

# 基于知识图谱全局和局部特征的复杂问答方法\*

陈跃鹤, 贾永辉, 谈川源, 陈文亮, 张民

(苏州大学 计算机科学与技术学院, 江苏 苏州 215008)

通信作者: 陈文亮, E-mail: [wlichen@suda.edu.cn](mailto:wlichen@suda.edu.cn)



**摘要:** 近年来, 研究者已经提出多种方法来解决知识库问答 (KBQA) 中的复杂问题, 并取得一定成果. 然而, 由于语义构成的复杂性以及可能存在推理路径的缺失, 复杂问题的求解效果依然不佳. 为了更好地解决这类问题, 提出基于知识图谱全局和局部特征的问答方法——CGL-KBQA. 所提方法利用知识嵌入技术提取知识图谱整体的拓扑结构和语义特征作为候选实体节点的全局特征, 根据实体表示和问句表示将复杂问答建模为复合的三元组分类任务. 同时, 将图谱在搜索过程生成的核心推导路径作为局部特征, 结合问句的语义相似性来构建候选实体不同维度特征, 最终形成混合特征评分器. 由于最终推理路径可能缺失, 采用基于无监督的多重聚类方法设计了聚类器模块, 进而根据候选实体的两类特征表示直接生成最终答案簇, 这使得非完全知识图谱问答成为可能. 实验结果表明, 所提方法在两个常见 KBQA 数据集上均取得不错的效果, 特别是在图谱知识不完全的情况下也具有非常好的效果.

**关键词:** 知识图谱问答; 信息检索; 知识图谱嵌入  
**中图法分类号:** TP18

中文引用格式: 陈跃鹤, 贾永辉, 谈川源, 陈文亮, 张民. 基于知识图谱全局和局部特征的复杂问答方法. 软件学报, 2023, 34(12): 5614–5628. <http://www.jos.org.cn/1000-9825/6799.htm>

英文引用格式: Chen YH, Jia YH, Tan CY, Chen WL, Zhang M. Method for Complex Question Answering Based on Global and Local Features of Knowledge Graph. Ruan Jian Xue Bao/Journal of Software, 2023, 34(12): 5614–5628 (in Chinese). <http://www.jos.org.cn/1000-9825/6799.htm>

## Method for Complex Question Answering Based on Global and Local Features of Knowledge Graph

CHEN Yue-He, JIA Yong-Hui, TAN Chuan-Yuan, CHEN Wen-Liang, ZHANG Min

(School of Computer Science and Technology, Soochow University, Suzhou 215008, China)

**Abstract:** Several methods have been proposed to address complex questions of knowledge base question answering (KBQA). However, the complex semantic composition and the possible absence of inference paths lead to the poor reasoning effect of complex questions. To this end, this study proposes the CGL-KBQA method based on the global and local features of knowledge graphs. The method employs the knowledge embedding technique to extract the topological structure and semantic features of knowledge graphs as the global features of the candidate entity node, and models the complex questions as a composite triple classification task based on the entity representation and question composition. At the same time, the core inference paths generated by graphs during the search process are utilized as local features, which are then combined with the semantic similarity of questions to construct different dimensional features of the candidate entities and finally form a hybrid feature scorer. Since the final inference paths may be missing, this study also designs a cluster module with unsupervised multi-clustering methods to select final answer clusters directly according to the feature representation of candidate entities, thereby making reasoning under incomplete KG possible. Experimental results show that the proposed method performs well on two common KBQA datasets, especially when KG is incomplete.

**Key words:** knowledge base question answering (KBQA); information retrieval; knowledge graph (KG) embedding

\* 基金项目: 国家自然科学基金 (61936010)

收稿时间: 2022-03-13; 修改时间: 2022-07-13, 2022-09-09; 采用时间: 2022-09-21; jos 在线出版时间: 2023-04-19

CNKI 网络首发时间: 2023-04-21

知识图谱 (knowledge graph, KG) 用于存储真实世界中的实体 (如人名, 地名等) 及实体间关系 (如出生地, 国籍等). 现有的大规模知识图谱包括 Freebase<sup>[1]</sup>, DBPedia<sup>[2]</sup> 和 NELL<sup>[3]</sup> 等. 而知识问答 (knowledge base question answering, KBQA) 作为知识图谱的应用, 在近几年来逐渐受到研究者们关注. 给定一个自然语言问句, 知识库问答的目的在于自动从上述知识库中找到问题的答案. 解决该任务的方法主要分为两类. 基于语义解析的知识库问答方法将自然语言问句转换为一种结构化查询语句, 并在知识库上执行得到最终答案<sup>[4-7]</sup>. 而基于信息检索的方法首先利用问句中的语义信息 (实体或关系) 从图谱中召回候选答案实体, 再通过排序方法从中选择最终答案<sup>[8-10]</sup>.

这两类方法在解决完全知识图谱上的复杂问答上取得一定进展, 然而在非完全知识图谱上仍存在问题. 非完全知识图谱 (简称非完全图谱) 是指由于知识图谱的静态特征和有限性使得在图谱中某些实体或实体间关系并不存在的情况, 即出现部分实体或关系缺失. 从严格意义来说, 现存大规模知识图谱都是非完全图谱, 差别在于信息缺失程度高低. 而复杂问句主要是指在推理过程中包含多跳关系, 或者因为推理的需要而考虑其他约束信息的问句, 图 1 说明了包含这两种情况的复杂问句样例. 在非完全图谱上的复杂问句推理主要存在两类问题: (1) 最佳或正确的推理路径可能在具体图谱中并不存在; (2) 问句复杂的语义构成在推理过程中难以捕捉. 比如在多跳类型的问句中, 由于非完全图谱中有效路径的缺失, 可能使模型倾向产生过长的推理路径, 而长路径的出现增加了表示和推理的难度. 在含约束类型的问句中, 约束信息在稀疏的图中难以将语义信息映射到推理过程. 对于这两个问题, 由于上述基于语义解析的方法在很大程度上依赖路径进行推理, 因此图谱的非完全性极大影响了其表现. 而基于信息检索的方法虽然较少受图谱固有的非完全性影响, 但如何充分地对接句中复杂的语义成分建模, 仍为亟待解决的问题<sup>[11]</sup>.

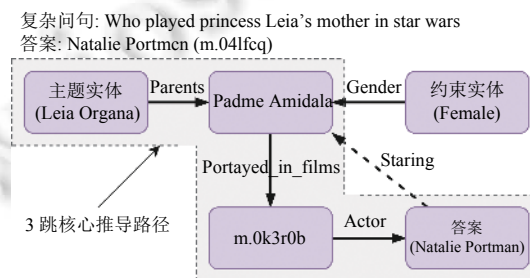


图 1 含约束的多跳复杂问句在稀疏和不完整的知识图谱上推理所面临的挑战

为了进一步说明这两类问题, 本文在图 1 中给出一个复杂问句样例. 该问句包含主题实体“Leia Organa”和约束实体“female”, 以及答案实体“Natalie Portman”. 从主题实体到达答案实体需要对问句进行 3 次推理, 从而得到 3 跳核心推导路径 (阴影区域), 此为前文提及的过长推理路径情况. 而从原问句中“mother”提取的约束实体“female”, 要求对问句构成进行充分解析从而解决复杂问句的约束情况. 在该例中, 如果关系边“staring (Natalie Portman, Padme Amidala)”存在于知识图谱中 (虚线表示), 那么这一复杂问句回答的难度将降低. 然而, 由于该关系边在实际的知识图谱中缺失, 系统不得不在图谱中较长的路径上进行推理 (实线表示). 这在实际问答过程中会产生指数增长的时空开销, 同时降低模型选择路径的准确率. 此外, 约束实体“female”, 作为复杂问句语义构成的一部分, 确定了中间实体的性别, 优化问题的推理逻辑, 从而使本文方法选择更加符合问句语义的答案.

针对上述问题, 本文提出 CGL-KBQA (combining global and local features-knowledge base questions answering), 利用知识图谱中的全局和局部特征获得候选实体表示, 并将问句复杂的语言特征建模到答案选择过程. 具体地, 本文采用知识图谱嵌入技术将知识图谱自身的拓扑结构和整体语义作为候选实体的全局特征, 而将图谱搜索过程产生的核心推导路径作为对应实体的局部特征. 在此基础上, 推理过程被转换为问句所隐含的三元组事实分类任务和问句与路径间的语义匹配任务. 利用上述两个特征, 本文构建了混合评分器模块, 从而充分解决复杂语义的建模问题. 由于最终用于搜索答案的路径可能缺失, 本文创新地设计了聚类器模块. 该模块利用评分器输出实体的两个不同维度特征建立数据空间, 采用基于无监督学习的多重聚类方法, 结合简单规则从数据空间中选择最符合问句表示的答案簇. 这使得本方法能够不依赖推理路径在非完全图谱上选择最终答案. 此外, 为了获得作为局部特征的

核心搜索路径并实现对候选实体的剪枝, 本文还设计了召回器模块. 该模块采用束搜索和优化的语义匹配模型, 在起始阶段召回图谱中与问句相关的路径和对应候选实体, 作为评分器模块的输入.

