于合龙, 沈金梦, 毕春光, 等. 基于知识图谱的水稻病虫害智能诊断系统 [J]. 华南农业大学学报, 2021, 42(5): 105-116.
YU Helong, SHEN Jinmeng, BI Chunguang, et al. Intelligent diagnostic system for rice diseases and pests based on knowledge graph[J]. Journal of South China Agricultural University, 2021, 42(5): 105-116.

基于知识图谱的水稻病虫害智能诊断系统

于合龙[1,2†]，沈金梦[1†]，毕春光[1,2]，梁　婕[3]，陈慧灵[4]
(1 吉林农业大学 信息技术学院，吉林 长春 130118； 2 吉林农业大学 智慧农业研究院，吉林 长春 130118；
3 悉尼科技大学 工程与信息技术学院，悉尼 2007； 4 温州大学 计算机与人工智能学院，浙江 温州 325035)

摘要:【目的】利用知识图谱对水稻病虫害领域复杂的异构数据信息进行结构化存储，建立病虫害间语义关系，为水稻病虫害关联检索及智能诊断提供理论依据。【方法】首先提出一种面向水稻病虫害的知识图谱构建方法和基于图的水稻病虫害检索算法，通过引入节气实体实现水稻病虫害的预警。其次提出基于确定性因子 (Certainty factor，CF) 模型和知识图谱相结合的知识推理方法，利用 CF 与水稻病株症状的结合实现水稻病虫害的诊断。【结果】利用命名实体识别模型，得出病、虫害名称及危害症状实体的准确率分别为 0.92、0.90 及 0.87，进一步构建包括 1 972 个实体及 5 226 个实体关系的垂直领域知识图谱。通过自主开发的智能诊断系统进行案例分析，试验表明，诊断算法正确率达到 86.25%。【结论】该系统有效地解决了水稻病虫害领域数据检索、预警与诊断中知识的复杂性及不确定性的问题，有较强的实用价值和推广前景。

关键词: 知识图谱；确定性因子模型；水稻病虫害；智能诊断
中图分类号: S435.11；TP182　　**文献标志码:** A　　**文章编号:** 1001-411X(2021)05-0105-12

Intelligent diagnostic system for rice diseases and pests based on knowledge graph

YU Helong[1,2†], SHEN Jinmeng[1†], BI Chunguang[1,2], LIANG Jie[3], CHEN Huiling[4]
(1 College of Information Technology, Jilin Agricultural University, Changchun 130118, China; 2 Institute of Smart Agriculture, Jilin Agricultural University, Changchun 130118, China; 3 College of Engineering and Information Technology, University of Technology Sydney, Sydney 2007, Australia; 4 College of Computer Science and Artificial Intelligence, Wenzhou University, Wenzhou 325035, China)

Abstract: 【Objective】To conduct structured storage of complex and heterogeneous data information in the field of rice diseases and pests using knowledge graphs, establish semantic relationships between diseases and pests, and provide a theoretical basis for rice diseases and pests association retrieval and intelligent diagnosis. 【Method】Firstly, a method of constructing a knowledge graph for rice diseases and pests was proposed. At the same time, a series of graph-based retrieval algorithms for rice diseases and pests were proposed for information mining, through introducing solar terms entities to achieve early warning of rice diseases and pests. Secondly, a knowledge reasoning method based on the combination of certainty factor (CF) model and

收稿日期:2021–01–06　　**网络首发时间:**2021–06–09 11:32:50
网络首发地址:https://kns.cnki.net/kcms/detail/44.1110.S.20210609.1114.002.html
作者简介:于合龙 (1974—)，男，教授，博士，E-mail: yuhelong@aliyun.com；沈金梦 (1995—)，女，硕士研究生，E-mail: 1757516665@qq.com；†表示同等贡献； 通信作者： 陈慧灵 (1983—)，男，副教授，博士，E-mail: chenhuiling.jlu@gmail.com
基金项目:国家自然科学基金 (U19A2061)；国家重点研发计划 (2019YFC1710700)；吉林省科技发展计划 (20190301024NY，20200301047RQ)

knowledge graph was proposed to realize the intelligent diagnosis of rice diseases and pests by combining CF with the symptom of diseased plant.【Result】The accuracy rates of named entity recognition model were 0.92, 0.90, and 0.87 in disease and pest name and hazard symptom entities. Further, a knowledge graph of rice disease and pest domain including 1 972 entities and 5 226 entity relationships was constructed. Through the self-developed intelligent diagnosis system, case analysis was conducted and the test showed that the correct rate of the diagnosis algorithm reached 86.25%.【Conclusion】This study effectively solves the complexity and uncertainty of knowledge in data retrieval, early warning and diagnosis in the field of rice diseases and pests, and has a strong practical value and extension prospects.

Key words: knowledge graph; certainty factor model; rice disease and pest; intelligent diagnosis

水稻作为中国主要的粮食作物之一，每年因病虫害造成的损失多达几百万吨[1]。各种水稻病虫害发生快、易扩散，且水稻受害植株常表现出相同的症状，让农民难以区分[2-3]。随着信息技术、光谱技术[4]、遥感技术[5] 等新方法不断出现，有的利用光谱的反射值或得到的图像特点来关联病虫害特征信息，构建病虫害模型，最终实现病虫害识别；有的通过图像技术、规则推理等方式进行作物病虫害诊断[6-11]，而在水稻领域常利用专家系统进行病虫害诊断[12-15]，它们大多从专家处获得规则算法，再根据受害症状等因素进行分类并判断。但上述研究并不能完全解决病虫害诊断中存在的问题，主要原因有以下三点：一是水稻数据种类多样，关系及属性复杂，数据间深层次的关联关系不易被挖掘，同时，也缺少相关推理过程的显示；二是传统专家系统大都是定量识别，存在推理可解释性弱的问题；三是很少有专家系统把节气与规则相结合进行病虫害预警。知识图谱以节点及边的形式将不同类型的实体、概念组合成巨大网络[16]，有利于以可视化形式展示知识的多种关系和结构，它为提高检索的可推理性及检索效率提供了机遇。随着农业信息化的普及，对农业领域异构数据的处理从原有的本体构建再到语义网，国内外学者都进行了探究[17-19]。此外，农业领域知识图谱也有成功的应用案例[20-22]，促进了知识图谱技术在农业领域的发展，初步实现了数据信息的有效利用。但知识图谱是网状结构，在实现知识推理时主要是定性地解决问题，且利用语义资源解决农业知识的应用相对不足，所以从大量繁琐数据中提取有用的农业知识、有效结合定量与定性，使推理更具可解释性的研究意义重大。本文主要工作是根据水稻病虫害数据，提出一种面向该领域的知识图谱构建方法，同时以图谱数据为支撑，提出系列基于图的水稻病虫害检索算法；然后将专家置信度确定性因子 (Certainty factor，CF) 融合到知识图谱中，解决图谱难以定量的问题，且增加图谱的可解释性，构建基于 CF 和知识图谱相结合的知识推理方法，并且通过自主开发的智能诊断系统，结合实际案例对文中方法进行分析，实现水稻领域病虫害的智能检索、预警及诊断。

1 水稻病虫害知识图谱构建

1.1 水稻病虫害知识图谱建模

水稻病虫害知识图谱建模是建立水稻病虫害的数据模型，即采取特定形式表示领域知识，并构建包含概念、属性及概念间关系的本体模型对水稻病虫害知识进行描述[23]。建模途径通常包括自顶向下和自底向上 2 种。本研究将采用 2 种途径相结合的方式，通过构建水稻病虫害领域本体模型[24]，映射到知识图谱的模式层，将从不同数据源中获取到的实体、关系、属性等知识进行融合，形成水稻病虫害知识图谱。

