信息管理
新视野论丛

UGC MO SHI XIA DE

UGC模式下的在线健康信息分析

ZAI XIAN JIAN KANG XIN XI FEN XI

许 鑫 施亦龙 著

上海科学技术文献出版社
Shanghai Scientific and Technological Literature Press

UGC MO SHI XIA DE

UGC模式下的在线健康信息分析

ZAI XIAN JIAN KANG XIN XI FEN XI

许　鑫　施亦龙　著

上海科学技术文献出版社
Shanghai Scientific and Technological Literature Press

图书在版编目（CIP）数据

UGC模式下的在线健康信息分析 / 许鑫，施亦龙著．—上海：上海科学技术文献出版社，2019
ISBN 978-7-5439-7793-8

Ⅰ．① U…　Ⅱ．①许…②施…　Ⅲ．①互联网络—应用—疾病—诊疗—卫生服务—研究—中国　Ⅳ．① R197.1

中国版本图书馆 CIP 数据核字 (2018) 第 289881 号

责任编辑：徐　静
封面设计：袁　力

UGC 模式下的在线健康信息分析
UGC MOSHIXIA DE ZAIXIAN JIANKANG XINXI FENXI
许　鑫　施亦龙　著
出版发行：上海科学技术文献出版社
地　　址：上海市长乐路 746 号
邮政编码：200040
经　　销：全国新华书店
印　　刷：常熟市文化印刷有限公司
开　　本：787×1092　1/16
印　　张：17.25
字　　数：409 000
版　　次：2019 年 1 月第 1 版　2019 年 1 月第 1 次印刷
书　　号：ISBN 978-7-5439-7793-8
定　　价：98.00 元
http://www.sstlp.com

目　录

引 言

Web2.0 阶段是互联网蓬勃发展历程中不可被忽视的重要阶段。从 Web1.0 到 Web 2.0，互联网已经从单一的浏览和接受信息时代跨入了全民织网、全民互动的时代，与之伴随而来的是另一个术语 UGC(user generated content，用户生成内容)的出现。UGC 描述的是一种用户使用互联网的行为方式，这种行为方式的出现，代表了由用户生成的以文字、图片、音视频等为载体的表达观点、需求、情绪的内容已经成为 Web2.0 时代的主体，这个时代更加关注用户的需求和体验。它有别于传统的以权威为中心，权威生成并向外辐射为传播形式，它更提倡用户除了是信息的消费者，更是信息的创造者、传播者和贡献者。

在过去的几十年里，越来越多的消费者主动参与到医疗健康领域中，他们希望掌控自身以及家人的健康。互联网的快速增长和普及使消费者更加方便地接触到各类健康信息，而大量良莠不齐的网络信息对于健康信息消费者来说可谓喜忧参半。在过去，健康信息主要通过医生或其他医疗相关机构的专业人员传递给患者。这种模式下，医生或医疗机构成为患者及其家属们获得健康信息的最主要的甚至是唯一的渠道，虽然健康信息的质量和准确性很高，但健康信息的数量、患者或健康消费者获取信息的途径和参与度都受到了大幅度限制。在今天，健康自我管理得到了广泛认知，让消费者参与到医疗健康过程中也变得越来越重要，而 UGC 模式下的在线健康信息成为这种以消费者为中心的新模式的重要推动力。

本书汇总了华东师范大学 iLab 实验室网络健康信息研究小组成员近 3 年来的研究成果，通过对研究方案不断地讨论、修改以及实施推进，最终从不同角度对 UGC 平台的在线健康信息进行了较为全面的分析和探究。参与本书相关研究的包括徐一方、苏晓兰、施亦龙、金碧漪、姜雯、于霜等，本书最终由许鑫、施亦龙撰写统稿，姚占雷协助校对。

第一章

绪　论

健康信息学已经成为国外图书情报领域研究的重要分支，而对于网络信息资源开发和利用的深入研究一直以来也是学界关注的热点。本书聚焦网络 UGC 模式下的在线健康信息，通过对其进行多方位的研究分析，探讨健康信息的问答特征、信息需求、用户交流模式、领域知识图谱、信息质量评价等方面的议题。本章从研究背景、研究意义、研究思路、研究内容、研究方法以及本书结构等 6 个方面加以概述。

1.1　研究背景

近 10 年以来，随着 Web2.0 技术和应用研究的不断深入，各类 UGC 网站如雨后春笋般兴盛发展并受到广泛关注，目前全世界范围内最受欢迎的 Top10 网站中（来自 Alexa 网站流量监测[1]）就有 5 个是以用户生成内容为主导的网站，这 5 个分别是 Facebook，Youtobe，Wikipedia，Yahoo，QQ。

