· 专论与综述 ·

农业知识驱动服务技术革新综述与前沿

王元胜[1,2]，吴华瑞[1,2]，赵春江[1,2※]

（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）

摘 要：农业知识驱动服务技术是指运用先进信息技术，科学、高效调配农业领域专业知识服务资源，为农业行业提供智能化知识服务的技术，在解决农业技术服务供需严重失衡等难点问题方面具有重要意义，日益成为支撑农业转型升级和高质量发展的重要引擎，代表着核心研究方向，伴随着技术发展全过程。目前农业行业迫切需要解决的是知识供给严重不足、服务效率不高的问题，农业知识驱动服务技术经历较长时间发展，在知识高效匹配和精准供给方面取得了较大进步，特别是 2022 年 11 月以来 ChatGPT 这类技术的出现，充分展现了超大规模预训练模型在知识智能服务方面的巨大潜力，这也是农业知识驱动服务可以取得突破的关键所在，可以在这方面发挥重要作用。该文在分析农业知识驱动服务相关技术现状的基础上，展望了农业领域可行的知识驱动服务技术路径，预测农业领域知识服务大模型研发构建会呈现参数由少到多、算力由弱趋强、强化训练逐渐加深的特点得到快速发展应用，未来将在专业技术指导、农业“装备-信息-农艺”融合、农业信息系统平台服务总线等方面系统升级现有农业知识服务范式，多模态服务将得到系统融合加深，人机交互模式将向“人性化”方向进一步黏合增强，从而为农业智能化转型升级提供全新的技术支撑，引领农业知识服务从数据检索、语义匹配迈向生成式知识驱动模式转变。

关键词：农业技术服务；知识驱动；ChatGPT；超大规模预训练模型；新范式

doi：10.11975/j.issn.1002-6819.202307106

中图分类号：TP311.5 **文献标志码**：A **文章编号**：1002-6819(2024)-07-0001-16

王元胜，吴华瑞，赵春江. 农业知识驱动服务技术革新综述与前沿[J]. 农业工程学报，2024，40(7)：1-16. doi: 10.11975/j.issn.1002-6819.202307106 http://www.tcsae.org

WANG Yuansheng, WU Huarui, ZHAO Chunjiang. Agricultural knowledge driven service technology innovation: Overview and frontiers[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2024, 40(7): 1-16. (in Chinese with English abstract) doi: 10.11975/j.issn.1002-6819.202307106 http://www.tcsae.org

0 引 言

农业是国民经济根本，其生产经营管理服务的正常运转离不开科技服务，然而农业领域产业众多、地域广袤、专业错综复杂，农业生产管理中遇到的问题需要专家、农技人员及时解答，提供专业的技术服务，但是面向数以亿计的农业从业人员提出的问题，运用信息化手段驱动知识进行高效解答始终是一个不断克服实践难题的动态发展过程，其中既有供需双方人员数量上的不对等问题，也有知识服务的资源供给不匹配、不及时的问题，主要是难以在最短的时间内将农业用户所需知识快速精准地匹配给对方，导致农业知识驱动服务效率低下。围绕信息资源精准匹配问题，近些年来科研人员开展了各种类型的研究和探索，从最初的基于数据库的结构化查询、基于关键字的数据检索，到基于规则的专家系统、基于语义相似度的智能搜索，再到近几年的基于神经网络和知识图谱的知识匹配，虽然在效率上逐渐提升，但在自动化、智能化精准匹配和推送方面仍然存在诸多不足，主要体现在机器的语言理解能力不足和知识匹配的精准度不高等方面，离知识驱动目标也较远。

ChatGPT 在自然语言理解驱动知识服务方面迈出了里程碑式的一步，彰显了大模型技术在知识服务方面的卓越优势[1]，其可借助大规模预训练模型系统和人类反馈强化学习的强大优势，在“人脑思维”的基础上加入“人类反馈系统”，更具备“拟人化”的能力，实现 AI 从感知理解世界到生成创造世界的跃迁，开启了人工智能新纪元[2-3]，使得让机器具备较强的文本理解能力和一定的语义推理能力、具有强大的零次/少次学习能力成为可能[4-5]，将其作为前沿知识服务驱动技术解决农业生产经营管理服务中问题具备现实条件。

但是，我们必须要看到 ChatGPT 类大模型技术巨大的算力资源和能耗一般单位或组织难以复制，本文从基于数据和规则的农业知识驱动服务技术、语义匹配、AI 大模型技术在知识驱动领域的研究新进展等方面进行综述，提出了农业知识驱动服务技术框架，展望了农业知识服务发展趋势，为开展知识生成式的农业知识驱动服务技术研究应用提供参考，可为解决当前农业技术服务中的难点问题提供思路和借鉴。

1 农业知识驱动服务技术框架

农业知识驱动服务技术主要来源于信息领域相关技术在农业领域加深应用，较完整的农业知识驱动服务技

收稿日期：2023-07-11 修订日期：2024-01-31
基金项目：科技创新 2030 重大项目（2021ZD0113604）
作者简介：王元胜，博士，研究员，研究方向为农业大数据与人工智能。Email：642634129@qq.com
※通信作者：赵春江，博士，研究员，中国工程院院士，研究方向为农业人工智能与知识服务。Email：zhaocj@nercita.org.cn

术架构如图 1 所示，主要包括数据来源、知识驱动和应用服务等层面。在数据来源方面，采用互联网挖掘、系统集成、资源整合等技术手段，获取农业网站、技术文章、文献资料、百科知识、公开数据集、农业信息系统、专家知识、问答数据等渠道知识；此外，物联网感知数据也是重要的数据来源，一般通过部署各类传感器获取农情监测数据，为知识服务提供生产应用现场的实时数据资源，使知识服务更具有针对性。在知识驱动方面，从技术应用发展的历程看包括数据检索、基于规则、语义匹配、知识图谱、农业大语言模型、向量数据库等技术。在应用服务方面，主要包括基于各渠道信息资源和知识驱动相关技术构建服务于农业全品全域知识服务系统、人机混合推荐引擎、智能问答系统等，以综合集成的信息系统或平台的方式，将专业知识高效应用于农业实践。其中知识驱动是其关键，其技术演进路径大致经历了数据检索、基于规则、语义匹配等主要发展阶段。

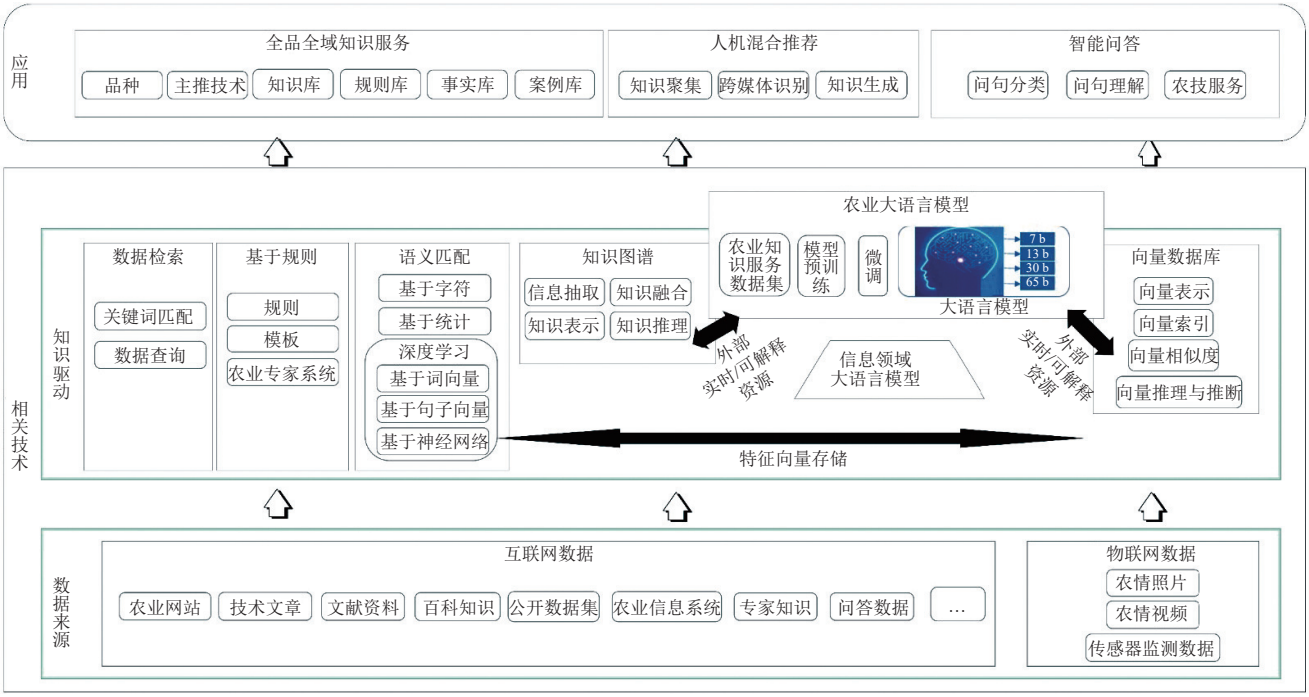


图 1 农业知识驱动服务技术框架

Fig.1 Technical framework for agricultural knowledge driven services

知识驱动技术的快速演进离不开人工智能技术的发展，从 20 世纪 60 年代专家系统的提出，到如今大语言模型的广泛应用，其主要发展进程如图 2 所示，每一项新技术的诞生都成功为农业垂直领域知识驱动发展赋能。

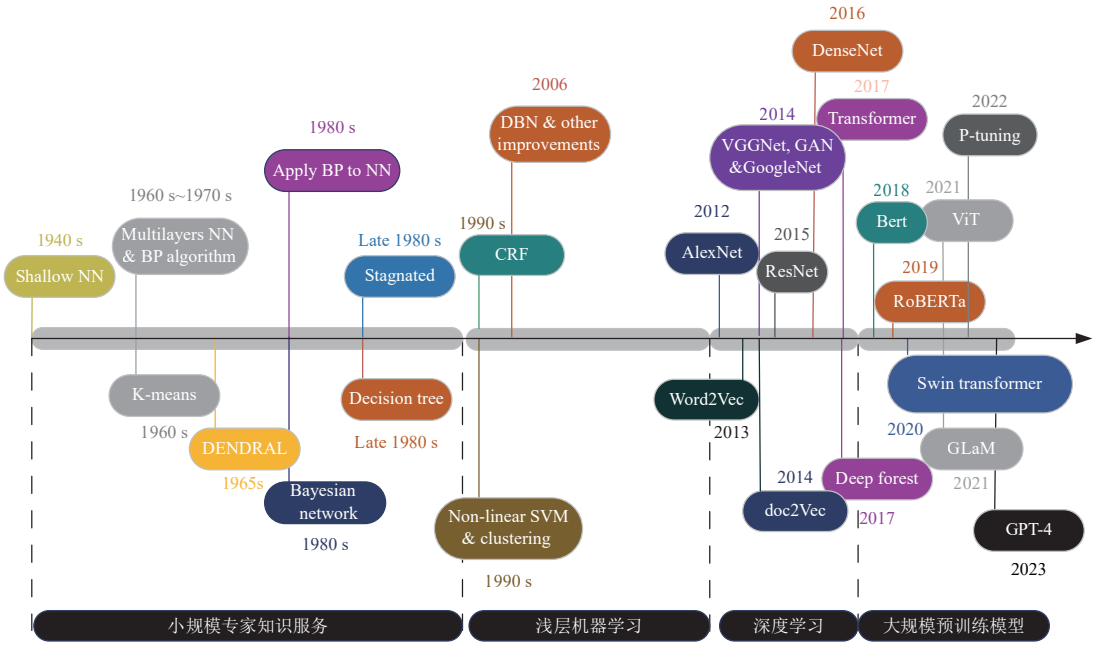


图 2 农业知识驱动服务相关的人工智能技术发展历程

Fig.2 Development history of artificial intelligence technology related to agricultural knowledge driven services

