基于ChatGPT API和提示词工程的专利知识图谱构建*

张玲玲 黄务兰

（上海商学院商务信息学院 上海 200235）

摘 要：[**研究目的**]在信息爆炸的时代背景下，专利数据的快速增长为知识管理和分析带来了新的挑战。该文旨在探讨利用ChatGPT从专利摘要中抽取信息，构建专利知识图谱，以提升知识管理和分析的效率和准确性。[**研究方法**]从中国知网专利数据库选取了智能驾驶领域的专利摘要，利用ChatGPT进行信息抽取。为实现高效批量处理，采用了ChatGPT API接口与模型进行交互。为确保信息抽取的准确性，多次迭代和优化提示词，设计了系统消息、助手消息及用户消息三种角色，通过模拟对话场景，引导模型精确抽取实体与关系。[**研究结果/结论**]研究结果表明，ChatGPT成功从1 126份专利摘要中提取了丰富的五元组信息，并以此为基础构建了专利知识图谱。与传统方法如Bert2Keras相比，ChatGPT在精确率、召回率及F1值等关键指标上均表现出明显优势，分别达到了88.2%、88.3%和88.3%，远超Bert2Keras的34.7%、9%和14.6%。最后，利用抽取的实体关系和Neo4j技术，成功地构建了知识图谱并完成了可视化展示，便于通过Cypher语句进行查询操作。该研究不仅证实了ChatGPT在专利知识图谱构建中的可行性，也为其在知识产权管理、技术研发及竞争情报分析等方面的智能化应用奠定了基础。

关键词：ChatGPT API；提示词工程；专利知识图谱；实体关系抽取；智能驾驶

中图分类号：G251 **文献标识码**：A **文章编号**：1002-1965(2025)03-0180-08

引用格式：张玲玲，黄务兰. 基于ChatGPT API和提示词工程的专利知识图谱构建[J]. 情报杂志，2025，44(3)：180-187.

DOI：10.3969/j.issn.1002-1965.2025.03.022

Construction of Patent Knowledge Graphs Based on ChatGPT API and Prompt Engineering

Zhang Lingling Huang Wulan

(School of Business Information, Shanghai Business School, Shanghai 200235)

Abstract: [**Research purpose**] In the context of the information explosion era, the rapid growth of patent data presents new challenges for knowledge management and analysis. This study aims to explore the use of ChatGPT for information extraction from patent abstracts to construct patent knowledge graphs, thereby enhancing the efficiency and accuracy of knowledge management and analysis. [**Research method**] Patent abstracts in the field of intelligent driving were selected from the China National Knowledge Infrastructure (CNKI) patent database, and ChatGPT was utilized for information extraction. To achieve efficient batch processing, the ChatGPT API was used for interaction. To ensure the accuracy of information extraction, multiple iterations and optimizations of prompts were performed, designing three roles: system messages, assistant messages, and user messages. These roles guided the model to accurately extract entities and relationships through simulated dialogue scenarios. [**Research result/conclusion**] The results indicate that ChatGPT successfully extracted rich quintuple information from 1126 patent abstracts and constructed a patent knowledge graph based on this information. Compared with traditional methods such as Bert2Keras, ChatGPT demonstrated significant advantages in key metrics such as precision, recall and F1-score, achieving 88.2%, 88.3% and 88.3%, respectively, far surpassing Bert2Keras's 34.7%, 9% and 14.6%. Finally, using the extracted entities, relationships and Neo4j technology, a knowledge graph was successfully constructed and visualized, facilitating query operations

收稿日期：2024-06-03 修回日期：2024-06-24

基金项目：国家社会科学基金一般项目“面向智库建设的图书馆知识服务模式和创新路径研究”(编号：18BTQ058)研究成果。

作者简介：张玲玲，女，1988年生，博士，讲师，研究方向：生成式人工智能、机器学习；黄务兰，女，1979年生，博士，副教授，研究方向：知识管理。

through Cypher statements. This study not only confirms the feasibility of using ChatGPT in patent knowledge graph construction but also lays the foundation for its intelligent application in intellectual property management, technological development, and competitive intelligence analysis.

Key words: ChatGPT API; prompt engineering; patent knowledge graph; entity and relation extraction; intelligent driving

在全球数字化迅速推进的背景下,知识产权保护的紧迫性和重要性日益凸显。为应对这一挑战,优化专利资助奖励政策及考核评价体系成为了关键策略,旨在激励高价值专利的创造与保护,并促进专利密集型产业的成长。这一举措深刻嵌入创新驱动发展战略中,强调了提升专利质量和加速成果转化对增强国家创新能力的核心作用。然而,随着专利数据的海量增长,如何有效地挖掘、组织和利用这些宝贵的知识资源,成为一个亟待解决的问题。传统的专利信息检索系统虽然提供了基本的查询服务,但在揭示专利知识内在关联和深度理解方面存在明显不足[1],难以满足用户对知识发现和技术趋势分析等高级需求。

