



“十二五”
国家重点图书

Handbook of Analytical Chemistry

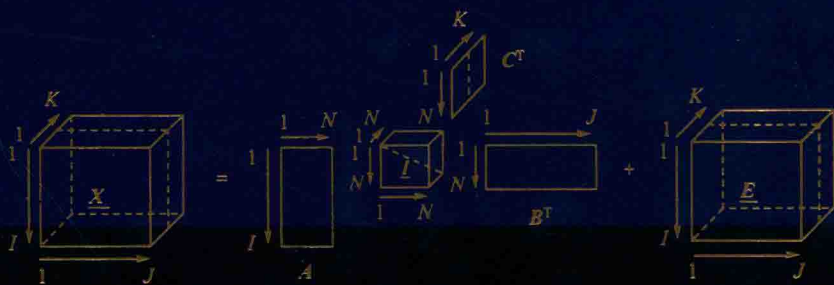
分析化学手册

第三版

10

化学计量学

梁逸曾 吴海龙 俞汝勤 主编



化学工业出版社

分析化学手册

第三版

10

化学计量学

梁逸曾 吴海龙 俞汝勤 主编



化学工业出版社

· 北京 ·

《分析化学手册》第三版在第二版的基础上作了较大幅度的增补和删减，保持原手册 10 个分册的基础上，将其中 3 个分册进行拆分，扩充为 6 册，最终形成 13 册。

本书的编写依照分析化学研究的全过程，从选择分析方法和采样开始，经化学量测的试验设计、信号预处理、定性定量分析的多元校正和多元分辨、再到有用决策信息的提取，包括化学模式识别、机理研究的数字模拟方法、化学构效关系研究直至人工智能与化学专家系统，试图覆盖分析化学计量学的全部内容。在讨论了化学计量学的基本概念和基本方法的数学定义、基本思路及算法的基础上，尽量同时给出相应的计算机程序和参考文献，方便读者通过本书所提供的定义或方法进行试验，并可进一步将其运用于实际问题的解决。

本书为从事分析化学研究的技术人员了解和使用化学计量学方法提供了重要的工具与参考资料。

图书在版编目 (CIP) 数据

分析化学手册. 10. 化学计量学/梁逸曾, 吴海龙, 俞汝勤主编. —3 版. —北京: 化学工业出版社, 2016. 11
ISBN 978-7-122-28118-0

I. ①分… II. ①梁…②吴…③俞… III. ①分析化学-手册②化学计量学-手册 IV. ①O65-62

中国版本图书馆 CIP 数据核字 (2016) 第 227580 号

责任编辑: 傅聪智 李晓红 任惠敏
责任校对: 王素芹

文字编辑: 孙凤英
装帧设计: 王晓宇

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)

印刷: 大厂聚鑫印刷有限责任公司

装订: 三河市胜利装订厂

787mm×1092mm 1/16 印张 54½ 字数 1408 千字 2016 年 11 月北京第 3 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: 190.00 元

版权所有 违者必究

《分析化学手册》(第三版)编委会

主任:汪尔康

副主任:江桂斌 陈洪渊 张玉奎

委员(按姓氏汉语拼音排序):

- | | |
|-----|------------------------------|
| 柴之芳 | 中国科学院院士
中国科学院高能物理研究所 |
| 陈洪渊 | 中国科学院院士
南京大学 |
| 陈焕文 | 东华理工大学 |
| 陈义 | 中国科学院化学研究所 |
| 丛浦珠 | 中国医学科学院药用植物研究所 |
| 邓勃 | 清华大学 |
| 董绍俊 | 发展中国家科学院院士
中国科学院长春应用化学研究所 |
| 郭伟强 | 浙江大学 |
| 江桂斌 | 中国科学院院士
中国科学院生态环境研究中心 |
| 江云宝 | 厦门大学 |
| 柯以侃 | 北京化工大学 |
| 梁逸曾 | 中南大学 |
| 刘振海 | 中国科学院长春应用化学研究所 |
| 庞代文 | 武汉大学 |
| 邵元华 | 北京大学 |
| 苏彬 | 浙江大学 |
| 汪尔康 | 中国科学院院士
中国科学院长春应用化学研究所 |
| 王敏 | 浙江大学 |

吴海龙	湖南大学
许国旺	中国科学院大连化学物理研究所
严秀平	南开大学
杨峻山	中国医学科学院药用植物研究所
杨芄原	复旦大学
杨秀荣	中国科学院院士 中国科学院长春应用化学研究所
姚守拙	中国科学院院士 湖南大学, 湖南师范大学
于德泉	中国工程院院士 中国医学科学院药物研究所
俞汝勤	中国科学院院士 湖南大学
张新荣	清华大学
张玉奎	中国科学院院士 中国科学院大连化学物理研究所
赵墨田	中国计量科学研究院
郑国经	北京首钢冶金研究院 (现北冶功能材料有限公司)
郑健	中华人民共和国科学技术部
朱俊杰	南京大学
庄乾坤	国家自然科学基金委员会化学科学部

