https://doi.org/10.1051/jnwpu/20213961387

基于多语义相似性的关系检测方法

谢金峰[1,2], 王羽[2], 葛唯益[2], 徐建[1]

(1.南京理工大学 计算机科学与工程学院, 江苏 南京 210094;
2.中国电子科技集团公司第二十八研究所 信息系统工程重点实验室, 江苏 南京 210007)

摘 要:关系检测是知识库问答的关键步骤,直接影响问答质量。现有方法中基于编码比较的方法提取文本整体语义进行匹配会丢失序列的局部信息,而基于交互的方法在文本低层表征层面进行比较会忽略全局语义。针对现有方法无法兼顾全局语义和局部语义信息的问题,提出了一种基于多语义相似性的关系检测模型,通过 BERT 模型分别对问题和关系进行语义表示,然后引入注意力机制、双向长短期记忆网络和多层感知机进行局部关联性分析;利用 BERT 计算出的句向量中含有序列的全局语义信息,设计了问题和关系句向量的全局相似度度量。在基准数据集 SimpleQuestions 和 WebQSP 上进行了实验验证,所提方法分别取得了 93.92%和 87.81%的准确率,优于其他现有的方法。

关 键 词:关系检测;语义相似性;双向长短期记忆网络;注意力机制
中图分类号:TP183 **文献标志码**:A **文章编号**:1000-2758(2021)06-1387-08

随着互联网发展,网络上的文本资源越来越丰富,结构化的信息和知识更易于管理和使用,为此构建出了大规模的开放域知识库,比如 Freebase[1]、Depdia[2]和 PkuBase。通常,采用三元组的形式表示知识库中的知识,形如〈entity, predicate, entity〉,其中 entity 表示特定实体,predicate 表示实体间的关系。知识库中包含大量的三元组,它们构成了一张庞大的关系图。面向知识库的智能问答是一个典型应用场景,被广泛应用于搜索引擎、智能聊天等实际领域中[3]。在面向知识库的问答系统中,查询一个以自然语言描述的问题,系统会在知识库中搜索对应的三元组,之后给出准确的答案。知识库问答包括 2 个子任务:①实体链接:检测出问题中的实体并链接到知识库中的特定实体;②关系检测:检测出问题中提及的关系(链)。关系检测步骤的准确率将会直接影响问答的质量。针对关系检测,以往的工作大致分为两类:将关系检测建模为分类任务;将关系检测建模为文本序列匹配任务。

将关系检测建模为分类任务[4-10]。知识库中会预定义一个关系集,基于分类的关系检测方法将每个关系视为一个类别,通过给问题分类确定与问题相关联的核心关系。例如文献[7]使用词汇映射的方法进行关系检测。之后文献[9]提出一种使用多通道卷积神经网络(CNN)提升分类方法的模型。基于分类的方法进行关系检测时不依赖于实体链接的结果,上一步骤的误差不会传递到这一步骤,但此类方法无法利用关系的语义信息,近年的研究中基于文本序列匹配的方法逐渐成为主流。

基于文本序列匹配的方法通过计算问题和关系的相似度,选择相似度最高的关系作为问题关联的核心关系。现有的方法大致可以分为 2 类:编码-比较型和交互型。基于编码-比较的方法先将输入问题和关系表示为向量序列,之后通过聚合操作将序列转化为定长的向量,使用一种距离计算公式来度量相似度,作为问题和关系的最终匹配得分[11-16]。例如文献[11]将关系视为单个标签,使用 TransE 学习的向量进行初始化。文献[12]提出的 MCCNNs 方法使用多通道 CNN 生成问题表示向量。文献[13]指出从多个粒度表示关系(关系粒度和词序列粒度)可以提升关系检测效果,提出的模型 HR-

收稿日期:2021-03-20 **基金项目**:国家自然科学基金(61872186)与信息系统工程重点实验室开放基金(05201901)资助
作者简介:谢金峰(1998—),南京理工大学硕士研究生,主要从事自然语言处理及深度学习研究。
通信作者:徐建(1979—),南京理工大学教授,主要从事数据挖掘及知识图谱研究。e-mail:dolphin.xu@ njust.edu.cn

