

# 基于文献的研究前沿特征与识别方法研究

中国医学科学院医学信息研究所 北京 100020

范少萍 安新颖 倪萍

**摘要** 针对现有研究前沿的研究中缺乏对前沿特征的系统描述与计算方法说明,在全面了解国内外有关研究前沿特征及其计算方法的研究内容与结论基础上,总结得出较为全面规范的研究前沿应具有的特征与特征计算方法,以期为今后开展研究前沿发现与识别研究提供参考。本文借鉴系统综述的思想与流程,选择 Web of Science 数据库、中国知网(CNKI)数据库以及万方数据库作为主要的文献数据来源,制定精确有效的检索式获取文献,依据文献遴选标准选择文献并进行精读。通过回顾已有研究,总结得出研究前沿主要具有时间上的新颖性、内容上的创新性、学科交叉性、高关注度 4 个特征,并给出不同数据源情况下,每种特征可选择的计算方法。

**关键词:** 研究前沿,系统综述,特征,计算方法

**中图分类号:** G350

## The Study on Characteristics and Identification Methods of Research Front Based on Literature

Chinese Academy of Medical Sciences-Institute of Medical Information & Library, Beijing 100020, China  
FAN ShaoPing AN XinYing NI Ping

**Abstract** Because the lack of systematic description of characteristics of research front and calculation methods in existing research fields, this paper aimed to obtain a set of comprehensive and standardized characteristics and calculation methods for research front by a comprehensive understanding of characteristics and identification methods of research front based on domestic and foreign research papers.

**基金项目:** 本文受国家自然科学基金项目:基于语义的医学领域前沿知识发现及演化机制研究(71303259)、中央级公益性科研院所基本科研业务费:基于统计和语义的医学文献主题新颖性探测方法研究(2016RC330004)的资助。

**作者简介:** 范少萍(1986-),助理研究员,博士,研究方向:情报分析方法与技术、科学计量与科技评价, email: fan.shaoping@imicams.ac.cn; 安新颖(1978-),通讯作者,副研究员,博士,研究方向:学科热点分析与机构评价; email: an.xinying@imicams.ac.cn; 倪萍(1991-),硕士研究生,研究方向:医学信息分析与科研评价, email: ni.ping@imicams.ac.cn。

We hope to provide a reference to the future relevant research through our work. This paper used systematic review to draw conclusion, this research selected Web of Science, China National Knowledge Infrastructure (CNKI), WANFANG DATA as the primary data sources, developed accurate and efficient retrieval formula to obtain literatures, and selected right paper to intensive reading according to the literature selection criteria. By reviewing existing researches, this study summarized the main characteristic of research front: they are novelty in time, innovation in content, interdisciplinary and high degree of concern. Moreover, this study also presented the calculation methods for each characteristic based on different data sources.

**Keywords:** Research front, systematic review, characteristics, identification methods

为了加快实施国家的创新驱动发展战略，中共中央国务院于2016年5月印发了《国家创新驱动发展战略纲要》<sup>[1]</sup>。在《纲要》的战略任务中，指出要“加强面向国家战略需求的基础前沿和高技术研究”、“面向科学前沿加强原始创新……”等。上述战略任务的提出，在国家战略高度指明了科学前沿研究的重要性与紧迫性，为我国科学研究指明方向的同时也提出了挑战。开展科学前沿研究首先需要明确科学前沿是什么，仅少数高层次科学家能够厘清领域知识结构，指出前沿方向，对于大部分研究人员而言，准确发现并定位领域研究前沿较为困难。

除领域专家的定性分析得出领域前沿外，借助科学计量学方法可以从定量角度辅助发现领域前沿，因此图书情报领域开展了大量研究前沿发现与识别的相关研究。由于研究前沿的前瞻性、理论性等特点，学术界对研究前沿的定义、特征分析、定量识别方法等方面还未达成统一，特别是在定量识别方法上，常常与新兴类、热点类研究识别存在交叉之处。部分国内学者，或将这三者不进行明显区分进行研究，或将前沿分为新兴前沿、热点前沿等，致使研究前沿的特征变得模糊不清。上述情况使得研

