

doi:10.6041/j.issn.1000-1298.2026.14.026

基于多教师蒸馏的猕猴桃种植知识图谱多跳问答模型

李书琴 杨晟沅

(西北农林科技大学信息工程学院, 陕西杨凌 712100)

摘要: 针对知识图谱多跳问答存在推理路径偏差导致推理性能下降的问题, 本文提出一种基于多教师蒸馏的猕猴桃种植知识图谱多跳问答模型。该模型由并行推理模型、混合推理模型和学生模型构成, 分别用于整合双向推理路径中的潜在信息、提供完整的上下文信息以及高效继承教师模型的推理能力。学生模型基于神经状态机实现, 通过置信度分布机制从多个教师模型及其中间推理过程中选择性学习信息, 并引入标准化的 Logits 函数动态调整蒸馏温度, 以平衡信息传递精度与模型学习效果。模型在自建的猕猴桃数据集 KiwiKG (Kiwi knowledge graph multi-hop dataset) 及 PQ (Pathquestion)、PQL (Pathquestion-large) 多跳数据集中开展试验, 试验结果表明: 该模型在 KiwiKG 上命中率 Hit@1 达到 90.92%, 与 EmbedKGQA 模型相比提高 1.82 个百分点; 在公开多跳数据集 PQ、PQL 也达到了最佳效果, 模型还兼顾推理性能与参数效率, 验证了其在多跳问答中的有效性与泛化能力。

关键词: 知识图谱; 猕猴桃; 多教师蒸馏; 多跳问答; 蒸馏温度

中图分类号: TP391.1 文献标识码: A 文章编号: 1000-1298(2026)14-0278-09 **OSID:** 

Multi-hop Question Answering Model for Kiwifruit Planting Knowledge Graph Based on Multi-teacher Distillation

Li Shuqin Yang Shengyun

(College of Information Engineering, Northwest A&F University, Yangling, Shaanxi 712100, China)

Abstract: Aiming to address the issue of inference performance degradation caused by reasoning path bias in multi-hop question answering over knowledge graphs, a novel model for multi-hop question answering in the domain of kiwifruit cultivation, was proposed based on multi-teacher knowledge distillation. The proposed model consisted of a parallel reasoning module, a hybrid reasoning module, and a student model, which were designed to integrate latent information from bidirectional reasoning paths, provide comprehensive contextual information, and efficiently inherit the reasoning capabilities of the teacher models, respectively. The student model was implemented by using a neural state machine, which selectively learned from multiple teacher models and their intermediate reasoning processes through a confidence-based distribution mechanism. Additionally, a normalized Logits function was introduced to dynamically adjust the distillation temperature, thereby balancing the precision of information transfer and the effectiveness of model learning. Experiments were conducted on the self-constructed kiwifruit dataset as kiwi knowledge graph multi-hop dataset (KiwiKG), as well as on public multi-hop question answering datasets as pathquestion (PQ) and pathquestion-large (PQL). The results demonstrated that the proposed model achieved a Hit@1 of 90.92% on KiwiKG, representing a 1.82 percentage point improvement over the EmbedKGQA model. Moreover, the model achieves state-of-the-art performance on the public datasets PQ and PQL. In addition to its strong reasoning capability, the model also maintained parameter efficiency, validating its effectiveness and generalization ability in multi-hop question answering tasks.

Key words: knowledge graph; kiwi; multi-teacher distillation; multi-hop question answering; distillation temperature

0 引言

近年来,随着猕猴桃种植规模不断扩大,农业生产对高效、精准的知识获取提出了更高要求^[1]。猕猴桃种植涉及病虫害防治、加工、质量评价等多方面复杂知识,且各领域信息存在高度异构性和关联性^[2]。知识图谱已在农业领域任务中展现出较好的应用潜力^[3-5],然而传统基于关键词或单跳推理的知识获取方式,已难以满足跨领域、多路径复杂推理的需求。因此研究面向猕猴桃种植知识图谱的多跳问答模型,对于推动领域智能问答系统发展具有重要意义。

知识图谱多跳问答作为解决复杂推理任务的核心挑战,其需通过关联多个三元组实现复杂信息推理,核心挑战在于有效捕捉多层次关系的逻辑链条^[6-7]。现有研究虽在实体关联方面取得进展,但对复杂的推理过程监督不足,导致推理路径易偏离正确方向^[8-9]。

传统方法依赖语义解析^[10-11]与信息检索技术,但前者易受误差累积影响推理精度,后者难以探索多跳路径的深层关联^[12]。针对这些问题,研究人员提出基于变分推理网络^[13]和键值记忆网络^[14]的方案优化知识表征能力,图卷积网络^[8]与图神经网络^[15]增强显式推理能力。强化学习框架也被应用于路径构建机制提升推理策略^[16-17],还有查询分解方法将复杂问题拆解为单跳子任务^[18-19]。尽管上述方法在性能上有所提高,但中间推理步骤监督信号的缺失仍是制约模型稳定性的关键瓶颈^[20]。

知识蒸馏技术将复杂教师模型的推理信息提供

给学生模型^[21-23],通过多教师蒸馏策略融合不同模型优势,如 Yang^[24]在 Web 问答中通过多教师知识融合提升推理能力,Li 等^[25]采用交互式双学生框架结合外部知识分解复杂问题,Shi 等^[26]将大模型生成-校正机制迁移至学生模型。

针对上述问题,本文提出多教师模型蒸馏的知识图谱多跳问答方法。首先识别问句汇总的关键实体,创建候选子图,使用并行推理教师模型和混合推理教师模型分别对候选子图进行推理,2 种教师模型从不同角度为中间推理步骤提供更可靠的实体分布。其次在知识蒸馏阶段,学生模型通过置信度分布的方法吸收教师模型的推理信息,并引入归一化 Logits 函数动态调整蒸馏温度,根据任务特点和模型需求平衡信息传递的精度与模型学习的有效性。最后,学生模型通过整合信息以生成推理答案。

