

# 我国图书馆学领域核心知识元及其组合的新颖度和贡献度分析<sup>\*</sup>

邵 阳

**摘 要** 文章以 6 种图书馆学核心期刊为数据源,抽取其 2012—2021 年刊载的代表性论文,分析论文中研究问题、研究方法的表述特点,构建问题、方法抽取规则,建立学科领域知识元词典;再抽取 2019—2021 年样本刊刊载的全部论文的问题、方法主题,通过编程建立问题、方法及其组合的共现关系,并测度分析问题、方法及其组合的新颖度与贡献度间的关系;最后抽取 2022 年至 2023 年 10 月样本刊刊载的相关论文,计算测度 2019—2021 年图书馆学领域研究问题、研究方法及其组合的新颖度在此阶段与贡献度之间的关系,以探究各类问题、方法及其组合的演化发展。研究发现,从问题、方法及其组合的知识元角度进行分析的方法能够有效发现学科领域的新问题、新方法,老问题、老方法,以及问题、方法的多种组合形式;各类问题、方法及其组合的演化发展为分析学科领域研究态势提供了良好的背景知识。

**关键词** 知识元;组合创新;新颖度;贡献度;图书馆学

**分类号** G250.1;G252.8

**本文引用格式**

邵阳.我国图书馆学领域核心知识元及其组合的新颖度和贡献度分析[J].图书馆工作与研究,2024(3):73-83.

## 1 引言

研究问题和研究方法是学术论文中相辅相成的两类核心知识元。研究问题指科学研究的对象,体现了研究的内容和目标;研究方法指解决问题的手段、途径和规则,体现了研究的层次和深度。科学研究是不断提出问题和解决问题的过程,研究问题与研究方法是学术进步和学科发展的内生动力,对学科发展发挥引导作用。因此,分析学科领域研究问题、研究方法(下文简称问题、方法)及其组合的研究现状,能够从知识元角度揭示学科领域的发展态势,不仅有利于科研人员识别科研创新机

会,而且有利于辅助科研管理部门把握科技发展趋势,提早科研布局,抢占创新先机。

19 世纪末 20 世纪初图书馆学从西方传入我国<sup>[1]</sup>。目前,该学科领域对问题与方法的研究均已形成一定体系。本研究以我国图书馆学核心期刊论文为数据源,在抽取问题(P)、方法(M)、问题—方法(P—M)的基础上,分析其研究的贡献度和新颖度,从论文知识元视角理清图书馆学研究问题与研究方法的构成,以期科研选题和科研决策提供参考。

## 2 相关研究概述

目前,学科领域研究现状分析已取得一定

<sup>\*</sup> 本文系 2021 年度黑龙江省省属高等学校基本科研业务费科研项目“新媒体全息技术推动高校深层学习的阅读路径探究”(项目编号:145109230)研究成果之一。

收稿日期:2023-03-29

本文编校:谷俊娟

成果,形成了一定的研究范式。笔者在中国知网进行文献检索发现,与学科领域研究态势分析相关的文献涉及图书情报档案、环境科学、经济学、计算机科学、医学等30多个学科。其研究范式主要为以学术文献、基金项目、科技发展规划文件等为依托,通过文献计量<sup>[2]</sup>、引文分析<sup>[3]</sup>、共现分析<sup>[4]</sup>、聚类分析<sup>[5]</sup>等方法揭示学科领域的研究现状、挖掘研究前沿热点、发现知识关联、预测未来研究方向、识别核心研究人员、评估学科影响力<sup>[6]</sup>等,并利用知识图谱展示其研究态势<sup>[7]</sup>。这种传统的研究范式仅通过文献的外在信息揭示学科发展态势,尚未深入文献的内容信息,因而研究结论偏差性较大。随着文本挖掘技术的成熟,相关研究逐渐深入至文本内容,利用主题模型抽取不同时间窗内文献的研究主题<sup>[8]</sup>、利用语义相似度挖掘相邻时间段的关联主题<sup>[9-11]</sup>。主题模型法将每篇文档视为由多个主题组成的概率分布,每个主题由一组词项的概率分布表示,利用文本中词项的共现特征发现文本的主题结构,进而发现领域研究态势。但由于缺乏领域词典或领域词典内容过于宽泛,导致利用主题模型法抽取的主题内容不够明确,从而影响了分析结果的精准性。

我国图书馆学研究态势相关研究文献主要针对图书馆学领域的某一子问题展开,如图书馆知识服务现状<sup>[12]</sup>、智慧图书馆研究态势<sup>[13]</sup>、古籍保护公众宣传现状<sup>[14]</sup>、阅读推广研究现状<sup>[15]</sup>、数据素养教育现状<sup>[16]</sup>、图书馆资源共享现状<sup>[17]</sup>等。也有研究关注图书馆学研究方法的识别、分类及应用现状<sup>[18]33-47</sup>、分析图书馆学研究方法面临的困境及出路<sup>[19]</sup>、构建图书馆学方法体系<sup>[20-21]</sup>、创新图书馆学研究方法论<sup>[22]</sup>等。但缺乏从P、M两类知识元角度,尤其是P—M组合的角度分析学科领域研究态势的成果。鉴于此,本研究基于P、M知识元抽取,通过计算P、M及P—M组合的新颖度和贡献度分析图书馆学领域的研究现状,为学科领域研究态势分析提供新方法和新视角。