究人员无法全面了解研究前沿的特征，并掌握识别方法，导致学术界缺乏一套权威有效的研究前沿识别指标体系。

系统综述（systematic review，SR）是一种被公认的客观评价和合成针对某一特定问题的研究证据的最佳手段，在医学研究领域得到了广泛应用<sup>[2]</sup>。本文拟借鉴系统综述的思想与流程，对国内外有关研究前沿特征与识别方法的研究文献进行综述，以期合理确定研究前沿的特征与识别方法，为后续相关研究提供参考与借鉴。

## 1 系统综述法概述

### 1.1 定义

系统综述的目的在于解决独立研究之间的矛盾结果，修正前人的结果，使结论趋于一致性，并可以解决以往研究未考虑的新问题。Cook, Mulrow 和 Haynes<sup>[3]</sup>将系统综述定义为：集合信息，对该信息进行批判性地评价，从多角度综合初始研究结果的一种二次文献研究方法。并指出一篇优秀的系统综述应该具备的三个特征，分别是：对待研究问题进行精确且显性的定义和陈述；可复制的检索策略（为此，系统综述文章中应包含所检索的数据库、检索词、对初

始研究年份、语言等的限制), 强调研究的可重复性; 使用预先设定的初始研究纳入和剔除准则。因此, 采用科学的方法产生的系统综述能够提供更加真实可靠的信息和知识, 而且高质量的系统综述能够对本学科产生深刻的影响。

系统综述包括两种类型: 定性系统综述和定量系统综述。当原始文献的研究结果被总结但未经过统计学合并时, 这种系统综述称作定性系统综述。定量系统综述, 又称 meta 分析, 即荟萃分析, 是应用统计学方法对若干个研究的结果进行定量统计合并的过程<sup>[4]</sup>。本文对有关研究前沿特征与识别方法的文献进行总结, 不进行统计学合并, 因此是定性系统综述。

## 1.2 系统综述的步骤

好的系统综述应该具备如下特征: 清楚地表明题目和目的、采用综合检索策略、明确研究入选和排除标准、列出所有入选的研究、清楚地表达每个入选研究的特点并对其方法学质量进行分析、阐明所有排除研究的原因等<sup>[2]</sup>, 主要步骤如下<sup>[5]</sup>。

### (1) 提出问题

不同于叙述性综述的资料导向性, 系统综述是问题导向型的, 任何系统综述都是从一个明确表达的问题开始的, 这是做好系统综述工作的第一步。系统综述的题目主要来源于不肯定、有重要争论的问题。

### (2) 检索文献

系统全面无偏差地收集与研究问题相关的文献信息是系统综述与叙述性综述的主要区别之一。由问题得出一系列的检索用词, 检索最相关的数据库, 并根据需要分别对发表年限、语种等进行限制。除了数据库的自动检索外, 还需要对特定

期刊进行手动检索, 最大限度地确保不遗漏文献。

### (3) 选择文献

选择文献是根据研究目的, 确定文献纳入和剔除的标准, 从收集到的相关文献中筛选符合要求的文献的过程。通过预先定义的文献选择准则, 可以很好的降低各种偏见对最后结果的影响, 尽量减少不必要的偏差。

### (4) 抽取数据

从符合纳入要求的文献中抽取用于系统综述的数据信息, 一般包括基本信息、研究特征和结果测量的内容, 所抽取的信息必须是可靠的、无偏倚的。

### (5) 结果分析与展示

对收集到的文献信息, 采用定性或定量的方法进行分析, 以得到相应的结果。

除上述五个必要步骤之外, 根据之前的分析, 系统综述还需要根据最新收集到的初始研究, 按前述步骤重新进行分析、评价以及时更新和补充新的信息, 对综述结果进行定期修正和更新, 进一步完善研究问题。此外, 系统综述还需要讨论该研究可能存在的各种偏倚, 指导以后的研究者避免相同的问题。

## 2 分析过程

### 2.1 研究前沿概念

目前国内外对“研究前沿”尚未进行统一界定, 现有研究对研究前沿的界定大致可以分为四类:

(1) 根据共被引强度来定义研究前沿。共被引强度是指用于描述两篇文献之间的关系, 用共被引的次数来表示, 当两篇文献共被引次

数越多则共引强度越高。普莱斯<sup>[6]</sup>对研究前沿的界定主要依据的就是共引强度，其将科学引文网络中被近期发表的文献引用的稍早些发表的文献集称为研究前沿，这些高频率共被引文献构成了研究前沿。

