

农作物病虫害知识图谱与问答系统构建研究

董文虎¹, 赵瑞雪^{1,2*}

(1. 中国农业科学院农业信息研究所, 北京 100081; 2. 国家新闻出版署农业融合出版知识挖掘与知识服务重点实验室, 北京 100081)

摘要: 农作物病虫害是制约农业发展、造成农作物减产与品质下降的重要限制条件, 构建知识图谱以合理整合农作物病虫害数据, 通过问答的方式解决农业生产过程中遇到的农作物病虫害问题, 为其提供科学、合理的病虫害防治知识, 对于实现农业信息化具有重要意义。本研究旨在概述农业知识图谱在农作物病虫害领域的应用现状及相关问答系统的构建现状及关键技术, 总结存在的问题; 通过构建病虫害知识图谱与问答系统进行实证分析与可视化展示, 完成农作物病虫害问答框架设计、自然语言理解配置及数据构建、对话管理设计、知识图谱查询以及原型系统展示, 实现了农作物病虫害多轮问答。针对现存问题进行分析并对农作物病虫害问答系统的未来发展方向进行展望, 为相关知识图谱及问答系统的构建提供参考。

关键词: 农作物病虫害; 知识图谱; 自然语言理解; 对话管理; 问答系统; Rasa

中图分类号: S433; TP391.1

文献标识码: A

文章编号: 2096-5877(2025)01-0069-09

Research on the Construction of Knowledge Graph and Q&A System for Crop Diseases and Pests

DONG Wenhui¹, ZHAO Ruixue^{1,2*}

(1. Agricultural Information Institute of CAAS, Beijing 100081; 2. Key Laboratory of Knowledge Mining and Knowledge Services in Agricultural Converging Publishing, National Press and Publication Administration, Beijing 100081, China)

Abstract: Crop pests and diseases are important limiting factors that restrict agricultural development and cause crop yield reduction and quality decline. Constructing knowledge graphs to reasonably integrate crop pest and disease data, solving crop pest and disease problems encountered in agricultural production through question-and-answer methods, and providing scientific and reasonable pest and disease prevention knowledge are of great significance for realizing agricultural informatization. This paper aims to summarize the current application status of agricultural knowledge graphs in the field of crop pests and diseases, the current status of the construction of related question-and-answer systems, and key technologies, and summarize the existing problems; by constructing pest and disease knowledge graphs and question-and-answer systems for empirical analysis and visualization, the design of the crop pest and disease question-and-answer framework, natural language understanding configuration and data construction, dialogue management design, knowledge graph query, and prototype system display are completed, and multi-round question-and-answer of crop pests and diseases is realized. The existing problems are analyzed and the future development direction of the crop pest and disease question-and-answer system is prospected, providing a reference for the construction of related knowledge graphs and question-and-answer systems.

Key words: Pests and diseases; Knowledge graph; Natural Language Understanding; Dialogue Management; Question and Answer system; Rasa

收稿日期: 2024-06-03

基金项目: 科学技术部科技创新 2030—新一代人工智能重大项目(2021ZD0113705)

作者简介: 董文虎(1998-), 男, 在读硕士, 主要从事知识图谱构建与问答研究。

通信作者: 赵瑞雪, 女, 博士, 研究员, E-mail: zhaoruixue@caas.cn

农作物病虫害是制约农业发展、造成农作物减产与品质下降的重要限制条件,据全国农技中心对2024年全国农作物重大病虫害发生趋势的预测显示,主要作物重大病虫害的发生呈重发态势,发生面积近年来不断增加,对农业生产造成重大威胁^[1]。因此,及时、准确地获取农作物病虫害相关知识与处理方法,明确病虫害的发生规律及危害情况,对于病虫害有效防治和绿色生产具有重要意义^[2]。区别于传统信息搜索引擎,问答系统能更好地满足用户的信息需求,及时高效地回答用户提出的问题^[3],在用户知识服务中扮演着重要角色。构建农作物病虫害问答系统,准确理解用户意图、将结构化的农作物病虫害防治知识传递给用户,能够提高病虫害防治的及时性、有效性、精准性,在实现农作物病虫害防治智能化与推动农业生产科研发展方面发挥重要作用^[4],是满足国家农作物病虫害防治科技创新、信息技术普及和推广应用要求的重要手段,有助于实现农作物病虫害防治的智能化、专业化和绿色化。

农作物病虫害知识图谱是农业领域知识图谱在病虫害领域的典型应用,在推动农业信息整合^[5],高效处理农业数据、挖掘农业知识,助力农业决策与知识服务等方面提供更全面、准确的支持^[6]。将农作物病虫害领域数据与知识图谱技术相结合,以图的形式存储、管理农作物病虫害中的实体与关系,进而更好地组织、管理和理解海量信息^[7]以及返回给用户更全面的知识体系^[8],能够解决病虫害领域数据分散、多样、孤岛化、数据价值利用不高的问题^[9],为问答系统提供丰富的知识和语义关系。通过建立基于农作物病虫害知识图谱的问答系统,增强系统问题语义和知识语义的理解和匹配能力^[10],以问答的方式为农业生产人员提供科学、合理的病虫害防治知识,对于农业信息化具有重要的理论与实践意义^[11]。

