doi:10.6041/j.issn.1000-1298.2026.03.026

设施园艺辅助生产的知识增强大语言模型 PengKGPT 研究

孙先鹏[1,2]　项颖峰[1]　付　颖[1]　张晨阳[3]　吴伟骏[1]

(1.西北农林科技大学园艺学院,陕西杨凌 712100;
2.西北农林科技大学 & 西安交通大学智能农机装备研究院,咸阳 710075;
3.加利福尼亚大学戴维斯分校农业与环境技术学院,戴维斯 95616)

摘要:中国设施园艺产业的高速发展,使得对智能知识服务需求激增。然而,当前碎片化、关联性低的设施园艺知识体系和低精度、低效率的知识服务手段在指导生产中存在较大的缺陷。此外,从业者对于问题的描述不够全面,进一步增加了设施园艺问题的解决难度。本文结合知识图谱(Knowledge graph,KG)和大语言模型(Large language model,LLM)的优势,提出了多源设施园艺知识增强的问答模型,用于分析解决设施园艺生产中的问题。首先,构建了一个包含 60 余种设施园艺常见种植品类,近 150 万字的设施园艺知识数据集,通过语义分割获得 26 349 个文本块存储于向量数据库,并提取数据集中与生产技术相关的文本知识构建了 KG。同时,提出了一个基于 KG 实体匹配的语义信息增强模型,挖掘了 KG 实体之间的潜在关联,通过实体匹配的方式,增强用户输入的语义信息。其次,本文设计了一种具有 KG 和向量数据库双重引导提示的检索增强生成方法,将 KG 和相关文本信息共同输入提示模板增强 LLM 的问题分析能力。此外,为了增强其在设施园艺领域的适应性,在相关问答语料上使用低阶适应(Low-rank adaptation,LoRA)微调了 LLM。基于此,开发了一个多源知识增强的 LLM(命名为 PengKGPT),用于对设施园艺生产中的问题进行推理和响应。它使用与生产相关的自然文本描述作为输入,并将多源知识作为额外的语料库。最后,案例研究表明,PengKGPT 的得分率和准确率分别达到 91.2% 和 82.10%,较基座模型提高 36.6 个百分点和 32.53 个百分点,增强了 LLM 对垂直领域问题的分析能力;与 ERNIE 4.0 Turbo 和 GPT-4o 经典商业模型相比,得分率分别提高 10.2、14 个百分点,准确率分别提高 10.04、12.69 个百分点,说明 PengKGPT 在解决设施园艺生产中的问题方面表现出更高的专业性和可靠性。结果表明,该模型可为设施园艺生产提供辅助作用。

关键词:设施园艺;大语言模型;知识图谱;异构网络;检索增强生成

中图分类号:S625　　**文献标识码**:A　　**文章编号**:1000-1298(2026)03-0270-14　　**OSID**:

PengKGPT: A Multi-source Knowledge-enhanced Large Language Model for Assisting in Protected Horticulture Production

SUN Xianpeng[1,2]　XIANG Yingfeng[1]　FU Ying[1]　ZHANG Chenyang[3]　WU Weijun[1]

(1. *College of Horticulture, Northwest Agriculture and Forestry University, Yangling, Shaanxi* 712100, *China*
2. *Northwest Agriculture and Forestry University & Xi'an Jiaotong University Intelligent Agricultural Machinery Equipment Research Institute, Xianyang* 710075, *China*
3. *College of Agricultural and Environmental Technology, University of California, Davis, Davis* 95616, *U. S. A.*)

Abstract: The rapid development of China's protected horticulture industry has led to a surge in demand for intelligent knowledge services. However, the current fragmented and loosely connected knowledge systems, along with imprecise and inefficient knowledge service methods, pose significant challenges in guiding production. Moreover, practitioners' descriptions of issues are often incomplete, further complicating the resolution of protected horticulture problems. To address these issues in protected horticulture production, integrating knowledge graph (KG) and large language model (LLM) were proposed to create a multi-source knowledge-enhanced question-answering model. Initially, a knowledge dataset for protected horticulture was constructed, encompassing over 60 commonly cultivated categories

收稿日期:2025-08-25　修回日期:2025-11-30
基金项目:陕西省重点研发计划项目(S2024-YF-ZDCXL-ZDLNY-0159)和 2024 年省级财政农业专项资金项目
作者简介:孙先鹏(1979—),男,副教授,博士,主要从事智慧设施农业研究,E-mail:sunxianpeng@nwsuaf.edu.cn

in protected horticulture and containing nearly 1.5 million words. Through semantic segmentation, totally 26 349 textual blocks were obtained and stored in a vector database. Additionally, textual knowledge related to production techniques was extracted from the dataset to construct a knowledge graph. Concurrently, a semantic information enhancement model was proposed based on KG entity matching. Subsequently, a retrieval-augmented generation method was designed, in which the KG and related textual information were input into the prompt template to improve the LLM's problem-analysis capabilities. Furthermore, to enhance its adaptability in the field of protected horticulture, the LLM was fine-tuned on relevant question-answering corpora by using low-rank adaptation (LoRA) method. Based on this, a multi-source knowledge-enhanced LLM (named PengKGPT) was developed to reason and respond to issues in protected horticulture production. Finally, the case studies revealed that PengKGPT attained score and accuracy rates of 91.2% and 82.10%, respectively, marking improvements of 36.6 and 32.53 percentage points compared with the base model. This enhancement significantly augmented the large language model's analytical capabilities for questions in vertical domains. When benchmarked against classic commercial models such as ERNIE 4.0 Turbo and GPT-4o, PengKGPT demonstrated increases of 10.2 and 14 percentage points in score rate, along with improvements of 10.4 and 12.69 percentage points in accuracy rate, respectively. These results indicated that PengKGPT exhibited superior professionalism and reliability in addressing challenges within protected horticulture production. The results indicated that this approach can provide auxiliary support for protected horticulture production.

Key words: protected horticulture; large language models; knowledge graphs; heterogeneous networks; retrieval-augmented generation

0　引言

中国设施园艺产业产值超过 1.4 万亿元，占中国农业总产值的 1/4 以上[1]。我国大力发展设施园艺，不仅有效保障了“菜篮子”工程，同时充分利用了非耕地及非耕地地区的资源，为节能减碳战略目标作出重大贡献[2]。近年来，中国设施园艺飞速发展，正处于由主观经验判断向智能决策规范化方向转变的重要阶段，其个性化、精细化、智能化的发展趋势使得从业者对于知识服务的需求激增[3]。

设施园艺生产涉及一个漫长而复杂的过程。单个过程本身包含多个生产步骤和技术要点，且之间存在着多维度关系。要实现指导设施园艺生产，需要高质量的知识体系。然而，当前阶段碎片化、低秩化的设施园艺知识使得从业者获取知识的渠道和方式较为局限[4]，同时由于知识之间缺乏有效的关联性，只能满足生产过程中的某一或某些环节的技术需求，很难实现生产全过程指导。此外，由于从业者在寻求知识服务过程中，对需求的描述较为片面，进一步增加了设施园艺问题的解决难度，制约了设施园艺生产转型升级。

因此，迫切需要整理和强化设施园艺生产各环节的知识，构建高质量的设施园艺知识体系。近年来，人工智能技术正在不断地拓展农业问题解决的边界[5]。具体而言，以深度学习为代表的数据科学技术在设施园艺知识的规范化组织管理及计算挖掘方面发挥了重要作用[6]。然而深度学习的黑箱特性导致计算结果偏差较大、可解释性较差，缺乏专家的专业知识。基于此，有学者在农业领域中引入知识图谱(Knowledge graph，KG)，以农业文本的形式建模和挖掘实践中的生产数据[7]。随后，在解决农业生产问题中通过知识抽取方法构建了 KG[8]，并应用于问答系统中。然而，来自农业文本的知识主要通过给出粗略的关系来相互链接[3]。随着设施园艺生产流程日益精细化，许多知识中隐含的关系描述仍然含糊不清，并且没有得到深入讨论，使得构建出的 KG 质量较低。然而，异构图算法为强化多关系和多实体的 KG 提供了可行思路[9]，它强大的多关系处理能力和上下文感知能力可以进一步挖掘 KG 中的隐含关系。因此，如果将异构图算法增强后的 KG 应用于丰富用户输入的语义信息，可以提高整个生产周期中技术指导的鲁棒性。

另一方面，在复杂流程和多因素影响下，如何利用关联知识有效推断并实现生产技术高效指导也是一个难点。传统上，通过查阅技术和经验手册或者咨询专家来解决生产问题效率很低。此后一些研究人员利用知识库及问答系统取得了更有希望的结果，但查询受到许多限制[10]。如今，大语言模型[11](Large language model，LLM)如 ChatGPT，因其在理解指令和生成类人响应方面的卓越性能而在突破人工智能能力的界限方面取得了进展。LLM 的出现使得更自然地指导设施园艺生产成为可能[12]。然而，在实践中，LLM 受限于训练数据更新滞后，难以及时获取并学习设施园艺领域的最新专业知识。且

在一些报告中也证实了 LLM 可能会产生模糊或虚假的关联结果,使得其在及时性和可靠性方面效果欠佳[13]。为了解决此问题,越来越多的研究者和实践者开始关注 LLM + KG[14]。KG 不仅可以在 LLM 的预训练和推理阶段提供外部知识,还可以为 LLM 的分析任务提供可解释性。部分研究者通过仅使用几百个高质量的示例数据微调 LLM,取得了与 MiniGPT - 4 或 GPT - 4 相比具有竞争力的结果[15]。这表明,通过使用较少但高质量的指令微调数据可以使 LLM 生成更好的响应。目前,KG 与 LLM 的合作研究有了一定的进展[16],因此,应用 LLM 技术并引入领域内高质量的 KG,为设施园艺中的生产指导提供解决方案成为可能。

