融合词汇语义功能的学科知识结构识别与演化分析*

王佳敏　房子辰　崔蕴学
（西安电子科技大学经济与管理学院，西安 710126）

摘要：开展学科知识结构识别和演化分析，有助于把握学科领域发展变化规律，进而推动学科发展。传统基于共词网络的研究通常将关键词视为同质性节点，导致分析粒度较粗、语义层次较低，且难以考察研究问题与研究方法等功能词汇之间的协同演化关系。针对上述局限，本研究通过词汇功能增强知识网络节点的语义信息，以提高知识结构识别和演化分析的精细化程度。首先，基于BERT模型识别学术文献中的研究问题与研究方法，构建融合词汇语义功能的异质性共现网络，并采用社区发现算法识别学科知识结构；其次，从知识结构整体演化与“问题–方法”二元演化两个维度开展主题强度量化分析，并以国际计算机学会数据集为例进行实证研究。实验结果表明，该方法有效识别出计算机学科领域的9个知识社区，系统揭示了各知识社区内研究问题、研究方法及其关联模式；融入二元属性的演化分析进一步捕捉了研究焦点的动态转变及其内在“问题–方法”关联逻辑。本研究提升了知识结构分析的细粒度和可解释程度，为学科知识结构挖掘提供了新的视角和方法。

关键词：词汇语义功能；共现网络；问题–方法；知识结构；演化分析；计算机学科
中图分类号：G350　DOI：10.3772/j.issn.1673–2286.2026.03.003
引文格式：王佳敏，房子辰，崔蕴学. 融合词汇语义功能的学科知识结构识别与演化分析[J]. 数字图书馆论坛，2026，22（3）：20–30.

数智时代，科技文献数据呈现爆发式增长[1]，学术信息的有效利用面临严峻挑战。在此背景下，识别学科知识结构的研究主题与演化规律[2]，已经成为科技情报研究的一项重要工作。作为文本挖掘研究的重要议题，主题识别与演化分析旨在通过解构文本内容规律以剖析未来趋势，其中知识结构识别是主题演化分析的基础，主题演化分析则是未来发展趋势预测的关键[3]。学科知识结构的识别和演化分析有助于研究者系统把握本领域的整体发展脉络与知识演进规律，助力科研管理部门实现科研资源的优化配置与高效利用，为国家科学制定学科发展战略、明确科研创新方向与攻坚重点提供重要依据[4]。

学术文本词汇功能是指词汇在文本中的语义角色，一般包含研究问题和研究方法[5–6]等，具有不同语义功能的词汇从不同角度反映文献的研究内容。词汇语义识别能够提升文献中所蕴含知识的挖掘与利用效果，从更细粒度、更深层次对文献内容进行组织和揭示[7]。近年来，词汇功能增强的知识网络研究得到广泛关注，研究者往往以特定语义功能的实体构建网络，探索实体之间的关联特征，如算法实体共现网络[8]、技术实体共现网络[9]，以及词汇功能敏感的关键词引用网络[10]等。然而，鲜有研究探讨不同语义功能词汇在学科知识结构分

收稿日期：2026–01–12
*本研究得到国家自然科学基金青年项目“基于文本语义理解的跨学科知识结构识别与演进路径研究”（编号：72304218）资助。

析及演化中的差异化作用，致使相关分析缺少细粒度的特征刻画。

共词网络作为学科知识结构挖掘与演化分析的一种重要方法，旨在通过词汇之间关联关系的变化情况揭示领域知识结构的演化过程，目前在相关研究中已经得到广泛的应用。但传统共词网络方法大多基于科技文献关键词构建，缺少对词汇语义功能的考虑，导致对知识结构及其演化结果的解读难以深入到文本语义特征层面，限制了知识结构分析的精确性和可解释程度。因此，本研究将研究问题词汇和研究方法词汇融入到共现网络，提出一种词汇语义功能增强的共词网络。与传统将关键词视为同质性节点的做法不同，本研究区分了研究问题词汇和研究方法词汇两类异质性节点，构建异质性共现网络，以期从二者的关联层面揭示学科知识结构的深层逻辑。在此基础上，进一步开展学科知识结构的识别与演化分析。在演化分析方面，本研究突破单一语义角色的局限，从知识结构的整体演化与“问题–方法”二元演化两个维度，系统考察主题强度的动态变化，并引入皮尔逊相关系数，量化二者之间的协同与消长关系，从而在机制层面揭示学科知识结构的演化逻辑，而非仅停留于主题构成的现象描述。

1　相关研究

1.1　学科知识结构识别

学科知识结构是指由特定学科中包含的知识要素（单元）及其相互关系组成的层次知识系统[11]。识别和分析学科知识结构，有助于把握领域发展态势，揭示知识组成规律，近年来受到国内外学者的广泛关注[12]。

现有研究主要沿3条技术路径展开。①基于科学知识网络的建模方法，其中共现网络[13–14]和引文网络[15–16]的应用更为广泛。共现网络通过词汇共现关系揭示研究主题的聚合特征，能有效剖析领域内主题簇的发展进程与成熟程度[17]。引文网络通过引用关系追踪知识流动与主题继承。然而，这些知识网络模型中的元素大多为文献、关键词等实体，缺少对文本深层语义信息的理解和融合，导致对学科知识结构的表示和揭示能力有限。②主题模型方法，以LDA（Latent Dirichlet Allocation）及其变体为代表，通过概率生成建模识别文献中的潜在主题结构[18]。近年来，BERTopic等结合预训练语言模型获取文本语义嵌入的方法进一步提升了主题识别的语义准确性[19]。但该类方法多面向宏观层面的主题分析，难以深入到知识单元之间的直接交流和互动，影响了分析结果的可解释性。③将语义信息引入传统知识网络或主题模型，构建语义加权引文网络[15]、语义增强的知识单元引用网络[20]等，以期同时捕捉结构关系与语义内容，在一定程度上提高了知识结构识别的精确性。然而，这类方法对词汇语义功能的区分仍较为粗放，尚未系统地将研究问题词汇和研究方法词汇作为两类异质性节点纳入统一的分析框架，导致对领域知识结构的解释仍停留在浅层关联层面。

