

基于词频突变和专家研判的研究前沿识别^{*}

——以光学学科为例

次雨桐¹ 黄进¹ 何益华¹ 方吉¹ 张虎²

(1. 华中科技大学图书馆, 武汉 430074; 2. 华中科技大学光学与电子信息学院, 武汉 430074)

摘要: 把握学科研究前沿是科研人员和科研管理者十分关注的问题, 本文提出从学科研究热点中识别研究前沿的方法, 帮助研判学科未来发展方向, 以期为学科发展决策提供参考。首先基于高被引论文, 采用共被引聚类算法来识别研究主题, 从中遴选出研究热点; 再获取与研究热点相关联的施引文献, 采用突变词检测算法结合专家研判来识别学科研究前沿; 运用此方法, 以光学学科为例, 从遴选出的9个光学研究热点中识别出若干学科研究前沿, 并借助资料验证法对识别结果的可靠性进行评估。结果表明: 本文提出的方法具有一定的可行性、前瞻性和可靠性, 识别结果能够为决策者提供参考。

关键词: 研究前沿; 共被引分析; 词频突变; 专家研判; 光学

中图分类号: G35 **DOI:** 10.3772/j.issn.1673-2286.2021.07.008

引文格式: 次雨桐, 黄进, 何益华, 等. 基于词频突变和专家研判的研究前沿识别——以光学学科为例[J]. 数字图书馆论坛, 2021 (7) : 54-61.

学科研究前沿是科研产生重大突破、引领科技发展的关键。世界各国从国家战略高度支持前沿研究, 以促进科技创新与发展。准确、及时识别并抢占研究前沿, 既是国家战略的宏观需求, 也是科研管理人员制定学科发展策略的根据, 又是科学研究人员跟踪创新趋势、把握学科发展方向的微观需要。然而, 如何快速、精准地识别研究前沿, 以尽早布局学科未来发展计划是目前科学难点和重点。

目前, 已有学者在研究前沿识别方面做出了贡献, 但仍然存在一些未解决、未明确的问题, 如研究前沿如何产生^[1]和如何发展演化^[2]。这些问题的解决需要相关理论的完善和方法的创新。基于此, 本文提出一种引文分析-文本分析-专家研判相结合的方法, 从研究热点中识别研究前沿, 并对此方法的可靠性进行评估、验证, 以期研究前沿的识别方法提供一个新的路径。

1 研究前沿的概念与识别方法

1.1 研究前沿的概念辨析

1965年, Price^[3]首次提出“研究前沿”的概念: “卓越科学家进行的领先研究, 是正在开发的研究领域”。此后, 诸多学者都提出了对研究前沿的理解。Small等^[4]认为研究前沿是一组高被引论文。Persson^[5]将研究前沿定义为与一组高被引论文相关联的施引文献。Kessler^[6]、Morris等^[7]认为研究前沿是一组具有耦合关系的文献。Garfield^[8-9]将研究前沿定义为一组高被引论文及其施引文献, 指出科学文献的引用分析能够跟踪并发现科学研究的新兴领域。Chen^[10]认为研究前沿是新兴的研究主题, 是一组增长率明显变化的突变词。目前, 学术界对研究前沿的定义还没有形成共识, 这一点可以从研究前沿、研究热点和新兴趋势等概念的交替使用上看出。但毋庸置疑的是, 研究前沿应该具有新颖性, 与新兴领域、新兴趋势和新兴主题在内涵上相差

^{*} 本研究得到国家自然科学基金青年项目“高强铝锂合金的激光选区熔化成形关键技术基础研究”(编号: 51805184)资助。

不大。

然而, 以往的文献中较少对研究热点与研究前沿的概念进行辨析, 常出现定义模糊、概念交替使用、将研究热点和研究焦点等误认为是研究前沿的情况。一般来说, 研究热点是某领域目前发展较热、关注度较高的研究主题, 与研究焦点内涵相同。而研究前沿是正在兴起、暂未引起广泛关注但极具突破性和引领性的研究。两者存在明显的区别并可以动态转化, 研究热点中可能发展出多个有潜力的前沿方向。另外, 随着研究前沿得到越来越多的同行关注和认可, 新的研究者加入后, 研究前沿也会转变成研究热点。对于学科研究前沿, 本文认为其是学科领域中近3年兴起的最具引领性、突破性和发展潜力的新进展、新成果和新问题。

