基于知识“非常规重组”的学术论文新颖性测度指标研究

梁国强、孙健、林歌歌、张硕

（北京工业大学经济与管理学院，北京，100124）

摘　要科学发展离不开对现有知识的重新组合，而新颖的研究通常依赖于对现有知识的“非常规重组”。本研究提出知识离心度指标，通过测度待评价成果对现有知识体系的偏离程度，对待评价成果的新颖性进行及时评价。本研究以2005年、2010年、2015年、2020年和2025年发表在《科学》和《自然》上的研究型论文、Top1%高被引论文和零被引论文作为样本数据，计算上述论文的知识离心度，并分析其潜在影响因素。结果显示，团队规模对论文新颖性有负向影响（β =-0.005，p<0.001），即团队规模越大，越不利于提升成果的新颖性；参考文献数量与论文新颖性呈正相关（β =0.027，p<0.001），即论文的参考文献数量越多，其新颖性在一定程度上越高。本研究提出的知识离心度指标具有较强的“及时性”和“可操作性”，能够在论文发表时及时评价其新颖性。

关键词　参考文献;知识离心度;非常规重组;团队规模

基于知识非典型重组的学术论文新颖性测度指标研究

梁国强、孙健、林格格、张硕（北京工业大学经济管理学院，中国北京100124）

摘要：科学的发展本质上依赖于既有知识的重组，而创新研究通常依赖于对既有知识库进行“非常规重组”。本研究提出知识偏心度，用于量化研究成果偏离既有知识体系的程度，从而及时评估其新颖性。实证分析选取了2005年、2010年、2015年、2020年和2025年发表于《科学》和《自然》的研究论文、排名前1%的高被引论文及零被引论文作为样本数据。我们计算了这些论文的知识离心度得分，并考察了其潜在影响因素。结果表明，团队规模对论文新颖性具有显著负向影响（β =-0.005，p<0.001），这意味着团队规模越大，越不利于提升研究的新颖性

产出。相反，参考文献数量与论文新颖性呈显著正相关（ β

收稿日期：XXXX-XX-XX；**修回日期**：XXXX-XX-XX基金项目：北京市自然科学基金面上项目“协同进化视角下研究前沿的涌现机制及识别方法研究”(9232002);国家自然科学基金“颠 覆性低碳技术识别及创新路径动态选择研究”(72304023)。作者简介：梁国强，男，1988 年生，博士，副教授，硕士生导师，主要研究领域为知识图谱、复杂网络、科学计量学；孙健，男，1979年生，博士，讲师，主要研究方向为大数据管理与应用；林歌歌，女，1992 年生，博士，主要研究方向为引文分析、文本挖掘、知识图谱；张硕，通讯作者，女，1992 年生，博士，副教授，主要研究方向为数据挖掘与技术预见、科技情报与科技评价，Email:zhangshuozs@bjut.edu.cn。

=0.027，p<0.001），这意味着参考文献数量越多，论文新颖性就会适度提升。所提出的知识偏心度具有较强的时效性和可操作性，可在论文发表后立即评估其新颖性。

关键词：参考文献；知识偏心度；非典型重组；团队规模

1 引言

对研究成果新颖性进行评价，筛选潜在变革性研究，加速前瞻培育和布局，提升基础研究的原始创新能力，是科研管理部门长久以来的夙愿和挑战。科学发展离不开对已有知识的重新组合，而新颖的研究通常依赖对现有知识的“非常规重组”[1-4]，涉及跨越传统学科边界，将原本互不关联的知识元、方法、理论或数据集进行创造性整合[5-8]。例如，Uzzi 等认为新颖性是对已有知识的非常规重组[1]，Wang 等认为新颖性是已有知识“意料之外”的组合[9]，Foster 等认为新颖性包括涉足人类尚未意识到的知识、首次将不同学科知识进行组合、将学科内尚未联系的知识进行组合三种类型[10]。

从“非常规重组”视角出发，测度学术成果新颖性的方法通常基于期刊、关键词、引文等。例如，Uzzi 的 z-score 指标及其衍生指标[1, 2, 9-11]、结构洞[12, 13]、颠覆性指数[14-16]、新词占比等[7]。Boudreau 等通过分析项目申报书中出现但未收录于 PubMed 数据库 MeSH 词中的词所占比例，来测度项目申报书的新颖性；他们发现，评审专家对新颖性较高的申报书评价较为负面[17]。Wu 等基于引文网络，提出了衡量现有科研成果对已有成果替代程度的指标（即颠覆性指数），认为当后续研究更多引用待评价成果、较少引用该成果的参考文献时，待评价成果的新颖性较强[14]。近两年来，随着深度学习技术的进步，基于语言大模型测度研究成果新颖性的方法也应运而生[18-20]。

尽管上述方法或指标在一定程度上反映了研究成果的新颖性，但在“及时性”和“可操作性”方面仍存在显著局限：一是及时评价面临困境。基于引用数据的评价指标需要依赖时间积累，难以即时判断新发表研究成果的新颖性，例如颠覆性指数及其衍生指标[21]。而且，一旦引入待评价成果的施引文献信息，该指标就容易被操纵。二是可操作性不强。被引次数、期刊影响因子、期刊分区等指标尽管存在较多弊端，但长期用于科研评价的原因之一是操作简单。Uzzi 的 z 分数指标及其衍生指标、结构洞、新词占比等指标虽然及时性较强，但由于计算复杂度较高，很难普及。

