

# MI

## 基于大数据挖掘的 医疗健康公共服务

Public Health Services based on Big Data Mining

■ 陆泉 陈静 刘婷 著



WUHAN UNIVERSITY PRESS  
武汉大学出版社

教育部人文社会科学重点研究基地重大项目“大数据资源的挖掘与服务研究——面向医疗健康领域”的研究成果（批准号：17JJD870002）



# 基于大数据挖掘的 医疗健康公共服务

Public Health Services based on Big Data Mining

■ 陆泉 陈静 刘婷 著



WUHAN UNIVERSITY PRESS  
武汉大学出版社

## 图书在版编目(CIP)数据

基于大数据挖掘的医疗健康公共服务/陆泉,陈静,刘婷著.—武汉:  
武汉大学出版社,2020.6

ISBN 978-7-307-21513-9

I.基… II.①陆… ②陈… ③刘… III.医疗卫生服务—研究—中国 IV.R199.2

中国版本图书馆CIP数据核字(2020)第084167号

责任编辑:詹蜜

责任校对:汪欣怡

版式设计:韩闻锦

---

出版发行:武汉大学出版社 (430072 武昌 珞珈山)

(电子邮箱:cbs22@whu.edu.cn 网址:www.wdp.com.cn)

印刷:武汉市金港彩印有限公司

开本:720×1000 1/16 印张:38.25 字数:550千字 插页:3

版次:2020年6月第1版 2020年6月第1次印刷

ISBN 978-7-307-21513-9 定价:118.00元

---

版权所有,不得翻印;凡购我社的图书,如有质量问题,请与当地图书销售部门联系调换。

# 前 言

随着信息技术在卫生领域的快速发展和普遍应用，医疗健康领域已经成为数据增长最快的领域之一。从医院医疗数据、卫生服务平台医疗健康数据、疾病监测数据到自我量化数据、网络数据和生物数据，海量、多源、异构的医疗健康大数据正在经历爆炸式的增长。在这一背景下，挖掘医疗健康大数据资源，开展医疗健康公共服务，可以有效利用医疗健康大数据资源满足社会重大急需。国务院《促进大数据发展行动纲要》提出要实施公共服务大数据工程，确立了通过大数据技术促进医疗健康公共服务发展的思路。国务院总理李克强也指出，发展健康医疗大数据产业，收集、分析医学大数据资源，不仅可以为用户提供更好的医疗服务，推动医疗健康产业发展，对于攻克高难疾病等工作也有重要意义。

医疗健康大数据对推进与变革我国的公共健康服务具有极其重要的意义与价值。医疗健康大数据的分析利用可改变传统的医疗实践，并全面提高医疗护理的质量和效率。国际医学领域还将电子病历与其他医疗信息的融合利用视为医学领域的第二次信息技术革命。近年来，快速发展的大数据分析理论与技术已经在推动医疗研究和实践方面发挥关键作用，并带来了显著的临床效益和成本效益。然而，医疗健康大数据也给公共健康服务研究与管理应用带来了严峻挑战，导致管理决策部门与广大研究人员难以系统认识基于大数据挖掘的医疗健康公共服务基本理论、发展现状、关键技术及应用模式，不利于医疗健康大数据研究与应用的科学谋划、统筹推

进与协调发展。

本书是笔者主持的教育部人文社会科学重点研究基地重大项目“大数据资源的挖掘与服务研究——面向医疗健康领域”（批准号：17JJD870002）的主要成果，梳理了重大项目完成的系列研究，从情报学、计算机科学与数据科学交叉视角，突出了大数据与人工智能背景下的健康领域特色，首先系统揭示了基于大数据挖掘的医疗健康公共服务的需求与环境，进而重点创新了基于大数据挖掘的医疗健康公共服务的理论方法，最后探索了典型的基于大数据挖掘的医疗健康公共服务应用模式，建立了基于大数据挖掘的医疗健康公共服务研究与应用的理论方法体系。