1.2 研究前沿的识别方法

对研究前沿的内涵理解不同, 便产生了不同的识别方法。目前有4类常用的研究前沿识别方法, 分别是引文分析、文本分析、专家研判和复合分析法^[11]。

引文分析是最传统、最经典的方法, 已逐步应用到实际工作中, 主要包括共被引分析法、耦合分析法和直接引用法。Small首次提出文献共被引的概念, 以测度文献间关系程度。当两篇论文同时被一篇文献引用, 认为这两篇论文共被引。Small认为通过统计文献共被引频次可以确定领域核心文献集, 共被引强度高的核心文献之间内容相似度高, 由此聚类形成的主题代表领域研究前沿。Garfield、Price和Small等最早通过共被引分析, 呈现科学结构和脉络, 跟踪科研进展; 李小涛等^[12]对医学信息学高被引论文进行共被引聚类, 识别出5个前沿方向; 中国科学院发布的《研究前沿》报告也采用共被引分析法识别学科研究前沿, 获得广泛关注。但共被引分析法存在引文时滞性问题, 导致识别出的研究前沿可能已经不能代表最新的研究方向。文献耦合的概念由Kessler提出, 认为两篇引用同一篇论文的论文称为耦合论文, 文献耦合强度越高, 主题相关性越高。此方法被Glänzel等^[13]、Morris等^[7]用来识别学科研究前沿, 并取得较好效果。文献耦合分析法采用施引文献, 在一定程度上克服了共被引分析法的时滞性问题, 但可能存在引用目的不同导致识别结果不精确的问题^[14]。直接引用法是将具有引用关系的相似文献聚类, 揭示领域的发展脉络, 这些新出现的并形成一定规模的相似文献的集合可以表征研究前沿。Shibata等^[15-16]利用直接引

用法识别研究前沿的文献, 并在氮化镓和复杂网络等领域进行了实证研究; 2017年, 爱思唯尔公司采用此方法, 对Scopus中的论文聚类, 识别出近9.6万个研究主题, 北京理工大学在此基础上遴选出材料科学领域研究前沿^[17]。目前应用直接引用法相对较少, 且统计过程较为复杂。

文本分析主要包括共词分析法、主题概率模型法和突变词检测算法3种。共词分析法在20世纪70年代被提出^[18], 认为一组相关联的词汇可以代表研究前沿的主题内容。该方法通过统计词汇在同一篇文献中同时出现的次数来计算关系强度, 进行聚类, 进而分析共词文献簇代表的主题, 从而识别学科前沿。曾硕勋等^[19]、张斌等^[20]分别采用此方法探析富勒烯和档案学领域研究前沿; 张洋等^[21]提出了共现网络构建模型。共词分析法由于缺少对低频词的关注, 可能无法识别新兴研究前沿^[22]。主题概率模型法(LDA)是Blei等^[23-24]提出的一种建立在概率层次下的主题识别方法, 通过对文本进行全文语义分析, 探测有价值的潜在主题。该方法利用文献中的词汇来生成主题, 同时揭示了词-主题-文献关系, 可以有效表达词汇之间语义关系, 有助于揭示研究前沿的内容。唐恒等^[25]、刘忠宝等^[26]运用LDA模型识别智能网联汽车和无人机领域新兴技术。突变词检测算法最早由Kleinberg^[27]提出, 根据研究前沿出现往往伴随词频改变这一规律现象, 该方法找出各时间段中增长率激增的突变词, 通过分析突变词状态的变化来发现前沿。CiteSpace软件中的Burst Detection功能采用该算法检测突变词, 识别领域研究前沿; 胡静等^[28]利用此方法, 挖掘阅读推广的前沿领域; 王曰芬团队^[29-32]基于国家、机构和作者等不同视角利用突变词检测算法探析人工智能研究前沿。突变词检测算法和主题概率模型法对表现形式有差异的相同概念认定为不同概念, 导致结果准确率降低。

专家研判是另一种识别研究前沿的有效方法, 凭借专家智慧、知识和经验, 综合多位专家的观点对问题作出判断和评估。如国际顶级期刊《科学》通过咨询专家建议, 提出125个前沿问题; 中国工程院发布的《全球工程前沿》报告, 主要以专家研判为核心, 融合数据分析, 遴选全球工程研究前沿。单纯依靠专家研判时, 费时费力且效率低下, 并容易受专家个人主观认识影响^[33]。

为了从多角度识别研究前沿, 有些学者提出了复合分析法, 通过将引文分析、文本分析、专家研判等互相组合分析, 以更准确定位研究前沿, 如Braam等^[34]、孙震

