

“十二五”国家重点图书出版规划项目

大数据技术与应用

丛书策划

上海大数据产业技术创新战略联盟（上海产业技术研究院）

上海市数据科学重点实验室（复旦大学）

丛书主编

朱扬勇 吴俊伟

Big Data

Technology and Application Series

于广军 杨佳泓

主编

医疗 大数据



上海科学技术出版社



大数据技术与应用

医疗大数据

于广军 杨佳泓
主编

上海科学技术出版社

上海世纪出版股份有限公司

图书在版编目(CIP)数据

医疗大数据/ 于广军,杨佳泓主编. —上海:上海科学技术出版社,2015.1(2015.1重印)
(大数据技术与应用)

ISBN 978-7-5478-2426-9

I. ①医... II. ①于... ②杨... III. ①医学—数据
IV. ①R

中国版本图书馆CIP数据核字(2014)第246119号

医疗大数据

于广军 杨佳泓 主编

上海世纪出版股份有限公司 出版
上海科学技术出版社

(上海钦州南路71号 邮政编码 200235)

上海世纪出版股份有限公司发行中心发行
200001 上海福建中路193号 www.ewen.co

苏州望电印刷有限公司印刷

开本 787×1092 1/16 印张 13.5

字数 300千字

2015年1月第1版 2015年1月第2次印刷

ISBN 978-7-5478-2426-9/TP·33

定价: 60.00元

本书如有缺页、错装或损坏等严重质量问题, 请向工厂联系调换

内容提要

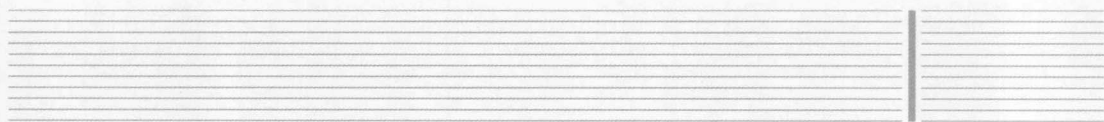
医学科学直接关乎全人类的生命健康,在大数据即将被纳入国家战略的今天,如何以数据创新探索未来的医学科学,如何在庞大的数据资源中快速获取信息、提升人类医疗集体经验,是亟待探讨的现实问题。

这本书汇集了中国医学科研前沿研究成果,首先从医疗信息化历史沿革入手,结合大数据时代下的国内外医疗现状,给出医疗大数据的基本概念和变革内容,包括定义、分类、描述和生命周期;其次,基于数据视角,研究了医疗大数据资源、医疗大数据安全和医疗大数据技术等问题;然后针对不同场景,介绍了临床、药学、中医、针灸和公共卫生等多种大数据应用实例,解析了区域医疗和健康物联两大主题中的大数据服务问题,据此对医疗大数据的未来应用进行全景式展望。

本书主要作为医学领域从事医疗大数据研究工作者的参考书,也可以为医疗领域的相关管理人士提供参考和启发。希望能对愿意参与到医疗领域大数据变革中来的读者有所启迪。

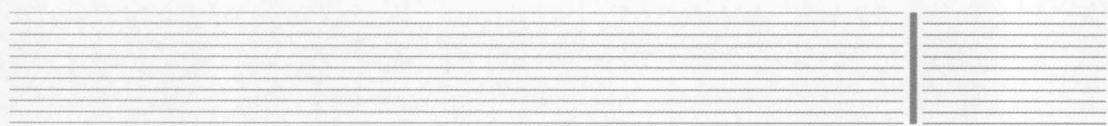
大数据技术与应用

学术顾问



中国工程院院士 邬江兴
中国科学院院士 梅 宏
中国科学院院士 金 力
教授, 博士生导师 温孚江
教授, 博士生导师 王晓阳
教授, 博士生导师 管海兵
教授, 博士生导师 顾君忠
教授, 博士生导师 乐嘉锦
研究员 史一兵

大数据技术与应用
编撰委员会



主任

朱扬勇 吴俊伟

委员

(以姓氏笔画为序)

于广军 朱扬勇 刘振宇 孙景乐 李光亚 李光耀 杨 丽
杨佳泓 吴俊伟 何 承 张鹏翥 陈 云 武 星 黄林鹏
童维勤 蔡立志

本书编委会

主 编

上海交通大学附属上海市儿童医院 于广军

上海申康医院发展中心 杨佳泓

编 委

(以姓氏笔画为序)

上海中医药大学附属岳阳中西医结合医院 史 晓

伊利诺伊大学芝加哥分校 吕 晖

上海交通大学医学院附属瑞金医院 朱立峰

上海交通大学附属上海市儿童医院 孙华君

上海申康医院发展中心 何 萍

复旦大学 余晓佳

复旦大学 应晓华

万达信息股份有限公司 张敬谊

国家卫生信息共享技术及应用工程技术研究中心 陈 诚

上海市数据科学重点实验室(复旦大学) 熊 贇

万达信息股份有限公司 潘 铮

前言

2014年1月,在人民大会堂举行的国家科技奖励大会上,上海医联工程项目获得了国家科技进步奖二等奖。“十年磨一剑”,医联工程项目自2005年酝酿、2006年启动、2007年获得国家科技部“十一五”支撑计划资助、2008年年底初步建成为国内“大规模,全覆盖”的区域医疗支撑系统、2009年进入深化应用阶段。医联工程所倡导的“互联互通、实用共享”的区域医疗信息化目标已写入国家医改方案,推动了全国以居民电子健康档案为基础的区域卫生信息系统的建设,促进了全程健康管理医疗模式的转变。