Bi-LSTM 利用双向长短期记忆网络(Bi-LSTM)学习不同层次的问题表示和不同粒度的关系表示。基于交互型的方法则假设文本的匹配度依赖于词级别局部的匹配度,先构建交互矩阵,之后对交互矩阵进行抽象表示,最后计算问题和关系的匹配得分。文献[17]提出的交互型方法使用注意力机制计算问题和关系的交互矩阵,之后利用多核 CNN 抽取特征计算相似度得分。由于关系自身存在语义,建模为文本序列匹配的模型可以识别出一些未见关系。

然而,文本匹配的关系检测方法存在如下问题:基于编码-比较的方法更重视全局语义信息,在聚合操作(最大池、平均池)将向量序列变为单个向量会丢失一部分文本的语义信息;基于交互的方法假设文本相似度基于序列局部相似度,因此交互型模型无法由局部匹配刻画全局信息,此类方法缺乏对全局语义的理解。

针对现有模型没有兼顾全局语义信息与局部语义信息的问题,本文提出一个新的关系检测模型,通过融合问题和关系的全局语义、局部语义信息进行关系检测,捕获多样化的特征来计算问题和关系的匹配度,提升关系检测精度。该模型主要分为 2 个部分:局部相似度度量模块和全局相似度度量模块。BERT[18]作为整个模型的文本编码层,获得问题和关系的向量化表示,局部相似度度量模块基于注意力机制,得到关系和问题之间的相似度权重后,使用 Bi-LSTM 分析差异,最后使用多层感知机计算 Q-R 局部相似度。全局相似度计算模块则使用 BERT 获得的句向量,将问题和关系句向量的余弦值作为 Q-R 的全局相似度。最终的匹配度等于全局相似度和局部相似度加权后的和。在广泛使用的 Simple-Question 和 WebQSP 数据集上评估所提出模型的性能,并与仅计算全局相似度或者仅计算局部相似度的模型进行对比分析,以验证本文提出的模型在关系检测任务上的有效性。

1 基于多语义相似性的关系检测模型

1.1 问题定义

图 1 为关系检测的步骤图,下面给出关系检测的形式化描述。

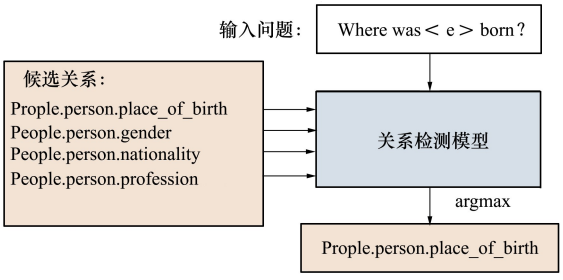

图 1 关系检测步骤图

给出知识库和一个问题,基于知识库的问答的第一步会检测出问题对应的实体,得到实体集合后,从知识库的三元组中选出与实体相联系的候选关系集合 $\boldsymbol{C}_r = \{r_1^{e_1}, r_2^{e_1}, \cdots, r_{|r_n|}^{e_n}\}$,简记为 $\boldsymbol{C}_r = \{r_1, r_2, \cdots, r_{|C|}\}$。其中,$r_i$ 表示候选关系集 $\boldsymbol{C}_r$ 中的一个关系(链),在简单的问答数据集中表示单个关系,在复杂的问答数据集中表示关系路径(从问题中的实体到知识库中实体的关系路径可能为单个关系,也可能为关系链)。第二步为关系检测,关系检测的目标是检测出与问题相关度最高的关系(链),也即是说,选出问题关联的核心关系(链)

$$r_q = \underset{r_i \in \boldsymbol{C}_r}{\operatorname{argmax}} S(r_i, q) \tag{1}$$

式中,$S(r_i, q)$ 表示关系 r 和 q 的匹配得分,也即是模型的计算目标。

1.2 关系检测模型

本节将会详细描述关系检测模型的 4 个模块:文本编码层,全局相似度度量模块、局部相似度度量模块和整体输出层。每一模块的作用如下:

1) 文本编码层:使用 BERT 作为整个模型的文本编码层,将问题和关系中的每个标签转换为向量,该向量含有标签的上下文信息。BERT 模型会在输入的文本前添加特殊的[CLS]标签,由于[CLS]标签自身无实际意义,在训练中能够学习到输入文本的全局语义信息,所以对应的向量可以作为文本的句向量使用。

2) 全局相似度度量模块:使用 BERT 得到的句向量,计算问题和关系句向量的余弦值作为全局相似度。

3) 局部相似度度量模块:使用 Soft Attention 计算问题和关系之间的注意力加权,而后使用 Bi-LSTM 提取差异,最后使用多层感知机计算局部相似度。

