

基于LDA-BERT模型的智慧图书馆主题演化研究

严素梅 海骏林峰

(华东理工大学科技信息研究所, 上海 200237)

摘要: 近年来, 图书馆正不断朝“智慧化”方向迈进, 智慧图书馆也逐渐走进研究者的视野。本文将LDA主题模型和BERT模型相结合, 构建主题相似性测度模型, 量化智慧图书馆领域主题之间的关联关系, 并以此为基础绘制主题演化图谱, 分析中外智慧图书馆领域研究主题的演化趋势, 探究中外研究的异同。基于主题演化图谱可以看出, 智慧图书馆领域的研究主要聚焦于智慧图书馆理论研究、智慧图书馆服务、智慧图书馆技术与实践研究、智慧图书馆管理、图书馆资源获取与知识产权等5个领域。其中, 中文研究侧重智慧图书馆服务; 外文研究则侧重图书馆的资源获取与管理。

关键词: 智慧图书馆; LDA-BERT模型; 主题演化; 关联分析

中图分类号: G353.1 **DOI:** 10.3772/j.issn.1673-2286.2023.02.008

引文格式: 严素梅, 海骏林峰. 基于LDA-BERT模型的智慧图书馆主题演化研究[J]. 数字图书馆论坛, 2023 (2) : 57-64.

党的二十大报告指出, 文化建设和社会建设是中国特色社会主义新时代的重要建设方向。图书馆是文化传播和交流的重要场所, 同时也是社会建设体系的重要环节。随着物联网和人工智能等信息技术的不断发展和普及, 现代技术与传统行业之间的结合现象愈发普遍, 传统图书馆正逐步向智慧图书馆转型^[1], 面向智慧图书馆展开的相关研究也逐年增多。作为当前智能化的信息交流中心, 智慧图书馆在我国文化建设和社会建设中有着十分重要的地位。本文旨在对智慧图书馆领域的相关研究内容进行分析 and 探讨。

1 相关研究

图书馆界对“智慧图书馆”的概念界定普遍基于Aittola等^[2]在2003年提出的“不受时空限制, 能够感知的移动图书馆服务”。早在2000年, 张洁等^[3]就提出了“智能图书馆”的概念, 将其描述成高新技术与建筑

艺术的有机结合, 但并未引起研究者的广泛关注。2010年, 严栋^[4]首次在国内提出“智慧图书馆”的概念, 将智慧图书馆等同于“图书馆+物联网+云计算+智慧化设备”, 随后引发国内学者研究智慧图书馆的热潮。随着概念的深化, 智慧图书馆的内涵已经在最初“移动图书馆服务”的基础上有了更多的拓展和延伸。2015年, 郑恽昕等^[5]从概念范畴、技术进展和应用实践等方面对国内外智慧图书馆研究现状和进展进行归纳总结。2018年, Zimmerman等^[6]对智慧图书馆领域出现的定性研究进行综合分析, 提出了智慧图书馆的建设方向。2020年, 赵丽梅等^[7]在全词共现网络分析的基础上进行文献计量, 揭示了我国智慧图书馆领域的研究热点, 并分析其内容结构和发展趋势。2022年, 胡媛等^[8]使用CiteSpace软件对智慧图书馆领域的研究文献进行计量和可视化分析, 揭示并分析了国内外智慧图书馆研究热点。

现有针对智慧图书馆领域的研究主要涉及概念界定、使用可视化软件进行关键词聚类、基于关键词词频

收稿日期: 2023-01-07

统计或词共现分析的热点识别等。针对该领域主题识别的相关研究多以关键词为切入点,未涉及文献的标题、摘要等更细粒度的文本特征,不能充分表征文本的语义。本文拟基于文献的标题、摘要和关键词字段,抽取特征词后进行主题挖掘,并结合深度学习方法,使用Python语言构建相似性测度模型,分析主题之间的关联关系,同时实现主题的识别和主题关联关系的量化测度。基于主题相似性测度结果,绘制智慧图书馆领域的主题演化知识图谱,并从主题内容、主题强度、主题演化和中外研究关联等方面对中外智慧图书馆领域的研究内容及其主题演化趋势展开分析,以期为我国智慧图书馆领域的研究提供科学的量化依据。