### 3 研究方法 with 计算指标

本研究按照“制定知识元抽取规则→构建知识元词典→抽取P、M知识元→建立P—M共现关系→计算知识元新颖度和贡献度→分析领域研究现状”的步骤展开。

#### 3.1 知识元抽取

知识元抽取是一种自然语言处理方法,是从信息源中识别、理解、筛选、归纳隐性知识的过程。目前,知识元抽取方法主要包括4种:①基于人工标注的方法<sup>[23]</sup>。该方法根据领域专家建立的文本语义标注规范为文本添加语义标签,经多轮交叉验证与修正获得标注语料。该方法对领域专家知识依赖性较强,缺乏大规模标注的语料库,因此,难以准确抽取知识元。②基于规则的方法<sup>[24]</sup>。该方法根据被抽取的对象及其上下文的模式特征或语法特征,借助句式结构、框架或符号的共性特征等建立规则表达式,以识别和抽取文本中的知识元。基于规则的方法可以提供清晰、易理解的特定域的语言,不足在于规则有限导致抽取结果不完整。③基于机器学习的方法<sup>[25]</sup>。该方法是在先学习一组数据样本的特征基础上,对另外一组数据进行抽取。机器学习方法往往需要大量的训练语料,且模型训练依赖大规模的标注语料,其鲁棒性受到考验。④基于深度学习的方法<sup>[26]</sup>。该方法对技术要求较高,目前的抽取方法尚存在建模过程中会遗失特征信息的问题。针对小文本量的处理,本研究提出基于规则与主题模型相结合的知识元抽取方法。具体步骤如下:

##### (1) 制定知识元抽取规则

采用等距抽样方法,从代表性期刊中选取每年度的第一期作为样本,分析所载论文摘要中P、M的表述特点。为提高规则的包容性,选择最大长度匹配模式,人工制定P、M的抽取规则。

##### (2) 构建知识元库

以学科领域对应的大量学术论文为数据源,将全部规则和文献摘要读入自编的 Python 程序。首先,识别文本中的标点符号,对所有摘要进行分句、断句处理;其次,命令每一条规则遍历每一个切分后的短句,通过字符串匹配的方式从文本中抽取 P、M 短语;再次,根据 P、M 的表述特点限制知识元字符串长度,自动抽取 P、M 知识元;最后,构建学科领域的专用 P 知识元库和 M 知识元库。

### (3) 借助 LDA 模型实现知识元抽取

由于文献摘要中 P、M 知识元表述方式多样,加之受人工制定规则的制约,基于规则的知识元抽取方法可能导致大量 P、M 知识元被遗漏。LDA 模型能够基于词典、利用概率方法生成与主题相关的 top-n 个主题词。本研究基于 LDA 模型的主题抽取原理,分别以构建的 P 知识元库和 M 知识元库为词典,利用 LDA 模型抽取文献摘要中的主题。该做法不仅能够确保得到的主题词是学科领域中 P、M 的集合,而且能够识别某摘要中不满足抽取规则的 P、M 知识元,从而有效弥补仅通过规则抽取方法可能遗漏部分知识元的缺陷。

## 3.2 贡献度和新颖度指标

### (1) 贡献度

贡献度最早是经济学领域经济效益的分析指标之一,指有效或有用成果数量与资源消耗及占用量之比,常用产出量(所得量)占投入量(消耗量、占用量)的比例表示,以度量目标对象的作用和影响力。本研究中的贡献度特指 P、M 知识元及其组合在学科领域中所占的比重,其表达式为:

$$P_{cd} = \frac{10 \times n(P_i)}{\sum n(P_i)} \quad (1)$$

$$M_{cd} = \frac{10 \times n(M_j)}{\sum n(M)} \quad (2)$$

$$(P-M)_{cd} = \frac{10 \times n(P-M)_k}{\sum n(P-M)} \quad (3)$$

在公式(1)–(3)中, $P_{cd}$ 、 $M_{cd}$ 和 $(P-M)_{cd}$ 分别为 P、M 和 P—M 的贡献度, $n(P_i)$ 、 $n(M_j)$ 、 $n(P-M)_k$ 分别为第 i 个 P、第 j 个 M 和第 k 对

P—M 出现的次数, $\sum n(P_i)$ 、 $\sum n(M)$ 、 $\sum n(P-M)$ 分别为某一时间窗口学科领域中相关 P、M 和 P—M 的出现次数。分子扩大 10 倍是为了避免贡献度过低不利于数据处理而设置的参数。贡献度反映当前 P、M 及 P—M 组合在学科领域中的作用和影响力,可通过贡献度分析透视学科领域 P、M 及 P—M 组合的研究热度。

### (2) 新颖度

新颖度强调学术研究的“新”,指前所未有的新出现的研究,其大小主要取决于研究最早出现的时间。P、M 及 P—M 组合出现的时间越晚,相关研究越新。因此,参考相关研究<sup>[27-28]</sup>,给出学科领域 P、M 及 P—M 组合的新颖度计算方法:

$$P_{Nd} = \frac{1}{(P_t - P_{t_0})} \quad (4)$$

$$M_{Nd} = \frac{1}{(M_t - M_{t_0})} \quad (5)$$

$$(P-M)_{Nd} = \frac{1}{[(P-M)_t - (P-M)_{t_0}]} \quad (6)$$

公式(4)–(6)中, $P_{Nd}$ 、 $M_{Nd}$ 、 $(P-M)_{Nd}$ 分别为知识元 P、M 和 P—M 的新颖度,t 为知识元的统计年(2022 年), $t_0$ 为知识元的首次出现年。新颖度能够反映 P、M 及 P—M 组合的创新性,通过新颖度分析能够发现学科领域 P、M 及 P—M 组合的创新研究机会。

## 4 我国图书馆学领域研究态势分析

### 4.1 数据来源

学科领域核心期刊刊载的论文质量普遍较高,能够代表学科的发展水平。另外,核心期刊论文格式规范,摘要中对 P、M 的表述明确,便于识别。因此,本研究选取图书馆学领域 6 种 CSSCI 来源期刊(《中国图书馆学报》《大学图书馆学报》《图书馆论坛》《图书馆杂志》《国家图书馆学刊》《图书馆建设》)为数据源,分别以期刊名作为“文献来源”,通过中国知网获取研究数据,并以 Endnote 格式提取期刊文献摘要。提取的摘要分为 3 类:①第一类