序

分析化学是人们获得物质组成、结构及相关信息的科学，即测量与表征的科学。其主要任务是鉴定物质的化学组成及含量测定、确定物质的结构形态及其与物质性质之间的关系。分析化学是一门社会和科技发展迫切需要的、多学科交叉结合的综合性科学。现代分析化学必须回答当代科学技术和社会需求对现存的方法和技术的挑战，因此实际上已发展成为“分析科学”。

《分析化学手册》是一套全面反映现代分析技术，供化学工作者使用的专业工具书。《分析化学手册》第一版于1979年出版，有6个分册；第二版扩充为10个分册，于1996年至2000年陆续出版。手册出版后，受到广大读者的欢迎，成为国内很多分析化验室和化学实验室的必备图书，对我国科技进步和社会发展都产生了重要作用。

进入21世纪，随着科技进步和社会发展对分析化学提出的种种要求，各种新的分析手段、仪器设备、信息技术的出现，极大地丰富了分析化学学科的内涵、促进了学科的发展。为更好总结这些进展，为广大读者服务，化学工业出版社自2010年起开始启动《分析化学手册》（第三版）的修订工作，成立了由分析化学界30余位专家组成的编委会，这些专家包括了10位中国科学院院士、中国工程院院士和发展中国家科学院院士，多位长江学者特聘教授和国家杰出青年基金获得者，以及各领域经验丰富的专家。在编委会的领导下，作者、编辑、编委通力合作，历时六年完成了这套1800余万字的大型工具书。

本次修订保持了第二版10分册的基本架构，将其中的3个分册进行拆分，扩充为6册，最终形成10分册13册的格局：

- | | | | |
|----|-----------|----|--------------|
| 1 | 基础知识与安全知识 | 7A | 氢-1核磁共振波谱分析 |
| 2 | 化学分析 | 7B | 碳-13核磁共振波谱分析 |
| 3A | 原子光谱分析 | 8 | 热分析与量热学 |
| 3B | 分子光谱分析 | 9A | 有机质谱分析 |
| 4 | 电分析化学 | 9B | 无机质谱分析 |
| 5 | 气相色谱分析 | 10 | 化学计量学 |
| 6 | 液相色谱分析 | | |

其中，原《光谱分析》拆分为《原子光谱分析》和《分子光谱分析》；《核磁共振波谱分析》拆分为《氢-1核磁共振波谱分析》和《碳-13核磁共振波谱分析》；《质谱分析》新增加了无机质谱分析的内容，拆分为《有机质谱分析》和《无机质谱分析》，并对仪器结构及方法原理进行了全面的更新。另外，《热分析》增加了量热学方面的内容，分册名变更为《热分析与量热学》。

本版修订秉承的宗旨：一、保持手册一贯的权威性和典型性，体现预见性和前瞻性，突出新颖性和实用性；二、继承手册的数据查阅功能，同时注重对分析方法和技术的介绍；三、着重收录了基础性理论和发展较成熟的方法与技术，删除已废弃的或过时的内容，更新有关数据，增补各领域近十年来的新方法、新成果，特别是计算机的应用、多种分析技术联用、分析技术在生命科学中的应用等方面的内容；四、在编排方式上，突出手册的可查阅性，各分册均编排主题词索引，与目录相互补充，对于数据表格、图谱比较多的分册，增加表索引和谱图索引，部分分册增设了符号与缩略语对照。

手册第三版获得了国家出版基金项目的支持，编写与修订工作得到了我国分析化学界同仁的大力支持，全套书的修订出版凝聚了他们大量的心血和期望，在此谨向他们，以及在编写过程中曾给予我们热情支持与帮助的有关院校、科研院所及厂矿企业的专家和同行，致以诚挚的谢意。同时我们也真诚期待广大读者的热情关注和批评指正。

《分析化学手册》(第三版)编委会
2016年4月

前 言

《分析化学手册》是化学工业出版社从 1979 年开始陆续出版的一套全面反映现代分析技术，供化学工作者使用的专业工具书。初版有六分册，1996—2000 年期间，出版了第二版，扩增为 10 个分册，其中第十分册即为 2000 年 12 月出版的《化学计量学》分册（梁逸曾 俞汝勤主编），这一新增的分册出版至今也已有 16 个年头了。大家知道，最近数十年是世界科技快速发展时期，《分析化学手册》可谓见证了这一关键过程。进入 21 世纪，随着科技进步和社会发展，各类分析新手段、新仪器设备、新信息技术的不断涌现，极大地丰富了分析化学学科内涵、促进了学科发展，同时化学计量学方法与近红外光谱结合应用，这一牛刀小试已充分证明了化学计量学的重要性和应用潜力。化学工业出版社自 2010 年起开始启动《分析化学手册》（第三版）的修订出版工作，《化学计量学》分册继续得到出版支持。

