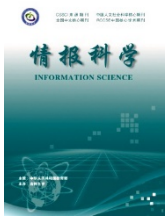

情报科学
Information Science
ISSN 1007-7634,CN 22-1264/G2

《情报科学》网络首发论文

题目：　跨学科知识元创新组合识别与学术创新机会发现研究
作者：　李秀霞，庞瑞欣
网络首发日期：　2024-12-27
引用格式：　李秀霞，庞瑞欣．跨学科知识元创新组合识别与学术创新机会发现研究[J/OL]．情报科学. https://link.cnki.net/urlid/22.1264.G2.20241226.1828.012

网络首发时间：2024-12-27 13:46:17　网络首发地址：https://link.cnki.net/urlid/22.1264.G2.20241226.1828.012

跨学科知识元创新组合识别与学术创新机会发现研究*

李秀霞，庞瑞欣

（曲阜师范大学传媒学院，山东日照　276826）

摘要：【目的/意义】学术创新机会发现是科研人员开展学术研究的前提，从跨学科知识元组合的角度发现学术创新机会，不仅拓宽了跨学科研究的视角，而且能为研究人员开展规范化的创新研究提供解决方案。【方法/过程】首先，结合规则匹配和主题提取方法抽取目标学科与源学科学术论文中的问题、方法知识元，建立学科内的“问题-方法”共现关系。结合共现关系和语义相似度计算，实现跨学科方法知识元的迁移。设计新颖性、融合度、重要性等跨学科“问题-方法”创新组合测度方法和创新系数指标，识别具有潜在创新机会的“问题-方法”组合。在此基础上，通过向量合成与凝聚层次聚类，详细分析多样化的跨学科创新机会。【结果/结论】通过实证研究，获得计算机学科方法知识元与情报学科问题知识元间丰富多样的创新组合关系。【创新/局限】将学术创新机会表示为“问题-方法”组合的形式，便于研究人员对创新机会的理解和接受；对识别出的“问题-方法”创新组合进行聚类，便于发现多样化的潜在创新机会，给研究人员开展创新研究提供了较大的自主选择空间。由于对计算机领域方法的理解有限，对得出的创新机会解释不够充分。

关键词：知识元抽取；知识元组合；创新组合聚类；创新机会发现

A Study of Interdisciplinary Knowledge Meta Innovation Combination Identification and Academic Innovation Opportunity Discovery

LI Xiuxia　PANG Ruixin

（School of Communication，Qufu Normal University，Rizhao 276826，China）

Abstract: 【Purpose/significance】The discovery of academic innovation opportunities is a prerequisite for researchers to conduct academic studies. Identifying these opportunities from the perspective of interdisciplinary knowledge element combinations not only broadens the perspective of interdisciplinary research, but also provides solutions for researchers to carry out standardized innovative research. 【Method/process】Initially, rule matching and topic extraction methods were employed to extract problem and method knowledge elements from academic papers in the target discipline and the source discipline. This process established the "problem-method" co-occurrence relationships within disciplines. By integrating these co-occurrence relationships with semantic similarity calculations, the transfer of method knowledge elements across disciplines was facilitated. Measurement methods and innovation coefficient indicators were developed for interdisciplinary "problem-method" combinations, focusing on the assessment of novelty, integration, and significance to identify potential innovation opportunities. Subsequently, a detailed analysis of diverse interdisciplinary innovation opportunities was conducted using vector synthesis and hierarchical clustering techniques. 【Result/conclusion】Through empirical research, this paper presents a rich and varied array of innovation combination relationships between the method knowledge elements of Computer Science and the problem knowledge elements of Information Science. 【Innovation/limitation】Representing academic innovation opportunities as "problem-method" combinations enhances the clarity and comprehensibility of these opportunities for researchers. Clustering the identified "problem-method" innovative combinations enables the discovery of a range of potential innovation opportunities, thereby offering researchers considerable flexibility in their innovative research pursuits. However, limitations in the understanding of methods within the Computer Science domain may affect the comprehensiveness of the explanations for the identified innovation opportunities.

Keywords：knowledge element extraction; knowledge element combination; innovative combination clustering; innovation opportunity discovery

---

*基金项目：国家社会科学基金资助项目“跨学科知识元迁移组合与学术创新机会发现研究"（22BTQ061）

作者简介：李秀霞(1971-)，女，山东日照人，教授，硕士，研究方向：数据挖掘与信息评价。庞瑞欣（2000-），女，山西太原人，硕士研究生，研究方向：数据挖掘与知识发现。

0 引言

在全球科技飞速发展和社会需求不断演变的背景下，创新已成为各国竞相追逐的新焦点，而跨学科研究是学术创新的重要途径之一。回顾科学发展的进程，新兴学科的出现、研究空白点的揭示以及颠覆性创新成果的产生等，往往源自学科间的交融与知识的渗透、流动与转移[1]。大科学时代，学科间的交流愈发密切，知识的交互催生了持续的创新发展，为学科进步和科技突破提供了强劲动力。因此，基于跨学科知识元的迁移，识别具有创新潜力的知识元组合，发现重要且有效的学术创新机会，对于科研机构和科研人员的学术创新规划、竞争优势获取具有重要的指导价值。

学术论文是学科知识的重要载体，深入分析其引用关系和引用内容能够揭示学术创新的生长点[2]。目前，学者们主要通过引文分析、关键词分析和主题分析等开展相关研究。而真正有价值的学科知识往往潜藏在学术论文的细粒度语义单元中[3]。因此，一些学者开始深入到学术论文的语义单元，识别跨学科知识元的关联关系，揭示学科间细粒度的知识交流特征[4]。问题、方法知识元是科技文献中的核心要素，二者之间存在天然的密切联系[5]，跨学科的“问题、方法”组合已成为科学创新的重要途径[6]。据此，本文基于知识元理论，聚焦学术论文中的问题、方法知识元，通过挖掘跨学科问题、方法知识元间的组合关系，探索跨学科知识元组合的创新潜力，为学术创新和科研决策提供参考支持。

1 相关研究

1.1 知识元抽取

知识元是表示、控制、管理和操作知识的基本单元，在文献中以词语、概念、术语来表征[7]。知识元的类型有很多，不同的分类形式对知识元的划分不同。按照知识元在科技文本中表示的内容，可以将知识元分为类别知识元、领域知识元、背景知识元、问题知识元、理论知识元等[8]，不同的知识元具有不同的组合创新潜力。近年来，知识元在探测知识的产生、传播与应用，追踪知识基础、知识中介和知识前沿，研究知识结构、知识演化与知识重组等方面均有应用，但在创新机会发现中的应用研究相对匮乏。

知识元抽取是根据知识元的属性和特征，按照一定的规则提取文献资源中各种类型的知识元。《知识元挖掘》一书在知识元抽取实践中起到奠基性作用[9]。知识元抽取方法有多种，主要有人工抽取方法、基于规则的抽取方法（又称模式匹配抽取）、基于统计的机器学习抽取方法和基于深度学习的抽取方法[10]。人工抽取是利用人工标注抽取知识元，因而具有较高的准确率[11]，但需要人工标注者具有较全面的领域知识，而且时间成本高，加上标注的文献数量有限，其分析结论往往并不具有通用性。基于统计的机器学习抽取方法是通过构建自动化体系，借用机器学习辅助抽取知识元，该方法通常结合机器学习和文本特征选择实现知识元的抽取，因此抽取效率较高，但需要领域专家根据学科特点设计不同类型的知识元模板[12]。基于深度学习的抽取方法或借助神经网络模型自动学习句子特征或采用自动化技术构建大规模标注语料库，自动化程度明显提高，克服了机器学习抽取方法中特征选择的问题[13]，但该方法在抽取过程中需要大量的训练语料，且对算法选择和参数设置具有较高的要求。基于规则的方法是根据文本句式结构或者符号特点人工构建抽取规则，依据规则抽取知识元，该方法对规则制定的要求较高，因而抽取结果的准确率较高，并且在文献知识元抽取中已有一定的实践基础，不足在于抽取质量受文本表述结构的影响较大。综上分析，根据所选文献资料的特点与规模，本文利用 LDA 主题模型完善基于规则的知识元抽取方法，实现对问题、方法知识元的抽取。