摘要,即6种样本刊2012—2021年每年度第一期刊载论文的摘要,用于制定知识元抽取规则;②第二类摘要,即6种样本刊2012—2021年刊载的所有论文的摘要,用于构建知识元库;③第三类摘要,即6种样本刊2019—2021年刊载的所有论文的摘要,用于抽取近年图书馆学领域的P、M知识元。

#### 4.2 知识元抽取

##### (1)构建知识元库

以第一类摘要为数据样本,根据各期刊P、M的表述特点,人工构建P、M抽取规则。以第二类摘要为数据源,自编Python程序,读入摘要和P、M描述规则:首先,识别标点符号,对所有摘要进行分句、断句处理;其次,运行每一条规则,遍历每一个经切分处理的摘要;再次,根据P、M的句法表达特点,将P对应的字符串位限定为5—12个字,将M对应的字符串位限定为4—6个字,运行程序得到P、M知识元;最后,针对机器抽取结果进行人工修正,去除无用词,形成初始P库和M库。知识经济时代,科学研究成果日益丰硕,学科领域的P、M数量不断增加,构建相对完备的学科领域知识元库需要整合学科领域多源数据中的P、M。本研究从博客和学术论坛中通过人工识别、筛选学科领域前沿P与M,进一步丰富P库和M库。

##### (2)抽取知识元

以第三类摘要为数据源,分别以构建的P库、M库为词典,利用LDA模型从摘要中抽取主题,得到图书馆学领域的P、M集合;经去重处理得到论文—P、论文—M的关联关系;根据P、M知识元的共现关系获得P—M组合。为避免因P—M组合数过多干扰后续分析,笔者参照吴丹等<sup>[18]36</sup>给出的方法词分类方案,将抽取到的P、M分别做归类处理,如将信息素养、数据素养、环境素养、媒介素养、职业素养、学术出版素养统一表示为素养研究;将SPSS、回归分析、因子分析、相关分析、时序分析、主成分分析等统一表示为统计分析法等。最终得到222对P—M组合、38个P和31个M。

#### 4.3 P、M及P—M组合的贡献度与新颖度

在中国知网数据库的文献分类目录下,选取“图书馆学、图书馆事业”学科类别,针对共现的P—M组合在主题下逐一检索,检索式为“问题”对应的词汇并“方法”对应的词汇,选取中文文献,根据“计量可视化分析”的结果,记录P—M共现的发文量(即共现次数)和首次共现年,并按发文量降序排序,结果如表1所示:

表1 P—M组合对应的发文量及最早发文时间(部分)

编号	问题	方法	发文量 (篇)	首次 共现年
1	图书馆服务	模型法	2199	1993
2	图书馆服务	问卷调查	1098	1978
3	阅读推广	问卷调查	1091	2009
4	知识组织与管理	问卷调查	902	2001
5	素养研究	问卷调查	694	2006
6	素养研究	调查研究法	660	2006
7	知识组织与管理	统计分析法	625	1983
8	图书馆建设	问卷调查	578	1981
9	用户研究	问卷调查	558	1995
10	图书馆空间	文献研究法	491	1987
11	公共图书馆	问卷调查	483	1991
12	图书馆服务	统计分析法	376	1985
13	数据开放共享	统计分析法	369	1991
14	用户研究	统计分析法	360	1987
15	图书馆服务	调查研究法	343	1982
16	素养研究	内容分析法	314	2005
17	图书馆建设	比较分析法	278	1976
18	图书馆服务平台	调查研究法	274	2003
19	科研合作与评价	关联分析法	260	1997
20	用户研究	比较分析法	257	1991

根据公式(1)—(6),结合表1中的P—M共现关系,计算P、M、P—M组合的新颖度和贡献度,并按新颖度降序排序,结果如表2、表3所示。

表2 P、M新颖度与贡献度(部分)

问题	问题 新颖度	问题 贡献度	方法	方法 新颖度	方法 贡献度
数字人文	0.100	0.003	行动者网络理论	0.143	0.000
智慧图书馆	0.083	0.007	用户画像	0.125	0.011

表 2 (续)

问题	问题 新颖度	问题 贡献度	方法	方法 新颖度	方法 贡献度
智库	0.077	0.001	情感分析	0.111	0.003
云计算	0.077	0.010	扎根理论	0.077	0.009
阅读推广	0.063	0.046	关联技术	0.059	0.011
公共文化建设	0.063	0.003	政策工具理论	0.056	0.001
机构知识库	0.053	0.005	共词分析	0.050	0.006
群体阅读研究	0.050	0.001	流程分析	0.048	0.002
隐私保护	0.048	0.002	数据挖掘	0.043	0.057
数字资源建设	0.045	0.006	定位技术	0.042	0.003
校企合作	0.045	0.001	关联分析	0.042	0.006
图书馆服务平台	0.043	0.020	机器学习	0.038	0.009
元数据研究	0.042	0.016	访谈	0.037	0.025
图书馆转型	0.036	0.002	词频分析	0.036	0.005
图书馆营销	0.034	0.005	本体	0.034	0.038

表 3 P—M 组合的新颖度和贡献度(部分)

P—M 组合	P—M 组合 新颖度	P—M 组合 贡献度
数字人文—行动者网络理论	0.500	0.001
智慧图书馆—扎根理论	0.500	0.002
信息素养—用户画像	0.333	0.005
公共文化建设—用户画像	0.333	0.005
数字人文—比较分析法	0.333	0.004
智库—访谈	0.333	0.004
数字人文—理性思辨法	0.333	0.001
公共文化建设—分类法	0.333	0.001
数字人文—访谈	0.250	0.004
数字人文—文献研究法	0.250	0.003
图书馆建设—用户画像	0.250	0.003
公共文化建设—机器学习	0.250	0.002
图书馆建设—扎根理论	0.250	0.002
文化保护与传承—关联分析法	0.250	0.001
图书馆服务—用户画像	0.200	0.066
数字人文—内容分析法	0.200	0.005
数字人文—政策工具理论	0.200	0.003
智库—理性思辨法	0.200	0.001

