

# 农业文本语义理解技术综述

吴华瑞<sup>1,2</sup> 郭威<sup>2,3</sup> 邓颖<sup>1,3</sup> 王郝日钦<sup>1</sup> 韩笑<sup>4</sup> 黄素芳<sup>5</sup>

(1. 国家农业信息化工程技术研究中心, 北京 100097; 2. 北京市农林科学院信息技术研究中心, 北京 100097;  
3. 农业农村部数字乡村技术重点实验室, 北京 100097; 4. 北京市农林科学院智能装备技术研究中心, 北京 100097;  
5. 沧州市农林科学院, 沧州 061001)

**摘要:** 随着互联网和人工智能技术的发展,农业知识智能化服务逐渐承担起为农业生产管理提供有效技术指导的作用。本文对农业文本语义理解中的关键技术及应用进行综述。首先按照自然语言处理中基于规则、机器学习和深度学习的语义处理方法介绍其在农业领域应用的进展;然后阐述了针对农业知识特性的语义分析方法,涵盖农业文本分析主要过程的储存、表达、计算,包括农业知识图谱的知识抽取、融合、表示、推理,TF-IDF、Word2Vec、BERT等农业文本表示模型与CNN、RNN、Attention等分类模型;阐述了可用于分词、向量化表达等的通用语料库和农业领域常用语料库;从农业智能问答、农业语义检索、农业智能管理决策方面阐述语义理解在农业领域中的应用;最后从农业语料库标准化构建、语义理解模型复杂度、多模态语义处理、多区域多语言语义理解等方面对农业文本的语义理解研究趋势进行了展望。

**关键词:** 农业知识智能化服务;深度学习;自然语言处理;农业文本;语义理解

**中图分类号:** TP183      **文献标识码:** A      **文章编号:** 1000-1298(2022)05-0001-16      **OSID:** 

## Review of Semantic Analysis Techniques of Agricultural Texts

WU Huarui<sup>1,2</sup> GUO Wei<sup>2,3</sup> DENG Ying<sup>1,3</sup> WANG Haoriqin<sup>1</sup> HAN Xiao<sup>4</sup> HUANG Sufang<sup>5</sup>

(1. National Engineering Research Center for Information Technology in Agriculture, Beijing 100097, China  
2. Beijing Research Center for Information Technology in Agriculture,  
Beijing Academy of Agriculture and Forestry Sciences, Beijing 100097, China  
3. Key Laboratory of Digital Village Technology, Ministry of Agriculture and Rural Affairs, Beijing 100097, China  
4. Beijing Research Center of Intelligent Equipment Technology for Agriculture,  
Beijing Academy of Agriculture and Forestry Sciences, Beijing 100097, China  
5. Cangzhou Academy of Agricultural and Forestry Sciences, Cangzhou 061001, China)

**Abstract:** With the development of Internet and artificial intelligence technology, agricultural knowledge intelligent services have gradually assumed the role of providing effective technical guidance for agricultural production management, especially during the epidemic. The key technologies and applications in the semantic understanding of agricultural knowledge service texts were reviewed. Firstly, its progress in agriculture was introduced according to the semantic processing methods based on rules, machine learning and deep learning in natural language processing. Then, the semantic analysis method for the characteristics of agricultural knowledge was introduced, covering the storage, expression and calculation of the main process of agricultural text analysis, including knowledge extraction, knowledge fusion, knowledge representation and knowledge inference of agricultural knowledge graph. The representation model of agricultural text such as TF-IDF, Word2Vec and BERT and classification models such as CNN, RNN and Attention were presented. Then the common corpus was described. The application of semantic understanding in agriculture from the aspects of agricultural intelligent question answering, agricultural semantic retrieval and agricultural intelligent management decision as well were

introduced. Finally, the research trend of agricultural text semantic understanding was prospected from the aspects of standardization construction of agricultural corpus, complexity of semantic understanding model, multi-modal semantic processing, multi-region and multi-language semantic understanding.

**Key words:** agricultural knowledge intelligent service; deep learning; natural language processing; agricultral text; semantic analysis

0 引言

目前互联网农业知识资源平台众多、信息量大且更新迭代快,但涉农资源整合程度较低、农业服务过程的质量较差,普遍存在农业专家/知识/技术进村下乡访达不及时、科技资源匹配不精确、技术服务标准不统一的问题,致使农民通过信息化手段获取有效的农业知识困难。为提升农业知识服务的范围、质量和效率,通过人工智能面向农业赋能是有效的方法之一,特别是随着计算机算力的提升,自然语言处理技术得到了空前发展,其中语义理解技术在知识服务方面应用最为广泛,目前在法律<sup>[1]</sup>、医学<sup>[2]</sup>、旅游<sup>[3]</sup>、农业等垂直领域主要是通过构建领域语料库并针对具体任务组合或改进通用模型,实现分析及处理方法的迁移,在局部语料中得到可观的效果,并以智能问答、知识百科、信息检索等形式进行综合应用,在实际场景中得到验证。

语义理解技术的发展分为 3 个阶段。第一阶段是基于规则的语义理解。主要是根据语言学理论建立语义生成的规则,描述各种语义成分及成分之间的结构关系和意义关系。发展至今,以专家系统为代表在农业知识服务领域已经取得了较显著成果<sup>[4-6]</sup>,根据领域专家知识推理和判断,模拟人类决策过程,以解决农业生产复杂问题。但是基于规则产生的方法对知识需求量极大,增加了成本和复杂性,且难以根据知识更新而学习。第二阶段是基于统计学的语义理解。主要利用机器学习的思想通过计算的手段利用经验来改善计算系统自身性能,先由专家事先根据任务目标对文本进行标注,并将这

些文本作为训练语料,让机器学习标注特征。其中 K 近邻、贝叶斯、支持向量机<sup>[7]</sup>、隐马尔科夫链在文本语义分类、命名实体识别方面取得较好的效果。第三阶段是基于神经网络的语义理解。深度学习本质上是一种特殊的机器学习,主要通过嵌套的概念层次来表示并实现巨大的功能灵活性,增加了运算层数,表现更为抽象,能够对语义数据进行表征学习,建立类似人脑的神经网络,模仿人脑的机制解释数据。特别随着 LSTM<sup>[8]</sup>、Transformer、BERT 等模型相继提出以及在农业领域的应用,加上迁移学习、知识蒸馏等学习方式与性能优化模型的熟化,将基于深度学习的语义理解推至一个新的高度。

在中英文语义理解任务中,二者最大的区别在于英文单词天然存在空格,可以非常容易的进行分词,而中文词间不存在区分符,且由多个字构成,所以中文文本的语义理解任务首先需要将文本进行分词。其次是词性差异,英文存在冠词和助动词,有助于语义的理解,因此,相较于英文,中文语义理解存在更大的难度。

农业领域语料的特殊性决定了面向农业知识时处理方法的差异,目前有关学者针对农业生产、加工、销售、技术服务等环节的知识服务开展了一系列的研究和应用,如基于知识图谱的农业品种、栽培、病虫害等知识百科,代替农业专家的智能问答机器人,农业标准化生产辅助决策系统等<sup>[9-15]</sup>。为了深入分析面向农业文本的语义理解技术和语义分析服务在农业领域当中的应用场景,如图 1 所示,本文对农业知识图谱、农业文本表示、农业文本分类等主要技术的发展加以总结和概括,对农业语料库、语义理解在农业领域的应用进行分析与阐述。

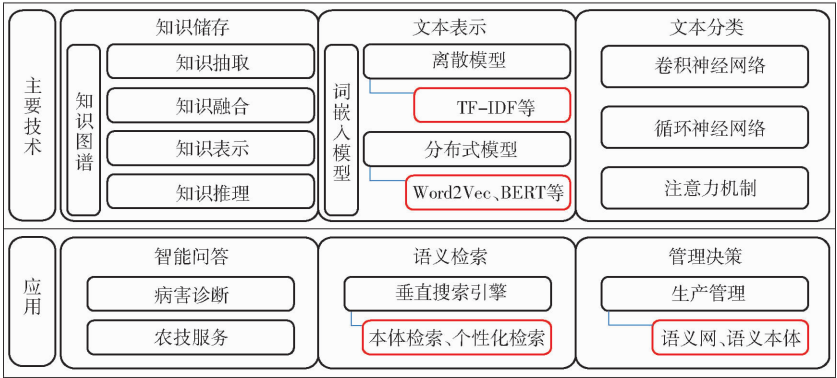


图 1 农业文本语义理解主要技术框架

Fig. 1 Framework of agricultural text semantic major technology

## 1 农业文本语义理解技术

农业文本语义理解技术从底层的农业知识存储、中间层的农业文本表示以及顶层的农业文本分类,实现了农业文本的人工智能理解全过程。其中,知识图谱是农业语义知识结构化智能存储的主要方式,通过对复杂的农业文本数据进行知识的抽取、融合、表示、推理,转化为全面表达领域知识信息的“实体-关系-实体”的三元组,实现知识的可视化表示。除此之外,还需要对人类的文字转化为计算机能够理解和计算的数据类型,则需要通过文本的表示技术,将文本数据通过词嵌入(Word embedding)方法在文本空间内进行量化的表示。形成可计算的文本向量后,计算机将载有文本特征的向量映射到多个类别上的过程,即为文本分类。

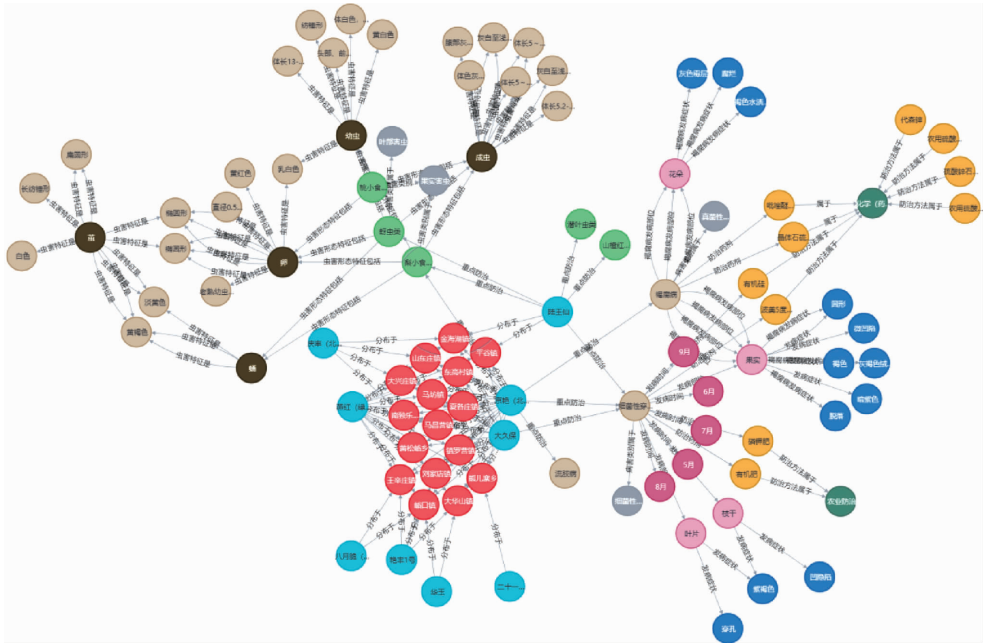


图2 农业知识图谱表示  
Fig. 2 Demonstration of agricultural knowledge graph

图谱对知识的覆盖不完整,并且依赖人工进行大量数据的标注,使知识图谱在农业中的应用服务面临困难。

如图3所示,农业知识来源主要包括表格、文本、数据库等。按照数据类型分为结构化数据、非结构化数据和半结构化数据。结构化的数据(表格、数据库等)可以直接用来构建知识图谱。非结构化

