

● 栾 锬, 李长玲, 王欣欣, 申力旭 (山东理工大学信息管理研究院, 山东 淄博 255049)

## 学科研究热点识别新视角: 主题动能<sup>\*</sup> ——以图书情报领域为例

**摘要:** [目的/意义] 从研究主题的知识生产与扩散两方面识别学科研究热点, 有利于明确学科重点, 把握学科发展动向, 为预测未来研究趋势提供研究基础。[方法/过程] 提出主题动能概念, 基于研究主题的生产与扩散指标构建主题动能模型, 综合分析某时间段某主题的研究质量 (横向热度) 与速度 (纵向变化), 识别学科研究热点; 并根据多个时间段的动能变化情况划分不同类型的研究热点, 描述其动态变化。[结果/结论] 选择图书情报领域 2017—2021 年研究文献与引证文献为例进行实证分析, 发现主题动能模型能够有效识别当前阶段的学科研究热点; 选择 2012—2016 年和 2017—2021 年两个时间段的相关数据分析主题动能变化, 能够划分学科“前沿型”“稳定型”“衰退型”三种类型研究热点, 实现对学科研究的动态描述。

**关键词:** 主题动能; 研究热点; 主题生产动能; 主题扩散动能; 识别方法

**DOI:** 10.16353/j.cnki.1000-7490.2023.03.003

**引用格式:** 栾锬, 李长玲, 王欣欣, 申力旭. 学科研究热点识别新视角: 主题动能——以图书情报领域为例 [J]. 情报理论与实践, 2023, 46 (3): 16-23.

### A New Perspective on Identifying Discipline Research Hotspots: Topic Kinetic Energy: Take LIS as an Example

**Abstract** [Purpose/significance] Identifying discipline research hotspots from the knowledge production and diffusion of research topics is conducive to clarifying the focus of current discipline research, grasping the development trend of discipline and providing a research basis for predicting future research direction. [Method/process] This paper proposes the concept of topic kinetic energy, constructs topic kinetic energy model based on the production and diffusion indicators of the research topic. By comprehensively analyzing the research quality (horizontal research heat) and research speed (longitudinal change) of the research topic in a certain period of time, the discipline research hotspots can be effectively identified. And according to the change of the topic kinetic energy in multiple time periods, different types of research hotspots are divided to describe its dynamic changes. [Result/conclusion] The empirical analysis on the literatures and citing literatures of LIS from two time periods of 2012-2016 and 2017-2021 shows that the topic kinetic energy model can effectively identify the current discipline research hotspots, and distinguish three types, namely frontier, stable and declining, to realize the dynamic description of discipline research.

**Keywords:** topic kinetic energy; research hotspots; topic production kinetic energy; topic diffusion kinetic energy; identification method

研究热点是科研共同体高度关注并积极开展研究的领域, 代表学科研究重点与发展方向<sup>[1]</sup>。利用科学方法识别学科研究热点及其变化情况, 是科研活动中探明知识基础、厘清学科重点、把握学科发展动向、发觉前沿趋势<sup>[2]</sup>的重要依据, 有利于科研机构或科研人员精准分配与投入研究资源, 为政府有关部门制定促进学科发展与优化学科发展环境等政策措施提供参考。期刊论文作为科研活动的

阶段性知识形态成果, 是最主要的成果表现形式<sup>[3]</sup>, 发文量反映研究主题的产出, 引用量反映研究主题传播扩散的影响力。因此, 本文基于期刊论文的发文量与被引量即研究主题的生产与扩散指标构建主题动能模型, 综合识别学科研究热点, 并根据热点主题的动能变化对其进行分类, 进一步发现学科研究的方向和趋势。

### 1 相关研究

当前, 国内外学者针对学科研究热点的识别开展了一系列研究。研究内容主要涉及以下几个方面。

<sup>\*</sup> 本文为国家社会科学基金重点项目“跨学科潜在知识生长点识别与创新趋势预测研究”的成果, 项目编号: 19ATQ006。

1) 基于引文分析的识别方法。基于文献之间的引用关系,构建引用网络,通过对共被引知识网络的可视化分析识别研究热点<sup>[4]</sup>。Chang等<sup>[5]</sup>采用文献耦合和共被引识别图书情报学的研究主题及演化趋势;Xie等<sup>[6]</sup>通过共被引可视化分析国际抗癌研究前沿;邓启平等<sup>[7]</sup>在构建领域文献引用网络的基础上,利用大规模网络聚类算法识别研究方向;候剑华等<sup>[8]</sup>利用引文网络结构变换模型方法,从共被引文献和施引文献的视角,对大数据领域的研究前沿和未来发展趋势进行预测。

2) 基于知识单元的识别方法。从词频与共词网络的角度识别研究热点。Liu等<sup>[9]</sup>通过词频分析识别坐骨神经损伤修复与保护的研究前沿和发展趋势;杨选辉等<sup>[10]</sup>运用突变检测与共词分析识别国内深阅读研究领域的新兴主题;荣国阳等<sup>[11]</sup>采用累积主题热度指数TP和主题加速度指数TAI,综合识别学科不同阶段的研究热点;Zhang等<sup>[12]</sup>采用关键词共现聚类方法识别分析我国农村普惠金融热点。

3) 基于时间序列的识别方法。在研究热点识别过程中,通过对整个分析对象的时间线进行划分,来探讨研究主题在整个生命周期中的演化过程<sup>[4]</sup>。李秀霞等<sup>[13]</sup>通过对不同时间段图书情报界个性化信息服务领域的研究文献进行聚类,揭示其研究热点、发展脉络及演化趋势;李长玲等<sup>[14]</sup>对期刊论文的被引频次赋予时间权重,改进z指数识别学科研究热点;奉国和等<sup>[15]</sup>构建时间加权关键词词频分析模型,计算关键词综合值,揭示学科研究热点及变化趋势;王康等<sup>[16]</sup>以时间加权修正后提取的关键词为知识单元进行科学主题演化分析,不仅能够突出“上升型”关键词,还能快速识别出突现词、新兴词和绝对高频词。

