

● 李昊城¹, 李长玲¹, 高峰², 刘嘉琪¹, 王浩¹

(1. 山东理工大学信息管理学院, 山东 淄博 255049; 2. 山东理工大学淄博发展研究院, 山东 淄博 255049)

构建动能模型识别学科新兴主题* ——以信息资源管理学科为例

摘要: [目的/意义] 从研究主题“新”与“兴”两个角度、知识生产和传播两个阶段, 识别学科新兴主题, 有利于跟踪学科发展最新动态, 为学术研究和科技创新提供参考。[方法/过程] 研究主题“新兴”过程与物体运动过程类似, 受引用推动力、创新驱动、被引牵引力等各种力的作用开始运动, 产生生产和传播动能。基于主题知识生产与传播两个方面、“新”与“兴”两个角度构建主题新兴动能模型, 多属性综合测度主题新兴程度, 识别目标学科新兴主题。[结果/结论] 选取信息资源管理学科 2018—2020 年和 2021—2023 年两个样本区间的研究文献及其被引数据为例进行实证研究, 发现模型能够有效识别学科新兴主题, 根据主题新兴动能变化, 划分为“新颖突现型”“稳定成长型”“成熟衰退型”三种类型, 以实现对新兴主题的深度分析。

关键词: 主题生产; 主题传播; 主题新兴动能; 新兴主题识别

DOI: 10.16353/j.cnki.1000-7490.2025.05.013

引用格式: 李昊城, 李长玲, 高峰, 刘嘉琪, 王浩. 构建动能模型识别学科新兴主题——以信息资源管理学科为例[J]. 情报理论与实践, 2025, 48 (5): 101-109.

Constructing a Kinetic Energy Model to Identify Emerging Topics in a Discipline: Information Resource Management as an Example

Li Haocheng¹, Li Changling¹, Gao Feng², Liu Jiaqi¹, Wang Hao¹

(1. School of Information Management, Shandong University of Technology, Shandong Zibo 255049;

2. Zibo Development Research Institute, Shandong University of Technology, Shandong Zibo 255049)

Abstract: [Purpose/significance] From the perspective of “newness” and “emergence” of research themes, and the two stages of knowledge production and dissemination, identifying the emerging themes of disciplines is conducive to tracking the latest development of disciplines, and providing references for academic research and scientific and technological innovation. [Method/process] The process of “emerging” research themes is similar to the process of object movement, which is driven by various forces such as citation driving force, innovation driving force, and attraction pulling force, etc., and starts to move, generating kinetic energy for production and dissemination. Based on the two aspects of theme knowledge production and dissemination, and the two perspectives of “newness” and “emergence”, we constructed the theme emergence kinetic energy model, measured the degree of theme emergence by multi-attribute synthesis, and identified the emerging themes of target disciplines. [Result/conclusion] The research literature and its citation data in the sample intervals of 2018 – 2020 and 2021 – 2023 in the discipline of information resources management are selected as examples for empirical research, and it is found that the model can effectively identify the emerging themes of the discipline, and according to the changes of the theme emerging kinetic energy, it can be divided into the “novel emergence”, “stable growth”, and “stable growth” categories. According to the changes in the emerging dynamics of the themes, the model can effectively identify the emerging themes and classify them into “novel and emergent”, “stable and growing”, and “declining and mature”, so as to realize the in-depth analysis of the emerging themes.

Keywords: thematic production; thematic communication; thematic emerging momentum; emerging thematic identity

* 本文为国家社会科学基金项目“跨学科知识流动中学科势能能对学科发展的影响研究”的成果之一, 项目编号: 23BTQ053。

学科新兴主题是科技竞争关注的焦点,对国家、企业及研发人员获取科技优势至关重要^[1]。在此背景下,识别学科新兴主题的重要性愈发凸显,传统新兴主题识别方法主要依赖静态或单一维度指标,往往难以全面反映主题的动态变化。因此,本文构建动能模型综合研究主题知识生产及传播阶段“新”与“兴”的相关特征指标,多属性综合捕捉学科新兴主题,有助于及时追踪科技发展趋势及前沿热点。

1 相关研究

新兴主题的概念最早于2002年提出^[2],被定义为新出现且重要的主题。随着对新兴主题内涵的不断理解,其被广泛定义为一种具有较大的科学影响力的、极具新颖性的、增长较快的主题^[3]。有学者认为它是目标学科中处于成长过程的重要主题^[4],应当强调时间维度上的“新”以及发展维度上的“兴”^[5]。因此,本文认为新兴主题的知识生产和传播两阶段都应当具备新颖性强、学科贡献度高和成长速度快的特点。

在前期研究中,常通过分析科技文献的关键词词频变化识别新兴主题,包括关键词出现频率急剧上升的程度^[6]、突现词及其作者数量变化^[7-8]等,这类方法虽然简单、及时,但仅从主题发文的角度出发,而忽略了被引对主题应用的描述,对新兴主题识别不够全面。

引文分析也是新兴主题识别常用研究方法,不同学者通过文献共被引与耦合关系、核心文档的相互引用和引文网络结构变化等方法识别特定领域的新兴主题^[9-12]。此外,还有学者通过随时间变化的引文分析,探讨主题在生命周期中的演化过程,识别新兴主题^[13]。这类方法从主题传播的角度,较为全面地衡量了新兴主题的受关注度,但却忽略了主题在知识生产阶段对新兴程度的表达。

综上所述,学科新兴主题识别主要采用词频突现、引文分析和时间序列等方法,从某一视角出发进行研究,识别结果具有一定的片面性。本研究团队在前期研究中借鉴物理学中的动能定理构建主题动能模型识别图书情报领域的热点主题^[14]。与之类似,在主题新兴过程中,其知识生产与传播同样具有动能属性,且动能与质量、速度有关,符合本文研究主题知识生产、传播的新颖性和学科贡献度的质量要求,也符合速度要求。因此,本文采用动能定理对主题新兴过程中的发文、引证数据,结合知识单元、引文分析、时间序列方法,量化研究主题的属性特征,细粒度识别新兴主题。

