

● 白如江¹, 冷伏海², 廖君华¹

(1. 山东理工大学科技信息研究所, 山东 淄博 255049; 2. 中国科学院科技战略咨询研究院, 北京 100190)

科学研究前沿探测主要方法比较与发展趋势研究^{*}

摘 要: [目的/意义] 梳理科学研究前沿的基本概念, 归纳目前科学研究前沿探测的主要方法以及这些方法的主要特点, 寻找影响科学研究前沿探测的主要因素并揭示未来发展趋势。[方法/过程] 利用文献调研方法对相关理论进行梳理分析。[结果/结论] 对比分析了目前科学研究前沿探测研究中主要使用的引文分析和主题词分析两种方法; 指出了目前存在的3个方面的问题, 即: 引文分析时滞性问题、缺乏语义信息支持问题以及分析数据源无法有效融合; 探讨了科学研究前沿探测未来发展方向。

关键词: 科学研究前沿; 研究方法; 发展趋势

Research on the Comparison of the Main Methods and Development Trends of Frontier Exploration in Scientific Research

Abstract [Purpose/significance] This paper reviews the basic concepts of the frontier of scientific research and sums up the main methods and their major features in current frontier of the scientific research. Based on that, the paper tries to find out the key factors which influence the frontier of the scientific research and show the development trends of the future. [Method/process] The paper uses literature investigation methods to sort out relevant theories. [Result/conclusion] The paper compares citation analysis method and subject words analysis method which are mainly used in the detection of the frontier of scientific study. Then, the paper points out three problems: time lag of citation analysis, deficiency of semantic information support and insufficient fusion of data source. At the same time, the paper discusses the future development of cutting-edge scientific research exploration direction.

Keywords: scientific research front; research method; development trend

《国家中长期科学和技术发展规划纲要(2006—2020)》把建设创新型国家作为面向未来的重大战略选择, 全面推进中国特色国家创新体系建设。中国科学院从2011年起正式实施《知识创新工程2020——科技创新跨越方案》(简称“创新2020”工程), 为创新型国家发展提供科技支撑^[1]。

在此背景下, 要实现以上科技发展目标, 支撑智库研判科技发展大势的情报分析能力尤为重要, 尤其是支持宏观科技决策、支撑重要领域科技创新的全局性、前瞻性、战略性的科技战略情报服务和研究。科学研究前沿探测研究是科技战略情报研究的重要方向之一。科学研究前沿探测更加注重揭示尽早发现、识别科技创新的新兴主题, 并评估其发展趋势, 从而为科技创新战略决策提供有效支撑。因此, 面对科技创新演变加剧和交叉融合加速的大环境, 如何有效地利用情报研究方法来尽早地识别科技创新

中的科学研究前沿, 评估未来发展趋势成为当前战略情报研究的重要任务之一。

本文全面梳理了目前科学研究前沿探测领域的主要理论和方法, 分析了目前存在的主要问题和挑战, 并在此基础上提出了未来的发展趋势。

1 科学研究前沿探测

对于什么是科研领域中的新鲜事物, 有多种说法, 如研究前沿 (Research Front)、新兴趋势 (Emerging Trend)、热点领域 (Hot Topic) 等。在科学研究前沿探测研究中, 对未来产生重大影响的科学研究领域, 往往表现出许多文献计量学特征。比如: 某一研究领域突然出现大量的新主题词; 某主题的文章发表数量在短时间内发生暴发式增长; 某研究领域的知识组织体系结构发生改变; 出现大量的代表新研究领域的新期刊; 主题词之间的关系或主题关系受到其他学科的影响而发生变化; 主题词的含义发生变化; 引文网络结构发生异常变化等。如何利用这些文献计量学特征进行科学研究领域方向识别, 许多专家学者进行了大量卓有成效的工作, 其中以利用引文分析和主题词分析研究最为常见。

^{*} 本文为国家社会科学基金项目“未来新兴科学研究前沿识别研究”(项目编号: 16BTQ083) 和教育部人文社会科学研究青年基金项目“基于引文内容分析的科技创新路径识别研究”(项目编号: 16YJC870008) 的成果之一。