4) 基于多源数据的识别方法。因为单一数据源信息量的有限性,越来越多的学者整合利用多源数据进行热点识别。Park<sup>[17]</sup>、裘惠麟等<sup>[18]</sup>融合专利文献和期刊论文数据识别能源领域与机器学习领域的热点主题;王晓光等<sup>[19]</sup>整合政府公报、行业新闻、官网、微博、微信动态等多源信息,对专业领域热点进行探测;李慧等<sup>[20]</sup>以科技文献、专利和网页新闻数据为文本语料库,构建结合主题重要性波动率与相似度的话题演化判定模型,识别热点主题及演化路径。

综上,学科研究热点识别主要采用引文分析、词频突现、共现聚类、时间序列、多源数据等方法,从某一视角识别学科研究热点,识别结果具有一定的片面性。基于此,本文引入物理学动能定理,提出主题动能的概念,整合研究主题的发文数据与引证数据,基于研究主题的生产质量与速度测度主题生产动能,基于研究主题的扩散质量

与速度测度主题扩散动能,从生产与扩散两个维度、质量与速度两个特征构建主题动能模型,综合知识单元、引文分析和时间序列方法,分析研究主题的横向热度与纵向发展,识别学科研究热点,并根据热点主题多个时间段的主题动能变化情况对其进行分类,以实现对学科研究的动态描述。

## 2 理论基础与模型构建

### 2.1 理论基础

物理学中,动能定理的表达式为:

$$W = FS = E_{k1} - E_{k0} = \frac{1}{2}mv_1^2 - \frac{1}{2}mv_0^2 \quad (1)$$

式中, $F$ 为力; $S$ 为距离; $W$ 为力 $F$ 的做功; $m$ 为物体的质量; $v_0$ 与 $v_1$ 分别为物体的初速度与末速度; $E_{k0}$ 与 $E_{k1}$ 分别为物体的初动能与末动能; $W$ 为动能变化量。因此, $W$ 既是力的做功,也是物体具有的动能变化量。

将物理学动能定理引入主题研究热度测度,其适用性主要表现为两个方面。一方面,动能为标量且恒大于零,适用于变力做功,研究主题的生长受到引用、创新、被引等多方面力的作用,各类作用力的大小随时间不规则变动。所以,动能定理具有应用于测度知识成长的可行性。另一方面,动能变化的本质是物体在力的作用下产生位移,即研究主题在作用力下不断产生知识创新。若将研究主题的发展划分为多个时间段,计算动能变化量,则可分析主题研究热度的变化情况。

因此,本文以动能理论为基础,将某研究主题因为受引用、知识创新、被引等力的作用产生进步而具有的能量定义为“主题动能”。通过测度研究主题的动能,分析其研究热度,识别学科研究热点。研究主题受力分析示意图见图1。

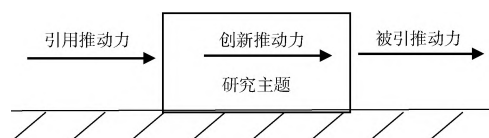


图1 研究主题的受力分析

Fig. 1 Force analysis of the research topic

研究主题在成长过程中受不同力的作用,会产生不同的动能。如果研究人员通过引用本学科或其他学科的参考文献,吸收各类知识,进行研究创新,使某研究主题的研究文献量从无到有、从少到多,就会具有生产动能;如果某研究主题的研究文献通过被引,使主题持续传承与发展,就会具有扩散动能。因此,本文通过构建主题动能模型,尝试从生产动能与扩散动能两方面测度研究主题在各种力的作用下的研究热度,识别学科研究热点。

## 2.2 模型构建

由于研究主题受力大小与位移多少无法计量,所以本文根据某研究主题的生产质量与速度、扩散质量与速度测度其生产动能与扩散动能,最终通过综合两方面的主题动能反映该主题的研究热度,识别学科研究热点并进行分类。

1) 基于引用与创新的主题生产动能。引用与创新是学科研究主题进步的基本推动力,二者共同促进了研究主题研究文献量的生产变化,直观反映其学科研究价值的肯定与热度。根据动能定理,构建某研究主题的生产动能  $E_p$  模型如下:

$$E_p = \frac{1}{2} m_p v_p^2 \quad (2)$$

式中,  $m_p$  为该研究主题的生产质量,即样本时间内的累积研究文献量;  $v_p$  为该研究主题的生产速度,即样本时间内其累计研究文献量在学科研究文献总量中占比的年均增速。其中:

$$m_p = \sum_{t=n}^i P_t \quad (3)$$

式中,  $t$  为年份;  $P_t$  为  $t$  年该研究主题的生产量(即研究文献量);  $n$  为该研究主题首次出现年份;  $i$  为数据截止年份。

在管理学中,投资在特定时期内的年均增长率为投资总增长率百分比的  $a$  次方根,即投资现有价值与基础价值的比值的  $a$  次方根,  $a$  为投资期间长度<sup>[21]</sup>。该年均增长率指标平滑了回报曲线,不会因为投资期间某时间点的回报巨变而迷失,因此本文引用投资年均增长率的计算思路,构建主题动能模型的速度指标。

生产速度用某研究主题研究文献量的学科占比的年均增速反映,一方面消除不同研究主题因学科每年研究文献总量不同导致的误差;另一方面有效识别相同比重的研究主题因累积时间不同导致的增速差异,以反映研究主题的新颖性。为避免占比增速出现负值影响后续动能计算结果,故取以  $e$  为底的指数函数。生产速度  $v_p$  公式如下:

$$v_p = e \left[ \left( \frac{m_p}{b v_p} \right)^{\frac{1}{i-n+1}} \right]^{-1} \quad (4)$$

