

高

等

学

校

教

材

计算机在材料科学与工程中的应用

杨明波 胡红军 唐丽文 编



化学工业出版社

本书是“计算机在材料科学与工程中的应用”系列教材之一，主要介绍计算机在材料科学与工程中的应用。全书共分8章，主要内容包括：计算机组成原理、操作系统、数据库系统、网络通信、材料数据库、材料模拟、材料工程应用等。本书可作为高等院校材料科学与工程类专业及相关专业的教材，也可供从事材料科学与工程工作的工程技术人员参考。

高等学校教材

计算机在材料科学与工程中的应用

杨明波 胡红军 唐丽文 编

本书在编写过程中参考了国内外有关文献，并得到了许多同行专家的指导和帮助，在此表示衷心的感谢。由于编者水平有限，书中难免存在不足之处，恳请读者批评指正。

本书可作为高等院校材料科学与工程类专业及相关专业的教材，也可供从事材料科学与工程工作的工程技术人员参考。

本书由化学工业出版社出版，书号为978-7-122-01200-2。



化学工业出版社

北京

本书为高等学校教材，主要介绍计算机在材料科学与工程中的应用。全书共分9章，其中第1章主要介绍材料科学与工程的基本知识和计算机在材料科学与工程中的应用概况；第2章主要介绍材料科学与工程中数据的计算机处理；第3章主要介绍数学模型的建立及数值求解；第4章主要介绍材料科学与工程中典型物理场的数值模拟；第5章主要介绍 ANSYS 软件及其在材料科学与工程中的应用；第6章主要介绍计算机在相图计算及材料设计中的应用；第7章主要介绍数据库及专家系统在材料科学与工程中的应用；第8章主要介绍神经网络及 Matlab 软件在材料科学与工程中的应用；第9章主要介绍材料加工成形过程的计算机模拟。

本书可作为材料科学与工程专业本科生及研究生的教学用书，也可供从事材料科学及材料加工研究、开发及应用的工程技术人员参考。

图书在版编目(CIP)数据

计算机在材料科学与工程中的应用 / 杨明波, 胡红军,
唐丽文编. —北京: 化学工业出版社, 2008.1
高等学校教材
ISBN 978-7-122-01706-2

I. 计… II. ①杨…②胡…③唐… III. 计算机应
用-材料科学-高等学校-教材 IV. TB3-39

中国版本图书馆 CIP 数据核字 (2007) 第 196840 号

责任编辑: 陶艳玲

文字编辑: 冯国庆

责任校对: 李林

装帧设计: 史利平

出版发行: 化学工业出版社 (北京市东城区青年湖南街 13 号 邮政编码 100011)

印 装: 化学工业出版社印刷厂

787mm×1092mm 1/16 印张 14 $\frac{3}{4}$ 字数 312 千字 2008 年 2 月北京第 1 版第 1 次印刷

购书咨询: 010-64518888 (传真: 010-64519686) 售后服务: 010-64518899

网 址: <http://www.cip.com.cn>

凡购买本书, 如有缺损质量问题, 本社销售中心负责调换。

定 价: 29.00 元

版权所有 违者必究

前 言

计算机作为一种现代工具在材料科学与工程中的应用已越来越广泛,从而极大地促进和推动了材料科学与工程研究的深入和发展。本书立足“材料科学与工程”一级学科,系统介绍了计算机在材料科学与工程中的应用,使读者初步掌握如何在材料科学与工程的学习及研究中更好地利用计算机这一工具。本书的最大特点在于注重理论知识讲解的同时,结合计算机在材料科学与工程中的应用实例讲解来培养学生的实际动手能力和创新意识。由于考虑到不同学校前导课程及教学进度的不同、本课程的学时不同、学生的计算机应用能力和材料科学与工程专业基础知识水平参差不齐,因此本书在编写时既注重基础理论知识的介绍,同时又注重实际例子的讲解,各章节涉及的内容较多、面较广,教师在教学时可根据本校的实际情况选用。

本书由重庆工学院杨明波、胡红军和唐丽文编写。其中第1章~第4章由杨明波和唐丽文编写,第5章~第9章由胡红军编写。全书由重庆大学张丁非教授主审。

由于条件所限,本书未能将所有参考文献一一列出,在此对所有参考文献的作者表示衷心的感谢。此外,本书在编写过程中得到了重庆大学汤爱涛副教授的大力指导,并得到了重庆工学院材料科学与工程学院及教务处的大力支持,在此也表示感谢。

由于计算机在材料科学与工程中的应用非常广泛,并且计算机技术的发展日新月异,材料科学与工程中新方法和新工艺不断出现,加之编者学识有限,书中难免有不足之处,敬请读者批评指正。