(2) 根据耦合强度定义研究前沿。所谓耦合强度是用相同引文的数量描述两篇文献之间关系的疏密，当两篇文献引用的相同文献越多说明文献的耦合强度越高。皮尔森<sup>[7]</sup>将研究前沿定义为引用高同被引文献的文献群，而将被引的文献称作研究前沿的知识基础。

(3) 基于词分析的研究前沿。2006年，陈超美<sup>[8]</sup>博士在回顾其他学者分析研究前沿的做法之后提出，使用从题目和摘要中提取的专业术语以及采用 Kleinberg 突发检测算法获得的频率激增的术语，能够识别新兴的研究前沿。

(4) 基于多维特征的研究前沿。许晓阳等<sup>[9]</sup>认为，在科学研究中最近出现、正在兴起的研究主题或研究领域就是研究前沿，并指出，从本质上讲，研究前沿必须是一种研究主题或研究领域，而且这种研究主题或领域往往来源于新的科学发现或研究进展，并在短时间内迅速引起领域内科学家的高度关注，其代表了科学发展的难点、热点与发展趋势。陈仕吉<sup>[10]</sup>认为研究前沿是科学研究中最先进、最新、最有发展潜力的研究主题或研究领域。

上述定义可以看出，研究前沿定义的落脚点主要是文献群、关键词/术语、研究主题或领域。其中，前沿文献群方便识别前沿文献的多种外部特征，但容易忽略对文献内容层面的分析；前沿关键词/术语反映的是文献研究主题核心和精髓，是对研究主题的高度概括和凝

练，但不能揭示词与词之间的关系，即使利用词共现关系也难以解释概念之间的逻辑；前沿主题阐述了前沿关键词之间的明确逻辑关系，得到的是研究方向和主要内容。

## 2.2 确定研究问题

本文的研究目的是通过综述国内外有关研究前沿的特征与计算方法的研究，总结得出研究前沿应具有的特征以及每项特征的计算方法，为今后开展研究前沿发现与识别研究提供较为全面的参考。

由于学术界目前还未对研究前沿形成统一的概念，因此，不同学者对研究前沿的内涵与外延有不同的理解与延伸，导致在研究前沿识别研究时无法准确描述前沿的特征。单一使用文献计量学的引文分析法、共词分析法或者社会网络分析法以及文本挖掘分析方法等，仅能揭示前沿某一方面的特征，不足以较为全面的展示研究前沿，并精准识别研究前沿<sup>[11]</sup>。为能更加全面广泛地确定与识别研究前沿特征，本文主要回顾从多方面、多角度、多维度对研究前沿及其相关概念(研究热点、热点主题、新兴主题、新兴趋势等)进行识别研究的文献的主要内容，总结文献中提出的前沿的特征与计算方法，明确方法的分析领域、适用范围、存在的问题等。

## 2.3 检索文献

为了解国内外有关研究前沿特征与识别方法的研究现状，国外数据库使用 Web of Science 核心合集数据库，国内数据库使用知网(CNKI)与万方2个数据库，以求更全面覆盖国内相关研究。

根据本文的研究问题，首先在3个数据库

检索“研究前沿”及相关主题的研究文献，发现很多文章与学位论文摘要中极易出现此类词语，用以说明研究的重要性与新颖性（如“……研究是领域的研究热点……”），但与本文的研究问题无关，增加了过多噪音。因此，在检索

词中增加“识别”、“探测”等主题词以及学科限制等，既精炼检索范围，突出研究问题，又在查全率的基础上，保证查准率。对各个数据库的检索时间、使用的检索式、检索结果等总结如表1所示。

