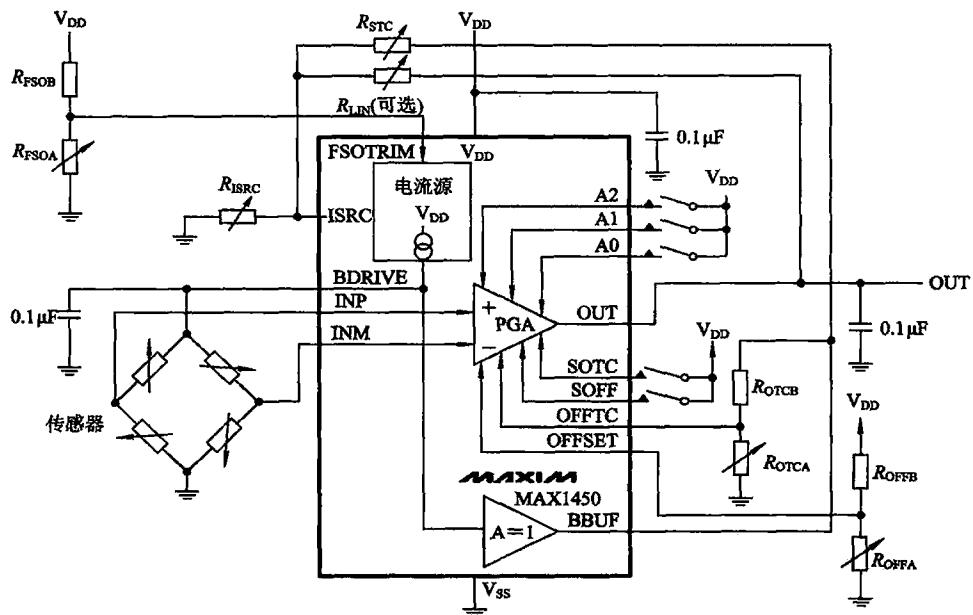


图 3-6 MAX1450 芯片的内部结构框图

2) 单元模块简介

MAX1450 芯片主要由可编程放大器 PGA 模块与可调节激励电流源 (CURRENT SOURCE) 及桥压缓冲输出级模块两个单元模块组成。

(1) PGA 模块。PGA 模块主要由三位增益可编程放大器 PGA、求和单元 Σ 及单位增益缓冲器 ($A=1$) 三部分组成，如图 3-7 所示。

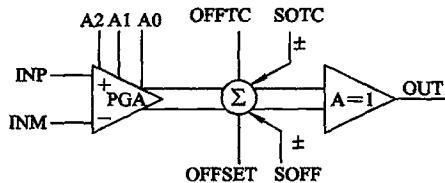


图 3-7 PGA 模块的主要组成部分示意图

① 三位增益可编程放大器 PGA。它的增益可由 A2、A1、A0 三位数码进行设置，数码值与相应增益的关系如下：

A2	A1	A0	增益
0	0	0	39
0	0	1	65
0	1	0	91
1	0	0	117
1	0	1	143
1	1	0	169
1	1	1	195

引脚 13(A2)、6(A1)、7(A0)均通过开关 S 接电源 V_{DD}，故当 S 闭合时为高电平 1，S 断开时悬空为低电平 0，从而 A2、A1、A0 组成不同数码。

PGA 的差分输入端为 INP(1 脚)、INM(20 脚)，差分输入电平可低至 10~30 mV，共模抑制比高达 90 dB；输入阻抗 1 MΩ。

② 求和单元 Σ。将放大输出的差分信号电压与由外部施加的信号控制的失调补偿电压、失调温度补偿电压及其符号进行求和相加。

③ 单位增益缓冲器 A。将双端差分输入转换为单端输出，由 14 脚 OUT 输出的信号可以是：经过零点(失调)补偿、温度补偿的；与差分输入信号成比例放大的信号。

3) 电流源(CURRENT SOURCE)及桥压缓冲输出级模块

该电流源是可调节的，调节范围为 0.1~2 mA。输出电流的大小由外接于 FSOTRIM 端(11 脚)的电压(由外接分压器的分压决定)以及接于 ISRC 与 V_{ss}端的电阻确定，用于驱动外部传感器桥。

3.3.2 MAX1450 的基本功能与补偿校准功能

1. 基本功能与初始工作状态的建立

MAX1450 的基本功能是：向传感器桥路提供工作电流；可对传感器桥路输出电压进行放大；放大倍数可由程控数字设定。实现以上基本功能的过程也是建立传感器与 MAX1450 的初始工作状态的过程。

1) 电桥型传感器与 MAX1450 的连接

(1) 传感器桥路供电端的连接：传感器桥路的两个供电端分别接 MAX1450 的 18 脚 BDRIVE 和 19 脚 V_{SS}(地)。

(2) 传感器桥路输出端的连接。传感器的两个不平衡电压输出端分别接 MAX1450 的 1 脚 INP 及 20 脚 INM。

2) 传感器供电电流的调定

(1) 用可调电阻器设置 R_{ISRC} 的初值。在 MAX1450 的 17 脚 ISRC 端与 19 脚 V_{SS}(地)之间接入电阻 R_{ISRC}，该电阻初值由下式确定：

$$R_{ISRC} \approx 13 \cdot R_b \quad (3-3)$$

式中 R_b 为传感器的输入/输出阻抗，在等臂电桥的情况下也是桥臂的电阻，可用万用表欧姆挡测量或由传感器技术指标资料获得。

(2) 调节分压比。由外接电阻 R_{FSOB} 与 R_{FSOA} 串联后接电源 V_{DD} 构成电压分压器，其分压端连至 MAX1450 的 11 脚——FSOTRIM 端。

观测传感器桥路供电端，也即 MAX1450 的 18 脚 BDRIVE 端的电压值 U_{BDRIVE}。该电压值与传感器输入阻抗 R_b 以及供电电流 I 有如下关系

$$U_{BDRIVE} \approx R_b \cdot I \quad (3-4)$$

当已知 R_b = 2 kΩ 时，若测得 U_{BDRIVE} ≈ 2.5 V，则可知工作电流 I = 1.25 mA，在 MAX1450 电流源可调节的范围(0.1~2 mA)内，则可视为已基本调定成功。

3) 放大倍数的设置与估算

(1) 放大倍数 A_{PGA} 的估算。估算公式如下：

$$A_{PGA} = \frac{U_{OUT\ max} - U_{OUT\ min}}{S(T) \cdot U_{BDRIVE} \cdot (P_{max} - P_{min})} \quad (3-5)$$

式中 S(T) 为室温下传感器灵敏度，表示传感器桥供电电压为 1 V 时，输入被测压力 1PSI 引起桥路输出的不平衡电压值。

U_{BDRIVE} 为传感器供电电流调定完毕之后实际的供电端电压值。

P_{max}、P_{min} 与 U_{OUT}_{max}、U_{OUT}_{min} 分别为传感器输入压力的最大、最小值，以及期望在 MAX1450 14 脚(OUT 端)产生的输出电压最大、最小值。

设 S(T) = 10 mV/(V · PSI)，U_{BDRIVE} = 2.5 V，P_{max} - P_{min} = 2PSI，U_{OUT}_{max} - U_{OUT}_{min} = 4 V，代入式(3-5)，可估算所需放大倍数 A_{PGA}：

$$A_{PGA} = \frac{4 \text{ V}}{10 \text{ mV}/(\text{V} \cdot \text{PSI}) \times 2.5 \text{ V} \times 2 \text{ PSI}} = 80$$

(2) 放大倍数的设置。经估算，所需的放大倍数 A_{PGA} = 80，可由数码[A2, A1, A0]予以设定，但因 MAXA1450 的可编程放大器 PGA 能设置的放大倍数只能是在 39~195 之间一些间隔为 26 的离散值，增益值应取小于但最接近所需值的数值。现取[A2, A1, A0] = [0 0 1]，即 65 < 80，以防过载。

2. 补偿校准功能

补偿校准项目有：满量程校准，满量程温度系数补偿，失调温度系数补偿，失调校准，线性校正等可选项。MAX1450 芯片的补偿校准功能为改善传感器静态性能、温度稳定性以及提高测量准确性提供了可能。

1) 参数初始化

在进行补偿校准操作前，需要设置待补偿校准传感器的原始参数，称为参数初始化。这些参数可在两个压力值(最小、最大值)和两个温度点测量获得，而参数的测量应是在传感器与 MAX1450 相连接完成了初始工作状态的建立后进行的。

(1) R_b : 室温下传感器的输入/输出阻抗($k\Omega$)，也是传感器桥路臂桥的电阻值。

(2) T_{CR} : 传感器输入/输出阻抗的温度系数($\text{ppm}/^\circ\text{C}$)。

(3) T_{CS} : 传感器灵敏度温度系数($\text{ppm}/^\circ\text{C}$)。

(4) $S(T)$: 室温下传感器灵敏度($\text{mV}/\text{V}/\text{PSI}$)；即传感器桥供电端为 1 V 时，输入压力为 1PSI 时的输出电压。

(5) $S(P)$: 室温下传感器灵敏度的非线性误差(% FSO)，或称灵敏度的线性度。

(6) P_{min} 、 P_{max} : 最小、最大输入压力，单位为 PSI，

$$1\text{PSI}(\text{磅力}/\text{英寸}^2) = 51.715 \text{ mmHg} = 6890(\text{Pa})$$

(7) $O(T)$: 室温下失调系数(mV/V)；在传感器桥供电端为 1 V，输入压力为零时输出的不平衡电压。

(8) OTC : 失调温度系数($\text{ppm FSO}/^\circ\text{C}$)。

2) 满量程校准

满量程校准可在室温 T_1 下进行。假设传感器桥压设置在 2.5 V，最小压力输入下芯片输出 $U_{OUT\ min}$ 为 0.5 V，最大压力输入下芯片输出 $U_{OUT\ max}$ 为 4.5 V，这时，应按如下步骤进行调整。

(1) 调节电阻 R_{FSOA} ，使芯片 FSOTRIM 端电压 $U_{FSOTRIM}$ 为 2.5 V，然后调节 R_{ISRC} ，使 U_{BBUF} 为 2.5 V。

(2) 调节 R_{OFFA} ，改变 OFFSET 端电压，使芯片在最小压力输入下失调输出为 0.5 V。

(3) 测最大压力输入下的输出 $U_{OUT\ FSO}$ 。

(4) 用下式计算电压 $U_{BIDEAL}(T_1=25^\circ\text{C})$ ：

$$U_{BIDEAL}(T_1=25^\circ\text{C}) = U_{FSOTRIM} \left[1 + \frac{U_{OUT\ max} - U_{OUT\ FSO}}{U_{OUT\ FSO}} \right]$$

如果计算出的 U_{BIDEAL} 超出桥路电压的允许摆幅($U_{SS}+1.3\text{ V} \sim U_{DD}-1.3\text{ V}$)，则需要重新设置 PGA 的增益。若 U_{BIDEAL} 太低，可降低一个步长的增益； U_{BIDEAL} 太高，可增加一个步长的增益，然后重新回到第一步，直到 U_{BIDEAL} 在规定的摆幅内。

(5) 调节电阻 R_{FSOA} ，使 $U_{FSOTRIM} \approx U_{BIDEAL}$ ，再调节电阻 R_{ISRC} ， $U_{BBUF} = U_{BIDEAL}(T_1)$ 。

重新调节 R_{OFFA} ，改变 OFFSET 端电压，使芯片在最小压力输入下失调输出为 0.5 V。

3) 满量程温度系数补偿

连接于 BBUF 和 ISRC 端的电阻 R_{STC} 被用于满量程温度系数补偿。对于大多数硅压力传感器，其典型值在 100 k Ω 左右。用可调电阻器设置 R_{STC} 的初值：

$$R_{STC} \approx \frac{500R_{ISRC}}{T_{CR} - T_{CS}} \quad (3-6)$$

式中， T_{CR} 为传感器输入/输出阻抗的温度系数； T_{CS} 为传感器灵敏度温度系数。

补偿过程调节如下：

(1) 在温度 T_2 环境下，测量满量程输出 $U_{OUT\ FSO}(T_2)$ 。

(2) 用公式计算 $U_{\text{BIDEAL}}(T_2) = U_{\text{FSOTRIM}} \times \left[1 + \frac{U_{\text{OUT max}} - U_{\text{OUT FSO}}}{U_{\text{OUT FSO}}} \right]$ 。

(3) 调节 R_{STC} , 直到 $U_{\text{BBUF}} = U_{\text{BIDEAL}}(T_2)$ 。

4) 失调温度系数补偿

连接于 BBUF 和 V_{ss} 端的分压电阻 R_{OTCB} 、 R_{OTCA} 用于调节温度系数补偿电压的输入值, SOTC 端的连接决定补偿的极性。

使 $R_{\text{OTCB}} = 100 \text{ k}\Omega$, 可变电阻 $R_{\text{OTCA}} = 200 \text{ k}\Omega$, 调整合适的阻值比, 并根据此比值选择 SOTC 端的连接。SOTC 端接 V_{dd} 时为正, SOTC 端接 V_{ss} 或悬空时为负。 R_{OTCB} 和 R_{OTCA} 的阻值比应满足下式要求:

$$\frac{R_{\text{OTCA}}}{R_{\text{OTCA}} + R_{\text{OTCB}}} \approx \frac{\text{OTC}(U_{\text{OUT max}} - U_{\text{OUT min}})}{1.15 T_{\text{CS}} U_{\text{BDRIVE}}} \quad (3-7)$$

式中, OTC 为失调温度系数, T_{CS} 为传感器灵敏度温度系数。

补偿过程调节如下:

(1) 在温度 T_1 (25°C)、 T_2 环境下, 分别测量输出失调电压 $U_{\text{OUT FSO}}$ 及 BDRIVE 端的电压 U_{BDRIVE} 值。

(2) 调节 R_{OTCA} , 使 OFFTC 端电压满足下式:

$$U_{\text{OFFTC}}(T_2) \approx \frac{U_{\text{OUT FSO}}(T_1) - U_{\text{OUT FSO}}(T_2)}{U_{\text{BDRIVE}}(T_1) - U_{\text{BDRIVE}}(T_2)} \times 1.15 \quad (3-8)$$

如果 $U_{\text{OFFTC}}(T_2)$ 为正, STOC 端接 V_{dd}; $U_{\text{OFFTC}}(T_2)$ 为负, 则 SOTC 端接地或悬空。

(3) 若 PGA 的增益改变, 以上步骤需重新进行。

5) 失调校准

连接于 V_{dd} 和 V_{ss} 端的分压电阻 R_{OFFA} 、 R_{OFFB} 提供一个分压, 接于 OFFSET 端以进行失调校准。SOFF 端的连接决定输入校准电压的极性。校准过程如下:

(1) 在温度 T_1 (25°C) 环境下, 先使 $U_{\text{OFFSET}} = 0 \text{ V}$, 测量最小压力输入下的输出失调电压 U_{ZERO} 。若 $U_{\text{ZERO}} < 0.5 \text{ V}$, SOFF 端接 V_{dd}; 若 $U_{\text{ZERO}} > 0.5 \text{ V}$, SOFF 端接 V_{ss} 或悬空。

(2) 调节 R_{OFFA} , 直到最小压力输入下的输出失调电压 $U_{\text{OUT min}} = 0.5 \text{ V}$ (期望失调输出)。

(3) 若 PGA 的增益改变, 以上步骤需重新进行。

6) 线性校正(可选)

线性校正是将输出电压 U_{OUT} 通过电阻 R_{LIN} 反馈到 ISRC 端, 用调节电流源的输出来实现的。因为线性校正需要的一些系统变量不易测量, 所以线性误差的校正只能通过经验估算法选取合适的 R_{LIN} 。

(1) 按如下经验公式选择可变电阻 R_{LIN} 的值:

$$R_{\text{LIN max}} \approx \frac{2R_{\text{ISRC}}R_{\text{STC}}}{(R_{\text{ISRC}} + R_{\text{STC}})S(p)} \quad (3-9)$$

式中 $S(p)$ 为室温下传感器灵敏度的非线性误差或称灵敏度的线性度。

(2) 在(50%~100%) $R_{\text{LIN max}}$ 范围内取 R_{LIN} 的几个典型值, 测量相应的非线性误差, 建立一个关系表, 从中选择合适的 R_{LIN} 大小。

注意: 线性校正必须在温度补偿完成后进行。线性校正完成后, 需要重新进行满量程校准。若不需要线性校正, 去掉电阻 R_{LIN} 即可。

3.4 适配压阻式传感器的单片集成调理电路 ——MAX1460 芯片

有多种集成化调理电路芯片，将调理电路与微处理系统集成在一块芯片上，片内有温度传感器，可以测量片外环境温度；更发挥其特点的是可以提供温度补偿所需的温度信号而无需用户另外连接片外温度传感器。有特别适合配接压阻式传感器的芯片 MAX1460，还有适合配接桥路型传感器或热电阻或热电偶、用作温度补偿和线性化处理的芯片 MAX1464 等多种调理电路芯片，本节简单介绍 MAX1460 芯片。

3.4.1 结构框图

MAX1460 器件的内部结构可分为四大主要组成部分：模拟前端模块、数字处理模块、输出模块以及测试接口部分。其内部功能结构框图如图 3-8 所示。

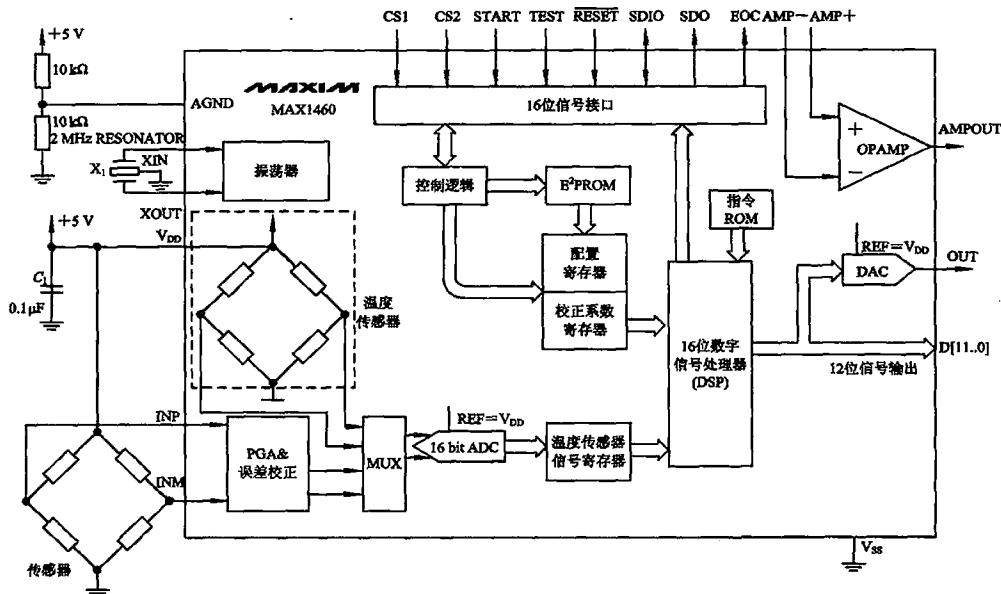


图 3-8 MAX1460 的内部功能结构框图

3.4.2 功能与原理简介

1. 模拟前端模块

MAX1460 的模拟前端模块主要包括：具有误差校正功能的可编程放大器(PGA& 误差校正)，温度传感器，多路开关(MUX)和模/数转换器(ADC)；

(1) 具有误差校正功能的可编程放大器(PGA& 误差校正)。其中的可编程放大器

PGA 可由二位数码($D_1 D_0$)来设置增益，对来自外部配接传感器的信号进行放大；误差校正功能是对输入信号进行粗偏置调整，由三位数码($D_2 D_1 D_0$)控制一个失调校正电路，使失调校正信号落在 ADC 的动态范围内，控制字来自片内的配置寄存器。

(2) 温度传感器。片内的温度传感器也是一个桥路，其四个桥臂电阻由半导体扩散工艺制作，它们的阻值都随温度变化。利用四个桥臂电阻随温度变化其阻值的不一致性产生不平衡输出电压，该电压是温度的函数，可用来测量温度。该温度信号也可用于温度补偿，芯片应尽量靠近被补偿的传感器放置，以使它们处于相同温度的热环境中。

温度传感器桥路输出的信号也经过一个三位数码控制字控制的失调校正电路(图 3-8 中未画出)来进行偏置校正，控制字也来自配置寄存器。

(3) 多路开关(MUX)。将已进行偏置调整设置的片内温度传感器的输出信号与片外传感器的输出信号依次导入模/数转换器。

(4) 16 bit ADC。这是一个 16 位 $\Sigma - \Delta$ 型模/数转换器，其动态范围为双极性 $-V_{DD} \sim +V_{DD}$ 。

当 ADC 的输入信号幅值在 $\pm 85\% V_{DD}$ 范围内时，有很高的模/数转换精度。

2. 引脚功能

MAX1460 的引脚图如图 3-9 所示。

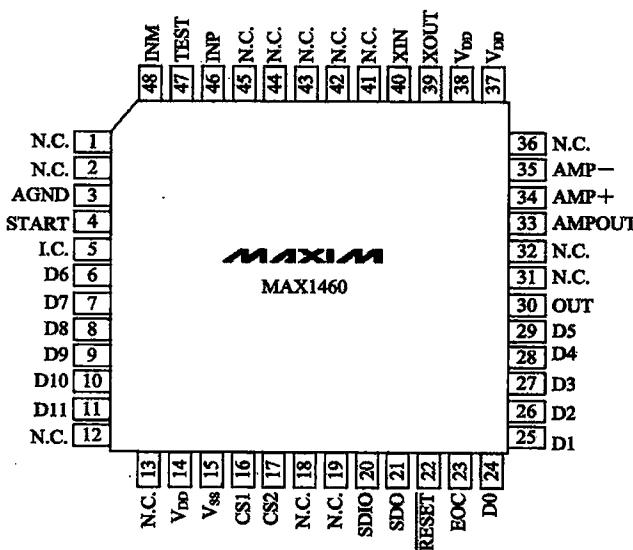


图 3-9 MAX1460 的引脚图

3. 数字处理模块及温度补偿

MAX1460 内有一个低功耗的 16 bit 数字信号处理器，即 DSP，因而可以按照用户确定的校正方程对模拟前端送来的量化后的输入信号 S 和温度信号 T 进行融合计算，对输入信号 S 进行温度补偿后再给予输出。

已知传感器灵敏度 G 定义式为

$$G = \frac{D}{S} \quad (3-10)$$

式中 D 为输出改变量或输出信号; S 为输入改变量或输入信号。

当在全量程温度变化范围内传感器的输入输出特性不变, G 为常量时, 输出信号 $D=G \cdot S$, 可以很方便地求出。如果灵敏度 G 不为常量, 随温度 T 呈非线性变化, 即

$$G = G_{\text{sin}}(1 + G_1 T + G_2 T^2) \quad (3-11)$$

式中 G_{sin} 为初始室温下传感器的灵敏度; G_1 、 G_2 分别为传感器灵敏度的一阶、二阶温度系数, 或称为一阶、二阶灵敏度温度系数。

另外, 在被测量为零时, 传感器桥路有不平衡电压输出, 该输出称为零位输出, 或称失调, 传感器的零位/失调输出电压也将随温度 T 呈非线性变化, 即

$$of = of_0 + of_1 T + of_2 T^2 \quad (3-12)$$

式中 of_0 、 of 分别为传感器在室温、温度 T 时的零位/失调电压值; of_1 、 of_2 分别为一阶、二阶零位/失调温度系数。

将式(3-11)与式(3-12)代入式(3-10), 可得外接传感器的温度补偿计算式, 即校正方程如下:

$$D = D_{\text{sin}}(1 + G_1 T + G_2 T^2)(S + of) + D_{\text{OFF}} \quad (3-13)$$

式中 D_{OFF} 是偏移调整量。

进行温度补偿的计算式(3-13)涉及 of_0 、 of_1 、 of_2 以及 G_1 、 G_2 等多个校正系数, 均由外接的具体传感器系统在不同温度条件下的输入输出特性来获得(详细讨论见第 5 章)并存储在片内的 128 bit E² PROM 中供 DSP 进行温度补偿计算时调用。由于温度补偿计算式采用了多项式拟合算法, 并对高阶分量也进行了补偿, 因此可使传感器精度提高 0.1%。

4. 输出模块

输出模块含 12 位并行数字输出口 $D[11..0]$ 、12 位数/模转换器 DAC、运算放大器 OPAMP 等。

经过校正的数字信号可以直接从 12 位并行数据口输出, 也可以经过 12 位 DAC 的数/模转换后再从模拟输出口输出。除此之外, 用户可以利用运放 OPAMP 来构成一个低通滤波器, 以对 DAC 的输出进行滤波。

5. 测试接口

对 MAX1460 的测试可以通过芯片的测试接口来实现。接口信号主要有片选信号(CS1 和 CS2)、启动信号(START)、测试使能信号(TEST)、复位信号(RESET)、串行数据信号(SDIO, SDO)和转换结束标志信号(EOC)。测试者可以按照给定的操作时序来通过此接口对芯片进行配置、写入校正系数(以确定校正方程)以及读出经 DSP 补偿校正的结果。

3.5 适配变压器式传感器的单片集成调理电路 ——AD698 芯片简介

AD698 是美国 Analog Devices 公司生产的与线位移差分变压器式传感器(Linear Variable Differential Transformer, LVDT)配接的单片集成调理电路芯片。

AD698 与 LVDT 配合，能够精确地将 LVDT 的机械位移转换成单极性或双极性的直流电压。AD698 具有所有必不可少的电路功能，只要增加几个外接无源元件来确定励磁频率和增益，就能把 LVDT 的次级输出信号按比例地转换成直流信号。

AD698 可配接多种形式的 LVDT：变压器次级线圈反向串联式、半桥式、全桥式。本节主要介绍变压器次级线圈反向串联式 LVDT 与 AD698 配接组成机械位移测量系统的有关问题。

3.5.1 LVDT 变压器式传感器简介

1. LVDT 变压器式位移传感器工作原理

LVDT 是一种将机械线位移 $x(t)$ 转换成电压输出的变压器式传感器，由一个初级线圈和两个相同的但反向串接的次级线圈组成。因两个次级线圈完全相同且反接，故磁芯位置居中时次端两个线圈感生电势相等，输出信号 $u_o(t)=0$ ；当磁芯离开平衡位置时改变了初、次级线圈的互感，从而在两个次级线圈中产生的感生电压数值不等，于是输出信号 $u_o(t) \neq 0$ 。输出信号 $u_o(t)$ 是一个调幅波，其幅值反映磁芯偏离中心平衡位置位移 $x(t)$ 的大小，其相位反映磁芯上、下移动的方向，数学表达式如下

$$u_o(t) = K \cdot x(t) \cdot \sin\omega_c(t) \quad (3-14)$$

式中： $\omega_c(t) = 2\pi f_c$ ，称为载波频率，是初级线圈激励信号 $e(t)$ 的角频率，选定后保持恒定； $x(t)$ 称为调制信号，是磁芯偏离平衡中心的位移； K 为常量，与变压器初级线圈有效电阻和自感以及与激励信号 $e(t)$ 的幅值（选定后保持恒定）有关。

2. LVDT 与 AD698 的连接

LVDT 的初级线圈与 AD698 的四个引脚相连：经引脚 2、3（EXC1、EXC2）由参考正弦波激励源 $e(t)$ 向 LVDT 初级线圈提供驱动功率；与引脚 10、11（±BIN）相连，从而通道 B 的输入信号为 LVDT 的初级激励源信号 $e(t)$ 。

LVDT 的次级线圈与 AD698 的两个引脚 12、13（±AIN）相连接，于是通道 A 的输入是 LVDT 的次级输出信号 $u_o(t)$ 。

3.5.2 AD698 的结构框图与工作原理

1. 结构框图

AD698 芯片内部主要由激励源、同步比率解调及滤波放大输出三大模块组成，其结构框图如图 3-10 所示。

2. 工作原理

通过介绍芯片内各模块的功能来介绍 AD698 的工作原理。

1) 激励源模块

该激励源的任务是向外接的 LVDT 传感器初级线圈提供驱动功率。激励源由电压参考、振荡器、放大器构成。

振荡器为多谐振荡器，该多谐振荡器产生一个三角波，并驱动正弦波发生器产生一个低失真的正弦波，正弦波的频率和幅值由外接电阻器 R_1 和外接电容器 C_1 决定。输出频率

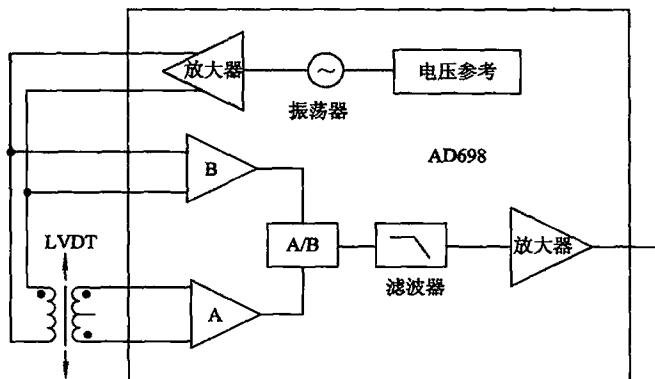


图 3-10 AD698 的结构框图

f_c 在 20 Hz~20 kHz 范围内可调，输出有效幅值在 2~24 V 可调。总谐波失真的典型值是 50 dB。振荡器输出信号经过功率放大器后获得足以驱动 LVDT 传感器的激励信号 $e(t)$ 。

2) 同步比率解调模块

同步比率解调模块由两个性能相同、相互独立的调幅波同步解调通道 A 与 B 以及占空比除法器组成。

(1) 双通道调幅波解调器。同步解调通道的主要环节是由相敏检波器与低通滤波器组成的调幅波解调器。尽管两个通道结构性能相同，但因输入的信号不同，故两个通道的输出信号 A 与 B 完全不同。

B 通道输入信号是 $e(t)$ ，它是频率在 20 Hz~20 kHz 范围的高频载波——等幅正弦波经相敏检波后为一全波整流信号，再经低通滤波后输出信号 B 为纹波很小的直流信号。

A 通道输入信号是 $u_o(t)$ ，它是幅度被 $x(t)$ 调制的调幅波。经相敏检波后信号幅值的包络线正比于检测位移信号 $x(t)$ ，即调制信号，其相位反映被测位移信号 $x(t)$ 的正、负极性。再经低通滤波后输出信号 B，即为待求的被测信号 $x(t)$ 。这时可以说 $x(t)$ 已被从载波 $e(t)$ 上提取出来。传统的 LVDT 位移传感器的调理电路就是这样一个相敏检波器与低通滤波器组成的单通道解调电路。单通道解调器易受电源电压波动或失调漂移影响，采用双通道同步解调模块可实现比值 A/B 与被测信号 $x(t)$ 成正比，可消除共模干扰影响，从而大大提高测量准确度。

(2) 占空比除法器。占空比除法器是将输入信号 A、B 执行比值运算的计算器。它输出一矩形脉冲波，矩形脉冲波的占空比与输入信号 A、B 的比值 A/B 成正比，故是一种脉宽调制信号。当 $A=B$ 时，占空比为 100%，这个矩形脉冲波也可作为输出信号，用于要求脉宽调制信号作为输出信号的场合。

3) 滤波、放大输出模块

该模块进一步对除法器输出信号进行滤波与放大。滤波器除了滤除激励源与高频噪声的干扰外，还将信号平滑为直流输出。放大器是一个电流放大器，占空比除法器输出的脉宽调制信号先被转换为一参考电流调制信号 i ，即 i 与矩形脉冲波的占空比成正比，也即与比值 A/B 成正比：

$$i = I_{ref} \cdot A/B \quad (3-15)$$

式中 $I_{ref} = 500 \mu A$, 是标定值为 $500 \mu A$ 的参考电流。

电流放大器将调制了的电流信号 i 转换为电压输出信号 U_{OUT} :

$$U_{OUT} = i \cdot R = I_{ref} \cdot A/B \cdot R_2 \quad (3-16)$$

于是, AD698 最终输出电压 U_{OUT} 与比值 A/B 成正比, 也即与被测信号 $x(t)$ 成正比。

3.5.3 主要参数设置与使用方法

以最常用的双电源对称供电(±15 V)为例来说明主要参数的设置。

1. 确定激励信号 $e(t)$ 的频率 f_{exc}

激励信号 $e(t)$ 的频率也即振荡器的振荡频率, 是根据系统的带宽来确定的。

(1) AD698 LVDT 位置测量子系统带宽 f_{SUB} 的确定与设置。带宽 f_{SUB} 根据被测机械位移信号 $x(t)$ 的最高频率来决定, 由外接电容 C_2 、 C_3 、 C_4 来设置。

当 AD698 LVDT 位置测量子系统带宽 f_{SUB} 确定后, 可由下式来设置 AD698 外接电容 C_2 、 C_3 、 C_4 的值:

$$C_2 = C_3 = C_4 = 10^{-4}/f_{SUB} \quad (\text{法拉}) \quad (3-17)$$

C_2 、 C_3 分别接在引脚 8、9 与引脚 16、17 之间, 是解调器通道 B、A 的滤波电容。

(2) LVDT 激励信号 $e(t)$ 频率 f_{exc} 最小值的确定与设置。 f_{exc} 的最小值应满足下式关系:

$$f_{exc} = 10f_{SUB} \quad (3-18)$$

当 $f_{SUB} = 250 \text{ Hz}$ 时, $f_{exc} \geq 2.5 \text{ kHz}$ 。

由外接电容 C_1 来设置所期望的激励频率 f_{exc} , C_1 的数值根据所要求的 f_{exc} 按下式来确定:

$$C_1 = \frac{35(\mu\text{F} \cdot \text{Hz})}{f_{exc}} \quad (\text{Hz}) \quad (3-19)$$

当 $f_{exc} = 2.5 \text{ Hz}$ 时, $C_1 = 0.014 \mu\text{F}$ 。

C_1 接于 AD698 引脚 6(FREQ1) 与引脚 7(FREQ2) 之间。

2. 确定激励信号的幅值

这样来确定激励信号的幅值: 使得当 LVDT 处在满量程位置时初级信号有效值(rms)为 1.0 V 至 3.5 V 范围, 同时次级信号有效值能在 0.25 V 至 3.5 V 范围。这是线性度与无噪声灵敏度的最佳值。该最佳值与 LVDT 的灵敏度 S 及传感器中变压器的变比 VTR 有关。

(1) LVDT 在满量程位置时变压器的变比 VTR(Voltage Transformation Ratio)。变比 VTR 由下式确定:

$$VTR = \frac{U_{sec}(\text{输出信号})}{U_{pri}(\text{输入信号})} = \frac{A}{B} \quad (3-20)$$

式中, U_{pri} 、 U_{sec} 分别是 LVDT 传感器在满量程位置时, 变压器的初级输入与次级输出电压信号。

(2) LVDT 变压器式位移传感器的灵敏度 S 。LVDT 的灵敏度 S , 在数值上表示当输入单位位移量时, 在初级输入信号为 1 V 条件下传感器次级输出信号的电压值(V)。 S 的表述式如下:

$$S = \frac{U_{\text{sec}}}{U_{\text{pri}}} \frac{1}{d} = \frac{\text{VTR}}{d} \quad (3-21)$$

式中, VTR 为 LVDT 中变压器的变比, 可由实验测得; d 为 LVDT 传感器满量程输入, 即磁心偏离零位的最大行程长度。

灵敏度 S 可由传感器使用手册资料获得, 如 E100 型传感器 $S=2.4 \text{ mV/V/mil}$ 。

(3) LVDT 最佳激励电压 U_{exc} 的确定。将式(3-21)代入式(3-20)中可得最佳激励电压 U_{exc} :

$$U_{\text{exc}} = U_{\text{sec}}/\text{VTR} = \frac{U_{\text{sec}}}{S \cdot d} \quad (3-22)$$

设选用 E100 型 LVDT 传感器, 可知 $S=2.4 \text{ mV/V/mil}$, 当满度位移输入值 $d=\pm 0.1 \text{ inch} (\pm 100 \text{ mil})$ 时, VTR 变比值为

$$\text{VTR} = S \cdot d = 2.4 \text{ mV/V/mil} \times 100 \text{ mil} = 0.24$$

故最佳激励电压 $U_{\text{exc}}=U_{\text{pri}}=3.5 \text{ V(rms)}$ 时, 最大次级输出电压将是

$$U_{\text{sec}} = U_{\text{exc}} \cdot \text{VTR} = 3.5 \times 0.24 = 0.84 \text{ V(rms)}$$

上述数值 $U_{\text{sec}}=0.84 \text{ V}$, 是在可接受的($0.25 \sim 3.5$) U_{max} 范围之内的。

(4) 激励电压 U_{exc} 的设置。根据需要的激励信号的幅值 $E_m=U_{\text{exc}}$ 来选择外接电阻 R_1 。
 U_{exc} 与 R_1 的关系为

$$12 \text{ V} \leqslant U_{\text{exc}} \leqslant 24 \text{ V}, \quad 0.1 \text{ k}\Omega \leqslant R_1 \leqslant 1 \text{ k}\Omega$$

$$5 \text{ V} \leqslant U_{\text{exc}} \leqslant 12 \text{ V}, \quad 1 \text{ k}\Omega \leqslant R_1 \leqslant 10 \text{ k}\Omega$$

$$2.5 \text{ V} \leqslant U_{\text{exc}} \leqslant 5 \text{ V}, \quad 10 \text{ k}\Omega \leqslant R_1 \leqslant 100 \text{ k}\Omega$$

3. 设置满量程输出电压 U_{out}

通过接入电阻 R_2 设定输出级电流放大器的增益来设置满量程输出电压 U_{out} , U_{out} 为 AD698 的最终电压输出值, 由 20 脚相对 21 脚量得。根据式(3-20)、式(3-21)可知变比 VTR 可表述为:

$$\text{VTR} = S \cdot d = \frac{A}{B} (\text{V/V}) \quad (3-23)$$

将式(3-16)变换形式可得

$$R_2 = \frac{U_{\text{out}}}{A/B \cdot I_{\text{ref}}} = \frac{U_{\text{out}}}{S \cdot d \cdot I_{\text{ref}}} \quad (3-24)$$

式中, R_2 为在引脚 20(SIG OUT)与引脚 19(FEEDBACK)之间接的电阻, 用来调节输出级电流放大器的增益; $I_{\text{ref}}=500 \mu\text{A}$, 是标定值为 $500 \mu\text{A}$ 的参考电流。

由式(3-24)可见, 当传感器 LVDT 的特性参数 S 、 d 确定后, 就可设定期望的满量程输出电压 U_{out} , 按式(3-24)计算出所需的增益电阻 R_2 。

例如: $U_{\text{out}}=\pm 10 \text{ V}$, 即满量程输出电压范围为 20 V ; $d=\pm 0.1 \text{ inch}$, 即传感器满量程输入位移范围为 $0.2 \text{ inch}=200 \text{ mil}$; $S=2.4 \text{ mV/V/mil}$, $I_{\text{ref}}=500 \times 10^{-6} \text{ A}$ 。将上述值代入式(3-24), 则得 $R_2=83.3 \text{ k}\Omega$ 。

4. AD698 外围电路连接图

单、双电源供电的 AD698 外围电路连接如图 3-11(a)、(b) 所示, 当不进行失调补偿时, R_3 与 R_4 开路。

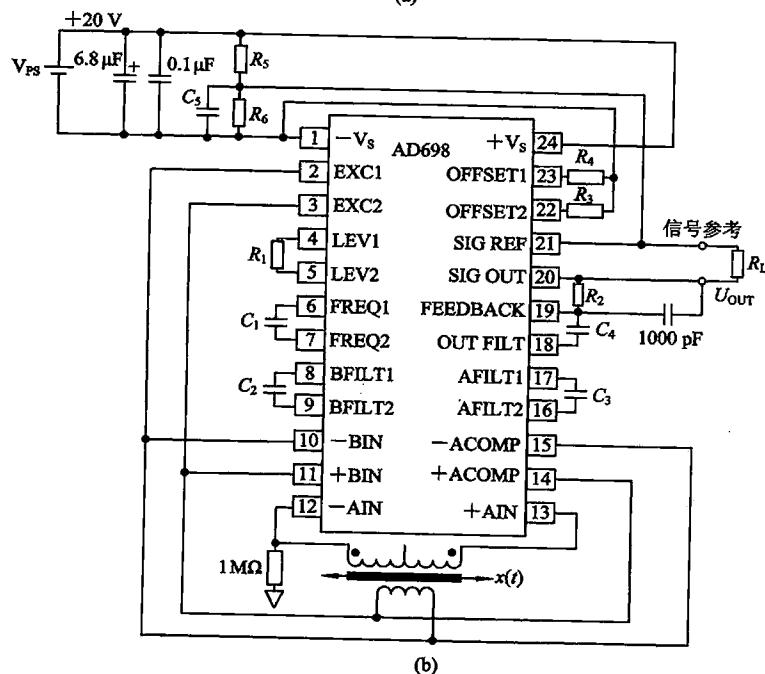
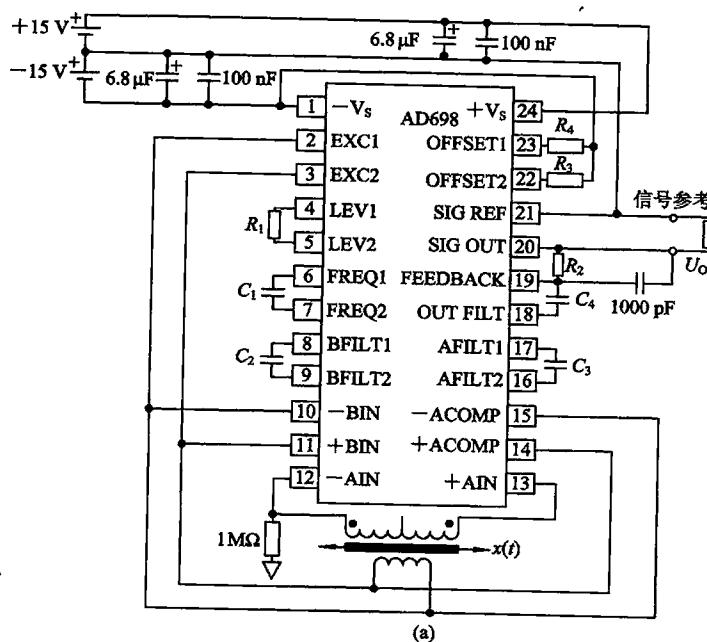


图 3-11 AD698 外围电路连接图

3.6 适配电容式传感器的集成调理电路——CAV414 芯片

电容式传感器具有很广泛的应用范围。德国 AMG 公司生产了一系列用于电容信号的转换、放大以及标准化输出的集成化调理电路，如 CAV404、CAV414 和 CAV424。该调理电路的输出电压可以进一步直接用来显示电容传感器的被测参量，可以与 A/D 转换器或微处理器相连构成电容式智能传感器系统，可以与一些工业标准化输出电路相连构成二线、三线制 0/4~20 mA 的变送器。

CAV414 与 CAV424 的结构原理基本相同，CAV424 还在片内提供了一个温度传感器可作为温度补偿用。本节介绍集成调理电路芯片 CAV414，它可检测的电容初值在 10 pF~1 nF 之间，电容的变化范围在 10.5 pF~2 nF 之间。

3.6.1 CAV414 的结构框图

CAV414 主要由基准振荡器、电压/电流基准模块、积分器模块、滤波器模块以及输出放大级模块等部分组成，结构框图如图 3-12 所示。

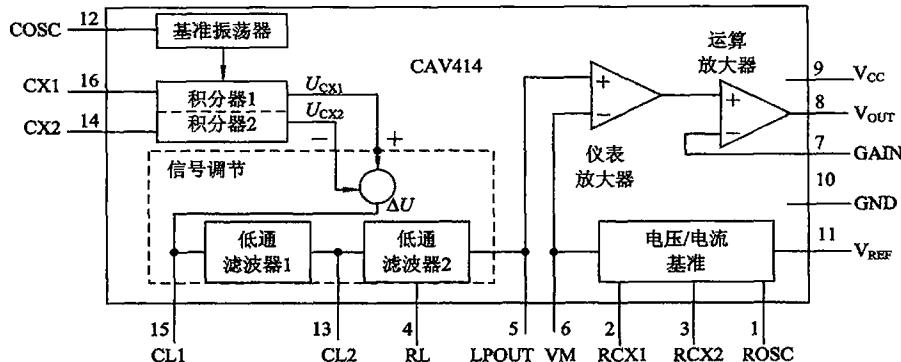


图 3-12 CAV414 的结构框图

3.6.2 工作原理与引脚连接

下面通过各模块的功能介绍来说明芯片的工作原理。

1. 积分器模块

积分器模块由两个积分器组成。它的功能是将电容量之差 $\Delta C = C_{x2} - C_{x1}$ 转换为一个相应的电压之差 ΔU 。 C_{x1} 与 C_{x2} 分别是积分器 1 与 2 的积分电容值。

16、14 脚 (C_{x1} 、 C_{x2} 端) 接差动电容传感器，或 C_{x1} 接参考电容， C_{x2} 接测量电容器；初始电容值为 10 pF~1 nF，可测 C_{x2} 的变化范围为 10.5 pF~2 nF。

2、3 脚 ($RCX1$ 、 $RCX2$ 端) 与地之间接阻值相同的电阻 $R_{CX1} = R_{CX2} = R_{Cx}$ 。

在 $T/2$ 时间内，两个积分器均同时对参考电压 U_M 进行积分，积分电容 C_{x1} 、 C_{x2} 两端充电电压 U_{cx1} 、 U_{cx2} 与充电电荷 Q_{cx1} 、 Q_{cx2} 成正比，与电容量 C_{x1} 、 C_{x2} 成反比：

$$U_{cx1} = \frac{Q_{cx1}}{C_{x1}} = \frac{\int_0^{\frac{T}{2}} I_{cx1} \cdot dt}{C_{x1}}$$

$$U_{cx2} = \frac{Q_{cx2}}{C_{x2}} = \frac{\int_0^{\frac{T}{2}} I_{cx2} \cdot dt}{C_{x2}}$$

其中, $I_{cx1} = U_M / R_{cx1}$, $I_{cx2} = U_M / R_{cx2}$, 为积分器的充电电流; 当 $R_{cx1} = R_{cx2} = R_{cx}$ 时, $I_{cx1} = I_{cx2}$; 当充电结束到 $T/2$ 时刻时, 充电电压达到最大值, 分别为

$$U_{cx1} = \frac{T}{2} \frac{U_M}{R_{cx} C_{x1}}, \quad U_{cx2} = \frac{T}{2} \frac{U_M}{R_{cx} C_{x2}}$$

当没有被测量输入时, $C_{x1} = C_{x2}$, 则 $U_{cx1} = U_{cx2}$, 二积分器充电电压最大值相等其差应为零; 当有被测量输入时, $C_{x1} \neq C_{x2}$, 积分器充电电压最大值之差反映被测量大小。

积分电容两端的电压 U_{cx1} 、 U_{cx2} 也分别是积分器 1、2 的输出电压, 故积分器模块完成了将电容量之差 $\Delta C = C_{x2} - C_{x1}$ 转换为相应电压之差 $\Delta U = U_{cx1} - U_{cx2}$ 。

2. 基准振荡器与电压/电流基准模块

(1) 基准振荡器提供一个基准振荡频率 $f_{osc} = 1/T$, 控制两个积分器在 $T/2$ 时间内同步充电, 在 $T/2 \sim 3T/2$ 时间内同步放电。

在引脚 1ROSC 与引脚 10GND(地)之间接入电阻 R_{osc} 。

在引脚 12COSC 与引脚 10GND(地)间接入电容 C_{osc} , 其值 $C_{osc} = 1.6C_{x1}$; 基准振荡频率 f_{osc} 由下式确定:

$$f_{osc} = \frac{U_M}{2 \cdot \Delta U_{osc} \cdot R_{osc} \cdot C_{osc}} \quad (3-25)$$

式中, ΔU_{osc} 为基准振荡器输出的波峰与波谷值之差, 它由内部电阻来确定, 通常为 2.1 V; $C_{osc} = C_{osc} + \Delta C_{osc}$, C_{osc} 为 12 脚的外接电容值, ΔC_{osc} 为寄生电容量。

(2) 电压/电流基准提供积分电压 U_M , 且 $I_{cx} = U_M / R_{cx}$ 。

3. 滤波器模块

滤波器模块由两级低通滤波器组成。它们的 3 dB 截止频率 f_{c1} 、 f_{c2} 由内部电阻 $R_{o1} = R_{o2}$ (典型值 20 kΩ) 及外接电容 C_{L1} 、 C_{L2} 来确定。

在引脚 15CL1 端与 10 脚 GND 地间接电容 C_{L1} ; 在引脚 13CL2 端与 10 脚 GND 地间接电容 C_{L2} , 通常取 $C_{L1} = C_{L2} = C_L = 200C_{x1}$ 。

因为低通滤波器的 3 dB 截止频率 $f_c = 1/(2\pi \cdot R_o \cdot C_L)$, 于是根据需要的 f_c 可确定外接电容 C_L 的值:

$$C_L = \frac{1}{2\pi \cdot R_o \cdot f_c} \quad (3-26)$$

滤波器模块的任务是要衰减基准振荡频率 f_{osc} 分量, 但要使被测信号的频率 f 分量通过, 故 f_c 、 f_{osc} 与 f 三个频率应满足如下关系

$$f < f_c \ll f_{osc} \quad (3-27)$$

由滤波器输出的电压信号 U_{LPOUT} 是反映被测量变化的低频信号, 可由引脚 5LPOUT 端输出。

4. 输出放大级模块

输出放大级由两级放大器构成：仪表放大器与运算放大器。

(1) 仪表放大器为高输入阻抗双端浮地差动放大器，输入两个信号：一是低通滤波器的输出信号 U_{LPOUT} ，该信号是反映被测量变化的信号，但含有直流分量 U_M ；另一输入信号是基准电压 U_M 。仪表放大器将双端输入信号转换为单端输出并消除了直流分量 U_M 。

(2) 运算放大器应构成一个同相放大器。放大器的输出端 V_{OUT} (引脚 8) 与 GND (10 脚 (地)) 之间接入两个相串联的电阻 R_1 、 R_2 。电阻 R_1 与 R_2 的连接点再与运算放大器的反相输入端，也即 7 脚 (GAIN) 端相连接。改变 R_2 的阻抗值可调节该同相放大器的增益。

3.6.3 [示例 3-2] 基于 CAV414 的电容式变送器的构成

CAV414 与集成电路 AM402 相配合组成二线制 4~20 mA、0~10 V 标准输出的电容式变送器。

1. 具体连接

CAV414 的引脚 5、6 分别与 AM402 的输入端 9 脚 IN+ 端、8 脚 IN- 端相连。电容式变送器的芯片与外围元件连接图如图 3-13 所示。

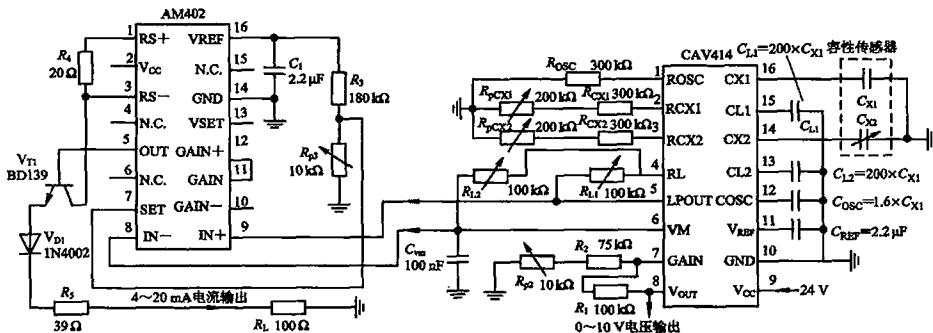


图 3-13 CAV414 与 AM402 组成的电容变送器芯片与外围元件连接图

2. 调试步骤

- (1) 无被测量输入时，测量电容传感器的电容初始值 C_{X1} ，并令 $C_{X1} = C_{X2}$ 。
- (2) 计算出振荡电容 C_{OSC} ($C_{OSC} = 1.6C_{X1}$)。
- (3) 使两个低通滤波电容的值相同 ($C_L = 1/(2\pi \cdot R_0 \cdot f_c)$)。
- (4) 通过 R_{pCX1} 、 R_{pCX2} 进行调零； R_{pCX1} 与 R_{pCX2} 均为 R_{CX1} 与 R_{CX2} 的微调部分。
- (5) 调节 R_{p2} 来改变输出电压的大小，使从 8 脚 V_{OUT} 端输出的电压最大为 10 V。
- (6) 调节电阻 R_4 ，使输出电流在 4~20 mA 间变化。 R_4 接于 AM402 芯片的 1、3 脚之间。

3.7 数据采集系统——单片集成接口芯片 ADuC812

3.7.1 简介

ADuC812 是一个单片全集成的 12 位数据采集系统。在一个芯片上大规模集成了片内

温度传感器、高性能多通道输入输出、非易失性快擦除存储器、以 8051 为内核的控制系统、电源管理系统及通信接口电路，具有 52 引脚方形扁平塑封和 56 引脚芯片级封装两种型号。

由于 ADuC812 带有片内温度传感器，而且数据采集和转换速度快，因此广泛用于可自校准或自调节的高智能化传感器系统，也可用于仪器仪表的数据采集和校正。另外，ADuC812 具有静态 CPU 标准、等待和掉电三种工作模式，其功耗很低，故特别适用于手提式 PC，便携式仪器仪表、监视器以及其他由电池供电的系统。200 kHz 的数据采集速度使其能够方便地用于需要瞬态捕获的信息数据采集系统。可以预测，在不远的将来，ADuC812 数据采集系统亦将大量应用于 DAS 以及各种场合的通信系统。

3.7.2 结构框图与组成模块的功能

数据采集系统芯片 ADuC812 的结构框图如图 3-14 所示。由 ADC 模/数转换器、DAC 数/模转换器、8 位可编程微控制器模块以及其他外围器件等模块组成。其功能简述如下。

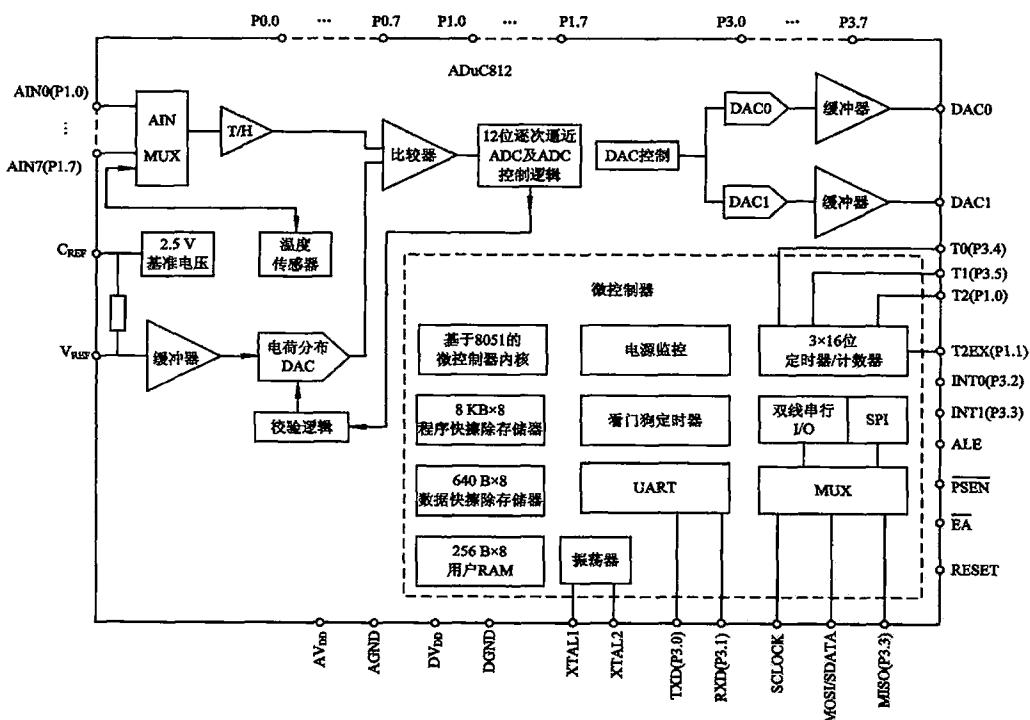


图 3-14 数据采集系统芯片 ADuC812 结构框图

1. ADC 模/数转换器

ADC 模块含有单电源供电的 12 位逐次逼近型快速模/数转换器、多路模拟切换开关

(MUX)、2.5 V 基准电压及温度传感器等，可实现多路输入、采样保持、模/数转换、自校验等多种功能。

(1) 模/数转换器。ADC 模块中的模/数转换器是由单电源供电的 12 位逐次逼近型快速模/数转换器，可以 200 kHz 的速率实现 12 位二进制快速转换，可转换 $0 \sim U_{ref}$ 的模拟信号。当 $U_{ref} = 2.5$ V 时，ADC 输出数码的最低位(量化值) $1\text{ LSB} = U_{ref}/2^{12} = 2.5\text{ V}/4096 = 0.61\text{ mV}$ 。另外，片内提供了高精度、低漂移的 2.5 V 基准电压。单步或连续转换方式可由软件或给外引脚加一个转换信号来启动，定时器 2 也可为模/数转换产生重复触发脉冲。模/数转换器也可设定在 DMA 模式下运行。在这种情况下，ADC 的连续转换、采样操作可与微控制器并行工作。模/数转换器还包括自校验和系统校验两种选择，以确保在长时间、高温下精确工作。从片内带隙基准源输出的与绝对温度成比例的电压，同样也能通过前端模/数转换器的多路转换开关取出，从而实现对环境温度的测量。

(2) 模拟切换开关 AIN MUX 可依次切换八个通道的模拟输入信号，外部信号输入引脚为 P1.0~P1.7。

(3) 温度传感器采用片内带隙温度传感器，可用来测量片外环境温度，它输出与绝对温度成比例的电压。信号由片内连接经过前端的多路切换开关 MUX 输入，从而实现对环境温度的测量。

(4) 2.5 V 基准电压是片内提供的高精度、低漂移的基准电压，以实现模/数转换器自校验/准和全系统校验/准，两种校验/准功能可选择。从 8 脚 V_{REF} 端接入电压值在 2.3 V~ U_{pp} 的外基准电压源也可实现模/数转换器的自校验；内基准源可接到 8 脚 V_{REF} 端，再经输入通道由多路切换开关导入，可实现全系统校验/准。

ADC 模块的操作，如控制采集和转换时间、控制输入通道选择以及用户自校验/准与系统校验/准选择等，均由微控制器中的三个特殊功能寄存器(SFR)控制。

2. DAC 数/模转换器

ADuC812 器件中有两个 12 位电压输出型数/模转换器：DAC0 与 DAC1，分别经各自的缓冲器接至输出端 DAC0 与 DAC1(9 脚、10 脚)，其可编程输出范围为 $0 \sim 2.5$ V 或 $0 \sim U_{pp}$ 。

3. 微控制器的基本模块

同所有的与 8051 微处理器相关的器件一样，ADuC812 器件中的微控制器的基本模块包含：基于 8051 的微控制器内核、独立的程序存储器与数据存储器，以及可以供用户使用的存储器等模块，构成八位可编程微控制器。

- (1) 基于 8051 的微控制器内核，具有与 8051 兼容的指令系统。
- (2) 存储器。ADuC812 器件片内微控制器内嵌入了两个快擦除存储器和一个 RAM。
 - 8 KB×8 程序快擦除存储器：为程序执行提供了方便，可由程序编辑器编程，也可在线编程。
 - 640 B×8 数据快擦除存储器：是供用户使用的片内 640 字节数据存储器。用户可用它作通用非易失性高速缓存区。可通过 6 个一组的特殊功能寄存器(SFR)访问此区。

上述快擦除存储器具有 E²PROM 灵活的在线可编程特点，并兼备 EPROM 的密度特

性，从而为用户提供了非易失性在线可编程代码及数据存储空间，具有高密度、低成本的特点。

- 256 B×8 用户 RAM：是片内数据静态存储器。其低 128 字节映像与 MCS-51 相同，高 128 字节映像为特殊功能寄存器(SFR)。

(3) 微控制器外围器件包括：并行、串行输入/输出口，定时器/计数器，电源监控以及看门狗定时器等电路模块。

① 并行输入/输出口：采用四个通用 8 位数据口(P0~P3)与外部器件交换数据，P1 口(P1.0~P1.7，引脚为 1~4、11~14)为单向输入口，控制八路被测信号输入通道的通断与通断的时间间隔。P0、P2 及 P3 口为双向口，接口的读写、输入/输出操作是通过与读写对应的特殊功能寄存器(SFR)进行的。所有口都含有一组输出锁存器、输出驱动器和输入缓冲器。

② 串行输入/输出口：ADuC812 具有三种串行输入/输出口，即通用串行异步接口、串行外围接口(SPI)和双线串行接口。

- 通用串行异步接口(UART)是全双工的，通过 RXD(P3.1)和 TXD(P3.0)，即 17、16 脚，实现同时收发信息。

• 串行外围接口(SPI)是一种工业标准同步串行口，它可同步传输并同时接收 8 位数据。该系统可分为为主和从两种操作，并由标准的 4 引脚组成：

MISO(P3.3, 19 脚)：主输入、从输出；

MOSI(P3.2, 18 脚)：主输出、从输入；

SCLOCK(26 脚)：串行时钟；

SS(13 脚)：从选择。

③ 定时器/计数器。ADuC812 采用了三组 16 位定时器/计数器，即定时器 0、1 和 2。片内含有定时器/计数器硬件，因此处理器就无需用软件来实现定时。定时器/计数器由双 8 位寄存器构成。

• 定时器功能：每个机器周期都被增 1，即它对机器周期进行计数，由于单个机器周期由 12 个振荡周期组成，则最大计数频率为振荡器频率的 1/12。

• 计数器功能：在 T0(P3.4)、T1(P3.5)或 T2(P1.0)外部输入脚从高电平 1 到低电平 0 的下降沿过程中增 1。

(4) 片内监控。ADuC812 具有片内监控功能，可在程序出现混乱或外系统错误时，使代码和数据错误最小。监控功能完全由特殊功能寄存器 SFR 控制。

(5) 看门狗定时器。当由于程序错误、电噪声等原因使 ADuC812 进入不稳定状态时，该看门狗定时器可在适当的时间内使系统复位。其复位间隔可在 16~2048 ms 内预置设定，也可设定看门狗永久性不起作用。

(6) 电源监控。当电源电压较低时，此监控功能保证用户保存工作寄存器内容，以避免可能出现的数据错误，同时，仅在“安全”电源时才执行代码。电源监控用于防止寄生噪声误触发中断，具有低功耗的使用三种电源工作模式的灵活控制功能：正常标准工作模式、空闲等待模式和掉电工作模式。

3.7.3 主要参数特点

1. 模拟输入/输出

- 8 通道, 12 位模数转换器(ADC);
- 具有片内基准电源与控制逻辑执行自校验操作, 以实现高精度测量;
- 200 kHz 高速采样;
- 片内 DMA 控制器可用于自动采集;
- 双 12 位电压输出型数/模转换器(DAC);
- 片内带有温度传感器。

2. 存储器

- 8 KB×8 片内程序快擦除存储器;
- 640 B×8 片内数据快擦除存储器;
- 256 B×8 片内用户 RAM(静态存储器);
- 64 KB 外部程序可寻址空间;
- 16 MB 外部数据可寻址空间。

3. 以 8051 为基础的内核

- 指令系统与 8051 兼容;
- 12 MHz 额定操作;
- 3 个 16 位定时器/计数器;
- 24 条可编程输入/输出线;
- 8 条可编程输入线;
- 9 个中断源, 2 个优先级。

4. 电源

- 3 V 和 5 V 两种规格电压工作的器件可供选择;
- 标准、等待和掉电三种工作模式, 控制方便灵活, 保证低功耗应用外围结构;
- 通用异步串行(UART)输入/输出口;
- 双线(与 I²C 兼容)和 SPI 串行接口;
- 看门狗定时器和电源监控功能。

3.7.4 应用举例——温度的测量与显示系统

1. 功能要求

- (1) 采用片内温度传感器测量环境温度, 外接热敏电阻温度传感器测量 8 个关键点的温度。
- (2) 温度测量值为高于摄氏零度的正值, 温度显示器分度值为 1℃。

2. 测温系统的实现

1) 系统的组成框图

温度测量与显示系统的组成框图如图 3-15 所示, 该系统由 8 个热敏电阻温度传感

器、ADuC812 及 MCS14495 构成的显示器控制接口电路组成。

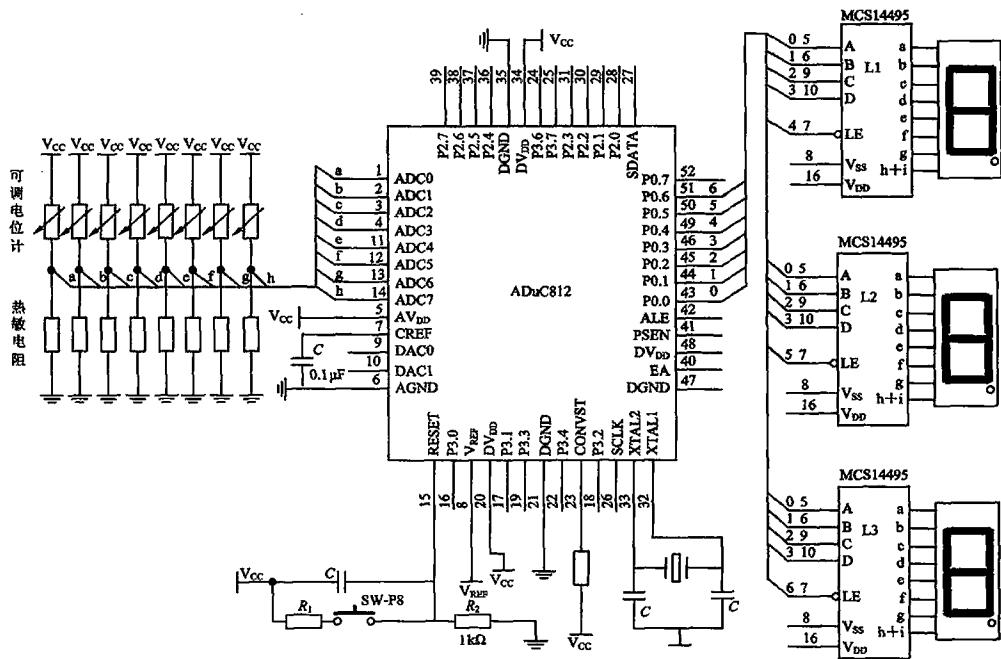


图 3-15 温度测量与显示系统的组成框图

2) 简要说明

(1) 温度传感器。片外温度传感器为热敏电阻，经可调电位分压后接至 ADuC812 的 P1 口(P1.0~P1.7)，对应的引脚为 1~4、11~14。选取高精度温度计作为标定基准，通过调节电位计改变分压值，使显示器显示的温度值与基准温度值相同。片内的温度传感器测量本地环境温度。

(2) ADuC812 器件的设置。ADuC812 的转换开始端 CONVST(引脚 23)，经电阻接电源 V_{cc} ，设置为高电平，由软件触发开始 A/D 转换。转换结果输出是在前一次转换结束与下次转换之前执行的。

XTAL1 与 XTAL2 端(引脚 32 与 33)，接 12 MHz 晶振。这就是 8052 内核的主时钟频率。

RESET 复位端(15 脚)接外界复位开关 SW。

CREF 参考电容端(7 脚)与模拟地 AGND 端(6 脚)之间接参考电容 $C=0.1 \mu F$ 。

DV_{DD}数字电源端(20 脚、34 脚)接数字电源 V_{cc} 。

DGND 数字地端(21 脚、35 脚)接数字地。

AV_{DD}模拟电源端(5 脚)接模拟电源 V_{cc} 。

(3) 显示器控制接口电路。用三片 MCS14495 组成对多位静态 LED 显示器的控制接口电路。它们与 ADuC812 的输出口 P0.0~P0.6(43 脚~51 脚)相连接，输出的 7 位二进

制数码对应被测温度范围为 $0 \sim 2^7$ ($0 \sim 128^\circ\text{C}$)；分别由三片 MCS14495，即 L1、L2、L3 控制显示个位、十位、百位温度值。

整个系统采用 C51 语言编程。注意：厂家提供一个上电初始化子程序，位于 ROM 空间，因此，从串行口加载用户程序或调用应用程序时，是从地址 FF00H 开始的，而不是 0000H。

3.8 温度传感器系统——全系统单片集成芯片 MAX6625

MAX6625 是美国 MAXIM 公司生产的一种新型智能温度传感器系统芯片。它将温度传感器、9 位 A/D 转换器、可编程温度界限报警和 I²C 总线串行接口等集成在同一个芯片中，在风扇控制、温度报警、系统温度测量及工业设置中得到广泛应用。

3.8.1 MAX6625 结构框图及引脚排列

MAX6625 型智能数字温度传感器系统芯片结构框图如图 3-16 所示，由带隙基准电压源及温度传感器、9 位模/数转换器、5 个控制寄存器、设置温度点比较器、故障排队计数器以及串行总线接口电路等部分组成。其引脚说明如下：

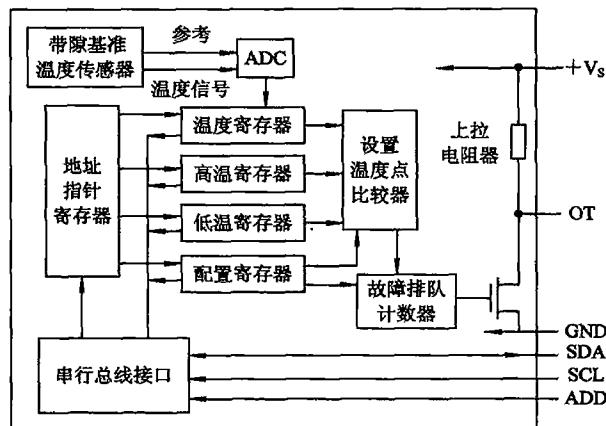


图 3-16 MAX6625 的结构框图

- SDA——I²C 总线兼容的串行双向数据线；
- SCL——I²C 总线兼容的串行时钟线；
- ADD——I²C 总线兼容的地址设置端；
- OT——温度报警输出端；
- Vs——电源端($+3 \sim 5.5\text{ V}$)；
- GND——地。

MAX6625 采用 SOT23-6 小型化封装，其引脚排列如图 3-17 所示。

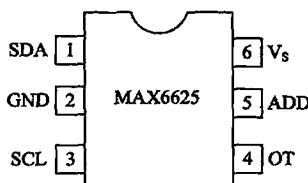


图 3-17 MAX6625 的引脚排列

1. 五个控制寄存器

(1) 地址指针寄存器。地址指针寄存器是一个 8 位寄存器，用于选择其他 4 个寄存器，对所选寄存器进行写入或读取。

通过 I²C 总线接到来自主控制器的指令地址数据，地址指针寄存器的状态即形成的地址码指向所选择的寄存器，从而可对该选中的寄存器进行读/写操作。地址指针寄存器的状态、形成的地址码数据格式与指向的寄存器如表 3-1 所示。

表 3-1 地址指针寄存器的状态、地址码数据格式及指向的寄存器

地址指针寄存器状态								地址码数据	指向的寄存器
D7	D6	D5	D4	D3	D2	D1	D0		
0	0	0	0	0	0	0	0	00000000	温度寄存器
						0	1	00000001	配置寄存器
						1	0	00000010	低温寄存器
						1	1	00000011	高温寄存器

(2) 温度寄存器。温度寄存器是一个 16 位寄存器，存放来自 A/D 转换器(ADC)最新的温度转换数据，是摄氏温度的补码，最低有效位 1LSB 表示 0.5°C，D15 为符号位，D14~D7 共 8 位为有效数据位，D6~D0 为 0。因此，可寄存的温度上限为 $2^8 \times 1\text{LSB} = 256 \times 0.5^\circ\text{C}$ ，为 128°C。

(3) 高/低温寄存器。高/低温寄存器也是 16 位寄存器，其数据格式与温度寄存器相同。

(4) 配置寄存器。配置寄存器是一个 8 位寄存器，其状态、数据格式和作用如表 3-2 所示。

表 3-2 配置寄存器的状态、数据格式及作用

D7	D6	D5	D4	D3	D2	D1	D0
0	0	0	故障排队深度选择位	OT 报警极性选择位	比较/中断模式选择位	掉电模式控制位	

(1) 故障排队深度选择位为 D4、D3：

[D4 D3]=[0 0]，故障深度为 1；

[D4 D3]=[0 1]，故障深度为 2；

[D4 D3]=[1 0]，故障深度为 4；

[D4 D3]=[1 1]，故障深度为 6。

(2) OT 报警极性选择位为 D2：

[D2]=[0]，低电平有效；

[D2]=[1]，高电平有效。

(3) 比较/中断模式选择位 D1：

[D1]=[0]，比较模式；

[D1]=[1]，中断模式。

(4) 掉电模式控制位为 D0：

[D0]=[0]，正常工作模式；

[D0]=[1]，掉电模式。

2. 温度传感器

温度传感器是由片内基准电压源供电的片内温度传感器。

3. 9 位 A/D 转换器(ADC)

ADC 将片内温度传感器的输出转换为数字量。其 D8 位为符号位，其余 8 位(D7~D0)表示被测温度值。量化值 $q=128/2^8 \approx 0.5^\circ\text{C}$ ，测温上限值为 128°C ，对应的温度分辨力为 0.5°C 。

设置温度点比较器和故障排队计数器执行温度超限报警功能，串行总线接口电路通过 I²C 串行总线与接口电路实现主控制器与 MAX6625 之间的通信。

3.8.2 主要功能与技术指标

MAX6625 的主要功能与技术指标如下：

(1) 由片内传感器进行环境对象的温度监测，由 9 位 A/D 转换器转换为 8 位数码及一位符号位表示温度值。测温范围为：

$-55^\circ\text{C} \sim +125^\circ\text{C}$ ，分辨力可达 0.5°C ；

$-40^\circ\text{C} \sim +80^\circ\text{C}$ 、 $-55^\circ\text{C} \sim +125^\circ\text{C}$ 对应的最小量程分别为 $\leq \pm 3^\circ\text{C}$ 与 $\leq \pm 4^\circ\text{C}$ 。

(2) MAX6625 为从器件，支持 I²C 总线的字节和字的读/写操作命令。由主控制器通过 I²C 总线接口进行编程，可实现温度测量值的读取与温度越限报警功能或控温。

(3) 带 I²C 串行接口，串行时钟频率范围在 $0 \sim 400 \text{ kHz}$ 内可调。

(4) 完成一次温度/数据转换约需 133 ms 。

(5) 电压范围： $+3 \sim +5.5 \text{ V}$ ；有静态工作电流约为 1 mA 或 $1 \mu\text{A}$ ，分别对应正常工作模式与掉电休眠两种可选模式。

(6) 在一条 I²C 串行总线上可挂接 4 个 MAX6625 器件，实现 1~4 个单点或多点温度测量、越限报警及温度控制。

3.8.3 工作原理

1. 多点温度测控与报警

当一条 I²C 串行总线挂接多个 MAX6625 时，可实现多点温度测控与越限报警，最多

挂接4个MAX6625，实现4点温度测控与越限报警。分别将器件的ADD(5脚)与不同引脚相连接，可构成它们各自的片选地址供主控制器辨识选用。ADD的连接与其相应片选地址见表3-3。

表3-3 ADD的连接与相应的片选地址

ADD(5脚)的连接	相应片选地址
ADD(5脚)接 GND(2脚)	1001000B
ADD(5脚)接 V _S (6脚)	1001001B
ADD(5脚)接 SDA(1脚)	1001010B
ADD(5脚)接 SCL(3脚)	1001011B

2. 越限报警信号的产生与输出

MAX6625器件具有温度越限报警功能，报警信号由OT端(4脚)输出。报警信号的产生与输出方式决定于配置寄存器状态对MAX6625初始状态的设置，以及高/低温寄存器中设置的 T_{HIGH}/T_{LOW} 温度值。

(1) 比较模式下越限报警信号的产生与输出。通过配置寄存器的状态设置将MAX6625的初始状态设置为

$$[D_7\ D_6\ D_5\ D_4\ D_3\ D_2\ D_1\ D_0] = [00010000]$$

即正常工作(D0=0)，比较工作模式(D1=0)，OT端输出的报警信号为低电平(D2=0)，故障深度为4([D4 D3]=[1 0])。

设置高温寄存器的高温点 T_{HIGH} 和低温寄存器的低温点 T_{LOW} 。

在上述设置情况下，以 $T_{HIGH}=+100^{\circ}\text{C}$, $T_{LOW}=+80^{\circ}\text{C}$ 为例说明越限报警信号的产生与输出：当A/D转换器(ADC)转换输出的温度值高于 $+100^{\circ}\text{C}$ 的次数连续有4次时(故障深度为4)，OT端才输出低电平报警信号；当A/D转换器输出的温度值低于 $+80^{\circ}\text{C}$ 的次数连续有4次时，OT端将退出报警，由低电平跳变为高电平。

MAX6625中的故障排队深度计数器用来计数超限的次数，当连续超限次数等于编程设置的故障排队深度时才触发报警，这样可以防止干扰引发波动的误报警。

(2) 中断模式下越限报警信号的产生与输出。这时配置寄存器的状态设置为

$$[D_7\ D_6\ D_5\ D_4\ D_3\ D_2\ D_1\ D_0] = [00010010]$$

再设置高/低温寄存器的 T_{HIGH} 与 T_{LOW} 。

在中断模式中，MAX6625既有高温越限报警，又有低温越限报警。报警信号的激活与前次越限报警有关。器件上电以后故障报警均被清除，待到读温度寄存器的操作后才开始越限计数，以故障排队深度为条件激活产生报警信号。

3.8.4 MAX6625的操作与使用

1. 写配置寄存器与高/低温寄存器

MAX6625器件执行温控或越限报警功能之前必须由主控制器经I²C总线发出一串时序数据脉冲。

1) 写配置寄存器

在串行时钟线SCL保持高电平时，在SDA串行数据线上产生一个START启动脉冲

之后与停止位脉冲之间有 27 个时序脉冲流传输。其内容包括：MAX6625 器件的片选地址、指针寄存器选配置寄存器的地址、配置寄存器的状态数据。其写入时序图如图 3-18 所示。

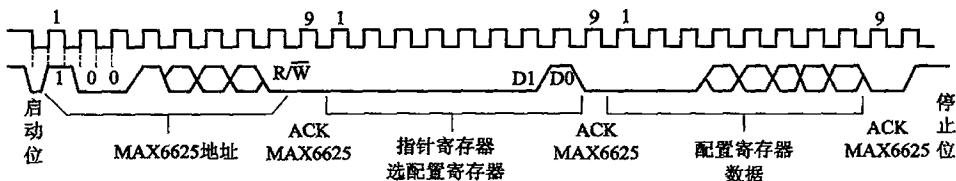


图 3-18 配置寄存器的写入时序图

- (1) MAX6625 器件的片选地址：7 位，1001xxx，由 7 个高低电平的脉冲表示。
- (2) R/W 位：1 位，由第 8 个脉冲表示，高电平 1 为读，低电平 0 为写入。R/W 决定数据流的传输方向是流向器件写还是流向器件读。
- (3) ACK 应答位：1 位，由第 9 个脉冲表示，低电平表示响应，即 I²C 总线上挂接的其它 MAX6625 器件处于空闲状态，选中的 MAX6625 器件处于等待写入/读出数据状态。
- (4) 地址指针寄存器选配置寄存器的地址：地址指针寄存器的状态数据应设置为所指向的配置寄存器的 8 位地址 [00000001]，由第 10~17 位脉冲电平来表示。
- (5) ACK 应答位：1 位，由第 18 位脉冲表示，低电平表示响应，即其它的寄存器处于空闲状态，选中的配置寄存器处于等待状态。
- (6) 配置寄存器的状态数据位：8 位，由第 19~26 位脉冲电平来表示。
- (7) ACK 应答位：1 位，由第 27 个脉冲表示，低电平表示响应。
- (8) 停止位：1 位，由第 28 个脉冲的高电平来表示。

2) 写高/低温寄存器

高/低温寄存器的写时序与配置寄存器的写时序基本相同。要注意的是，第 10~17 位时序脉冲电平应表示地址指针寄存器指向高/低温寄存器的 8 位地址 [00000011]/[00000010]。

第 19~34 位脉冲，共 16 个脉冲的电平表示高/低温寄存器的数据 T_{HIGH}/T_{LOW} 。其中符号位为 1 位，温度数值用 8 位表示，后面的 7 个脉冲均为低电平。

2. 读温度寄存器

在温度测量与显示应用场合，由片内温度传感器检测到的温度值需要从 MAX6625 器件中的温度寄存器读出它的当前值。与写时序不同的是，R/W 位为高电平 1 表示读出，数据流从器件流向主控制器。

3.8.5 MAX6625 器件的编程举例

4 点温度测量系统，采用 4 片 MAX6625 测温，用 AT89C51 单片机作微控制器，线路连接如图 3-19 所示。

测温前要对每片 MAX6625 写入初始设置。写入内容包括：欲选择的 MAX6625 片选地址，配置寄存器的地址及其状态位数据，高/低温寄存器的 T_{HIGH}/T_{LOW} 。初始设置写入程序的编程举例分述如下。

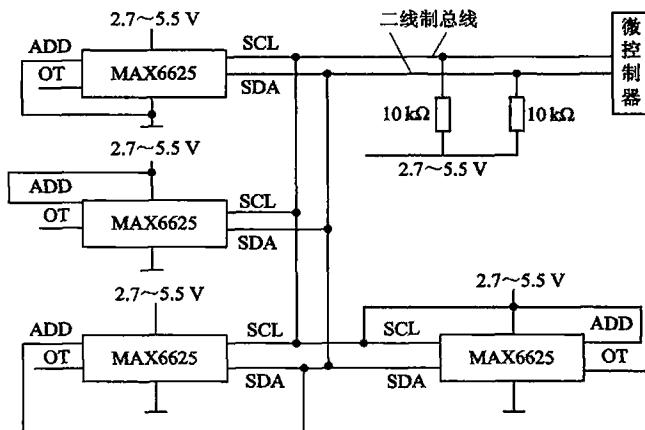


图 3-19 4 片 MAX6625 与单片机微控制器的连接

1. 初始设置程序的编程举例

编程要求 对 MAX6625#1 器件中的配置寄存器写入如下初始状态：器件正常工作时采用比较工作模式，OT 端报警信号低电平有效，故障深度为 4。

解 由图 3-19 连接图可知，MAX6625 的片选地址为 [10010000]；由编程要求知配置寄存器的状态位应为 [00010000]；又知配置寄存器的地址为 [00000001]。

由于 AT89C51 无专用的 I²C 总线接口，因而这里模拟 I²C 接口进行编程。

(1) 配置寄存器的写入程序：

```

WRITE: LCALL START; 产生起始位
MOV A, #10010000B; MAX6625 器件 1 片选地址, R/W=0
LCALL WRBYT;       写器件地址
LCALL CACK;        查询 MAX6625 应答
JB F0, WRITE;      无应答重新开始
MOV A, #00000001B; 配置寄存器地址
LCALL WRBYT;       写入配置寄存器地址
LCALL CACK
JB F0, WRITE
MOV A, #00010000B; 写配置寄存器数据
LCALL WRBYT
LCALL CACK
JB F0, WRITE
LCALL STOP;        产生停止位
RET

```

(2) 高/低温寄存器的写入程序。

编程要求 要求高温寄存器 $T_{HIGH} = 100^{\circ}\text{C}$ 。

解 已知高温寄存器地址为 [00000011]，低温寄存器地址为 [00000010]，无符号 8 位二进制数码 [00101000] = 80°C, [00101010] = 100°C。

以下为高温寄存器的写入程序：

WRITE1, LCALL START;	产生起始位
MOV A, #10010000B;	MAX6625 器件 1 的片选地址, R/W=0
LCALL WRBYT;	写器件片选地址
LCALL CACK;	查询 MAX6625 应答
JB F0, WRITE1;	无应答重新开始
MOV A, #00000011B;	高温寄存器地址
LCALL WRBYT;	写入高温寄存器地址
LCALL CACK	
JB F0, WRITE1	
MOV A, #01010000B;	写高温寄存器高 8 位
LCALL WRBYT	
LCALL CACK	
JB F0, WRITE1	
MOV A, #00000000B;	写高温寄存器低 8 位
LCALL WRBYT	
LCALL CACK	
JB F0, WRITE1	
LCALL STOP	
RET	

低温寄存器的编程设置只要改变地址和数据即可。

2. MAX6625 寄存器读程序的编程举例

MAX6625 的寄存器读程序包括温度寄存器的读、高/低温寄存器的读和配置寄存器的读。

温度寄存器的读程序如下：

READ: LCALL START	
MOV A, #10010000B;	MAX6625 器件 1 的片选地址
LCALL WRBYT;	写 MAX6625 器件 1 的地址
LCALL CACK	
JB F0, READ	
MOV A, #00000000B;	温度寄存器的地址
LCALL WRBYT;	写温度寄存器的地址
LCALL CACK	
JB F0, READ	
LCALL START	
MOV A, #10010001B;	R/W=1, MAX6625 器件 2 的片选地址
LCALL WRBYT;	写 MAX6625 器件 2 的地址
LCALL CACK	
JB F0, READ	
LCALL RD BYT;	读温度寄存器高 8 位
LCALL MACK;	主控制器回答
LCALL RD BYT;	读温度寄存器低 8 位
LCALL MNACK;	主控制器不回答
LCALL STOP	
RET	

3.9 XTR101 4~20 mA 回路变送器芯片

3.9.1 简介

XTR101 是一个单片精密、低漂移的两线制变送器，它可以把微弱的电压信号进行放大并转换成 4~20 mA DC 电流信号后进行远距离传送。它由一个高精度的仪表放大器、压控输出电流源和两个精密 1 mA 电流源组成。它可作为电阻型、电压型及电桥型传感器的信号调理变送电路，被广泛用于工业过程控制、生产自动化、压力和温度等非电量变换、远距离测量以及监控等系统中。

3.9.2 结构框图及引脚功能

XTR101 的内部电路结构如图 3-20 所示，主要由信号放大器、压控电流源和激励恒流源三部分组成。两路 1 mA 恒流源为传感器提供激励和零点迁移，由 A_1 、 A_2 单电源仪表放大器构成传感器信号放大器，由它们控制 A_3 与 V_1 构成的电流源，实现电压/电流变换输出。

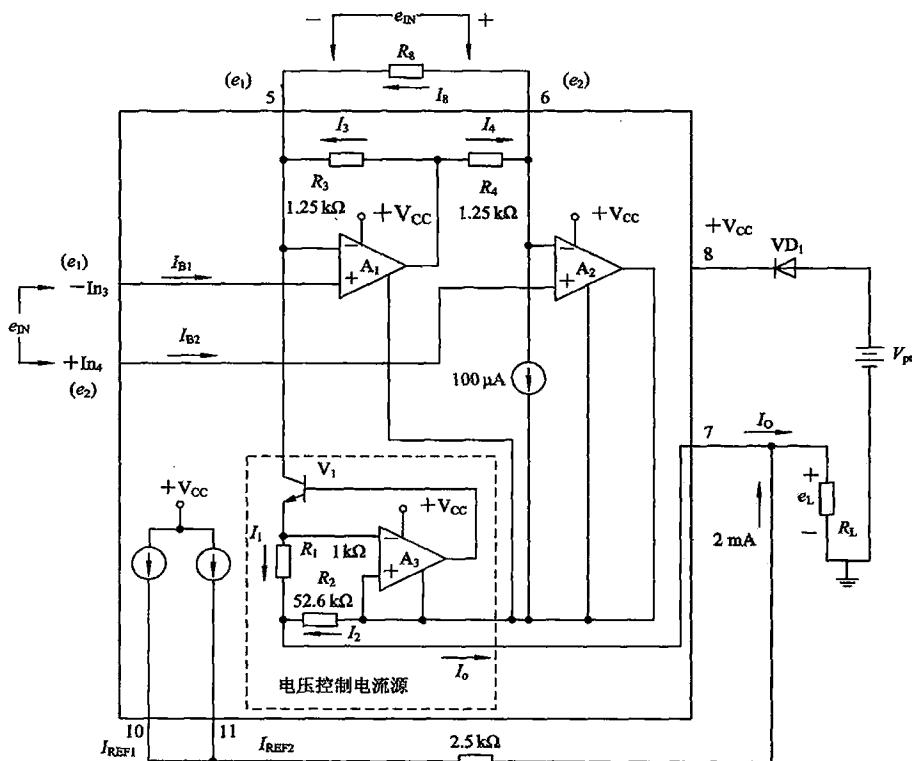


图 3-20 XTR101 原理框图

根据图中的元件参数和到放大器的虚短和虚断特性，可得出输出电流 I_o 的表达式为

$$I_o = (40/R_s + 0.016/1 \Omega)e_{IN} + 4 \text{ mA}$$

式中， R_s 为量程电阻； e_{IN} 为输入毫伏信号，即传感器输出信号。

由上式可见，当 $e_{IN}=0$ 时， $I_o=4 \text{ mA}$ ；当 e_{IN} 最大时，调节 R_s 可使 $I_o=20 \text{ mA}$ 。

3.9.3 主要参数

- (1) 总误差： $<\pm 1\% (-40^\circ\text{C} \sim +85^\circ\text{C})$ ；
- (2) 输入失调电压： $<\pm 2.5 \text{ mV}$ ；
- (3) 失调电压温漂： $<\pm 5 \mu\text{V}/^\circ\text{C}$ ；
- (4) 输入失调电流： $<20 \text{ nA}$ ；
- (5) 失调电流温漂： $<\pm 0.25 \text{ nA}/^\circ\text{C}$ ；
- (6) 差模输入阻抗： $>3 \text{ G}\Omega$ ；
- (7) 共模输入阻抗： $>0.5 \text{ G}\Omega$ ；
- (8) 共模抑制比： $\geq 80 \text{ dB}$ ；
- (9) 电源抑制比： $\geq 110 \text{ dB}$ ；
- (10) 工作电源电压： $+9 \sim +40 \text{ V}$ ；
- (11) 工作方式：两线或三线制。

3.9.4 应用方法举例

1. 带热电阻冷端补偿的热电偶输入温度变送器

电路如图 3-21 所示。一路 1 mA 恒流电流流过 20Ω 调零电位器，实现零点调节，从而使输入信号最小时，输出为 4 mA 电流信号。另一路 1 mA 恒流电流流过并联的 100Ω 的铂电阻 RTD 和 15Ω 的锰铜电阻，产生的电势正好对 J 型热电偶进行了冷端补偿，热电偶的输出信号、冷端补偿信号加调零信号作为 e_{IN} ，送入 XTR101 输入端(3、4 脚)。

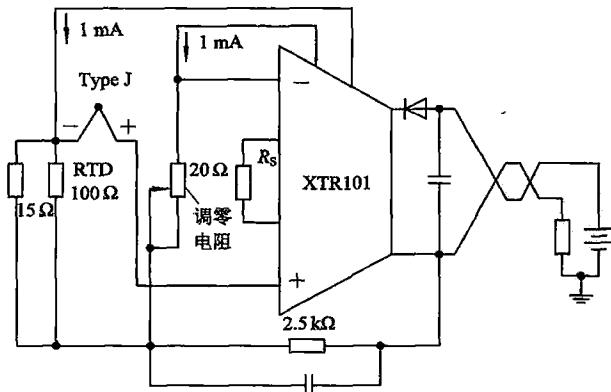


图 3-21 带热电阻冷端补偿的热电偶输入温度变送器

R_s 为增益调节电阻，调节 R_s 可使输入电压 e_{IN} 在从最小值变到最大值时输出电流 I_o

从4 mA变到20 mA。即 $\Delta e_{IN} = e_{IN\ max} - e_{IN\ min}$ 对应 $\Delta I = 16$ mA的输出电流。

由于XTR101使用的是单电源，因此在正常工作时，信号输入端应加+5 V左右的直流偏置电压。见图3-20中，两路1 mA激励源流过2.5 kΩ电阻，形成5 V的直流偏置电压，确保放大器正常工作。由于两个输入端都存在直流偏压，这就相当于在放大器的输入端存在一个共模电压，XTR101的技术指标中已经包含了这部分误差。如果偏置不是5 V，而是另外一个共模电压CMV，则会在输入端引入 $(CMV - 5)/CMRR$ 的失调误差(CMRR是共模抑制比)。

芯片采用9~40 V直流电压源供电，两条电源线同时作为输出电流的信号线。通常在供电回路中加一个防电源反偏二极管，以便在电源不慎接反的情况下保护芯片免受损坏。

2. 电桥输入型压力变送器

电路如图3-22所示。 R_s 为量程电阻， R_1 、 R_2 为电桥平衡电阻，即调零电阻。电桥激励由1 mA恒流源提供，另一路1 mA恒流源和 R_1 、 R_2 电位器构成调零电路。

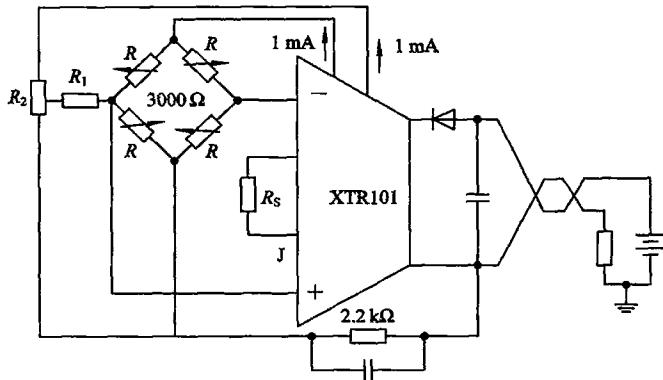


图3-22 电桥输入型压力变送器

电桥平衡电阻 R_1 、 R_2 的确定：对于电桥电阻大于2750 Ω的应用场合，调整 R_2 值使在量程下限时，电路输出4 mA，调零的同时也调整了电路内部的失调电压。兼顾调节范围大小和调节灵敏度， R_1 取电桥电阻的5~10倍， R_2 略大于 R_1 。若输入电桥电阻小于2750 Ω，则需要加以相应的变换。

3.10 适配桥路式传感器的信号调理电路 ——ZMD31050芯片

3.10.1 简介

ZMD31050是一款高精度桥式传感器信号调理电路，可以直接以数字信号输出，只需简单的外围电路，即可构成单片数字传感器。

该芯片的特点是高精度数字输出，外接电路简单，通用性强，多种输出方式可选择。

ZMD31050对传感器的偏移、灵敏度、温漂和非线性可以有效进行补偿和校正，目前

基于 ZMD31050 的数字式气压传感器系统已经在地震前兆辅助观测领域中投入使用，其设计经验值得在其他桥式传感器设计中推广应用。

3.10.2 功能框图与引脚排列

ZMD31050 功能框图如图 3-23 所示，其封装及引脚定义如图 3-24 所示。

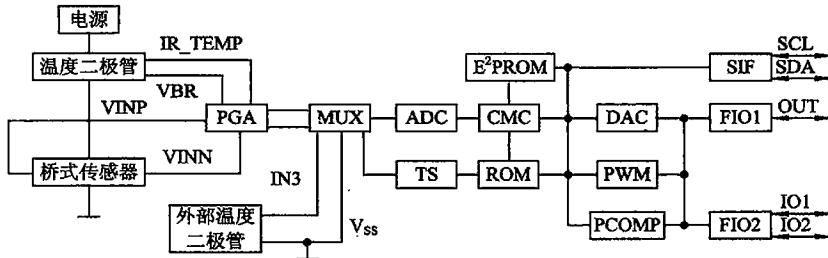


图 3-23 ZMD31050 功能框图

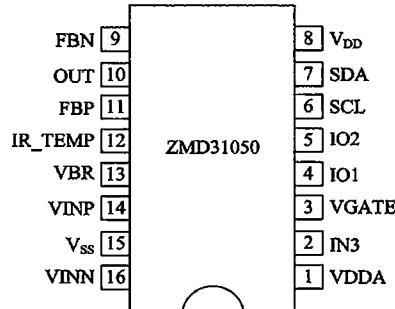


图 3-24 ZMD31050 封装外形及引脚定义

ZMD31050 具有内部 E²PROM，共包含 32 个 16 位地址空间。ZMD31050 工作所必需的 32 个参数存储在 E²PROM 中。其中，22 个校准常量用于传感器信号的计算校准，7 个参数用于配置应用程序，1 个 CRC 字(word)用于检查 E²PROM 内容的正确性，另外还有 2 个 16 位的字供用户自由使用。

每次上电后，E²PROM 中的内容被复制到 RAM 中，根据 RAM 中的配置，器件自动完成信号调理过程。首先，桥式传感器传递出的信号在 PGA 中进行预放大，MUX 将该信号与外部二极管或分立温度传感器信号按照某种序列传送给 ADC 单元，ADC 单元对这些信号进行 A/D 转换。然后，CMC 根据 ROM 中存放的校正公式和 E²PROM 中存放的校准参数对信号进行校正。根据设置，传感器信号以模拟量、数字量或 PWM 的形式输出，输出信号由串行接口及 FIO1、FIO2 提供。

图 3-23 中，ZMD31050 的各组成部分说明如下：

PGA——可编程增益放大器；

MUX——多路器；

ADC——A/D 转换器；

CMC——校准控制器；

DAC——D/A 转换器；
FIO1——可选 IO1，模拟输出(电压/电流)，PWM2 (脉宽调制)，ZACwireTM (一线接口)；
FIO2——可选 IO2，PWM1，SPI，SPI 从机选择，Alarm1、Alarm2 串行接口，I²C，时钟；
PCOMP——可编程比较器，提供报警输出信号；
E²PROM——用于校准参数设置和校准配置；
TS——片上温度传感器；
ROM——存放校准模式；
PWM——脉宽调制模式。

3.10.3 主要参数

ZMD31050 的主要参数如下：

- 电源电压：+2.7~+5.5 V；
- 输入信号：1~275 mV/V；
- 多种输出方式可供选择：电压(0~5 V)，电流(4~20 mA)，PWM，I²C，SPI，ZACwireTM(一线接口)，报警输出；
- 桥式传感器的激励源可选：比例电压、恒压模式或恒流模式；
- 高精度：-25°C~+85°C 时的误差为 0.1%，-40°C~+125°C 时的误差为 0.25%；
- 可对传感器的偏移、灵敏度、温漂和非线性进行数字补偿；
- 输出分辨率最高为 15 位，可选择相对应的采样频率(最多 3.9 kHz)；
- PC 通过数字接口实现器件的配置和校准。

3.10.4 应用举例

ZMD31050 接收来自前端桥式传感器的微弱模拟信号，将这一信号放大，经 A/D 转换、补偿与校正后以数字信号形式传给后端微处理器；微处理器获取信号并进行处理；串口电平转换器 MAX202 完成电平转换，从而实现系统与 PC 的通信。

图 3-25 所示为基于 ZMD31050 的数字式气压传感器应用系统框图。J₁ 为桥式压力传感器与 ZMD31050 的接口，J₂ 是系统供电电源接口，J₃ 为系统与 PC 的接口。桥式压力传感器与 ZMD31050 简单连接即构成了数字式气压传感器，主要应用于观测气体压力的变化。其中，桥式压力传感器为美国 Silicon Microstructures 公司的 SMI5502-015-A。该器件量程为 1 标准大气压，满足目标测量要求。采用恒压源供电，输出电压信号，便于整个系统设计。本数字式气压传感器对温度无特殊要求，温度测量选择内置温度二极管。经 ZMD31050 调理过的信号通过数字串口以 15 位数字信号的形式输出。

外接的微控制器与其外围的晶体振荡器、看门狗、MAX202 构成微控制系统。其中，外接微控制器为 51 系列单片机中的 AT89C4051，可满足全部数据处理和通信的要求。晶振 Y₁ 选择 11.0592 MHz，为 AT89C4051 提供时钟信号。X5045 为 AT89C4051 工作产生复位信号。MAX202 完成电平转换，实现 AT89C4051 与上位机通信。

数字式气压传感器与微控制系统构成数字式气压传感器应用系统，整个系统采用 +5 V 电源供电。数字式气压传感器支持 I²C 通信模式，由 SDA、SCL 线经上拉后与后端微控制器系统连接。通信过程中，ZMD31050 为从机模式，其默认通信地址为 0x78。

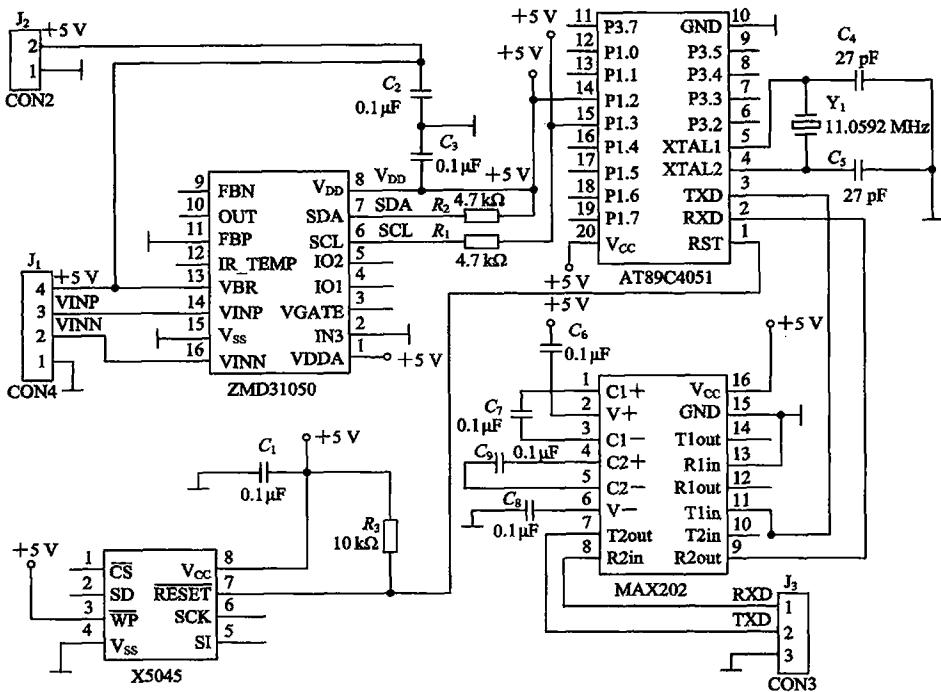


图 3-25 基于 ZMD31050 设计的数字式气压传感器应用系统框图

3.11 双轴加速度传感器系统——全系统单片集成芯片 ADXL202

3.11.1 简介

ADXL202 是美国 ADI 公司出品的低成本、低功耗、功能完善的双轴加速度传感器。它有以下的功能特点：

- (1) 将敏感元件和调理电路集成在一个芯片上，小巧紧凑，提高了可靠性，可广泛地应用在斜度测量、惯性导航、地震监测装置和交通安全系统等领域。
- (2) 既能测量动态加速度(如振动加速度)，又能测量静态加速度(如重力加速度)。
- (3) 在 3 dB 带宽情况下的频响达 5 kHz。
- (4) 与电解质型、水银型、热敏型等传感器相比，ADXL202 响应速度更快。
- (5) 耐冲击性，可耐高达 1000g 的冲击。

3.11.2 结构框图与功能

ADXL202 结构框图如图 3-26 所示，它是基于单块集成电路的完善的双轴加速度测量系统，采用以多晶硅为表面的微电机传感器和信号控制环路来执行操作的开环加速测量结构。对每根轴而言，输出环路将模拟信号转换为脉宽占空比的数字信号。这些数字信号直接与微处理器接口。ADXL202 可测量正负加速度，其最大测量范围为 $\pm 2g$ 。ADXL202

也可测量静态加速度，亦可用作斜度测量。传感器采用在硅片上经表面微加工的多晶硅结构，用多晶硅的弹性元件支撑它并提供平衡加速度所需的阻力。结构偏转是通过由独立的固定极板和附在移动物体上的中央极板组成的可变电容来测量的。固定极板通过方波的每 π 个相位控制。加速度计受到加速度力后改变了可变电容的平衡，使输出方波的振幅与加速度成正比。而相位解调技术用来提取信息，判断加速方向。解调器的输出通过 $32\text{ k}\Omega$ 的固定电阻输出到脉宽占空比解调器。这时，允许用户改变滤波电容的大小来设置输出信号的带宽。这种滤波提高了测量的精度，并有效地防止频率混叠。经过低通滤波后，模拟信号由 DCM（脉宽占空比解调器）转换为脉宽占空比信号。通过一个电阻 R_{SET} 将 T_2 设定在 $0.5\sim10\text{ ms}$ 范围内，在 0 加速度时使输出占空比为 50% 。加速度可由一计数器/计时器或低功耗的微控制器通过测量 T_1 、 T_2 来测得。模拟输出信号可通过以下两种方法获得：一种从 XFILT 和 YFILT 引脚得到；一种是通过 RC 滤波器对 XOUT 和 YOUT 引脚输出的脉冲信号滤波后得到的 DC 值推算。

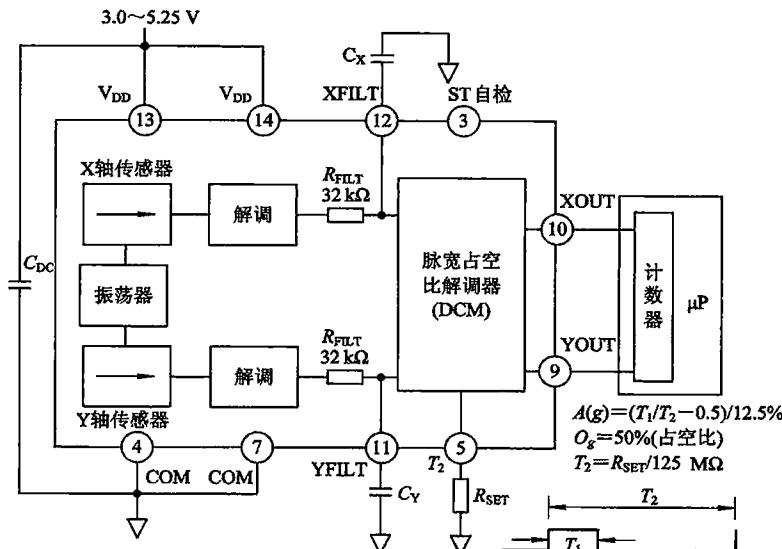


图 3-26 ADXL202 结构框图

ADXL202 的引脚排列如图 3-27 所示，其引脚定义及功能见表 3-4。



图 3-27 ADXL202 引脚图

表 3-4 ADXL202 引脚定义及功能

引脚序号	名 称	功 能
1, 6, 8	N. C.	空置端
2	VTP	检测端
3	ST	自我测试端
4, 7	COM	公共接地端
5	T2	外接 R_{SET} , 可设定 T_2 周期
9	YOUT	Y 轴脉宽信号输出端
10	XOUT	X 轴脉宽信号输出端
11	YFILT	连接 Y 滤波电容
12	XFILT	连接 X 滤波电容
13	V _{DD}	电源端, 接 +3~+5.25 V, 与 14 引脚连接
14	V _{DD}	电源端, 接 +3~+5.25 V, 与 13 引脚连接

3.11.3 主要参数特点

ADSL202 的主要参数特点如下。

- (1) 两种工作温度范围：商业温度范围为 0°C~70°C，工业温度范围为 -40°C~+85°C。
- (2) 具有下面两种输出信号：
 - ① 从 XFILT 端与 YFILT 端输出直流信号；
 - ② 从 XOUT 端与 YOUT 端输出一定占空比的数字量信号。
- (3) 灵敏度、重力、加速度的检测分辨率优于 $5 \times 10^{-3} g$ ，直流输出灵敏度为 312 mV/g。
- (4) 可以承受电源开关反复接通断开的冲击，电源开关断开后再接通时的恢复时间约 1.6 ms。
- (5) 低功耗(<0.6 mA)。
- (6) 每根轴的带宽均可通过电容调整。
- (7) 直流工作电压为 +3~+5.25 V。
- (8) 可承受 1000g 的剧烈冲击。

3.11.4 使用说明及典型应用举例

1. ADXL202 使用注意事项

1) 传感器过载的影响

ADXL202 可输出数字信号并直接传输给微处理器，而无需 A/D 转换或附加其他电路。如果把传感器输出数字信号的两个引脚 XOUT、YOUT 直接接到单片机的两个 I/O 引脚上，ADXL202 的静态电源电流为 0.6 mA，而单片机 I/O 引脚对地之间的电阻为 15 kΩ 左右，在 5 V 的电压下会产生 0.33 mA 的电流，同时施加两路信号就会产生“过载”现象。为了解决这一问题，可以用运放构建一个电压跟随器，过载问题就可得到较好的解决。

2) 传感器信号干扰的影响

ADXL202 数据资料中已经明确指出，传感器与微处理器共用电源时会引起干扰，解决干扰的方案如图 3-28 所示。同时应该注意：① 传感器要采用独立的电池供电；② 在布线时要把器件和调理信号的电阻、电容放在一块单独的板上或用粗地线将其围在线路板的某一区域，才能得到较好的抑制干扰的效果。

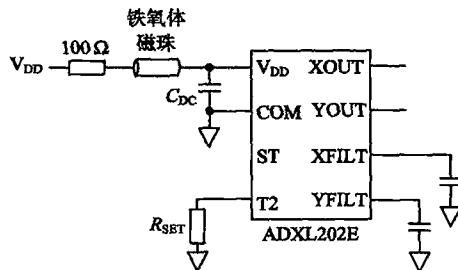


图 3-28 噪声干扰抑制电路

3) 传感器精度、分辨率的影响

传感器的分辨率主要取决于 XFILT 和 YFILT 设定的滤波带宽、微控制器的计算速度。ADXL202 的脉宽占空比信号具有 14 位的分辨率，如果设置好合适的滤波带宽，并且微处理器运行速度足够快，就可以得到很高的分辨率。实验发现传感器的绝对误差在 $\pm 0.005g$ ，而传感器的量程为 $\pm 2g$ ，9 位的分辨率就已经足够了，采用 14 位的分辨率意义不大。另外，传感器的电源电压可以在 3~5.25 V 范围内变化，但电源电压的变化是否会对数字信号的输出产生一定的影响，这个问题在器件数据资料中并没有说明。如果是长期工作，又采用电池供电的方式，则应对该问题引起重视。

4) 传感器温度变化的影响

根据 ADXL202 数据资料，温度变化只会对两个通道的脉宽占空周期 T_2 产生缓慢影响，只需温度变化后适时修正，不必每次都测量 T_2 。实验发现，这种说法只在温度缓慢变化的条件下成立，如果环境温度发生突变（例如夏天从有冷气空调的环境突变到炎热的户外环境）， T_2 就会发生毫无规律的剧烈变化，至少需要 30 分钟以上的稳定期，传感器才会恢复正常，使用中应注意这一点。

5) 传感器标定的影响

根据 ADXL202 数据资料，理想状态下传感器在感应到一个正向的重力加速度时，占空比 T_1/T_2 应为 62.5%；感应到一个反向的重力加速度时， T_1/T_2 应该为 37.5%；不承受加速度时， T_1/T_2 应为 50%。但是，测试发现每一个样品的试验值都与理想值差距很大，而且各不相同。为了提高数据精度，分别对每一个传感器进行标定。先记下一个正向重力加速度时的占空比，如 55%，再记下一个反向重力加速度时的占空比，如 32%。根据数据资料的说明，此时的灵敏度 $S = (55\% - 32\%) / 2g = 11.5\% / g$ 。如果线性度比较好，把传感器水平放置不让其承受加速度，此时的占空比为 $(55\% - 11.5\%) / 2g = 43.5\% / 2g$ ，但实验的结果并非如此。在对不同的传感器进行数据比较时，发现测量到的动态加速度误差比静态加速度误差值大得多。在所有试验参数都相同的情况下，测得重力加速度在敏感轴上的

分量绝对误差为 $\pm 0.005g$ ，而在振动条件下同一时刻不同传感器加速度的最大差值达 $0.04g$ 。在使用时，应注意这一问题。

2. 典型应用——ADXL202 加速度计在振动测试中的应用

ADXL202 用于振动测试的测量电路原理图如图 3-29 所示。

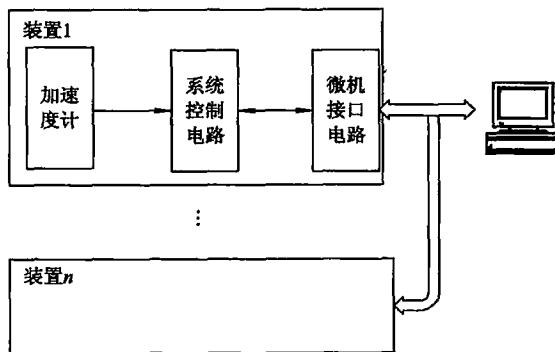


图 3-29 ADXL202 用于振动测试的测量电路原理图

其中，加速度计部分电路主要由 ADXL202 构成，并辅助以一些滤波及调节电路；系统控制电路主要由 AT89C2051 单片机构成，利用它处理 ADXL202 产生的占空比调制信号；微机接口电路主要实现单片机与 PC 之间的串行通信接口，通过该接口，测量结果可以传送到 PC 中，同时 PC 可以通过命令来控制测量电路的工作。此外，在控制电路中还使用了串行 E²PROM 来存储本测量装置的标号、采样参数以及与占空比计算相关的参数，这样有利于提高测量准确度，并有利于多点测量。

在单片机控制中使用中断方式进行串行通信，使用定时器产生一个时间基准来记录 ADXL202 输出信号。当单片机接收到 PC 发送的指令后，单片机开始执行数据采集及计算，并将计算结果以及运行状态存放在指定单元中，PC 可以随时读取。此外，单片机可以执行 PC 发送的多种命令，如存取标定参数，存取指定伴置数据，更改测试终端的编号等。由于 T_s 是由 R_s 确定的，在测量过程中变化很小，利用这一特点可以简化采样程序，提高占空比计算速度。在 PC 中，使用 VC6.0 编制控制程序。

参 考 文 献

- [1] 刘君华. 基于 LabVIEW 的虚拟仪器设计[M]. 北京：电子工业出版社，2003
- [2] 刘君华. 基于 LabWindows/CVI 的虚拟仪器设计[M]. 北京：电子工业出版社，2003
- [3] Torg Stecker, Helmut Kremer. 德国 AMG 公司，德国美茵茨. 用于电容传感器信号转换的集成电路 CAV424[J]. 仪表技术与仪器，2003, 40~48
- [4] 刘君华, 贾惠芹, 丁晖, 等. 虚拟仪器图形化编程语言 LabVIEW 教程[M]. 西安：西安电子科技大学出版社，2001
- [5] 刘君华. 虚拟仪器编程语言 LabWindows/CVI 教程[M]. 北京：电子工业出版社，

2001

- [6] Maxim Integrated Products. Signal Conditioner. MAX1450. <http://www.maxim-ic.com>
- [7] Maxim Integrated Products. Signal Conditioner. MAX1460. <http://www.maxim-ic.com>
- [8] Analog Devices. ADuC812 单片智能转换器. <http://www.analog.com>
- [9] Maxim Integrated Products. MAX6625/MAX6626 9-bit/12-bit Temperature Sensors with I²C-Compatible Serial Interface in a SOT23. <http://www.maxim-ic.com>
- [10] Burr-Brown Corporation. XTR101: Precision, Low Drift 4 – 20 mA Two-wire transmitter. <http://www.burr-brown.com>
- [11] ZMD AG. ZMD31050 Advanced Differential Sensor Signal Conditioner Datasheet. <http://www.zmd.biz>
- [12] Analog Devices. ADXL202/ADXL210 Low Cost 2 g/10 g Dual AxisMEMS Accelerometers with Digital Output. <http://www.analog.com>
- [13] Analog Microelectronics. Converter IC for Capacitive Signals CAV414. <http://www.analogmicro.de>
- [14] Analog Devices. AD698 Universal LVDT Signal Conditioner. <http://www.analog.com>

第4章 基本智能化功能与其软件实现

本章内容

- 改善线性度及智能化非线性刻度转换功能
- 改善静态性能提高系统精度及智能化自校零、自校准功能
- 改善稳定性抑制交叉敏感及智能化多传感器数据融合功能
- 改善动态性能扩展频带及智能化频率自补偿功能
- 提高信噪比与分辨力及智能化信号提取与消噪功能
- 增强自我管理与自适应能力及智能化控制功能

实现传感器各项智能化的功能和建立智能传感器系统，是传感器克服自身不足，获得高稳定性、高可靠性、高精度、高分辨力与高自适应能力的必由之路与必然趋势。不论非集成化实现方式还是集成化实现方式，或是混合实现方式，传感器与微计算机/微处理器赋予智能的结合所实现的智能传感器系统，都是在最少硬件条件基础上采用强大的软件优势来“赋予”智能化功能的。本章所介绍的内容，仅是实现部分基本的智能化功能常采用的智能化技术。

4.1 改善线性度及智能化非线性刻度转换功能

测量系统的静态性能由其静态输入输出特性来表征，它的质量指标将决定测量系统的精度，测量系统的线性度指标是影响系统精度的重要指标之一。其中，处于测量系统前端的传感器，其输入输出特性的非线性是使得测量系统输入输出特性具有非线性的主要原因。

传感器及其调理电路的输出量多是电学量，传统测量仪器系统的基本功能就是要将传感器及其调理电路输出的电学量转换为被测量，以便输出显示，称为刻度转换。如果按照线性关系进行刻度转换，就会引入非线性误差，降低线性度指标，因为前端待转换的关系是非线性的。当然，人们期望传感器本身的输入输出特性具有良好的线性，为此传感器工作者一直进行着不懈的努力。但是，传感器静态特性的非线性却总是存在的。与此同时，人们从电路方面精心设计非线性校正器以期改善系统的非线性。所谓非线性校正器，就是可以按某种非线性关系来进行刻度转换的环节。系统中有了这种非线性刻度转换环节，全系统输入输出特性将逼近直线。由于各个传感器非线性特性的不一致性，因此用硬件电路实现非线性校正的刻度转换存在很大难度与局限性。

智能传感器系统是通过软件来进行非线性刻度转换的，在实现智能化刻度转换功能的

同时，也实现了非线性自校正功能，从而改善了系统的静态性能，提高了系统的测量精度。由于软件的灵活性，智能传感器系统丝毫不介意系统前端的正模型有多么严重的非线性。所谓正模型，是传感器及其调理电路的输入输出特性($x-u$)，如图4-1(b)所示。智能传感器能自动按图4-1(c)所示的逆模型进行刻度转换，输出系统的被测量值 y ，实现系统的输出 y 与输入 x 呈理想直线关系，如图4-1(d)所示。

所谓逆模型，是指正模型

$$u = f(x) \quad (4-1)$$

的反非线性特性

$$y = x = f(u) \quad (4-2)$$

式中： x 为系统的被测输入量； u 为传感器及其调理电路的输出量，又是存放在微机中非线性校正器软件模块的输入； $y=x$ 为非线性校正器软件模块的输出，也即系统的总输出。

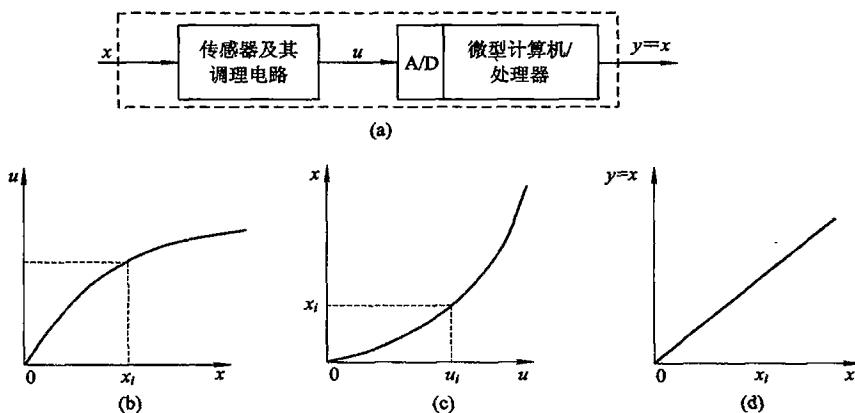


图 4-1 智能传感器系统

(a) 智能传感器系统框图；(b) 正模型；(c) 逆模型；
(d) 智能传感器系统的输入(x)输出(y)特性

智能传感器系统采用软件既灵活又简便地实现了非线性自校正功能后，就不必再为改善系统中每一环节的非线性而耗费精力，其所要求的条件仅仅是：前端正模型($x-u$ 特性)具有重复性。

采用智能化非线性自校正模块以实现刻度转换的编程方法有多种，常用的有查表法、曲线拟合法，近年来又发展了神经网络法及支持向量机法等多种方法。它们都具有极强的非线性映射能力，都能通过学习训练后进行所需逆模型关系的映射，在刻度转换过程中不仅能改善非线性，还能改善系统稳定性抑制交叉敏感，这一点将在相应章节中介绍。本章重点介绍常用的查表法与曲线拟合法。