2　基于数据和规则的农业知识驱动服务技术

当前，我国农业科技成果转化率远远落后于发达国家，农业技术服务能力也相对不足，其中农技人员的专业知识无法与日益上升的农业技术发展需求相匹配，知识驱动服务技术手段明显不足是导致上述问题的关键因素[6]，研究和应用先进信息技术推动知识高效服务是农业知识驱动服务的核心内容，研究人员在这方面开展了大量工作，推进相关技术不断发展。

2.1　数据检索

早期受计算机发展水平限制，农业知识服务系统主要将知识存储在数据库中，主要依赖于基于关键词的数据检索技术提供服务，系统通过匹配用户提出的问题，与预先建立的问答数据库中的关键词进行匹配返回相关的答案或信息。关键词检索执行方式一般是后台采用数据库的 sql 语句结构化查询、搜索引擎的倒排序索引检索或类似技术实现，这种方式通常会返回大量信息，答案模糊且冗余[7]。

2.2　基于规则

为了提高服务的准确度和灵活性，农业知识服务系统中开始引入规则或规则库来驱动问答过程，通过人工编写规则和模板，指定问题和答案之间的关联，系统通过匹配问题和规则来确定最合适的答案反馈给用户。这种方法通过一系列规则的定义，以及如何根据问题类型和模式来生成答案，在一定程度上提高了问答的准确性和灵活性[8-9]。

农业专家系统中利用专家知识和规则来回答用户的问题是基于规则驱动的重要体现，这种方式模拟了领域专家的推理过程，通过推理和逻辑推断来提供问题的解决方案，早期在国内外得到广泛应用，在农业领域服务时间较长。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[9]。

3　语义匹配

早期农业知识服务方法主要依赖于人工规则和专家知识，并且对问题的解析和语义理解能力有限。随着自然语言处理和机器学习的发展，语义匹配是让计算机理解人类输入信息意图中的重要手段，也是近年来知识服务技术发展的热点和难点问题[10]。现代农业知识服务系统采用了更多语义匹配方法，如基于机器学习和深度学习的技术，以提高问答系统的准确性和智能化程度[11]。语义匹配是串联早期知识驱动服务和基于 AI 知识驱动的核心主线，可以从基于字符、基于统计和基于深度学习方面采用一系列算法或模型来实现语义匹配目标（表 1）。

表 1　语义匹配方法

Table 1　Semantic matching method

目标 Target	原理 Theory	类型 Type	名称 Name	用途 Use
文本语义相似度分析 Text semantic similarity analysis	基于字符 Character-based	字面文本相似度	编辑距离	编辑距离（levenshtein distance）计算两个字符串之间的相似性，可以用于拼写纠正、文本匹配等任务[12]
			LCS	LCS（longest common subsequence）计算两个字符串之间最长的公共子序列，常用于文本相似度和比较任务[13]
			Jaccard	用于衡量两个集合之间的相似性， 可用于文本聚类和分类任务[14]
	基于统计 Statistic-based	基于 向量空间	TF-IDF	TF-IDF (term frequency-inverse document frequency) 评估一个词在文本中的重要性，常用于信息检索和文本分类任务[15]
		基于主题模型	LSA	LSA (latent semantic analysis) 是一种基于奇异值分解的文本特征降维方法，用于捕捉文本的潜在语义信息[16]
			PLSA	PLSA (probabilistic latent semantic analysis) 是一种概率模型， 用于对文本进行主题建模和文档分类[17]
			LDA	LDA (latent dirichlet allocation) 是一种概率模型， 用于对文本进 行主题建模和文档分类，与 PLSA 类似但引入了 Dirichlet 先验分布[18]
		基于概率模型	BM25	BM25 (Best match 25) 是一种用于信息检索的评分函数， 考虑了词频和文档长度等因素，常用于文本检索和排名任务[19]
	基于深度学习 Based on deep learning	分布式词向量	Word2vec	是一种将词在连续向量空间中的表示、有效地获取词语间语义关系的词嵌入技术[20]
			BERT	双向编码器表征量（bidirectional encoder representations from transformers,BERT）是谷歌 2018 年提出一种基于 Transformer 进行双向编码、提升句子级别语义匹配度的预训练模型，被广泛应用于句子语义理解与表示任务中[21]
		无监督方法	doc2vec	是一种文档嵌入技术，用于将整个文档映射到连续向量空间， 能 够捕捉文档的语义信息[22]
			SIF	SIF (smooth inverse frequency) 是一种用于降低词语频率对词向 量的影响的方法，常用于文本表示和聚类任务[23]
		有监督方法	孪生网络架构	是一种神经网络架构，用于比较两个输入之间的相似性， 常用于文本匹配和相似度计算任务。如 HUANG 等提出的“双塔模型”（Deep Structured Semantic Model, DSSM）是最早孪生网络架构模型之一[24-25]
			交互模型架构	是指结合多种模型或技术进行信息交互和融合的架构， 用于提高文本处理和理解的效果和性能。如结合 CNN（卷积神经网络）和 DSSM（深度语义匹配模型）的 CDSSM 即为此类型，利用神经网络对文本多层次语义分析与特征向量表示优势，实现更为精准的候选集合推荐[8]

其中，TF-IDF 作为经典的特征表示方法，仍然在信息检索和文本分类中得到应用，但随着深度学习技术的发展，基于神经网络的词嵌入方法（如 Word2vec 和 BERT）逐渐成为主流，其能够克服 TF-IDF 仅统计词在该文档和词在整个文档集的频率信息方面弊端，更好地捕捉语义关系和上下文信息[26]。LSA 和 PLSA 等概率主题模型在文本主题建模方面有一定的应用，但受限于其对数据稀疏性的敏感性和复杂性，后续发展中逐渐被更先进的模型如 LDA 和 BERT 所取代；BM25 作为信息检索的评分函数，仍然有一定的应用，在问答系统的农业

知识筛查和路径规划中仍然有应用[27]。然而，随着深度学习技术的兴起，使用神经网络进行信息检索和排名的方法也在不断发展，例如使用 Transformer 模型进行文档和查询的交互[28]。在词嵌入技术方面，Word2vec 在将词语映射到向量空间方面取得了重要的突破，但在处理复杂语义和上下文信息时存在局限性，随着深度预训练模型的出现，如 BERT，基于注意力机制，采用双向 Transformer 编码器进行编码，能够更好地获得上下文语义更全面地理解词语和句子的语义，因此在自然语言处理领域得到广泛研究应用；doc2vec 作为文档嵌入技术，能够捕捉整个文档的语义信息，也在一定程度上被深度预训练模型所取代，如 BERT 可以通过加入特殊标记来处理整个文档级别的任务；SIF 作为一种降低词频影响的方法，仍然有一定的应用，但随着深度预训练模型的发展，其效果逐渐被更强大的模型所取代；孪生网络架构在以相似度匹配文本给出候选集方面发挥着重要作用，但由于模型是分别对句子进行独立编码容易产生语义分离而引起匹配层精确度降低，由此，大量借助注意力编码的交互模型被应用到改进技术路线中，取得了更好的匹配效果[29-30]。

在这条技术发展主线中，还有一些如 GPT-3、XLNet、ERNIE、RoBERTa、GPT-4 等大模型，可在大规模语料上进行训练，能够更好地理解语义、生成文本和解决自然语言处理任务，在语言生成、语言理解和预训练等方面取得了显著进展，正在与农业结合逐渐加深应用，并且不断推动着知识驱动技术的发展[31]。

文本语义表示是语义驱动的核心，现有应用于农业的研究主要聚焦在基于词向量、基于句子向量和基于神经网络等方式（表 2），实际应用中往往会多种方式并存，并且近年来的研究热点多集中在基于神经网络的相关语义匹配技术方面，分类识别准确率有了显著提升。

表 2 基于语义匹配的农业服务研究

Table 2 Research on agricultural services based on semantic matching

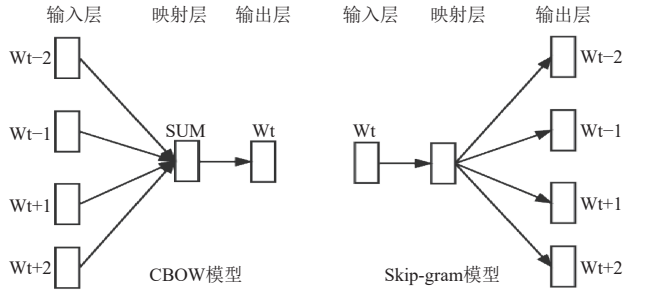
匹配方式类型 Matching method type	年份 Year	技术名称 Technical name	研究对象 Research object	效果 Effect	文献 References
词向量、神经网络 Word vectors, neural networks	2022	BERT、BiLSTM-CRF	番茄病虫害实体识别	番茄病虫害命名实体识别精度达到 85.63%	[32]
词向量、句子向量 Word vectors, sentence vectors	2018	Word2 Vec、LSTM	水稻 FAQ 问答系统语义匹配	句子相似度计算准确率达到了 93.1%	[33]
词向量、句子向量、神经网络 Word vectors, sentence vectors, neural networks	2022	Word2 Vec、BiLSTM	中国农技推广信息平台问答社区水稻问题匹配	BiLSTM-CNN 模型可高效提取文本不同粒度的特征，模型准确率和 F1 值达到 98.2% 和 88.75%	[34]
句子向量、神经网络 Sentence vectors and neural networks	2021	Bert+DSSM	作物种植问答系统句子相似度匹配	算法性能（准确率为 0.716，比 DSSM 和 ESIM 分别提升了 0.104 和 0.050 8）	[27]
句子向量、神经网络 Sentence vectors and neural networks	2023	CDSSM（CNN+DSSM）	作物病害处方推荐	模型病害诊断正确率为 71%，处方推荐准确率为 82%，优于其他 5 种作物病害处方推荐模型	[8]
基于神经网络 Based on neural networks	2022	DAMM（基于 Bert）	“中国农业技术推广信息平台”与稻米相关的问答社区中实现正确答案的智能检测	Bert 能有效解决高维稀疏问题，与其他 6 种答案选择模型相比，DAMM 在稻米相关答案选择数据集中表现最佳，MAP（Mean Average Precision）和 MRR（Mean Reciprocal Rank）分别提高了 85.7% 和 88.9%	[35]
基于神经网络 Based on neural networks	2023	EGC（基于 ERNIE+DPCNN+BiGRU）	农业新闻文本分类	相较 ERNIE 提升 1.47% 精确率、1.29% 召回率和 1.42% 的 F1 值，优于传统模型	[36]
基于神经网络 Based on neural networks	2022	BERT、Textcnn	农业问句分类	F1 值达 93.32%，比 Word2 Vec 提高 2.1%	[37]
基于神经网络 Based on neural networks	2022	TF-IDF、双向门控循环神经网络、多粒度卷积神经网络、Bert	以“中国农技推广信息平台”农技问答社区短文本问答数据语义特征和丰富语义特征提取	结合双向门控循环单元神经网络和多粒度卷积神经网络的农业问答文本分类模型表现出色，文本分类模型的分类正确率高达 95.9%，文本匹配模型在语义相似度判断方面的正确率达 94.15%，农业病虫害命名实体识别模型在农业实体的识别中取得了 92.07% 的正确率	[38]
基于神经网络 Based on neural networks	2023	BERT-LAD	农业命名实体识别	在六类农业命名实体识别中 F1 值达到 80.43%	[39]
基于神经网络 Based on neural networks	2022	BERT	农业命名实体识别	通过微调的双向编码器表示，动态生成上下文嵌入，并从 BERT 层中获得多粒度信息，F1 值达到 95.02%	[40]
基于神经网络 Based on neural networks	2023	BERT、BiLSTM-CRF	农业病害命名实体识别	准确率达 94.23%。F1 值为 80.48%	[41]