1.1 引文分析法

1965 年,普赖斯观察到科学家似乎倾向于引用最新发表的文章,并将其称为即时因子 (Immediacy Factor)。普赖斯认为某个领域的研究前沿是由科学家积极引用的文章所体现的。稍早发表的一些论文会被一组新近发表的论文通过参考文献紧密联系起来,这些稍早发表的论文被看作是活跃的研究前沿^[2]。

自普赖斯提出该研究方法以来,许多科学家根据普赖斯的思想进行了深入研究和拓展,并逐步形成了引文分析方法体系。归纳起来主要有共被引分析和文献耦合分析两种方法。

1) 共被引分析方法。共被引分析方法 (Co-Citation),是指两篇以上论文共同被后来的一篇或多篇文献所引用,则称这两篇引文共被引。以引用它们的论文数量的多少为测度,这种测度称为共被引强度 (Co-Citation Strength)。共被引强度越大,即同时引用这两篇论文的文獻越多,说明它们之间的关系越密切。

Small 等在普赖斯的思想基础上,把科学研究前沿表述为同被引文献簇。提出了共被引聚类方法来分析特定领域内基础性的、较早发表的核心论文。他认为共被引文章聚类表征着当前活跃的研究领域^[3]。

加菲尔德提出共被引文献簇及其施引文献簇共同组成的文献簇表征出科学研究前沿^[4]。Persson 认为高频共被引文献簇的施引文献簇才是研究前沿^[5]。

近年来,共被引分析方法得到情报分析人员的广泛关注,并且在不同程度上改进了共被引分析方法,并逐步应用到实际工作中^[6-9]。比如:美国科技信息研究所 (Institute for Scientific Information, ISI) 利用共引方法在科学前沿探测中做了大量的工作。在 ISI 的基本科学指标库 (Essential Science Indicators, ESI) 中对高被引论文进行同被引聚类处理,将内容联系密切的高被引论文收敛为若干簇,对每簇内部的论文进行半自动处理,得到了相关前沿领域^[10]。

利用共被引分析进行科学研究前沿识别遇到的最大挑战就是时滞性。一篇文献从发表到有被引信息需要一定时间积累,所以会影响在科学研究前沿探测过程具有潜在的科学研究前沿识别。

2) 文献耦合分析方法。“文献耦合” (Bibliographic Coupling) 这一术语是 1963 年美国麻省理工学院教授 M. M. Kessle 最先提出来的。他在对《物理评论》 (Physical Review) 刊出的论文进行引文分析研究时发现,越是学科、专业内容相近的论文,它们参考文献中的相同文献的数量就越多^[11]。

Morris (2003) 将持续被一组固定的、与时间无关的

基本文章引用的大量文章定义为研究前沿。Morris 在文献耦合聚类的基础上添加了时间轴,绘制了研究前沿时间线索 (Time-Line) 可视图^[12]。该可视图可以展现出研究前沿主题的出现和消失、潜在新兴研究前沿主题及其基础文档。

文献耦合分析方法虽然在一定程度上克服了共被引分析方法时滞性的缺点,但是,在文献耦合进行研究前沿识别时遇到的主要问题是数据集的固定性。因为文献耦合分析不像共被引分析那样数据集随时间变化而不断变化,一篇文章一旦发表,它的参考文献是固定的,不会再有改变。所以在用文献耦合分析时就限制了在研究前沿主题演化方面的研究。另外一个问题是,两篇文献虽然引用了同一篇文献,但有可能这两篇文献引用的是文献的不同部分,也就是说引用的目的和在文章中发挥的作用是不同的。针对这些缺点,许多学者也对其进行了改进^[13]。

不管是共被引分析还是文献耦合分析,它们共同的缺点是无法自动对筛选出的论文进行主题描述,目前主要是通过筛选出论文题目或关键词来标识,或者是通过专家判读的方式进行。

1.2 主题词分析法

新兴的科学研究前沿主题往往表现出大量涌现的新主题词以及主题词含义的变化发展等。因此,尽早识别出文献集中大量突发的新主题词及相关属性的变化可以有效识别出研究前沿主题的诞生。基于主题词的研究前沿识别主要有以下 4 种方法:

1) 突发词检测算法。考虑到某一新兴科学研究前沿出现,可能会伴随着词频密度的改变,Kleinberg 于 2002 年提出了突发词检测算法 (Burst Detection)^[14]。突发词检测算法的基本思想是观察在一定时间内词的突然变化情况,根据词的大量突然涌现判定一个新兴研究主题的产生。陈超美教授在 CiteSpace II 软件中采用了 Kleinberg 突发词检测算法来获取某一研究前沿主题描述^[5]。Mane 等利用突发词算法选择高频词来做共词分析,并绘制了 PNAS 杂志从 1982 年到 2001 年发表的全部论文的知识图谱,借助该图谱来识别主要的科学研究主题及其发展趋势^[15]。

2) 共词分析方法。研究研究前沿的主题描述通常不止是一个或几个分散的主题词能够说明清楚的。如果使用一组相互关联的词汇来描述会更加清晰和明确。共词分析可以根据论文中不同词汇之间共同出现的机会进行聚类分析,生成一系列的词汇簇。根据这些词汇簇在论文发表的不同时期揭示期研究前沿主题发展变化情况。

共词分析的思想最初是在 20 世纪 70 年代由法国文献

计量学家提出的，1986 年，法国科研中心（Centre National de la Recherche Scientifique, CNRS）的 M. Callon, J. Law 和 A. Rip 出版了 “Mapping the Dynamics of Science and Technology”，提出了 “LEXIMAPPE”^[16-17]。由于在结果分析方面具有得天独厚的直接性，很快引起研究者的高度关注。

Kostoff 于 20 世纪 90 年代左右提出了数据库内容结构分析法（Database Tom ography, DT），通过对共词分析窗口位置、窗口大小、窗口内容 3 个方面的调整，实现了对共词分析方法的改进。并利用此方法实行了技术竞争情报分析和高技术领域前沿分析等应用^[18]。

3) 非相关知识发现方法。1986 年，美国芝加哥大学的 D. R. Swanson 教授首次提出 “基于非相关文献的知识发现法”^[19-20]。所谓基于非相关文献的知识发现法就是从表面上没有任何联系的文献内容之间识别出有效的、新颖的、潜在有用的以及最终可理解的知识的的情报研究方法。该方法可辅助科研人员发现潜在的关联，进而促进新知识的产生，有助于推动科学的发展^[21]。

非相关知识发现方法摒弃了传统的引文分析方法，利用自然语言处理技术对科技文献内容进行深入分析，从中发现相关知识点，进而发现潜在的知识关联。国内外学者对非相关知识发现理论进行了研究并拓展了相关理论与方法^[22-23]。

4) 基于概率主题模型方法。共词分析方法利用词语相互出现的位置这一关系进行科学研究主题的描述，而大多数主题的描述词语会高频率地出现在文献中，并且可能会均匀分布在文章中的不同位置。基于概率的主题模型方法运用概率统计方法对科技文献的主题进行科学研究前沿探测，并取得不错的效果。

1996 年，美国国防高级研究计划署（DARPA）希望开发出一种自动新闻主题发现系统。根据这一需求，科学家们提出了主题探测与追踪模型^[24]。该模型的主要目标是在时序标签的文本流中自动探测其中包含的主题^[29,31]。在科学研究前沿识别研究中，可以利用主题探测与追踪模

型根据科技文本的内容探测其隐含的科学研究主题的变化，进而识别科学研究前沿。

最开始的主题探测与追踪模型主要使用基于规则的方法进行探测^[25]，基于规则的探测方法不仅在规则编制过程中费时费力，而且如果文本内容改变规则也要相应改变，鲁棒性不好。基于概率的主题模型的出现改变了这一现状。2003 年，D. M. Blei 等提出了 LDA 模型，用于发现文本中的主题^[26]。2005 年 J. Zhang 等^[27]使用概率模型对在线文档进行了创新性主题探测与跟踪。D. M. Blei 等于 2006 年提出了动态主题模型。主要研究如何将 LDA 模型进行扩展，让动态 LDA 模型可以处理具有时间戳的文档数据集，实现动态主题探测与追踪^[28]。

由于基于概率的主题模型不仅使用多个主题词描述同一主题，而且给出了每个主题词对这一主题的贡献度。此外，还可以通过调整阈值的设置改变主题词的个数。因此基于概率的主题模型在科学研究前沿探测中受到广泛青睐。表 1 为科学研究前沿探测主要理论方法的总结。