针对上述问题,本文提出一个用于辅助设施园艺生产的知识增强 LLM,称为 PengKGPT。将设施园艺相关的自然文本作为输入信息,首先增强其语义信息,之后检索获得相关文本信息,并输入到提示模板中。在其引导提示的支持下,通过推理回答设施园艺相关问题的可能解决方案。

1 文献综述

目前,农业生产中积累了来自科技文献、科学数据、历史统计和领域专家等的海量数据[3],其作为智能服务体系的底层逻辑支撑,使得组织这些多源异构数据显得极其重要。国家农业科学数据中心、中国农技推广网站、美国农业部、FAOSTAT 等构建了包括农作物品种、农作物病虫害、农业技术服务等类型的知识库。陈雷等[17]利用网络爬虫技术获得了原始数据包括 11 个农业大类、共计 8 481 个小类的条目数据。然而,网络爬虫组织手段和构建传统的知识库准确性较低且耗时又复杂,同时还会丢失知识之间的关联。基于此,有学者引入 KG 用来存储和挖掘农业知识。陈明等[18]通过人工收集相关数据并手动构建了花卉病虫害 KG。这种基于模板匹配法的手动构建方法,虽然提高了 KG 的准确性,但是并未深入挖掘数据潜在的关联。韦紫君等[19]通过 BERT、BiLSTM、CRF 等深度学习算法抽取农业知识构建了 KG。然而使用深度学习方法抽取知识构建 KG 会限制实体和关系类型的数量,不仅需要大量的人工干预,而且针对自然语言文本信息长距离上下文关联复杂、歧义、多义等问题效果较差。基于上述问题,郭旺等[10]提出使用 LLM 构建 KG,可以减少人工的干预并提高 KG 的质量。随后,杨波等[20]使用大模型实现了医疗领域命名实体识别。同时,蚂蚁集团与浙江大学联合构建并升级了蚂蚁百灵大模型对于知识抽取任务的能力,并发布大模型知识抽取框架 OneKE[21],进一步助力了大模型构建 KG。

同时,在 KG 应用方面,也有较多学者进行研究。HAIDAR 等[22]提出一种基于框架和规则的 KG 推理方法用来确定花生植株是否感染某些疾病。SATHYANARAYANAN 等[23]提出了基于关联规则挖掘的 KG 增强方法,CEN 等[24]基于异构图提出了一种用于图链接预测的图神经网络模型。通过增强 KG,进一步挖掘了 KG 信息,提高了下游任务的鲁棒性。此外,越来越多的学者开始研究如何利用 KG 中的知识对文本的语义表示进行扩展。ZHANG 等[25]先从文本中提取实体并获取其向量表示,随后通过设计知识编码层将实体知识有效融入模型框架,实现实体知识的结构化融合。ANNERVAZ 等[26]利用注意力机制来对齐文本与实体,并使用卷积神经网络编码外部知识,获取文本的目标三元组,将其作为外部知识拼接到文本表示。LIU 等[27]提出了 KG - BART 模型,通过融合 KG 信息和文本信息,生成高质量的句子,纯文本作为结构化知识的补充。可见,KG 可通过多种技术手段应用于文本语义扩展方面。

此外,以 OpenAI 提出的 GPT 系列模型[28]为代表的指令跟随大型语言模型的出现,例如 ChatGPT,给各行业的发展带来了变革性影响。特别是开源模型,如 LLaMA 和 ChatGLM[11],其训练成本相对较低,这使得 LLM 成为农业领域智能知识服务的潜在解决方案。然而,LLM 针对复杂问题推理能力有限,并且在特定领域中表现较差[13],生成内容存在虚构倾向。目前,仅依靠 LLM 进行行业应用不够可靠。KG 以其结构化的知识表达、可推理性和强解释性,正逐渐成为 LLM 技术研究的重要支撑手段。KG 在增强 LLM 进行问答任务时,发挥了不同的作用[16]。在构建复杂问答系统时,KG 通常作为外部背景知识载体发挥作用。KG^{2}RAG 从 KG 中检索相关子图,并利用检索到的 KG 扩展文本块以增强生成效果[29]。同时,KG 能够为 LLM 提供推理指引,使其基于事实证据获取精确知识。ToG - 2 以实体为中间表示,通过文档与 KG 之间的迭代检索来引导 LLM 获得准确答案[30]。此外,KG 可作为优化器与验证器,用于提升 LLM 的输出质量。可见,在 LLM 与 KG 协同技术研究中,研究者仅关注 KG 可以提供“精准、结构化、可信”的知识及其可推理性,忽略了 KG 在异构网络表征与关系建模方面的优势带来的知识增强潜力。本研究将这一潜力应用于用户语义信息增强中,获得良好的效果。

尽管 KG 在增强 LLM 进行问答任务时获得了

较好的表现，但是在垂直行业应用中，LLM 能够及时准确地获得丰富的行业知识仍然至关重要。结合特定领域的知识来微调 LLM 在垂直领域的应用是一种可行的低成本解决方案。其中，LoRA（Low-rank adaptation）微调方法的可扩展性和灵活性受到广大研究者使用[31]。同时，MAO 等[32]提出的检索增强生成（Retrieval- augmented generation，RAG）技术，通过检索垂直领域的知识库获得垂直行业知识以增强生成过程，显著提升了垂直领域问答的效果。此后，该技术逐渐被引入垂直领域的大模型应用研究中。MATSUMOTO 等[33]设计了一个 KG 增强检索增强生成框架用于解决生物医学问题。可见，检索增强生成提供了一种新的、有效的替代方法。然而，这一技术在设施园艺复杂问答任务中，存在检索质量不足和推理能力差的问题。因此，本研究将 KG 和检索增强生成的优势结合，为设施园艺领域的问答任务提供了解决方案。

农业 LLM 发布者为了迅速将大模型技术引入到农业领域中，主要利用现有技术如 RAG 技术、LoRA 微调技术等搭建了农业综合领域或细分领域的大模型，这极大地促进了农业领域的智慧化发展。近年来，中国农业大学、西北农林科技大学、北京市农林科学院、浙江大学、商汤科技开发有限公司、中国中化控股有限责任公司等机构陆续发布了农业领域的 LLM。国内农业 LLM 的研究主要侧重于农业综合知识理解、生产决策支持与多源数据融合分析，旨在构建具备农业整体领域或者分领域的知识推理与任务执行能力的智能农业认知系统；国外农业大模型的发展则更加聚焦于具体的业务应用场景，例如 John Deere 聚焦于农业设备的自动化作业决策分析（https：// www. deere. com/en/digital-tools），KrishiAI 聚焦于农业领域的知识问答[34]。可见，国内外均尚未有学者将大模型技术应用于设施园艺领域的智能知识服务中。

总体而言，上述方法使得 KG 和 LLM 服务于设施园艺领域成为可能。此外，它还通过将设施园艺专业知识紧密集成到 LLM 中，为辅助设施园艺生产提供了可行的解决方案。

2　多源知识增强 LLM 构建

本文提出的方法流程如图 1 所示，主要分为 2 个阶段。

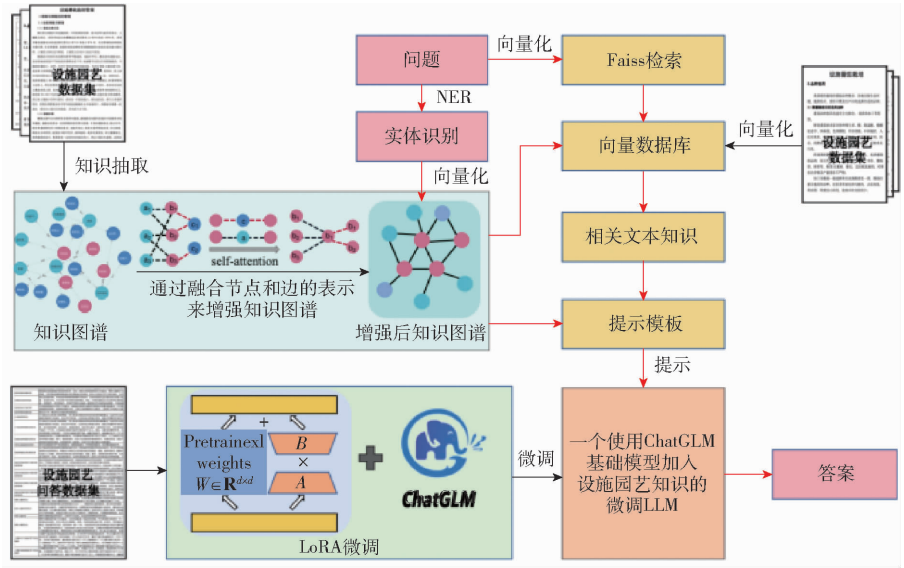

图 1　PengKGPT 的微调与增强流程图

Fig. 1　Flowchart of fine-tuning and enhancement of PengKGPT

（1）收集并整理了设施园艺领域内的文本数据，构建设施园艺数据集；进行专家鉴定，以保证文本数据的真实性和可靠性。文本数据通过文本分割器获得文本块，并转换为向量表示存储于向量数据库。从多源数据中分离出生产技术相关的文本数据并通过知识抽取的方式构建了 KG。同时，提出了一个基于 KG 实体匹配的语义信息增强模型，该模型深入挖掘 KG 实体间的潜在关联，通过实体匹配方法，实现对用户输入语义信息的有效增强。