4.1.1 查表法

查表法是一种分段线性插值法，根据精度要求对反非线性曲线(如图4-2)进行分段，用若干段折线逼近曲线，将折点坐标值存入数据表中，测量时首先要明确对应输入被测量

x_i 的电压值 u_i 是在哪一段；然后根据那一段的斜率进行线性插值，即得输出值 $y_i = x_i$ 。

下面以三段为例，折点坐标值为

横坐标： u_1, u_2, u_3, u_4 ；

纵坐标： x_1, x_2, x_3, x_4 。

各线性段的输出表达式为

第Ⅰ段

$$y(\text{I}) = x(\text{I}) = x_1 + \frac{x_2 - x_1}{u_2 - u_1}(u_i - u_1)$$

第Ⅱ段

$$y(\text{II}) = x(\text{II}) = x_2 + \frac{x_3 - x_2}{u_3 - u_2}(u_i - u_2)$$

第Ⅲ段

$$y(\text{III}) = x(\text{III}) = x_3 + \frac{x_4 - x_3}{u_4 - u_3}(u_i - u_3)$$

输出 $y=x$ 表达式的通式为

$$y = x = x_k + \frac{x_{k+1} - x_k}{u_{k+1} - u_k}(u_i - u_k) \quad (4-3)$$

式中： k 为折点的序数，3 条折线有 4 个折点， $k=1, 2, 3, 4$ 。

由电压值 u_i 求取被测量 x_i 的程序框图如图 4-3 所示。

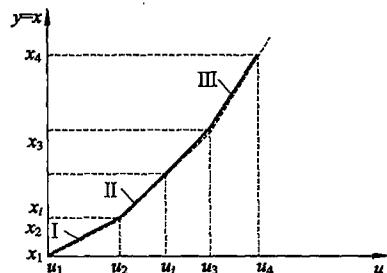


图 4-2 反非线性的折线逼近

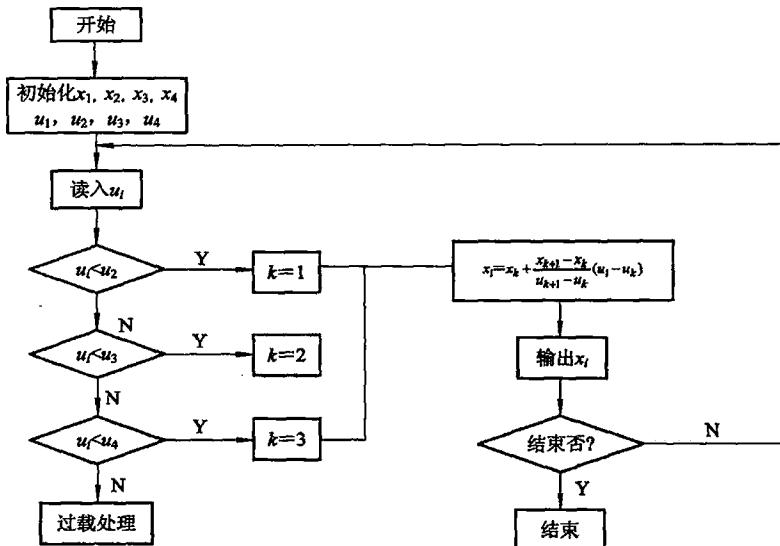


图 4-3 非线性自校正流程图

折线与折点的确定有两种方法： Δ 近似法与截线近似法，如图 4-4 所示。不论哪种方法，所确定的折线段与折点坐标值都与所要逼近的曲线之间存在误差 Δ ，按照精度要求，各点误差 Δ_i 都不得超过允许的最大误差界 Δ_m ，即 $\Delta_i \leq \Delta_m$ 。

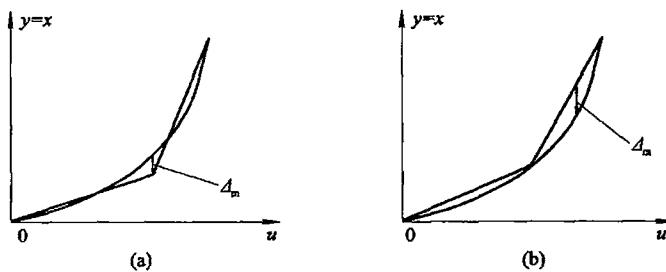


图 4-4 曲线的折线逼近
(a) Δ 近似法; (b) 截线近似法

1. Δ 近似法

折点处误差最大，折点在 $\pm \Delta_m$ 误差界上。折线与逼近的曲线之间的误差最大值为 Δ_m ，且有正有负。

2. 截线近似法

折点在曲线上且误差最小，这是利用标定值作为折点的坐标值。折线与被逼近的曲线之间的最大误差在折线段中部，应控制该误差值小于允许的误差界 Δ_m 。各折线段的误差符号相同，或全部为正，或全部为负。

4.1.2 曲线拟合法

曲线拟合法采用 n 次多项式来逼近反非线性曲线，该多项式方程的各个系数由最小二乘法确定，具体步骤如下。

1. 列出逼近反非线性曲线的多项式方程

(1) 对传感器及其调理电路进行静态实验标定，得校准曲线。标定点的数据为

$$\left. \begin{array}{l} \text{输入 } x_i: x_1, x_2, x_3, \dots, x_N \\ \text{输出 } u_i: u_1, u_2, u_3, \dots, u_N \end{array} \right\} i = 1, 2, \dots, N \quad (N \text{ 为标定点个数})$$

(2) 假设反非线性特性拟合方程为

$$x_i(u_i) = a_0 + a_1 u_i + a_2 u_i^2 + a_3 u_i^3 + \dots + a_n u_i^n$$

n 的数值由所要求的精度来定。若 $n=3$ ，则

$$x_i(u_i) = a_0 + a_1 u_i + a_2 u_i^2 + a_3 u_i^3 \quad (4-4)$$

式中： a_0, a_1, a_2, a_3 为待定常数。

(3) 求解待定常数 a_0, a_1, a_2, a_3 。根据最小二乘法原则来确定待定常数 a_0, a_1, a_2, a_3 的基本思想是，由多项式方程式(4-4)确定的各个 $x_i(u_i)$ 值，与各个点的标定值 x_i 之均方差应最小，即

$$\sum_{i=1}^N [x_i(u_i) - x_i]^2 = \sum_{i=1}^N [(a_0 + a_1 u_i + a_2 u_i^2 + a_3 u_i^3) - x_i]^2 \\ = \text{最小值} = F(a_0, a_1, a_2, a_3) \quad (4-5)$$

式(4-5)是待定常数 a_0, a_1, a_2, a_3 的函数。为了求得函数 $F(a_0, a_1, a_2, a_3)$ 最小值时的常数 a_0, a_1, a_2, a_3 ，我们对函数求导并令它为零，即

令 $\frac{\partial F(a_0, a_1, a_2, a_3)}{\partial a_0} = 0$, 得

$$\sum_{i=1}^N [(a_0 + a_1 u_i + a_2 u_i^2 + a_3 u_i^3) - x_i] \times 1 = 0$$

令 $\frac{\partial F(a_0, a_1, a_2, a_3)}{\partial a_1} = 0$, 得

$$\sum_{i=1}^N [(a_0 + a_1 u_i + a_2 u_i^2 + a_3 u_i^3) - x_i] u_i = 0$$

令 $\frac{\partial F(a_0, a_1, a_2, a_3)}{\partial a_2} = 0$, 得

$$\sum_{i=1}^N [(a_0 + a_1 u_i + a_2 u_i^2 + a_3 u_i^3) - x_i] u_i^2 = 0$$

令 $\frac{\partial F(a_0, a_1, a_2, a_3)}{\partial a_3} = 0$, 得

$$\sum_{i=1}^N [(a_0 + a_1 u_i + a_2 u_i^2 + a_3 u_i^3) - x_i] u_i^3 = 0$$

经整理后得矩阵方程

$$\begin{cases} a_0 N + a_1 H + a_2 I + a_3 J = D \\ a_0 H + a_1 I + a_2 J + a_3 K = E \\ a_0 I + a_1 J + a_2 K + a_3 L = F \\ a_0 J + a_1 K + a_2 L + a_3 M = G \end{cases} \quad (4-6)$$

式中: N 为实验标定点个数; $H = \sum_{i=1}^N u_i$; $I = \sum_{i=1}^N u_i^2$; $J = \sum_{i=1}^N u_i^3$; $K = \sum_{i=1}^N u_i^4$; $L = \sum_{i=1}^N u_i^5$;

$$M = \sum_{i=1}^N u_i^6; D = \sum_{i=1}^N x_i; E = \sum_{i=1}^N x_i u_i; F = \sum_{i=1}^N x_i u_i^2; G = \sum_{i=1}^N x_i u_i^3.$$

通过求解式(4-6)的矩阵方程可得待定常数 a_0 、 a_1 、 a_2 、 a_3 。

2. 将所求得的常系数 $a_0 \sim a_3$ 存入内存

将已知的反非线性特性拟合方程式(4-4)写成下列形式:

$$\begin{aligned} x(u) &= a_3 u^3 + a_2 u^2 + a_1 u + a_0 \\ &= [(a_3 u + a_2) u + a_1] u + a_0 \end{aligned} \quad (4-7)$$

为了求取对应电压为 u 的输入被测值 x , 每次只需将采样值 u 代入式(4-7)中进行三次 $(b+a_i)u$ 的循环运算, 再加上常数 a_0 即可。

这种非线性校正法的缺点在于当有噪声存在时, 可能在求解方程时遇到矩阵病态情况, 而使求解受阻。为了克服这一缺点, 可采用神经网络法、支持向量机等算法。

4.1.3 [示例 4-1]与铂电阻配用的智能化刻度转换模块的设计(曲线拟合法)

要求 测温范围 $0 \sim 500^\circ\text{C}$, 刻度转换模块的绝对偏差小于 0.5°C 。

解 (1) 在 $0 \sim 500^\circ\text{C}$ 范围内从标准分度表中取 $N=11$ 个标准分度值, 如表 4-1 所示。

表 4-1 给出了铂电阻 Pt100 的正模型, 即输入(T)输出(R)特性。

表 4-1 Pt100 铂电阻 $N=11$ 个分度值(0~500°C, 间隔 50°C)

输入 $T/^\circ\text{C}$	输出 R/Ω	输入 $T/^\circ\text{C}$	输出 R/Ω
0.00	100.00	300.00	212.05
50.00	119.40	350.00	229.72
100.00	138.51	400.00	247.09
150.00	157.33	450.00	264.18
200.00	175.86	500.00	280.98
250.00	194.10		

(2) 逆模型的数学表达式设为三阶四项多项式:

$$T = a_0 + a_1 R + a_2 R^2 + a_3 R^3 \quad (4-8)$$

(3) 待定常数 a_0 、 a_1 、 a_2 、 a_3 的确定。根据式(4-5)、式(4-6)求得 $a_0 \sim a_3$ 的数值为

$$a_0 = -247.89, \quad a_1 = 2.4077, \quad a_2 = 0.000\ 602\ 53, \quad a_3 = 1.072 \times 10^{-6}$$

具有上述常系数数值的式(4-8)的编程算式就成为智能化刻度转换模块。

(4) 逆模型的检验。向逆模型输入电阻 R , 比较标准分度值的温度 T 与逆模型计算(输出)值 T' , 其偏差 $\Delta = T' - T$, 结果列入表 4-2。

表 4-2 智能化刻度转换模块式(4-8)的检验结果

输入电阻 R/Ω	输出温度/ $^\circ\text{C}$		偏差 $ \Delta $ $/^\circ\text{C}$	输入电阻 R/Ω	输出温度/ $^\circ\text{C}$		偏差 $ \Delta $ $/^\circ\text{C}$
	分度值 T	计算值 T'			分度值 T	计算值 T'	
100	0	-0.01	0.01	200	266.35	266.34	0.01
110	25.68	25.68	0	210	294.25	294.24	0.01
120	51.57	51.57	0	230	350.81	350.81	0
130	77.65	77.66	0.01	250	408.45	408.46	0.01
140	103.94	103.95	0.01	260.72	440.00	439.80	0.20
150	130.45	130.45	0	280.23	498.00	497.73	0.27
160	157.17	157.17	0				

由表 4-2 所示的检验结果可见, 在 0~500°C 温度范围内, 由标准分度表给出的 T 与逆模型计算的输出值 T' 之差的最大值为 $|0.27|^\circ\text{C}$, 小于 0.5°C 的允许偏差, 故刻度转换用逆模型式(4-8)满足要求。若转换允许偏差减小, 则视情况需增加多项式的阶次及项数。

(5) 线性度改善情况的评价。

① 改善前测温系统的线性度。改善前测温系统的线性度由 Pt100 铂电阻测温传感器的正模型的线性度决定, 其最小二乘法线性度求取步骤如下:

- 拟合直线。由表 4-1 给出的标准分度关系, 根据第 2 章式(2-9)可计算得到拟合直线的两个常系数 k 和 b , 从而最小二乘拟合直线方程为

$$R = 102.169 + 0.36195 T \quad (4-9)$$

- 最大拟合偏差 $\Delta L_m = \Delta R_m$ 。在 $0 \sim 500^\circ\text{C}$ 范围内，式(4-9)根据温度 T 计算所得的 R (计)与相同温度 T 由标准分度表给出的 R (标)之差即为拟合偏差，该拟合偏差的最大值在零度， $|\Delta R_m| = 2.169 \approx 2.17 \Omega$ (合温度偏差约 7°C)。

- 最小二乘法线性度。根据第2章式(2-6)的定义式

$$\delta_L = \frac{|\Delta L_m|}{Y(\text{FS})} \times 100\%$$

式中， $|\Delta L_m| = |\Delta R_m| = 2.17 \Omega$ ，为最大拟合偏差。

$Y(\text{FS}) = R(\text{FS})$ 为满量程输出值，代入测温的上限(500°C)、下限(0°C)值，由式(4-9)可求得 $R(\text{FS}) = R(T=500^\circ\text{C}) - R(T=0^\circ\text{C}) = 283.144 - 102.169 = 180.975 \Omega$ ，于是得

$$\delta_L = \frac{2.17}{180.975} = 1.2\%$$

② 改善后系统的线性度。

- 拟合直线。为简单起见，智能传感器系统的拟合直线可选为理想直线方程

$$T = kT$$

式中， $k=1$ 。

• 拟合偏差。根据表4-2列出的智能传感器系统输出值 T' 与系统输入值 T 应呈线性关系，但却有偏差 $\Delta T = T' - T$ ，从表4-2中可得最大拟合偏差 $|\Delta_m| = 0.27^\circ\text{C}$ 。

- 理论线性度。量程为 $Y(\text{FS}) = 500^\circ\text{C} - 0^\circ\text{C} = 500^\circ\text{C}$ ，理论线性度为

$$\delta_L = \frac{|\Delta_m|}{Y(\text{FS})} = \frac{0.27}{500} = 0.054\%$$

与改善前的 1.2% 相比，经智能化刻度转换模块进行非线性自校正后，全系统线性度提高 22 倍，非线性误差减小 22 倍。

4.2 改善静态性能提高系统精度及智能化 自校零、自校准功能

自校零与自校准功能的核心思想是：不论何种因素，如温度、电源电压波动或自身的老化，引起了传感器输入输出特性发生漂移，偏离了初始标定曲线，只要现场实时进行标定实验，测出漂移后的输入输出特性，并按其进行刻度转换，就能消除特性漂移引入的测量误差，输出的被测量值更接近实际的真实值。智能传感器系统具有自校零与自校准功能，不仅可以消除零点漂移、灵敏度漂移，而且同时进行非线自校正的刻度转换，因此系统稳定性与综合测量精度均大大提高。

这种智能化技术，可以实现采用低精度、低重复性、低稳定性的测量系统而获得高精度的测量结果，其测量精度仅取决于作为标准量的基准。这样，可以不再为使测量系统中的每一个测量环节都具有高精度、高稳定性与高重复性而耗费精力，而只需将主要精力集中在获得高精度、高稳定性的参考基准上面。

根据现场实时欲建立的输入输出特性是线性的还是非线性的，所需标定的点数与所需的基准数目各不相同。两基准法需有两个基准，适用于建立具有线性特性的系统；多基准(至少三个基准)适用于建立具有非线性特性的系统。

4.2.1 两基准法

两基准法又称三步测量法，适用于测量系统的正模型可用线性方程表示的系统。

假设一传感器系统经标定实验得到的静态输出(y)输入(x)特性为如下线性方程：

$$y = a_0 + a_1 x \quad (4-10)$$

式中： a_0 为零位值，即当输入 $x=0$ 时之输出值； a_1 为灵敏度，又称传感器系统的转换增益。

对于一个理想的传感器系统， a_0 与 a_1 应为保持恒定不变的常量。但是实际上，由于各种内在和外来因素的影响， a_0 、 a_1 都不可能保持恒定不变。譬如，决定放大器增益的外接电阻的阻值会因温度变化而变化，并引起放大器增益改变，从而使得传感器系统总增益改变，也就是系统总的灵敏度发生变化。设 $a_1 = S + \Delta a_1$ ，其中 S 为增益的恒定部分， Δa_1 为变化量；又设 $a_0 = P + \Delta a_0$ ， P 为零位值的恒定部分， Δa_0 为变化量，则

$$y = (P + \Delta a_0) + (S + \Delta a_1)x \quad (4-11)$$

式中： Δa_0 为零位漂移； Δa_1 为灵敏度漂移。

由式(4-11)清楚可见，由零位漂移与灵敏度漂移会引入测量误差 Δa_0 与 $\Delta a_1 x$ 。

传统的传感器技术一直追求精心设计与制作、严格挑选高质量的材料及元器件，以期将 Δa_1 及 Δa_0 控制在某一限度内。但这需要以高成本作为代价。

智能传感器系统是传感器与微处理器赋以智能的结合，它通过两个基准对系统进行实时标定能够自动校正因零位漂移、灵敏度漂移而引入的误差。

以对压力传感器进行自校准为例，智能压力传感器系统实现自校准功能的原理框图如图 4-5 所示。微处理器系统在每一特定的周期内发出指令，控制多路转换器执行三步测量法(见 2.4.4 节)，使自校准环节接通不同的输入信号。因为本系统的输入信号为压力，故多路转换器是一个压力扫描阀。

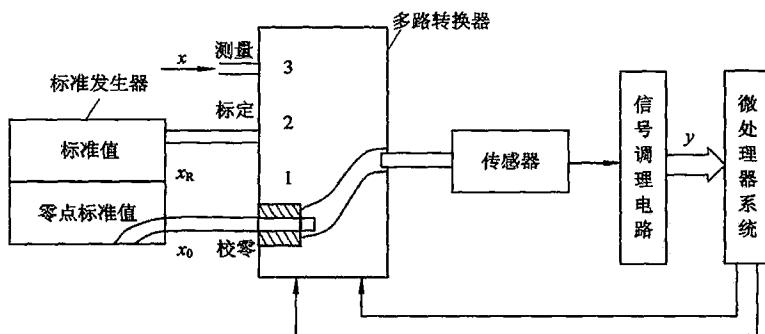


图 4-5 智能传感器系统实现自校准功能原理框图

1) 第一步：校零

测量系统的零点。输入信号是零点标准值 x_0 。若压力传感器测量的是相对大气压 P_B 的表压 ΔP ，那么零点标准值就是通大气 P_B ，从而保证压力测量系统为零输入 $x_0 = P_B - P_B = 0$ ，测量系统的输出值 $y_0 = a_0$ 。在零输入条件下系统的输出值不为零值，必由系

统的误差源所产生：

$$y_0 = a_1 \cdot E_0$$

式中： a_1 为系统的增益； E_0 为系统的误差源。

2) 第二步：标定

实时测量系统的增益/灵敏度 a_1 。输入信号为标准值，由标准压力发生器产生标准压力值 $P_R = x_R$ ，系统的输出值为 y_R 。于是被校准系统的增益/灵敏度 a_1 为

$$a_1 = S + \Delta a = \frac{y_R - y_0}{x_R} \quad (4-12)$$

因为输出值 y_R 也含有误差源 E_0 的影响， $y_R = x_R \cdot a_1 + E_0 \cdot a_1$ ，故差值 $y_R - y_0 = x_R \cdot a_1$ ，即消除了误差源 E_0 的影响。

3) 第三步：测量

输入信号为被测目标参量压力 $P = x$ ，测量系统相应的输出值为 y_x 。因为

$$y_x = P \cdot a_1 + E_0 \cdot a_1 = P \cdot a_1 + y_0$$

故

$$P = x = \frac{y_x - y_0}{a_1} = \frac{y_x - y_0}{y_R - y_0} \cdot x_R \quad (4-13)$$

整个传感器系统的精度由标准发生器产生的标准值的精度来决定，只要求被校系统的各环节，如传感器、放大器、A/D 转换器等在三步测量所需时间内保持短暂稳定。在三步测量所需时间间隔之前和之后产生的零点漂移、灵敏度时间漂移、温度漂移等都不会引入测量误差。这种实时在线自校准可以采用低精度的传感器、放大器、A/D 转换器等环节，达到高精度测量的目的。因此，具有自校准功能的智能传感器系统实现了高精度。

上面所述实现自校准功能的方法要求被校系统的输出输入特性呈线性，即具有式(4-10)线性方程所描述的特性，这样就仅需两个标准值(其中一个为零点标准值)就能完善地标定系统的增益/灵敏度。然而，对于输入输出特性呈非线性的系统，只采用有两个标准值的三步测量法来进行自校准是不够完善的。

4.2.2 多基准法

多基准法适用于测量系统的正模型是非线性方程的系统。

1. 工作原理与实施步骤

要求在现场，在测量当时的工作条件下对传感器系统进行实时在线标定实验，确定出当时的输出输入特性，及其反非线性特性拟合方程式，并按其读数就可以消除干扰的影响。这是智能传感器系统实现自校准功能的最完善、最理想的一种方法。为了缩短实时在线标定的时间，标定点数不能多，但又要反映出输入输出特性的非线性，为此标定点不能少于三点。因此，要求标准发生器至少提供三个标准值。实时在线自校准的实施过程是：

(1) 对传感器系统进行现场、在线、测量前的实时三点标定，即依次输入三个标准值 x_{R1} 、 x_{R2} 、 x_{R3} ，测得相应输出值 y_{R1} 、 y_{R2} 、 y_{R3} 。

(2) 列出反非线性特性拟合方程式(二阶三项多项式)：

$$x(y) = C_0 + C_1 y + C_2 y^2 \quad (4-14)$$

(3) 由标定值求反非线性特性曲线拟合方程的系数 C_0 、 C_1 、 C_2 。按照最小二乘法原则，即方差最小

$$\sum_{i=1}^3 [(C_0 + C_1 y_{Ri} + C_2 y_{Ri}^2) - x_{Ri}]^2 = F(C_0, C_1, C_2) = \text{最小}$$

根据函数求极值(最小值)条件,令偏导数为零,然后再经整理后得矩阵方程

$$\begin{cases} C_0 N + C_1 P + C_2 Q = D \\ C_0 P + C_1 Q + C_2 R = E \\ C_0 Q + C_1 R + C_2 S = F \end{cases} \quad (4-15)$$

式中, $N=3$, 为在线实时标定点个数; 其余

$$\begin{aligned} P &= \sum_{i=1}^3 y_{Ri}; & Q &= \sum_{i=1}^3 y_{Ri}^2; & R &= \sum_{i=1}^3 y_{Ri}^3 \\ S &= \sum_{i=1}^3 y_{Ri}^4; & D &= \sum_{i=1}^3 x_{Ri}; & E &= \sum_{i=1}^3 x_{Ri} y_{Ri}; & F &= \sum_{i=1}^3 x_{Ri} y_{Ri}^2 \end{aligned}$$

由标定值计算出 P 、 Q 、 R 、 S 、 D 、 E 、 F 后,解式(4-15)矩阵方程可得待定常系数 C_0 、 C_1 、 C_2 。

已知 C_0 、 C_1 、 C_2 数值后,反非线性特性拟合方程式(4-14)即被确定,智能传感器系统可将测量 x 时传感器的输出值 y 按式(4-14)求出输出值 $x(y)$,即代表系统测出的输入待测目标参量 x 。因此,只要传感器系统在实时标定与测量期间保持输出输入特性不变,传感器系统的测量精度就决定于实时标定的精度,其它任何时间特性的漂移带来的不稳定性都不会引入误差。

实验证明,对一个具有零漂、温漂且在 100°C 温度变动范围内总误差达 $\pm 1\%$ 的压力传感器系统,采用满度值精度为 $\pm 0.02\%$ 的标准压力值进行实时三点校准/标定,可达到整个系统的短时精度优于 $\pm 0.1\%FS$ 。

2. 780B(PCU)压力自校准系统

780B(PCU)压力自校准系统原理示意图如图 4-6 所示。

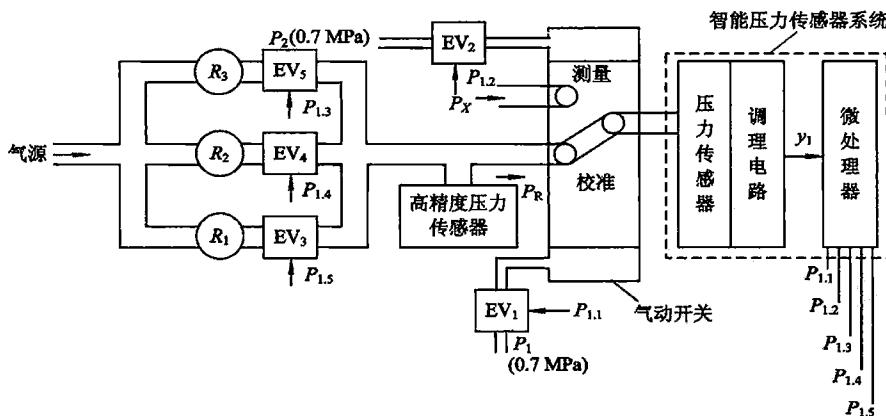


图 4-6 780B(PCU)压力自校准系统原理框图

图 4-6 中, $EV_1 \sim EV_5$ 为电驱动阀门,由微处理器经控制线 $P_{1.1}$ 、 $P_{1.2}$ 、 $P_{1.3}$ 、 $P_{1.4}$ 、 $P_{1.5}$ 发出控制信号来控制气路是“通”还是“断”。气动开关受气动控制压力 P_1 、 P_2 的控制。 P_1 与 P_2 均是 0.7 MPa 的压力。在程序控制指令控制下控制压力 P_1 工作,推动气动开关

使压力传感器与校准管路接通，处于校准状态。三个标准压力值分别为 P_{R1} 、 P_{R2} 、 P_{R3} ，按顺序施加到被校传感器上。校准结束时，控制压力 P_2 工作(P_1 放掉)，推动气动开关使传感器回到测量状态，测压孔与被测压力 P_x 接通。 R_1 、 R_2 、 R_3 是三个压力调节器，将它们事先调节到合适位置，则可得到三个不同数值的标准压力值： P_{R1} 、 P_{R2} 、 P_{R3} 。它们的精确值由高精度压力传感器读出。

测量系统的正模型用非线性方程来表示的系统，要实现自校准功能需要一套可现场实施标定实验的专用自校准系统。自校准系统应能提供与被测目标参量同类型的多个(大于等于 3 个)标准量值，且还有多点切换器依次将各个标准量值输入被校系统。在现场建立这样一个多基准可切换输入的系统对于大多数非电参量不是一件容易的事，故目前只有压力等少数几个参量的测量系统能够实现自校零与自校准功能。

对于那些尚不能实施这种完善的自校准技术的目标参量和场合，为了消除零点、灵敏度漂移以及因各种干扰因素产生的不稳定的影响，可以采用多传感器数据融合补偿技术对系统静态特性和稳定性进行综合改善，详见 4.3 节。

4.3 改善稳定性抑制交叉敏感及智能化多传感器 数据融合功能

当今，多传感器智能化技术迅速发展，已成为改善传感器系统性能的最有效的手段。多传感器智能化技术包括两大方面：

其一，将多个传感器与计算机(或微处理器)组建智能化多传感器系统；其深刻内涵是提高某点位置处(单点)某一个参量(单参量) x_1 的测量准确度，而不是一般意义的多点多参量测量系统。

其二，将多个传感器获得多个信息的数据进行融合处理，实现某种改善传感器性能的智能化功能，在抑制交叉敏感改善传感器稳定性的同时，系统的线性度也得到改善。

4.3.1 单传感器系统

通常的测量系统都是由单传感器系统组成的。它有两个基本组成部分：传感器与数据处理，其框图如图 4-7 所示。

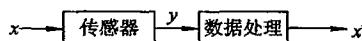


图 4-7 单传感器测量系统框图

1. 单传感器系统的正模型与逆模型

表征单传感器系统两个基本部分的输入输出关系的数学描述分别称为正、逆模型。

(1) 传感器部分含传感器及其调理电路，执行获取信息的任务。配接了调理电路的传感器部分感知并检测物理量 x ，将 x 按一定规律转换为有用输出量 y 。有用输出量 y 是指便于远距离传输的量，目前多为电量，如电压、电流、电脉冲的频率。输入 x 与输出 y 遵从一定规律是指具有一定重复性，且可用数学表达式来描述。描述传感器部分输入输出

$(x-y)$ 关系的数学表达式就称为传感器系统的正模型，单传感器系统的正模型 $y=f(x)$ 为一元多项式。

(2) 数据处理部分完成信息处理、分析及显示功能，就是从传感器输出信号的数据中将代表被测量 x_1 的有效信息提取出来并给出显示。直白地说，其最基本的功能就是将传感器部分的输出量 y 转换为被测量 x ，并给出显示。显示的值 x' 与真值 x 之间有一定偏差，希望这个偏差尽量小。数据处理部分的这个基本功能称为“刻度转换”。对于智能传感器系统，刻度转换是在计算机或微处理器中由软件实现的(详见 4.1 节)。数据处理部分输出 x' 与输入 y 关系的数学描述称为逆模型。对于单传感器系统，其逆模型为正模型 $y=f(x)$ 的反函数 $x=f^{-1}(y)$ ，也是一个一元多项式。

2. 单传感器系统的应用

单传感器系统既可用于单点单参量测量系统，也可组建多点多参量测量系统。

(1) 单点单参量测量系统：选用标称的目标参量与被测参量 x 同名的传感器，如压力传感器可用来测量压力一个参量，而且仅是压力传感器所在位置(单点)处的压力参量。通常的压力测量系统、位移测量系统、液位测量系统等一般都是单传感器系统。

(2) 多点多参量测量系统：由多个单传感器测量系统组成，其原理框图如图 4-8 所示。

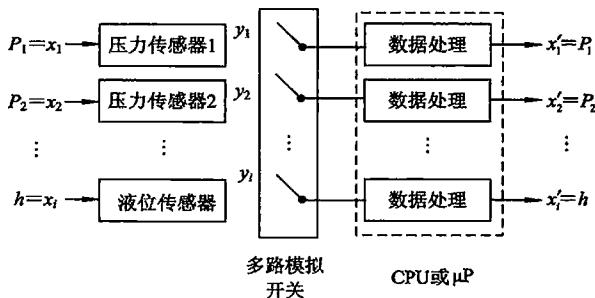


图 4-8 由单传感器系统组成的多点多参量测量系统

图 4-8 系统中的每个传感器只能完成单点单参数测量，两个位置点的压力需要两个压力传感单元来检测。通过一个多路模拟开关将各点传感器输出的信号 y_1, y_2, \dots, y_i 按一定顺序或同时并行导入计算机或微处理器，进行独立的数据处理按各自的逆模型完成各自的刻度转换，并输出被测参量 P_1, P_2, \dots, h 。图 4-8 中各单传感器系统的正模型 $y_1 = f(x_1 = P_1), y_2 = f(x_2 = P_2), y_i = f(x_i = h)$ ，与各自的逆模型 $P_1 = f^{-1}(y_1), P_2 = f^{-1}(y_2), h = f^{-1}(y_i)$ 相对应。模型的数学表达式均为一元多项式，故图 4-8 所示的通常意义的多点多参量测量系统其实质仍然是单传感器系统。

4.3.2 交叉敏感与传感器系统的稳定性

1. 交叉敏感现象

不论是传统工艺制作的经典传感器，还是半导体工艺制作的现代传感器，都存在交叉敏感。交叉敏感是引起单传感器系统不稳定的主要因素，表现为传感器标称的目标参量恒

定不变，而其它非目标参量变化时，该传感器的输出值发生变化。如半导体氢传感器，其标称的目标参量——氢气浓度恒定，而如果一氧化碳气体浓度发生变化，则其输出值也会发生变化。这就是氢传感器存在对非目标参量——一氧化碳气体的交叉敏感。

几乎所有的传感器都存在对温度的交叉敏感且不仅仅有一个交叉敏感量。以压力传感器为例，其标称的目标参量——压力恒定，而传感器的环境温度 T 或供电电压 U /电流 I 变化时，其输出电压值发生变化，表明压力传感器存在对环境温度 T 及供电电压 U /电流 I 两个非目标参量的交叉敏感。

2. 交叉敏感带来的问题

由交叉敏感现象说明，作为单传感器系统的正、逆模型

$$y = f(x), \quad x = f^{-1}(y)$$

用一元多项式方程来表征是不完备的。上述对温度 T 、供电电流 I 具有交叉敏感的压力传感器，其正模型至少应用三元多项式来表征：

$$y = u = f(P = x, T, I) \quad (4-16)$$

相对应的逆模型

$$P = x = f^{-1}(y = u, T, I). \quad (4-17)$$

也应用三元多项式表征才较完备。否则，由于正模型不能完备代表多元交叉敏感的实际传感器系统，再根据不完备的正模型建立的逆模型获得的被测量值将会有很大的误差。以一个干扰量为例，上述压力传感器当供电电流 I 恒定时，不同温度条件下的正、逆模型 $u=f(P)$ 与 $P=f^{-1}(u)$ 如图 4-9 所示。

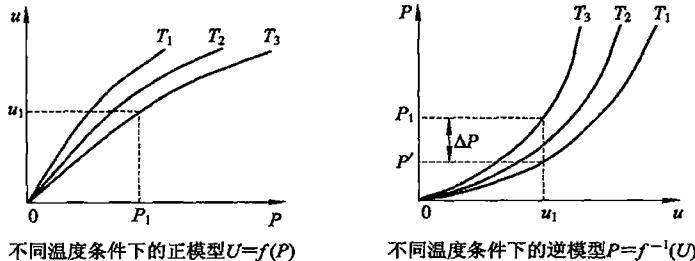


图 4-9 不同温度条件下压力传感器的正模型与逆模型

由图 4-9 清楚可见，当传感器系统工作环境温度从 T_1 变至 T_3 ，其特性也随之漂移，若仍按 T_1 时之正、逆模型求取测量值 P' ，则与实际值 P_1 之间存在很大偏离 ΔP 。存在交叉敏感的传感器系统性能不稳定，准确性差。这是常规单传感器系统普遍存在的问题。

4.3.3 多传感器技术改善传感器系统性能的基本方法

多传感器技术改善传感器系统性能的基本方法有二：模型法和冗余法。

1. 模型法

以改善传感器稳定性，消除干扰量交叉敏感的影响为例，模型法的基本思路是：当主测参量为 x_1 的传感器存在干扰量 x_2 时，若欲消除干扰量 x_2 的影响，则需监测该干扰参量 x_2 ，从而建立测量 x_1 与 x_2 的多(2个)传感器系统；若欲消除 n 个干扰量的影响，则需建立

测量 $n+1$ 个参量的多($n+1$ 个)传感器系统。基于模型法改善稳定性，消除两个干扰量影响的三传感器—智能传感器系统框图如图 4-10 所示。

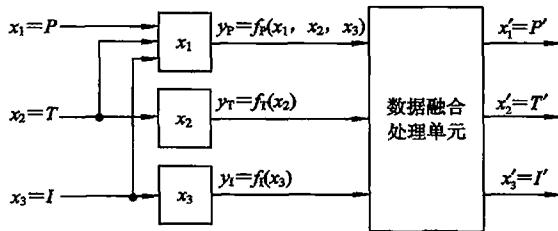


图 4-10 基于模型法的三传感器—智能传感器系统

(1) 传感器单元： x_1 为主传感器及其调理电路单元。设其目标参量为压力 $x_1 = P$ 。 x_2 、 x_3 分别为辅传感器及其调理电路单元，它们的目标参量分别是温度 $x_2 = T$ 与电流 $x_3 = I$ ，这些参量是主传感器的干扰量。每个传感器的输出分别为：

主传感器(压力) $y_P = f_P(x_1, x_2, x_3) = f_P(P, T, I)$ ，是三元函数模型；

辅传感器(温度) $y_T = f_T(x_1, x_2, x_3) = f_T(T)$ ，可用一元函数模型近似；

辅传感器(电流) $y_I = f_I(x_1, x_2, x_3) = f_I(I)$ ，可用一元函数模型近似。

图 4-10 中的多传感器—智能传感器系统，是为消除 $n=2$ 个干扰量(温度 T 、电流 I)改善压力传感器 $x_1=P$ (压力)而建立的($m=3$)三传感器—智能传感器系统。系统中传感器的总数

$$m = n + 1 \quad (4-18)$$

其中 n 为欲消除的干扰量数。

(2) 数据融合处理单元：图 4-10 中的数据融合处理单元是存入计算机内进行数据融合的智能化软件模块。该模块实现由 $m=3$ 个传感器输出的数据 y_P 、 y_T 、 y_I 求目标参量 $x_1=P$ 的某种融合算法。根据已建立的逆模型

$$x_1 = P = g(y_P, y_T, y_I) \quad (4-19)$$

计算被测目标参量，其计算所得的值 $P' = x_1'$ 消除了干扰量 T 与 I 的影响，更接实际值 P 。

不仅如此，模型法在消除交叉敏感提高传感器系统稳定性的同时，也进行了非线性校正，系统的线性度也得到改善(详见[示例 5-1])。

建立逆模型的方法有多种，本书介绍常用的两种方法。一种是回归分析法(在第 5 章详述)；另一种姑且称为机器学习算法，如神经网络机器学习算法(在第 6 章详述)、支持向量机学习算法(在第 7 章中详述)。根据不同的神经网络又有 BP 误差反向传播神经网络、RBF 径向基函数神经网络、GANN 遗传神经网络等及其派生算法。神经网络与支持向量机的结构优化方法有多种，在第 8 章将介绍一种采用粒子群算法的最新优化方法。

2. 冗余法

采用冗余法消除干扰量影响，改善传感器稳定性基本思路是：不去监测主测参量为 x_1 的传感器的干扰量，不去探究干扰量对主测参量 x_1 传感器的影响规律，而是采用与主测参量 x_1 同类的多个(至少 3 个)传感器建立测量主测参量 x_1 的多传感器系统。基于冗余

法消除传感器漂移改善稳定性的多传感器—智能传感器系统框图如图 4-11 所示。

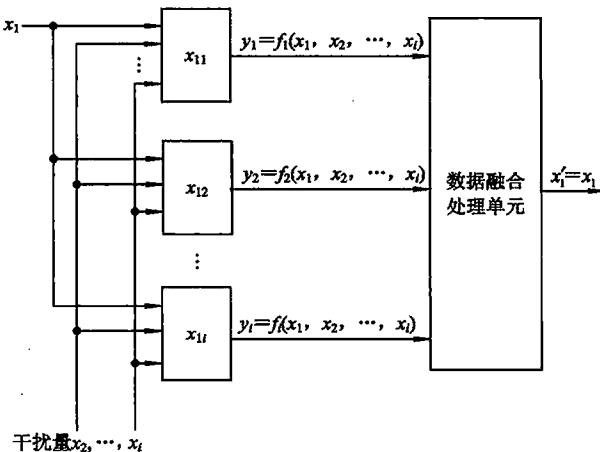


图 4-11 基于冗余法的三传感器—智能传感器系统框图

(1) 传感器单元均为主测同一参量 x_1 的传感器，它们的输出均受干扰量 x_2, \dots, x_i 的影响，每个传感器的输出分别为：

$$\text{传感器 } x_{11} \quad y_1 = f_1(x_1, x_2, \dots, x_i)$$

$$\text{传感器 } x_{12} \quad y_2 = f_2(x_1, x_2, \dots, x_i)$$

⋮

$$\text{传感器 } x_{1i} \quad y_i = f_i(x_1, x_2, \dots, x_i)$$

传感器数量 $i \geq 3$

(2) 数据融合处理单元。图 4-11 中的数据融合处理单元是在计算机中进行数据融合处理的智能化软件模块，本书第 9 章将介绍为消除传感器性能漂移、提高稳定性而采用的多传感器数据主元分析融合算法。

4.4 改善动态性能扩展频带及智能化频率自补偿功能

我们希望传感器输入时变信号 $x(t)$ 与其响应输出信号 $y(t)$ 二者的幅值比在不同的频率都保持相等，各次谐波分量的相位差均相等。这样，由输出信号 $y(t)$ 才能无失真地反映被测输入信号 $x(t)$ 。但是，实际的传感器系统 $x(t)$ 与 $y(t)$ 只能在一定频率范围内，在允许的动态误差条件下保持所谓的一致。回顾 2.1.2 节中所述，常见的一阶、二阶传感器系统在不同频率 $\omega=2\pi f$, $y(t)$ 与 $x(t)$ 因幅值比不同产生的动态幅值误差如表 4-3 所示。表中 ω 为被测量的角频率； $\omega_c=1/\tau$ 为一阶传感器转折角频率， τ 为其时间常数； ω_0 为二阶传感器系统无阻尼固有振荡角频率。

表 4-3 信号频率与动态幅值误差的关系

一阶系统	频率比 ω/ω_c	1/10	1/7	1/6	1/5	1
	动态幅值误差 $ \gamma $	0.5%	1%	1.4%	2%	29.3%
二阶系统 $0 < \zeta < 1$	频率比 ω/ω_0	1/10	1/7	1/6	1/5	1
	动态幅值误差 $ \gamma $	1%	2%	3%	5%	10%

由表 4-3 可见, 如果想保证信号通过传感器系统后产生的幅值误差 $|\gamma| < 2\%$, 则二阶系统的无阻尼固有振荡频率 f_0 应比信号频率 f 大 7 倍, 即 $\omega/\omega_0 = f/f_0 = 1/7$; 若是一阶系统, 则转折频率 f_c 应比信号频率 f 大 5 倍, 即 $\omega/\omega_c = f/f_c = 1/5$ 。所以, 当信号的频率高, 而传感器系统的工作频带不能满足测量允许误差的要求时, 就希望扩展系统的频带以改善系统的动态性能。以计算机为中心的数据采集系统与传感器相结合的现代测控系统具有强大的软件优势, 能够补偿原传感器系统动态性能的不足, 将系统频带扩展。其主要采用两种方法实现频率自补偿: 数字滤波法与频域校正法。与硬件扩展频带技术相比, 软件智能化频率自补偿技术方法简便、灵活、调试容易。

利用计算机完成数字信号处理, 从广义上讲就是对输入的数字信号进行数字滤波, 数字滤波可看成是一种将一组输入的数字序列通过 Z 域的传递函数 $W(z)$ 作用后转变成另一组数字序列输出的装置, 称为数字滤波器。数字滤波器不仅仅用于计算机内的软件对采集到的传感器信号进行滤波消噪, 而且还可用于构建一种预期频率特性以补偿原传感器频率特性频带窄的不足, 实现传感器系统频带扩展的作用, 这些都需要在计算机中设计一个有预期频率特性其模型由 Z 域传递函数 $W(z)$ 表征的数字滤波器。

为了设计一个能对输入数字信号进行传递、转换与处理的数字滤波器, 需要有一定 Z 变换的基础知识。

4.4.1 数字滤波器的数学基础——Z 变换简介

广义地说, 任一系统都可以看成一个滤波器。它按所要求的功能执行对输入信号的传递转换与处理。

对于连续时间系统, 如图 4-12 所示, 它的输入(激励)信号 $x(t)$ 与其输出(响应)信号 $y(t)$ 的关系在时域中可由微分方程确立。因此, 在时域中是通过解微分方程来求解系统对激励的时间响应 $y(t)$ 的。通过拉氏变换 $X(s)=L[x(t)]$, $Y(s)=L[y(t)]$, 建立输入 $X(s)$ 与输出 $Y(s)$ 的关系, 在 $s=\sigma+j\omega$ 域中微分方程变为代数方程:

$$Y(s) = W(s)X(s) \quad (4-20)$$

当 $\sigma=0$ 时, 则为频域 $s=j\omega$, 式(4-20)变为

$$Y(j\omega) = W(j\omega)X(j\omega) \quad (4-21)$$

$X(j\omega)$ 、 $Y(j\omega)$ 分别是 $x(t)$ 、 $y(t)$ 的傅里叶正变换。

$W(s)$ 、 $W(j\omega)$ 分别称为系统的传递函数与频率特性。广义地说, 也称为模拟滤波器特

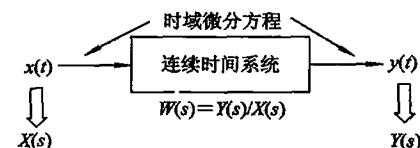


图 4-12 连续时间系统

性。 s 域是复频域，拉普拉斯变换是处理分析连续时间系统的有力工具。

对于具有采样/保持(S/H)功能的计算机系统，输入的连续时间信号 $x(t)$ 变为时间序列 $x(nT)$ ，计算机的输出也是时间序列 $y(nT)$ 。由于 $x(nT)$ 、 $y(nT)$ 都是不连续的，故微分方程不能使用，当时间间隔 T 很小时，可近似由差分方程描述。

通过求 Z 变换， $X(z) = Z[x(nT)]$ ，
 $Y(z) = Z[y(nT)]$ ，则在 Z 域中有关系：

$$W(z) = \frac{Y(z)}{X(z)} \quad (4-22)$$

式中， $W(z)$ 称为离散时间系统(见图 4-13)

的传递函数，也是广义离散时间滤波器特性。

当幅值也被量化时(系统中有 A/D)，则称其为数字滤波器。

Z 变换是处理、分析离散时间系统的有力工具，其作用与拉氏变换在连续时间系统中的作用类似。在线性离散时间(测控)系统中，线性差分方程表征系统动力学性质。

1. 数值序列 $x(nT)$ 或 $x(n)$ 的 Z 变换

对函数 $x(t)$ 求 Z 变换，就是对 $t=0, 1, \dots, nT$ 离散时刻处的离散时间序列 $x(nT)$ 求 Z 变换，记为 $X(z)$ ，其定义式为

$$\begin{aligned} X(z) &= Z[x(t)] = Z[x(nT)] = Z[x(n)] = \sum_{n=-\infty}^{\infty} x(n)z^{-n} \\ &= \sum_{n=-\infty}^{-1} x(n)z^{-n} + \sum_{n=0}^{\infty} x(n)z^{-n} \end{aligned} \quad (4-23)$$

式中 $z = e^{j\omega T}$ 为复变量，代入上式可得

$$X(z) = X[e^{j\omega t}] = \sum_{n=-\infty}^{\infty} x(n)e^{-j\omega nT} \quad (4-24)$$

式(4-24)就是傅里叶变换的离散形式。

2. 某些典型函数的 Z 变换

(1) 单位阶跃函数的 Z 变换。如图 4-14 所示单位阶跃函数：

$$x(t) = \begin{cases} 1, & t \geq 0 \\ 0, & t < 0 \end{cases}$$

由定义式得

$$\begin{aligned} x(z) &= Z[x(t)] = \sum_{n=0}^{\infty} x(n)z^{-n} \\ &= x(0)z^0 + x(1)z^{-1} + x(2)z^{-2} + \dots \\ &= 1 + z^{-1} + z^{-2} + \dots = \frac{1}{1 - z^{-1}} = \frac{z}{z - 1} \end{aligned}$$

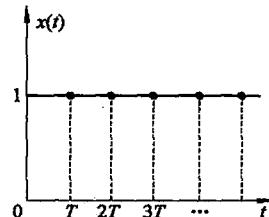


图 4-14 单位阶跃函数

(2) 单位斜坡函数 $x(t)$ 的 Z 变换。如图 4-15 所示单位斜坡函数：

$$x(t) = \begin{cases} t, & t \geq 0 \\ 0, & t < 0 \end{cases}$$

因为

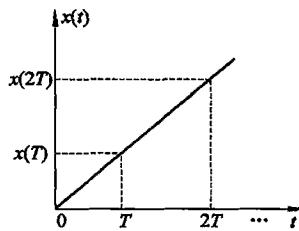


图 4-15 单位斜坡函数

$$x(nT) = nT, \quad n = 0, 1, 2, \dots$$

故

$$\begin{aligned} X(z) &= Z[x(t)] = \sum_{n=0}^{\infty} x(n) z^{-n} \\ &= x(0)z^{-0} + x(1)z^{-1} + \dots \\ &= T(z^{-1} + 2z^{-2} + 3z^{-3} + \dots) \\ &= T \frac{z^{-1}}{(1-z^{-1})^2} = \frac{Tz}{(z-1)^2} \end{aligned}$$

(3) 指数函数 $x(t)$ 的 Z 变换。对于指数函数：

$$x(t) = \begin{cases} e^{-at}, & t \geq 0 \\ 0, & t < 0 \end{cases}$$

因为

$$x(nT) = e^{-anT}, \quad n = 0, 1, 2, \dots$$

所以有

$$\begin{aligned} x(z) &= Z[e^{-at}] = \sum_{n=0}^{\infty} x(nT) z^{-n} \\ &= 1 + e^{-aT} z^{-1} + e^{-2aT} z^{-2} + e^{-3aT} z^{-3} + \dots \\ &= \frac{1}{1 - e^{-aT} z^{-1}} = \frac{z}{z - e^{-aT}} \end{aligned}$$

(4) 余弦函数的 Z 变换。对于余弦函数：

$$x(t) = \begin{cases} \cos \omega t, & t \geq 0 \\ 0, & t < 0 \end{cases}$$

因为

$$\cos \omega t = \frac{1}{2} (e^{j\omega t} + e^{-j\omega t})$$

所以有

$$\begin{aligned} X(z) &= Z[\cos \omega t] = \frac{1}{2} \left[\frac{1}{1 - e^{j\omega T} z^{-1}} + \frac{1}{1 - e^{-j\omega T} z^{-1}} \right] \\ &= \frac{1}{2} \frac{2 - (e^{-j\omega T} + e^{j\omega T}) z^{-1}}{1 - (e^{-j\omega T} + e^{j\omega T}) z^{-1} + z^{-2}} = \frac{1 - z^{-1} \cos \omega T}{1 - 2z^{-1} \cos \omega T + z^{-2}} \end{aligned}$$

(5) 求 Z 变换：

$$X(s) = \frac{1}{s(s+1)}$$

从 s 域求相应 Z 变换的途径是把 $X(s)$ 反变回 $x(t)$, 然后再求 $x(t)$ 的 Z 变换。

对 $X(s) = \frac{1}{s(s+1)}$ 求拉普拉斯反变换, 得原函数

$$x(t) = \begin{cases} 1 - e^{-t}, & t \geq 0 \\ 0, & t < 0 \end{cases}$$

故有

$$X(z) = Z[1 - e^{-t}] = \frac{1}{1 - z^{-1}} - \frac{1}{1 - e^{-T}z^{-1}}$$

应当注意: 虽然 Z 变换 $X(z)$ 与它的 Z 反变换 $x(n)$ 一一对应, 但对应不是单值的, 即若 $x(t)$ 的 Z 变换是 $x(z)$, 其 Z 反变换未必是 $x(t)$ 。如单位阶跃函数的 Z 变换是 $1/(1 - z^{-1})$, 然而 $1/(1 - z^{-1})$ 的 Z 反变换也可能是任一时间函数, 但该时间函数在 $t=0, T, 2T, 3T, \dots$ 处却也都是有相同的值 1, 在满足采样定理的条件下, 通过 $x(n)$ 那些点的平滑曲线可近似为连续时间函数 $x(t)$, 如图 4-16 所示。

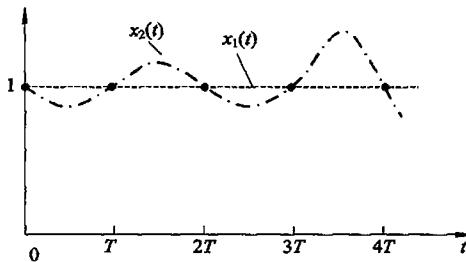


图 4-16 不同的时间函数 $x_1(t)$ 、 $x_2(t)$ 有相同的时间序列值

3. Z 反变换

$X(z)$ 的 Z 反变换产生相应的时间序列值 $x(n)$, 记为 z^{-1} 。作为由 $X(z)$ 到 $x(n)$ 的一种运算, 除了可以查表之外, 常有 4 种方法。

1) 直接除法

直接除法又称幂级数展开法, 它只能求得 $x(n)$ 的前几项数值, 得不出 $x(n)$ 的通项表达式。直接除法源于 Z 变换定义式:

$$X(z) = \sum_{n=0}^{+\infty} x(n)z^{-n} = x(0) + x(T)z^{-1} + x(2T)z^{-2} + \dots + x(nT)z^{-n} + \dots$$

当把 $X(z)$ 展开为 z^{-1} 的幂级数时, $x(n)$ 的数值可通过对照方法确定。

[例 1] 试求 $X(z)$ 的 Z 反变换 $x(n)$, $n=0, 1, 2, 3, 4$ 。

$$X(z) = \frac{10z + 5}{(z - 1)(z - 0.2)}$$

解 (1) 先将 $X(z)$ 写成 z^{-1} 的多项式之比:

$$X(z) = \frac{10z + 5}{(z - 1)(z - 0.2)} = \frac{10z^{-1} + 5z^{-2}}{(1 - z^{-1})(1 - 0.2z^{-1})} = \frac{10z^{-1} + 5z^{-2}}{1 - 1.2z^{-1} + 0.2z^{-2}}$$

(2) 分子被分母除, 得

$$X(z) = 10z^{-1} + 17z^{-2} + 18.4z^{-3} + 18.68z^{-4} + \dots$$

(3) 由对比可得

$$x(0) = 0, \quad x(1) = 10, \quad x(2) = 17, \quad x(3) = 18.4, \quad x(4) = 18.68$$

2) 留数法(积分反演法)

这是求解 Z 反变换最一般的方法。

$$x(n) = K_1 + K_2 + \cdots + K_m = \sum_{i=1}^m R_{es}[X(z)z^{n-1} \text{ 在其极点 } z = z_i \text{ 处的留数}] \quad (4-25)$$

(1) 如果 $X(z)z^{n-1}$ 分母包含一个单极点 $z = z_i$ 时, 则对应的留数 K 为

$$K = \lim_{z \rightarrow z_i} (z - z_i) X(z) z^{n-1} \quad (4-26)$$

(2) 若 $X(z)z^{n-1}$ 含有 m 阶重极点 z_i , 则其留数 K 为

$$K = \frac{1}{(m-1)!} \lim_{z \rightarrow z_i} \frac{d^{m-1}}{dz^{m-1}} [(z - z_i)^m X(z) z^{n-1}] \quad (4-27)$$

应当指出, 只要 $X(z)z^{n-1}$ 在原点 $z=0$ 处无极点, 则留数法是求取 Z 反变换的一种非常简便的方法。

[例 2] 求 $X(z) = \frac{z^2}{(z-1)(z-0.5)}$, $|z| > 1$ 的反变换。

解 (1) $X(z)z^{n-1} = \frac{z^{n+1}}{(z-1)(z-0.5)}$, 极点: $z_1 = 1$, $z_2 = 0.5$, 均为单极点。

(2) 单极点的留数 K_1 、 K_2 由式(4-26)有:

$$K_1 = [\text{在单极点 } z = 1 \text{ 处的留数}] = \lim_{z \rightarrow 1} \left[(z-1) \frac{z^{n+1}}{(z-1)(z-0.5)} \right] = 2$$

$$\begin{aligned} K_2 &= [\text{在单极点 } z = 0.5 \text{ 处的留数}] = \lim_{z \rightarrow 0.5} \left[(z-0.5) \frac{z^{n+1}}{(z-1)(z-0.5)} \right] \\ &= \frac{(0.5)^{n+1}}{-0.5} = -\left(\frac{1}{2}\right)^n \end{aligned}$$

$$x(n) = K_1 + K_2 = 2 - \left(\frac{1}{2}\right)^n$$

故

$$x(0) = 1, x(1) = 1.5, x(2) = 1.75, x(3) = 1.875, \dots$$

3) 部分分式展开法

如果在坐标原点 ($z=0$) 处, $X(z)$ 有一个或多个零点的话, 则由部分分式展开法将 $X(z)/z$ 或 $X(z)$ 展开成简单的一阶或二阶项之和, 再查 Z 变换表找出每个展开项对应的 Z 反变换。这是利用了 Z 变换的线性性质, 各部分分式 Z 反变换之和即为 $X(z)$ 的 Z 反变换。讨论由下式给出的 $X(z)$:

$$X(z) = \frac{b_0 z^m + b_1 z^{m-1} + \cdots + b_{m-1} z + b_m}{z^n + a_1 z^{n-1} + \cdots + a_{n-1} z + a_n}, \quad m \leq n \quad (4-28)$$

首先对 $X(z)$ 的分母多项式进行因式分解, 并求其极点:

$$X(z) = \frac{b_0 z^m + b_1 z^{m-1} + \cdots + b_{m-1} z + b_m}{(z - p_1)(z - p_2) \cdots (z - p_n)}$$

在 $X(z)$ 全部是低阶极点, 并且至少有一个零点是在原点 ($b_m = 0$) 的情况下, 一般的步骤是将 $X(z)/z$ 展开成部分分式如下:

$$\frac{X(z)}{z} = \frac{a_1}{z-p_1} + \frac{a_2}{z-p_2} + \cdots + \frac{a_n}{z-p_n} = \sum_{i=1}^n \frac{a_i}{z-p_i}$$

解得

$$a_i = \left[(a - p_i) \frac{X(z)}{z} \right]_{z=p_i} \quad (4-29)$$

为各极点的留数。

若在 $z = p_i$ 处 $\frac{X(z)}{z}$ 有二重极点，并且无其他极点，那么

$$\frac{X(z)}{z} = \frac{C_1}{(z-p_1)^2} + \frac{C_2}{z-p_1}$$

则

$$C_2 = \left[(z-p_1) \frac{X(z)}{z} \right]_{z=p_1}, \quad C_1 = \left[\frac{d}{dz} (z-p_1)^2 X(z)/z \right]_{z=p_1}$$

[例 3] 求 $X(z) = \frac{10z}{(z-1)(z-0.2)}$ 的 Z 反变换。

$$\text{解 } (1) \frac{X(z)}{z} = \frac{10}{(z-1)(z-0.2)} = \frac{a_1}{z-1} + \frac{a_2}{z-0.2}$$

(2) 求 a_i :

$$a_1 = \lim_{z \rightarrow 1} (z-1) \frac{10}{(z-1)(z-0.2)} = \frac{10}{0.8} = 12.5$$

$$a_2 = \lim_{z \rightarrow 0.2} (z-0.2) \frac{10}{(z-1)(z-0.2)} = -12.5$$

(3)

$$\frac{X(z)}{z} = \frac{12.5}{z-1} - \frac{12.5}{z-0.5}$$

所以

$$X(z) = 12.5 \left(\frac{z}{z-1} - \frac{z}{z-0.5} \right)$$

(4) 查表可得

$$x(n) = 12.5(1^n - 0.2^n), \quad n \geq 0$$

$$x(0) = 0, x(1) = 10, x(2) = 12, x(3) = 12.48, \dots$$

4. Z 变换的某些基本性质和定理

借助 Z 变换基本性质(定理)，使得应用 Z 变换法分析离散时间系统问题更加方便。

1) 线性特性

若时间序列 $x_1(n)$ 的 Z 变换为 $Z[x_1(n)] = X_1(z)$ ， $x_2(n)$ 的 Z 变换为 $Z[x_2(n)] = X_2(z)$ ，则

$$Z[x(n)] = Z[\alpha_1 x_1(n) + \alpha_2 x_2(n)] = X(z) = \alpha_1 X_1(z) + \alpha_2 X_2(z) \quad (4-30)$$

式中 α_1 、 α_2 为常数。

2) 延时性质

若 $X(z) = Z[x(n)]$ ，则

$$(1) \quad Z[x(n-m)] = z^{-m} X(z) = X_1(z) \quad (4-31)$$

$$(2) \quad Z[x(n+m)] = z^m \left[X(z) - \sum_{n=0}^{m-1} x(n)z^{-n} \right] \quad (4-32)$$

证明：

$$\begin{aligned} (1) \quad Z[x(n-m)T] &= z^{-m} \sum_{n=0}^{\infty} x(n-m)z^{-n}z^m = z^{-m} \sum_{n=0}^{\infty} x(n-m)z^{-(n-m)} \\ &= z^{-m} \sum_{k=-m}^{\infty} x(k)z^{-k} \quad (\text{设 } k = n - m) \\ &= z^{-m} \sum_{k=0}^{\infty} x(k)z^{-k} + z^{-m} \sum_{k=-m}^{-1} x(k)z^{-k} \quad (\text{设 } k < 0 \text{ 时, } x(k) = 0) \\ &= z^{-m} X(z) \end{aligned}$$

所以, Z 变换乘以 z^{-m} 后, 时间序列 $x(n)$ 将延迟一段时间(mT 个周期), 见图 4-17。

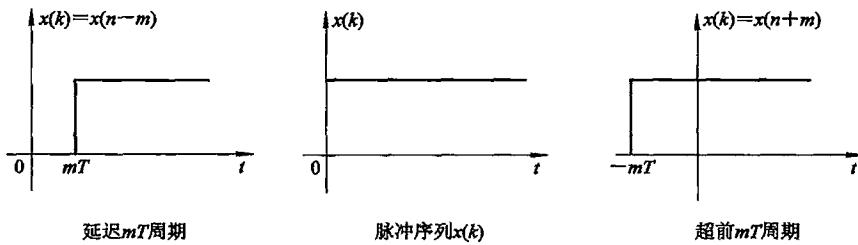


图 4-17 脉冲序列 $x(k)$ 的超前与延迟

$$\begin{aligned} (2) \quad Z[x(n+m)T] &= z^m \sum_{n=0}^{\infty} x(n+m)z^{-n}z^{-m} = z^m \sum_{n=0}^{\infty} x(n+m)z^{-(n+m)} \\ &= z^m \left[\sum_{k=-m}^{\infty} x(k)z^{-k} \right] \quad (\text{设 } k = n + m, n = 0 \text{ 时 } k = m) \\ &= z^m \left[\sum_{k=-m}^{\infty} x(k)z^{-k} + \sum_{k=0}^{m-1} x(k)z^{-k} - \sum_{k=0}^{m-1} x(k)z^{-k} \right] \\ &= z^m \left[\sum_{k=0}^{\infty} x(k)z^{-k} - \sum_{k=0}^{m-1} x(k)z^{-k} \right] \\ &= z^m \left[X(z) - \sum_{n=0}^{m-1} x(n)z^{-n} \right] \end{aligned}$$

即

$$Z[x(n+m)] = z^m X(z) - z^m x(0) - z^{m-1} x(1) - \cdots - zx(m-1)$$

当 $m=1$ 时, $Z[x(n+1)] = zX(z) - zx(0)$;

当 $m=2$ 时, $Z[x(n+2)] = z^2 X(z) - z^2 x(0) - zx(1)$ 。

3) 卷积定理

(1) 离散卷积分。两个时间序列 $x_1(n)$ 、 $x_2(n)$ 的卷积分为

$$x(n) = \sum_{l=-\infty}^{\infty} x_1(l)x_2(n-l) = \sum_{l=-\infty}^{\infty} x_2(l)x_1(n-l) \quad (4-33)$$

两个序列的卷积与进行卷积的两个序列的次序无关, 记为 $x(n) = x_1(n) * x_2(n)$ 。

(2) 时域卷积定理。若

$$Z[x_1(n)] = X_1(z)$$

$$\begin{aligned} Z[x_2(n)] &= X_2(z) \\ \text{则 } Z[x_1(n) * x_2(n)] &= X_1(z)X_2(z) \end{aligned} \quad (4-34)$$

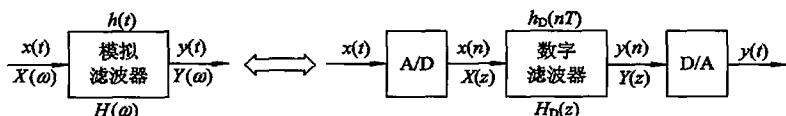
式(4-34)表明序列 $x_1(n)$ 和 $x_2(n)$ 卷积的 Z 变换等于两序列 Z 变换之乘积。典型序列的 Z 变换如表 4-4 所示。

表 4-4 典型序列的 Z 变换

序号	$x(n)$ ($n > 0$)	$X(z)$	序号	$x(n)$ ($n > 0$)	$X(z)$
1	$\delta(n)$	1	7	$ne^{-\alpha nT}$	$(ze^{-\alpha T})(z - e^{-\alpha T})^2$
2	a^n	$\frac{z}{z-a}$	8	$\sin\omega nT$	$\frac{z \sin\omega T}{z^2 - 2z \cos\omega T + 1}$
3	$u(n)$	$\frac{z}{z-1}$	9	$\cos\omega nT$	$\frac{z(z - \cos\omega T)}{z^2 - 2z \cos\omega T + 1}$
4	n	$\frac{z}{(z-1)^2}$	10	$e^{-\alpha nT} \sin\omega nT$	$\frac{ze^{-\alpha T} \sin\omega T}{z^2 - 2ze^{-\alpha T} \cos\omega T + e^{-2\alpha T}}$
5	$e^{-\alpha nT}$	$\frac{z}{z - e^{-\alpha T}}$	11	$e^{-\alpha nT} \cos\omega nT$	$\frac{z(z - e^{-\alpha T} \cos\omega T)}{z^2 - 2ze^{-\alpha T} \cos\omega T + e^{-2\alpha T}}$
6	na^n	$\frac{az}{(z-a)^2}$	12	$a^{n-1}, n=1, 2, \dots$	$\frac{z^{-1}}{1 - az^{-1}}$

5. 连续时间(t)滤波器 $H(s)$ 与其等效离散时间(nT)滤波器 $H_D(z)$

幅值整量化了的离散时间滤波器称为数字滤波器 $H_D(z)$ 。连续时间滤波器即为模拟滤波器 $H(s)$ 。求图 4-18 的 $H(s)$ 的等效 $H_D(z)$ 的方法有多种, 详见 4.5 节, 本节仅给出两种方法的关键步骤。

图 4-18 $H(s)$ 与 $H_D(z)$ 的等效

(1) 后向差分法求等效 $H_D(z)$ 的方法是: 令 $H(s)$ 中的 s 为

$$s = \frac{1 - z^{-1}}{T} \quad (T \text{ 为采样间隔}) \quad (4-35)$$

(2) 双线性变换法求等效 $H_D(z)$ 的方法是令 $H(s)$ 中的 s 为

$$s = \frac{2}{T} \cdot \frac{1 - z^{-1}}{1 + z^{-1}} = \frac{2}{T} \cdot \frac{z - 1}{z + 1} \quad (T \text{ 为采样间隔}) \quad (4-36)$$

4.4.2 扩展频带的数字滤波法

1. 数字滤波法扩展频带实现频率自补偿功能的思路

数字滤波法的补偿思路是: 给现有传递函数为 $W(s)$ 的待补偿系统串接一个传递函数为 $H(s)$ 的环节, 于是系统总传递函数 $I(s) = W(s)H(s)$ 满足动态性能的要求。其实质就是

一种“拼凑”法。这个需要附加的串联环节 $H(s)$ ，由软件编程设计的等效数字滤波器实现。下面以一阶环节为例说明数字滤波法实现扩展频带的原理。

欲将某一阶环节的传感器频带扩展 A 倍以上，其传递函数 $W(s)$ 和频率特性 $W(j\omega)$ 为

$$W(s) = \frac{1}{1 + \tau s}, \quad W(j\omega) = \frac{1}{1 + j\omega\tau} \quad (4-37)$$

现欲将其频带扩展 A 倍，即扩展后转折角频率 ω'_r 为

$$\omega'_r = A\omega_r$$

也就是时间常数减小 A 倍，即

$$\tau' = \frac{\tau}{A}$$

我们通过附加一个称为校正环节的串联环节达到上述目的。

2. 校正环节传递函数 $H(s)$ 的确定

串入校正环节 $H(s)$ 后，与原传感器 $W(s)$ 组成一个新环节 $I(s)$ ，如图 4-19 所示。 $I(s)$ 应具有所希望的动态特性，即

$$I(s) = \frac{Y(s)}{X(s)} = \frac{Y(s)U(s)}{U(s)X(s)} = W(s)H(s) = \frac{1}{1 + \tau' s} \quad (4-38)$$

于是，得校正环节的传递函数 $H(s)$ 为

$$H(s) = \frac{I(s)}{W(s)} = \frac{1 + \tau s}{1 + \tau' s} \quad (4-39)$$

式中， $\tau' \leq \frac{\tau}{A}$ 为串联了校正环节 $H(s)$ 后总的时间常数。

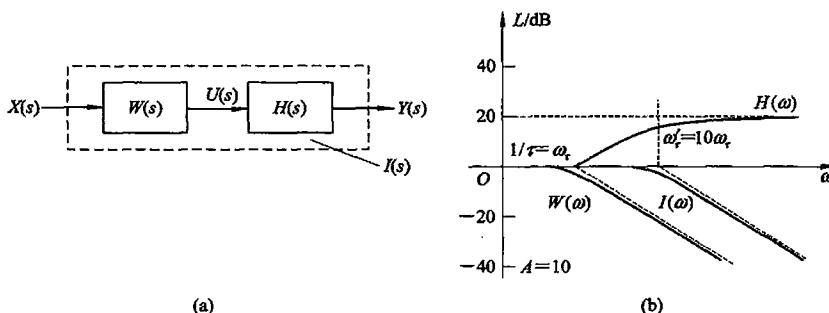


图 4-19 串联校正环节
(a) 系统框图；(b) $H(\omega)$ 、 $W(\omega)$ 、 $I(\omega)$ 对数幅频图

3. $H(s)$ 等效数字滤波器 $H(z)$ 的设计

1) 等效数字滤波器 $H(z)$ 的 Z 域传递函数表达式

采用后向差分法，令

$$s = \frac{1}{T}(1 - z^{-1})$$

代入式(4-39)可得模拟滤波器 $H(s)$ 的等效数字滤波器 $H(z)$ 如下：

$$H(z) = \frac{Y(z)}{U(z)} = H(s) \Big|_{s=\frac{1}{T}(1-z^{-1})} = \frac{1 + \tau s}{1 + \tau' s} \Big|_{s=\frac{1}{T}(1-z^{-1})}$$

整理后得

$$H(z) = A \frac{1 + cT - z^{-1}}{1 + bT - z^{-1}} = \frac{Y(z)}{U(z)}, \quad c = \frac{1}{\tau}, \quad b = Ac \quad (4-40)$$

2) 等效数字滤波器 $H(z)$ 的计算机编程实现

将式(4-40)交叉相乘

$$A(1 + cT - z^{-1})U(z) = (1 + bT - z^{-1})Y(z)$$

等式两边求 Z 反变换

$$A(1 + cT) \cdot u(k) - Au(k-1) = (1 + bT)y(k) - y(k-1)$$

于是得差分方程如下

$$\begin{aligned} y(k) &= \frac{A(1 + cT)}{1 + bT}u(k) - \frac{Au(k-1)}{1 + bT} + \frac{y(k-1)}{1 + bT} \\ y(k) &= p \left[\frac{A}{q}u(k) - Au(k-1) + y(k-1) \right] \end{aligned} \quad (4-41)$$

式中, $p = 1/(1+bT)$; $b = A/\tau = 2\pi f_b$; $q = 1/(1+cT)$; $c = 1/\tau = 2\pi f_c$; f_c 为扩展频带前传感器原有的转折频率; f_b 为扩展频带后传感器系统的转折频率; $f_b = Af_c$; k 为采样时序号; T 为采样间隔。

式(4-41)给出了 $H(s)$ 的等效数字滤波器 $H(z)$ 的编程算式, 当前时刻 kT 的输出由三部分组成:

$p \frac{A}{q}u(k)$ ——当前时刻 kT 的输入值 $u(k)$ 乘以系数 $p \frac{A}{q}$;

$pAu(k-1)$ ——当前时刻之前一时刻 $(k-1)T$ 的输入值 $u(k-1)$ 乘以系数 pA ;

$py(k-1)$ ——当前时刻之前一时刻 $(k-1)T$ 的输出值 $y(k-1)$ 乘以系数 p 。

实现编程算式(4-41)就实现了需要串联校正环节的等效数字滤波器, 但是必须已知待扩展频带环节原有的动态特性, 即必须已知表征一阶环节动态特性的特征参数 τ 。确定 τ 值的实验方法有两种: 一是频率特性法, 要求输入信号是频率可调、幅值恒定的正弦波信号; 二是阶跃响应法, 要求输入信号为阶跃信号。对于测量非电量的传感器, 多采用阶跃响应法, 因为获得非电量, 如温度、压力的阶跃信号比获得它们的正弦信号要容易得多。

4. 采用数字滤波法扩展频带应注意的问题

1) 待补偿系统参数值不准确对补偿效果的影响

以一阶系统为例, 其特征参数 τ 的测定值 $\tau(\text{测})$ 与其实际值 $\tau(\text{实})$ 之间肯定存在偏差, 从而获得补偿后系统的传递函数不可能是式(4-38), 而是如下的表达式

$$I(s) = \frac{1}{1 + \tau(\text{实}) \cdot s} \cdot \frac{1 + \tau(\text{测}) \cdot s}{1 + \frac{\tau(\text{测})}{A} \cdot s} = \frac{1}{1 + \tau s} \cdot \frac{1 + \tau'' s}{1 + \frac{\tau''}{A} s} \quad (4-42)$$

其中 $\tau = \tau(\text{实})$, 是传感器系统实际的时间常数; $\tau'' = \tau(\text{测})$ 是时间常数测定值。

可以看到, 补偿后的系统成了一个二阶系统。可能会出现频带扩展了但却有过冲或不出现过冲但频带未能扩展到预期目标的现象(见[示例 4-2])。

2) 频带扩展倍数 A 的受限因素

频带扩展倍数 A 要受系统硬件频带如采样频率的限制, 也就是软件补偿效果受限于硬件条件。一个离散时间系统必须遵守奈奎斯特采样定理, 也就是说, 其采样频率 f_s 必须大

于系统的截止频率理论上为2倍，实际上需要8~20倍。故，若系统补偿前的带宽为 $f_c=1\text{ kHz}$ ，数采卡的采样频率 $f_s=100\text{ kHz}$ ，那么频带扩展倍数 $A=\frac{1}{2\sim 20} \cdot \frac{f_s}{f_c}=50\sim 5$ 。

4.4.3 扩展频带的频域校正法。

频域校正法的校正思想是：在已知系统传递函数 $W(s)$ 的前提下，把由于受限于系统频带而畸变的 $y(t)$ 经过处理，找到被测输入信号 $x(t)$ 的频谱 $X(m)$ ，再通过傅里叶反变换进而获得被测信号的真值 $x(t)$ ，相当于将系统频带扩展。

图4-20给出了频域校正法的过程，与数字滤波法一样，它也必须已知系统的传递函数，如果不知道的话，则需要事先测定表征动态特性的特征参数，从而得出传递函数 $W(s)$ 或频率特性 $W(j\omega)$ ，然后再开始用软件实现频域校正，详细的校正步骤介绍如下。

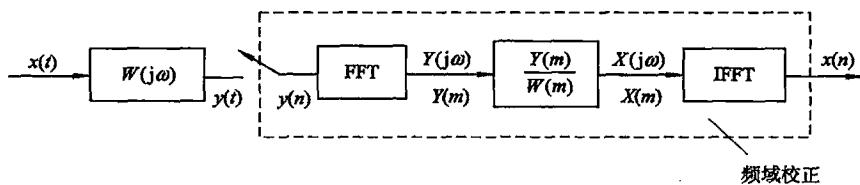


图4-20 频域校正过程示意图

1. 采样

对输入信号 $x(t)$ 的输出响应信号 $y(t)$ 进行采样，得时间序列 $y(n)$ ($n=0, 1, 2, \dots, N-1$)，信号记录长度 $t_p=NT_s$ ， T_s 为采样间隔， $\frac{1}{T_s}=f_s$ 为采样频率，必须满足采样定理

$$f_s > 2f_m$$

式中， f_m 为输入信号 $x(t)$ 的最高频率。

2. 频谱分析

对采样信号 $y(n)$ 进行频谱分析，即进行快速傅里叶变换(FFT)，得出它的频谱 $Y(m)$ ($m=0, 1, 2, \dots, N/2$)，其基波频率为 $1/t_p=\Omega/2\pi$ ， $\omega=m\Omega$ 。

3. 做复数除法运算

已知系统频率特性 $W(j\omega)$

$$W(j\omega) = \frac{Y(j\omega)}{X(j\omega)}$$

式中： $Y(j\omega)$ 为系统输出信号的频谱； $X(j\omega)$ 为系统输入信号的频谱。

因为计算机是离散时间系统，只能得到离散的谱线，即

$$\omega = m\Omega$$

式中： $\Omega=2\pi/t_p$ 为基波频率； m 为谱线序号， $m=0, 1, 2, \dots, N/2$ 。故系统频率特性的离散时间表达式为

$$W(m) = \frac{Y(m)}{X(m)}$$

将 $W(m)$ 与 $Y(m)$ 做复数除法，可得系统被测输入信号的频谱：

$$X(m) = \frac{Y(m)}{W(m)} \quad (4-43)$$

由式(4-43)运算所得的 $X(m)$, 正是被测输入信号的频谱。

对频谱 $X(m)$ 进行傅里叶反变换(IFFT)即可得原函数 $x(t)$ 的离散时间序列 $x(n)$ ($n=0, 1, 2, \dots, N-1$)。这个原函数 $x(t)$ 正是我们要测量的系统的输入信号的真值。这就意味着: 若不进行频域校正, 传感器系统输出的响应信号 $y(t)$ 是畸变了的, 它用畸变了的 $y(t)$ 代表被测的输入信号 $x(t)$, 这当然就存在误差。频域校正把畸变的 $y(t)$ 经过处理找到被测输入信号 $x(t)$ 的频谱 $X(m)$, 进而获得了被测信号 $x(t)$ 的真值, 于是便消除了误差。

4.4.4 [示例 4-2]采用数字滤波法将测温传感器(一阶系统)频带扩展

$A \geq 10$ 倍

第一步: 时间常数测定。将温度传感器放在冰水混合物的冰瓶中, 待温度平衡后迅速将温度传感器提出冰瓶, 由此对温度传感器输入一个从 0℃ 至室温(39℃)的温度阶跃信号, 数据采集系统同步采集系统的响应输出信号 $u(t)$, 如图 4-21 所示。

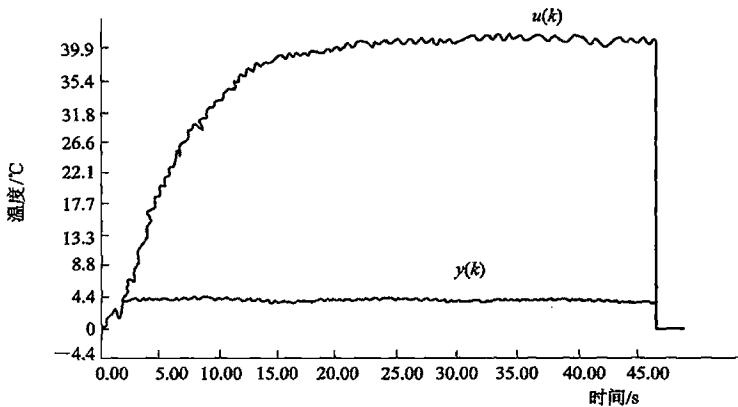


图 4-21 温度传感器系统对阶跃输入温度信号的响应

经计算得 $\tau = 2.9$ s, $f_c = \frac{1}{2\pi\tau} = 0.0557$ Hz。

第二步: 执行校正环节 $H(s)$ 等效数字滤波 $H(z)$ 的编程算式(4-41)。根据已测定出的 τ 值与所要求的频带扩展倍数 $A \geq 10$, 计算出传感器频带扩展前后的转折频率 f_c 、 f_b 以及 p 、 q 值, 编程算式(4-41)即可执行。

第三步: 最佳补偿效果的判断与调节实现。由于时间常数 τ 的测定必然存在误差, 不可能绝对准确, 故设置的 f_c (转折频率)会出现偏差, 从而会发生 f_c 偏大时补偿不足, f_c 偏小时补偿过分的现象。这个现象可以通过以下方法进行补偿。观察阶跃响应 $u(t)$ 的采样信号 $u(k)$, 以及通过数字滤波器后的阶跃响应信号 $y(k)$, 如图 4-21 中的曲线 $u(k)$ 为数字滤波前的温度阶跃响应信号, 设置不同 f_c 值滤波器输出的温度阶跃响应 y_1 、 y_2 、 y_3 , 如图 4-22 所示。曲线 y_3 对应于 $f_c = 0.0557$ Hz; 曲线 y_1 对应于 $f'_c = 0.0265$ Hz $< f_c = 0.0557$ Hz, 出现补偿过冲; 曲线 y_2 对应于 $f''_c = 0.557$ Hz $> f_c = 0.0557$ Hz, 出现补偿不

足，上升前沿还不够快。比较曲线 y_1 、 y_2 、 y_3 可见，曲线 y_3 为最佳补偿效果，上升沿有明显改善但又无太大过冲。由此可见，通过对数字滤波器的软件调试，很容易调整 f_c 值，达到最佳补偿效果。

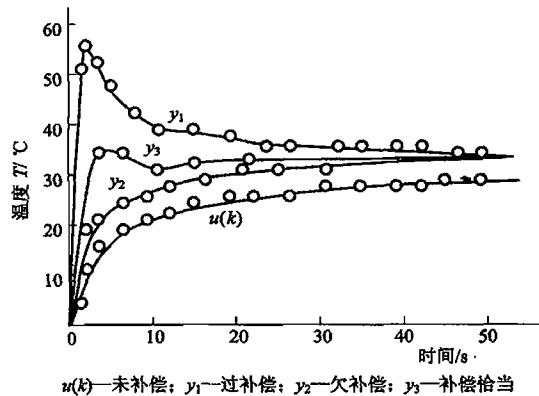


图 4-22 不同补偿效果的图示

4.5 提高信噪比与分辨力及智能化信号提取与消噪功能

传感器获得的信号中常常夹杂着各种干扰信号。兼有信息获取与处理功能的智能传感器系统，其信息处理的基本功能是能自动准确地将有用信号从噪声中提取出来。因为只有有用信息才表征被测对象的特征。信号提取与消噪技术有多种，如滤波器技术、频域消噪技术、相关分析技术等。它们各有特点，经典滤波器一般只适用于信号频谱和噪声频谱不重合的情况；信号频谱与噪声频谱重叠时相关分析具有优势；小波消噪技术发展极快，我们将将在第 10 章介绍，本节简介数字滤波、相关及频域谱分析消噪技术。

4.5.1 数字滤波技术

本书涉及到智能传感器系统中的数字滤波器，其功用主要有三个：一是抑制噪声干扰信号，起消噪作用；二是提取某频段频率分量的信号；三是作频率补偿器使测量系统频带扩展（详见 4.4 节）。本小节从信号提取与消噪功能的角度讨论数字滤波器。

1. 滤波器的特征频带与分类

滤波器有多种分类方式，主要可划分为经典与现代两大类，其中经典滤波器只适用于噪声频谱和信号频谱不重叠的场合，对信号与噪声有分离消噪与信号提取的作用。经典滤波器进而又分为模拟与数字滤波器两种。考虑逼近理想滤波特性采用的逼近准则不同，常见的滤波器有巴特沃斯与切比雪夫滤波器。按当前输出与前一时刻输入与输出的关系来分，有无限冲激响应(IIR)与有限冲激响应滤波器(FIR)。现代滤波器把信号与噪声都视为随机信号，不在意二者频谱是否重叠，其研究重点是从含有噪声的数据记录中估计出信号的某些特征或信号本身，主要有自适应滤波、小波分析等方法。本节讨论经典滤波器。

1) 滤波器的特征频带

滤波器的类型可以不同，但其频率特性 $H(\omega)$ 都有三个特征段：通带、阻带及过渡带。

(1) 通带。在这段频率范围，滤波器输入信号的幅值 $|X(\omega)|$ 与其输出信号的幅值 $|Y(\omega)|$ 之比

$$A = \frac{|Y(\omega)|}{|X(\omega)|} \approx \text{常量}$$

近似为一常量，信号可不受或受很小的影响通过滤波器。理想带通的 A 为常量，其幅频特性是一条平直线。

(2) 阻带。在这段频率范围内 $A \approx 0$ ，信号受很大衰减而被阻止。理想阻带的 A 为 0，信号完全被阻止。

(3) 过渡带。这是位于通带与阻带之间的一段频带，幅值比值 A 不为常量，随信号频率的变化而改变。

2) 滤波器的分类

根据上述三种特征频带在滤波器频率特性频段中所处位置区间的不同，滤波器可分为四类，如图 4-23 所示。

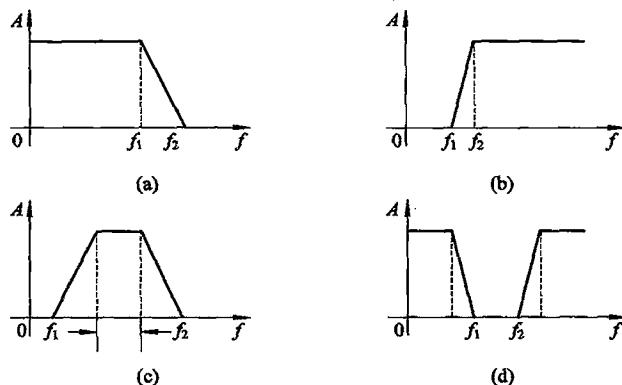


图 4-23 四类滤波器的频谱

(a) 低通滤波器；(b) 高通滤波器；(c) 带通滤波器；(d) 带阻滤波器

(1) 低通滤波器，通带在低频段，为 $0 \sim f_1$ ；过渡带为 $f_1 \sim f_2$ ， $f_2 > f_1$ 。

(2) 高通滤波器，通带在高频段，为 $f_2 \sim \infty$ ；过渡带为 $f_1 \sim f_2$ ， $f_2 > f_1$ 。

(3) 带通滤波器，通带在中频段，为 $f_1 \sim f_2$ ；阻带在通带两侧，有两个过渡带。

(4) 带阻滤波器，阻带在中频段，为 $f_1 \sim f_2$ ；阻带两侧均为通带，有两个过渡带。

考虑到信号提取与抑制噪声，希望欲提取信号的频段落在通带，欲消除的噪声的频带落在阻带，且不希望存在过渡带。没有过渡带的滤波器是理想滤波器。

2. 巴特沃斯滤波器与切比雪夫滤波器

没有过渡带的理想滤波器在技术上是难以实现的。实际设计滤波器的过程就是寻找逼近理想滤波器特性的过程，即寻求一个可实现的衰减或上升特性来逼近理想特性，使衰减或上升变化率处在所规定的限度内。

根据不同的逼近准则，就有不同的衰减或上升特性，从而形成不同的滤波器。不同的逼近准则形成的依据为滤波器的传递函数。例如一般的低通频率响应特性，其传递函数为

$$H(s) = \frac{Y(s)}{X(s)} = \frac{Kb_0}{s^n + b_{n-1}s^{n-1} + \dots + b_1s + b_0} \quad (4-44)$$

式中， $X(s)$ 、 $Y(s)$ 分别为输入 $x(t)$ 、输出 $y(t)$ 信号的拉普拉斯变换； $b_0, b_1, b_2, \dots, b_{n-1}$ 为供适当选择的常数； K 为直流放大倍数； n 为滤波器阶次。

在很多逼近准则中，常用的是巴特沃斯(Butterworth)和切比雪夫(Chebyshev)准则。它们的传递函数都是式(4-44)的变形，差别仅在常数 $b_0, b_1, b_2, \dots, b_{n-1}$ 的选择。巴特沃斯滤波器有所谓最平的通带，而切比雪夫滤波器的通带有“波纹”。

1) 巴特沃斯低通滤波器

(1) 幅频特性的一般表达式。巴特沃斯低通滤波器是一种以所谓最平通带特性逼近理想低通滤波器的滤波器。其幅频特性为输出 $|Y(\omega)|$ 与输入 $|X(\omega)|$ 幅值之比，表达式为

$$|H(\omega)| = \frac{|Y(\omega)|}{|X(\omega)|} = \frac{1}{\sqrt{1 + \left(\frac{\omega}{\omega_c}\right)^{2n}}} \quad (4-45)$$

式中： $n=1, 2, 3, \dots$ 为滤波器的阶次； ω_c 为截止角频率，是幅值 $|H(\omega)|$ 下降至最大值的 $1/\sqrt{2}=0.707$ 或 3 dB 时对应的角频率值。

将 ω/ω_c 不同值代入式(4-45)，就可计算得巴特沃斯滤波器的输出 $|Y(\omega)|$ 与输入 $|X(\omega)|$ 幅值比 $A=|H(\omega)|$ 随信号频率 ω/ω_c 变化的规律。 $|H(\omega)|-\omega/\omega_c$ 关系就是幅频特性，如图 4-24 所示。 n 等于 1 和 2 时的幅频特性的相应数值列入表 4-5 中。

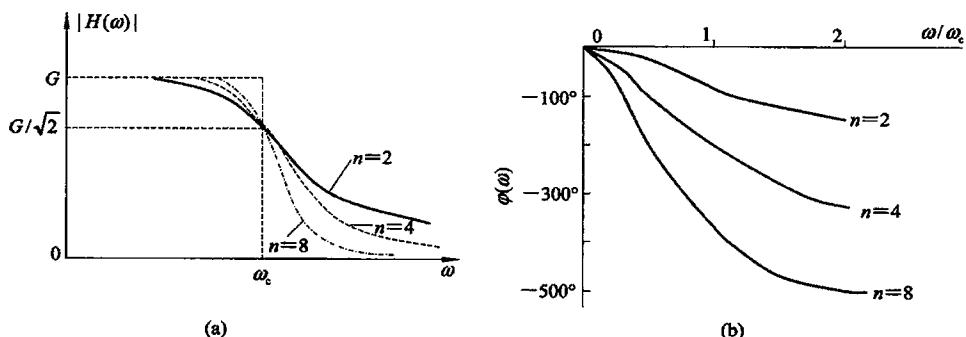


图 4-24 n 值不同时巴特沃斯滤波器幅频特性与相频特性

(a) 幅频特性；(b) 相频特性

表 4-5 $n=1, 2$ 时，巴特沃斯低通滤波器 $|H(\omega)|=A$ 与 ω/ω_c

n	ω/ω_c	0	0.1	0.5	1.0	5.0	10
1	$A=\frac{1}{\sqrt{1+\left(\frac{\omega}{\omega_c}\right)^2}}$	1	0.9950371	0.8944272	0.7071068	0.1961161	0.0995037 ≈ 0.1
2	$A=\frac{1}{\sqrt{1+\left[\left(\frac{\omega}{\omega_c}\right)^2\right]^2}}$	1	0.9999500	0.9701425	0.7071068	0.0399680	0.0099995 ≈ 0.01

理想低通滤波器在通带 $|H(\omega)| = 1$, 幅频特性应是不随信号频率变化的常量, 但实际上 $|H(\omega)|$ 随 ω 的增大而逐渐下降, 称 $|H(\omega)|$ 与 1 之差为衰减率或称衰减度 δ , 即

$$\delta = 1 - A = 1 - |H(\omega)| \quad (4-46)$$

(2) 一阶巴特沃斯低通滤波器。

① 传递函数:

$$H(s) = \frac{Kb_0}{b_1 s + b_0} = \frac{K}{\tau s + 1} \quad (4-47a)$$

式中, $\tau = b_1 / b_0$; $s = \sigma + j\omega$ 。若令 $\sigma = 0$, 得频率特性如下:

$$H(j\omega) = H(\omega) = \frac{K}{\tau j\omega + 1} \quad (4-47b)$$

② 幅频特性:

$$A(\omega) = |H(\omega)| = \frac{K}{\sqrt{1 + \left(\frac{\omega}{\omega_c}\right)^2}} \quad (4-48)$$

式中, $\omega_c = 1/\tau$ 为转折角频率或称截止角频率。与式(4-45)相比, 式(4-48)中 $n=1$ 。

③ 相频特性:

$$\theta(j\omega) = \theta(\omega) = -\arctan \omega \tau = -\arctan \frac{\omega}{\omega_c} \quad (4-49)$$

通常将具有式(4-47)特性的系统称为一阶系统, 也即一阶系统具有一阶巴特沃斯低通滤波器的特性。作为低通滤波器 $n=1$ 时, 其低通特性是很差的, n 值越大就越接近理想特性。

2) 切比雪夫低通滤波器

切比雪夫低通滤波器的幅频特性定义如下:

$$|H(\omega)| = \frac{K}{\sqrt{1 + \epsilon^2 C_n^2 \left(\frac{\omega}{\omega_c}\right)}} \quad (4-50)$$

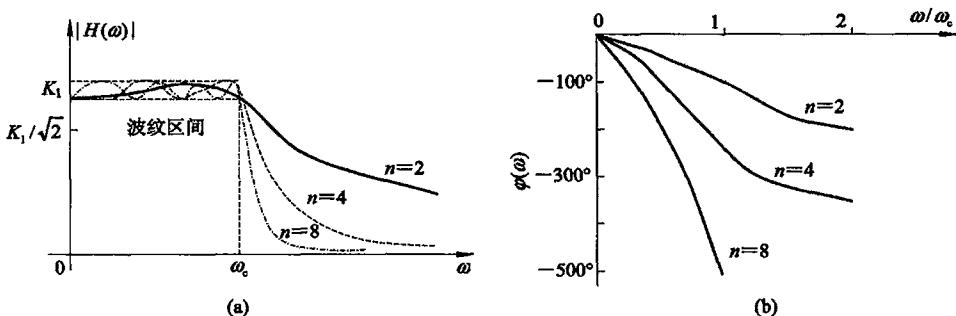
式中, K 为直流放大倍数; ϵ 为与波纹大小有关的常数; C_n 为切比雪夫多项式, $n=1, 2, \dots$ 为滤波器的阶数; ω_c 为波纹终止点对应的角频率, 当只有 3 dB 切比雪夫滤波器时, 等纹波区间的终止角频率 ω_c 才指通常的 3 dB 处的截止角频率。

当 $\epsilon=0.0885$, 通常波纹幅值 A 为 1 dB=20 lgA 时, 切比雪夫低通滤波器的传递函数为

$n=1$ 时

$$H(s) = \frac{K}{\frac{s}{\omega_c} + 1.96523} \quad (4-51)$$

如图 4-25(a)、(b)所示分别给出了相同 ϵ 值、不同 n 值时的幅频特性与相频特性。由图 4-25 可见, 在通带内, 具有相等幅度的波纹, 随着 n 值的增加, 波纹数目相应增加, 同时阻带内衰减也增加, 与理想特性近似越好。

图 4-25 n 值不同时切比雪夫低通滤波器的幅频特性与相频特性

(a) 幅频特性; (b) 相频特性

3. 模拟滤波器的设计

以巴特沃斯模拟滤波器设计为例介绍一种滤波器设计方法。

1) 巴特沃斯低通滤波器的设计

设计目标是确定阶次 n 与截止角频率 ω_c 以满足设计要求。设计要求通常是对幅频特性的过渡段特性提出要求，如：

信号频率 $f_1 (\omega_1 = 2\pi f_1)$ 时，对应的衰减率为 δ_1 ；

信号频率 $f_2 (\omega_2 = 2\pi f_2)$ 时，对应的衰减率为 δ_2 。

由式(4-46)已知衰减率 δ 与幅值比 $A = |H(\omega)|$ 的关系，再写如下

$$\delta = 1 - A \quad (4-52)$$

(1) 低通滤波器阶次 n 的确定。将式(4-45)与设计要求的 f_1 、 f_2 及其 δ_1 、 δ_2 代入式(4-52)可得

$$\delta_1 = 1 - \frac{1}{\sqrt{1 + \left(\frac{\omega_1}{\omega_c}\right)^{2n}}} \quad (4-53a)$$

$$\delta_2 = 1 - \frac{1}{\sqrt{1 + \left(\frac{\omega_2}{\omega_c}\right)^{2n}}} \quad (4-53b)$$

将式(4-53)移项、平方并整理后，可得

$$\left(\frac{\omega_1}{\omega_c}\right)^{2n} = \frac{1}{(1 - \delta_1)^2} - 1 = \frac{1}{A_1^2} - 1 \quad (4-54a)$$

$$\left(\frac{\omega_2}{\omega_c}\right)^{2n} = \frac{1}{(1 - \delta_2)^2} - 1 = \frac{1}{A_2^2} - 1 \quad (4-54b)$$

求式(4-54)中两式之比，有

$$\left(\frac{\omega_2}{\omega_1}\right)^{2n} = \left(\frac{f_2}{f_1}\right)^{2n} = \frac{\left(\frac{1}{A_2^2} - 1\right)}{\left(\frac{1}{A_1^2} - 1\right)}$$

即

$$\left(\frac{f_2}{f_1}\right)^n = \sqrt{\frac{\left(\frac{1}{A_2^2} - 1\right)}{\left(\frac{1}{A_1^2} - 1\right)}} \quad (4-55)$$

为使等式(4-55)成立, n 值可能不为整数, 这时取 n 为大于零的最小整数值。

(2) 低通滤波器截止角频率 ω_c 的确定。根据已确定的阶次 n , 由式(4-54)两式中的任一个式子都可求出 ω_c 。

(3) 巴特沃斯低通滤波器的传递函数 $H(s)$ 与幅频特性 $|H(\omega)|$:

$n=1$ 时,

$$H(s) = \frac{1}{\frac{1}{\omega_c}s + 1}, \quad |H(\omega)| = \frac{1}{\sqrt{1 + \left(\frac{\omega}{\omega_c}\right)^2}} \quad (4-56a)$$

$n=2$ 时,

$$H(s) = \frac{1}{\left(\frac{s}{\omega_c}\right)^2 + \sqrt{2}\left(\frac{s}{\omega_c}\right) + 1}, \quad |H(\omega)| = \frac{1}{\sqrt{1 + \left(\frac{\omega}{\omega_c}\right)^{2 \times 2}}} \quad (4-56b)$$

$n=3$ 时,

$$H(s) = \frac{1}{\left(\frac{s}{\omega_c}\right)^3 + 2\left(\frac{s}{\omega_c}\right)^2 + 2\left(\frac{s}{\omega_c}\right) + 1}, \quad |H(\omega)| = \frac{1}{\sqrt{1 + \left(\frac{\omega}{\omega_c}\right)^{2 \times 3}}} \quad (4-56c)$$

式中, ω_c 为截止角频率; 直流放大倍数 $K=1$ 。

2) 巴特沃斯高通滤波器的设计

巴特沃斯低通滤波器传递函数 $H_d(s)$ 的 s 用 $1/s$ 代替可得阶次 n 相同的巴特沃斯高通滤波器 $H_g(s)$ 。下面以 $n=1$ 为例介绍其设计方法。

(1) 一阶巴特沃斯高通滤波器的传递函数 $H_g(s)$ 。由式(4-56)可知, 当 $n=1$ 时的一阶低通滤波器传递函数 $H_d(s)$:

$$H_d(s) = \frac{1}{\frac{1}{\omega_{cd}}s + 1} = \frac{1}{a_d s + 1}, \quad a_d = \frac{1}{\omega_{cd}}$$

$H_d(s)$ 中的 s 由 $1/s$ 代替, 且再将 a_d 由 $1/a_g$ 代替, 即得一阶高通滤波器 $H_g(s)$:

$$H_g(s) = \frac{1}{a_d s + 1} \Big|_{s \rightarrow \frac{1}{s}} = \frac{s}{s + a_d} \Big|_{a_d \rightarrow \frac{1}{a_g}} = \frac{a_g s}{1 + a_g s} \quad (4-57)$$

式中 $a_g = \frac{1}{\omega_{cg}}$, ω_{cg} 为高通滤波器通带截止角频率。

(2) 一阶巴特沃斯高通滤波器的频率特性 $H_g(\omega)$ 及幅频特性 $|H_g(\omega)|$:

$$H_g(\omega) = \frac{j a_g \omega}{1 + j a_g \omega} = \frac{j \omega / \omega_{cg}}{1 + j \omega / \omega_{cg}}$$

$$|H_g(\omega)| = \frac{\omega / \omega_{cg}}{\sqrt{1 + \left(\frac{\omega}{\omega_{cg}}\right)^2}}$$

令 $\omega / \omega_{cg} = 0, 0.1, 0.5, 1.0, 5.0, 10.0$ 等不同值, 可计算得到相应的幅值比 $|H_g(\omega)|$ 与 ω / ω_{cg} 的关系, 见表 4-6。

表 4-6 巴特沃斯高通滤波器幅频特性 $|H_g(\omega)| - \omega/\omega_{cg}$

ω/ω_{cg}	0	0.1	0.5	1.0	5.0	10.0	50.0
$\frac{1}{\sqrt{1+(\frac{\omega}{\omega_{cg}})^2}}$	1	0.995 037 19	0.894 427 19	0.707 106 78	0.196 116 135	0.099 503 719	0.019 996 001
$ H_g(\omega) = \frac{\omega/\omega_{cg}}{\sqrt{1+(\frac{\omega}{\omega_{cg}})^2}}$	0	0.099 503 71	0.447 213 59	0.707 106 78	0.980 580 675	0.995 037 19	0.999 800 06

由表 4-6 数据规律可见, 当 $\omega/\omega_{cg}=0$ 时, $|H_g(\omega)|=0$; 并随 ω 频率增大 $|H_g(\omega)|$ 增大并趋近 1, $|H_g(\omega)| - \omega/\omega_{cg}$ 关系呈现高频特性, 故是高通滤波器, 且当 $\omega=\omega_{cg}$ 时 $|H_g(\omega)|=0.7071$, 衰减率 $\delta_g=1-|H_g(\omega)| \approx 0.293$, 可见 ω_{cg} 为高通滤波器的截止角频率。高通滤波器幅频特性如图 4-26 所示。

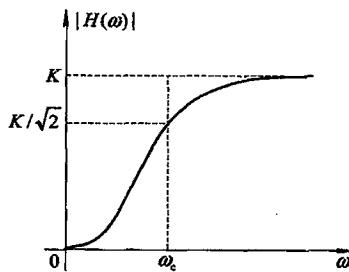


图 4-26 高通滤波器幅频特性

3) 巴特沃斯带通滤波器的设计

巴特沃斯带通滤波器的幅频特性如图 4-27 所示。图中 ω_0 为中频角频率; $B=\omega_2-\omega_1$ 为带宽, ω_1 与 ω_2 为通常定义的截止角频率。由其幅频特性可见, 带通滤波器可由一个截止频率为 $\omega_{cd}=\omega_2$ 的低通滤波器与一截止频率 $\omega_{cg}=\omega_1$ 的高通滤波器串联而成, 下面以一阶为例进行设计。

(1) 低通滤波器的传递函数 $H_d(s)$:

$$H_d(s) = \frac{1}{\frac{1}{\omega_{cd}}s + 1} = \frac{1}{a_d s + 1}, \quad a_d = \frac{1}{\omega_{cd}} = \frac{1}{\omega_2}$$

(2) 高通滤波器的传递函数 $H_g(s)$:

$$H_g(s) = \frac{a_g s}{1+a_g s}, \quad a_g = \frac{1}{\omega_{cg}} = \frac{1}{\omega_1}$$

(3) 带通滤波器的传递函数 $H_b(s)$:

$$H_b(s) = H_d(s) \cdot H_g(s) = \frac{1}{a_d s + 1} \cdot \frac{a_g s}{1+a_g s}$$

$$H_b(s) = \frac{a_g s}{a_d \cdot a_g s^2 + (a_d + a_g)s + 1} \quad (4-58a)$$

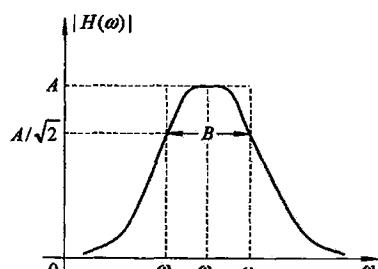


图 4-27 带通滤波器幅频特性

令 $s=j\omega$, 可得带通滤波器的频率特性 $H_b(\omega)$ 和幅频特性 $|H_b(\omega)|$:

$$H_b(\omega) = \frac{1}{\frac{j\omega}{\omega_{cd}} + 1} \frac{j\omega/\omega_{cg}}{1 + j\omega/\omega_{cg}} \quad (4-58b)$$

$$|H_b(\omega)| = |H_d(\omega)| \cdot |H_g(\omega)|$$

式中,

$$|H_d(\omega)| = \frac{1}{\sqrt{1 + \left(\frac{\omega}{\omega_{cd}}\right)^2}}, \quad |H_g(\omega)| = \frac{\omega/\omega_{cg}}{\sqrt{1 + \left(\frac{\omega}{\omega_{cg}}\right)^2}}$$

要说明的是, 欲使带通滤波器的低截止角频率 ω_1 与高截止角频率 ω_2 分别由一阶高通低通滤波器的截止角频率 ω_{cg} 、 ω_{cd} 来决定, 即

$$\omega_1 = \omega_{cg}, \quad \omega_2 = \omega_{cd} \text{ 或 } f_1 = f_{cg}, \quad f_2 = f_{cd}$$

这就要求通带宽 B 不能太窄, 否则 ω_1 与 ω_2 不是通常意义的截止角频率, 即 $|H(\omega_1)|$ 与 $|H(\omega_2)|$ 小于 0.707。如: $f_1 = 7$ Hz, $f_2 = 350$ Hz 时, 可以达到 $|H(\omega = 2\pi f_1)| = |H(\omega = 2\pi f_2)| = 0.7096 \approx 0.707$ (-3 dB), 但当 $f_1 = 7$ Hz, $f_2 = 70$ Hz 时, 却有 $|H(\omega = 2\pi f_1)| = |H(\omega = 2\pi f_2)| = 0.7036$, 均小于 0.707, 故要设计带宽窄的带通滤波器, 需要 n 值大的高阶低通与高通滤波器进行串联组合。

4) 巴特沃斯带阻滤波器的设计

巴特沃斯带阻滤波器的幅频特性如图 4-28 所示, 图中 ω_0 为中频角频率, $B = \omega_2 - \omega_1$ 为带宽, ω_1 与 ω_2 为通常定义的截止角频率。由其幅频特性可见, 带阻滤波器可由一个截止角频率为 $\omega_{cd} = \omega_1$ 的低通滤波器与一截止角频率 $\omega_{cg} = \omega_2$ 的高通滤波器串联而成, 以一阶为例, 带阻滤波器的传递函数为 $H(s)$

$$H(s) = H_d(s) \cdot H_g(s) = \frac{1}{a_d s + 1} \cdot \frac{a_g s}{1 + a_g s} \quad (4-59)$$

式中: $a_d = 1/\omega_{cd}$, $a_g = 1/\omega_{cg}$; 在 B 足够宽时 $\omega_{cd} = \omega_1$, $\omega_{cg} = \omega_2$ 。

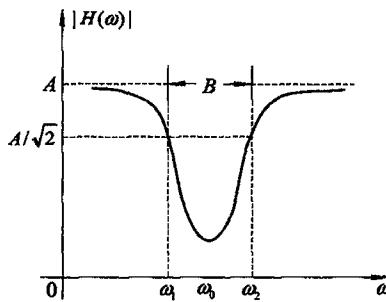
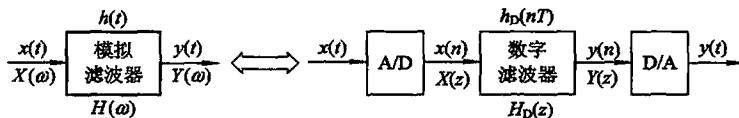


图 4-28 带阻滤波器幅频特性

4. 连续时间(t)滤波器 $H(s)$ 的离散时间(nT)等效滤波器 $H(z)$

连续时间滤波器又称模拟滤波器。经过采样/保持器后进入计算机的信号是在时间为离散时刻 ($t = nT$, $n = 0, 1, 2, \dots$) 上具有数值, 而数值又是整量化了的时间序列。在计算机中处理这种幅值量化了的离散时间序列信号的滤波器称为数字滤波器。

通常的情况是首先确定一个期望的模拟滤波器 $H(s)$: 被测信号的频率范围应在其通带内, 欲清除的干扰噪声的频率范围应落在阻带, 从而滤波器的类型及其相应的截止频率就被确定; 然后下一步的主要任务就是设计软件, 实现与 $H(s)$ 等效的数字滤波器 $H_D(z)$ 。等效示意框图如图 4-29 所示。等效方法有多种, 如脉冲响应不变法(Z 变换法)、后向差分法、双线性变换法、频率预曲折双线性变换法等。下面只介绍两种方法: 后向差分法与双线性变换法。

图 4-29 $H(s)$ 与 $H_D(z)$ 的等效

1) 后向差分法

后向差分法是一种数值积分法。下面以最简单的一阶系统为例, 介绍由微分方程建立差分方程, 再对差分方程进行 Z 变换, 进而获得连续时间滤波器 $H(s)$ 的等效数字滤波器 $H_D(z)$ 。

(1) 推导差分方程。

$$\text{微分方程} \quad \frac{dy(t)}{dt} = -ay(t) + ax(t), \quad a = \frac{1}{\tau} \quad (\tau \text{ 为时间常数})$$

$$\text{传递函数} \quad H(s) = \frac{Y(s)}{X(s)} = \frac{1}{1+as} = \frac{a}{s+a}$$

将上面微分方程的等号两边从 0 到 t 进行积分。

$$\int_0^t \frac{dy(t)}{dt} dt = -a \int_0^t y(t) dt + a \int_0^t x(t) dt$$

若求解每个采样周期 T 时 $y(t)$ 的值, 将 $t=kT$ 代入上式得

$$\begin{aligned} \int_0^{kT} \frac{dy(t)}{dt} dt &= -a \int_0^{kT} y(t) dt + a \int_0^{kT} x(t) dt \\ y(kT) - y(0) &= -a \int_0^{kT} y(t) dt + a \int_0^{kT} x(t) dt \end{aligned} \quad (4-60)$$

同理, 从 $0 \rightarrow t=(k-1)T$ 积分, 则有:

$$\begin{aligned} \int_0^{(k-1)T} \frac{dy(t)}{dt} dt &= -a \int_0^{(k-1)T} y(t) dt + a \int_0^{(k-1)T} x(t) dt \\ y((k-1)T) - y(0) &= -a \int_0^{(k-1)T} y(t) dt + a \int_0^{(k-1)T} x(t) dt \end{aligned} \quad (4-61)$$

将式(4-60)和式(4-61)相减后可得

$$y(kT) - y((k-1)T) = -a \int_{(k-1)T}^{kT} y(t) dt + a \int_{(k-1)T}^{kT} x(t) dt \quad (4-62)$$

式(4-62)右侧两项在数值上可由各种方法积分, 运用后向差分法积分就是矩形面积近似曲线 $y(t)$ 下的面积, 如图 4-30 所示, 则式(4-62)变为

$$y(kT) - y((k-1)T) = -aTy(kT) + aTx(kT)$$

整理得

$$y(k) = qx(k) + py(k-1) \quad (4-63)$$

式中: $q = \frac{aT}{1+aT} = \frac{T}{T+\tau} < 1$; $p = \frac{1}{1+aT} = \frac{\tau}{T+\tau} < 1$; $q+p=1$ 。

式(4-63)就是后向差分法实现模拟低通滤波器 $H(s)=a/(s+a)$, $a=1/\tau=\omega_c$ 的等效数字滤波器的计算机编程算式。当前时刻 kT 的输出 $y(k)$ 由两部分组成:

- $qx(k)$ ——当前时刻的输入值 $x(k)$ 乘以 q ;
- $py(k-1)$ ——当前时刻的前一时刻 $(k-1)T$ 的输出值 $y(k-1)$ 乘以 p 。

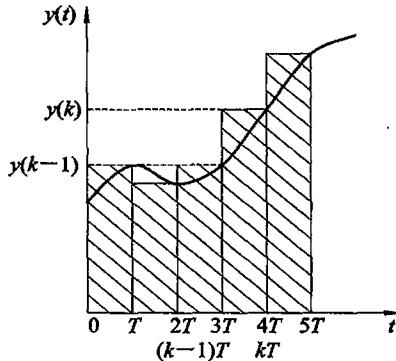


图 4-30 后向差分法面积近似图

由式(4-63)实现的 $H(s)$, 可以通过 p 、 q 的值很方便地调节截止频率 $\omega_c=a$, 例如, 当 $q=0.2$, $p=0.8$ 时, 对应的时间常数 $\tau=1/a=4T$; 而当 $q=0.9$, $p=0.1$ 时, $\tau=T/9$ 。由硬件制作的滤波器要想调节它的截止频率 $\omega_c=a=1/\tau$ 是很不方便的。

(2) 由 $H(s)$ 求等效 $H_D(z)$ 的后向差分法。具体方法是令 $H(s)$ 中的 s 为

$$s = \frac{1 - z^{-1}}{T}$$

即

$$H_D(z) = H(s) \Big|_{s=\frac{1-z^{-1}}{T}} = \frac{a}{s+a} \Big|_{s=\frac{1-z^{-1}}{T}} = \frac{a}{\frac{1-z^{-1}}{T} + a}$$

于是有

$$H_D(z) = \frac{Y(z)}{X(z)} = \frac{aT}{1+aT - z^{-1}}$$

$$aTX(z) = (1 + aT - z^{-1})Y(z)$$

求 Z 反变换得差分方程

$$aTx(k) = (1 + aT)y(k) - y(k-1)$$

故

$$y(k) = \frac{1}{1+aT}y(k-1) + \frac{aT}{1+aT}x(k) = qx(k) + py(k-1)$$

上述差分方程与式(4-63)相同。因此, 由后向差分法求 $H(s)$ 的等效数字滤波器 $H_D(z)$, 只需令 $H(s)$ 中的 $s=(1-z^{-1})/T$ 即可。

2) 双线性变换法(梯形积分法或 Tustin 变换法)

(1) 推导差分方程。如图 4-31 所示的曲线 $y(t)$ 下的面积由梯形面积近似, 即

$$\int_{(k-1)T}^{kT} y(t) dt = \frac{1}{2} [y(kT) + y((k-1)T)]T$$

$$\int_{(k-1)T}^{kT} x(t) dt = \frac{1}{2} [x(kT) + x((k-1)T)]T$$

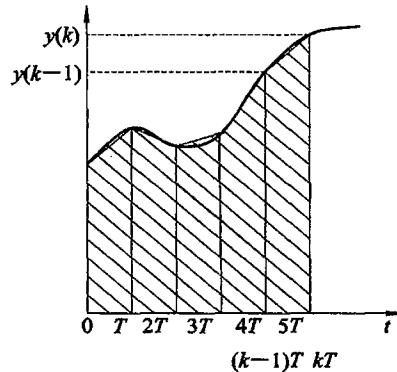


图 4-31 双线性变换法面积近似图(梯形近似)

于是, 由式(4-62)得差分方程如下

$$y(kT) = y((k-1)T) - \frac{aT}{2} [y(kT) + y((k-1)T)] + \frac{aT}{2} [x(kT) + x((k-1)T)] \quad (4-64)$$

经整理后可得

$$y(k) = py(k-1) + qx(k) - qx(k-1) \quad (4-65)$$

式中

$$p = \frac{1-aT/2}{1+aT/2}, \quad q = \frac{aT/2}{1+aT/2}$$

式(4-64)就是双线性变换法实现模拟低通滤波器 $H(s) = a/(a+s)$, $a = 1/\tau = \omega_c$ 的等效数字滤波器的计算机编程算式。当前时刻 kT 的输出 $y(k)$ 由三部分组成:

- $py(k-1)$ ——当前时刻的前一时刻 $(k-1)T$ 的输出值 $y(k-1)$ 乘以系数 p ;
- $qx(k)$ ——当前时刻的输入值 $x(k)$ 乘以系数 q ;
- $qx(k-1)$ ——当前时刻的前一时刻 $(k-1)T$ 的输入值 $x(k-1)$ 乘以系数 q 。

(2) 由 $H(s)$ 求等效 $H_D(z)$ 的双线性变换法。具体做法是令 $H(s)$ 中的 s 为

$$s = \frac{2}{T} \frac{1-z^{-1}}{1+z^{-1}} = \frac{2}{T} \frac{z-1}{z+1} \quad (4-66)$$

即

$$H_D(z) = H(s) \Big|_{s=\frac{2}{T}\frac{1-z^{-1}}{1+z^{-1}}} = \frac{a}{s+a} \Big|_{s=\frac{2}{T}\frac{1-z^{-1}}{1+z^{-1}}} = \frac{a}{\frac{2}{T}\frac{1-z^{-1}}{1+z^{-1}} + a}$$

于是有

$$H_D(z) = \frac{Y(z)}{X(z)} = \frac{a}{\frac{2}{T}\frac{1-z^{-1}}{1+z^{-1}} + a} = \frac{\frac{Ta}{2}(1+z^{-1})}{2(1-z^{-1}) + aT(1+z^{-1})}$$

$$(1 - z^{-1})Y(z) + \frac{aT}{2}(1 + z^{-1})Y(z) = \frac{aT}{2}(1 + z^{-1})X(z)$$

再对上式求 Z 反变换

$$y(k) - y(k-1) + \frac{aT}{2}y(k) + \frac{aT}{2}y(k-1) = \frac{aT}{2}x(k) + \frac{aT}{2}x(k-1)$$

整理后得差分方程

$$y(k) = \frac{1 - aT/2}{1 + aT/2}y(k-1) + \frac{aT/2}{1 + aT/2}x(k) + \frac{aT/2}{1 + aT/2}x(k-1)$$

上述差分方程与式(4-65)相同，因此由双线性变换法求 $H(s)$ 的等效数字滤波器 $H_D(z)$ ，只需令 $H(s)$ 中的 $s = 2(1 - z^{-1}) / (1 + z^{-1})T$ 即可。

由式(4-63)、式(4-64)都等效于同一模拟滤波器 $H(s)$ ，说明不同的等效方法获得的等效数字滤波器 $H_D(z)$ 有不完全相同的表达式。

5. 有限冲激响应与无限冲激响应数字滤波器

通常，数字滤波器的 Z 域脉冲传递函数 $H(z)$ 有两种类型，相应的输入、输出时间序列 $x(n)$ 与 $y(n)$ 关系的差分方程也有两种类型，这两种滤波器分别称为有限冲激响应滤波器(FIR)和无限冲激滤波器(IIR)。

1) 有限冲激响应滤波器(FIR)的数学模型

有限冲激响应滤波器的脉冲传递函数表达式为

$$H(z) = \frac{Y(z)}{X(z)} = a_0 + a_1z^{-1} + a_2z^{-2} + \cdots + a_Nz^{-N} = \sum_{n=0}^N a_n z^{-n} \quad (4-67)$$

于是有

$$Y(z) = (a_0 + a_1z^{-1} + a_2z^{-2} + \cdots + a_Nz^{-N})X(z)$$

求 Z 反变换得差分方程

$$y(n) = a_0x(n) + a_1x(n-1) + \cdots + a_nx(n-N) \quad [1] \text{ 即 } y(n) = \sum_{r=0}^N a_r x(n-r) \quad (4-68)$$

式(4-68)给出的差分方程表明：当前时刻的输出 $y(n)$ 由一系列的（包括当前时刻和历史时刻）输入值 $x(n-r)$ 乘以相应系数 a_r 决定。

2) 无限冲激响应滤波器(IIR)的数学模型

无限冲激响应滤波器的脉冲传递函数表达式为

$$H(z) = \frac{Y(z)}{X(z)} = \frac{\sum_{r=0}^N a_r z^{-r}}{1 + \sum_{k=1}^M b_k z^{-k}} \quad (4-69)$$

展开得

$$\frac{a_0 + a_1z^{-1} + a_2z^{-2} + \cdots + a_Nz^{-N}}{1 + b_1z^{-1} + b_2z^{-2} + \cdots + b_Mz^{-M}} = \frac{Y(z)}{X(z)}$$

于是有

$$(1 + b_1z^{-1} + b_2z^{-2} + \cdots + b_Mz^{-M})Y(z) = (a_0 + a_1z^{-1} + a_2z^{-2} + \cdots + a_Nz^{-N})X(z)$$

求反变换得差分方程

$$\begin{aligned} & y(n) + b_1 y(n-1) + b_2 y(n-2) + \cdots + b_M y(n-M) \\ & = a_0 x(n) + a_1 x(n-1) + \cdots + a_N x(n-N) \end{aligned}$$

故

$$y(n) = \sum_{r=0}^N a_r x(n-r) - \sum_{k=1}^M b_k y(n-k) \quad (4-70)$$

式(4-70)给出的差分方程表明当前时刻的输出 $y(n)$ 由两部分决定：

- 一系列的(含当前时刻和历史时刻)输入值 $x(n-r)$ 乘以相应的系数 a_r , $r = 0, 1, 2, \dots, N$;

- 一系列的历史时刻输出值 $y(n-k)$, 乘以相应的系数 b_k , $k = 1, 2, \dots, M$ 。

式(4-69)、式(4-70)所表征的均是无限冲激响应滤波器(IIR)的数学模型。

总结本节介绍的数字滤波器 $H(z)$ 的设计方法是：首先，在 S 域里先设计一个与模拟滤波器 $H(s)$ 等效的数字滤波的脉冲传递函数 $H(z)$, 以及相应的输出 $y(n)$ 与输入 $x(n)$ 序列的差分方程式，即得无限冲激响应数字滤波器(IIR)的计算机编程算式。

要注意，由 $H(s)$ 求 $H(z)$ 的等效方法有多种，任选一种都可。但因逼近 $|H(\omega)|$ 的方式与着重点有所不同，故采用不同等效方法求得的滤波器系数也不尽相同。也就是说，写成式(4-70)一般表达式后，对应的系数 a_r ($r = 0, 1, 2, \dots, N$) 与 b_k ($k = 1, 2, \dots, M$) 的数值可能不同。此外， a_r 与 b_k 的数值与采样间隔 T 有关。因此，在不同的(含 A/D 硬件与软件平台)计算机系统上面，必须保证相同的采样间隔 T ，才能实现相同滤波器系数 a_r 与 b_k 的数字滤波器具有相同的滤波器特性。

在 LabVIEW 软件平台上可以根据幅频特性高、低截止频率(带通、带阻)或高截止频率(低通、高通)以及阶次 n 来设计相应的滤波器，并直接获得滤波器系数 a_r 与 b_k ，详见[示例 4-6]。同样道理，在不同的计算机硬、软件平台上利用所获得系数 a_r 、 b_k 来实现数字滤波器，只有采用相同的采样隔离 T ，才会有一致的滤波器特性。