鉴于此,本文将目光投向了新兴的人工智能技术,特别是以 ChatGPT API 和提示词工程为基础的专利知识图谱构建技术。专利知识图谱是一种图形化的知识组织形式,能够清晰地展现专利之间的复杂关系。通过实体识别、关系抽取等技术手段,专利文档中的技术概念、发明人、申请机构、引用关系等关键信息被转化为节点和边,从而构建出一个多维度、动态更新的知识网络[2]。而 ChatGPT API 作为 OpenAI 推出的自然语言处理接口,凭借其卓越的语言理解和生成能力,为自动解析专利文本、提取关键信息提供了强大的技术支持。同时,提示词工程作为一种优化模型输入策略的方法,通过精心设计的提示词,能够引导模型更加精准地执行特定任务。在专利知识图谱的构建过程中,提示词工程有望进一步提高信息抽取的准确性和针对性。

本研究以 ChatGPT API 为基础,以公开的专利数据为研究对象,结合提示词工程进行实体关系联合抽取,旨在探索构建更加精细化、智能化的专利知识图谱。期望通过本研究,为专利知识的深度挖掘与高效利用提供一条创新的技术路径,以助力企业创新决策、学术研究和政策制定。同时,本研究也将为人工智能在复杂知识组织领域的应用提供新的实践案例和理论启示。

1　相关工作

1.1　专利知识图谱

专利知识图谱是专门针对专利数据构建的一种知识图谱。它通过图谱的方式,有序地组织和清晰展示专利信息,同时精准识别出专利数据中的关键实体及其相互关系,为深入理解和分析专利信息提供了一种新方法[2]。该图谱不仅简化了复杂专利信息的理解过程,还提供了深入挖掘和利用专利数据的有效手段。在推动技术创新方面,专利知识图谱发挥着举足轻重的作用。它能够帮助研究人员和企业迅速把握相关技术领域的最新发展动态,明确研发方向,避免不必要的重复研发,从而加速创新进程[3]。此外,专利知识图谱还为市场分析和知识管理提供了有力支持。通过深入分析专利布局和技术演进路径,政策制定者及企业能够做出更加明智和科学的战略决策[4]。更值得一提的是,专利知识图谱还能促进跨领域的研究合作和协同创新,揭示不同技术领域和研究领域内实体与关系的联系,为发现新的研究方向和合作机会提供了可能[5]。

1.2　专利知识图谱构建方法

专利知识图谱的构建方法经历了从人工处理逐渐转变为深度学习技术的应用。在早期,构建专利知识图谱主要依赖于手动操作。赖朝安等基于关键词共现法构建了移动医疗知识图谱[3],邵泽宇等则通过 Citespace 和手工代码构建了区块链专利知识网络图谱[5]。尽管这些方法具有开创性,但它们面临着人力成本高、数据处理效率低以及数据覆盖范围小等挑战。随着深度学习技术的出现,专利知识图谱的构建方法取得了显著进步。吕向如将知识图谱的构建过程分为三个阶段,并运用了 BiLSTM-CRF 模型、基于注意力机制的 BiLSTM 及关键词策略,和 BERT-BiGRU-CRF 模型,从而成功构建了新能源汽车领域的专利知识图谱[2]。马国斌利用 Bert-BiLSTM-CRF 模型和 Bert 模型分别识别专利摘要中的实体并提取实体间的关系,构建了面向知识检索的制造业专利知识图谱[4]。曹树金等则结合了 BERT 模型、LDA 主题模型以及 Bert4keras,构建了面向创新的教育机器人专利知识图谱[6]。何玉等提出了 SpERT-Aggcn 模型,该模型通过提取嵌套实体并引入完整的依存文法信息,提高了关系抽取的精度,从而构建了绿色合作专利知识图谱[7]。

尽管深度学习技术在专利知识图谱的构建中展现了显著效果,推动了方法的自动化,并提升了效率和质量,但仍存在一些不足。首先,深度学习模型依赖于大量的标注数据进行训练,这在某些专利领域可能并不容易获取。其次,这类模型通常需要大量的计算资源,这在资源受限的环境中限制了其应用。最后,深度学习模型的解释性相对较弱,因此在需要深入理解模型

决策过程的场景中,这可能成为一个问题。

1.3 ChatGPT 在自然语言处理中的应用现状

ChatGPT 的出现为专利知识图谱构建提供了新的解决方案。作为一种先进的人工智能模型,ChatGPT 自发布以来就因其强大的知识学习和复杂语言逻辑理解能力而受到广泛关注。ChatGPT 拥有高达 1 750 亿的模型参数,在专利文献的理解和生成方面,ChatGPT 展现出了明显的优势。此外,ChatGPT 采用的完全端到端训练模式,无需人工标注数据,非常适合进行大规模无监督预训练,从而大幅减轻了人工操作的负担。通过预训练中的对话、问答、文本生成等多种任务,ChatGPT 已经获得了出色的语言理解与生成能力,能够有效应对专利文献中的复杂语言,并准确地抽取实体和关系。同时,ChatGPT 对长篇幅文本具有出色的语义理解能力,能够捕捉文本的上下文逻辑关系[8],这对于理解复杂的专利文档内容具有重要意义。

