基于多角度注意力机制的单一事实知识库问答方法

罗 达 苏锦钿 李鹏飞
(华南理工大学计算机科学与工程学院 广州 511400)

摘 要 近年来,基于知识库的问答受到了广泛的关注,成为了一个重要的自然语言处理任务。在基于知识库的问答任务中,简单问题是指能够通过知识库的单一事实进行回答的问题。针对简单问题的回答,现有的解决方法主要是将问题和知识库事实映射到同一向量空间中,然后通过计算问题和事实之间的相似度来得到答案,但这种方法会损失原始单词的部分语义交互信息。针对该问题,文中提出了一种基于多角度注意力机制的关系检测模型,从多个角度对问题和知识库关系的相关性进行了建模,从而保留了更多的原始交互信息,并提高了模型的准确率。此外,为了减小噪音的影响并提高实体识别的准确率,在实体链接过程中提出结合基于语言模型的动态词向量和单词的词性特征对问题进行表征。实验结果表明,所提方法在基于 FB2M 和 FB5M 的 SimpleQuestions 数据集上分别获得了 78.9%和 78.3%的准确率,能够很好地反映问题与知识库关系之间的语义相关性,并提升了单一事实知识库问答的准确率。

关键词 问答,知识库,深度学习,注意力机制,自然语言处理
中图法分类号 TP183 **文献标识码** A **DOI** 10.11896/jsjkx.190400071

Multi-view Attentional Approach to Single-fact Knowledge-based Question Answering

LUO Da SU Jin-dian LI Peng-fei
(School of Computer Science and Engineering, South China University of Technology, Guangzhou 511400, China)

Abstract Knowledge base question answering (KB-QA) has received extensive attention in recent years, and becomes an important natural language processing task. In the knowledge base question answering task, simple question refers to the question that can be answered by a single-fact of the knowledge base. For this task, the existing approaches mainly map the question and the KB fact into a common vector space and calculate their similarity to get the answer. But this approach would lose part of the semantic interaction information of the original words. To solve this problem, a multi-view attention-based relation detection approach was proposed, which aims to model the correlation between the question and the KB relation from multiple perspectives, and preserves more original interaction information so as to improve the accuracy of the approach. In addition, in order to alleviate the impacts of noisy data and improve the accuracy of entity recognition, this paper also encoded the question by combining the dynamic word vectors based on the language model and the part-of-speech feature of the word during the process of entity linking. Finally, experiments conducted on the SimpleQuestions dataset based on FB2M and FB5M achieve the accuracy results of 78.9% and 78.3%, respectively, which illustrative the effectiveness of the proposed approach for reflecting the semantic correlation between the question and the KB relation, and reflect the improvements of the accuracy for single-fact KBQA.

Keywords Question answering, Knowledge base, Deep learning, Attention mechanism, Natural language processing

1 引言

近年来,随着 YAGO[1],Freebase[2] 和 Dbpedia[3] 等大规模知识库的出现,基于知识库的问答系统(Knowledge Based Question Answering,KB-QA)受到了学术界和工业界的广泛关注,其主要原因是可以将自然语言问题转化为三元组,并通过查询知识库的方式来获得直接的答案。但由于基于自然语言表示的问题与知识库中的事实表示往往存在较大的语义鸿沟,将问题映射到知识库中的特定三元组存在困难,因此该任务仍面临很多问题有待解决和完善。

目前,基于知识库的问答研究主要存在两种方法:基于语义解析的方法和基于深度学习的方法。基于语义解析的方法首先将自然语言问题映射为人工定义的逻辑形式,然后再将其转化为特定的查询逻辑语言(如 SPARQL),最后在知识库

到稿日期:2019-04-11 返修日期:2019-06-28 本文受广东省自然科学基金(2015A030310318),广东省科技厅应用型科技研发专项资金项目(20168010124010),广东省医学科学技术研究基金项目(A2015065)资助。

罗 达(1994—),男,硕士生,主要研究方向为机器学习和自然语言处理,E-mail:472447327@qq.com;苏锦钿(1980—),男,博士,副教授,CCF 会员,主要研究方向为机器学习和自然语言处理,E-mail:sujd@scut.edu.cn(通信作者);李鹏飞(1993—),男,硕士生,主要研究方向为机器学习、自然语言处理。

中查询并得出答案,相关的研究请参见文献[4-6]。基于语义解析的方法虽然比较符合人类的直觉,但是需要依赖较多的人工标注数据和模板,缺乏灵活性和通用性。而基于深度学习的方法则具有更强的适应性,其核心是通过使用编码器对问题和知识库的事实进行编码,使得正确的事实与问题在编码空间上更加相近,最后根据相似度计算得到答案。目前,大部分基于深度学习的 KB-QA 系统首先使用 n-gram 方法或者序列标注方法对问题的文本进行实体抽取,然后查询知识库以得到一系列候选的知识库事实,并通过训练深度神经网络模型来计算问题和知识库事实间的相关性,最后结合以上步骤得到最终答案。总体来说,基于知识库的问答系统主要包含两个子任务:1)实体链接,即抽取问题中出现的实体文本,并将其链接到知识库中的实体对象;2)关系检测,即检测候选关系与问题的相关性。