在中国，各类 UGC 网站及手机应用也是百花齐放，具体可见于 Kantar Media CIC 发布的 2016 年中国社会化媒体格局图（图 1-1）。百度（Baidu）、阿里巴巴（Alibaba）、腾讯（Tencent）和新浪（Sina）（总称为 BATS）旗下拥有 8 个社交及电商品牌，每一个都拥有数亿的活跃用户，这些平台无疑是中国社会化格局的核心力量。BATS 的壮大使得互联网具有快速传播（viral）、大信息量（informative）及高实用性（practical）的特点。其中，I 和 P 两点尤其重要且特征显著。目前许多充斥在社交网络上的信息、新闻及网民评论等相关内容通常较难在其他途径被发现。相比于全球市场，在这样的情况下，社交媒体在中国显得越来越重要。

如今，大众所生活的时代是一个城市化进程进一步深化的时代，亚健康的状态也似乎成了大众健康的常态。大众在健康管理、医疗保健以及饮食控制上开始逐渐重视，用户对于生活方方面面的关注都会体现在互联网中。对于健康保健的理念和行为习惯也随着互联网的渗透而改变。来自 PEW 研究中心的一份报告显示，80%的网络用户会在互联网上搜寻与健康主题相关的信息[2]。根据 2015 年 9 月中国科协发布的第九次中国公民科学素质调查结果显示，公民利用互联网及移动互联网获取科技信息的比例达到 53.4%，比 2010 年的

[1] Alexa. The top 500 sites on the web[EB/OL]. [2017-01-31]. http://www.alexa.com/topsites.

[2] FOX S. The Social Life of Health Information, 2011[EB/OL]. [2015-06-20]. http://www.pewinternet.org/2011/05/12/the-social-life-of-health-information-2011/.

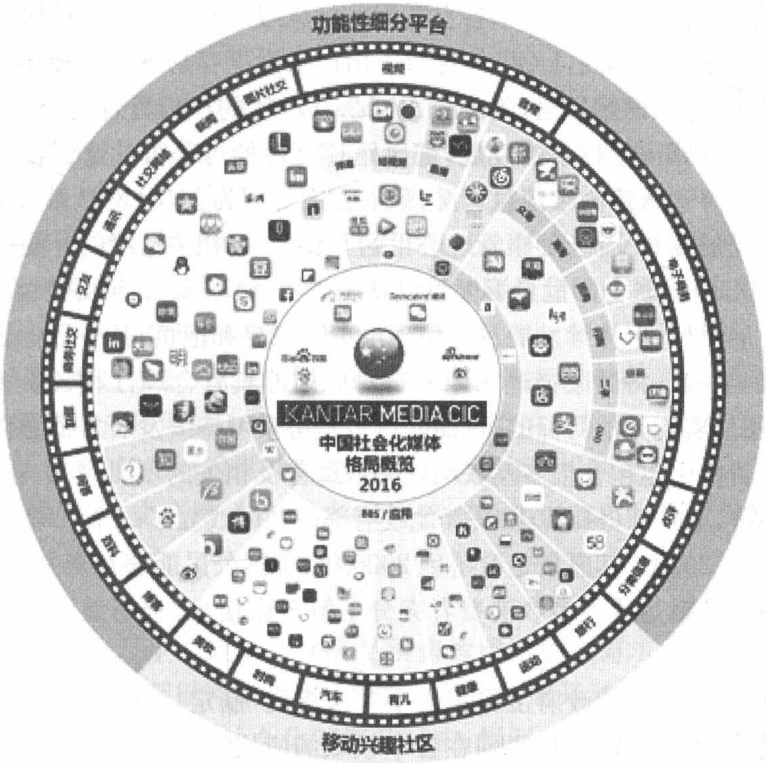

图 1-1 2016 年中国社会化媒体格局图

资料来源：CIC. 2015 年中国社会化媒体格[EB/OL]. [2015-12-14]. http://www.seeisee.com/index.php/2015/05/28/p8478.