表1 三大检索数据库的检索式与检索结果表

数据库名称	检索时间	检索式	检索结果
Web of Science	2016年6月1日	TS=(("research front*" NEAR (detect* or identif*)) OR ("hot topic*" or "hot research*" NEAR (detect* or identif*)) OR ("emerging topic*" or "emerging research*" NEAR (detect* or identif*))) AND 语种:(English) 时间跨度:所有年份。索引:SCI-EXPANDED, CPCI-S。精炼依据:Web of Science 类别:(MATHEMATICS APPLIED OR COMPUTER SCIENCE INFORMATION SYSTEMS OR MATHEMATICAL COMPUTATIONAL BIOLOGY OR COMPUTER SCIENCE ARTIFICIAL INTELLIGENCE OR COMPUTER SCIENCE THEORY METHODS OR COMPUTER SCIENCE INTERDISCIPLINARY APPLICATIONS OR INFORMATION SCIENCE LIBRARY SCIENCE OR COMPUTER SCIENCE SOFTWARE ENGINEERING OR MATHEMATICS OR COMPUTER SCIENCE HARDWARE ARCHITECTURE OR COMPUTER SCIENCE CYBERNETICS OR MATHEMATICS INTERDISCIPLINARY APPLICATIONS OR SOCIAL SCIENCES MATHEMATICAL METHODS OR PHYSICS MATHEMATICAL)	264条。
CNKI	2016年5月30日	SU=(“研究前沿”+“前沿主题”+“前沿领域”+“科学前沿”+“前沿研究”+“新兴主题”+“新兴趋势”+“热点主题”+“热点话题”+“研究热点”)AND SU=(“识别”+“探测”)	15332条，计算机软件及计算机应用学科文献数量最多，共5182条，图书情报与数字图书馆学科下共118条。
万方	2016年5月30日	主题=(“研究前沿”+“前沿主题”+“前沿领域”+“科学前沿”+“前沿研究”+“新兴主题”+“新兴趋势”+“热点主题”+“热点话题”+“研究热点”)*主题=(“识别”+“探测”)	12358条，其中期刊论文2180条。

由于Web of Science数据库初步检索出文献较多，为进一步聚焦研究问题，找出更加合理的研究文献，限定Web of Science类别主要为计算机科学、信息科学和图书情报学，最终得到264篇文献。上述文献中研究主题多为交通领域、模式识别、图像识别、微博热点话题与新话题识别（多结合用户）、复杂网络的社区识别、指纹识别、作者识别等，真正聚焦多维度科学文献前沿热点研究的文献较少。

本文主要关注主题类前沿的发现与识别研究，对前沿技术、新兴技术等领域暂未涉及，因此，在CNKI与万方的中文检索式未将有关技术类的关键词列出。虽然2个主要中文数据

库检索出文献数量较多，但存在较大范围的数据重复，且涉及学科领域较多。人工精选最相关研究文献，并二次检索获得其他相关文献。

除上述3大主要检索数据库外，本文还收集了其他网络资源中，如Google Scholar等，有关研究前沿识别方法的研究文献，以期全面的了解该主题的研究现状。

## 2.4 选择文献与结果解读

### (1) 选择标准

根据本文的研究问题与研究目的，对检索结果进行人工筛选。首先通过阅读文献的标题、摘要与关键词等主要字段遴选描述研究前沿多维特

征, 并给出主要计算方法与思路的文献。对于仅采用一种方法, 如图书情报领域常用的引文分析法、共词分析法等进行研究前沿识别, 或对该方法进行改进与完善的研究不纳入精读范围内。同时, 虽然对新闻媒体、微博领域热点主题挖掘的研究较多, 但此类分析对象与科技领域的主题特