早期的实体链接方法主要是使用 n-gram 方法对问题中的词组进行遍历,在知识库中查询 name 属性包含该词组的实体作为候选实体[7-9]。但是,该方法在查询知识库时耗时过长,且容易引入过多不相关候选实体而产生噪声数据。为了减少候选实体的数量,Hakimov 等[10]使用一个标准 LSTM(Long Short-Term Memory)序列标注模型对问题进行标注,然后在知识库中查询实体名或实体别名与标注区域匹配的候选实体。Dai 等[11]和 Yin 等[12]均使用一个双向 LSTM+CRF 的标注模型来确定问题中的实体提及范围,以 n-gram 的方式遍历标注结果,并在知识库中查询相应实体,最终得到候选实体。上述方法虽然能够显著减少不相关实体的数量,但是序列标注错误容易导致正确实体无法召回。此外,Yih 等[13]使用一种对带噪音的短文本提取主题词的工具 S-MART[14]来提取问题中的实体。而 Qu 等[15]在使用序列标注模型识别实体的基础上,还结合启发式算法来解决序列标注模型标注不准确所带来的问题,从而提高实体链接的召回率。从上述工作可以看出,基于序列标注的方法已在实体链接任务中取得了较好的效果。但对于知识库问答任务来说,问题文本的用词和表达方式是非常多样的,而当前的研究并没有充分利用外部信息来进一步提升模型的性能。针对该不足,本文提出一种结合 ELMo(Embeddings from Language Models)语言模型和词性特征的双向 GRU(Gated Recurrent Units)网络,其通过引入丰富的外部语义知识来提升序列标注的准确率以及序列标注结果对实体的召回能力。同时,为了减小序列标注错误所带来的影响,在无法根据实体提及范围查询到候选实体时提出使用滑动窗口的方法来进一步扩大搜索范围。

KB-QA 中的关系检测与传统的关系检测有一定的区别:传统的关系检测任务中关系数量一般小于 100;而 KB-QA 任务中即使是较小的知识库 Freebase2M[2],其关系的数量也高达 6000。Yu[16]等对双向多角度匹配(Bilateral Multiple Perspective Matching,BiMPM)[17]模型进行了修改,并将其应用于关系检测任务中。BiMPM 虽然对问题和关系进行了双向的特征提取,但在其上下层中都添加了一个 LSTM 网络以提取更高级的特征,因此可能造成原始信息的损失。Zhang 等[18]提出了基于注意力机制的单词级别的交互模型,并在 SimpleQuestions[8]和 WebQuestions[19]数据集上证明了该模型的有效性。Golub 等[9]提出了一种基于字符级别的方法,并在网络结构中使用注意力机制来更好地处理长问题序列问题。Qu 等[15]使用一个基于注意力机制的 RNN(Recurrent Neural Network)模型来获得问题模式和关系间的语义级别相关性,并通过 CNN 网络获得问题模式与关系间的单词级别相关性,从而从语义级别和单词级别两方面对问题模式和关系的相关性进行建模。上述工作均证明了基于注意力机制的 RNN 可以显著提升关系检测的准确率。但是单一的注意力机制难以刻画问题模式和知识库关系间复杂和高级的语义联系,因此本文提出一种基于多角度注意力机制的关系检测模型。所提模型结合多层 GRU 和 LSTM 网络,利用注意力机制从多个角度对不同粒度的文本进行建模,目的是提取问题和关系之间更丰富的语义关系,从而提高实体关系检测的准确率。

综上所述,本文的主要贡献有:

1)提出了一种结合 ELMo 语言模型的动态词向量以及词性特征的实体链接方法,将实体链接任务视为序列标注任务,并通过一种滑动窗口的方法来减小序列标注错误带来的影响;

2)提出了一种利用多角度注意力机制的关系检测方法,利用注意力机制从不同的角度提取问题模式和关系的不同层次语义关系,提高关系检测模型的准确率;

3)将上述两种方法结合后应用于基于单一事实的 KB-QA 任务上,最后利用 SimpleQuestions 数据集验证其有效性。

2 模型实现

2.1 任务定义

本文基于结构化知识库 Freebase,对知识库实体和关系进行查询。Freebase 中的事实是以(subject,relation,object)的三元组形式进行描述,如(m/011152n,music/artist/album,m/02z5mj8)为一条事实,其中第一项和第三项分别为头实体和尾实体,用唯一的实体 id 表示,第二项为它们之间的关系。根据知识库查询结果可知,头实体"m/011152n"对应的实体为"atlantic starr",尾实体"m/02z5mj8"对应的实体为"brilliance"。因此,该条知识可用自然语言描述为"atlantic starr 发行过 brilliance 这一张音乐专辑"。

SimpleQuestions 是 Facebook 于 2015 年公开的一份针对单一事实的知识库问答语料,来自于人工标注的数据。语料中的每一个问题对应着 Freebase 知识库中的一条事实。

简单问题是能够通过知识库中的一个三元组来回答的问题,因此只需对问题和三元组中的两个元素(subject,relation)进行匹配,若匹配结果正确,则可通过知识库查询确定唯一的第三个元素 object,从而得到正确答案。然而,由于知识库存在大量的实体和关系,直接对问题和(subject,relation)对进行匹配并不现实,因此可将该任务分解成以下两个子任务。

1)实体链接。给定一个问题 Q,首先找出问题的实体提及 X,接着从知识库中查询全名或别名和实体提及 X 匹配的所有实体,然后加入候选实体集 S,并将所有包含候选实体的三元组中的关系提取出来后加入候选关系集 R。