1 模型结构

本文模型分为混合教师模型、并行教师模型以及学生模型 3 部分,整体架构如图 1 所示。模型对输入问句进行关键实体识别,并以此生成候选子图。随后,并行推理教师模型和混合推理教师模型分别进行计算,并通过预测特征与真实标签的相似度为教师预测标签分配权重,指导学生模型的学习。学生模型吸收教师模型的信息,并通过标准化的 Logits 函数动态调整蒸馏温度以适应不同任务需求,平衡信息传递精度与学习能力提升。最终,学生模型整合学习到的信息生成答案,实现高效的问答推理能力。

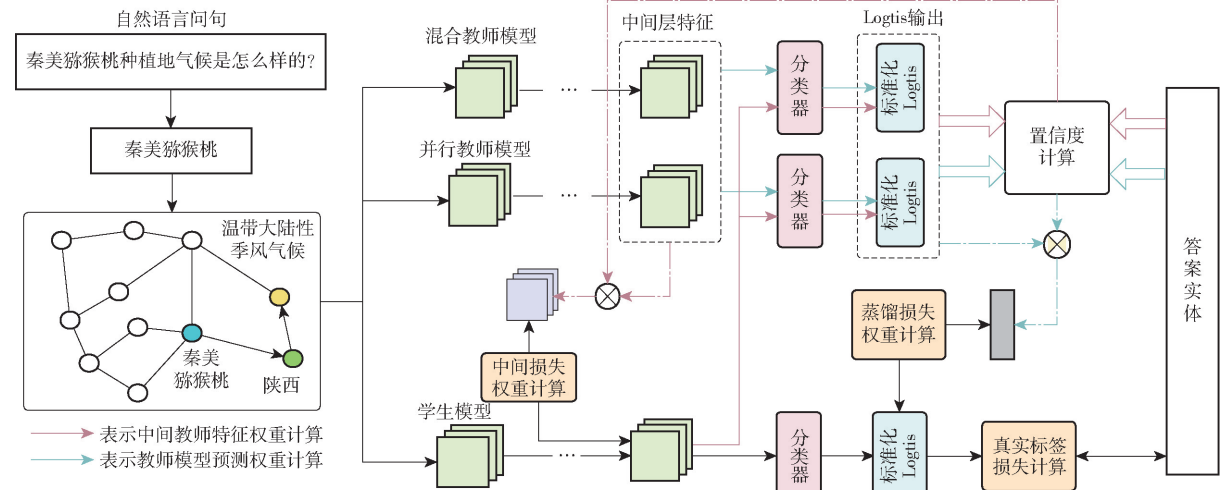


图 1 基于多教师蒸馏的多跳问答模型架构

Fig. 1 Multi-hop question answering model architecture based on multi-teacher distillation

1.1 学生模型

在猕猴桃种植知识图谱多跳问答模型中,学生

模型通过学习教师模型提供的中间实体分布,进行多跳推理并学习实体表示,从而提升问答准确性和

推理能力。学生模型基于神经状态机^[27]实现,实体分布和实体嵌入的学习与图神经网络的“传播-聚合”更新机制相一致。因此,神经符号机(Neural symbolic machines, NSM)可以在每一步推理时更新实体的表示,直到最终得到正确答案^[28]。NSM 主要由指令组件和推理组件组成。指令组件将指令向量发送到推理组件,而推理组件推断实体分布并学习实体表示。其整体结构如图 2 所示。

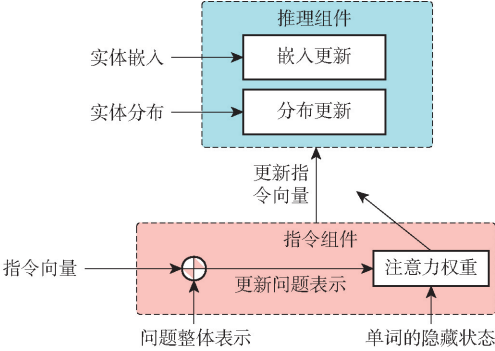


图 2 NSM 状态机模型结构

Fig. 2 NSM state machine model structure

指令组件将自然语言问题转换为一系列指令向量,以控制推理过程。首先使用 GloVe^[29]获取查询词嵌入,并通过 LSTM 编码器^[30]生成隐藏状态集合。最终隐藏状态被视为问题的表示 q 。知识图谱中三元组记为 $\{ \langle e, r, e' \rangle | e, e' \in E, r \in \mathcal{R} \}$, 式中 e 为头实体, r 为实体间关系, e' 为尾实体, $\mathcal{E}$ 和 $\mathcal{R}$ 分布是实体集和关系集。 $E \in \mathbf{R}^{d \times |E|}$ 和 $R \in \mathbf{R}^{d \times |R|}$ 是实体与关系的嵌入矩阵。

指令向量 $i^{(k)}$ 为指令嵌入向量,在第 k 个推理步骤通过公式可得到

$$i^{(k)} = \sum_{j=1}^l a_j^{(k)} h_j \tag{1}$$

其中 $a_j^{(k)} = \text{softmax}(W_a(q^{(k)} \odot h_j) + b_a)$ (2)

$$q^{(k)} = W^k [i^{(k-1)}; q + b^k] \tag{3}$$

式中 $\{h_j\}_{j=1}^l$ 是采用标准的 LSTM 编码器获得的一组隐藏状态, l 是查询的长度,注意力权重 $a_j^{(k)}$ 在不同时间步聚焦查询的特定部分, $q^{(k)}$ 为第 k 步的问句表示,第 k 步权重矩阵 $W^k \in \mathbf{R}^{d \times 2d}$, 注意力矩阵 $W_a \in \mathbf{R}^{d \times d}$, $b_a, b^k \in \mathbf{R}^d$ 是要学习的参数。每一步会动态更新查询表示,使其整合前一指令向量的信息,从而生成一系列指令向量 $\{i^{(k)}\}_{k=1}^n$ 。在获得指令向量 $i^{(k)}$ 后,可将其作为当前步骤指导信号。推理组件输入包括当前步骤指令向量、实体分布 $p^{(k)}$ 和上一步实体嵌入,计算公式为

