

• 数据分析与大数据挖掘 •

基于数据特征的在线健康社区信息融合模式研究

王 萍¹ 牟冬梅^{1,2*} 杨鑫禹¹ 李 桦^{1,2} 彭 浩¹ 李 茵¹

(1. 吉林大学公共卫生学院, 吉林 长春 130021; 2. 吉林大学第一医院, 吉林 长春 130021)

摘 要: [目的/意义] 探索基于数据特征的在线健康社区信息融合模式, 推动多源异构在线健康社区信息向健康知识的转化。[方法/过程] 在梳理相关研究现状的基础上, 解析当前在线健康社区数据特征给信息融合利用带来的挑战, 以序化论、转化论和融合论为理论基础, 构建在线健康社区信息融合模式。[结果/结论] 研究构建的在线健康社区信息融合模式, 包括 1 个核心目标、4 个融合层级、4 个构成要素和 4 个融合阶段。

关键词: 在线健康社区; 健康信息; 信息融合; 诊疗推荐

DOI: 10.3969/j.issn.1008-0821.2022.08.004

(中图分类号) G203 (文献标识码) A (文章编号) 1008-0821 (2022) 08-0028-09

Research on Information Fusion Mode of Online Health Community Based on Data Features

Wang Ping¹ Mu Dongmei^{1,2*} Yang Xinyu¹ Li Hua^{1,2} Peng Hao¹ Li Yin¹

(1. School of Public Health, Jilin University, Changchun 130021, China;

2. The First Hospital, Jilin University, Changchun 130021, China)

Abstract [Purpose/Significance] Explore the data feature-oriented online health community information fusion model, and promote the conversion of multi-source heterogeneous online health community information into health knowledge. [Method/Process] On the basis of combing the current research status, the study analyzed the challenges brought by the current characteristics of online health community data used in information fusion, and constructed an online health community information fusion model based on the sequence theory, transformation theory and integration theory. [Result/Conclusion] The online health community information fusion model includes a core goal, four integration levels, four components and four integration stages.

Key words: online health community; health information; information fusion; diagnosis and treatment recommendations

当前世界正经历着以数字化、网络化和智能化为显著标志的信息技术革命, 相关技术的成熟, 推动在线健康社区(Online Health Communities, OHCs)的诞生与蓬勃发展, 使其成为人们获得健康信息的重要渠道。2017 年发布的《中国网民科普需求搜

索行为报告》结果显示, 关键词“健康与医疗”是中国网民用户最关注的科普主题^[1]。为了满足人们日益增长的健康信息服务需求, 国家相继发布了多项政策推动互联网医疗体系的建设。2018 年 4 月 28 日, 国务院发布了《国务院办公厅关于促进

收稿日期: 2021-11-22

基金项目: 国家自然科学基金项目“信息链视域下电子病历数据驱动健康服务供给侧决策的路径与模式研究”(项目编号: 71974074); 吉林省科技发展计划项目“面向精准医疗的多模态医学数据融合与挖掘团队”(项目编号: 20200301004RQ)。

作者简介: 王萍(1990-), 女, 博士研究生, 研究方向: 医学数据整合与智慧服务。杨鑫禹(1996-), 女, 博士研究生, 研究方向: 医学数据整合与智慧服务。李桦(1981-), 女, 博士研究生, 研究方向: 医学数据整合与智慧服务。彭浩(1997-), 男, 硕士研究生, 研究方向: 医学数据整合与智慧服务。李茵(1983-), 女, 博士, 研究方向: 医学数据整合与智慧服务。

通讯作者: 牟冬梅(1970-), 女, 教授, 博士生导师, 研究方向: 医学数据整合与智慧服务。

“互联网+医疗健康”发展的意见》，意见中鼓励医疗机构运用互联网技术搭建线上线下一体化的医疗服务模式。2020 年 12 月，国家卫生健康委、国家医保局以及国家中医药管理局，联合发布《关于深入推进“互联网+医疗健康”“五个一”服务行动的通知》，以期推动解决人们就医过程中存在的“急难愁盼”问题。

综上，伴随大数据战略在医疗领域的深入贯彻和落实，促进了海量多源异构的在线健康信息的广泛积累，但也引发人们难以从海量复杂的健康信息中获取所需要的健康知识的问题，形成了健康信息供给泛滥与用户对精准健康知识需求之间的矛盾。目前的在线健康信息研究仍停留在个例事件，且碎片化，未上升到理论层，更未深入探讨信息融合过程的序化、转化、融合形成健康知识。因此，探索基于数据特征的在线健康社区信息融合，构建适应“互联网+医疗健康”新业态环境的在线健康社区信息融合模式，是实现用户健康信息需求与健康信息供给精准匹配的重要途径之一，有助于推动多源异构的在线健康信息向健康知识的转化。

1 相关研究现状

1.1 在线健康社区研究现状

近年来，在线健康社区作为用户分享和获取健康信息的重要途径，已经吸引国内外学者展开了多个视角的研究，分别围绕用户健康信息需求、健康信息搜寻行为和健康知识交互共享等方面展开研究。①用户健康信息需求方面：有研究者利用潜在语义索引模型构建用户的信息需求框架^[2]，也有学者探究用户行为特征和需求偏好^[3]，或利用医生的用户画像揭示患者对医生的需求关注点^[4]，以及患者寻求第二医疗意见需求的关键因素^[5]；②用户的健康信息搜寻行为方面：有研究者使用 Tobbi 眼动仪分析用户思维方式对信息搜寻行为的调节作用^[6]，分析用户参与行为动机^[7]。信息交互模式方面，发现用户与医生间的对话呈现螺旋式和直线式的交互模式^[8]，也有研究者探究用户信息搜寻过程的情感变化^[9-10]；③用户健康知识共享与传播方面：有学者探讨在线健康社区中专业的医学用户的健康知识共享意愿^[11]、知识共享型社区的构成^[12]、用户信息交互的网络拓扑结构^[13]、知识服务的主题^[14]、

