

# 基于信息匹配方法的中文知识库问答系统

彭 怀<sup>1</sup>, 宋井宽<sup>2</sup>, 唐向红<sup>1</sup>

(1. 贵州大学, 贵州 贵阳 550025;  
2. 电子科技大学, 四川 成都 610054)

**摘 要:** 知识库问答任务是自然语言处理领域中的研究热点之一, 目前国内外学者对知识库问答方法的研究大多数是基于英文数据, 基于中文数据的研究非常少。由于中文存在语言多变性、语法不明确性、语言歧义性等特点, 导致很多英文知识库问答研究方法很难应用于中文数据。针对以上问题, 该文提出一种基于信息匹配的中文知识库问答研究方法, 探索方法在中文数据上的效果。首先对问题进行主语实体识别和属性值识别; 其次将问句中的实体链接到知识库中的实体, 使用逻辑回归对候选实体进行筛选; 再次抽取其两跳内关系作为候选查询路径, 将候选查询路径和问题进行相似度匹配得到匹配度最高的候选路径; 最后使用实体拼接来得到多实体情况的查询路径, 查询知识库获得最终答案。该方法在 CCKS2019 CKBQA 测试集上的 F 值达到了 75.6%。

**关键词:** 知识库问答; 自然语言处理; 实体识别; 实体链接; 预训练模型; 文本匹配

中图分类号: TP18

文献标识码: A

文章编号: 1673-629X(2022)02-0014-06

doi: 10.3969/j.issn.1673-629X.2022.02.002

## Question Answering System of Chinese Knowledge Base Based on Information Matching Method

PENG Huai<sup>1</sup>, SONG Jing-kuan<sup>2</sup>, TANG Xiang-hong<sup>1</sup>

(1. Guizhou University, Guiyang 550025, China;  
2. University of Electronic Science and Technology of China, Chengdu 610054, China)

**Abstract:** Knowledge base question answering is one of the hot topics in the field of natural language processing. At present, most of the researches on knowledge base question answering are based on English data, but few are based on Chinese data. Due to the language variability, grammar ambiguity and language ambiguity, many question answering methods of English knowledge base are difficult to be applied to Chinese data. To solve the above problems, we propose a question-and-answer research method based on information matching for Chinese knowledge base to explore its effect on Chinese data. Firstly, subject entity identification and attribute value identification are carried out. Secondly, the entity in the question is linked to the entity in the knowledge base, and the candidate entity is screened by logistic regression. Thirdly, the two-hop internal relationship is extracted as the candidate query path, and the text matching method based on the pre-trained language model is used to select the candidate path with the highest similarity to the question. Finally, the query path of multi-entity case is obtained by using entity splicing, and the final answer is obtained by querying the knowledge base. The proposed method achieves an F value of 75.6% on the CCKS2019 CKBQA test set.

**Key words:** knowledge base question answering; natural language processing; entity recognition; entity linking; pre-training model; text matching

## 0 引言

知识库问答任务是自然语言处理学术界和工业界的热门研究方向。知识库是知识表现和存储的载体, 目前知识库主要通过三元组表示(头实体1, 关系, 尾实体)实体与实体之间可能存在的语义关系, 例如: 博

尔赫是阿根廷人, 可以表示为:(博尔赫斯, 出生地, 阿根廷)。知识库问答任务是识别自然语言处理问题中所包含的实体、实体关系、实体类型、实体组合后, 通过知识库查询语言到知识库中查询答案。

目前在英文数据集上主要有两类知识库问答方

收稿日期: 2021-03-09

修回日期: 2021-07-10

基金项目: 贵州省留学回国人员科技活动择优资助项目-优秀类项目(2018.0002)

作者简介: 彭 怀(1991-), 男, 硕士研究生, 研究方向为知识图谱; 宋井宽, 教授, 博导, 博士, 研究方向为人工智能、多媒体等; 唐向红, 教授, 博导, 博士, 研究方向为大数据技术与智能制造。