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 为物体所受作用力; S 为物体运动距离; m 为物体质量; v_0 与 v_1 为物体运动的初速度与末速度; E_{k0} 与 E_{k1} 为物体的初始动能与末动能; W 既是物理动能变化量,也是物体受力做功。

在研究主题成长过程中,可将其视作具有质量、速度及能量属性的物体^[14],在引用推动力、自身创新驱动力和被引牵引力作用下做功,使主题不断创新、成长与传播。研究主题受力分析如图1所示。

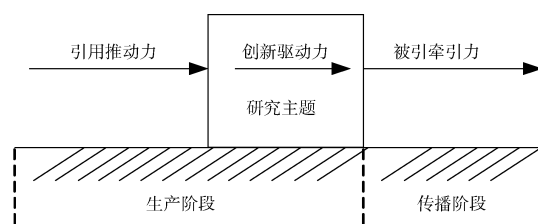


图1 主题受力分析

Fig. 1 Force analysis of the research topic

研究主题的“新”与“兴”往往伴随着主题知识产生、传播两个过程。期刊论文作为科研活动的阶段性知识产物,是最主要的成果表现形式^[15]。其中,论文发表是知识生产阶段,论文被引是知识传播阶段。

在生产阶段,研究人员引用本学科或其他学科的参考文献进行某主题的科学研究。一方面,引用行为使参考文献的知识输出推动这个主题的知识创新、成果产出,产生推动力;另一方面,研究人员自发的科研创新行为为主题生产提供内生驱动力。研究主题在引用推动力、创新驱动力的作用下产出新的研究成果,使这个主题的研究文献从无到有、从少到多,形成生产动能。

在传播阶段,主题吸引更多学者参与研究,相关研究文献被引用进行知识传播,对主题发展形成牵引作用。在被引牵引力的作用下,主题传播的广度和深度不断增加,形成传播动能。

通过测度研究主题生产、传播阶段的动能,可在一定程度上反映研究主题“新”与“兴”的程度。因此,本文以动能理论为基础,将研究主题受引用推动力、创新驱动力的作用使主题“新兴”的能量定义为“主题生产新兴动能”,研究主题受被引牵引力作用使主题“新兴”的能量定义为“主题传播新兴动能”。分别测度两部分能量,最终计算得到“主题新兴动能”,以识别学科新兴主题。

2.2 主题新兴动能模型构建

由于研究主题受力大小与位移多少无法量化,本文从新颖性和学科贡献度两个方面反映研究主题知识生产和传播的质量,结合速度指标,分别构建主题新兴生产、传播

动能模型,并计算主题新兴动能,识别学科新兴主题。

在新兴主题识别研究中,新颖性是最难量化的指标^[16]。在相关研究中,大多采用研究主题出现时长来表示其新颖性^[17]。但这种新颖性量化方式,无法区分同年出现主题的新颖程度,也忽略了新研究文献分布的时间区间、篇均生产时间等体现新颖性的指标。因此,本文构建主题生产、传播阶段的时长新颖度、分布新颖度和篇均新颖度指标,多属性综合测度研究主题的新颖性。

一方面,由于主题生产、传播阶段的新颖性相关指标计算需要统计其首次出现与被引时间,需要足够跨度的大样本数据,所以将1993—2023年定义为大样本区间;另一方面,新兴研究主题的“新”与“兴”需要重点分析其近年表现,因此选择2018—2020年和2021—2023年为两个实证样本区间,且以2021—2023年数据为主,2018—2020年数据为辅,分析研究主题的新兴趋势,对识别结果进行分类及深度分析。

2.2.1 主题生产新兴动能 研究主题生产阶段,引用推动力与创新驱动力的合力作用下促进了研究文献量的增长。其新颖度越大,学科贡献度越高,生产速度越快,则主题在样本区间的新兴程度越高。根据动能定理,构建研究主题生产新兴动能模型为:

$$E_{qp} = \frac{1}{2} m_{qp} v_{qp}^2 \quad (2)$$

式中, m_{qp} 为某研究主题 q 的生产新兴质量,由实证样本区间内主题 q 的文献生产新颖度与学科贡献度构成; v_{qp} 为主题 q 的生产速度。

1) 主题生产新兴质量。主题生产新兴质量的计算公式为:

$$m_{qp} = \left(\frac{1}{TP_n - TP_0 + 1} \times \frac{NP}{HP} \times \frac{\sum_{i=1}^{i=NP} (PT_i - ST) + 1}{NP} \right) \times \frac{NP}{FN} \quad (3)$$

式中, TP_0 为主题 q 在大样本中首次出现时间; TP_n 为主题在实证样本区间内最终出现时间; NP 为其在实证样本区间研究文献总量; HP 为其在大样本中首次出现时间至实证样本截止时间主题累积文献总量; $\frac{NP}{HP}$ 为主题 q 新研究文献量的占比; PT_i 为主题 q 在实证样本区间内第 i 篇文献的发表时间; ST 为实证样本区间起始时间; FN 为实证样本区间内目标学科文献总量; $\frac{NP}{FN}$ 为主题 q 的生产学科贡献度。

2) 主题生产速度。在统计学中,平均环比增长率能够平滑单个时间段内的异常对增长率的影响,使得测度的增长趋势更加稳定可靠,减少了由于短期波动引起的误

差,更好地反映增长趋势。基于此,构建主题生产速度模型为:

$$v_{qp} = \frac{1}{1 + e^{-\left\{ \left[\prod_{i=1}^j \left(1 + \frac{pn_i - pn_{i-1}}{pn_i} \right) \right]^{\frac{1}{j}} - 1 \right\}}} \quad (4)$$

式中, pn_i 与 pn_{i-1} 分别为某研究主题 q 在实证样本区间第 t 年、第 $t-1$ 年时的生产文献量; j 为实证样本时间跨度;

$\left\{ \left[\prod_{i=1}^j \left(1 + \frac{pn_i - pn_{i-1}}{pn_i} \right) \right]^{\frac{1}{j}} - 1 \right\}$ 为主题 q 生产环比增长率。由于主题生产文献量不可能为负这一特性,其生产速度恒大于等于0。为避免出现负值,将主题生产环比增长率映射到在实数区间值域为 $(0,1)$ 的 Logistic 函数,二者均为单调增函数且保证了主题的生产速度为正。其中,

