

中国科学院科学出版基金资助出版

计算机化学化工丛书

# 模式识别方法在化学化工中的应用

陈念貽 钦 佩 著  
陈瑞亮 陆文聪

科学出版社

2000

## 内 容 简 介

本书全面介绍了化学、化工中复杂数据信息处理的模式识别方法,以及与之结合的人工神经网络、遗传算法和非线性回归方法,列举了大批实际应用成果,并提供读者上机练习的机会,做到学以致用。

本书可供化学、化工及相关领域的科研人员和工程技术人员阅读,亦可作为高等学校的教学参考书。

### 图书在版编目(CIP)数据

模式识别方法在化学化工中的应用/陈念贻等著.-北京:科学出版社,2000

(计算机化学化工丛书/许志宏主编)

ISBN 7-03-008393-8

I. 模… II. 陈… III. 模式识别-应用-化学系统工程  
IV. TQ021.8

中国版本图书馆 CIP 数据核字(2000)第 05391 号

科学出版社 出版

北京东黄城根北街 16 号  
邮政编码:100717

新蕾印刷厂印刷

科学出版社发行 各地新华书店经销

\*

2000 年 9 月第 一 版 开本:850×1168 1/32

2000 年 9 月第一次印刷 印张:10 3/4

印数:1—2500 字数:271 000

定价:27.00 元

(如有印装质量问题,我社负责调换〈北燕〉)

《计算机化学化工丛书》  
编 委 会

主 编 许志宏

副主编 杨小震

编 委 (以姓氏笔画为序)

|     |     |     |     |
|-----|-----|-----|-----|
| 马沛生 | 王淀佐 | 王 夔 | 许 禄 |
| 李 科 | 来鲁华 | 陈丙珍 | 陈冀胜 |
| 陈凯先 | 陈念贻 | 陈敏伯 | 陈维明 |
| 杨友麒 | 严新建 | 林少凡 | 郑崇直 |
| 周家驹 | 胡鑫尧 | 俞汝勤 | 郭传杰 |
| 郭 力 | 袁身刚 | 麻德贤 | 惠永正 |
| 潘忠孝 |     |     |     |

## 《计算机化学化工丛书》序

化学化工是信息量特别大的一门学科。到 2000 年 4 月底，在 CAS 登录的化学物质约 2400 万种，它们的各种性质，加上多元体的性质，可以说是一个无边无际的数据海洋。于是，化学数据库的建设就成为 20 世纪后 20 年国际上的一件大事。中国科学院从 1979 年开始建设化学数据库，迄今已经整整 20 年。其间，多次得到国家和中国科学院的奖励。

长期以来，人们不仅希望能定性地掌握而且希望能定量地了解化学化工学科的规律，而计算机的能力及相关技术高速发展，正在帮助人们一步一步地实现着这个愿望。从理论化学的计算、谱图解析、化学计量学、化工过程模拟、热力学的复杂计算，都在计算机的帮助下得到了很大的发展。

上述基础工作为化学化工领域的工作者增加了很大的自由度：可以用计算机帮助对化合物的谱图解析，帮助选择合成路线，帮助进行药物分子设计，可以进行新过程、新技术的开发，可以进行大型工业装置的设计，可以对工厂的生产过程进行优化，……等等。

在改革开放的 20 年里，我们的计算机化学从无到有，形成了一支生机勃勃的研究队伍，也形成了一个很大的用户和读者群体。到 21 世纪，这个群体更需要有能力利用计算机帮助自己的工作，所以本套丛书中也包含一些计算机化学化工的教材，以利于化学化工本科生和研究生的培养和工程人员自学。

所以，我们希望通过本套丛书介绍一些解决问题的方法，帮助读者在遇到问题时，知道如何去解决问题。为此，要求作者在自己的著作中，要给出软件、数据的出处、网络地址或光盘。时代发展得很快，仅做到这一点显然还是不够的。我们特别注意到

近年来, Internet 网络的高速发展已经给我们的时代带来了巨大变化。到 1999 年 7 月, Internet 已经是一个连接 5600 多万个节点的一个网络系统, 它上面的文本信息已经超过 600 亿字节。这些信息一个最大的长处是时间的滞后最小, 易于通过计算机获取。