法,第一种是语义解析方法,该方法是直接通过编写规则库、辅助词典、人工推理、机器学习、深度学习手段从问句中识别实体、实体关系、实体组合。Wang 等人<sup>[1]</sup>使用序列标注模型识别问题中的实体,使用序列到序列模型预测问题中的关系序列,并使用答案验证机制和循环训练方式提升模型的性能,在英文多关系问题数据集 WebQuestion 上达到了先进水平。Hu 等人<sup>[2]</sup>提出了一种状态转移的框架,设计了四种状态转移动作和限制条件,结合多通道卷积神经网络等多种方法,在英文复杂问题数据集 ComplexQuestion 上达到了最先进水平。基于语义解析的方法通常使用分类模型进行关系的预测,面临着未登录关系的问题,即训练集未出现的关系难以被预测出来。在中文数据上通常包含几千种以上的关系,语义解析方法在关系数量非常大的情况下效果往往都不太好,使得语义解析方法在应用于中文知识库问答 (Chinese knowledge based question answering, CKBQA) 上受到了极大限制。第二种是信息检索的方法,首先通过实体识别技术、实体词典等其他方式识别问句中的候选实体集合,之后根据预定义的逻辑形式,从知识库中查询候选实体在知识库中的所有一跳或多跳关系,从而得到候选查询路径集合。最后通过计算候选查询路径与问句的相似度获得匹配度最高的候选查询路径,到知识库中查询答案。Yu 等人<sup>[3]</sup>提出了一种增强关系匹配的方法,使用二层 BILSTM 与候选关系进行多层次的匹配,并使用关系匹配对实体链接结果进行重排序,在英文多关系问题数据集上取得了最先进水平。目前在中文领域知识库问答方法主要是基于信息检索和向量建模两种方法进行改进。如 Lai 等人<sup>[4]</sup>使用卷积神经网络识别问句中语义特征,并通过答案和问句匹配度确定结果;周博通等人<sup>[5]</sup>提出一种方式,首先进行命名实体识别,之后通过基于注意力机制的双向 LSTM<sup>[6]</sup>进行属性映射,最后基于前两步的结果从知识库中选择答案;张芳容等<sup>[7]</sup>提出一种融合人工规则的关系抽取方法,提高了关系识别准确率;段江丽等<sup>[8]</sup>提出基于依赖结构的语义关系识别方法,通过依赖结构从问句中挖掘有价

值的语义信息。

在 CKBQA 任务上,Yang 等人<sup>[9]</sup>提出了一种联合抽取实体的关系的流水线方法,在 CCKS2018 COQA 任务上取得了第二名的成绩。参考流水线方法,针对 CKBQA 任务,该文提出一种信息匹配的方法:先进行实体和属性值识别,再进行实体链接,进而从知识库中抽取候选查询路径,使用文本匹配模型选择与问题最相似的候选路径,最后使用实体拼接技术探索多实体情况的可能结果。该方法在 CCKS2019 CKBQA 测试集上的 F 值达到了 75.6%。

## 1 相关工作

### 1.1 整体流程

该文使用的信息匹配方法的主要流程:多种辅助词典构建、实体与实体属性值识别、实体链接与筛选、候选查询路径生成与文本匹配、实体拼接与答案检索。模型流程:

(1)通过 CCKS 官方提供的数据和搜狗词典文件构造实体分词词典、实体链接词典、实体词频词典、实体属性词典辅助后续流程工作;

(2)使用命名实体识别技术识别出问句中的实体、使用知识库实体词表识别问句中的实体,将前面识别出的实体放入候选实体列表,之后通过属性识别得到最终的候选实体列表;

(3)通过分析实体、知识库实体的特点构建特征值,使用机器学习的方式训练模型将实体链接到知识库中的实体,从而获得候选实体列表。该文使用逻辑回归算法训练实体链接模型;

(4)通过知识库查询语句查询候选实体在知识库中的一度、二度关系获取候选查询路径列表,之后通过训练好的文本匹配模型预测候选路径和问句的匹配度获得前  $N$  个候选查询路径;