$$e^{(0)} = \sigma \left(\sum_{(r, e') \in N_e} r W_T \right) \tag{4}$$

式中初始实体嵌入向量 $e^{(0)}$, $W_T \in \mathbf{R}^{d \times d}$ 为关系类型

变换矩阵参数, σ 为 sigmoid 激活函数,用于利用关系类型编码实体。

对于每一个三元组通过与关系向量 r 匹配计算匹配向量 $m_{(e', r, e)}^{(k)}$, 将相邻三元组的匹配信息进行聚合并赋予权重,并通过前馈层更新实体嵌入 $e^{(k)}$, 在第 k 步推理过程汇总并使用计算实体分布, 计算式为

$$m_{(e', r, e)}^{(k)} = \sigma(i^{(k)} \odot W_r) \tag{5}$$

$$\hat{e}^{(k)} = \sum_{(r, e') \in N_e} p_{e'}^{(k-1)} m_{(e', r, e)}^{(k)} \tag{6}$$

$$e^{(k)} = \text{FFN}([e^{(k-1)}, \hat{e}^{(k)}]) \tag{7}$$

$$p^{(k)} = \text{softmax}((E^{(k)})^T \omega) \tag{8}$$

式中 $\hat{e}^{(k)}$ ——第 k 步聚合后的实体表示

N_e ——实体 e 的邻居集合

$p_{e'}^{(k-1)}$ ——第 $k-1$ 步邻居实体 e' 的概率

$\text{FFN}(\cdot)$ ——前馈神经网络

$E^{(k)}$ ——包含第 k 步所有实体嵌入的矩阵

ω ——实体分布计算参数

1.2 教师模型

教师模型被训练在中间推理步骤提供监督信号,即任务中推断的实体分布以改进学生模型。为增强推理准确性,本文在 NSM 框架基础上引入双向推理机制,使正向与反向路径在中间步骤保持一致性,从而提升实体分布的稳定性与可靠性。教师模型可以在处理多跳知识图谱问答任务时,通过聚合涉及实体的关系嵌入来初始化节点嵌入,有效减少了噪声实体的影响,突出重要关系的语义特征^[25]。在更新实体嵌入时,教师模型对先前的嵌入和关系聚合嵌入进行统一更新,相较于普通 NSM 模型,能够更有效地推导出节点嵌入,从而提升推理性能。

因此,针对不同的推理特征使用并行推理教师模型和混合推理教师模型 2 种不同推理架构用于不同的教师模型。

1.2.1 并行推理教师模型

并行推理教师模型为正向推理和反向推理分别设置 2 个独立 NSM 网络,整体推理结构如图 3 所示。2 个 NSM 完全独立,彼此之间没有共享任何参数,旨在分别处理正向推理和反向推理任务。每个网络专注于其推理方向,以充分挖掘各自方向的潜在信息。为增强推理的准确性,在推理的中间步骤加入一致性约束,用于对正向推理生成的中间实体分布 $p_f^{(k)}$ 和反向推理生成的中间实体分布 $p_b^{(n-k)}$ 进行对齐和优化,即 $p_f^{(k)} \approx p_b^{(n-k)}$ 。这种一致性约束要求 2 个推理方向的中间分布相似,从而确保 2 个推理过程在探索路径时保持信息同步性。由于 2 个

网络之间没有参数共享,它们的交互限于一致性约束,而没有更深层的信息交换。因此,并行推理架构更倾向于松散整合,主要通过约束机制提高推理效果。

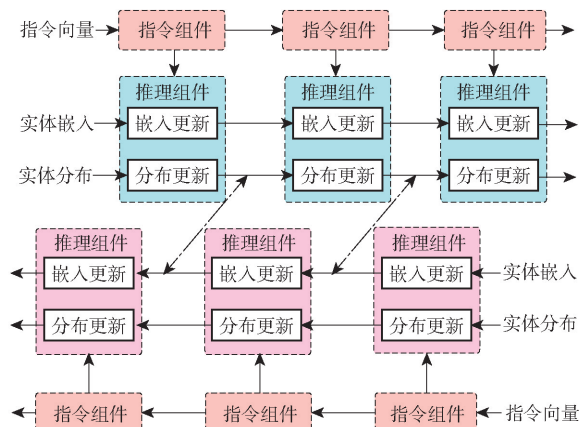


图 3 并行推理结构教师模型

Fig. 3 Parallel reasoning structure teacher model

1.2.2 混合推理教师模型

混合推理教师模型共享了正向推理和反向推理的指令组件,并将 2 个推理过程设计为一个循环流水线,整体推理结构如图 4 所示。在这种架构下,除了引入中间实体分布的一致性约束,还通过深度信息交互进一步增强推理效果。正向推理最终步骤的推导结果会被用作反向推理的初始信息,即正向推理的最终状态被循环利用,用于初始化反向推理。形式化的约束关系为

$$p_f^{(n)} = p_b^{(0)} E_f^{(n)} = E_b^{(0)} i_f^{(n+1-k)} = i_b^{(k)} \quad (k=1, 2, \dots, n) \quad (9)$$

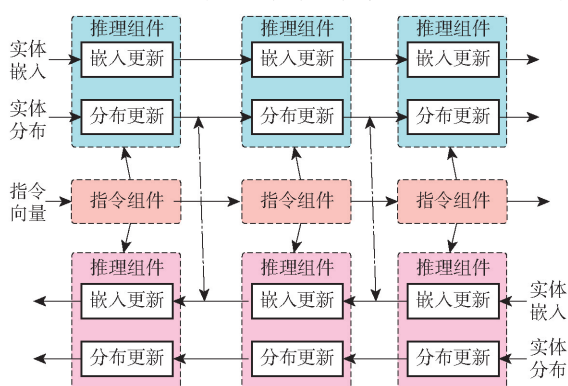


图 4 混合推理结构教师模型

Fig. 4 Hybrid reasoning structure teacher model

在这种方式下,反向推理不仅共享了正向推理的指令向量,还直接继承了正向推理的最终信息,从而能够更好地追溯正向推理的路径。这种信息交互能够弥补反向推理无法完全模仿正向推理逆向过程的不足。通过共享指令向量和循环利用最终状态,混合推理实现了更深层的信息交互,为反向推理提供了更丰富的上下文信息提升推理效果。