此外,ChatGPT 在自然语言处理领域已证实具有强大能力,并在多个相关任务中取得了显著的应用成果。作为一款生成式模型,ChatGPT 在学术论文写作和创新性评价领域取得了成功应用。它不仅能够高质量地生成学术论文的引言[9]和中英文摘要[10-11],还能有效地评估论文的创新性[12],标题的语句流畅度和语义相关性[13]。同时,ChatGPT 在论文摘要语法矫正、学术用语简化和规范方面也提供了有力支持[14]。在文本数据增强领域,ChatGPT 也发挥了重要作用。它通过生成与原始文本概念相似但语义不同的样本,丰富了训练数据的多样性和内容[15-16],从而提升了模型训练的性能和准确率。此外,ChatGPT 还能自动化标注数据并构建数据集,大幅降低了人力和时间成本,同时保证了数据集的质量和实用性[17]。

在知识管理和组织方面,ChatGPT 展现了出色的自适应性,能够自动生成和组织知识内容[18],为知识管理提供了新的解决方案,并推动了知识的高效应用。ChatGPT 在处理复杂数据方面也表现出色,如在电信诈骗案件的影响力评估中显示了其强大的分析能力[19]。在情感计算[20]、立场检测[21]和隐含仇恨言论检测[22]等文本分析任务上,ChatGPT 的应用也体现了其对文本情绪和倾向性的深入理解能力,为这些领域提供了新的视角和方法。

另外,ChatGPT 在实体提取方面展现了明显的优势。在专利技术功效实体抽取方面,结合 Prompt 方法,ChatGPT 能有效识别和提取专利技术词、功效词及其二元组,不仅提升了专利技术功效矩阵的构建质量,也进一步展示了 ChatGPT 在处理复杂专利数据、进行跨领域和跨语言分析方面的强大能力[23]。此外,还有研究利用 ChatGPT 从司法文本中提取命名实体,为司法智能化提供了有力支持[24]。然而,这些研究并未涉及实体间关系的抽取,也未进行实体关系联合抽取或构建专利知识图谱。

综上所述,尽管 ChatGPT 在自然语言处理领域的应用已展现出广泛的覆盖范围和显著的优势,为学术研究、知识管理、技术应用、数据处理、文本分析和内容创作等多个方面提供了高效且创新的解决方案,但在利用 ChatGPT 构建专利知识图谱方面的研究仍需深入。目前,虽有研究探讨了利用 ChatGPT 构建法律案件知识图谱,并利用该图谱评估电信诈骗案例的影响力[19],但其核心在于法律案例影响力的评估,而非知识图谱的构建,且其研究对象是法律案例,与专利的特点存在较大差异。因此,有必要进一步探究 ChatGPT 在专利知识图谱构建中的具体应用效果。

本研究从专利知识图谱自动化构建的角度出发,利用 ChatGPT API 和提示词工程,精准抽取专利实体及其关系,并把 ChatGPT 信息提取能力与其他实体关系抽取技术进行对比,最终实现了专利知识图谱的可视化。通过本研究,期望能揭示 ChatGPT 在专利知识图谱构建中的独特优势和可能存在的局限性,从而为专利分析和知识管理领域提供更全面、高效的技术支持,并为其带来崭新的视角。

2 研究设计

专利知识图谱构建的关键在于从非结构化数据中提取实体和关系。本文利用 ChatGPT API 和提示词来提取信息。研究框架如图 1 所示。

2.1 数据采集与预处理

在知网专利数据库中以“智能驾驶”为关键词选择所有专利数据 1 537 条。去除重复数据后,剩下 1 126 条。专利相关数据有 “作者”“申请人”“题名”“公开号”“公开日期”“摘要”等信息。除“摘要”信息外,都是结构化数据,因此,把“摘要”作为待处理内容,让 ChatGPT 提取信息。

2.2 专利知识图谱构建

2.2.1 ChatGPT API

本文收集了 1 126 条专利摘要,且专利摘要的平均长度为 255.833 字,总长度为 288 324 字,远远超过了 ChatGPT 交互窗口中一次能输入的文本长度。为实现自动化专利摘要信息提取功能,本文通过 API 接口实现与 ChatGPT 的交互[25]。API 编程方式灵活性强,提供了很多自定义选项,提高了对结果的控制能力。模型设置上选择性能较好的“gpt4”模型;为减少随机性,保持一贯性,温度控制设置为 0。