在本书开篇的需求与环境部分，通过智慧医疗政策文本分析、健康行业上市公司调查、自闭症信息服务市场调查与用户需求发现以及网络健康社区中的用户信息需求挖掘，系统揭示了我国的医疗健康政策与市场现状；通过处理流程视角下的大数据技术分析、医疗健康大数据研究剖析、MIMIC-III 电子病历数据集及其研究分析、网络医学图像信息资源组织调查以及医疗健康研究跨学科知识图谱构建分析，系统揭示了医疗健康大数据的资源与技术基础及研究脉络。

在本书核心的理论方法部分，通过在线医疗咨询中的医患知识不对称规律挖掘、在线医疗社区患者择医行为影响因素揭示以及社交问答用户健康信息行为及其社会资本关系研究，系统探索了医疗健康公共服务用户行为的理论规律；通过电子病历大数据组织与知识抽取、医学文献内图像多标签分类以及医学课程体系知识挖掘与图谱构建，系统创新了医疗健康大数据组织与挖掘的基础理论方法；通过疾病演化分析、疾病指标预测、疾病危重度动态预测以及手术预后时间预测，系统构建了电子病历这一核心医疗健康大数据的分析与挖掘方法；通过基于模糊本体融合与推理的药物相互作用知识发现、基于网络评论的药物满意度预测、基于电子病历主题挖掘的疾病规则发现以及基于简单时间问题理论的临床路径融合，深入构建了医疗健康公共服务知识发现的理论方法体系。

在本书结尾部分，针对当前医疗健康数据与知识利用的“最

最后一公里”现象，通过颅脑健康知识库、多医学文档内容重构辅助阅读系统以及生物医学知识的可视化教学模式的设计、开发与评价，从满足不同主体对不同资源的不同需求角度，探索与验证了典型的医疗健康知识公共服务应用模式。

本书突破了现有的信息资源管理理论、医疗健康公共服务理论及大数据理论之间的壁垒，将信息资源特性及大数据技术研究深化到医疗健康领域的知识挖掘与需求满足层面，由此产出了一系列新的基础理论、科学方法与应用模式，其学术观点与理论体系具有基础性与原创性。本书具有重要的学术价值与应用参考价值，对大数据与医疗健康公共服务研究有重要指导价值，对我国的信息服务与医疗健康公共服务事业发展有重要推动作用，对促进信息科学、数据科学、公共健康与医学科学的融合发展有重要启发意义。

武汉大学信息资源研究中心主任李纲教授，武汉大学大数据研究院院长马费成教授、副院长夏义堃教授为项目研究提供了诸多支持和指导。中国社会科学院学部委员黄长著研究员、武汉大学人文社科资深教授马费成教授、华中师范大学副校长夏立新教授、中国人民大学信息资源管理学院周晓英教授、中国科学院大学管理学院霍国庆教授、南京大学信息管理学院苏新宁教授、武汉大学研究生院院长陈传夫教授、武汉大学信息管理学院院长方卿教授、副院长黄如花教授、副院长陆伟教授给项目提供了宝贵的专家指导意见。华中师范大学信息管理学院陈静教授、贵州省人民医院俞思伟博士、武汉大学信息管理学院范昊教授、邓胜利教授、孙永强教授等团队成员为项目研究做出了巨大贡献。南京大学人文社会科学大数据研究院裴雷副院长、南京大学江苏紫金传媒智库颜嘉麒研究员、福建东南肝胆健康大数据研究院骈文景副院长、中南财经政法大学信息与安全工程学院余传明教授、华中师范大学信息管理学院董庆星副教授等提供了非常具体的研究建议。借此对各位专家学者表示衷心感谢！

武汉大学的刘婷、丁凤一、江超、朱安琪、陈仕、郝志同、曹思曼等三十余位研究生与本科同学参与了项目研究及本书写作。项目研究还受到武汉大学图书情报实验教学示范中心与武汉大学超算