在上海,医联工程横向覆盖全市三级医院,纵向联通各区属医疗机构,已覆盖3900万患者,建成国内最大的患者诊疗档案库,拥有8.2亿条医嘱、1.8亿个病案、8100万份检验检查报告和107太字节医学影像数据;医联工程在服务医改、支撑资源整合、分布式影像网络会诊、三级医院诊疗信息社区调阅等方面,大幅度提升了区域卫生服务水平。

医联工程建成以来,对于患者,就医更方便,“一院办卡,跨院就医”“就医一站式付费”,共发放1300余万张医联卡,每月提供70万专家预约号源,人均节约就诊时间60min、排队时间45min;对于医生,服务看诊、提高医疗质量,支持每日5000人次实时诊疗档案调阅、1.25万人次重复医疗智能提醒,节约大量医疗费用;对于管理者,实现精细化管理,建成集医疗管理、绩效考核和统计分析为一体的整合平台,为医院管理提供决策支持。

同样,我们也是医疗领域大数据的倡导者,每天生产的1000万条数据、1亿余次调阅,是我们亟待数据创新的源动力。2012年,我们建议并获得了上海市科学技术委员会的首个医疗大数据项目的资助。

这本书的意图是成为大数据时代医疗领域的启发式读本,我们试图借助现有医学

各种应用场景,具体说明当前医疗领域数据资源分布状况、已使用或能使用的数据科技,以帮助理解未来的医学科学和医疗模式。

这段时间以来,编委的每位成员都付诸了努力,在此对大家辛勤的工作表示感谢。我们的工作需要特别感谢哈佛大学医学院的汤春蕾博士,她提供了许多观点与资料,其专著《数据产业》也为我们提供了借鉴。章节编写工作具体说明如下:于广军负责第1章第1小节和第16章;第1章的后面两小节由何萍和杨佳泓完成,何萍还撰写了第14章,杨佳泓还撰写了第5章;第2章和第4章由熊赅完成;第3章和第11章由余晓佳完成;第6章由陈诚完成;第7章由张敬谊完成;第8章由朱立峰完成;第9章由孙华君完成;第10章由史晓完成;第12章由吕晖完成;第13章由应晓华完成;第15章由潘铮完成。

最后,要衷心感谢的是阅读这本书的读者,感谢你们积极投身于医疗的大数据变革,愿我们更好地利用大数据,“让数据产生价值,让创新改变未来,让生活更加美好”。

于广军

目 录

第1章 医疗的大数据时代	1
• 1.1 国内外医疗现状	2
1.1.1 美国奥巴马医改的原因：医疗保险	2
1.1.2 医疗体制存在的问题：医患矛盾	4
• 1.2 医疗信息化的历史沿革	5
1.2.1 医疗卫生信息化概念	5
1.2.2 医疗机构信息化发展	6
1.2.3 区域医疗信息化建设	8
• 1.3 大数据时代的医疗变革	9
1.3.1 医疗大数据的研究内容	9
1.3.2 医疗服务的未来：你是自己最好的医生	11
第2章 医疗大数据基本概念	13
• 2.1 医疗大数据定义	14
• 2.2 医疗大数据分类	16
• 2.3 医疗大数据描述	17
2.3.1 元数据	17

2.3.2	本体	19
• 2.4	医疗大数据生命周期	24
	参考文献	25
第3章	医疗大数据资源	27
• 3.1	领域内数据资源	28
3.1.1	电子病历	28
3.1.2	医学影像	29
3.1.3	临床检验	30
3.1.4	医患行为	30
• 3.2	行业相关数据资源	31
3.2.1	医保政务	31
3.2.2	医学文献	32
3.2.3	制药行业	33
3.2.4	医药销售	33
• 3.3	学科相关数据资源	34
3.3.1	生命科学	34
3.3.2	人口学	35
3.3.3	环境科学	36
• 3.4	互联网数据资源	36
3.4.1	互联网	36
3.4.2	社交媒体	37
	参考文献	38
第4章	医疗大数据技术应用现状	41
• 4.1	已有的应用	42
4.1.1	大数据在医疗行业已有应用概述	42
4.1.2	大数据在智慧医疗中的应用案例	42

