doi:10.6041/j.issn.1000-1298.2026.17.028

基于大模型驱动的水稻种植领域知识图谱构建

陈　明[1,2]　郭钦鸿[1]　吴桓宇[1]

(1.上海海洋大学信息学院,上海 201306;2.农业农村部渔业信息重点实验室,上海 201306)

摘要:传统水稻种植知识管理存在知识碎片化、检索困难和更新滞后等问题,难以满足现代精准农业的信息需求。为此提出了基于大模型驱动的水稻种植领域知识图谱构建方法。构建了包含 51 篇期刊论文和 2 本专业书籍的 RC-PKBD 数据集(40.98 万字符),基于改进的 LightRAG 框架设计专业提示工程策略,定义 15 类实体类型和 12 类关系类型,实现专业知识的高效抽取。建立了基于图结构的双层次检索机制,成功构建包含 10 723 个节点和 10 955 条边的知识图谱。质量评估显示,具体型问题准确性达 91.73%,抽象型问题覆盖度达 90.20%,相比 GraphRAG 方法准确性提升 36.60 个百分点。本研究为水稻种植领域知识数字化提供了有效解决方案,对推动智慧农业发展具有重要价值。

关键词:水稻种植;大语言模型;知识图谱;图检索;增量更新

中图分类号:TP391　　**文献标识码**:A　　**文章编号**:1000-1298(2026)17-0299-13　　**OSID**:

Large Model-driven Knowledge Graph Construction for Rice Cultivation Domain

Chen Ming[1,2]　Guo Qinhong[1]　Wu Huanyu[1]

(1. *College of Information Technology*, *Shanghai Ocean University*, *Shanghai* 201306, *China*

2. *Key Laboratory of Fisheries Information*, *Ministry of Agriculture and Rural Affairs*, *Shanghai* 201306, *China*)

Abstract: Traditional rice cultivation knowledge management faces challenges, including knowledge fragmentation, retrieval difficulties, and outdated information, which cannot meet the demands of modern precision agriculture. A large language model-driven knowledge graph construction method for rice cultivation domain was proposed. The RC-PKBD dataset containing 409 800 Chinese characters from 51 journal articles and 2 professional books were constructed. Based on an improved LightRAG framework with domain-specific prompt engineering strategies, totally 15 entity types and 12 relationship types were defined to achieve efficient extraction of professional knowledge. The construction process involved four key steps: text processing and chunking, entity and relationship extraction using large language models, entity deduplication and alignment through vector similarity calculation, and graph structure construction with storage. A dual-level retrieval mechanism based on graph structure was established, successfully constructing a knowledge graph containing 10 723 nodes and 10 955 edges. The system demonstrated significant advantages over traditional deep learning methods in semantic understanding depth and domain adaptability. Quality evaluation showed that accuracy reached 91.73% for specific questions and coverage achieved 90.20% for abstract questions, with a 36.60 percentage points improvement in accuracy compared with GraphRAG. The research result can provide an effective solution for rice cultivation knowledge digitalization and has important value for advancing smart agriculture development, while offering a replicable technical approach for knowledge services in other agricultural domains.

Key words: rice cultivation; large language models; knowledge graph; graph retrieval; incremental updating

收稿日期:2025-04-29　修回日期:2025-09-02

基金项目:中国(上海)自由贸易试验区临港新片区专项发展资金项目(SH-LG-GK-2020-02-19)

作者简介:陈明(1966—),男,教授,博士,主要从事物联网技术、农业信息技术和知识图谱等研究,E-mail:mchen@shou.edu.cn

0 引言

随着人工智能技术的迅猛发展,大型语言模型(Large language model, LLM)在各领域应用日益广泛[1]。作为全球主要粮食作物之一[2],水稻种植知识的数字化与智能化表示对保障粮食安全和提高生产效率具有战略意义[3]。传统水稻种植知识管理主要依赖纸质资料和经验传承,存在知识碎片化、检索困难和更新滞后等问题,难以满足现代精准农业对个性化信息服务的迫切需求[4]。构建水稻种植领域知识图谱,实现专业知识的系统化组织与智能检索,对推动智慧农业发展具有实践价值。

知识图谱概念由 Google 于 2012 年首次提出[5],旨在构建描述现实世界实体及关系的语义网络。经过十余年发展,知识图谱已经从通用知识库扩展至专业领域,在教育[6]、工业[7]、医疗[8]和农业[9]等领域都有很广的应用。在农业领域,知识图谱技术已在多个子领域展现出强大的应用潜力。从应用层面看,农业知识图谱主要涵盖 5 个核心方向:信息检索、智能问答系统、推荐系统、专家诊断系统和作物预测[9]。在病虫害防控方面,已有研究构建了基于深度学习的作物病虫害可视化知识图谱,实现病虫害智能诊断和防治推荐[10];在作物管理方面,玉米、小麦等主要作物的知识图谱构建技术逐渐成熟[11];在花卉园艺领域,也出现了专门的病虫害知识管理系统[12];在农用电机故障诊断方面,也已有研究构建了面向故障识别与诊断的知识图谱[13]。

然而,在水稻种植这一重要领域,现有研究仍显不足。虽然宋璐璐[14]利用 BERT - BiLSTM - CRF 组合模型构建了水稻病虫害的知识图谱;高锐涛等[15]也利用相同的方法构建了水稻种植知识图谱,但该方法主要局限于传统深度学习技术,在处理专业农业文献时仍面临语义理解不足、复杂关系抽取不完整等挑战。此外,现有水稻种植知识图谱研究多集中在特定子问题,缺乏对水稻全生命周期知识的系统性整合。

近年来,大模型的快速发展为知识图谱构建提供了新路径。与传统方法相比,大模型在农业垂直领域具有独特优势。首先,强大的语义理解能力可以准确识别农业专业术语。其次,多轮对话机制能够处理复杂的知识抽取任务。此外,零样本学习能力有效降低了领域适应成本。

在技术框架方面,Guo 等[16]提出的轻量级检索增强生成(Light retrieval-augmented generation, LightRAG)框架通过引入图结构到检索过程,有效提升了复杂信息检索精度;微软设计的图检索增强生成(Graph retrieval-augmented generation, GraphRAG)[17]利用社区结构增强检索效果;ToG(Think-on-Graph)系列技术[18-20]实现了动态检索与知识整合。

目前,国内也有相关学者利用大模型对专业领域的知识构建了知识图谱:韦一金等[21]利用 GraphRAG 构建了中国马铃薯新品种的知识图谱;朱晓俊等[22]也利用大模型构建了机器人研制场景的知识图谱。但目前在水稻种植领域还没有相关的专业研究。

水稻种植知识具有显著的跨学科性和地域差异性,涉及农学、生物学、气象学等多个领域,同时包含大量生产实践经验[23]。这种多源异构的知识特性对知识库构建提出了更高要求,需要有效整合不同来源的专业知识,建立它们之间的语义关联。当前水稻种植领域面临的关键挑战是:如何利用大模型的强大语义理解能力,从非结构化文本中高效提取专业实体与关系,构建系统化、规范化的知识表示框架,并支持知识的动态更新与智能检索[24]。鉴于此,本文旨在解决水稻种植领域知识分散及非结构化数据利用难的问题。主要研究内容包括:整合多源数据,构建高质量的水稻种植领域专用数据集;提出一种融合大语言模型语义理解能力的知识图谱构建与检索框架(如 LightRAG),优化从非结构化文本到结构化知识的转化流程;通过对比实验与消融实验,验证所提方法在复杂农业问答任务中的准确性与鲁棒性。

1 研究方法

1.1 数据集构建

为确保构建的知识库能够准确、全面地反映水稻种植领域的专业知识,本研究设计水稻种植专业知识库数据集(Rice cultivation professional knowledge base dataset, RC - PKBD)时遵循权威性、全面性、实用性和时效性原则。权威性原则确保选取行业公认的权威资料作为数据来源;全面性原则保证覆盖水稻种植全生命周期的各环节;实用性原则注重选取具有实际指导意义的内容;时效性原则兼顾基础理论与前沿技术发展。