1 农作物病虫害知识图谱及问答应用现状

1.1 农作物病虫害知识图谱

知识图谱在2012年被谷歌公司提出,旨在提高搜索结果质量并提升用户的搜索体验^[12],可分为通用知识图谱与领域知识图谱。领域知识图谱面向特定领域,突出专业深度,对特定行业场景的多源异构信息进行整合和分析,在军事、金融、医疗、情报、农业等领域广泛应用^[13],并在智能搜

索、智能问答、智能推荐、情报分析等方面发挥了重要作用^[14]。农业知识图谱是知识图谱技术在农业领域的重要应用,为农业知识工程的发展带来了新的思路和机遇,目前在作物病虫害、种质资源库、畜禽疫病、林业、渔业、中草药以及经济作物等农业细分领域知识库建设以及病虫害可视化分析等方面发挥了重要作用^[15]。农作物病虫害具有以下特性:多样性,农作物病虫害种类繁多、表现形式多样;多源异构,数据源、数据结构繁多;复杂性,数据需要整合多方面信息,关系复杂;危害性,病虫害对农业生产具有直接危害,数据整合与应用落地具有实践意义。

知识图谱由数据层和模式层两部分构成^[16]。数据层是实际存储在知识图谱中的实体、属性和关系等具体的数据,模式层是用来描述知识图谱中实体、属性和关系的语义结构。目前多采用本体作为知识图谱的模式层,借助本体定义的规则和公理约束知识图谱的数据层^[17]。在农作物病虫害领域,部分研究面向特定农作物,探讨某作物具有的不同的特性,致力于解决该作物知识图谱构建细粒度定义问题,如花卉^[18]、樱桃^[19]、苹果^[20]、小麦^[21]等。另一部分研究未对作物明确区分,其面向知识图谱本体与数据层构建的共性问题,提高知识图谱的借鉴性和可移植性,如李悦等^[22]借鉴融合多来源、多载体的专业化高质量知识资源将农作物病虫害领域本体与通用的科技文献本体以及传统的知识组织体系进行深度融合以构建病虫害本体。

在知识图谱数据层面,农作物病虫害数据的主要来源有结构化知识、半结构化知识、非结构化知识。在结构化知识方面,其数据来源主要为实验室自有病虫害数据集、以关系型数据库等方式保存的病虫害数据等。在半结构化知识方面,相关研究数据来源主要为网络数据库,利用爬虫等技术爬取相关网站中的数据,如百科、中国作物种质信息网、中国农业信息网等。在非结构化知识方面,相关研究主要利用深度学习方法进行抽取,数据来源为作物病虫害防治书籍等,如周焯等^[23]利用基于FastBert模型针对水稻病虫害的中文实体关系进行抽取;吴赛赛等^[24]利用Bert-BiLSTM-CRF进行抽取,提升了三元组提取效果;刘合兵等^[25]构建了小麦病虫害命名实体识别语料库,提出了基于规则修正的中文命名实体识别模型WPD-RA进行小麦病虫害实体识别,为相关领域实体识别提供了借鉴。

1.2 病虫害知识图谱驱动的问答系统应用

使用知识图谱技术构建面向具体农业任务的

问答系统有助于帮助农户快速、精准地解决农业领域内的专业问题,将农作物病虫害知识图谱技术应用于一代农业系统知识问答产品,有助于加速农业创新,满足农业知识自动化获取需求^[26]。最初农业知识问答系统通过计算实体相似度实现,需要匹配大量的农业知识问答库,效率较低。现在随着知识融合与知识推理技术在农业中的不断发展,农业知识问答系统具有了一定的扩展能力,在问答效率提升的同时保证了结果的准确率。如郭旭超等^[27]提出了一种基于意图嵌入信息和槽位门控机制的意图识别与槽-位填充联合模型(AgIG-IDSF),在共享编码模块引入了注意力机制用于丰富上下文语义特征,同时,提出了一种融合意图嵌入表示和槽位门控机制的意图-槽位交互方法用以增强意图信息指导槽位填充任务的能力,进而提高模型的整体识别性能;赵赛等^[28]设计实现了包含闲聊、症状、防治、属性、周期和其他等共6类意图的马铃薯病虫害问答系统;郑泳智等^[29]在荔枝与龙眼知识图谱的基础上,利用AC自动机、Bert+BiLSTM+CRF、Bert+RNN等完成了病虫害知识问答系统;赵纪帅^[30]采集玉米、水稻和小麦的病虫害相关数据构建存储于Neo4j的知识图谱,利用训练好ALBERT-BiLSTM-CRF模型提取问句的关键实体,然后结合ALBERT-TextCNN模型的问句分类结果,生成对应的查询语句

以实现问答功能等。

综合来看,目前农业问答领域的工作主要集中在以下几个部分:问答数据集构建、实体识别与意图识别任务、知识存储与检索、答案生成与推荐。然而,现有研究在问答系统的通用性、可移植性、多轮对话等方面仍存在不足,本研究在此基础上结合知识图谱与Rasa框架,完成了问答系统框架设计、农作物病虫害数据集定义、对话管理及知识图谱查询与回复语句构建等关键任务,针对农作物病虫害问答场景,设计了相关意图与数据集,实现了农作物病虫害多轮问答并进行可视化展示,为相关系统设计与实现提供了理论参考。