中心的支持。书中也多次引用国内外同行的研究成果。在此一并表示感谢。

基于大数据挖掘的医疗健康公共服务是一个有着深刻内涵与广阔前景的新兴交叉研究与应用领域，由于作者水平与研究条件有限，本书存在诸多不足，敬请读者批评指正。

陆 泉

2019年6月于珞珈山

# 目 录

第 1 章 医疗健康政策与市场	1
1.1 中国智慧医疗政策分析	1
1.1.1 智慧医疗	2
1.1.2 智慧医疗政策文本分析	6
1.1.3 政策分析结果	19
1.2 健康行业上市公司调查分析	21
1.2.1 健康行业与上市公司	22
1.2.2 市场环境分析	25
1.2.3 细分领域分析	28
1.2.4 上市公司运行分析	32
1.3 中国自闭症信息服务市场与用户需求	43
1.3.1 自闭症与健康信息学	43
1.3.2 自闭症信息服务市场	45
1.3.3 自闭症患者家长需求发现	52
1.3.4 自闭症信息服务不足与发展建议	66
1.4 网络健康社区中的用户信息需求挖掘	70
1.4.1 概述	71
1.4.2 用户信息需求挖掘方法	73
1.4.3 用户信息需求剖析	79
1.5 本章小结	87

<b>第 2 章 医疗健康大数据基础与研究脉络</b> .....	88
2.1 处理流程视角下的大数据技术 .....	88
2.1.1 大数据技术工具 .....	89
2.1.2 大数据技术的主要问题 .....	103
2.1.3 大数据技术发展趋势 .....	109
2.2 医疗健康大数据研究剖析 .....	113
2.2.1 医疗健康大数据 .....	113
2.2.2 医疗健康大数据研究主体 .....	114
2.2.3 医疗健康大数据研究对象 .....	122
2.2.4 医疗健康大数据研究的类型与主题 .....	127
2.2.5 研究全景 .....	138
2.3 MIMIC-III 电子病历数据集及其研究 .....	139
2.3.1 MIMIC-III 数据集剖析 .....	140
2.3.2 MIMIC 数据集相关研究分析 .....	149
2.3.3 结论与建议 .....	157
2.4 网络医学图像信息资源组织 .....	158
2.4.1 网络医学图像信息资源组织方式调查 .....	158
2.4.2 网络医学图像信息资源组织发展策略 .....	169
2.5 医疗健康研究跨学科知识图谱——以自闭症为例 .....	170
2.5.1 自闭症跨学科研究 .....	171
2.5.2 自闭症跨学科知识图谱 .....	174
2.5.3 自闭症跨学科研究特征 .....	193
2.6 本章小结 .....	196
<b>第 3 章 医疗健康公共服务用户行为</b> .....	197
3.1 在线医疗咨询中的医患知识不对称 .....	198
3.1.1 医患沟通与知识不对称 .....	198
3.1.2 医患知识不对称分析框架 .....	202
3.1.3 医患知识不对称规律 .....	210
3.2 在线医疗社区患者择医行为影响因素 .....	221
3.2.1 在线医疗社区与付费咨询业务 .....	222

3.2.2 患者择医行为影响因素研究模型 .....	223
3.2.3 在线医疗社区数据挖掘方法及步骤 .....	228
3.2.4 患者择医行为影响因素揭示 .....	236
3.3 社交问答用户健康信息行为及其社会资本 .....	239
3.3.1 概述 .....	239
3.3.2 健康信息行为与社会资本关系研究模型 .....	244
3.3.3 健康信息行为与社会资本关系模型验证 .....	248
3.4 本章小结 .....	259
<b>第4章 医疗健康大数据组织与挖掘</b> .....	<b>260</b>
4.1 电子病历大数据组织 .....	260
4.1.1 概述 .....	261
4.1.2 电子病历大数据组织模型 .....	264
4.1.3 电子病历大数据组织方法实现 .....	266
4.2 医学文献内图像多标签分类 .....	276
4.2.1 概述 .....	277
4.2.2 基于迁移学习的医学文献内图像多标签 分类模型构建 .....	281
4.2.3 模型验证 .....	287
4.3 医学课程体系知识挖掘与图谱构建 .....	299
4.3.1 概述 .....	299
4.3.2 医学课程——知识主题图谱模型构建 .....	302
4.3.3 模型实证 .....	310
4.4 本章小结 .....	315
<b>第5章 电子病历数据分析与挖掘</b> .....	<b>316</b>
5.1 疾病演化分析 .....	316
5.1.1 概述 .....	317
5.1.2 疾病演化分析过程与方法 .....	321
5.1.3 疾病演化分析评价 .....	328
5.2 疾病指标预测 .....	334