问答可视化的知识图谱^[15]、信息分享过程的主题分布及主题动态^[16]、健康知识交流效率的差异^[17]和为用户提供健康知识推荐服务^[18]。

1.2 信息融合研究现状

早在 1973 年，美国国防部就开始资助声纳信号处理系统相关研究^[19]。信息融合模型是基于 JDL (Joint Directors of Laboratories) 数据融合模型演变而来^[20]，该模型将信息融合概括分为 3 层，即数据层、特征层、决策层^[21]。①数据层：主要探究为完成数据融合而专门设计的操作规则，例如：匹配连接^[22]、数据冲突^[23]、概率分类器^[24]、画像构建^[25]等；②特征层：顾伟等在多特征融合的基础上提出，多通道特征模型通过人体整体信息和人体部位的语义信息来检测行人^[26]；③决策层：有学者将其归纳为目标提炼、态势分析、威胁估计及过程精炼的四级模型^[27]，随后，又拓展到情境感知决策支持系统^[28]、网络大数据知识融合模型^[29-30]、面向应急决策全流程的信息融合框架^[31]。

通过梳理国内外的相关研究，发现学者从不同的视角对在线健康社区信息展开定性和定量的研究。从研究内容角度看，国内外学者围绕在线健康社区信息生态系统用户、信息内容和信息环境 3 个要素展开多方面的研究；从研究内容的发展趋势来看，国外对在线健康社区的应用和研究比国内稍早，从早期对在线健康社区的应用推广的接受程度研究，逐渐发展深入到健康知识服务效果和健康知识发现的研究。纵览国内外在线健康社区的研究成果，健康信息需求和行为方面的研究已经积累了丰富的研究基础，但有关在线健康社区信息融合利用方面的研究仍有研究空间。

2 在线健康社区数据特征为融合利用带来的挑战

在线健康社区种类繁多、用户数量庞大且增长速度快的特征，使得在线健康社区信息产生信息过载问题、组织结构复杂的规范性低问题、来源渠道广而可信度下降问题和信息内容价值密度稀疏的信息降级问题，都为在线健康社区信息的融合利用带来了挑战。

2.1 在线健康社区信息组织结构复杂且规范性低问题

由于在线健康社区自由开放的特征，平台并不

会对用户描述健康信息的方式、信息架构进行约束,形成在线健康社区信息组织架构的多样性,也带来了信息结构复杂、规范性差、利用难度高等诸多问题。在线健康社区具有社交媒体数据形式的不规范性、多样性,表现为在线健康社区用户生成内容的异构性。在线健康社区信息组织结构规范性低,为信息的融合利用带来消极影响,包括:①健康信息的可读性降低,当面对组织结构混乱的健康信息时,会加重用户的认知负荷,产生一系列健康信息加工处理的问题;②健康信息的可共享性降低,由于健康信息的可传递是建立在其科学性和规范性的基础上,也就意味着越不规范和不科学的信息,则越不容易被传递共享;③健康信息的可处理性降低,当健康信息处于混沌无序的状态时,信息的存储、降噪、语义消歧等处理工作都将面临巨大的挑战,使健康信息加工的时间成本、人力成本上升,降低产生新健康知识的效率。

2.2 在线健康社区信息积累体量庞大的信息过载问题

信息过载已经成为当代社会各领域中普遍存在的问题,该现象在社交媒体环境中尤为显著。在线健康社区作为依托社交媒体平台而存在的一种虚拟社交平台,致使在线健康社区中的用户所接收到的信息是无穷无尽的,其信息发送频率远远高于用户处理信息的认知能力。互联网时代下的在线健康社区信息过载,导致用户要面对越来越多嘈杂的、琐碎的、低价值的信息,但随着信息数量的增长,其边际价值和意义反而降低了。正因为在线健康社区信息存在过载问题,为用户利用信息带来了诸多的负面影响。一方面,因海量信息呈现出“先发布后过滤”的模式,很大程度上降低了用户对价值密度较高的健康信息的感知能力;另一方面,数据的体量巨大,也增加了用户获取高质量健康信息的时间成本、精力成本和财务成本。

2.3 在线健康社区信息价值密度稀疏的信息降级问题

Klapp O E 指出,信息过载不仅仅是信息数量的增加,更重要的是将会导致信息质量层面的降级^[32]。信息降级的表现有多个方面,如信息的噪音化、信息无用且过剩。社交媒体环境下的在线健康社区信息的来源渠道广、复杂异构、更新速度快

等特征,都是导致信息价值降级的原因。在线健康社区信息价值质量降级,为信息的利用带来了诸多挑战。首先,在线健康社区信息的主题特征减弱,信息价值密度低的最直观表现是冗余信息过多,该因素直接影响在线健康社区信息主题识别过程,包括医学领域实体识别、评估映射、映射模型构建过程;其次,用户受到虚假健康信息欺骗的概率增高,虚假信息是诱导人们形成错误健康认知、采取不恰当的健康行为的主要危险因素,给用户的身心健康和心理健康带来不可逆的危害;再者,数据分析利用的结果准确率降低,当输入的样本为海量高维的网络健康数据时,添加了许多与结局变量弱相关的混杂因素,增加机器学习运算时间和复杂度的同时,也降低了数据分析结果的精确度。

3 在线健康社区信息融合模式构建

针对在线健康社区信息特征为其融合利用带来的挑战,基于情报学研究的序化论、转化论和融合论,构建在线健康社区信息融合模式,包括1个核心目标、4个融合层级和4个构成要素。

3.1 在线健康社区信息融合的目标

在线健康社区信息融合是以提高信息的利用率为核心,以为用户提供精准高质量的健康知识服务为目标,探讨在线健康社区信息的序化、转化和融合过程,实现异构网络健康知识深度融合。在线健康社区信息的融合,是实现主观经验和客观知识的一体化,满足用户健康知识服务的需求,推动传统的基于用户个体经验的健康知识构建范式向动态的群体健康知识构建范式的演进。在线健康社区信息融合的原则,是探究泛在的在线健康社区信息与用户精准知识服务需求的供需平衡问题。在线健康社区信息融合的框架设计遵循科学性、系统性和有序性原则,综合实现各种健康知识链接的形式各尽其能,即融合的过程充分考虑健康信息与健康信息的实体链接关系、用户与在线健康社区信息实体间的知识链接。