1.2 跨学科组合与学术创新

学术创新是在已沉寂的研究领域提出新思想，在活跃的研究领域取得重大进展，或将彼此分离的研究领域融合起来[14]。学术创新机会是指在特定时期某学术领域在学术思路、学术观点、研究方法和研究方向等方面的潜在发展点；学术创新机会为学者提供了探索未知、解决问题和推动学科发展的可能性，是学术研究持续发展的动力源泉。学术创新机会具有学科前沿性、学科交叉性和社会价值性，不同学科、不同领域之间的知识交叉、碰撞和融合是创新思想的主要来源之一。来自外部领域的知识对于知识系统的作用是能够使领域内的知识产生突变[15]。李长玲[16]将跨学科知识组合分为强组合关系、弱组合关系，弱组合关系往往存在潜在的知识生长点，比强组合关系更能跨越领域界限获得创新机会。目前，已有的

跨学科组合关系，多是基于引文关系发现跨学科知识创新点[17]，也有基于共词网络发现跨学科新兴主题的研究[18]。近年来，随着知识信息的表达和组织由物理层次的文献单元逐渐向认知层面的知识单元转化，研究人员开始从文献知识元角度发现跨学科创新点。曹树金等[19]通过抽取学术论文的研究问题、研究方法、理论原理等实体及其关系，构建领域知识图谱，根据跨学科知识图谱洞察不同学科知识间的潜在关联，发现针对相似问题的跨学科解决方案，该项研究为科研创新机会发现提供了新颖的借鉴视角。

1.3 评述

综上所述，无论是知识元抽取还是跨学科知识的创新，其研究内容均相对丰富，并形成了系统的研究体系。在知识元抽取方面，其发展趋势之一是基于深度学习的抽取方法逐渐成为主流，另一趋势是多方法结合的抽取策略。而跨学科知识创新发展的轨迹正从文献层面的粗粒度组合创新逐步转向更为细粒度的知识组合创新。学术研究是一个提出问题和解决问题的过程，因此，在多种类型的学术论文知识元中，最能体现学术创新价值的知识元是问题、方法知识元[5]。目前，基于知识元的跨学科研究尚显不足，为数不多的研究主要集中在分析学科内部问题、方法知识元关系网络的形成及演化机制[20]，或通过构建跨学科知识元的知识图谱来揭示不同学科知识联系的内在结构和模式[19]。适应学术创新发展的需求，本文拟采用规则匹配与主题提取相结合的方法抽取学术论文中的问题、方法知识元；在跨学科知识元迁移的基础上，识别跨学科知识元的创新组合，进而发现学术研究机会，为开展学术创新研究提供选题参考。

2 研究框架与研究方法

本研究的实现目标是识别跨学科知识元的创新组合、发现学术创新机会，拟解决的核心问题是：（1）如何抽取不同学科领域的问题、方法知识元；（2）如何识别跨学科知识元的创新组合；（3）怎样通过跨学科知识元的创新组合发现学术创新机会。为解决上述问题，设计图 1 所示的研究路线。

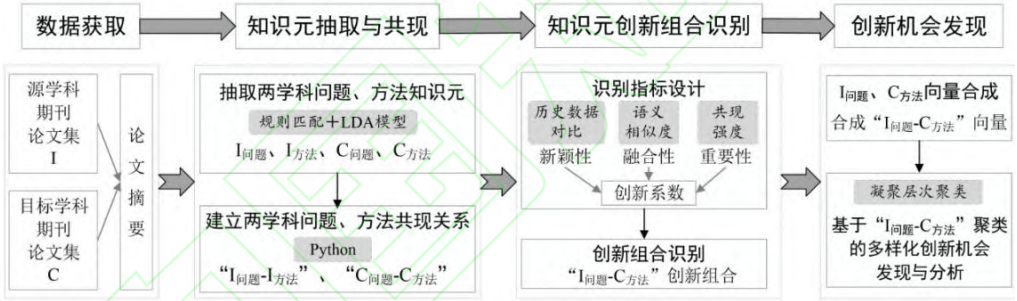

图 1　研究路线

Figure 1 Research route

实现跨学科知识迁移组合首先需要确定源学科与目标学科，要求源学科和目标学科的问题与方法知识元间存在一定的关联性、相似性，且在特定环境下能够将源学科知识元的相关特征映射到目标学科。计算机科学是一门研究程序系统、信息处理与人工智能的系统学科，为信息分析与服务提供了有力的技术、工具和方法，极大提升了信息获取与分析的效率。情报学作为一门以信息为主要研究对象的综合性学科，对技术、方法具有天然的亲和力和敏感性，往往用计算机技术辅助解决信息收集、处理、加工、整理等问题，因此与计算机科学的交叉关系极为密切。由此，本文选取计算机科学为源学科，情报学为目标学科。

2.1 知识元抽取与共现

本研究结合规则匹配和 LDA 模型抽取学科领域的知识元[21]。规则匹配的方法要求被抽取的对象及其上下文具有明显的模式特征或语法特征，可借助句式结构、框架或者符号的共性特征等建立规则表达式来识别、抽取文本中的知识元。具体过程是：采用等距抽样的方法，精读样本论文的摘要内容，并归纳制定论文中问题、方法知识元的描述规则。为使规则有较高的包容性，要求构建的正则表达式的描述规则尽可能完备，并契合语料集特征。分析句子模式，归纳总结句法结构，将一些相似度较大、重合度较高的句子模式合并，以提升规则的结构化程度。通过 Python 程序将提取的样本论文摘要和规则读入，先根据标点符号将所有样本论文的摘要进行分句、断句处理，再令每一条规则遍历每一个切分后的短句，

适当限制知识元字符串的长度，通过字符串匹配方式从文本中获取知识元。在人工干预下构建学科领域内的问题、方法知识元库，并将其引入分词系统。最后，以构建的知识元库替换全局语料库，利用 LDA 模型分别抽取目标学科、源学科领域的问题、方法知识元。

构建学科内部的“问题-方法”共现关系，明晰学科内部问题与方法的对应关系。首先，对每篇论文抽取出的问题、方法做去重处理，以确保一篇论文中每种问题、方法只出现一次。然后，调用 Python 程序中的 Numpy 工具包，按照论文抽取顺序依次将抽取出的问题、方法表示为数组方式，并以每篇论文为单位将问题、方法对齐，建立学科领域内“问题-方法”的共现关系。最后，使用程序遍历每一篇论文，统计“问题-方法”共现的次数。

2.2 跨学科知识元创新组合识别方法

在建立学科领域内“问题-方法”共现关系的基础上，结合知识元的相似关系实现跨学科方法知识元的迁移，以丰富目标学科研究问题的解决方法，为研究人员提供学术创新机会。但不同的跨学科“问题-方法”组合其创新程度不同，解决问题的能力也存在差异，而且还可能产生一些无价值的组合[22]。由此，下面设计跨学科知识元组合的融合性、重要性、新颖性测度方法以及创新系数指标，以识别具有创新价值的知识元组合。

为方便解释说明，将目标学科中的问题、方法知识元用$I_{问题}$、$I_{方法}$来表示，将源学科中的问题、方法知识元用$C_{问题}$、$C_{方法}$来表示，跨学科的“问题-方法”组合用“$I_{问题}-C_{方法}$”来表示。

