

融合子图信息和规则路径的不确定知识图谱补全模型^①



蒋娅娅, 李贯峰, 王世卓, 尹仁蕊

(宁夏大学 信息工程学院, 银川 750021)

通信作者: 李贯峰, E-mail: ligf@nxu.edu.cn

摘要: 随着不确定知识图谱 (uncertain knowledge graph, UKG) 在开放世界知识表示中的广泛应用, 其内部所建模的概率型关系日益复杂. 仅依赖嵌入表示或简单的模式匹配方法, 已难以满足对高质量推理结果的需求. 因此, 引入高置信、可解释的规则挖掘机制, 对于提升 UKG 的推理能力与知识可解释性具有重要意义. 为此本文提出了一种基于 BERT 语义建模与结构路径规则挖掘的不确定知识图谱补全模型 UBERT-RM (uncertain knowledge graph BERT-rule mining). UBERT-RM 构建了一个端到端的统一框架, 将子图建模、路径生成与置信度预测有机融合于一体. 模型利用 BERT 提取三元组的上下文语义表征, 路径生成模块采用 Transformer 解码器结构, 以自回归方式逐步生成高置信度的关系路径. 规则解析模块中引入动态置信度阈值机制, 对节点进行筛选与解析, 确保最终的推理路径在保持语义连贯的同时具备良好的可信度与可解释性. 在置信度预测部分, 模型将生成的规则与目标三元组共同编码, 通过多头自注意力机制进行深层语义交互与信息聚合, 并引入多层感知机实现对尾实体置信度的回归建模, 从而完成从路径挖掘到置信度预测的闭环推理过程. 在 CN15k 和 NL27k 数据集上的实验结果表明, UBERT-RM 在链接预测任务中的效果都达到了最佳.

关键词: 不确定知识图谱补全; 语义建模; 规则路径; 链接预测

引用格式: 蒋娅娅, 李贯峰, 王世卓, 尹仁蕊. 融合子图信息和规则路径的不确定知识图谱补全模型. 计算机系统应用. <http://www.c-s-a.org.cn/1003-3254/10123.html>

Uncertain Knowledge Graph Completion Model Integrating Subgraph Information and Rule Path

JIANG Ya-Ya, LI Guan-Feng, WANG Shi-Zhuo, YIN Ren-Rui

(School of Information Engineering, Ningxia University, Yinchuan 750021, China)

Abstract: With the widespread application of uncertain knowledge graph (UKG) in open-world knowledge representation, the probabilistic relations they model have become increasingly complex. Sole reliance on embedding representations or simple pattern-matching methods is now insufficient to meet the demand for high-quality reasoning results. Therefore, the introduction of high-confidence, interpretable rule mining mechanisms is essential for enhancing both the reasoning capability and knowledge interpretability of UKGs. To this end, an uncertain knowledge graph completion model named UBERT-RM (uncertain knowledge graph BERT-rule mining) is proposed, which integrates BERT-based semantic modeling with structural path rule mining. UBERT-RM constructs a unified end-to-end framework that organically combines subgraph modeling, path generation, and confidence prediction. Contextual semantic representations of triples are extracted using BERT, while the path generation module employs a Transformer decoder structure to autoregressively generate high-confidence relation paths. In the rule parsing module, a dynamic confidence threshold mechanism is introduced to filter and parse nodes, ensuring that the final reasoning paths remain semantic

^① 基金项目: 国家自然科学基金 (62066038); 宁夏自然科学基金 (2024AAC03098); 宁夏全职引进高层次人才科研启动项目 (2023BSB03066)

收稿时间: 2025-09-28; 修改时间: 2025-11-03; 采用时间: 2025-11-17; csa 在线出版时间: 2026-03-02

coherence while achieving high reliability and interpretability. For confidence prediction, the generated rules and target triples are jointly encoded, followed by deep semantic interaction and information aggregation through a multi-head self-attention mechanism. A multilayer perceptron is then employed to perform regression modeling of the tail-entity confidence, thereby completing a closed-loop reasoning process from path mining to confidence prediction. Experimental results on the CN15k and NL27k datasets demonstrate that the proposed UBERT-RM achieves state-of-the-art performance on the link prediction task.

Key words: uncertain knowledge graph completion (UKGC); semantic modeling; rule path; link prediction

在构建开放世界知识库的过程中,知识图谱(knowledge graph, KG)作为一种结构化的知识表示形式,已广泛应用于问答系统、推荐系统与推理等任务中^[1]. 传统的确定性知识图谱通常以三元组(h, r, t)的形式表示事实,其中 h 和 t 表示头实体和尾实体, r 表示连接两者的关系. 目前主流的公开确定性知识图谱有Freebase^[2]、DBpedia^[3]、YAGO^[4]和Wikidata^[5]等. 然而,在实际应用场景中,尤其是在自动化信息抽取和众包构建过程中,获得的知识往往带有不确定性,存在噪声或语义模糊等问题. 为应对这一挑战,不确定知识图谱(uncertain knowledge graph, UKG)应运而生,与传统知识图谱不同,UKG在每个三元组中附加了一个置信度分数,用以刻画该知识事实成立的概率,从而更好地支持在不完全或含噪信息下的智能推理与决策. 目前已有多个代表性UKG数据集被广泛研究和应用,如NELL^[6]和ConceptNet^[7]等.

在开放世界设定下,即使是规模非常大的KG也难以覆盖全部知识内容,因此KG中存在很多缺失的事实,很多实体之间的隐含关系没有被充分挖掘,导致KG的结构和内容不完整^[8]. 因此,研究者们提出了KG补全研究,通过预测现有实体之间的潜在关系和发现新的关系事实来补全KG并扩大其规模^[9]. 现有的KG补全方法可以分为两类:基于嵌入的方法和基于规则的方法^[10]. 基于嵌入的方法通过将高维的离散数据映射到低维连续空间中,利用参数化表示学习构建实体与关系的分布式语义表征,继而设计适配KG内在代数或几何结构的评分函数,通过距离度量或张量分解等数学框架量化三元组的语义合理性. 代表性模型包括TransE^[11]、DistMult^[12]和Complex^[13]等. 这类方法在效率和可扩展性方面表现优异,但由于其计算过程难以解释,往往在推理透明性方面存在不足. 相比之下,基于规则的方法以图结构中的路径与模式为依据,归纳出一系列逻辑规

则用于三元组预测. 其关键在于对知识图谱进行系统遍历提炼规则,并据此完成推断和补全. 规则挖掘的代表方法有AMIE^[14]和AnyBURL^[15]等. 这类方法在每一次的推理过程中都对一条显式的规则链条,使得结果不仅可追溯,也便于人工验证和知识回溯. 这种特性对于可信人工智能和知识驱动系统至关重要,尤其适用于对逻辑因果结构有严格要求的场景,如医疗、法律、文化遗产等领域. 尽管规则方法在可解释性方面具有天然优势,但其性能与适用性仍受到计算开销、语义建模能力和数据稀疏性的多重制约. 近年来,pLogicNet^[16]、RNNLogic^[17]和Rule^[18]等工作尝试融合神经表示与符号逻辑,以兼顾推理的表达能力与可解释性,成为KG补全的一种发展趋势. 然而,上述研究大多聚焦于确定性KG,不能直接应用到UKG中.

为应对UKG中复杂的概率关系与语义不确定性,本文提出了一种名为UBERT-RM(uncertain knowledge graph BERT-rule mining)的UKG补全框架,UBERT-RM由3个相互协同的关键模块构成:首先,通过置信度加权聚合机制构建邻域感知的实体表示,有效捕捉局部图结构中的语义与置信度信息;其次,设计了基于交叉注意力的自回归路径生成模块,该模块能够动态整合历史路径序列与头实体子图信息,生成高可信度的关系推理链;最后,引入基于统计特性的动态阈值筛选与逆向解析机制,实现高质量规则路径的自适应挖掘. 实验结果表明,UBERT-RM通过深度融合路径结构、图语义与置信度信息,在不确定知识环境下显著提升了推理性能与可解释性,为知识图谱补全研究提供了新的思路.

