

# 学术成果主题新颖性与学术引用的相关关系研究

逯万辉<sup>1,2</sup> 苏金燕<sup>1</sup> 余倩<sup>1</sup>

(<sup>1</sup>中国社会科学院中国社会科学评价研究院 北京 100732)

(<sup>2</sup>中国科学院文献情报中心 北京 100190)

**摘要** 文章通过对学术成果主题新颖性测度与其被引情况的统计分析,研究和对比了不同层次主题新颖性的学术成果在刊出之后的学术传播与影响关系。结果表明学术成果主题新颖性与学术被引之间存在显著的相关关系,不同层次主题新颖性成果在刊出后其被引生命周期存在较大差异。

**关键字** 主题新颖性 学术引用 生命周期

## Research on Correlation of Academic Achievement Theme Novelty and Academic Citation

Lu Wanhui<sup>1,2</sup> Su Jinyan<sup>1</sup> Yu Qian<sup>1</sup>

(Chinese Academy of Social Science Evaluation Studies, CASS, Beijing, 100732)

(National Science Library, Chinese Academy of Sciences, Beijing, 100190)

**Abstract** Through measure the novelty of academic achievement and its cited situation, the paper studied the novelty and influence from the different levels of the academic achievements. The results show that there is a significant correlation between the novelty of academic achievement and academic citations, and the cited life cycle in different levels of novelty is significantly different.

**Keywords** novelty of theme, academic citation, citation life cycle

## 1 引言

学术成果作为科研活动的重要学术产出,是衡量科研水平的一项重要指标,但是对其学术质量的评判往往呈现出滞后性的特征,特别是在基础研究领域,重要的理论创新必须经过时间和实践的检验后其学术价值才能得以彰显。而这里说到的学术质量与学术价值又是一个特别复杂的概念,现有的评价系统都较难以对其进行直接地度量 and 客观地评判。因此众多的评价主体都将评价对象进行了分解和重构,以期通过对某些指标的测度来探讨学术质量,在我国自然科学领域中,曾长期把新颖性、先进性和实用性作为评价科研成果学术质量的主要指标,并且认为新颖性是先进性与实用性的基础<sup>[1]</sup>,根据《科技查新规范》的释义解释,新

颖性是指在查新委托日以前查新项目的科学技术内容部分或者全部没有在国内外出版物上公开发表过,因此新颖性就具有较强的时序特征<sup>[2]</sup>,目前基于新颖性的查新服务已经在自然科学领域产生了广泛应用。同时该标准也同样适用于技术领域,依据我国《专利法》第22条之规定,申请专利的发明或者实用新型满足新颖性的标准,必须不同于现有技术,同时也不得出现抵触申请,因此,申请发明和实用新型专利应具备新颖性是授予其专利权的必要条件之一<sup>[3]</sup>。在社会科学领域,学术成果的新颖性也是其评价的一项重要指标,陈永胜和孙昌玲<sup>[4]</sup>将社科研究成果的新颖性归结为“事实”“观点”“解释”“体系”“方法”五个维度,并试图构建了社会科学查新工作指标体系的结构模型。综上所述,不管是自然科学领域、技术领域还是人文社会科学领域,在

开展学术评价与学术成果质量判断时,新颖性都是其中的一个重要方面。与此同时,基于学术引用的文献计量学方法作为当前学术评价体系中的一项重要指标,在学术评价活动中发挥了重要的作用,也衍生出了一系列评价指标,如影响因子、H指数、G指数等,从学术引用的角度对学术活动的主体和被引对象进行了描述<sup>5-6</sup>。

总的来说,学术引用情况能够从一定程度上反映学术成果的传播能力和影响力,是后续学者针对前期成果开展的更加深入或系统的研究探索,而主题新颖性则能够反映学术成果在某种程度上的创新性和前沿性。因此,如果能综合两种测度方法,对学术成果的创新性评价将会更加科学。但是从另一个角度来说,这两种方法在对学术成果评价上是否具有一致性,主题新颖性高的成果是否能在发布之后的一段时期内产生较高的学术引用与学术影响,这也是当前需要论证的一个问题。两种测度方法计算结果的相关性是实现数据融合的关键前提和基础,基于此,本文将针对学术成果主题新颖性与其学术引用情况之间的相关关系这一命题进行研究和探索,以为后续研究提供参考和基础支撑。