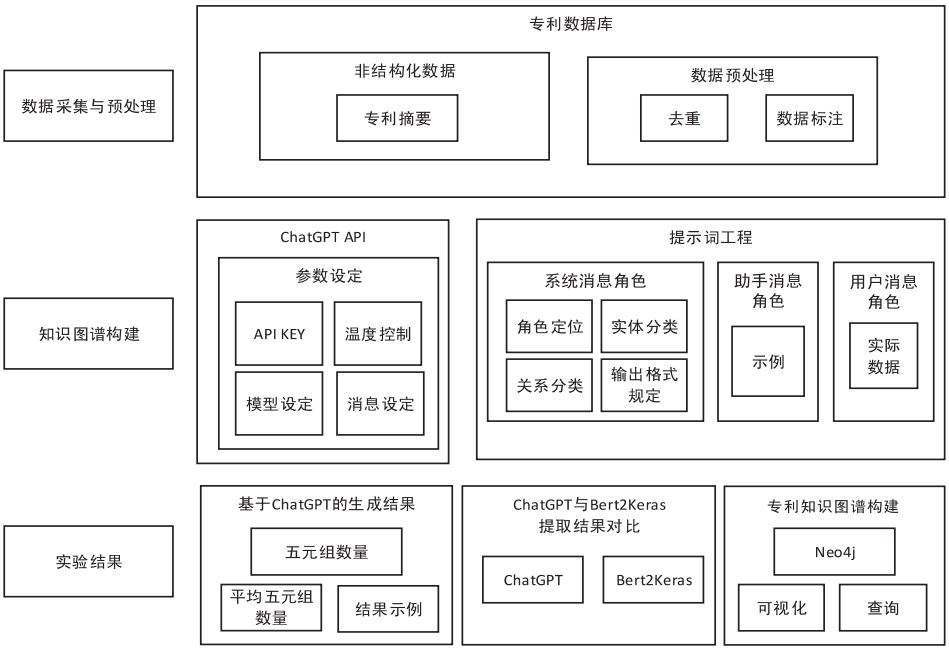

图 1　研究框架

ChatGPT 中 API 使用方法如图 2 所示,使用 Python 语言通过 OpenAI 的 API 与 ChatGPT 模型进行交互。首先,导入 OpenAI 库以便能够调用 API。openai_api_key 变量存储了 API 密钥,这是使用 OpenAI 服务所必需的。在 get_completion 函数中,创建了 OpenAI 的客户端实例,并使用 chat. completions. create 方法发送聊天信息给 ChatGPT 模型,该方法需要指定模型名称、温度参数以及对话消息。对话消息设置了系统消息、助手消息和用户消息三个角色,构建一个模拟对话场景。三个角色的设置有助于提供上下文信息、模拟用户交互,并指导 GPT 模型进行更准确的响应。其中,系统消息为对话提供了背景信息和系统级的指令,有助于模型理解整体任务的框架和内容。助手消息扮演助手角色,提供了 few-shot 示例,指导模型提取信息。用户消息是用户角色,接受用户要发送给 ChatGPT 的文本。这个过程使得用户能够与 ChatGPT 模型进行交互,获取基于输入文本生成回复文本。

```
from openai import OpenAI
openai_api_key ="sk-***"
def get_completion(text，model="gpt-4"):
    client = OpenAI(api_key=openai_api_key)
    response = client.chat.completions.create(
        model=model,
        temperature=0,
        messages=[
            {"role": "system","content": system_message()},
            {"role": "assistant","content": assistant_message()},
            {"role": "user","content": user_message(text=text)}
        ]
    )
    return response
```

图 2　ChatGPT API 使用方法

2.2.2　提示词工程

在构造与 ChatGPT 模型交互的 API 接口函数时,本文设置了系统消息、助手消息和用户消息三个角色。其中,系统消息负责提供对话的背景信息和系统级指令,是提示词工程的主要应用场景;助手消息提供了信息抽取示例,利用 few-shot 示例能显著提高与大语言模型交互的效率和准确率[26]。因此,本文在系统消息中设计了提示词,在助手消息中嵌入了 few-shot 示例,在用户消息中提供了具体的专利摘要。

提示词工程作为一个设计、优化和细化输入提示的过程,其目标在于确保用户意图能够有效传递给如 ChatGPT 这样的大语言模型。它通过设计、改进和实施提示或指令的实践,引导模型输出,以助力完成各种任务[27]。提示词工程对于从模型中获取准确、相关和连贯的回应至关重要,已成为用户充分利用大语言模型并在广泛应用中取得最佳结果的关键技能[28]。在设计系统消息中的提示词时,本文遵循了设定角色、阐述背景、定义目标、给定条件[29]及规定输出格式[28]等多个原则。此外,本文还强调链式思考(Chain of Thought)原则的运用。链式思考是一种引导大型语言模型执行复杂推理任务的有效方法。它通过在提示词中融入逐步解决问题的逻辑,指导模型生成一系列反映问题逻辑结构的中间推理步骤,进而提高答案的准确性,同时使问题解决过程更加透明和可解释[27]。

基于以上原则,本文设计了如图 3 所示的提示词。首先,基于设定角色,阐述背景和定义目标的原则,在任务描述中,将执行任务的角色设定为知识图谱领域的专家,确保任务执行者能从专业的视角出发处理问题。背景被设定为智能驾驶领域的专利摘要处理,为任务提供了明确的专业背景和必要的上下文。定义了一个清晰的目标,即从专利摘要中提取实体和关系,实