点不同, 不能完全将其分析结果纳入精读范围, 还需结合文献的内容进一步确定。

根据上述选择原则与标准, 由 2 位研究人员进行“背靠背”阅读文献。当筛选结果不一致时, 邀请第 3 位研究人员讨论并达成一致意见。具体流程如图 1 所示。

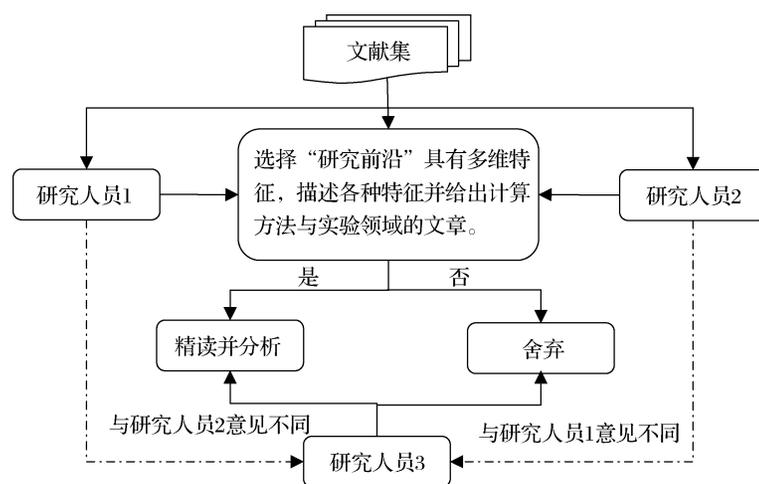


图1 文献筛选流程图

## 2.5 结果解读

本文重点阅读的文献中, 最早从多个维度进行趋势探测的文章是发表在 2005 年的 *Detecting emerging trends from scientific corpora*<sup>[12]</sup> 一文, 该文章从受关注度 (interest) 与有用性 (utility) 两个维度识别新兴趋势, 研究认为受关注程度和有用性的衡量值均大于 0 为新兴研究趋势; 受关注程度大于 0 而有用性小于等于 0 则为潜在的新兴研究趋势; 受关注程度小于等于 0 而有用性大于 0 则为不显著但对研究有用的趋势; 受关注程度和有用性的衡量值均小于等于 0 为陈旧的趋势。2007 年, Chen 等<sup>[13]</sup> 发表的 *Hot topic extraction based on timeline analysis and multidimensional sentence modeling* 一文采用《华盛顿邮报》、路透社报道、CNN 报道等新闻报道,

从普遍性 (Pervasiveness) 与时事性 (Topicality) 两方面揭示热点话题。Tu 等<sup>[14]</sup> 主要利用年份特征, 识别期刊与会议文献中的新兴主题。根据 (计算年 - 研究主题出现年 + 1) 计算主题新颖性, 根据 (计算年累积发文量 / 统计时间段内总发文量) 计算发表数量指数。文章使用年份代表文献的产生时间, 但文献通常按月出版, 使得计算结果过于粗糙, 精度较差。与 Tu 等利用时间信息类似, Jarić 等<sup>[15]</sup> 使用参考文献平均年龄以及 2 年前发表的参考文献的比例<sup>[16]</sup> 两个指标, 识别新兴研究领域。利用参考文献年龄的优势在于不存在时间滞后性, 可以及时了解现有研究。

*Development and verification of a bibliometric model for the identification of frontier research*<sup>[17]</sup>

一文利用文献计量学理论与方法,采用国家项目数据,进行研究前沿发现研究。该研究认为,研究前沿应具有时效性(Timeliness)、创新性(Innovativeness)、风险性(Risk)、应用性(Applicability)与学科交叉性(Interdisciplinary),且模型计算结果表明,创新性与学科交叉性都是最显著有效的,而其他3项指标,对项目的通过概率无显著影响。与此类似,Zhang等<sup>[18]</sup>参考现有文献计量领域在探测研究前沿时所采用的指标,利用突破性、继承性、时效性和跨学科性四个指标构建综合前沿性指数。通过实验发现,突破性指数对前沿影响最大,其次为跨学科性和时效性,继承性对前沿的影响最小。

国内在多维度识别研究前沿或新兴研究领域方面,起步交晚,已具有一定的研究成果。如2009年,殷蜀梅<sup>[19]</sup>就提出,应该从政策和投入、实践应用、科研活动活跃度、网络资源兴起和利用程度等维度判断新兴研究趋势,但随着科技迅猛发展,以上维度显然过于粗糙且难以定量度量。刘红霞等<sup>[20]</sup>基于汉语科技词系统,从兴趣度和有用性两个维度,给出了新兴研究趋势的监测方法,但没有进行实验验证,方法的效果无法考证。郭涵宁<sup>[21,22]</sup>认为,新兴研究领域可以使用新作者数量、突现主题词共现、突现重要文献数量、引文年龄、学科跨度指数、基金资助项目数量与额度等七个维度进行识别与探索,并在量子计算、RNA干扰、语义网等多个领域进行实证发现,构建的多元科学指标可以从不同角度对识别新兴研究领域及其成熟程度进行分析与判断。盛立<sup>[23]</sup>聚焦于生物医学领域,从前瞻性、时效性、新颖性和灰色性四个维度识别研究前沿,对神经性毒剂医学防护研究领域等生物医学领域进行实证,验