## 2 相关研究

在学术成果主题新颖性测度方法上,主要有专家定性评价方法和计算机辅助定量评价方法两种思路。基于专家定性评价的主题新颖性测度依托于学术同行的专业知识和对学术前沿的把握能力,以同行评议的形式对学术成果的学术价值进行判断,其中同行评议这一概念来源于公元9世纪阿拉伯医学家Al-Ruhawi所著的《医师伦理》一书,并在之后的学术活动中被不断地丰富和完善<sup>7-8</sup>。近代以来形成的同行评议联盟对其进行了规范性表达,认为学术工作中的同行评议是同一学术研究领域的一些专家长期从事的一个职业,其主要任务是在学术研究成果发表前对其所包含的信息进行核查与验证<sup>9-10</sup>。我国学者在研究中对其功能又进行了扩展,赋予了同行评议更多的评价功能,叶继元等<sup>11-12</sup>认为同行评议是由具体从事某领域研究的专家根据一定的标准和程序对同一领域的科学研究活动及相关要素进行评价的方法和评价活动。总的来说,同行评议的重点在于选取合适的评价主体,即构建出科学、合理、公开、透明的领域小同行是实施同行评议的前提和基础。

基于计算机辅助的定量测度方法则侧重于采用技术的手段和方法,通过引文分析、文本语义挖掘等方法,对学术成果发布后的学术传播与影响、学术研究主题的演化关系等进行量化的测度和分析,从而揭示学术研究成果在特定数据集内的学术扩散能力与学术影响力。在具体操作过程中,学者们对引文分析方法也进行了细化,如Leydesdorff等<sup>13</sup>提出学术引用应该被进一步区分为长期引用和短期引用,其中长期引用能够体现论文的创新性与学术影响力,重在反映学术研究的深度;短期引用则可以说明学术成果处于研究前沿或研究热点之中,反映了研究主题的先进性,也即新颖性。但是严格来说,基于引文分析的学术评价活动实际上只是学术成果在学术影响力上的表层反映或者说不完全反映,过于重视引文的评价属性可能会导致其偏离学术成果的实际内容越来越远<sup>14</sup>。由于上述基于引文的学术评价的局限性,学术评价领域的研究人员开始将目光转向学术成果的内容本身,通过对学术成果的内容分析与挖掘来构建学术成果的研究主题新颖性测度指标,从而展示学术研究成果内容间的差异,与此同时也产生了一系列相关的测度算法。如沈阳<sup>15</sup>基于关键词在文中出现的频率、被用户检索的频率、使用的时间跨度周期及用户对关键词的创新度评价等维度构建了文献的创新度测算指标。杨建林和钱玲飞<sup>16</sup>在此基础上进一步提出了可以通过关键词的交叉率和共现关系来量化研究学科的创新能力,构建了基于关键词对逆文档频率的主题新颖度度量方法,并进行了实证研究。这些基于关键词的分析方法虽然可以在一定程度上反映或揭示学术成果中所包含的重要主题、概念、思想、模型或研究方法,但是由于作者在标注关键词时会存在词语表达的随意性,关键词中不规范或不合理之处比较常见,而这些不规范的关键词会对主题新颖度量化测度结果的准确性产生影响。随着当前计算能力的大幅提高,在大数据环境下采用深度学习的方法进行文本特征的语义扩展和内容计算已经成为可能,因此,有学者构建了基于深度学习的主题新颖性测度算法,通过对文本语义相似度的计算构建相似度矩阵,并在此基础上对各文本间的相似性强度关系的挖掘,进行主题新颖性测度<sup>17</sup>。从一定程度上来说,基于内容的文本语义分析技术能够较为客观地刻画和描述文本内容信息,并在此基础上可以实现无监督的文本主题新颖性的量化实时测度。

