面向 AI4Science 的科学论文图像语义描述框架体系构建研究*

张逸勤，邓三鸿，巩洪村，杨　杰，刘　浏

摘　要　科学论文图像的语义描述与标注是提升文献知识挖掘和智能信息管理能力的关键。本文基于学术图像的元数据结构和信息需求,提出了一个多层次的科学论文图像语义描述框架(SDF-SLI),旨在实现 AI4Science 背景下科学文献图像内容的精确语义解析。该框架通过基础标识层、内容层、语义层及关系层 4 个相互关联的层级,分别从图像的基本信息、视觉内容、语义内涵和层次关联 4 个维度解析科学论文图像内容。本文同时构建了全面的本体模型,系统地映射了不同框架层之间的语义关系,利用多模态大型语言模型,在图书情报领域论文进行了实证案例验证,证明了该框架的实用性与有效性。本文讨论了科学论文图像语义描述的关键挑战与未来机遇,为科学知识管理的智能化发展提供了参考。

关键词　科学论文图像，语义描述框架，本体模型，知识服务

DOI: 10.13663/j.cnki.lj.2026.05.002

Research on the Construction of a Semantic Description Framework for Scientific Document Images Oriented toward AI4Science

Zhang Yiqin, Deng Sanhong, Gong Hongcun, Yang Jie, Liu Liu

Abstract　The semantic description and annotation of scientific document images are essential for enhancing knowledge mining and intelligent information management of academic literature. Based on metadata structures and information requirements of academic images, this paper proposes a multi-level semantic description framework for scientific paper images (SDF-SLI), aimed at achieving precise semantic parsing of scientific document image content in the context of AI4Science. The framework analyzes scientific document images through four interconnected layers—basic identification layer, content layer, semantic layer, and relationship layer—which examine basic information, visual content, semantic connotation, and hierarchical relationships. This research also establishes a comprehensive ontology model that systematically maps semantic relationships between these framework layers, utilizing multimodal large language models for implementation. Through empirical case studies in library and information science literature, the framework demonstrates its practicality and effectiveness. The paper further discusses key challenges and future opportunities in semantic description of scientific paper images, offering insights into the intelligent development of scientific knowledge management.

Keywords　Scientific document images, Semantic description framework, Ontology model, Knowledge services

0　引言

科学论文作为学术研究成果的重要表现形式之一,已逐渐成为科研人员获取前沿知识、进行学术交流和创新活动的重要载体[1]。相较于

* 本文系国家自然科学基金青年学生基础研究项目(博士研究生)“融合创新测度和生成模型的科学技术知识空间预测研究”(项目编号:724B2015)的研究成果之一。
通信作者:邓三鸿, E-mail: sanhong@nju.edu.cn

单一模态的文本内容,科学论文所包含的图像在复杂信息的传达上具有形象、简洁且高效的特点[2],可以清晰地展示数据关系、技术流程、实验结果等重要内容[3]。随着面向科技创新的知识服务逐渐向多源异构[4]、内容组织细粒度[5]、跨模态语义关联[6]、深度挖掘分析与智能化应用[7]的方向发展,传统以科学论文文本为核心的知识发现模式正面临异构数据载体和新型服务需求的挑战。在此研究背景下,科学论文图像作为重要的异构数据形式,如何全面深入地准确解析图像所承载的丰富信息,已成为提升科技文献知识发现效率和质量的关键问题。

对图像内容进行有效的语义描述与表示,是弥补科学论文中图像在知识发现、科学研究和技术传播中应用不足的重要手段。现有的图像语义标注与描述框架及其衍生模型主要应用于艺术、历史、建筑设计、地理、生物医学和医疗健康等领域[8-10],为这些领域提供了包括底层视觉特征和语义特征在内的图像内容描述方法。然而,科学论文中图像具有高度结构化和专业化的特点,通常包含复杂的数据关系、技术流程和符号表达,并且与文献文本内容紧密相关[11]。这些特性使得现有的描述框架难以适用于科学论文图像的处理,无法充分捕捉其中的专业语义及其与文本的跨模态关联。现有方法在科学论文图像的细粒度检索、深度语义组织和多模态知识发现中表现有限,未能有效发挥这些图像的知识潜力[12]165-170。因此,构建一个适用于科技文献图像的语义描述框架,以提升其在知识检索、语义分析和多模态关联中的表现,已成为亟待解决的研究课题。

基于上述研究背景,本文梳理了科学论文图像的语义描述需求及其在知识发现和信息检索中的应用场景,提出了科学论文图像语义描述框架(Semantic Description Framework for Scientific Literature Images, SDF-SLI),旨在实现科学论文图像与文本的深度语义关联。本文通过应用实例验证了该框架在科学论文图像语义描述中的有效性,讨论了科学论文图像知识组织与管理中关键技术和主要挑战,以期为实现科学论文图像内容自动标注、多模态知识库构建及智能化科学信息管理提供参考和支持。

1 相关研究

1.1 图像内容的语义表示与描述

图像内容的语义表示与描述是人工智能领域中计算机视觉(Computer Vision)与自然语言处理(Natural Language Processing)交叉的前沿方向之一。对图像内容组件进行语义标注与分类识别,有助于弥补图像视觉特征与文本特征间存在的“语义鸿沟”,从而实现科技文献的跨模态知识组织与更高效的资源聚合[13]。

现有图像语义标注模型在图像检索和自动标注方面已取得较大进展,形成了多个具有代表性的层次化模型。在图像内容描述方面,于永新[14]构建了包含图像特征层、实体层、实体关系层和图像语义层的四层图像语义模型;王晓光等[15]将术语词表和图像元数据融合至“情感、行为/活动、场景、对象空间、对象”等高层语义表达中,构建了敦煌壁画数字图像语义描述层次模型;雷珏莹等[16]围绕图像的解读与意义建构选择叙事结构和逻辑,提出了图像数字叙事的基本框架;齐小英等[17]从知识描述与知识服务的用户需求视角出发,建立了历史地图知识与语义组织框架。这些模型通过分层描述图像的底层特征(如颜色、纹理和形状)和高层语义特征(如场景、行为和情感)提升了图像语义表达的深度和准确性。