式中,  ${}_{i} p v_p = \frac{\sum_{t=n}^i P_t}{\sum_{t=o}^i A P_t}$  为该研究主题的现有生产占比,即

样本时间内该研究主题累积研究文献量  $\sum_{t=n}^i P_t$  在学科研究文献总量  $\sum_{t=o}^i A P_t$  中的占比,  $o$  为样本数据起始年份;

$b v_p = \frac{P_n}{A P_n}$  为该研究主题的基础生产占比,即该研究主题首次出现年份  $n$  的研究文献量  $P_n$  在该年学科研究文献总

量  $A P_n$  中的占比。

2) 基于被引的主题扩散动能。被引是学科研究主题进步的后续推动力,研究主题的扩散范围越广,被引频次越高,说明该研究主题受到的关注度或认可度越高。根据动能定理,构建某研究主题的扩散动能  $E_s$  模型如下:

$$E_s = \frac{1}{2} m_s v_s^2 \quad (5)$$

式中,  $m_s$  为该研究主题的扩散质量,即样本时间内的累积引证文献量;  $v_s$  为该研究主题的扩散速度,即样本时间内其累计引证文献量在学科引证文献总量中占比的年均增速。其中:

$$m_s = \sum_{t=n}^i S_t \quad (6)$$

式中,  $t$  为年份;  $S_t$  为  $t$  年该研究主题的扩散量(即引证文献量);  $n$  为该研究主题首次被引年份;  $i$  为数据截止年份。

$$v_s = e \left[ \left( \frac{m_s}{b v_s} \right)^{\frac{1}{i-n+1}} \right]^{-1} \quad (7)$$

式中,  ${}_{i} p v_s = \frac{\sum_{t=n}^i S_t}{\sum_{t=o}^i A S_t}$  为该研究主题的现有扩散占比,即

样本时间内该研究主题累积引证文献量  $\sum_{t=n}^i S_t$  在学科引证文献总量  $\sum_{t=o}^i A S_t$  中的占比,  $o$  为样本数据起始年份;

$b v_s = \frac{S_n}{A S_n}$  为该研究主题的基础扩散占比,即该研究主题首次被引年份  $n$  的引证文献量  $S_n$  在该年学科引证文献总量  $A S_n$  中的占比。同理,取以  $e$  为底的指数函数。

3) 热点识别的主题动能。研究主题成长受引用创新推动力与被引推动力的合力作用,其主题动能  $E$  即为主题生产动能与主题扩散动能的代数和。考虑到二者不同的数据规模,故对  $E_p$  与  $E_s$  进行标准化。本文选用离差标准化方法,将数值映射到  $[0, 1]$  之间,计算公式如下:

$$x' = \frac{x_i - \min_{1 \leq j \leq n} \{x_j\}}{\max_{1 \leq j \leq n} \{x_j\} - \min_{1 \leq j \leq n} \{x_j\}} \quad (8)$$

经标准化后,  $E_p$  与  $E_s$  均转化为统一的标准化格式,分别用  $E_p'$  与  $E_s'$  表示。样本时间内研究主题的主题动能  $E$  如公式(9)所示:

$$E = E_p' + E_s' \quad (9)$$

## 3 实证研究——以图书情报领域为例

### 3.1 数据来源与预处理

本文以图书情报领域为例进行实证分析,以 CNKI 期刊数据库与中国引文数据库为数据源,样本来源为 2021 年综合影响因子排名前 10 的 CSSCI 期刊,包括《中国图书馆学报》《情报学报》《图书情报知识》《情报理论与实践

践》《图书与情报》《大学图书馆学报》《情报资料工作》《图书情报工作》《情报科学》《情报杂志》。由于 CNKI 期刊数据库缺少《情报学报》2012 年的数据，以维普中文期刊服务平台的数据作为补充。采集上述 10 种期刊 2012—2021 年 24432 篇载文的题录数据以及对应的 332786 条引证文献数据，数据采集时间为 2022 年 3 月 27 日。去除 24432 篇载文中的卷首语、通知、会议纪要、书评等非研究论文 1580 篇，得到有效样本 22852 篇。

本文选择以文献作者标识的关键词为该文献的研究主题，用以检验模型的有效性、识别图书情报领域的研究热点。样本文献数据预处理流程见图 2。

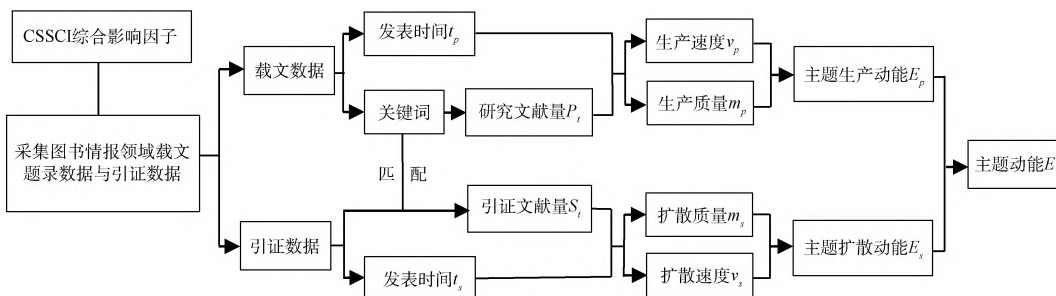


图 2 数据预处理流程

Fig. 2 Data preprocessing process

1) 获取主题生产动能  $E_p$  指标数据。①统计图书情报领域每年的研究文献量  $AP_i$ 。②用 Bibexcel 对载文数据的关键词字段进行拆分，统计关键词每年的研究文献量  $P_i$  和关键词首次出现年份；筛选有效词，去除研究目的不明确、去除表示研究背景的词，合并同义近义词；保留 2017—2021 年 5 年内累积词频高于 5 次的关键词共 1072 个。