如果能将科学数据库在网上的功能, 由数据的存取扩大到运算、绘图、模拟等多方面, 必将极大地推动科学数据库工作的发展和广泛使用。在 21 世纪, 人们将逐步可以做到在用户端将数据从库中取出, 在服务器端程序系统上计算, 结果以图形或多媒体方式输出到用户端。据了解, 我们有些作者在自己的工作中已经能够在网上实现数据查询、计算、绘图、三维图形显示等。

进入 21 世纪, Internet 网络系统的应用将更加普及, Internet 网络的客户/服务器的应用将进入千家万户, 进入教室和办公室的各个角落。所以, 如果能将科学数据库和计算程序库在网络服务器上实现, 那么它的普及应用, 将会随着计算机网络的推广而推广。

如果有的丛书作者, 目前仅能给出单机版本的软件, 也欢迎他们能再作一点工作, 很快能达到上网服务的目的。相信进入 21 世纪不久, 在用户端上人们就有可能逐步享受到多种媒体的全方位的科学信息的服务。

这套丛书是我国多位化学化工学科的专家、教授、学者多年辛勤劳动的成果, 也是科学出版社、国家自然科学基金委员会优秀研究成果专著出版基金和中国科学院科学出版基金大力支持的产物, 希望它的出版能促进我国 21 世纪计算机化学化工学科的发展, 并有助于相关学科发展。

《计算机化学化工丛书》编委会

2000 年 5 月

## 前 言

数据信息采掘和信息融合(Data Mining and Data Fusion)是当今信息科学一个新热点。其涵义是综合运用多种算法,对从多种渠道来的大量数据进行计算机处理,通过去粗取精、去伪存真、由此及彼、由表及里的信息加工,抽提有用信息,发现自然规律。化学、化工是实验科学,是积存数据的“大户”,是数据信息采掘和融合技术大有可为的“用武之地”。它不仅可用于化工生产优化、辅助新产品试制,而且在分子设计、材料设计、化合物和物系的性质预报等基础研究中也很有用。此外,环境保护、地质勘探、医疗诊断和生物信息科学研究中,都可能利用它打开新局面。

我和我的同事从事化学化工数据信息采掘研究和应用 20 余年,我们开发的算法和软件技术,广泛应用于宝山钢铁公司、南京炼油厂、郑州铝厂等数十家大、中型企业,取得了很好的效益。在 863 高科技项目中,辅助多种新材料研制取得实效。与此同时,我们还将这些算法用于相图计算和物性预报,开拓了新的研究领域。近年来,我们开发的新算法和软件已开始在美国福特公司以及瑞士、新加坡、巴西等国的企业和科研单位推广应用,并取得效果。我们编写本书的目的就是希望推动我国更多的化学、化工、冶金等领域的科研、技术人员掌握和运用这一新技术,为生产和研究工作做出贡献。

基于上述目的,我们试图将本书编成“学以致用”的书,不但介绍算法原理和应用成果,而且能指导读者掌握使用我们的软件。读者可从因特网免费下载我们软件的演示版,也可以低价从代销点取得我们软件的简易版。本书附录载有若干练习题,读者可利用演示版或简易版针对练习题进行不同程度的上机实习。如果读者需要,也可从代销点购买我们的软件正版及其说明书,将本书与

说明书参照阅读,可收到理论联系实际之效果。本书附录为读者提供免费下载演示版和购买正版、简易版的网址,以便读者阅读本书时,可以边读边上机,提高学习的兴趣。

本书包括的科学研究成果,得到过 863 计划、国家自然科学基金和美国福特公司的资助,本书的出版得到中国科学院出版基金的支持,在此一并致谢。

化学、化工的数据信息采集是一门边缘科学技术,涉及面很广。本书虽然是在我们 20 余年工作基础上编写的,但也难免有“挂一漏万”之处,我们欢迎广大读者在阅读、使用过程中,向我们提出宝贵意见,以便今后改进。