然而现有模型多用于自然图像资源的语义标注,未能充分满足科技文献图像的专业性、结构化和多模态信息整合需求[18]。科学论文图像具有以下3个显著特征:专业化的学科领域知识整合、结构化的图像表现形式以及多维度信息的高度融合[19]。这些特性使其语义标注需求超出通用图像标注模型的适用范围。因此,构建针对科技文献图像的语义标注描述框架,结合领域知识和多模态信息进行精确的语义解析,对实现科技文献知识的深度挖掘与多维检索具有重要意义。

1.2 面向AI4Science的图像语义研究

随着科研活动日益数据密集化和智能化,AI4Science正在重塑科技文献挖掘的研究范式[20]。传统以文本为中心的知识服务模式已难以满足跨模态、细粒度的科研需求。现有研究主要聚焦于文本层面的语义解析与信息提

取，如王晓光等[21]以功能单元理论为基础，验证了科学论文功能单元本体 FUO 在揭示论文的情报结构和功能特征上的可用性；张敏等[22]构建了包含文献、句子、知识对象、知识关系在内的多维语义索引，实现了文献内部细粒度知识的语义检索；石栖等[23]围绕科学实验对科学问题、原理、实验步骤等相关知识对象进行表征与建模，形成学科领域科学实验知识图谱（Scientific Experiment Knowledge Graph, SEKG）；沈思等[24]960-970基于宏观、中观和微观3个维度设计了学术全文本知识图谱概念框架，筛选出9种关键知识元进行多维知识元的学术图谱构建。这些工作为科技文献的智能分析奠定了重要基础，但在多模态融合方面仍有待深化。伴随视觉语言大模型（Vision-Language Models, VLMs）和多模态大模型的快速发展，科学论文图像作为承载专业知识的重要载体，其语义理解与知识表示的研究价值日益凸显[25]。

然而对于科学论文中的图像语义研究尚存在显著不足，现有智能知识服务系统主要依赖于文本内容进行信息组织和检索[26]，而对科学论文图像的深层语义理解与挖掘未给予足够重视。这种局限性导致图像中的信息未能充分融入知识服务体系，使得图像在实现细粒度检索、跨模态知识发现和深度语义组织方面存在挑战[27]。基于此，在 AI4Science 范式下，构建融合视觉语言大模型的科学论文图像语义描述框架，不仅能够提升图像语义理解的准确度，还将为跨模态知识图谱构建、智能检索系统开发和知识服务创新提供新的方法论支撑，对推动科技文献多模态知识服务体系的演进具有重要的理论意义和实践价值。

2　科学论文图像语义描述框架的层次结构分析

科学论文图像的语义描述是提高科技文献信息可读性、可检索性和自动化处理能力的重要环节。本文提出了基于层次结构的 SDF-SLI，该框架通过4个相互关联的层次实现对图像的系统化描述，如图1所示。基础标识层负责提供图像的规范化标识信息以支撑上层分析；内容层基于标识信息进行结构化解析，为语义理解提供客观依据；语义层在此基础上进行深层解读，揭示科学发现；关系层则通过建立多维关联网络，实现知识的系统化组织。各层次间通过信息流动与反馈形成有机整体，构建起

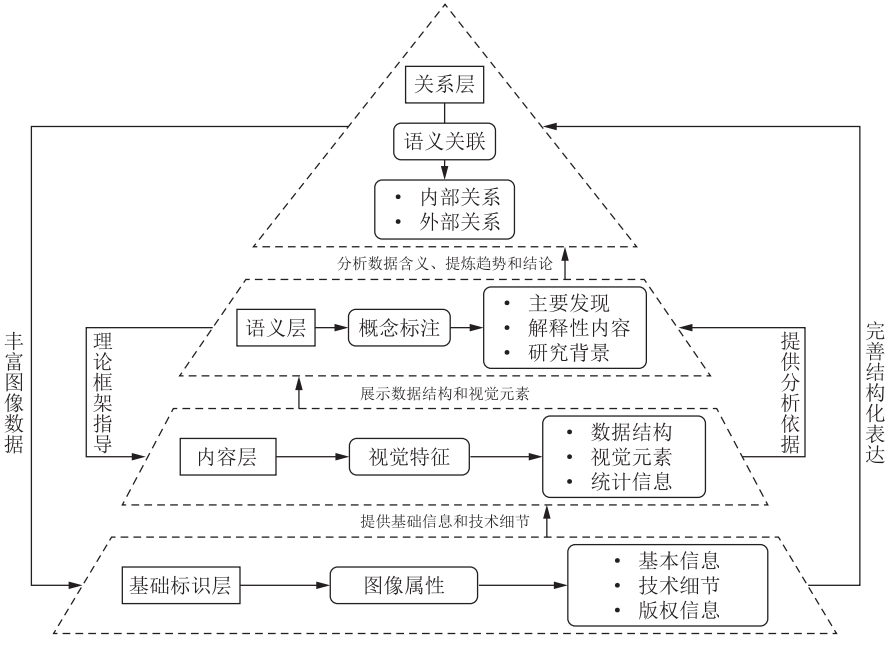


图1　科学论文图像语义层次框架

从表层特征到深层语义的完整描述链条。每一层次聚焦图像不同方面的信息,以层次递进的方式实现对图像语义的全面描述。

2.1 基础标识层

在科学论文图像语义描述框架中,基础标识层主要关注图像的基本属性和外部描述,为后续层次分析提供必要的背景信息和支持。该层分为以下3部分:

(1)基本信息是图像的核心标识性数据,包括图表标题、作者信息、创建日期及图表标识符等。这些数据作为图像的直接附加说明,不仅帮助读者快速理解图像的背景来源,还在语义描述框架中为图像与文献上下文的关联奠定了基础。

