

# 基于LDA主题模型的信息服务文献主题 提取与演变研究\*

钱旦敏<sup>1</sup> 郑建明<sup>2</sup>

(1. 南通大学医学院, 南通 226001; 2. 南京大学信息管理学院, 南京 210023)

**摘要:** 运用LDA主题模型梳理近22年来国内信息服务研究主题演化情况, 为该领域的可持续发展提供借鉴和指导。以中国知网 (CNKI) 为信息源, 以信息服务为主题检索CSSCI期刊, 排除不符合论文, 共检索到8 908篇论文, 并按年文献量趋势划分为4个阶段。基于TF-IDF构建LDA主题模型, 利用MindMapper绘制主题演变图, 比较分析信息服务研究主题的持续、弱化、转移与新兴。模型结果显示, 1998年以来我国信息服务领域持续研究主题包括图书馆服务、信息产业服务、资源服务及学科服务; 图书馆服务模式和信息伦理研究逐渐弱化、转移; 移动信息服务、智库大数据、专业领域信息以及图情计量成为近年来该领域研究热点。

**关键词:** 信息服务; LDA主题建模; TF-IDF; 主题演变

**中图分类号:** G203

**DOI:** 10.3772/j.issn.1673-2286.2019.10.003

据《2019年上半年互联网和相关服务业运行情况》报道, 2019年上半年信息服务收入规模达3 703亿元, 同比增长23%, 增速较一季度提高5.8个百分点, 占互联网业务收入比重为68.5%<sup>[1]</sup>。随着信息服务规模增速持续提升, 国家工信部对信息服务业投入大量关注度, 连续发布一系列有关信息安全、信息服务管理、信息服务辅助技术要求、数据集要求等与信息服务相关的政策文件和行业标准。

同时, 随着信息技术的不断发展以及大数据、智能时代的到来, 各行各业都在信息空间留下了“数据脚印”, 信息服务作为各产业的“耳目”“尖兵”“参谋”, 其服务领域不仅涉及图书情报、计算机科学、信息经济、新闻传媒等行业, 而且涉足反恐行动<sup>[2]</sup>、扶贫项目<sup>[3]</sup>等社会研究领域, 引起了不同领域专家和学者的高度关注。

研究信息服务领域的主题演变有利于相关学者与科研人员前瞻、准确地把握前沿领域的竞争态势和重

大趋势, 有效地规划研究主题, 实现科技创新。因此, 本文对信息服务主题演变趋势进行研究。

## 1 主题演变研究简介

主题演变是主题随时间逐渐发生改变的过程, 包括主题的新兴、消亡、强度演化、内容向其他主题迁移或渗透, 以及主题本身的特征研究等。对期刊主题演变展开研究的方法众多, 包括从定性到定量分析, 从简单图表到信息可视化分析, 从基于词频统计到基于模型的机器学习方法等。无论哪种方法, 其主要目的是获取文本集合的关键信息。

国内利用主题模型对信息服务类文献进行演变研究不多。何伟林等<sup>[4]</sup>基于改进的主题模型CSToT, 以国内情报学领域9种期刊作为数据来源, 分析国内情报学领域2012—2016年的研究主题结构以及其演化过程; 熊回香等<sup>[5]</sup>构建基于社会化标签的单用户兴趣模型和

\*本研究得到教育部人文社会科学研究基金项目“新市民公共卫生信息精准化服务模式研究” (编号: 17YJCZH140)、江苏省哲学社会科学基金项目“居民健康信息精准服务模式研究” (编号: 18SHB004)、江苏省高校哲学社会科学基金项目“新型城镇化进程中公共卫生信息精准化服务模式研究” (编号: 2017SJB1211) 和南通大学博士基金“健康管理智慧信息服务模式研究” (编号: 03082161) 资助。

群用户兴趣模型;董克等<sup>[6]</sup>运用主题模型与时间序列分析等方法,以档案学领域的CSSCI来源期刊《档案学通讯》和《档案学研究》刊载的近10年论文为分析对象进行文本内容挖掘。