1.1.1 *水稻病虫害本体建模* 建模过程分为明确目标领域本体及任务、模型复用、罗列本体中涉及到的领域元素、明确分类体系、定义相关属性与关系及定义约束条件 6 个主要步骤。在构建相关本体层类别概念过程中，利用网络本体语言建立知识模型，根据国家农业科学数据共享中心数据库对病虫害进行分类，结合水稻专家的指导，得到本体层类别以及概念之间的关系，如图 1 所示。

确定水稻病虫害领域类的集合。实际的概念一般用类形容，如发病阶段、危害症状、节气等称为类。用三元组表示为 (危害部位，rdfs:hasclass，Owl:thing)。

确定水稻病虫害领域内概念的关系集合。本体概念间关系分为子类间关系、实例及类间关系和参照关系 3 种。其中 subClassOf (病斑颜色，病害症状) 表示病斑颜色是病害症状的子类；type(暗绿色，病斑颜色) 表示暗绿色属于病斑颜色类。

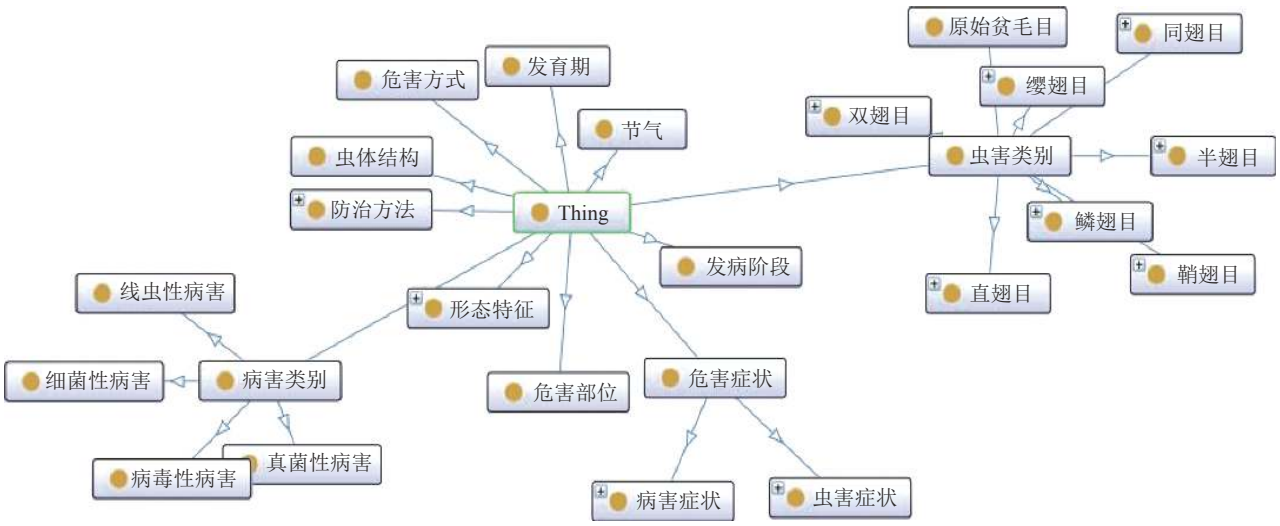


图 1　水稻病虫害本体构建图

Fig. 1　The ontogeny construction figure of rice diseases and pests

确定水稻病虫害领域的属性集合。本体概念中包括对象属性及数据属性 2 种。用三元组表示为(花足秆，纹枯病，Owl:dataProperty)。其中：花足秆是纹枯病的数据属性，即花足秆是纹枯病的别名；水稻病虫害本体概念间存在约束关系的特殊属性。三元组 (病害类别，病害症状，Owl:objectProperty) 表示病害类别和病害症状之间的关系。

确定水稻病虫害领域的实例集合。从语义角度分析实例代表对象，如每个具体病虫害都是病虫害类的实例，用三元组 (纹枯病，rdfs:type，Owl:individual) 表示。

1.1.2　水稻病虫害知识表示　水稻病虫害知识图谱的结构关系分为 2 大类：一类为概念层级关系 (GM_R)、另一类为实体关系 (GE_R) [25]。概念层级关系图 GM_R=<CM_R, RM_R>，其中 CM_R 代表图中出现的概念节点，RM_R 则代表被多条边连着的概念间的关系边。实体关系图 GE_R=<EE_R, RE_R>，其中 EE_R 代表图中出现的实体节点，RE_R 代表被多条边连着的实体间的关系边。以水稻病害为例，展示知识图谱模式层与数据层之间的相互对应关系，如图 2

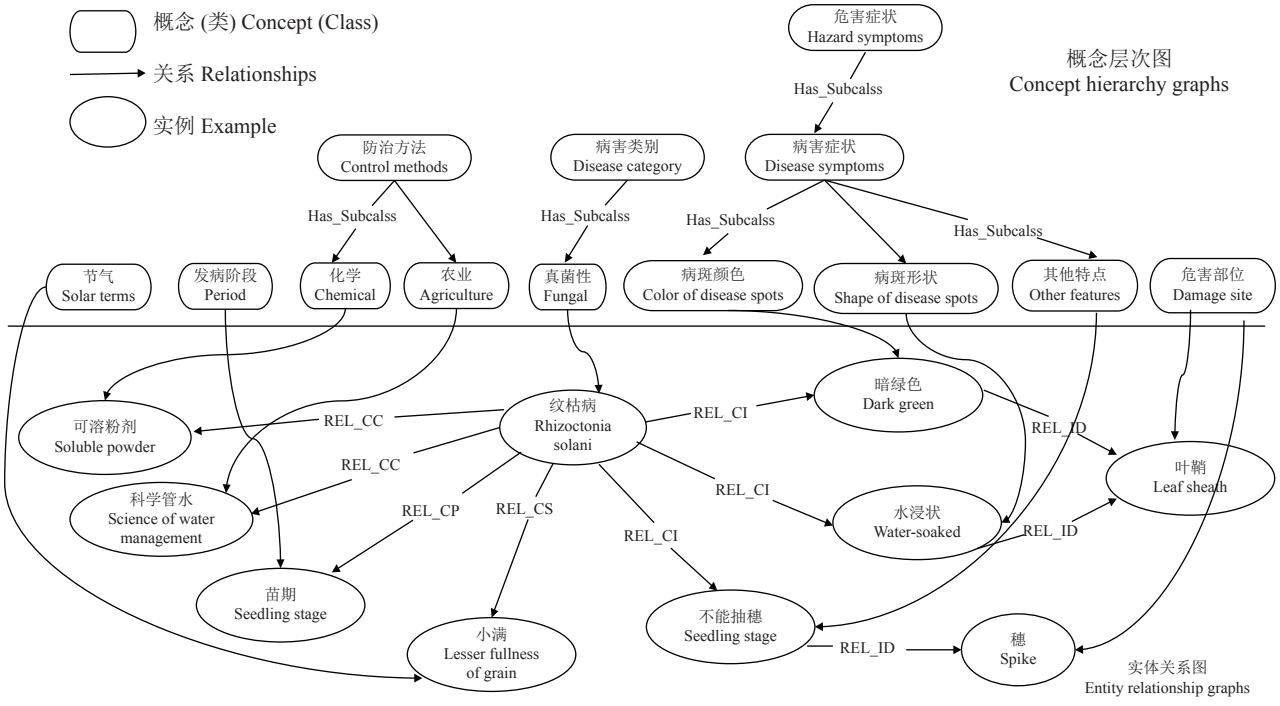


图 2　知识图谱概念层次图及实体关系图

Fig. 2　Conceptual hierarchy graph and entity relationship graph of knowledge graph

所示。其中水稻病虫害领域知识中的各种概念、实体以及连接它们的关系由节点、边分别表示，即防治方法、发病阶段、节气等概念节点和科学管水、苗期、立夏等实体节点，可并称为节点集合；Has_Subclass(概念间) 关系和 REL_PH(虫害实体和危害方式) 关系、REL_CI(病害实体与危害症状实体) 关系等，并称为边集合。

