

TEXTBOOK OF MEDICAL INFORMATICS

医学信息学教程

李 毅 赵乐平 主编



北京大学医学出版社

医学信息学教程

TEXTBOOK OF MEDICAL INFORMATICS

主 编 李 毅 赵乐平
编写人员 李 毅 北京大学医学部
赵乐平 美国华盛顿大学
于 娜 北京大学医学部
陶东华 美国圣路易斯大学
李 维 北京大学医学部

北京大学医学出版社

YIXUE XINXIXUE JIAOCHENG

图书在版编目 (CIP) 数据

医学信息学教程 / 李毅, 赵乐平主编. —北京: 北京大学医学出版社, 2016.1
ISBN 978-7-5659-1318-1

I. ①医… II. ①李… ②赵… III. ①医学信息—医学院校—教材 IV. ①R-0

中国版本图书馆CIP数据核字 (2015) 第322179号

医学信息学教程

主 编: 李 毅 赵乐平

出版发行: 北京大学医学出版社

地 址: (100191) 北京市海淀区学院路 38 号 北京大学医学部院内

电 话: 发行部 010-82802230; 图书邮购 010-82802495

网 址: <http://www.pumpress.com.cn>

E-mail: booksale@bjmu.edu.cn

印 刷: 北京瑞达方舟印务有限公司

经 销: 新华书店

责任编辑: 马联华 责任校对: 金彤文 责任印制: 李 啸

开 本: 889 mm × 1194mm 1/16 印张: 18.75 字数: 589千字

版 次: 2016年1月第1版 2016年1月第1次印刷

书 号: ISBN 978-7-5659-1318-1

定 价: 49.00 元

版权所有, 违者必究

(凡属质量问题请与本社发行部联系退换)

本书由

北京大学医学科学出版基金
资助出版

有了一本教科书，具有创造性的科学家就可以从教科书达不到的地方开始研究，从而可以高度集中到科学界所关心的最微妙、最深奥的自然现象中去。

新的理论总是同它在某一具体自然现象领域的应用一起产生的；离开应用，理论不会被人们接受。被接受以后，这些相关的应用就随着理论一起进入教科书，未来的工作者即由此而学到他们的专业。

随着学生选修大学一年级课程到通过博士论文答辩，被指派的问题变得越来越复杂，越来越缺乏可供遵循的先例。

Thomas S. Kuhn
The Structure of Scientific Revolutions
The University of Chicago Press
1962

医学信息学 (Medical Informatics) 是一门以促进人类健康为宗旨, 研究如何在计算机网络环境下, 不断提高数据、信息和知识的采集、处理、分析和利用效率, 为医药卫生领域的疾病防控、临床诊治、康复保健、科研教学和管理决策提供支持与服务交叉学科。

人类的活动与信息密不可分。现代医学对信息技术的依赖程度越来越高。随着大数据时代的到来, 医药卫生领域对医学信息学专业人才的需求越来越迫切。

本教程注重思想性、科学性、先进性、协同性、创新性和实用性, 采用多维的模块化设计方法构建了全新的医学信息学课程体系。从医学信息环境的四个重要支撑 (标准、安全、伦理和法律) 到医学信息管理流程中的四个主要环节 (信息采集、处理、分析和应用), 从适用于独立开展研究的医学决策分析、机器学习、数据挖掘、本体论、自然语言处理、认知心理和人机交互到必须通过团队协同合作才能完成的医学信息管理和各类医学信息系统建设, 从以政府为主导的医学信息化发展规划到以市场为主导的医学信息产业发展等内容, 本教程中均有所涉及。本教程的教学目的是培养学生的医学信息素养、独立思考和创新能力以及团队协作精神, 使学生熟练掌握医学信息学的基本知识、基本理论和基本技能, 全面了解国际医学信息学最新的研究成果和未来发展方向。本教程符合本科生教学评估的要求。

本教程共分三篇。第一篇绪论: 介绍医学信息学的相关概念, 医学信息学的历史、现状和未来发展趋势。第二篇理论、技术和方法: 介绍医学信息标准、安全、伦理和法律, 医学决策分析、机器学习和数据挖掘、本体论和本体工程、自然语言处理、认知心理和人机交互。第三篇研发和应用: 介绍基于医学信息管理业务流程的医学信息采集和存储、处理、搜索、分析和评价、传播和交流, 医学信息系统的模型、种类、研发和评价, 国内外医学信息化和信息产业的发展状况。