本文的贡献可以总结为以下4点.

(1) 提出了一种端到端的不确定知识图谱补全模型UBERT-RM,该模型在统一框架下融合了子图信息、路径推理与基于规则的置信度预测机制. 通过在

路径推理过程中引入 BERT 结构, 模型能够同时建模实体的上下文语义与推理路径特征, 从而显著增强对复杂概率关系的表征能力与推理性能。

(2) 在路径生成过程中引入显式交叉注意力机制, 将已生成关系序列与头实体上下文表示联合建模, 有效提升了生成路径的结构合理性与语义一致性。同时提出基于统计阈值的筛选与逆向贪心解析算法, 自适应地挖掘高置信规则, 增强补全过程的可解释性。

(3) 将规则与三元组作为共同输入, 通过自注意力交互机制融合语义与路径信息, 采用多层感知机进行置信度预测, 构建路径引导下的实体得分矩阵, 实现从路径生成到尾实体预测的闭环优化。

(4) 验证了模型在两个 UKG 数据集上的优越性能: 在 CN15k 和 NL27k 两个不确定知识图谱数据集上进行的研究表明, UBERT-RM 在链接预测与置信度估计两项核心任务中均显著优于现有主流基线模型。

1 相关工作

1.1 确定性知识图谱补全

传统的补全方法主要依赖于构建知识嵌入, 将实体和关系嵌入低维向量空间, 通过优化特定的评分函数确定三元组的可信度^[19]。嵌入模型可分为翻译类模型(如 TransD^[20])、张量分解类模型(如 DualE^[21])和卷积神经网络模型(如 ConEx^[22])。与传统方法不同, 图神经网络^[23]通过邻域聚合的方式建模实体和关系间的复杂交互, 使模型具备跨多跳推理的能力。典型的模型有图卷积网络模型^[24]、注意力图网络模型^[25]以及结合预训练语言模型的图谱补全方法。近年来, 预训练语言模型与知识图谱不断融合, 形成了以文本增强、联合训练与生成式补全为代表的多种新范式。Yao 等人^[26]提出的 KG-BERT 模型利用 BERT 从自由文本中提取丰富语义模式, 将知识图谱补全任务转化为序列分类问题, 通过将三元组视为单个序列, 计算其合理性。BERTL^[27]将 KG 中的局部子图线性化为推理路径, 并与目标三元组结合输入到 BERT 中进行微调。

基于规则的知识图谱补全通过挖掘和利用 KG 中隐含的逻辑规则来推导新的三元组, 从而实现 KG 的扩充与完善。与基于嵌入的方法相比, 基于规则的方法具有更好的可解释性和透明度, 能够提供推理过程的清晰逻辑链, 便于人类理解和验证。如 RuleFormer^[28]是基于 Transformer 架构的规则挖掘方法。RUGE^[29]通过

迭代过程将软规则指导知识图谱嵌入。但由于它们不能捕获知识的不确定性, 因此这些基于规则的方法不能直接应用于 UKG 补全。

1.2 不确定知识图谱补全

在 UKG 研究中, 现有研究方法大多采用嵌入方法进行补全, 模型 UKGE^[30]在经典的知识图谱嵌入模型上引入置信度建模机制, 通过设计特定的损失函数, 使模型预测出的三元组置信度尽可能贴近其真实置信度标签。同时在训练过程中引入模糊逻辑规则, 通过规则推理生成额外三元组, 以优化嵌入。BEUrRE^[31]采用盒嵌入的方式, 通过建模实体之间的边界及其重叠区域来表达实体的不确定性与包容关系, 并通过逻辑约束项提升语义一致性。GMUC^[32]则引入高斯混合机制, 刻画关系中可能存在的不确定性分布, 提升了模型对复杂语义的表示能力。SUKE^[33]先利用概率软逻辑推理未知事实, 并以这些推理得到的事实来训练评估器; 同时, 通过双向量表示和双模块架构解耦不同信息流, 从而在不确定知识条件下稳定提升知识图谱补全性能。MUKGE^[34]提出 URR 算法整合全局不确定信息, 并利用循环相关嵌入区分非对称关系, 在不确定知识图谱的嵌入与推理中实现了更高精度和更强的表达能力。

以上方法都是基于嵌入的补全, 缺乏可解释性。基于规则的方法因其形式化的逻辑表示能力和良好的可解释性, 近年来受到研究者的重视, 其中, 代表方法 UKRM^[35]利用 Transformer 解码器生成关系路径, 实现了对 KG 中多跳推理路径的建模, 有效提升了模型在复杂结构下的推理能力。然而, 该方法主要侧重于结构层面的路径模式构建, 缺乏对语义上下文信息的显式建模与融合, 导致其所生成规则在语义覆盖度和泛化能力方面存在局限。

2 问题定义

为了更好地刻画事实的可靠性, UKG 在传统三元组的基础上附加一个置信度分数, 从而形成四元组: $(h, r, t, s), s \in [0, 1]$, 其中 h 表示头实体, r 表示关系, t 表示尾实体, s 表示 (h, r, t) 成立的概率。这种形式使得 UKG 能够在含噪、不完全和语义模糊的环境下, 更真实地表示世界知识。

在本文中, 不确定知识图谱补全 (uncertain knowledge graph completion, UKGC) 任务可以定义为: 在给

定部分已知的四元组集合的情况下,对查询\$(h,r,?,?)\$进行预测,推理出潜在尾实体\$t\$,并同时估计其置信度分数\$s\$.进一步引入一阶逻辑规则作为推理依据,一阶逻辑规则可由关系路径逆向解析得到,其一般形式为:

$$r_1(X,Z_1) \wedge r_2(Z_1,Z_2) \wedge \dots \wedge r_k(Z_{k-1},Y) \Rightarrow r_h(X,Y) \quad (1)$$

其中,\$k\$是规则的长度,规则左侧是规则体,规则右侧是规则头,\$X\$、\$Y\$和\$Z_i\$是可以被特定实体替换的变量,以此来得到规则主体,继而得到规则头.

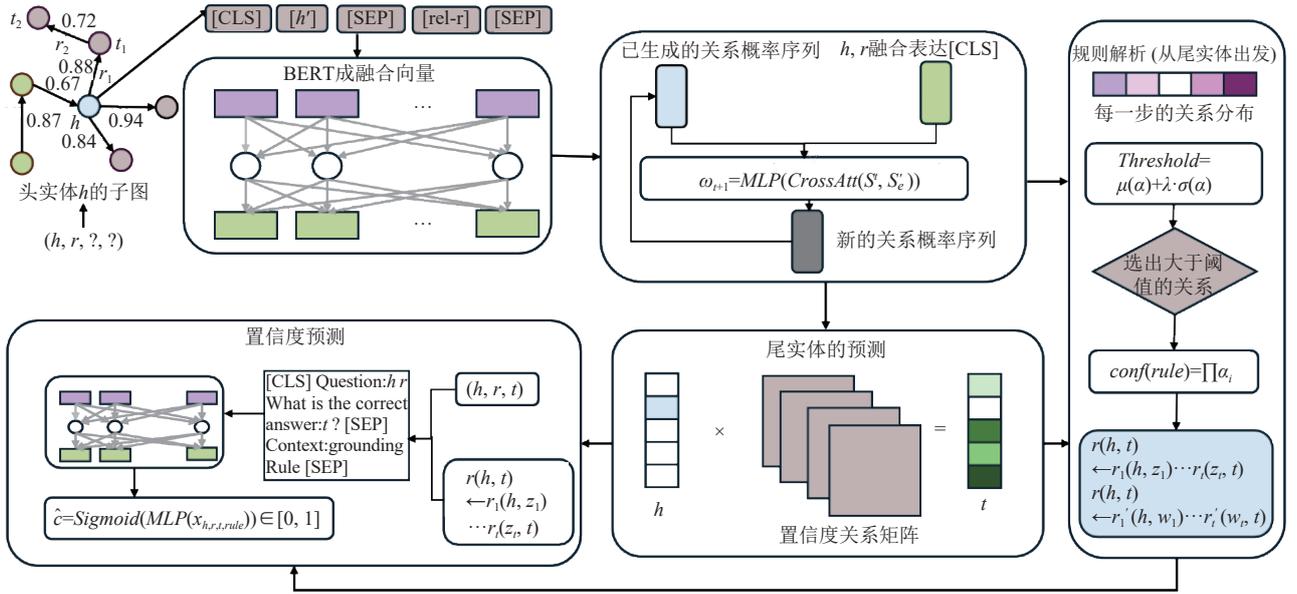


图1 UBERT-RM 模型图

3.2 查询语义建模

为了更充分地捕获查询三元组中头实体的上下文信息并兼顾关系的不确定性,本文设计了一种基于子图的实体表示方法.对于每个查询,模型首先从图中提取以头实体\$h\$为中心的\$k\$跳邻居子图,记为\$G_h\$.该子图包含所有\$h\$与在\$k\$跳范围内连接的实体节点及对应边信息,用以提供头实体的局部环境.