证了方法的有效性。范云满等<sup>[24]</sup>借鉴文献16的指标,同时增加了被引量指标,将主题映射为文本进行累积计算。但实验过程中作者发现,利用所提方法进行新兴主题预测时,可能有较大的失真。蒋玉桃<sup>[25]</sup>从主题频度、主题被引频次、主题影响因子三个维度对热点主题进行探测研究,并在数学领域进行了验证。郑彦宁等<sup>[26]</sup>基于关键词共现,从新颖性和研究主题关注者数量两个维度识别研究前沿,在研究结论中,作者也指出,仅使用上述两个指标不足以全面描述研究前沿,还需要在此基础上进一步补充性的指标并进行验证。范少萍等<sup>[11]</sup>虽然经过分析较为全面的总结了前沿知识应具有生命周期特征、普遍认可性、权威性、创新性和学科交叉性等特征,但由于指标计算问题,尚未能通过实验验证所提方法的有效性。

国内从多维度进行研究前沿或新兴主题探测研究较多的是北京工业大学黄鲁成教授团队。在研究前沿探测研究中,该团队主要使用创新性与学科交叉性两个主要指标进行探测模型构建,在智能材料领域验证了所提探测模型的实用性和合理性<sup>[27]</sup>;在新兴主题探测研究中,使用关注度、成长潜力和关联度组合<sup>[28]</sup>或关注度与实用性组合<sup>[29]</sup>在不同领域进行识别研究。其中,两个关注度采用了不同的计算方法,第一种组合中的关注度主要使用关键词在所有文章中的TFIDF值之和进行表示,第二种组合中的关注度使用共词分析的方法对渐增型突现文献的施引文献的高频关键词进行共现聚类,找到新兴主题。

在实验领域的选择方面,上述主要研究涉及多个研究领域,如生物医学领域的神经性毒剂医学防护研究领域、合成生物学研究、军队

健康管理相关领域、生物材料等；物理领域的LED、精密单点定位技术、智能材料、传感器等；计算机领域的机器学习、复杂网络、语义网等；生态学领域的水体富营养化研究、鲟鱼研究、渔业研究和一般生态研究等；数学学科以及其他自然学科领域。虽然尚无法证明各种不同指标方法对所有领域都适用，但可以看出，对研究前沿特征与计算方法的研究已延伸到多个学科领域。

国内外多位学者从多种不同维度对研究前沿、新兴主题或热点主题等进行了探测研究，但较多研究也指出了方法存在的不足之处，如时间精度不够、指标不够全面、缺乏实证研究等。可见，研究人员需针对所选领域特征与需求，设计科学合理的研究前沿特征及计算方法，进而通过实验验证方法的有效性与可行性。

### 3 前沿知识特征与计算方法总结

总结 2.5 中各特征的本质与主要计算方法，得出研究前沿的主要特征有：时间上的新颖性、内容上的创新性、学科交叉性、高关注度。以上 4 个特征从定量角度较为全面的总结了研究前沿的特点，便于后续开展相关定量分析研究。

(1) 时间上的新颖性。即研究前沿需是最新出现的当前时间段内的前沿。该特征从时间维度描述研究前沿，认为越是新出现的主题越容易包含最新的研究内容，也更有可能是研究前沿。研究前沿的这一特征与“新兴类主题”相同，但潜在的含义是最新的研究前沿内容也越新，而新兴主题侧重时间新，并不一定要考察主题的新颖性。此外，时效性特征可以将领域的“研究前沿”与“知识基础”区分开来。

(2) 内容上的创新性。即研究前沿所涉及的研究内容应处于创造新知识和开发新认知的最前端，对引领该领域发展具有重要作用。该特征从内容层面描述研究前沿，从根本上将其与“新兴类主题”、“热点类主题”进行区分，突出强调研究前沿内容上的突破性与引领性。此外，该特征还可以将领域的“研究前沿”与“知识基础”区分开来。

(3) 学科交叉性。即组成研究前沿的知识基础来源于多个不同学科。多学科的交叉促使科学研究未来发展的不确定性增强，新成果跨领域应用的机会增加，出现创新性影响的概率加大。因此，研究前沿不仅要考虑创新性，还要测度知识的学科交叉性。学科交叉不仅推动了科研世界复杂问题的解决，而且促进了不同领域研究者们的交流，激发新知识、新思维和新技术的产生口，进而形成新的科学问题，促使研究前沿的改变。