本教程由李毅、赵乐平、于娜、陶东华、李维共同完成。具体的分工是: 第一章基本概念: 李毅、李维、赵乐平; 第二章医学信息学的历史、现状和未来: 李毅、李维、赵乐平; 第三章医学信息标准: 李毅、李维; 第四章医学信息安全: 李毅、陶东华; 第五章医学信息伦理: 李毅、陶东华; 第六章医学信息法律: 李毅; 第七章医学决策分析: 李毅、陶东华; 第八章机器学习和数据挖掘: 李毅、于娜; 第九章本体论和本体工程: 李毅、于娜; 第十章自然语言处理: 李毅、于娜; 第十一章认知心理和人机交互: 李毅、陶东华; 第十二章医学信息管理: 李毅、于娜、陶东华; 第十三章医学信息系统: 李毅、于娜、陶东华; 第十四章医学信息化和信息产业: 李毅。统稿由李毅、赵乐平负责。

本教程主要作为医学院校本科生医学信息学课程教材, 也可以作为医学院校研究生、专升本学生以及医疗卫生信息领域从业人员的参考书或培训教材。

建议课时安排 72 学时:

第一章 基本概念 1.5 学时

第二章 医学信息学的历史、现状和未来 1.5 学时

第三章 医学信息标准 3 学时

第四章 医学信息安全 3 学时

第五章 医学信息伦理 3 学时

第六章 医学信息法律 3 学时

第七章 医学决策分析 8 学时

第八章 机器学习和数据挖掘 8 学时

第九章 本体论和本体工程 7 学时

第十章 自然语言处理 7 学时

第十一章 认知心理和人机交互 6 学时

第十二章 医学信息管理 8 学时

第十三章 医学信息系统 8 学时

第十四章 医学信息化和信息产业 5 学时

本教程参阅并借鉴了许多医学信息相关的图书、期刊、报纸和网络文献资料。在此，向这些文献资料的作者和出版发布机构表示衷心感谢。限于篇幅，本教程中仅列出了部分主要文献的出处。同时，衷心感谢北京大学医学科学出版基金的支持和北京大学医学出版社在本教程出版发行过程中给予的耐心、细致的帮助。

医学信息学发展速度较快，学科体系庞杂，内容涉及面广，各相关专业领域相互融合、深度交叉。在本教程编写过程中，作者虽然力求系统全面、重点突出、难易结合，但由于水平有限，内容上难免存在错漏和不妥之处，敬请各位读者批评指正。

李 毅 赵乐平

2015 年 5 月北京

第一篇 绪论	1	第五章 医学信息伦理	89
第一章 基本概念	3	第一节 伦理学、生命伦理学	
第一节 什么是医学信息学.....	3	信息伦理学.....	89
第二节 医学信息学相关基本概念.....	5	第二节 医学信息学中的伦理问题.....	94
文献导读.....	17	第三节 解决医学信息伦理问题的相关对策.....	96
思考与练习.....	17	文献导读.....	98
第二章 医学信息学的历史、现状和		思考与练习.....	99
未来	18	第六章 医学信息法律	100
第一节 医学信息学的历史.....	18	第一节 法律和信息安全.....	100
第二节 医学信息学的现状.....	28	第二节 隐私权.....	102
第三节 医学信息学的未来.....	36	第三节 证据和电子签名.....	104
文献导读.....	40	第四节 远程医疗的法律关系与法律责任.....	105
思考与练习.....	41	第五节 医学信息相关法律.....	106
第二篇 理论、技术和方法	43	文献导读.....	116
第三章 医学信息标准	45	思考与练习.....	117
第一节 标准化.....	45	第七章 医学决策分析	119
第二节 标准.....	47	第一节 决策分析.....	119
第三节 标准体系.....	52	第二节 临床决策分析.....	123
第四节 医学信息相关的国际标准化机构和		第三节 工具软件: TreeAge Pro.....	128
组织	55	文献导读.....	131
第五节 主要的医学信息标准.....	59	思考与练习.....	132
文献导读.....	72	第八章 机器学习和数据挖掘	133
思考与练习.....	73	第一节 机器学习.....	133
第四章 医学信息安全	74	第二节 数据挖掘.....	135
第一节 医学信息系统自身安全.....	74	第三节 经典算法和实例.....	136
第二节 医学信息系统应用给健康带来的		第四节 工具软件: Weka.....	139
安全隐患	85	文献导读.....	146
文献导读.....	87	思考与练习.....	147
思考与练习.....	87	第九章 本体论和本体工程	148
		第一节 本体论.....	148
		第二节 本体工程.....	153