1.3 多教师蒸馏策略

多教师知识蒸馏并非通过整合来自多个教师的预测提高蒸馏的有效性。现有多种蒸馏方案,例如简单地为不同的教师分配平均或其他固定权重^[31-33],并根据梯度空间中的熵^[34]、潜在因素^[35]或多目标优化计算权重^[36]。然而,存在低质量预测的情况下,这些无标签策略可能会误导学生训练。例如,基于熵的策略会更喜欢盲目信任的模型,因为其有利于方差较低的预测;基于优化的策略有利于多数意见,易被噪声数据误导。

在每个样本上,学生模型和教师模型都会生成一个关于答案实体类别标签的概率分布,软标签损失函数的计算即为衡量 2 个分布之间的相似性^[37]。需要计算学生模型的预测输出与教师模型预测输出之间的交叉熵。最终模型的预测在应用 Softmax 和 Kullback-Leibler (KL) 散度之前对 Logits 进行标准化预处理^[38],将 Logits 映射到零均值和有限方差的范围内,可保留教师 Logits 的内在关系,而无需学生在数值上严格匹配教师。

为有效聚合多个教师模型预测分布,本文模型为每个教师分配不同权重 ω_{KD} ,这些权重反映它们在样本上的置信度,通过计算教师预测和真实标签之间的交叉熵损失 $\mathcal{L}_{CEKD}^k$ 获得,蒸馏损失可通过 2 个教师权重聚合得到,计算式为

$$\mathcal{L}_{CEKD}^k = - \sum_{c=1}^C y^c \ln(\sigma(z_{T_k}^c)) \quad (10)$$

$$\mathcal{L}_{KD} = -(\omega_{KD}^1 + \omega_{KD}^2) \sum_{c=1}^C z_{T_k}^c \ln(\sigma(z_S^c)) \quad (11)$$

其中

$$\omega_{KD}^k = 1 - \frac{\exp(Z(x; \tau)_k)}{\sum_{c=1}^K \exp(Z(x; \tau)_c)} \quad (12)$$

$$Z(x; \tau) = \frac{x - \bar{x}}{\tau \sigma(x)} \quad (13)$$

式中 T_k ——第 k 个教师模型

y^c ——第 c 类真实标签

$z_{T_k}^c$ ——学生模型对第 c 类的 Logit 输出

$Z(x; \tau)$ ——归一化后的 Logit 值

x ——原始 Logit 值

$\bar{x}$ ——Logit 均值

$\sigma(x)$ ——所有 Logit 的标准差

τ ——蒸馏温度

$\exp(Z(x; \tau)_c)$ ——第 k 个教师的蒸馏温度

$D = \{(x_i, y_i)\}_{i=1}^N$ ——输入的数据及其标签

N ——样本数量

$F \in \mathbf{R}^{h \times \omega \times c}$ ——最后一个模型网络结构的输出

$\mathbf{z} = [z^1 \quad z^2 \quad \cdots \quad z^C]$ ——Logit 输出矩阵
 C ——类别数
 $\mathcal{L}_{KD}$ ——蒸馏总损失
 z_s ——学生模型经第 k 个教师分类器后的
Logits

预测更接近真实答案路径的教师模型将被赋予更高的权重,因为其具有足够的置信度做出准确判断,能够提供准确的多跳推理路径指导,从而帮助学生模型更有效地完成复杂问答任务。相比之下,如果通过计算教师模型预测的熵获取权重,当输出分布较尖锐时,权重会变大,而不考虑最高概率类别是否正确,这些偏差较大的目标可能会误导学生模型推理结果,进而对模型问答性能产生负面影响。

中间层同样能够学习结构化知识^[34],因此将方法扩展至中间层以挖掘更多信息,并进行特征匹配计算。整体公式为

$$z_s \rightarrow T_k = W_{T_k} \mathbf{h}_s \tag{14}$$

$$\mathcal{L}_{CE_{inter}}^k = - \sum_{c=1}^C y^c \ln(\sigma(z_s^c \rightarrow T_k)) \tag{15}$$

$$\omega_{inter}^k = 1 - \frac{\exp(\mathcal{L}_{CE_{inter}}^k)}{\sum_j \exp(\mathcal{L}_{CE_{inter}}^j)} \tag{16}$$

式中, W_{T_k} 是第 k 个教师模型的最终分类器, $\mathbf{h}_s \in \mathbf{R}^C$ 表示学生模型的最后 1 层特征向量,通过将 $\mathbf{h}_s$ 传入每个教师模型的分​​类器后计算 $\mathcal{L}_{CEKD}^k$, ω_{inter}^k 为第 k 个教师的中间层蒸馏权重,作为衡量相似性的指标,表示教师模型分类器在学生特征空间中的区分能力。在模型汇总仅采用最后 1 个特征块的输出特征以避免带来过高的计算成本,能够更好地捕捉知识图谱中的多跳推理关系,有效增强了学生模型对教师模型中间层特征的学习能力,从而优化了知识迁移整体性能。

综上,中间层特征损失函数定义为

$$\mathcal{L}_{inter} = \sum_{k=1}^K \omega_{inter}^k \|F_{T_k} - r(F_s)\|_2^2 \tag{15}$$