2 研究方法与数据获取

2.1 研究方法

进行主题演化研究时,首先要解决的问题是主题相似性的测度。LDA (Latent Dirichlet Allocation) 模型^[9]是当前自然语言处理领域挖掘文本主题最常用的方法,其通过吉布斯采样 (Gibbs Sampling) 算法识别大规模文本中潜在的主题信息,同时可以较好处理一词多义的情况。在使用LDA模型挖掘文本主题后,得到主题对应的主题词概率矩阵以及文献对应的主题概率矩阵。通过分析挖掘出的主题,结合对主题所属重点文献 (高被引量) 的人工判读,命名的主题可以更

全面地表征主题所包含文献的研究内容。但LDA模型在挖掘主题过程中使用的词袋方法在分析文献内容时会忽略词之间的关系,无法结合语境输出文档主题。

BERT (Bidirectional Encoder Representations from Transformers) 模型^[10]的出现为自然语言处理任务提供了一个泛用型很强的预训练模型。同时,BERT模型可以很好体现词汇与上下文之间的关系,是现阶段自然语言处理领域性能强大的深度学习模型之一。当前已有将LDA与BERT模型结合运用于情感主题分类^[11]、文本主题聚类^[12]的相关研究,这些研究证明通过引入BERT模型,可以有效提升LDA模型主题抽取的效果,同时可以弥补LDA模型由于运用词袋方法导致的语义缺失问题。

因此,本文构建了基于LDA-BERT模型的相似性测度方法:在LDA模型的基础上嵌入BERT模型,对挖掘出的主题进行向量化处理,使LDA模型挖掘出的主题充分考虑词汇与上下文之间的语义关系。在向量表示的基础上进行相似性测度可以更好地体现出主题之间的关联关系。

本文的研究框架如图1所示,共分为三步。第一步是文本预处理。从CNKI和Web of Science数据库获取国内外智慧图书馆领域的相关文献,并结合TF-IDF和Textrank方法对文献的标题和摘要字段进行特征词提取。将提取出的特征词与关键词字段合并去重,再使用自定义词典对合并后的结果去除停用词并合并同义词,剔除不相关词的干扰。第二步是相似性测度。用LDA模

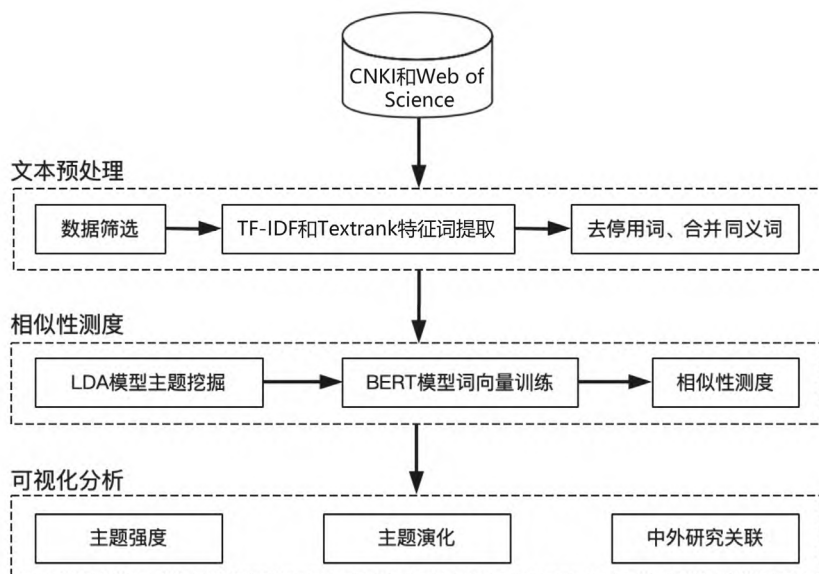


图1 研究框架图

型挖掘中外智慧图书馆领域的研究主题, 将BERT模型嵌入LDA主题模型进行词向量转换, 使主题表示为实数值的向量形式, 最后基于向量使用余弦相似度算法计算主题之间的相似度。第三步是可视化分析。采用可视化方法从主题强度、主题演化、中外研究关联三个方面展开分析, 将LDA模型输出的概率作为主题向量的权重, 对数据集进行向量化表示, 计算不同阶段数据集之间的关联, 并使用可视化方法分析中外智慧图书馆领域研究内容的关联关系。