陈念贻

1999 年 9 月于上海

# 目 录

|                                  |       |
|----------------------------------|-------|
| 《计算机化学化工丛书》序 .....               | (iii) |
| 前言 .....                         | (v)   |
| 第一章 化学、化工的复杂数据处理 .....           | (1)   |
| 1.1 化学、化工中的几个共性课题 .....          | (1)   |
| 1.2 复杂数据处理的困难和对策 .....           | (4)   |
| 1.2.1 复杂数据处理的困难 .....            | (4)   |
| 1.2.2 复杂数据处理的对策 .....            | (8)   |
| 1.3 化学、化工复杂数据处理的意义和价值 .....      | (11)  |
| 1.3.1 复杂数据处理是化学科学的一个薄弱环节 .....   | (11)  |
| 1.3.2 复杂数据信息采掘是改进化工生产的捷径 .....   | (12)  |
| 1.3.3 处理复杂数据必须综合应用多种计算方法 .....   | (13)  |
| 1.4 复杂数据信息采掘的信息处理流程 .....        | (20)  |
| 参考文献 .....                       | (22)  |
| 第二章 复杂数据信息采掘的主要方法之一 ——模式识别 ..... | (23)  |
| 2.1 模式识别方法的原理和基本概念 .....         | (23)  |
| 2.2 数据文件的标准化 .....               | (25)  |
| 2.3 主成分分析及其衍生方法 .....            | (25)  |
| 2.3.1 主成分分析的原理及 KL 变换 .....      | (25)  |
| 2.3.2 主成分的特性 .....               | (28)  |
| 2.3.3 主成分的贡献率 .....              | (29)  |
| 2.3.4 主成分算法步骤 .....              | (30)  |
| 2.3.5 限值响应问题 .....               | (31)  |
| 2.4 白化变换——线性投影法(LMAP) .....      | (31)  |
| 2.4.1 LMAP 的原理 .....             | (31)  |

|        |                            |      |
|--------|----------------------------|------|
| 2.4.2  | 算法步骤                       | (33) |
| 2.5    | 最优判别平面方法                   | (35) |
| 2.5.1  | ODP 的原理                    | (35) |
| 2.5.2  | 讨论                         | (39) |
| 2.5.3  | ODP 算法步骤                   | (40) |
| 2.6    | 偏最小二乘法                     | (41) |
| 2.6.1  | 主成分的 NIPALS 算法             | (42) |
| 2.6.2  | PLS 算法步骤和原理                | (46) |
| 2.6.3  | PLS 的若干性质                  | (48) |
| 2.6.4  | PLS 预报步骤                   | (50) |
| 2.6.5  | PLS 成分数目的确定                | (50) |
| 2.7    | 非线性映照                      | (52) |
| 2.7.1  | 线性映照的局限性                   | (52) |
| 2.7.2  | NLM 原理                     | (53) |
| 2.7.3  | PCA-NLM、LMAP-NLM 和 PLS-NLM | (54) |
| 2.7.4  | NLM 的计算步骤                  | (55) |
| 2.8    | 相似分析法                      | (56) |
| 2.8.1  | SIMCA 的基本原理                | (56) |
| 2.8.2  | SIMCA 信息分析                 | (58) |
| 2.8.3  | SIMCA 计算步骤                 | (59) |
| 2.9    | KNN 法及其衍生方法                | (60) |
| 2.10   | 聚类分析方法                     | (62) |
| 2.10.1 | 分级聚类方法                     | (63) |
| 2.10.2 | 最小生成树法                     | (64) |
| 2.10.3 | 最短生成路径法                    | (65) |
| 2.10.4 | 判别聚类的势函数法                  | (66) |
| 2.11   | 模式识别的逆映照方法                 | (67) |
| 2.11.1 | 线性逆映照(LIM)                 | (68) |
| 2.11.2 | 非线性逆映照(NLIM)               | (69) |
|        | 参考文献                       | (69) |