(2)技术细节部分关注图像的制作和展示方式,这包括图表类型、使用的软件或工具、图像格式、分辨率以及文件大小等数据[28]。这些信息决定了图像的视觉呈现形式,为图像解读的策略提供参考[29]。不同图表类型在信息传达和语义描述侧重点上各有差异,如柱状图通常用于数据对比、折线图则适于表示数据变化趋势。图像的生成工具在一定程度上反映了其技术背景,包含实验设计、数据分析以及结论呈现的方法选择路径。

(3)版权信息部分确保图像的合规使用,涵盖了许可证类型、使用限制、来源文献和DOI标识符等信息[30]。这些信息为图像的合法使用提供保障,增强图像可追溯性。版权信息约束图像使用范围,确保其在学术传播中的合规性;而来源文献和DOI标识符则为进一步查阅提供了有效路径,保障图像在二次利用和学术共享中的合法性。

2.2 内容层

内容层是解析图像实际数据和视觉元素的核心,细化为3个关键组成部分:

(1)数据结构揭示图像中的变量定义、数据类型、数据范围和单位等基本信息[31]。变量定义与单位直接决定了数据点的意义,确保描述者理解图像表达的信息。在复合科学图表中,不同变量的定义及单位明确了每个数据点的实际含义。数据预处理方法(如归一化处理或缺失值填补)对数据呈现结果产生直接影响,能够揭示数据修正过程,提供数据可靠性信息支持,确保语义描述更加客观、严谨。

(2)视觉元素包括坐标轴信息、图例说明、颜色编码、数据点和线条样式,以及注释和标签等视觉特征[32],这些元素提供理解图像内容的直观线索。坐标轴标记影响数据趋势的理解;颜色编码帮助区分不同数据集或数值区间;注释和标签用于对特定数据点或关键趋势的详细解释。通过结合各类视觉元素信息,语义描述框架能够生成精确的描述,使读者全面地掌握图像所传达的科学信息。

(3)统计信息提供了定量分析依据,包括描述性统计、统计检验方法、置信区间和误差范围等[33]。描述性统计信息揭示数据基本特征,置信区间与误差范围为数据的稳健性和结果的可靠性提供量化支持,尤其在描述包含复杂统计方法的图表时,帮助说明研究结果的可信度,使语义描述更加准确、严谨。

2.3 语义层

语义层旨在揭示图像中的核心发现、解释性内容和研究背景,深入探究内涵的逻辑推理和科学意义,通过联系文献中的研究问题、研究方法、研究结果等知识元[24]970-975进行高层次抽象和推理。

(1)核心发现回答了图像“说明了什么”的问题,描述图像中的关键趋势、异常值、重要模式和核心结论,揭示科学发现[34]。在描述实验结果图像时,语义描述需明确指出关键发现,包括是否符合预期、是否有显著变化以及与研究假设的相关性等核心信息。

(2)解释性内容通过对数据的解释、趋势说明、潜在原因分析和影响因素的讨论,丰富了图像的语义层次[35]。结合领域知识,对数据背后的原因进行推测与分析,如针对异常值分析其可能源于实验设计或外部环境因素的干扰。解释性内容不仅帮助读者理解数据的呈现原因,更能增强图像在科学研究中的信息深度,使图像不再仅限于数据呈现而是提升到解释数据现象的层面。

(3)研究背景提供了广泛的上下文信息,包括研究的目的、理论框架、研究假设及研究限制[12]170-177。这部分将论文图像嵌入更宏观的

研究框架之中,确保图像语义描述与研究整体的连贯性。通过将图像放置于研究背景中,语义层的描述不仅能够揭示图像与研究目的的联系,还能说明其在研究流程中的具体作用和意义。

2.4　关系层

关系层通过分析图像内部和外部的关联性,揭示图像内容与更广泛研究的联系,细分为内部关系和外部关系 2 部分:

(1) 内部关系关注图像内部元素之间的相互作用和依存关系,包括变量之间、组件之间的关系,以及因果关系和层次关系[36]。在多变量图表中,变量间的相关性或因果性分析是图像语义描述的重点。这些内部关系揭示了图像内容的内在逻辑结构,帮助呈现复杂的交互模式;因果关系解释实验现象的产生机制,层次关系揭示数据或概念间的从属和分类结构。通过内部关联分析,语义描述框架可以系统化地呈现复杂信息,使读者深入理解图像中的科学逻辑。

(2) 外部关系解析图像与其他研究内容之间的关联性,包括图像与其他图表、文本内容、研究问题以及理论框架的多重联系[37]。图像可能用来补充验证文本中的实验结果或与其他图像协同展现研究的不同维度。通过明确图像与其他研究内容的交互关系,语义描述框架能够生成更加全面的图像描述,使读者清晰了解图像在整体研究中的地位和作用。外部关系解析有助于读者将图像内容与研究关键问题相互联系,确保图像在语义表达中与研究主题保持一致性,增加科研成果的系统性和说服力。

3　基于科学论文图像语义描述框架的知识表示模型构建

3.1　模型构建目标与需求评估

本研究采用系统化的本体工程方法进行设计与构建,旨在准确表达学术图像的语义内容。通过分析现有图像语义描述本体,结合学术图像特定需求,构建多层次、可扩展的语义描述框架,以支持图像内容的多维度解析与知识发现。

经前期调研,确定了 4 个核心应用场景:(1)学术图像的自动标注与语义提取;(2)基于语义的精确检索与知识推荐;(3)跨模态学术知识图谱的构建;(4)学术文献的智能化分析与挖掘。在本体复用方面,本研究重点分析了以下相关本体:(1)都柏林核心元数据(Dublin Core Meta Data, DC)提供的基础元数据描述框架;(2)VRA Core[38]针对视觉资源的专业描述框架;(3)CIDOC CRM[39]在文化遗产领域的语义建模。经评估后,选择性地整合了 Dublin Core 的核心元数据属性以及 VRA Core 中适用于学术图像的视觉描述元素,并参考 CIDOC CRM 的事件建模方法构建了图像与研究活动的关联模式。

在应用层面,该模型需要有效支持知识组织、信息检索和语义推理。知识组织包括构建知识图谱、建立概念体系及关系网络分析;信息检索支持多维度的查询、关系推理和知识发现。模型在实现上推荐标准化的数据建模,设置关系权重并保持扩展性,从而确保语义框架的兼容性和应用的可持续性。系统实现方面则需分层构建、聚焦于核心关系、确保数据质量,为“AI4Science”智能化知识服务提供基础支撑。