本研究在数据集构建中主要采用“时效性与传统性并重”的策略,以确保知识图谱既能反映最新技术发展又能保留经典栽培经验。数据来源按知识类型分层选择:基础理论知识(40%)主要来源于《水稻高产栽培技术》[25-26]等经典专业书籍,涵盖水稻生物学特性、基本栽培原理等稳定性知识;技术应用知识(35%)选择 2022—2024 年期刊文献,重

点关注新品种、新技术和管理创新；实践经验知识（25%）兼顾传统农艺经验和现代实证研究。通过这种分层构建方式，RC－PKBD 数据集实现了传统知识完整性与技术前沿性的有机结合，字符分布为传统基础知识 16.39 万字符、近期技术文献 14.34 万字符、实践经验知识 10.25 万字符，确保知识图谱既有深厚的理论根基又具备良好的技术时效性。

本研究采用系统化的三阶段筛选流程：

初筛阶段：以“水稻种植”“水稻栽培”“水稻管理”等为关键词，在中国知网核心期刊数据库中检索相关文献，获得候选文献 680 篇。

精筛阶段：基于以下标准进行筛选：内容权威性——优先选择国家级科研机构、重点高校和行业领军专家发表的成果；实用指导性——侧重于对实际生产具有直接指导价值的内容；技术先进性——关注能够提高水稻产量、质量和资源利用效率的创新技术；叙述规范性——选择结构清晰、专业术语使用准确的文本资料。经此阶段筛选，保留文献 156 篇。

终筛阶段：通过专家评审和内容重复性检查，最终确定 51 篇核心期刊论文作为研究语料。

通过系统筛选和整理，本研究共收集了 51 篇核心期刊论文和 2 本专业书籍，形成了包含 40.98 万字符的综合语料库。从内容覆盖角度，数据集包含水稻种植全过程的关键环节，具体涵盖品种选择与育苗技术、土壤准备与肥料管理、水分管理与灌溉技术、病虫害识别与防治、生长调节与产量提升、收获与储存处理以及绿色有机种植技术等方面。

为确保数据质量，本研究采取了多源交叉验证、数据清洗与预处理、数据更新机制以及数据标注质量控制等措施。多源交叉验证对同一知识点从不同来源进行比对，确保内容一致性；数据清洗与预处理去除重复内容、纠正错误信息、标准化格式；数据更新机制建立定期更新流程，确保数据的时效性；数据标注质量控制则采用双人标注与第三方审核相结合的方式，保证标注一致性和准确性。

1.2　基于 LightRAG 的图结构知识表示方法

LightRAG 在处理实体之间复杂关系方面具有明显优势。其核心原理是利用图结构来捕捉和表示实体间的多维关联，通过双层检索范式增强信息获取的全面性。图结构文本索引将非结构化文本转化为包含实体和关系的图结构；双层次检索策略结合低层次实体检索和高层次主题检索，实现全面信息获取；图向量表示的集成则将图结构与向量表示相结合，提升检索效率。

水稻种植领域知识具有高度的专业性、系统性和关联性，这与图结构的特点高度契合。图结构能够自然地表示水稻品种、生长阶段、环境因素、管理措施等实体，以及它们之间的各种复杂关系。

1.3　实验方法和领域适配性优化

本研究基于 LightRAG 框架，设计了一套适合水稻种植领域的知识图谱构建方法，整体流程如图 1 所示。

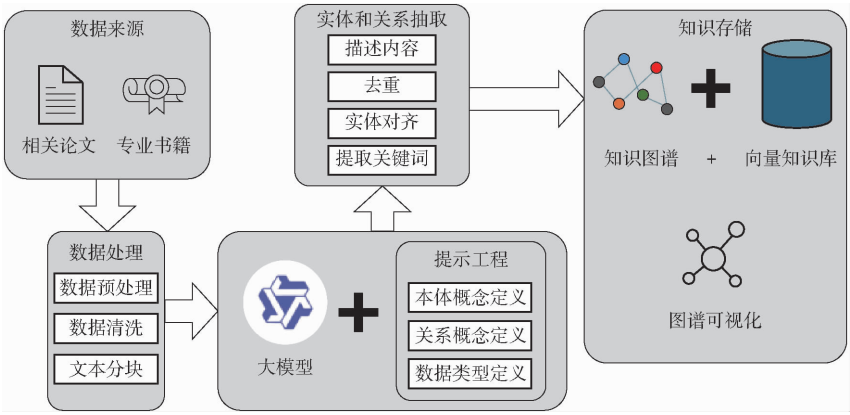

图 1　基于改进 LightRAG 的水稻种植领域知识图谱构建流程图

Fig. 1　Improved LightRAG-based rice cultivation knowledge graph construction flowchart

流程主要包括 4 个关键步骤：

（1）文本处理与分块：将水稻种植领域文献按照语义完整性原则分割为适当大小的文本块，便于后续处理。

（2）实体与关系抽取：利用专业提示工程设计，引导大模型从文本块中识别水稻种植领域的关键实体和关系，包括品种、技术、环境、病虫害等实体类型，以及它们之间的适宜性、影响、应用等关系类型。

（3）实体对齐与去重：对提取的实体和关系进行去重和对齐处理，基于向量相似度和专业知识合并语义相同但表达不同的实体，建立规范化的知识表示。

（4）图结构构建与存储：将处理后的实体和关系构建为完整的知识图谱，并通过 GraphML 格式存储，同时将实体和关系的向量表示存储在向量数据库中，支持后续的检索应用。

基于图结构的知识表示具有三大优势:将分散的水稻种植知识连接成有机整体,表达实体间因果关系和条件依赖;支持从品种、技术、环境等多角度组织和检索知识;支持新知识动态融合,无需重建整个索引。

1.3.1 模型配置与参数设定

本研究采用 Qwen 2.5 - 72B - Instruct 作为核心抽取大语言模型,选取 BGE - M3(BGE - M3 - Embedding)[27]作为嵌入模型。具体的相关配置见表 1。

表 1 大语言模型和嵌入模型主要配置

Tab.1 Main configurations of LLM and embedding model

配置类别	参数名称	参数值	说明
大语言模型配置	Temperature	0.3	保证输出的稳定性和一致性
	Max Tokens	4 096	满足复杂农业文本处理需求
	Top-p	0.9	平衡输出多样性与准确性
	Frequency Penalty	0.1	减少重复输出
嵌入模型配置	向量维度	1 024	保证语义表示精度
	批处理大小	32	平衡处理效率与内存占用
	相似度阈值	0.85	用于实体去重与对齐

1.3.2 水稻种植领域提示词设计

针对农业文本的专业性和复杂性,本研究设计了多层次提示词策略。

基于水稻种植全生命周期,定义 15 类核心实体以及 12 类专业关系类型。并且通过多轮对话抽取策略,设计“实体识别→关系抽取→一致性检查→补充抽取”的 4 步骤提示流程,要求模型确保每个实体至少与一个其他实体建立关系,避免孤立节点。并且采用了专业术语引导策略和示例驱动学习,提供了完整的抽取示例;同时设计了容错与扩展机制,允许模型抽取除预定义类型之外的实体和关系,以应对农业领域知识的复杂性。

通过上述提示词设计策略,确保了对水稻种植专业知识的高质量抽取。

1.3.3 农业场景适配性优化

针对水稻种植领域术语多样性,建立专业词典并设计同义词识别规则。例如,“分蘖期”“分蘖阶段”“分蘖生长期”统一为“分蘖期”。

相比通用知识图谱,农业知识具有明显的时序性、地域性和条件依赖性特点。本研究特别设计了 3 种条件约束关系表示:时序约束,如“播种→育苗→移栽→分蘖→抽穗→成熟”的生长阶段序列;地域约束,如品种与气候区域的适应性关系;条件约束,如病虫害与环境条件的因果关系。

另外,考虑到农业技术快速发展,也设计了专门的知识冲突检测与解决机制。当新旧知识冲突时,综合考虑信息来源权威性、时间新近性和内容完整性进行融合决策。

2 方法与过程

2.1 文本处理与分块

文本处理与分块是知识图谱构建的基础环节,直接影响后续实体抽取的质量和效率。本研究采用以下策略进行文本处理与分块。

格式标准化处理:对原始文档进行统一的格式标准化处理,包括将其统一转为 txt 文件、去除冗余的格式标记、统一标点符号、规范化数字表示等。对于表格和图像中的关键信息,通过 OCR(Optical character recognition)技术提取并转换为结构化文本,确保重要的视觉信息不会在处理过程中丢失。

