



普通高等教育“十一五”国家级规划教材

全国高等学校教材 • 供药学类专业用

计算机 在药学中的应用

主编 董鸿晔



人民卫生出版社



中国医药出版社

医药类核心期刊 · 医药类核心期刊

计算机 在药学中的应用

主编 王小明

中国医药出版社

普通高等教育“十一五”国家级规划教材

全国高等学校教材 · 供药学类专业用

计算机 在药学中的应用

主 编 董鸿晔
编 者 (以姓氏笔画为序)
于 净 (沈阳药科大学)
卢虹冰 (第四军医大学)
李祥生 (山西医科大学)
金 杰 (沈阳药科大学)
周 怡 (广东药学院)
晏峻峰 (湖南中医药大学)
海 滨 (中国药科大学)
董鸿晔 (沈阳药科大学)

人民卫生出版社

图书在版编目 (CIP) 数据

计算机在药学中的应用 / 董鸿晔主编. —北京: 人民卫生出版社, 2010.8

ISBN 978-7-117-13193-3

I. ①计… II. ①董… III. ①计算机应用—药学—医学院校—教材 IV. ①R9-39

中国版本图书馆 CIP 数据核字 (2010) 第 124525 号

门户网: www.pmph.com	出版物查询、网上书店
卫人网: www.ipmph.com	护士、医师、药师、中医师、卫生资格考试培训

版权所有, 侵权必究!

本书本印次封底贴有防伪标, 请注意识别。

计算机在药学中的应用

主 编: 董鸿晔

出版发行: 人民卫生出版社 (中继线 010-59780011)

地 址: 北京市朝阳区潘家园南里 19 号

邮 编: 100021

E - mail: pmph@pmph.com

购书热线: 010-67605754 010-65264830

010-59787586 010-59787592

印 刷: 三河市富华印刷包装有限公司

经 销: 新华书店

开 本: 787×1092 1/16 印张: 11.5

字 数: 278 千字

版 次: 2010 年 8 月第 1 版 2010 年 8 月第 1 版第 1 次印刷

标准书号: ISBN 978-7-117-13193-3/R·13194

定 价: 22.00 元

打击盗版举报电话: 010-59787491 E-mail: WQ@pmph.com

(凡属印装质量问题请与本社销售中心联系退换)

本教材为普通高等教育“十一五”国家级规划教材,是一本结合药学专业实际情况,重点讲授计算机的具体应用,以及培养学生使用和开发药学应用软件能力的入门教材。教材共分为9章,重点讲述药学试验设计与优化、药学计算可视化、药学实验数据统计分析、计算机辅助测量方法及其应用、计算机仿真技术及其应用、人工神经网络技术及其应用、数据挖掘技术及其应用、药学信息学及网络药学等计算机在药学中的应用。本书作为高等学校药学类各专业本科高年级学生及研究生教材,同时也可作为药学工作者的参考用书。

本教材的主要内容都有相应的教学演示文档、实例数据和部分应用软件提供。实验指南作为附录,放在教材内容末。

前言

随着现代科学技术的发展和计算机的广泛使用,各学科对计算机的依赖程度越来越高,药学领域也不例外。从实验数据的处理及拟合、模型参数的确定、非线性方程的求解到药学过程模拟,均离不开计算机的帮助,对这方面内容计算机主要发挥的是高速的数值计算功能;另一方面,我们还要利用计算机进行药学信息的发布、药学信息检索、药学数据挖掘等一系列其他非计算性的工作,同时还需利用计算机进行药学实用软件的开发工作。这些需求促进了《计算机在药学中的应用》这门课程的发展,国内部分高等医药院校开设这门课程已有二十余年的历史了。

《计算机在药学中的应用》是一门旨在提高药学各专业大学生计算机应用水平的专业课程。尽管学生在基础阶段的学习中已经学过《大学计算机基础》、《计算机程序设计》等有关计算机的基础课程,但在毕业设计阶段还经常碰到有些学生无法利用计算机开展毕业设计有关工作的问题,如:利用计算机进行网上文献检索、实验模型参数的确定、药动学模型计算、药学专业论文的编辑、药学信息的多媒体发布、常用药学计算软件应用,以及实用药学历程设计或软件的开发。产生上述问题的主要原因是基础阶段的学习中讲授的仅是计算机的基本理论和基本知识,没有讲授这些理论和知识在具体专业中的应用。而《计算机在药学中的应用》正是结合专业的实际情况讲授计算机的具体应用,是培养学生使用药学应用软件和开发药学应用软件的入门课程。

