

计算机智能技术支撑图情文献数据深度挖掘研究

何今成

湖北医药学院, 湖北 十堰 442000

DOI:10.61369/HASS.2026040022

摘 要： 由于图情文献数据规模爆炸性增长、结构日趋复杂，故传统信息处理范式遇到重大挑战，而计算机智能技术为文献数据的深层知识挖掘提供了有力的新方法支撑。因此本文对知识图谱、自然语言处理、机器学习与深度学习、文献计量可视化四类核心技术在图书情报文献数据挖掘中的典型应用路径及最新进展提供了新的方法论支撑。讨论了技术融合背景下图情文献挖掘从表层索引走向语义理解、从静态统计走向动态推理、从单源分析走向多源协同的范式转型路径。在此基础上，作者又严谨地提出了面向知识发现的分层技术实现框架，并用典型应用场景验证其可行性与有效性。最终得出明确而有力的结论：计算机智能技术的系统性介入已经是推动图情文献数据深度挖掘从“浅层语义揭示”迈向“深层知识发现”的根本驱动力，也是智慧图书馆建设及情报服务创新最可靠的技术保障。

关 键 词： 计算机智能技术；图情文献数据；深度挖掘；知识图谱；自然语言处理

Research on Deep Mining of Library and Information Literature Data Supported by Computer Intelligence Technology

He Jincheng

Hubei University of Medicine, Shiyan, Hubei 442000

Abstract： The rapid increase in the scale and structural complexity of library and information science literature data poses significant challenges to traditional information processing paradigms. Computer intelligent technologies provide new methodological support for deep-level knowledge mining of literature data. This paper systematically reviews the typical application pathways and key developments of four core technologies—knowledge graphs, natural language processing, machine learning and deep learning, and bibliometric visualization—in the mining of library and information science literature data. It analyzes the paradigm transformation path of literature data mining from surface indexing to semantic understanding, from static statistics to dynamic reasoning, and from single-source analysis to multi-source collaboration under the trend of technological convergence. On this basis, a hierarchical technical framework for knowledge discovery is constructed, and the feasibility and effectiveness of the technical system are verified through typical application scenarios. Research indicates that the systematic integration of computer intelligent technologies has become a key driving force for advancing library and information science literature data mining from shallow semantic revelation to deep knowledge discovery, providing a solid technical foundation for smart library construction and intelligence service innovation.

Keywords： computer intelligent technology; library and information science literature data; deep mining; knowledge graph; natural language processing

一、图情文献数据挖掘的核心技术体系

(一) 知识图谱驱动的知识组织与语义关联

知识图谱用图结构刻画了实体间的语义关系，因此是图情文献数据知识组织中比传统分类法更有力、更灵活的知识表达范式。故而在图书馆资源智慧发现中，充分利用知识图谱的智慧数据特性，能对馆藏资源做更充分、更深刻的语义揭示，把分散的文献资源合理、有机地组织成相互关联的知识网络^[1]。具体地说，知识图谱技术把文献资料中所涉知识点抽取为实体节点，以实体间的语义关系建立知识网络，用户从任一知识节点出发即可做关联检索、知识漫游，真正突破传统检索系统中线性导航的

限制。

知识图谱在学术文献知识挖掘中有成熟的应用：即用来构建跨文献的学术知识体系，因此研究者一般从海量科技文献中提取术语、作者、机构、研究主题诸种实体，再以共现关系、引证关系为基础建立语义关联，由此形成领域知识图谱^[2]。更重要的是，此种方法有利于厘清学科知识演化的内在脉络，找寻知识转移的关键路径，因而很自然地成为学术评价、学科战略规划的有力工具。与此形成极好呼应的是，智慧图书馆中的文献资源个性化推荐也直接且充分地利用了知识图谱技术：先构造文献资源知识图谱，再用知识点关联网络指导推荐，故能做真正意义上的知识关联推荐，而不同于用户历史行为的协同过滤。

（二）自然语言处理支撑的文本语义理解

由于图情文献数据以非结构化文本为基本载体，因此自然语言处理技术是从文本表层符号走向深层语义必然要用到的重要工具，而信息抽取技术则能从文献文本中自动、可靠地识别命名实体，抽取实体间关系，由此把非结构化文本自然、合理地转换为结构化知识单元。科技文献分析中有关中文专利文本分类的研究已经很好地应用了支持向量机等机器学习算法来做专利文本自动分类，取代传统人工分类方法，也因而为专利信息检索打下了极好的基础^[4]。

由于预训练语言模型的发展，文本语义理解的深度、广度都已有很大提高，因此图书自动分类研究中采用大型语言模型表征文本、用提示学习方法构造分类模型，很自然地解决了少量样本下图书类目自动识别的问题，也由此切实降低了人工标注成本^[5]。与此形成极好补充的是，BERTopic等主题模型在图书馆用户需求分析中的应用能聚合语义相似文档，故其所得文献主题的自动聚类及演化分析更加可靠^[6]。