第三节 医学本体	156	第四节 医学信息的采集和存储	206
第四节 工具软件: Protégé	158	第五节 医学信息的处理	208
文献导读	169	第六节 医学信息的搜索	215
思考与练习	170	第七节 医学信息的分析和评价	223
第十章 自然语言处理	171	第八节 医学信息的传播和交流	233
第一节 概述	171	文献导读	242
第二节 理论和方法	173	思考与练习	243
第三节 工具软件: GATE	181	第十三章 医学信息系统	244
文献导读	187	第一节 系统科学和信息系统概述	244
思考与练习	188	第二节 医学信息系统的层次模型和 主要功能	250
第十一章 认知心理和人机交互	189	第三节 医学信息系统的研发	256
第一节 认知心理学	189	第四节 医学信息系统评价	264
第二节 人机交互	192	文献导读	268
第三节 认知信息学在医学领域中的应用 ..	196	思考与练习	268
文献导读	198	第十四章 医学信息化和信息产业	270
思考与练习	199	第一节 信息化	270
第三篇 研发和应用	201	第二节 信息产业	272
第十二章 医学信息管理	203	第三节 医药卫生信息化	274
第一节 信息管理的概念、内容、流派和 理论	203	第四节 医药卫生信息产业	280
第二节 医学信息管理的概念、范围和 性质	204	文献导读	282
第三节 医学信息的活动周期和级次	205	思考与练习	283
		主要参考书目	285

图1	3T转化路径图	4	图27	Weka中利用Apriori对数据进行关联分析的结果	146
图2	数据元的元数据总体模型	9	图28	本体论词源	149
图3	SECI模型	15	图29	本体谱	152
图4	数据、信息、知识、智能在决策-行动过程中的相互关系	16	图30	语义网模型	153
图5	医学信息学总体框架	27	图31	美国国家生物医学本体中心 (NCBO) 门户网站BioPortal	157
图6	美国高德纳公司发布的技术成熟度曲线模型	36	图32	Protégé界面	159
图7	GB/T 13017-2008中企业标准体系表的层次结构图	54	图33	graphviz-2.34配置界面	159
图8	以产品为中心的企业技术标准序列结构图	54	图34	用Protégé浏览美国NCI叙词表	160
图9	信息系统安全保障评估的基本流程	81	图35	Protégé中的概念注释	160
图10	可能性表达的概率范围	125	图36	本体概念与概念之间关系的可视化呈现	161
图11	可能性与概率的对应关系	126	图37	owl文档格式	161
图12	膝关节置换术决策分析后的决策树	127	图38	本体类和类层次的构建	164
图13	TreeAge Pro界面	129	图39	类的公理构建过程	164
图14	梗死治疗决策树	129	图40	完成类的公理的构建	164
图15	经过Roll Back选择溶栓方案	130	图41	对象属性is_part_of的传递性属性	165
图16	建立投资决策树分支和结局	130	图42	对象属性eat的主体的类是Animal	165
图17	输入事件发生概率值的投资决策树	131	图43	对象属性eated的逆关系是eat	166
图18	Roll Back后的投资决策树	131	图44	通过关系建立类与类的关联 (树枝是树的一部分)	166
图19	Weka的主界面	140	图45	通过关系建立类与类的关联 (树叶是树枝的一部分)	167
图20	Weka的预处理界面	141	图46	通过关系建立类与类的联系 (素食动物吃植物)	167
图21	arff文件中数据的二维表形式	142	图47	通过关系建立类与类的关联 (长颈鹿吃树叶)	168
图22	Weka中的Apriori算法关联分析结果	143	图48	建立maineat对象属性和Carnivore类	168
图23	Weka中打开labor.arff文件的界面	143	图49	通过关系建立类与类的关联 (X主要吃素食动物)	169
图24	Weka分类实验结果 (1)	144			
图25	Weka分类实验结果 (2)	145			
图26	Weka中的可视化分类树	145			