3.1 基于词向量

词向量或称为词嵌入，是语义表示的基础方法[42]。它通过将文本数据中的词汇转化为向量形式表示，使得计算机能够理解和处理文本数据。在语义驱动中，词向量扮演着重要的角色，因为它能够捕捉词汇之间的语义关系，从而更好地理解文本数据的含义。Word2vec 是早期主流的词向量训练模型（图 3），它使用 CBOW 和 Skip-gram 两种词向量模型来捕捉训练过程中词汇之间的语义关系，从而提高文本分类、聚类和语义匹配等任务的效果。因此，词向量在语义驱动中发挥着至关重要的作用[37,43]。

农业知识驱动服务技术相关研究中 Word2vec 词向量在语义匹配方面研究较多，常用于局部语料特征提取。近年来，深度双向注意力模型如 BERT、RoBERTa 等也成为主流的词向量嵌入方法。这些模型通过大规模的预训练学习，能够更准确地捕捉词汇之间的语义关系和上

下文信息，在自然语言处理任务中性能取得显著提升。梁敬东等运用 Word2vec 和 LSTM 进行语义匹配构建水稻 FAQ 问答系统，准确率达到了 93.1%[33]，刘志超等[34]在中国农技推广平台问答社区服务中运用 word2vec 对水稻问句表示初始化问句表征，然后使用 BiLSTM 长短期神经网络提取语义时序特征，最后在语义层次上使用一种包含语义信息的余弦函数计算问句相似度取得较高的准确率。

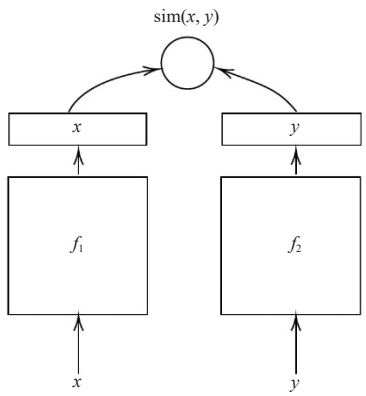


注：Wt-1 为当前词的前一个词，Wt-2 为当前词的前两个词，Wt+1 为当前词的后一个词，Wt+2 为当前词的后两个词，SUM 为上下文中所有词向量的加和，Wt 为当前词的词向量表示。
Note: Wt-1 is the previous word of the current word, Wt-2 is the first two words of the current word, Wt+1 is the last word of the current word, Wt+2 is the last two words of the current word, SUM is the sum of all word vectors in the context, and Wt is the word vector representation of the current word.

图 3　Word2vec 模型结构
Fig.3　Word2vec model structure

3.2　基于句子向量

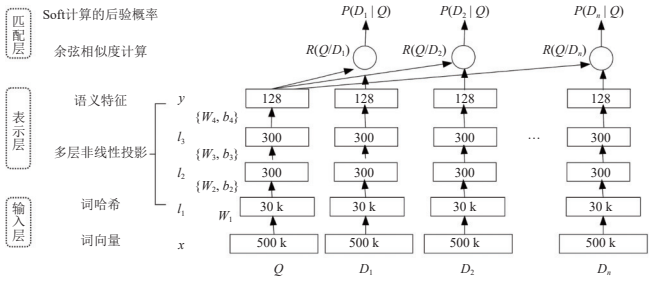
农业领域问答方面的知识服务较多，这类知识驱动一般是基于句子的语义匹配方式，代表模型是微软 HUANG 等[15,25]于 2013 年提出的“双塔模型”（DSSM）（图 4），由两个深度神经网络（f_1 和 f_2）组成，用于计算输入 x 和输出 y 的相似性。DSSM 已成为语义相似度计算的主流方法，被广泛应用于问答领域[37,44]，在农业领域作为基准模型研究比较多，取得了比较好的应用效果[8]。



注：x 为查询（query）或输入文本的表示，y 为文档（document）或目标文本的表示，f_1 为第一个神经网络塔（tower）中的特征向量，f_2 为第二个神经网络塔中的特征向量，sim(x,y) 为 x 和 y 之间的相似度。
Note: x represents the query or input text representation, y represents the document or target text representation, f_1 represents the feature vector in the first neural network tower, f_2 represents the feature vector in the second neural network tower, and sim (x, y) represents the similarity between x and y.

图 4　双塔模型结构
Fig.4　Double tower model structure

DSSM 模型结构（图 5）主要包括输入、表示和匹配 3 层，后来也被广泛采用于基于孪生网络的模型构建中。在这三层中分别采用了 trigram 高维句子向量文本转换、多个全连接层构成前馈神经网络的多层感知机（muti layer perception，MLP）、基于低维句子向量余弦的两个句子相似度计算等处理。尽管 DSSM 模型在文本匹配任务方面表现出色，但算法无法覆盖到文本的语序和上下文方面的重要信息[8,24-25]。



注：x 为查询（query）或输入文本的表示，y 为文档（document）或目标文本的表示；l_1、l_2、l_3 为神经网络中的隐藏层；Q 为查询（query）文本的集合，D_1、$D2$、D_n 为文档（document）的集合，n 表示文档的数量；P(D_1|Q)、P(D_2|Q)、P(D_n|Q) 为给定查询文本 Q 条件下，文档 D_1、D_2、D_n 的概率分布，即文档与查询的相关性概率；R(Q,D_1)、R(Q,D_2)、R(Q,D_n) 为查询 Q 与文档 D_1、D_2、D_n 之间的相关性得分或相似度。
Note: x represents the query or input text representation, y represents the document or target text representation;l_1, l_2, and l_3 are hidden layers in the neural network; Q is the set of query texts, D_1, D_2, and D_n are the sets of documents, and n represents the number of documents; P (D_1 | Q), P (D_2 | Q), P (D_n | Q) are the probability distributions of documents D_1, D_2, and D_n under the given query text Q conditions, that is, the correlation probability between the document and the query; R (Q, D_1), R (Q, D_2), R (Q, D_n) are the correlation scores or similarities between query Q and documents D_1, D_2, and D_n.

图 5　DSSM 模型结构
Fig.5　DSSM model structure

DSSM 用字向量作为输入，克服了传统单词向量（如 Word2 Vec 或 LDA）简单相加或拼接的无监督训练所引入的误差问题，不仅降低了对分词的依赖性，还提升了模型的泛化性能；同时，DSSM 整体上在各层分别采取的处理方法，解决了棘手的 OOV（out Of vocabulary）问题[45-46]。但是 DSSM 也有缺点，就是采用了词袋模型（BOW），而且采用的是弱监督、端到端模型，丧失了语序信息和上下文信息，预测结果存在不可控性。

DSSM 模型在句子向量级别提供了强大的语义表示，非常适合句子或短文本相似性计算和匹配任务，农业中运用 DSSM 技术在语义匹配的文本相似度计算方面应用较多。李菲[27]在 DSSM 模型基础上，使用 Bert 模型来替换词袋模型，句子向量表示中加入了上下文信息与位置信息，最终向量表示能更全面地表达句子的含义，进一步提升了作物种植知识相似度计算算法性能（准确率为 0.716，比 DSSM 和 ESIM 分别提升了 0.104 和 0.050 8）。张领先采用 CDSSM（CNN+DSSM）算法用于作物病害处方推荐，模型病害诊断正确率为 71%，处方推荐准确率为 82%，优于 DSSM、DSSM LSTM、Cosine、Jaccard、BM25 等 5 种模型[8]。

3.3　基于神经网络

词向量和基于句子向量的方法在一定程度上可以捕

捉词汇和句子的语义信息，但它们都很难处理长文本或序列数据中的上下文信息，无法很好地对句子或序列中的顺序关系和依赖关系建模。词向量的语义匹配方法忽略了词语在不同上下文中的多义性和歧义性，无法准确捕捉这种上下文语境的变化，句子向量往往采用简单的平均或拼接等方式来得到句子的表示，无法充分考虑句子中词语之间的复杂关系和语义组合，而基于神经网络的语义匹配方法通过引入深度学习模型和神经网络结构，可以通过多层堆叠、注意力机制、卷积神经网络、循环神经网络等来处理上下文信息、学习句子之间的关系，并通过端到端的学习来优化模型参数，这使得基于神经网络的语义匹配模型在建模复杂语义关系、处理长文本和更好地捕捉上下文信息等方面具有优势，因此近年来基于神经网络的语义匹配成为热点在较多的研究应用。

近年来，农业领域使用卷积神经网络（convolutional neural network，CNN）、循环神经网络（recurrent neural network，RNN）或 Transformer 等深度学习模型，将文本表示为高维特征向量，并通过神经网络的输出来判断文本之间的匹配方面研究比较多，并且在多种模型的信息交互和融合方面研究呈上升趋势。

针对农业智能问答知识精准匹配问题，吴华瑞、WANG 等运用基于 Attention_DenseCNN 的水稻文本分类、双向门控循环单元和多粒度卷积神经网络农业问答分类等技术，文本特征提取准确率达到了 95% 以上[35,47-48]。鲍彤等采用 BERT 和 TextCNN 结合，解决传统词向量误差，农业问句分类 F1 值达 93.32%，超参数设定优越，满足农业智能问答需求[37]。SUKUMAR 等围绕印度农业知识服务，通过基于前馈神经网络多层感知机（muti layer perception，MLP）和循环神经网络（RNN）的自然语言处理（natural language processing，NLP）技术创建农业聊天机器人，帮助偏远地区没有互联网接入的农民更好地了解作物科学种植，语义匹配准确率可以达到 97.83%[49]。MOHAMMAD 等提出一个基于深度学习的面向农业部门的鲁棒查询响应框架，并利用聚类算法将查询的相似性答案组成聚类[50]。

深度学习将多层模型应用神经网络隐藏层，自动学习和提取特征，解决了特征工程或传统机器学习中大量手工标注或操作带来效率低下问题，农业领域应用深度神经网络（Deep Neural Networks，DNN）、卷积神经网络（CNN）、递归神经网络或 Transformer 等深度学习模型大大提高了农业知识驱动服务中语义匹配的效率和准确性。

3.4 知识图谱

知识图谱以结构化的方式呈现现实世界，作为连接网络信息和人类认知的重要桥梁，在知识服务方面发挥着重要作用，基于知识图谱的语义匹配技术在语义理解和语义推理方面具有独特的优势，主要体现在基于图形式组织语义知识具有丰富的语义关系表示，能够借助结构化存储表达实体之间的层次关系、属性约束和语义关联，更准确地捕捉实体之间的语义关系和上下文信息，有助于准确匹配和推理，整合图像、视频、声音、文本等多种形式多模态的语义信息提供更全面和准确的语义匹配结果，可支持实体的关联推理、属性约束推理、路径推理等高级的推理和语义推断，与预训练语言模型成为近年来人工智能领域备受关注的热点技术[51]。

比较经典的知识图谱架构如图 6 所示，基于知识图谱的问答系统等服务是人工智能快速发展背景下产生的新技术，可以快捷地获取知识，是近年来研究的热点问题，在快速构建丰富的领域知识库，精准匹配推荐知识，提供准确的答案和个性化服务等方面发挥重要作用，IBM、微软、百度、谷歌推出的 Watson 系统、小娜、小度、谷歌搜索等典型应用都用到了知识图谱技术和数据资源，以提供更智能、准确和个性化的服务[52]。

知识图谱问答系统构建在丰富的知识库之上，是典型的知识驱动型应用，可以利用实体识别与链接、关系抽取与建模、语义表示与推理、查询与检索、可视化与交互等技术来提供精准解答，包含了对简单、多跳和聚合等多种类型问题的处理方法[52-53]。

其中，简单问题只需一个三元组回答（一跳），复杂问题需要多个相连三元组回答（多跳）。例如，“北京的气候类型是什么”是一个简单问题。而“北京适合种植哪些农作物”是一个需要多次跳跃的复杂问题。另外，聚合问题需要对多个三元组集合进行操作，如求交集、求并集、比较大小、统计数量等，以获得答案。例如，“同时适合在北京、天津和河北种植的农作物有哪些”是一个聚合问题[53]。

