

人工智能技术在文学领域研究中的应用

The Application of Artificial Intelligence Technology in Literary Research

陈艺心 /CHEN Yixin

(北京科技大学马克思主义学院, 北京, 100083)
(School of Marxism, University of Science and Technology Beijing, Beijing, 100083)

摘要: 文学一直是人文研究的中心, 随着科技发展, 人工智能技术在文学研究中崭露头角。从文本挖掘到情感分析, 人工智能为文学研究提供了全新的视角和强大的分析工具。这一交叉融合使得传统的文学研究在数字时代焕发出新的创意和潜力, 推动着文学与技术共同发展。人工智能在文学领域的应用不仅丰富了研究方法, 还拓展了研究范围, 使研究者能够更深入地理解文学作品背后的文化、社会和心理因素。但也面临挑战, 如算法的可解释性和文学创造性的复杂性。尽管如此, 人工智能与文学的融合为学术界提供了更广阔的空间。近年来人工智能技术在文学研究中的应用, 揭示了人工智能技术如何促进文学研究的创新和发展, 以期为学术界在文学与人工智能交叉领域提供启示。

关键词: 人工智能 文学 自然语言处理 知识图谱 数字人文

Abstract: Literature has always been the center of humanistic research, and with the development of technology, artificial intelligence technology has emerged in literary research. From text mining to sentiment analysis, artificial intelligence provides a new perspective and powerful analytical tools for literary research. This cross fusion has brought new creativity and potential to traditional literary research in the digital age, promoting the common development of literature and technology. The application of artificial intelligence in the field of literature not only enriches research methods, but also expands the scope of research, enabling researchers to gain a deeper understanding of the cultural, social, and psychological factors behind literary works. But it also faces challenges, such as the interpretability of algorithms and the complexity of literary creativity. However, the integration of artificial intelligence and literature provides a broader space for the academic community. In recent years, the application of artificial intelligence technology in literary research has revealed how AI technology promotes innovation and development in literary research, in order to provide inspiration for the academic community in the intersection of literature and artificial intelligence.

Key Words: Artificial Intelligence; Literature; Natural language processing; Knowledge graph; Digital humanities

中图分类号: I0-05; TP18 DOI: 10.15994/j.1000-0763.2025.08.010 CSTR: 32281.14.jdn.2025.08.010

文学领域, 作为一门博大精深的人文学科, 专注于文学作品的创作、解读与批判。在这一领域中, 学术性语言占据着举足轻重的地位。通过对文学作品的细致剖析与深入批评, 我们

能够洞察其背后所蕴含的文化底蕴、社会心理以及意识形态等深层次内涵。文学评论家们运用多种理论框架, 诸如结构主义、后现代主义以及女性主义等, 对文本进行精准解读, 从而

基金项目: 教育部人文社会科学规划基金项目“新时代弘扬雷锋精神研究”(项目编号: 23YJA710020)。

收稿日期: 2023年12月27日; **返修日期:** 2024年12月21日

作者简介: 陈艺心(1987-)女, 北京人, 北京科技大学马克思主义学院讲师, 研究方向为马克思主义原理、文化批判、思想教育研究。Email: 18600967111@163.com

挖掘出作者的创作意图和作品所承载的多重意义。同时,随着数字化时代的迅猛发展,文学作品的数量呈现出爆炸性增长的态势,这给大规模文本数据的处理与挖掘带来了前所未有的挑战。通过对不同时期文学发展的系统分析,我们能够清晰地描绘出文学演进的脉络与轨迹。如何迅速而准确地检索文学作品、深入剖析作品的风格与结构,进而辅助并加速文学研究的进程,已成为当前亟待解决的重要问题。