1.1.3 水稻病虫害本体与知识图谱映射匹配机制 构建水稻病虫害知识图谱的首要工作是对该领域本体与知识图谱间匹配机制的确认。本体实质是使用层次化抽象方法进行关系和实体的表示[26]。本体概念层级结构被当成树，其概念、实例及关系通过树形节点和树间的连线来体现。而知识图谱本质被认为是一张庞大的语义知识网络[27]，其概念层级关系图被比作树，树的节点表示其概念节点，实体节点和实体间关系就是图谱中实体关系图中的节点及连线。因此，可根据树与树、树与图之间的映射表示本体和知识图谱的本体映射匹配模式。

把水稻病虫害本体中防治方法、危害症状、节气等概念当作树的节点，而知识图谱中，它们也对应地成为知识图谱概念层级中树的相关节点，同时，纹枯病、水浸状等实例作为知识图谱中实体关系图的节点，具体如图 3 所示。

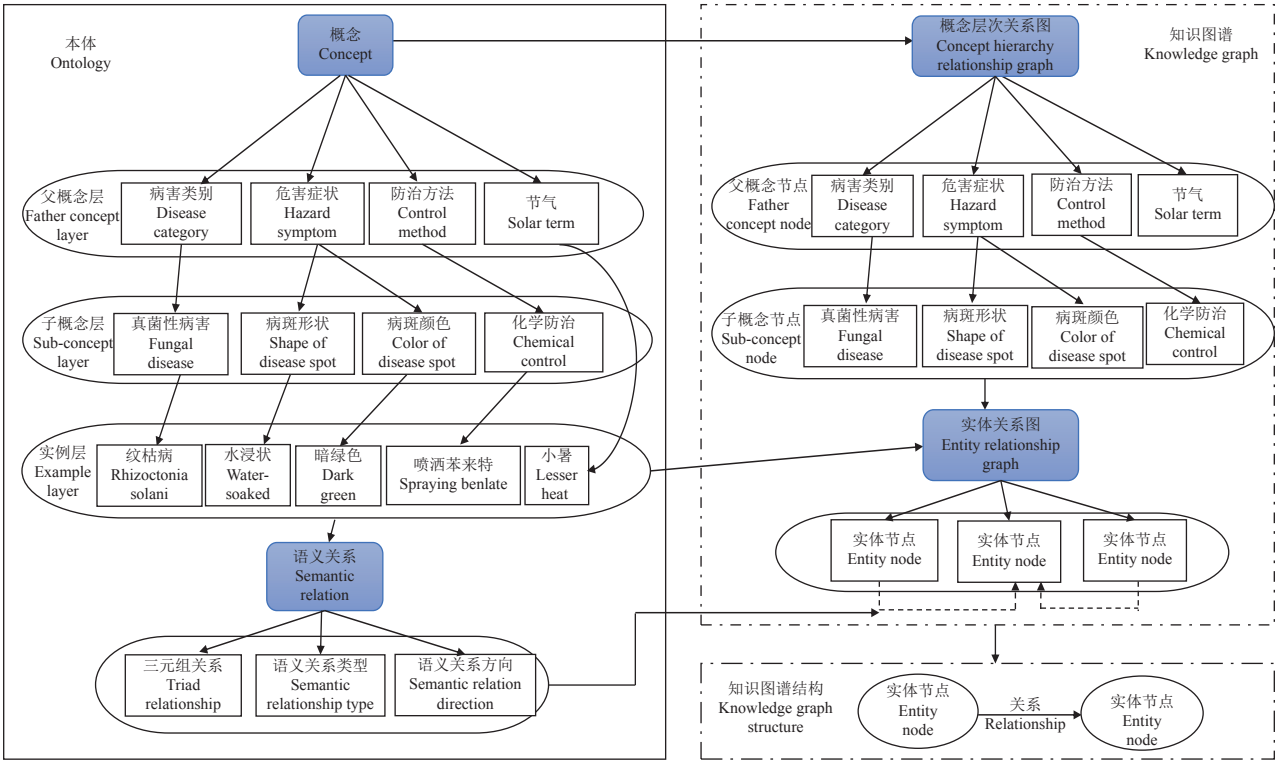


图 3 本体与知识图谱映射图

Fig. 3 The mapping picture from ontology to knowledge graph

1.2 水稻病虫害知识获取

1.2.1 水稻病虫害数据获取与预处理 目前，水稻领域暂时没有公开数据库可作为试验语料，本研究首先利用 python 爬取网络上水稻病虫害文献 (包括南方农业、农家参谋、农民致富之友、乡村科技、植物医生等) 资料，获取到 576 篇水稻相关文本，共 4140 个语句。然后通过正则表达式、规范字符格式等一系列数据预处理操作，删除非文本数据，获得规范化的水稻病虫害语料库。

一般分词都被当作是命名实体识别的基础，但是分词会产生各类不同的错误。例如，水稻病害实体“叶鞘腐败病”分词后为叶鞘/腐败/病。由于错误分词的情况会导致实体的特征表示出错，基于字符的实体识别可减少此类问题的出现。研究中利用字向量作为模型的最初输入，使用预训练的方式，以字为单位切割，进而得到实体的特征表示。

1.2.2 水稻病虫害知识抽取模型框架 本研究采用 Bi-LSTM-CRF 模型[28] 进行水稻病虫害实体的抽取，其框架包括 3 个部分，如图 4 所示。

表示层：水稻文本数据需进行文本向量化，将相应字符映射为一定维度的实数向量，才能被计算机处理。CBOW 模型[29] 是依据当前字的前后各

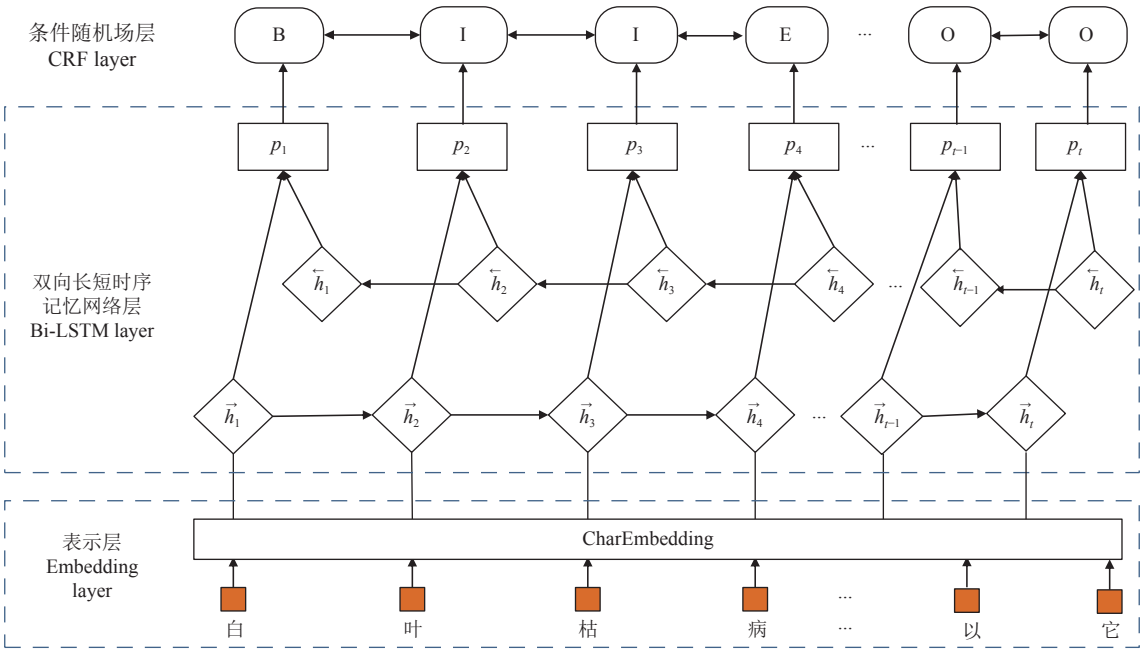


图 4　Bi-LSTM-CRF 模型

Fig. 4　Bi-LSTM-CRF model

2 个字预测当前字出现的概率。例如水稻病害实体“菌核秆腐病”，通过“秆”字的前后“菌”“核”“腐”“病”4 个字，来预测“秆”这个位置出现各个字的概率，其中给定的“秆”字出现的概率最大。