2 科学研究前沿特征及影响因素分析

欧洲研究理事会（European Research Council, ERC）认为，“前沿研究”这个词更能体现对当前科学研究的最新理解，因为，当前基础研究和应用研究之间的传统区别已经不再明显，许多新兴科学和技术领域（例如，生物技术、信息与通信技术、材料和纳米技术以及认知科学等）经常兼具二者的本质特征。因此，“前沿研究”能更好地反映当前科学研究的特点。科学“前沿研究”应具有以下几个方面的特性：

- 1) 前沿研究立足于创造新知识和发展新认知的最前沿，相关研究往往会带来基础性的新发现，以及在理论和经验理解方面的进展，甚至偶尔可能取得一些革命性突破，这将彻底改变我们对整个世界的认识。
- 2) 前沿研究本质上是一种风险性的尝试。在新的最激动人心的研究领域，通常无法知道什么是最有成效的发展方法和研究途径，研究人员必须足够勇敢，并准备随时

表 1 科学研究前沿探测主要理论与方法

		理论基础	代表性成果
引文分析法	共被引分析	共被引分析方法，是指两篇以上论文共同被后来的一篇或多篇文献所引用，则称这两篇引文共被引	1973 年，Small 把科学研究前沿表述为同被引文献簇。提出了共引聚类方法来分析特定领域内基础性的、较早发表的核心论文
			1994 年，加菲尔德提出研究前沿是由共被引文献簇及其施引文献簇共同组成的文献簇
			1994 年，Persson 提出研究前沿就是高频共被引文献簇的施引文献簇
	文献耦合分析	两篇同时引用一篇论文的论文称之耦合论文，并把它们之间的这种关系称为文献耦合	1963 年，美国麻省理工学院教授 M. M. Kessler 提出两篇同时引用一篇论文的论文称之耦合论文（ Coupled Papers）， 并把它们之间的这种关系称为文献耦合
			1994 年，Persson 利用文献耦合分析基于同被引分析识别出的知识基础来确定研究前沿
			2003 年，Morris 等采用创新性的时间线方法来分析和展现研究前沿。可以在图形上展现研究前沿出现和消失的时间、潜在的新兴研究前沿及其基础文档

表 1 (续表)

	理论基础	代表性成果
主题词分析法	词频的波动与社会、情报现象之间存在内在联系,从而透过词频现象看内容本质的科学方法	齐普夫词频分布定律揭示语言统计规律 1997 年,加拿大蒙特利尔大学的 R. Dalpé 教授通过分析词频规律,完成了一份关于世界纳米科技研究状况的计量分析报告
	统计在同一篇文献中两个词汇共同出现的次数,以此为基础对这些词进行聚类分析,从而反映出这些词之间的亲疏关系,进而分析这些词所代表的学科和主题的结构变化	早期的共词聚类理论探讨集中在关键词间相关程度的测试,如包容指数 (Inclusion Index) 和临近指数 (Proximity Index) 等指标 Callon (1988) 提出用密度和向心度来测度聚类簇与簇之间、簇内部联系程度的强弱,分别计算出各簇 (研究主题) 密度和向心度,把所有主题展现在战略坐标图中 Kosstoff 博士将基于关键词或主题词的共词分析进一步扩大到基于全文本自动抽取词汇的共词分析——数据库内容结构分析技术 (Database Tomography, DT)
	爆发词检测	J. Kleinberg 在 2002 年提出了爆发监测的算法,其主要思想就是寻找文献集合中突然增加的词汇,用来揭示新出现的研究前沿主题
	概率主题模型	D. M. Blei 等 (2003) 提出了 LDA 模型,通过对文档中的词语分布情况,发现其中的研究主题 D. M. Blei 等 (2006) 提出了 DLDA 模型。该模型主要是对 LDA 模型进行了扩展,加入时间标签,使得 DLDA 模型可以处理具有时间戳记的文档数据集,实现动态研究主题的探测与追踪

去冒险,所以真正的研究前沿并不一定出现在资助机构确定资助的优先领域。

3) 前沿研究探讨的问题往往不考虑学科之间已有的界限,很有可能涉及多学科、交叉学科或者跨学科的研究,汇集不同学科背景、不同理论及思维方法、技术、方法论和仪器设备,甚至可能是不同目的和动机的研究人员从事的联合研究。