在工业4.0革命中,人工智能逐渐融入各个领域。尤其在学术研究中,智能技术为研究提供工具和方法,帮助解决复杂问题、处理大数据,并提升研究效能与质量。这也使得文学与人工智能的融合推动了计算文学和数字化人文学的发展。传统文学研究正经历数字化变革,借助自然语言处理、机器学习等技术,文本分析更精细,大规模作品研究成为可能。计算文学利用数学建模和算法,探索文学现象,提供精确、可量化的研究手段。数字化人文学则促进跨学科合作,推动文学数字化转型。尽管人工智能为文学研究带来机遇,但也面临算法偏见、文学创造性复杂性等挑战。总体而言,这一交叉融合为学术界提供了新的研究范式,推动着计算文学和数字化人文学在数字时代的持续演进。^[1]人工智能技术在文学领域的应用凸显了自然语言处理的重要性,包括文本挖掘、意见挖掘和实体识别,这些技术深化了对文学作品的理解。同时,智能化知识图谱构建与文学关系的探讨展示了人工智能技术在建立复杂文学知识网络方面的潜力,进一步增强了文学作品间的关联研究。此外,计算文学与数字化人文学的交叉领域,以及虚拟现实和增强现实在文学领域的应用,为文学研究带来了新视角和互动体验。

一、自然语言处理技术与文学

在文学领域,数据与文本挖掘、意见挖掘、实体识别以及自然语言处理技术的结合,通过深入分析文本数据,揭示情感倾向、主题趋势,实现意见挖掘。实体识别有助于识别文学作品中的关键人物、地点或事件。自然语言处理技

术通过语言模型和语法分析提升对文学作品的理解与解释,为文学研究提供新视角,促进作品内涵的深入挖掘与理解。

1. 数据与文本挖掘

数据与文本挖掘研究包含有文本表示和模型构建,特征提取中的数据降维研究,文本分类^[2]和聚类^[3]挖掘算法研究,基于自然语言过程的深度语义挖掘等。

文学领域的数据与文本挖掘被用于文本分析、主题识别和风格分析等。挖掘大量文学作品,研究人员可揭示作品间关联、作者独特风格和文学流派。情感分析技术也有助于理解读者对作品的情感反应。数据与文本挖掘为文学研究提供了全新视角,促使深入探讨作品背后的隐藏信息,丰富了文学理论和批评的研究方法。在工作中通过构建一种术语连接图,在单个术语级别上自动分析大量全文文章。解释了该方法及其与其他方法如共被引分析和文本挖掘的关系,并将其应用于“Journal of Artificial Societies and Social Simulation”以进行数据挖掘与分析。

2. 意见挖掘

分析文学作品中的观点、情感极性(正面、负面、中性)和作者态度对读者理解和主题挖掘、情感驱动具有重要作用,有助于改善作品或适应读者需求。

例如,在研究中将空间意见挖掘应用于文学作品中,以探讨作家对空间和地点事件的体验方式,主要比较了情感分析和话题建模两种主要技术,并分析各自局限性。还有学者立足于作品“The Sympathizer”,^[4]为了检验所描绘的空间关系以及探究其中潜在意见,基于TextBlob(基于python的自然语言处理库),分别使用潜在狄利克雷分配(Latent Dirichlet Allocation, LDA)和潜在语义分析(Latent Semantic Analysis, LSA)进行主题建模,以帮助文学学者更准确地分析大量作品中的空间关系。同时还有一种有针对性的情感分类模型,最初提取特定tweet的目标,并根据提取的目标确定所述tweet的极性。采用双向门控递归神经网络进行分类。在其他工作中,还通过整

合表层与深层特征来增强情感分类,而提出一种集成情感分析研究方式。

3. 实体识别

实体识别(Named Entity Recognition, NER)是自然语言处理领域的重要任务,用于从文本中识别出命名实体,如人名、地名、组织机构等。该领域的研究关注于开发算法和模型,利用机器学习技术自动进行实体识别,为信息提取、问答系统等提供基础支持。在文学领域,实体识别技术可以辨识文学作品中的事物、地点和事件,有助于深入研究故事结构、角色关系和文学主题。这一技术为研究者提供了高效的分析工具,促进了文学作品的全面解读,并为文学研究提供了精准的数据支持。