在微处理器应用中还有不少行之有效的平均滤波方法，如同步加算平均法、移动平衡法、中值滤波法等，具体内容可参考有关书籍，这里不再赘述。

4.5.2 相关

当信号淹没在噪声中并且信号和噪声的频带重叠时，相关技术是提高信噪比，把信号从噪声中提取出来的有力工具。

所谓相关，是指有明显的规律性，即在时间轴上任意间隔为 τ 的两个时刻的取值

$$x(t_1) \text{ 与 } x(t_1 + \tau)$$

$$x(t_2) \text{ 与 } x(t_2 + \tau)$$

$$x(t_3) \text{ 与 } x(t_3 + \tau)$$

⋮

$$x(t_m) \text{ 与 } x(t_m + \tau)$$

均有相同符号，或两者均大于平均值 m 或两者均小于平均值 m ，两两相乘总为正，总和相加后求平均 $R_x(\tau)$ 数值就大，故相关性好。反之，若间隔 τ 的两个时刻的取值 $x(t)$ 、 $x(t+\tau)$ 没有明显的规律性，即符号有时相同有时相反，则两两相乘有时为正有时为负，总和相加相互抵消后再求平均 $R_x(\tau)$ 的数值就小，故相关性不好。当 $\tau \rightarrow 0$ 时才呈现相关性的信号

$x(t)$, 就是一个完全无规则的随机信号。

1. 相关函数的定义式

1) 自相关函数 $R_x(\tau)$

信号 $x(t)$ 的自相关函数定义式

$$R_x(\tau) = E[x(t)x(t+\tau)] = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T x(t)x(t+\tau) dt \quad (4-71)$$

它描述一个随机过程在相隔 τ 的两个不同时刻取值的相关程度。

2) 互相关函数

两个信号 $x(t)$ 、 $y(t)$ 的互相关函数 $R_{xy}(\tau)$ 的定义式如下：

$$R_{xy}(\tau) = E[x(t)y(t+\tau)] = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T x(t)y(t+\tau) dt \quad (4-72)$$

它描述了两个不同的随机过程在相隔 τ 的两个不同时刻取值的相关程度。

2. 自相关函数的性质与特点

1) 自相关函数的性质

(1) 自相关函数是偶函数, 即

$$R_x(\tau) = R_x(-\tau)$$

证明：因为

$$R_x(-\tau) = E[x(t)x(t-\tau)]$$

令 $t-\tau=t'$, 则 $t=t'+\tau$ 代入上式

$$R_x(-\tau) = E[x(t'+\tau)x(t')] = R_x(\tau)$$

证毕。可知 $R_x(\tau)$ 的波形以 τ 为横轴, 以 $\tau=0$ 处的纵坐标对称。

(2) $\tau=0$ 时的自相关值 $R_x(0)$ 是自相关函数 $R_x(\tau)$ 的最大值且等于均方值 ψ^2 。

证明：根据自相关定义及均方值 ψ^2 的定义式有

$$R_x(\tau=0) = E[x(t)x(t+\tau)] = E[x(t)x(t)] = E[x^2(t)] = \psi^2$$

故有下述关系成立

$$R_x(\tau=0) = \psi^2 = \sigma^2 + m^2 = R_x(0)$$

式中, σ^2 为方差; m 为平均值。

为了证明自相关函数的最大值在 $\tau=0$ 处, 即要证明

$$R_x(0) \geq R_x(\tau)$$

因为已知两个量的差的平方必为正数, 对于一样本集合有下式成立:

$$\begin{aligned} E\{[x(0)-x(\tau)]^2\} &\geq 0 \\ E\{x^2(0)+x^2(\tau)-2x(0)x(\tau)\} &\geq 0 \\ E[x^2(0)]+E[x^2(\tau)]-2E[x(0)x(\tau)] &\geq 0 \end{aligned} \quad (4-73)$$

因为各态历经的平稳随机过程统计平均值与时间无关, 故

$$E[x^2(0)] = E[x^2(\tau)] = R_x(0)$$

而 $E[x(0)x(\tau)] = R_x(\tau)$, 将上述结果代入式(4-73)中, 则有

$$R_x(0) - R_x(\tau) \geq 0$$

故

$$R_x(0) \geq R_x(\tau)$$

证毕。

(3) 平稳随机过程的自相关函数 $R_x(\tau)$ 不反映相位信息, 只与两个时刻的延时间隔 τ 有关。

(4) 在 $\tau=\infty$ 处的自相关函数 $R_x(\tau=\infty)$ 的数值为自相关函数的最小值, 并且等于均值的平方 m^2 , 即

$$R_x(\tau = \infty) = m^2$$

根据定义, 有

$$\lim_{\tau \rightarrow \infty} R_x(\tau) = E[x(t)x(t+\tau)]$$

可以认为, 当 $\tau \rightarrow \infty$ 时随机变量不相关, 表现为 $x(t)$ 与 $x(t+\tau)$ 数值有大有小, 符号有正有负, 如图 4-32 所示。当样本数目 N 足够大时, $x(t)$ 与 $x(t+\tau)$ 取值的坐标点相对均值 m_x 、 m'_x 呈对称分布。

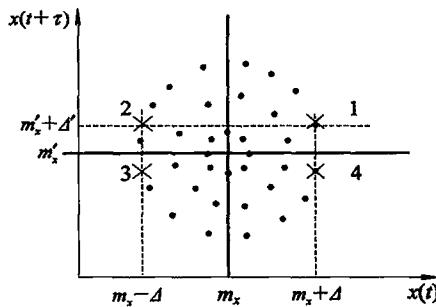


图 4-32 不相关图示

因为

$$m_x = E[x(t)] = m$$

$$m'_x = E[x(t+\tau)] = E[x(t)] = m$$

将 $x(t)$ 、 $x(t+\tau)$ 用平均值 m 与差值 Δ 或 Δ' 来表示。在图 4-32 中, 以取对称的四个坐标点为例, 说明如下:

	$x_i(t)$	$x_i(t+\tau)$	$x_i(t)x_i(t+\tau)$
点 1	$m+\Delta$	$m+\Delta'$	$m^2+m(\Delta+\Delta')+\Delta\Delta'$
点 2	$m-\Delta$	$m+\Delta'$	$m^2-m(\Delta-\Delta')-\Delta\Delta'$
点 3	$m-\Delta$	$m-\Delta'$	$m^2-m(\Delta+\Delta')+\Delta\Delta'$
点 4	$m+\Delta$	$m-\Delta'$	$m^2+m(\Delta-\Delta')-\Delta\Delta'$

将该四个对称点坐标值之积相加, 并平均, 则得

$$\frac{1}{4} \sum_{i=1}^4 x_i(t)x_i(t+\tau) = m^2$$

故 N 足够大时, 图 4-32 中的坐标点均相对平均值 m_x 、 m'_x 对称分布, 因此当 $\tau \rightarrow \infty$ 时, 样本总体有

$$\lim_{\tau \rightarrow \infty} R_x(\tau) = E[x(t)x(t+\tau)] = m^2$$

且为最小值。

2) 自相关函数的特点

(1) 确定性信号在所有 τ 值, 包括 $\tau \rightarrow \infty$, 都有自相关函数值不为平均值 m^2 的数值存在, 即 $R_x(\tau \rightarrow \infty) \neq m^2$ 。而随机信号在很小的 τ 值, 其自相关函数就降到平均值 m^2 。

(2) 确定性信号 $x(t)$ 是周期信号, 则它的自相关函数 $R_x(\tau)$ 也是周期函数, 其周期与信号 $x(t)$ 的周期相同, 下面举例说明。

[例] 求正弦函数 $x(t) = A \sin(\omega t + \varphi)$ 的自相关函数, 该正弦波的周期 $T = 2\pi/\omega$ 。

解 根据定义式有

$$\begin{aligned}\hat{R}_x(\tau) &= \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} x(t)x(t+\tau) dt \\ &= \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} A \sin(\omega t + \varphi) A \sin[\omega(t+\tau) + \varphi] dt\end{aligned}\quad (4-74)$$

令 $\omega t + \varphi = a$, 则 $dt = da/\omega$, 式(4-74)可写为

$$\begin{aligned}\hat{R}_x(\tau) &= \frac{A^2}{\omega T} \left[\int_{-\pi}^{\pi} \sin a \sin(a + \omega\tau) da \right] \\ &= \frac{A^2}{\omega T} \left[\int_{-\pi}^{\pi} \cos\omega\tau \sin^2 a da + \int_{-\pi}^{\pi} \sin\omega\tau \cos a \sin a da \right] \\ &= \frac{A^2}{\omega T} \left[\cos\omega\tau \int_{-\pi}^{\pi} \sin^2 a da + \sin\omega\tau \int_{-\pi}^{\pi} \cos a \sin a da \right] \\ &= \frac{A^2}{2} \cos\omega\tau\end{aligned}\quad (4-75)$$

式(4-75)表明, 正弦函数 $x(t)$ 的自相关函数 $R_x(\tau)$ 是与初相位 φ 无关的余弦函数。余弦函数的周期 $T = 2\pi/\omega$ 与正弦函数 $x(t)$ 相同, $R_x(\tau)$ 的波形如图 4-33 所示。

3) 自相关函数的应用举例

利用自相关函数有利于检测和识别淹没在随机噪声中的周期信号。现在求正弦波加随机噪声的自相关函数 $R_x(\tau)$ 。设输入信号为

$$x(t) = x_0(t) + N(t)$$

式中: $x_0(t)$ 为正弦波信号; $N(t)$ 为干扰噪声。

$x(t)$ 的自相关函数 $R_x(\tau)$ 为

$$\begin{aligned}R_x(\tau) &= \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T x(t)x(t+\tau) dt = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T [x_0(t) + N(t)] \times [x_0(t+\tau) + N(t+\tau)] dt \\ &= \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T [x_0(t)x_0(t+\tau) + x_0(t)N(t+\tau) + N(t)x_0(t+\tau) + N(t)N(t+\tau)] dt\end{aligned}$$

因为干扰噪声 $N(t)$ 、 $N(t+\tau)$ 与正弦信号 $x_0(t)$ 、 $x_0(t+\tau)$ 不相关, 故相乘项的积分为零。于是上式可写为

$$\begin{aligned}R_x(\tau) &= \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T [x_0(t)x_0(t+\tau) + N(t)N(t+\tau)] dt \\ &= R_{x_0}(\tau) + R_N(\tau)\end{aligned}\quad (4-76)$$

式中: $R_{x_0}(\tau)$ 为正弦波的自相关函数; $R_N(\tau)$ 为噪声的自相关函数。

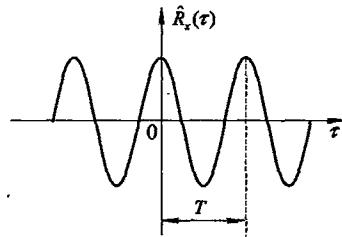


图 4-33 正弦函数的自相关函数波形图

由式(4-76)可见, 正弦波加噪声的自相关函数是正弦波自相关函数与随机噪声自相关函数之和。

图 4-34 分别给出了两个输入信号 $x_1(t)$ 与 $x_2(t)$ 的自相关函数图 $R_{x_1}(\tau)$, $R_{x_2}(\tau)$ 的波形。

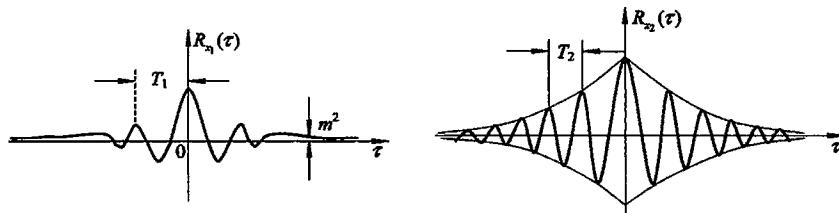


图 4-34 $x_1(t)$ 和 $x_2(t)$ 的自相关函数图

由图 4-34 可知, 信号 $x_1(t)$ 、 $x_2(t)$ 中均含有周期为 T_1 、 T_2 的正弦信号。其中信号 $x_1(t)$ 的均方值为 $\phi_1^2 = R_{x_1}(0)$, 且它具有直流分量, $m^2 \neq 0$; 而信号 $x_2(t)$ 的均方值为 $\phi_2^2 = R_{x_2}(0)$, 它没有直流分量, $m^2 = 0$ 。由 $R_{x_2}(\tau)$ 随 τ 的增大而衰减的速率较 $R_{x_1}(\tau)$ 衰减慢, 可知 $x_2(t)$ 中含有频带较窄的窄带噪声, 而 $x_1(t)$ 中含有频带较宽的宽带噪声。

图 4-35(a)记录的是放在汽车某处的垂直加速度传感器的输出波形 $x(t)$ 。汽车是在沙石搓板路上高速行驶的, 欲知汽车车身是否存在有规律的周期振动。为此, 我们计算记录信号 $x(t)$ 的自相关函数 $R_x(\tau)$, 如图 4-35(b)所示。 $R_x(\tau)$ 为一周期性衰减的偶函数, 从横坐标可读出周期 $T=0.1$ s, 则频率 $f=10$ Hz, 这说明汽车车身存在一个频率为 10 Hz 的有规律性的周期性振动。

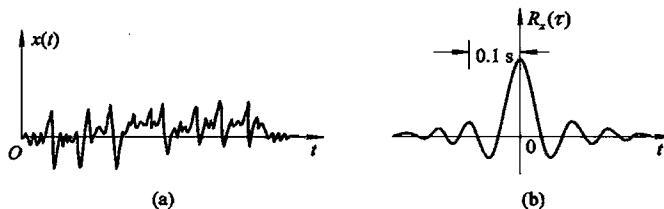


图 4-35 垂直加速度传感器的输出波形

3. 互相关函数的性质与特点

1) 性质

(1) 互相关函数不是偶函数。根据式(4-72)给出的定义式, 信号 $x(t)$ 、 $y(t)$ 的互相关函数为 $R_{xy}(\tau)$

$$R_{xy}(\tau) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T x(t)y(t+\tau) dt \quad (4-77)$$

令 $t' = t + \tau$, 则 $t = t' - \tau$, $dt = dt'$, 式(4-77)可写为

$$R_{xy}(\tau) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T y(t')x(t' - \tau) dt'$$

于是有

$$R_{xy}(\tau) = R_{yx}(-\tau)$$

但上式通常不成立, $R_{xy}(-\tau) \neq R_{xy}(\tau)$, $R_{xy}(\tau) \neq R_{yx}(\tau)$, $R_{yx}(-\tau) \neq R_{xy}(-\tau)$, 故

$$R_{xy}(\tau) \neq R_{yx}(-\tau)$$

所以, 互相关函数不是偶函数, 互相关函数与两信号的相位有关。

(2) 互相关函数 $R_{xy}(\tau)$ 在 $\tau=0$ 处的值 $R_{xy}(0)$ 不具有特征性。

(3) 互相关函数 $R_{xy}(\tau)$ 有时也有最大值, 即将两个信号中的某一个在时间上延迟 t_0 后, 则两信号最相关或最相似, 但最大值的位置与具体情况有关。

2) 互相关函数的特点

(1) 互相关函数的计算要用到时间延迟信息, 因而可用来测量有时间延迟信息的物理量(如速度、流量等), 而构成相关流量计, 这是一种先进的流速、流量测量方法。

(2) 基于相关原理可以将混叠在噪声中的特定频率的微弱有用信号提取出来, 或者说将噪声或不需要的信号从有用信号中分离出去。

3) 互相关函数在信号提取与消噪方面的应用

基于互相关原理进行信号提取与消噪的相关仪原理框图如图 4-36 所示。

相关仪的输入信号有两个: 一个是 $s(t)$, 是由干扰噪声 $n(t)$ 与淹没在其中的微弱被测信号 $c(t)$ 组成的:

$$s(t) = c(t) + n(t)$$

另一个是 $y(t)$, 是参考信号, 其频率应与所欲提取的有用信号的频率相同

$$y(t) = B \cos \omega_0 t$$

$s(t)$ 与 $y(t)$ 相乘后输出 $x(t) = s(t) \cdot y(t)$, 再经积分器输出 $U(t)$, 故 $U(t)$ 表达式如下

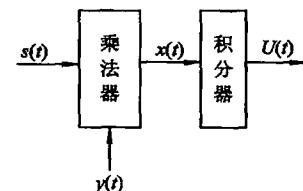


图 4-36 相关仪原理框图

$$U(t) = \frac{1}{T} \int_0^T s(t) \cdot y(t) dt \quad (4-78)$$

式(4-78)就是延时 $\tau=0$ 时, $s(t)$ 与 $y(t)$ 的互相关函数 $\hat{R}_{sy}(0)$, 即

$$\hat{R}_{sy}(0) = U(t) \quad (4-79)$$

故相关仪由乘法器与积分器组成, 其输出信号 $U(t)$ 为两个输入信号($s(t)$ 与 $y(t)$)在延时 $\tau=0$ 时的互相关函数。如何能将干扰噪声 $n(t)$ 排除并将有用被测信号 $c(t)$ 提出, 可分以下几种情况举例说明。

(1) 特定频率($\omega_0=2\pi f_0$)的有用信号淹没在随机干扰噪声 $n(t)$ 中, 在这种情况下

$$\begin{aligned} s(t) &= c(t) + n(t) = A \cos \omega_0 t + n(t), & c(t) &= A \cos \omega_0 t \\ y(t) &= B \cos(\omega_0 t + \varphi) \end{aligned}$$

将 $s(t)$ 与 $y(t)$ 表达式代入式(4-78), 得

$$\begin{aligned} U(t) &= \hat{R}_{sy}(0) = \frac{1}{T} \int_0^T s(t) \cdot y(t) dt \\ &= \frac{1}{T} \int_0^T [A \cos \omega_0 t + n(t)] \cdot B \cos(\omega_0 t + \varphi) dt \\ &= \frac{1}{T} \int_0^T A \cdot B \cdot \cos \omega_0 t \cdot \cos(\omega_0 t + \varphi) dt \\ &\quad + \frac{1}{T} \int_0^T B \cdot \cos(\omega_0 t + \varphi) \cdot n(t) dt \end{aligned}$$

因为随机噪声 $n(t)$ 与 $\cos(\omega_0 t + \varphi)$ 不相关, 上式第二项为零, 故

$$\begin{aligned} U(t) &= \hat{R}_{yy}(0) = \frac{AB}{T} \int_0^T \cos \omega_0 t \cdot \cos(\omega_0 t + \varphi) dt \\ &= \frac{AB}{T} \int_0^T \cos \omega_0 t \cdot [\cos \omega_0 t \cdot \cos \varphi - \sin \omega_0 t \cdot \sin \varphi] dt \\ &= \frac{AB}{T} \left[\cos \varphi \cdot \int_0^T \cos^2 \omega_0 t dt - \sin \varphi \int_0^T \cos \omega_0 t \cdot \sin \omega_0 t dt \right] \end{aligned}$$

由于 $\cos(\omega_0 t)$ 与 $\sin(\omega_0 t)$ 具有正交性, 上式第二项为零, 则

$$\begin{aligned} U(t) &= \hat{R}_{yy}(0) = \frac{AB}{T} \cos \varphi \cdot \int_0^T \cos^2 \omega_0 t dt = \frac{AB}{2} \cos \varphi \\ A &= \eta \cdot \hat{R}_{yy}(0) \end{aligned} \quad (4-80)$$

式中: $\eta = 2/B \cos \varphi$ 为常量; A 为被测有用信号 $c(t)$ 的幅值。

式(4-80)说明: 被测有用信号的幅值 A 与相关仪输出的相关值成正比, 比例常数 η 可由实验标定来确定。相关仪可以由硬件乘法器与积分器组成, 由积分器输出的信号 $U(t)$ 就是相关值 $\hat{R}_{yy}(0)$, 该信号与被测有用信号 $c(t)$ 的幅值 A 成比例, 且消除了随机噪声的影响。相关仪也可由软件模块来实现, 通过数采卡采集 $s(t)$ 与 $y(t)$ 两路信号, 获得离散时间序列值 $s(n)$ 、 $y(n)$, 相关值 $\hat{R}_{yy}(0)$ 由离散时间表达式来计算:

$$\hat{R}_{yy}(0) = \frac{1}{k} \sum_0^{k-1} s(n) \cdot y(n)$$

式中: k 为一周期 T 内的采样点数, $n=0, 1, \dots, k-1$ 。

(2) 特定频率($\omega_0 = 2\pi f_0$)的有用信号淹没在高频($\omega > \omega_0$)干扰噪声 $n(t)$ 中。在这种情况下, 高频干扰噪声可以有多个, 为简单计以一个高频干扰信号为例进行计算:

$$\begin{aligned} n(t) &= A_n \cdot \cos \omega t, \quad c(t) = A \cos \omega_0 t, \quad \frac{\omega}{\omega_0} \geq 5 \sim 10 \\ s(t) &= A \cos \omega_0 t + A_n \cos \omega t \\ y(t) &= B \cdot \cos(\omega_0 t + \varphi) \end{aligned}$$

将 $s(t)$ 、 $y(t)$ 表达式代入(4-78)中, 得

$$\begin{aligned} U(t) &= \hat{R}_{yy}(0) = \frac{1}{T} \int_0^T [A \cos \omega_0 t + A_n \cos \omega t] \cdot B \cos(\omega_0 t + \varphi) dt \\ &= P + Q \end{aligned}$$

其中 P 为第一部分:

$$P = \frac{1}{T} \int_0^T A \cdot B \cos \omega_0 t \cdot \cos(\omega_0 t + \varphi) \cdot dt = \frac{A}{\eta} = \frac{AB}{2} \cos \varphi \quad (\text{已证})$$

Q 为第二部分:

$$Q = \frac{1}{T} \int_0^T A_n \cdot \cos \omega t \cdot B \cos(\omega_0 t + \varphi) \cdot dt \quad (4-81)$$

$$Q = \frac{1}{T} \int_0^T A_n \cdot B \cos \omega t (\cos \omega_0 t \cdot \cos \varphi - \sin \omega_0 t \cdot \sin \varphi) \cdot dt$$

$$Q = \frac{1}{T} \cos \varphi \int_0^T A_n \cdot B \cos \omega t \cdot \cos \omega_0 t dt - \frac{A_n B}{T} \sin \varphi \int_0^T \cos \omega t \cdot \sin \omega_0 t dt \quad (4-82)$$

再由三角函数积化和差公式进行分项，得

$$Q = \frac{A_n B \cos \varphi}{2T} \int_0^T [\cos(\omega + \omega_0)t + \cos(\omega - \omega_0)t] dt$$

$$- \frac{A_n B \sin \varphi}{2T} \int_0^T [\sin(\omega + \omega_0)t - \sin(\omega - \omega_0)t] dt$$

于是积分后有：

$$Q = \frac{A_n B \cos \varphi}{2T} \left[\frac{1}{\omega + \omega_0} \sin(\omega + \omega_0)t + \frac{1}{\omega - \omega_0} \sin(\omega - \omega_0)t \right]$$

$$+ \frac{A_n B \sin \varphi}{2T} \left[\frac{1}{\omega + \omega_0} \cos(\omega + \omega_0)t - \frac{1}{\omega - \omega_0} \cos(\omega - \omega_0)t \right] \quad (4-83)$$

由式(4-83)可见，当 ω 足够大时，可使 $Q \ll P$ 至允许的误差范围，于是

$$U(t) = P = \frac{A}{\eta}, \quad A = \eta \cdot U(t) = \eta \cdot \hat{R}_{yy}(0) \quad (4-84)$$

式中 $\eta = 2/B \cos \varphi$ 为一常量， A 为被测信号 $c(t)$ 的幅值。

相关仪能够对大于被测信号频率的高频干扰信号有不同程度的衰减，频率越高的干扰信号衰减率越大，近似以与高频 ω 成反比即 ω^{-1} 规律衰减。

(3) 对调幅波进行解调。

① 调幅波。当角频率之比 $\omega/\Omega \geq 5 \sim 10$ 的两个信号相乘时，可实现低频(Ω)信号 $c(t)$ 对高频(ω)信号幅值的调制，调制波的数学表述为

$$P(t) = c(t) \cdot e(t) = P_m \cdot \cos(\omega t + \varphi), \quad P_m = E \cdot c(t) \quad (4-85)$$

式中 $e(t) = E \cos(\omega t + \varphi)$ 为高频信号，称为载波； $c(t)$ 为被测低频($\omega/\Omega \geq 5 \sim 10$)信号，称为调制波； P_m 为被 $c(t)$ 调制的调幅波的幅值。

② 基于相关法的调幅波解调器。采用如图 4-36 所示的相关仪可将调制波 $c(t)$ 从调幅波 $P(t)$ 中分离解调出来，从而被测低频信号 $c(t)$ 得以测量。用于此目的的相关仪称为相关解调器。

输入相关解调器的两个信号分别是 $s(t)$ 与 $y(t)$ 。调幅波 $P(t)$ 淹没在随机噪声 $n(t)$ 中，参考信号 $y(t)$ 的频率与载波频率(ω)相同，有

$$s(t) = P(t) + n(t) = P_m \cos(\omega t + \varphi) + n(t) = E \cdot c(t) \cdot \cos(\omega t + \varphi) + n(t)$$

$$y(t) = B \cdot \cos(\omega t)$$

将 $s(t)$ 与 $y(t)$ 代入式(4-78)中，得

$$U(t) = \hat{R}_{yy}(0) = \frac{1}{T} \int_0^T s(t) \cdot y(t) dt$$

$$= \frac{1}{T} \int_0^T [E \cdot c(t) \cdot \cos(\omega t + \varphi) + n(t)] \cdot B \cdot \cos(\omega t) dt$$

$$= \frac{1}{T} \left[\int_0^T E \cdot B \cdot c(t) \cdot \cos(\omega t + \varphi) \cdot \cos(\omega t) dt + \int_0^T B \cdot n(t) \cdot \cos(\omega t) dt \right]$$

因为 $\cos(\omega t)$ 与随机噪声 $n(t)$ 不相关，上式第二项为零，故

$$U(t) = \hat{R}_{yy}(0) = \frac{E \cdot B}{T} \int_0^T c(t) \cdot \cos(\omega t + \varphi) \cdot \cos(\omega t) dt \quad (4-86)$$

在 $c(t)$ 随时间变化的角频率 $\Omega \ll \omega$ 的条件下，一般至少满足 $\omega/\Omega \geq 5 \sim 10$ ， $c(t)$ 与 $\cos(\omega t + \varphi)$ 或 $\cos(\omega t)$ 相比较可近似为一常量，因而可从积分号内提出来，故

$$U(t) = \hat{R}_{xy}(0) = \frac{E \cdot B}{T} \cdot c(t) \int_0^T \cos(\omega t + \varphi) \cdot \cos \omega t \cdot dt \quad (4-87)$$

求解上式积分可仿照式(4-80)进行, 最后可得

$$\begin{aligned} U(t) &= \hat{R}_{xy}(0) = \frac{E \cdot B}{2} \cos \varphi \cdot c(t) \\ c(t) &= \xi \cdot \hat{R}_{xy}(0) \end{aligned} \quad (4-88)$$

式中 $\xi = 2/(EB \cos \varphi)$ 为常量; $c(t)$ 为被测有用信号。

(4) 相关仪中的积分器。根据前述三种情况的讨论可见, 按照式(4-80)、(4-84)、(4-88), 通过计算互相关 $\hat{R}_{xy}(0)$ 可以获得被测有用信号 $c(t)$ 或其幅值 A , 从随机噪声、高频干扰或载波信号中把微弱信号提出来。由式(4-78)给出的积分式也可以是多重积分, 常量照样可以从 n 重积分号中提出来; 正交分量的 n 重积分及不相关项的 n 重积分仍为零, 对高频 ω 干扰项呈 ω^{-n} 衰减, 这时的 n 重积分器就是 n 阶低通滤波器。相关仪中的积分器就是低通滤波器, 在低于截止频率的通带以允许的衰减率通过低频信号, 大于截止频率的高频干扰信号将受到不同程度的呈 ω^{-n} 规律的衰减。低通滤波器的设计参见[示例 4-3]。

4.5.3 频域谱分析法

利用快速傅里叶变换算法(FFT)对输入信号进行离散傅里叶变换得到时域输入信号 $x(t)$ 的频谱 $X(m)$ 。 $m=0$ 时的谱线 $X(0)$ 为直流分量; $m=1$ 时的谱线 $X(1)$ 为基波分量; $m>1$ 时的谱线 $X(m)$ 为谐波分量。基波频率 f_1 为

$$m = 1, \quad f_1 = \frac{1}{t_p} = \frac{1}{NT} \quad (4-89)$$

式中: T 为采样时间间隔; N 为采样点数; t_p 为信号采样记录时间。

根据频谱的谐波性可知, $m>1$ 时的谐波频率 f_m 为

$$m > 1, \quad f_m = mf_1 = \frac{m}{NT} \quad (4-90)$$

m 为谱线序数, $m=1, 2, \dots, N-1$, 总计有 N 条谱线, 其中有 $N/2$ 条是有效的。

设图 4-37(a)所示为某一信号 $x(t)$ 的频谱 $X(m)$ 。它是由对 $x(t)$ 进行采样, 得离散时间序列 $x(n)$ ($n=0, 1, 2, \dots, N-1$), 再对 $x(n)$ 进行快速傅里叶变换(FFT)得到的。对 $x(n)$ 进行低通、高通、带通滤波的实现方法分述如下。

1. 低通滤波器的实现

首先设定低通滤波器的截止频率 f_c , 将 f_c 的值与 $X(m)$ 的各条谱线的频率值作比较, 若

$$f_{j-1} < f_c < f_j$$

则令 $m \geq j$ 的谐波幅值为零, 即

$$X(m \geq j) = 0$$

保留 $m < j$ 的各条谱线幅值不变, 从而得到图 4-37(b)所示的频谱 $X_L(m)$ ($m=0, 1, 2, \dots, j-1$)。这就是经过理想低通滤波器后的信号 $x_L(t)$ ($x_L(n)$) 的频谱, 对 $X_L(m)$ 进行傅里叶反变换(IFFT)就得到离散序列 $x_L(n)$ 。

2. 高通滤波器的实现

首先设定高通滤波器的截止频率 f_H , 将 f_H 的值与 $X(m)$ 的各条谱线的频率值作比较, 若

$$f_{i-1} < f_H < f_i$$

则令 $m \leq i-1$ 的各次谐波幅值为零, 即

$$X(m \leq i-1) = 0$$

但保留 $m \geq i$ 的各次谐波幅值不变, 从而得到图 4-37(c) 所示的频谱 $X_H(m)$ ($m = i, i+1, \dots, N/2-1$)。这就是经过理想高通滤波器的信号 $x_H(t)$ ($x_H(n)$) 的频谱。对 $X_H(m)$ 进行傅里叶反变换(IFFT)就得到离散序列 $x_H(n)$ 。

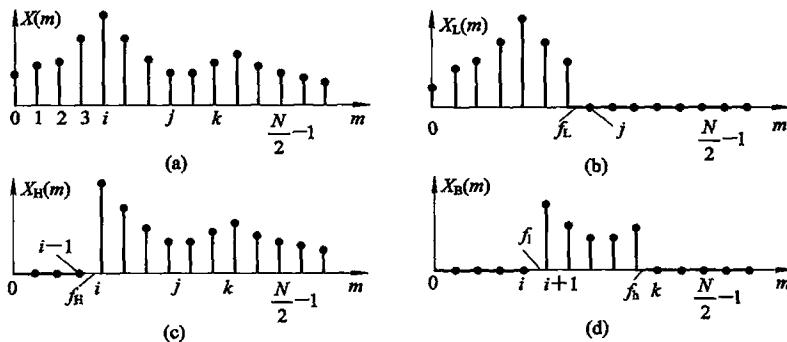


图 4-37 $x(t)$ 的频谱及滤波后频谱

3. 带通滤波器的实现

将带通滤波器的上、下限截止频率 f_L 、 f_U 与 $X(m)$ 的各条谱线的频率值作比较, 若上限:

$$f_{k-1} < f_L < f_k$$

下限:

$$f_i < f_U < f_{i+1}$$

则令 $m \leq i$ 的各次谐波幅值及 $m \geq k$ 的各次谐波幅值为零, 而保留 $i < m < k$ 的各条谱线幅值不变, 从而得到图 4-37(d) 所示的频谱 $X_B(m)$, $i < m < k$, 这就是经过理想带通滤波器后的信号 $x_B(t)$ ($x_B(n)$) 的频谱。欲得其时域函数值 $x_B(n)$, 对其频谱 $X_B(m)$ 进行傅里叶反变换(IFFT)即可。

综上所述可见, 频域谱分析法进行滤波, 实现起来非常方便灵活。对输入信号采样后进行快速傅里叶变换获得其频谱 $X(m)$, 去掉不希望的谱线后即构成所希望的频率特性 $X'(m)$, 再对其进行傅里叶反变换(IFT)即可获得时域信号。该时域信号就是经过滤波后的输出信号。由频域谱分析法实现的低、高、带通滤波器均是锐截止的理想滤波器。

4.5.4 [示例 4-3] 交流电桥调幅波解调器中滤波器参数的确定

功能: 该滤波器与乘法器相配合构成调幅波解调器, 可用于对交流电桥等输出调幅波进行解调的场合。

设已知交流电桥供电电源频率 $f_0 = 10 \text{ kHz}$, 被测信号的最高频率小于 2 kHz 。

要求：

- 试确定滤波器的参数与类型；
- 该滤波器应有使低频信号通过的能力，对频率小于2 kHz的被测信号衰减率不能大于30%；
- 该滤波器对高频信号有足够的衰减能力，使通过滤波器的高频信号的衰减率至少达到90%以上；
- 分析乘法器输出信号的频率成分。

1. 不平衡交流电桥调制器

交流电桥电路输出的不平衡电压 U_{BD} (如图4-38所示)是一调幅波，故交流电桥为调幅波调制器。桥路的四个桥臂(或两个桥臂)为电参数型变换器。电参数的变化量 ΔR 与待测参量(目标参量) $c(t)$ 的变化成比例。根据电桥理论，在恒压源供电情况下，输出的不平衡电压 U_{BD} 与供电电压 $E=E_m \cos \omega_0 t$ 及桥臂阻抗变化成比例：

$$U_{BD} \propto E \cdot \frac{\Delta R}{R}$$

又桥臂电参数变化与被测参量 $c(t)$ 成比例：

$$\frac{\Delta R}{R} \propto c(t)$$

故有

$$U_{BD} \propto E_m c(t) \cdot \cos \omega_0 t \quad (4-91)$$

式(4-91)与调幅波表达式(4-85)形式相同，故 U_{BD} 为一调幅波，供电电源电压为载波， $c(t)$ 为调制波。

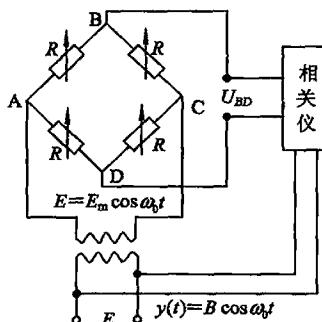


图4-38 不平衡交流电桥调制器

2. 调幅波的频率成分

假设被测参量 $c(t)$ 为单频信号， $c(t)=A \cos \Omega t$ 。则

$$U_{BD} = A \cos \Omega t \cdot E_m \cos \omega_0 t$$

利用三角函数积化和差公式将上式分项，可得

$$\begin{aligned} U_{BD} &= \frac{A \cdot E_m}{2} [\cos(\omega_0 + \Omega)t + \cos(\omega_0 - \Omega)t] \\ &= \frac{A \cdot E_m}{2} [\cos 2\pi(f_0 + f)t + \cos 2\pi(f_0 - f)t] \end{aligned} \quad (4-92)$$

已知 $f_0 = 10 \text{ kHz}$, $f \leq 2 \text{ kHz}$, 所以电桥输出的调幅波 U_{BD} 由两个频率信号合成: 一个是 $f_0 + f = 12 \text{ kHz}$, 另一个是 $f_0 - f = 8 \text{ kHz}$ 。

3. 乘法器输出信号的频率成分

采用图 4-36 所示相关仪作调幅波解调器, 两个输入信号分别是

$$s(t) = U_{BD}$$

$$y(t) = B \cos\omega_0 t$$

则乘法器输出 $x(t)$ 为

$$\begin{aligned} x(t) &= s(t) \cdot y(t) \\ &= \frac{ABE_m}{2} [\cos(\omega_0 + \Omega)t + \cos(\omega_0 - \Omega)t] \cdot B \cos\omega_0 t \\ &= \frac{ABE_m}{2} [\cos(\omega_0 + \Omega)t \cos\omega_0 t + \cos(\omega_0 - \Omega)t \cos\omega_0 t] \end{aligned}$$

利用三角函数积化和差公式将上式分项可得

$$x(t) = \frac{ABE_m}{2} \cos\Omega t + \frac{ABE_m}{4} [\cos(2\omega_0 + \Omega)t + \cos(2\omega_0 - \Omega)t] \quad (4-93)$$

由式(4-93)可知, 乘法器输出信号 $x(t)$ 由三个频率信号合成: 其一是 $2f_0 + f = 22 \text{ kHz}$, 其二是 $2f_0 - f = 18 \text{ kHz}$, 其三是被测低频信号 $c(t)$ 的频率 $\Omega (2 \text{ kHz})$ 。

4. 低通滤波器参数的确定

低通滤波器的转折频率 $f_1 > 2 \text{ kHz}$, 满足对被测低频信号 $c(t)$ 通过的要求(在积分时可以从积分号中提出来); 若令 $f_1 = 2 \text{ kHz}$ 时, 其衰减率为 29.3%, 满足不大于 30% 的要求。

如果采用一阶积分器, 过渡带幅频特性的下降规律是: $10f_1$ 处幅值下降至 $1/10$, 即 $22 \text{ kHz} > f_1 = 20 \text{ kHz}$ 频率分量幅值衰减率可达到 90%。但对于 18 kHz 信号分量的幅值衰减率达不到 90%, 若欲达到更高的衰减率则应采用二阶或二阶以上的低通滤波器。滤波器参数初定后可进行仿真设计及检验, 合格后方可最后确定。

4.5.5 [示例 4-4] 设计一个巴特沃斯低通数字滤波器

要求: 该低通数字滤波器等效模拟滤波器 $H_d(s)$ 幅频特性过渡段特征是:

- 对信号频率 $f_1 = 70 \text{ Hz}$ 的衰减率 $\delta_1 \leq 0.293$;
- 对信号频率 $f_2 = 350 \text{ Hz}$ 的衰减率 $\delta_2 \geq 0.8$ 。

写出巴特沃斯低通数字滤波器 $H_d(z)$ 的计算机编程算式 $y(j)$ 。

1. 等效模拟低通滤波器传递函数 $H(s)$ 的确定

分别求出阶次 n 及截止角频率 ω_c , 则 $H(s)$ 即可确定。

(1) 确定等效模拟低通滤波器阶次 n 。

根据式(4-53)至式(4-55)的推导, 阶次 n 应满足式(4-55), 即

$$\left(\frac{f_2}{f_1}\right)^n = \sqrt{\frac{\left(\frac{1}{A_2^2} - 1\right)}{\left(\frac{1}{A_1^2} - 1\right)}}$$

由式(4-52)可知幅值比 A_1 、 A_2 :

$$A_1 = |H(\omega_1)| = 1 - \delta_1 \geq 0.707$$

$$A_2 = |H(\omega_2)| = 1 - \delta_2 \leq 0.2$$

将 A_1 、 A_2 的值代入式(4-55)，有

$$\left(\frac{f_2}{f_1}\right)^n = \left(\frac{350}{70}\right)^n \geq \frac{4.898\ 978\ 486}{1.000\ 302\ 146} = 4.897\ 499\ 218$$

因为 $f_2/f_1=5$ 且大于 4.897 492 18，故取 $n=1$ 。

(2) 确定等效模拟低通滤波器 $H(s)$ 的截止角频率 ω_c 。

已知 $n=1$, $f_1=70$ Hz 及 $f_2=350$ Hz, 代入式(4-54)中任一式中均可求出 ω_c 。现由式(4-54a)求解 ω_c :

$$\left(\frac{\omega_1}{\omega_c}\right)^{2n} = \frac{1}{A_1^2} - 1$$

有

$$\left(\frac{\omega_1}{\omega_c}\right)^2 = \frac{1}{(0.707)^2} - 1 = 1.000\ 604\ 182$$

故 $\omega_c \approx \omega_1 = 2\pi f_1 = 2\pi \times 70 = 439.8$ rad/s。

(3) 求模拟滤波器的传递函数 $H(s)$ 。根据式(4-46)可得:

$$H(s) = \frac{1}{\tau s + 1}$$

式中 $\tau = \frac{1}{\omega_c} = \frac{1}{439.8}$ s = 2.274 ms。

2. 等效数字滤波器 $H_D(z)$

(1) 采用后向差分法，令 $s = \frac{1-z^{-1}}{T}$ ，则

$$H_D(z) = H(s) \Big|_{s=\frac{1-z^{-1}}{T}} = \frac{1}{\tau s + 1} = \frac{\omega_c}{s + \omega_c} \Big|_{s=\frac{1-z^{-1}}{T}}$$

$$H_D(z) = \frac{\omega_c T}{1 + \omega_c T - z^{-1}} = \frac{Y(z)}{X(z)}$$

经 Z 反变换得差分方程后，等效数字滤波器 $H_D(z)$ 的计算机编程算式如下：

$$y(j) = qx(j) + py(j-1)$$

式中， $q = \frac{\omega_c T}{1 + \omega_c T} = \frac{T}{\tau + T}$; $p = \frac{1}{1 + \omega_c T} = \frac{\tau}{\tau + T}$; T 为采样间隔。与式(4-70)比较： $a_0 = q$, $a_1 = 0$; $b_1 = p$, $b_2 = b_3 = 0$ 。

(2) 采用双线性变换法，令 $s = \frac{2z-1}{Tz+1}$ ，可得等效数字滤波器的编程算式:

$$y(j) = \frac{1 - \frac{\omega_c T}{2}}{1 + \frac{\omega_c T}{2}} y(j-1) + \frac{\frac{\omega_c T}{2}}{1 + \frac{\omega_c T}{2}} x(j) + \frac{\frac{\omega_c T}{2}}{1 + \frac{\omega_c T}{2}} x(j-1)$$

与式(4-70)比较：

$$a_0 = \frac{\frac{\omega_c T}{2}}{1 + \frac{\omega_c T}{2}}; \quad a_1 = a_0; \quad b_1 = \frac{1 - \frac{\omega_c T}{2}}{1 + \frac{\omega_c T}{2}}$$

要注意：

- 不同等效方法求得的无限冲激响应滤波器(IIR)模型的系数 a_r ($r=0, 1, 2, \dots, N$)， b_k ($k=1, 2, \dots, M$)并不相同，但都满足设计要求的巴特沃斯低通滤波器特性。
- 不论是仿真时对信号的采样间隔 T ，还是数据采集卡对信号的采样间隔 T ，对滤波器系数 a_r 、 b_k 都有影响。因此在仿真设计时的 T 值与实际数据采集系统的数据采集间隔 T 必须相同，才能保证滤波器系数一致时实际的智能系统中使用的滤波器与仿真设计时的滤波器有相同的滤波器功能。

4.5.6 [示例 4-5] 设计一个巴特沃斯高通数字滤波器 $H_g(z)$

要求：该高通数字滤波器等效模拟滤波器 $H_g(z)$ 幅频特性过渡段特征是：

- 对信号频率 $f_1 = 4.0$ Hz 的衰减率要达到 $\delta_{1g} \geq 0.707$ ；
- 对信号频率 $f_2 = 40$ Hz 的衰减率要达到 $\delta_{2g} \leq 0.2$ 。

写出巴特沃斯高通数字滤波器 $H_g(z)$ 的计算机编程算式 $y(k)$ 。

1. 等效模拟高通滤波器传递函数 $H_g(s)$ 的确定

(1) 确定 $H_g(s)$ 的阶数。为此先设计一个模拟低通滤波器 $H_d(s)$ ，其幅频特性过渡段特征是：

- 对信号频率 $f_1 = 4.0$ Hz 的衰减率 $\delta_{1d} \leq 1 - \delta_{1g} \leq 0.293$ ；
- 对信号频率 $f_2 = 40$ Hz 的衰减率 $\delta_{2d} \geq 1 - \delta_{2g} \geq 0.8$ 。

设计步骤与[示例 4-4]相同，可得巴特沃斯滤波器阶数 n 应满足不等式：

$$\left(\frac{f_2}{f_1}\right)^n = \left(\frac{40}{4}\right)^n \geq 4.897\ 500\ 2$$

因为 $f_2/f_1 = 40/4 = 10 > 4.897\ 500\ 2$ ，故只需 $n=1$ 即可。

于是一阶模拟巴特沃斯低通滤波器的传递函数 $H_d(s)$ 为

$$H_d(s) = \frac{b_0}{b_1 s + b_0} = \frac{1}{as + 1}$$

$$a = \frac{b_1}{b_0} = \frac{1}{\omega_{cd}}$$

(2) 确定等效模拟高通滤波器的传递函数 $H_g(s)$ 。由已确定的阶次 $n=1$ ，根据式(4-57)，可得

$$H_g(s) = \frac{a_g s}{1 + a_g s}, \quad a_g = \frac{1}{\omega_{cg}}$$

(3) 确定等效模拟高通滤波器的截止角频率 ω_{cg} 。根据对 $\omega_1 = 2\pi f_1$ ， $\omega_2 = 2\pi f_2$ 处衰减率的要求，可计算相应的幅值 A_1 与 A_2 ：

$$A_1 = \frac{f_1/f_{cg}}{\sqrt{1 + (f_1/f_{cg})^2}} = 1 - \delta_{1g} \leq 0.293 \quad ①$$

$$A_2 = \frac{f_2/f_{cg}}{\sqrt{1 + (f_2/f_{cg})^2}} = 1 - \delta_{2g} \geq 0.8 \quad ②$$

采用上述①、②两式中任一式，均可求得

$$f_{cg} \geq 13 \text{ Hz}, \quad \omega_{cg} \geq 81.7 \text{ rad/s}$$

2. 确定与 $H_g(s)$ 等效的数字滤波器 $H_g(z)$ 的计算机编程算式

(1) 双线性变换法：

$$y(j) = \frac{1}{2+aT} [(2-aT)y(j-1) + 2x(j) - 2x(j-1)]$$

式中，无限冲激响应(IIR)滤波器系数 $a_0 = \frac{2}{2+aT}$; $a_1 = -\frac{2}{2+aT}$; $b_1 = \frac{2-aT}{2+aT}$ 。

(2) 后向差分法：

$$y(j) = \frac{1}{1+aT} [y(j-1) + x(j) - x(j-1)]$$

式中，无限冲激响应(IIR)滤波器系数 $a_0 = \frac{1}{1+aT}$; $a_1 = -\frac{1}{1+aT}$; $b_1 = \frac{1}{1+aT}$, T 为采样间隔。

两种变换法所得输出脉冲序列 $y(j)$ ($j=1, 2, \dots$) 都对应模拟高通滤波器：

$$H_g(s) = \frac{a_g s}{s + a_g}, \quad a_g = \frac{1}{\omega_{cg}}$$

式中， ω_{cg} 为高通滤波器的截止角频率。

4.5.7 [示例 4-6]二阶带通数字滤波器在 LabVIEW 平台上的实现、检验及向 MPU 移植与复现

设计任务：在 LabVIEW 平台上设计一个二阶带通数字滤波器 $H_D(z)$ 与二阶模拟带通滤波器 $H(s)$ 等效。 $H(s)$ 的类型为带通巴特沃兹滤波器。低、高截止频率分别为 40 Hz 与 60 Hz。过渡带的斜率不大于 -40 dB/10 倍频程。

要求：仿真检验所设计的滤波器 $H_D(z)$ 是否与 $H(s)$ 等效。

给出该等效数字滤波器向微处理器中复现移植的编程算式。

1. 在基于 LabVIEW 的“虚拟巴特沃兹滤波器设计仪”平台上设计数字滤波器

1) “虚拟巴特沃兹滤波器设计仪”基本功能

虚拟巴特沃兹滤波器设计仪具有如下主要功能：

(1) 滤波器参数设置。在滤波器参数设置框内有五个数字输入控件，用于对五个参数进行设置，以便于进行优化设计。可设置：

- 采样频率 f_s (Hz)、低截止频率 f_l (Hz)、高截止频率 f_h (Hz)、滤波器阶次 n (必须大于零)；

- 滤波器类型有低通(形参 f_l 无效)、高通(形参 f_h 无效)、带通和带阻四类可选。

(2) 输出功能。输出功能有两种形式：

- 可对滤波器前后输入、输出的时序信号进行波形显示，以便于对滤波器进行检验；
- 可输出滤波器系数 a_k ($k=1, 2, \dots, N$) 和 b_r ($r=0, 1, \dots, M$)，以供移植时使用。

(3) 仿真自检功能。提供一个正弦波发生器，其频率、相位、幅值、采样频率可选，以便向所设计的滤波器提供一个幅值恒定、频率可调的正弦波，基于扫频法以检验滤波器的幅频特性是否符合预期要求。

基本功能(1)、(2)部分的前面板如图 4-39 所示。

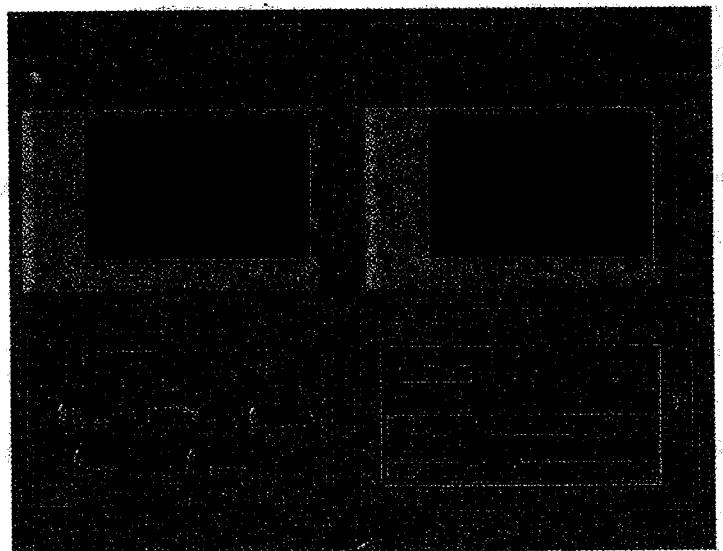


图 4-39 虚拟巴特沃兹滤波器设计仪前面板

2) 设计提示

(1) 巴特沃兹滤波器图标的调用路径:

[Functions] → [Analyze] → [Signal Processing] → [Filters] → [Butterworth Filter. vi]

(2) 获取巴特沃兹滤波器系数的图标及调用路径:

[Functions] → [Analyze] → [Signal Processing] → [Filters] → [Advanced IIR Filtering] → [Butterworth Coefficients. vi]

以上两个图标分别如图 4-40(a)、(b)所示。

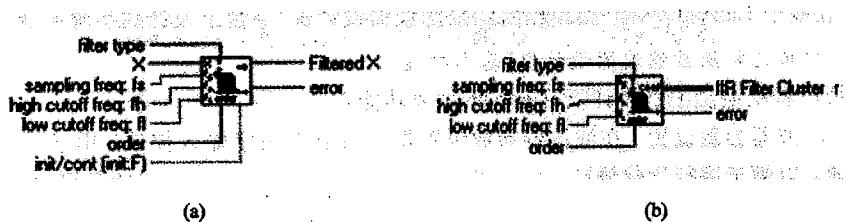


图 4-40 LabVIEW 中巴特沃兹滤波器图标

(a) 巴特沃兹滤波图标; (b) 获取巴特沃兹数字滤波器系数的图标

(3) 流程图。巴特沃兹滤波器的流程图编辑窗口如图 4-41 所示。其中 x 是实际采集到的外部信号，也可以是内部正弦波发生器提供的正弦信号。用 Butterworth Filter. vi 图标可实现对信号 x 的巴特沃兹滤波，用 Butterworth Coefficients. vi 图标可输出所设计巴特沃兹滤波器的滤波系数 a_k ($k = 1, 2, \dots, N$) 和 b_r ($r = 0, 1, \dots, M$)。

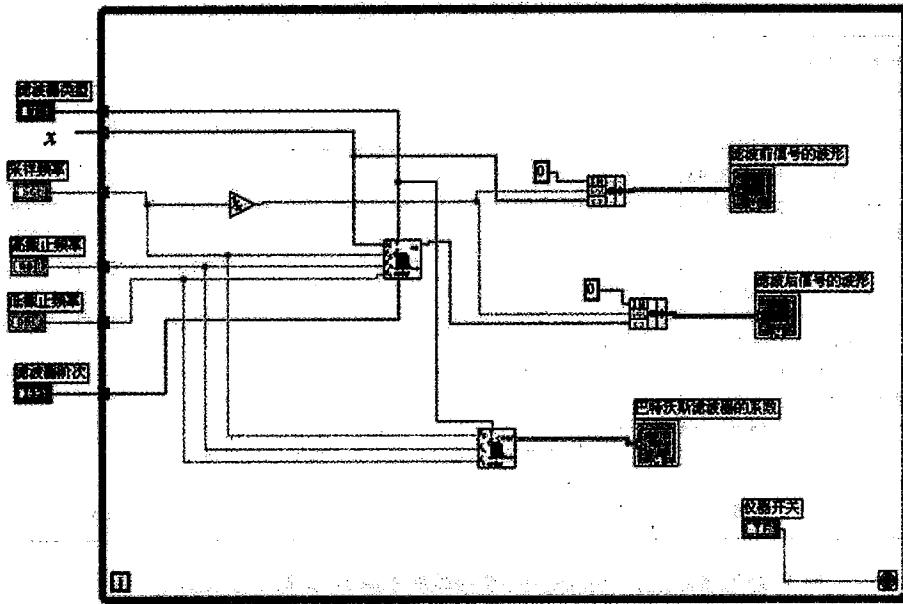


图 4-41 虚拟巴特沃兹滤波器流程图编辑窗口

2. 在虚拟巴特沃兹滤波器设计仪上设计检验二阶带通数字滤波器

设计数字滤波器时，数字滤波器模型即滤波器系数以及高低截止频率、滤波器阶数与采样频率有关，因此设计检验好的滤波器模型在移植检验时应采用相同的采样频率。

1) 二阶带通数字滤波器的设计

按照设计任务的要求对滤波器参数进行设置。

- (1) 低截止频率 $f_l = 40 \text{ Hz}$ 。
- (2) 高截止频率 $f_h = 60 \text{ Hz}$ 。
- (3) 滤波器阶数：根据过渡带斜率要求不高于 $-40 \text{ dB}/10$ 倍频程，确定阶数 $n = 2$ 。
- (4) 采样频率 f_s ：根据采样定理采样频率 f_s 至少应为被测信号的二倍，一般取 20 倍。由于带通滤波器的带宽上限为 60 Hz，故应 $f_s > 20 \times 60 = 1200 \text{ Hz}$ ，设置为 1400 Hz。

2) 二阶带通数字滤波器的检验

基于扫频法测定待检数字滤波器的幅频特性可利用自检仿真功能进行。记录不同信号频率 f 时滤波器输入、输出正弦波的幅值，并计算幅值比 $A = |\text{输出}| / |\text{输入}|$ 、对数幅值比 $L = 20 \lg A$ ，画出 $A - f$ 、 $L - f$ 关系图。输入信号幅值恒定为 5 V，重点观测数字滤波器的输出，在频率 $f_l = 40 \text{ Hz}$, $f_h = 60 \text{ Hz}$ 时是否幅值比 $A \geq 0.707$? 在 10 倍频程 $10f = 10 \times 60 = 600 \text{ Hz}$ 时，幅值比是否达到 0.01?

当频率 f 变化时，幅值比与频率的关系如表 4-7 所示。当频率为 40 Hz 及 60 Hz 时，幅值比均为 0.7，对数幅值均为 -3.10 dB ；当频率为 600 Hz 时，幅值比为 8×10^{-5} ，对数幅值为 -81.94 dB ，带宽与过渡带均满足要求。幅值比 A 与频率 f 的关系图即幅频特性如

图 4-42 所示。

表 4-7 二阶带通数字滤波器幅值与频率关系

频率 f/Hz	0	5	10	15	20	25	30	35
A	0	0.0018	0.0076	0.0192	0.04	0.08	0.158	0.335
L/dB	-113.98	-54.99	-42.38	-34.33	-27.96	-21.94	-16.03	-9.50
频率 f/Hz	40	45	47	49	51	53	55	60
A	0.7	0.996	1.0	1.0	1.0	1.0	1.0	0.7
L/dB	-3.10	-0.035	0	0	0	0	0	-3.10
频率 f/Hz	65	70	75	500	550	600	650	700
A	0.45	0.3	0.205	0.0002	0.0001	0.0000	0.0000	0.0000
L/dB	-6.94	-10.46	-13.76	-72.40	-75.92	-81.94	-93.98	-113.98

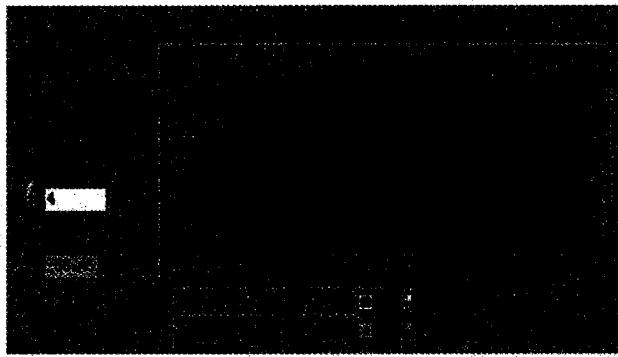


图 4-42 LabVIEW 平台下二阶带通数字滤波器幅频特性图

3. 二阶带通数字滤波器的移植与复制

1) 获取带通滤波器的系数

滤波器的幅频特性经检验合格后记录该滤波器的系数 a_k 、 b_j 分别为

$$a_1 = -3.7799, a_2 = 5.4485, a_3 = -3.5472, a_4 = 0.8808$$

$$b_0 = 0.0019, b_1 = 0.0000, b_2 = -0.0038, b_3 = 0.0000, b_4 = 0.0019$$

2) 写出移植算式

移植算式为

$$y_i = \frac{1}{a_0} \left[\sum_{j=0}^{N_b-1} b_j x_{i-j} - \sum_{k=1}^{N_a-1} a_k y_{i-k} \right] \quad (4-94)$$

式(4-94)是 LabVIEW 平台上的数字滤波器差分方程，其中 N_b 是正向系数个数， b_j 是正向系数， N_a 是反向系数个数， a_k 是反向系数。

又有

$$y(n) = \sum_{r=0}^N a_r x(n-r) - \sum_{k=1}^M b_k y(n-k) \quad (4-95)$$

式(4-95)是 Matlab 环境下的数字滤波器差分方程，其中 a_r 是正向系数， b_k 是反向系数。

经过比较分析，上述两个差分方程虽然形式不一样，但是实际内涵是一样的。

将求得的系数数值代入式(4-95)，该带通滤波器的编程式为

$$\begin{aligned} y(i) = & 3.7799y(i-1) - 5.4485y(i-2) - 3.5472y(i-3) - 0.8808y(i-4) \\ & + 0.0019x(i) - 0.0038x(i-2) + 0.0019x(i-4) \end{aligned} \quad (4-96)$$

3) 在 Matlab 环境中对移植算式进行检验

在 Matlab 环境中制作正弦波的仿真信号，基于扫频法获取移植算法的幅频特性检验是否符合设计任务书的要求，在 Matlab 环境中对编程算式进一步进行幅频特性检验合格后，就可将式(4-96)写入需使用该低通滤波器的场合。

通过以下程序可在 Matlab 环境中画出数字滤波器的幅频特性图，如图 4-43 所示。

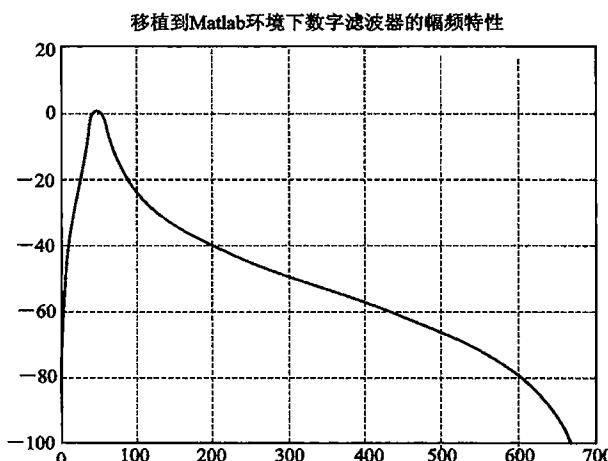


图 4-43 Matlab 环境下二阶带通数字滤波器幅频特性图

在把 LabVIEW 上获得的反向系数 a 写入到 Matlab 程序时，第一个系数是 1，然后再依次从 LabVIEW 上获得反向系数，正向系数不需加 1，依次为从 Labview 上获得的正向系数；设置的采样频率和采样点数需和 LabVIEW 平台上设置的一致：

```

a=[1 -3.7799 5.4485 -3.5472 0.8808]; %反向系数
b=[0.0019 0 -0.0038 0 0.0019]; %正向系数
[h w]=freqz(b,a,1024,1400); %作幅频特性，采样点数为 1024,
                               %采样频率为 1400 Hz
plot(w,20*log10(abs(h)));grid on;
title('移植到 Matlab 环境下数字滤波器的幅频特性')

```

$f = 40$ Hz 时， $20 * \log_{10}(\text{abs}(h))$ 为 -3.0238 dB； $f = 50$ Hz 时， $20 * \log_{10}(\text{abs}(h))$ 为 -0.4538 dB； $f = 60$ Hz 时， $20 * \log_{10}(\text{abs}(h))$ 为 -3.0248 dB； $f = 600$ Hz 时，

$20 * \log_{10}(\text{abs}(h))$ 为 -79.3513 dB 。与图 4-42 所示的虚拟巴特沃兹滤波器获得的二阶带通数字滤波器幅频特性对比，说明移植后的滤波器达到了最初设计要求。

4.6 增强自我管理与自适应能力及智能化控制功能

控制功能是体现测控系统具有判断、决策、自我管理与自适应能力的一种智能化功能。一个自动化控制系统由传感器、控制器及执行器三大环节组成。在如图 1-1 所示的现代自动化控制系统中，由安装了控制软件模块的计算机担任控制器，代替了由比率放大、积分、微分器组成传统硬件 PID 控制器，使得参数的改变更灵活便捷，调节效率更快更高。

现场总线控制系统(FCS)中的智能设备都具有微处理器和数字接口，并挂接在现场总线上。但没有仅安装控制模块的微处理器作为独立的控制器智能设备。起调节控制作用的软件控制模块可以安装在智能传感器/变送器的微处理器中，也可以安装在智能执行器的微处理器中。由上位机经现场总线下达控制参量的设定值，智能传感器/变送器及执行器之间的检测信息与控制指令也经现场总线传递。

随着自动化技术的发展，闭环自动控制模块的算法极其发达，多种多样，常用的有 PID 算法、模糊算法(详见第 12 章)等。本章简介 PID 控制模式的软件编程算法。

广义地说，具有控制功能的控制器也是一种滤波器，要实现控制器的控制功能就是要求取与模拟 PID 控制器等效的数字 PID 控制器的脉冲传递函数。

4.6.1 模拟 PID 控制器的传递函数

1. 模拟 PID 控制器的控制作用

由三种控制作用的适当配合来控制一个变量：

- 比例控制(控制作用与偏差动作信号 $e(t)$ 成比例)；
- 积分控制(控制作用与偏差动作信号 $e(t)$ 的积分成比例)；
- 微分控制(控制作用与偏差动作信号 $e(t)$ 的微分成比例)。

2. 模拟 PID 控制作用的数学表达式

模拟 PID 控制系统框图如图 4-44(a)所示。

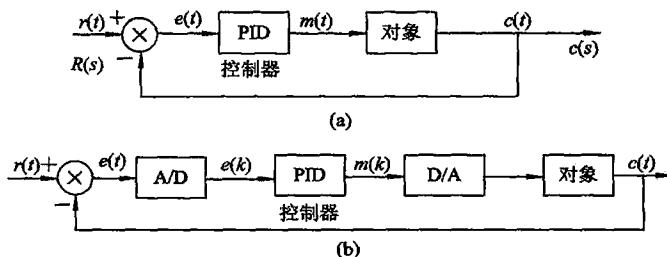


图 4-44 闭环控制系统中的 PID 控制器

(a) 模拟式；(b) 数字式

模拟 PID 控制作用的数学表达式如下：

$$m(t) = K \left[e(t) + \frac{1}{T_i} \int_0^t e(t) dt + T_d \frac{de(t)}{dt} \right] \quad (4-97)$$

式中： $e(t)$ 为控制器的输入信号（偏差动作信号）； $m(t)$ 为控制器输出信号（控制信号）； K 为比例增益； T_i 为积分时间； T_d 为微分时间。

3. 模拟控制器(PID)的传递函数

对式(4-97)微分方程两边求拉氏变换可得模拟控制器(PID)的传递函数是：

$$G(s) = \frac{M(s)}{E(s)} = K \left[1 + \frac{1}{T_i s} + T_d s \right]$$

4.6.2 数字 PID 控制器脉冲传递函数

数字 PID 控制系统框图如图 4-44(b) 所示。

将式(4-97)离散化，利用梯形求和近似积分项(参见图 4-45)：

$$\begin{aligned} m(kT) &= K \left\{ e(kT) + \frac{T}{T_i} \left[\frac{e(0) + e(T)}{2} + \frac{e(T) + e(2T)}{2} + \dots + \frac{e[(k-1)T] + e(kT)}{2} \right] \right. \\ &\quad \left. + T_d \frac{e(kT) - e[(k-1)T]}{T} \right\} \end{aligned}$$

则有

$$m(kT) = K \left\{ e(kT) + \frac{T}{T_i} \sum_{h=1}^k \frac{e[(h-1)T] + e(hT)}{2} + \frac{T_d}{T} [e(kT) - e(k-1)T] \right\} \quad (4-98)$$

定义 $\frac{e[(h-1)T] + e(hT)}{2} = f(hT), \quad f(0) = 0 \quad (4-99)$

则 $\sum_{h=1}^k \frac{e[(h-1)T] + e(hT)}{2} = \sum_{h=1}^k f(hT) = y(kT) \quad (4-100)$

故有 $k = 1, y(T) = f(T)$

$k = 2, y(2T) = f(T) + f(2T)$

$k = 3, y(3T) = f(T) + f(2T) + f(3T)$

⋮

$k = k, y(kT) = f(T) + f(2T) + \dots + f(kT)$

于是 $y(kT) - y[(k-1)T] = f(kT)$

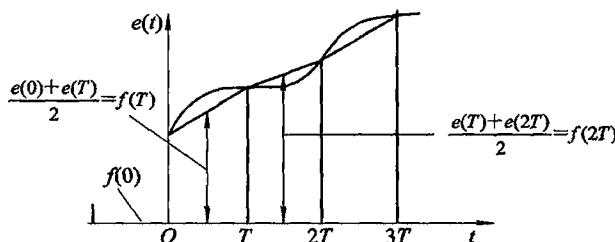


图 4-45 利用梯形求和近似积分项

对上式求 Z 变换：

$$Y(z) - z^{-1}Y(z) = F(z)$$

可得

$$Y(z) = \frac{1}{1-z^{-1}}F(z)$$

将式(4-100)取 Z 变换，有

$$Z\left[\sum_{h=1}^k \frac{e(h-1)T + e(hT)}{2}\right] = Z\left[\sum_{h=1}^k f(hT)\right] = Z[y(kT)] = Y(z) = \frac{F(z)}{1-z^{-1}}$$

将式(4-99)取 Z 变换有

$$\frac{1}{2}[E(z) + z^{-1}E(z)] = F(z) = \frac{1+z^{-1}}{2}E(z)$$

所以

$$Z\left[\sum_{h=1}^k \frac{e(h-1)T + e(hT)}{2}\right] = \frac{1+z^{-1}}{2(1-z^{-1})}E(z)$$

对式(4-98)取 Z 变换，得

$$\begin{aligned} M(z) &= K\left[E(z) + \frac{T}{2T_i} \frac{1+z^{-1}}{1-z^{-1}}E(z) + \frac{T_d}{T}(1-z^{-1})E(z)\right] \\ &= K\left[1 - \frac{T}{2T_i} + \frac{T}{T_i} \frac{1}{1-z^{-1}} + \frac{T_d}{T}(1-z^{-1})\right]E(z) \\ &= \left[K_p + K_i \frac{1}{1-z^{-1}} + K_d(1-z^{-1})\right]E(z) \end{aligned}$$

式中 $K_p = K\left(1 - \frac{T}{2T_i}\right) = K - \frac{K_i}{2}$ 为比例增益； $K_i = K \frac{T}{T_i}$ 为积分增益； $K_d = K \frac{T_d}{T}$ 为微分增益。

于是可得数字 PID 控制器的脉冲传递函数为

$$G_D(z) = \frac{M(z)}{E(z)} = K_p + K_i \frac{1}{1-z^{-1}} + K_d(1-z^{-1}) \quad (4-101)$$

上式称为位置式 PID 控制器的脉冲传递函数。

对于实现数字 PID 控制器的计算机编程算式，由式(4-101)可得

$$(1-z^{-1})M(z) = [K_p(1-z^{-1}) + K_i + K_d(1-2z^{-1}+z^{-2})]E(z)$$

$$M(z) - z^{-1}M(z) = (K_p + K_i + K_d)E(z) - (K_p + 2K_d)z^{-1}E(z) + K_dz^{-2}E(z)$$

求上式的 Z 反变换得差分方程：

$$m(k) - m(k-1) = (K_p + K_i + K_d)e(k) - (K_p + 2K_d)e(k-1) + K_d e(k-2)$$

故

$$m(k) = (K_p + K_i + K_d)e(k) + m(k-1) - (K_p + 2K_d)e(k-1) + K_d e(k-2) \quad (4-102)$$

式(4-102)表明，当前 k 时刻控制器的输出 $m(k)$ 由以下三部分来决定：

- 当前时刻的输入 $(K_p + K_i + K_d)e(k)$ ；
- 当前时刻之前一时刻的输出 $m(k-1)$ 与输入 $(K_p + 2K_d)e(k-1)$ 之差；
- 当前时刻之前一时刻的前一时刻的输入 $K_d e(k-2)$ 。

式(4-102)就是实现数字 PID 控制器的计算机编程算式。

参 考 文 献

- [1] 肖丹, 张勇, 刘君华. 基于 LabVIEW 的相关信号提取仪的设计[C]. 第二届全国虚拟仪器学术交流大会, 2007
- [2] 刘君华, 申忠如, 郭福田. 现代测试技术与系统集成[M]. 北京: 电子工业出版社, 2005
- [3] 钱凌影, 刘君华. 智能化模块——数字滤波器网络虚拟样机的设计[C]. 第九届全国敏感元件与传感器学术会议论文集. 北京: 国防工业出版社, 2005: 650 - 652
- [4] 雷大宇, 刘君华, 周金林, 等. 基于 Labwindows/CVI 的电涡流定位测距与相关检测系统[J]. 仪器仪表学报, 2003 年第 4 期(增刊)
- [5] 梁晋涛, 阎晓燕, 刘君华. 伪随机虚拟相关辨识仪的研究[J]. 测试技术学报, 2002, 16(专刊)
- [6] 张永怀, 杨帮华, 刘君华. 虚拟数字滤波提高红外气体分析系统精度的应用研究[J]. 化工自动化及仪表, 2002, 29(6): 61 - 63
- [7] 丁晖, 申忠如, 刘君华. 基于小波和相关分析的虚拟介损在线监测仪[J]. 电压技术, 2000, 12
- [8] Liu Junhua, et al. A New Gain Calibration Method for Measuring Instrument with Multiple - Stage Gains. Journal of Electronic[J]. Measurement and Instrument, 1997, 9(9):217 - 20
- [9] Bowen M and Smith G. Considerations for the Design of Smart Sensors[J]. Sensors and Actuators(A), 1995, 47(19): 516 - 520
- [10] [日]绪方胜彦. 离散时间控制系统[M]. 刘君华, 等译. 西安: 西安交通大学出版社, 1990
- [11] 周舒梅. 动态信号分析和仪器[M]. 北京: 机械工业出版社, 1990
- [12] Wilmshurst TH. Signal Recovery from Noise in Electronic Instrumentation[C]. 2nd ed. Bristol: IOP Publishing Ltd Techno House, 1990: 29 - 37
- [13] 浦申臣, 顾炎君. 780B 压力测量系统[J]. 测控技术, 1989(1): 46 - 47
- [14] 刘君华. 采用压力传感器同时测量压力与温度[J]. 微电子学与计算机, 1990(2): 8 - 11
- [15] 刘君华, 侯军. 计算机频率补偿法在动态测量中的应用[J]. 电子测量技术, 1986. 12: 1 - 7

第5章 多元回归分析法及其在智能传感器系统中的应用

本章内容

- 多元回归分析法与定常系数多元回归方程
- 回归分析法与可变系数回归方程
- [示例 5-1] 基于回归分析模型法消除一个干扰量影响的智能化软件模块设计
- [示例 5-2] 基于回归分析模型法降低两个干扰量影响的智能化软件模块设计

表征正、逆模型的回归方程式，有一元、二元或多元回归方程。其多项式的系数有定常系数和可变系数。采用一元多项式回归方程来表征正、逆模型，虽然可以很好地拟合传感器的输入输出非线性特性，但却是很不完备的。因为实际的传感器系统存在干扰量，即应是一个多输入单输出系统，故采用二元或多元回归方程来表征正、逆模型的多输入单输出特性才是较为完备的。本章结合改善传感器稳定性消除交叉敏感影响的(参数)模型法，介绍分别采用二元或多元定常回归方程、变系数回归方程分析法来建立消除干扰量影响的逆模型。

5.1 多元回归分析法与定常系数多元回归方程

多元回归分析模型法建立逆模型的核心思想是：欲消除 $n \geq 1$ 个干扰量对主传感器测量目标参量的影响，就要设置 $n \geq 1$ 个监测干扰量的辅助传感器，进而建立更完备的逆模型的 $m = n + 1$ 元常系数高阶回归方程。阶数由满足允许的误差来决定。

消除一个干扰量时 $n=1$ ，需要建立二传感器智能传感器系统，逆模型为二元回归方程，进行二传感器数据融合；消除 $n=2$ 个干扰量时，需建立三传感器智能传感器系统，逆模型为三元回归方程，进行三传感器数据融合。

5.1.1 二传感器数据融合——二元回归分析法

二传感器可测量两个参量，得到两个参量的信息。两个信息的融合算法可以有多种，曲面拟合算法是其中之一，也就是二元回归分析法。下面均以压阻式压力传感器为例来说明数据融合算法。

1. 二元回归分析法基本原理

已知压力传感器输出是电压 U_P ，并且存在对温度的交叉灵敏度。因此如果按照传统的方法，只对压力传感器进行一维标定实验，获得输入(压力 P)输出(电压 U_P)特性曲线，并由此来求取被测压力值会有较大的误差。因为被测量 P 不是输出值 U_P 的一元函数。现在

由另一温度传感器输出电压 U_T , 代表温度信息 T , 则压力参量 P 可以用 U_P 及 U_T 的二元函数表示才较完备, 即

$$P = f(U_P, U_T)$$

由二维坐标(U_P , U_T)决定的 P_i 在一平面上, 可以利用二次曲面拟合方程, 即二元回归方程描述:

$$P = \alpha_0 + \alpha_1 U_P + \alpha_2 U_T + \alpha_3 U_P^2 + \alpha_4 U_P U_T + \alpha_5 U_T^2 + \varepsilon_1 \quad (5-1)$$

同理有

$$T = \beta_0 + \beta_1 U_P + \beta_2 U_T + \beta_3 U_P^2 + \beta_4 U_P U_T + \beta_5 U_T^2 + \varepsilon_2 \quad (5-2)$$

式中, $\alpha_0 \sim \alpha_5$ 、 $\beta_0 \sim \beta_5$ 为常系数; ε_1 、 ε_2 为高阶无穷小。方程式项数 t 依允许的误差 ε 来取, 式中 $t=5$ 。

如果式(5-1)、式(5-2)中各常系数已知, 那么式(5-1)与式(5-2)就是消除交叉敏感, 求取被测量 P 或 T 更完备的逆模型。为此, 首先要进行二维标定实验, 由最小二乘法原理确定均方误差最小条件下的常系数。

2. 实验标定

在压力传感器的量程范围内确定 n 个压力标定点, 在工作温度范围内确定 m 个温度标定点, 于是由压力 P 与温度 T 标准值发生器产生在各个标定点的标准输入值为

$$P_k: P_1, P_2, P_3, \dots, P_n$$

$$T_j: T_1, T_2, T_3, \dots, T_m$$

对应于上述各个标定点的标准输入值读取相应的输出值 U_{Pk} 和 U_{Tj} , 这样, 我们在 m 个不同温度状态对压力传感器进行静态标定, 共计有 $s=m \times n$ 个标定点, 每个标定点同时对应有 4 个标定值, 也称一个样本对: 两个传感器的输入量, 即主测量压力 P 与辅参量温度 T , 以及二者相应的输出量, 分别为 U_P 和 U_T 。共获得了对应 m 个不同温度状态的 m 条输入输出特性, 即 $P-U_P$ 特性簇, 如图 5-1(a) 所示。同时我们也获得对应于不同压力状态的温度传感器的 n 条输入输出特性($T-U_T$), 即 $T-U_T$ 特性簇, 如图 5-1(b) 所示。

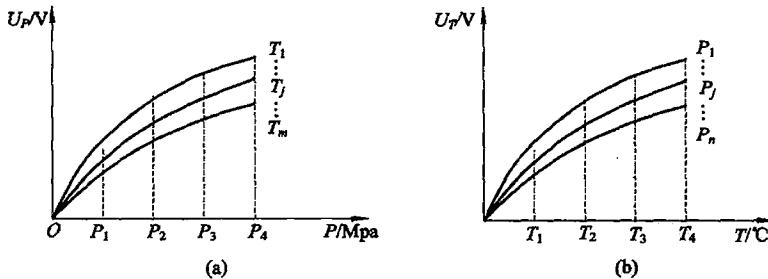


图 5-1 压力传感器输入输出特性
(a) $P-U_P$ 特性簇; (b) $T-U_T$ 特性簇

3. 二元回归方程待定常数的确定

为确定式(5-1)所表征的二元回归方程式的常系数, 通常根据最小二乘法原理, 求得的系数值满足均方误差最小条件。系数 $\alpha_0 \sim \alpha_5$ 和 $\beta_0 \sim \beta_5$ 的求法相同。下面以 $\alpha_0 \sim \alpha_5$ 为例说明求取步骤。

(1) 第 i 个标定点的压力数据计算值(以下简称计算值) $P(U_{Pi}, U_{Ti})$, 根据式(5-1)为

$$P(U_{Pi}, U_{Ti}) = \alpha_0 + \alpha_1 U_{Pi} + \alpha_2 U_{Ti} + \alpha_3 U_{Pi}^2 + \alpha_4 U_{Pi} U_{Ti} + \alpha_5 U_{Ti}^2 + \epsilon_i$$

(2) 第 i 个标定点的压力标定值 P_i 与计算值 $P(U_{Pi}, U_{Ti})$ 之间存在误差 γ_i , 其方差为 γ_i^2 :

$$\text{差值: } \gamma_i = P(U_{Pi}, U_{Ti}) - P_i$$

$$\text{方差: } \gamma_i^2 = [P(U_{Pi}, U_{Ti}) - P_i]^2$$

$$\begin{aligned} \gamma_i^2 &= [(\alpha_0 + \alpha_1 U_{Pi} + \alpha_2 U_{Ti} + \alpha_3 U_{Pi}^2 + \alpha_4 U_{Pi} U_{Ti} + \alpha_5 U_{Ti}^2 + \epsilon_i) - P_i]^2 \\ &= \sum_{k=0}^t [\alpha_k h_{ik} - P_i]^2 \end{aligned} \quad (5-3)$$

式中 $t=5$ 为回归方程式的项数, $h_{i0} = 1$, $h_{i1} = U_{Pi}$, $h_{i2} = U_{Ti}$, $h_{i3} = U_{Pi}^2$, $h_{i4} = U_{Pi} \cdot U_{Ti}$, $h_{i5} = U_{Ti}^2$ 。

(3) 全部标定点压力标定值与计算值之差的平方和 I_s 为

$$I_s = \sum_{i=1}^{s=m \times n} \gamma_i^2 = \sum_{i=1}^{s=m \times n} \sum_{k=0}^t [\alpha_k h_{ik} - P_i]^2 = I(\alpha_0, \dots, \alpha_k, \dots, \alpha_{t=5}) \quad (5-4)$$

式中, $s=m \times n$ 为标定点总数, 当压力标定点数 $m=6$, 温度标定点数 $n=6$ 时, $s=36$ 。

由式(5-4)可见, I_s 是常系数 $\alpha_0, \alpha_1, \dots, \alpha_5$ 的多元函数。

(4) 回归方程待定系数 $\alpha_0 \sim \alpha_5$ 的最小二乘最优解。根据多元函数求极值条件求 $\alpha_0 \sim \alpha_5$ 的最优解, 首先令下列各偏导数为零, 即 $\frac{\partial I_s}{\partial \alpha_k} = 0, k=0, 1, \dots, t=5$; 展开可得以下六个方程:

$$\left. \begin{aligned} \frac{\partial I_s}{\partial \alpha_0} &= 2 \sum_{i=1}^s \left[\sum_{k=0}^5 \alpha_k h_{ik} - P_i \right] \times h_{i0} = 0, h_{i0} = 1 \\ \frac{\partial I_s}{\partial \alpha_1} &= 2 \sum_{i=1}^s \left[\sum_{k=0}^5 \alpha_k h_{ik} - P_i \right] \times h_{i1} = 0, h_{i1} = U_{Pi} \\ \frac{\partial I_s}{\partial \alpha_2} &= 2 \sum_{i=1}^s \left[\sum_{k=0}^5 \alpha_k h_{ik} - P_i \right] \times h_{i2} = 0, h_{i2} = U_{Ti} \\ \frac{\partial I_s}{\partial \alpha_3} &= 2 \sum_{i=1}^s \left[\sum_{k=0}^5 \alpha_k h_{ik} - P_i \right] \times h_{i3} = 0, h_{i3} = U_{Pi}^2 \\ \frac{\partial I_s}{\partial \alpha_4} &= 2 \sum_{i=1}^s \left[\sum_{k=0}^5 \alpha_k h_{ik} - P_i \right] \times h_{i4} = 0, h_{i4} = U_{Pi} U_{Ti} \\ \frac{\partial I_s}{\partial \alpha_5} &= 2 \sum_{i=1}^s \left[\sum_{k=0}^5 \alpha_k h_{ik} - P_i \right] \times h_{i5} = 0, h_{i5} = U_{Ti}^2 \end{aligned} \right\} \quad (5-5a)$$

由方程(5-5a)可得

$$\sum_{i=1}^s \left[\sum_{k=0}^5 \alpha_k h_{ik} \right] \times h_{ik} = \sum_{i=1}^s P_i \times h_{ik} \quad (5-5b)$$

式中, $i=1, 2, \dots, s=m \times n=36$; $k=0, 1, 2, \dots, t=5$ 。

由线性代数知识, 可将式(5-5b)写成矩阵形式:

$$\boldsymbol{\alpha} \cdot \mathbf{H} \times \mathbf{H}^T = \mathbf{P} \cdot \mathbf{H}^T \quad (5-6)$$

式中, $\boldsymbol{\alpha} \cdot \mathbf{H} = \sum_{k=0}^t \alpha_k \times h_{ik}$, $\mathbf{P} \times \mathbf{H}^T = \sum_{i=1}^s P_i \times h_{ik}$, $s=m \times n=6 \times 6=36$, $t=5$ 。

于是, 回归方程待定常系数 $\alpha_0, \alpha_1, \dots, \alpha_5$ 的最小二乘最优解的求解式为

$$\boldsymbol{\alpha} = \mathbf{P} \cdot \mathbf{H}^T \cdot (\mathbf{H} \times \mathbf{H}^T)^{-1} \quad (5-7)$$

这就是求解最小二乘问题的方程组法。

(5) 最优解式(5-7)算法的实现。Matlab 软件具有强大的数值计算和矩阵计算功能, 可以很轻松地求解式(5-7)的矩阵计算, 参见[示例 5-1]和[示例 5-2]。

5.1.2 三传感器数据融合——三元回归分析法

1. 单一功能(只测一个目标参量)的三传感器数据融合

通过监测两个干扰量, 即两个非目标参数, 可以消除这两个干扰量的影响, 提高该单功能传感器对被测目标参量的测量精度。监测干扰量的传感器, 可以选用能够测量这两个干扰参量的任何形式的传感器, 只需把它们放置在同一测量场中, 与测量目标参量的传感器经受同样强度干扰量的影响。

仍以压阻式压力传感器为例, 其输出不仅受到工作环境温度 T 的影响, 而且还受到电源供电电流 I 的影响。为了消除这两个参量的影响, 需要对 T 和 I 分别进行监测, 建立如图 5-2 所示的三传感器数据融合智能传感器系统, 进行三维标定实验, 确立三元回归方程:

$$P = \alpha_0 + \alpha_1 U_P + \alpha_2 U_T + \alpha_3 U_I + \alpha_4 U_P^2 + \alpha_5 U_T^2 + \alpha_6 U_I^2 + \alpha_7 U_P U_T + \alpha_8 U_P U_I + \alpha_9 U_T U_I + \epsilon_p \quad (5-8)$$

式中, P 为规定的被测参量压力; U_P 为压阻式压力传感器输入压力为 P 时的输出电压值; U_T 为监测工作环境温度用温度传感器的输出; U_I 为监测供电电流用传感器的输出; ϵ_p 为可忽略的高阶(大于二阶)无穷小量。

根据三维标定实验, 按照均方误差最小原则确定式(5-8)中的常系数 $\alpha_0 \sim \alpha_9$, 从而式(5-8)可以用来建立如图 5-2 所示的三传感器数据融合智能传感器系统, 以抑制对两个干扰量的交叉敏感, 提高原传感器系统对温度、电源波动的稳定性。

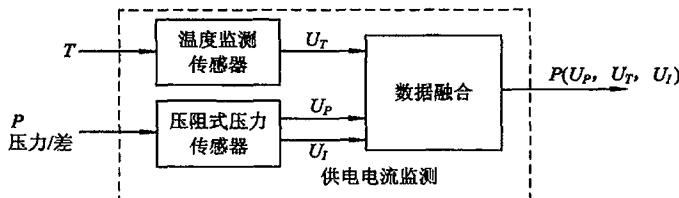


图 5-2 三传感器数据融合智能传感器系统框图

2. 三功能(测量三个参量)的传感器数据融合——美国霍尼韦尔公司的 ST-3000 型智能变送器

美国霍尼韦尔(Honeywell)公司于 1983 年率先推出的 ST-3000 型变送器, 是世界上第一台智能化的压力变送器。其敏感元件在同一硅片上($130 \times 175 \text{ mil}^2$ (1 mil (密尔)=0.001 inch(英寸)), 采用离子注入等 IC 技术, 配置压差、静压和温度三种传感器, 从而有效地解决了静压、压差及温度之间交叉敏感对测量的影响问题, 使之具有高精度、高稳定性等特点。ST-3000 变送器的内部除传感器的调理电路外, 还带有微处理器、存储器以及 I/O 接口

等，具有双向通信能力和完善的自诊断功能。变送器的输出有两种形式：一为标准的4~20 mA的模拟信号输出；二为数字信号输出。

1) ST - 3000 智能变送器结构概况

ST - 3000 智能变送器是能实现测量压力(差) P 、静压 S_P 、温度 T 三个参量的三功能传感器，它的传感芯片电路如图 5 - 3 所示。

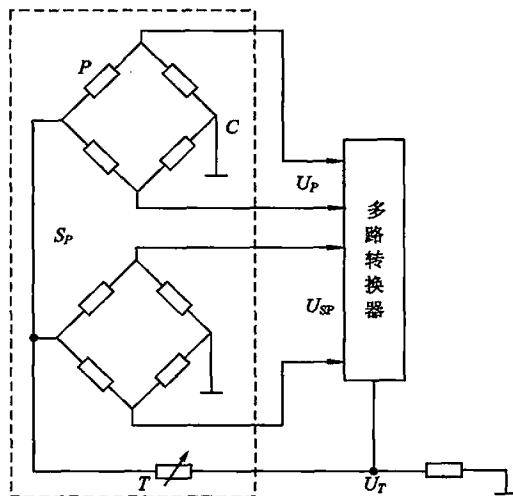


图 5 - 3 传感芯片的电路图

2) 数据融合

三个传感器相互之间存在交叉灵敏度，每个传感器进行刻度转换的逆模型都应是三元回归方程，即：

$$P = f(U_P, U_{SP}, U_T), \quad S_P = h(U_P, U_{SP}, U_T), \quad T = g(U_P, U_{SP}, U_T)$$

上述三个方程都如式(5 - 8)所示，共有 3×10 个未知待定常数，需要由三维标定实验数据来确定。

为简化处理起见，我们首先进行降维处理。由于对静压的测量精度要求不高，可将它作为一元函数来对待：

$$S_P = h(U_{SP})$$

又因静压主要影响压力(差)的零点输出，产生的干扰量用 U'_0 表示。 U'_0 与静压输出 U_{SP} 的关系由 n 阶多项式方程描述：

$$U'_0 = \gamma_0 + \gamma_1 U_{SP} + \gamma_2 U_{SP}^2 + \gamma_3 U_{SP}^3 + \cdots + \gamma_n U_{SP}^n \quad (5 - 9)$$

式中， $\gamma_0, \gamma_1, \gamma_2, \dots, \gamma_n$ 为待定常系数，通过标定实验来确定。选定 n 个不同静压值 S_{Pi} ($i=1, 2, 3, \dots, n$)，测定相应压力(差)的零点 U'_{0j} ($j=1, 2, 3, \dots, n$)，即

$$S_{Pi}: S_{P1}, S_{P2}, S_{P3}, \dots, S_{Pn}$$

$$U'_{0j}: U'_{01}, U'_{02}, U'_{03}, \dots, U'_{0n}$$

根据最小二乘法原理和利用标定值求解矩阵方程，可求得常系数 $\gamma_0 \sim \gamma_n$ ，从而方程式(5 - 9)得以确定，常系数 $\gamma_0 \sim \gamma_n$ 存入内存。测量时，对与三个输入量 P, S_P, T 相应的三

个输出量 U_P 、 U_{SP} 、 U_T 进行采样。首先根据采样值 U_{SP} 代入式(5-9)计算 U' ，然后再与采样值 U_{SP} 做减法，得

$$U = U_P - U' \quad (5-10)$$

式中， U 是消除了零点干扰量后的压力(差)输出值。于是被测压力(差)值就降为二元函数

$$P = f(U, U_T) \quad (5-11)$$

因为静压 U_{SP} 对温度输出 U_T 基本上没有影响，故被测温度由二元函数表示已足够：

$$T = h(U_P, U_T) \quad (5-12)$$

降元后的 P 与 T 就可以采用二传感器数据融合技术来处理了，即可以采用式(5-1)、式(5-2)来进行数据融合处理。

5.2 回归分析法与可变系数回归方程

本节仍以压阻式压力传感器为例，说明通过建立可变系数回归方程作为消除一个干扰量——温度的逆模型进行数据融合的方法。

5.2.1 工作原理

我们已知经典传感器的输入输出特性是由式(2-1)所给出的一元多项式回归方程，对输出被测量进行刻度转换用的模型是传感器输入输出特性($P-U$)的反非线性特性($U-P$)，逆模型也是一个一元多项式回归方程：

$$P = A_0(T) + A_1(T)U + A_2(T)U^2 + \cdots + A_5(T)U^5 + \epsilon_P \quad (5-13)$$

式中： P 、 U 是压力传感器的输入压力与相应输出电压； ϵ_P 为高阶无穷小量； $A_0(T)$ ， $A_1(T)$ ， $A_2(T)$ ，…， $A_5(T)$ 为多项式的系数，它们都随温度 T 而变化。

对不同的工作温度 T_i ，压力传感器有不同的输入输出($P-U$)特性，对应也有其不同的反非线性特性($U-P$)逆模型，其表征为式(5-13)有不同的系数 $A_0(T_i)$ ， $A_1(T_i)$ ， $A_2(T_i)$ ，…， $A_5(T_i)$ 。

可变系数回归分析法消除温度干扰的基本思路是：找出系数 $A_0(T) \sim A_5(T)$ 随温度 T 变化的规律性；设置温度传感器监测压力传感器的工作温度；当测出当前工作温度 T_i 时，即可确定出当前温度 T_i 时的系数 $A_0(T_i)$ ， $A_1(T_i)$ ， $A_2(T_i)$ ，…， $A_5(T_i)$ ，也即压力传感器当前温度 T_i 时输入输出特性所对应的逆模型即可确定。按照 T_i 时的逆模型进行刻度转换计算得到被测目标参量 P ，从而可以避免引入附加温度误差。

5.2.2 回归方程可变系数 $A_0(T) \sim A_5(T)$ 的确定

回归方程可变系数 $A_0(T) \sim A_5(T)$ 的确定主要分两个阶段：一是前期准备，二是分别建立各系数 $A_0(T) \sim A_5(T)$ 随温度 T 变化的关系式。

1. 前期准备

1) 实验标定

标定方法见 5.2.1 节，实验标定数据如表 5-1 所示。

表 5-1 不同温度 T 时压力传感器的输入 P 与对应输出电压 U 的实验标定值

输出 U /mV	输入 P	输入压力 $P \times 10^4$ Pa					
		P_1	P_2	P_3	P_4	P_5	P_6
温度 T /°C	T_1	$U(P_1, T_1)$	$U(P_2, T_1)$	$U(P_3, T_1)$	$U(P_4, T_1)$	$U(P_5, T_1)$	$U(P_6, T_1)$
	T_2	$U(P_1, T_2)$	$U(P_2, T_2)$	$U(P_3, T_2)$	$U(P_4, T_2)$	$U(P_5, T_2)$	$U(P_6, T_2)$
	T_3	$U(P_1, T_3)$	$U(P_2, T_3)$	$U(P_3, T_3)$	$U(P_4, T_3)$	$U(P_5, T_3)$	$U(P_6, T_3)$
	T_4	$U(P_1, T_4)$	$U(P_2, T_4)$	$U(P_3, T_4)$	$U(P_4, T_4)$	$U(P_5, T_4)$	$U(P_6, T_4)$
	T_5	$U(P_1, T_5)$	$U(P_2, T_5)$	$U(P_3, T_5)$	$U(P_4, T_5)$	$U(P_5, T_5)$	$U(P_6, T_5)$
	T_6	$U(P_1, T_6)$	$U(P_2, T_6)$	$U(P_3, T_6)$	$U(P_4, T_6)$	$U(P_5, T_6)$	$U(P_6, T_6)$

2) 建立各个标定温度时的逆模型

逆模型用五阶六项式一元回归方程来逼近:

$$\left. \begin{aligned} T = T_1, P = A_0(T_1) + A_1(T_1)U + A_2(T_1)U^2 + A_3(T_1)U^3 + A_4(T_1)U^4 + A_5(T_1)U^5 \\ T = T_2, P = A_0(T_2) + A_1(T_2)U + A_2(T_2)U^2 + A_3(T_2)U^3 + A_4(T_2)U^4 + A_5(T_2)U^5 \\ T = T_3, P = A_0(T_3) + A_1(T_3)U + A_2(T_3)U^2 + A_3(T_3)U^3 + A_4(T_3)U^4 + A_5(T_3)U^5 \\ \vdots \\ T = T_6, P = A_0(T_6) + A_1(T_6)U + A_2(T_6)U^2 + A_3(T_6)U^3 + A_4(T_6)U^4 + A_5(T_6)U^5 \end{aligned} \right\} \quad (5-14)$$

式中共有 6×6 个系数, 其中 $A_0(T_i)$ 为标定温度 T_i 时的零位值 ($i=1, 2, 3, \dots, 6$); $A_1(T_i)$ 为标定温度 T_i 时的灵敏度 ($i=1, 2, 3, \dots, 6$); $A_2(T_i) \sim A_5(T_i)$ 分别是标定温度 T_i 时各高阶非线性灵敏度系数 ($i=1, 2, 3, \dots, 6$)。

如果上述 6×6 个系数 $A_0(T_i), A_1(T_i), A_2(T_i), \dots, A_5(T_i)$ ($i=1, 2, 3, \dots, 6$) 已知, 那么方程组 (5-14) 即可确定, 各标定温度的逆模型也就确立了。

(1) 求 T_1 时的系数 $A_0(T_1), A_1(T_1), \dots, A_5(T_1)$ 。

求 6 个未知数需要建立 6 个方程。利用 T_1 时的 6 对标定值 P_i 及 $U(P_i, T_1)$ 可建立 6 个方程式, 如下:

$$\left. \begin{aligned} P_1 = A_0(T_1) + A_1(T_1)U(P_1, T_1) + A_2(T_1)U^2(P_1, T_1) + A_3(T_1)U^3(P_1, T_1) \\ + A_4(T_1)U^4(P_1, T_1) + A_5(T_1)U^5(P_1, T_1) \\ P_2 = A_0(T_1) + A_1(T_1)U(P_2, T_1) + A_2(T_1)U^2(P_2, T_1) + A_3(T_1)U^3(P_2, T_1) \\ + A_4(T_1)U^4(P_2, T_1) + A_5(T_1)U^5(P_2, T_1) \\ P_3 = A_0(T_1) + A_1(T_1)U(P_3, T_1) + A_2(T_1)U^2(P_3, T_1) + A_3(T_1)U^3(P_3, T_1) \\ + A_4(T_1)U^4(P_3, T_1) + A_5(T_1)U^5(P_3, T_1) \\ \vdots \\ P_6 = A_0(T_1) + A_1(T_1)U(P_6, T_1) + A_2(T_1)U^2(P_6, T_1) + A_3(T_1)U^3(P_6, T_1) \\ + A_4(T_1)U^4(P_6, T_1) + A_5(T_1)U^5(P_6, T_1) \end{aligned} \right\} \quad (5-15)$$

求解矩阵方程组 (5-15), 可得 T_1 时的 6 个系数 $A_0(T_1), A_1(T_1), \dots, A_5(T_1)$ 。

(2) 其它温度点 T_i ($i=2, 3, \dots, 6$) 的系数 $A_0(T_i), A_1(T_i), \dots, A_5(T_i)$ 的确定。

利用 $T=T_2$ 时的 6 对标定值 P_i 及 $U(P_i, T_2)$ 建立 6 个方程, 可解得 $A_0(T_2) \sim A_5(T_2)$; 同理, 可确定其它各系数。

2. 建立各系数 $A_0(T), A_1(T), \dots, A_5(T)$ 与温度的关系式

各系数随温度变化一般而言也是非线性的, 其非线性程度取决于实际的传感器, 我们仍用五阶六项多项式来逼近。

1) 系数 $A_0(T)$ 与温度 T 关系式的确立

$A_0(T)-T$ 关系式用五阶六项多项式逼近如下:

$$A_0(T) = a_0 + a_1 T + a_2 T^2 + a_3 T^3 + a_4 T^4 + a_5 T^5 \quad (5-16)$$

其中 a_0, a_1, \dots, a_5 为待定常数, 若已知, 则式(5-16)确定。

利用 6 个温度值 T_1, T_2, \dots, T_6 与已知的 $A_0(T_1), A_0(T_2), \dots, A_0(T_6)$ 建立 6 个方程式, 可求解常系数 $a_0 \sim a_5$, 方程式如下:

$$\left. \begin{aligned} A_0(T_1) &= a_0 + a_1 T_1 + a_2 T_1^2 + a_3 T_1^3 + a_4 T_1^4 + a_5 T_1^5 \\ A_0(T_2) &= a_0 + a_1 T_2 + a_2 T_2^2 + a_3 T_2^3 + a_4 T_2^4 + a_5 T_2^5 \\ &\vdots \\ A_0(T_6) &= a_0 + a_1 T_6 + a_2 T_6^2 + a_3 T_6^3 + a_4 T_6^4 + a_5 T_6^5 \end{aligned} \right\} \quad (5-17)$$

求解方程组(5-17), 则未知待定常数 $a_0 \sim a_5$ 得以确定。

2) 系数 $A_1(T)$ 与温度 T 关系式的确立

$A_1(T)-T$ 关系式仍由五阶六项多项式表示如下:

$$A_1(T) = b_0 + b_1 T + b_2 T^2 + b_3 T^3 + b_4 T^4 + b_5 T^5 \quad (5-18)$$

利用 6 个温度值 T_1, T_2, \dots, T_6 与已知的 $A_1(T_1), A_1(T_2), \dots, A_1(T_6)$ 建立 6 个方程式, 可求解常系数 $b_0 \sim b_5$ 。

3) 其它高次项 $A_i(T)$ ($i=2, 3, 4, 5$) 与温度 T 关系式的确立

它们的关系式仍用五阶六项多项式表示:

$$A_2(T) = c_0 + c_1 T + c_2 T^2 + c_3 T^3 + c_4 T^4 + c_5 T^5 \quad (5-19)$$

$$A_3(T) = d_0 + d_1 T + d_2 T^2 + d_3 T^3 + d_4 T^4 + d_5 T^5 \quad (5-20)$$

$$A_4(T) = e_0 + e_1 T + e_2 T^2 + e_3 T^3 + e_4 T^4 + e_5 T^5 \quad (5-21)$$

$$A_5(T) = f_0 + f_1 T + f_2 T^2 + f_3 T^3 + f_4 T^4 + f_5 T^5 \quad (5-22)$$

由 6 个温度值 T_i 与 6 个二次项系数 $A_2(T_i)$ ($i=1, 2, \dots, 6$), 建立 6 个方程式, 可解得 $c_0 \sim c_5$; 由 6 个温度值 T_i 与 6 个三次项系数 $A_3(T_i)$ ($i=1, 2, \dots, 6$), 建立 6 个方程式, 可解得 $d_0 \sim d_5$; 同理, 由 6 个温度值 T_i 分别与 6 个四次项系数 $A_4(T_i)$ 、五次项系数 $A_5(T_i)$ 所建立的各自的 6 个方程式, 可解得常系数 $e_0 \sim e_5, f_0 \sim f_5$, 从而式(5-13)各次项系数与温度的关系式确立。

至此, 当测得传感器工作温度 $T=T_k$ 时, 就可由式(5-16)、式(5-18)~(5-22)求得该温度 T_k 时的各次项系数 $A_0(T_k), A_1(T_k), \dots, A_5(T_k)$, 于是便找出当时温度 T_k 的逆模型:

$$P = A_0(T_k) + A_1(T_k)U + A_2(T_k)U^2 + \dots + A_5(T_k)U^5 \quad (5-23)$$

根据式(5-23)逆模型将传感器输出电压值进行转换, 求得 T_k 时的目标参量压力 P , 就避免了附加温度误差的引入。

5.3 应用举例

本例介绍一种由摩托罗拉公司 1997 年开发的 CMOS 单片集成压力传感器的非线性自校正与温度自补偿功能。

1. 系统设置概述

整个集成压力传感器系统包括：压阻式桥路压力传感器、温度传感器、CMOS 模拟信号调理电路、稳压供电电源和稳流供电电源、8 位微处理器 MCU(68H05)、10 位模数转换器(A/D)、8 位数模转换器(D/A)、2K 字节 EPROM、128 字节 RAM、系统引导程序存储器(BOOT ROM)以及数字通信外围电路接口(SPI)。整个系统的电路结构框图如图 5-4 所示。所有电路在 $10 \mu\text{m}$ 厚的 SOI 封底硅片上制成。

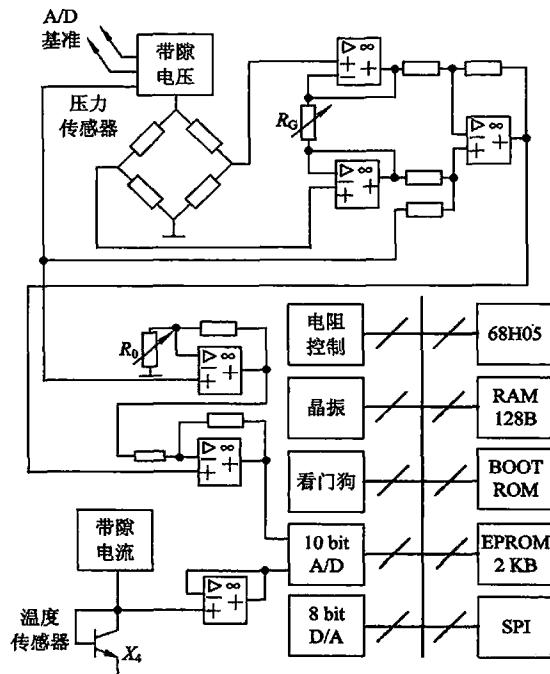


图 5-4 传感器系统的电路结构框图

图中可变电阻 R_G 用来调节放大倍数， R_0 用来调节传感器系统的零点。 R_G 与 R_0 均由微处理器 MCU 的程序控制进行自动调节。晶体管温度传感器由恒流源供电，用来监测工作环境的温度变化；带隙恒压源为压力传感器、调理放大电路以及 A/D 转换器等供电。

2. 校正及补偿方法

(1) 非线性自校正。采用一元二阶三项式逆模型进行刻度转换，即式(5-13)只取前三项：

$$P = A_0(T) + A_1(T)U + A_2(T)U^2 \quad (5-24)$$

式中： U 为校正时压力传感器的输出； P 为数据融合补偿后输出的被测目标参量(压力)；

$A_0(T)$ 、 $A_1(T)$ 、 $A_2(T)$ 为随温度变化的系数。

(2) 温度自补偿。式(5-24)中的多项式系数 $A_0(T)$ 、 $A_1(T)$ 、 $A_2(T)$ 也采用一元二阶三项式来逼近, 即

$$\begin{aligned} A_0(T) &= a_0 + a_1 T + a_2 T^2 \\ A_1(T) &= b_0 + b_1 T + b_2 T^2 \\ A_2(T) &= c_0 + c_1 T + c_2 T^2 \end{aligned} \quad (5-25)$$

式中待定常数有 9 个: a_0 、 a_1 、 a_2 、 b_0 、 b_1 、 b_2 及 c_0 、 c_1 、 c_2 。式(5-25)中的温度 T 也可转换为测温传感器的输出值 U_T 。为求解 9 个待定常数需要建立 9 个方程。通过二维标定实验, 压力标定点的个数至少取三个, 同样温度标定点的个数也至少取三个, 于是至少可以获得 $3 \times 3 = 9$ 个标定值, 从而可建立 9 个方程式, 解得 9 个待定未知常数。监测得到温度值后由式(5-25)可求得式(5-24)中的各项系数 $A_0(T)$ 、 $A_1(T)$ 、 $A_2(T)$ 在温度 T 时的数值; 代入主测参量压力传感器当时的输出电压值 U , 即可由式(5-24)计算获得当时温度 T 的输入目标参量压力 P 。

3. 专用校准系统

专用校准系统由以下几部分组成: 温度控制环境室, 用来给定不同环境温度的标准值; 标准压力源, 用来给定压力的不同标准值; 电源; 用于控制和计算的 PC; 传感器外围接口(SPI)与 PC 之间的接口板。

专用设备的功能包括: 检验压力及温度传感器的输出值, 对压力传感器进行现场的二维标定实验, 控制设置调理放大电路的增益和零点偏移, 清除 EEPROM 内容, 向 EEPROM 写入新的有关系数, 并按 MCU 中 EEPROM 所存入的方程式(5-25)和(5-24)计算融合补偿后的输出值(P)等, 以及提供校准过程中所需的一切服务。

5.4 示例

5.4.1 [示例 5-1] 基于回归分析模型法降低一个干扰量影响的智能化软件模块设计

要求:

- 基于回归分析模型法设计一个数据融合软件模块, 使压阻式压力传感器具有温度自补偿功能, 可提高抗一个干扰量(温度)的稳定性;
- 综合评价该压力传感器在配备了温度自补偿模块融合前、后的性能;
- 写出数据融合模块逆模型的编程代码。

该数据融合模块的输入信号来自如图 5-5 所示二传感器智能传感器系统中的两个传感器: 其一是被补偿的主传感器——压力传感器, 另一个是监测温度干扰量的辅传感器。抗一个温度干扰量的回归模型分析法就是将式(5-1)作为逆模型进行两传感器数据融合。具体设计步骤如下。

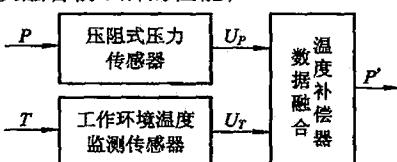


图 5-5 二传感器温度自补偿智能传感器系统

1. 二维标定实验

被补偿的主传感器为 JCY-101 型压阻式压力传感器，其输入与输出量分别为 P 、 U ；监测干扰量温度的辅传感器，其输入、输出量分别为 T 、 U_T 。在工作温度范围内($21.5^{\circ}\text{C} \sim 70^{\circ}\text{C}$)选定 $n=6$ 个不同的温度状态，测定被补偿压力传感器的静态特性，即输入(压力 P)输出(电压 U_P)关系，压力 P 也在量程范围内取 $m=6$ 个标定值：

$$P_i (10^4 \text{ Pa}) : 0, 1.0, 2.0, 3.0, 4.0, 5.0, i=1, 2, \dots, m=6;$$

$$T_j (\text{ }^{\circ}\text{C}) : 21.5, 28.0, 34.0, 44.0, 50.0, 70.0, j=1, 2, \dots, n=6.$$

2. 数据处理

1) 计算矩阵方程的常系数 $\alpha_0 \sim \alpha_5$, $\beta_0 \sim \beta_5$

利用表 5-2 的实验标定数据，求解矩阵方程组(5-5)或方程组(5-7)。

表 5-2 JCY-101 型压力传感器的二维实验标定数据

P/MPa		0	1.0	2.0	3.0	4.0	5.0
$T/\text{ }^{\circ}\text{C}$							
21.5	U_P/mV	-13.84	10.69	28.88	47.05	65.19	83.36
	U_T/mV	27.64	26.95	26.43	25.92	25.45	24.94
28.0	U_P/mV	-13.49	9.32	26.34	43.12	59.99	76.82
	U_T/mV	34.41	33.93	33.47	32.93	32.47	31.91
34.0	U_P/mV	-10.80	7.54	24.84	42.05	59.25	76.38
	U_T/mV	37.76	36.92	36.44	35.97	35.39	35.09
44.0	U_P/mV	-9.72	6.56	23.87	41.21	58.58	75.87
	U_T/mV	54.88	53.97	52.87	52.41	51.93	51.55
50.0	U_P/mV	-8.62	4.86	21.84	38.70	56.32	73.75
	U_T/mV	65.77	64.79	63.84	62.91	61.99	61.06
70.0	U_P/mV	-7.72	3.72	21.25	38.60	55.56	73.28
	U_T/mV	86.12	84.94	83.78	82.65	81.55	80.45

2) 建立二传感器系统的输入 P 、 T 与输出 U 、 U_T 的数学表达式

将求得的常系数 $\alpha_0 \sim \alpha_5$ 或 $\beta_0 \sim \beta_5$ 的数值代入式(5-1)和式(5-2)中即获得用于消除交叉敏感，用二元回归方程描述的静态逆模型：

$$P = \alpha_0 + \alpha_1 U + \alpha_2 U_T + \alpha_3 U^2 + \alpha_4 UU_T + \alpha_5 U_T^2 + \epsilon_P \quad (5-26)$$

$$T = \beta_0 + \beta_1 U + \beta_2 U_T + \beta_3 U^2 + \beta_4 UU_T + \beta_5 U_T^2 + \epsilon_T \quad (5-27)$$

式中， ϵ_P ， ϵ_T 是误差允许范围内可忽略的无穷小误差项，否则应增加多项式的项数。

计算得各常系数的值如下：

$$\alpha_0 = 0.27869353627411; \alpha_1 = 0.05038287227223; \alpha_2 = 0.01282223268423;$$

$$\alpha_3 = 1.697833602 \times 10^{-5}; \alpha_4 = 1.2230607108 \times 10^{-4}; \alpha_5 = -1.0676913988 \times 10^{-4}$$

由上述各常系数值 $\alpha_0 \sim \alpha_5$ 确定的逆模型，其软件编程算式即可用于实现基于回归分析法消除一个干扰量影响的智能化模块。

3) 融合计算结果

由上述常系数值确立的逆模型(式5-26)计算所得目标参量P, 即融合结果值列入表5-3。

表5-3 不同温度条件下压力标定值与融合处理输出的压力计算值(MPa)

序号	目标参量 温度/℃							平均值
		21.5	28.0	34.0	44.0	50.0	70.0	
1	标定值	0.00						
	计算值	-0.19	-0.14	0.02	0.11	0.16	-0.19	-0.038
	偏差 ΔP	-0.19	-0.14	0.02	0.11	0.16	-0.19	-0.038
2	标定值	1.00						
	计算值	1.12	1.10	1.02	1.03	0.95	1.12	1.056
	偏差 ΔP	0.12	0.10	0.02	0.03	-0.05	0.12	0.056
3	标定值	2.00						
	计算值	2.11	2.03	1.98	2.02	1.94	2.11	2.030
	偏差 ΔP	0.11	0.03	-0.02	0.02	-0.06	0.11	0.030
4	标定值	3.00						
	计算值	3.10	2.96	2.94	3.03	2.94	3.10	3.01
	偏差 ΔP	0.10	-0.04	-0.06	0.03	-0.06	0.10	0.01
5	标定值	4.00						
	计算值	4.10	3.90	3.90	4.04	3.98	4.10	4.003
	偏差 ΔP	0.10	-0.10	-0.10	0.04	-0.02	0.10	0.003
6	标定值	5.00						
	计算值	5.10	4.85	4.87	5.05	5.02	5.10	4.998
	偏差 ΔP	0.10	-0.15	-0.13	0.05	0.02	0.10	-0.002

3. 数据融合处理后JCY-101型压力传感器性能的综合评价

用融合处理前后的线性度指标来评价静态性能的改善程度, 用融合处理前后的零位灵敏度系数指标来评价温度稳定性的改善程度。

1) 线性度

(1) 融合处理前: 用21.5℃标定的静态特性计算最小二乘法线性度。拟合直线方程为

$$P(U) = b + kU = 0.57214 + 0.052262U$$

由上述直线方程计算得到的压力拟合值P(U)、标定值P(标)与其拟合偏差 ΔP 列入表5-4(a)。

表5-4(a) 融合处理前拟合值P(U)、标定值P(标)与其拟合偏差 ΔP

标定值 P(标)	0	1.0	2.0	3.0	4.0	5.0
拟合值 P(U)	-0.151	1.131	2.081	3.301	3.979	4.929
拟合偏差 ΔP	-0.151	0.131	0.081	0.031	-0.021	-0.071

单位: MPa;

量程: $Y_{FS} = P_{FS} = 5.0 \text{ MPa}$ 。

由表 5-4(a)数据可得最大拟合偏差 $|\Delta P_m|=0.151 \text{ MPa}$ 。故最小二乘法线性度为

$$\delta_L = \frac{|\Delta P_m|}{P_{fs}} = \frac{0.151}{5.0} = 3\%$$

(2) 融合处理后：为简便起见采用理论线性度来评价，理论线性度方程为

$$P(\bar{U}) = b + k\bar{P} = \bar{P}$$

式中， $b=0$, $k=1$ 。

由上述理论直线方程计算得到的压力拟合值 $P(U)$ 、标定值 $P(\text{标})$ 与其偏差 ΔP 列入表 5-4(b)。

表 5-4(b) 融合处理后拟合值 $P(U)$ 、标定值 $P(\text{标})$ 与其拟合偏差 ΔP

标定值 $P(\text{标})$	0	1.0	2.0	3.0	4.0	5.0
拟合值 $P(\bar{P})$	0.038	1.056	2.030	3.01	4.003	4.998
拟合偏差 ΔP	0.038	0.056	0.030	0.01	0.003	-0.002

单位：MPa；

量程： $Y_{fs} = P_{fs} = 5.0 \text{ MPa}$ 。

由表 5-4(b)数据可得最大拟合偏差 $|\Delta P_m|=0.056 \text{ MPa}$ ，故理论线性度为

$$\delta_L = \frac{|\Delta P_m|}{P_{fs}} = \frac{0.056}{5.0} = 1\%$$

理论线性度的数值比最小二乘法线性度数值要大，故融合后线性度指标可由 3% 提高到优于 1%。

2) 温度影响系数

(1) 零位温度系数。

融合处理前：

$$\alpha_0 = \frac{|\Delta U_{0m}|}{U_{fs}} \cdot \frac{1}{\Delta T}$$

融合处理后：

$$\alpha_0 = \frac{|\Delta P_{0m}|}{P_{fs}} \cdot \frac{1}{\Delta T} \quad (5-28)$$

式中： $\Delta T = T_2 - T_1$ 为工作温度变化范围； P_{fs} 、 U_{fs} 为压力传感器满量程输入与输出值； ΔU_{0m} 为工作温度变化 ΔT 范围内，压力传感器零点漂移最大值； ΔP_{0m} 为逆模型，即式(5-26)融合计算在 ΔT 范围内的零点压力最大偏差。

由表 5-2 所列实验标定数据及表 5-3 融合处理后数据分别可知：

$$\Delta T = 70.0^\circ\text{C} - 21.5^\circ\text{C} = 48.5^\circ\text{C}$$

$$U_{fs} = 83.36 \text{ mV}, \quad P_{fs} = 5.0 \text{ MPa}$$

$$\Delta U_{0m} = |-13.84 - (-7.72)| = 6.12 \text{ mV}$$

$$|\Delta P_{0m}| = 0.038 \text{ MPa}$$

于是可得

融合处理前：

$$\alpha_0 = \frac{|\Delta U_{0m}|}{U_{fs}} \cdot \frac{1}{\Delta T} = \frac{6.12}{83.36 \times 48.5} = 1.5 \times 10^{-3} (\text{}/^\circ\text{C})$$

融合处理后：

$$\alpha_0 = \frac{|\Delta P_{0m}|}{P_{fs}} \cdot \frac{1}{\Delta T} = \frac{0.038}{5.0 \times 48.5} = 1.6 \times 10^{-4} (^{\circ}\text{C})$$

(2) 灵敏度温度系数。灵敏度温度系数的计算公式为

融合前：

$$\alpha_s = \frac{U(T_1) - U(T_2)}{U(T_1)\Delta T}$$

融合后：

$$\alpha_s = \frac{P(T_1) - P(T_2)}{P(T_1)\Delta T} \quad (5-29)$$

式中： $P(T_2)$ 、 $P(T_1)$ 及 $U(T_2)$ 、 $U(T_1)$ 分别为同一输入压力作用下，工作温度为 T_2 、 T_1 时，压力传感器的输入值和输出值； $\Delta T = T_2 - T_1$ ，为工作温度变化范围。

由表 5-2 所列标定数据可知 $\Delta T = 70.0 - 21.5 = 48.5^{\circ}\text{C}$ ；压力传感器的输出电压信号 U 随工作温度升高而减小，在满量程压力值(5.0 MPa)输入时，输出电压值随温度变化有最大改变量为 ΔU_{max} ， $\Delta U_{max} = |U(T_2) - U(T_1)| = |73.28 - 83.36| = 10.08 \text{ mV}$ ；且 $U(T_1) = U(21.5^{\circ}\text{C}) = 83.36 \text{ mV}$ ，则由式(5-29)计算可得融合前灵敏度温度系数为

融合前：

$$\alpha_s = \frac{U(T_1) - U(T_2)}{U(T_1)\Delta T} = \frac{10.08}{83.36 \times 48.5} = 2.49 \times 10^{-3} / ^{\circ}\text{C}$$

由表 5-3 融合处理后数据可知，在 $\Delta T = 48.5^{\circ}\text{C}$ 温度范围内，融合计算值不存在随温度变化单调上升或下降的规律，而是围绕期望值(压力标定值)随机偏离，在满量程 $P_{fs} = 5.0 \text{ MPa}$ 时，两个温度点融合计算压力值的最大偏差 $\Delta P_m = P(T_2) - P(T_1) = 4.85 - 5.1 = -0.25 \text{ MPa}$ ，代入式(5-29)得

融合后：

$$\alpha_s = \frac{P(T_1) - P(T_2)}{P(T_1)\Delta T} = \frac{-0.25}{5.0 \times 48.5} = -1.0 \times 10^{-3} / ^{\circ}\text{C}$$

可见，由以上二元二阶六项多项式表征的逆模型进行处理融合，处理前后的数据表明：线性度指标可由 3% 提高到优于 1%；零位温度系数由 $1.5 \times 10^{-3} / ^{\circ}\text{C}$ 提高到 $1.6 \times 10^{-4} / ^{\circ}\text{C}$ ；灵敏度温度系数由 $2.49 \times 10^{-3} / ^{\circ}\text{C}$ 提高到 $1.0 \times 10^{-3} / ^{\circ}\text{C}$ 。故传感器的静态性能与温度稳定性均得到一定程度的改善。若想得到更好的效果，可以尝试增加逆模型多项式的项数(即减小误差项 ϵ_P 的数值)，或改用其它融合算法。

5.4.2 [示例 5-2] 基于回归分析法模型降低两个干扰量影响的智能化软件模块设计

要求：采用式(5-8)的三元回归方程作为逆模型，改善压阻式压力传感器(JCY-201)的温度稳定性与恒流源供电电流的稳定性，并对改善前后的稳定性作出评价。

监测干扰量温度 T ，温度传感器的输出电压为 U_T ；监测干扰量恒流源供电电流 I ，电流传感器输出电压为 U_I ；构建的三传感器数据融合智能传感器系统框图如图 5-2 所示。

1. 三维标定实验

为了确定刻度转换用三元回归方程表示的逆模型式(5-8)中的 10 个待定常数 $\alpha_0 \sim \alpha_9$ ，

原则上需要建立 10 个方程式，各标定点的数量应能满足此要求。但是为了全量程范围内全面检验融合后稳定性的改善效果，实际上共标定 $m \times n \times w = 216$ 个标定点，各标定点取值如下：

物理量	各物理量的标定点的数值						标定点个数
压力 P/MPa	0 0.1 0.2 0.3 0.4 0.5						$m=6$
度 $T/^\circ\text{C}$	25 35 39 49 55 64						$n=6$
电流 I/mA	6 7 8 9 10 11						$w=6$