26.6% 提高了一倍多，其中公民对医学新进展感兴趣的比例为 69.8%[1]。由此可见，互联网已经成为用户表达健康信息需求，进行健康信息搜寻的一个重要载体。

目前互联网上存在着数量庞大、种类繁多的健康网站，大致包括汇总各类疾病知识、健康信息、诊断防治管理信息的综合类健康信息网站和针对某种疾病提供各种解决方案的专业性的健康网站两种类型。这些网站虽然提供了广泛的信息资源，但是仍旧不能满足人们对于健康信息的需求，可能的原因主要有以下两点：(1) 即便网站的目标服务对象就是一般民众，网站在导航设计、信息资源内容组织时，通常过于专业，使得普通大众会遭遇到难以快速查找到符合需求的信息资源，以及查找到以后难以正确理解和遵循等尴尬情况。(2) 网站提供的信息资源虽然广泛，但缺乏针对性，每一个人的身体状况都有可能因自身身体条件、所处环境和心理状态等因素而显得千差万别，这些细小的差别难以在健康资源网站上体现出来，因此大众也就不能获得建设性的意见和建议[2]。

大众迫切需要更能发挥主观能动性的方式来表达需求，获得健康信息资源。UGC 的平台为大众提供了能够建立社区、互相交流、提供情感慰藉的空间。在 UGC 模式下，人们从互联网获取网络健康信息的渠道，从最早的发布健康信息的各类专业医疗网站，发展为论坛、博客、问答社区、社交媒体等。人们如今不仅可以单向地浏览专业医疗网站发布的健康信

[1] 中国科协发布第九次中国公民科学素质调查结果[EB/OL]. [2016-03-22]. http://education.news.cn/2015-09/19/c_128247007.htm.

[2] 沈光宝. Internet 上药学信息资源的开发利用及评价[J]. 情报科学，2002，20(9)：961-964.

息，还可以与其他网友进行互动，分享各自掌握的信息或经验。与专业的医疗网站发布的信息和公开发表的学术文献不同，UGC信息对于用户来说更容易理解和使用。医疗保健观念从过去的被动就医，甚至讳疾忌医，到如今Web2.0时代的积极管理，主动分享。这样的改变一方面是由于互联网信息技术的发展，健康医疗资源的可获得性增强，与过去相比较，用户可以更便捷地搜寻到医疗健康信息资源；另一方面是因为UGC网站最大限度地鼓励了人们对彼此健康状况的互相交流，健康管理心得的互相分享，相关情绪的互相抚慰。

2015年9月25日，联合国世界卫生组织发表声明，启动2030年可持续发展目标。相比较以前的千年发展目标，可持续发展目标呈现出更高眼界和抱负，其中目标之一是确保健康生活与促进全人类福祉，将健康描述为因其自身原因即具有合理性和必要性的目标。更为重要的是，健康还是实现其他目标的组成部分，也是考察可持续发展目标总体进展的可靠指标。这份最新的目标中也包含很多互联网健康信息普及相关的项目[1]。在美国，美国联邦机构工作组(FIW)、卫生部(HHS)以及其他政府部门共同成立了一个“2020健康人群”的项目(Healthy People 2020)，其愿景是希望到2020年，美国公民能够长寿并过上健康的生活，其中主要的一项任务就是让大多数消费者充分了解健康信息，利用健康信息来预防和治疗各种疾病[2]。

李克强总理在2015年的政府工作报告中首次提出“制定‘互联网+’行动计划，大力推动移动互联网、云计算、大数据、物联网在各个领域的应用”。具体到医疗健康领域，报告中提到2018年的目标是社会服务进一步便捷普惠，健康医疗等民生领域互联网应用更加丰富，公共服务更加多元，线上线下结合更加紧密。社会服务资源配置不断优化，公众享受到更加公平、高效、优质、便捷的服务[3]

在这样的大背景下，互联网健康医疗必定会蓬勃发展，大量互联网健康医疗的服务和工具也将会大量涌现，进而人们将有机会接触到越来越多的网络健康信息。

1.2 研究意义

国内外对消费者健康信息需求的关注度日益升温，消费者健康信息的相关研究在这样的时代背景下显得格外重要和富有价值。UGC模式下的在线健康信息能真实地反映健康信息的消费者对于健康信息的需求和认知，对这些数据进行集中采集、挖掘和分析对比有利于更好地把握健康信息需求和明晰消费者健康知识结构体系，进一步发现更加隐蔽的信息需求点和知识点，促进“互联网+”计划在健康医疗方面的建设。然而，目前网络UGC数据存在非结构化、信息源形式多样、健康信息主题不明朗、消费者健康知识体系不清晰等问题。因此，如何从这些多元的非结构化数据中抽取蕴含着的健康信息，如何通过同源纵向挖掘和多源横向比较的分析研究来较为全面地了解不同用户对健康信息的认知、需求以及其行为模式，进而揭示UGC模式下健康信息的特征特点和质量评价，这必然是一个非常有意义的

---

〔1〕确保健康生活与促进全人类福祉[EB/OL]. [2016-04-20]. http://www.who.int/mediacentre/news/statements/2015/healthy-lives/zh/.

〔2〕About-Healthy-People[EB/OL]. [2016-04-20]. http://www.healthypeople.gov/2020/About-Healthy-People.