总的来说，尽管上述方法在知识结构识别方面取得了显著进展，但其对词汇语义功能的考虑普遍有限。多数研究将关键词视为同质性节点，忽略了作为研究问题与研究方法不同功能的词汇在文本中承担的不同语义角色，影响了对学科知识结构的细粒度揭示。当然，部分学者在融合语义信息的领域知识结构分析方面的探索，为本研究引入词汇语义功能增强共现网络提供了一定的启发。

1.2　学科知识结构演化分析

学科知识结构演化旨在揭示同一学科领域内不同主题内容的发展历程和变化规律。探索学科知识结构演化有助于快速了解学科发展历程，为理论研究和应用实践提供学理依据。当前研究主要从主题强度、主题结构和主题内容等维度开展学科知识结构演化分析[21]。①在主题强度维度，研究者通过文档主题分布[22]或特征词频次[23]测度主题热度的变化，揭示主题演变过程。②在主题结构维度，一般将主题群落的结构演化划分为新生、消亡、合并、分裂、增长和收缩等基本类型，以此来揭示群落进化的基本特征。刘晋霞等[24]利用不同时间窗口的子话题的新生、继承、合并、分裂、消亡的关联关系来反映热点主题内各个部分间的关系，体现其结构演化过程。③在主题内容维度，余弦相似度[24]、词嵌入[25]等方法常被用于量化主题语义的漂移与继承关系。在实际应用中，往往综合其中多个维度开展领域主题演化分析，例如，Wang等[19]从主题强度和主题内容的演变两个方面揭示了图书情报领域的跨学科性。

学科主题演化分析多在宏观层面展开，难以深入到语义和微观知识关联层面，对学科知识演化过程的解释能力有限。近年来，部分研究开始将词汇语义功能

引入主题演化分析，以期在知识单元层面揭示学科知识结构演化模式。例如：刘自强等[20]利用人工标注的方法将文献核心关键词标注为核心研究问题、主要研究方法、关键技术与工具3类语义角色，构建包含单一语义角色的共词网络，并从主题强度、主题结构和主题内容3个维度展示了2006—2015年我国图书情报领域大数据研究的主题演化情况；陈果等[26]将基于词嵌入类比方案训练所得的集成模型应用于人工智能文献语料集，抽取了“问题–方法”实体关系，从宏观、中观、微观3个层面展示了领域知识演化规律；俞琰等[27]提出一种优化的提示学习方法，使用少量样本对关键词的问题与方法两个语义功能进行识别与分析从而获得更加准确的领域主题演化分析结果。然而，上述研究仍存在局限：①在主题识别阶段，多数研究将问题词与方法词分别聚类或分别构建网络，割裂了二者在真实知识结构中的内在关系，难以揭示“问题–方法”在同一主题内部的动态关联逻辑；②演化分析多在单一语义角色层面展开，或分别考察问题与方法的强度变化，未能在统一的网络框架下量化二者的协同与消长关系。

针对上述不足，本研究将研究问题词汇和研究方法词汇两类异质性节点嵌入同一共词网络，通过社区发现算法实现知识结构的识别，并从整体演化与“问题–方法”二元演化两个维度开展主题强度分析，以量化主题在不同发展阶段的重要性及其内部语义关系的紧密程度。

2　研究框架与方法

2.1　研究框架

本研究的研究框架如图1所示。首先，收集特定领域的文献数据，并进行词汇功能识别、同义词合并等预处理，从而得到研究问题词汇和研究方法词汇；其次，构建词汇语义功能增强的共词网络，利用社区发现算法识别领域知识结构，并使用*Z*-Score值识别各社区的核心节点，进而开展网络结构分析和知识结构主题分析；最后，按年份划分时间窗口，分别计算各知识结构在整体、问题、方法3个层面的主题强度，进而从知识结构整体演化与“问题–方法”二元演化两个维度开展主题演化分析，并进一步分析“问题–方法”的协同关系。

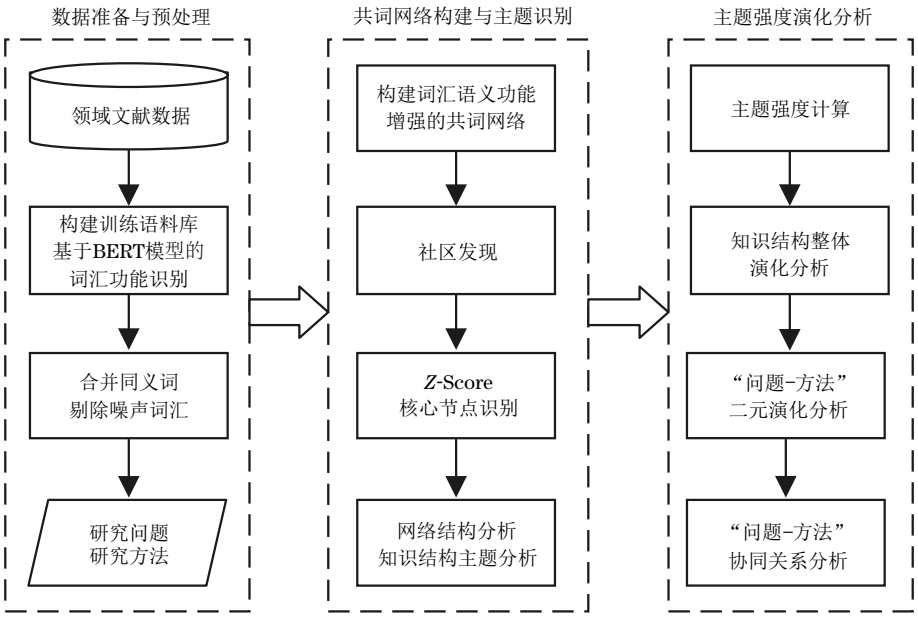


图1　研究框架

2.2　研究方法

2.1.1　词汇功能识别方法

本研究采用Wang等[10]提出的基于BERT模型的方法识别数据集中文献的词汇功能。该方法分为两个步骤。一是训练语料库的自动构建。筛选标题格式为“A based on B”“A using B”“A for B”的论文，利用Stanford NLP工具进行分词、词性标注与实体提取后，采用Knuth-Morris-Pratt算法从标题中提取研究问题词