（1）“$I_{问题}-C_{方法}$”组合的新颖性

“$I_{问题}-C_{方法}$”组合的新颖性是指新出现的或前所未有的组合，论文中关键要素组合的新颖性能够衡量研究成果的创新性[23]。陆泉[24]、钱佳佳等[25]根据截至成果发表时论文中“问题-方法”组合在同领域出现的频数测度组合的新颖性；Ren H 等[26]认为，新颖的知识元组合应该是以前没有或很少在出版物上公开发表过。借鉴前人的思想，以统计年t_0为界限，通过历史数据对比法对比“$I_{问题}-C_{方法}$”组合在t_0年之前出现的频次$f_{(I_{问题}-C_{方法})_{t0-}}$与$t_0$年之后出现的频次$f_{(I_{问题}-C_{方法})_{t0+}}$，以此筛选高新颖性（$Nov_k$）的“$I_{问题}-C_{方法}$”组合。即：

$$\begin{cases} f_{(I_{问题}-C_{方法})_{t0+}} > f_{(I_{问题}-C_{方法})_{t0-}} \text{时，} Nov_k \text{高} \\ f_{(I_{问题}-C_{方法})_{t0+}} < f_{(I_{问题}-C_{方法})_{t0-}} \text{时，} Nov_k \text{低} \end{cases} \quad (1)$$

t_{0-}、t_{0+}分别为t_0年之前、之后的时间。为确保“$I_{问题}-C_{方法}$”组合的新颖性，保留高Nov_k的“$I_{问题}-C_{方法}$”组合，删除低Nov_k的“$I_{问题}-C_{方法}$”组合。

（2）“$I_{问题}-C_{方法}$”组合的融合性

融合性是指“$I_{问题}-C_{方法}$”组合的合理性、匹配性，能够反应“$I_{问题}-C_{方法}$”组合中$I_{问题}$与$C_{方法}$作为一个整体相互协调、创造新知识的潜力，$I_{问题}$与$C_{方法}$的融合性决定着创新研究的成败与质量。“$I_{问题}-C_{方法}$”组合中$I_{问题}$、$C_{方法}$分别与$C_{问题}$、$I_{方法}$的语义相似度越高，则$I_{问题}$与$C_{方法}$的兼容性越强，“$I_{问题}-C_{方法}$”组合的融合性越高。

本文采用语义模型 Word2Vec 表示两学科$I_{问题}$、$I_{方法}$、$C_{问题}$、$C_{方法}$的词向量。Word2Vec 能够理解词在特定上下文中的意义，能更好地表示词与词之间的类比关系和相似关系。Word2Vec 包含 CBOW 和 Skip-Gram 两种训练方法，CBOW 是通过上下文来预测当前词语的概率，侧重于表示“语法功能相似度”；Skip-Gram 是通过当前词语来预测上下文词语的概率，侧重于表示“语义主题相似度”[27]，而且 Skip-Gram 模型在处理专业领域文本方面更优越，因此，选择 Skip-Gram 模型负采样方法进行训练词向量。

基于$I_{问题}$、$I_{方法}$、$C_{问题}$、$C_{方法}$的词向量，利用余弦相似度计算“$I_{问题}-C_{方法}$”组合的融合性，计算公式为：

$$fus_k = \frac{Sim(I_{问题k}, C_{问题k}) + Sim(I_{方法k}, C_{方法k})}{2} \quad (2)$$

$$Sim(I_{问题k}, C_{问题k}) = \frac{v_{I_{问题k}} \cdot v_{C_{问题k}}}{\|v_{I_{问题k}}\| \times \|v_{C_{问题k}}\|} \qquad Sim(I_{方法k}, C_{方法k}) = \frac{v_{I_{方法k}} \cdot v_{C_{方法k}}}{\|v_{I_{方法k}}\| \times \|v_{C_{方法k}}\|} \quad (3)$$

式中$v_{I_{问题}}$、$v_{C_{问题}}$与$v_{I_{方法}}$、$v_{C_{方法}}$分别代表两学科中$I_{问题}$、$C_{问题}$与$I_{方法}$、$C_{方法}$对应的词向量，$Sim(I_{问题k},C_{问题k})$、$Sim(I_{方法k},C_{方法k})$则表示两学科间的问题语义相似度和方法语义相似度。两学科间的$Sim(I_{问题k},C_{问题k})$、$Sim(I_{方法k},C_{方法k})$值越高，其$C_{方法}$与$I_{问题}$对齐的可能性越大[28]，“$I_{问题}-C_{方法}$”组合的融合性越高。研究中根据具体情况设定相似度阈值。

（3）“$I_{问题}-C_{方法}$”组合的重要性

知识元组合的重要性反映“$I_{问题}-C_{方法}$”组合解决问题、促进学科创新的程度。一般来说，在目标学科中，$I_{问题}$如果对应多种解决方案，说明该$I_{问题}$是本学科的研究热点，得到本学科的高度重视，这类$I_{问题}$通过本学科领域的$I_{方法}$实现创新的空间已经不大，而利用跨学科的$C_{方法}$来解决，则可能带来突破性的创新[29]。对于研究量较少的$I_{问题}$，利用本领域成熟的$I_{方法}$就能实现创新，没必要借鉴其他领域的方法。在源学科中，解决问题较多的$C_{方法}$通常有更多的成功案例和证据支持，是更成熟、更被广泛接受的$C_{方法}$，较容易在新的学科中找到应用场景，移植到目标学科后解决$I_{问题}$可能比较理想，风险也较低。而应用较少的$C_{方法}$需要更多的实践检验，在其他学科应用的成功率尚需推敲。因此，在目标学科中，$I_{问题}$与$I_{方法}$共现的次数越多，$I_{问题}$越重要，越需要借助$C_{方法}$实现创新；在源学科中，$C_{方法}$与$C_{问题}$共现的次数越多，$C_{方法}$越重要、越成熟；这样的$I_{问题}$与$C_{方法}$组合后将具有重要的学术价值。由此，设计“$I_{问题}-C_{方法}$”组合的重要性计算公式为：

$$imp_k=\frac{\sum_{i=1}^{n}(f_{(I_{问题k}-I_{方法i})})}{\sum_{i=1}^{n}f_{I_{方法i}}}+\frac{\sum_{j=1}^{m}(f_{(C_{方法k}-C_{问题j})})}{\sum_{j=1}^{m}f_{C_{问题j}}} \qquad (4)$$

$f_{(I_{问题k}-I_{方法i})}$、$f_{(C_{方法k}-C_{问题j})}$分别为目标学科与源学科中$I_{问题}$与$I_{方法}$、$C_{方法}$与$C_{问题}$的共现次数。

（4）知识元组合的创新系数ICO_k

根据“$I_{问题}-C_{方法}$”组合的新颖性、融合性、重要性，设计创新系数指标，以全面评估“$I_{问题}-C_{方法}$”组合的创新程度、$C_{方法}$与$I_{问题}$融合的紧密性以及对目标学科创新发展的重要性。创新系数ICO_k的计算见公式（5）。权重系数α、β、γ可依据专家意见和对创新目标的具体要求来确定。

$$ICO_k=\alpha*fus_k+\beta*imp_k+\gamma*Nov_k \qquad (5)$$

2.3 基于“$I_{问题}-C_{方法}$”组合的多样化创新机会发现

“$I_{问题}-C_{方法}$”的创新组合为开展学术创新提供了有价值的知识关系基础。但研究方案的形成需要建立在系统的知识背景上，学术研究中的问题是复杂的、多维度的，同时，研究者的知识基础有别，研究兴趣多样，仅靠一组“$I_{问题}-C_{方法}$”组合难以确定学术创新方案。而将“$I_{问题}-C_{方法}$”创新组合聚类，便于研究者比较$I_{问题}$、$C_{方法}$及“$I_{问题}-C_{方法}$”之间的相似性和差异性，启发研究者从宏观的角度审视$I_{问题}$、$C_{方法}$，理解$I_{问题}$、$C_{方法}$的逻辑关系和知识背景，从“$I_{问题}-C_{方法}$”组合的聚类结果中系统地挖掘出更多自主化、多样化的创新机会。为此，通过聚类算法将得到的“$I_{问题}-C_{方法}$”组合聚为不同的类型，使$I_{问题}$、$C_{方法}$之间的关系更加系统。首先，将“$I_{问题}-C_{方法}$”创新组合中$I_{问题}$、$C_{方法}$对应的词向量进行降维处理，在 2 维空间将两者合成为一个“$I_{问题}-C_{方法}$”向量。然后，利用凝聚层次聚类法对得到的“$I_{问题}-C_{方法}$”创新组合进行聚类，在此基础上，分析基于“$I_{问题}-C_{方法}$”创新组合的创新机会。