作为学术评价的一个重要维度,基于引文分析与基于内容挖掘的主题新颖性测度方法都从不同维度上研究和展示了学术成果主题间的演化关系,但是较高主题新颖性的学术成果是否在后续学科发展过程中能产生较强的学术影响,其被引生命周期将会如何演变?研究和回答这一问题,对进一步研究学科发展、开展学术评价、探测学术前沿与趋势等都将有一定的支撑作用。

### 3 研究数据与方法

#### 3.1 数据选择

由于不同学科的学术资料使用习惯、学术引用对象、学科引用半衰期等方面存在较大差异<sup>[18-19]</sup>,因此本文在数据源及测度对象的选择和实验设计上,就需要综合考虑这些影响因素。本文研究依托于中国社会科学评价研究院开发的中国人文社会科学期刊引文数据库进行学术被引情况统计,由于该数据库是基于期刊的引文数据库,因此,为了准确地测度学术成果主题新颖性与学术引用之间的相关关系,在研究对象的选择上,就需要选择期刊使用率较高的学科。据统计,在我国人文社会科学领域,图书情报学领域期刊利用率相对较高,该学科期刊所刊载的论文的参考文献中期刊论文的比例为64.32%<sup>[19]</sup>,此外,从各学科学术引用周期特征来看,图书情报学的平均引文半衰期为2.8年<sup>[18]</sup>,因此,为了最大限度地考虑学术成果在发布后的被引情况及新颖主题在较长时间周期内的变化情况,进而研究学术成果主题新颖性与学术引用之间的相关关系,本文选取图书情报学期刊论文实验数据源,选取该学科部分期刊2010~2016年发表的论文为测度对象。通过对该数据源中期刊在2010年所刊载的学术成果的主题新颖性进行测度,并对这些学术成果在2011~2016被引情况进行统计,来探讨学术主题新颖性与学术引用两者之间的关系。数据测度区间关系示意图如图1所示。

#### 3.2 研究方法

为了实现本文的研究目标,本文首先需要对学术成果的主题新颖性进行量化测度。在主题新颖性测度

方法的选择上,文献[17]通过对现有学术成果主题新颖性测度方法梳理,构建了基于深度学习算法和隐马尔科夫链的主题新颖性测度模型并与已有测度方法进行了对比验证,其实验结果表现出了较好的可解释性,不同学术成果的主题新颖性的区分度较为明显,同时,也克服了需要人为设置相似度阈值来进行主题新颖性测度的参数调试过程。因此,本文采用该研究成果中的主题新颖性测度模型进行学术成果主题新颖性测度。该模型首先采用Doc2Vec算法<sup>[20]</sup>对学术成果的文本主题相似性进行计算并构建相似性拓扑网络,在此基础上采用隐马尔科夫链算法<sup>[21]</sup>,整体考虑文档测试集内所有文本间的主题相似度关系,将其视作拓扑网络中主题关系转移矩阵,计算出该矩阵的特征因子,将主题新颖性测度从文本主题相似度计算和内容特征因子计算转化为文本主题新颖度测度指标与内容特征因子的单调递减函数关系。其具体思路表现为在测试集内的文本相似度网络和主题转移矩阵中,通过综合考虑某一文本主题与其他文本主题之间相似度高低和相似文本的数量大小,来判断该文本的主题在测试集内的主题差异性。如果网络中与该节点相似的节点数量越少、相似度越低,则该文本主题与其他文本主题间的差异性越高,计算结果中该节点的内容特征因子越低,其主题新颖性越高。这一思路的公式化表示为:

$$Cef(D_i) = \sum_{D_j \in M} \frac{\text{weight}(D_j) \times Cef(D_j)}{\text{degree}(D_i)} \quad \text{公式(1)}$$

$$\text{Nov}(D) = e^{-Cef(D)} \quad \text{公式(2)}$$

其中,Cef(D<sub>i</sub>)表示文档D<sub>i</sub>的内容特征因子(Content Eigenfactor,简称Cef),M是所有与文档D<sub>i</sub>存在相似度关系的文档集合,weight(D<sub>j</sub>)是边关系(D<sub>i</sub>, D<sub>j</sub>)间的权重,degree(D<sub>i</sub>)是转移矩阵中文档D<sub>i</sub>的度数。Nov(D)表示主题新颖性指数(Novelty of Theme,简称Nov)。