2.2 数据获取

分别选择CNKI数据库、Web of Science核心合集对智慧图书馆领域发表的中文和外文文献进行检索。

智慧图书馆的研究以智能图书馆为起点。因此, 中文检索式为“TKA=(智慧图书馆+智能图书馆) AND SU=(智慧图书馆+智能图书馆)”, 限制来源数据库为北大核心和CSSCI, 限制学科为图书情报与数字图书馆, 限制文献类型为学术期刊。由于国内智慧图书馆概念首次提出为2010年, 故选择2010年之后发表的文献进行分析, 设定检索时间为2010年1月1日—2022年9月30日, 对文献被引量数据进行人工补充。经人工筛选, 得到666条中文文献。

外文检索式为“TS=((smart* AND librar*) OR (intelligent* AND librar*) OR (intellectual* AND librar*) OR smartlibrar* OR smarterlibrar*)”, 限制来源数据库为SCI-Expanded和SSCI, 限制研究方向为“Information Science Library Science”, 限制文献类型为论文。由于Aittola于2003年首次提出智慧图书馆的概念, 故外文文献选择2003年及以后发表的文献进行分析, 设定检索时间为2003年1月1日—2022年9月30日。在外文检索过程中, 如果将智慧图书馆限制为词组, 检索结果较少, 存在漏检的情况, 因此放宽了检索条件, 以词形式进行检索。经人工筛选, 得到569条外文文献。

3 主题演化分析

由于外文文献的发文量相对较低, 在多次测试主题抽取结果后, 以5年为时间节点, 按时间段分为4个数据集(ET1、ET2、ET3、ET4)进行主题演化分析。此外, 有学者将国内智慧图书馆的发展分为四个阶段^[8]:

2010—2013年为萌芽阶段; 2013—2016年为初步发展阶段; 2016—2019年为快速发展阶段; 2019—2020年为平稳发展阶段。笔者借鉴这一阶段划分, 并考虑国内智慧图书馆领域发文数量, 经多次测试主题抽取结果后, 以3年为时间节点, 按时间段分为4个数据集(CT1、CT2、CT3、CT4)进行主题演化分析(总年数为13, 故CT1的时间段只能是4年)。中外文献每个数据集所对应的文献量如表1所示。

表1 中外文献数据集划分及发文量统计

发展阶段	外文数据集	文献量/篇	中文数据集	文献量/篇
萌芽阶段	ET1 (2003—2007)	95	CT1 (2010—2013)	45
初步发展阶段	ET2 (2008—2012)	114	CT2 (2014—2016)	94
快速发展阶段	ET3 (2013—2017)	165	CT3 (2017—2019)	203
平稳发展阶段	ET4 (2018—2022)	195	CT4 (2020—2022)	324

3.1 主题挖掘

在进行主题挖掘之前, 通过计算语义一致性指标获得不同数据集的最优主题数。语义一致性是一种可以捕捉到模型中语义信息的评价方法。通过建立词袋, 基于单词的共现统计度量, 有效地识别部分低质量词汇, 提高主题识别的质量。随着主题数增加, 一致性结果呈现波动上升趋势。语义一致性在一定范围内最大时, 确定的主题数挖掘出的主题可以较好地表征数据集的主题。经计算, 数据集对应不同主题数时的语义一致性结果如图2所示。

折线图中不同样式的线段对应不同的数据集, 横坐标表示主题数, 纵坐标表示语义一致性结果。根据计算结果, 中文数据集最优主题数分别为8个、6个、8个、6个; 外文数据集最优主题数分别为5个、8个、7个、5个。其中, 数据集CT1的文献量较小, 当主题数为4时, 主题的表达效果更好, 故选择主题数4对数据集CT1进行主题挖掘。主题挖掘结果如表2所示。

结合主题挖掘结果, 可以将中外智慧图书馆领域的研究归纳为5个方向: 智慧图书馆理论研究、智慧图书馆服务、智慧图书馆技术与实践研究、智慧图书馆管理、图书馆资源获取与知识产权。