本文的主要贡献如下.

(1) 利用知识图谱中的全局特征和局部路径特征得到候选实体表示, 并将问句复杂的语义特征建模到候选实体的选择过程中, 提出一种新的实体评分机制.

(2) 基于候选实体表示, 进一步采用基于无监督的多重聚类融合方法生成最终答案簇, 可以较好回答存在缺失路径的复杂问句.

(3) CGL-KBQA 是一个全新的针对复杂问句的知识库问答方法, 采用基于信息检索的逻辑实现. 通过对数据集 WebQuestionsSP 和 ComplexWebQuestions 的实验, 证明了该方法的有效性.

本文第 1 节介绍知识图谱问答和推理的相关工作. 第 2 节介绍相关概念. 第 3 节介绍本文提出的基于知识图谱全局和局部信息知识库问答方法. 第 4 节给出了相关的实验结果和分析. 第 5 节总结全文.

## 1 相关工作

知识库问答是自然语言处理重要研究领域, 受到学术界和工业界的广泛关注, 本节将介绍完全知识图谱上的传统知识库问答方法和非完全知识图谱上的问句推理.

### 1.1 传统知识库问答

传统知识库问答方法主要分为两类, 分别是基于语义解析和基于信息检索的方法, 并在此基础上衍生出许多有效的方法.

基于语义解析的知识库问答方法着重于构建一个语义解析树或等效的中间查询结构以表示问句语义, 进而在知识库中执行寻找最终答案. 近年来, 在这一方向研究者提出了许多方法, 如 Berant 等人<sup>[12]</sup>使用短语和关系谓词之间的粗粒度对齐, 从统计角度上实现问句向中间结构的映射. Berant 等人<sup>[13]</sup>不使用大量统计文本作提前对齐, 而是将问句语义消歧的步骤设计在对查询语句的排序阶段, 使得问答方法在很大程度上得到简化. 而 Xu 等人<sup>[14]</sup>选择关注问句自身, 利用了问句中丰富的句法信息来将问句解析成以依存句法树为基础架构的中间表示. 同样的, Sun 等人<sup>[5]</sup>也从句法出发, 进而提出问句的多层次骨架结构解析方法, 该方法充分考虑问句的层次信息使得问句的语义表示更为立体. 值得注意的是, Yih 等人<sup>[15]</sup>, Luo 等人<sup>[16]</sup>, Lan 等人<sup>[6]</sup>结合知识库自身的图结构特性, 将问句解析为更符合图谱表示的查询图结构来表示问句的语义, 进而使多跳问句和约束问句被更好的建模, 这些方法使知识库问答方法的整体性能极大提高. 而在对生成的中间形式排序的过程中, 吴锷等人<sup>[17]</sup>和宋鹏程等人<sup>[18]</sup>提出优化的语义相似度模型训练方法, 并结合剪枝来提高中间查询格式和问句的相关性. 然而, 大多数基于语义解析方法或多或少依赖于人工规则或特征, 这限制了其可扩展性和可迁移性. 同时, 知识图谱的完全性要求也对这类方法产生影响, 因为这些方法都极度依赖于图谱中的具体的路径进行推理.

不同于基于语义解析方法, 基于信息检索的方法首先利用问句的语义信息(实体或关系)从图谱中召回候选答案实体, 再通过排序方法从中选择最优答案. 从推理逻辑来看, 基于信息检索的方法专注于将问句和答案映射到相同的向量空间中. 这使得这些方法可以查询任何独立于其结构特征和构成关系的知识库, 而不借助任何语法或词汇表. 在之前的研究工作中, Bordes 等人<sup>[19]</sup>首个将基于嵌入表示的方法应用于知识库问答当中, 通过向量空间内实体和问句表示之间的平移关系来建模答案选择过程. 而在后续工作中, Jain<sup>[20]</sup>和 Xu 等人<sup>[8]</sup>利用可迭代访问的记忆网络来存储候选实体以及其对应向量编码, 以模仿复杂问句的推理. 这使得通过基于信息检索的方法对解决多跳问句成为可能. 与上述主要使用词袋模型来编码问句和图谱内元素的方法不同, Hao 等人<sup>[21]</sup>则最早提出利用深层网络(如 LSTM)来编码问句, 使模型能学习到更符合图谱特征的代表. 此外, 多资源知识的混合方法也被提出, Das 等人<sup>[22]</sup>将知识图谱和外部文本事实投影到相同数据空间, 实现利用额外知识源对图谱内结构性知识的补充. 这类方法虽然在答案选择上具有更好的灵活性, 但在解析问句的语义构成上仍存在缺陷. 此外, 一些基于信息检索的问答方法只能预测唯一答案, 这对于完整的知识库问答任务来说仍存在问题.



和上述工作相比, 本文的主要区别和创新在于: 1) 对于语义解析方法, 本文方法并不生成中间语义结构, 而是将问句和预测实体联合建模. 通过将问句自身的句法构成和实体特征, 与预测实体构成三元组事实来表示问答过程. 2) 对于信息检索方法, 本文将知识图谱的拓扑结构和整体语义作为全局特征, 并且将知识图谱搜索过程产生的核心推导路径作为对应实体的局部特征, 从多个维度表达候选实体的特性. 最终本文引入答案聚类以缓解推理过程对具体路径的极度依赖, 同时问答不再限于唯一答案, 使得问题的解决过程更加完整.

## 1.2 非完全知识图谱上的推理

在复杂问句的推理过程中, 可能由于知识图谱的稀疏或者不完全, 因而无法在特定的图谱中找到正确或最佳的推理路径. 为了应对这一挑战, Sun 等人<sup>[9]</sup>提出利用外部的问句相关文本来补充从非完全图谱中提取的子图. 与之前方法不同, 其将补充文本和图谱的子图重构为一个异质图并在其上推理. 而 Xiong 等人<sup>[23]</sup>和 Han 等人<sup>[24]</sup>提出将外部的补充文本信息融合到实体表征中来实现知识补充, 这种语义层面的补充极大启发了之后的研究. 除了外部的文本语料库, Saxena 等人<sup>[10]</sup>提出利用预先训练好的知识库嵌入模型将答案选择的过程建模为链接预测任务, 更加充分地利用了图谱自身的特征信息. 张天杭等人<sup>[25]</sup>则将这类方法进一步改进, 并将其应用在中文的知识库问答实现当中. 而 Bhutani 等人<sup>[4]</sup>将来自不同类型知识图谱的信息相互补充, 利用图谱对齐方法来解决非完全的问题. 然而, 这些方法仍然存在两个问题. 其一, 外部语料库缺少泛化性, 在问答过程中可能并不适用于所有问句. 其二, 这些方法既没有考虑到问句本身的信息(如句法特征), 也没有考虑到推理过程中可能被利用的路径信息.

和上述工作相比, 本文提出方法并未引入外部数据, 因而更具有泛化性和迁移特性. 同时该方法通过问句和路径间的语义匹配任务, 在推理过程中强化了知识图谱内部的局部特征即从主题实体到候选答案实体的潜在路径, 使得候选实体的多维特征更为立体, 进而提高问题解决效率.

## 2 方法实现

### 2.1 任务定义

#### 2.1.1 知识图谱

给定实体集  $\mathcal{E}$  和关系集  $\mathcal{R}$ , 知识图谱  $\mathcal{G}$  被定义为一组三元组  $(s, r, o)$  的组合, 其中  $(s, r, o) \in \mathcal{E} \times \mathcal{R} \times \mathcal{E}$ . 每个三元组中存在头部实体和尾部实体  $s, o \in \mathcal{E}$ , 以及实体间关系  $r \in \mathcal{R}$ .

非完全知识图谱  $\mathcal{G}'$  表示为由于知识图谱的静态特征和有限性使得在图谱中某些实体或实体间关系并不存在的情况. 其与完全知识图谱的关系表示为  $\mathcal{G}' \subseteq \mathcal{G}$ .

#### 2.1.2 知识图谱嵌入

知识图谱嵌入也称知识图谱表示学习, 其目的在于为知识图谱  $\mathcal{G}$  内的每个实体  $e \in \mathcal{E}$  和关系  $r \in \mathcal{R}$  学习语义信息的低维表示. 对于每个图谱嵌入模型, 需要定义得分函数  $\phi(s, r, o)$ , 以评估三元组语义上的正确或错误. 事实三元组的概率对数表示为:

$$P(Y_{(s,r,o)} = 1) = \sigma(\phi(s, r, o; \Theta)) \quad (1)$$

其中,  $Y = 1$  表示三元组为真,  $\Theta$  表示模型参数.

三元组分类任务就是在训练好的知识图谱嵌入模型上, 给定一个完整三元组, 利用打分函数判断三元组的真实性. 三元组的得分越高, 其合理性越高.