大家知道，根据国际化学计量学学会给化学计量学作出的定义：化学计量学是一门通过统计学或数学方法将对化学体系的量测值与体系的状态之间建立联系的学科。它应用数学、统计学与其他方法和手段（包括计算机）选择最优试验设计和量测方法，并通过对方量测数据的处理和解析，最大限度地获取有关物质系统的成分、结构及其他相关信息。

化学计量学是瑞典 Umea 大学 S. 沃尔德（S. Wold）在 1971 年首先提出来的。1974 年美国 B. R. 科瓦斯基和沃尔德共同倡议成立了国际化学计量学学会。化学计量学在 20 世纪 80 年代有了较大的发展，各种新的化学计量学算法的基础及应用研究取得了长足的进步，成为化学与分析化学发展的重要前沿领域。它的兴起有力地推动了化学和分析化学的发展，为分析化学工作者优化试验设计和量测方法、科学处理和解析数据并从中提取有用信息，开拓了新的思路，提供了新的手段。进入新世纪后，化学计量学受不断涌现的现代分析仪器发展使用和复杂体系定性定量分析问题的促进驱动，又获得新的进展，如形成了“数学分离分析”这一绿色分析化学新思路，较系统地发展完善了化学多维校正理论及其在分析领域的新应用，目前正在获得越来越广泛的关注和应用等。

化学计量学的基本任务是研究有关化学量测的基础理论与方法学，其应用数学、统计学与信息理论和方法、计算机科学的方法和手段，科学地设计化学实验，选择最优的量测方法，最有效地获取体系有用的特征数据，并通过解析量测数据最大限度地从中提取有关物质的定性、定量、形态、结构等信息。化学计量学是一门正在发展的新兴学科，其主要研究内容包括统计学与统计方法、校正理论、模型估计和参数估计、实验设计和优化方法、分析信号处理、化学模式识别、定量构效关系、人工智能和专家系统、软件和库检索等。

《化学计量学》分册的修编将继续按照分析化学全过程而展开，从选择分析方法和采样开始，经化学量测的实验设计、信号预处理、定性定量分析的多元校正和多元分辨，再到有用决策信号的提取，包括化学模式识别、机理研究数字模拟方法、化学定量构效关系研究直至人工智能及专家系统等，试图覆盖化学计量学的全部内容，在讨论化学计量学基本概念和

基本方法的数学定义、基本思路和算法的基础上, 尽量同时给出计算机程序及相应的参考文献, 使读者有所收获, 并可进一步用于具体实际问题的解决。

在第三版中, 《化学计量学》分册基本保持第二版的结构框架, 对具体内容有较大幅度的增删, 做了较多的重组、充实和改进等。已删除第二版中第二章, 将相关内容充实于相应的章节。新版的第二、三、四章分别作为采样、实验优化设计以及量测数据的统计评价与质量控制专章, 其中第二版第五章第一节与第二节的内容, 已扩增为新版的第四章, 并突出描述了化学量测与质量控制的关系等。新版第五章新增了分析信号的“背景扣除方法”及“漂移校准方法”两节。第六章在第二版第六章基础上有较大幅度的扩增, 尤其是对灰色分析系统来说, 新增了分量较重的二阶校正及三阶校正等内容, 包括新增了其相应的理论、方法、算法以及应用实例等, 通过引入“数学分离分析”这一绿色分析化学思路, 已真正实现复杂化学体系多目标物的同时、快速、精准之定量分析。第七章新增了“基于机器学习的分类回归方法”一节。第八章为新增的“计算机数字模拟方法”, 第九章也对原化学构效(QSAR)研究方法一章作了较大幅度的增补, 尤其是三维模型等内容的添加, 更有利于 QSAR 研究的深入开展。附录部分内容较为丰富, 除了化学计量学中常用名字的中英文对照表和符号说明外, 还将第二版第二篇内容(第十二、十三、十四、十五章共四章)进行整合, 同时新增增加了不少新方法的 MATLAB 源程序。

第二版中《化学计量学》分册由梁逸曾和俞汝勤担任主编, 第三版中分册主编增添了吴海龙。梁逸曾为本手册的编写, 精心组织、呕心沥血多年, 可不幸于 2016 年 10 月 19 日过劳且患不治之症仙逝, 留下遗憾。在本版前期编著过程中, 梁逸曾从编著大纲的修订到各章节的具体编著, 倾注了大量的心血, 完成了其中大部分的工作量。这里, 除了原先得到杜一平、李晓宁、徐承建、许青松、李博彦等同志的帮助外, 还得到了曹东升、李洪东、范伟、张良晓、张志敏、卢红梅、易伦朝、曾仲达、龚范等同志的帮助和协力, 湖南大学的部分博士生和硕士生胡勇、方焕、刘志、谢丽霞、万瑶、尹小丽、谷惠文、丁玉洁、张燕、张晓华、刘亚娟、夏慧、朱丽、王丽、孙小东、王童、肖蓉等也付出了很多心血, 谨在此一并表示衷心的感谢! 吴海龙继承梁逸曾教授遗志, 不辱使命, 与出版社编辑紧密合作, 终于完成本版的编著。俞汝勤先生从一开始就高度重视本版的编著, 从大纲制订到具体内容的取舍都亲自把关, 倾注了大量心血。本分册第三版的出版, 是对我国著名分析化学与化学计量学家、国际“化学计量学终身成就奖获得者”梁逸曾教授的最好纪念!