本研究基于论文的参考文献信息，从已有知识“非常规重组”视角出发，提出了一种衡量论文新颖性的指标：知识离心度（Knowledge Eccentricity，KE）。该指标可在论文发表时即对其新颖性进行测度。为提高 KE 指标的可操作性，本研究开发了基于 DeepSeek 的 KE 指标计算智能体和批量计算平台。科研管理部门只需提供待评价成果的 DOI，即可迅速计算该成果的新颖性，且计算过程直观、透明。

2 新颖性指标设计与研究思路

2.1 指标设计

本研究认为，待评价论文所引用的参考文献之间集聚程度较强，表明待评价论文的知识离心度较弱，新颖性不强。也就是说，如果一篇学术论文引用的参考文献彼此密集交织，那么该篇学术论文中的知识并未偏离原有的知识基础，而是依赖已有知识进行常规组合。反之，则属于新颖性较强的成果，具有潜在的颠覆性。为测度学术论文 v 的知识离心度（Knowledge Eccentricity，KE_v），我们提出以下公式：

$$KE_v = 1 - \sqrt[3]{\frac{2L}{N*(N-1)}}$$

其中，L 表示论文 v 的参考文献之间存在引用关系的数量，N 表示论文 v 的参考文献数量。图 1 展示了三种不同类型的知识组合模式，其中橙色节点 v 表示待评价论文，蓝色节点表示待评价论文 v 的参考文献。具体计算步骤为：第一步：获取待评价论文的参考文献集合 R={Ref_1, Ref_2,...,Ref_n}。第二步：获取集合 R 中每篇参考文献的参考文献集合 RR={$RRef_1$, $RRef_2$,...,$RRef_n$}。第三步：计算集合 R 中的参考文献配对是否出现在集合 RR 中（记作 l_v），出现则记为 1，否则记为 0。第四步：将全部 l_v 累加，L=$\sum_n l_v$，最终根据公式计算 KE_v，概念模型见图 1*。

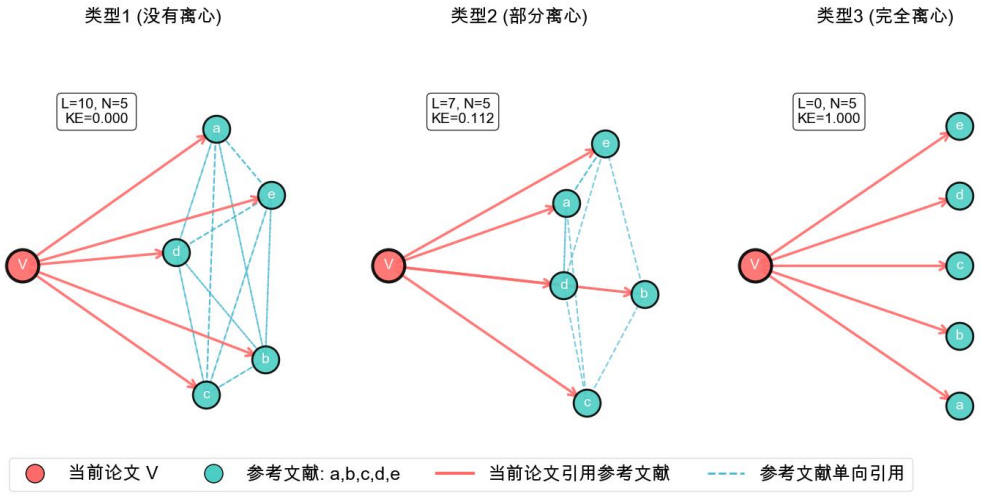

图 1 三种不同类型的知识组合情况。Type I 中，全部参考文献均存在引用行为(KE_v=0)；Type II 中，部分参考文献存在引用行为(KE_v=0.112)；Type III 中，参考文献之间不存在引用行为(KE_v=1)。

1.2 研究思路

《科学》和《自然》两本期刊在学术界享有较高声誉，成果发表于上述两本期刊是一种荣誉，有助于带来更多关注和引文影响力。根据引文规范理论，高被引论文是具有较强引文影响力的论文，而零被引论文则是长期以来未被科学界认可的成果。基于此，本研究收集了2005年、2010年、2015年、2020年和2025年发表在《科学》和《自然》上的研究型论文、Top 1%高被引论文和零被引论文作为样本数据，计算上述论文的KE值，并对其潜在影响因素进行分析，研究思路见图2。

* KE 指标在线计算：https://www.coze.cn/space/7446604282029588489/bot/7581092641727758345

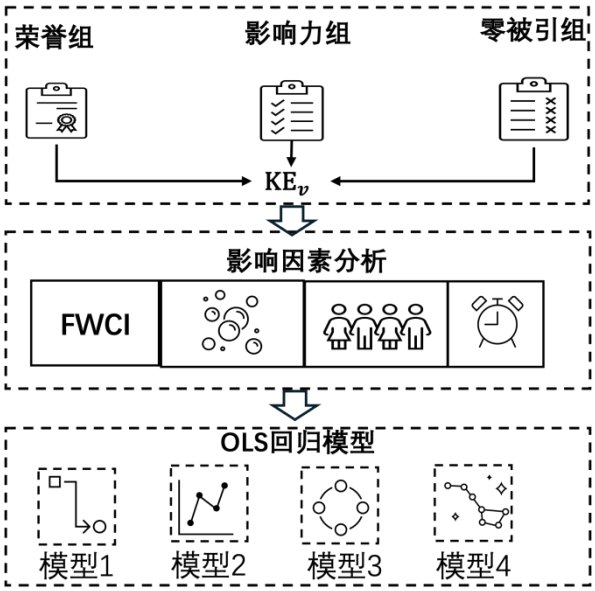

图 2 研究思路

2 研究结果

本研究的数据来源于OpenAlex数据库，该数据库涵盖广泛的学术出版物、作者、机构和引用关系。在删除参考文献为0的论文后，三组论文合计20184篇。下面分别从发表时间、研究领域、团队规模等方面分析其对新颖性指标的影响。