#### 2.1.3 句子级语义匹配任务

给定两个句子  $s^a = \{w_1^a, w_2^a, \dots, w_{l_a}^a\}$  和  $s^b = \{w_1^b, w_2^b, \dots, w_{l_b}^b\}$ , 其中  $w_i$  表示为句子中的词. 句子级匹配任务的目的就是训练一个评分器  $\xi$ , 用于评价  $s^a$  和  $s^b$  之间的语义关联  $y = \xi(s^a, s^b)$ . 在深度学习背景下, 主要通过将句子转换为低维的嵌入表示, 再通过数据空间内嵌入表示之间的距离关系来表达句子的相似度.

### 2.2 方法概述

给定一个知识图谱  $\mathcal{G}$ , 其中包含实体集合  $\mathcal{E}$  和关系集合  $\mathcal{R}$ . 输入自然语言问句  $Q$ , 召回器模块通过剪枝搜索方法结合用于排序搜索路径的语义匹配模型从  $\mathcal{G}$  中提取候选实体集  $\mathcal{E}_c \subseteq \mathcal{E}$ , 并与对应的搜索路径形成数据对传入评

分器模块. 进而, 评分器模块利用知识图谱全局和局部特征结合问句语义和句法信息建立评分机制, 对所有  $e_{\text{cand}} \in \mathcal{E}_c$  生成两个不同维度的特征得分, 得到候选实体表示. 最后, 聚类器模块将评分模块生成的实体表示映射到二维特征数据空间, 利用多聚类融合方法捕捉候选实体的空间分布, 结合简单规则从  $\mathcal{E}_c$  中选取正确答案簇  $\text{ans}$ . 整体框架如图 2 所示. 其中, 实线为模块间数据流, 粗虚线为模块内数据交互, 点虚线为模型作用.

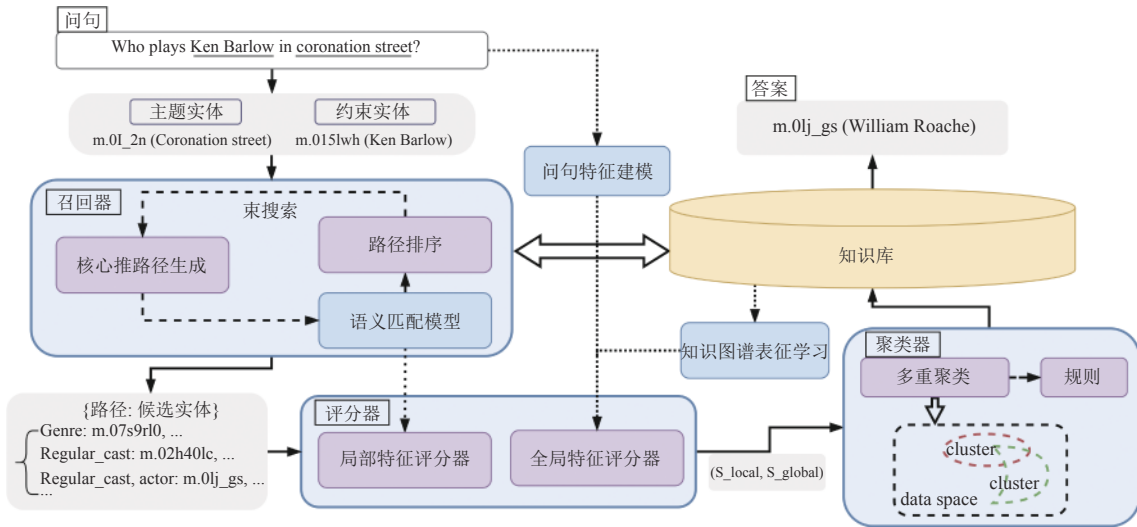


图 2 整体框架

### 2.3 召回器模块

召回器模块具有两个目的, 一方面生成局部搜索路径, 将路径对应实体召回作为候选实体, 另一方面对路径排序从而减少候选实体空间. 受基于语义解析方法<sup>[6,15]</sup>启发, 本文将问句和图谱中路径的相似性构建为句子级的语义匹配任务, 结合束搜索方法生成核心推理路径. 模块内工作如图 2 (召回器) 所示. 算法的执行过程如算法 1 所示.

#### 算法 1. 召回器模块算法.

输入: 主题实体  $TopicEntity$ , 束大小  $N$ , 跳数限制  $L$ ;

输出: 候选路径集合  $P_{t+1}$ .

```

 $P_0 = \{\}$ 
 $\mathcal{E}'_0 = \{TopicEntity\}$ 
for  $t: 0 \rightarrow L$  do
  for  $p$  in  $P_t$  do
     $\mathcal{E}'_t = \text{实体召回}(p)$ 
  end for
  for  $e$  in  $\mathcal{E}'_t$  do
     $P'_{t+1} = P'_{t+1} \text{ add } 1 \text{ 跳搜索}(e)$ 
  end for
   $P_{t+1} = \text{路径排序}(P'_{t+1}, N)$ 
end for
return  $P_{t+1}$ 

```

● 核心推理路径生成. 束搜索通常在图的解空间较大时用于缩减搜索的时空开销. 在该方法中, 束搜索起到了剪枝的作用, 即在路径的生成的过程中启发式的删除概率过低的推理路径. 本文假设束大小为  $N$ , 在搜索过程的第  $t$  次迭代时产生一个由主题实体组成的长度为  $N$  的查询路径集合, 表示为  $P_{t+1}$ . 在第  $(t+1)$  次迭代过程中, 对于每个  $p \in P_t$ , 本文从路径  $p$  中召回所有实体构成实体集  $\mathcal{E}'_t$ , 并根据  $\mathcal{E}'_t$  中的实体进一步做 1 跳搜索 (路径召回). 对所有  $p \in P_t$  重复这一操作, 并让  $P'_{t+1}$  表示操作后所有结果查询路径的集合. 然后, 本文利用语义匹配模型训练的评分函数对  $P'_{t+1}$  中所有查询路径进行排序, 并将 top- $N$  个路径放入  $P_{t+1}$  中. 不断继续迭代这一过程. 直到搜索跳数受限.

● 路径排序. 在第  $t$  次迭代结束时, 通过计算候选查询路径  $p$  与问句  $Q$  之间语义相似度  $prob(p|Q)$ , 再对  $P_t$  中的所有候选路径进行排序. 具体地, 本文通过基于 BERT<sup>[26]</sup> 的语义匹配模型来实现语义相似度的计算过程. 在训练过程中, 本文将真实知识图谱内非正确的推理路径作为负例采样, 并且在一个训练批次内仅放入唯一正确路径作为正例, 进而结合 listwise 的排序策略<sup>[27]</sup> 权衡批次内所有训练数据的全局关系. 该排序策略是一种基于对比学习的排序策略, 目的在于最大化正例和负例之间的语义差异. 问句和核心推理路径拼接为  $\{que : path\} = \{Q : TopicEntity, p, < pad \>\}$  的形式作为模型的输入, 而将模型最后一层 CLS 标签所指示的向量输出, 作为输入问句路径对的语义相似度表示.

### 2.4 评分器模块

本节中介绍如何利用知识图谱中的全局特征和局部路径特征获得候选实体表示, 并将问句复杂的语义特征建模到候选实体的选择过程. 具体实现如图 3 所示.