其中, $\mathcal{L}_{inter}$ 用于对齐学生模型和教师模型特征维度, $\mathcal{L}_2$ 损失函数被用于衡量学生特征映射到教师特征空间后的结果 $r(F_s)$ 与中间层特征 F_{T_k} 之间的距离。

2 试验结果与分析

2.1 数据集

为确保猕猴桃知识图谱数据的专业性和精准性,采用《猕猴桃高效栽培》等专业书籍作为主要数据来源,通过 scrapy 爬取相关农业网站、农业专业百科等相关数据。将纸质书籍通过 OCR 技术转化为电子化图像格式;随后,对图像格式的书籍内容进行

人工校准,获得文本格式文档;对数据进行无效信息和重复信息删改,完成基本的数据清洗工作。

采用 BIESO 策略进行标注,BIESO 将文本中每个字符标注为特定类别,以识别文本中实体。“B”“I”“E”“S”“O”分别表示字符开始、内部、结束、单个字符,以及不属于任何实体的其他字符。最终提炼出猕猴桃种植环节所需的知识结构,涵盖猕猴桃的形态特征、生长习性、种植管理流程、病虫害防治、加工、评价标准等方面,具体内容如表 1 所示。

表 1 猕猴桃知识图谱类别定义及示例

Tab.1 Kiwi fruit knowledge graph categories, definitions and examples

实体类别	符号含义	类别定义	示例
VAR	品种	猕猴桃品种名称	美味猕猴桃
DIS	病害	遭受的病害	根腐病、疫霉病
PEST	虫害	受到的虫害	金龟甲类
PLAC	产地	猕猴桃产地	云南、广西
CLI	气候	不同区域气候	温带大陆气候
MED	防治药剂	处理病虫害的药剂	150 倍五氯酚钠
PMTH	种植方法	猕猴桃种植方法	手握采摘
PTIME	采摘时间	猕猴桃采摘时间	2、3 月
MMTH	加工方法	猕猴桃加工方法	猕猴桃发酵果酒
PRO	工艺流程	猕猴桃加工工艺流程	榨汁、过滤
CA	认证标准	国家猕猴桃行业标准	ISO 22000《食品安全管理体系 适用于食品链中各类组织的要求》
SA	安全标准	猕猴桃质检安全标准	GB 2762—2017《食品安全国家标准 食品中污染物限量》
QA	质量标准	猕猴桃等级	GB/T 32714—2016《猕猴桃质量等级》

共获得 12 468 条三元组,8 553 个实体,采用模板构建问句从中提取知识图谱中含有的所有两跳的实体和关系,采用模板构建问句共获得 5 013 条问答对,构建数据集 KiwiKG。为验证模型的泛化性,除自建数据集外还使用以 FreeBase 数据库为基础的多跳任务专用数据集 PQ、PQL^[39]进行模型训练和评估,数据集详情如表 2 所示。

表 2 数据集参数

Tab.2 Parameters of dataset

数据集	实体数	关系类型	三元组数	问答对数
KiwiKG	8 553	18	12 468	5 013
PQ-2H	1 056	14	1 211	1 908
PQ-3H	1 836	14	2 839	5 198
PQL-2H	5 034	363	4 247	1 594
PQL-3H	6 505	411	5 597	1 031

2.2 试验环境及模型参数

软件环境为 PyTorch 1. 10. 0,Python 3. 7,CUDA

11.4,支持 GPU 加速。在训练学生网络之前,在多跳推理任务上预训练教师网络。为避免过度拟合,训练过程中每 5 个训练轮次在验证集上评估一次 Hits@1 指标,采用早停策略。使用 Adam 优化器优化所有模型,批量大小为 40。学习率在 {0.01, 0.005,0.001,0.000 5,0.000 1} 之间进行调整,推理步骤设置为 3。

2.3 评价指标

使用 Hit@1 指标进行评估,Hit@1 衡量模型预测结果的前 1 个候选项中是否包含正确答案。为进一步衡量模型实用性,采用推理时间和模型参数量 2 个指标对模型效率进行评估。

2.4 结果与分析

将本文模型与 SRN、GraftNet、KV - Mem、EmbedKGQA、VRN 5 种在知识图谱多跳问答领域具有代表性的模型,分别在数据集 KiwiKG、PQ - 2H、PQ - 3H、PQL - 2H 和 PQL - 3H 进行测试。为证明本文模型合理性,通过消融试验验证教师模型设置以及蒸馏温度标准化的有效性。

2.4.1 不同模型性能测试

表 3 为不同模型性能测试结果,可以看到本文提出的模型在自建数据集 KiwiKG、PQ 多跳问答数据集、PQL 多跳问答数据集上均展现了更优越的性能。

表 3 不同模型性能测试结果
Tab.3 Experimental results of different models performance

模型	Hit@1/%					推理时间/s	参数量
	KiwiKG	PQ - 2H	PQ - 3H	PQL - 2H	PQL - 3H		
KV - Mem ^[14]	81.72	90.53	79.42	70.57	61.49	182.4	5.910 × 10 ⁶
VRN ^[13]	80.74	91.71	82.29	72.34	68.37	498.3	1.618 × 10 ⁷
GraftNet ^[8]	86.96	92.07	83.75	71.87	67.41	313.5	1.018 × 10 ⁷
SRN ^[40]	88.58	92.35	89.28	78.68	72.58	392.8	1.276 × 10 ⁷
EmbedKGQA ^[40]	89.10	92.52	90.32	79.30	74.43	207.6	6.740 × 10 ⁶
本文	90.92	93.98	91.37	80.69	75.95	231.2	7.510 × 10 ⁶

