·计量与评价·

基于科学论文多源数据的研究前沿集成识别模型研究

孙 震[1,2]

(1. 中国科学院文献情报中心 北京 100190; 2. 中国科学院大学 北京 100049)

摘 要 [目的/意义] 拟探讨设计一种研究前沿的集成识别模型,以期为研究前沿的相关实践提供一些借鉴和参考。[方法/过程] 对研究前沿的基本概念和识别方法进行详细梳理,总结现有问题和局限,利用科学论文的发文引用数据、下载使用数据、替代计量数据等多种数据,结合引文分析、词频分析、共词分析等多种工具和方法,以"神经网络计算"领域为例,构建了研究前沿的集成识别模型。[结果/结论] 根据科学论文不同类型和不同时期数据的特点,集成识别模型可以集合各种数据方法的优势,弥补各种数据方法的不足,提高最终识别结果的可靠性和准确度,具有一定的创新可行性。

关键词 研究前沿 引文分析 词频分析 共词分析 集成识别模型

中图分类号 G350 **文献标识码** A **文章编号** 1002-1965(2016)08-0095-06

引用格式 孙 震. 基于科学论文多源数据的研究前沿集成识别模型研究[J]. 情报杂志,2016,35(8):95-100.

DOI 10.3969/j.issn.1002-1965.2016.08.017

Study on the Integrated Detection Model of Research Front Based on the Multi-source Data of Scientific Papers

Sun Zhen[1,2]

(1. National Science Library, Chinese Academy of Sciences, Beijing 100190;
2. University of Chinese Academy of Sciences, Beijing 100049)

Abstract [Purpose/Significance] This paper intends to design an integrated detection model of research front in order to provide some references for the relevant research practice. [Method/Process] On the basis of the detailed analysis of the basic concepts and detection methods of research front, the problems and limitations are summarized. Using the "neural network computing" field as the case study, it uses publicity data, citation data, usage data and altmetrics data of scientific papers, combined with citation analysis, word frequency analysis and co-word analysis methods, to build the integrated detection model. [Result/Conclusion] According to the different types and features of scientific papers, the integrated detection model can make full use of the advantages of different data and methods, thus improve the reliability and accuracy of the final detection results, and it is proved that the model is innovative and feasible.

Key words research front citation analysis word frequency analysis co-word analysis integrated detection model

在科学研究主题的不断动态演变中,及时探测、识别并追踪目标研究领域或研究主题的最新前沿,既是科研人员必须具备的基本素养,又是科研管理者进行科学决策的重要依据。此外,围绕"科学前沿"开展研究也一直是小到科研人员、大到国家政府均十分关注的课题和方向。

科研领域中与研究前沿(Research Front)相似或相近的概念有很多,如研究热点(Research Focus)、新兴趋势(Emerging Trend)、新兴主题(Emerging Topic)、新兴研究领域(Emerging Field, Emerging Research Area)等。虽然至今没有形成统一的定义,但有别于研究热点,研究前沿普遍被认为是科学研究中最新、最先进、最有发展潜力的研究主题或研究领域,它来自于科学发现,代表了科学发展的难点、重点以及发展趋势,

收稿日期: 2016-04-20 **修回日期**: 2016-05-17

作者简介: 孙 震(ORCID: 0000-0002-9840-0541),男,1988年生,博士研究生,研究方向: 情报研究方法与技术。

具有前瞻性[1]。相应地,对于研究前沿的识别方法学者们也进行过各种不同的探索。

在对研究前沿的基本概念和识别方法进行详细梳理的基础上,总结存在的问题和局限,拟探讨构建一种基于科学论文多源数据的研究前沿集成识别模型,以期为研究前沿的实践提供一些借鉴和尝试。