3.2　语义框架层次映射关系分析

知识表示模型的设计是实现语义描述的关键环节,本研究基于图像语义描述框架的层次结构,采用本体模型作为核心技术手段,通过将不同层次的语义信息映射到相应的本体类,实现对图像内容的精准、系统化描述。本文从 4 个层次展开分析,阐述各层次的映射关系及其在知识表示中的作用,探讨层次间的逻辑关系与模型的整合价值。

如图 2 所示,基础标识层通过提供图像的技术规格、格式参数和创建信息等基础数据,为内容层的深度解析提供必要的技术保障和约束条件,确保数据提取的准确性和可靠性。内容层与语义层形成双向的深度交互机制,其中内容层通过对数据结构、视觉特征和统计信息的系统化分析,为语义层的解读提供了客观依据,而语义层的理论框架指导内容层关注特定的数据特征和视觉元素,形成相互促进、不断深化的信息增强循环。语义层与关系层的协同体现了知识组织的高层抽象,语义层通过对研究目的、

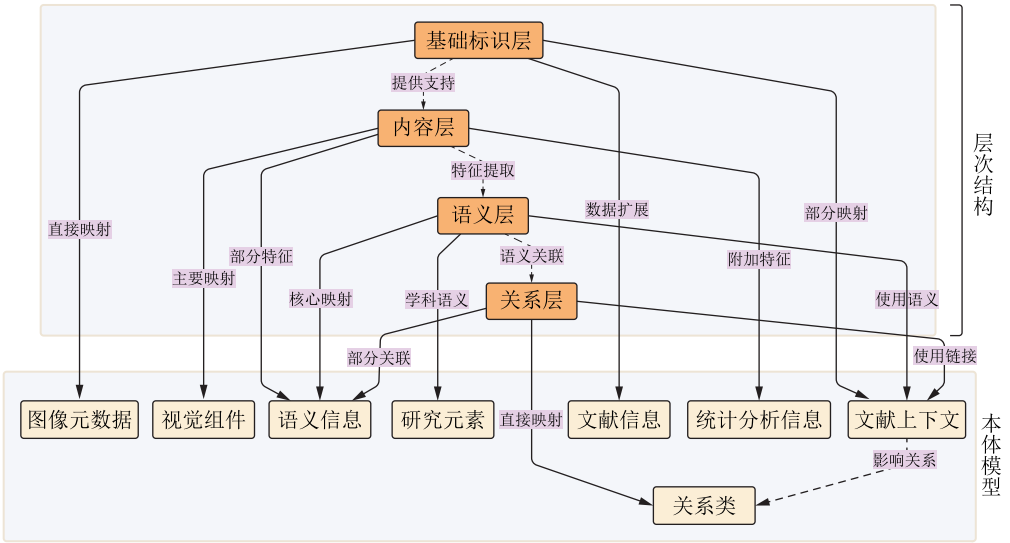


图2 语义框架与本体模型映射分析

理论框架和研究发现的深入理解,为关系层构建准确的逻辑关联提供理论基础,同时关系层建立的多维关联网络反向强化了其他层次的描述深度。通过外部关系丰富基础数据信息,内部关系完善结构化表达,跨模态关联深化语义解释。这种跨层级的信息流动机制实现了从基础描述到高层语义的完整表达链条,各层级在保持相对独立功能定位的同时,通过紧密的关联形成了高度协同的有机整体。层次化的框架设计确保了语义描述的系统性和完整性,为基于学术图像的知识发现和智能应用提供了坚实的理论基础,充分体现了本体模型在支持学术图像语义描述和知识服务方面的科学价值。

3.3 科学论文图像语义描述本体构建

本研究采用迭代式设计方法构建科学论文图像语义描述本体,通过系统化的领域分析提取核心概念类。本体模型包含7个主要类簇:文献类(Document)、图像类(Image)、内容类(Content)、语义类(Semantics)、统计分析类(Statistical Analysis)、视觉组件类(Visual Component)和关系类(Relationship)。在概念提取基础上,结合语义框架层次映射关系,明确了各层次间的交互机制,构建了完整的语义表达链条,见图3。

文献类涵盖不同的科学论文形式;图像类则包含统计图、模型图、示意图和实验图像等不同类型的学术图像;内容类主要描述图像的基本结构要素,如数据结构、数据系列、数据点和变量等;语义类聚焦于图像的深层语义表达,包括研究目的、研究假设、研究背景、主要发现、理论框架、研究限制和解释性内容;统计分析类专注于图像中的定量特征,包括描述性统计、统计检验和置信区间;视觉组件类则描述图像的视觉表现形式,包括坐标轴、图例、标签、注释等元素;关系类通过内部关系和外部关系刻画图像要素间的语义联系。

本体的属性设计采用数据属性与对象属性相结合的方式。数据属性涵盖基本信息(如标题、创建时间)、技术细节(如格式、分辨率)、版权信息、统计信息等属性字段。对象属性则定义了类之间的语义关联,如图像与文献的归属关系、内容要素间的组织关系、语义概念间的逻辑关系等。为确保本体模型的严谨性与可用性,研究运用OWL本体语言的约束机制,通过owl: Restriction对属性进行了严格的定义域(rdfs: domain)、值域(rdfs: range)和基数限制。例如,将hasBasicInformation属性的定义域限定为Image类,值域限定为Metadata类,并通过owl: Reflexive标注其自反性,以保证每个图像实例都必须具备基本信息描述。同时,通过设