由表 3 可知,KV - Mem 模型表现较差,尤其在 PQL - 3H 数据集中 Hit@1 仅为 61.49%,因为该模型依赖于键值存储机制来处理知识图谱中的信息,在处理高维度和稀疏的知识图谱数据时存在困难,尤其在复杂多跳推理任务的处理能力上较差,易丧失关键信息造成推理结果的不准确;VRN 模型则引入了可变关系的概念来学习关系的动态变化,但该模型在处理节点之间变化不大的情况下会受语义关系影响而降低表现效果,在复杂的图结构中难以有效地选择正确推理路径,使得在猕猴桃知识图谱任务中 Hit@1 仅为 80.74%;GraftNet 在 PQ 多跳问答数据集和 KiwiKG 数据集中表现较好,但在更为复杂的 PQL 多跳问答数据集中的表现略有下降;SRN 模型表现相对优异,在多个数据集中其 Hit@1 相对 GraftNet 有明显提升,这一结果可以归因于 SRN 采用逐步推理的策略,虽然在简单的推理任务中能够充分捕捉关系信息,但在面对更复杂的多跳推理任务时,模型推理能力受到一定限制,导致其在 PQL - 3H 任务中表现不如其他模型;EmbedKGQA 模型通过嵌入知识图谱的信息,在所有任务中表现均衡且远高于其他模型,该模型的改进在于其能够有效地利用知识图谱中的信息进行多跳推理,从而在复杂任务中获得更好结果。

本文模型先对输入的问句进行关键实体识别,

并基于此生成候选子图。随后,并行教师模型和混合推理教师模型并行进行推理计算,通过预测特征与真实标签的相似度来为教师模型的预测标签分配权重,从而指导学生模型的学习。学生模型采用蒸馏策略吸收教师模型的信息,并通过标准化的 Logit 函数动态调整蒸馏温度,以满足不同任务需求。最终,学生模型整合推理路径中的信息生成答案,从而实现高效的问答推理能力。在自建猕猴桃数据集 KiwiKG Hit @ 1 达到 90.92%,较基线模型 EmbedKGQA 提升 1.82 个百分点;在 PQ 多跳问答数据集上也表现较优,在二跳和三跳问题上分别达到 93.98% 和 91.37%;而在更为复杂且难以推理的 PQL 数据集,较基线模型 EmbedKGQA 在 PQL - 2H 和 PQL - 3H 提升 1.39、1.52 个百分点。

从效率角度分析,推理时间和参数量在一定程度上反映模型在实际应用中的可部署性与计算开销。KV - Mem 模型结构较为简洁,采用键值存储机制实现信息检索,因此模型参数最少、推理速度较快,但由于缺乏深度语义建模能力,难以应对复杂的多跳推理任务;VRN 模型结构复杂,参数规模庞大,推理时间难以满足要求;SRN 在准确性上有所提升,但其推理开销大;EmbedKGQA 模型的设计上较为轻量,推理速度快。本文模型使用了蒸馏策略,在保持较小参数规模的同时,仅需较短推理时间即可完成

复杂多跳问答任务。试验结果表明,本文模型在保持较高准确率的同时,也在推理效率和模型规模上展现出较强优势,体现了良好的综合性能和实际应用潜力。

综合来看,本文模型在多个指标均展现了较强的竞争力,在处理猕猴桃知识图谱数据集的任务时,能够较其他模型提供更为精确和稳定的性能表现,且在 PQ、PQL 多跳问答数据集上的优秀表现也证明了本文模型具有较好的泛化性。

2.4.2 消融试验

为进一步分析使用并行教师模型、混合教师模型以及多教师模型蒸馏方法的有效性,进行消融试验,Hit@1 结果如表 4 所示。

表 4 不同教师模型消融试验结果
Tab.4 Ablation experiment results of different teacher models

模型	KiwiKG	PQ-2H	PQ-3H
本文	90.92	93.98	91.37
NSM _{+p}	89.47	90.53	87.20
NSM _{+h}	89.32	92.16	89.35
NSM	83.92	88.48	78.61

本文模型在 PQ-2H 和 PQ-3H 任务分别提升 1.82、2.02 个百分点,相较 NSM_{+p} 表现出更高精度。在 KiwiKG 任务上,本文模型同样相对 NSM_{+p} 模型提升 1.45 个百分点,相对 NSM_{+h} 模型提升 1.6 个百分点。这一结果表明,结合现有的多教师蒸馏模型,能够有效整合 2 种教师模型的优点,进一步提升模型的整体表现。

本文模型在猕猴桃数据集及 PQ 多跳数据集上均表现出了优异的性能,对比使用并行教师模型、混合教师模型和 NSM 模型,本文模型在所有指标都展现了明显的优势,证明多教师蒸馏机制的有效性。

2.5 案例分析

选取问句“秦美猕猴桃种植地的气候是怎么样的?”作为示例,对模型结果进行展示。问句正确回答为“温带大陆性气候”,正确推理路径为“秦美猕猴桃→陕西→温带大陆性气候”。

在推理执行过程中,学生模型由 2 类教师模型共同引导:混合教师模型提供丰富的上下文结构信息,增强路径整体语义一致性;并行教师模型则重点挖掘双向路径中的潜在推理线索,拓宽推理空间。在 2 类教师模型输出的候选路径指导下,学生模型对路径中间节点实施动态加权学习,形成更具鲁棒性的路径表示,并在多跳路径选择中实现有效的信息融合与推理动作选择。

为进一步提升学生模型对高质量路径的偏好能

力与整体泛化性能,引入置信度引导的选择性蒸馏机制,根据路径的置信度进行加权蒸馏优化,降低低效路径对中间实体分布学习的干扰。同时,通过标准化的 Logit 温度调控函数,使模型在处理不同长度路径与推理复杂度时能够灵活适配,实现精度与稳定性的平衡。在该实际案例中,模型准确推理出“秦美猕猴桃”的产地气候,验证了本文模型在多跳问答任务中的应用价值及其实用性。

如图 5 所示,针对主题实体“秦美猕猴桃”以图谱中实际构建的图结构为基础,首先识别出与主题实体直接相关的一跳候选节点,如“陕西省”“猕猴桃果汁”“斑衣蜡蝉”“9 月上旬”等,并进一步通过路径生成后续子图构建出包含两跳关系的潜在推理路径。具体而言,路径如“秦美猕猴桃→产地在→陕西省→气温是→温带大陆性季风气候”被模型赋予的置信度高达 0.946,表明学生模型能够有效进行多跳推理。