4) 整体输出层:输出最终的相似度计算结果。

模型结构如图 2 所示。

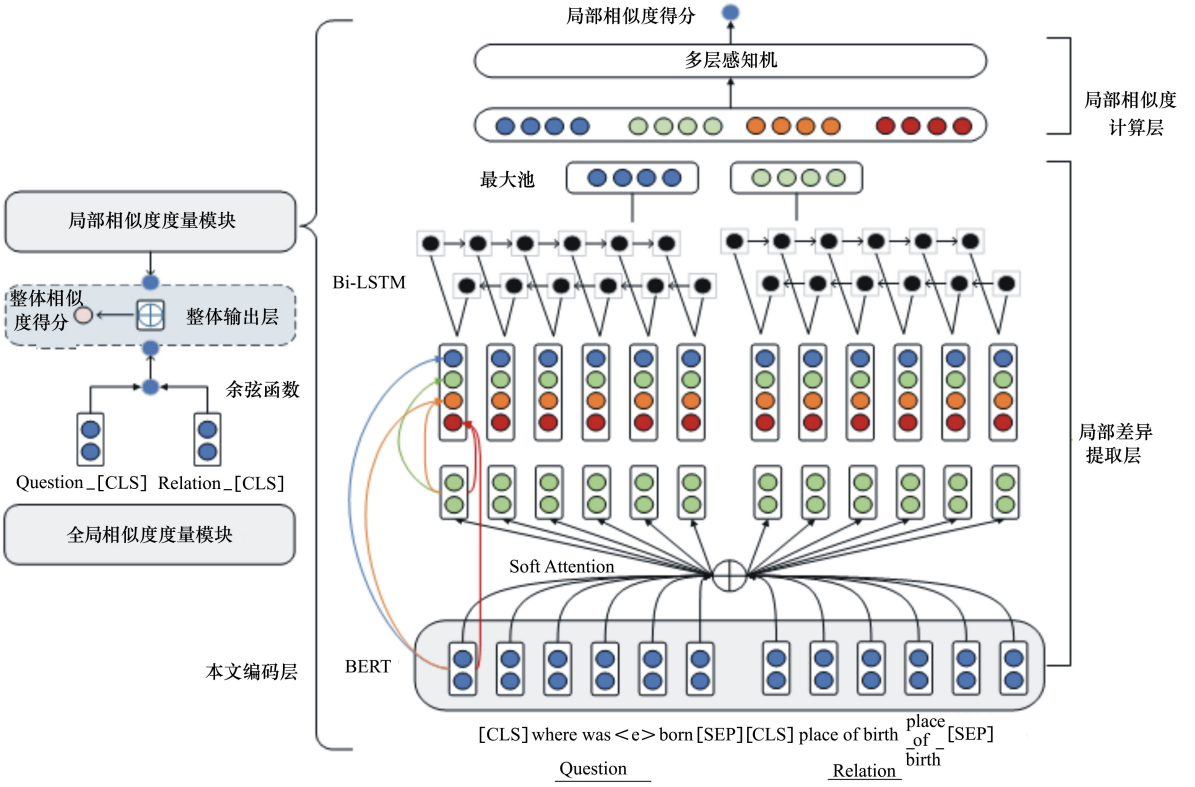

图 2　关系检测模型结构

1.2.1　文本编码层

1) 关系表示层

通常,从词级别和关系级别来表示一个关系(链),假设模型此时正在处理含有 $|r|$ 个关系的关系链 $\boldsymbol{r}=\{r_1,\cdots,r_{|r|}\}$,关系的表示为

$$\boldsymbol{r}=\{r_1^{\text{word}},\cdots,r_{|r|}^{\text{word}}\}\cup\{r_1^{\text{rel}},\cdots,r_{|r|}^{\text{rel}}\} \tag{2}$$

式中,r_i^{word} 表示单个关系 r_i 中的词部分,r_i^{rel} 表示单个关系 r_i 中关系名称部分,在本次研究中,关系链中关系数量 $|r|\leqslant 2$,将所有标记的总数量记为 $|R|$。以 WebQSP 中的一个关系链为例,关系链为 $\{r_1=\text{film},r_2=\text{character\_note}\}$,将其表示为 $\{\text{film},\text{character},\text{note}\}\cup\{\text{character\_note}\}=\{\text{film},\text{character},\text{note},\text{character\_note}\}$。为了方便记录,将上述关系表示记为 $\boldsymbol{r}=\{r_1,\cdots,r_{|R|}\}$。