2.1 时间与分组对论文新颖性的影响

结果显示，不同时间发表的论文，其新颖性指标整体上介于 0.530-0.602 之间，未产生较大波动，但存在轻微时间变化趋势，即发表越早的论文整体的新颖性偏低。组间比较结果显示，荣誉组和影响力组($t = -22.724$, $p<0.001$)、荣誉组和零被引组($t = -20.611$, $p<0.001$)、影响力组和零被引组($t = -8.121$, $p<0.001$)论文的新颖性存在统计学差异，整体呈现出“零被引组>影响力组>荣誉组”的趋势。

KE 值分布的多样性可以在一定程度上反映不同组内论文新颖性程度是否一致。由于香农指数是衡量系统多样性的常用指标[22]，因此，本研究采用香农指数对各组论文 KE 值分布的多样性进行分析，发现高影响力组的多样性最高（香农指数为 1.278），其次是零被引组（香农指数为 1.214）和荣誉组（香农指数为 1.099），与采用辛普森指数和基尼-辛普森指数计算结果的趋势一致（见附录 A），提示高影响力论文的新颖性变动较大，其次是零被引论文，而发表在《科学》或《自然》上的论文新颖性变动较小。

以三组论文新颖性的均值 0.581 为阈值，荣誉组中 KE≥0.581 的论文占比最低（35.2%），其次是影响力组（51.5%）和零被引组（56.8%）。此外，荣誉组和影响力组论文的新颖性分布与零被引组明显不同。零被引组论文的新颖性呈现明显的极化趋势，即 KE=0和 1 的论文数量多于另外两组，特别是 KE=1的论文数量，提示新颖性过高的研究通常不会在短时间内得到认可，见图 3。

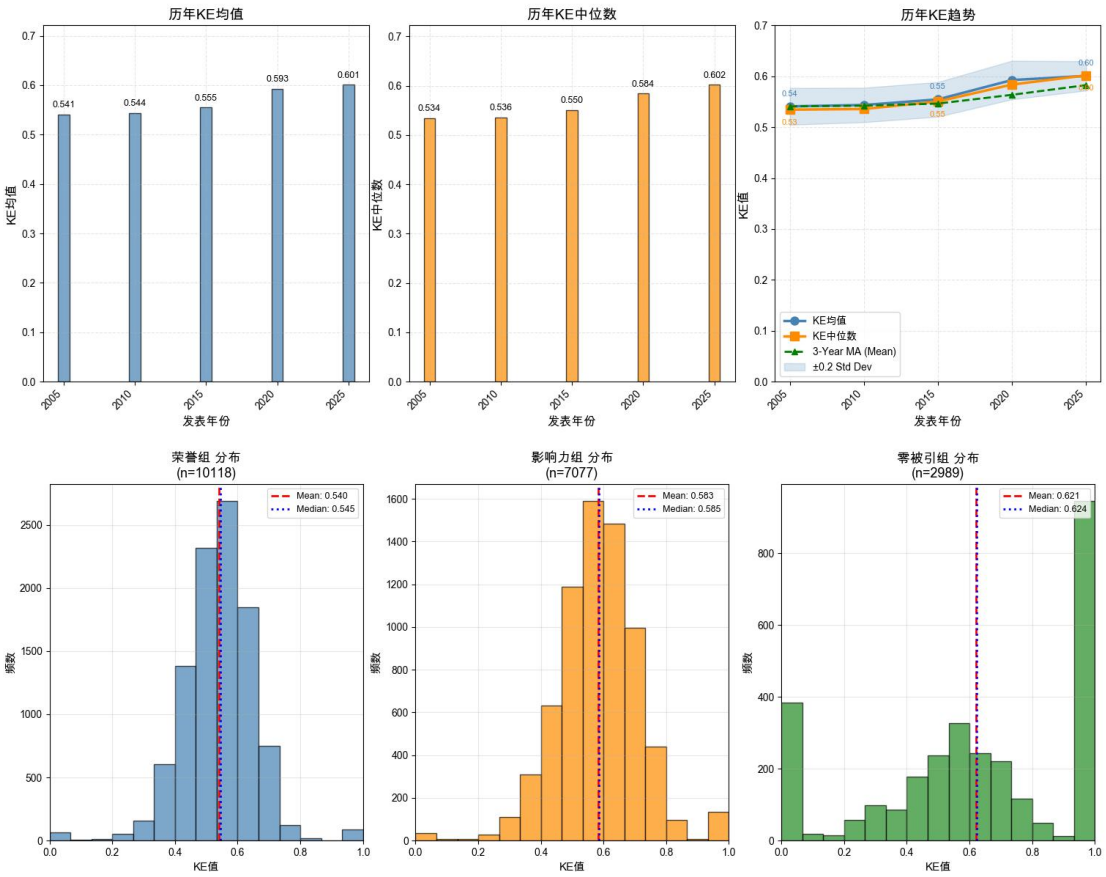

图 3 KE 值在不同时间和组间的分布图

2.2 研究领域对论文新颖性的影响

对研究领域大类进行统计后发现，数据源于社会科学（Social Sciences）、健康科学（Health Sciences）、物理科学（Physical Sciences）和生命科学（Life Sciences）四大领域。荣誉组、影响力组和零被引组在各领域的构成等详细信息见附录 B。