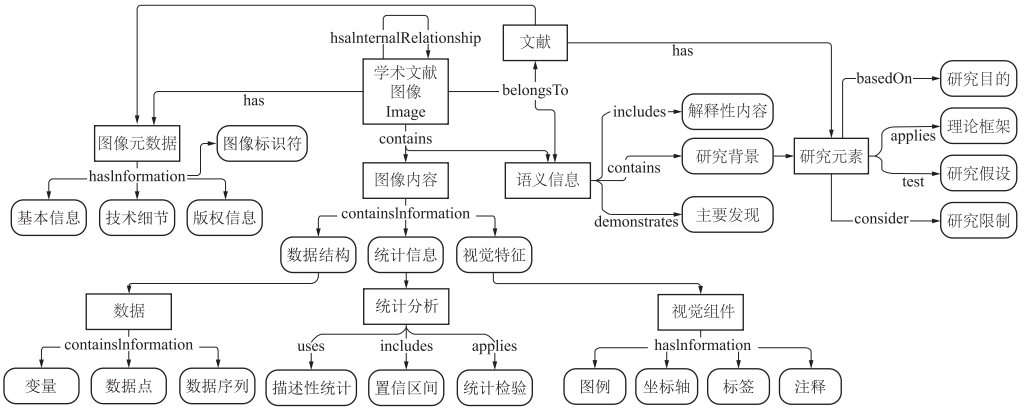


图 3　科学论文图像概念类本体模型图



图 4　科学论文图像语义描述本体模型的概念类、对象属性与数据属性

置propertyChain等高级特性,实现了复杂语义关系的传递表达,增强了本体的推理能力。图4展示了本体模型的概念类、对象属性与数据属性的完整构建结果。

4 面向AI4Science的科学论文图像语义描述框架验证

4.1 科学论文图像语义描述框架应用流程

科学论文图像语义描述框架应用流程如图5所示,具体包括:获取科学论文数据集,将原始科学论文文档经过预处理形成标准化的学术图像数据集与图像相关上下文文本数据集;应用科学论文图像语义描述框架SDF-SFI分别进行图像语义表示与文本语义表示,基于所提取出的图像特征与文本特征进行标注结果规范化处理,以增强图像和文本内容的一致性和关联性;在语义信息结构化和描述质量评估与优化后,进行学术论文图像语义知识标注结果的存储与分析。经过上述应用流程,实现对科学论文的多维度深度挖掘,支持知识发现、趋势分析和多层次的学术研究应用。

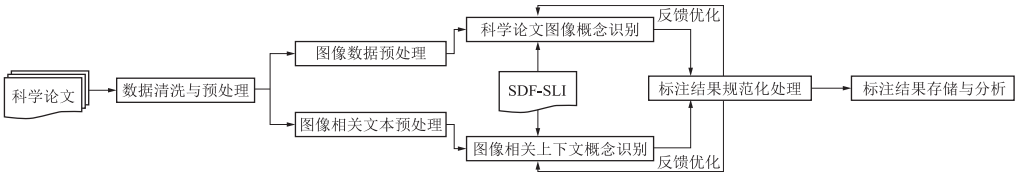


图5 科学论文图像语义描述框架应用流程

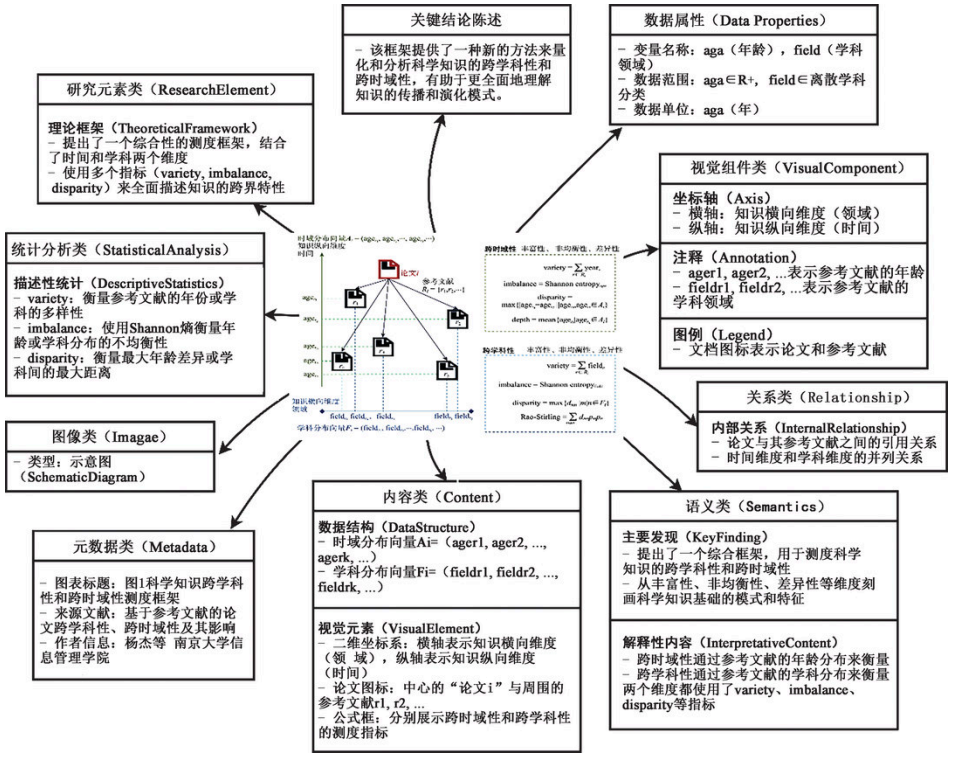
为了充分验证提出的科学论文图像语义描述框架SDF-SFI,本研究使用多模态大语言模型GPT-4o[40]以验证所构建的科学论文图像语义描述本体的有效性,选择图书情报与档案管理领域中文核心期刊所发表的论文作为研究对象,将结果应用于具体的科学论文示例[41]中。