3 实证研究

3.1 数据获取与处理

选取情报学领域跨学科知识吸收能力较强的 5 种 CSSCI 期刊（《情报学报》《情报科学》《情报理论与实践》《情报杂志》和《数据分析与知识发现》），在维普数据库下载上述各期刊 2021 年的文献共 1344 篇，批量导出引证参考文献共 12826 篇，从中确定知识流动性强的计算机学期刊。利用 Python 程序提取引证参考文献中的期刊名称，统计被引期刊的引用频次并对其排序。根据期刊引用频次的顺序，筛选出情报学领域引用的计算机学科频次较高的 5 种期刊（《中文信息学报》《计算机科学》《计算机

应用研究》《计算机学报》《计算机工程与应用》），以此作为本研究的源学科期刊。摘要是期刊论文核心内容的缩影，凝聚了论文的精华，通常包含问题、方法等关键知识元。因此，导出两个学科上述期刊中的摘要作为知识元抽取的文本资料，以提高文本处理的效率和质量。

3.2 知识元抽取与共现关系发现

利用加载了自定义词典的 jieba 工具对摘要进行分词处理，并用前期研究中构建的知识元库[21]替换全局语料库，利用 LDA 模型抽取两个学科领域的问题、方法知识元。利用 excel 中的 VBA 程序和人工检查对其进行消歧、去重、合并和矫正等处理。最终从情报学 4990 篇有效文献中共抽取问题知识元 8768 次，1831 个；方法知识元 5056 次，共 447 种；从计算机科学 10336 篇有效文献中共抽取问题知识元 9267 次，2150 个；方法知识元 6894 次，819 种。调用 Python 程序中的 Numpy 工具包，按照抽取顺序依次将问题、方法表示为数组方式，并对齐每篇论文中的问题、方法。通过遍历每一篇论文摘要，统计问题与方法的共现关系和共现次数。最终获得情报学领域“问题-方法”共现关系 7928 对，计算机学领域“问题-方法”共现关系 4800 对，部分结果见表 1、表 2。

表 1　情报学“问题-方法”共现关系（部分）

Table 1 Information science "problem-method" co-occurrence (partial)

问题	方法						
文本挖掘	数据生命周期	TRIZ 理论	系统动力学	……	LSTM	ATM 模型	PLDA 模型
竞争情报	指数随机图模型	WSR 系统方法	数据包络分析	……	PEST 分析法	神经网络模型	德尔菲法
用户行为	系统动力学	形式概念分析	马尔可夫模型	……	关联规则挖掘	社会网络分析	结构方程模型
智库建设	SWOT 分析法	理论推演	熵值法		社会网络分析法	Kano 模型	信息生态理论
……	……	……	……	……	……	……	……
个性化推荐	卷积神经网络	协同过滤	主题模型	……	语义网	本体建模	知识图谱

表 2　计算机科学“问题-方法”共现关系（部分）

Table 2 Computer science "problem-method" co-occurrence (partial)

问题	方法						
文本聚类	Canopy 聚类算法	BA 算法	最近邻算法	……	CFSFDP 聚类算法	Adam 一阶优化算法	混沌粒子群优化算法
目标搜索	GeoHash 算法	李雅谱诺夫方法	粒子群优化	……	人工蜂群算法	梯度下降法	蜂群算法
资源分配	混合蜂群算法	网络优化模型	凸优化理论	……	蜂群算法	动态博弈理论	蝗虫优化算法
信息推荐	社会正则化算法	隐语义模型	协同过滤	……	LSTM	贪婪算法	词向量模型
……	……	……	……	……	……	……	……
多目标优化	混合变异杂草算法	KM 匹配	隐语义模型	……	多目标粒子群优化	动态博弈理论	BP 神经网络

3.3 跨学科知识元创新组合识别

根据前文提出的“$I_{问题}-C_{方法}$”创新组合的融合度、新颖性、重要性测度方法和创新系数指标，识别具有创新潜力的“$I_{问题}-C_{方法}$”组合。

（1）“$I_{问题}-C_{方法}$”组合的新颖性

计算机科学与情报学之间知识交流密切，一些方法已为两学科共享，如 SVM、Word2Vec、LSTM 等，这些方法与情报学中的$I_{问题}$组合，其创新性并不显著。因此，本文将未在目标学科中出现过的“$I_{问题}-C_{方法}$”组合视为具有较高新颖性的组合。通过历史数据对比法，将新发现的组合与目标学科数据集中已有的“$I_{问题}-C_{方法}$”组合进行对比，如果新组合在目标学科数据集中未曾出现，则认为该组合具有新颖性。具体过程为：筛选出在两个学科数据集中都存在的方法，并在目标学科数据集中删除与这些方法对应的“$I_{问题}-C_{方法}$”组合，以保障“$I_{问题}-C_{方法}$”在情报学领域数据集中是首次出现的组合，确保组合的新颖性和创新力。经过筛选，剩余 72975 对“$I_{问题}-C_{方法}$”组合。

（2）“$I_{问题}-C_{方法}$”组合的融合性

“$I_{问题}-C_{方法}$”组合的融合性既要求组合中的$I_{问题}$、$C_{方法}$分别与$C_{问题}$、$I_{方法}$具有较高的语义相似性，又要求$I_{方法}$与组合中的$I_{问题}$存在共现关系，以确保$C_{方法}$与$I_{问题}$在语义上具有关联性。首先，查询与组合中的$I_{问题}$不存在共现关系的$I_{方法}$，查询与该$I_{方法}$语义相似度高的$C_{方法}$，并删除对应的“$I_{问题}-C_{方法}$”。然后，采用 Word2vec

中的 Skip-Gram 模型负采样方法训练两学科的知识元语料库（即两学科经过去重、勘误等操作后的文献摘要）。借助 Python 的 Jieba 库，添加自定义词典（由两学科筛选出的论文关键词组成）与哈工大停用词表，以提高分词效果。模型对应的训练参数设置如下：n_dim=100，min_count=5，window=5，alpha=0.025、epochs=5、sg=1。使用训练好的模型将源学科和目标学科的知识元映射到统一的向量空间。利用公式（2）、（3）计算 $Sim(I_{问题k}, C_{问题k})$、$Sim(I_{方法k}, C_{方法k})$，进而计算“$I_{问题}-C_{方法}$”组合的 fus_k 值，结果见表 4。

（3）“$I_{问题}-C_{方法}$”组合的重要性

在目标学科中，与多种方法存在共现关系的 $I_{问题}$ 是学科内重点关注的问题，也是学科研究的热点问题。这类 $I_{问题}$ 利用本学科内的方法开展研究，创新的空间已经不大；而利用跨学科的有效方法来解决，则能实现突破性的创新。在源学科中，与更多问题存在共现关系的 $C_{方法}$ 是源学科内相对成熟的方法，这些方法迁移至目标学科时解决 $I_{问题}$ 的成功率通常较高，因此是跨学科研究中极具价值的方法。按照公式（4）计算“$I_{问题}-C_{方法}$”组合中的重要性。结果见表 4。