4.4 P、M 及 P—M 组合新颖度和贡献度分析

4.4.1 P 的新颖度和贡献度

根据二八关键少数法则,将 38 个研究问题按新颖度高低近似分为 9 个新问题和 29 个老问题。

新问题及其首篇论文发表年为:数字人文(2012 年)、智慧图书馆(2010 年)、智库(2009

年)、云计算(2009 年)、阅读推广(2006 年)、公共文化建设(2006 年)、群体阅读研究(2006 年)、机构知识库(2003 年)、隐私保护(2001 年)。不难发现,新问题多为数字化技术和数字化环境的产物,数字化技术为图书馆资源建设提供了机遇,学者们重点关注特色资源、古籍资源的收集与整理以及典藏标准构建、数据保护、数据库建设与利用等问题。网络技术、人工智能技术、机器学习技术等为图书馆高质量服务与智能化建设提供了强劲动力,学者们逐渐关注智慧图书馆、智库、分布式计算等问题,相关研究主要包括建设方案与策略、建设标准与要求、智能化环境下图书馆员的职责、服务模式等主题。新媒体环境为阅读推广活动注入新活力,学者们针对不同类型的读者,从阅读内容的生产形式、实体形式、传播形式等方面展开研究。新技术环境为图书馆学研究带来机遇,针对新问题的研究方法亦逐渐增多,如针对数字人文的研究方法包括比较分析、理性思辨、行动者网络理论、文献研究、统计分析、政策工具理论、评价研究、关联分析、案例分析等。但新问题的贡献度不高,由表 2 可知,新颖度为 0.1 的“数字人文”(研究问题的新颖度最大值为 0.1),其贡献度很低,仅为 0.003(贡献度最小值为 0.001),“智慧图书馆”研究的新颖度为 0.083,贡献度仅为 0.007,说明针对新问题的研究方法虽多,但总体研究量并不大。

典型的老问题及其首篇论文发表年为:图书馆服务(1957 年)、古籍研究(1957 年)、图书馆建设(1978 年)、阅读研究(1981 年)、用户研究(1981 年)、需求研究(1981 年)、图书馆史(1981 年)、图书馆空间(1982 年)、社区图书馆(1982 年)、知识传播与流动(1983 年)等,其中贡献度较高的老问题是图书馆服务、图书馆建设等。服务是图书馆的工作职责,也是其核心价值的体现,因此,图书馆服务是研究者持续关注的核心问题。作为软文化载体,图书馆普遍注重加强自身建设,包括服务资源建设、馆

员队伍建设、服务和管理系统建设、馆舍空间和环境建设等。在数字化、网络化、智能化环境下,人们对信息、知识的需求不断增长,对服务方式、服务渠道、服务水平的要求不断提升,因此,图书馆建设一直是学者关注的热点问题,研究内容包括新建设方案、建设标准、建设质量等。老问题采用的研究方法较多,如扎根理论、用户画像、引文分析、关联分析、自然语言处理、词频分析、调查访谈、内容分析、文献研究、案例分析、本体、比较分析、理性思辨、计量分析、统计分析等方法以及模型法、分类法等。老问题的贡献度相对较高,如“图书馆服务”的贡献度最高,为 0.273,但其新颖度最低,仅为 0.015;图书馆建设的贡献度位列第 3,但其新颖度位列倒数第 3。

利用 Pearson 相关系数计算研究问题的新颖度和贡献度之间的相关性。38 个研究问题的新颖度和贡献度总体呈负相关,相关系数为 -0.252,从 9 个新问题的首篇论文发表年看,多集中于 2000—2012 年,距今已 10 余年。由此可见,图书馆学领域的研究问题虽有自身特色,但学者对新问题的关注度不高,多针对老问题开展研究。

#### 4.4.2 M 的新颖度和贡献度

根据二八关键少数法则,将 31 种研究方法按新颖度高低近似分为 7 种新方法和 24 种老方法。由表 2 可知,新颖度较高的研究方法及其首篇论文发表年为:行动者网络理论(2015 年)、用户画像(2014 年)、情感分析(2013 年)、扎根理论(2009 年)、关联技术(2005 年)、政策工具理论(2004 年)、共词分析(2002 年)。新方法出现于 2001—2015 年,且多数方法来自其他学科,如用户画像始于经济营销领域,情感分析、扎根理论方法始于医疗、教育领域,关联技术则始于生物化学领域。图书馆学研究应用较多的老方法及其首篇论文发表年为分类法(1956 年)、调查研究法(1979 年)、统计分析法(1980 年)、实验观察法(1980 年)、结构方程模型(1980 年)、文献研究法(1981 年)、引文分析法

(1982 年)等。其中,分类法、文献研究法、引文分析法等是图书馆学领域专用研究方法<sup>[18]37</sup>,而调查研究法、统计分析法、实验观察法、结构方程模型则是社会学研究的通用方法。

根据 Pearson 相关系数计算,所有方法的新颖度和贡献度之间的相关系数是 -0.296,总体呈负相关。可见,目前图书馆学领域多数研究采用老方法,图书馆学专有方法的创新性不足。但在新技术环境下,部分学者正逐步吸收其他学科领域的技术方法用于研究本学科领域的问题。