2 农作物病虫害知识图谱与问答系统构建实证

2.1 农作物病虫害知识图谱

2.1.1 农作物病虫害本体

本研究参考了部分研究提出的细粒度农作物病虫害知识图谱本体,同时在此基础上,根据问答系统的具体任务以及数据源数据结构对农作物病虫害涉及的概念类型及关系属性等进行了提炼与总结,确定了农作物病虫害领域的知识和范围,设计的本体概要模型如图1所示。

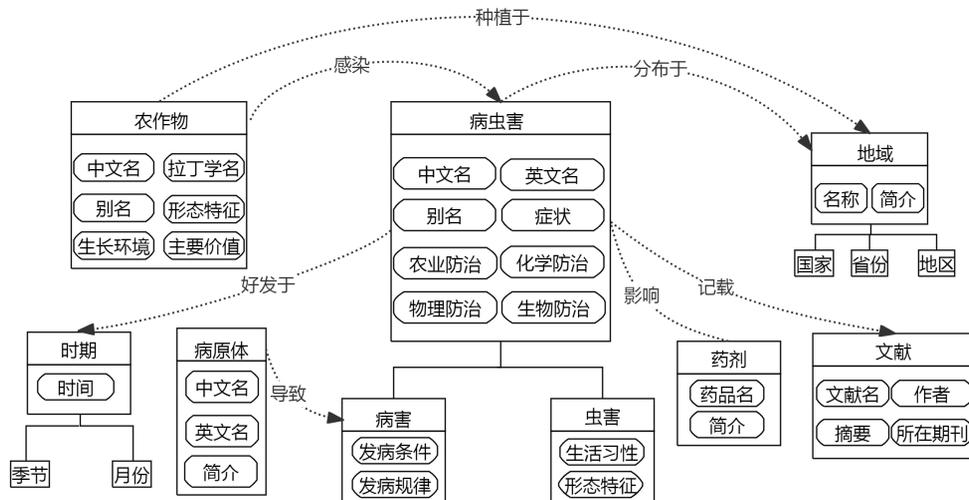


图1 农作物病虫害本体设计

由该本体概要模型可知,该本体从核心对象病虫害出发,将其分为病害与虫害两类,病虫害类均具有中英文名称、别名、症状等不同属性;病害可由不同的病原体导致,有不同的发病规律、发病条件等,虫害有相应的生活习性、形态特征;病虫害的为害主体是农作物,农作物具有名称、

形态特征、生长环境等不同属性;病虫害与农作物分布于不同的地域,包括国家、省份、地区等;不同的病虫害好发于不同的时期,如不同月份、季节等;可利用不同的药剂进行防治,同一种药剂又可防治数种病虫害;关于病虫害的研究文献可在对应的文献中进行查找,文献本身具有名

称、摘要等信息。不同实体之间采用“实体-关系-实体”的方式对其之间的关系进行描述,如水稻会感染水稻黄叶病,其中“水稻”及“水稻黄叶病”为实体,“感染”为它们之间的关系,不同的实体包含不同的属性,采用“实体-属性-属性值”对该实体进行具体描述。

2.1.2 农作物病虫害数据

本研究知识图谱数据层构建参考了部分农业病虫害领域专业书籍,包括系统划分并概括组织了农业学科主题概念的《农业科学叙词表》,以专业叙词表的形式组织了农业学科领域相关病虫害主题概念的《中国农作物病虫害(第三版)》以及相关的学术期刊文献等,保证了病虫害数据内容的准确与权威性。依据已有的原始数据,确定了常见的农作物,在此基础上利用Scrapy爬取相关的数据,完成数据集构建。病虫害数据依据“农

业病虫害云数据库”中的数据以及百科中的数据等,针对每种农作物常见病虫害与相关数据进行筛选与扩充。农作物病虫害文献数据来源为农业专业知识服务系统数据,通过检索下载,获取某病虫害相关的文献名、摘要、期刊等相关信息。

2.1.3 农作物病虫害知识图谱存储

本研究使用Neo4j存储农作物病虫害知识图谱,其作为图数据库使用图形结构来存储数据,支持高效的图形遍历和查询^[31],能够直接表现数据的关联特性;利用专门用于查询数据库中图形数据的cypher语言来进行操作,可以更轻松地查询和操作图形数据^[32]。病虫害数据导入利用py2neo,将抽取好的实体、关系及属性等数据通过连接Neo4j数据库,以cypher语句完成构建工作,实现知识图谱的可视化展示,部分知识图谱如图2所示。

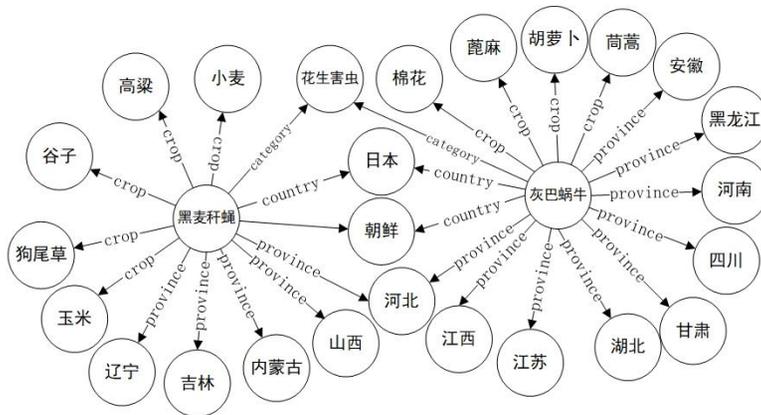


图2 农作物病虫害知识图谱范例图(部分)