现了对任务目标的明确界定。

图 3 提示词工程

其次，基于给定条件原则，为确保实体和关系定义和分类的准确性，充分参考资料，深入了解专利摘要。结合中国汽车工程学术研究综述[30]对智能驾驶的分析，利用 LDA 主题模型[31]，参考智能驾驶领域专家建议把智能驾驶领域的实体分成专利、硬件技术、软件技术、环境感知与定位、决策与规划、运动控制、车路协同、安全技术、人机交互、测试与评价、技术创新、其他实体 12 种不同的实体。

在确定实体类型后，需要明确实体间的关系。根据对专利数据的词频统计结果，专利摘要中包含了 323 种不同的关系表述。为有效归类和整理这些关系，本文采取了以下处理步骤。第一，利用 Word2vec 模型对专利文本进行词向量训练。Word2vec 是一种通过训练神经网络学习词向量表示的模型，能够将词汇转化为向量形式，捕捉词语间的语义关系[32]。第二，基于 Word2vec 的训练结果，计算了关系间的相似度，设定 0.8 为相似度阈值。对相似度高于此阈值的关系进行合并，以减少关系冗余和重叠。第三，采用 K-Means 算法[33]对合并后的关系进行聚类分析。聚类结果如图 4 所示，当聚类数 K 为 6 时，聚类内误差平方和（WCSS）的下降速度显著放缓。

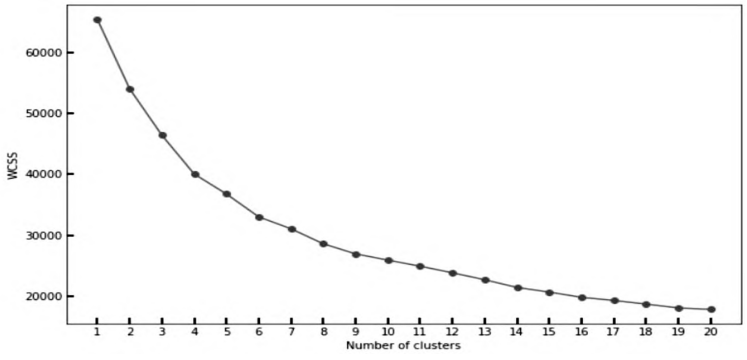

图 4 Kmeans 聚类的结果

根据“肘部法则”，这表明聚类数 K 为 6 时，聚类效果最优，能清晰反映关系的内在结构。但考虑到可能存在未被明确归类的关系或未来潜在的新关系类型，本文增设了一个“其他关系”的类别。最终，确定了 7 种不同的关系类型，分别为组成部分、实现途径、流程步骤、功能用途、物理结构、创新点，以及其他关

系。这些实体和关系的定义和分类作为限定条件,为信息提取和处理提供了明确指导,确保了信息处理的准确性和一致性。

最后,要求以特定 JSON 格式输出提取的实体和关系,规范了数据的表现形式,确保了数据的后续处理和有效利用。

2.3　评价指标

基于 ChatGPT API 和提示词工程提取的五元组信息可靠性与否通过精确率(Precision)、召回率(Recall)和 F1 值来评估,计算公式如式(1)~(3)所示。与一般的分类模型不同,信息提取任务是提取句子中{"subject"、"subject_type"、"object"、"object_type"、"predicate"}这五元组,不是二分类或多分类问题,没有假正例(False Positives)或假反例(False Negatives)。因此,五元组任务的精确率衡量了预测正确的五元组数量占预测总五元组数量的比例;召回率衡量了模型能够正确预测出多少真正的正例;而 F1 值是精确率和召回率的调和平均数,用于综合评估模型的性能,考虑了精确率和召回率的平衡。

$$Precision = \frac{TruePositives}{TotalPredicatedPositives} \tag{1}$$

$$Recall = \frac{TruePositives}{TotalActualPositives} \tag{2}$$

$$\mathrm{F1} = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{3}$$

式中:*TruePositives* 表示真正例;*TotalPredicatedPositives* 表示总的预测正例;*TotalActualPositives* 表示总的实际正例。

3　实验结果

3.1　基于 ChatGPT 的生成结果

基于 ChatGPT 对 1 126 个专利摘要生成了 12 148 个五元组信息,平均每个专利生成了约 11 个五元组,生成的内容较为丰富。生成的部分结果如表 1 所示。从表 1 可知,ChatGPT 生成结果与标注结果在描述上基本一致,ChatGPT 对每个句子都生成了信息,subject_type 和 predicate 的生成结果和标注结果重合度较高。生成结果中的"换道轨迹拟合方程即为本智能驾驶换道轨迹生成方法的最终输出结果"的 object_type 被分为"运动控制",而标注结果中则将其归类为"软件技术"。这可能是因为生成结果更侧重于结果的描述,而标注结果则更强调其在智能驾驶系统中的作用。