隐狄利克雷分配模型(Latent Dirichlet Allocation, LDA)在算法复杂度和展示效果方面表现优越,它的参数不会随文档数的增长而线性增长,有很好的泛化能力,常被用来模拟大规模语料的语义信息<sup>[7]</sup>。词频-反文档频率(Term Frequency-Inverse Document Frequency, TF-IDF)是最早的文本降维技术。本文将时间因素引入LDA建模中,采用TF-IDF频率计算方法确定最优主题数目,以信息服务相关文章为实验材料,阐述如何使用LDA主题建模研究某一邻域的主题演变特征。

## 2 基于LDA主题模型的信息服务主题提取与演变方法

### 2.1 数据预处理

本文主要通过主题提取、聚类、主题演变3种技术对信息服务类文献进行研究。本研究以中国知网(CNKI)为数据源,检索以信息服务为主题的CSSCI来源期刊,发表时间设定为全部,截至2019年7月17日,共检索到相关文献9 060篇,其中最早的文献发表于1998年。经人工排检,剔除非正式论文、期刊征稿通知、重复论文等文献,最终得到符合研究的相关文献8 908篇。

由于关键词是直接揭示论文主题的最核心词汇,某一关键词在其所属领域文献中反复出现,则可反映出该关键词所表征的研究主题是该领域的研究热点。因此,本研究提取文献中的关键词作为主要研究数据<sup>[8]</sup>。首先对关键词进行规范化处理,合并同义词。

### 2.2 词频统计及TF-IDF计算

TF-IDF是一种常用于信息处理和数据挖掘的加权技术,主要涉及词频和逆文档频率2个影响因素。词频(Term Frequency, TF)表示某个关键词出现的频率。逆文档频率(Invers Document Frequency, IDF)用于衡量词语普遍重要性,词语出现在越多的文档中,该词语的区分度越低,重要性越低。TF-IDF值越大表示该

关键词对该文献的重要性越大。

本文在数据预处理的基础上,将预处理后的关键词用TF-IDF权重确定关键词的重要程度,为下一步LDA主题模型的聚类提供数据支持。

### 2.3 LDA主题模型的建立

LDA主题模型是一种生成主题概率模型,是一个三层贝叶斯概率模型。LDA使用词袋的思想,对词汇进行模糊聚类,聚集到一类的词可以间接地表示一个隐含的主题。此模型挖掘文本信息,能用来衡量不同文档间的潜在关系,也能通过某一类词来表达文档中隐藏的主题。

本文结合TF-IDF值,利用LDA主题模型对关键词集合进行建模,得出主题-文档的强度矩阵。首先确定LDA主题模型中的超参数 $\alpha$ 、 $\beta$ 以及主题数目 $T$ 的最佳值。结合国内外相关文献知识,本文取值 $\alpha = \frac{50}{T}$ , $\beta = 0.01$ 。利用Python中的LDA工具包对文档数据进行模型训练,选择不同主题数量 $K$ 对不同时间段分别进行多次聚类,对比不同主题数量情况下各主题中关键词间相关性,选择最为合理的主题数进行研究。

## 3 信息服务文献LDA主题模型结果分析

对论文主题演变的研究主要有3种类型:①研究某一主题在不同时间轴上的强度变化;②研究主题相关社会网络,如论文作者或研究机构的变化;③研究主题内容的变化规律,包括主题内容持续、转移、弱化、兴起等现象<sup>[9]</sup>。通过文献梳理发现,该领域的主题研究大多为某一时间段内的主题呈现,而对不同时间段内主题的对比研究相对较少,没有时间段的延续性研究。本文旨在研究信息服务领域期刊文献的研究主题,故选取第3种演变模式进行探讨。

### 3.1 发展阶段划分

本研究通过对研究时间段的切分,从纵向上比较信息服务主题的演变历程。在时间的划分上以每年文献数为依据,将粒度减小,对所有时间段主题演变进行对比分析<sup>[10]</sup>。