1 研究前沿相关进展与现状

1.1 研究前沿的基本概念 研究前沿通常被认作是某时期内最具发展潜力的新兴研究领域或研究主题,而这些研究领域和主题表现为科学新发现和新进展时,往往会伴随着某种或强或弱的计量学信号。例如:某领域论文数量的增加,新作者新期刊的出现,主题词词频、含义或词间关系的变化,引文网络结构及引用关系的变化等[1]。因此,对这些计量学信号加以跟踪和利用就能探测和识别这些新兴的研究前沿和方向。

1965年,D. J. Price将引文网络中一篇施引文献近期频繁引用的30-50篇文献集定义为研究前沿(Research Front)[2],是计量学领域对研究前沿的最早定义。H. Small在1973年提出同被引分析法,并将同被引文献簇定义为研究前沿[3]。O. Persson在1994年将与高共被引文献簇相关联的施引文献群定义为研究前沿[4]。同年,E. Garfield将同被引聚类核心论文及其施引文献共同定义为研究前沿[5]。2003年,S. A. Morris又将利用文献耦合聚类所得到的、被固定文章所引用的文献集定义为研究前沿[6]。2006年,陈超美将一群突现的动态概念及潜在的研究主题定义为研究前沿[7]。

综上所述,计量学领域的研究前沿(Research Front)常被定义为一簇高被引文献或施引文献、一群突现概念或主题等,Research Front概念的提出是为了揭示某科研领域发展趋势的瞬时性成分,这种趋势现阶段是最新且吸引眼球的,但也是动态的、暂时的、不稳定的。因此,Research Front所代表的科研新趋势在未来可能扎根成为该领域的重要基础研究,也可能因缺乏长期的科研实际价值不被科学家所认可而最终蒸发消亡[8]。那么,如何准确描述"研究前沿"的真正概念?如何定义科学家眼中具有实际价值和真正效用的科研新趋势?有学者提出利用Research Frontier的概念定义现阶段具有重要价值和发展潜力、但相关研究尚未完全展开,未来有较大概率转化成重点科研选题的"真正"研究前沿,认为Research Front和Research Frontier是一种"期望值"与"观察值"、"先验评价"与"后验分析"、"潜在价值"和"实际价值"的关系[9]。由此看来,计量学领域的Research Front是情报人员对Research Frontier进行的预判,是为了帮助科学家提前识别和尽早确认某科研领域的Research Frontier。此外,Research Frontier也还需要时间来检验,并通过同行专家或输入源(即引文,Cited Reference)的最新施引文献(Citing Paper)的被引频次等方法特征进行最终确认。Research Front和Research Frontier概念的辨析和区分,对于厘清科技情报人员和科学家眼中对"研究前沿"概念的不同理解,解决计量学领域与科学共同体长期以来对"研究前沿"认知的混淆和矛盾具有重要意义。

1.2 研究前沿的识别方法 研究前沿的识别方法大体分为定性和定量两种。定性识别方法主要围绕专家展开,利用专家调查问卷或咨询会议等方式,结合德尔菲法等方法,整理提炼不同专家观点,形成代表该领域发展趋势的前沿预测报告,识别当前制约该领域发展的重大基础问题[10]。定性识别方法具有操作简便、直接、灵活等优点,但易受专家知识背景和经验等主观因素制约,也存在人力物力成本较大等缺陷。定量的识别方法是采用最多的方式,通过跟踪某领域或主题出现新进展时呈现的文献计量特征,以文献为核心对其内部知识结构进行挖掘、对其发展规律和特点进行探测并予以揭示,进而识别该领域或主题的发展趋势和动向。