2) 获取主题扩散动能  $E_s$  指标数据。①统计图书情报领域每年的引证文献量  $AS_i$ 。②利用函数提取引证文献数据中的发表时间，筛除 2022 年的引证文献；将载文对应的被引频次信息与其关键词对应，利用数据透视表和函数计算汇总关键词每年的引证文献量  $S_i$  并统计关键词首次被引年份。

### 3.2 主题动能计算与研究热点识别

1) 主题生产动能  $E_p$  计算。将研究主题 2017—2021 年的累积研究文献量  $\sum_{i=t_n}^i P_i$ ，首次出现年份的研究文献量  $P_n$ ，首次出现距离数据截止年份的时间差以及图书情报领域每年的研究文献量  $AP_i$  代入公式 (2) ~ 公式 (4)，计算主题生产速度  $v_p$  和主题生产动能  $E_p$ ，部分计算结果见表 1 第 7 列与第 8 列。

2) 主题扩散动能  $E_s$  计算。将研究主题 2017—2021 年的累积引证文献量  $\sum_{i=t_n}^i S_i$ ，首次被引年份的引证文献

量  $S_n$ ，首次被引距离数据截止年份的时间差以及图书情报领域每年的引证文献量  $AS_i$  代入公式 (5) ~ 公式 (7)，计算主题扩散速度  $v_s$  和主题扩散动能  $E_s$ ，部分计算结果见表 1 第 13 列与第 14 列。

3) 主题动能  $E$  计算。将  $E_p$  与  $E_s$  代入公式 (8) 进行标准化处理，将标准化后的  $E'_p$  与  $E'_s$  代入公式 (9)，计算 2017—2021 年的主题动能  $E$ ，部分计算结果见表 1 第 17 列。

本文根据帕累托法则，选取主题动能  $E$  由高到低排名前 20% 的研究主题作为识别得到的当前图书情报领域研究热点，相关指标的部分数据结果如表 1 所示。

表 1 为根据 2017—2021 年主题动能识别到的排名靠前的图书情报领域研究热点，表中识别结果“突发公共卫生事件”“科学数据”“文旅融合”“新冠肺炎”

“数字人文”等都是近几年突发的图书情报领域新颖研究热点，说明主题动能模型识别学科研究热点是有效的。为了更好地验证模型有效性，细化识别结果，本文对比分析表 1 研究热点在 2012—2016 年与 2017—2021 年两个时间段内的主题动能变化，测度研究热点的变化趋势并进行分类。

### 3.3 学科研究热点分类原则与结果

首先，计算图书情报领域研究热点 2012—2016 年的主题动能  $E_{(2012-2016)}$ ，部分计算结果见表 2 第 2 列；其次，考虑到引用指标的延迟特性，本文将 2012—2016 年研究文献在 2017—2021 年的被引情况纳入 2017—2021 年的主题动能  $E_{(2017-2021)}$  计算，以测度研究热点的长期变化情况，部分计算结果见表 2 第 5 列，因为被引数据的变化，所以表 2 中研究热点 2017—2021 年的主题动能与表 1 计算结果不同；最后，根据公式 (1) 计算研究热点的主题动能  $E_{(2017-2021)}$  与  $E_{(2012-2016)}$  的差量，即研究热点的主题动能变化量  $W$ ，部分计算结果见表 2 第 6 列。

为了对基于主题动能识别到的研究热点进行更好地区分，本文根据表 2 中的指标数据，依据以下原则分析研究热点研究热度（即主题动能）的趋势变化，对识别结果进行分类，以准确把握学科研究现状与未来发展趋势。

“前沿型”研究热点应当具备新颖性、增长迅速且影响力高的特点。本文将“前沿型”研究热点界定为

表1 2017—2021年主题动能计算与研究热点识别结果(部分)

Tab. 1 Topic kinetic energy calculation results and research hotspot identification results from 2017 to 2021 (part)