在 UKG 中,每条边都伴随有置信度值以量化其存在的可信程度.为提升局部结构质量,在构建子图时,模型引入置信度阈值机制,仅保留置信度不低于 0.7 的边,以筛除潜在噪声边.由此得到的子图既包含了高置信关系,又增强了头实体对其可靠邻居的感知能力.不同于将子图节点序列化输入 BERT,本文在子图层面先融合邻居信息,生成上下文增强的头实体表示\$h'\$.具体定义如下:

$$h' = \sum_{(r,h) \in N_{in}(h)} \alpha_r^{in} r_{ran} + \sum_{(h,r) \in N_{out}(h)} \alpha_r^{out} r_{dom} + y_h \quad (2)$$

3 UBERT-RM 模型

3.1 模型概述

本文提出了一种基于编码器-解码器的端到端图谱补全模型——UBERT-RM.该模型整体架构如图1所示,该模型包括4个核心模块:查询语义建模、规则路径生成、规则路径解析与置信度预测,旨在通过结构信息、语义上下文与置信度的深度融合,提升推理性能和可解释性.

其中,\$N_{in}(h)\$和\$N_{out}(h)\$分别为指向外出和入内到头实体\$h\$的高置信度关系集合,\$r_{ran}\$和\$r_{dom}\$表示关系\$r\$的range和domain嵌入,\$y_h\$为头实体自身的个体特征嵌入.模型引入关系嵌入,显式编码结构的方向性与语义角色.\$\alpha_r^{in}\$和\$\alpha_r^{out}\$分别为出边和入边方向上关系的归一化置信度权重,其计算方式为:

$$\alpha_r^{in} = \frac{conf(r,h) \cdot n_r^{in}}{\sum_{r_j \in N_{in}(h)} conf(r_j,h) \cdot n_{r_j}^{in}} \quad (3)$$

$$\alpha_r^{out} = \frac{conf(h,r) \cdot n_r^{out}}{\sum_{r_j \in N_{out}(h)} conf(h,r_j) \cdot n_{r_j}^{out}} \quad (4)$$

其中,\$conf(r,h)\$和\$conf(h,r)\$分别对应边的置信度,\$n_r^{in}\$与\$n_r^{out}\$为关系作为入边、出边在子图中出现的次数.\$N_{in}(h)\$与\$N_{out}(h)\$为子图中指向实体\$h\$的所有入边关系和出边关系集合.模型以置信度对邻居边的贡献加权,使高置信度关系集合.

信边在实体表示中占据更大权重;同时采用由置信度与边频次共同驱动的联合归一化,平衡不同关系对表示的影响,缓解因关系数量或置信度分布不均造成的偏置,提升表示的稳健性与多样性.

3.3 关系路径生成器

路径生成模块旨在给定查询对的前提下,动态生成连接潜在尾实体的多跳关系序列,从而构建可解释的推理路径.该模块基于 Encoder-Decoder 框架实现,其中编码器利用预训练语言模型 BERT 提取查询语义,解码器则采用标准 Transformer 架构实现路径的逐步生成.

在编码器部分,首先将头实体 h' 和关系 r 的标记拼接,构造 BERT 输入序列: $[\text{CLS}], h', [\text{SEP}], r, [\text{SEP}]$, 并送入 BERT 经过多层双向自注意力机制编码,得到上下文增强的全局表示 $[\text{CLS}]$ 向量.在生成路径阶段,解码器以 $[\text{BOS}]$ 作为初始输入,结合编码器输出的 $[\text{CLS}]$, 递归生成关系路径.在上述子图聚合得到上下文增强的头实体表示的基础上,该阶段将查询语义 $[\text{CLS}]$ 作为语义先验输入至解码器的 Cross-Attention 中,驱动关系分布 p_t 的生成; p_t 再通过可谓传播作用于子图结构,形成了由“语义-结构”的前向链路.随后,链接预测损失对尾实体分布 u_t 的梯度将经由 p_t 到查询语义向量反传至解码器并进一步影响 BERT 的交叉注意力机制,使得结构到语义的闭环得以建立.

在每个时间步 t , 解码器首先使用 Masked Self-Attention 来聚合已生成的关系序列:

$$\text{SelfAtt}(r \leq t-1) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d}} + M\right)V \quad (5)$$

其中, M 为 mask matrix. Q 、 K 、 V 分别是已生成关系嵌入序列,并经过线性变换后的查询、键和值向量矩阵.在获得已生成路径的上下文表示后,解码器通过 Cross-Attention 融合编码器输出的全局语义向量预测下一个关系 r_t .具体步骤如下:

$$\omega_t = \sigma\left(\text{MLP}\left(\text{CrossAtt}\left(S'_r, S'_e\right)\right)\right) \quad (6)$$

其中, S'_r 表示解码器当前已生成的关系序列片段, S'_e 为编码器的输出.每一步选取概率最高的关系作为当前时间步生成的关系,并作为下一个时间步的输入 token.该过程持续进行,直到达到最大路径长度 k .

在得到每一步关系的概率分布后,模型基于可微矩阵推理机制对尾实体进行预测,以保证推理路径的可

解释性.头实体被表示为一个 one-hot 向量 e_h .对于推理路径中的每一步,第 t 步的关系生成概率分布为 α_t ,对于置信度关系矩阵集合为 M_j ,其中每个 $M_j \in [0, 1]^{|E| \times |E|}$ 表示关系 r_j 在 KG 中的置信度矩阵,即 $M_j[i, j]$ 表示 (e_i, r_j, e_j) 三元组的置信度, E 为图谱中实体数量.在每一步中,模型对所有关系矩阵按生成概率加权求和:

$$M_t = \sum_j \alpha_{t,j} M_j \quad (7)$$

其中, M_t 即为第 t 步的综合关系矩阵.在推理过程中,连续地将每一步的综合关系矩阵与前一步的实体权重相乘,逐步传播实体权重.最终的尾实体预测过程表示为:

$$Y = \left(\prod_{t=1}^T \sum_j \alpha_{t,j} M_j \right) e_h \quad (8)$$

其中, $Y \in \mathbb{R}^{|E|}$ 是连续值向量,每个元素 Y_i 表示头实体 e_h 通过当前生成路径推导到实体 e_i 的推理权重.这一过程是在沿生成路径连续叠加各个三元组置信度信息,从头实体逐步将置信度传递至候选尾实体.最终,权重最大的实体即为模型预测的尾实体,同时推理路径可溯源至每一步的关系概率分布和对应的关系矩阵,实现了符号逻辑推理的可解释性.为了促使模型将更高的权重分配给真实尾实体,本文采用负对数形式的损失函数,通过最大化目标实体在预测向量中的得分,提升推理的准确性.具体的损失如式 (9) 所示:

$$\text{loss}_{\text{UBERT-RM}} = -\log\left(\max(\gamma, Y_{\text{target}})\right) \quad (9)$$