本分册涉及的相关科研成果, 主要取自湖南大学化学生物传感与计量学国家重点实验室和中南大学梁逸曾课题组。这些成果的取得多年来曾得到了国家自然科学基金委员会、国家科学技术部和教育部、湖南省科学技术厅和教育厅以及湖南大学、中南大学的大力支持和帮助。谨在此, 一并表示衷心的感谢!

尽管我们在编著过程中尽了最大努力, 希望能以最佳状态呈现在广大读者面前, 但由于知识面及能力的限制, 书中一定会有许多不足之处, 恳请批评指正。

吴海龙 俞汝勤
2016 年 11 月于岳麓山下

目 录

第一章 现代分析化学的信息理论	1	参考文献	35
第一节 化学量测与化学信息	1	第三章 化学试验设计与优化方法	36
第二节 熵——化学量测的“不确定度”的 定量度量	2	第一节 因子设计及其析因分析方法	37
一、分析试验与“不确定度”	2	一、主效应的估价	38
二、“不确定度”与仙农熵	2	二、交叉效应的估价	39
三、仙农熵的性质	3	三、效应及残差正态图	41
四、条件熵与可疑度	3	第二节 部分因子设计	44
五、可疑度、互信息与散度	4	一、半因子设计法	44
第三节 定性分析的信息理论和方法	5	二、半因子设计的产生方法及四分之一 因子设计法	48
一、不同定性分析鉴定方法的信息量 估价	5	三、因子设计分辨率	50
二、仪器定性分析的信息量	6	第三节 正交试验设计和正交设计表	51
第四节 定量分析的信息理论和方法	13	一、正交实验设计	51
一、定量测定的信息量	13	二、正交表及其交互效应表	52
二、提高分析精密性与准确度的信息量 ..	14	三、正交设计表的线性图及其应用	53
三、痕量分析的信息量	14	四、常用正交表及相应交互效应表和线 性图	54
四、质量控制分析的信息量	15	第四节 均匀试验设计及均匀设计表	64
第五节 分析仪器的信息功能	16	一、均匀设计表的构造	64
一、分析仪器与分析方法的供信能力	16	二、均匀性准则和使用表的产生	67
二、分析仪器的信道容量	18	三、拟水平均匀设计	70
三、分析方法的信息效率	20	四、均匀设计表	72
参考文献	22	第五节 最优试验设计	79
第二章 分析采样理论和方法	23	一、最优设计的各种准则	79
第一节 采样的基本概念和理论	23	二、 D -最优设计	81
一、随机采样	23	三、等价定理及 D -最优设计的构造	82
二、系统采样	24	四、构造 D -最优设计的数值方法	89
三、分层采样	24	五、饱和 D -最优设计	92
四、代表性采样	25	六、几种常用最优混合设计表	96
五、最小采样数目的估计	26	第六节 单纯形试验设计法	97
六、采样常数	27	第七节 混料试验设计方法	100
第二节 非均匀体系建模方法及大批物质 的采样误差	28	一、单纯形格子点	102
一、固体物质的采样理论和方法	28	二、单纯形格子设计法	104
二、颗粒性质因子 (G_y 理论)	29	三、单纯形重心设计	106
三、动态过程的采样理论和方法	30	第八节 化学中常用优化方法	109
第三节 质量检验的采样方法	31	一、最优化研究中的基本概念	109
一、计量抽样检验	31	二、局部优化算法	112
二、计数抽样检验	33	三、全局优化算法	121
		四、拉格朗日乘子法	132
		参考文献	133

第四章 分析测量数据的统计评价与质量控制	135
第一节 化学测量的特点及其不确定性的表征与溯源分析	135
一、化学测量的特点	135
二、测量不确定性的定义	136
三、测量不确定性的来源	136
四、测量不确定度的概率性质	137
五、测量不确定度的估计方法	138
第二节 量测误差与质量控制	154
一、量测误差与不确定性	155
二、化学测量中数据特征描述	156
三、化学测量的质量保证与质量控制	160
第三节 分析方法的品质因数及校验方法	169
一、部分分析方法的品质因数	169
二、分析方法的校验	172
三、分析方法的可追溯性	174
第四节 分析量测的数据统计评价与假设检验	177
一、分析结果的两类错误及其统计判决	177
二、分析方法的检测下限	179
三、化学测量数据的统计检验	182
参考文献	187
第五章 分析信号处理方法	190
第一节 分析信号的平滑方法	190
一、移动窗口平均法	190
二、移动窗口多项式最小二乘平滑法	190
三、移动窗口中位数稳健平滑法	194
四、粗糙惩罚平滑法	195
第二节 分析信号的求导方法	198
一、直接差分法	198
二、多项式最小二乘拟合求导法	199
三、基于小波分析的求导方法	204
第三节 分析信号的变换方法	207
一、卷积运算的物理意义	208
二、光谱分析的多重性效益与 Hadamard 变换	209
三、傅里叶变换用于分析信号处理	212
四、小波多分辨变换法及其用于分析信号处理	220
第四节 分析信号的背景扣除方法	232
一、自适应迭代加权惩罚最小二乘背景扣除方法	232
二、分位数回归背景扣除算法	235