Logistic 公式为 $f(x) = \frac{1}{1 + e^{-x}}$,将主题生产环比增长率代入 x 后的计算结果作为主题生产速度。

3) 主题生产新兴动能各变量含义。上文中主要变量及其释义见表1。

2.2.2 主题传播新兴动能 依据上文主题生产新兴动能模型构建思路,构建被引牵引力作用下的主题传播新兴动能模型。其公式为:

$$E_{qd} = \frac{1}{2} m_{qd} v_{qd}^2 \quad (5)$$

式中, m_{qd} 为研究主题 q 的传播新兴质量,由实证样本区间其传播阶段的新颖度与学科贡献度构成; v_{qd} 为主题 q 的传播速度。

1) 主题传播新兴质量。主题传播新兴质量的计算公式为:

$$m_{qd} = \left(\frac{1}{DP_n - DP_0 + 1} \times \frac{ND}{HD} \times \frac{\sum_{i=1}^{i=ND} (DT_i - ST) + 1}{ND} \right) \times \frac{ND}{SN} \quad (6)$$

式中, DP_0 为主题 q 在大样本中首次被引时间; DP_n 为其在实证样本区间内最终被引时间; ND 为实证样本区间内主题 q 被引文献总量; HD 为其在大样本主题首次被引时间至实证样本截止时间的累积被引文献量; DT_i 为主题 q 在实证样本区间内第 i 篇引证文献的发表时间; SN 为实证样本区间内目标学科文献被引总量。

2) 主题传播速度。主题传播速度的计算公式为:

$$v_{qd} = \frac{1}{1 + e^{-\left\{ \left[\prod_{i=1}^j \left(1 + \frac{m_i - m_{i-1}}{m_i} \right) \right]^{\frac{1}{j}} - 1 \right\}}} \quad (7)$$

式中, m_i 与 m_{i-1} 分别为主题 q 在实证样本区间第 t 年、第 $t-1$ 年的被引文献量; j 为实证样本时间跨度。

3) 主题传播新兴动能各变量含义。上文中主要变量及其释义见表2。

表1 主题生产新兴动能模型中各变量含义
Tab.1 Meaning of variables in the model of emerging kinetic energy of production

| 变量 | 变量名 | 变量说明 | 含义 |
|---|-----------|--|---|
| $\frac{1}{TP_n - TP_0 + 1}$ | 主题生产时长新颖度 | 为实证样本区间内主题 q 最终出现时间与与其在大样本中首次出现时间之差的倒数 | 其值越大,说明主题出现时间越短,新颖性越强,为避免分母为0,加1处理 |
| $\frac{NP}{HP}$ | 主题生产分布新颖度 | 为实证样本区间内主题 q 研究文献量与在大样本中主题首次出现时间至实证样本区间截止时间研究文献总量的比值 | 其值越大,说明文献越集中在实证样本区间,主题新颖性越强 |
| $\frac{\sum_{i=1}^{i=NP} (PT_i - ST) + 1}{NP}$ | 主题篇均生产新颖度 | 为实证样本区间内主题 q 每篇文献发表时间与其在实证样本起始时间之差的均值 | 其值越大,说明生产时间越接近实证样本结束时间,新颖性越强。为避免分子为0影响计算结果,加1处理 |
| $\frac{NP}{FN}$ | 主题生产学科贡献度 | 为实证样本区间内主题 q 研究文献量与所在学科总生产量的比值 | 其值越大,主题在目标学科的贡献度越大 |
| $\frac{1}{1 + e^{-\{ [\prod_{t=1}^j (1 + \frac{pn_t - pm_{t-1}}{pm_t})]^{\frac{1}{j-1}} \}}}$ | 主题生产速度 | 为实证样本区间内主题 q 研究文献量的平均环比增长率映射到 Logistic 函数的结果 | 其值越大,主题在样本区间内的生产速度越快 |

表2 主题传播新兴动能模型中各变量含义
Tab.2 Meaning of variables in the emerging kinetic energy model of topic propagation

| 变量 | 变量名 | 变量说明 | 含义 |
|---|-----------|--|--|
| $\frac{1}{DP_n - DP_0 + 1}$ | 主题传播时长新颖度 | 为实证样本区间内主题 q 最终被引时间与在大样本中首次被引时间之差的倒数 | 其值越大,说明其被引的时间跨度越短,传播新颖性越强,为避免分母为0,加1处理 |
| $\frac{ND}{HD}$ | 主题传播分布新颖度 | 为实证样本区间内主题 q 被引文献量与在大样本中主题首次被引时间至实证样本截止时间主题总被引量的比值 | 其值越大,说明主题被引文献越集中在实证样本区间,主题传播的新颖性越强 |
| $\frac{\sum_{i=1}^{i=ND} (DT_i - ST) + 1}{ND}$ | 主题篇均传播新颖度 | 为实证样本区间内主题 q 每篇引证文献发表时间与实证样本起始时间之差的均值 | 其值越大说明,主题被引时间越接近样本结束时间,主题传播的新颖性越强。为避免分子为0影响计算结果,加1处理 |
| $\frac{ND}{SN}$ | 主题传播学科贡献度 | 为实证样本区间内主题 q 被引量与学科总被引量的比值 | 其值越大,主题在所在学科的知识传播贡献度越大 |
| $\frac{1}{1 + e^{-\{ [\prod_{t=1}^j (1 + \frac{rn_t - rm_{t-1}}{rm_t})]^{\frac{1}{j-1}} \}}}$ | 主题传播速度 | 为实证样本区间内主题 q 被引量的平均环比增长率映射到 Logistic 函数的结果 | 其值越大,主题在样本区间内的传播速度越快 |

2.2.3 主题新兴动能 研究主题生产和传播阶段并非孤立存在,生产阶段的新兴知识需要传播阶段的被引来验证和推广,而传播阶段的反馈又促进主题进一步的研究,推动主题生产阶段的深入发展,二者相辅相成,同样重要,共同促进了研究主题的发展。