#### 4.4.3 P—M 组合的新颖度和贡献度

根据新颖度、贡献度的不同,按照二八关键少数法则,结合部分  $(P-M)_{Nd}$ 、 $(P-M)_{cd}$  值相同的数据特点,在 222 对 P—M 组合中将  $(P-M)_{Nd} \geq 0.111$  的 59 对组合列为高新颖度 P—M 组合,其余为低新颖度 P—M 组合;将 53 对  $(P-M)_{cd} \geq 0.05$  的组合列为高贡献度 P—M 组合,其余为低贡献度 P—M 组合。在此基础上,将 222 对 P—M 组合划分为 4 种可能的状态:①新兴组合。P—M 组合的新颖度高、贡献度低,说明 P—M 组合是新近提出的,处于研究初期,尚未引起更多研究者的关注。②新兴热门组合。P—M 组合的新颖度、贡献度均较高,说明对应的 P—M 组合是近期提出的,已有一定的研究热度。③过热研究组合。P—M 组合的新颖度低、贡献度较高,说明对应的 P—M 组合相对陈旧,但研究热度不减。④退化组合。P—M 组合的相关研究新颖度、贡献度均很低,说明对应的 P—M 组合比较陈旧,且极少被研究者关注。

根据表 3 中 P—M 组合的新颖度和贡献度制作散点图,并建立坐标系。根据前文按照二八关键少数法则划分高低新颖度、高低贡献度的界限,选取坐标原点为(0.111,0.050);根据散点的分布特点,将 P—M 组合划分为 A 区(高新颖度、低贡献度)、B 区(高新颖度、高贡献度)、C 区(低新颖度、高贡献度)、D 区(低新颖度、低贡献度)4 个区域。具体如图 1 所示。

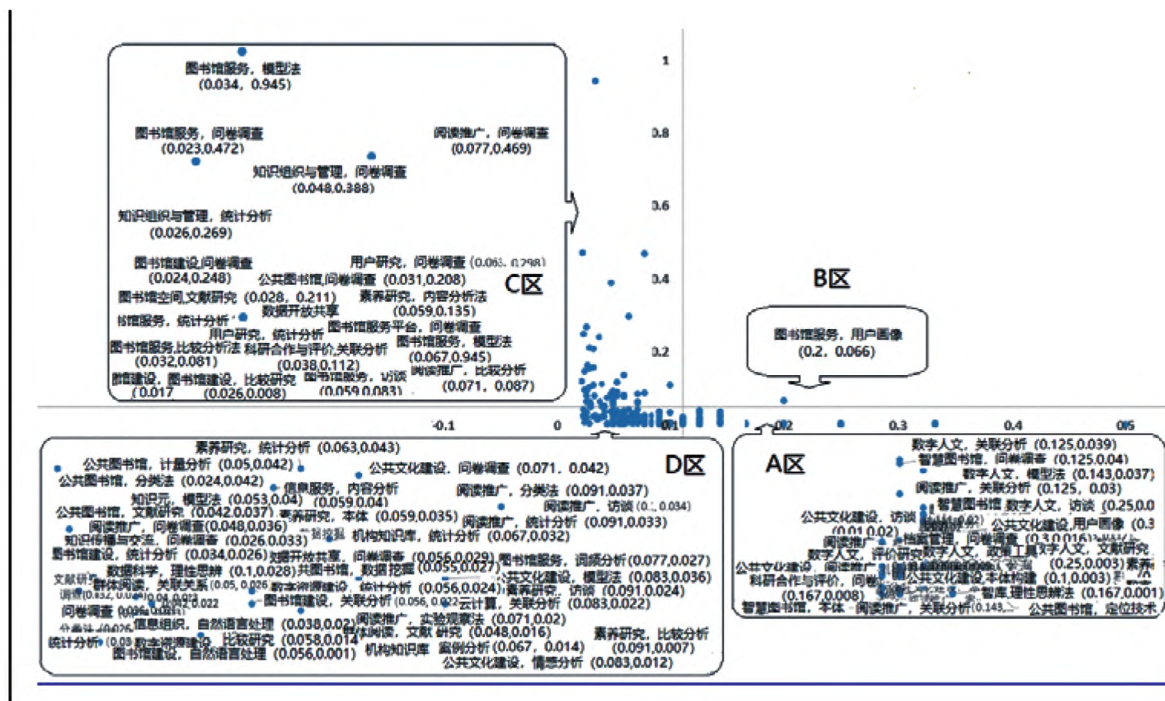


图1 图书馆学研究P—M组合的新颖度、贡献度分布

根据表2、表3,结合图1的P—M组合区间分布,可以得出图书馆学领域P—M组合的4种研究态势。

#### (1)新兴P—M组合

A区是高新颖度、低贡献度的新兴P—M组合,包括智慧图书馆—扎根理论、数字人文—行动者网络理论、图书馆建设—用户画像、图书馆服务—用户画像、公共文化建设—用户画像等。根据问题、方法的各种可能组合,将其分为新问题—新方法、新问题—老方法、老问题—新方法和老问题—老方法4类组合。

其一,新问题—新方法组合。该类组合中P、M均具有较高的新颖度,其中典型组合是数字人文—行动者网络理论,其新颖度为0.5。数字人文是新颖度最高的研究问题,其新颖度为0.1。数字人文又称人文计算,是全数字化环境和全人文化环境的产物,相关研究始于2011年美国斯坦福大学举办的“数字人文国际大会”,同年被引入我国图书馆学领域,属于较新的研究问题。同时,数字人文又是研究方

法,其涉及对电子信息的调查、分析、综合和表达<sup>[29]</sup>,如何利用媒体影响人文学科,以及人文学科对知识计算的贡献等。因此,相关研究方法较多,从应用时间看,先后出现了政策工具理论、内容分析法、文献研究法、访谈、理性思辨法、行动者网络理论等。其中,行动者网络理论的应用最晚,始于2013年<sup>[30]</sup>。数字人文—行动者网络理论是新颖度最高的P—M组合,相关研究包括通过行动者网络理论挖掘古籍内容关联关系、分析古籍网络中的人物身份和社交方式、探析古籍文化资本与社会资本的转化关系等<sup>[31]</sup>,该对P—M组合的应用有效拓展了图书馆资源组织和管理的方式。