文档分块策略:根据语义完整性原则,将原始文本分割为 500 ~ 1 000 词的段落单元。这一规模既保证了单个块内语义的完整性,又控制了信息密度,便于大模型进行有效处理。在分块过程中,优先以自然段落为界,同时考虑章节标题、小标题等结构元素,确保每个块具有相对独立且完整的语义。

通过上述处理,原始的非结构化水稻种植文本被转化为标准化、分块化的文本单元,为下一步的实体与关系抽取奠定了基础。

2.2 实体与关系抽取

实体与关系抽取是知识图谱构建的核心环节,直接决定了知识表示的质量和完整性。本研究设计了系统的实体类型体系和关系类型体系,并采用大模型驱动的抽取技术进行高质量的知识提取。

基于对水稻种植领域知识结构的分析,本研究定义了 15 类核心实体类型,覆盖了水稻种植全过程的关键要素,具体内容见表 2。

对于每类实体,进一步定义了其属性详细描述体系,如水稻品种的生育期长度、产量潜力、抗性特点等,使实体表示更加丰富和精确。为描述实体间的多样化联系,本研究定义了 12 类核心关系类型,具体内容见表 3。

每种关系类型都定义了描述信息、强度、关键词等属性,增强了关系表示的精确性。

基于水稻种植领域特性,设计了专门的提示模板,引导大模型准确识别专业实体和关系。并且使用大模型为抽取的实体和关系生成详细描述,丰富知识表示。模型不仅需识别实体和关系,还需生成

表 2　实体类型及其定义

Tab. 2　Entity types and their definitions

实体类型	实体定义
作物品种	各类水稻品种
生长阶段	水稻生长的各个阶段
种植技术	水稻种植采用的各类农艺技术
设备工具	水稻种植过程中使用的机械设备和工具
环境因素	影响水稻生长的环境条件
病虫害	水稻常见的病害和虫害
营养元素	稻生长所需的各类营养物质
水分管理	水稻种植过程中的水分管理措施
地理位置	与水稻种植相关的地域信息
相关机构	与水稻研究、推广相关的组织机构
产量指标	衡量水稻产量的各项参数
研究方法	水稻研究的科学方法
时间周期	水稻生长或农事活动的时间段
经济效益	水稻种植的经济收益参数
土壤特性	水稻种植土壤的物理化学特性

表 3　关系类型及其定义

Tab. 3　Relationship types and their definitions

关系类型	关系说明
品种特性	水稻品种与其特有特性间的关系
生长阶段	水稻不同生长阶段之间的先后顺序及其管理要求的关系
技术应用	种植技术对水稻生长、产量或品质的影响关系
环境适应	水稻品种与特定环境条件的适应性关系
病虫害防控	病虫害与防控措施、症状或影响之间的关系
营养需求	水稻与特定营养元素需求、缺乏症状或施肥管理间的关系
水分管理	水稻生长阶段与水分需求或管理策略间的关系
地域适应	水稻品种与适宜种植区域间的关系
产量构成	水稻产量与其构成要素间的关系
季节时序	水稻种植与季节、时间节点的关系
经济效益	水稻品种或技术与经济收益间的关系
土壤影响	土壤特性对水稻生长、管理措施的影响关系

结构化的描述内容，包括核心定义、关键特征和典型应用。提示模板包含领域知识引导、任务定义、实体类型说明、样例展示等要素，有效提升了大模型对农业领域专业术语的理解能力。

设计的提示词结构如图 2 所示，主要包含 4 个核心组成部分：①任务定义与背景说明：明确告知模型当前任务是从水稻种植文本中抽取实体和关系，并提供领域背景知识，使模型理解农业文本的专业特性。②实体类型定义：详细列出 15 种预定义实体类型（如作物品种、生长阶段、种植技术等），并给出每种类型的具体定义和典型示例，引导模型准确识别不同类型的农业实体。③关系类型说明：阐述 12 种关系类型（如技术应用、环境适应、病虫害防控等）的含义和判断标准，帮助模型理解实体间的语义关联。④输出格式规范：规定结构化的 JSON 输

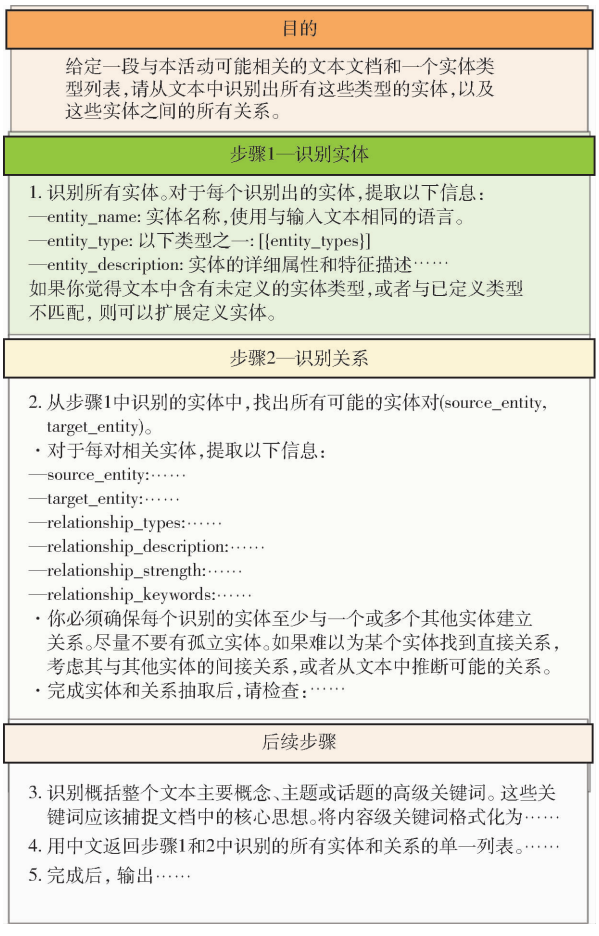

图 2　提取实体和关系的部分提示词

Fig. 2　Extract partial prompt words for entities and relationships

出格式，确保抽取结果的一致性和可处理性，便于后续的知识图谱构建。这种结构化的提示词设计充分利用了大模型的上下文学习能力，通过领域知识引导和示例驱动，显著提升了水稻种植专业知识的抽取质量。

另外，实验允许大模型抽取除了定义之外的实体和关系。因为考虑到数据集的广度和深度，在面对 40 多万字符庞大的数据集时，难免会遗漏部分信息。后面的实验也证明了这一举措的重要性。在抽取实体和关系的同时，也构建了相应的向量表示，实现了双重索引机制。使用嵌入模型 bge－m3（BGE M3－Embedding）将实体和关系转换为高维向量表示，支持语义相似度计算和相关性检索：实体向量←嵌入模型（实体名称＋实体描述）；关系向量←嵌入模型（关系描述＋相关实体信息）。这些向量表示被存储在向量数据库中，用于后续的语义相似度检索。与传统仅基于关键词匹配的检索方法相比，向量检索能够捕捉语义层面的隐式联系，大幅提高检索的召回率和相关性。实体和关系的向量 $\boldsymbol{v}_e$、$\boldsymbol{v}_r$ 表示计算公式为

$$\boldsymbol{v}_e = \mathrm{Encoder}(\mathrm{name}(e) + \mathrm{description}(e)) \quad (1)$$

$$\boldsymbol{v}_r = \text{Encoder}(\text{description}(r) + \text{source}(r) + \text{target}(r)) \tag{2}$$

式中 Encoder——预训练语言模型编码器

name(e)——实体名称

description(e)——实体描述

source(r)、target(r)——关系的源实体、目标实体

2.3 去重与实体对齐

在多源文本知识抽取过程中,同一实体或关系可能以不同表述形式多次出现,导致知识冗余和不一致。通过计算实体间的余弦相似度 s 来识别潜在的重复项,即

$$s(e_i, e_j) = \cos(\boldsymbol{v}_{e_i}, \boldsymbol{v}_{e_j}) \tag{3}$$

式中,$\boldsymbol{v}_{e_i}$ 和 $\boldsymbol{v}_{e_j}$ 分别为实体 e_i 和 e_j 的向量表示,通过将实体名称和描述输入到语言模型中获得。当相似度超过预设阈值 θ(本研究设定为 0.85)时,系统将这些实体标记为潜在重复项。