将处理后的8 908篇文献按每年检索到的文献数量进行统计,得到近22年被CNKI期刊库收录的CSSCI

来源期刊,且以“信息服务”为主题的文献量统计(见图1)。

可见,“信息服务”主题研究大致可分为4个阶段:1998—2006年是起始阶段,相关主题的文献研究平均数量在350篇/年,呈小幅度波动式上升;2007—2010年属于发展阶段,4年文献数量急速上升,涨幅近1倍;

2010年开始趋向成熟阶段,从2012年开始,期刊文献研究数量逐渐减少,截至2017年文献数量与起始阶段较为相似;考虑到2018年与2019年的文献可能未收录完全,文献实际数量有较大偏差,将这两年的文献单独作为第4阶段进行分析。以起始、发展、成熟阶段为例,聚类结果如表1所示。

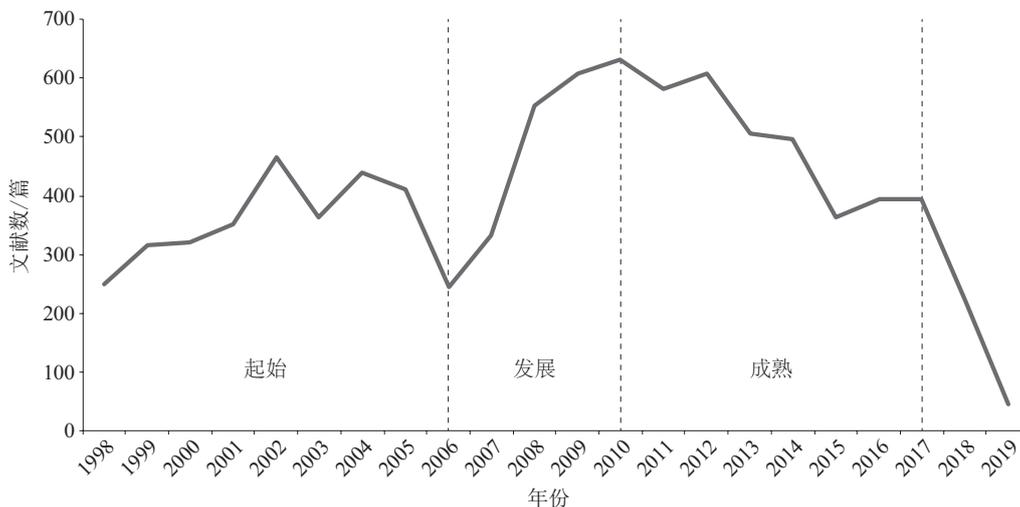


图1 1998—2019年“信息服务”文献量统计

表1 各时间段的高频主题词

	起始阶段 (1998—2006年)	发展阶段 (2007—2010年)	成熟阶段 (2011—2017年)
主题1	图书馆	社区图书馆	智慧图书馆
主题2	公共图书馆	服务模式	信息服务
主题3	MyLibrary	知识服务	服务模式
主题4	信息服务	创新	大数据
主题5	互联网	农村信息化	互联网
主题6	信息资源	信息资源共享	数字参考咨询
主题7	虚拟参考咨询	学科服务	情报学
主题8	个性化	政府信息	学科服务
主题9	信息检索	学科馆员	健康
主题10	知识管理	著作权	智库

务模式和信息伦理;新兴的研究主题有移动信息服务、智库大数据、专业领域信息以及图情计量研究。

### 3.2.1 持续研究主题

(1) 图书馆服务。1998—2006年,图书馆服务的研究内容主要是环境与用户行为对个性化服务的影响,以及个性化服务技术等;研究主题围绕在图书馆建筑和公共图书馆资源两个方面,并有少量关于图书馆个性化服务的研究。随着互联网信息时代的到来,越来越多的学者开始关注数字图书馆的服务研究。2007年以后,互联网信息服务与图书馆服务逐渐融合,研究主题开始转向图书馆信息化服务。