2)关系排序。首先将问题中的实体提及用特殊符号进行替换,得到问题模式 Q_p,从而屏蔽实体提及对关系排序模块的影响;然后对问题模式 Q_p 和 R 中的所有候选关系的相关性进行建模打分;最后选出与 Q_p 相关性最高的候选关系 r_f。

3)答案生成。若候选实体集 S 中只有一个候选实体 s_f 连接候选关系 r_f,则把根据 s_f 和 r_f 所确定的三元组(s_f, r_f, o_f)作为答案;若有多个候选实体均连接候选关系 r_f,则选取包含关系数最多的实体作为最终实体 s_f,并确定最终答案。

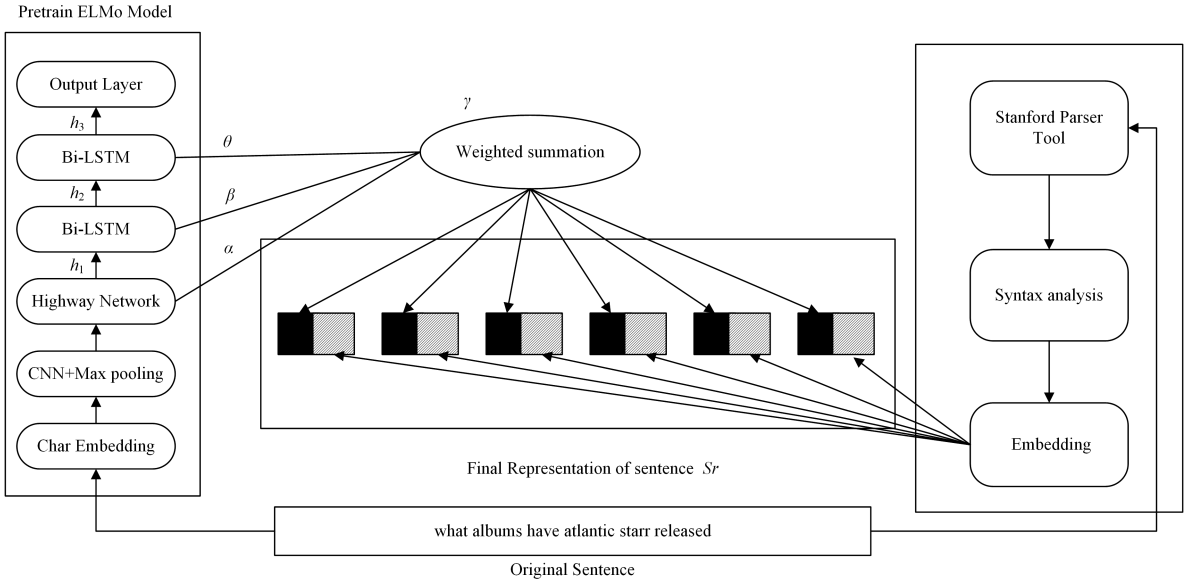

图 1　结合 ELMo 预训练词向量和词性特征的词表示层

Fig. 1　Word representation layer with ELMo pre-trained word vector and part of speech

2.2　实体链接模型

实体链接模型的任务是从问题 Q 中提取出实体提及 X,并在知识库中查询名字为 X 的知识库事实,从而确定候选实体集 H 和候选关系集 R。由于实体提及是由若干个连续单词组成的,因此实体链接模型任务可看作是一个预测问题 Q 中的每一个单词是否为实体提及的 0-1 序列标注问题。考虑到问题文本用词和表达方式的多样性,本文使用一个结合 ELMo 语言模型和词性(Part-of-Speech, POS)特征的双向 GRU 网络。

单词的表示由两部分组成,如图 1 所示,句子的最终表示为:

$$S_r=[R_1,R_1,\cdots,R_n] \tag{1}$$

其中,n 为句子单词数,每个单词的表示 R_i 由两部分拼接而成,分别为预训练 ELMo 模型中的动态词向量 R_i^{ELMo} 和由 Stanford Parser 工具解析出的词性对应的特征向量 R_i^{POS},即:

$$R_i=[R_i^{\mathrm{ELMo}};R_i^{\mathrm{POS}}] \tag{2}$$

预训练 ELMo 语言模型如图 1 中的左侧方框所示。为了使词向量包含不同粒度的词义信息,本文使用 ELMo 模型内部的三层隐状态,分别是 Highway 网络的输出层 h_1 和两层双向 LSTM 网络的隐状态输出 h_2,h_3,并用 3 个可学习调整的参数 α,β,θ 来对这 3 个输出进行加权组合,同时在最外层乘以缩放参数 γ 以适应整个模型的数据分布,将最终结果作为单词的表示。第 i 个词的 ELMo 向量计算公式如下:

$$R_i^{\mathrm{ELMo}}=\gamma\cdot(\alpha\cdot h_{1,i}+\beta\cdot h_{2,i}+\theta\cdot h_{3,i}) \tag{3}$$

单词词性特征的提取如图 1 中的右侧方框所示。首先使用 Stanford Parser 工具对句子进行句法分析,生成句法分析树,句法树的叶子结点即为对应单词的词性;然后将词性映射到随机初始化的向量空间中,将其作为单词的词性特征;最后将拼接起来的单词特征 R_i 传入一个 3 层的双向 GRU 网络进行序列标注。