汇和研究方法词汇，并计算其与论文关键词的相似度，排除相似度过低的词汇，由此得到训练语料库。随机抽取1 000篇文献进行人工验证，精确度达95%。二是基于BERT模型的词汇功能识别。首先，以训练语料库中的论文摘要与关键词拼接作为输入，以研究问题词汇和研究方法词汇作为输出，训练BERT模型，结果在研究问题词汇和研究方法词汇识别上的F1值均达到0.90；其次，将训练好的模型应用于原始数据集中没有特定标题样式的论文，并随机抽取2 000篇文献进行人工验证，精确度均超过85%，表明该方法具有较高的可靠性。通过上述处理，本研究获得了数据集中每篇论文的研究问题词汇和研究方法词汇。

2.2.2　知识结构主题识别方法

利用学术文献构建的共词网络内存在社区现象，同一社区内的节点关联更紧密，可表征学科领域的研究主题[28]。因此，本研究采用社区发现算法进行学科知识结构识别。首先，基于识别出的每篇论文的研究问题词汇和研究方法词汇构建词汇共现对；其次，计算这些共现对的频次和关系形成共现网络；最后，采用Louvain社区发现算法[29]识别学科知识结构主题，并结合Z-Score值确定社区核心节点。沿用前人研究的做法[30−31]，利用Z-Score指标为每个社区内的节点排序，由于本研究中社区内的节点具有语义功能，故分别选取Z-Score值最高的一个研究问题词汇和一个研究方法词汇共同作为社区的主题标签，具体计算方法如式（1）~式（3）所示。

$$B=\sum_{i\in M}N_M^i \tag{1}$$

其中，B表示社区M内所有节点与本社区内其他节点连边的权重和的总和，N_M^i表示社区M中第i个节点与本社区内其他节点连边的权重之和。

$$Q=\sum_{i\in M}\left(N_M^i\right)^2 \tag{2}$$

其中，Q表示社区M中所有节点与本社区内其他节点连边的权重和的平方的总和。

$$Z_j=\frac{N_M^j-B/M^0}{\sqrt{Q/M^0-\left(B/M^0\right)^2}} \tag{3}$$

其中，Z_j表示社区M中第j个节点的Z-Score值，M^0表示社区M的节点数，N_M^j表示社区M的第j个节点与社区M其他节点连边的权重之和。节点的Z-Score值越高，说明该节点与社区内其他节点的关系越紧密，越能代表整个社区。

2.2.3　主题强度演化分析方法

分析学科主题演化的过程有助于理解该学科的发展脉络，并揭示主题在动态变化中的演变规律。为此，本研究将从两个方面展开领域知识结构演化分析。①整体演化分析。通过整合主题内词汇的主题强度，系统揭示领域知识结构的整体发展态势、关键节点分布及演化脉络，为宏观把握主题演进规律提供全景视角。②“问题–方法”二元演化分析。聚焦知识结构内研究问题词汇和研究方法词汇的主题强度动态对比，解析不同时间窗口下二者的强度消长及关联模式，揭示领域中问题聚焦与方法应用的演化特征，为研判领域知识结构的未来发展方向、优化研究策略提供依据。

其中，主题强度（Topic Intensity）是指某一主题在特定时期内的受关注度与研究热度，通常可通过主题词频次、发文量等指标进行测度。考虑到不同年份的发文量存在差异，本研究将主题强度定义为：在时间窗口t内，主题c内部词汇的总频次与该时间窗口总发文量之比。主题强度的具体计算方法如式（4）所示。

$$TI_{ct}=\frac{\sum k_c}{\sum t} \tag{4}$$

其中，TI_{ct}表示主题c在时间t内的主题强度，$\sum k_c$表示主题c内的词汇在时间t内的总频次，$\sum t$表示时间t内的总发文量。

3　实验与结果

3.1　数据集及预处理

国际计算机学会（Association for Computing Machinery，ACM）是全球计算机领域极具影响力的专业学术组织。ACM Digital Library数据库收录了自20世纪50年代以来该学会旗下的期刊论文和会议论文，数据覆盖范围广泛且质量较高，能够客观反映计算机学科的研究全貌。因此，本文选取该数据库作为数据来源，构建的数据集共包含254 028篇文献，时间跨度为2000—2019年。各年份发表的文献数量如图2所示，整体呈增长趋势。

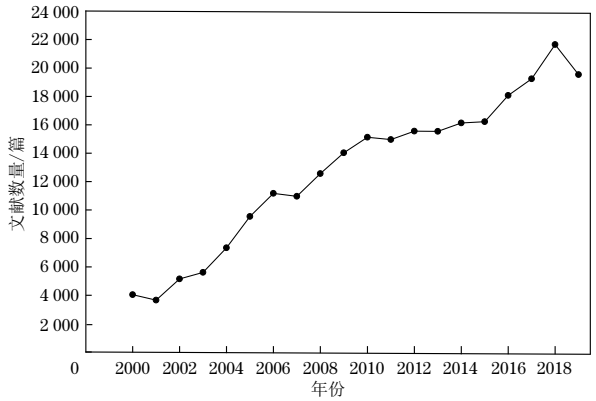


图2　ACM论文数量分布情况

基于前述词汇功能识别方法，本研究共获得508 056个关键词，其中研究问题词汇和研究方法词汇各254 028个。针对识别结果中存在的词形不完整、同义词等问题，进行人工处理。①合并同义词，包括将复数形式转为单数、统一缩写与全称等。例如将“recurrent neural network”与“rnn”统一为“recurrent neural network”。②剔除噪声词汇。对存在歧义的词汇，由2位研究人员独立判断，意见不一致时协商确定。处理后，共获得175 844个不重复的研究问题词汇和172 131个不重复的研究方法词汇。研究问题词汇和研究方法词汇的频次统计情况如表1所示，绝大多数词汇仅出现1次；但高频词汇高度集中，呈现出明显的长尾分布，且研究问题词汇和研究方法词汇在各级频次的占比高度趋近；频次超过9次的研究问题词汇和研究方法词汇占比分别为0.93%、1.00%，均接近1.00%，相关词汇反映了学科领域内的核心议题。

3.2　知识结构识别

3.2.1　网络结构分析

本研究使用Gephi 0.10.1构建词汇语义功能增强的共词网络，识别计算机学科领域的知识结构。原始共现网络规模庞大，难以有效揭示网络的结构特征，且存在大量度数较小的节点，易影响对领域关键知识的把握。因此，本研究综合考虑网络节点规模与分析目的，以节点度数为5作为阈值进行剪枝，即过滤掉度数小于5的节点。据此计算剪枝后知识网络的拓扑结构指标，结果如表2所示。

