

# 基于创新知识元谱系的学术论文新颖性测度研究\*

■ 戎军涛 索传军 周彦廷 李艺亭

中国人民大学信息资源管理学院 北京 100872

**摘要:** [目的/意义] 基于“频次”与基于“引文”的评价方法本质上是对学术论文“影响力”的评价,与学术创新性并不完全对等,从内容层次和语义层次对学术论文进行新颖性测度能够为学术论文创新性评价提供新思路。[方法/过程] 研究遵循数据驱动的学术评价范式路径。首先,基于分类比较的思想构建创新知识元谱系的概念模型;其次,基于创新知识元谱系构建学术论文新颖性测度模型;最后,通过实证验证模型的适用性。[结果/结论] 实证结果表明,研究所提出的新颖性测度模型借助创新知识元谱系作为参照系,采用分类比较思想进行评价,能够实现从语义内容层面测度学术论文的新颖性。但该模型同时存在谱系适用性的局限。

**关键词:** 创新知识元谱系 学术评价 创新性评价 新颖性测度

**分类号:** G250

**DOI:** 10.13266/j.issn.0252-3116.2024.01.003

## 1 引言

广大科研人员对学术成果的需求问题,已经从过去的发现与获取困难,转变为目前的判断与选择困难。帮助科研人员从庞杂的海量文献中选择出高质量的文献,特别是选择出具有创新性的知识元,节省科研人员的知识查找、选择、阅读和利用的时间,是学术论文创新性评价的直接目的和重要价值<sup>[1]</sup>。从学术论文的内在属性来看,学术论文评价的本质体现为学术论文创新性的测度<sup>[2]</sup>。

对学术论文的创新性评价最终要落实到对论文所承载的创新知识元的计量和评价上,即借助知识元最终实现对论文作者思想和观点的认知和评价。如何从学术论文内容层面对知识元进行新颖性判断和测度,是困扰学术论文创新评价的重要问题。创新知识元是创新性科学知识的基本构成单元。因此,对学术论文评价的实质是对论文中创新知识元的识别、计量和评价。对知识元如何做到准确计量和解读,是学术论文创新性评价的关键。但是时至今日,知识计量依然是情报学和计算机科学研究的难题,目前还没有确定的、统一的、有效的知识元用于知识的直接计量。

数据智能时代,学术论文是一组知识元的逻辑组合,表现为一组相关的数据对象集合,这为基于内容的学术论文创新性评价提供了可能的实现路径。索传

军<sup>[3]</sup>教授提出了数据驱动的学术评价范式,明确了学术论文的创新性评价路径。具体来说,数据驱动的学术评价路径是指:将学术论文中的“创新知识元”识别与抽出,进行语义标注,形成以创新知识元为核心的学术论文语义内容表示形式;构建经过形式化表示的创新知识元谱系为参照系,通过识别待评价知识元在创新知识元谱系中的位置与重要度,与创新知识元谱系中相关节点的分析、推理,输出具有一定概率的确定性创新测度结果。文献的数据化、知识元语义描述、创新知识元谱系构建和新颖性测度计算就构成了学术论文创新性评价的关键环节。

本研究是数据驱动范式下学术论文创新性评价的尝试性探索,立足从语义层面探索创新知识元的新颖性测度问题,通过将待测评知识元与创新知识元谱系进行语义比对,进而判定和测度知识元的新颖性。具体来说,创新知识元新颖性测度的关键问题主要有两个方面:一是创新知识元谱系的构建;二是学术论文新颖性测度模型的构建。

## 2 相关研究工作

目前,基于内容的学术论文创新性评价还存在诸多困难,现有研究对新颖性测度普遍采用的是比较方法,也就是通过比较两个内容对象的属性,来判断学

\* 本文系国家自然科学基金项目“数据驱动的学术评价范式研究”(项目编号:21BTQ008)研究成果之一。

作者简介:戎军涛,博士研究生;索传军,教授,博士生导师,通信作者,E-mail: suocj@ruc.edu.cn;周彦廷,博士研究生;李艺亭,博士研究生。

收稿日期:2023-08-10 修回日期:2023-11-03 本文起止页码:27-38 本文责任编辑:徐健

术论文内容的新颖性,有两种方式:①不同的评价对象之间进行相互比较;②确定标尺,将待评价对象与标尺进行对比。根据内容对象的类型,可以将新颖性测度方法分为基于文本特征的比较方法和基于知识图谱的比较方法。

### 2.1 基于文本特征的比较方法

当前研究普遍用关键词和创新语句来代替文本内容,通过关键词及其相关因素的计算来测度论文内容的新颖性<sup>[4]</sup>,典型方法有词频统计法、关键词交叉测度法和语义相似度计算法。

词频统计方法是关键词词频与所有文档的关键词频进行比较,关键词的频率越高,时间跨度越长,则关键词的创新度越低<sup>[5]</sup>。同样,通过计算词频差异程度较高的突变词,可以有效探测突破性创新<sup>[6]</sup>。虽然关键词可以在一定程度上反映学术成果中包含的重要主题、概念、思想、模型或方法,但是使用单个关键词仍然不足以准确刻画和反映学术成果的主题新颖性。将词频与引文网络结合进行创新测度,可以一定程度上降低关键词词频统计方法的局限<sup>[7]</sup>。近年来,基于共现思想的关键词对频率计算方法被用来测度主题的新颖性<sup>[8]</sup>。利用关键词组合来识别突破性论文已经成为新的技术路径<sup>[9]</sup>。但是该方法在计算关键词词频或者相似度的时候并没有考虑关键词的语义功能信息,容易导致语义功能不同的词新颖值相互影响。将关键词按照语义类别划分成问题词、方法词后再进行词对组合计算,可以一定程度上弥补该方法的局限<sup>[10]</sup>。

交叉测度法通过测定关键词的交叉程度来确定学科的创新力。该方法认为从不同学科而言,学科之间交叉越多,越容易出新成果,学科创新力越强,从学科内部而言,学科内部不同研究领域的交叉越多,该学科创新潜力越强<sup>[11]</sup>。

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相似度计算已经成为可能<sup>[12]</sup>。语义相似度计算思路通常有两种路径:①将论文的研究主题与科学研究前沿主题进行相似度计算,其测度结果认为如果某篇学术论文的研究主题和当前的科学研究前沿主题相契合,同时发表在影响因子较高的期刊上,这篇论文即具有较高的创新力<sup>[13]</sup>。②利用 Word2Vec 工具包将文本特征的计算从词语层面扩展到句子层面,通过计算文本相似度来判断文本的新颖性<sup>[14]</sup>。该技术方法将句子看成词典集合,本质上依然是基于词的语义相似度计算。

### 2.2 基于知识图谱的比较方法

知识图谱作为一种揭示实体间关系的语义网络,是知识结构的最新表示形式,也可以看作“认知地图”的高级形态。基于知识图谱的比较思想是利用知识图谱挖掘和呈现特定领域现有科技论文中的知识元素,作为新发表论文创新点识别的比对库,发现论文中新出现的实体或关系,即创新点。通过与知识图谱的对比可以判断最新研究成果中是否出现了新的观点、技术或方法,重复创新点的识别与填充,能够动态监测该领域的科研创新。最重要的是,研究人员可以根据知识图谱中实体及其相互联系,发现创新的突破口<sup>[15]</sup>。

以上研究提供了重要的理论依据和方法参考。当前对于知识的计量仍处于词汇层面,大都仅考虑了主题词、短语等孤立的元素,以主题词、关键词等部分属性代替知识进行评价,未考虑到它们之间的语义关系。词频统计更多地反映研究热点、相似度计算反映主题差异,并不适合直接用来判断创新性,也不能历时、动态地监测论文创新性。创新知识元相对完整、集中的反映了学术论文的创新点,是不同于主题词、关键词的知识表达形式。本研究将通过语义比较,实现创新知识元的新颖性测度。