在上述各温度状态和供电电流条件下，测定 JCY - 201 压阻式压力传感器的输入输出特性 ($P - U_P$)。相应的三维标定值有 $m \times n \times w = 216$ 个，在表 5 - 5 中仅列出其中的部分值。

表 5 - 5 JCY - 201 型压力传感器三维标定实验标定数据

序号	P /MPa	I /mA	T /°C	U_I /V	U_T /V	U_P /V	序号	P /MPa	I /mA	T /°C	U_I /V	U_T /V	U_P /V
1	0	6.0	25.0	0.6	5.62	0.325	109	0	6.0	49.0	0.6	5.82	0.154
2	0.1	6.0	25.0	0.6	5.62	0.834	110	0.1	6.0	49.0	0.6	5.82	0.747
3	0.2	6.0	25.0	0.6	5.62	1.329	111	0.2	6.0	49.0	0.6	5.82	1.243
4	0.3	6.0	25.0	0.6	5.62	1.825	112	0.3	6.0	49.0	0.6	5.82	1.739
5	0.4	6.0	25.0	0.6	5.62	2.232	113	0.4	6.0	49.0	0.6	5.82	2.237
6	0.5	6.0	25.0	0.6	5.62	2.819	114	0.5	6.0	49.0	0.6	5.82	2.736
7	0	7.0	25.0	0.7	6.56	0.321	115	0	7.0	49.0	0.7	6.8	0.13
8	0.1	7.0	25.0	0.7	6.56	0.925	116	0.1	7.0	49.0	0.7	6.8	0.825
9	0.2	7.0	25.0	0.7	6.56	1.503	117	0.2	7.0	49.0	0.7	6.8	1.403
10	0.3	7.0	25.0	0.7	6.56	2.08	118	0.3	7.0	49.0	0.7	6.8	1.983
11	0.4	7.0	25.0	0.7	6.56	2.66	119	0.4	7.0	49.0	0.7	6.8	2.562
12	0.5	7.0	25.0	0.7	6.56	3.24	120	0.5	7.0	49.0	0.7	6.8	3.145
13	0	8.0	25.0	0.8	7.51	0.317	121	0	8.0	49.0	0.8	7.76	0.106
14	0.1	8.0	25.0	0.8	7.51	1.017	122	0.1	8.0	49.0	0.8	7.76	0.899
15	0.2	8.0	25.0	0.8	7.51	1.677	123	0.2	8.0	49.0	0.8	7.76	1.561
:	:	:	:	:	:	:	:	:	:	:	:	:	:
105	0.2	11.0	39.0	1.1	10.54	2.163	213	0.2	11.0	64.0	1.1	10.92	1.978
106	0.3	11.0	39.0	1.1	10.54	3.073	214	0.3	11.0	64.0	1.1	10.92	2.889
107	0.4	11.0	39.0	1.1	10.54	3.985	215	0.4	11.0	64.0	1.1	10.92	3.802
108	0.5	11.0	39.0	1.1	10.54	4.898	216	0.5	11.0	64.0	1.1	10.92	4.718

注： U_I 为电流传感器输出电压， U_T 为温度传感器输出电压， U_P 为压力传感器输出电压。

2. 数据处理

1) 逆模型的数学表达式

重写由三元回归方程作为逆模型的方程式如下：

$$\begin{aligned} P(U_P, U_I, U_T) = & \alpha_0 + \alpha_1 U_P + \alpha_2 U_I + \alpha_3 U_T + \alpha_4 U_P^2 + \alpha_5 U_I^2 + \alpha_6 U_T^2 \\ & + \alpha_7 U_P U_I + \alpha_8 U_P U_T + \alpha_9 U_I U_T + \varepsilon_P \end{aligned} \quad (5-30)$$

式中 $P(U_P, U_I, U_T)$ 为由式(5-30)融合计算输出的待测目标参量压力值； U_P, U_I, U_T 分别为三个传感器(压力、电流、温度)的输出电压值。

2) 待定常数 $\alpha_0 \sim \alpha_9$ 的确定

将标定实验得到 $s=m \times n \times l$ (m 为压力标定点的个数， n 为温度标定点的个数， l 为电流标定点的个数) 组数据 $(U_{P_i}, U_{I_i}, U_{T_i})$, $i=1, 2, \dots, s$ 全部代入式(5-30)可以得到 $s=m \times n \times l$ 组方程，以及包含 $t+1$ 个未知数、 s 个方程的方程组：

$$\left\{ \begin{array}{l} P_1 = \alpha_0 + \alpha_1 U_{P_1} + \alpha_2 U_{I_1} + \alpha_3 U_{T_1} + \alpha_4 U_{P_1}^2 + \alpha_5 U_{I_1}^2 + \alpha_6 U_{T_1}^2 + \alpha_7 U_{P_1} U_{I_1} \\ \quad + \alpha_8 U_{P_1} U_{T_1} + \alpha_9 U_{I_1} U_{T_1} \\ P_2 = \alpha_0 + \alpha_1 U_{P_2} + \alpha_2 U_{I_2} + \alpha_3 U_{T_2} + \alpha_4 U_{P_2}^2 + \alpha_5 U_{I_2}^2 + \alpha_6 U_{T_2}^2 + \alpha_7 U_{P_2} U_{I_2} \\ \quad + \alpha_8 U_{P_2} U_{T_2} + \alpha_9 U_{I_2} U_{T_2} \\ \vdots \\ P_s = \alpha_0 + \alpha_1 U_{P_s} + \alpha_2 U_{I_s} + \alpha_3 U_{T_s} + \alpha_4 U_{P_s}^2 + \alpha_5 U_{I_s}^2 + \alpha_6 U_{T_s}^2 + \alpha_7 U_{P_s} U_{I_s} \\ \quad + \alpha_8 U_{P_s} U_{T_s} + \alpha_9 U_{I_s} U_{T_s} \end{array} \right.$$

改写为矩阵表示形式为

$$\mathbf{P} = \mathbf{H} \times \boldsymbol{\alpha} + \boldsymbol{\varepsilon}_P$$

其中压力矩阵 \mathbf{P} 为由标定压力值构成的 $s \times 1$ 维列矩阵 $[P_1, P_2, \dots, P_s]^T$, s 为标定数据的总数； \mathbf{H} 为 $s \times t$ 维系数矩阵， $\mathbf{H} = [1, U_P, U_I, U_T, U_P^2, U_I^2, U_T^2, U_P U_I, U_P U_T, U_I U_T]$ ， $t+1$ 为待定常系数的个数。 $\boldsymbol{\alpha}$ 为 $t \times 1$ 待求常系数矩阵， $\boldsymbol{\alpha} = [\alpha_0, \alpha_1, \dots, \alpha_t]^T$ 。 $\boldsymbol{\varepsilon}_P$ 为高阶无穷小量组成的矩阵。

3. 线性回归分析的矩阵方法 Matlab 源程序

线性回归分析的矩阵方法 Matlab 源程序如下：

```

clc
clear
close all %程序运行前, 清屏及其它变量
disp('欢迎使用基于线性回归法的数据融合软件!');
disp('若要融合自己制作的样本数据, 请参照示例程序中的数据格式要求修改程序, 然后融合!');
% 定义输入输出变量
raw_data=[P Ut Ui Up]
P=raw_data(:, 1);
Ui=raw_data(:, 2);
Ut=raw_data(:, 3);
Up=raw_data(:, 4);

```

```

%原始标定数据的输入，可以将数据直接输入至程序中，也可以将数据先输入至文件，然后在
%程序中读入
%数据文件。数据中的 P、Ut、Ui、Up 均为 n×1 维列向量，n 为标定数据点的个数

U0=ones(n, 1); %函数 ones(n, 1)产生 n×1 维列矩阵，矩阵中的每一项值均为 1;
U1=Up; U2=Ui; U3=Ut; U4=Up.*Up;
U5=Ui.*Ui; U6=Ut.*Ut; U7=Up.*Ui;
U8=Up.*Ut; U9=Ui.*Ut; %程序用了 10 个未知参数，如果用 20 个未知参数，需要补充
%更多的项;
%U10=Up.*Up.*Up; U11=Ui.*Ui.*Ui; U12=Ut.*Ut.*Ut; U13=Up.*Up.*
%Ui;
U14=Up.*Up.*Ut;
%U15=Up.*Ui.*Ui; U16=Up.*Ut.*Ut; U17=Up.*Ui.*Ut; U18=Ui.*Ut.*Ut;
U19=Ui.*Ui.*Ut;

H=[U0 U1 U2 U3 U4 U5 U6 U7 U8 U9]; %H 为 10 参数的系数矩阵
E=P;
Z=inv(H'*H); %Z 为系数矩阵 H 的自相关阵 H'*H 的逆矩阵。Inv(H)是 Matlab 中求矩
%阵逆矩阵的函数
alfa=Z*H'*E; %alfa 为所求待定系数矩阵

disp('融合出的回归方程各系数见 F1 后的数据, a0 a1 a2 a3 a4 a5 a6 a7 a8 a9;');
alfa1=alfa(:, 1) %待求系数
disp('压力标定值见 P 后的数据:'); P %压力标定值输出
disp('压力值融合结果见 P2 后的数据:'); P1=H*alfa; P2=P1 %融合后的压力输出值
P3=P1-P %融合后的压力输出值与标定压力值的偏差
disp('请记录回归方程各系数与偏差, 然后使用线性回归检验软件进行检验!');

```

4. 融合计算结果

待定常系数矩阵 $\alpha = [\alpha_0, \alpha_1, \dots, \alpha_9]^T$ 是通过在 Matlab 环境下编程计算求得的。首先将三维标定实验得到的数据 P_i 和 (U_{Pi}, U_{Ui}, U_{Ti}) , $i=1, 2, \dots, s$ 输入到程序中, 所有的压力标定数据 P_i 构成矩阵 $P = [P_1, P_2, \dots, P_s]^T$, 所有的压力传感器输出数据 U_{Pi} 构成矩阵 $U_P = [U_{P1}, U_{P2}, \dots, U_{Ps}]^T$, 同样有 $U_i = [U_{i1}, U_{i2}, \dots, U_{is}]^T$ 和 $U_T = [U_{T1}, U_{T2}, \dots, U_{Ts}]^T$ 。计算系数矩阵 H , 然后将 P 和 H 代入式 $\alpha = (H^T H)^{-1} H^T P$ 中, 就可以解得满足均方误差最小条件的三元二次方程的常系数 $\alpha_0 \sim \alpha_9$ 的具体数值如下:

$$\begin{aligned}
\alpha_0 &= 0.1680, \alpha_1 = 0.2873, \alpha_2 = -1.748 \\
\alpha_3 &= 0.1252, \alpha_4 = 0.0015, \alpha_5 = 1.6087 \\
\alpha_6 &= 0.0072, \alpha_7 = -0.1416, \alpha_8 = -0.0033 \\
\alpha_9 &= -0.2000
\end{aligned}$$

将 $\alpha_0 \sim \alpha_9$ 的值代入式(5-30), 就可由测量值 U_P 、 U_i 、 U_T 计算待求压力值 \hat{P} , 计算结果列入表 5-6。

表 5-6 数据融合处理后压力计算数据 \hat{P} 及与标定值 P 之间的偏差 Δ (压力单位: MPa)

I/mA		6		7		8		10		11	
$T/\text{°C}$	P	\hat{P}	Δ	\hat{P}	Δ	\hat{P}	Δ	\hat{P}	Δ	\hat{P}	Δ
25°C	0	0.0134	0.0134	-0.0029	-0.0029	-0.0112	-0.0112	-0.0072	-0.0072	0.0066	0.0066
	0.5000	0.4835	-0.0165	0.4987	-0.0013	0.5081	0.0081	0.5049	0.0049	0.4934	-0.0066
35°C	0	0.0106	0.0106	-0.0044	-0.0044	-0.0107	-0.0107	-0.0022	-0.0022	0.0108	0.0108
	0.5000	0.4904	-0.0096	0.5045	0.0045	0.5133	0.0133	0.5100	0.0100	0.4956	-0.0044
39°C	0	0.0232	0.0232	0.0090	0.0090	0.0014	0.0014	0.0062	0.0062	0.0195	0.0195
	0.5000	0.4900	-0.0100	0.5059	0.0059	0.5137	0.0137	0.4993	-0.0007	0.4970	-0.0030
49°C	0	-0.0008	-0.0008	-0.0156	-0.0156	-0.0243	-0.0243	-0.0161	-0.0161	-0.0019	-0.0019
	0.5000	0.4831	-0.0169	0.4988	-0.0012	0.5059	0.0059	0.5023	0.0023	0.4884	-0.0116
55°C	0	0.0160	0.0160	0.0005	0.0005	-0.0060	-0.0060	-0.0018	-0.0018	0.0116	0.0116
	0.5000	0.4838	-0.0162	0.4976	-0.0024	0.5079	0.0079	0.5031	0.0031	0.4902	-0.0098
64°C	0	0.0117	0.0117	-0.0043	-0.0043	-0.0118	-0.0118	-0.0059	-0.0059	0.0077	0.0077
	0.5000	0.4870	-0.0130	0.5018	0.0018	0.5101	0.0101	0.5065	0.0065	0.4934	-0.0066

注: 表中 T 表示工作温度, I 表示工作电流, P 表示压力标定值, \hat{P} 表示压力计算值, 偏差 $\Delta = \hat{P} - P$ 。

5. 融合处理效果评价

降低温度与电源波动两个干扰量的影响, 传感器系统稳定性的改善程度分别用融合前后的零位、灵敏度温度系数以及电流影响系数来评价。

由表 5-6 标定实验数据可得: 满量程输出值 $U_{\text{FS}} = 4.928 \text{ V}$, 当温度的变化范围为 $\Delta T = 39^\circ\text{C}$ (由 25°C 至 64°C), 电流由 6 mA 至 11 mA, 变动 $\Delta I = 5 \text{ mA}$ 时, 零点值的最大变化范围 $\Delta U_{0m} = 0.356 \text{ V}$ (由 0.328 V 至 -0.028 V), 满量程输出值由 4.928 V 下降到 2.710 V, 输出变化范围 $\Delta U_m = -2.218 \text{ V}$ 。

1) 融合前

(1) 零位温度系数

$$\alpha_0 = \frac{|\Delta U_{0m}|}{U_{\text{FS}} \times \Delta T} = \frac{0.356}{4.928 \times 39} \approx 1.9 \times 10^{-3} (\text{/}^\circ\text{C})$$

(2) 灵敏度温度系数

$$\alpha_s = \frac{|\Delta U_m|}{U_{\text{FS}} \times \Delta T} = \frac{|2.710 - 4.928|}{4.928 \times 39} \approx 1.2 \times 10^{-2} (\text{/}^\circ\text{C})$$

(3) 电流影响系数

$$\alpha_I = \frac{|\Delta U_m|}{U_{\text{FS}} \times \Delta I} = \frac{|2.710 - 4.928|}{4.928 \times 5} \approx 9.0 \times 10^{-2} (\text{/mA})$$

2) 融合后

在 $\Delta T = 39^\circ\text{C}$, $\Delta I = 5 \text{ mA}$ 变化范围内, 零点融合计算值的最大偏差 $|\Delta P'_{0m}| = 0.0243 \text{ MPa}$; 满量程压力 $P_{\text{FS}} = 0.5 \text{ MPa}$, 其融合计算值的最大偏差量 $|\Delta P'_m| = 0.0169 \text{ MPa}$, 则有:

$$\alpha_0 = \frac{|\Delta P'_{0m}|}{P_{\text{FS}} \times \Delta T} = \frac{0.0243}{0.5 \times 39} \approx 1.2 \times 10^{-3} (\text{/}^\circ\text{C})$$

$$\alpha_s = \frac{|\Delta P'_m|}{P_{FS} \times \Delta T} = \frac{0.0169}{0.5 \times 39} \approx 8.7 \times 10^{-4} (^{\circ}\text{C})$$

$$\alpha_I = \frac{|\Delta P'_m|}{P_{FS} \times \Delta I} = \frac{0.0169}{0.5 \times 5} \approx 6.8 \times 10^{-3} (\text{/mA})$$

将上述各系数的计算结果综合于表 5-7。

表 5-7 融合前后各评价系数的计算值

评价参数	零位温度系数 α_0 ($^{\circ}\text{C}$)	灵敏度温度系数 α_s ($^{\circ}\text{C}$)	电流影响系数 α_I (/mA)
融合前计算值	1.9×10^{-3}	1.2×10^{-2}	9.0×10^{-2}
融合后计算值	1.2×10^{-3}	8.7×10^{-4}	6.8×10^{-3}

由上表中融合前后的各系数值可以看出，融合后温度灵敏度系数 α_s 提高近两个数量级，电流影响系数 α_I 提高一个数量级，而零位温度系数 α_0 也改善了 $1/3$ 。

参 考 文 献

- [1] Yoshiharu Yoshi, et al. 1 Chip Integrated Softwear Calibrated CMOS Pressure Sensor with MCU, A/D Convertor, D/A Convertor, Digital Communication Port, Signal Conditioning Circuit and Temperature Sensor[C], Proc. Transducers'97, Chicago, June 16 - 19, 1997, 1485 - 1487
- [2] 刘君华, 张重斌, 朱长纯. 硅压阻式压力传感器温度系数的在线实时补偿[J]. 传感技术学报, 1989(2)
- [3] 刘君华. 采用压力传感器同时测量动态压力与温度[J]. 微电子与计算机, 1990(2): 8 - 11
- [4] 常炳国, 刘君华. 采用神经网络技术获取高抗干扰的传感器系统特性[J]. 传感技术学报, 2000(1): 33 - 37
- [5] G H Golub, C F Van Loan. 矩阵计算[M]. 袁亚湘, 等译. 北京: 科学出版社, 2001
- [6] 李世维, 刘君华, 等. 降低多传感器交叉敏感的法方程组法[J]. 自然版. 西安: 西安交通大学出版社, 2009

第6章 神经网络技术与其在智能传感器系统中的应用

本章内容

- 人工神经网络基础知识
- 前馈 BP、径向基 RBF 神经网络
- [示例 6-1]基于 BP 神经网络模型法温度自补偿智能化逆模型的设计
- [示例 6-2]基于 RBF 神经网络模型法抗两个干扰量影响的智能化逆模型的设计

神经网络在智能传感器系统中已得到了广泛成功的应用。近年来，它也成为智能传感器系统实现智能化功能的一种极其有效的智能化技术手段。

基于改善传感器稳定性、抑制交叉敏感的多传感器技术模型法(详见 4.3.3 小节)，本章着重讨论如何应用 BP、RBF 神经网络建立消除干扰量影响的智能化逆模型；通过神经网络在传感器智能化功能中的应用，介绍其改善传感器性能的智能化技术工作原理与使用方法。

为了避免繁琐的神经网络编程，本章所用神经网络取自 Matlab 神经网络工具箱，故也简要介绍 Matlab 神经网络工具箱中网络的结构体系和使用函数。

6.1 概 述

人工神经网络(Artificial Neuron Networks, ANN)是由大量的处理单元组成的非线性大规模自适应动力系统。它是在现代神经生理科学研究成果的基础上提出来的，是人们试图通过模拟大脑神经网络处理、记忆信息的方式设计的一种使之具有人脑那样的信息处理能力的新“机器”。Kononen 曾给出一个神经网络的定义，他指出神经网络是由一些简单的(通常为自适应的)单元及其层次组织的大规模并行联结构造的网络，它致力于按照生物神经系统的同样方式处理真实世界的客观事物，从而反映了人脑功能的若干特征，但并非神经系统的真实描写，而是它的简化、抽象和模拟。

神经网络的研究可以追溯到 19 世纪 40 年代。1943 年，心理学家麦克洛奇(W. McCulloch)和数理逻辑学家皮兹(W. Pizz)在《数学生物物理公报(Bulletin of Mathematical Biophysics)》上发表了关于神经网络的数学模型。这个模型后来一直被称为 M-P 神经网络模型。他们总结了神经元的一些基本生理特征，提出神经元形式化的数学描述和网络的结构方法，神经网络的研究从此开始了。到 19 世纪 70 年代，虽然由于人工智能、专家系统的发展，使得神经网络的发展一度出现低潮，但神经网络的研究并没有因此停滞不前，80 年代神经网络又重新兴起。目前神经网络广泛应用于传感器信息处理、信号处理、自动控制、知识处理、运输与通信等领域。

6.2 神经网络基础知识

6.2.1 神经网络结构

如果将大量功能简单的基本神经元通过一定的拓扑结构组织起来，构成群体并行分布式处理的计算结构，那么这种结构就是人工神经网络。

根据神经元之间连接的拓扑结构的不同，可将神经网络分为两大类：分层网络和相互连接型网络。

1. 分层网络结构

分层网络将一个神经网络模型中的所有神经元按功能分成若干层，通常有输入层、隐层（中间层）和输出层，各层按顺序连接，如图 6-1 所示。输入层是与外部激励打交道的界面，它接受外部输入模式，并由各输入单元传给与之相连的隐层各神经元；隐层（中间层，可以不只一层）是网络内部处理单元的工作区域。不同模型的处理功能差别主要反映在对中间层的处理上。输出层将网络计算结果输出，是与外部显示设备或执行机构打交道的界面。同层之间神经元互不相连，相邻层神经元之间的连接强度由连接权值表示。 w_{ji} 为隐层第 j 个神经元与输入层第 i 个神经元之间的连接权值； v_{kj} 为输出层第 k 个神经元与隐层第 j 个神经元之间的连接权值。

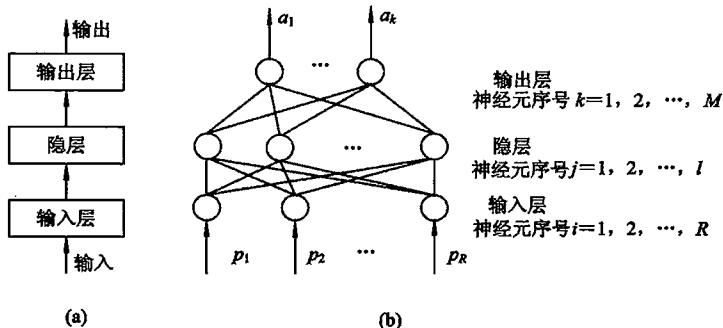


图 6-1 分层网络功能层次

2. 相互连接型网络结构

所谓相互连接型网络，是指网络中任意两个单元之间是可达的，即存在连接路径，如图 6-2 所示。在该网络结构中，对于给定的某一输入模式，由某一初始网络参数出发，在一段给定的时间内网络处于不断改变输出模式的动态变化之中。最后，网络可能会产生某一稳定输出模式，但也有可能进入周期性振荡状态。因此，相互连接型网络可以认为是一种非线

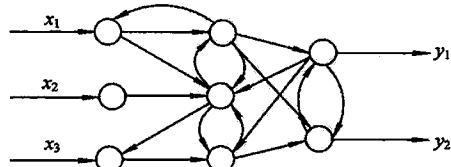


图 6-2 相互连接型网络

性动力学系统。

6.2.2 神经元模型

在人脑中，神经细胞之间的信息传递并非是一个神经细胞把自己所接收到的信息原封不动地传给另一个细胞，而是对所接收到的信息进行一定的处理，将处理过的信息传递给另一个神经细胞。作为对人脑神经细胞的一种模拟，人工神经元也必须具有一定的信息处理能力。

神经元是人工神经网络的基本单元，图 6-3 所示为一个简单的人工神经元模型。 $p_1, p_2, \dots, p_i, \dots, p_R$ 表示神经元(序号 l)的 R 个输入； w_{il} 表示该神经元(序号为 l)与前层第 i 个神经元的连接权值； b 为偏置值，又称阈值； a 为神经元的输出，神经元的输入与输出关系的一般数学表达式如下：

$$\left. \begin{aligned} n &= b + \sum_{i=1}^R p_i \cdot w_{il} \\ a &= f(n) \end{aligned} \right\} \quad (6-1)$$

其中： n 为该神经元(序号 l)的总输入； $f(n)$ 为表示神经元输入输出关系的函数，称为作用函数、响应函数或传递函数。当 $b=0$ 时，称为无偏置/无阈值神经元；当 $b \neq 0$ 时，表示当神经元所接受的输入达到一定强度后，才能被激活，称为有偏置/阈值神经元；当 $R=1$ 时，为单输入神经元；当 $R>1$ 时，称为多输入或矢量输入神经元，此时连接权 $w_{11}, w_{12}, \dots, w_{1R}$ 组成一矢量。

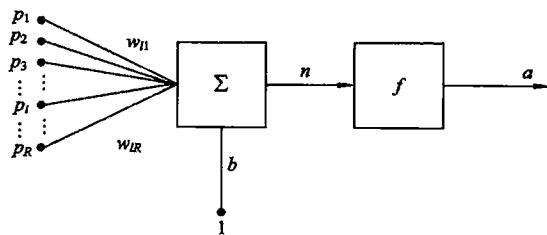


图 6-3 神经元模型

通过式(6-1)所表示的神经元的输入输出关系，还可以看出神经元的两个基本特性：输出是各个输入综合作用的结果；神经元具有可塑性，即它的输出是可通过改变连接权值 w_{il} 来调节的。

6.2.3 神经元作用函数

神经元的作用函数有多种形式，最常用的有 S型(sigmoid)、线性型(line)、硬限幅型(hardlim)、高斯函数型(Gauss)等。

1. S型作用函数

S型作用函数反映了神经元的非线性输入输出特性，它又分为对数型、正切型等多个类型。

(1) 对数型(logsig)：输入输出特性采用对数函数表示，其函数曲线如图 6-4(a)所示，

数学表达式如下：

$$a = f(n) = \frac{1}{1 + e^{-n}} \quad (6-2)$$

其中： n 表示神经元的总输入； a 为神经元的输出。函数 $f(n)$ 的值域为 $(0, 1)$ ，是一个单边函数；如果将 $f(n)$ 减去 0.5 就可得到一个双边函数。

(2) 正切型(tansig)：函数曲线如图 6-4(b)所示，数学表达式如下：

$$a = \frac{2}{1 + e^{-2 \cdot n}} - 1 \quad (6-3)$$

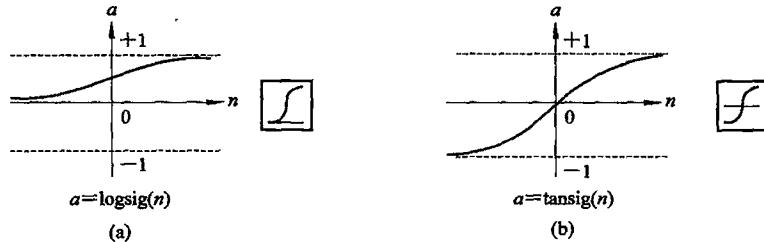


图 6-4 S 型神经元作用函数
(a) 对数型 sig 函数；(b) 正切型 sig 函数

2. 线性型(line)

(1) 饱和线性函数：函数曲线如图 6-5(a)所示，数学表达式如下：

$$a = \begin{cases} 0, & \text{如果 } n < 0 \\ n, & \text{如果 } 0 \leq n \leq 1 \\ 1, & \text{如果 } n > 1 \end{cases} \quad (6-4)$$

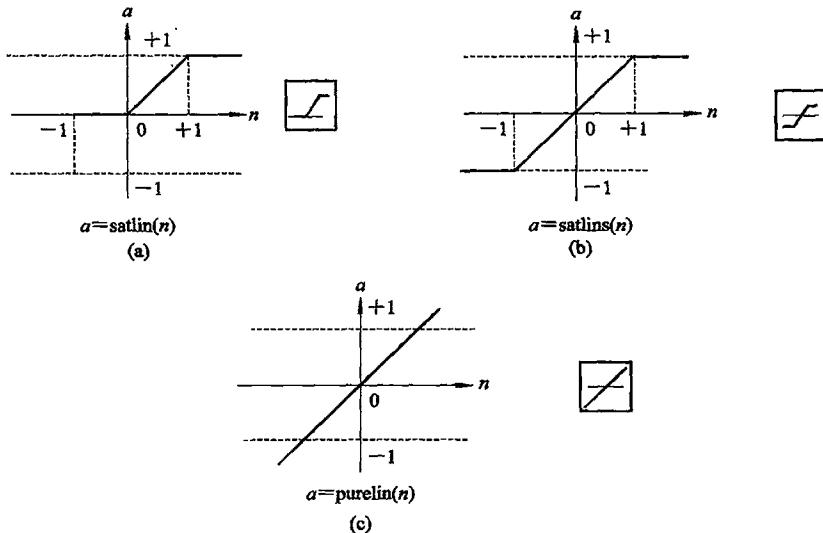


图 6-5 线性作用函数
(a) 饱和线性函数；(b) 对称饱和线性函数；(c) 纯线性函数

(2) 对称饱和线性函数：函数曲线如图 6-5(b)所示，数学表达式如下：

$$a = \begin{cases} -1, & \text{如果 } n < -1 \\ n, & \text{如果 } -1 \leq n \leq 1 \\ 1, & \text{如果 } n > 1 \end{cases} \quad (6-5)$$

(3) 纯线性函数：函数曲线如图 6-5(c)所示，数学表达式如下：

$$a = n \quad (6-6)$$

3. 硬限幅型(hardlim)

(1) 硬限幅函数：函数曲线如图 6-6(a)所示，数学表达式如下：

$$a = \begin{cases} 1, & \text{若 } n \geq 0 \\ 0, & \text{其它} \end{cases} \quad (6-7)$$

(2) 对称硬限幅函数：函数曲线如图 6-6(b)所示，数学表达式如下：

$$a = \begin{cases} 1, & \text{若 } n \geq 0 \\ 0, & \text{其它} \end{cases} \quad (6-8)$$

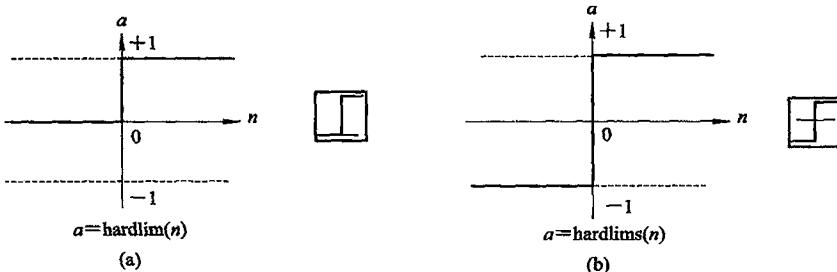


图 6-6 硬限幅作用函数

(a) 硬限幅函数；(b) 对称硬限幅函数

4. 高斯型函数(Gauss)

高斯型函数的函数曲线如图 6-7 所示，数学表达式如下：

$$a = e^{-\frac{n^2}{2}} \quad (6-9)$$

高斯函数的输出为 1 和 0.5 时所对应的输入之间的差值称为函数的分散度(spread)，明显地，对应于式(6-9)的分散度为 0.833。高斯神经元函数通常用作 RBF 网络的隐层传递函数。

对于神经元作用函数的选择，目前还没有定性的法则，一般根据应用情况的不同而定，但对于 S 型非线性函数，一般用于多层神经网络的隐层，而线性函数和限幅函数则多用于神经网络的输出层。

大脑中的各个神经元之间的连接方式多种多样。根据大脑中神经元连接方式的不同，我们同样可以构造出各种各样的人工神经网络模型。就目前常用的比较成熟的网络而言，

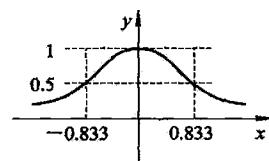


图 6-7 高斯型函数

可以从不同的角度进行分类。若从网络的结构上进行分类，可将神经网络分为前馈神经网络和反馈神经网络。有代表性的网络模型有感知器、多层映射 BP、RBF、双向联想记忆(BAM)、Hopfield 模型等，它们是从不同的角度对生物神经系统不同层次的描述和模拟。利用这些网络模型可实现函数逼近、数据采集聚类、模式分类、优化计算等功能。因此，神经网络广泛应用于人工智能、自动控制、机器人、统计学等领域的信息处理中。

前馈神经网络主要包括感知器神经网络、BP 神经网络和 RBF 神经网络等，本节主要介绍 BP 神经网络和 RBF 神经网络。

6.2.4 BP 神经网络

1. BP 神经网络

BP 神经网络由于其权值采用反向传播(Back Propagation)的学习算法而得名。Rumelhart、McClelland 和他们的同事于 1982 年成立了一个 PDP 小组，研究并行分布信息处理方法，探索人类认知的微结构，经过三年的努力，于 1985 年发展了 BP 神经网络学习算法。BP 网络是一种多层前馈神经网络，其神经元的变换函数是 S 型函数，因此其输出量为 0 到 1 之间的连续量，它可以实现从输入到输出的任意的非线性映射。在确定了 BP 网络的结构后，利用输入输出样本集对其进行训练，也即对网络的权值和偏置值(bias)进行学习和调整，以使网络实现给定的输入输出映射关系。经过训练的 BP 网络，对于不是样本集中的输入，也能给出合适的输出，这种性质称为泛化(generalization)功能。因此，BP 神经网络具有拉格朗日(Lagrange)插值法、牛顿(Newton)插值法等类似的插值功能，只是拉格朗日插值法和牛顿插值法只能用于二维空间的曲线插值，而 BP 神经网络可实现多维空间的曲面插值。

2. BP 网络结构

BP 神经网络通常有一个或多个隐层。在实际应用中，用得最多的是三层 BP 神经网络。图 6-8 和图 6-9 所示为一个简单的三层 BP 网络模型。网络的输入层包含 $i(i=3)$ 个节点，隐层包含 $j(j=4)$ 个节点，输出层有 $k(k=2)$ 个节点。连接权值 Iw_{ji} 表示输入层第 i 个节点与隐层的第 j 个节点的连接权值，共有 4×3 个权值。 Lw_{kj} 表示隐层第 j 个节点与输出层第 k 个节点的连接权值，共有 2×4 个连接权值。

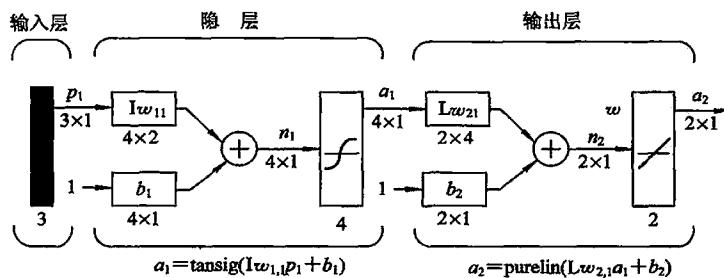


图 6-8 BP 网络模型

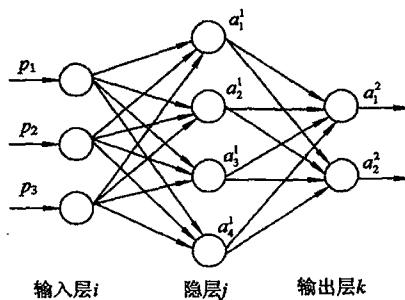


图 6-9 BP 网络模型示意图

3. BP 神经网络的神经元模型

BP 神经网络一般由多层神经元构成, 因此它的神经元可以有多种类型, 其神经元的选用需要视具体情况而定。由于 BP 神经网络是通过误差反向传播来实现的, 因此, BP 神经网络中的神经元必须是连续可微的, 所以, BP 神经网络的神经元函数不能选用限幅函数。对于输出范围比较小的网络, 可以将其所有的神经元全部选为 S 型函数; 若网络的输出范围比较大, 则一般把隐含层神经元选为 S 型函数, 而把输出层神经元选为纯线性函数, 从理论上讲, 这样选择神经元函数可以任意精度逼近任意的一个平滑函数。对于图 6-9 所示 BP 网络模型的神经作用函数选用情况如下:

(1) 输入层神经元作用函数。输入层神经元作用函数选为纯线性函数, 故节点 i 的输出为

$$O_i = p_i \quad (6-10)$$

其中, p_i 为第 i 个节点的输入。

(2) 隐层神经元作用函数。隐层神经元作用函数选用对数 S 型函数, 故节点 j 的输出为

$$O_{1j} = f(n_{1j}) = \frac{1}{1 + e^{-n_{1j}}} = a_j^1 \quad (6-11)$$

节点 j 的总输入:

$$n_{1j} = \sum_{i=1}^{R=3} O_i \cdot Iw_{ji} + b_{1j} \quad (6-12)$$

(3) 输出层神经元作用函数。输出层神经元作用函数选对数 S 型函数。

节点 k 的输出:

$$O_{2k} = f(n_{2k}) = \frac{1}{1 + e^{-n_{2k}}} \quad (6-13)$$

节点 k 的总输入:

$$n_{2k} = \sum_{j=1}^{l=4} O_{2j} \cdot Lw_{kj} + b_{2k} \quad (6-14)$$

4. BP 神经网络的学习算法

当权值 Iw_{ji} ($l \times R$ 个)、 Lw_{kj} ($m \times l$ 个)与阈值 b_{1j} (l 个)、 b_{2k} (m 个)随机赋予初始值、确定分组输入 p_1, p_2, \dots, p_R 后, 根据式(6-10)、(6-11)、(6-13)、(6-14)进行计算, 就可得出输出层节点 k 的输出 O_{2k} 与期望输出 d_k 存在误差, 输出层 m 个节点的总误差 E 取为

$$E = \frac{1}{2} \sum_{k=1}^{m-2} [d_k - O_{2k}]^2 \quad (6-15)$$

网络学习也称为网络训练，就是通过反复的计算，求取 E ，根据 E 的大小调整网络参数，最终使得误差 E 足够小。网络权值参数的修正数学表达式求取所遵循的规则称为学习规则，其基本思想是：使权值沿误差函数 E 的负梯度 $-\frac{\partial E}{\partial w}$ 方向改变，即

$$\Delta Lw_{kj} = Lw_{kj}(t+1) - Lw_{kj}(t) = -\eta \frac{\partial E}{\partial Lw_{kj}} \quad (6-16)$$

$$\Delta Iw_{ji} = Iw_{ji}(t+1) - Iw_{ji}(t) = -\eta \frac{\partial E}{\partial Iw_{ji}} \quad (6-17)$$

其中， η 为学习因子，又称步长。

按照误差反向传播算法，分别求取输出层训练误差 δ_{2k} 、隐层训练误差 δ_{1j} ，最后得出权值修正公式。

(1) 输出层训练误差 δ_{2k} ：

$$\delta_{2k} = -\frac{\partial E}{\partial n_{2k}} = -\frac{\partial E}{\partial O_{2k}} \cdot \frac{\partial O_{2k}}{\partial n_{2k}} \quad (6-18)$$

根据误差定义式(6-15)及式(6-13)、(6-14)，可求得

$$\frac{\partial E}{\partial O_{2k}} = -(d_k - O_{2k}) = [d_k - f(n_{2k})] \quad (6-19)$$

根据式(6-13)、(6-14)可得

$$\frac{\partial O_{2k}}{\partial n_{2k}} = f'(n_{2k}) = f(n_{2k}) \cdot [1 - f(n_{2k})] \quad (6-20)$$

将式(6-19)、(6-20)代入式(6-18)，得

$$\delta_{2k} = -f(n_{2k})[1 - f(n_{2k})] \cdot [d_k - f(n_{2k})] \quad (6-21)$$

(2) 隐层训练误差 δ_{1j} ：

$$\delta_{1j} = -\frac{\partial E}{\partial n_{1j}} = -\frac{\partial E}{\partial O_{1j}} \cdot \frac{\partial O_{1j}}{\partial n_{1j}} \quad (6-22)$$

根据式(6-11)可得

$$\frac{\partial O_{1j}}{\partial n_{1j}} = f'(n_{1j}) = f(n_{1j}) \cdot [1 - f(n_{1j})] \quad (6-23)$$

根据式(6-18)、(6-14)可得

$$\frac{\partial E}{\partial O_{1j}} = \frac{\partial E}{\partial n_{1k}} \cdot \frac{\partial n_{1k}}{\partial O_{1j}} = -\delta_{1k} \cdot \sum_{j=1}^{l=4} Iw_{kj} \quad (6-24)$$

将式(6-23)、(6-24)代入式(6-22)，得

$$\delta_{1j} = -f(n_{1j}) \cdot [1 - f(n_{1j})] \cdot \sum_{j=1}^{l=4} \delta_{1k} \cdot Iw_{kj} \quad (6-25)$$

(3) 权值修正公式。

① Lw_{kj} 的修正公式：将式(6-16)变换为

$$\Delta Lw_{kj} = -\eta \cdot \frac{\partial E}{\partial Lw_{kj}} = -\eta \cdot \frac{\partial E}{\partial n_{2k}} \cdot \frac{\partial n_{2k}}{\partial Lw_{kj}} = \eta \cdot \delta_{2k} \cdot O_{2j} \quad (6-26)$$

其中

$$\delta_{2k} = -\frac{\partial E}{\partial n_{2k}}; \quad O_{2j} = \frac{\partial n_{2k}}{\partial Lw_{kj}}$$

则有

$$Lw_{kj}(t+1) = Lw_{kj}(t) + \eta \delta_{2k} O_{2j} \quad (6-27)$$

② Iw_{ji} 的修正公式：将式(6-17)变换为

$$\Delta Iw_{ji} = Iw_{ji}(t+1) - Iw_{ji}(t) = -\eta \frac{\partial E}{\partial n_{ij}} \frac{\partial n_{ij}}{\partial Iw_{ji}} = \eta \delta_{1j} O_{1i} \quad (6-28)$$

其中

$$\delta_{1j} = -\frac{\partial E}{\partial n_{ij}}, \quad O_{1i} = \frac{\partial n_{ij}}{\partial Iw_{ji}}$$

则有

$$Iw_{ji}(t+1) = Iw_{ji}(t) + \eta \delta_{1j} O_{1i} \quad (6-29)$$

引入势态因子 α ，最后修正公式变为

$$Lw_{kj} = Lw_{kj}(t) + \eta \delta_{2k} O_{2j} + \alpha [Lw_{kj}(t) - Lw_{kj}(t-1)] \quad (6-30)$$

$$Iw_{ji} = Iw_{ji}(t) + \eta \delta_{1j} O_{1i} + \alpha [Iw_{ji}(t) - Iw_{ji}(t-1)]$$

(4) 学习流程图。BP 网络的一个样本的学习流程如图 6-10 所示。

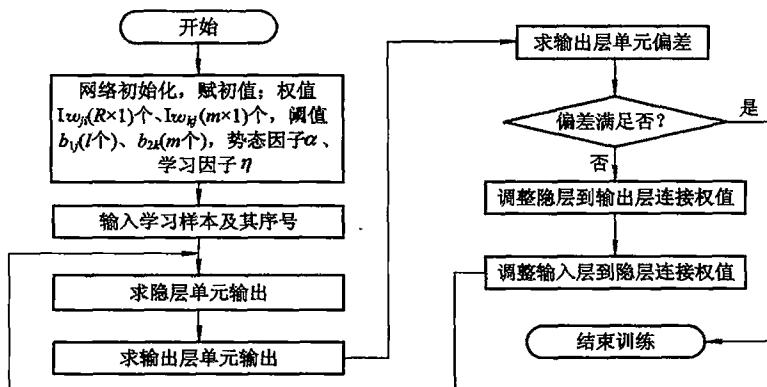


图 6-10 BP 网络训练过程及算法流程

- ① 网络初始化，随机设定连接权值 Iw_{ji} 、 Lw_{kj} ，阈值 b_{1j} 、 b_{2k} ，学习因子 η ，势态因子 α 。
- ② 向具有上述初始值的神经网络提供输入学习样本和序号。
- ③ 计算隐层单元输出值。
- ④ 计算输出层单元的输出。
- ⑤ 计算输出层和隐层训练误差 δ_{2k} 、 δ_{1j} 。
- ⑥ 修正权值。
- ⑦ 判断均方误差 e 是否满足给定允许偏差 ϵ 。当满足时转到⑧，否则转向⑤、⑥和⑦。
- ⑧ 结束训练。

6.2.5 径向基(RBF)神经网络

从网络的函数逼近角度来说，神经网络可分为全局逼近神经网络和局部逼近神经网络。前面谈到的 BP 神经网络由于对每个输入输出数据对，网络的每一个权值均需要调整，因此是全局逼近网络。RBF 神经网络对于每个输入输出对，只有少量的权值需要进行调

整，所以是局部逼近神经网络。也正因为 RBF 网络的局部逼近特性，使得它在逼近能力、分类能力和学习速度等方面均优于 BP 网络。

在 Matlab 环境中，可以用 newff 函数来创建一个 BP 神经网络。

1. RBF 神经网络模型

RBF 神经网络的神经元作用函数采用高斯型函数，其神经元模型如图 6-11 所示。它同前面所谈到的神经元有所不同，前面谈到的 BP 网络的神经元的总输入是对各输入和偏置值(bias)进行加权求和得到的，而如图 6-11 所示的 RBF 网络的神经元的总输入是权值矩阵的行向量与输入向量的向量距与偏置值(bias)的乘积，其数学表达式为

$$n_i = b_i \sqrt{\sum_j (w_{ij} - p_{ji})^2} \quad (6-31)$$

其中： n_i 为网络隐层第 i 个神经元的总输入； b_i 为第 i 个神经元的偏置值(bias)； w_{ij} 为隐层权值矩阵的第 i 个行向量的第 j 个元素； p_{ji} 是第 j 个输入向量的 i 时刻输入值。

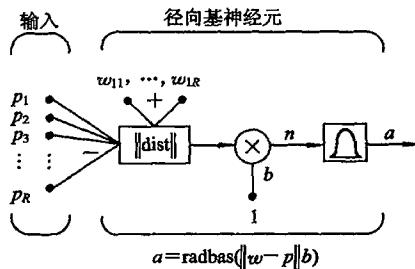


图 6-11 径向基神经元

RBF 神经网络模型同 BP 网络类似，通常其输出层是纯线性神经元，只是其隐层神经元是称之为 radbas 型的神经元。RBF 神经网络模型如图 6-12 所示。

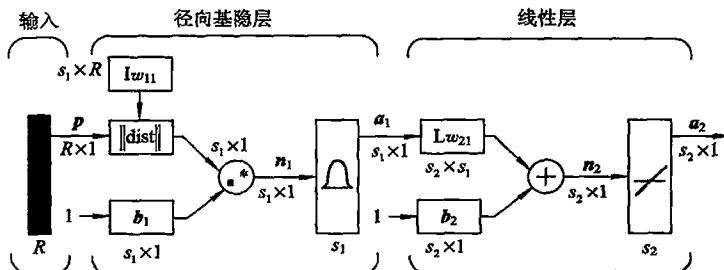


图 6-12 径向基神经网络模型

2. RBF 神经网络的训练

由高斯函数的函数表达式可知，其输出最大值为 1，当输入向量与权值向量的向量距减小时，神经元的输出增大。偏置 b 用来调节高斯函数对输入的灵敏度， b 的绝对值越大，神经元对输入越灵敏，也就是说，神经元的响应函数曲线越“宽”。对应于图 6-7，也就是函数的输出为 0.5 时的两个总输入之差的绝对值越大。对于 b 的取值，一般由训练样本的样本距和样本的范围决定， b 的取值大于两个相邻样本点的最大距离，而小于任意两个样

本的最大距离。例如，对于一个单输入单输出的 RBF 神经网络，输入的样本为 $\{-6, -4, 0, 2, 4, 7\}$ ，那么 b 的取值应大于 4 而小于 13。

RBF 神经网络中，其隐层神经元的数量可以说是由样本点的数量来决定的，有多少个输入样本点，就有多少个隐层神经元。对于每个隐层神经元的输入，其输出满足下列条件：

(1) 若是其对应的样本点，我们也称其为该神经元的特征输入点，那么其对应的输出应趋于 1；

(2) 对于非样本点输入，输入与特征输入的点距离越远，则神经元的输出越小。

因此，RBF 神经网络的输入权值是由样本决定的，而与期望输出并没有太大关系。

在输入权值、隐层神经元的偏置值(bias)全部确定好之后，隐层的输出也就确定了。由于 RBF 神经网络的输出层神经元的响应函数是纯线性函数，因此，在选定输出层神经元之后，隐层与输出层之间的神经元连接权值可以由下式确定：

$$\begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1s_1} & b_1 \\ w_{21} & w_{22} & \cdots & w_{2s_1} & b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ w_{s_21} & w_{s_22} & \cdots & w_{s_2s_1} & b_{s_2} \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_{s_2} \\ 1 \end{bmatrix} = T \quad (6-32)$$

其中： w_{ij} 为输出层第 i 个神经元与隐层第 j 个神经元的连接权值； b_i 为输出层第 i 个神经元的偏置值(bias)； a_j 为隐层第 j 个神经元的输出向量； T 为理想输出矩阵。求解式(6-32)，即可得到输出层与隐层的连接权值。

在 Matlab 环境中，可以用 newrb 函数来创建一个 RBF 神经网络。

6.3 示例

本节的示例均基于 4.3.3 节多传感器技术改善传感器性能的模型法，均以消除工作环境温度 T 或供电电压 U 、电流 I 等干扰量对压力传感器交叉敏感的影响为例，说明基于神经网络的智能化软件模块的设计方法与步骤。

6.3.1 [示例 6-1] 基于神经网络模型法温度自补偿智能化模块的设计

要求：

在 Matlab 环境下设计一个具有如下功能的 BP 神经网络：

- 具有温度补偿功能，可对易受环境温度影响的传感器，如压阻式压力传感器进行温度补偿；
- 该补偿器中的神经网络模块可更换学习样本进行再训练以适应不同力量程的传感器在不同工作环境温度影响下进行温度补偿，提高温度稳定性；
- 网络训练完毕后，输入压力传感器量程范围内的任何输出电压值 U_P 与其工作环境温度传感器的输出电压值 U_T ，补偿模块可给出对应的压力值，该压力值 P' 在消除温度影响的同时也进行了零点及非线性补偿。
- 写出 BPNN(BP 神经网络)温度自补偿模块向其他计算机系统移植复现的编程算式。

- 综合评价压力传感器配备了BPNN温度自补偿模块后的性能。

解：构建一个具有温度自补偿功能的二传感器数据融合智能传感器系统，其中一个是辅助传感器，用来监测干扰量温度 T ；另一个是被补偿的主传感器——压力传感器。系统如图 6-13 所示。

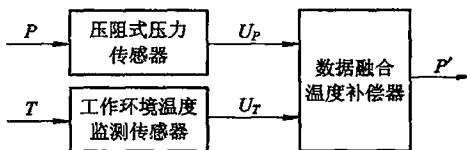


图 6-13 二传感器数据融合智能传感器系统

1. 二维标定实验

在工作温度范围内选定多个不同的温度状态对被补偿的压力传感器进行标定实验，表 6-1 列出了在 $21.5^{\circ}\text{C} \sim 70^{\circ}\text{C}$ 间 6 个温度状态的静态标定数据，在压力传感器量程范围内选了 6 个标定值，总计有 36 个标定点，可获得 6 个温度状态对应的 6 条压力传感器的输入 (P) 输出 (U_P) 静态特性，也称这些标定值为样本数据。

从表 6-1 的标定值可以看出，在输入压力值不变的情况下，工作环境温度不同，压力传感器的输出电压值 U_P 也随之改变。

表 6-1 不同温度状态下的压力传感器静态标定数据——二维实验标定数据

压力 / $\times 10^4 \text{ Pa}$		0	1.0	2.0	3.0	4.0	5.0
温度 / $^{\circ}\text{C}$		$U_P, U_T / \text{mV}$					
21.5	U_P	-13.84	10.69	28.88	47.05	65.19	83.36
	U_T	27.64	26.95	26.43	25.92	25.45	24.94
28	U_P	-13.49	9.32	26.34	43.12	59.99	76.82
	U_T	34.41	33.93	33.47	32.93	32.47	31.91
34	U_P	-10.80	7.54	24.84	42.05	59.25	76.38
	U_T	37.76	36.92	36.44	35.97	35.39	35.09
44	U_P	-9.72	6.56	23.87	41.21	58.58	75.87
	U_T	54.88	53.97	52.87	52.41	51.93	51.55
50	U_P	-8.62	4.86	21.84	38.70	56.32	73.75
	U_T	65.77	64.79	63.84	62.91	61.99	61.06
70	U_P	-7.72	3.72	21.25	38.6	55.56	73.28
	U_T	86.12	84.94	83.78	82.65	81.55	80.45

2. 神经网络样本文件的制作

神经网络样本文件有两种：分别用以对神经网络进行训练或检验用的训练样本文件和

检验样本文件。

1) 样本数据对

由二维标定实验中的 36 个标定点，获得 36 组实验标定数据或者说获得 36 组样本数据对，如表 6-2 所示。样本数据对(以下简称样本对)由输入量与对应输出量两部分组成，表 6-2 中每个序号的样本对含有三个输入量： U_P 、 U_T 与 T ，对应期望输出量为 P 。

表 6-2 用于消除压力传感器对温度交叉敏感的神经网络样本数据对

序号	1	2	3	...	33	34	35	36
U_P/mV	-13.84	10.69	28.88	...	21.25	38.6	55.56	73.28
U_T/mV	27.64	26.95	26.43	...	83.78	82.65	81.55	80.45
$T/^\circ C$	21.5	21.5	21.5	...	70	70	70	70
$P/10^4 Pa$	0	1.0	2.0	...	2.0	3.0	4.0	5.0

2) 样本文件的制作

通常，将标定实验获得的样本数据对总数(本实验的样本对总数为 36)中的 1/2~2/3 用做神经网络的训练形成训练样本文件，形成网络结构及权值；再用余下的 1/2~1/3 的样本数据对形成检验样本文件，进行神经网络的检验。

(1) 训练样本文件。训练样本文件由 b1.txt 与 b2.txt 两个文件组成，装入了表 6-2 中除了序号为 1、8、15、18、19、22、29、36 八组样本对以外的其余 28 组样本对，形成 28 列 3 行训练用输入样本文件 b1.txt，装载的是神经网络的输入数据 U_T 与 U_P ；b2.txt 文件有一行 28 列数据，是与输入样本数据相对应的神经网络的期望输出值 P 。两个文件的每列是一组样本数据对。文件中每行的样本数据要求等间隔排列，以便程序能够正确读取样本文件数据：

```
b1.txt: 21.5 21.5 21.5 21.5 21.5 28 28 28 ... 70 70 70 70  
10.69 28.88 47.05 65.19 83.36 -13.49 26.34 43.12 ... 3.72 21.25 38.6 55.56  
26.95 26.43 25.92 25.45 24.94 34.41 33.47 32.93 ... 84.94 83.78 82.65 81.55  
b2.txt: 1.0 2.0 3.0 4.0 5.0 0.0 2.0 3.0 ... 1.0 2.0 3.0 4.0
```

(2) 检验样本文件。检验样本文件由 b3.txt 与 b4.txt 两个文件组成，装载了表 6-2 中序号为 1、8、15、18、19、22、29、36 共计 $M=8$ 个样本数据对：

```
b3.txt: 21.5 28 34 34 44 44 50 70  
-13.84 9.32 24.84 76.38 -9.72 41.21 56.32 73.28  
27.64 33.93 36.44 35.09 54.88 52.41 61.99 80.45  
b4.txt: 0.0 1.0 2.0 5.0 0.0 3.0 4.0 5.0
```

3. 样本文件数据的归一化处理

经过归一化处理后的数据在一 $1\sim+1$ 或者 $0\sim1$ 之间，这样可以有更好的数据融合效果。数据归一化的公式有多种，如：

$$\bar{X} = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \quad (6-33)$$

或
$$\bar{X} = \frac{0.9(X - X_{\min})}{X_{\max} - X_{\min}} + 0.05 \quad (6-34)$$

式中 X 、 \bar{X} 分别为归一化前后的样本数据， X_{\min} 、 X_{\max} 为 X 所在行的最小、最大值。本例采

用公式(6-33)。

当采用归一化样本数据用于神经网络的训练与检验，并且将训练好的网络移植到单片机中进行测量时，测量用的样本必须进行相同的归一化处理。如果训练检验的样本没有进行归一化，则测量样本也不用进行归一化处理。

4. 在 Matlab 环境下的 BPNN 设计流程

在 Matlab 环境下的 BPNN 设计流程如图 6-14 所示。

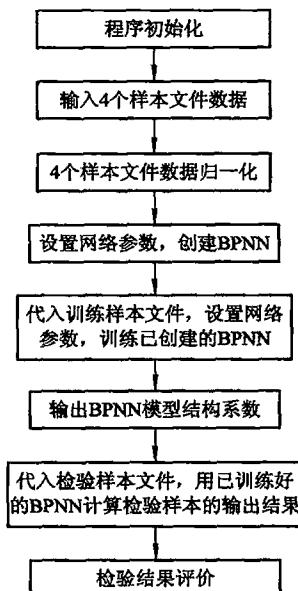


图 6-14 在 Matlab 环境下 BPNN 设计流程图

按照上述流程可设计出以下程序供参考：

(1) 程序初始化。用函数 clear all 初始化程序，将所有变量、函数清零，举例如下：

```
Clear all;
```

(2) 输入 4 个样本文件数据。用函数 load 输入样本文件数据，程序举例如下，程序中%为注释内容的开始标记，注释内容不参与程序的运行：

```
%调入样本文件
FID1=load('b1.txt'); Xuexiyangben=FID1'; %输入学习输入样本
FID2=load('b2.txt'); XueDesire=FID2'; %输入学习输出样本
FID3=load('b3.txt'); Ceshiyangben=FID3'; %输入检验输入样本
FID4=load('b4.txt'); Ceshiqiwang=FID4'; %输入检验输出样本
```

(3) 4 个样本文件数据归一化。分别用函数 min 和 max 寻找样本文件中每一行数据的最小值和最大值，然后编程实现不同变量所在的每行数据的归一化，程序举例如下：

```
p1=Xuexiyangben'; %将学习输入样本赋给数组 p1;
minp1=min(p1); %取 p1 中的最小值
maxp1=max(p1); %取 p1 中的最大值
```

```
%下面进行学习样本 3 行 28 列数据的归一化处理
for j=1:28
    for i=1:3
        xueguiyi(i, j)=(p1(i, j)-minp1(j))/(maxp1(j)-minp1(j));
    end
end
p2=Ceshiyangben';          %将检验输入样本赋给数组 p2;
minp2=min(p2);             %取 p1 中的最小值
maxp2=max(p2);             %取 p1 中的最大值
%下面进行检验样本 3 行 8 列数据的归一化处理
for j=1:8
    for i=1:3
        testguiyi(i, j)=(p2(i, j)-minp2(j))/(maxp2(j)-minp2(j));
    end
end
```

(4) 设置网络参数, 创建 BPNN。用函数 newff 创建 BPNN:

```
net=newff(PR, [S1, S2, …, SN1], {TF1, TF2, …, TFN1}, BTF, BLF, PF)
```

参数说明如下:

① PR——R 个输入的最小、最大值构成的 $R \times 2$ 矩阵, 存放训练输入样本文件每行的最大与最小值, 本示例为 3×2 矩阵。

② Si——SNI 层网络第 i 层的神经元个数, 输入层神经元个数自动获取, 不在这里设置。本示例有隐层和输出层两层神经元, 每层神经元个数分别为 6 和 1。

③ TFi——第 i 层的传递函数, 可以是任意可导函数, 默认为“tansig”, 可设置为 logsig, purelin 等; 输入层的输入信号直接输入隐层的输入端; 本示例分别采用 logsig 和 purelin 函数作为隐层和输出层的传递函数。

④ BTF——反向传播网络训练函数, 默认为“trainlm”, 可设置为 trainbfg, trainrp, traingd 等; 本示例采用默认函数 trainlm。

⑤ BLF——反向传播权值、阈值学习函数, 默认为“learngdm”; 本示例采用默认函数 learngdm。

⑥ PF——功能函数, 即训练样本训练结果的目标值, 默认为“mse”, 即训练样本训练结果的均方差值; 本示例采用默认函数 mse。

使用该函数首先要设置隐层与输出层神经元数量以及相应的传递函数, 确定训练函数、学习函数及功能函数或采用默认函数。其中权值与阈值的默认函数给权值、阈值、学习因子、势态因子自动赋初值, 例如权值初始值为小于 1 的随机数, 学习因子及势态因子的默认值分别为 0.01 和 0.9; 并按照规定的功能函数对训练结果进行评价, 默认功能函数为均方差函数。程序举例如下:

```
net=newff(minmax(p1), [6, 1], {'logsig', 'purelin'});
```

其中, minmax(p1)用来获得训练输入样本文件每行的最大与最小值。 $[6, 1]$ 用来设置 BPNN 隐层与输出层神经元节点数, 隐层节点数为 6, 输出层节点数为 1。输入层节点数与训练输入样本文件的行数相同, 自动获取。 $\{\text{'logsig'}, \text{'purelin'}\}$ 用来设置隐层与输出层的

传递函数，隐层传递函数采用 logsig 函数，输出层采用 purelin 纯线性函数。输入层传递函数默认为纯线性函数，即输出等于输入。在 newff 中没有出现的参数均采用默认函数或默认值。

(5) 代入训练样本文件，设置网络参数，训练已创建的 BPNN。用函数 train 训练 BPNN：

```
[net, tr] = train(net, P, T, Pi, Ai, VV, TV)
```

train 函数根据 net.trainFcn 和 net.trainParam 训练网络 net。

输入参数：

- ① P——网络的输入，即训练样本文件数据。
- ② T——网络训练的目标值，默认值为 0；通常是训练结果均方差值的期望值。
- ③ Pi——初始输入延迟，可选项，只在有输入延迟时需要，一般不用，默认值为 0。
- ④ Ai——初始层延迟，可选项，只在有层延迟时需要，默认值为 0。
- ⑤ VV——确认向量结构，默认为空矩阵。
- ⑥ TV——测试向量结构，默认为空矩阵。

输出参数：

- ① net——训练好返回的网络结构。
- ② tr——网络训练步数和性能。

train 函数的信号格式为阵列或矩阵。在训练网络过程中进行网络测试，确认向量用来及时终止训练，以免过训练损害网络的泛化能力。本示例中采用的程序如下：

```
%bp 神经网络训练
net.trainParam.epochs=1000; %训练次数为 1000
net.trainParam.goal=0; %训练目标误差为 0
[net, tr]=train(net, xueguiyi, XueDesire); %进行神经网络训练
```

(6) 输出 BPNN 模型结构参数。用函数 net.IW、net.LW 分别获取输入层与隐层间的权值及隐层与输出层间的权值，用 net.b 可分别获取隐层与输出层的阈值。程序举例如下：

```
%以下显示权值和阈值
iw1=net.IW{1} %显示输入层与隐层间的权值
b1=net.b{1}' %显示隐层的阈值
lw2=net.LW{2}' %显示隐层与输出层间的权值
b2=net.b{2} %显示输出层的阈值
```

(7) 代入检验样本文件，用已训练好的 BPNN 计算检验样本的输出结果。用函数 sim 检验已经训练好的 BPNN，计算检验样本的输出结果。

函数功能：神经网络仿真函数。只有在创建好一个网络后才能进行网络仿真。

调用格式：

```
[T, X, Y] = sim('model', TIMESPAN, OPTIONS, UT)
[T, X, Y1, ..., Yn] = sim('model', TIMESPAN, OPTIONS, UT)
```

输出参数：

- ① T——返回的时间步长向量；
- ② X——返回的矩阵状态或结构格式状态；
- ③ Y——返回的输出，即检验样本的输出结果；

④ Y_1, \dots, Y_n ——只能用于框图模式, 这里 n 为根输出块的个数。

输入参数:

① model——框图模型名称, 即仿真的网络名, 它必须在进行网络仿真前已经创建好;

② TIMESSPAN——训练所需时间范围, 写成[TStart TFinal]格式或[TStart OutputTimes TFinal]格式, 其中 TStart 为起始时间, TFinal 为终止时间, OutputTimes 为时间步长 T;

③ OPTIONS——可选仿真参数, 这是由 SIMSET 创建的一个结构体, 一般不用;

④ UT——可选外部输入, $UT = [T, U_1, \dots, U_n]$, 或者是一个 $u=UT(t)$ 的函数表达式, 其中 $T = [t_1, \dots, t_m]'$ 。

本示例中的程序举例如下:

```
% 测试数据
Temp=sim(net, testguiyi)      % 利用训练好的神经网络测试数据
e=Ceshiqiwang' - Temp        % 求出测试数据的测试偏差
```

(8) 检验结果评价。计算检验结果的绝对偏差或者引用误差, 判断结果能否满足实用要求。程序如下:

```
% 输出测试均方差
perf=mse(e, net)            % 输出测试数据的均方差
Ceshiqiwang1=Ceshiqiwang';  % 输出测试期望
```

5. 训练检验

(1) 条件: 输入节点数 3, 隐层节点数 6, 输出层节点数 1, 训练迭代次数 1000。

(2) 训练与测试: 运行编好的程序, 程序将自动输入网络训练与检验所需要的 4 个样本文件, 并进行 1000 次训练与检验。1000 次训练检验结束后, 检验样本的网络输出如表 6-3 所示, 通过与期望输出的对比, 可以看到检验结果与期望输出结果的误差很小。其中, 训练样本的均方差为 1.2640×10^{-4} , 测试样本的均方差为 9.7957×10^{-4} 。

表 6-3 检验样本期望输出和实际输出对照表

期望输出 $/\times 10^4$ Pa	0.0	1.0	2.0	5.0	0.0	3.0	4.0	5.0
检验结果 $/\times 10^4$ Pa	-0.035 158	1.001 453	2.014 487	4.946 656	0.045 219	3.015 784	3.989 751	5.033 822
绝对偏差 $/\times 10^4$ Pa	0.035158	0.001453	0.014487	0.053344	0.045219	0.015784	0.010249	0.033822

(3) 训练检验好的 BPNN 模型结构参数。

- 输入层与隐层之间的 3×6 个连接权值:

```
4.940832, -0.061951, -3.447318,
0.14874, 1.685793, 0.589081,
-0.392504, 0.02056, 0.263814,
-0.005577, 0.004517, 0.004535,
-0.423756, 0.024349, 0.278386,
-0.947187, 0.432573, -1.487912
```

- 隐层的 6 个阈值：

-21.780019, -8.052591, 6.290255, -0.392, 5.82355, 1.377625

- 隐层与输出层的 6 个连接权值：

0.451025, -0.286421, -57.773565, 50.896771, 23.380664, 0.375225

- 输出层的 1 个阈值：

14.476775

上述模型参数可移植到其它软件环境、DSP、单片机或者其它 PC 机中，消除温度的影响，实现压力传感器对压力值的准确测量。

6. 采用 BPNN 消除压力传感器交叉敏感的效果评价

(1) BPNN 融合后压力传感器的静态特性。将表 6-2 中 36 组样本数据同时作为训练样本与检验样本进行 BPNN 融合，融合后的压力值 P' 如表 6-4 所示。

表 6-4 融合后压力传感器在不同温度下输出的压力值 P' 单位： $\times 10^4$ Pa

P'	0	1.0	2.0	3.0	4.0	5.0
21.5°C	-0.000 281	1.001 555	1.998 981	3.001 02	3.997 072	5.001 683
28°C	0.000 304	0.999 882	1.999 519	2.999 735	4.000 652	4.999 89
34°C	0.000 052	0.998 146	1.999 523	3.003 013	4.002 921	4.996 294
44°C	-0.000 106	1.001 133	1.998 724	2.998 367	4.002 073	5.000 002
50°C	0.000 047	0.999 997	2.000 255	2.997 016	4.005 147	4.997 356
70°C	0.000 014	0.999 889	1.998 828	3.007 456	3.989 324	5.004 545

(2) 最小二乘法线性度的计算。

- 融合前：最小二乘法直线拟合方程为

$$P = 0.0587 \times u + 0.6287$$

将 u 代入直线拟合方程得到各个 P 值，并与期望输出相比较，可得出最大拟合偏差 $\Delta P_{\max} = 0.1843 \times 10^4$ Pa，则最小二乘法线性度为

$$\delta_L = \frac{|\Delta P_{\max}|}{P_{\max}} \times 100\% = \frac{0.1843}{5} \times 100\% = 3.7\%$$

- 融合后：最小二乘法直线拟合方程为

$$P' = 0.9970 \times P + 0.0091$$

当 P 为 0 时，有最大拟合偏差

$$\Delta P_{\max} = |0.0091 - 0| = 0.0091 \times 10^4$$
 Pa

则最小二乘法线性度为

$$\delta_L = \frac{|\Delta P_{\max}|}{P_{\max}} \times 100\% = \frac{0.0091}{5} \times 100\% = 0.18\%$$

(3) 温度稳定性参数的计算。

- 融合前：

零位温度系数

$$\alpha_0 = \frac{|-13.84 - (-7.72)|}{(70 - 21.5) \times 83.36} = 1.51 \times 10^{-3} \text{ } (^{\circ}\text{C})$$

温度灵敏度系数

$$\alpha_t = \frac{|73.28 - 83.36|}{(70 - 21.5) \times 83.36} = 2.49 \times 10^{-3} \text{ } (^{\circ}\text{C})$$

- 融合后：

零位温度系数

$$\alpha_0 = \frac{0.000304}{(70 - 21.5) \times 5} = 1.25 \times 10^{-6} \text{ } (^{\circ}\text{C})$$

温度灵敏度系数

$$\alpha_t = \frac{|5 - 5.004545|}{(70 - 21.5) \times 5} = 1.87 \times 10^{-5} \text{ } (^{\circ}\text{C})$$

通过线性度及温度稳定性参数的计算，可以看到经过 BPNN 的数据融合，融合后的线性度从原来的 3.7% 减小到 0.18%，融合后的零位温度系数及温度灵敏度系数分别减小了 3 个数量级及 2 个数量级，表明融合后传感器的线性度及温度稳定性都得到了不同程度的改善。

6.3.2 [示例 6-2] 基于 RBF 神经网络法抗两个干扰量影响的智能化软件模块的设计

要求：采用 RBF 神经网络作为逆模型，改善压阻式压力传感器(JCY-201)的温度稳定性与恒流源供电电流的稳定性，并对改善前后的稳定性作出评价。

同第 5 章的多元线性回归法，监测干扰量温度 T 的温度传感器的输出电压为 U_T ，监测干扰量恒流源供电电流 I 的电流传感器的输出电压为 U_I ，构建基于 RBF 神经网络的三传感器数据融合智能传感器系统框图如图 5-2 所示。

1. 三维标定实验

为了训练 RBF 神经网络，确定网络的各权值和偏置的具体数值，使得网络的输出值与标定值之差能够达到误差精度的要求且具有一定的推广能力，各标定点的数量应能满足神经网络训练的要求。为了全量程范围内全面检验融合后稳定性的改善效果，实际上共标定 $m \times n \times w = 360$ 个标定点，选其中 7 组不同温度的标定点(共 252 组数据)作为神经网络的学习(训练)样本，其余的 3 组不同温度的标定点(共 108 组数据)作为网络的检验样本。

学习样本的各物理量取值如下：

物理量	各物理量的标定点的数值						标定点个数
压力 P/MPa	0 0.1 0.2 0.3 0.4 0.5						$m=6$
温度 $T/\text{ }^{\circ}\text{C}$	25 31 35 39 49 55 64						$n=7$
电流 I/mA	6 7 8 9 10 11						$w=6$

学习样本各标定点的具体数值见表 6-5(只列出部分数据，下表同)。

检验样本的各物理量取值如下：

物理量	各物理量的标定点的数值					标定点个数
压力 P/MPa	0 0.1 0.2 0.3 0.4 0.5					$m=6$
温度 $T/^\circ\text{C}$	43 53 58					$n=3$
电流 I/mA	6 7 8 9 10 11					$w=6$

检验样本各标定点的具体数值见表 6-6。

表 6-5 RBF 神经网络用 JCY-201 型压力传感器三维标定实验数据——学习样本

序号	P /MPa	I /mA	T /°C	U_I /V	U_T /V	U_P /V	序号	P /MPa	I /mA	T /°C	U_I /V	U_T /V	U_P /V
1	0	6.0	25.0	0.6	5.62	0.325	127	0	9.0	39.0	0.9	8.61	0.334
2	0.1	6.0	25.0	0.6	5.62	0.834	128	0.1	9.0	39.0	0.9	8.61	1.083
3	0.2	6.0	25.0	0.6	5.62	1.329	129	0.2	9.0	39.0	0.9	8.61	1.825
4	0.3	6.0	25.0	0.6	5.62	1.825	130	0.3	9.0	39.0	0.9	8.61	2.496
5	0.4	6.0	25.0	0.6	5.62	2.232	131	0.4	9.0	39.0	0.9	8.61	3.248
6	0.5	6.0	25.0	0.6	5.62	2.819	132	0.5	9.0	39.0	0.9	8.61	3.985
7	0	7.0	25.0	0.7	6.56	0.321	133	0	10.0	39.0	1.0	9.57	0.332
8	0.1	7.0	25.0	0.7	6.56	0.925	134	0.1	10.0	39.0	1.0	9.57	1.169
9	0.2	7.0	25.0	0.7	6.56	1.503	135	0.2	10.0	39.0	1.0	9.57	1.994
10	0.3	7.0	25.0	0.7	6.56	2.08	136	0.3	10.0	39.0	1.0	9.57	2.821
11	0.4	7.0	25.0	0.7	6.56	2.66	137	0.4	10.0	39.0	1.0	9.57	3.65
12	0.5	7.0	25.0	0.7	6.56	3.24	138	0.5	10.0	39.0	1.0	9.57	4.402
13	0	8.0	25.0	0.8	7.51	0.317	139	0	11.0	39.0	1.1	10.54	0.33
14	0.1	8.0	25.0	0.8	7.51	1.017	140	0.1	11.0	39.0	1.1	10.54	1.255
15	0.2	8.0	25.0	0.8	7.51	1.677	141	0.2	11.0	39.0	1.1	10.54	2.163
16	0.3	8.0	25.0	0.8	7.51	2.338	142	0.3	11.0	39.0	1.1	10.54	3.073
:	:	:	:	:	:	:	:	:	:	:	:	:	:
123	0.2	8.0	39.0	0.8	7.65	1.656	249	0.2	11.0	64.0	1.1	10.92	1.978
124	0.3	8.0	39.0	0.8	7.65	2.319	250	0.3	11.0	64.0	1.1	10.92	2.889
125	0.4	8.0	39.0	0.8	7.65	2.981	251	0.4	11.0	64.0	1.1	10.92	3.802
126	0.5	8.0	39.0	0.8	7.65	3.645	252	0.5	11.0	64.0	1.1	10.92	4.718

注： U_I 为电流传感器输出电压， U_T 为温度传感器输出电压， U_P 为压力传感器输出电压，下表同。

表 6-6 RBF 神经网络用 JCY-201 型压力传感器三维标定实验数据——检验样本

序号	P /MPa	I /mA	T /°C	U_I /V	U_T /V	U_P /V	序号	P /MPa	I /mA	T /°C	U_I /V	U_T /V	U_P /V
1	0	6.0	43.0	0.6	5.76	0.16	55	0	9.0	53.0	0.9	8.79	0.213
2	0.1	6.0	43.0	0.6	5.76	0.762	56	0.1	9.0	53.0	0.9	8.79	0.97
3	0.2	6.0	43.0	0.6	5.76	1.258	57	0.2	9.0	53.0	0.9	8.79	1.713
4	0.3	6.0	43.0	0.6	5.76	1.755	58	0.3	9.0	53.0	0.9	8.79	2.457
5	0.4	6.0	43.0	0.6	5.76	2.251	59	0.4	9.0	53.0	0.9	8.79	3.203
6	0.5	6.0	43.0	0.6	5.76	2.75	60	0.5	9.0	53.0	0.9	8.79	3.95
7	0	7.0	43.0	0.7	6.73	0.145	61	0	10.0	53.0	1.0	9.78	0.21
8	0.1	7.0	43.0	0.7	6.73	0.842	62	0.1	10.0	53.0	1.0	9.78	1.048
9	0.2	7.0	43.0	0.7	6.73	1.42	63	0.2	10.0	53.0	1.0	9.78	1.872
10	0.3	7.0	43.0	0.7	6.73	2.0	64	0.3	10.0	53.0	1.0	9.78	2.699
11	0.4	7.0	43.0	0.7	6.73	2.581	65	0.4	10.0	53.0	1.0	9.78	3.527
12	0.5	7.0	43.0	0.7	6.73	3.162	66	0.5	10.0	53.0	1.0	9.78	4.358
13	0	8.0	43.0	0.8	7.69	0.12	67	0	11.0	53.0	1.1	10.74	0.214
14	0.1	8.0	43.0	0.8	7.69	0.918	68	0.1	11.0	53.0	1.1	10.74	1.13
15	0.2	8.0	43.0	0.8	7.69	1.58	69	0.2	11.0	53.0	1.1	10.74	2.036
16	0.3	8.0	43.0	0.8	7.69	2.242	70	0.3	11.0	53.0	1.1	10.74	2.944
:	:	:	:	:	:	:	:	:	:	:	:	:	:
51	0.2	8.0	53.0	0.8	7.8	1.555	105	0.2	11.0	58.0	1.1	10.82	1.992
52	0.3	8.0	53.0	0.8	7.8	2.217	106	0.3	11.0	58.0	1.1	10.82	2.904
53	0.4	8.0	53.0	0.8	7.8	2.879	107	0.4	11.0	58.0	1.1	10.82	3.817
54	0.5	8.0	53.0	0.8	7.8	3.544	108	0.5	11.0	58.0	1.1	10.82	4.731

2. RBF 神经网络 Matlab 源程序

RBF 神经网络 Matlab 源程序如下：

```

clc
clear
close all %程序开始前, 清屏, 清除其它变量
%输入原始数据, 每一列为一个变量, 行数为样本的个数。原始标定数据的输入, 可以将数据
直接输入至程序中, 也可以将数据先输入至文件, 然后在程序中读入数据文件。数据中的 P、
 $U_I$ 、 $U_T$ 、 $U_P$  均为  $n \times 1$  维列向量,  $n$  为标定数据点的个数
raw_data=[P Ut Ui Up]; %学习(训练)样本
data_test=[P_test Ut_test Ui_test Up_test]; %测试样本
%根据原始数据制作学习(训练)样本与测试样本, 学习(训练)样本与测试样本的数据格式相同
P=raw_data(:, 1); Ui=raw_data(:, 2); Ut=raw_data(:, 3); Up=raw_data(:, 4);

```

```

P_test= data_test(:, 1); Ui= data_test (:, 2); Ut= data_test (:, 3); Up= data_test (:,
4);
P1=[Up Ui Ut]'; %学习样本输入
T1=P'; %学习样本期望输出
P2=[Up_test Ui_test Ut_test]'; %测试样本输入
T2=P_test'; %测试样本输出
%将输入数据进行归一化
[PN1, minp, maxp, TN1, mint, maxt] = premnmx(P1, T1);
PN2 = tramnmx(P2, minp, maxp);
TN2 = tramnmx(T2, mint, maxt);

%训练
switch 2
    case 1 %神经元数是训练样本个数
        spread = 0.1; %此值越大，覆盖的函数值就越大(默认为 1)
        net = newrbe(PN1, TN1, spread);

    case 2 %神经元数逐步增加，最多就是训练样本个数
        goal = 1e-8; %训练误差的平方和(默认为 0)
        spread = 0.1; %此值越大，需要的神经元就越少(默认为 1)
        MN = size(PN1, 2); %最大神经元数(默认为训练样本个数)
        DF = 2; %显示间隔(默认为 25)
        net = newrb(PN1, TN1, goal, spread, MN, DF);

    case 3 %此值越大，需要的神经元就越少(默认为 1)
        spread = 0.1;
        net = newgrnn(PN1, TN1, spread);
end
%测试
YN1 = sim(net, PN1); %学习(训练)样本实际输出
YN2 = sim(net, PN2); %测试样本实际输出

MSE1 = mean((TN1-YN1).^2) %学习(训练)样本实际输出的均方误差
%n1=length(TN1); % e1=sum((TN1-YN1).^2)/n1 %e1 与 MSE1 相等
MSETD1 = std((TN1-YN1).^2)

MSE2 = mean((TN2-YN2).^2) %测试样本实际输出均方误差
%n2=length(TN2); % e2=sum((TN2-YN2).^2)/n2 %e2 与 MSE2 相等
MSETD2 = std((TN2-YN2).^2)
%反归一化
Y1 = postmnmx(YN1, mint, maxt)' %将归一化的学习(训练)输出数据恢复为原始数据
E1=Y1-P %学习(训练)输出数据与标定值之间的偏差
Y2 = postmnmx(YN2, mint, maxt)' %将归一化的测试输出数据恢复为原始数据

```

```

E2=Y2-P_test           %测试输出数据与标定值之间的偏差
max(E2)
min(E2)

```

3. 采用经过数据归一化处理的程序

神经网络采用 Matlab 中的 newrb() 函数, neurons = 252。学习样本经 RBF 神经网络训练后的输出值与相应的压力标定值之间的最大偏差绝对值为 $1.0396e-011$, 均方误差 $MSE1=5.6613e-023$, 均方误差的标准差 $MSETD1=2.4666e-022$ 。这里省略学习样本的预测输出表。可见, 采用 RBF 神经网络对学习样本的输出已经消除了压力传感器的交叉敏感的影响(在误差允许的范围内)。

检验样本经 RBF 神经网络后的输出值与相应的压力标定值之间最大偏差绝对值为 $\max=0.0113$, 均方误差 $MSE2=3.8568e-005$, 均方误差的标准差 $MSETD2=2.1727e-004$ 。输出样本的详细情况见表 6-7(只列出部分数据)。

表 6-7 输出样本

T/°C	I/mA	I=6		I=7		I=8		I=10		I=11	
		P	\hat{P}	Δ	\hat{P}	Δ	\hat{P}	Δ	\hat{P}	Δ	\hat{P}
43°C	0	0.0040	0.0040	0.0033	0.0033	0.0014	0.0014	0.0013	0.0013	0.0018	0.0018
	0.5	0.5001	0.0001	0.4999	-0.0001	0.5005	0.0005	0.4999	-0.0001	0.5002	0.0002
53°C	0	-0.0003	-0.0003	-0.0006	-0.0006	-0.0002	-0.0002	0.0004	0.0004	0.0015	0.0015
	0.5	0.5002	0.0002	0.5001	0.0001	0.5000	-0.0000	0.5002	0.0002	0.5001	0.0001
58°C	0	-0.0113	-0.0113	-0.0075	-0.0075	-0.0041	-0.0041	-0.0009	-0.0009	-0.0023	-0.0023
	0.5	0.5000	-0.0000	0.5003	0.0003	0.4993	-0.0007	0.4999	-0.0001	0.4997	-0.0003

注: 表中 T 表示工作温度, I 表示工作电流, P 表示压力标定值, \hat{P} 表示压力计算值, 偏差 $\Delta=\hat{P}-P$, 单位都是 V。

4. 数据融合融合效果评价

由于学习(训练)样本的网络输出值与标准值之间的偏差很小(均小于 $1.0e-011$), 如果根据这些数据来计算融合后的效果, 可以认为已经完全消除交叉敏感。

通过表 6-7 可知, 本例中检验样本的网络输出值与标准值之间的偏差与学习样本的偏差相比, 偏差较大, 故按检验样本计算数值来评价数据融合效果。

利用检验样本计算的融合后的零位温度系数、灵敏度温度系数以及电流影响系数如下:

在 $\Delta T=39^\circ\text{C}$, $\Delta I=5 \text{ mA}$ 变化范围内, 零点融合计算值的最大偏差 $|\Delta P'_{0m}|=0.0113 \text{ MPa}$; 满量程压力 $P_{fs}=0.5 \text{ MPa}$, 其融合计算值的最大偏差量 $|\Delta P'_m|=0.0007 \text{ MPa}$, 则有

$$\alpha_0 = \frac{|\Delta P'_{0m}|}{P_{fs} \times \Delta T} = \frac{0.0113}{0.5 \times 39} = 5.79 \times 10^{-4} \quad (\text{}/^\circ\text{C})$$

$$\alpha_s = \frac{|\Delta P'_m|}{P_{fs} \times \Delta T} = \frac{0.0007}{0.5 \times 39} = 3.59 \times 10^{-5} \quad (\text{}/^\circ\text{C})$$

$$\alpha_I = \frac{|\Delta P'_m|}{P_{fs} \times \Delta I} = \frac{0.0007}{0.5 \times 5} = 2.8 \times 10^{-4} \quad (\text{/mA})$$

融合前传感器的零位温度系数、灵敏度温度系数以及电流影响系数见第 5 章[示例]

5-2]。将融合前、线性回归法融合后和 RBF 神经网络融合后的各参数列于表 6-8 中。

表 6-8 融合前、线性回归法融合后和 RBF 神经网络融合后的各参数

评价参数	零位温度系数 α_0 (/ $^{\circ}$ C)	灵敏度温度系数 α_s (/ $^{\circ}$ C)	电流影响系数 α_I (/mA)
融合前计算值	1.852×10^{-3}	1.154×10^{-2}	9.002×10^{-2}
融合后计算值(回归法)	1.246×10^{-3}	8.667×10^{-4}	6.760×10^{-3}
融合后计算值(RBF)	5.794×10^{-4}	3.590×10^{-5}	2.8×10^{-4}

从上表可见, 经 RBF 神经网络融合后, 零位温度系数降低到融合前的 1/3, 灵敏度温度系数降低约 3 个数量级, 电流影响系数降低约 2 个数量级。

参 考 文 献

- [1] 张永怀, 刘君华, 林继鹏. OGSA 结合 RBF 网络数据融合技术在红外气体分析中的应用[J]. 仪器仪表学报, 2005(1)
- [2] 邵军, 刘君华, 乔学光, 等. 利用 BP 神经网络提高光纤光栅压力传感器的选择性[J]. 传感技术学报, 2007, 20(7): 1531-1534
- [3] 张永怀, 刘君华. 基于神经网络技术提高红外气体分析器的选择性[J]. 西安交通大学学报, 2003, 37(8): 787-790
- [4] Zhang Y, Liu J H, Zhang Y H, et al. Cross sensitivity reduction of gas sensors using genetic neural network[J]. Optical Engineering, 2002, 41(3): 615-625
- [5] 张永怀, 刘君华. 采用 BP 神经网络及其改进算法改善传感器特性[J]. 传感器技术学报, 2002, 15(3): 185-188
- [6] 司端锋, 常炳国, 刘君华. 基于 BP 神经网络的传感器特性补偿新算法的研究[J]. 仪表技术与传感器. 2002(1): 59-61
- [7] 丁晖, 刘君华, 申忠如. 基于径向基函数神经网络预测器的多传感器故障在线检测和信号恢复的研究[J]. 模式识别与人工智能, 2000, 13(4): 429-433
- [8] 常炳国, 吴浩扬, 刘君华, 等. 基于径向基函数神经网络识别法识别变压器油中微量特征气体[J]. 西安交通大学学报, 1999, 33(2): 14-16
- [9] 常炳国, 刘君华. 基于神经网络的传感器特性补偿法[C]. STC99, 1999, Beijing: 427-430
- [10] 司端锋, 常炳国, 刘君华, 等. 采用计算机软件方法提高传感器的输出稳定性研究[C]. 第五届全国计算机应用联合学术会议论文集, 北京: 电子工业出版社, 1999

第7章 支持向量机技术在智能传感器系统中的应用

本章内容

- 关于统计学习理论与支持向量机的基础知识
- 支持向量机的训练、检验与测量
- [示例 7-1]降低两个干扰量影响的 SVM 智能化软件模块的设计
- [示例 7-2]使用 SVC(支持向量分类)对两组分混合气体进行定性识别

支持向量机(Support Vector Machines, SVM)技术已在很多领域得到广泛和成功的应用，近年来也已被引入智能传感器系统，用来实现传感器的智能化功能，成为继神经网络技术之后另一种更为有效的智能化技术手段。目前，SVM 技术已用于传感器的动态建模、故障诊断、气体辨识和交叉敏感消除等。

在 SVM 方法中，通过定义不同的内积函数，就可以实现多项式逼近、贝叶斯分类器、径向基函数(Radial Basic Function, RBF)、多层感知器等许多现有学习算法的功能。

本章以支持向量机建立消除传感器交叉敏感干扰量影响的智能化逆模型为例，着重介绍支持向量机实现传感器智能化功能的工作原理，以及多参量定量测量及定性识别中的使用方法。

在应用多传感器逆模型法(详见 4.3.3 节)抑制交叉敏感改善传感器稳定性的方法中，支持向量机技术与多元回归分析法不同的是，它不必建立包括待消除的非目标参量在内的函数解析式，这一点与神经网络方法相同。三者的相同点是它们的训练样本和检验样本都是由多维标定实验数据提供的，都是为了消除 k 个非目标参量的影响，需要检测这 k 个非目标参量，进行 $k+1$ 维标定实验。

7.1 统计学习理论与支持向量机的基础知识

7.1.1 统计学习理论

长期以来，构造可以从经验(例子或数据)中学习的机器是哲学和科技领域研究的目标之一。其中，基于数据的机器学习是现代智能技术领域的一个重要方面，研究的目的是从观测数据(样本或例子)出发寻找规律，并利用这些规律进行建模，用于对未来数据或无法观测的数据进行预测。

迄今为止，机器学习的实现方法大致可以分为三种。

第一种是经典的(参数)统计估计方法。现有机器学习方法共同的重要理论基础之一是统计学。参数方法的基础是传统的统计学，在这种方法中，参数的相关形式是已知的，训

练习样本用来估计参数的值。这种方法的局限性在于，首先，它需要已知样本的分布形式，这需要花费很大代价。其次，传统统计学是研究样本数目趋于无穷大时的渐近理论，即研究当样本数量趋于无穷大时的极限特性，现有学习方法有很多也基于此假设。但在实际问题中，样本数目往往是有限的，因此一些理论上很优秀的学习方法往往在实际中的表现却不尽人意。

第二种方法是经验非线性方法，如人工神经网络(Artificial Neural Network, ANN)。这种方法利用已知样本建立非线性模型，克服了传统参数估计方法的困难。但是，这种方法存在两个方面的问题：首先是到目前为止，缺乏一种统一的数学理论；其次，由于所建立系统模型的性能最终决定于观测数据(样本)的数量与质量，而在实际应用过程中所获得的原始观察数据(样本)数量往往是有限的，而且是采样(非连续)的，通常这些采样没有统一的格式，同时由于所研究问题本身的高维本质，将会在输入空间中仅仅形成一个稀疏分布，从而导致问题往往接近于 Hadamard 意义上的病态，造成传统神经网络方法的泛化(generalization)问题，会产生过拟合数据的模型。为了克服这一问题，机器学习产生了两个研究方向：一是参数选择的优化算法；二是选择“最优”模型的统计测量方法。

第三种方法是基于统计学习理论(Statistical Learning Theory, SLT)的学习方法，如支持向量机和基于核的方法。与传统统计学相比，统计学习理论是一种专门研究小样本数据量情况下机器学习规律的理论。该理论针对小样本统计问题建立了一套新的理论体系，在这种体系下的统计推理规则不仅考虑了对渐近性能的要求，而且追求在现有有限信息的条件下得到最优结果。其创立者 V. Vapnik 等人从上个世纪六七十年代开始致力于此方面的研究，到九十年代中期，随着其理论的不断发展和成熟，也由于神经网络等学习方法在理论上缺乏实质性进展，统计学习理论开始受到越来越广泛的重视。

统计学习理论研究的内容包括经典统计学中很多重要的课题，特别是判别分析(模式识别)、回归分析和密度估计等问题，可参阅 Vladimir N. Vapnik 著的《统计学习理论》(Statistical Learning Theory, 1998)和《统计学习理论的本质》(the Nature of Statistical Learning Theory)等著作。在这里我们对其中的一些概念和内容进行简要的介绍。

统计学习理论是一种研究小样本估计和预测的理论。它从理论上系统地研究了经验风险最小化原则成立的条件、有限样本下经验风险与期望风险的关系，以及如何利用这些理论找到新的学习原则和方法。统计学习理论研究的主要问题包括以下四个：

- 一个基于经验风险最小化(Empirical Risk Minimization, ERM)原则的学习过程一致性的充分必要条件是什么？
- 这个学习过程的收敛速度有多快？
- 如何控制这个学习过程的收敛速度(推广能力)？
- 如何构造能够控制推广能力的算法？

与之相对应的统计学习理论研究的基本内容包括：

- 学习过程一致性的理论——经验风险最小化准则下统计学习一致性的条件；
- 学习过程收敛速度的非渐近理论——在这些一致性条件下，关于统计学习方法推广性的界的理论；
- 控制学习过程推广能力的理论——在这些界的基本上，建立小样本归纳推理准则；
- 构造学习算法的理论——实现这些新的准则的设计方法(算法)。

1. 学习问题的一般性表示

在统计学习理论中，通常把学习问题看做是利用有限数量的观测来寻找待求的依赖关系的问题。学习问题可以一般地表示如下：设 $F(z)$ 为定义在空间 z 上的概率测度，所考虑的函数集合为 $Q(z, \alpha)$ ， $\alpha = [\alpha_1, \alpha_2, \dots]$ ， $\alpha \in \Lambda$ ， Λ 是参数集合，学习的目标是在给定了一定量但概率测度 $F(z)$ 未知的独立同分布数据：

$$z_1, z_2, \dots, z_l \quad (7-1)$$

的条件下，来最小化风险泛函：

$$R(\alpha) = \int Q(z, \alpha) dF(z), \quad \alpha \in \Lambda \quad (7-2)$$

其解记为 $Q(z, \alpha_0)$ 。

因此，学习的一般问题就是如何在经验数据式(7-1)的基础上最小化风险泛函式(7-2)，其中 z 代表数据对 (x, y) ， $Q(z, \alpha)$ 即特定的损失函数(与具体的问题有关)。

独立同分布数据也可以表述为由一个信源给定的 l 个观测量，其中每个观测量由一个数据对 (x_i, y_i) 组成： $x_i \in \mathbb{R}^d$ ($i=1, 2, \dots, l$) 和相对应的“真值(或期望值)” $y_i \in \mathbb{R}$ 。假定这些数据是从一些未知概率分布 $P(x, y)$ (P 表示概率分布， p 表示分布的密度) 中随机抽取的独立同分布(independently and identically distributed, i. i. d.) 数据。学习机器的任务就是学习 $x_i \rightarrow y_i$ 的映射关系。学习机器实际上就是一个可能的映射集合 $x \rightarrow f(x, \alpha)$ ，其中函数 $f(x, \alpha)$ 是由可调整的参数 α 标识的。当 α 选定后，称学习机器已经被“训练”了。当学习机器确定后，即对于给定的输入 x 和一组选定的 α ，它总是能够给出相同的输出 $f(x, \alpha)$ 。如对于一个神经网络， α 对应神经网络的权值和偏置值， α 确定意味着其权值和偏置值已经确定。

一个经过训练的机器，其测试误差的期望为

$$R(\alpha) = \int \frac{1}{2} |y - f(x, \alpha)| dP(x, y) \quad (7-3)$$

式中， $R(\alpha)$ 被称为期望风险(expected risk)， $P(x, y)$ 未知。若密度 $p(x, y)$ 存在， $dP(x, y)$ 可以写为 $p(x, y) dx dy$ 。然而对于概率分布 $P(x, y)$ 未知的情况，要使得期望风险最小化，就只有样本信息可以利用，这也就导致式(7-3)定义的期望风险是无法直接计算和最小化的。因此在传统的机器学习方法中，采用了经验风险最小化(Empirical Risk Minimization, ERM)的准则来逼近期望风险式(7-3)。

经验风险 $R_{\text{emp}}(\alpha)$ 的定义为在一个观测数量有限的训练集上的被测平均误差率：

$$R_{\text{emp}}(\alpha) = \frac{1}{2l} \sum_{i=1}^l |y_i - f(x_i, \alpha)| \quad (7-4)$$

从式(7-4)可以看出，在经验风险 $R_{\text{emp}}(\alpha)$ 的定义中没有概率分布，当给定具体的训练集 $\{(x_i, y_i)\}$ 和选定的 α 值后， $R_{\text{emp}}(\alpha)$ 则为一个固定的数。用经验风险最小化表示期望风险最小化并没有可靠的理论依据，二者之间事实上存在着一个学习的一致性问题。

式(7-3)中的量 $|y - f(x, \alpha)|/2$ 被称为损失，对于二值分类的问题来说，它只能取 0 (或 -1) 或 1 两个数。若选取某参数 η ， $0 \leq \eta \leq 1$ ，对于损失 $|y - f(x, \alpha)|/2$ ，下面的界以 $1 - \eta$ 概率成立：

$$R(\alpha) \leq R_{\text{emp}}(\alpha) + \sqrt{\frac{h[\ln(2l/h) + 1] - \ln(\eta/4)}{l}} \quad (7-5)$$

式中, h 是一个非负整数, 称为 Vapnik Chervonenkis(VC) 维, 它表征学习机器的复杂性, h 越大, 学习机器的复杂性越高, 能区分的类别越多, 相应地, 出现所谓的“过拟合”的可能性也越高。式(7-5)右边被称为“风险界”(risk bound), 右边第二项被称为“VC 置信范围”(VC confidence)。置信范围反映了真实风险和经验风险差值的上界, 即结构的复杂性所带来的风险, 它与学习机器的 VC 维 h 及训练样本数量 l 有关。因此, 在传统的机器学习方法中所普遍采用的经验风险最小化原则在样本数据有限的情况下是不合理的。如果要求学习的风险最小, 就需要权衡式(7-5)右边的两项, 使其共同趋于极小。在获得学习模型经验风险最小的同时, 希望其推广能力尽可能大, 这就需要 h 尽可能小。

2. VC 维

VC 维是一个标量值, 用于测量函数列的容量, 其概念是建立在点集被“打散(shatter)”的概念基础上的。一函数集的 VC 维是 h , 当且仅当存在一点列 $\{x^i\}_{i=1}^h$, 这些点列能共以 2^h 种方式可分, 且当 $q > h$ 时, 没有序列 $\{x^i\}_{i=1}^q$ 满足这个性质。即

$$\text{VC dim}(F) = \max\{h : N(F, h) = 2^h\}$$

式中, $N(F, h)$ 定义为 h 维向量中不同向量的个数。

图 7-1 显示平面中的三个点是如何被线性指示函数集分开的, 然而四个点的情况则不能区分。对于这两种情况, VC 维与自由参数的个数相等, 但是在一般情况下, VC 维与自由参数的个数不等, 如函数 $A \sin(bx)$ 只有两个参数, 但有无限的 VC 维(Vapnik, 1995)。 n 维空间中的线性指示函数序列其 VC 维等于 $n+1$ 。

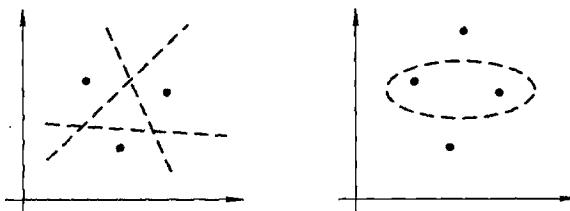


图 7-1 VC 维示意图

3. 结构风险最小化(Structural Risk Minimization, SRM)

根据式(7-5), 如果固定训练样本数目 l 的大小, 那么控制风险 $R(\alpha)$ 的参量有两个: $R_{\text{emp}}(\alpha)$ 和 VC 维 h 。其中, 经验风险依赖于学习机器所选定的函数 $f(x, \alpha)$, 这样就可以通过控制 α 来控制经验风险。而 VC 维 h 依赖于学习机器在学习过程中可选函数的集合。

为了获得对 h 的控制, 可以将函数集合结构化, 建立 h 与各函数子结构之间的关系, 通过控制对函数结构的选择来达到控制函数 VC 维 h 的目的。

如图 7-2 所示, 设函数 $Q(z, \alpha)$, $\alpha \in \Lambda$ 的集合 S 具有一定的结构, 这个结构是由一系列嵌套的函数子集 $S_k = \{Q(z, \alpha), \alpha \in \Lambda_k\}$ 组成的, 满足:

$$S_1 \subset S_2 \subset \cdots \subset S_n \subset \cdots \subset$$

其中, 结构的元素满足如下两个性质:

- 每个函数集 S_k 的 VC 维 h_k 是有限的, 且有

$$h_1 \leq h_2 \leq \cdots \leq h_n \leq \cdots$$

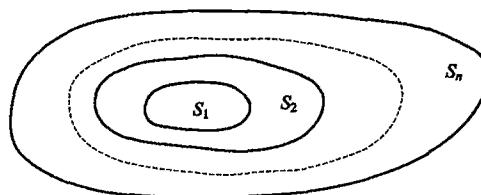


图 7-2 由嵌套的函数子集确定的函数集结构

• (结构中的任何元素 S_k 所包含性质)整个完全有界函数的集合满足:

$$0 \leq Q(z, \alpha) \leq B_k, \quad \alpha \in \Lambda_k$$

或者对一定的(p, τ_k), 函数集合满足下列不等式:

$$\sup_{\alpha \in \Lambda_k} \frac{\left(\int Q^p(z, \alpha) dF(z) \right)^{1/p}}{\int Q(z, \alpha) dF(z)} \leq \tau_k, \quad p > 2$$

称这种结构为容许(或可同伦)结构, 式中 \sup 为取上限运算符号。

结构风险最小化原则就是在容许结构的嵌套函数集 S_k 中寻找一个合适的子集 S^* , 使结构风险达到最小, 即:

$$\min_{S_k} R_{\text{emp}}[f] + \sqrt{\frac{h \ln\left(\frac{2l}{h} + 1\right) - \ln\left(\frac{\delta}{4}\right)}{l}} \quad (7-6)$$

由式(7-6)可知, 学习风险由两个部分组成, 即经验风险和置信范围。对于给定的样本数目 l , 随着 VC 维 h 的增加, 经验风险逐渐变小, 而置信范围逐渐递增。置信范围与经验风险随 h 的变化趋势如图 7-3 所示。

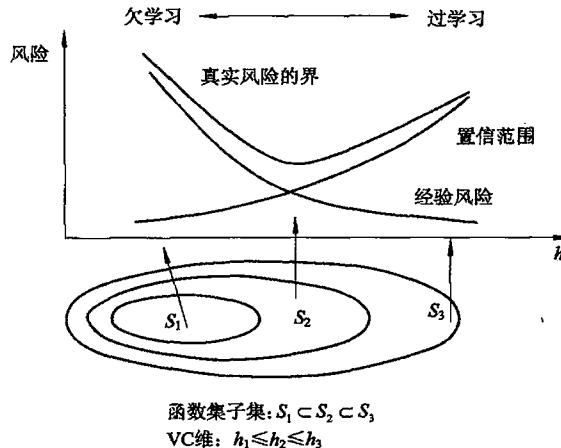


图 7-3 结构风险最小化原理图

图 7-3 表明, 真实风险的界是经验风险和置信范围之和。随着结构元素序号的增加, 经验风险将减少, 而置信范围将增加。最小的真实风险的上界是在结构的某个适当的元素

上取得的。综合考虑经验风险与置信范围的变化，可以求得最小的风险边界，它所对应的函数集的中间子集 S^* 可作为具有最佳泛化能力的函数集合。

综上所述，结构风险最小化就是根据函数集的性质将其划分为一系列的嵌套子集，学习问题就是根据推广能力选择其中最好的子集，并根据经验风险在子集中选择最好的函数。SRM 原则定义了在对给定数据逼近的准确性与逼近函数的复杂性之间的一种折中。

4. 建模误差

机器学习中建模的目标就是从假设模型空间中选择一个模型，使其能够最接近于（从一些误差测度的角度上说）目标空间中的基础函数（underlying function）。这种做法产生的误差主要源于两个方面：逼近误差和估计误差。

逼近误差的产生是由于假设空间小于目标空间，因此基础函数可能会位于假设空间之外。模型空间的选择不当将会产生很大的逼近误差，这种情况被称为模型失配。估计误差的产生是由于学习过程中从假设空间中选择非优化模型所造成的。这两个误差合在一起形成泛化误差。

7.1.2 支持向量机

统计学习理论建立在一套较坚实的数学理论基础之上，并为解决有限样本的学习问题提供了一个统一的框架。它能将很多现有方法纳入其中，有望帮助解决许多原来难以解决的问题（比如神经网络结构选择问题、局部极小点问题、过拟合问题等）；同时，在这一理论基础上还发展了一种新的通用学习方法——支持向量机（Support Vector Machine, SVM），其已初步表现出很多优于已有方法的性能，并推动着机器学习理论和技术的发展。

支持向量机是基于统计学习理论的一种新的通用机器学习方法，其基本思想是通过用内积函数定义的非线性变换将输入空间转换到一个高维特征空间（Feature Space），在这个高维空间中使用线性函数假设空间来寻找输入变量和输出变量之间的一种非线性关系。其学习训练是通过源于最优化理论的算法来实现由统计学习理论导出的学习偏置，它采用结构风险最小化原则，比采用经验风险最小化原则的神经网络有更好的泛化能力。

1. 支持向量机方法的主要优点

(1) 它是专门针对有限样本情况的，其目标是得到现有信息条件下的最优解而不仅仅是样本数趋于无穷大时的最优值。

(2) 支持向量机的算法最终将转化成为一个二次型寻优问题，从理论上说，得到的将是全局最优点，解决了在神经网络方法中无法避免的局部极值问题。

(3) 支持向量机的算法将实际问题通过非线性变换转换到高维的特征空间，在高维空间中构造线性判别函数来实现原空间中的非线性判别，这种特殊性质能保证机器有较好的推广能力，同时它巧妙地解决了维数问题，使其算法复杂度与样本维数无关。

2. 支持向量机的结构

支持向量机结构如图 7-4 所示。图 7-4 中的变量 $\alpha_i (i=1, 2, \dots, s)$ 是拉格朗日乘子， s 是乘子的数量； b 为阈值或偏移量； $K(x, x'_i)$ 为一个支持向量机的核函数； $x'_i (i=1, 2, \dots, s)$ 为 SVM 的支持向量； x 为训练样本、检验样本或实测样本中的某个向量； $y(x)$ 为对应 x 的输出量。

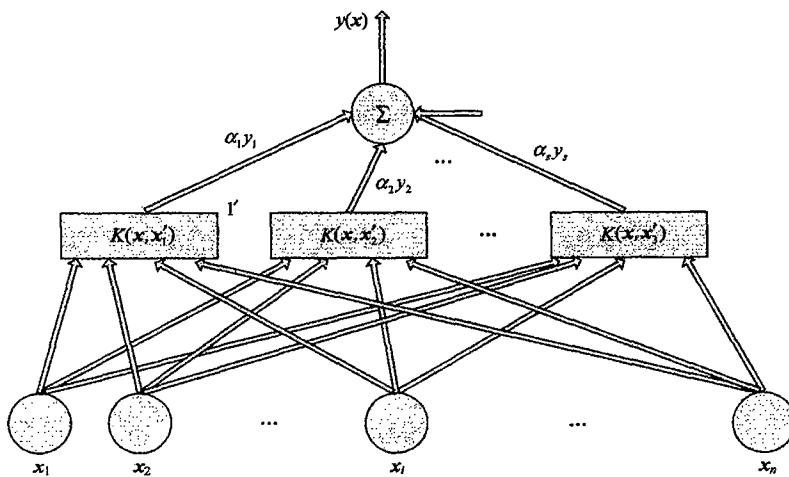


图 7-4 支持向量机结构示意图

在支持向量机结构示意图中,由原始的观测数据构成支持向量机的输入空间,通过某种关系将输入空间的数据映射到高维的特征空间。支持向量机是通过核函数来实现这种映射关系的。在特征空间 F 中,支持向量机通过线性回归函数

$$f(x) = wK(x, x')^T + b \quad (7-7)$$

来进行数据分类或拟合。

3. 支持向量机的核函数

核函数理论建立在再生核的希尔伯特空间(Reproducing Kernel Hilbert Spaces, RKHS)的基础上。核是一个函数 K ,对于所有的 $x, x' \in X$,满足: $K(x, x') = \langle \varphi(x), \varphi(x') \rangle$,这里的 φ 是从输入空间 X 到(内积)特征空间 F 的映射。核的名字最早出现在积分算子理论中,该理论以核与其相关特征空间的关系为理论基础,通过对偶表达使得高维特征空间的维数不再影响计算,即通过计算输入空间的核函数的值来计算内积,从而在特征空间中不需要进行内积的计算,这就提供了一种有效避免“维数灾难”的方法。

若特征空间的内积与输入空间的核函数等价,即

$$K(x, x') = \langle \varphi(x), \varphi(x') \rangle = \langle \varphi(x'), \varphi(x) \rangle \quad (7-8)$$

满足 Mercer 条件,也即核函数 K 是一个对称正定的函数:

$$K(x, x') = \sum_m a_m \varphi_m(x) \varphi_m(x'), \quad a_m \geq 0 \quad (7-9)$$

$$\iint K(x, x') g(x) g(x') dx dx' > 0, \quad g \in L_2 \quad (7-10)$$

那么核函数就能表示为特征空间中一个内积。满足这个性质的核函数通常称为 Mercer 核。

4. 核函数的种类

支持向量机的核函数有多种形式,例如线性核函数、多项式核函数及 RBF 核函数等。SVM 的核函数不同,则 SVM 输出表达式及输出结果不同。若有 n 组实验标定样本数据 $[x_{i1}, x_{i2}, \dots, x_{im}] (i=1, 2, \dots, n)$,每组样本数据中输入量(支持向量)为 m 维,则对于一

组输入量 $x = [x_1, x_2, \dots, x_m]$, 其输出值 y 在不同的核函数下的表达式如下。

1) 线性核(Linear)

线性核的核函数为 $K(x, x_i) = \langle x, x_i \rangle$, $\langle x, x_i \rangle$ 表示两向量的内积, 即 $K(x, x_i) = \sum_{j=1}^m x_j x_{ij}$ 。SVM 的输出值 y 为

$$\begin{aligned} y &= b + \sum_{i=1}^n w_i \langle x, x_i \rangle = b + \sum_{i=1}^n w_i x x_i^T \\ &= b + \sum_{i=1}^n w_i \begin{bmatrix} x_1 & x_2 & \cdots & x_m \end{bmatrix} \begin{bmatrix} x_{i1} \\ x_{i2} \\ \vdots \\ x_{im} \end{bmatrix} \\ &= b + \sum_{i=1}^n w_i \left(\sum_{j=1}^m x_j x_{ij} \right) \end{aligned} \quad (7-11)$$

对于第 5 章所介绍的压力传感器, 当 $m=3$ 时, 上式 SVM 的输出值为

$$y_P = b + \sum_{i=1}^n \left[w_i \sum_{j=1}^3 x_j x_{ij} \right]$$

2) 多项式核(Polynomial)

多项式核的核函数为 $K(x, x_i) = (\langle x, x_i \rangle + p_2)^{p_1}$, p_1 和 p_2 为核函数参数, 调整这两个参数可以改善支持向量机的预测准确度。SVM 的输出值 y 为

$$\begin{aligned} y &= b + \sum_{i=1}^n w_i (\langle x, x_i \rangle + p_2)^{p_1} = b + \sum_{i=1}^n w_i (x x_i^T + p_2)^{p_1} \\ &= b + \sum_{i=1}^n w_i \left\{ \begin{bmatrix} x_1 & x_2 & \cdots & x_m \end{bmatrix} \begin{bmatrix} x_{i1} \\ x_{i2} \\ \vdots \\ x_{im} \end{bmatrix} + p_2 \right\}^{p_1} \\ &= b + \sum_{i=1}^n w_i \left(\sum_{j=1}^m x_j x_{ij} + p_2 \right)^{p_1} \end{aligned} \quad (7-12)$$

同理, 当 $m=3$ 时, 由上式可以获得压力传感器的 SVM 输出表达式:

$$y_P = b + \sum_{i=1}^n \left[w_i \left(\sum_{j=1}^3 x_j x_{ij} + p_2 \right)^{p_1} \right]$$

3) RBF 核——高斯型径向基函数(Gaussian Radial Basis Function)

RBF 核的核函数为 $K(x, x_i) = e^{-\frac{\|x-x_i\|^2}{2\sigma^2}}$, $\|x-x_i\|$ 表示两个向量 x 、 x_i 取差后求模。 σ 为核函数参数, 调整 σ 可改善支持向量机的预测准确度。SVM 的输出值 y 为

$$\begin{aligned} y &= b + \sum_{i=1}^n w_i e^{-\frac{\|x-x_i\|^2}{2\sigma^2}} = b + \sum_{i=1}^n w_i \exp \left\{ -\frac{\| \begin{bmatrix} x_1 & x_2 & \cdots & x_m \end{bmatrix} - \begin{bmatrix} x_{i1} & x_{i2} & \cdots & x_{im} \end{bmatrix} \|^2}{2\sigma^2} \right\} \\ &= b + \sum_{i=1}^n w_i \cdot \exp \left\{ -\frac{\sum_{j=1}^m (x_j - x_{ij})^2}{2\sigma^2} \right\} \end{aligned} \quad (7-13)$$

同理, 当 $m=3$ 时, 由上式可以获得压力传感器的 SVM 输出表达式:

$$y_p = b + \sum_{i=1}^n w_i \exp \left\{ -\frac{\sum_{j=1}^3 (x_j - x_{ij})^2}{2\sigma^2} \right\}$$

式中 $\exp(a) = e^a$ 。

4) Sigmoid 核(也称为多层感知器核, Multi-Layer Perception)

Sigmoid 核的核函数为 $K(x, x_i) = \tanh(p_1 \langle x, x_i \rangle + p_2)$, $\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$, p_1 和 p_2 为核函数参数, 调整这两个参数可以改善支持向量机的预测精度。SVM 的输出值 y 为

$$\begin{aligned} y &= b + \sum_{i=1}^n w_i \tanh(p_1 \langle x, x_i \rangle + p_2) = b + \tanh(p_1 \mathbf{x} \mathbf{x}_i^T + p_2) \\ &= b + \sum_{i=1}^n w_i \tanh \left(p_1 [x_1 \quad x_2 \quad \cdots \quad x_m] \begin{bmatrix} x_{i1} \\ x_{i2} \\ \vdots \\ x_{im} \end{bmatrix} + p_2 \right) \\ &= b + \sum_{i=1}^n w_i \tanh(p_1 \sum_{j=1}^m x_j x_{ij} + p_2) \end{aligned} \quad (7-14)$$

当 $m=3$ 时, 由式(7-14)可以获得压力传感器的 SVM 输出表达式:

$$y_p = b + \sum_{i=1}^n w_i \tanh(p_1 \sum_{j=1}^3 x_j x_{ij} + p_2)$$

5) 张量积(Tensor Product)核

多维核的核函数可以通过形成张量积核来获得(Aronszajn, 1950)。其核函数的表达式为 $K(x, x_i) = \prod_{i=1}^n K_i(x, x_i)$, 这种核函数对于多维样条核的构建特别有用, 可以通过单变量核的积的形式简单地获得。

此外, 还有很多其它形式的核函数, 包括指类型径向基函数(Exponential Radial Basis Function)核、傅里叶级数(Fourier Series)核、样条(Splines)核、B 样条(B Splines)核、附加核(Additive Kernels)、框架核、小波(Wavelet)核和张量积(Tensor Product)核等。上面的几种核函数还可以进行组合, 如张量积的小波核等。

5. 支持向量机的种类

支持向量机按照用途的不同可以分为支持向量分类(Support Vector Classification, SVC)、支持向量回归(Support Vector Regression, SVR)和支持向量聚类等。

1) SVC

支持向量分类是指用支持向量的方法描述分类问题。为便于理解, 这里按照线性可分、线性不可分和非线性分类三种情况逐次进行说明。

(1) 线性可分情况。线性可分情况不具有一般性。考虑数据集的两类问题:

$$D = \{(x_1, y_1), \dots, (x_t, y_t)\}, \quad x \in \mathbb{R}^d, y \in \{-1, 1\} \quad (7-15)$$

如果这些数据是线性可分的, 其分割超平面为

$$\langle w, x \rangle + b = 0 \quad (7-16)$$

如果超平面无误差地将矢量集分离, 且距超平面最近的矢量与超平面之间的距离最大, 那

么就认为矢量集被超平面最优分离。如果式(7-16)没有泛化损失，那么它就是一个规范超平面，其中的参数 w 、 b 的约束为

$$\min_i |\langle w, x_i \rangle + b| = 1 \quad (7-17)$$

具体可表述为：权矢量的范数等于训练数据集中的点距超平面最近距离的倒数。图 7-5 显示距离每个超平面的最近点。

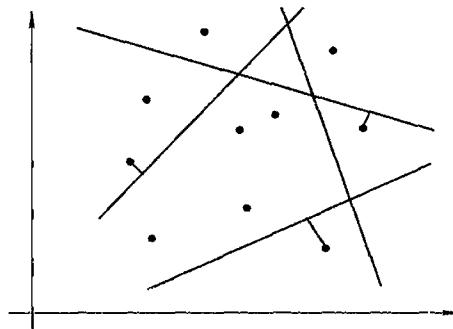


图 7-5 规范超平面

一个形式规范的分割超平面必须满足下面的约束：

$$y_i[\langle w, x_i \rangle + b] \geq 1, \quad i = 1, 2, \dots, l \quad (7-18)$$

那么点 x 距离超平面 (w, b) 的距离 $d(w, b; x)$ 为：

$$d(w, b; x) = \frac{|\langle w, x_i \rangle + b|}{\|w\|} \quad (7-19)$$

对于式(7-19)，其意义是明显的。在正确分类情况下，对于输出 y_i ，如果 $y_i=1$ ，与它对应的输入 x_i 在分类线上，那么至少应满足 $wx_i^T + b = 1$ ，于是 $y_i[wx_i^T + b] - 1 = 0$ 。反之，如果 $y_i=-1$ ，与它对应的输入 x_i 在分类线上，那么至少应满足 $wx_i^T + b = -1$ ，同样得到 $y_i[wx_i^T + b] - 1 = 0$ ，这称为 Kuhn-Tucker 条件。如果 $y_i=1$ ，与它对应的输入 x_i 不在分类线上，那么 $wx_i^T + b > 1$ ，于是 $y_i[wx_i^T + b] - 1 > 0$ ；如果 $y_i=-1$ ，与它对应的输入 x_i 不在分类线上，那么 $wx_i^T + b < -1$ ，同样得到 $y_i[wx_i^T + b] - 1 > 0$ 。最优超平面由最大的间隔 ρ （每一类之间最近点之间的距离）给定，并服从于式(7-19)的约束。间隔 $\rho(w, b)$ 定义为

$$\begin{aligned} \rho(w, b) &= \min_{x_i: y_i=-1} d(w, b; x_i) + \min_{x_j: y_j=1} d(w, b; x_j) \\ &= \min_{x_i: y_i=-1} \frac{|\langle w, x_i \rangle + b|}{\|w\|} + \min_{x_j: y_j=1} \frac{|\langle w, x_j \rangle + b|}{\|w\|} \\ &= \frac{1}{\|w\|} (\min_{x_i: y_i=1} |\langle w, x_i \rangle + b| + \min_{x_j: y_j=1} |\langle w, x_j \rangle + b|) \\ &= \frac{2}{\|w\|} \end{aligned} \quad (7-20)$$

因此，最优分离数据的超平面是通过最小化

$$\Phi(w) = \frac{1}{2} \|w\|^2 \quad (7-21)$$

得到的。因为满足式(7-18)（它是一个分离超平面），所以 b 是独立的。改变 b ，超平面将

沿着最优超平面的法方向移动。因此，间隔保持不变，但是超平面不再是唯一的，因为超平面距离一类比另一类更近。需要考虑的是如何最小化方程(7-21)，使其等价为实现结构风险最小(SRM)原则(因为前提是线性可分，所以不需要考虑经验风险)。为此，首先假设下面的界成立：

$$\|w\| < A \quad (7-22)$$

那么由方程(7-19)和(7-20)有，任何数据点与超平面之间的距离都大于 $\frac{1}{A}$ ：

$$d(w, b; x) \geq \frac{1}{A} \quad (7-23)$$

在 n 维空间中的规范超平面集的 VC 维是 h ，其界为

$$h \leq \min[R^2 A^2, n] + 1 \quad (7-24)$$

此处， R 为包围所有数据点的超球面的半径。

因此，最小化方程(7-21)等价于最小化 VC 维的上界。在约束方程(7-19)的条件下，求方程(7-21)的最优化问题的解由拉格朗日泛函的鞍点给出：

$$\Phi(w, b, \alpha) = \frac{1}{2} \|w\|^2 - \sum_{i=1}^l \alpha_i [y_i (\langle w, x_i \rangle + b) - 1] \quad (7-25)$$

式中， $\alpha = [\alpha_1, \alpha_2, \dots, \alpha_l]$ 是拉格朗日乘子。

拉格朗日泛函必须最小化相应的 w, b ，最大化相应的 $\alpha_i \geq 0 (i=1, 2, \dots, l)$ 。经典的拉格朗日泛函的对偶性可以将原始问题方程(7-25)转换为它的对偶问题，求解就很容易了。对偶问题：

$$\max_{\alpha} w(\alpha) = \max_{\alpha} (\min_{w, b} \Phi(w, b, \alpha)) \quad (7-26)$$

最小化拉格朗日泛函相应的 w 和 b ：

$$\frac{\partial \Phi}{\partial b} = 0 \Rightarrow \sum_{i=1}^l \alpha_i y_i = 0 \quad (7-27a)$$

$$\frac{\partial \Phi}{\partial w} = 0 \Rightarrow w = \sum_{i=1}^l \alpha_i y_i x_i \quad (7-27b)$$

因此，把方程(7-27)代入式(7-26)可得对偶问题为

$$\max_{\alpha} w(\alpha) = \max_{\alpha} \left[-\frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l \alpha_i \alpha_j y_i y_j \langle x_i, x_j \rangle + \sum_{k=1}^l \alpha_k \right] \quad (7-28)$$

问题的解为

$$\alpha^* = \operatorname{argmin}_{\alpha} \left[\sum_{i=1}^l \sum_{j=1}^l \alpha_i \alpha_j y_i y_j \langle x_i, x_j \rangle - \sum_{k=1}^l \alpha_k \right] \quad (7-29)$$

相应的约束为

$$\alpha_i \geq 0, \quad i = 1, 2, \dots, l; \quad \sum_{j=1}^l \alpha_j y_j = 0 \quad (7-30)$$