利用文献的引用关系可以发现科学家们对某研究领域或主题的关注深度与交流密度,进而揭示科学发展规律,确定前沿发展态势与方向。因此,基于直接引用分析[11-12]、同被引分析[13-14]、文献耦合分析[15-16]等引文分析的方法已被广泛应用于研究前沿的探测和识别中。此外,为了克服引文分析的间接性和时滞性等缺陷,学者们也在尝试利用词频分析[17-18]和共词分析[19-20]等能基于直接反映研究内容主题词汇的方法进行研究前沿的识别。也有学者认为,计量学领域的某些文献计量指标,例如高被引文献集的稳定指数(Stability Index)[21]、文献引用半衰期(Citation Half Life)和中介中心性(Betweenness Centrality)[22]、创新指数(Novelty Index)和发表容量指数(Published Volume Index)[23]等,也可以从一定角度揭示研究领域的发展态势,用于研究前沿的探测。近年来,又出现了利用社区结构探测[24]、离群数据挖掘[25-26]、直接聚类分析[27]、语义计算应用于科技规划文本和项目数据[28]、科学文献的动态用户数据[29-30]等新方法和新数据进行研究前沿识别的案例。

各种研究前沿识别方法具有各自特点和优势,可以根据不同应用背景选择不同的识别方法。引文分析方法中,直接引用分析度量文献的直接关联,能在微观上更早揭示领域内部知识结构特征和发展态势;同被

引分析度量动态变化的被引文献关联,虽具有时滞性,但能宏观监测大范围甚至整个科学研究领域的内部知识结构演变; 文献耦合分析度量施引文献关联,耦合数据在文献发表后即可获得且是静态不变的。引文分析作为一种间接识别方法,由于其引用滞后性和对数据库的依赖,存在自身固有的缺陷,而基于主题词汇的直接识别方法,能更直观揭示研究领域和主题的内容特征,但由于句子才是文献文本结构的最小意义单元,词和共词并不足以描绘科学发展的全貌[31]。文献计量指标能在一定程度上反映研究前沿特征,但可能并不具有普遍适用性与代表性。近年兴起的利用科技规划文本和项目数据、科学文献动态用户数据等识别新方法,可以弥补传统方法数据的静态性、滞后性等缺陷,但也存在数据源不齐全、采集处理难度较大等制约因素。可以看出,单一的研究前沿识别方法,总会不可避免的存在这样或那样的缺陷,整合利用多样化数据源、综合运用多种识别方法、领域专家介入进行辅助判读等方式可以最大程度提高研究前沿识别的可靠性和精确度。

2　基于科学论文多源数据的研究前沿集成识别模型

2.1　研究前沿集成识别模型数据源的选择　一般认为,前沿的概念可以总体划分为科学研究前沿(Science Research Front)和技术创新前沿(Technology Innovation Front)。科学研究前沿简称"研究前沿",它以科学论文及其相关数据作为研究对象与中心,数据类型主要包括科学论文的发文数据、引用数据、使用数据和替代计量数据等[32],本文所讨论的研究前沿概念及其数据即来源于此; 技术创新前沿的基本研究数据则是围绕专利及其相关引文等展开的,其具体内容不再赘述。

科学论文的发文数据即文献题录等元数据,自论文发表便已产生,是生成其他数据的基础,其数据结构是静止不变的; 科研人员的引证行为催生了引用数据(施引数据和被引数据),但这些引用信息的形成需要时间积累,具有时间滞后性; 科学论文在网络数据库出版发表后被科研人员浏览(Html)、下载(PDF 和 XML),产生了论文的使用数据,在短时间内呈明显增长态势,可对其实时追踪研究,在引用数据形成前发挥作用; 科研人员在浏览论文后,可能会将其转发到 Mendeley、ResearchGate、科学网等学术社交媒体网络平台中,产生论文的替代计量数据,能一定程度上反映科研大众对该科研成果的态度。

前文梳理已知,现有研究前沿识别理论与实践往往只是单独利用了科学论文的一种数据,或者是基于文献的发文与引用数据,或者是基于科学论文的使用数据,或者是其他的项目数据等,很少有融合多种数据开展研究前沿识别的实践。因此,基于科学论文不同类型和不同时期数据的特点,本文拟结合科学论文的多种数据,综合运用多种工具和方法,以"神经网络计算"领域为例,对其研究前沿进行识别和探索,构建研究前沿的集成识别模型。