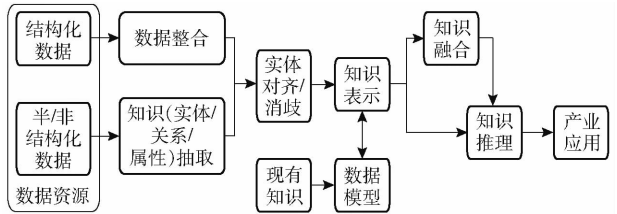


图3 知识图谱技术路线图

Fig. 3 Technology roadmap of knowledge graph

### 1.1 农业知识图谱

知识图谱的本质是一种语义网络,它是一种实体-关系-实体的三元组表示形式,2012年由Google<sup>[16]</sup>提出,最初是通过其大规模的知识表达网络来优化搜索引擎,提高搜索质量以及用户使用体验。目前,随着人工智能技术的发展,众多的智能应用、智能服务相继涌现,知识图谱逐渐开始被应用于智能搜索、知识百科、智能问答、个性化推荐、辅助决策等方面。用户搜索不再通过简单的关键词模糊匹配,而是对用户搜索内容进行语义分析理解,推理用户的实际意图,使搜索结果更具有逻辑层次,更符合用户的意图。

农业大数据存在多源异构的特点,数据分散无序,知识图谱能够有效拼接知识碎片信息(图2),在农业大数据融合中起到关键的作用,但现有的知识

的数据(文本、音频、视频、图像等)、半结构化数据则需要预先进行知识抽取,再经过知识融合,利用知识表示技术,构建可视化的知识图谱。除此之外,通过知识推理能够获得新的知识,对现有的知识图谱进行迭代更新,使知识图谱更加完善。农业领域知识图谱的研究主要集中在知识抽取、知识融合、知识表达、知识推理等方面。

#### 1.1.1 知识抽取

知识抽取是从不同来源、不同结构的数据中提取知识,形成结构化数据存入到知识图谱的过程。包含实体抽取、关系抽取、属性抽取3方面的内容,它可以克服农业领域数据存在存储的分散性和结构的不统一性问题,受到了越来越多的关注。早期的

农业知识抽取<sup>[17]</sup>是基于规则的,需要具有专业知识的专家进行人工编写三要素的抽取规则,然后通过模式匹配的方式进行实体、关系、属性的挖掘,时间成本和人力成本巨大,且农业领域本体知识众多,本体之间关系复杂,不同时空条件下相同本体拥有各异的属性,导致人工编写实体抽取规则的可扩展性较差。

针对上述问题,多项研究提出了自动化和半自动化的农业领域知识抽取方法。BiLSTM 是目前主流的知识抽取模型,也是知识抽取冷启动的基础模型,加入 CRF 之后,利用其状态转移矩阵来约束错误的标签,可以使模型的 F1 值有明显的提升。宋林鹏等<sup>[18]</sup>提出基于神经网络的词向量 + BiLSTM + CRF 的农业实体提取方法,实验证明该方法具有更好的特征抽象能力和更高的农业实体识别精度,减少了对人工特征定义的依赖。

BiLSTM 模型存在长序列前端语义稀释导致信息丢失,引入注意力机制,通过生成不同的语义向量,使注意力集中在问题的关键部位,忽略次要部分,可有效地解决问题。赵鹏飞等<sup>[19]</sup>提出在 BiLSTM + CRF 的基础上,通过注意力机制 (Attention) 获取不同语境下的实体标签,以构建农业实体识别模型,该研究解决了传统的农业命名实体识别方法对人工特征标注依赖性强、语义特征信息提取不全、实体名称不统一等问题。

BERT 是采用 Transformers 进行特征提取的深度双向预训练语义理解模型,能进一步提升语义模型的效果。袁培森等<sup>[20]</sup>采用 BERT 模型对特征向量的训练实现了对水稻表型 7 类实体关系抽取。李悦<sup>[21]</sup>、吴赛赛等<sup>[22]</sup>将 BERT 与 BiLSTM CRF 相结合,进行结构化、非结构化数据的半自动知识抽取、知识融合,并运用 Neo4j 进行知识存储,实现了农业病虫害知识的抽取和知识图谱的可视化表达。

BERT + BiLSTM + CRF 模型在农业知识抽取任务上取得了巨大的进展,但是由于网络结构庞大,参数众多,模型训练、运算耗时较长。李亮德等<sup>[23]</sup>结合知识蒸馏方法,以 BERT - ALA + BiLSTM + CRF 为教师模型,以 BiLSTM + CRF 为学生模型进行模型的蒸馏,进行农业实体抽取模型的训练,该方法解决了人工特征标注的低精确度问题,减少了深度神经网络复杂度和参数量,实现了低延迟、高精度的农业实体识别。

知识抽取对于农业知识图谱的构建具有重要的意义,同时也面临着巨大的挑战。近年来,预训练语言模型的性能得到很大程度的提升,加上深度迁移学习的发展,预训练模型到农业领域模型的迁移训练变得更加高效。随着注意力机制、Transformer、知识蒸馏等技术的提出,基于弱监督学习的农业知识智能抽取技术,

得到了快速发展,已逐渐替代了基于规则的知识抽取方法,实现对语义特征、实体信息的高效提取。

### 1.1.2 知识融合

知识融合是对不同数据源进行整合,使知识库、知识图谱的实体信息更加全面具体的技术,它包括了本体对齐、实体对齐、实体消歧、记录链接、本体匹配,其本质都是从多源信息中将相同的本体、实体进行融合。目前,国内外研究机构依据不同农产品、不同时空范围、不同生产加工流程,构建了大量的农业领域相关的知识图谱,而因为传统知识对齐等工作的人工投入成本巨大,因此始终未能形成统一的大规模农业知识图谱,亦未能实现对数据的有效利用。因此,自动、批量化的知识融合研究对于农业大数据的整合、数据资源的利用、农业决策模型的开发等具有重要的意义。实体链接是通过实体识别技术对文本中的实体进行检测,将其对应信息与知识图谱中对应实体进行链接,并加入到已有的知识图谱/知识库中,实现知识图谱智能融合的技术。夏迎春<sup>[24]</sup>在构建病虫害知识图谱的过程中,提出基于主题模型与实体链接算法 (Entity linking algorithm based on topic model and graph, ELTMG),通过构建候选实体集、构建实体相关图、计算最优链接实体 3 个步骤进行知识库融合,在 AGDISTIS 算法的基础上 F1 值提升了 5.2%,获得了更好的知识库融合的效果。创建大型知识库方面尚缺少跨库融合应用,大多研究仍在处理特定的小样本知识阶段。

现阶段,随着 NLP 技术的发展,知识融合在中文、英文等单语言的知识图谱中已获得了较好的应用成效,但在多语言的知识图谱融合上还有待研究和探索,成为未来知识融合的一个重要方向,将世界不同语言不同国家的开源知识库整合,打通语言限制,实现知识的世界范围共享。

### 1.1.3 知识表示

要运用知识图谱中的信息,需要借助知识表示。农业知识表示的内容是农业生产经验、自然规律等,以本体为核心,以 RDF 三元组为框架,表达实体、标签、属性、关系等多层语义关系。农业上对知识表示开展了多项研究,主要采用逻辑表示法、框架表示法、语义网等方法进行农业知识的描述。

卢山<sup>[25]</sup>对产生式表示法、逻辑表示法、框架表示法、面向对象表示法、语义网表示法进行比对分析,结合玉米收获机割台设计知识的特点,采用本体描述语言 OWL 进行知识表示,实现了对玉米收割台知识间复杂关系清晰的形式化表达。张熔<sup>[26]</sup>通过对各种方法的对比分析,根据水稻领域知识的复杂特征,采用框架表示法实现了基于语义的水稻病

虫害知识表示。苑超<sup>[27]</sup>通过 Hadoop 分布式计算框架运行水稻领域知识语义网,实现了云端的语义表达、查询和推理,具有快速准确的优势。

合理优化设计知识表示方案,能更好地表达关系复杂的多维度农业信息,能决定下游的知识推理和上游的知识获取的形式和难度。因此,知识表示对农业知识图谱的构建和应用都有至关重要的作用。

#### 1.1.4 知识推理

知识推理是通过已有知识推断出未知知识的过程。知识图谱中的推理主要针对实体关系进行推理,能够辅助推理出新的事实、新的关系、新的公理以及新的规则,并以此对知识图谱进行补全。知识推理主要基于逻辑规则、图结构、分布式表示、神经网络等方法。在农业领域,基于逻辑规则的农业知识推理研究较为普遍。

李雪梅<sup>[28]</sup>构建了农业科技信息资源本体,借助 Jena 推理机和推理规则,提出农业科技信息资源本体的语义推理框架,实现了农业信息资源的有效推理。杨金桂<sup>[29]</sup>通过 Cloud - OWL 构建云本体,对茶园气象知识进行表示,采用语义网规则语言 SWRL 构建推理规则,结合描述逻辑推理和语义推理进行农业云本体的语义推理,建立基于云本体农业知识服务,实现农业领域不确定性知识的高效复用。

目前,知识推理主要以提升规则挖掘效率和准确度为目标,农业领域大多数研究都采用基于规则的推理方法,而人工的规则制定对专家知识、人力及时间的消耗巨大,而随着深度网络技术的日益成熟,神经网络代替基于规则和图的推理将是未来研究的发展方向。

### 1.2 农业文本表示

文本表示是自然语言处理中的基础工作,文本表示的性能直接影响到整个自然语言处理系统的性能。文本向量化就是将文本表示成一系列能够表达文本语义的向量,是文本表示的一种重要方式。传统方法是通过构建语义词典,比较两个词拥有同义词或者上位词集的相似性来判断语义是否相似,常见的有 WordNet<sup>[30]</sup>、Probase<sup>[31]</sup>等,但构建词典存在人力物力消耗巨大、覆盖范围有限、无法及时更新的问题,垂直领域构建难度过大,应用较少;独热表示法(One - hot representation)将单词表现成一个与词典大小一致的特征向量,将只有单词对应的位置设为 1,其他位置均为 0,由于该方法本质上是一个词袋模型,不考虑词与词之间的顺序,且存在特征离散稀疏问题,对噪声非常敏感;HARRIS<sup>[32]</sup>在 1954 年提出分布假说理论,说明出现在相同上下文的词语语义相似,并由 FIRTH<sup>[33]</sup>在 1957 年进行了更加明

确的阐述,词的语义由其上下文刻画,依据该假说的词向量表示分为基于矩阵的表示、基于聚类的表示和基于神经网络的表示。HINTON 等<sup>[34]</sup>在 1986 年提出分布表示,通过训练将某种语言的每一个词映射到一个固定长度的短向量,根据词间距离判断语法、语义相似度。随着算力的提升,基于神经网络的深度学习技术在自然语言处理中逐渐占据主流,典型的包括 C&W 模型<sup>[35]</sup>、CBOW 模型<sup>[36]</sup>、Skip - Gram 模型、基于负采样的模型等。