编 者
2007年10月

目 录

第 1 章 绪论	1
1.1 材料科学与工程的概念	1
1.1.1 材料的作用及分类	1
1.1.2 材料科学与工程的内容	1
1.2 计算机在材料科学与工程中的应用简介	2
1.2.1 用于新材料和新合金的设计	2
1.2.2 用于材料科学研究中的模拟	3
1.2.3 用于材料工艺过程的优化及自动控制	5
1.2.4 用于材料组成和微观结构的表征	5
1.2.5 用于数据和图像处理及其他	5
第 2 章 材料科学与工程中数据的计算机处理	7
2.1 数据处理的基本理论	7
2.1.1 曲线拟合与最小二乘法	7
2.1.2 插值法	17
2.2 Origin 软件在数据处理中的应用	18
2.2.1 Origin 软件介绍	18
2.2.2 Origin 在数据处理中的应用实例	18
2.3 Excel 软件在数据处理中的应用	28
2.3.1 计算功能	28
2.3.2 参数估计	30
思考题与上机操作实验题	35
第 3 章 数学模型的建立及数值求解	39
3.1 数学模型介绍	39
3.1.1 数学模型的含义	39
3.1.2 数学模型分类	39
3.1.3 数学模型的建立步骤	40
3.1.4 其他概念	41
3.2 数学模型的建立方法及实例	41
3.2.1 理论分析法	41
3.2.2 数值模拟的方法	43
3.2.3 类比分析法	47
3.2.4 数据分析法	49

3.2.5 利用计算机软件 (Origin 软件) 建立数学模型	50
3.3 数学模型的数值求解	53
3.3.1 常用的计算方法	54
3.3.2 有限差分法	58
3.3.3 有限元法	64
思考题与上机操作实验题	78
第 4 章 材料科学与工程中典型物理场的数值模拟	79
4.1 温度场数学模型及求解	79
4.1.1 温度场的基本知识	79
4.1.2 温度场数学模型的建立	82
4.1.3 温度场的有限差分求解	86
4.2 铸件充型过程的通用数学模型	100
4.3 浓度场的通用数学模型与计算	102
4.3.1 菲克第一定律	102
4.3.2 菲克第二定律	103
4.3.3 渗碳浓度场的有限差分求解	103
思考题与上机操作实验题	110
第 5 章 ANSYS 软件及其在材料科学与工程中的应用	112
5.1 ANSYS 软件介绍	112
5.2 ANSYS 在材料科学与工程中的应用实例	113
5.2.1 铸件温度场分布及模具内部应力场分布	113
5.2.2 热处理淬火过程的温度场模拟	118
5.2.3 热喷涂过程的温度场和应力场分析	126
5.2.4 铝材挤压过程的应力应变分析	131
5.2.5 焊接过程中的应力和变形分析	138
思考题与上机操作实验题	158
第 6 章 计算机在相图计算及材料设计中的应用	162
6.1 相图计算的理论与实践	162
6.1.1 相图计算过程及特点	162
6.1.2 THERMO-CALC 相图计算软件介绍及使用	164
6.2 材料设计的理论与实践	167
6.2.1 材料设计介绍	167
6.2.2 Material studio 软件及其在材料设计中的应用	172
思考题与上机操作实验题	176
第 7 章 数据库及专家系统在材料科学与工程中的应用	177
7.1 数据库在材料科学与工程中的应用	177

7.1.1 数据库的组成与结构	177
7.1.2 材料科学与工程数据库的发展	178
7.2 专家系统及其在材料科学与工程中的应用	179
7.2.1 专家系统基本知识	179
7.2.2 材料科学与工程中的部分专家系统介绍	181
思考题	185
第 8 章 神经网络及 Matlab 软件在材料科学与工程中的应用	186
8.1 神经网络与 Matlab 软件简介	186
8.1.1 神经网络介绍	186
8.1.2 线性神经网络	189
8.1.3 BP 神经网络	190
8.2 Matlab 软件在材料科学与工程中的应用	193
8.2.1 数表公式化——曲线拟合	193
8.2.2 焊接神经网络模型建立	194
8.2.3 混凝土板瞬态温度场算法分析	195
思考题	199
第 9 章 材料加工成形过程的计算机模拟	200
9.1 概述	200
9.2 MoldFlow 塑料注塑成形过程模拟软件介绍及使用	201
9.2.1 MoldFlow 软件介绍	201
9.2.2 MoldFlow 软件操作步骤及分析实例	202
9.3 Deform 体积成形分析软件介绍及使用	208
9.3.1 Deform 分析软件介绍	208
9.3.2 Deform 软件操作步骤及实例	209
9.4 ProCAST 铸造成形过程模拟软件介绍及使用	217
9.4.1 ProCAST 软件介绍	217
9.4.2 ProCAST 软件的操作步骤	220
9.4.3 ProCAST 软件应用实例	224
思考题与上机操作实验题	229
参考文献	230