### 第三章 复杂数据信息采掘的主要方法之二——人工神经网络

|                              |       |
|------------------------------|-------|
| 网络 .....                     | (71)  |
| 3.1 人工神经网络模型 .....           | (71)  |
| 3.1.1 人工神经网络的节点 .....        | (71)  |
| 3.1.2 人工神经网络的拓扑结构 .....      | (74)  |
| 3.1.3 人工神经网络的运行 .....        | (76)  |
| 3.1.4 人工神经网络的性质和特点 .....     | (76)  |
| 3.1.5 人工神经网络的学习与训练 .....     | (78)  |
| 3.2 误差逆传播神经网络 .....          | (80)  |
| 3.2.1 误差逆传播(BP)学习算法的提出 ..... | (80)  |
| 3.2.2 BP网络结构与学习规则 .....      | (80)  |
| 3.2.3 BP算法学习规则的数学推导 .....    | (84)  |
| 3.2.4 BP网络的简单评价 .....        | (88)  |
| 3.3 多层前馈网络的 SABP 算法 .....    | (89)  |
| 3.3.1 传统 BP 算法主要缺点及改进 .....  | (89)  |
| 3.3.2 模拟退火算法 .....           | (90)  |
| 3.3.3 三层前馈网络 SABP 算法原理 ..... | (94)  |
| 3.4 自组织映射神经网络 .....          | (97)  |
| 3.4.1 自组织特征映射 .....          | (97)  |
| 3.4.2 算法设计 .....             | (99)  |
| 3.5 人工神经网络结果的二维图象显示 .....    | (100) |
| 参考文献 .....                   | (101) |

### 第四章 复杂数据信息采掘的主要方法之三——遗传算法

|                               |       |
|-------------------------------|-------|
| .....                         | (103) |
| 4.1 演化算法 .....                | (103) |
| 4.1.1 概述 .....                | (103) |
| 4.1.2 自组织、自适应和自学习性(智能性) ..... | (105) |
| 4.1.3 本质并行性 .....             | (105) |
| 4.2 遗传算法概述 .....              | (106) |
| 4.2.1 基本概念 .....              | (106) |

|            |                                          |              |
|------------|------------------------------------------|--------------|
| 4.2.2      | 模式定理 .....                               | (107)        |
| 4.3        | 遗传算法设计 .....                             | (110)        |
| 4.3.1      | 遗传算法的基本结构 .....                          | (110)        |
| 4.3.2      | 设计遗传算法的基本步骤 .....                        | (111)        |
| 4.3.3      | 编码方案 .....                               | (112)        |
| 4.3.4      | 适应度 .....                                | (116)        |
| 4.3.5      | 选择策略 .....                               | (118)        |
| 4.3.6      | 遗传算子的设计 .....                            | (122)        |
|            | 参考文献 .....                               | (126)        |
| <b>第五章</b> | <b>数据文件的建立、评估和数据类型考查 .....</b>           | <b>(128)</b> |
| 5.1        | 数据文件的格式要求 .....                          | (128)        |
| 5.1.1      | 格式和要求 .....                              | (128)        |
| 5.1.2      | 多目标问题 .....                              | (129)        |
| 5.1.3      | 预加工 .....                                | (130)        |
| 5.2        | 数据文件的评估原理 .....                          | (131)        |
| 5.2.1      | 超多面体判据 .....                             | (132)        |
| 5.2.2      | KNN 留一法判据 .....                          | (133)        |
| 5.2.3      | 回归法判据 .....                              | (134)        |
| 5.3        | 数据评估的做法和标准 .....                         | (135)        |
| 5.3.1      | 超多面体判据 .....                             | (135)        |
| 5.3.2      | KNN 留一法的判据 .....                         | (135)        |
| 5.3.3      | 回归法的判据 .....                             | (136)        |
| 5.4        | 数据结构的初步分析 .....                          | (136)        |
| 5.4.1      | 近邻分析(nearest neighbor analysis) .....    | (136)        |
| 5.4.2      | 拓扑分析(topological type analysis) .....    | (136)        |
| 5.4.3      | 近线性分析(near-linearity analysis) .....     | (139)        |
| 5.4.4      | 时间序列分析(time series analysis) .....       | (140)        |
| 5.4.5      | Fisher 指数分析(fisher index analysis) ..... | (141)        |
|            | 参考文献 .....                               | (142)        |
| <b>第六章</b> | <b>数据的相关分析方法 .....</b>                   | <b>(143)</b> |