2011—2016年,图书馆服务模式是该领域学者的研究热点,从2017年起逐渐出现智慧图书馆和信息化服务模式两个分支。传统图书馆的服务模式开始转变为数字图书馆和智慧图书馆,着重研究智慧图书馆的服务模式、技术应用,以及智慧图书馆馆员及其发展趋势等。同时,虚拟参考咨询服务伴随图书馆的服务模式改变而发展,它是一个国际性的潮流,是图书馆业务发展到电子网络信息时代的必然产物,标志着图书馆正在利用网上平台拓展自身的专业性服务。

## 3.2 研究主题演变

为更好地观察研究主题的演变,本研究利用MindMapper软件以可视化的形式展示4个时间段的研究主题(见图2)。通过横向与竖向的对比使文献主题演变趋势更加直观、清晰。

可见,1998—2019年,我国信息服务领域的持续研究主题分布在图书馆服务、信息产业服务、资源服务及学科服务4个方面;弱化及转移的研究主题有图书馆服

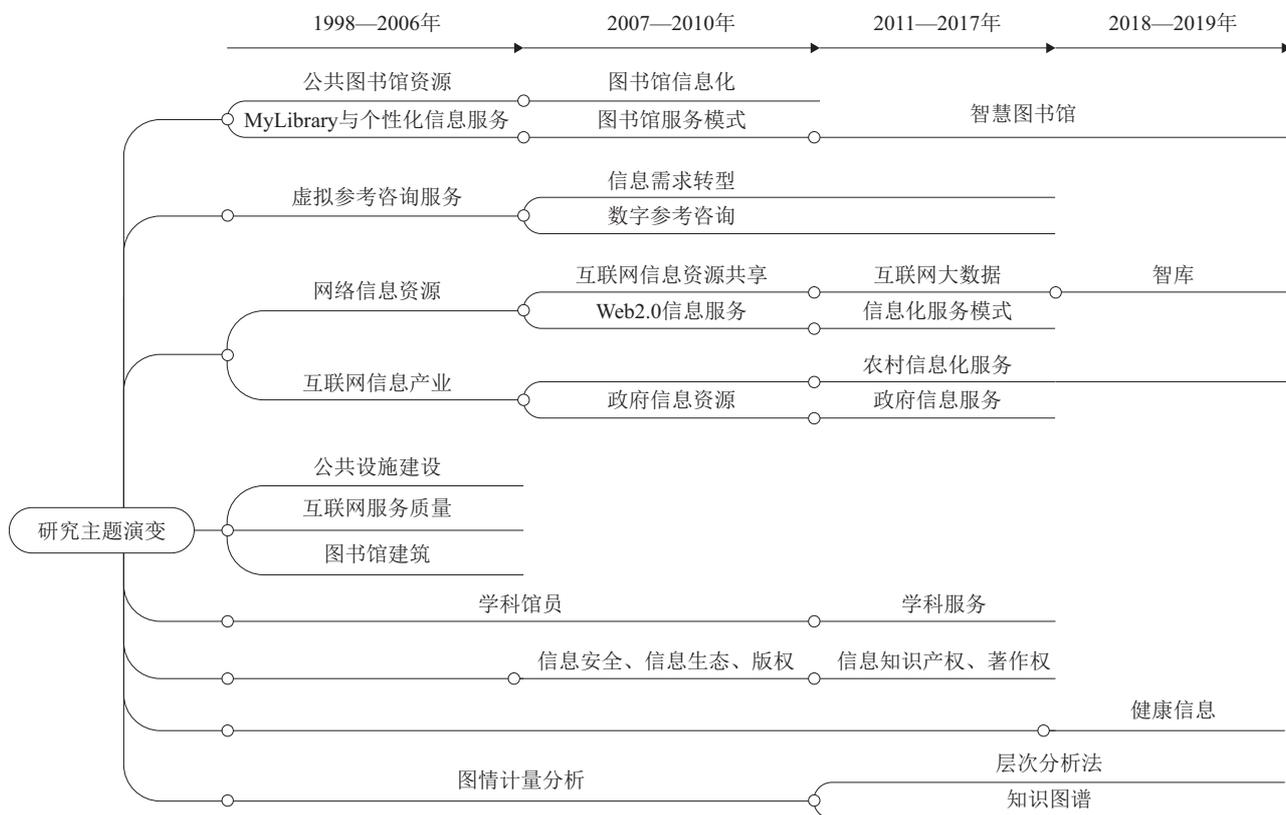


图2 1998—2019年4个阶段研究主题演变图

由以上分析可知,近22年来我国图书馆服务已由传统的图书馆服务转变为互联网环境下的数字图书馆和智慧图书馆服务,且基本围绕服务技术和 service 创新展开相关研究。

(2) 信息产业服务。信息服务以现代信息技术为手段服务于全社会,使人类更及时、有效和充分地利用信息,完善人类社会生活服务。1998—2006年信息产业服务相关研究主要针对图书馆个性化、网络信息资源的研究。2007—2010年,研究开始涉及政府信息资源与信息需求转型,并关注于农村信息化服务的发展。自1980年涉农信息服务政策不断演变<sup>[11]</sup>,尤其2013年,习近平总书记提出了“精准扶贫”重要思想,学界陆续涌现出“精准扶贫需要用好大数据”<sup>[12]</sup>、“农业扶贫重在信息扶贫”<sup>[13]</sup>等思潮,不少学者已经意识到农村信息资源匮乏是我国社会主义新农村发展的巨大瓶颈,因此,越来越多的学者加大了对农业信息化服务的研究。