经过序列标注模型后,若序列标注的结果无法匹配到知识库中的实体,则通过以下方法确定问题的实体提及以及候选实体集 H。

1)将输出标签为 1 的相邻单词连起来,得到候选实体字符串 S。若识别出多个候选实体字符串,则取最后一个。

2)在知识库查询全名或别名为字符串 S 的所有实体 id,并将结果添加到候选实体集 H 中。

3)若经过步骤 2)后候选实体集 H 仍然为空,则以候选实体字符串 S 为中心,用大小等于 S 的窗口按照[−1,+1,−2,+2]的距离进行滑动("−"表示向左滑动,"+"表示向右滑动),每滑动一次,得到一个新的字符串 S',并查询知识库中全名或别名为 S' 的实体 id;若查询结果不为空,则将结果添加到候选实体集 H 中,并停止滑动,同时将 S' 作为实体提及。

在得到问题的实体提及后,用 $\#entity\#$ 替换问题 Q 中的实体字符串 S,得到问题模式 Q_p,将 H 中所有实体链出的关系添加到候选关系集 R 中。

2.3　关系检测模型

关系检测模型的任务是对候选关系集 R 中的关系进行排序,选取最有可能的关系作为知识库查询三元组的谓语,排序的依据是问题模式 Q_p 和候选关系 R_c 的相关程度。本文提出的模型使用注意力机制从单词级别和关系级别两个粒度对问题模式和候选关系的相关性进行建模(见图 2)。模型结合了单词级别的相关性计算 $score_1$,$score_2$,$score_3$ 和关系级别的相关性计算 $score_4$,最后用一个线性层学习 4 个分数的权

重，并对 4 个分数加权相加，得到关系的最终得分。下文将详细介绍模型的细节。

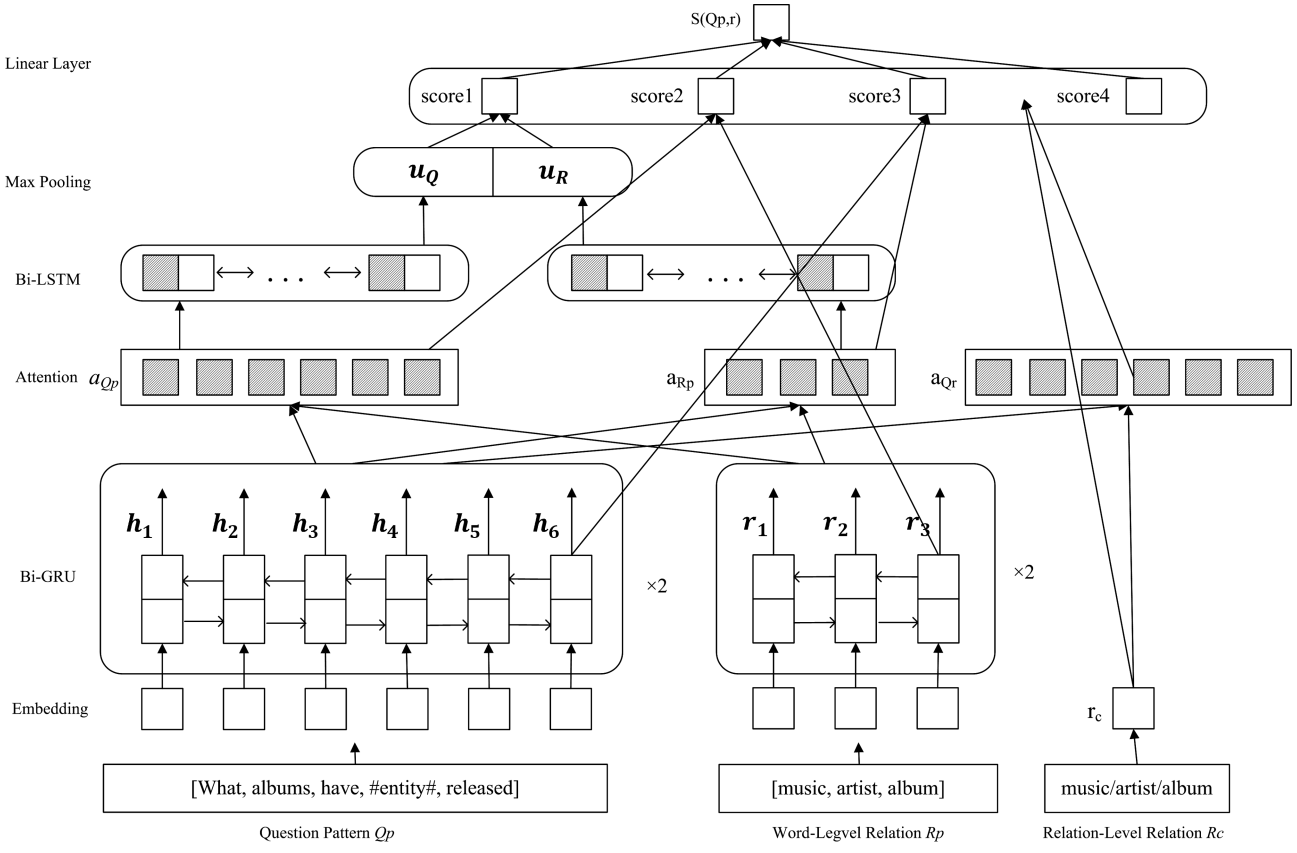

图 2 基于多角度注意力机制的关系检测模型

Fig. 2 Relation detection with multi-view attention mechanism

2.3.1 表示层和编码层

对于问题模式 Q_p，先将每个单词用 GloVe 预训练词向量表示，再将词向量传入一个两层双向 GRU 网络进行编码。对于单词级别的关系 R_p，与问题模式 Q_p 一样，先将分词后的关系用预训练词向量表示，再传入另一个两层双向 GRU 网络进行编码。对于关系级别的关系 R_c，不对其进行分词，直接将关系 id 映射为随机初始化的可训练关系向量。