在传统的基于机器学习的文本分类方法中,独热表示法是一种常用的文本表示方法。该方法将文本中的每个单词表示为一个向量,其维度是预处理后的文本中词汇的数量。但是,这种方法有明显的局限性。一方面,如果整体数据较大,词汇表中包含大量单词,则文本向量维数会过高,严重影响计算效率。另一方面,one - hot 忽略了上下文的语义信息,造成了严重的信息丢失。为了克服上述缺陷,HINTON 提出了词嵌入的概念。词嵌入是一种分布式表示。该方法的主要思想是将单词从高维空间映射到低维空间,解决了向量稀疏性问题。而映射到低维空间后,不同词对应的词向量之间的位置关系反映了它们的语境语义信息。为了更快、更有效地训练词嵌入,MIKOLOV 提出了两种神经网络语言模型:CBOW 和 Skip - Gram。CBOW 是根据上下文预测当前的单词,而 Skip - Gram 是根据当前的单词预测上下文。2017 年,华盛顿大学团队开发了一种基于 3 层双向 LSTM 的语境嵌入模型 ELMo,它具备捕获上下文信息的能力,比 Word2Vec 效果表现更加优秀。2018 年,OpenAI 开始使用 Transformer 构建嵌入模型,是谷歌开发的一种新的神经网络架构。Transformer 完全基于注意力机制,大大提高了 TPU 上大规模模型训练的效率,第一个模型称为 GPT。同年,谷歌开发了基于双向变压器的 BERT。BERT 使用 33 亿个单词进行训练,是目前最先进的嵌入模型。使用更大模型和更多训练数据的趋势仍在继续。OpenAI 最新的 GPT - 3 模型包含 1 700 亿个参数,谷歌的 GShard 包含 6 000 亿个参数。

近年,在农业领域常见的文本表示模型有 TF - IDF、Word2Vec、BERT 等。

#### 1.2.1 TF - IDF

词频 - 逆文件频率(Term frequency - inverse document frequency, TF - IDF)<sup>[37]</sup>是一种用于资讯检索与资讯探勘的常用加权技术,也是一种非常有效的特征提取算法。TF - IDF 是一种统计方法,用以评估字词对于一个文件集或一个语料库中的其中一份文件的重要程度。字词的重要性随着它在文件中出现



的次数成正比增加,但同时会随着它在语料库中出现的频率成反比下降。TF-IDF 方法可保留文本中具有代表性的低频词语,去除区分度低的高频词。

在农业文本处理领域,魏芳芳等<sup>[38]</sup>利用选择好的特征词进行 TF-IDF 权重计算建立相应的文本表示模型,用于衡量该特征词的重要程度。使用已经编号的类别和特征向量,对每个文档计算 TF-IDF 值,然后通过特征选择、特征项权值计算处理,最后采用 SVM 算法获得农业文本分类器,实现了中文农业文本的精确自动分类,准确率达到 95.6%,召回率达到了 96.4%。杜亚茹等<sup>[39]</sup>将浅层句法分析等语言学方法与 TF-IDF 和 C-value 等统计学方法相结合进行概念抽取;在分类关系抽取时,基于目标本体的已知一个分支,采用余弦距离计算概念与已知分支概念的语义距离,并结合概念之间的共现频度来确定层次及上下位关系。与目前中文本体的代表性方法相比,文中提出的方法在查全率和查准率方面有明显的提高。郑丽敏等<sup>[40]</sup>针对传统的 TF-IDF 没有考虑特征词对类间分布状况影响的问题,在 TF-IDF 中引入特征选择效果较好的卡方统计量(Chi-square, CHI)方法进行修正。利用改进的特征加权处理方法提高分类精度,使 FSE\_ERE 方法在高质量食品安全事件新闻文本中完成实体关系抽取工作。段青玲等<sup>[41]</sup>将 TF-IDF 方法优化及改进,进行特征项权重计算。该方法不仅考虑特征词在整个语料集中的重要程度,而且考虑特征词在各个类别之间以及各个类别内的差异性。采用基于信息熵的方法对每个类别分别提取热词候选词,最后采用基于时间变化的方法进行候选词热度计算,根据候选词热度排序结果得到热词。该方法能够有效地提取农业热词,为不同农业用户群体发现和分析产业热点提供帮助。

TF-IDF 的优点是简单快速,而且容易理解。缺点是有时候用词频来衡量文章中的一个词的重要性不够全面,有时候重要的词出现的可能不够多,而且这种计算无法体现位置信息,无法体现词在上下文的重要性。为了体现词的上下文结构,Word2Vec 算法应运而生。

### 1.2.2 Word2Vec

Word2Vec 可以提供一个高效的实现,即架构连续字包(CBOW)和 Skip-Gram 来计算字的向量表示,这些表示可以用于语言中的各种任务处理。CBOW 架构根据上下文预测当前单词,而 Skip-Gram 架构预测单词围绕当前给出的单词。Word2Vec 在给定上下文中具有相似含义的单词显示出很近的距离,从而理解并向量化文档中单词的

含义。图 4 为 MIKOLOV 提出的 Word2Vec 学习算法 CBOW 和 Skip-Gram 的模型架构。由输入层、投影层和输出层 3 部分组成,它们的输出过程不同。输入层接收  $W_{(t)} = \{W_{(t-2)}, W_{(t-1)}, W_{(t+1)}, W_{(t+2)}\}$  作为参数,其中  $W_i$  表示单词。投影层对应于多维向量的数组,并存储多个向量的总和。输出层对应于从投影层输出向量结果的层。具体而言,CBOW 类似于前馈神经网络语言模型(NNLM),并预测来自其他词向量的输出词。CBOW 的基本原理是通过分析相邻单词来预测某个单词何时出现。CBOW 的投影层将所有单词投影到同一位置,因此,所有单词的向量保持平均值并共享所有单词的位置。CBOW 的结构展示了统一组织分布在数据集中的信息的优势。相反,Skip-Gram 展示了一种从一个单词预测其他单词向量的结构。Skip-Gram 的基本原理是预测某个单词周围出现的其他单词。Skip-Gram 的投影层插入到输入层的单词周围的相邻单词。跳转图的结构显示了当新单词出现时矢量化优势。根据 MIKOLOV 的研究,当数据量较大时,CBOW 比 Skip-Gram 更快、更适合学习,而 Skip-Gram 在学习新单词时比 CBOW 表现出更好的性能。然而,其他比较 CBOW 和 Skip-Gram 性能的研究表明 Skip-Gram 的性能超过了 CBOW。

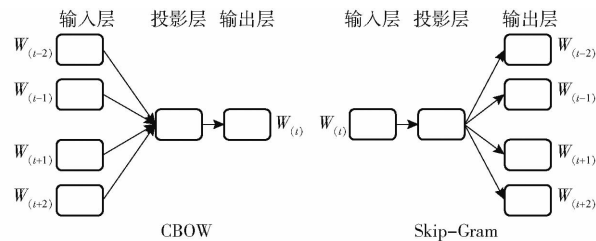


图 4 CBOW 和 Skip-Gram 架构

Fig. 4 Framework of CBOW and Skip-Gram

在农业领域,研究者针对农业文本所具有的特性,使用 Word2Vec 对农业文本进行向量化处理,王郝日钦等<sup>[42]</sup>根据水稻文本具备的特征,采用 Word2Vec 方法对文本数据进行处理与分析,能够有效地解决文本的高维性和稀疏性问题,并结合农业分词词典对文本数据进行向量化处理,然后使用注意力机制和密集连接的卷积神经网络提取文本特征,解决了农业问答社区中水稻提问数据快速自动分类的问题。赵明等<sup>[43]</sup>针对问答系统对用户问句的语义信息有较高要求的特点,首先利用 Word2Vec 将句子中的词转换为具有语法、语义信息的词向量,利用训练得到的词向量和 BIGRU 神经网络进行问句分类模型的训练,实现了对番茄病虫害问句的快速自动分类。陈瑛等<sup>[44]</sup>采用 Word2Vec 中的 Skip-Gram 模型进行训练,得到每个词的向量表示,采用

Lucene 全文检索架构和长短期记忆神经网络 (Long short-term memory, LSTM) 构建了食品安全自动问答系统。金宁等<sup>[45]</sup>运用 TF-IDF 算法拓展文本特征, 采用 Word2Vec 方法的 Skip-Gram 模型训练分词结果, 将中文词语转换为低维、连续的词向量。为进一步突出不同词语对问句含义的贡献程度, 将词语的 TF-IDF 值与 Word2Vec 词向量的乘积作为该词语的加权词向量。然后构建混合神经网络模型进行多粒度的特征提取, 实现了农业问答社区中农业问句的精确快速分类。

1.2.3 BERT

Word2Vec 产生的词向量是静态的, 不考虑上下文信息。而一些词语往往存在一词多义的现象, 因此在文本向量化过程中需要的不仅仅是一个词到向量的映射, 而应该学习一个考虑上下文的模型, BERT 预训练模型相比于 Word2Vec 为代表的词嵌入方法, 突出的进步就是更动态, 能解决一词多义的现象。

BERT (Bidirectional encoder representations from transformers) 文本预训练模型作为文本向量化转化工具获得文本特征表示, 既能获得文本语义特征, 又能解决 Word2Vec 忽略一词多义的现象。BERT 使用 Transformer 中的编码器作为特征提取器, 这种方法对上下文有很好的利用, 不需要像 BiLSTM 那样双向堆叠。配合 MLM 这样的降噪目标在大规模语料上进行训练, 根据特定领域任务进行微调, 具有良好的效果。

BERT 是一种遮蔽语言模型, 在获取词向量的过程中随机遮蔽一些词语, 然后在预训练过程中在原始词汇的位置进行预测。对于 BERT 模型的输入, 每一个词语的表示都由词语向量、段向量和位置向量共同组成, 其中, 标记 [CLS] 代表一个句子的开始, 标记 [SEP] 代表一个句子的结束。如图 5 所示。

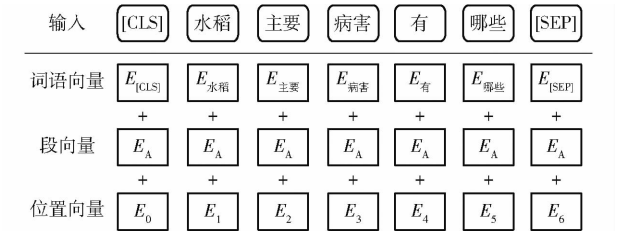


图 5 BERT 文本输入示例

Fig. 5 Demonstration of input text of BERT

在农业文本处理领域, 研究者使用 BERT 模型在农业语料库上进行训练, 取得了良好的效果。杨国峰等<sup>[46]</sup>对问句数据集进行预处理, 分别构建双向长短期记忆自注意力网络分类模型、Transformer 分类模型和基于 BERT 的微调分类模型, 并利用 3 种模型提取问句信息, 进行问句分类模型的训练。实