2) 问题表示层

问题表示为词序列,将问题中所有标记的总数量记为 $|Q|$,则问题 $\boldsymbol{q}=\{w_1,w_2,\cdots,w_{|Q|}\}$。

3) BERT

使用预训练语言模型 BERT 将句子转换为向量序列,BERT 首先从输入的文本数据中得到 3 个向量:标记嵌入、位置嵌入和段嵌入。此外,由于后续需要文本的句向量进行计算,使用 BERT 的[CLS]和[SEP]标签。

1) 标记嵌入:使用 BERT 词表将输入 BERT 的每个标签转为定长向量,得到标记嵌入。

2) 段嵌入:在 BERT 中用于区分不同的句子,由于 BEMSM 的输入为单句,所以这个值全设为 0。

3) 位置嵌入:为了得到输入序列的顺序信息,BERT 对出现在不同位置的词分别附加一个不同的向量用来区分。

问题序列 $\boldsymbol{q}=\{w_1,\cdots,w_{|Q|}\}$ 和关系序列 $\boldsymbol{r}=\{r_1,\cdots,r_{|R|}\}$ 加上[CLS]和[SEP]标签,变为 $\boldsymbol{q}=\{w_{[\text{CLS}]},w_1,\cdots,w_{|Q|},w_{[\text{SEP}]}\}$ 和 $\boldsymbol{r}=\{r_{[\text{CLS}]},r_1,\cdots,r_{|R|},r_{[\text{SEP}]}\}$ 后,经过 BERT 结构,输出为

$$\boldsymbol{E}_q=\{E_{q_{[\text{CLS}]}},E_{q_1},\cdots,E_{q_{|q|}},E_{q_{[\text{SEP}]}}\} \tag{3}$$

$$\boldsymbol{E}_r=\{E_{r_{[\text{CLS}]}},E_{r_1},\cdots,E_{r_{|r|}},E_{r_{[\text{SEP}]}}\} \tag{4}$$

1.2.2　全局相似度度量模块

这一模块的目的是计算问题和关系的全局相似

度,在训练过程中,BERT 的[CLS]标签对应的向量通过学习包含句子的全局信息,可以作为句向量使用。使用问题和关系的[CLS]标签对应的输出向量来计算全局语义相似度,得到

$$S_{\text{global}} = \cos(E_{q_{[\text{CLS}]}}, E_{r_{[\text{CLS}]}}) \tag{5}$$

1.2.3 *局部相似度度量模块*

局部相似度度量模块的目的是计算问题序列和关系序列的局部语义相似度。注意力机制的加入使该模块能够注意输入文本的局部语义信息。该模块输入为文本编码层的输出部分 $\boldsymbol{E}_q$ 和 $\boldsymbol{E}_r$。其中 $\boldsymbol{E}_q \in \mathbb{R}^{(|Q|+2)\times d_e}$,$\boldsymbol{E}_r \in \mathbb{R}^{(|R|+2)\times d_e}$,$d_e$ 为 BERT 的输出向量维度,为了方便后续表示,将 BERT 的输出记为 Q 和 R。通过 BERT,将文本序列转换为带有上下文信息的实值向量序列。

1) 局部差异提取:

之后,进行局部差异提取,局部差异提取层的目的是分析2个序列 $\boldsymbol{Q} = (q_1, \cdots, q_{|Q|+2})$ 与 $\boldsymbol{R} = (r_1, \cdots, r_{|R|+2})$ 的关联程度,对 $\boldsymbol{Q}$ 中的每个词 q_i,计算 q_i 和 $\mathbb{R}$ 中的词 r_j 之间的注意力权重

$$\alpha_{i,j} = q_i^{\mathrm{T}} r_j \tag{6}$$

利用 Softmax 公式计算 $\boldsymbol{Q}$ 关于 $\boldsymbol{R}$ 的注意力加权值,以及 $\boldsymbol{R}$ 关于 $\boldsymbol{Q}$ 的注意力加权值,得到

$$\overline{q_i} = \sum_{j=1}^{l_r} \frac{\exp(\alpha_{i,j})}{\sum_{k=1}^{l_r} \exp(\alpha_{i,k})} r_j \tag{7}$$

$$\overline{r_j} = \sum_{i=1}^{l_q} \frac{\exp(\alpha_{i,j})}{\sum_{k=1}^{l_q} \exp(\alpha_{k,j})} q_i \tag{8}$$