(5)使用实体拼接方式生成多实体候选查询路径,最后通过候选查询路径和问句的字符重合度得到重合度最高的候选查询路径,查询知识库获得答案。流程如图 1 所示。

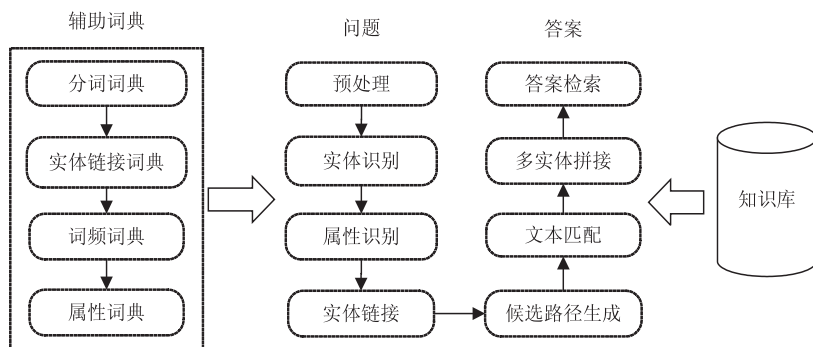


图 1 问答流程

## 1.2 技术介绍

BERT<sup>[10]</sup>是谷歌在2018年提出的一个自然语言处理预训练模型,在很多NLP任务中都取得了很好的效果,例如:命名实体识别、阅读理解、文本匹配、文本分类等。BERT内部是使用双向Transformer<sup>[11]</sup>网络,是真正意义上第一个可以双向提取特征的模型,如图2所示。Transformer是谷歌在2017年提出的一种网络结构,每个Transformer模块由编码器和解码器构成,它是基于自注意力机制的,与RNN不同的是,它采用并行结构可以加快计算,其输入由词向量、位置向量、注意力权重向量构成。因此,它既有CNN可以并行计算的优点,同时也可以捕获词在句子中的位置信息,还可以通过自注意机制调整词对句子结果的权重。

LSTM模型是对循环神经网络(recurrent neural

network, RNN)模型的改进,也是深度学习方法的卓越代表之一。RNN模型在传统神经网络和隐马尔可夫模型上,为神经网络中各层的隐层单元增加时间序列特征,之后通过隐藏层权重将上一个时间点的神经单元的值传递至当前的神经单元,从而使神经网络具备了记忆功能。LSTM模型是在RNN模型的基础上进行了改进,RNN模型存在时间序列比较长的时候,当前神经元很难捕获之前较远的时间序列信息。为解决RNN的缺陷,LSTM在RNN每层中增加了记忆单元,同时LSTM还增加了门控机制,在隐藏层中各单元传递信息是通过几个可控门来控制每层信息输入、遗忘、传递的比例。从而使RNN模型同时具备长期存储信息和控制信息输入输出的功能。

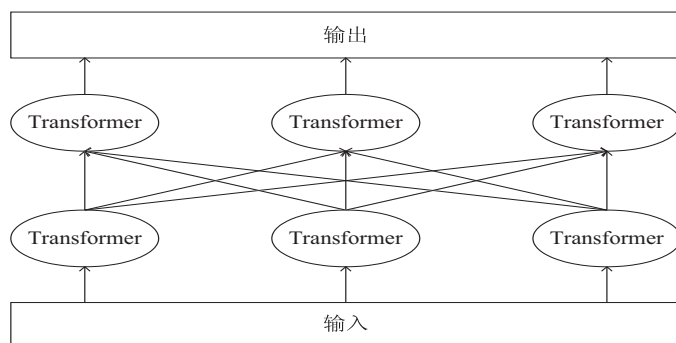


图2 BERT核心结构

## 2 模型和方法

### 2.1 辅助词典构建

该方法在流程中需要多个词典用于分词、计算词频等,均来自于PKUBase知识库或外部资源,词典介绍如下:

(1)实体链接词典:此词典是问句中的实体到知识库实体的映射,由CCKS2019 CKBQA主办方提供;