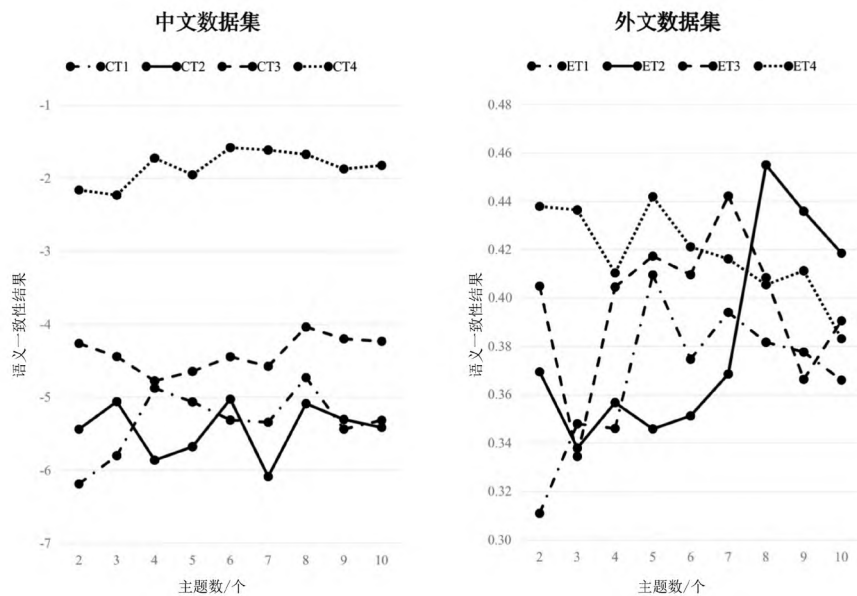


图2 中外文献数据集“主题-语义”一致性折线图

表2 中外智慧图书馆领域主题分布

文献类型	数据集	时间跨度/年	主题数/个	主题标签
中文	CT1	2010—2013	4	智能信息技术服务、数字图书馆、物联网服务转型、智慧服务理论研究
	CT2	2014—2016	6	个性化智慧服务、数字图书馆读者服务、高校图书馆服务管理、物联网技术服务、信息系统与数据服务、智慧馆员与图书馆服务
	CT3	2017—2019	8	图书馆技术服务管理、人工智能服务技术、智慧图书馆理论研究、图书馆读者服务、数据技术与智慧服务、读者智慧服务、高校图书馆科研服务、智慧服务与知识融合
	CT4	2020—2022	6	读者服务技术创新、智慧服务、图书馆技术服务平台、数据服务技术、智慧图书馆文献综述、数字服务体系
外文	ET1	2003—2007	5	图书馆理论研究、数字图书馆知识产权、图书馆学术访问、图书馆信息共享、图书馆数字资源管理
	ET2	2008—2012	8	图书馆与信息科学、文献计量分析、德尔菲法、互联网数字图书馆、信息科学应用研究、目录与标签分类、数字知识产权、数据索引与知识组织
	ET3	2013—2017	7	图书馆资源获取、图书馆学理论研究、信息科学文献计量分析、人工智能图书馆、图书馆数据管理、图书馆学比较研究、移动图书馆
	ET4	2018—2022	5	图书馆电子资源获取、知识管理文献计量分析、数字图书馆信息管理、图书馆技术服务、图书馆管理

3.2 主题强度分析

通过LDA主题模型输出文献的“文档-主题”概率矩阵，结合文献对应主题的发文量以及被引次数表征学科主题强度；结合主题篇均被引量指标，关注主题随着时间推移的关注度以及影响力的变化。绘制中外智慧图书馆领域研究主题强度和篇均被引量气泡图，如图3所示。

气泡图中不同纹理图案的气泡分别代表不同的数据集，横坐标表示主题，主题顺序与表2中外智慧图书馆领域主题标签的顺序相同；纵坐标表示主题强度/篇均被引量，主题强度/篇均被引量越大，气泡面积越大。主题强度的计算值为相对值，即每一个数据集中所有主题的主题强度之和相等，数据集中的主题越少，主题强度分布越集中；主题越多，主题强度分布越分散。篇均被引量与文献被引量直接相关，其计算结果为绝对值。