进行方差齐性检验后，发现不同研究领域具有可比性，因此采用单因素方差分析进行多组比较。结果显示，不同研究领域论文的新颖性存在显著的统计学差异（$p<0.001$），见表 1。为进一步分析各组间差异，采用 TukeyHSD 检验进行两两比较。结果显示，社会科学领域论文的新颖性显著高于其他领域（$p<0.001$）；健康科学和物理科学两个领域论文的新颖性差异没有显著差异（$p>0.05$）；生命科学较健康科学、物理科学两个领域论文的新颖性低（$p<0.001$），见表 2。

表 1 不同学科论文新颖性的基本情况

研究领域	样本量	均值	标准差	中位数	下四分位数	上四分位数
物理科学	8509	0.561	0.157	0.558	0.482	0.637
生命科学	5101	0.538	0.118	0.544	0.471	0.610
健康科学	3511	0.555	0.161	0.557	0.478	0.638
社会科学	2903	0.649	0.264	0.646	0.545	0.793

表 2 不同领域论文新颖性的 Tukey HSD 检验

组 1	组 2	均值差*	校正后的 p 值	下限	上限	拒绝原假设
健康科学	生命科学	-0.018	0.000	-0.027	-0.008	TRUE

健康科学 物理科学	0.006	0.322	-0.003	0.015	FALSE
健康科学 社会科学	0.094	0.000	0.083	0.105	TRUE
生命科学 物理科学	0.024	0.000	0.016	0.031	TRUE
生命科学 社会科学	0.112	0.000	0.102	0.122	TRUE
物理科学 社会科学	0.088	0.000	0.079	0.098	TRUE

*均值差为组 2 的均值减去组 1 的均值。

2.3 团队规模、参考文献数量和 FWCI对论文新颖性的影响

科研团队是为实现特定科学目标，由具备不同知识、技能和专长的科研人员组成的协作群体。随着科学活动专业程度日趋提升以及跨学科解决问题的需求增加，大科研团队比例逐渐增多，发表在高期刊分区上的论文尤为明显[23, 24]。领域权重引用影响力（FWCI）是文献实际被引用次数与该领域预期引用总量平均值的比值，旨在对不同领域间的论文影响力进行横向比较[25]。当 FWCI>1时，表示论文被引次数超过全球预期引用量的平均水平。

分析团队规模、参考文献数量和 FWCI 对论文新颖性的影响，并按照四分位数法，进一步将团队规模分为小团队(Q_1,P_{25})、中小团队(Q_2,P_{50})、中大团队(Q_3,P_{75})和大团队(Q_4)四组，将参考文献数量划分为少量(Q_1,P_{25})、较少(Q_2,P_{50})、较多(Q_3,P_{75})和大量参考文献(Q_4)四组。由于 FWCI 存在等于 0 的情况，为确保组内样本量均衡，将 FWCI 进一步分为零、低、中低、中高和高领域权重引用影响力组。剔除作者缺失、FWCI 缺失等字段后，最终分析了 20158 篇论文的团队规模、FWCI 与论文新颖性之间的关系，以及 20149 篇论文的参考文献数量与论文新颖性之间的关系。

结果显示，团队规模与论文新颖性呈弱负相关（r=-0.023，p<0.001），参考文献数量与论文新颖性呈弱正相关（r=0.160，p<0.001），FWCI与论文新颖性呈弱正相关（r=0.051，p<0.001），且各变量内的分组间存在显著的统计学差异（p<0.001）。上述结果提示：小团队和大团队、少量参考文献和大量参考文献、零FWCI和高领域权重引用影响力的论文通常具有较高的新颖性，见图4。