验结果表明采用基于 BERT 的微调常见作物病害问句分类模型, 其分类准确率、精确率、召回率、精确率和召回率加权调和平均值分别高于双向长短期记忆自注意力网络模型和 Transformer 分类模型 2~5 个百分点。袁培森等<sup>[20]</sup>获取水稻表型组学数据, 并进行标注和分类; 随后, 提取关系数据集中的词向量、位置向量及句子向量, 基于双向转换编码表示模型 (BERT) 构建水稻表型组学关系抽取模型; 最后, 将 BERT 模型与卷积神经网络模型、分段卷积网络模型进行结果比较。结果表明, 在 3 种关系抽取模型中, BERT 模型表现更佳, 精度达 95.11%、F1 值为 95.85%。王郝日钦等<sup>[47]</sup>为了解决问答社区中相同语义问句文本的快速自动检测, 提出一种基于 BERT 的 Attention-DenseBiGRU 的农业问句相似度匹配模型。针对农业文本具备的特征, 采用 12 层的中文 BERT 文本预训练模型对文本数据进行向量化处理, 并与 Word2Vec、Glove、TF-IDF 方法进行对比分析, 得出 BERT 方法能够有效地解决农业文本的高维性和稀疏性问题, 并且解决多义词在不同语境下具有不同含义的问题。

为减少不必要的算力消耗, 扩展使用场景, 以 BERT 为基础的轻量模型应运而生, 包括利用知识蒸馏技术的 DistilBERT<sup>[48]</sup>、AlBERT<sup>[49]</sup> 和 TINYBERT<sup>[50]</sup>, 通过减少预训练模型的参数降低模型的复杂度, 在文本向量化可达到显著提高文本向量化的效果。

1.3 农业文本分类

文本分类主要包括文本特征的提取和分类模型的训练。在基于机器学习的文本分类方法中, 特征提取和分类模型是两个完全独立的过程。传统的特征提取方法需要人工提取特征, 提取过程复杂, 准确率较低, 经过优化和改进, 研究者在传统的机器学习农业文本分类上取得了突破。魏芳芳等<sup>[38]</sup>通过构建农业行业关键词库、特征词选择和权重计算, 构建 SVM 农业文本分类模型, 模型准确率达 96.5%。段青玲等<sup>[51]</sup>基于 SVM 对自动抓取的农业 Web 数据进行文本分类, 实现了农业信息的自动采集和分类, 分类准确率达到 92.5%。杜若鹏等<sup>[52]</sup>在 TF-IDF 的基础上引入卡方检验值, 通过特征词频因子修正, 利用朴素贝叶斯算法进行农业科技文献文本分类, 取得了 94% 的平均准确率。而与传统方法相比, 基于深度学习的文本分类特征提取是通过多层复杂的人工神经网络特征提取得到的, 可以达到更高的准确率、更快的训练速度和更强的解释性。

近年来, 学者们已经将文本分类的重点从传统的机器学习转移到人工神经网络。人工神经网络能够从复杂的原始数据中提取抽象的层次特征, 并具

有很强的非线性映射能力。使用神经网络进行文本分类的优点之一是不需要在特征提取和选择上花费大量的时间,并且将单词的分布式表示作为特征输入到网络中。然后,神经网络可以自动提取有价值的信息用于文本分类任务。目前,基于深度学习的文本分类模型有很多,包括基于 CNN 的文本分类模型、基于 RNN 的文本分类模型以及基于注意机制的文本分类模型。

### 1.3.1 基于 CNN 的文本分类模型

卷积神经网络是一种多层复杂的神经网络结构,在图像识别领域,WANG 等<sup>[53]</sup>提出了基于深度 CNN 的 CAPTCHA 识别方法。PAN 等<sup>[54]</sup>提出了一种基于 CNN 的食物识别算法。此外,PAN 等<sup>[55]</sup>也将 CNN 与农产品相结合,提出了一种针对农产品的疾病监测系统。在文本分类领域,KIM<sup>[56]</sup>将 CNN 与自然语言相结合,提出了一种有效的文本分类方法。使用带有卷积层的 CNN 进行文本分类,并比较了不同的方法,如随机初始化、预处理词嵌入、静态输入矩阵和动态输入矩阵,最后得出静态输入矩阵分类效果最好的结论。KALCHBRENNER 等<sup>[57]</sup>提出了一个类似的模型,称为动态卷积神经网络(Dynamic convolutional neural network, DCNN)。与 KIM 提出的 CNN 方法不同,DCNN 包含 5 个卷积层和多个临时 k-max 池化层,k-max 池化层从一系列卷积滤波器中提取  $k$  个顶点值,并确保输出长度是固定的。HUANG 等<sup>[58]</sup>将字符级卷积网络进行中文的文本分类实证研究,证明了字符级卷积网络可以达到具有竞争力的分类效果。由于 CNN 和 RNN 在计算机视觉领域的结合已经取得了很好的效果,所以 XIAO 等<sup>[59]</sup>在句子分类方面将 RNN 和 CNN 结合,使用了一个 5 层的卷积网络提取高级文本特征,这些高级特征也被用作 LSTM 的输入。

在之前的文本分类中,CNN 使用了一种简单的架构。由于浅层 CNN 只能在限制窗口大小的情况下提取局部特征,CONNEAU 等<sup>[60]</sup>提出了一种深度的 CNN 来提取文本分类中的分层局部特征。它们的卷积层深度达到了 29。该模型在 8 个免费的大规模数据集上实现了稳定的性能。这是第一次证明深度对卷积神经网络的性能有提升。类似地,JOHNSON 等<sup>[61]</sup>提出了一种深度金字塔卷积神经网络(Deep pyramid convolutional neural network, DPCNN),该网络细致研究了单词级 CNN 的深度。这种新型的 DPCNN 结构能够有效地提取远程关联的特征,获得更多的全局信息。首先,该模型输入一句话到文本区域嵌入层,该层使用单词嵌入为句子中的每个单词生成向量表示。接下来是两个卷积块

的叠加和一个快捷方式。他们将特征映射的数量固定为 250 个,内核大小固定为 3 个。利用预激活的  $W\sigma(x) + b$  和身份映射的快捷连接使能深度网络训练。下采样可以有效地表示文本中更多的全局信息。在该模型中,下采样的步长为 2。该方法利用无监督嵌入训练文本区域嵌入,提高了文本区域嵌入的精度,减少了训练时间。

然而,大多数基于 CNN 的方法使用固定的窗口大小,因此无法提取可变的 n-gram 特征。WANG 等<sup>[62]</sup>提出了一种具有多尺度特征的密集连接 CNN,提取可变 n-gram 特征用于文本分类。密集连接之所以能够在上下游卷积块之间创建快捷路径,是因为将较小尺度的特征组合成大尺度的特征,从而产生可变的 n-gram 特征。虽然基于 CNN 的方法在提取可变 n-gram 特征方面发挥了很大的优势,但它们只关注局部连续词序列,而忽略了语料库中的全局词共现信息。此外,CNN 提取的局部语义特征也暴露出了其冗余性的缺点。YAO 等<sup>[63]</sup>提出了一种用于文本分类的新型图卷积网络(Graph convolutional network, GCN)。GCN 可以捕获文档和词的关系,以及全局词共现信息。

在农业领域,研究者们针对农业特定领域研究卷积神经网络在农业文本分类的应用,张明岳等<sup>[64]</sup>提出了一种基于卷积神经网络的农业问答情感极性特征抽取分析模型,结合农业分词字典,利用批规范后的卷积神经网络对数据集进行训练。实验结果表明,该方法能够准确识别测试样例集中的冗余队列,首次提出了一种农业文本二分类的解决方案。冯帅等<sup>[65]</sup>根据上述农业文本二分类的卷积神经网络模型,对卷积神经网络模型进行了优化,提出基于深度卷积神经网络的水稻知识文本分类方法,采用优选出的 4 层残差模块结构作为基本结构,使用胶囊网络(Capsule network, CapsNet)替代其池化层,设计了水稻知识文本分类模型,能够实现准确、高效的水稻知识文本分类。提出了一种水稻文本四分类的解决方案。金宁等<sup>[45]</sup>为了解决农业文本多分类问题,提出了一种农业文本十二分类的解决方案,利用双向门控循环单元神经网络获取输入词向量的上下文特征信息,构建多尺度并行卷积神经网络,进行多粒度的特征提取,实验结果表明,基于混合神经网络的短文本分类模型可以优化文本表示和文本特征提取,能够准确地对用户提问进行自动分类。

### 1.3.2 基于 RNN 的文本分类模型

递归神经网络(RNN)将双向递归结构引入神经网络,解决了输入信息之间的相互关系问题。RNN 在对文本序列进行顺序建模时具有很大的优



势。文本分类的主要应用模型是双向递归神经网络 (Bidirectional recursive neural network, BRNN),是由 SOCHER 等<sup>[66]</sup>提出的。双向递归结构假设当前输出与前面的信息和后面的信息相关,这些信息可以捕获全局的长期依赖关系。因此,RNN 在文本分类方面具有多变量模型。长短期记忆网络 (LSTM) 是 RNN 的一种改进,可以解决长期依赖问题。LSTM 通过门结构对 cell 状态进行删除或添加信息来更新每一层的隐藏状态。TANG 等<sup>[67]</sup>提出了门控循环网络模型来学习句子的语义及其上下文关系,首先通过 CNN 或 LSTM 学习文本表示,然后利用门控循环神经网络结构,将句子的语义及其关系编码成文本表示。LAI 等<sup>[68]</sup>设计了更复杂的网络结构,提出了一种递归卷积神经网络(RCNN),将 RNN 与 CNN 结合,使用双向 LSTM 来获取每个单词的上下文表示。

在农业领域,研究者针对农业特定领域研究循环神经网络在农业文本分类的应用,赵明等<sup>[69]</sup>为了对番茄病虫害智能问答系统用户问句进行高效分类,构建了基于 Word2Vec 和双向门控循环单元 (Bidirectional gated recurrent unit, BIGRU) 神经网络的番茄病虫害问句分类模型。针对问答系统对用户问句的语义信息有较高要求的特点,利用训练得到的词向量和 BIGRU 神经网络进行问句分类模型的训练。结果表明,在 2 000 条番茄病虫害数据集上,采用 BIGRU 的番茄病虫害问句分类模型,可以快速准确的进行番茄病害和番茄虫害的二分类。赵明等<sup>[70]</sup>为了对饮食文本信息高效分类,相比于上述文献的数据集,构建了 48 000 条饮食文本数据集,建立一种基于 Word2Vec 和长短期记忆网络的分类模型。由 Word2Vec 构建文本向量作为 LSTM 的初始输入,训练 LSTM 分类模型,自动提取特征,进行饮食宜、忌的文本分类。利用该方法能够高质量地对饮食文本自动分类,帮助人们有效地利用健康饮食信息。梁敬东等<sup>[71]</sup>构建一个基于 Word2Vec 和 LSTM 神经网络,包括输入层、嵌入层、LSTM 层、全连接层和输出层的句子相似度模型。构建的模型显著提升了句子相似度计算的准确率,基于该模型开发的水稻 FAQ 问答系统,能够准确匹配用户问题和水稻 FAQ 中的问题,帮助农户更好地解决水稻生产中遇到的问题。首次在农业文本领域,将深度学习模型与农业文本相似度进行结合。

1.3.3 基于注意力机制的文本分类模型

CNN 和 RNN 在文本分类任务中可以取得很好的结果,但它们的缺点是不够直观,可解释性不佳。因此研究者在上述架构的基础上加入了注意力机制。注意力机制是自然语言处理领域中常见的长期