〔3〕国务院关于积极推进“互联网+”行动的指导意见[EB/OL]. [2016-04-20]. http://cpc.people.com.cn/n/2015/0705/c64387-27255409.html.

研究课题。

1. 理论意义

从学科发展角度来看，健康信息学(health informatics)在国外已经是一个较为成熟的学科，相关研究数量比较多，范围也比较广，本书研究能丰富和发展消费者健康信息学领域的现有数据。非医学专业人士逐渐并大量成为医学信息服务的消费者，消费者健康信息学作为医学信息学的分支在这一过程中得以产生和发展。对消费者自身生成的健康信息加以深入的分析、知识发现及再组织，有利于促进消费者健康信息学的发展，有利于医学信息学、图书情报学、计算机科学等学科的各自发展和交叉融合。

从网络信息研究角度来看，以往的相关研究主体视角多为专家角度，而本书研究关注的是如今越来越多网络用户使用的 UGC 信息，评价视角为用户角度。全面直观地感知大众对于健康主题的理解程度，能够对提供针对性的健康信息有更好的把握。同时，更加深入理解不同社会化媒体上的信息交流模式的异同，对完善健康知识体系和提升健康信息服务都具有借鉴意义。

2. 现实意义

从用户角度，研究能更全面地掌握并满足消费者健康信息需求。通过对消费者健康信息需求的了解和掌握，能够帮助网络浏览者更加全面地了解某种疾病，满足自身潜在需求的同时及时做好疾病预防措施。在社会化问答社区中，提问者可以更加有针对性地提出自己在健康方面的需求，回答者也可以真正有的放矢地解答。

从平台建设角度，了解网络用户的信息需求和交互特征，关注网络健康信息质量评价中的指标因素，可以为平台建设、服务优化、质量控制等机制设计提供参考，同时给予 UGC 信息提供者和平台设计者更好的创作指导，从整体上提高信息质量，优化用户体验，从而增强消费者对平台的黏性。

从国家健康事业角度，政府可以利用在线健康信息提高公众的健康信息素养和健康水平，从而降低医疗费用开支。消费者通过浏览相关疾病健康信息，可以从各个方面了解某种疾病，可以和政府、社会组织一起更加有效地做好疾病的预防、诊断、治疗工作。政府可以利用健康信息和社会化平台，促进公众提高自身健康信息素养和健康水平，同时也能促进整个国家的医疗健康发展。

1.3 研究思路

随着大数据时代的到来，信息浪潮席卷了社会生活的方方面面，互联网的进一步发展使得人们获取各类信息变得更加便捷，这也使越来越多的人选择通过网络来满足其信息需求。

健康是高品质生活的基础，健康信息是与每个人都休戚相关的重要资源。互联网因其具有隐蔽性、便捷性等优点，正逐渐成为人们获取健康信息的高效渠道。人们从互联网获取网络健康信息的渠道从最早的发布健康信息的各类专业医疗健康网站，发展为论坛、博客、问答社区、社交媒体等。人们如今不仅可以单向浏览专业医疗健康网站发布的信息，还可以与其他网友进行互动，分享各自掌握的信息或经验。与专业的医疗健康网站发布的信息和公开发表的学术文献不同，UGC 信息对于用户来说更容易理解和使用，通过对 UGC 信息的深入分析可以获取更为直接和真实的用户需求、行为特征和交流模式等多方面的内容。然

而，UGC健康信息由于是用户自由创作的，缺乏相对统一的标准、监管和控制，导致其信息质量良莠不齐，因此UGC模式下的在线健康信息的质量问题尤为重要。

本书首先介绍了研究背景、意义、思路、内容以及后文实证分析章节中所涉及的不同研究方法，主要包括内容分析法、文本挖掘法、社会网络分析法、问卷调查法、机器学习法、知识图谱法、统计分析法等。由于在线UGC健康信息数据来源多样，本书主要选择了比较普遍的一些渠道，例如社会化问答社区、网络论坛、博客社区、社交论坛等，在数据分析上采用多种研究方法相结合的方式。

其次，梳理了消费者健康信息需求、在线健康信息以及网络健康信息质量评价等方面的研究，综合借鉴以往的相关研究，为下文的实证分析提供理论基础。

再次，在实证分析的章节部分，先对数据量较大的单个数据源采用不同方法进行多角度分析，例如内容特征、需求模型、交流模式等，希望能够挖掘出数据背后的深层信息。随后根据现有的单数据源研究进行延伸，将UGC数据特点和自动问答、知识服务、附加情感特征的信息评价体系应用分析相结合。最后通过对多源数据的综合比较，分析UGC数据的涉及主题、质量评价和内容特点的异同。最后，对研究得出的UGC模式下的在线健康信息特点进行概括，并对UGC模式下的健康相关工作的开展提出了建议。本书的研究思路总框架图见图1-2。