## 3 创新知识元谱系构建

### 3.1 创新知识元谱系

知识元是学术文献中相对独立的、具有完备知识表达、最小粒度的语义单元<sup>[16]</sup>。创新知识元是知识元的一种类型,通常指学术成果中表达作者在探究事物或现象本质和规律基础上所获取的知识,以及作者在传播和运用已知知识过程中所获取的新发现的知识单元<sup>[17]</sup>。简单的说,创新知识元就是学术成果中由创新点构成的基本单元。

创新知识元谱系是学术论文新颖性比较和判断的“评价参照系”,属于学术谱系的一种类型。从学术史角度看,概括地说,学术谱系就是基于时间序列与某一研究主题或问题相关的一组研究发现构成的网络图谱,其基本特征是学术传承<sup>[18]</sup>。图谱中的节点是各类学术要素,如文献、作者、概念、创新点、研究主题、研究问题,等等。图谱中的边是节点间的关系,包括历时性的继承关系、顺承关系以及共时性的并列关系、相关关系等,可以用这种形式化的方式描述科学成果之间的发展演变关系。创新知识元谱系就是以创新知识元为节点,以知识元之间的创新演化

发展关系为边的主题演变脉络图谱。

### 3.2 创新知识元谱系的概念模型

任何学术成果都有明确的研究主题, 每个研究主题都蕴含着一组随着时间变化的相互关联的问题。因而, 任何问题都不是孤立存在的, 它们是相互联系的, 是一个具有一定数据结构的问题网络, 称为主题问题谱系<sup>[19]</sup>。任何学术论文所研究的问题, 都是其所在主题的问题网络中的一个节点。因此, 主题—问题框架就构成了创新知识元谱系的概念模型的基础架构, 其控制着创新知识元谱系的实例数据, 即创新知识元实例。为方便创新评价的分类比较, 本研究将创新知识元谱系的概念模型设计成分类分面形式, 主题框架部分包括创新对象、主题对象、主题域, 问题框架部分包括问题及问题链, 具体结构关系如图 1 所示:

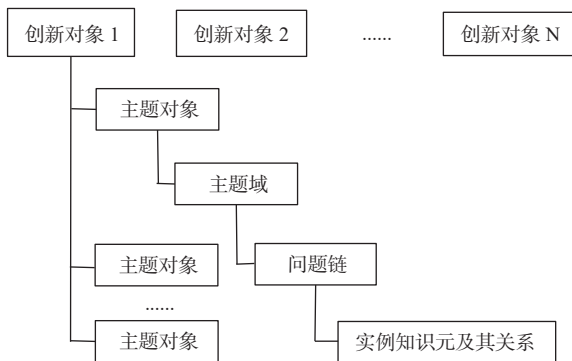


图 1 创新知识元谱系概念模型

Figure 1 Conceptual model of the innovative knowledge element genealogy

#### 3.2.1 创新对象

创新知识元谱系存在多种类型。评价只能在同一类型的创新知识元谱系中进行。创新对象是创新知识元谱系依据创新要素所划分的创新类型。学术论文的评价要与相应创新对象门类下的创新知识元谱系进行比对。通常, 学术论文创新对象的类型分为事实材料、理论问题和研究方法<sup>[20]</sup>。根据学术论文的结构或逻辑架构, 可将创新知识元谱系的创新对象划分进一步扩展为问题创新、原理创新、方法创新、数据创新和应用创新。

#### 3.2.2 主题对象

创新对象门类下包含若干研究主题。主题对象指学术论文研究主题所指涉的研究客体。主题对象可分为实体性研究对象和构念型研究对象。实体型研究对象指自然界、人类社会中客观存在的物体、现象、行为、活动等事物对象; 构念型研究对象指以非实体形式存在的事实、事理、观念、思维等理念对象。

#### 3.2.3 主题域

主题域指学术论文所论述的研究对象的具体指向和方面。通常, 一个研究对象会有多个主题域, 不同的主题域之间是并列关系。主题域通常包括材料、性质、结构、内外关系、影响因素、发展趋势、动力机制、演化路径、因果关系、历史背景、意义、目的、价值、策略等因素。

#### 3.2.4 问题链

主题域中包含着若干研究问题。研究问题是学术论文主题中所要探索的未知事项, 通常蕴涵着问题的指向、研究的目标和求解的应答域。问题链是问题之间依据内在逻辑发展关系形成的学术链条。根据特定主题领域内具体问题的研究发展程度, 可以形成不同发展阶段的问题集合。将这些问题集合依据内在逻辑关系连接起来就构成了问题链, 可以显示学术发展的脉络。问题链直接控制着实例知识元, 具体表现为实例知识元之间的发展、演化关系链。

### 3.3 创新知识元谱系中的节点关系

创新知识元谱系的节点是创新知识元, 节点关系主要围绕特定主题领域知识元之间的创新发展演化的脉络关联来构建, 而并非是对知识元的所有语义关系进行构建。通过对创新方式进行分类, 来获取创新知识元之间的创新发展关系。国外学界通常将创新划分为破坏性创新和连续性创新<sup>[21]</sup>。国内学界则根据创新的原创程度, 通常将创新划分为原始创新 (0—1 的创新) 和累积创新 (1—10 的创新)<sup>[22]</sup>。按照知识转移的方式和创新程度, 累积创新可进一步分为借鉴式创新、继承式创新和发展式创新<sup>[23]</sup>。本研究采用国内关于创新的划分思想; 创新知识元谱系中的节点关系主要指累积创新型语义关系。

原始创新是指在解释新现象或解决新、老问题过程中获得全新的科学发现, 开辟一个新的研究方向、新的研究领域或创立、构建一套新的理论体系、学说, 或第一次得到全新的知识发现或技术发明、创造, 其结果是全新的、自主的、首创的。依据创新的发展阶段和范围, 原始创新具体可分为起点原创、局部原创和整体原创三种类型。累积创新也叫渐进式创新, 是指在原有创新基础上的进步, 一般指对原有理论、方法、技术、产品的改良、改进性发展, 以使其更加完善, 性能或能力得到提升。依据创新的演化程度, 累积创新具体可分为继承、演进和跃迁三种类型。创新类型与方式如表 1 所示:



表1 创新类型与方式划分  
Table 1 Classification of innovation types and methods

创新类型	创新方式	解释说明
原始创新	整体原创	整体原创是一种普遍创新,指在其领域内取得的全局性的、革命性的、历史性的、质变性的创新成果,属于问题解决阶段的突破式或颠覆式创新
	局部原创	局部原创是一种阶段性创新,指在其领域内取得的具有局部性、革新性、当下性、量变性质的阶段性创新成果,属于问题探索阶段的创新
	起点原创	起点原创是指仅提出全新的科学理论观点、范畴或见解,但并未具体展开论证或构建理论学说体系,属于问题提出阶段的创新
累积创新	继承式创新	继承式创新是指在前人的创新成果基础上对其局部进行补充、丰富、发展和扬弃,包括修正、改进、优化等
	演进式创新	演进式创新是在原有理论框架下发现了新原理、新规律,衍生出了新的结论与推断,是已有成果在新的情境或场景下的拓展、延伸和再创造
	跃迁式创新	跃迁式创新指在演进创新基础上取得了很大进展和突破,但总体上还处于量变积累阶段,并未形成颠覆性质变

## 4 学术论文新颖性测度模型

新颖性是创新性的重要测度指标。学术论文的新颖性是与其之前发表的学术论文的研究结论进行比较的结果。因此学术论文的新颖性评价需要依靠一定的参照系作为评价标准。具体来说,学术论文的创新性体现为其包含的创新知识元的新颖性。依据创新知识元谱系作为参照系,可以测度创新知识元的新颖性。