记忆机制模型。与 CNN 和 RNN 最大的不同是,基于注意力机制的方法可以直观地呈现每个单词对结果的贡献。DU 等<sup>[72]</sup>提出了一种新的注意模式,将 RNN 和基于 CNN 的注意模型结合起来。该方法首先利用卷积运算获得注意力信号,每个注意力信号代表一个词上下文的局部语义信息;然后使用 RNN 来创建带有注意力信号的文本。一个词的注意力权重越高,它所包含的信息就越有价值,在文本构建过程中就越重要。ZHOU 等<sup>[73]</sup>也提出了一种基于注意力的双向长短期记忆网络 (Att-BLSTM)。该模型最大的优点是将神经网络注意机制与 BiLSTM 相结合,捕捉句子中最重要的语义信息。MA 等<sup>[74]</sup>提出了 Global-local mutual attention (GLMA)模型,该模型优点是能够有效地捕获局部语义特征,有效地解决全局长期依赖关系。相互注意机制包括局部引导的全局注意和全局引导的局部注意。局部引导的全局注意保留全局长期依赖的有用信息,全局引导的局部注意提取最有用、信息量最大的局部语义特征。YANG 等<sup>[75]</sup>也提出了基于 RNN 的分层注意网络 (Hierarchical attention network, HAN)模型,可以解决文本长期依赖的问题。该模型在句子级和文档级增加了注意机制,对高度重要的内容分别表示不同的权重。它可以缓解 RNN 获取文档序列信息时的梯度消失问题。然而,HANs 的训练速度要慢得多,因为它们利用了 RNN。GAO 等<sup>[76]</sup>提出了一种分层卷积注意力网络 (Hierarchical convolutional attention network, HCAN),这是一种基于自注意力机制的结构,可以在 RNN 这样的长序列中捕获语义关系,也可以在文本分类任务中实现像 CNN 那样的快速和准确性能。实验还表明,基于自注意力机制的模型可以取代基于 RNN 的模型,在降低准确率的情况下减少训练时间。在农业领域,王郝日钦等<sup>[42]</sup>对卷积神经网络 (CNN) 上下游卷积块之间建立一条稠密的链接,并结合注意力机制 (Attention),使文本中的关键词特征得以充分体现,使文本分类模型具有更好的文本特征提取精度,从而提高了分类精确率。

2 语料库

语料库的构建是所有语义分析处理的前提,大规模、高质量的语料以及知识库构建结构与可扩展性决定着语义理解技术面向农业领域任务能否实现和达到效果。

2.1 大规模通用语料库

通用型语料库体量庞大,在大型科技公司服务过程发挥重要作用,表 1 收录了常用的开源通用语料库信息,如基于知识工程构建的 FreeBase,谷歌提

出的知识图谱是该知识库典型应用,基于语义网构建的 DBpedia,融合维基百科和专家知识;国内院校及科技公司构建了北京大学 CCL 语料库、哈尔滨工业大学同义词林、搜狗互联网语料库 SogouT 等。由

于自然语言的表达方式相对一致,跨行业语料处理具有泛化性,在农业语义理解研究过程中,部分处理方式是基于大型语料库的处理而迁移获得,如:对农业文本分类、知识抽取等任务。

表 1 自然语言语料信息  
Tab.1 Natural language corpus information

| 数据集名称                           | 类型       | 体量                              |
|---------------------------------|----------|---------------------------------|
| 北京大学 CCL 语料库 <sup>[77]</sup>    | 综合       | 字符数:783463175                   |
| 搜狗互联网语料库 SogouT <sup>[78]</sup> | 综合       | 4 000 万张互联网网页                   |
| 人民日报标注语料库 <sup>[79]</sup>       | 综合       | 228 960 个句子,共 420 万字            |
| 清华 THCHS-30 语料库 <sup>[80]</sup> | 综合       | 500 条原始音频和 500 条 copy-move 篡改音频 |
| 哈尔滨工业大学同义词林 <sup>[81]</sup>     | 综合       | 77 343 条                        |
| 今日头条短文本分类数据集 <sup>[82]</sup>    | 综合       | 共 382 688 条,分布于 15 个分类中         |
| MSRA 微软亚洲研究院数据集 <sup>[83]</sup> | 地点、机构、人物 | 训练集(46 364),测试集(4 365)          |
| DBpedia <sup>[84]</sup>         | 百科       | 900 万以上的百科实体以及 6 700 万以上的三元组关系  |

2.2 农业领域语料库

由于通用语料库中垂直领域文本数据量有限,针对性不强,大部分情况下无法解决农业特定领域的问题,影响语义服务的精确度。因此,农业科研工作者在问答系统构建、模型训练过程中,通常需要针对实际情况构建特定领域语料库,如表 2 所示,农业领域目前公开的有农作物品种、农作物病虫害、农业技术服务等类型的知识库。

表 2 农业垂直领域语料  
Tab.2 Agricultural vertical corpus

| 数据集名称                      | 类型 | 体量                    |
|----------------------------|----|-----------------------|
| 中国农技推广 <sup>[85]</sup>     | 综合 | 14 401 092 条农作物的问答对数据 |
| 国家农业科学数据中心 <sup>[86]</sup> | 综合 | 756 个数据集,数据总体量 760 TB |
| 美国农业部 <sup>[87]</sup>      | 综合 | 30 亿条记录和 150 个数据来源    |
| FAOSTAT <sup>[88]</sup>    | 综合 | 粮食及农业数据总体量 996 MB     |
| 农业科技网络书屋 <sup>[89]</sup>   | 综合 | 36 694 条农作物的问答对数据     |
| 农业专业知识服务系统 <sup>[90]</sup> | 综合 | 3 亿条农业知识数据量           |

随着深度学习领域的发展,图像、文本、视频等多媒体处理边界逐渐呈现模糊化的形式,学习模型也逐渐呈现多任务处理的形式,因此语料库构建逐渐从单一类型的语料向多模态发展,如图像-文字语料库、视频-文字语料库。

3 语义理解在农业领域应用

3.1 农业智能问答

20 世纪 90 年代之后,随着互联网的发展,数据的获取变得简单,检索式的问答技术快速发展,基于

逻辑推理、模板匹配、机器学习、数据冗余性的方法相继被提出,根据问句的浅层语义去检索答案。但是检索式的问答存在答案和问题需要存在共同关键词的局限性,随着百科类网站的兴起,高质量结构化的数据获取更加方便,大量知识库被建立起来,加上机器学习技术的兴起,推动了基于知识库的问答系统研究。在农业领域,智能研究起步较晚,2007 年前后,才开始出现基于本体、知识库的农业智能问答的研究<sup>[91]</sup>。

知识库问答 (Knowledge base question & answering, KBQA) 是以自然语言的形式给出问题,通过对问题进行语义理解和解析,进而利用知识库进行查询、推理得出答案。基于知识库的农业智能问答是充分利用知识库中的数据解决问题的一项重要研究任务,其实现过程分为问题分析、文本信息检索、答案生成 3 个模块<sup>[92]</sup>,基本架构如图 6 所示。

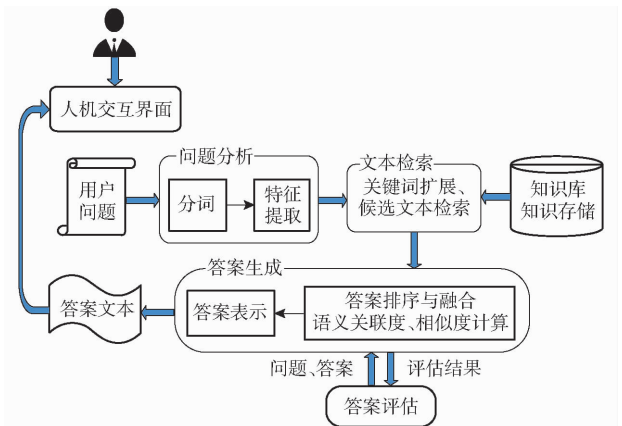


图 6 基于知识库问答系统基本架构  
Fig.6 Framework of Q&A system based on knowledge base

随着知识图谱的提出和普及,越来越多的学者将注意力放到知识图谱的农业问答系统研究中,在本体层的基础上构建数据层,利用知识图谱将结构

化和非结构化的数据进行数据抽取、融合,形成一种具有语义的知识库,更好地表达实体之间的关联性,实现实体间的上下文会话识别与推理,为智能问答的应用提供了新的知识管理途径。夏迎春<sup>[24]</sup>构建了基于 Neo4j 数据库的农业病虫害知识图谱,提出基于主题模型与图的实体链接算法,并设计研发作物病虫害知识问答系统,实现了农业病虫害知识问答。为了丰富数据源,提高泛化性,吴茜<sup>[93]</sup>收集了农作物数据、农作物病害数据和农药产品数据,通过 Protégé 工具构建了农业知识图谱,然后提出多特征的条件随机场命名实体识别算法和基于双向长短期记忆网络的属性链接方法,实现了交互式农业知识问答系统。

农业知识图谱用于理解问题的深层语义信息,满足用户的精细化需求,对实现农业知识智能化服务、农业信息化发展有一定的意义,但对于语义复杂、开放性问题则难以准确回答用户。研究开放检索生成式问答方法,通过模型训练最终生成合适的答案,对于农业问答系统的答案自动生成具有重要指导意义。王郝日钦等<sup>[47]</sup>构建的基于 BERT - Attention - DenseBiGRU 的神经网络,采用中文 BERT 预训练模型对 3 万对农业问答数据进行训练,获得农业问句的向量化表示,并输入 DenseBiGRU 和协同注意力机制模型,提取不同粒度的农业文本特征,得到适用于农业文本相似度的神经网络参数,问答匹配精确率达到了 97.2%,实现农业问句相似度的精确判断,满足了农业问答社区的需求。张明岳等<sup>[64]</sup>采用卷积神经网络模型对 8 000 条农业问句文本信息进行特征提取及分类,经过不断迭代的训练之后,得到了用于判断农业无效问句的神经网络参数。实现了农业问答文本特征抽取的任务,准确率达到了 82.7%。金宁等<sup>[45]</sup>对 12 个类别的 20 000 条农业问句进行分类,采用 TF - IDF 与 Word2Vec 相结合的方法对农业问句进行向量化处理,然后构建了基于 BiGRU 与多尺度并行的卷积神经网络模型进一步提取农业问句语义特征,模型准确率达到了 95.9%,实现了准确的农业问句分类,满足农技问答社区的需求。此外, WANG 等<sup>[94]</sup>采用 Albert + Match - LSTM 农业问答语义分类匹配方法,通过注意力机制和卷积核引入使准确率达到了 96.9%,大大降低了模型时间复杂度。

### 3.2 农业语义检索

搜索引擎在互联网信息检索中发挥着主导地位,信息检索已成为从海量信息资源中获取知识并解决问题的主要途径。随着互联网上的信息指数级增长,农业信息也随之快速膨胀,传统的关键词匹配