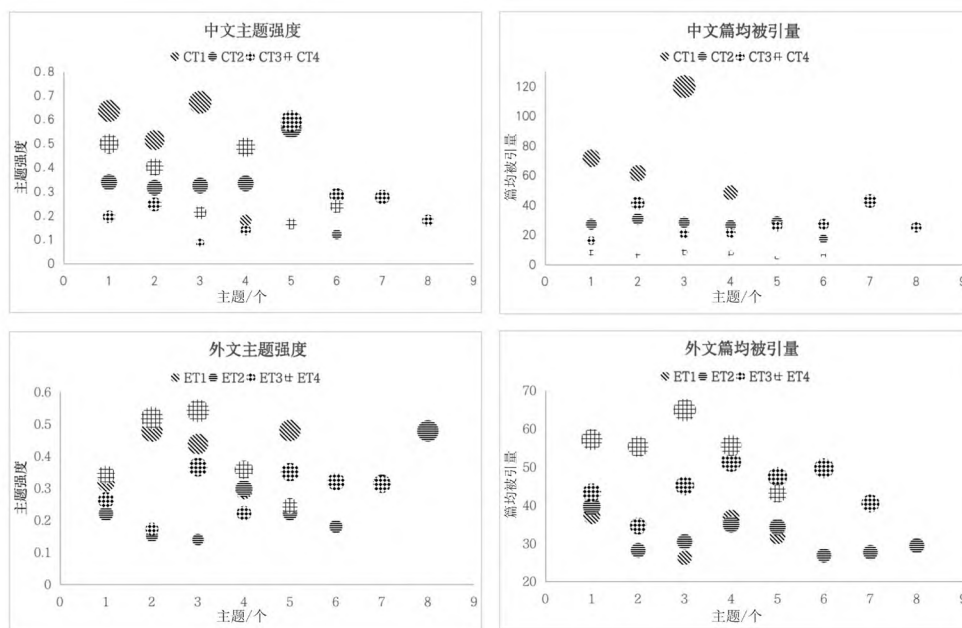


图3 中外文献主题强度及篇均被引量气泡图

从中文主题强度和篇均被引量气泡图中可以看出主题CT1-1(智能信息技术服务)、CT1-3(物联网服务转型)、CT2-5(信息系统与数据服务)、CT3-5(数据技术与智慧服务)、CT4-1(读者服务技术创新)、CT4-4(数据服务技术)相较于同数据集中的其他主题,主题强度和篇均被引量相对突出。通过分析归属该主题的文献可知,热度较高的文献均与信息技术和数据服务相关,这与智慧图书馆的理念相同,即通过信息技术为用户提供智慧化服务,以提升用户在图书馆中的体验。外文主题强度和篇均被引量气泡图中,主题ET1-2(数字图书馆知识产权)、ET1-5(图书馆数字资源管理)、ET2-8(数据索引与知识组织)、ET3-3(信息科学文献计量分析)、ET3-5(图书馆数据管理)、ET4-2(知识管理文献计量分析)、ET4-3(数字图书馆信息管理)相较于数据集中的其他主题,主题强度和篇均被引量较为突出。通过分析归属该主题的文献可知,外文研究更倾向于数字资源的管理,涵盖知识产权、信息管理、数字资源等细分内容,说明国外智慧图书馆领域更关注信息资源所有者的权益。

从整体上看,外文研究主题的篇均被引量趋势与国内相反,由远及近呈现增长的趋势,2018—2022年的篇均被引量相对较高。由此可见国外在进行智慧图书馆领域的研究时,更倾向关注最新的研究成果,因此知识概念的更新迭代速度更快。

3.3 主题演化分析

对计算出的相邻数据集间的主题相似度矩阵进行归一化处理,将筛选结果导入Origin软件,绘制智慧图书馆领域中外文献研究主题的演化路径桑基图,如图4、图5所示。

两张图直观地展示了4个数据集之间主题的“流动”情况,即智慧图书馆领域文献的四个研究阶段研究主题的发展趋势与走向。可以看出中文研究中数字图书馆、智慧服务、智能信息技术等主题,外文研究中知识产权、数字资源管理等主题存在继承、分化和合并的情况,并以此构成整个智慧图书馆领域的主题演化路径。从整体上看,智慧图书馆领域文献研究主题的演化路