（2）构建了具有真实用户问答语料的数据集，并基于此微调了 ChatGLM。设计了一种具有 KG 和向量数据库双重引导提示的检索增强生成方法，进一步提升了模型推理的完整性和可靠性。在此基础上，开发了 PengKGPT 模型。该模型通过结合用户查询内容，通过相关文本信息和实体引导提示驱动 PengKGPT，实现对设施园艺生产中相关问题的分析解决。

2.1 数据集构建

设施园艺领域尚未有学者或机构构建大规模文本数据集。因此本研究组建了设施园艺学专业团队,参考了科技文献、科学数据、历史统计、领域专家、互联网咨询和基于众包技术获取的海量文本、音频、图像、视频等设施园艺数据[3],构建设施园艺领域文本数据集。

考虑到有较多作物在设施园艺领域与非设施园艺领域中均可种植,但是其种植技术、方法和时间不同。然而通过爬虫技术或其他手段获取数据时,会忽略上述问题。因此,为了保证数据集的准确性和前沿性,采用人工收集后专家审阅的方式构建了近150 万字的设施园艺领域文本数据集(Protected Horticulture Field Text Dataset,简称 PHD)。包含果树、蔬菜、食用菌、花卉、瓜类等 60 余种常见设施园艺种植品类的特性介绍、品种分类、生产技术等。

同时,依托杨凌棚掌柜信息科技有限责任公司服务真实设施园艺从业者时产生的生产相关的问答数据,精心挑选了 1 500 个问答对组成了设施园艺问答数据集(Protected Horticulture Q&A Dataset,简称 PHQAD),用于支持 LLM 的微调。

2.2 知识图谱构建

KG 作为 LLM 检索增强内容的一部分,应该具有准确性高、关联性强的特点。但是,PHD 的内容复杂,关联性差。如果基于此构建 KG,不仅会增加构建的难度和成本,同时会降低 KG 的准确性,淡化其在辅助设施园艺生产中的作用。然而,辅助设施园艺生产的核心作用是解决生产技术相关的问题,为了增强这一核心作用,将 PHD 中生产技术相关的内容分离并提取,基于此,构建生产技术相关的 KG。

为了构建高质量的 KG 并挖掘知识要素间的逻辑关联,采用由蚂蚁集团和浙江大学联合研发的大模型知识抽取框架 OneKE[21],其具体的构建流程如图 2 所示。输入相应的指令信息到 OneKE 框架,通过轮询方式实现命名实体识别(NER)和关系提取(RE)。抽取后得到的实体集和关系集存储于 Neo4J 并实现可视化。OneKE 中的指令格式采用字典类型的字符串,它由以下 3 个字段构成:①instruction,即任务描述,以自然语言指定模型扮演的角色以及需要完成的任务。②schema,这是一份需提取的标签列表,明确指出了待抽取信息的关键字段。③input,指的是用于信息抽取的源文本。详细的 instruction 和 input 指令信息见图 2。schema指令信息主要包括实体类型列表或者关系类型列表。由于目前未有学者针对设施园艺领域构建本体模型,因此本文邀请专家对 PHQAD 中的问句进行分析,根据问句中的实体和意图,定义了 13 种与生产技术相关的实体类型:作物、地区、品种、设施类型、时期、环境管理、水肥管理、田间管理、植株形态、部位、灾害、药剂、发生原因;13 种与生产技术相关的关系类型:品种类别、防治、适用、抗病品种、易生灾害、原因、发育时期、发生时期、表现、侵害、表现部位、管理、管理部位。技术相关的本体及关系模型如图 3 所示。鉴于设施园艺领域内信息抽取的复杂性和对提示的依赖程度较高,在 schema 指令中融入对其的描述和示例来提升抽取任务的效果。

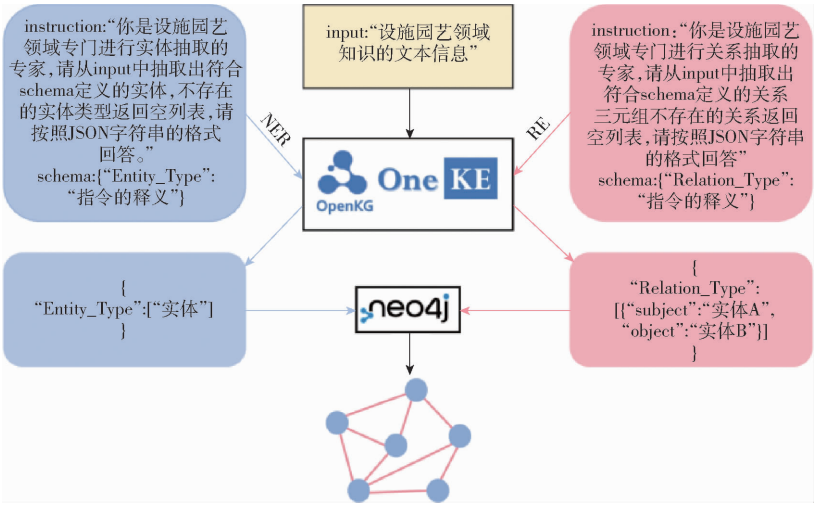

图 2 OneKE 框架构建设施园艺 KG 流程图

Fig. 2 Workflow for constructing a protected horticulture KG using OneKE framework

2.3 基于 KG 实体匹配的语义信息增强模型

在构建 PHQAD 过程中,发现设施园艺从业者对相关问题的文本描述较为简洁。为了增强用户输入的语义信息,提出了一个基于 KG 实体匹配的语义

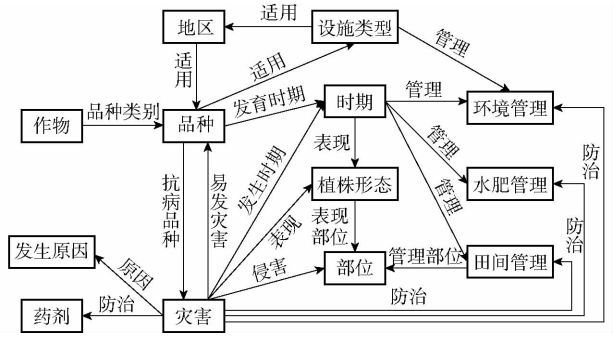

图 3　KG 本体及关系模型图

Fig. 3　Ontology and relationship model of KG

信息增强模型，如图 4 所示（图中 V 为一组 n 个节点，$V=V_1\cup V_2\cup V_3$，K_1、K_2、K_3 为指定节点 v 分别在每个节点类型上的邻域大小）。该模型的核心部分为 GATNE－T 算法[24]，使用 GATNE－T 算法对 KG 进行建模，获得节点在各种关系下的初始化向量化表示。执行基于不同边类型下的元路径的随机游走得到节点序列，将其输入 Skip－Gram 模型，更新模型参数，得到最终的节点向量，并将其作为实体特征。计算 KG 实体与用户输入信息中实体的匹配度，将匹配得到的 KG 实体信息添加到用户输入信

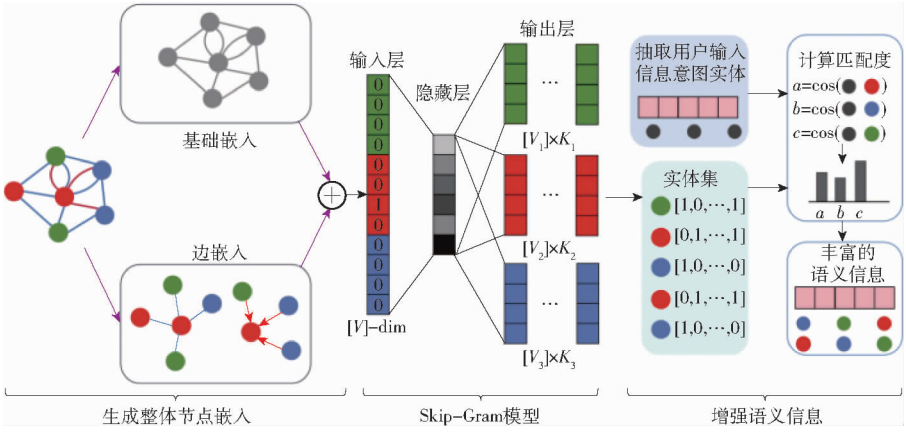

图 4　基于 KG 实体匹配的语义信息增强模型

Fig. 4　Semantic information enhancement model based on KG entity matching

息中，实现用户输入语义信息的增强。

具体来说，在 GATNE－T 中，将某一节点 v_i 在每一种边类型 r 上的整体嵌入分为 2 部分：基本嵌入和边嵌入。节点 v_i 的基本嵌入在不同边类型之间是共享的。节点 v_i 对边类型 r 的邻居进行聚合，随机初始化节点 v_i 在边类型 r 下的初始化边向量 $\boldsymbol{u}_{j,r}^{(0)}$，得到节点 v_i 在边类型 r 下的第 k 层边向量 $\boldsymbol{u}_{i,r}^{(k)}\in\mathbf{R}^s$，$N_{i,r}$ 是节点 v_i 边类型为 r 的邻居。计算式为

$$\boldsymbol{u}_{i,r}^{(k)}=\sigma\left(\hat{\boldsymbol{W}}^{(k)}\operatorname{mean}\left(\left\{\boldsymbol{u}_{j,r}^{(k-1)},\forall v_j\in N_{i,r}\right\}\right)\right)\quad(1)$$