|            |                         |              |
|------------|-------------------------|--------------|
| 6.1        | 相关分析的价值和局限性 .....       | (143)        |
| 6.2        | 单因子相关分析和 t-f 图 .....    | (145)        |
| 6.3        | 双因子分析和 f-f 图 .....      | (154)        |
| 6.4        | 三因子分析及三维图的显示 .....      | (157)        |
| 6.5        | 数据变换与相关分析相结合的算法 .....   | (158)        |
| 6.6        | f-f 图的分级投影方法 .....      | (160)        |
|            | 参考文献 .....              | (161)        |
| <b>第七章</b> | <b>数据文件的样本筛选 .....</b>  | <b>(162)</b> |
| 7.1        | 数据文件可分性不好的三个原因 .....    | (162)        |
| 7.2        | 子空间局部考查 .....           | (163)        |
| 7.3        | 添加自变量影响的考查 .....        | (166)        |
| 7.4        | 离群点的删除 .....            | (168)        |
|            | 参考文献 .....              | (172)        |
| <b>第八章</b> | <b>数据文件的自变量筛选 .....</b> | <b>(173)</b> |
| 8.1        | 自变量筛选的意义 .....          | (173)        |
| 8.2        | 自变量筛选的多义性 .....         | (175)        |
| 8.3        | 相关分析的应用和局限性 .....       | (176)        |
| 8.4        | 有关变量的共线性检查 .....        | (178)        |
| 8.5        | 近线性数据文件的自变量筛选 .....     | (178)        |
| 8.6        | 偏置型数据集的自变量筛选 .....      | (180)        |
| 8.7        | 包容型数据集的自变量筛选 .....      | (185)        |
| 8.8        | 子空间局部考查与自变量筛选 .....     | (186)        |
| 8.9        | 自变量筛选必须结合专业知识进行 .....   | (187)        |
|            | 参考文献 .....              | (189)        |
| <b>第九章</b> | <b>数据文件的实用建模 .....</b>  | <b>(190)</b> |
| 9.1        | 实用建模的要求和目标 .....        | (190)        |
| 9.2        | 分类判别问题的超多面体模型 .....     | (192)        |
| 9.3        | 最佳投影-自动矩形-分级投影方法 .....  | (196)        |
| 9.4        | 增补测试样本的算法及应用 .....      | (196)        |
| 9.5        | 实用建模中外推的方法 .....        | (200)        |

|             |                                |              |
|-------------|--------------------------------|--------------|
| 9.6         | 实用建模中的回归方法 .....               | (202)        |
| 9.7         | 人工神经网络的实用建模 .....              | (203)        |
| 9.8         | 模式识别与人工神经网络相结合的方法 .....        | (205)        |
| 9.9         | 限值响应问题的实用建模 .....              | (206)        |
| 9.10        | 多目标优化模型的建立 .....               | (210)        |
|             | 参考文献 .....                     | (211)        |
| <b>第十章</b>  | <b>原子参数和分子参数 .....</b>         | <b>(212)</b> |
| 10.1        | 原子和分子参数选择的原理 .....             | (212)        |
| 10.2        | 原子的价电子数( $z$ ) .....           | (212)        |
| 10.3        | 原子的电离势( $I$ ) .....            | (213)        |
| 10.4        | 原子半径和离子半径( $R$ ) .....         | (213)        |
| 10.5        | 电负性 .....                      | (220)        |
| 10.6        | 分子的拓扑参数和原子集团参数 .....           | (224)        |
| 10.7        | 离子键化合物及物系的物性表征参数 .....         | (226)        |
| 10.8        | 金属键化合物及物系的物性表征参数 .....         | (228)        |
| 10.9        | 共价化合物及物系的物性表征参数 .....          | (230)        |
|             | 参考文献 .....                     | (230)        |
| <b>第十一章</b> | <b>数据信息采掘在物性预报中的应用 .....</b>   | <b>(231)</b> |
| 11.1        | 原子-分子参数-数据信息采掘方法 .....         | (231)        |
| 11.2        | 若干热力学性质的计算机预报 .....            | (231)        |
| 11.2.1      | 化合物熔点的计算机预报 .....              | (231)        |
| 11.2.2      | 化合物包晶分解温度的计算机预报 .....          | (233)        |
| 11.2.3      | 液态合金混合熵的计算机预报 .....            | (236)        |
| 11.3        | 若干物理性质的计算机预报 .....             | (237)        |
|             | 参考文献 .....                     | (238)        |
| <b>第十二章</b> | <b>数据信息采掘在相图计算中的应用 .....</b>   | <b>(239)</b> |
| 12.1        | 相图计算的意义和相图计算的原子参数-模式识别方法 ..... | (239)        |
| 12.2        | 原子参数-模式识别方法概述 .....            | (240)        |
| 12.3        | 二元合金相图中间相的形成规律 .....           | (242)        |