表 1　ChatGPT 生成五元组和人工标注五元组

项目	说明
原文	本发明涉及智能驾驶换道轨迹生成方法,包含步骤:判定目标车道;计算换道轨迹拟合方程;换道轨迹拟合方程即为本智能驾驶换道轨迹生成方法的最终输出结果
人工标注五元组	[{"subject":"发明","subject_type":"专利","predicate":"组成部分","object":"智能驾驶换道轨迹生成方法","object_type":"软件技术"},{"subject":"智能驾驶换道轨迹生成方法","subject_type":"软件技术","predicate":"流程步骤","object":"判定目标车道","object_type":"决策与规划"},{"subject":"智能驾驶换道轨迹生成方法","subject_type":"软件技术","predicate":"流程步骤","object":"计算换道轨迹拟合方程","object_type":"决策与规划"},{"subject":"智能驾驶换道轨迹生成方法","subject_type":"软件技术","predicate":"功能用途","object":"换道轨迹拟合方程即为本智能驾驶换道轨迹生成方法的最终输出结果","object_type":"软件技术"}]
GPT 生成五元组	[{"subject":"发明","subject_type":"专利","predicate":"组成部分","object":"智能驾驶换道轨迹生成方法","object_type":"软件技术"},{"subject":"智能驾驶换道轨迹生成方法","subject_type":"软件技术","predicate":"流程步骤","object":"计算换道轨迹拟合方程","object_type":"决策与规划"},{"subject":"智能驾驶换道轨迹生成方法","subject_type":"软件技术","predicate":"功能用途","object":"换道轨迹拟合方程即为本智能驾驶换道轨迹生成方法的最终输出结果","object_type":"运动控制"}]

3.2　ChatGPT 与 Bert2Keras 提取结果对比

为客观评价 ChatGPT 在实体关系抽取任务上的性能,探讨其在构建知识图谱的可行性,本文选择 Bert2Keras 作为基线模型进行对比分析。BERT,作为一种基于 Transformers 架构的预训练模型,已在大规模无标注语料上进行了深入训练,能够捕捉到深层次的语义信息[34]。Bert2Keras 库[35]是基于 Keras 的便捷工具,用于加载和应用 BERT 模型。本文使用 Bert2Keras 库实现实体与关系联合提取,并对比了 ChatGPT 与 Bert2Keras 在精确率(Precision)、召回率(Recall)和 F1 值这 3 个关键指标上的表现。在数据划分方面,把原始数据集按 8∶2 的比例划分,训练集 914 个,测试集 212 个。两个模型在测试集上的对比结果如表 2 所示。由表 2 可知,ChatGPT 在所有 3 个指标上的表现均明显优于 Bert2Keras,特别是在召回率和 F1 两个指标上展现出卓越性能。研究结果表明,ChatGPT 在理解复杂任务以及提取实体和关系方面具备显著优势,因此,利用 ChatGPT 来构建知识图谱是完全可行的。

表 2　ChatGPT 与 Bert2Keras 性能对比

模型	Precison	Recall	F_1
ChatGPT	0.882	0.883	0.883
Bert2Keras	0.341	0.094	0.147

3.3　专利知识图谱构建

使用 ChatGPT 进行实体关系抽取,得到 12 148 个包含 subject、subject_type、object、object_type、predi-

cate 信息的五元组。为实现专利核心信息的可视化，从五元组中提取 subject、predicate、object 三元组，剔除类别信息，只保留关键的实体和关系信息。随后，把三元组转换为 csv 格式，导入到 Neo4j，得到专利知识图谱（如图 5 所示）。专利包含了丰富的实体和关系，通过知识图谱将实体信息通过关系链接起来，形成一个直观的网络结构。利用图谱，可以更容易地发现数据之间的隐藏模式和关系。此外，Neo4j 支持 Cypher 语句高效进行图遍历操作，对于复杂的查询特别有优势。传统的关键词搜索可能不足以准确反映复杂的查询意图，而知识图谱能提供语义上的联系和上下文，帮助改进搜索结果。

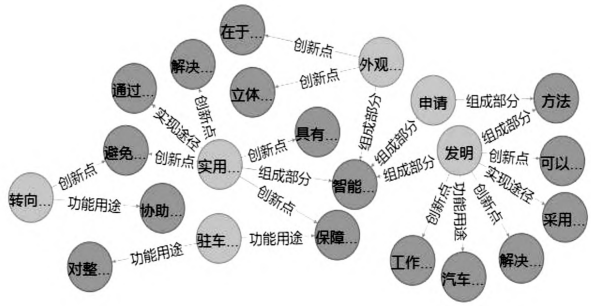

图 5 专利知识图谱

本文利用 ChatGPT 对专利文献进行深度理解、抽取关键信息并构建专利知识图谱，实验结果显示，ChatGPT 在概念识别以及非结构化数据生成等方面表现出显著优势。相较于 Bert2Keras 的传统方法，ChatGPT 能够更准确地理解和提取专利文档中的创新点，实现途径，组成部分，功能用途等多元信息，利用 ChatGPT 能高效构建出高质量的知识图谱。尤其在处理复杂专利描述文本时，其强大的自然语言处理能力有助于减少人工干预，提高了知识图谱构建的自动化程度和准确性。