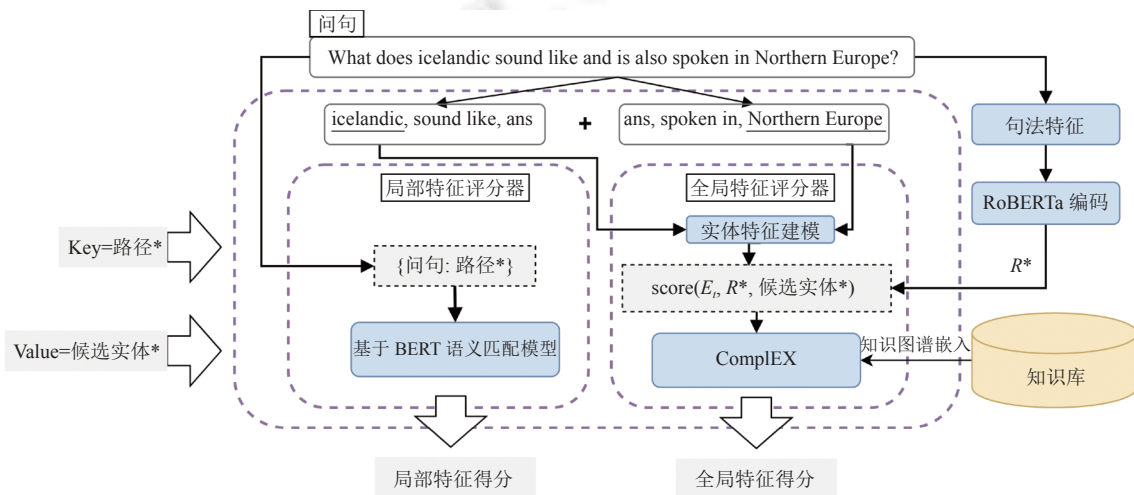


图 3 评分器模块具体实现

#### 2.4.1 问句特征建模

● 实体特征. 对于含约束类型的复杂问句, 比如“*What does icelandic sound like and is also spoken in Northern Europe?*”, 能够提取出主题实体“*icelandic*”以及约束实体“*Northern Europe*”. 在基于语义解析的方法中, 这种约束信息可以通过预定义的模板操作融入到中间表示. 但对于基于信息检索的方法, 这类语义信息由于难以建模而倾向被忽略. 事实上, 该复杂问句的语义在逻辑上可以解构为“*icelandic sound like?*”和“*what spoken in Northern Europe?*”两个子问句的复合, 如图 3 中所示. 根据语义平移思想, 对原问句的推理可以转换为子问句的语义复合, 即  $(Icelandic + Northern Europe) \odot (sound\ like + spoken\ in) \Rightarrow ans$ . 该式前部分表示实体特征融合, 而中间部分则近似于去除实体的问句本身. 为了实现实体特征融合, 本文将主题实体, 约束实体和问句映射到同一维度的空间, 分别表示为  $e_t, e_c, e_q$ . 并将约束实体在问句表征向量方向上的投影视为对主题实体的语义增强, 如公式 (2) 所示. 由于主题实体和约束实体的特征表示均来自同一个预训练的知识图谱嵌入模型, 而问句则通过 RoBERTa 模型<sup>[28]</sup> 训练以拟合嵌入模型所学习到的关系表示 (详见第 3.3.2 节). 因此, 该策略能在建立实体间语义相关的同时有效抑制噪声.

$$E_i = \frac{1}{1+n} \left[ e_i + \alpha \sum_{i=1}^n \varphi(e_q, e_c(i)) \right] \quad (2)$$

其中,  $\varphi$  是矢量映射函数,  $n$  表示约束实体数量.

• 句法特征. 受 Luo 等人<sup>[16]</sup>启发, 本文在语义和句法两个层面对问句进行编码. 但之前工作不同的是, 本文希望在句法树编码的基础上引入问句中词的相对位置信息, 来加强问句中各语义成分的关联. 当子问句转换为陈述形式“icelandic sound like <ans>”, 实体“icelandic”和关系谓词“like”在句法树上的相对位置信息可以帮助更好建立问句表示. 因此, 本文在基于 Tree-GRU (tree-structured gated recurrent unit) 特征基础上, 进一步考虑基于 TPF (tree position feature) 特征. 在基于 TPF 的句法特征中, 对于问句中的每个词得到其词嵌入表示  $w_i$  和相应句法树位置表示  $d_i$ . 本文将  $w_i$  和  $d_i$  连接, 得到  $w_i$  的相对位置表达  $z_i = \{w_i; d_i\}$ . 进而利用 BiGRU 模型对整个问句进行编码, 最终得到基于 TPF 的位置编码, 表示为  $r_{\text{emb}}^{\text{TPF}}$ . 本文将基于 Tree-GRU 的特征表示为  $r_{\text{emb}}^{\text{TGRU}}$ . 这两类句法特征和问句通过 RoBERTa 模型获得的嵌入表示做权重和作为问句最终表示.

#### 2.4.2 全局评分器

全局评分器旨在表示推理过程的全局事实意义, 并生成候选实体的全局特征表示, 如图 3 所示. 本文采用 ComplEx 知识图谱嵌入模型<sup>[29]</sup>, 在真实的知识图谱子集上预先学习实体节点的语义信息以及实体相关的拓扑结构. 进而, 受 Saxena 等人<sup>[10]</sup>的启发, 本文将候选答案节点的全局特征建模为知识图谱补全中的三元组分类任务, 即利用得分函数  $\phi(o, r, s)$  来判断三元组  $(o, r, s)$  的真实性, 在图 3 中表示为  $\text{Score}()$ . 在第 3.3.1 节, 问句被解构为子问句的复合, 而子问句分别可以表示为事实三元组 (icelandic, sound like, <ans>) 和 (Northern Europe, be spoken in, <ans>). 因此, 在全局评分器中, 实体的融合特征将作为三元组头部表示, 而原问句在语义和句法上的融合表示则近似为子问句的关系复合作为三元组关系表示, 而来自图谱嵌入模型的候选实体表征则构成了三元组尾部表示. 复合的三元组分类任务建模如公式 (3) 所示. 比起前者工作, 该方法充分捕捉了问句的语义和句法多层次特征, 同时将问句的复杂语义构成完整的建模到复合三元组表示, 进而加强了实体与关系以及答案之间的逻辑联系.

$$S_{\text{global}} = \phi(o, r, s) = \phi(E_i, r_q \oplus [r_{\text{emb}}^{\text{TGRU}}; r_{\text{emb}}^{\text{TPF}}], E_c) \quad (3)$$

其中,  $\phi$  代表三元组分类任务中的评分函数.  $E_i$  为实体特征融合表示,  $r_q$  为问句的 RoBERTa 编码表示,  $E_c$  为候选实体表征,  $\oplus$  为权重和.

#### 2.4.3 局部评分器

全局评分器虽然能够从图谱宏观上表示推理过程, 但忽略了知识图谱内的局部特征即从主题实体到候选答案实体的潜在路径. 而局部评分器旨在捕捉实体在图谱中的路径信息, 并转换为候选实体的特征表示. 事实上, 局部搜索路径与候选实体的对应关系为局部评分器提供了两方面信息. 其一, 如果候选实体对应的搜索路径存在, 可以作为隐含特征融入推理过程. 其二, 如果判断为错误路径, 可以对排序对应的非正确候选实体起到抑制作用.

具体来说, 本文使用第 3.2 节路径排序中相同的基于 BERT 语义匹配模型作为基准模型, 对从主题实体出发在图谱中搜索得到的局部查询路径  $p$  和问句  $Q$  进行语义相似度计算, 得到  $\text{prob}(p|Q)$  作为路径  $p$  的存在概率. 如公式 (5) 所示. 基于这一概率, 本文设置存在性阈值  $K$ , 以确定搜索路径是否有意义. 如果该路径具有高存在概率, 问句的局部特征得分将通过偏置函数与全局特征得分映射到同一数据空间中. 偏置函数是一个启发式函数, 旨在使两类特征在数据空间内的分布更为合理. 反之, 如果该路径存在概率很低, 则通过奖惩系数在局部特征维度抑制该路径对应的候选实体的整体得分.

$$S_{\text{local}} = \text{prob}(p|Q) + f_{\text{bias}}(\text{prob}(p|Q), K, \gamma) \quad (4)$$

其中,

$$\text{prob}(p|Q) = \sigma(w^T \text{BERT}_{\text{CLS}}(\{que : path\})) \quad (5)$$

其中,  $\sigma$  为 Sigmoid 函数,  $\{que : path\}$  为问句与路径的拼接形式.  $f_{\text{bias}}$  为偏置函数,  $\gamma$  为奖惩系数.



## 2.5 聚类器模块

经过混合评分模块的处理, 候选实体表示已经包含全局和局部两个维度特征. 进而, 本文通过聚类模块, 将候选实体表示映射到特征空间, 在不依赖推理路径的情况下, 捕捉候选实体的空间分布生成最终答案簇.

- **K-means**<sup>[30]</sup> 给定一组数据点和所需的聚类数量  $K$  (用户指定), 该算法根据预先设计的距离函数将数据迭代分成  $K$  个聚类.

- **DBSCAN**<sup>[31]</sup> 是一种基于密度的聚类算法, 使用一组参数来描述样本分布的邻近程度. DBSCAN 将聚类定义为由密度可达性关系得出的最大密集连接样本集合.

多重聚类方法. 与传统的聚类任务不同, 该方法并不注重于数据簇的划分, 而是期望从数据空间中选择与问句最相关的唯一答案聚类. 然而, 具体实验中生成的特征空间容易出现集群密度不均和数据量较小的情况. 虽然 K-means 可以快速地从数据空间中找到答案聚类, 但由于存在需要人工预定义聚类数量, 易受到异常值影响以及对不规则形状不敏感等问题, 往往导致部分正确实体被过滤. 同样, 虽然 DBSCAN 无需指定聚类数量且能够在避免异常值影响的同时生成任意形状的聚类, 但其容易因采纳过多边缘点而引入实体误差. 因此, 本文结合两类方法的不同特点, 在通过 DBSCAN 对答案实体进行扩召回的同时, 利用 K-means 做准确性上的提高. 具体而言, 多重聚类方法首先将评分器生成的候选实体特征映射到二维数据空间中, 建立候选答案点集  $D = \{(x, y) | x \in S_{\text{global}}, y \in S_{\text{local}}\}$ . 利用上述两类方法对  $D$  中的所有数据分别生成形态不同的聚类簇, 表示为  $C_{KM} = \{c_{km}^1, c_{km}^2, \dots, c_{km}^K\}$  和  $C_{DB} = \{c_{db}^1, c_{db}^2, \dots, c_{db}^h\}$ . 选择数据空间  $D$  中距原点曼哈顿距离最远的特征点  $h$  作为求解中心, 进而分别从两个不同的答案聚类中抽取包含  $h$  的答案簇  $c_{km}^h \subset C_{KM}$  和  $c_{db}^h \subset C_{DB}$ . 最后, 根据答案实体在聚类空间中的分布情况, 对  $c_{km}^h$  和  $c_{db}^h$  取交集或并集.