图50	经推理得出X为肉食动物的结果.....	169	图65	面向问题解决的医学知识认识论框架..	197
图51	Thu-nlp分词结果.....	182	图66	发现与纠正医疗差错的七阶段模型.....	198
图52	NLPIR分词结果.....	182	图67	医学信息的活动周期.....	205
图53	不同自然语言处理平台分词效果的 对比.....	183	图68	Alexa在2010年1月全球网站 排名排在前10位.....	219
图54	GATE界面.....	183	图69	临床证据的等级(新九级).....	230
图55	语料调用.....	184	图70	信息传播的基本模式.....	234
图56	运行中文命名实体插件.....	184	图71	“大众传播场”模式.....	234
图57	命名实体识别.....	185	图72	信息交流的基本模式.....	236
图58	新建语料库添加语料.....	186	图73	广义的科学交流系统.....	237
图59	运行命名实体识别插件.....	186	图74	系统的主要部件及相互关系.....	247
图60	出院小结中手术术式的识别.....	187	图75	医学信息系统的层次模型.....	251
图61	ACT-R符号系统的内部结构.....	190	图76	FEA架构.....	258
图62	ACT-R 5.0的信息组织.....	191	图77	RUP开发阶段示意图.....	261
图63	诺曼的行动七阶段模型.....	194	图78	泛在网分层概念架构.....	272
图64	认知信息学与其他学科的关系.....	196			

表目录

表1	卫生信息数据集核心元数据内容.....	10	表12	针对四类不同风险等级的医学信息系统的不同的监控和规制建议方案.....	86
表2	使用计算机处理生物或医学数据和信息时的术语.....	19	表13	可能性描述用语的概率表达 (对52名全科医生的调查).....	126
表3	IMIA历年医学信息学年鉴主题.....	24	表14	临床证据的等级(老五级).....	229
表4	IMIA的专业工作组设置情况.....	28	表15	临床证据的等级(新五级).....	230
表5	医学信息学有关的大型会议.....	35	表16	临床证据的等级 (GRADE的证据质量).....	230
表6	未来医疗信息技术发展方向.....	37	表17	临床证据的等级 (GRADE的推荐强度等级).....	230
表7	ASTM E31委员会分会和相关的 医疗信息标准.....	58	表18	《美国联邦政府医疗信息化战略规划 (2015—2020年)》的目标.....	276
表8	HITSP确定的用例及其内容.....	59			
表9	ICD三位编码所对应的类别、名称、 范围、归类.....	60			
表10	专家词典工具.....	67			
表11	不同安全等级的技术类安全要求.....	82			

第一篇 绪 论

第一章

基本概念

学习目的

理解医学信息学的概念、性质；了解医学信息学的交叉学科特征；了解医学发展模式和信息学的基本层次；掌握数据元、数据集、元数据和大数据的概念；了解数据、信息、知识、智能的概念和关系。

学习重点

医学信息学概念的内涵；数据元、数据集、元数据和大数据的概念；数据、信息、知识、智能的相互关系。

第一节 什么是医学信息学

一、医学信息学的概念

医学信息学（Medical Informatics）是一门以促进人类健康为宗旨，研究如何在计算机网络环境下，不断提高数据、信息和知识的采集、处理、分析和利用效率，为医药卫生领域的疾病防控、临床诊治、康复保健、科研教学和管理决策提供支持与服务交叉学科。

二、对医学信息学概念的理解

（一）医学信息学的主体框架

医学信息学是由医学和信息学两大主体交叉融合而成的。

1. 什么是医学？

医学（Medicine）是旨在保护和加强人类健康、预防和治疗疾病的科学知识体系和实践活动。现代医学不仅有众多的技术目标，还有关于疾病、痛苦、残障、健康的社会共识，以及人道主义的崇高愿景。现代医学的发展模式正逐步从“生物-心理-社会”模式^[1]向现代整体医学模式转变^[2]。现代整体医学（Modern Holistic Medicine）模式提倡要研究自然、社会、精神、心理因素对人的影响，提出医学要以研究群体健康

^[1] 许树强,张铁山,韩鹏.健康医学模式与未来医院发展的新思路.中华医院管理杂志,2013,29(6):426-429.

