

## 摘要

### 虚拟健康社区用户生成内容主题发现及个性化推荐研究

近年来,随着互联网技术的蓬勃发展,网络资源在生活、工作和学习中发挥重要作用,越来越多的用户为高效获取信息和解决问题选择在网络媒体上互动。Web2.0 的到来,改变了人们利用信息的模式,用户不仅仅是被动接受信息内容的“受众”,而且可以主动参与到信息生产和传播的过程中,所以,虚拟社区的用户既是信息文化内容的消费者,也是生产者和提供者,虚拟社区因其交流快捷、可编辑性强、使用简单方便受到广大用户的持续关注,为很多有医疗卫生需求的人群提供了帮助。我国经济高速发展,人民生活节奏不断加快,虚拟健康社区因其能够有效帮助用户解决日常生活中健康类需求,成为不少人了解、关注、咨询健康问题的平台,“互联网+医疗”模式将线下资源搬移到线上,患者和医生之间的信息交互在线上平台中完成,大大节省了用户时间,节约线下资源。信息经济发展迅猛,互联网中信息呈爆炸式增长,虚拟健康社区涵盖了大量有价值的信息和知识,为医疗健康领域的知识发现提供了新的研究对象,大数据技术的成熟发展为研究提供了可靠的工具,据此可实现网络环境下虚拟健康社区用户生成内容的主题发现及个性化推荐研究。

如何对虚拟健康社区用户生成内容进行知识管理和组织,对用户需求进行挖掘,创新虚拟健康社区服务新模式,为虚拟健康社区用户提供高质量服务,成为虚拟健康社区研究面临的新问题。鉴于此,本文将主题发现理论和方法引入到虚拟健康社区用户生成内容研究中去,采集“好大夫在线”平台眼科学用户提问数据,提出了基于用户咨询内容的热点问题发现和知识推荐服务模型,通过挖掘热点问题以及问题主题相关关系,构建语义关联主题图谱并搭建了个性化推荐模型,实现了从用户提问出发研究提问主题特征,为虚拟健康社区平台研究提供了新的视角,丰富了该领域的研究体系结构。

实验结果表明,本文所提出的研究思路和研究方法能够对虚拟健康社区用户生成内容进行主题识别,发现虚拟健康社区用户的关注点和兴趣主题;基于K-Means 聚类技术的需聚合能够将虚拟健康社区用户的需求挖掘出来,帮助平台高效聚合需求;基于主题语义相似度,将虚拟健康社区用户生成内容的主题词用图谱的形式变现出来,知识图谱构建有效获取主题词间的语义联系;个性化推荐模型的构建帮助平台将适合用户需求的产品和内容信息精准推荐给相关用户,有效解决用户提问难提问不准确等问题。

**关键词:**

虚拟健康社区, 用户生成内容, 主题发现, 需求聚合, 个性化推荐

## Abstract

Topic discovery and personalized recommendation of user generated content in virtual health community

In recent years, with the rapid development of Internet technology, network resources play an important role in life, work and learning. More and more users choose to interact on the network media for efficient access to information and problem solving. With the advent of Web2.0, the mode of people using information has changed. Users can not only passively accept information content, but also participate in the process of information production and dissemination. Users can select information content in a large range, also can produce information content by themselves, screen out the information they want through information retrieval and other functions, or through feedback function They express their opinions on content products, so users of virtual community are not only consumers of information and cultural content, but also producers and providers. Because of their fast communication, strong editing and simple use, virtual community has been continuously concerned by the majority of users, which provides help to many people with medical and health needs. With the rapid development of our economy and the accelerating pace of people's lives, the virtual health community has become a platform for many people to understand, care and consult health problems because of its effective help users in solving daily health needs. "Internet plus medical" mode moves offline resources to online, and the information interaction between patients and doctors is completed on the online platform, which greatly saves. User time, save offline resources. With the rapid development of information economy, information in the

Internet is explosive growth. Virtual health community covers a large number of valuable information and knowledge, which provides new research objects for knowledge discovery in medical and health field. The mature development of big data technology provides reliable tools for research. Therefore, the theme discovery and content generated by users in virtual health community can be realized in the network environment Sex recommendation research.

How to manage and organize the content generated by users in virtual health community, mining the user needs, innovating the new model of virtual health community service, providing high-quality service for users in virtual health community has become a new problem for the research of virtual health community. In view of this, this paper introduces the theory and method of topic discovery into the research of user generated content in virtual health community, and proposes a hot topic discovery and knowledge recommendation service model based on user consulting content. By mining hot issues and topic related relationships, the semantic association theme map is constructed and personalized recommendation model is built, which can be researched from user questions. The research on the theme features of the question provides a new perspective for the research of virtual health community platform, enriching the research architecture in this field.

The experimental results show that the research ideas and methods proposed in this paper can identify the topics generated by users in virtual health community, and find the concerns and interest topics of users in virtual health community; the demand aggregation based on K-means clustering technology can mine the needs of users in virtual health community, and help the platform to efficiently aggregate the needs; based on the semantic similarity of topics, The topic words of user generated content in virtual health community are realized in the form of map, and

the knowledge map construction can effectively obtain the semantic connection between the topic words; the construction of personalized recommendation model helps the platform accurately recommend the product and content information suitable for the user's needs to the relevant users, and effectively solve the problems of difficult and inaccurate questions.

**KeyWords :**

Virtual health community ,User generated content , Topic discovery ,Demand aggregation ,Personalized recommendation

# 目 录

摘要.....	I
Abstract.....	III
<b>第 1 章 绪论</b> .....	1
1.1 研究背景及意义.....	1
1.1.1 研究背景.....	1
1.1.2 研究意义.....	2
1.2 国内外研究综述.....	3
1.2.1 虚拟健康社区研究现状.....	3
1.2.2 用户生成内容研究现状.....	4
1.2.3 知识图谱研究现状.....	5
1.2.4 研究现状述评.....	6
1.3 研究内容与研究方法.....	6
1.3.1 研究内容.....	6
1.3.2 研究方法.....	7
1.3.3 论文框架.....	8
1.4 论文创新点.....	9
<b>第 2 章 相关概念及理论基础</b> .....	11
2.1 虚拟健康社区的含义和特点.....	11
2.1.1 虚拟健康社区的含义.....	11
2.1.2 虚拟健康社区的特点.....	11
2.2 用户生成内容的概念和特点.....	12
2.2.1 用户生成内容的含义.....	12
2.2.2 用户生成内容的特点.....	12
2.3 知识图谱概述.....	13
2.3.1 知识图谱的概念.....	13
2.3.2 知识图谱的特点.....	13
2.3.3 知识图谱的应用.....	14

2.4 相关技术理论模型.....	15
2.4.1 Glove 词向量模型.....	15
2.4.2 LDA 主题模型.....	15
2.4.3 K-Means 聚类.....	16
<b>第3章 虚拟健康社区用户需求及用户交互分析.....</b>	<b>17</b>
3.1 虚拟健康社区用户的内涵和类型.....	17
3.1.1 虚拟健康社区用户的内涵.....	17
3.1.2 虚拟健康社区用户的类型.....	17
3.2 虚拟健康社区用户需求.....	18
3.2.1 虚拟健康社区用户需求的内涵.....	18
3.2.2 虚拟健康社区用户需求的特点.....	18
3.2.3 虚拟健康社区用户需求的类型.....	19
3.3 虚拟健康社区用户信息交互.....	19
3.3.1 虚拟健康社区用户信息交互的含义.....	19
3.3.2 虚拟健康社区用户信息交互的特征.....	19
3.3.3 虚拟健康社区用户信息交互的形式.....	20
3.4 本章小结.....	20
<b>第4章 虚拟健康社区用户生成内容主题识别.....</b>	<b>23</b>
4.1 虚拟健康社区用户生成内容主题识别的目的及意义.....	23
4.1.1 虚拟健康社区用户生成内容主题识别的目的.....	23
4.1.2 虚拟健康社区用户生成内容主题识别的意义.....	23
4.2 虚拟健康社区用户生成内容主题识别的方法及过程.....	24
4.2.1 虚拟健康社区用户生成内容主题识别的方法.....	24
4.2.2 虚拟健康社区用户生成内容主题识别的过程.....	24
4.3 基于语义相似度的关联主题图谱构建.....	24
4.3.1 语义关联主题图谱的提出.....	24
4.3.2 语义关联主题图谱的构建方法.....	25
4.4 实验研究.....	25
4.4.1 数据采集与数据预处理.....	25
4.4.2 基于 LDA 主题模型的主题识别.....	26