由于主题生产新兴动能 E_{qp} 与主题传播新兴动能 E_{qd} 两者的数据规模不同,故对二者进行 Min-Max 标准化处理,将计算结果映射至 (0,1) 区间,标准化公式为:

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

综合主题新兴过程中形成的生产新兴动能和传播新兴动能,构建的主题新兴动能模型为:

$$E_q = E_{qp}' + E_{qd}' \tag{9}$$

3 实证研究

3.1 数据来源

本文以信息资源管理学为例,选择中国知网 (CNKI)

期刊综合影响因子排名前10的期刊,包括《中国图书馆学报》《情报学报》《图书与情报》《现代情报》《情报科学》《情报理论与实践》《情报资料工作》《图书情报知识》《图书情报工作》《情报杂志》,在中国引文数据库 (CCD) 采集10种期刊于1993—2023年共31年的72595篇文献题录及其相关数据,去除会议纪要、书评、通知等非研究论文后,得到总计69932篇有效样本数据。其中,2018—2020年有5441篇,2021—2023年有5193篇,数据采集时间为2024年1月15日,文献题录信息包含作者、题名、关键词、参考文献、刊名和发表年份等字段,被引数据包含被引文献题名、被引时间等。

实证环节选择信息资源管理学在2018—2020年、2021—2023年两个实证样本区间的相关文献及被引数据作为研究对象。作者标注的关键词是文献内容的凝练^[18],故本文选择以文献作者标注的关键词代表文献的研究主题,用以检验模型的有效性,识别信息资源管理学新兴主题。数据处理与研究流程见图2。

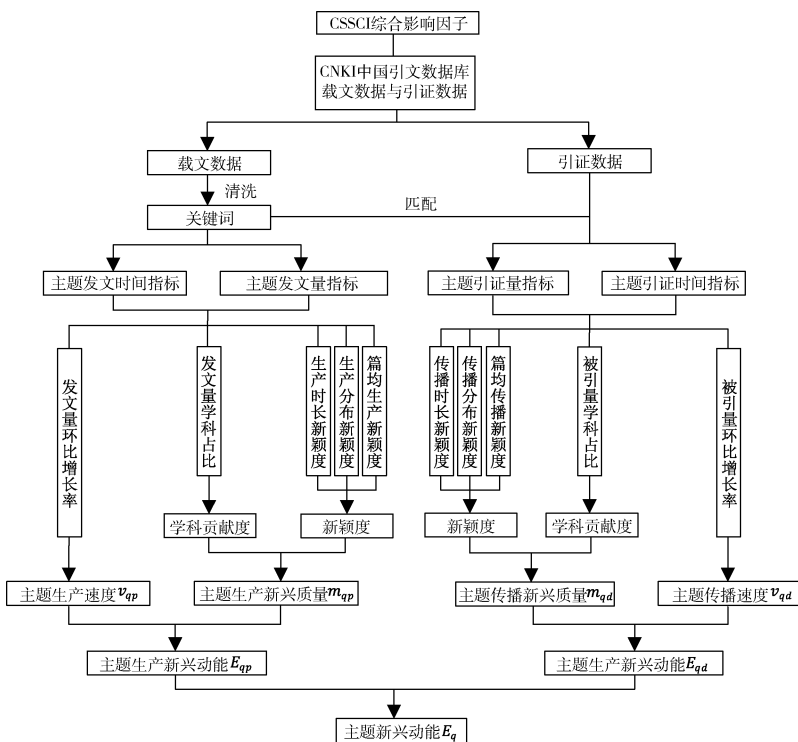


图2 研究思路流程

Fig. 2 Research idea flow chart

3.2 模型计算结果

依据图2流程对获取的数据进行处理与计算。

1) 数据预处理。将获取的数据导入MySQL数据库,区分发文数据和引证数据,编写Python代码对发文数据进行清洗,对关键词字段进行同义词、近义词合并,去除表示研究背景及研究意义不明的词。选取实证样本区间2018—2020年及2021—2023年出现频次大于等于5的关键词作为研究对象。

2) 研究主题生产与传播新兴动能计算。对清洗得到的关键词,获取发文时间及发文量指标,计算得到主题生产新颖度、生产学科贡献度和速度。将主题生产新颖度、学科贡献度代入公式(3),计算得到主题生产新兴质量 m_{qp} ,将其与主题生产速度 v_{qp} 代入公式(2)计算得到主题生产新兴动能 E_{qp} ,最新样本区间2021—2023年的生产质量、速度、动能计算结果分别见表3的第3列、第4列、第5列。

匹配载文数据中的关键词,获取主题文献被引时间及被引量指标。通过相关指标计算得到主题传播新颖度、学科贡献度和传播速度,将主题传播新颖度、贡献度代入公式(6)计算得到主题传播新兴质量 m_{qd} ,再将其与主题传播速度 v_{qd} 代入公式(5)计算主题传播新兴动能 E_{qd} 。主题生产、传播新兴质量相对于其速度数值较小,导致主题新兴动能没有较为明显的区分度,并对 m_{qp} , m_{qd} 放大千

倍处理。最新样本区间2021—2023年的传播质量、速度、动能计算结果分别见表3的第7列、第8列、第9列。

3) 研究主题新兴动能计算及新兴主题识别。将主题生产新兴动能 E_{qp} 及主题传播新兴动能 E_{qd} 分别代入公式(8)后得到标准化的生产与传播新兴动能 E_{qp}' 和 E_{qd}' ,再代入公式(9)计算主题新兴动能 E_q ,最新样本区间2021—2023年的结果分别见表3的第6列、第10列、第11列。

为有效分析新兴主题生产与传播趋势,增加2018—2020年样本区间主题新兴动能的计算结果列于表3的第12列。根据2021—2023年主题新兴动能 E_q 由高到低的排序结果(部分)见表3。

表3为2021—2023阶段根据主题新兴动能模型的部分识别结果,根据帕累托法则,选取降序排列前20%作为新兴主题识别结果,共得到95个。由于新兴主题识别是一种预测性任务,目前尚无统