^[2] 潘秋子,王敏,崔小希,等.医学模式的现状和未来发展模式探索.西昌学院学报·自然科学版,2015,29(1):80-83.

为目标, 实施预防疾病和促进健康的干预措施^[3]。

对“21世纪的医学”有多种不同的提法, 为人熟知的4P和TIDEST都力图反映新特点, 引领新方向。其中, 4P是指预测(Prediction)、预防(Prevention)、参与(Participation)、个性化(Personalization); TIDEST是指靶向(Targeted)、整合(Integrated)、基于数据的(Data-based)、基于证据的(Evidence-based)、系统医学(Systems Medicine)和转化医学(Translational Medicine)。

2003年, 美国国立卫生研究院(the National Institute Health, NIH)在其研究路线图(the NIH Roadmap)中提出将转化医学作为主要指导思想。转化医学的核心是: 将医学生物学基础研究成果迅速、有效地转化为可在临床实际应用的理论、技术、方法和药物。转化医学是一个把生物基础研究的最新成果快速有效地转化为临床医学技术的过程, 即从实验室到病床(Bench to Bedside), 再从病床到实验室(Bedside to Bench)的连续、双向的过程。

2008年, 丹尼斯·多尔蒂(Denise Dougherty)等在《美国医学会杂志》(JAMA)中提出了著名的3T转化路径(图1)^[4]。每个转化步骤都对之前的研究结果进行检验, 范围逐渐扩大, 从基础科学研究的发现, 到临床研究, 最后扩展为医疗服务体系的转型。

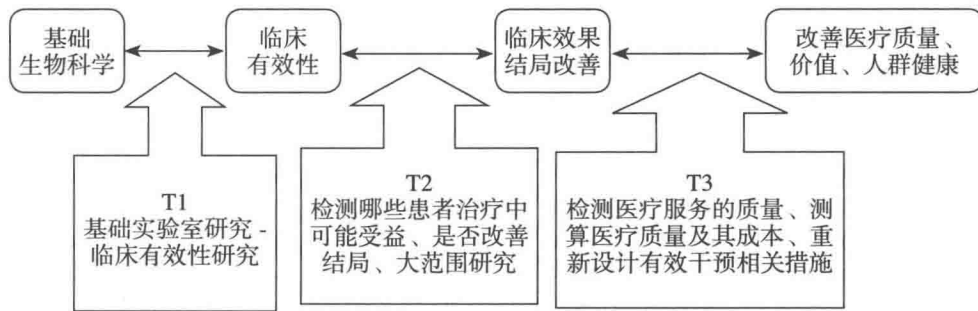


图1 3T转化路径图

2011年11月, 美国医学研究所(the Institutes of Medicine, IOM)发表了《向着精准医学迈进》(Towards Precision Medicine)的报告, 第一次对精准医学做了全面、详细的叙述, 其要点是: 在对疾病进行重新“分类”的基础上的“对症下药”, 创建生物医学的知识网络(knowledge network)和疾病的新的分类分型^[4]。

2015年1月, 美国总统奥巴马在发表国情咨文时提出启动一个新的“精准医学计划”(Precision Medicine Initiative)。其中, 包括四个要素: 精确(the right treatment)、准时(at the right time)、共享(give all of us access)、个体化(personalized information)^[5]。

精准医学对4P、靶向和基于数据的医学兼容并蓄, 与整合、循证、系统医学的提法也不冲突。

2. 什么是信息学?

