



“十二五”国家重点图书 现代化学与应用丛书
普通高等教育化学类专业规划教材

基础化学计量学 及其应用

倪力军 张立国 ◎编 著



华东理工大学出版社
EAST CHINA UNIVERSITY OF SCIENCE AND TECHNOLOGY PRESS

基础化学计量学 及其应用

张立国 ◎编 著

图书在版编目(CIP)数据

基础化学计量学及其应用/倪力军,张立国编著.
—上海:华东理工大学出版社,2011.7
普通高等教育化学类专业规划教材
ISBN 978-7-5628-3011-5

I. ①基… II. ①倪… ②张… III. ①化学计量学—
高等学校—教材 IV. ①06-04

中国版本图书馆 CIP 数据核字(2011)第 086437 号

“十二五”国家重点图书·现代化学与应用丛书·

普通高等教育化学类专业规划教材

基础化学计量学及其应用

编 著 / 倪力军 张立国

策划编辑 / 刘 强

责任编辑 / 刘 强

责任校对 / 金慧娟

封面设计 / 陆丽君 裘幼华

出版发行 / 华东理工大学出版社

社 址:上海市梅陇路 130 号,200237

电 话:(021)64250306(营销部) 64251837(编辑室)

传 真:(021)64252707

网 址:press.ecust.edu.cn

印 刷 / 常熟华顺印刷有限公司

开 本 / 890 mm×1240 mm 1/32

印 张 / 13

字 数 / 369

版 次 / 2011 年 7 月第 1 版

印 次 / 2011 年 7 月第 1 次

书 号 / ISBN 978 7-5628-3011-5/O·237

定 价 / 48.00 元

(本书如有印装质量问题,请到出版社营销部调换。)

“十二五”国家重点图书·现代化学与应用丛书· 编委会成员名单

顾问委员

郑兰荪(中国科学院院士,厦门大学)

张玉奎(中国科学院院士,中国科学院大连化学物理研究所)

刘若庄(中国科学院院士,北京师范大学)

魏可镁(中国工程院院士,福州大学)

陈小明(中国科学院院士,中山大学)

杨锦宗(中国工程院院士,大连理工大学)

刘新厚(中国科学院理化技术研究所)

梁逸曾(中南大学)

主任委员

陈光巨(北京师范大学)

陈六平(中山大学)

冯亚青(天津大学)

龚跃法(华中科技大学)

胡常伟(四川大学)

黄元河(北京师范大学)

李浩然(浙江大学)

林贤福(浙江大学)

聂进(华中科技大学)

荣国斌(华东理工大学)

王海水(华南理工大学)

谢如刚(四川大学)

袁履冰(大连理工大学)

陈建中(福州大学)

杜一平(华东理工大学)

高士祥(南京大学)

何仁(大连理工大学)

黄可龙(中南大学)

蓝闽波(华东理工大学)

李祥高(天津大学)

穆劲(华东理工大学)

宁桂玲(大连理工大学)

童叶翔(中山大学)

王利民(华东理工大学)

杨光富(华中师范大学)

张维冰(中国科学院大连化学物理研究所)

张文清(华东理工大学)

委 员

安永林(大连理工大学)

陈国荣(华东理工大学)

陈若愚(江苏工业学院)

方德彩(北京师范大学)

高绍康(福州大学)

杭义萍(华南理工大学)

康诗钊(华东理工大学)

李桂玲(华中科技大学)

李 奇(北京师范大学)

李向清(华东理工大学)

刘海燕(华东理工大学)

刘淑芹(大连理工大学)

鲁礼林(武汉科技大学)

罗曦芸(上海博物馆)

马思渝(北京师范大学)

潘铁英(华东理工大学)

邵超英(东华大学)

宋慧宇(华南理工大学)

唐 乾(华中科技大学)

陶晓春(华东理工大学)

王芳辉(北京化工大学)

王立世(华南理工大学)

王 氢(华东理工大学)

王世荣(天津大学)

陈东红(华中科技大学)