等^[35]在共被引分析的基础上结合共词分析,分析研究前沿的演变规律;王兴旺等^[36]结合词频分析和专家研判确定儿童早期发展研究前沿。

本文提出一种基于引文分析-文本分析-专家研判相结合的新思路,将共被引分析、突变词检测算法和专家研判3种方法有机结合,通过共被引分析法关联相似文献,揭示研究主题和热点,利用突变词检测算法识别研究热点发展演变过程中新出现的突发主题,改善共被引分析法的滞后性,再引入专家研判法,使得识别结果更加可靠。这种复合性分析方法发挥了单一方法的优势,同时弥补了单一方法存在的滞后性、精确性差等缺点,有助于科学、精确地识别研究前沿。

2 基于词频突变和专家研判的研究前沿识别方法

2.1 数据来源

在以往的研究中,通常基于某一特定学科领域的论文探测学科研究前沿。然而,在跨学科研究热潮之下,学科研究前沿往往产生于学科交叉融合处。因此,从各学科高被引论文的关联中更容易产生研究热点和前沿突破。基于对学科研究前沿的概念界定,本文尝试将研究热点作为知识基础,通过追踪研究热点的后续进展,从研究热点中探测研究前沿。

本文数据来自Web of Science核心合集的SCIE和SSCI数据库,数据检索时间为2020年12月。首先,获取

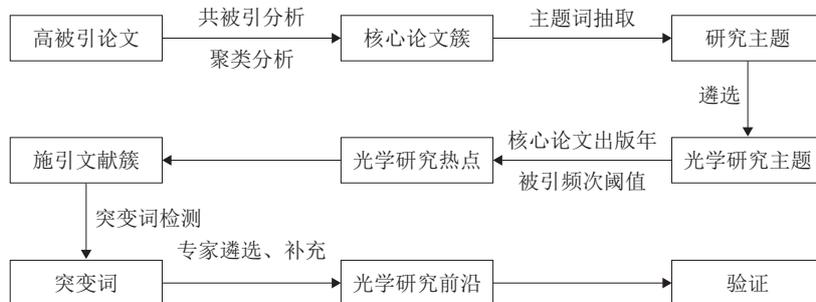


图1 学科研究前沿识别方法与过程

2.2.1 基于共被引关系的主题聚类算法

研究主题形成首先通过共被引关系形成共被引论文对,然后将共被引论文对聚类成不同文献簇(研究主题)。当两篇论文被一篇文献同时引用,则这两篇论

2012—2017年出版的所有学科的高被引论文(被引频次在同学科领域、同出版年排名前1%的论文),通过聚类筛选出研究主题,每个研究主题包含一组高被引论文;其次,根据高被引论文所属学科,筛选出包含光学(optics)学科高被引论文的研究主题,根据高被引论文的总被引频次和平均出版年2个指标识别出最具影响力且最“年轻”的研究主题作为光学研究热点;再次,获取光学研究热点中高被引论文的施引文献(施引文献出版年为2012—2020年),从施引文献中检测出在近3年(2018—2020年)产生突变的突变词;最后,结合专家研判识别出学科研究前沿。

考虑到后期需要对研究前沿识别结果可靠性进行验证,故将高被引论文出版年截止日期设为2017年,将2012—2017年出版的高被引论文作为知识基础,将其在2018—2020年的最新进展作为研究前沿。

2.2 学科研究前沿识别方法与过程

本文通过对高被引论文进行共被引分析,聚类出研究热点后,追踪研究热点的后续发展状态,对高被引论文的施引文献进行突变词检测,得到一组突变词。突变词可能变成研究前沿,也有可能趋弱或沉寂^[37]。基于此考虑,本文在定量方法基础上,辅助专家研判,通过定量、定性相结合,集成共被引分析方法-突变词检测算法-专家研判法,从学科研究热点中识别研究前沿,将在更大程度上保证识别结果的可靠性(见图1)。

文共被引,形成一个共被引论文对。创建共被引论文对后,将共被引频次进行归一化,采用余弦相似性计算共被引相似度,再通过单链接算法将各个共被引论文对进行聚类。

本文设置共被引相似度的阈值为0.2,从某一个满

足阈值(共被引相似度大于或等于0.2)的共被引论文对(如C1)开始,将其作为初始集合,找到与初始集合有共被引关系的其他共被引论文对(如C2),计算两个集合的最相似成员之间的相似度,相似度大于或等于0.2,则合并两个集合,一直持续到没有论文可以添加到集合中,由此产生的集合形成一个聚类,即研究主题。聚类时设置每个聚类或研究主题至少包含2篇高被引论文,最多包含50篇高被引论文。共被引论文对C1和C2的相似度 $S(C1, C2)$ 由两个集合中的最相近论文的相似度决定。