|             |                                   |              |
|-------------|-----------------------------------|--------------|
| 12.4        | 原子参数与三元合金相形成的关系——取代对<br>的概念·····  | (243)        |
| 12.5        | 非过渡金属间三元化合物的形成规律·····             | (244)        |
| 12.6        | 过渡金属间三元化合物的形成规律·····              | (246)        |
| 12.7        | 过渡金属和非过渡金属间三元化合物的形成规<br>律·····    | (247)        |
| 12.8        | 氧化物系相图中的中间化合物的形成规律·····           | (248)        |
| 12.9        | 二元液态合金的相互作用系数和液相分层的计<br>算机预报····· | (249)        |
| 12.10       | 相图中间相熔化类型的判别·····                 | (250)        |
| 12.11       | 三元相图液相面的计算机预报·····                | (250)        |
|             | 参考文献·····                         | (253)        |
| <b>第十三章</b> | <b>数据信息采掘在新材料、新产品研制中的应用·····</b>  | <b>(254)</b> |
| 13.1        | 材料设计和分子设计的意义·····                 | (254)        |
| 13.2        | 材料设计专家系统·····                     | (255)        |
| 13.3        | 材料设计专家系统用于已有数据的加工·····            | (256)        |
| 13.4        | 材料设计专家系统辅助实验探索·····               | (258)        |
| 13.5        | 材料设计辅助材料智能加工·····                 | (263)        |
|             | 参考文献·····                         | (264)        |
| <b>第十四章</b> | <b>数据信息采掘在化工生产优化中的应用·····</b>     | <b>(266)</b> |
| 14.1        | 化工过程复杂反应体系的量纲分析·····              | (266)        |
| 14.2        | 数据信息采掘和优化建模在炼油工业中的应用<br>·····     | (268)        |
| 14.3        | 数据信息采掘在高分子材料生产优化中的应用<br>·····     | (273)        |
| 14.4        | 数据信息采掘在染料色光控制中的应用·····            | (277)        |
| 14.5        | 数据信息采掘在醋酸乙烯催化合成中的应用·····          | (278)        |
| 14.6        | 数据信息采掘用于提高有机合成产率·····             | (279)        |
| 14.7        | 数据信息采掘在化工环保中的应用·····              | (281)        |
| 14.8        | 数据信息采掘在产品检验自动化中的应用·····           | (283)        |

|             |                               |              |
|-------------|-------------------------------|--------------|
| 14.9        | 数据信息采掘在化工设备防腐方面的应用·····       | (284)        |
| 14.10       | 数据信息采掘在其他化工过程中的应用·····        | (284)        |
|             | 参考文献·····                     | (285)        |
| <b>第十五章</b> | <b>数据信息采掘在冶金生产优化中的应用·····</b> | <b>(286)</b> |
| 15.1        | 炼焦配煤的优化·····                  | (286)        |
| 15.2        | 模式识别在降低焦比中的应用·····            | (287)        |
| 15.3        | 模式识别方法分析炼钢转炉炉龄的影响因素···        | (288)        |
| 15.4        | 连铸坯表面质量的模式识别分析·····           | (288)        |
| 15.5        | 模式识别在合金钢生产中的应用·····           | (289)        |
| 15.6        | 联合法生产氧化铝净溶出率的模式识别优化···        | (291)        |
| 15.7        | 烧结法碳酸化分解终点的优化控制·····          | (292)        |
| 15.8        | 电解铝电流效率的优化模型·····             | (292)        |
| 15.9        | 电解铝阳极导电合金成分优化·····            | (293)        |
| 15.10       | 钢铁表面氮化过程的质量优化·····            | (294)        |
| 15.11       | 汽车零件光亮镀铬的质量优化·····            | (295)        |
| 15.12       | 热法炼镁质量与配料比的关系·····            | (296)        |
|             | 参考文献·····                     | (296)        |
| <b>附录</b>   | ·····                         | <b>(297)</b> |
| A           | 为初学者按“向导方式”上机实习的操作指南·····     | (297)        |
| B           | 上机实习·····                     | (302)        |
| C           | 附表·····                       | (303)        |