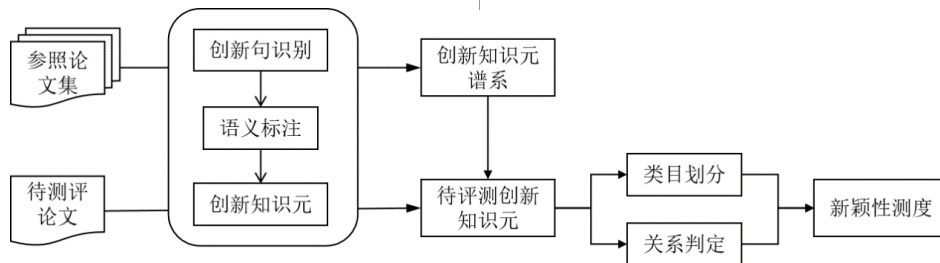


图2 创新知识元新颖性测度模型

Figure 2 Model of novelty measurement of innovation knowledge element

首先,对待测评论文进行创新知识元抽取和语义标注,获得创新知识元的语义三元组。

其次,根据知识元语义标注的结果,将知识元进行分类识别,确定其创新类型、研究对象、主题域和问题域,并将其定位到相应创新知识元谱系类下的问题链,并与该问题链中的创新知识元进行匹配。如果没有可匹配的节点,可将待测评创新知识元的类型定位为“原始创新”。

再者,将待测知识元与创新知识元谱系中的实例知识元进行比对,确定两个知识元之间的关系权值。具体来说,可以依据知识元所属的领域语义词典(即知识元标注用的描述词汇表)中的概念类的层次关系进行关系判断。领域语义词典是控制术语的概念分类层级系统,可以通过复用相关语义本体或自定义来构建;常见的语义词典有 WordNet<sup>[24]</sup>、同义词词林<sup>[25]</sup>、HowNet<sup>[26]</sup>等。通过计算术语或知识实体在概念层级

### 4.1 新颖性测度模型

新颖性评判通常是对学术论文创新情况的定性描述,是指作者在科技论文中所表达的科技成果或学术内容及其创新程度、创新类别和创新价值等方面的属性。从学术论文内容的角度看,含有创新知识元就表示有新颖性。将创新知识元与创新知识元谱系进行相似度比较是创新性测度的核心思想。创新知识元谱系是创新性评判的参照系,也是创新性评判的关键环节。基于创新知识元谱系的论文新颖性具体评测模型如图2所示:

中的位置、路径来测度两个知识元实体之间的相似或相关程度,是一种语义层面的概念比较方法。表2描述了知识元之间语义比较的关系判定规则和创新方式类型的判定。

表2 语义比较判定规则  
Table 2 Semantic comparison judgment rules

概念体系中的关系	关系内涵说明	创新方式
相近关系	领域的语义词典中两类知识元属于同一个分类体系下,即同族体系,那么类目之间相似度较高,关系为相近关系	继承
相关关系	领域的语义词典中两类知识元分属于两个分类体系下,即异族体系,那么两者之间为相关关系	演进
相异关系	两个知识元分属不同的体系且互不相关,则关系为相异关系	突破

最后,根据待测评知识元在创新知识元谱系中所属的位置进行计算。根据知识元之间的关系,参照新颖性测度指标体系(见表3)获取关系权值。依据新颖性测度函数,对待测评知识元进行新颖度计算。

表 3 新颖性测度指标体系  
Table 3 Novelty measurement indicator system

创新类型	原始创新			累积创新		
	整体原创	局部原创	起点原创	突破创新	演进创新	继承创新
问题发现类	1	0.8	0.6	0.5	0.3	0.1
理论构建类	1	0.8	0.6	0.5	0.3	0.1
方法创新类	1	0.8	0.6	0.5	0.3	0.1
事实数据类	1	0.8	0.6	0.5	0.3	0.1
应用发明类	1	0.8	0.6	0.5	0.3	0.1

## 4.2 新颖性测度函数

### 4.2.1 新颖性测度思想

新颖性测度是对学术论文创新情况的定量描述, 通常指对学术论文的整体创新所达到的程度或水平的测度。创新知识元的新颖度依据新颖性测度函数计算。根据创新知识元在谱系中的位置和语义关系来构建新颖性测度函数。

首先, 需要确定当前待测评的创新知识元所在谱系中的问题链以及其所在问题链中的位置, 确定根节点 (即源头起点) 的创新度, 也就是原始创新的新颖度。其次, 构建问题链条中的累积创新的语义关系的权重。根据累积创新的程度, 一般采用专家调研法来确定语义关系的权重。设定因素集为  $U=\{\text{问题发现类, 理论构建类, 方法创新类, 事实数据类, 应用发明类}\}$ 。依据创新程度的高低将新颖度的权重粗粒度设定为 0—1 区间, 即建立权重集为  $A=(1, 0.8, 0.6, 0.5, 0.3, 0.1)$ 。值越接近 1, 新颖度越高。将创新类型设定为评价集, 可设评价集  $V=\{\text{整体原创, 局部原创, 起点原创, 突破创新, 演进创新, 继承创新}\}$ 。以创新知识元类型为因素集, 以原始创新的初始值和累积创新的语义关系权重为评价集, 可以构建新颖性测度指标体系 (见表 3), 作为新颖性测度所需数值的计算依据。最后, 该问题链条上源头起点的原始创新度与当前位置的创新知识元之间路径的语义关系权重的乘积就是创新知识元的新颖度。创新知识元的新颖性会随着问题链的增长而降低, 这也符合创新性扩散的一般原理。

### 4.2.2 函数构建

创新知识元的新颖度可以用绝对新颖度和相对新颖度表示。绝对新颖度是当前知识元相对整个创新链条上的知识元所具有的新颖程度。一般来说, 原始创新用绝对新颖度表示。创新链条上源头起点的原始创新度与当前位置的创新知识元之间路径的语义关系权重的乘积就是创新知识元的绝对新颖度。相对新

颖度是指当前创新知识元新颖度与创新链条中的当前创新知识元之前所有创新知识元的绝对新颖度累积量的比值, 是一个可以用来比较参照的相对值。

根据上述测度思想, 笔者采用相对新颖度来设定新颖性测度函数, 如公式 (1) 所示:

$$\text{Nov}(\text{KE}) = \sigma \frac{P \prod_{n=1}^i S_i}{P + P \sum_{n=1}^i S_i} \quad \text{公式 (1)}$$

其中,  $\text{Nov}(\text{KE})$  表示创新知识元的相对新颖度;  $P \prod_{n=1}^i S_i$  表示当前创新知识元的绝对新颖度;  $P + P \sum_{n=1}^i S_i$  表示创新链条上当前创新知识元之前所有创新知识元的绝对新颖度的总和。 $\sigma$  表示新颖性系数, 一般为该主题领域相关问题的发文数量的倒数, 为方便计算可忽略;  $P$  表示问题链中源头起点的原始创新初始值, 代表原始创新的绝对新颖度;  $S_i$  表示创新知识元谱系问题链中第  $i$  个创新知识元的语义关系权重。初始值和语义关系权重可依据新颖性测度指标体系 (见表 3) 来获取。

## 5 实证分析

图书馆学研究对象问题是图书馆学理论研究领域的经典元问题。在图书馆学发展的两百年历史进程中, 中外各国图书馆学研究对象的表述就有数百种。特别是建国后, 我国图书馆学理论研究快速发展, 关于研究对象的学说先后有“要素说”“事业说”“规律说”“矛盾说”“交流说”“知识说”等众多理论观点, 彼此相互批判、融合, 形成了一幅流派众多、源流清晰的学术发展谱系。因此, 本研究以国内建国后“图书馆学研究对象”为主题, 进行实证分析, 来判断、测度相关知识元的新颖度, 从而验证学术论文新颖性测度模型的适用性和有效性。