3.4 规则解析器

在 UBERT-RM 中,为提升规则的可解释性和推理路径的有效性,本文设计了一个逆向动态阈值筛选的规则解析器.该模块基于解码器输出的每一步关系生成概率分布以及最终生成的尾实体分布,进行逆向解析,从而提取出一组高置信、有效的符号规则,用于后续的置信度预测.具体而言,解析过程从尾实体出发,逐层回溯生成路径.在第 t 层,设 $h_t \in \mathbb{R}^{|E|}$ 为当前实体的权重向量, $R_j \in \mathbb{R}^{|E| \times |E|}$ 为关系 r_j 的置信度邻接矩阵,则该关系对目标实体传播的结构贡献可定义为:

$$c_{t,j}[u] = R_j h_t \quad (10)$$

其中, $c_{t,j}[u]$ 表示通过关系 r_j 逆向回溯时,头实体 u 对当前目标实体集合的传播强度,综合反映了图结构连通性与不确定边置信度的作用.在此基础上,为结合结构贡献与解码器的语义偏好,本文进一步引入解码器在

第 t 步生成关系 r_j 的概率 $p_t(j)$, 并定义联合评分:

$$g_{t,j} = p_t(j) \cdot \phi(c_{t,j}) \quad (11)$$

其中, $\phi(\cdot)$ 为对向量 $c_{t,j}$ 的聚合函数. 该联合评分既衡了解码器语义预测下关系被选中的可能性, 又反映了关系在不确定图结构中对目标传播的实际贡献, 从而实现语义与结构的有效融合. 我们引入动态阈值机制, 与传统固定阈值相比, 本文的动态阈值机制具备自适应性和统计动机. 该设计为一种基于概率分布形态的风险平衡策略: 当生成概率分布标准差较大时, 说明某些关系置信度远高于均值, 分布离散, 有明显的高置信关系, 模型可以适当地提高阈值. 反之说明关系置信度集中, 分布平滑, 可以适当地降低阈值, 提高召回率, λ 控制筛选宽松度. 该机制也等价于在低计算开销下实现了按分布动态调节风险地贝叶斯决策近似. 具体做法是: 基于当前关系概率分布, 计算其置信度均值和标准差, 并据此得到阈值 τ , 其计算公式如下:

$$\tau_t = \mu(\omega_t) + \lambda \cdot \sigma(\omega_t) \quad (12)$$

其中, $\mu(\omega_t)$ 为生成概率均值, 体现的是当前生成概率分布的整体平均水平, $\sigma(\omega_t)$ 为标准差, 体现的是各关系偏离均值的程度. 从理论上讲, 当该阈值由解码器当前步的概率分布 p_t 计算, 因此它直接依赖于 BERT 提供的语义上下文信息; 被阈值筛选后的高置信度关系又影响规则解析与置信度回归模块的输入, 从而形成语义到结构、再到置信预测的协同链路.

为了在训练阶段实现可微的关系筛选, 让阈值机制能参与端到端反向传播. 本文设计了软门控开关:

$$m_t(r) = \sigma\left(\frac{\omega_t(r) - \tau_t}{\alpha}\right) \quad (13)$$

其中, $\omega_t(r)$ 是解码器在第 t 步对关系 r 输出的归一化分数, 软门控 $m_t(r)$ 是一个介于(0,1)的连续权重, 它衡量了关系 r 的激活强度. 动态阈值 τ_t 决定当前时间步的选拔标准. 软门控 $m_t(r)$ 决定每条候选关系是否、多大程度通过这个门. 它是 Sigmoid 平滑近似的通行率. 两者协同形成动态筛选机制. 动态阈值负责感知整体不确定性, 软门控负责平滑执行筛选决策.

当开始逆向解析时, 模型从尾实体出发, 解析器在每一层根据动态阈值与软门控筛选出高置信度的关系和中间实体, 沿逆向路径逐步扩展直至回溯到头实体. 最终完整解析出的规则以显式的关系序列和实体序列

形式存储, 并基于路径中各三元组置信度对规则整体置信度进行计算:

$$\text{conf}(\text{rule}) = \prod_{(e_i, r, e_j) \in \text{path}} M_r[e_i, e_j] \quad (14)$$

3.5 置信度评估器

在完成规则解析和路径推理后, 为进一步量化规则推理生成三元组的置信度, 本文设计一个基于 BERT 的置信度预测模块. 模型首先将推理生成的三元组及其对应的推理路径拼接成自然语言提示序列, 并结合预训练语言模型对其置信度进行回归预测.

具体而言, 该模块首先将三元组 (h, r, t) 与对应规则拼接为模板化输入序列, 示例如图 1 中的置信度预测模块中的构造格式. 该序列输入 BERT 编码器后, BERT 在每一层的交互中, 利用序列首 token 的[CLS]与其他 token 进行自注意力交互, 实现信息聚合. 通过逐层聚合机制, [CLS]向量融合了整个序列的上下文信息. BERT 输出的[CLS]表示作为置信度预测的全局特征表示, 输入至多层感知机中, 回归得到置信度分数:

$$c = \sigma(\text{MLP}(x_{h,r,t,\text{rule}})) \quad (15)$$

其中, $x_{h,r,t,\text{rule}}$ 表示 BERT 聚合后的序列嵌入, σ 表示 Sigmoid 激活函数, 用以保证置信度输出范围为[0,1]. 训练时, 模型最小化预测置信度与真实置信度标签之间的均方误差, 损失函数定义为:

$$\text{loss} = \frac{1}{n} \sum_{i=1}^n (c - c^*)^2 \quad (16)$$

其中, c 是来自本模型置信度预测部分得到的置信度, c^* 是三元组置信度的真实值.

4 实验结果与分析

4.1 实验设置

实验在 Windows Server 2022 Standard 操作系统下进行, 硬件配置使用的 CPU 为 Intel Xeon Gold 6248R CPU@3.0 GHz, 使用的 GPU 为 NVIDIA A100-PCIE-40 GB, 软件框架是基于 PyTorch 1.12.1 的深度学习统一训练框架, 开发语言为 Python.

4.1.1 数据集

为评估 UBERT-RM 模型在不确定性知识图谱补全任务中的效果, 本文选取两个公开基准数据集: CN15k 与 NL27k, 这两个数据集均是在 2019 年由 Chen 等人^[30]

提出的不确定知识图谱表示学习数据集. CN15k 的数据来源于常识知识图谱 ConceptNet^[7], 其中包含 241 158 条不确定三元组, 每条三元组的置信度分数由人工众包标注产生, 原置信度分数区间为[0.1, 22], 构造后的三元组置信度落在[0, 1]区间之内. NL27k 的数据来源于持续开放网页抽取知识图谱 NELL^[6], 其中包含 175412 条不确定三元组, 图谱中的置信度由抽取系统 EM 算法迭代产生, 并辅以人工众包标注方法进行调整, 原始置信度分数区间为[0.1, 0.9], 构造后的三元组置信度落在[0.1, 1]区间之内. 表 1 总结了数据集相关的内容.

表 1 数据集统计结果

统计项	CN15k	NL27k
实体	15000	27221
关系	36	404
训练集	204984	149001
验证集	16881	12278
测试集	19293	14034

实验在 CN15k 和 NL27k 两个数据集上进行, 分别代表概念性 KG 与实例性 KG. 本文将三元组划分为: 85% 用于训练, 10% 用于测试, 5% 用于验证. 如表 1 所示, CN15k 包含 15 000 个实体和 36 种关系, NL27k 则涵盖 27 221 个实体和 404 种关系.