4.4.3 基于 K-Means 的聚类分析.....	27
4.4.4 基于语义相似度的关联主题图谱构建.....	28
4.5 结论与分析.....	30
4.6 本章小结.....	31
<b>第5章 虚拟健康社区用户生成内容主题语义关联个性化推荐.....</b>	<b>33</b>
5.1 个性化推荐的目的及类型.....	33
5.1.1 个性化推荐的目的.....	33
5.1.2 个性化推荐的类型.....	33
5.2 用户生成内容主题语义关联个性化推荐过程.....	34
5.3 用户生成内容主题语义关联个性化推荐模型.....	35
5.4 本章小结.....	38
<b>第6章 结论与展望.....</b>	<b>39</b>
6.1 研究结论.....	39
6.2 研究不足与展望.....	40
参考文献.....	43
作者简介及在校期间所取得的科研成果.....	49
致谢.....	51



# 第1章 绪论

## 1.1 研究背景及意义

### 1.1.1 研究背景

近年来,随着互联网技术的蓬勃发展,网络资源在生活、工作和学习中发挥重要作用,越来越多的用户为高效获取信息和解决问题选择在网络媒体上互动,互联网注册人数大幅度增加,2020年9月29日,中国互联网信息中心(CNNIC)在京发布的第46次《中国互联网发展状况统计报告》显示,截止到2020年6月,我国网民规模达9.40亿,互联网普及率达67.0%<sup>[1]</sup>。

Web2.0的到来,改变了人们利用信息的模式,用户不仅仅是被动接受信息内容的“受众”,而且可以主动参与到信息生产和传播的过程中,用户可以在较大范围内选择信息内容,还可以自身生产信息内容,通过信息检索等功能筛选出自己想要的信息,或通过反馈功能发表他们对内容产品的意见,所以,虚拟社区的用户既是信息文化内容的消费者,也是生产者和提供者,虚拟社区因其交流快捷、可编辑性强、使用简单方便受到广大用户的持续关注。社交媒体是人们用来创作、分享、交流意见、观点及经验,允许用户生产内容,能够通过文本、图像、音乐和视频等方式呈现的虚拟社区和网络平台<sup>[2]</sup>,由于用户信息需求多样化,因此社交媒体的种类繁多,例如学术类虚拟社区小木虫、经管之家等,运动类APPkeep、咕咚、每日瑜伽等,健康类平安好医生、春雨医生、好大夫在线、糖友网等,用户可以根据自身的信息需求选择相应的交流社区。可以看出,现实生活中有很多实际问题都可以在网络中解决,由此可以节约用户大量时间,提高效率。在互联网背景下,在线医疗健康领域的应用越来越广泛,与此同时,随着居民生活水平的提高和健康管理意识的增强,人们通过网络获取健康信息的需求越来越来大<sup>[3]</sup>,医疗保健和互联网的交集为促进健康服务发挥了巨大的潜力,更重要的是他为许多无法寻求有效治疗的人提供了及时帮助<sup>[4]</sup>。虚拟健康社区将医生丰富的就诊资源搬移到网上,一定程度上缓解医疗资源分配不均的现状,同时解决了一部分患者看病难的问题。

我国经济高速发展,人民生活节奏不断加快,虚拟健康社区因其能够有效帮助用户解决日常生活中健康类需求,成为不少人了解、关注、咨询健康问题的平台,“互联网+医疗”模式将线下资源搬移到线上,患者和医生之间的信息交互在线上平台中完成,大大节省了用户时间,节约线下资源。信息经济发展迅猛,互联网中信息呈爆炸式增长,虚拟健康社区已经成为当前非常流行的健康信息交流

平台，人们可以从他人那里获取有用的信息并和她们交流经验<sup>[5]</sup>，虚拟健康社区涵盖了大量有价值的信息和知识，为医疗健康领域的知识发现提供了新的研究对象，大数据技术的成熟发展为研究提供了可靠的工具，据此可实现网络环境下虚拟健康社区用户生成内容的主题发现及个性化推荐研究。

### 1.1.2 研究意义

本研究具有理论意义和实践意义。

在理论层面：

(1) 扩展了虚拟健康社区研究内容体系。

本文研究聚焦虚拟健康社区用户生成内容，提出了基于用户咨询内容的热点问题发现和知识推荐服务模型，通过挖掘热点问题以及问题主题间相关关系，构建语义关联主题图谱并搭建了个性化推荐模型，实现了从用户提问出发研究提问主题特征，为虚拟健康社区平台研究提供了新的视角，丰富了该领域的研究体系结构。

(2) 丰富用户需求挖掘方法和平台管理手段。

用户信息需求不仅仅受主观因素影响，还会受到周围环境的制约，本文通过对虚拟健康用户咨询内容的分析，一方面通过数据挖掘和知识图谱等技术，获取用户生成内容主题并基于语义相似性得到主题关联图谱，充分挖掘用户需求，通过聚类研究得到用户咨询提问簇，将类簇可视化，便于直观展示。另一方面通过构建个性化知识推荐模型，能够有效帮助平台管理者及时将相关需求信息推荐给用户并进行需求预测，丰富平台管理手段，最终实现了复杂知识有序化，平台推荐精准化，改善平台服务水平，提升服务能力。

实践意义：

(1) 为虚拟健康社区用户生成内容主题发现提供了方法

本研究基于主题模型和词向量模型的主题挖掘方法，聚类技术方法，可视化展示方法，通过采集虚拟健康社区用户咨询内容，挖掘用户咨询热点主题获取用户现实需求，并得到了基于语义关联的主题图谱，为虚拟健康社区用户生成内容主题挖掘提供了一定的技术支持。

(2) 为虚拟健康社区知识推荐提供了参考模型

虚拟社区爆炸式的信息产出，使得用户不得不花费更多的时间搜寻信息，而平台运营者也需要更多的精力在海量数据中寻找用户的需求，难以做到及时准确掌握用户关注动态，导致平台发展缓慢，服务能力有限。因此本研究提出了基于主题发现的个性化推荐方法，并构建相应的推荐模型，为用户高效获取信息提供了捷径，同时帮助平台更好的管理知识，创新平台服务模式。

## 1.2 国内外研究综述

### 1.2.1 虚拟健康社区研究现状

随着“互联网+”概念的兴起，虚拟健康社区平台出现在人们的视野里，利用互联网技术将线下的医疗资源搬到线上，聚集了不同的用户，为用户提供各种医疗信息服务，在人们的生活中发挥着重要作用。虚拟健康社区的研究是学术界的一个热点前沿话题，国内外学者针对虚拟健康社区的研究体系逐渐完善，对其研究成果颇丰。