现在,经过全国部分高等医药院校的同仁共同努力,《计算机在药学中的应用》一书纳入了普通高等教育“十一五”国家级规划教材,即将正式出版了。本书力求遵循简明、实用的原则,对药学实验数据处理、常用药学计算及模拟等需要用到复杂数学知识的内容,以简单实用的形式呈现给读者,并提供立即可以应用的程序;对一些常用软件及药学应用软件的介绍采用实例应用的形式,对于新开发的药学应用软件,着重于介绍软件开发的背景、环境及方法,力争为读者提供一种药学软件开发的基本思路。

我们倾听国内部分高等药学院系计算机基础教学第一线的声音,设计了本书的编写大纲,并请多年从事教学工作、经验丰富的教师参加编写。我们尽力从计算机在药学中的应用实际出发,认真综合计算机较新的技术,构成了本书的主线,力图为我国药学高级专门人才的培养做出一点努力。本书得到国内许多专家学者的关注和支持,参考了一大批他们的专著。本书许多内容都经过了多年教学和科研应用,有很好的实践基础。

本书第1章和第9章由董鸿晔编写,第2章由董鸿晔、海滨编写,第3章由于净编写,第4章由董鸿晔、李祥生编写,第5章由金杰编写,第6章由董鸿晔、晏峻峰编写,第7章由卢虹冰、董鸿晔编写,第8章由董鸿晔、周怡编写。全书最后由董鸿晔定稿。每章后面都有少量思考题。实验指南由董鸿晔、于净编写,作为附录内容,放在本书最后。使用本教材建议安排教学36~40学时、实验28~32学时。

本教材的主要内容都有相应的教学演示文档和相应的实例数据、应用软件及用法介绍,同时也提供了部分可执行的应用程序,有助于加深读者对教材知识的理解,而且也为药实验数据处理及模拟提供了帮助,增加本教材的实用性。有需要这些材料的读者可与作者取得联系(donghy001@163.com)。

几十年前,我们刚刚接触计算机的时候,做什么事情都必须自己进行程序设计,哪怕只要显示一个汉字。随着计算机技术的发展,适应各行各业各种需求的软件大量涌现出来,十几年来我们编写的药计算应用软件就有一个足够大的集合了。计算机在药中的应用已经不再需要完全依赖自己编程了。因此本书在编写方式上改变了策略,一方面介绍计算机在药中的应用的基本原理、特点和方法,不过多要求掌握复杂的算法;另一方面主要配合实例讲解具体软件的使用方法。从使用的目的出发,研究使用的方法,引导读者亲自去使用,达到灵活使用解决实际问题的目标。

本书所引实例,大多是作者及许多同仁多年完成国家自然科学基金等计算机、药学科研课题的经验总结,反映了计算机科学对自然科学和人类社会产生的变革在药学领域的巨大影响。我们深信,计算机在药中的应用正在形成一门科学,它的理论方法和本书介绍的软件用法用例必将成为广大药学工作者必不可少的工具。我们愿将本书献给我国广大药学工作者,希望本书成为广大药学工作者爱不释手的伙伴。本书既可作为普通高等学校药类各专业的研究生和本科高年级专业教学用书,也可作为广大药学工作者的参考用书。

本书编写过程中,作者的单位沈阳药科大学、中国药科大学、广东药学院、第四军医大学、山西医科大学和湖南中医药大学都给予了大量支持,在此一并表示感谢。沈阳药科大学药信息学重点实验室(筹)的全体研究生参加了本教材中所有实例的验证工作。由于作者水平有限,教材中难免有缺点和错误,恳切希望应用本教材的广大师生和读者指正。