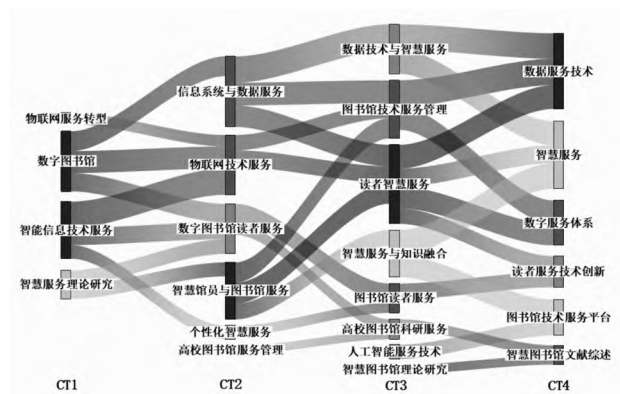


图4 中文研究主题演化图谱



图5 外文研究主题演化图谱

径相对复杂，多个主题之间存在关联，研究主题相互融合，从时序上表现为研究主题由少变多，并呈现出收紧的趋势。

结合图4中的信息分析可知，由于智慧图书馆在国内发展的时间相对较短，中文文献的研究内容在演化过程中并未出现消亡的主题；新生主题有高校图书馆服务管理、人工智能服务技术、智慧图书馆理论研究等。其中高校图书馆服务管理的出现，体现出随着智慧图书馆概念的提出，研究内容从最初的技术研究逐渐扩展到智慧图书馆的管理，通过提升图书馆的智能化管理水平增强智慧图书馆馆员的职业成就感、职业自豪感、职业安全感和职业自信心，实现智慧图书馆的创新发展和可持续发展^[13]。人工智能服务技术主题的出现代表着智慧图书馆的应用技术从最初的物联网、RFID逐渐扩展到人工智能领域，也意味着智慧图书馆领域研究内容更加广泛，应用技术更加多样化。

相较于国内智慧图书馆领域的研究，国外研究历程相对较长，图5中出现了图书馆学理论研究、图书馆资源获取两个消亡主题。通过查看主题包含的文献，消亡主题的研究内容多与传统图书馆相关，即传统图书馆相关的理论研究及图书馆的纸质馆藏资源获取。由此可以看出，面向传统图书馆的研究已经发展到较为成熟的阶段，学者的研究重点也已转移到新兴的智慧图书馆领域。

3.4 中外研究关联分析

基于主题挖掘的结果，将主题强度作为权重赋予主题，得到每个数据集基于主题的向量表示。以此为基础计算中外不同研究阶段之间的关联，绘制内容关联桑基图（见图6）。

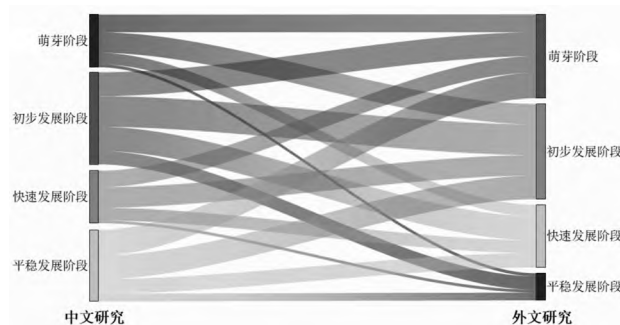


图6 中外研究关联

图中连线的宽度体现不同数据集之间的关联程度，关联性越高连线越宽。从时间维度上看，中文研究相较于外文研究存在一定的滞后性，其中中文研究在平稳发展阶段的研究内容表现尤为明显，与外文研究的萌芽阶段和初步发展阶段关联性更强，与外文研究的快速发展阶段和平稳发展阶段关联性则相对较低。

从图中还可以看出，中外研究均是初步发展阶段的研究内容与其他阶段的研究内容关联性更强。在初步发展阶段，智慧图书馆属于新兴领域，此时的研究重在探索，主要探究各类信息技术如何与传统图书馆相结合。因此该阶段的研究范围相对较广，存在关联的可能性也更大。随着研究的不断深入，对于智慧图书馆的概念和定义逐渐明晰，研究内容也会从广度研究转向深度、细化的研究，此时的研究从探寻如何将信息技术应用于图书馆转变为将某一特定技术与图书馆建设深入结合，以提升图书馆服务的智慧化水平为目标。

4 结论

本文基于LDA-BERT主题相似性测度模型，对智慧图书馆领域的中外文献进行主题挖掘，并通过主题强度分析、主题演化分析及热点主题识别对智慧图书馆领域的研究内容进行归纳总结，得到以下结论。

（1）从主题挖掘的结果来看，智慧图书馆领域的研究主题共分为5个方向，但中外文献研究的侧重点不同。中文研究偏向于将信息技术运用于提升图书馆的智慧化服务水平，以提高用户的服务体验；外文研究偏向于运用信息技术提升图书馆获取信息资源及管理信息资源的能力，从信息资源方面提升用户在图书馆的体验。

（2）从不同发展阶段的主题强度来看，主题强度较高的中文研究多与技术服务和数据服务相关，侧重点依然是运用信息技术提升图书馆的服务水平；外文研究

的侧重点则是图书馆的信息管理和知识产权服务,旨在维护用户信息产出方面的权益。这与主题挖掘的结果类似。

(3) 从主题演化的趋势来看,中外研究主题数量呈现先增多后减少的趋势。智慧图书馆领域在初步发展阶段更注重广度的研究,随着研究不断深入,研究内容不断细分,逐渐转向面向某一技术的深度应用研究。中文研究目前还处于由广度向深度不断探索和转变的阶段;外文研究已步入深度应用研究阶段,研究内容相对固定。