一的标准来衡量识别结果的准确性^[19],故采用资料分析法验证本文识别结果。其中“ChatGPT”和“生成式人工智能”是众所周知的新兴主题,同时“新冠病毒感染疫情”“突发公共卫生事件”“区块链”“算法治理”“数字记忆”等与已有相关研究识别到的新兴主题^[20-22]一致,说明动能模型是可行有效的。

3.3 学科新兴主题分类原则与结果

研究主题新兴动能变化量反映主题新兴变化程度。选择2018—2020年新兴主题动能相关指标作为基准,分析2021—2023年区间主题新兴动能相对于过去的变化,以揭示主题增长或衰退的趋势,把握学科发展现状与动向。

核密度估计(Kernel Density Estimation),是一种重要的非参数统计方法,通过平滑数据分布,不需要依赖特定分布假设。其生成的密度曲线考虑了数据点在整个分布中的相对密度,相较于简单的算数平均,通过其曲线计算加权平均值更能反映数据特征,能够更合理确定研究主题新兴程度显著变化临界点。所以,使用核密度估计对2018—2020年新兴主题动能数据进行计算,得到其加权平均值 $A=0.1365$ 。

设某研究主题 q 两个区间的新兴动能差值 $\Delta E = E_{2021-2023} - E_{2018-2020}$,2018—2020年研究主题总量为 N ,主题在此区间新兴动能排名为 R 。经过不断调整阈值,观察数据分布并结合专家经验,最终确定 ΔE 等于 A 和 $-A$ 时

表3 主题新兴动能计算与新兴主题识别结果(部分)
Tab.3 Thematic energy calculation and emerging theme identification results (part)

| 排序 | 关键词 | 主题生产新兴动能 (2021—2023) | | | | 主题传播新兴动能 (2021—2023) | | | | 主题新兴动能 (两阶段) | |
|-----|----------|----------------------|----------|----------|-----------|----------------------|----------|----------|-----------|--------------------|--------------------|
| | | m_{qp} | v_{qp} | E_{qp} | E_{qp}' | m_{qd} | v_{qd} | E_{qd} | E_{qd}' | $E_{q(2021-2023)}$ | $E_{q(2018-2020)}$ |
| 1 | ChatGPT | 6.3815 | 1.0000 | 3.1907 | 1.0000 | 0.6400 | 1.0000 | 0.3200 | 0.9020 | 1.9020 | 0.0000 |
| 2 | 生成式人工智能 | 4.7940 | 0.9998 | 2.3960 | 0.7509 | 0.4051 | 0.9998 | 0.2026 | 0.5709 | 1.3219 | 0.0000 |
| 3 | 元宇宙 | 1.7198 | 0.7773 | 0.5196 | 0.1628 | 0.7095 | 0.9996 | 0.3547 | 1.0000 | 1.1628 | 0.0000 |
| 4 | 课程思政 | 0.5431 | 0.7432 | 0.1500 | 0.0470 | 0.1339 | 0.9998 | 0.0670 | 0.1887 | 0.2357 | 0.0000 |
| 5 | 新冠病毒感染疫情 | 0.3239 | 0.3793 | 0.0233 | 0.0073 | 0.6988 | 0.4783 | 0.0799 | 0.2252 | 0.2325 | 2.0000 |
| 6 | 算法治理 | 1.0862 | 0.8808 | 0.4213 | 0.1321 | 0.0640 | 1.0000 | 0.0320 | 0.0900 | 0.2221 | 0.0000 |
| 7 | 后疫情时代 | 0.1222 | 0.3439 | 0.0072 | 0.0023 | 0.1058 | 0.9742 | 0.0502 | 0.1414 | 0.1436 | 0.0000 |
| 8 | 突发公共卫生事件 | 0.3397 | 0.4136 | 0.0291 | 0.0091 | 0.3454 | 0.5232 | 0.0473 | 0.1331 | 0.1422 | 0.3425 |
| 9 | 区块链 | 0.5681 | 0.4569 | 0.0593 | 0.0186 | 0.3005 | 0.5257 | 0.0415 | 0.1169 | 0.1355 | 0.4400 |
| 10 | 跨境数据流动 | 0.4946 | 0.5246 | 0.0681 | 0.0213 | 0.1477 | 0.7402 | 0.0405 | 0.1139 | 0.1352 | 0.0000 |
| 11 | 数字孪生 | 0.1091 | 0.5045 | 0.0139 | 0.0043 | 0.1584 | 0.7637 | 0.0462 | 0.1301 | 0.1344 | 0.0000 |
| 12 | 数据要素 | 0.4553 | 0.7408 | 0.1249 | 0.0391 | 0.1181 | 0.7156 | 0.0302 | 0.0851 | 0.1242 | 0.0000 |
| 13 | 新文科 | 0.2833 | 0.3702 | 0.0194 | 0.0061 | 0.2032 | 0.6277 | 0.0400 | 0.1127 | 0.1188 | 0.0986 |
| 14 | BERT | 0.7770 | 0.4055 | 0.0639 | 0.0200 | 0.1939 | 0.5945 | 0.0343 | 0.0965 | 0.1165 | 0.1809 |
| 15 | 信息迷雾 | 0.4979 | 0.4396 | 0.0481 | 0.0151 | 0.0818 | 0.9302 | 0.0354 | 0.0996 | 0.1147 | 0.0000 |
| 16 | 信息茧房 | 0.3504 | 0.4744 | 0.0394 | 0.0124 | 0.1701 | 0.6502 | 0.0360 | 0.1012 | 0.1136 | 0.0544 |
| 17 | 数据治理 | 0.2911 | 0.4357 | 0.0276 | 0.0087 | 0.2183 | 0.5627 | 0.0346 | 0.0973 | 0.1060 | 0.1455 |
| 18 | 数字人文 | 0.4171 | 0.4656 | 0.0452 | 0.0142 | 0.2298 | 0.5285 | 0.0321 | 0.0903 | 0.1045 | 0.1717 |
| 19 | 数字记忆 | 0.4391 | 0.9313 | 0.1904 | 0.0597 | 0.0703 | 0.6624 | 0.0154 | 0.0433 | 0.1030 | 0.0000 |
| 20 | 数字政府 | 0.2934 | 0.4750 | 0.0331 | 0.0104 | 0.1114 | 0.7662 | 0.0327 | 0.0920 | 0.1024 | 0.0000 |
| 21 | 公共数据 | 0.2328 | 0.8972 | 0.0937 | 0.0294 | 0.0572 | 0.9308 | 0.0248 | 0.0697 | 0.0990 | 0.0000 |
| 22 | 颠覆性技术 | 0.3920 | 0.4113 | 0.0332 | 0.0104 | 0.1193 | 0.7244 | 0.0313 | 0.0881 | 0.0985 | 0.0658 |
| 23 | 智慧图书馆 | 0.2242 | 0.4899 | 0.0269 | 0.0084 | 0.2365 | 0.5166 | 0.0316 | 0.0888 | 0.0973 | 0.1798 |
| 24 | 扎根理论 | 0.2911 | 0.4280 | 0.0267 | 0.0083 | 0.2346 | 0.5118 | 0.0307 | 0.0863 | 0.0946 | 0.1756 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |

注：表中数据均保留4位小数。

为研究主题新兴程度显著增长和衰退的临界点。

为对新兴主题更好地进行区分，本文根据以下原则，对识别结果进行分类。

一是“新颖突现型”新兴主题，具有在目标学科内近期出现、短时间新兴程度显著提升的特点，因此本文将这类新兴主题划分标准界定为：出现时间小于等于3年， $\Delta E > A$ 的识别结果。

二是“稳定成长型”新兴主题，具有生产、传播阶段能量较高且发展稳定的特点，因此本文将这类新兴主题划分标准界定为：出现时间大于3年，在两个实证样本区间内皆为新兴主题，且 ΔE 在 $-A \sim A$ 之间波动的识别结果。

三是“成熟衰退型”新兴主题，具有发展时间较长，在目标学科内已经积累了一定的研究成果，主题前期新兴程度较高，但近期新兴程度下降趋势明显的特点，因此本文将该类新兴主题划分标准界定为：出现时间大于3年，在两个实证样本区间内皆为新兴主题，但 $\Delta E < -A$ 的识别结果。

识别结果的分类原则见图3。

依据表3 计算结果与图3 所示分类原则，将表3 中信

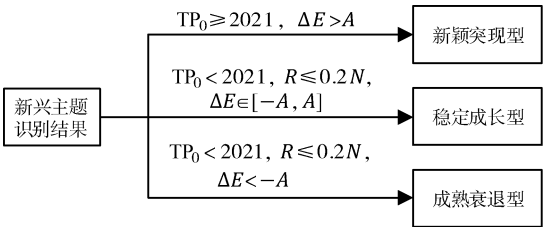


图3 新兴主题分类原则
Fig.3 Classification principle of emerging theme

息资源管理学科新兴主题识别结果区分为三类。部分分类结果如表4 所示。

表4 新兴主题识别结果分类(部分)
Tab.4 Classification principle of emerging identification (part)

| 主题类型 | 新兴主题 |
|-------|--------------------------------------|
| 新颖突现型 | ChatGPT、生成式人工智能、元宇宙、课程思政、算法治理、后疫情时代等 |
| 稳定成长型 | 数字人文、新文科、数据要素、数字孪生、跨境数据流动、BERT 等 |
| 成熟衰退型 | 新冠病毒感染疫情、突发公共卫生事件、区块链、在线健康社区、网络舆情等 |

3.4 结果分析及启示

相较于前期的相关研究，本文构建的主题新兴动能模

型具有以下优势。

1) 模型从研究主题知识生产和传播两个维度测度其新兴程度。从生产和传播两个维度测度主题的新兴程度,既能提高新兴主题识别的准确性,又能揭示主题在不同维度上的成长态势,避免片面数据源导致识别结果的偏差。

以表3中“数据要素”与“颠覆性技术”为例,在2021—2023年区间内,传播新兴动能几乎相同,分别为0.0302,0.0313,说明二者在传播方面的新兴程度相近,但从生产角度分析,“数据要素”的生产新兴动能为0.1249,大幅领先于“颠覆性技术”的0.0332。最终从主题新兴动能排名上,“数据要素”较“颠覆性技术”领先10个名次。

“数据要素”成为新兴主题的关键在于政策驱动和急剧增长的社会需求。2020年4月29日,中共中央、国务院印发《关于构建更加完善的要素市场化配置体制机制的意见》,明确“数据”为继土地、劳动力、资本、技术后的第5大生产要素,这一政策首次将数据从技术领域提升至国家战略层面。2023年,再次印发《“数据要素×”三年行动计划(2024—2026年)》,政策的连续推动,催生了一系列与“数据要素”强关联的新兴主题,包括“公共数据”“数据治理”“数字政府”“跨境数据流动”等。这些主题的新兴程度在本文的识别结果中都是较为显著的。这一实际现象,也印证了本文的识别结果。

2) 模型能够较好地描述研究主题的“新”与“兴”。对于新兴主题的“新”,模型对较难量化的新颖性指标进行了较好的改进;对于新兴主题的“兴”,基于主题的学科贡献度以及速度构建了相关指标;综合主题的“新”与“兴”,结合动能模型,进行了较为全面的评价。

通过主题三个新颖性指标能够比较全面地衡量主题的“新”。在表3中,“数字人文”与“扎根理论”的生产速度分别为0.4656与0.4280,后台数据中二者学科贡献度分别为0.0128与0.0130,衡量主题生产“兴”的两个指标数据都很接近。但在主题“新”的维度上,二者在大样本中首次出现时间接近,分别为2011年与2010年,而“数字人文”的生产分布新颖度,篇均生产新颖度皆高于“扎根理论”,二者的生产新颖度分别为0.0327与0.0216,说明在近年研究中“数字人文”较“扎根理论”的新成果产出更加持续和集中,整体表现出更强的新颖性。最终,“数字人文”生产新兴动能排名领先“扎根理论”17个位次,说明模型通过多个指标对主题的新颖度进行量化,能够较为全面地测度主题的“新”。