例如,在范·达伦-奥斯卡姆(Van K. Dalen-Oskam)等的工作中,将现有的命名实体识别模型应用于现代荷兰小说,并通过将名称链接到维基百科条目来执行命名实体解析。^[5]一些学者为了深入了解小说中所描绘的社区结构和社会互动,提出了一种用于从小说中自动提取社会网络及其网络结构的自然语言处理工具,推动了人工智能系统进一步具有文化意识。名词学作为名称研究学科的一个学科分支,学者们的目标是分析和比较文学作品中名称的用法和功能,戴斯(D. E. Does)等针对该问题,研究了多种NER模型,以对文学作品中的名词进行自动化分析,并将单个文本中的网元以及语料库中网元进行可视化,推动实体识别在文学领域的应用。^[6]

二、智能化知识图谱构建与文学关系的探讨

1. 知识图谱的定义与构建

知识图谱是将每个项目、实体和用户表示为节点,并通过边将相互交互的这些节点链接起来而构建的图。知识图谱将信息表示为语义图的功能引起了工业界和学术界的广泛关注。它提供语义结构化信息的特性为包括问答、推荐和信息检索在内的许多任务带来了重要的可能解决方案,并被许多研究人员认为构建更

智能的机器提供了巨大的希望。自“大数据”一词诞生以来,知识图谱为各种领域的“大数据”应用提供支持。

正如某些相关领域学者所介绍的理论^[7]以及古铁雷斯(H. Gutierrez)所提到的内容^[8]一样,人工智能知识图谱的工作与语义网、数据库、知识工程、自然语言处理和机器学习方面的科学进步密切相关。在过去的几十年里,本体谱的方法为许多研究人员建立了构建词汇表、模式和本体的路线图,以满足各种应用程序。

自20世纪50年代数字计算机开始应用以来,科学家意识到了在数据收集和管理中使用机器可读标准的重要性。许多出版物讨论了与知识图谱相关的主题,例如数据模型的共识、语义符号和网络、受控词汇、^{[9], [10]}空间数据操作规则等等。如今,在互联网和Web时代,知识图谱在数据管理中仍然发挥着重要作用,知识图谱在开放、开放数据与遵循FAIR原则的数据中应用也取得了新的进展。它们发展到现在,有两种构建知识图谱的主要方式:人工手动构建和智能自动构建。以下将从两种方式的优缺点进行总结:

(1) 人工手动构建

在知识图谱发展的早期阶段,依靠人工进行构建,主要优势在于人工对于信息的筛选能够智能进行判别,但也面临巨大的挑战,最直接的挑战是会耗费大量的时间和资源,并且由于人的意识的主观性会出现知识遗漏或者误判的情况。因此,提升效率和准确度,需要先进的算法辅助人们进行知识图谱的构建。

(2) 智能自动构建

随着科技的进步,知识图谱构建智能化已成为主要的趋势,它相对于人工手动构建,效率大幅提升。它可以利用较为新型的算法进行实体关系的抽取,通过关系抽取,可以节约劳动力,减少人力的投入。因此,智能构建知识图谱可以被大范围应用。通过实体识别、关系抽取和信息处理,可构建更宏大的知识图谱。所以,对于文学作品而言,首先需要对文本以及语料库进行加工与处理,从公开的SCC语料库里进行加工与处理,构建所需要的数据集,

并对数据集进行标注。其中SCC语料库中,根据文学作品类型的不同,分为戏剧,小说,民俗以及诗歌等。

2. 实体抽取与关系建模

实体抽取是从文本中识别并提取出具有特定意义的实体的过程。首先,需要对文学作品的分词与词性标注结果进行处理,将文学作品没有意义的词汇,如虚词等进行删除,并将余下的词汇进行错误修正,对文学作品中的人名进行匹配与替换,即简称替换为完整的姓名。

其次,根据词性积习难改匹配抽取实体,根据词性标签定义几种实体,如人物等实体,并人工进行实体间关系的定义,需指明,在实体抽取中需要手动或者半自动标注人物实体的开始和结束的位置。通过实体抽取识别文学作品中的事物,地点后进行关系建模,利用上一步实体抽取的结果建立人物之间的关系,这需要对实体表示进行进一步的处理,然后输入到关系建模中的模块。