2.2.2 基于施引文献的突变词检测算法

Kleinberg提出突变词检测算法,检测使用频次在某些出版年份骤增或在较短时间内突然出现的术语。CiteSpace软件中的Burst Detection模块集成了该算法的功能,本文使用CiteSpace软件,从高被引论文的施引文献中初步识别出一组突变词,基于突变词的突变时间段初步筛选出候选的学科研究前沿列表。

2.2.3 专家研判法

专家研判法作为一种定性分析法,是对共被引分析和突变词检测算法等定量分析法的重要补充,是识别研究前沿的重要手段。领域专家能够较为准确地把握学科领域发展方向,因此,本文邀请每个研究热点相关领域的多位专家对突变词检测算法识别出的候选研究前沿进行判断、甄别、补充,得到最终的学科研究前沿。

3 光学学科研究前沿识别

3.1 研究热点

首先,基于Web of Science中SCIE和SSCI数据库收录的2012—2017年出版的高被引论文,计算高被引论文之间的相似度,利用单链接算法进行聚类,形成研究主题,每个研究主题包含一组高被引论文;其次,从研究主题中筛选出包含“optics”学科类别高被引论文的研究主题200个,作为光学研究主题;再次,将200个光学研究主题,按照高被引论文总被引频次进行降序排列,提取排在前20%的研究主题,再根据高被引论文的

平均出版年降序排列,遴选出高被引论文的平均出版年份在2014年及其之后的研究主题,得到9个光学研究热点;最后,对9个光学研究热点包含的高被引论文的题目进行分析,进而命名研究热点。9个光学研究热点分别为有机聚合物太阳能电池、宇称-时间对称、等离激元光子学、二维材料、光学相干断层成像、有机发光二极管、量子点太阳能电池、量子相干、量子自旋。其中,最小的一个聚类是有机聚合物太阳能电池研究热点,包含3篇高被引论文;最大的一个聚类是光学相干断层成像研究热点,包含39篇高被引论文。

每个研究热点包含一组高被引论文,每篇高被引论文在Web of Science核心合集数据库中都有一个唯一的入藏号,即UT号。将每个研究热点中包含的高被引论文的UT号之间用“OR”连接,构建检索式,在Web of Science核心合集的高级检索功能中进行检索。得到检索结果后,对这组高被引论文进行引文分析,通过Web of Science核心合集中的“引文报告”功能,获得每个研究热点中包含的高被引论文的施引文献,并下载施引文献的全部著录项,包含标题、摘要、关键词、出版年、来源出版物和UT号等多个字段。依据此方法,分别获取光学9个研究热点的施引文献数据集。

3.2 研究前沿初步识别

因有机聚合物太阳能电池研究热点体量较小,更容易阐述,故本文以有机聚合物太阳能电池这一个研究热点为例,展示研究前沿的初步识别、命名方法和专家研判过程。

首先,获取有机聚合物太阳能电池研究热点涉及的3篇高被引论文的施引文献,共2 071篇。通过Kleinberg提出的突变词检测算法,从施引文献的标题、摘要和关键词中初步识别得到突变词36个。其次,剔除无意义和较宽泛的突变词。再次,筛选突变时间在2018—2020年的词。最后,将突变词进行整理、合并,得到9个突变词,即non-fullerene(非富勒烯)、power conversion efficiency(能量转换效率)、all-polymer solar cell(全聚合物太阳能电池)、stability(稳定性)、flexible solar cell(柔性太阳能电池)、ternary polymer solar cell(三元有机太阳能电池)、bulk heterojunction solar cell(异质结太阳能电池)、interfacial material(界面材料)、active layer(活性层)。其中,有些突变词需要经过整理或合并之后再请专家研判,例如13 percent

efficiency和11 percent efficiency这两个突变词都表示有机聚合物太阳能电池的能量转换效率,因此可以合并为power conversion efficiency。