### 5.1 数据集获取

本次实证研究采集了中国知网 (CNKI) 1949—2023 年 5 月、主题为“图书馆学研究对象”的论文 151 篇, 获取文献的标题、关键词、结构化摘要、作者、文献来源、发表时间、引言、结论、参考文献等信息作为基础数据集。经由人工判读对基础数据进行清洗。笔者主要选取研究结论或观点明确提出“图书馆学研究对象是什么”的理论型文献, 去除不同来源的重复文献、会议通知、报道、综述以及其他不相关的理论分析型文献, 进一步精炼、获取主题相关性更高的核心文献 50 篇。

5.2 创新知识元标注

学术论文文本内容具有很强的规则性，模式特征和语法特征比较明显，因此本研究选取了基于规则的

抽取方法。以索传军<sup>[27]</sup>、王晓光<sup>[28]</sup>、孙建军<sup>[29]</sup>、宋宁远<sup>[30]</sup>等提出的语义标注思想为标准进行创新知识元的实例标注，流程如图3所示：

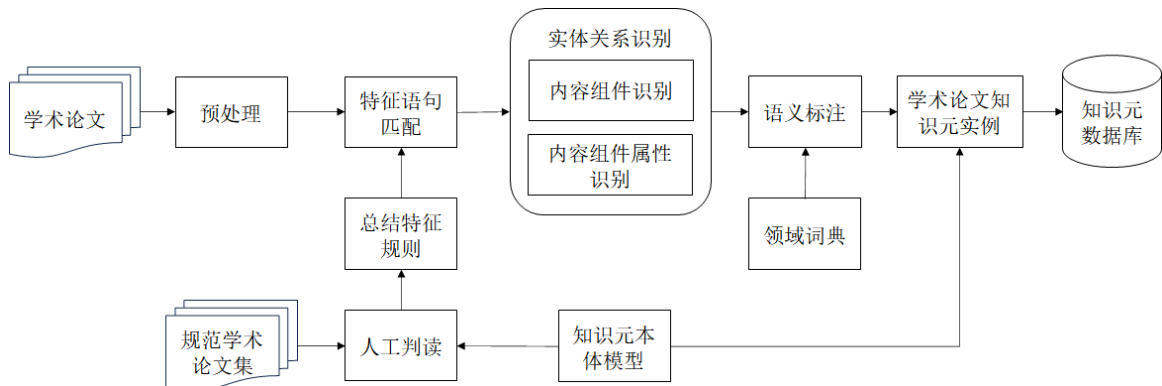


图3 知识元标注流程

Figure 3 Knowledge element annotation process

通过制定知识元内容的抽取规则来识别、获取知识元语句。抽取规则通过人工阅读进行构建。本研究关注的是关于“图书馆学研究对象”的观点、思想、理论、模型等内容，回答的是“图书馆学研究对象是什么”的问题。包含这类问题的创新知识元语句表达形式较为明显、规范，一般表现为“……是……”的形式。根据化柏林的抽取规则研究<sup>[31]</sup>，这类定义的主要组成为被定义体和内涵描述部分。根据这两部分所处位置，可以分为两种不同的模式：被定义项前置模式和被定义项后置模式。在参照索传军<sup>[32]</sup>、李贺<sup>[33]</sup>、赵蓉英<sup>[34]</sup>的知识元抽取规则方法基础上，设定了“图书馆学研究对象”领域的创新知识元语句抽取规则，如表4所示：

表4 “图书馆学研究对象”创新知识元抽取规则  
Table 4 Rules for extracting innovative knowledge elements on “objects of library science”

定义模式	抽取规则	实例
被定义项前置模式	图书馆学 + 研究对象   对象   + 是   应转向   定义为   表述是 + (……)	图书馆学所研究的对象就是图书馆事业及其各个组成要素
被定义项后置模式	(……) + 是   构成   作为 + 图书馆学 + 研究对象   对象	文献信息的开发与利用，便构成了图书馆学的研究对象

以自定义的创新知识元标注本体作为规范的领域词典，依据纳米出版物语义标注模型<sup>[35]</sup>来进行标注。经过标注后的三元组，不再是字符串形式的词汇，而是一个语义实体类。具体标注实例见表5。实例知识元内容表达的描述语句见图4。

表5 创新知识元标注实例

Table 5 Example of innovative knowledge element annotation

标注项	标注内容
知识元名称	“符号信息说”
知识元 ID	http://www.semanticweb.org/shiyanshi1/ontologies/2023/4/untitled-ontology-4#符号信息说
知识元标签	SIT (symbolic information theory)
内容文字描述	“符号信息是图书馆学的基本研究对象”
内容三元组表达	S:[# 符号信息说] P:[研究对象] O:[# 符号信息]

```
<!-- http://www.semanticweb.org/shiyanshi1/ontologies/2023/4/untitled-ontology-4#符号信息 -->
<owl:NamedIndividual rdf:about="http://www.semanticweb.org/shiyanshi1/ontologies/2023/4/untitled-ontology-4#符号信息">
  <rdf:type rdf:resource="http://www.semanticweb.org/shiyanshi1/ontologies/2023/4/untitled-ontology-4#符号信息"/>
</owl:NamedIndividual>

<!-- http://www.semanticweb.org/shiyanshi1/ontologies/2023/4/untitled-ontology-4#符号信息说 -->
<owl:NamedIndividual rdf:about="http://www.semanticweb.org/shiyanshi1/ontologies/2023/4/untitled-ontology-4#符号信息说">
  <rdf:type rdf:resource="http://www.semanticweb.org/shiyanshi1/ontologies/2023/4/untitled-ontology-4#符号信息说"/>
  <研究对象 rdf:resource="http://www.semanticweb.org/shiyanshi1/ontologies/2023/4/untitled-ontology-4#符号信息"/>
</owl:NamedIndividual>
</rdf:RDF>
```

图4 实例知识元内容表达的描述语句

Figure 4 Description statement for the expression of instance knowledge element content



### 5.3 创新知识元谱系的构建

学术传承的脉络梳理,需要重视学术传承过程中出现的标志性节点以及节点之间的关系。依据《中国图书馆学科发展史》<sup>[36]</sup>,并参考马恒通<sup>[37]</sup>、王子舟<sup>[38]</sup>、徐引篪<sup>[39]</sup>等相关综述理论文章,选取重要文献的代表性观点作为创新知识元节点,并进一步判断、构建创新知识元之间的继承、发展、演化关系。首先,依据上文提出的创新知识元谱系数据结构,对其进行属性标注。“图书馆学研究对象”问题属于图书馆学理论范畴,谱系的创新对象属于“理论创新”,“主题对象”属于“图书馆学”,“主题域”是“图书馆学研究对象”,“问题”是“图书馆学研究对象是什么”,“问题链”由关于“图书馆学研究对象”的各种学说、思想等创新知识元节点连接而成。其次,依据图书馆学研究对象各流派思想的关系,确定以“图书馆”说、“情

报交流”说、“信息资源”说、“知识”说为根节点,也就是原始创新节点。再者,按照思想观点的发生发展和源流演化关系逐步向后添加相关知识元节点。

通过梳理图书馆学研究对象领域在发展过程中的关键里程碑式的节点及它们之间的演变关系,得到整个领域的发展脉络。经专家判断、审定后,最终形成由4个主干谱系和16条枝干谱系组成的“图书馆学研究对象”领域创新知识元谱系图(见图5),图中每个节点上都标有依据新颖性测度函数测算得出的新颖度值。4个主干谱系分别是以“图书馆”说为根节点的谱系,以“情报交流说”为根节点的谱系,以“知识交流说”为根节点的谱系和以“知识说”为根节点的谱系。每个主干谱系下都有若干条问题链,例如位于图5上方的<“图书馆”说—“图书馆事业”说—“事业规律”说>问题链。