最后,即可通过标注的数据进行模型的训练,在此过程中,可以正确地抽取人物实体与建模任务之间的关系,并通过另一组标注好的数据进行模型的评估,检查文学作品中实体抽取与关系建模任务中的准确性。

3. 语义推理与图谱扩展

通过实体抽取与关系建模后,构建一个初始的文学图谱,此图谱可以是一个简单的图结构,其中节点表示实体,边表示关系。而语义推理是利用自然语言处理进一步推导出文学作品中隐含的各种信息,包括但不限于通过上下文推断出文学作品中人物的情感关系,以及预测故事中未提及的事件等。而图谱扩展是根据推理结果,扩展图谱,添加新的实体和关系。例如,如果推理表明两个人物之间存在未明确提及的关系,可以将这个关系添加到图谱中。最终即可得到知识图谱在文学中的应用。

三、计算文学与数字化人文学探讨

1. 数字化人文学

数字信息技术的兴起以及随之而来的“计

算转向”从根本上改变了我们的研究方式。人文学科使用人工智能进行研究也被称为数字人文(Digital Humanities, DH)。这是一门利用计算机和数字技术来研究文学、历史、艺术等人文学科的学科。罗斯(C. D. Roth)将当前存在的三个子领域确定为DH的独特且不重叠的社区:“数字化人文学科”,涉及数字化资源的创建、使用和分析;“数值人文”也称为“计算人文”,通过计算模型来解决人文研究问题;以及“数字人文”,致力于从人文角度研究在线交流等数字现象。^[11]伯格哈特(M. J. Burghardt)扩展了罗斯的分类法,添加了“公共人文学科”,涉及人文学科、数字出版和电子学习中的学术交流。^[12]

(1) 语料库创建和采样

a. 数据获取

DH在文学领域的研究,首先需要对数据进行获取,语料库的核心由三本在社区中颇有声望且已经存在了一段时间的DH期刊组成。^[13]除了DH文章的子语料库之外,还可以进一步纳入以某种方式与DH相关的其他学科的子语料库,总共有34,041篇文章,超过2.99亿个token。通过一些学者的现有研究对学科选择进行指导,因为它全面概述了在*Computers and the Humanities*上发表文章的作者及其与核心学科的关系。^[14]卢曼(J. Luhmann)使用GROBID9将每个PDF文件转换为XML,从中提取文章文本^[12]。DHQ文章以原生XML文件形式从其官方网站下载。为了平衡学科之间文章的长度,我们只保留长度在1,000到20,000个令牌之间的文章。我们尝试自动删除非研究性文档,这些文档由通用标题标识,例如“社论”“书评”“简介”和“新闻和笔记”并删除了那些作者未明确记录在元数据中的文章。

b. 预处理与数据清洗

他们指出所有文章均需要使用spaCy进行标记化、词性(POS)标记、词形还原和小写化。关于词形还原,对几个特定领域的仅复数名词(人文、语言学、数据、媒体等)需要进行一些手动更正。通过手动检查判断,可能从收集的文本会包含一些来自错误光学字符识

别 (Optical Character Recognition, OCR) 的噪音, 主要是由于原始来源中的单词连字符造成的。如果要清除这些错误, 需要连接任何二元组, 包括连字符, 在这些情况下, 二元组在语料库中出现的频率低于串联的一元组。例如, 二字母连字符比一字母连字符出现的频率要低得多。因此, 每次出现连字符都会被替换。不存在连字符的二元组 (例如, 纪律、信息) 保持不变, 因为这些无法明确解决, 这种情况下产生的最常见的词粒子 (-tion、-ity、-ing、-sion 等) 被 LDA 过滤掉, 被分组为一个单独的主题。除此之外, OCR 噪声对于分析的主题的质量没有任何明显的负面影响。