2.2 问答系统构建

2.2.1 问答系统框架

问答系统构建结合Rasa框架实现,优点有更加节省资源并提高计算效率、通过有限的训练数

据进行模型训练,更加灵活和便捷。针对农作物病虫害智能问答场景,结合农业领域用户需求,设计如图3所示问答系统架构流程,问答系统主要由人机交互模块、自然语言理解模块、对话管

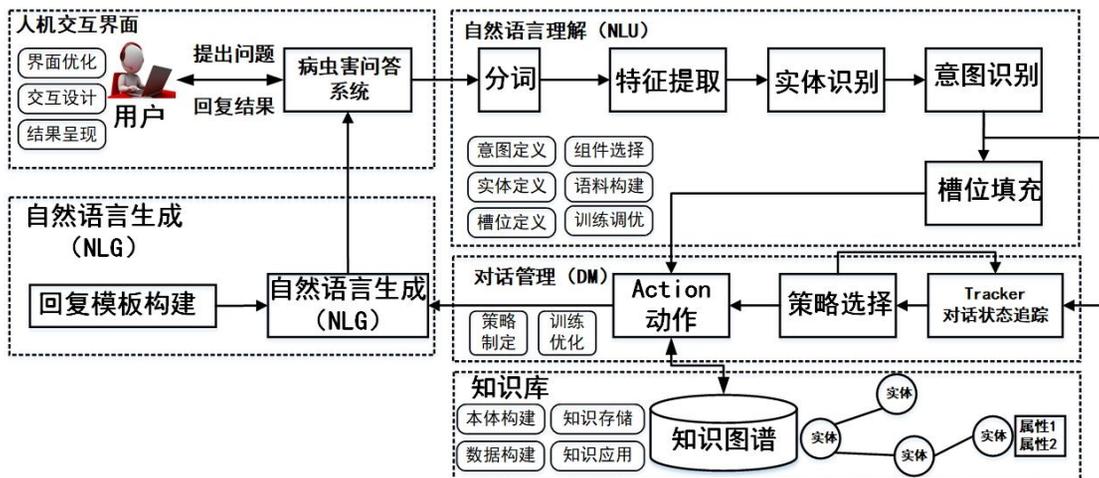


图3 问答系统原型设计图

理模块、知识库模块、自然语言生成等模块组成。

人机交互模块为问答系统的用户在进行使用时展示给用户的界面,其作用是接收用户输入的信息以及将问答系统的结果返回给用户,其任务有交互界面的优化设计、将结果以合理的方式展现给用户。自然语言理解模块基于Rasa框架的Rasa NLU模块实现,需要接收用户输入的信息,经过训练好的农作物病虫害NLU模型后结构化输出槽位、意图等信息以供对话管理模块使用。意图及实体识别是自然语言模块的主要任务,通过需求调研、问答分析理清意图、实体及槽位定义,构建合适的训练与测试集,选择合理的预训练语言模型,通过训练与评估模型,达到较好的识别效果。对话管理模块主要由Rasa Core完成,Rasa Core负责处理对话流程和决策,与用户进行交互,并根据上下文推断下一步应该采取的动作

或响应。针对农作物病虫害问答系统面向的服务人员为获取相关农作物病虫害信息目的明确的特点,在设计实现问答系统时,采用基于模板的自然语言生成方式,通过构建回复模板,将从Neo4j数据库中得到的信息整合到模板中,返回到人机交互模块以提供给用户。

2.2.2 意图及实体定义

本研究根据对中国农技推广信息服务平台、惠农网、农业种植网等网站中用户的不同设问方式的调研工作,参考了其他病虫害问答系统以及农作物病虫害知识图谱结构,共确定了三类用户意图(表1)。第一类意图为保证问答系统的可用性与完备性而设计的问答系统的基础意图(编号:1~3),如打招呼意图、无法识别的意图等,该类意图通常不涉及实体识别任务,用户问句一般也可以穷举,属于比较简单的意图分类任务;第