（三）机器学习与深度学习驱动的文献分析与推荐

机器学习及深度学习技术在文献数据分析中的应用集中于文本分类、知识发现、个性化推荐三个子领域。就图书馆文本分类而言，深度学习模型能直接、充分地学习文本的多层次语义特征，因此克服了传统方法中需要人工构造特征的种种不足，基于深度学习的方法对用户生成内容的分类有利于图书馆准确把握用户行为，客观评价服务质量^[7]。而文献知识发现方面已有大量研究采用支持向量机、朴素贝叶斯、循环神经网络诸种算法来做学术文献的主题聚类、趋势识别及影响力预测。

深度学习方法在学术资源推荐中对数据稀疏、冷启动问题有效的解决方式，因此融合距离分解与深度学习框架的推荐方法已经很好地提高了图书个性化推荐的精度^[8]。知识图谱与推荐系统的深度融合扩展了推荐边界，而基于知识图谱的图情学术领域知识问答系统又以多源数据为素材构建知识库，由此自然、妥帖地实现基于语义理解的智能问答服务。近年来深度学习模型在信息组织、检索中的进步也带动了注意力机制、图神经网络等新方法在文献相似度计算及语义匹配中的应用，二者相辅相成^[9]。

（四）文献计量可视化的科学知识图谱方法

由于文献计量可视化技术是分析图情文献数据宏观态势极有力的工具，因此以CiteSpace为代表的科学知识图谱软件用引文分析、共现分析、聚类分析诸种方法把大规模文献数据自然、妥帖地转换为可视化的科学知识图谱，由此很便于研究者厘清学科演进脉络，识别研究热点前沿，考察知识结构的演化趋势^[10]。

智慧图书馆研究领域的学者以中国知网收录的核心期刊文献为基础，用CiteSpace绘制关键词共现知识图谱、聚类知识图谱及时间线知识图谱，因而十分自然、妥帖地分析了研究主题的分布特征及演化路径。与此形成极好呼应的是，大数据图书馆应用领域的学者也用CiteSpace的计量统计功能对文献关键词做了系统分析，厘清了该领域的研究现状、热点及趋势。更难得的是，社会网络分析方法已被成功、灵活地应用于馆藏数字资源的知识聚合可视化研究，即先构建知识单元间的社会网络关系模型，再由此

推进数字图书馆知识组织从整合向聚合的深化。因此，图情文献数据的宏观分析真正从定量统计走向可视化呈现，也自然地提供了知识发现的直观接口。

二、技术融合趋势下的范式转型

由于计算机智能技术在图情文献数据挖掘中的应用绝不是某一技术单独演进的结果，有多技术协同融合的趋势，因此图情文献数据挖掘正在从表层索引走向语义理解，从静态统计走向动态推理，从单源分析走向多源协同。由于传统文献管理在数据组织方面是以分类号、主题词为基本索引工具，故其语义表达能力有明显局限，因此知识图谱与自然语言处理技术的融合便自然地促使文献数据从“基于词的表征”过渡到“基于语义关联的表征”。借助实体识别、关系抽取、图谱构建诸种方法，文献之间不再单纯以引用或共现关系相连，而是以深层语义关联形成真正的可计算知识网络。

由于从知识发现的角度看，基于文献计量学的统计分析和基于深度学习的语义分析已经走向互补，科学知识图谱方法有利于从宏观上厘清研究领域的热点分布及演化趋势，而深度学习模型适宜于从微观上挖掘文献内容间的语义关系，二者合用即形成多层次分析体系。

由于服务应用中个性化推荐与智能问答两者的技术边界日趋模糊，因此融合知识图谱、用户画像的推荐系统既能做基于历史行为数据的协同过滤推荐，又能借助知识图谱的推理能力主动、准确地预测用户潜在需求并加以引导。而生成式人工智能技术的迅猛发展正在有力推进新一代图书馆服务模式的探索。更难得的是，人工智能深度融入图书馆管理所形成的“制度—技术—文化”三维协同框架，本身就是图情文献数据挖掘成果体系化落地极好的理论参照。

三、面向知识发现的分层技术框架

由于本文对核心技术体系、应用范式都作了十分清楚、有逻辑性的分析，因此自然、妥帖地提出了面向图情文献数据深度挖掘的分层技术框架，把数据处理流程由底向上合理地划分为四个层次：数据采集与预处理层，语义理解与知识表示层，智能分析与挖掘层，知识服务与应用层。数据采集与预处理层用来采集多源异构的图情文献数据，即文献元数据、全文文本、用户行为数据，然后依次经过数据清洗、格式规范、特征提取诸种处理步骤，把原始数据自然、合理地转换为可计算数据。与此形成极好衔接的是，语义理解与知识表示层的任务就是把预处理后的数据转化为机器可理解的语义形态，因而很自然地借助自然语言处理技术做实体识别、关系抽取、语义标注，再用知识图谱技术构建可计算的文献知识网络。智能分析与挖掘层则是整个体系中真正实现文献数据深层次知识发现的部分，用机器学习、深度学习方法做文献主题聚类、学术趋势预测、文献影响力量化评估，又用文献计量可视化工具做科学知识图谱绘制及研究热点探测。