陈立功(天津大学)

戴启广(华东理工大学)

冯文芳(华中科技大学)

郭璇华(华南理工大学)

胡 坪(华东理工大学)

李方实(南京工业大学)

李明慧(大连工业大学)

李硕凡(华南理工大学)

林树坤(福州大学)

刘建宇(华南理工大学)

卢 怡(华东理工大学)

罗 钊(华中科技大学)

吕玄文(华南理工大学)

倪力军(华东理工大学)

钱 枫(华东理工大学)

舒谋海(上海交通大学)

唐明生(郑州大学)

唐燕辉(华东理工大学)

童晓峰(华东理工大学)

王 磊(华东理工大学)

王 敏(华南理工大学)

王全瑞(复旦大学)

王文锦(华南理工大学)

王幸宜(华东理工大学)
王 燕(华东理工大学)
伍新燕(华东理工大学)
许 琳(华南理工大学)
杨铁金(齐齐哈尔大学)
于建国(北京师范大学)
袁红玲(华中科技大学)
张春梅(华东理工大学)
张 敏(华东理工大学)
张绍文(北京理工大学)
张玉兰(华东理工大学)
张兆国(上海交通大学)
赵 平(华东理工大学)
仲剑初(大连理工大学)
周志彬(华中科技大学)
朱龙观(浙江大学)

王亚光(华东理工大学)
王朝霞(华东理工大学)
徐志珍(华东理工大学)
许艳杰(天津大学)
杨 毅(大连工业大学)
俞开潮(华中科技大学)
袁天佑(广西大学)
张大德(华东理工大学)
张立国(华东理工大学)
张小平(北京师范大学)
张玉良(华东理工大学)
张正波(华中科技大学)
郑炎松(华中科技大学)
周丽绘(华东理工大学)
朱 红(北京交通大学)
邹 刚(华东理工大学)

“十二五”国家重点图书《现代化学与应用丛书》联系邮箱 hxyyyhx@163.com

前言

化学计量学是连接化学、数学、计算机技术的交叉学科,在高校化学类专业中作为必修课或选修课。化学计量学中的很多方法来自多元统计学,采用矩阵、向量来描述相关化学原理和定律。笔者在从事化学计量学本科生及研究生教学的过程中深切体会到,虽然大多数理工科大学生在大一、大二阶段学过线性代数或概率统计课程,但并不了解相关数学概念、定理与化学信息、化学定律有什么关系,缺乏将化学信息转化为数学方法描述的过渡和训练,通常感到这门课的内容抽象、难懂,难以适应与理解化学计量学中的相关内容和方法。其次,化学计量学中的方法大多需要通过计算机来实现,对于只学过基本计算机语言但缺乏相应锻炼的化学类专业的学生而言,化学计量学算法的计算机实现也是一桩难事,需要采用合适的计算机平台开展本科化学计量学教学与练习。

本书是笔者十年来在华东理工大学化学类专业开设的化学计量学、化学信息的解析及其应用讲义和从事化学计量学应用和理论研究的基础上修订、补充完成的,目的是想提供一本适合化学类专业本科生、便于教学和学生自学并练习的教材。

本书具有以下几个特色。

一是在编写过程中注意各知识点和方法间的联系、不同章节间有呼应。有时同一个例子在不同的章节中都会用到,以便学生通过例题关注和理解不同知识点、方法和概念间的联系及相互关系,而不是孤立地去学习化学计量学的内容。

二是在内容和例题的选编上贯穿“授人以鱼不如授人以渔”的理念,采用恰当的例题为载体,将化学中的问题转化成抽象的数学问题

来描述,再用合适的数学方法来解决该数学问题,在这个过程中将相关概念和方法讲透,并联系实际应用。当学生掌握和理解采用数学工具来描述和解决化学问题的思想和方法后,便不难理解和运用教材中的各种方法。因此,本书的第2章用了很大篇幅进行数学基础知识的回顾,讲解线性代数和数理统计中的基本概念、理论与化学现象、信息的关系,通过大量生动的例子演示数学概念和知识如何用于化学数据、信息及规律的描述。使学生能够适应线性代数的描述方法、理解数理统计中的基本概念并掌握其应用。在打好这些基础后,后面的主成分分析、多元校正、回归分析及模式识别等化学计量学基本内容和方法的讲解与学习就会比较轻松易懂。