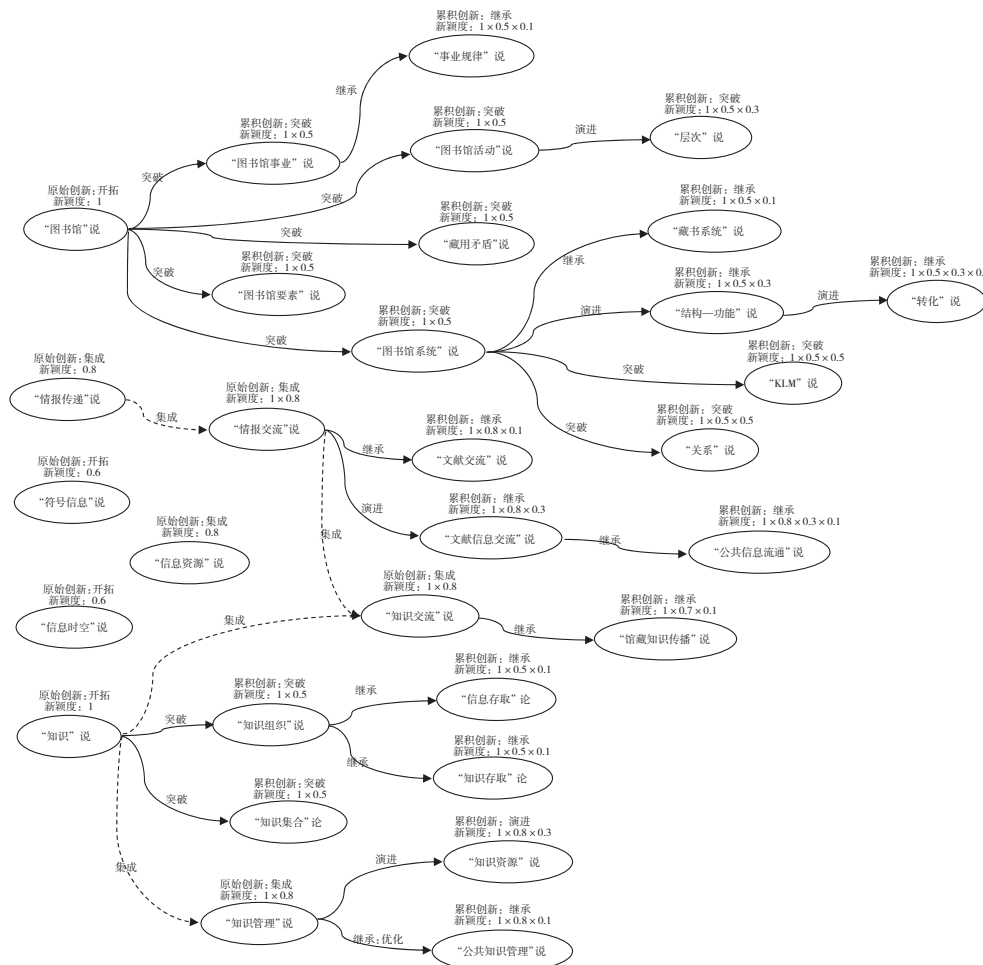


图5 “图书馆学研究对象”主题领域创新知识元谱系

Figure 5 Innovation knowledge element genealogy of the “research objects of library science”

创新知识元谱系从空间维度展示了“图书馆学研究对象”相关的各种思想、学说、观点等创新知识元之间的位置和关系,从时间维度展示了这些知

识之间的演进和发展。创新知识元谱系的知识元节点及其之间的关系可以进一步按照知识元标注模型进行语义三元组标注,以方便计算机进行识别、理解、

比较和计算。

#### 5.4 新颖性评判与测度

依据本研究提出的新颖性测度模型，以创新知识

元谱系为评价参照系，利用新颖性测度函数对“图书馆学研究对象”主题领域的 20 篇论文进行评判与计算。新颖性测度结果如表 6 所示：

表 6 创新知识元新颖性测度结果  
Table 6 Results of novelty measurement of innovative knowledge elements

序号	创新知识元语句	创新类型	新颖度	年份	来源文献题名
1	图书馆学就是研究收藏、提供一定范围书刊资料与人们对书刊资料一定需要之间的矛盾运动规律的科学	无	0	1980	图书馆及其研究对象
2	图书馆学的研究对象是图书馆工作	继承	0.002 9	1982	关于图书馆学的研究对象和内容
3	图书馆学的研究对象是图书馆系统的整体联系，亦即图书馆系统的结构与功能以及二者之间的辩证关系，或简称为结构—功能说	继承	0.083 3	1984	图书馆学研究对象及其理论体系逻辑起点
4	文献信息的开发与利用，便构成了图书馆学的研究对象	观点创新	0.600 0	1986	现代图书馆学的特点及其研究对象
5	(1) 图书馆学的研究对象就是图书馆 (2) 我们把这个对象析为质、态、能 3 个互相紧密联系在一个统一体中的对象	演化	0.021 4	1993	论图书馆学和文献信息学的研究对象
6	“读者(含用户)服务及其相关的图书馆活动”(下面简称“读”)应作为新图书馆学研究对象中的主要内容	继承	0.031 3	1997	图书馆学研究对象的新思考
7	图书馆学的研究对象是多层次综合整体的图书馆活动。它的层次体系是：①客观具体层次；②科学抽象层次；③思维具体层次	继承	0.007 9	1997	图书馆学研究对象新论
8	控制说的表述：研究社会信息及过程之控制	演进	0.218 2	1999	图书馆学研究对象述评—兼论控制说
9	文献群中知识单元的可获得性是图书馆学的研究对象	突破	0.333 3	1998	图书馆学的核心问题和研究对象新见
10	本文将图书馆定位于公共信息的特殊流通形态，明确图书馆学的研究对象是公共信息的流通，而基本概念就是公共信息	继承	0.173 9	1998	从图书馆的形态研究看图书馆学的研究对象、基本概念
11	文章提出了关于图书馆学研究对象的新观点，认为图书馆学的研究对象是如何把文献信息进行科学重组后及时提供给读者，并通过他们物化成社会主义物质文明，活化为社会主义精神文明，最后转化为生产力，简称为“转化说”	继承	0.011 9	2000	转化说：图书馆学研究对象的新探讨
12	图书馆学的研究对象是图书馆与读者在知识传播中的有机结合	继承	0.004 0	2003	关于图书馆学研究对象的再思考—兼与王子舟先生商榷
13	现代图书馆学研究对象是知识信息的组织、传播与利用	继承	0.004 2	2003	知识管理对现代图书馆学研究对象的创新
14	图书馆学的研究对象是：“图书馆为需求者提供文献信息保障的全部过程及其相互关系。”	继承	0.035 3	2006	究竟什么是图书馆学的研究对象
15	文献信息资源的整序与传递过程是图书馆的本质内核，是图书馆学的研究对象，图书馆学是研究文献信息资源整序与传递过程体系的科学	继承	0.004 0	2006	对图书馆学研究对象的再认识
16	图书馆学的研究对象是：主客观知识、知识集合、知识受众及三者之间的关系	无	0	2009	逼近图书馆学研究对象的“斯芬克司”之谜—兼与王子舟先生商榷
17	载体知识的收藏、传播和利用是图书馆学的研究对象	无	0	2011	载体知识的收藏、传播和利用是图书馆学的研究对象
18	关于图书馆学研究对象的表述是：研究社会大系统之下的图书馆子系统的“矛盾群”的产生、发展过程，即影响图书馆事业发展的众多矛盾的产生条件及其运动规律—合力	继承	0.031 3	2002	对图书馆学研究对象不同表述的思考—兼评“矛盾说”
19	“文献状态相互转化”作为图书馆学的研究对象，具有科学对象的一般性和特殊性，能够合理解释图书馆学基本原理和图书馆的常态化工作	继承	0.083 3	2016	从科学问题的分类看图书馆学的研究对象
20	人类社会的公共知识流现象是图书馆学的研究对象	演进	0.010 9	2016	基于公共知识流视角的图书馆学研究对象及其基本定律研究