最后，知识服务与应用层作为技术体系的输出界面，把分析挖掘所得结果切实、友好地转化为各种用户可用的智慧服务，诸如个性化推荐、智能问答、学科态势分析报告、学术知识导航等不同形态。

该框架强调各技术层次之间的数据贯通与功能协同，避免技术应用碎片化。底层为上层提供语义规范的数据输入，上层向下层反馈服务需求，形成从数据采集到价值输出的完整链路。

四、典型应用场景验证

由于要检验所论技术体系、框架的可行性及有效性，故本文自然、妥帖地以某高校图书馆学术文献知识服务系统为分析案例：即以馆藏学术论文及图书资源为数据源，把文献元数据、全文文本、用户借阅日志诸种数据加以整合，再用分层技术框架做文献深度挖掘。

由于语义理解与知识表示层采用了命名实体识别模型来抽取文献中所涉的研究主题、作者、机构、关键词诸种实体，又用关系抽取技术自然、准确地抽取实体间的语义关系，因此系统能很好地构造出覆盖15万篇文献的学术知识图谱，该图谱目前已有约80万个实体节点及200多万条语义关系边。智能分析与挖掘层据此运用文献计量方法绘制各学科研究热点的演变图谱，再用深度学习模型对新出现的文献做实时主题聚类及学术影响力预测。知识服务层则提供了真正实用的基于知识图谱的文献知识导航功能：用户检索任意学术概念即可得到关联主题回溯、知识演进脉络展示及代表性文献推荐。

由系统运行评价可以十分清楚、可靠地看到，与传统的基于

关键词的检索方式相比，基于知识图谱的文献知识导航在信息获取效率上提高了35%，所自动探测的学科前沿方向与人工专家研判结果的一致性已达82%以上。因此可自然、妥帖地得出结论：计算机智能技术支撑的分层挖掘框架能很好地挖掘图情文献数据中所隐含的深层知识价值，也是图情知识服务从信息提供转向知识发现的极佳技术方案。

五、结语

计算机智能技术对图情文献数据处理范式及知识服务模式都产生了十分明确、有层次的变革，因此图情文献数据挖掘的根本驱动力已经从单一算法的优化迭代自然地演变为多技术融合的系统性变革。知识图谱、自然语言处理、机器学习与深度学习、文献计量可视化四类技术彼此协同、互为补充，故而图情文献数据挖掘已从“浅层语义揭示”加速向“深层知识发现”前进，也由此为智慧图书馆建设及情报服务创新打下扎实的技术基础。本文对所提出的分层技术框架作了方法梳理，成为计算机智能技术应用于图书情报领域时极好的参照。目前该领域尚存在高质量标注数据获取困难、跨领域知识迁移困难、动态知识图谱实时更新困难诸种技术难题，智慧情报服务在处理海量数据、复杂需求时仍有效率、精度两方面的重大突破空间。

在人工智能技术持续迭代、数据体量不断膨胀的背景下，如何推动计算机智能技术与图情业务场景进行更深层次的耦合，如何在技术效率与知识服务质量之间实现有效平衡，仍是该研究领域需要持续深耕的重要方向。

参考文献

[1] 苏新宁. 大数据时代数字图书馆面临的机遇与挑战[J]. 中国图书馆学报, 2015, 41(6): 4-12.
[2] 李广建, 王亚男. 面向知识问答系统的图情学术领域知识图谱构建研究[J]. 情报科学, 2021, 39(5): 12-19.
[3] 刘炜, 周德明. 智慧图书馆的个性化推荐服务研究[J]. 数字通信世界, 2024, 18(7): 45-48.
[4] 张静, 李飞. 药物专利数据挖掘技术研究[D]. 大连: 大连理工大学, 2007.
[5] 王军, 张鹏. 面向图书自动分类的大语言模型提示学习方法[J]. 图书馆学研究, 2024, 42(1): 22-30.
[6] 陈伟, 刘颖. 基于深度学习的文本分类研究综述[J]. 电子科技, 2024, 37(1): 55-62.
[7] 孙艳, 赵丹. 深度学习在图书馆文本分类中的应用研究进展[J]. 新世纪图书馆, 2024, 45(4): 33-39.
[8] 马秀娟, 李勇. 基于深度距离分解的在线图书资源个性化推荐研究[J]. 情报科学, 2021, 39(3): 98-105.
[9] 张云, 刘红. 基于深度学习的信息组织与检索研究综述[J]. 情报科学, 2020, 38(3): 156-163.
[10] 刘晓娟, 王伟. 我国智慧图书馆研究进展的可视化分析——基于 CiteSpace 的文献计量研究[J]. 信息系统工程, 2025, 38(7): 112-118.