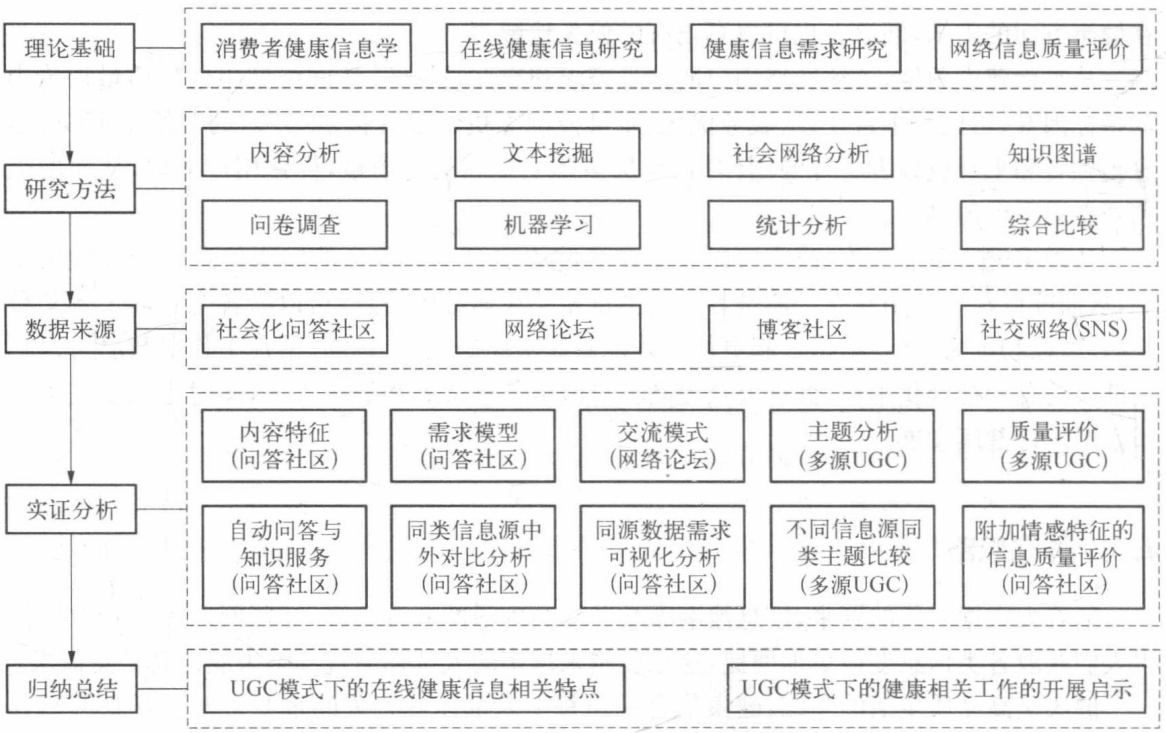

图1-2　研究思路总框架图

消费者的医疗保健观念从过去的被动就医，转向积极管理、主动分享，就医模式也从原来的医院挂号、临床就医，向着关注预防和日常保健、康复方向转变。学术界、医学界以及商业领域企业都在积极探索如何才能为普通消费者提供更加精准、便捷和全面有效的医疗保健服务。本书通过对网络UGC社区数据的分析和挖掘，围绕在线健康信息开展研

究，从理论探讨和实践应用两个方面加以探索，以期能对国内相关领域研究起到抛砖引玉的作用。

1.4 研究内容

由于网络 UGC 平台类型多样，涉及的数据内容和可能的研究视角也呈现多元化特点，本书对 UGC 模式下的在线健康信息的内容特征、需求模型、交流模式、主题分析和质量评价等方面分别进行了研究。

社会化问答社区是以交互性、社区化、共享性、知识性等特点为核心的网络时代的产物。本书第三章采用内容分析法对美国最大的社会化问答社区 Yahoo! Answers 中的健康栏目下相关的原始问答记录进行编码。研究从用户角度出发，分析消费者对于网络健康信息的需求特点，随后通过比较不同类型疾病提问者在健康信息需求上的差异，对如何更针对性地回答问题，满足疾病患者的信息需求提供借鉴，最后探究了提问与回答之间的交叉性关系。