随着云时代的来临,移动互联网、智能硬件和物联网的快速普及,大数据吸引了越来越多学者的关注,“互联网+大数据”逐渐成为研究热点。2011年以后,信息产业与大数据广泛结合,帮助企业、学者更好地适应变化并做出更明智的决策。

(3) 资源服务。1998—2006年,公共资源服务尚未完全普及,政府对图书馆等基础设施的建设考虑较多。由于信息在网络中的流动性强,网络信息资源作为一种新型数字化资源,取代传统的物流方式,加上无线电和卫星通信技术的充分运用,上传到网上的任何信息资源都可以极快地传递到用户手中;另外,由于政府公务处理电子化成为发展趋势,信息资源共享成为政府关注的内容之一,2007年以后,出现了更多关于网络信息资源共享研究和政府信息资源管理相关文献。

(4) 学科服务。以学科馆员为主题开展的学科服务是传统图书馆服务的拓展与深化,是图书馆走向知识化、个性化、泛在化、智能化的必然要求。1998—2019年,学者们对学科服务这一主题的关注只增不减。前12年主要针对学科馆员、信息资源建设及学科服务模式进行研究<sup>[14]</sup>;近9年主要针对学科馆员的技术需求、职业资格做深入研究。“第二代”学科馆员及嵌入式学科服务成为新的研究热点<sup>[15]</sup>。

### 3.2.2 研究主题的弱化及转移

(1) 图书馆服务模式。1998—2006年,我国部分

城市及偏远地区的图书馆覆盖率低,与图书馆有关的主题研究主要在于政府等机构的公共图书馆资源建设与个性化服务,以满足读者的需求。2007年以后,互联网飞速发展,人们获取信息的渠道和阅读方式呈现多样化特点。随着电子阅读日渐普及,传统的实体图书馆利用率下降,以传统图书馆为主题的研究逐渐减少。2017年以后与图书馆有关的研究主题发生转移,建立实体图书馆与互联网虚拟图书馆互联互通、共建共享的图书馆服务体系成为热点,文献研究主题更加关注智慧图书馆及电子阅读的建设与发展。