式中　σ——非线性激活函数，选择 ReLU 函数

$\hat{\boldsymbol{W}}^{(k)}$——可学习参数

$\operatorname{mean}(\ )$——均值聚合器

将第 K 层边向量 $\boldsymbol{u}_{i,r}^{(k)}$ 表示为 $\boldsymbol{U}_{i,r}$，拼接节点 v_i 所有的边向量为 $\boldsymbol{U}_i$，计算式为

$$\boldsymbol{U}_i=(\boldsymbol{U}_{i,1},\boldsymbol{U}_{i,2},\cdots,\boldsymbol{U}_{i,m})\quad(2)$$

使用 self-attention 机制来计算边类型 r 在向量线性组合 $\boldsymbol{U}_i$ 中的权重系数 $a_{i,r}\in\mathbf{R}^m$，其中，$\boldsymbol{w}_r\in\mathbf{R}^{d_a}$ 和 $\boldsymbol{W}_r\in\mathbf{R}^{d_a\times s}$ 为在边类型 r 下要学习的参数。计算式为

$$a_{i,r}=\operatorname{softmax}\left(\boldsymbol{w}_r^{\mathrm{T}}\tanh\left(\boldsymbol{W}_r\boldsymbol{U}_i\right)\right)^{\mathrm{T}}\quad(3)$$

节点 v_i 在边类型 r 下的最终向量表示为

$$\boldsymbol{v}_{i,r}=\boldsymbol{b}_i+\alpha_r\boldsymbol{M}_r^{\mathrm{T}}\boldsymbol{U}_i a_{i,r}\quad(4)$$

式中　$\boldsymbol{b}_i$——节点 v_i 的基向量

α_r——边向量在整体向量中的超参数

$\boldsymbol{M}_r$——可训练转换矩阵，$\boldsymbol{M}_r\in\mathbf{R}^{d_a\times s}$

使用基于元路径的随机游走生成节点序列，随后应用 Skip－Gram 模型学习节点嵌入及参数。给定关系 r 和网络 $G_r=(V,E_r,A)$，及长度为 l 的元路径 $\Gamma:\rightarrow V_1\rightarrow V_2\cdots\rightarrow V_t\rightarrow\cdots\rightarrow V_l$，在第 t 步转移概率 p 公式为

$$p(v_j|v_i,\Gamma)=\begin{cases}\dfrac{1}{|N_{i,r}\cap V_{t+1}|} & ((v_i,v_j)\in E_r,v_j\in V_{t+1})\\ 0 & ((v_i,v_j)\in E_r,v_j\notin V_{t+1})\\ 0 & ((v_i,v_j)\notin E_r)\end{cases}\quad(5)$$

式中　V_{t+1}——第 $t+1$ 步时，邻域范围内包含的所有实体节点

只有邻接节点符合路径定义的类型，下一步才有可能被采样到。采用随机游走方法得到遵循路径 $P=\{v_{p1},v_{p2},\cdots,v_{pl}\}$ 的长度为 l 的节点系列，将节点 v_{pt} 的上下文节点定义为 $C=\{v_{pk}|v_{pk}\in P,|k-t|\leqslant c,t\neq k\}$，其中，$c$ 是窗口的半径，窗口的半径越大，节点 v_{pt} 和上下文节点组成的训练对就越多，准确率也就越高。在本研究中，防止特定元路径可能导致信息丢失和模型性能的下降，不设置元路径，随机游走不受限制，从而能够更全面地捕捉网络中的多层次关

系和潜在模式。

得到节点 v_i 及其上下文节点 C 后，使用 Skip-Gram模型学习节点表示。目标为最小化负对数似然函数，计算式为

$$-\ln P_\theta(\{v_j \mid v_j \in C\} \mid v_i) = \sum_{v_j \in C} -\ln P_\theta(v_j \mid v_i) \tag{6}$$

式中 θ——所有参数

$P_\theta(v_j \mid v_i)$——在中心节点为 v_i 的情况下，上下文节点中真实存在 v_j 的概率

随后使用异构函数对该概率进行归一化

$$P_\theta(v_j \mid v_i) = \frac{\ln(\boldsymbol{c}_j^{\mathrm{T}} \boldsymbol{v}_{i,r})}{\sum_{k \in V_t} \ln(\boldsymbol{c}_k^{\mathrm{T}} \boldsymbol{v}_{i,r})} \tag{7}$$

式中 $\boldsymbol{c}_j$——节点 v_j 的上下文嵌入

$\boldsymbol{c}_k$——节点 v_k 的上下文嵌入

最后，使用异构负采样将每个节点对 (v_i, v_j) 的目标函数 $-\ln P_\theta(v_j \mid v_i)$ 近似表示为

$$E = -\ln\sigma(\boldsymbol{c}_j^{\mathrm{T}} v_i) - \sum_{l=1}^{L} E_{v_k \sim P_t(v)}[\ln\sigma(-\boldsymbol{c}_k^{\mathrm{T}} v_i)] \tag{8}$$

式中 σ——sigmoid 函数　　L——负样本数

$P_t(v)$——节点集合的噪声分布

v_k——负样本

模型训练参数设置如表 1 所示。模型参数更新后，得到最终 n 维嵌入节点。此时，KG 中的节点均有语义丰富的向量表示。选用余弦相似度来衡量 KG 实体 x 与用户输入信息中的实体 y 的匹配度。余弦相似度的值越接近 1，表明 2 个实体匹配程度越大。

表 1　模型参数设置表

Tab. 1　Model parameters settings

参数	数值/类型
节点嵌入维度 d	200
边嵌入维度 s	10
随机游走的路径长度	10
样本窗口大小	5
负样本数 L	5
系数 α_r	1
优化器	Adam
学习率	0.001

针对用户输入信息中的各实体，从 KG 实体集中，选择与其匹配程度最高的若干实体添加到用户输入信息中，实现对用户输入语义信息的增强。

2.4　设施园艺知识增强 LLM

为了设计一个离线 LLM，使其在设施园艺生产中提供更具有竞争力的回答，本文使用了专为中文场景设计的对话模型 ChatGLM 作为基础模型。另外，使用 PHD、KG 作为外部语料数据。同时，通过微调方法使得 ChatGLM 学习 PHQAD 中的语料，使其适应设施园艺领域。并设计了一种 PHD 和 KG 双重引导提示的检索增强生成方法。基于上述研究，获得了一个名为 PengKGPT 模型的离线 LLM，如图 5 所示，提出了多源知识增强 LLM 框架。

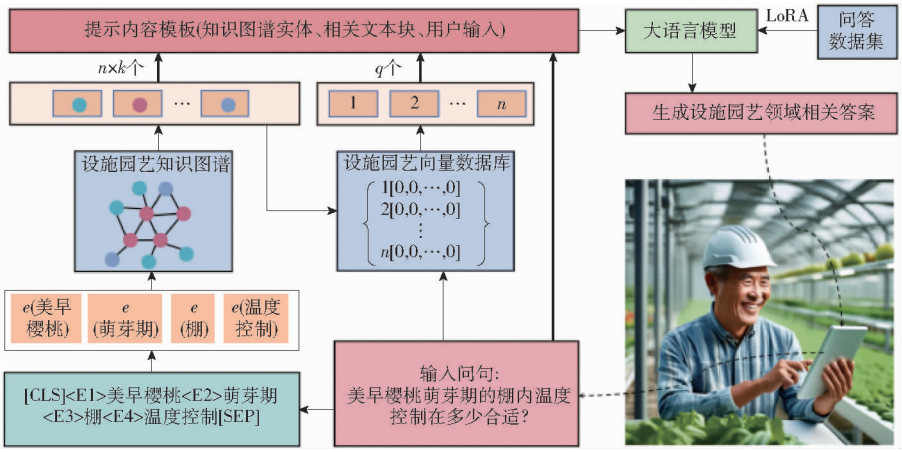

图 5　多源知识增强 LLM 框架

Fig. 5　Framework of multi-source knowledge-enhanced LLM

首先，根据用户输入信息，抽取得到意图实体。从 KG 中，选择与用户输入信息中的实体匹配程度最高的若干实体添加到用户输入信息中，得到增强后的用户输入信息。并将其与向量数据库匹配，得到若干关联知识。最后，将增强后的用户输入信息和向量数据库匹配得到的知识输入到提示模板中，以增强 LLM 的推理能力。使用 LoRA 方法微调后的 LLM，由多源知识引导的提示驱动，从而能够分析设施园艺生产中的相关问题。

2.4.1　数据向量表示

PHD 通过数据加载器处理，并通过文本分割器按照段落模式获得设施园艺知识文本块。此外，通

过文本嵌入模型 bge－large－zh－v1.5[35] 实现文本向量化，并存储于向量数据库 D_{VD}

$$D_{VD}=\{\boldsymbol{e}_d(1),\boldsymbol{e}_d(2),\cdots,\boldsymbol{e}_d(w)\} \tag{9}$$