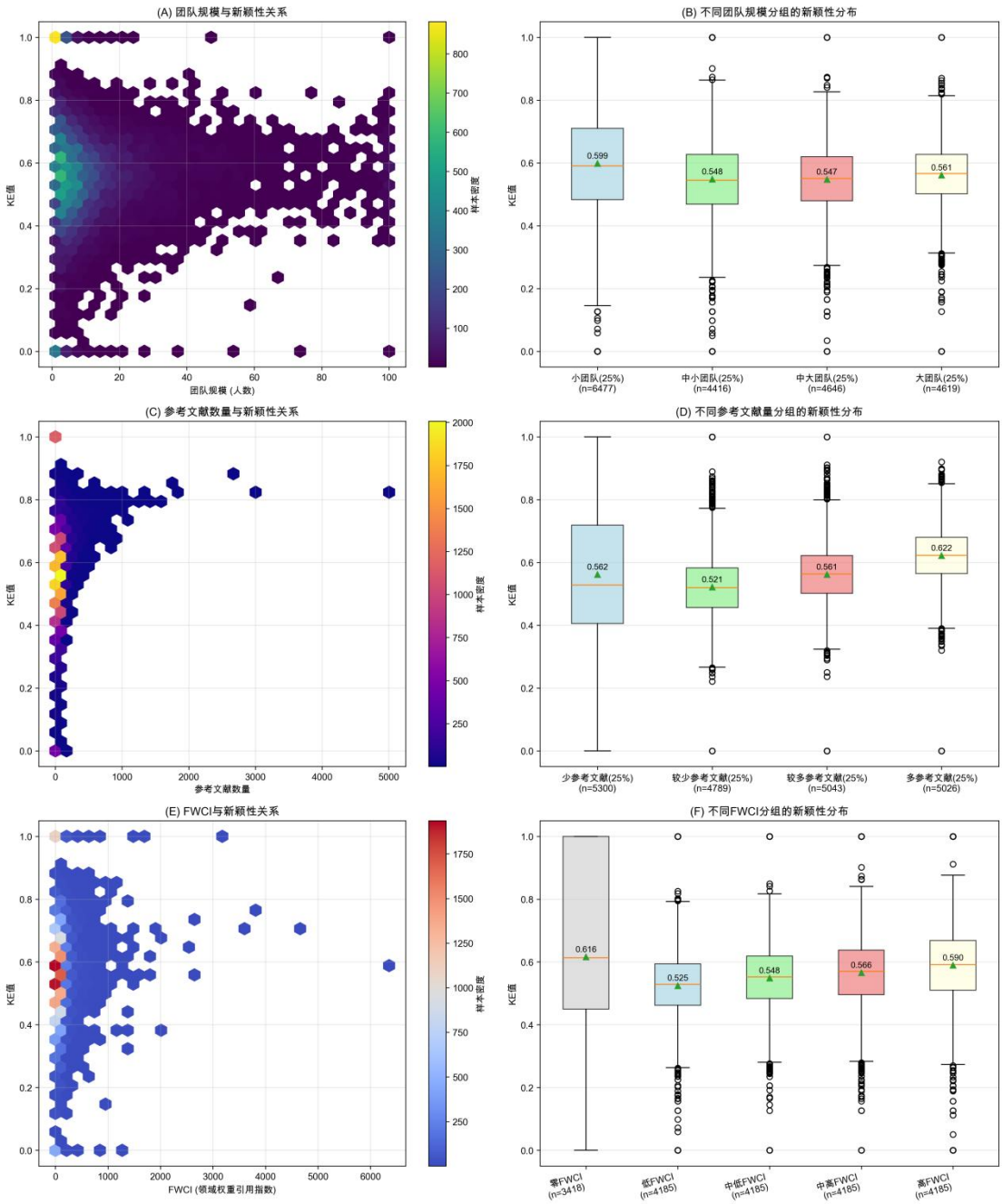

图4 团队规模、参考文献数量和 FWCI 对论文新颖性的影响

2.4 论文新颖性影响因素的 OLS 回归分析

将论文被引次数、参考文献数量、FWCI、团队规模作为自变量，论文新颖性作为因变量，发表年份和研究领域作为控制变量，进行标准化OLS回归分析。对自变量进行标准化和一致性处理，并采用方差膨胀因子（VIF）进行多重共线性检验，结果显示VIF介于1-1.5之间，提示上述自变量的共线性较轻。回归模型分析结果显示，被引次数对KE具有微弱但显著的正向影响（$\beta=0.004$，$p<0.05$），提示学术认可度可能促进知识创新（模型1）。模型2引入参考文献数量后，其表现出极强的正向效应（$\beta=0.027$，$p<0.001$），而此前被引次数的影响不再显著，提示参考文献数量对论文新颖性的解释力较强。模型3加入FWCI值，但该变量未呈现显著影响，提示学科间相对影响力的作用有限。模型4进一步纳入团队规模，发现其具有显著的负向效应（$\beta=-0.005$，$p<0.001$），提示团队规模越大，越不利于成果的新颖性，而参考文献数量的增加对论文新颖性具

具有正向作用，见表3。

从模型整体拟合度看，调整后的 R^2 从 0.061 逐步提升至 0.086，提示自变量对论文新颖性具有一定解释力。其中参考文献数量的稳健显著性尤为突出，强调了知识基础对创新的关键作用，为后续创新研究提供了理论依据。上述结果与本研究随机森林特征重要性排序结果一致（见附录 C）。

表3 论文新颖性影响因素的 OLS回归分析

	模型 1	模型 2	模型 3	模型 4
被引次数	0.004** (0.002)	0.001 (0.321)	0.002 (0.195)	0.002 (0.179)
参考文献数量		0.027*** (0.000)	0.028*** (0.000)	0.028*** (0.000)
领域加权引文影响力值			-0.001 (0.382)	-0.001 (0.576)
团队规模				-0.005*** (0.000)
N	20184	20184	20184	20184
R^2	0.062	0.086	0.086	0.087
调整R^2	0.061	0.086	0.086	0.086

*控制变量为研究领域、发表年份,(*ɲp< 0.05, **ɲp< 0.01, ***ɲp< 0.001)。

3. 结论与讨论

传统科研评价过度依赖论文数量、被引次数、影响因子等指标，在衡量学术成果新颖性方面存在不足。科学创新的核心机制之一在于对现有知识进行新颖且非常规的组合。尽管学术界普遍存在依赖成熟知识的路径倾向，但突破性创新往往源于对不熟悉主题的探索及跨领域的知识整合[10, 21]。然而，高度新颖性的研究伴随较强的学科交叉性和知识结构“非常规”等特征，往往挑战现有范式，因而难以在传统评价体系中获得及时认可。例如，辛顿（Geoffrey E. Hinton）关于神经网络的早期研究、麦克林托克（Barbara McClintock）的“跳跃基因”假说，以及安布罗斯（Victor Ambros）对微小核糖核酸（miRNA）及基因调控机制的探索等。