3.3 专家研判确定研究前沿

将9个突变词(候选研究前沿)通过邮件发送给有机聚合物太阳能电池领域的3位专家,当一个突变词获

得2位及以上专家认可,本文即认为是研究前沿;同时,也接受专家提名研究前沿,如果一个突变词获得2位及以上专家提名,本文即认为是研究前沿。最终,从有机聚合物太阳能电池研究热点中识别出8个研究前沿(见表1)。随后,在Web of Science数据库中检索有机聚合物太阳能电池8个研究前沿的相关文献,统计从2012—2020年的逐年发文量,结果见图2。

表1 专家研判的有机聚合物太阳能电池的研究前沿

序号	突变词	专家1	专家2	专家3	研究前沿	研究前沿命名
A1	non-fullerene	✓	✓	✓	是	非富勒烯聚合物太阳能电池
A2	power conversion efficiency	✓	✓	×	是	有机聚合物太阳能电池的能量转换效率
A3	all-polymer solar cell	✓	×	✓	是	全聚合物太阳能电池
A4	stability	✓	✓	✓	是	有机聚合物太阳能电池的稳定性
A5	flexible solar cell	✓	✓	✓	是	柔性有机太阳能电池
A6	ternary polymer solar cell	×	✓	✓	是	三元有机太阳能电池
A7	bulk heterojunction solar cell	✓	×	×	否	异质结太阳能电池
A8	interfacial material	✓	×	×	否	界面材料研究
A9	active layer	×	×	×	否	活性层形态
B1	energy loss in polymer solar cell	提名		提名	是	有机聚合物太阳能电池内部的能量损失
B2	tandem solar cells	提名		提名	是	叠层太阳能电池

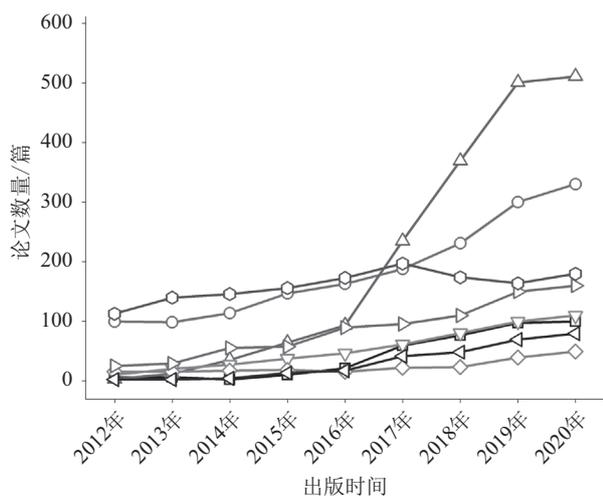


图2 有机聚合物太阳能电池领域研究前沿

从表1可知,3位专家均认为非富勒烯聚合物太阳能电池(A1)、有机聚合物太阳能电池的稳定性(A4)、柔性有机太阳能电池(A5)是有机聚合物太阳能电池的研究前沿。从图2文献统计趋势发现,非富勒烯聚合物太阳能电池(A1)是近年来有机聚合物太阳能电池领域中最热的前沿方向。不同的非富勒烯材料以及给体材料的合成是近年来有机聚合物太阳能电池能量转换效率(A2)提升最主要的原因。在2016年以前,聚合物太阳能电池的受体主要以富勒烯为主,电池能量转换效率在10%徘徊。2016年,基于非富勒烯材料ITIC与聚合物给体PBDB-T的电池效率突破11%之后,不同的非富勒烯受体以及与之相匹配的给体相继问世,在短短的3年时间里,有机聚合物太阳能电池的效率达到17.4%。有机聚合物太阳能电池的稳定性(A4)也是目前研究重点。为了实现应用,电池的稳定性直接决定了太阳能电池的使用寿命。另外,相较于其他种类电池,有机太阳能电池机械性能高、质量轻,使得它在柔性器件方面有着天然优势,高效率柔性器件应用广泛,商业化前景强,因而柔性有机太阳能电池(A5)被广泛研究。

另外, 获得了两个由专家提名的研究前沿, 分别是有机聚合物太阳能电池内部的能量损失 (B1) 和叠层太阳能电池 (B2), 从文献量来看, 两个研究前沿的文献量逐年上升。

研究前沿的命名不仅考虑突变词本身的含义, 也要考虑其与研究热点的关联。例如, 对于突变词 non-fullerene, 本义是非富勒烯, 定位其关联的论文并阅读论文标题, 结合网络搜索和专家意见, 确定该突变词代表