三、自动双边指数平滑算法 (ATEB)	243
第五节 色谱分析信号的漂移校准方法	254
一、基于小波模式匹配及微分进化方法的自动色谱谱峰校准方法	256
二、多尺度色谱谱峰校准方法	265
三、基于组分波谱信息的色谱谱峰校准方法	276
四、质谱辅助色谱谱峰校准	281
参考文献	289
第六章 多元校正与多元分辨	294
第一节 概论	294
一、现代分析化学面临的机遇与挑战	295
二、化学计量学的多变量解析思路	297
三、化学计量学中的软模型和硬模型	302
第二节 分析化学中的校正理论	304
一、张量校正理论	304
二、黑、白、灰多组分体系及仪器分析策略	306
第三节 白色分析体系的多元校正方法	308
一、直接校正方法	308
二、间接校正方法	319
三、通用标准加入法	326
四、广义内标法	328
五、非线性体系的校正方法	334
六、病态体系和有偏估计方法	349
七、多元校正的分析化学品质因数和可靠性分析	353
第四节 灰色分析体系的多元校正方法	356
一、矢量校正方法	356
二、矩阵(非迭代类二阶张量)校正方法	367
三、三维(迭代类二阶张量)校正方法	387
四、四维校正(三阶校正)	405
第五节 黑色分析体系的多元分辨方法	412
一、基于主成分分析的体系组分数确定方法	413
二、矩阵分辨方法	426
三、张量分辨方法	480
第六节 广义灰色分析体系的多元校正模型	492
一、近红外光谱与广义灰色分析体系	492
二、广义灰色分析体系的模型校验方法	493
三、广义灰色分析体系的常用多元校正方法	504
第七节 回归建模中的稳健方法	510
一、回归诊断方法 (regression diagnostic methods)	511

二、稳健回归方法	517	Carlo 计算机模拟方法	608
参考文献	523	四、误差分析的 Monte Carlo 模拟 算法	610
第七章 化学模式识别	531	第二节 基于微分方程数字解法的计算机 模拟方法	612
第一节 模式空间的几种距离与相似性 度量	531	第三节 电化学过程的数字模拟	616
一、化学模式空间	531	参考文献	621
二、模式空间的距离与相似性度量	532	第九章 化学构效关系的研究方法	622
第二节 特征抽取方法	534	第一节 化学构效关系研究起源——有机 反应性相关分析方法简介	622
第三节 模式识别的数据预处理方法	535	一、线性自由能概念及其经典 Hammett σ 常数	623
第四节 有监督的模式识别方法——判别 分析法	536	二、取代基电子效应常数 (σ 常数) 的 应用与拓展	623
一、距离判别法	536	三、取代基的立体效应常数	625
二、Bayes 判别法	537	四、取代基电子效应常数与立体参数的 协同效应	632
三、Fisher 判别法	539	第二节 化学构效关系研究的基本假设及 基本过程	632
四、逐步判别分析法	542	一、化学图论基本概念	633
五、线性学习机	548	二、几种重要的化学图矩阵	634
六、K-最近邻法	549	第三节 分子描述子	636
七、势函数判别法	550	一、分子描述子的定义	636
八、人工神经网络判别法	551	二、分子描述子的分类	636
第五节 无监督的模式识别方法——聚类 分析法	553	第四节 化合物结构的拓扑指数表征	637
一、系统聚类法	553	一、Wiener 拓扑指数	638
二、一分为二法	560	二、Randic 分支指数和分子连接性 指数	639
三、最小生成树法	561	三、分子识别数	645
四、K-均值聚类法	562	四、苏尔兹分子拓扑指数	648
五、基于全局寻优的聚类法	563	五、回归距离和 (RDS) 及回归顶点点价 (RVD)	652
六、模糊聚类法	567	六、分子形状及灵活性指数 (Kappa) 指数	653
第六节 基于投影的模式识别方法	571	七、电子拓扑状态——QSAR 及数据库分析 中的结构模拟参数	655
一、基于主成分分析的投影判别法	571	八、邻接复杂度的信息量指数及其 应用	662
二、基于主成分分析的 SIMCA 分类法	576	第五节 化学构效关系 (QSAR) 研究的 三维模型	667
三、基于偏最小二乘的投影判别法	584	一、距离几何的 3D-QSAR 方法	669
第七节 基于机器学习的分类回归方法	589	二、分子形态分析方法	671
一、支持向量机	590	三、比较分子场 QSAR 分析方法	673
二、分类回归树	593	四、虚拟受体方法	675
三、助推建模	596	第六节 QSAR/QSPR 中的建模方法及其 应用	678
四、随机森林	600		
参考文献	601		
第八章 计算机数字模拟法	604		
第一节 基于统计机理的 Monte Carlo 数字 模拟法	604		
一、伪随机数的产生方法	604		
二、化学动力学系统的 Monte Carlo 模拟 算法	606		
三、适用于所有均相化学反应体系的 Monte			