## 3 实验

### 3.1 数据集

WebQuestionsSP (WebQSP) 数据集<sup>[32]</sup>主要包含可通过 Freebase 知识图谱回答且相对简单的 1 跳和 2 跳问句, 同时包含少量复杂问句, 数据统计如表 1 所示. 由于完整的 Freebase 知识图谱包含近 20 亿三元组, 想要在此基础上训练知识图谱嵌入模型从时间和硬件角度上无法实现. 因此, 本文按照 Sun 等人<sup>[9]</sup>的方法提取出知识图谱的一个子集, 其中包含 WebQSP 的问句中提及的所有实体 2 跳以内的部分事实三元组. 最终生成的子集中包含 180 万个实体和 570 万个三元组. 本文在此基础上构建知识图谱全局评分器. 知识图谱子集的具体生成方式见第 3.2 节.

表 1 数据集统计

数据集	训练集	验证集	测试集	复杂问句 (%)	约束问句 (%)
WebQuestionsSP (WebQSP)	2 848	250	1 639	40.8	25.6
ComplexWebQuestions (CWQ)	27 623	3 518	3 531	88.5	64.5

ComplexWebQuestions (CWQ)<sup>[33]</sup>是根据 WebQSP 数据集内问句生成的复杂问句数据集. 该数据集通过扩展问句实体或对答案添加约束, 以构建语义更为复杂的多跳问句. 其中包含 4 种类型的复杂问句: conjunction (45%), composition (45%), comparison (5%), superlative (5%). 这 4 类问句可以总结为含约束问句, 多跳问句, 比较类型问句和次序问句, 本文注重解决前两类问句, 且这两类问句往往在该数据集中存在复合的情况. 本文采用与 WebQSP 数相同的知识图谱子集生成方法, 该子集分别包含 340 万个实体和 810 万个三元组.

### 3.2 基线方法及相关设置

本文将提出的方法与几种典型知识图谱问答系统在 WebQSP 和 CWQ 两个数据集上进行比较.

- 基于信息检索的方法. Key-value memory network (KV-Mem)<sup>[34]</sup>维护一个存储知识图谱事实和文本编码为键值对的内存表, 并在其上检索. GRAFT-Net<sup>[9]</sup>和 Pull-Net<sup>[9]</sup>创建从问句出发的具体子图, 进而使用图卷积神经网络



络来进行推理. EmbedKGQA<sup>[10]</sup>使用预先训练的知识图谱嵌入模型来将非完全图谱上的答案选择构建为简单的链接预测任务. BERT+RM<sup>[35]</sup>提出 3 个辅助任务来对问句与关系间的表示通过多任务学习加强. 此外我们还提供了 CGL-Baseline (没有使用全局和局部特征以及多重聚类方法) 作为基线系统, 以显示本文方法的有效性.

- 基于语义解析的方法. UHop<sup>[36]</sup>在不考虑约束的情况下, 使用束大小为 1 的束搜索来处理多跳问句. TextRay<sup>[37]</sup>将复杂问句分解为多个简单问句, 进而通过复杂问句的构成特征将简单问句答案的组合作为问答结果. SPARQA<sup>[5]</sup>提出基于句法的层次架构来解析复杂问句. MULTIQUE<sup>[4]</sup>使用一系列的构成规则将复杂问句映射为简单查询路径的组合. GrailQA RANKING<sup>[38]</sup>是一个基于 BERT 实现的问答方法, 具有很高的泛化能力.

对于 WebQSP 数据集, 完全知识图谱的设定是易于实现的. 这是因为其包含的问句的推理跳数都在 2 跳以内, 使得在 Freebase 知识图谱中不太可能出现路径缺失的情况, 这也保证了通过第 4.1 节中提及方法创建的知识图谱子图不会遗漏数据集中问句的任何实体节点和推理必须的关系. 而为了实现非完全图谱上的推理, 本文跟随 Sun 等人<sup>[9]</sup>的工作, 通过丢弃图谱子集中的一些路径来模拟这一情况. 具体地, 本文以概率  $p = 0.5$  随机地从构建的知识图谱子集中删除三元组, 这会模拟推理过程中关系或实体缺失的情况, 在文中称这种设置为 KG-50. 而前者称为 KG-full.

对于 CWQ 数据集, 其包含更多复杂的问句, 而这类问句往往需要 3 跳及以上的推理路径. 对这类问句来说, 本身就可能是因为 Freebase 知识图谱的不完全而产生这样过长的推理路径, 同时因为计算资源限制和解析困难, 所以不总像 WebQSP 数据集中问句能在图谱中找到有效的推理元素 (实体或关系) 进行推理. 这使得图谱全局特征评分器难以提取一个包含所有实体节点和推理必须关系的知识图谱子集. 因此, 该数据集主要用于说明本文方法在完全图谱上推理的能力. 本文将 CWQ 数据集的完全图谱设置为 KG-full. 同时, 为了与其他基线方法比较本文采用和 WebQSP 数据集同样的思路也构建了一个知识图谱子集, 用于更加全面的说明本文的方法, 实验设置为 KG-50.

### 3.3 评价指标

本文使用 Hits@1 和 Macro-F1 作为评估性能的指标, 该评价指标也广泛用于之前的研究工作. Hits@1 定义为所有数据中 top-1 答案推理出正确答案的比例. 形式上, 如果模型预测问句的 top-1 答案存在标注答案中, 则该问句 Hits@1 为 1, 高 Hits@1 值说明模型具有较高的预测性能. 而 Macro-F1 定义为准确率和召回率的调和平均数. 该指标统计模型对各个问句预测结果的准确率和召回率, 得到各自的 F1 值, 然后取平均值得到 Macro-F1. 在后文中, Macro-F1 用 F1 表示, 高 F1 值说明模型具有较高的预测性能.

### 3.4 结果与分析

#### 3.4.1 WebQuestionsSP 分析

从数据上看, WebQSP 是一个相对较小的数据集. 表 2 显示了 WebQSP 数据集上各方法的总体比较. 在基于信息检索的方法中, 本文方法在 KG-50 设置和 KG-full 设置中都优于所有基线方法, 这其中包括 Pull-Net 和 EmbedKGQA 以及 BERT+RM. 具体而言, 在 KG-50 设置下, Hits@1 指标提高 1.4% 而 F1 指标则提高 4.4%. 而在 KG-full 设置下, Hits@1 与 F1 则分别提高了 0.5% 和 2.7%. 同时, 本文在基线方法 CGL-Baseline 和 CGL-KBQA 方法之间进一步加入显著性测试, 以说明本文方法带来的提升的鲁棒性, 最终在配对卡方检验下 KG-50 和 KG-full 设置的结果均为  $p < 1-E5$ . 这验证了本文的假设, 在不借助外部文本补充仅对知识图进行深度利用的情况下, 本文方法能够通过结合知识图谱的全局和局部特征对候选实体表示进行充分建模, 而问句复杂的语义构成也能被该方法所捕捉, 所以无论在结果准确率还是问答完成性上都有可观的提高.

在与基于语义解析的方法的比较中, 对于 KG-full 设置, 虽然本文方法与 UHop 和 GrailQA Ranking 在 F1 指标上存在微小差距, 但在 Hits@1 指标上同样取得最好效果. 本文认为这一差距产生的主要原因可能是由于在聚类阶段, 数据空间中候选实体的全局和局部特征得分分布不均匀, 导致并不能产生高召回的答案簇. 而在 KG-50 设置上, 本文方法也极大地超过了其他方法的表现. 从整体上看, 无论是在完全知识图谱还是在非完全知识图谱的问答任务中, CGL-KBQA 都极具竞争力.

表 2 WebQSP 数据集的结果 (%)

分类	方法	KG-50		KG-full	
		Hits@1	F1	Hits@1	F1
基于信息检索方法	KB-Men	32.7	23.6	46.7	38.6
	GTAFT-Net	52.3	37.4	68.7	62.8
	Pull-Net	50.1	—	68.1	—
	EmebdKGQA	53.2	—	66.6	—
	BERT+RM	59.0	43.1	72.9	64.5
	CGL-Baseline	46.9	38.9	59.1	48.2
基于语义解析方法	UHop	—	—	—	68.5
	TextRay	—	—	—	60.3
	MULTIQUE	50.8	41.7	69.8	61.2
	GrailQA Ranking	—	—	—	<b>67.0</b>
	CGL-KBQA	<b>60.4</b>	<b>47.5</b>	<b>73.3</b>	66.3

### 3.4.2 ComplexWebQuestions 分析

与 WebQSP 数据集相比, CWQ 数据集具有更多训练数据, 这使得本文方法能更好地学习知识图谱中包含的全局和局部特征, 因而更加合理的构建候选实体的特征环境. 然而, CWQ 数据集中所提供的问句语义构成更为复杂, 这会造成极大的推理困难.