编码层的最终输出为问题模式 Q_p 的表示 $H_{1:n}=[h_1,h_2,\cdots,h_n]$、单词级别关系 R_p 的表示 $R_{1:m}=[r_1,r_2,\cdots,r_m]$，以及关系级别关系 R_c 的表示 r_c，其中 n 和 m 分别为 Q_p 和 R_p 中的单词数量。

2.3.2 注意力层

在注意力层，以相同的计算方式分别计算 Q_p 对于 R_p、R_p 对于 Q_p 以及 Q_p 对于 R_c 这 3 个基于注意力机制的表示。以 Q_p 的编码层输出对于单词级别关系 R_p 的编码层输出的注意力机制表示为例，对于问题模式中的第 i 个单词，关系 R_p 的注意力表示 p_i 的计算如下：

$$p_i=\sum_{j=1}^{m}a_{ij}r_j \tag{4}$$

$$a_{ij}=\frac{\exp(w_{ij})}{\sum_{k=1}^{m}\exp(w_{ik})} \tag{5}$$

$$w_{ij}=\boldsymbol{v}^{\mathrm{T}}\cdot h_i\cdot\boldsymbol{W}^{\mathrm{T}}\cdot r_j \tag{6}$$

其中，a_{ij} 表示 Q_p 中第 i 个单词对 R_c 中第 j 个单词的注意力权重；w_{ij} 表示 Q_p 中第 i 个单词 h_i 和 R_c 中第 j 个单词 r_j 的点乘相似度；v 和 W 是两个可学习的参数矩阵，用于对 h_i 和 r_j 进行线性变换。

注意力层的最终输出为 Q_p 对 R_p 的注意力表示 $a_{Qp}=[q_{p1},q_{p2},\cdots,q_{pn}]$、$R_p$ 对 Q_p 的注意力表示 $a_{Rp}=[q_{r1},q_{r2},\cdots,q_{rm}]$，以及 Q_p 对 R_c 的注意力表示 $a_{Qc}=[q_{c1},q_{c2},\cdots,q_{cn}]$。

沿用本节的符号定义，对编码层的输出与注意力层相应的输出进行相似度计算，得到 $score_2$，$score_3$，$score_4$，具体计算公式如下：

$$score_2=r_m\cdot\sum_{i=1}^{n}q_{pi} \tag{7}$$

$$score_3=h_n\cdot\sum_{j=1}^{m}q_{ri} \tag{8}$$

$$score_4=r_c\cdot\sum_{i=1}^{n}q_{ci} \tag{9}$$

2.3.3 交互层

为了提取编码层和注意力层的高级交互特征，同时保留一定的原始信息，首先对 Q_p 和 R_c 的编码层和注意力层的输出进行保留、相减和相乘的交互操作，并将结果级联拼接在一起，从而形成新的特征编码 V^Q 和 V^R，其中第 i 项的计算公式如下：

$$v_i^Q=[h_i;q_{pi};(h_i-q_{pi});(h_i*q_{pi})] \tag{10}$$

$$v_i^R=[r_i;q_{ri};(r_i-q_{ri});(r_i*q_{ri})] \tag{11}$$

得到基于文本交互的特征编码 V^Q 和 V^R 之后，将其传入一个双向 LSTM 网络以捕获高层次的上下文交互信息。然后，再连接一层最大池化层，使得不定长的文本序列转化为定长的最终表示 u_Q 和 u_R。最后，对 u_Q 和 u_R 进行点乘以计算相似度，得到 $score_1$。

2.3.4　组合层

经过不同粒度的相似度计算之后,可以得到 4 个问题模式和关系间的相关性分数($score_1$,$score_2$,$score_3$,$score_4$),其分别代表了不同粒度的相似度。最后将这 4 个分数级联拼接起来,并利用一个线性层来学习它们对总分数的贡献度,从而得到最终的分数 $S(Q_p,r)$。

本文采用 marginRankingLoss 损失函数,目的是最大化正确关系和其他候选关系的差距:

$$loss(Q_p,r^+,r^-)=\sum_{(Q_p,r^+)\in D}\max(0,\gamma+S(Q_p,r^-)-S(Q_p,r^+)) \tag{12}$$

其中,γ 为超参数,表示正负样本的优化间隔。

3　实验与结果

本节利用 SimpleQuestions 数据集来验证本文所提方法的总体效果;同时,还对实体链接模型和关系检测模型进行实验分析,并证明其有效性。

3.1　数据集和模型参数的设置

SimpleQuestions 数据集为一个针对单一事实的知识库问答数据集,共包含 108442 条人工撰写的问题,这些问题是基于 Freebase 知识库中的事实所构建的,每一个问题的答案对应知识库中的一个三元组。将数据集的 70%(75910 条数据)作为训练集,10%(10845 条数据)作为验证集,20%(21687 条数据)作为测试集。使用的知识库为 Freebase 知识库的子集 FB2M 和 FB5M,前者约包含 215 万个实体,后者约包含 490 万个实体。