• 4.2 国内外技术现状	44
4.2.1 疾病预测与诊断	45
4.2.2 患者监控及预警	48
4.2.3 药物警戒	49
4.2.4 电子病历处理	51
• 4.3 面临的挑战	52
参考文献	54
第5章 医疗大数据安全	59
• 5.1 医疗大数据安全的界定	60
• 5.2 人的安全	60
5.2.1 医生隐私	61
5.2.2 患者隐私	61
5.2.3 现有隐私法律法规	62
5.2.4 医疗数据隐私探讨	63
• 5.3 数据安全	63
5.3.1 数据资源共享	63
5.3.2 数据资产界定	64
5.3.3 数据真假判断	65
参考文献	65
第6章 医疗大数据技术	67
• 6.1 医疗大数据存储与管理技术	69
6.1.1 分布式存储系统	69
6.1.2 主要的数据库存储方案	70
6.1.3 不同数据存储方案的选择	71
• 6.2 医疗大数据处理技术	73
6.2.1 基于并行计算的分布式数据处理技术	73
6.2.2 分布式流处理技术	74
6.2.3 内存计算处理技术	74

• 6.3	医疗大数据分析 with 挖掘技术	74
6.3.1	分析挖掘的工具集	74
6.3.2	分类挖掘算法	75
6.3.3	文本挖掘算法	75
<hr/>		
• 6.4	医疗大数据整合	76
6.4.1	相关术语标准	77
6.4.2	居民主索引技术	80
6.4.3	数据清洗技术	81
6.4.4	数据质量控制技术	82
	参考文献	84
<hr/>		
第7章	医疗大数据应用开发	85
<hr/>		
• 7.1	面向医疗大数据应用的逻辑参考架构	86
• 7.2	医疗大数据管理层	87
7.2.1	医疗大数据资源	87
7.2.2	医疗大数据融合处理	88
<hr/>		
• 7.3	医疗大数据整合层	89
• 7.4	医疗大数据处理层	92
• 7.5	医疗大数据应用支撑层	93
• 7.6	医疗大数据应用层	94
<hr/>		
第8章	临床大数据应用	95
<hr/>		
• 8.1	临床大数据综述	96
8.1.1	定义	96
8.1.2	特点	97
• 8.2	临床大数据应用案例	98
8.2.1	临床决策支持	98
8.2.2	自我治疗	99
8.2.3	临床患者行为管理	100
	参考文献	101

第9章 药学大数据应用	103
• 9.1 药学数据资源	104
9.1.1 药学数据源	104
9.1.2 药学数据特征	107
9.1.3 药学数据常用分析方法和应用方向	108
• 9.2 药学大数据应用实例	110
9.2.1 新药研发中的实践	110
9.2.2 药学文献挖掘	110
9.2.3 药品不良反应监测	113
9.2.4 临床药事大数据分析	114
参考文献	115
第10章 中医大数据应用	117
• 10.1 中医的大数据契机	118
10.1.1 中医在现代医学主导下的不足	118
10.1.2 中医诊疗体系的大数据探索	118
• 10.2 中医数据资源	119
10.2.1 中医典籍	119
10.2.2 诊断数据——四诊仪	120
10.2.3 医案数据	121
• 10.3 中医数据的典型应用	122
10.3.1 中医专家库系统	122
10.3.2 中医健康服务网络	122
参考文献	123
第11章 针灸大数据应用	125
• 11.1 针灸数据探索	126
11.1.1 针灸数据源	126

11.1.2	针灸数据特征	129
• 11.2	针灸领域的已有应用	130
11.2.1	针刺手法虚拟化	130
11.2.2	腧穴经络力学探索	131
11.2.3	临床治疗方案推荐	131
	参考文献	132
第12章	基因大数据应用	135
• 12.1	组学数据	136
12.1.1	基因序列数据	137
12.1.2	转录组学数据	138
12.1.3	蛋白质组学数据	139
12.1.4	其他组学数据	139
• 12.2	组学数据在医学中的应用案例	140
12.2.1	基因芯片与测序技术在遗传性疾病诊断中的应用	141
12.2.2	全基因组关联性分析	142
12.2.3	疾病网络模式发现	144
	参考文献	145
第13章	公共卫生大数据应用	149
• 13.1	公共卫生的内涵	150
• 13.2	公共卫生大数据	151
13.2.1	公共卫生数据源	151
13.2.2	公共卫生数据特征	153
• 13.3	公共卫生大数据应用	154
13.3.1	应用需求	154
13.3.2	传染病监测与管理	155
13.3.3	慢性病监测与管理	156
	参考文献	157

第14章 区域医疗中的大数据应用	159
• 14.1 区域医疗大数据	160
14.1.1 数据来源	161
14.1.2 主要特征	163
• 14.2 区域医疗大数据的典型应用	165
14.2.1 病种质量分析	166
14.2.2 影像数据分析	168
参考文献	171
第15章 健康物联中的大数据应用	173
• 15.1 健康物联数据解析	174
15.1.1 数据源和数据来源	175
15.1.2 常见数据类型及其处理	178
• 15.2 健康物联的典型应用	179
15.2.1 个人健康管理	179
15.2.2 第三方卫生服务	181
参考文献	184
第16章 医疗大数据的未来展望	185
• 16.1 社会化医学	186
16.1.1 政务	187
16.1.2 教育	188
16.1.3 商业	188
• 16.2 个性化医疗	189
参考文献	191
索引	193