（4）创新系数 ICO_k 计算

在新颖性、融合度、重要性测度的基础上，按照公式（5）计算“$I_{问题}-C_{方法}$”组合的 ICO_k。根据数据结果，结合专家建议，设定 fus_k、imp_k、Nov_k 三者的权重系数分别为 0.4、0.3、0.3。这种分配是为了凸显融合性的作用，对研究者而言，“$I_{问题}-C_{方法}$”组合的 fus_k 越高，利用其开展创新研究的成功率越高，这是学术创新的前提保障；同时也不忽视重要性和新颖性在评估“$I_{问题}-C_{方法}$”组合创新中的作用。由于迁移至情报学的 $C_{方法}$ 是目前情报学数据集中没有出现的方法，所以统一取“$I_{问题}-C_{方法}$”组合的 Nov_k 值为最大值 1。根据数据特征和专家建议，选取 ICO_k 大于 0.6 的“$I_{问题}-C_{方法}$”组合作为创新组合，共识别出“$I_{问题}-C_{方法}$”创新组合 441 对，其中，$I_{问题}$ 有 47 个，$C_{方法}$ 有 51 种。部分结果见表 3。

表 3 “$I_{问题}-C_{方法}$”组合创新得分（按 ICO_k 由大到小排序）

Table 3 "$I_{问题}-C_{方法}$" Combination Innovation Score (in descending order of ICO_k)

$I_{问题}$	$C_{方法}$	fus_k	imp_k	Nov_k	ICO_k
指标体系	蜂群算法	0.616	0.595	0.300	0.725
指标体系	人工蜂群算法	0.619	0.549	0.300	0.712
指标体系	GeoHash 算法	0.625	0.491	0.300	0.697
网络模型	粒子群优化	0.615	0.414	0.300	0.670
知识发现	Canopy 聚类算法	0.641	0.373	0.300	0.668
信息生态	蜂群算法	0.622	0.362	0.300	0.657
技术创新	人工蜂群算法	0.638	0.331	0.300	0.655
专利分析	粒子群优化	0.665	0.296	0.300	0.655
文本挖掘	混沌粒子群优化算法	0.645	0.320	0.300	0.654
知识组织	粒子群优化算法	0.609	0.367	0.300	0.654
文本挖掘	协同进化	0.627	0.340	0.300	0.653
文本挖掘	混合蜂群算法	0.637	0.326	0.300	0.653
知识管理	高斯模型	0.629	0.334	0.300	0.652
……	……	……	……	……	……

分析 441 对“$I_{问题}-C_{方法}$”创新组合中 $I_{问题}$ 和 $C_{方法}$ 的具体内涵，发现 $C_{方法}$ 主要有以下几类：特征选择、关系抽取方法，聚类、分类优化算法，目标搜索、检测与监控方法，趋势预测、时序分析方法，数据处理中的参数优化法，以及路径优化、资源管理优化、决策优化等方法。$I_{问题}$ 大致包括（数据、信息、知识等）组织与管理、（企业、组织、个人、文献等）评估预测、（语义网、本体、社会网络）知识图谱、（资源、舆情、智库等）管理与决策优化以及文本挖掘与知识发现等。而且，“$I_{问题}-C_{方法}$”创新组合中 $I_{问题}$、$C_{方法}$ 之间的关系比较匹配。同时，创新组合中的 $I_{问题}$、$C_{方法}$ 间存在多对多的对应关系，为跨学科学术创新提供了多种机会。

3.4 基于“$I_{问题}-C_{方法}$”创新组合的多样化创新机会发现

3.4.1 “$I_{问题}-C_{方法}$”创新组合聚类

研究问题的解决方案不是唯一的，研究者的研究范式和观察视角又具有多样性，为从“$I_{问题}-C_{方法}$”创新组合中更好地发现有代表性的创新方案，使研究者在熟悉的方向有更多的选择机会，在“$I_{问题}-C_{方法}$”创新组合的基础上，对“$I_{问题}-C_{方法}$”创新组合做进一步的聚类分析。首先，将 441 对“$I_{问题}-C_{方法}$”创新组合中的 47 个$I_{问题}$、51 种$C_{方法}$的词向量降为 2 维；之后，将每对“$I_{问题}-C_{方法}$”中的$I_{问题}$向量和$C_{方法}$向量合成一个“$I_{问题}-C_{方法}$”向量；最后，对得到的 441 个“$I_{问题}-C_{方法}$”向量进行凝聚层次聚类。根据创新组合中$I_{问题}$与$C_{方法}$的词义将“$I_{问题}-C_{方法}$”创新组合划分为 5 类，并为每类“$I_{问题}-C_{方法}$”创新组合加注命名标签，分别为：特征识别与关系发现、信息检索与目标发现、组合优化与结构发现、路径优化与结构发现、类别划分与管理优化，见图 2。

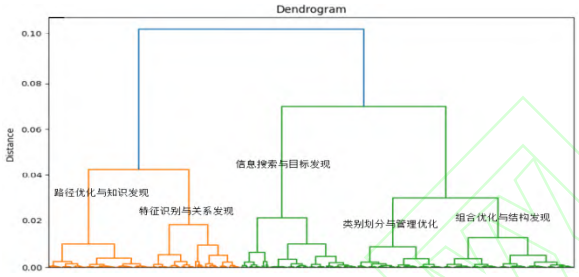

图 2 “$I_{问题}-C_{方法}$”创新组合的聚类结果

Figure 2 Clustering results of "$I_{问题}-C_{方法}$" Combination

为便于解释分析，将各类“$I_{问题}-C_{方法}$”创新组合列于表 4。

表 4 “$I_{问题}-C_{方法}$”创新组合的聚类结果（部分）

Table 4 Clustering results of "$I_{问题}-C_{方法}$" Innovation Combination (partial)

组合聚类名称	“$I_{问题}-C_{方法}$”组合
Type1: 特征识别与关系发现（18 个$I_{问题}$、21 种$C_{方法}$）	信息生态-蜂群算法、信息生态-粒子滤波算法、科技创新-蜂群算法、信息生态-协同进化、指标权重-蜂群算法、专利分析-蜂群算法、信息生态-帝国竞争算法、信息融合-蜂群算法、信息生态-Kruskal 算法、信息生态-LD 算法、信息生态-DBN 模型、信息资源管理-蜂群算法、企业创新-蜂群算法、专利分析-混合蜂群算法、新型智库-蜂群算法、专利分析-粒子滤波算法、信息资源管理-粒子滤波算法、专利分析-帝国竞争算法、协同过滤推荐-蜂群算法、专利分析-DQN 算法、企业创新-协同进化、风险评估-蜂群算法、专利分析-随机梯度方法、信息资源管理-梯度下降法、新型智库-粒子滤波算法、信息资源管理-帝国竞争算法、信息资源管理-动态博弈理论、企业创新-梯度下降法、隐私保护-帝国竞争算法、信息资源管理-GSO 算法、企业创新-混合蜂群算法、专利分析-DBN 模型、信息资源管理-匈牙利算法、舆情传播-蜂群算法、信息资源管理-随机梯度方法、信息资源管理-DBN 模型、……
Type2: 信息搜索 与目标发现（24 个$I_{问题}$、18 种$C_{方法}$）	文本挖掘-改进人工蜂群算法、文本挖掘-混沌粒子群优化算法、文本挖掘-GeoHash 算法、数据开放-人工蜂群算法、数据开放-蝙蝠算法、舆情管理-改进人工蜂群算法、舆情管理-蚁群优化算法、数据开放-布谷鸟算法、舆情管理-人工蜂群算法、核心竞争力-人工蜂群算法、数据开放-烟花算法、舆情管理-高斯模型、语义网-改进人工蜂群算法、隐性知识-改进人工蜂群算法、公共政策-改进人工蜂群算法、核心竞争力-改进人工蜂群算法、研究热点-改进人工蜂群算法、核心竞争力-高斯模型、公共政策-蝙蝠算法、公共政策-布谷鸟算法、知识检索-人工蜂群算法、核心竞争力-烟花算法、核心竞争力-萤火虫算法、智库建设-改进人工蜂群算法、智库建设-差分进化算法、核心竞争力-鸟群算法、智库建设-蝙蝠算法、公共政策-差分进化算法、公共政策-萤火虫算法、知识检索-差分进化算法、公共政策-正弦余弦算法、公共政策-鸟群算法、公共政策-混沌粒子群优化算法、信息资源管理-狼群算法、……
Type3: 组合优化与结构发现（32 个$I_{问题}$、33 种$C_{方法}$）	指标体系-GeoHash 算法、网络模型-蜂群算法、文本挖掘-混合蜂群算法、知识管理-高斯模型、文本挖掘-SLAM 方法、知识管理-正弦余弦算法、文本挖掘-DQN 算法、文本挖掘-动态博弈理论、技术创新-GeoHash 算法、信息融合-人工蜂群算法、核心竞争力-蜂群算法、指标权重-人工蜂群算法、知识组织-萤火虫算法、知识检索-蜂群算法、专利分析-GeoHash 算法、舆情管理-SLAM 方法、专利分析-布谷鸟算法、信息资源管理-人工蜂群算法、数据开放-动态博弈理论、核心竞争力-贪婪算法、信息融合-萤火虫算法、核心竞争力-粒子滤波算法、核心竞争力-混合蜂群算法、信息资源管理-蝙蝠算