基本构成数据的选择分为两大部分: 传统的发文数据与引用数据,在 WOS(Web of Science)数据库输入 neural network comput* OR neural comput* OR comput* neuroscience OR neurocomput*,以主题字段进行检索,收集每篇文献的题录与引文数据; 科学论文的使用数据和替代计量数据,则依赖各数据库出版商提供的论文下载情况和 Mendeley、ResearchGate 等学术媒体社交平台的监测记录。其中,由于学术出版商公开的下载使用数据只有少数满足条件,选择 Springer、MIT Press 和 Elsevier 下的 Journal of Computational Neuroscience、Neural Computing and Applications、Neuroinformatics、Neural Processing Letters、Neural Computation、Neurocomputing、Neural Networks 等 7 种计算神经领域期刊的论文下载使用数据进行收集和处理; 由于相关研究[32]已证明 Mendeley 这类学术社交媒体的替代计量数据与 Facebook、Twitter 等社会大众社交平台相比,其数据变化规律与引用数据更为相似,只是出现时间点比引用数据更早,位置比引用数据有一定程度前移,因此选择 Mendeley、ResearchGate 学术社交媒体平台并以相关科学论文的收藏、分享、转发情况作为替代计量数据的主要收集来源。

2.2　研究前沿集成识别模型方法与工具的选择　国内外学者对研究前沿识别方法的优劣并没有达成统一定论,但基于引文和词汇的复合识别方法却被普遍认为能较好的用于研究前沿的实践[33-34]。另外,有学者对引文分析中直接引用、同被引分析、文献耦合等方法进行过比较,发现基于直接引证的引文网络具有更高的内容相似性[35],当识别对象不是大规模文献集时,能在微观上较早探测到较新、较大的类簇,能更直接揭示研究领域的内容结构特征和发展趋势,不容易遗漏新出现的研究前沿且识别效果较好[36-37]。因此,识别模型采用引文分析中直接引文网络、词频分析和共词分析的复合方法对论文的发文与引用数据进行分析。

SCI 创始人 E. Garfield 在历史演化图谱(Historiography Mapping)基础上,2006 年推出了引文编年可视化软件 HistCite[38],它能用图示方式展现某领域文献之间的引证关系,快速描绘该领域演进历史和脉络,发掘领域内的科学前沿。识别模型利用 HistCite 软件

进行引文分析，由于 HistCite 的 LCS（local citation score）比 GCS（global citation score）更能准确定位研究脉络，因此选取内部被引频次（LCS）达到一定阈值的 WOS 文献数据绘制引文编年图，进行神经网络计算领域的发展脉络、研究方向与主路径等探测分析。词频分析与共词分析则主要基于 WOS 数据利用书目共现分析软件 Bicomb[39] 和 SPSS 开展。

科学论文下载使用数据和替代计量数据的收集和处理则相对复杂，参考文献[29]和[30]的相关理论，基本原理是利用科学论文的下载使用次数和替代计量收藏转发热度来探测科学研究前沿。由于传统纸质期刊存在明显的滞后性且获取成本较大，外加近年 OA 开放获取和科研网络媒介的迅速普及发展，科学家在开展某新兴领域或主题相关研究之前，准备工作的第一步往往就是下载该领域相关文献、浏览该领域网络学术媒介相关专家的观点倾向等。相关研究也已证实，科学论文的下载使用数据与引用数据存在着某种关联，可以用来预测文献的未来引用影响力[40-41]。因此，如果某研究主题在一定时间段内被频繁地下载、讨论、转发和分享，说明这类新兴主题正在被科学家们大量关注，极有可能是目前的研究前沿，因为只有科学家们感兴趣的、与自身研究方向相符的主题才会被持续关注。利用科学论文的下载使用数据可以弥补引文分析等传统识别方法的时滞性、静态性等缺陷，而学术社交媒体的替代计量数据包含了非正式发表的学术信息和科研动向，是传统科学共同体的有力数据补充。