式中　$\boldsymbol{e}_d(w)$——文本块的向量化表示

KG 通过增强，每一节点获得嵌入表示。给定一个设施园艺相关问题，采用信息提取模型[36] 来识别句子中的实体 $E(w)$ 和关系 $R(w)$。并通过文本嵌入模型 bge－large－zh－v1.5 获得实体向量集合 E_V 和关系向量集合 R_V

$$Q=\{E(1),E(2),\cdots,E(w),R(1),R(2),\cdots,R(w)\} \tag{10}$$

$$E_V=\{\boldsymbol{e}_e(1),\boldsymbol{e}_e(2),\cdots,\boldsymbol{e}_e(w)\} \tag{11}$$

$$R_V=\{\boldsymbol{e}_r(1),\boldsymbol{e}_r(2),\cdots,\boldsymbol{e}_r(w)\} \tag{12}$$

式中　Q——输入问句中的实体和关系集合

$\boldsymbol{e}_e(w)$——对应实体的向量表示

$\boldsymbol{e}_r(w)$——对应关系的向量表示

2.4.2　双重引导提示的检索增强生成方法

双重引导提示的检索增强生成方法中，提示模板由用户输入描述、PHD 中相关文本块和 KG 中相关实体构成。其中，用户输入信息为用户和 LLM 搭建了桥梁。由于 LLM 预训练时缺乏设施园艺领域语料，因此，本文利用 RAG 技术，将 PHD 中相关文本块以提示的方式供 LLM 使用，这不仅为 LLM 的推理提供了更加准确的信息，也提供了更专业的领域语料，减少了语言模型中的幻觉问题。尽管 PHD 为 LLM 提供了更加准确的信息，但是 LLM 训练时可能会存在固有偏见，导致文本块检索偏颇，缺乏多样性。同时，由于文本块信息之间较多的潜在关联未被挖掘，使得推理缺乏完整性。因此，本文加入了可解释性更强的 KG，其明确的关系链为 LLM 的推理提供了更丰富的信息，提高了推理任务的鲁棒性。具体构建流程如下：

用户输入设施园艺相关信息后，抽取并获得实体和关系的向量表示。首先，从 KG 中，选择与用户输入信息中的实体匹配程度最高的 p 个实体添加到用户输入信息中，得到增强后的用户输入信息，并输入到提示模板中。将增强后的用户输入信息与向量数据库匹配，得到匹配程度最高的 q 个关联知识，同样输入到提示模板中。

本研究中设定 q 为 3，p 为 2。考虑到提示模板的严谨性，q 及 p 设定较大，会导致相似度较低的信息混入，导致 LLM 产生错误的推理；q 及 p 设定较小，无法为 LLM 提供足够的提示，导致推理结果不理想，甚至无法区别于其他 LLM。

2.4.3　利用 LoRA 微调大模型

基础模型 ChatGLM 在训练时，农业领域尤其是设施园艺领域语料数据较少，可能会导致其在该领域存在偏颇，使得其在设施园艺领域的推理能力较差。尽管在提示模板中，通过 RAG 技术为 LLM 提供了更专业的领域语料。为了让 LLM 能够适应设施园艺领域任务，通过引入算力要求低、灵活性高的微调技术 LoRA，让其学习了 PHQAD 中的知识，适应设施园艺领域的推理任务。

微调过程中，将 2 000 个设施园艺生产相关的问答对，按比例 9∶1 随机划分数据集为训练集和测试集。并通过向量化模型 bge－base－zh－v1.5 将测试集中的答案和 LLM 微调后生成的答案向量表示，然后计算两者的相似度。将测试集中的 200 条数据相似度结果求均值，得到最终的准确率。微调训练参数设置如表 2 所示。

表 2　训练参数设置

Tab.2　Training parameters settings

参数	数值/类型
模型输入最大长度	512
模型输出最大长度	1 024
训练最大步数	3 000
每个设备的训练批量大小	4
LoRA 秩	8
LoRA 放大系数	32
LoRA 层 dropout 率	0.1
训练轮数	3
学习率	5×10^{-5}
梯度累加的步数	2
优化器	AdamW

3　模型结果与分析

3.1　构建 KG

KG 中的实体及关系抽取使用 2.2 节中的方法，在一台 NVIDIA－A100(40G)服务器上实现。在 KG 中，本文定义了 13 类实体和 13 类关系，其具体数量如表 3 所示。由于存在着不同的头实体和尾实体由同一关系类型连接，因此构成的三元组有 22 类。

本文通过三维图表直观地说明这些三元组的组成关系及数量，如图 6 所示。

具体来说，在三维图表中，x 轴的标签表示三元组中的头实体类型，y 轴中的标签表示三元组中的尾实体类型，x 轴和 y 轴在三维图表底部的交点上的柱体表示关系类型，z 轴表示关系类型的数量，每个柱体的顶部数字表示该三元组类型的数量。从图表呈现的信息可以清晰地看出，KG 中实体之间的关系具有高度的复杂性。

本研究运用简单随机抽样与两阶段加权聚类抽样方法，对 KG 的准确性展开系统性评估，具体评估

表 3 技术 KG 的实体及关系数量

Tab. 3 Quantities of entities and relationships in the KG

实体	实体数量	关系	关系数量
作物	121	品种类别	496
品种	573	适用	1 025
时期	202	发育时期	596
地区	122	管理	6 002
设施类型	40	表现	5 478
环境管理	3 293	表现部位	336
水肥管理	2 878	管理部位	2 021
田间管理	4 141	抗病品种	1 171
植株形态	3 122	发生时期	413
部位	163	易生灾害	906
灾害	392	侵害	765
药剂	532	原因	2 927
发生原因	2 406	防治	4 113
总计	17 985	总计	26 249

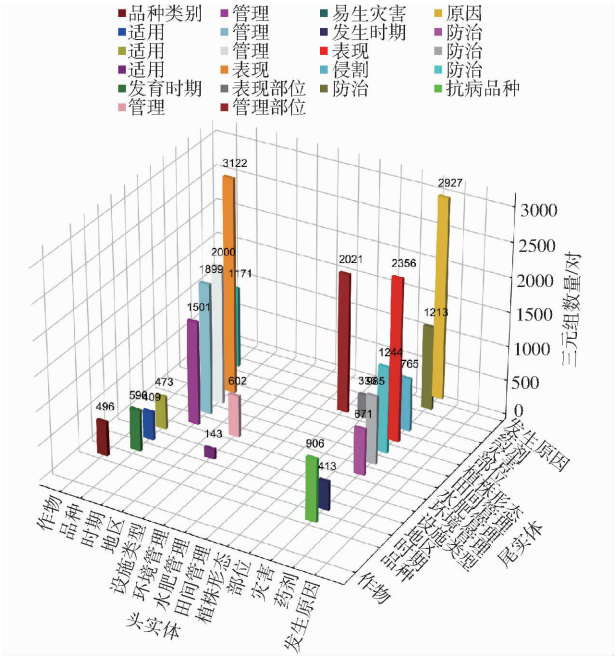

图 6 实体和关系展示及数据统计

Fig. 6 Entity and relationship display with statistical data

结果详见表 4。通过整合 2 种抽样策略的置信区间分析,可判定 KG 的质量指标落在 0.80 ~ 0.84 区间,表明在既定置信水平和误差范围内,该 KG 具备相对较高的准确性。KG 准确性较低的主要原因在于设施园艺领域知识数据集较复杂,且文本序列较长,捕获复杂语义存在一定困难。

表 4 KG 评估结果

Tab. 4 Knowledge graph evaluation results

抽样方法	准确率落在置信区间的概率	实体抽取准确率
简单随机抽样(SRS)	0.84	0.82
两阶段加权聚类抽样(TWCS)	0.81	0.80

3.2 基于 KG 实体匹配的语义信息增强分析

以用户输入“美早樱桃萌芽期的棚内温度控制在多少合适?”为例。表 5 为该问句经过命名实体识别得到的实体名称和利用余弦相似度与 KG 匹配得到的实体名称及其对应的相似度。由表 5 可知,该问句共识别得到实体 4 个,每个实体与增强后的 KG 匹配到 2 个相似度最高的实体。“美早樱桃”匹配到“美早”和“甜樱桃”,对应到 KG 中,“美早”属于“品种”,“甜樱桃”属于“作物”。“萌芽期”匹配到“萌芽期”和“萌芽前后”,对应到 KG 中,“萌芽期”和“萌芽前后”均属于“时期”。“棚”匹配到“棚室”和“大棚”,对应到 KG 中,“棚室”和“大棚”均属于“设施类型”。“温度控制”匹配到“温度控制”和“温湿度控制”,对应到 KG 中,“温度控制”和“温湿度控制”均属于“环境管理”。尽管用户输入与增强后匹配得到的“萌芽期”“温度控制”在文本形式上保持一致,其相似度仍仅为 0.902 9 和 0.902 1。这表明语义增强模型在引入 KG 结构信息后,能够区分同名概念在不同语义关联下的细微差异,从而使节点表示更具区分度与结构敏感性。

表 5 实体名称及相似度

Tab. 5 Entity names and corresponding similarity values

实体名称(用户输入)	实体名称及相似度
美早樱桃	美早(0.663 2)、甜樱桃(0.634 8)
萌芽期	萌芽期(0. 902 9)、萌芽前后(0.783 9)
棚	棚室(0.797 1)、大棚(0.736 8)
温度控制	温度控制(0.902 1)、温湿度控制(0.791 9)