董鸿晔

2010年3月4日于沈水湾

目 录

第一章 计算机在药学中的应用概述	1
第一节 计算机科学与药学研究	1
第二节 药学应用软件与工具简介	2
一、Excel 软件简介	3
二、MATLAB 工具箱简介	5
三、SQL Server 2005 数据挖掘工具简介	6
四、JMP 应用软件简介	8
第三节 如何开发满足药学应用特殊需求的软件	10
一、应用软件开发流程	10
二、软件开发语言工具	10
思考题	11
第二章 药学试验设计与优化	12
第一节 试验设计与优化方法概述	12
一、试验设计中的基本概念	12
二、常用试验设计与优化方法	12
三、试验设计与优化的基本步骤	13
第二节 基于计算机的试验设计与优化方法的实现	13
一、正交试验设计	13
二、均匀设计	17
三、响应曲面法	19
思考题	22
第三章 药学计算可视化	23
第一节 药学计算可视化方法概述	23
第二节 药学计算可视化方法的设计与实现	24
一、基于 MATLAB 的概率单位正规法半数致死量计算	24
二、基于 VB6.0 的固体制剂体外溶出试验数据计算	28
思考题	32
第四章 药学实验数据统计分析	33
第一节 药学实验数据统计分析方法概述	33

第二节 基于 Excel 的药学实验数据统计分析	33
一、描述统计分析	34
二、假设检验	37
三、回归分析	39
第三节 基于 MATLAB 的药学实验数据统计分析	41
一、聚类分析	41
二、判别分析	44
三、主成分分析	47
思考题	51
 第五章 计算机辅助测量方法及其应用	53
第一节 计算机辅助测量技术概述	53
第二节 基于图像的计算机辅助测量方法的设计与实现	53
一、系统组成及工作原理	54
二、软件设计	54
第三节 计算机辅助测量技术在药学中的应用	55
一、抗生素效价的计算机辅助测量	55
二、药物粒径的计算机辅助测量	60
三、病理组织切片的计算机辅助测量	65
思考题	69
 第六章 计算机仿真技术及其应用	70
第一节 计算机仿真技术概述	70
第二节 计算机仿真系统的设计与实现方法	70
一、Simulink 仿真工具箱简介	71
二、计算机仿真模型设计	71
第三节 药动力学参数计算仿真模型设计与应用	74
一、药动力学算法描述	74
二、药动力学参数计算仿真模型设计	76
思考题	85
 第七章 人工神经网络技术及其应用	86
第一节 人工神经网络技术概述	86
一、人工神经网络基本原理	86
二、典型人工神经网络模型结构	88
三、人工神经网络在药学中的应用	92
第二节 基于 MATLAB 的人工神经网络模型设计	92
一、MATLAB 神经网络工具箱使用简介	93
二、人工神经网络模型设计与应用	94

思考题.....	102
第八章 数据挖掘技术及其应用.....	103
第一节 数据挖掘技术概述.....	103
一、数据挖掘定义及含义	103
二、数据挖掘方法	103
三、数据挖掘过程	104
四、数据挖掘的应用	105
第二节 基于 SQL Server 2005 数据挖掘工具的数据挖掘	106
一、SQL Server 2005 数据挖掘工具简介	106
二、数据仓库、联机分析处理与数据挖掘	106
三、基于 SQL Server 2005 的数据挖掘解决方案	107
第三节 基于 SQL Server 2005 的中药指纹图谱数据挖掘	107
一、中药指纹图谱数据仓库的构建	108
二、数据准备	108
三、创建指纹图谱数据库, 数据结构化	112
四、建立数据挖掘方案	115
思考题.....	126
第九章 药学信息学及网络药学.....	127
第一节 药学信息学及网络药学与药物科技创新.....	127
一、药学信息学及网络药学概述	127
二、促进药物科技创新的途径	128
第二节 基于药学信息学的药学计算网络平台设计与应用.....	129
一、开发环境的构建与开发流程	129
二、系统设计	130
三、网页界面设计	139
思考题.....	143
附: 实验指南	144
实验一 药学试验设计与优化.....	144
实验二 药学计算可视化.....	149
实验三 基于 Excel 的药学实验数据分析	152
实验四 基于 MATLAB 的药学实验数据分析	155
实验五 基于 MATLAB 神经网络工具箱的药物处方优化	159
实验六 基于 MATLAB 的药动力学参数计算仿真模型设计	163
实验七 基于 SQL Server 2005 的药学数据挖掘	170