智慧图书馆—扎根理论也是新问题—新方法组合,其新颖度为0.5。智慧图书馆是一种高度自动化管理的移动图书馆,能够提供智能共享空间、特色文化空间<sup>[32]</sup>以及智慧化服务与管理<sup>[33]</sup>。智慧图书馆的新颖度较高,为0.083,对应的研究方法较多,其中扎根理论是新颖度(0.077)最高的研究方法。扎根理论由社会学家 Barney Glaser 与 Anselm Strauss 于

2024年3月 March, 2024



1967年提出,是从实际观察入手,从第一手资料(访谈、调查等)中归纳经验,然后上升到理论的一种研究方法,适合研究新问题。但扎根理论直到2009年才被应用于图书馆学研究<sup>[34]</sup>,因此,在该领域属于新颖度较高的研究方法。扎根理论的应用为智慧图书馆服务要素的构建、服务模式的实现及服务能力的评价提供了方法基础。

其二,新问题—老方法组合。随着物联网和人工智能技术的发展,图书馆学研究问题呈现智能化趋势,出现了智慧图书馆、智库建设等新的研究问题。目前,我国智慧图书馆建设尚处于探索阶段,缺乏科学合理、完备细化的建设标准和服务体系,学者们主要通过扎根理论(0.077)、关联分析(0.042)、比较分析(0.027)、理性思辨(0.025)、调查研究(0.023)等方法分析智慧图书馆的概念体系和功能实现,探索智慧图书馆的理论框架、技术架构、系统平台、服务模式等,除扎根理论外,其他方法的新颖度均不高。智库聚合专家学者的智慧,为社会各领域提供满意的决策、方案、设计等。中国最早的智库是“稷下学宫”<sup>[35]</sup>,但我国图情领域对智库的关注始于2009年<sup>[36]</sup>,之后学者们从智库知识服务、智库功能、智库协同、智库机构服务、智库运行机制、智库建设、智库馆员、智库影响力等方面展开研究,对应的研究方法较为传统,包括访谈(0.037)、案例分析(0.028)、比较分析(0.027)、理性思辨(0.025)等方法。对新问题智库的研究虽然多采用老方法,但作为P—M组合,却具有较高的新颖度,其中新颖度最高的新问题—老方法组合是智库—访谈,新颖度为0.333。新问题的提出和研究极大地提升了图书馆的组织、管理和服务能力。

其三,老问题—新方法组合。新方法的应用为老问题的创新研究提供了有力支撑。图书馆建设—用户画像和素养研究—用户画像是图书馆研究中老问题—新方法组合的代表。用户画像在图书馆学领域早期研究出现于李

业根<sup>[37]</sup>发表的《基于大数据的图书馆信息营销策略》一文。用户画像包括个体用户画像和群组用户画像,是从收集的用户数据(如用户属性数据、用户行为数据)中抽象出标签化的用户模型,便于机器理解和处理用户信息。用户属性(社会属性、行业属性、生活习惯、兴趣特征等)数据和用户行为(阅读行为、访问行为、检索行为等)数据能够反映用户信息和其知识需求。相关研究基于用户数据,借助新技术和新手段构建融合多源信息的用户画像,包括用户行为建模、用户画像构建,以及用户画像的拓展、聚类、演绎、追踪等内容。图书馆借助用户画像能够精准预测用户需求,生成全内容的知识产品,实现全过程的服务优化,包括图书馆推荐系统构建、图书馆资源建设、创客空间建设、用户精准知识服务、读者位置定位、读者阅读障碍治疗、用户素养教育等。可见,用户画像这一研究方法的引入给图书馆建设、图书馆服务、图书馆素养培育等传统研究问题带来新的发展生机。

其四,老问题—老方法组合。典型的老问题—老方法组合是文化保护与传承—关联分析。我国的相关研究始于1995年吴建中与国际图联文献保护计划国际中心主任瓦拉莫夫的文献保护专题讨论<sup>[38]</sup>。2006年,该类研究开始升温,至今热度不减,研究问题包括历史文化、红色文化、民族文化、传统文化、区域地方特色文化、乡村特色文化等,相关研究方法包括文献调研、对比分析、访谈、案例分析、理性思辨、深度访谈、计量分析、关联分析等。其中新颖度最高的组合是文化保护与传承—关联分析,新颖度为0.25。关联分析属于因素分析法,其通过挖掘存在于大量数据集中的关联性或相关性,定量分析不同事物、方案的优劣,描述事物发展的规律和模式。1990年李荫涛<sup>[39]</sup>探究了关联分析法在情报学中的应用,图书流通管理则是关联分析法在图书馆较早的应用领域<sup>[40]</sup>。数字环境下,多源异构数据的融合成为增强知识服务、知识组合与管理效能的有效



途径,通过关联分析可以实现文化遗产的有效保护、扩充、利用与传承。因此,关联分析成为近年研究文化遗产与保护问题的有效手段,也促使对老问题的研究迈入一个新的发展阶段。

该区的 P—M 组合表明,问题与方法的有效组合能够产生新知识,知识组合是实现研究创新的有效途径。

#### (2) 新兴热门 P—M 组合

B 区为高新颖度、高贡献度的 P—M 组合,但该区域仅有图书馆服务—用户画像一对 P—M 组合。服务用户是图书馆一切工作的核心,也是研究者持续关注主题,因此对图书馆服务的研究热度不会减退;用户画像是获取用户多元信息、实现精准服务的新技术、新方法,备受研究者青睐。这对组合的新颖度为 0.2,贡献度为 0.066,该组合在图 1 中的位置可以看出,虽然其属于新兴热门 P—M 组合,但与 A 区中 P—M 组合的整体新颖度(区间在 0.111—0.5)相比,其新颖度不算太高;与 A 区中的 P—M 组合的整体贡献度(区间在 0.05—0.945)相比,其贡献度也偏低。综上可知,作为新兴热门 P—M 组合在很长一段时间内其研究热度将会持续,但研究者需避免重复研究。