3.2 在线健康社区信息融合模式的层级结构

基于在线健康社区融合利用所面临的挑战,研究构建了在线健康社区信息多层级融合模式,从健康信息集成层、序化层、转化层和融合层面探究数据融合的框架。研究构建的在线健康社区信息多层级融合模式,如图1所示。

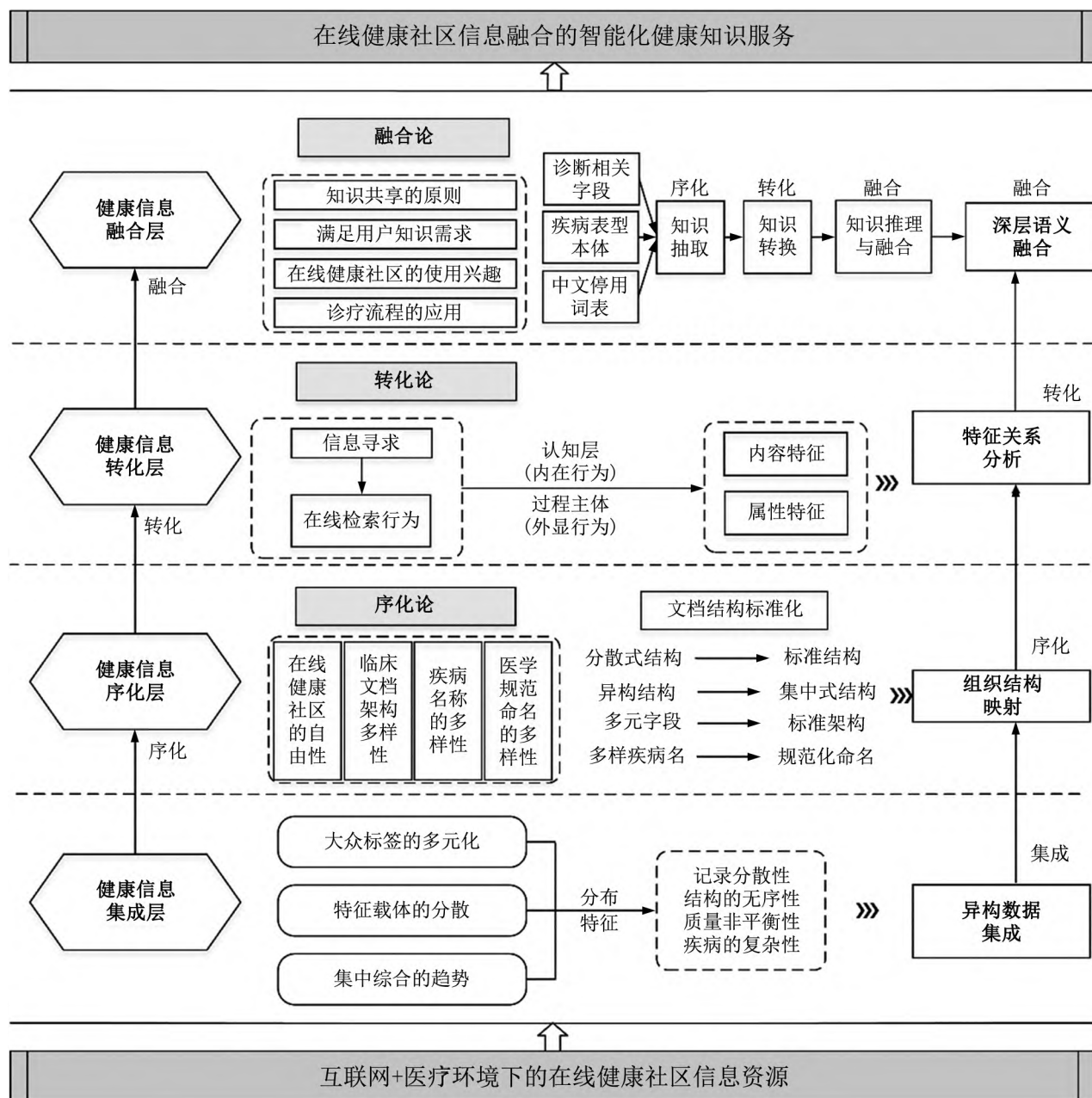


图 1 在线健康社区信息多层级融合模式

1) 在线健康社区信息的集成层

在线健康社区信息的数据汇集集成是打破互联网环境下“信息孤岛”问题的重要途径，为实现对分散在不同位置的同型异源、异质异构、多模态信息的集成整合，就是将不同时间空间的异构信息源、同构信息源进行物理层面的整合，基于某种共性化的特征或主题的精细化组织，将具有不同属性的在线健康社区信息，按主题、时间、模态等外在特征或内容特征进行二次组织、排序聚合奠定基础。在线健康社区信息的集成是推进智能化健康知识服务的数据基础，是建立在互联网环境下在线健

康社区信息破碎化的背景下，以及用户对海量异构健康信息的序化、转化和融合的需求基础上。

2) 在线健康社区信息的序化层

在线健康社区信息的组织结构映射，是实现信息的序化。在线健康社区用户的多样性，以及健康信息的自组织性，造成在线健康社区信息的组织结构和表现形式复杂多样，为数据的进一步分析利用带来了挑战。因此，在线健康社区信息的组织重构，是在线健康社区信息深度挖掘利用的必要环节，以解决在线健康社区信息规范性差的问题。在线健康社区信息组织结构重构的目的是将动态获取

的多源数据,借助一定工具、规则和方法,将数据的结构进行映射组织,形成新的规范化的在线健康社区信息结构,创建新的信息对象。目的是消除分散异构在线健康社区信息冗余和不确定性,解决同源异构、异源异构的在线健康社区信息的结构不规范化,难以进行数据分析利用的问题。在线健康社区信息的组织结构重构,就是要对混乱无序的信息进行序化的组织,是对数据的存储结构、映射过程和映射结果进行规范化。