第1章 绪论

1.1 材料科学与工程的概念

1.1.1 材料的作用及分类

材料是人类生产和生活水平提高的物质基础，是人类文明的重要支柱和进步的里程碑。材料的进步取决于社会生产力和科学技术的进步，同时材料的发展又推动了社会经济和科学技术的发展。因此，从某种角度讲，人类的文明史同时也是一部材料发展史。但是人类在漫长的历史发展中大都是依靠自然的恩赐，仅仅停留在利用天然材料的状态。自从19世纪以来，随着社会的发展和科学技术的进步，人们对材料不断提出新的要求，有些要求完全超出天然材料所能提供的性能，从而促进了人类开始对材料从依靠到创造的转变，对材料的认识也逐渐发展到形成一门科学。如图1-1所示，人类在经历了石器、铜器和铁器时代后，人们在20世纪60年代把材料、能源和信息称为当代文明的三大支柱；在20世纪70年代又把新材料、信息技术和生物技术看成新技术革命的主要标志。这表明材料的发展与社会文明的进步有着非常密切的关系。而现代科学技术的发展历程也充分证明了这一点。目前，人们已经逐渐掌握了材料的组成、结构和性能之间的内在关系，能够按照使用要求对材料的性能进行设计创造。在20世纪下半叶逐渐形成了以新材料技术为基础的信息技术、新能源技术、生物工程、空间技术和海洋开发技术的新技术群，更使材料科学得到飞速发展。

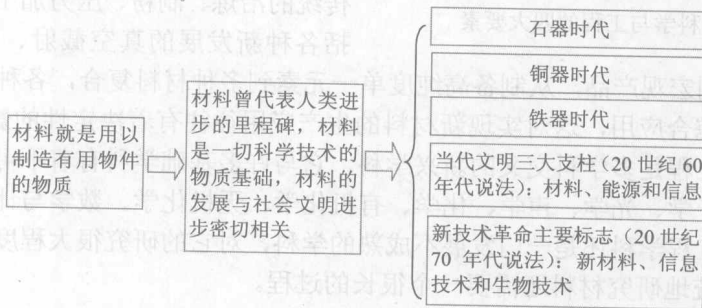


图1-1 材料的作用

材料分类方法很多，根据其组成与结构可以分为金属材料、无机非金属材料、有机高分子材料和复合材料等。根据其性能特征和作用分为结构材料和功能材料。根据用途还可以分为建筑材料、能源材料、电子材料、耐火材料、医用材料和耐腐蚀材料等。图1-2显示了材料的简单分类。

1.1.2 材料科学与工程的内容

材料科学与工程是关于材料组成、结构、制备工艺与其性能及使用过程间相互关系的知识开发及应用的学科。研究的内容包括材料的组织、结构、杂质、缺陷与性能之间的关系，材料的形成机理和制备方法，材料在加工、使用过程中的变化和失效机理，材料性能的测试

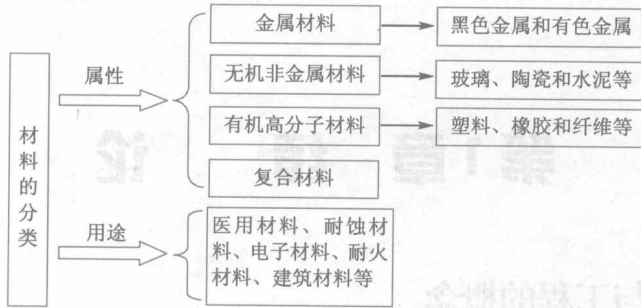


图 1-2 材料的分类

和材料的工程应用等。与其他的事物发展过程一样，材料科学的发展经历了一个由简单到复杂，由以经验为主到以材料知识为主，逐步形成了独立的材料科学与工程体系的学科。其中材料科学侧重于发现和揭示四要素之间的关系，提出新概念和新理论；而材料工程则侧重于寻求新手段实现新材料的设计思想并使之投入应用，两者相互依存。图 1-3 显示了材料科学与工程的四大要素

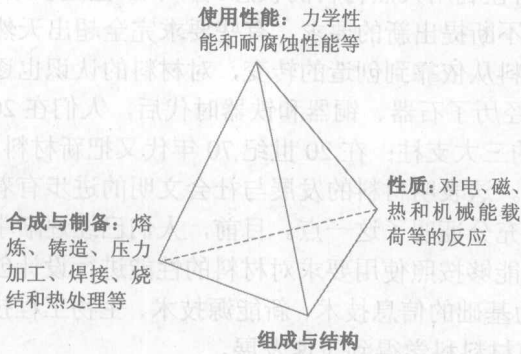


图 1-3 材料科学与工程的四大要素

与工程的四大要素，图中材料的性质主要是指材料对电、磁、光、热、机械载荷的反应，而这些性质主要取决于材料的组成与结构。使用性能是材料在使用状态下表现出来的行为。它与设计和工程环境密切相关，有些材料在实验室环境下表现出很好的性能，但在特定的使用条件下，如氧化与摩擦、疲劳及其他复杂载荷条件下，就不能满足使用要求。此外，使用性能还包括可靠性、耐用性、寿命预测和延寿措施。而材料的合成与制备过程内容很多，包括传统的冶炼、制粉、压力加工和焊接等，也包括各种新发展的真空喷射、气相沉积等新工

艺；从微观水平到宏观产品，从制备高纯度单一元素到多种材料复合，各种化学、物理、机械加工方法均应综合应用，这对实现新材料的生产应用往往有着决定性的影响。

材料科学与工程是多学科交叉的新兴学科，它与许多基础学科有着不可分割的联系，如固体物理学、电子学、光学、声学、化学、有机化学、无机化学、数学与计算机技术等。目前，材料科学与工程学科还是一门发展不成熟的学科，对它的研究很大程度上还基于事实和经验的积累，系统地研究材料还需要一个很长的过程。