在主题新颖性测度的基础上,本文着重研究学术成果主题新颖性与学术引用之间的相关关系,为了使测度结果准确可靠且具有可比性,本文在对学术成果被引情况进行统计时,以学科被引半衰期为统计数据周期,统计各个学术成果在刊出之后的整体被引变化情况及其在半衰期前后的变化情况,作为其学术传播和影响力的具体体现,在此基础上研究学术成果主题新颖性与其学术引用的相关关系,并对比不同层次主题新颖性成

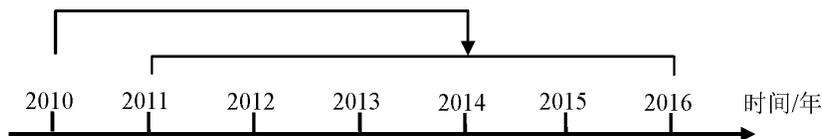


图1 图书情报学数据源选取及测度区间

果被引生命周期的变化情况。

#### 4 主题新颖性与引用关系研究

##### 4.1 学术成果主题新颖性测度及其被引情况统计

基于上述研究方法,本文选取图书情报学部分核心期刊作为研究对象,通过测度其2010年发表学术论文的主题新颖性及其在2011~2016年被该学科期刊群的引用情况,来进行主题新颖性与被引的相关关系研究。该数据源期刊及其所载学术论文数量如表1所示。

表1 2010年数据源期刊及其刊载论文情况

刊名	2010年发文量
大学图书馆学报	132
情报科学	393
情报理论与实践	368
情报学报	137
情报资料工作	145
图书情报工作	784
图书情报知识	95
图书与情报	217
现代图书情报技术	146
中国图书馆学报	91

以上期刊中2010年所载的2508篇文章为研究对象,通过采用Doc2vec算法两两计算其相似度,构建学术论文的相似度矩阵,在其基础上采用上文公式(1)和公式(2)的算法计算该矩阵所构成的拓扑网络中的节点的特征因子,从而得到每个节点的内容特征新颖性,即每篇文章的主题新颖性,部分计算结果见表2所示。

表2 主题新颖性计算结果

序号	题目	新颖性
1	期刊核心竞争力的形成机理研究	0.5691
2	知识服务与服务科学	0.4806
3	区域图书馆联盟发展趋势及经验分析	0.4710
4	高校图书馆服务质量对感知有用性的影响研究	0.4661
5	基于读者隐性知识的图书馆知识管理	0.4055
6	基于OWL-S API的数字图书馆服务组合应用研究	0.4006
7	学科馆员集群的形成与凸显	0.3852
8	基于本体的CSSCI学术资源网络模型构建及其应用研究	0.3675
9	图书馆组织对象及其层次研究	0.3628
10	知识链接的构建方式研究	0.3457

在该数据源期刊内学术论文主题新颖性测度之后,本文对这些学术论文在2011~2016年被本学科期刊的引用情况进行了统计。由于篇幅所限,这里以论文所在期刊在各年的被引分布进行展示,计算结果如表3所示。从该表中可以看出,10种期刊中除《中国图书馆学报》之外,其余9种期刊2010年刊载的论文均在

刊出的第三年达到被引高峰,即该学科学术论文的被引峰值年为2012年。这一结论将被引半衰期一起作为重要时间节点在下文主题新颖性与被引关系的研究中进行讨论。

表3 论文所在期刊的历年被引情况

刊名	2011年 篇均被引	2012年 篇均被引	2013年 篇均被引	2014年 篇均被引	2015年 篇均被引	2016年 篇均被引
大学图书馆学报	0.61	0.70	0.61	0.55	0.37	0.25
情报科学	0.20	0.27	0.22	0.21	0.14	0.15
情报理论与实践	0.36	0.41	0.33	0.33	0.22	0.24
情报学报	0.43	0.69	0.62	0.58	0.31	0.26
情报资料工作	0.31	0.43	0.31	0.20	0.28	0.11
图书情报工作	0.34	0.41	0.30	0.30	0.24	0.20
图书情报知识	0.52	0.51	0.32	0.36	0.28	0.22
图书与情报	0.38	0.37	0.31	0.25	0.20	0.18
现代图书情报技术	0.35	0.44	0.44	0.31	0.29	0.15
中国图书馆学报	1.60	1.55	1.44	1.58	1.37	0.96