3) 在线健康社区信息的转化层

在线健康社区信息的特征关系分析,是为了实现信息的转化。充分了解在线健康社区信息的整体特征和局部特征,有助于数据分析者更全面深入地了解数据,进而选取合适的数据分析方法,制定合理的数据分析流程,解决多源异构数据的信息过载问题,也能精准地满足用户个性化的健康知识需求。全面地认识在线健康社区信息的特征,有助于实现知识深层融合过程中具备更好的适应能力和知识筛选能力,推动以问题解决为逻辑起点的在线健康社区信息融合系统的构建。在线健康社区信息的特征体现在多个方面,包括数据的来源多样、质量差异特征、健康话题的主题特征、用户的交互行为特征、空间关系、时序关系和主题关系等。其表现的特征关系是对信息质量的全面性和可信度的评估,根据特征去评估在线健康社区开放、共享的群体智慧信息,能够解决用户健康有关的信息需求。

4) 在线健康社区信息的融合层

在线健康社区信息是在经过异构数据集成、组织结构映射、特征关系分析的基础上,深层的语义融合对信息内容的深度挖掘,解决因信息价值密度低导致的信息降级问题。首先,是根据在线健康社区中多模态信息的特征,将碎片化的网络健康知识进行关联重构、多层次的融合分析,完成多模态在线健康社区信息的底层特征和高层语义内涵之间的映射;然后,根据一定的推理规则对在线健康社区中的领域专家知识进行实体抽取、属性抽取、关系抽取,并进行命名实体的对齐、整合和验证;最后,借助结构化描述体系中的医学领域命名实体的融合、关联关系的融合和语义融合,实现在线健康社区信息的深层语义融合。

3.3 在线健康社区信息融合模式的构成要素

在线健康社区信息融合模式的构成要素包括4类,分别是用户主体要素、数据客体要素、技术方法要素和应用环境要素。主体要素是指在线健康社区中健康信息的生产者、健康知识的传递者、健康知识的消费者;客体要素是指健康信息的内容和载体,包括不同模态的健康科普数据、健康问答数据、临床病例数据等;技术方法要素是在线健康社区信息的存储技术、信息推荐技术、数据挖掘技术和自然语言处理技术等;环境要素包括互联网+医疗政策环境、社会环境、健康知识产生环境、社交媒体传播环境和健康知识应用环境等。在线健康社区信息融合模式的构成要素示意图,如图2所示。

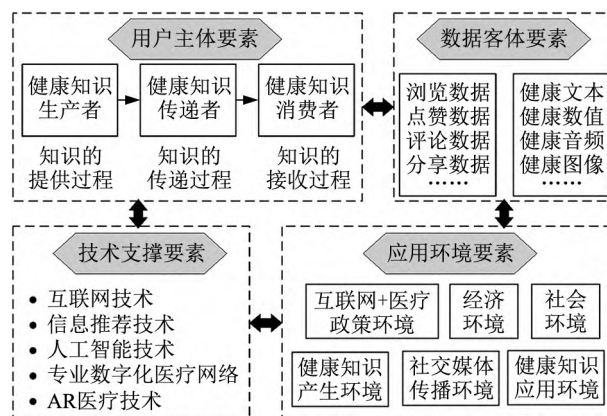


图2 在线健康社区信息融合模式的构成要素

1) 在线健康社区信息融合的用户主体要素

互联网环境下的在线健康社区信息融合的主体要素,是参与健康知识的生产、传递和利用的各类用户,扮演着健康知识的生产者、传递者和消费者角色。用户作为健康知识依附的主体,串联起了在线健康社区信息在不同群体之间的知识提供、知识传播到知识利用的过程,推动健康知识在各个用户主体之间的定向流动。随着用户对健康知识需求的改变,在线健康社区信息的知识流动方向也会不断地发生改变。根据在线健康社区中用户的角色特征,用户可以划分为多种类型:医生、患者、家属等。

2) 在线健康社区信息融合的数据客体要素

在线健康社区信息融合的客体要素是社区中积累的健康相关数据,包括健康相关的疾病症状体征、健康检查结果、疾病名称等个人健康信息,也包括健康信息搜寻行为、健康信息交流互动等社交行为数据。在线健康社区积累的海量信息是为用户

提供健康知识服务的数据基础,支撑着知识服务过程中的用户需求精准定位、知识筛选和协同过滤推荐。通过在线健康社区信息融合的结果,形成的信息流是知识服务活动的动力,链接起健康知识提供者和健康知识消费者。在线健康社区信息融合后形成的知识具有密度和粘性等特征,这些特征是影响用户知识交流感知的动力因素和障碍因素。在线健康社区信息数据类型有多种分类,按照信息的内容可以分为健康科普类信息、个人健康信息、健康信息沟通数据;按照信息的知识服务功能可以分为医学保健常识信息、医学病例讨论信息、医学药物相关信息等;按照信息的来源渠道可以分为医疗机构发布的信息、用户个人发布的健康相关信息等。

3) 在线健康社区信息融合的技术支撑要素

技术要素是在线健康社区信息融合的必要支撑,包括基础支撑技术、信息融合技术和应用技术。①基础支撑技术是搭建各种类型在线健康社区的基石。在线健康社区构建需要网站设计技术,搭建起承载信息的平台,用 MySQL 等数据库构建技术搭建海量健康信息的存储数据库;②信息融合技术是健康知识发现的核心技术,借助于深度学习和自然语言处理等技术,推动多源异构的在线健康社区信息的深度语义融合,以满足用户健康知识服务的需求及数据序化、转化和融合的发展需求;③应用技术是健康信息融合后为用户提供知识服务的关键技术,包括协同过滤推荐技术、人工智能技术和知识图谱技术等。信息推荐技术可以根据用户对健康信息的兴趣度、相关度和意外度进行信息推荐,提高用户对健康知识服务的满意度,并有助于打破信息过载带来的负面影响。