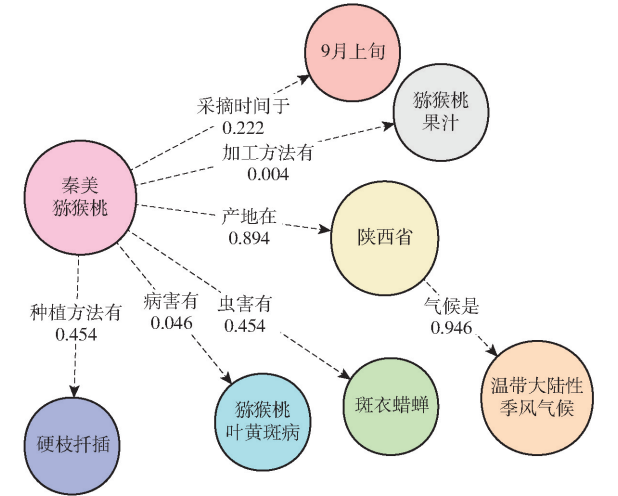


图 5 多跳推理过程示意图

Fig.5 Schematic of multi-hop reasoning process

3 结论

(1)面向猕猴桃种植场景,本文设计并行推理、混合推理 2 类教师模型;并行推理模型可双向挖掘推理隐含信息,混合推理模型能够充足上下文特征。学生模型借助 2 类教师模型,在多步推理过程中习得稳健实体分布,并基于教师置信度分布选择性蒸馏知识。

(2)将本文模型与主流的知识图谱领域多跳问答模型进行对比,验证本文模型在完成多跳问答的同时,与各数据集进行对比,在自建猕猴桃数据集 KiwiKG 及公开知识图谱多跳问答数据集 PQ、PQL 中均优于其他模型,在 KiwiKG 上 Hit@1 达到 90.92%,优于最佳基线模型 EmbedKGQA 1.82 个百分点;在公开数据集 PQ、PQL 中,2 跳子集 PQ-2H、

PQL-2H 的问答 Hit@1 分别达到 93.98% 和 93.42%, 相较 EmbedKGQA 分别提升 1.46、1.39 个百分点;3 跳子集 PQ-3H、PQL-3H 的 Hit@1 分别为 91.37%、90.81%, 相较 EmbedKGQA 分别提升 1.05、1.52 个百分点。表明本文模型在多跳问答任务中的有效性。本文模型使用了蒸馏策略,在保持较小参数规模的同时,仅需较短的推理时间即可完成复杂的多跳问答任务。

参 考 文 献

[1] 郭明月, 刘雅晨, 李伟夫, 等. 基于视频跟踪算法的果园猕猴桃产量实时预估[J]. 农业机械学报, 2023,54(6):178-185.
Guo Mingyue, Liu Yachen, Li Weifu, et al. Real-time estimation of orchard kiwifruit yield based on video tracking algorithms [J]. Transactions of the Chinese Society for Agricultural Machinery, 2023,54(6):178-185. (in Chinese)

[2] 李书琴, 张明美, 刘斌. 融合字词语义信息的猕猴桃种植领域命名实体识别研究[J]. 农业机械学报, 2022,53(12):323-331.
Li Shuqin, Zhang Mingmei, Liu Bin. Research on named entity recognition in kiwifruit planting field integrating word semantic information[J]. Transactions of the Chinese Society for Agricultural Machinery, 2022,53(12):323-331. (in Chinese)

[3] 侯琛, 牛培宇. 农业知识图谱技术研究现状与展望[J]. 农业机械学报, 2024,55(6):1-17.
Hou Chen, Niu Peiyu. Review of research status and prospects of agricultural knowledge graphs[J]. Transactions of the Chinese Society for Agricultural Machinery, 2024,55(6):1-17. (in Chinese)

[4] 李书琴, 高云帆. 基于知识图谱和语义与结构融合的羊病诊断方法[J]. 农业机械学报, 2025,56(7):575-584.
Li Shuqin, Gao Yunfan. Diagnosis method of sheep disease based on knowledge graph and fusion of semantics and structure [J]. Transactions of the Chinese Society for Agricultural Machinery, 2025,56(7):575-584. (in Chinese)

[5] 王博, 丁俊琦, 吴启锋, 等. 基于多源异构数据和知识图谱的作物病害关联分析[J]. 农业机械学报, 2025,56(12):546-559.
Wang Bo, Ding Junqi, Wu Qifeng, et al. Association analysis of crop diseases based on multi-source heterogeneous data and knowledge graph[J]. Transactions of the Chinese Society for Agricultural Machinery, 2025,56(12):546-559. (in Chinese)

[6] 吴天波, 周欣, 程军军, 等. 基于位置和注意力联合表示的知识图谱问答[J]. 计算机工程, 2022,48(8):98-104.
Wu Tianbo, Zhou Xin, Cheng Junjun, et al. Knowledge graph question answering based on joint representation of position and attention[J]. Computer Engineering, 2022,48(8):98-104. (in Chinese)

[7] 晋艳峰, 黄海来, 林沿铮, 等. 基于知识表示学习的 KBQA 答案推理重排序算法[J]. 计算机应用研究, 2024,41(7):1983-1991.
Jin Yanfeng, Huang Hailai, Lin Yanzheng, et al. KBQA answer reasoning re-ranking algorithm based on knowledge representation learning[J]. Computer Application Research, 2024,41(7):1983-1991. (in Chinese)

[8] Sun H, Dhingra B, Zaheer M, et al. Open domain question answering using early fusion of knowledge bases and text[C]// Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, 2018:4231-4242.

[9] Yih W, Richardson M, Meek C, et al. The value of semantic parse labeling for knowledge base question answering[C]// Meeting of the Association for Computational Linguistics, 2016:201-206.