得到文本编码层的输出 $\boldsymbol{Q}$ 和 $\boldsymbol{R}$,以及注意力加权后的 $\overline{\boldsymbol{Q}} = (\overline{q_1}, \cdots, \overline{q_{|Q|+2}})$ 和 $\overline{\boldsymbol{R}} = (\overline{r_1}, \cdots, \overline{r_{|Q|+2}})$ 后,使用文献[19]提出的表达式对这些向量进行差异表示

$$\boldsymbol{m}_{q_i} = [q_i, \overline{q_i}, (q_i - \overline{q_i}), (q_i \odot \overline{q_i})] \tag{9}$$

$$\boldsymbol{m}_{r_i} = [r_i, \overline{r_i}, (r_i - \overline{r_i}), (r_i \odot \overline{r_i})] \tag{10}$$

式中:$-$ 表示逐元素减法;$\odot$ 表示逐元素乘法;$\boldsymbol{m}_{q_i} \in \mathbb{R}^{4\times d_e}$;$\boldsymbol{m}_{r_i} \in \mathbb{R}^{4\times d_e}$;[,] 表示连接操作。之后,使用双向长短期记忆网络(Bi-LSTM)处理信息进行整体分析,使用 2 个 Bi-LSTM 来分析差异,得到对序列分析后的表示

$$\boldsymbol{v}_q = \text{Bi-LSTM}_1(\boldsymbol{m}_q) \tag{11}$$

$$\boldsymbol{v}_r = \text{Bi-LSTM}_2(\boldsymbol{m}_r) \tag{12}$$

式中,$v_q \in \mathbb{R}^{(|Q|+2)*d_r}$,$v_r \in \mathbb{R}^{(|R|+2)*d_r}$,$d_r$ 为 Bi-LSTM 的输出向量维度。由于整体模型中的输入序列不能保证具有相同的长度,使用聚合操作(全局最大池)提取序列中的重要特征,将上述2个序列转为定长的向量 $\boldsymbol{g}_q \in \mathbb{R}^{d_r}$ 和 $\boldsymbol{g}_r \in \mathbb{R}^{d_r}$。至此,模型完成了分析序列差异的过程,提取出了问题和关系序列的特征。

2) 局部相似度计算层:

差异计算层的输入是差异提取层输出的 2 个特征序列 $\boldsymbol{g}_q$ 和 $\boldsymbol{g}_r$,目的是计算问题和关系的局部相似度

$$f = H([\boldsymbol{g}_q, \boldsymbol{g}_r, (\boldsymbol{g}_q - \boldsymbol{g}_r), (\boldsymbol{g}_q \odot \boldsymbol{g}_r)]) \tag{13}$$

$$S_{\text{local}} = w_o * f \tag{14}$$

式中:H 代表一个 3 层的感知机;$w_o \in \mathbb{R}^{d_f}$ 为全连接层的参数,模块最终的输出为问题和关系的局部语义相似度 S_{local}。

1.2.4 *整体输出层*

计算最终关系和问题的匹配得分,权重由对比实验得出,对比了全局相似度,局部相似度权重分配为(0.25,0.75),(0.5,0.5),(0.75,0.25) 以及自动分配的结果,最终选定(0.75,0.25) 这组权重为最终的分配

$$S = 0.75 * S_{\text{global}} + 0.25 * S_{\text{local}} \tag{15}$$

2 实 验

2.1 实验设置

为了验证模型有效性,在 SimpleQuestions 和 WebQSP 数据集上进行实验,相关的关系检测数据集在文献[20]和文献[13]中发布,数据集的相关信息见表 1,为了表示方便,在后续将提出的模型记作 BEMSM。

表 1 数据集

数据集	训练/验证/测试集规模	源知识库	关系数量
SQ	72 238/10 309/20 609	Freebase	6 701
WebQSP	3 116/-/1 649	Freebase	4 536