4) 在线健康社区信息融合的应用环境要素

在线健康社区信息融合之后的应用效果,受到在线健康社区的外部环境和内部环境的共同影响。外部环境通常包括“互联网+医疗”政策环境、行业技术环境、社会环境,如在经济环境的影响下,当某一地区的经济条件较差导致医疗卫生资源缺乏时,用户通过在线健康社区获取健康信息的意愿则更强烈,用户对在线健康知识的应用程度则越高;内部环境则包括健康知识产生环境、社交媒体传播环境和健康知识应用环境。在线健康社区信息的内

部环境是影响知识产生、流动和应用的重要因素,由于健康知识单元不都是以游离或孤立的形式存在,而是常常基于关联关系呈现出具有一定特征的团簇性或者群体属性,当有高质量信息提供者和高效的信息传播途径时,在线健康社区信息的应用范围会逐渐扩张,拓展信息融合后的应用环境。在外部环境和内部环境的共同驱动作用下,推动在线健康社区信息的多层级融合,使得为用户提供更高质量的健康知识服务更具有可行性。

4 在线健康社区信息融合利用的阶段划分

在线健康社区信息融合利用的过程,从数据形态的变化和价值密度的提升视角可以划分为 5 个阶段,即用户网络健康知识服务需求定位阶段、多源在线健康社区信息获取阶段、领域健康知识抽取阶段、领域健康知识生成阶段和健康知识融合利用阶段。将为用户提供高质量的健康知识服务作为核心目标,研究定义了在线健康社区信息融合各个过程阶段的作用与重要任务。在线健康社区信息融合的过程阶段,如图 3 所示。

4.1 用户健康知识服务需求定位阶段

随着社会公众生活水平的提高,越来越关注健康相关信息的获取,在线健康社区逐渐成为公众获取在线健康社区信息的主要渠道。虽然公众对健康信息的需求逐渐增高,但当前在线健康社区建设侧重于平台的开发建设,较少考虑用户的健康信息需求的精准性^[33],由此造成用户获取和利用在线健康社区信息的困难和障碍,并由此导致用户信息搜寻过程中产生焦虑情绪^[34]。因此,精准定位用户的健康信息需求和信息搜寻动机,是提升信息融合的知识服务效果的前提。识别出用户的健康信息需求,可以提升健康知识服务结果匹配性、健康知识推荐内容有用性、健康知识推荐过程精准性、健康知识推荐服务效率和健康知识推荐服务满意度。在线健康社区用户的需求可以有多种定位方法,既可以根据用户所处的特定情景(健康保健、慢性病管理、急症咨询等),又可以根据用户的健康信息兴趣(减肥、美容、健身等),还可以根据用户自身的特征(年龄、所属地区范围、职业、性别等)以及在线健康社区信息的主题(孕前检查、母婴护理、婴幼儿看护等)。

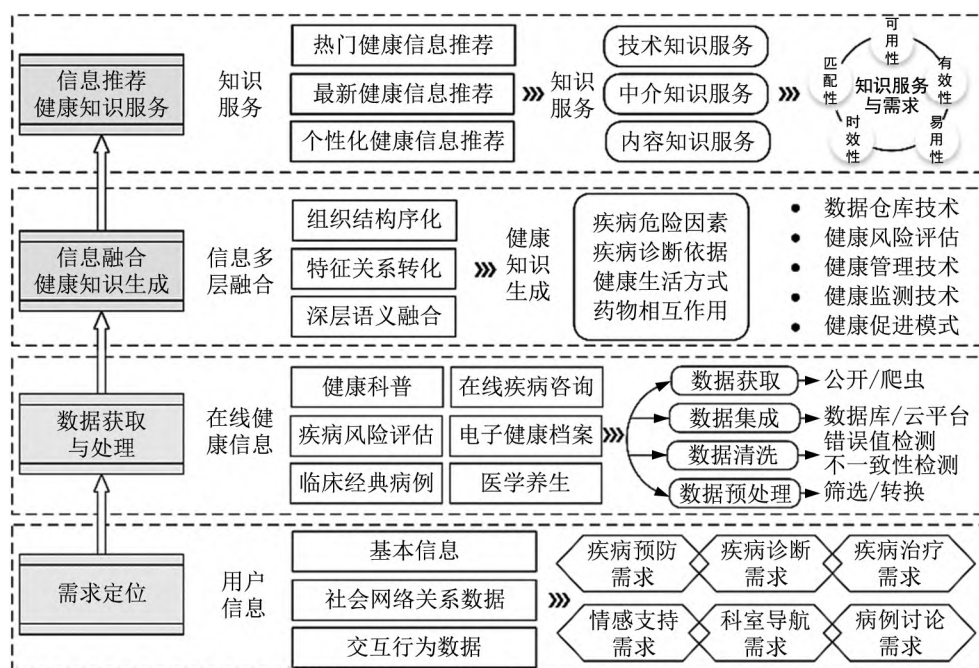


图3 在线健康社区信息融合的阶段

4.2 在线健康社区信息的获取与处理阶段

在线健康社区信息的获取阶段，就是根据用户的健康信息需求采集散落在网络中的各类基础数据，为后期的健康知识服务奠定数据基础。在线健康社区具有来源广且分散异构的特征，储存了用户通过在线健康社区平台进行健康信息交流的静态数据和动态数据。数据采集的过程，首先要确保数据采集的全面性，保持数据采集格式的统一，采用全链路的健康信息采集、校验和统一规则的集成整合，并采集特定的数据指标和元素，如健康问答话题的标题、时间、回答者、提问者等信息，便于后期的数据追踪溯源、实时更新。在线健康社区信息的处理，需要对原始数据进行整合、清洗、特征分析等信息处理工作，实现打通底层健康数据间的壁垒，以提高在线健康社区信息向知识的转化能力。

4.3 多层级融合的健康信息融合阶段

在线健康社区信息多层级融合是以满足用户健康知识服务需求为目标，解决网络健康大数据的组织结构异构、内容价值密度低和知识服务效果差的问题。多层级的融合就是通过对异构的在线健康社区信息结构的映射、知识抽取、知识转换、知识的融合，实现在线健康社区中健康医疗数据中知识的序化、转化和融合。融合后形成的健康知识库，进一步借助人工智能技术，叠加自然语言处理技术和