表1 用户意图定义

| 类别 | 编号 | 意图 | 说明 | 涉及槽位 |
|----------|--------|---------------|--------------|---------------|
| 基础意图 | 1 | 问候 | 询问系统能实现的功能 | - |
| | 2 | 未知意图 | 无法理解的意图 | - |
| | 3 | 再见 | 用户终止对话 | - |
| 与病虫害相关意图 | 4 | 相关信息 | 病虫害的相关信息 | 病害、虫害 |
| | 5 | 影响作物 | 病虫害会影响哪些作物 | 病害、虫害 |
| | 6 | 别名 | 病虫害的别名 | 病害、虫害 |
| | 7 | 症状 | 病虫害的症状 | 病害、虫害 |
| | 8 | 防治方法 | 病虫害的全部防治方法 | 病害、虫害 |
| | 9 | 防治方法 农业防治方法 | 特指农业防治方法 | 病害、虫害 |
| | 10 | 防治方法 化学防治方法 | 特指化学防治方法 | 病害、虫害 |
| | 11 | 防治方法 物理防治方法 | 特指物理防治方法 | 病害、虫害 |
| | 12 | 防治方法 生物防治方法 | 特指生物防治方法 | 病害、虫害 |
| | 13 | 时期 | 病虫害的好发时期 | 病害、虫害 |
| | 14 | 病虫害分布地域 | 病虫害主要分布在哪些地区 | 病害、虫害 |
| | 15 | 相关文献 | 病虫害相关文献 | 病害、虫害 |
| | 16 | 发病条件 | 病害的发病条件 | 病害 |
| | 17 | 发病规律 | 病害的发病规律 | 病害 |
| | 18 | 病原体 | 引发某病害的病原体 | 病害 |
| | 19 | 生活习性 | 虫害的生活习性 | 虫害 |
| | 20 | 形态特征 | 虫害的形态特征 | 虫害 |
| | 其他相关意图 | 21 | 易感病虫害 | 某作物可能受哪些病虫害感染 |
| 22 | | 生长环境 | 某作物的生长环境 | 农作物 |
| 23 | | 作物价值 | 某作物的主要价值 | 农作物 |
| 24 | | 地域 病虫害 | 某地域需要预防的病虫害 | 地域 |
| 25 | | 某时期易发病 | 某时期易发什么病虫害 | 时期 |
| 26 | | 药剂 | 某药剂可以治疗哪些病虫害 | 药剂 |

二类为查询农作物病虫害相关知识而设计与病虫害相关意图(编号:4~20),如查询某病害相关的实体等等,该类意图在农作物病虫害问答情景中较为常见,是较为常见的问答任务;第三类为其他相关意图(编号:21~26),如查询某作物可能会感染的病虫害等;共计26种用户意图,与意图相关的槽位共计6种,分别为病害、虫害、农作物、地域、时期、药剂。

2.2.3 自然语言理解

自然语言理解模块的主要功能为将自然语言文本转化为计算机可以理解和处理的形式,包含两个子任务:命名实体识别与意图识别。该部分任务通过 Rasa NLU 模块实现,以组件的形式完成的各部分功能,利用 pipeline 个性化的指定预训练模型以实现相关功能。其设计思路如图4所示。

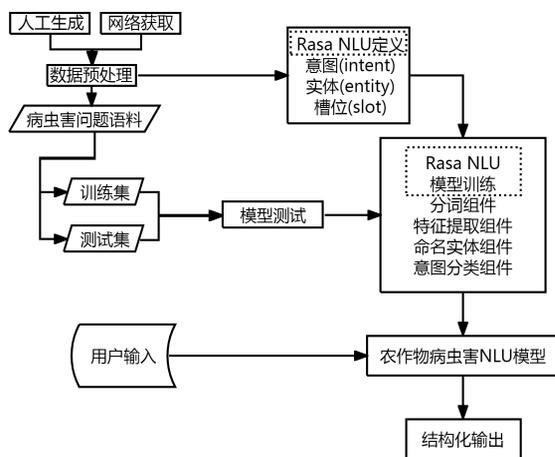


图4 自然语言理解设计流程图

首先收集农作物病虫害问答语料,语料数据集用以完成自然语言理解模型的训练及测试。通过问句及系统分析确定相关问答场景,进而确定用户在农作物病虫害场景下意图及实体类别,通过域文件(domain)完成 Rasa NLU 中意图、实体及槽位的定义。

命名实体识别及意图识别利用组件的形式实现,并利用不同的组件实现其他功能,如利用 Fallback Classifier,参数设置为 threshold: 0.5, ambiguity_threshold: 0.1,当意图识别效果为最高的意图概率低于0.5或排名最高的两个意图之间的概率差值低于0.1时判定用户语句存在歧义,要求用户提供更多信息。利用 EntitySynonymMapper 完成同义词的替换,针对同一种农作物不同的表达方式消除歧义,实现实体规范化。

2.2.4 意图及实体数据集构建

意图及实体识别数据集的构建在调研相关网

站及有关文献的基础上,采用人工编写与生成工具扩充的方式进行,主要有三部分来源。

第一部分为将用户提问数据利用开源文本标注工具 Rasa-NLU-Trainer 进行标注,数据标注方式的数据来源为利用爬虫爬取中国农技推广信息服务平台与农业种植网的问答栏目。第二部分为手工编写,针对基础意图,可利用手工编写的方式对设问方式进行穷举,尽可能全地将用户问句列出。手工编写完成的意图大多为较为简单的意图,如问候等基础意图,该部分意图语句较短、用户问法较为单一,且一般不涉及槽位填充任务,可以通过枚举的方式列出。第三部分为生成工具 Chatito 扩充,其思想是将用户某意图的提问拆解为不同的构成成分,如“实体+问句关键词+助词”的形式,通过排列组合以形成大量的问答数据集。生成并输出的 json 训练集格式示例如下:

```
{ "text": "你好,发生稻褐蚜会导致作物出现什么症状", "intent": "search_symptom ", "entities": [{"end": 8, "entity": "disease", "start": 5, "value": "稻褐蚜"}]}
```