拉格朗日乘子决定于带有约束式(7-30)的方程式(7-29)，则最优分离超平面可写为

$$\begin{aligned} w^* &= \sum_{i=1}^l \alpha_i y_i x_i \\ b^* &= -\frac{1}{2} \langle w^*, x_r + x_s \rangle \end{aligned} \quad (7-31)$$

这里 x_r 和 x_s 是每一类中的任意支持向量，满足条件：

$$\alpha_r, \alpha_s > 0, y_r = -1, y_s = 1 \quad (7-32)$$

那么，硬分类器为

$$f(x) = \text{sgn}(\langle w^*, x \rangle + b) \quad (7-33)$$

其中 $\text{sgn}(a)$ 表示取符号函数。 $a > 0$ 时， $\text{sgn}(a) = 1$ ， $a < 0$ 时， $\text{sgn}(a) = -1$ 。另外，当存在线性内插间隔的时候，会用软分类器

$$f(x) = h(\langle w^*, x \rangle + b), \quad h(z) = \begin{cases} -1, & z < -1 \\ z, & -1 \leq z \leq 1 \\ 1, & z > 1 \end{cases} \quad (7-34)$$

软分类器比硬分类器更合适，这是因为当在间隔内部没有训练数据时，分类器能够输出 -1 到 1 之间的实数值。根据 Kuhn-Tucker 条件，

$$\alpha_i(y_i[\langle w, x_i \rangle + b] - 1) = 0, \quad i = 1, \dots, l \quad (7-35)$$

仅有满足以下条件的点才有非零的拉格朗日乘子：

$$y_i[\langle w, x_i \rangle + b] = 1 \quad (7-36)$$

这些点被称为支持向量(Support Vector, SV)。如果数据被线性分离，那么所有的 SV 都在间隔上，数目也会比较少。这样超平面只取决于训练集的一个很小的子集，从训练集中去除其它点后重新计算得到的超平面会得到与之前相同的结果。

如果数据是线性可分的，那么下面的等式成立：

$$\|w\|^2 = \sum_{i=1}^l \alpha_i = \sum_{i \in \text{SV}} \alpha_i = \sum_{i \in \text{SV}} \sum_{j \in \text{SV}} \alpha_i \alpha_j y_i y_j \langle x_i, x_j \rangle \quad (7-37)$$

因此，根据方程式(7-24)，分类器的 VC 维的界为

$$h \leq \min \left[R^2 \sum_{i \in \text{SV}}, n \right] + 1 \quad (7-38)$$

且如果训练数据 x 被正则化于单位球内，则

$$h \leq 1 + \min \left[\sum_{i \in \text{SV}}, n \right] \quad (7-39)$$

(2) 线性不可分情况。上述分类器仅限于线性可分情况，然而一般的情况并非如此。对这种问题的归纳有两种方法：依赖于问题的先验知识和数据的噪声估计。在期望能用超平面正确分离数据的情况下，引入一个额外的与错误分类相关联的损失函数是比较合适的。正如前所述，可以用一个更为复杂的函数来描述这个界。如果说在线性可分情况下所需要做的是最小化分类器的结构风险，那么在线性不可分情况下，需要做的是最小化分类器结构风险与分类结果的经验风险的综合风险。为了使得最优可分超平面一般化，Cortes 和 Vapnik(1995)引入了非负变量 $\xi_i \geq 0$ 和惩罚函数：

$$F_\sigma(\xi) = \sum_i \xi_i, \quad \sigma > 0, \text{ 常取 } \sigma = 1$$

式中， ξ_i 是错误分类的量测。最优化问题转换为最小化分类器的 VC 维的同时最小化分类误差，于是式(7-18)在不可分情况下变为

$$y_i[\langle w, x_i \rangle + b] - 1 + \xi_i \geq 0, \quad i = 1, 2, \dots, l \quad (7-40)$$

式中 $\xi_i \geq 0$ ，广义最优分类超平面是在上式约束下由最小化函数 $\Phi(w, \xi) = \frac{1}{2} \|w\|^2 +$

$C \sum_i \xi_i$ (C 为一个给定的值) 的向量 w 确定的。上式在条件 $y_i [w x_i^T + b] - 1 + \xi_i \geq 0$ ($i = 1, 2, \dots, l$) 限制下的优化问题的解由 Lagrangian 的鞍点给定：

$$\Phi(w, x, \alpha) = \frac{1}{2} \|w\|^2 - \sum_{i=1}^l \alpha_i (y_i [w x_i^T + b] - 1 + \xi_i) - \sum_{i=1}^l \beta_i \xi_i \quad (7-41)$$

式中 α, β 是 Lagrange 乘子。Lagrangian 对 w, b, x 最小化，相对于 α, β 最大化。同样地，它可以转换为对偶问题。于是除了得到式(7-27)外，还有：

$$\frac{\partial \Phi}{\partial \xi_i} = 0 \Rightarrow \alpha_i + \beta_i = C \quad (7-42)$$

于是对偶问题变成了

$$\max_w w(\alpha) = \max_{\alpha} \left[-\frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l \alpha_i \alpha_j y_i y_j \langle x_i, x_j \rangle + \sum_{k=1}^l \alpha_k \right] \quad (7-43)$$

其解为

$$\alpha^* = \arg \min_{\alpha} \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l \alpha_i \alpha_j y_i y_j \langle x_i, x_j \rangle - \sum_{k=1}^l \alpha_k \quad (7-44)$$

约束条件为

$$0 \leq \alpha_i \leq C, \quad i = 1, 2, \dots, l; \quad \sum_{i=1}^l \alpha_i y_i = 0 \quad (7-45)$$

比较式(7-30)和(7-45)可知，除了 Lagrange 乘子 α_i 的边界之外，这里的对偶问题与可分情况是一样的。Cortes 法的不确定部分是需要确定系数 C ，它在分类器中引入了额外的容量控制。 C 可以直接选择为 5，但是最终选择的 C 的值必须反映数据中的噪声。

(3) 非线性 SVC。在分类的实际应用中，简单的线性分类并不能满足实际需要，更为一般化的分类器必须具有非线性分类能力。如本节的核函数介绍部分所述，如果把输入空间通过非线性函数映射到更高维的特征空间，则可在特征空间用线性分类的方法实现输入空间的非线性分类。因此，直接选取一个合适的核函数取代式(7-44)中的 $\langle x_i, x_j \rangle$ ，然后把输入向量通过相应的分线性变换映射到特征空间，再代入到特征空间的线性分类器即可。

2) SVR

支持向量回归是指用支持向量的方法描述回归问题。由于损失函数的引入，SVM 也可以应用到回归问题(Smola, 1996)。修正的损失函数必须包括一个距离的测量。图 7-6 给出了四个可能的损失函数。

图 7-6(a) 的损失函数对应着传统的最小二乘误差标准。图 7-6(b) 的损失函数是拉普拉斯型损失函数，不如二次型的损失函数敏感。Huber 提出的损失函数如图 7-6(c) 所示，当数据的分布未知时，可作为具有最优性质的鲁棒损失函数。这三个损失函数不能产生稀疏的支持向量。为了解决这个问题，Vapnik 提出如图 7-6(d) 所示的损失函数作为 Huber 损失函数的逼近，而且能够获得稀疏的支持向量集。

(1) 线性回归。线性回归是一个数据集的逼近问题。给定一列数据：

$$D = \{(x_1, y_1), \dots, (x_l, y_l)\}, x \in \mathbb{R}^d, y \in \mathbb{R} \quad (7-46)$$

具有线性函数：

$$f(x) = \langle w, x \rangle + b \quad (7-47)$$

最优化回归函数是通过最小化以下泛函式获得的：

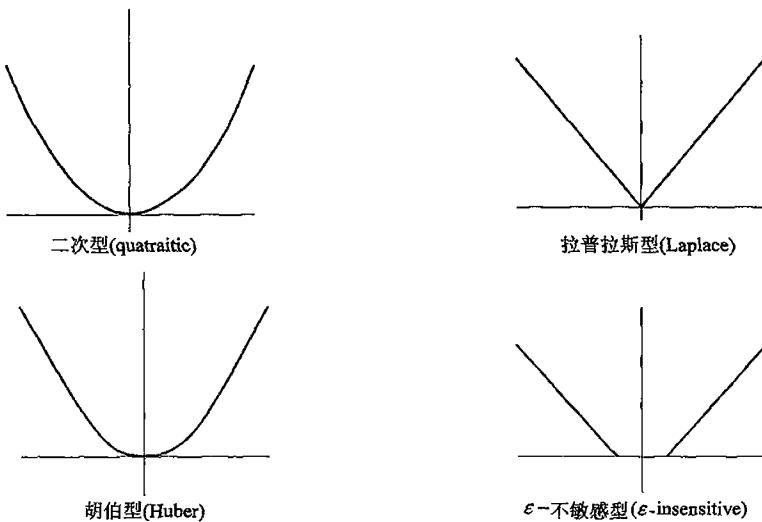


图 7-6 各种损失函数

$$\Phi(w, \xi) = \frac{1}{2} \|w\|^2 + C \sum_i (\xi_i^- + \xi_i^+) \quad (7-48)$$

式中: C 为惩罚因子, 是一预指定的值; ξ^- 和 ξ^+ 为松弛变量, 分别表示约束的上界和下界。

使用一个 ϵ -不敏感损失函数(见图 7-6(d)):

$$L_\epsilon(y) = \begin{cases} 0, & |f(x) - y| < \epsilon \\ |f(x) - y| - \epsilon, & |f(x) - y| \geq \epsilon \end{cases} \quad (7-49)$$

所求问题的解为

$$\max_{\alpha, \alpha^*} w(\alpha, \alpha^*) = \max_{\alpha, \alpha^*} \left[-\frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) \langle x_i, x_j \rangle + \sum_{i=1}^l \alpha_i (y_i - \epsilon) - \alpha_i^* (y_i + \epsilon) \right] \quad (7-50)$$

或者

$$\bar{\alpha}, \bar{\alpha}^* = \operatorname{argmin}_{\alpha, \alpha^*} \left[\frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) \langle x_i, x_j \rangle - \sum_{i=1}^l (\alpha_i - \alpha_i^*) y_i + \sum_{i=1}^l (\alpha_i + \alpha_i^*) \epsilon \right] \quad (7-51)$$

其约束条件为

$$\begin{aligned} 0 \leq \alpha_i, \alpha_i^* \leq C, \quad i = 1, \dots, l, \\ \sum_{i=1}^l (\alpha_i - \alpha_i^*) = 0 \end{aligned} \quad (7-52)$$

求解带有约束条件式(7-52)的方程式(7-51)得到拉格朗日乘子 α 和 α^* , 由方程(7-47)得到回归方程, 此处,

$$\begin{aligned} \bar{w} &= \sum_{i=1}^l (\alpha_i - \alpha_i^*) x_i \\ \bar{b} &= -\frac{1}{2} \langle \bar{w}, (x_r + x_s) \rangle \end{aligned} \quad (7-53)$$

解满足 Karush - Kuhn - Tucker(KKT) 条件：

$$\bar{\alpha}_i \bar{\alpha}_i^* = 0, \quad i = 1, \dots, l \quad (7-54)$$

因此，支持向量是拉格朗日乘子远大于零的点。当 $\epsilon=0$ 时，我们得到 L_1 损失函数和简化的最优化问题：

$$\min_{\beta} \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l \beta_i \beta_j \langle \mathbf{x}_i, \mathbf{x}_j \rangle - \sum_{i=1}^l \beta_i y_i \quad (7-55)$$

约束条件为

$$\begin{aligned} -C &\leq \beta_i \leq C, \quad i = 1, \dots, l \\ \sum_{i=1}^l \beta_i &= 0 \end{aligned} \quad (7-56)$$

式中 $\beta_i = \alpha_i - \alpha_i^*$ 。由方程(7-47)给出的回归函数中：

$$\begin{aligned} \bar{\mathbf{w}} &= \sum_{i=1}^l \beta_i \mathbf{x}_i \\ \bar{b} &= -\frac{1}{2} \langle \bar{\mathbf{w}}, (\mathbf{x}_r + \mathbf{x}_s) \rangle \end{aligned} \quad (7-57)$$

二次型的损失函数如图 7-6(a) 所示，有

$$L_{\text{quadratic}}(f(x) - y) = (f(x) - y)^2 \quad (7-58)$$

解的确定，或者说优化问题转换为

$$\min \frac{1}{2} \| \mathbf{w} \|^2 + \frac{C}{2} \sum_{i=1}^l \xi_i^2 \quad (7-59)$$

式中， C 为惩罚因子， ξ_i 为松弛因子，约束条件为

$$y_i = \langle \mathbf{w}, \mathbf{x}_i \rangle + b + \xi_i, \quad i = 1, 2, \dots, l \quad (7-60)$$

转换为对偶问题得到：

$$\begin{aligned} \max_{\alpha, \alpha^*} w(\alpha, \alpha^*) &= \max_{\alpha, \alpha^*} \left[-\frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) \langle \mathbf{x}_i, \mathbf{x}_j \rangle \right. \\ &\quad \left. + \sum_{i=1}^l (\alpha_i - \alpha_i^*) y_i - \frac{1}{2C} \sum_{i=1}^l (\alpha_i^2 + (\alpha_i^*)^2) \right] \end{aligned} \quad (7-61)$$

使用 KKT 条件，相应的最优化问题可以简化且有 $\beta_i^* = |\beta_i|$ ，由此产生的最优化问题

$$\min_{\beta} \left[\frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l \beta_i \beta_j \langle \mathbf{x}_i, \mathbf{x}_j \rangle - \sum_{i=1}^l \beta_i y_i + \frac{1}{2C} \sum_{i=1}^l \beta_i^2 \right] \quad (7-62)$$

的约束条件为

$$\sum_{i=1}^l \beta_i = 0 \quad (7-63)$$

回归函数由方程式(7-47)和(7-57)确定。

对于二次型的损失函数，也可以这样理解，对式(7-59)引入拉格朗日函数：

$$L = \frac{1}{2} \| \mathbf{w} \|^2 + \frac{C}{2} \sum_{i=1}^l \xi_i^2 - \sum_{i=1}^l \alpha_i (\mathbf{w} \mathbf{x}_i^T + b + \xi_i - y_i) \quad (7-64)$$

式中 $\alpha_i (i=1, 2, \dots, l)$ 为拉格朗日乘子。分别求解式(7-64)的关于 \mathbf{w} 、 ξ_i 、 b 、 α_i 的偏微分可得：

$$\begin{cases} \frac{\partial L}{\partial w} = w^T - \sum_{i=1}^n \alpha_i x_i^T = 0 \\ \frac{\partial L}{\partial \xi_i} = C\xi_i - \alpha_i = 0 \\ \frac{\partial L}{\partial b} = \sum_{i=1}^n \alpha_i = 0 \\ \frac{\partial L}{\partial \alpha_i} = w x_i^T + b + \xi_i - y_i = 0 \end{cases} \quad (7-65)$$

由式(7-65)消去 w 、 ξ_i 则可得：

$$\begin{bmatrix} 0 & 1 & \cdots & 1 \\ 1 & x_1 x_1^T + \frac{1}{C} & \cdots & x_1 x_i^T + \frac{1}{C} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_i x_1^T + \frac{1}{C} & \cdots & x_i x_i^T + \frac{1}{C} \end{bmatrix} \begin{bmatrix} b \\ \alpha_1 \\ \vdots \\ \alpha_i \\ \vdots \\ \alpha_n \end{bmatrix} = \begin{bmatrix} 0 \\ y_1 \\ \vdots \\ y_i \\ \vdots \\ y_n \end{bmatrix}$$

求解上式可得到 LS-SVM 的回归方程，得到 α_i 和 b ，于是可得到回归方程为

$$f(x) = \sum_{i=1}^l \alpha_i \langle x_i, x \rangle + b \quad (7-66)$$

(2) 非线性回归。同分类问题类似，非线性建模通常要求足够多的建模数据。采用与非线性 SVC 相同的方法，可以用非线性映射来把数据映射到高维特征空间，然后在该空间进行线性回归。维数灾难问题依然用核方法来解决。例如，用核函数 $K(x, x_i)$ 取代式(7-66)中的内积 $\langle x, x_i \rangle$ 即可得到非线性回归

$$f(x) = \sum_{i=1}^n \alpha_i K(x, x_i) + b \quad (7-67)$$

7.2 支持向量机的训练、检验与测量

7.2.1 训练样本及检验样本的制备

在第 6 章已经介绍了用于神经网络学习的样本制备方法及步骤，支持向量机学习所需要的样本的制备方法与第 6 章相同。取实验标定获得的样本数据对中的一部分 (N_1 组样本对，约占总样本数的 $1/2 \sim 2/3$) 作训练样本，剩余部分 (N_2 组样本对) 作为检验样本， $N_1 + N_2 = N$ 。训练样本及检验样本数据的数量可能不同，但是格式相同，由 SVM 的输入样本 (即支持向量机的输入向量) 和期望输出样本 (即期望输出向量) 组成，格式为 $\{[x_{i1}, x_{i2}, \dots, x_{in}], y_i\}$ 。

7.2.2 支持向量机的训练

按照以下步骤对支持向量机进行训练：

(1) 输入训练样本：将全部训练样本的输入向量 $[x_{i1}, x_{i2}, \dots, x_{in}] (i=1, 2, \dots, N_1)$

作为支持向量 x_i , $y = [P'_i](i=1, 2, \dots, N_1)$ 作为期望输出向量 y , 一次并行输入支持向量机, 并将全部检验样本的输入向量 $[x_{i1}, x_{i2}, \dots, x_{im}](i=1, 2, \dots, N_2)$ 作为特征空间向量 x 依次串行输入支持向量机。

(2) 设置 SVM 学习参数及核函数参数。

(3) SVM 训练: 基于训练样本及真实风险最小化原则, 求出 SVM 结构参数(权系数 $\alpha_1, \alpha_2, \dots, \alpha_s$ 和偏移量 b), 使输出向量 $y(x)$ 与训练样本中的期望输出向量 y 的偏差最小, 从而表明支持向量机的训练结束, 训练好的支持向量机能否用于实测还需要进行检验。

7.2.3 支持向量机的检验

将检验样本作为输入向量 x 输入已经训练好的支持向量机中, 获得检验样本的输出结果。当检验样本结果满足实测要求时, 表明训练好的支持向量机可进行准确的识别和回归; 反之, 则要调整支持向量机训练参数进行新的训练, 直至支持向量机检验结果偏差满足实测要求为止。通过对支持向量机的检验, 确定代表支持向量机结构的权系数 $\alpha_1, \alpha_2, \dots, \alpha_s$ 和偏移量 b , 根据所选的核函数, 并带入相应的 SVM 输出表达式(公式(7-12)~公式(7-15)或其它式)中, 就可以获得降低或消除压力传感器交叉敏感的回归模型, 得到准确的智能压力传感器系统的输入输出关系。

7.2.4 测量

将被测量样本 $[x_{i1}, x_{i2}, \dots, x_{im}]$ (既不是训练用的样本也不是检验用的样本)输入相应的已经训练好的回归模型——SVM 的输出表达式中, 就可以由输出结果获得准确的待测量。

7.2.5 移植

将已经训练完毕且检验合格可以实现某种智能化功能的支持向量机结构, 即已知权系数 $\alpha_1, \alpha_2, \dots, \alpha_s$ 和偏移量 b 的回归模型——SVM 的输出表达式, 移植到另一个硬件环境(如单片机、DSP 或测量用的计算机)或已知的另一个软件环境中, 也就是训练检验与测量是在不同的编程语言环境(如 C 语言、LabVIEW、Matlab)中进行的。之后, 若将测量样本数据输入到 SVM 的输出表达式中进行被测量的测量, 仍然可以达到同样的提高整个系统稳定性和准确性的该种智能化功能的效果。

7.3 基于 SVM 方法的三传感器数据融合原理

7.3.1 三传感器数据融合的智能传感器系统的组成

采用 SVM 技术进行三传感器数据融合的智能传感器系统由传感器模块和支持向量机模块两部分组成, 如图 7-7 所示。

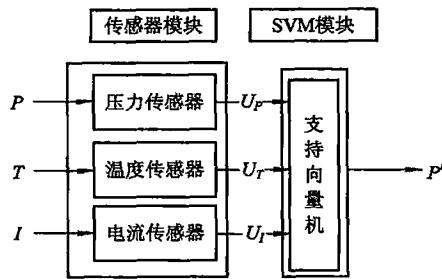


图 7-7 采用 SVM 进行三传感器数据融合的智能压力传感器系统框图

1. 传感器模块

如图 7-7 所示，传感器模块输出三个电压信号，其中 U_P 为被测压力 P （目标参量）的电压输出信号， U_I 、 U_T 为两个非目标参量的检测电压信号。传感器模块的检测电路如图 7-8 所示。

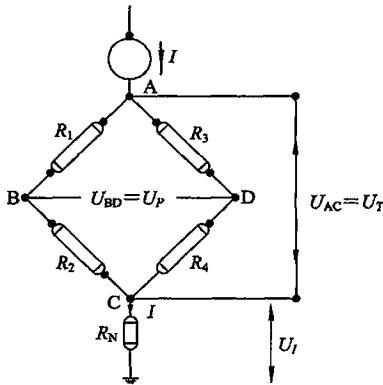


图 7-8 传感器电路原理图

(1) 压力传感器：实验采用 CYJ-201 型压阻式压力传感器，其输出电压 U_P 对应被测压力 P （目标参量）。该传感器包含 4 个压敏电阻，这 4 个电阻封装在一起，其中 R_1 和 R_4 完全一样，其电阻随着压力的增大而增大， R_2 和 R_3 完全一样，其电阻随着压力的增大而减小。一个理想的压力传感器，其输出电压 U_P 应为输入压力 P 的一元单值函数，即 $U_P = f(P)$ ，其反函数为 $P = f^{-1}(U_P)$ 。而在实际应用中，该传感器受工作温度 T 和电源供电电流 I 的影响，其输出电压 U_P 也将发生变化，实际上是一个三元函数，即 $U_P = f(P, I, T)$ 。

(2) 温度传感器：温度传感器将工作温度 T 转换为电压信号 U_T 。由于图 7-8 中的 4 个电阻的温度特性是一样的，都是随着温度的升高而电阻增大，因此，采用恒流源供电的压阻式压力传感器，其供电端(AC 两端)电压 U_{AC} 即为 U_T 。

(3) 电流传感器：电流传感器将电流信号 I 转换为电压信号 U_I ，如图 7-8 所示。采用标准恒定电阻 R_N 与压力传感器相串联， R_N 两端电压为 $U_I = IR_N$ 。

2. 支持向量机模块

支持向量机(SVM)模块是由软件编程实现的一种算法。在本实验中，网络的三个输入量 x_1, x_2, x_3 分别对应着 U_p, U_t, U_r ，输出量为 P' 。 P' 亦是智能压力传感器系统的总输出量。通过压力传感器的多维标定实验获得 n 组样本 $\{(U_{p_i}, U_{t_i}, \dots, U_{r_i}), P'_i\} (i=1, 2, \dots, n)$ ，记为 n 个样本点： (x_i, y_i) ， $(x_i \in \mathbb{R}^m, y_i \in \mathbb{R}, m=3)$ ， x_i 代表输入向量 $(U_{p_i}, U_{t_i}, \dots, U_{r_i})$ ； m 代表输入向量的维数，对于本实验的问题而言， $m=3$ ； y_i 代表期望输出标量 P'_i 。基于 SVM 进行数据融合的目的是来拟合输入 x 与输出 y 之间的关系：

$$y(x) = w^T x + b = \sum_{i=1}^s \alpha_i K(x, x_i) + b \quad (7-68)$$

式中： $x_i' (i=1, 2, \dots, s)$ 是支持向量； s 为支持向量的数量； x 是被测输入向量； b 是 SVM 阈值或偏移量； w 是 SVM 的权值系数，其数量与支持向量数量相同，即 $w=[w_1, w_2, \dots, w_s, \dots, w_n]$ ； α_i 是与 SVM 的权值系数相对应的拉格朗日乘子； $K(x, x_i)$ 为 SVM 的核函数。

7.3.2 [示例 7-1]降低两个干扰量影响的 SVM 智能化软件模块的设计

要求：采用 Gaussian 型 RBF 核函数的 SVM 作为多传感器交叉敏感的逆模型，改善压阻式压力传感器(JCY-201)的温度稳定性与恒流源供电电流的稳定性，并对改善前后的稳定性作出评价。

监测干扰量温度 T 的温度传感器的输出电压为 U_T ；监测干扰量恒流源供电电流 I 的电流传感器输出电压为 U_I ；构建 Gaussian 型 RBF 核函数的 SVM 的三传感器数据融合智能传感器系统框图如图 7-2(支持向量机的图)所示。

1. 样本的制作

三维标定实验及获得的样本数据与第 5 章表 5-2 相同。支持向量机的学习样本与测试样本与 RBF 网络的样本完全相同，可以在程序中直接输入，也可以先制作成文件。文件类型的学习(训练)样本和测试(检验)样本，包含训练样本的输入样本文件 pressure_train_in.txt 和期望输出样本文件 pressure_train_out.txt，检验样本的输入样本文件 pressure_test_in.txt 和期望输出样本文件 pressure_test_out.txt，一共四个.txt 文件。这里的输入样本是三维的，有三个输入变量 U_p, U_t, U_r ，期望输出样本是一维的，是一列与输入变量相对应的压力值 P 。

2. 用于消除交叉敏感的支持向量机 Matlab 源程序

程序如下：

```
%程序开始前，清屏，清其它变量
clc; clear; close all;

%输入原始数据，每一列为一个变量，行数为样本的个数。原始标定数据的输入，可以将数据
直接输入至程序中，也可以将数据先输入至文件，然后在程序中读入数据文件。数据中的 P、
Ut、Ui、Up 均为 n×1 维列向量，n 为标定数据点的个数
raw_data=[P Ut Ui Up]; %学习(训练)样本输入
data_test=[P_test Ut_test Ui_test Up_test]; %测试(检验)样本输入
```

```

%根据原始数据制作学习(训练)样本与测试样本, 学习(训练)样本与测试样本的数据格式相同
P=raw_data(:, 1); Ui=raw_data(:, 2);
Ut=raw_data(:, 3); Up=raw_data(:, 4);
P_test= data_test(:, 1);Ui= data_test (:, 2);
Ut= data_test (:, 3); Up= data_test (:, 4);

P1=[Up Ui Ut]'; %学习(训练)样本输入向量
T1=P'; %学习(训练)样本期望输出向量
P2=[Up_test Ui_test Ut_test]'; %测试(检验)样本输入向量
T2=P_test'; %测试(检验)样本输出向量

%将输入数据进行归一化处理
%[PN1, minp, maxp, TN1, mint, maxt] = premnmx(P1, T1);
%PN2 = tramnmx(P2, minp, maxp);
%TN2 = tramnmx(T2, mint, maxt);

%如果对数据不进行归一化处理, 选下面的程序
PN1=P1; TN1=T1; PN2=P2; TN2=T2;

%支持向量机的学习参数以及核函数的选择
C = 500;
lambda = 1e-10;
epsilon = 1e-6;
%kerneloption = 0.115; %对输入数据进行归一化处理所采用的值
kerneloption = 0.338; %对输入数据未进行归一化处理所采用的值

kernel='gaussian'; %高斯型的 RBF 核
verbose=1;

%函数 svmreg()为支持向量机的学习函数
%输出确定输入空间的支持向量 xsup 和 ysup 以及支持向量机的权系数 w 和偏移量 w0
[xsup, ysup, w, w0] = svmreg(PN1', TN1', C, epsilon, kernel, kerneloption, lambda,
verbose);

%函数 svmval()为支持向量机的检验函数
%ypred1 为学习样本的支持向量预测输出, ypred2 为测试样本的支持向量预测输出
ypred1 = svmval(PN1', xsup, w, w0, kernel, kerneloption);
ypred2 = svmval(PN2', xsup, w, w0, kernel, kerneloption);
disp('学习样本的压力标定值见 P 后的数据:');
P
disp('学习样本经 SVR 后的输出值见 Y1 后的数据:');
Y1=postmnmx(ypred1, mint, maxt) %若程序开始时对数据进行归一化处理, 选择该语句
Y1=ypred1 %如果程序开始时对数据没有进行归一化处理, 选择该语句
delta1=Y1-P
disp('学习样本经 SVR 后的输出值与其压力标定值之间的最大、最小偏差:');
max(delta1)

```

```

min(delta1)
MSE1 = mean(delta1.^2) %训练均方误差
MSETD1 = std(delta1.^2)

disp('测试样本的压力标定值见 P_test 后的数据：');
P_test
disp('测试样本经 SVR 后的输出值见 Y2 后的数据：');
%Y2 = postmnmx(ypred2, mint, maxt)S %若程序开始时对数据进行归一化处理,
选择该语句
Y2=ypred2 %如果程序开始时对数据没有进行归一化处理, 选择该语句

disp('测试样本经 SVR 后的输出值与其压力标定值之间的最大、最小偏差：');
delta2=Y2-P_test
max(delta2)
min(delta2)
MSE2 = mean(delta2.^2) %测试均方误差
MSETD2 = std(delta2.^2)

```

3. 程序说明

(1) 数据预处理：如果有些支持向量机对所处理数据的范围有要求，如数据必须在区间 $[-1, 1]$ 内，那么就需要对原始数据进行归一化处理。对于本书提供的程序，可以进行归一化处理，也可不进行归一化处理，具体见程序中的说明。

(2) 支持向量机的参数： C 为拉格朗日乘子的界， λ 为 QP(二次规划)方法的条件参数， ϵ 为所求解周围的 ϵ 邻域(也称试管)和 kerneloption，对 RBF 核函数来说，kerneloption 相当于参数 σ (见式 7-6)。

(3) 参数调整对输出结果的影响：上述的这些核参数除了 kerneloption 对输出结果有显著的影响外，其它几个参数 C 、 λ 和 ϵ 对结果输出影响不大，其中 ϵ 决定输出结果的准确度，一般小于 10^{-6} 即可。参数 kerneloption 的数值需要进行反复试验，从而找到合适的支撑向量机结构参数 w 和 b ，使输出结果最佳(一般为测试样本的最大输出值和最小输出值的绝对值和最小)。

(4) 对输入样本(学习样本和测试样本)是否作归一化处理，使输出结果最佳的参数 kerneloption 的最优值一般不一样，这在实验中需要注意。

(5) 程序中的 SVM 的相关函数的编写以及后面的程序所用到的支持向量机核函数，具体可以参考法国学者 Alain Rakotomamonjy 编写的 SVM and Kernel Methods Matlab Toolbox，网址为 <http://asi.insa-rouen.fr/enseignants/~arakotom/toolbox/index.html>，以及核机器学习网站 <http://www.kernel-machines.org/>，这里就不提供相应函数的源程序了。

4. SVM 融合输出结果

用于消除三传感器交叉敏感 SVM 的核函数，本实验选用 Gaussian 型 RBF 核函数。支持向量机实现的程序环境为 Matlab 2007b。程序中的 SVM 学习函数为 svmreg()，SVM 的仿真测试函数为 svmval()。 C 为拉格朗日乘子的界，程序中 $C = 500$ ； λ 为 QP 方法的条

件参数，程序中 $\lambda = 1e-10$ ； ϵ 为所求解附近的 ϵ 邻域，程序 $\epsilon = 1e-7$ ；
 kerneloption 为核参数，程序中 $\text{kerneloption} = 0.338$ 。程序中对输入样本未进行归一化处理。

采用 Gaussian 型 RBF 核函数的 SVM 输出结果：学习样本经 SVR 后的输出值与其压力标定值之间最大偏差的绝对值为 $2.5111e-007$ ，均方误差为 $MSE1 = 2.5024e-010$ ，均方误差的标准差 $MSETD1 = 3.9722e-009$ 。由上面数据可见，在一定误差许可的条件下，采用 Gaussian 型 RBF 核函数的 SVM，对学习样本的输出已经极大地消除了交叉敏感的影响。检验样本经 SVR 后的输出值与其压力标定值之间的最大偏差的绝对值为 0.0041 ，均方误差为 $MSE2 = 6.0730e-007$ ，均方误差的标准差为 $MSETD2 = 2.4253e-006$ 。检验样本的输出值(部分)见表 7-1 所示。

表 7-1 $\sigma = 0.338$ 时，Gaussian 型 RBF 核函数的 SVM 检验样本输出结果

I/mA	6				7				8				10				11			
$T/^\circ\text{C}$	P	\hat{P}	Δ		\hat{P}	Δ		\hat{P}	Δ	\hat{P}	Δ		\hat{P}	Δ		\hat{P}	Δ			
43	0	0.0015	0.0015		0.0009	0.0009		0.0004	0.0004	-0.0001	-0.0001		0.0021	0.0021						
	0.5	0.5001	0.0001		0.4993	-0.0007		0.5006	0.0006	0.4996	-0.0004		0.5009	0.0009						
53	0	-0.0001	-0.0001		-0.0002	-0.0002		-0.0000	-0.0000	0.0008	0.0008		0.0029	0.0029						
	0.5	0.5003	0.0003		0.5003	0.0003		0.4999	-0.0001	0.5009	0.0009		0.5003	0.0003						
58	0	-0.0041	-0.0041		-0.0014	-0.0014		-0.0002	-0.0002	0.0041	0.0041		-0.0005	-0.0005						
	0.5	0.4998	-0.0002		0.5006	0.0006		0.4995	-0.0005	0.4995	-0.0005		0.4997	-0.0003						

注：表中 T 表示工作温度， I 表示工作电流， P 表示压力标定值， \hat{P} 表示压力计算值，偏差 $\Delta = \hat{P} - P$ ，单位为 MPa。

同样，如果 kerneloption 的数值选择不好，比如在程序中令 $\text{kerneloption}=10$ ，其它参数保持不变，则输出结果学习样本经 SVR 后的输出值与其压力标定值之间的最大偏差绝对值为 0.0277 ，均方误差 $MSE1 = 4.3901e-005$ ， $MSETD1 = 1.0083e-004$ 。检验样本的输出值与对应的压力标定值之间的最大偏差绝对值为 0.016 ，均方误差 $MSE2 = 3.1914e-005$ ， $MSETD2 = 5.962e-004$ 。检验样本部分输出值见表 7-2。

表 7-2 $\sigma = 10$ 时，Gaussian 型 RBF 核函数的 SVM 检验样本输出结果

I/mA	6				7				8				10				11			
$T/^\circ\text{C}$	P	\hat{P}	Δ		\hat{P}	Δ		\hat{P}	Δ	\hat{P}	Δ		\hat{P}	Δ		\hat{P}	Δ			
43	0	-0.0144	-0.0144		-0.0156	-0.0156		-0.0160	-0.0160	-0.0138	-0.0138		-0.0136	-0.0136						
	0.5	0.4913	-0.0087		0.5002	0.0002		0.5003	0.0003	0.4961	-0.0039		0.4965	-0.0035						
53	0	0.0051	0.0051		0.0049	0.0049		0.0062	0.0062	0.0116	0.0116		0.0112	0.0112						
	0.5	0.4907	-0.0093		0.4988	-0.0012		0.5000	0.0000	0.4983	-0.0017		0.4980	-0.0020						
58	0	-0.0118	-0.0118		-0.0134	-0.0134		-0.0129	-0.0129	-0.0096	-0.0096		-0.0101	-0.0101						
	0.5	0.4882	-0.0118		0.4974	-0.0026		0.4986	-0.0014	0.4972	-0.0028		0.4976	-0.0024						

注：同表 7-1。

对比表 7-1 和表 7-2 的数据可明显看出，虽然选取 $\sigma = 10$ (即程序中参数 $\text{kerneloption}=10$)也可以对数据进行融合，但是融合的结果明显比选 $\sigma = 0.338$ 时的输出结果差很多。因此，需要对这些参数进行仔细的选择、调试，最后选取最优的结果。

5. 融合后效果评价

(1) 当取 $\sigma = 0.338$ 时, SVM 输出效果评价。在 $\Delta T = 39^\circ\text{C}$, $\Delta I = 5 \text{ mA}$ 变化范围内, 零点计算值的最大偏差 $|\Delta P'_{0m}| = 0.0041 \text{ MPa}$; 满量程压力 $P_{\text{fs}} = 0.5 \text{ MPa}$, 其计算值的最大偏差量 $|\Delta P'_m| = 0.0009 \text{ MPa}$, 则有:

$$\alpha_0 = \frac{|\Delta P'_{0m}|}{P_{\text{fs}}} \cdot \frac{1}{\Delta T} = \frac{0.0041}{0.5} \cdot \frac{1}{39} = 2.1 \times 10^{-4} \quad (\text{}/^\circ\text{C})$$

$$\alpha_s = \frac{|\Delta P'_m|}{P_{\text{fs}} \cdot \Delta T} = \frac{0.0009}{0.5} \cdot \frac{1}{39} = 4.6 \times 10^{-5} \quad (\text{}/^\circ\text{C})$$

$$\alpha_I = \frac{|\Delta P'_m|}{P_{\text{fs}} \cdot \Delta I} = \frac{0.0009}{0.5 \times 5} = 3.6 \times 10^{-4} \quad (\text{/mA})$$

(2) 当取 $\sigma = 10$ 时, SVM 的效果评价。同样, 在 $\Delta T = 39^\circ\text{C}$, $\Delta I = 5 \text{ mA}$ 变化范围内, 零点计算值的最大偏差 $|\Delta P'_{0m}| = 0.016 \text{ MPa}$; 满量程压力 $P_{\text{fs}} = 0.5 \text{ MPa}$, 其计算值的最大偏差量 $|\Delta P'_m| = 0.0118 \text{ MPa}$, 则有:

$$\alpha_0 = \frac{|\Delta P'_{0m}|}{P_{\text{fs}}} \cdot \frac{1}{\Delta T} = \frac{0.016}{0.5} \cdot \frac{1}{39} = 8.2 \times 10^{-3} \quad (\text{}/^\circ\text{C})$$

$$\alpha_s = \frac{|\Delta P'_m|}{P_{\text{fs}} \cdot \Delta T} = \frac{0.0118}{0.5} \cdot \frac{1}{39} = 6.1 \times 10^{-4} \quad (\text{}/^\circ\text{C})$$

$$\alpha_I = \frac{|\Delta P'_m|}{P_{\text{fs}} \cdot \Delta I} = \frac{0.0118}{0.5 \times 5} = 4.7 \times 10^{-3} \quad (\text{/mA})$$

(3) 融合效果比较。需要说明的是, 在使用 Gaussian 型 RBF 核函数的 SVM 过程中, 其核函数的参数 kerneloption 的数值需要进行反复试验, 从而找到合适的支撑向量机结构 w 和 b , 使得输出结果最佳。

融合前的计算值同第 5 章[示例 5-1], 融合效果比较见表 7-3。

表 7-3 融合后的各参数

评价参数	零位温度系数 $\alpha_0 (\text{}/^\circ\text{C})$	灵敏度温度系数 $\alpha_s (\text{}/^\circ\text{C})$	电流影响系数 $\alpha_I (\text{/mA})$
融合前计算值	1.9×10^{-3}	1.2×10^{-2}	9.0×10^{-2}
$\sigma = 0.338$, Gaussian 型 RBF 核融合	2.1×10^{-4}	4.6×10^{-4}	3.6×10^{-4}
$\sigma = 10$, Gaussian 型 RBF 核融合	8.2×10^{-3}	6.1×10^{-4}	4.7×10^{-3}

6. 移植

由于压力传感器在现场实际使用的过程中, 传感器的实际测量输出一般通过单片机或 DSP 进行数据处理, 但是在能进行浮点运算的单片机或者 DSP 的硬件环境中, 不可能安装 Matlab 等大型数值计算软件, 因此在实际的使用过程中可以这样来进行移植: 先在 PC 计算机上训练支持向量 SVM, 然后再将训练合格的支持向量 SVM 的结构参数、权系数 w 、偏置 b 、RBF 核函数的 σ 直接固化到单片机或者 DSP 中, 这样在单片机或者 DSP 中只需进行式(7-8)所示简单的乘法、加法、指数和矩阵乘法等运算, 即可在现场完成传感器数据融合输出。

采用 Gaussian 型 RBF 核函数的 SVM, 经过学习样本训练后的权系数和偏置详见表 7-4。

表 7-4 经过 SVM 学习后的支持向量机的权系数与偏移量

权系数	权系数数值									
	$w_0 \sim w_9$	$w_{10} \sim w_{19}$	$w_{20} \sim w_{29}$	$w_{30} \sim w_{39}$	$w_{40} \sim w_{49}$	$w_{50} \sim w_{59}$	$w_{60} \sim w_{69}$	$w_{70} \sim w_{79}$	$w_{80} \sim w_{89}$	$w_{90} \sim w_{99}$
$w_0 \sim w_9$	-5.1515	-3.8024	-1.2988	-11.6221	27.5267	-15.0892	-3.7380	-5.0454	-6.9700	-11.6971
$w_{10} \sim w_{19}$	-9.5427	7.4446	-1.0642	-1.6607	-4.1339	-6.4829	-4.1580	2.7147	-0.7087	-1.8437
$w_{20} \sim w_{29}$	-1.7696	-3.5045	-4.3461	0.7136	-0.5763	-0.7901	-1.0978	-1.5796	-3.2021	1.8062
$w_{30} \sim w_{39}$	-1.4001	-1.2582	-1.1418	-1.1885	-1.3536	2.3264	-9.1751	27.1525	20.3028	25.4071
$w_{40} \sim w_{49}$	-96.0319	8.5147	-3.2784	14.0125	14.6797	15.2737	13.5109	7.6891	-0.8762	2.4023
$w_{50} \sim w_{59}$	3.0268	3.1486	0.7916	-2.0326	-0.5045	1.3205	1.4326	7.7889	8.0394	4.7125
$w_{60} \sim w_{69}$	-0.2801	0.819	1.1424	1.2868	2.0808	-2.8462	-0.3079	0.5698	0.4417	0.3623
$w_{70} \sim w_{79}$	0.6677	-1.1348	49.7563	-21.7346	4.6875	79.4249	244.1223	123.8275	16.1915	-26.8498
$w_{80} \sim w_{89}$	-16.8914	-6.5081	-6.6867	-61.9567	4.6115	1.0873	7.9518	17.5118	14.7643	-6.8367
$w_{90} \sim w_{99}$	2.6709	3.4342	2.2598	2.9697	5.0258	-0.521	2.1304	0.9691	-0.0029	0.8396
$w_{100} \sim w_{109}$	3.0741	-1.2916	2.6722	1.6764	1.4347	1.4825	0.9984	-2.824	-39.0524	-17.5276
$w_{110} \sim w_{119}$	-56.1417	-127.3354	-279.041	-131.7276	-10.4841	34.8978	22.4497	17.0127	18.4504	68.6956
$w_{120} \sim w_{129}$	-3.0465	-1.8324	-7.1531	-14.8276	-13.0771	5.3204	-1.8495	-2.9887	-1.6207	-13.7176
$w_{130} \sim w_{139}$	-16.0891	-16.2098	-1.5847	-0.8532	0.6528	0.0280	-1.0797	2.0842	-1.5038	-0.9569
$w_{140} \sim w_{149}$	-0.7002	-0.6064	-0.1231	1.3271	-6.9806	54.9870	150.016	111.7165	500.00	0.0805
$w_{150} \sim w_{159}$	-1.3075	-119.9906	-88.6865	-87.6463	-97.9011	-106.9821	-0.5503	-1.1049	-0.6923	-2.2102
$w_{160} \sim w_{169}$	5.0442	12.2895	-0.2987	-2.0143	-2.1873	19.079	23.3299	43.9412	-0.3753	-2.0201
$w_{170} \sim w_{179}$	-3.2796	-2.5916	-4.5214	6.0827	-0.0092	-1.5352	-0.9152	-0.4796	-0.4561	3.6526
$w_{180} \sim w_{189}$	4.5878	-44.7552	-137.9638	-87.6705	-456.7712	20.6037	0.4045	111.1792	79.7309	76.084
$w_{190} \sim w_{199}$	85.5564	90.2245	0.3836	1.2242	0.5592	2.943	-5.4761	-16.6696	0.3733	2.516
$w_{200} \sim w_{209}$	1.9836	-15.5552	-20.1018	-41.0525	0.3668	2.2417	2.8076	1.9352	3.5926	-7.6293
$w_{210} \sim w_{219}$	0.3752	1.8845	1.0263	0.4105	0.0399	-4.114	1.0598	4.8254	20.1577	10.6364
$w_{220} \sim w_{229}$	64.8241	-4.174	0.4727	-8.3724	-4.5622	-2.8619	-2.997	-3.8505	-0.057	-0.1644
$w_{230} \sim w_{239}$	0.3678	-0.1556	2.3482	5.9196	-0.1494	-0.6142	-0.2097	2.9924	4.3048	9.2507
$w_{240} \sim w_{249}$	-0.055	-0.5264	-0.3484	-0.0119	-0.1217	2.3162	-0.3634	-0.6017	-0.2603	0.0201
$w_{250} \sim w_{251}$	0.2776	1.2409					偏移量 $b=0.2383$			

移植用的 Matlab 程序(部分):

```
%程序开始
%初始化
xsup=[Up Ui Ut]; % xsup 为支持向量, 本实验中, 支持向量和学习输入样本相同
x=[Up_meas Ui_meas Ut_meas]; % x 为待融合数据, 即传感器现场测量的数据
kerneloption=0.338; % kerneloption 是指 Gaussian 型 RBF 函数中的 sigma, 带宽
[n1 n2]=size(x); % 计算待融合数据的尺寸, n1 为待融合数据的组数, n2 为待融合
% 数据中所包含变量的个数, 这里的 n2 应该为 3, ( $U_p$ ,  $U_i$ ,  $U_T$ )
% 三个变量
[n n3]=size(xsup); % 计算支持向量的尺寸, n 为支持向量的个数, n3 为支持向量中所
% 包含变量的个数, 这里的 n3 亦为 3, 也是( $U_p$ ,  $U_i$ ,  $U_T$ )三个变量。
```

```
%下面的程序为求取 RBF 函数中的指数
ps = zeros(n1, n);
kerneloption=ones(1, n2) * kerneloption;
metric = diag(1./kerneloption.^2);
ps = x * metric * xsup';
[nps, pps]=size(ps);
normx = sum(x.^2 * metric, 2);
normxsup = sum(xsup.^2 * metric, 2);
ps = -2 * ps + repmat(normx, 1, pps) + repmat(normxsup', nps, 1);
%ps 为 RBF 函数的指数值
K = exp(-ps/2); %得到 RBF 函数的具体数值

w='训练 RBF 核支持向量机得到的权系数的计算值'
b=0.2383; %支持向量机的偏置
y=K * w+b; %根据公式得到融合后的计算值
程序结束
```

说明：

(1) 根据 SVM 训练的结果，在程序中输入支持向量 xsup，在本程序中，支持向量就是学习样本的输入向量，即 $xsup=[U_p \quad U_i \quad U_t]$ ，但是在其它的 SVM 中则不一定，请读者注意。

(2) 输入待融合向量 x，即传感器现场实际测量的值， $x=[U_{p_meas} \quad U_{i_meas} \quad U_{t_meas}]$ ；

(3) 根据 RBF 核函数 $K(x, x') = \exp\left(-\frac{\|x-x'\|^2}{2\sigma^2}\right)$ ，求取核函数矩阵的数值。式中 $x=x$ ，即待融合向量； $x'=xsup$ ，为支持向量； $\sigma=kerneloption$ ，即 RBF 的带宽。

(4) 根据支持向量机训练得到的权系数矩阵 w、偏置 b 和公式 $y=w \cdot x + b$ ，求取融合后的数值。

程序是在 Matlab 环境下编写的，移植则需要将上面的程序转化为 C 语言的程序，具体可根据所移植的环境，将其改写为 C51 的程序(单片机)或者 DSP 的 C 语言程序。

7.3.3 [示例 7-2] 使用 SVC(支持向量分类)对两组分混合气体进行定性识别

要求：以红外气体传感器为例，采用 SVC 对 SO_2 与 NO_2 两组分混合气体种类进行定性识别。

SO_2 与 NO_2 两种气体的谱线在其红外吸收区域存在交叉现象，即测量这两种气体的红外传感器存在交叉敏感。这种现象在实验标定数据中表现为当 SO_2 浓度不变而 NO_2 浓度变化时，不仅 NO_2 红外传感器的输出电压发生变化， SO_2 红外传感器的输出电压也发生变化。气体传感器的交叉敏感会给气体种类识别结果带来误差。

气体的定性识别，就是判定气体的有无。当气体的浓度小于规定的分辨率时，就认为这种气体不存在；当气体浓度大于规定的分辨率时，就认为这种气体存在。当然，气体传感器的分辨率应该满足所需分辨率的要求。本实验中，在标定时设定了每种气体浓度的阈

值，当气体浓度小于该阈值时，就认为没有该种气体，反之则认为存在该种气体。

因所识别的是两种气体的有无，其结果的组合有四种，所以这是一个多类问题。

1. 实验标定

利用红外气体传感器对 SO_2 与 NO_2 两组分混合气体进行实验标定，获得实验标定数据，即不同气体浓度条件下测定两种传感器输出电压 U_{SO_2} 与 U_{NO_2} 。

2. SVC 的训练与检验样本的制备

根据实验标定的数据，制备 SO_2 （气体 1）与 NO_2 （气体 2）两组分混合气体模式识别样本，训练样本与检验样本格式一致。每种气体的有无是根据标定时该种气体的浓度阈值确定的。对于 SO_2 和 NO_2 气体来说，我们分别设定其阈值浓度为 30 ppm 和 5 ppm，当被测气体浓度小于该阈值时认为没有这些气体存在；当其浓度分别大于其浓度阈值时，则认为存在该种气体。详细的训练样本数据见表 7-5，检验样本的数据见表 7-6。

表 7-5 SO_2 和 NO_2 两种气体传感器输出电压与气体有无识别的标定数据(训练样本)

序号	U_{SO_2}	U_{NO_2}	O_{SO_2}	O_{NO_2}	序号	U_{SO_2}	U_{NO_2}	O_{SO_2}	O_{NO_2}	序号	U_{SO_2}	U_{NO_2}	O_{SO_2}	O_{NO_2}
01	0.00	0.00	-1	-1	23	11.02	25.53	-1	1	45	38.02	31.66	1	1
02	0.59	4.52	-1	-1	24	11.20	26.92	-1	1	46	38.48	35.18	1	1
03	0.94	7.25	-1	1	25	22.65	9.74	1	-1	47	38.78	37.47	1	1
04	1.57	12.09	-1	1	26	23.24	14.26	1	-1	48	38.96	38.85	1	1
05	2.11	16.22	-1	1	27	23.60	16.99	1	1	49	37.56	16.15	1	-1
06	2.57	19.74	-1	1	28	24.23	21.83	1	1	50	38.15	20.67	1	-1
07	2.86	22.03	-1	1	29	24.76	25.96	1	1	51	38.50	23.40	1	1
08	3.04	23.41	-1	1	30	25.22	29.48	1	1	52	39.13	28.24	1	1
09	4.28	1.84	-1	-1	31	25.52	31.77	1	1	53	39.67	32.37	1	1
10	4.87	6.36	-1	-1	32	25.70	33.15	1	1	54	40.13	35.89	1	1
11	5.22	9.09	-1	1	33	31.45	13.52	1	-1	55	40.43	38.18	1	1
12	5.85	13.93	-1	1	34	32.03	18.05	1	-1	56	40.60	39.56	1	1
13	6.39	18.06	-1	1	35	32.39	20.77	1	1	57	40.01	17.21	1	-1
14	6.85	21.58	-1	1	36	33.02	25.61	1	1	58	40.60	21.73	1	-1
15	7.15	23.87	-1	1	37	33.55	29.74	1	1	59	40.96	24.46	1	1
16	7.33	25.25	-1	1	38	34.01	33.26	1	1	60	41.59	29.30	1	1
17	8.16	3.51	-1	-1	39	34.31	35.55	1	1	61	42.12	33.43	1	1
18	8.75	8.03	-1	-1	40	34.49	36.93	1	1	62	42.58	36.94	1	1
19	9.10	10.76	-1	1	41	35.91	15.44	1	-1	63	42.88	39.23	1	1
20	9.73	15.60	-1	1	42	36.50	19.97	1	-1	64	43.06	40.61	1	1
21	10.27	19.73	-1	1	43	36.86	22.69	1	1					
22	10.72	23.24	-1	1	44	37.49	27.54	1	1					

注：(1) U_{SO_2} 与 U_{NO_2} 分别为 SO_2 和 NO_2 两种气体传感器的输出电压(mV)，作为 SVC 的学习样本输入数据；

(2) O_{SO_2} 与 O_{NO_2} 分别为判定 SO_2 和 NO_2 两种气体有无的数据，-1 表示没有该种气体，1 表示存在该种气体，作为学习样本的期望输出数据。

表 7-6 SO_2 和 NO_2 两种气体传感器输出电压与气体有无识别的标定数据(检验样本)

序号	U_{SO_2}	U_{NO_2}	O_{SO_2}	O_{NO_2}	序号	U_{SO_2}	U_{NO_2}	O_{SO_2}	O_{NO_2}
01	1.97	2.33	-1	-1	07	5.22	9.09	-1	1
02	3.21	5.65	-1	-1	08	7.95	24.21	-1	1
03	2.71	8.01	-1	1	09	37.96	19.23	1	-1
04	4.53	22.05	-1	1	10	39.85	23.98	1	1
05	4.49	3.41	-1	-1	11	42.44	39.71	1	1
06	4.27	4.74	-1	-1					

注: (1) U_{SO_2} 与 U_{NO_2} 分别为 SO_2 和 NO_2 两种气体传感器的输出电压(mV), 作为 SVC 的检验样本输入数据;

(2) O_{SO_2} 与 O_{NO_2} 分别为判定 SO_2 和 NO_2 两种气体有无的数据, -1 表示没有该种气体, 1 表示存在该种气体, 作为检验样本的期望输出数据。

从上面两个表格中的数据可以看出, 两组分混合气体有四种模式(气体 1 和 2 都没有、有气体 1 和没有气体 2、没有气体 1 和有气体 2、气体 1 和 2 都有), 即

(1) 模式 1: $\varphi(\text{SO}_2) < 30 \text{ ppm}$, $\varphi(\text{NO}_2) < 5 \text{ ppm}$, 即 SO_2 和 NO_2 都没有, 输出结果为 $(-1, -1)$;

(2) 模式 2: $\varphi(\text{SO}_2) < 30 \text{ ppm}$, $\varphi(\text{NO}_2) \geq 5 \text{ ppm}$, 即没有 SO_2 和只有 NO_2 , 输出结果为 $(-1, 1)$;

(3) 模式 3: $\varphi(\text{SO}_2) \geq 30 \text{ ppm}$, $\varphi(\text{NO}_2) < 5 \text{ ppm}$, 即只有 SO_2 和没有 NO_2 , 输出结果为 $(1, -1)$;

(4) 模式 4: $\varphi(\text{SO}_2) \geq 30 \text{ ppm}$, $\varphi(\text{NO}_2) \geq 5 \text{ ppm}$, 即 SO_2 和 NO_2 都有, 输出结果为 $(1, 1)$ 。

根据上述四种模式的输出结果(O_1, O_2), 可以将两组分混合气体的四种模式编成四类, 见表 7-7。

表 7-7 两个 SVC 的 O 值选取原则

两种气体的浓度阈值	$\varphi(1) < 30$ $\varphi(2) < 5$	$\varphi(1) < 30$ $\varphi(2) \geq 5$	$\varphi(1) \geq 30$ $\varphi(2) < 5$	$\varphi(1) \geq 30$ $\varphi(2) \geq 5$
输出结果	$(-1, -1)$	$(-1, 1)$	$(1, -1)$	$(1, 1)$
SVM 分类标识	1	2	3	4

由上面的叙述可以看出, 这是一个多类问题。目前用支持向量机处理多类问题主要采用两种策略, 一种是“一对一(one against one, 1A1)”, 另一种是“一对多(one against all, 1AA)”。

这两种策略都采用“一个机器”的方法, 即通过求解单个最优化问题来构建一个多类的 SVC, 采用的都是“分划和获胜”方法, 将一个多类问题分解为几个二值子问题, 并为每个子问题构建一个标准的 SVC。二者不同的是, 1AA 方法是为每一类构建一个 SVC, 从所有其它类的样本中训练区分属于一个单类的样本。1A1 方法是为每个“类对”构建一个 SVC。对于 N 类的问题, 1A1 的方法共要产生 $N(N-1)/2$ 个分类器。当用在测试的时候, 对于一个测试数据, 每个分类器会对该数据给出一个投票, 最后这个数据属于获得最多选票的那一类, 这一过程相当于加权投票的过程。

本实验分别采用一种 1A1 和 1AA 策略的多类 SVC(Multi SVC, MSVC)来实现两组分混合气体四种模式的识别。采用 MSVC 可消除气体传感器的交叉敏感，提高识别准确率。

3. SVC 结构的确定

将训练样本和检验样本分别输入 SVC 中，当 SVC 训练完毕并且检验结果的误差满足实测要求时，表明 SVC 可用来实测；反之，则要调整 SVC 核函数参数，重新确定 SVC 结构。当 SVC 结构确定之后，表明 SVC 的乘子及偏移量已经确定，即用于模式识别的回归模型已经确定。

4. 两组分混合气体四种模式的识别

将 SO_2 与 NO_2 两种红外气体传感器的实测输出电压代入已经确定的回归模型，即可获得被测两组分混合气体成分的模式。使用多类 SVC 进行气体识别的流程图如图 7-9 所示。

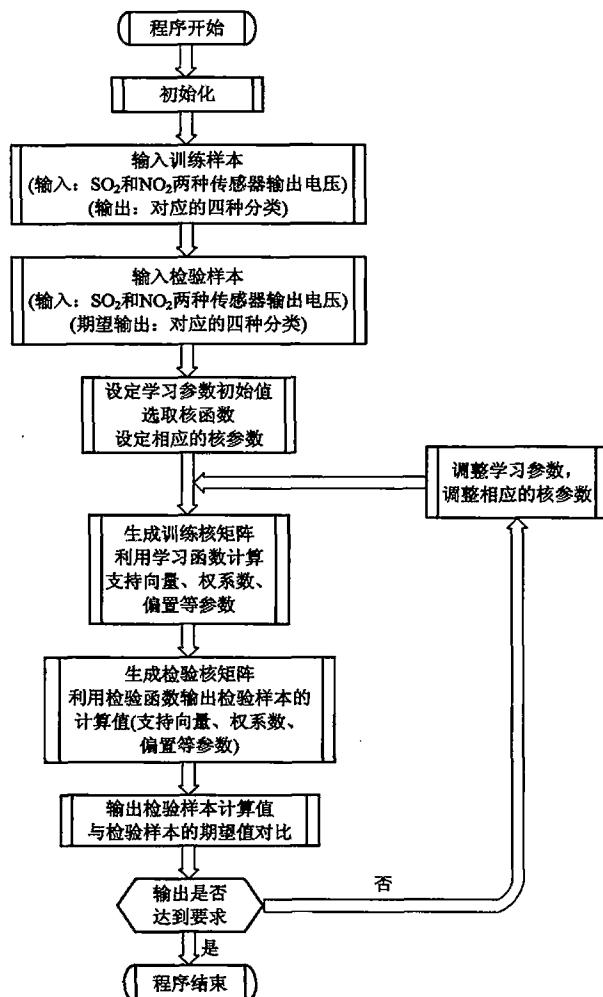


图 7-9 使用多类 SVC 进行气体识别流程图

用于两种混合气体定性识别的支持向量机 Matlab 源程序有以下两个。

程序 1 采用 one against one 策略的多类 SVC 源程序：

```
%SVM MultiClassification, 支持向量机的多类识别
%one against one strategy
%程序初始化
clc; close all; clear all;
%----- Training Data Set -----
%输入训练数据
%红外气体传感器的输出电压, 单位为 mV, 前面为 SO2, 后面为 NO2。S_output 作为训练输入样本
S_output=[0.00 0.00; 0.59 4.52; 0.94 7.25; 1.57 12.09; 2.11 16.22; 2.57 19.74; 2.86 22.03;...];
%前面数值表示 SO2 气体的识别情况, 以 30 ppm 为界, -1 表示没有该种气体, 1 表示有该种气体; 后面数值表示 NO2 气体的识别情况, 以 5 ppm 为界, -1 表示没有该种气体, 1 表示有该种气体
%上述气体的分类情况: 根据分类规则, 共可分为 4 类, 分别用(-1, -1), (-1, 1), (1, -1), (1, 1)表示, (-1, -1)表示 SO2 和 NO2 都没有, 为种类 1; (-1, 1)表示 SO2 无, NO2 有, 为种类 2; %(1, -1)表示 SO2 有, NO2 无, 为种类 3; (1, 1)表示 SO2 和 NO2 都有, 为种类 4
G_con=[-1-1; -1-1; -1 1; -1 1; -1 1; -1 1; -1 1;-1-1; -1-1;...];
%G_class 表示训练输出样本, 即训练输出的期望值
G_class=[1 1 2 2 2 2 2 2 1 1 ...]';
xapp=S_output;
yapp=G_class;
%学习过程与学习参数
c = 1000;
lambda = 1e-7;
%kerneloption=1; kernel='poly';
%当选用多项式核时的设置
kerneloption=10; %选高斯核函数时的标准偏差。当选用其它核函数的时候, Kerneloption 的数值为 1
kernel='gaussian'; %高斯核函数
verbose = 1;
nbclass=4; %共分为 4 类
kerneloptionm. matrix=svmkernel(S_output, kernel, kerneloption);
[xsup, w, b, nbsv, classifier, pos]=svmmulticlass ([] , G_class, nbclass, c, lambda, 'numerical', kerneloptionm, verbose);
%测试数据
S_test=[1.97 2.33;3.21 5.65;2.71 8.01;4.53 22.05;4.49 3.41;4.27 4.74;5.22 9.09;7.95
24.21;37.96 19.23;39.85 23.98;42.44 39.71];
G_test=[-1-1; -1-1; -1 1; -1 1;-1-1;-1-1; -1 1; -1 1; 1-1;1 1;1 1];
Gtest_expect=[1 1 2 2 1 1 2 2 3 4 4]';
```

```
%训练样本预测输出结果
kerneloptionml.matrix=svmkernel(S_output, kernel, kerneloption, xapp(pos, :));
%检验样本预测输出结果
[ypred1, maxi] = svmmultival([], [], w, b, nbsv, 'numerical', kerneloptionml);
kerneloptionm2.matrix=svmkernel(S_test, kernel, kerneloption, xapp(pos, :));
[ypred2, maxi] = svmmultival([], [], w, b, nbsv, 'numerical', kerneloptionm2);
%训练和检验样本的期望输出数据与预测输出数据
Result1=[yapp ypred1]
Result2=[Gtest_expect ypred2]
%程序结束
```

程序 2 采用 one against all 策略的多类 SVC 源程序：

```
%程序初始化与训练样本输入数据同程序 1
nbclass=4;
[n1, n2]=size(xapp);
%Learning and Learning Parameters
c = 500;
lambda = 1e-7;
kerneloption=1; kernel='poly'; %多项式核;
%kerneloption=10; kernel='gaussian';
verbose = 1;
%----- one against all algorithms -----
[xsup, w, b, nbsv]=svmmulticlassoneagainstall(xapp, yapp, nbclass, c, lambda, kernel, kerneloption, verbose);
%测试数据, 同程序 1
[ypred1, maxi] = svmmultival(xapp, xsup, w, b, nbsv, kernel, kerneloption);
[ypred2, maxi] = svmmultival(S_test, xsup, w, b, nbsv, kernel, kerneloption);
Result1=[yapp ypred1]
Result2=[Gtest_expect ypred2]
disp('这是一个采用 "one against all"策略的多类分类算法');
%程序结束
```

5. SVM 融合输出结果

采用多项式核的检验样本输出结果：

期望输出	1	1	2	2	1	1	2	2	3	4	4
预测输出	1	1	2	2	1	1	2	2	3	4	4

采用 Gaussian 核的检验样本输出结果：

期望输出	1	1	2	2	1	1	2	2	3	4	4
预测输出	1	1	2	2	1	1	2	2	3	4	4

由以上两个表格可以看出，使用多项式核和 Gaussian 核的 SVM 都可以对两种气体进行很好的辨识。需要注意的是，采用 Gaussian 核的时候，需要调整核函数的带宽 σ 。 σ 取值

不当, SVM 的分类效果会不好, 如:

期望输出		1	1	2	2	1	1	2	2	3	4	4
预测输出	$\sigma=100$	1	1	2	2	1	1	2	2	3	4	4
	$\sigma=10$	1	1	2	2	1	1	2	2	3	4	4
	$\sigma=2$	1	1	2	2	1	1	2	2	3	4	4
	$\sigma=1.5$	1	1	2	2	1	1	2	2	3	4	4
	$\sigma=1.1$	4	4	2	2	1	4	2	2	3	4	4
	$\sigma=1$	4	4	2	2	4	4	2	2	3	4	4
	$\sigma=0.5$	4	4	4	4	4	4	2	2	4	4	4

采用 1A1 和 1AA 两种策略的 MSVC 方法, 对于本实验来说输出相同, 都可以对两种气体进行很好的定性识别。

参 考 文 献

- [1] 白鹏, 谢文俊, 刘君华. 基于 SVM 的混合气体红外光谱分析关键技术研究[J]. 光子学报, 2008, 37(3): 566 - 572
- [2] Haifeng Ge, Junhua Liu. Identification of gas mixtures by a distributed support vector machine network and wavelet decomposition from temperature modulated semiconductor gas sensor[J]. Sensors and Actuators B, 2006, 117: 408 - 414
- [3] 葛海峰, 林继鹏, 刘君华, 等. 基于支持向量机和小波分析的气体辨识[J]. 仪器仪表学报 2006, 27(6): 573 - 578
- [4] Bai Peng, Liu junhua. Mixture gas component concentration analysis based on support vector machine and infrared spectrum[J]. Chinese Optics Letters, 2006, 4 (4): 243 - 246
- [5] 白鹏, 谢文俊, 刘君华. 支持向量机在混合气体种类光谱识别中的应用[J]. 光电工程, 2006, 33(8): 37 - 43
- [6] 白鹏, 刘君华. SVM 在混合气体光谱分析中的应用[J]. 仪器仪表学报, 2006, 27 (10): 1242 - 1247
- [7] Li Jun, Liu Junhua. Identification of dynamic systems using support vector regression neural network[C]. 2006, LNCS3972: 934 - 942
- [8] 肖丹, 张勇, 刘君华. 基于支持向量机的数据融合仪的远程共享[J]. 电子测量与仪器学报, 2006 年增刊
- [9] 白鹏, 刘君华. Compensated algorithm of sensor characteristic based support vector machine[C]. The 6th International Symposium on Test and Measurement, 2005, 6
- [10] 刘玉烨, 林继鹏, 刘君华. SVM(支持向量机)在智能传感器系统中实现的研究[C]. 第九届全国敏感元件与传感器学术会议论文集[C]. 北京: 国防工业出版社, 2005:

656 – 659

- [11] 林继鹏, 刘君华, 凌振宝. 并行支持向量机算法及其应用[J]. 吉林大学学报(信息科学版), 2004, 22(5): 453 – 457
- [12] 周鸣争, 汪军. 基于支持向量机的传感器非线性误差校正[J]. 电子科技大学学报, 2006(2): 242 – 245
- [13] 吴德会. 基于最小二乘支持向量机的传感器非线性动态补偿[J]. 仪器仪表学报, 2007(6): 1018 – 1023

第8章 粒子群优化算法与其在智能传感器系统中的应用

本章内容

- 粒子群优化算法发展与应用概况
- 粒子群优化算法的基础知识
- 粒子群优化算法的发展与改进简介
- [示例 8-1]遍历优化 LS-SVM 模型参数
- [示例 8-2]基于粒子群优化算法优化一个参数
- [示例 8-3]基于粒子群优化算法优化两个参数

最优化问题，是研究在满足一定约束条件下，寻找一组参数以便某些最优性能指标得以实现，也就是在众多可行方案中找到一个最优的方案。最优化问题广泛存在于各个领域，如在智能传感器系统中建立逆模型的神经网络法、最小二乘支持向量机算法等，都存在最优化问题。其中神经网络算法的隐层节点数需要优化选取，最小二乘支持向量机算法中的惩罚因子及核函数参数均需优化选择。

粒子群优化(Particle Swarm Optimization)算法是近年迅速发展起来的一种智能优化算法。该算法和遗传算法有一些相似的地方，比如二者都是基于群体进行优化的方法，都是将系统初始化为一组随机解，通过迭代搜索最优值。但是，粒子群优化算法要比遗传算法更为简单、更易实现，需要调整的参数较少，因此粒子群优化算法从它诞生之初就引起学者的广泛关注，成为国际上一个新的研究热点。

本章结合改善传感器稳定性/抗干扰能力的多传感器技术模型法，介绍了基于粒子群优化算法结合最小二乘支持向量机的一种智能化技术的工作原理与使用方法，着重讨论粒子群优化算法与最小二乘支持向量机应用于传感器智能化功能的实现，如何将粒子群优化算法与最小二乘支持向量机相结合建立逆模型以降低干扰量的影响。

8.1 粒子群优化算法发展与应用概况

8.1.1 群智能

一只蜜蜂或蚂蚁的行动能力非常有限，它几乎不可能独立存在于自然世界中，而多个蜜蜂或蚂蚁形成的群体则具有非常强的生存能力，且这种能力不是多个个体之间的能力通过简单叠加所获得的。社会性动物群体所拥有的这种特性能帮助个体很好地适应环境，通过群体内部个体之间的信息交互，个体所获得的信息远多于自身感官直接获取的信息，而且群内个体还能处理信息，并根据所获得的信息(包括环境信息和附近其他个体的信息)改

变自身的一些行为模式和规范。这样就使得群体涌现出一些单个个体所不具备的能力和特性，尤其是对环境的适应能力。这种对环境变化所具有的适应能力可以被认为是一种智能，也就是说动物个体通过聚集成群而涌现出了智能。因此，Bonabeau 将群智能(Swarm Intelligence, SI)定义为：无智能或简单智能的主体通过任何形式的聚集协同而表现出智能行为的特性。这里我们关心的不是个体之间的竞争，而是它们之间的协同。

8.1.2 群智能的主要算法

目前，群智能研究领域有两种主要的算法：蚁群优化算法和粒子群优化算法。蚁群优化算法(Ant Colony Optimization, ACO)由 Colorni A、Dovigo M 和 Maniezzo V 于 1991 年提出，是仿照蚁群在寻找食物过程中表现出来的寻优方式得到的一种优化算法，已被广泛应用于组合优化问题中。在车辆调度、机器人路径规划、路由算法设计领域均取得了良好效果，且不断取得新的成果。ACO 算法具有广泛实用价值，曾一度成为群智能的代名词。

粒子群优化算法源于 1987 年 Reynolds 对鸟群社会系统的仿真研究，在 1995 年意大利的 Kennedy 和 Eberhart 创造性地完成了粒子群优化算法(Particle Swarm Optimization, PSO)。PSO 的优势在于简单且容易实现，同时又有深刻的智能背景。短短几年时间，PSO 算法获得了很大发展，并在一些领域得到应用，如 PID 调速、调温，液压伺服等控制系统，与各种神经网络结合(遗传、模糊、径向基、小波等)进行火灾图像识别、故障诊断等，目前已被“国际演化计算会议(CEC)”列为讨论专题之一。

8.1.3 群智能算法的特点

群智能方法易于实现，算法中仅涉及各种基本数学操作，其数据处理过程对 CPU 和内存要求不高，且该方法只需目标函数的输出值，而无需其梯度信息。发展与应用证明群智能方法是一种能够有效解决大多数全局优化问题的新方法。

总体而言，群智能算法仍然处于新兴发展与改进完善阶段。目前其算法的数学理论基础还相对薄弱，涉及的各种参数设置仍多用经验方法确定，对具体问题与应用环境的依赖性较大。提高算法的性能一直是研究者最为关心的问题之一，将两个甚至两个以上的群智能算法相结合形成群智能混合算法，或者将群智能与其他先进技术，如神经网络、模糊逻辑、紧急搜索和最小二乘支持向量机相融合，以改善其自身或相应技术方法的性能正越来越受到重视。

事实上，群智能方法能够用于解决大多数优化问题，或者将其转化为优化求解的问题。其应用领域已扩展到多目标优化、数据分类、聚类、模式识别、流程规划、信号处理、系统辨识、控制与决策等方面。

8.2 粒子群优化算法的基础知识

自然界中各种生物体均具有一定的群体行为，而人工生命的主要研究领域之一就是探索自然界生物的群体行为从而在计算机上构建其群体模型。通常，群体行为可以由几条简单的规则进行建模。如鱼群、鸟群等，虽然每一个个体具有非常简单的行为规则，但群体的行为却非常复杂。Reynolds 将这种类型的个体称为 boids，并使用计算机图形动画对复杂的群体行为进行仿真。他在仿真中采用了下列三条简单规则：

- (1) 飞离最近的个体，以避免碰撞；
- (2) 飞向目标；
- (3) 飞向群体的中心。

群体内每个个体的行为可采用上述规则进行描述，这是微粒群算法的基本概念之一。

Boyd 和 Richerson 在研究人类的决策过程时，提出了个体学习和文化传递的概念。根据他们的研究结果，人们在决策过程中使用两类重要的信息：一是自身的经验，二是其他人的经验。也就是说，人们根据自身的经验和他人的经验进行自己的决策。这是微粒群算法的另一基本概念。

微粒群算法最早是在 1995 年由美国社会心理学家 James Kennedy 和电气工程师 Russell Eberhart 共同提出的，其基本思想是受他们早期对许多鸟类的群体行为进行建模与仿真研究结果的启发，而他们的模型及仿真算法主要利用了生物学家 Frank Heppner 的模型。

Frank Heppner 的鸟类模型与其他类模型在反映群体行为方面有许多相同之处，所不同之处在于：鸟类被吸引飞向栖息地。在仿真中，一开始每一只鸟均无特定目标进行飞行，直到有一只鸟飞到栖息地，当设置期望栖息比期望留在鸟群中具有较大的适应值时，每一只鸟都将离开群体而飞向栖息地，随后就自然地形成了鸟群。

由于鸟类使用简单的规则确定自己的飞行方向与飞行速度（实质上，每一只鸟都试图停在鸟群中而又不相互碰撞），当一只鸟飞离鸟群而飞向栖息地时，将导致它周围的其他鸟也飞向栖息地。这些鸟一旦发现栖息地，将降落在此，驱使更多的鸟落在栖息地，直到整个鸟群都落在栖息地。

由于 James Kennedy 和 Russell Eberhart 所具有的专业背景，很容易理解他们为什么会对 Heppner 的鸟类模型感兴趣。鸟类寻找栖息地与对一个特定问题寻找解很类似，已经找到栖息地的鸟引导它周围的鸟飞向栖息地的方式，增加了整个鸟群都找到栖息地的可能性，也符合信念的社会认知观点。

Eberhart 和 Kennedy 对 Heppner 的模型进行了修正，以使微粒能够飞向解空间并在最好解处降落。其关键在于如何保证微粒降落在最好解处而不降落在其他解处，这就是信念的社会性及智能性所在。

信念具有社会性的实质在于个体向它周围的成功者学习。个体与周围的其他‘同类’比较，并模仿其优秀者的行为。将这种思想用算法实现能够形成一种新的最优化算法。

要解决上述问题，关键在于在探索（寻找一个好解）和开发（利用一个好解）之间寻找一个好的平衡。太小的探索导致算法收敛于早期所遇到的好解处，而太大的开发会使算法不收敛。

另一方面，需要在个性与社会性之间寻求平衡，也就是说，既希望个体具有个性化，像鸟类模型中的鸟不互相碰撞，又希望其知道其他个体已经找到的好解并向他们学习，即社会性。

Eberhart 和 Kennedy 较好地解决了上述问题，提出了微粒群优化算法（Particle Swarm Optimization, PSO）。

8.2.1 基本粒子群优化算法

Kennedy 和 Eberhart 在 1995 年的 IEEE 国际神经网络学术会议上正式发表了题为《Particle Swarm Optimization》的文章，标志着微粒群算法的诞生。

对粒子 i 在第 j 维子空间中运动的速度，由下述基本微粒群算法进行调整：

$$v_{ij}(t+1) = v_{ij}(t) + c_1 r_{1j}(t)[p_{ij}(t) - x_{ij}(t)] + c_2 r_{2j}(t)[p_{ej}(t) - x_{ij}(t)] \quad (8-1)$$

$$x_{ij}(t+1) = x_{ij}(t) + v_{ij}(t+1) \quad (8-2)$$

式中： $i=1, 2, \dots, M$, M 是该群体中粒子的总数，“ i ”表示第几个粒子；下标“ j ”表示微粒的第 j 维，即算法所优化的第 j 个参数； t 表示此时优化的代数(次数)； $x_{ij}(t)$ 为 t 时刻粒子 i 在 j 维子空间的位置； $v_{ij}(t)$ 表示 t 时刻粒子 i 在 j 维子空间的速度，定义为每次迭代中粒子移动的距离； c_1 、 c_2 为加速因子，通常在 $0 \sim 2$ 间取值，其中 c_1 称为认知参数， c_2 称为社会参数； r_{1j} 、 r_{2j} 为两个 $[0 \sim 1]$ 之间变化的相对独立的随机函数； $p_{ij}(t)$ 为粒子个体 i 的历史最好解(个体最优位置)的 j 维值，即单个粒子 i 在所优化的第 j 个参数中的历史最优解； $p_{ej}(t)=\min\{p_{ij}(t)\}$ 为所有粒子在 t 时刻的历史最好解(群体最优位置)的 j 维值，即所有粒子在所优化的第 j 个参数中的历史最优解。

式中第二项是将当前粒子的位置与该粒子的历史最优位置之差用 r_{1j} 随机函数进行一定程度的随机化，作为改变粒子当前位置向自身历史最优位置运动的调整分量。第三项是将当前粒子的位置与整个群体的历史最优位置之差用 r_{2j} 随机函数进行一定程度的随机化，作为改变当前粒子位置向群体最优位置运动的调整分量。

在每一次迭代过程中，每个粒子都需要根据目标函数来计算其适应值大小，目标函数可以用均方误差、方差、标准差等形式来表示，其数值称为适应值。再根据适应值来确定当前粒子最优位置 $p_{ij}(t)$ 及群体最优位置 $p_{ej}(t)$ ，然后根据式(8-1)、式(8-2)调整各个粒子的速度及位置。其结束条件为迭代次数达到设定值或者群体迄今为止搜索到的最优位置满足预设最小适应值。一般情况下，最大迭代次数设定为 100，预设适应值为零。

设 $f(x)$ 为最小化的目标函数，则微粒 i 的当前最好位置由下式确定：

$$p_i(t+1) = \begin{cases} p_i(t), & f(x_i(t+1)) \geq f(p_i(t)) \\ x_i(t+1), & f(x_i(t+1)) < f(p_i(t)) \end{cases} \quad (8-3)$$

设群体中的微粒数为 M ，则群体中所有微粒所经历过的最好位置 $p_g(t)$ ，即群体最优位置可表示为：

$$\begin{aligned} p_g(t) &\in \{p_0(t), p_1(t), \dots, p_M(t)\} \mid f(p_g(t)) \\ &= \min\{f(p_0(t)), f(p_1(t)), \dots, f(p_M(t))\} \end{aligned} \quad (8-4)$$

从上述微粒进化方程可以看出， c_1 调节微粒飞向自身最好位置方向的步长， c_2 调节微粒向全局最好位置飞行的步长。为了减少在进化过程中微粒离开搜索空间的可能性， v_{ij} 通常限定于一定范围内，即 $v_{ij} \in [-v_{\max}, v_{\max}]$ 。如果问题的搜索空间限定在 $[-x_{\max}, x_{\max}]$ 内，则可设定 $v_{\max}=k \cdot x_{\max}$ ， $0.1 \leq k \leq 1.0$ 。

8.2.2 标准粒子群优化算法

为了改善基本 PSO 算法的收敛性能，Y. Shi 与 R. C. Eberhart 在 1998 年的 IEEE 国际进化计算学术会议上发表了题为《A Modified Particle Swarm Optimizer》的论文，首次在速度进化方程中引入惯性权重，即

$$v_{ij}(t+1) = wv_{ij}(t) + c_1 r_{1j}(t)[p_{ij}(t) - x_{ij}(t)] + c_2 r_{2j}(t)[p_{ej}(t) - x_{ij}(t)] \quad (8-5)$$

式中 w 为惯性权重因子，其值非负，值的大小影响整体寻优能力。因此，基本 PSO 算法是惯性权重 $w=1$ 的特殊情况。惯性权重 w 使微粒保持运动惯性， w 与 v_{ij} 的乘积表示粒子依据自身的速度进行惯性运动所占的比重。

引入惯性权重 w 可清除基本 PSO 算法对 v_{\max} 的需要, 因为 w 本身具有维护全局和局部搜索能力的平衡的作用。这样, 当 v_{\max} 增加时, 可通过减少 w 来达到平衡搜索; 而 w 的减少可使得所需的迭代次数变小。从这个意义上讲, 可以将 v_{\max} 固定为每维变量的变化范围, 只对 w 进行调节。

对全局搜索而言, 好的方法通常是在前期有较高的探索能力以得到合适的种子, 而在后期有较高的开发能力以加快收敛速度。为此, 可将 w 设定为随着进化而线性减小, 例如由 0.9 减小到 0.4 等。Y. Shi 和 R. C. Eberhart 的仿真实验结果也表明, w 线性减小取得了较好的实验结果。

目前, 有关 PSO 算法的研究大多以带惯性权重的 PSO 算法为基础进行扩展和修正。为此, 在大多文献中将带惯性权重的 PSO 算法称为 PSO 算法的标准版本, 或简称标准 PSO 算法, 而将基本 PSO 算法称为 PSO 的初始版本。

8.2.3 粒子群优化算法流程

基本微粒群算法的流程如图 8-1 所示。

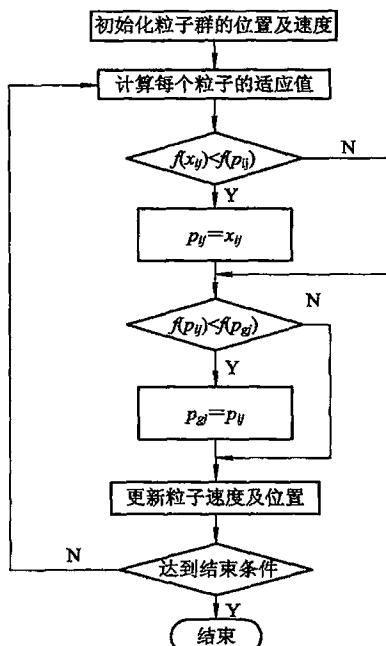


图 8-1 微粒群算法的流程图

第一步: 初始化所有粒子(群体规模为 M)。在允许范围内随机设置粒子的初始位置、速度及群体最优值, p_{ij} 表示各粒子的起始位置, 群体最优值为 p_{gi} 。

第二步: 评价每个粒子的适应值, 即分别对每个粒子计算目标函数值, 目标函数可以是均方误差、方差、标准差(均方根误差)等。

第三步: 对每个粒子, 将其适应值 $f(x_{ij})$ 与其历史最优位置 p_{ij} 的适应值 $f(p_{ij})$ 进行比较, 如果较好, 即 $f(x_{ij}) < f(p_{ij})$, 则将其作为当前最优位置 p_{ij} , 即令 $p_{ij} = x_{ij}$ 。

第四步：对每个粒子，将当前最优位置 p_{ij} 的适应值 $f(p_{ij})$ 与群体历史最优位置 p_{gi} 的适应值 $f(p_{gi})$ 进行比较，如果优于 p_{gi} ，即 $f(p_{ij}) < f(p_{gi})$ ，则将其作为群体最优位置 p_{gi} ，令 $p_{gi} = p_{ij}$ ，并重新设置 p_{gi} 的索引号。

第五步：根据粒子群速度和位置更新方程，即式(8-1)、式(8-2)，来调整粒子的速度和位置。

第六步：检查终止条件（通常为达到最大迭代次数或者足够好的适应值），若满足终止条件，或者最优解停止不再变化，终止迭代，否则返回第二步。

8.3 粒子群优化算法的发展与改进简介

在微粒群算法的改进方面，首先是由 Kennedy 和 Eberhart 在 1997 年提出的二进制 PSO 算法，为 PSO 算法与遗传算法的性能比较提供了一个有用的方式，该方法可用于神经网络的结构优化。

其次，为了提高算法的收敛性能，Shi 和 Eberhart 于 1998 年对 PSO 算法的速度项引入了惯性权重 w ，并提出在进化过程中动态调整惯性权重以平衡收敛的全局性和收敛速度，该进化方程已被相关学者称之为标准 PSO 算法。Clerc 于 1999 年在进化方程中引入收缩因子以保证算法的收敛性，同时使得速度的限制放松。有关学者已通过代数方法对此方法进行了详细的算法分析，并给出了参数选择的指导性建议。

Angeline 于 1999 年借鉴进化计算中的选择概念，将其引入 PSO 算法中。通过比较各个微粒的适应值淘汰掉差的微粒，而将具有较高适应值的微粒进行复制以产生等数额的微粒来提高算法的收敛性。而 Lovbjerg 等人进一步将进化计算机制应用于 PSO 算法，如复制、交叉等，给出了算法交叉的具体形式，并通过典型测试函数的仿真实验说明了算法的有效性。

为了提高 PSO 算法收敛的全局性，保证微粒的多样性是其关键。为了保证进化过程中群体中微粒的多样性，Suganthan 在标准 PSO 算法中引入了空间邻域的概念，将处于同一个空间领域的微粒构成一个子微粒群分别进行进化，并随着进化动态地改变选择阈值以保证群体的多样性。Kennedy 引入邻域拓扑的概念来调整邻域的动态选择，同时引入社会信念将空间邻域与邻域拓扑中的环拓扑相结合以增加邻域间的信息交流，提高群体的多样性。Lovbjerg 等人于 2001 年将遗传算法中的子群体概念引入 PSO 算法中，同时引入繁殖算子以进行子群体的信息交流。

人们在 PSO 算法的行为分析和收敛性分析方面进行了大量的研究工作。首先是采用代数方法对几种典型的 PSO 算法的运行轨迹进行了分析，给出了保证收敛性的参数选择范围。在收敛性方面，Frans van den Bergh 引用 Solis 和 Wets 关于随机性算法的收敛准则，证明了标准 PSO 算法不能收敛于全局最优解，甚至于局部最优解，同时证明了保证收敛的 PSO 算法能够收敛于局部最优解，而不能保证收敛于全局最优解。

PSO 算法最早应用于人工神经网络的训练方法，Kennedy 和 Eberhart 成功地将 PSO 算法应用于分类 XOR 问题的神经网络训练。随后，PSO 算法在函数优化、约束优化、极大极小问题、多目标优化等问题中均得到了成功的应用，特别是在电力系统、集成电路设计、

系统辨识、状态估计等问题中的应用均有报导。

从国内外关于微粒群算法研究的相关文献以及进化算法领域的发展趋势的分析来看，微粒群算法目前主要有以下几个研究方向：

(1) 微粒群算法的改进。标准微粒群算法主要适用于连续空间函数的优化问题，如何将 PSO 算法应用于离散空间优化问题，特别是一类非数值优化问题，将是微粒群算法的主要研究方向。另外，充分吸引其他‘进化类算法的优势，以改进 PSO 算法存在的不足也是值得研究的问题。

(2) 微粒群算法的理论分析。到目前为止，PSO 算法的分析方法还很不成熟，也不够系统，存在许多不完善和未涉及到的问题。如何利用有效的数学工具对 PSO 算法的运行行为、收敛性及计算复杂性进行分析是目前的研究热点之一。

(3) 微粒群算法的生物学基础。如何根据群体进化行为完善 PSO 算法，同时分析群体智能行为；如何将其引入 PSO 算法中，以充分借鉴生物群体进化规律和进化的智能性也是目前的研究方向之一。

(4) 微粒群算法与其他‘类进化算法的比较研究。

(5) 微粒群算法的应用。算法研究的目的是应用，如何将 PSO 算法应用于更多领域，同时研究应用中存在的问题也是值得关注的热点。

8.4 应用示例

8.4.1 待优化的参数与待改善的传感器

在第 5、6、7 章中均以压阻式压力传感器受环境温度及工作电流波动干扰量的影响，致使传感器稳定性差为典型实例，讨论了基于多传感器数据融合建立逆模型的多种处理方法，以抑制对干扰量的交叉敏感，改善传感器的稳定性。第 6 章介绍的神经网络与第 7 章介绍的支持向量机技术，均存在对某些参数的优化选取问题。其中神经网络算法的收敛及预测精度与其初始权值的选择有关，因此神经网络的输出具有不可预测性和不一致性，而且容易陷入局部最小，要找到最优模型需要对权值进行优化选取。支持向量机法及最小二乘支持向量机算法虽然可以避免上述缺点，但其算法参数的选取会直接影响到模型的预测精度，而目前多采取试凑法或遍历优化进行参数选择，如高斯核函数在 0~20、惩罚因子在 0~100000 范围内优化选取，通常采用遍历试探法来进行优化选取，即分别确定一系列不同步长后，按每一个从 0~20、0~100000 范围内取二者不同的组合进行试算，将最佳结果时的二者组合数值确定为参数的优化结果。因此优化选取的计算过程很耗时耗力而且找到的未必是全局最优解。本节采用粒子群优化算法来优化选取最小二乘法支持向量机的上述两个参数，可以大幅度节省时间，高效快速达到优化目的。

1. 待优化的参数

最小二乘支持向量机(LS-SVM)与标准支持向量机(SVM)所不同的是，优化目标的损失函数是以误差的二范数来表示的，用等式约束代替了 SVM 中的不等式约束条件，提高了收敛速度。

优化问题转换为

$$\min \frac{1}{2} \|w\|^2 + \sum \frac{c}{2} \sum_{i=1}^n \xi_i^2 \quad (8-6)$$

约束条件为

$$y_i = \langle w \cdot X \rangle + b + \xi_i, \quad i = 1, 2, \dots, n \quad (8-7)$$

其中 c 为惩罚因子，控制对样本超出计算误差的惩罚程度。 ξ_i 为松弛因子。

通过引入拉格朗日函数，根据 KKT 条件求解可得到 LS-SVM 的回归函数模型为

$$f(X) = \sum_{i=1}^n a_i K(X, X_i) + b \quad (8-8)$$

其中 a_i 为拉格朗日乘子； $K(X, X_i)$ 为核函数，主要完成对样本数据的内积运算。在利用 LS-SVM 建立回归模型的时候，需要对惩罚因子 c 及核函数参数进行合理选取。

2. 待改善的传感器

实验采用压阻式压力传感器，在 25°C ~ 64°C 温度变化范围内取 10 个标定点，在 5~11 mA 供电电流变化范围内取 7 个标定点。各标定点取值如下：

物理量	各物理量的标定点的数值	标定点个数
温度 $T/^\circ\text{C}$	25 31 35 39 43 49 53 55 58 64	$m=10$
压力 P/MPa	0 0.1 0.2 0.3 0.4 0.5 0.6	$n=7$
电流 I/mA	5 6 7 8 9 10 11	$w=7$

在上述各温度状态和供电电流条件下，测定压阻式压力传感器的输入输出特性 ($P - U_P$)。相应的三维标定值有 $m \times n \times w = 490$ 个，在表 8-1 中仅列出其中的部分值。得到的压力传感器的输出电压等温曲面图如图 8-2 所示。

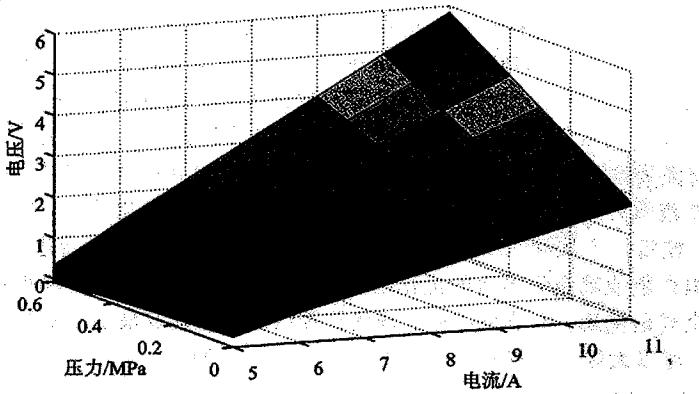


图 8-2 压力传感器的等温曲面图

表 8-1 部分标定实验数据

序号	温度 /℃	标定压力 /MPa	电流 /mA	电压 /V	序号	温度 /℃	标定压力 /MPa	电流 /mA	电压 /V
1	58	0	9	0.016	21	39	0	9	0.334
2	64	0	11	0.022	22	25	0.1	6	0.834
3	58	0	10	0.028	23	43	0.1	8	0.918
4	31	0	11	0.028	24	31	0.1	8	0.94
5	58	0	8	0.056	25	64	0.2	8	0.52
6	58	0	11	0.076	26	58	0.2	5	1.068
7	31	0	9	0.082	27	55	0.3	5	1.492
8	58	0	7	0.092	28	53	0.3	5	1.498
9	31	0	8	0.109	29	53	0.4	5	1.912
10	55	0	11	0.12	30	35	0.4	10	3.674
11	64	0	6	0.168	31	55	0.5	6	2.727
12	53	0	10	0.21	32	35	0.5	9	4.079
13	53	0	8	0.22	33	64	0.6	5	2.719
14	53	0	5	0.245	34	31	0.6	5	2.758
15	35	0	8	0.276	35	49	0.6	7	3.727
16	25	0	8	0.317	36	39	0.6	8	4.309
17	25	0	5	0.326	37	25	0.6	10	5.337
18	39	0	6	0.327	38	31	0.6	11	5.74
19	39	0	5	0.328	39	39	0.6	11	5.812
20	39	0	11	0.33	40	25	0.6	11	5.841

由图 8-2 中的等温曲面可以看出, 压力传感器的输出电压与供电电流及标定压力之间的非线性关系是很明显的。而且随着温度的变化, 压力传感器的输出也存在一定的非线性漂移, 此时压力传感器的稳定性可以用温度灵敏度系数、电流灵敏度系数和零位温度系数来衡量。

由表 8-1 标定实验数据可得: 满量程输出值 $U_{FS} = 5.841 \text{ V}$, 当温度的变化范围为 $\Delta T = 39^\circ\text{C}$ (由 25°C 至 64°C), 电流由 5 mA 至 11 mA , 变动 $\Delta I = 6 \text{ mA}$ 时, 零点值的最大变化范围 $\Delta U_{0m} = 0.318 \text{ V}$ (由 0.334 V 至 0.016 V), 满量程输出值由 5.841 V 下降到 2.719 V , 输出变化范围 $\Delta U_m = 3.122 \text{ V}$ 。

模型校正前压力传感器的零位温度系数 α_0 、灵敏度温度系数(也称温度灵敏度系数) α_t 及电流影响系数(也称电流灵敏度系数) α_I 分别为

$$\alpha_0 = \frac{|\Delta U_{0m}|}{U_{FS} \times \Delta T} = \frac{0.318}{5.841 \times 39} \approx 1.40 \times 10^{-3} (\text{}/^\circ\text{C})$$

$$\alpha_s = \frac{|\Delta U_m|}{U_{fs} \times \Delta T} = \frac{|2.719 - 5.841|}{5.841 \times 39} \approx 1.37 \times 10^{-2} (\text{}/\text{°C})$$

$$\alpha_I = \frac{|\Delta U_m|}{U_{fs} \times \Delta I} = \frac{|2.719 - 5.841|}{5.841 \times 6} \approx 8.90 \times 10^{-2} (\text{/mA})$$

8.4.2 [示例 8-1] 遍历优化 LS-SVM 模型参数

本示例以图 7-8 所示的压阻式压力传感器为例，在环境温度及工作电流双重条件改变下，利用 LS-SVM 技术建立压力传感器的逆模型，选用了高斯核函数，详见式(7-14)。为了提高模型的预测准确度，采用遍历优化方法对惩罚因子 c 及核函数参数 σ 进行优化选取，对于惩罚因子 c 在 $0 \sim 100000$ 优化范围内选择遍历优化步长为 100，核函数参数 σ 在 $0 \sim 20$ 优化范围内（由于不能取零，因此初始值选为 0.01），选取优化步长为 1。优化过程如下：

(1) 将 (c, σ) 初始化为 $(0, 0.01)$ 代入 LS-SVM 建立传感器回归模型，根据检验样本计算模型预测结果的均方误差记为 error_{\min} ，将优化参数记为 $(c_{\text{best}}, \sigma_{\text{best}})$ 。

(2) 判断 σ 是否超出最大范围，若未超出，则将 σ 的值加 1；若超出，则将 σ 重新初始化为 0.01，同时将 c 的值加 100。

(3) 将改变之后的 (c, σ) 代入 LS-SVM 重建回归模型，将计算得到的模型预测结果的均方误差与 error_{\min} 进行比较，如果小于它，则将当前误差记为 error_{\min} ，同时将当前参数记为 $(c_{\text{best}}, \sigma_{\text{best}})$ 。

(4) 未达到结束条件转第(2)步。

优化过程部分程序如下：

```
%% 样本输入
trainin=...; % 训练样本输入
trainout=...; % 训练样本输出
testin=...; % 检验样本输入
testout=...; % 检验样本输出
errormin=100; % 设定参考误差
kernel = 'RBF_kernel'; % 核函数形式
type = 'f'; % 回归模型
for c=0:100:100000 % c 搜索范围
    for delta=0.01:1:20
        model = initlssvm(trainin, trainout, type, c, delta, kernel); % 模型初始化
        model = trainlssvm(model); % 训练
        Yt = simlssvm(model, testin); % 回归
        error=mse(Yt - testout);
        if error<errormin
            errormin=error;
            cbest=c;
            deltabest=delta;
        end
    end
end
```

```

end
end

```

此过程的优化流程图如图 8-3 所示。

此过程结束条件为预测结果均方误差为零或参数(c, σ)均超出最大范围。根据优化过程计算可以得到优化误差曲线如图 8-4 所示，横轴为优化步数，纵轴为训练样本的模型预测结果均方误差。

由图 8-4 可以看出，整个优化过程共需 2×10^4 步才能结束。所需时间约为 21 599 秒，得到 c 为 60 801， σ 为 0.01，根据所得参数重建传感器回归模型，可得到测试样本的预测结果均方误差为 2.74×10^{-5} 。模型预测结果如图 8-5 所示，部分预测结果数据见表 8-2。

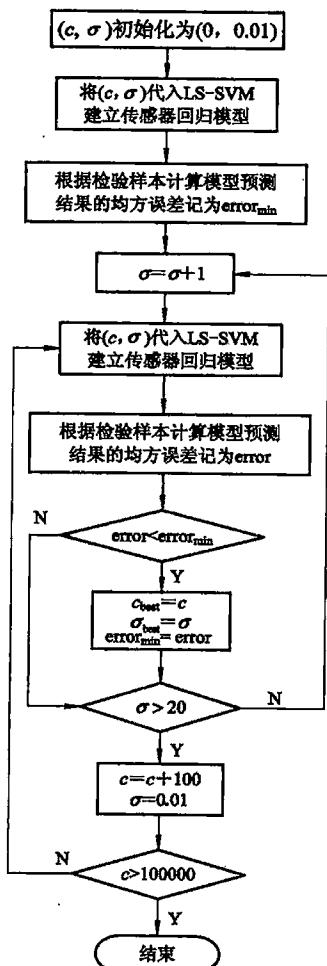


图 8-3 遍历优化过程优化流程图

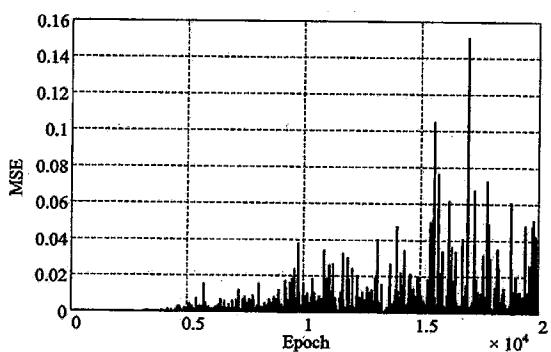


图 8-4 遍历优化误差曲线

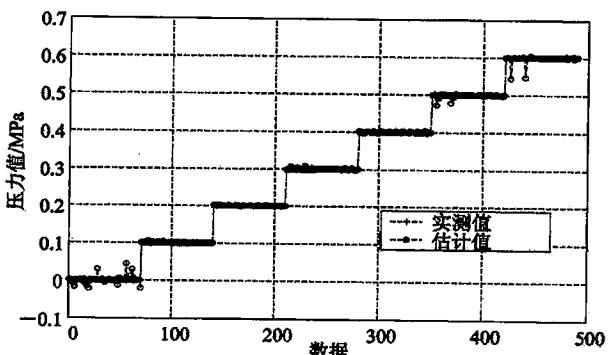


图 8-5 遍历优化所得模型预测结果

表 8-2 部分标定实验数据

序号	温度 /℃	标定压力 /MPa	电流 /mA	电压 /V	预测结果 /MPa	序号	温度 /℃	标定压力 /MPa	电流 /mA	电压 /V	预测结果 /MPa
1	58	0	9	0.016	0.0016	21	39	0	9	0.334	0.0021
2	64	0	11	0.022	0.0198	22	25	0.1	6	0.834	0.0998
3	58	0	10	0.028	0.0287	23	43	0.1	8	0.918	0.1002
4	31	0	11	0.028	0.0017	24	31	0.1	8	0.94	0.1002
5	58	0	8	0.056	0.0118	25	64	0.2	8	0.52	0.1979
6	58	0	11	0.076	0.0055	26	58	0.2	5	1.068	0.2015
7	31	0	9	0.082	0.0015	27	55	0.3	5	1.492	0.3005
8	58	0	7	0.092	0.0000	28	53	0.3	5	1.498	0.3000
9	31	0	8	0.109	0.0005	29	53	0.4	5	1.912	0.4027
10	55	0	11	0.12	0.0434	30	35	0.4	10	3.674	0.3982
11	64	0	6	0.168	0.0004	31	55	0.5	6	2.727	0.4995
12	53	0	10	0.21	0.0130	32	35	0.5	9	4.079	0.4783
13	53	0	8	0.22	0.0041	33	64	0.6	5	2.719	0.5998
14	53	0	5	0.245	0.0003	34	31	0.6	5	2.758	0.6000
15	35	0	8	0.276	0.0165	35	49	0.6	7	3.727	0.6007
16	25	0	8	0.317	0.0078	36	39	0.6	8	4.309	0.6037
17	25	0	5	0.326	0.0018	37	25	0.6	10	5.337	0.5442
18	39	0	6	0.327	0.0004	38	31	0.6	11	5.74	0.6019
19	39	0	5	0.328	0.0017	39	39	0.6	11	5.812	0.6011
20	39	0	11	0.33	0.0293	40	25	0.6	11	5.841	0.6002

由表 8-2 的预测结果可以计算出经过粒子群算法优化模型校正之后, 压力传感器在 $\Delta T=39^{\circ}\text{C}$, $\Delta I=6 \text{ mA}$ 变化范围内, 模型计算零点值的最大偏差 $|\Delta P'_{0m}|=0.0434 \text{ MPa}$; 满量程压力 $P_{\text{fs}}=0.6 \text{ MPa}$, 其模型计算值的最大偏差量 $|\Delta P'_{\text{m}}|=0.0595 \text{ MPa}$, 则模型校正后压力传感器的零位温度系数、温度灵敏度系数及电流灵敏度系数分别为:

$$\alpha_0 = \frac{|\Delta P'_{0m}|}{P_{\text{fs}} \times \Delta T} = \frac{0.0434}{0.6 \times 39} \approx 1.85 \times 10^{-3} (\text{/}^{\circ}\text{C})$$

$$\alpha_s = \frac{|\Delta P'_{\text{m}}|}{P_{\text{fs}} \times \Delta T} = \frac{0.0595}{0.6 \times 39} \approx 2.54 \times 10^{-3} (\text{/}^{\circ}\text{C})$$

$$\alpha_I = \frac{|\Delta P'_{\text{m}}|}{P_{\text{fs}} \times \Delta I} = \frac{0.0595}{0.6 \times 6} \approx 1.65 \times 10^{-2} (\text{/mA})$$

8.4.3 [示例 8-2]采用粒子群优化算法优化一个参数

采用 LS-SVM 技术建立压阻式压力传感器的逆模型，由以上分析可知，建模过程中，需要对其中的惩罚因子 c 及核函数参数 σ 进行优化选取。在本示例中，主要对核函数参数进行了优化选择，而惩罚因子 c 固定不变， c 取值为 100。优化过程如下：

- (1) 取实验数据中的一半作训练样本，全部数据作测试样本。
- (2) 初始化参数 σ, c 固定取值为 100，建立 LS-SVM 回归模型。
- (3) 由于只优化一个参数，因此设置粒子群维数为 1，粒子群中粒子的数目选取 10~30 个为宜，在本实验中选取 10 个，迭代次数选为 100。粒子群优化的参数为 σ 。根据优化参数的优化范围对粒子群的初始位置即 σ 及速度进行初始化。
- (4) 设置目标函数为模型检验样本预测结果 y_{ij} 与 LV-SVM 建立的压力传感器逆模型的期望输出 \hat{y}_{ij} 的均方误差。有：

$$f_i = \frac{1}{l} \sum_{j=1}^l (y_{ij} - \hat{y}_{ij})^2 \quad (8-9)$$

式中 f_i 表示第 i 个粒子的适应值，在本例中取值为 1~10； l 表示样本的个数，在本例中取值为 490； y_{ij} 是第 i 个粒子的第 j 个样本的模型预测值，即被测压力； \hat{y}_{ij} 是第 j 个样本的模型期望值，即对应的压力标定值。根据每个微粒值的大小即 (c, σ) 代入 LS-SVM 重建回归模型，根据检验样本的计算结果，由式(8-9)即可得到每个微粒对应的适应值。

根据式(8-3)比较各个粒子的适应值大小，即可得到当前粒子的最优位置 $p_i(t)$ ，即能够使得当前计算结果最优的 σ 的值。由式(8-4)即可得到当前群体的最优位置 $p_g(t)$ ，即最优参数。

- (5) 根据式(8-5)及式(8-2)调整粒子的速度及位置。

- (6) 未达到结束条件则转第(4)步。

优化过程部分程序如下：

```
%%样本输入
trainin=...; %训练样本输入
trainout=...; %训练样本输出
testin=...; %检验样本输入
testout=...; %检验样本输出
errormin=100; %设定参比误差
.....
%将粒子的位置 X 及速度 V 随机初始化
X=a+(b-a)*rand(N, D, 1);
%a, b 是优化参数的取值范围, N 是粒子个数, D 是粒子维数, 本例中取 1
V=m+(n-m)*rand(N, D, 1); %m, n 是速度的飞行速度限制范围
fitness=fitcal(X, trainin, trainout); %计算适应值
[C, I]=min(fitness(:, 1, j)); %查找适应值最小点
gbest(1, :, j)=X(I, :, j); %记录当前群体最优解
if B(1, 1, j)<=C
    gbest(1, :, j)=gbest(1, :, j); %记录群体历史最优解
else
```

```

gbest(1, :, j)=gbest(1, :, D);
end
for p=1:N
G(p, :, j)=gbest(1, :, j); %记录群体最优解
end
for i=1:N;
[C, I]=min(L(i, 1, :));
if L(i, 1, j)<=C
pbest(i, :, j)=X(i, :, j); %记录各个粒子历史最优解
else
pbest(i, :, j)=X(i, :, D);
end
end
V(:, :, j+1)=W(j) * V(:, :, j)+c1 * rand * (pbest(:, :, j)-X(:, :, j))+c2 * rand * (G(:, :, j)-X(:, :, j));
%速度调整
X(:, :, j+1)=X(:, :, j)+V(:, :, j+1); %微粒位置调整
.....
function fitval = fitcal(pm, Ptrain, Ttrain) %适应度函数
[x, y, z]=size(pm);
for j=1:x
delta=pm(j, 1, z);
type = 'f'; %回归
kernel = 'RBF_kernel';
model = initlssvm(Ptrain, Ttrain, type, 100, delta, kernel); %模型初始化
model = trainlssvm(model); %训练
Yd = simlssvm(model, testin); %回归
error=mse(Yd-testout); %记录每个粒子的适应值
fitval(j, 1, 1)=error;
end

```

上述优化过程流程图如图 8-6 所示。

上述优化过程的结束条件为模型预测结果均方误差为零或粒子迭代次数达到设定值。将优化结束之后得到的粒子群最优位置(c, σ)赋予 LS-SVM 利用测试样本重建的回归模型，即可得到测试样本的模型预测结果。

选取粒子群数目为 1，粒子群中粒子数目为 10 个，粒子群优化代数为 100，惯性权重因子 w 取初值为 0.9，终止值为 0.3 进行逐代调整，学习因子取 $c_1=c_2=2$ ，粒子群的优化参数 σ 的优化范围为 0~20，可得到如图 8-7 所示粒子群优化误差曲线。横轴为学习步数，纵轴为模型预测结果均方误差。由图 8-7 可以看出，粒子群优化算法只需 30 多步即可找到全局最优解，收敛速度非常快。优化 100 步所用时间约为 986 秒，得到的惩罚因子 c 为 100，核函数参数 σ 为 0.0523，预测结果均方误差为 2.13×10^{-5} 。模型预测结果如图 8-8 所示，部分测试结果如表 8-3 所示。

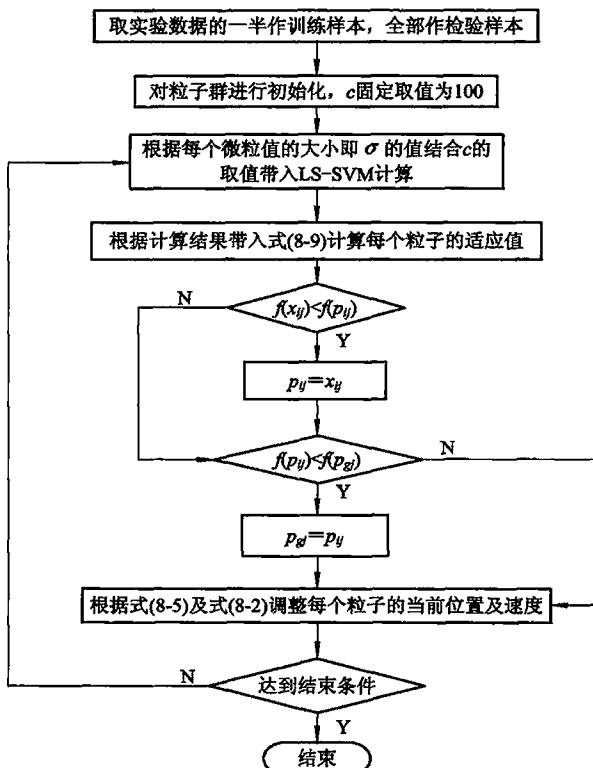


图 8-6 单参数优化过程流程图

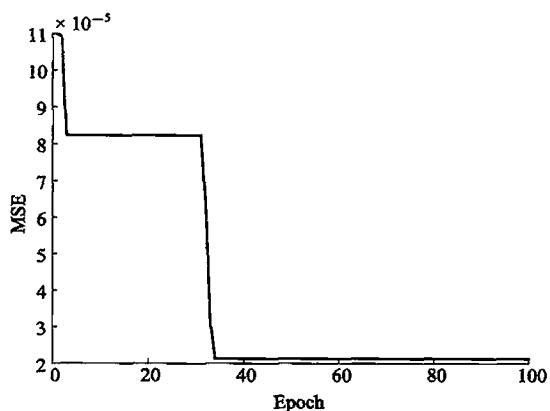


图 8-7 粒子群优化误差曲线

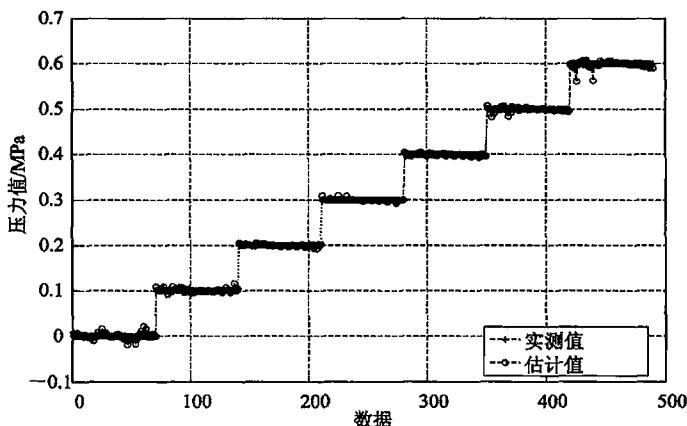


图 8-8 单参数校正模型预测结果

表 8-3 部分标定数据及预测结果

序号	温度 /°C	标定压力 /MPa	电流 /mA	电压 /V	预测结果 /MPa	序号	温度 /°C	标定压力 /MPa	电流 /mA	电压 /V	预测结果 /MPa
1	58	0	9	0.016	0.0048	21	39	0	9	0.334	0.0071
2	64	0	11	0.022	0.0008	22	25	0.1	6	0.834	0.1022
3	58	0	10	0.028	0.0158	23	43	0.1	8	0.918	0.0941
4	31	0	11	0.028	0.0014	24	31	0.1	8	0.94	0.0911
5	58	0	8	0.056	0.0218	25	64	0.2	8	0.52	0.1922
6	58	0	11	0.076	0.0038	26	58	0.2	5	1.068	0.1948
7	31	0	9	0.082	0.0019	27	55	0.3	5	1.492	0.2969
8	58	0	7	0.092	0.0053	28	53	0.3	5	1.498	0.2980
9	31	0	8	0.109	0.0029	29	53	0.4	5	1.912	0.3965
10	55	0	11	0.12	0.0007	30	35	0.4	10	3.674	0.3963
11	64	0	6	0.168	0.0002	31	55	0.5	6	2.727	0.4985
12	53	0	10	0.21	0.0042	32	35	0.5	9	4.079	0.4834
13	53	0	8	0.22	0.0186	33	64	0.6	5	2.719	0.5952
14	53	0	5	0.245	0.0048	34	31	0.6	5	2.758	0.6014
15	35	0	8	0.276	0.0104	35	49	0.6	7	3.727	0.6011
16	25	0	8	0.317	0.0073	36	39	0.6	8	4.309	0.5954
17	25	0	5	0.326	0.0077	37	25	0.6	10	5.337	0.5603
18	39	0	6	0.327	0.0048	38	31	0.6	11	5.74	0.6080
19	39	0	5	0.328	0.0082	39	39	0.6	11	5.812	0.6030
20	39	0	11	0.33	0.0070	40	25	0.6	11	5.841	0.5968

由表 8-3 的预测结果可以计算出经过粒子群算法优化模型校正之后, 压力传感器在 $\Delta T = 39^\circ\text{C}$, $\Delta I = 6 \text{ mA}$ 变化范围内, 模型计算零点值的最大偏差 $|\Delta P'_{0m}| = 0.0216 \text{ MPa}$; 满量程压力 $P_{fs} = 0.6 \text{ MPa}$, 其模型计算值的最大偏差量 $|\Delta P'_m| = 0.0477 \text{ MPa}$, 则模型校正后压力传感器的零位温度系数、温度灵敏度系数及电流灵敏度系数分别为:

$$\alpha_0 = \frac{|\Delta P'_{0m}|}{P_{fs} \times \Delta T} = \frac{0.0216}{0.6 \times 39} \approx 9.23 \times 10^{-4} (\text{}/\text{ }^\circ\text{C})$$

$$\alpha_t = \frac{|\Delta P'_m|}{P_{fs} \times \Delta T} = \frac{0.0477}{0.6 \times 39} \approx 2.04 \times 10^{-3} (\text{}/\text{ }^\circ\text{C})$$

$$\alpha_I = \frac{|\Delta P'_m|}{P_{fs} \times \Delta I} = \frac{0.0477}{0.6 \times 6} \approx 1.32 \times 10^{-2} (\text{/mA})$$

8.4.4 [示例 8-3]采用粒子群优化算法优化两个参数

在本示例中同样采用最小二乘支持向量机技术建立传感器的逆模型。为了提高模型的预测准确度及模型的抗干扰能力, 利用粒子群优化算法对最小二乘支持向量机中的核函数参数及惩罚因子进行优化选取。与[示例 8-2]不同的是此时优化的参数为两个, 即 c 及 σ 。优化过程如下:

(1) 由于每个粒子群只能优化一个参数, 因此设置粒子群维数为 2, 每维粒子群中粒子的数目选取 10 个, 迭代次数选为 100。其中粒子群 p_1 优化的参数为 c , 粒子群 p_2 优化的参数为 σ 。根据两个优化参数的优化范围对两个粒子群的初始位置及速度进行初始化。

(2) 设置目标函数为模型检验样本预测结果 y_i 与 LV-SVM 建立的压力传感器逆模型的期望输出 \hat{y}_i 的均方误差, 如式(8-9)所示。将每个微粒值的大小即(c, σ)代入 LS-SVM 重建回归模型, 根据检验样本的计算结果, 由式(8-9)即可得到每个微粒对应的适应值。

(3) 对粒子群 p_1 , 根据式(8-3)可以确定其中适应值最小的那个粒子的值, 将其记为 $p_{11}(t)$, 即能够使得当前计算结果最优的 c 的值; 同理可以得到 $p_{12}(t)$, 即能够使得当前计算结果最优的 σ 的值。此即为当前粒子的最优位置 $p_{ij}(t)$ 。根据式(8-4)可得到当前群体的最优位置 $p_{g1}(t)$ 、 $p_{g2}(t)$, 即群体最优位置 $p_{gi}(t)$ 。

(4) 根据式(8-5)及式(8-2)调整粒子的速度及位置。

(5) 未达到结束条件则转第(2)步。

优化过程部分程序同示例 8-2, 只是粒子维数及适应度函数有所变化:

```
function fitval = fitcal(pm, Ptrain, Ttrain)
[x, y, z]=size(pm);
for j=1:x
    c=pm(j, 1, z);
    delta=pm(j, 2, z);
    type = 'f';                                     % 回归
    kernel = 'RBF_kernel';
    model = initlssvm(Ptrain, Ttrain, type, c, delta, kernel); % 模型初始化
    model = trainlssvm(model);                      % 训练
```

```

Yd = simlssvm(model, testin); % 回归
error=mse(Yd-testout);
fitval(j, 1, 1)=error; % 记录每个粒子适应值
end

```

上述优化过程流程图如图 8-9 所示。

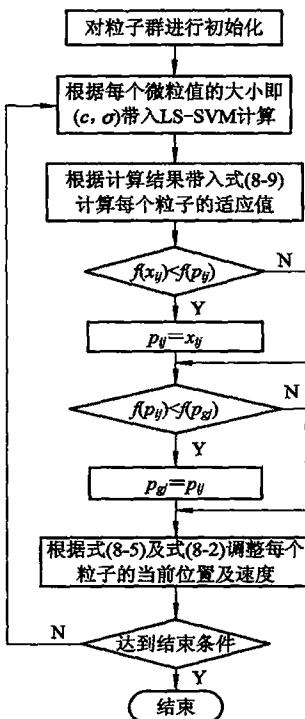


图 8-9 双参数优化过程流程图

上述优化过程的结束条件为 LS-SVM 建立的传感器回归逆模型预测结果均方误差为零或粒子迭代次数达到设定值。将优化结束之后得到的粒子群最优位置 (c, σ) 赋予 LS-SVM，利用测试样本重建回归模型，即可得到测试样本的模型预测结果。

选取粒子群数目为 2，粒子群中粒子数目为 10 个，粒子群优化代数为 100，惯性权重因子 w 取初值为 0.9，终止值为 0.3 进行逐代调整，学习因子取 $c_1=c_2=2$ ，两维粒子的优化范围分别为 0~100000 和 0~20。根据上述的模型优化过程计算，可得到如图 8-10 所示双参数粒子群优化误差曲线，横轴为学习代数，纵轴为模型预测结果均方误差。

由图 8-10 中可以看出，粒子群优化算法只需 30 多步即可找到全局最优解，收敛速度非常快。优化 100 步所用时间约为 1676 秒，得到的惩罚因子 c 为 23 359.39，核函数参数为 0.0364，预测结果均方误差为 1.25×10^{-6} 。模型预测结果如图 8-11 所示，预测结果数据见表 8-4。

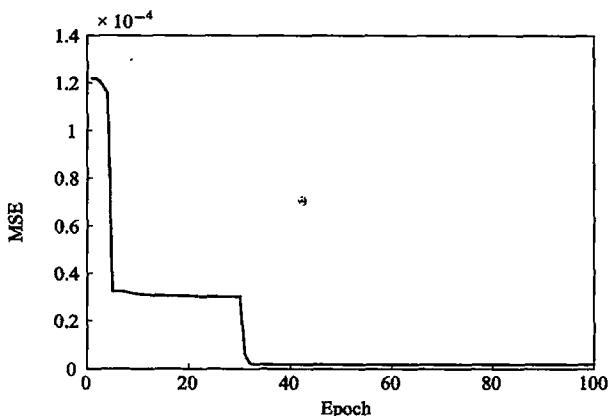


图 8-10 双参数粒子群优化误差曲线

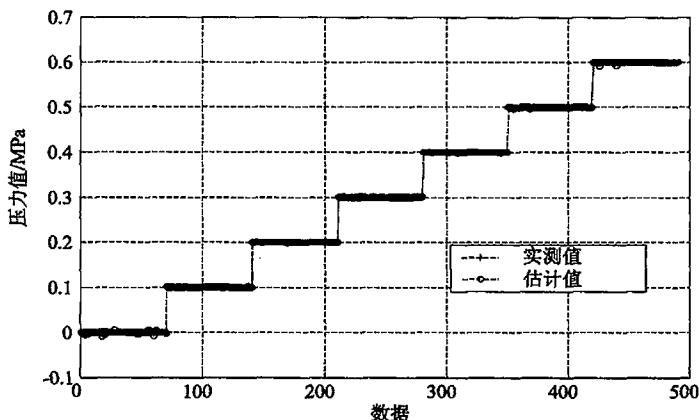


图 8-11 双参数校正模型预测结果

由表 8-4 的预测结果可以计算出经过粒子群算法优化模型校正之后压力传感器在 $\Delta T = 39^\circ\text{C}$, $\Delta I = 6 \text{ mA}$ 变化范围内, 零点模型计算值的最大偏差 $|\Delta P'_{0m}| = 0.0105 \text{ MPa}$; 满量程压力 $P_{\text{FS}} = 0.6 \text{ MPa}$, 其模型计算值的最大偏差量 $|\Delta P'_m| = 0.0091 \text{ MPa}$, 则模型校正后压力传感器的零位温度系数、温度灵敏度系数及电流灵敏度系数分别为

$$\alpha_0 = \frac{|\Delta P'_{0m}|}{P_{\text{FS}} \times \Delta T} = \frac{0.0105}{0.6 \times 39} \approx 4.49 \times 10^{-4} (\text{/}^\circ\text{C})$$

$$\alpha_s = \frac{|\Delta P'_m|}{P_{\text{FS}} \times \Delta T} = \frac{0.0091}{0.6 \times 39} \approx 3.89 \times 10^{-4} (\text{/}^\circ\text{C})$$

$$\alpha_i = \frac{|\Delta P'_{0m}|}{P_{\text{FS}} \times \Delta I} = \frac{0.0091}{0.6 \times 6} \approx 2.53 \times 10^{-3} (\text{/mA})$$