图 7 是提示模板对用户输入的内部响应过程。提示内容中的“萌芽前后”和“温湿度控制”是生成的回答含有较丰富信息的主要贡献部分(生成的回

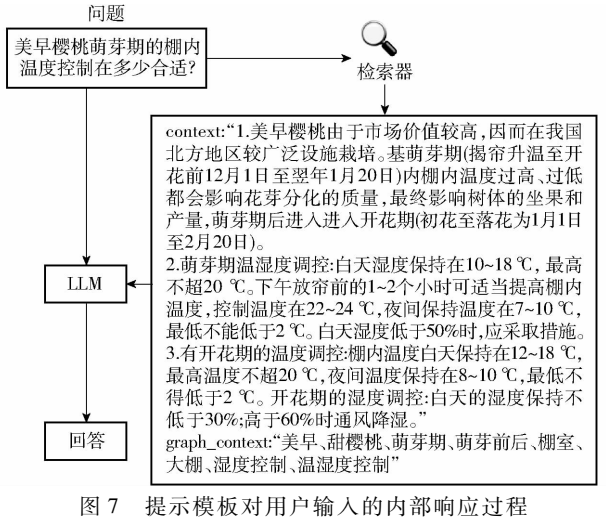

图 7 提示模板对用户输入的内部响应过程(包含向量数据库和 KG)

Fig. 7 Internal response process of prompt template to user input, including vector database and KG

答见图 8），这与图 9 展示的仅包含向量数据库信息的提示模板对用户输入的内部响应过程及答案形成明显对比。可见，本研究提到的基于 KG 实体匹配的语义信息增强模型在解决设施园艺生产问题时发挥了重要的作用。这说明，在 KG 增强 LLM 的研究中，不应该只关注 KG 能够提供精准、结构化、可信的知识及其可推理性，同时要探索 KG 网络的结构优势。

模型	问题:美早樱桃萌芽期的棚内温度控制在多少合适?	回答分析点评
GPT-4o	美早樱桃的萌芽期对温度有一定的要求，以促进其正常生长和发育、在棚内温度控制方面，一般建议如下： 白天温度：保持在15~20℃之间。这个温度范围可以促进芽的萌发和新梢的生长。 夜间温度：保持在5~10℃之间。夜间较低的温度有助于抑制过快的生长，避免徒长。 需要注意的是，不同地区的气候条件可能会有所不同，因此在实际操作中还就根据具体情况进行适当调整。同时，温度的控制应结合湿度、光照等因素，以保证樱桃树的整体健康生长。	**优势**:语言表达流畅、描述通用。回答提供了基本的温度范围(白天15~20℃，夜间5~10℃)，并强调了根据地区调整和结合其他因素(如湿度，光照)，体现了模型的通用知识库和适应性。适合作为对非专业读者的概览性建议。 **劣势**:缺乏精准细节。温度范围较宽泛，未针对美早樱桃的品种特性或设施栽培场景细化；缺少具体时间、湿度控制等补充信息，实用性较低。
ERNIE4.0 Turbo	美早樱桃萌芽期的的棚内温度控制对樱桃的生长至关重要。具体来说，为了促进美早樱桃顺利萌芽，棚内温度应控制在以下范围内： 白天温度：建议保持在18~20℃，这个温度范围有利于樱桃树体的生理活动和萌芽进程。 夜间温度：不应低于5~7℃，以确保樱桃树在夜间也能维持正常的生理机能，避免低温对萌芽造成不利影响。 此外，萌芽期还需注意棚内湿度控制，建议保持相对湿度在80%左右，以促进萌芽整齐和提高坐果率。同时，根据天气情况和樱桃树的生长状态，适时进行通风换气，以保持棚内空气的新鲜和适宜的生长环境。 综上所述，合理控制棚内温度和湿度是美早樱桃萌芽期管理的关键措施之一。	**优势**:补充信息丰富。回答不仅提供了温度范围(白天18~20℃，夜间5~7℃)，还详细说明了湿度控制和通风换气的重要性增强了实用性。 **劣势**:精准性不足，温度范围仍较笼统，未涉及具体时间或设施类型，且夜间温度范围错误(专家及专业书籍建议夜间温度不得低于7℃)；湿度建议过于统一，未考虑不同阶段的差异，深度有限。
ChatGLM-6B	萌芽期温度管理，升温不可过急，一般要控制在白天18~20℃，夜间8~10℃。	**优势**:简洁明了，回答直接给出了温度范围(白天18~20℃，夜间8~10℃)，适合快速参考。 **劣势**:过于简略。缺乏背景信息、湿度控制或其他管理细节，深度不足；未针对美早樱桃品种定制，可能不适用于实际设施栽培场景。
PengKGPT	美早是樱桃的一个品种，多以设施栽培为主，其设施栽培的温室类型有两种：塑料大棚和日光温室。美早樱桃温室栽培的萌芽期为揭帘升温至开花前，大约时间12月1日到翌年1月20日。这段时间内的温度控制建议如下： 白天适宜温度在10~18℃，最高不超20℃，下午放帘前的1~2个小时可控制在22~24℃。夜间适宜温度在7~10℃，最低不能低于2℃。同时，白天湿度不低于50%，低于50%时，向地面喷水增加空气湿度，夜间则顺其自然。 而在此之后的开花期温度控制为：白天适宜温度在12~18℃，最高温度不超20℃，夜间适宜温度在8~10℃，最低不应低于2℃。白天的湿度保持不低于30%，低于30%时向地面洒水，增加空气湿度；高于60%时通风降湿，尽量避免棚膜滴水。开花期注意通风，保持适宜的湿度，防止花腐病等病害。	**优势**:精准性和深度突出。 精准性：提供了具体的温度范围(白天10~18℃，夜间7~10℃)和时间段(12月1日至翌年1月20日)，并细化了下午放帘前的临时温度控制(22~24℃)，数据更贴近实际栽培需求。 深度：不仅涵盖萌芽期，还扩展至开花期的温湿度控制(白天12~18℃，夜间8~10℃)，并详细说明湿度管理(如喷水、通风)和病害预防，体现了对樱桃生长周期的全面理解。 背景信息：开头介绍了美早樱桃的设施栽培类型(塑料大棚和日光温室)，增强了回答的针对性和可信度。 **总体优势**:PengKGPT的回答更具实践指导价值，通过具体数据和阶段划分，直观展示了其在农业专业知识上的深度和准确性，优于其他模型的泛泛而谈。

图 8　不同模型的输出对比

Fig. 8　Comparison of outputs from different models

3.3　双重引导提示的新型检索增强生成方法效果

图 7 为包含向量数据库和 KG 信息的提示模板对用户输入的内部响应过程，图 9 为仅包含向量数据库信息的提示模板对用户输入的内部响应过程。

图 9 展示了用户输入问题后，直接检索向量数据库信息并输入到提示模板中，经过 LLM 推理给出答案。图 9 中的 context 包含的 3 条相关文本信息均与用户输入的问题相关；答案内容也直接地回答了用户输入的问题。图 7 展示了用户输入问题后，通过多源知识增强 LLM 框架，将检索到的向量数据库和 KG 中的信息输入到提示模板中，经过 LLM 推理给出答案。图 7 中的 context 包含的 3 条相关文本信息，其中有 1 条信息（第 3 条）关于开花期的温湿度控制，并没有和用户输入的问题有直接关联；答案内容不仅直接地回答了用户输入的问题，还包含较多的拓展信息，如增加了开花期和 2 个时期的湿度控制信息。与图 9 中的答案相比，图 7 中的答案展现出更为详尽且经过深思熟虑的特性，表现出更高的全面性。

通过对设施园艺领域的专业知识分析可知，樱桃的花蕾和叶芽发育处于同一时间段内，这段时间内樱桃的生长对于温度和湿度比较敏感。然而，温室内的环境因子温度和湿度关联性极强，温度变化必定会引起湿度变化。因此，在环境调控中，应该充分考虑温度和湿度阈值及其之间的关系。樱桃生长萌芽期结束后，立即进入开花期，这 2 个时期的分界线在生产中并不明显。因此，在考虑萌芽期的环境控制时，应考虑下一时期的环境需求，做到对樱桃生长中各个环节的全面把控。图 7 中的答案完美地满足了上述需求。可见，PengKGPT 在解决用户输入问题的同时，适度地提醒了用户其他环节的注意事项，展示出了更强的关联信息推理与扩展能力，其回答更具全面性和预见性。这充分证明了本研究中提

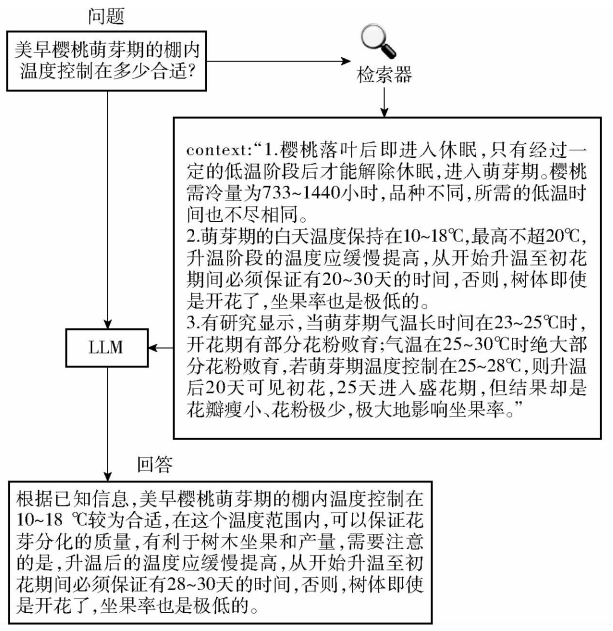

图 9 提示模板对用户输入的内部响应过程
(仅包含向量数据库)