从已有知识“非常规重组”视角出发，本研究认为，当论文引用的参考文献紧密聚集在一起时，其知识基础往往较为成熟，因此该论文整合跨学科知识的能力有所欠缺。基于此，本文提出了用于测度知识非常规重组的知识离心度指标，并分析了其影响因素。研究发现，不考虑学科领域和发表年份时，零被引论文的整体新颖性高于引文影响力较高的论文；而发表于《科学》和《自然》的论文，其新颖性普遍不如零被引论文和高影响力论文。从各组论文新颖性指标的多样性分布来看，零被引组论文的新颖性波动相对较大（香农指数为1.214）。这既与零被引组中社会科学领域论文的占比（42.8%）高于荣誉组（3.9%）和影响力组（17.8%）有关，也与各组新颖性等于1的论文数量有关：零被引组中此类论文的数量远高于其他两组。这在一定程度上印证了高度新颖的研究往往无法及时获得认可，呈现出“睡美人”特征。此外，高影响力论文的新颖性变动最大，而发表于《科学》或《自然》的论文新颖性变动最小。这表明研究成果的引文影响力与新颖性并不完全一致；发表于《科学》或《自然》的论文，因其极为严格的同行评审和发表压力，往往能在知识重组过程中保持必要的张力[26]。

从研究领域对论文新颖性的影响来看，不同研究领域论文的新颖性不存在可比性。例如，社会科学领域论文的新颖性普遍高于自然科学领域，这提示在采用 KE 值进行论文新颖性测度和跨领域横向比较时，要控制研究领域的影响。另外，研究认为小团队的研究成果通常具有最强的新颖性，这与 Wu 等[14]人的研究结论基本一致。FWCI 与论文新颖性呈弱正相关，反映出被引次数与论文新颖性指标之间存在一定的正相关关系。

控制研究领域和发表年份后，回归结果发现，团队规模和参考文献数量是影响论文新颖性的主要因素。团队规模对论文新颖性的影响为负向，即团队规模越大，越不利于提升论文的新颖性；而参考文献数量与论文

新颖性呈正相关，即论文的参考文献数量越多，其新颖性也会在一定程度上提升，这表明新颖性研究大多建立在大量已有知识基础之上。

综上所述，本研究提出了一种论文新颖性测度的新指标，并分析了其影响因素。研究认为，该指标能够及时、便捷地评价论文新颖性，为识别零被引论文的新颖性提供了新思路。研究成果新颖性受团队规模和已有知识基础的影响，与引文及其相关衍生指标无关，如被引次数、FWCI等。值得注意的是，知识“非常规重组”只是知识创新的一种典型方式，既不是知识创新的充分条件，也不是必要条件；也就是说，并非所有知识的“非常规重组”都能产生新颖的研究，而新颖的研究也不全是知识“非常规重组”的结果。

参 考 文 献

[1] UZZI B, MUKHERJEE S, STRINGER M, 等. 非典型组合与科学影响力 [ ] 2013, J . 《科学》, 342 6157 468 72 ( ): - . [2] LEE Y-N, WALSH J P, WANG J. 科学团队中的创造力：剖析新颖性与影响力 [J]. 《研究政策》，2015年，44(3)：684-97. [3] SHI F, EVANS J. 研究内容与情境的意外组合与影响力相关，并由来自不同学科的科学界外人士促成 [J]. 《自然·通讯》，2023年，14(1)：1641. [4] SCHILLEBEECKX S J D, LIN Y, GEORGE G, 等. 知识重组与发明者网络：嵌入性对知识再利用和影响力的非对称效应 [J]. 《管理学杂志》，2020年，47(4)：838-66.[5] SHIBAYAMA S, WU Z, YIN D, 等. 新颖性指标的最新进展 [J]. SSRN，2025年。[6] ZHAO Y, ZHANG C. 学术论文新颖性测度综述 [J]. 《科学计量学》，2025年，130(2)：727-53.[7] BORNMANN L, TEKLES A, ZHANG H H, 等. 分析被引参考文献的非常规组合时，我们测量的是新颖性吗？基于F1000Prime数据对文献计量新颖性指标的验证研究 [J]. 《信息计量学杂志》，2019年，13(4). [8] WAGNER C S, WHETSELL T A, MUKHERJEE S. 国际科研合作：知识重组中的新颖性、常规性与非典型性 [J]. 《研究政策》，2019年，48(5)：1260-70. [9] WANG J, VEUGELERS R, STEPHAN P. 科学中的新颖性偏见：文献计量指标使用者应引以为戒 [J]. 《研究政策》，2017年，46(8)：1416-36. [10] FOSTER J G, RZHETSKY A, EVANS J A. 科学家研究策略中的传统与创新 [J]. 《美国社会学评论》，2015年，80(5)：875-908. [11] LUO Z, LU W, HE J, 等. 研究问题与方法的组合：科学新颖性的一种新测度 [J]. 《信息计量学杂志》，2022年，16(2). [12] CHEN C, DUBIN R, KIM M C. 再生医学的新兴趋势与最新进展：科学计量学更新（2000—2014） [J]. 《生物治疗专家意见》，2014年，14(9)：1295-317. [13] RZHETSKY A, FOSTER J G, FOSTER I T, 等. 选择实验以加速集体发现 [J]. 《美国国家科学院院刊》，2015年，112(47)：14569-74. [14] WU L, WANG D, EVANS J A. 大型团队推动科学与技术的发展，小型团队则颠覆科学与技术 [ ] 2019, J . 《自然》, 566 7744 378 82 ( ): - . [15] LEAHEY E, LEE J, FUNK R J. 哪些类型的新颖性最具颠覆性？ [J]. 《美国社会学评论》，2023年，88(3)：562-97. [16] LIN Y, EVANS J A, WU L. 科学新方向源于脱节与不协调 [J]. 《信息计量学杂志》，2022年，16(1). [17] BOUDREAU K J, GUINAN E C, LAKHANI K R, 等. 审视知识前沿内外：科学中的知识距离、新颖性与资源配置 [J]. 《管理科学》，2016年，62(10)：2765-83. [18] WANG Z, WANG Z, ZHANG G, 等. 一种基于混合图和大语言模型的方法：通过知识重组与传播测量科学新颖性 [J]. 《应用专家系统》，2026年，298. [19] HUANG S, LU W, XU Z, 等. 基于比较能力的大模型识别潜在颠覆性研究 [J]. 《信息处理与管理》，2025年，62(6). [20] WANG Z, ZHANG H, CHEN J, 等. 一种通过整合主题建模与云模型测量科学论文新颖性的有效框架 [J]. 《信息计量学杂志》，2024年，18(4).[21] 杨杰,邓三鸿,王昊. 科学研究的颠覆性创新测度——相对颠覆性指数 [J]. 情报学报, 2023, 42(9): 1052-64.[22] SHANNON C E. 通信的数学理论 [J]. 《贝尔系统技术杂志》，1948年，27(3)：379-423. [23] TANG X, SHI W, WU R, 等. 图书馆与信息科学（LIS）领域团队规模的扩大：规模越大就一定越好吗？ [J]. 《信息科学杂志》，2023年。[24] LIN Y, LI L, WU L. 团队规模及其对颠覆性指数的负面影响 [ ] 2025, J . 《信息计量学杂志》, 19 3 ( ). [25] SCELLES N, TEIXEIRA DA SILVA J A. 改进领域加权引文影响力（FWCI），使同一领域内的出版物影响力具有可比性：以体育管理为例 [J]. 《科学计量学》，2025年，130(3)：1571-86.