5.2.1	概述 .....	335
5.2.2	疾病指标预测系统分析与设计 .....	337
5.2.3	疾病指标预测系统实现与评价 .....	341
5.3	疾病危重度动态预测 .....	351
5.3.1	概述 .....	351
5.3.2	疾病危重度动态预测模型构建 .....	353
5.3.3	区分特征筛选 .....	356
5.3.4	预测效果分析 .....	360
5.4	手术预后时间预测 .....	362
5.4.1	概述 .....	363
5.4.2	手术预后时间预测系统设计 .....	365
5.4.3	手术预后时间预测系统实现与评价 .....	373
5.5	本章小结 .....	384
<b>第 6 章</b>	<b>医疗健康公共服务知识发现</b> .....	<b>385</b>
6.1	基于模糊本体融合与推理的药物相互作用知识发现 ..	385
6.1.1	模糊本体 .....	386
6.1.2	模糊本体构建与融合 .....	397
6.1.3	模糊本体知识推理 .....	402
6.1.4	药物相互作用知识发现 .....	408
6.2	基于网络评论的药物满意度预测 .....	416
6.2.1	概述 .....	417
6.2.2	领域迁移学习 .....	418
6.2.3	药物满意度预测 .....	419
6.2.4	跨领域药物满意度预测 .....	432
6.3	基于电子病历主题挖掘的疾病规则发现 .....	436
6.3.1	概述 .....	436
6.3.2	疾病规则发现与自动诊断模型 .....	442
6.3.3	模型评价 .....	456
6.4	基于简单时间问题理论的临床路径融合 .....	459
6.4.1	临床路径融合概述 .....	460

---

6.4.2	临床路径融合方法 .....	463
6.4.3	临床路径融合实现与评价 .....	471
6.5	本章小结 .....	494
<b>第7章</b>	<b>医疗健康知识公共服务</b> .....	<b>495</b>
7.1	颅脑健康知识库 .....	495
7.1.1	颅脑健康知识服务需求 .....	496
7.1.2	颅脑健康知识库系统设计 .....	497
7.1.3	颅脑健康知识库系统评价 .....	507
7.2	多医学文档内容重构辅助阅读系统 .....	513
7.2.1	多医学文档内容重构辅助阅读需求 .....	514
7.2.2	多医学文档内容重构辅助阅读系统设计 .....	519
7.2.3	系统开发 .....	525
7.2.4	系统评估 .....	528
7.3	生物医学知识的可视化教学模式 .....	536
7.3.1	生物医学知识教学不足分析 .....	537
7.3.2	可视化教学模式——以“细胞的癌变”为例 .....	540
7.3.3	可视化教学模式的优势 .....	549
7.3.4	问题与展望 .....	552
7.4	本章小结 .....	554
	<b>参考文献</b> .....	<b>555</b>

# 第 1 章 医疗健康政策与市场

近年来，随着我国经济的不断发展以及人们对健康的愈加关注，医疗健康行业持续快速发展。本章主要从政策分析、企业调查分析、信息需求挖掘及信息服务等方面，介绍基于大数据挖掘的医疗健康公共服务的政策背景与市场环境。首先对中国智慧医疗政策进行文本分析，为医疗健康政策制订提供指导；其次通过分析健康行业市场环境、细分领域以及上市公司的运行状况了解健康行业的发展趋势；然后通过分析自闭症信息服务网站了解自闭症信息服务的市场现状与自闭症家长的信息需求，为我国自闭症信息服务的发展提出建议；最后针对网络健康社区的用户需求进行挖掘，了解网络健康社区用户的需求，为有针对性的健康信息服务提供方向指引。



## 1.1 中国智慧医疗政策分析

我国作为人口大国，医疗资源严重匮乏，长期以来面对着医疗资源分配不均匀、配置不合理等日益突出的供求问题，“就医难”成为整个医疗体系都不容易解决的难题。但是随着新一代的信息技术和智慧城市的发展，智慧医疗也开始慢慢发展起来，走进人们视野，它具有高效、高质量和可负担的优点，不仅可以高效的提高医疗的质量，而且能有效地抑制不断上涨的医疗费用，是解决我国医