以 4、14、16 号知识元为例进行说明。4 号知识元其语义三元组标注为“S: 图书馆学”“P: 研究对象”“O: 文献献信息的开发利用”。经比对，在创新知识元谱系中没有与之相匹配的节点，由此判断该知识元为原始创新。因为文献并未系统论述其理论，所以该知识元为观点型创新，新颖度取值为 0.6。测度完成后，该知识元节点作为新节点置于谱系中。其后继节点为 14 号知识元，语义标注为“S: 图书馆

学”“P: 研究对象”“O: 文献信息保障”。依据知识实体“文献信息保障”可以判断其属于文献信息类的问题链条，所以将其与图谱中“文献信息开发与利用”节点进行语义比较。在自定义的语义词典（即本体描述词汇表）中，“文献信息保障”与“文献信息开发与利用”同属文献信息活动类，因此二者是相似关系，语义关系判断为“继承”，依据新颖性测度函数计算其新颖度为 0.053 5。



20 号知识元认为“公共知识流”是图书馆学研究对象,其更接近公共知识管理说。在语义词典中,“公共知识流”属于“对象类”,而“公共知识管理”属于“活动类”,两者不属于同一类,语义存在较大差异,因此两者关系判断为“演进”,其新颖度取值为 0.010 9。16 号知识元由于其与王子舟的知识集合论的观点相重复,所以其新颖度为 0。

需要强调的是,具有新颖性的文献不一定就具有高水平和高影响力,但其在科研过程中仍具备一定的科研价值,从中可以发现最新研究进展,了解学科主题的发展趋势<sup>[40]</sup>。创新知识元的新颖值均是与文章发表之前领域内的知识元对比计算而得的,讨论的是文章发表时的新颖值,而非其他时间点上的新颖值。

## 6 学术论文新颖性测度模型的评价

### 6.1 测度模型的特点

#### 6.1.1 基于分类进行比较

科学评价的实质是分类基础上的同类比较<sup>[41]</sup>。比较方法是关于学术成果创新性评价研究普遍采用的技术路径,但是鲜有采用建立在分类基础上的比较方式。本研究提出的新颖性测度路径采用基于分类的比较方式,也就是先分类、再比较。对某项知识元进行评价,就是要基于它所在的学科、所属的主题,也就是将其放在这个学科的该主题学术成果集合中加以考察,从学科发展脉络和相关问题谱系分析、判断其潜在的研究价值。采用创新知识元谱系作为参照,绝不意味着要把知识元放在庞大的学术问题谱系中进行比较,而是要将其放在与其问题相关的问题谱系中进行比较。因此,分类就是将知识元放到与其相关的问题谱系中,是进行比较的前置环节。

#### 6.1.2 借助学术谱系进行参照

创新知识元谱系的构建是从内容层面实现以“创新”为价值取向的学术论文价值评价的关键。无论是基于形式的学术评价还是基于内容层面的学术评价,都没有从内容层面构建明确的、形式化的参照系。本研究明确提出以创新知识元谱系为参照系。创新知识元谱系是学术谱系的一种,是以认识对象和与之相关的科学问题为核心,以问题解决过程中所获得的创新知识元为节点,根据认识层次和认识深度,按时间顺序将相关的创新知识元进行纵向地历史性关联,同时,根据逻辑或语义关系,将彼此关联的创新知识元进行横向延伸,从而形成纵横交错、关联联结的知识网络。创新知识元谱系为评审专家判断学术论文价

值提供了统一的、可进行比较的参照系,弥补了评审专家认知结构和认知能力的天然局限性,有利于减少专家同行评议的主观性,促进基于内容的学术论文价值智能化评价的实现<sup>[42]</sup>。一旦完成了学科主题领域创新知识元谱系的构建,即可通过待评价的创新知识元与相应学科主题领域创新知识元谱系的比较来判断是否真正具有创新。根据与谱系中其他创新知识元之间的关系,观察其对谱系中哪一科学知识做出了发展和超越,对他人产生了哪些启迪,启迪产生了哪些新的科学发现,据此分析判断待评价知识元的价值。

#### 6.1.3 针对内容进行评价

当前基于内容的创新性评价研究采用关键词、创新语句来代替文本内容,并没有区分其语义功能。虽然关键词可以一定程度上可以反映学术文本的内容,但是其并不能完全表达文本的内容含义。在数据网络环境下,学术成果不再是一个简单的“文本”,而是一组逻辑相关的数字对象集合,每个数字对象都是独立的、完整的语义内容单元,即知识元。本研究以知识元为基础提出的学术论文新颖性测度思想属于基于知识元的科学计量方法,能够离析出科学文献中所蕴含的知识元成分,实现对某学科领域专业属性知识框架的揭示,进而有助于从实质性专业知识贡献的角度,对该领域的知识等进行直接的科学计量评估与评价<sup>[43]</sup>。

### 6.2 局限性

学科主题领域的创新知识元谱系从历史角度揭示了学科知识的生产、发展和演进脉络。但是,创新知识元谱系并不是一个普适的、标准的、统一的学术谱系,是根据评价对象的主题域和问题特征来动态构建的。特定问题的创新知识元谱系只能适用于特定问题的评价,不能用来作为其他问题的参照系。因此,并不存在一个统一的、标准的知识元谱系来作为普适的参照系。为了确保评价的针对性和准确性,每次评价都需要构建契合自身领域和问题特征的创新知识元谱系。

本研究提出的评价方法更适合于追求创新价值,且创新要素特征明显的理论、方法型论文的新颖性测度,不适用于技术应用型论文的评价。从基于创新知识元谱系的评价过程来看,创新知识元标注、创新知识元谱系构建等前期工作较多,增加了方法使用的复杂性。当前的自然语言处理、语义本体、知识图谱、人工智能等领域知识能够为知识元标注、新颖性测度计算提供技术支撑,但是创新知识元谱系是一个涉及

多层次、多层面的复杂问题,受到价值尺度、评价标准、价值取向等因素的影响<sup>[44]</sup>。因此,创新知识元谱系的构建仍然需要较多的人工干预,这在一定程度上影响了本方法大规模使用。

## 7 结语

学术论文作为科学活动的重要学术成果,其核心是创新性知识。一切科学活动的评价最终都要落到对学术论文的评价。对学术论文的创新性进行评价是长期困扰学术评价领域的重要问题,也是争议最多的问题。学术论文的定量评价方法力图弥补定性评价的不足,正在成为评价的主流方法。当前,如何利用定量方法分析、发现学术论文的新颖性,成为学术评价领域的热点和难点。

本研究遵循数据驱动的学术评价范式路径,基于分类比较的思想,提出了创新知识元谱系的概念模型;从语义内容层面提出了基于创新知识元谱系的学术论文新颖性测度模型;基于创新的累积性特征,构建了创新知识元的新颖性测度函数。通过图书馆学基础理论研究领域的“图书馆学研究对象”问题的实证研究,进一步验证了测度模型的适用性和可行性。实证结果表明,本研究所提出的新颖性测度模型借助创新知识元谱系作为参照系,采用分类比较思想进行评价,能够实现从语义内容层面测度学术论文的新颖性。但该模型同时存在谱系适用性的局限。