实体链接模型的参数如表 1 所列,其中词性向量是随机初始化的。

表 1　实体链接模型的参数设置

Table 1　Parameter setting of entity linking model

参数	值
ELmo 词向量维度	1024
词性特征向量维度	32
BiGRU 层数	3
BiGRU 隐藏层大小	128
drop-out 比例	0.5
Batch size	1
优化器	Adam

关系检测模型的参数如表 2 所列,其中关系向量是随机初始化的。

表 2　关系检测模型参数设置

Table 2　Parameter setting of relation detecting model

参数	值
GloVe 词向量维度	300
问题模式 GRU 层数	2
问题模式 GRU 隐藏层维度	128
关系 GRU 层数	2
关系 GRU 隐藏层维度	128
LSTM1、LSTM2 层数	1
LSTM1、LSTM2 隐藏层维度	128
关系向量维度	300
Batch size	50
负采样数量	64
优化器	Adam

3.2　模型的总体效果

将本文总体模型与近年来在 SimpleQuestions 数据集上表现较好的 6 个典型模型进行对比,结果如表 3 所列。这些模型分别为 SimpleQuestions 数据集发布者公开的 Baseline 模型——MemNN[8]、Golub 等提出的基于字符级别注意力的模型[9]、Dai 等提出的 Cfo 模型[11]、Yin 等提出的 AMPCNN 模型[12]、Qu 等提出的 AR-SMCNN 模型[15],以及 Hao 等提出的基于联合事实模型[20]。

表 3　不同模型在 SimpleQuestions 数据集上的准确率

Table 3　Accuracies of different models on SimpleQuestions

(单位:%)

模型	FB2M	FB5M
Bordes et al.(2015)	63.9	62.7
Golub and He(2016)	70.9	70.3
Dai et al.(2016)	—	75.7
Yin et al.(2016)	76.4	75.9
Qu et al.(2018)	77.9	76.8
Hao et al.(2018)	**80.2**	—
本文方法	78.9	**78.3**

表 3 中各对比模型的结果均为对应文献中公开的结果。从中可以看出:1)本文方法在 FB2M 知识库的背景下取得了当前第二高的准确率 78.9%,在 FB5M 知识库的背景下取得了当前最高的准确率 78.3%。Hao 等采用联合事实共同训练模型,这在一定程度上减少了实体链接和关系排序两个阶段的"错误累计",并在 FB2M 知识库的背景下取得了最高的准确率 80.2%。而本文针对两个阶段分别提出了相应的模型,虽然最终的准确率为 78.9%,比 80.2%低了 1.3%,但本文方法将知识库问答任务进行解耦,更具有通用性,既可组合起来解决 KB-QA 的问题,也可单独解决实体识别和语义匹配问题,另外也有助于定位模型的问题,从而进行针对性的优化。2)在实体链接过程中,Bordes 等的方法和 Golub 等使用结合规则的 n-gram 遍历方法,Yin 等、Qu 等的方法和本文方法均使用序列标注神经网络。显然,后者的准确率比前者的准确率有大幅提高,这说明 n-gram 方法会导致候选实体集数量过大,从而带来干扰项,在很大程度上影响了整体系统的准确率。3)在关系排序过程中,Golub 等、Yin 等、Qu 等的方法和本文方法均使用了注意力机制,但本文方法的准确率最高。与前三者的单一注意力机制相比,本文使用了 3 种注意力机制,这使得两段文本进行了多层次的交互,表征更为丰富,从而提升了效果;此外,Golub 等和 Yin 等还分别基于字符级别对句子进行编码,并解决了 OOV 问题。相比之下,本文方法虽然没有使用字符级别的编码,但多角度注意力机制弥补了这个缺陷,取得了更高的准确率。

3.3　实体链接模型的效果

本文的实体链接模型为一个序列标注模型。为验证模型的效果,下面分别对比序列标注的准确率以及实体的召回率。

对比的基准模型包括:Dai 等提出的 Cfo 模型[11]和 Qu 等提出的 AR-SMCNN[15]。其中,Cfo-RNN-CRF 和 Cfo-CRF 的结果来自文献[11]。AR-SMCNN -Bi-GRU 的实验结果是根据文献[15]提供的开源代码复现所得。为了对比不同层数 GRU 网络的效果,用 Ours-N-layer(N=1,2,3,4,5)分别表示

本文的实体链接模型中所堆叠的不同 GRU 层数。各序列标注模型的准确率如表 4 所列。

表 4 序列标注模型的准确率

Table 4 Accuracies of sequence labeling models

模型	句子级别准确率/%
Cfo-CRF	91.2
Cfo-RNN-CRF	95.5
AR-SMCNN-Bi-GRU	93.8
Ours-1-layer	95.8
Ours-2-layer	96.2
Ours-3-layer(本文模型)	**96.5**
Ours-4-layer	96.4
Ours-5-layer	96.4

从表 4 可以看出:1)本文提出的结合 ELMo 和词性特征的序列标注模型 Ours-3-layer 取得了 96.5%的准确率,超过了传统方法(Cfo-CRF)和基于深度学习的方法(Cfo-RNN-CRF,AR-SMCNN-Bi-GRU),证明了本文实体链接模型的有效性。2)当 GRU 网络的堆叠层数从 1 变化到 3 时,模型准确率不断提高,这说明深层次的网络能获得更丰富的句子表征,从而提升了任务效果。但随着层数继续从 3 增加到 5,模型准确率并没有继续提高,反而下降了 0.1%,说明 GRU 网络的层次并不是越多越好,因为不同任务对句子表征粒度的需求不同。从实验结果来看,本任务最适合的层数为 3。