一、基于回归分析的建模方法及其应用	680
二、基于模式识别的建模方法及其应用	694
参考文献	706

第十章 人工智能与化学专家系统

方法	711
第一节 启发式分类与搜索方法	711
一、广度优先搜索	713
二、深度优先搜索	713
三、启发式搜索方法	713
第二节 知识表达技术	714
一、逻辑表达方法	714
二、语义网络表达法	714
三、产生式规则表达法	715
第三节 化学专家系统简介	716
一、DENDRAL 质谱、核磁共振谱图解析专家系统	717
二、PLATO 数据解析专家系统	719
三、高效液相色谱专家系统	720
四、ESESOC 有机化合物结构解析专家系统	721
参考文献	722

第十一章 统计学和应用数学基础

知识	723
第一节 必要统计学基础知识	723
一、随机事件的概率公式	723
二、随机变量及其分布	725
三、随机变量的数值特征	729
第二节 必要应用数学基础知识	730
一、矢量及其运算	730
二、矩阵及其运算	732
三、独立性、正交性和子空间	737
四、矢量范数和矩阵范数	739
五、张量	740
参考文献	740

附录	741
I 化学计量学中常用名词及符号说明	741
一、化学计量学中常用名词中英文对照表	741
二、符号说明	751
II 使用 MATLAB 语言进行化学计量学研究编程运算简介	752
一、MATLAB 语言简介	752
二、MATLAB 与化学计量学	752
三、MATLAB 系统简介	753
四、MATLAB 的工作环境	754
五、MATLAB 的一般运算符号和矩阵运算功能	756
六、MATLAB 的绘图功能	773
七、ATLD 算法的 MATLAB 程序	777
八、SWATLD 算法的 MATLAB 程序	779
九、APQLD 算法的 MATLAB 程序	781
十、APQLD 算法的 MATLAB 程序	784
III R 语言及其在化学计量学中的应用	789
一、R 是什么	789
二、R 的安装	789
三、加载包	790
四、在 R 中如何获取帮助	792
五、R 中的数据	792
六、R 语言的基础运算操作	794
七、R 语言中的常用函数	805
八、R 语言的基础画图	808
九、R 中的化学计量学	814
IV Python 语言及其在化学计量学中的应用	822
一、Python 语言简介	822
二、Python 语言中的基础运算操作	824
三、R 与 Python 语言在数据分析方面的对比分析	831
四、Python 与 MATLAB 语言线性代数计算方面的对比分析	831
五、Python 语言中的基础绘图操作	836
六、常用化学计量学算法的 Python 语言源代码	844
主题词索引	854

第一章 现代分析化学的信息理论

第一节 化学量测与化学信息

近年来随着计算机，特别是微型计算机大量进入化学特别是分析化学实验室，一门新的化学分支学科——化学计量学应运而生，为有效进行化学量测和提供化学信息开辟了新通路，为分析化学的发展提供了新机遇。化学计量学是一门交叉学科，它应用数学、统计学与计算机科学工具和手段及其最新成果来设计或选择最优化学量测方法，并通过解析化学量测数据以最大限度地获取化学及其相关信息，自然，它首先就在分析化学中得到了普遍认同。自20世纪70年代中期诞生以来，在80年代得到长足发展，至今已日趋成熟。现在，让我们来看看化学计量学与分析化学作为一门化学量测和化学信息学科之间的关系。

从图1-1可以看出，对于化学量测的每一步，化学计量学都有相应的理论和方法学，研究如何来使化学量测和数据解析过程变得更有效，实质上，化学计量学就是一门关于化学量测的理论基础和方法学的化学分支学科^[1]。

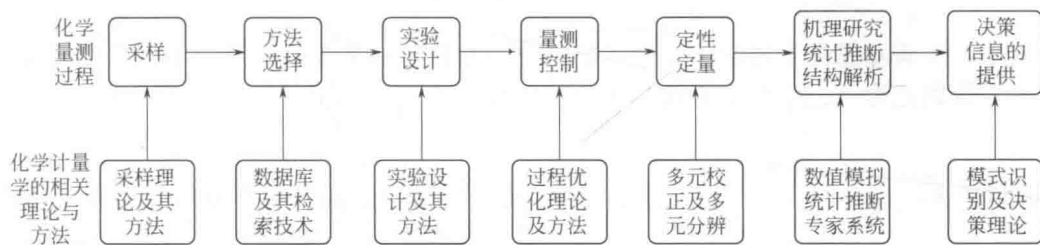


图 1-1 化学量测流程与化学计量学方法的关系示意图

分析化学作为一门化学量测和化学信息学科，对其量测过程的有效性、效率的估计和评价就显得十分重要了。分析信息理论从信息理论的角度来研究化学量测过程。如果将通信信息处理的过程与化学分析中的化学量测过程来进行比较，就可容易发现这两个过程十分相似（参见图1-2）^[2]。