(4) 从研究内容的关联来看,国内智慧图书馆领域研究起步较晚,导致中文研究在发展前期存在一定程度的滞后性。但现阶段国内对于智慧图书馆领域的关注度更高,在公开发表的外文文献中,国内学者发表的文献数量呈现上升趋势。通过对中外文献的研究分析发现,国内对于新兴技术在智慧图书馆领域的应用研究范围更广,国内学者正积极地探索适合我国智慧图书馆建设的发展道路。由此可见,国内智慧图书馆领域还有很大的发展空间。

根据挖掘出的主题,结合研究主题的强度及演化趋势,从各个主题的主题词及主题内容深度探究,概括中外智慧图书馆领域五大方向的研究呈如下发展态势。

(1) 智慧图书馆理论研究。此方向的文献主要研究集中于智慧图书馆的概念界定、基于已有研究的文献计量等方面。智慧图书馆概念在2003年被提出后,国外便开始出现智慧图书馆领域的相关研究。2009年IBM提出智慧地球概念,智慧图书馆领域的关注度也有所上升。随着智慧图书馆领域文献的增加,基于文献计量的综述类文章逐渐增多,通过对智慧图书馆相关研究的总结与分析,围绕智慧图书馆的未来发展方向以及研究热点进行讨论。国内智慧图书馆领域兴起后,又陆续出现一些学者对中外智慧图书馆领域的比较研究;随着智慧图书馆概念的完善,智慧图书馆领域的细分研究增多,基于计量学的研究文献也逐步扩展到细分领域。

(2) 智慧图书馆服务。从总体上看,中文文献更多集中于智慧图书馆服务,外文文献中对于此方向的研究则相对较少。智慧图书馆概念的提出,引起众多学者以各类技术为基础,以给读者提供个性化、定制化的服务为目标展开研究。通过精准定位,挖掘用户的需求,满足用户的需要,使用户在接受图书馆服务的过程中有更好的体验。乌恩^[14]在知识服务的基础上提出了图书馆

服务的第四个阶段——智慧服务,认为智慧图书馆的服务模式是基于图书馆馆员智慧的知识服务。随着研究的不断深入,智慧服务所涉及的技术领域不断扩展,面向对象也更加丰富。

(3) 智慧图书馆技术与实践研究。智慧图书馆作为图书馆与信息技术结合的交叉学科,即采用各类现代化技术来提升图书馆的智慧化水平。其中,涉及的技术包括但不限于:物联网、RFID、云计算、5G、人工智能、区块链等。通过此类技术与图书馆的不同功能结合,可以为用户提供更好的服务体验。此方向的文献多从理论层面出发,分析不同技术在图书馆各类功能中的适用性,并结合实践对图书馆功能进行更新升级,来验证技术结合的可靠性与实用性。中外文献在该方向的研究方法大致相同,但从主题分类中可以看出,外文研究中信息科学出现的频率较高,这与国外的学科划分相关。图书馆与信息科学是美国高校自20世纪20年代起开设的硕士专业,在不断演变过程中,倾向于数据分析、计算机技术在图书馆领域的应用研究。

(4) 智慧图书馆管理。智慧图书馆的管理不再局限于纸质资源管理和用户管理,此阶段在传统管理的基础上,重点面向数字资源的管理、智慧图书馆应用技术管理以及智慧馆员的管理。中文文献中该方向的研究多集中于智慧馆员的管理。通过培养智慧馆员,提升智慧馆员的能力,使馆员可以更好地借助智能技术面向用户提供智慧服务。外文文献中的管理则更倾向于面向图书馆数据资源的管理。如何防止数据泄露、高效地处理数据以便图书馆更好地提供数据服务是国外智慧图书馆领域研究的重点。

(5) 图书馆资源获取与知识产权。该研究方向是外文研究关注的重点,中文研究相对较少。资源获取与知识产权是图书馆领域一直关注的问题,而资源获取也是用户在图书馆的首要目的。2003年,Dilevko等^[15]通过文本挖掘技术,从知识获取角度识别图书馆资源库中非英文小说的关键词等,协助图书管理员主动向用户推荐携带书评的书籍的详细信息。发展到智慧图书馆阶段,向用户提供信息的技术已相对成熟,图书馆已有资源的管理成为研究重点。信息的指数化增长,使数字资源的存储形式逐渐多样化,伴随而来的是资源泄露风险的增加及知识产权纠纷的相关问题。因此,如何规范资源获取的途径,防止信息泄露导致知识产权纠纷是当前国外智慧图书馆领域关注的重点。