计算机在药学中的应用概述

第一节 计算机科学与药学研究

21 世纪是一个信息时代,信息在医药领域的地位和作用已日益引起人们的重视,开发利用现代药学信息技术、加快医药科技创新,已成为医药现代化发展的一个不可缺少的条件。现代信息技术对药物研究的促进作用是加快实现药物研究科技创新的途径。目前,医药科技和信息科技是当今世界发展最快和影响最大的学科。受其自身发展内在逻辑力量的强力驱动,医药科技和信息科技正从宏观和微观两个方向不断深入拓展。药物、化学和信息等学科交叉汇聚,融合加速,学科界限日趋模糊,思路不断突破,逐渐形成新的科技体系,进入了前所未有的多学科集群创新时代。当代医药研究的领先权,在很大程度上取决于信息技术手段的先进程度。因此,信息处理技术理所当然地成为药学及生物医药等领域的重要研究前沿和技术支撑点。

药物研究与开发是一个复杂而漫长的过程。一般要经过药物资源调查、药物设计与筛选、化学合成与改造、药效学、药物代谢、安全性评价、工艺及制剂、质量检测与控制、临床评价、中试放大、产业化生产、市场反馈等多个阶段。在整个开发过程中,药学科研工作者需要针对各种复杂的对象体系构建多种研究方案,需要完成不同阶段的各种实验和测试,产生大量的实验数据和研究信息,从而形成药物研究信息流程。随着数据库技术的发展,各研究机构及企事业单位建立了相应的药学信息数据库,但由于缺少分析方法及信息处理平台,虽然数据越来越多,而数据带给研究者的辅助决策信息却很贫乏。如何将大量的数据转化为辅助决策信息,如何对数据进行挖掘以便分析评价实验结果,获取药物研发过程的相关信息,以及总结和探索研究对象的规律性等已成为药学科研领域亟待解决的关键问题和研究热点。

随着药学科研工作更加深入和广泛的开展,药学科研工作者需要使用各种新型的仪器设备与先进的分析技术,需要缩短新药开发周期和处理大量的多变量数据。昔日以化学分析为主的经典分析化学,已发展成为一门以众多仪器分析(包括:色谱、光谱、质谱、核磁共振及各类仪器联用等)为主的现代分析化学。计算机技术与分析仪器的结合促进了化学及药学量测数据获取方法的重大进步,这一进步不仅实现了分析仪器数据的自动采集、传递和储存,而且使分析仪器的自动化操作成为现实。在药学科研领域不断提出的越来越高的各种分析要求面前,人们认识到计算机科学和融入了信息技术的新型分析工具和分析方法,将为解决药学及生命科学等许多学科所提出的复杂研究体系辨识难题及自动化地提取和解

析仪器分析实验数据,提供强有力的技术工具。

为应对人类疾病谱变化及健康观念转变,医药科学和药学研究正面临前所未有的严峻挑战和重大发展契机,对中药及天然药物、化学合成药物、生物技术药物、药物制剂等新药的开发创新工作提出了迫切要求。为提高药物研发速度、解决药学研究难题,代谢组学、基因组学、蛋白质组学及化学组学等“组学”正在信息学旗帜下汇集,各类医药研究数据的信息量和复杂性正在迅速增加,这一学科潮流及技术走势驱使现代药学向信息化、数字化和计算智能化发展。药学信息学应运而生。

药物信息学(pharmainformatics)是药学与计算科学等各相关学科交叉融合而产生的新兴边缘学科,它是运用信息技术及计算分析方法收集、管理、辨析和处理药学研究数据,获取相关信息,进而解释药物实验现象、发现新的知识及规律、指导药物研究、加速新药创制以及确保合理用药的科学。药物信息学顺应药物学科自身的需求而带有鲜明的药学特点,具有广阔的研究范围及发展前景。