MercyMpinganjira 在文献<sup>[6]</sup>中运用资源交换理论和社会交换的影响理论，探讨了社会资本、感恩和负债在虚拟健康社区中对互惠意愿水平的解释作用；ZhiJunYan 等在文献<sup>[7]</sup>中使用收益与成本模型来检验职业健康委员会成员如何分享一般和特定的知识并利用结构方程分析了 323 名在线健康社区用户数据，揭示了个人利益促进知识共享，而成本阻碍了知识共享；刘萌萌在文献<sup>[8]</sup>中利用文献研究法将相关文献梳理后确定研究内容，问卷调查法确定受试样本数据，最后使用 SPSS 对收集的数据进行数据质量分析、基本人口统计学分析、各变量得分的方差分析，AMOS 对假设模型进行拟合度分析，利用多重线性回归方法进行路径分析，建立在线健康社区用户使用以及信息采纳行为综合模型，对模型进行实证研究，提出相应的建议和对策，为促进在线健康社区用户健康信息采纳行为意向的做法提供了依据；周涛等<sup>[9]</sup>认为在线健康社区的快速发展为用户获取医疗健康信息和服务提供了重要平台。但用户主要以搜索和浏览信息为主，缺乏知识分享意愿，这将影响社区的可持续发展。整合信息系统成功模型和社会支持两个视角，研究了在线健康社区用户知识分享行为，对收集到的 326 份有效问卷采用结构方程模型进行分析，研究了在线健康社区用户知识分享行为，研究结果发现，信息质量和服务质量显著影响用户对社区的信任，信息支持和情感支持显著影响用户对其他成员信任，信任和隐私风险决定用户知识分享的意愿。研究结果启示在线健康社区应重视信息服务质量，为社区用户创造可信的互动环境，减少用户隐私担忧，从而促进成员的知识分享意愿和行为；潘建鹏等<sup>[10]</sup>通过对知识融合相关理论、方法和技术进行梳理和总结形成虚拟健康社区专家知识融合的系统框架，以社区的疾病百科、科普文章为基础并结合相关标准文档构建高血压疾病本体知识库，通过对医患平台的文本数据进行知识抽取和知识融合实现专家知识的有效扩充并进行了相关知识推荐，充分揭示了虚拟健康社区专家知识组织、组织知识融合与知识服务间关联；宋拓在其论文<sup>[11]</sup>中通过从用户需求的角度来解决知识聚合的高效化和精准化，用户画像的分析得出虚拟健康社区用户的需求，构建了虚拟健康社区的知识聚合模型并给予谱聚类算法实现虚拟健康社区知识聚合，基于“好大夫”健康平台展开分析，最后对虚拟健康社区的只是聚合效果

进行了评价研究,通过理论分析和技术实现,提出来虚拟健康社区只是聚合的优化策略,一次全方面提升虚拟健康社区中知识资源给用户带来的愉悦体验;杨瑞仙在文献<sup>[12]</sup>中针对目前在线健康社区知识交流效率研究未剔除环境因素影响的问题,在分析环境因素对在线健康社区知识交流效率影响的基础上,剔除外生因素对在线健康社区知识交流效率的影响,为客观评价在线健康社区的知识交流效率提供参考。

### 1.2.2 用户生成内容研究现状

Web2.0 的到来,信息技术飞速发展,用户生成内容发挥的作用越来越重大,应用领域愈发广泛,电商平台的消费者通过其他购买并使用过某种产品的消费者所创造的内容可以更加准确客观地了解到产品的质量,辅助其做出判断选择,更多社交媒体的用户生成内容也会成为其他用户决策的重要依据。此用户生成内容的应用拓展到各个行业领域,因此其发挥的作用也越来越有价值,国内外学者对用户生成内容的研究涉及到方方面面,研究成果意义重大。

MarioL 等<sup>[13]</sup>通过对社交媒体网站上用户生成内容中用户满意度体验的调查,确定葡萄酒在高档餐厅顾客满意度中的作用;ImtiazArif 等<sup>[14]</sup>通过评估 Facebook 品牌相关用户生成内容对消费者行为的影响,研究结果显示,UGC 会影响顾客的反应,促使顾客传递有关品牌的信息,购买产品或打算购买产品并与品牌保持联系以获得最细的更新信息;QiangYe 等<sup>[15]</sup>实证研究了旅行者评论对在线销售的影响,研究结果强调了用户在线生成的评论对旅游业经营业绩的重要性;Assaker<sup>[16]</sup>采用结构方程模型研究了信任、专业化、感知有用性和感知易用性对女性和男性年轻人和老年人对用户生成内容和在线评论使用意愿的影响,并以旅游网站旅游信息为你研究样本,研究结果表明在线旅游评论中性别和年龄的影响提供了额外的贡献,有助于理论和实践的发展;陶兴等<sup>[17]</sup>针对当前学术问答社区内用户生成内容的自动摘要生成问题,提出改进的 W2V-MMR 自动摘要生成算法,利用基于深度学习的 Word2Vec 词向量生成模型,优化摘要句信息质量;引入最大边界相关(MMR)的思想,对学术问答社区内的用户生成问答文本进行自动摘要,为社区内的学术用户提供高效、准确的知识聚合服务;杜永峰<sup>[18]</sup>基于 CiteSpace 以 2007-2019 年国内外发表的关于用户生成内容的文章为数据来源,对其热点和作者进行可视化分析,通过统计研究领域的发文信息、研究热点、研究作者进行梳理和分析,将结果利用 CiteSpace 软件绘制图谱进行可视化分析结果显示用户生成内容研究热点有一部分一直与食品、媒体等实际应用相关,研究热点不断细分并扩展到不同领域;刘婷艳等在文献<sup>[19]</sup>中通过分析和收集国内外视频网站用户生成内容相关文献,以了解其在该领域的研究进展及研究趋势,为其在学术领

域与商业领域等研究和应用提供借鉴与参考，运用文献及知识图谱分析方法，对国内外视频网站用户生成内容的起源与发展、研究热点、未来发展趋势进行对比和可视化分析；刘莉等在文献<sup>[20]</sup>中以社会影响理论为基础，引入信任维度作为中介变量，构建用户发表内容特征对社会化电商消费者购买意愿的概念模型，并采用探索性因子分析和线性回归分析对问卷调查数据进行分析。

### 1.2.3 知识图谱研究现状

知识图谱是一个蕴含丰富语言信息的有向异构网络结构图，可以辅助补充数据之间的关系，近年来，知识图谱在搜索引擎和自然语言处理的研究引起了学者们的关注，研究者们将知识图谱技术应用到各个领域，包含教育、医疗、电力、交通等行业领域，为行业发展提供了便利，促进了技术的发展和创

新。SunHaixia 等<sup>[21]</sup>开发了一种利用医学知识图谱识别临床可疑索赔的自动化方法，以便进行 FWA 检测，实验结果表明，医学知识图谱有助于发现 70% 的疑似索赔，该方法可成功地从索赔文件中识别出疑似 FWA 病例，有助于提高索赔处理的效率；ShujunHuang<sup>[22]</sup>提出了一种基于知识图谱的智能油气搜索引擎，与传统服务相比，它在搜索油气信息方面有更多的优势，更容易了解用户的搜索意图，有效提高了搜索质量；HanchuanXu 等<sup>[23]</sup>将知识图谱应用在养老服务领域，提出了新的知识处理和容和方法构建养老服务知识图谱，实验结果表明该方法能快速构建知识图谱，提升养老服务质量；严春来<sup>[24]</sup>和徐嘉成<sup>[25]</sup>分别从图书馆视角出发，利用 Citespace 可视化软件对科学数据热点和趋势进行了图谱分析和通过 CiteSpace 对 CNKI 中文期刊“会展旅游”相关 737 篇有效文献进行分析，对会展旅游产业关系、发展方法、产品构建作出探讨；孙鑫瑞等<sup>[26]</sup>基于开源数据构建交通知识图谱和事理图谱，针对微博文本部分使用交通知识图谱进行交通事件的识别；针对微博中的图片使用目标检测进一步提升三类事件的识别准确率利用交通知识图谱和目标检测的方法实现微博中交通事件的识别，辅助解决交通问题；王琼等<sup>[27]</sup>通过对系统数据的挖掘，并将人工智能、知识图谱及自然语言等技术与其相融合，构建电力行业的知识图谱，根据电力行业的特点采用人工智能进行分析，从而对电力行业的发展进行预测。知识图谱技术的推广应用，可以推进电力行业的科学化发展和信息化建设，从而实现智能电网与用户的双向友好互动，保证各环节移动互联网，扩大智能电网的应用领域，具有很好的应用前景和现实的应用价值；赵一鸣等在文献<sup>[28]</sup>中基于百科知识图谱 CN-DBpedia 构建以文本中的命名实体为节点、实体之间的语义关系为边的语义图，然后引入概念图谱 CN-Probase，实现实体和概念之间的映射，进而生成融入概念知识图谱的增强型中文文本语义图，最后以新闻文本的模式发现任务为例对本文提出的方法进行了