4 结 语

本文利用 ChatGPT API 和提示词工程成功构建了专利知识图谱，结果表明 ChatGPT 拥有出色的语义理解和知识建模能力，非常适用于专利知识图谱的自动化构建。这一成果不仅为大型语言模型的应用开辟了新领域，还为专利信息的有效整合与利用提供了创新思路和技术支持。借助 ChatGPT 的强大功能，专利知识图谱构建技术将得到显著提升，进而推动知识产权管理、技术研发及市场竞争情报分析等相关工作的智能化发展，为企业和科研机构制定更精准的战略决策提供有力支撑。尽管 ChatGPT 在构建专利知识图谱方面取得了积极成果，但仍存在一些挑战。首先，为了更充分地利用和挖掘专利知识图谱的价值，需要将其应用进一步扩展到个性化推荐、知识问答、知识推理等更多场景。其次，由于专利文献具有高度的专业性和多样性，这就要求模型必须具备更强的领域适应能力。此外，对于模糊表述或隐含关系的深入挖掘，以及大规模专利知识图谱的实时更新问题，也亟待解决。针对这些挑战，未来的研究方向可以考虑结合预训练模型与领域专业知识进行微调，研发更为精确的上下文感知关系抽取算法，并探索增量式图谱构建策略，以不断提升专利知识图谱的构建质量与应用效果。

参 考 文 献

[1] Siddharth L, Li G, Luo J. Enhancingpatent retrieval using text and knowledge graph embeddings: A technical note[J]. Journal of Engineering Design, 2022, 33(8/9): 670-683.

[2] 吕向如. 中文专利知识图谱构建研究[D]. 北京：北京信息科技大学, 2019.

[3] 赖朝安, 钱 娇. 基于知识图谱的专利挖掘方法及其应用[J]. 科研管理, 2017, 38(1): 333-341.

[4] 马国斌. 基于知识图谱的专利知识检索研究[D]. 哈尔滨：哈尔滨工业大学, 2021.

[5] 邵泽宇, 孟天宇. 基于知识图谱的区块链专利数据挖掘[J]. 技术与创新管理, 2020, 41(6): 588-595.

[6] 曹树金, 李睿婧. 基于专利文献摘要的创新知识图谱构建与应用[J]. 情报理论与实践, 2022, 45(11): 21-28.

[7] 何 玉, 张晓冬, 郑 鑫. 基于 SpERT-Aggcn 模型的专利知识图谱构建研究[J]. 数据分析与知识发现, 2024, 8(1): 146-156.

[8] OpenAI. Chat GPToverview[EB/OL]. [2024-03-29]. https://openai.com/research/overview.

[9] 郭 鑫, 王一博, 王继民. ChatGPT 生成中文学术内容分析——以情报学领域为例[J]. 图书馆论坛, 2024, 44(3): 134-143.

[10] 王一博, 郭 鑫, 刘智锋, 等. AI 生成与学者撰写中文论文摘要的检测与差异性比较研究[J]. 情报杂志, 2023, 42(9): 127-134.

[11] Gao C A, Howard F M, Markov N S, et al. Comparingscientific abstracts generated by ChatGPT to original abstracts using an artificial intelligence output detector, plagiarism detector and blinded human reviewers[J]. BioRxiv, 2022: 2022.521610.

[12] 王雅琪, 曹树金. ChatGPT 用于论文创新性评价的效果及可行性分析[J]. 情报资料工作, 2023, 44(5): 28-38.

[13] 吴 娜, 沈 思, 王东波. 基于开源 LLMs 的中文学术文本标题生成研究—以人文社科领域为例[J]. 情报科学, 2024, 42(7): 137-145

[14] 李侗桐, 高瑞婧, 田 佳. ChatGPT 在中文科技期刊摘要文字编辑中的实用性测试与分析[J]. 中国科技期刊研究, 2023, 34(8): 1014-1019.

[15] Dai H, Liu Z, Liao W, et al. Auggpt: Leveraging ChatGPT for text data Augmentation[J]. ArXiv Preprint ArXiv, 2023: 2302.13007.

[16] 张 恒, 赵 毅, 章成志. 基于 SciBERT 与 ChatGPT 数据增强的研究流程段落识别[J]. 情报理论与实践, 2024, 47(1): 164-172, 153.