对于识别出的同义实体,采用综合合并策略。首先从多个同义表述中选择最规范、最常用的名称作为主要标识;然后整合各表述中的属性信息,形成更完整的实体描述;最后将同义实体的关系转移到合并后的实体,并处理可能的关系冲突。为确保溯源性,保留了所有原始表述及其出处,便于后续查证。

本研究基于 LightRAG 原有的去重机制 $D(\cdot)$ 进行水稻种植领域的适应性优化。该机制在原始 LightRAG 中定义为

$$\hat{D} = (\hat{V}, \hat{E}) = \text{Dedupe} \circ \text{Prof}(V, E) \tag{4}$$

其中,$\hat{D}$ 表示经过去重后的知识图谱,$\hat{V}$ 和 $\hat{E}$ 分别表示去重后的实体节点集合和关系边集合。V 和 E 表示原始的实体节点和关系边集合。Dedupe 代表去重函数,用于识别并合并相同或高度相似的实体和关系;Prof 代表剖析函数,用于为每个实体和关系生成详细描述和特征。运算符 $\circ$ 表示函数组合,即先执行 Prof,再执行 Dedupe。

对该机制进行了领域适应性优化,主要包括:引入水稻种植领域词典,提高专业术语归一化的准确性;考虑水稻种植专业术语在不同上下文中的特定含义,避免错误合并;根据实体类型和重要性,采用不同的合并标准和策略。例如,对于水稻品种这类核心实体,设定了更严格的合并条件,要求不仅名称相似,其关键特性(如生育期长度、抗性特点)也需具有高度一致性;而对于一般性描述实体,则采用相对宽松的合并标准。

2.4 图结构构建与存储

完成实体与关系抽取、去重与对齐后,进入图结构构建与存储阶段,此阶段将结构化知识元素转化为完整的知识图谱,并实现高效存储与访问。

在节点构建过程中,将识别出的实体作为图的节点,每个节点包含实体名称、类型、描述等属性。同时,为每个节点分配唯一标识符,便于后续的检索和引用。根据抽取的关系信息构建图的边,每条边代表两个实体间的语义关联,包含关系类型、描述、强度等属性。

在图数据存储方面,实体和关系的向量表示被存储在向量数据库中,支持相似度检索。

使用 Neo4j 图数据库进行知识图谱的管理和可视化,充分利用其高效的图遍历和查询能力。Neo4j 提供了强大的 Cypher 查询语言,能够方便执行复杂的图模式匹配和路径分析[28],为后续的检索和问答提供支持。

2.5 增量更新

水稻种植技术不断发展,新的研究成果和栽培经验持续产生,这要求知识库具备高效的增量更新能力。基于 LightRAG 框架的增量更新算法理念,设计了适合水稻种植领域的知识更新机制,允许系统在不重建整个索引的情况下整合新数据。

当有新的水稻种植文献需要整合时,系统首先对新文献执行与初始构建相同的处理流程,生成新的实体、关系及其向量表示。然后,通过实体对齐机制将新抽取的知识元素与已有知识库进行比对融合,如果抽取的内容在知识库中已经存在,则将不会进行后续的操作,避免知识冗余和产生前后内容不一致的现象。

增量更新的核心是识别和处理知识冲突。当新文献中的信息与已有知识存在冲突时,系统基于信息来源的权威性、时间新近性和内容完整性进行权衡,确定最终的知识表示。这种机制确保了知识库在保持一致性的同时,能够及时反映最新的研究进展。

以下是其伪代码:

输入:新实体集合 E_{merged},现有实体集合 E_{old},阈值 θ = 0.85

输出:去重后的实体集合 E_{merged}

初始化 E_{merged} 为 E_{old} 的副本

```
for each e_new in E_new:
    e_old ← E_merged 中与 e_new 最相似的实体
    if similarity(e_new, e_old) > θ:
        if e_new.type is key type:
            if key_attributes_match(e_new, e_old):
                e_merged ← merge(e_new, e_old)
                E_merged ← (E_merged - {e_old}) ∪
```

$\{e_merged\}$

8：　　else：

9：　　$e_merged \leftarrow merge(e_new, e_old)$

10：　　$E_merged \leftarrow (E_merged - \{e_old\}) \cup \{e_merged\}$

11：　　else：

12：　　$E_merged \leftarrow E_merged \cup \{e_new\}$

13：return E_merged

为验证增量更新机制的实际效果，本研究以“超级杂交水稻”新技术文献（据文献[2－4，25－26]整理而成）整合为例展示完整执行过程。当新抽取实体“超级杂交水稻”与现有“杂交水稻”计算相似度为0.87（大于阈值0.85）时，系统通过关键属性匹配验证两者技术特征一致但产量潜力存在提升关系，决策合并为增强版实体。合并过程包括保留新实体名称、整合属性信息（产量潜力从“12 吨/公顷”更新为“12～15 吨/公顷”）、迁移原实体的47条关系，并重新计算向量表示。整个更新过程耗时约85 s，实现了知识的有机整合。

3　实验结果

3.1　实验平台

本研究在以下硬件和软件环境中进行：CPU 处理器 Intel（R） Core（TM） i9－14900K 3.20 GHz，GPU 为 NVIDIA GeForce RTX 4090D，内存容量为64 GB。软件环境包括 Python 3.12，CUDA 12.4，Ollama 框架和 Neo4j 图数据库。实验采用的大模型为通过硅基流动（SiliconFlow）云服务平台调用 API 形式的 Qwen 2.5－72B－Instruct 模型。硅基流动是一个专业的大模型云服务平台，提供多种开源大模型的 API 接口服务，支持高并发调用和稳定的推理性能。嵌入模型为在 Ollama 框架下调用的 bge－m3。

3.2　知识图谱构建结果

本研究通过大模型驱动的知识抽取方法，从水稻种植领域文献中成功构建了规模可观的知识图谱。经统计分析，图谱共包含 10 723 个节点（nodes）和 10 955 条边（edges），构成了完整的水稻种植领域知识网络。完整知识图谱构建过程共消耗约 461 万 tokens。按类型分布如图 3 所示。

知识图谱中的 10 723 个节点，以种植技术为核心，占 23.6%（2526 个），体现了实用栽培技术的主导地位；生长阶段（9.6%）和环境因素（7.9%）次之，反映了水稻生长发育和环境适应的重要性；作物品种、病虫害、产量指标等关键要素分布合理，形成完整的知识体系。

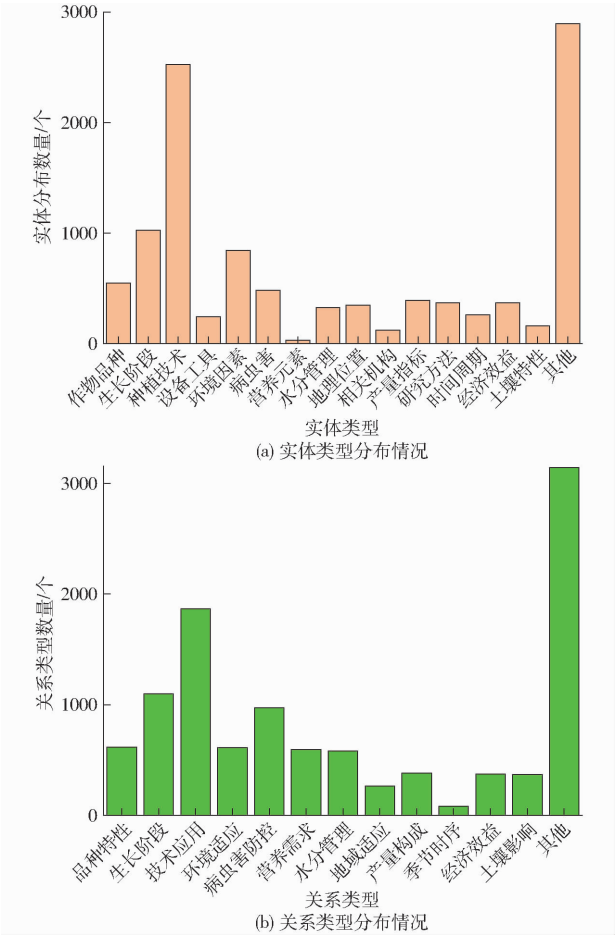

图 3　水稻种植领域知识图谱节点类型与关系类型分布

Fig. 3　Distribution of node and relationship types in rice cultivation domain knowledge graph