c. 特征选择和短语连接

对数据进行清洗后, 需要选择专门对名词、专有名词和名词短语进行主题建模。通常需要提供降噪功能, 可以消除与识别主题无关的所有功能词。正如一些学者在文献中所提及, 仅使用名词的方法可以带来更好的可解释主题, 同时还可以加快 LDA 训练过程。涉及概念或实体的技术术语大多是遵循某些词性模式的单个名词和名词短语。^{[15], [16]} 为了提取名词短语, 需要连接多次出现并显示出高关联值 (≥ 0.3) 的 2-、3- 和 4-gram, 通过标准化逐点互信息测量 (Normalized Pointwise Mutual Information, NPMI)。然后按照原始 POS 模式过滤候选者。为了进一步减少噪音, 删除了包含少于三个字母字符或任何标点符号 (撇号和连字符除外) 的术语, 以及手动定义的高频词停止列表中的术语。最后, 需要限制用于对具有最高平均 tf-idf 得分 (术语频率 - 逆文档频率) 的术语进行主题建模。

(2) 采样

由于文章的数量差异很大, 为了确保每个学科和十年对主题具有相同的影响, 并且期刊的影响力不会历时变化, 卢曼每次运行主题建模都是使用分层的文章样本作为训练数据来执行。需要对一些文本进行适度随机过采样, 对其他的文本进行随机欠采样。最终聚合主题模型的所有主题模型和基于概率分布的统计都是根据该方案在语料样本的基础上计算的^[12]。

(3) 主题建模

主题建模是一种无监督机器学习方法, 用于根据单词共现模式发现大量文档中潜在的隐藏语义结构 (主题)。通过对语料库进行分析, 使用主题建模方法, 其中 LDA 方法是最流行的主题建模技术之一。而 LDA 实现通常应用自然语言处理工具包 MALLET 的标准模型, 它使用吉布斯采样^[17]来训练模型, 并且还具有非对称 α 和对称 β 狄利克雷先验优化方法。

(4) 主题聚合

近年来, 通过协调多个模型的主题来确保 LDA 可靠性得到了进一步发展,^[18] 相关领域的研究人员引入使用聚类技术聚合多个主题模型的方法。^[19] 首先使用先前确定的超参数和不同的随机状态获得了 N 个 LDA 模型, 每个模型都适合新的语料库样本。根据术语概率分布之间的余弦距离, 使用分层凝聚聚类对它们进行聚类。

2. 计算文学

文学研究是涉及文学文本和文学相关文物分析的学术领域。重点通常是具有艺术意图的文本, 即或多或少偏离日常语言的文本。那么, 该学科的主题也是文学的接受情况、文本总体趋势的分析以及文学生产的条件。因此, 该学科涵盖了从戏剧到小说和诗歌的广泛文本形式, 并且还可以依赖其他数据。计算文学研究 (Computational Literary Studies, CLS) 是使用计算方法进行分析的学科——这是可以从相对简单的字数统计开始或主要基于手动注释的研究方法。

基于此, CLS 方法的一个关键挑战是运用传统的文学研究类别和理论进行计算分析, 也就是对抽象的类别进行建模, 从文本表面自动做出推断。与此同时, 许多 CLS 方法需要使机器语言处理支持的研究结果可用于文学研究中的解释。也就是, 对数值、定量结果的解释汇集了作为自然科学中心目标的“解释”和作为人文科学目标的“理解”。因此, 可解释性问题也可以更广泛地理解为关于计算方法产生的结果到传统非计算文学理论领域的“可翻译性”问题。为了弥合自然科学与人文科学之间的差

距,不仅结果必须易于理解,而且获得结果的过程也必须易于理解。所以,CLS方法包含解释、理解和解释模型输出。

四、虚拟现实与增强现实

创建沉浸式的文学体验,可以使读者更直观地参与到故事中。例如,虚拟现实技术可以用于创造文学作品的虚拟环境,增强现实可以与纸质书籍或电子书结合,提供额外的信息或交互元素。