1.2 计算机在材料科学与工程中的应用简介

计算机作为一种现代工具，在当今世界的各个领域日益发挥巨大的作用，它已渗透到各门学科领域以及日常生活中成为现代化的标志。在材料科学与工程领域，计算机也正在逐渐成为极其重要的工具，计算机在材料科学与工程中的应用正是材料研究和开发飞速发展的重要原因之一。目前，计算机在材料科学与工程中的应用主要表现在以下几个方面。

1.2.1 用于新材料和新合金的设计

新材料和新合金的设计开发，长期以来采用的是配方方法，有人比作“炒菜式”的方法。一般需经对成分-组织-性能关系的调整做多次反复实验，即“炒作”才能获得较满意的结果。

这种方法有相当大的盲目性，费功、费时、经济损耗大。为此，人们期望从实验比较和总结归纳的研制方式走向演绎计算的方法，而计算机技术的飞速发展恰恰迎合了这一发展趋势，即按所需材料性能来设计、制备新材料和新合金，并使所设计的合金成分、组织或工艺达到最佳配合。这种设计的基本原理是基于已有的大量数据和经验事实出发，利用已有的各种不同结构层次的数学模型，如合金的成分、组织、结构与性能关系的数学模型及相关数据理论，通过计算机运算对比和推理思维来完成优选新合金和新材料的设计过程，其中引入了数学上的最优化理论来获得最佳方案的材料配方及生产工艺。

近年来，又有人提出材料设计的专家系统。如图1-4所示，在专家系统中两个最重要的部分是材料数据库和材料知识库。其中材料数据库中存储的是具体有关材料的数据值，它只能进行查询而不能推理，而材料知识库存储的是规则。当从数据库中查询不到相应的性能值时，知识库却能通过推理机构以一定的可信度给出性能的估算值，从而实现性能预测功能。同时，也可用知识库进行组分和工艺设计。目前，人工神经网络研究的突破又为新材料和新合金的设计提供了一种新的思路。人工神经网络 ANN 是用工程技术手段模拟生物神经网络的结构和功能特征的一类人工系统，它的特点是：①既可解决定性问题又可解决定量问题；②擅长于处理复杂的、多维的非线性问题③具有自学习能力，即从已有的实验数据中自动总结规律，而不依赖于“专家”头脑。目前进行材料设计的方法都涉及材料的组分、工艺性能和使用之间的关系，而当前材料特别是新材料的内在规律尚不甚清楚。人工神经网络的自学习功能正好适用于材料设计或性能预测这一类问题。

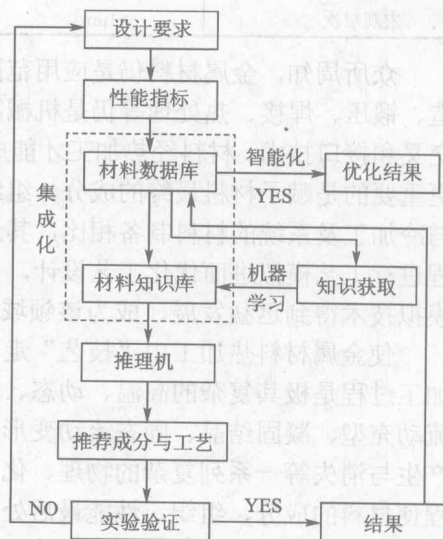


图 1-4 材料设计专家系统流程图

1.2.2 用于材料科学研究中的模拟

采用各种新颖算法的模拟技术，并结合运算功能强大的计算机，人们能够细致、精确地对物质内部状况进行研究。这导致计算机模拟在材料科学中的应用越来越广泛，并由此产生了一门新的材料研究分支——计算材料科学（computational materials science）。采用模拟技术进行材料研究的优势在于它不但能够模拟各类实验过程，了解材料的内部微观性质及其宏观力学行为，并且在没有实际制备出这些新材料前就能预测它们的性能，为设计出优异性能的新型结构材料提供强有力的理论指导。材料科学研究中的模拟“实验”比实物实验更高效、经济、灵活，并且在实验很困难或不能进行的场合仍可进行模拟“实验”，特别是在对微观状态与过程的了解方面，模拟“实验”更有其独特性甚至有不可替代的作用。一般而言，材料研究主要针对三类不同的尺度范围。

- ① 原子结构层次 主要是凝聚态物理学家和量子化学家处理这一微观尺度范围。
- ② 介观层次 即介于原子和宏观之间的中间尺度，在这一尺度范围主要由材料学家、冶金学家、陶瓷学家处理。
- ③ 最后是宏观尺寸 此时大块材料的性能被用作制造过程，机械工程师、制造工程师等分别在这一尺度范围进行处理。

既然材料性质的研究是在不同尺度层次上进行的，那么，计算机模拟也可根据模拟对象

的尺度范围而划分为若干层次,见表1-1。下面以热加工过程的计算机模拟为例,简单介绍宏观层次的模拟。

表 1-1 计算机模拟的层次

模拟层次	空间尺度	模拟对象
电子层次	0.1~1nm	电子结构
原子分子层次	1~10nm	结构、力学性能、热力学和动力学性能
微观结构层次	约1 μ m	晶粒生长、烧结、位错、粗化和织构
宏观层次	>1 μ m	铸造、焊接和锻造和化学气相沉积等