疗卫生领域的难题的有益举措。本章节从政策文本外部特征和政策工具这两个视角来具体细化分析中国智慧医疗政策的外部政策属性特征及政策工具运用的组合协调情况，通过政策工具编码的形式对智慧医疗政策的协调性进行分析和探讨，得出中国智慧医疗政策工具运用的可取之处以及存在的问题，为以后智慧医疗政策的制定提供一定的指导。本节主体内容选自项目研究成果“基于政策文献分析的中国智慧医疗研究”<sup>①</sup>。

### 1.1.1 智慧医疗

#### (1) 智慧医疗定义

目前对智慧医疗的定义和概念说法不一，关于它的讨论不绝于耳，但是整体上还在一个摸索的过程，每个概念都有它说的比较合理的一面。而目前有三种比较流行的说法普遍被学者们所接受：第一种说法是将智慧医疗整个庞大的系统看作是生态系统，系统的主要任务是利用高速传输的信息网络给医疗机构提供优质共享的信息化医疗服务；还有一种说法把智慧医疗看作是能建立长效协作关系的工作合作伙伴，并且能在预测与预防相关问题时候帮助相关使用人员科学合理的作出决策；最后一种比较可靠的说法是通过医疗信息化的建设，智慧医疗可以帮助每位用户建立一份专属于自己的电子健康信息档案，保存于医疗信息资源平台上，并按照严格的标准和规范对这些信息进行管理，推进智慧医疗的建设进程。笔者认为，智慧医疗和智慧城市的发展亦步亦趋，共同进步，而新兴技术的发展更是为智慧医疗的发展插上技术的翅膀，把民众的健康管理与医疗信息化的融合推入高阶的发展阶段。从广义上来讲，智慧医疗实际上是在概念这一维度上，将健康这一理念进行内涵和外延的拓展，将人们的身体健康状况和活力状况作为将要达到的目标，将

---

<sup>①</sup> 魏莹莹. 基于政策文献分析的中国智慧医疗研究 [D]. 武汉大学, 2019.

创新的产品、新型的商业模式及灵活的制度有机融合在一起，把社会上的医疗服务资源进行充分整合与调动，把全社会对于智慧社区医疗服务的热情给激发出来，实现方便快捷、持续有效的精准医疗健康服务。换一个狭义的角度来讲，智慧医疗是将新一代的信息技术及生物技术，综合运用云计算、物联网和大数据等技术，整合各个政府部门、卫生部门及智慧医院和社区卫生等相关医疗机构，创新出健康管理的新的医疗服务内容和形式，最后可以达到对用户的健康状况进行一个全息全程的动态监测和及时反馈的服务体系。

## (2) 智慧医疗特点

智慧城市是一个巨大的工程系统，而智慧医疗是其中的一个部分，它主要是通过医疗物联网、云计算、移动互联网、可移动穿戴设备、数据的整理融合及挖掘，将原来医疗的基础设施及设备与IT的高新技术进行整合，在此基础上进行智能的选择决策。它的优点在于可以突破时间和空间的局限性，呈现出智慧医疗服务的最好水平。智慧医疗是以居民的健康水平作为重心，通过建设统一的医疗卫生信息化平台，将原来杂乱无章的医疗信息整合成为有逻辑性和条理性的信息整体，并能从中挖掘出来有价值的信息，将信息转化成能满足用户需求的有价值数据。整体来看，智慧医疗主要有协作性、互联性、及时性、普及性、可靠性、预防性以及创新性等特点<sup>①</sup>。

## (3) 智慧医疗总框架

智慧医疗其实是我国医疗信息化不断深入和医学人性化发展，自然而然形成的一种新的医疗健康服务形态，同时也是智慧城市建设的重点之一，是信息便利民众的重要举措和手段。智慧城市的整体技术体系对智慧医疗建设的总体框架有着很深厚的影响，同时还

---

<sup>①</sup> 裘加林，田华，郑杰等．智慧医疗 [M]．北京：清华大学出版社，2015：33-34.