检索筛选信息不聚焦,搜索结果查全率低,排序依据不足。语义分析可以解析用户意图,突破关键词查询的局限性。因此,基于语义的检索方法已逐渐成为农业领域检索研究的热点。

现阶段,农业领域的语义检索主要是基于本体以及基于用户行为习惯两方面开展研究。基于本体的语义检索<sup>[95-99]</sup>在构建农业垂直领域本体库的基础上,标注本体信息,确定搜索词句和本体之间的相似、相关度,以此为据对候选搜索结果进行排序;基于用户行为的语义检索<sup>[100-102]</sup>在通过计算本体相似度的基础上加入了用户行为习惯、时间遗忘曲线等多维参数,对用户的搜索意图进行辅助定位,增加检索的准确率。特定本体库的优势在于能清晰表达领域知识的概念、结构、关系,形成具有一定结构化的数据字典工具,在这样的工具中进行检索可以使结果精度更高。要实现一个集成度高、覆盖面广、综合性强的农业全域检索方法或系统,需要构建一个大型的本体库,传统方法需要投入大量的专家资源进行人工标注和构建,难度较高,目前国内外尚未形成此类成果。因此,现阶段自动构建本体在农业领域成为研究热点,借鉴知识图谱中知识抽取相似技术,通过基于自然语言规则的模型,抽取、分析本体概念间的潜在关系,实现本体库的自动构建,但现在的研究成果离优良的理解性还有很大的差距,随着研究的不断深入,知识蒸馏、迁移学习、注意力机制、Transformer 等技术的提出,这种现状有望得到改善。

### 3.3 农业管理决策

农业管理决策语义服务能够帮助农民收集和整合生产所需的信息,通过分析提供最佳的决策方案,为农民增收致富提供技术支持,有利于提高农业生产的产量和质量。传统的决策系统通常是人工录入条件和决策数据,将用户的条件因子与系统数据库中的条件进行匹配,选取对应最匹配的解决方案,而农业的地区化、多样化导致人工录入的数据耗时耗力,且覆盖面不全,难以满足广大农业生产者的需求。近年来,越来越多的学者将语义技术引入农业决策支持系统中,通过语义理解整合互联网、物联网数据以及已有专家系统、书籍中的生产管理信息,智能化匹配用户需求,生成个性化的精准生产管理决策方案。

现阶段面向农业管理决策语义服务的研究主要可以归纳为基于语义网和语义本体两类。语义网是互联网信息实时共享的最新发展,提供了一种通用机制,允许跨不同应用程序、企业和社区共享数据,孙想等<sup>[14]</sup>、NASEEM 等<sup>[103]</sup>利用农业语义网技术,构建农业生产决策系统,克服农业多源异构数据整合

困难,解决生成决策方案不准确的问题。另一方面,王艺等<sup>[104]</sup>、WANG 等<sup>[105]</sup>、韩乐<sup>[106]</sup>通过构建语义本体,结合专家知识,利用农作物生长信息及气象因子,为管理人员提供综合信息服务和辅助型决策,实现异构、多源农业信息的整合,开发本地化的农业资源,为个体农户提供个性化、主动的信息决策服务,为种植业、养殖业等生产过程提供科学指导依据。不论是基于语义网还是语义本体,农业智能决策系统通过对农业文本的语义分析,形成适应复杂生产环境的农业生产管理决策模型,辅助农民决策及实时诊断调控,可以减少农业生产成本和环境污染,提高经济效益。

## 4 展望

语义理解技术已经广泛应用到农业知识服务领域,移动终端的广泛使用也提升了用户对农业问题精准答案的需求。在现有研究成果的基础上,总结农业语义理解研究领域的重点问题和发展趋势,认为该领域还存在如下具有挑战性的研究内容。

(1) 针对农业数据源标准化程度低的问题,面向互联网数据、专家知识数据、农业百科数据等,建立农业语义数据表达方式统一化过滤机制,构建统一标准的知识库,从而解决数据爆炸、存储滥用等问题。

(2) 农业文本信息的标注方式仍然以人工辅助

标注为主,需要大量的监督,耗时费力,半监督或者无监督的模型成为主要发展方向之一。降低语义理解模型处理的复杂度,提升模型处理的效率,根据农业知识服务应用的实际情况,结合终端处理性能提供边缘计算或者经过蒸馏后的模型,提升模型的普适性,进而全面服务基层农业科技人员和农民。

(3) 对于多模态语义处理问题,进一步研究集成图像-文本、视频-文本以及多模态组合内容的分析机制,通过统一维度的映射与模型构建,完成复杂语义的处理。

(4) 农业知识库构建完毕,面向社会提供全天候实时服务,避免垃圾信息、违规信息的注入对知识库数据安全提出新的要求,其安全贯穿在模型训练、模型预测以及服务整个过程中,此外,知识的获取途径的版权问题需要同步考虑,通过系统规则以及法律约束的方式需并行实施,也可尝试利用区块链等防篡改技术。

(5) 面向农业知识的跨区域服务,基于人工智能语义的翻译需求逐年提升,包括我国不同民族间的相互翻译、国际语言的翻译,其核心实现方式是通过系统化工程组合,将任务处理模型与翻译模型相互融合。

随着算力的提升、自然语言处理技术的发展及移动网络技术的快速升级,面向农业知识的智能化服务势必会有更加广阔的发展空间与应用价值。

## 参 考 文 献

- [1] 黄都培. 基于本体的法律信息语义检索[J]. 计算机工程与应用, 2008, 44(28): 196 - 199.  
HUANG Dupei. Ontology-based semantic retrieval for legal information[J]. Computer Engineering and Applications, 2008, 44(28): 196 - 199. (in Chinese)
- [2] 董文波, 孙仕亮, 殷敏智. 医学知识推理研究现状与发展[J/OL]. 计算机科学与探索[2022-02-11]. <http://kns.cnki.net/kcms/detail/11.5602.TP.20220125.1849.004.html>.  
DONG Wenbo, SUN Shiliang, YIN Minzhi. Research and development of medical knowledge graph reasoning[J/OL]. Journal of Frontiers of Computer Science and Technology[2022-02-11]. <http://kns.cnki.net/kcms/detail/11.5602.TP.20220125.1849.004.html>. (in Chinese)
- [3] 刘济源. 旅游领域知识图谱的构建及应用研究[D]. 杭州: 浙江大学, 2019.  
LIU Jiyuan. Research on the construction and application of knowledge graph in tourism domain[D]. Hangzhou: Zhejiang University, 2019. (in Chinese)
- [4] 孙想, 冯臣, 吴华瑞. 基于语义 Web 的农业生产知识集成技术[J]. 农业工程学报, 2008, 24(增刊 2): 186 - 190.  
SUN Xiang, FENG Chen, WU Huarui. Technology of agricultural production knowledge integration based on semantic Web[J]. Transactions of the CSAE, 2008, 24(Sup. 2): 186 - 190. (in Chinese)
- [5] 吴华瑞, 赵春江, 王纪华, 等. 基于规范化软件技术的农业专家系统开发平台研究[J]. 计算机与农业(综合版), 2003(1): 15 - 19.  
WU Huarui, ZHAO Chunjiang, WANG Jihua, et al. The research of a development platform for agricultural expert system based on standard techno-componentware[J]. Computer and Agriculture, 2003(1): 15 - 19. (in Chinese)
- [6] 杨宝祝, 赵春江, 李爱平, 等. 网络化、构件化农业专家系统开发平台(PAID)的研究与应用[J]. 高技术通讯, 2002(3): 5 - 9.  
YANG Baozhu, ZHAO Chunjiang, LI Aiping, et al. Research and application of platform for agricultural intelligent-system development[J]. High Technology Letters, 2002(3): 5 - 9. (in Chinese)
- [7] 李伟. 基于语义扩散核与支持向量机的半监督农业文本分类研究[D]. 赣州: 赣南师范大学, 2018.  
LI Wei. Semi-supervised agricultural text classification based on semantic kernel and support vector machines[D]. Ganzhou: Gannan Normal University, 2018. (in Chinese)
- [8] 赵燕. 基于机器学习的农业文本分类方法研究[D]. 沈阳: 沈阳农业大学, 2018.  
ZHAO Yan. Research on agricultural text classification method based on machine learning[D]. Shenyang: Shenyang

- Agricultural University, 2018. (in Chinese)
- [9] 吴赛赛. 基于知识图谱的作物病虫害智能问答系统设计与实现[D]. 北京:中国农业科学院,2021.  
WU Saisai. Design and implementation of intelligent question and answering system for crop disease and pests based on knowledge graph[D]. Beijing:Chinese Academy of Agricultural Sciences, 2021. (in Chinese)
- [10] 赵鹏飞,赵春江,吴华瑞,等. 基于 BERT 的多特征融合农业命名实体识别[J]. 农业工程学报,2022,38(3):112–118.  
ZHAO Pengfei, ZHAO Chunjiang, WU Huarui, et al. Recognition of the agricultural named entities with multi-feature fusion based on BERT[J]. Transactions of the CSAE, 2022, 38(3): 112–118. (in Chinese)
- [11] 张青岭,李显正,李航宇,等. 知识图谱在农业中的应用[J]. 电子技术与软件工程,2019(7):245–247.  
ZHANG Qingling, LI Xianzheng, LI Hangyu, et al. Application of knowledge graph in agriculture[J]. Information Technology, 2019(7):245–247. (in Chinese)
- [12] 胡德鹏. 基于农业本体问句分析的问答系统研究与架构设计[D]. 北京:中国农业科学院,2013.  
HU Depeng. The research of question analysis based on ontology and architecture design for question answering system in agriculture[D]. Beijing:Chinese Academy of Agricultural Sciences, 2013. (in Chinese)
- [13] 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.  
WANG Yi, WANG Ying, YUAN Ye, et al. A decision support system for fertilization and irrigation management of citrus based on semantic ontology[J]. Transactions of the CSAE, 2014, 30(9): 93–101. (in Chinese)
- [14] 孙想,吴华瑞,朱华吉,等. 基于语义 Web 的农业生产协同决策服务机制研究[J]. 农机化研究,2011,33(3):34–38.  
SUN Xiang, WU Huarui, ZHU Huaji, et al. Research on the mechanism of agricultural collaborative decision-making service based on semantic Web[J]. Journal of Agricultural Mechanization Research, 2011, 33(3):34–38. (in Chinese)
- [15] 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.  
SUO Junfeng, LIU Yong. Semantic similarity algorithm based on agricultural ontology and its application on crop ontology[J]. Transactions of the CSAE, 2016, 32(16): 175–182. (in Chinese)
- [16] AMIT S. Introducing the knowledge graph[R]. America: Official Blog of Google, 2012.
- [17] 赵明,杜亚茹,杜会芳,等. 植物领域知识图谱构建中本体非分类关系提取方法[J]. 农业机械学报,2016,47(9):278–284.  
ZHAO Ming, DU Yaru, DU Hui Fang, et al. Research on ontology non-taxonomic relations extraction in plant domain knowledge graph construction[J]. Transactions of the Chinese Society for Agricultural Machinery, 2016, 47(9): 278–284. (in Chinese)
- [18] 宋林鹏,刘世洪,王翠. 基于词向量 + BiLSTM + CRF 的农业技术需求文本实体提取[J]. 江苏农业科学,2021,49(5):186–193.  
SONG Linpeng, LIU Shihong, WANG Cui. Extraction of text entity of agricultural technology demand based on Word Vector + BiLSTM + CRF[J]. Jiangsu Agricultural Sciences, 2021, 49(5):186–193. (in Chinese)
- [19] 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.  
ZHAO Pengfei, ZHAO Chunjiang, WU Huarui, et al. Named entity recognition of Chinese agricultural text based on attention mechanism[J]. Transactions of the Chinese Society for Agricultural Machinery, 2021, 52(1):185–192. (in Chinese)
- [20] 袁培森,李润隆,王翀,等. 基于 BERT 的水稻表型知识图谱实体关系抽取研究[J]. 农业机械学报,2021,52(5):151–158.  
YUAN Peisen, LI Runlong, WANG Chong, et al. Entity relationship extraction from rice phenotype knowledge graph based on BERT[J]. Transactions of the Chinese Society for Agricultural Machinery, 2021, 52(5):151–158. (in Chinese)
- [21] 李悦. 农作物病虫害知识图谱构建研究[D]. 北京:中国农业科学院,2021.  
LI Yue. Research on the construction of knowledge graph of crop diseases and pests[D]. Beijing: Chinese Academy of Agricultural Sciences, 2021. (in Chinese)
- [22] 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.  
WU Saisai, ZHOU Ailian, XIE Nengfu, et al. Construction of visualization domain-specific knowledge graph of crop diseases and pests based on deep learning[J]. Transactions of the CSAE, 2020, 36(24):177–185. (in Chinese)
- [23] 李亮德,王秀娟,康孟珍,等. 基于语义融合与模型蒸馏的农业实体识别[J]. 智慧农业(中英文),2021,3(1):118–128.  
LI Liangde, WANG Xiujian, KANG Mengzhen, et al. Agricultural named entity recognition based on semantic aggregation and model distillation[J]. Smart Agriculture, 2021, 3(1): 118–128. (in Chinese)
- [24] 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.  
XIA Yingchun. Agriculture knowledge service system based on knowledge graph[D]. Hefei: Anhui Agricultural University, 2018. (in Chinese)
- [25] 卢山. 基于本体的玉米收获机割台设计知识融合表达方法研究[D]. 泰安:山东农业大学,2020.  
LU Shan. Research on knowledge fusion representation method of corn harvester header design based on ontology[D]. Taian: Shandong Agricultural University, 2020. (in Chinese)
- [26] 张榕. 基于语义的水稻领域知识表示分析与研究[D]. 长沙:湖南农业大学,2007.  
ZHANG Rong. Research and analysis based on semantics of rice domain knowledge expression[D]. Changsha: Hunan Agricultural University, 2007. (in Chinese)
- [27] 苑超. 云环境下基于语义网的水稻领域知识表示分析与研究[D]. 长春:吉林农业大学,2016.  
YUAN Chao. Analysis and research of rice domain knowledge representation based on Semantic Web in cloud environment[D]. Changchun: Jilin Agricultural University, 2016. (in Chinese)
- [28] 李雪梅. 基于语义的农业科技信息资源本体构建与推理服务研究[D]. 昆明:云南师范大学,2015.  
LI Xuemei. Research on ontology construction and reasoning service of agricultural science and technology information resources based on semantics[D]. Kunming: Yunnan Normal University, 2015. (in Chinese)
- [29] 杨金桂. 面向农业云本体的语义推理方法研究[D]. 合肥:安徽农业大学,2015.  
YANG Jingui. Research on the semantic reasoning method of agricultural cloud ontology[D]. Hefei: Anhui Agricultural University, 2015. (in Chinese)