目前对消费者健康信息需求的研究通常聚焦于某一种疾病且研究方法多以问卷调查、焦点小组访谈等定性研究方法为主，导致研究的样本数量有限，对某类疾病的研究结果很难适用于其他疾病。本书第四章同样是基于社会化问答社区数据，研究结合归纳法的思想，利用文本挖掘分析方法，通过对社会化问答社区中的健康信息分析构建了消费者健康信息需求模型，并对模型进行了优化，讨论了模型的具体应用场景。

伴随着互联网的深入发展，专业性网络论坛已逐渐成为学术科研工作者重要的非正式交流场所。本书第五章为了增进对专业性网络论坛中用户交流的特点及模式的了解，采用社会网络分析法围绕交流主体(用户)、交流客体(主题)和交流方式三个方面分别探讨了社会化网络论坛中 5 种用户交流模式，即一对一、单中心、多中心、发布式和跨领域交流模式，并对用户交流特点和影响用户交流的主要因素进行了总结。

确认消费者对健康信息关注的主题是把握消费者健康信息需求，进而提供精准医疗保健服务的先决条件。本书第六章结合定性和定量方法，对来自为大众普遍使用的不同社交媒体上的多种疾病数据进行采集分析，提炼健康主题，提取特征词汇及特征词间关系，最终构建消费者健康知识图谱，深入分析和讨论了知识图谱对于消费者健康信息素养提升和健康信息系统设计的启示，并探索了知识图谱的具体应用场景。

UGC 在线健康信息的质量是保证实证分析能够顺利进行以及结论有效性的基础。本书第七章综合借鉴以往的研究，并结合健康信息的特点初步构建了 UGC 在线健康信息质量评价指标体系。在研究中首先使用问卷调查法，根据用户的评分进行模型的修正和提炼，随后通过层次分析法赋予权重，形成最终的 UGC 在线健康信息质量评价指标体系，最后利用研究形成的评价指标体系对不同健康领域、不同来源的 UGC 在线健康信息进行了实证应用，探讨了不同健康领域、不同来源的 UGC 健康信息的质量差异。

除了对现有数据进行分析外，本书第八章还根据分析进行了延伸的应用研究。

(1) 利用国外发展成熟的自闭症问答社区的数据构建了一个自闭症问答知识库，并通过其提供的知识服务来满足患者的多重信息需求。

(2) 通过对中美两大社会化问答社区中随机抽取的健康问答数据进行对比分析，对中美用户在疾病的认识、应对、发展方面存在的异同作出概括，了解到中美社会对于精神性疾

病患者在教育、医疗、就业等体系的建设水平存在一定的差异。

(3) 对健康信息提问所使用的词汇及其模式加以揭示，分析了糖尿病消费者在表达健康信息需求时的语用习惯和词汇特征，并通过可视化展现增强消费者获取健康信息资源的可达性。

(4) 探究不同类型网络社区中健康主题特征分布情况，以及在此基础上如何对健康网站上的信息进行组织以及站内检索的优化。

(5) 通过考察加入情感特征后其对自动化预测问答社区信息质量的影响，发现UGC在线健康信息与传统的在线健康信息相比包含更多的情感因素。

本书的研究内容涉及多方面：从数据来源来看，分别选择了社会化问答社区、网络论坛、博客社区和社交网络等UGC平台；从数据类型来看，既有原始问答数据和帖子记录，也有基于这些原始数据的编码数据，还包括用户问卷调研数据；从关注范围来看，研究基于国外和国内具有代表性的UGC平台数据展开，同时也对国内外同类型平台数据作了对比分析；从用户类型来看，这些数据的产生既来源于专业医学人员，也有普通非专业人士产生的。本书希望通过多视角多维度的分析，能够较全面地展现网络UGC平台在线健康信息的方方面面，探讨如何有利于更好地引导和服务网络环境下的健康知识传播。

1.5 研究方法

本书的研究中涉及大量不同信息源的在线健康信息数据，综合运用了多种研究方法，主要包括以下几种。

(1) 内容分析法　内容分析法(content analysis)是一种对于传播内容进行客观、系统和定量的描述的研究方法，其实质是对传播内容所含信息量及其变化的分析，即由表征的有意义的词句推断出准确意义的过程。内容分析法通过把文本转化为数据化形式，最后形成用做统计分析的评判记录，在具体分析时还可以结合原始文本记录加以举例或者解读。如果在内容分析法转化编码过程中涉及两个或两个以上的研究者，还需要进行信度分析，以保证他们能够按照相同的分析维度，对同一材料进行评判的结果具有一致性，这也是保证内容分析结果可靠性、客观性的重要指标。