众所周知,金属材料仍是应用范围最为广泛的机械工程材料,而材料的热加工工艺如铸造、锻压、焊接、热处理等仍是机械制造业重要的加工工序,其也是材料与制造两大行业的交叉和接口技术。材料经热加工才能成为零件或毛坯,它不仅使材料获得一定的形状、尺寸,更重要的是赋予材料最终的成分、组织与性能。由于热加工兼有成形和改性两个功能,因而与冷加工及系统的材料制备相比,其过程质量控制具有更大的难度。因此,对材料热加工过程进行工艺模拟进而优化工艺设计,具有更为迫切的需求。近二十多年来,材料热加工工艺模拟技术得到迅猛发展,成为该领域最为活跃的研究热点及技术前沿。

使金属材料热加工由“技艺”走向“科学”,彻底改变热加工的落后面貌。金属材料热加工过程是极其复杂的高温、动态、瞬时过程,难以直接观察。在这个过程中,材料经液态流动充型、凝固结晶、固态流动变形、相变、再结晶和重结晶等多种微观组织变化及缺陷的产生与消失等一系列复杂的物理、化学、冶金变化而最后成为毛坯或构件。必须控制这个过程使材料的成分、组织、性能最后处于最佳状态,必须使缺陷减到最小或将它驱赶到危害最小的地方去。但这一切都不能直接观察到,间接测试也十分困难。长期以来,基础学科的理论知识难以定量指导材料加工过程,材料热加工工艺设计只能建立在“经验”基础上。近年来,随着试验技术及计算机技术的发展和材料成形理论的深化,材料成形过程工艺设计方法正在发生着质的改变。材料热加工工艺模拟技术就是在材料热加工理论指导下,通过数值模拟和物理模拟,在试验室动态仿真材料的热加工过程,预测实际工艺条件下材料的最后组织、性能和质量,进而实现热加工工艺的优化设计。它将使材料热加工沿此方向由“技艺”走向“科学”,并为实现虚拟制造迈出第一步,使机械制造业的技术水平产生质的飞跃。

计算机模拟是预测并保证材料热加工过程质量的先进手段,特别对确保关键大件一次制造成功,具有重大的应用背景和效益。我国重大机电设备研制、生产的一个难点是大件制造;大件制造的关键又是热加工。我国在2015年以前,水电、火电、核电、冶金、矿山、石化等重大机电设备对关键大件制造均有迫切的需求。以三峡水电机组为例,单机容量达70万千瓦,五大部件(转轮、蜗壳、主轴、座环、顶盖)的质量和尺寸均居世界第一。其转轮直径达9.8m,质量达500t,采用铸焊结构,制造难度很大。由于大件形大体重,品种多,批量小,生产周期长,造价高,迫切要求“一次制造成功”,一旦报废,在经济和时间上都损失惨重,无法挽回。由于传统的热加工工艺设计只能凭经验,采用试错法(test and error method)无法对材料内部宏观、微观结构的演化进行理想控制,因而发生多次大件报废的惨痛事故,投入使用的大件,也难以消除缩孔、缩松、夹杂、偏析、热裂、冷裂、混晶等缺陷,很多大件“带伤”运行。建立在工艺模拟、优化基础上的热加工工艺设计技术,可以将“隐患”消灭在计算机拟实加工的反复比较中,从而确保关键大件一次制造成功。这已为国内外不少应用实例所证实。

经多年研究开发,已经形成一批热加工工艺商业软件,主要有MAGMA、PROCAST、SIMULOR、SOLDIA、SOLSIAR、AFS Solidification System3D(铸造)、DEFORM

AUTOFORGE、SUPERFORGE（体积塑性成形）、DYNA3D、PAM-STAMP、ANSYS（板料塑性成形）、ABAQUS（焊接）等。已在铸造、锻压行业生产中得到较广泛应用，如日本已有约10%铸造工厂采用此项技术；美国福特、通用汽车公司在开发新车型时，已将板材冲压过程的数值模拟作为一个重要技术环节；法国应用此技术对400t重的核电转子锻件的锻造工艺进行了校核、优化，确保了一次制造成功。

数值模拟已逐步成为新工艺研究开发的重要手段和方法。在工业发达国家（如美国），应用商业软件进行数值模拟已成为与实验同样重要的实现技术创新、开发新工艺的基本研究手段。选择合适的商业软件为软件平台，结合具体问题，进行改进提高研究，逐步成为多、快、好、省的研究方法。具体方式有：①对现有软件的某些技术问题进行理论研究；②为解决具体问题插入自编软件模块；③应用理论分析补偿法、实验补偿法等，找出并消除商用软件的误差，使模拟结果更精确；④与软件公司合作，增加软件功能，实现软件升级。