##### 4.2 主题新颖性与学术引用关系研究

在学术论文主题新颖性测度与被引情况统计的基础上,通过采用SPSS对学术论文主题新颖性与学术被引关系进行研究,同时为了进一步探讨学术成果的主题新颖性及其被引关系随时间的变化情况,本文以年为单位,研究不同新颖性程度的学术成果在之后的被引情况。经统计,主题新颖性成果与被引关系存在显著相关性,表现为0.01水平上的显著相关,具体计算结果见表4。

表4 主题新颖性与被引相关性

主题 新颖性	Pearson 相关性 显著性 (双侧)	刊出1年 后总被引	刊出2年 后总被引	刊出3年 后总被引	刊出4年 后总被引	刊出5年 后总被引	刊出6年 后总被引
		.058**	.066**	.062**	.057**	.058**	.056**
	.004	.001	.002	.005	.004	.005	

注:“\*”在.01水平(双侧)上显著相关。

上述结果可以表明主题新颖性与学术引用之间在整体上存在显著的相关关系。进一步分析该相关系数和P值可以发现,主题新颖性与被引的关系在学术成果在刊出2年后达到相关系数最大值、P值为极小值,之后有相关系数有所降低,但是在5年后相关性又有所回升,具体表现相关系数有所提升、P值有所减小,这一结果也可以从一定程度上表明主题新颖性较高的学术论文的生命周期相对更长。为了进一步验证这一假设,本文选取主题新颖性排名前1%和后1%的论文(各25篇)、主题新颖性排名前5%和后5%的论文(各125篇)在刊出后各年的变化情况进行论证。

基于上文对学术论文在各年的被引情况的统计分

析,结果显示2010年学术论文一般在刊出后的第二年即2012年被引量达到高峰,因此我们将该年标注为被引峰值年,在对比不同主题新颖性的学术成果的被引情况之外,也对同一类型的学术成果在被引峰值年前后的被引情况进行对比,来揭示主题新颖性与被引之间的关系。

从图2和图3可以明显看出,主题新颖性较高的学术论文的被引情况要远高于主题新颖性较低的学术论文,可以从一定程度上说明在这一学科领域中,主题新颖性较高的学术论文更容易激发学者的研究思路或者在研究共同体内找到共鸣,从而激发出在此成果基础上新的研究成果。同时可以看出,在论文被引峰值年之前,主题新颖性较高的学术论文与主题新颖性较低的学术论文的被引量都表现出较快增长速度,但是在峰值年之后,主题新颖性较高的被引量增长速度要高于主题新颖性较低的学术论文的被引量,具体可以通过图中2012~2013年的被引量曲线斜率来反映这一现象。但是在半衰期之后,主题新颖性较高的学术论文仍表现出较强的被引生命力,被引曲线仍以较高的斜率向上增长,而主题新颖性较低的学术论文其被引曲线基本趋于平稳,增长量几乎为0,被引生命基本停止,其学术传播与影响力几乎消失。