随着用户对知识服务的要求越来越高，传统的简单事实性知识图谱问答仅依赖单个三元组作为知识载体，无法满足这种复杂需求，通常需要利用知识图谱的属性和关系，并进行特定约束条件的检查、大规模三元组的检索以及系列关系的推理等操作才能解答。这些操作既涉及三元组检索，也涉及逻辑操作。因此，面对难度不断上升的复杂问题，需要从庞大的知识图谱中找到相应的三元组，并进行推理、统计、数值计算和聚合等多种操作及逻辑操作，使得问答任务变得更加复杂[54]。

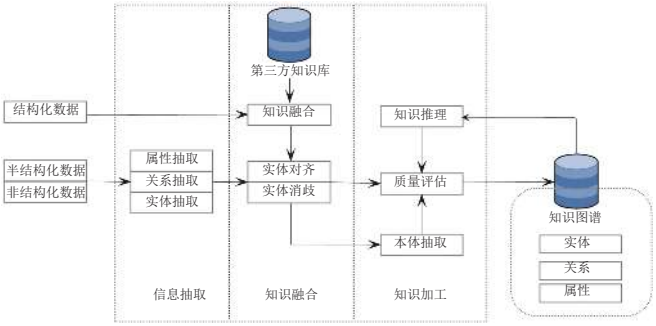


图 6 知识图谱架构

Fig.6 Knowledge graph architecture

目前，信息领域 Freebase、Wikidata、Cyc、WordNet、DBpedia、YAGO、XLORE、CN-DBpedia、Zhishi.me 等通用知识图谱构建方面，通用知识库建设已经具备一定规模，并且相应知识驱动应用研究比较多[54-56]，农业领

域围绕知识图谱的知识驱动技术研究与应用在实体识别、实体链接、关系抽取与建模、语义表示与推理、图谱查询检索、可视化与交互等方面都有涉及，在逻辑推理、规则推理、关联规则挖掘、机器学习和深度学习等知识推理技术方面研究在逐渐加深，趋势是从个别作物的局部环节开始，逐步向更多产业、更多环节全面融合渗透[57-58]，预期今后可能在与大语言模型结合的研究应用中有比较急迫的需求和发展空间[59-69]。

知识图谱在农业知识服务中研究应用主要体现在信息决策、知识问答和资源推荐等方面。

农业决策方面，利用知识图谱的信息整合匹配能力，可通过图查询语言快速匹配相关知识，助力农业人员精准决策。CHHETRI 等[70] 以木薯为研究对象，结合知识图谱和深度学习，提出低延迟的疾病识别方法，有效解决了语义技术局限性。WANG 等[71] 利用知识图谱推理提取奶牛病害隐性特征，输入 BiLSTM-CNN 神经网络进行诊断决策，性能显著优于 XGBoost 和改进 CNN。HE 等[72] 基于知识图谱开发最优施肥决策系统，提供准确的施肥方案。王艺等[73] 设计柑橘肥水管理决策支持系统，基于语义本体为果农提供精准建议，为智慧农场构建提供支撑。

知识图谱在农业智能问答系统中作为知识库提供全面的信息支持，相较于检索提供更为直观简练的答案。NIZAR 等[74] 利用作物价值链数据构建了作物自动选择工具，在全球地理范围内根据关系数据模型筛选合适作物。LI 等[75] 通过大量文献和专家知识构建农作物病虫害知识图谱，基于 Pyqt5 框架实现作物问答算法。随着知识图谱技术发展，多模态问答、多跳推理等新技术在农业发展中逐渐发挥重要作用，杨硕等[76] 引入上下文注意力和 CNN 网络，获取多模态图谱特征，提高问答准确率。谷刘涛等[77] 提出基于知识图谱嵌入和路径推理的多跳问答模型，设计水稻病虫害问答原型系统。

知识图谱推荐系统通过用户历史行为和兴趣分析，过滤冗余信息，为农业人员提供个性化服务。XIE 等[78] 提出基于关注因子分解器和知识图谱的农产品推荐算法，通过 MLP 集成农产品特征向量和用户嵌入向量，实现用户点击率预测。ZOU 等[79] 使用 RippleNet 模型构建玉米新品种推广网络，考虑气象因子实现县域范围内精准推荐。LEI 等[80] 创建多模态分层食谱知识图谱，结合用户偏好提供可靠食谱推荐。WANG 等[81] 融合遥感图像的知识图谱通过图卷积网络和深度图注意力网络改进冷启动推荐准确性。

总体而言，知识图谱是通过统一的知识组织形成的明确、有机结构的知识库。结合图谱数据库技术，它具备强大的知识推理能力，为各领域的知识驱动应用提供了条件。谷歌能够利用知识图谱的语义搜索能力，具备一定的“联想”功能，根据知识图谱的结构、查询情境和查询意图，返回与搜索词相关的实体、概念等知识，以知识卡片的形式提供给用户，为其提供相对详细的答案[53]。农业领域知识图谱应用目前在信息抽取、知识融合层面较多，受知识库资源等因素限制，大规模调度三元组进行复杂推理方面的应用研究还比较少。其中在自然语言处理方面正在迅速发展图神经网络可以处理复杂的知识图谱、捕捉实体之间的关系以及进行推理，预期可能会在农业问答系统研究中发挥重要作用[82-85]。

4　AI 大模型技术在知识驱动领域的研究进展

近年来，随着大数据和云计算的快速发展，深度学习和神经网络在自然语言领域的突破，以及增强学习方法的兴起，以 ChatGPT 类 AI 模型为代表的对话系统的研究应用进入了一个新的高潮，核心是借助超大规模的预训练模型，从海量数据中提炼出隐性知识库用于知识服务，特别是在问答服务方面具有流畅的语言组织能力、拟人化的文本水平和强大的逻辑表达能力等优点，代表了当前技术问答在知识驱动方面的重要发展方向[1]。最有影响力、热度较高的技术包括大规模预训练模型、Transformer、向量数据库等，其中 Transformer 是底座式模型技术，贯穿在最新知识服务的各个方面，无论是信息领域，还是包括农业在内的其他各垂直领域，近年来的知识服务大多都会应用到 Transformer 模型。

4.1　大规模预训练模型

超大规模预训练模型可以概括为是具有极大参数量和计算资源要求的深度学习模型，通常通过在大规模数据集上进行自监督或无监督的预训练来获得强大的表示学习和迁移学习能力，近年来发展迅速，呈现出以大数据驱动为特征的趋势，能够在小样本或零样本情况下适应，同时还能够实现跨模态的关联，引领了在深度学习领域的发展方向[86]。

这类模型通过学习大量的公开数据，获得了语言处理、上下文理解、推理、多模态处理以及知识管理等多种能力，并且通过数据收集、预处理、预训练和微调等过程，将学习到的丰富知识存储在庞大的参数中，能够根据用户的输入，凭借强大的语言理解和语义表示能力，从模型参数中推导出接近人类响应的最佳答案或回复[1]。因此，大规模预训练模型显著降低了人工智能应用的门槛，成为人工智能最新里程碑式技术，是当前技术问答等知识服务中可能迅速成为热点的知识驱动技术[87-90]。

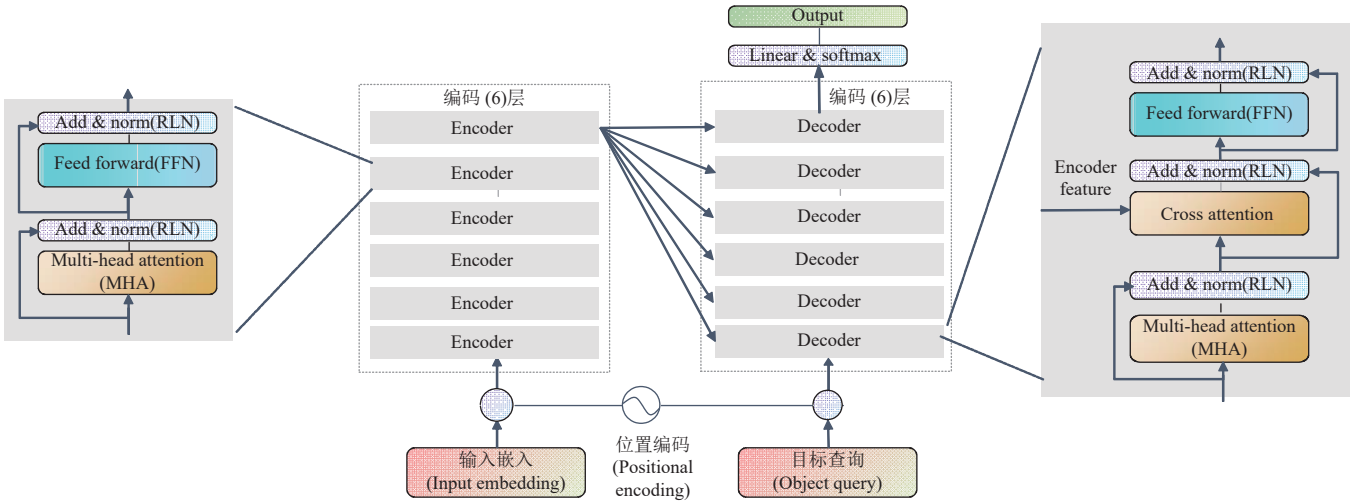
4.1.1 Transformer

当前计算机算力和大数据积累已经达到可观的量级，VASWANI 等[91] 基于自注意力机制的 Transformer 模型的提出加速了大模型发展进程。

Transformer 模型通过定义注意力机制来实现对一组键值对的映射，从而取代了 RNN 等时序神经网络结构，以处理序列信息。它利用自注意力机制和前馈神经网络对特征进行自我学习和自我调整，具备出色的并行计算能力，并在一定程度上缓解了特征信息丢失的问题。计算原理涉及计算输入词向量之间的关联程度，并使用关联关系来分配权重，以准确反映不同词向量的重要性。除了考虑词向量的个体特征，还融合了词向量与其他样本词向量之间的关系，以提高并行计算效率和降低计算复杂度，从而实现大规模 AI 模型的训练[92-93]。

Transformer 已经成为自然语言处理领域基础性架构，是支撑 ChatGPT 完全基于注意力机制摆脱了人工标注数据集缺陷的底层技术，在 ChatGPT 等大规模语言模型中扮演着重要的角色。

Transformer 网络由可训练的位置编码（position encoding）方法、可进行全局交互的自注意力机制（self-attention）、可捕获多方面信息的多头注意力机制（multi-head attention）、对特征进行非线性转换和映射的前馈神经网络（feed forward network）组成（图 7），基于包含 6 层编码器和 6 层解码器的编码器－解码器架构实现特征提取，嵌入（input embedding）引入位置编码后，通过 6 个编码器层的迭代，得到编码特征，这些特征在解码器的交叉注意力模块中与目标查询进行注意力计算，得到每个目标查询与整体的关注程度。在 6 个解码器层的迭代后，通过线性分类器和归一化（liner and softmax）操作，获得最终的输出结果[91,94]。



注：RLN 为相对位置编码，FFN 为前馈神经网络，MHA 为多头注意力机制。
Note: RLN represents relative position encoding, FFN represents feedforward neural network, and MHA represents multi head attention mechanism.