的研究前沿应命名为“非富勒烯聚合物太阳能电池”。

3.4 研究前沿汇总

获取光学研究热点的施引文献, 从中检测突变词, 并结合专家研判, 最终识别出51个光学研究前沿, 具体内容见表2。

表2 施引文献突变词检测结合专家研判决别出的光学研究前沿

研究热点	研究前沿	研究前沿数量/个
有机聚合物太阳能电池	非富勒烯聚合物太阳能电池; 有机聚合物太阳能电池的能量转换效率; 全聚合物太阳能电池; 有机聚合物太阳能电池的稳定性; 柔性有机太阳能电池; 三元有机太阳能电池; 有机聚合物太阳能电池内部的能量损失; 叠层太阳能电池	8
宇称-时间对称	光的非互易传输; 超表面; 纳米光子学; 手性; 回音壁模式光学微腔; 光学频率梳; 拓扑光子学; 高阶拓扑模式	8
等离激元光子学	表面等离激元微纳结构光热效应; 海水淡化; 表面等离激元热电子; 等离激元超材料; 等离激元纳米结构光催化剂	5
二维材料	二维材料可饱和吸收体的锁模光纤激光器; 黑磷的大规模制备及其光电应用; 大尺寸石墨烯单晶薄膜的可控制备及转移技术; 二维IV-VI族半导体在光电子器件中的应用; 过渡金属硫化物的光电性能调控与应用	5
光学相干断层成像	光学相干断层扫描血管造影; 扫频源相干光断层扫描; 眼科成像; 糖尿病性视网膜病变; 深度学习算法	5
有机发光二极管	高效纯蓝TADF; 高效红色TADF发光材料; 钙钛矿量子点; QLED; 蓝色OLED材料	5
量子点太阳能电池	量子点太阳能电池的能量转化效率; 多重激子效应; PbSe量子点不稳定性的改善; 通过配体工程调控量子点的表面状态; 碘化铅配体钝化	5
量子相干	量子相干度量; 量子相干动力学研究; 量子相干在光合作用能量传输中的作用; 量子计算机; 量子退相干管控	5
量子自旋	量子自旋液体材料; 拓扑绝缘体和拓扑表面态研究; 量子精密测量; 多体局域化系统的研究; 量子多体系统的非平衡动力学	5

4 光学学科研究前沿识别结果的评估

在以往的一些研究中, 较少对研究前沿识别结果的可靠性进行评估。资料验证法指利用已公布的重要奖项或已发表的文献对研究结果的可靠性进行验证^[38]。例如, Small等^[39]利用诺贝尔奖、IEEE奖励计划和权威潜力研究方向列表 (如《科学》杂志公布的年度十大科学突破) 等作为参考, 评估识别的新兴话题, 验证结果显示, 1/3的识别结果可以得到相关奖项的支持。本文借鉴Small的资料验证法来评估研究前沿识别结果可靠性。

4.1 资料验证

本文搜集了美国光学学会发布的 *optics in 2019*、*optics in 2020* 等年度光学进展, 《科学》杂志每年度发

布的 *Science: Breakthrough of the year*, 诺贝尔物理学奖, 英国物理学会主办的《物理世界》每年公布的国际物理学年度十大突破, 丹尼斯·加博尔奖和欧洲物理学会菲涅尔奖等; 此外, 也搜集了《中国激光》杂志社“中国光学十大进展”和国家科学技术三大奖的获奖名单。获取上述所有奖项近3年的获奖名单作为参考。

4.2 结果分析

将表2中51个光学研究前沿与近3年获奖名单进行比对后发现, 在51个光学研究前沿中, 17个研究前沿得到相关的奖项支持, 即1/3的研究前沿可以得到相关的奖项支持。虽然这些奖项存在不完整性, 但可以在一定程度表明与研究前沿相关的研究进展已经获得较权威机构的认可。例如, “光的非互易传输”在2018年和2020年均入围美国光学学会年度光学进展。“量子计

算机”和“量子精密测量”相关研究成果分别入选2019年和2020年度国际物理学十大突破。

5 总结

本文提出了一种将共被引分析法、突变词检测算法和专家研判法相结合的复合分析法,来从研究热点中识别研究前沿,并以光学学科为研究目标,展示了识别学科研究前沿的路径和过程。本文的方法从理论上补充和丰富了研究前沿识别方法,具有一定的科学价值,并且在实践上具有较强的可操作性。从识别效果的评估和验证表明,该方法也具有一定的可靠性和认可度,能够对学科发展方向的研判提供参考。