我国目前的研究工作，有一些已接近或达到世界先进水平。如焊接凝固裂纹精确评价技术及开裂判据；焊接氢致裂纹精确评价技术及开裂判据；伴随有动态再结晶过程的金属热塑性本构关系；三维塑性成形晶粒度演化模拟及组织预测；板料成形模拟的半显示时间积分的有限元算法；金属材料准固相区热应力本构方程及模拟仿真；电渣熔铸工艺过程三维模拟及优化；球墨铸铁及镍基合金的微观组织模拟；固态相变条件下弹塑应力场应变分量的理论分析及模拟；并行工程环境下金属热成形模拟仿真。

1.2.3 用于材料工艺过程的优化及自动控制

材料工艺过程的优化及自动控制、材料加工技术的发展主要体现在控制技术的飞速发展。微型计算机和可编程控制器在材料加工过程中的应用正体现了这种发展和趋势。在材料加工过程中应用计算机不仅能减轻劳动强度，而且能改善产品质量和精度，提高产量。

用计算机可以对材料加工工艺过程进行优化控制。如在计算机模拟和对工艺过程的数学模拟进行研究的基础上，可以用计算机对渗碳、渗氮全过程进行控制，可以用计算机精密控制注塑机的注射速度。计算机技术、微电子技术和自动控制技术相结合，使工艺设备、检测手段的准确性和精确度等大大提高。以在热处理中的应用为例，计算机首先应用于炉温控制，其后迅速扩展到气氛控制，真空热处理控制，气体渗碳、渗氮控制，离子化学热处理控制，激光热处理的控制，渗碳、淬火、清洗和回火的整个生产过程的控制等。控制技术也由最初的简单顺序控制发展到数学模型在线控制和统计过程控制，由分散的个别设备的控制发展到计算机综合管理与控制，控制水平提高，可靠性得到充分保证。计算机在材料加工中的应用不仅可以减轻劳动强度，而且可以改善产品质量和精度，提高产量。

1.2.4 用于材料组成和微观结构的表征

目前，材料组成和结构表征研究主要采用各种大型分析设备进行，如扫描电镜（SEM）、透射电镜（TEM）、分析电镜（AEM）、扫描探针显微镜（SPM）等各种电镜，以及可见光谱、红外光谱、拉曼光谱、原子吸收光谱、等离子体发射光谱、荧光光谱等各种谱仪和X射线衍射、电子衍射、中子衍射等各种衍射仪。这些大型分析设备几乎无一例外的是在计算机的控制之下完成分析工作的。这些分析设备提供有不同的分析模拟软件以及相应的数据库，而且这些分析模拟软件的功能非常强大，大大减轻了数据处理的工作量，可以给出各种图表。

1.2.5 用于数据和图像处理及其他

材料科学研究在实验中可以获得大量的实验数据，借助计算机的存储设备，可以大量保存数据，并对这些数据进行处理（计算、绘图，拟合分析）和快速查询等。同时，材料的性

第2章 材料科学与工程中数据的计算机处理

材料科学与工程研究中获得的大量原始数据需要经过处理才能得到所需要的结果并加以保存, 计算机的飞速发展使得不但可以利用计算机大量保存并方便快速查找实验数据, 而且可以对数据进行进一步的后续处理(如计算、绘图、拟合分析等)。

本章在论述数据处理基本理论的基础上(包括最小二乘法 and 插值法), 引入了数据处理功能强大的 Origin7.0 软件和 Excel2003 软件, 结合大量的应用实例, 介绍了软件在材料科学与工程中的应用方法和步骤。

2.1 数据处理的基本理论

2.1.1 曲线拟合与最小二乘法

在科学研究和实际工作中, 常常会遇到这样的问题: 给定两个变量 x 、 y 的 m 组实验数据 $(x_1, y_1), (x_2, y_2), \dots, (x_m, y_m)$, 如何从中找出这两个变量间的函数关系的近似解析表达式(也称为经验公式), 使得能对 x 与 y 之间的除了实验数据外的对应情况作出某种判断。

这样的问题一般可以分为两类: 一类是对 x 与 y 之间所存在的对应规律一无所知, 这时要从实验数据中找出切合实际的近似解析表达式是相当困难的, 俗称这类问题为黑箱问题; 另一类是依据对问题所做的分析, 通过数学建模或者通过整理归纳实验数据, 能够判定出 x 与 y 之间满足或大体上满足某种类型的函数关系式 $y=f(x, a)$, 其中 $a=(a_1, a_2, \dots, a_n)$, 是 n 个待定的参数, 这些参数的值可以通过 m 组实验数据来确定(一般要求 $m > n$), 这类问题称为灰箱问题。解决灰箱问题的原则通常是使拟合函数在 x_i 处的值与实验数值的偏差平方

和最小, 即 $\sum_{i=1}^n [f(x_i, a) - y_i]^2$ 取得最小值。这种在方差意义下

对实验数据实现最佳拟合的方法称为“最小二乘法”, a_1, a_2, \dots, a_n 称为最小二乘解, $y=f(x, a)$ 称为拟合函数。

曲线拟合(fitting a curve): 根据一组数据, 即若若干点, 要求确定一个函数, 即曲线, 使这些点与曲线总体来说尽量接近。曲线拟合的目的: 根据实验获得的数据去建立因变量与自变量之间有效的经验函数关系, 为进一步的深入研究提供线索。