近年来,随着计算机分子图形学的发展,新药化合物设计已进入三维定量构效关系技术发展阶段,可对配体和受体三维结构进行识别和优化计算,直接为药物分子定量优化设计提供关键技术,形成了计算化学、组合化学和高通量筛选三位一体的计算机辅助药物分子设计技术,其发展趋势是将信息技术全面和系统地应用于药物研发、药品生产、药物作用机制研究、药品质量控制、临床合理用药等各个方面。例如:应用动力学原理和数学建模方法定量描述药动学;将定量构效关系、虚拟药动学、计算毒理学等各种计算药学技术集成,发展虚拟筛选技术,创建基于知识的计算工具来解决药物开发中的高通量筛选问题;运用模式识别、数据挖掘、机器学习等计算智能方法,建立计算机辅助三维药物设计技术,采用分子对接、构效关系、分子类药性、多样性、虚拟筛选等方法进行药物分子设计;在药物临床应用领域,研究个体化给药方案的构建方法和“数字药房”的解决方案以及各种专家系统,并用信息化方法对药物应用进行经济学评估等。

从某种意义上看,新药开发能力与药物研究科技创新活动的水平取决于人们对信息科学的掌握和利用程度。近年来,我国逐步认识到利用药学信息技术实现药学资源共享、获得决策支持、提供药学服务的重要性,但由于药学信息学是一个新兴的领域,其发展具有很大的复杂性,仍有许多应用需要深入研究,特别是药学信息学应用平台的建设是目前亟待解决的重大关键问题。依托药学信息学应用平台,构建创新药研究过程数据仓库,对实验数据及研究过程进行数据挖掘和知识发现,将数据转化为辅助决策信息,不仅可以积极促进药学信息资源的开发利用,加强各国药学信息中心及医药企业信息化建设,而且可以提高药学科人员的信息意识,促进创新药物的研制,这将对药学研究领域、医药产业化、中药现代化等多方面产生积极影响。

第二节 药学应用软件与工具简介

随着计算机技术发展以及应用范围的普及,大量的计算机应用软件不断问世,应用软件涵盖了数据处理、数值计算、统计分析、优化设计、图形图像处理、智能控制、多媒体、网络通信、数据库及信息管理等众多应用领域。这些应用软件的种类繁多,

普遍具有功能多、通用性强、算法成熟、资源丰富等特点,用以满足不同专业及用户的需求,深受广大科技工作者的欢迎。下面针对药学应用实际情况介绍几个常用软件与工具。

一、Excel 软件简介

著名的 Microsoft Excel 为众多药学工作者广泛使用。作为普通的用户,人们往往简单使用其丰富的电子表格功能,对其内置的“工作表函数”、“数据分析”等自动化工具了解不够。由于药学研究离不开实验,实验数据处理工作量大且需要一定的统计学基础,利用 Excel 的“工作表函数”和“数据分析”工具可以在药学科研工作中实现数据处理的自动化和流程化。特别是利用 Excel 创建的工作表存储数据具有良好的兼容性,可实现与众多的应用软件的数据直接交互访问。因此,Excel 是一种应用广泛、高效率,高精确度而又经济的自动化数据处理工具软件。

(一) Excel 工作表函数

Excel 工作表函数是指为了达到某些专业计算目的,由系统预定义的特定代数化公式,多为较复杂的长公式串。系统会自动把源数据(参数)代入公式,并按照既定的计算流程(Excel 称其为语法)进行运算,返回求解的计算结果或做出逻辑判断。

1. Excel 内置的统计函数 Excel 内置了数学函数、统计函数、工程函数等 13 大类函数,共 300 余种,药学工作中常用的统计函数也有几十种,例如 LINEST(线性回归函数)、CORREL(相关系数)、LOGEST(指数回归拟合曲线方程)、NORMDISI(正态分布函数)、STDEVA(给定样本的标准差函数)、STDEVAP(给定样本的总体标准差函数),等等。其中 10 种最常用的工作表函数被置于 Excel 公式编辑栏的下拉列表框内,使用时可信手拈来,其他的统计函数可通过加载或在 Excel 的 VBA 中使用。药学中经常使用的 Excel 统计函数见表 1-1。