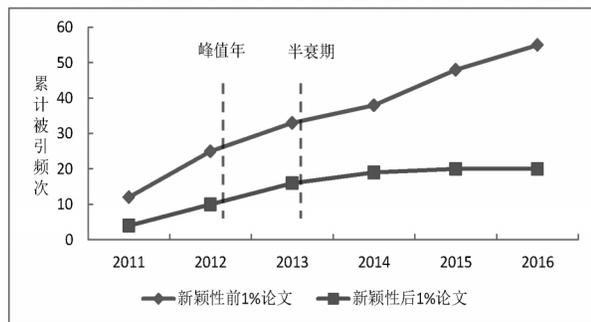


图2 主题新颖性前后1%的学术论文的被引情况

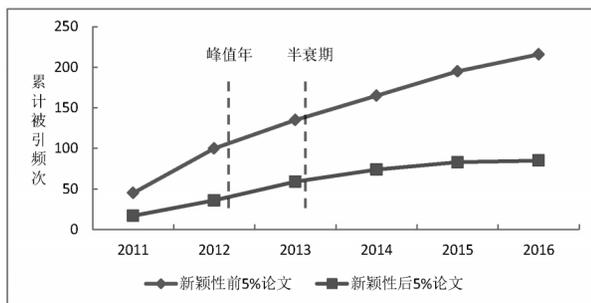


图3 主题新颖性前后5%的学术论文的被引情况

进一步对比主题新颖性前后1%和5%的学术论文的篇均被引情况,我们发现,主题新颖性测度结果中排名靠后的论文其篇均被引情况保持较强的一致性,从图4中可以看出,主题新颖性较低的后1%和5%论文的篇均被引曲线几乎完全重合,且在半衰期之后的被引增量几乎为0,学术传播与影响力基本消失。该现象说明,从分层对比的结果来看,不同层次上主题新颖性较低的学术论文的被引生命周期存在一致性。而再对比主题新颖性较高的前1%和前5%的学术论文的被引情况,我们发现,在被引半衰期之前,包括被引峰值年在内,主题新颖性较高的前1%和前5%论文的篇均被引情况存在一致性,但是在半衰期之后的两年内,主题新颖性较高的前5%论文的篇均被引情况要高于前1%的论文的篇均被引情况,但是从半衰期之后的第三年起,主题新颖性较高的前1%的论文的篇均被引又开始快速增长,这一现象似乎可以说明,主题新颖性较高的学术论文中也存在一定数量的冷门主题或者说是前沿主题,这些主题可以在更长的时间周期后产生影响,对这类论文进行探测和研究,可以从一定程度上预测本学科领域的研究趋势。

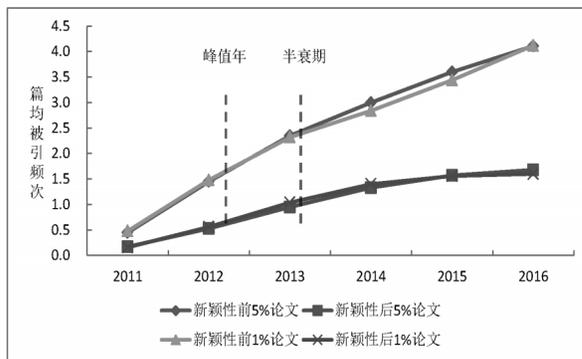


图4 主题新颖性前后1%和5%的学术论文的篇均被引情况

## 5 结论与不足

本文在对学术成果主题新颖性测度研究方法现状进行调研的基础上,采用了基于Doc2vec算法和隐马尔科夫链模型所构建的学术成果主题新颖性测度模型,对我国图书情报学领域部分学术期刊上所刊载的学术论文的主题新颖性进行测度,对这些论文在刊出之后的被引情况进行统计分析,在此基础上,通过对学术成果主题新颖性与学术引用之间的相关关系进行研究,得出以下结论:

(1)学术成果主题新颖性与学术引用之间存在显

著的正相关关系,主题新颖性较高的学术成果在刊出后其被引情况高于主题新颖性较低的学术成果。

(2)主题新颖性较高的学术论文的生命周期相对更长,在论文被引峰值年之前,主题新颖性较高的学术论文与主题新颖性较低的学术论文的被引量都表现出较快增长速度,但是在峰值年之后,主题新颖性较高的被引量增长速度要高于主题新颖性较低的学术论文的被引量。在被引半衰期之后,主题新颖性较高的学术论文仍表现出较强的被引生命力,被引曲线仍以较高的斜率向上增长,而主题新颖性较低的学术论文其被引曲线基本趋于平稳,增长量几乎为0,被引生命基本停止,学术传播与影响力几乎消失。