机器学习方法，实现挖掘其中蕴含的显性和隐性健康知识。在线健康社区信息融合，可以打通底层的多源异构网络健康大数据和用户精准健康知识服务需求，以多源异构的在线健康社区信息驱动精准的健康知识服务。

4.4 健康信息推荐与知识服务阶段

健康信息的推荐与知识服务是在线健康社区融合的最终目标，是数据集成、融合效果的评价。基于在线健康社区信息多层级融合的信息推荐，按照推荐目的和动机可以分为热门健康信息推荐、最新发布的健康资讯、个性化健康信息推荐等。健康信息推荐，使用户获取健康知识能够按照自身的认知进行健康知识体系重构，达到剔除错误的健康知识，吸收并弥补自身健康知识的短板，并运用到日常生活中的健康保健或解决自身实际遇见的健康问题。健康知识服务就是服务者通过用户已经发布的信息需求，选择科学的且符合用户需要的健康知识，定向推荐给用户。用户接收到健康知识后可以根据对知识的有效性、有用性、准确性、匹配性和满意度等进行评价，评价结果对网络健康知识的获取、处理和融合过程提供正反馈和负反馈的指导。

5 在线健康社区信息融合利用案例分析

5.1 案例选择与数据收集

选择北京大学国际医院微信公众平台的 AI 智

能导诊为在线健康信息融合的应用案例,验证在线健康社区信息融合的可行性和现实应用价值。其 AI 智能导诊模块是百度 AI 开放平台开发的智能分诊解决方案的实践应用,被中国医院协会信息专业委员会(China Hospital Information Management Association, CHIMA)作为智慧医院建设的经典案例进行分享^[35]。本文从北京大学国际医院的微信公众号、中国医院协会信息专业委员会、百度 AI 开放平台、媒体报道、期刊论文等途径搜集案例相关数据^[36]。

5.2 案例描述与分析

北京大学国际医院的微信公众平台智能分诊功能,就是通过在线健康平台搜集患者所述症状、基本信息和疾病信息,利用自然语言处理技术和机器学习方法融合所有数据,最终提供最佳的就诊科室,以此减少患者因缺少医学知识而导致的挂号科室错误问题,提高患者就诊的效率,优化就诊流程,减少医疗资源的浪费。基于在线健康信息融合的北京大学国际医院智能分诊过程,如图 4 所示。

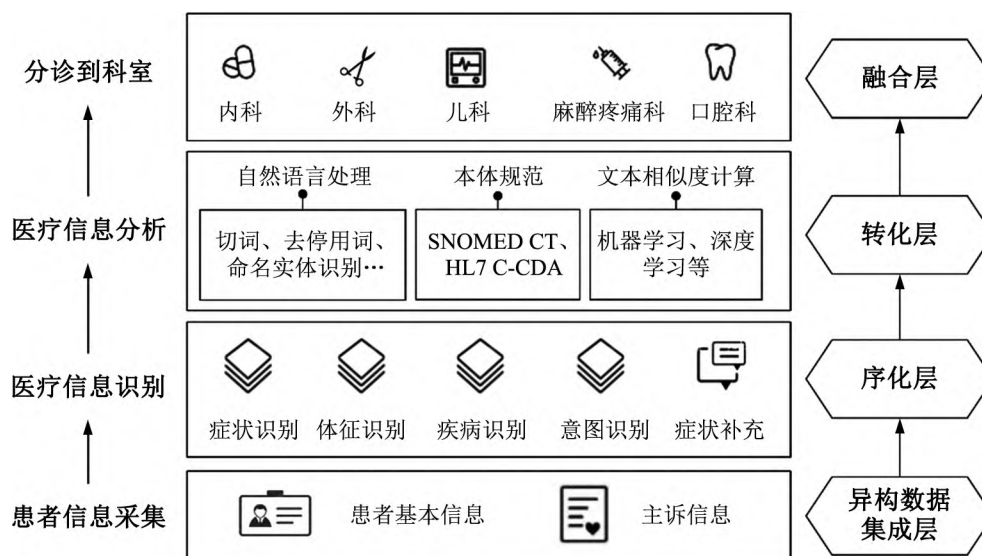


图 4 基于在线健康信息融合的北京大学国际医院智能分诊过程

1) 北京大学国际医院的线上智能分诊的异构信息集成

智能分诊第一步是患者信息的采集,患者输入基本信息和主诉,形成异构数据的语料库。患者的基本信息包括患者的性别、年龄、地址等,患者的主诉信息包括患者的症状、体征,信息的采集过程支持自然语言输入和选项选择两种模式。当患者输入的信息不足以实现分诊到某一科室时,系统会智能地推荐患者可能出现的症状选项,请患者再次确认信息,进行关联症状候补,直至匹配对应的诊疗科室。这一信息采集过程得到的信息是非结构化、口语化且无序的,需要进一步地序化和转化才能实现诊疗科室推荐。

2) 北京大学国际医院的线上智能分诊的信息序化阶段

在线健康平台对患者信息的序化阶段,是利用自然语言处理技术将无序信息有序化的过程。这个

过程首先识别判断患者主诉是否有分诊的需求,然后,再利用分词、切词和去停用词,抽取其中的医疗信息,并将其序化分类为症状类、体征类、疾病类和意图类等。通过序化,使患者输入的信息拆分并归纳入对应的类别,进一步形成可以被机器理解和分析的内容。

3) 北京大学国际医院的线上智能分诊的信息转化阶段

患者信息转化的过程是利用自然语言处理技术,将识别的患者输入的医学术语转化为专业的医学用语,将口语化的医学词汇转化为标准的医学症状、体征、疾病名称等,以提高信息融合的准确率。采用的方法包括切词、去停用词、命名实体识别等,使用的医学标准包括 SNOMED CT、HL7 C-CDA、OMAHA 医学本体等,对健康信息内容进一步规范,并分析信息的内容特征和属性特征,为输入的患者健康相关信息的语义融合奠定基础。