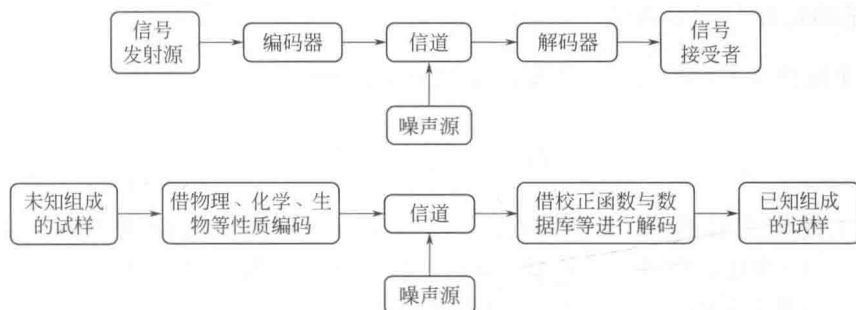


图 1-2 通信信息处理过程与化学量测过程的比较

实际上,进行化学量测的目的本质上就在于消除或减少被量测的化学系统在化学成分、结构及其他相关信息的“不确定度”。分析信息理论和方法可提供相应的概念和方法来定量表征化学量测系统的“不确定度”及化学量测过程中体系之“不确定度”的消除或减少的定量度量(或称为化学量测过程的信息量之获取)。它不但可以对不同的分析过程,如分离、定性鉴定、定量测定等进行信息理论分析,还可对现代分析化学中各类复杂分析仪器的提供信息之能力进行合理评价,为分析化学量测提供了一套选择不同分析过程或不同分析仪器的理论基础和具体定量方法。本章将系统地介绍这些概念和方法。

第二节 熵——化学量测的“不确定度”的定量度量

一、分析试验与“不确定度”

分析工作的目的是取得有关未知试样的化学成分与结构的相关信息。所以,在进行分析测试之前,必然存在某种不确定性或称“不确定度”。设有一份试样,其中可能含有 k 种离子中的某一种,定性分析的任务就是确定这一离子是何种离子。在分析测试之前,存在有 k 种可能性,或者称分析实验有 k 个可能的结局。如将上述情况用数学式表示,并设上述 k 种可能结局为 a_1, a_2, \dots, a_k ,发生这 k 种可能结局的概率为 P_1, P_2, \dots, P_k ,则有

$$A = \frac{a_1, a_2, \dots, a_k}{P_1, P_2, \dots, P_k}$$

式中, A 表示发生上述定性分析实验的一个事件。如果存在有两个不同定性分析实验的事件,分别记为 A_1 和 A_2 ,并由以下的两个数表表示出,即

$$A_1 = \begin{matrix} a_1, a_2, \dots, a_k \\ 1, 0, \dots, 0 \end{matrix}; \quad A_2 = \begin{matrix} a_1, a_2, \dots, a_k \\ 1/k, 1/k, \dots, 1/k \end{matrix}$$

很明显,因为对于事件 A_1 ,它发生的可能性实际上只有一种,即只可能发生 a_1 ,而发生 a_2, \dots, a_k 的概率都等于零,所以,此事件的“不确定度”实际上不存在,如果存在这样一种“不确定度”的定量度量标准的话,对于事件 A_1 ,其值应为零;然而,对于事件 A_2 ,它发生的可能性就有 k 种,即发生 a_1, a_2, \dots, a_k 的概率都相等,故此事件的“不确定度”很大,如存在一种“不确定度”的度量标准的话,对于事件 A_2 ,其值应该很大或至少要大于零。以下将要讨论的熵的概念,就是这样“不确定度”的一种定量度量。

二、“不确定度”与仙农熵

对于一个随机事件的数表,仙农熵的定义如下:

$$H = - \sum_{i=1}^k P_i \log P_i \quad (1-1)$$

仙农给出了上述定义并称 H 为熵。 \log 一般表示以 e 为底,此时所得熵的单位为奈特(nat);如果以10为底,其单位为特(dit);以2为底,其单位为比特(bit)。从上式可知,如对前述的两个实例,用式(1-1)计算可得

$$H(A_1) = -1 \log 1 = 0$$

$$H(A_2) = - \sum (1/k) \log(1/k) = - \log(1/k) = \log(k)$$

从上述结果明显可知, 事件 A_2 的不确定度大于事件 A_1 的不确定度。

三、仙农熵的性质

(1) 非负性, 即

$$H(P_1, P_2, \dots, P_k) \geq 0$$

这是因为 $P_i (i = 1, 2, \dots, k)$ 为概率, 故有 $1 \geq P_i \geq 0$, 所以 $\log(P_i)$ 不可能取正值, 即 $(-\sum P_i \log P_i)$ 才取正值。