已知数据对 $(x_j, y_j)(j=1, 2, \dots, n)$ 如图 2-1 所示, 求多项式

$$y = f(x) \quad (2-1)$$

使得

$$\delta_i = y_i - \hat{y}_i$$

$$Q = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n [y_i - f(x)]^2 \quad (2-2)$$

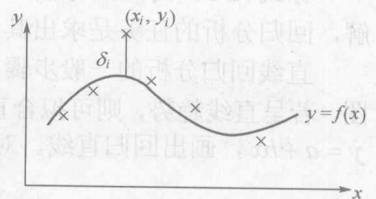


图 2-1 最小二乘法原理图

求解 (1, 2) 为最小, 这就是一个最小二乘问题。

(1) 一元线性拟合

一元线性回归分析是处理两个变量之间关系的最简单模型, 它所研究的对象是两个变量之间的线性相关关系。一元线性拟合是用直线回归方程表示两个数量变量间依存关系的统计分析方法, 属双变量分析的范畴。如果某一个变量随着另一个变量的变化而变化, 并且它们的变化在直角坐标系中呈直线趋势, 就可以用一个直线方程来定量地描述它们之间的数量依存关系, 这就是直线回归分析。直线回归分析中两个变量的地位不同, 其中一个变量是依赖另一个变量而变化的, 因此分别称为因变量 (dependent variable) 和自变量 (independent variable), 习惯上分别用 y 和 x 来表示。

例 1: 为了研究氮含量对铁合金溶液初生奥氏体析出温度的影响, 测定了不同氮含量时铁合金溶液初生奥氏体析出温度, 得到表 2-1 给出的 5 组数据。

表 2-1 氮含量与灰铸铁初生奥氏体析出温度测试数据

序号	氮含量 $x/\%$	初生奥氏体析出温度 $y/^\circ\text{C}$	序号	氮含量 $x/\%$	初生奥氏体析出温度 $y/^\circ\text{C}$
1	0.0043	1220	4	0.0100	1208
2	0.0077	1217	5	0.0110	1205
3	0.0087	1215			

如果把氮含量作为横坐标, 把初生奥氏体析出温度作为纵坐标, 将这些数据标在平面直角坐标上, 则得图 2-2, 这个图称为散点图。

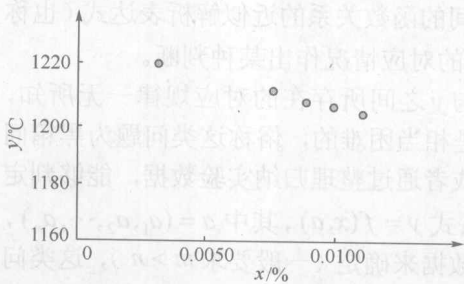


图 2-2 氮含量与灰铸铁初生奥氏体析出温度

从图 2-2 可以看出, 数据点基本落在一条直线附近。这告诉人们, 变量 x 与 y 的关系大致可看作是线性关系, 即它们之间的相互关系可以用线性关系来描述。但是由于并非所有的数据点完全落在一条直线上, 因此 x 与 y 的关系并没有确切到可以唯一地由一个 x 值确定一个 y 值的程度。其他因素, 诸如其他微量元素的含量以及测试误差等都会影响 y 的测试结果。如果要研究 x 与 y 的关系, 可以作线性拟合。

$$\hat{y} = a + bx \tag{2-3}$$

称式(2-3)为回归方程, a 与 b 是待定常数, 称为回归系数。从理论上讲, 有无穷多组解, 回归分析的任务是求出其最佳的线性拟合。

直线回归分析的一般步骤: 将 n 个观察单位的变量对 (x, y) 在直角坐标系中绘制散点图, 若呈直线趋势, 则可拟合直线回归方程, 求回归方程的回归系数和截距。写出回归方程, $\hat{y} = a + bx$, 画出回归直线。对回归方程进行假设检验。

$$Q = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n [y_i - (a + bx_i)]^2 = \min \tag{2-4}$$

$$\frac{\partial Q}{\partial a} = -2 \sum_{i=1}^n (y_i - a - bx_i) = 0 \quad \frac{\partial Q}{\partial b} = -2 \sum_{i=1}^n x_i (y_i - a - bx_i) = 0 \tag{2-5}$$

其中

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i, \quad \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i \tag{2-6}$$

$$l_{xy} = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) \quad (2-7)$$

$$l_{xx} = \sum_{i=1}^n (x_i - \bar{x})^2 \quad (2-8)$$

$$a = \bar{y} - b\bar{x}, \quad b = \frac{l_{xy}}{l_{xx}} \quad (2-9)$$

一元线性拟合精度：相关系数 γ

$$\gamma = \frac{l_{xy}}{\sqrt{l_{xx}l_{yy}}} \quad (2-10)$$

$$l_{yy} = \sum_{i=1}^n (y_i - \bar{y})^2 \quad (2-11)$$

$\gamma = 1$ ，存在线性关系，无实验误差；

$\gamma = 0$ ，毫无线性关系。

剩余平方和

$$Q = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n [y_i - (a + bx_i)]^2 \quad (2-12)$$