三是采用 MATLAB 平台实现化学计量学方法,将学生从繁琐的编程中解放出来。本书绝大部分例题都附有 MATLAB 实现的命令和注释,书中的例题和练习题中的数据均做成电子文档备份在配套光盘中,以便于学生根据教材进行练习。本书中用到的常用 MATLAB 函数、自编的 M 文件和 MATLAB 函数均收录在附录或相关章节中,并附有注释、调用说明及例子,可减轻学生在学习和应用化学计量学方法时计算机编程的困难,非常便于初学者使用。

四是在进行基础知识的讲解和介绍中强调化学计量学方法的应用及注意事项,以便于读者正确使用相关方法。书中例题尽可能采用生产、科研实践或大学生课外创新实践的数据,有明确的化学或应用背景,可使读者感受到化学计量学不仅仅是抽象的数学方法、理论,而且是有广阔应用背景的学科。

五是本书配有 32 学时的教学课件,可供高等院校教师教学时参考。

在本书付梓之际,感谢参考文献中所列的同行和前辈,是他们的工作将笔者引进化学计量学学科大门;感谢中南大学梁逸曾教授 2002 年 7 月在首都师范大学举办的第二届全国化学计量学培训与交流会议期间的讲座,使笔者体会到了如何通过好的例子架起化学与数学间的桥梁。本课题组所指导的部分研究生及参加创新实践的本科生为本书提供了部分算例,进行了校稿、例题与程序验证等工作,在此一并致谢。

由于笔者水平有限,书中的不足之处在所难免,敬请读者和同行批评指正。

目 录

第 1 章 绪论

- 1.1 为什么学习化学计量学 / 1
- 1.2 化学计量学的定义 / 3
- 1.3 化学计量学解决问题的方法 / 4
- 1.4 化学计量学的研究对象与内容 / 5
- 1.5 化学计量学与现代分析化学的关系 / 8
- 1.6 化学计量学教学及其目的 / 9
- 思考题 / 11
- 参考文献 / 11

第 2 章 数学基础知识与 MATLAB 简介

- 2.1 线性代数基础知识 / 12
 - 2.1.1 向量和矩阵 / 12
 - 2.1.2 矩阵的加法和数乘 / 13
 - 2.1.3 矩阵乘法 / 13
 - 2.1.4 矩阵的转置和对称性 / 14
 - 2.1.5 某些特殊矩阵 / 14
 - 2.1.6 矩阵的逆 / 15
 - 2.1.7 矩阵的秩 / 15
 - 2.1.8 矩阵的特征值和特征向量 / 16
 - 2.1.9 线性相关与线性无关 / 16
- 2.2 矩阵概念与化学信息 / 17