Fig. 9 Internal response process of prompt template to user input (vector database only)

出的双重引导提示的新型检索增强生成方法在辅助设施园艺生产中做出了重要贡献。

3.4 辅助设施园艺生产的多源知识增强 LLM

目前,国内的如文心一言和国外的如 GPT-4o 等 LLM 在许多任务中表现良好。然而,这些商业 LLM 都是非开源模型,无法对其定量测量,从而它们的结果只能作为参考[37]。因此,为了进行公平的比较,根据最近研究中对 LLM 评价的调查[38],采用对相同特定问题的回答结果作为评价所选模型的基准,包括 GPT-4o、ERNIE 4.0 Turbo、ChatGLM-6B 和 PengKGPT。在这方面,本文邀请了 5 位设施园艺领域专业人员对 50 个与设施园艺生产相关的模型的回答进行排名(5 位专家包括 2 名设施园艺领域教授,1 名农业工作站研究员和 2 名一线工作 30 余年的从业者。50 个问题包括:病虫害防治、温室设施环境调控建议、农事操作建议、设施选型及作物选品、品种介绍、水肥管理等,基本覆盖设施园艺生产程序。50 个问题的选择标准为:①领域覆盖性。基本覆盖设施园艺生产程序,且各场景问题占比均衡。②任务多样性。包括描述性问答、推理性判断及策略性决策。③难度分级性。每一场景问题选取不同难度梯度的问题,以平衡评测任务的层次性)。对于每个问题,模型回答的打分从 1 到 10 分(单题分级:1~5 分为差,6~7 分为良好,8~10 分为优秀;单题总分为 500 分),求其平均值获得每个模型的评分(评估标准为:①准确性。答案是否符合专业知识。②完整性。回答是否覆盖问题关键要点。③实用性。答案在实际场景决策或应用执行中的可用价值。综合考量这 3 类标准进行评估。为验证人工评测结果的一致性,计算了评估者间信度,采用 Fleiss' Kappa 系数进行统计检验)。本文用中文评估了模型对不同问题的响应,最终 Fleiss' Kappa 系数为 0.62,表明评估者间具有良好一致性,验证了人工打分结果的可靠性。

ERNIE 4.0 Turbo、GPT-4o、ChatGLM-6B、PengKGPT 模型的总得分率分别为 81%、77.2%、54.6%、91.2%。与 ChatGLM-6B 基础模型相比,PengKGPT 生成的答案总得分率提高了 36.6 个百分点,表现出优越性能,这表明多源知识增强 LLM 的方法起到显著作用。GPT-4o 和 ERNIE 4.0 Turbo 在回答设施园艺相关问题时也表现良好,50 道问答的总得分率分别仅比 PengKGPT 低 14 个百分点和 10.2 个百分点。尤其是 ERNIE 4.0 Turbo,这是由于其训练语料库中中文信息更丰富。但是由于 GPT-4o 和 ERNIE 4.0 Turbo 未开源,且需要月费才可以使用,使得使用门槛显著高于 PengKGPT。

由于本文的模型主要服务对象为中国的设施园艺从业者,因此使用中文评估模型的响应。图 8 给出了问答的具体实例及分析点评,结果是从 3 轮问答的最佳结果中选出。本文用中文提出问题,所有的回答都是真实测试给出的。可以看出,这些模型在回答设施园艺相关问题方面展示出一定程度的专业知识。与其他 LLM 相比,ChatGLM-6B 的答案不仅简短,而且脱离了正确答案范围,可靠性较差。GPT-4o 和 ERNIE 4.0 Turbo 可以回答设施园艺中的具体问题,但回答中大多数内容没有价值,更加笼统,不够集中。也就是说,该回答不仅可以回答上述品种的问题,还可以应用于其他品种同样的问题。相比之下,PengKGPT 的回答更加详细、精准,同时还会将与此问题相关联的信息添加在回答中。这是因为 PengKGPT 的提示模板中加入了技术相关 KG,使得答案不局限于问题,回答更加全面,更具有指导生产的意义。这也表明,本文提出的双重引导提示的检索增强生成方法,有助于提高 LLM 在设施园艺领域解决相关问题的能力。

本文采用 2.4.3 节中的准确率计算方法来评价各模型答案的效果。表 6 为各模型生成答案的准确率对比。PengKGPT 模型的准确率较基座大模型提高了 32.53 个百分点,较微调后的基座大模型提高了 18.89 个百分点;微调后的基座大模型准确率较没微调的基座大模型提高了 13.64 个百分点。表明 LoRA 微调方法对 LLM 领域适应性方面带来了大幅度的改进,同时本文提出的框架进一步提高了

表 6　各模型生成答案的准确率对比

Tab. 6　Comparison of answer accuracy among different models　%

模型	准确率
GPT－4o	69.41
ERNIE 4.0 Turbo	72.06
ChatGLM－6B	49.57
ChatGLM－6B(微调后)	63.21
PengKGPT	82.10

PengKGPT 模型的准确率。PengKGPT 模型的准确率最高，达到了 82.10%，与 ERNIE 4.0 Turbo 和 GPT4o 经典商业模型相比，准确率分别提高 10.04 个百分点和 12.69 个百分点，表明其更适用于设施园艺领域的任务，这与专业人员对模型的评分结果一样。

3.5　讨论

PengKGPT 是人工智能技术应用于设施园艺领域内的一次良好表现。为中国设施园艺产业智慧化发展探索了一条道路，可为未来具身智能技术在设施园艺领域应用提供基础。但是，本文的数据集还在不断扩大中，目前仅包含了种植面积较大的设施园艺作物的数据，对于其他设施园艺栽培的作物数据，还在收集中。

本文提出的方法在自建设施园艺数据上响应较好，但它仍存在一定的局限性。在构建 KG 时，尽管经过了准确性评估，但大模型抽取知识因领域文本跨句指代、术语多义等问题，致实体相关判定有偏差；通用大模型在该领域细粒度识别与抽取能力有限，在面对复杂句式、隐含因果关系及条件语义时容易产生误判；自动构建的三元组尚缺乏完备的逻辑一致性与语义约束校验，部分复合操作被错误抽象为因果关系，从而引发局部语义偏差；进而使得构建完成的 KG 仍存在少量错误实体或关系。

这类噪声在 KG 增强 LLM 的链路中可能被放大，对最终回答质量产生影响。例如，在检索阶段，错误实体关系会让 LLM 检索到不相关或错误信息，且错误会在后续传播中进一步影响模型理解与回答。具体而言，如果 KG 中存在错误的“属于”关系，将某个实体错误地关联到某个类别中，那么 LLM 在回答涉及该实体的问题时，就可能给出错误的答案。系统对这类潜在噪声的鲁棒性取决于多个因素。一方面，如果模型在训练过程中接触过大量包含噪声的数据，并且 LLM 具有一定的噪声过滤能力，那么模型可能对 KG 中的小范围噪声具有一定的容忍度。另一方面，KG 的结构和规模也会影响噪声的传播范围。一个结构清晰、规模适中的 KG，可能更容易定位和纠正噪声，从而减少其对模型回答质量的影响。因此，下一步工作中，在知识检索阶段，进行多源一致性验证，在生成阶段，进行语义冗余校正，同时加入反馈式自校正机制，进一步提高系统的鲁棒性。LLM 的响应水平与数据质量有较大的关系，因此，应拓展更多的高质量数据，并构建相关的评价数据及评价标准，进一步检验模型的性能。同时收集更多的问答数据，进一步提高 PengKGPT 的性能。

4　结论

(1)本文提出了一个多源知识增强的大语言模型 PengKGPT，用于辅助设施园艺生产。首先构建了设施园艺领域数据集，并基于此构建了 KG 和向量数据库。设计了基于 KG 实体匹配的语义信息增强模型，挖掘了 KG 实体之间的潜在关联，通过实体匹配的方式，增强用户输入的语义信息。同时，为了提高相关问题分析的专业性和可靠性，构建真实用户问答语料的数据集，并基于此微调了开源模型 ChatGLM。通过设计具有 KG 和向量数据库双重引导提示的检索增强生成方法，进一步提升了模型推理的完整性和可靠性。

(2)结果表明，基于 KG 实体匹配的语义信息增强模型可以改变原 KG 的实体向量，即增强了实体的语义信息，起到增强语义信息的作用。此外，在设施园艺领域问答结果对比中，PengKGPT 的得分率和准确率最高，达到了 91.2% 和 82.10%，基座模型 ChatGLM－6B 得分率和准确率只有 54.6% 和 49.57%，而经典商业模型 GPT－4o 和 ERNIE 4.0 Turbo 得分率分别为 77.2% 和 81.0%，准确率分别为 69.41% 和 72.06%。表明 PengKGPT 的回答更具有价值，可以解决生产中的相关问题，实现辅助设施园艺生产。

参　考　文　献

[1]　李天来，齐明芳，孟思达. 中国设施园艺发展 60 年成就与展望[J]. 园艺学报，2022，49(10)：2119－2130.
LI Tianlai, QI Mingfang, MENG Sida. Sixty years of facility horticulture development in China: achievements and prospects [J]. Acta Horticulturae Sinica, 2022, 49(10): 2119－2130. (in Chinese)

[2]　张晓文，杨晓霞，刘艳鹏. 大食物观下中国非耕地设施农业建设与发展[J]. 蔬菜，2024(4)：1－12.
ZHANG Xiaowen, YANG Xiaoxia, LIU Yanpeng. Construction and development of facility agriculture of nonarable land in