Bi-LSTM 层：Bi-LSTM 由 LSTM 演化而来，LSTM 是缓解循环神经网络在面对长序列会出现梯度消失的问题提出的[30]。LSTM 是包括遗忘门、输出门和输入门的网络模型，各门的结构如下：

$$f_t = \sigma\left[\boldsymbol{w}_f(h_{t-1}, x_t) + \boldsymbol{b}_f\right], \tag{1}$$

$$i_t = \sigma[\boldsymbol{w}_i(h_{t-1}, x_t) + \boldsymbol{b}_i], \tag{2}$$

$$\tilde{C}_t = \tanh[\boldsymbol{w}_c(h_{t-1}, x_t) + \boldsymbol{b}_c], \tag{3}$$

$$C_t = f_t \otimes C_{t-1} + i_t \otimes \tilde{C}_t, \tag{4}$$

$$o_t = \sigma[\boldsymbol{w}_o(h_{t-1}, x_t) + \boldsymbol{b}_o], \tag{5}$$

$$h_t = o_t \otimes \tanh C_t \tag{6}$$

其中：f_t、i_t、o_t分别为遗忘门、输入门、输出门；$\boldsymbol{w}$代表权重矩阵；$\boldsymbol{b}$代表偏置向量；σ代表 sigmoid 函数；C_t代表当前细胞的状态，$\otimes$代表矩阵按位相乘。

单向 LSTM 网络在进行训练时只考虑到句子的时序信息而忽略了上下文之间的关系，因此，多句对话时，往往不能取得很好的效果。例如：水稻病害实体“白叶枯病”，LSTM 只能访问“叶”的前一个字“白”的特征而不能预测下一个字“枯”的出现。Bi-LSTM 模型将前向的 LSTM 和后向的 LSTM 结合，可以充分利用序列的上下文信息，具有能够捕获前后信息特征的作用。

CRF 层：CRF 是一种序列建模算法[31]，它综合了隐马尔可夫模型和最大熵模型的优点。它由既定的观察序列进一步推测出对应的状态序列，可以利用相邻前后的标签关系获取当前最优的标记。

1.2.3　*水稻病虫害知识抽取过程与结果分析*　本研究采用 Bi-LSTM-CRF 模型进行水稻病虫害命名实体识别，要识别出病害名称、虫害名称和危害症状 3 种不同的命名实体。语料中训练集和测试集按 7∶3 划分。主要的流程如下：

首先，利用 CBOW 模型，通过对语料进行无监督训练，对不同维度字向量进行对比，最终得到 100 维度时模型性能最好。通过预训练方式，得到水稻病虫害文本 100 维度的字向量特征，应用在水稻病虫害领域的命名实体识别。其次，字嵌入层的向量 $\boldsymbol{x}$ 将作为 t 时刻 Bi-LSTM 层的输入，通过正向 LSTM 输出特征序列和反向输出序列，获得隐藏层拼接的向量，通过 tanh 激活函数的加权求得最终的输出。最后，把 Bi-LSTM 的输出作为 CRF 层的输入，采用状态转换矩阵预测当前标签，通过利用 Softmax 函数，得到最终序列的条件概率。使用 Viterbi 算法[32] 将得分最高的序列作为模型最终的标注结果。通过多次重复试验得到模型的参数，如表 1 所示。

表 1 Bi-LSTM-CRF 模型参数设置
Table 1 Parameter settings of Bi-LSTM-CRF model

参数 Parameter	参数值 Parameter value
字向量维度 Word vector dimension	100
隐藏层维数 Hidden layer dimension	128
学习率 Learning rate	0.001
批尺寸 Batch_size	32
学习衰减率 Dropout rate	0.75
迭代次数 Epoch	50

测评常用 *P*(准确率)、*R*(召回率) 和 *F*1 值来评价试验结果。其中，*P* 为模型正确识别水稻病虫害实体数与识别出水稻病虫害实体总数的比值，*R* 为系统正确识别水稻病虫害实体数与数据集中存在的水稻病虫害实体总数的比值，*F*1 值的计算公式为 $F1 = 2PR/(P+R)$。通过试验得出 3 类实体识别的准确率、召回率、*F*1 值，如表 2 所示。

表 2 命名实体试验结果
Table 2 Named entity experimental results

实体类型 Entity type	准确率 Accuracy rate	召回率 Recall rate	*F*1值 *F*1 value
病害名称 Disease name	0.92	0.88	0.90
虫害名称 Pest name	0.90	0.87	0.88
危害症状 Hazard symptom	0.87	0.84	0.85

1.3 水稻病虫害知识图谱存储

本研究运用当前流行的图数据库 Neo4j[33] 进行水稻病虫害领域知识的存储。根据模式层的防治方法 (Control methods)、发病阶段 (Period)、危害部位 (Damage site) 等 11 大类知识进行水稻病虫害数据分析，将获取的知识以三元组的形式进行表达，即相同类型实例放在一个表中，表的每列代表该类实体的具体值，每行存储该类实体的实例。运用机器学习算法提取病虫害名称实体、危害症状实体及它们之间的关系数据，并进行结构化处理，同时定义好对应的标签。在此基础上，再请该领域专家进行实体的补充和修正，由专家辅助建立关系。最终将全部内容建立对应的实体及三元组关系表，表 3 为部分三元组示例。通过搭建近义词映射表实现水稻病虫害领域异构数据源的实体对齐正确性，并在完成对齐后进行去重操作。水稻病虫害三元组数量较多，需要进行批量入库。先将实体表、三元组关系表进行整合，采用“LOAD”的方式，将转化后的 CSV 文件写入到 Neo4j 中构建知识图谱。

表 3 水稻病害实体关系部分三元组示例
Table 3 Examples of triples in the entity relationship of rice diseases

实体 Entity	关系 Relationship	实体 Entity
水稻纹枯病 *Rhizoctonia solani*	REL_CI	椭圆形 Oval shape
水稻纹枯病 *R. solani*	REL_CI	暗绿色 Dark green
水稻纹枯病 *R. solani*	REL_CP	苗期 Seedling stage
水稻纹枯病 *R. solani*	REL_CC	及时拔除病株 Promptly pull up diseased plants
椭圆形 Oval shape	REL_ID	叶片 Leaf blade

中心节点表示病害类别标签下的水稻纹枯病实体；外侧节点表示水稻纹枯病的 27 个病斑颜色和病斑形状，Rel_CI 表示病害类别和危害症状的关系，relciName 和 CF 表示关系的属性值，如图 5 所示。