2.3 误差与数理统计基本知识 / 24

2.3.1 误差的定义与类型 / 24

2.3.2 精密度和准确度 / 25

2.3.3 概率与数理统计的基本概念 / 27

2.3.4 统计描述 / 28

2.3.5 统计分析 / 33

2.3.6 数理统计在数据分析中的应用 / 35

2.4 MATLAB 平台简介 / 42

2.4.1 MATLAB 系统 / 42

2.4.2 MATLAB 工作环境 / 44

2.4.3 MATLAB 的矩阵运算 / 44

2.4.4 MATLAB 的常用函数 / 46

2.4.5 数据图形化的常用指令与图形的简单修饰 / 52

2.4.6 指令集的函数化(M 文件) / 54

2.4.7 数据特征的度量及其 MATLAB 函数 / 55

2.5 化学数据的预处理方法 / 56

2.5.1 数据预处理的目的 / 56

2.5.2 离散数据的常用预处理方法 / 56

2.5.3 光谱数据的常用预处理方法 / 64

思考题 / 70**练习题 / 71****参考文献 / 71**

第 3 章 相关分析及其应用

3.1 简单线性相关分析 / 74**3.2 相关关系与函数关系 / 81****3.3 复相关分析 / 82****3.4 典型相关分析 / 82**

3.4.1 典型相关分析原理及应用简介 / 83

3.4.2 典型相关分析的数学实现 / 84

3.4.3 典型相关分析中各变量的特点及其相互关系 / 86

3.4.4 典型相关分析中的统计检验 / 86

3.4.5 典型相关分析的应用 / 87

练习题 / 97

参考文献 / 99

第 4 章 回归分析及其应用

4.1 回归分析的概念与原理 / 101

4.2 一元线性回归分析 / 103

4.2.1 一元线性回归方程的建立 / 104

4.2.2 一元线性回归模型的检验 / 105

4.2.3 一元线性回归分析的应用与 MATLAB 实现 / 108

4.3 多元线性回归分析 / 111

4.3.1 多元线性回归方程的建立 / 111

4.3.2 多元线性回归模型的检验 / 113

4.3.3 多元线性回归模型的模型参数检验及 MATLAB
实现 / 117

4.4 可化成多元线性回归的问题 / 119

4.4.1 多项式函数的线性回归 / 119

4.4.2 多元幂函数的线性回归 / 120

4.4.3 指数函数的线性回归 / 120

4.4.4 多元对数函数的线性回归 / 120

4.5 二次回归模型及最优回归模型 / 120

4.6 回归分析中必须注意的问题 / 133

4.6.1 回归分析中自变量与因变量的确定 / 133

4.6.2 自变量的筛选 / 137

4.6.3 多元线性回归的前提 / 138

4.7 相关分析与回归分析的区别和联系 / 139

思考题 / 141

练习题 / 142

参考文献 / 143

第5章 实验设计与实验数据的处理

5.1 基本概念介绍 / 144

- 5.1.1 实验设计的定义 / 144
- 5.1.2 实验的因素和水平 / 147
- 5.1.3 因素间的交互作用 / 147
- 5.1.4 全面实验和多次单因素实验 / 149
- 5.1.5 实验设计方法的分类 / 152

5.2 正交实验设计 / 152

- 5.2.1 关于正交的直观解释 / 152
- 5.2.2 正交表的特点 / 153
- 5.2.3 正交实验设计的主要步骤 / 154
- 5.2.4 正交实验设计结果的方差分析 / 159
- 5.2.5 正交实验数据的回归分析 / 165

5.3 均匀实验设计 / 167

- 5.3.1 均匀设计的原理和特点 / 168
- 5.3.2 均匀实验设计的主要步骤 / 170
- 5.3.3 均匀实验设计的应用 / 170
- 5.3.4 均匀设计表的设计* / 172