表 3 表明各方法在 CWQ 数据集上的结果, 其中\*表示开发集结果. 从表 3 结果来看, 本文方法和其他知识图谱问答方法虽然在精度指标 Hits@1 上表现略有欠缺, 但在整体指标 F1 上表现最好, 显示出相当的性能. 对于 Hits@1 指标, 本文方法与基于信息检索的方法 PullNet 略有差距, 这是因为后者引入了实体相关的文本信息来作为知识图谱的外部补充, 从而能够选择更加贴合问句本身的语义的 top-1 答案实体. 而与基于语义解析的方法 TextRay 和 MULTIQUE 的差距也说明结合构成树对问句进行语义解析, 能够更好地排序图谱中的搜索路径, 从而得到高准确率. 但对于 F1 指标, 本文方法的效果得到了明显的改善, 并超过了其他的所有方法的表现. 该综合指标表明, 一方面本文提出的评分器可以有效地表示候选答案在全局和局部两个层面的特征, 另一方面聚类器能够根据候选实体的特征分布探测到正确的答案簇. 总的来说, 本文方法能够减少推理过程中对路径的依赖, 同时在有限的实体召回中找到最优答案, 具有不错的整体效果.

表 3 CWQ 数据集的结果 (%)

分类	方法	KG-50		KG-full	
		Hits@1	F1	Hits@1	F1
基于信息检索方法	KB-Men	14.8*	—	21.1	—
	GTAFT-Net	26.1*	—	32.8	26.0
	Pull-Net	<b>31.5*</b>	—	<b>45.9</b>	—
基于语义解析方法	UHop	—	—	30.5	29.8
	TextRay	—	—	40.8	33.9
	MULTIQUE	—	—	41.2	34.6
	GrailQA Ranking	—	—	31.2	—
	SPARQA	—	—	31.6	—
	CGL-KBQA	29.7*	<b>28.0*</b>	40.1	<b>37.9</b>

### 3.4.3 非完全图谱上的召回

为了进一步说明在非完全图谱上的推理, 本文增加召回模块在两个数据集上对候选答案的召回效果实验. 在 KG-full 设置上, 召回过程在完全知识图谱上正常执行. 而对于 KG-50 设置, 本文在核心推导链的生成过程中随机丢弃部分召回的路径和实体, 而这一概率也被设置为 0.5.

如表 4 所示, 召回器模块在图谱缺失情况下仍能有效地从知识图谱中召回候选实体. 在 WebQSP 数据集中, 即使丢弃掉一半的路径及对应实体, 仍能达到 62% 的答案召回率. 结合前文表 1 中结果, 也说明评分器模块的设计能够很好地对候选实体表示进行多层次建模. 而对于更加复杂的 CWQ 数据集, 虽然回答问句所需推理路径往往扩展到 3 跳及以上, 但召回模块在巨大的求解压力下仍能在图谱完全时取得尚可结果. 而在非完全图谱上, 路径越长, 丢弃路径造成的误差不断传递, 使得最终的实体召回结果相比于 WebQSP 数据集有着更高比例的下降. 结合上述两个实验, 虽然召回效果在不同的实验设置下表现不同, 但本文方法在两个数据集上的整体表现仍呈现优势, 这表明本文方法在非完全图谱推理上具有可行性和有效性.

表 4 不同设置下的候选实体召回效果 (%)

Dataset	KG-50	KG-full
WebQuestionsSP	62.0	78.4
ComplexWebQuestions	34.5	44.3

#### 3.4.4 消融实验

在本节中, 本文在两个数据集上测试了 CGL-KBQA 方法 5 个不同的版本, 并在每个版本中都去掉一种文中提及的设置, 以说明整个方法结构的有效性. 对于所有实验, 本文都应用  $F1$  的评价指标, 实验结果如表 5 所示.

表 5 消融实验 (%)

Method	WebQSP		CWQ	
	KG-50	KG-full	KG-50	KG-full
CGL-KBQA	<b>47.5</b>	<b>66.3</b>	<b>28.0</b>	<b>37.9</b>
w/o 召回器模块	39.4	50.0	25.8	32.3
w/o 全局评分器	28.2	48.6	22.1	27.6
w/o 局部评分器	36.9	45.3	24.0	32.3
w/o 聚类器模块	36.4	47.8	23.9	33.8
w/o 问句特征建模	46.3	63.7	26.9	35.2

表 5 中, “w/o 召回器模块”表示不使用召回器模块来生成搜索路径候选实体对, 而直接使用知识图谱中 1 跳内的所有路径及对应实体作为评分模块的输入. 如表 5 所示, 包含召回模块的 CGL-KBQA 在两个数据集上的多个设置上均有明显的性能提升, 尤其在 KG-full 的设置下提升分别为 15.8% 和 5.6%. 这意味着策略性地生成高语义相关性搜索路径同时减少候选实体空间可以有效提高系统整体效果. “w/o 全局评分器”表示在评分器模块建模候选实体表示时不使用图谱内的全局特征, 只考虑应用局部评分器. 而“w/o 局部评分器”则相反. 结果表明, 在 WebQSP 数据集上去除这两部分会使系统性能在 KG-full 设置下分别下降 17.7% 和 21.0%. 而对于 CWQ 数据集则分别下降 10.3% 和 5.6%. 但这种效果的下降在 KG-50 的设置下更为明显. 这说明本文提出的对候选实体从全局和局部两方面建模的方式能使得实体表示更加立体, 从而稳定提高问答任务中对答案选择的准确率. 而对于更为复杂的问句, 因为获取局部路径特征的难度增加, 候选实体的全局特征表现出更有效的作用. “w/o 聚类器模块”表示不使用多重聚类方法, 只返回得分最高的 top-1 实体作为最终答案. 多重聚类方法的应用, 在 WebQSP 数据集上的两个设置下分别得到 11.1% 和 18.5% 的提高, 而在 CWQ 数据集上分别得到 4.1% 和 4.1% 的提高. 这意味着聚类模块在完善问答任务的同时也能捕捉正确候选实体在特征数据空间内的分布特性. “w/o 问句特征建模”表示在对问句的建模过程中不使用约束实体信息和句法信息. 结果表明, 本文提出的问句建模方法极为稳定且有效. 结合所有实验, 表明 CGL-KBQA 方法的结构是有效的, 且每个模块都提高该方法整体效果.

#### 3.4.5 实体特征权重分析

在本节中, 本文对第 2.4.1 节实体特征中  $\alpha$  超参进行进一步实验分析, 从而说明约束实体特征对文本方法的影响, 结果如图 4 所示.

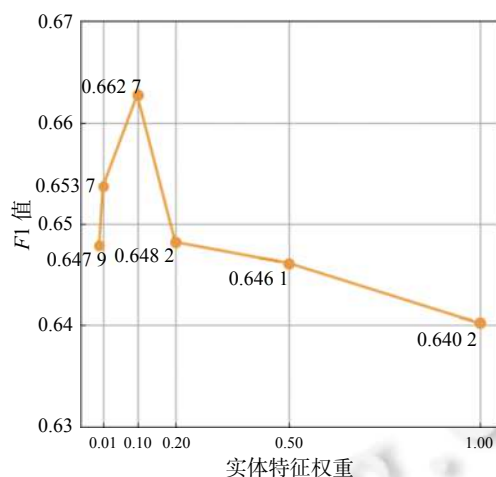


图4 实体特征权重参数的结果

在本文中设计超参目的在于引入约束实体特征同时, 减弱其噪声带来的影响. 如图4所示, 在 $\alpha = 0.10$ 时, CGL-KBQA 在问答上表现出最好的效果, 同时效果又随着参数的不断加大而有所降低, 本文认为这是因为实体特征虽然对于构建三元组分类任务来说非常重要, 但其权重的增加在引入噪声的同时也会减弱主题实体的重要性. 而在 $\alpha < 0.10$ 时的效果增加则认为特征是信息的未引入, 使得方法还未去捕捉这种事实特征.

#### 3.4.6 案例分析

在本节中, 为了进一步说明评分器生成的数据空间中实体分布情况以及聚类模块的实现和效果, 本文选择了预测样例“what countries are part of the UK?”的答案实体空间分布, 如图5所示. 其中红色点代表 CGL-KBQA 聚类器模块最终的答案, 而蓝色点代表其余的候选实体. 同时图中用两种颜色的候选空间分别表示本文采用的不同聚类方法所得到的答案空间 (头部聚类答案簇), 黄色代表 K-means, 蓝色代表 DBSCAN.