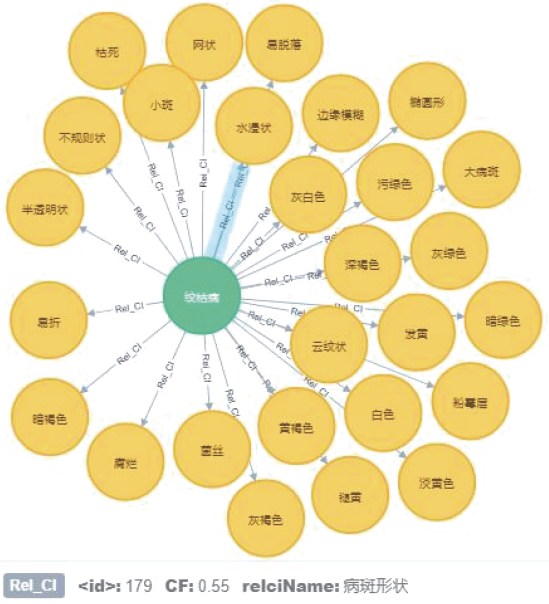


图 5 知识图谱可视化
Fig. 5 Knowledge graph visualization

水稻病虫害所有数据整合到图数据库后的具体信息，如表 4、表 5 所示。

表 4　知识图谱实体类型及属性

Table 4　Knowledge graph entity types and properties

实体类型 Entity type	实体属性 Entity property	举例 Example
Category	Disease name	水稻纹枯病、水稻白叶枯病 *Rhizoctonia solani*, *Xanthomonas campestris*
Control methods	Control methods name	浅水勤灌、无色防虫网 Shallow water and diligent irrigation, colorless insect-proof net
Period	Period name	苗期、分蘖期 Seedling stage, tillering stage
Damage site	Damage site name	根、茎 Root, stem
Ill spot	Ill spot name	暗绿色、水浸状 Dark green, water-soaked
Pest category	Pest name	二化螟、稻纵卷叶螟 *Chilo suppressalis*, *Cnaphalocrocis medinalis*
Insect structure	Insect structure name	头部、胸部 Head, thorax
Morphological characteristic	Morphological characteristic name	灰白色、鱼鳞状 Grayish white, fish scale shape
Developmental stage	Developmental stage name	幼虫、蛹 Larvae, chrysalis
Hazard pattern	Hazard pattern name	刺吸、食叶 Prickly suction, leaf-eating
Solar term	Solar term name	春分、小暑 The beginning of spring, lesser heat
总计/个 Total		1972

表 5　知识图谱实体关系类型及属性

Table 5　Knowledge graph entity relationship types and properties

实体关系类型 Entity relationship type	实体关系属性 Entity relationship property	举例 Example
REL_CC	Relcc name	水稻纹枯病防治方法是科学灌溉 *Rhizoctonia solani* was prevented and controlled by scientific irrigation
REL_CP	Relcp name	水稻纹枯病发病阶段是苗期 The onset stage of *R. solani* is seedling stage
REL_CI	relic name、CF	水稻纹枯病的症状为有暗绿色、水浸状病斑 The symptoms of *R. solani* include dark green, water-soaked spots
REL_ID	Relid name	暗绿色出现的部位是叶鞘 The part that appears dark green is the leaf sheath
REL_PH	Relph name	二化螟危害方式是钻蛀 The damage method of *Chilo suppressalis* is borer
REL_PI	Relpi name	二化螟症状是有枯黄色斑点 The *C. suppressalis* symptoms include withered yellow spot
REL_PM	Relpm name、CF	二化螟形态特征为有暗褐色纵线 The *C. suppressalis* morphological characteristics include dark brown longitudinal lines
REL_MI	Relmi name	头部表现出铜绿色、近三角形 The head exhibits copper-green, sub-triangular shape
REL_MD	Relmd name	暗褐色纵线出现的发育期是幼虫期 The developmental stage that appears dark brown longitudinal lines is larvae
REL_CS	Relcs name	水稻纹枯病出现的节气为立夏 The solar terms for the emergence of *R. solani* is the beginning of summer
总计/个 Total		5 226

2　水稻病虫害知识图谱推理

2.1　基于图挖掘算法的水稻病虫害信息检索

实际水稻病虫害检索应用时，通过图挖掘算法分析图谱中的知识数据，对其中已有的知识深层次分析后得到潜藏在数据内部的新知识[34]，以实现用户不同的检索需求。其一，用户要获取实体节点间的离散距离，即探索出数据间最佳路径，提出基于最短路径的水稻病虫害检索算法；其二，用户想明确哪些实体隶属一类的问题，提出基于连通组件的

水稻病虫害检索算法；其三，用户要找到与已知实体节点最为相似的其余节点，提出基于杰卡德相似度的水稻病虫害检索算法。系统会根据用户的输入自动匹配最佳算法，为用户高效检索出最佳答案，图 6 为实现数据检索算法的详细流程。

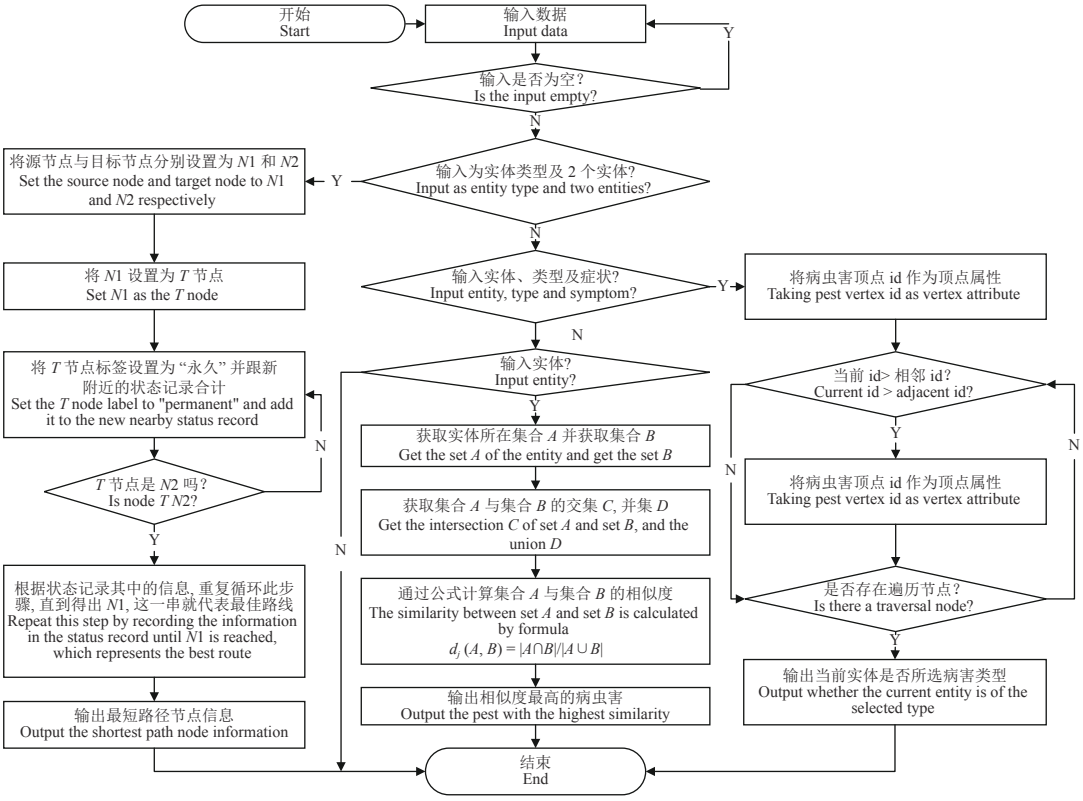


图 6 检索流程图

Fig. 6 Search flow chart

文中提出的系列水稻病虫害检索算法结合图谱的优势，不但能帮助用户更好地理解知识图谱结构，当有新的子图需要频繁更新时，也可以快速显示新节点信息。同时也提升了系统检索效率，对系统应用具有重要作用。

2.2 基于节气的水稻病虫害预警

节气能反应季节不同变化，常用作指导农业耕种。它具有很强的区域性，不同农作物、不同地域具备不同的气象信息。其中针对节气的时间概念是指节气前后的天数，即节气所处阳历月份的上半月或下半月。本研究主要针对吉林省长春市各县单季水稻的相关农事活动。节气作为水稻预警的重要影响因素，通过病虫害不同的发病阶段，得出不同发病阶段的月份时间范围，再根据文献 [35] 中节气和月份对应的时间段，利用月份映射出发病阶段和节气间的对应关系，发现该节气可能出现的病虫害，进而增加目标病虫害正确识别的概率，进一步实现预警。