图 7 Transformer 网络结构示意图
Fig.7 Knowledge graph Architecture Transformer Network Structure Diagram

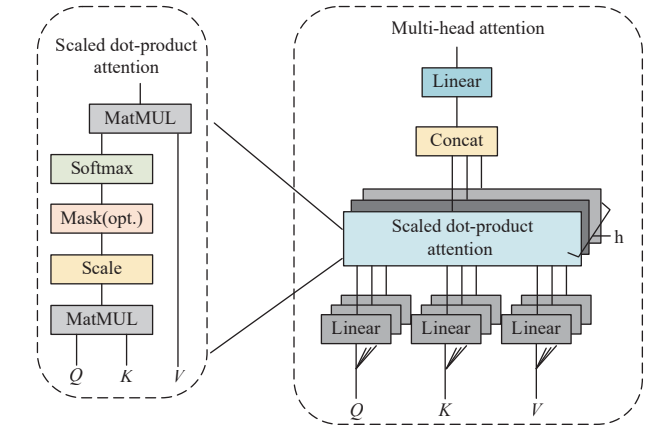
1）位置编码器

为解决放弃了循环结构后 Transformer 无法直接捕捉文本中词语位置信息的问题，为每个词的位置分配一个对应的向量，利用正弦、余弦函数进行编码，记录序列数据之间位置方面的相关性，同时处理全局信息，无论元素位于序列中距离有多远，都不需要关注顺序关系而直接对数据进行并行处理，使得计算速度和存储效率都得到显著提升[94-95]。

2）多头注意力机制

Transformer 引入自注意力机制取代以往循环递归等结构，由多个自注意力模块实现序列全局建模。文献 [83] 中采用缩放点积注意力，用点积计算注意力，然后进行尺度缩放。相比通常的自注意力模块，它降低了计算量，提高了速度和空间利用效率。基本结构如图 8 所示。计算步骤是，将输入 x 转换为 3 个不同向量：查询向量 $\boldsymbol{q}$，键向量 $\boldsymbol{k}$ 和值向量 $\boldsymbol{v}$；对于一组输入的向量，将对应的 $\boldsymbol{q}$，$\boldsymbol{k}$，$\boldsymbol{v}$ 向量合成 $\boldsymbol{Q}$，$\boldsymbol{K}$，$\boldsymbol{V}$ 矩阵；使用 $\boldsymbol{Q}$，$\boldsymbol{K}$ 矩阵相乘计算得到分数矩阵 Score；因矩阵乘积带来参数膨胀，为使梯度稳定，对 Score 除以 d_k；Score 矩阵使用 softmax 归一化；将得到的 Score 矩阵与 $\boldsymbol{V}$ 矩阵进行点乘，得到自注意力层输出结果[91,96]。

通过自注意力机制，Transformer 计算每个词与全部词之间的注意力，以获得全局语义信息。这使得模型能够捕捉到长距离的依赖关系，超越一般的感受野范围[97-99]。



a. 缩放点积注意力机制
b. Scaling dot product attention mechanism

b. 多头注意力机制
b. Multi head attention mechanism

注：Q、K、V 为查询向量、键向量和值向量，h 为注意力头。
Note: Q, K, and V represent query vectors, key vectors, and value vectors, while h represents attention head.

图 8 Transformer 的注意力机制
Fig.8 The Attention Mechanism of Transformer

多头注意力机制将 $\boldsymbol{Q}$（查询向量）、$\boldsymbol{K}$（键向量）、$\boldsymbol{V}$（值向量）矩阵拆分为多个子空间，并将低维参数映射到高维不同子空间，计算注意力权重。最后，连接所有子空间的注意力信息并与权重矩阵相乘，得到最终的注意力结果，在不增加总参数的前提下，使得模型能够同时关注输入序列的多个方面内容，通过降低维度和并

行计算，提高自注意力机制的性能，使得 Transformer 模型能够捕捉更全面、丰富的特征。通过多头注意力，模型可以同时捕捉全局和局部相关性，从而有效建模序列中不同范围的关联关系，以便更好地编码多个关系和微小差异（图 8 b）。这种机制极大地增强了模型的表达能力，并为模型提供了更丰富的特征表示。

3）其他关键模块

Transformer 中 6 个相同的编码层都由多头自注意力机制和全连接的前馈神经网络两个子层组成，每个子层上都有一个残差连接和标准化层。6 个相同的解码层都包含自注意图机制和前馈神经网络，此外在两层之间包含一个多头自注意力层。随着 Transformer 网络深度的增加，特征的数值分布可能发生数量级上的巨大差异，为了让各个特征均能发挥一定的作用，保证特征分布的稳定性，需要去除数据在数量级上的差异，因此引入了残差网络模型（图 7 中 RLN），以防止整个训练的错误率可能会随着梯度的消失而无法持续下降[91,100]。

在每个编码器层和解码器层中的自我注意层之后加入了一个简单的前馈神经网络，由两个线性变换层和其中的一个非线性激活函数组成[87]。

在解码器层的交叉注意力模块中（图 7 中的 Cross Attention），使用编码器的输出作为 $\boldsymbol{K}$，$\boldsymbol{V}$ 矩阵与查询向量组成的 $\boldsymbol{Q}$ 矩阵进行计算，由此步骤获得查询向量与全局特征之间的注意力 [87]。

综合分析，Transformer 在 ChatGPT 以及问答这类技术中具备以下几个方面的优势[21,91,101]：

1）序列建模：Transformer 通过自注意力机制（self-Attention）实现了对输入序列的建模。它能够捕捉到序列中不同位置之间的关系，从而更好地理解上下文和语义。在 ChatGPT 中，这种序列建模的能力使得模型能够理解用户输入的完整语境，并生成连贯、一致的回答。

2）上下文编码：Transformer 的编码器部分用于将输入序列（如对话历史）转化为高维表示。这种表示捕捉了序列中每个元素的上下文信息，并将其用于生成响应。通过对上下文进行编码，ChatGPT 能够理解先前的对话内容，从而生成相关的回答。

3）生成语言模型：Transformer 在 ChatGPT 中用作生成语言模型。通过自回归的方式，模型逐步生成下一个词或字符，基于之前生成的内容和上下文进行预测。Transformer 的解码器部分在每个时间步都生成一个新的词或字符，使得模型能够自动、连贯地生成回答。

4）多层次表示：Transformer 使用多层的编码器和解码器，每一层都能够对输入进行更深入的抽象和建模。这种多层次的表示能够捕捉不同级别的语义和结构信息，使得 ChatGPT 在生成回答时具有更好的语义准确性和流畅度。

因此，相比之前基于 Word2vec 等词嵌入方式的局部语义相似度匹配的知识驱动方式而言，Transformer 提取语义特征更加灵活，借助 Transformer，ChatGPT 可方便地实现对上下文的建模、序列的编码和生成语言模型等关键任务，知识驱动更加高效。Transformer 的自注意力机制和多层次表示能力使得 ChatGPT 能够生成自然、连贯的对话回答，并提供人性化的交互体验。

4.2　GPT

除基础模型外，支撑技术问答取得最新突破还得益于生成式预训练模型（generative pre-training transformer，GPT），是 OpenAI 实验室在 2018 年提出的基于 Transformer 架构的自动生成模型，与 BERT 双向编码所不同的是，GPT 只采用自左向右的单向解码器，较新 GPT3 网络结构如图 9 所示，本质上是根据前文预测下一个最有可能出现的单词，非常适合问答。OpenAI 在 GPT 模型基础上引入了指令学习、有监督精调以及基于人类反馈的强化学习等技术，经过多次迭代升级后，于 2022 年 11 月 30 日推出了全新的对话式通用人工智能工具 ChatGPT（chat generative pre-trained transformer），是 AIGC 应用的重大突破，被认为是继数据库和搜索引擎之后的全新一代的“知识表示和调用方式”[102]。

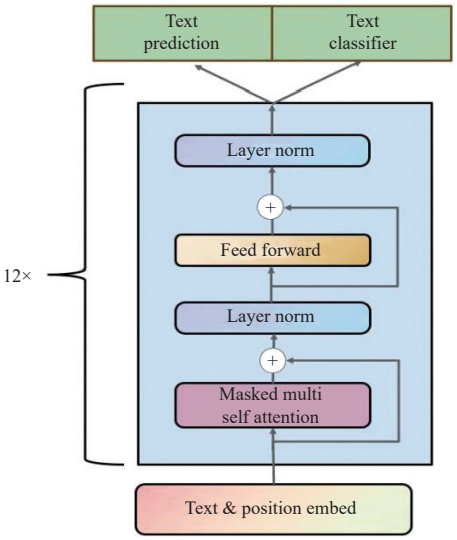


图 9　GPT-3 网络结构图

Fig.9　GPT-3 network structure diagram

4.3　大语言模型

最新知识驱动服务引擎集中体现在大语言模型上，从近年来发展动态看这点尤为突出。国外大模型研究起步较早，以 OpenAI、谷歌等公司为代表经过多年研究，大模型版本不断迭代、规模逐渐扩大、性能不断提升，近 5 年推出了 GPT1-4、Bert、GlaM、PaLM、LLaMa 等大模型，国内企业和科研院所积极跟进，近 3 年相继推出“文心一言”“盘古”“悟道”等超大规模预训练模型，力图缩小与国外差距。农业领域从零开始构建大语言模型不现实，这些为农业知识服务培育了良好的应用生态[103]，如表 3 所示。

4.4　向量数据库

还有一种值得注意的前沿技术向量数据库，是将数据存储为高维向量的数据库，是一种存储和检索基于向量表示的数据系统，用于高效的向量索引、相似度计算和向量推理与推断。这些向量是特征或属性的数学表示。

每个向量都有一定数量的维数，根据数据的复杂性和粒度，其范围可以从数万到数千个不等。这些向量通常是通过对原始数据应用某种转换或嵌入函数来生成的，如文本、图像、音频、视频等。该嵌入函数可以基于机器学习模型、单词嵌入、特征提取算法等多种方法。

向量数据库的主要优点是可以根据向量距离或相似度对数据进行快速、准确的相似度搜索和检索，不一定要使用传统方法基于精确匹配或预定义标准来查询数据库，而是可以使用向量数据库来根据其语义或上下文信息来查找最相似或最相关的数据，应用场景涉及自然语言处理（natural language processing，NLP）、计算机视觉（computer vision，CV）、推荐系统（recommendation system，RS）和其他需要语义理解和数据匹配的领域[110-111]，这方面的优势应用到农业领域知识服务中，必然会让 AI 服务更贴近应用实际，服务以单纯文本方式提升为多模态方式，更加智能高效。

表 3 主要大语言模型

Table 3 Main large language models

区域 Region	模型名称 Model name	发布机构 Publishing agency	发布时间 Release time	参数规模 Parameter scale	特点 Characteristic	说明 Illustrate
国外 Foreign	GPT-1	OpenAI	2018 年 6 月	1.17 亿	单模态	基于 Transformer 的 Decoder 架构[101]
	GPT-2	OpenAI	2019 年 2 月	15 亿	单模态	基于 Transformer 的 Decoder 架构[104]
	GPT-3 Davinci	OpenAI	2020 年 5 月	1 750 亿	单模态	基于 Transformer 的 Decoder 架构[105]
	GPT-4	OpenAI	2023 年 3 月	万亿级	多模态	具体参数未公开。基于 Transformer 的 Decoder 架构[106]
	Bert	谷歌	2018 年 10	3.4 亿	单模态	基于 Transformer 的双向编解码架构[21]
	Switch Trans-former	谷歌	2021 年 1 月	1.6 万亿	单模态	首个万亿级语言模型，在 transformer 结构中引入了一个全局模块，解决处理长序列时效率低下的问题
	GLaM	谷歌	2021 年 12 月	1 200 亿	单模态	用图结构建模文本的关系和语义信 息，但输入和输出仍然是基于文本的[107]
	PLaM2	谷歌	2023 年 6 月	3.6 万亿	多模态	增强实验室机器人 Bard 的数学能力、逻辑推理能力。2022 年 4 月发布 5 400 亿参数 PLaM，基于 Transformer 的 decoder 架构
	LLaMA2	Meta	2023 年 7 月	700 亿	多模态	基于 Transformer 的 decoder 架构，开源\可商用[108]
国内 Domestic	文心一言	百度	2023 年 3 月 16 日	2 600 亿	多模态	基于文心 ERNIE（基于 Transformer）
	盘古	华为	预计 2023 年 7 月	1.085 万亿	多模态	基础结构是 Transformer 的 Decoder 架构
	通义千问	阿里	2023 年 4 月	超过 10 万亿	多模态	技术底座来自 transformer
	紫东太初 2.0	中科院自 动化所	2023 年 5 月	千亿级	多模态	基于华为全栈国产化软硬件平台昇腾 AI 与昇思 MindSpore
	MOSS	复旦大学	2023 年 2 月	160 亿	多模态	基于 Tranformer 架构，开源
	ChatGLM	清华大学、智谱 AI	2023 年 3 月	1 300 亿	单模态	基于 Transformer，是自回归空格填充 的 GML（general language model pre- training with autoregressive Blank In- filling）架构，开源，允许免费商用[109]
	ChatGLM2-6B	清华大学	2023 年 5 月	78 亿（文本 62 亿/图片 16 亿）	多模态	文本基于 GML 架构图形部分采用了 Swin Tranformer 架构，开源，允许免费商用
	星火	讯飞	2023 年 5 月	超过 1 000 亿个参数）	多模态	基于 Transformer 架构
	混元	腾讯	2023 年 6 月	万亿级	多模态	采用腾讯太极机器学习平台自研的训练框架 AngelPTM