#### (3) 过热 P—M 组合

C 区为低新颖度、高贡献度的 P—M 组合,从组合的贡献度看,由高到低依次为图书馆服务—模型法(0.945)、图书馆服务—问卷调查(0.472)、阅读推广—问卷调查(0.469)、知识组织管理—问卷调查(0.388)、用户研究—问卷调查(0.298)、知识组织与管理—统计分析(0.269)、图书馆建设—问卷调查(0.248)等。用户是图书馆服务的对象,围绕服务用户这一核心问题,学者们积极开展用户研究、资源建设研究、用户与馆员素养研究等。资源是图书馆服务的内容,包括资源的选择、收集、加工、整理、存储、优化、共享等,对资源进行有效组织与管理是为用户提供优质服务的保证。上述问题在图书馆学领域一直备受重视,研究热度持续不减。相当一部分研究采用传统的问

卷调查法、访谈法,结合结构方程模型、多维指标模型、统计分析等方法获取用户的兴趣爱好、行为习惯等信息,以提高服务的精准性和用户的满意度。但由于研究方法的新颖度较低,导致该类组合的新颖度偏低,在 0.022—0.1 之间。因此,在数智时代,应着力探讨应用新方法解决图书馆的核心问题,使老问题研究焕发新的生机和活力,以推动图书馆服务与建设的高质量发展。

#### (4) 退化的 P—M 组合

D 区为退化的 P—M 组合,包括图书馆史研究—问卷调查、图书馆建设—文献研究、图书馆服务—文献研究、知识传播与交流—问卷调查、古籍研究—统计分析、图书馆建设—统计分析、公共图书馆—分类法等。上述组合中,研究问题是图书馆关注的老问题,利用的研究方法也较为传统,其组合的新颖度很低,在 0.024—0.1 之间。该区域的 P—M 组合虽然较多(占总 P—M 组合的 50%),但每一对组合的贡献度(0.001—0.043)均较低,说明研究者已意识到利用本领域的老方法解决老问题难以实现研究创新。

针对退化的 P—M 组合,学者们或依托当前技术环境发现新问题,或改进研究方法实现老问题创新研究。如针对图书馆建设—文献研究、图书馆服务—文献研究等退化的 P—M 组合,通过改进研究方法实现创新,出现了图书馆建设—关联分析、图书馆服务—用户画像等组合;针对古籍研究—统计分析、档案管理—统计分析、图书馆史研究—问卷调查等组合,通过利用传统方法解决新问题实现创新,出现了数字人文—比较研究新组合。

#### 4.4.4 P、M 以及 P—M 组合的新颖度与贡献度之间的关系分析

新颖度代表研究问题、研究方法的创新性,贡献度反映研究问题、研究方法的热度。根据前文所述,2019—2021 年(前期)图书馆学领域 P、M、P—M 的新颖度和贡献度总体均成负相关,相关系数分别为 -0.252、-0.296、

—0.271。由此可知,前期高新颖度的知识元虽是当时学科领域研究的创新点,但同时期的贡献度并不高,即未形成研究热点。为追踪探究前期高新颖度的知识元经过一段时间研究积累后的演化趋势,笔者在中国知网检索获取2022年至2023年10月(后期)样本刊刊载的与P、M相关的论文,根据公式(1)、(2)、(3)计算P、M以及P—M组合在该时间段的贡献度,并计算前期P、M及P—M组合的新颖度与后期相应P、M及P—M组合的贡献度之间的Pearson相关系数,发现P、M及P—M组合新颖度与贡献度的相关系数分别为:0.353、0.227、0.061。可见,前期知识元的新颖度与后期对应知识元的贡献度之间呈正相关,且相关性呈现以下规律:前期的P知识元的新颖度与后期对应的贡献度之间的相关性最高,M知识元次之,P—M知识元最小。此结果表明,前期具有一定创新性的P、M及P—M组合知识元受到研究者的关注,现已成为热点研究内容。因此,本研究提出的P、M及P—M组合知识元分析方法能够探究学科领域的研究态势,发现有可能成为未来研究热点的研究机会。

## 5 结语

本研究提出一种细粒度分析学科领域研究现状的新视角。基于主题提取与规则结合的知识元抽取方法,通过自编程序构建P—M的共现关系,并给出P、M及其组合的新颖度、贡献度计算指标。再以我国图书馆学领域为例,运用该方法测度分析该领域近年的P、M、P—M组合及其相应的新颖度与贡献度的关系,从而验证了该方法在分析学科领域研究态势及发现新的研究问题方面的有效性。