排序	研究热点	主题生产动能						主题扩散动能						标准化与主题动能		
		$m_p$	$P_n$	$pv_p$	$bv_p$	$v_p$	$E_p$	$m_s$	$S_n$	$pv_s$	$bv_s$	$v_s$	$E_s$	$E'_p$	$E'_s$	$E$
1	大数据	354	88	0.0348	0.0393	0.9765	168.7871	5428	92	0.0588	0.0828	0.9359	2377.1036	0.9011	1.0000	1.9011
2	高校图书馆	392	97	0.0385	0.0433	0.9774	187.2410	4253	63	0.0461	0.0567	0.9600	1959.9771	1.0000	0.8244	1.8244
3	突发公共卫生事件	105	2	0.0103	0.0009	1.8811	185.7791	467	3	0.0051	0.0003	2.6641	1657.2348	0.9922	0.6970	1.6892
4	网络舆情	212	44	0.0208	0.0196	1.0122	108.6003	3924	48	0.0425	0.0432	0.9967	1948.9487	0.5785	0.8198	1.3983
5	图书馆	277	73	0.0272	0.0326	0.9655	129.1127	3284	37	0.0356	0.0333	1.0133	1685.8933	0.6884	0.7091	1.3975
6	人工智能	156	20	0.0153	0.0089	1.1213	98.0693	2283	11	0.0247	0.0099	1.2224	1705.6702	0.5220	0.7174	1.2394
7	深度学习	123	10	0.0121	0.0045	1.2470	95.6395	1240	3	0.0134	0.0027	1.4596	1320.9552	0.5090	0.5554	1.0644
8	用户画像	72	3	0.0071	0.0013	1.4850	79.3847	1589	51	0.0172	0.0056	1.3830	1519.6571	0.4219	0.6391	1.0610
9	科学数据	176	16	0.0173	0.0071	1.2139	129.6637	1341	10	0.0145	0.0090	1.1056	819.5474	0.6914	0.3444	1.0358
10	智慧图书馆	87	11	0.0086	0.0049	1.1248	55.0326	1754	7	0.0190	0.0063	1.2801	1437.0362	0.2914	0.6043	0.8957
11	文旅融合	16	7	0.0016	0.0035	0.7885	4.9742	284	5	0.0031	0.0002	3.7768	2025.5651	0.0230	0.8520	0.8750
12	数字人文	134	16	0.0132	0.0071	1.1393	86.9726	1754	17	0.0190	0.0153	1.0451	957.9792	0.4626	0.4027	0.8653
13	新冠肺炎	106	19	0.0104	0.0093	1.0631	59.8984	949	140	0.0103	0.0047	1.6253	1253.3632	0.3174	0.5270	0.8444
14	扎根理论	136	21	0.0134	0.0094	1.0766	78.8134	1338	8	0.0145	0.0072	1.1619	903.1804	0.4188	0.3796	0.7984
15	公共图书馆	144	28	0.0142	0.0125	1.0257	75.7528	1314	7	0.0142	0.0063	1.1935	935.9309	0.4024	0.3934	0.7958
16	智库	169	28	0.0166	0.0125	1.0605	95.0387	1267	19	0.0137	0.0171	0.9578	581.1510	0.5058	0.2441	0.7499
17	情报学	160	34	0.0157	0.0152	1.0074	81.1855	1539	23	0.0167	0.0207	0.9584	706.8392	0.4315	0.2970	0.7285
18	阅读推广	130	27	0.0128	0.0120	1.0120	66.5757	1312	8	0.0142	0.0072	1.1567	877.6866	0.3532	0.3689	0.7221
19	情感分析	100	19	0.0098	0.0085	1.0306	53.1086	1342	6	0.0145	0.0054	1.2447	1039.5939	0.2810	0.4370	0.7180
20	突发事件	133	33	0.0131	0.0147	0.9769	63.4610	1585	14	0.0172	0.0126	1.0658	900.2301	0.3365	0.3784	0.7149
21	知识图谱	145	32	0.0143	0.0143	0.9998	72.4660	1484	20	0.0161	0.0180	0.9778	709.3972	0.3848	0.2980	0.6828
22	智慧服务	43	2	0.0042	0.0009	1.4406	44.6169	845	2	0.0091	0.0018	1.4685	911.1425	0.2355	0.3830	0.6185
23	政策工具	41	2	0.0040	0.0009	1.4220	41.4547	645	12	0.0070	0.0013	1.6781	908.2011	0.2186	0.3817	0.6003
24	情报工作	76	5	0.0075	0.0022	1.3147	65.6756	806	6	0.0087	0.0054	1.1060	492.9612	0.3484	0.2069	0.5553
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...

注:表中数据保留4位小数。

2012—2016年与2017—2021年两阶段主题动能变化量  $W$  排名前20%, 2017—2021年占比增速大于10%, 即生产速度  $v_p$  与扩散速度  $v_s$  均大于1.1052的研究热点。

“稳定型”研究热点具有宏观上热度相对稳定、微观上波动发展的特点<sup>[11]</sup>, 是图书情报领域较为稳固的核心研究内容。本文将“稳定型”研究热点界定为2012—2016年与2017—2021年两阶段主题动能变化量  $W$  大于0, 2012—2016年主题动能  $E_{(2012-2016)}$  排名前20%, 2017—2021年占比增速波动不超过5%, 即生产速度  $v_p$  与扩散速度  $v_s$  在0.9512~1.0513之间波动的研究热点。

“衰退型”研究热点近期关注度不高, 呈现热度逐年下降的趋势。该类研究热点在发展过程中已经积累一定的研究成果, 近年研究或纵向深入, 或横向拓展联系相近领域。本文将“衰退型”研究热点界定为2012—2016年与2017—2021年两阶段主题动能变化量  $W$  小于等于0, 2012—2016年主题动能  $E_{(2012-2016)}$  排名前20%且呈现负增长, 即2017—2021年生产速度  $v_p$  与扩散速度  $v_s$  均小于1的研究热点。

设目标学科研究热点总量为  $N$ , 某热点主题在某指标

的排名为  $K$ , 那么某热点主题在某时间段的主题动能指标中的排名为  $K_E$ 、在主题动能变化量指标中的排名为  $K_W$ , 分类原则设计为图3所示。

依据表2计算结果与图3所示分类原则, 将表1中图书情报领域研究热点识别结果区分为3类, 部分分类结果如表3所示。

### 3.4 结果分析

本文识别到的“大数据”“人工智能”“数字人文”“科学数据”“突发公共卫生事件”等与已有研究<sup>[11, 22-24]</sup>的识别结果相同或相近, 说明本文构建的主题动能模型识别学科研究热点是可行有效的。同时, 克服前期研究成果单指标、绝对量的研究局限, 从研究主题的研究文献量及其占比增速、引证文献量及其占比增速两个维度, 较全面地评价主题研究热度。

1) 主题动能模型较好地实现了质量和速度评价指标的平衡, 既能有效识别稳定高热度的学科研究热点, 也能识别到增速较快的新兴学科研究热点。与已有相关研究文献采用累计热度指标识别学科研究热点<sup>[22-24]</sup>相比, 本文采取衡量累积热度与热度增速的主题动能指标, 综合研究

表2 2012—2016年与2017—2021年两时间段

研究热点主题动能变化量计算结果(部分)

Tab. 2 Calculation results of topic kinetic energy change of research hotspot in 2012-2016 and 2017-2021 (part)