在实体召回率的实验中,我们对比了几个做了类似实验的模型,包括 Golub[9],Yin[12],Qu[15]等所提出的模型。表 5 列出了取 Top K 的候选实体所取得的召回率结果。

表 5 Top K 候选实体的召回率

Table 5 Recall of Top K candidate entities

(单位:%)

Top	Golub(2016)	Yin(2016)	Qu(2018)	本文方法
1	52.9	73.6	74.5	**74.6**
5	—	85.0	86.0	**86.4**
10	74.0	87.4	88.5	**89.2**
20	77.8	88.8	90.2	**91.3**
50	82.0	90.4	91.9	**93.1**
100	85.4	91.6	93.1	**94.3**

为了对比实体链接模型的效果,与 Qu 等的方法一样,本文根据实体链出关系的数量对实体进行了降序排序。从表 5 可以看出:1)本文的实体抽取方法在 Top1－Top100 的召回率上均取得了最佳成绩,充分证明了本文序列标注模型＋启发式算法召回实体的有效性。2)Yin 等、Qu 等的方法和本文方法均使用了基于神经网络的序列标注模型来抽取实体,得到的 Top1－Top100 的实体召回率均比 Golub 等的 n-gram 方法的召回率高出许多。这证明了序列标注模型的有效性。3)理论上,只要使用所有可能的 n-gram 进行问句遍历,就可以取得 100%的实体召回率。但一方面这样操作的计算代价太大,另一方面会引入大量数据噪声。Golub 等在 n-gram 遍历的基础上,根据自定义规则过滤掉某些实体。而本文方法则在采用模型进行序列标注后,一旦匹配到知识库中的实体,就停止添加候选实体。这不仅提升了模型的计算速度,还避免了召回过多噪声实体。

3.4 关系检测模型的效果

为了验证关系检测任务,Yin 等[12]基于 SimpleQuestions 中的问题构建了一个单独的用于关系检测任务的数据集,并通过关系检测准确率来评估模型效果。

基于该数据集,本文与几个当前在该数据集中表现最好的模型进行对比,包括 Yih 等[13]提出的 BiCNN、Yin 等[12]提出的 AMPCNN、Yu 等[21]提出的 HR-BiLSTM、Qu 等[15]提出的 AR-SMCNN。实验的对比结果如表 6 所列。

表 6 SimpleQuestion 关系检测任务的准确率

Table 6 Accuracies on SimpleQuestions relation detection tasks

模型	关系编码粒度	准确率/%
BiCNN	字符级别	90.0
AMPCNN	单词级别	91.3
HR-BiLSTM	单词级别＋关系级别	93.3
AR-SMCNN	单词级别＋切分的关系级别	**93.7**
Ours (score1＋score2＋score3＋score4)	单词级别＋关系级别	**93.5**
Ours(without score1)	单词级别＋关系级别	92.5
Ours(without score2 and score3)	单词级别＋关系级别	93.1
Ours(without score4)	单词级别＋关系级别	92.9

从表 6 可以看出:1)本文提出的基于多角度注意力机制的关系检测模型取得了近年来第二好的成绩,准确率比当前最佳模型 AR-SMCNN 的结果低 0.2%。两者均对关系进行了切分,不同的是 AR-SMCNN 额外将关系按自定义规则切分成两部分,这虽然使得模型表现更适应 SimpleQuestions 数据集上的任务,但该切分方式缺乏通用性。而本文模型只是将关系切分成单词的组合,能适用于所有知识库。此外,该数据集中平均每个问题对应 7.2 个候选关系,而本文实体链接模型的结果中每个问题平均对应 18.8 个候选关系,这对关系检测模型的去噪能力提出了更严格的要求。表 7 进一步对比了本文的基于多角度注意力机制的关系检测模型和 AR-SMCNN,其中 AR-SMCNN 的结果是由基于作者开源代码中的关系检测代码所得到的。

表 7 不同关系检测模型在本文实体链接结果中的准确率

Table 7 Accuracies of different relation detection models on entity-linking results

模型	准确率/%
AR-SMCNN	90.5
本文的关系检测模型	90.9

从表 7 可以看出,由于本文提出的实体链接方法对关系检测的去噪能力要求较高,在候选关系更多的情况下,本文的关系检测模型优于 AR-SMCNN 模型,其准确率提升了 0.4%。我们认为这主要是因为利用基于多角度注意力机制的关系检测模型能从多个角度对关系和问题的相关性进行建模,从而更好地分辨出噪声关系。2)与 BiCNN 相比,本文的关系检测模型没有使用字符级别的编码,可能存在 OOV 的问题,但模型利用 3 个注意力单元对问题和关系的相关性进行了不同角度的建模,从而得到互补和层次丰富的语义信息,这在一定程度上弥补了这方面的缺陷,其准确率比 BiCNN 高 3.5%。另外,本文的关系检测模型比采用单一注意力机制的 AMPCNN 的准确率高 2.2%,这也证明了多角度注意力机制的有效性。3)本文的关系检测模型在去除任意一个注意力单元后,准确率都有所下降,这说明多角度注意力机制对模型而言是互补的,不同单元之间具有一定的差异性,而多角度注意

力机制能刻画文本之间不同语义的相关性。此外,去除 $score_1$ 时,对应的注意力机制的模型准确率下降得最多,达到 0.9%,这说明该部分对模型的影响最大。如 3.3 节所述,此注意力机制融合了更高层次的交互特征,说明它更能消除问题和关系表达之间的语义鸿沟。