通过学科贡献度及速度指标两个维度能够较为全面衡量主题的“兴”。表3中的“课程思政”与“信息迷雾”的生产新颖度分别为0.3958、0.3889,数值接近。二者的

生产贡献度分别为0.0013、0.0015,差距较小。但其生产速度分别为0.7432、0.4396,“课程思政”大幅高于“信息迷雾”。“课程思政”作为教育领域的重要改革方向,得益于近年来一系列政策文件的出台和落实,相关政策从顶层设计到具体实施,对其产生与发展形成了强力驱动。最终,“课程思政”生产新兴动能排名领先“信息迷雾”11个位次,说明模型通过计算比较学科贡献度及速度,能够较为全面地测度主题的“兴”。

3) 模型能根据主题新兴动能变化情况细粒度划分不同发展趋势的新兴主题类型。本文根据主题在两个实证样本区间的动能变化情况,将新兴主题识别结果划分为“新颖突现型”“稳定成长型”“成熟衰退型”三种类型,它们分别处于生命周期的不同发展阶段。其生命周期如图4所示。

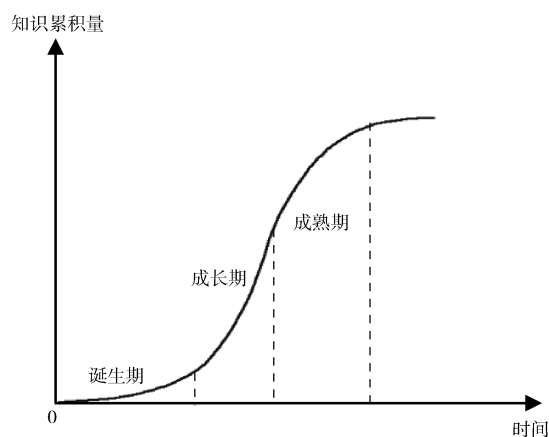


图4 生命周期示意
Fig. 4 Life cycle diagram

“新颖突现型”新兴主题,如表4中“ChatGPT”“课程思政”等处于“诞生期”,通常具有研究零散分布、新颖性高、短时间热度高但发展稳定性不足的典型特点。一方面,可能受新技术出现或政策变化等的影响,导致研究主题被替代或衰退;另一方面,随着研究力量的不断集中,可能继续发展为“稳定成长型”新兴主题。

“稳定成长型”新兴主题,处于生命周期中的“成长阶段”,研究文献快速增长,是学科发展的稳定力量。如表4中“数字人文”,在其概念形成后,信息资源管理学科基于资源和方法论优势,迅速成为数字人文研究的主力军^[23],相关创新研究持续突破,表现出稳定的创新趋势和影响力。

“成熟衰退型”新兴主题,具有“成熟期”的特点,表现为前期维持较高的研究热度,但近期下降趋势明显。在表4中,“新冠病毒感染疫情”“突发公共卫生事件”等由于宏观政策转变导致研究方向热度下降,“区块链”

“在线健康社区”等因为核心问题得到解决、创新空间减少,研究力量逐渐退出,主题“新”与“兴”的程度减弱。“成熟衰退型”新兴主题常随时间推移逐步衰退为一般性研究主题。

总之,信息资源管理学科的新兴主题多数为学科交叉性强的技术类和社会政策类问题。“ChatGPT”等新兴技术提高了信息处理的效率和精度,推动了学科的创新,带来了新的研究方向和机会。同时,信息资源管理学科必须确保坚守自身的理论特色和研究核心,使新兴技术服务于学科核心问题,与学科的理论基础相结合,关注学科长期的可持续发展,避免在技术浪潮中“迷失”自我。对于社会政策类主题,如“数据要素”“课程思政”等,反映了国家政策驱动下的研究需求,信息资源管理学科在发挥学科优势的同时,应当积极推动与其他领域知识的深度交叉融合,为创新研究提供新思路和新方案,在促进政策有效推进的同时,进一步提升学科影响力。

4 结束语

本文结合物理学中的动能定理构建主题新兴动能模型,从主题的生产与传播两个过程、新与兴两个角度,综合主题的新颖性、学科贡献度和成长速度多属性综合测度主题的新兴程度;基于主题相邻样本区间新兴动能变化情况,构建新兴主题分类模型,实现对新兴主题的动态描述。以信息资源管理学领域为例进行实证研究,结果表明模型能够有效识别学科新兴主题,并将识别结果划分为“新颖突现型”“稳定成长型”“成熟衰退型”三类,结合生命周期理论,进一步探讨了不同类型的新兴主题的发展规律和特点。

虽然本文构建的主题新兴动能模型在信息资源管理学领域得到了有效验证,但在其他学科应用的普适性还需进一步验证。此外,未来将从标题、摘要、全文中提取关键词对研究内容进行补充,扩大数据集范围,综合模型在不同学科中的验证结果,继续优化模型,使学科新兴主题识别更加准确、全面。□