欧盟在未来和新兴技术项目 (Future and Emerging Technologies, FTE) (<https://ec.europa.eu/programmes/horizon2020/en/node/791>) 资助项目申请时,重点考虑该项目是否属于科学“研究前沿”的主要因素有:创新性、风险性、基础性、长期性等,如图 1 所示 (图片来源:<http://ec.europa.eu/programmes/horizon2020/en/h2020-section/future-and-emerging-technologies>)。

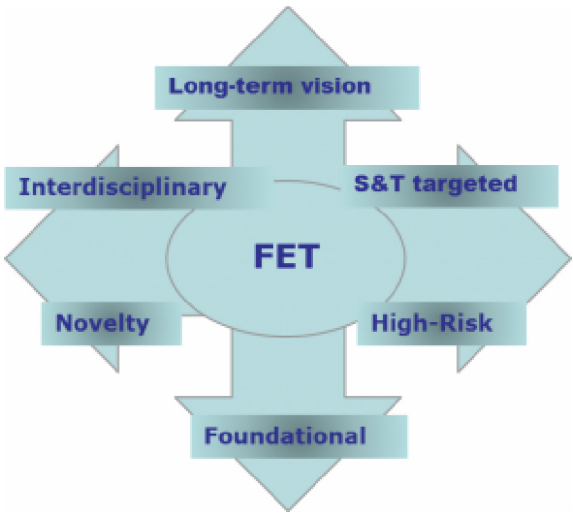


图 1 FTE 项目考虑的“研究前沿”识别因素

根据“科学研究前沿”上述几个方面的特征和欧盟给出的“科学研究前沿”项目考虑因素,本文总结归纳了以下可能影响科学研究前沿识别的因素,见表 2。

表 2 科学研究前沿影响因素

影响因素	关键属性	文献计量学指标	实现方法
研究前沿应该是创造新知识的研究领域	创新性	时效性 (Timeliness)	通过文献的参考文献或项目申请时间来计算
		与当前研究领域的相似性 (Similarity)	通过自然语言处理技术计算科技文献与领域研究方向之间的相似度
研究前沿领域本质上是具有风险性的,研究者应该了解并能有效控制该风险	研究者往往开展与当前学科独立的研究或者全新开辟一个研究领域	风险性 (Risk)	通过自然语言处理技术计算当前研究主题是否前期有研究基础,评估该研究与前期研究的原始创新能力
研究前沿不仅关注新的知识更应该注重潜在有用知识的产生,即考虑基础研究也考虑应用研究	该研究的可用性	可用性 (Applicability)	该研究的可用性
研究前沿应该考虑跨学科、跨领域性,组织不同学科背景的研究者,运用不同的研究方法达到各自不同的研究目的	学科的交叉性	学科的交叉性 (Interdisciplinarity)	不同学科的交叉性

上述指标有些可以通过数学模型计算进行量化,如时效性指标、风险性指标等,有些则无法通过具体的量化值进行量化如可用性指标等,这些指标可以进行定性研究。

3 当前面临的挑战与未来发展方向

关于科学研究前沿识别的研究国内外相关专家学者从

不同角度和思路进行了深入分析,提出了许多创造性的方法,各种方法具有自身的特点。但是,受数据源和分析原理的影响,会存在着一些固有的缺陷。总结起来主要有以下几个方面的问题:

1) 引文分析难以逾越时滞性问题。任何一篇论文如果达到一定引用频次,必定需要一定时间的积累来完成,因此,利用引文分析方法识别出的研究前沿在时间上必然显得滞后了。

此外,引文分析方法只是利用文献外在引用指标进行科学研究前沿探测,对识别出的研究前沿难以对其主题进行自动描述。

文献耦合分析方法的主要弊端是一篇文章的参考文献是固定的,不能像共被引分析那样动态揭示出学科发展状态。

2) 主题词分析缺乏语义信息支持。虽然基于主题词的研究前沿识别方法可以有效克服引文方法的时滞性,但是主题词分析方法面临的主要挑战是主题词的选取和如何保证主题词之间的语义信息。比如,“气相沉积法”这个词本身的含义是一种碳纳米管制备方法,利用“气相沉积法”可以制备高强度的碳纳米管。在科学研究中往往需要了解用什么实验方法得到什么实验效果。但是,目前的主题词分析方法还不能有效地解决这个问题。