为进一步增强方法的适用性,下一步的研究工作包括:优秀论文语料库的构建、创新知识元模型的完善、创新知识元描述规则库的构建、创新知识元语料库的构建、创新知识元谱系的优化等。在学术论文新颖性测度的基础上,结合认知计算技术进一步构建能够智能识别学术研究创新点和重大技术突破点的相关模型,是未来基于人工智能的学术创新评价需要进一步探索和研究的空間。

### 参考文献:

- [1] 索传军,于果鑫. 学术论文研究亮点的语言学特征与分布规律研究[J]. 图书情报工作, 2020, 64(9): 104-113. (SUO C J, YU G X. Exploration of the research "highlights" in academic papers[J]. Library and information service, 2020, 64(9): 104-113.)
- [2] 索传军,盖双双. 单篇学术论文的评价本质、问题及新视角分析[J]. 情报杂志, 2018, 37(6): 102-107. (SUO C J, GAI S S. Analysis on the essence, problems and new perspective of individual academic paper evaluation[J]. Journal of intelligence, 2018, 37(6): 102-107.)
- [3] 索传军. 学术评价论[M]. 北京: 科学技术文献出版社, 2020: 198-207. (SUO C J. On academic evaluation [M]. Beijing: Scientific and Technical Documentation Press, 2020: 198-207.)
- [4] 曹树金,曹茹烨. 情报学论文创新性评价研究——LDA和SVM融合方法的应用[J]. 图书情报知识, 2022, 39(4): 56-67. (CAO S J, CAO R Y. Evaluation of paper innovativeness in information science by a method integrated of LDA and SVM[J]. Documentation, information & knowledge, 2022, 39(4): 56-67.)
- [5] 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, 30(1): 125-127. (SHEN Y. A keyword-based innovation evaluation method[J]. Information studies: theory & application, 2007, 30(1): 125-127.)
- [6] 张金柱,张晓林. 利用引用科学知识突变识别突破性创新[J]. 情报学报, 2014, 33(3): 259-266. (ZHANG J Z, ZHANG X L. Identification of radical innovation based on mutation of cited scientific knowledge[J]. Journal of the China Society for Scientific and Technical Information, 2014, 33(3): 259-266.)
- [7] 谢珍,马建霞,胡文静. 面向代表作评价的学术论文创新性测度方法[J]. 情报理论与实践, 2022, 45(7): 81-88. (The innovation measurement method for academic papers oriented to the evaluation of representative works[J]. Information studies: theory & application, 2022, 45(7): 81-88.)
- [8] 杨建林,钱玲飞. 基于关键词对逆文档频率的主题新颖度度量方法[J]. 情报理论与实践, 2013, 36(3): 99-102. (YANG J L, QIAN L F. Theme novelty measurement based on inverse document frequency of keyword pairs[J]. Information studies: theory & application, 2013, 36(3): 99-102.)
- [9] 迟培娟,丁洁兰,冷伏海. 突破性论文的三元计量特征及识别研究——以生物医学领域为例[J]. 情报学报, 2022, 41(7): 663-675. (CHI P J, DING J L, LENG F H. Characteristics of the technology impact of breakthrough papers in biology and medicine [J]. Journal of the China Society for Scientific and Technical Information, 2022, 41(7): 663-675.)
- [10] 钱佳佳,罗卓然,陆伟. 基于问题—方法组合的科技论文新颖性度量与创新类型识别[J]. 图书情报工作, 2021, 65(14): 82-89. (QIAN J J, LUO Z R, LU W. Novelty measurement and innovation type identification of scientific literature based on question-method combination[J]. Library and information service, 2021, 65(14): 82-89.)
- [11] 钱玲飞,杨建林,张莉. 基于关键词分析的学科创新力比较——以情报学图书馆学为例[J]. 情报理论与实践, 2011, 34(1): 117-120. (QIAN L F, YANG J L, ZHANG L. Disciplinary creativity comparison based on keyword analysis[J]. Information studies: theory & application, 2011, 34(1): 117-120.)
- [12] 侯剑华,王东毅. 基于SAO-ADV模型的学术论文创新性的测度方法研究[J]. 情报理论与实践, 2020, 43(11): 129-136. (HOU J H, WANG D Y. Research on innovation measurement method of academic papers based on SAO-ADV model[J]. Information studies: theory & application, 2020, 43(11): 129-136.)

- [13] 杨京, 王芳, 白如江. 一种基于研究主题对比的单篇学术论文创新力评价方法 [J]. 图书情报工作, 2018, 62(17): 75-83. (YANG J, WANG F, BAI R J. A method to evaluate academic papers' innovation based on the research theme comparing [J]. Library and information service, 2018, 62(17): 75-83.)
- [14] 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, 2(3): 22-29. (LU W F, TAN Z Y. Measuring novelty of scholarly articles [J]. Data analysis and knowledge discovery, 2018, 2(3): 22-29.)
- [15] 曹树金, 曹茹烨. 基于知识图谱的科技论文创新点动态识别研究 [J]. 现代情报, 2022, 42(12): 22-41, 82. (CAO S J, CAO R Y. Research on dynamic recognition of innovation points in scientific papers based on knowledge graph[J]. Journal of modern information, 2022, 42(12): 22-41, 82.)
- [16] 戎军涛, 吴鹏飞, 李华. 面向知识发现的学术文献内容情景模型构建研究 [J]. 情报理论与实践, 2019, 42(3): 153-159. (RONG J T, WU P F, LI H. Construction of academic literature content scenario model for knowledge discovery[J]. Information studies: theory & application, 2019, 42(3): 153-159.)
- [17] 索传军. 知识转移视角下的学术论文老化与创新研究 [J]. 图书情报工作, 2014, 58(5): 5-12. (SUO C J. Study on obsolescence and innovation of academic papers from perspective of knowledge transfer[J]. Library and information service, 2014, 58(5): 5-12.)
- [18] 索传军, 张璇, 李木子. 论评价参照系的内涵、作用与构建 —— 兼论学术谱系的功能与作用 [J]. 中国人民大学学报, 2022(4): 180-190. (SUO C J, ZHANG X, LI M Z. On the referential frame of academic evaluation and the function of academic genealogy[J]. Journal of Renmin university of China, 2022(4): 180-190.)
- [19] 刘益东. 问题谱系比较评议法 —— 同行评议方法新探 [J]. 自然辩证法研究, 1998, 14(10): 31-35. (LIU Y D. Comparative evaluation method of problem genealogy: a new exploration of peer evaluation method[J]. Studies in dialectics of nature, 1998, 14(10): 31-35.)
- [20] 汪信砚. 哲学社会科学创新论 [M]. 北京: 中国社会科学出版社, 2014: 61-64. (WANG X Y. Innovation in philosophy and social sciences [M]. Beijing: China Social Sciences Press, 2014: 61-64.)
- [21] CHRISTENSEN C M. The innovator's dilemma: when new technologies cause great firms to fail[M]. Boston: Harvard Business School Press, 1997.
- [22] 索传军, 盖双双, 周志超. 认知计算 —— 单篇学术论文评价的新视角 [J]. 中国图书馆学报, 2018, 44(1): 50-61. (SUO C J, GAI S S, ZHOU Z C. Cognitive computing: a new perspective for evaluating the individual academic paper[J]. Journal of library science in China, 2018, 44(1): 50-61.)
- [23] 索传军, 王春明. 基于引文的知识转移类型与方式分析 [J]. 情报理论与实践, 2023, 46(4): 114-121. (SUO C J, WANG C M. Analysis of knowledge transfer types and modes based on citations[J]. Information studies: theory & application, 2023, 46(4): 114-121.)
- [24] MILLER G A. WordNet: a lexical database for English[J]. Communications of the ACM, 1995, 38(11): 39-41.
- [25] 梅家驹, 竺一鸣, 高蕴琦, 等. 同义词词林 [M]. 上海: 上海辞书出版社, 1983. (MEI J J, ZHU Y M, GAO Y Q, et al. Chinese Thesaurus [M]. Shanghai: Shanghai Lexicographical Publishing House, 1983.)
- [26] 董振东. 语义关系的表达和知识系统的建造 [J]. 语言文字应用, 1998(3): 76-82. (DONG Z D. The expression of semantic relationships and the construction of knowledge systems[J]. Applied linguistics, 1998(3): 76-82.)
- [27] 索传军, 盖双双. 知识元的内涵、结构与描述模型研究 [J]. 中国图书馆学报, 2018, 44(4): 54-72. (SUO C J, GAI S S. The connotation, structure and description model of knowledge Unit[J]. Journal of library science in China, 2018, 44(4): 54-72.)
- [28] 王晓光, 孙梦琳, 宋宁远. 科学论文功能单元本体设计与设计应用实验 [J]. 中国图书馆学报, 2018, 44(4): 73-88. (WANG X G, SUN M L, SONG N N. Design and application of scientific paper functional units ontology[J]. Journal of library science in China, 2018, 44(4): 73-88.)
- [29] 孙建军, 裴雷, 蒋婷. 面向学科领域的学术文献语义标注框架研究 [J]. 情报学报, 2018, 37(11): 1077-1086. (SUN J J, PEI L, JIANG T. Research on semantic annotation in academic literature[J]. Journal of the China Society for Scientific and Technical Information, 2018, 37(11): 1077-1086.)
- [30] 李春秋, 徐增旭林, 宋宁远, 等. 基于纳米出版物的中文学位论文语义组织研究 [J]. 中国图书馆学报, 2021, 47(5): 97-115. (LI C Q, XU Z X L, SONG N Y, WANG X G. Nanopublication-based semantic organization of Chinese dissertation[J]. Journal of library science in China, 2021, 47(5): 97-115.)
- [31] 化柏林, 刘一宁, 郑彦宁. 针对学术定义的抽取规则构建方法研究 [J]. 情报理论与实践, 2011, 34(12): 5-9. (HUA B L, LIU Y N, ZHENG Y N. Studies on methods of formulating rules for academic definition extraction[J]. Information studies: theory & application, 2011, 34(12): 5-9.)
- [32] 索传军, 赖海媚. 学术论文问题知识元的类型与描述规则 [J]. 中国图书馆学报, 2021, 47(2): 95-109. (SUO C J, LAI H M. Types and description rules of problem knowledge units in academic papers[J]. Journal of library science in China, 2021, 47(2): 95-109.)
- [33] 李贺, 杜杏叶. 基于知识元的学术论文内容创新性智能化评价研究 [J]. 图书情报工作, 2020, 64(1): 93-104. (LI H, DU X Y. Research on intelligent evaluation for the content innovation of academic papers [J]. Library and information service, 2020, 64(1): 93-104.)
- [34] 赵蓉英, 张心源. 基于知识元抽取的中文智库成果描述规则研究 [J]. 图书与情报, 2017(1): 119-127. (ZHAO R Y, ZHANG X Y. On descriptive rules of Chinese think-tank products based