验证。

## 1.2.4 研究现状述评

通过以上研究现状梳理,可以看出虚拟健康社区相关研究是国内外学者的一个前沿热点话题,且普遍通过对用户数据进行挖掘研究,研究重点在用户使用满意度、知识共享、专家知识推荐等方面,总体来说,国内针对虚拟健康社区的研究还处于初级阶段,集中于概念,社区用户行为等方面的讨论,但是也有近几年研究有偏重于应用实践,学者们通过构建一定的模型探讨虚拟健康社区更深层次的应用。在用户生成内容研究方面,国内外学者研究领域比较宽泛,但更多研究集中在电子商务领域,通过用户生成内容直观揭示用户满意度,研判用户生成内容对顾客购买意愿的影响,研究 UGC 对用户行为的影响贡献,与此同时,国内学者还从用户生成内容数据中挖掘更加有价值的情报,例如知识推荐,知识聚合研究等,总之,用户生成内容研究逐渐细分到各个行业领域,研究价值越来越高。知识图谱是近几年学者们较为青睐的研究手段,应用非常广泛,国外学者更加知识图谱的注重实践应用,例如在医学、油气搜索、养老等领域的个性化推荐应用大受欢迎,而国内学者在注重实践应用的同时也在关注理论研究,大部分理论研究是基于知识图谱的研究热点挖掘,趋势预判等,实践应用也涉及较多,例如舆情分析,话题检测、情报推荐等。以上分析可以看出,在虚拟健康社区研究领域,研究理论体系不够完善,缺少从用户角度出发,深层次挖掘用户创造内容的成果,知识图谱虽然应用领域较为宽泛,表达倾向性行强,但普遍集中在知识发现研究,缺少对信息的整合讨论,少有从网络用户角度出发,挖掘用户需求的研究成果。

鉴于此,本文在已有的国内外研究基础上,提出了主题发现、需求挖掘、需求聚类研究方法,深度挖掘虚拟健康社区用户生成内容主题,并构建基于用户生成内容咨询主题的知识推荐模型,探讨虚拟健康社区服务新模式,帮助平台提升服务质量和效率。

## 1.3 研究内容与研究方法

### 1.3.1 研究内容

本文针对虚拟社区中用户提问咨询进行主题挖掘和推荐服务研究,主要研究内容和章节安排如下:

第 1 章绪论。本章主要介绍本文的选题背景和研究意义,梳理了国内外研究现状,提出了本文的研究内容和研究方法,制定了本文研究框架,指出了本文创新点。

第 2 章相关概念及理论基础。本章第一节介绍了虚拟健康社区的概念和特点,

第二节对用户生成内容的含义和特点进行了深层次剖析,其次阐述了知识图谱的相关概念、特点和应用,最后介绍了本文的技术方法。

第3章虚拟健康社区用户需求及用户交互分析。本章主要分析了虚拟健康社区用户的概念和特点,阐述了用户的分类,其次研究了虚拟健康社区用户的信息需求,包括需求的内涵、特征、类型,最从用户交互的概念、形式和影响因素分析了虚拟健康社区用户信息交互。

第4章虚拟健康社区用户生成内容主题识别。本章首先介绍了用户生成内容主题识别的目的和意义以及主题识别的过程,提出了基于用户生成内容主题识别方法,通过采集虚拟健康社区中用户咨询文本数据,利用LDA主题识别模型挖掘用户需求,通过Glove词向量模型将文本数据向量化,基于K-Means将用户生成内容进行类别划分,再基于语义相似度构建关联主题图谱,为后续基于主题语义关联个性化推荐提供研究基础。

第5章虚拟健康社区用户生成内容主题语义关联个性化推荐过程及方法。本章首先分析了主题语义关联个性化推荐的目的和原则,其次阐述了推荐过程和推荐方法,最后重点构建了虚拟健康社区基于用户生成内容主题语义关联个性化推荐模型。

第6章结论与展望。本章总结了全文观点和结论,分析了本研究存在的不足和未来的研究方向,为以后的研究指出方向。

### 1.3.2 研究方法

本研究过程中主要采用的研究虚拟健康社区用户生成内容法有:

(1)文献分析法。文献分析法主要用于分析虚拟健康社区,用户生成内容,知识图谱等有关文献,整理了国内外学者当前的研究成果,总结了研究现状,同时基于以上文献梳理提出了本文的研究对象,研究目的和意义,此外通过文献分析,为本文的研究方法和技术奠定了坚实的基础。

(2)实例分析法。本文研究虚拟健康社区用户生成内容,经过整理和调研分析,选用“好大夫在线”平台的用户咨询提问为研究对象,采集了“好大夫在线”平台上眼科学模块用户的提问相关数据内容,验证本文所提方法和技术的可行性和有效性,为同类研究对象的研究提供方法借鉴和技术支持。

(3)文本挖掘法。文本挖掘法主要用于对虚拟健康社区用户生成内容的深层次分析,通过数据挖掘技术对虚拟健康社区用户生成内容进行主题识别分析,发现用户生成内容的主题,利用Glove词向量模型将文本数据向量化,再通过余弦相似度计算与主题词相似的前4个相似词,为主题图谱构建做好相应的准备。

(4)可视化分析法。可视化技术是近年来备受青睐的技术,应用非常广泛,本文通过可视化软件—Gephi软件对处理分析后的数据进行可视化展示,直观反

映基于 LDA 主题识别和 Glove 词向量模型处理之后的用户生成内容主题词间关联关系，为虚拟健康社区个性化推荐模型构建提供理论基础。

(5) 聚类分析法。聚类来源于人们对事物形成的合理猜想——一个领域中存在某种内因使一些事物彼此相似且有别于其他事物。聚类技术就是致力于发现这种内在特征，并使类内部尽量紧密、类之间尽可能分离。文档聚类是指把文档集内某些具有相似性的个体组织在一起构成子集，每个子集称为一个簇，力图使相同簇中的文档彼此相似，不同簇的文档彼此不相似。本文使用 K-Means 聚类技术将虚拟健康社区用户的需求进行聚类分析，便于高效聚合用户需求，深层次挖掘用户需求。

### 1.3.3 论文框架

本文首先进行虚拟健康社区用户生成内容主题识别工作，利用 LDA 主题模型挖掘出用户提问咨询文本中的主题词，得到用户的需求和关注点，其次根据 Glove 词向量模型将文本向量化。

在主题识别和向量化数据的基础上，根据余弦相似度得到与主题词相似的前 4 个关键词，可以深层次挖掘用户需求主题，预测未来某一段时间的需求，并将主题词和相似词借助 Gephi 软件利用主题图谱方式直观可视化表示出来，有助于简单快速发现用户需求和关注点。

提出了基于 K-means 均值聚类的方法，将用户需求聚合分析，可使原本杂乱无序的文本数据被合理划分到不同的类簇，再使用词云图可视化展示不同类簇，将用户的高频需求抽取出来，便于清晰准确的分析用户需求，利用价值更高。