3) 数据源无法交叉融合。以往的科学研究前沿探测研究在数据源选择上主要是利用已经发表的论文数据进行分析,而本文认为:科学“前沿领域”应包含三部分内容,一部分是引起世界科学家高度关注的对未解的科学问题所做的种种探索并取得了重大突破或一定进步,这部分内容多数包含在已经发表的学术论文或研究报告中;另一部分是面对未解决的问题,近期正在进行,但已明确研究目标和方法路径的科学前沿探索,该部分内容包含在各类科研项目申请文本中;还有一部分前瞻若干年内的科学研究前沿方向,这部分内容多数包含在科技规划文本中。

本文认为即使引文分析和主题词分析两种方法不断改进,但是,由于它们依赖的数据文本均是“过去时”记载科技研究成果的科技文献,因此其识别的科学研究前沿的“前瞻价值”也备受质疑。

结合以上分析,本文认为科学研究前沿探测研究未来主要会集中在以下几个方面:

1) 突破引文时滞性障碍,逐步转向基于文本内容挖掘的研究前沿识别研究。

基于文本内容挖掘的科学研究前沿探测方法可以在第一时间内对科技文本内容进行深入挖掘分析,识别出文献内包含的科学研究前沿信息,随着文本挖掘技术(比如条件随机场模型、语义标注技术等)的迅速发展,因此利

用文本挖掘技术进行科学研究前沿探测将会逐步成为未来主要研究方法。

2) 由单一数据源分析向混合数据源对比分析转变。正如前文提到的,目前科学研究前沿探测的分析数据源主要是利用已经发表的论文数据,而科技规划文本和正在研发的项目数据相比论文数据往往蕴含着更多的近期和远期的前瞻科学研究前沿思想信息。因此,综合利用科技规划文本、项目数据和论文数据将是未来科学研究前沿探测研究的主要分析数据源。

3) 由单一类型研究前沿识别向多类型研究前沿识别发展。根据影响科学研究前沿的主要因素分析,结合不同科学研究前沿主题类型,未来科学研究前沿探测会呈现多种类型的科学研究前沿类型,比如:未来潜在研究前沿、新兴研究前沿、热点研究前沿、高风险高回报研究前沿等。

4 结束语

本文总结归纳了科学研究前沿探测研究中主要运用的引文分析和主题词分析两种主要探测方法,及其不同分析方法的代表性研究成果;然后,分析了科学研究前沿探测主要特征和影响因素;随后,指出了目前存在的3个主要问题,即引文分析时滞性问题、缺乏语义信息支持问题以及分析数据源无法有效融合的问题;最后,论文分析了科学研究前沿探测未来发展方向。通过本文对科学研究前沿探测相关理论与方法的梳理,以期对未来科学研究前沿探测研究有所启迪。□

参考文献

- [1] 杨卫. 引领新常态的基础研究——基金委主任杨卫代表展望2020年的“三个并行” [EB/OL]. [2015-04-13]. <http://www.nsf.gov.cn/publish/portal0/tab38/info47990.htm>.
- [2] PRICE D D. Networks of scientific papers [J]. Science, 1965, 149 (3683): 510-515.
- [3] SMALL H, GRIFFITH B C. The structure of scientific literatures I: identifying and graphing specialties [J]. Science Studies, 1974, 4 (1): 17-40.
- [4] GARFIELD E. Research fronts [J]. Current Contents, 1994 (41): 3-7.
- [5] 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.
- [6] SMALL H. Tracking and predicting growth areas in science [J]. Scientometrics, 2006, 68 (3): 595-610.
- [7] SMALL H, UPHAM P. Citation structure of an emerging research area on the verge of application [J]. Scientometrics, 2009, 79 (2): 365-375.