5.4 序贯实验设计方法 / 175

- 5.4.1 单纯形 / 175
- 5.4.2 单纯形寻优过程图解 / 176
- 5.4.3 m 个因素的单纯形优化实验设计步骤 / 176

5.5 本章小结 / 184

思考题 / 186

练习题 / 186

参考文献 / 189

第 6 章 主成分分析及其在回归分析中的应用

- 6.1 主成分分析 / 190
 - 6.1.1 主成分分析的直观描述 / 193
 - 6.1.2 主成分分析中的一些基本概念 / 195
 - 6.1.3 主成分分析的求解步骤 / 198
 - 6.1.4 主成分的求解原理* / 200
 - 6.1.5 原始变量用主成分得分的近似表示 / 203
 - 6.1.6 主成分分析的特点及其几何解释 / 203
 - 6.1.7 主成分个数确定的原则 / 205
 - 6.1.8 主成分分析的 NIPALS 算法 / 206
- 6.2 主成分分析的实例及其在 MATLAB 平台的实现 / 209
- 6.3 主成分回归 / 221
 - 6.3.1 问题的提出 / 221
 - 6.3.2 主成分回归(PCR)及其在 MATLAB 平台的实现 / 222
- 6.4 偏最小二乘回归(PLSR) / 231
 - 6.4.1 PLSR 的原理 / 231
 - 6.4.2 PLSR 的算法步骤 / 232
 - 6.4.3 PLSR 的若干性质* / 236
 - 6.4.4 PLSR 的预测步骤 / 236
 - 6.4.5 PLSR 中潜变量个数的确定 / 237
 - 6.4.6 PLSR 的 MATLAB 实施与应用 / 238
- 6.5 本章小结 / 246
- 思考题 / 249
- 练习题 / 250
- 参考文献 / 253

第 7 章 化学校正理论及光谱定量分析技术

- 7.1 基本概念 / 254

- 7.2 化学量测数据的矩阵表示 / 255
- 7.3 多元化学校正方法 / 255
 - 7.3.1 分析化学中的 MLR 方法 / 255
 - 7.3.2 分析化学中的 K 矩阵法 / 259
 - 7.3.3 分析化学中的 P 矩阵法 / 260
 - 7.3.4 基于主成分分析的 PCR 与 PLSR 多元校正方法 / 260
- 7.4 多元校正方法在近红外定量分析技术中的应用 / 260
 - 7.4.1 近红外多元校正方法的原理和应用 / 261
 - 7.4.2 影响近红外定量分析模型的因素 / 266
- 7.5 光谱多元校正定量分析技术 / 271
- 练习题 / 273
- 参考文献 / 274

第 8 章 化学模式识别及其应用

- 8.1 基本概念 / 278
 - 8.1.1 化学模式识别 / 278
 - 8.1.2 特征(Features) / 282
 - 8.1.3 模式(Patterns)与模式空间 / 283
 - 8.1.4 模式识别方法的分类 / 283
 - 8.1.5 分类方法的检验和评估 / 285
 - 8.1.6 模式识别研究中的几个重要问题 / 285
- 8.2 模式识别的数据准备 / 286
 - 8.2.1 数据预处理 / 286
 - 8.2.2 特征抽提 / 286
- 8.3 模式间相似程度的度量方式 / 289
 - 8.3.1 距离系数 / 289
 - 8.3.2 相似系数 / 293
- 8.4 判别分析(Discrimination Analysis) / 300
 - 8.4.1 KNN 法及其衍生方法 / 300
 - 8.4.2 二元线性分类器及其训练 / 307

- 8.4.3 Fisher 判别法 / 312
- 8.4.4 偏最小二乘判别分析(PLS-DA)方法 / 315
- 8.5 无监督模式识别法 / 320**
 - 8.5.1 一次计算形成法 / 321
 - 8.5.2 最大生成树法(Largest Spanning Tree) / 328
 - 8.5.3 最小生成树法(Minimal Spanning Tree) / 330
 - 8.5.4 特征投影的模式识别方法 / 333
- 8.6 模式识别方法在近红外定性分析技术中的应用 / 337**
 - 8.6.1 简化与改进的 KNN 模式识别方法(IS-KNN) / 338
 - 8.6.2 影响近红外定性分析模型的因素 / 343
- 8.7 本章小结 / 344**
 - 8.7.1 模式识别的基本步骤 / 344
 - 8.7.2 影响模式识别结果的因素 / 345
- 思考题 / 346**
- 练习题 / 347**
- 参考文献 / 350**

第 9 章 如何用好化学计量学

- 9.1 避免采用不稳定的数据建模 / 351
- 9.2 避免采用离群点和异常点建模 / 351
- 9.3 谨慎进行数据的转换和确认 / 352
- 9.4 对原始数据进行必要的分析,了解数据的基本特点 / 352
- 9.5 充分利用专业背景知识、慎重评价信息解析得到的结果 / 353
- 9.6 建模与验证相结合,不可片面追求拟合效果 / 353
- 9.7 根据具体问题慎重选择适用的方法 / 354