4.2 科学论文图像语义描述框架的实证案例

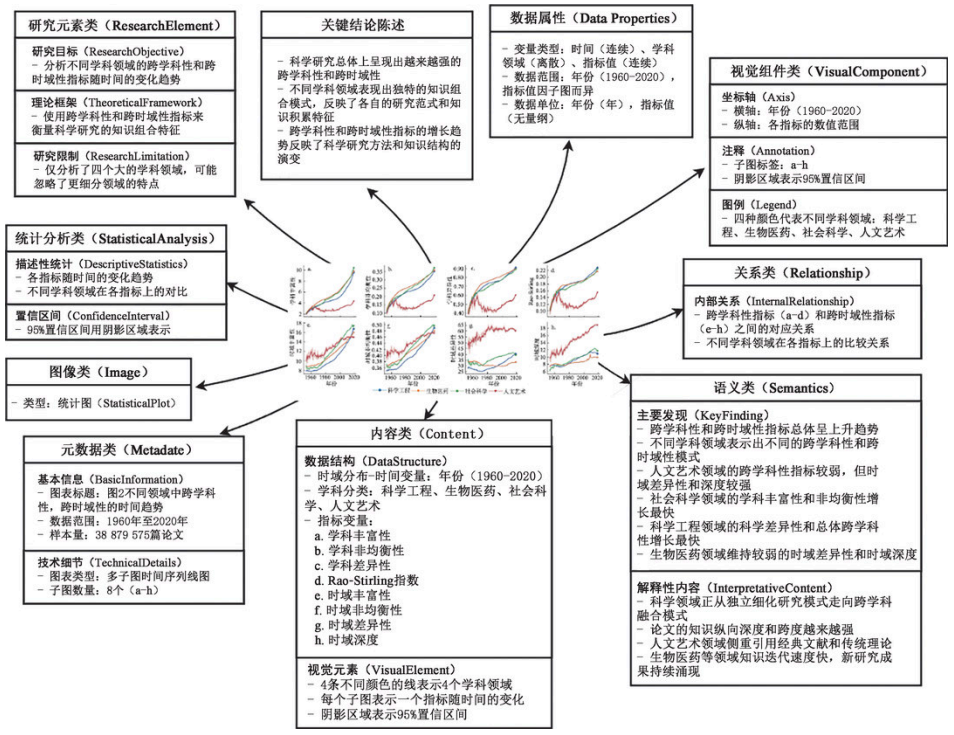
本研究针对学术图像语义标注的实证研究设计了系统化的验证流程,重点探索了基于多模态大语言模型GPT-4o的标注方法及其在语义描述框架各层次中的具体应用。在实践过程中构建了标准化的提示模板(Prompt Template),该模板以图像输入、上下文信息和分析层级为核心要素,系统约束模型从元数、内容解析、语义关系及研究解释等多个维度生成结构化结果,并以统一的JSON格式输出,从而提升标注过程的规范性、一致性与可复用性。

该模板基于语义描述框架的4个层次进行结构化设计,包括基础标识层、内容层、语义层和关系层的具体任务要求。实验采用了来自高被引学术论文的多张典型图像作为验证样本,涵盖了折线图、散点图、模型图等多种类型。在具体的标注过程中,针对每张图像,首先将其与上下文文本一起输入GPT-4o,通过预先设计的提示模板引导模型按照框架层次进行系统化解析。研究采用了三层递进的验证机制以保证标注质量:首先,通过进行单个案例的深度分析,详细记录框架在每个层次的应用效果以评估模型输出的完整性和准确性;其次,由2位图情领域博士生对标注结果进行人工审核和优化;最后,邀请领域专家对处理结果进行评估,从专业角度验证框架的普适性。评估结果显示,框架在概念准确性、语义完整性和关系描述3个维度的平均得分均超过4.2(满分5分),表明了框架良好的通用适用性;实验结果表明,基于GPT-4o的语义标注方法在保证效率的同时,也实现了较高的标注质量,在处理复杂的学术图像时也表现出强大的跨模态理解能力。

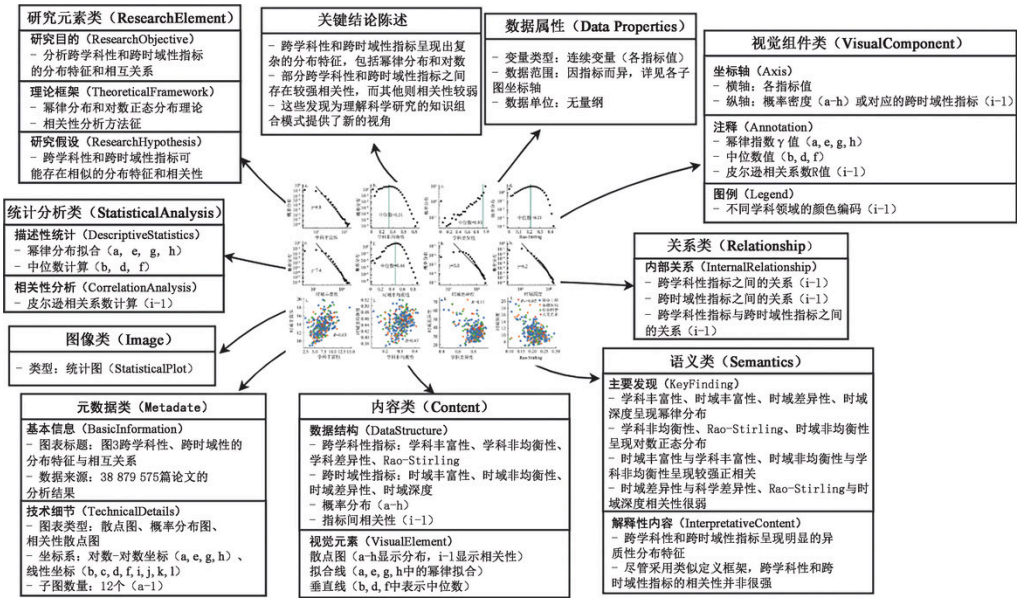
具体应用案例如图6所示,模型对学术论文图像的基本标识信息进行了提取,包括图表标题、创建时间、作者信息等,为图像内容的溯源和背景理解提供了基础。在图像内容结构和视觉元素分析方面,模型能够识别并解读图像中的数据结构(如变量、数据点、数据系列等)和视觉组件(如坐标轴、图例、注释等),确保了对图像内容的精确解析。同时,模型进一步揭示