最后本文提出了虚拟健康社区用生成内容个性化推荐研究，旨在通过对用户提问咨询的主题挖掘和分析，提炼出用户关注的高频需求和预测未来用户需求，将高频需求推荐给平台上其他用户，免去了其他用户浏览检索的时间，大大提高查询效率，节约时间，同时提升平台服务水平。

本文的研究路线图参考图 1.1 所示。



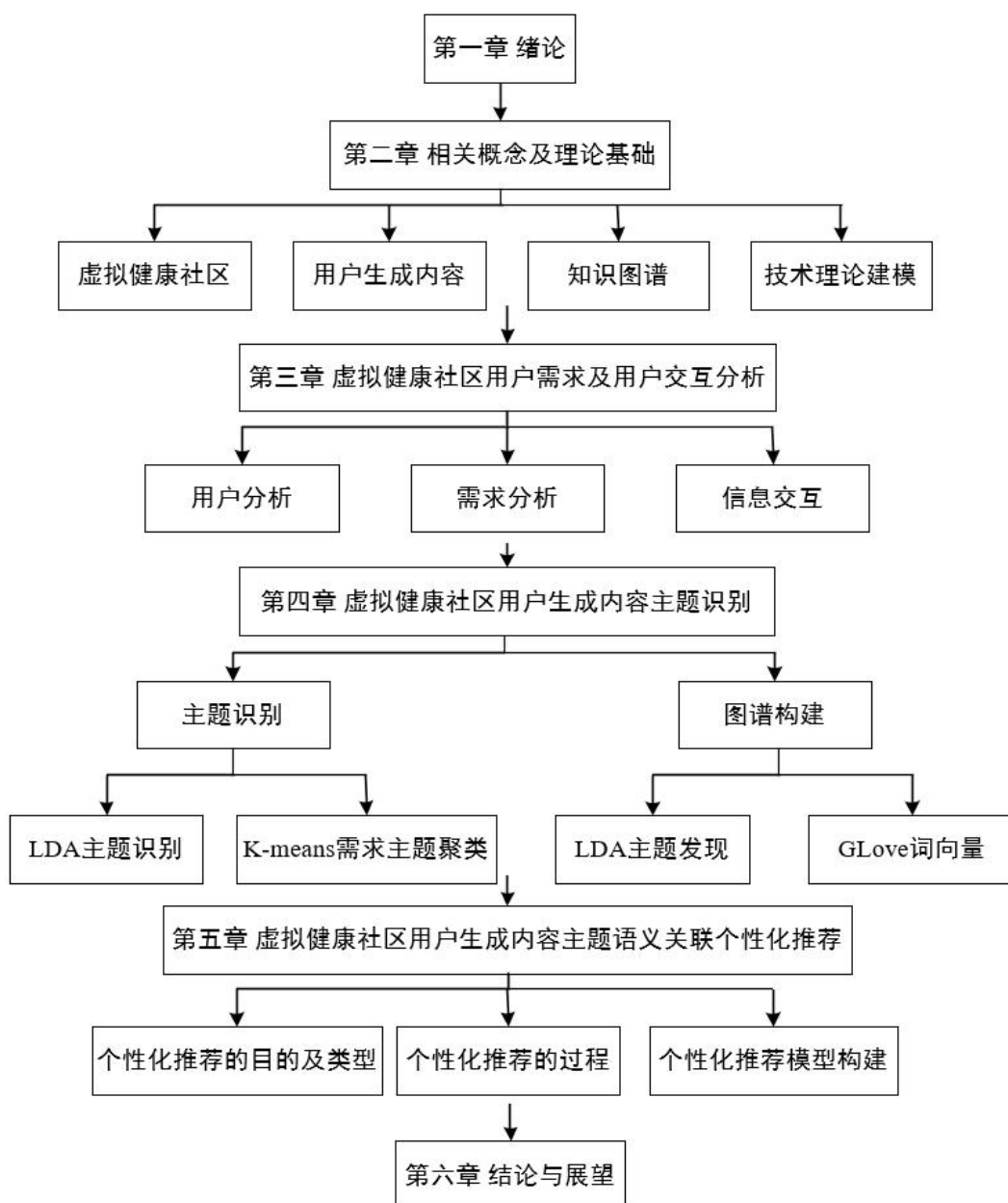


图 1.1 本文研究路线图

## 1.4 论文创新点

本文的创新点体现在以下几个方面：

(1) 本文选取当下热门社交媒体—虚拟健康社区为研究平台，选用用户提问咨询为研究对象，分析了平台用户交互和需求的特点。利用大数据技术，从海量数据中挖掘用户关注点和需求，角度新颖，方法便捷。

(2) 提出虚拟健康社区用户生成内容主题发现研究方法，通过 LDA 主题模型获取虚拟健康社区用户提问咨询主题，再利用 GloVe 词向量模型扩展了主题词，利用知识图谱直观展示了主题词间关联关系，为深层次挖掘用户生成内容的主题，

从非结构化客观数据中挖掘用户关注主题及主题间关联关系提供了很好的借鉴。其次将用户生成内容进行聚类分析并可视化展示，直观反映用户需求和关注点。

(3) 提出了虚拟健康社区基于用户生成内容主题语义关联的推荐方法，构建了基于主题发现的个性化推荐模型，实现虚拟健康社区用基于用户生成内容分析的精准推荐，为虚拟健康社区用户提供更高效的服务。

## 第2章 相关概念及理论基础

### 2.1 虚拟健康社区的含义和特点

#### 2.1.1 虚拟健康社区的含义

医疗资源分布不均衡，医疗水平差异明显，看病难看病贵等问题一直是我国重视且不断改进的医疗问题，随着数字时代的到来，互联网技术的快速发展，传统的就医方式发生了改变，线上线下相结合的就医模式给患者和医生都提供了极大便利。与此同时，移动技术的成熟发展，社交媒体成为人们交流的便捷工具，虚拟健康社区是一种常见的，使用率较高的社交媒体类型，也是学术界关注较多的研究对象，因其属于社交媒体的一种类型，所以大部分学者不再赋予其新的定义，鉴于此，本文引用文献<sup>[29]</sup>概念，认为虚拟健康社区是“互联网+医疗”模式下服务大众的窗口，通过盘活、整合现有线上线下资源，构建患者和医生之间的医疗知识交互平台，常见的虚拟健康社区有“好大夫在线”平台，春雨医生、平安好医生，虚拟健康社区话题论坛等。虚拟健康社区包含较多有价值的医疗信息，用户的健康咨询问诊，医护人员针对医疗卫生问题的答疑解惑，虚拟社区开放自由的性质方便用户快速搜寻查找相关症状和解决方案，用户不必重复提问咨询，同一问题不需要医生多次解答，通过虚拟健康社区交流，有效减少就医成本，实现医患间高效率沟通，改善传统就医弊端，创新医患交流模式。

#### 2.1.2 虚拟健康社区的特点

虚拟健康社区的产生对大家的生活观念，工作方式和日常行为产生越来越深刻的影响，越来越多的用户选择使用虚拟健康社区平台关注自己和家人的身体健康情况，了解虚拟健康社区的特点有助于我们更好地利用虚拟健康社区解决生活中的医疗问题，帮助我们较为全面客观的挖掘平台信息内容。本文总结了以往学者们对虚拟健康社区的研究成果，概括虚拟健康社区的特点有：

(1) 用户交流开放化。虚拟健康社区平台与现实场所不一样，它具有虚拟社区的特点，是开放自由的，没有“围墙”限制，不设置“通行证”，任何用户都可以参与注册使用，是给用户的信息交流的一种手段，所有用户都可以在平台上咨询问题，提出意见，用户之间也可以评论交流参与讨论。