(2) 信息伦理。信息伦理是指涉及信息开发、信息传播、信息管理和利用等方面的伦理要求、伦理准则、伦理规约,是调整人与人之间以及个人与社会之间信息关系的行为规范的总和<sup>[16]</sup>,与信息生态主题发展关系密切。2007—2010年,信息伦理主题研究关注个人信息道德,如信息知识产权、著作权。2017年以后,信息伦理的关注重点从个人信息道德转向社会信息道德,如信息安全、信息生态、版权。拉斐尔·卡普罗教授<sup>[17]</sup>的《21世纪信息社会的伦理挑战》论述了信息社会的伦理问题,特别讨论了网络环境下的信息伦理问题,将信息伦理学从计算机伦理学中区分出来,反映出信息伦理学理论的发展和变化。

### 3.2.3 研究主题的新兴

(1) 移动信息服务。移动信息服务是从移动用户信息需求出发,利用移动通信技术和移动网络平台开展的信息组织、传递、开发、利用等社会化服务,是对传统信息服务的整合与拓展<sup>[18]</sup>。移动信息服务是现代移动信息技术与用户需求协同驱动的产物。2017年,针对智慧图书馆、移动图书馆及电子阅读的移动信息服务得到大范围推广,相关研究也相继出现,表现出研究主题与领域应用紧密结合的特点。研究重点主要是移动信息服务的理论基础、模式构建及服务质量评价,且多与新媒体环境相关联<sup>[19]</sup>。

(2) 智库大数据。随着云时代的来临,大数据吸引了越来越多科研人员的关注。2011年开始,部分学者的文献主题研究关注于互联网大数据的建设、创新和发展。近年来,大数据产业保持着高速增长态势,从单一的技术概念逐渐转化为新要素、新战略、新思维。大数据与传统智库结合,建设成新型智库。因此,从2017年开始的研究重点趋向于智库的创新建设,助力中国从

“智库大国”向“智库强国”转变<sup>[20]</sup>,大力建设高水平、国际化的智库,逐步为人类创造更多的价值。

(3) 专业领域信息。随着大数据、云计算、移动互联网、人工智能等现代信息技术的发展,各专业领域逐渐向信息化发展。2017年开始,以健康、健康信息为研究主题的文献逐渐增多。健康医疗领域信息化对优化健康医疗资源配置、创新健康医疗服务的内容与形式产生重要影响。研究内容主要包括基于可穿戴医疗设备等收集个人健康数据,基于云平台的分布式存储与并行计算、临床决策和精准医学,这为健康医疗服务推向智能化时代提供了有力支持。

(4) 图情计量研究。通过文献梳理可知,图书情报学计量研究在2011年以前已经存在,但通过对比发现,已有研究中使用聚类分析等方法的文献数量较少,未成体系。2011年以后,计量研究方法用于图情领域的文献越来越多<sup>[21]</sup>;从2017年起使用较多的方法有CiteSpace、VOSviewer、层次分析法、Louvain算法等。计量研究的目的是为分析研究对象领域的现状及发展情况,是领域研究自我检测、自我认知的重要手段。

研究主题的弱化、转移及兴起在各研究领域皆有可能出现,分析其原因有以下两点:一是受社会大环境的影响,如新生的科学技术带动传统产业发展、人们对相关服务的选择发生改变;二是受该研究主题及其所属领域发展情况的影响。弱化的主题并不意味着被学者完全摒弃,而大多是转移或融入其他研究主题中,以适应社会发展及时代需求。因此,研究主题应在不同时期寻找不同的增长点,以保证学科领域的可持续发展。

### 3.3 研究主题未来趋势

根据研究主题演变趋势可知,各领域、各行业都向着大数据、信息化、互联网等方向发展,由此信息服务已向以物联网为核心的智慧服务升级<sup>[22-23]</sup>。它是利用信息技术对传统服务业进行改造提升,同时也包括其他因信息经济发展催生的各类新兴服务业。