SimpleQuestions(SQ)[21]是一个简单问答数据集,其关系检测数据中的每个问题只与知识库的一个三元组相关,也就是说,关系链的长度等于 1。

WebQSP(WQ)[7]是一个复杂问答数据集,包含

的问题可能与知识库中的多个三元组相关,数据集中包含的关系链长度小于等于 2。

模型的超参数设置见表 2,在实验中,BERT 作为整个模型的文本编码层,语言模型使用的是 Google 官方提供的预训练模型 BERT-Base,由 12 层的 transformer 编码结构组成,隐藏层向量维度为 768 维,BERT 模型包含 1 亿多参数量,输出词向量维度为 768 维。由于大部分关系名称不在 BERT 词表中,需要扩展 BERT 的原始词表,表中添加形如“place_of_birth”的关系名,对应的向量为关系名包含单词向量的均值。训练模型时,先训练只有全局相似度度量模块的模型,之后,BEMSM 加载 BERT 训练后的权重,这部分权重不再更新,后续继续训练 10 个 epoch。将问题和关系的最大句子长度设置为 30,超过 30 的部分会被舍弃,长度不足 30 的序列用 0 进行填充,经过 BERT 处理后,问题和关系的序列长度为 32。在每个 Bi-LSTM 层后设置 dropout,防止模型过拟合。

表 2　模型参数集

参数类型	参数值
BERT 模型隐藏层维度	768
LSTM 的隐藏层	256
LSTM 层数	1
多层感知机每层参数	[256,128,64]
Dropout 参数	0.3
批大小	64/32
学习率	2×10^{-5}
Loss 函数中的边界阈值	0.1
句子最大长度	30

2.2　评价指标

为了评估 BEMSM 模型在关系检测任务上的有效性,沿用之前关系检测工作采用的评价指标,使用精确率 P 来衡量模型的结果:

$$P = \frac{N_c}{N} \tag{16}$$

式中:N 是测试集中问题的数量;N_c 是被准确识别出关系的问题数量。

2.3　实验结果

2.3.1　局部相似度模块和全局相似度模块的影响

本实验的目的在于验证 BEMSM 模型中局部相似度度量模块和全局相似度度量模块对于整个模型的影响。

为了验证各个模块的效果,将 BEMSM 与单独的全局相似度度量模块、单独的局部相似度度量模块对比。除此之外,还对比使用 GloVe 词向量和 Bi-LSTM 为文本编码层的模型,在初始化时,不在 GloVe 词表中的词的向量从均匀分布(-0.5, 0.5)中随机采样生成,在训练过程中词嵌入层不更新。局部相似度度量模块和全局相似度度量模块的影响如表 3 所示。

表 3　局部相似度和全局相似度模块的影响

文本编码层	设置	SQ/%	WebQSP/%
GloVe+Bi-LSTM	全局相似度模块	92.74	84.60
GloVe+Bi-LSTM	局部相似度模块	93.56	84.29
GloVe+Bi-LSTM	全局+局部	93.70	85.75
BERT	全局相似度模块	93.69	85.81
BERT	局部相似度模块	93.74	85.99
BERT	BEMSM	93.92	87.81

从表 3 中的数据可以看出,综合学习问题和关系的局部信息和全局信息对关系检测任务是有效的,单一的局部相似度度量模块和全局相似度度量模块得到的指标都低于综合模型,在 WebQSP 数据集上差异更明显,BEMSM 模型获得精确率为 87.81%,高于仅考虑全局相似性的 85.81%和局部相似性的 85.99%。

使用 BERT 作为文本编码层的模型所表现的性能也要优于 GloVe 词向量结合 Bi-LSTM 的文本编码层,使用 BERT 的所有模型准确率都有提升。在综合模型上,使用 BERT 模型的指标分别提升了 0.2%和 2.06%,这得益于 BERT 充足的预训练过程和优秀的表征能力,BERT 能捕捉 Bi-LSTM 容易忽视的长距离依赖,进一步增强了关系检测模型的识别能力。

2.3.2　注意力机制对 BEMSM 模型的影响

本实验的目的是验证注意力机制对模型的影响。由于全局相似度模块没有使用注意力信息,所以在局部相似度度量模块中分析注意力影响。将局部相似度度量模块与去掉注意力权重的模型进行对比实验。将注意力矩阵中的值设置为 1 来消除注意力的影响。同样在基于 BERT 的模型和基于 GloVe 结合 Bi-LSTM 的模型上进行上述实验。实验结果

如表 4 所示。

表 4 注意力机制对 BEMEM 模型的影响

文本编码层	设置	SQ/%	WebQSP/%
Bi-LSTM	去注意力	93.15	82.41
Bi-LSTM	局部相似	93.56	84.29
BERT	去注意力	93.14	84.78
BERT	局部相似	93.74	85.99