处理分析时首先将下载使用数据、替代计量数据分别处理：科学论文的下载使用数据利用学术出版商提供的论文下载统计工具获得（见表1），监测、记录并计算7种期刊所含论文主题词一段时期内的下载次数（T_1）；Mendeley、ResearchGate 学术社交媒体的网络替代计量数据，则通过检索或爬取学术社交媒体中包含的神经网络计算领域相关论文收藏和转发次数总和，依据论文收藏分享次数对应计算得出论文所含主题词的出现次数（T_2）。其次，汇总计算下载使用数据和替代计量数据合并后各主题词的出现总次数（T）。再次，利用各主题词在 WOS 数据库检索并计算含有该主题词的已发表文章总数（P）。最后，利用如下公式计算神经网络计算领域的研究前沿识别系数（R）：

$$T = T_1 + T_2 \quad (1)$$

$$R = \frac{T}{P} \quad (2)$$

将主题词出现总次数（T）除以与该主题相关的已发表论文数（P），与含有主题词的论文发表时间进行对比，如果含有主题词的论文发表时间较新，并且除商研究前沿识别系数（R）为较高数值，那么该主题则可代表研究前沿。意即：某研究领域或主题已发表论文数不多，但发表时间较新，证明该领域或主题近年刚被提出，并且如果该主题的相关论文或研究内容近期被频繁下载、收藏、分享和转发（尽管出现总次数不是所有主题中最高的），则证明该新兴主题正引起科研人员强烈关注，很有可能成为该领域的前沿方向[30]。

表1 主要学术出版商提供的论文下载统计工具[30]

出版商	论文下载统计工具	数据形式
Springer	Most downloaded articles	下载最多的5篇论文，按7/30/90天统计
Elsevier	Top 25 Hottest Articles	下载最多的25篇论文，按季度统计
Wiley	Most Accessed	下载最多的10篇论文，按月统计
Sage	Most Read	下载最多的50篇论文，按月统计
Taylor & Francis	Most read articles	下载最多的20篇论文，按天统计
MIT Press	Most Downloaded Articles	下载最多的20篇论文，按天统计
ACM DL	Bibliometrics	每篇论文均提供被下载情况的数据，6周统计
PLOS	Metrics	每篇论文均提供被下载情况的数据，按天统计
Nature	Top content	被下载、被email转发最多的论文统计，按天统计
Nature	Nature metrics	每篇论文均提供被下载情况的数据，按天统计
PNAS	Metrics	每篇论文均提供被下载情况的数据，按月统计

因此，识别模型通过三个层次来基于科学论文的用户下载使用数据和替代计量数据进行研究前沿识别：主题词隶属于近年新发表的研究文章或者新出现的学术讨论；主题词被下载和被收藏转发的次数总和相对较高并达到一定阈值；研究前沿的识别系数数值较大并达到一定阈值。

2.3 研究前沿集成识别模型的框架流程 研究前沿集成识别模型的框架和具体流程见图1。“神经网络计算”领域研究前沿的集成识别结果（Research Fronts）即来自基于发文引用数据和基于下载使用、替代计量数据识别结果的综合集成。另外，为了使识别模型的最终识别结果（Research Frontier）与科学家眼中的实际科学前沿更为吻合，也为了从文末引文与关键词/主题词的考究回归到最新施引文献的“学术价值预判”，识别模型在集成识别结果的基础上，引入“神经网络计算”领域 WOS 论文数据施引文献的被引频次、领域同行专家研读判断两种方式，用以保证集成识别模型对 Research Frontier 的最终可靠识别和确认。