4.1.2 对比模型

在与本文模型进行对比之前, 本文将对对比模型分为 4 类, 以全面覆盖不同类型的知识图谱补全方法. 第 1 类为确定性知识图谱表示学习方法, 包括经典的 TransE^[11]和 RotatE^[36]这两种模型均在结构建模方法具有良好的泛化能力, 用于验证传统方法在图谱结构语义建模方面的性能. 第 2 类为不确定性嵌入模型, 涵盖 UKGE^[30]、BEUrRE^[31]、MUKGE^[34]、SUK E^[33]、UKGsE^[37]以及 PASSLEAF^[38], 这些方法代表了当前基于嵌入方法的 UKG 补全技术. 第 3 类为基于 UKG 的

规则学习模型, 包括 UKRM^[35]. 第 4 类为 UBERT-RM 模型以及 3 个消融变体, 分别为 UBERT w/o BERT、UBERT w/o rule 和 UBERT w/o graph, 旨在进一步探究 UBERT-RM 中语义建模模块和规则路径引导机制的有效性贡献.

4.2 实验对比与分析

4.2.1 链接预测

链接预测任务旨在对给定的不完整查询三元组 $(h, r, ?)$ 进行补全, 预测最可能的尾实体列表. 本文采用了排名方式对候选尾实体排序, 最终评估模型在知识推理中的检索性能.

本文在该任务中使用了标准的评估指标: 平均倒数排名 MRR 和 Hits@N ($N \in \{1, 3, 10\}$). MRR 用来衡量正确实体在所有候选实体中平均的排名情况, Hits@N 表示正确尾实体排在前 N 的比例, 两个指标越高表明模型的推理性能越优.

(1) 参数设置

在 BERT 编码过程中, 本文使用 HuggingFace 提供的 BERT-base-chinese 作为实体对查询的编辑器, 编码器层数为 12, 最大序列长度设为 64. BERT 输出维度为 768. 我们采用网格搜索策略在验证集上进行超参数调优, 学习率 lr 为 $\{0.0005, 0.001, 0.01\}$, 优化器采用 Adam. 为确保与 BERT 解码器的一致性, 解码器的维度设为 768, 批大小为 $\{64, 128, 256\}$. 实验最佳超参数为 $lr=0.001$, 批大小为 128.

(2) 实验结果分析

实验结果如表 2 所示. 可以观察到, 所提出的 UBERT-RM 模型在 CN15k 与 NL27k 两个不确定知识图谱数据集上几乎在所有指标上均超越现有主流对比方法, 尤其在 MRR 与 Hits@1 等核心指标上表现尤为突出, 验证了引入语义上下文建模与置信度机制后的规则生成方法的有效性. 具体分析如下.

表 2 链接预测结果

模型	CN15k				NL27k			
	MRR	Hits@1	Hits@3	Hits@10	MRR	Hits@1	Hits@3	Hits@10
TransE	0.115	0.035	0.137	0.277	0.261	0.270	0.397	0.555
RotatE	0.126	0.042	0.148	0.297	0.297	0.349	0.484	0.600
UKGE	0.142	0.104	0.168	0.209	0.440	0.295	0.358	0.588
BEUrRE	0.103	0.063	0.137	0.248	0.308	0.284	0.493	0.560
UKGsE	0.010	0.002	0.008	0.020	0.057	0.042	0.082	0.108
PASSLEAF	0.122	0.062	0.159	0.262	0.363	0.412	0.532	0.626
UKRM	<u>0.189</u>	<u>0.142</u>	<u>0.221</u>	<u>0.339</u>	<u>0.632</u>	<u>0.522</u>	<u>0.631</u>	<u>0.743</u>
UBERT-RM	0.194	0.156	0.243	0.347	0.649	0.541	0.652	0.773

注: 最优结果用粗体显示, 次优结果用下划线显示

首先,在 CN15k 数据集上,与嵌入式方法 UKGE 相比,UBERT-RM 在 MRR 和 Hits@10 上分别取得了 36.6% 和 66% 的性能提升;其中 Hits@1 从 0.104 提升至 0.156,增幅达 50%,显示出 UBERT-RM 能够显著提升尾实体预测的精度,弥补结构稀疏导致的信息缺失。其次,与当前基于规则路径的代表模型 UKRM 相比,UBERT-RM 在 CN15k 上的 MRR 和 Hits@10 分别提升 2.6% 和 2.4%,而 Hits@1 指标从 0.142 提升至 0.156,增长 9.9%。这说明 UBERT-RM 不仅在语义建模方面更具优势,其引入的动态阈值机制与结构矩阵加权推理策略,也进一步增强了规则路径的判别性与多样性。

在结构复杂、实体规模更大的 NL27k 数据集中,UBERT-RM 同样表现出极强的鲁棒性。相较于 UKGE 模型,其在 MRR 和 Hits@10 上分别提升了 47.5% 和 31.4%,其中 Hits@1 从 0.295 提升至 0.541,增幅为 83.3%,验证了 UBERT-RM 对复杂推理路径的建模能力和高召回场景下的性能稳定性。与 UKRM 模型的对比中,UBERT-RM 在 NL27k 上的 MRR、Hits@1 和 Hits@10 指标分别提升 2.7%、3.6% 和 4%,说明在引入 BERT 编码与置信度感知机制后,UBERT-RM 能够更加精准地捕捉三元组的上下文语义,并生成更可靠的推理路径。

综合来看,UBERT-RM 在链接预测任务中的出色表现可归因于以下 3 点改进机制:(1) 基于 BERT 的查询语义建模有效提升了实体关系表示能力,缓解了传统结构方法在弱监督场景下的稀疏性问题;(2) 路径生成结合 Transformer 解码器,可动态生成可解释的多跳关系路径;(3) 融合动态阈值机制与结构推理权重,提升了路径筛选的鲁棒性与可解释性。综上所述,UBERT-RM 不仅在预测性能上全面超越对比模型,也在模型设计上实现了从语义到结构、从生成到评估的深度融合,为不确定知识图谱下的规则学习与实体预测提供了更加可靠与可解释的范式。

4.2.2 置信度预测

在不确定性知识图谱中,三元组本身就带有概率性,因此除了链接预测外,更核心的任务是对已生成或观察到的三元组进行可信度估计。本任务采用均方误差 (MSE) 与平均绝对误差 (MAE) 作为性能评估指标,MSE 衡量预测值与真实值之间误差的平方的平均,用于反映整体预测偏差的大小;MAE 则计算预测值与真

实值之间误差的绝对值平均,相较于 MSE 对异常值不敏感,更具稳定性和直观解释性。这两个指标越小,表明模型预测出的置信度越接近真实标注,性能越优。

(1) 参数设置

该编码器采用与主模型相同的 BERT-base-Chinese 权重初始化,但参数独立训练。我们设定最大输入长度为 128,输出维度为 768,Dropout 设置为 0.1,优化器采用 Adam。

(2) 实验结果分析

实验结果如表 3 所示,在 UKG 中的置信度预测任务上,UBERT-RM 模型在 CN15k 和 NL27k 两个数据集上均显著优于现有主流方法,在 MSE 和 MAE 两项指标上均取得最优结果,表现出更强的预测精度与稳定性。具体而言,在 CN15k 数据集上,UBERT-RM 的 MSE 与 MAE 分别达到 0.03941 和 0.16653,相较于基线模型 UKGE 与 PASSLEAF,分别在 MSE 上减少了 80.3% 与 57.9%,在 MAE 上减少了 54.0% 与 32.9%。相较于 UKRM,MSE 也有所降低。同样地,在 NL27k 数据集上,UBERT-RM 的 MSE 和 MAE 分别为 0.01367 和 0.04991,相较于表现次优的 UKRM,MSE 降低了 16.5%,MAE 降低了 9.8%。这些结果充分说明 UBERT-RM 在复杂图谱结构下依然能够精准建模三元组的置信度标签。

表 3 置信度预测结果

模型	CN15k		NL27k	
	MSE	MAE	MSE	MAE
UKGE	0.19988	0.36206	0.02861	0.05967
BEURRE	0.11656	0.28259	0.08920	0.22194
PASSLEAF	0.09375	0.24801	0.01949	0.06253
UKGsE	0.10274	0.25564	0.12202	0.27605
MUKGE	0.08250	0.20960	0.02690	0.08220
SUKE	0.05120	<u>0.17820</u>	0.07700	0.31900
UKRM	<u>0.04456</u>	0.18765	<u>0.01638</u>	<u>0.05532</u>
UBERT-RM	0.03941	0.16653	0.01367	0.04991