[26] HOYNINGEN-HUENE P, LOHSE S. 关于库恩“本质张力”的自然化 [J].《科学史与科学哲学研究》A辑，2011年，42(1)：215-8.

附录

附录 **A**

表 1 三组论文的多样性分布

组别	样本量	唯一领域数	香农指数	辛普森指数	基尼-辛普森指数
荣誉组	10008	4	1.099	0.625	0.625
影响力组	7077	4	1.278	0.694	0.694
零被引组	2939	4	1.214	0.673	0.673

附录 **B**

表 2 不同年份和领域论文量

发表_年份	物理科学	生命科学	健康科学	社会科学
2005	1667	1120	644	640
2010	1553	1113	595	552
2015	1695	1072	701	501
2020	1501	900	1089	852
2025	2093	896	482	358
	研究领域	文献数量	组内占比 (%)	总占比 (%)
影响力组	物理科学	3094	43.72	15.45
	健康科学	1848	26.11	9.23
	社会科学	1258	17.78	6.28
	生命科学	877	12.39	4.38
	研究领域	文献数量	组内占比 (%)	总占比 (%)
荣誉组	物理科学	4453	44.49	22.24
	生命科学	4037	40.34	20.16
	健康科学	1130	11.29	5.64
	社会科学	388	3.88	1.94
	研究领域	文献数量	组内占比 (%)	总占比 (%)
零被引组	社会科学	1257	42.77	6.28
	物理科学	962	32.73	4.8
	健康科学	533	18.14	2.66
	生命科学	187	6.36	0.93