(2) 文本挖掘法　文本挖掘(text mining)是以文本为研究对象，涉及统计分析、数据挖掘、NLP(自然语言处理)、数据库等多种技术方法的跨学科的知识和技术。文本挖掘是指为了发现知识，从文本数据中抽取隐含的、前所未知的、具有价值的模式的过程[1]。文本挖掘在预处理过程中的重点是自然语言特征的识别与抽取，这一操作将非结构化数据转化并存储为更为显著的结构化形式。文本挖掘是一个抽取文本信息、分析文本数据，从而发现文本知识的过程，一般流程主要分为4步，即确认问题及设计方案、文本数据的获取、文本特征词提取、文本特征词分析。

(3) 社会网络分析法　“社会网络”是指社会行动者(actor)及其间的关系的集合，即一

[1] TAN A H. Text Mining: The State of the Art and the Challenges[C]//In: Proc of PAKDD Workshop on Knowledge Discovery from Advanced Databases. Beijing, China: 1999: 65-70.

个社会网络是由多个点(社会行动者)和各点之间的连线(行动者之间的关系)组成的集合[1]。社会网络中各点之间的连线可以分为有向的和无向的。社会网络分析(social network analysis, SNA)就是对这一社会网络中行为者之间的关系进行量化的分析,主要包括密度分析、中心性分析、凝聚子群分析等。其中网络密度指群体成员间彼此的联系程度[2],还反映了网络的连通性[3]。中心性包括中心度和中心势两个概念,中心度刻画单个行动者在网络中所处的核心位置,而中心势刻画的则是一个网络所具有的中心趋势[4],中心性指标主要包括点度中心性、中间中心性和接近中心性。

(4) 知识图谱法 知识图谱(mapping knowledge domain)是一种以可视化的方式展示信息中包含的知识要点、核心结构、整体知识架构的技术,在图书情报界也被称为知识域可视化或知识领域映射地图。它是显示知识发展进程与结构关系的一系列各种不同的图形,用可视化技术描述知识资源及其载体,挖掘、分析、构建、绘制和显示知识及它们之间的相互联系。利用可视化的图谱可以形象地把复杂的知识领域通过数据挖掘、信息处理、知识计量和图形绘制而显示出来,揭示知识领域的动态发展规律,为学科研究提供切实的、有价值的参考。常用的可视化软件有 Ucinet、Pajek 和 Citespace 等。

(5) 问卷调查法 问卷调查法(questionnaires method)即用调查表格的形式间接搜集研究材料和研究数据的一种调查方法[5]。通过向被调查者线上或线下发出提前设计好的调查表格,请他们填写对有关问题的意见、对某个事物的标准进行评分等来获得研究者所需的信息和数据的一种研究方法。本书在研究中设计了 UGC 在线健康信息质量评价用户调查问卷,使用问卷星进行了线上投放(包括问卷星的推荐服务和朋友之间扩散两种方式),获得了用户对 UGC 在线健康信息质量评价指标重要性的评分。

(6) 机器学习法 机器学习(machine learning)是指采用某些算法指导计算机利用已知数据得出适当的模型,并利用此模型对新的情境给出判断的过程。在此过程中计算机不断模拟和学习人类的行为,以获取新的知识或技能,建立数据库,重新组织已有的知识结构使之不断改善自身的性能。在整个机器学习的过程中,样本环境向系统的学习部分提供某些信息,学习部分利用这些信息修改知识库,以增进系统执行部分完成任务的效能,执行部分根据知识库完成任务,同时把获得的信息反馈给学习部分。影响学习系统设计的最重要的因素是样本环境向系统提供的信息,或者更具体地说是被模拟学的信息质量。

(7) 统计分析法 统计分析(statistical analysis)是指运用数学方式,建立数学模型,对通过调查获取的各种数据及资料的规模、速度、范围、程度等数量关系进行数理统计和分析,形成定量的结论,认识和揭示事物间的相互关系、变化规律和发展趋势,借以达到对事物的正确解释和预测的一种研究方法。它是继统计设计、统计调查、统计整理之后的一项十分重要的工作,是在前几个阶段工作的基础上通过分析从而达到对研究对象更为深刻的认识。常被运用对数据进行基本的统计分析和信度效度检验。

---

〔1〕朱庆华,李亮. 社会网络分析法及其在情报学中的应用[J]. 情报理论与实践,2008,31(2):179-183.

〔2〕李培林,覃方明. 社会学:理论与经验[M]. 北京:社会科学文献出版社,2005:102-115.

〔3〕邱均平,李佳靓. 基于社会网络分析的作者合作网络对比研究——以《情报学报》、《JASIST》和《光子学报》为例[J]. 情报杂志,2009,29(11):1-5.

〔4〕刘军. 社会网络分析导论[M]. 北京:社会科学文献出版社,2004:16-131.