注:最优结果用粗体显示,次优结果用下划线显示

4.2.3 消融实验

为深入探究 UBERT-RM 模型中各关键模块对整体性能的贡献,本文设计了 3 组消融实验,分别从语义建模能力与结构规则引导机制两个角度进行剖析。具体地,我们构造了 3 个变体模型:UBERT-RM w/o BERT、UBERT-RM w/o graph 和 UBERT-RM w/o rule。w/o BERT 移除原模型中用于对查询头实体及其关联关系进行编码的 BERT,直接将实体关系对的表示输入到解码器进

行预测任务,旨在评估预训练语义建模模块对实验结果的影响。w/o graph 移除了输入编码阶段对查询头实体邻居节点信息的结构融合操作,即不再构造以头实体为中心的子图上下文特征,直接对原始的实体-关系对进行编码。该设置旨在评估结构上下文在丰富查询表示、提升路径生成质量的贡献。w/o rule 在置信度预测阶段不再引入规则,仅对三元组进行编码,旨在评估规则路径在三元组可信度建模中的作用。

从表4所示链接预测结果可知,在CN15k和NL27k两个数据集上,当移除BERT编码器后,模型在Hits@10上分别下降了5.6%和17.5%,MRR分别下降了19.6%和18.8%,这说明BERT编码器在建模实体上下文语义方面具有显著作用,能够为规则路径生成提供更稳定且语义一致的输入表示,从而提升尾实体预测的精度;当去除实体子图融合模块后,CN15k上Hits@10与MRR分别下降了6.9%与6.2%,NL27k上则下降了19.7%与8.6%,表明实体邻居子图所提供的结构上下文信息对增强实体表示与路径生成具有重要意义,尤其在结构更复杂的大型图谱中,该影响更为显著。

表4 消融实验链接预测结果

模型	CN15k		NL27k	
	Hits@10	MRR	Hits@10	MRR
w/o BERT	0.291	0.156	0.638	0.527
w/o graph	0.323	0.182	0.621	0.593
UBERT-RM	0.347	0.194	0.773	0.649

进一步在表5所示置信度预测任务中,若移除语义编码器,模型在CN15k上的MSE从0.03941上升至0.07231,MAE从0.16653上升至0.22463;在NL27k上,MSE与MAE分别上升了43.7%与16.3%,表明语义表示不仅提升了路径生成质量,同时也增强了三元组置信度回归的准确性与稳定性;而移除规则路径模块,在CN15k上MSE与MAE分别上升了47.1%和19.0%,在NL27k上分别增加了54.0%与22.4%,进一步印证了引入规则路径所带来的结构约束有助于提升可信度估计的判别能力与可解释性。

表5 消融实验置信度预测结果

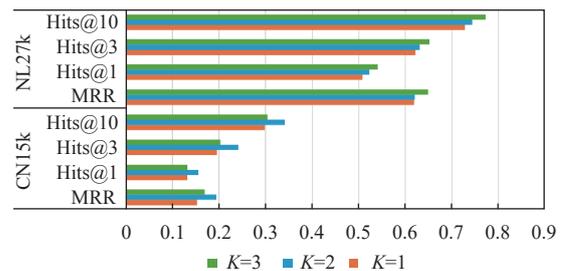
模型	CN15k		NL27k	
	MSE	MAE	MSE	MAE
w/o BERT	0.07231	0.22463	0.01964	0.05807
w/o rule	0.05796	0.19821	0.02105	0.06108
UBERT-RM	0.03941	0.16653	0.01367	0.04991

综上,UBERT-RM在语义建模、结构引导与规则融合机制上的协同设计对于实现高质量推理与精确置

信预测起到了关键作用,充分说明所提出各模块在任务性能中是不可缺少的。

4.2.4 敏感性参数实验

为了进一步分析模型在关键超参数下的性能表现,本文针对子图选取跳数 K -hop和路径采样跳数 T 进行分析。在子图选取跳数的实验中,本文在 $K \in \{1, 2, 3\}$ 这3种跳数范围内进行对比实验,探讨邻居信息采集范围内对路径生成于置信度预测性能的影响。可以从图2中看到,在CN15k数据集中,当 $K=1$ 时,实体上下文信息不足,路径生成受限,导致推理能力较弱。当 $K=3$ 时,尽管上下文更丰富,但引入的噪声边增多,训练开销上升,同时部分低置信度边会对表示带来干扰。当 $K=2$ 时,MRR取得较优的平衡,既保证了上下文信息的完整性,又有效抑制了噪声。在NL27k数据集上,模型表现出不同的趋势。由于该数据集的实体与关系种类更多、连接更稠密,较小跳数下获取的信息仍不足以覆盖完整语义上下文。随着跳数的增加,模型能获取更全面的结构信息,进而提升路径生成的质量和置信预测的准确性。因此,在NL27k上,MRR随着跳数的增长而稳步提升,表明模型在大规模稠密图中对扩展邻域的利用更为显著。综上,不同跳数对模型性能的影响具有数据依赖性。在结构简单或稀疏的知识图谱中,合理控制跳数有助于避免噪声干扰;而在关系复杂、实体众多的图谱中,适当扩大跳数范围有利于信息充分建模,从而提升整体性能。

图2 各指标随子图跳数 K 变化图

路径采样跳数 T 控制生成推理链的最大长度,本实验设置 $T \in \{2, 3, 4\}$,在保证计算成本可控的同时,探索不同推理深度对性能的影响。从图3中可以看出,在CN15k数据集上,尽管该图谱结构相对稀疏,跳数从2增加到4后,各个指标均有小幅增长。表明在相对小规模图谱中,适当扩展邻域范围可以增强模型对实体上下文的建模能力。在NL27k数据集上,性能增长更为显著。该数据集实体种类与关系数量更多,图结构更

复杂,模型在跳数提升至4时取得了最优表现.特别是 Hits@10 和 MRR 指标,展现出明显优势.这说明更大范围的语义结构信息对模型捕捉潜在路径规律、学习关系间置信度具有积极作用.

为了进一步分析模型在不同 λ 值下的性能表现, λ 控制了模型在规则解析阶段的筛选强度,其变化直接影响规则集合的稀疏性、推理精度以及计算效率.实验设置中我们对 λ 取值进行了以下设置: $\lambda \in \{0, 0.5, 0.75, 1.0\}$, MSE 和 MAE 被选为衡量规则集合质量的标准.

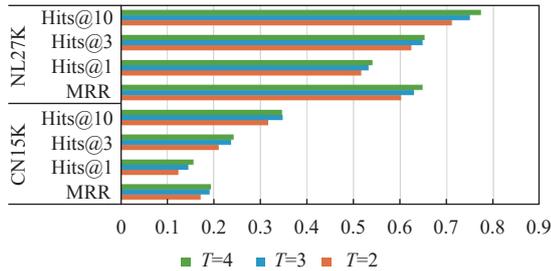


图3 各指标随路径采样跳数 T 变化图

从表6中的实验结果可以看出,在CN15k数据集上,当 $\lambda=0.5$ 时,模型的性能达到最优平衡.较低的 λ 值虽然计算速度较快,但模型无法有效区分高置信度和低置信度的关系,导致MAE和MSE均较高.