信息学(Informatics)是以信息和信息能作为研究对象, 以信息的本质特征和信息的运动规律作为研究内容, 以信息方法为手段, 以扩展人类智能为主要研究目标的一门科学。

信息学讨论信息对象的存在方式和变化规律, 并由信息哲学提供本体论支持、方法论指导和价值论约束。

信息学可以分为两个基本的层次, 即理论信息学(Theoretical Informatics)和应用信息学(Applied Informatics)。理论信息学是对一切信息现象、对每一门信息学都适用的公共理论, 它通用于解决每一个具体领域的信息问题, 是整个信息科学的基础和核心。各门应用信息学各自面向特定的学科, 在某个具体的

^[3] 耿庆山, 整体医学对未来医学发展的影响, 现代医院, 2006, 6(9): 1-2.

^[4] Dougherty D, Conway PH. The “3T’s” road map to transform US health care: the “how” of high-quality care. JAMA, 2008, 299(19): 2319-2321.

^[5] 杨焕明, 奥巴马版“精准医学”的“精准”解读, 中国医药生物技术, 2015, (03): 1-9.

信息领域中起作用,而且主要是相关领域信息的综合^[6]。

3. 什么是交叉学科?

交叉学科(interdisciplinary)是指不同学科之间相互交叉、融合、渗透而出现的新兴学科。

交叉学科是横跨两个或多个一级学科门类的专业学科,或者是兼顾两个方向明显有区别的二级学科。

交叉学科可以是自然科学与人文社会科学之间的交叉而形成的新兴学科,也可以是自然科学和人文社会科学内部不同分支学科的交叉而形成的新兴学科,还可以是技术科学和人文社会科学内部不同分支学科的交叉而形成的新兴学科。

根据教育部2012年最新的学科划分,我国有哲学、经济学、法学、教育学、文学、历史学、理学、工学、农学、医学、管理学、艺术学12个学科门类。医学门类下设11个专业,它们分别是基础医学、临床医学、口腔医学、公共卫生与预防医学、中医学、中西医结合、药学、中药学、法医学、医学技术、护理学^[7]。医学信息学的任务就是要为这11个专业领域的科研、医疗、保健、教学和管理提供保障和服务。

交叉学科是以问题研究为导向的、通过融合不同学科的知识体系以产生新的知识体系而形成的学科。在研究环节中,来自不同学科背景的交叉学科研究人员共同协作、调适各自的研究途径,以取得对问题的更准确的切入。

与单学科、跨学科、超学科、多学科和无学科等研究模式相比,交叉学科研究人员在实际工作中最有可能通过融合不同学科间离散的知识体系而创造新的知识体系。

(二) 医学信息学的性质

医学信息学实质是一门采用信息学理论、技术和方法,研究与人类健康、疾病预防和治疗相关的数据、信息和知识的获取、处理、传递和利用,为改善医疗质量、保障人们健康生活而提供信息化产品和服务的交叉学科。

第二节 医学信息学相关基本概念

一、数据、数据元、数据集、元数据和大数据

(一) 数据

1. 数据的概念

数据(Data)是存储在某种介质上能够被识别的物理符号。

国际标准化组织(the International Organization for Standardization, ISO)对数据给出了更为严格的定义:“数据是对事实、概念或指令的一种特殊表达形式,这种特殊表达形式可以用人工的方式或用自动化的装置进行通信、翻译转换或进行加工处理。”

数据可以分为数值数据和非数值数据两种类型。数值数据分为有符号数据和无符号数据。非数值数据包括字符、图形、图像、语言以及逻辑数据等。

2. 医学数据的特征

医学领域存在着大量的数据,包括完整的人类遗传密码的信息,关于患者的病史、诊断、检验和治疗的临床信息,药品管理信息,医院管理信息等。这些医学数据具有以下特点:

(1) 多样性

由于大量的医学数据是从医学影像、实验数据以及医生与患者的交流中获得的,所以原始的医学数据具有多种形式。医学数据包括影像、信号、纯数值、文字等。医学数据的多样性是它区别于其他领域数据的最显著特征。

^[6] 李宗荣,张勇传,周建中,等.理论信息学:概念、原理与方法.医学信息,2004,17(12):773-785.

^[7] 教育部关于印发《普通高等学校本科专业目录(2012年)》《普通高等学校本科专业设置管理规定》等文件的通知.http://www.moe.edu.cn/business/htmlfiles/moe/s3882/201210/xxgk_143152.html.