# 第一章 化学、化工的复杂数据处理

## 1.1 化学、化工中的几个共性课题

迄今为止,化学、化工都是以实验工作为主的学科。这是由其研究对象的特点决定的。化学、化工研究的对象多为由众多原子组成的复杂体系,原子间的作用复杂多样,难于用简单的解析方程求解。众多原子的统计力学方程也极为繁杂,宏观物系中通常发生传热、传质、流体流动过程,有时还有相变、界面现象和电流、电磁场的作用。要从“第一原理”推算和把握如此复杂的体系和过程,在可以预见的未来尚难办到。与此同时,这类体系在尺寸上与人类活动范围相近,其变化过程所需能量在地球表面环境下易于供应,这就使得这类靠解析方程求解极为困难的体系,做实验取数据较为容易。于是化学、化工领域的数据积累在各门学科中居于前列,如何从其中总结出经验或半经验规律,据以指导实验或提供线索,以便能用较小的工作量、较少的盲目性解决各种实际问题,就成为普遍的要求。

化学家的一项重要任务是发明各种有用的新物质(例如构建计算机用的高性能半导体,航天工业用的耐高温高强度合金,以至高效的抗癌药,色泽鲜艳不易褪色的染料等)。这里一个带共性的课题是:用何种原子(或元素)可堆成何种结构,具有何种物理或化学性质或生物活性,即结构-性质(性能)关系问题。

虽然上述问题原则上可通过量子力学、统计力学的方程(Schrodinger 方程, Liouville 方程)求解,但许多实际问题体系过于复杂,单靠“第一原理”求解不可能,但可利用“第一原理”或基础实验找出与目标值(性质或性能)有关的参数(通常称为原子参数、分子参数或化学键参数),据此处理已知实验结果(如各种物理、化学

性质的数据),总结经验或半经验规律<sup>[1-3]</sup>。

化学家经常遇到的一个共性问题就是要寻求各种物质的原子参数和该物质性质的关系,这些数学关系一般是非线性、高噪声、多因子的复杂关系。这方面的课题大体上可表示如图 1.1。

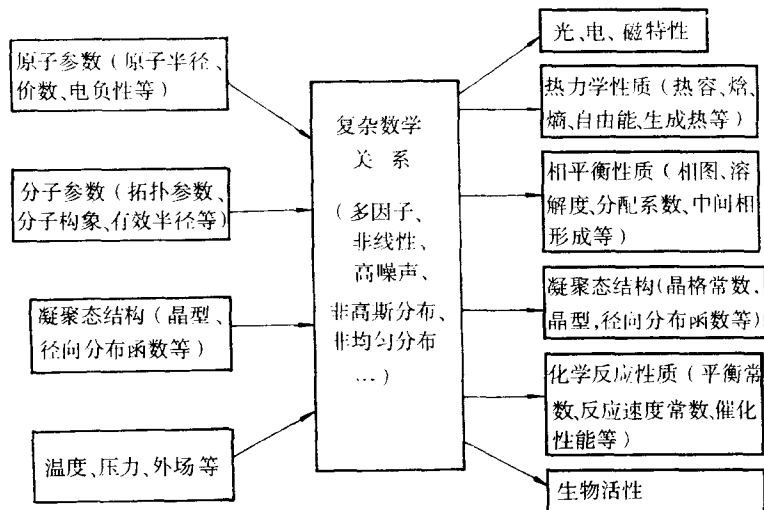


图 1.1 物质结构和性质的关系

化学、化工专家的另一重要任务是发明或改进生产方法,将有用的物质大量、高质量、低能耗、低成本地生产出来。这通常牵涉到研究并掌握一系列传热、传质、流体流动和化学反应网络的复杂过程。尽管已知传热方程、扩散方程,流体力学 Navier-Stokes 方程和描述反应速度的化学动力学方程可联立求解,但多数实用的体系过于复杂,求解十分困难。可举石油化工中加氢裂化(hydrocracking)为例:原料油中有成千种化合物,且其间比例常常改变,多孔的固体催化剂细孔分布亦不均匀,加氢反应包括一个复杂的反应网络。各局部的传热、传质条件各不相同,生产结果又随操作参数(温度、压力、循环量等)和原料纯度(硫含量等)而变,十分明显,欲解决生产实际问题(如产品航空煤油的质量指标或液体产品