该语句对意图、实体等信息进行了标注,表明了实体开始与结束的位置以及意图以供训练使用。其中“text”为待处理问句的句子示例,表明用户的提问方式;“intent”为在该问句下用户的意图是什么;“entities”表明了该问句中的实体,中括号内的内容注明了该实体的开始及结束位置以及实体类型。

通过人工审核编辑的形式,对形成的数据集进行筛选,并将其转换为 yaml 格式,最终完成数据集的构建工作。以查询“发病症状”为意图的部分问句训练集如表2所示。

表2 病虫害症状相关问题集(部分)

| 用户意图 | 问句示例 |
|---------|----------------------------------|
| 病虫害症状查询 | 你好,发生[花生灰霉病](disease)会导致作物出现什么症状 |
| | 发生[水稻恶苗病](disease)有哪些症状 |
| | [小麦条锈病](disease)的症状有哪些 |
| | 出现啥情况我能判断是发生了[水稻细菌性条斑病](disease) |
| | [小麦赤霉病](disease)的特征是什么 |
| | 出现[小麦网腥黑穗病](disease)时会有怎样的症状表现 |
| | |

2.2.5 对话管理

对话管理模块主要由 Rasa core 完成,它负责处理对话流程和决策,与用户进行交互,并根据

上下文推断下一步应该采取的动作或响应。其设计构建如图 5 所示。

在构建农作物病虫害问答系统时,需要先将

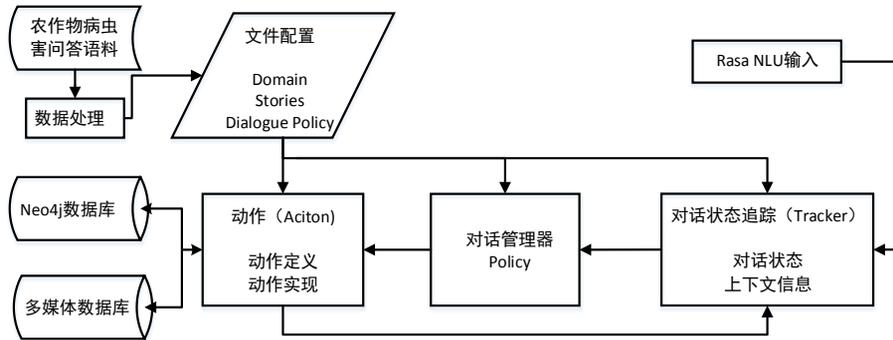


图5 对话管理设计流程图

asa core 进行配置,域文件的配置与自然语言理解任务相关,用于定义问答系统的核心元素,如意图、实体等,为问答系统提供了关键的配置信息。故事文件是问答系统用以学习对话流程的相关数据,利用 story 中给定的农作物病虫害问答步骤,Rasa core 便可以根据相应的策略学习需要执行哪些动作。故事文件可以根据调研用户问答流程、真实问答场景模拟及与问答系统交互来编写并完善。在自定义动作中,需要完成具体操作,如连接 Neo4j 数据库,利用 cypher 语句查询数据库,获取相应的数据。

对话策略用于决定在给定对话状态下问答系统应该采取什么行动,通过对故事的学习以预测需要执行的动作。本研究的对话策略采用 TED-Policy、MemoizationPolicy、RulePolicy 以及 Form-Policy。

TEDPolicy 用以预测问答系统动作。TED 策略是基于 Transformer 架构的问答策略,它综合考虑了对话历史状态、词槽、用户输入的意图及实体,将当前对话映射为一个对话向量,学习如何

从当前对话状态预测下一步行动。Memoization-Policy 会记录每个对话状态中用户的输入以及对应的对话动作,根据历史对话记录来预测下一个对话动作,在预测下一个对话动作时,如果当前对话状态与已记录的对话状态完全匹配,则会选择与该状态相关联的对话动作作为预测结果。RulePolicy 使用规则策略来处理特定的场景或用户请求。FormPolicy 用于处理用户输入的表单信息,当用户需要提供多个信息时,表单策略可以引导用户一步一步地提供这些信息,根据用户在对话中提供的信息,逐步构建表单,并最终完成填写。

2.2.6 知识图谱查询方法

在基于知识图谱的农作物病虫害问答系统中,需要完成对知识图谱的查询功能,具体实现流程如图 6 所示。

首先,通过 py2neo 连接知识图谱,Rasa core 根据用户意图、实体以及历史会话信息等选择系统将执行的具体动作,在该动作中定义了查询知识图谱的 cypher 语句以及获取到相关信息后的处理

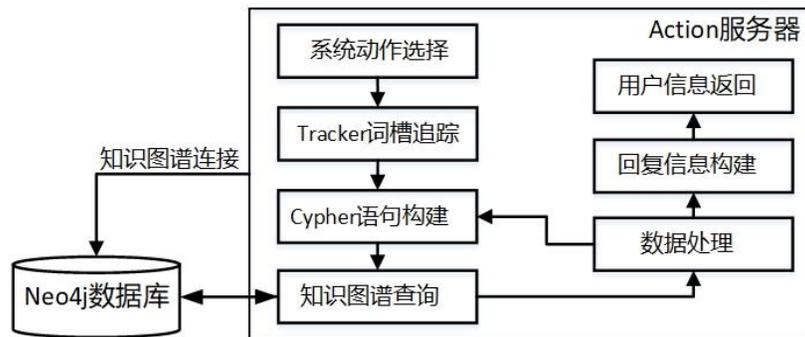


图6 知识图谱查询流程

方式。通过 tracker 对象获取用户词槽中存储的实体,进而对 cypher 中的占位符进行补充,进而构建完整的 cypher 查询语句,通过构建的 cypher 语句进行知识图谱中相关信息的查询,并将查询结果返回系统。例如查询某病害可能会影响哪些农作物,通过 MATCH(p:disease{name:\$disease})-[r:disease_crop]->(s:crop)RETURN s.name 语句,将病害名填入 cypher 语句中,进而获取其影响农作物的名称。