表 8-4 标定数据及预测结果

序号	温度 /℃	标定压力 /MPa	电流 /mA	电压 /V	预测结果 /MPa	序号	温度 /℃	标定压力 /MPa	电流 /mA	电压 /V	预测结果 /MPa
1	58	0	9	0.016	0.0005	21	39	0	9	0.334	0.0006
2	64	0	11	0.022	0.0032	22	25	0.1	6	0.834	0.1004
3	58	0	10	0.028	0.0049	23	43	0.1	8	0.918	0.1003
4	31	0	11	0.028	0.0002	24	31	0.1	8	0.94	0.1011
5	58	0	8	0.056	0.0087	25	64	0.2	8	0.52	0.2002
6	58	0	11	0.076	0.0002	26	58	0.2	5	1.068	0.1995
7	31	0	9	0.082	0.0001	27	55	0.3	5	1.492	0.3003
8	58	0	7	0.092	0.0000	28	53	0.3	5	1.498	0.3002
9	31	0	8	0.109	0.0003	29	53	0.4	5	1.912	0.3995
10	55	0	11	0.12	0.0052	30	35	0.4	10	3.674	0.3997
11	64	0	6	0.168	0.0000	31	55	0.5	6	2.727	0.5001
12	53	0	10	0.21	0.0031	32	35	0.5	9	4.079	0.4970
13	53	0	8	0.22	0.0035	33	64	0.6	5	2.719	0.5993
14	53	0	5	0.245	0.0001	34	31	0.6	5	2.758	0.6003
15	35	0	8	0.276	0.0105	35	49	0.6	7	3.727	0.5997
16	25	0	8	0.317	0.0056	36	39	0.6	8	4.309	0.5997
17	25	0	5	0.326	0.0005	37	25	0.6	10	5.337	0.5915
18	39	0	6	0.327	0.0000	38	31	0.6	11	5.74	0.6000
19	39	0	5	0.328	0.0006	39	39	0.6	11	5.812	0.6006
20	39	0	11	0.33	0.0062	40	25	0.6	11	5.841	0.5998

表 8-5 几种优化方法比较

评价参数	零位温度系数 α_0 (/℃)	灵敏度温度系数 α_s (/℃)	电流影响系数 α_i (/mA)
模型校正前	1.4×10^{-3}	1.32×10^{-2}	8.91×10^{-2}
遍历优化	1.85×10^{-3}	2.54×10^{-3}	1.65×10^{-2}
PSO 优化一个参数	9.23×10^{-4}	2.04×10^{-3}	1.32×10^{-2}
PSO 优化两个参数	4.49×10^{-4}	3.89×10^{-4}	2.53×10^{-3}

几种优化方法的比较见表 8-5。由表 8-5 可见，经过遍历优化之后模型的温度灵敏度系数及电流灵敏度系数都得到明显改善，但零位温度系数并没有得到改善，而且变差了，说明此时的模型在零点的计算结果偏差过大，此时的模型参数并不是最好的。利用 PSO 优化一个参数之后模型的零位温度系数、温度灵敏度系数及电流灵敏度系数均得到大幅度改善，说明此时的模型比遍历优化得到的模型好，性能也得到较好改善。经过 PSO 优

化两个参数之后模型的各评价参数均较之前方法有了更大幅度提高，均改善了近一个数量级，说明此时的模型在三种方法获取模型中是最佳的。PSO 在全局优化及收敛速度方面具有较大优势。

参 考 文 献

- [1] 曾建潮, 介婧, 崔志华. 微粒群算法[M]. 北京: 科学出版社, 2004
- [2] K W Chau. Particle swarm optimization training algorithm for ANNs in stage prediction of Shing Mun River[J]. Journal of Hydrology, 2006, 329(3–4): 363–367
- [3] Hui Pan, Ling Wang, Bo Liu. Particle swarm optimization for function optimization in noisy environment[J]. Applied Mathematics and Computation, 2006, 181(2): 908–919
- [4] Yan Jiang, Tiesong Hu, ChongChao Huang, et al. An improved particle swarm optimization algorithm[J]. Applied Mathematics and Computation, 2007, 193(1): 231–239
- [5] 刘瑞兰, 牟盛静, 苏宏业, 等. 基于支持向量机和粒子群算法的软测量建模[J]. 控制理论与应用, 2006(6). 895–906
- [6] 陈如清, 俞金寿. 基于粒子群最小二乘支持向量机的软测量建模[J]. 系统仿真学报, 2007(22): 5307–5310
- [7] 吴启迪, 汪镭. 智能微粒群算法研究及应用[M]. 南京: 江苏教育出版社, 2005: 15–16
- [8] 冯静, 苏宁. 群智能理论及应用研究[J]. 计算机工程与应用, 2006(17): 31–34
- [9] 高尚, 杨静宇. 群智能算法及其应用[M]. 北京: 中国水利水电出版社, 2006: 1–2
- [10] Yujun Li, Xiaojun Tang, Junhua Liu. Application of Particle Swarm Optimization Algorithm in Soft Sensor Modeling [C]. 2009 9th International Conference on Electronic Measurment and Instruments(ICEMI), 2009. 8

第9章 主成分分析及其在智能传感器系统中的应用

本章内容

- 主成分分析法
- PCA 算法在消除传感器漂移中的应用
- [示例 9-1] 建立传感器阵列，获取关联数据与漂移数据
- [示例 9-2] 传感器发生漂移的识别

主成分分析(Principal Component Analysis, PCA)可以达到的主要目的是使数据降维，将原数据进行转换，获得少数几个新变量，这些新变量为原变量的线性组合，同时，这些变量要尽可能多地表征原变量的数据特征而不丢失信息。PCA 已在实现数据的简化和压缩、建模、奇异值检测、特征变量的提取与选择、分类和预报等广大领域获得了成功应用。

本章首先介绍了主成分分析法的基础知识，然后以基于冗余法消除传感器漂移、改善传感器稳定性为例，举例说明主成分分析法在智能传感器系统中的应用。

9.1 主成分分析法

主成分分析最初是由 Pearson 在 1901 年提出来的，也称主元分析。当时的 PCA 主要用于对空间的一些点进行直线和平面的最佳拟合。1933 年，Hotelling 对 PCA 方法进行了改进，使其成为目前被广泛使用的方法。在 70 多年的发展中，不同领域的学者从不同的角度对 PCA 方法进行了研究，并冠以不同的名称：在数值分析领域，PCA 被称为奇异值分解(Singular Value Decomposition)和 Karhunen - Loeve 扩展；在物理学领域，PCA 被称为本征矢量投影(Eigen Vector Projection)；在化学领域，PCA 则被称为主要因素分析(Principal Factor Analysis)等。主成分分析可以实现数据的简化和压缩、建模、奇异值检测、特征变量的提取与选择、分类和预报等目的。

9.1.1 二维空间中的 PCA

PCA 的操作涉及到多维空间中的投影概念。不失一般性，为说明简单起见，这里以二维空间中的主成分分析为例来说明 PCA 的算法思想。假定在二维空间中有一组测试点 (y_{1i}, y_{2i}) ($i=1, 2, \dots, m$)，如图 9-1 所示。

如果将二维数据降至一维数据，也就是将二维空间的点投影到一维空间的一条线上，在没有任何约束条件的情况下，其投影的方向有无穷多个，这是没有意义的。PCA 操作采用如下约束条件：在一维空间中的这条直线必须包含原数据的最大方差，即沿着这条直线

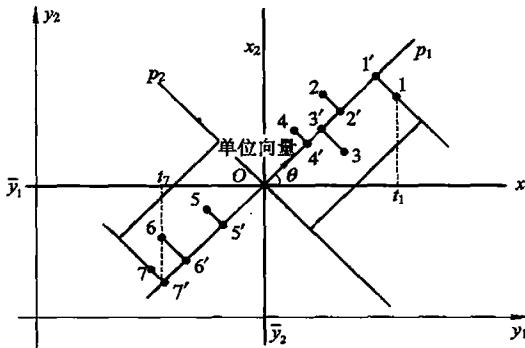


图 9-1 二维空间中的 PCA 示意图

使原数据的方差达到最大。图 9-1 中点 i ($i=1, 2, \dots, 7$) 向直线 p_1 投影为点 i' ($i'=1, 2, \dots, 7$)，这些点的重心为 O ，其分布可用它们到中心点 O 的距离的平方和表示。原数据点的距离分布为

$$S_2 = |O_1|^2 + |O_2|^2 + \dots + |O_7|^2 \quad (9-1)$$

如果用 p_1 上的投影表示，则

$$|O_i|^2 = |O_{i'}|^2 + |ii'|^2 \quad (9-2)$$

所以

$$S_2 = |O_{1'}|^2 + |O_{2'}|^2 + \dots + |O_{7'}|^2 + |11'|^2 + |22'|^2 + \dots + |77'|^2 \quad (9-3)$$

PCA 选择投影直线 p_1 使式(9-3)中 S_2 的值最大。这条直线也正好是这些原数据点的最好拟合线，它使得所有的原始数据点到 p_1 直线上对应投影点垂直距离的平方和最小。

称 p_1 为主成分空间，图中箭头表示该空间中的单位向量，即载荷向量。如点 1 和点 7 在 p_1 空间中的投影点分别为 $1'$ 和 $7'$ ，它们在 p_1 空间中的坐标分别为 t_1 和 t_7 ，即在 p_1 空间中用载荷向量对投影点距重心点距离度量的得分。

上述例子中，使用一维新变量 p_1 表征二维的原数据 (y_{1i}, y_{2i}) ($i=1, 2, \dots, m$) 的结构特征，新变量包含了原数据中绝大部分的信息特征，称为第一主成分。还有部分剩余的信息没有被包含进来，可以使用与选取第一主成分相同的方法，再选出第二主成分来描述这剩余信息部分。第二主成分应在与第一主成分不相关的其余变量中能包含最大的方差。对于多维空间，依此类推，可以选出第三、第四等主成分。

其实，PCA 对原变量的变换得到的新变量就是原变量的线性组合，如图 9-1 所示。原坐标系的原点经过转换后，放到重心 O 处。根据几何规则，新变量可以由原数据以线性组合的形式表示：

$$\begin{bmatrix} p_1 \\ p_2 \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \quad (9-4)$$

$$\begin{aligned} p_1 &= ax_1 + bx_2 = x_1 \cos\theta + x_2 \sin\theta \\ p_2 &= cx_1 + dx_2 = x_1 (-\sin\theta) + x_2 \cos\theta \\ x_{1i} &= y_{1i} - \bar{y}_1 \\ x_{2i} &= y_{2i} - \bar{y}_2 \end{aligned} \quad (9-5)$$

其中 $a^2 + b^2 = 1$, $c^2 + d^2 = 1$; 主成分 1 为 $\begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix}$, 主成分 2 为 $\begin{bmatrix} c \\ d \end{bmatrix} = \begin{bmatrix} -\sin\theta \\ \cos\theta \end{bmatrix}$ 。

将二维空间的 PCA 算法扩展到多维, 就是我们常用的 PCA 算法。

9.1.2 PCA 算法

假设 $x = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nm} \end{bmatrix}$ 是一个 $n \times m$ 的数据矩阵, 其中每一列对应一个变量, 每一行对应一个样本。例如, 表 8-1 表示的压阻式压力传感器标定数据中, $m=3$, 表示压力、温度和电流三个传感器, $n=28$, 表示共有 28 组样本。 x 可以分解为 m 个向量的外积之和, 即

$$x = t_1 p_1^T + t_2 p_2^T + \cdots + t_m p_m^T \quad (9-6)$$

其中 $t_i \in \mathbb{R}^n$ 为得分向量, $p_i \in \mathbb{R}^m$ 为载荷向量。 x 的得分向量也叫 x 的主成分。式(9-6)可以写成矩阵形式:

$$x = TP^T \quad (9-7)$$

其中 $T = [t_1, t_2, \dots, t_n]$ 为得分矩阵, $P = [p_1, p_2, \dots, p_m]$ 为载荷矩阵。

各个得分向量之间是正交的, 即对任何 i 和 j , 当 $i \neq j$ 时, 满足 $t_i^T t_j = 0$ 。各个载荷向量之间也是互相正交的, 同时每个载荷向量的长度都为 1, 即

$$\begin{aligned} p_i^T p_j &= 0 \quad i \neq j \\ p_i^T p_i &= 1 \quad i = j \end{aligned} \quad (9-8)$$

将式(9-6)两侧同时右乘 p_1 , 可以得到下式

$$xp_1 = t_1 p_1^T p_1 + t_2 p_2^T p_1 + \cdots + t_m p_m^T p_1 \quad (9-9)$$

将式(9-8)代入式(9-9), 可得:

$$t_1 = xp_1 \quad (9-10)$$

式(9-10)说明, 每一个得分向量实际上是数据矩阵 x 在这个得分向量 t_1 相对应的载荷向量方向 p_1 上的投影。向量 t_1 的长度反映了数据矩阵 x 在 p_1 方向上的覆盖程度。它的长度越大, x 在 p_1 方向上的覆盖程度或变化范围越大。如果将得分向量按其长度做以下排列:

$$\|t_1\| > \|t_2\| > \cdots > \|t_m\|$$

那么载荷向量 p_1 将代表数据 x 变化最大的方向。 p_2 与 p_1 垂直并代表数据 x 变化的第二大方向, p_m 将代表数据 x 变化最小的方向。

当矩阵 x 中的变量间存在一定程度的线性相关时, 数据 x 的变化将主要体现在最前面的几个载荷向量方向上, 数据矩阵 x 在最后面的几个载荷向量上的投影将会很小, 它们主要是由于测量噪声引起的。这样就可以将矩阵 x 进行主元分解后写成下式

$$x = t_1 p_1^T + t_2 p_2^T + \cdots + t_k p_k^T + E \quad (9-11)$$

式中 E 为误差矩阵, 代表 x 在 p_{k+1} 到 p_m 等载荷向量方向上的变化。很多实际应用中, k 往往要比 m 小得多。由于误差矩阵 E 主要是由于测量噪声引起的, 所以将 E 忽略掉往往回起

到清除测量噪声的效果，不会引起数据有用信息的明显损失。

9.2 PCA 算法在消除传感器漂移中的应用

传感器特性漂移，表现为传感器性能不稳定，这种现象普遍存在。传感器性能不稳定已成为实时在线监测系统的瓶颈。例如在电力系统中，如果一个实时在线监测系统在一年中出现1~2次误报，那么这个监测系统将是不可信任而必须撤出的。因此，实时在线监测系统对其中的传感器的稳定性提出了更严格的要求，务必杜绝因传感器本身特性漂移而产生误报的现象。

冗余法是传感器故障诊断及漂移消除的一种有效方法。其基本思路是：不去探究引起传感器漂移的是哪种干扰量，以及干扰量对传感器漂移产生怎样的影响。监测一个参量本来只需要一个传感器，而冗余法则采用多个目标参量相同的传感器（至少三个）来监测同一个被测量，建立监测同一被测量的多传感器系统，其系统框图如图9-2所示（同第4章图4-11）。需要对多路同种传感器的输出信号进一步进行数据融合处理，以识别并克服传感器的漂移。本节将介绍基于主成分分析融合算法的多传感器数据融合处理法在消除传感器漂移中的应用。

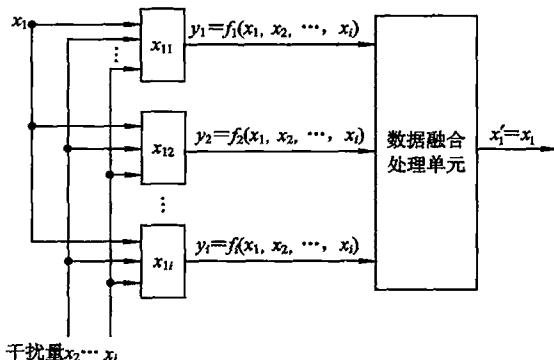


图9-2 基于冗余法的三传感器(监测一个参量)的智能传感器系统

9.2.1 PCA算法实现传感器故障检测的思想

首先逐个建立传感器正模型来表达传感器正常工作时输入与输出的特征关系，它们对被测量的变化具有相关关联的响应特性。而在工作一定时间之后，当某个传感器输入输出特性发生漂移时，其输出测量值就会出现与该传感器模型正常输出不相符的现象，且与其他传感器的输出的关联性降低。利用已建立的传感器正模型对各传感器实际运行中的输出测量值进行分析，通过对分析结果进行分析和评价从而判断各传感器是否正常工作。平方预报误差(Squared Prediction Error, SPE)为评价传感器是否正常工作的统计量。

1. 故障检测的判断依据

- (1) 计算平方预报误差。测得一组数据向量 $x = [x_1, x_2, \dots, x_m]$ ，其中 x_i 为第 i 个

传感器的输出数据，则利用主元分析技术可获得向量 x 的近似值或估计值为

$$\hat{x} = tP_h^T = xP_hP_h^T = xC \quad (9-12)$$

式中： $C = P_hP_h^T$ ， $P_h = [p_1, p_2, \dots, p_h]$ ， h 为选定的主元个数； t 为 PCA 主元向量，且：

$$t = xP_h$$

真实值 x 和估计值 \hat{x} 之差为

$$\Delta x = \hat{x} - x = x(I - C)$$

则平方预报误差 SPE 为

$$SPE(x) = \|x\|^2 \quad (9-13)$$

SPE 统计量代表的是数据中没有被 PCA 模型所解释的变化，在正常情况下，SPE 的值比较小。当某个传感器发生故障时，该传感器输出与阵列中其它传感器输出之间的关联关系将发生改变。此时，由正常运行数据建立起来的 PCA 模型将产生很大的 SPE 值。由此可以指示出某传感器发生故障。利用各传感器对 SPE 的贡献量，可基本判定是哪个传感器发生了故障，一般认为贡献量最大的传感器发生了故障。

(2) 确定 SPE 控制限。当传感器发生漂移时，SPE 值会随着漂移值的增大而增大，根据这一特性，可用交叉假设法来确定 SPE 的控制限。设置控制限就是设置一个阈值，当漂移值达到设定的阈值时，系统发出故障报警或从传感器阵列中剔除故障传感器，以保证监测系统整体工作正常。

传感器阵列在正常情况下的输出数据矩阵 $x \in \mathbb{R}^{n \times m}$ ，其中， n 为数据组数(或样本数)， m 为传感器个数。 x 的表达式如下：

$$x = [x_1, x_2, \dots, x_m]$$

式中 x_i 为第 i 个传感器的输出数据。

首先利用 x 建立主元分析模型，得到选取主元所对应的载荷矩阵 P_h ，然后依次假设传感器 $1, 2, \dots, m$ 发生漂移，漂移量为 $a\%$ ，则可得到 m 个数据矩阵 $x_i (i=1, 2, \dots, m)$ ，表达式如下：

$$x_i = [x_1, x_2, \dots, x_{i-1}, x_i + a\% * L, x_{i+1}, \dots, x_m] \quad (9-14)$$

式中： L 为传感器的量程。

接着计算数据矩阵 x_i 的 SPE 值，则可得到 $n \times m$ 个 SPE 值，对这些 SPE 值求平均，得到的平均值可认为是传感器漂移 $a\%$ 时的 SPE 值的控制限。

2. 传感器故障检测方法

传感器故障检测主要分两部分：

(1) 建立 PCA 传感器模型，以反映传感器的正常运行状况。

① 在传感器的量程范围内，收集传感器在正常情况下的输出数据：

$$x = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nm} \end{bmatrix}$$

其中： n 为样本数， m 为传感器个数。

② 对 x 进行如下归一化处理, 目的是消除由于不同量纲所造成的虚假变异影响:

$$\mathbf{x}_s = [\mathbf{x} - (1 \ 1 \cdots 1)^T \mathbf{M}] \operatorname{diag} \frac{\frac{1}{s_1}, \frac{1}{s_2}, \dots, \frac{1}{s_m}}{\sqrt{n-1}} \quad (9-15)$$

式中: $\mathbf{M} = [m_1, m_2, \dots, m_m]$ 为变量 x 的均值; $\operatorname{diag}(\cdot)$ 为对角矩阵; 对角元素为变量的标准差的倒数, 即 $\frac{1}{s_1}, \frac{1}{s_2}, \dots, \frac{1}{s_m}$ 。

③ 对 \mathbf{x}_s 进行奇异分解:

$$\mathbf{x}_s = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T + \sigma_2 \mathbf{u}_2 \mathbf{v}_2^T + \cdots + \sigma_m \mathbf{u}_m \mathbf{v}_m^T \quad (9-16)$$

式中: $\sigma_i \mathbf{u}_i$ 记为第 i 个主元的分向量 t_i , \mathbf{v}_i 为 x 的第 i 个载荷向量。

④ 计算 \mathbf{x}_s 的特征值:

$$\lambda_j = \sigma_j^2, \quad (j = 1, 2, \dots, m) \quad (9-17)$$

⑤ 计算解释度:

$$S = \frac{\sum_{j=1}^h \lambda_j}{\sum_{j=1}^m \lambda_j} \quad (9-18)$$

⑥ 根据解释度大小确定主元个数。

⑦ 确定 SPE 的控制限。

(2) 利用 PCA 模型进行传感器故障检测。

① 采集传感器阵列的当前输出数据 $\mathbf{x} = [x_1, x_2, \dots, x_m]$, 其中 x_i 为第 i 个传感器的输出数据。利用已建立的 PCA 模型计算 \mathbf{x} 的近似值(或称为估计值):

$$\hat{\mathbf{x}} = \mathbf{t} \mathbf{P}_h^T = \mathbf{x} \mathbf{P}_h \mathbf{P}_h^T = \mathbf{x} \mathbf{C} \quad (9-19)$$

式中: $\mathbf{C} = \mathbf{P}_h \mathbf{P}_h^T$, $\mathbf{P}_h = [p_1, p_2, \dots, p_h]$, h 为选定的主元个数。

② 计算实际采样值与估计值之差:

$$\Delta \mathbf{x} = \hat{\mathbf{x}} - \mathbf{x} = \mathbf{x} (\mathbf{C} - \mathbf{I}) = [\Delta x_1, \Delta x_2, \dots, \Delta x_m] \quad (9-20)$$

则 SPE 值为

$$\operatorname{SPE}(\mathbf{x}) = \|\Delta \mathbf{x}\|^2$$

③ 比较当前 SPE 值与 SPE 值控制限, 若 SPE 值大于 SPE 值控制限, 则认为传感器阵列中有传感器发生故障。

④ 当前 SPE 值超出了 SPE 值控制限, 计算各传感器对 SPE 的贡献量(即比较式(9-20)中的 Δx_i 的绝对值的大小), 并认为其中贡献量最大的传感器发生了故障。

9.2.2 [示例 9-1]建立传感器阵列、获取关联数据与漂移数据

传感器发生漂移的过程是一个非常缓慢的过程, 可能在一个传感器阵列中某传感器发生漂移需要几个月甚至更长的时间。采用本示例的方法可以更方便快捷、不受时间限制地获得传感器漂移数据, 以便学习改善传感器稳定性的主成分分析法。

任务: 建立一个模拟传感器阵列, 由该传感器阵列可获取关联数据与漂移数据; 测量模拟传感器阵列中每个传感器的正模型, 即输入输出特性; 获取相对被测参量的关联数

据；令某个传感器发生漂移，并采集其漂移数据。

1. 模拟传感器阵列系统设计

模拟传感器阵列原理图如图 9-3 所示。采用一个分压器代表一个传感器，设置不同的分压比代表传感器之间的差异，电源电压 U_{in} 模拟被测参量；改变电源电压 U_{in} ，代表被测参量改变，且各传感器输入相同的被测量；各个分压器的输出电压 U_{out} 代表传感器的输出，不同的分压比由式(9-21)中的电阻 R_1 与电阻 R_2 阻值确定

$$U_{out} = \frac{R_2 + R_y}{R_1 + R_2 + R_x + R_y} U_{in} \quad (9-21)$$

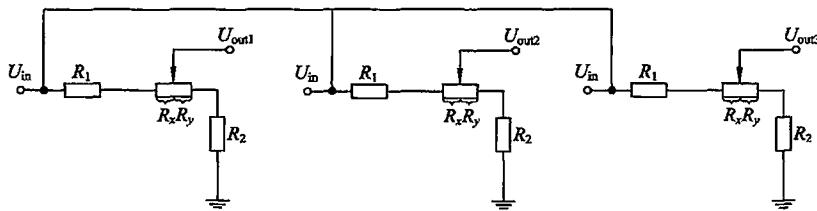


图 9-3 模拟传感器阵列原理图

电位器的满量程为 $R_x + R_y$ ，由式(9-21)可知，通过调节电位器，即改变 R_x 和 R_y 两者之间的电阻比，模拟传感器的漂移。这样，传感器的漂移可以不受时间的限制而人为地控制。

本节设计的模拟传感器阵列含有三个传感器。各传感器的输入电压 U_{in} 范围为 1~10 V，输出电压范围为 0~10 V。

2. 标定实验与关联数据的获取

模拟传感器系统的标定实验为：在多个不同电压 U_{in} 测量各个分压器（模拟传感器阵列）的输出电压 U_{out} ，即得相对同一输入电压 U_{in} （代表被测参量）的关联数据，也即模拟传感器阵列的输入输出特性。

图 9-4 所示特性与如表 9-1 所示标定数据为三个实际电压传感器的输入输出特性。

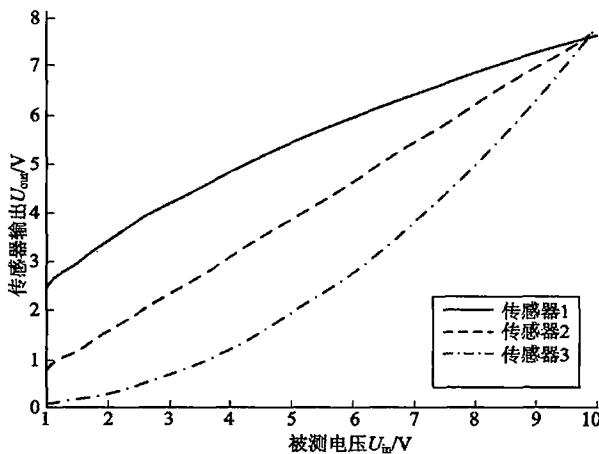


图 9-4 实际电压传感器阵列各传感器输入输出特性曲线

表 9-1 实际电压传感器阵列标定数据

样本序号	输入量 U_{in}/V	各传感器的输出数据/V			样本序号	输入量 U_{in}/V	各传感器的输出数据/V		
		传感器 1	传感器 2	传感器 3			传感器 1	传感器 2	传感器 3
1	1.04	2.464	0.806	0.084	18	5.80	5.821	4.487	2.616
2	1.32	2.777	1.023	0.136	19	6.08	5.960	4.704	2.875
3	1.60	3.059	1.240	0.200	20	6.36	6.096	4.920	3.146
4	1.88	3.316	1.457	0.275	21	6.64	6.229	5.138	3.429
5	2.16	3.555	1.674	0.363	22	6.92	6.356	5.349	3.716
6	2.44	3.779	1.891	0.464	23	7.2	6.484	5.565	4.023
7	2.72	3.983	2.102	0.574	24	7.48	6.610	5.782	4.344
8	3.00	4.184	2.319	0.699	25	7.76	6.733	5.999	4.674
9	3.28	4.374	2.536	0.835	26	8.04	6.854	6.216	5.018
10	3.56	4.558	2.752	0.984	27	8.32	6.973	6.432	5.374
11	3.84	4.734	2.969	1.115	28	8.6	7.089	6.648	5.742
12	4.12	4.904	3.185	1.319	29	8.88	7.205	6.866	6.123
13	4.40	5.068	3.403	1.504	30	9.16	7.318	7.083	6.516
14	4.68	5.228	3.619	1.702	31	9.44	7.429	7.299	6.920
15	4.96	5.382	3.837	1.912	32	9.72	7.539	7.516	7.338
16	5.24	5.532	4.054	2.135	33	10.00	7.648	7.733	7.767
17	5.52	5.679	4.270	2.369					

对模拟传感器标定获得的 33 组数据进行多项式插值，共获得 128 组标定数据。

3. 标定数据矩阵描述

设传感器阵列的标定(输出)数据矩阵用 x 表示， $x \in \mathbb{R}^{128 \times 3}$ ，其中，128 为数据组数(或称样本数)，3 为传感器个数，即

$$x = \begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ \vdots & \vdots & \vdots \\ x_{n1} & x_{n2} & x_{n3} \end{bmatrix}$$

其中： $n=128$ ； x_{ij} 表示第 i 组数据、第 j 个传感器的标定值。

4. 漂移数据的获得

1) 漂移实验 1

用模拟传感器阵列对 5 V 电压进行测量。在第 67 次采样时刻，将被测电压 U_{in} 调整到 8 V，用以模拟被测量改变；从第 172 次采样时刻起，调节模拟传感器 1 中的电位器，使传感器 1 发生漂移；从第 290 次采样时刻起，调节模拟传感器 3 中的电位器，使传感器 3 发生漂移，各传感器的输出结果如图 9-5 所示。

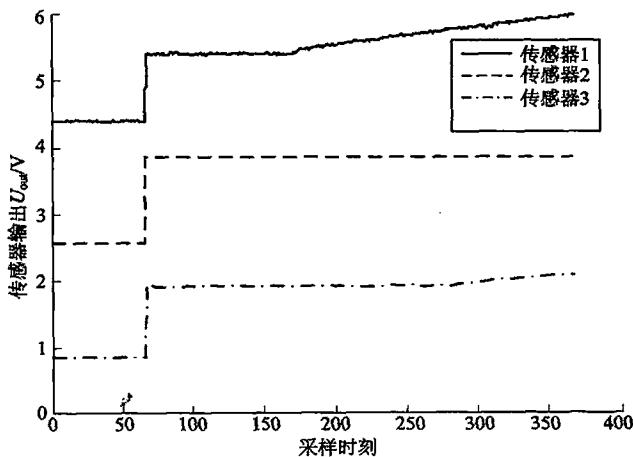


图 9-5 各传感器输出与获得的漂移数据

2) 漂移实验 2: 模拟传感器阵列输出电压含有高斯白噪声

被测电压 U_{in} 为 5.5 V, 高斯白噪声的均值为 0, 标准差为 0.05。从第 140 次采样时刻起, 调节模拟传感器 2 中的电位器, 使传感器 2 发生漂移; 从第 219 次采样时刻起, 改变调节传感器 2 中的电位器, 使传感器 2 的漂移规律发生改变, 各模拟传感器的输出结果如图 9-6 所示。

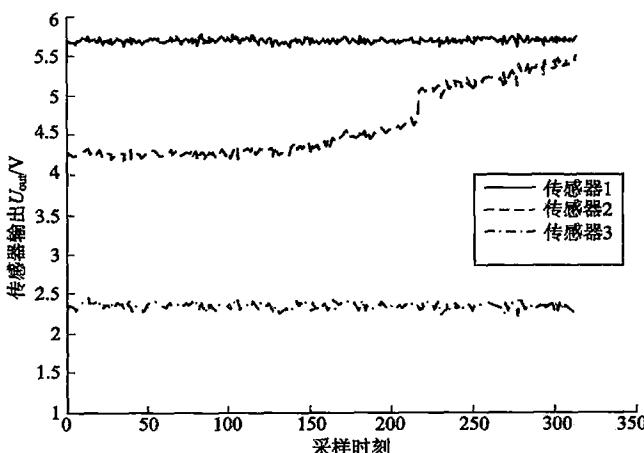


图 9-6 含有高斯白噪声的各传感器输出结果与漂移数据

9.2.3 [示例 9-2] 传感器发生漂移的识别

1. 建立传感器漂移 PCA 模型

仿照表 9-1 的格式将模拟传感器标定的 128 个数据构成的矩阵 x 进行主元分析, 获取如下参数:

$$\text{均值 } M = [5.5159 \quad 4.2955 \quad 2.9154]$$

标准差 $s = [5.6989 \quad 4.7477 \quad 3.7094]$

利用式(9-15)对 x 进行归一化处理得到数据矩阵 x_s , 对 x_s 进行奇异分解可得:
 x_s 的特征值向量

$$\lambda = [8547 \quad 230.81 \quad 2.973]$$

x_s 的载荷矩阵

$$P = \begin{bmatrix} -0.68987 & 0.60947 & -0.39068 \\ -0.57919 & -0.14092 & 0.80292 \\ -0.4343 & -0.78019 & -0.45021 \end{bmatrix}$$

三个主元的解释度为 97.338%, 2.629%, 0.034%, 因此, 确定主元个数为 1。第一主元所对应的载荷矩阵为

$$P_h = [-0.68987 \quad -0.57919 \quad -0.4343]^T$$

根据式(9-14), 传感器漂移 1% 时式(9-14)表示的 x_s 的 SPE 控制限为 5.3×10^{-5} 。

2. 漂移传感器的检测与辨识

根据式(9-19)计算漂移数据的估计值 \hat{x} 。

根据式(9-20)计算实际采样值与估计值之差。

求取 SEP 并与 SPE 的控制限 5.3×10^{-5} 相比较, 大于控制限的传感器为漂移传感器, 在实际应用中不考虑其输出数值, 取消对漂移传感器的数据检测。

3. 检测结果

漂移实验 1 的检测结果如图 9-7 所示, 在第 187 次采样时刻, SPE 值超出控制限; 在第 317 次采样时刻, SPE 值超出控制限; 而在第 70 次采样时刻, 由于传感器输出的变化是由正常输入响应引起的, 故未发生报警。图 9-8 给出了上述两个采样时刻各传感器对 SPE 的贡献量。由图可知, 在第 187 次采样时刻, 传感器 1 对 SPE 的贡献量最大, 即传感器 1 发生了漂移; 在第 317 次采样时刻, 传感器 3 对 SPE 的贡献量最大, 即传感器 3 发生了漂移。该检测结果与实际漂移数据相符。

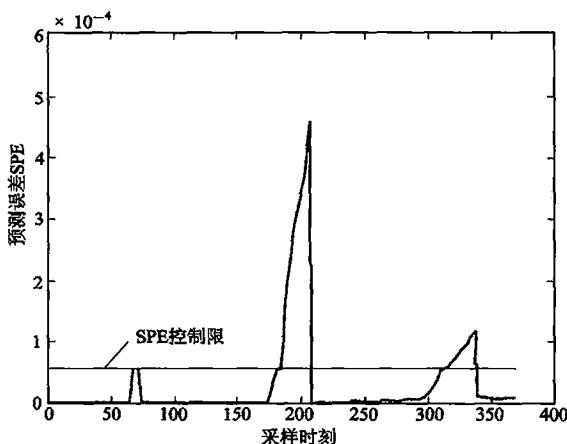


图 9-7 漂移实验 1 传感器漂移报警图(1)

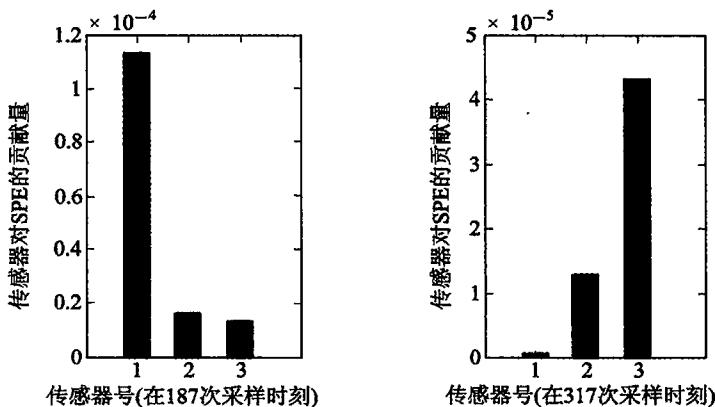


图 9-8 漂移实验 1 传感器漂移报警图(2)

4. 程序清单

```

clc                                     %清除内存中的所有变量和函数
clear

%主元分析程序
%载入传感器阵列的标定数据。其中，sample.txt 为正常工作的传感器阵列的标定数据，数据
以矩阵形式存储在 sample.txt 文件中
data=load('sample.txt');
%对 data 进行奇异分解，u 为归一化矩阵；s 为与 data 维数相同且降序排列的非负对角矩阵；
load 为载荷矩阵
[u,s,load]=svd(data);
%描述矩阵 s 的行数和列数，其中，row 为 s 的行数；column 为 s 的列数
[row,column]=size(s);

for(i=1:column)
    lmda(i)=s(i,i)^2;                      %计算特征向量
end

%定义主元矩阵
score=[];
ratio_temp=0;
for i=1:column
    score_temp=u(:,i)*s(i,i);
    score=[score score_temp];                %计算主元矩阵
    ratio_temp=lmda(i)+ratio_temp;
end

%定义解析度

```

```
present_ratio=[];
for jj=1: column
    ratio=lmda(jj)/ratio_temp;
    present_ratio=[present_ratio ratio]; %计算解析度
end

%传感器漂移检测
%漂移实验 1
%载入漂移实验 1 中各传感器输出数据
drift1=load('drift1.txt');
[row1 column1]=size(drift1);

%利用 PCA 模型计算传感器阵列的估计值
predict_x1=drift1 * load * load';

%计算真实值与估计值之差
error1=predict_x1-drift1;

%计算 SPE
for i=1:row1
    for j=1: column1
        drift1_spe(i,j)=(error1(i,j))^2;
    end
end

%计算各传感器对 SPE 的贡献量
spe_contribution=error1.^2;

%比较各 SPE 与 SPE 限，判断漂移传感器
for i=1: column1
    if drift1_spe(i)>5.3e-5
        disp(i)
        disp('该传感器出现漂移，请勿采信其输出数据');
    end

%漂移实验 2
%载入漂移实验 2 中各传感器输出数据
drift2=load('drift2.txt');
[row2 column2]=size(drift2);

%利用 PCA 模型计算传感器阵列的估计值
predict_x2=drift2 * load * load';
```

```
%计算真实值与估计值之差
error2=predict_x2-drift2;

%计算 SPE
for i=1: column2
    drift2_spe(i)=norm(error2(:,i))^2;
end

%比较各 SPE 与 SPE 限, 判断漂移传感器
for i=1: column2
    if drift2_spe(i)>5.3e-5
        disp(i)
        disp('该传感器出现漂移, 请勿采信其输出数据');
    end

```

参 考 文 献

- [1] 王颖, 张勇, 刘君华. 基于虚拟仪器的 PCA 模型构建仪设计[C]. 第二届全国虚拟仪器学术交流大会, 2007
- [2] 郝惠敏, 汤晓君, 白鹏, 等. 基于核主成分分析和支持向量回归的红外光谱多组分混合气体定量分析[J]. 光谱学与光谱分析, 2008, 28(6): 1286 - 1289
- [3] 谭应朝, 刘君华. 抑制传感器漂移特性智能化模块的模拟实验研究与实现[D]. 西安交通大学硕士学位论文, 2006
- [4] S Capone, M Epifani, F Quaranta, et al. Monitoring of rancidity of milk by means of an electronic nose and a dynamic PCA analysis[J]. Sensors and Actuators B, 2002, 78: 174 - 179
- [5] R Shao, F Jia, E B Martin, et al. Wavelet and non-linear principal components analysis for process monitoring[J]. Control Engineering Practice, 1999, 7: 865 - 879
- [6] Rongfu Luo, Manish Misra, David M Himmelblau. Sensor Fault Detection via multi-scale analysis and dynamic PCA[J]. Ind. Eng. Chem. Res, 1999, 28: 1489 - 1495
- [7] F Jia, E B Martin, A J Morris. Non-linear Principal Components Analysis for Process Fault Detection [J]. Computers & Chemical Engineering, 1998, 22: 851 - 854
- [8] Bhavik R Bakshi. Multiscale PCA with application to multivariate statistical process monitoring. American Institute of Chemical Engineers[J]. AIChE Journal. 1998, 44(7): 1596 - 1606
- [9] Mark A Kramer. Non-linear principal component analysis using auto-associative neural networks[J]. AIChE Journal, 1991, 37(2): 233 - 243

第 10 章 小波分析及其在智能传感器系统中的应用

本章内容

- 小波分析基础
- Matlab 工具箱中小波分析函数
- [示例 10-1] 小波数字滤波器的实现

自 1986 年以来，小波分析得到了迅速的发展，其应用也逐渐变得越来越广泛，在某些领域中取得了一些超乎寻常的成果，特别是在噪声消除、特征信号的提取、图像处理等方面。因此，小波分析越来越得到广大科技工作者的青睐。

本章首先从短时傅里叶变换入手，导出小波变换的变换公式，建立小波变换与傅里叶变换的关系，重点介绍正交小波分解与重构的方法、小波包及正交小波包以及正交小波包的分解与重构的方法。在了解小波分析基础知识的基础上，介绍 Matlab 工具箱中小波分析函数的功能及使用方法。在本章的最后，给出了用于智能传感器系统的小波数字滤波器示例。

10.1 小波分析基础

1807 年，Fourier 曾断言，任何一个周期为 2π 的函数 $f(x)$ 都是 Fourier 级数的和。绝大多数周期为 2π 的连续函数满足 Fourier 的断言，但在 1873 年 P. DuBois 和 Reymond 构造了一个实变量 x 的 2π 周期连续函数，它的 Fourier 级数在给定的点是发散的，这在当时的数学界引起了轩然大波，也引起了众多数学家的兴趣。在那时，数学界有三个未开发的领域：

- (1) 修改函数的一般要求，并求得适合于 Fourier 级数的函数；
- (2) 修改 Fourier 级数收敛的定义；
- (3) 求得另外的正交族，使得对于在三角族情形发生的 P. DuBois-Reymond 发散现象在此时不至于发散。

这三个方向后来导致了重要的结果。其中，第三个方向导致了小波分析的产生。

本节首先从 2π 变换入手，在推导出小波变换定义的基础上，重点介绍小波分析所涉及到的小波函数、小波分解与重构等内容。

10.1.1 小波分析与短时 Fourier 变换

1822 年，法国工程师 Fourier 指出，一个任意函数 $x(t)$ 都可以分解为无穷多个不同频率正弦信号的和，这即是谐波分析的基本概念。由于正弦信号有着广泛的应用背景，因此，傅里叶分析很快得到了发展，并广泛应用于电学、声学、机械学、生物医学等众多工程。遗憾的是，要在计算机中对某一时域信号作 Fourier 分析，必须将该时域信号全部采样后再进行

Fourier 分析，才能得到完整的频域信号，这使得 Fourier 分析的实时性非常差。有时我们关心的是局部时间信号的频率含量，而无需得到整个时间信号的频率含量。为了使得 Fourier 分析具有一定的实时性，通常对时域信号加窗，即截取一小段时域信号进行 Fourier 分析。

1. Fourier 变换

对于 $f(t) \in L$ 空间的能量信号 $f(t)$ ，即时间连续信号 $f(t)$ ，它的 Fourier 变换定义为

$$F(\omega) = \int_{-\infty}^{\infty} f(t) e^{-j\omega t} dt \quad (10-1)$$

$F(\omega)$ 的逆 Fourier 变换定义为

$$f(t) = \int_{-\infty}^{\infty} F(\omega) e^{j\omega t} d\omega \quad (10-2)$$

实际应用中，为了计算 Fourier 变换及其逆 Fourier 变换，需要用数值积分，取 $f(t)$ 在实数集 \mathbf{R} 上的离散点值来计算这个积分。设 $f(t)$ 由采样得到，采样间隔为 Δt ， $f(n\Delta t)$ ($n=0, 1, \dots, N-1$) 为采样值，对应的离散 Fourier 变换和逆 Fourier 变换为

$$F(\omega) = \sum_{n=0}^{N-1} f(n\Delta t) e^{-j\omega n\Delta t} \quad (10-3)$$

$$f(n) = \sum_{k=0}^{N-1} F(k) e^{j\frac{2\pi}{N} nk} \quad (10-4)$$

2. 短时 Fourier 变换

短时 Fourier 变换即时间信号加窗后的 Fourier 变换，其定义为

$$w_b F(\omega) = \int_{-\infty}^{\infty} e^{-j\omega t} f(t) \overline{w(t-b)} dt \quad (10-5)$$

式中， $w(t)$ 为一个窗口函数。窗口函数 $w(t)$ 的中心 t^* 与半径 $\Delta\omega$ 分别定义为

$$t^* = \frac{1}{\|w\|_2^2} \int_{-\infty}^{\infty} t |w(t)|^2 dt \quad (10-6)$$

$$\Delta\omega = \frac{1}{\|w\|_2} \left\{ \int_{-\infty}^{\infty} (t - t^*) |w(t)|^2 dt \right\}^{1/2} \quad (10-7)$$

这时， $w_b F(\omega)$ 给出了时间信号在时间窗

$$[t^* + b - \Delta\omega, t^* + b + \Delta\omega] \quad (10-8)$$

的局部信息，时间信号 $f(t)$ 的加窗过程如图 10-1 所示。

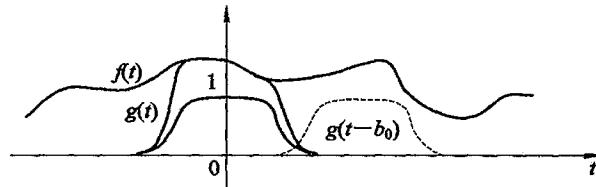


图 10-1 窗口 Fourier 变换

3. 小波分析

如果把短时 Fourier 变换中的窗口函数 $w_{a,b}(t)$ 替代为 $\psi_{a,b}(t)$ ，其中：

$$\psi_{a,b}(t) = |\alpha|^{-1/2} \psi\left(\frac{t-b}{\alpha}\right) \quad (10-9)$$

那么式(10-5)变为

$$w_\phi f(a, b) = |a|^{-1/2} \int_{-\infty}^{\infty} f(t) \overline{\phi\left(\frac{t-b}{a}\right)} dt \quad (10-10)$$

该式即为小波变换定义式。

对应于式(10-10)，小波逆变换为

$$f(t) = \frac{1}{C_\phi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{a^2} w_\phi f(a, b) \phi\left(\frac{t-b}{a}\right) db da \quad (10-11)$$

比较式(10-5)与式(10-10)，可以看到短时 Fourier 变换与小波变换之间的类似性，它们都是函数 $f(t)$ 与另一个具有两个指标函数族的内积。

对于 $\phi(t)$ 的一个典型的选择是：

$$\phi(t) = (1 - t^2) \exp\left(-\frac{t^2}{2}\right) \quad (10-12)$$

它是 Gauss 函数二阶导数，有时称这个函数为墨西哥帽函数，因为它像墨西哥帽的截面。墨西哥帽函数在时间域与频率域都有很好的局部化功能，函数图形如图 10-2 所示。

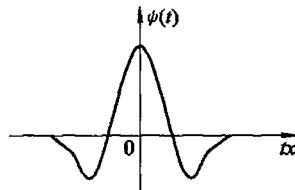


图 10-2 墨西哥帽函数图形

短时 Fourier 变换与小波变换之间的不同可由窗口函数的图形来说明，如图 10-3 所示。对于 $w_{\omega, b}$ ，不管 ω 值的大小，具有同样的宽度。相比之下， $\phi_{a, b}$ 在高频($1/a$ 相当于 Fourier 变换中的 ω)， a 越大，频率越低)时很窄，低频时很宽。因此，在很短暂的高频信号上，小波变换能比窗口 Fourier 变换更好地进行“移近”观察。

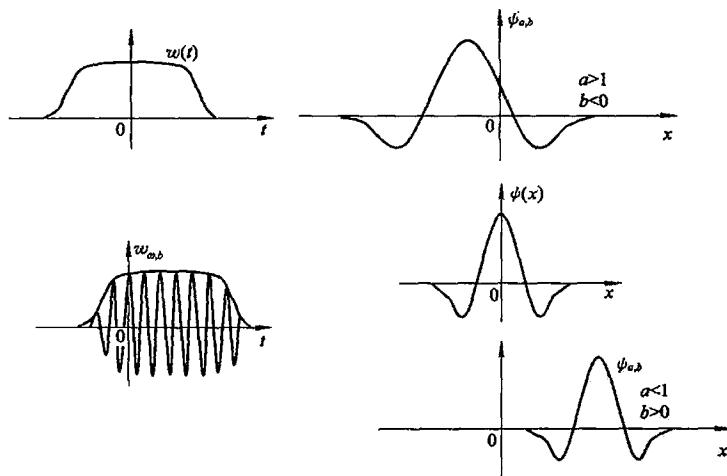


图 10-3

(a) 窗口 Fourier 变换函数 $w_{\omega, b}$ 的形状；(b) 小波 $\phi_{a, b}$ 的形状

10.1.2 离散小波

如果 a, b 都是离散值。这时，对于固定的伸缩步长 $a_0 \neq 0$ ，可选取 $a=a_0^m, m \in \mathbb{Z}$ ，不失一般性，可假设 $a_0 > 0$ 或 $a_0 < 0$ 。在 $m=0$ 时，取固定的 $b_0 (b_0 > 0)$ 整数倍离散化 b ，选取 b_0

使 $\psi(x - nb_0)$ 覆盖整个实轴。选取 $a = a_0^m$, $b = nb_0 a_0^m$, 其中 m, n 取遍整个整数域, 而 $a_0 > 1$, $b_0 > 0$ 是固定的。于是, 相应的离散小波函数族为

$$\psi_{m,n}(t) = a_0^{-m/2} \psi\left(\frac{x - nb_0 a_0^m}{a_0^m}\right) = a_0^{-m/2} \psi(a_0^{-m}x - nb_0) \quad (10-13)$$

对应的离散小波变换系数为

$$C_{m,n} = \int_0^\infty f(t) \psi_{m,n}^*(t) dt$$

离散小波逆变换为

$$f(t) = C \sum_{-\infty}^{\infty} \sum_{-\infty}^{\infty} C_{m,n} \psi_{m,n}(t)$$

式中, C 为一常数。

10.1.3 小波级数

对应于 Fourier 级数的定义:

$$f(x) = \sum_{k=-\infty}^{\infty} F(k\omega_0) e^{jk\omega_0 t} \quad k = 0, \pm 1, \dots, \pm \infty \quad (10-14)$$

式中, $F(k\omega_0) = \frac{1}{T} \int_t^{t+T} f(x) e^{-jk\omega_0 t} dt$ 。

同样可以定义小波级数:

$$f(x) = \sum_{j,k \in \mathbb{Z}} c_{j,k} \psi_{j,k}(x) = \sum_{k \in \mathbb{Z}} d_{j,k} \bar{\psi}_{j,k}(x) \quad (10-15)$$

式中, $\begin{cases} c_{j,k} = \langle f, \bar{\psi}_{j,k} \rangle \\ d_{j,k} = \langle f, \psi_{j,k} \rangle \end{cases}$, 称这两个无限级数为“小波级数”, 并且是 $L^2(R)$ 收敛的, 即 $c_{j,k}$

和 $d_{j,k}$ 的绝对值随着 j 和 k 的增大, 最终趋于 0; $f(x)$ 在实数域内能量有限。

10.1.4 多分辨分析

1. 多分辨分析的概念

如何由 $\phi(x) \in L^2(R)$ 出发, 使由 $\phi_{k,n}(x)$ 张成 $L^2(R)$ 的闭子空间

$$V_k = \text{clos}_{L^2(R)} \langle \phi_{k,n}(x) : n \in \mathbb{Z} \rangle \quad (10-16)$$

$\{\phi(x-n) : n \in \mathbb{Z}\}$ 是 V_0 的一个 Riezz 基, $\phi(x)$ 称为尺度函数, 这就是多分辨分析。

设 $\phi(x)$ 生成一个多分辨分析 $\{V_k\}$, 由于 $\phi(x) \in V_0 \subset V_1$, 所以 $\phi(x)$ 可以用 V_1 的基底 $\{\phi_{1,n} : n \in \mathbb{Z}\}$ 表示。由于 $\{\phi_{1,n} : n \in \mathbb{Z}\}$ 是 V_1 的一个 Riezz 基, 所以存在唯一 l^2 序列 $\{p_n\}$, 即离散的, 且其平方和为有限值的 $\{p_n\}$, 使

$$\phi(x) = \sum_{n=-\infty}^{\infty} p_n \phi(2x - n) \quad (10-17)$$

式(10-17)即为函数 $\phi(x)$ 的两尺度关系, 系列 $\{p_n\}$ 称为两尺度序列。

对于模为 1 的复数 z , 引入如下记号

$$P(z) = \frac{1}{2} \sum_{n=-\infty}^{\infty} p_n z^n \quad (10-18)$$

称为序列 $\{p_n\}$ 的符号。对式(10-17)两边作 Fourier 变换, 则得到两尺度关系式

$$\hat{\phi}(\omega) = P(z) \hat{\phi}\left(\frac{\omega}{2}\right), \quad z = e^{-j\omega/2} \quad (10-19)$$

同样地, 由于 $\psi(x) \in W_0 \subset V_1$, 所以存在唯一 L^2 序列 $\{q_n\}$, 使

$$\psi(x) = \sum_{n=-\infty}^{\infty} q_n \phi(2x - n) \quad (10-20)$$

引入序列 $\{q_n\}$ 的符号

$$Q(z) = \frac{1}{2} \sum_{n=-\infty}^{\infty} q_n z^n \quad (10-21)$$

对式(10-20)两边作 Fourier 变换, 类似地得到

$$\hat{\psi}(\omega) = Q(z) \hat{\phi}\left(\frac{\omega}{2}\right), \quad z = e^{-j\omega/2} \quad (10-22)$$

2. 分解算法与重构算法

由前所述可知, 对于 $f(x) \in L^2(R)$, 它有唯一分解:

$$f(x) = \sum_{k=-\infty}^{\infty} g_k(x) = \dots + g_{-1}(x) + g_0(x) + g_1(x) + \dots \quad (10-23)$$

式中, $g_k(x) \in W_k$. 令 $f_k(x) \in V_k$, 则有:

$$f_k = g_{k-1}(x) + g_{k-2}(x) + \dots \quad (10-24)$$

并且

$$f_k(x) = g_{k-1}(x) + f_{k-1}(x) \quad (10-25)$$

令

$$M(z) = \begin{bmatrix} P(z) & P(-z) \\ Q(z) & Q(-z) \end{bmatrix} \quad (10-26)$$

在 $|z|=1$ 上, 作函数

$$G(z) = \frac{Q(-z)}{\det M(z)}, \quad H(z) = \frac{-P(z)}{\det M(z)} \quad (10-27)$$

则

$$M^T(z)^{-1} = \begin{bmatrix} G(z) & G(-z) \\ H(z) & H(-z) \end{bmatrix} \quad (10-28)$$

对于符号 $G(z)$ 、 $H(z)$ 的序列 $\{g_n\}$, $\{h_n\} \in l^1$, 存在如下的分解关系式:

$$\phi(2x - l) = \frac{1}{2} \sum_{n=-\infty}^{\infty} \{g_{2n-l} \phi(x - n) + h_{2n-l} \psi(x - n)\}, \quad l \in ZZ \quad (10-29)$$

若令 $a_n = g_{-n}/2$, $b_n = h_{-n}/2$, 则(10-29)式变成:

$$\phi(2x - l) = \sum_{n=-\infty}^{\infty} \{a_{l-2n} \phi(x - n) + b_{l-2n} \psi(x - n)\}, \quad l = 0, \pm 1, \pm 2, \dots \quad (10-30)$$

为计算方便及以免产生混淆, 有:

$$f_k(x) = \sum_{j=-\infty}^{\infty} c_{k,j} \phi(2^k x - j) \quad (10-31)$$

$$g_k(x) = \sum_{j=-\infty}^{\infty} d_{k,j} \phi(2^k x - j) \quad (10-32)$$

在 $c_{k,j}$, $d_{k,j}$ 中, k 代表分解的“水平”, 也即分解的层次。

对于每个 $f(x) \in L^2(\mathbb{R})$, 固定 $N \in \mathbb{Z}$, 设 f_N 是 f 在空间 V_N 上的投影, 有:

$$f_N = \text{proj}_{V_N} f \quad (10-33)$$

可以把 V_N 看做是“抽样空间”, 而把 f_N 看做 f 在 V_N 上的“数据”(或者说测量采样值)。由于:

$$V_N = W_{N-1} + V_{N-1} = W_{N-1} + W_{N-2} + \cdots + W_{N-M} + V_{N-M} \quad (10-34)$$

所以, $f_N(x)$ 有唯一分解

$$f_N(x) = g_{N-1}(x) + g_{N-2}(x) + \cdots + g_{N-M} + f_{N-M} \quad (10-35)$$

对于固定的 k , 由 $\{c_{k+1,n}\}$ 求 $\{c_{k,n}\}$ 、 $\{d_{k,n}\}$ 的算法称为分解算法。应用分解关系式 (10-30) 有

$$\begin{aligned} f_{k+1}(x) &= \sum_{l=-\infty}^{\infty} c_{k+1,l} \phi(2^{k+1}x - l) \\ &= \sum_l c_{k+1,l} \left[\sum_n (a_{l-2n} \phi(2^k x - n) + b_{l-2n} \psi(2^k x - n)) \right] \\ &= \sum_n \left\{ \sum_l a_{l-2n} c_{k+1,l} \right\} \phi(2^k x - n) + \sum_n \left\{ \sum_l b_{l-2n} c_{k+1,l} \right\} \psi(2^k x - n) \end{aligned}$$

分解 $f_{k+1}(x) = f_k(x) + g_k(x)$, 得到:

$$\sum_n \left\{ c_{k,n} - \sum_l a_{l-2n} c_{k+1,l} \right\} \phi(2^k x - n) + \sum_n \left\{ d_{k,n} - \sum_l b_{l-2n} d_{k+1,l} \right\} \psi(2^k x - n) = 0$$

所以, 由 $\{\phi_{k,n} : n \in \mathbb{Z}\}$, $\{\psi_{k,n} : n \in \mathbb{Z}\}$ 的线性无关性以及 $V_k \cap W_k = \{0\}$, 得到分解算法:

$$\begin{cases} c_{k,n} = \sum_l a_{l-2n} c_{k+1,l} \\ d_{k,n} = \sum_l b_{l-2n} d_{k+1,l} \end{cases} \quad (10-36)$$

分解过程如图 10-4 所示。

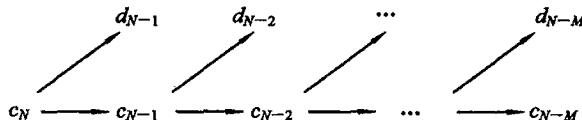


图 10-4 分解过程

在实际计算中, 假定取值点所对应的 $f(x)$ 的水平为 N , 即

$$f(x) \approx f_N$$

对于某个正数 N ($0 \leq M \leq N$), 信号由 N 水平分解到 $N-M$ 水平, 即已知 $\{c_{N,n}\}$, 求 $\{d_{k,n}\}$ 及 $\{c_{k,n}\}$, $k = N-1, \dots, N-M$ 。

同样地, 固定 k , 由 $\{c_{k,n}\}$ 、 $\{d_{k,n}\}$ 求 $\{c_{k+1,n}\}$ 的算法称为重构算法。应用两尺度关系有

$$\begin{aligned} f_k(x) + g_k(x) &= \sum_l c_{k,l} \phi(2^k x - l) + \sum_l d_{k,l} \psi(2^k x - l) \\ &= \sum_l c_{k,l} \sum_n p_n \phi(2^{k+1} x - 2l - n) + \sum_l d_{k,l} \sum_n q_n \psi(2^{k+1} x - 2l - n) \\ &= \sum_l \sum_n (c_{k,l} p_{n-2l} + d_{k,l} q_{n-2l}) \phi(2^{k+1} x - n) \\ &= \sum_n \left\{ \sum_l (p_{n-2l} c_{k,l} + q_{n-2l} d_{k,l}) \right\} \phi(2^{k+1} x - n) \end{aligned}$$

因为 $f_k(x) + g_k(x) = f_{k+1}(x)$, 有 $f_{k+1} = \sum_n c_{k+1} \phi(2^{k+1}x - n)$ 及 $\{\phi_{k+1,n} : n \in \mathbb{Z}\}$ 的线性无关性, 得到重构算法:

$$c_{k+1,n} = \sum_l (p_{n-2l} c_{k,l} + q_{n-2l} d_{k,l}) \quad (10-37)$$

重构过程如图 10-5 所示。

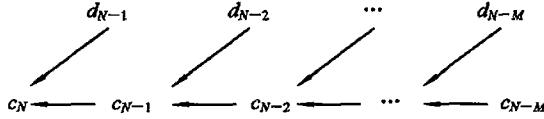


图 10-5 重构过程

10.1.5 小波包分析

小波包分析是多分辨分析的推广, 它提供了更为丰富和精确的信号分析方法。小波包元素是由三个参数来确定的一个波形, 这三个最基本的参数是: 位置、尺度(与一般小波分解一样)和频率。在正交小波分解过程中, 一般的方法是将低频系数向量继续分解成两部分, 高频系数不再分解。小波分解过程中, 系数 c_k 与 d_k ($k=N, N-1, \dots, 1$) 所对应的频域段如图 10-6(a) 所示。而在小波包分解中, 每一个高频系数向量也像低频部分的分解一样, 被分解成两部分, 因而, 它提供了更丰富的分析方法。在一维情况下, 它产生一个完整的二叉树; 在二维情况下, 它产生一个完整的四叉树。小波包分解过程中, 系数 c_k 与 d_k ($k=N, N-1, \dots, 1$) 所对应的频域段如图 10-6(b) 所示。

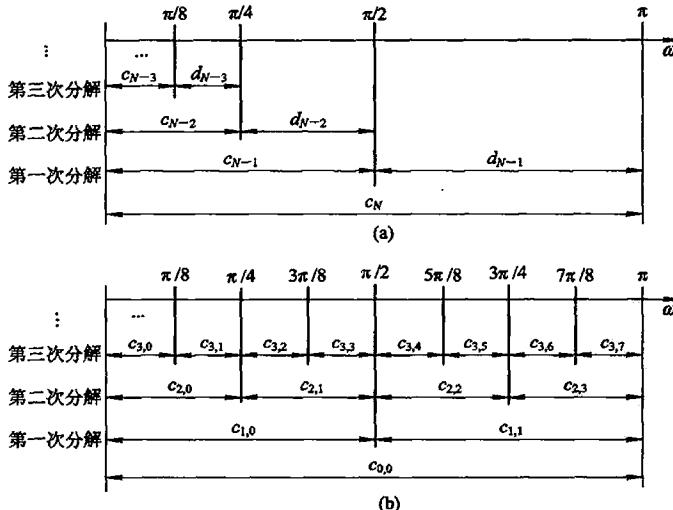


图 10-6 小波分析与小波包分析中系数对应的频域段

(a) 小波分析; (b) 小波包分析

限于篇幅, 小波包分析的理论在此不再详述, 有关 Matlab 中小波包分析函数的介绍详见 10.2 节。

10.2 Matlab 工具箱中小波分析函数

在一维信号的分析中, Matlab 工具箱提供了许多小波分析功能函数, 为方便读者理解本章后面的小波分析示例, 更快地掌握用小波分析进行一维信号处理的方法, 本节给出了一些一维信号处理中比较有用的函数的功能和应用方法。为保持与 Matlab 中编程的符号一致, 本章的所有符号, 不论是变量还是常量、向量、参数, 字体均为正体, 且不加粗。

10.2.1 小波分析函数

1. orthfilt 函数

orthfilt 函数用于来计算与某一小波对应的尺度滤波器相关的四个滤波器, 调用格式为

$[Lo_D, Hi_D, Lo_R, Hi_R] = \text{orthfilt}(W)$

其中, Lo_D 为分解的低通滤波器, 对应于式(10-36)中的 $a(n)$; Hi_D 为分解的高通滤波器, 对应于式(10-36)中的 $b(n)$; Lo_R 为重构的低通滤波器, 对应于式(10-37)中的 $p(n)$; Hi_R 为重构的高通滤波器, 对应于式(10-37)中的 $q(n)$ 。

2. wfilters 函数

wfilters 函数用于计算正交或双正交小波 wname 相关的四个滤波器, 调用格式为

格式 1 $[Lo_D, Hi_D, Lo_R, Hi_R] = \text{wfilters}('wname')$

格式 2 $[F1, F2] = \text{wfilters}('wname', 'type')$

格式 1 返回的四个滤波器分别是 Lo_D (分解的低通滤波器)、 Hi_D (分解的高通滤波器)、 Lo_R (重构的低通滤波器)和 Hi_R (重构的高通滤波器)。其中 Lo_D 和 Hi_D 是 Lo_R 和 Hi_R 的对偶算子, 也可以分别理解为 Lo_R 和 Hi_R 的共轭转置矩阵。在信号的处理中, 它们的作用分别对应于分解图 10-4 和重构图 10-5。以分解层数为 4 层为例, 算子的作用可以用图 10-7 表示。

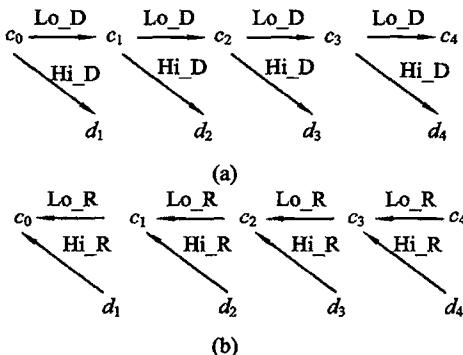


图 10-7 算子的作用

(a) 分解; (b) 重构

格式 2 根据 type 的设定值来计算不同的正交或双正交小波滤波器，具体设置如表 10-1 所示。

表 10-1 type 设定值

type 的设定值	返回滤波器序列	滤波器含义
D	Lo_D, Hi_D	分解滤波器
R	Lo_R, Hi_R	重构滤波器
L	Lo_D, Lo_R	低通滤波器
H	Hi_D, Hi_R	高通滤波器

3. dwt 函数

dwt 函数用于单尺度一维离散小波变换，调用格式为

格式 1 [CA, CD]=dwt(X, 'wname')

格式 2 [CA, CD]=dwt(X, Lo_D, Hi_D)

格式 3 [CA, CD] = dwt(X, 'wname', 'mode', mode)

格式 4 [CA, CD] = dwt(X, Lo_D, Hi_D, 'mode', mode)

该函数是一个一维小波分析函数，它用来计算信号 X 的周期小波分解的低频向量 CA 和高频系数向量 CD。在分解时，可以采用小波函数进行分解（如格式 1 和格式 3），也可以采用分解滤波器进行分解（如格式 2 和格式 4）。格式 3 和格式 4 中的模式选择详见 dwtmode 函数的介绍。

若 mode='per'，则 CA 与 CD 向量的长度为

X 的长度 length(X) 为奇数

$$\text{length}(\text{CA}) = \text{length}(\text{CD}) = \frac{\text{length}(X) + 1}{2}$$

X 的长度 length(X) 为偶数

$$\text{length}(\text{CA}) = \text{length}(\text{CD}) = \frac{\text{length}(X)}{2}$$

否则，CA 与 CD 向量的长度为

$$\text{length}(\text{CA}) = \text{length}(\text{CD}) = \text{floor} \frac{\text{length}(X) + \text{length}(\text{Lo}_R) - 1}{2}$$

4. dwtmode 函数

dwtmode 函数用于设置离散小波变换拓展模式，调用格式为

格式 1 dwtmode

格式 2 dwtmode('mode')

当对信号或图像的边缘进行处理时，需要对信号的边缘进行拓展。一般说来，拓展模式有三种。dwtmode 函数的功能就是在对信号（或图像）进行离散小波变换或小波包变换时进行模式拓展设置，不同的模式代表了对信号（图像）边缘不同的处理方法。各种模式的具体意义如表 10-2 所示。

表 10-2 模式类型

模式类型	类型说明
zpd	补零模式，这是一种缺省的类型模式
sym	对称延拓模式，即把边缘值进行复制
spd	1 阶平滑模式，即对边缘进行一次插值平滑处理
sp0	0 阶平滑模式，即将边界值进行恒定值扩展
ppd	周期延拓模式，即在边界进行周期延拓
per	周期化模式，这使得 dwt 的功能同 dwtperr 相同

5. idwt

idwt 函数用于单尺度一维离散小波逆变换，调用格式为

格式 1 $X = \text{idwt}(\text{CA}, \text{CD}, \text{'wname'})$

格式 2 $X = \text{idwt}(\text{CA}, \text{CD}, \text{Lo_R}, \text{Hi_R})$

格式 3 $X = \text{idwt}(\text{CA}, \text{CD}, \text{'wname'}, \text{L})$

格式 4 $X = \text{idwt}(\text{CA}, \text{CD}, \text{Lo_R}, \text{Hi_R}, \text{L})$

格式 5 $X = \text{idwt}(\text{CA}, \text{CD}, \text{'wname'}, \text{'mode'}, \text{mode})$

格式 6 $X = \text{idwt}(\text{CA}, \text{CD}, \text{Lo_R}, \text{Hi_R}, \text{'mode'}, \text{mode})$

格式 7 $X = \text{idwt}(\text{CA}, \text{CD}, \text{'wname'}, \text{L}, \text{'mode'}, \text{mode})$

格式 8 $X = \text{idwt}(\text{CA}, \text{CD}, \text{Lo_R}, \text{Hi_R}, \text{L}, \text{'mode'}, \text{mode})$

格式 1 是由向量 CA、CD 和小波名称重建除向量 X；格式 2 是由向量 CA、CD 及重建滤波器 Lo_R 和 Hi_R 重建除向量 X；格式 3、4 同格式 1、2 相比，其返回的向量是格式 1、2 返回的向量 X 中长度为 L 的中间部分；格式 5、6、7、8 可以通过设置变换模式来进行小波变换，mode 的取值见表 10-2。当小波变换模式为周期模式（mode='per'）时，length(X) = 2length(CA)；当小波变换模式为其它模式时，length(X) = 2length(CA) - length(Lo_R) + 2。

6. dwtperr 函数

dwtperr 函数用于单尺度一维离散小波变换（周期性），调用格式为

格式 1 $[\text{CA}, \text{CD}] = \text{dwtperr}(\text{X}, \text{'wname'})$

格式 2 $[\text{CA}, \text{CD}] = \text{dwtperr}(\text{X}, \text{Lo_D}, \text{Hi_D})$

这个函数的功能同 dwt 函数类似，只不过信号 X 需要经过周期性地延拓，更确切地说，当 length(X) 为偶数时，信号拓展成为

$\text{extX} = [\text{X}(\text{length}(\text{X}) - \text{length}(\text{Lo_D}) + 1), \text{X}, \text{X}(1:\text{length}(\text{Lo_D}))]$

信号拓展后，用一般的卷积和抽样进行运算，最后只保留中间部分，长度为 $\text{ceil}(\text{length}(\text{X}))$ 。

7. idwtperr 函数

idwtperr 函数用于单尺度一维离散小波重构（周期性），调用格式为

格式 1 $X = \text{idwtper}(\text{CA}, \text{CD}, \text{'wname'})$
 格式 2 $X = \text{idwtper}(\text{CA}, \text{CD}, \text{Lo_R}, \text{Hi_R})$
 格式 3 $X = \text{idwtper}(\text{CA}, \text{CD}, \text{'wname'}, \text{L})$
 格式 4 $X = \text{idwtper}(\text{CA}, \text{CD}, \text{Lo_R}, \text{Hi_R}, \text{L})$

格式 1、3 是用小波函数进行重构，格式 2、4 是用重构滤波器进行重构，其中 CA 和 CD 的长度是相等的，Lo_R 和 Hi_R 的长度也是相等的，返回系数 X 为重构后信号的向量。如果 CA 的长度为 la，则 X 的长度为 $\text{length}(X) = 2 * la$ 。对于格式 2、4，则是对信号中间长度为 L 的部分进行重构， $L \leq \text{length}(X)$ 。

8. appcoef 函数

appcoef 函数用于提取一维小波变换低频系数，调用格式为

格式 1 $A = \text{appcoef}(\text{C}, \text{L}, \text{'wname'}, \text{N})$
 格式 2 $A = \text{appcoef}(\text{C}, \text{L}, \text{'wname'})$
 格式 3 $A = \text{appcoef}(\text{C}, \text{L}, \text{Lo_R}, \text{Hi_R})$
 格式 4 $A = \text{appcoef}(\text{C}, \text{L}, \text{Lo_R}, \text{Hi_R}, \text{N})$

这个函数用于从小波分解结构 [C, L] 中提取一维信号的低频系数，格式 1、4 中的 N 为尺度，且是满足关系式 $0 \leq N \leq \text{length}(\text{L}) - 2$ 的一个整数。格式 2、3 用于提取最后一尺度 ($N = \text{length}(\text{L}) - 2$) 的小波变换低频系数。返回参数 A 是一个向量，其长度为 $\text{length}(\text{S})/2^N$ 。

9. detcoef 函数

detcoef 函数用于提取一维小波变换高频系数，调用格式为

$D = \text{detcoef}(\text{C}, \text{L}, \text{N})$
 $D = \text{detcoef}(\text{C}, \text{L})$

10. upwlev 函数

upwlev 函数用于单尺度一维小波分解的重构，调用格式为

格式 1 $[\text{NC}, \text{NL}, \text{CA}] = \text{upwlev}(\text{C}, \text{L}, \text{'wname'})$
 格式 2 $[\text{NC}, \text{NL}, \text{CA}] = \text{upwlev}(\text{C}, \text{L}, \text{Lo_R}, \text{Hi_R})$

这个函数用于对小波分解结构 [C, L] 进行单尺度重构，返回上一尺度的分解结构 [NC, NL]，并提取最后一尺度的低频系数向量 CA，即如果 [C, L] 是尺度 N 的分解结构，则 [NC, NL] 是尺度 N-1 的分解结构。

11. wrcoef 函数

wrcoef 函数用于对一维小波系数进行单支重构，调用格式为

格式 1 $X = \text{wrcoef}(\text{'type'}, \text{C}, \text{L}, \text{'wname'}, \text{N})$
 格式 2 $X = \text{wrcoef}(\text{'type'}, \text{C}, \text{L}, \text{Lo_R}, \text{Hi_R}, \text{N})$

函数对一维信号的分解结果 [C, L] 用指定的小波函数重构滤波器进行重构。格式中的 type 的取值有 a, d 两种：

当 type='a' 时，指对低频部分进行重构，此时 N 的取值范围为 $0 \leq N \leq \text{length}(\text{L}) - 2$ ；
 当 type='d' 时，指对高频部分进行重构，此时 N 的取值范围为 $0 < N \leq \text{length}(\text{L}) - 2$ 。

12. upcoef 函数

upcoef 函数用于一维系数的直接小波重构，调用格式为

格式 1 $Y = \text{upcoef}(O, X, 'wname', N)$

格式 2 $Y = \text{upcoef}(O, X, 'wname', N, L)$

格式 3 $Y = \text{upcoef}(O, X, Lo_R, Hi_R, N)$

格式 4 $Y = \text{upcoef}(O, X, Lo_R, Hi_R, N, L)$

函数计算向量 X 向上 N 步的重构小波系数，其中 N 是严格的正数， $O='a'$ 或 ' d '。当 $O='a'$ 时，函数对低频系数重构；当 $O='d'$ 时，函数对高频系数重构。格式 2、4 中的 L 表示只对向量 X 中间长度为 L 的部分进行重构。格式 1、3 中的 N 的默认值为 1。

13. wavedec 函数

wavedec 函数用于多尺度一维小波分解，调用格式为

格式 1 $[C, L] = \text{wavedec}(X, N, 'wname')$

格式 2 $[C, L] = \text{wavedec}(X, N, Lo_D, Hi_D)$

其中，wavedec 的功能是函数用小波分解滤波器完成对信号 X 的一维多尺度分解， N 为尺度，且是严格的正数。输出参数 C 是由 $[CA_N, CD_N, CD_{N-1}, \dots, CD_1]$ 组成的分解向量， L 是由 $[CA_N$ 的长度， CD_N 的长度， CD_{N-1} 的长度， \dots ， CD_1 的长度， X 的长度] 组成的记录簿向量。以一个 3 尺度分解为例，其分解结构的组织形式如图 10-8 所示。

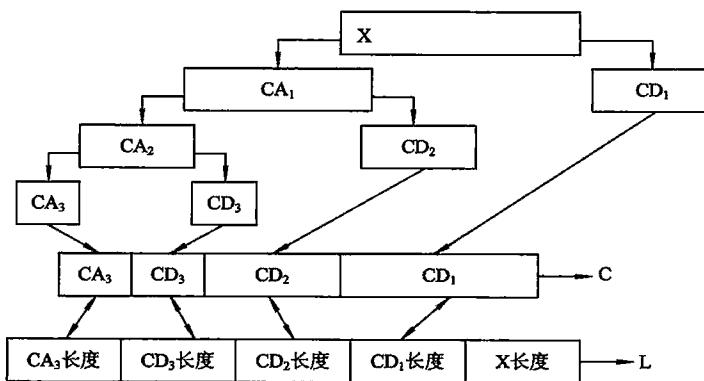


图 10-8 分解结构的组织形式

如果格式 2 中的分解滤波器向量 Lo_D 和 Hi_D 刚好对应于格式 1 中的滤波，其名称为 $wname$ ，且分解层数 N 也设置成同格式 1 一样，则格式 2 返回的结果同格式 1 完全相同。

给定一个长度为 N 的信号 S ，分解后的分解系数由两部分组成：低频系数向量 CA_1 和高频系数向量 CD_1 ，向量 CA_1 是由信号 S 与低通分解滤波器 Lo_D 经过卷积运算得到的，向量 CD_1 是由信号 S 与高通分解滤波器 Hi_D 经过卷积运算得到的。其分解图如图 10-9 所示。

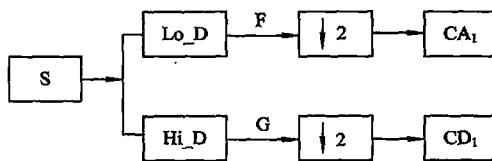


图 10-9 分解图

在图 10-9 中, 如果滤波器的长度均为 $2M$, $n = \text{length}(S)$, 则信号 F 和 G 的长度为 $n + 2M - 1$, 所以系数 CA_1 和 CD_1 的长度为 $\text{floor}\left(\frac{n-1}{2}\right) + M$ 。

再进行下一步分解, 用同样的方法把低频系数 CA_1 分成两部分, 即把上面方框图中的 S 用 CA_1 代替, 分解后返回低频系数 CA_2 和次高频系数 CD_2 。往下分解依此类推。

14. waverec 函数

waverec 函数用于多尺度一维小波重构, 调用格式为

格式 1 $X = \text{waverec}(C, L, 'wname')$

格式 2 $X = \text{waverec}(C, L, Lo_R, Hi_R)$

该函数用指定的小波函数或重构滤波器对小波分解结构 $[C, L]$ 进行多尺度一维小波重构, 它是 waverec 函数的逆函数, 有:

$X = \text{waverec}(\text{wavedec}(X, N, 'wname'), 'wname')$

格式 1 是用小波函数进行重构, 格式 2 是用重构滤波器进行重构。

15. wden 函数

调用该函数可以直接对一维信号消噪, 其调用格式为

格式 1 $[XD, CXD, LXD] = \text{wden}(X, T PTR, SORH, SCAL, N, 'wname')$

格式 2 $[XD, CXD, LXD] = \text{wden}(C, L, T PTR, SORH, SCAL, N, 'wname')$

它的返回值是经过对原始信号 X 进行消噪处理后的信号 XD 及其分解结构 $[CXD, LXD]$ 。另外, $SORH$ 指定软阈值 ($SORH = 's'$) 或硬阈值 ($SORH = 'h'$) 的选择; $T PTR$ 指定阈值选取的规则, 它有四种选项, 见表 10-3; N 为小波分解的层数; $wname$ 指定分解时所用的小波; 格式 2 中的 C, L 为输入信号在所选用小波 ' $wname$ ' 时的分解结构; 参数 $SCAL$ 是阈值尺度改变的比例, 它有三种选项, 见表 10-4。

表 10-3 参数 T PTR 的选项

T PTR 选项	阈值类型
rigrsure	采用无偏估计原则进行适应性阈值选择
heursure	选用首次选择的启发式变量作为阈值
sqtwoolog	选用 $\sqrt{2 * \log(\text{length}(X))}$ 作为阈值
minimaxi	鞍点阈值

表 10-4 参数 SCAL 的选项

SCAL 的选项	相应的模式
'one'	基本模式
'sln'	未知尺度的基本模式
'mln'	非白噪声的基本模式

若采用与格式 1 相同的设置，那么格式 2 所得到的结果与格式 1 完全相同。

16. wdencmp 函数

小波消噪的另一种更普遍的函数是 wdencmp，它可以直接对一维信号或二维信号进行消噪或压缩处理，方法也是通过对消波分解系数进行阈值量化来实现的，它可以让用户选择自己的阈值量化方案，其使用方式为

- (1) [XC, CXC, LXC, PERFL0, PERFL2]=wdencmp('opt', X, 'wname', N, THR, SORH, KEEPAPP)
- (2) [XC, CXC, LXC, PERFL0, PERFL2]=wdencmp ('opt', C, L, W, N, THR, SORH, KEEPAPP)

其中，各参数说明如下：

- opt='gbl'，并且 THR 是一个正的实数，则阈值是全局阈值；
- opt='lvd'，并且 THR 是向量，表示各层大小不同的阈值，长度为 N；
- X 是待处理的输入信号；
- wname 为所选择的小波的名称；
- N 为分解层次；
- SORH 为软阈值或硬阈值选择项，当 SORH='s' 时表示软阈值，当 SORH='h' 时表示硬阈值；
- KEEPAPP=1，将小波分解的低频系数不作任何处理，KEEPAPP=0，对小波分解的低频系数也进行阈值量化处理；
- XC 为消噪后的输出向量(输入是一维向量)或压缩输出矩阵；
- CXC、LXC 为 XC 的小波分解结构；
- PERFL2 和 PERFL0 为重构、压缩百分率：

$$\text{PERFL2} = 100 * (\text{vector - norm of CXC} / \text{vector - norm of C})^2$$
- [C, L] 为 X 的分解结构。

17. wnoise 函数

wnoise 函数用于产生噪声小波测试数据，调用格式为

格式 1 X = wnoise(FUN, N)

格式 2 [X, XN] = wnoise(FUN, N, SNRAT)

格式 3 [X, XN] = wnoise(FUN, N, SNRAT, INIT)

N 表示数据的长度为 2^N 。参数 XN 返回的是信号 X 与噪声的叠加信号，噪声的产生由噪声种子 INIT 设定，格式 2 默认噪声种子为 1。参数 SNRAT 设置返回信号 XN 的信噪

比,信号X的幅值随SNRAT设置值的绝对值增大而增大,并且具有正负特性。当SNRAT=0时,X是恒为0的0值信号。参数FUN设置所产生的信号类型,其设置的数值为1~6,分别对应于上述六个数值的信号波形如图10-10左边的一列所示,图中右边的一列分别对应于左边的信号在信噪比为4时的含噪声信号XN。

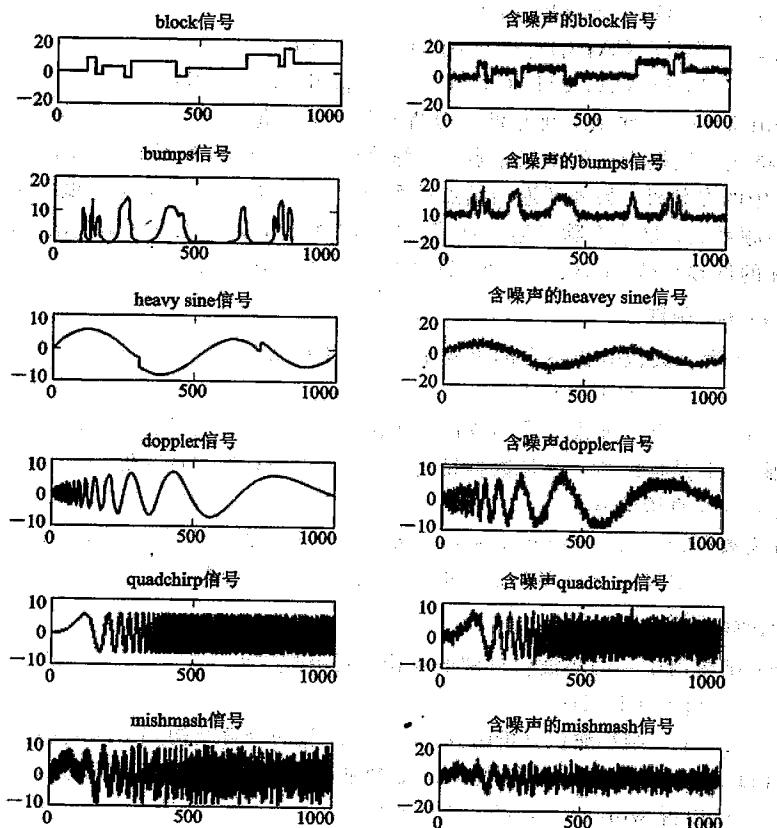


图10-10 不同类型信号及含噪信号的波形

10.2.2 小波包函数

1. wpdec 函数

一维小波包的分解函数,调用格式为:

格式1 $[T, D] = \text{wpdec}(X, N, 'wname', E, P)$

格式2 $[T, D, RN] = \text{wpdec}(X, N, 'wname')$

对于格式1,它根据小波包函数的'wname'(参见wfilters函数)、熵标准E和参数P对信号X进行N层小波包分解,并返回小波包分解结构[T,D](T为树结构,D为数据结构)。其中E用来指定熵标准,E的类型可以有'shannon'、'threshold'、'norm'、'log energy'、'sure'或'user'。P是一个可选的参数,它的选择根据参数E的值来确定。

格式 2 等价于 $[T, D, RN] = wpdec(X, N, 'wname', 'shannon')$, 即默认为 'shannon' 熵标准。

2. wpsplt 函数

wpsplt 为分割(分解)小波包函数, 调用格式为

格式 1 $[T, D] = wpsplt(T, D, N)$

格式 2 $[T, D, CA, CD] = wpsplt(T, D, N)$

格式 3 $[T, D, CA, CH, CV, CD] = wpsplt(T, D, N)$

wpsplt 是一个一维或二维小波包分析函数。它在重组一个结点后, 便更新树结构和数据结构。格式 1 根据指定的重组结点 N, 修改计算树结构 T 和数据结构 D; 格式 2 除了返回格式 1 中的参数外, 另外还返回结点的系数, 其中, CA 是结点 N 的低频系数, CD 是结点 N 的高频系数; 格式 3 除了返回格式 1 中的参数外, 另外还返回结点的系数, 其中, CA 是结点 N 的低频系数, CH、CV 和 CD 是结点 N 的高频系数。

3. wprcoef 函数

wprcoef 为小波包分析系数的重构函数, 调用格式为

$X = wprcoef(T, D, N)$

wprcoef 是一个一维或二维小波包分析函数, 它计算结点 N 的小波包分解系数的重构信号(图像), T 是树结构, D 是数据结构, $X = wprcoef(T, D)$ 等价于 $X = wprcoef(T, D, 0)$, 即完全重构原始信号。

另外需要强调一点, wprcoef 函数一次只能对一个结点进行重构, 而不能同时对多个结点进行重构。如果想要对多个结点进行重构, 需要多次调用该函数。

4. wprec 函数

wprec 函数用于一维小波包分解的重构, 调用格式为

$X = wprec(T, D)$

wprec 是一个一维小波包分析函数, 它对小波包的分解结构 $[T, D]$ 进行重构, 并返回重构后的向量 X, 其中 T 是树结构, D 是数据结构。因为有 $X = wprec(wpdec(X, N, 'wname'))$, 所以说 wprec 是 wpdec 的反函数。

5. wpjoin 函数

wpjoin 函数用于重新组合小波包, 调用格式为

格式 1 $[T, D] = wpjoin(T, D, N)$

格式 2 $[T, D, X] = wpjoin(T, D, N)$

格式 3 $[T, D] = wpjoin(T, D)$

格式 4 $[T, D, X] = wpjoin(T, D)$

wpjoin 是一个一维或二维小波包分析函数, 用来重新组合小波包, 即把结点 N(结点可以以索引的形式表示, 也可以用深度-位置的形式表示)以下的二叉子树去掉(更新后的小波包分解树不对结点 N 作进一步的分解)后, 返回一个更新后的小波包分解结构。

格式 1 根据指定的结点 N, 去掉结点 N 以下的二叉数以后, 返回修改计算后的小波包分解树结构 T 和数据结构 D。

格式 2 除了返回格式 1 中的参数外, 另外还返回结点的小波包分解系数。

格式3 等价于 $[T, D] = \text{wpjoin}(T, D, 0)$ ，它去掉了根结点以下的所有二叉子树，更新后的小波包分解结构实质上对信号没有进行任何的小波包分解。

格式4 等价于 $[T, D, X] = \text{wpjoin}(T, D, 0)$ ，更新后的小波包分解结构实质上对信号没有进行任何尺度的小波包分解，返回后的小波包分解系数 X 同原始信号一样。

6. besttree 函数

besttree 函数用于计算最佳(优)树 TT ，调用格式为

格式1 $[TT, DD] = \text{besttree}(T, D)$

格式2 $[TT, DD, E] = \text{besttree}(T, D)$

格式3 $[TT, DD, E, N] = \text{besttree}(T, D)$

besttree 是一个一维或二维小波包分析函数。它能根据一个熵标准来计算初始树的最佳子树，算出的子树比初始树小得多。

根据小波包的组织方式，对于一个给定的正交小波，一个长度为 $N=2^L$ 的信号最多可以有 2^L 种不同的分解方式，这恰好是一个深度为 L 的完整的二叉子树的数目，是一个非常庞大的数目，一般来说，用枚举法进行一一列举是难以想象的。由于我们感兴趣的是根据一个简单而又可行的标准来寻找一个最佳的分解方式(即最佳分解树结构或最佳小波包基)和一个有效的算法，所以我们可以根据最小熵标准来进行处理。

对于格式1，它根据小波包分解结构 $[T, D]$ 来计算小波包分解的最佳树 TT 及相应的数据结构 DD 。

对于格式2，它返回计算后的最佳树结构 TT 、数据结构 DD 以及初始树每个结点的熵值向量 E (向量元素的顺序与结点的索引序号依次对应，即索引为 0, 1, 2, … 的结点熵值依次对应向量中第 1, 2, 3, … 个元素值)。

对于格式3，它除了返回格式2中所返回的参数外，同时还返回最佳树与初始树相比，所有被合并结点的索引序号的向量 N 。例如，若返回向量 $N=[2, 6]$ ，则表示初始树中索引号为 2 和 6 的结点以下的二叉子树被合并，即最佳树与初始树相比，索引为 2 和 6 的结点不再分解。

7. bestlevt 函数

bestlevt 函数用于计算最佳完整小波包树，调用格式为

格式1 $[TT, DD] = \text{bestlevt}(T, D)$

格式2 $[TT, DD, E] = \text{bestlevt}(T, D)$

bestlevt 是一个一维或二维的小波包分析函数，它可以根据一种熵标准化计算出初始树的最佳完整子树，这个完整的子树比初始树的深度小一些。格式1 根据最优深度小波包树的分解，修改计算小波包分解结构 $[T, D]$ ，并返回新的小波包分解结构 $[TT, DD]$ 。格式2 除了返回格式1中的参数外，还返回最优熵值 E 。

bestlevt 函数在功能上与 besttree 函数类似，它们之间的不同之处是：① bestlevt 函数返回的树结构 TT 首先是一个完整的二叉子树，然后在完整的基础上，它是一个最佳小波包分解树；besttree 函数返回的是一个从熵标准角度出发最佳的小波包分解树，但它不一定是一个完整的二叉子树。② bestlevt 函数返回的最佳熵值 E 是最佳分解树的总体熵，是一个数值；besttree 函数返回的是初始树每一个结点的熵值，是一个向量。

8. wentropy 函数

wentropy 函数用于计算小波包的熵，调用格式为

格式 1 $E = \text{wentropy}(X, T, P)$

格式 2 $E = \text{wentropy}(X, T)$

wentropy 是一个一维或二维的小波包分析函数。对于格式 1，它返回一个向量或矩阵 X 的熵值 E ，在 X 为向量或矩阵的情况下，输出 E 都是一个实数，参数 T 用来指定一个熵标准，它有 'shannon'、'threshold'、'norm'、'log energy'、'sure' 和 'user' 这几种类型， P 是一个可选的参数，它的选择根据参数 E 的值来决定。

对于格式 2，它等价于 $E = \text{wentropy}(X, T, 0)$ 。

9. entrupt 函数

entrupt 函数用于更新小波包的熵值，调用格式为

格式 1 $NDATA = \text{entrupt}(\text{TREE}, \text{DATA}, T)$

格式 2 $NDATA = \text{entrupt}(\text{TREE}, \text{DATA}, T, P)$

格式 1 根据一个给定的小波包分解结构 [TREE, DATA] 和熵标准 T ，返回更新后小波包分解数据结构 $NDATA$ 。此时，各结点的熵值发生了变化，可以用 wdatamgr 函数读取各结点的熵值大小。

格式 2 和格式 1 相比，多了一个可选择输入参数 P 。它返回的也是一个更新后的小波包分解数据结构 $NDATA$ ，同样可以用 wdatamgr 函数读取各结点的熵值大小。

10. ddencmp

ddencmp 函数返回消噪与压缩的默认值，调用格式为

$[THR, SORH, KEEAPP, CRIT] = \text{ddencmp}(\text{IN1}, \text{IN2}, X)$

参数说明：

- X : 输入信号；
- $IN1$: 指明是消噪还是压缩，若为 "den"，则表示消噪，若为 "cmp"，则表示压缩；
- $IN2$: 指明是使用小波还是小波包，若为 "wv"，则表示使用小波，若为 "wp"，则表示使用小波包。
- THR : 用小波或小波包进行消噪(压缩)所采用的阈值，取值为 "soft" 或 "hard"；
- $KEEPAPP$: 容许保持近似系数；
- $CRIT$: 平均信息量名称，只用于小波包，因此对于小波来说，本函数的输出参数实际上只有前三个。

ddencmp 函数返回输入信号 X 用小波或小波包进行压缩或消噪的默认值，输入信号可以是一维信号(向量)，也可以是二维信号(矩阵)。

11. wpthcoef 函数

wpthcoef 为小波包系数的阈值处理函数，调用格式为

$NDATA = \text{wpthcoef}(\text{DATA}, \text{TREE}, \text{KEEPAPP}, \text{SORH}, \text{THR})$

wpthcoef 函数通过系数阈值处理由小波包分解结构 [DATA, TREE] 获得的新的数据结构。如果 $KEEPAPP=1$ ，低频系数不作阈值处理，否则，进行阈值处理。如果 $SORF= 's'$ ，则使用软阈值；如果 $SORF= 'h'$ ，则使用硬阈值。THR 为阈值的大小，它是一个全局阈值。

12. wpdenccmp 函数

wpdenccmp 函数的功能为使用小波包进行信号消噪或信号压缩，调用格式为

$[XD, TREED, DATAD, PERFO, PERFL2] = \text{wpdenccmp}(X, SORH, N, 'wname', CRIT, PAR, KEEAPP)$

该函数通过小波包系数阈值处理对输入信号 X 进行信号消噪或压缩，并返回消噪(压缩)信号 XD。

13. wpcoef 函数

wpcoef 为小波包系数求取函数，调用格式为：

格式 1 $X = \text{wpcoef}(T, D, N)$

格式 2 $X = \text{wpcoef}(T, D)$

wpcoef 是一个一维或二维的小波包分析函数。格式 1 返回与结点 N 对应的系数，其中 T 是树结构，D 是数据结构；如果 N 不存在，则 $X = []$ 。格式 2 等价于 $X = \text{wpcoef}(T, D, 0)$ 。

14. wpfun 函数

wpfun 为小波包函数，调用格式为

格式 1 $[WPWS, X] = \text{wpfun}('wname', NUM, PREC)$

格式 2 $[WPWS, X] = \text{wpfun}('wname', NUM)$

wpfun 是一个小波包分析函数。格式 1 计算指定小波 'wname' 的小波包，且时间长度为二进制时间间隔(长度为 2^{prec} ，prec 必须是正整数)。输出矩阵 WPWS 包含从 0 到 NUM 的小波包函数，且按 $[W_0; W_1; \dots; W_{NUM}]$ 的顺序排列，输出向量 X 是普通的 X-网络向量。格式 2 等价于 $[WPWS, X] = \text{wpfun}('wname', NUM, 7)$ 。

当用正交小波时，产生小波包的方法是很简单的。首先，我们根据相应的小波，用两个长为 $2N$ 的滤波器($h(n)$)和($g(n)$)开始，它们分别是低通分解滤波器和高通滤波器各自除以 $\sqrt{2}$ 后的重构滤波器。

我们先定义下面的函数序列($W_n(x)$ ， $n=0, 1, 2, \dots$)

$$W_{2n} = 2 \sum_{k=0}^{2N-1} h(k) W_n(2x - k)$$

$$W_{2n+1} = 2 \sum_{k=0}^{2N-1} g(k) W_n(2x - k)$$

其中： $W_0(x) = \phi(x)$ 是尺度函数； $W_1(x) = \psi(x)$ 是小波函数。

例如，对 haar 小波尺度函数， $W_1(x) = \psi(x)$ 是小波函数，两个函数的支撑长度均在区间 $[0, 1]$ 上。从上式我们可以看出，可以通过把支撑区间分别在 $[0, 1/2]$ 和 $[1/2, 1]$ 内两个 $1/2$ 尺度的 W_n 加起来获得 W_{2n} 函数。同样，也可以通过把支撑区间分别在 $[0, 1/2]$ 和 $[1/2, 1]$ 内两个 $1/2$ 尺度的 W_n 进行加减来获得 W_{2n+1} 函数。

对于更规则的小波，用相似的构造方法可以获得光滑的小波包函数序列，且具有的支撑为 $[0, 2N-1]$ 。

15. wpcutree 函数

wpcutree 函数用于剪切小波包分解树，调用格式为

格式 1 $[T, D] = \text{wpcutree}(T, D, L)$

格式 2 $[T, D, RN] = wpcutree(T, D, L)$

`wpcutree` 是一个一维或二维小波包分析函数，它将小波包分解树第 L 层以下的所有二叉子树全部剪掉，并返回经过修改计算后的小波包分解结构 $[T, D]$ 。其中， L 是小波包分解树层数。

对于格式 1，它将剪切掉小波包分解第 L 层以下所有的二叉子树，并返回经过修改计算后的相应小波包分解结构 $[T, D]$ 。

对于格式 2，它除了返回格式 1 中的变量外，还返回初始小波包分解树经过重组的树结点索引序号 RN 。

16. `wp2wtree` 函数

`wp2wtree` 函数用于从小波包中提取小波树，调用格式为

$[T, D] = wp2wtree(T, D)$

`wp2wtree` 是一个一维或二维的小波包分析函数，它根据小波包分解树，修改计算树结构 T 和数据结构 D ，并返回小波树 $[T, D]$ 。

10.3 [示例 10-1] 小波数字滤波的实现

滤波是信号处理中最为重要的内容之一。由前述有关基于傅里叶变换的滤波器设计的章节可知，经典的滤波设计方法为按照截止频率和相应的信号衰减度等参数指标来设计满足要求的滤波器，设计步骤非常明确，常用的滤波器甚至不需要设计，直接通过查表即可获得滤波器参数，技术非常成熟。而通过小波分析来实现信号的滤波时，由于小波的种类多，灵活性强，滤波器的设计有别于常规的滤波器设计方法。下面通过一个实例来说明如何用小波分析来实现信号的滤波。

本例中，首先生成一个离散混合信号 $x(n)$ ，采样频率为 20000 Hz，信号包含一个低频正弦波，一个中频正弦波，还有高频随机噪声。其中低频正弦波频率为 250 Hz，相位为 0；中频正弦波频率为 1000 Hz，相位为 0.2 弧度。

要求：

- (1) 通过小波分析将低频 250 Hz 信号从 $x(n)$ 中提取出来；
- (2) 通过小波分析将中频 1000 Hz 信号从 $x(n)$ 中提取出来。

1. 工作原理

小波变换在信号消噪中的应用思路同傅里叶变换滤波的应用思路相似，只不过傅里叶变换的数字滤波是等步长频谱滤波，而小波变换消噪则是二等分频谱滤波，只有进行小波包分解才能实现等步长频谱滤波。由于变换的基波不一样，经典的滤波效果和小波消噪的效果也不一样。在小波消噪处理中，选用的小波不同，消噪效果也不一样。

应用小波分析进行消噪主要涉及到小波的分解与重构，下面以一维信号为例来介绍小波消噪的原理。

含有噪声的一维信号可以表示成如下的形式：

$$s(i) = f(i) + e(i), \quad i = 0, 1, 2, \dots, n-1$$

其中 $f(i)$ 为真实信号; $e(i)$ 为高斯白噪声, 噪声级为 1; $s(i)$ 为含噪声的信号。

对信号 $s(i)$ 进行消噪的目的就是要抑制信号中的噪声部分, 从而在 $s(i)$ 中恢复出真实信号 $f(i)$ 。在实际工程中, 有用信号通常表现为低频信号或是一些比较平稳的信号, 而噪声信号则通常表现为高频信号。一般来说, 一维信号的消噪算法可以分为三个步骤进行。

(1) 对信号进行小波分解。

选择一个小波并确定一个小波分解的层次 N , 然后对信号 s 进行 N 层小波分解。分解过程如图 10-4 和图 10-7(a) 所示, 图 10-6 是对应的频域段, 分解算法见式(10-36)。

信号处理与分析的实质是信号与不同频率基波的相关运算, 滤波也不例外。经典滤波器的设计是基于傅里叶变换的, 其基波是正弦波。而我们知道, 由于正弦波的正交特性, 只有相同频率的正弦波的相关函数值不全为 0, 因此, 信号经过经典滤波器滤波后, 只有通带范围内的频率分量得以保留。用小波分析来进行滤波也是一样, 唯一的区别在于把正弦波改成了所选的小波。小波种类多, 选择范围广, 一方面, 使得小波分析灵活性强, 但另一方面, 所选小波对最后的滤波结果带来直接的影响, 若选择不好, 其滤波效果就不会很理想。既然滤波的本质也是基于信号的相关运算, 那么所选的小波的波形自然越接近期望信号的波形就越好。对于波形平滑的期望信号, 应选择波形平滑的小波, 如 morlet 小波、db5 小波等。而对于波形变化本来就比较剧烈的信号, 如矩形波信号, 则应选择波形剧烈的小波, 如 Haar 小波。对于本例而言, 不妨选取 db5 小波。

对于经典滤波器来说, 所有信号分量的频率可以折算成相对频率, 采样频率对应于 2π , 离散信号必须满足采样定律, 因此最高频率为 π 。经典滤波器的截止频率可以选择为任意频率。离散小波分析目前还没有如此成熟的技术, 对于一般的小波分析, 其高频与低频的划分总是通过二分法进行的, 即对于某频率范围内的信号进行一次分解, 总是将信号分量等分为高频和低频两部分。因此, 其频率不可以设置为任意的频率。如果非要设置某一特定的截止频率, 则需要用到小波包分析。用小波分析进行滤波, 需要进行分解和重构, 至少需要两次卷积计算, 分解层次越多, 所要进行的运算量越大, 而经典滤波则只需要一次卷积运算, 因此小波分析的运算量相对较大。若用小波包分析进行滤波, 由于它将每次分解后的高频部分也进一步用二分法分解, 因此其计算量将进一步增大。另外, 小波分析中基波是所选小波, 因此, 小波分析中的频率是相对于所选小波而言的, 用小波对以采样频率为 f_s 获得的含有噪声的频率为 $f_s/4$ 的正弦波信号分解一次以后, 不能认为正弦波的信号依然完全在低频段, 这在后续的实例中可以看出。因此, 用小波分析来进行信号的滤波, 其分解层次与重构系数的选择需要通过试验来确定, 无法像经典滤波器设计那样按照一定的步骤来计算。在本例中, 分别选择不同分解层次与重构系数来进行滤波, 以对比滤波的效果。

(2) 系数的阈值量化处理。

信号在经过小波分解后, 虽然不同的分解系数对应于不同的频段, 从理论上来说, 用重构滤波器将某频率段的系数重构即可得到该频段的信号, 但在信噪比较高高的信号的小波分析中, 如果某一个小范围的频率分量分布在另一个频率段内, 则这种方法显然欠妥。此时, 可以通过对系数的阈值化处理来解决这样的问题。阈值化处理有软阈值化处理和硬阈值化处理两种。硬阈值化处理就是将其它频段中绝对值高于阈值的系数保留, 其它设为 0; 软阈值化处理就是将其它频段中的系数做柔性处理, 所有的系数乘上一个系数, 分解系

数绝对值越小，所乘的系数也越小，否则就越大，但不超过 1。关于阈值与分解系数所乘的系数的确定方法，参见 wden 函数的介绍。

小波分析的这一灵活性使得它对信噪比较高的信号可以同时实现低通滤波、带通滤波、带阻滤波，灵活性更强，甚至不需要知道进行低通滤波还是高通滤波，只需要关心（设定）阈值和阈值化处理的方法即可。

(3) 对信号进行重构。

用重构滤波器对小波分解后的某些层的系数进行小波的重构，即可以达到消噪的目的。重构的方式如图 10-5、图 10-7(b) 所示。小波分析的重构算法见式(10-37)。

2. 滤波结果与分析

虽然信号的滤波可以采用 wden 和 wdencmp 函数直接实现，但这两个函数往往只能进行低通滤波。为了了解清楚采用小波分析实现滤波的过程，本例采用最初级的小波分解和小波重构函数。整个滤波实例的 Matlab 程序代码如下，得到的结果如图 10-11 所示。

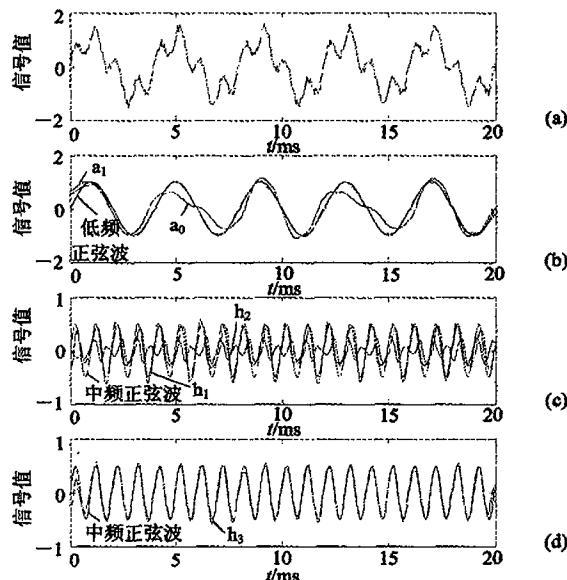


图 10-11 小波信号提取结果

- (a) 初始信号；(b) 低频正弦波及小波低频系数重构信号 a_0 、 a_1 波形；
- (c) 中频正弦波及小波高频系数重构信号 h_1 、 h_2 波形；
- (d) 中频正弦波及小波高频系数重构信号 h_3 波形

```
%生成模拟混合信号
t=0:0.05:20; %生成时间向量
x0=sin(0.5*pi*t); %生成低频正弦信号，幅值为 1
x1=0.5*sin(2*pi*t+0.2); %生成中频正弦信号，幅值为 0.5
ns=0.3*rand(1,length(t))-0.5;%生成高频随机噪声，幅值为 0.15
```

```

x=x0+x1+ns; %合成信号

%小波分析
[C, L]=wavedec(x, 5, 'db5'); %选择“db5”小波进行小波分解，分解层数为 5
a0 = wrcoef('a', C, L, 'db5', 5); %对第 5 层的低频系数进行重构
a1 = wrcoef('a', C, L, 'db5', 4); %对第 4 层的低频系数进行重构
h1 = wrcoef('d', C, L, 'db5', 4); %对第 4 层的高频系数进行重构
h2 = wrcoef('d', C, L, 'db5', 3); %对第 3 层的高频系数进行重构
h3= wrcoef('d', C, L, 'db5', 3)+wrcoef('d', C, L, 'db5', 4); %对第 3、4 层的高频系数进
行重构

%显示信号波形
figure;
subplot(4, 1, 1); %在 4 行 1 列的图的第一个图中显示初始信号
plot(t, x, 'k');
subplot(4, 1, 2); %在第二个图中用点划线显示低频正弦波，用实线显示
a0, 用虚线显示 a1
plot(t, x0, 'k', t, a0, 'b', t, a1, 'c');
subplot(4, 1, 3); %在第三个图中用点划线显示中频正弦波，用实线显示
h1, 用虚线显示 h2
plot(t, x1, 'k', t, h1, 'c', t, h2, 'b');
subplot(4, 1, 4); %在第 4 个图中用点划线显示低频正弦波，用实线显示 h3
plot(t, x1, 'k', t, h3, 'b');

```

由图 10-11(b)可以看出, a_1 的滤波效果比 a_0 好, a_0 的波形与低频正弦波有较大的差异。从经典滤波技术的角度来说, 采样频率为 20 000 Hz, 进行 5 次分解后, 低频段的频率范围应为 $0 \sim 20\ 000\ Hz / 2^5 = 625\ Hz$, 大于低频正弦波频率 250 Hz 的两倍, 因此, a_0 应该与低频正弦波整体重合, 噪声比 a_1 少。而实际上并非如此, 其原因就在于经典滤波技术中的单一频率在 db5 小波的频率体系中具有较宽的频率范围。由图 10-11(c)可以看出, h_1 整体波形与 x_1 一致, 但有较大偏差, h_2 则与 x_1 明显频率不同, 因此, 不论 h_1 还是 h_2 , 都不是 x_1 的理想逼近, 也就是说, 单纯重构第 4 层或第 3 层的高频系数, 作为实现提取 x_1 的小波分析带通滤波器不够理想。但从图 10-11(c)还可以发现, 当 h_1 的值比 x_1 的值小的时候, h_2 基本是正的, 而当 h_1 大于 x_1 的时候, h_2 大多情况下是负的, 因此, 将 h_1 和 h_2 相加, 或许是 x_1 的一个更好的逼近, 也就是说, 选择第 4 层和第 3 层的高频系数进行重构, 会是提取 x_1 的一个更好的小波带通滤波器。 h_1 和 h_2 相加得到的 h_3 的波形如图 10-11(d)所示。从图 10-11(d)可以看出, 结果的确如此, h_3 几乎与 x_1 重合, 5~15 ms 内, 两者最大偏差为 0.2208, 明显小于 h_1 与 x_1 之间的最大偏差 0.3768。从图 10-11(b)、(c)、(d)中可以看出, 重构信号两端与期望值都有较大差异, 这是小波分析的边界效应所致, 因此, 用小波分析进行滤波, 边界部分不可用。要得到较好的滤波效果, 输出值前后都必须有足够的数据, 这就使得用小波分析进行滤波和经典滤波一样, 都存在相位滞后, 不具有实时性。关于小波选择对滤波结果的影响, 读者不妨在上述源代码中将 db5 改为 db2, 然后运行程序, 将得到的结果与图 10-11 做一比较。

参 考 文 献

- [1] Li Jun, Liu Junhua. Identification of dynamic systems using support vector regression neural network[C]. 2006 LNCS3972: 934 – 942
- [2] Haifeng Ge, Junhua Liu. Identification of gas mixtures by a distributed support vector machine network and wavelet decomposition from temperature modulated semiconductor gas sensor[J]. Sensors and Actuators B, 2006, 117: 408 – 414
- [3] 葛海峰, 丁晖, 刘君华. Identification of CO、H₂、C₂H₂ and their mixtures by wavelet transformation and support vector machine[C]. The 6th International Symposium on Test and Measurement, 2005, 6
- [4] 葛海峰, 丁晖, 刘君华. Gas identification by wavelet transformation: based fast feature extraction and support machine from temperature on solid state sensors[C]. The 13th international conferences on solid state sensors, actuators and Microsystems, Seoul, Korea, June, 2005
- [5] Ding hui, Liu Junhua, Shen zhongru. Drift reduction of gas sensor by wavelet and principal component analysis[J]. Sensors and Actuators B, 2003, (96): 354 – 363
- [6] Lu Wenke, Zhu changchun, Liu Junhua. Implementing wavelet transform with SAW elements[J]. SCIENCE IN CHINA(Series E) , 2003, 46(6): 627 – 638
- [7] 张永怀, 白鹏, 周金林, 等. 小波分析与虚拟仪器在红外气体分析器中的应用[J]. 化工自动化及仪表, 2003, 30(3): 66 – 68
- [8] 冉启文, 谭立英. 小波分析与分数傅里叶变换及应用[M]. 北京: 国防工业出版社, 2002, 4
- [9] 张勇, 刘君华, 刘月明. 小波分析在光激光拾硅微谐振传感器微弱信号检测中的应用[J]. 仪器仪表学报, 2002, 23(4): 342 – 346
- [10] 丁晖, 刘君华, 申忠如. 小波技术在氧化锌避雷器泄露电流在线检测中的应用[J]. 西安交通大学学报, 2001, 6
- [11] 丁晖, 申忠如, 刘君华. 基于小波和相关分析的虚拟介损在线监测仪[J]. 高电压技术, 2000, 12
- [12] [美]崔锦泰. 小波分析导论[M]. 西安: 西安交通大学出版社, 1995, 1
- [13] 程正兴. 小波分析算法与应用[M], 西安: 西安交通大学出版社, 1998, 5
- [14] 胡昌华, 张军波, 夏军, 等. 基于 Matlab 的系统分析与设计——小波分析[J]. 西安: 西安电子科技大学出版社, 2000

第 11 章 线性相位滤波器与自适应滤波器

本章内容

- 线性相位滤波器基础知识
- 线性相位有限冲击响应滤波器的数学模型
- 自适应滤波器
- [示例 11-1] 用窗口法设计一个线性相位低通 FIR 滤波器
- [示例 11-2] 线性相位 FIR 滤波器在 Matlab 中的实现
- [示例 11-3] 自适应滤波在系统辨识中的应用

若滤波器不具有线性相位，信号中不同频率的分量虽然都落在滤波器通带内，滤波器对它们的衰减率相同，可以保证通过滤波器后各分量幅值的比例保持不变，但各频率分量滞后的时间不一致，以至于滤波前后各频率分量之间的相位差发生变化，滤波前后合成的信号波形不能保持一致，从而产生波形失真。

线性相位滤波器是这样一种滤波器：既保证信号中不同频率的分量都落在滤波器通带内，各分量幅值的比例滤波前后保持不变，而且滤波后各频率分量滞后的时间保持一致。

线性相位是滤波器所期望的。严格说来，在要求滤波后既消除了噪声，又能够保证信号波形不产生失真的精密测量场合，应该采用线性相位滤波器。

自适应滤波器能够自动调节当前时刻的滤波器参数在未知信号和未知噪声条件下，实现最优滤波。

线性相位滤波器与自适应滤波器在工程实践和信息处理技术中应用十分广泛。

11.1 线性相位滤波器

11.1.1 线性相位与线性相位滤波器

1. 线性相位图例

设滤波前初始信号 $y(t)$ 有三个频率分量 $y_1(t)$ 、 $y_2(t)$ 和 $y_3(t)$ ， $y(t) = y_1(t) + y_2(t) + y_3(t)$ ，三个分量的频率依次为 $\omega_1 = 2\pi f_1$ ， $\omega_2 = 2\pi f_2$ 和 $\omega_3 = 2\pi f_3$ ，且 $f_2 = 2f_1$ ， $f_3 = 3f_1$ 。于是，令 $\omega = \omega_1$ ，有

$$y(t) = \sin\omega_1 t + \sin\omega_2 t + \sin\omega_3 t = \sin\omega t + \sin 2\omega t + \sin 3\omega t$$

$y(t)$ 及其分量的波形如图 11-1(a) 所示。

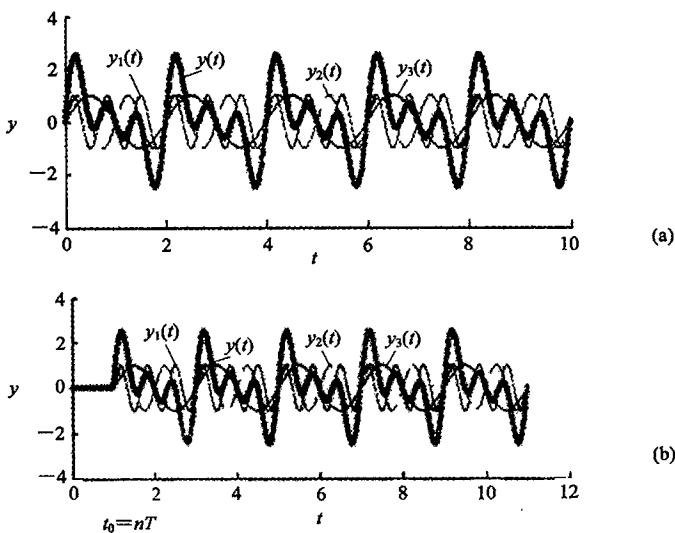


图 11-1 线性相位图示

(a) 滤波前的三个频率分量; (b) 滤波后的三个频率分量

经过滤波后, 基波分量 \$y_1(t) = \sin\omega t\$ 产生滞后时间 \$t_0\$, 折合成相位为

$$\varphi_1 = \omega_1 t_0 = \omega t_0 \quad (11-1)$$

对于二次谐波分量 \$y_2(t) = \sin 2\omega t\$ 与三次谐波分量 \$y_3(t) = \sin 3\omega t\$ 而言, 它们也必须滞后相同的时间 \$t_0\$, 才能保证合成的信号 \$y(t)\$ 的波形不会发生畸变, 如图 11-1(b)所示。同样的时间 \$t_0\$, 对于不同的频率 \$\omega_2\$ 和 \$\omega_3\$ 有不同的相位。

二次谐波 \$y_2(t)\$ 的相位滞后 \$\varphi_2\$ 为

$$\varphi_2 = \omega_2 t_0 = 2\omega t_0 \quad (11-2)$$

三次谐波 \$y_3(t)\$ 的相位滞后 \$\varphi_3\$ 为

$$\varphi_3 = \omega_3 t_0 = 3\omega t_0 \quad (11-3)$$

推论: 若将一信号以采样间隔 \$T_0\$ 进行采样, 共采样 \$N\$ 点, 经 FFT(快速傅里叶变换)得到 \$N\$ 条谱线, 其中有效谱线 \$N/2\$ 条, 序号 \$k\$ 为 \$0, 1, \dots, N/2-1\$; 则基波 \$k=1\$ 和 \$k \geq 2\$ 的谐波必须有线性相移

$$\varphi_k = \omega_3 t_0 = k\omega t_0 \quad \text{谐波次数 } k = 1, \dots, \frac{N}{2}-1 \quad (11-4)$$

于是, 滤波前后信号波形保持不变。

线性相位移的条件是: 各次分量波形移位延迟时间 \$t_0\$ 相同时, 各次分量波形的相位与其频率成线性改变。

2. 线性相位滤波器概念

线性滤波器对各次谐波分量产生相同时间 \$t_0\$ 的滞后, 各次谐波 \$k\$ 相移有如式(11-4)所示的与频率 \$\omega\$ 的线性比例关系。

11.1.2 线性相位有限冲击响应滤波器的数学模型

FIR 脉冲传递函数表达式及其对应差分方程已由 4.5 节式(4-67)和式(4-68)给出, 脉冲传递函数及对应差分方程重写如下:

$$H(z) = \frac{Y(z)}{X(z)} = C_0 + C_1 z^{-1} + \cdots + C_{N-1} z^{-(N-1)} = \sum_{n=0}^{N-1} C_n z^{-n} \quad (11-5a)$$

$$\begin{aligned} y(n) &= c_0 x(n) + c_1 x(n-1) + c_2 x(n-2) + \cdots + c_{N-1} x(n-N+1) \\ &= \sum_{r=0}^{N-1} c_r x(n-r) \end{aligned} \quad (11-5b)$$

根据 4.5 节内容可知, 由一个等效模拟滤波器 $H(s)$ 经采用不同的变换法, 如后向差分法、双线性变换法等, 就得到了 Z 域的脉冲传递函数 $H(z)$:

$$H(z) = \frac{Y(z)}{X(z)} = \frac{b_0 + b_1 z^{-1} + b_2 z^{-2} + \cdots}{a_0 + a_1 z^{-1} + a_2 z^{-2} + \cdots} \quad (11-6)$$

但求得的是无限冲击响应(IIR)滤波器的脉冲传递函数, 对式(11-6)进行多项式除法, 保留前 N 项, 即可获得如式(11-5)所示的 $N-1$ 阶 FIR 滤波器。

1. 一阶低通模拟滤波器生成的 FIR 滤波器

一阶低通模拟滤波器的 S 域传递函数 $H_d(s)$ 为

$$H_d(s) = \frac{a}{s+a} = \frac{100}{s+100} \quad (11-7)$$

式中 $a = \omega_{cd} = 100 \text{ rad/s}$, 为截止角频率, 相应的截止频率 $f_{cd} = \omega_{cd}/2\pi = 15.915 \text{ Hz}$ 。采用双线性变换法求与 $H_d(s)$ 等效的 Z 域脉冲传递函数 $H_d(z)$, 也就是把 $s = \frac{2}{T} \frac{1-z^{-1}}{1+z^{-1}}$ 代入到式(11-7)中, 得到:

$$\begin{aligned} H_d(z) &= H_d(s) \Big|_{s=\frac{2}{T} \frac{1-z^{-1}}{1+z^{-1}}} = \frac{a}{s+a} \Big|_{s=\frac{2}{T} \frac{1-z^{-1}}{1+z^{-1}}} \\ &= \frac{a}{\frac{2}{T} \frac{1-z^{-1}}{1+z^{-1}} + a} \end{aligned} \quad (11-8)$$