(2)分词词典:分词词典参照Yang等人<sup>[9]</sup>的方法,通过实体链接词典中的所有实体,以及知识库中所有实体的主干成分构建。例如<红楼梦\_(中国古典长篇小说四大名著之一)>这个实体,只保留下划线之前的部分“红楼梦”;

(3)词频词典:该词典用于计算实体和属性值的词频特征,使用搜狗开源的中文词频词典构建;

(4)属性词典:该词典用于识别属性值的模糊匹配,使用知识库中所有属性值,构建字到词的映射。

### 2.2 实体识别和属性值识别

#### 2.2.1 实体识别

该文采用词典分词和神经网络模型结合进行实体识别。首先将分词词典导入分词工具,对自然语言问题进行分词,如果问句中的分词在词表中就将分词加

入候选实体中,中文分词可能会存在一定错误,并且还存在嵌套实体问题,这种情况下只会保留最长的实体,比如问句“华为的董事长是谁?”,正确的分词结果应当为“华为|的|董事长|是|谁|?”,但词典中存在“华为的董事长”这种更长的实体,所以实际的分词结果为“华为的董事长|是|谁|?”,进而得到错误的实体。针对这样的问题,该文基于预训练语言模型BERT,将训练集的标注实体还原为问句实体,训练一个命名实体识别模型,之后通过模型识别问句中的实体,将识别出的实体加入候选实体列表中。实体识别模型具体流程:(1)对问句进行实体和非实体标注,目前标注主要有IO、BIO、BIEO、BIOES这几种方式,通过测试BIO标注在数据集中效果最好,因此,该文采用BIO标注。B表示实体的头部位置,I表示实体非头部位置,O表示问句中非实体部分。(2)使用BERT-LSTM-CRF模型进行命名实体识别,可分为特征提取和实体标注两部分。在特征提取部分中,长度为 $m$ 的输入问句被分割成词的序列 $\{w_1, w_2, \dots, w_n\}$ 送入BERT网络中,经分词及词嵌后得到 $m$ 个词向量。将词向量经过 $N$ 层的Transformer模块进行上下特征提取后,得到一个[句子长度,隐藏层大小]的特征矩阵,即完成了问句的特征提取。实体识别部分该文采用BiLSTM-CRF



模型,首先将特征矩阵作为输入放入 BI-LSTM 层,通过双向 LSTM 层进行前后向语义特征提取,之后输出带语义信息的特征向量,此时特征向量隐藏层包含前向和后向 LSTM 层信息。将特征向量依次经过 dropout 层、全连接层、线性层之后获得的特征向量作为 CRF 层的输入。该文采用 BIO 标注,实体识别本质上是一个三分类问题,B、I、O 代表词的三种类型。在 CRF 中,是通过维比特算法算出每个词的最大概率,从而识别词的类型。CRF 还可约束词的类型,例如 B 后面只能接 I、不能接 O。

### 2.2.2 属性值识别

问题中包含的属性值规范性较低,可能是很长的字序列,也可能没办法直接与知识库实体进行对应,仅通过分词词典会忽略一些实体。因此针对大部分实体的属性值,使用特殊方式进行识别:

(1)特殊数字、别名、简称、书名等,构建规则库,判断匹配结果是否在知识库的属性值中,在则加入候选属性值;

(2)时间属性:构建正则表达式,将其还原为知识库中规范的时间表达,如“2009 年 6 月”还原为“2009.06”,加入候选属性值;

(3)模糊匹配属性:得到问题中每个字对应的所有属性值,统计每个属性值的次数,选择 top3 的属性加入候选属性值。

### 2.3 实体链接及筛选

对于 2.2 部分得到候选实体列表中的每个实体进行过滤,先判断实体词性是否是名词,删除掉所有非名词的实体。之后通过构建好的实体链接词典,将问句中实体可以连接的知识库实体加入到候选实体中。平均每个问题初步得到的候选实体数量为 12.6,多余的候选实体会引入干扰,同时增加后续步骤的时间成本。因此,参考 Yang 等人<sup>[9]</sup>的方法,根据实体特点为每个实体计算一些特征。