从表2可知，精简后的语义功能增强的共词网络包含8 524个节点和14 420条边，网络直径为14，图密度为0.001，并且平均路径长度较长，而平均聚类系数为0.000 1，表明该网络规模较大且结构稀疏。这主要是因为区分词汇功能后，不同功能的词汇在语义关联上具有较强的特异性和分散性，使得词汇间的连接相对减少，难以形成紧密的聚类结构。节点平均度为3.336，表明网络中的每个研究问题词汇或研究方法词汇平均与其他3个词汇存在共现关系。节点平均加权度为3.547，表明每个研究问题词汇或研究方法词汇平均与其他词汇产生约3.5次共现关系，反映出该领域的知识共现强度较低。模块度为0.553，已达到理想阈值[32]，说明网络中存在明显的社区结构，适合进行社区发现。

表1　研究问题词汇和研究方法词汇的频次统计情况

频次	研究问题词汇		研究方法词汇	
	数量/个	占比/%	数量/个	占比/%
＞9	1 631	0.93	1 729	1.00
9	249	0.14	272	0.16
8	346	0.20	325	0.19
7	493	0.28	414	0.24
6	642	0.37	665	0.39
5	948	0.54	999	0.58
4	1 600	0.91	1 672	0.97
3	3 394	1.93	3 472	2.02
2	11 083	6.30	11 542	6.71
1	155 462	88.41	151 045	87.75

表2　词汇语义功能增强的共词网络结构指标

网络结构指标（单位）	数值
节点（个）	8 524
边（条）	14 420
平均度	3.336
平均加权度	3.547
网络直径（步）	14
图密度	0.001
平均路径长度（步）	5.372
平均聚类系数	0.000 1
模块度	0.553
社区数（个）	1 446

3.2.2　知识结构主题分析

在初始的语义功能增强共词网络中，共识别出1 446个社区。该网络规模庞大、结构复杂，可视化效果欠佳，不利于后续分析。为此，本研究根据节点度数对网络进行剪枝，保留度数大于90的节点，得到一个包

含100个节点和193条边的网络图。随后，提取该网络中的最大连通子图，最终得到90个节点和178条边，并将其作为后续分析对象。采用Gephi 0.10.1进行网络社区划分，并计算得到模块度为0.489，说明社区划分效果较好。

语义功能增强的共词网络被划分为9个知识社区，既梳理出计算机学科的整体知识结构，也细致展现了各子领域的研究问题与研究方法，有效提升了学科知识结构解析的完整性与精准度。

（1）知识社区1：人机交互主题。该社区中人机交互（human-computer interaction）、普适计算（ubiquitous computing）等研究问题主要通过设计模式（design pattern）、用户界面（user interface）等研究方法解决，反映出人机交互领域的研究在注重界面可用性与标准化的同时，也开始关注情境感知与隐式交互设计。

（2）知识社区2：教育与软件工程主题。计算机科学教育（computer science education）、软件工程（software engineering）是该社区的重点研究问题词汇，演化计算（evolutionary computation）、软件工程、社交媒体（social media）、物联网（Internet of Things）等为该社区的基本研究方法词汇。软件工程同时扮演教学内容与方法工具的双重角色，分类法（taxonomy）用于知识组织，演化计算代表算法教学，社交媒体、物联网等作为技术工具引入教育实践，呈现出学以致用的双向融合特征。

（3）知识社区3：系统开发主题。人工智能（artificial intelligence）、大数据（big data）、模型检测（model checking）、编程语言（java）是该社区基本研究方法词汇，嵌入式系统（embedded system）、信息物理系统（cyber-physical system）、面向方面程序设计（aspect-oriented program）等是该社区的研究问题词汇。嵌入式与信息物理系统的开发，正从单一的可靠性验证（如模型检测），向融合大数据分析与人工智能决策的综合方法体系演进。

（4）知识社区4：网络搜索主题。协同过滤（collaborative filtering）、图像检索（image retrieval）、人机交互（human-robot interaction）等研究问题与主动学习（active learning）、强化学习（reinforcement learning）、自然语言处理（natural language processing）等研究方法形成多重匹配关系，其中自然语言处理技术因契合语义搜索需求而成为该社区的核心方法，反映出方法选择与问题特性之间的匹配关系。

（5）知识社区5：推荐系统主题。推荐系统（recommendation system）、异常检测（anomaly detection）、多媒体应用（multimedia application）等研究问题词汇与深度学习（deep learning）、神经网络（neural network）、卷积神经网络（convolutional neural network）等研究方法词汇形成密切关联，表明深度学习已成为推荐系统领域的主流技术路线。

（6）知识社区6：移动设备与虚拟现实主题。研究问题主要包括移动电话（mobile phone）、增强现实（augmented reality）、社交网络（social network）及物联网（Internet of Things）；研究方法则涵盖增强现实、推荐系统（recommendation system）、计算机视觉（computer vision）、移动设备（mobile device）、眼动追踪（eye-tracking）及区块链（blockchain）等。值得注意的是，增强现实同时扮演研究问题与研究方法双重角色，表明其既作为应用场景被研究（如用户体验、交互设计），也作为一种视觉交互手段服务于其他问题。

（7）知识社区7：信息检索主题。研究问题主要包括信息检索（information retrieval）、软件开发（software development）、遗传编程（genetic programming）及活动识别（activity recognition）；研究方法涵盖机器学习（machine learning）、遗传编程、进化算法（evolutionary algorithm）、众包（crowdsourcing）及游戏化（gamification）等。机器学习、进化算法构成智能计算基础，众包、游戏化等体现社会化协作路径，遗传编程同时扮演研究问题与研究方法双重角色，GPU提供算力支撑，服务于软件开发、移动设备（mobile device）、活动识别等应用场景，呈现出“算法智能与人本计算深度融合”的特征。

（8）知识社区8：无线网络与云计算主题。无线网络（wireless sensor network）、虚拟环境（virtual environment）、Web应用（web application）、移动应用（mobile application）、云计算（cloud computing）等研究问题与云计算、聚类（clustering）、虚拟现实（virtual reality）、设计准则（design principle）等研究方法共同构成其知识基础。该社区呈现出“云–边–端协同”的核心特征：无线网络作为边缘感知与传输层，云计算作为集中处理与服务平台，两者相互依存，共同支撑虚拟环境、Web/移动应用等上层服务。