要注重与智慧城市的其他领域进行对接。用户层在整个系统中处于顶层的地位，用户主要有服务者、服务对象、管理决策者和别的相关用户。服务者主要是医生、护士等直接向病人提供医疗健康服务的人员。服务对象主要是病人还有不同健康程度的健康人。管理决策者主要是指医院、卫生部门、药品监督部门的管理负责人。其他相关用户主要是社区养老机构或者房地产企业提供的医疗健康服务机构。而智慧医疗的总体框架的核心部分主要是从城市医疗健康信息化的整体建设角度出发，由四个层次要素和三个支撑体系组成。四个层次要素从下往上依次是物联网感知层、网络通信层、数据及服务的支撑层、智慧医疗的应用层面。它们之间的关系主要是：处于横向的要素而下方对其他要素又互相依赖的关系，处于上方的四个横向支撑要素被纵向支撑体系所制约，具有相互约束的关系。

①物联网感知层。它的主要作用是为智慧医疗的健康卫生环境提供较为智能的感知能力，围绕着物联网技术，以各种生物纳米高新技术为辅助，利用各种输出设备，如智能芯片、智能生物传感器、无线射频识别技术、高清摄像头等技术装置，可以对医疗信息化建设所需要的信息进行识别搜集和处理，实现对城市范围内人、医院、救护车、健康流动站、家庭、居民社区、家庭及社区的养老院等地方进行数据监测，为以后的智慧医疗建设提供智能数据信息支持与服务。

②网络通信层。它主要是存在于智慧城市网络通信中。而智慧城市的网络通信层是以建立具有普及性、适应性、高效型、共享型、便捷性的网络通信基础为主要的目的，并且要为以后各个地方的信息建设、共享和互通互联提供相应的设施和设备基础，重点是做到三种网络的融合，即是移动互联网、电视广播网和电信网，最后将建成高速光纤网络和无线宽带网络的全覆盖，具有容量大、速度快、可靠性高等特点，其通信层的重点在于医疗卫生专网的建设之中。

③数据及服务支撑层。现在是信息的时代，是大数据的时代，数据的重要性不言而喻。在现在这个时代，除了物质财富和脑力财富外，排名第三重要的非数据信息莫属，是比较重要的战略性资

源，而如何做到数据的互享互通互融才是让我们的医疗产业变得智能与聪慧的重点，因此，医疗信息化的数据和服务支撑层在一定程度上是医疗建设的重中之重。该层主要采用 SOA、云计算、大数据等技术，对城市各级医疗健康数据信息资源进行聚合、共享及共用，同时与智慧城市的其他领域的的数据资源互联互通，为各类智慧医疗应用提供支撑。

④智慧医疗应用层。这一层主要依靠上文所提到的三个技术层面作为第四层的基础，然后开发和设计出各种各样智能的智慧医疗应用系统。智慧医疗应用层的主要目的是满足各个具体行业和领域的业务需求，采用智慧分析、辅助统计、预测、仿真等手段，对及时掌握的各类感知到的信息进行综合加工，建立各类医疗健康应用。整个系统的基础是电子病历及健康档案，系统建成有助平衡城乡之间在医疗资源方面分布不均的问题，有助于破除各个医疗机构之间的消息壁垒，促进我国医疗信息及资源的互通，还可以把各级医院的资源进行调动，进一步推动智慧卫生应急平台等一系列智慧应用的进步与发展。这些智慧应用为医疗健康服务对象、服务者和管理决策者等提供了便捷、联动、高效的整体信息化应用和服务，同时也带动城市医疗健康产业体系的发展。

⑤支撑体系。它其实具有指导性和导向性，对智慧医疗的核心部分的设计和框架性的部分进行设计，并且对智慧医疗的建设运行和管理等环节进行及时的指导，保障智慧医疗健康系统有序、高效运行，及时提供优质的医疗健康服务。支撑体系主要包括标准规范体系、安全保障体系、建设管理体系。标准规范体系将会从技术和服务等多角度对医疗信息化建设的总体进程和运营模式进行规范性的指示，充分发展建设过程中的柔性、开放性及拓展性，确保与智慧城市其他领域能否互联互通。安全保障体系广义上来讲就是为我们城市有效运行提供一个安全有保障的信息安全环境，该体系可以把城市的基础信息网络水平进行提升和改进、把整个系统的安全水平控制在一个安全可控的范围之内。以技术的角度分析，核心依然在于统一建成一个信息保障平台，实现统一入口、统一认