图书馆智慧服务、智慧情报服务、政府智慧服务、新型科技智库服务等新兴服务在国内逐步兴起并开展,一方面,通过信息技术推动服务业重构变革与创新,不断孕育、催生服务新模式;另一方面,通过“互联网+”持续推动各行各业服务模式的改造升级,促进产业价值链向高附加值、高技术含量环节攀升,提升政府、图

书馆等部门或机构的智能化服务水平。当前,我国智慧服务业发展在国际上还处于相对较低水平,随着我国在电子政务/电子商务、数据开放、隐私保护、数据安全等领域的不断规范和完善,智慧服务业将蕴藏巨大发展潜力,让“智慧”服务生活,为中国经济实现赶超式、跨越式发展提供核心动力。

## 4 结论

本研究基于CNKI期刊库,使用信息服务相关主题的CSSCI期刊文献数据,将LDA主题模型与TF-IDF结合,利用时间段的连续性得到主题-文档的相关性和强度规律,并按每年期刊文献数量将1998—2019年划分为4个时间段,缩小时间粒度,对现有信息服务文献相关研究主题演变趋势进行探讨,跟踪其研究主题的持续、弱化、转移与新兴。

经分析,本文得出以下研究结论:①1998年以来,我国信息服务领域的热点研究主题主要有图书馆服务、信息产业服务、资源服务及学科服务;②随着时代演变与技术变迁,产生了很多新兴的研究主题,如移动信息服务、智库大数据、专业领域信息以及图情计量研究等;③图书馆服务与互联网一直是研究热点,且发展日趋成熟。信息服务是改造提升传统产业和提高各行各业技术水平、效率、竞争力的重要手段。本文根据实际背景对演化结果进行合理的讨论以正确把握信息服务的研究方向及演变过程,希望提出的信息服务发展脉络对信息服务领域可持续发展提供一定的借鉴和指导。

由于LDA是词袋模型,且是一种双重稀疏模型<sup>[24]</sup>,仅考虑摘要中主题词出现的频率,未考虑主题词权重,因此对于分析结果有影响。后期将进一步优化LDA模型,注重不同主题词的权重和同一主题词在不同语境下内涵的差异,提升信息服务模式主题演变分析结果的准确性。