知识图谱共包含 10 955 条关系边，技术应用关系占主导地位（17.0%，1867 条），突出了栽培技术的核心作用；生长阶段关系（10.0%）和病虫害防控关系（8.9%）紧随其后；其它关系类别占 28.7%（3 147 条），这一现象反映了水稻种植领域知识关系的复杂多样性。通过对这些关系的人工抽样分析（随机抽取 200 条），发现主要包括：时序依赖关系（32%）：如“施肥后需间隔 3 天才能灌水”；条件制约关系（28%）：如“高温条件下不宜使用某种农药”；效果评价关系（24%）：如“该技术显著提高产量”；操作细节关系（16%）：如“播种深度应控制在2～3 厘米”。这些发现表明现有本体设计仍需进一步细化。未来研究将基于这些分析结果，扩展关系本体至 20～25 个类别，提升知识表示的精确性。

3.3　典型子图案例解析

使用 Neo4j 图数据库进行知识图谱的可视化[29]，为后续的检索和问答提供支持。

在构建的水稻种植领域知识图谱中，土壤特性子图（图 4）呈现了水稻种植中与土壤相关的复杂知识网络。该子图以土壤特性为核心节点，向外辐射

连接了土壤肥力、土壤酸碱性、土壤质地等关键属性节点,形成了一个包含 43 个节点和 67 条边的知识网络。这种放射状结构不仅展示了各土壤因素的独立特性,还通过节点间的交叉连接反映了土壤系统的复杂性和各因素间的相互影响关系。



图 4 土壤特性子图结构展示

Fig. 4 Subgraph structure display of soil characteristics

土壤特性子图涵盖了水稻种植中土壤相关的核心知识体系,包括土壤肥力系统、土壤改良技术集、环境适应性知识以及水分-土壤互动关系等维度。例如,在土壤肥力系统中,有机质含量通过促进关系连接到微生物活性,再通过影响关系关联到养分释放速度,形成完整的因果链;在土壤改良技术集中,各种改良措施如秸秆还田、深耕、稻菜轮作等都明确连接到对应的土壤问题,提供了针对性的解决方案;而环境适应性知识则展示了不同土壤条件下适宜种植的水稻品种选择指导,如“早籼 430”品种与“土壤疏松”条件间的适应关系 。

这一结构化的土壤特性知识子图在实际应用中具有重要价值,能够支持精准种植决策:通过土壤肥力→微生物活性→养分释放的关联链,快速定位土壤问题根源;提供问题诊断与解决方案:针对不同土壤问题(如酸化、板结),系统提供对应的改良措施(秸秆还田、深耕等);也能基于土壤特性自动推荐适宜品种:如“早籼 430”与“土壤疏松”的适应关系。该子图将原本碎片化的土壤管理知识连接成系统化的决策树,使农户能够基于具体土壤条件快速获得科学的种植建议,显著提升了专业知识在农业生产中的可用性和实用价值。

3.4 知识图谱质量评估

为确保构建的水稻种植知识图谱的准确性和实用性,设计了全面的质量评估体系,从多个维度对知识图谱进行评价。

评估指标体系包括 4 个主要维度:

覆盖度:评估知识图谱对水稻种植领域核心知识的覆盖情况,包括实体类型覆盖率、关系类型覆盖率和关键概念覆盖率。

准确性:评估实体和关系的准确程度,包括实体描述准确率、关系描述准确率和属性值准确率。

一致性:评估知识图谱内部的逻辑一致性,包括实体命名一致性、关系定义一致性和本体约束满足度。

规范性:评估知识表示的规范化程度,包括术语使用规范性、关系描述规范性和数据格式规范性。

首先构建了测试问答数据集:利用 Claude - 3.7 -Sonnet[30] 大语言模型基于 RC - PKBD 中的部分内容生成了 300 个问答对。这些问答对包括两类问题:具体型问题(150 个)和抽象型问题(150 个)。具体型问题如“水稻需要的微量元素有哪些?”,主要测试系统对特定知识点的检索能力;抽象型问题如“气温骤变时应该如何管理水稻田?”,则测试系统对多源知识综合理解和整合的能力。对于生成的回答,严格要求大模型按数据集中明文出现的内容作为答案。不得出现数据集中没有的回答。并且经过人工审核生成最后的问答集。

随后按照 LightRAG 框架实现了双层检索策略,双层检索机制的工作流程如图 5 所示。首先,系统接收用户查询后,同时启动两个并行的检索过程:低层次实体检索通过向量相似度匹配直接相关的实体节点,获取精确的事实性知识;高层次关系检索则通过图遍历算法探索实体间的复杂关联,发现隐含的语义联系。两个检索结果经过融合处理后,形成全面的知识检索结果,既保证了查询的准确性,又增强了答案的完整性和关联性。

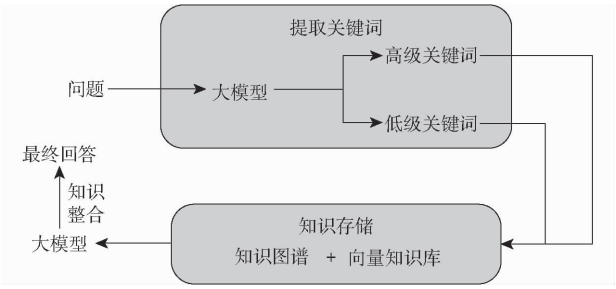

图 5 双层检索流程图

Fig. 5 Two-level retrieval flowchart

回答统一采用混合检索,即将低层次实体检索和高层次关系检索的答案融合,从而形成一个全面的,综合的答案。从 4 个维度评估系统回答质量,其中具体型问题主要用于评估准确性,抽象型问题主要用于评估覆盖度,而逻辑性和实用性则通过两类问题综合评估。

本研究采用“大模型自动评分 + 专家人工评估”的混合评估方法。评估所采取的大模型有两个，分别为 Claude - 3.7 - Sonnet[30] 和 Deepseek - R1[31]，这两个模型目前所表现的性能较为强大[32]。两个模型来源不同且技术路径各异，可以减少单一模型评估的偏差，提高评估结果的客观性。随后进行专家人工评估，邀请了 2 名相关领域的专家和 3 名一线农业生产者组成评估团队。每个维度按 1 ~ 5 分评分(1 分为非常差，5 分为非常好)，大模型评分和专家评分各占 50% 权重，最终转换为百分比。具体评分标准如下：

覆盖度评分标准：5 分(实体、关系和概念覆盖率 90% 以上)；4 分(覆盖率 80% ~90%)；3 分(覆盖率 70% ~80%)；2 分(覆盖率 60% ~70%)；1 分(覆盖率 0 ~60%)。

准确性评分标准：5 分(准确率 95% 以上)；4 分(准确率 90% ~95%)；3 分(准确率 85% ~90%)；2 分(准确率 80% ~85%)；1 分(准确率 0 ~80%)。

一致性评分标准：5 分(无矛盾，完全一致)；4 分(极少矛盾，基本一致)；3 分(有少量矛盾)；2 分(矛盾明显)；1 分(严重矛盾)。

规范性评分标准：5 分(完全符合规范)；4 分(基本符合规范)；3 分(部分不规范)；2 分(多处不规范)；1 分(严重不规范)。

根据表 4 的评估结果，水稻种植领域知识图谱展现出了良好的性能表现。具体型问题的准确性达 91.73%，表明系统能够精确检索和提供水稻种植领域的专业知识；抽象型问题的覆盖度达 90.20%，显示图谱全面涵盖了水稻种植各方面知识。在一致性方面，具体型和抽象型问题分别获得 88.93% 和 86.27% 的评分，反映知识图谱内部结构保持较高的逻辑一致性。规范性评分分别为 90.46% 和 87.73%，说明图谱符合行业术语和表述标准。综合得分方面，具体型问题和抽象型问题分别达到 90.44% 和 87.87%，证明系统能够有效支持不同类型的查询需求。

表 4　具体型与抽象型问题评估维度表现

Tab.4　Performance of specific and abstract questions across evaluation dimensions　%