4) 北京大学国际医院的线上智能分诊的信息融合阶段

智能分诊是患者健康相关信息经过序化、转化后进一步融合的结果,该结果能够为患者提供智能化的线上诊疗科室推荐服务。在这一阶段,通过机器学习和深度学习方法对输入数据进行语义层面的深度融合,输出就诊建议,然后将其映射到医院的分科规则,进而匹配到医院的对应科室。匹配科室的基础上,根据医生的学科背景和专长,进一步匹配医生的用户画像,为患者推荐该学科的医生,提高医生和患者的匹配度。

6 结 语

本文根据在线健康社区的信息环境,归纳总结在线健康社区数据特征为其融合利用带来的规范性低、信息过载和信息降级问题。由此,根据情报学研究中的序化论、转化论和融合论,探索在线健康社区信息融合模型的目标和原则,并将在线健康社区融合层级划分为异构数据的集成层、组织结构的映射层、特征关系的分析层和文本语义的融合层面,探究信息融合的模式。在研究创新性地提出面向在线健康社区信息特征的信息融合模式,有助于推动在线健康社区信息向健康知识的转化,为面向用户提供精准高质量的健康知识服务提供一定的参考依据。

但本研究仅从理论层面探究了在线健康社区信息融合模型,尚未进行实践应用效果的实证研究。因此,下一步,本研究将根据构建的在线健康社区信息融合模型进行实证研究,探索融合模型在健康信息推荐和健康知识服务中的应用效果。

参 考 文 献

- [1] 科普中国. 2017 年度中国网民科普需求搜索行为报告 [EB/OL]. https://www.cast.org.cn/art/2017/11/15/art_338_47195.html, 2019-03-23.
- [2] 陆泉,朱安琪,张霁月,等. 中文网络健康社区中的用户信息需求挖掘研究——以求医网肿瘤板块数据为例 [J]. 数据分析与知识发现, 2019, 3 (4): 22-32.
- [3] 李泽中,张海涛,张鑫蕊,等. 融合用户社交与情境信息的虚拟知识社区个性化知识推荐研究 [J]. 情报理论与实践, 2020, 43 (4): 152-158.
- [4] 李岩,郭凤英,翟兴,等. 基于 Jieba 中文分词的在线医疗网站医生画像研究 [J]. 医学信息学杂志, 2020, 41 (7): 14-18.
- [5] 李立萍,吴红,沈丽宁. 在线健康社区中患者寻求第二医疗意见影响因素探析 [J]. 医学信息学杂志, 2020, 41 (7): 9-13.
- [6] 姜婷婷,郭倩,徐亚苹,等. 证据类型对在线健康信息标题选择的影响: 眼动实验与启示 [J]. 图书情报工作, 2020, 64 (19): 61-70.
- [7] 张帅,王文韬,谢阳群. 在线健康社区用户持续参与行为的演化规律及动力机制 [J]. 现代情报, 2021, 41 (5): 59-66.
- [8] 李月琳,张建伟,张姍. 螺旋式与直线式: 在线健康医疗平台用户与医生交互模式研究 [J]. 情报学报, 2021, 40 (1): 88-100.
- [9] Park A, Conway M. Longitudinal Changes in Psychological States in Online Health Community Members: Understanding the Long-term Effects of Participating in an Online Depression Community [J]. Journal of Medical Internet Research, 2017, 19 (3): e71.
- [10] Park K A, Eum S Y, Oh H, et al. Factors Affecting Online Health Community Participation Behavior in Patients with Thyroid Cancer [J]. Plos One, 2020, 15 (6): e0235056.
- [11] 彭昱欣,邓朝华,吴江. 基于社会资本与动机理论的在线健康社区医学专业用户知识共享行为分析 [J]. 数据分析与知识发现, 2019, 3 (4): 63-70.
- [12] Okun S, Goodwin K. Building a Learning Health Community: By the People, for the People [J]. Learning Health Systems, 2017, 1 (3): e10028.
- [13] 尹慧子,张海涛,马婷婷,等. 智慧医疗情境下信息交互行为及拓扑结构研究 [J]. 现代情报, 2020, 40 (3): 137-147.
- [14] Litchman M L, Rothwell E, Edelman L S. The Diabetes Online Community: Older Adults Supporting Self-care Through Peer Health [J]. Patient Education and Counseling, 2018, 101 (3): 518-523.
- [15] 廖开际,黄琼影,席运江. 在线医疗社区问答文本的知识图谱构建研究 [J]. 情报科学, 2021, 39 (3): 51-59, 75.
- [16] Zhang S, Grave E, Sklar E, et al. Longitudinal Analysis of Discussion Topics in an Online Breast Cancer Community Using Convolutional Neural Networks [J]. Journal of Biomedical Informatics, 2017, 69: 1-9.
- [17] 胡德华,张又月,罗爱静. 基于遗传投影寻踪算法的学术虚拟社区知识交流效率研究 [J]. 图书馆论坛, 2019, 39 (4): 67-73, 83.
- [18] 翟姗姗,胡畔,潘英增,等. 融合知识图谱与用户病情画像的在线医疗社区场景化信息推荐研究 [J]. 情报科学, 2021, 39 (5): 97-105.
- [19] Bleiholder J, Naumann F. Data Fusion [J]. ACM Computing Surveys (CSUR), 2009, 41 (1): 1-41.
- [20] Hall D L, Llinas J. An Introduction to Multisensor Data Fusion [J]. Proceedings of the IEEE, 1997, 85 (1): 6-23.
- [21] 李洋,赵鸣,徐梦瑶,等. 多源信息融合技术研究综述 [J]. 智能计算机与应用, 2019, 9 (5): 186-189.