（9）知识社区9：入侵检测主题。在各知识社区

中规模最小，数据挖掘（data mining）是解决入侵检测（intrusion detection）、软件产品线（software product line）问题的基础方法，隐马尔科夫模型（hidden Markov model）作为一种经典的机器学习模型，也被广泛用于入侵检测，整体呈现出数据驱动的异常检测特征。

3.3 主题强度演化分析

在知识结构识别的基础上，本研究进一步分析各知识社区所代表的学科子领域的主题强度演化趋势。依据主题强度分析框架及其计算方法，借助Python编程实现各知识结构的整体主题强度，以及研究问题词汇、研究方法词汇的主题强度在不同时间窗口的强度计算，据此绘制主题强度演化曲线。

3.3.1 知识结构整体演化分析

知识结构主题强度整体演化过程可视化结果如图3所示，主题1~9分别对应上文识别出的9个知识社区所代表的主题。不同主题强度的演变趋势不同，总体可分为增长趋势、平稳趋势和衰败趋势3种。

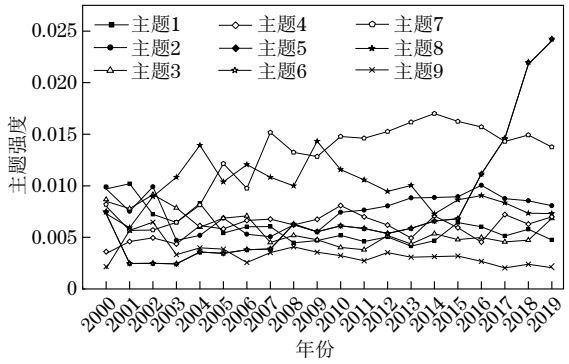


图3 知识结构整体演化趋势

（1）增长趋势：主题4、5、6、7总体呈增长趋势。主题4是网络搜索相关研究，在2000—2010年主题强度稳步增长，2010—2013年迅速下降，2014年前后自然语言处理领域进入神经网络时代，推动互联网搜索从关键词匹配向语义理解转变[33]，拉动该领域研究热度回暖并稳步上行，最终呈现增长趋势。主题5是推荐系统相关研究，主题6是移动设备与虚拟现实相关研究，均持续受到关注，2015年以后，深度学习在推荐系统[34]、移动智能应用[35]中得到广泛部署。主题7是信息检索相关研究，呈现先迅速增长后缓慢下降的发展趋势，新算法的持续引入为该领域带来了新的发展空间。

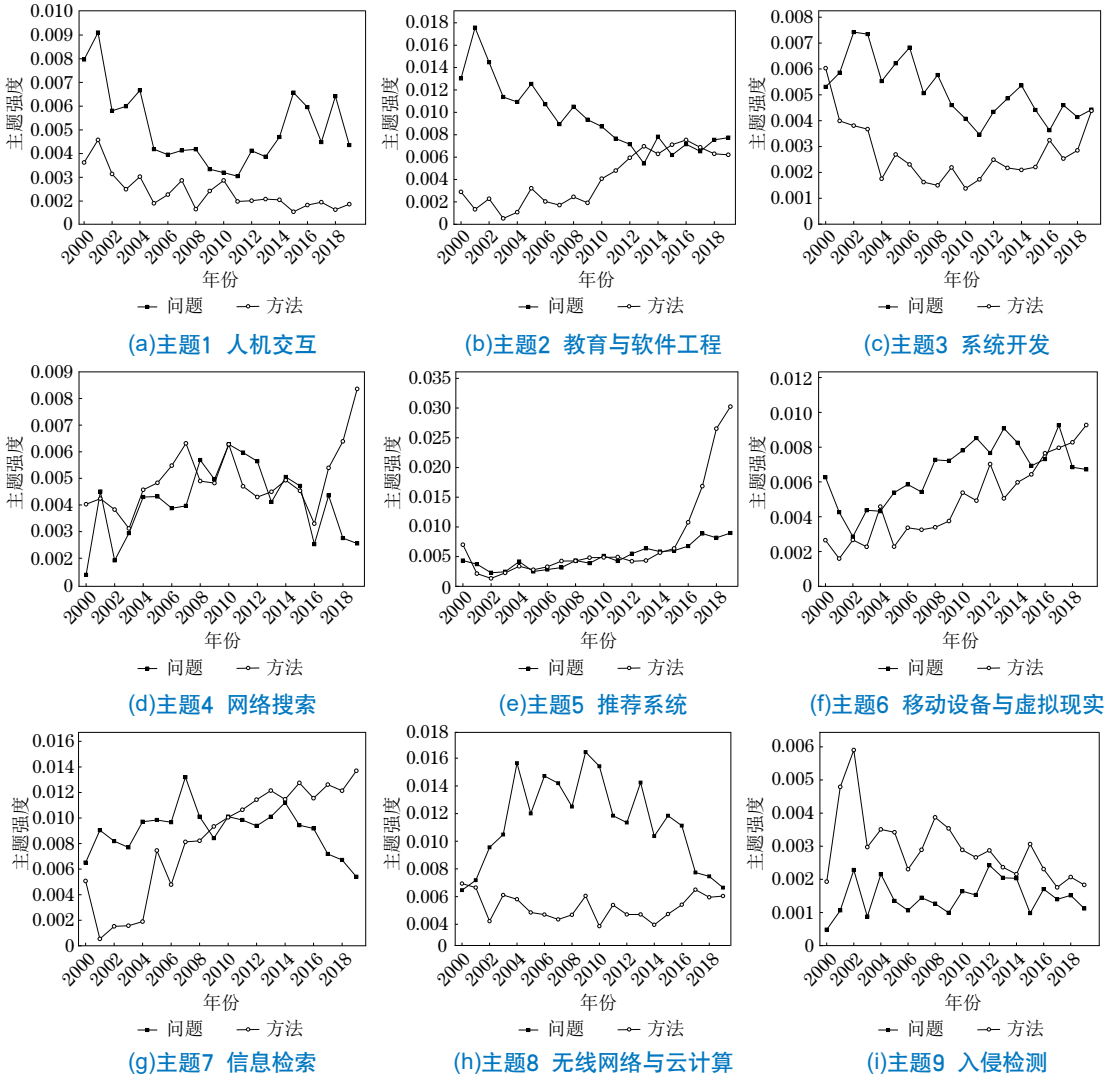
（2）平稳趋势：主题2、8、9趋势相对平稳。主题2是教育与软件工程相关研究，呈现先下降后上升的趋势，其中软计算方法（如演化计算）在高等教育中的应用研究从2010—2019年呈现显著增长，研究主题从理论向实践应用转变[36]。主题8是无线网络与云计算相关研究，呈现先上升后下降的趋势，无线网络、虚拟环境、云计算等是信息技术不断发展的产物，反映出该领域经历了从概念探索到技术成熟的过程，峰值之后相关研究不再作为独立热点集中出现。主题9是入侵检测相关研究，主题强度整体较低，仅在2001年、2002年出现短暂峰值，其他年份相差较小。隐马尔可夫模型作为关键性方法，在过去几十年中被成功用于识别各类恶意活动，基于该模型构建的入侵检测系统检测准确率超过94%[37]。但其高精度并未持续驱动主题热度，该方法在证明有效后迅速成为入侵检测领域的常规工具，相关研究从“方法创新”转向“工程应用”，不再作为独立主题热点被广泛讨论。