(2) 用户需求比较集中。虚拟健康社区属于医疗服务平台，用户大都为医生患者或对医疗卫生比较关注的人群，群聚性高，都希望能够享受好的持续性的医疗服务，高质量的服务水平会增加用户黏性，吸引更多的用户参与使用，因此大家的需求也比几种，更多

集中在医疗卫生相关方面，

(3) 社区管理松散化。虚拟平台是没有限制的，没有现实中各种管理人员，人们可以漫无目的的任意进出，用户交流具有超时空性，通过移动网络，人们之间的交流不受地域的限制，不受时间以及用户身份的限制，每个人都可以参与交流，其管理是松散自由的。

## 2.2 用户生成内容的概念和特点

### 2.2.1 用户生成内容的含义

虚拟社会是随着网络的发展以及网络与社会的相互作用而出现的一个新概念，是对应于现实社会的，并与现实社会生活系统有着较大区别的一个系统，虚拟社区是现有虚拟社会的一种表现形态，用户交互是在社交媒体这样的虚拟社区中进行的。随着 web2.0 技术的发展，用户不再是被动地接受各种信息和知识的受众，而是逐渐向主动创造信息方向发展，主动生产问题、参与评论和转发、自由点赞与关注收藏，在交互过程中，人们的信息身份是不断转变的，既是信息知识的受众，也是生产信息的创造者，同样是信息知识的搬运者，这也是 web2.0 时代最突出的特点。在虚拟社区中用户主动创造的信息内容就是用户生成内容 (UserGeneratedContent, UGC)，用户生成内容又被称为用户创造内容，UGC 最早起源于互联网领域，即用户将自己原创的内容通过互联网平台进行展示或者提供给其他用户。UGC 是伴随着以提倡个性化为主要特点的 Web2.0 概念兴起的，其最大的特点就是个性化创造。在越来越开放的网络环境下，这种提倡用户参与其中的模式很快得到了网民的青睐，不仅仅可以查找网站本身的内容而且可以参与到网页内容的补充中，自己创造内容提供给其他用户浏览或评论，用户由原先的以下载为主变成下载和上传并重，从 2005 年起，伴随着视频分享网站 YouTube 的诞生，UGC 模式迅速兴起发展，学界对 UGC 的研究与探讨从未停止，研究成果亦颇有价值<sup>[30]</sup>，但是用户生成内容至今尚未有一个统一的定义。

用户生成内容的研究已是学术界的一个热点主题，主要理论研究涵盖用户生成内容概念研究、用户生成内容类型研究、用户生成内容动机研究、用户生成内容质量研究、用户生成内容引发的相关法律研究、用户生成内容数据挖掘研究，主要应用研究有用户生成内容在教育中的应用、用户生成内容在地理定位中的应用、用户生成内容在电子商务中的应用、用户生成内容在民主政治中的应用、用户生成内容在公共没你中的应用、用户生成内容在网络游戏中的应用、用户生成内容在图书馆、博物馆中的应用<sup>[31]</sup>，可以看出学者们对 web2.0 环境下应运而生的用户生成内容相对重视，研究视角广泛，研究内容多样，研究成果渗透到较多领域，UGC 在未来的互联网空间领域还有巨大的开发潜力<sup>[30]</sup>。

### 2.2.2 用户生成内容的特点

通过梳理以往的研究成果，学者们在总结用户生成内容特征时往往根据情境的不同，提炼用户生成内容的不同特征，例如刘雪纯在其论文<sup>[32]</sup>中研究的是社会化电商情境下的

用户生成内容，即依托电商类平台进行创作和发布的关于产品或服务信息的原创内容，通过总结以往的研究成果，探讨了电商情境下用户生成内容的特征。本文结合虚拟健康社区的特点，认为虚拟健康社区用户生成内容有以下特点：

(1) 内容个性化。个性化是用户生成内容最本质的特征，虚拟健康社区是开放自由的，用户生成内容就是用户自己发表的言论，每个人的语言特点，语言风格，用语习惯都是不一样的，因此不同用户的言论都有自身独有的特点，个性化程度较高。

(2) 内容质量高低不一。用户的职业、年龄、受教育程度的不同会影响用户生成内容的质量，有的人会认真交流心得，探讨专业问题，了解并使用专业术语，而有些用户只是简单随意地发表想法，甚至还有带着一定目的而来的广告人员在虚拟健康社区平台发布小广告。由此可见，虚拟健康社区用户生成内容因为用户之间的差异，

(3) 文本非结构化。用户生成内容最大的一个特点就是由用户自身发表的，口头表达主要思想，不需要考虑逻辑是否合理、格式框架是否正确，语义语序是否有问题，语言风格口语化，没有固定的模板和要求，因此用户生成内容数据呈现非结构化特点，表现形式风格各异。

## 2.3 知识图谱概述

### 2.3.1 知识图谱的概念

知识图谱是近年来学术界使用频率较高的一种科研工具，各行各业都在使用。知识图谱（KnowledgeGraph）又称科学知识图谱，是一种知识库，最早是由 Google 公司在 2012 年提出来的，目的在于提升搜索引擎的搜索质量，增强用户搜索体验<sup>[33]</sup>，知识图谱在近几年得到迅速发展，应用领域非常广泛，如数据分析、个性化推荐、智能问答等，在图书情报领域，知识图谱普遍被用来绘制、分析研究主体之间的相关关系，并将其关系进行展示的可视化工具。

大数据的精准分析需要知识图谱，数据时代，各行各业都累积了规模可观的大数据，但是这些数据并没有被很好地利用起来，未发挥其应用的价值，这是由于当前的机器缺乏知识图谱这类背景知识，无法准确深入理解数据的内在含义，限制了对这些大数据的精准与精细分析<sup>[34]</sup>，数据只是被杂乱存放起来，并没有凸显出本身的价值。本文利用知识图谱将虚拟健康社区用户生成内容主题关联起来，帮助平台有效挖掘用户需求。

### 2.3.2 知识图谱的特点

知识图谱最明显的特征就是结果呈现可视化，除此之外，本文参考文献<sup>[35-36]</sup>总结归纳知识图谱特点如下：

(1) 可扩展性。知识图谱实质上是一个结构化的语义库，基本组成单位是“实体—关系—实体”或“实体—属性—属性值”的三元组，实体是知识图谱的基本组成元素，关

系是指两个基本元素之间的语义相关关系，属性是对实体的揭示说明，是实体与属性值之间的映射关系，通常来讲，在知识图谱的表示中，节点代表实体，边代表实体间语义关系或实体的属性值，实体间通过关系相互联结，构成网络状的知识结构，可以根据实际需要删减或增加实体，扩展性强。

(2) 知识发现性。知识图谱所呈现的不仅仅是学科知识的基本情况，还需要揭示实体间背后的关联关系、规律和演化趋势，在揭示现有知识关系的过程中从而生成新的知识，这也是知识图谱很重要的一个特点。

(3) 结构简单。知识图谱旨在说明实体与实体之间的关联关系，没有复杂的影射关系，表达简洁清晰，实体含义简单，节点描述通俗易懂，组成结构简单。

(4) 表现形式多样。知识图谱简单来说是一种可视化工具，其展现形式既可以是二维的平面图形，也可以是立体图形，既可以是静止的状态，也可以动态揭示实体间的演变趋势。

知识图谱依据不同标准可忽而分为不同的类型，本文参考文献<sup>[37-38]</sup>总结了知识图谱的几种类型，如下：从构建过程是否依赖自动抽取技术，将知识图谱分为早期专业人士手工构建的知识图谱和从互联网中自动抽取实体与关系构建的知识图谱，前者规模小，质量高，能够确保准确性和完整性，后者规模较大，但是存在大量不完整的信息；从覆盖范围，知识图谱可分为通用知识图谱和行业知识图谱两类，通用知识图谱着重于知识的广度，对知识的精准度要求不高，多描述常识性的知识，而专业知识图谱则强调某一专业领域的学科知识，对行业的专业性和准确度要求比较高。