	法、信息融合-高斯模型、核心竞争力-梯度下降法、信息资源管理-布谷鸟算法、资源共享-布谷鸟算法、资源共享-烟花算法、信息资源管理-混沌粒子群优化算法、隐性知识-GeoHash 算法、知识检索-粒子滤波算法、信息化建设-梯度下降法、信息资源管理-ELM 模型、信息化建设-DE 算法、风险评估-人工蜂群算法、……
Type4: 路径优化与知识发现（31 个$I_{问题}$、32 种$C_{方法}$）	技术创新-蜂群算法、文本挖掘-贪婪算法、知识组织-蜂群算法、文本挖掘-匈牙利算法、文本挖掘-蚁群优化算法、知识关联-蜂群算法、专利分析-协同进化、隐私保护-蜂群算法、语义检索-人工蜂群算法、数据开放-帝国竞争算法、语义网-粒子滤波算法、语义网-贪婪算法、信息资源管理-蚁群优化算法、信息资源管理-高斯模型、数据开放-随机梯度方法、信息可视化-人工蜂群算法、专利分析-ACS 算法、智库建设-贪婪算法、信息资源管理-协同进化、语义检索-蚁群优化算法、信息可视化-蚁群优化算法、信息资源管理-烟花算法、信息资源管理-萤火虫算法、信息资源管理-鸟群算法、信息可视化-高斯模型、网络信息传播-蜂群算法、区块链-协同进化、语义检索-烟花算法、智库建设-帝国竞争算法、信息可视化-蝙蝠算法、新型智库-协同进化、信息资源管理-ACS 算法、语义检索-高斯模型、语义检索-GeoHash 算法、智库建设-匈牙利算法、语义检索-萤火虫算法、信息可视化-烟花算法、知识检索-梯度下降法、……
Type5:类别划分与管理优化（28 个$I_{问题}$、27 种$C_{方法}$）	指标体系-人工蜂群算法、文本挖掘-协同进化、文本挖掘-帝国竞争算法、知识管理-差分进化算法、文本挖掘-高斯模型、知识组织-人工蜂群算法、网络模型-萤火虫算法、文本挖掘-鸟群算法、文本挖掘-烟花算法、文本挖掘-ACS 算法、语义网-人工蜂群算法、知识组织-混沌粒子群优化算法、知识关联-人工蜂群算法、语义网-高斯模型、信息资源管理-改进人工蜂群算法、核心竞争力-协同进化、信息融合-混沌粒子群优化算法、信息融合-差分进化算法、企业创新-改进人工蜂群算法、新型智库-改进人工蜂群算法、知识关联-萤火虫算法、文本聚类-萤火虫算法、智库建设-蚁群优化算法、智库建设-高斯模型、核心竞争力-DQN 算法、信息化建设-协同进化、信息化建设-粒子滤波算法、智库建设-鸟群算法、信息资源管理-人工鱼群算法、核心竞争力-DBN 模型、智库建设-正弦余弦算法、风险评估-改进人工蜂群算法、信息化建设-DBN 模型、信息化建设-DQN 算法、精准营销-蜂群算法、……

根据表 4 中的聚类结果能够发现多样化的创新机会。研究人员可以根据创新组合的具体类型，或是依据创新组合中的特定$I_{问题}$，结合自己的研究背景和研究兴趣，选定合适的研究方案。限于篇幅，下面分别以表 4 中的任一类“$I_{问题}-C_{方法}$”创新组合以及在情报学领域研究量相对较多的一个$I_{问题}$为例，对其进行简要分析，探讨表 4 中的“$I_{问题}-C_{方法}$”创新组合为情报学研究提供的创新机会。

（1）基于“$I_{问题}-C_{方法}$”创新组合类型的创新机会发现（以 Type2 类为例）

Type2 中的“$I_{问题}-C_{方法}$”是信息搜索与目标发现类创新组合。信息搜索与目标定位在情报分析、知识发现、决策支持、网络安全、竞争情报等方向的具体任务是：发现数据中的模式、趋势和关联，提供实时、准确的情报；从大量数据中识别和分析潜在的威胁和机会，监测和防御网络攻击；监测竞争对手，分析市场趋势，预测潜在的竞争威胁等[30]。可用的研究方法相对丰富，既有直接的搜索方法，如布尔搜索、通配符搜索等，又有辅助目标定位的算法，如机器学习算法（决策树、SVM、KNN 等）、深度学习算法（CNN、RNN、LSTM 等）。随着大数据技术和计算机技术的快速发展，人们对信息搜索与目标定位效率和准确性的要求不断提升。应对方案有：采用实时响应能力强的搜索与定位技术，快速定位动态变化的信息或资源；或通过多源数据融合技术实现全面的数据洞察。而计算机学科已经具备较多的相关技术。

Type2 类中可借鉴的$C_{方法}$有：人工蜂群算法、改进人工蜂群算法、混沌粒子群优化算法、蝙蝠算法、布谷鸟算法、差分进化算法、烟花算法、高斯模型等。人工蜂群算法、改进人工蜂群算法、萤火虫算法、蝙蝠算法等具有多目标动态优化能力，能够从多维数据中搜索有价值的信息。在$I_{问题}$中，舆情管理与舆情事件的社会背景、舆情环境、信息技术、网络文化、传媒格局、个人行为等高度相关；智库建设受内外环境因素、组织管理因素、资源因素等多方面的影响；评价核心竞争力也需综合考虑评价对象的诸多因素。基于不同因素参数的多目标函数，选择人工蜂群算法、改进人工蜂群算法、萤火虫算法、蝙蝠算法对其进行多目标动态搜索与优化，可实现有效的舆情管理、智库建设和核心竞争力评价。同样，可利用人工蜂群算法、蝙蝠算法、布谷鸟算法、烟花算法等的信息搜索与定位能力，准确定位开放数据中的产权数据、隐私数据，辅助知识产权保护和隐私保护等问题的研究。利用混沌粒子群优化算法、布谷鸟