a）科学知识跨学科性和跨时域性测度框架示例



b）不同领域中跨学科性、跨时域性的时间趋势示例



c）跨学科性、跨时域性的分布特征与相互关系示例

图6 科学论文图像的多维度语义标注示例组图

了图像的语义含义和主要发现，包括对图像中展示的关键趋势、异常值和重要模式的解读，这些信息对于读者快速理解图像所承载的科学信息具有重要意义。此外，模型还对图像的关系属性进行了详细分析，包括内部关系（如图表中不同变量之间的因果关联）和外部关系（图像与文献文本内容以及其他图表的关联）。这一关系识别增强了图像与整体研究主题的连贯性，为跨文献知识的综合分析提供了支持。在数据结构和统计分析方面，模型识别并描述了图像中所使用的统计信息（如均值、标准差、p值、置信区间等），进一步丰富了图像的定量分析能力。最终，模型成功地将图像内容与更广泛的研究背景相联系，解释了图像在特定研究目的、理论框架和研究假设下的科学含义，为图像在学术研究中的应用提供了背景支持。

5 讨论与展望

5.1 面向不同学科的图像语义描述挑战

本研究提出的语义描述框架通过多层次语义描述体系，实现了从基础属性到高层语义的完整覆盖，建立了清晰的概念层级结构，确保了语义描述的逻辑性和系统性。框架的模块化设计支持根据不同学科需求进行定制化扩展。然而，在学术图像语义描述的实践中仍然面临挑战。

不同学科领域展现出显著的特征差异，高度专业化的学科图像可能需要额外的领域知识支持。理工科领域的学术论文图像强调定量数据的精确表达，如实验数据图表、工程设计图纸等；生物医学领域则以显微图像、解剖结构图、分子模型图为主，需要专业的医学知识支持；人文社科领域多采用概念模型图和关系网络图，侧重定性分析的可视化表达。这种学科差异性不仅体现在图像的视觉特征上，更反映在专业术语、研究方法和理论框架等深层语义的表达方式上。构建统一而灵活的跨学科语义描述体系，将成为未来科学论文图像语义框架拓展的首要挑战。

5.2 面向图像理解的关键技术挑战

图像理解技术的核心在于准确识别和解读学术图表的类型与结构，未来需要开发更精准的图像分析工具，提升针对复杂学术图表的识别和解读能力。文本识别与抽取也是科学论文图像语义描述中的关键技术点，包括从图表中提取和理解嵌入的文字信息、从复杂背景中准确提取文字以及处理专业术语和公式等内容。

同时学术论文图像需要与其所在研究的更广泛背景联系起来,这不仅要求对图像本身有深入的理解,还应对相关研究问题、应用方法、分析结论和研究创新有着充分的认识。因此,面向智能化学术知识服务的技术突破主要体现在以下 2 个方面:跨模态融合技术的发展支持学术图像视觉特征与语义信息的深度整合,知识表示技术的完善则为跨模态、跨学科的知识发现提供理论基础[42]。

5.3　未来应用展望

面向 AI4Science 的时代需求,自动图表生成、智能搜索与检索以及图表理解和智能分析将会是未来 3 个主要的研究应用方向[43]。自动图表生成旨在根据语义描述自动生成图表,在提高科研效率的同时,确保生成的图表准确反映出预定义的语义规范。智能搜索与检索将支持通过基于图像语义内容的高级搜索功能,使研究人员能够更准确地找到相关图表资料,支持跨语言、跨学科、跨模态的科技文献信息检索服务[22]。在图表理解和智能分析方面,未来知识服务系统可辅助更好地理解图表内容,支持自动化的知识元分析和系统化的科学评价[26]。本文提出的科学论文图像语义描述框架更侧重于图像内容的语义理解与知识表达,未来可考虑将 IIIF(International Image Interoperability Framework)的互操作标准[44]与本文的语义描述框架相结合,在保证资源互操作性的基础上实现更深层的语义理解与知识服务。

(本文数据链接地址:http://hdl.handle.net/20.500.12304/13383)

参考文献

[1] Huang J, Chen H, Yu F, et al. From detection to application: Recent advances in understanding scientific tables and figures[J]. ACM Computing Surveys, 2024, 56(10): 1-39.
[2] Burri R V, Dumit J. 13 social studies of scientific imaging and visualization[J]. The Handbook of Science and Technology Studies, 2008, 297: 297-317.
[3] Ariga K, Tashiro M. Change in the graphics of journal articles in the life sciences field: analysis of figures and tables in the journal "Cell"[J]. History and Philosophy of the Life Sciences, 2022, 44(3): 1-34.
[4] 邓三鸿,杨杰,王昊,等.多源异构数据视角下的学术评价:内涵、进展与展望[J].科技情报研究,2023, 5(4): 42-56.
[5] 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.
[6] 张鹤,黄倩.多模态信息融合的知识服务[J].中国科技资源导刊,2016, 48(6): 50-54.
[7] 黄永文,孙坦,赵瑞雪,等.大数据与人工智能背景下新型知识服务研究与实践[J].图书情报工作,2022, 66(19): 36-46.
[8] Eakins J P. Automatic image content retrieval-are we getting anywhere?[C]//Proceeding of Third International Conference on E-lectronic Library and Visual Information Research. De Mont fort U-niversity. Milton Keynes: Aslib, 1996: 123-125.
[9] Jaimes A, Chang S F. Model-based classification of visual information for content-based retrieval[C]//Proceedings of SPIE-The International Society for Optical Engineering, 1998: 402-414.
[10] Kong H, Hwang M, Kim P. The study on the semantic image retrieval based on the personalized ontology[J]. International Journal of Information Technology, 2006, 12(2): 35-46.
[11] Siegel N, Horvitz Z, Levin R, et al. Figure seer: parsing result-Figures in research papers[C]//European Conference on Computer Vision, 2016: 664-680.
[12] 丁培,叶兰.科技文献中学术图表标注研究进展[J].现代情报,2021, 41(4): 165-177.
[13] 高隽,谢昭,张骏,等.图像语义分析与理解综述[J].模式识别与人工智能,2010, 23(2): 191-202.
[14] 于永新.基于本体的图像语义识别和检索研究[D].天津:天津大学,2009.
[15] 王晓光,徐雷,李纲.敦煌壁画数字图像语义描述方法研究[J].中国图书馆学报,2014, 40(1): 50-59.
[16] 雷珏莹,王晓光,侯西龙.基于 IIIF 的图像数字叙事策略研究[J].图书情报工作,2023, 67(21): 25-34.
[17] 齐小英,杨海平.历史地图知识组织:需求、框架及实践[J].中国图书馆学报,2024, 50(3): 82-95.
[18] 陈金菊,欧石燕.数字图像语义标注模型比较与分析[J].图书情报工作,2018, 62(6): 116-124.
[19] Jwalin Bhatt, Khurram Azeem Hashmi, Muhammad Zeshan Afzal, et al. A survey of graphical page object detection with deep neural networks[J]. Applied Sciences, 2021, 11(12): 5344.
[20] 徐雷,张亚菲,叶均玲.科技文献创新内容的识别、组织与应用进展[J].情报学报,2024, 43(2): 237-250.