### 2.3.3 知识图谱的应用

知识图谱能够提升知识管理水平，优化知识管理模式，近年来，在搜索引擎和自然语言处理的研究引起了学者们的关注，应用领域非常广泛，涵盖社会各行各业，可用于领域热点分析<sup>[39]</sup>，知识服务<sup>[40]</sup>，医学决策支持<sup>[41]</sup>，智能问答<sup>[42]</sup>，军事领域研究<sup>[43]</sup>、科技资源领域<sup>[44]</sup>、金融领域<sup>[45]</sup>、个性化推荐研究<sup>[46]</sup>、教育教学<sup>[47]</sup>等。在用户行为分析、信息智能化处理等方面知识图谱具有很好的发展前景<sup>[48]</sup>。知识图谱的应用是当前学术研究的热点和重点，也是促进人工智能发展的基础技术之一。

知识图谱是运用结构化的语言，描述现实世界中存在的实体或概念间关系的知识库集合，被看作是一张具有知识结构的网络图<sup>[48]</sup>，因此，在现实生活中，大部分数据都可以以图谱的方式表达出来，只不过有的实体关系比较简单，用知识图谱表现出来更为清晰明了，而有的实体关系比较复杂，相对来说操作难度大。

## 2.4 相关技术理论模型

### 2.4.1 Glove 词向量模型

在深度学习文本分类模型中，一个重要的预处理步骤是对单词进行词向量表示，词向量是使用数字编码代替单词，目的是使计算机能够“认识”这些经过编码后的单词，传统的 One-hot 词向量表示方法忽略了词与词之间的空间距离关系<sup>[49]</sup>。针对这个问题，Google 团队在 2013 年发布了 Word2vec 工具用来处理自然语言任务重文本词向量表示问题，利用大量文本语料库训练出来的文本词向量表示工具能够较好地反应词与词之间的距离关系<sup>[50]</sup>。PenningtonJ 等于 2014 年提出 GloVe (GlobalWordVectors) 模型<sup>[51]</sup>，该模型通过词共现矩阵得到词向量，是一种无监督模型，将每个词语语义信息表示为数值向量，使得向量之间尽可能多的蕴涵语义和语法信息，与 Word2vec 词向量模型一样可以根据训练大规模语料得到词语的语义信息，Word2vec 词向量模型近年来在自然语言处理领域得到了非常广泛的应用，而 Glove 词向量在自然语言处理中很少用到，该模型不仅仅只关注上下文语义，而是用到了全局文本信息，根据全文文本分析语义，克服了 Word2vec 对多义词处理效果不好的问题<sup>[52]</sup>。GloVe 词向量模型结合了 word2vec 和 LDA 的优点，融入全局的先验统计信息，可以加快模型训练速度，因此，在本文研究中，采用 GloVe 模型训练词语的词向量，Glove 模型能够避免过多人工参与的主观性，避免只考虑局部语义信息的不完整性，且能综合考虑整体评论数据。本文在对虚拟健康社区提问数据经过预处理后，利用 GloVe 词向量模型将评论语词转化成 300 维的词向量便于后续处理，限于篇幅，本文不再进行相关展示。

### 2.4.2 LDA 主题模型

主题模型近年来在图书情报与档案管理领域的应用非常广泛，主题模型使用统计推断方法自动发现大规模语料库中隐含的规律，并赋予语义解释，这是实现文本挖掘和分析的重要手段<sup>[53]</sup>，主题模型最早可追溯到潜在语义分析 (latent semantic analysis, LSA)<sup>[54]</sup>，LSA 使用奇异值分解 (singular valuedecomposition, SVD) 来实现从词项空间到语义空间的映射，概率潜在语义索引 (probabilistic latent semantic indexing, pLSA)<sup>[55]</sup>对 LSA 加以概率化，对其生成过程给出了清晰的概率解释。LDA 隐含狄利克雷分布 (英语: LatentDirichlet allocation, 简称 LDA)，是一种主题模型，它可以将文档中每篇文档的主题按照概率分布的形式给出，是包含“文档—主题—词语”的三层结构，因此也叫三层贝叶斯概率模型。LDA 的优点是，对于每一个主题均可以找出一些词语来描述它，其在自然语言处理领域，包括主题识别、文本分类以及文本相似度计算均有较好应用。LDA 是一个非常典型的词袋模型，该模型认为一篇文档是由一组词构成的一个集合，词和词之间没有顺序以及先后关系，在处理文档时，只需考虑词汇是否出现，而不必考虑词汇出现的前

后顺序，一篇文档可以包含多个主题，文档中每一个词都由其中的一个主题生成。此外，随着训练文档数量的增加，LDA 模型的参数空间不会增加，因此，它具有出色的语义挖掘和主题分析能力，尤其适用于具有大量文本的数据分析<sup>[53]</sup>，其生成过程如下<sup>[56]</sup>：

- (1) 对每一篇文档，从主题分布中抽取一个主题；
- (2) 从上述抽取的主题所对应的单词分布中抽取一个单词；
- (3) 重复以上过程直至遍历文档中每一个单词。

采用 LDA 模型挖掘数据主题有以下优势：(1) 能够处理较海量的非结构化文本，对于大量数据的文本有较好的处理能力。(2) 泛化能力好，LDA 主题模型应用广泛，可依据不同的实际需要应用到各领域。(3) 扩展性好<sup>[57]</sup>。LDA 结合大数据应用，扩展到社会网络分析、情感分析方面。

### 2.4.3 K-Means 聚类

聚类算法属于非监督算法，其主要目的是把相似的对象聚到一起，因此，一个聚类算法通常不需要使用训练数据进行学习，而只需知道如何计算相似度即可<sup>[58]</sup>，K-Means 聚类也叫 K-均值聚类或 K-平均算法，是一种经典的聚类算法，典型的基于划分的聚类算法，该算法通过计算对象之间的空间距离来进行聚类，通过迭代计算簇中心的方式找到稳定中心，从而确定每个簇及其成员。K 其基本思想是首先需要划分簇的个数 K 值，从含有 n 个数据对象的数据集中选取 K 个数据对象作为初始聚类的中心，然后根据距离公式相似度来计算其余各个数据的对象到 K 个初始聚类中心的距离，根据近邻原则将数据对象划分到距离它最近的那个中心所在的簇中，调整生成新的簇，分别集散各个新生成的簇中数据对象的均值作为各簇的中心，比较这两个聚类中心，如果没有变化，则算法结束，如果不一样，就要重新对数据对象进行划分，直到完成收敛条件<sup>[59]</sup>。其步骤如下：(1) 首先确定聚类要形成的簇的个数 k，这是事先需要指定的参数。(2) 随机选取 k 个点作为 k 个簇的中心。(3) 根据欧式距离，度量所有文档和各类簇中心的距离，按照其远近将所有文档分配到最靠近的簇，得到 k 个簇。(4) 计算每个簇的质心，也就是其成员向量的平均值。这些质心称为各簇的心的中心值。(5) 返回执行 3、4 步，直到每个簇上分到的点与在上一轮分到的点相同，这时说明簇的中心稳定了，也就是簇中的成员确定了。

K-Means 聚类方法简单，可操作性强，聚类效果较好，而且将质心作为簇的中心可使簇中的每个点到中心的距离平方和达到最小。但这种方法也存在缺陷，上述最小值只是局部最小值而并非全局最小值。最终的簇对初试中心很敏感，初试簇中心指定不同的点，结果产生的簇可能就不同。为尽可能贴近全局最优结果，往往要换用不同的初始点多次执行 K-均值算法，然后从中选择一个相对最佳的结果，即距离平方和最小的那个。鉴于 K-Means 聚类的优秀算法，本文选用 K-Means 聚类完成虚拟健康社区用户需求聚合，得到用户的需求，便于平台进行整合和制定相应的改进方案。