（3）衰败趋势：主题1、3处在发展过程中，主题强度起伏较大，但总体呈衰败趋势。人机交互领域的核心方法（如用户中心设计）已高度成熟并融入常规开发流程，相关研究不再作为独立热点集中发文，导致主题强度整体走低[38]。系统开发领域，系统之系统（SoS）的出现对传统软件工程方法提出了新的挑战，原有主题的研究热度被新主题分流[39]，反映出该领域从传统单一向复杂系统演进的趋势。

3.3.2 “问题–方法”二元演化分析

知识结构中“问题–方法”二元演化趋势如图4所示。根据二者的动态关联及演化特征，可将9个研究主题划分为3种类型（共同增长型、问题主导型、方法主导型），以揭示不同类型主题的研究重心差异。共同增长型指主题内问题强度与方法强度均呈上升趋势，且二者数值接近；问题主导型指问题强度长期高于方法强度；方法主导型指方法强度长期高于问题强度。

（1）共同增长型：主题6。其中的研究问题词汇和研究方法词汇的主题强度同步增长且数值趋近，从这一词汇关联特征可以反映出，该主题所属的研究领域在问题探究过程中，方法创新紧密跟进，二者形成良性互动，共同驱动研究发展，为该主题的持续活跃提供了有力支撑。



(a)主题1　人机交互　(b)主题2　教育与软件工程　(c)主题3　系统开发

(d)主题4　网络搜索　(e)主题5　推荐系统　(f)主题6　移动设备与虚拟现实

(g)主题7　信息检索　(h)主题8　无线网络与云计算　(i)主题9　入侵检测

图4　“问题–方法”二元演化趋势

（2）问题主导型：主题1、2、3、8。主题1中的研究问题词汇和研究方法词汇的主题强度均呈现下降趋势，表明人机交互技术正经历从“人机协同”向“人智协同”的范式转变，传统基于语音、手势、眼动等方法的交互框架正在被新型人智交互范式所取代[40]。主题2中，早期针对教育资源数字化、课程管理系统开发的需求催生大量相关研究问题，但随着慕课平台、分布式系统的普及，核心问题逐步明晰并趋于稳定，慕课已从早期以资源共享为特征的在线课程，逐步演化为融合人工智能、元宇宙等技术的智慧慕课新形态，技术赋能与问题导向共同驱动着在线教育领域的持续演进[41]。主题3则受软件开发工业化趋势驱动，推动问题强度因标准化流程建立而下降，方法强度因工程化需求增长而上升。主题8中问题强度长期高于方法强度，且呈现问题强度增长方法强度下降、问题强度下降方法强度增长的研究态势。2000—2010年，无线网络频谱分配等问题集中涌现，推动问题强度快速上升，随着5G、边缘计算、软件定义网络等技术的发展，研究重点从频谱分配转向系统集成与性能优化[42]。

（3）方法主导型：主题4、5、7、9。主题4的问题强度与方法强度曲线在2000—2015年呈现交错上升态势，反映出链接分析等方法创新与用户搜索需求的良性互动。2015年以后出现问题强度下降方法强度上升的趋势。一方面，传统基于关键词的搜索难以支持多步推理和探索式学习任务[43]；另一方面，深度学习驱动的语义解析持续迭代推高方法强度。2015年以后，主题5的方法强度随着深度学习的广泛应用显著超越问题强度，深度学习技术在推荐系统中的应用持续深化，但多数研究关注算法改进，可能导致方法创新与现实需求适配性下降[44]。主题7在2010年以前

问题强度高于方法强度，2010年以后方法强度高于问题强度，反映出传统基于关键词匹配的技术体系难以满足复杂检索需求。主题9呈现问题强度与方法强度同步波动的特征，2015年后针对边缘计算环境、物联网异构设备集群等高动态分布式场景的资源有限[45]，相关研究的发展速度相对滞后。

此外，为进一步揭示研究问题与研究方法的协同关系，本研究计算了各主题研究问题强度与研究方法强度的皮尔逊相关系数，反映二者在时序上的线性关联程度。相关系数为正且绝对值较大，表明研究问题与研究方法同步变化、协同发展；相关系数为负且绝对值较大，表明二者此消彼长、存在替代关系；相关系数接近0，则表明二者变化相对独立。各主题的相关系数计算结果如表3所示。

表3　各主题“问题–方法”相关系数

主题	相关系数
1	0.56
2	−0.77
3	0.25
4	0.09
5	0.83
6	0.60
7	−0.03
8	−0.53
9	0.21

从表3可以看出，主题5的相关系数为0.83，呈强正相关，表明尽管该知识社区方法强度显著超越问题强度，但方法创新与问题探索仍保持高度协同，并非脱离问题独立发展。主题2的相关系数为−0.77，呈强负相关，反映出该知识社区问题与方法存在此消彼长的替代关系，与问题主导型特征一致。主题6的相关系数为0.60，呈中等正相关，印证了其共同增长型特征。9个主题所体现的“问题–方法”强度结构差异与相关系数分析结果，在一定程度上揭示了各知识社区所处的发展阶段及其内部研究问题与研究方法的协同特征，为理解各主题内部研究问题与研究方法的协同程度提供了量化依据。