(4) 高关注度。即研究前沿受到科学界的高度关注和普遍认可。研究前沿的普遍认可性将其与“新兴类主题”进行区分，新兴类主题，主要侧重主题在出现时间上是最新的，代表的是一个年轻的甚至是处于“萌芽”状态的领域，因此科学界对该领域的关注与认可程度存在不确定性，但研究前沿所代表的必须是受到广泛关注并且普遍认可的主题。

本文认为，研究前沿是具备上述所有特征的一类知识，特别是创新性，是研究前沿应具有的重要特征。研究前沿正是因为内容上与已有知识存在显著差异，才更有可能成为领域发展的新方向与新突破，从而产生新知识。

由于作者思路、分析数据源、数据结构等不同，导致每种特征有多种计算方法，本文总

结现有研究涉及的研究前沿的特征及主要计算方法如表 2 所示。在今后的研究前沿识别中，

可根据实际需求与研究目的，选择合理的计算方法。

表2 研究前沿特征与计算方法汇总表

序号	特征	计算方法	数据来源
1	新颖性	1/(计算年 - 研究主题出现年 + 1)。	文献出版时间
		引文年龄的平均值(引文年龄 = 项目提交年 - 参考文献出版年)。	文献出版时间与参考文献出版时间
		主题中包含的所有论文的平均发表时间。	文献出版时间
		论文发表年份 - 参考文献发表年份。 研究主题年龄。	文献出版时间与参考文献出版时间 文献出版时间
2	创新性	基于文本内容的时序聚类分析。	文献关键词
		该主题与前一个时间窗所有主题的主题相似度(欧式距离)的平均值。	研究主题
		突现主题词共现网络。	文献关键词
		关键词突发检测。 高质量学术文献。	文献关键词 文献
3	学科交叉性	参考文献跨学科性的变化。	参考文献所属学科
		其他学科关键词数量与比例。	学科关键词
		计算参考文献的赫芬达尔 — 赫希曼指数。 Rao-Stirling 指数。	参考文献所属学科 参考文献所属学科
		文献的学科交叉广度。	文献所属学科
4	关注度	某年主题被提及频次; 某年主题总被引频次; 某年主题内作者声誉(作者发表文献数量); 某年主题来源期刊或会议权重。 在指定年份,第一次在该领域发表论文的作者数量; 计算时间内,某一研究领域获得基金资助的数量与金额。 TF-IDF 算法,关键词在所有文章中的 TFIDF 值之和。 研究主题关注作者数量的比例变化。	文献元数据 文献元数据 文献关键词 作者

## 4 结语

为全面了解国内外有关研究前沿特征及其计算方法的研究现状,从而总结得出研究前沿的特征与特征计算方法,本文借鉴系统综述的思想与流程,选择 Web of Science 数据库、中国知网(CNKI)数据库以及万方数据库作为主要的文献数据来源,制定精确有效的检索式获取文献。依据文献遴选标准选择文献并进行精读。通过回顾已有研究,总结得出研究前沿主要具有时间上的新颖性、内容上的创新性、学科交叉性、高关注度 4 个特征,并给出不同数据来源情况下,每种特征可选择的计算方法。本

文的研究结论为今后开展研究前沿发现与识别研究提供了较为全面的参考。

## 参考文献

- [1] 中共中央国务院. 印发《国家创新驱动发展战略纲要》[EB/OL]. [2016-07-04]. [http://news.xinhuanet.com/politics/2016-05/19/c\\_1118898033.htm](http://news.xinhuanet.com/politics/2016-05/19/c_1118898033.htm)
- [2] 詹思延. 如何做一个好的系统综述和 Meta 分析[J]. 北京大学学报医学版, 2010, 42(6): 644-647.
- [3] Cook D J, Mulrow C D, Haynes R B. Systematic Reviews: Synthesis of Best Evidence for Clinical Decisions.[J]. Annals of Internal Medicine, 1997, 126(5): 376-380.
- [4] Glass G V. Primary, Secondary, and Meta-Analysis

of Research1[J]. Educational Researcher, 1976, 5(10): 3-8.

[5] 邱璇. 系统综述——一种更科学和客观的综述方法[J]. 图书情报知识, 2010(1): 15-19.