2.2.7 回复模板构建方法

回复语句构建主要采取了响应模板的方式和自定义操作的方式实现。响应模板的方式在域文件中进行了定义,包括意图及相应的回复语句,该方法需要对用户提问后回复的内容进行手工编写,一般而言回复都比较固定。

针对简单闲聊意图,因不涉及信息填充等任务,用户的意图比较明确,因此可以在域文件的 response 中对回复语句进行定义,在定义时,需要遵循简洁化、精准化、多样化的原则。简洁化即回复语句简洁,不宜包含太多杂乱内容;精准化为对该意图下用户所需内容的判断,将所需信息提供给用户;多样化指回复内容多样,通过构建不同的回复语句,使得问答系统从相应内容中随机选取回复语句。

自定义操作能够填充响应模板中的动态变量,也可以将相应的结果直接发送给用户。定义回复模板,如 template = "感染 {disease} 的症状如下: {symptom}", 在查询到某病害症状后,填充相关的病虫害名称信息与其对应的症状信息,将构建的信息返回给用户。若没有查询到相关症状,则选择“病虫害名称+未查询到相关结果”的方式返回给用户。通过上述构建方式针对系统定义的意图进行相关回复模板,完成回复语句构建。

2.2.8 问答系统实现结果

农作物病虫害问答系统部署采用 B/S 架构,将系统分为客户端与服务器两部分。其中服务器部分包括 Rasa 服务器与自定义动作服务器,负责接收用户的需求信息并将结果发送给用户。客户端利用 Rasa Chat Widget,部署在网页上,收集用户输入信息以及向用户展示问答结果。

问答界面如图 7 所示。首先进行打招呼等问候,因无需查询知识图谱,因此系统能够较快返回构建模板结果。对“麦叶蜂”相关信息进行问答,系统能够识别用户意图,在知识图谱中查询有关信息,并将结果返回给用户。

利用客户端对问答系统展开询问,针对不同



图 7 问答系统界面展示

询问记录其能否实现预期结果。从测试结果来看,共进行了 300 条数据测试,达到预期结果的共 262 条,系统平均准确率达到 87.33%,能够较为出色地实现农作物病虫害问答任务,完成基于知识图谱的问答功能,具有实际应用价值。

3 总结与展望

3.1 总结

本研究对农作物病虫害知识图谱与问答系统应用现状进行了总结,并进行了实证探讨,构建了农作物病虫害知识图谱,提出了与 Rasa 结合的病虫害多轮问答构建路径,设计实现了农作物病虫害问答系统,为相关领域问答系统的设计与实现提供了参考。问答系统是知识服务的高级形式,用户对农业知识问答服务需求的内因及相关技术发展的外因推动着索引式信息检索方式向问答式知识应答方式的转变^[33],领域知识图谱驱动的问答系统研究与应用于相关领域的研究人员及用户提供了更具专业性的信息。目前农业知识服务智能化、精准化继续深化,构建农业领域知识图谱,提高农业问答系统智能化水平,既是农业用户信息检索服务的需求,也是农业智能技术研究应用的需求。探索在农业垂直领域知识图谱构建的基础上实现问答系统,对领域知识图谱驱动的问答系统实现中的关键问题进行分析,以用户的现实需求为导向、以新兴科学技术为支撑,能够为用户提供精准、丰富的信息资源与农业知识,为相关农业生产研究提供专业资源,具有十分重要的现实意义。

3.2 展望

本研究仍有一定的不足之处,还有改进空间。首先,用户提问方式复杂多变,不同地域的用户设问习惯不同,目前构建的问答语料集未必能覆盖全部设问方式,仍需要收集相关语料,更贴近用户的提问语境,构建更加完善的语料库。其次,诸如大语言模型、领域知识注入等方法将能够有效提高系统在专业领域的适配性,还需要进一步探讨利用大语言模型等提升系统性能、理解用户的自然语言输入,更准确地解析意图和细节问题。第三,问答系统能实现的功能依赖于对其意图等相关信息的定义,而知识图谱经过合理的推理、判断等能够实现不同的功能,比如“稻弄蝶是否也会影响玉米”等推理判断类问题,显然系统目前未设置该类意图,无法回答此类问题。针对该类型的问题,应探索不同的实现方式,完善问答场景。第四,模板形式反馈给用户相关信息不够灵活,无法充分满足不同用户的需求,此外,未来还应考虑多模态的交互方式,探索更加合理的人机交互形式。第五,随着知识图谱的更新以及用户需求增加,系统功能也应随之改变,以满足用户的需求,适应相关技术的发展,为专业领域的知识问答提供更加权威、准确的服务。