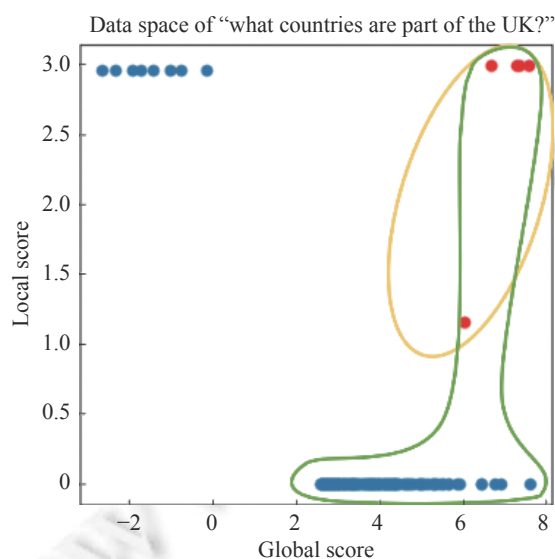


图5 样例问句的答案空间分布

首先从问句出发可以清晰感知到其存在多答案, 因此为了真实地解决问题, 聚类模块的能力就必然不满足于仅返回 top-1 答案, 而应该去返回符合问句语义描述的“头部聚类”结果. 从图5中可以看到, K-means 方法可以快



速且在小范围内准确的确定答案簇,而 DBSCAN 则更倾向于去实现点的可达性.根据两种方法生成的不同头部聚类结果以及其余特征点的分布,在该问句下最终选择交集来返回答案,保证回答的预测准确率.

## 4 总 结

本文提出了一种新的知识图谱问答方法 CKL-KBQA,其利用知识图谱中的全局和局部特征,结合问句复杂的语义构成,重新建模了候选答案的选择过程.对于整个方法而言,本文采用了 3 个模块即召回器,评分器和聚类器.首先,通过搜索路径和问句的相似性,结合基于剪枝的搜索方法,产生有效的候选实体集.进而,从图谱的全局特征和局部推理路径出发,结合复杂问句的语义特征和句法特征,建立有效的混合评分机制.最后,采用基于无监督的多聚类方法,不依赖推理路径而直接从数据空间中准确生成最优答案簇,使非完全图谱上的推理成为可能.在 WebQSP 和 CWQ 两个数据集上的实验结果证明了本文方法的有效性,尤其是在知识图谱非完全的情况下.在未来的工作中,会进一步考虑引入外部信息加强对候选实体建模,进一步优化图谱内实体的特征表示,以达到更好的问答效果.

## References:

- [1] Bollacker K, Evans C, Paritosh P, Sturge T, Taylor J. Freebase: A collaboratively created graph database for structuring human knowledge. In: Proc. of the 2008 ACM SIGMOD Int'l Conf. on Management of Data. Vancouver: ACM, 2008. 1247–1250. [doi: [10.1145/1376616.1376746](https://doi.org/10.1145/1376616.1376746)]
- [2] Lehmann J, Isele R, Jakob M, Jentzsch A, Kontokostas D, Mendes PN, Hellmann S, Morsey M, van Kleef P, Auer S, Bizer C. DBpedia—A large-scale, multilingual knowledge base extracted from Wikipedia. Semantic Web, 2015, 6(2): 167–195. [doi: [10.3233/SW-140134](https://doi.org/10.3233/SW-140134)]
- [3] Mitchell T, Cohen WW, Hruschka ER, Talukdar PP, Yang BS, Betteridge J, Carlson A, Dalvi BB, Gardner M, Kisiel B, Krishnamurthy J, Lao N, Mazaitis K, Mohamed TP, Nakashole N, Antonios E, Ritter A, Samadi M, Settles B, Wang R, Wijaya DT, Gupta A, Chen X, Saparov A, Greaves M, Welling JS. Never-ending learning. Communications of the ACM, 2018, 61(5): 103–115. [doi: [10.1145/3191513](https://doi.org/10.1145/3191513)]
- [4] Bhutani N, Zheng XY, Qian K, Li YY, Jagadish H. Answering complex questions by combining information from curated and extracted knowledge bases. In: Proc. of the 1st Workshop on Natural Language Interfaces. ACL, 2020. 1–10. [doi: [10.18653/v1/2020.nli-1.1](https://doi.org/10.18653/v1/2020.nli-1.1)]
- [5] Sun YW, Zhang LL, Cheng G, Qu YZ. Sparqa: Skeleton-based semantic parsing for complex questions over knowledge bases. In: Proc. of the 34th AAAI Conf. on Artificial Intelligence. New York: AAAI Press, 2020. 8952–8959. [doi: [10.1609/aaai.v34i05.6426](https://doi.org/10.1609/aaai.v34i05.6426)]
- [6] Lan YS, Jiang J. Query graph generation for answering multi-hop complex questions from knowledge bases. In: Proc. of the 58th Annual Meeting of the Association for Computational Linguistics. ACL, 2020. 969–974. [doi: [10.18653/v1/2020.acl-main.91](https://doi.org/10.18653/v1/2020.acl-main.91)]
- [7] Kapanipathi P, Abdelaziz I, Ravishankar S, et al. Leveraging abstract meaning representation for knowledge base question answering. In: Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021. Association for Computational Linguistics, 2020. 38847–38894. [doi: [10.18653/v1/2021.findings-acl.339](https://doi.org/10.18653/v1/2021.findings-acl.339)]
- [8] Xu K, Lai YX, Feng YS, Wang ZG. Enhancing key-value memory neural networks for knowledge based question answering. In: Proc. of the 2019 Conf. of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Vol. 1 (Long and Short Papers). Minneapolis: Association for Computational Linguistics, 2019. 2937–2947. [doi: [10.18653/v1/N19-1301](https://doi.org/10.18653/v1/N19-1301)]
- [9] Sun HT, Bedrax-Weiss T, Cohen W. PullNet: Open domain question answering with iterative retrieval on knowledge bases and text. In: Proc. of the 2019 Conf. on Empirical Methods in Natural Language Processing and the 9th Int'l Joint Conf. on Natural Language Processing (EMNLP-IJCNLP). Hong Kong: Association for Computational Linguistics, 2019. 2380–2390. [doi: [10.18653/v1/D19-1242](https://doi.org/10.18653/v1/D19-1242)]
- [10] Saxena A, Tripathi A, Talukdar P. Improving multi-hop question answering over knowledge graphs using knowledge base embeddings. In: Proc. of the 58th Annual Meeting of the Association for Computational Linguistics. ACL, 2020. 4498–4507. [doi: [10.18653/v1/2020.acl-main.412](https://doi.org/10.18653/v1/2020.acl-main.412)]
- [11] Lan YS, He GL, Jiang JH, Jiang J, Zhao WX, Wen JR. A survey on complex knowledge base question answering: Methods, challenges and solutions. In: Proc. of the 30th Int'l Joint Conf. on Artificial Intelligence. Montreal: IJCAI.org, 2021. 4483–4491. [doi: [10.24963/ijcai.2021/611](https://doi.org/10.24963/ijcai.2021/611)]
- [12] Berant J, Chou A, Frostig R, Liang P. Semantic parsing on freebase from question-answer pairs. In: Proc. of the 2013 Conf. on Empirical Methods in Natural Language Processing. Seattle: Association for Computational Linguistics, 2013. 1533–1544.
- [13] Berant J, Liang P. Imitation learning of agenda-based semantic parsers. Trans. of the Association for Computational Linguistics, 2015, 3: 545–558. [doi: [10.1162/tacl\\_a\\_00157](https://doi.org/10.1162/tacl_a_00157)]