本文基于LDA-BERT模型对中外智慧图书馆领域

的文献进行分析,从主题强度、主题演化趋势及中外研究关联等方面进行分析,以期为我国未来的智慧图书馆建设提供理论参考。但本文还存在一些不足,如在进行文献检索时将研究范围限定在核心期刊,可能无法完全代表智慧图书馆领域的所有研究,希望能在后续的研究过程中完善。

参考文献

- [1] 谭春辉,周一夫.近十年我国智慧图书馆研究态势[J].图书馆论坛,2022,42(1):69-79.
- [2] AITTOLA M, RYHÄNEN T, OJALA T. SmartLibrary - Location-aware mobile library service [C] // 5th International Symposium on Human-Computer Interaction with Mobile Devices and Services Mobile HCI 2003, 2003: 411-416.
- [3] 张洁,李瑾.智能图书馆[J].图书馆理论与实践,2000(6):12-13,31.
- [4] 严栋.基于物联网的智慧图书馆[J].图书馆学刊,2010,32(7):8-10.
- [5] 郑恽昕,包平.智慧图书馆理论与实践进展研究[J].图书馆工作与研究,2015(7):36-39.
- [6] ZIMMERMAN T, CHANG H C. Getting smarter: definition, scope, and implications of smart libraries [C] // Proceedings of the 18th ACM/IEEE Joint Conference on Digital Libraries, 2018: 403-404.
- [7] 赵丽梅,廖文杰.基于全词共现网络的我国智慧图书馆研究热点及内容分析[J].情报科学,2020,38(10):23-28.
- [8] 胡媛,邹小敏,朱益平.国内外智慧图书馆研究热点与主题分析[J].图书馆理论与实践,2022(5):77-84.
- [9] BLEI D M, NG A Y, JORDAN M I. Latent Dirichlet Allocation [J]. Journal of Machine Learning Research, 2003, 3(4/5):993-1022.
- [10] DEVLIN J, CHANG M W, LEE K, et al. BERT: Pre-training of deep bidirectional transformers for language understanding [C] // Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2019: 4171-4186.
- [11] 丁申宇.基于BERT-LDA的在线评论细粒度情感分析[D].兰州:兰州财经大学,2022.
- [12] 庄穆妮,李勇,谭旭,等.基于BERT-LDA模型的新冠肺炎疫情网络舆情演化仿真[J].系统仿真学报,2021,33(1):24-36.
- [13] 王瑞珍.智慧图书馆员情绪调查与管理策略分析[J].图书馆工作与研究,2014(5):61-64.
- [14] 乌恩.智慧图书馆及其服务模式的构建[J].情报资料工作,2012(5):102-104.
- [15] DILEVSKO J, DALI K. Electronic databases for readers' advisory services and intellectual access to translated fiction not originally written in English [J]. Library Resources and Technical Services, 2003, 47(3):80-95.

作者简介

严素梅,女,1975年生,硕士,副研究馆员,研究方向:文献计量、创新评价、竞争情报,E-mail:smyan@ecust.edu.cn。
海骏林峰,男,1998年生,硕士,研究方向:数据分析、知识发现。

Research on Theme Evolution of Smart Library Based on LDA-BERT Model

YAN SuMei HAI JunLinFeng

(Institute of Science and Technology Information, East China University of Science and Technology, Shanghai 200237, P. R. China)

Abstract: In recent years, the library is constantly moving towards the "intelligent", the smart library has gradually come into the vision of researchers. This paper combines the LDA topic model and BERT model, builds a topic similarity measurement model, quantifies the correlation between topics in the field of smart library, and draws the topic evolution map based on this, analyzes the evolution trend of topics in the field of smart library at home and abroad, and explores the similarities and differences between Chinese and foreign studies. Based on the theme evolution graph, we can see that the research in the field of smart library mainly focuses on the theory research of smart library, smart library service, smart library technology and practice research, smart library management, library resource acquisition and intellectual property rights. Among them, Chinese research focuses on the smart library service; Foreign language studies focus on the acquisition and management of library resources.

Keywords: Smart Library; LDA - BERT Model; Theme Evolution; Analysis of Association

(责任编辑:王玮)