研究热点	2012—2016		2017—2021		W
	$E_{(2012-2016)}$	$v_p$	$v_s$	$E_{(2017-2021)}$	
大数据	1.0001	0.9765	0.9815	1.0068	0.0067
高校图书馆	0.6948	0.9774	0.9595	1.0843	0.3895
突发公共卫生事件	0.0056	1.8811	1.9125	1.0072	1.0016
网络舆情	0.3757	1.0122	1.0070	0.6563	0.2806
图书馆	0.7438	0.9655	0.9453	0.7570	0.0132
人工智能	0.0027	1.1213	2.5494	0.6329	0.6302
深度学习	0.0050	1.2470	1.8704	0.5435	0.5385
用户画像	0.0000	1.4850	1.7314	0.4572	0.4572
科学数据	0.1653	1.2139	1.3165	0.7283	0.5630
智慧图书馆	0.0229	1.1248	1.1234	0.3168	0.2939
文旅融合	0.0000	0.7885	3.7295	0.0521	0.0521
数字人文	0.0183	1.1393	1.3355	0.4935	0.4752
新冠肺炎	0.0000	1.0631	1.2955	0.3291	0.3291
扎根理论	0.0572	1.0766	1.1721	0.4400	0.3828
公共图书馆	0.2654	1.0257	0.9677	0.4279	0.1625
智库	0.0842	1.0605	1.0606	0.5243	0.4401
情报学	0.2020	1.0074	1.0463	0.4620	0.2600
阅读推广	0.1818	1.0120	0.9684	0.3863	0.2045
情感分析	0.0558	1.0306	1.1763	0.3051	0.2493
突发事件	0.1358	0.9769	1.0412	0.3660	0.2302
知识图谱	0.1767	0.9998	1.0437	0.4179	0.2412
智慧服务	0.0085	1.4406	1.1373	0.2471	0.2386
政策工具	0.0094	1.4220	1.4696	0.2308	0.2214
情报工作	0.0200	1.3147	1.3273	0.3617	0.3417
...	...	...	...	...	...

注:表中数据保留4位小数。

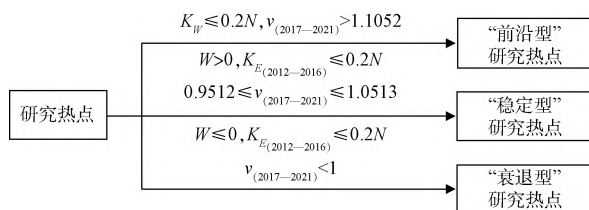


图3 学科研究热点分类原则

Fig. 3 Classification principle of discipline research hotspot

表3 研究热点识别结果分类(部分)

Tab. 3 Classification of research hotspot identification results (part)

热点类型	研究热点
前沿型	突发公共卫生事件、人工智能、用户画像、科学数据、数字人文、智慧服务、情报工作、学科建设
稳定型	大数据、高校图书馆、情报学、网络舆情、知识服务、文献计量、知识图谱、引文分析
衰退型	竞争情报、信息服务、专利分析、知识管理、共词分析、电子政务、社会网络、信息检索

主题的横向相对热度与纵向热度变化识别学科研究热点。一方面,主题动能模型能够识别持续高热度的研究热点。

例如,表1中热度排名第1的研究热点“大数据”,生产速度  $v_p = 0.9765$ 、扩散速度  $v_s = 0.9359$ ,速度排名分别为574与718,但其生产质量  $m_p = 354$ 、扩散质量  $m_s = 5428$ ,质量排名分别为2与1,综合质量与速度指标后,“大数据”的主题动能  $E$  最大,成为图书情报领域最热的研究主题。2015年武汉大学召开“数字时代图书馆学情报学教育国际研讨会”<sup>[25]</sup>,以信息数据管理为核心的图书情报学科引入大数据的研究,并研究热度猛增,较好地印证了本文的识别结果。另一方面,主题动能模型能有效识别增速较快的新兴突现研究热点。例如,表1中热度排名第3的研究热点“突发公共事件”,在样本期间的生产质量  $m_p = 105$ 、扩散质量  $m_s = 467$ ,质量排名分别为19与84,但其生产速度  $v_p = 1.8811$ 、扩散速度  $v_s = 2.6641$ ,速度排名分别为1与11,综合质量和速度指标后,“突发公共事件”成为增速较快的新兴学科研究热点。2020年新冠病毒疫情席卷全球,这一“黑天鹅事件”持续发酵并在短期内被世界卫生组织定性为“国际关注的突发公共卫生事件”<sup>[26]</sup>,对社会产生巨大的冲击力和震撼力,迅速成为国家重视、学者关注的热点问题,与本文研究结论相符合。因此,主题动能模型较好地均衡了质量与速度指标,从热度与增速两方面综合识别学科研究热点。

2) 主题动能模型从主题的生产与扩散两个维度评价主题研究热度,整合研究主题的发文数据与引证数据,避免因单一数据源的信息量有限影响研究热点的识别效力。与已有相关研究文献采用研究主题发文数据识别研究热点<sup>[11, 22-24]</sup>相比,主题动能模型考虑到研究主题受到的后续被引带来的推动力,在发文数据的基础上增加来自外部的引证数据,构建基于主题生产动能与主题扩散动能的主题动能模型,避免单一数据源评价研究主题的有限性。例如,本文与同期研究<sup>[11]</sup>共同识别到的研究热点“网络舆情”,在文献[11]的识别结果中“网络舆情”研究热度排名9,在本文识别结果中“网络舆情”基于发文数据的主题生产动能排名6,基于引证数据的主题扩散动能排名4,综合该研究主题的发文与引证数据后,其主题动能排名4,相比排名上升,较全面地评价主题研究热度。因此,主题动能模型整合研究主题的多源数据,测度结果更加全面、客观。