- [14] Xu K, Wu LF, Wang ZG, Yu M, Chen LW, Sheinin V. Exploiting rich syntactic information for semantic parsing with graph-to-sequence model. In: Proc. of the 2018 Conf. on Empirical Methods in Natural Language Processing. Brussels: Association for Computational Linguistics, 2018. 918–924. [doi: [10.18653/v1/D18-1110](https://doi.org/10.18653/v1/D18-1110)]
- [15] Yih WT, Chang MW, He XD, Gao JF. Semantic parsing via staged query graph generation: Question answering with knowledge base. In: Proc. of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th Int'l Joint Conf. on Natural Language Processing (Vol. 1: Long Papers). Beijing: ACL, 2015. 1321–1331. [doi: [10.3115/v1/P15-1128](https://doi.org/10.3115/v1/P15-1128)]
- [16] Luo KQ, Lin FL, Luo XS, Zhu K. Knowledge base question answering via encoding of complex query graphs. In: Proc. of the 2018 Conf. on Empirical Methods in Natural Language Processing. Brussels: Association for Computational Linguistics, 2018. 2185–2194. [doi: [10.18653/v1/D18-1242](https://doi.org/10.18653/v1/D18-1242)]
- [17] Wu K, Zhou XB, Li ZH, Liang XW, Chen WL. Path selection for Chinese knowledge base question answering. Journal of Chinese Information Processing, 2021, 35(9): 113–122 (in Chinese with English abstract). [doi: [10.3969/j.issn.1003-0077.2021.09.011](https://doi.org/10.3969/j.issn.1003-0077.2021.09.011)]
- [18] Song PC, Shan LL, Sun CJ, Lin L. A knowledge base question answering system based on query path ranking. Journal of Chinese Information Processing, 2021, 35(11): 109–117, 126 (in Chinese with English abstract). [doi: [10.3969/j.issn.1003-0077.2021.11.011](https://doi.org/10.3969/j.issn.1003-0077.2021.11.011)]
- [19] Bordes A, Weston J, Usunier N. Open question answering with weakly supervised embedding models. In: Proc. of the 2014 Joint European Conf. on Machine Learning and Knowledge Discovery in Databases. Nancy: Springer, 2014. 165–180. [doi: [10.1007/978-3-662-44848-9\\_11](https://doi.org/10.1007/978-3-662-44848-9_11)]
- [20] Jain S. Question answering over knowledge base using factual memory networks. In: Proc. of the 2016 NAACL Student Research Workshop. San Diego: Association for Computational Linguistics, 2016. 109–115. [doi: [10.18653/v1/N16-2016](https://doi.org/10.18653/v1/N16-2016)]
- [21] Hao YC, Zhang YZ, Liu K, He SZ, Liu ZY, Wu H, Zhao J. An end-to-end model for question answering over knowledge base with cross-attention combining global knowledge. In: Proc. of the 55th Annual Meeting of the Association for Computational Linguistics (Vol. 1: Long Papers). Vancouver: ACL, 2017. 221–231. [doi: [10.18653/v1/P17-1021](https://doi.org/10.18653/v1/P17-1021)]
- [22] Das R, Zaheer M, Reddy S, McCallum A. Question answering on knowledge bases and text using universal schema and memory networks. In: Proc. of the 55th Annual Meeting of the Association for Computational Linguistics (Vol. 2: Short Papers). Vancouver: Association for Computational Linguistics, 2017. 358–365. [doi: [10.18653/v1/P17-2057](https://doi.org/10.18653/v1/P17-2057)]
- [23] Xiong WH, Yu M, Chang SY, Guo XX, Wang WY. Improving question answering over incomplete KBs with knowledge-aware reader. In: Proc. of the 57th Annual Meeting of the Association for Computational Linguistics. Florence: Association for Computational Linguistics, 2019. 4258–4264. [doi: [10.18653/v1/P19-1417](https://doi.org/10.18653/v1/P19-1417)]
- [24] Han JL, Cheng B, Wang X. Two-phase hypergraph based reasoning with dynamic relations for multi-hop KBQA. In: Proc. of the 29th Int'l Joint Conf. on Artificial Intelligence. Yokohama: IJCAI.org, 2021. 500.
- [25] Zhang TH, Li TT, Zhang YK. Multi-hop Chinese knowledge question answering method based on knowledge graph embedding. Journal of Jilin University (Science Edition), 2022, 60(1): 119–126 (in Chinese with English abstract). [doi: [10.13413/j.cnki.jdxblxb.2020417](https://doi.org/10.13413/j.cnki.jdxblxb.2020417)]
- [26] Devlin J, Chang MW, Lee K, Toutanova K. BERT: Pre-training of deep bidirectional transformers for language understanding. In: Proc. of the 2019 Conf. of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Vol. 1 (Long and Short Papers). Minneapolis: Association for Computational Linguistics, 2018. 4171–4186. [doi: [10.18653/v1/N19-1423](https://doi.org/10.18653/v1/N19-1423)]
- [27] Cao Z, Qin T, Liu TY, Tsai MF, Li H. Learning to rank: From pairwise approach to listwise approach. In: Proc. of the 24th Int'l Conf. on Machine Learning. Corvalis: ACM, 2007. 129–136. [doi: [10.1145/1273496.1273513](https://doi.org/10.1145/1273496.1273513)]
- [28] Liu YH, Ott M, Goyal N, Du JF, Joshi M, Chen DQ, Levy O, Lewis M, Zettlemoyer L, Stoyanov V. RoBERTa: A robustly optimized BERT pretraining approach. arXiv:1907.11692, 2019.
- [29] Trouillon T, Welbl J, Riedel S, Gaussier É, Bouchard G. Complex embeddings for simple link prediction. In: Proc. of the 33rd Int'l Conf. on Machine Learning. New York: JMLR.org, 2016. 2071–2080.
- [30] Hartigan JA, Wong MA. Algorithm AS 136: A K-means clustering algorithm. Journal of the Royal Statistical Society. Series C (Applied Statistics), 1979, 28(1): 100–108.
- [31] Ester M, Kriegel HP, Sander J, Xu XW. A density-based algorithm for discovering clusters in large spatial databases with noise. In: Proc. of the 2nd Int'l Conf. on Knowledge Discovery and Data Mining. Portland: AAAI Press, 1996. 226–231.
- [32] Yih WT, Richardson M, Meek C, Chang MW, Suh J. The value of semantic parse labeling for knowledge base question answering. In: Proc. of the 54th Annual Meeting of the Association for Computational Linguistics (Vol. 2: Short Papers). Berlin: ACL, 2016. 201–206. [doi: [10.18653/v1/P16-2033](https://doi.org/10.18653/v1/P16-2033)]
- [33] Talmor A, Berant J. The web as a knowledge-base for answering complex questions. In: Proc. of the 2018 Conf. of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Vol. 1 (Long Papers). New Orleans: Association for Computational Linguistics, 2018. 641–651. [doi: [10.18653/v1/N18-1059](https://doi.org/10.18653/v1/N18-1059)]

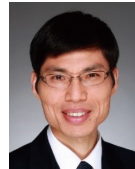
- [34] Miller A, Fisch A, Dodge J, Karimi AH, Bordes A, Weston J. Key-value memory networks for directly reading documents. In: Proc. of the 2016 Conf. on Empirical Methods in Natural Language Processing. Austin: Association for Computational Linguistics, 2016. 1400–1409. [doi: [10.18653/v1/D16-1147](https://doi.org/10.18653/v1/D16-1147)]
- [35] Yan YM, Li RM, Wang SR, Zhang HZ, Zan DG, Zhang FZ, Wu W, Xu WR. Large-scale relation learning for question answering over knowledge bases with pre-trained language models. In: Proc. of the 2021 Conf. on Empirical Methods in Natural Language Processing. Punta Cana: Association for Computational Linguistics, 2021. 3653–3660. [doi: [10.18653/v1/2021.emnlp-main.296](https://doi.org/10.18653/v1/2021.emnlp-main.296)]
- [36] Chen ZY, Chang CH, Chen YP, Nayak J, Ku LW. UHop: An unrestricted-hop relation extraction framework for knowledge-based question answering. In: Proc. of the 2019 Conf. of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Vol. 1 (Long and Short Papers). Minneapolis: Association for Computational Linguistics, 2019. 345–356. [doi: [10.18653/v1/N19-1031](https://doi.org/10.18653/v1/N19-1031)]
- [37] Bhutani N, Zheng XY, Jagadish HV. Learning to answer complex questions over knowledge bases with query composition. In: Proc. of the 28th ACM Int'l Conf. on Information and Knowledge Management. Beijing: ACM, 2019. 739–748. [doi: [10.1145/3357384.3358033](https://doi.org/10.1145/3357384.3358033)]
- [38] Gu Y, Kase S, Vanni M, Sadler B, Liang P, Yan XF, Su Y. Beyond I.I.D.: Three levels of generalization for question answering on knowledge bases. In: Proc. of the 2021 Web Conf. Ljubljana: ACM, 2021. 3477–3488. [doi: [10.1145/3442381.3449992](https://doi.org/10.1145/3442381.3449992)]

#### 附中文参考文献:

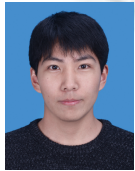
- [17] 吴锬, 周夏冰, 李正华, 梁兴伟, 陈文亮. 中文知识库问答中的路径选择. 中文信息学报, 2021, 35(9): 113–122. [doi: [10.3969/j.issn.1003-0077.2021.09.011](https://doi.org/10.3969/j.issn.1003-0077.2021.09.011)]
- [18] 宋鹏程, 单丽莉, 孙承杰, 林磊. 基于查询路径排序的知识库问答系统. 中文信息学报, 2021, 35(11): 109–117, 126. [doi: [10.3969/j.issn.1003-0077.2021.11.011](https://doi.org/10.3969/j.issn.1003-0077.2021.11.011)]
- [25] 张天杭, 李婷婷, 张永刚. 基于知识图谱嵌入的多跳中文知识问答方法. 吉林大学学报(理学版), 2022, 60(1): 119–126. [doi: [10.13413/j.cnki.jdxblxb.2020417](https://doi.org/10.13413/j.cnki.jdxblxb.2020417)]



陈跃鹤(1998—), 男, 硕士生, CCF 学生会会员, 主要研究领域为知识图谱问答, 自然语言处理.



陈文亮(1977—), 男, 博士, 教授, 博士生导师, CCF 专业会员, 主要研究领域为自然语言处理, 知识图谱.



贾永辉(1997—), 男, 硕士, 主要研究领域为知识图谱问答, 自然语言处理.



张民(1970—), 男, 博士, 教授, 博士生导师, CCF 高级会员, 主要研究领域为自然语言处理, 机器翻译, 人工智能.



谈川源(1999—), 男, 硕士生, CCF 学生会会员, 主要研究领域为知识图谱问答, 自然语言处理.