[10] Berant J, Chou A, Frostig R, et al. Semantic parsing on freebase from question-answer pairs[C]// Proceedings of the Conference on Empirical Methods in Natural Language Processing, 2013.

[11] Chen L, Berant J, Le Q, et al. Neural symbolic machines: learning semantic parsers on freebase with weak supervision[C]// Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, 2017:22-23.

[12] Lan Y, He G, Jiang J, et al. Complex knowledge base question answering: a survey[J]. IEEE Transactions on Knowledge and Data Engineering, 2023,35(11):11196-11215.

[13] Zhang Y, Dai H, Kozareva Z, et al. Variational reasoning for question answering with knowledge graph[C]// Proceedings of the AAAI Conference on Artificial Intelligence, 2018:1533-1544.

[14] Miller A, Fisch A, Dodge J, et al. Key-value memory networks for directly reading documents[C]// Proceedings of the Conference on Empirical Methods in Natural Language Processing, 2016:1400-1409.

[15] Hu R, Rohrbach A, Darrell T, et al. Language-conditioned graph networks for relational reasoning[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019:10294-10303.

[16] Das R, Dhuliawala S, Zaheer M, et al. Go for a walk and arrive at the answer: reasoning over paths in knowledge bases using reinforcement learning[C]// International Conference on Learning Representations, 2018:10294-10303.

[17] Lin X, Socher R, Xiong C. Multi-hop knowledge graph reasoning with reward shaping[C]// Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, 2018:3243-3253.

[18] Min S, Zhong V, Zettlemoyer L, et al. Multi-hop reading comprehension through question decomposition and rescoring[C]// Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, 2019:6097-6109.

[19] Hudson D, Manning C. Compositional attention networks for machine reasoning[C]// Proceedings of the International Conference on Learning Representations, 2018.

[20] Zheng S, Chen W, Wang W, et al. Multi-hop knowledge graph reasoning in few-shot scenarios[J]. IEEE Transactions on

- Knowledge and Data Engineering, 2024,36(4):1713 – 1727.
- [21] Hinton G. Distilling the knowledge in a neural network[PP/OL]. V1. arXiv (2015-03-09)[2025-03-10]. <https://doi.org/10.48550/arXiv.1503.02531>.
- [22] Furlanello T, Lipton Z, Tschannen M, et al. Born again neural networks[C]//Proceedings of the 35th International Conference on Machine Learning,2018;1607 – 1616.
- [23] Zhang Y, Xiang T, Hospedales T M, et al. Deep mutual learning[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,2018;4320 – 4328.
- [24] Yang Z, Shou L, Gong M, et al. Model compression with two-stage multi-teacher knowledge distillation for web question answering system[C]//Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing,2020.
- [25] Li X, He S, Lei F, et al. Teaching small language models to reason for knowledge-intensive multi-Hop question answering[C]//Findings of the Association for Computational Linguistics, 2024;7804 – 7816.
- [26] Shi Z, Sun W, Gao S, et al. Generate-then-ground in retrieval-augmented generation for multi-hop question answering[PP/OL]. V2. arXiv (2024-09-16)[2025-03-10]. <https://doi.org/10.48550/arXiv.2406.14891>.
- [27] Hudson D, Manning C. Learning by abstraction: the neural state machine[C]//Proceedings of the 33rd Conference on Neural Information Processing Systems, 2019;5901 – 5914.
- [28] He G, Lan Y, Jiang J, et al. Improving multi-hop knowledge base question answering by learning intermediate supervision signals[C]//Proceedings of the 14th ACM International Conference on Web Search and Data Mining, 2021;553 – 561.
- [29] Pennington J, Socher R, Manning C D. Glove: global vectors for word representation[C]//Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing,2014;1532 – 1543.
- [30] Srivastava N, Mansimov E, Salakhudinov R. Unsupervised learning of video representations using LSTM[C]//Proceedings of the 32nd International Conference on Machine Learning, 2015;843 – 852.
- [31] Wu M C, Chiu C T, Wu K H. Multi-teacher knowledge distillation for compressed video action recognition on deep neural networks[C]//Proceedings of the 2019 IEEE International Conference on Acoustics, Speech and Signal Processing, 2019.
- [32] Fukuda T, Suzuki M, Kurata G, et al. Efficient knowledge distillation from an ensemble of teachers[C]//Proceedings of the 18th Annual Conference of the International Speech Communication Association,2017;3697 – 3701.
- [33] You S, Xu C, Xu C, et al. Learning from multiple teacher networks[C]//Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining,2017;1285 – 1294.
- [34] Kwon K, Na H, Lee H, et al. Adaptive knowledge distillation based on entropy[C]//Proceedings of the 2020 IEEE International Conference on Acoustics, Speech and Signal Processing,2020;7409 – 7413.
- [35] Liu Y, Zhang W, Wang J. Adaptive multi-teacher multi-level knowledge distillation[J]. Neurocomputing, 2020, 415: 106 – 113.
- [36] Du S, You S, Li X, et al. Agree to disagree: adaptive ensemble knowledge distillation in gradient space, online[C]//Proceedings of the 34th Conference on Neural Information Processing Systems,2020;12345 – 12355.
- [37] Zhang H, Chen D, Wang C. Confidence-aware multi-teacher knowledge distillation[C]//Proceedings of the 2022 IEEE International Conference on Acoustics, Speech and Signal Processing,2022;4498 – 4502.
- [38] Sun S, Ren W, Li J, et al. Logit standardization in knowledge distillation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,2024;15731 – 15740.
- [39] Zhou M, Huang M, Zhu X. An interpretable reasoning network for multi-relation question answering[C]//Proceedings of the 27th International Conference on Computational Linguistics,2018;2010 – 2022.
- [40] Qiu Y, Wang Y, Jin X, et al. Stepwise reasoning for multi-relation question answering over knowledge graph with weak supervision[C]//Proceedings of the 13th ACM International Conference on Web Search and Data Mining,2020;474 – 482.