附录 1 关于马氏距离的定理与证明 / 356

附录 2 本书常用 MATLAB 函数释义 / 360

附录 3 本书自编 MATLAB 函数的源代码及注释 / 376

第1章 绪 论

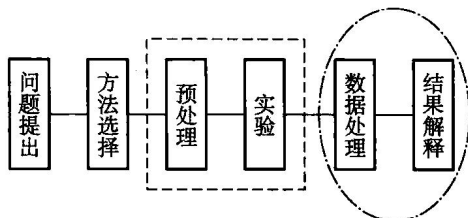
1.1 为什么学习化学计量学

现代科学技术、计算机和互联网、多媒体的飞速发展与普及,给人类生活、工作和科学研究带来极大便利的同时也使我们面临着信息爆炸和信息泛滥的困惑。在各行各业的生产实践与各种实验性学科研究中,经常会产生大量类型不同的实验或测试数据。例如,一个现代化化工厂就有成千上万个传感器不断记录生产数据;医疗、环保、地质勘探领域每天也在积累大量数据。随着计算机技术的进步与互联网的普及,大批数据库(Data base)已经建立。这些数据中蕴藏着大量宝贵信息,为我们认识事物的性质、规律提供了丰富的材料与科学依据,犹如有价“矿藏”有待开发挖掘。它能够快速有效地从铺天盖地的数据与信息中抽提有用的、重要的信息与规律,以指导生产实践与经济、科研活动,成为现代科技人才应当具备的重要技能和素质。

化学是一门理论与实验并重的学科,化学领域工作者每天会面对大量的实验数据。实验数据的种类很多,按基本物性分为物理和化学两大类。如物质的硬度、尺寸、密度、质量、沸点、压强等属于物理性能数据;化学数据主要反映关于化学性能方面的性质,如样品中所含物质的结构、组成(含量)、相对分子质量、化学反应速率常数等。此外,既有以离散形式体现的化学信息,如活化能、化合物(物质)含量等;也有以连续的谱图体现的化学信息,如红外、紫外光谱等。在样品的质量分析、检测过程中大量使用各种光谱、色谱甚至联用仪器方法来获取样品的化学信息,这些化学信息通常以兆为单位,以向量或矩阵、张量等形式存储在计算机中。采用简单的分析方法很难从

中找到有效解决方案和所需的信息,化学计量学则可以为这类复杂数据的解析和信息抽提提供解决方法。

一个化学问题的研究过程一般如下。



以往的化学工作者对研究过程中的预处理及实验过程研究较多(即虚线方框内的内容),其主要精力放在如何获得实验数据上。而对方法选择、数据处理及结果解释(椭圆框内的内容)方面的研究较少,往往不能很好地从实验数据中提取有用的化学信息。传统的化学教育侧重于对化学原理、知识、方法的传授对学生实验技能的培养、实验数据的获取上,在数据的处理、结果分析和解释方面的相关教育和培训较少,学生往往不能很好地从实验及测试数据中提取有用的信息。

在实际化学问题的解决过程中,实验只是其中的一个步骤,与之同样重要的还有实验方案的设计与选择、数据处理和实验结果中化学信息的提取等。实验结果的不合理解释会歪曲隐含在实验数据中的化学信息,导致错误的结论。

此外,目前各类现代分析仪器都配有专门的软件处理和分析测试数据。如果对软件中所用的数据处理方法缺乏必要的了解,不但难以理解、利用和解释软件提供的分析结果而且不易掌握软件中各个功能、参数的使用与设置。此外,商业软件有时不能满足分析者的特定需求或并不能提供用户所需要和关心的某些特定信息(或某些中间过程的信息),需要分析人员对相关数据另外进行分析处理。所以,如何从大量化学实验及测试数据中获取有用的信息,为生产实践、科研和质量控制提供服务是化学工作者及相关领域技术人员所面临的共同难题。

本书主要介绍以数字形式为载体或可化为数字形式的化学实验