经整理可得

$$H_d(z) = \frac{aT + aTz^{-1}}{(2+aT) + (aT-2)z^{-1}} \quad (11-9)$$

对式(11-9)进行多项式直接相除, 其商的表达式即为 FIR 的一阶低通滤波器的脉冲传递函数

$$H_d(z) = c_0 + c_1 z^{-1} + c_2 z^{-2} + \cdots + c_{N-1} z^{-(N-1)} \quad (11-10)$$

式中 $c_0 = aT/(2+aT)$, $c_1 = 4aT/(2+aT)^2$, $c_2 = -4aT(aT-2)/(2+aT)^3$, ..., $c_{N-1} = -4aT(aT-2)^{N-2}/(2+aT)^N$ 。

当采样周期 $T=1.0 \text{ ms}$ 时, $aT=0.1$, 有

$$c_0 = \frac{0.1}{2.1}, c_1 = \frac{0.4}{2.1^2}, c_2 = \frac{(-0.4) \times (-1.9)}{2.1^3}, \dots, c_{N-1} = \frac{(-0.4) \times (-1.9)^{N-2}}{2.1^N}$$

把 $c_i (i=0, 1, \dots, N-1)$ 的值代入式(11-5)得:

$$\begin{aligned}
 y(n) &= c_0x(n) + c_1x(n-1) + \cdots + c_Nx(n-N+1) = \sum_{r=0}^{N-1}c_rx(n-r) \\
 &= \frac{1}{21}x(n) + \frac{0.4}{2 \cdot 1^2}x(n-1) + \frac{0.4 \times 1.9}{2 \cdot 1^3}x(n-2) + \cdots \\
 &\quad + \frac{-0.4 \times 1.9^{N-2}}{2 \cdot 1^N}x(n-N+1)
 \end{aligned}$$

2. 一阶高通模拟滤波器生成的 FIR 滤波器

一阶高通模拟滤波器的 S 域传递函数 $H_g(s)$ 为

$$H_g(s) = \frac{s}{s+a} = \frac{s}{s+100} \quad (11-11)$$

式中 $a=\omega_{cd}=100 \text{ rad/s}$, 为截止角频率, 相应的截止频率 $f_{cd}=\omega_{cd}/2\pi=15.915 \text{ Hz}$ 。采用双线性变换法, 也就是把 $s=\frac{2(1-z^{-1})}{T(1+z^{-1})}$ 代入到式(11-11)中, 得到与 $H_g(s)$ 等效的 Z 域脉冲传递函数 $H_g(z)$:

$$H_g(z) = \left. \frac{s}{s+a} \right|_{s=\frac{2(1-z^{-1})}{T(1+z^{-1})}} = \frac{2-az^{-1}}{(2+aT)-(2-aT)z^{-1}} \quad (11-12)$$

与一阶低通滤波器 FIR 系数的求取方法一样, 对式(11-12)进行多项式直接相除, 其商的表达式即为 FIR 一阶高通滤波器的脉冲传递函数, 其表达式的形式也与式(11-10)完全一样, 只是此时的 $c_i(i=0, 1, \dots, N-1)$ 分别为 $c_0=2/(2+aT)$, $c_1=-4aT/(2-aT)^2$, $c_2=-4aT(2-aT)/(2+aT)^3$, ..., $c_{N-1}=-4aT(2-aT)^{N-2}/(2+aT)^N$ 。当采样周期为 $T=1.0 \text{ ms}$ 时, $aT=0.1$, 一阶高通 FIR 滤波器的系数分别为

$$c_0 = \frac{2}{2.1}, c_1 = \frac{-0.4}{2 \cdot 1^2}, c_2 = \frac{-0.4 \times 1.9}{2 \cdot 1^3}, \dots, c_{N-1} = \frac{-0.4 \times 1.9^{N-2}}{2 \cdot 1^N}$$

3. 有限冲激响应滤波器(FIR)的频率特性 $H(\omega)$

由 Z 变换知识 $z=e^{j\omega T}$, 将其代入式(11-5a)可得

$$H(\omega) = H(e^{j\omega T}) = C_0 + C_1 e^{-j\omega T} + \cdots + C_{N-1} e^{-j(N-1)\omega T} = \sum_{n=0}^{N-1} C_n e^{-jn\omega T} \quad (11-13)$$

而时间序列 $h(n)$ 的离散傅里叶变换式为

$$H(\omega) = \sum_{n=-\infty}^{+\infty} h(n) e^{-jn\omega T} \quad (11-14)$$

式(11-14)与式(11-13)形式相同, 故式(11-14)中的傅里叶系数 $h(n)$ 就是 FIR 系统的滤波系数 C_n , 即

$$h(0) = C_0, h(1) = C_1, \dots, h(N-1) = C_{N-1}; N \rightarrow \infty$$

4. 生成线性相位 FIR 滤波器需要考虑的问题

1) 无限长时间序列的不可实现与有限截取

由式(11-5b)差分方程可见, 当前时刻(nT)的输出 $y(n)$ 是由当前时刻的输入值 $x(n)$ 乘以系数 C_0 , 以及无限长历史时刻的输入值 $x(n-1), x(n-2), \dots, x(n-N+1)$ 乘以相应的系数组成的无限长序列。实际上, 若 N 趋于无穷大, 这是不可能实现的, 故必须对输入 $x(nT)$ 进行有限截取, 截取长度 N 应为有限值。

2) 具有线性相位 FIR 滤波器的形式

当 FIR 系统的系数 $C_n(h(n))$ 对称时, 滤波器将具有线性相位。对称有偶对称和奇对称两种情况。对于偶对称, 即 FIR 滤波器系数 h 满足关系式 $h(N-1-n)=h(n)$; 而对于奇对称, 即 FIR 滤波器系数 h 满足关系式 $h(N-1-n)=-h(n)$ 。另外, FIR 系统系数的长度有奇数个和偶数个之分, 因此需要分为四种情况来论证。

(1) 奇数长度偶对称。奇数长度偶对称的滤波器有奇数个参数, 可以分成三部分, 前 $(N-3)/2$ 个参数为一部分, 后 $(N-3)/2$ 个参数为一部分, 中间的参数 $h((N-1)/2)$ 为一部分。于是, 该滤波器的离散傅里叶变换可写为

$$H(j\omega) = \sum_{n=0}^{N-1} h(n)e^{-j\omega n} = \sum_{n=0}^{(N-3)/2} h(n)e^{-j\omega n} + h((N-1)/2)e^{-j\omega(N-1)/2} + \sum_{n=(N+1)/2}^{N-1} h(n)e^{-j\omega n} \quad (11-15)$$

令 $n=m+(N-1)/2$, 则 $h(n)e^{-j\omega n}=h((N-1)/2+m)e^{-j\omega[m+(N-1)/2]}$, 于是

$$\sum_{n=(N+1)/2}^{N-1} h(n)e^{-j\omega n} = \sum_{m=1}^{(N-1)/2} h((N-1)/2+m)e^{-j\omega[m+(N-1)/2]}$$

把 m 改为 n , 即得到 $\sum_{n=1}^{(N-1)/2} h((N-1)/2+n)e^{-j\omega[n+(N-1)/2]}$ 。对于一个类似 $1+2+\cdots+100$ 的求和计算, 我们知道, 它等同于 $100+99+\cdots+1$ 的求和计算。因此, 对于 $\sum_{n=0}^{(N-3)/2} h(n)e^{-j\omega n}$, 有:

$$\sum_{n=0}^{(N-3)/2} h(n)e^{-j\omega n} = \sum_{n=1}^{(N-1)/2} h((N-1)/2-n)e^{-j\omega[(N-1)/2-n]}$$

再考虑到滤波器是偶对称的, $h(N-1-n)=h(n)$, 因此, 把上式中的 $h((N-1)/2-n)$ 替换为 $h(N-1-[n-(N-1)/2])=h((N-1)/2+n)$, 得到:

$$\sum_{n=1}^{(N-1)/2} h((N-1)/2-n)e^{-j\omega[(N-1)/2-n]} = \sum_{n=1}^{(N-1)/2} h((N-1)/2+n)e^{-j\omega[(N-1)/2-n]}$$

于是, 式(11-15)可以重写为

$$H(j\omega) = \sum_{n=1}^{(N-1)/2} h((N-1)/2+n)e^{-j\omega[(N-1)/2-n]} + h((N-1)/2)e^{-j\omega(N-1)/2} + \sum_{n=1}^{(N-1)/2} h((N-1)/2+n)e^{-j\omega[n+(N-1)/2]}$$

从上式右边三项中提取 $e^{-j\omega(N-1)/2}$ 得到:

$$H(j\omega) = \sum_{n=0}^{N-1} h(n)e^{-j\omega n} = e^{-j\omega(N-1)/2} \left[\sum_{n=1}^{(N-1)/2} h((N-1)/2+n)(e^{-j\omega n} + e^{j\omega n}) + h((N-1)/2) \right] \quad (11-16)$$

考虑到 $e^{j\omega n} + e^{-j\omega n} = 2 \cos(\omega n)$, 式(11-16)可以写作

$$H(j\omega) = e^{-j\omega(N-1)/2} \left[\sum_{n=1}^{(N-1)/2} 2h((N-1)/2+n)\cos(n\omega) + h((N-1)/2) \right] \quad (11-17)$$

由式(11-17)中可以看出, $H(j\omega)$ 的幅值为 $\sum_{n=1}^{(N-1)/2} 2h\left(\frac{N-1}{2}+n\right)\cos(n\omega) + h\left(\frac{N-1}{2}\right)$, 对于给定的 ω , 它是一个确定的数值。 $H(j\omega)$ 的相位 $\theta = -\omega(N-1)/2$, 负号的存在表明相位

是滞后的。当 y 和 x 满足形如 $y=ax+b$ 的关系式时，我们称 y 和 x 是线性的，其中 x 为自变量， y 为因变量， a 为比例系数， b 为截长。明显地， $H(j\omega)$ 的相位 θ 与频率 ω 满足此条件，自变量为 ω ，因变量为 θ ，比例系数为 $-(N-1)/2$ ，截长为 0，因此，相位与频率满足线性关系，也就是说 $H(j\omega)$ 具有线性相位。

在采用形如式(11-17)的滤波器滤波后的信号中，所有的频率分量的滞后时间为 $\Delta t=\theta/\omega=(N-1)/2$ 。注意到离散傅里叶变换中，所选时间窗长度为 $N-1$ ，忽略采样时间参数 T_s ，实为 $(N-1)T_s$ ，因此，经滤波后的信号中，各分量的滞后时间均为 $T_s(N-1)/2$ 。对于选定长度的滤波器，它是一个常数，经滤波后的信号中，各频率分量合成的信号等同于原信号的通带范围内各频率分量分别滞后 $T_s(N-1)/2$ 后重新合成的信号，也等同于原信号的通带范围内各频率分量的合成信号整体滞后 $T_s(N-1)/2$ ，因此不会发生畸变。当然，由于滤波器不是理想的，过渡带中的频率分量没有完全衰减到 0，通带内接近截止频率分量也有一定程度的衰减，就算用线性相位滤波器滤波得到的信号，依然有一定程度的畸变。但相对于非线性相位滤波器来说，其畸变程度要小得多。

(2) 偶数长度偶对称。偶数长度的滤波器系数可分成前半部分和后半部分两部分，因此其傅里叶变换可写作

$$H(j\omega) = \sum_{n=0}^{N-1} h(n)e^{-j\omega n} = \sum_{n=0}^{N/2-1} h(n)e^{-j\omega n} + \sum_{n=N/2}^{N-1} h(n)e^{-j\omega n} \quad (11-18)$$

把 $\sum_{n=N/2}^{N-1} h(n)e^{-j\omega n}$ 写作 $\sum_{n=1}^{(N-1)/2} h(N/2-1+n)e^{-j\omega[n+N/2-1]}$ ，考虑到 $h(N-1-n)=h(n)$ ，可以把 $\sum_{n=0}^{N/2-1} h(n)e^{-j\omega n}$ 写作 $\sum_{n=1}^{N/2-1} h(N/2-1-n)e^{-j\omega[N/2-1-n]}$ 。于是，从式(11-18)右边两项中提取 $e^{-j\omega(N/2-1)}$ 得到：

$$H(j\omega) = \sum_{n=0}^{N-1} h(n)e^{-j\omega n} = e^{-j\omega(N/2-1)} \left[\sum_{n=1}^{N/2} h(N/2-1+n)(e^{-j\omega n} + e^{j\omega n}) \right] \quad (11-19)$$

同样考虑到 $e^{j\omega n} + e^{-j\omega n} = 2 \cos(\omega n)$ ，于是可知 $H(j\omega)$ 的滞后相位为 $-\omega(N/2-1)$ ，它也是与 ω 满足线性关系的。

(3) 奇数长度奇对称。有

$$H(j\omega) = \sum_{n=0}^{N-1} h(n)e^{-j\omega n} = \sum_{n=0}^{(N-3)/2} h(n)e^{-j\omega n} + h((N-1)/2)e^{-j\omega(N-1)/2} + \sum_{n=(N+1)/2}^{N-1} h(n)e^{-j\omega n} \quad (11-20)$$

把 $\sum_{n=(N+1)/2}^{N-1} h(n)e^{-j\omega n}$ 写作 $\sum_{n=1}^{(N-1)/2} h((N-1)/2+n)e^{-j\omega[n+(N-1)/2]}$ ，考虑到 $h(N-1-n)=-h(n)$ ，可以把 $\sum_{n=0}^{(N-3)/2} h(n)e^{-j\omega n}$ 写作 $\sum_{n=1}^{(N-1)/2} -h((N-1)/2+n)e^{j\omega(N-1)/2+n}$ 。于是，从式(11-20)右边三项中提取 $e^{-j\omega(N-1)/2}$ 得到：

$$H(j\omega) = \sum_{n=0}^{N-1} h(n)e^{-j\omega n} = e^{-j\omega(N-1)/2} \left[\sum_{n=1}^{(N-1)/2} h((N-1)/2+n)(e^{-j\omega n} - e^{j\omega n}) + h((N-1)/2) \right] \quad (11-21)$$

考虑到 $e^{j\omega n} - e^{-j\omega n} = j2 \sin(\omega n)$ ，上式可以写作

$$H(j\omega) = e^{-j\omega(N-1)/2-\pi/2} \left[\sum_{n=1}^{(N-1)/2} 2h((N-1)/2+n) \sin(n\omega) + h((N-1)/2) \right] \quad (11-22)$$

明显地, $H(j\omega)$ 的滞后相位为 $\omega(N-1)/2+\pi/2$, 它也是线性的。

(4) 偶数长度奇对称。有

$$H(j\omega) = \sum_{n=0}^{N-1} h(n)e^{-j\omega n} = \sum_{n=0}^{N/2-1} h(n)e^{-j\omega n} + \sum_{n=N/2}^{N-1} h(n)e^{-j\omega n} \quad (11-23)$$

同奇数长度奇对称情况一样, 把 $\sum_{n=N/2}^{N-1} h(n)e^{-j\omega n}$ 写作 $\sum_{n=1}^{(N-1)/2} h(N/2-1+n)e^{-j\omega[n+N/2-1]}$, 考虑到 $h(N-1-n) = -h(n)$, 可以把 $\sum_{n=0}^{N/2-1} h(n)e^{-j\omega n}$ 写作 $\sum_{n=1}^{N/2-1} -h(N/2-1-n)e^{-j\omega[N/2-1-n]}$ 。于是, 从式(11-23)右边两项中提取 $e^{-j\omega(N/2-1)}$ 得到:

$$H(j\omega) = \sum_{n=0}^{N-1} h(n)e^{-j\omega n} = e^{-j\omega(N/2-1)-\pi/2} \left[\sum_{n=1}^{N/2} h(N/2-1+n)(e^{-j\omega n} - e^{j\omega n}) \right] \quad (11-24)$$

明显地, $H(j\omega)$ 的滞后相位为 $\omega(N/2-1)+\pi/2$, 它同样是线性的。

11.1.3 线性相位 FIR 滤波器的窗口设计

1. 线性相位 FIR 滤波器窗口设计的基本思路

线性相位 FIR 滤波器窗口设计的基本思路是: 瞄准目标是理想频率特性, 可以是理想低通、高通、带通或带阻, 根据需要而定。

第一步, 由傅里叶反变换求其频率特性 $H_d(\omega)$ 的离散时间序列 $h_d(n)$ 。

第二步, 选择窗口的大小, 也就是 FIR 滤波器系数的个数 N , 并截取 $h_d(n)$ 的以 $n=0$ 为中心的中间 N 项, 将时间序列右移延时 $\text{floor}(N/2)$ 。函数 $\text{floor}(x)$ 表示不大于 x 的整数。

第三步, 选择窗函数, 并求取长度为 N 的窗函数系数。将 $h_d(n)$ 的中间 N 项与窗函数系数对应项相乘, 即得到 FIR 滤波器的系数 $h(n)$ 。

第四步, 求 $h(n)$ 的频谱 $H(\omega)$, 该频谱与期望的理想频率特性有差异, 差异与截取的长度和所选择的窗函数有关。

第五步, 根据采用窗函数的特点分析获得的频谱过渡带斜率及截止角频率 ω_c , 若不满足设计要求, 则调整窗函数重复第二步至第四步, 直至满足要求。

上述窗口设计法又称傅里叶级数法。

2. 线性相位低通滤波器的设计

期望的理想低通滤波器频率特性 $H_d(\omega)$ 为

$$H_d(j\omega) = \begin{cases} 0 & |\omega| > \omega_c \\ 1 & |\omega| \leq \omega_c \end{cases} \quad (\omega_c \text{ 为角截止频率}) \quad (11-25)$$

(1) 对 $H_d(\omega)$ 进行傅里叶反变换, 以获取时间序列 $h_d(n)$ 。离散傅里叶反变换的形式为

$$h_d(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} H_d(e^{j\omega}) e^{j\omega n} d\omega \quad (11-26)$$

将频域扩展到 $-\omega_c$ ，式(11-25)的值代入式(11-26)可得：

$$\begin{aligned} h_d(n) &= \frac{1}{2\pi} \int_{-\omega_c}^{\omega_c} e^{j\omega n} d\omega = \frac{1}{j2\pi n} \int_{-\omega_c}^{\omega_c} e^{j\omega n} d(j\omega n) = \frac{1}{j2\pi n} e^{j\omega n} \Big|_{-\omega_c}^{\omega_c} = \frac{\sin(\omega_c n)}{\pi n} \\ &= \frac{\omega_c}{\pi} \frac{\sin(\omega_c n)}{\omega_c n} = \frac{1}{T_c} \frac{\sin(\omega_c n)}{\omega_c n} = \frac{1}{T_c} \text{sinc}(\omega_c n) \end{aligned} \quad (11-27)$$

式中 $\text{sinc}(n\omega_c) = \sin(n\omega_c)/n\omega_c$ ，称之为 sinc 函数，它是以 2π 为周期的。当 $0 \leq n < \pi/\omega_c$ 时，sinc 函数大于 0；当 $n = \pi/\omega_c$ 时，sinc 函数第一次经过 0 点；当 $2\pi/\omega_c > n > \pi/\omega_c$ 时，sinc 函数小于 0。由于随着 n 的增大，sinc 函数的分母成比例增大，于是 sinc 函数随着 n 的增加而作衰减振荡。另外，对于给定的截止频率 ω_c ， T_c 是一个常数，因此， $h_d(n)$ 是以 $n=0$ 为中心对称，其幅值呈 sinc 函数形式衰减的无限长时间序列，其形状如图 11-2(b)所示。

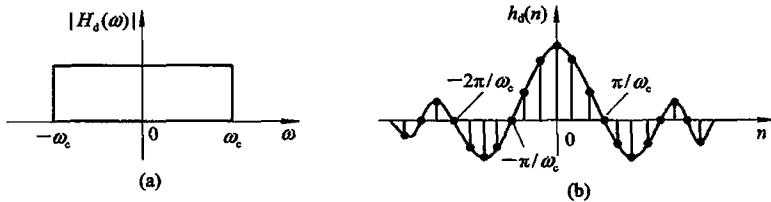


图 11-2 理想矩形幅频特性及无限长序列的插入

(a) 理想矩形幅频特性 $|H_d(\omega)|$ ；(b) $|H_d(\omega)|$ 对应放入无限长时间序列 $h_d(n)$

(2) 将无限长序列 $h_d(n)$ 有限截短，并向右移延时。如图 11-2(b)所示，已知 $h_d(n)$ 是以 $n=0$ 为中心对称的无限长序列，因此物理实现是不可能的。为了让 FIR 滤波成为可能，需要从 $h_d(n)$ 中提取一段有限个系数来近似 $h_d(n)$ ，这个过程称之为截短。从图 11-2(b)还可以看出， $n=0$ 时， $h_d(n)$ 的绝对值最大，然后向 n 的正负方向振荡衰减，直至趋于 0，因此，截短一般是截取以 $n=0$ 为中心的一段长度为 N 的系数 $h_N(n)$ 。由于 $h_d(n)$ 是关于 $h_d(0)$ 偶对称的， N 一般取为正的奇数。另外，负时间($-n$)的序列 $h(-n)$ 与正时间(n)的序列 $h(n)$ 以 $n=0$ 为中心对称，但负时间序列是真实物理世界不存在的，且又不可见，这称之为非因果的，也是不可实现的，因此需要对截短后的序列 $h_N(n)$ 向右移动 $(N-1)/2$ ，也就是将原来的序列 $h_N[(N-1)/2], \dots, h_N(-1), h_N(0), h_N(1), \dots, h_N[(N-1)/2]$ 变成 $h_N(0), h_N(1), \dots, h_N(N-1)$ 。

(3) 加窗。 $h_d(n)$ 的有限截短实际上是对 $h_d(n)$ 加了一个矩形窗，在这个窗之内的系数可以全部看到，在这个窗之外的系数全部被置为 0，这等效于给 FIR 滤波器系数另外再乘了一个系数，只是在窗内的系数为 1，窗外的系数为 0。在数学上，这个过程可以写作

$$h(n) = h_N\left(n - \frac{N-1}{2}\right) \cdot R_N(n) \quad (11-28)$$

式中 $R_N(n)$ 为矩形窗函数，其函数式为

$$R_N(n) = \begin{cases} 1, & N-1 \geq n \geq 0 \\ 0, & \text{其它} \end{cases} \quad (11-29)$$

除了矩形窗之外，为了获得不同指标的滤波器，有关研究工作者对矩形窗做了改进，窗内的系数并非全为 1，而是从中心点开始，滤波器系数向左、向右逐步衰减，从而获得了其它形式的窗函数。对于选用矩形窗来说，加窗的过程实质上在截短的过程中已经完成。

若选用的是其它窗函数，则首先要求取窗函数的系数，然后代入到式(11-28)中，才能求得滤波器系数 $h(n)$ 。不同形式的窗函数有各自的优缺点，在实际应用中需要根据设计指标选择合适的窗函数。有关窗函数的内容将在下一小节介绍。

(4) 时间序列 $h(n)$ 的频率 $H(\omega)$ 。根据傅里叶变换中频域卷积特性可知，时域两个函数相乘，其积的频谱是各自频谱函数的卷积，因此，经过将无限长时间序列 $h_d[(N-1)/2]$ 与窗函数相乘后，被截短为有限长时间序列 $h(n)$ ，其频谱 $H(\omega)$ 可由频域卷积定理来求。下面分别求时域两函数各自的频谱函数，再求它们频谱函数的卷积。

① 无限长时间序列 $h_d\left(n-\frac{N-1}{2}\right)$ 的频谱函数。根据傅里叶变换中的时移特性，将信号在时域中沿时间轴平移一个常值而成为时延信号，在频域其幅值不变，而相频谱中相位角的改变与频率成正比。已知未延时时间序列的频谱为 $H_d(\omega)$ 时，时延序列 $h_d\left(n-\frac{N-1}{2}\right)$ 的频谱为

$$H_d(\omega) = |H_d(\omega)| e^{-j\varphi'(\omega)} \quad (11-30)$$

其中 $\varphi'(\omega) = \varphi(\omega) + \omega(N-1)/2$ ，相位角的改变量为 $\Delta\varphi = \varphi'(\omega) - \varphi(\omega) = \omega(N-1)/2$ ，当 N 为常数时，相位角的改变与频率 ω 成正比。当初始相位角 $\varphi(\omega) = 0$ 时有

$$h_d(n) \leftrightarrow H_d(\omega) = |H_d(\omega)| e^{-j\varphi(\omega)} = |H_d(\omega)| \quad (11-31a)$$

$$h_d[n-(N-1)/2] \leftrightarrow H_d'(\omega) = |H_d(\omega)| e^{-j\varphi'(\omega)} = |H_d(\omega)| e^{-j\omega(N-1)/2} \quad (11-31b)$$

特别地，理想低通滤波器的频谱函数具有式(11-27)所示的形式。

② 窗函数的频谱。已知未时延窗函数时间序列

$$w_N(n) \leftrightarrow |W_N(\omega)| e^{-j\varphi(\omega)} = |W_N(\omega)|$$

即 $\varphi(\omega) = 0$ 。经时延 $(N-1)/2$ 后有

$$W_N[n-(N-1)/2] \leftrightarrow W_N'(\omega) = |W_N(\omega)| e^{-j\varphi'(\omega)} = |W_N(\omega)| e^{-j\omega(N-1)/2} \quad (11-32)$$

按傅里叶正变换定义式可求 $W_N(\omega)$ 。特别地，对于矩形窗，时延 $(N-1)/2$ 的频谱为

$$W_R(\omega) = \sum_{n=-\infty}^{+\infty} R_N(n) e^{-j\omega n} = \sum_{n=0}^{N-1} e^{-j\omega n} = \frac{1 - e^{-j\omega N}}{1 - e^{-j\omega}} = \frac{\sin(\omega N/2)}{\sin(\omega/2)} e^{-j\omega \frac{N-1}{2}} \quad (11-33)$$

矩形窗函数及其幅值谱如图 11-3 所示。

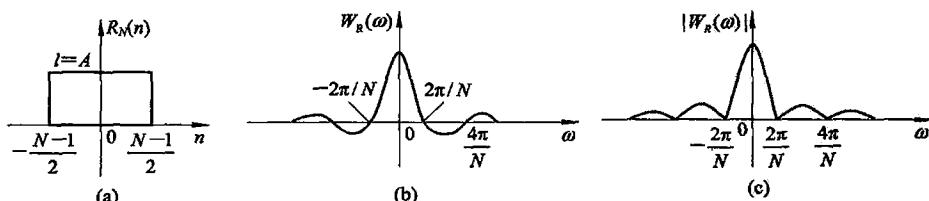


图 11-3 矩形窗函数及幅值谱

(a) 矩形窗；(b) 矩形窗频谱；(c) 矩形窗幅值谱

③ 求两时间序列 $h_d(n)$ 与 $w_N(n)$ 之乘积 $h(n)$ 的频谱。根据傅里叶变换中的频域卷积特

性, 将该两时间序列的频谱进行卷积。已知式(11-31)和式(11-32)成立, 则乘积

$$h(n) = h_d\left(n - \frac{N-1}{2}\right)w_N\left(n - \frac{N-1}{2}\right) \leftrightarrow H(\omega) = \frac{1}{2\pi}H'_d(\omega) * W_R(\omega) \quad (11-34)$$

将式(11-31)和式(11-32)代入式(11-34), 由卷积定义有

$$\begin{aligned} H(\omega) &= \frac{1}{2\pi} \int_{-\pi}^{\pi} H'_d(\omega) W'_R(\omega - \theta) d\theta \\ &= \frac{1}{2\pi} \int_{-\omega_c}^{\omega_c} |H_d(\omega)| e^{-j\omega\frac{N-1}{2}} |W_R(\omega)| e^{-j(\omega-\theta)\frac{N-1}{2}} d\theta \end{aligned} \quad (11-35a)$$

特别地, 根据式(11-25)可知, 在 $-\omega_c$ 至 ω_c 区间, $H_d(\omega) = 1$, 将式(11-31a)和式(11-33)代入式(11-35a)可得窗函数为矩形窗的低通滤波器系数 $h(n)$ 的频谱为

$$\begin{aligned} H(\omega) &= \frac{1}{2\pi} \int_{-\omega_c}^{\omega_c} e^{-j\omega\frac{N-1}{2}} e^{-j\omega\frac{N-1}{2}(\omega-\theta)} \frac{\sin[(\omega-\theta)N/2]}{\sin[(\omega-\theta)/2]} d\theta \\ &= e^{-j\omega\frac{N-1}{2}} \frac{1}{2\pi} \int_{-\omega_c}^{\omega_c} W_R(\omega-\theta) d\theta \end{aligned} \quad (11-35b)$$

式(11-35b)卷积的结果如图 11-4 所示, 图中曲线表示离散时刻幅值的包络。

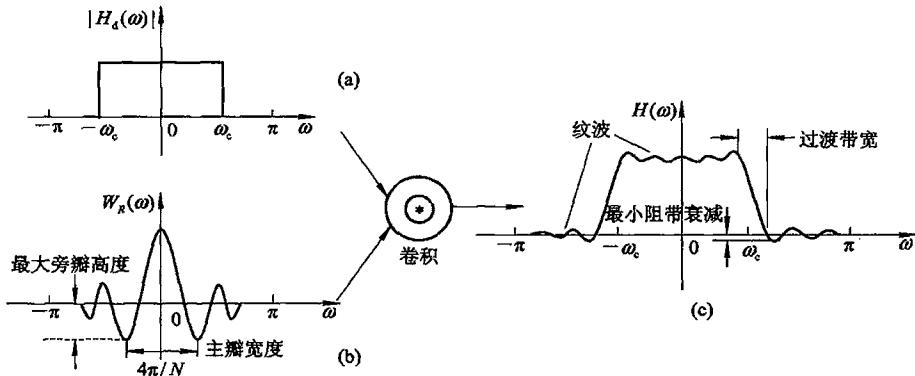


图 11-4 卷积结果图

- (a) $H_d(\omega)$ ——无限长时间序列 $h_d(n)$ 的频谱;
- (b) $W_R(\omega)$ ——矩形窗的频谱;
- (c) $H(\omega)$ ——有限长时间序列 $h(n)$, $n=0, 1, \dots, N-1$ 的频谱

至此, 我们得到了一个物理真实系统可实现的, 由 N 个偶对称脉冲序列 $h(0), h(1), \dots, h(N-1)$ 体现的线性相移 FIR 滤波器, 该滤波器的频率域特性为 $H(\omega)$ 。如图 11-4(c)所示的低通滤波器特性与原期望的如图 11-4(a)所示的理想频率特性 $H_d(\omega)$ 存在差异, 其原因是 $H_d(\omega)$ 是无限长时间序列 $h_d(n)$ 的频谱, 差异程度由截短采用的窗口函数 $w_N(n)$ 所决定。

(5) 有限长时间序列 $h(n)$ 频谱 $H(\omega)$ 的相关参数。由图 11-4 可看出, 加矩形窗后使实际频率响应偏离理想频率响应, 主要影响有三个方面: 首先是理想幅频特性陡直边缘处形成过渡带, 过渡带宽取决于矩形窗函数频率响应的主瓣宽度; 其次是过渡带两侧形成肩峰和波纹, 这是矩形窗函数频率响应的旁瓣引起的, 旁瓣相对值越大, 旁瓣越多, 波纹越多; 第三是随窗函数宽度 N 的增大, 矩形窗函数频率响应的主瓣宽度减小, 但不改变旁瓣的相对值。

为了改善 FIR 滤波器性能, 要求窗函数的主瓣宽度尽可能窄, 以获得较窄的过渡带; 旁瓣相对值尽可能小, 数量尽可能少, 以获得通带波纹小, 阻带衰减大, 在通带和阻带内均平稳的特点, 这样可使滤波器实际频率响应更好地逼近理想频率响应。

以矩形窗为例, 从图 11-4(c)所示的 $H(\omega)$ 可以看出, 当窗函数宽度为 N 时, 加窗截取后的频谱图存在过渡带及泄漏(即截止频率之外的幅频特性不为 0)。

① 过渡带宽定义为

$$\Delta\omega = \omega_s - \omega_p$$

式中 ω_p 为通带截止频率, ω_s 为阻带允许最小衰减频率。

② 相对衰耗定义为

$$A(\omega) = 20 \lg[|H(e^{j\omega})| / |H(e^{j0})|] = 20 \lg[H(\omega)/H(0)] \quad (11-36)$$

滤波器的几乎所有重要指标都是由窗函数决定的, 改进滤波器的关键在于改进窗函数。为了便于研究, 我们把窗谱 $W(\omega)$ 也化为相对衰耗 $20 \lg[W(e^{j\omega})/W(e^{j0})]$ 的形式。

窗谱的两个最重要的指标是主瓣宽度和旁瓣峰值衰耗。旁瓣峰值衰耗定义为

$$\text{旁瓣峰值衰耗} = 20 \lg(\text{第一旁瓣峰值} / \text{主瓣峰值}) \quad (11-37)$$

对于矩形窗而言, 旁瓣峰值衰耗为 -13 dB。

旁瓣峰值衰耗与阻带最小衰耗有联系, 但不是同一个概念。旁瓣峰值衰耗适用于窗函数, 它是窗谱主副瓣幅度之比, 而阻带最小衰耗适用于滤波器, 当滤波器是用窗口法得出时, 阻带最小衰耗取决于窗谱主副瓣面积之比。

为了改善 FIR 滤波器性能, 必须修改窗函数, 使其具有更好的窗谱。一个好的窗谱, 应满足以下两方面的条件:

- 主瓣尽可能窄, 以使设计出来的滤波器有较陡的过渡带;
- 第一副瓣面积相对主瓣面积尽可能小, 即能量尽可能集中在主瓣, 外泄少, 这样设计出来的滤波器才能尖峰和余振小。

对任一个具体的窗函数而言, 以上两个条件互相矛盾, 不能同时满足。我们所能做的是根据具体设计指标, 选择一种能兼顾各项指标的相对最佳的窗口。

3. 几种常用的窗口函数简介

(1) 矩形窗:

$$w(n) = R_N(n) \quad (11-38)$$

矩形窗的特点是旁瓣较大, 尤其是第一个旁瓣峰太高, 达到主瓣高度的 21% , 所以泄漏很大。矩形窗的优点是容易获得, 而且主瓣宽度小, 其等效带宽为 $1/T$ 。

(2) 三角窗:

$$w(n) = \begin{cases} \frac{2n}{N-1}, & 0 \leq n \leq \frac{N-1}{2} \\ 2 - \frac{2n}{N-1}, & \frac{N-1}{2} \leq n \leq N-1 \end{cases} \quad (11-39)$$

它是由两个长度为 $N/2$ 的矩形窗进行线性卷积而得到的。

(3) 汉宁(Hanning)窗, 又称为余弦窗:

$$w(n) = \frac{1}{2} \left[1 - \cos \frac{2n\pi}{N-1} \right], \quad 0 \leq n \leq N-1 \quad (11-40)$$

汉宁窗的主要思路是：通过矩形窗谱的合理叠加减小旁瓣面积。因此可写出

$$w(n) = \frac{1}{2}R_N(n) - \frac{1}{2} \cdot \frac{1}{2} [e^{j\frac{2\pi}{N-1}} + e^{-j\frac{2\pi}{N-1}}]R_N(n) \quad (11-41)$$

其窗谱

$$W(e^{j\omega}) = 0.5W_R(e^{j\omega}) - 0.25W_R(e^{j\omega-\frac{2\pi}{N-1}}) - 0.25W_R(e^{j\omega+\frac{2\pi}{N-1}}) \quad (11-42)$$

式中 $W_R(e^{j\omega})$ 为矩形窗谱。当 N 较大时, $\frac{2\pi}{N-1} \approx \frac{2\pi}{N}$, 于是 $W(e^{j\omega})$ 可看成是三个不同位置矩形窗谱的叠加。叠加结果付出的代价是主瓣增宽一倍, 得到的好处是旁瓣峰值衰耗由 -13 dB 增加到 -31 dB。汉宁窗的旁瓣峰值较小, 衰减较快, 主瓣宽度为 $1.5/T$, 比矩形窗的主瓣宽, 但总泄漏比矩形窗小得多。由于汉宁窗比较容易获得, 因此是经常使用的一种时间窗。

(4) 海明(Hamming)窗：

$$w(n) = 0.54 - 0.46 \cos \frac{2n\pi}{N-1}, \quad 0 \leq n \leq N-1 \quad (11-43)$$

式(11-43)可转化为

$$w(n) = 0.08 + 0.92 \left[\frac{1}{2} \left(1 - \cos \frac{2n\pi}{N} \right) \right]$$

由此式可知, 海明窗是在矩形窗上拼接一个汉宁窗而形成的, 它包括一个高为 0.08 的矩形窗和一个最大高度为 0.92 的汉宁窗。由于汉宁窗的主瓣比矩形窗主瓣宽, 利用矩形窗的第二个旁瓣是正值, 使部分抵消汉宁窗的第一旁瓣负值。所以海明窗的第一旁瓣峰值非常小, 但其它旁瓣的衰减没有汉宁窗快, 因为这些旁瓣受海明窗函数中的矩形窗函数支配。海明窗主瓣等效宽度由于矩形窗第一旁瓣负值的部分抵消作用而略优于汉宁窗, 为 $1.4/T$ 。海明窗泄漏很小, 而且也不难获得, 因此海明窗也是常用的时间窗之一。

(5) 布莱克曼(Blackman)窗：

$$w(n) = 0.42 - 0.5 \cos \frac{2\pi}{N-1} + 0.08 \cos \frac{4\pi}{N-1}, \quad 0 \leq n \leq N-1 \quad (11-44)$$

它是 5 个矩形窗谱的叠加。前两项与汉宁窗类似, 是两个矩形窗谱的叠加; 后一项构成两个调整副谱, 使旁瓣进一步压低, 但主瓣宽度增加到 $12\pi/N$ 。

(6) 凯塞(Kaiser)窗的表达式是：

$$w(n) = \frac{I_0 \left[\beta \sqrt{1 - \left(\frac{2n}{N-1} - 1 \right)^2} \right]}{I_0(\beta)}, \quad 0 \leq n \leq N-1 \quad (11-45)$$

式中, $I_0[\cdot]$ 是第一类修正的零阶贝塞尔函数, 可以用级数展开来计算它的值:

$$I_0(x) = 1 + \sum_{k=1}^{\infty} \left[\frac{(x/2)^k}{k!} \right]^2 \quad (11-46)$$

一般取 15~25 次计算, 便可达到足够精度。 $I_0[x]$ 在 $[0, \infty]$ 上是单调增函数, $I_0[0]=1$ 。 β 是决定窗函数形状从而决定窗函数特性的参数。不管 β 是多少, 窗口的中心点 $n=(N-1)/2$ 处总是 1: $w(N-1)/2 = I_0[\beta]/I_0[0] = 1$ 。而在窗的两边 $n=0$ 和 $n=N-1$ 处, $w(0)=w(N-1)=1/I_0[\beta]$ 。 β 越大, $I_0[\beta]$ 就越大, 窗包络两边 $w(0)=w(N-1)$ 就压得越低。改变 β 值, 就能得到不同特性的窗: $\beta=0$ 时凯塞窗相当于矩形窗, $\beta=5.44$ 时相当于海明窗,

$\beta=8.5$ 时接近于布莱克曼窗。从频谱看, β 越大, 主瓣越宽, 旁瓣峰值衰耗越大, 实践中常取 $4 < \beta < 9$ 。

几种常用窗口的包络形状见图 11-5, 性能比较见表 11-1。其中, 旁瓣衰减速度(第一旁瓣高度/主瓣高度)是指旁瓣峰值下降的渐进速度, 单位是 dB/oct(倍频程)。从表 11-1 可明显地看出: 窗函数的主瓣宽度和旁瓣峰值衰耗是矛盾的, 一项指标的提高总是以另一项指标的下降为代价, 窗口选择实际上是对两项指标作权衡。以上介绍的前五种窗中, 这两项指标是跳变的。于是有人提出可调整窗, 适当修改参数, 可在这两项指标间作连续的选择。常用的可调整窗是凯塞窗。

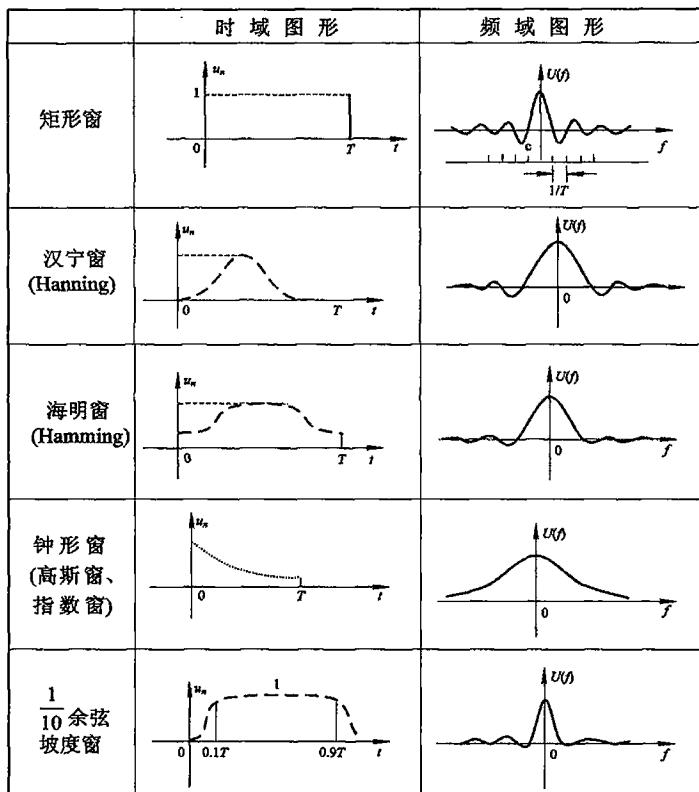


图 11-5 五种窗函数波形及其窗谱

表 11-1 五种窗函数的特性

名 称	主 瓣 宽 度	第一旁瓣高度/主瓣高度
矩形窗	$1/T$	21%
汉宁窗	$1.5/T$	2.5%
海明窗	$1.4/T$	0.8%
钟形窗	$1.9/T$	无旁瓣
$\frac{1}{10}$ 余弦坡度窗	小	很小

11.1.4 [示例 11-1]用窗口法设计一个线性相位低通 FIR 滤波器

要求：截止频率为 f_c ，采样频率是 $8f_c$ ；通带范围内，衰减度不超过 5.8 dB。

(1) 写出理想的频响：

$$H_d(e^{j\omega}) = \begin{cases} e^{-j\omega}, & |\omega| \leq \omega_c \\ 0, & \omega_c < |\omega| \leq \pi \end{cases} \quad (11-47)$$

式中 $\omega_c = 2\pi f_c T = 2\pi f_c / f_s = 0.25\pi$ ； $\alpha = (N-1)/2$ 。

(2) 选窗的形状与大小：令 $N=11$ ，此时，

① 如用矩形窗

$$w_R(n) = R_N(n)$$

② 如用海明窗，把 $N=11$ 和 $n=0, 1, \dots, 10$ 代入式(11-43)

$$w(n) = 0.54 - 0.46 \cos \frac{2n\pi}{N-1} \quad 0 \leq n \leq N-1$$

得

$$w(n) = \{0.08, 0.0168, 0.399, 0.682, 0.912, 1, 0.912, 0.682, 0.399, 0.168, 0.008\} \quad (11-48)$$

(3) 加窗：

$$h(n) = h_d(n) \cdot w(n)$$

序列相乘等于对应项相乘。

如用矩形窗，由于矩形窗的系数均为 1，因此直接取式(11-27)的中间 11 项即可：

$$h(n) = \{-0.045, 0, 0.075, 0.159, 0.225, 0.25, 0.225, 0.159, 0.075, 0, -0.045\} \quad (11-49)$$

如用海明窗，则用式(11-48)和式(11-47)的中间 11 项的对应项相乘，例如 $-0.045 \times 0.08 = -0.0036$ ，于是得到：

$$h(n) = \{-0.0036, 0, 0.03, 0.108, 0.205, 0.25, 0.205, 0.108, 0.03, 0, -0.0036\} \quad (11-50)$$

$h(n)$ 就是所要求的 FIR 滤波器系数。从式(11-49)和(11-50)中可以看出，这两个滤波器的系数的确是奇数偶对称的。

(4) 检验。 $h(n)$ 是否合乎要求需要通过检验来判别，特别地，还要检验滤波器在截止频率处的幅频特性是否满足要求。于是，令 $\omega=0.25\pi$ ，分别把式(11-49)和式(11-50)代入 $H(j\omega) = \sum_{n=0}^{N-1} h(n)e^{-j\omega n}$ ，并计算 $-20 \lg 10(|H(e^{j\omega})|)$ ，分别得到 5.58 dB 和 5.98 dB。因此，矩形窗设计的滤波器满足要求，海明窗不满足要求，若选用海明窗，需要增大滤波器长度。

关于滤波器的检验，可以在 Matlab 中用 fvtool 函数绘出滤波器的幅频特性和相频特性，并进行对比检验。fvtool 函数功能与调用方法的简介详见下一小节。

11.1.5 [示例 11-2]线性相位 FIR 滤波器在 Matlab 中的实现

本小节首先介绍几个 Matlab 中有关线性相位 FIR 滤波器设计与检验的函数功能与使

用方法，然后举例说明如何采用这些函数进行线性相位 FIR 滤波器设计与检验。鉴于在 Matlab 环境下的编程中，参数无法按照规范的书写方法赋予斜体、正体等格式，为与 Matlab 保持一致，本节中所有的参数，不论是常量还是变量，不论是向量还是参量，都写作正体，且不加黑。

1. 函数介绍

1) firceqrip 函数

`firceqrip` 函数用来设计 n 阶具有线性相位的有限冲击响应滤波器，该滤波器在阻带具有等纹波。 n 等于滤波器长度减去 1。调用格式有多种，其调用格式与参数意义分别如下：

(1) $h = \text{firceqrip}(n, wo, del)$ ：用来设计 n 阶具有线性相位的有限冲击响应低通滤波器，该滤波器在阻带具有等纹波。参数 wo 指定截止频率； $del = [d1, d2]$ 指定峰值或者通带与阻带的最大允许误差，其中 $d1$ 设置通带误差， $d2$ 指定阻带误差。由于 `firceqrip` 采用的是正则化频率，也就是采样频率为 f_s ，低于采样频率的信号频率 f 正则化为 $w = 2f/f_s$ 。显然，根据采样定律，信号频率不能高于采样频率的 $1/2$ ，即 $w \leq 1$ ，因此 wo 的值必须设置在 $(0, 1)$ 范围内。

(2) $h = \text{firceqrip}(\dots, 'slope', r)$ ：用输入关键词 '`slope`' 和 r 设计阻带不具有等纹波特性的滤波器。 r 以 dB 为单位，确定阻带的倾斜度， $r > 0$ 。

(3) $h = \text{firceqrip}(\dots, 'passedge')$ ：指定通带起始频率。

(4) $h = \text{firceqrip}(\dots, 'stopedge')$ ：指定阻带起始频率。

(5) $h = \text{firceqrip}(\dots, 'high')$ ：设计高通滤波器。

(6) $h = \text{firceqrip}(\dots, 'min')$ ：设计最小相位 FIR 滤波器。

(7) $h = \text{firceqrip}(\dots, 'invsinc', c)$ ：设计具有 sinc 函数形状的低通滤波器，关键词 `invsinc` 采用逆 sinc 函数，它由 c 是标量还是二元素向量来确定。

- 当 c 为标量时，通带采用函数 $1/\sin(c * w)$ ，其中 w 是正则化频率。
- 当 c 为二元素向量 $[c_1 c_2]$ 时，通带采用函数 $1/\sin(c_1 * w)c_2$ ，其中 w 是正则化频率。

2) filter 函数

`filter` 函数根据滤波器参数对信号进行滤波，并返回滤波后得到的信号。`filter` 函数也有多种调用格式，分别介绍如下。

(1) $y = \text{filter}(b, a, x)$ ：参数 a 和 b 分别指定滤波器分母和分子中的系数向量。该函数的作用实际上就是实现如下运算：

$$\begin{aligned} a(1)y(n) &= b(1)x(n) + b(2)x(n-1) + \cdots + b(nb+1)x(n-nb) \\ &\quad - a(2)y(n-1) - \cdots - a(na+1)y(n-na) \end{aligned}$$

如果 $a(1)$ 不等于 1，滤波器对滤波器系数正则化；如果 $a(1)=0$ ，返回错误。如果 x 是矩阵，那么该函数对矩阵的每一列进行滤波；如果 x 是一个多维阵列，那么函数沿着 x 的第一个非单维阵列滤波。

(2) $[y, zf] = \text{filter}(b, a, x)$ 和 $[y, zf] = \text{filter}(b, a, x, zi)$ ：滤波后产生一个附加输出 zf ，包含从 0 初始状态计算得到的最后状态向量。

(3) $y = \text{filter}(Hq, x)$ ：用滤波器 Hq 对输入数据 x 进行滤波，得到滤波后的输出数据 y ，向量 x 和 y 具有相同的长度。如果 x 是矩阵，那么该函数对矩阵的每一列进行滤波；如果 x 是一个多维阵列，那么函数沿着 x 的第一个非单维阵列滤波。

(4) $[y, zf] = \text{filter}(Hq, x)$: 滤波后产生一个附加输出 zf , 包含从 0 初始状态计算得到的最后状态向量。

3) fvtool 函数

`fvtool` 函数可用来分析数字滤波器的图形化用户界面, 它也有多种调用格式。

(1) `fvtool(B, A)`: 装载滤波器可视化工具, 并计算滤波器的幅频响应, 公式如下:

$$\frac{B(e)^{j\omega}}{A(e)^{j\omega}} = \frac{b(1)e^{-j\omega} + b(2)e^{-j2\omega} + \dots + b(m+1)e^{-j(m+1)\omega}}{a(1) + a(2)e^{-j\omega} + \dots + a(n+1)e^{-j(n+1)\omega}}$$

分子和分母的系数分别在向量 B 和 A 中。

(2) `fvtool(B, A, B1, A1, ...)`: 实现多滤波器分析, 如果用来实现有限冲击响应滤波器的分析, 把 $A_i(i=1, 2, \dots)$ 设为 1 即可。如果只对一个有限冲击响应滤波器进行分析, 可直接写作 `fvtool(h)`, h 为 FIR 滤波器的系数向量。

数字滤波器的图形化用户界面如图 11-6 所示, 点击图标 , 只显示滤波器的幅频特性曲线, 点击 只显示相频特性, 点击 则同时显示滤波器的幅频特性曲线和相频特性曲线。点击 显示滤波器的冲击响应, 点击 则显示滤波器的阶跃响应, 点击 显示滤波器的零点和极点, 点击 则可查看滤波器的系数。

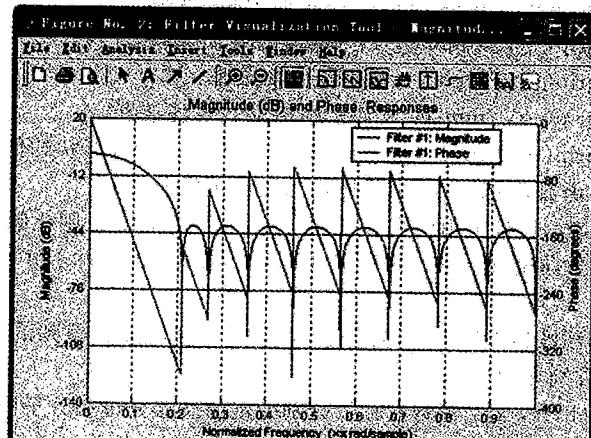


图 11-6 数字滤波器的图形化用户界面

从图 11-6 中还可以看出, 相位曲线成锯齿波状, 这是将相位归算到 $0 \sim -360^\circ$ 的结果。从左到右, 每多一个锯齿, 其实际相位是显示的相位多增加一次 360° , 即从左边开始, 第 $i(i=1, 2, 3, \dots)$ 个锯齿中显示的相位为 θ , 其实际相位为 $-(i-1) \times 360^\circ + \theta$ 。例如, 若第二个锯齿中显示相位为 -180° , 其实际滞后的相位为 $-(2-1) \times 360^\circ - 180^\circ = -540^\circ$ 。另外, 除了第一锯齿以外, 其它锯齿的最大值往往不是 0, 最小值也不是 -360° , 这是由于滤波器长度有限所致。上一个锯齿的结束点和下一个锯齿的起始点, 这两个相邻离散化相位之间有一定的间隔, 这个间隔分散在上一个锯齿的结束处和下一个锯齿的起始处。滤波器长度越大, 则间隔越小, 锯齿的起始点越接近 0, 结束点越接近 -360° 。

2. 应用实例

本实例首先生成一个含有随机噪声的方波，然后设定滤波器参数，获得线性相位滤波器系数，再通过卷积运算实现滤波。

程序如下：

```
%生成方波
t=0:0.01:20;
for j=1:length(t)
    for i=1:10
        if t(j)<=2*i&t(j)>2*i-1
            a(j)=1+0.2*(rand-0.5);%produce random noise whose amplitude is 0.2
        end
        if t(j)<=2*i-1&t(j)>2*i-2
            a(j)=-1+0.2*(rand-0.5);
        end
    end
end
%设定滤波器参数，获得线性相位滤波器系数
n=19;           %设定滤波器阶次为 19
wo=0.1;          %设定截止频率为 0.1
del=[0.001, 0.01]; %指定通带与阻带的最大允许误差分别为 0.001, 0.01
h=firceqrrip(n, wo, del);
%得到的滤波器系数为：[-0.0015  0.0091  0.0158  0.0269  0.0407  0.0561
% 0.0716  0.0852  0.0955  0.1010  0.1010  0.0955  0.0852  0.0716  0.0561
% 0.0407  0.0269  0.0158  0.0091  -0.0015];
%卷积实现滤波
for i=1:length(t)
    c=0;
    for j=1:n+1
        if i>j
            c=c+h(j)*a(i-j+1);
        end
    end
    b(i)=c;          %b 为滤波后得到的信号
end
%直接调用函数 filter 实现滤波
y=filter(h, 1, a);
%显示信号
figure(1)
fvtool(h)          %显示滤波器的幅频特性和相频特性
figure(2)
subplot(4, 1, 1)    %在同一张图中设定 4 行 1 列的 4 个子图，并指定第一行的子图
plot(t, a, 'r')     %在第一行的子图中用红色线条('r'指定红色)绘出方波 a 的波形。横坐
```

```

    标为时间轴 t
    subplot(4, 1, 2)
    plot(t, y, 'b')      %在第二行的子图中用蓝色线条('b'指定蓝色)绘出用 filter 函数滤波后的
                          %信号 y 的波形, 横坐标为时间轴 t
    subplot(4, 1, 3)
    plot(t, b, 'c')      %在第三行的子图中用青色线条('c'指定青色)绘出自编程滤波后的信
                          %号 b 的波形, 横坐标为时间轴 t
    subplot(4, 1, 4)
    plot(t, b-y, 'k')    %在第四行的子图中用黑色线条('k'指定黑色)绘出两种滤波方法得到
                          %的信号的差异, 横坐标为时间轴 t

```

最后得到的滤波器幅频特性(蓝线)与相频特性(绿线)结果如图 11-6 所示。从图中可以看出, 在频率高于 0.27 以后, 即在阻带区, 滤波器具有等纹波, 而在通带内, 滤波器具有线性相位。

初始方波信号、滤波后的信号以及两种实现滤波计算得到的滤波信号的差异如图 11-7 所示。对比图 11-7(a)与(b)和(c)可以看出, 滤波后的信号具有一定度的信号滞后。而由图 11-7(d)可以看出, 两种实现滤波计算的方法得到的信号的差异非常小, 其幅值不及初始信号幅值的 $1/10^{15}$, 这说明 filter 函数所完成的工作实质上就是滤波器系数与初始信号的卷积过程。两种实现滤波计算方法得到的结果之间的差异是由于计算的截断误差所带来的。

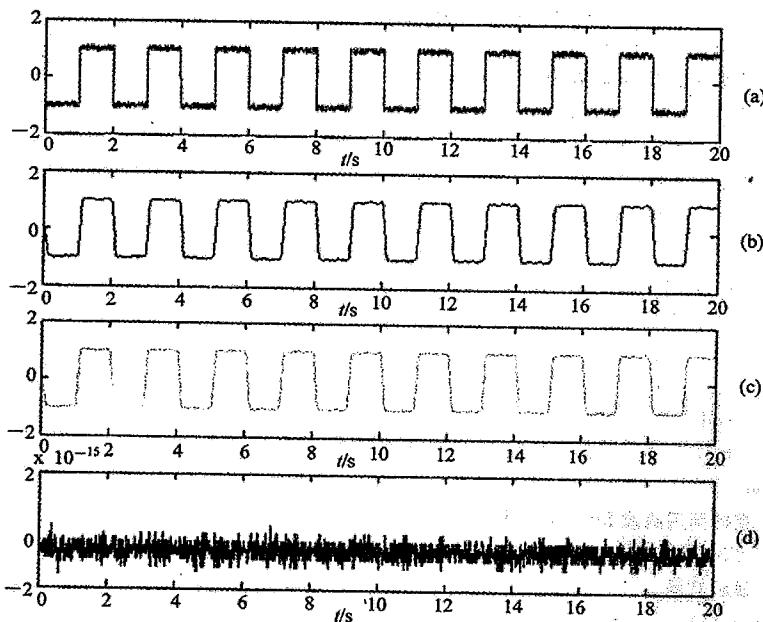


图 11-7 滤波结果

- (a) 初始方波信号;
- (b) 用 filter 函数计算得到的滤波信号;
- (c) 自行编程计算得到的滤波信号;
- (d) 两种滤波信号的差异

11.2 自适应滤波器

所谓的自适应滤波，就是利用前一时刻获得的滤波器参数的结果，自动调节当前时刻的滤波器参数，以适应信号和噪声未知的或随时间变化的统计特性，从而实现最优滤波。自适应滤波是在计算机技术、维纳(Wiener)滤波、卡尔曼(Kalman)滤波等线性滤波基础上发展起来的一种“最佳”滤波方法。这里所谓的“最佳”，是指滤波后的信号和期望输出信号之间的偏差的能量最小。由于自适应滤波具有更强的适应性和更优的滤波性能，从而在工程实际中，尤其在信息处理技术中得到广泛的应用。

自适应滤波把研究对象看做是不确定的系统或信息过程。“不确定”是指所研究的处理信息过程及其环境的数学模型不是完全确定的，其中包含一些未知因数和随机因数。实质上，任何一个实际的信息过程都具有不同程度的不确定性，这些不确定性有时表现在过程内部，有时表现在过程外部。从过程内部来讲，描述研究对象即信息动态过程的数学模型的结构和参数，我们事先不知道。而外部环境对信息过程的影响可以等效地用扰动来表示，但这些扰动通常也是不可测的，它们可能是确定的，也可能是随机的。此外，一些测量噪音也以不同的途径影响信息过程。这些扰动和噪声的统计特性常常是未知的。面对这些客观存在的各种不确定性，如何综合处理信息过程，并使某一些指定的性能指标达到最优或近似最优，这就是自适应滤波所要解决的问题。

自适应滤波器不需要关于输入信号的先验知识，计算量小，特别适用于实时处理。自适应滤波器的特性变化是由自适应算法通过调整滤波器系数来实现的。一般而言，自适应滤波器由两部分组成：一是滤波器结构；二是调整滤波器系数的自适应算法。

11.2.1 自适应滤波器的结构

自适应滤波器的结构采用 FIR 或 IIR 结构均可，由于 IIR 滤波器存在稳定性问题，因此一般采用 FIR 滤波器作为自适应滤波器的组成结构。图 11-8 示出了 FIR 自适应滤波器的一般结构。图中的 $x(n)$ 为滤波器输入信号，也就是初始的测试信号， $y(n)$ 为滤波器输出信号， $d(n)$ 为参考信号或期望信号， $e(n)$ 则是 $d(n)$ 和 $y(n)$ 的误差信号。自适应滤波器的滤波器系数受误差信号 $e(n)$ 控制，根据 $e(n)$ 的值和自适应算法自动调整。

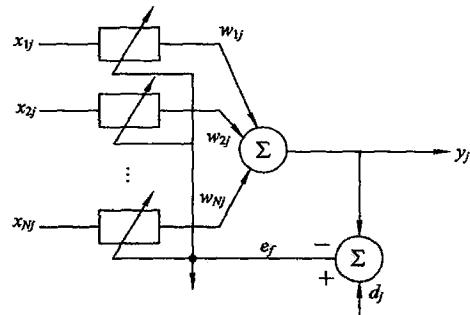


图 11-8 FIR 自适应滤波器结构

11.2.2 自适应滤波理论与算法

对于一个含有噪声的观测信号，它可以表示为

$$x(n) = s(n) + v(n) \quad (11-51)$$

式中 $s(n)$ 为信号的真值， $v(n)$ 为噪声信号的值， $x(n)$ 为含有噪声的观测值。自适应滤波器

本身的形式与前述的 FIR 响应滤波器形式并无二样，即 $y(n) = \sum_{i=0}^{N-1} w_i(n)x(n-i)$ 。与经典滤波的区别在于：经典滤波的前提是把 $v(n)$ 看做频率可预测的，而且是一般不会变换的，其幅值往往是有一定范围的，而自适应滤波则把 $v(n)$ 看做未知的，其频率、幅值都是时变的。这一区别最终导致滤波器系数 w_i 的求取方法完全不同。经典滤波器的设计往往是由设定的滤波器指标来进行的，而自适应滤波器中权值的获得是根据测试数据与期望输出数据通过一定的算法来实现的。

在自适应滤波中，最广泛采用的目标函数之一是均方误差(MSE)，其定义为

$$F(e(n)) = \xi(n) = E(e^2(n)) = E(d^2(n) - 2d(n)y(n) + y^2(n)) \quad (11-52)$$

假设自适应滤波器是由线性组合器构成的，即输出信号是由来自于阵列的信号的线性组合构成，在许多应用中，输入信号向量的每个元素是由同一个信号的时延形式构成的，此时，信号 $y(n)$ 是输入信号经过 FIR 滤波的结果：

$$y(n) = \sum_{i=0}^{N-1} w_i(n)x(n-i) = w^T(n)x(n) \quad (11-53)$$

式中： $x(n) = [x(n), x(n-1), \dots, x(n-N)]^T$ ， $w(n) = [w_0(n), w_1(n), \dots, w_N(n)]^T$ 。此时，

$$\begin{aligned} E(e^2(n)) &= \xi(n) = E[d^2(n) - 2d(n)w^T(n)x(n) + w^T(n)x(n)x^T(n)w(n)] \\ &= E[d^2(n)] - 2E[d(n)w^T(n)x(n)] + E[w^T(n)x(n)x^T(n)w(n)] \end{aligned} \quad (11-54)$$

对于固定系数的滤波器而言，MSE 函数为

$$\begin{aligned} \xi(n) &= E[d^2(n)] - 2w^T E[d(n)x(n)] + w^T E[x(n)x^T(n)]w \\ &= E[d^2(n)] - 2w^T P + w^T R w \end{aligned} \quad (11-55)$$

式中 $P = E[d(n)x(n)]$ ，是输入信号与期望信号之间的互相关向量； $R = E[x(n)x^T(n)]$ ，是输入信号的自相关矩阵。MSE 函数对向量 w 的梯度向量为

$$g_w = \frac{\partial \xi}{\partial w} = \left[\frac{\partial \xi}{\partial w_0} \quad \frac{\partial \xi}{\partial w_1} \quad \dots \quad \frac{\partial \xi}{\partial w_N} \right]^T = -2P + 2Rw \quad (11-56)$$

如果 P 向量和矩阵 R 已知，令该梯度向量为 0，可以直接求解 w ：

$$w_0 = R^{-1}P \quad (11-57)$$

上式中的解称为维纳解。遗憾的是， P 向量和矩阵 R 很难精确估计，当输入和期望信号都是遍历性过程时，能利用时间平均估计 R 和 P ，大多数自适应算法都隐含地利用了这一点。

理论上讲，自适应滤波问题没有唯一的解。为了得到自适应滤波器及其应用系统，可以采用各种不同的递推算法，这些自适应算法都具有各自的特点，适用于不同场合。常用的自适应滤波算法包括基于最小二乘准则的方法、基于卡尔曼滤波理论的方法、基于维纳滤波理论的方法和基于神经网络理论的方法。限于篇幅，本章简要介绍最小均方算法和基于最小二乘准则的算法。

1. 最小均方算法

最小均方算法(Least Mean Square, LMS)是由 Widrow 和 Hoff(1960)开发并命名的，它是随机梯度算法族中的一员。LMS 算法的一个显著特点是它的简单性，此外，它不需要计算有关的相关函数，也不需要矩阵求逆运算，这使得它成为其它线性自适应滤波算法的

参照标准。

LMS 算法包含两个过程：

(1) 滤波过程，包括计算线性滤波器输出对输入信号的响应以及通过比较输出结果与期望响应产生的估计误差。

(2) 自适应过程，即根据估计误差自动调整滤波器参数。

这两个过程一起工作组成一个反馈环。首先，我们有一个用来完成滤波过程的横向滤波器；其次，我们有一个调节滤波器系数的自适应控制算法。

对于式(11-55)，如果可以得到 \mathbf{P} 向量和矩阵 \mathbf{R} 的较好的估计，分别记为 $\hat{\mathbf{P}}(n)$ 和 $\hat{\mathbf{R}}(n)$ ，则可以利用如下最陡下降算法搜索维纳解：

$$\mathbf{w}(n+1) = \mathbf{w}(n) - \mu \hat{\mathbf{g}}_w(n) = \mathbf{w}(n) + 2\mu(\hat{\mathbf{P}}(n) - \hat{\mathbf{R}}(n)\mathbf{w}(n)) \quad (11-58)$$

式中 $\hat{\mathbf{g}}_w(n)$ 表示目标函数对滤波器系数的梯度向量估计值。一种可能的解是利用 \mathbf{R} 和 \mathbf{P} 的瞬态估计值来估计梯度向量，即

$$\hat{\mathbf{R}}(n) = \mathbf{x}(n)\mathbf{x}^T(n) \quad (11-59)$$

$$\hat{\mathbf{P}}(n) = d(n)\mathbf{x}(n) \quad (11-60)$$

于是得到的梯度估计值为

$$\begin{aligned} \hat{\mathbf{g}}_w(n) &= -2d(n)\mathbf{x}(n) + 2\mathbf{x}(n)\mathbf{x}^T(n)\mathbf{w}(n) \\ &= 2\mathbf{x}(n)[\mathbf{x}^T(n)\mathbf{w}(n) - d(n)] = -2e(n)\mathbf{x}(n) \end{aligned} \quad (11-61)$$

如果用瞬态平方误差 $e^2(n)$ 代替 MSE 作为目标函数，梯度估计值则代表了真实的梯度向量，因为

$$\begin{aligned} \frac{\partial e^2(n)}{\partial \mathbf{w}} &= \left[2e(n) \frac{\partial e(k)}{\partial w_0(n)} \quad 2e(n) \frac{\partial e(k)}{\partial w_1(n)} \quad \dots \quad 2e(n) \frac{\partial e(k)}{\partial w_N(n)} \right]^T \\ &= -2e(n)\mathbf{w}(n) = \hat{\mathbf{g}}_w(n) \end{aligned} \quad (11-62)$$

由于得到的梯度算法使得平方误差的均值最小化，因此被称为 LMS 算法，其更新方程为

$$\mathbf{w}(n+1) = \mathbf{w}(n) + 2\mu e(n)\mathbf{x}(n) \quad (11-63)$$

式中，收敛因子 μ 应该在一个小的范围内取值，以保证算法的收敛性。

2. RLS 算法

最小二乘(Least Square, LS)算法旨在使期望信号与模型滤波器输出之差的平方和达到最小。当在每次迭代中接收到输入信号的新采样值时，可采用递归形式求解最小二乘问题，得到递归最小二乘(Recursive Least Square, RLS)算法。

RLS 算法能实现快速收敛，即使是在输入信号相关矩阵的特征值扩展比较大的情况下。当工作于时变环境中时，这类算法具有极好的性能，但其实现都以增加计算复杂度和稳定性问题为代价，而这些问题对于基于 LMS 准则的算法来说却并不重要。

对于最小二乘算法，目标函数是确定性的，并且由下式给出：

$$\xi^d(n) = \sum_{i=0}^n \lambda^{n-i} e^2(i) = \sum_{i=0}^n \lambda^{n-i} [d(i) - \mathbf{x}^T(i)\mathbf{w}(n)]^2 \quad (11-64)$$

式中 λ 为指数加权因子或遗忘因子，其值应选择在 $[0, 1]$ 范围内。

应该注意的是，在推导 LMS 算法和基于 RLS 准则的算法时，利用了先念误差。在 RLS 算法中，用 $e(n)$ 表示后念误差， $e'(n)$ 表示先念误差，在推导基于 RLS 准则算法的过程中，将首先选择后念误差。

每一个误差都是由期望信号和采用最近的系数 $w(n)$ 得到的滤波器输出之差所组成的。将 $\xi^d(n)$ 相对于 $w(n)$ 求导，可以得到：

$$\frac{\partial \xi^d(n)}{\partial w(n)} = -2 \sum_{i=0}^n \lambda^{n-i} x(i) [d(i) - x^T(i)w(n)] \quad (11-65)$$

令式(11-65)等于 0，则可以通过如下关系式找到使得最小二乘误差最小的最优向量 $w(n)$ ：

$$-\sum_{i=0}^n \lambda^{n-i} x(i) x^T(i) w(n) + \sum_{i=0}^n \lambda^{n-i} x(i) d(i) = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \quad (11-66)$$

从而得到最优系数向量 $w(n)$ 的表达式为

$$w(n) = \left[\sum_{i=0}^n \lambda^{n-i} x(i) x^T(i) \right]^{-1} \sum_{i=0}^n \lambda^{n-i} x(i) d(i) = R_D^{-1}(n) P_D(n) \quad (11-67)$$

式中： $P_D(n) = \sum_{i=0}^n \lambda^{n-i} x(i) d(i)$ ，称为输入信号与期望信号的确定性互相关向量；

$$\begin{aligned} R_D(n) &= \sum_{i=0}^n \lambda^{n-i} x(i) x^T(i) = x(n) x^T(n) + \lambda \sum_{i=0}^{n-1} \lambda^{n-i} x(i) x^T(i) \\ &= x(n) x^T(n) + \lambda R_D(n-1) \end{aligned} \quad (11-68)$$

称为输入信号的确定性相关矩阵。

为了避免求矩阵 $R_D(n)$ 的逆，我们可以利用如下的矩阵求逆引理：

$$[A + BCD]^{-1} = A^{-1} - A^{-1}B[D^{-1}B + C^{-1}]^{-1}D^{-1} \quad (11-69)$$

式中 A 、 B 、 C 和 D 是具有合适维数的矩阵，并且矩阵 A 和 C 是非奇异矩阵。上述关系可以简单地通过将右边表达式左乘 $A + BCD$ 得到一个恒等式来证明。如果我们选取 $A = \hat{R}_D(n-1)$ ， $B = D^T = x(n)$ ， $C = 1$ ，由式(11-66)可知， $\hat{R}^{-1}(n) = [A + BCD]^{-1}$ 。于是由矩阵求逆引理可得：

$$\begin{aligned} \hat{R}_D^{-1}(n) &= A^{-1} - A^{-1}B[D^{-1}B + C^{-1}]^{-1}D^{-1} \\ &= [\lambda \hat{R}_D(n-1)]^{-1} - [\lambda \hat{R}_D(n-1)]^{-1} x(n) [x^T(n) [\lambda \hat{R}_D(n-1)]^{-1} x(n) + 1]^{-1} \\ &\quad \cdot x^T(n) [\lambda \hat{R}_D(n-1)]^{-1} \\ &= \frac{1}{\lambda} \left[\hat{R}_D^{-1}(n-1) - \frac{\hat{R}_D^{-1}(n-1) x(n) x^T(n) \hat{R}_D^{-1}(n-1)}{\lambda + x^T(n) \hat{R}_D^{-1}(n-1) x(n)} \right] \end{aligned} \quad (11-70)$$

利用式(11-70)来计算 $\hat{R}^{-1}(n)$ ，与每次迭代过程中直接计算 $\hat{R}(n)$ 的逆矩阵（乘法次数在 N^3 数量级）相比，更新方程的计算复杂度更低（乘法次数在 N^2 数量级）。由于每次计算 $\hat{R}^{-1}(n)$ 时，总是用到 $\hat{R}^{-1}(n-1)$ ，整个滤波过程中求取 $\hat{R}^{-1}(n)$ 就成了一个递归形式，因此

这种最小二乘法成为递归最小二乘法。但是，当 $n < N$ 时（即初始化阶段）， $\hat{\mathbf{R}}^{-1}(n)$ 总是奇异的，这时需要通过所谓的后向代入法来进行初始化，即从 $n=0$ 到 $n=N$ 的时期，可以不用任何矩阵求逆运算而准确地得到 $w_i(n)$ 。

当 $n=0$ 时，只要 $x(0) \neq 0$ ，则有：

$$w_0(0) = \frac{d(0)}{x(0)}$$

当 $n=0$ 时，有：

$$w_0(1) = \frac{d(0)}{x(0)}, \quad w_1(1) = \frac{-x(1)w_0(1) + d(1)}{x(0)}$$

当 $n=0$ 时，有：

$$\begin{aligned} w_0(2) &= \frac{d(0)}{x(0)}, \quad w_1(1) = \frac{-x(1)w_0(2) + d(1)}{x(0)} \\ w_2(2) &= \frac{-x(2)w_0(2) - x(1)w_1(2) + d(2)}{x(0)} \end{aligned}$$

利用归纳法，可以证明在任意时刻 n ，有：

$$w_i(n) = \frac{-\sum_{j=1}^i x(j)w_{i-j}(n) + d(i)}{x(0)} \quad (11-71)$$

这样，在计算初始的 $N+1$ 个滤波器系数时，采用式 (11-71) 即可，不需要计算式 (11-70)。

11.2.3 Matlab 中的自适应滤波函数

1. `s = initlms(w0, mu, zi, lf)`

该函数返回完整的种群结构 `s`，`s` 在调用 `adaptlms` 函数时需要用到。向量 `w0` 包含滤波器系数的初始值，它的长度应等于自适应滤波器长度加 1。`mu` 是最小均方 (LMS) 法的步长，所指定的步长决定了 LMS 算法收敛到解所用的时间和解的准确度。一般而言，小的步长适应更慢，但更精确；大的步长适应运算快，但是误差较大。

LMS 算法的矩阵形式是：

$$\mathbf{w}(k+1) = \mathbf{w}(k) + \mu \mathbf{e}(k) \mathbf{x}(k)$$

为了确保良好的收敛速度和稳定性，`mu` 一般选择在如下范围内：

$$0 < \mu < \frac{1}{N(\text{输入信号功率})}$$

式中 `N` 为信号中样本的数量。

输入参数 `zi` 指定滤波器初始条件，若忽略 `zi` 或者指定为空，`initkalman` 默认 `zi` 为 0 向量，其长度等于 `length(w0)-1`。对于限定的处理条件，例如 `for` 循环中使用 `adaptlms`，指定初始条件非常重要，LMS 算法的每一次迭代都使用上一次迭代的权值，由于提供了初始条件，因此第一次迭代具有一组先验滤波器权值，所以开始进行迭代运算。

输入参数 `lf` 为泄漏因子。指定泄漏因子可以改善算法的特性。泄漏权值 $w(k)$ 强制算法进一步适应运算，就算已经达到了最小值。这可能意味着泄漏 LMS 没有获得非常精确的最小均方误差的测度时，使用泄漏因子算法可以减小误差的敏感性，或者对输入的小数

值的敏感性。典型地, lf 设置为 0.9 至 1.0 之间, 表示没有泄漏。如果指定 lf 为空, 它默认为 1。

在使用 initlms 函数后, 如果核对 s 的内容, Matlab 会显示结构元素, 而不是输入参数的名称。为了帮助记住各元素对应的 initlms 的输入参数, 表 11-2 给出了该映射关系。

表 11-2 Initlms 结构参数

Initlms 参数	结构域	参数内容
wo	s.coeffs	LMS FIR 滤波器系数。在进行自适应运算之前, 应给 FIR 滤波器初始化系数。这些系数应输入到 s.coeffs 中(长度滤波器阶数加 1)。当把 s 用作一个输出参数的时候, 更新后的滤波器系数返回到 s.coeffs 中
mu	s.step	设置 LMS 算法步长。它决定了自适应滤波器逼近滤波器的解的速度与逼近程度
zi	s.states	自适应运算后, 返回 FIR 滤波器的状态。这是一个可选项, 如果忽略它, 它默认为一个 0 向量, 其长度等于滤波器的阶数。在循环结构中使用 adaptlms 时, 用这个元素指定自适应 FIR 滤波器的初始滤波状态
if	s.leakage	指定 LMS 泄漏参数。允许执行泄漏 LMS 算法。泄漏因子的存在可以达到最小值以后, 通过强行使 LMS 算法继续自适应运算, 以改善算法的结果。这是一个可选项, 如果忽略(不指定或者设置为空), 则默认为 1
	s.iter	自适应滤波器运行中总的迭代次数。虽然这可以在 s 中设置, 但不要这么做, 最好把它看作一个只读参数

2. $[y, e, s] = \text{adaptlms}(x, d, s)$

该函数的功能是: 用 FIR 最小均方自适应滤波器作用于 x 和期望信号 d, 滤波后的信号返回给 y。s 是一个结构体, 它包含定义所采用的 LMS 的初始化设置, 这同一些滤波器适应过程的输出一样。表 11-2 详细列出了 s 的内容, 包括输入和输出。标注为 initlms 的那一列中给出了 s 中相对应于输入参数的元素。

返回参数中的 e 表示预测误差, 它表明的是滤波器使得输入数据适应于期望信号的程度, 或者说 y 逼近 d 的程度; s 即为更新的结构 s。

3. $s = \text{initnlms}(w_0, mu, zi, lf, offset)$

该函数用来构建一个用于调用 adaptnlms 函数的完整的种群结构 s。同 initlms 函数相比, 一方面, 正则化最小均方算法的向量形式为

$$w(k+1) = w(k) + \mu_n e(k)x(k)$$

式中: $\mu_n = 1 / (\epsilon + \|x(k)\|^2)$ 。

另一方面, initnlms 增加了一个参数 offset。它是一个可选的正则化项, 这可以避免输入数据的模的平方非常大的时候被 0(或者非常小的数)除的条件。如果 offset 指定为空, 则默认为 0。与输入参数 offset 相对应, 种群结构 s 中有一个元素 s.offset。

4. $[y, e, s] = \text{adaptnlms}(x, d, s)$

该函数为正则化最小均方(Normalized Least Mean Squared, NLMS)FIR 自适应滤波函数, 滤波后的数据返回到 y 中, 输入参数结构体 s 包含了定义所采用的 NLMS 自适应算法的初始设置, 更新后的参数在输出参数结构体 s 中, 它的各个元素的含义见 initnlms 函数。

数简介。输出参数 e 表示预测误差，它表明了 y 对 d 的近似程度。

5. $s = \text{initkalman}(w0, k0, qm, qp, zi)$

该函数可构建一个用作 adaptkalman 函数的输入参数的初始化结构。向量 w0 包含滤波器系数的初始值，其长度等于自适应 FIR 滤波器阶数加 1。k0 包含初始状态误差协方差矩阵，它应为维数等于 w0 长度的 Hermitian 对称方阵。qm 是测试噪声方差，qp 是过程噪声协方差。zi 用来指定滤波器初始条件，若忽略它或者设置为空，函数默认它为 0 向量，其长度等于 w0 的长度减去 1。如果在诸如 for 循环的程序结构中使用 adaptkalman 函数，初始条件提供第一次循环迭代的滤波器权值。Kalman 滤波器算法的每次迭代都调用前一次迭代的权值。如果没有初始条件，第一次迭代就没有输入可用，后面的每次循环迭代都会发生同样的问题，以至于滤波器无法自适应到某一个结果。

同 initlms 函数一样，在使用 initkalman 函数后再核对 s 的内容，Matlab 会显示结构元素，表 11-3 是 s 的元素与输入参数的映射。

表 11-3 initkalman 结构参数

initkalman 参数	结构域	参数内容
w0	s.coeffs	Kalman 自适应滤波器系数。在进行自适应运算之前，应给 FIR 滤波器初始化系数。在把 s 用作一个输出参数的时候，更新后的滤波器系数返回到 s.coeff 中
k0	s.errcov	状态误差协方差矩阵。用初始误差状态协方差矩阵初始化这个元素。把它用作一个输出参数时，更新后的矩阵返回到该参数中
qm	s.measvar	包含测试噪声方差矩阵
qp	s.procov	包含过程噪声协方差矩阵
	s.states	返回 FIR 滤波器的状态。这是一个可选项，如果忽略，它默认为 0 向量，向量长度为滤波器阶次
	s.gain	Kalman 增益向量。对此参数无要求，但每次迭代后，都会计算并返回
	s.iter	自适应滤波器运行中总的迭代次数。它是一个只读参数

11.2.4 [示例 11-3]自适应滤波在系统辨识中的应用

要求：

- (1) 首先生成一组随机变量，然后用常规低通滤波器进行滤波；
- (2) 将滤波后的信号作为期望信号，用自适应滤波法对随机变量进行自适应滤波；
- (3) 比较自适应滤波器系数与常规低通滤波器系数，以认识自适应滤波器的自适应能力。

首先用 randn() 函数生成随机变量 x，并用窗法设计 FIR 滤波器 b，对生成的随机变量进行滤波得到信号 d，然后用最小均方算法对随机变量进行自适应滤波。待自适应过程结束后，绘出得到的自适应滤波器系数与 FIR 滤波器 b。

整个过程的源代码如下：

```
x = 0.1 * randn(1, 500); %Input to the filter
b = fir1(31, 0.5); %FIR system to be identified
```

```
%fir1 函数采用窗法设计 FIR 滤波器，b= fir1(N, Wn)设计一个 N 阶低通 FIR 数字滤波器，并返回滤波器系数，其长度为 N+1。截止频率 Wn 必须设置在 0~1 之间，1 对应于采样频率的一半。滤波器是实数的，并且具有线性相位，在 Wn 处的正则化增益为 -6 dB
d = filter(b, 1, x); %Desired signal
w0 = zeros(1, 32); %Initial filter coefficients
mu = 0.8; %LMS step size.
s = initlms(w0, mu);
[y, e, s] = adaptlms(x, d, s);
stem([b, '], s.coeffs, ']'); %plot coefficients of FIR filter and that of adaptive filter
legend('Actual', 'Estimated');
title('System Identification of an FIR filter'); grid on;
```

辨识结果如图 11-9 所示。其中星号表示实际的滤波器系数，圆圈表示辨识得到的滤波器系数。从图中可以看出，两个滤波器的系数几乎完全相等，最大差别出现在第 16 和 17 个系数，实际参数均为 0.4501，辨识结果分别为 0.4387 和 0.4382。由此可知，自适应滤波算法具有很强的自适应能力，在获得测试信号后，若还能知道期望信号，采用自适应滤波器可以得到比较理想的滤波器，使得采用该滤波器滤波得到的信号具有高的信噪比，小的相位滞后。

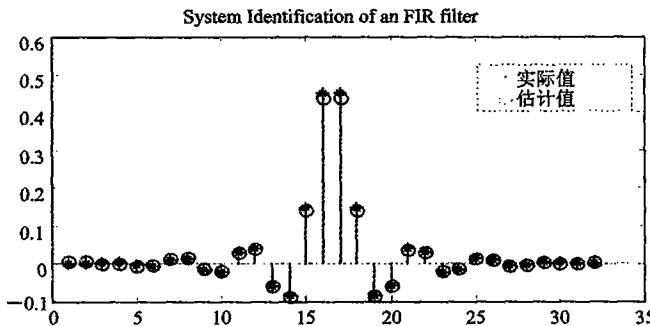


图 11-9 滤波器系数对应关系图

参 考 文 献

- [1] 胡广书. 数字信号处理：理论、算法与实现[M]. 2 版. 北京：清华大学出版社，2003.8
- [2] 陈怀琛. 数字信号处理教程 MATLAB 释义与实现[M]. 2 版. 北京：电子工业出版社，2008.10
- [3] Paulo S R Diniz. 自适应滤波算法与实现[M]. 2 版. 刘郁林，景晓军，谭刚兵，译. 北京：电子工业出版社，2004.7
- [4] 杨晓源. 数字地震观测中的线性和最小相位滤波[J]. 四川地震，2005，2；1-5
- [5] 刘益成，易碧金，巩庆钢，等. 地震数据采集系统相位特性与测试方法[J]. 石油物

探, 2009, 48(2): 168 - 172

- [6] 宫晓琳, 房建成, 盛蔚. 一种 GPS 与高精度气压高度表在线互标定方法[J]. 电子与信息学报, 2009, 31(4): 818 - 821.
- [7] 黄粉平, 张玲, 郑恩让. 快速自适应滤波的图像增强方法[J]. 西安科技大学学报, 2008, 28(4): 762 - 765
- [8] 司锡才, 张雯雯, 李利, 等. 一种新的自适应消噪方法[J]. 宇航学报, 2008, 9(6): 2013 - 2018

第 12 章 模糊智能传感器系统

本章内容

- 模糊理论概述
- 模糊传感器的测量原理
- 用 LabVIEW 实现的模糊智能传感器系统举例
- [示例 12-1] 模糊温度测试仪
- [示例 12-2] 模糊温度控制仪

模糊传感器是顺应人类的生活实践、生产和科学实践的需要而提出并得到迅速发展的。在很多实际问题中，只获得某个物理量的精确数值去对被测对象进行描述仍然是很不完备的。需要进一步根据不同对象，综合有关专家经验知识，进行更高智能的推理、判断，得出用人类语言符号对被测对象进行的定性描述。如变压器油温测得 85℃ 还是 85.1℃ 并不重要，更重要的是，应给出“正常”还是给出“偏高”或“偏低”的结论。于是就很自然地提出两个新问题：一个是测量结果的语言符号化表示和符号化测量系统；另一个是由传统传感器测量得到的数值结果如何通过专家智能转换为用人类语言符号来表示。

模糊传感器就是采用模糊集合理论方法，来实现测量的数值结果转换为人类语言符号表示的一种新型智能传感器系统。其中，数值/符号转换是核心，模糊集合理论则是实现模糊测量的基础。基于计算机的虚拟测量系统的迅速发展，为在测量过程中使用各种复杂的信息处理算法提供了便利。

本章试图应用模糊理论建立虚拟仪器形式的模糊智能传感器系统，并介绍了“模糊温度测试仪”与“模糊温度控制仪”的工作原理及设计方法。

12.1 模糊集合理论概述

模糊理论对系统的描述和刻画建立在自然语言的基础上，而人类历经几千年的历史发展形成的自然语言是人类最方便、最有效的表达方式。模糊理论能快速方便地描述与处理问题主要基于以下特点：

- 模糊逻辑基于自然语言的描述；
- 模糊逻辑可以建立在专家经验的基础上；
- 模糊逻辑容许使用不精确的数据；
- 模糊概念在概念上易于理解；
- 模糊逻辑可以对任意复杂的非线性函数建模。

正是基于上述特点使得模糊逻辑获得广泛应用，如用于家用电器、智能控制、C³I、医

疗诊断、气象预报甚至经济管理等领域，其中最著名的模糊应用是日本仙台市地铁系统的模糊控制。

12.1.1 模糊集合的定义及其表示方法

所谓模糊，通常是指对概念的定义以及语言意义理解上的不确定性。显然，这种模糊性主要体现在主观理解上，这也是人类社会生活和生产过程中经常遇到的，它是定性分析与定量分析、主观分析与客观分析、模糊性与精确性之间的一个人为的折衷。模糊数学正是为解决这类问题而发展起来的，而模糊集合理论则是其基础。但是我们不能产生这样一种思想，即认为模糊数学是模糊的概念。事实恰恰相反，模糊数学是借助定量的方法研究模糊现象的工具，它是精确的。

1. 模糊集合与经典集合

1965年，美国加州伯克利大学的查德教授(L. A. Zadeh)发表了里程碑性的文章《模糊集合》。在这篇文章里他第一次用“模糊”(fuzzy)这个词表示技术文献中的“不分明性”(vague)，由此开创了模糊数学及其应用的新纪元。模糊集合是一种特别定义的集合，它与普通集合既有联系又有区别。对于普通集合而言，任何一个元素要么属于该集合，要么不属于该集合，非此即彼，界限分明；而对于模糊集合，一个元素可以既属于该集合又不属于该集合，亦此亦彼，界限模糊。例如对于“青年”这一概念，在普通集合中需要定义一个明确的边界，如30岁以下才可以肯定为青年；而在模糊集合中，青年的定义集合没有一个明确的边界，30岁以下是青年，40岁也属于青年，50岁在一定程度上也属于青年，只是他们属于“青年”这一集合的程度不同而已。

2. 模糊集合的定义

在介绍模糊集合的定义之前，需要明确与其密切相关的论域的概念。简单来说，所谓论域，就是指所讨论变量的取值范围，就像函数的自变量的取值范围一样。根据所解决问题的需要，论域既可以为连续的，也可以是离散的，由有限个元素构成。

在此基础上，模糊集合的严格数学定义如下：

设给定论域 X ， X 到 $[0, 1]$ 闭区间的任一映射 μ_A

$$\begin{aligned} \mu_A : X &\rightarrow [0, 1] \\ x &\rightarrow \mu_A(x) \end{aligned} \quad (12-1)$$

都确定 X 的一个模糊子集 A ， μ_A 称为模糊子集的隶属函数， $\mu_A(x)$ 称为 x 相对于 A 的隶属度。从定义不难看出，论域 X 的模糊子集 A 由隶属函数 μ_A 来表征， $\mu_A(x)$ 取值范围为闭区间 $[0, 1]$ ， $\mu_A(x)$ 的大小反映了 x 对于模糊子集的从属程度。 $\mu_A(x)$ 的值接近于 1，表示 x 隶属于 A 的程度很高； $\mu_A(x)$ 的值接近于 0，表示 x 隶属于 A 的程度很低。由此可见模糊子集 A 完全由隶属函数 μ_A 所描述。在有些著作中，论域上定义的模糊子集有时也被称为模糊集合，简称为模糊集。

根据上述定义，若以人的年龄的集合 $X = \{x | 0 \leq x \leq 200\}$ 作为论域，即用人的生理年龄($x \in X$)作为元素构成的一个集合，再将“年老”和“年轻”两个模糊概念分别用模糊集合 O 和 Y 表示，那么对于任一 $x \in X$ ，都将以不同的程度隶属于这两个模糊集合。令 $\mu_O(x)$ 、 $\mu_Y(x)$ 分别表示这两个模糊集合的隶属函数，其定义见式(12-2)，对应的隶属函数曲线如

图 12-1 所示。

$$\mu_O(x) = \begin{cases} 0, & 0 \leq x \leq 50 \\ \frac{1}{1 + \left(\frac{5}{x-50}\right)^2}, & 50 < x \leq 200 \end{cases}$$

$$\mu_Y(x) = \begin{cases} 1, & 0 \leq x \leq 25 \\ \frac{1}{1 + \left(\frac{x-25}{5}\right)^2}, & 25 < x \leq 200 \end{cases}$$

(12-2)

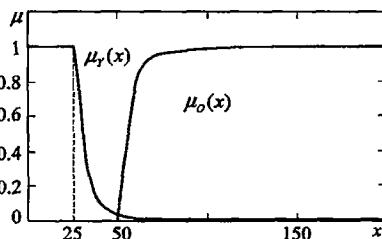


图 12-1 “年轻”和“年老”的隶属函数曲线

从图 12-1 中可以看出，模糊集合 O 和 Y 可完全由其隶属函数刻画。所以对于论域的模糊集合来说，一旦给定它的隶属函数，那么它就完全确定了。不同的隶属函数所确定的模糊集合也不同。从函数的角度而言，隶属函数定义了从论域到单位闭区间 $[0, 1]$ 的一个映射。

此外，对于同一论域，可以定义多个不同的模糊集合，集合数要根据具体情况而定，不能一概而论。例如，以人类的年龄所构成的论域 X 为例，它不但可以划分为{“年老”、“年轻”}，还可以进一步细分为{“婴儿”，“幼儿”，“少年”，“青年”，“中年”，“中老年”，“老年”}等集合。诸如“年老”或“年轻”等，是定义在语言域中的语言变量“年龄”的取值，这些概念在论域 X 上将有确定的模糊集合与其对应。

3. 模糊集合的表示方法

令 A 表示在论域 X 中定义的模糊集合，它的表示有以下三种常用形式：

(1) Zadeh 表示法：

$$A = \frac{\mu_A(x_1)}{x_1} + \frac{\mu_A(x_2)}{x_2} + \cdots + \frac{\mu_A(x_n)}{x_n} = \sum_{i=1}^n \frac{\mu_A(x_i)}{x_i} \quad (12-3)$$

或

$$A = \int \frac{\mu_A(x)}{x}, \quad x \in X \quad (12-4)$$

(2) 向量表示法：

$$A = [\mu_A(x_1), \mu_A(x_2), \dots, \mu_A(x_n)] \quad (12-5)$$

(3) 序偶表示法：

$$A = \{x, \mu_A(x) \mid x \in X\} \quad (12-6)$$

说明：式(12-4)中的积分号并不表示实际的积分运算，只不过是为了区别论域为离散形式时的情形(式(12-3))。

12.1.2 隶属函数的确定方法及常用形式

由模糊集合的定义不难看出隶属函数的重要作用，借助于隶属函数可以把人类模糊语言描述定量化；进而基于计算机实现模仿人类思维、推理和判断的活动。对于在论域 X 上确定的模糊集合的隶属函数，其映射方式是多种多样的，没有统一的模式，一般是根据经验或统计进行确定，也可由专家、权威给出。对于同一模糊概念，不同的人因理解或认识上的差异可能建立完全不同的隶属函数。尽管建立的隶属函数可以有不同的形式，但必须反映客观实际。

1. 确定隶属函数的一般原则

(1) 若模糊集合反映的是社会的一般意识, 是大量的可重复表达的个别意识的平均结果, 例如年轻人、经济增长快、生产正常等, 则此时采用模糊统计法来确定隶属函数是一种切实可行的方法, 不足之处是工作量较大。

(2) 如果模糊集合反映的是某个时间段内的个别意识、经验和判断, 例如, 某专家对某个项目的可行性评价, 那么, 对这类问题可采用 Delphi 法。

(3) 模糊集合反映的模糊概念已有相应成熟的指标, 若这种指标经过长期实践检验已经成为公认的对事物是真实的又是本质的刻画, 则可直接采用这种指标, 或者通过某种检验方式将这种指标转化为隶属函数。

(4) 对某些模糊概念, 虽然直接给出其隶属函数比较困难, 但可以比较两个元素相应的隶属度, 此时可用相对选择法求得其隶属函数。

(5) 若一个模糊概念是由若干个模糊因素复合而成的, 则可以先求单个因素的隶属函数, 再综合出模糊概念的隶属函数。

大多数情况下, 经常是初步确定粗略的隶属函数, 然后再通过“学习”和实践检验逐步修改和完善, 而实际效果正是检验和调整隶属函数的依据。对于模糊统计法、相对选择法、因素加权综合法等具体的隶属函数确定方法, 这里就不详细论述了, 有兴趣的读者可参考有关资料。

2. 常见的隶属函数举例

在以下各例子中, 均设定论域 $X = \{x | 0 \leq x \leq 10\}$, 定义在其上的模糊集合为 A , 隶属度为 $\mu_A(x)$ 。

(1) 高斯(Gaussian)型隶属函数:

$$f(x, \sigma, c) = e^{-\frac{(x-c)^2}{2\sigma^2}} \quad (12-7)$$

它有两个特征参数 σ 和 c 。 $\sigma=2$, $c=5$ 时隶属函数曲线如图 12-2 所示。

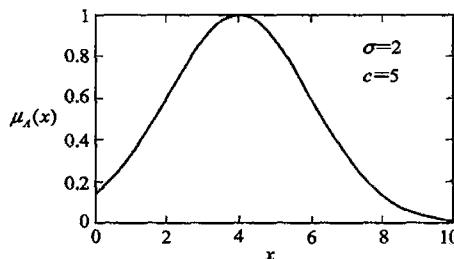


图 12-2 高斯型隶属函数

(2) 钟形隶属函数:

$$f(x, a, b, c) = \frac{1}{1 + \left(\frac{x-c}{a}\right)^{2b}} \quad (12-8)$$

特征参数为 a 、 b 和 c , 由于隶属函数的形状如钟形, 故得其名。如当 $x \in [0, 10]$, $a=2$, $b=4$, $c=6$ 时隶属函数曲线如图 12-3 所示。

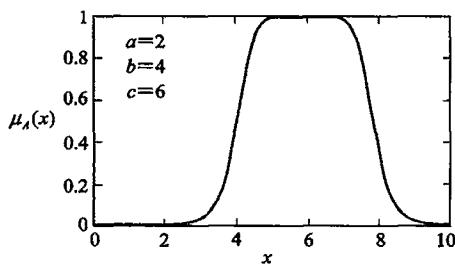


图 12-3 钟形隶属函数

(3) Sigmoid 函数形隶属函数：

$$f(x, a, c) = \frac{1}{1 + e^{-a(x-c)}} \quad (12-9)$$

特征参数为 \$a\$ 和 \$c\$。当 \$x \in [0, 10]\$，\$a=2\$，\$c=4\$ 时对应的隶属函数如图 12-4 所示。

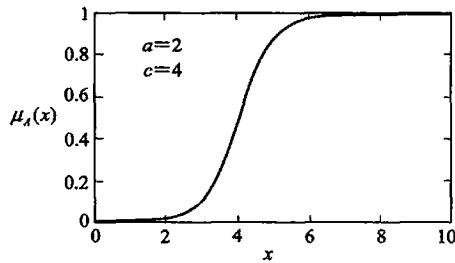


图 12-4 Sigmoid 函数形隶属函数

(4) 差型 Sigmoid 隶属函数：

$$f(x, a_1, c_1, a_2, c_2) = \frac{1}{1 + e^{-a_1(x-c_1)}} - \frac{1}{1 + e^{-a_2(x-c_2)}} \quad (12-10)$$

由两个 S 形隶属函数的差构成的隶属函数。该函数具有四个特征参数 \$a_1\$、\$c_1\$、\$a_2\$、\$c_2\$。当 \$x \in [0, 10]\$，\$a_1=5\$，\$c_1=2\$，\$a_2=5\$，\$c_2=7\$ 时对应的隶属函数如图 12-5 所示。

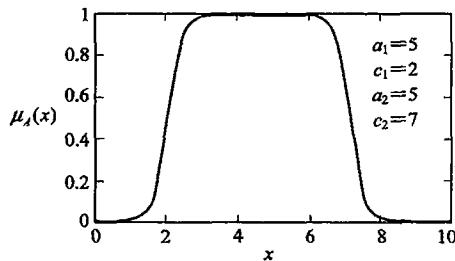


图 12-5 差型 Sigmoid 隶属函数

(5) 梯形隶属函数:

$$f(x, a, b, c, d) = \begin{cases} 0, & x \leq a \\ \frac{x-a}{b-a}, & a \leq x \leq b \\ 1, & b \leq x \leq c \\ \frac{d-x}{d-c}, & c \leq x \leq d \\ 0, & x \geq d \end{cases} \quad (12-11)$$

特征参数为 a, b, c, d 。当 $x \in [0, 10]$, $a=1, b=4, c=6, d=8$ 时对应的隶属函数如图 12-6 所示。另外当 $b=c$ 时, 梯形隶属函数就演变成了下面的三角形隶属函数。

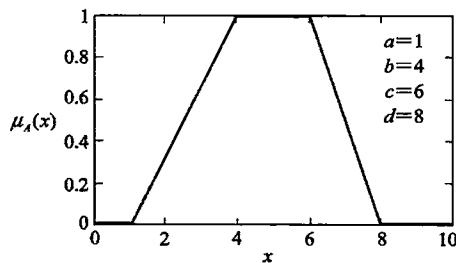


图 12-6 梯形隶属函数

(6) 三角形隶属函数:

$$f(x, a, b, c, d) = \begin{cases} 0, & x \leq a \\ \frac{x-a}{b-a}, & a \leq x \leq b \\ \frac{c-x}{c-b}, & b \leq x \leq c \\ 0, & c \leq x \end{cases} \quad (12-12)$$

特征参数为 a, b, c 。当 $x \in [0, 10]$, $a=2, b=5, c=8$ 时对应的隶属函数如图 12-7 所示。

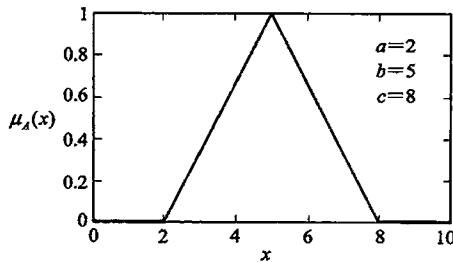


图 12-7 三角形隶属函数

当然, 隶属函数并不限于上述几种形式, 选用何种形式的隶属函数要根据实际应用场合而定。

12.1.3 模糊集合的基本运算

1. 模糊集合的相等

若有两个模糊集合 A 和 B , 对于所有的 $x \in X$, 均有 $\mu_A(x) = \mu_B(x)$, 则称模糊集合 A 与 B 相等, 记作 $A=B$ 。

2. 模糊集合的包含关系

若有两个模糊集合 A 和 B , 对于所有的 $x \in X$, 均有 $\mu_A(x) \leq \mu_B(x)$, 则称模糊集合 A 包含于 B , 记作 $A \subseteq B$ 。

3. 模糊空集

若对所有 $x \in X$, 均有 $\mu_A(x) = 0$, 则称 A 为模糊空集, 记作 $A=\emptyset$ 。

4. 模糊集合的并集

若有三个模糊集合 A 、 B 和 C , 对于所有的 $x \in X$, 均有

$$\mu_C(x) = \mu_A(x) \vee \mu_B(x) = \max[\mu_A(x), \mu_B(x)] \quad (12-13)$$

则称 C 为 A 与 B 的并集, 记为 $C=A \cup B$ 。符号 \max 表示最大算子, 即取两个隶属度中的较大值, 常用符号“ \vee ”表示。

5. 模糊集合的交集

若有三个模糊集合 A 、 B 和 C , 对于所有的 $x \in X$, 均有

$$\mu_C(x) = \mu_A(x) \wedge \mu_B(x) = \min[\mu_A(x), \mu_B(x)] \quad (12-14)$$

则称 C 为 A 与 B 的交集, 记为 $C=A \cap B$ 。符号 \min 表示最小算子, 即取两个隶属度中的较小值, 常用符号“ \wedge ”表示。

6. 模糊集合的补集

若有二个模糊集合 A 和 B , 对于所有的 $x \in X$, 均有

$$\mu_B(x) = 1 - \mu_A(x) \quad (12-15)$$

则称 B 为 A 的补集。

7. 模糊集合的直积

若有二个模糊集合 A 和 B , 其论域分别为 X 和 Y , 则定义在积空间 $X \times Y$ 上的模糊集合 $A \times B$ 为 A 和 B 的直积, 其隶属函数为

$$\mu_{A \times B} = \min[\mu_A(x), \mu_B(y)] \quad (12-16)$$

或者

$$\mu_{A \times B}(x, y) = \mu_A(x)\mu_B(y) \quad (12-17)$$

并且两个模糊集合的直积的概念可以很容易地推广到多个集合中去。

12.1.4 模糊关系的定义及合成

模糊关系是普通关系的推广, 普通关系描述元素之间的关联关系, 而模糊关系则描述的是模糊前提与模糊结论之间的模糊倾向性。日常生活中我们经常听到诸如“ A 和 B 很相似”、“ X 比 Y 大很多”等描述模糊关系的语句。借助于模糊集合理论, 可以定量地来描述这

些模糊关系。

1. 模糊关系的定义

定义：设 X, Y 是两个非空集合，则直积

$$X \times Y = \{(x, y) \mid x \in X, y \in Y\} \quad (12-18)$$

中的一个模糊子集 R 称为 X 到 Y 的一个模糊关系，并且描述为

$$\mu_R: X \times Y \rightarrow [0, 1] \quad (12-19)$$

式中映射 μ_R 将集合 X 和集合 Y 的直积 $X \times Y$ 与模糊关系 R 联系起来。当论域 X 和 Y 都是有限集时，模糊关系可以用模糊矩阵来表示。设 $X = \{x_1, x_2, \dots, x_n\}$, $Y = \{y_1, y_2, \dots, y_n\}$ 时，模糊关系可用如下的 $n \times m$ 阶矩阵来表示。

$$R = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1m} \\ r_{21} & r_{22} & \cdots & r_{2m} \\ \vdots & \vdots & & \vdots \\ r_{n1} & r_{n2} & \cdots & r_{nm} \end{bmatrix} \quad (12-20)$$

矩阵中的元素 r_{ij} 表示 x_i 与 y_j 对于关系 R 的隶属程度。

上述定义的模糊关系又称为二元模糊关系，当论域为 n 个集合的直积 $X_1 \times X_2 \times \dots \times X_n$ 时，它对应的为 n 元模糊关系。一般情况下模糊关系都指的是二元模糊关系。

2. 模糊关系的合成

模糊关系的合成定义：设 X, Y, Z 是论域， R 是 X 到 Y 的一个模糊关系， S 是 Y 到 Z 的一个模糊关系，则 R 到 S 的合成 T 也是一个模糊关系，记为 $T = R \circ S$ ，它具有隶属度

$$\mu_{R \circ S} = \bigvee_{y \in Y} (\mu_R(x, y) \wedge \mu_S(y, z)) \quad (12-21)$$

其中“ \vee ”和“ \wedge ”分别表示并运算和交运算。这种方式合成也被称为最大最小合成，也是最常用的一种合成方式。

当论域 X, Y, Z 都是有限时，模糊关系的合成可用模糊矩阵的合成来表示。设 R, S, T 三个模糊关系对应的模糊矩阵分别为

$$R = (r_{ij})_{n \times m}, \quad S = (s_{jk})_{m \times l}, \quad T = (t_{ik})_{n \times l}$$

则有

$$t_{ik} = \bigvee_{j=1}^m (r_{ij} \wedge s_{jk}) \quad (12-22)$$

对于不能用模糊矩阵表达的模糊关系也可以用上述定义的最大最小运算进行合成运算。

12.1.5 语言变量与模糊推理

1. 语言变量

语言是人们进行思维和信息交流的重要工具。语言分为两种：自然语言和形式语言。日常人们所说的语言属于自然语言，其特点是语义丰富、灵活，同时具有模糊性。形式语言有严格的语法规则和语义，不存在任何的模糊性和歧义，计算机语言就是这样，如 C 语言等。带模糊性的语言称为模糊语言，如长、短、年轻、年老和极老等。

语言变量是自然语言中的词或句，它的取值不是通常的数，而是用模糊语言表示的模

糊集合。例如，若把“气温”作为一个模糊语言变量，则它的取值不是具体的温度，而是诸如“冷”、“凉”、“适宜”、“热”、“很热”等用模糊语言表示的模糊集合。

查德(Zadah)将语言变量定义如下：语言变量由一个 5 元体($x, T(x), U, G, M$)来表征。其中 x 是变量的名称； U 是 x 的论域； $T(x)$ 是语言变量值的集合，每个语言变量值是定义在论域 U 上的一个模糊集合； G 是语法规则，用以产生语言变量 x 的值的名称； M 是语义规则，用以产生模糊集合的隶属函数。

例如，若定义“气温”为语言变量，则 $T(\text{气温})$ 可能为

$$T(\text{气温}) = \{\text{冷}, \text{凉}, \text{暖}, \text{热}\}$$

上述每个模糊语言如“冷”、“凉”、“暖”等是定义在论域 U 上的一个模糊集合。设论域 $U = [-10, 35]^\circ\text{C}$ ，则可大致认为低于 5°C 为“冷”，大于 0°C 小于 15°C 为“凉”，依此类推。采用这些模糊集合的隶属函数如图 12-8 所示。

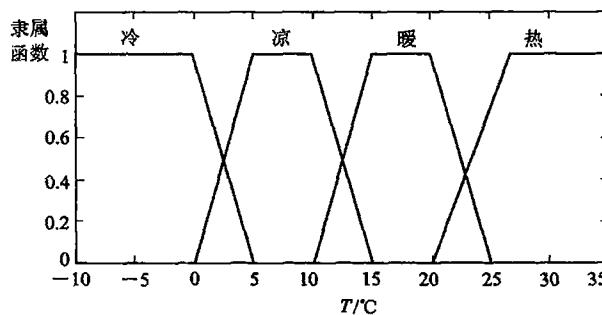


图 12-8 气温的隶属函数

2. 模糊推理

模糊逻辑的诱人之处在于它能够模拟人类大脑的推理机制。通过模糊逻辑能够很方便地将基于人类语言或专家的控制规则转变为模糊规则，从而借助于日益普及的计算机进行自动控制。

1) 模糊规则的基本形式

模糊规则由大量的 if … then 语句构成，它的基本形式如下：

if \tilde{u}_1 is \tilde{A}_1^1 and \tilde{u}_2 is \tilde{A}_2^2 and … and \tilde{u}_n is \tilde{A}_n^n , then \tilde{y}_1 is \tilde{B}_1^1 … and \tilde{y}_n is \tilde{B}_n^n

其中： $\tilde{u}_1, \tilde{u}_2, \dots, \tilde{u}_n$ 表示输入变量； $\tilde{y}_1, \tilde{y}_2, \dots, \tilde{y}_n$ 表示输出变量； $\tilde{A}_i = \{\tilde{A}_i^j : j = 1, 2, \dots, N_i\}$ 表示与每个输入变量对应的语言变量的语言值； $\tilde{B}_i = \{\tilde{B}_i^s : s = 1, 2, \dots, M_i\}$ 表示与输出变量相对应的语言变量的取值。

模糊规则的 if 部分通常被称为规则的前件(Premise)，then 部分则被称为后件(Consequent)。模糊推理过程中使用的规则库(Rule Base)就是由大量的这类规则构成的，规则库的大小根据所考虑问题的复杂程度差别很大。一般来说，当问题复杂时，规则库会很大。但是我们不能想当然地认为规则越多，问题求解的精度会越高。事实上，对输入空间划分过细会引起规则爆炸，此时推理过程变得极其缓慢烦琐，其也就失去了使用价值。

2) 模糊推理的一般过程

模糊推理是基于模糊规则，采用模糊逻辑由给定的输入到输出的映射过程。它包括如

下五个方面的内容：

- (1) 输入变量模糊化(Fuzzification)，即把确定的输入转化为由隶属度描述的模糊集。
- (2) 在模糊规则的前件中应用模糊算子(与、或、非)。
- (3) 根据模糊蕴含运算由前提推断结论。
- (4) 合成每一个规则的结论部分，得出总的结论。
- (5) 反模糊化(Defuzzification)，即把输出的模糊量转化为确定的输出。

其中在反模糊化中可以使用的常规方法有中心法、二分法等。有关模糊推理的具体过程和原理读者可参考相关文献。

12.2 模糊传感器系统

模糊传感器的研究最早出现在 20 世纪 80 年代末，它是模糊集合理论应用中发展较晚的一个领域，从此国内外的许多专家学者开始致力于这方面的研究，并且获得许多理论和应用成果。

E. Benoit 教授认为，模糊传感器必须依据数值测量关系，并且可以重新构造其结构，以适应不同的测量要求。D. Stipanicer 教授介绍了一种叫做“模糊眼”的模糊视觉传感器，对其原理和结构进行了研究，并将它成功地应用于一个位置控制系统，该系统可以捕捉、跟踪光源，实现手眼协同。H. Schodel 利用模糊集合理论，探讨了模糊传感器的非确定性信息传播、自校正、人机接口和语义划分等问题，并且利用模糊传感器对水中油污进行了测量。另外还有 Mauris、L. Foulloy 等人也都在模糊传感器方面取得了重要的研究成果。

对于模糊传感器系统而言，其测量结果的表示是一种基于语言符号化描述的符号测量系统。它是“数值测量”与“语言符号化表示”二者优势互补的一体化符号测量系统，是基于模糊集合理论实现数值/符号转换的一种智能传感器系统，目前已经成为测量领域的最新研究方向之一。

12.2.1 测量结果“符号化表示”的概念

传统测量旨在追求被测量与标准量(单位)的比值的数值准确性，以比值的数值与单位二者相联合来表示测量结果。其测量结果的表示是一种数值符号描述，也就是说，传统测量系统是测量结果的表示为一种数值符号描述的符号测量系统。这种测量系统对被测对象给以定量描述，具有精确性、严密性等诸多优点。传统的测量方法在人类的文明发展进程中发挥了巨大作用，并且今后仍然是绝对不可能被完全取代的。

随着科学技术的飞速发展，人们发现只进行传统的那种单纯的数值测量，其结果单纯以数值符号化来描述，在很多情况下都是不完备的。例如，由传统测量系统测得齿轮箱加速度为 $28g$ 时，需要进一步明确其所处的状态是“强振”、“中振”、“微振”还是“正常”，是否必须立即停机检修等。这种例子是非常多的，它表明测量结果的单纯数值表示是不完备的，需要在数值描述的基础上进一步给出对象所处状态的语言描述。语言描述是对被测对象进行的定性描述，是对被测对象内在变化规律更深刻的揭示，因此其难度是很大的。它往往需要进行多点多参数测量，经过具有丰富经验、深厚知识的专家分析、判断和推理之

后才能得到最终结果。由于其难度大，数值测量与语言符号化表示是分离进行的。如当测得养鱼池的水温是30℃，变压器中的油温是90℃时，需要根据知识与经验进一步进行推理判断，并给出它们的温度状态是“高”、“适中”或是“低”等形式的语言描述。

12.2.2 模糊传感器的基本概念和功能

1. 模糊传感器的基本概念

模糊传感器是将数值测量与语言符号表示二者相结合而构成的一体化符号测量系统，是在传统数值测量基础上进一步给出拟人类语言符号描述的智能传感器系统。其中的核心环节是数值/语言符号转换环节。实现数值/语言符号转换功能的方式有多种，即由数值域到语言域的映射关系有多种形式存在形式。模糊传感器系统基于模糊集合理论进行数值/语言符号转换。根据上述基本概念，模糊传感器的原理框图如图12-9所示。

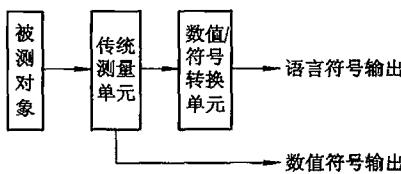


图12-9 模糊传感器原理框图

传统测量单元完成传统的数值测量，给出测量结果的数值符号描述；数值/符号转换单元是核心单元，它基于模糊集合理论来完成将测量数值结果转换为拟人类语言描述。

2. 模糊传感器的基本功能

作为一种新型的智能传感器，模糊传感器不但具备智能传感器的一般特点和功能，同时也具有自己独特的功能。

(1) 学习功能。模糊传感器的学习功能是其最重要的一种功能。例如模糊血压计，要使其直接反映出血压的“正常”和“不正常”，该模糊血压计首先要积累大量的反映血压正常的相关知识，其次还要将测量结果用人类所能接受的语言表达出来。从这个意义上讲，模糊血压计必须具备学习功能。

模糊传感器能够实现在专家指导下学习或者无需专家指导的自组织学习，并且能够针对不同的测量任务要求，选择合适的测量方案。从某种意义上来说，模糊传感器可以认为是一个完成特殊任务的小型专家系统。

(2) 推理功能。模糊传感器在接收到外界信息后，可以通过对人类知识的集成而生成的模糊推理规则实现传感器信息的综合处理，对被测量的测量值进行拟人类自然语言的表达等。对于模糊血压计来说，当它测到一个血压值后，首先通过推理，判断该值是否正常，然后用人类理解的语言，即“正常”或“不正常”来表达。为实现这一功能，推理机制和知识库(存放基本模糊推理规则)是必不可少的。

(3) 感知功能。模糊传感器与传统传感器一样可以感知敏感元件确定的被测量，但是模糊传感器不仅可以输出数量值，而且可以输出易于人类理解和掌握的自然语言符号量，这是模糊传感器的最大特点。

(4) 通信功能。模糊传感器具有自组织能力，不仅可以进行自检测、自校正、自诊断

等，而且可以与上级系统进行信息交换。

12.2.3 模糊传感器的结构

1. 基本逻辑结构

基于模糊测量原理，模糊传感器的基本逻辑结构由信号提取、信号处理、数值/符号转换和模糊概念合成四部分组成，如图 12-10 所示。

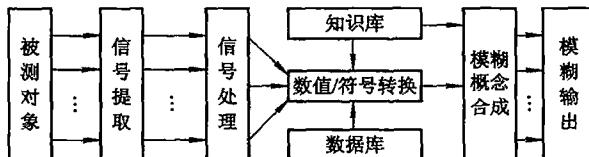


图 12-10 模糊传感器的基本逻辑结构

(1) 信号提取模块基于普通敏感探头获取被测物理量的参量值，完成待测量系统的信号检测任务。

(2) 信号处理模块的基本处理任务有三个：其一是对信号进行放大、滤波；其二是基于多传感器多信息融合算法，获得高选择性、高稳定性的测量值；其三是进行模/数转换，与计算机之间进行信息传输。

(3) 数值/符号转换单元是实现模糊测量的核心，由计算机完成。数值/符号转换单元是模糊传感器实现模糊测量的关键技术。模糊传感器测量被测物理量的准确性很大程度上取决于知识库与数据库。知识库包含两方面的内容：一方面为模糊集隶属函数的知识以及确定元素属于模糊集合的隶属度，另一方面为模糊蕴涵推理规则。通常意义上，这一环节是由该领域的专家来完成的。但是不容忽视的是，专家的研究成果和丰富经验往往不易以严格的规则形式描述。另外，如果所描述的模糊蕴涵关系复杂，则需要相当规模的模糊推理规则，实现模糊推理的运算量相当大。

(4) 模糊概念合成模块根据知识库和数值/符号转换单元的输出进行模糊推理和合成，得出正确的拟人类语言测量结果。

2. 基本物理结构与软件结构

根据上述模糊传感器的基本逻辑结构，可以设计出如图 12-11 和图 12-12 所示的一种模糊传感器基本物理结构与软件结构。

(1) 信息提取单元负责将与被测对象有关的测量信息通过数据采集电路输入计算机，其中包括作用在被测对象上的干扰信息。

(2) 信息交换单元主要进行人机交互、录入专家知识以建立规则库，同时通过系统总线提供通信接口，这些是模糊智能传感器系统所必须具备的功能。

(3) 测量前，计算机处理单元首先根据从信息交换单元得到的领域专家知识，经过规则学习模块、隶属函数生成模块和模糊蕴涵规则库等处理过程，生成进行模糊测量所必需的隶属函数和相应的模糊蕴涵规则库。测量时，首先通过信息提取单元得到测量值，然后根据前面建立的模糊蕴涵规则库，经过数值处理模块、数值/符号转换模块和模糊推理模块的处理，最后得到模糊化的语言符号测量结果。为提高传感器的稳定性和选择性，测量

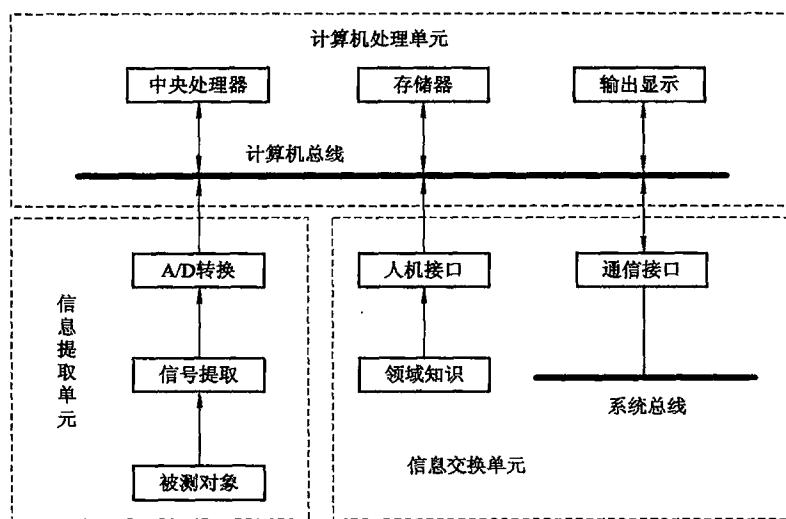


图 12-11 模糊传感器的基本物理结构

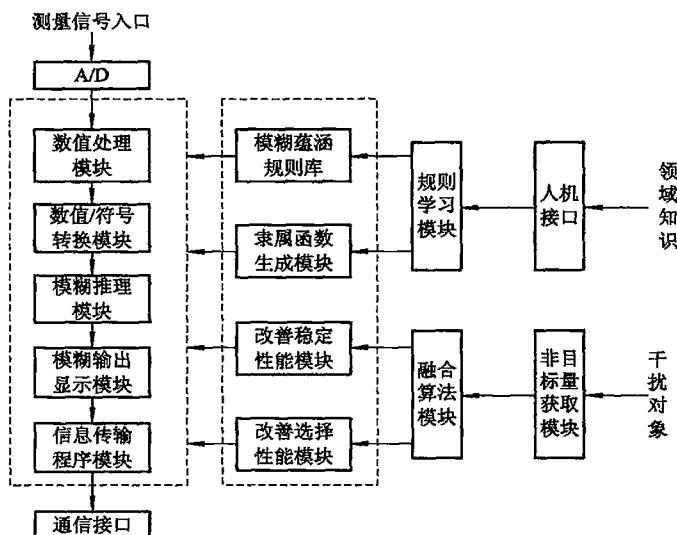


图 12-12 模糊传感器的基本软件结构

时还要进行多个非目标参量(即干扰对象)的同步测量,然后经过多传感器数据融合算法模块、改善稳定性能模块和改善选择性能模块等处理,得到高质量的数值测量结果。根据测量的需要,模糊传感器的测量结果还可以通过信息传输程序模块进行传输,实现信息共享。