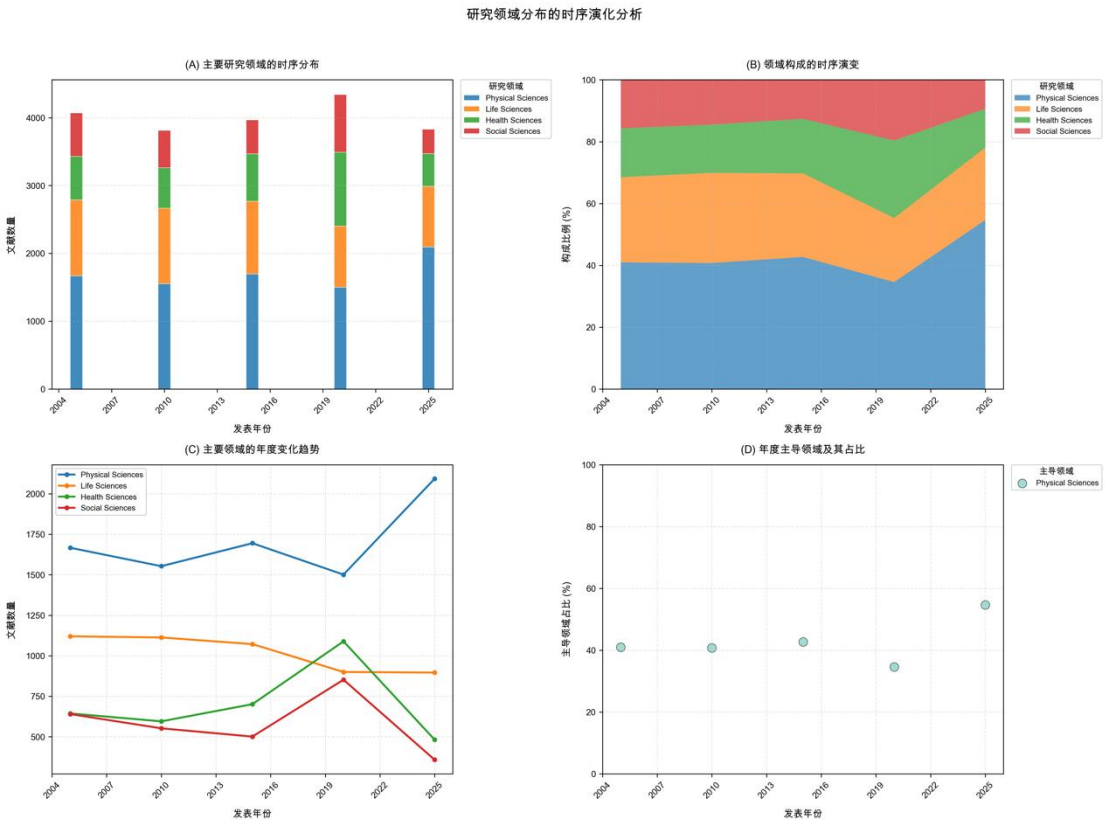

图 1 不同年份和领域的文献数量

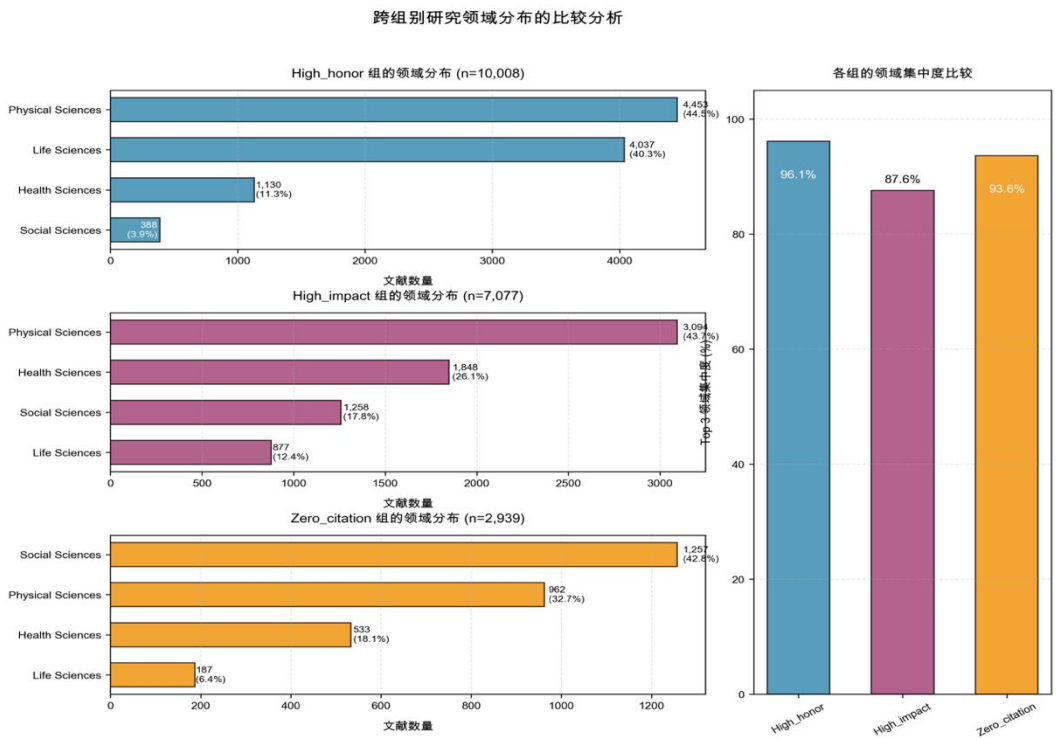

图 2 不同年份、组别、研究领域的文献数量

附录 C

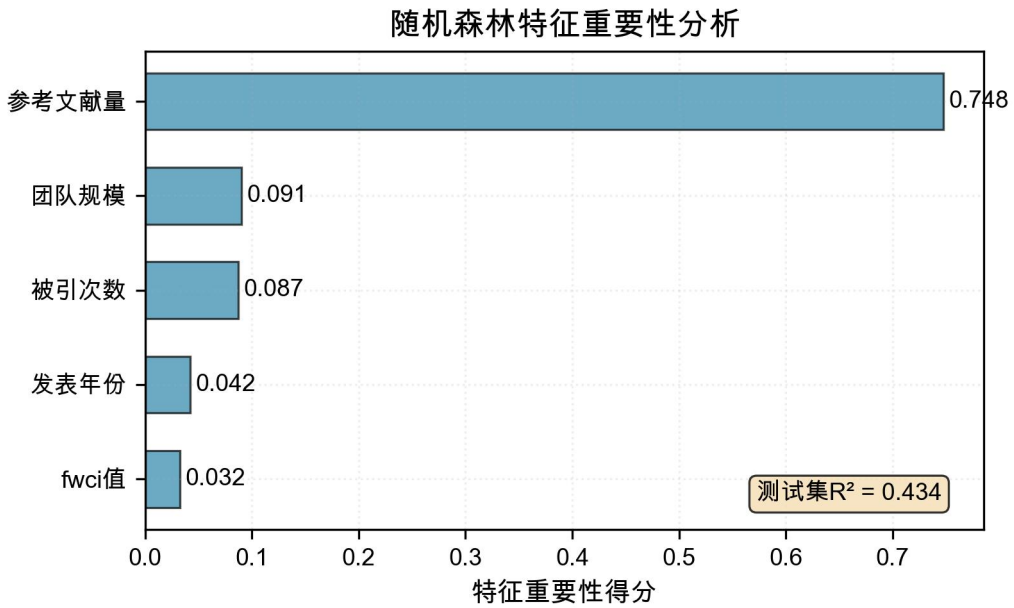

图 3 随机森林特征重要性得分

（责任编辑 XXX）