(1)问句中实体的长度:实体的长度,例如:华为,长度为 2;

(2)问句中实体的词频:实体在搜狗词典中词频数;

(3)实体在问句中的位置:问句实体离句首的位置距离;

(4)知识库实体两跳内关系和问句中词的重叠数量;

(5)知识库实体在实体链接词典中的排序,序列越小,实体链接概率越高。

构建实体链接数据集,实体链接分类器该文使用支撑向量机模型进行训练,实体特征包括问句中实体的长度  $X$ 、问句中实体的词频、实体在问句中的位置、

知识库实体两跳内关系和问句中词的重叠数量、知识库实体在实体链接词典中的排序,  $Y$  为 0、1,0 代表问句中实体没有正确链接到知识库实体,1 代表问句中实体正确链接到知识库实体。通过实体链接分类器得到分数排名前  $N$  个候选知识库实体。

### 2.4 候选查询路径生成及文本匹配

在 CCKS2019 中文知识库问答任务中提供的数据集,大部分的问题都是单实体单关系,单实体双关系这种情况,更复杂的情况也可以由简单问题拼接得到。该文采用查询候选实体在知识库中单跳路径和两跳路径结果作为候选查询路径,查询形式为(实体,关系)或者(实体,关系 1,关系 2)。用文本匹配模型(如 ESIM<sup>[12]</sup>)进行用户问句和候选查询路径相似度匹配,但是此模型非常依赖大量标注好的数据集,导致在一些小样本的数据集上效果很不理想。因此,该文使用 BERT 预训练模型来降低对大量标注数据的依赖,经过测试,在少量数据集上使用 BERT 模型进行文本匹配的效果要远远好于 ESIM 模型。BERT 模型使用 mask 机制对上下文单词进行预测、句子对分类等无需人工标注的监督学习任务,学习到词级别、句子级别的信息。将预训练语言模型迁移到下游自然语言处理任务,作用类似于扩大了语料,增加了模型的性能和泛化能力。目前典型的预训练模型有 EMLO<sup>[13]</sup>、BERT、ROBERTA<sup>[14]</sup>、ALBERT<sup>[15]</sup>、ELECTRA<sup>[16]</sup>。

该文基于预训练的 BERT 模型,使用训练集进行文本匹配的微调,在验证集和测试集上,使用该模型计算问题和候选查询路径的相似度。在训练中,文本匹配模型是基于符合自然语言语义问题数据集训练的,但是生成的候选查询路径是不符合问题语义的。针对这种情况,该文对候选路径进行一些特殊处理,例如:(周杰伦,血型)改为“周杰伦的血型?”,在训练集上使用一些多负例的手段,通过增加负例的数量提升模型的泛化能力,训练集中一个正例数据对应 4 个负例数据。使用训练好的文本匹配模型对问句和改进后的候选路径进行打分。

### 2.5 实体拼接及答案检索

上述 2.4 节描述的方法只适用于单实体的情况,实际上,仍然有一部分问题包含两个及以上的主语实体,例如“北京大学出了哪些哲学家”。因此,该文采用实体拼接的方式,探索每个问题作为双实体问题的候选答案。对于每个问题,首先对 2.4 节打分后的候选查询路径进行排序,选取前 10 个单关系查询路径,之后到知识库中对这些查询路径进行检索,通过查询结果判断这些单关系路径是否可以拼接为多关系查询路径,将可以拼接后的多关系路径加入候选查询路径集合,最后,将 2.4 节获得的候选查询路径和本节得到

拼接后的查询路径,同问句进行重叠字的计算,选择字数最多的作为查询答案路径。

### 3 实验与分析

#### 3.1 实验设置

实验运行环境:操作系统 ubuntu18.04,显卡 NVIDIA TITAN Xp 12 GB 显存,Python 版本 3.6.10、pytorch 版本 1.2.0、pytorch-transformers 版本 1.2.0、scikit-learn 版本 0.20.3、torchtext 版本 0.6.0、torchvision 版本 0.4.0。