算法、差分进化算法、烟花算法等搜索政策文件中的关键信息，能够辅助政策文件的精准分析和有效传播；利用人工蜂群算法、改进人工蜂群算法寻找政策文件中最优的信息组合，能够充分挖掘政策文件的应用潜力和价值。借助蝙蝠算法发现政策文本中重要的信息模式，可支持政策分析和决策制定；借用布谷鸟算法、差分进化算法，实现语义网中关键知识的高效查询与搜索，用以辅助知识服务与智能推荐系统的创新发展。

（2）基于“$I_{问题}-C_{方法}$”创新组合中$I_{问题}$的创新机会（以“信息资源管理”为例）

同一$I_{问题}$往往涉及多个研究层面，而在每个不同的层面，采用的$C_{方法}$通常不同。即便是同一个研究层面的$I_{问题}$也可能对应多种不同的$C_{方法}$。研究中可根据具体的$I_{问题}$和需求，从不同类型的“$I_{问题}-C_{方法}$”创新组合中灵活选择$C_{方法}$。下面仅以“信息资源管理”为例，说明如何选择多样化的学术创新机会。

“信息资源管理”的目的是实现对各种信息资源的有效开发和高效利用，涉及信息的获取、加工、组织、存储、传播、服务以及信息检索与利用、信息分析与预测等方面的内容[31]。情报学作为信息服务的重要学科，在信息资源管理研究中已取得丰硕成果，研究方法相对成熟。当前，信息资源的异质性和多源性特征愈发突出，用户对信息质量和服务水平的期望日益提升。面对这一趋势，对“信息资源管理”的研究需要不断创新，积极探索新的技术路径，以提高信息资源管理的效率，增强用户满意度。

表 4 中的各类“$I_{问题}-C_{方法}$”创新组合中都有针对“信息资源管理”开展创新研究的$C_{方法}$。

Type1 中的$C_{方法}$能够从大量资源中提取特征信息，发现资源之间的关键关系，实现对多模态复杂数据的有效管理。如借鉴蜂群算法的群体智能，以及混合蜂群算法和 GSO 算法的协同搜索机制，提取信息资源的核心特征，进而识别其中的层次结构、网络结构、周期性、关联性等复杂模式。基于 DBN 模型的多层神经网络结构，利用 DBN 模型的学习机制，能够从原始数据中学习到从底层到高层的多级特征，发现隐藏在数字资源内部的复杂结构，例如，在文本数据中识别出主题分布，在图像数据中提取出对象的组成元素。粒子滤波算法具有强大的状态估计与跟踪能力，借助该算法能够识别和提取信息资源随时间变化的模式和行为（如数据流的趋势、周期性特征、异常点等），分析不同时间点或不同条件下信息资源状态之间的相互影响，揭示信息资源状态随时间变化的规律以及随时间转换的内在逻辑和机制。匈牙利算法在求解二分图最优匹配问题上具有独特优势，该算法能够在所有可能的匹配关系中提取到起关键作用的特征，因此，可以借助匈牙利算法从信息资源中选择代表性的特征，实现资源在不同需求间的均衡配置，提高资源的利用率。也可利用随机梯度方法学习并提取大规模信息资源数据的关键特征，追踪权重更新过程中的梯度信息，通过优化模型参数，深入挖掘并揭示特征之间的内在联系，发现信息资源的结构关系。

借助 Type2 中的$C_{方法}$能够实现对信息资源的快速检索和精准定位。这类组合中的$C_{方法}$主要是狼群算法。可利用狼群算法模拟狼群在复杂环境中快速定位的能力，对信息资源进行高效动态搜索，帮助在复杂的信息空间快速定位到有价值的信息资源；利用狼群算法模拟狼群的“围攻”行为，帮助确定信息资源的最优检索策略，通过不断调整搜索参数，找到最符合用户需求的资源位置，提高信息资源的检索效果。还可利用狼群算法模仿狼群在狩猎过程中的等级关系、协作和信息共享机制，实现信息资源的有效管理。

利用 Type3 中的$C_{方法}$对信息资源进行组合优化，揭示其关系结构，实现信息资源的高效组织与优化管理。相应的算法有人工蜂群算法、蝙蝠算法、LD 算法、布谷鸟算法、混沌粒子群优化算法、ELM 模型等。上述不同方法可以结合使用，通过双重优化策略提升信息资源组合、管理与检索的效率和质量。如使用人工蜂群算法初始化信息资源的特征空间，通过蜜蜂的搜索行为发现资源间的潜在关系；结合蝙蝠算法对搜索结果进一步优化，通过蝙蝠的回声定位特性加强局部搜索能力，细化资源分类，提高资源分类的准确性，优化资源结构。应用 LD 算法中逐渐降低的“温度”机制搜索解空间，有效处理信息资源中的噪声和不确定性；为提高搜索的质量和避免局部最优，可利用布谷鸟算法进一步优化 LD 算法得到的结果，根据其巢寄生和随机游走行为模式发现未被充分利用或认知的信息资源。还可以结合混沌粒子群优化和 ELM 模型寻找最优或近似最优的资源组合方案，利用混沌粒子群优化算法选择信息资源的

特征作为 ELM 模型的输入，用混沌粒子群优化算法指导 ELM 模型参数的设置，使用选出的最优参数训练 ELM 模型，实现信息资源的优化组合。

利用 Type4 中的$C_{方法}$优化资源搜索路径，发现资源之间的结构关系。对应的$C_{方法}$有蚁群优化算法、协同进化、烟花算法、萤火虫算法、鸟群算法、正弦余弦算法、ACS 算法、高斯模型。这些算法基于自然界的群体行为或自然规律，通过个体之间的协同合作与信息交流，采用启发式和自适应的搜索策略，具有较好的全局搜索能力。在信息资源管理中，借助上述算法能够发现信息资源的相似性、差异性和关联性，发现资源之间的结构关系、用户需求和资源之间的匹配关系；还能够优化资源分配和调度，优化资源的搜索路径和搜索效果，提高资源的利用效率。

Type5 中的$C_{方法}$主要是根据群体协作优化策略，通过优劣对比，实现对大量资源的有序管理。可利用的$C_{方法}$有人工鱼群算法、改进的人工蜂群算法。人工鱼群算法借鉴鱼群的觅食行为模式，通过不断地搜索和评估，能够识别出对资源分类最为关键的代表性特征，并根据这些特征的重要性和相关性，为不同的资源分配最佳权重，以确保资源分类结果的准确性和可靠性；还可以利用该算法分析用户的需求和资源特征，搜索出与用户需求最匹配的资源或资源集合，服务于个性化资源推荐系统、资源发现和智能检索等应用场景。改进的人工蜂群算法通过迭代优化的方式分析和评估大量的数据特征，动态调整特征组合，确定信息资源最优的特征组合。算法不仅能够识别对分类结果贡献较大的特征，还能够根据每个特征对分类结果的影响程度，确定这些特征的最优权重分配，提高信息资源分类的准确性；同时，还可以借助该算法分析资源特征之间的相互作用，揭示资源之间的潜在联系，为资源的整合和关联分析提供依据。

近年来，计算机科学领域涌现出了众多高效且成熟的优化算法，涉及信息的搜索、检测、组合、管理、识别等方面。适应跨学科创新研究的趋势，情报学研究人员应当具有敏锐的洞察力和前瞻性，主动汲取和整合这些先进的算法与技术，以提高情报分析的技术精度和解决实际问题的能力，推动情报学的创新发展。