- on knowledge element extraction[J]. Library & information, 2017(1): 119-127.)
- [35] 牛丽慧, 欧石燕. 纳米出版物及其应用研究进展 [J]. 图书情报工作, 2018, 62(7): 125-133. (NIU L H, OU S Y. Research advances of nanopublication and its applications[J]. Library and information service, 2018, 62(7): 125-133.)
- [36] 中国图书馆学会. 中国图书馆学学科史 [M]. 北京: 中国科学技术出版社, 2014: 184-185. (Library Society of China. History of library science in China[M]. Beijing: China Science and Technology Press, 2014: 184-185.)
- [37] 马恒通. 新中国图书馆学研究对象争鸣 50 年 (1949—1999) [J]. 图书馆, 2000(1): 22-27, 37. (MA H T. On the contention of study object of library science in China in the past 60 years. [J]. Library, 2000(1): 22-27, 37.)
- [38] 王子舟. 图书馆学研究对象的认知过程及范式特征 [J]. 江西图书馆学刊, 2002(3): 4-8. (WANG Z Z. Cognitive process and paradigm characteristics of the research object in library science[J]. Journal of the library science in Jiangxi, 2002(3): 4-8.)
- [39] 徐引麓, 霍国庆. 图书馆学研究对象的认知过程 —— 兼论资源说 [J]. 中国图书馆学报, 1998(3): 3-13. (XU Y C, HUO G Q. Cognition process of research objects in library science: on the resource[J]. Journal of library science in China, 1998(3): 3-13.)
- [40] 许丹, 徐爽, 陈斯斯, 等. 基于自然语言词对法的文献主题新颖性探测研究 [J]. 图书情报工作, 2018, 62(8): 130-138. (XU D, XU S, CHEN S S. Document theme novelty detection research based on natural language pairs[J]. Library and information service, 2018, 62(8): 130-138.)

- [41] 叶继元. 人文社会科学评价体系探讨 [J]. 南京大学学报 (哲学·人文科学·社会科学), 2010(1): 97-110, 160. (YE J Y. Approaching evaluation system in humanities and social sciences[J]. Journal of Nanjing University (philosophy humanities and social sciences), 2010(1): 97-110, 160.)
- [42] 盖双双. 学术论文的价值评价研究 [D]. 北京: 中国人民大学, 2020: 143-144. (GAI S S. Research on the value evaluation of academic papers[D]. Beijing: Renmin University of China, 2020: 143-144.)
- [43] 孙震, 冷伏海. 基于知识元的新型科学计量范式探析 [J]. 情报学报, 2017, 36(6): 555-564. (SUN Z, LENG F H. Research on the new paradigm of scientometrics based on knowledge element[J]. Journal of the China Society for Scientific and Technical Information, 2017, 36(6): 555-564.)
- [44] 索传军. 论学术评价的价值尺度 —— 兼谈“唯论文”问题的根源 [J]. 中国社会科学评价, 2021(1): 122-131, 160. (SUO C J. The value yardstick for academic evaluation: the root of the problem of “only publications count”[J]. China social science review, 2021(1): 122-131, 160.)

#### 作者贡献说明:

戎军涛: 提出研究思路, 设计研究框架, 调研与整理文献、数据, 起草论文;

索传军: 提出研究选题, 指导论文撰写与修改论文;

周彦廷: 构建测度函数, 整理、分析数据;

李艺亭: 文献调研, 整理、分析数据。

### Novelty Measurement of Academic Literature Based on the Innovative Knowledge Elements Genealogy

Rong Juntao Suo Chuanjun Zhou Yanting Li Yiting

School of Information Resource Management, Renmin University of China, Beijing 100872

**Abstract: [Purpose/Significance]** The paper evaluation based on “frequency” and “citation” essentially evaluates the “influence” of academic papers, which is not completely equivalent to academic innovation. Measuring the novelty from the content and semantic levels is a new approach for evaluating the innovation of academic papers. **[Method/Process]** This study followed the data-driven academic evaluation paradigm. Firstly, it constructed a conceptual model of the innovative knowledge elements genealogy on classification and comparison. Secondly, it proposed a model for measuring the novelty of academic papers on the innovative knowledge elements genealogy. Finally, it empirically verified the applicability of the model. **[Result/Conclusion]** The empirical results indicate that the novelty measurement model is able to measure the novelty of academic papers at the level of semantic content on the innovative knowledge elements genealogy and the idea of classification and comparison. However, the model also suffers from the limitation of genealogical applicability.

**Keywords:** the genealogy of innovative knowledge elements academic evaluation innovative evaluation novelty measurement