2.3 基于确定性因子模型和知识图谱结合的水稻病虫害诊断

2.3.1 确定性因子模型 基于确定性因子模型的不确定推理是对不确定知识应用与处理的一种方法，它的典型代表是 MYCIN 专家系统。水稻病虫害诊断中，每个症状对于病害或虫害为真的程度，专家都赋予其 CF 值，其表示形式为：

$$\text{IF}\quad A\quad \text{THEN}\quad B\quad \text{CF}(B|A)$$

其中 A、B 分别为前提条件和结论。

确定性因子的定义[36] 如下：

$$\text{CF}(B|A)=\begin{cases}\dfrac{P(B|A)-P(B)}{1-P(B)}, & P(B|A)>P(B)\\ 0, & P(B|A)=P(B)\\ \dfrac{P(B|A)-P(B)}{P(B)}, & P(B|A)<P(B)\end{cases}。\quad(7)$$

在水稻病虫害诊断中，需要知道某种证据在某个结论中出现的概率，所以必须把确定性因子转化成概率[37]。由上面公式 (7) 得：

$$P(B|A)=\begin{cases}\text{CF}(B|A)(1-P(B))+P(B), & \text{CF}(B|A)\geqslant 0\\ (\text{CF}(B|A)+1)P(B), & \text{CF}(B|A)<0\end{cases}。\quad(8)$$

要根据 $P(B)$，即 B 的先验概率，才能得出最终概率。$P(B)$ 可以通过专家经验或者文献给出，还可

直接视为无知，也就是 $P(B)$=0.5。

2.3.2　基于确定性因子模型和知识图谱结合的知识推理方法　通过知识间关联关系，根据 CF 的定量特点及属性图模型的特性，把专家赋予的 CF 值作为 REL_CI 关系的特殊属性引入知识图谱，且病害实体与危害症状实体间只对应一个 CF 值。但经常会出现一种病害或虫害共同拥有多个症状的情况。对于症状的不确定性而言，当有多条规则支持结论时，那么结论的确定性因子计算公式为：

$$\mathrm{CF}(B,A)=\mathrm{CF}(B,A_1)+\mathrm{CF}(B,A_2)-\mathrm{CF}(B,A_1)\times\mathrm{CF}(B,A_2)。\tag{9}$$

在利用知识图谱进行水稻病虫害诊断过程中，不会存在证据为假的情况，即$\mathrm{CF}(B|A)<0$不存在，当出现多个症状和病虫害有关联时，选择公式 (9) 进行并行 CF 值计算。在推理诊断过程中，当选择某个或某几个症状时，无法判断知识图谱中相互关联的该症状究竟是属于哪种病害或者虫害，采用 CF 和知识图谱结合的方法解决这个不确定性问题，该算法的实现流程如下：

输入: ListSymptom = {*M1,M2,M3*} 3 种症状；

输出: *M1,M2,M3* 所有组合症状中，患病概率由高到低的前 3 个疾病；

步骤 1: 对 ListSymptom {*M1,M2,M3*}中症状进行组合，得到列表 L；

步骤 2: 对组合数据进行遍历，列表 L 索引值 i=0；

步骤 3: while i 小于列表 L 的长度；

步骤 4: 根据集合 $L[i]$ 的症状与该症状对应的关系 R，在数据库中查找是否存在相关病虫害的集合 D；

步骤 5: if D 不为空 then；

步骤 6: 在数据库中查找出对应病虫害的可信度 CF_n；

步骤 7: 由公式 (8) 算出发病的概率；

步骤 8: 将症状、病虫害与概率记录到字典 RecDict 中；

步骤 9: 根据概率值对结果列表进行由高到低的排序；

步骤 10：取出结果列表中前 3 组数据，写入结果字典 ResDict 中；

步骤 11: return ResDict。

2.3.3　水稻病虫害诊断　由于不需要向用户询问相关信息，只需要用户对所观察到的水稻病虫害症状做出尽量具体且多的选择。所以用户界面设计十分简单，但它要求用户能够对观察到的水稻病虫害的表现症状进行正确的描述。

使用基于 CF 和知识图谱结合的知识推理方法进行诊断，表 6 为 1 个诊断实例。

表 6　诊断实例表

Table 6　Table of examples of diagnoses

ID	症状名称 Symptom name	症状名称代码 Symptom name tag	病害名称 Disease name	病害名称代码 Disease name tag	确定性因子 Certainty factor (CF)
1	暗绿色 Dark green	H1	水稻纹枯病 *Rhizoctonia solani*	M1	0.55
2	水浸状 Water-soaked	H2	水稻纹枯病 *R. solani*	M1	0.55
3	暗绿色 Dark green	H1	细菌性条斑病 *Xanthomonas oryzae*	M2	0.36
4	水浸状 Water-soaked	H2	细菌性条斑病 *X. oryzae*	M2	0.36
5	卷曲 Curl	H3	细菌性条斑病 *X. oryzae*	M2	0.36
6	暗绿色 Curl	H1	水稻白叶枯病 *X. campestris*	M3	0.47
7	水浸状 Water-soaked	H2	水稻白叶枯病 *X. campestris*	M3	0.47
8	卷曲 Curl	H3	水稻白叶枯病 *X. campestris*	M3	0.47

用户对出现的症状进行选择，利用 CF 模型与知识图谱结合的算法中 $P(B)$设为无知，即$P(B)$=0.5。具体计算步骤如下：

第 1 步：经过推理得出水稻纹枯病、细菌性条斑病和水稻白叶枯病的 CF 值分别为 0.7975、0.7379 和 0.8511。

水稻纹枯病：$\mathrm{CF}(M1|H)$=0.55+0.55−0.55×0.55=0.7975；

细菌性条斑病：$\mathrm{CF}(M2|H1H2)$=0.36+0.36−0.36×0.36=0.5904，

$\mathrm{CF}(M2|H)$=0.5904+0.36−0.5904×0.36≈0.7379；

水稻白叶枯病：$\mathrm{CF}(M3|H1H2)$=0.47+0.47−0.47×0.47=0.7191，

$\mathrm{CF}(M3|H)$=0.7191+0.47−0.7191×0.47≈0.8511；

第 2 步：由于用户要判断出哪一种病害出现的概率更高，所以利用确定性因子模型，求出每种病害出现的概率。通过计算求得 3 种病的概率分别为 0.8988、0.8690 和 0.9256。

水稻纹枯病：$P(M1|H)=\mathrm{CF}(M1|H)(1-P(B))+P(B)=0.7975\times(1-0.5)+0.5\approx0.8988$；

细菌性条斑病：$P(M2|H)=\mathrm{CF}(M2|H)(1-P(B))+P(B)=0.7379\times(1-0.5)+0.5\approx0.8690$；

白叶枯病：$P(M3|H)=\mathrm{CF}(\mathrm{M3}|\mathrm{H})(1-\mathrm{P}(\mathrm{B}))+\mathrm{P}(\mathrm{B})=0.8511\times(1-0.5)+0.5\approx0.9256$。

病害出现的概率越大，说明水稻发生该病害的可能性越大，因此最可能发生的病害为水稻白叶枯病。

2.4 算法性能评估

算法的性能根据诊断结果的正确率 (Accuracy) 进行判断。利用诊断系统对长春市某县 4—9 月份 8 种常见病害进行实际测试，由水稻领域专家最终核实 80 例样本，并表现在图 7 的 8×8 混淆矩阵中。