〔5〕陶永明. 问卷调查法应用中的注意事项[J]. 中国城市经济,2011,9(20):305-306.

1.6 本书结构

本书共分为9章，对UGC模式下的在线健康信息展开多角度的分析，主要结构如下。

第一章，阐述了研究的背景和意义，明确了本书的研究思路，在此基础上细分了研究内容和介绍了研究方法，最后是本书结构。

第二章，介绍了消费者健康信息学、在线健康信息研究、健康信息需求和网络信息质量评价的相关理论，主要包括消费者健康信息、在线健康信息学特点、健康信息需求特征和网络信息质量评价指标等内容，为后文开展研究奠定了理论基础。

第三章到第八章是实证分析部分。

第三章运用内容分析法和统计分析法，针对国外社会化问答社区中的问答数据分别从提问、回答以及两者相关性三方面做了信息特征分析，讨论了社会化问答社区中不同提问需求和最佳答案的特点，以及相互之间的影响关系内容特征。

第四章，针对第三章研究采用的同源数据进行了特征词提取，利用文本挖掘法进行分析后构建了消费者健康信息需求模型，随后针对模型存在的不足予以优化，最后介绍了该模型的三类具体应用场景。

第五章，通过社会网络分析法对国内主要面向专业人士的网络论坛中的用户交流特点及模式进行了探讨，构建了5种用户主要交流模式，并对用户交流的特点和影响用户交流的主要因素进行了总结与归纳。

第六章，对4组来自不同网络UGC平台的数据进行健康知识图谱分析和探讨，分别对展现不同主题间联系的知识图谱、来自不同社交媒体网站图谱和消费者健康信息素养等方面作了探讨，并在此基础上延伸了对于提升消费者健康信息素养以及健康信息系统设计的启示。

第七章，通过问卷调查法先对原有的UGC在线健康信息质量评价指标进行筛选，随后使用指标体系对不同UGC来源、不同健康领域的信息进行了实证应用，并对评价结果进行了探讨。

第八章，在前几章的基础上进行了延伸性的应用研究，采用不同视角对国内外不同UGC数据源进行综合比较分析，包括构建问答知识库、着重分析情感因素对于问答的影响、健康信息需求可视化、同一疾病在不同类型网络社区中的主题特征以及中美在线健康信息用户的认知差异等内容。

第九章，对本书所做的研究工作进行总结，并提出在UGC模式下开展健康相关工作的建议。

参考文献

[1] Alexa. The top 500 sites on the web[EB/OL]. [2017-01-31]. http://www.alexa.com/topsites.

[2] CIC. 2015年中国社会化媒体格[EB/OL]. [2015-12-14]. http://www.seeisee.com/index.php/2015/05/28/p8478.

[3] FOX S. The Social Life of Health Information, 2011[EB/OL]. [2015-06-20]. http://www.pewinternet.org/2011/05/12/the-social-life-of-health-information-2011/.

[4] 中国科协发布第九次中国公民科学素质调查结果[EB/OL]. [2016-03-22]. http://education.news.cn/2015-09/19/c_128247007.htm.
[5] 沈光宝. Internet 上药学信息资源的开发利用及评价[J]. 情报科学,2002,20(9): 961-964.
[6] 确保健康生活与促进全人类福祉[EB/OL]. [2016-04-20]. http://www.who.int/mediacentre/news/statements/2015/healthy-lives/zh/.
[7] About-Healthy-People[EB/OL]. [2016-04-20]. http://www.healthypeople.gov/2020/About-Healthy-People.
[8] 国务院关于积极推进"互联网+"行动的指导意见[EB/OL]. [2016-04-20]. http://cpc.people.com.cn/n/2015/0705/c64387-27255409.html.
[9] TAN A H. Text Mining: The State of the Art and the Challenges[C]//In: Proc of PAKDD Workshop on Knowledge Discovery from Advanced Databases. Beijing, China, 1999: 65-70.
[10] 朱庆华,李亮. 社会网络分析法及其在情报学中的应用[J]. 情报理论与实践,2008(2): 179-183.
[11] 李培林,覃方明. 社会学. 理论与经验[M]. 北京: 社会科学文献出版社,2005.
[12] 邱均平,李佳靓. 基于社会网络分析的作者合作网络对比研究——以《情报学报》、《JASIST》和《光子学报》为例[J]. 情报杂志,2009,29(11): 1-5.
[13] 刘军. 社会网络分析导论[M]. 北京: 社会科学文献出版社,2004: 16-131.
[14] 陶永明. 问卷调查法应用中的注意事项[J]. 中国城市经济,2011(20): 305-306.