回归平方和

$$U = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 = \sum_{i=1}^n [(a + bx_i) - \bar{y}]^2 \quad (2-13)$$

离差平方和

$$S = \sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 + \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 = Q + U \quad (2-14)$$

应用直线回归的注意事项如下。

① 作回归分析要有实际意义，不能把毫无关联的两种现象，随意进行回归分析，忽视事物现象间的内在联系和规律；如对儿童身高与小树的生长数据进行回归分析既无道理也无用途。另外，即使两个变量间存在回归关系时，也不一定是因果关系，必须结合专业知识作出合理解释和结论。

② 直线回归分析的资料，一般要求应变量 y 是来自正态总体的随机变量，自变量 x 可以是正态随机变量，也可以是精确测量和严密控制的值。若稍偏离要求时，一般对回归方程中参数的估计影响不大，但可能影响到标准差的估计，也会影响假设检验时 P 值的真实性。

③ 进行回归分析时，应先绘制散点图 (scatter plot)。若提示有直线趋势存在时，可作直线回归分析；若提示无明显线性趋势，则应根据散点分布类型，选择合适的曲线模型 (curvilinear modal)，经数据变换后，化为线性回归来解决。一般说，不满足线性条件的情形下去计算回归方程会毫无意义，最好采用非线性回归方程的方法进行分析。

④ 绘制散点图后，若出现一些特大特小的离群值 (异常点)，则应及时复核检查，对由于测定、记录或计算机录入的错误数据，应予以修正和剔除。否则，异常点的存在会对回归方程中的系数 a 、 b 的估计产生较大影响。

⑤ 回归直线不要外延。直线回归的适用范围一般以自变量取值范围为限，在此范围内求出的估计值 \hat{y} 称为内插 (interpolation)；超过自变量取值范围所计算的 \hat{y} 称为外延 (extrapolation)。若无充足理由证明，超出自变量取值范围后直线回归关系仍成立时，应该避

免随意外延。

(2) 多元线性回归

讨论了因变量 y 只与一个自变量 x 有关的一元线性回归问题,但在实际中常常会遇到因变量 y 与多个自变量 x_1, x_2, \dots, x_p 有关的情况,这就向人们提出了多元回归分析的问题。直线回归研究的是一个因变量与一个自变量之间的回归问题,但在畜禽、水产科学领域的许多实际问题中,影响因变量的自变量往往不止一个,而是多个,比如绵羊的产毛量这一变量同时受到绵羊体重、胸围、体长等多个变量的影响,因此需要进行一个依变量与多个自变量间的回归分析,即多元回归分析,而其中最为简单、常用并且具有基础性的是多元线性回归分析,许多非线性回归和多项式回归都可以化为多元线性回归来解决,因而多元线性回归分析有着广泛的应用。研究多元线性回归分析的思想、方法和原理与直线回归分析基本相同,但是其中要涉及到一些新的概念以及进行更细致的分析,特别是在计算上要比直线回归分析复杂得多,当自变量较多时,需要应用电子计算机进行计算。

$$\hat{y} = a + b_1x_1 + b_2x_2 + \dots + b_mx_m \quad (2-15)$$

假设随机变量 y 与 p 个自变量 x_1, x_2, \dots, x_p 之间存在着线性相关关系,实际样本量为 n , 其第 i 次观测值为

$$x_{i1}, x_{i2}, x_{i3}, \dots, x_{ip}; y_1, y_2, y_3, \dots, y_i \quad (2-16)$$

则其 n 次观测值可写为如下形式。

$$\begin{cases} y_1 = \beta_0 + \beta_1x_{11} + \beta_2x_{12} + \dots + \beta_px_{1p} + \varepsilon_1 \\ y_2 = \beta_0 + \beta_1x_{21} + \beta_2x_{22} + \dots + \beta_px_{2p} + \varepsilon_2 \\ \vdots \\ y_n = \beta_0 + \beta_1x_{n1} + \beta_2x_{n2} + \dots + \beta_px_{np} + \varepsilon_n \end{cases} \quad (2-17)$$

式中, $\beta_0, \beta_1, \beta_2, \dots, \beta_p$ 是未知参数; x_1, x_2, \dots, x_p 是 p 个可以精确测量并可控制的一般变量; $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$ 是随机误差。和一元线性回归分析一样,假定 ε_i 是相互独立且服从同一正态分布 $N(0, \sigma)$ 的随机变量。

若将式(2-17)用矩阵表示,则有

$$Y = X\beta + \varepsilon \quad (2-18)$$

式中

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \quad X = \begin{pmatrix} 1 & x_{11} & x_{12} & \dots & x_{1p} \\ 1 & x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{np} \end{pmatrix}$$

$$\beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix} \quad \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

多元线性回归分析的首要任务就是通过寻求 β 的估计值 b , 建立多元线性回归方程

$$\hat{y} = b_0 + b_1x_1 + b_2x_2 + \dots + b_px_p \quad (2-19)$$

来描述多元线性模型

$$y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \dots + \beta_px_p \quad (2-20)$$