同时,本文提出的方法也存在一定的局限性,如本文从高被引论文出发识别研究前沿,由于引文的时滞性,导致最终识别出的研究前沿是近年来正在兴起的研究主题,并不能预测未来的研究前沿;此外,没有与其他研究前沿识别方法的效果进行比较。这些问题都需要在后续进行更深入的研究。

参考文献

- [1] MARRONE M. Application of entity linking to identify research fronts and trends [J]. *Scientometrics*, 2020, 122 (1): 357-379.
- [2] XU Y, ZHANG S, ZHANG W, et al. Research front detection and topic evolution based on topological structure and the pagerank algorithm [J]. *Symmetry*, 2019, 11 (3): 310.
- [3] PRICE D. Networks of scientific papers [J]. *Science*, 1965, 149 (3583): 510-515.
- [4] SMALL H G, GRIFFITH B C. The structure of scientific literatures I: Identifying and graphing specialties [J]. *Science Studies*, 1974, 4 (1): 17-40.
- [5] PERSSON O. The intellectual base and research fronts of JASIS 1986—1990 [J]. *Journal of the Association for Information Science & Technology*, 1994, 45 (1): 19-27.
- [6] KESSLER M M. Bibliographic coupling between scientific papers [J]. *American Documentation*, 1963, 14 (1): 10-25.
- [7] MORRIS S A, YEN G, WU Z, et al. Time line visualization of research fronts [J]. *American Society for Information Science and Technology*, 2003, 54 (5): 413-422.
- [8] GARFIELD E. Research fronts [J]. *Current Contents*, 1994, 41 (10): 3-7.
- [9] GARFIELD E. Citation indexes for science: a new dimension in documentation through association of ideas [J]. *Science*, 1955, 122 (3159): 108-111.
- [10] CHEN C M. CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature [J]. *JASIST*, 2006, 57 (3): 359-377.
- [11] 黄晓斌, 吴高. 学科领域研究前沿探测方法研究述评 [J]. *情报学报*, 2019, 38 (8): 872-880.
- [12] 李小涛, 金心怡, 李艳, 等. 基于ESI高被引论文的医学信息学研究前沿可视化分析 [J]. *现代情报*, 2018, 38 (12): 120-125.
- [13] GLÄNZEL W, CZERWON H J. A new methodological approach to bibliographic coupling and its application to the national, regional and institutional level [J]. *Scientometrics*, 1996, 37 (2): 195-221.
- [14] BOYACK KW, KLAVANS R. Co-citation analysis, bibliographic coupling, and direct citation: Which citation approach represents the research front most accurately? [J]. *Journal of the American Society for Information Science & Technology*, 2010, 61 (12): 2389-2404.
- [15] SHIBATA N, KAJIKAWA Y, TAKEDA Y, et al. Detecting emerging research fronts based on topological measures in citation networks of scientific publications [J]. *Technovation*, 2008, 28 (11): 758-775.
- [16] 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.
- [17] 崔宇红, 王飒, 高晓巍, 等. 基于全域微观模型的研究前沿主题探测和特征分析 [J]. *图书情报工作*, 2018, 62 (15): 75-82.
- [18] LE MARC M, COURTIAL J P, SENKOVSKA E D, et al. The dynamics of research in the psychology of work from 1973 to 1987: From the study of companies to the study of professions [J]. *Scientometrics*, 1991, 21 (1): 69-86.
- [19] 曾硕勋, 冯敏, 姜玲. 材料科学研究热点与前沿挖掘——以富勒烯为例 [J]. *中国材料进展*, 2019, 38 (2): 161-166.
- [20] 张斌, 杨文. 中国档案学研究热点与前沿问题探讨 [J]. *图书情报知识*, 2020 (3): 28-40, 62.
- [21] 张洋, 赵镇宁. 共现科学知识图谱构建技术与工具研究 [J]. *图书情报知识*, 2019 (1): 119-129.
- [22] 华旦草, 安培浚, 肖仙桃. 基于文本内容的科学前沿探测方法研究进展 [J]. *现代情报*, 2021, 41 (1): 169-177.