[6] Price D J. Networks of Scientific Papers[J]. Science, 1965, 149(3683): 510-515.

[7] Persson O. The Intellectual Base and Research Fronts of JASIS 1986–1990[J]. Journal of the Association for Information Science and Technology, 1994, 45(1): 31-38.

[8] Chen C. CiteSpace II: Detecting and Visualizing Emerging Trends and Transient Patterns in Scientific Literature[J]. Journal of the Association for Information Science and Technology, 2006, 57(3): 359-377.

[9] 许晓阳, 郑彦宁, 赵筱媛, 等. 研究前沿识别方法的研究进展[J]. 情报理论与实践, 2014, 37(6): 139-144.

[10] 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, 25(9): 28-33.

[11] 范少萍, 安新颖, 赵迎光. 基于多维特征识别的医学领域前沿知识发现研究框架[J]. 中华医学图书情报杂志, 2016, 25(3): 1-7.

[12] Le M H, Ho T B, Nakamori Y. Detecting Emerging Trends from Scientific Corpora[J]. 2005.

[13] Chen K Y, Luesukprasert L, Chou S T. Hot Topic Extraction Based on Timeline Analysis and Multidimensional Sentence Modeling[J]. IEEE Transactions on Knowledge & Data Engineering, 2007, 19(8): 1016-1025.

[14] Tu Y N, Seng J L. Indices of Novelty for Emerging Topic Detection[J]. Information Processing & Management, 2012, 48(2): 303-325.

[15] Jarić I, Knežević-Jarić J, Lenhardt M, et al. Relative Age of References as A Tool to Identify Emerging research Fields with an Application to the Field of Ecology and Environmental Sciences[J]. Scientometrics, 2014, 100(2): 519-529.

[16] Glänzel W, Schoepflin U. A Bibliometric Ageing Study Based on Serial and Non-serial Reference Literature in the Sciences[J]. Scientific Literature,

1995(1995).

[17] Development and Verification of A Bibliometric Model for the Identification of Frontier Research-DBF. [EB/OL]. [2015-12-04]. <http://www.ait.ac.at/departments/innovation-systems/development-and-verification-of-a-bibliometric-model-for-the-identification-of-frontier-research-dbf/>.

[18] Lihua, ZHANG, Zhiqiang, et al. Detection of Early-stage Research Fronts—An Example of Complex Networks Research[J]. 中国文献情报: 英文版, 2014, 7(4): 77-94.

[19] 殷蜀梅. 判断新兴研究趋势的技术框架研究[J]. 图书情报知识, 2008, 2008(3): 76-80.

[20] 刘红霞, 张运良, 乔晓东, 等. 基于汉语科技词系统的新兴科研趋势监测研究[J]. 数字图书馆论坛, 2010(5).

[21] Guo H, Weingart S, Rner K. Mixed-indicators Model for Identifying Emerging Research Areas[J]. Scientometrics, 2011, 89(1): 421-435.

[22] 郭涵宁. 多元科学指标视角下的新兴研究领域识别探索[D]. 大连: 大连理工大学, 2013.

[23] 盛立. 生物医学领域研究前沿识别与趋势预测[D]. 北京: 中国人民解放军军事医学科学院, 2013.

[24] 范云满, 马建霞. 基于 LDA 与新兴主题特征分析的新兴主题探测研究[J]. 情报学报, 2014, 33(7): 698-711.

[25] 蒋玉桃. 热点主题的探测系统研究及技术实现[D]. 成都: 电子科技大学, 2015.

[26] 郑彦宁, 许晓阳, 刘志辉. 基于关键词共现的研究前沿识别方法研究[J]. 图书情报工作, 2016, 60(4): 85-92.

[27] 吴菲菲, 杨梓, 黄鲁成. 基于创新性和学科交叉性的研究前沿探测模型——以智能材料领域研究前沿探测为例[J]. 科学学研究, 2015, 33(1): 11-20.

[28] 黄鲁成, 唐月强, 吴菲菲, 等. 基于文献多属性程度的新兴主题识别方法研究[J]. 科学学与科学技术管理, 2015(2): 34-43.

[29] 黄鲁成, 王静静, 李欣, 等. 基于文献计量的新兴趋势分析——以生物材料为例[J]. 情报杂志, 2015(7): 58-64.