## 第3章 虚拟健康社区用户需求及用户交互分析

### 3.1 虚拟健康社区用户的内涵和类型

#### 3.1.1 虚拟健康社区用户的内涵

用户是信息需求的出发者和信息的而最终接受者,一般而言,只有具备信息需求、信息利用能力和接受信息服务行动的人员才能称为信息用户,信息用户通常是指科研、技术、生产、管理、文化等各种活动中一切需求与利用信息的个人或团体。鉴于此,本文认为虚拟健康社区用户是指在虚拟健康社区进行浏览、咨询、交流、学习等一系列活动的个人或群体。用户通常在虚拟健康社区发表问题咨询,相关用户会进行解答,大部分是从事专业领域的医生和有医疗卫生需求的患者,其次,用户之间也会进行交流和讨论,例如某一患者A曾经在社区平台咨询过,具有相同需求的其他患者看到后会选择直接询问患者A,这样就不需要医生重复诊断,用户之间的交流讨论大大节省看病时间。

#### 3.1.2 虚拟健康社区用户的类型

信息用户依照不同的标准可划分为不同的类型,这些标准设计信息用户的实践活动、信息需求个人或团体的自然属性和研究目的,常见的划分标准有按行业类型划分、学科类型、服务方式类型、职业类型的等标准。鉴于此,本文依照虚拟健康社区的属性特征,将虚拟健康社区用户划分为四种类型,分别是咨询型用户、服务型用户、浏览型用户和综合型用户。

(1) 咨询型用户。这部分用户由于自身有某些确定性的健康类问题需求希望得到解决,因而在虚拟健康社区提问咨询,这部分用户在使用虚拟健康社区平台时会直接本着自己已有的目标选择,有针对性地交流讨论,虚拟健康社区咨询型类型的用户占比较大。

(2) 服务型用户。服务型用户顾名思义主要是提供服务的,这部分用户主要解决用户健康咨询问题,他们本身从事医疗工作或者扮演医护人员角色的个人或团体,为有需求的患者或其他咨询人员提供服务帮助,这部分用户主要解答咨询型用户的提问。

(3) 浏览型用户。此部分用户并没有体现出自身的显性需求,属于对虚拟健康社区平台信息知识的潜在需求,这种需求是用户存在但没有意识到的,或意识到而没有表达出来或不愿意表达的信息需求,但是显性需求往往是由隐性需求转化而来的,所以浏览型用户有很大的机会转化为咨询型用户。

(4) 综合型用户。综合性用户的表现方式多样,这部分用户扮演多种角色,综合性用户既在平台上提问,也会解答其他用户的问题,或者是有某种隐藏的需求没有表达出来,又或者给用户一定服务帮助,这部分用户可能既是咨询型用户又是服务型用户,又或者是浏览型用户,综合型用户在平台的活跃性程度相对较高。

## 3.2 虚拟健康社区用户需求

### 3.2.1 虚拟健康社区用户需求的内涵

用户信息需求是当个体遇到问题的一种心理状态,一般是指对事物的欲望和要求,用户购买和使用产品的过程就是在满足需求<sup>[60]</sup>。鉴于此,本文认为虚拟健康社区用户需求主要指虚拟健康社区用户为了解决自己现有的医疗问题,或在使用健康社区时对社区平台服务和社区功能以及自身要求的一种亟需满足的心理状态,既包含现有的必要的产品和功能需求,也包含未曾满足期待进一步完善改正地或增加的需求。信息需求是用户信息行为的起点,信息服务商提供商能否正确了解用户信息需求的特点和规律影响着最终的服务效果,尤其是在不断变化的网络环境下,电子信息资源数量剧增,挖掘和分析虚拟健康社区用户需求是服务商今后提升平台竞争力的努力方向之一。

### 3.2.2 虚拟健康社区用户需求的特点

不同的信息用户在面临不同的外在问题时具有不同的需求特点,信息需求不仅受主观因素的影响,而且受社会环境、生产条件的制约。本文借鉴文献<sup>[61]</sup>移动环境下用户需求特征的思路,将虚拟健康社区用户的需求特征概括一下三点:

(1) 需求广泛。虚拟健康社区用户多呈现个体性,类型多样,有的是长期访问的用户,有的是短期或单次体验用户,用户需求不尽相同,表现为不同用户的实际健康需求是不一样的,同时,用户对服务内容本身的需求也是碎片化的。第一印象往往成为用户是否再次使用的关键,而良好的初次体验会增加用户印象,进而增强用户的使用意愿,这就要求,虚拟健康社区应该高度重视用户需求的碎片化特征,广泛全面挖掘用户信息需求。

(2) 需求动态化。虚拟健康社区的用户需求是比较强烈的,随着时间的推移,用户生活和学习环境的变化,需求也在动态性变化,与此同时,信息资源类型的多样性也不断影响着用户需求的内容和方式。因此,虚拟健康社区平台应该及时了解用户变化的动态需求,提升服务质量。

(3) 需求超前性。用户需求往往追求新颖超前,快速及时,用户的目的是为了解决科研或者生活学习中亟待解决的问题,希望尽快得到结果,尤其在健康

类社区，用户更加迫切想要得到解决问题的方案，因此能否快速及时地获取信息是用户信息需求的首要条件。

### 3.2.3 虚拟健康社区用户需求的类型

用户需求有不同的分类标准，本文结合虚拟健康社区的特点将虚拟健康社区用户需求分为两种主要类型，一种是对社区内容的需求，一种是对平台服务的需求。

(1) 基于社区内容的需求。基于虚拟健康社区的内容需求主要指用户对平台资源本身的需求，平台资源类型是否丰富，涵盖面是否广泛，信息资源质量的优劣，知识的可获取性等都是用户所关注的重点，也是用户反馈需求的重要出发点。

(2) 基于平台服务的需求。基于平台服务的需求主要指用户根据平台的使用体验所反馈的功能性需求。例如平台的界面设计是否友好，选用功能是否全面，是否经常闪退，程序进程缓慢，收费项目等都在影响用户的体验，用户体验良好与否直接决定了平台的使用率。

## 3.3 虚拟健康社区用户信息交互

### 3.3.1 虚拟健康社区用户信息交互的含义

网络环境下用户信息交互行为是“媒介—社会—人”三者循环互动的关键<sup>[62]</sup>。用户的交互行为渗透在社会活动的方方面面，互联网正经历从“个体交互”到“群体交互”的过渡，将进入一个“人人时代”<sup>[63]</sup>，在虚拟健康社区平台上，用户交互环境不仅仅是依托网页的连接，而更加导向用户之间的连接，用户在平台上创作和使用信息，整合各类信息资源，用户之间相互交流信息，平台服务商提供交互大环境，进而形成“人—环境—信息”三位一体的关系网络，虚拟健康社区用户信息交互是用户生成内容的前提，正是通过用户信息之间的交互，才有了用户创造的信息，因此分析用信息交互是有必要的。

### 3.3.2 虚拟健康社区用户信息交互的特征

用户可以通过留言评论、点赞收藏等渠道交流信息，在与其他用户的交互过程中获取到有价值的知识，用户的评论是参与平台交互的主要形式，通过交互行为，吸收他人创造内容，从而充实自己的知识体系，完善自己的认知。根据交互的含义，本文总结虚拟健康社区用户信息交互主要有以下两大特征：

(1) 自由交互特征。虚拟健康社区识别开放的社区，它不受时间和空间的限制，用户只要利用网络，就可以与不同身份的人人交流，既可以面对面交流也

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：<https://d.book118.com/365302243212011300>