(3)从学术成果主题新颖性分层对比的结果来看,不同层次上主题新颖性较低的学术论文的被引生命周期存在一致性;而在不同层次上主题新颖性较高的学术论文的被引关系上却存在不同,主题新颖性较高的学术论文在半衰期之后的被引生命周期相对更长,同时不同层次主题新颖性的学术论文的被引生命周期也表现出一定的差异性。

由于篇幅所限,本文未对主题新颖性较高的论文的主题分布情况进行详细研究和阐述。在后续研究中,需要进一步分析不同层次主题新颖性的学术论文的被引生命周期存在差异性的原因,并对更长时间周期内主题新颖性较高的论文中主题演化关系进行研究,从而对其中的冷门主题和前沿主题进行识别。在此基础上可以总结学科领域当前的研究热点,探测学科领域的研究前沿,并对未来的研究趋势进行预测。

### 参考文献

- [1] 张京辉,胡淑礼,王亚非,等.软科学成果查新研究[J].软科学,1993(4):28-33.
- [2] 《科技查新教程》编写组.科技查新教程[M].北京:机械工业出版社,2001.
- [3] 尹新天.中国专利法详解[M].北京:知识产权出版社,2011.
- [4] 陈永胜,孙昌玲.社会科学查新工作指标体系的结构模型与价值分析[J].情报资料工作,2002(6):6-8.
- [5] Bertoli-Barsotti L, Lando T. The h-index as an almost-exact function of some basic statistics[J]. Scientometrics, 2017,113(2):1209-1228.
- [6] Abbas A M. Bounds and inequalities relating h-index, g-index, e-index and generalized impact factor: an improvement over existing models[J]. PLoS One, 2012,7(4):e33699.

- [7] Farrell P R, Magida F L, Farrell M K. Ancient texts to PubMed: a brief history of the peer-review process[J]. J Perinatol, 2017, 37(1):13-15.
- [8] Spier R. The history of the peer-review process[J]. Trends Biotechnol, 2002,20(8):357-358.
- [9] Stamps A E. Advances in peer review research: an introduction [J]. Science & Engineering Ethics, 1997,3(1):3-10.
- [10] Peer Review of Scholarly Journals[EB/OL]. [2018-04-17]. http://peereviewer.com/freelance-peer-review-scholarly-journals/.
- [11] 叶继元.引文法既是定量又是定性的评价法[J].图书馆,2005(1):43-45.
- [12] 贺颖,刘慧,刘友存,等.基于可视化ACA技术的同行评议专家学术范式辨别研究[J].图书情报工作,2010(14):28-31.
- [13] Leydesdorff L, Borrmann L, Comins J, et al. Citations: indicators of quality? The impact fallacy[J]. Frontiers in Research Metrics and Analysis, 2016(1):1-26.
- [14] Hirsch J E. An index to quantify an individual's scientific research output[J]. Proc Natl Acad Sci USA, 2005,102(46):16569-16572.
- [15] 沈阳.一种基于关键词的创新度评价方法[J].情报理论与实践,2007(1):125-127.
- [16] 杨建林,钱玲飞.基于关键词对逆文档频率的主题新颖度量方法[J].情报理论与实践,2013(3):99-102.
- [17] 逯万辉,谭宗颖.学术成果主题新颖性测度方法研究——基于Doc2Vec和HMM算法[J].数据分析与知识发现,2018(3):22-29.
- [18] 任全娥,郝若扬.基于文献引证关系的人文社会科学论文评价[J].大学图书馆学报,2012(3):111-118.
- [19] 苏金燕.我国人文社会科学领域文献使用特征研究[J].图书情报工作,2015(22):118-123.
- [20] Le Q V, Mikolov T. Distributed representations of Sentences and Documents[J]. Computer Science, 2014(4):1188-1196.
- [21] Sheldon D. Manipulation of Pagerank and collective hidden Markov models[D]. Cornell University, 2010.

【作者简介】逯万辉,男,1987年生,中国社会科学院中国社会科学评价研究院助理研究员,中国科学院文献情报中心博士研究生。

苏金燕,女,1979年生,中国社会科学院中国社会科学评价研究院副研究员。

余倩,女,1982年生,中国社会科学院中国社会科学评价研究院助理研究员。

收稿日期:2018-06-12