4 研究结论与意义

跨学科知识组合是创造新知识的有效途径之一，识别知识元的跨学科组合关系有助于发现学术创新机会。本文提出了一种基于跨学科知识元组合的学术创新机会识别方法。主要研究结论为：（1）结合规则匹配和主题提取的方法能够有效抽取学术论文中的问题、方法知识元。基于规则抽取知识元具有较高的准确率，但由于学术论文在行文方式上具有灵活性，句式表达多变，而规则描述受限，导致抽取的知识元不够完善[21]。以构建的学科领域知识元库作为词典，结合 LDA 主题模型抽取知识元，可以有效弥补因规则有限而遗漏知识元的不足，提高知识元抽取的查全率。（2）综合新颖性、融合性、重要性三个维度能够有效识别跨学科知识元的创新组合。已有的创新性评估方法主要考虑新颖性维度[24][25]；也有根据共现关系评估创新组合融合性的研究，但缺乏语义上的关联性[32]。已有的评估方法都忽视了创新组合的重要性，而重要性体现了创新成果的实际应用价值，应用价值是知识元组合具有创新性的必要条件。新颖性、融合性、重要性三个指标不仅体现了知识元组合创新的内涵，也保证了学术创新的合理性和价值性，由此遴选出的“$I_{问题}-C_{方法}$”创新组合能够为学术创新活动提供基础保障。（3）相比通过方法迁移识别创新机会的研究[33]，将创新机会表示为“$I_{问题}-C_{方法}$”知识组合的形式，能够帮助研究人员更直观地发现创新机会。而且，通过对“$I_{问题}-C_{方法}$”创新组合的聚类，可以揭示$I_{问题}$、$C_{方法}$之间的多种关联关系，发现多种潜在的创新机会，可为研究人员根据自己的研究基础、研究倾向开展创新研究提供了较大的自主选择空间。另外，通过实证研究，本文呈现了丰富且详细的计算机科学、情报学合作创新的知识元及其组合关系。随着网络化和智能化技术的不断发展，情报学研究的范畴不断拓展，对高效技术方法的需求日益迫切。实证研究部分获取到从计算机科学迁移至情报学的知识元，并成功识别出具有创新潜能的“$I_{问题}-C_{方法}$”组合，研究结果不仅为情报学研究提供了强有力的技术支持，也为研究人员开展学术创新研究提供了宝贵的参考资源。

本研究的局限性在于仅利用$I_{问题}$、$C_{方法}$之间的关系识别创新组合。事实上，学术论文中与学术创新

有关的知识元不止问题、方法两种，还包括数据对象、研究对象、研究工具等，未来研究将考虑抽取文献中的这些知识元，并基于多元组合关系探索更详细的学术创新机会。另外，本文实证部分识别出了较多的用于情报学开展创新研究的$C_{方法}$，限于对计算机领域方法的理解，对得出的创新机会解释不够充分，未来研究将致力于对$C_{方法}$的学习，加深对$C_{方法}$的理解，以提供更为深入和全面的创新机会分析。

参考文献

[1] 徐晴.图书情报学跨学科知识转移研究[D].武汉:武汉大学, 2016.

[2] 李长玲,高峰,牌艳欣.试论跨学科潜在知识生长点及其识别方法[J].科学学研究,2021,39(6): 1007-1014.

[3] 马费成.情报学的进展与深化[J].情报学报,1996,(5):22-28.

[4] 黄晓捷,熊回香,肖兵等.基于"问题—方法"知识元挖掘的学科知识流动研究[J].图书情报工作,2024,68(8):80-96.

[5] 杨金庆,庞业佳,刘智锋等.“问题—方法”关联视角下领域知识创新网络演化机制研究——以信息资源管理学科群为例[J].图书情报工作,2024, 68(10):97-108.

[6] 唐晓波,向莉丽,牟昊.基于研究问题与研究方法贡献的论文学术价值早期识别方法[J].情报科学,2022,40(9):3-11,19.

[7] 高继平,丁堃, 潘云涛等. 知识元研究述评[J]. 情报理论与实践,2015,38(7):134-138,133.

[8] 秦春秀,杨智娟, 赵捧未等. 面向科技文献知识表示的知识元本体模型[J].图书情报工作, 2018, 62(3): 94-103.

[9] 温有奎,徐国华,赖伯年等. 知识元挖掘[M]. 西安:西安电子科技大学出版社, 2005.

[10] 王忠义,沈雪莹,黄京. 科技文献资源中方法知识元的抽取研究[J].情报科学, 2021, 39(1): 13-20.

[11] CHU H, KE Q. Research methods: What's in the name? [J]. Library & Information Science Research, 2017, 39(4): 284-294.

[12] LIN W, JI D, LU Y. Disorder recognition in clinical texts usingmulti-label structured SVM [J]. BMC informatics, 2017, 18(1): 75.

[13] 余丽,钱力,付常雷等.基于深度学习的文本中细粒度知识元抽取方法研究[J].数据分析与知识发现,2019,3(1):38-45.

[14] 胡凡刚.教育虚拟社区助学者伦理规范[M].北京:科学出版社, 2024.36-43.

[15] 李长玲,刘小慧,刘运梅等.基于开放式非相关知识发现的潜在跨学科合作研究主题识别——以情报学与计算机科学为例[J].情报理论与实践,2018, 41(2): 100-104.

[16] 李长玲,高峰,牌艳欣.试论跨学科潜在知识生长点及其识别方法[J].科学学研究,2021,39(6):1007-1014.

[17] RICHARD N V. Interdisciplinary research by the numbers [J]. Nature. 2015, 525: 306.

[18] 马费成,刘旻璇.知识网络的结构、演化及热点探测——CSSCI(1998-2011)经济学文献计量分析[J].情报科学,2014, 32(7): 3-8.

[19] 曹树金,曹茹烨. 基于知识图谱支持科研创新的跨学科知识发现研究[J]. 情报理论与实践,2022,45(11):10-20.

[20] 杨金庆,庞业佳,刘智锋,等.“问题—方法”关联视角下领域知识创新网络演化机制研究——以信息资源管理学科群为例[J].图书情报工作,2024,68(10):97-108

[21] 邹洋杰,李秀霞,王晓璎.基于知识元抽取的不同学科领域研究方法交流态势分析——以情报学与计算机科学学为例[J].情报杂志,2023,42(7):154-160.

[22] PREEZ D T G, PISTORIUS W C.. Analyzing technological threats and opportunities in wireless data services[J]. Technological Forecasting and Social Change, 2003,70(1):1–20.

[23] WANG J, VEUGELERS R, STEPHAN P. Bias against novelty in science: a cautionary tale for users of bibliometric indicators[J]. Research policy, 2017, 46(8): 1416-1436.

[24] 陆泉,秦雨萱,陈静.跨学科"技术—主题"创新组合识别——以人工智能技术驱动图情领域创新为例[J].图书情报工作, 2024, 68(2):50-61.

[25] 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.
[26] REN H, ZHAO Y. Technology opportunity discovery based on constructing, evaluating, and searching knowledge networks[J].Technovation, 2020(1):102196.
[27] 张剑,屈丹,李真.基于词向量特征的循环神经网络语言模型[J].模式识别与人工智能,2015,28(4):299-305.
[28] 李秀霞,邵作运.基于离群主题词跨学科组合的学术创新机会发现研究[J].情报理论与实践,2023,46(12):122-130.
[29] LEE J , KO N , YOON J ,et al. An approach for discovering firm-specific technology opportunities: Application of link prediction to F-term networks[J].Technological Forecasting and Social Change, 2021, 168(1):120746.
[30] 王洪林,刘伟.基于粒子群优化算法的中小企业竞争情报搜集系统模型[J].科技管理研究.2021,41(21):196-203.
[31] 马捷,白宇婷,韩佳霖.基于 i Schools 多维数据分析的我国信息资源管理学科国际化发展路径研究[J/OL].图书馆建设. https://link.cnki.net/urlid/23.1331.G2.20240715.0858.002.
[32] ROH T , YOON B. Discovering technology and science innovation opportunity based on sentence generation algorithm[J]. Informetrics, 2023, 17:101403.
[33] 庞瑞欣,李秀霞.基于知识元迁移的学科领域方法库构建研究[J].情报理论与实践, 2024, 47(5):204-212.