模型配置:预训练模型使用 BERT 基础版,不区分大小,隐藏层维度大小为 512,隐藏层有 12 层,隐藏层激活函数使用 gelu,词表大小 30 522。LSTM 模型输入维度为 300,输出维度为 300。GRU 模型输入维度为 300,输出维度为 300。

该文使用由北京大学和恒生电子公司共同发布的中文开放领域知识库问答数据集。该任务中问题的标注 SQL 语句均来自于 PKUBase 知识库 (<http://pkubase.gstore-pku.com/>)。数据集的数据统计如表 1 所示。

表 1 语料集数据统计

问题类型	训练集	验证集	测试集
单实体单关系	1 159	476	484
单实体多关系	682	156	160
多实体	356	133	121
总数	2 297	765	765

#### 3.2 命名实体识别结果

对于实体识别和属性环节,该文在测试上针对不同实体识别模型进行消融实验,并且记录了保留不同数量的候选实体的召回率,实验结果如表 2 所示,其中  $f1@n$  表示在保留前  $n$  个候选实体情况下所有问题标注实体的  $f1$  值。

表 2 测试集上实体识别结果

模型	保留数量	$f1@n/\%$
BERT+LSTM+CRF	全部(平均 8.9)	90.3
BERT+LSTM+CRF	1	85.6
BERT+LSTM+CRF	3	88.9
BERT+LSTM+CRF	5	90.1
BERT+LSTM	3	88.6
BERT+CRF	3	88.4
LSTM+CRF	3	80.6
BERT	3	85.2
LSTM	3	79.6
CRF	3	76.5

结果表明:(1)LSTM、GRU、CRF 模型对候选实体

的筛选均有促进作用;(2)BERT 预训练模型在命名实体识别任务上相对于其他模型有着巨大优势。

#### 3.3 实体链接结果

对于实体链接环节,在测试集上针对 5 种特征进行了消融实验,并且记录了保留不同数量的候选实体的召回率。实验结果如表 3 所示,Recall@n 表示在保留前  $n$  个候选实体情况下所有问题标注实体的召回率。

表 3 测试集上实体链接结果

特征	保留数量	Recall@n/%
问句实体+实体特征	全部(平均 12.6)	95.3
问句实体+实体特征	1	80.5
问句实体+实体特征	3	92.6
问句实体+实体特征	5	93.4
问句实体+实体特征	10	94.7
仅实体特征	3	88.4
仅问句实体特征	3	70.3

结果表明:(1)选择的问句实体特征和知识库实体的特征对实体链接准确度有很大影响;(2)从实验结果来看,仅保留前 5 的候选实体就可以达到接近全部数量的结果,同时选择仅保留前五的实体还可以降低训练时间、数据噪音。

#### 3.4 知识库问答结果

进一步,在测试集上计算了文本匹配环节使用不同数量负例及不同检索方案的 F 值。该文对比了三种方案的性能:(1)直接选择文本匹配后相似度最高的查询路径;(2)对所有问题使用桥接获得可能的多实体情况查询路径,对于可以获得多实体查询路径的问题,直接覆盖方案一的路径;(3)对文本匹配排名前 3 的路径和多实体路径和问题重新进行重叠字数的匹配,选择字面上最相近的作为最终查询路径。

表 4 测试集上知识库问答结果

检索方式	负例数量	F-score
仅文本匹配结果	3	68.7
+实体拼接	3	70.5
+实体拼接和字面匹配	3	75.6
+实体拼接和字面匹配	1	74.1
+实体拼接和字面匹配	5	72.1
+实体拼接和字面匹配	10	66.7

从表 4 的实验结果及分析可以得到:在文本匹配环节上,合适数量的负例可以获得更好的学习文本相似性,本任务上 3 个负例效果最佳;实体拼接可以考虑多实体的情况,但会引入一些错误,即一些实际为单实体的问题得到了多实体情况的查询路径,而重叠字数匹配可以有效缓解该问题。