- [23] BLEI D M, ANDREW Y, JORDAN M, et al. Latent Dirichlet allocation [J]. *Journal of machine learning research*, 2003, 3 (4): 993-1022.
- [24] BLEI D M, LAFFERTY J. Dynamic topic models [C] // *Proceedings of the 23rd International Conference on Achine Learning*. NewYork: ACM, 2006: 113-120.
- [25] 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.
- [26] 刘忠宝, 康嘉琦, 张静. 基于主题突变检测的颠覆性技术识别——以无人机技术领域为例 [J]. *科技导报*, 2020, 38 (20): 97-105.
- [27] KLEINBERG J. Bursty and hierarchical structure in streams [J]. *Data Mining and Knowledge Discovery*, 2003, 7 (4): 373-397.
- [28] 胡静, 李璐. 基于词频突变的我国阅读推广研究前沿挖掘 [J]. *情报科学*, 2017, 35 (10): 75-78.
- [29] 王曰芬, 曹嘉君, 余厚强, 等. 人工智能研究前沿识别与分析: 基于领域全局演化研究视角 [J]. *情报理论与实践*, 2019, 42 (9): 1-7.
- [30] 姜宇星, 王曰芬, 范丽鹏, 等. 人工智能研究前沿识别与分析: 基于主要国家(地区)对比研究视角 [J]. *情报理论与实践*, 2019, 42 (9): 8-15.
- [31] 邹本涛, 王曰芬, 曹嘉君, 等. 人工智能研究前沿识别与分析: 基于高产作者多属性综合研究视角 [J]. *情报理论与实践*, 2019, 42 (9): 22-27.
- [32] 范丽鹏, 余厚强, 姜宇星, 等. 人工智能研究前沿识别与分析: 基于高产机构对比研究视角 [J]. *情报理论与实践*, 2019, 42 (9): 16-21.
- [33] 王山. 研究前沿探测方法进展 [J]. *情报科学*, 2019, 37 (10): 164-169.
- [34] BRAAM R R, MOED H F, VAN RAAN A F J. Mapping of science by combined co-citation and word analysis. I. Structure aspects [J]. *Journal of the American Society for Information Science and Technology*, 1991, 42 (4): 233-251.
- [35] 孙震, 冷伏海. 一种基于知识元共现的ESI研究前沿知识演进分析方法 [J]. *情报学报*, 2018, 37 (11): 1095-1113.
- [36] 王兴旺, 余婷婷, 张云婷. 同行认可视角下的科学研究前沿探测方法研究 [J]. *情报杂志*, 2019, 38 (8): 63-67, 81.
- [37] 王梦婷. 基于突变检测的主题突变分析研究 [J]. *情报科学*, 2016, 34 (12): 36-39.
- [38] 卢超, 侯海燕, DING Y, 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.
- [39] SMALL H, BOYACK K W, KLAVANS R. Identifying emerging topics in science and technology [J]. *Research Policy*, 2014, 43 (8): 1450-1467.

作者简介

次雨桐, 女, 1993年生, 硕士, 馆员, 研究方向: 情报研究和学科服务, E-mail: cba@hust.edu.cn。

黄进, 女, 1971年生, 本科, 馆员, 研究方向: 情报研究和学科服务。

何益华, 女, 1969年生, 本科, 副研究馆员, 研究方向: 情报研究和学科服务。

方吉, 男, 1982年生, 本科, 馆员, 研究方向: 情报研究和学科服务。

张虎, 男, 1986年生, 博士, 讲师, 研究方向: 激光材料及加工。

Research Fronts Prediction Based on Burst Detection and Expert Judgment: Taking Optical Discipline as An Example

CI YuTong¹ HUANG Jin¹ HE YiHua¹ FANG Ji¹ ZHANG Hu²

(1. Library, Huazhong University of Science & Technology, Wuhan 430074, China; 2. School of Optical and Electronic Information, Huazhong University of Science & Technology, Wuhan 430074, China)

Abstract: Grasping discipline research fronts is a matter of great concern to researchers and scientific research managers. In this paper, research fronts come from research hotspots. This study aims to help judge development direction of discipline and provide a reference for decision-making of discipline development. Based on highly cited papers, the co-citation analysis method is used to identify research topics and hotspots. In addition, citing papers related to the research hotspots are obtained. This study proposes a burst detection algorithm and expert judgment based on citing papers for identifying research fronts. Several research fronts are identified from nine optical research hotspots. This study evaluates the reliability of the recognition results with data verification method. The results show that the method proposed in this study is feasible, forward-looking and reliable, which means the identification results can provide a reference for decision makers.

Keywords: Research Fronts; Co-citation Analysis; Burst Detection; Expert Judgment; Optics

(收稿日期: 2021-06-12)