评估维度	具体型问题	抽象型问题
准确性	91.73	-
覆盖度	-	90.20
一致性	88.93	86.27
规范性	90.46	87.73
综合得分	90.44	87.87

系统在处理简单直接的查询和复杂抽象的问题时均表现良好。抽象型问题虽然难度更高，但系统仍然保持了较高的性能水平。评估方法存在一定局限性，主要体现在以下几个方面：

模型自评分偏差：由于专家时间和成本约束，人工评估的样本规模相对有限，可能无法完全覆盖所有知识点和查询类型。

地域代表性：参与评估的专家和生产者主要来自华东地区，对于不同气候条件和种植制度下的适用性评估可能存在局限。

评估标准主观性：尽管制定了详细的评分标准，但专家评估仍具有一定主观性，不同专家的知识背景和经验可能影响评分结果。

动态验证不足：当前评估主要针对静态知识内容，缺乏对知识在实际生产过程中动态应用效果的长期追踪验证。

3.5　对比实验

为了更全面地评估本研究方法的有效性，进一步与图增强 RAG 系统 GraphRAG 进行了详细对比。

3.5.1　GraphRAG 实验配置

为确保对比实验的公平性和可重复性，严格控制了实验变量，具体配置如下：使用相同的 RC - PKBD 数据集，采用相同的文本分块策略(500 ~ 1 000 词段落单元)，确保两种方法处理的原始数据完全一致。

模型配置方面，采用的核心大语言模型统一为 Qwen 2.5 - 72B - Instruct，嵌入模型统一使用 BGE - M3 - Embedding。具体的参数配置与表 1 保持一致。

在查询处理流程方面，本方法采用改进的 LightRAG 双层检索策略；对应地，GraphRAG 也采用混合检索策略(局部搜索 + 全局搜索)，局部搜索针对具体实体，全局搜索基于社区报告。

在与前文相同的评价指标体系下，对比结果如表 5 所示。

表 5　与 GraphRAG 的对比结果

Tab.5　Comparison results with GraphRAG　%

评估维度	问题类型	本方法	GraphRAG
准确性	具体型问题	92.80	56.20
覆盖度	抽象型问题	90.40	81.80
一致性	具体型问题	89.60	74.80
	抽象型问题	86.93	72.40
规范性	具体型问题	91.47	78.20
	抽象型问题	88.27	75.20
综合得分	具体型问题	91.29	69.73
	抽象型问题	88.53	76.47

根据表 5 中的数据，可以清晰地看到本文所采

用的研究方法有着显著的准确性优势，相比于 GraphRAG 的 56.20% 提升了 36.60 个百分点。这主要归因于针对水稻种植领域设计的专业提示词策略，以及专门针对农业场景的条件约束关系表示，提高了知识抽取精度。

在抽象型问题的覆盖度评估中，本方法达到 90.40%，比 GraphRAG 的 81.80% 提升了 8.60 个百分点。这主要归功于多轮对话抽取策略，以及 15 类实体和 12 类关系的系统化本体模型提供了更全面的知识覆盖。

本方法在一致性和规范性两个维度上均显著优于 GraphRAG，这得益于：建立的专业词典和同义词识别规则确保术语一致性，以及多重验证保证了知识表示的规范性。

3.5.2 统计显著性分析

为验证本方法相比 GraphRAG 的性能提升是否具有统计显著性，对实验结果进行了统计检验分析。具体的方法为：采用配对样本 t 检验（Paired Sample t-test）分析两种方法在相同测试集上的性能差异。由于每个问题都有对应的评分对比，符合配对样本的条件。显著性水平设定为 $\alpha = 0.05$。

配对样本 t 检验分析显示，本方法在所有评估维度上均显著优于 GraphRAG。具体而言，在准确性维度（具体型问题），95% 置信区间为［32.21%，40.99%］，t 值为 37.69（$p < 0.001$）；在覆盖度维度（抽象型问题），95% 置信区间为［7.39%，9.81%］，t 值为 14.10（$p < 0.001$）。一致性和规范性维度的提升同样达到极显著水平，具体型问题和抽象型问题的平均提升均超过 13 个百分点，所有检验的 p 值均小于 0.001，表明性能差异具有极高的统计显著性。

为评估性能提升的实际意义，计算了 Cohen's d 效应量。结果显示，准确性提升的效应量最大（解等于 4.35），属于极大效应量；覆盖度提升的效应量为 1.63，一致性和规范性维度的效应量均超过 1.3，这些数值远大于 Cohen 标准中的大效应量阈值（解等于 0.8），表明本方法的性能提升不仅统计显著，而且具有实际应用价值。

3.5.3 典型问答案例分析

为更直观地展示本方法相比 GraphRAG 的性能优势，本节通过一个典型的水稻种植专业问答案例进行深入分析。

如图 6 所示，选取“什么育秧方式的生长特点是利于腐霉菌的生长，软腐缺氧，呼吸道阻，影响机体的正常代谢，降低生命活力，抵御不良环境的能力变差？”这一问题进行测试，其标准答案是：“水育秧（只有这一个答案，没有更多。）”。

GraphRAG 的回答存在问题：冗余信息过多，系

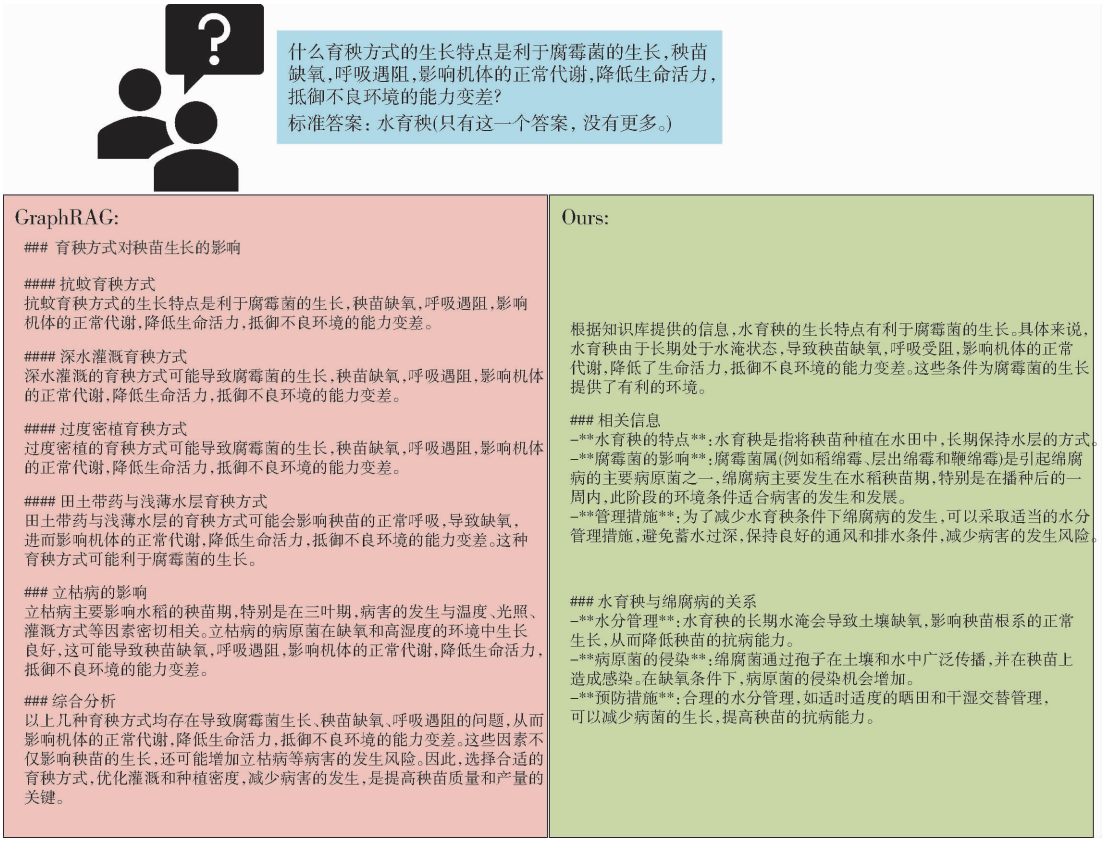

图 6 本方法与 GraphRAG 问答效果对比案例

Fig. 6 Comparison case of question-answering performance between proposed method and GraphRAG