[17] 商锦铃, 张建勇. 基于 ChatGPT 和提示工程的查询式摘要数据集 AMTQFSum 构建研究[J]. 数据分析与知识发现,2024, 8(Z1):122-132.
[18] 曹茹烨, 曹树金. ChatGPT 完成知识组织任务的效果及启示[J]. 情报资料工作, 2023, 44(5): 18-27.
[19] 裴炳森, 李　欣, 吴　越. 基于 ChatGPT 的电信诈骗案件类型影响力评估[J]. 计算机科学与探索, 2023, 17(10): 2413-2425.
[20] Amin M M, Cambria E, Schuller B. Willaffective computing emerge from foundation models and general AI: A first evaluation on ChatGPT [J]. ArXiv Preprint ArXiv, 2023: abs/2303.03186.
[21] Zhang B, Ding D, Jing L. Howwould stance detection techniques evolve after the launch of ChatGPT? [J]. ArXiv Preprint ArXiv, 2022:2212.14548.
[22] Huang F, Kwak H, An J. Is ChatGPTbetter than human annotators? potential and limitations of ChatGPT in explaining implicit hate speech [C]//Companion Proceedings of the ACM Web Conference 2023, 2023: 294-297.
[23] 白如江, 陈启明, 张玉洁, 等. 基于 ChatGPT+Prompt 的专利技术功效实体自动生成研究[J]. 数据分析与知识发现, 2024,8(4):14-25.
[24] 田萍芳,刘恒永,高　峰,等.基于大语言模型的本体提示指导的司法命名实体识别[J/OL].武汉大学学报(理学版),2024:1-13[2024-06-19]. https://doi.org/10.14188/j.1671-8836.2024.0027.
[25] OpenAI API. Introduction docs for openAI aPI[EB/OL].[2024-03-25]. https://platform.openai.com/docs/introduction.
[26] Brown T, Mann B, Ryder N, et al. Languagemodels are few-shot learners [J]. Advances in Neural Information Processing Systems, 2020, 33: 1877-1901.
[27] Wei J, Wang X, Schuurmans D, et al. Chain-of-thought prompting elicits reasoning in large language models [J]. Advances in Neural Information Processing Systems, 2022, 35: 24824-24837.
[28] Ekin S. Promptengineering for ChatGPT: A quick guide to techniques, tips and best practices [J/OL]. [2024-03-29]. https://www.techrxiv.org/doi/full/10.36227/techrxiv.22683919.v2.
[29] Meskó B. Promptengineering as an important emerging skill for medical professionals: Tutorial[J]. Journal of Medical Internet Research, 2023, 25: e50638.
[30] 《中国公路》编辑部. 中国汽车工程学术研究综述·2023[J]. 中国公路学报,2023,36(11):1-192.
[31] Blei D M, Ng A Y, Jordan M I. Latentdirichlet allocation[J]. Journal of Machine Learning Research, 2003, 3(1): 993-1022.
[32] Mikolov T, Chen K, Corrado G, et al. Efficientestimation of word representations in vector space[J]. ArXiv Preprint ArXiv, 2013:1301.3781.
[33] Kodinariya T M, Makwana P R. Review ondetermining number of cluster in K-means clustering [J]. International Journal, 2013, 1(6): 90-95.
[34] Devlin J, Chang M W, Lee K, et al. BERT: Pre-training of deep bidirectional transformers for language understanding [J]. ArXiv Preprint ArXiv, 2018:1810.04805.
[35] SU J. Bert4Keras[EB/OL].[2024-03-25]. https://github.com/bojone/bert4keras

(责编:王育英;校对:刘影梅)

(上接第 152 页)

[6] 苗　艳. 从"媒介化"到"事件化":媒介事件研究范式的拓展与变化 [J]. 西南民族大学学报(人文社科版), 2017, 38(4): 159-163.
[7] 彭　兰.碎片化社会背景下的碎片化传播及其价值实现[J].今传媒, 2011, 10(19):15-17.
[8] 丹尼尔·戴杨,伊莱休·卡茨.媒介事件:历史的现场直播[M].北京:北京广播学院出版社,2000:3-8.
[9] 邱林川,陈韬文.迈向新媒介事件研究[J].传播社会学刊,2009(9):19-37.
[10] 苏　晨,许永超. 媒介事件的本土化:中国内地媒介事件研究再回顾[J]. 未来传播, 2019, 26(2): 72-77.
[11] Katz E, Liebes T. How disaster, terror and war have upstaged media events[J]. International Journal of Communication, 2007(1):157-166.
[12] 丹尼尔·戴扬,邱林川,陈韬文."媒介事件"概念的演变[J].传播与社会学刊,2009(9):1-17.
[13] 周　翔,李　镓. 网络社会中的"媒介化"问题:理论、实践与展望 [J]. 国际新闻界, 2017, 39(4): 137-154.
[14] 自国天然. 情之所向:数字媒介实践的情感维度 [J]. 新闻记者, 2020(5): 41-49.
[15] Fuller. The new media event:discourse,publics and celebrity fandom as connective action [J]. Communication Research and Practice,2018,4(2):167-182.
[16] 柳红兵. 新媒介事件的传播机制研究[D].西安:西北大学,2011.
[17] 方　洁. 被裹挟与被规制:从新媒体与大众媒体的框架建构看新媒介事件的消解 [J]. 国际新闻界, 2014, 36(11): 6-18.
[18] 梁建恕.融合传播视阈下的"新媒介事件"策略——以紫金山新闻客户端建设云上交互平台为例[J].城市党报研究,2022(8):47-50.
[19] 荆学民. 微观政治传播论纲 [J]. 现代传播(中国传媒大学学报), 2021, 43(7): 16-27.

(责编:王育英;校对:王菊)