(2) 对于确定性事件, 其熵为零。即

$$H(1, 0, \dots, 0) = 0$$

所谓确定性事件, 即为该事件只有一种可能结果, 如用数学语言表达则为, 对于发生某一可能结果的概率为 1, 而发生其他可能结果的概率都为零。前面讨论的事件 A_1 就是确定性事件的一个例子。注意在此实际上是引入了一个人为的假设, 即

$$\lim_{P \rightarrow 0} (-P \log P) = 0 \quad (1-2)$$

式中, P 为一表示概率的实数。

引入这一假设是因为 $\log(0)$ 在数学上为一无意义的数。可是, 如从极限的角度来看, 引入式 (1-2) 在数学上是完全合理的。如假设 $P = 1/e^n$, 此时

$$-\log P = n$$

而

$$-P \log P = n/e^n \rightarrow 0 \quad (1-3)$$

随着 n 的增大, n/e^n 将很快接近于零。实际上因为此时 P 是一个比 $(-\log P)$ 更高阶的无穷小, 所以上式实际上是用极限的概念来避免对 $\log(0)$ 的直接计算。有了式 (1-3) 的定义, 所以有

$$H(1, 0, \dots, 0) = -1 \log(1) - \sum 0 \log 0 = -1 \log(1) = 0$$

(3) 对于等概率结果的事件, 其熵最大。即

$$H(P_1, P_2, \dots, P_k) \leq H(1/k, 1/k, \dots, 1/k) = \log(k) \quad (1-4)$$

所谓等概率结果的事件, 即该事件有 k 种可能结果, 而且, 发生这 k 种结果的概率都相等, 即发生某一可能结果的概率都为 $(1/k)$ 。由此性质可知, 等概率结果的事件的不确定度最大。

四、条件熵与可疑度

今考察两个随机试验 A 与 B, 其可能结果为

$$A: A_1, A_2, \dots, A_n$$

$$B: B_1, B_2, \dots, B_m$$

则两事件同时发生时的熵为

$$H_{n,m}(AB) = - \sum_{i=1}^n \sum_{j=1}^m P(A_i B_j) \log[P(A_i B_j)] \quad (1-5)$$

条件熵的定义：如果试验 A 出现结果 A_k 时，试验 B 在此条件下的熵为

$$H_m(B|A_k) = - \sum P(B_i | A_k) \log[P(B_i | A_k)]$$

相对应也可定义试验 B 出现结果 B_k 时，试验 A 在此条件下的熵

$$H_n(A|B_k) = - \sum P(A_i | B_k) \log[P(A_i | B_k)]$$

更进一步，还可定义在进行试验 A 的前提下试验 B 的条件熵

$$\begin{aligned} H(B|A) &= \sum P(A_k) H_m(B|A_k) \\ &= \sum P(A_k) [- \sum P(B_i | A_k) \log[P(B_i | A_k)]] \\ &= - \sum \sum P(A_k) P(B_i | A_k) \log[P(B_i | A_k)] \end{aligned}$$

或在进行试验 B 的前提下试验 A 的条件熵

$$H(A|B) = - \sum \sum P(B_k) P(A_i | B_k) \log[P(A_i | B_k)] \quad (1-6)$$

条件熵的性质：

$$\textcircled{1} H(A|B) = H_{n,m}(AB) - H_m(B)$$

$$H(B|A) = H_{n,m}(AB) - H_n(A)$$

此性质可由积事件（即同时发生的事件）的概率公式（参阅第十一章）直接导出。

$$\textcircled{2} H(A|B) \leq H_n(A)$$

$$H(B|A) \leq H_m(B)$$

此性质的证明可参阅文献 [2]。

③ 如果随机试验 A 与 B 相互独立，此时 $P(A|B) = P(A)$ （参阅第十一章），亦即分析试验 B 与待解决的分析课题（试验 A）毫不相干，当然做试验 B 将得不到试验 A 的任何信息，也无从减少关于试验 A 的“不确定度”。此时，有

$$H_{n,m}(AB) = H_n(A) + H_m(B) \quad (1-7)$$

五、可疑度、互信息与散度

可疑度的定义：从条件熵的定义可知， $H(A|B)$ 实际表述了在进行了试验 B 以后的试验 A 的“不确定度”，所以， $H(A|B)$ 又称为可疑度。这样定义的可疑度，实际上可以说是反映了分析仪器或方法提供信息的能力与解决给定分析课题的需要之间的差距的定量度量。

互信息的定义：

$$I(A;B) = H_n(A) - H_n(A|B) = H_m(B) - H_m(B|A) = I(B;A) \quad (1-8)$$