(下转第 167 页)

- 选择及行为路径 [J]. 情报学报, 2021, 40 (1): 77-87.
- [34] 刘艳华, 华薇娜, 钱爱兵. 国际健康信息研究主题的演化路径分析 [J]. 现代情报, 2018, 38 (7): 135-141.
- [35] Chen L, Baird A, Straub D. An Analysis of the Evolving Intellectual Structure of Health Information Systems Research in the Information Systems Discipline [J]. Journal of the Association for Information Systems, 2019, 20 (8): 1023-1074.
- [36] 吴江, 王凯利, 董克, 等. 信息计量领域网络分析方法应用研究综述 [J]. 情报学报, 2021, 40 (10): 1118-1128.
- [37] 钱宇星, 周华阳, 周利琴, 等. 老年在线社区用户健康信息需求挖掘研究 [J]. 现代情报, 2019, 39 (6): 59-69.
- [38] 徐中阳, 石艳霞, 尚珊. “健康中国”背景下的图书馆健康信息服务研究综述 [J]. 图书馆, 2020, (7): 38-44.
- [39] Perez-Dasilva J A, Meso-Ayerdi K, Mendiguren-Galdosin T. Fake News and Coronavirus: Detecting Key Players and Trends Through Analysis of Twitter Conversations [J]. Profesional De La Informacion, 2020, 29 (3): e290308.
- [40] 张洋, 张磊. 网络信息资源评价研究综述 [J]. 中国图书馆学报, 2010, 36 (5): 75-89.
- [41] 成全, 王火秀, 骈文景. 基于证据推理的医疗健康网站信息质量综合评价研究 [J]. 数字图书馆论坛, 2020, (4): 53-59.
- [42] 刘冰, 张文珏. 基于用户视角的网络健康信息服务质量评价体系构建研究 [J]. 情报科学, 2019, 37 (12): 40-46.
- [43] 张馨遥, 沈涌, 张健, 等. 基于模糊综合评价的在线健康信息服务用户满意度分析 [J]. 情报科学, 2018, 36 (6): 73-78, 99.
- [44] Wu D, Xu H, Wang Y, et al. Quality of Government Health Data in Covid-19: Definition and Testing of An Open Government Health Data Quality Evaluation Framework [J]. Library Hi Tech, 2022, 40 (2): 516-534.
- [45] Han J, Kamber M. 数据挖掘概念与技术 [M]. 范明, 等译. 北京: 机械工业出版社, 2012: 1-4.
- [46] 盛姝, 黄奇, 郑姝雅, 等. 在线健康社区中用户画像及主题特征分布下信息需求研究——以医享网结肠癌数据为例 [J]. 情报学报, 2021, 40 (3): 308-320.
- [47] 位志广, 宋小康, 朱庆华, 等. 基于随机森林的健康谣言分享意愿研究 [J]. 现代情报, 2020, 40 (5): 78-87.
- [48] 米国伟, 先祖权, 王琳, 等. 新冠肺炎疫情期间公众心理健康信息需求研究——以社会化问答平台“知乎”为例 [J]. 现代情报, 2021, 41 (6): 108-117.
- [49] Chai Y, Bian Y, Liu H, et al. Glaucoma Diagnosis in The Chinese Context: An Uncertainty Information-Centric Bayesian Deep Learning Model [J]. Information Processing & Management, 2021, 58 (2): 102454.
- [50] Grundmeier R W, Xiao R, Ross R K, et al. Identifying Surgical Site Infections in Electronic Health Data Using Predictive Models [J]. Journal of the American Medical Informatics Association, 2018, 25 (9): 1160-1166.

(责任编辑: 郭沫含)

(上接第 36 页)

- [22] Yan L L, Özsu M T. Conflict Tolerant Queries in AURO-RA [J]. IEEE Computer Society, 1999, (1): 279.
- [23] Naumann F, Bilke A, Bleiholder J, et al. Data Fusion in Three Steps: Resolving Inconsistencies at Schema-, Tuple-, and Value-level [J]. IEEE Data Engineering Bulletin, 2006, 29 (2): 21-31.
- [24] Fisch D, Kalkowski E, Sick B. Knowledge Fusion for Probabilistic Generative Classifiers with Data Mining Applications [J]. IEEE Transactions on Knowledge and Data Engineering, 2014, 26 (3): 652-666.
- [25] 黄晓斌, 张明鑫. 融合多源数据的企业竞争对手画像构建 [J]. 现代情报, 2020, 40 (11): 13-21, 33.
- [26] 顾伟, 李菲菲, 陈虬. 基于多特征融合的行人检测方法 [J]. 电子科技, 2021, 34 (5): 29-34.
- [27] Hall D L, Llinas J. An Introduction to Multisensor Data Fusion [J]. Proceedings of the IEEE, 1997, 85 (1): 6-23.
- [28] Smirnov A, Levashova T, Shilov N. Patterns for Context-based Knowledge Fusion in Decision Support Systems [J]. Information Fusion, 2015, 21: 114-129.
- [29] 唐晓波, 朱娟, 杨丰华. 大数据环境下的知识融合框架模型研究 [J]. 图书馆学研究, 2016, (1): 32-35, 18.
- [30] 周利琴, 范昊, 潘建鹏. 网络大数据中的知识融合框架研究 [J]. 情报杂志, 2018, 37 (1): 145-150, 197.
- [31] 操玉杰, 李纲, 毛进, 等. 大数据环境下面向决策全流程的应急信息融合研究 [J]. 图书情报知识, 2018, (5): 95-104.
- [32] Klapp O E. Opening and Closing in Open Systems [J]. Behavioral Science, 1975, 20 (4): 251-257.
- [33] 王志红. 图情领域聚合概念溯源及信息聚合研究进展 [J]. 图书馆论坛, 2019, 39 (1): 100-110.
- [34] 罗晓兰, 韩景侗, 樊卫国, 等. 互联网时代的健康信息与健康焦虑 [J]. 情报资料工作, 2019, 40 (2): 76-86.
- [35] 杨永燕. 北京大学国际医院智慧医院建设实践和思考 [EB/OL]. <https://chima.org.cn/Html/News/Articles/8858.html>, 2021-12-08.
- [36] 百度 AI 开放平台. 智能分诊解决方案 [EB/OL]. <https://ai.baidu.com/solution/aifenzhen>, 2021-12-08.

(责任编辑: 郭沫含)