总体上，国内外知识驱动相关技术发展趋势呈现出越来越强的智能化、垂直化和个性化趋势，其中大语言模型作为当前热点关键技术发挥着重要作用。大语言模型的出现和快速发展使得知识服务系统能够更好地理解和生成自然语言，提供更准确、全面的答案，并具备了更强大的语境理解和推理能力。大语言模型的应用进一步促进了知识图谱、深度学习模型、向量数据库等在相关领域的交叉发展，在智能助理、智能客服、知识服务等多个领域展现了巨大的潜力和广阔的前景。其中，向量数据库助力基于神经网络的语义理解代替了过去的关键字搜索，在消除大语言模型“幻觉”、实时性不足、可解释性差等问题方面将发挥重要作用[112-113]。

在这种趋势下，农业领域已经开始有快速步入大模型应用进程方面案例，美国农业科技公司农民商业网络（farmers business network）推出了 Norm，可通过 GPT-3.5 技术从多个在线数据源获取信息，提供农业生产的各种解决方案[114]。印度农业部门创造了占 54% 的国家劳动力的巨大的就业机会，但是一直受知识和基础设施短缺的困扰，特别是在农村地区，所依赖的 Kisan 呼叫中心存在网络拥堵和代表知识不完全等限制，农业部门利用呼叫中心数据集训练自然语言模型，研发基于 WhatsApp 的聊天机器人 RASA，基于语义相似度匹配驱动知识与农民轻松沟通，在起步条件下取得了比较好的预期目标[115]。最近印度已经研发出 KissanGPT，利用 ChatGPT 技术和自己的知识库来提供准确和有用的答案，主要目的是弥补印度农业部门技术服务不足问题，提供最新人工智能服务，并帮助农民最大限度地提高作物产量[31]。

中国是农业大国，农业技术服务经历了数据库、专家系统和云服务平台等主要服务形式，支撑服务的关键技术也由原始的关键词检索，到基于规则推理，发展至近几年的语义匹配，当前正在结合国家重点研究计划和行业发展需求，推进大语言模型与农业应用的结合，特

别是在农业知识驱动服务方面将会有新的应用成果服务于农业产业[116]。

总结来说，农业领域知识服务方面技术发展目前有几个明显的趋势，一是在计算机领域流行的大语言模型正在加速向农业领域渗透，也是农业知识高效服务的潮流和热点；二是大量高质量数据积累已成为共识[117]，推进农业领域知识服务高质量数据集进程会加快；三是在克服时效性不足方面，大语言模型与知识图谱、向量数据库和其他自有知识库相结合也是必然趋势；四是农业知识服务中多模态需求日益突出，基于多模态的大模型综合服务技术将会得到深入研究应用。

5　结论与展望

随着全球人工智能发展步伐加快，感知智能日趋成熟，由大语言模型为引领的认知智能正在发生翻天覆地的变化，人工智能加速知识驱动相关技术发展越来越快。从小规模专家知识服务的 40 年、到浅层机器学习算法研究的 20 年，到深度学习算法研究的 7 年，再到最近大规模预训练模型研究的 5 年的发展历程，可以看出大规模语言模型已经处在 AI 发展进程中的自主智能化新阶段，以自然语言为基础和核心的知识服务迎来前所未有的新机遇。农业知识驱动服务技术必然会搭乘这趟快车实现技术革新和范式升级，加快推进现代农业跨越式发展。

5.1　知识积累受到空前重视并加速发展

大模型能够成功研发构建的前提是需要从海量数据学习中获得通用语言理解能力和推理能力，农业领域的垂直化知识要靠自身结合产业实际完成，这将催生一批高质量的农业知识服务数据集加速形成，知识图谱这类来源可靠、构建过程有严格质量把关、有知识专家和领域专家共同参与的知识库建设将得到重视和加强，并且随着大模型服务的深入，将会不断促进和带动知识的更新迭代，推进知识循环进入良性发展轨道。

5.2　大语言模型将得到快速发展

大语言模型加快农业领域知识服务高质量发展进程已毋庸置疑，首先需要解决的是低成本问题，目前信息领域已经有较多大语言模型诞生，其中有不少是开源模型，并且针对大模型的微调、强化学习等知识迁移技术同步得到快速发展，为农业领域的公益知识服务奠定了很好的基础，预计农业领域知识服务大模型会走参数由少到多、算力由弱趋强、强化训练逐渐增强的实用主义发展路线，目标是提供低成本、精准、高效的农业技术服务。

5.3　集成应用更加新颖全面

在技术的不断推陈出新中，符合人工智能从半结构化、非结构化数据中提取数字特征并进行存储、加工利用需求的向量数据库，可弥补大语言模型时效性不足的知识图谱等技术会得到加强，农业知识驱动相关的系统平台在技术集成上将更加新颖全面，知识服务将更加精准、高效、灵活、全面。

5.4　产学研用结合更加系统深入

农业知识服务需要全面系统的知识，同时在服务的过程中也会产生大量数据和知识，这些正是新技术能够提供服务和进一步升级演进所需要的，在新范式知识服务推动下，农业领域产学研用的结合将更加紧密，形成相互融合促进的技术服务进化模式。

[参　考　文　献]

[1] VAN DIS E A M, BOLLEN J, ZUIDEMA W, et al. ChatGPT: five priorities for research[J]. Nature, 2023, 614(7947): 224-226.

[2] FUI-HOON NAH F, ZHENG R, CAI J, et al. Generative AI and ChatGPT: Applications, challenges, and AI-human collaboration[J]. Journal of Information Technology Case and Application Research, 2023, 25(3): 277-304.

[3] KASNECI E, SEßLER K, KÜCHEMANN S, et al. ChatGPT for good? On opportunities and challenges of large language models for education[J]. Learning and Individual Differences, 2023, 103: 102274.

[4] WU T, HE S, LIU J, et al. A brief overview of ChatGPT: The history, status quo and potential future development[J]. IEEE/CAA Journal of Automatica Sinica, 2023, 10(5): 1122-1136.

[5] CHOWDHERY A, NARANG S, DEVLIN J, et al. Palm: Scaling language modeling with pathways[J]. Journal of Machine Learning Research, 2023, 24(240): 1-113.

[6] MCCABE M, TRIPATHI S, DOOLEY D M, et al. CPO: a crop planning and production process ontology and knowledge graph[J]. Knowledge Graph Technologies: the Next Frontier of the Food, Agriculture, and Water Domains, 2023, 104: 1187090.

[7] SHEN X, JIA A L, SHEN S, et al. Helping the ineloquent farmers: Finding experts for questions with limited text in agricultural Q&A communities[J]. IEEE Access, 2020, 8: 62238-62247.

[8] 张领先，赵聃桐，丁俊琦，等. 基于 CDSSM 的作物病害处方推荐方法[J]. 农业机械学报，2023，54(3)：308-317.
ZHANG Lingxian, ZHAO Dantong, DING Junqi, et al. Recommendation method of crop disease prescription based on CDSSM[J]. Journal of Agricultural Machinery, Transactions of the Chinese Society for Agricultural Machinery, 2023, 54(3): 308-317. (in Chinese with English abstract)

[9] YOGISH D, MANJUNATH T N, HEGADI R S. Survey on trends and methods of an intelligent answering system[C]//2017 International Conference on Electrical, Electronics, Communication, Computer, and Optimization Techniques (ICEECCOT). 2017: 346-353.

[10] GIUNCHIGLIA F, SHVAIKO P. Semantic matching[J]. The Knowledge Engineering Review, 2003, 18(3): 265-280.

[11] ROY P K, SAUMYA S, SINGH J P, et al. Analysis of community question‐answering issues via machine learning and deep learning: State‐of‐the‐art review[J]. CAAI Transactions on Intelligence Technology, 2023, 8(1): 95-117.

[12] LI Y, LIU B. A normalized Levenshtein distance metric[J]. IEEE transactions on pattern analysis and machine intelligence, 2007, 29(6): 1091-1095.

[13] BERGROTH L, HAKONEN H, RAITA T. A survey of longest common subsequence algorithms[C]//Proceedings Seventh International Symposium on String Processing and Information Retrieval, SPIRE, 2000: 39-48.

[14] WU S, LIU F, ZHANG K. Short text similarity calculation based on jaccard and semantic mixture[C]//In Bio-Inspired Computing: Theories and Applications. Singapore, Springer, 2021: 37–45.

[15] HUANG C H, YIN J, HOU F. A text similarity measurement combining word semantic information with TF-IDF method[J]. Chinese Journal of Computers, 2011, 34(5): 856-864.

[16] 朱颖东，钟勇. 结合优化的文档频和 LSA 的特征选择方法[J]. 计算机工程与应用，2009，45（34）：121-123，143.
ZHU Haodong, ZHONG Yong. Feature selection method combined on optimized document frequency with LSA[J]. Computer Engineering and Applications, 2009, 45(34): 121-123, 143. (in Chinese with English abstract)

[17] HOFMANN, T. Unsupervised learning by probabilistic latent semantic analysis[J]. Machine learning, 42(2001): 177-196.

[18] BLEI D M, NG A Y, JORDAN M I. Latent dirichlet allocation[J]. Journal of Machine Learning Research, 2003, 3: 993-1022.

[19] 范晨熙，黄理灿，李雪利. 基于 Lucene 的 BM25 模型的评分机制的研究[J]. 工业控制计算机，2013，26(3)：78-79.
FAN Chenxi, HUANG Lican, LI Xueli. Research on scoring mechanism of BM25 model based on lucene[J]. Industrial Control Computer, 2013, 26(3): 78-79. (in Chinese with English abstract)

[20] CHURCH, K. W. Word2Vec[J]. Natural Language Engineering. 2017, 23(1), 155-162.

[21] DEVLIN, J. , CHANG, M. W. , LEE, K, et al. Bert: Pre-training of deep bidirectional transformers for language understanding[EB/OL]. [2019-05-24]https://doi.org/10.48550/arXiv.1810.04805.

[22] LAU J H, BALDWIN T. An empirical evaluation of doc2vec with practical insights into document embedding generation[EB/OL]. [2016-07-19] https://doi.org/10.48550/arXiv.1810.04805.

[23] PELLEGRINI V. Self-Supervised Fine-Tuning of sentence embedding models using a Smooth Inverse Frequency model: Automatic creation of labels with Smooth Inverse Frequency model[D]. Turin:Politecnico di Torino, 2023.

[24] LIN G, SHEN C, VAN DEN HENGEL A, et al. Efficient piecewise training of deep structured models for semantic segmentation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016: 3194-3203.