结束语　针对单一事实知识库问答问题,本文提出了一种基于神经网络注意力机制的方法,将总任务分解为实体链接和关系检测两个子任务。在实体链接任务中,提出了结合语言模型和词性特征的实体识别模型,通过对形式多变的问句进行解析,提取出对应知识库的实体;在关系检测任务中,提出了基于多角度注意力机制的模型,通过对问题和知识库关系的相关性进行建模,捕获了层次丰富的语义信息。实验结果表明,本文的子任务模型分别取得了接近当前最佳模型的效果,并且在总任务中表现出了很好的效果,其中基于 FB2M 知识库的总任务取得了 78.9%的准确率,接近当前最好的结果 80.2%;基于 FB5M 知识库的总任务取得了 78.3%的准确率,为当前最好结果。

在下一步的工作中,我们将继续探讨如何从知识库中引入更多的实体信息,并针对同名实体的问题展开研究。

参 考 文 献

[1] SUCHANEK F M,KASNECI G,WEIKUM G. Yago:a core of semantic knowledge[C]// Proceedings of the 16th international conference on World Wide Web. New York,NY:ACM,2007:697-706.

[2] BOLLACKER K,EVANS C,PARITOSH P,et al. Freebase:a collaboratively created graph database for structuring human knowledge[C]// Proceedings of the 2008 ACM SIGMOD international conference on Management of data. New York,NY:ACM,2008:1247-1250.

[3] LEHMANN J,ISELE R,JAKOB M,et al. DBpedia-a large-scale,multilingual knowledge base extracted from Wikipedia[J]. Semantic Web,2015,6(2):167-195.

[4] BERANT J,LIANG P. Semantic parsing via paraphrasing[C]// Proceedings of the 52nd Annual Meeting of ACL (Volume 1:Long Papers). Stroudsburg,PA:ACL,2014:1415-1425.

[5] REDDY S,LAPATA M,STEEDMAN M. Large-scale semantic parsing without question-answer pairs[J]. Transactions of the Association of Computational Linguistics,2014,2(1):377-392.

[6] BERANT J,CHOU A,FROSTIG R,et al. Semantic parsing on freebase from question-answer pairs[C]// Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing. 2013:1533-1544.

[7] BORDES A,CHOPRA S,WESTON J. Question Answering with Subgraph Embeddings[C]// Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing,A meeting of SIGDAT,a Special Interest Group of the ACL. Doha,Qatar,View at Publisher View at Google Scholar,2014:615-620.

[8] BORDES A,USUNIER N,CHOPRA S,et al. Large-scale simple question answering with memory networks[J]. arXiv:1506.02075,2015.

[9] HE X,GOLUB D. Character-Level Question Answering with Attention[C]// Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing. Stroudsburg,PA:ALL,2016:1598-1607.

[10] HAKIMOV S,JEBBARA S,CIMIANO P. Evaluating Architectural Choices for Deep Learning Approaches for Question Answering Over Knowledge Bases[C]// ICSC. IEEE,2019:110-113.

[11] DAI Z,LI L,XU W. CFO:Conditional Focused Neural Question Answering with Large-scale Knowledge Bases[C]// Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics. Stroudsburg,PA:ACL,2016:800-810.

[12] YIN W,YU M,XIANG B,et al. Simple Question Answering by Attentive Convolutional Neural Network[C]// COLING 2016. New York:ACM,2016:1746-1756.

[13] YIH W T,CHANG M W,HE X,et al. Semantic Parsing via Staged Query Graph Generation:Question Answering with Knowledge Base[C]// Meeting of the Association for Computational Linguistics & the International Joint Conference on Natural Language Processing. Stroudsburg,PA:ALL,2015:1321-1331.

[14] YANG Y,CHANG M W. S-MART:Novel Tree-based Structured Learning Algorithms Applied to Tweet Entity Linking[C]// ACL 2015. Stroudsburg,PA:ACL,2015:504-513.

[15] QU Y,LIU J,KANG L,et al. Question Answering over Freebase via Attentive RNN with Similarity Matrix based CNN[J]. arXiv:1804.03317,2018.

[16] YU Y,HASAN K S,YU M,et al. Knowledge Base Relation Detection via Multi-View Matching[C]// ADBIS (Short Papers and Workshops). Berlin:Springer,2018:286-294.

[17] WANG Z,HAMZA W,FLORIAN R. Bilateral Multi-Perspective Matching for Natural Language Sentences[C]// IJCAI 2017. Cambridge,MA:MIT,2017:4144-4150.

[18] ZHANG H,XU G,LIANG X,et al. An Attention-Based Word-Level Interaction Model:Relation Detection for Knowledge Base Question Answering[J]. IEEE Access,2018:1-1.

[19] BERANT J,CHOU A,FROSTIG R,et al. Semantic parsing on freebase from question-answer pairs[C]// Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing. Stroudsburg,PA:ALL,2013:1533-1544.

[20] HAO Y,LIU H,HE S,et al. Pattern-revising Enhanced Simple Question Answering over Knowledge Bases[C]// Proceedings of the 27th International Conference on Computational Linguistics. New York:ACM,2018:3272-3282.

[21] YU M,YIN W,HASAN K S,et al. Improved neural relation detection for knowledge base question answering[C]// Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics. Stroudsburg,PA:ACL,2017:571-581.