3) 主题动能模型利用相对量构建速度指标,能够消除因各年文献总量不同而产生的测度误差,有效测度学科研究主题在某时间的相对热度。与已有相关研究文献采用绝对发文量指标识别研究热点<sup>[22-24]</sup>相比,本文利用相对量构建速度指标既能够测度研究主题在某时间内的相对热度,反映多个时间段内的热度变化,又可以测度在研究主题的研究文献量或引证文献量相同的前提下,学科每年的

文献总量差异导致的研究主题占比差异,使得研究主题的热度测度更加客观。例如,表1中热度排名13的研究热点“新冠肺炎”和排名19的研究热点“情感分析”,二者首次出现年份的研究文献量相同 $P_n = 19$ ,但“情感分析”在样本期间的首年2017年已有研究文献,该年学科研究文献总量为2242篇,“新冠肺炎”在2020年出现研究文献,该年学科研究文献总量为2053篇,“新冠肺炎”的基础生产占比 $b_p = 0.0093$ 高于“情感分析”的0.0085,表明“新冠肺炎”这一学科研究热点在首次出现年份的研究热度高于“情感分析”,体现其受社会、学者的关注程度。因此,相对量的指标构建降低了绝对量的影响,有效测度学科研究主题在某时间的相对热度。

4) 根据主题动能变化情况可以有效区分不同发展趋势的学科研究热点。本文根据2012—2016年与2017—2021年两时间段内主题动能变化量 $W$ 划分不同发展趋势的研究热点。“前沿型”研究热点是近期研究热度比较高、增长迅速的研究热点,具有更好的研究潜力,更能体现学科研究的新方向,主要包括“突发公共卫生事件”“人工智能”“数字人文”“科学数据”等这些学科交叉性强、突现新颖的热点主题。“稳定型”研究热点具有目前阶段呈现稳定发展、随时间发展热度波动不大的特点,主要包括“高校图书馆”“情报学”“知识服务”“文献计量”等图书情报领域较为专业核心性的热点主题,也包括“大数据”这类近几年持续高热的热点主题。“衰退型”研究热点近期关注度不高,呈现热度逐年下降的趋势,该类研究热点在发展过程中积累了一定的研究成果,趋于饱和,主要包括“竞争情报”“知识管理”“电子政务”“信息服务”等热点主题。基于多时间段的主题动能变化细分学科研究热点的变化趋势,结果符合学科研究现状,近似其他学者的研究成果<sup>[11 24 27]</sup>,区分效果较好。

因此,通过测度某时间段学科研究主题的主题动能可以识别该阶段的学科研究热点,有效把握学科研究现状;根据多时间段的主题动能变化情况可以测度热点主题的发展变化,实现对学科研究的动态分析。

#### 4 结束语

本文将动能理论引入学科研究热点识别,构建主题动能模型,从生产与扩散两个方面、质量与速度两个特征综合评价研究主题的学术表现,识别学科研究热点;基于主题动能变化构建研究热点分类模型,实现对研究主题热度的动态描述。以图书情报领域为例进行实证研究,发现主题动能模型能够有效识别学科研究热点,“前沿型”“稳定型”“衰退型”的分类进一步发现学科研究的方向和趋势。

本文构建的主题动能模型在图书情报领域得到有效验证,但尚存在不足。在后续研究中,一方面需要对模型进行进一步的验证和完善,提高模型在其他学科、多学科应用的普适性;另一方面需要通过全文分词识别论文研究主题,避免关键词选取的主观性,使学科热点识别结果更加全面、有效。□

#### 参考文献

- [1] 卢超,侯海燕, DING Ying, 等. 国外新兴研究话题发现研究综述 [J]. 情报学报, 2019, 38 (1): 97-110.
- [2] 马费成. 关注学科热点透视学术进步 [J]. 情报资料工作, 2022, 43 (1): 13-14, 22.
- [3] 索传军, 盖双双. 单篇学术论文的评价本质、问题及新视角分析 [J]. 情报杂志, 2018, 37 (6): 102-107.
- [4] 罗瑞, 许海云, 董坤. 领域前沿识别方法综述 [J]. 图书情报工作, 2018, 62 (23): 119-131.
- [5] CHANG Yuwei, HUANG Muhsuan, LIN Chiaowen. Evolution of research subjects in library and information science based on keyword, bibliographical coupling, and co-citation analyses [J]. Scientometrics, 2015, 105 (3): 2071-2087.
- [6] XIE Ping. Study of international anticancer research trends via co-word and document co-citation visualization analysis [J]. Scientometrics, 2015, 105 (1): 611-622.
- [7] 邓启平, 王小梅. 科研领域重点研究方向判别方法研究 [J]. 图书情报工作, 2018, 62 (17): 60-67.
- [8] 侯剑华, 李蓬姬, 杨秀财. 基于引文网络结构变换的大数据研究前沿预测 [J]. 情报科学, 2018, 36 (6): 142-148.
- [9] LIU Guangyao, JIANG Rui, JIN Yan. Sciatic nerve injury repair: a visualized analysis of research fronts and development trends [J]. Neural Regeneration Research, 2014, 9 (18): 1716-1722.
- [10] 杨选辉, 杜心雨, 蔡志强. 基于突变检测与共词分析的深阅读新兴趋势分析 [J]. 图书馆建设, 2018 (5): 48-53.
- [11] 荣国阳, 李长玲, 范晴晴, 等. 主题热度加速度指数——学科研究热点识别新方法 [J]. 图书情报工作, 2021, 65 (20): 59-67.
- [12] ZHANG Mei, SU Huihui, WEN Jinghua. Hotspot analysis of rural inclusive finance based on keyword co-occurrence clustering [J]. International Journal of Applied Decision Sciences, 2022, 15 (1): 100-116.
- [13] 李秀霞, 邵作运, 郑春厚. 我国图书情报界 PIS 研究的共词可视化分析 [J]. 情报杂志, 2012, 31 (8): 109-113.
- [14] 李长玲, 牌艳欣, 相富钟, 等. 改进  $z$  指数的高被引学科研究热点识别方法探讨 [J]. 情报理论与实践, 2020, 43 (6): 69-75.