### 参考文献:

- [1]肖希明,沈玲.中国特色图书馆学基础理论体系的历史发展与当代构建[J].中国图书馆学报,2021(3):4-22.
- [2]史静,朱薇薇,王鑫,等.近5年国际地球科学领域发展态势文献计量分析[J].地质学报,2017(12):2881-2888.
- [3]杨爱青.我国图书馆情报学科研究态势的可视化研究[D].曲阜:曲阜师范大学,2012.
- [4]刘亚文,张红燕,阳灵燕.共词分析国内外生物信息学领域研究态势[J].生物信息学,2020(3):186-194.
- [5]朱宗乾,刘庆.ERP实施领域研究热点国内外比较研究[J].情报杂志,2015(3):125-131.
- [6]张超星.学科领域发展态势分析:五维分析法[J].情报学报,2020(11):1204-1213.
- [7]王雪梅,张志强,李新.基于知识图谱和GIS分析的国际青藏高原研究态势分析[J].情报学报,2015(9):930-937.
- [8]周健,张杰,屈冉,等.基于LDA的国内外区块链主题挖掘与演化分析[J].情报杂志,2021(9):161-169.
- [9]熊文靓,付慧真.基于主题挖掘模型的跨学科性研究主题及其演化研究[J].情报科学,2021(11):117-126.
- [10]李慧,胡吉霞,佟志颖.面向多源数据的学科主题挖掘与演化分析[J].数据分析与知识发现,2022(7):44-55.
- [11]孙震,冷伏海.一种基于知识元迁移的ESI研究前沿知识演进分析方法[J].情报学报,2021(10):1027-1042.
- [12]丁志健,王焕景.基于文献计量的国内高校图书馆知识服务研究现状评析[J].图书馆工作与研究,2015(6):53-58.
- [13]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(1):69-79.
- [14]陆思晓,倪焯慧,景欣媛.公共图书馆古籍保护公众宣传现状调研[J].四川图书馆学报,2022(2):9-15.
- [15]解晓婷.长沙市图书馆儿童数字阅读推广实证分析[J].图书馆研究与工作,2022(4):60-65.
- [16]万文娟.“双一流”高校图书馆数据素养教育现状与发展策略分析[J].图书馆学研究,2021(21):56-61.
- [17]高波,黄洁晶.中国大陆图书馆资源共享现状调查[J].中国图书馆学报,2013(1):71-78.
- [18]吴丹,樊舒,李秀园.中国情境下图书馆学研究方法的识别、分类及应用[J].中国图书馆学报,2021(5).
- [19]李林华.图书馆学研究方法:困境及其出路[J].图书情报工作,2008(8):66-68,76.
- [20]李明杰.关于图书馆学方法论的哲学思考[J].图书馆理论与实践,1999(4):8-10.
- [21]焦运立.图书馆学研究方法体系和方法论体系的辨析——兼谈21世纪这两个体系的构建与整合[J].图书馆学研究,2010(7):2-6,36.
- [22]杨斌.基于研究过程的图书馆学方法论创新研究[J].图书馆学研究,2007(1):2-4.
- [23]Zadeh B Q, Schumann A K. The ACL RD-TEC2.0: a language resource for evaluating term extraction and entity recognition methods[C]//Proceedings of the Tenth International Conference on Language Resources and Evaluation.

- Portorož; European Language Resources Association, 2016: 1862-1868.
- [24]周宁,余肖生,刘玮,等.基于XML平台的知识元表示与抽取研究[J].中国图书馆学报,2006(3):41-45,13.
- [25]谈春梅,颜世伟,刘子牧.网络专题知识组织知识元自动抽取系统的设计与实现[J].数据分析与知识发现,2008(3):62-67.
- [26]余丽,钱力,付常雷,等.基于深度学习的文本中细粒度知识元抽取方法研究[J].数据分析与知识发现,2019(1):38-45.
- [27]钱佳佳,罗卓然,陆伟.基于问题—方法组合的科技论文新颖性度量与创新类型识别[J].图书情报工作,2021,65(14):82-89.
- [28]唐晓波,向莉丽,牟昊.基于研究问题与研究方法贡献的论文学术价值早期识别方法[J].情报科学,2022(9):3-11,19.
- [29]陈力.数字人文视域下的古籍数字化与古典知识库建设问题[J].中国图书馆学报,2022(2):36-46.
- [30]王广雷,吴晓伟,楼文高,等.基于人际竞争情报分析的产业集群信息服务机制研究[J].情报杂志,2013(4):16-21.
- [31]徐嘉弈.明清之际江南地区藏书家的书籍收藏阅读与书籍社交网络研究——以祁彪佳为例[D].曲阜:曲阜师范大学,2021.
- [32]李玉海,金喆,李佳会,等.我国智慧图书馆建设面临的五大问题[J].中国图书馆学报,2020(2):17-26.
- [33]初景利,段美珍.智慧图书馆与智慧服务[J].图书馆建设,2018(4):85-90,95.
- [34]邱建玲,刘占辉.图书馆馆员心理问题成因及对策的质性分析[J].中国电力教育,2009(6):204-205.
- [35]李亮.稷下学宫对当代智库参政议政的启示[J].智库理论与实践,2021(1):49-55.
- [36]张燕蕾.智库:图书馆发展的新机遇[J].图书馆学研究,2009(11):5-7.
- [37]李业根.基于大数据的图书馆信息营销策略[J].图书馆学刊,2014(10):7-9.
- [38]吴建中,瓦拉莫夫.保护文化遗产 抢救文献资源——关于图书馆未来的对话之三[J].图书馆杂志,1995(5):34-36.
- [39]李荫涛.关联分析法在情报学中的应用[J].情报理论与实践,1990(1):5-8.
- [40]吕文可.灰关联分析法在图书流通管理中的应用[J].图书馆,1998(2):35-36,43.

作者简介:

邵 阳(1985—),女,馆员,齐齐哈尔大学图书馆,黑龙江,齐齐哈尔,161000。

## Analysis of the Novelty and Contribution of Core Knowledge Elements and Their Combinations in the Field of Library Science in China

Shao Yang

**Abstract** This article takes six core journals in library science as the data source, extracts representative papers published in them from 2012 to 2021, analyzes the expression characteristics of research questions and research methods in the papers, constructs rules for extracting questions and methods, and establishes a knowledge element dictionary in the discipline field. Then, it extracts the research questions and method topics of all papers published in the sample journals from 2019 to 2021, establishes the co-occurrence relationship between questions, methods and their combinations through programming, and measures and analyzes the relationship between the novelty and contribution of questions, methods, and their combinations. Finally, it extracts relevant papers published in the sample journals from 2022 to October 2023, calculates and measures the relationship between the novelty and contribution of research questions, research methods, and their combinations in the field of library science from 2019 to 2021, and explores the evolution and development of various types of questions, methods, and their combinations. The study found that the method of analyzing from the perspective of knowledge elements of questions, methods, and their combinations can effectively discover new problems, new methods, old problems, old methods, and various combinations of questions and methods in the discipline field; the evolution and development of various types of questions, methods and their combinations provide a good background knowledge for analyzing the research trend in the discipline field.

**Keywords** Knowledge element; Combination innovation; Novelty; Contribution; Library science

**Class Number** G250.1; G252.8

2024年3月 March, 2024