- [8] 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.
- [9] 韩涛. 知识结构演化深度分析的方法及其实现 [D]. 北京: 中国科学院研究生院, 2008.
- [10] <http://portal.isiknowledge.com/portal.cgi?DestApp=ESI&Func=Frame>.
- [11] KESSIER M M. Bibliographic coupling between scientific papers [J]. American Documentation, 1963, 14: 10-25.
- [12] MORRIS S A, YEN G, WU Z, et al. TimeLine visualization of research fronts [J]. Journal of the American Society for Information Science and Technology, 2003, 55 (5): 413-422.
- [13] 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.
- [14] KLEINBERG J. Bursty and hierarchical structure in streams [J]. Data Mining and Knowledge Discovery, 2003, 7 (4): 373-397.
- [15] MANE K K, BORNER K. Mapping topics and topic bursts in PNAS [J]. Proceedings of the National Academy of Sciences of the United States of America, 2004, 101 (Suppl 1): 5287-5290.
- [16] QIN H. Knowledge discovery through co-word analysis [J]. Library Trends, 1999, 48 (1): 133-159.
- [17] CALLON M, LAW J, RIP A. Mapping the dynamics of science and technology: sociology of science in the real world [M]. [S. l.]: Macmillan, 1986.
- [18] KOSTOFF R N. Database tomography for technical intelligence [J]. Competitive Intelligence Review, 1993, 4 (1): 38-431.
- [19] SWANSON D R. Undiscovered public knowledge [J]. Library Quarterly, 1986, 56: 103-118.
- [20] SWANSON D R. Fish oil, Raynaud's Syndrome, and undiscovered public knowledge [J]. Perspectives in Biology and Medicine, 1986, 30: 7-18.
- [21] 安新颖, 冷伏海. 基于非相关文献的知识发现原理研究 [J]. 情报学报, 2006 (1): 87-93.
- [22] 李勇, 冷伏海, 王林. 基于非相关文献的三阶知识发现方法探讨 [J]. 中国图书馆学报, 2011 (4): 21-26, 69.
- [23] 钱庆, 洪娜, 李勇, 安新颖. 中文非相关文献知识发现系统 CmedLBKD 构建 [J]. 情报理论与实践, 2012 (4): 109-113.
- [24] 张云秋, 冷伏海. 非相关文献知识发现的理论基础研究 [J]. 中国图书馆学报, 2009 (4): 25-30.
- [25] 黄水清, 马俊岭. 汉语社会科学文献非相关文献知识发现的实证研究——以农业经济学文献为例 [J]. 中国图书馆学报, 2009 (4): 31-38.
- [26] 张云秋, 冷伏海. 基于非相关文献知识发现中的文本挖掘研究 [J]. 情报理论与实践, 2007 (2): 194-197.
- [27] 曹志杰, 冷伏海. 非相关文献知识发现方法在航天科技情报研究中的应用分析 [J]. 情报理论与实践, 2008 (4): 569-572.
- [28] ZHANG J, SONG Y, ZHANG C, LIU S. Evolutionary hierarchical dirichlet processes for multiple correlated time-varying corpora [C] // Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '10, New York, NY, USA, 2010: 1079-1088.
- [29] ALLAN J. Topic detection and tracking: event-based information organization [M]. Kluwer Academic Publishers, Norwell, MA, USA, 2002.
- [30] BLEI D M, LAFFERTY J D. Dynamic topic models [C] // Proceedings of the 23rd international conference on Machine learning ICML '06, New York, NY, USA. ACM, 2006: 113-120.
- [31] ZHANG J, GHAMRANI Z, YANG Y. A probabilistic model for online document clustering with application to novelty detection [C] // SAUL L K. NIPS, 2004: 1617-1624.
- [32] WANG C, BLEI D M, HECKERMAN D. Continuous time dynamic topic models [C] // Proceedings of the 24th Conference on Uncertainty in Artificial Intelligence, UAI '08, 2008: 579-586.
- 作者简介:** 白如江, 男, 1979 年生, 副研究馆员。研究方向: 文本数据挖掘和科技情报。
- 冷伏海, 男, 1963 年生, 研究员。研究方向:** 竞争情报和高科技信息分析。
- 廖君华, 女, 1979 年生, 讲师。研究方向:** 文本数据挖掘和信息管理系统。
- 作者贡献声明:** 白如江: 论文思路设计、资料收集、论文撰写、纠错和修订。
- 冷伏海:** 研究思路指导。
- 廖君华:** 部分资料搜集、整理。
- 录用日期:** 2016-10-14