3 讨 论

3.1 可能存在的误差和局限 虽然基于科学论文的多源数据、综合运用多种方法和工具，可以集合各种

数据方法的优势、弥补各种数据方法的不足,提高最终识别结果的准确度和可靠性,但也存在很多可能引起误差的因素和局限。

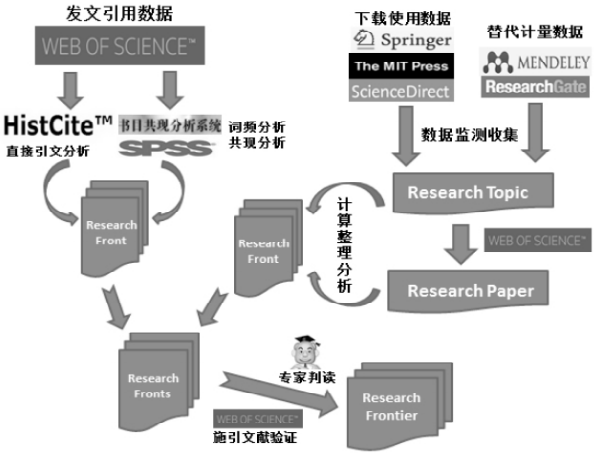

图 1　研究前沿集成识别模型框架流程

集成识别方法与数据来源本身相对复杂,运行工作量较大,收集数据的准确性和完整性要求较高; 利用 HistCite 进行引文分析可以展现各年代文献之间的引证关系,却无法反映文献内容之间的关联程度; 词频分析操作简便,能直观揭示研究领域的内容特征,但词频具有时间波动性,可能由于人为阈值选择而遗漏; 共词分析更关注词间关系,减少了人为干预,但无法反映尚未形成热点的潜在前沿主题,脱离了语境的关键词不能完全表达词间语义关系; 科学论文的下载使用数据,可以克服传统识别方法的时间滞后性等明显缺陷,但很大程度上依赖网络出版商的用户使用数据提供程度,而目前提供这项服务的出版商只有少数,提供的数据也不够全面; 科学论文的替代计量数据是对传统科学共同体的有力数据补充,可以提前挖掘科研大众对某科研主题的实时态度,但此类数据获取难度大、技术要求高; 利用 WOS 数据施引文献的被引频次有利于对 Research Frontier 的最终准确识别,但由于出版商提供数据的限制,获得难度较大。

3.2　进一步的研究展望　探讨构建了一种利用科学论文的发文引用数据、下载使用数据、替代计量数据等多种数据,结合引文分析、词频分析、共词分析等多种方法的研究前沿集成识别模型,并辅以专家判读、施引文献被引次数等方式保证最终识别结果的准确可靠输出。能够对“神经网络计算”领域的研究前沿进行不同层面的识别和探测,且有利于对该领域前沿方向和发展趋势予以最终识别确认。基于引文分析、词频分析、共词分析的研究前沿识别探测并不十分鲜见,已有很多学者进行过相关实践,但作为本文的主要创新点,基于论文下载使用数据与替代计量数据的前沿识别方法既是亮点,又是进一步实证的重点和难点。

下载使用数据需要依赖学术出版商的完全公开支持,虽然国内 CNKI 已经提供论文的下载数据,国外类似 PLoS、Science 等出版商也开始提供不同格式的详尽用户使用数据,但毕竟只占少数,且提供的下载数据规模、格式、内容也十分有限。因此在进一步实证研究时,要不断调试并选取最为适合和可用的论文下载数据供应平台,此外,也可以尝试联系学术出版商,利用签订科研合作协议等方式获取全面详细的用户动态下载使用数据,开展更为深入具体的科学实证研究。

替代计量数据又可细分为 4 类: Facebook、Twitter、新浪微博等大众社交媒体数据,报纸、门户网站等传统主流媒体数据,Mendeley、ResearchGate、科学网等学术社交媒体数据,新浪博客、腾讯博客等网络博客媒体数据[32]。除学术社交媒体数据能用于识别研究前沿外,其他类型替代计量数据也涵盖着大量的科学前沿信息。作为传统论文数据的有力补充,替代计量数据在研究前沿识别上蕴含着巨大潜力,虽然现阶段对其进行挖掘、抽取与利用的难度较大,但相信随着学术出版商与网络媒介平台的持续紧密合作,外加本体、语义网、大数据、云技术、移动互联等网络信息技术的不断升级和革新,替代计量数据必将在研究前沿的探测识别上发挥更大的实际效用。