- [30] FELLBAUM C, MILLER G. WordNet: an electronic lexical database[M]. MIT Press, 1998.
- [31] WU Wentao, LI Hongsong, WANG Haixun, et al. Probbase: a probabilistic taxonomy for text understanding[C]//SIGMOD, 2012: 481–492.
- [32] HARRIS Z S. Distributional structure[J]. Word, 1954, 10:146–162.
- [33] FIRTH J R. A synopsis of linguistic theory 1930—1955: Studies in Linguistic Analysis (Special Volume of the Philological Society)[C]. Oxford: Blackwell, 1957.
- [34] HINTON G E, MCCLELLAND J L, RUMELHART D E. Distributed representations[C]//Parallel Distributed Processing: Explorations in the Microstructure of Cognition, 1986: 77–109.
- [35] COLLBERT R, WESTON J. A unified architecture for neural language processing: deep neural networks with multitask learning[C]//Proceedings of the 25th International Conference on Machine Learning, 2008: 160–167.
- [36] MIKOLOV T, CHEN K, CORRADO G, et al. Efficient estimation of word representations in vector space[J/OL]. arXiv Preprint, arXiv: 1301.3781, 2013. <https://arxiv.org/abs/1301.3781>.
- [37] RAMOS J. Using TF-IDF to determine word relevance in document queries[C]//Proceedings of the First Instructional Conference on Machine Learning, 2003: 1–4.
- [38] 魏芳芳,段青玲,肖晓琰,等.基于支持向量机的中文农业文本分类技术研究[J].农业机械学报,2015,46(增刊):174–179.  
WEI Fangfang, DUAN Qingling, XIAO Xiaoyan, et al. Classification technique of Chinese agricultural text information based on SVM[J]. Transactions of the Chinese Society for Agricultural Machinery, 2015, 46(Supp.): 174–179. (in Chinese)
- [39] 杜亚茹,孙琛,赵明.基于Web的蔬菜领域本体概念和分类关系学习方法研究[J].计算机与数字工程,2015,43(10):1761–1766,1785.  
DU Yaru, SUN Chen, ZHAO Ming. Learning of concepts and taxonomic relation for vegetable domain ontology based on Chinese wiki encyclopedia[J]. Computer & Digital Engineering, 2015, 43(10): 1761–1766, 1785. (in Chinese)
- [40] 郑丽敏,齐珊珊,田立军,等.面向食品安全事件新闻文本的实体关系抽取研究[J].农业机械学报,2020,51(7):244–253.  
ZHENG Limin, QI Shanshan, TIAN Lijun, et al. Entity relation extraction of news texts for food safety events[J]. Transactions of the Chinese Society for Agricultural Machinery, 2020, 51(7): 244–253. (in Chinese)
- [41] 段青玲,张璐,刘怡然,等.基于农业网络信息分类的热词自动提取方法[J].农业机械学报,2018,49(7):160–167.  
DUAN Qingling, ZHANG Lu, LIU Yiran, et al. Automatic extraction method of hot words based on agricultural network information classification[J]. Transactions of the Chinese Society for Agricultural Machinery, 2018, 49(7): 160–167. (in Chinese)
- [42] 王郝日钦,吴华瑞,冯帅,等.基于Attention\_DenseCNN的水稻问答系统问句分类[J].农业机械学报,2021,52(7):237–243.  
WANG Haoriquin, WU Huarui, FENG Shuai, et al. Classification technology of rice questions in question answer system based on Attention\_DenseCNN[J]. Transactions of the Chinese Society for Agricultural Machinery, 2021, 52(7): 237–243. (in Chinese)
- [43] 许童羽,赵冬雪,周云成,等.基于word2vec和Attention-Seq2Seq的水稻病虫害智能问答方法研究[J].沈阳农业大学学报,2019,50(3):378–384.  
XU Tongyu, ZHAO Dongxue, ZHOU Yuncheng, et al. Research on method of intelligent Q & A for rice pests and diseases based on word2vec and Attention-Seq2Seq[J]. Journal of Shenyang Agricultural University, 2019, 50(3): 378–384. (in Chinese)
- [44] 陈瑛,陈昂轩,董玉博,等.基于LSTM的食品安全问题自动问答系统方法研究[J].农业机械学报,2019,50(增刊):380–384.  
CHEN Ying, CHEN Angxuan, DONG Yubo, et al. Methods of food safety question answering system based on LSTM[J]. Transactions of the Chinese Society for Agricultural Machinery, 2019, 50(Supp.): 380–384. (in Chinese)
- [45] 金宁,赵春江,吴华瑞,等.基于BiGRU\_MulCNN的农业问答问句分类技术研究[J].农业机械学报,2020,51(5):199–206.  
JIN Ning, ZHAO Chunjiang, WU Huarui, et al. Classification technology of agricultural questions based on BiGRU\_MulCNN[J]. Transactions of the Chinese Society for Agricultural Machinery, 2020, 51(5): 199–206. (in Chinese)
- [46] 杨国峰,杨勇.基于BERT的常见作物病害问答系统问句分类[J].计算机应用,2020,40(6):1580–1586.  
YANG Guofeng, YANG Yong. Question classification of common crop disease question answering system based on BERT[J]. Journal of Computer Applications, 2020, 40(6): 1580–1586. (in Chinese)
- [47] 王郝日钦,王晓敏,缪伟晟,等.基于BERT-Attention-DenseBiGRU的农业问答社区问句相似度匹配研究[J].农业机械学报,2022,53(1):244–252.  
WANG Haoriquin, WANG Xiaomin, MIAO Yisheng, et al. Densely connected BiGRU neural network based on BERT and Attention mechanism for Chinese agriculture-related question similarity matching[J]. Transactions of the Chinese Society for Agricultural Machinery, 2022, 53(1): 244–252. (in Chinese)
- [48] SANH V, DEBUT L, CHAUMOND J, et al. DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter[J/OL]. arXiv Preprint, arXiv:1910.01108, 2019. <https://arxiv.org/abs/1910.01108v4>.
- [49] LAN Z, CHEN M, GOODMAN S, et al. Albert: a lite bert for self-supervised learning of language representations[J/OL]. arXiv Preprint, arXiv:1909.11942, 2019. <https://arxiv.org/abs/1909.11942>.
- [50] JIAO X, YIN Y, SHANG L, et al. Tinybert: distilling bert for natural language understanding[J/OL]. arXiv Preprint, arXiv:1909.10351, 2019. <https://arxiv.org/abs/1909.10351>.
- [51] 段青玲,魏芳芳,张磊,等.基于Web数据的农业网络信息自动采集与分类系统[J].农业工程学报,2016,32(12):172–178.  
DUAN Qingling, WEI Fangfang, ZHANG Lei, et al. Automatic acquisition and classification system for agricultural network information based on Web data[J]. Transactions of the CSAE, 2016, 32(12): 172–178. (in Chinese)
- [52] 杜若鹏,鲜国建,寇远涛.基于改进TF-IDF-CHI算法的农业科技文献文本特征抽取[J].数字图书馆论坛,2019(8):18–24.  
DU Ruopeng, XIAN Guojian, KOU Yuantao. Improvement and application of TF-IDF-CHI in agricultural science text feature extraction[J]. Digital Library Forum, 2019(8): 18–24. (in Chinese)