参考文献

- [1] ROTOLO D, HICKS D, MARTIN B R. What is an emerging technology? [J]. Research Policy, 2015, 44 (10): 1827-1843.
- [2] MATSUMURA N, MATSUO Y, OHSAWA Y, et al. Discovering emerging topics from WWW [J]. Journal of Contingencies and Crisis Management, 2002, 10 (2): 73-81.
- [3] WANG Qi. A bibliometric model for identifying emerging research topics [J]. Journal of the Association for Information Science and Technology, 2018, 69 (2): 290-304.
- [4] TU Yining, SENG Jialang. Indices of novelty for emerging topic detection [J]. Information Processing & Management, 2012, 48 (2): 303-325.
- [5] 柴文越, 刘小平, 梁爽. 新兴主题识别方法研究综述 [J]. 现代情报, 2023, 43 (12): 164-177. (CHAI Wenyue, LIU Xiaoping, LIANG Shuang. Review of emerging topic identification methods [J]. Journal of Modern Information, 2023, 43 (12): 164-177.)
- [6] OHNIWA R L, HIBINO A. Generating process of emerging topics in the life sciences [J]. Scientometrics, 2019, 121 (3): 1549-1561.
- [7] YANG Jinqing, LU Wei, HU Jiming, et al. A novel emerging topic detection method: a knowledge ecology perspective [J]. Information Processing and Management, 2022, 59 (2): 102843.
- [8] 唐恒, 邱悦文. 多源信息视角下的多指标新兴技术主题识别研究——以智能网联汽车领域为例 [J]. 情报杂志, 2021, 40 (3): 81-88. (TANG Heng, QIU Yuewen. Emerging technology topic identification based on multi-source information: intelligent connected vehicle as an example [J]. Journal of Intelligence, 2021, 40 (3): 81-88.)
- [9] XU Haiyun, WINNINK J, YUE Zenghui, et al. Multidimensional scientometric indicators for the detection of emerging research topics [J]. Technological Forecasting and Social Change, 2021, 163 (1): 120490.
- [10] SASAKI H, FUGETSU B, SAKATA I. Emerging scientific field detection using citation networks and topic models—a case study of the Nano carbon field [J]. Applied System Innovation, 2020, 3 (3): 1-17.
- [11] 侯剑华, 李莲姬, 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. (HOU Jianhua, LI Lianji, YANG Xiucui. Frontier prediction of big data based on citation network structure transformation [J]. Information Science, 2018, 36 (6): 142-148.)
- [12] 曹艺文, 许海云, 武华维, 等. 基于引文曲线拟合的新兴技术主题的突破性预测——以干细胞领域为例 [J]. 图书情报工作, 2020, 64 (5): 100-113. (CAO Yiwen, XU Haiyun, WU Huawei, et al. Study on radical innovation prediction to emerging technology topics based on citation curve fitting: taking the field of stem cells as an example [J]. Library and Information Service, 2020, 64 (5): 100-113.)
- [13] 葛菲, 谭宗颖. 学科领域主题新兴趋势探测方法研究——基于关键词生命周期和引文分析 [J]. 情报理论与实践, 2013, 36 (9): 78-82. (GE Fei, TAN Zongying. Research on the emerging trend detection method of the discipline area themes [J]. Information Studies: Theory & Application,

- 2013, 36 (9): 78-82.)
- [14] 栾锟, 李长玲, 王欣欣, 等. 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. (LUAN Kun, LI Changling, WANG Xinxin, et al. A new perspective on identifying discipline research hotspots: topic kinetic energy: take LIS as an example [J]. Information Studies: Theory & Application, 2023, 46 (3): 16-23.)
- [15] 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. (MA Feicheng. Focusing on disciplinary hotspots to gain insight into academic progress [J]. Information and Documentation Services, 2022, 43 (1): 13-14, 22.)
- [16] XU Shuo, HAO Liyuan, AN Xin, et al. Emerging research topics detection with multiple machine learning models [J]. Journal of Informetrics, 2019, 13 (4): 10.1016.
- [17] 邓启平, 柯佳秀. 基于基金项目数据的新兴交叉主题识别——以量子技术为例 [J]. 图书情报工作, 2023, 67 (20): 130-141. (DENG Qiping, KE Jiaxiu. Identifying emerging interdisciplinary topics based on the fund project data: a case study of quantum technology [J]. Library and Information Service, 2023, 67 (20): 130-141.)
- [18] 黄如花, 赵洋, 黄雨婷. 国际开放科学研究进展 [J]. 图书情报工作, 2021, 65 (1): 140-149. (HUANG Ruhua, ZHAO Yang, HUANG Yuting. Research progress in the international research of open science [J]. Library and Information Service, 2021, 65 (1): 140-149.)
- [19] 许海云, 张慧玲, 武华维, 等. 新兴研究主题在演化路径上的关键时间点研究 [J]. 图书情报工作, 2021, 65 (8): 51-64. (XU Haiyun, ZHANG Huiling, WU Huawei, et al. Key time-points of emerging research topic on their evolution path [J]. Library and Information Service, 2021, 65 (8): 51-64.)
- [20] 段庆锋, 陈红, 刘东霞, 等. 基于 LSTM 模型与加权链路预测的学科新兴主题成长性识别研究 [J]. 现代情报, 2022, 42 (9): 37-48, 142. (DUAN Qingfeng, CHEN Hong, LIU Dongxia, et al. Identifying growth of discipline topics using LSTM and weighted link prediction [J]. Journal of Modern Information, 2022, 42 (9): 37-48, 142.)
- [21] 钱旦敏, 楼筱湾, 王华麟, 等. 我国信息资源管理学科及其邻近学科视角下的新兴主题识别 [J]. 图书馆论坛, 2023, 43 (9): 54-64. (QIAN Danmin, LOU Xiaowan, WANG Hualin, et al. Identification of the emerging themes from the perspective of information resource management and its adjacent disciplines in China [J]. Library Tribune, 2023, 43 (9): 54-64.)
- [22] 段庆锋, 陈红, 闫绪娴. 等. 基于知识结构突变的学科新兴主题识别研究 [J]. 情报学报, 2023, 42 (9): 1018-1028. (DUAN Qingfeng, CHEN Hong, YAN Xuxian, et al. Identifying emerging scientific topics by abrupt change of knowledge structure [J]. Journal of the China Society for Scientific and Technical Information, 2023, 42 (9): 1018-1028.)
- [23] 李娜. 国际数字人文研究的演化路径与热点主题分析 [J]. 图书馆, 2021 (5): 59-67, 73. (LI Na. Analysis on the evolution path and hot topics of international digital humanities research [J]. Library, 2021 (5): 59-67, 73.)
- 作者简介:** 李昊城, 硕士生。李长玲 (通信作者, Email: lichl69@163.com), 教授, 博士生导师。高峰, 博士, 教授, 硕士生导师。刘嘉琪, 硕士生。王浩, 硕士生。
- 作者贡献声明:** 李昊城, 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。高峰, 框架设计与方法指导。刘嘉琪, 论文修改与校对。王浩, 论文修改与校对。
- 录用日期:** 2024-12-12