## 参考文献

- [1] 运行监测协调局. 2019年上半年互联网和相关服务业运行情况 [EB/OL]. [2019-07-30]. <http://www.miit.gov.cn/n1146312/n1146904/n1648355/c7228214/content.html>.
- [2] 李阳. 适应新时代背景的反恐情报信息服务: 体系构建与实施路径 [J]. 图书情报知识, 2019 (4): 62-70.
- [3] 郭蕾, 余波, 张妍妍, 等. 精准扶贫中的信息需求及其服务策略 [J]. 图书馆论坛, 2018, 38 (4): 39-48.
- [4] 何伟林, 奉国和, 谢红玲. 基于CSToT模型的科技文献主题发现与演化研究 [J]. 数据分析与知识发现, 2018, 2 (11): 64-72.
- [5] 熊回香, 杨雪萍, 高连花. 基于用户兴趣主题模型的个性化推荐研究 [J]. 情报学报, 2017, 36 (9): 916-929.
- [6] 董克, 韩宇姝. 基于Topic Model的我国档案学主题结构与演化研究 [J]. 信息资源管理学报, 2017, 7 (3): 97-105.
- [7] 刘自强, 王效岳, 白如江. 语义分类的学科主题演化分析方法研究——以我国图书情报领域大数据研究为例 [J]. 图书情报工作, 2016, 60 (15): 76-85, 93.
- [8] 曲佳彬, 欧石燕. 基于主题过滤与主题关联的学科主题演化分析 [J]. 数据分析与知识发现, 2018, 2 (1): 64-75.
- [9] 李湘东, 张娇, 袁满. 基于LDA模型的科技期刊主题演化研究 [J]. 情报杂志, 2014, 33 (7): 115-121.
- [10] 王婷婷, 王宇, 秦琳杰. 基于动态主题模型的时间窗口划分研究 [J]. 数据分析与知识发现, 2018, 2 (10): 54-64.
- [11] 汪汉清, 樊振佳. 我国涉农信息服务政策核心话语演变分析: 以1980年以来中央“一号文件”为例 [J]. 图书情报工作, 2019, 63 (8): 96-106.
- [12] 石欣鹭, 田晓膺. 精准扶贫需用好大数据 [J]. 人民论坛, 2018 (22): 66-67.
- [13] 张琴. 农业扶贫重在信息扶贫 [J]. 中国国情国力, 2004 (12): 36-37.
- [14] 杨俊丽. 国内学科服务创新发展研究 [J]. 图书馆工作与研究, 2012 (4): 21-24.
- [15] 王焕景, 孙兴敏. 中外图书馆学科服务研究对比分析 [J]. 图书情报工作, 2017, 61 (8): 43-53.
- [16] 孟令权. 我国信息伦理理论研究现状述评 [J]. 图书馆学研究, 2010 (10): 23-27.
- [17] 杨智慧. 规范网络行为的信息伦理 [EB/OL]. [2019-10-22]. 学习时报, [http://www.china.com.cn/xsbs/txt/2004-10/27/content\\_5689552.htm](http://www.china.com.cn/xsbs/txt/2004-10/27/content_5689552.htm).
- [18] 赵杨, 王娟. 基于用户体验的移动信息服务运作机制探究 [J]. 情报资料工作, 2013 (2): 89-93.
- [19] 王博雅, 邓仲华. 移动阅读服务质量测评——基于ANP-模糊综合评价法 [J]. 图书馆论坛, 2018, 38 (2): 71-80.
- [20] 陈潭. 从大数据到大智库: 大数据时代的智库建设 [J]. 中国行政管理, 2017 (12): 42-45.
- [21] 郭卫宁. 国内图书情报学领域知识服务研究脉络探析 [J]. 图书馆学研究, 2016 (4): 2-9.
- [22] 罗立群, 李广建. 智慧情报服务与知识融合 [J]. 情报资料工

- 作, 2019, 40 (2): 87-94.
- [23] 温有奎. 从数字信息服务到智慧服务——以“淘智”为例 [J]. 数字图书馆论坛, 2015 (10): 2-7.
- [24] 张卫卫, 胡亚琦, 翟广宇, 等. 基于LDA模型和Doc2vec的学术摘要聚类方法 [J/OL]. 计算机工程与应用: 1-8 [2019-10-24]. <http://kns.cnki.net/kcms/detail/11.2127.TP.20191010.1344.010.html>.

## 作者简介

钱旦敏, 女, 1981年生, 博士, 副教授, 研究方向: 数据建模, E-mail: qdm11@163.com。  
郑建明, 男, 1960年生, 博士, 教授, 研究方向: 信息资源管理。

Research on Subject Extraction and Evolution of Information Service Literature Based on LDA Topic Model

QIAN DanMin<sup>1</sup> ZHENG JianMing<sup>2</sup>

( 1. Department of Medical Informatics of Medical School, Nantong University, Nantong 226001, China;  
2. School of Information Management, Nanjing University, Nanjing 210023, China )

Abstract: LDA topic model is used to sort out the evolution of domestic research subject of information service, and provide reference and guidance for sustainable development in this field. Using the CNKI journal database as the information source, the CSSCI journals are searched for the topic of information service, and the non-conformity papers are excluded. A total of 8 908 papers are retrieved and divided into four stages according to the annual literature volume trend. Building an LDA topic model based on TF-IDF, MindMapper is used to draw the subject evolution map to compare and analyze the persistence, weakening, transfer and emerging of the research topic. The results of the model show that since 1998, the main research topics in the field of information services in China include library services, information industry services, resource services and subject services; library service models and information ethics research have gradually weakened and shifted; mobile information services, think tanks data, professional field information, and graph measurement have become hotspots in this field in recent years.

Keywords: Information Service; LDA Topic Modeling; TF-IDF; Theme Evolution

(收稿日期: 2019-10-08)