[25] HUANG P S, HE X, GAO J, et al. Learning deep structured semantic models for web search using clickthrough data[C]//Proceedings of the 22nd ACM international conference on Information & Knowledge Management. 2013: 2333-2338.

[26] YADAV D, DESAI J, YADAV A K. Automatic text summarization methods: A comprehensive review[EB/OL].[2022-10-28] https://doi.org/10.1007/s42979-022-01446-w.

[27] 李菲，基于数据增广的作物种植问答系统研究[D]. 合肥：安徽农业大学，2021.
Li Fei. Research on Crop Planting Question-Answering System Based on Data Augmentation[D]. Hefei: Anhui Agricultural University, 2021. (in Chinese with English abstract)

[28] WU C, WU F, QI T, et al. Hi-Transformer: hierarchical interactive transformer for efficient and effective long document modeling[EB/OL]. [2021-12-09]https://doi.org/10.48550/arXiv.2106.01040.

[29] CHANDRASEKARAN D, MAGO V. Evolution of semantic similarity—a survey[J]. ACM Computing Surveys (CSUR), 2021, 54(2): 1-37.

[30] HAN M, ZHANG X, YUAN X, et al. A survey on the techniques, applications, and performance of short text semantic similarity[J]. Concurrency and Computation: Practice and Experience 2021, 33(5): e5971.

[31] Farm Smarter Not Harder: KissanGPT for revolutionizing Indian farming[M/OL].[2023-04-15] https://kisanudyojak.in/kissangpt-smart-farming-tool-for-indian-farmers.

[32] ZHANG W, WANG C, WU H, et al. Research on the Chinese named-entity–relation-extraction method for crop diseases based on BERT[J]. Agronomy, 2022, 12(9): 2130.

[33] 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 with English abstract)

[34] 刘志超，王晓敏，吴华瑞，等. 基于 BiLSTM-CNN 的水稻问句相似度匹配方法研究[J]. 中国农机化学报，2022，43(12)：125-132.
LIU Zhichao, WANG Xiaomin, WU Huarui, et al. Research on rice question-and sentences similarity matching method base on BiLSTM-CNN[J]. Journal of Chinese Agricultural Mechanization, 2022, 43(12): 125-132. (in Chinese with English abstract)

[35] WANG H, WU H, ZHU H, et al. A residual LSTM and Seq2seq neural network based on GPT for Chinese rice-related question and answer system[J]. Agriculture-Basel, 2022, 12(6): 1-19

[36] 杨森淇，段旭良，肖展，等. 基于 ERNIE+DPCNN+BiGRU 的农业新闻文本分类[J]. 计算机应用，2023，43(5)：1461-1466.
YANG Senqi, DUAN Xuliang, XIAO Zhan, et al. Text classification of agricultural news based on ERNIE+DPCNN+BiGRU[J]. Journal of Computer Applications, 2023, 43(5): 1461-1466. (in Chinese with English abstract)

[37] 鲍彤，罗瑞，郭婷，等，基于 BERT 字向量和 Textcnn 的农业问句分类模型分析[J]. 南方农业学报，2022，

53（7）：2068-2076.
BAO Tong, LUO Rui, GUO Ting, et al. Agricultural question classification model based on BERT word vector and TextCNN[J]. Journal of Southern Agriculture, 2022, 53(7): 2068-2076 (in Chinese with English abstract)

[38] 金宁. 基于深度学习的农业短文本挖掘技术研究[D]. 沈阳：沈阳农业大学，2022.
ingJin N. Research on Agro-short Text Mining Method Using Deep Learning[D]. Shenyang: Shenyang Agricultural University, 2022 (in Chinese with English abstract)

[39] VEENA G, KANJIRANGAT V, GUPTA D. Agroner: An unsupervised agriculture named entity recognition using weighted distributional semantic model[J]. Expert Systems with Applications, 2023, 229: 120440.

[40] GUO X, LU S, TANG Z, et al. CG-ANER: Enhanced contextual embeddings and glyph features-based agricultural named entity recognition[J]. Computers and Electronics in Agriculture, 2022, 194: 106776.

[41] LIU Y, WEI S, HUANG H, et al. Naming entity recognition of citrus pests and diseases based on the BERT-BiLSTM-CRF model[J]. Expert Systems with Applications, 2023, 234: 121103.

[42] Hinton G E, MCCLELLAND J L, RUMELHART D E. Distributed representations[M]. Pittsburgh: Carnegie Mellon University, 1986: 77–109.

[43] MAO R, HE K, ZHANG X, et al. A survey on semantic processing techniques[J]. Information Fusion, 2024, 101: 101988.

[44] WANG J, DONG Y. Measurement of text similarity: A survey[J]. Information, 2020, 11(9): 421.

[45] WANG S, ZHANG S, SHEN Y, et al. Unsupervised deep structured semantic models for commonsense reasoning[EB/OL]. [2019-04-03]https://doi.org/10.48550/arXiv.1904.01938.

[46] CHENG Y, CHEN R, YUAN X, et al. Overview of long-form document matching: Survey of existing models and their challenges[C]//Journal of Physics: Conference Series, IOP Publishing, 2022, 2171(1): 012059.

[47] 吴华瑞，郭威，邓颖，等. 农业文本语义理解技术综述[J]. 农业机械学报，2022，53(5)：1-16.
WU Huarui, GUO Wei, DENG Ying, et al. Review of semantic analysis techniques of agricultural texts[J]. Transactions of the Chinese Society for Agricultural Machinery, 2022, 53(5): 1-16. (in Chinese with English abstract)

[48] WANG H, ZHU H, WU H, et al. A densely connected GRU neural network based on coattention Mechanism for Chinese rice-related question similarity matching[J]. Agronomy, 2021, 11(7): 1307.

[49] SUKUMAR R, HEMALATHA N, SARIN S, et al. Text based smart answering system in agriculture using RNN[C]// Proceedings of the 18th International Conference on Natural Language Processing (ICON), 2021: 663–669.

[50] REHMAN M Z U, RAGHUVANSHI D, KUMAR N. KisanQRS: A deep learning-based automated query-response system for agricultural decision-making[J]. Computers and Electronics in Agriculture, 2023, 213: 108180.

[51] AGARWAL O, GE H, SHAKERI S, et al. Knowledge graph based synthetic corpus generation for knowledge-enhanced language model pre-training[EB/OL]. [2021-05-13]https://doi.org/10.48550/arXiv.2010.12688.

[52] ZOU X. A survey on application of knowledge graph[C]//Journal of Physics: Conference Series, IOP Publishing, 2020, 1487(1): 012016.

[53] WANG T, HUANG R, WANG H, et al. Multi-Hop knowledge graph question answer method based on relation knowledge enhancement[J]. Electronics. 2023, 12(8): 1905.

[54] CHEN X, JIA S, YANG Y. A review: Knowledge reasoning over knowledge graph[J]. Expert Systems with Applications, 2020, 141: 112948.

[55] CHEN Z, WANG Y, ZHAO B, et al. Knowledge graph completion: A review[J]. Ieee Access, 2020, 8: 192435-192456.

[56] WANG Q, MAO Z, WANG B, et al. Knowledge graph embedding: A survey of approaches and applications[J]. IEEE Transactions on Knowledge and Data Engineering, 2017, 29(12): 2724-2743.

[57] 张宇，赵春江，林森，等. 基于 Penman-Monteith 模型和路径排序算法相结合的草莓灌溉方法与验证[J]. 智慧农业（中英文），2021，3(3)：116-128.
ZHANG Yu, ZHAO Chunjiang, LIN Sen, et al. Irrigation method and verification of strawberry based on Penman-Monteith model and path ranking algorith[J]. Smart Agriculture, 2021, 3(3): 116-128. (in Chinese with English abstract)

[58] 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.
TANG Wentao, HU Zelin. Survey of Agricultural Knowledge Graph[J]. Computer Engineering and Applications, 2024, 60(2): 63-76. (in Chinese with English abstract)

[59] 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 Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2022, 38(3): 112-118. (in Chinese with English abstract)

[60] 聂啸林，张礼麟，牛当当，等. 面向葡萄知识图谱构建的多特征融合命名实体识别[J]. 农业工程学报，2024，40(3)：201-210.
NIE Xiaolin, ZHANG Lilin, NIU Dangdang, et al. Multi-feature fusion named entity recognition method for grape knowledge graph construction[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2024, 40(3): 201-210. (in Chinese with English abstract)

[61] 姜丽华，赵瑞雪，董春岩，等. 基于深度学习的水产病害可视化知识图谱构建与验证[J]. 农业工程学报，2023，39(15)：259-267.

JIANG Lihua, ZHAO Ruixun, DONG Chunyan, et al. Construction and verification of the visual knowledge map of aquatic diseases based on deep learning[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2023, 39(15): 259-267. (in Chinese with English abstract)

[62] 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 Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2020, 36(24): 177-185. (in Chinese with English abstract)

[63] 郝志刚，刘冲，秦丽. 基于中文字邻接图的食品抽检公告实体及关系联合抽取[J]. 农业工程学报，2023，39(14)：283-292.
HAO Zhigang, LIU Chong, QIN Li. Entity and relationship joint extraction model of food inspection announcement based on Chinese character adjacency graph[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2023, 39(14): 283-292. (in Chinese with English abstract)

[64] 穆维松，刘天琪，苗子溦，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.
MU Weisong, LIU Tianqi, MIAO Ziwei, et al. Research progress on knowledge graph technology and its application in agriculture[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2023, 39(16): 1-12. (in Chinese with English abstract)

[65] SEN P, MAVADIA S, SAFFARI A. Knowledge graph-augmented language models for complex question answering[C]//ACL 2023 Workshop on Natural Language Reasoning andStructured Explanations, Toronto, Canada, 2023.

[66] PAN S, LUO L, WANG Y, et al. Unifying large language models and knowledge graphs: A roadmap[J]. IEEE Transactions on Knowledge and Data Engineering, 2024, 10(1): 1-20.

[67] COLON-HERNANDEZ P, HAVASI C, ALONSO J, et al. Combining pre-trained language models and structured knowledge[EB/OL]. [2021-02-05]https://doi.org/10.48550/arXiv.2101.12294.

[68] CHOUDHARY N, REDDYC K. Complex logical reasoning over knowledge graphs using large language models[EB/OL]. [2021-02-05] https://doi.org/10.48550/arXiv.2305.01157.

[69] PENG B, GALLEY M, HE P, et al. Check your facts and try again: Improving large language models with exter-nal knowledge and automated feedback[EB/OL]. [2023-03-08] https://doi.org/10.48550/arXiv.2302.12813.

[70] CHHETRI T R, HOHENEGGER A, FENSEL A, et al. Towards improving prediction accuracy and user-level explainability using deep learning and knowledge graphs: A study on cassava disease[J]. Expert Systems with Applications, 2023, 233: 120955.

[71] WANG H, SHEN W, ZHANG Y, et al. Diagnosis of dairy cow diseases by knowledge-driven deep learning based on the text reports of illness state[J]. Computers and Electronics in Agriculture, 2023, 205: 107564.

[72] HE J, Wang J, HE D, et al. The design and implementation of an integrated optimal fertilization decision support system[J]. Mathematical and Computer Modelling, 2011, 54(3/4): 1167-1174

[73] 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 Yin, YUAN Ye, et al. A decision support system for fertilization and irrigation management of citrus based on semantic ontology[J]. Transactions of the Chinese Society of Agricultural Engineering, 2014, 30(9): 93-101. (in Chinese with English abstract)

[74] NIZAR N M M, JAHANSHIRI E, THARMANDRAM A S, et al. Underutilised crops database for supporting agricultural diversification[J]. Computers and Electronics in Agriculture, 2021, 180: 105920.

[75] LI X, ZHANG H. Research on crop planting problem automatic answering system based on knowledge graph[C]//2023 IEEE International Conference on Image Processing and Computer Applications (ICIPCA). IEEE, 2023: 1-6.