1. 虚拟现实

虚拟现实技术是一种将计算机图形学、多媒体技术、传感技术等多学科领域的技术进行有机整合而形成的一种综合性的新兴技术,被广泛应用于各个领域,包括游戏、运动教学、抑郁情绪干预、图书馆服务、高校思政课教学、教育领域以及中医经络教学等各个领域。在文学领域,文学作品与虚拟现实技术结合会催生一种具身性的沉浸式体验,这对促进文学的发展具有重要的意义。所以,如何实现虚拟现实技术与文学作品结合就变为核心问题。这些核心问题包括如何解决开放性与多样性解读的问题——自然语言与最终界面怎样实现、如何克服VR眼镜造成的眩晕感等。这些难题的解决,会推动VR技术与文学的共同发展。文学与VR技术的融合面临诸多挑战,然而,随着这些挑战的逐步克服,文学领域有望迎来一场深刻的变革。

2. 增强现实

当计算机生成的数字化虚拟图像与真实物理世界在同一时空并存时,增强现实技术诞生。增强现实技术是指计算机通过摄像头对真实世界进行捕捉,将虚拟世界与真实世界实时融合,并由媒介输出展现的技术。这种技术的到来刺激了人类的感官神经,并以不可阻挡的势头被应用到各行各业中,实现多用户交互仿真。

在文学领域的应用主要包含如下几个方面:文学创作与增强现实、读者体验的增强、文学教育与增强现实、文学场景的虚拟化和文学分析工具的创新。^[20]文学创作方面,作家可

利用增强现实技术创造虚拟世界、交互式故事体验等文学作品,改善读者的文学体验,如虚拟书签、交互式阅读应用等。同时,在文学教育中,增强现实技术可通过虚拟场景帮助学生更好地理解文学作品。该技术为教学提供了沉浸和体验的学习环境,促进了学习者的学习和知识建构。作为主流教学形式的补充,增强现实技术不仅丰富了教学形式,还为开发者提供了理论依据和实践经验。最后,该技术也可用于开发文学分析工具,以更深入、更交互的方式理解和研究文学作品。

五、前景与挑战

人工智能技术在文学研究中潜力巨大,能拓展研究深度和广度,分析作品,挖掘隐藏主题、情感和关系,助力理解文学作品,为研究者提供广阔视野。虚拟现实和增强现实技术为文学创作者提供创新方式,读者也能更互动、沉浸式参与故事,为文学界带来创新和变革。自然语言处理技术则使从大量文学作品中提取信息更高效,加速研究过程,提高准确性。

但人工智能技术在文学研究中也面临挑战。目前,该技术难以解析文学作品的深层含义和隐喻,因为文学作品通常蕴含丰富的情感、文化和历史背景,超出了当前人工智能技术的理解范畴。因此,提高人工智能技术在解读文学作品深层含义方面的能力成了一个迫切的问题。

此外,人工智能技术在文学创作中也面临着创造性和独特性的考验。文学创作是一种极具个性化的活动,需要独特的想象力、情感和观点。尽管人工智能技术可以通过学习海量的文学作品来生成新的文本,但这些生成的文本往往缺乏真正的创造性和独特性,因为它们仅仅是基于已有的知识和模式进行组合和重构。为了应对这些挑战,首先通过优化人工智能技术的算法和模型,可以更好地理解和解读文学作品的深层含义。例如,开发更先进的自然语言处理技术,使人工智能技术能够更准确地把握文本中的语境和语义。其次,尝试结合人类

的创造力和人工智能技术的生成能力,以提升文学创作的独特性和创造性。例如,让人工智能技术提供创作的灵感和素材,再由人类作家在此基础上进行创作和修改。这样,人工智能技术和人类可以携手合作,共同创作出既具创造性又独特的文学作品。

尽管人工智能在文学数据处理方面有所进展,但仍存在技术局限性,需要更先进的技术来应对文学作品的复杂性、多样性和创造性。在数字化人文学和计算文学的交汇点,需保持对文学本质的理解,避免过度强调技术而忽略文学的人文性和艺术性。

综上所述,人工智能技术在文学研究中的应用,重点发力点应在自然语言处理技术与文学、智能化知识图谱构建与文学关系、计算文学与数字化人文学、虚拟现实(Virtual Reality, VR)和增强现实(Augmented Reality, AR)在文学领域应用等方面。尽管人工智能技术在文学研究中有许多挑战,但只要我们保持探索和创新,就有可能找到克服这些挑战的方法。随着科技的进步和人类对人工智能技术的深入理解,人工智能技术将在文学研究中发挥越来越重要的作用。