- [53] WANG J, QIN J H, XIANG X Y, et al. CAPTCHA recognition based on deep convolutional neural network[J]. Mathematical Biosciences and Engineering, 2019,16(5):5851–5861.
- [54] PAN L, QIN J, CHEN H, et al. Image augmentation-based food recognition with convolutional neural networks[J]. Computers Materials & Continua, 2019,59(1):297–313.
- [55] PAN W, QIN J, XINAG X, et al. A smart mobile diagnosis system for citrus diseases based on densely connected convolutional networks[J]. IEEE Access, 2019,7: 87534–87542.
- [56] KIM Y. Convolutional neural networks for sentence classification[J/OL]. arXiv Preprint, arXiv: 1408.5882, 2014. <https://arxiv.org/abs/1408.5882>.
- [57] KALCHBRENNER N, GREFENSTETTE E, BLUNSON P. A convolutional neural network for modeling sentences[J/OL]. arXiv Preprint, arXiv: 1404.2188, 2014. <https://arxiv.org/abs/1404.2188>.
- [58] HUANG W, WANG J. Character-level convolutional network for text classification applied to Chinese corpus[J/OL]. arXiv Preprint, arXiv:1611.04358v2, 2016. <https://doi.org/10.48550/arXiv.1611.04358>.
- [59] XIAO Y, CHO K. Efficient character-level document classification by combining convolution and recurrent layers[J/OL]. arXiv Preprint, arXiv:1602.00367,2016. <https://arxiv.org/abs/1602.00367v1>.
- [60] CONNEAU A, SCHWENK H, BARRAULT L, et al. Very deep convolutional networks for text classification[J/OL]. arXiv Preprint, arXiv:1606.01781, 2017. <https://arxiv.org/abs/1606.01781>.
- [61] JOHNSON R, ZHANG T. Deep pyramid convolutional neural networks for text categorization[C]//Proceedings of the Annual Meeting of the Association for Computational Linguistics, 2017: 562–570.
- [62] WANG S, HUANG M, DENG Z. Densely connected CNN with multi-scale feature attention for text classification[C]//27th International Joint Conference on Artificial Intelligence, AAAI Press,2018: 4468–4474.
- [63] YAO L, MAO C, LUO Y. Graph convolutional networks for text classification[C]//Proceedings of the AAAI Conference on Artificial Intelligence, 2019: 7370–7377.
- [64] 张明岳,吴华瑞,朱华吉.基于卷积模型的农业问答语性特征抽取分析[J].农业机械学报,2018,49(12):203–210.  
ZHANG Mingyue, WU Huarui, ZHU Huaji. Analysis of extraction of semantic feature in agricultural question and answer based on convolutional model[J]. Transactions of the Chinese Society for Agricultural Machinery, 2018,49(12):203–210. (in Chinese)
- [65] 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 convolution neural network[J]. Transactions of the Chinese Society for Agricultural Machinery,2021,52(3):257–264. (in Chinese)
- [66] SOCHER R, PENNINGTON J, HUANG E, et al. Semi-supervised recursive autoencoders for predicting sentiment distributions [C]//Proceedings of the Conference on Empirical Methods in Natural Language Processing, 2011: 151–161.
- [67] TANG D, QIN B, LIU T. Document modeling with gated recurrent neural network for sentiment classification [C] // Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, 2015: 1422–1432.
- [68] LAI S, XU L, LIU K, et al. Recurrent convolutional neural networks for text classification [C] // Twenty-ninth AAAI Conference on Artificial Intelligence, 2015.
- [69] 赵明,董翠翠,董乔雪,等.基于BIGRU的番茄病虫害问答系统问句分类研究[J].农业机械学报,2018,49(5):271–276.  
ZHAO Ming, DONG Cuicui, DONG Qiaoxue, et al. Question classification of tomato pests and diseases question answering system based on BIGRU[J]. Transactions of the Chinese Society for Agricultural Machinery, 2018,49(5):271–276. (in Chinese)
- [70] 赵明,杜会芳,董翠翠,等.基于word2vec和LSTM的饮食健康文本分类研究[J].农业机械学报,2017,48(10):202–208.  
ZHAO Ming, DU Huifang, DONG Cuicui, et al. Diet health text classification based on word2vec and LSTM[J]. Transactions of the Chinese Society for Agricultural Machinery,2017,48(10):202–208. (in Chinese)
- [71] 梁敬东,崔丙剑,姜海燕,等.基于word2vec和LSTM的句子相似度计算及其在水稻FAQ问答系统中的应用[J].南京农业大学学报,2018,41(5):946–953.  
LIANG Jingdong, CUI Bingjian, JIANG Haiyan, et al. Sentence similarity computing based on word2vec and LSTM and its application in rice FAQ question-answering system[J]. Journal of Nanjing Agricultural University, 2018, 41(5): 946–953. (in Chinese)
- [72] DU J, GUI L, XU R, et al. A convolutional attention model for text classification[C]//National CCF Conference on Natural Language Processing and Chinese Computing,2017:183–195.
- [73] ZHOU P, SHI W, TIAN J, et al. Attention-based bidirectional long short-term memory networks for relation classification[C]//Proceedings of the Annual Meeting of the Association for Computational Linguistics, 2016: 207–212.
- [74] MA Q, YU L, TIAN S, et al. Global-local mutual attention model for text classification[J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2019,27(12): 2127–2139.
- [75] YANG Z, YANG D, DYER C, et al. Hierarchical attention networks for document classification[C]//Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2016: 1480–1489.
- [76] GAO S, RAMANATHAN A, TOURASSI G. Hierarchical convolutional attention networks for text classification[C]//3rd Workshop on Representation Learning for NLP 2018, 2018: 11–23.
- [77] 詹卫东,郭锐,常宝宝,等.北京大学CCL语料库的研制[J].语料库语言学,2019,6(1):71–86,116.  
ZHAN Weidong, GUO Rui, CHANG Baobao, et al. The building of the CCL corpus: its design and implementation[J]. Corpus Linguistics, 2019, 6(1):71–86,116. (in Chinese)
- [78] 搜狗互联网语料库[DB/OL]. [2017–03–01]. <http://www.sogou.com/labs/>.
- [79] 黄水清,王东波.新时代人民日报分词语料库构建、性能及应用(一)——语料库构建及测评[J].图书情报工作,2019,63(22):5–12.

- HUANG Shuiqing, WANG Dongbo. Construction, performance and application of new era people's daily segmented corpus (I)—construction and evaluation of corpus[J]. Library and Information Service, 2019, 63(22): 5–12. (in Chinese)
- [80] WANG Dong, ZHANG Xuwei, ZHANG Zhiyong. Thchs-30: a free Chinese speech corpus[J/OL]. arXiv Preprint, arXiv: 1512.01882, 2015. <https://arxiv.org/abs/1512.01882>.
- [81] 哈工大同义词林 [DB/OL]. [2022-03-10]. [http://ir.hit.edu.cn/demo/ltp/Sharing\\_Plan.htm](http://ir.hit.edu.cn/demo/ltp/Sharing_Plan.htm).
- [82] 邵云飞, 刘东苏. 基于类别特征扩展的短文本分类方法研究[J]. 数据分析与知识发现, 2019, 3(9): 60–67. SHAO Yunfei, LIU Dongsu. Classifying short-texts with class feature extension[J]. Data Analysis and Knowledge Discovery, 2019, 3(9): 60–67. (in Chinese)
- [83] 祖木然提古丽·库尔班, 艾山·吾买尔. 中文命名实体识别模型对比分析[J]. 现代计算机, 2019(14): 3–7. ZUMURANTIGULI Kuerban, AISHAN Wumaier. Analysis and comparison of Chinese named entity recognition model[J]. Modern Computer, 2019(14): 3–7. (in Chinese)
- [84] BO X, YONG X, LIANG J, et al. CN-DBpedia: a never-ending Chinese knowledge extraction system[C]//International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems, Springer, Cham, 2017.
- [85] 中国农技推广 [DB/OL]. [2022-03-10]. <http://njtg.nercita.org.cn/user/index.shtml>.
- [86] 国家农业科学数据中心 [DB/OL]. [2022-03-10]. <https://www.agridata.cn/#/home>.
- [87] 王禹, 许世卫, 李哲敏. 美国农业部 (USDA) 组织架构和职能概况[J]. 世界农业, 2015(6): 145–149. WANG Yu, XU Shiwei, LI Zhemin. USDA organization structure and function overview[J]. World Agriculture, 2015(6): 145–149. (in Chinese)
- [88] TUBIELLO F N, SALVATORE M, ROSSI S, et al. The FAOSTAT database of greenhouse gas emissions from agriculture[J]. Environmental Research Letters, 2013, 8(1): 015009.
- [89] 佚名. 农业科技网络书屋·专家在线系统[J]. 基层农技推广, 2019, 7(4): 2.
- [90] 农业专业知识服务系统 [DB/OL]. [2022-03-10]. <http://agri.ckcest.cn/index.html>.
- [91] 钟娟, 谢能付. 基于 Web 的农业知识问答系统研究[J]. 农业网络信息, 2007(7): 21–22, 75. ZHONG Juan, XIE Nengfu. Research on Web-based agricultural knowledge question answering system[J]. Agriculture Network Information, 2007(7): 21–22, 75. (in Chinese)
- [92] DWIVEDI S K, SINGH V. Research and reviews in question answering system[J]. Procedia Technology, 2013(10): 417–424.
- [93] 吴茜. 基于知识图谱的农业智能问答系统设计与实现[D]. 厦门: 厦门大学, 2019. WU Qian. Design and implementation of agriculture intelligent Q&A system based on knowledge graph[D]. Xiamen: Xiamen University, 2019. (in Chinese)
- [94] WANG X, WANG H, ZHAO G, et al. ALBERT over Match-LSTM network for intelligent questions classification in Chinese[J]. Agronomy, 2021, 11(8): 1530.
- [95] 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)
- [96] YANG Y, XIONG J, WANG S. A semantic search engine based on SKOS model ontology in agriculture[C]//International Conference on Computer and Computing Technologies in Agriculture, 2010: 110–118.
- [97] QIN X, ZHANG H, ZHENG H. Research on intelligent retrieval system for agricultural information resources based on ontology[J]. Journal of Physics Conference Series, 2019, 1168(2): 022041.
- [98] 陈叶旺, 李海波, 余金山. 一种基于农业领域本体的语义检索模型[J]. 华侨大学学报(自然科学版), 2012, 33(1): 27–32. CHEN Yewang, LI Haibo, YU Jinshan. A semantic retrieval model based on agricultural field ontology[J]. Journal of Huaqiao University(Natural Science), 2012, 33(1): 27–32. (in Chinese)
- [99] 张智慧. 基于 Nutch 的农业垂直搜索引擎的研究[D]. 保定: 河北农业大学, 2014. ZHANG Zhihui. Research on agricultural vertical search engine based on Nutch[D]. Baoding: Hebei Agricultural University, 2014. (in Chinese)
- [100] 蒋阳辉. 基于用户兴趣模型的个性化农业搜索引擎的研究与实现[D]. 杨凌: 西北农林科技大学, 2016. JIANG Yanghui. Research and implementation of personalized agricultural search engine based on user interest model[D]. Yangling: Northwest A&F University, 2016. (in Chinese)
- [101] 邢智. 内蒙古农资服务平台 Web 搜索智能优化研究[D]. 呼和浩特: 内蒙古工业大学, 2019. XING Zhi. Research on intelligent optimization of Web search on Inner Mongolia agricultural materials service platform[D]. Huhhot: Inner Mongolia University of Technology, 2019. (in Chinese)
- [102] BENEVENTANO D, BERGAMASCHI S, MARTOGLIA R. Exploiting semantics for searching agricultural bibliographic data[J]. Journal of Information Science, 2016, 42(6): 748–762.
- [103] NASEEM J, BURNEY S M A, MEHMOOD N, et al. An efficient rule based decision support system using semantic web technology[J]. International Journal of Advanced Computer Science and Applications (IJACSA), 2018, 9(12): 346–350.
- [104] 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. WANG Yi, WANG Ying, YUAN Ye, et al. A decision support system for fertilization and irrigation management of citrus based on semantic ontology[J]. Transactions of the CSAE, 2014, 30(9): 93–101. (in Chinese)
- [105] WANG Y, WANG Y, WANG J, et al. An ontology-based approach to integration of hilly citrus production knowledge[J]. Computers and Electronics in Agriculture, 2015, 113: 24–43.
- [106] 韩乐. 基于语义技术的生猪疾病诊断决策支持系统研究[D]. 重庆: 西南大学, 2016. HAN Le. Research on decision support system for pig disease diagnosis based on semantic technology[D]. Chongqing: Southwest University, 2016. (in Chinese)