- [15] 奉国和, 孔泳欣. 基于时间加权关键词词频分析的学科热点研究 [J]. 情报学报, 2020, 39 (1): 100-110.
- [16] 王康, 陈悦, 苏成, 等. 多维视角下科学主题演化分析框架 [J]. 情报学报, 2021, 40 (3): 297-307.
- [17] PARK I, LEE K, YOON B. Exploring promising research frontiers based on knowledge maps in the solar cell technology field [J]. Sustainability, 2015, 7 (10): 13660-13689.
- [18] 裘惠麟, 邵波. 多源数据环境下科研热点识别方法研究 [J]. 图书情报工作, 2020, 64 (5): 78-88.
- [19] 王晓光, 王宏宇, 黄茵. 基于多源数据的专业领域热点探测模型研究 [J]. 图书情报工作, 2019, 63 (14): 52-61.
- [20] 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, 6 (7): 44-55.
- [21] 朱诒照, 王德萍. 资本运营管理 [M]. 上海: 复旦大学出版社, 2010: 87.
- [22] 马赫, 关心惠, 沈思. 图书情报学项目研究现状与热点: 基于“十三五”时期国家社科基金年度与青年项目的分析 [J]. 情报科学, 2022, 40 (4): 186-192.
- [23] 闫慧, 陈慧彤. 国内外图情档领域学术热点比较研究——兼论学术热点与学科发展的协同关系 [J]. 情报资料工作, 2022, 43 (2): 5-13.
- [24] 潘颖, 郑建明, 孙红蕾. 图书情报学科的研究特征与学科生长点透视——近5年国家社科基金课题指南与立项数据对比 [J]. 图书馆论坛, 2020, 40 (7): 28-35.
- [25] 郭晓婉, 冉从敬, 吴丹, 等. 大数据时代图书情报学理论与教育发展对策——第四届中美数字时代图书馆学情报学教育国际研讨会综述 [J]. 图书情报知识, 2016 (1): 116-121.
- [26] 世卫组织将新型冠状病毒疫情列为国际关注的突发公共卫生事件 [EB/OL]. [2022-05-25]. <http://world.people.com.cn/n1/2020/0131/c1002-31565415.html>.
- [27] 温有奎, 乔晓东, 张富财. 中国情报学期刊论文热点关键词演变轨迹与预测 [J]. 情报杂志, 2022, 41 (1): 164-169.
- 作者简介: 栾锟 (ORCID: 0000-0002-6284-6605), 硕士生。李长玲 (ORCID: 0000-0001-6266-4820, 通信作者, Email: lichl69@163.com), 教授, 硕士生导师。王欣欣 (ORCID: 0000-0002-9039-8432), 硕士生。申力旭 (ORCID: 0000-0003-4714-8152), 硕士生。
- 作者贡献声明: 栾锟, 理论与方法设计, 数据处理, 论文撰写与修改。李长玲, 框架设计与论文指导, 论文修改与审定。王欣欣, 论文修改与校对。申力旭, 论文修改与校对。
- 录用日期: 2022-10-21

(上接第29页)

- [14] 郑美莺, 梁飞豹. 基于引用期刊质量的影响因子修正法 [J]. 编辑学报, 2015, 27 (1): 19-21.
- [15] 郑德俊, 叶继元. 期刊引用认同及其被引评价新指标有效性分析 [J]. 中国图书馆学报, 2010, 36 (4): 106-114.
- [16] 李晓红, 于善清, 胡春霞, 等. 科技期刊评价中应重视“基金论文比”的作用 [J]. 科技管理研究, 2005, 25 (10): 138-139.
- [17] 刘睿远, 刘雪立, 王璞, 等. 基金论文比作为科技期刊评价指标的合理性——基于SCI数据库中眼科学期刊的实证研究 [J]. 中国科技期刊研究, 2013, 24 (3): 472-476.
- [18] 中国科学技术信息研究所. 2019版中国科技期刊引证报告(核心版社会科学卷) [M]. 北京: 科学技术文献出版社, 2020.
- [19] 中国科学技术信息研究所. 2019版中国科技期刊引证报告(核心版自然科学卷) [M]. 北京: 科学技术文献出版社, 2020.
- [20] 俞立平, 庞如超, 周娟美. 学术期刊学术传播水平评价——期刊传播因子 [J]. 信息资源管理学报, 2021, 11 (4): 133-140.
- [21] 王伟明, 徐海燕, 张发明. 基于双激励模型的学术期刊动态综合评价研究 [J]. 情报资料工作, 2020, 41 (5): 65-71.
- [22] 中国商办工业杂志社. 2017—2018年度CSSCI来源期刊目录公布 [EB/OL]. [2022-08-03]. [https://www.sohu.com/a/124501206\\_285117](https://www.sohu.com/a/124501206_285117).
- [23] 孙曦媚, 高秀香. 统计学原理 [M]. 北京: 北京理工大学出版社, 2017: 230.
- [24] 李跃艳, 熊回香, 李晓敏. 基于主成分分析法的期刊评价模型构建 [J]. 情报杂志, 2019, 38 (7): 199-207.
- 作者简介: 王欣欣 (ORCID: 0000-0002-9039-8432), 硕士生。李长玲 (ORCID: 0000-0001-6266-4820, 通信作者, Email: lichl69@163.com), 教授, 硕士生导师。栾锟 (ORCID: 0000-0002-6284-6605), 硕士生。申力旭 (ORCID: 0000-0003-4714-8152), 硕士生。
- 作者贡献声明: 王欣欣, 论文撰写与数据处理。李长玲, 研究设计与修改指导。栾锟, 论文修改与校对。申力旭, 论文修改与校对。
- 录用日期: 2022-11-08