从表中数据可以看出,去掉注意力权重的影响后,基于 GloVe 结合 Bi-LSTM 的模型和基于 BERT 的模型的准确度指标都有小幅度下降,在 WebQSP 数据集上更为明显(84.29% 和 82.41%,85.99% 和 84.78%)。WebQSP 数据集中,问题和关系的序列长度更长,注意力机制使模型能够将注意力关注到语义最相关的地方,所以影响更为明显。本文模型去掉注意力权重后,由于差异提取层不只依赖于注意力机制产生的结果,模型仍然能取得不错的效果。

2.3.3 模型整体表现

将模型与近年的几个基准进行比较,这些基准属于基于文本序列匹配的关系检测方法,模型的输出为问题和关系的相似度得分,实验结果见表 5。

表 5 BEMSM 在 SQ 和 WebQSP 上的实验结果

模型	关系输入粒度	SQ/%	WebQSP/%
Bi-LSTM	words	91.2	79.32
BiCNN	char-3-gram	90.0	77.74
HR-Bi-LSTM	words+relations	93.3	82.53
ABWIM	words+relations	93.5	85.32
DAM	words+relations	93.3	84.10
BEMSM	words+relations	93.9	87.81

1) BiLSTM:使用 2 个 Bi-LSTM 分别表示问题和关系,将最后一个时间步的输出作为句向量来比较相似度。

2) BiCNN[22]:使用 3-gram 表示问题和关系,比如“who”表示为#-w-h, w-h-o, h-o-#。使用 CNN 作为编码层,池化后计算相似度。

3) HR-Bi-LSTM[13]:使用词序列+关系名表示关系,使用 HR-Bi-LSTM 得到问题的抽象层次,将问题表示和关系表示通过聚合操作固定为相同维度向量后,将它们之间的余弦值作为相似度。

4) DAM[14]:基于 transformer 结构的模型,使用 transformer 结构得到关系和问题表示,经过聚合操作后,计算向量余弦值作为相似度得分。

5) ABWIM[17]:使用词序列和关系名来表示关系,通过注意力机制将问题和关系软对齐,之后使用 CNN 计算相似度得分。

BEMSM 模型整体表现的实验结果见表 5,从实验对比结果可以看出,在 SimpleQuestion 数据集上,提出的模型取得了与基线相似的准确度,在 WebQSP 上,提出的方法比最好的基线方法高 2.49%。实验结果验证了模型的有效性,BEMSM 模型使用 BERT 作为文本编码层,可以捕捉到 Bi-LSTM 未能捕捉的长距离依赖,之后综合学习问题和关系的全局语义信息和局部语义信息进行关系检测,可以更全面地计算语义相关性,性能要优于只计算单语义信息的基线方法。

3 结 论

针对现有的关系检测方法无法兼顾问题、关系间全局语义和局部语义信息的问题,提出了一种基于多语义相似度的关系检测模型。该模型将 BERT 作为整个模型的文本编码层,使用 Soft Attention 和 Bi-LSTM 进行局部相似度计算;使用距离计算公式度量问题和关系句向量的全局相似度。最后综合这 2 种相似度得到整体的相似度。在 SimpleQuestions 和 WebQSP 基准数据集上进行实验,分析局部对比相似度模块和全局相似度模块的效果、注意力机制的效果以及模型整体效果,验证了 BEMSM 模型的有效性。

参考文献:

[1] 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, 2008

[2] 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

[3] 徐增林, 盛泳潘, 贺丽荣, 等. 知识图谱技术综述[J]. 电子科技大学学报, 2016, 45(4): 589-606
XU Zenglin, SHENG Yongpan, HE Lirong, et al. Review on knowledge graph techniques[J]. Journal of University of Electronic Science and Technology of China, 2016, 45(4): 589-606 (in Chinese)

[4] KWIATKOWKSi T, ZETTLEMOYER L, GOLDWATER S, et al. Inducing probabilistic CCG grammars from logical form with higher-order unification[C]//Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing, 2010

[5] LIANG P. Lambda dependency-based compositional semantics[EB/OL]. (2013-09-17)[2021-06-07]. https://arxiv.org/abs/1309.4408

[6] BERANT J, LIANG P. Semantic parsing via paraphrasing[C]//Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, 2014

[7] 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

[8] YAO X, VAN B. Information extraction over structured data: question answering with freebase[C]//Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, 2014

[9] XU K, REDDY S, FENG Y, et al. Question answering on freebase via relation extraction and textual evidence[C]//Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, 2016