统返回了包含营养钵育秧方式、机械化育秧技术等大量相关但不精确的内容。其次，准确性不足：虽然涉及多种育秧技术，但未能准确识别出问题所指向的特定育秧方式。但本文所提到的方法可以精确指出“水育秧”这一标准答案，并且系统解释了水育秧环境下腐霉菌生长、软腐缺氧等问题的形成机理，在提供精确答案的同时给出必要的专业解释，避免了信息冗余。其回答内容紧密围绕问题描述的关键特征，体现良好的语义理解能力。

3.6　大模型驱动与传统深度学习的知识图谱构建方法对比

为了更全面地评估本研究方法的优势，将大模型驱动的知识图谱构建方法与传统的基于 BERT－BiLSTM－CRF 的深度学习方法进行了系统对比。具体信息见表 6。这种对比有助于明确大模型在农业领域知识图谱构建中的独特价值和应用前景。

传统的 BERT－BiLSTM－CRF 方法虽然在特定任务上表现良好，但在处理复杂农业文本时存在明显局限。相比之下，大模型驱动的方法在语义理解深度、知识抽取完整性、系统适应性和实际应用便利性等方面具有显著优势。特别是在处理水稻种植这类专业性强、知识结构复杂的领域时，大模型的优势更加明显。

表 6　大模型驱动与传统深度学习知识图谱构建方法对比

Tab. 6　Comparison of LLM-driven and traditional deep learning methods for knowledge graph construction

对比维度	大模型驱动方法	传统深度学习方法
数据需求	无需大量标注数据，具有零样本和少样本学习能力，可直接处理原始农业文本	需要大量人工标注的训练数据，标注成本高，数据准备周期长
语义理解深度	具备强大的语义理解能力，能够理解复杂的农业专业术语和上下文关系，支持多轮推理	语义理解相对局限，主要依赖预训练表示，对复杂语义关系理解不够深入
知识抽取完整性	能够同时抽取实体、关系和属性，支持复杂关系识别，抽取结果更加全面	主要关注实体识别，关系抽取能力相对有限，难以处理复杂的多元关系
领域适应性	通过提示工程快速适应农业领域，无需重新训练，适应成本低	需要针对特定领域重新训练或微调，适应新领域成本较高
实体关系质量	能够识别隐含的语义关系，关系类型丰富，质量较高	主要识别显式关系，对隐含关系识别能力有限
更新维护便利性	支持增量更新，新知识融合便利，无需重新训练整个模型	更新需要重新训练，维护成本高，更新周期长
处理文本复杂度	能够处理长文本和复杂句式，对文本结构要求较低	对文本长度和复杂度敏感，处理长文本效果下降
应用场景适用性	适用于知识快速变化的农业领域，特别适合专业知识密集型场景	适用于固定任务和稳定领域，在传统 NLP 任务中表现稳定

4　结论

（1）本研究基于大模型驱动的知识图谱构建技术，成功构建了面向水稻种植领域的专业知识图谱系统，为农业知识的数字化转型提供了新的技术路径。

（2）研究的主要创新性为：首先，提出了适用于农业专业文本的大模型驱动知识抽取方法，通过专业提示工程和多轮对话策略，有效解决了传统方法在处理农业专业术语时语义理解不足的问题；其次，设计了覆盖水稻种植全生命周期的多维度知识表示框架，实现了从品种选择到收获储存各环节知识的系统化组织；最后，构建了基于图结构的双层次检索机制，显著提升了复杂农业问题的查询效果。

（3）实验验证表明，与现有先进方法相比，本研究方法在准确性、覆盖度、一致性和规范性等关键指标上均实现了显著提升，统计检验证实了这些改进的可靠性。构建的知识图谱不仅规模可观，更重要的是实现了农业专业知识从碎片化向系统化、从静态存储向智能检索的转变。

参　考　文　献

[1]　王悦. AI 大模型赋能教辅图书智能化升级：以教辅智能体为例[J]. 编辑学刊，2025(2)：116－120.

[2]　唐文丽. 水稻种植效益提升的有效措施[J]. 种子科技，2025，43(2)：207－209.

[3]　林彦宇，刘淇，武靖韬. 智能化水稻浸种装置与浸种液在现代农业中的发展现状及趋势分析[J]. 南方农机，2024，55(24)：7－10，19.

[4]　陈世军. 我国水稻栽培技术现状及提高种植效益的解决措施[J]. 种子科技，2025，43(2)：219－221.

[5]　Fensel D，Şimsek U，Angele K，et al. Introduction：what is a knowledge graph？[M]//Knowledge graphs：methodology，tools and selected use cases. Cham：Springer，2020：1－10.

[6]　贺建，张定林，武丽萍，等. 知识网络构建策略在有机化学教学中的应用及效果分析[J]. 大学化学，2025，40(8)：66－71.
He Jian，Zhang Dinglin，Wu Liping，et al. Application and effectiveness analysis of knowledge network construction strategies in organic chemistry teaching[J]. University Chemistry，2025，40(8)：66－71.（in Chinese）

[7] 李雯静, 万尧, 马倩, 等. 目标检测和知识图谱联合应用的矿山事故隐患识别推理[J]. 金属矿山, 2026(1): 188 - 197.
Li Wenjing, Wan Yao, Ma Qian, et al. Mining accident hazard identification and reasoning through the combined application of object detection and knowledge graph[J]. Metal Mine, 2026(1): 188 - 197. (in Chinese)

[8] 何美娜,胡慧,毛树松,等.护理领域知识图谱研究进展[J].护理研究,2025,39(8):1402 - 1408.
He Meina,Hu Hui,Mao Shusong, et al. Research progress of knowledge graph in nursing[J]. Chinese Nursing Research, 2025, 39(8): 1402 - 1408. (in Chinese)

[9] 侯琛,牛培宇.农业知识图谱技术研究现状与展望[J].农业机械学报,2024,55(6):1 - 17.
Hou Chen, Niu Peiyu. Review of research status and prospects of agricultural knowledge graphs[J]. Transactions of the Chinese Society for Agricultural Machinery, 2024, 55(6): 1 - 17. (in Chinese)

[10] 张嘉宇,郭玫,张永亮,等.细粒度苹果病虫害知识图谱构建研究[J].计算机工程与应用,2023,59(5):270 - 280.
Zhang Jiayu,Guo Mei,Zhang Yongliang,et al. Research on construction of fine-grained knowledge graph of apple diseases and pests[J]. Computer Engineering and Applications, 2023,59(5):270 - 280. (in Chinese)

[11] 董文虎,赵瑞雪.农作物病虫害知识图谱与问答系统构建研究[J].东北农业科学,2025,50(1):69 - 77.
Dong Wenhu,Zhao Ruixue. Research on the construction of knowledge graph and Q&A system for cropdiseases and pests[J]. Journal of Northeast Agricultural Sciences, 2025,50(1):69 - 77. (in Chinese)

[12] 陈明, 朱珏樟, 席晓桃. 基于知识图谱的花卉病虫害知识管理方法[J]. 农业机械学报, 2023, 54(3): 291 - 300.
Chen Ming, Zhu Juezhang, Xi Xiaotao. Knowledge management method of flower disease and pests based on knowledge graph [J]. Transactions of the Chinese Society for Agricultural Machinery, 2023, 54(3): 291 - 300. (in Chinese)

[13] 王耀君,徐国威,朱建军,等. 农业领域大语言模型研究进展[J]. 农业机械学报,2025,56(9):240 - 256.
Wang Yaojun, Xu Guowei, Zhu Jianjun, et al. Survey of research on large language models in agriculture[J]. Transactions of the Chinese Society for Agricultural Machinery, 2025,56(9):240 - 256. (in Chinese)

[14] 宋璐璐.基于知识图谱的水稻病虫害问答系统的设计与实现[D].雅安:四川农业大学,2023.
Song Lulu. Design and implementation of a question and answer system for rice diseases and pests based on knowledge graph [D]. Yaan: Sichuan Agricultural University,2023. (in Chinese)

[15] 高锐涛,林达伟,郭亮,等.基于知识图谱的水稻种植智能问答系统设计与实现[J].计算机工程,2024,50(12):133 - 141.
Gao Ruitao, Lin Dawei, Guo Liang, et al. Design and implementation of rice planting intelligent question-answering system based on knowledge graph[J]. Computer Engineering, 2024, 50(12): 133 - 141. (in Chinese)

[16] Guo Z, Xia L, Yu Y, et al. Lightrag: simple and fast retrieval-augmented generation[PP/OL]. V3. arXiv (2025 - 04 - 28) [2025 - 09 - 02]. https://doi.org/10.48550/arXiv.2410.05779.