表6 λ 参数敏感性实验结果

λ	CN15k		NL27k	
	MSE	MAE	MSE	MAE
0	0.04668	0.18743	0.01693	0.05547
0.5	0.03914	0.16653	0.01456	0.051163
0.75	0.04563	0.17592	0.01367	0.04991
1	0.04991	0.20133	0.01801	0.05596

在NL27k数据集上, λ 的变化对性能的影响更为显著.当 $\lambda=0.75$ 时,模型表现出最佳的MSE和MAE.与CN15k数据集相似,较小的 λ 值导致模型的精度降低,适度地增大能够有效抑制噪声,提升模型对高置信规则的识别能力.整体来看,对相对稀疏一点的图谱,适中的 λ 值能较好地平衡规则精度与计算效率;对于稠密一点的图谱,适当增大 λ 值能够有效提升性能并减少噪声.通过调节 λ 值,模型能够根据不同数据集的特征自动调节阈值筛选强度,优化推理过程和规则集质量.

4.2.5 动态阈值效率分析

为验证动态阈值筛选与逆向贪心解析策略在复杂图谱中的计算效率,本文设计了针对不同规模和复杂度图谱的效率对比实验.分别在CN15k和NL27k两个

数据集上,构造了不同密度级别的子图,评估模型在处理时间和内存消耗方面的性能表现.实验中,本文将动态阈值筛选策略与固定阈值策略进行对比.固定阈值策略采用数据集全局统计的置信度中位数作为阈值,不随路径生成过程动态调整;而动态阈值策略则基于式(12),根据每一步生成关系的概率分布特征自适应调整阈值.实验在相同的硬件环境下进行,使用NVIDIA A100 GPU和Intel Xeon Gold 6248R CPU,以确保结果的可比性.

如表7所示,随着图谱规模和复杂度的增加,动态阈值筛选机制相较于固定阈值策略展现出更明显的效率优势.特别是在节点数量超过10k且平均度大于6的NL27k数据集上,动态阈值策略带来的加速比可达1.49倍,表明其在复杂图谱中具有良好的计算效率.进一步分析发现,该加速主要归因于动态阈值能够自适应地调整筛选标准,在高置信度关系分布相对集中的情况下适当提高阈值,有效减少了候选路径的探索空间.

表7 不同图谱复杂度下策略效率对比

数据集	图谱节点数 规模(k)	平均度	动态阈值		固定阈值		加速 比(倍)	内存消 耗(MB)
			耗时(s)	耗时(s)	耗时(s)	耗时(s)		
CN15k-small	5	3.2	0.42	0.48	1.14	246		
CN15k-full	15	4.7	1.35	1.62	1.34	712		
NL27k-small	10	6.3	2.14	2.86	1.34	1245		
NL27k-full	27	8.1	4.86	7.24	1.49	2879		

为进一步探究路径复杂度对策略效率的影响,本文在NL27k数据集上测试了不同最大路径长度设置下的性能变化.实验中分别设置路径长度从2到4,记录动态阈值策略与固定阈值策略的执行时间、内存消耗及筛选后的平均候选路径数量.

根据表8可知,在2跳场景下,动态阈值策略平均生成4.8条候选路径,而固定阈值策略生成5.7条,差异相对较小;而到4跳时,这一差异扩大至51%.值得注意的是,路径数量增长比与执行时间加速比呈现高度一致的趋势,证实了路径筛选效率是影响整体计算性能的关键因素.从2跳到4跳,固定阈值策略的执行时间从2.54s增至7.24s;而动态阈值策略从2.13s增至4.86s.这表明动态阈值策略不仅在绝对执行时间上有优势,在面对路径长度增加时,其计算复杂度的增长率也明显较低.

4.2.6 模型复杂分析

为评估模型的推理效率与资源消耗,本节从时间

复杂度与空间复杂度两个方面进行分析,展示模型在可解释性与效率间的平衡.

表8 不同路径长度下的策略效率指标

路径长度	动态阈值 耗时 (s)	固定阈值 耗时 (s)	加速 比 (倍)	动态阈值 平均候选 路径数	固定阈值 平均候选 路径数	路径数 增长比 (倍)
2	2.13	2.54	1.19	4.8	5.7	1.19
3	3.42	4.35	1.27	7.2	9.5	1.32
4	4.86	7.24	1.49	12.4	18.7	1.51

设输入序列长度为 L ,路径步长为 T ,隐藏维度为 d ,关系集合规模为 $|R|$,诱导子图边数为 $|E_s|$.UBERT-RM的整体计算主要包含BERT编码、Transformer解码、动态阈值及软门控筛选、规则解析与置信度预测这5个阶段.

在BERT编码阶段,复杂度为 $O(N_B(L^2d + Ld^2))$,其中 N_B 为编码层数;解码阶段每步包含自注意力、跨注意力及MLP投影操作,整体复杂度为 $O(T^2d + LTd + Td|R|)$.动态阈值与软门控仅涉及分布统计与激活运算,复杂度为 $O(Tk|E_s|)$,其中 $k \ll |R|$ 表示经动态筛选后的平均候选关系数.置信度回归模块基于短序列BERT建模,计算量较小. $O(N_B(L^2d + Ld^2)) + O(T^2d + LTd + Td|R|) + O(Tk|E_s|)$ 组成了模型整体的时间复杂度,可以看出主导项来自解码与矩阵传播部分.

模型空间开销主要来源于BERT参数、解码器注意力缓存以及规则解析过程的关系矩阵存储.BERT参数存储量为 $O(N_Bd^2)$.解码器注意力层数缓存及临时张量开销为 $O(d(L+T))$.在规则解析阶段,动态阈值与软门控仅保留高置信度边与路径,使得稀疏邻接矩阵由 $O(|E| \cdot |R|)$ 降为 $O(k|E_s|)$,显著降低存储成本.模型总体空间复杂度可表示为 $O(d(L+T)) + O(k|E_s|)$.总体而言,UBERT-RM在时间和空间复杂度上均呈线性增长,通过动态阈值与软门控机制有效减少冗余计算,实现了推理效率与模型可解释性的双重提升.

5 总结

本文提出了一种面向不确定知识图谱的规则挖掘与置信度预测模型UBERT-RM,该模型有效融合查询语义、结构信息与不确定性特征,构建了端到端的可解释推理框架.模型通过引入高置信邻居信息增强实体表示,并结合动态阈值机制与逆向贪心解析策略,实现了可解释且高置信的路径规则生成.在置信度建模方面,UBERT-RM设计了基于BERT的置信度评估模

块,结合路径与三元组信息,实现对推理结果置信度的细粒度回归预测.UBERT-RM在CN15k和NL27k两个公开数据集上,在链路预测和置信度回归任务中均显著优于现有主流模型,验证了其在不确定环境下的推理能力与可解释性.未来工作可推动模型在医学、金融、法律等专业领域知识图谱中的应用,利用其可解释推理能力,为临床辅助诊断、金融风控及法律条款检索等任务提供高置信度支持.

参考文献

- Liang K, Meng LY, Liu M, *et al.* A survey of knowledge graph reasoning on graph types: Static, dynamic, and multi-modal. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024, 46(12): 9456–9478. [doi: [10.1109/TPAMI.2024.3417451](https://doi.org/10.1109/TPAMI.2024.3417451)]
- Bollacker K, Evans C, Paritosh P, *et al.* Freebase: A collaboratively created graph database for structuring human knowledge. *Proceedings of the 2008 ACM SIGMOD International Conference on Management of Data*. Vancouver: ACM, 2008. 1247–1250.
- Lehmann J, Isele R, Jakob M, *et al.* DBpedia—A large-scale, multilingual knowledge base extracted from Wikipedia. *Semantic Web Journal*, 2015, 6(2): 167–195.
- Suchanek FM, Kasneci G, Weikum G. YAGO: A core of semantic knowledge. *Proceedings of the 16th International Conference on World Wide Web*. Banff: ACM, 2007. 697–706.
- Vrandečić D, Krötzsch M. Wikidata: A free collaborative knowledgebase. *Communications of the ACM*, 2014, 57(10): 78–85. [doi: [10.1145/2629489](https://doi.org/10.1145/2629489)]
- Carlson A, Betteridge J, Kisiel B, *et al.* Toward an architecture for never-ending language learning. *Proceedings of the 24th AAAI Conference on Artificial Intelligence*. Atlanta: AAAI Press, 2010. 1306–1313.
- Speer R, Chin J, Havasi C. ConceptNet 5.5: An open multilingual graph of general knowledge. *Proceedings of the 31st AAAI Conference on Artificial Intelligence*. San Francisco: AAAI Press, 2017. 4444–4451.
- Shen T, Zhang F, Cheng JW. A comprehensive overview of knowledge graph completion. *Knowledge-based Systems*, 2022, 255: 109597. [doi: [10.1016/j.knosys.2022.109597](https://doi.org/10.1016/j.knosys.2022.109597)]
- Lin Y, Liu Z, Sun M, *et al.* Learning entity and relation embeddings for knowledge graph completion. *Proceedings of the 29th AAAI Conference on Artificial Intelligence*. Austin: AAAI Press, 2015. 2181–2187.