表 1-1 药学中常用的 Excel 统计函数

序号	函数	功能
1	AVEDEV	返回一组数据点到其平均值的绝对偏差的平均值
2	AVERAGE	返回参数的平均值
3	AVERAGEA	返回参数的平均值,包括数字、文本和逻辑值
4	BETADIST	返回 Beta 分布的累积函数
5	BINOMDIST	返回一元二项式分布的概率值
6	CHIDIST	返回与 χ^2 分布相关的单尾概率值
7	CHITEST	返回独立性检验值
8	CONFIDENCE	返回样本总体平均值的置信区间
9	CORREL	返回两组数据集合的相关系数
10	COVAR	返回协方差,即每对偏差乘积的平均值
11	CRITBINOM	返回一个使二项式分布的函数值小于或等于临界值的最小数值
12	DEVSQ	返回偏差平方和

续表

序号	函数	功能
13	EXPONDIST	返回指数分布
14	FDIST	返回 F 概率分布
15	FORECAST	根据通过给定数据的线性趋势线返回一个值
16	FTES	返回 F - 检验的结果
17	GEOMEAN	返回几何平均数
18	GROWTH	根据通过给定数据的指数趋势线返回一个值
19	HARMEAN	返回调和平均值
20	INTERCEPT	返回通过给定数据的线性回归线的截距
21	KURT	返回一数据集的峰值
22	LARGE	返回一数据集中的第 k 个最大值
23	LINEST	返回通过给定数据的线性趋势线的参数
24	LOGEST	返回通过给定数据的指数趋势线的参数
25	LOGINV	返回对数正态分布的逆函数
26	LOGNORMDIST	返回对数正态分布的累积函数
27	MAX	返回参数清单中的最大值
28	MEDIAN	返回一组给定数字的中数(中位数)
29	MIN	返回参数清单中的最小值
30	MODE	返回一数据集中出现最多的值
31	NORMDIST	返回正态分布的累积函数
32	NORMINV	返回正态分布累积函数的逆函数
33	NORMSDIST	返回标准正态分布的累积函数
34	PERCENTRANK	返回数值区域中某数的百分比排位
35	POISSON	返回泊松分布
36	PROB	返回区域中落在上下限之间的数值所对应的概率
37	QUARTILE	返回数据集中的四分位点
38	SLOPE	返回通过给定数据的线性回归线的斜率
39	SMALL	返回数据集中的第 k 个最小值
40	STANDARDIZE	返回一个正态化的数值
41	STDEV	估计样本的标准偏差
42	STDEVP	计算样本总体的标准偏差
43	STEYX	返回用线性回归计算出的与每个 x 值对应的 y 预测值的标准误差
44	TREND	返回通过给定数据的线性趋势线上的数值
45	VAR	估计样本的方差
46	VARP	计算样本的总体方差
47	WEIBULL	返回韦伯分布
48	ZTEST	返回 z - 检验的双尾 P 值

2. Excel 工作表函数的使用 Excel 工作表函数是以“工作表”为操作平台来使用的,源数据要根据其属性按行、列分别组织在工作表内,计算结果将返回到选定的工作表单元格区域内或新的工作表中。

(二) Excel 数据分析工具

Microsoft Excel 提供了 19 种常用数据分析工具,称为“分析工具库”,用于对庞杂的样本实验数据进行统计学分析。只要为分析工具提供必要的源数据,系统即会按照指定的统计函数的计算流程对源数据(组)进行分析,显示出相应的分析结果,并同时生成图表。

在药学科研工作中,“数据分析工具”多用于“计量资料”的统计学分析。如定量描述一批样本数据的频数特征(集中趋势、离散趋势),由样本推断总体(如方差分析),多因素之间的关联性(如回归分析、相关分析),等等。

分析工具库是 Microsoft Excel 提供的一个自定义命令或自定义功能的加载项程序,要在 Excel 中使用它,需要先进行加载,其方法如下:

(1) 在“工具”菜单上,单击“加载宏”。

(2) “可用加载宏”框中,选中“分析工具库”旁边的复选框,然后单击“确定”。

提示:如果“分析工具库”未列出,请单击“浏览”进行查找。

(3) 如果出现一条消息,指出您的计算机上当前没有安装分析工具库,请单击“是”进行安装。

(4) 单击菜单栏上的“工具”。加载分析工具库后,“数据分析”命令会添加到“工具”菜单中。

(5) 数据分析工具可由图 1-1 所示的数据分析列表选择。

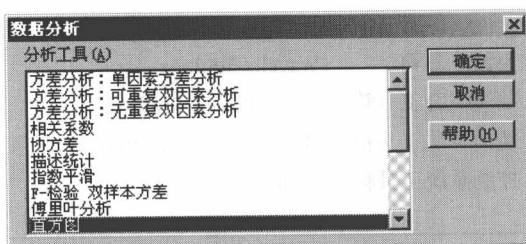


图 1-1 数据分析列表

二、MATLAB 工具箱简介

MATLAB 是矩阵实验室(matrix laboratory)的简称,是美国 MathWorks 公司出品的大型商业数学应用软件,是目前国际最流行的一种科学计算软件。MATLAB 支持多平台操作系统,是一种简单、易学、高效的编程语言,广泛用于算法开发、数据可视化、数据分析以及数值计算、工程与科学绘图、建模和仿真等。MATLAB 包含 600 多个常用算法内建函数,有众多面向具体应用的工具箱, MATLAB 的开放式结构使 MATLAB 产品很容易针对特定的需求进行扩充,提供与其他面向对象的高级语言进行混合编程的接口。

MATLAB 工具箱分为功能型工具箱和领域型工具箱两类。功能型工具箱主要用来扩充 MATLAB 的符号计算功能、图形建模仿真功能、文字处理功能以及与硬件实时交互功能。领域型工具箱的专业性很强,用户可以根据本专业的需求直接进行下载使用。药学中常用的 MATLAB 工具箱见表 1-2。

表 1-2 药学中常用的 MATLAB 工具箱

工具箱中文名称	工具箱英文名称	工具箱用途
MATLAB 主工具箱	MATLAB Main Toolbox	包含有多个函数库,用于数值计算、符号运算、绘图等
统计工具箱	Statistics Toolbox	概率分布、方差分析、假设检验、非参数检验、回归分析、判别分析、主成分分析、聚类分析、试验设计等
优化工具箱	Optimization Toolbox	线性规划和二次规划、求函数极值、多目标优化、非线性方程求解
模型预测控制工具箱	Model Predictive Control Toolbox	模型预测控制器设计、不同模型之间的转换、模型辨识等
神经网络工具箱	Neural Network Toolbox	感知器、BP 神经网络、径向基神经网络、自主组织映射神经网络等
动态仿真工具箱	Simulink Toolbox	系统可视化建模、动态模型仿真、数据分析等
图像处理工具箱	Image Processing Toolbox	图像恢复、图形绘制、图像几何变换、图像增强、二值分析等
信号处理工具箱	Signal Processing Toolbox	数字和模拟滤波器设计,仿真谱分析和估计 FFT、DCT 等变换,参数化模型
系统辨识工具箱	System Identification Toolbox	参数/非参数化模型辨识、模型验证、递推参数估计、各种模型类的建立和转换函数
小波工具箱	Wavelet Toolbox	小波分析与变换、小波包变换、信号处理等
曲线拟合工具箱	Curve Fitting Toolbox	实现多种类型的线性、非线性曲线拟合
遗传算法工具箱	Genetic Algorithm Toolbox	函数组合优化、机器学习等
控制系统工具箱	Control System Toolbox	连续系统设计和离散系统设计、状态空间和传递函数以及模型转换、时域响应、频域响应等

添加工具箱的方法如下:如果是 MATLAB 安装光盘上的工具箱,重新执行安装程序,启动 MATLAB 后, MATLAB 缺省目录是 bin,回到 MATLAB 目录,在 MATLAB\toolbox 下,得到各种工具箱。如果是单独下载的工具箱,则需要把新的工具箱解压到 toolbox 目录下,然后用 addpath 或者 pathtool 把该工具箱的路径添加到 MATLAB 的搜索路径中,最后用 which newtoolbox_command.m 来检验是否可以访问。如果能够显示新设置的路径,则表明该工具箱可以使用了。Math Works 公司提供的商品化 MATLAB 工具箱有 30 多种,toolbox 目录下的子目录数量是随安装情况而变的。