[21] 王晓光,李梦琳,宋宁远.科学论文功能单元本体设计与标引应用实验[J].中国图书馆学报,2018,44(4):73-88.
[22] 张敏,丁良萍,刘欢.面向科技文献的多维语义索引构建思路及实现[J].情报理论与实践,2021,44(8):139-145.
[23] 石栖,陈文杰,胡正银,等.面向知识发现的科学实验知识图谱构建研究[J/OL].数据分析与知识发现,1-18[2024-10-28]. http://kns.cnki.net/kcms/detail/10.1478.G2.20240626.0951.002.html.
[24] 沈思,朱雨菲.面向学术全文本多维知识元的学术图谱构建研究[J].情报学报,2024,43(8):960-975.
[25] 丁培.学术图表知识发现技术框架及研究进展[J].图书情报工作,2021,65(23):136-148.
[26] 范昊,郑小川,热孜亚·艾海提,等.基于知识图谱的标准文献多维知识发现研究[J].情报理论与实践,2023,46(9):175-184.
[27] 徐雷,秦翠玉,李娇.科技文献数据化及组织呈现路径研究[J].中国图书馆学报,2022,48(3):25-42.
[28] Dai Wenjing, Wang Meng, Niu Zhibin, et al. Chart decoder: generating textual and numeric information from chart images automatically[J]. Journal of Visual Languages & Computing, 2018, 10(48): 101-109.
[29] Cliche M, Rosenberg D, Madeka D, et al. Scatteract: Automated extraction of data from scatter plots[C]// Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2017, Skopje, Macedonia, September 18-22, 2017, Proceedings, Part I 10. Springer International Publishing, 2017: 135-150.
[30] Rights Link for Scientific Communications[EB/OL]. [2024-10-28]. https://www.copyright.com/solutions-rightslink-scientific-communications/.
[31] Jaradeh M Y, Oelen A, Farfar K E, et al. Open research knowledge graph: next generation infrastructure for semantic scholarly knowledge[C]// Proceedings of the 10th International Conference on Knowledge Capture. 2019: 243-246.
[32] Hashmi K A, Liwicki M, Stricker D, et al. Current status and performance analysis of table recognition in document images with deep neural networks[J]. IEEE Access, 2021, 9: 87663-87685.
[33] Davila K, Setlur S, Doermann D, et al. Chart mining: a survey of methods for automated chart analysis[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 43(11): 3799-3819.
[34] Berkley C, Bowers S, Jones M B, et al. Improving data discovery for metadata repositories through semantic search[C]//International Conference on Complex, Intelligent and Software Intensive Systems. Fukuoka: IEEE, 2009: 1152-1159.
[35] Takis J, Islam A Q M, Lange C, et al. Crowdsourced semantic annotation of scientific publications and tabular data in PDF[C]//Proceedings of the 11th International Conference on Semantic Systems. ACM, 2015: 1-8.
[36] Shi X, Wu Y, Cao H, et al. Layout-aware subfigure decomposition for complex figures in the biomedical literature[C]//ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2019: 1343-1347.
[37] Wang N X R, Mahajan D, Danilevsky M, et al. Semeval-2021 task 9: fact verification and evidence finding for tabular data in scientific documents[EB/OL].[2024-10-31] https://aclanthology.org/2021.semeval-1.39/.
[38] Vra Core[EB/OL]. [2024-10-31]. https://www.loc.gov/standards/vracore/.
[39] CIDOC-CRM[EB/OL]. [2024-10-31]. https://cidoc-crm.org/sites/default/files/cidoc_crm_version_7.3.pdf.
[40] OpenAI. ChatGPT[EB/OL].[2024-07-27]. https://chat.openai.com/.
[41] 杨杰,王左戎,邓三鸿,等.基于参考文献的论文跨学科性、跨时域性及其影响力研究[J].情报学报,2024,43(9):1003-1014.
[42] 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.
[43] 赖茂生.数智时代的知识组织[J].科技情报研究,2024,6(2):1-10.
[44] 陈涛,李惠,张永娟,等.LIBRA技术理论及其在史料图像资源中的应用[J].大学图书馆学报,2022,40(4):64-74.

张逸勤 南京大学信息管理学院,博士研究生。研究方向:数字人文与文本挖掘。作者贡献:设计研究方法、论文起草。E-mail: yiqin.zhang@smail.nju.edu.cn 江苏南京 210023
邓三鸿 南京大学信息管理学院,教授,博士生导师。研究方向:信息检索、知识管理。作者贡献:提出研究框架、论文修订。 江苏南京 210023
巩洪村 南京大学信息管理学院,博士研究生。研究方向:数据挖掘、自然语言处理。作者贡献:数据标注与处理、模型构建。 江苏南京 210023
杨 杰 南京大学信息管理学院,博士研究生。研究方向:科学计量与评价。作者贡献:数据分析、模型校验。 江苏南京 210023
刘 浏 南京农业大学信息管理学院,副教授。研究方向:数据挖掘、自然语言处理。作者贡献:论文修订。 江苏南京 210000

(收稿日期:2024-11-20 修回日期:2025-03-01)