另外,近年来兴起的利用科技规划文本数据、项目数据等进行科技前沿识别的方法,可以克服科学论文下载使用数据、替代计量数据存在的“是否被科学共同体最终真实应用”的诸多疑问和局限,也可以在下一步研究实践中予以补充并加以应用,使研究前沿集成识别模型的识别结果更加客观、准确和可靠。

参 考 文 献

[1] 陈仕吉. 科学研究前沿探测方法综述[J]. 现代图书情报技术, 2009(9): 28 - 33.

[2] Price D J D. Networks of scientific papers[J]. Science, 1965, 149(3683): 510 - 515.

[3] Small H. Co - citation in the scientific literature: A new measure of the relationship between two documents[J]. Journal of the American Society for Information Science, 1973, 24(4): 265 - 269.

[4] Persson O. The intellectual base and research fronts of JASIS 1986 - 1990[J]. Journal of the American Society for Information Science, 1994, 45(1): 31 - 38.

[5] Garfield E. Research fronts[J]. Current Contents, 1994, 41(10): 3 - 7.

[6] Morris S A, Yen G, Wu Z, et al. Time line visualization of research fronts[J]. Journal of the American Society for Information Science and Technology, 2003, 54(5): 413 - 422.

[7] Chen C. Citespace II: Detecting and visualizing emerging trends and transient patterns in scientific literature[J]. Journal of the American Society for Information Science and Technology, 2006, 57(3): 359 - 377.

[8] 张丽华. 研究前沿探测及其演化分析方法与实证研究 [D]. 北京: 中国科学院大学,2015.

[9] 钟 镇. 从高被引与零被引论文的引文结构差异看 Research Front 与 Research Frontier 的区别 [J]. 图书情报工作,2015,59(8):87 - 96.

[10] 刘小平,冷伏海,李泽霞. 国际科技前沿分析的方法和途径 [J]. 图书情报工作,2012,56(12):60 - 65.

[11] Shibata N,Kajikawa Y,Takeda Y,et al. Detecting emerging research fronts in regenerative medicine by the citation network analysis of scientific publications [J]. Technological Forecasting and Social Change,2011,78(2):274 - 282.

[12] Liu J S,Lu L Y Y,Lu W M. Research fronts in data envelopment analysis [J]. Omega,2016,58:33 - 45.

[13] Ding W,Chen C. Dynamic topic detection and tracking: A comparison of HDP,c - word,and cocitation methods [J]. Journal of the Association for Information Science and Technology,2014,65(10):2084 - 2097.

[14] Zhao D,Strotmann A. The knowledge base and research front of information science 2006 - 2010: An author cocitation and bibliographic coupling analysis [J]. Journal of the Association for Information Science and Technology,2014,65(5):995 - 1006.

[15] 马瑞敏,倪超群. 作者耦合分析: 一种新学科知识结构发现方法的探索性研究 [J]. 中国图书馆学报,2012(2):4 - 11.

[16] Huang M H,Chang C P. Detecting research fronts in OLED field using bibliographic coupling qith sliding window [J]. Scientometrics,2014,98(3):1721 - 1744.

[17] Lucio - Arias D,Leydesdorff L. An indicator of research front activity: measuring intellectual organization qs uncertainty reduction In document sets [J]. Journal of the American Society for Information Science and Technology,2009,60(12):2488 - 2498.

[18] Lv P H,Wang G F,Wan Y,et al. Bibliometric trend analysis on global graphene research [J]. Scientometrics,2011,88(2):399 - 419.

[19] Ohniwa R,Hibino A,Takeyasu K. Trends in research foci in life science fields over the last 30 years monitored by emerging topics [J]. Scientometrics,2010,85(1):111 - 127.