[17] Edge D, Trinh H, Cheng N, et al. From local to global: a graph RAG approach to query-focused summarization[PP/OL]. V2. arXiv (2025 - 02 - 19)[2025 - 09 - 02]. https://doi.org/10.48550/arXiv.2404.16130.

[18] Sun J, Xu C, Tang L, et al. Think-on-Graph: deep and responsible reasoning of large language model on knowledge graph [PP/OL]. V6. arXiv (2024 - 03 - 24)[2025 - 09 - 02]. https://doi.org/10.48550/arXiv.2307.07697.

[19] Ma S, Xu C, Jiang X, et al. Think-on-Graph 2.0: deep and interpretable large language model reasoning with knowledge graph-guided retrieval[PP/OL]. V7. arXiv (2025 - 02 - 20)[2025 - 09 - 02]. https://doi.org/10.48550/arXiv.2407.10805.

[20] Liang X, Gu Z. Fast Think-on-Graph: wider, deeper and faster reasoning of large language model on knowledge graph[PP/OL]. V1. arXiv (2025 - 01 - 24)[2025 - 09 - 02]. https://doi.org/10.48550/arXiv.2501.14300.

[21] 韦一金,任有强,赵慧,等.基于 GraphRAG 的中国马铃薯新品种知识图谱构建[J].植物遗传资源学报,2025,26(6):1229 - 1241.
Wei Yijin, Ren Youqiang, Zhao Hui, et al. Construction of a knowledge graph for selection and breedingresearch of new potato varieties in China based on GraphRAG[J]. Journal of Plant Genetic Resources, 2025, 26(6):1229 - 1241. (in Chinese)

[22] 朱晓俊,黄慧欣,汪鹏,等.基于大模型的机器人研制场景知识图谱构建方法[J].控制工程,2025,32(8):1405 - 1416.
Zhu Xiaojun, Huang Huixin, Wang Peng, et al. A method for constructing knowledge graph of robot development scenariosbased on large language model[J]. Control Engineering of China, 2025, 32(8): 1405 - 1416. (in Chinese)

[23] 冯帅,许童羽,周云成,等.基于深度卷积神经网络的水稻知识文本分类方法[J].农业机械学报,2021,52(3):257 - 264.
Feng Shuai, Xu Tongyu, Zhou Yuncheng, et al. Rice knowledge text classification based on deep convolutionneural network [J]. Transactions of the Chinese Society for Agricultural Machinery, 2021, 52(3): 257 - 264. (in Chinese)

[24] 张坤丽,王影,付文慧,等.大语言模型驱动下知识图谱的构建及应用综述[J].郑州大学学报(理学版),2026,58(2):1 - 9.
Zhang Kunli, Wang Ying, Fu Wenhui, et al. A survey on the construction and application of knowledge graphs driven by large language models[J]. Journal of Zhengzhou University(Natural Science Edition), 2026, 58(2): 1 - 9. (in Chinese)

[25] 王守林. 水稻高产栽培技术[M]. 北京:中国盲文出版社,2000.
[26] 熊玉唐. 杂交水稻高产栽培技术[M]. 贵州:贵州科技出版社, 2007.
[27] Chen J, Xiao S, Zhang P, et al. M3-Embedding: multi-lingual, multi-functionality, multi-granularity text embeddings through self-knowledge distillation[PP/OL]. V4. arXiv (2024 - 06 - 28)[2025 - 09 - 02]. https://doi.org/10.48550/arXiv.2402.03216.
[28] 许浩翔,余敦辉,邓怡辰,等.基于分层强化学习的知识图谱约束问答模型[J].计算机应用,2025,45(12):3764 - 3770.
Xu Haoxiang, Yu Dunhui, Deng Yichen, et al. Knowledge graph constrained question answering model basedon hierarchical reinforcement learning[J]. Journal of Computer Applications, 2025, 45(12): 3764 - 3770. (in Chinese)
[29] 陈汇远.基于知识图谱的列控车载设备故障诊断的研究[J].机车电传动,2025(2):129 - 136.
Chen Huiyuan. Research on fault diagnosis of CTCS on-board equipment based on knowledge graph[J]. Electric Drive for Locomotives,2025(2): 129 - 136. (in Chinese)
[30] Anthropic. Claude 3.7 Sonnet[EB/OL]. [2025 - 02 - 19]. https://www.anthropic.com/claude.
[31] Guo D, Yang D, Zhang H, et al. Deepseek - R1: incentivizing reasoning capability in llms via reinforcement learning[PP/OL]. V1. arXiv (2025 - 01 - 22)[2025 - 09 - 02]. https://doi.org/10.48550/arXiv.2501.12948.
[32] Yang C, Wang R, Jiang J, et al. Nondeterministic polynomial-time problem challenge: an ever-scaling reasoning benchmark for LLMs[PP/OL]. V1. arXiv (2025 - 04 - 15)[2025 - 09 - 02]. https://doi.org/10.48550/arXiv.2504.11239.

(上接第 298 页)

[18] 赵朗月,吴一全.基于机器视觉的表面缺陷检测方法研究进展[J].仪器仪表学报,2022,43(1):198 - 219.
Zhao Langyue,Wu Yiquan. Research progress of surface defect detection methods based on machine vision[J]. Chinese Journal of Scientific Instrument,2022,43(1):198 - 219. (in Chinese)
[19] 李道亮,赵晔,杜壮壮.农业领域多模态融合技术方法与应用研究进展[J].农业机械学报,2025,56(1):1 - 15.
Li Daoliang,Zhao Ye,Du Zhuangzhuang. Research progress and application of multimodal fusion technology in agriculture[J]. Transactions of the Chinese Society for Agricultural Machinery,2025,56(1):1 - 15. (in Chinese)
[20] 秦韬,庄卫东.机器视觉技术在现代农业生产中的研究现状[J].现代化农业,2024(4):90 - 93.
Qin Tao,Zhuang Weidong. Research status of machine vision technology in modern agricultural production[J]. Modernizing Agriculture,2024(4):90 - 93. (in Chinese)
[21] 张智淇.基于表型检测与综合环境参数的香菇产量预测方法研究[D].沈阳:沈阳农业大学,2025.
Zhang Zhiqi. Research on lentinus edodes yield prediction methods based on phenotypic detection and comprehensive environmental parameters[D]. Shenyang: Shengyang Agricultural University, 2025. (in Chinese)
[22] Li Y, Che Y, Zhang H, et al. Wheat growth stage identification method based on multimodal data[J]. European Journal of Agronomy,2025,162:127423.
[23] Jia N, Zheng C. Multimodal rapid identification of growth stages and discrimination of growth status for Morchella[J]. Smart Agricultural Technology,2024,9:100507.
[24] 田爽,张云志,马庆芳,等.关于金耳的研究进展分析[J].中国林副特产,2024(4):101 - 102.
Tian Shuang,Zhang Yunzhi,Ma Qingfang,et al. Analysis of research progress on *Tremella aurantialba*[J]. Forest By-Product and Speciality in China,2024(4):101 - 102. (in Chinese)
[25] 刘源,王民乐,陈靖,等.冀北地区冷棚栽培金耳关键技术[J].中国食用菌,2021,40(12):89 - 92.
Liu Yuan,Wang Minle,Chen Jing,et al. Key cultivation technologies of *Tremella aurantialba* in cold greenhouses in northern Hebei Province[J]. Edible Fungi of China,2021,40(12):89 - 92. (in Chinese)
[26] 沈真辉,罗祥英,赵仙伟,等.利用转录组分析光照对金耳子实体转色的影响[J].食用菌学报,2023,30(6):12 - 27.
Shen Zhenhui, Luo Xiangying, Zhao Xianwei, et al. Transcriptome analysis of the effect of light on color transformation of *Tremella aurantialba* fruiting bodies[J]. Acta Edulis Fungi,2023,30(6):12 - 27. (in Chinese)