[76] 杨硕，李书琴. 多模态知识图谱增强葡萄种植问答对的答案选择模型[J]. 农业工程学报，2023，39(14)：207-214.
YANG Shuo, LI Shuqin. Enhancing answer selection model of grape planting using multimodal knowledge graph[J]. Transactions of the Chinese Society of Agricultural Engineering, 2023, 39(14): 207-214. (in Chinese with English abstract)

[77] 谷刘涛. 基于知识图谱嵌入和路径推理的农业多跳问答模型研究[D]. 合肥：安徽农业大学，2023.
GU Liutao. Research on Multi-hop Question-Answering Model in Agriculture Based on Knowledge Graph Embedding and Path Reasoning[D]. Hefei: Anhui Agricultural University, 2023 (in Chinese with English abstract)

[78] XIE H, YANG J, HUANG C, et al. Recommendation algorithm for agricultural products based on attention factor decomposer and knowledge graph[C]//2022 Asia Conference on Algorithms, Computing and Machine Learning (CACML). IEEE, 2022: 626-631.

[79] ZOU Y, PAN S, YANG F, et al. Precise recommendation method of suitable planting areas of maize varieties based on knowledge graph[J]. Agriculture, 2023, 13(3): 526.

[80] LEI Z, HAQ A U, ZEB A , et al. Is the suggested food your desired?: Multi-modal recipe recommendation with demand-based knowledge graph[J]. Expert Systems with Applications, 2021, 186(6): 115708.

[81] WANG F, ZHU X, CHENG X, et al. MMKDGAT: Multi-modal knowledge graph-aware deep graph attention network for remote sensing image recommendation[J]. Expert Systems with Applications, 2024, 235: 121278.

[82] NIRANJAN P Y, RAJPUROHIT V S, SANNAKKI S S. Question answering system for agriculture domain using

machine learning techniques: literature survey and challenges[J]. International Journal of Computational Systems Engineering, 2020, 6(2): 91-99.

[83] JAIN N, JAIN P, KAYAL P, et al. AgriBot: Agriculture-specific question answer system[EB/OL]. [2019-01-17] IndiaRxiv Preprints, https://doi.org/10.35543/osf.io/3qp98.

[84] LI Y. A knowledge graph-based Q and A system for agricultural planting technology[C]//Second International Conference on Electronic Information Engineering and Computer Communication (EIECC 2022). SPIE, 2023, 12594: 235-240.

[85] LU G, LI S, MAI G, et al. AGI for Agriculture[EB/OL]. [2023-04-12] https://doi.org/10.48550/arXiv.2304.06136.

[86] 卢经纬，郭超，戴星原，等. 问答 Chatgpt 之后：超大预训练模型的机遇和挑战[J]. 自动化学报，2023，49（4）：705-717.
LU Jingwei, GUO Chaoxingyuan, DAI X, et al. The ChatGPT after: Opportunities and challenges of very large scale pre-trained models. Acta Automatica Sinica, 2023, 49(4): 705−717 (in Chinese with English abstract)

[87] WEI J, BOSMA M, ZHAO V Y, et al. Finetuned language models are zero-shot learners[EB/OL]. [2022-02-08] https://doi.org/10.48550/arXiv.2109.01652.

[88] SANH V, WEBSON A, RAFFEL C, et al. Multitask prompted train-ing enables zero-shot task generalization[EB/OL]. [2022-03-17] https://doi.org/10.48550/arXiv.2110.08207.

[89] CHUNG H W, HOU L, LONGPRE S, et al. Scaling instruction-finetuned language models[J]. Journal of Machine Learning Research, 2024, 25(70): 1-53.

[90] SHEN S, HOU L, ZHOU Y, et al. Flan-MoE: Scaling instruction-finetuned language models with sparse mixture of experts[EB/OL].[2023-05-01]https://doi.org/10.48550/arXiv.2305.14705.

[91] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]// Advances in Neural Information Processing Systems. 2017: 5998-6008.

[92] ZHANG H, SONG H, LI S, et al. A survey of controllable text generation using transformer-based pre-trained language models[J]. ACM Computing Surveys, 2023, 56(3): 1-37.

[93] 王琴，王鑫，颜靖柯，等. 融合空间位置注意力机制的英语题注生成模型[J]. 计算机工程与应用，2022，58(12)：139-148.
WANG Qin, WANG Xin, YAN Jingke, ZHONG M, et al. English caption generation model fused with attention mechanism of spatial position[J]. Computer Engineering and Applications, 2022, 58(12): 139-148. (in Chinese with English abstract)

[94] LIN T, WANG Y, LIU X, et al. A survey of transformers[J]. AI Open, 2022, 3: 111-132.

[95] LIU X, YU H F, DHILLON I, et al. Learning to encode position for transformer with continuous dynamical model[C]//International Conference on Machine Learning. PMLR, 2020: 6327-6335.

[96] CORDONNIER J B, LOUKAS A, JAGGI M. Multi-head attention: Collaborate instead of concatenate[EB/OL]. [2021-05-20], https://doi.org/10.48550/arXiv.2006.16362.

[97] SHAW P, USZKOREIT J, VASWANI A. Self-attention with relative position representations[EB/OL]. [2018-04-12] , https://doi.org/10.48550/arXiv.1803.02155.

[98] GHOJOGH B, GHODSI A. Attention mechanism, transformers, BERT, and GPT: tutorial and survey[J]. 2020. https://doi.org/10.31219/osf.io/m6gcn.

[99] HAO Y, DONG L, WEI F, et al. Self-attention attribution: Interpreting information interactions inside transformer[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2021, 35(14): 12963-12971.

[100] CHITTY-VENKATA K T, EMANI M, VISHWANATH V, et al. Neural architecture search for transformers: A survey[J]. IEEE Access, 2022, 10: 108374-108412.

[101] RADFORD A, NARASIMHAN K, SALIMANS T, et al. Improving language understanding by generative pre-training[EB/OL].[2021-05-23] https://arxiv.org/pdf/2304.04309v1.

[102] 哈尔滨工业大学. ChatGPT 调研报告[R/OL]. 2023-03-06. http：//wu-kongzhiku.com/hangyechanye/109956.html

[103] 张振乾，汪澍，宋琦，等，人工智能大模型在智慧农业领域的应用[J]. 智慧农业导刊，2023，3（10）：9-12,17.
ZHANG Zhenqian, WANG Shu, SONG Qi, et al. The application of artificial intelligence big models in the field of smart agriculture[J]. Journal of Smart Agriculture, 2023, 3(10): 9-12,17. (in Chinese with English abstract)

[104] BROWN T B, MANN B, RYDER N, et al. Language models are few-shot learners[J]. Advances in Neural Information Processing Systems, 2020, 33: 1877-1901.

[105] RADFORD A, WU J, CHILD R, et al. Language models are un-supervised multitask learners[J]. OpenAl Blog. 2019, 1(8): 9.

[106] ACHIAM J, ADLER S, AGARWAL S, et al. Gpt-4 technical report[EB/OL]. [2024-01-04]https://doi.org/10.48550/arXiv.2303.08774

[107] DU N, HUANG Y, DAI A M, et al. Glam: Efficient Scaling of language models with miture-of-experts CW PMLR[C]//International Conference on Machine Learning: PMLR, 2022: 5547-5569.

[108] TOUVRON H, LAVRIL T, IZACARD G, et al. LLaMA: Open and efficient foundation language models[EB/OL]. [2023-02-27]https://doi.org/10.48550/arXiv.2302.13971.

[109] DU Z, QIAN Y, LIU X, et al. GLM: General language model pretraining with autoregressive blank infilling[C]// Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers) Dublin, Ireland,2022: 320-335.

[110] What is A Vector Database?[Z/OL]. [2023-03-18]https://learn.microsoft.com/en-us/semantic-kerne1/memories/vector-db.

[111] HAN Y, LIU C, WANG P. A comprehensive survey on vector database: storage and retrieval technique, challenge[EB/OL].[2023-10-18] https://doi.org/10.48550/arXiv.2310.11703.

[112] GUO R, LUAN X, XIANG L, et al. Manu: A cloud native

vector database management system[EB/OL]. [2022-06-28] https://doi.org/10.48550/arXiv.2206.13843.

[113] WOODIE A. Vector databases emerge to fill critical role in AI[M/OL][2023-03-27]https://wwwdatanami.com/2023/03/27/vector-databases-emerge-to-fill-critic al-role-in-ai.

[114] FBNSM. Meet Norm, the World's First Al Ag Advi-sor[M/OL]. [2023-04-14]https://www. fbn. com/comm unity/blog/norm-first-ai-ag-advisor

[115] DARAPANENI N, TIWARI R, PADURI A R, et al. Farmer-bot: An interactive bot for farmers[EB/OL].[2022-04-07] https://doi.org/10.48550/arXiv.2204.07032.

[116] 赵春江. 农业知识智能服务技术综述[J]. 智慧农业（中英文）2023,5(2): 126-148.
ZHAO Chunjiang. Agricultural knowledge intelligent service technology: A review[J]. Smart Agriculture, 2023, 5(2): 126-148. (in Chinese with English abstract)

[117] OSINGA S A, PAUDEL D, MOUZAKITIS S A, et al. Big data in agriculture: Between opportunity and solution[J]. Agricultural Systems, 2022, 195: 103298.

Agricultural knowledge driven service technology innovation: Overview and frontiers

WANG Yuansheng[1,2] , WU Huarui[1,2] , ZHAO Chunjiang[1,2※]

(1. *National Engineering Research Center for Information Technology In Agriculture*, *Beijing* 100097, *China*; 2. *Information Technology Research Center*, *Beijing Academy of Agriculture and Forestry Sciences*, *Beijing* 100097, *China*)

Abstract: Agricultural knowledge-driven service technology (AKDST) can enable to generate natural language using promising artificial intelligence (AI). Various domains and formats of intelligent and personalized knowledge can be provided from crop production to food processing in the agricultural industry. AKDST can serve as a promising knowledge service to promote the quality and productivity in modern agriculture. This is also the main goal of current agriculture to fully meet the essential requirements of modern society. As such, there is great potential and opportunities for AKDST in the frontier of agricultural research. The whole process of technology development can be covered from the conception design, implementation, evaluation, application, and dissemination. Meanwhile, it is urgently necessary to require sufficient and efficient knowledge services in the agricultural industry at present. The current knowledge service can be improved to realize the short waiting time with low cost, high coverage, and accuracy. AKDST can be expected to translate the great progress in personalized and customized knowledge services. The most relevant and useful knowledge can also be found in the preferred modalities and formats, according to the needs and preferences. Especially, the advanced ChatGPT has been released to provide interactive and participatory knowledge services since November 2022. The large-scale pre-trained models can be potential for agricultural knowledge-intelligent services. The existing knowledge can be easily accessed to share the innovative technology. ChatGPT can serve as the prime example to generate fluent and coherent dialogues with technical support and feedback in the AKDST advancement. This review aims to analyze the current status and trend of AKDST-related technologies, and then prospect the potential of AKDST in the field of agriculture. Future research was also recommended to design and implement the large-scale pre-trained models. The more powerful and versatile AKDST was achieved in the large model, performance, and learning. In addition, the current mode of agricultural knowledge service was updated from the data retrieval, semantic matching, and the passive and static knowledge bases. Furthermore, technical support was combined with the agricultural machinery, information technology, agronomic practices, and communication channels for different components in the agricultural information system. Multimodal service was integrated with the text, image, voice, and video. The human-machine interaction was further enhanced suitable for human behaviors, habits, and cultures, considering human needs, preferences, emotions, human values, rights, and dignity. Technical support was also provided in the intelligence of agriculture, leading to the transformation from the agricultural knowledge service to the generative knowledge-driven mode. New knowledge was created using existing knowledge. Novel and diverse knowledge was output, such as summaries, explanations, suggestions, and evaluations.

Keywords: agricultural technology services; knowledge-driven; ChatGPT; large-scale pre-training model; new paradigm