China from the perspective of an all-encompassing approach to food[J]. Vegetables, 2024(4): 1 - 12. (in Chinese)
[3] 赵瑞雪,杨晨雪,郑建华,等. 农业智能知识服务研究现状及展望[J]. 智慧农业(中英文),2022,4(4):105 - 125.
ZHAO Ruixue, YANG Chenxue, ZHENG Jianhua, et al. Agricultural intelligent knowledge service: overview and future perspectives[J]. Smart Agriculture, 2022, 4(4): 105 - 125. (in Chinese)
[4] 赵春江. 农业知识智能服务技术综述[J]. 智慧农业(中英文),2023,5(2):126 - 148.
ZHAO Chunjiang. Agricultural knowledge intelligent service technology: a review[J]. Smart Agriculture, 2023, 5(2): 126 - 148. (in Chinese)
[5] 吕璐成,韩涛. 人工智能赋能知识服务,开启智能数字农业未来——2020 全国图书情报青年学术论坛会议综述[J]. 农业图书情报学报,2021,33(12):83 - 88.
LÜ Lucheng, HAN Tao. Artificial intelligence enables knowledge service and opens up the future of intelligent agriculture: review of 2020 national library and information youth academic forum[J]. Journal of Library and Information Science in Agriculture, 2021, 33(12): 83 - 88. (in Chinese)
[6] 傅隆生,宋珍珍,ZHANG Xin,等. 深度学习方法在农业信息中的研究进展与应用现状[J]. 中国农业大学学报,2020,25(2):105 - 120.
FU Longsheng, SONG Zhenzhen, ZHANG Xin, et al. Applications and research progress of deep learning in agriculture[J]. Journal of China Agricultural University, 2020, 25(2): 105 - 120. (in Chinese)
[7] 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 the research status and prospects of agricultural knowledge graphs[J]. Transactions of the Chinese Society for Agricultural Machinery, 2024, 55(6):1 - 17. (in Chinese)
[8] 张海瑜,陈庆龙,张斯静,等. 基于语义知识图谱的农业知识智能检索方法[J]. 农业机械学报,2021,52(增刊):156 - 163.
ZHANG Haiyu, CHEN Qinglong, ZHANG Sijing, et al. Intelligent retrieval method of agricultural knowledge based on semantic knowledge graph[J]. Transactions of the Chinese Society for Agricultural Machinery, 2021, 52(Supp.): 156 - 163. (in Chinese)
[9] LI Z, LIU H, ZHANG Z, et al. Learning knowledge graph embedding with heterogeneous relation attention networks[J]. IEEE Transactions on Neural Networks and Learning Systems, 2022, 33(8): 3961 - 3973.
[10] 郭旺,杨雨森,吴华瑞,等. 农业大模型:关键技术、应用分析与发展方向[J]. 智慧农业(中英文),2024,6(2):1 - 13.
GUO Wang, YANG Yusen, WU Huarui, et al. Big models in agriculture: key technologies, application and future directions [J]. Smart Agriculture, 2024, 6(2): 1 - 13. (in Chinese)
[11] DU Z, QIAN Y, LIU X, et al. GLM: general language model pretraining with autoregressive blank infilling[C]// Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Dublin, Ireland: Association for Computational Linguistics, 2022: 320 - 335.
[12] KUSKA M T, WAHABZADA M, PAULUS S. AI for crop production-where can large language models (LLMs) provide substantial value? [J]. Computers and Electronics in Agriculture, 2024, 221: 108924.
[13] LEE P, BUBECK S, PETRO J. Benefits, limits, and risks of GPT - 4 as an AI chatbot for medicine[J]. New England Journal of Medicine, 2023, 388(13): 1233 - 1239.
[14] KIM J, KWON Y, JO Y, et al. KG - GPT: a general framework for reasoning on knowledge graphs using large language models[J]. arXiv Preprint, arXiv:2310.11220, 2023.
[15] ZHOU C, LIU P, XU P, et al. Lima: less is more for alignment[J]. Advances in Neural Information Processing Systems, 2023, 36: 55006 - 55021.
[16] MA C, CHEN Y, WU T, et al. Large language models meet knowledge graphs for question answering: synthesis and opportunities[J]. arXiv Preprint, arXiv:2505.20099, 2025.
[17] 陈雷, 周娜, 朱鈡璇, 等. 农业知识图谱构建数据集[J]. 农业大数据学报, 2024, 6(1): 1 - 8.
CHEN Lei, ZHOU Na, ZHU Pengxuan, et al. A dataset for constructing agricultural knowledge graph[J]. Journal of Agricultural Big Data, 2024, 6(1): 1 - 8. (in Chinese)
[18] 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 diseases and pests based on knowledge graph[J]. Transactions of the Chinese Society for Agricultural Machinery, 2023, 54(3): 291 - 300. (in Chinese)
[19] 韦紫君,宋玲,胡小春,等. 基于实体级遮蔽 BERT 与 BiLSTM - CRF 的农业命名实体识别[J]. 农业工程学报,2022,38(15):195 - 203.
WEI Zijun, SONG Ling, HU Xiaochun, et al. Named entity recognition of agricultural based entity-level masking BERT and BiLSTM - CRF[J]. Transactions of the CSAE, 2022, 38(15):195 - 203. (in Chinese)
[20] 杨波,孙晓虎,党佳怡,等. 面向医疗问答系统的大语言模型命名实体识别方法[J]. 计算机科学与探索,2023,17(10):2389 - 2402.
YANG Bo, SUN Xiaohu, DANG Jiayi, et al. Named entity recognition method of large language model for medical question answering system[J]. Journal of Frontiers of Computer Science and Technology, 2023, 17(10): 2389 - 2402. (in Chinese)

[21] GUI H, YUAN L, YE H, et al. IEPile: unearthing large scale schema-conditioned information extraction corpus[C]// Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), 2024: 127 – 146.

[22] HAIDAR L R, SUBROTO I M I, MARWANTO A. Expert system on soybean disease using knowledge representation method [J]. Journal of Telematics and Informatics, 2017, 5(1): 36 – 46.

[23] SATHYANARAYANAN D, KRISHNAMURTHY M, et al. Efficient intelligent generic recommendation knowledge graph in education domain using association rule mining and machine learning[J]. International Journal of Innovative Technology and Exploring Engineering, 2019, 8(10): 2291 – 2296.

[24] CEN Y, ZOU X, ZHANG J, et al. Representation learning for attributed multiplex heterogeneous network[C]// Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2019: 1358 – 1368

[25] ZHANG Z, HAN X, LIU Z, et al. ERNIE: enhanced language representation with informative entities[J]. CoRR, 2019, abs/1905.07129.

[26] ANNERVAZ K M, BASU ROY CHOWDHURY S, DUKKIPATI A. Learning beyond datasets: knowledge graph augmented neural networks for natural language processing[C]// Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, (Volume 1 Long Papers). New Orleans, Louisiana: Association for Computational Linguistics, 2018: 313 – 322.

[27] LIU Y, WAN Y, HE L, et al. KG – BART: knowledge graph-augmented BART for generative commonsense reasoning[C]// Proceedings of the AAAI Conference on Artificial Intelligence, 2021, 35(7): 6418 – 6425.

[28] OUYANG L, WU J, JIANG X, et al. Training language models to follow instructions with human feedback[J]. Advances in Neural Information Processing Systems, 2022, 35: 27730 – 27744.

[29] ZHU X, XIE Y, LIU Y, et al. Knowledge graph-guided retrieval augmented generation[J]. arXiv Preprint, arXiv:2502.06864, 2025.

[30] MA S, XU C, JIANG X, et al. Think-on-Graph 2.0: deep and faithful large language model reasoning with knowledge-guided retrieval augmented generation[C]// ICLR, 2025.

[31] HU E J, SHEN Y, WALLIS P, et al. LoRA: low-rank adaptation of large language models[C]// ICLR, 2022.

[32] MAO Y, HE P, LIU X, et al. Generation-augmented retrieval for open-domain question answering[C]// ACL‖IJCNLP, 2021: 4089 – 4100.

[33] MATSUMOTO N, MORAN J, CHOI H, et al. KRAGEN: a knowledge graph-enhanced RAG framework for biomedical problem solving using large language models[J]. Bioinformatics, 2024, 40(6): btae353.

[34] KUMAR M, WILSON S, GOEL N, et al. KrishiAI: architecture for harnessing capabilities of LLMs for delivery of accurate cultivation information to farmers[M]// SAINI M K, GOEL N, MIGUEZ M, et al. Agricultural-centric Computation. Cham, Switzerland: Springer Nature, 2025: 68 – 79.

[35] XIAO S, LIU Z, ZHANG P, et al. C – Pack: packed resources for general Chinese embeddings[C]// Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval. Washington DC, USA: Association for Computing Machinery, 2024: 641 – 649.

[36] ZHOU B, BAO J,LIU Y, et al. BA – IKG: BiLSTM embedded ALBERT for industrial knowledge graph generation and reuse [C]// 2020 IEEE 18th International Conference on Industrial Informatics (INDIN), Warwick, United Kingdom, 2020: 63 – 69.

[37] WU C, ZHANG X, ZHANG Y, et al. PMC – LLaMA: further finetuning LLaMA on medical papers[J]. arXiv Preprint, arXiv:2304.14454, 2023, 2(5): 6.

[38] PENG B, LI C, HE P, et al. Instruction tuning with GPT – 4[J]. arXiv Preprint, arXiv:2304.03277, 2023.