## 4 结束语

该文提出了一种信息匹配的模型,依次对问题进行实体及属性识别、实体链接及筛选、文本匹配和答案检索等,验证了预训练语言模型在知识库问答上的性能,在 CCKS2019 CKBQA 测试集上的 F 值达到了 75.6%。模型优点:(1)使用预训练模型和知识库分词技术大大提升了问句主题词的识别准确率;(2)使用文本匹配技术将问句与实体在知识库中的查询路径进行匹配,避免存在未登录关系的问题;(3)使用实体拼接探索多实体多关系问题。模型缺陷:(1)基于机器学习的实体链接技术比较依赖问句实体、知识库实体特征;(2)产生了过多的候选查询路径,影响了模型运行效率。因此,笔者认为未来可以使用深度学习技术进行实体链接,减少特征依赖,提升准确率;在问句中增加实体类型、实体数量信息以进一步提升多实体多关系问题的准确率。

### 参考文献:

- [1] WANG Y, ZHANG R, XU C, et al. The apva-turbo approach to question answering in knowledge base [C]//Proceedings of the 27th international conference on computational linguistics. Santa Fe, New Mexico, USA: Association for Computational Linguistics, 2018: 1998–2009.
- [2] HU S, ZOU L, ZHANG X. A state-transition framework to answer complex questions over knowledge base [C]//Proceedings of the 2018 conference on empirical methods in natural language processing. Brussels, Belgium: Association for Computational Linguistics, 2018: 2098–2108.
- [3] YU M, YIN W, HASAN K S, et al. Improved neural relation detection for knowledge base question answering [J]. arXiv: 1704. 06194, 2017.
- [4] LAI Y, JIA Y, LIN Y, et al. A chinese question answering system for single-relation factoid questions [C]//National CCF conference on natural language processing and Chinese computing. Dalian, China: Springer, 2017: 124–135.
- [5] 周博通, 孙承杰, 林 磊, 等. 基于 LSTM 的大规模知识库自动问答 [J]. 北京大学学报: 自然科学版, 2018, 54 (2): 286–292.
- [6] HOCHREITER S, SCHMIDHUBER J. Long short-term memory [J]. Neural Computation, 1997, 9 (8): 1735–1780.
- [7] 张芳容, 杨 青. 知识库问答系统中实体关系抽取方法研究 [J]. 计算机工程与应用, 2020, 56 (11): 219–224.
- [8] 段江丽, 胡 新. 自然语言问答中的语义关系识别 [J]. 山东大学学报: 工学版, 2020, 50 (3): 1–7.
- [9] LI Y, MIAO Q, YIN C X, et al. A joint model of entity linking and predicate recognition for knowledge base question answering [C]//China conference on knowledge graph and semantic computing. Tianjing: [s. n.], 2018: 95–100.
- [10] DEVLIN J, CHANG M W, LEE K, et al. Bert: pre-training of deep bidirectional transformers for language understanding [J]. arXiv: 1810. 04805, 2018.
- [11] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [J]. arXiv: 1706. 03762, 2017.
- [12] CHEN Q, ZHU X, LING Z, et al. Enhanced lstm for natural language inference [J]. arXiv: 1609. 06038, 2016.
- [13] PETERS M E, NEUMANN M, IYYER M, et al. Deep contextualized word representations [J]. arXiv: 1802. 05365, 2018.
- [14] LIU Y, OTT M, GOYAL N, et al. Roberta: a robustly optimized bert pretraining approach [J]. arXiv: 1907. 11692, 2019.
- [15] LAN Z, CHEN M, GOODMAN S, et al. A lite BERT for self-supervised learning of language representations [J]. arXiv: 1909. 11942, 2019.
- [16] CLARK K, LUONG M T, LE Q V, et al. Electra: pre-training text encoders as discriminators rather than generators [J]. arXiv: 2003. 10555, 2020.