三、SQL Server 2005 数据挖掘工具简介

数据挖掘(data mining, DM)又称知识发现(knowledge discovery in database, KDD),是从大量的、不完全的、有噪声的、模糊的、随机的数据中,提取隐含在其中的、人们事先不知道的但又是潜在有用的信息和知识的过程,是统计学、数据库技术和人工智能技术的综合应用,其目标是在大量的数据中发现令人感兴趣的模式。目前,国外有许多研究机构、公司和学术组织从事数据挖掘工具的研制和开发,经过近十多年的发展,数据挖掘技术已趋于成熟,并已投入商业应用。世界上比较有影响的典型数据挖掘系统

有 SAS 公司的 Enterprise Miner、IBM 公司的 Intelligent Miner、SGL 公司的 SetMiner、SPSS 公司的 Clementine、Sybase 公司的 Warehouse Studio 等,但由于其中很多工具对操作者提出了诸如精通统计学、数据挖掘算法等较高的要求,限制了这些工具的推广使用。

SQL Server 2005 作为新一代的数据挖掘工具,因其具有优于传统数据挖掘应用程序的众多优势,改变了这一现状。图 1-2 所示的是数据挖掘的三个重要的方面,即数据报表、数据分析和数据挖掘。其中,数据报表模块解决了“它现在是什么样”的问题,数据分析模块解释了“它为什么是这样”的原因,而数据挖掘模块则为企业提供“它以后会是什么样”的预测。

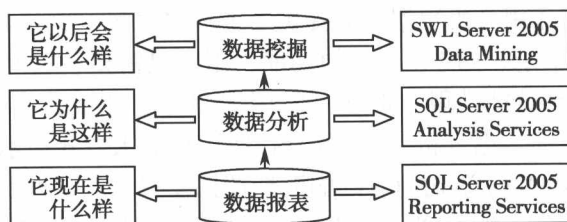


图 1-2 SQL Server 2005 数据挖掘模块

（一）SQL Server 2005 数据挖掘功能

SQL Server 2005 的数据挖掘功能,并非一个独立应用程序,而是开发智能应用程序的平台,包括集成服务 SQL Server Integration Services (SSIS) 和分析服务 Analysis Services 两大功能。

1. 集成服务 SSIS 结构及功能 数据预处理是数据挖掘之前的一项准备工作,需要花费较多的时间和精力。SQL Server 2005 中的集成服务器提供了数据抽取、转换、下载等功能,简化了数据预处理阶段的任务。SSIS 提供控制流(control flow)和数据流(data flow)两种工作流环境。一个 SSIS 项目中最基本的部署和执行单位被称为“包”,它是装载 SSIS 工作流的容器。一个 SSIS 项目可以包含多个包,一个包只能包含一个控制流,但可以含有一个或多个数据流。控制流也叫任务流,每个节点就是一个任务,这些任务定义“包”需要完成的任务,这些任务按照事先定义的顺序执行,可以根据执行结果的不同有不同的分支,比如下载数据、执行 SQL 语句、激活其他程序等。数据流是专门针对数据操作的工作流,一个数据流又称为一个管道,其中的每个节点被称为一次转换。数据流包含于任务流中,要使用数据流,必须先定义任务流中的数据流任务。

2. 分析服务 Analysis Services 结构及功能 SQL Server 2005 Analysis Services 是用于联机分析处理(OLAP)和数据挖掘的中层服务器。该系统包括一个服务器,可以构造用于分析的数据多维数据集,同时该系统还提供对多维数据集信息的快速客户端访问。Analysis Services 将数据仓库中的数据组织成包含预先计算聚合数据的多维数据集,以便为复杂的分析查询提供快速解答和从多维关系数据源中创建数据挖掘模型。

SQL Server 2005 Analysis Services 数据挖掘模块提供了多种有效的数据挖掘算法,主要包括决策树算法、回归分析算法、时间序列算法、聚类分析算法、关联分析算法、神经网络算法等。