[10] 胡宝顺, 王大玲, 于戈,等. 基于句法结构特征分析及分类技术的答案提取算法[J]. 计算机学报, 2008, 31(4): 662-676
HU Baoshun, WANG Daling, YU Ge, et al. An answer extraction algorithm based on syntax structure feature parsing and classification[J]. Chinese Journal of Computers, 2008, 31(4): 662-676 (in Chinese)

[11] DAI Zihang, LI Lei, XU Wei. CFO: conditional focused neural question answering with large-scale knowledge bases[C]//Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, 2016

[12] DONG Li, WEI Furu, ZHOU Ming, et al. Question answering over freebase with multi-column convolutional neural networks[C]//Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing, 2015

[13] YU Mo, YIN Wenpeng, 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, 2017

[14] CHEN Yongrui, LI Huiying. DAM: transformer-based relation detection for question answering over knowledge base[J]. Knowledge-Based Systems, 2020(201): 106077

[15] 张仰森, 王胜, 魏文杰, 等. 融合语义信息与问题关键信息的多阶段注意力答案选取模型[J]. 计算机学报,2021,44(3): 491-507
ZHANG Yangsen, WANG Sheng, WEI Wenjie, et al. An Answer selection model based on multi-stage attention mechanism with combination of semantic information and key information of the question[J]. Computer Science, 2021, 44(3): 491-507 (in Chinese)

[16] QIU Yunqi, LI Manling, WANG Yuanzhuo, et al. Hierarchical type constrained topic entity detection for knowledge base question answering[C]//Companion Proceedings of the Web Conference, 2018

[17] ZHANG Hongzhi, XU Guandong, LIANG Xiao, et al. An attention-based word-level interaction model for knowledge base relation detection[J]. IEEE Access, 2018, 6: 75429-75441

[18] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]//Advances in Neural Information Processing Systems, 2017

[19] MOU Lili, MEN Rui, LI Ge, et al. Natural language inference by tree-based convolution and heuristic matching[C]//Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, 2016

[20] YIN Wenpeng, YU Mo, XIANG Bing, et al. Simple question answering by attentive convolutional neural network[C]//The 26th International Conference on Computational Linguistics, 2016

[21] BORDES A, USUNIER N, CHOPRA S, et al. Large-scale simple question answering with memory networks[EB/OL]. (2015-06-05)[2021-06-07]. https://arxiv.org/abs/1506.02075

[22] YIH T, CHANG Mingwei, HE Xiaodong, et al. Semantic parsing via staged query graph generation: question answering with knowledge base[C]//Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics, 2015

A relation detection method based on multi semantic similarity

XIE Jinfeng[1,2], WANG Yu[2], GE Weiyi[2], XU Jian[1]

(1.School of Computer Science & Engineering, Nanjing University of Science & Technology, Nanjing 210094, China;
2.Science and Technology on Information Systems Engineering Laboratory, the 28th Research Institute of China Electronics Technology Group Corporation, Nanjing 210007, China)

Abstract: Relation detection is a critical step of knowledge base question answering, which directly affects the quality of question answering. Among the existing methods, the encoding-comparison method extracts text global semantic information for matching, which often ignores the local semantic feature of text sequence. The interaction approach performs the comparison on low-level representations based on the sequence local information, which fails to consider the global semantic information of the input sequences. To solve the issues, this paper proposes a relation detection model based on Bert model and multi-semantic similarity considering global and local semantic information. First, our model introduces Bert as a text encoding layer to represent questions and relations as sequences of vectors. And then, a bi-directional long short-term memory (Bi-LSTM) layer with the attention mechanism is used to analyze the local semantic relevance and calculate the local similarity. Finally, our model uses a distance calculation formula to measure the global semantic relevance between questions and relations. The experimental results on two benchmark datasets, SimpleQuestions and WebQSP, show that the proposed model achieves the accuracy of 93.92% and 87.81% respectively, performs better than state-of-the-art approaches.

Keywords: relation detection; semantic similarity; Bi-LSTM; attention mechanism

引用格式: *谢金峰, 王羽, 葛唯益, 等. 基于多语义相似性的关系检测方法[J]. 西北工业大学学报,2021,39(6): 1387-1394*

XIE Jinfeng, WANG Yu, GE Weiyi, et al. A relation detection method based on multi semantic similarity[J]. Journal of Northwestern Polytechnical University, 2021, 39(6): 1387-1394 (in Chinese)