[20] 叶春蕾,冷伏海. 基于共词分析的学科主题演化方法改进研究 [J]. 情报理论与实践,2012,35(3):79 - 82.

[21] Small H G. A co - citation model of a scientific specialty: a longitudinal study of collagen research [J]. Social Studies of Science,1977:139 - 166.

[22] Chen C. Measuring The movement of a research paradigm [C] // Electronic Imaging 2005. International Society For Optics And Photonics,2005:63 - 76.

[23] Tu Y N,Seng J L. Indices of novelty for emerging topic detection [J]. Information Processing & Management,2012,48(2):303 - 325.

[24] Mccain K W. Assessing an author's influence using time series historiographic mapping: The oeuvre of conrad hal waddington (1905 - 1975) [J]. Journal of the American Society for Information Science and Technology,2008,59(4):510 - 525.

[25] 张英杰. 科技领域前沿计量探测方法研究 [D]. 北京: 中国科学院研究生院,2011.

[26] 王莉亚. 基于离群数据的主题演化研究 [D]. 北京: 中国科学院研究生院,2012.

[27] 方 丽,崔 雷. 利用双聚类算法探测学科前沿及知识基础——以 h 指数研究领域为例 [J]. 情报理论与实践,2014,37(11):55 - 60.

[28] 白如江. 基于语义计算的科学研究前沿识别研究 [D]. 北京: 中国科学院大学,2015.

[29] Wang X,Wang Z,Xu S. Tracing scientist's research trends realtimely [J]. Scientometrics,2013,95(2):717 - 729.

[30] 王贤文,毛文莉,王 治. 基于论文下载数据的科研新趋势实时探测与追踪 [J]. 科学学与科学技术管理,2014,35(4):3 - 9.

[31] 王立学,冷伏海. 简论研究前沿及其文献计量识别方法 [J]. 情报理论与实践,2010(3):54 - 58.

[32] 王贤文,方志超,胡志刚. 科学论文的科学计量分析: 数据、方法与用途的整合框架 [J]. 图书情报工作,2015,59(16):74 - 82.

[33] Van Den Besselaar P,Heimeriks G. Mapping research topics using word - reference co - occurrences: A method and an exploratory case study [J]. Scientometrics,2006,68(3):377 - 393.

[34] Zitt M,Lelu A,Bassecoulard E. Hybrid citation - word representations in science mapping: Portolan charts of research fields? [J]. Journal of the American Society for Information Science and Technology,2011,62(1):19 - 39.

[35] 宫 雪,崔 雷. 利用不同类型引文探测研究前沿及比较研究 [J]. 中华医学图书情报杂志,2010,19(4):8 - 10,31.

[36] Shibata N,Kajikawa Y,Takeda Y,et al. Comparative study on methods of detecting research fronts using different types of citation [J]. Journal of the American Society for Information Science and Technology,2009,60(3):571 - 580.

[37] Boyack K W,Klavans R. Co - citation analysis,bibliographic coupling,and direct citation: Which citation approach represents the research front mosta accurately? [J]. Journal of the American Society for Information Science and Technology,2010,61(12):2389 - 2404.

[38] Garfield E,Paris S,Stock W G. Histcite[tm]: A software tool for informetric analysis of citation linkage [J]. Information - wissenschaft & Praxis,2006,57(8):391.

[39] 崔 雷,刘 伟,闫 雷,等. 文献数据库中书目信息共现挖掘系统的开发 [J]. 现代图书情报技术,2008,24(8):70 - 75.

[40] Brody T,Harnad S,Carr L. Earlier web usage statistics as predictors of later citation impact [J]. Journal of the American Society for Information Science and Technology,2006,57(8):1060 - 1072.

[41] Jahandideh S,Abdolmaleki P,Asadabadi E B. Prediction of future citations of a research paper from number of its internet downloads [J]. Medical Hypotheses,2007,69(2):458 - 459.

(责编: 刘影梅; 校对: 王平军)