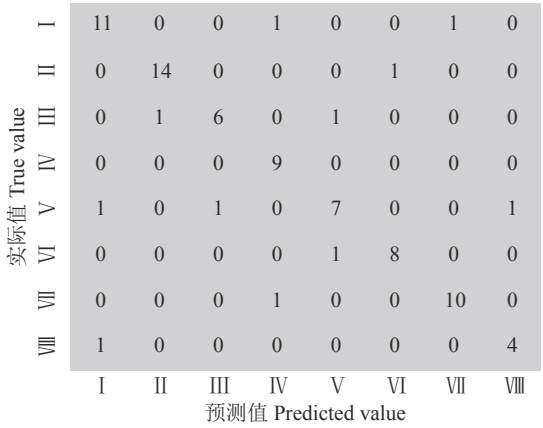
通过真正例 (True positive，TP)、假正例 (False positive，FP)、假负例 (False negative，FN)、真负例 (True negative，TN)4 个基础指标得到评估诊断正确率的计算公式：

$$\text{正确率}=\frac{\mathrm{TP}+\mathrm{TN}}{\mathrm{TP}+\mathrm{TN}+\mathrm{FP}+\mathrm{FN}}。\tag{10}$$

由式 (10) 计算 80 例样本的正确率，并依据混淆矩阵进行验证：

$$\text{正确率}=\frac{11+14+6+9+7+8+10+4}{80}=86.25\%。$$

综上，本研究提出的基于知识图谱和确定性因子模型结合的诊断方法的正确率为 86.25%。



Ⅰ：水稻纹枯病 *Rhizoctonia solani*；Ⅱ：稻瘟病 *Pyricularia oryzae*；Ⅲ：水稻霉霜病 *Sclerophthora macrospora*；Ⅳ：稻曲病 *Ustilaginoidea virens*；Ⅴ：烂秧病 *Fusarium graminearum*；Ⅵ：水稻白叶枯病 *Xanthomonas campestris*；Ⅶ：水稻恶苗病 *Fusarium moniliforme*；Ⅷ：水稻窄条斑病 *Cercospora oryzae*

图 7 8 种常见病害的系统诊断性能评价结果

Fig. 7 Evaluation results of system diagnostic performance of eight common diseases

3 水稻病虫害知识图谱应用

3.1 水稻病虫害智能诊断整体架构

面对水稻病虫害海量知识冗余、分散、多源的特点，建立水稻病虫害范畴的知识图谱，用户不仅可以检索水稻病虫害的相关知识，而且可以通过症状诊断水稻所患病虫害，从而有效拓展用户的知识面，提升水稻生产水平与质量。本研究开发了包括 80 多种病虫害的智能诊断系统，其整体架构由水稻病虫害知识图谱和智能推理 2 大部分组成，如图 8 所示。本系统的 PC 端访问地址为：http://d236424e.

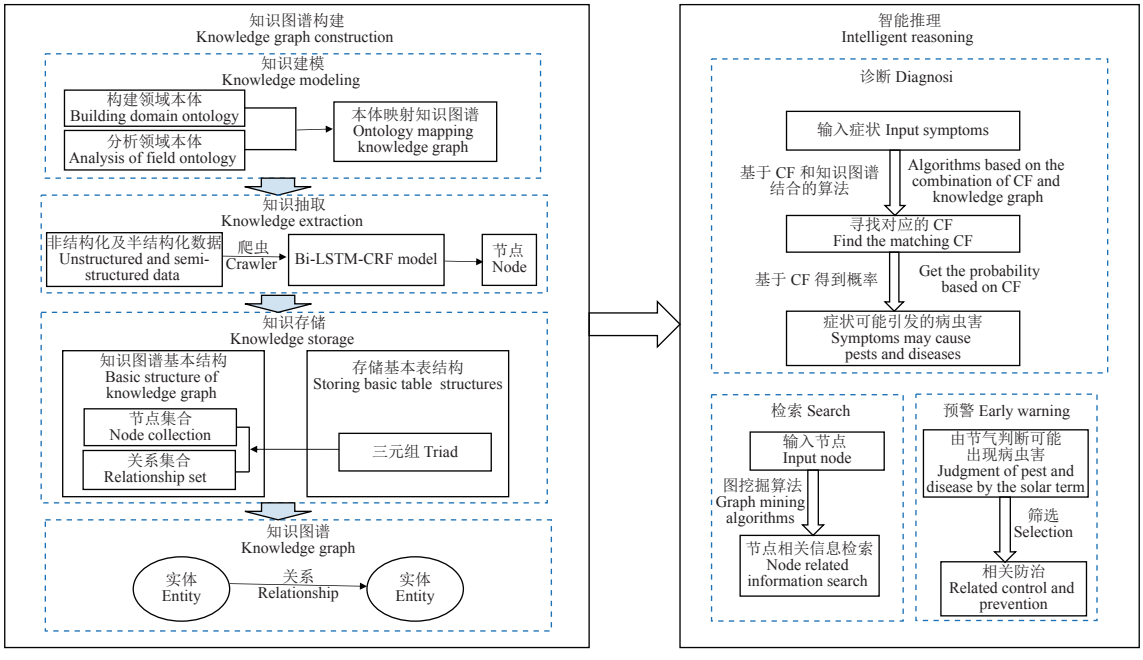


图 8 智能诊断系统整体架构

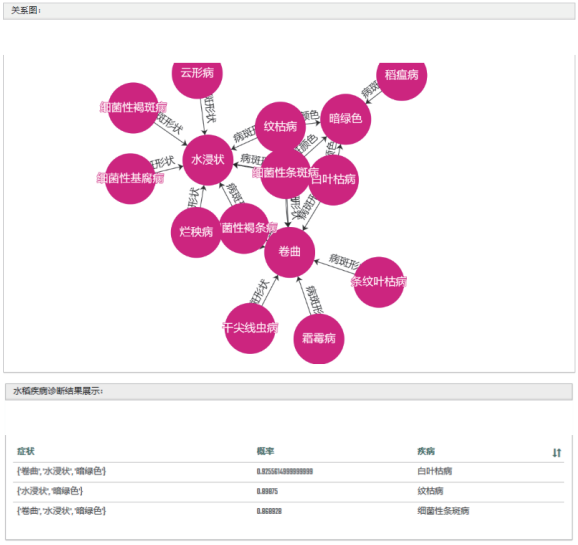
Fig. 8 Overall architecture of an intelligent diagnostic system

nat3.z100.vip。

3.2　水稻病虫害知识图谱案例分析

为确保用户更深层次地了解水稻病虫害间的潜在关系，以便对其病虫害有更进一步的诊断和预测，开发了以 ECharts 可视化为基础，包含知识展示、检索、诊断及预警 4 个功能模块的智能诊断系统，如图 9a 所示。系统在吉林省某农村进行了应用。知识展示模块的目的是让用户对知识图谱有更深刻直观的理解，也为后续的推理提供有力的数据支撑。当用户选择查询推理时，通过查询的方式实现信息的检索，例如查询“腐烂”症状，可得到与腐烂有直接关联的烂秧病、纹枯病等病害信息。图 9b 以病害为例，用户选择“水浸状”、“卷曲”、“暗绿色”3 个症状，不仅可以得到与其关联的所有病害实体节点，还可根据 CF 模型与知识图谱结合的算法，诊断出在 3 种症状发生的情况下，最可能出现的病害是水稻白叶枯病，并链接判断后的目标病害图片，为用户检查提供依据。预警模块根据节气对用户进行农事指导，通过建立病虫害与节气的关联关系，把节气作为实体节点引入知识图谱中，增加了图谱的完整性。例如 5 月下旬长春某地区，小满节气水稻可能会出现恶苗病、霜霉病、稻瘟病、胡麻斑病、纹枯病等病害，系统预警模块则展示对应病害图片，提高用户正确识别目标病害的概率。

a: 水稻病虫害诊断首页
a: The home of rice disease and pest diagnosis



b: 水稻病虫害诊断推理可视化界面
b: The visual interface of rice disease and pest diagnosis and reasoning