参考文献:

- [1] 刘杰,曾娟,黄冲,等. 2024年全国农作物重大病虫害发生趋势预报[J]. 中国植保导刊, 2024, 44(1): 37-40.
- [2] 高芑,周淑香,赫思聪,等. 吉林省中部地区露地十字花科蔬菜主要害虫及其发生动态研究[J]. 东北农业科学, 2024, 49(1): 75-79.
- [3] 郑实福,刘挺,秦兵,等. 自动问答综述[J]. 中文信息学报, 2002(6): 46-52.
- [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.
- [5] 陈娉婷,官波,沈祥成,等. 大数据时代开放式农业信息知识库构建研究[J]. 东北农业科学, 2018, 43(5): 60-64.
- [6] 唐闻涛,胡泽林. 农业知识图谱研究综述[J]. 计算机工程与应用, 2024, 60(2): 63-76.
- [7] 王守会,覃颀. 知识库问答系统研究进展[J]. 小型微型计算机系统, 2021, 42(9): 1793-1801.
- [8] 李涛,王次臣,李华康. 知识图谱的发展与构建[J]. 南京理工大学学报, 2017, 41(1): 22-34.
- [9] 穆维松,刘天琪,苗子激,等. 知识图谱技术及其在农业领域应用研究进展[J]. 农业工程学报, 2023, 39(16): 1-12.
- [10] 李贺,刘嘉宇,李世钰,等. 基于疾病知识图谱的自动问答系统优化研究[J]. 数据分析与知识发现, 2021, 5(5): 115-126.
- [11] 罗治情,官波,陈娉婷,等. 大数据背景下农业农村信息化建设研究[J]. 东北农业科学, 2022, 47(4): 147-150.
- [12] 侯梦薇,卫荣,陆亮,等. 知识图谱研究综述及其在医疗领域的应用[J]. 计算机研究与发展, 2018, 55(12): 2587-2599.
- [13] 田玲,张谨川,张晋豪,等. 知识图谱综述—表示、构建、推理与知识超图理论[J]. 计算机应用, 2021, 41(8): 2161-2186.
- [14] 全国知识图谱与语义计算大会. 知识图谱发展报告(2022)[R]. 北京:高等教育出版社, 2022: 193-203.
- [15] 曹丽英,邸玉琦,陈帅. 基于知识图谱的玉米病虫害研究可视化分析[J]. 东北农业科学, 2022, 47(1): 145-150.
- [16] 徐增林,盛泳潘,贺丽荣,等. 知识图谱技术综述[J]. 电子科技大学学报, 2016, 45(4): 589-606.
- [17] 黄恒琪,于娟,廖晓,等. 知识图谱研究综述[J]. 计算机系统应用, 2019, 28(6): 1-12.
- [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.
- [19] 王栋,周菲,李颖芳,等. 我国甜樱桃产业知识图谱构建研究[J]. 中国果树, 2023(1): 104-108.
- [20] 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.
- [21] 朱越. 基于小麦病虫害知识图谱的推荐技术研究[D]. 郑州:河南农业大学, 2023.
- [22] 李悦,孙坦,鲜国建,等. 面向多源数据深度融合的农作物病虫害本体构建研究[J]. 数字图书馆论坛, 2021(2): 2-10.
- [23] 周焯,徐向英,章永龙,等. 基于FastBert的水稻病虫害实体关系抽取研究[J]. 南京师范大学学报(工程技术版), 2023, 23(1): 33-38.
- [24] 吴赛赛,周爱莲,谢能付,等. 基于深度学习的作物病虫害可视化知识图谱构建[J]. 农业工程学报, 2020, 36(24): 177-185.
- [25] 刘合兵,张德梦,熊蜀峰,等. 融合ALBERT与规则的小麦病虫害命名实体识别[J]. 计算机科学与探索, 2023, 17(6): 1395-1404.
- [26] Liu X X, Bai X S, Wang L H, et al. Review and trend analysis of knowledge graphs for crop pest and diseases[J]. IEEE Access, 2019, 7: 62251-62264.
- [27] 郭旭超,郝霞,姚晓闯,等. 农业病虫害知识问答意图识别与槽位填充联合模型研究[J]. 农业机械学报, 2023, 54(1): 205-215.
- [28] 赵赛,杨婉霞,王巧珍,等. 基于马铃薯病虫害知识图谱的问答系统[J]. 农业工程, 2023, 13(8): 29-37.
- [29] 郑泳智,吴惠彝,朱定局,等. 基于荔枝和龙眼病虫害知识图谱的问答系统[J]. 计算机与数字工程, 2021, 49(12): 2618-2622.
- [30] 赵纪帅. 基于ALBERT的农作物病虫害知识图谱构建与应用[D]. 合肥:安徽农业大学, 2023.
- [31] 李金阳. 图数据库在图书馆的应用研究[J]. 图书馆, 2020(11): 109-115.
- [32] Green A, Guagliardo P, Libkin L, et al. Updating graph databases with Cypher[J]. Proceedings of the VLDB Endowment, 2019, 12(12): 2242-2254.
- [33] 张智雄,于改红,刘熠,等. ChatGPT对文献情报工作的影响[J]. 数据分析与知识发现, 2023, 7(3): 36-42.

(责任编辑:范杰英)