4　结语

本研究构建了词汇语义功能增强的共词网络，基于ACM数据集识别学科知识结构，并从整体演化和“问题–方法”二元演化两个视角探究主题强度演化过程。研究有效识别出计算机学科领域的9个知识社区，系统揭示了各社区内研究问题与研究方法的关联关系及其演变特征。本研究的特色和贡献在于：在理论层面，将词汇语义功能作为知识结构的基本分析单元，通过构建研究问题与研究方法两类异质性节点，揭示了二者动态匹配所形成的学科知识结构；在方法层面，构建了“问题–方法”共现网络，并引入皮尔逊相关系数量化二者的协同与消长关系，将“问题–方法”二元关系作为演化分析的核心对象，进一步揭示了研究问题与研究方法的匹配关系，为理解学科发展的内在动力机制提供了可量化的分析工具；在实践层面，通过对计算机领域文献的实证分析，识别出9个主要知识社区，揭示了各社区内部研究问题与研究方法的不同演化模式，可为科研管理提供决策参考。

本研究存在一定的局限性：首先，仅将问题词汇和方法词汇作为共现网络节点，尚未考虑其他语义功能的词汇；其次，演化分析主要依赖主题强度曲线与文献支撑进行归因推断，演化证据链仍有待加强；最后，实证分析采用的数据时间窗口为2000—2019年，未包含2019年以后的文献。未来将引入高频词时序变化、关键年份突变检测等量化分析方法，并考虑多种类型的词汇语义功能，进一步丰富演化分析的多维证据支撑，提升归因推断的严谨性。尽管数据窗口的选取不影响方法本身的通用性与有效性，但后续研究将把数据延伸至论文最新发表年份，以揭示大模型、生成式人工智能等新兴主题带来的知识结构变化。

参考文献

[1] 刘春江，刘自强，方曙. 基于SAO的技术主题创新演化路径识别及其可视化研究[J]. 情报学报，2023，42（2）：164−175.

[2] ZHAO X H，WANG X F，LIU Y Q，et al. Evolution of journal preference based on topic focus: a case study in the technology innovation management field[J]. Scientometrics，2025，130（4）：2137−2166.

[3] 苏婉，于森，禚传阳. 我国数据要素政策主题演化与识别分析[J]. 图书情报工作，2024，68（16）：90−103.

[4] 梁爽，刘小平. 基于文本挖掘的科技文献主题演化研究进展[J]. 图书情报工作，2022，66（13）：138−149.

[5] 王佳敏，陆伟，程齐凯，等. 基于细粒度关键词引用网络的领域知识多维分析[J]. 情报学报，2022，41（7）：733−744.

[6] 程齐凯，李鹏程，张国标，等. 学术文本词汇功能识别：基于标题生成策略和注意力机制的问题方法抽取[J]. 情报学报，2021，40（1）：43−52.

[7] 黄红，陈翀，张婧莹. 科技文献内容语义识别研究综述[J]. 情报学报，2022，41（9）：991−1002.

[8] WANG Y Z，ZHANG C Z，SONG M，et al. Exploring academic influence of algorithms by co-occurrence network based on full-text of academic papers[J]. Aslib Journal of Information Management，2025，77（4）：651−680.

[9] ZHANG H，ZHANG C Z，WANG Y Z. Revealing the technology development of natural language processing: a scientific entity-centric perspective[J]. Information Processing & Management，2024，61（1）：103574.

[10] WANG J M，CHENG Q K，LU W，et al. A term function-aware keyword citation network method for science mapping analysis[J]. Information Processing & Management，2023，60（4）：103405.

[11] CHENG Q K，WANG J M，LU W，et al. Keyword-citation-keyword network: a new perspective of discipline knowledge structure analysis[J]. Scientometrics，2020，124（3）：1923−1943.

[12] 刘非凡，张爽，罗双玲，等. 基于深度图神经网络方法的领域知识结构探测[J]. 情报学报，2021，40（11）：1209−1220.

[13] ZHANG X Y，XIE Q，SONG C，et al. Mining the evolutionary process of knowledge through multiple relationships between keywords[J]. Scientometrics，2022，127（4）：2023−2053.

[14] BA Z C，CAO Y J，MAO J，et al. A hierarchical approach to analyzing knowledge integration between two fields: a case study on medical informatics and computer science[J]. Scientometrics，2019，119（3）：1455−1486.

[15] 王伟，梁继文，杨建林. 基于引文网络的领域主题层次结构识别方法研究[J]. 图书情报工作，2022，66（17）：81−92.

[16] FAN Y L，LEHMANN S，BLOK A. Extracting the interdisciplinary specialty structures in social media data-based research: a clustering-based network approach[J]. Journal of Informetrics，2022，16（3）：101310.

[17] HOSSEINI E，TAGHIZADEH MILANI K，SABETNASAB M S. Development and maturity of co-word thematic clusters: the field of linked data[J]. Library Hi Tech，2023，43（1）：81−113.

[18] 胡泽文，韩雅蓉，王梦雅. 基于LDA−Word2vec的图书情报领域机器学习研究主题演化与热点主题识别[J]. 现代情报，2024，44（4）：154−167.

[19] WANG Z Y，CHEN J，CHEN J P，et al. Identifying interdisciplinary topics and their evolution based on BERTopic[J]. Scientometrics，2024，129（11）：7359−7384.

[20] 王佳敏，房子辰，窦永香. 引用功能感知的知识单元引用网络构建与多维分析[J]. 图书情报工作，2024，68（21）：133−144.

[21] 刘自强，王效岳，白如江. 多维度视角下学科主题演化可视化分析方法研究：以我国图书情报领域大数据研究为例[J]. 中国图书馆学报，2016，42（6）：67−84.

[22] LI Y T，CHEN Y，WANG Q Y. Evolution and diffusion of information literacy topics[J]. Scientometrics，2021，126（5）：4195−4224.

[23] ZHU H M，LI Q，WANG Q，et al. Evolution analysis of online topics based on “word-topic” coupling network[J]. Scientometrics，2022，127（7）：3767−3792.

[24] 刘晋霞，侯倩倩，杜静，等. 子话题及词汇关联视角下的新兴领域热点主题演化研究[J]. 情报杂志，2023，42（3）：123−129.