图 9　系统界面图

Fig. 9　System interface figure

4　结论

本研究首先提出水稻病虫害知识图谱构建方法，解决了数据多样及孤岛化问题，面向特定的水稻病虫害领域知识图谱有助于充分发挥该领域数据的实用价值。随后，提出系列基于图的水稻病虫害检索算法，增加了系统查询的多样性。然后，又提出确定性因子模型和知识图谱结合的知识推理方法，实现定量和定性的结合，正确率达 86.25%，增加了诊断的准确性和可解释性。最后，系统以 ECharts 为可视化展示工具，实现了水稻病虫害数据在特定场景下的应用，有利于用户挖掘隐藏在图谱中的实体及关系；通过引入节气实体，增加图谱的完整性，提升系统病虫害诊断的正确率。本研究实现了基于 WEB 的吉林省长春市各县水稻病虫害智能诊断系统，具有用户界面友好、操作方式简单、便捷、准确的优点，为植物病虫害智能诊断提供了一种高效的新途径。后期将扩大农作物应用区域，探究更多农作物病虫害特征，构建更加丰富的农作物病虫害知识图谱，并在实际应用中不断完善和补充现有方法，以期用户能在不同地区访问系统，得到符合该地区节气的不同识别结果，提高诊断准确率。

参考文献：

[1] 刘万才, 陆明红, 黄冲, 等. 水稻重大病虫害跨境跨区域监测预警体系的构建与应用[J]. 植物保护, 2020, 46(1): 87-92.

[2] 刘明辉, 沈佐锐, 高灵旺, 等. 基于 WebGIS 的农业病虫害预测预报专家系统[J]. 农业机械学报, 2009, 40(7): 180-186.

[3] HU X F, CHENG C, LUO F, et al. Effects of different fertilization practices on the incidence of rice pests and diseases: A three-year case study in Shanghai, in subtropical southeastern China[J]. Field Crops Research, 2016, 196: 33-50.

[4] 许童羽, 郭忠辉, 于丰华, 等. 采用 GA-ELM 的寒地水稻缺氮量诊断方法[J]. 农业工程学报, 2020, 36(2): 209-218.

[5] 孙娟, 蔡银杰, 冯成玉, 等. 无人机施药防治水稻病虫害参数组合初选[J]. 中国植保导刊, 2018, 38(12): 72-73.

[6] 王艺, 王英, 原野, 等. 基于语义本体的柑橘肥水管理决策支持系统[J]. 农业工程学报, 2014, 30(9): 93-101.

[7] 姚青, 张超, 王正, 等. 分布式移动农业病虫害图像采集与诊断系统设计与试验[J]. 农业工程学报, 2017, 33(S1): 184-191.

[8] 戴建国, 赖军臣. 基于图像规则与 Android 手机的棉花病虫害诊断系统[J]. 农业机械学报, 2015, 46(1): 35-44.
[9] 杨林楠, 郜鲁涛, 林尔升, 等. 基于 Android 系统手机的甜玉米病虫害智能诊断系统[J]. 农业工程学报, 2012, 28(18): 163-168.
[10] 马浚诚, 温皓杰, 李鑫星, 等. 基于图像处理的温室黄瓜霜霉病诊断系统[J]. 农业机械学报, 2017, 48(2): 195-202.
[11] 温皓杰, 张领先, 傅泽田, 等. 基于 Web 的黄瓜病害诊断系统设计[J]. 农业机械学报, 2010, 41(12): 178-182.
[12] SARMA S K, SINGH K R, ABHIJEET S. An expert system for diagnosis of disease in rice plant[J]. International Journal of Artificial Intelligence, 2010, 1(2): 26-31.
[13] HONGGOWIBOWO A S. A web-based rice plant expert system using rule-based reasoning[J]. Telkomnika, 2009, 7(3): 187-194.
[14] SHARMA R, CHANDERMOHAN, SINGH H, et al. Development of an image based expert system for identification of rice diseases and their management[J]. Plant Disease Research, 2012, 27(2): 158-161.
[15] MUHIBUDDIN A, AIRLANGGA P, SULTHONI M M, et al. Implementing backward chaining method in expert system to detect and treat rice, chilli, and corn plant's pests and diseases[J]. Journal of Information Technology and Computer Engineering, 2018, 2(2): 71-75.
[16] 王昊奋, 丁军, 胡芳槐, 等. 大规模企业级知识图谱实践综述[J]. 计算机工程, 2020, 46(7): 1-13.
[17] SU X L, LI J, CUI Y P, et al. Review on the work of agriculture ontology research group[J]. Journal of Integrative Agriculture, 2012, 11(5): 720-730.
[18] 索俊锋, 刘勇. 基于农业本体的语义相似度算法及其在农作物本体中的应用[J]. 农业工程学报, 2016, 32(16): 175-182.
[19] DRURY B, FERNANDES R, MOURA M F, et al. A survey of semantic web technology for agriculture[J]. Information Processing in Agriculture, 2019, 6(4): 487-501.
[20] 夏迎春. 基于知识图谱的农业知识服务系统研究[D]. 合肥: 安徽农业大学, 2018.
[21] 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.
[22] 王娟. 基于案例推理和 KG 的烟草病害防控模型研究[D]. 合肥: 安徽农业大学, 2016.
[23] YANG Y, CHEN K, CHAO L, et al. Research on hierarchy structure generation method of ontology knowledge pan-concept in agriculture[J]. Advances in Robotics & Automation, 2017, 6(3): 1-6.
[24] 余凡. 领域本体构建方法及实证研究[M]. 武汉: 武汉大学出版社, 2015: 75-79.
[25] GUAN N, SONG D, LIAO L. Knowledge graph embedding with concepts[J]. Knowledge-Based Systems, 2019, 164(3): 38-44.
[26] AYDIN S, AYDIN M N. Ontology-based data acquisition model development for agricultural open data platforms and implementation of OWL2MVC tool[J]. Computers and Electronics in Agriculture, 2020, 175(3): 1-9.
[27] 李涓子, 侯磊. 知识图谱研究综述[J]. 山西大学学报(自然科学版), 2017, 40(3): 454-459.
[28] 王莉军, 周越, 桂婕, 等. 基于 BiLSTM-CRF 的中医文言文文献分词模型研究[J]. 计算机应用研究, 2020, 37(11): 3359-3362.
[29] 赵鹏飞, 赵春江, 吴华瑞, 等. 基于注意力机制的农业文本命名实体识别[J]. 农业机械学报, 2021, 52(1): 185-192.
[30] 张善文, 王振, 王祖良. 结合知识图谱与双向长短时记忆网络的小麦条锈病预测[J]. 农业工程学报, 2020, 36(12): 172-178.
[31] 李想, 魏小红, 贾璐, 等. 基于条件随机场的农作物病虫害及农药命名实体识别[J]. 农业机械学报, 2017, 48(S1): 178-185.
[32] KHAN N A, MOHAMMADI M. A Modified viterbi algorithm-based if estimation algorithm for adaptive directional time–frequency distributions[J]. Circuits, Systems, and Signal Processing, 2018, 38(5): 2227-2244.
[33] PRAVEENA R K S, JUSTUS S. Concept relation knowledge visualization with CR logic using Neo4j[J]. International Journal of Recent Technology and Engineering (IJRTE), 2019, 8(4): 8475-8480.
[34] 吴运兵, 杨帆, 赖国华, 等. 知识图谱学习和推理研究进展[J]. 小型微型计算机系统, 2016, 37(9): 2007-2013.
[35] 于合龙, 陈程程, 林楠, 等. 互联网+农业科技服务云平台构建与农业时空推荐算法研究[J]. 吉林农业大学学报, 2019, 41(4): 495-504.
[36] 刘欣. 基于确定性因子理论的肺癌诊断 Web 专家系统的研究与实现[D]. 长春: 吉林大学, 2017.
[37] 靳留乾, 徐扬, 方新, 等. 基于证据理论的不确定性推理方法及其应用[J]. 计算机工程与应用, 2015, 51(10): 6-11.

【责任编辑 李晓卉】