[参考文献]

- [1] Principe, V. A. 'A Computational Literature Review of Football Performance Analysis Through Probabilistic Topic Modeling'[J]. *Artificial Intelligence Review*, 2022, 55(2): 51-71.
- [2] 李文波、孙乐、张大鲲. 基于Labeled-LDA模型的文本分类新算法[J]. *计算机学报*, 2008, 31(4): 620-627.
- [3] Yong, Z., You, W. L., Shi, X. X. 'An Improved Knn Text Classification Algorithm Based on Clustering'[J]. *Journal of Computers*, 2009, 4(3): 230-237.
- [4] Yun, Y. S., Keikhosrokiani, P. Pouryaasl, M. 'Opinion Mining on Viet Thanh Nguyen's The Sympathizer Using Topic Modelling and Sentiment Analysis'[J]. *Journal of Information Technology Management*, 2022, 14: 163-183.
- [5] Van Dalenokam, K., Dedoes, J. 'Named Entity Recognition and Resolution for Literary Studies'[J]. *Computational Linguistics in the Netherlands Journal*, 2014, 4: 121-36.
- [6] DE Does, J., Depuydt, K., VAN Dalen-Oskam, K., et al. *CLARIN in the Low Countries*[M]. UK: Ubiquity Press, 2017, 361-370.
- [7] Abu-Salih, B. 'Domain-Specific Knowledge Graphs: A Survey'[J]. *Journal of Network and Computer Applications*, 2021, 185: 103076.
- [8] Hagra, H. 'Toward Human-Understandable, Explainable AI'[J]. *Computer*, 2018, 51(9): 28-36.
- [9] Hubaux, A. 'Description of Geological Objects'[J]. *Journal of the International Association for Mathematical Geology*, 1970, 2: 89-95.
- [10] Rassam, G., Gravestijn, J. 'Cross-Database, Cross-National Geologic Indexing: Problems and Solutions'[J]. *Geology*, 1982, 10: 600-603.
- [11] Roth, C. 'Digital, Digitized, and Numerical Humanities'[J]. *Digital Scholarship in the Humanities*, 2019, 34: 616-632.
- [12] Luhmann, J., Burghardt, M. 'Digital Humanities—A Discipline in Its Own Right? An Analysis of the Role and Position of Digital Humanities in the Academic Landscape'[J]. *Journal of the Association for Information Science and Technology*, 2021, 72(3): 314-328.
- [13] Schreiban, S., Siemens, R., Unsworth, J. *A Companion to Digital Humanities*[M]. Oxford: Wiley-Blackwell, 2004, Xxiii-Xxvii.
- [14] Sula, C. A., Hill, H. V. 'The Early History of Digital Humanities: An Analysis of Computers and The Humanities (1966-2004) and Literary and Linguistic Computing (1986-2004)'[J]. *Digital Scholarship In The Humanities*, 2019, 34: I190-I206.
- [15] Justeson, J. S., Katz, S. M. 'Technical Terminology: Some Linguistic Properties and an Algorithm for Identification in Text'[J]. *Natural Language Engineering*, 1995, 1: 9-27.
- [16] Weikum, G. 'Foundations of Statistical Natural Language Processing'[J]. *Acm Sigmod Record*, 2002, 31: 7-8.
- [17] Blei, D. M. 'Probabilistic Topic Models'[J]. *Communications of the Acm*, 2012, 55: 77-84.
- [18] Griffiths, T. L., Steyvers, M. 'Finding Scientific Topics'[J]. *Proceedings of the National Academy of Sciences*, 2004, 101: 28-35.
- [19] Blair, S. J., Bi, Y., Mulvanna, M. D. 'Aggregated Topic Models for Increasing Social Media Topic Coherence'[J]. *Applied Intelligence*, 2020, 50: 138-156.
- [20] 韩模永. 增强现实与空间转向——网络文学的场景书写及其审美变革[J]. *文艺理论研究*, 2019, 39(4): 33-38.