[25] GAO Q，HUANG X，DONG K，et al. Semantic-enhanced topic evolution analysis: a combination of the dynamic topic model and word2vec[J]. Scientometrics，2022，127（3）：1543−1563.

[26] 陈果，彭家彬，肖璐. 基于“问题−方法”知识抽取的科研领域知识演化研究：以人工智能为例[J]. 情报理论与实践，2022，45（6）：32−38.

[27] 俞琰，郑斯煜，葛萌. 基于关键词语义功能的领域主题演化分析[J]. 情报杂志，2025，44（1）：187−197.

[28] 程齐凯，王晓光. 一种基于共词网络社区的科研主题演化分析框架[J]. 图书情报工作，2013，57（8）：91−96.

[29] BLONDEL V D，GUILLAUME J L，LAMBIOTTE R，et al. Fast unfolding of communities in large networks[J]. Journal of Statistical Mechanics：Theory and Experiment，2008（10）：P10008.

[30] WANG X G，CHENG Q K，LU W. Analyzing evolution of research topics with NEViewer: a new method based on dynamic co-word networks[J]. Scientometrics，2014，101（2）：1253−1271.

[31] 陈翔，黄璐，倪兴兴，等. 基于动态语义网络分析的主题演化路径识别研究[J]. 情报学报，2021，40（5）：500−512.

[32] NEWMAN M. Fast algorithm for detecting community structure in networks[J]. Physical Review E-Statistical，Nonlinear，and Soft Matter Physics，2004，69（6）：066133.

[33] WILDHABER B，GÖLDA A，RIETSCHE R. Search-augmented large language models: Transforming internet search: a systematic literature review[C]//Proceedings of the Thirty-Third European Conference on Information Systems（ECIS 2025）. Atlanta：AIS eLibrary，2025：1750.

[34] ALAM M M，AHMED M. Deep learning-based recommendation systems: Review and critical analysis[C]//Proceedings of International Conference on Data Analytics & Management. Singapore：Springer Singapore，2023：39–55.

[35] WANG Y C，WANG J Y，ZHANG W Z，et al. A survey on deploying mobile deep learning applications: a systemic and technical perspective[J]. Digital Communications and Networks，2022，8（1）：1–17.

[36] VERMA A. Bibliometric analysis of soft computing’s evolution in higher education: Comprehensive insights and future directions[J]. Journal of Educators Online，2025，22（4）：n4.

[37] NIKOLOVA E. Markov models for malware and intrusion detection: a survey[J]. Serdica Journal of Computing，2021，15（2）：129–147.

[38] SAYAGO S，SENDÍN M，TEIXIDÓ M. Future perspectives on human-computer interaction research: towards the year 2030[M]. Cham：Springer，2025.

[39] CAVALCANTE E，BATISTA T，OQUENDO F. Looking back and forward: a retrospective and future directions on software engineering for systems-of-systems[J]. Journal of Software：Evolution and Process，2024，36（10）：e2697.

[40] 王镇远，田东，董禹，等. 多模态交互：从人机协同迈向人智协同[J]. 数据与计算发展前沿，2025，7（3）：81–93.

[41] 徐晓飞，张龙，张策，等. 智慧教育与智慧慕课的新形态、新模式与新行动[J]. 中国教育信息化，2025，31（12）：3–16.

[42] KUMAR K，NAYAK M，KUMAR K J，et al. Wireless networks and cloud: a new era of connectivity[C]// Proceedings of 2025 International Conference on Computing Technologies & Data Communication（ICCTDC）. Piscataway：IEEE，2025：1–6.

[43] GUAN Z T，WANG Y. Evolving paradigms in task-based search and learning: a comparative analysis of traditional search engine with LLM-enhanced conversational search system[EB/OL]. [2026–01–25]. https://arxiv.org/abs/2512.00313.

[44] ASRI B，QASSIMI S，RAKRAK S. Adaptive personalized recommendation systems: a systematic review[J]. Information Systems，2026，135：102594.

[45] MAAD M A，HAMOUSH M，ABD EL SALAM A，et al. A lightweight deep learning-based intrusion detection framework for IoT edge networks: Evaluation on the AWID dataset[C]// Proceedings of 2025 IEEE International Conference on Emerging Trends in Engineering and Computing（ETECOM）. Piscataway：IEEE，2025：1–6.

作者简介

王佳敏，男，博士，副教授，硕士生导师，研究方向：智能情报与知识服务、数据挖掘与知识发现，E-mail：wangjiamin@xidian.edu.cn。
房子辰，女，硕士研究生，研究方向：知识组织与知识发现。
崔蕴学，男，博士，助理研究员，研究方向：信息资源组织与检索。

Integrating Term Function for Identification and Evolutionary Analysis of Discipline Knowledge Structure

WANG JiaMin FANG ZiChen CUI YunXue
(School of Economics and Management, Xidian University, Xi’an 710126, P. R. China)

Abstract: Identifying and analyzing the knowledge structure and its evolution in a discipline helps to grasp the patterns of development and change in a field, thereby promoting and guiding its progress. Traditional co-word network-based studies typically treat keywords as homogeneous nodes, leading to coarse analytical granularity and low semantic depth, while also failing to capture the co-evolutionary relationships between functional terms, such as research questions and methods. To address these limitations, this study enhances the semantic information of knowledge network nodes through term function to improve the granularity of knowledge structure identification and evolution analysis. First, the BERT model is employed to identify research questions and methods in academic literature, based on which a heterogeneous co-occurrence network incorporating term functions is constructed, and a community detection algorithm is applied to identify the disciplinary knowledge structure. Second, topic strength quantification is conducted from two dimensions: the overall evolution of the knowledge structure and the binary evolution of question-method pairs, with an empirical study using the ACM (Association for Computing Machinery) dataset. Experimental results show that the proposed method effectively identifies nine knowledge communities in the field of computer science, systematically revealing the research questions, methods, and their association patterns within each community. Furthermore, the evolution analysis incorporating binary attributes captures the dynamic shifts in research foci and the intrinsic question-method association logic. This study enhances the granularity and interpretability of knowledge structure analysis, providing a new perspective and methodology for disciplinary knowledge structure mining.

Keywords: Term Function; Co-occurrence Network; Question-Method; Knowledge Structure; Evolution Analysis; Computer Science

（责任编辑：王玮）