- 10 Zhang W, Paudel B, Wang L, *et al.* Iteratively learning embeddings and rules for knowledge graph reasoning. Proceedings of the 2019 World Wide Web Conference. San Francisco: ACM, 2019. 2366–2377.
- 11 Bordes A, Usunier N, Garcia-Durán A, *et al.* Translating embeddings for modeling multi-relational data. Proceedings of the 27th International Conference on Neural Information Processing Systems. Lake Tahoe: Curran Associates Inc., 2013. 2787–2795.
- 12 Yang BS, Yih WT, He XD, *et al.* Embedding entities and relations for learning and inference in knowledge bases. Proceedings of the 3rd International Conference on Learning Representations. San Diego, 2015.
- 13 Trouillon T, Welbl J, Riedel S, *et al.* Complex embeddings for simple link prediction. Proceedings of the 33rd International Conference on International Conference on Machine Learning. New York: JMLR.org, 2016. 2071–2080.
- 14 Galárraga LA, Teflioudi C, Hose K, *et al.* AMIE: Association rule mining under incomplete evidence in ontological knowledge bases. Proceedings of the 22nd International Conference on World Wide Web. Rio de Janeiro: ACM, 2013. 413–422.
- 15 Meilicke C, Chekol MW, Ruffinelli D, *et al.* Anytime bottom-up rule learning for knowledge graph completion. Proceedings of the 28th International Joint Conference on Artificial Intelligence. Macao: AAAI Press, 2019. 3137–3143.
- 16 Qu M, Tang J. Probabilistic logic neural networks for reasoning. Proceedings of the 33rd International Conference on Neural Information Processing Systems. Vancouver: Curran Associates Inc., 2019. 693.
- 17 Qu M, Chen JK, Xhonneux LP, *et al.* RNNLogic: Learning logic rules for reasoning on knowledge graphs. Proceedings of the 9th International Conference on Learning Representations. Vienna: OpenReview.net, 2021.
- 18 Tang XJ, Zhu SC, Liang YT, *et al.* RuleE: Knowledge graph reasoning with rule embedding. Findings of the Association for Computational Linguistics: ACL 2024. Bangkok: Association for Computational Linguistics, 2024. 4316–4335.
- 19 Chen Z, Wang YH, Zhao B, *et al.* Knowledge graph completion: A review. IEEE Access, 2020, 8: 192435–192456. [doi: [10.1109/ACCESS.2020.3030076](https://doi.org/10.1109/ACCESS.2020.3030076)]
- 20 Ji GL, He SZ, Xu LH, *et al.* Knowledge graph embedding via dynamic mapping matrix. Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing. Beijing: Association for Computational Linguistics, 2015. 687–696.
- 21 Iyer RG, Bai YS, Wang W, *et al.* Dual-geometric space embedding model for two-view knowledge graphs. Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. Washington: ACM, 2022. 676–686.
- 22 Demir C, Ngomo ACN. Convolutional complex knowledge graph embeddings. Proceedings of the 18th International Conference on the Semantic Web. Springer, 2021. 409–424.
- 23 Kong XX, Hu FR, Song YC, *et al.* DGE-ASR: Two-view graph neural networks for knowledge graph completion. Proceedings of the 4th International Conference on Artificial Intelligence, Robotics, and Communication. Xiamen: IEEE, 2024. 1021–1024.
- 24 Zhang ZQ, Wang J, Ye JP, *et al.* Rethinking graph convolutional networks in knowledge graph completion. Proceedings of the 2022 ACM Web Conference 2022. Lyon: ACM, 2022. 798–807.
- 25 Zhang X, Zhang CX, Guo JT, *et al.* Graph attention network with dynamic representation of relations for knowledge graph completion. Expert Systems with Applications, 2023, 219: 119616. [doi: [10.1016/j.eswa.2023.119616](https://doi.org/10.1016/j.eswa.2023.119616)]
- 26 Yao L, Mao CS, Luo Y. KG-BERT: BERT for knowledge graph completion. arXiv:1909.03193, 2019.
- 27 Zha HW, Chen ZY, Yan XF. Inductive relation prediction by BERT. Proceedings of the 36th AAAI Conference on Artificial Intelligence. AAAI Press, 2022. 5923–5931.
- 28 Xu ZZ, Ye P, Chen H, *et al.* RuleFormer: Context-aware rule mining over knowledge graph. Proceedings of the 29th International Conference on Computational Linguistics. Gyeongju: International Committee on Computational Linguistics, 2022. 2551–2560.
- 29 Guo S, Wang Q, Wang LH, *et al.* Knowledge graph embedding with iterative guidance from soft rules. Proceedings of the 32nd AAAI Conference on Artificial Intelligence. New Orleans: AAAI Press, 2018. 590.
- 30 Chen XL, Chen MH, Shi WJ, *et al.* Embedding uncertain knowledge graphs. Proceedings of the 33rd AAAI Conference on Artificial Intelligence. Honolulu: AAAI Press, 2019. 413.
- 31 Chen XL, Boratko M, Chen MH, *et al.* Probabilistic box embeddings for uncertain knowledge graph reasoning. Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics:

- Human Language Technologies. Association for Computational Linguistics, 2021. 882–893.
- 32 Zhang JT, Wu TX, Qi GL. Gaussian metric learning for few-shot uncertain knowledge graph completion. Proceedings of the 26th International Conference on Database Systems for Advanced Applications. Taipei: Springer, 2021. 256–271.
- 33 Wang JB, Nie K, Chen XY, *et al.* SUKE: Embedding model for prediction in uncertain knowledge graph. IEEE Access, 2021, 9: 3871–3879. [doi: [10.1109/ACCESS.2020.3047086](https://doi.org/10.1109/ACCESS.2020.3047086)]
- 34 Liu Q, Zhang QH, Zhao F, *et al.* Uncertain knowledge graph embedding: An effective method combining multi-relation and multi-path. Frontiers of Computer Science, 2024, 18(3): 183311. [doi: [10.1007/s11704-023-2427-z](https://doi.org/10.1007/s11704-023-2427-z)]
- 35 Chen YL, Wu TX, Liu YC, *et al.* Uncertain knowledge graph completion with rule mining. Proceedings of the 21st International Conference on Web Information Systems and Applications. Yinchuan: Springer, 2024. 100–112.
- 36 Sun ZQ, Deng ZH, Nie JY, *et al.* RotatE: Knowledge graph embedding by relational rotation in complex space. Proceedings of the 7th International Conference on Learning Representations. New Orleans: OpenReview.net, 2019.
- 37 Yang SH, Zhang WY, Tang R, *et al.* Approximate inferring with confidence predicting based on uncertain knowledge graph embedding. Information Sciences, 2022, 609: 679–690. [doi: [10.1016/j.ins.2022.07.098](https://doi.org/10.1016/j.ins.2022.07.098)]
- 38 Chen ZM, Yeh MY, Kuo TW. PASSLEAF: A pool-based semi-supervised learning framework for uncertain knowledge graph embedding. Proceedings of the AAAI Conference on Artificial Intelligence, 35(5), 4019–4026. [doi: [10.1609/aaai.v35i5.16522](https://doi.org/10.1609/aaai.v35i5.16522)]

(校对责编: 张重毅)