根据被测量的多少,模糊传感器可以分为一维模糊传感器和多维模糊传感器。一维模糊传感器结构通常比较简单,仅有一个被测量的被测对象。但是,绝大多数情况下,被测对象不仅所处环境复杂,易受多种干扰,而且有时需要测量的几个被测量相互关联,需要

采用几个传感器同时测量，从而构成多维模糊传感器。显然一维模糊传感器是多维模糊传感器的特例。多维模糊传感器的测量相对比较复杂，通常要采用基于多传感器的信息融合算法进行数据处理，在这方面已经有不少成熟的理论和方法，而且还在不断地完善之中。

12.2.4 模糊传感器语言描述的产生方法

如前所述，模糊传感器的作用是提供数值测量的语言描述，因此产生语言描述是模糊传感器的重要功能之一。产生语言描述的常用方法有两种：一种是运用语言间的语义关系产生概念而后产生相应的语言描述；另一种则是根据特定点测量值的语言描述，通过插值的方法产生其它测量点的语言描述。

1. 通过语义关系产生概念

模糊传感器可以输出多个语言描述，这些语言描述通过它们语义间的关系相联系。以一个简单的温度测量系统为例。语言描述“热”(hot)和“很热”(very-hot)间的语义关系可归因于语言域 Y 上的顺序关系，该关系又同数值域 N 上的大小关系相对应，并表示为

$$\text{hot} \leqslant \text{very-hot}$$

所有概念间的关系由传感器自身管理。首先，我们要定义一个特殊概念，称为属概念 (Generic Concept)。所谓属概念，是指对应于数值域中那些最具有代表性的测量点或测量范围的语言描述。譬如，电冰箱的温度通常保持在 $-5 \sim 15^{\circ}\text{C}$ 范围内，那么可认为 $0 \sim 5^{\circ}\text{C}$ 为最适宜的温度范围，而 $0 \sim 5^{\circ}\text{C}$ 在语言域中用“适中”语言概念来描述。于是我们可定义“适中”这个语言概念为属概念。

此外，产生新概念还需要给出其它语言描述和含义。Benoit 教授定义了稍高(more-than)、稍低(less-than)、高(above)、低(below)等模糊算子来产生模糊顺序分度，以此来产生新概念。如定义属概念为“适中”，根据上述模糊算子可产生新概念“热”、“很热”、“冷”、“很冷”，表示为

适中(mild)(属概念)

热(hot)=more-than(mild)

很热(very-hot)=above(hot)

冷(cold)=less-than(mild)

很冷(very-cold)=below(cold)

这五个概念的隶属函数形式如图 12-13 所示。当然，还可以用其它形式表示各概念的隶属函数，如三角形、柯西形等。使用中如果模糊传感器已将这些隶属函数存储在其数据库中，则只需要修改属概念参数就可以自动修正其形状，使其符合测量要求。

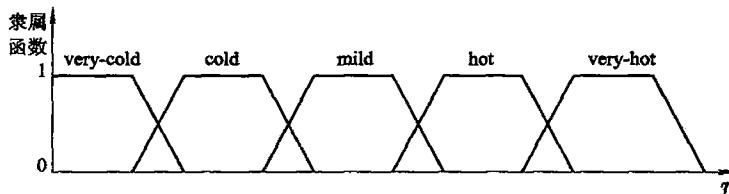


图 12-13 根据语义关系产生概念的隶属函数

以温度测量为例来说明新语言概念的产生过程。设论域 $U = [0, 1]$ 表示温度测量归一化处理后的范围, 语言域 $S = \{\text{非常冷}, \text{冷}, \text{热}, \text{非常热}\}$, 那么, 产生新概念的实质在于确定语言域 S 中新生概念相应的隶属函数。

首先, 定义属概念为“冷”(用 c_1 表示)和“热”(用 c_2 表示), 其相应的隶属函数为

$$\text{对于 } x \in U, \mu_R(c_2, x) = a \quad (12-23)$$

$$\text{则 } x \in U, \mu_R(c_1, x) = 1 - a \quad (12-24)$$

上式表明, 如果对于论域上的元素 x 隶属于“热 c_2 ”的程度(用 $\mu_R(c_2, x)$ 表示)为 a 的话, 那么, 它隶属于“冷 c_1 ”的程度(用 $\mu_R(c_1, x)$ 表示)必然为 $1 - a$ 。这种关系可用图 12-14 表示。

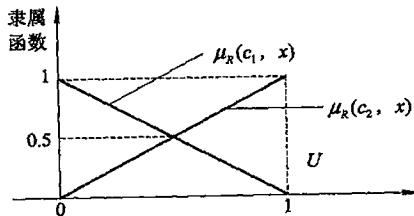


图 12-14 属概念 c_1 、 c_2 隶属函数曲线

属概念及其隶属函数确定后, 就可通过模糊算子产生新的模糊概念。我们定义“非常”(very)模糊算子, 则 $\text{very}(c_1)$ 表示“非常冷”, 而 $\text{very}(c_2)$ 表示“非常热”。因此 $\{\text{very}(c_1), c_1, c_2, \text{very}(c_2)\}$ 构成了论域 U 上基于属概念 c_1 和 c_2 的新的语言域。我们把 x 隶属于新生概念“ $\text{very}(c_1)$ ”和“ $\text{very}(c_2)$ ”的程度, 即隶属函数 $\mu_R(\text{very}(c_1), x)$ 和 $\mu_R(\text{very}(c_2), x)$ 表示为属概念隶属度 $\mu_R(c_1, x)$ 和 $\mu_R(c_2, x)$ 的函数形式, 写成下列关系式:

$$\mu_R(\text{very}(c_1), x) = f[\mu_R(c_1, x)] \quad (12-25)$$

$$\mu_R(\text{very}(c_2), x) = f[\mu_R(c_2, x)] \quad (12-26)$$

这里显然有:

$$\text{若 } \mu_R(c_1, x) > 0.5, \text{ 则 } \mu_R(\text{very}(c_1), x) < \mu_R(c_1, x) \quad (12-27)$$

$$\text{若 } \mu_R(c_1, x) < 0.5, \text{ 则 } \mu_R(\text{very}(c_1), x) > \mu_R(c_1, x) \quad (12-28)$$

在满足上述条件下, 可选择函数形式为

$$f(\xi) = \xi(1 - \sin(k\pi(\xi - 0.5))) \quad (12-29)$$

式中 ξ 为属概念隶属函数, k 为修正因子, 满足 $0 < k < 1$ 。

至此, 在论域 U 上基于属概念“冷”和“热”生成了两个新概念“非常冷”和“非常热”。显然新概念如果不符合测量要求, 就需要通过适当的训练算法修正隶属函数。

总之, 通过语义关系产生新概念的方法可描述为: 首先定义属概念及其隶属函数; 其次, 利用存储在模糊传感器中的模糊算子, 产生新的模糊概念; 最后, 利用属概念隶属函数得出新生概念的隶属函数。如果新概念不符合测量要求, 则可通过训练算法修正其隶属函数, 直至满足要求为止。

2. 通过插值法产生新概念

插值法产生新概念的原理如下: 对于数值域中特定的元素, 我们称之为特征测量量 (Characteristic Measurements), 用 v_i 来表示。对于每个 v_i 其数值域模糊集合表示为

$F(v_i)$, 则 v_i 隶属于 $F(v_i)$ 的程度等于 1, 即 $\mu_{F(v_i)}(v_i) = 1$; 而其它的特征测量量用 $v_j (j \neq i)$ 表示, 其数值域模糊集合表示为 $F(v_j)$, 显然, 其隶属于模糊集合的程度为 0, 即 $\mu_{F(v_j)}(v_i) = 0$ 。那么对于任意一点 $v \in [v_i, v_j]$, 其隶属于模糊集合 $F(v_i)$ 和 $F(v_j)$ 的隶属度分别为

$$\mu_{F(v_i)}(v) = \frac{d(v_i, v)}{d(v_i, v_j)} \quad (12-30)$$

$$\mu_{F(v_j)}(v) = \frac{d(v_j, v)}{d(v_i, v_j)} \quad (12-31)$$

它们之间关系满足

$$\mu_{F(v_i)}(v_i) = 1, \quad \mu_{F(v_i)}(v_j) = 0 \quad (12-32)$$

$$\mu_{F(v_j)}(v_i) = 0, \quad \mu_{F(v_j)}(v_j) = 1 \quad (12-33)$$

其中式(12-30)、式(12-31)两式中 d 表示两点距离, 如图 12-15 所示。该距离应当满足下述条件

$$\mu_{F(v_i)}(v) + \mu_{F(v_j)}(v) = 1 \Leftrightarrow d(v_i, v) + d(v_j, v) = d(v_i, v_j) \quad (12-34)$$

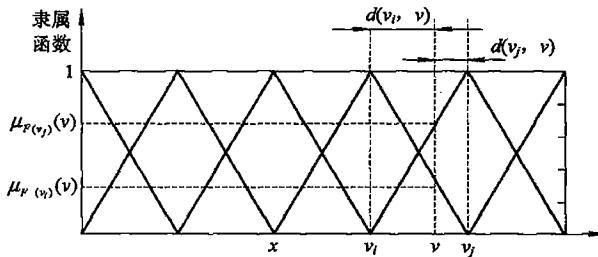


图 12-15 插值法产生概念示意图

最简单的距离可表示为

$$d(v_i, v) = |v_i - v| \quad (12-35)$$

如果隶属函数的形状已知, 则可定义更一般的形式

$$\begin{aligned} \forall x \in [v_i, v_j], \quad \mu_{F(v_i)}(x) &= f[d(v_j, v)/d(v_i, v_j)] \\ \mu_{F(v_j)}(x) &= f[d(v_i, v)/d(v_i, v_j)] \end{aligned} \quad (12-36)$$

f 是 $[0, 1]$ 上的增函数, 且满足

$$f(0) = 0, \quad f(1) = 1 \quad \text{且} \quad f(1-a) = 1-f(a) \quad (12-37)$$

如函数 $f(x)=3x^2-2x^3$ 满足上述条件, 那么其隶属函数形式如图 12-16 所示。

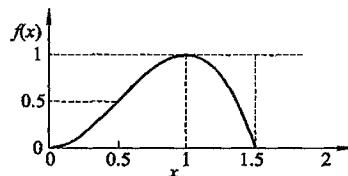


图 12-16 函数 $f(x)$ 隶属函数示意图

另外要提及的是, 通过定义特征测量量来产生新语言值的这种方法可以应用到多维模糊测量中去。

12.2.5 模糊传感器对测量环境的适应性

在温度测量问题中，模糊传感器可将 $T=15^{\circ}\text{C}$ 解释为“冷”，但是按照不同的背景，这个描述可能是不确切的。例如在游泳池中这个温度可能是冷的，但在电冰箱中它又是很热的。因此有必要根据不同的背景知识来修正模糊传感器的输出。

通常，在考虑测量背景知识时，应对数值测量描述进行适应性处理。有两种处理方法比较适用：一种是基于适应函数的处理方法；另一种是基于专家定性学习的方法。

1. 基于适应函数的处理方法

假设定义属概念和产生新概念的数值域 N 叫做标准数值域(Ideal Universe)，而实际测量数值域用 N' 表示，那么 N 和 N' 之间的不一致性可用适应函数 h 来修正。从 N 到 N' 的映射如图 12-17 所示。

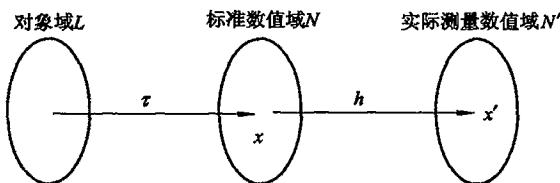


图 12-17 适应函数示意图

对于一个确定的测量对象，其数值测量可描述为

$$L = \tau(h(x)) \quad x \in \tau^{-1}(L) \quad (12-38)$$

式中： τ 即 $L \rightarrow x$ （表示对象域 L 到数值域 x 的映射）； τ^{-1} 即 $x \rightarrow L$ （表示数值域 x 到对象域 L 的逆映射）。

为了实现这种处理方法，通常要把适应函数(Adaptive Function) h 存放在模糊传感器的知识库中。

另外，与属概念对应的特征测量点 M_c 不应随适应函数的变化而变化，表示为 $h(M_c)=M_c$ ，而且对特征测量点应保持线性，表示为

$$h'(M_c) = k \quad (12-39)$$

式中： $h'(M_c)$ 表示适应函数的导数， k 为常数。

选择一个较为理想的适应函数是相当复杂的，有兴趣的读者可查阅有关资料。

2. 专家指导下的定性学习方法

模糊传感器的学习功能是通过比较专家和模糊传感器对同一被测量定性描述的差异来实现的。设对于同一被测量 x ，专家给出的语言描述表示为 $l(x)$ ，模糊传感器输出的语言描述表示为 $l'(x)$ ， e 表示 $l(x)$ 和 $l'(x)$ 之间的定性差异，则修正规则如下：

- (1) 若 e 为正向定性差异，表示为 $l(x) > l'(x)$ ，则可通过增加模糊算子调整该差异；
- (2) 若没有定性差异，表示为 $l(x) = l'(x)$ ，则模糊算子不变；
- (3) 若 e 为负向定性差异，表示为 $l(x) < l'(x)$ ，则可通过减小模糊算子来调整该差异。

上述“增加”、“减小”、“不变”均指模糊算子。通常，增加算子是指将隶属函数曲线向

数值量小的方向平移或扩展；减少算子是指将隶属函数曲线向数值大的方向平移或扩展；而算子不变是指隶属函数保持不变。

对于温度 T_0 ，学习前描述为

$$I(T_0) = \{0.6 / \text{冷}, 0.4 / \text{适中}\}$$

学习后可以得到下列描述：

$$I(T_0) = \{0.4 / \text{冷}, 0.6 / \text{适中}\}$$

增加模糊算子的动作示意如图 12-18 所示。

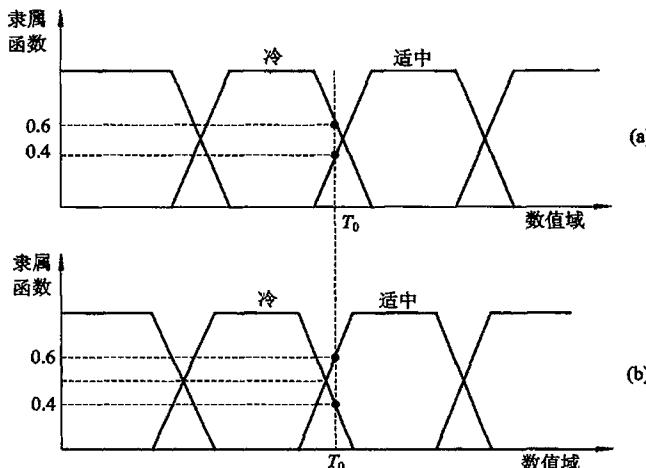


图 12-18 增加模糊算子示意图

(a) 学习前; (b) 学习后

训练样本可由专家经验知识确定。专家经验知识的获取，即专家信息的输入可由下列步骤实现：

- (1) 确定测量范围的上下限；
- (2) 确定论域 U 上描述被测量数量值的个数；
- (3) 确定表征每个被训练概念(包括属概念和新概念)的模糊子集；
- (4) 通过采样输入对应被训练概念(包括属概念和新概念)隶属度为 1 的采样值；
- (5) 通过相关训练算法产生被训练概念(包括属概念和新概念)对应的隶属度。

12.2.6 模糊传感器隶属函数的训练算法

语言概念产生之后，还必须对这些概念的隶属函数进行训练，以符合人类对该概念的描述。另外，为了增强模糊传感器适应不同测量要求的能力，也必须对生成概念的隶属函数进行训练。原则上讲，隶属函数曲线通过训练可以调整到任意形状。然而，有时这并非必要且实现较难，通常可将隶属函数分为连续隶属函数和分段隶属函数两种情况。

1. 连续隶属函数训练方法

这里以柯西形隶属函数为例进行语言概念训练。训练最终体现在对式(12-29)中修正因子 k 的调整上。

在论域 $U=[0, 1]$ 上生成语言概念集合 $S=\{\text{very}(c_1), c_1, c_2, \text{very}(c_2)\}$, 每个概念在数值域上所对应的数值范围如图 12-19 所示。

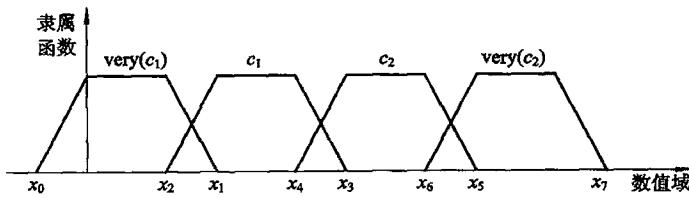


图 12-19 $\text{very}(c_1)$ 、 c_1 、 c_2 、 $\text{very}(c_2)$ 在数值域上对应的数值范围示意图

其训练步骤如下：

(1) 对“ $\text{very}(c_1)$ ”概念的训练。设训练样本 $x \in [x_0, x_1]$ 。

① 由经验曲线或前次训练后生成的曲线计算 $\mu_R(\text{very}(c_1), x)$ 和 $\mu_R(c_1, x)$ 。

② 如果 $\mu_R(\text{very}(c_1), x) \geq \mu_R(c_1, x)$,

即该语言概念 $\text{very}(c_1)$ 同训练样本的状态是相符的。这时若 $\Delta = x_{01} - x$ 小于一给定阈值 δ , 则该语言概念的修正可以结束, 否则增加式(12-29)中的修正因子 k , 并转到第①步。其中 x_{01} 为 $\text{very}(c_1)$ 与 c_1 两概念的隶属函数曲线的交点对应论域上的值, 如图 12-20 所示。

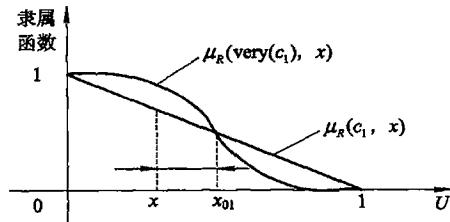


图 12-20 新概念隶属函数

③ 如果 $\mu_R(\text{very}(c_1), x) < \mu_R(c_1, x)$,

即该语言概念 $\text{very}(c_1)$ 同训练样本的状态是不相符的, 则应减小式(12-29)中的修正因子 k , 并且转到第①步。

(2) 对“ c_1 ”概念的训练。由于概念 c_1 介于概念 $\text{very}(c_1)$ 和 c_2 之间, 故对 c_1 的训练涉及到 $\text{very}(c_1)$ 和 c_2 , 因此首先要计算 c_1 的重心 g_1 , 并以此点为界分左右两端训练概念 c_1 。设训练样本 $x \in [x_2, x_3]$ 。

① 计算概念 c_1 的重心 g_1 :

$$g_1 = \frac{\int \mu_R(c_1, x)x \, dx}{\int \mu_R(c_1, x) \, dx} \quad (12-40)$$

② 如果 $x \leq g_1$, 则计算 $\mu_R(c_1, x)$, $\mu_R(\text{very}(c_1), x)$ 。若

$$\mu_R(\text{very}(c_1), x) \leq \mu_R(c_1, x) \quad (12-41)$$

则语言概念 c_1 同训练样本 x 的状态是一致的。这时若 Δ 大于一给定阈值 δ , 则应当增大式(12-29)中的修正因子 k , 转到①, 否则结束训练。

$$\text{如果 } \mu_R(\text{very}(c_1), x) > \mu_R(c_1, x) \quad (12-42)$$

此时该语言概念 c_1 同训练样本 x 的状态不一致, 应当增大式(12-29)中的修正因子 k , 转到①。

③ 如果 $x \geq g_1$, 计算 $\mu_R(c_1, x)$, $\mu_R(c_2, x)$ 。如果

$$\mu_R(c_1, x) \geq \mu_R(c_2, x) \quad (12-43)$$

此时该语言概念 c_1 同训练样本 x 的状态是相符的，则训练结束。

$$\text{如果 } \mu_R(c_1, x) < \mu_R(c_2, x) \quad (12-44)$$

此时该语言概念 c_1 同训练样本 x 的状态是不相符的，则增大式(12-29)中的修正因子 k ，转到①。

(3) 对“ c_2 ”和“very(c_2)”概念的训练。概念“ c_2 ”和“very(c_2)”的训练方法同 very(c_1) 和 c_1 的训练方法相同，此处不再重复。

2. 分段隶属函数训练方法

这里依然以温度测量为例，采用梯形隶属函数。设在论域 U 上，语言域 $S = \{\text{冷}, \text{温}, \text{热}\}$ ，其分段隶属函数如图 12-21 所示。

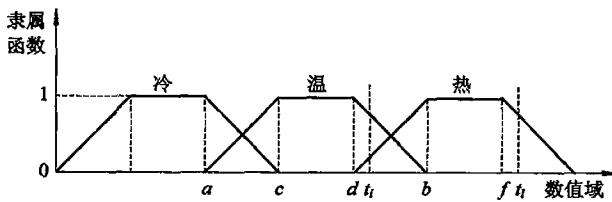


图 12-21 分段隶属函数

首先定义概念“温”为属概念，其余概念通过语言关系产生。

1) 使 t_i 属于“温”概念的训练

首先利用给定训练样本值 $\{t_i, i=1, 2, \dots, n\}$ ，按图 12-21 所示的分段隶属函数计算隶属度。

(1) 当 $c < t_i < d$ 时，表示由隶属函数所确定的概念“温”与语义概念“温”相符，则不修正。

(2) 若 $t_i > d$ ，且 $\mu_{\text{温}}(t_i) < \mu_{\text{热}}(t_i)$ ，表明此时隶属函数所确定的概念为“热”，与语义概念“温”不符合，则需修正隶属函数。

① $d^{(j)} = k(t_i - d^{(j-1)}) + d^{(j-1)}$ ，其中 $d^{(j)}$ 为速度修正因子。若 $k=0$ ， $d^{(j)} = d^{(j-1)}$ ，表示不修正；若 $k=1$ ，则 $d^{(j)} = t_i$ 表示修正最大。 $d^{(j)}$ 表示端点 d 的第 j 次修正值。

②

$$c^{(j)} = \begin{cases} c^{(j-1)} + \Delta^{(j)}, & \mu_{\text{温}}(c^{j-1} - \Delta^j) > 0.5 \\ c^{(j-1)}, & \text{其它} \end{cases} \quad (12-45)$$

其中： $\Delta^{(j)} = k'(t_i - d^{(j-1)})$ ，表示修正 d 对 c 产生影响的相关因子， $0 < k' < k$ 为偏移修正因子； $c^{(j)}$ 表示点 c 的第 j 次修正值。

(3) 若 $t_i > d$ ，且 $\mu_{\text{温}}(t_i) \geq \mu_{\text{热}}(t_i)$ ，则不修正。

(4) 若 $t_i < c$ ，且 $\mu_{\text{温}}(t_i) \geq \mu_{\text{冷}}(t_i)$ ，修正方法同上。

2) 使 t_i 属于“热”概念的训练

利用给定的训练样本值 $\{t_l, l=1, 2, \dots, m\}$ ，按图 12-21 分段隶属函数计算隶属度。

(1) 若 $t_l > b$ ，不必修正。

(2) 若 $t_l < b$ ，且 $\mu_{\text{温}}(t_l) > \mu_{\text{热}}(t_l)$ ，则

$$b^{(j)} = \begin{cases} 2t_l - d, & \mu_{\text{温}}(t_l) > 0.5 \\ b^{(j-1)}, & \text{其它} \end{cases} \quad (12-46)$$

(3) 若 $t_l < b$, 且 $\mu_{\text{温}}(t_l) < \mu_{\text{热}}(t_l)$, 则不必修正。

(4) 若 $b < t_l < f$, 则不必修正。

利用给定训练样本值对“冷”概念的训练同上, 此处不再赘述。

12.3 示例

12.3.1 [示例 12-1] 模糊温度测试仪

要求: 模糊温度测试仪是在传统的温度测量基础上, 利用模糊测量原理构造的新型智能传感器系统, 用于对培育豆芽菜的温室温度进行测量。豆芽菜生长时的最低温度为 10℃, 最适宜温度为 21~27℃, 最高温度为 28~30℃, 不宜超过 32℃。该测试仪的具体要求如下:

- 能对[0, 40]℃范围内的温室温度自动给出相应的语言描述: “温度过低”、“温度偏低”、“温度正常”、“温度偏高”或“温度过高”;
- 能对温室温度变化和当前温度分别进行曲线显示与数字量显示。

在该系统中, 采用模糊测量的方法对培育豆芽菜温室的温度进行测量, 并对当前温度进行拟人类语言显示。整个温度测试系统的系统框图如图 12-22 所示。

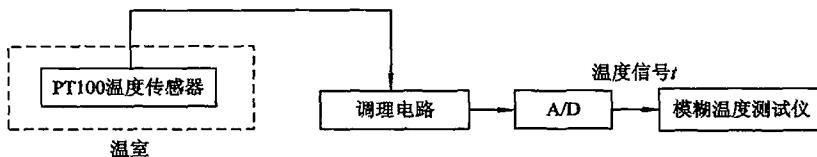


图 12-22 温度测试系统的系统框图

整个系统由温度传感器、调理电路、数据采集卡 A/D、模糊温度测试仪组成。温度传感器将所测的温室温度 t 通过调理电路和数据采集卡 A/D 进行转换并送入模糊温度测试仪, 模糊温度测试仪根据模糊原理对温室当前温度进行判断, 并对温室温度变化曲线和当前温度的数字量进行显示。

1. 模糊温度测试仪工作原理

(1) Matlab 中的梯形隶属函数 trapmf。

格式: $y = \text{trapmf}(x, [a \ b \ c \ d])$

说明: 梯形曲线可由四个参数 a, b, c, d 确定, 参见公式(12-11)。式(12-11)也可以更紧凑地表示为

$$f(x, a, b, c, d) = \max\left(\min\left(\frac{x-a}{b-a}, 1, \frac{d-x}{d-c}\right), 0\right) \quad (12-47)$$

参数 a 和 d 确定梯形的“脚”, 而参数 b 和 c 确定梯形的“肩膀”。

示例：

```
x=0:0.1:10;
y=trapmf(x, [1 4 6 8]);
plot(x, y);
xlabel('trapmf, p=[1 4 6 8]');
```

(2) 隶属函数 $\mu_N(t)$ 、 $\mu_M(t)$ 、 $\mu_H(t)$ 的确定。培育豆芽菜的温室温度值的集合 $U = \{t | 0^\circ\text{C} \leq t \leq 40^\circ\text{C}\}$ 作为论域，论域中任一元素 t (如为 25°C) 将以不同的程度隶属于五个模糊子集 NL(温度过低)、NS(温度偏低)、ZO(温度正常)、PS(温度偏高)、PL(温度过高)。温室温度的模糊集合 $T(t) = \{ \text{NL}, \text{NS}, \text{ZO}, \text{PS}, \text{PL} \}$ 中的五个模糊子集 NL、NS、ZO、PS、PL 分别由隶属函数 $\mu_{NL}(t)$ 、 $\mu_{NS}(t)$ 、 $\mu_{ZO}(t)$ 、 $\mu_{PS}(t)$ 、 $\mu_{PL}(t)$ 确定，隶属函数采用梯形隶属函数。并且根据专家以及豆芽菜生长的实际温度情况，可以确定上述五个模糊子集对应的隶属函数参数，如表 12-1 所示，其相应的温度 t 的隶属函数如图 12-23 所示。

表 12-1 与 $\mu_{NL}(t)$ 、 $\mu_{NS}(t)$ 、 $\mu_{ZO}(t)$ 、 $\mu_{PS}(t)$ 、 $\mu_{PL}(t)$ 对应的梯形隶属函数参数值

参数值 ${}^\circ\text{C}$ 模糊子集	a	b	c	d
NL	0	0	8	10
NS	8	10	20	21
ZO	20	21	27	28
PS	27	28	30	32
PL	30	32	40	40

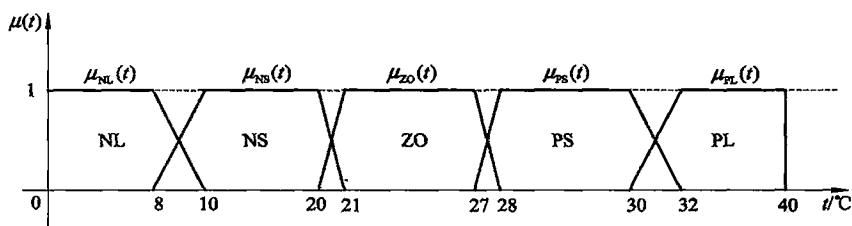


图 12-23 $\mu_{NL}(t)$ 、 $\mu_{NS}(t)$ 、 $\mu_{ZO}(t)$ 、 $\mu_{PS}(t)$ 和 $\mu_{PL}(t)$

(3) 模糊推理的实现与最大隶属度的求取。这部分的实现过程如图 12-24 所示，图中变量 temp_Negative Large、temp_Negative Small、temp_Zero、temp_Positive Small、temp_Positive Large 分别对应于隶属度 $\mu_{NL}(t)$ 、 $\mu_{NS}(t)$ 、 $\mu_{ZO}(t)$ 、 $\mu_{PS}(t)$ 、 $\mu_{PL}(t)$ ；最大隶属度 index 的取值定义为 {0, 1, 2, 3, 4}，用于指示相应的拟人类语言输出集合 {"温度过低"、"温度偏低"、"温度正常"、"温度偏高"、"温度过高"}。这样，当由控件键入温度值 t_1 后，首先计算温度 t_1 对于五个模糊集的隶属度 $\{\mu_{NL}(t_1), \mu_{NS}(t_1), \mu_{ZO}(t_1), \mu_{PS}(t_1), \mu_{PL}(t_1)\}$ ，然后利用模糊集合的基本运算中的并运算(由函数 max() 实现)求取 $\{\mu_{NL}(t_1), \mu_{NS}(t_1), \mu_{ZO}(t_1), \mu_{PS}(t_1), \mu_{PL}(t_1)\}$ 中的最大值，即求取最大隶属度，同时置标志 index。

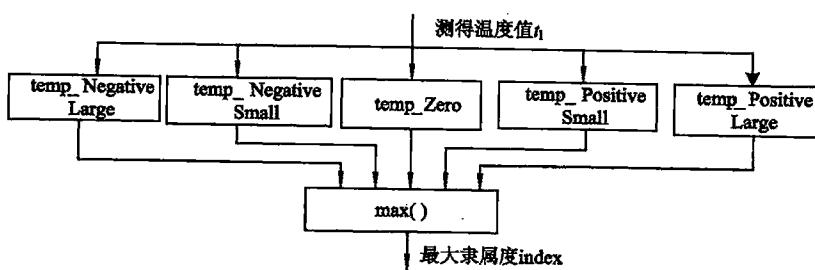


图 12-24 模糊推理的实现与最大隶属度的求取

(4) 语言符号输出。这是模糊测量的最后步骤, 根据上一步得到的最大隶属度 index 的取值 {0, 1, 2, 3, 4}, 确定输入温度所属的模糊集合, 从前述已经定义好的拟人类语言输出集合 {“温度过低”、“温度偏低”、“温度正常”、“温度偏高”、“温度过高”} 中找到对应的语言输出值。

2. 模糊温度测试仪的设计步骤

(1) 面板设计。启动 LabVIEW, 进入仪器开发环境, 创建如图 12-25 所示的面板。

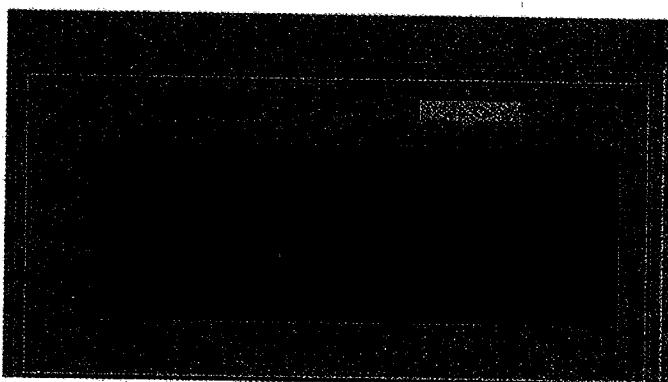


图 12-25 模糊温度测试仪面板

① 放置数字显示控件 1: 执行 All Controls→Numeric→Numeric Indicator 操作。

该控件为输出显示型控件, 用于输出温度传感器所测得的温室当前温度值, 标记(Label)为“温度”。

② 放置字符串显示控件 2: 执行 All Controls→String&Table→String Indicator 操作。

该控件为输出显示型控件, 用来输出温室温度的状态, 有 5 种可能的输出结果: “温度过低”、“温度偏低”、“温度正常”、“温度偏高”、“温度过高”。标记(Label)为“温度状态显示”。

③ 放置图形显示控件 3: 执行 All Controls→Graph→Waveform Chart 操作。

该控件为输出显示型控件, 用来显示温室的温度变化曲线。

(2) 流程图设计。

模糊温度测试仪主要是对温室当前的温度进行测量, 得出温室当前的温度值; 对测得

的温度进行模糊推理判断，得出温室的温度状态，并进行拟人类语言显示。具体程序设计如下：

- ① 打开流程图编辑窗口“Diagram”。
- ② 温度采集。将传感器经调理电路、数据采集卡采集入计算机的电压数据，经过刻度转换成温度值，并在前面板上显示温度。
 - 用 LabVIEW 软件里面的 AI Acquire Waveform. vi(调用路径为 All Functions→NI Measurements→Data Acquisition→Analog Input→AI Acquire Waveform. vi)获得采样信号，其中设置设备号为 1，通道号(channel)为 0；
 - 采集 1000 组电压信号，输出其平均值；
 - 进行电压信号和 PT100 铂电阻阻值的转换；
 - 利用曲线拟合实现 PT100 铂电阻阻值和温度的刻度转换；
 - 显示转换后的温度。

整个温度采集程序如图 12-26 所示。

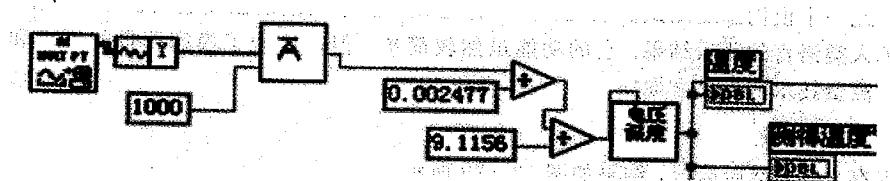


图 12-26 模糊温度测试仪中的温度采集程序

- ③ 温度模糊判断。本模块对刻度转换后的温度值进行模糊判断，它有五个子 VI，分别管理 5 个模糊域，当温度值处于其中一个模糊域或处于这个模糊域的隶属度大于其他模糊域时，这个模糊域就生效，会在前面板上显示这个模糊域相应的文字符号。模糊判断程序如图 12-27 所示。

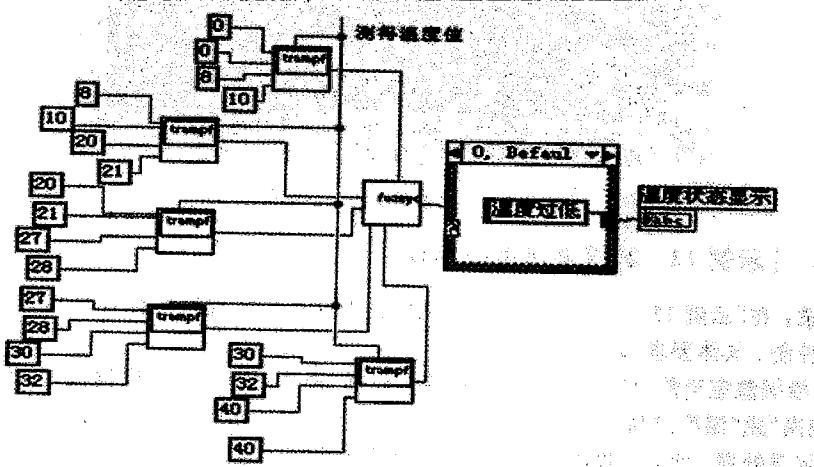


图 12-27 模糊温度测试仪中的温度模糊判断程序

- ④ 连线。完成编辑的流程图如图 12-28 所示。

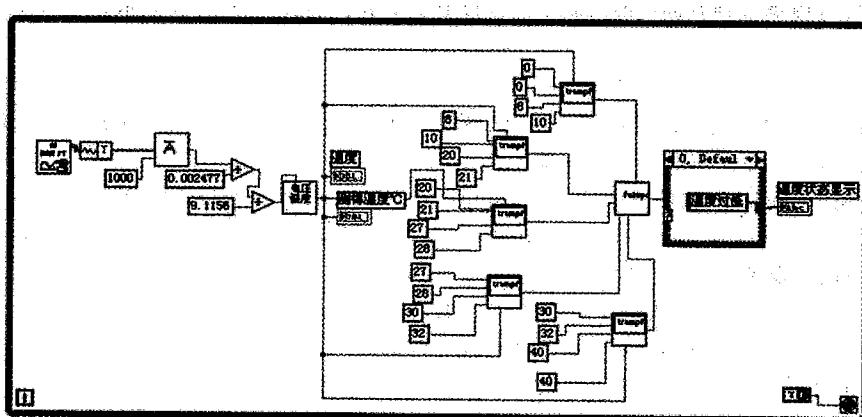


图 12-28 模糊温度测试仪程序图

至此一个模糊温度测试仪就完成了。它能够对温室当前的温度进行测量，并给出数字式和拟人类语言的测量结果。它的功能虽然较简单，但却体现了模糊测量的原理和方法，代表了测量技术的发展方向。

3. 运行检验

保存文件，然后运行，结果如图 12-29 所示。



图 12-29 模糊温度测试仪运行结果

12.3.2 [示例 12-2]模糊温度控制仪

要求：在[示例 12-1]所完成的基本功能的基础上，实现对培育豆芽菜的温室温度进行模糊控制。具体要求如下：

- 根据温室当前温度的模糊判断结果，包括“温度过低”、“温度偏低”、“温度正常”、“温度偏高”或“温度过高”等，产生相应的控制信号，驱动相应的温度控制装置对温室进行升温或降温处理，使其温度处于豆芽菜生长的最适值；
- 温度控制装置可以手动控制也可以自动控制，它们的工作状态可以显示。

1. 模糊控制系统的建立

该系统是在[示例 12-1]的基础上增加了温度控制装置，采用模糊控制的方法对温度

进行控制，最终实现培育豆芽菜的温室温度的模糊监控。该系统的原理框图如图 12-30 所示。

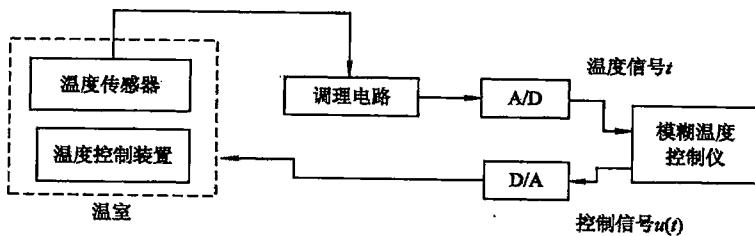


图 12-30 温度模糊控制系统的原理框图

整个系统由温度传感器、调理电路、数据采集卡(包括 A/D 和 D/A)、模糊温度控制仪和温度控制装置组成。模糊温度控制仪根据当前温度判断的结果产生相应的控制信号 $u(t)$ ，通过 D/A 输出给温度控制装置，温度控制装置对温室进行升温或降温处理，使其温度处于豆芽菜生长的最适值。

2. 模糊温度控制仪的设计

1) 面板设计

启动 LabVIEW 之后，进入仪器设计环境，创建如图 12-31 所示的面板。该面板主要由四部分组成：第一部分用于显示温室当前温度、温度状态以及温室温度的变化曲线，与 [示例 12-1] 相同。第二部分用于完成温室温度的控制，包括控制模式的选择、两种不同控制模式下相应控制参数或控制开关的选择。第三部分用于显示温度控制器的工作状态。第

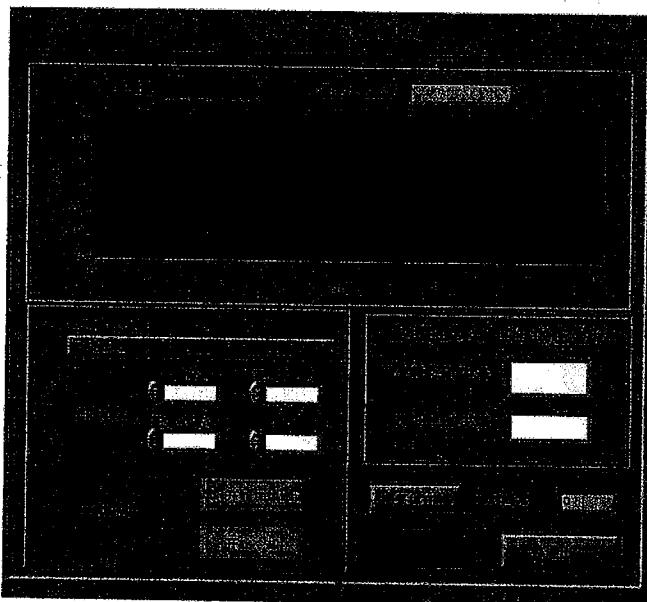


图 12-31 模糊温度控制仪面板

四部分用于计算温度控制系统的相关参数，并且控制此仪器的开或关。

(1) 放置数字显示控件 1、2：执行 All Controls→Numeric→Numeric Indicator 操作两次。

这两个控件均为输出显示型控件，用于输出温度传感器所测得的温室当前温度值和系统的时间常数，将其标记(Label)分别改为“温度”、“时间常数”。

(2) 放置字符串显示控件 3、4、5：执行 All Controls→String&Table→String Indicator 操作三次。

这三个控件均为输出显示型控件，用来输出温室温度的状态和温度控制器工作状态。将它们的标记(Label)分别改为“温度状态显示”、“冷却泵工作状态”、“加热器工作状态”。“温度状态显示”有五种可能的输出结果：“温度过低”、“温度偏低”、“温度正常”、“温度偏高”、“温度过高”；“冷却泵工作状态”有两种输出结果：“正在冷却”、“停止冷却”；“加热器工作状态”有两种输出结果：“正在加热”、“停止加热”。

(3) 放置图形显示控件 6：执行 All Controls→Graph→Waveform Chart 操作。

该控件为输出显示型控件，用来显示温室的温度变化曲线。

(4) 放置数字输入控件 7、8、9、10：执行 All Controls→Numeric→Numeric Control 操作四次。

这四个控件为输入控制型控件，在温室温度自动控制时，用于键入模糊温度域的边界值。将它们的标记(Label)分别改为“偏低”、“正常(低)”、“正常(高)”、“偏高”。

(5) 放置列表控件 11：执行 All Controls→List&Ring→Menu Ring 操作。

然后对控件增加两个选项(Item)，分别是自动控制和手动控制，其标记(Label)为“控制模式选择”。

(6) 放置开关控件 12、13、14、15、16：执行 All Controls→Boolean→Labeled Oblong Button 五次。

这两个控件都为开关型控件，分别用于温室温度自动控制时的自动加热和手动控制方式的手动冷却、启动采样、启动数据处理过程和关闭仪器。

2) 流程图设计

模糊温度控制仪的主程序流程图如图 12-32 所示，其核心就是模糊控制器的设计。

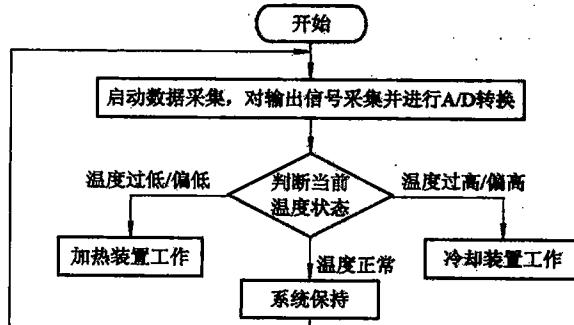


图 12-32 模糊温度控制仪主程序流程图

具体设计步骤如下：

(1) 模糊控制器结构的选择。模糊控制器以当前的温度为输入，以温室温度控制装置

的控制信号为输出，所设计的模糊控制器为单输入单输出模糊控制器。

(2) 模糊语言变量选择。选取模糊语言变量值时要考虑到控制规则的灵活与细致性，又要兼顾简单易行的要求。针对温度监控，本文将输入温度 t 的语言值和控制信号 $u(t)$ 的语言值定义为{过低、偏低、正常、偏高、过高}，用{NL、NS、ZO、PS、PL}表示。

(3) 模糊隶属函数确定。隶属函数有很多种，包括三角形、梯形、矩形函数等。选择隶属函数时，应根据实际情况选择。输入温度的隶属函数选用的是梯形隶属函数，具体过程与[示例 12-1]相同。

控制信号 $u(t)$ 的隶属函数由温度控制装置决定，将其论域定为 $[-2, -1, 0, 1, 2]$ ，输出为负时温度控制装置对温室升温，为正时温度控制装置对温室降温。所选择的隶属函数如图 12-33 所示。

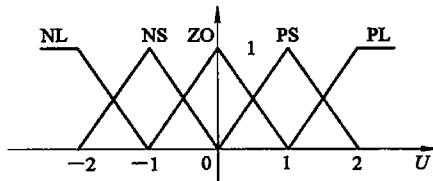


图 12-33 输出控制信号的隶属函数曲线

Matlab 中的三角形隶属函数 trimf 格式为 $y = \text{trimf}(x, [a \ b \ c])$ ，三个参数为 a 、 b 、 c ，参见公式(12-12)，或者更紧凑地表示为

$$f(x, a, b, c) = \max\left(\min\left(\frac{x-a}{b-a}, \frac{c-x}{c-b}\right), 0\right) \quad (12-48)$$

参数 a 和 c 确定三角形的“脚”，而参数 b 确定三角形的“峰”。

示例：

```
x=0:0.1:10;
y=trimf(x, [3 6 8]);
plot(x, y);
xlabel('trimf, P=[3 6 8]');
```

(4) 模糊推理规则确定。在温度监控仪中，根据输入温度的状态来决定控制信号的输出，从而让温控装置作出相应的动作。具体的规则如下：

温度过低——控制信号输出值 -2 ，温度控制装置以最大功率让温室升温；

温度偏低——控制信号输出值 -1 ，温度控制装置以较大功率让温室升温；

温度正常——控制信号输出值 0 ，温度控制装置不动作；

温度偏高——控制信号输出值 1 ，温度控制装置以较大功率让温室降温；

温度过高——控制信号输出值 2 ，温度控制装置以最大功率让温室降温。

根据上面分析的控制规则，以 if …then…的形式形成系统的规则语句，例如 if (T) is (PB) then (U) is (NB)。在 LabVIEW 中采用相应的模块来实现 if …then…结构。

通过这四步，就完成了模糊温度控制仪的主要程序设计。

3. 运行检验

首先保存文件，然后编译运行。例如系统处于自动控制状态，检测到温室当前“温度过

低”，则加热器正在工作，温室在升温，目前已升温到“温度偏低”状态，对应的运行结果如图 12-34 所示；当监控仪选为手动控制方式时，需要根据当前的温度状态来控制温度控制装置的动作，如图 12-35 所示，图中为检测到温室当前“温度过高”，冷却泵正在工作，温室在降温，当前温室温度已降到“温度偏高”状态。

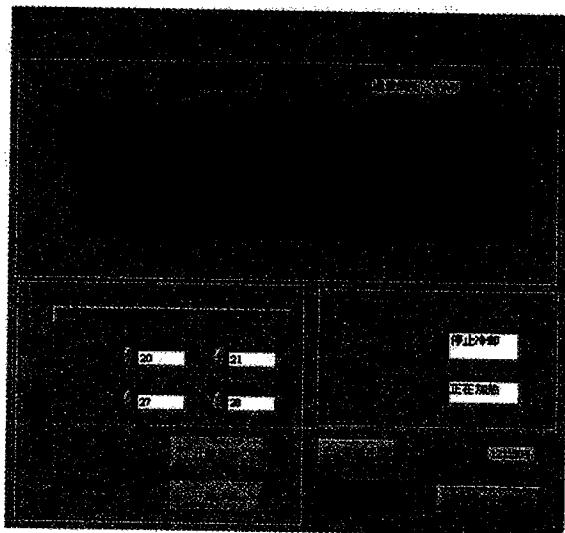


图 12-34 温度监控仪自动加热状态

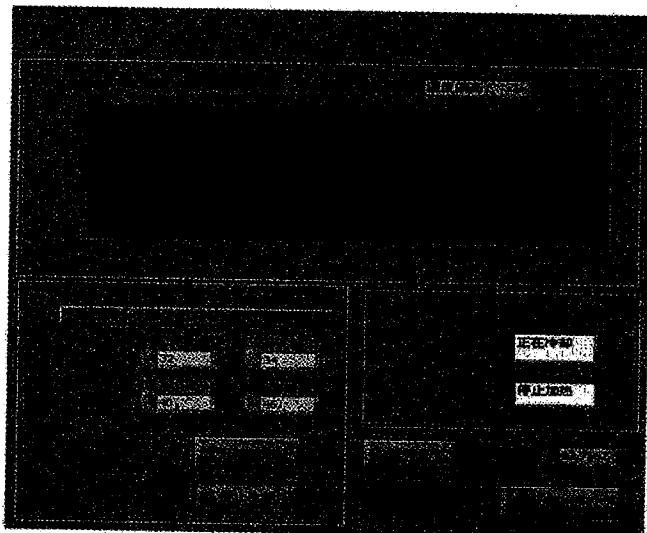


图 12-35 温度监控仪手动冷却状态

参 考 文 献

- [1] 常炳国, 刘君华. 模糊神经网络观测器在变压器状态监控中的应用[J]. 清华大学学报. 2003, 39(1): 35 - 39
- [2] 沈毅, 刘宜平, 刘志言. 模糊神经网络在多传感器故障检测与诊断中的应用[J]. 中国机械工程, 2001, 12(10): 1176 - 1178
- [3] 常炳国, 刘君华, 吴浩扬. 监测变压器绕组热点温度智能模糊传感器系统的研究[J]. 中国电机工程学报, 2000, 20(8): 14 - 17
- [4] 常炳国, 刘君华. 基于模糊集合理论的变压器绕组热点研究[J]. 中国电机工程学报. 1999, 19(7): 22 - 25
- [5] 洪文学, 韩峻峰, 周少敏. 模糊传感器研究现状与展望[J]. 传感器技术. 1996, (5): 1 - 4
- [6] 韩峻峰. 模糊传感器基本理论及其应用的研究[D]. 哈工大博士论文, 1996
- [7] 孟凡振, 周少敏, 洪文学, 等. 模糊传感器一般结构的探讨[C]. 北京: 非电量电测技术与传感技术学术研讨会, 1996, 46 ~ 50
- [8] Russo F. Fuzzy System in Instrumentation[J]. IEEE Trans., 1996, 45(2): 683 - 689
- [9] Han Junfeng. Neural Network-based Fuzzy Sensors for Gas Identification [J]. Journal of Electronic Measurement and instrument, 1995, 9(9): 417 - 419
- [10] Schodel H. Utilization of Fuzzy Techniques in Intelligent Sensors[J]. Fuzzy Sets and Systems, 1994, 63(6): 271 - 272
- [11] Benoit E, Mauris G, Foulloy L. A Fuzzy Color Sensor[C]. Proc. Of the 13th IMEKO World Congress, 1994, 12(9): 1015 - 1020
- [12] Foulloy L, Galichet S. Fuzzy Sensors for Fuzzy Control [J]. Fuzziness and Knowledge-based systems, 1994, 2(1): 13 - 24

第 13 章 无线传感器网络

本章内容

- 无线传感器网络与 Internet 的互联
- 无线传感器网络
- [示例 13-1]设计一个“远程无线传感器网络监测平台”的虚拟仪器程序
- [示例 13-2]将本地 VI“远程无线传感器网络监测平台”制作成为网络化 VI
- [示例 13-3]工业监测用无线传感器网络的硬件芯片与引脚连接
- [示例 13-4]工业监测用无线传感器网络的软件设计流程举例

随着自动化技术的推动，尤其是现场总线控制系统(Fieldbus Control System, FCS)发展的要求，目前已发展出了多种通信模式的现场总线网络化智能传感器/变送器。

随着社会的进步与发展，人们在更广泛的领域提出传感器系统的网络化需求，如大型机械的多点远程监测、环境地区的多点监测、危重病人的多点监测与远程会诊、电能的自动实时抄表系统以及远程教学实验等，无线传感器网络的重要性日益凸显。

无线传感器网络是由部署在监测区域内大量的传感器节点以自组织方式构成的无线网络，目的是感知、采集和处理网络覆盖的地理区域中感知对象的信息，并发布给观察者。其由于展开快速、抗毁性强、监测精度高、覆盖区域大等特点在军事、航空航天、工农业、环境监测多方面产生了广阔的应用前景，成为当前信息领域的研究热点。

传感器正向集成化、微型化、智能化、网络化方向发展。网络化是智能传感器系统的重要发展方向。本章将通过四个示例介绍一种远程无线传感器网络建立的过程与方法。

13.1 概 述

13.1.1 无线传感器网络研究与应用状况

无线传感器网络的研究起步于 20 世纪 90 年代末期。从 21 世纪开始，传感器网络引起了学术界、军事界和工业界的极大关注，美国和欧洲相继启动了许多无线传感器网络的研究计划。特别是美国通过国家自然基金委、国防部等多种渠道投入巨资支持传感器网络技术的研究。

无线传感器网络不同于传统数据网络，它对无线传感器网络的设计与实现提出了新的挑战，主要的要求有：低能耗、低成本、通用性、网络拓扑、安全、实时性、以数据为中心等。

1998 年，UCLA 和 Rockwell 研究中心在 DARPA 支持下进行的 WINS (Wireless

Integrated Sensor Network)是传感器网络研究的开端，其主要目的是为嵌入仪器、设备以及环境中的传感器、执行机构和处理器构建一个分布式网络环境，提供访问 Internet 的能力。之后，又相继出现了一些新的演示系统，比如 Smart Dust，它主要强调利用 MEMS 技术设计微型化的传感器节点，总体积约 100 立方毫米的节点完全有可能像尘埃一样悬浮在空气中。加州大学伯克利分校的 Macro Motes 在形如硬币大小的节点中配置了温度、湿度、压力、磁场等多种传感器，在车辆跟踪、动物学家关于水鸟活动的研究中应用非常成功。

NASA 的 JPL(Jet Propulsion Lab)实验室研制的 Sensorwebs 是为将来的火星探测进行技术准备的。2000 年，美国国防部把 Smart Sensorweb 定为国防部科学技术五个尖端领域之一。Smart Sensorweb 的基本思想是在整个作战空间中放置大量的传感器节点，来收集、传递信息，并且将信息汇集到融合点，在那里综合成一张图片，并分发给需要该信息的作战者。

在民用方面，Intel、Crossbow、Freescale、Ember 等美国公司，以及欧洲的一些公司如 Philips、Siemens、Ericsson、Chipcon 等也纷纷设立或启动相应的行动计划。

美国的很多大学都已开展无线传感器网络的研究。加州大学伯克利分校研制的传感器系统 Mica、Mica2、Mica2Dot 已被广泛地用于低能耗无线传感器网络的研究和开发；麻省理工学院(Massachusetts Institute of Technology)致力于基础知识的信号处理技术；哈佛大学(Harvard University)研究无线传感器网络通信的理论基础等。

在国内，无线传感器网络领域的研究也跟进很快，已经在很多研究所和高校广泛展开。其研究的热点、难点包括：设计小型化的节点设备；开发适合传感器节点的嵌入式实时操作系统；无线传感器网络体系结构及各层协议；时间同步机制与算法、传感器节点的自身定位算法和以其为基础的外部目标定位算法等。

特别是进入 21 世纪后，对无线传感器网络的核心问题有了许多新颖的解决方案，但是，这个领域从总体上来说尚属于起步阶段，目前还有许多问题亟待解决。

随着传感器技术、嵌入式技术、分布式信息处理技术和无线通信技术的发展，以大量的具有微处理能力的微型传感器节点组成的无线传感器网络(WSN)逐渐成为研究热点问题。与传统无线通信网络 Ad Hoc 网络相比，WSN 的自组织性、动态性、可靠性和以数据为中心等特点，使其可以应用到人员无法到达的地方，比如战场、沙漠等。因此，可以断定未来无线传感器网络将有更为广泛的前景。

13.1.2 无线传感器网络通信协议

随着 20 世纪 90 年代末 RF 与蓝牙(Bluetooth)等短距离无线通信技术的发展，无线个域网 WPAN(Wireless Personal Application Network)开始发展起来。目前无线个域网标准化组织 IEEE 802.15 工作组已完成了以下标准的制定：

中速无线个域网标准 IEEE 802.15.1——蓝牙；

高速无线个域网标准 IEEE 802.15.3——超宽带(UWB)；

低速无线个域网标准 IEEE 802.15.4。

低速无线个域网主要为电源能力受限的、吞吐量要求较低的无线应用提供简单的低成本网络连接，主要目标是以简单灵活的协议构建一种安装布置合理、数据传输可靠、设备成本极低、能量消耗较小的短距离无线通信网络。

低速无线个域网符合无线传感器网络关于低能耗、低成本、通用性、网络拓扑、安全、实时性、以数据为中心等要求，因此目前研究、应用的无线传感器网络的物理层及 MAC 层协议多采用 IEEE 802.15.4 标准。

基于 IEEE 802.15.4 标准的网络层协议主要有 2001 年 9 月成立的 ZigBee 联盟提出的 ZigBee 协议栈及适用于无线传感器网络节点的嵌入式微型 IPv6 协议栈。其中，ZigBee 协议以其低成本、不同厂商生产的产品可兼容等特点得到广泛的研究与应用。

13.1.3 无线传感器网络与 Internet 的互联

目前，国内外对无线传感器网络与 Internet 互联的研究尚未成熟。归纳已有的研究成果，无线传感器网络与 Internet 互联的主要内容是：利用网关或 IP 节点，屏蔽下层无线传感器网络，向远端的 Internet 用户提供实时的信息服务，并且实现互操作。

已有两种解决无线传感器网络与 Internet 互联的方案：同构网络和异构网络。

同构网络引入一个或几个无线传感器网络传感器节点作为独立的网关节点并以此为接口接入互联网，即把与互联网标准 IP 协议的接口置于无线传感器网络外部的网关节点。这样做比较符合无线传感器网络的数据流模式，易于管理，无需对无线传感器网络本身进行大的调整；缺点是会使得网关附近的节点能量消耗过快并可能会造成一定程度的信息冗余。

异构网络的特点是：部分能量高的节点被赋予 IP 地址，作为与互联网标准 IP 协议的接口。这些高能力节点可以完成复杂的任务，承担更多的负荷，难点在于无法对节点的所谓“高能力”有一个明确的定义。同时，如何使得 IP 节点之间通过其它普通节点进行通信也是一个技术难题。

综上所述，同构网络较易实现，即在无线传感器网络外部建立一个网关节点。

13.1.4 实现远程监测的无线传感器网络系统的典型结构

采用同构网络实现远程监测的无线传感器网络系统典型结构如图 13-1 所示，由传感器节点、汇聚节点、服务器端的 PC 和客户端的 PC 四大硬件环节组成，各组成环节功能如下。

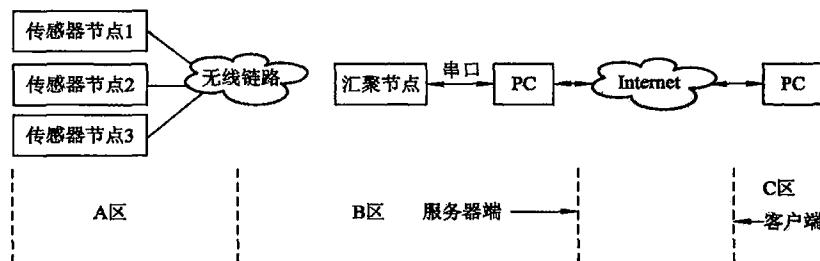


图 13-1 远程监测无线传感器网络系统结构框图

- **传感器节点：**部署在监测区域(A 区)，通过自组织方式构成无线网络。传感器节点监测的数据沿着其它节点逐跳进行无线传输，经过多跳后达到汇聚节点(B 区)。

- **汇聚节点：**是一个网络协调器，负责无线网络的组建，再将传感器节点无线传输进来的信息与数据通过 SCI(Serial Communication Interface，串行通信接口)传送至服务器端 PC。

• 服务器端 PC：是一个位于 B 区的管理节点，也是独立的 Internet 网关节点。在 LabVIEW 软件平台上面有两个软件：一是对传感器无线网络进行监测管理的软件平台 VI，即一个监测传感器无线网络的虚拟仪器 VI；二是 Web Server 软件模块和远程面板技术(Remote Panel)，可实现传感器无线网络与 Internet 的连接。

• 客户端 PC：客户端 PC 上无需进行任何软件设计，在浏览器中就可调用服务器 PC 中无线传感器网络监测虚拟仪器的前面板，实现远程异地(C 区)对传感器无线网络(A 区)的监测与管理。

本章将主要介绍传感器节点、汇聚节点以及传感器无线网络与 Internet 的互联。

13.2 无线传感器网络与 Internet 的互联

13.2.1 基于 LabVIEW 虚拟仪器的网络化方法

在 LabVIEW 6.1 以前的版本中，已经提供了相关的虚拟仪器网络化功能，例如 TCP/IP、Remote Data Acquisition(RDA)、Internet Toolkit、VI Server、Front Panel Web Publishing、DataSocket 等。当然也有其他第三方公司的一些 Toolkits 提供了 LabVIEW 的网络功能，如 AppletVIEW。NI 公司在总结上述功能的基础上，从 LabVIEW 6.1 版本开始，提供了全新的 Web Publish Tools 工具。使用该工具，用户可以轻松地将自己的虚拟仪器(VI)发布到网络上。用户可以通过一个计算机(无论近端或远程)来操作另一台不同计算机上的 VI。服务器端，用户无需特别的配置服务器程序，LabVIEW 已经提供了完整的 Web Server 服务器程序模块(Web 服务器)，LabVIEW 的远程面板(Remote Panel)技术可将服务器端运行的 VI 面板嵌入 HTML 网页，发布上网。客户端所需的组件只不过是一个 IE 浏览器以及安装 LabVIEW 的运行环境 run-time engine 而已。

利用 LabVIEW 设计网络化的 VI 分为三个步骤。

(1) 制作本地可运行的 LabVIEW 程序 VI。在传感器无线网络与 Internet 互联的场合，这个 LabVIEW VI 就是在 LabVIEW 环境中设计的名为“远程无线传感器网络监测平台”的虚拟仪器程序。这是位于 B 区的传感器无线网络的管理节点。该监测管理程序以 LabVIEW VI 的形式存于服务器端 PC 中。

(2) 配置网络服务器。配置网络服务器是为了使本地计算机成为一台可响应远程用户操作、提供本地 VI 响应数据的服务器，成为一个独立的 Internet 网关节点。

配置网络服务器的内容包括注册 Web Server，定义访问权限，定义可访问的 VI 列表三个部分。

(3) 发布嵌入 VI 前面板的 HTML 网页。分两步进行：一是设置网页参数；二是生成 HTML 网页。

所有在 LabVIEW 平台上设计的 VI 均可以通过上述三个步骤制作成为网络化 VI。网络化 VI 构架如图 13-2 所示。当用户在浏览器下键入 HTML 文件的网址后，便可在浏览器内对 VI 前面板上的按钮等输入控件进行操作，控制服务器端 VI 的运行，观察前面板上输出控件的显示，获得 VI 运行结果。

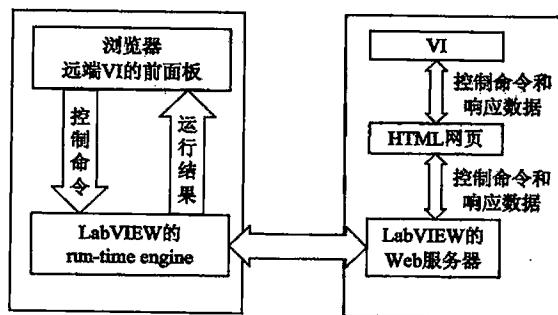


图 13-2 LabVIEW 的网络化 VI 构架

13.2.2 [示例 13-1]设计“远程无线传感器网络监测平台”虚拟仪器程序

要求：

- (1) 该 VI 经过 PC 机的串口与下位机进行通信，下达指令或接收数据，下位机为图 13-1 所示无线传感器网络的汇聚节点，即协调器中的微处理器；
- (2) 该 VI 对无线网络有操作功能，如“检索网络”、“发送数据”、“自动发送”等；
- (3) 该 VI 对无线网络节点地址等具有显示功能；
- (4) 该 VI 对无网络状态具有显示功能，如显示网络中的传感器节点是否入网、无线网络是否组建成功等；
- (5) 该 VI 对接收到的传感器历史数据与当前数据具有显示功能。

由五大模块来实现上述要求：串口配置按钮模块、网络操作按钮模块、网络状态指示灯模块、串口接收数据显示框模块、传感器数据显示模块。其前面板与流程图分别如图 13-3、图 13-4 所示。

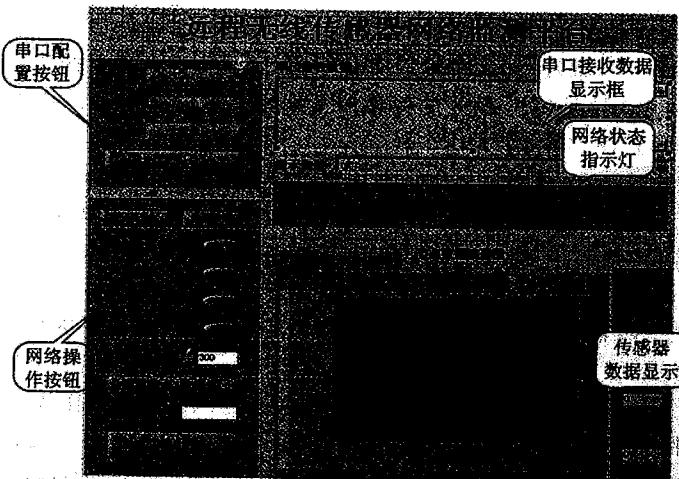


图 13-3 “远程无线传感器网络监测平台”VI 面板

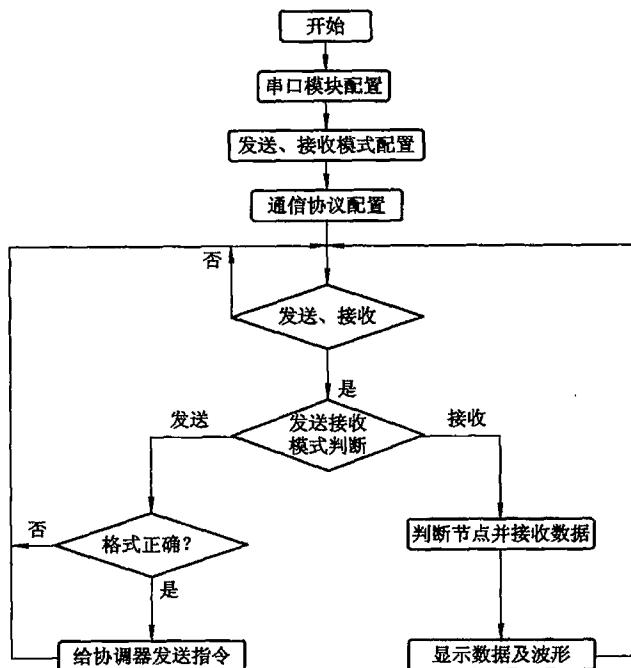


图 13-4 “远程无线传感器网络监测平台”流程图

1. 串口配置按钮模块

在此模块中使用 LabVIEW 中的 VISA Configure Serial Port 模块，由面板上的三个按钮(通讯口、波特率及校验位)实现对 PC 机串口的配置，从而服务器端 PC 与下位机即作为无线网络汇聚节点的协调器之间可通过 PC 串口进行通信。

2. 网络操作按钮模块

此模块有八个按钮、一个 Ring 控件、一个 Numeric 控件及一个 String 控件，实现对无线传感器网络的操作，如发送数据、自动发送、检索网络、读取节点数据等。

3. 串口接收数据显示框模块

在此模块中使用 String 控件显示监测平台通过 PC 机的串口接收到的无线传感器网络信息，如网络各节点的长短地址等。

4. 网络状态指示灯模块

此模块中包括四个指示灯，指示无线传感器网络各节点的运行状态，如网络是否组建成功、传感器节点 $n(=1, 2, 3)$ 是否入网。

5. 传感器数据显示模块

此模块中设置三个 Numeric Indicator 控件分别显示三个传感器节点的传感器数据；将三个 Graph 控件拖入一个 Tab Control 控件中实现三种传感器历史数据的重叠选择显示。

在图 13-3 所示前面板上，点击相应的串口配置按钮即可完成对串口的配置。无线传感器网络信息在串口接收数据框中显示，若网络出现故障，则在提示信息中显示出错信息。网络组建成功及节点成功加入后会点亮相应的指示灯。点击图中左下角的按钮，可完成相应的指令发送及数据接收操作。接收的节点数据将分别以图形和数据的形式在右下角的图形显示框和数据显示框中显示，其中，图形显示框为三个节点复用，随着读取数据的节点不同而转换。

在图 13-4 所示的流程图中，先配置串口模块，再确定发送、接收模式，同时确定与下位机的通信协议。当有发送或接收控制键动作时，程序首先判断是给下位机发送指令，还是接收下位机回传的数据。若是发送指令，则判断所发送指令是否符合格式要求，不符合则退出发送；若是接收指令则接收数据，判断是哪个节点数据后再分别在相应节点的显示框中显示波形和数据值。上述设计好的本地 VI 可进一步向网上发布，成为网络化 VI，远方 C 区的使用者可调用该本地 VI 的前面板，就像在本地 B 区操作该 VI 前面板一样对传感器无线网络进行操作与管理。

13.2.3 [示例 13-2] 将本地“远程无线传感器网络监测平台”VI 制作成为网络化 VI

要求：通过客户端浏览器可调用服务器端本地“远程无线传感器网络监测平台”VI 的前面板，并在该前面板上操作就可实现对无线传感器网络经 Internet 进行远程操作与管理。

所有在 LabVIEW 平台上设计的 VI 都是按如下相同的步骤制作成网络化 VI 的。

1. 配置网络服务器

(1) 注册 Web Server。注册 Web Server 需要定义 Web Server 的根目录路径、HTTP 协议端口、控制 VI 的时间以及 Web Server 的日志文件。打开 LabVIEW，选择 Tools→Options，弹出 LabVIEW 的设置对话框，在设置对话框中选择 Web Server Configuration，如图 13-5 所示，对 LabVIEW 的服务器进行设置。其中：

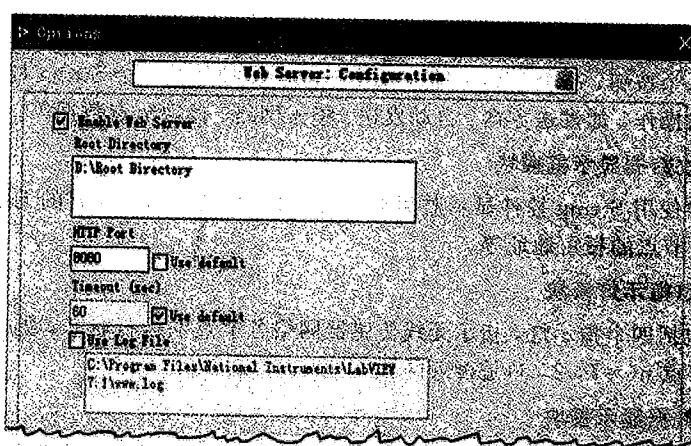


图 13-5 LabVIEW 的 Web Server 设置面板

① Root Directory: 用于输入网页和 VI 程序的根目录，供 Web 服务器查找 VI 使用。例如本例中应为 D:\Root Directory。

② HTTP Port: Web 服务器的 HTTP 协议端口，默认端口为 80，端口数值范围从 1 到 65535。注意，如果在服务器计算机上不仅运行有 LabVIEW 的 Web 服务器，还运行其他的 Web 服务器，为了避免因 LabVIEW 的 Web 服务器的 HTTP 协议端口与其他的 Web 服务器端口冲突，而无法使网络用户在远端正确访问服务器上的 LabVIEW 程序，需要将 Web 服务器的 HTTP 协议端口定义为不同的数值，例如 8080 或其他值。如果不存在其他 Web 服务器端口，则可采用默认值而不予设置。

③ Timeout(sec): 设置网络用户控制服务器上的 VI 程序的时间，单位为秒，默认时间为 60 秒。它是远程网络上的 VI 面板与服务器上的 VI 程序的有效连接时间。每次远程用户打开一个新的网络虚拟仪器页面时，服务器开始计时，到 60 秒后，页面和服务器的联系中断。用户继续操作面板时，会重新连接服务器。

④ Use Log File: 使用日志记录。该日志文件记录远端用户访问的情况，便于管理人员进行维护工作。

(2) 定义访问权限。定义访问权限是指定义一个管理操作 Web Server 的用户列表，并且分别对每一个用户做出访问权限设置。访问权限有三种，分别是“具备观看和控制权限(Allow Viewing and Controlling)”，“只具备观看权限(Allow Viewing)”或“列为拒绝户(Deny Access)”。定义访问权限列表的工作可以在“Web Server: Browser Access”中设置。选择 Tools→Options→Web Server: Browser Access，弹出如图 13-6 所示面板。

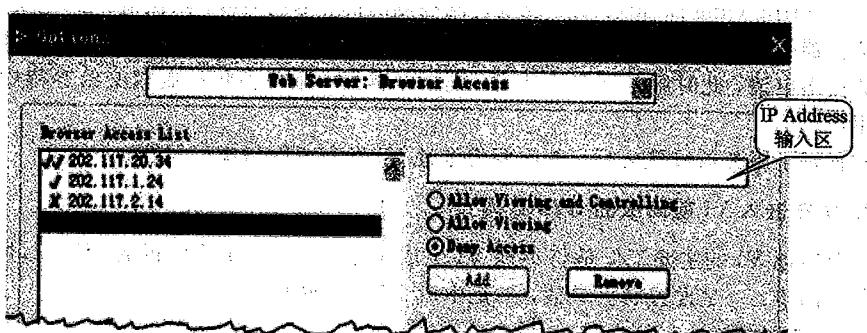


图 13-6 定义访问权限面板

其中：

- ① Browser Access List: 访问列表显示区，显示所有已经定义的 IP Address 列表。
- ② Allow Viewing and Controlling: “具备观看和控制权限”选择框，如 202.117.20.34；
- ③ Allow Viewing: “只具备观看权限”选择框，如 202.117.1.24；
- ④ Deny Access: “列为拒绝户”选择框，如 202.117.2.14。

管理员在 IP Address 输入区中输入用户 IP，然后选择控制权限，点击“Add”，即可将用户 IP 加入“Browser Access List”中。

(3) 定义可访问的 VI 列表。定义可访问的 VI 列表是指定义一个可以被远程用户访问

的服务器端 VI 列表，以及针对每个 VI 设定控制时间长短，防止 VI 被某一个使用者长时间控制，造成其他使用者无法使用。上述功能可以选择 Tools→Options→Web Server: Visible VIs，通过如图 13-7 所示面板进行设置。

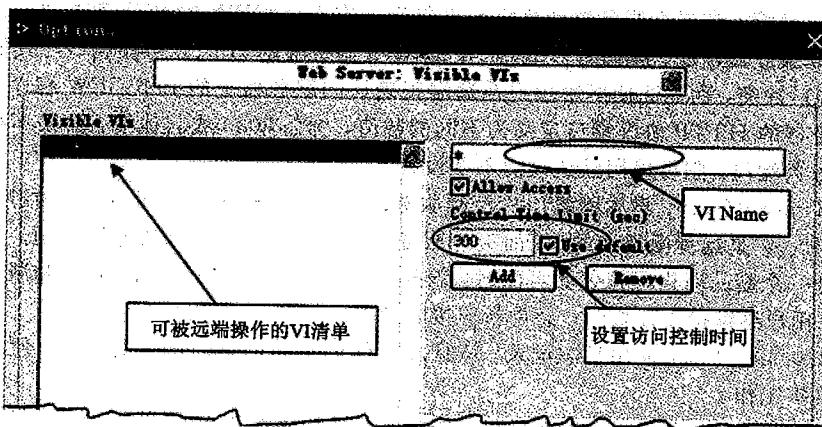


图 13-7 定义可访问的 VI 列表的面板

① **Visible VIs**: 可以访问的服务器端的 VI 列表，默认值是“*”，表示服务器端的所有 VI 均可以被网络用户访问到。

② **Control Time Limit(sec)**: 设置访问控制时间，单位为秒，默认值为 300 秒。一个网络虚拟仪器被一个网络用户占用时，另一个网络用户试图访问同一个网络虚拟仪器，并向服务器提出申请，此时服务器开始计时，到 300 秒后，服务器将控制权交给等待中的另外一个人。

配置好了服务器设置之后，就需要将本地 VI 发布成网络虚拟仪器。

2. 发布嵌入 VI 前面板的 HTML 网页

发布本地 VI 的主要工作是将本地 VI 生成一个包含该 VI 前面板的 HTML 网页。该任务分为两步：一是设置网页参数，二是生成 HTML 网页。

1) 设置网页参数

设置网页参数是指配置将要生成的网络虚拟仪器的页面参数以及与本地 VI 进行关联。选择 Tools→Web Publish Tool，弹出网页发布对话框，如图 13-8 所示。

① **Document Title**: 输入待发布的 HTML 网页标题，本例为“远程无线传感器网络监测平台”。

② **Header**: 在网页中位于面板前面的文本，标识前面板开始的位置，本例输入“远程无线传感器网络监测平台开始位置”。

③ **Footer**: 在网页中位于面板后面的文本，标识前面板结束的位置，本例输入“远程无线传感器网络监测平台结束位置”。

④ **VI Name**: 将要发布的 VI 名称，本例中本地 VI 程序名称是“Remotel.vi”。

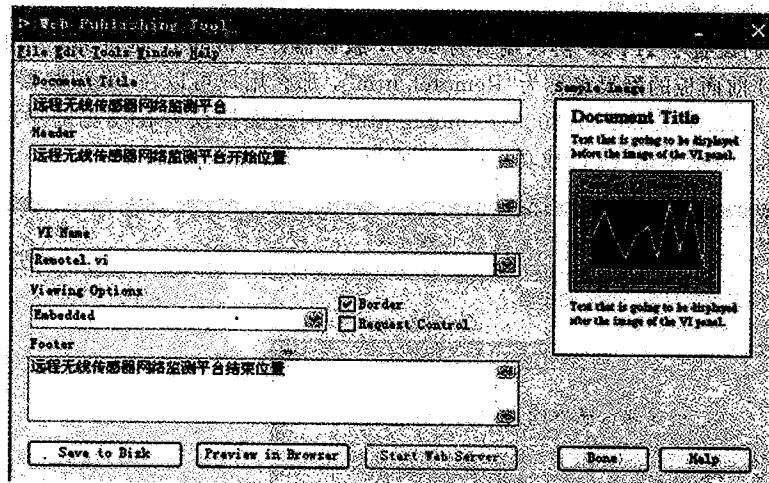


图 13-8 网页发布对话框

2) 生成 HTML 网页

用户设置好面板上的参数后，在图 13-8 所示的面板上点击“Start Web Server”按钮，使 LabVIEW 的 Web 服务器开始运行。若点击“Save to Disk”，则弹出的对话框如图 13-9 所示。

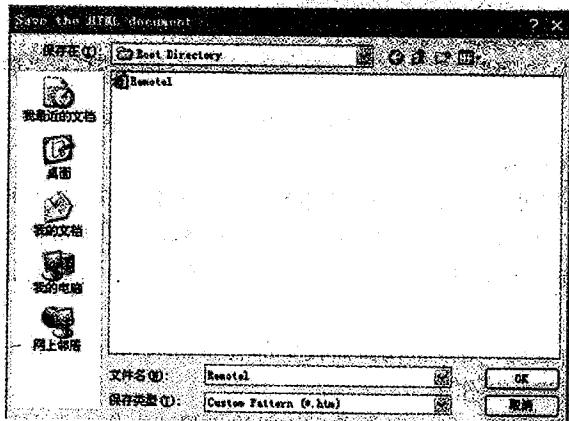


图 13-9 生成 HTML 网页

在该对话框中，用户可以将计算机自动生成的含有 VI 前面板的 HTML 网页命名并保存在 Web 服务器的根目录中，本例中网页名称为 Remotel.htm，保存在 D:\Root Directory 中。

3. 网络虚拟仪器实际操作测试

进行网络虚拟仪器实际操作测试是为了检验生成的网络虚拟仪器是否可以正常运行和操作，一般包括两步工作：一是浏览网络虚拟仪器；二是操作网络虚拟仪器。

1) 浏览网络虚拟仪器

本例我们设置了 HTTP 协议端口为 8080, 服务器计算机的 IP 地址为 202.117.20.33, 包含远程 VI 前面板的网页名称为“Remotel.htm”, 那么用户可以在 IE 浏览器的地址栏中输入 <http://202.117.20.33:8080/Remotel.htm>, 网络虚拟仪器在浏览器中应显示如图 13-10 所示, 表示本地 VI 已成功上网, 已成为一个网络虚拟仪器。

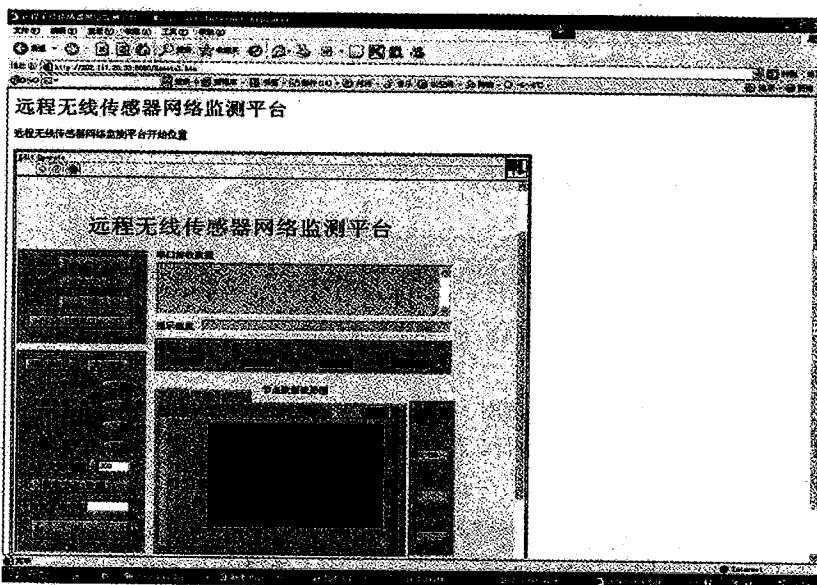


图 13-10 网页上的网络虚拟仪器界面

2) 操作网络虚拟仪器

操作网络虚拟仪器可分为三步: 申请控制权、操作网络虚拟仪器获得数据和释放控制权。

(1) 申请控制权。申请控制权是指建立客户端网络虚拟仪器和服务器端 VI 的连接。在图 13-10 中, 用鼠标左键点击网页上的 VI 面板区域, 在弹出的对话框中选择“Request Control of VI”。

如果控制权申请成功, 则虚拟仪器面板中部会弹出“Control Granted”对话框, 然后用户就可以对虚拟仪器进行操作了。

(2) 操作网络虚拟仪器。操作网络虚拟仪器的步骤和操作本地 VI 一样。

(3) 释放控制权。释放控制权是指撤销客户端网络虚拟仪器与服务器端的连接, 以便其他的网络用户使用该网络虚拟仪器。用鼠标左键点击网页上的 VI 面板区域, 在弹出的对话框中选择“Release Control of VI”, 即可释放对服务器端 VI 的控制权。

LabVIEW 内定可以连接的 Remote Panel 的数目为一个。用户可以向 NI 公司订购额外的授权以使更多的用户端可以同时连接到 Web Server 端。

几乎所有的 LabVIEW 程序都可以通过上述网络发布过程制作成网络虚拟仪器, LabVIEW 和网络的结合给虚拟仪器的网络化提供了一条简便快捷的道路, 也给无线传感器网络远程监控提供了捷径。

13.3 无线传感器网络

一个无线传感器网络一般由传感器节点、汇聚节点与管理节点三大硬件部分组成，传感器节点与汇聚节点间是通过无线通信设施进行联系的。图 13-11 所示是一种经济型无线传感器网络。受传感器节点发射能力的限制，在更远距离的测试研究中心需要采集远程分布的传感器节点的信息时，则可由管理节点与 Internet 相连，把无线传输的传感器节点信息进一步通过 Internet 传至世界各地。上一节已经介绍了经管理节点将无线传感器网络与 Internet 互联，本节介绍无线传感器网络的传感器节点、汇聚节点以及它们之间的无线通信问题。

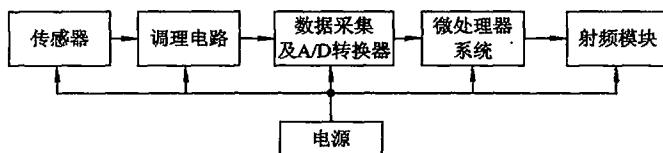


图 13-11 无线网络传感器节点的组成环节框图

13.3.1 无线传感器网络中的传感器节点

无线传感器网络中的每一个传感器节点都是一个具有无线通信功能的智能传感器系统，因此无线传感器网络中的传感器节点除了通常智能传感器系统所具有的环节，如传感器及其调理电路、数据采集与 A/D 转换、微处理器系统之外，还特有射频模块以实现无线通信智能化功能。

1. 传感器及其调理电路

应根据无线传感器网络所在的地区环境特点来选择传感器，以适应环境温度变化范围、尺寸体积等特殊要求。传感器所配接的调理电路将传感器输出的变化量转换成能与 A/D 转换器相适配的 0~2.5 V 或 0~5 V 的电压信号。当处于无电网供电地区时，传感器及其调理电路都应是低功耗的。

2. 数据采集及 A/D 转换器与微处理器系统

传感器节点中的计算机系统是低功耗的单片微处理器系统，可以适应远离测试中心、偏远地区恶劣环境的工作条件。如美国德克萨斯州仪器(TI)公司生产的 MSP430F149A 超低功耗混合信号处理器(Mixed Signal Processor)，它内部自带采样/保持器和 12 位 A/D 转换器，可对信号进行采集、转换以及对全节点系统进行指令控制和数据处理。

3. 射频模块

射频模块接收外部无线指令并将传感器检测到的被测参量数据信息无线发送出去，如 TI 公司的 CC2420 无线收发芯片。

4. 电源

无线传感器网络中对传感器节点的供电是一个极具特殊性的正处于研究热点的技术问

题。若节点处于远离电网的偏远地区，一般采用电池供电或无线射频供电方式。当然这种方式供电产生的电功率有限，因而节点发射功率有限，无线信号传输的距离也受限制。采用太阳能或风能作为电源可以产生大的电功率，但也需要进行相应的基础设施的建设与投资。因此，采用何种方式供电，视无线传感器网络的监测目的、地区环境等各项因素而定。

13.3.2 无线传感器网络中的汇聚节点

图 13-1 中的无线传感器网络汇聚节点是一个网络协调器，操作 PC 中监测管理软件平台的面板控件，在其指令下负责执行无线传感器网络的配置与组建，并将接收到的传感器节点无线传输的数据信息再传至 PC 机。通常协调器主要由微处理器系统、射频模块、通信接口以及电源四个部分组成，其硬件组成框图如图 13-12 所示。

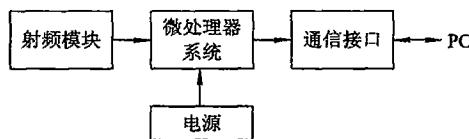


图 13-12 无线网络协调器硬件组成框图

1. 通信接口

协调器中的通信接口负责与 PC 机进行通信。一方面，当操作 PC 机中无线传感器网络监测平台 VI 前面板上的相应控件时，通信接口负责传递下达的相应指令，如检索网络、发送数据等；另一方面，协调器接收到传感器节点无线发送的数据信息时，也将其通过通信接口上传到 PC 中。

2. 微处理器系统

协调器中的微处理器是整个无线传感器网络的主控制器，是协调器的核心。

3. 射频模块

该射频模块将接收传感器节点无线发送的数据信息，经通信接口上传至 PC；另一方面，以无线传输方式下达 PC 对传感器节点的操作指令。

协调器与传感器节点中的微处理器、射频模块以及电源模块等有许多共同处，可以采用相同的芯片进行设计。

13.3.3 [示例 13-3]工业监测用无线传感器网络的硬件芯片与引脚连接

要求：

- (1) 该无线传感器网络的结构如图 13-1 所示。
- (2) 该无线传感器网络至少有三个测量温度(0~200°C)、分辨力为 0.1°C 的传感器节点；
- (3) 测点分布在高温强噪声工业现场，要求无线信号传输距离达到 20 m；
- (4) 下位机(协调器)与上位机(PC)之间采用串口通信方式(SCI)。

该工业监测用无线传感器网络结构如图 13-1 所示，主要由传感器节点、网络协调器(汇聚节点)及 PC 三部分组成，下面只讨论传感器节点和网络协调器所涉及的硬件、芯片

电路及其引脚连接。

1. 传感器及其调理电路

传感器采用工业测温铂电阻 Pt100，它的稳定性好，适宜长期监测。调理电路采用双恒流源仪用放大器电路，如图 13-13 所示。

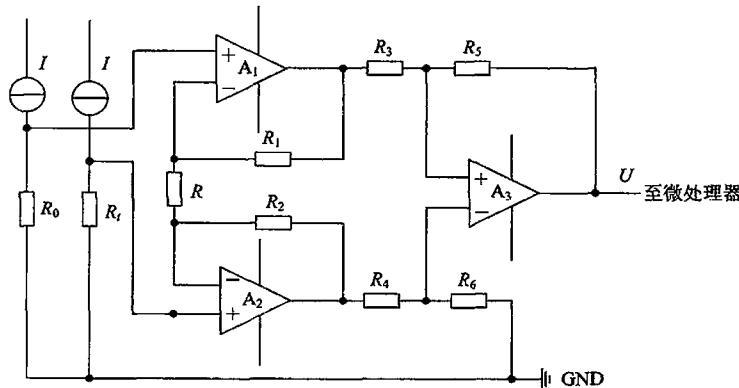


图 13-13 测温铂电阻 Pt100 调理电路图

图 13-13 中 R_t 为测温铂电阻 Pt100； $R_0 = 100 \Omega$ 为精密电阻； $I = 1 \text{ mA}$ 为恒流源供电电流； U 为调理电路输出电压，且

$$U = K(IR_t - IR_0) \quad (13-1)$$

可求放大倍数 K 为

$$K = \frac{U}{I}(R_t - R_0)^{-1} \quad (13-2)$$

当 $U=2.5 \text{ V}$, $I=1 \text{ mA}$ 时，在 $0 \sim 200^\circ\text{C}$ 范围内 $R_t=175.86 \Omega$ ，代入式(13-2)得 $K=32.96$ ，可取 $K=30$ ，这时最大输出电压 $U_{\max}=2.2758 \text{ V}$ ，小于 2.5 V ，即小于 A/D 转换器允许输入的最大电压值。

根据分辨力 0.1°C 与测温上限 200°C 的要求，A/D 转换器的位数 b 需满足

$$2^b \geq \frac{200}{0.1} = 2000 \quad (13-3)$$

即 $b \geq 11$ 均可满足分辨力的要求。本例中单片机 A/D 转换器是 12 位，故满足要求。

2. 微处理器模块

微处理器模块选用美国 TI 公司生产的 MSP430F149A 单片机，它是一种 16 位超低功耗的混合信号处理器(Mixed Signal Processor)，在无线传感器网络中作为网络协调器与传感器节点的主控器。

1) MSP430F149A 单片机的特点

MSP430F149A 单片机具有以下特点：

- 超低功耗；
- 强大的处理能力；
- 高性能模拟技术及丰富的片上外围模块；

- 系统工作稳定；
- 方便高效的开发环境。

2) MSP430F149A 单片机的主要性能

MSP430F149A 单片机的主要性能如下。

- 低电源电压范围：1.8~3.6 V。
- 超低功耗： $2.5 \mu\text{A}$ @4 kHz, 2.2 V; $280 \mu\text{A}$ @1 kHz, 2.2 V。
- 5 种节电模式：等待方式 $1.6 \mu\text{A}$, RAM 保持的节电方式 $0.1 \mu\text{A}$ 。
- 从等待方式唤醒时间： $6 \mu\text{s}$ 。
- 16 位 RISC 结构，125 ns 指令周期。
- 基本时钟模块配置(XX 时钟搭配使用可降低单片机功耗)：高速晶体(最高 8 MHz), 低速晶体(32 768 Hz)。
- 12 位 200 kb/s 的 A/D 转换器，自带采集保持，故满足测温分辨力 0.1°C 需要位数 $b \geq 11$ 的要求。
- 内部温度传感器。
- 具有 3 个捕获/比较寄存器的 16 位定时器 Timer_A、Timer_B。
- 两通道串行通信接口可用于异步或同步(UART/SPI)模式。
- 6 个 8 位并行口，且 2 个 8 位端口有中断能力。
- 硬件乘法器。
- 多达 60 KB FLASH 和 2 KB RAM。
- 串行在线系统编程。
- 保密熔丝的程序代码保护。

3) 无线传感器网络中 MSP430F149A 单片机的引脚连接

在传感器节点与网络协调器中，MSP430F149A 的引脚连接分别如图 13-14、图 13-15 所示。

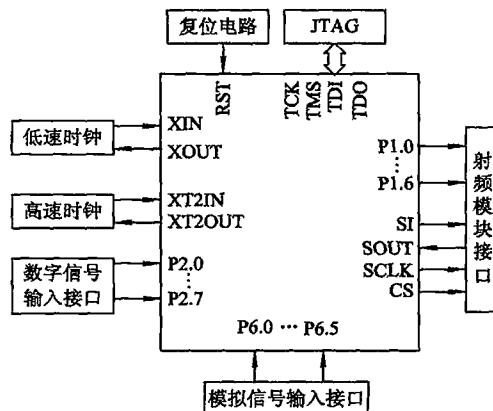


图 13-14 传感器节点中 MSP430F149A 单片机引脚分配图

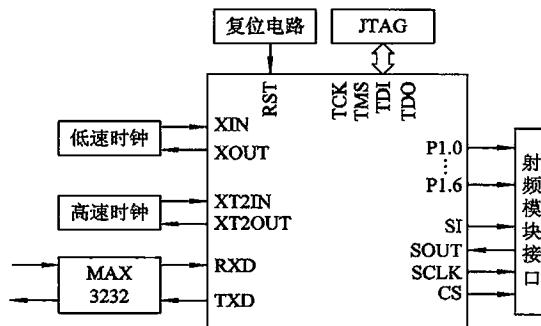


图 13-15 网络协调器中 MSP430F149A 单片机引脚分配图

(1) 图 13-14 与图 13-15 引脚连接相同的部分有：

① 时钟电路，向单片机提供必需的时钟信号，采用高、低双时钟配置模式可降低功耗。

- 引脚 XIN、XOUT 是低速时钟电路的接入端，采用 32 768 Hz 频率的晶振；
 - 引脚 XT2IN、XT2OUT 是高速时钟电路接入端，可采用的晶振频率最高为 8 MHz。
- ② 复位电路，采用上电模式向单片机最小系统提供正常稳定工作必需的复位信号。

- 引脚 RST 为复位信号接入端；

③ JTAG 是单片机仿真口数据通信接口，有四个引脚：TDO、TDI、TMS 与 TCK，通过这些引脚可以将程序烧录在单片机中；

④ 与射频模块的连接。

- 引脚 SI、SOUT、SCLK 和 CS 是单片机与射频模块之间进行 SPI 通信用连接脚；
- 引脚 P1.0～P1.6 是单片机对射频模块进行配置、操作，实现无线传输的连接脚。

(2) 图 13-14 与图 13-15 引脚连接不同的部分有：

① 在图 13-14 所示的传感器节点中，

• 引脚 P6.0～P6.5 是 6 路模拟信号输入通道，与传感器的模拟信号输出端相连，其中有 2 路 4～20 mA 标准电流信号输入通道及 4 路 0～2.5 V 电压信号输入通道；
 • 引脚 P2.0～P2.7 是 8 路数字量输入通道，与数字式传感器输出端相连。

本例采用图 13-13 所示传感器及其调理电路，其输出的模拟电压 U 可指定 P6.0～P6.5 中任一个通道输入。

② 在图 13-15 所示的网络协调器中，

• 引脚 RXD、TXD 分别是单片机串口通信的数据接收端和发送端，是与上位机——PC 串口通信的接口，用于网络协调器与 PC 之间进行串口通信，通信接口电路如图 13-16 所示。

通信接口电路的任务是进行电平转换，本例采用一片 MAX3232 电平转换电路芯片作为通信接口电路。

MAX3232 兼容 5 V 与 3.3 V 的 CMOS 电平输入，有两个收发器，最高传输速度为 250 kb/s，满足系统对 SCI 通信的要求。在 PC 向协调器发数据时，其输入的电平通过

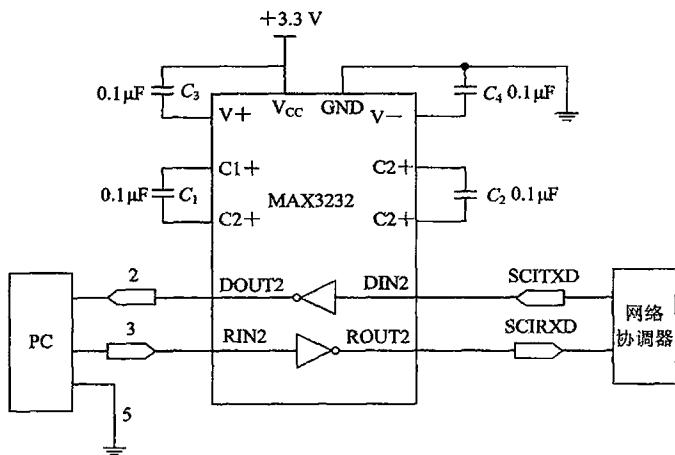


图 13-16 SCI 通信接口电路

MAX3232 转换为 3.3 V 电平；在协调器向 PC 发送数据时，其输入的电平通过 MAX3232 转换为 RS232-C 通信标准规定的电平。

RS232-C 是美国电子工业协会正式公布的，已经应用于各个领域的一种异步串行通信标准。它的逻辑电平以公共地为对称，其逻辑“0”电平规定在 +3~+15 V 之间，逻辑“1”电平则在 -3~-15 V 之间，因而它不仅要使用正负极性的双电源，而且与传统的 TTL 等数字电路的逻辑电平不兼容，连接时必须进行电平转换。本例中，由于 MSP430F149A 的 SCI 的逻辑电平为 3.3 V，而计算机的串口电平为 -15~+15 V，因此必须在通信的时候进行电平转换。

3. 射频模块

1) 射频无线收发芯片简介

本例选用 TI 公司的 CC2420 无线收发芯片为射频模块的核心。CC2420 是 2003 年底推出的符合 IEEE 802.15.4 标准的无线收发芯片。它基于 Smart RF 03 技术，使用 0.18 μm CMOS 工艺生产，具有很高的集成度。

CC2420 是一个半双工的 RF 芯片，它具有完全集成的压控振荡器，只需天线、16 MHz 晶振等非常少的外围电路就能在 2.4 GHz 频段上工作。CC2420 的选择性和敏感性指数超过了 IEEE 802.15.4 标准的要求，可确保短距离通信的有效性和可靠性。芯片体积小、成本低、功耗小，具有硬件加密、安全可靠、组网灵活、抗毁性强等特点，非常适合于工业监控系统。

2) 射频无线收发芯片 CC2420 的内部功能模块

CC2420 的内部功能模块如图 13-17 所示。CC2420 用无线方式接收到射频信号（模拟信号），首先经过低噪声放大器（Low Noise Amplifier, LNA），然后正交下变频到 2 MHz 的中频上，形成中频信号的同相分量和正交分量。两路信号经过滤波和放大后，直接通过模/数转换器（Analog to Digital Converter, ADC）转换成数字信号。后继的处理，如自动增益控制、最终信道选择、解扩以及字节同步等，都是以数字信号的形式进行处理的。

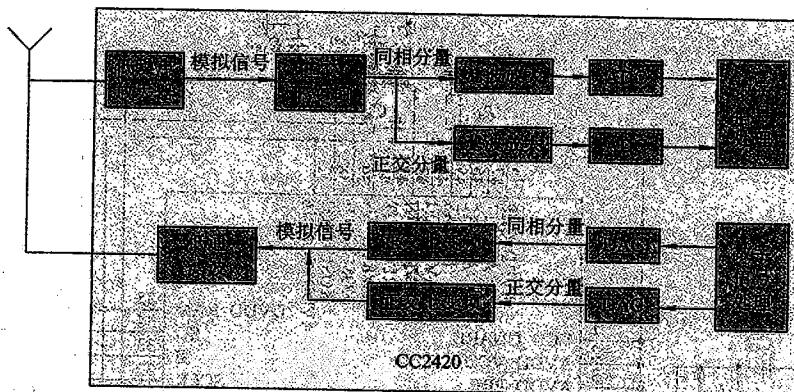


图 13-17 CC2420 无线收发芯片内部功能模块

CC2420 发送数据时, 使用直接正交上变频。基带信号的同相分量和正交分量直接被数/模转换器转换为模拟信号, 通过低通滤波器后, 直接变频到设定的信道上。

3) 无线收发芯片 CC2420 与微处理器接口

CC2420 与微处理器之间的接口包含 11 个引脚, 如图 13-18 所示。其中, CC2420 通过一个 SPI 接口与微处理器进行通信, SPI 接口由 CS_n、SI、SO 及 SCLK 四个引脚与微处理器的 P1 口相连, 实现微处理器对 CC2420 的片选、复位操作, 并将 CC2420 的状态反馈给微处理器。MSP430F149 单片机内部集成了硬件 SPI 控制器, 可以方便地与 CC2420 连接。简单的外围电路和处理器接口, 使得 CC2420 可方便地运用于各种设备。

CC2420 与传感器节点中的微处理器相连, 承担传感器节点向网络协调器无线发送传感器数据和接收来自协调器的无线指令信息的任务。

CC2420 与网络协调器——汇聚节点中的微处理器相连接, 承担汇聚节点向传感器节点无线发送指令信息和接收来自传感器节点的无线数据信息的任务。

4) CC2420 外围电路

CC2420 内部使用 1.8 V 工作电压, 因而功耗很低, 适合于电池供电的设备。外部数字 I/O 接口使用 3.3 V 电压, 这样可以保持和 3.3 V 逻辑器件的兼容性。它在片上集成了一个直流稳压器, 能够把 3.3 V 电压转换成 1.8 V 电压, 对于只有 3.3 V 电源的设备, 不需额外的电压转换电路就能正常工作。图 13-18 也给出了 CC2420 的典型外围电路连接。

CC2420 需要有 16 MHz 的参考时钟用于 250 kb/s 数据的收发。这个参考时钟可以来自外部时钟源, 也可以由内部晶体振荡器产生。如果使用外部时钟, 直接从 XOSC16_Q1 引脚引入, XOSC16_Q2 引脚保持悬空; 如果使用内部晶体振荡器, 晶体接在 XOSC16_Q1 和 XOSC16_Q2 引脚之间。CC2420 要求时钟源的精准度应该在 $\pm 40 \times 10^{-6}$ 以内。

4. 传感器节点的构成

将图 13-13、图 13-14 与图 13-18 相连接即构成图 13-11 中的一个传感器节点。本例要求有三个传感器节点。

5. 协调器——汇聚节点的构成

将图 13-15 与图 13-18 相连接, 即构成图 13-12 中的无线网络协调器——汇聚节点。

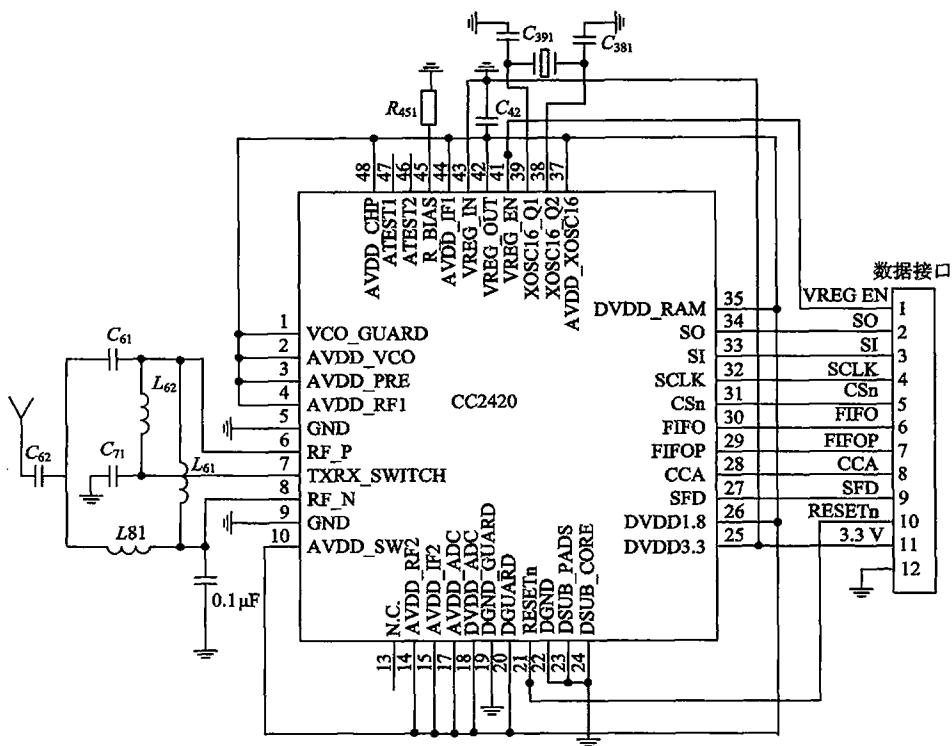


图 13-18 无线收发芯片与微处理器、典型外围电路连接图

13.3.4 [示例 13-4] 工业监测用无线传感器网络的软件设计流程举例

要求：针对[示例 13-1]中工业监测用无线传感器网络硬件系统进行软件设计；概述软件设计内容或设计流程图。

工业监测用无线传感器网络的硬件系统如图 13-1 所示，包括无线传感器网络协调器、传感器节点和 PC 三种不同的硬件，负责网络组建的不同任务。其中传感器节点分布于监测区域内，执行数据采集、处理和通信等工作。网络协调器(ZigBee 协调器)负责无线传感器网络的组建并通过 SCI 方式与 PC 进行通信，将各节点采集的数据信息汇总到 PC 上，利用 PC 与 Internet 连接，实现远程监测。因此，系统的软件设计包括网络协调器软件设计、传感器节点软件设计、PC(上位机)软件设计三部分。其中 PC 软件设计在 13.1 节中已作介绍，故本示例仅说明网络协调器与传感器节点的软件设计。

网络协调器及传感器节点的软件设计即是实现基于 IEEE 802.15.4 网络协议及 ZigBee 协议的无线传感器网络功能的软件设计过程。网络协调器与传感器节点这两种网络节点在网络中有不同功能，软件设计在符合 IEEE 802.15.4 网络协议及 ZigBee 协议的基础上有所不同，但其软件结构均符合 ZigBee 协议栈的软件结构，故先介绍基本软件结构——ZigBee 协议栈的软件结构。

1. ZigBee 协议栈软件结构

ZigBee 协议栈采用分层结构，每一层为上层提供一系列特殊的服务。本文中的网络协调器与传感器节点的软件设计均在此结构的基础上编制相应的应用程序，以对其功能进行相应配置，并根据系统需要对 ZigBee 协议栈进行相应的裁剪，其软件基本结构如图 13-19 所示。整个协议栈分为四个部分：硬件驱动设计、物理层设计、MAC 层设计、网络层设计。

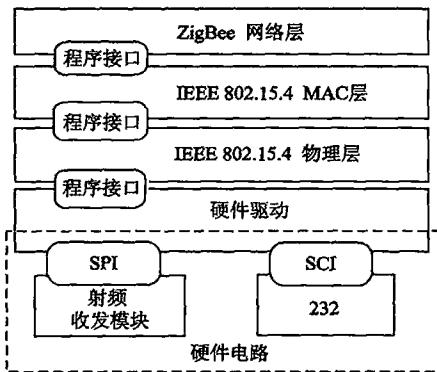


图 13-19 ZigBee 协议栈软件结构

1) 硬件驱动设计

硬件驱动定义了微处理器及无线传输模块的功能，完成硬件初始化。

硬件的操作主要是对微处理器及 CC2420 片内寄存器和 RAM 的读写，硬件驱动软件设计主要包括微处理器的 SPI 口驱动、ADC 驱动、中断驱动、RS232 驱动等。本例中，有关硬件驱动的简单操作均以宏定义的形式实现，以提高程序的效率。

在硬件驱动软件设计中，通过微处理器 SPI 口驱动软件实现对 CC2420 的操作。CC2420 是无线收发模块的核心，微处理器通过 SPI 接口访问 CC2420 内部寄存器和 RAM 存储区。在访问过程中，CC2420 是 SPI 接口的从设备，接收来自处理器的时钟信号和片选信号，并在处理器的控制下执行输入输出操作。SPI 接口的接收或发送都与时钟下降沿对齐。

2) 物理层软件设计

物理层主要通过控制 CC2420 实现数据的收发。

IEEE 802.15.4 的物理层主要负责射频信道的管理、信道能量检测及数据收发等。IEEE 802.15.4/ZigBee 通信协议为分层协议，层与层之间是通过服务接入点(SAP)相连接的。每一层都可以通过本层与下一层的 SAP 调用下层所提供的服务，同时通过与上层的 SAP 为上层提供相应服务。SAP 是层与层之间的唯一接口，而具体的服务是以通信原语的形式供上层调用的。在调用下层服务时，只需要遵循统一的原语规范，并不需要去了解如何处理原语。这样就做到了数据层与层之间的透明传输。

3) MAC 层设计

MAC 层负责实现两个功能：数据的收发和信道评估。信道评估用于判断信道是否空

闲，完成信道评估功能的是信道扫描原语。

MAC 层提供共享媒介访问控制功能，它是多跳共享的无线广播信道。这种信道具有空间复用特性，充分利用该特性可以提高信道利用率。因此 MAC 层的设计关系到整个系统的性能。IEEE 802.15.4 MAC 层定义原语操作来实现与物理层和网络层的接口，每一个原语对应一个程序。MAC 层和物理层一样也分为数据操作和管理信息操作。对于数据的收发，是通过数据操作原语实现的；设计管理操作完成系统内部消息的管理。

4) 网络层设计

网络层负责完成网络组建和路由管理任务。

网络层协议是通信网络协议的核心，它的主要目标是建立并维护一个无线传感器网络，并根据节点地址信息决定是否接收和发送数据包，同时向上层（应用层）提供一个简单易用的软件接口，实现各传感器与观察者之间的通信。本例中组建的是简单的星型 ZigBee 网络，网络中存在网络协调器和网络传感器节点，这两种节点的网络层功能分别着重于协调器对网络的组建及传感器节点申请网络加入两个方面。

(1) 网络的组建。在网络协调器组网之前要先调用 MAC 层信道扫描子程序（原语）进行信道扫描。处理器根据返回的扫描结果给出信道使用与空闲情况并选取信道，然后设置系统信息：PAN ID 及逻辑地址。此时协调器可以接收入网请求，给新加入的设备分配逻辑地址。组网子程序流程图如图 13-20 所示。



图 13-20 协调器组网子程序流程图

(2) 入网申请。网络传感器节点上电，执行了初始化过程以后，便开始搜索其无线覆盖范围之内的网络信息，如果存在一个已由网络协调器建立的无线网络，则执行网络加入过程。同时，也存在节点发出入网请求被协调器拒绝的情形，例如地址分配满，此时节点会重新选择其他存在的无线网络发起关联请求，若仍入网失败，则回到休眠状态。网络传感器节点申请入网子程序流程图如图 13-21 所示。

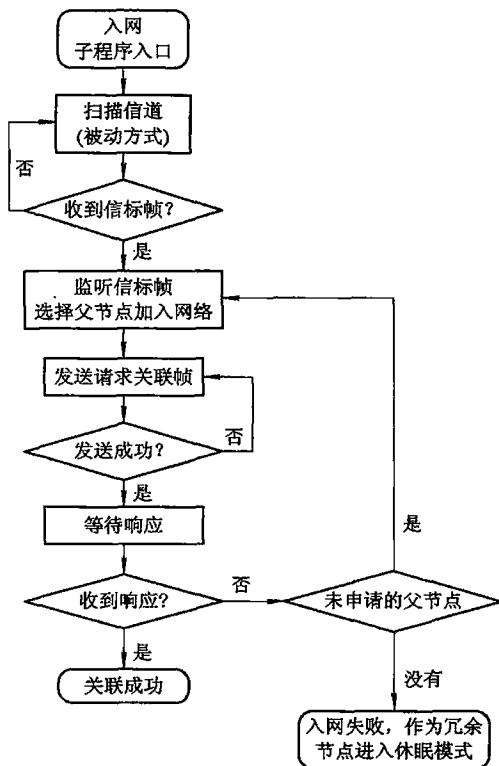


图 13-21 网络传感器节点申请入网子程序流程图

2. 网络协调器软件设计

网络协调器在整个无线传感器网络中负责网络的组建并通过 SCI 方式与 PC 进行通信，接收 PC 指令并依照相应指令将各节点采集的数据信息汇总到 PC 上。协调器软件设计主要包括系统的初始化程序、网络的组建程序及 SCI 通信程序。

1) 网络协调器的主程序流程

网络协调器上电后，首先进行系统硬件和软件的初始化，初始化结构后，开启全部中断，调用组网程序，组建一个无线 ZigBee 网络，等待节点的加入。判断网络组建成功后，设置网络的工作状态，并进入等待 PC(人机接口)命令的状态。若接收到完整的串口命令(命令格式由程序设定)，首先判断命令类型。如果串口命令为“D”(Debug)，则对网络进行调试，即检索网络的节点个数及其状态；如果串口命令为“S”(Send)，则调用数据传输子程序，向节点传输数据。串口接收数据调用串口中断程序完成。主程序流程如图 13-22 所示。

2) 系统初始化

初始化过程包括硬件初始化、软件初始化、微处理器时钟初始化、SCI(串口)通信初始化、SPI 通信初始化、无线传输模块 CC2420 初始化、网络地址初始化和 ZigBee 协议栈初始化。

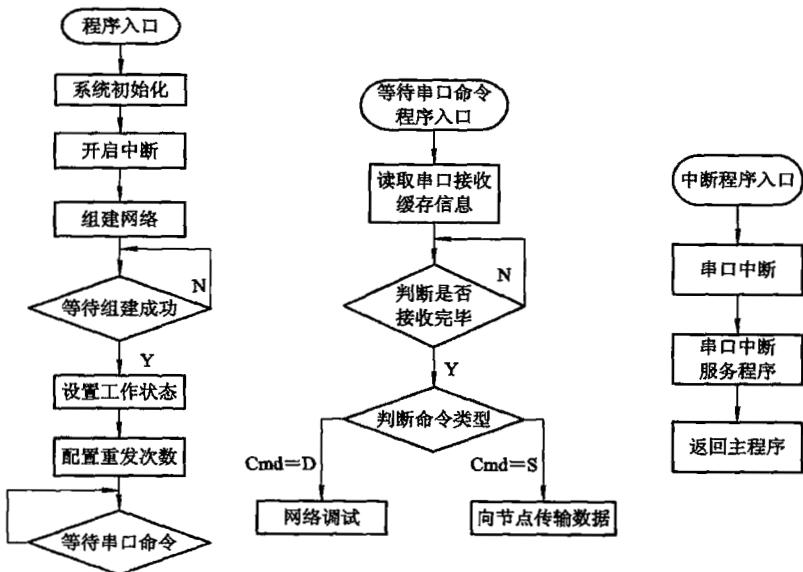


图 13-22 网络协调器主程序流程图

3) SCI 通信程序

SCI 通信采用中断方式实现。本例中，串口中断程序的主要作用为判断命令的格式，接收指令并提供给主程序。当串口中断产生时，软件首先判断指令的第一位是否为“#”，是则继续接收指令；不是则退出中断服务程序。然后判断指令是否是“0X0d”，不是则表示本条指令未结束，继续接收；是则表示本条指令接收完毕，放入接收缓存，退出中断程序。放入串口接收缓存的数据将实时被主程序读取使用。其流程如图 13-23 所示。SCI 通信（串口通信）是协调器软件设计很重要的一个环节，网络操作命令的接收、网络节点数据的上传、网络工作状态的查询都是通过串口通信实现的。

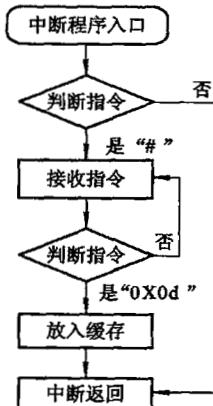


图 13-23 实现 SCI 通信串口中断服务程序流程图

3. 传感器节点程序设计

传感器节点软件设计主要完成数据采集以及通过无线通信模块将采集到的数据无线传送至协调器，通过协调器传送至上位机。节点遵循睡眠—唤醒—正常工作的工作模式。在睡眠状态下，处理器停止工作，而 SRAM、SPI 端口、定时器以及中断系统继续工作，无线模块处于低电流接收状态。传感器节点接收到协调器传输的上位机指令后，根据指令的内容，将相应的数据打包通过无线网络向协调器传输。传感器节点的具体软件流程如图 13-24 所示。

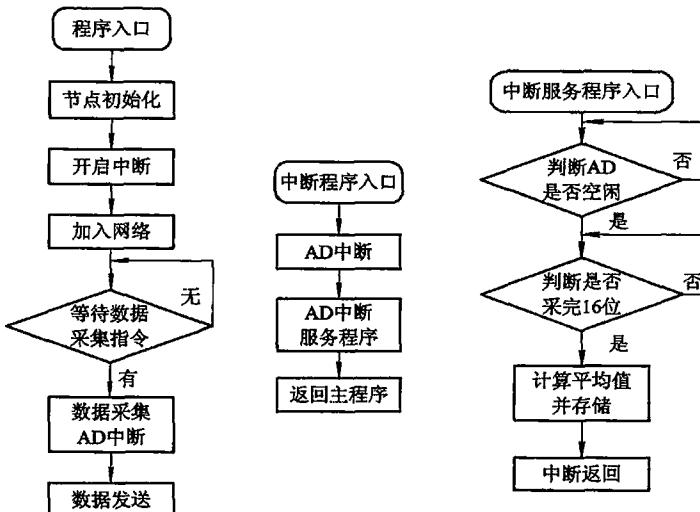


图 13-24 传感器节点程序流程图

传感器节点的初始化与网络协调器初始化流程基本相同，区别在于：节点中无 SCI 通信初始化，增加了 ADC 初始化；另外，节点无需网络地址的初始化，节点的地址由网络协调器自动分配。节点上电后，自动搜索网络，加入网络后，即等待协调器发送上位机指令。如果接收到上位机索要监测数据的指令，则进入 AD 中断服务程序。最后，将采集到的数据打包发送给协调器，由协调器上传给上位机。

参 考 文 献

- [1] 王颖. 远程无线传感器网络监测系统的研究与实现. 西安交通大学硕士学位论文, 2008
- [2] Edgar H, Callaway, Jr. 无线传感器网络: 体系结构与协议 [M]. 王永斌, 屈晓旭, 译. 北京: 电子工业出版社, 2007: 35-40
- [3] 瞿雷, 刘盛德, 胡咸斌. ZigBee 技术及应用 [M]. 北京: 北京航空航天大学出版社, 2007

- [4] 于海斌, 等. 智能无线传感器网络系统[M]. 北京: 科学出版社, 2006: 151~158
- [5] ZigBee Alliance. Network Specification. Version 1.00 ZigBee Document 02130r10, December 14th, 2006
- [6] 钱凌影, 刘君华. 实验数据远程采集和传输技术[C]. 2005 年全国虚拟仪器学术交流大会论文集. 北京: 电子测量技术杂志社, 2005: 79~80
- [7] 林继鹏, 刘君华. 互联网上仪器及开发技术[C]. 2005 年全国虚拟仪器学术交流大会论文集, 2005
- [8] Lingyin Qian, Junhua Liu. The Realization Technology of Remote Measurement & Control Laboratory. Proceeding of Regional Inter-University Postgraduate Electrical and Electronic Engineering Conference[C]. Hong Kong: City University of Hong Kong, 2005, 52
- [9] Cui Xiangwu, Liu Junhua. Remote Measurement of Carbon Oxide Based on the Technology of Virtual Instrument[C]. Proceedings of International Postgraduate Conference on Electronic Engineering. Xi'an, July12~15, 2004: 467~469
- [10] Texas Instruments Inc. Chipcon AS Smart RF CC2420 Preliminary Datasheet(rev 1.2), 2004
- [11] 夏刚, 陈曦, 赵千川. 无线传感器网络及应用简介[J]. 自动化博览, 2004, 21(1): 34~37
- [12] 郭会军, 贾惠芹, 刘君华. 基于 ActiveX 的网络虚拟仪器实验室[J]. 西安交通大学学报, 2003, 37(2): 219~220
- [13] Sichitiu M L, Veerarittiphan C. Simple, accurate time synchronization for wireless sensor networks[A]. IEEE Wireless Communications and Networking Conference[C]. WCNC2003, New Orleans
- [14] 王颖, 张勇, 刘君华. 基于 LabVIEW 的无线传感器网络监测平台设计[C]. 第三届虚拟仪器学术交流大会, 2005