



融合外部语义知识的中文文本蕴含识别

李世宝, 李 贺, 赵庆帅, 殷乐乐, 刘建航, 黄庭培

(中国石油大学(华东)海洋与空间信息学院, 山东 青岛 266580)

摘 要: 基于神经网络的文本蕴含识别模型通常仅从训练数据中学习推理知识, 导致模型泛化能力较弱。提出一种融合外部语义知识的中文知识增强推理模型(CKEIM)。根据知网知识库的特点提取词级语义知识特征以构建注意力权重矩阵, 同时从同义词词林知识库中选取词语相似度特征和上下位特征组成特征向量, 并将注意力权重矩阵、特征向量与编码后的文本向量相结合融入神经网络的模型训练过程, 实现中文文本蕴含的增强识别。实验结果表明, 与增强序列推理模型相比, CKEIM在15%、50%和100%数据规模的CNLI训练集下识别准确率分别提升了3.7%、1.5%和0.9%, 具有更好的中文文本蕴含识别性能和泛化能力。

关键词: 中文文本蕴含; 自然语言推理; 注意力机制; 双向长短期记忆网络; 知网; 词林

开放科学(资源服务)标志码(OSID):



中文引用格式: 李世宝, 李贺, 赵庆帅, 等. 融合外部语义知识的中文文本蕴含识别[J]. 计算机工程, 2021, 47(1): 44-49.

英文引用格式: LI Shibao, LI He, ZHAO Qingshuai, et al. Chinese textual entailment recognition fused with external semantic knowledge[J]. Computer Engineering, 2021, 47(1): 44-49.

Chinese Textual Entailment Recognition Fused with External Semantic Knowledge

LI Shibao, LI He, ZHAO Qingshuai, YIN Lele, LIU Jianhang, HUANG Tingpei

(College of Oceanography and Space Informatics, China University of Petroleum (East China), Qingdao, Shandong 266580, China)

[Abstract] The textual entailment recognition model based on neural network learns inference knowledge only from training data, which leads to the weak generalization ability of the model. This paper proposes a Chinese Knowledge Enhanced Inference Model (CKEIM) fused with external semantic knowledge. Based on the features of the HowNet knowledge base, the features of word-level semantic knowledge are extracted to construct an attention weight matrix. At the same time, the semantic similarity features of words and hyponymy features are selected from the CiLin knowledge base of synonyms to form the feature vector. Finally, the attention weight matrix, the feature vector and the encoded text vectors are integrated into the training of the neural network model to implement enhanced recognition of Chinese textual entailment. Experimental results show that compared with the Enhanced Sequential Inference Model (ESIM), CKEIM improves the recognition accuracy by 3.7%, 1.5% and 0.9% respectively on CNLI training sets of 15%, 50% and 100% data scales, which demonstrates that it has better Chinese textual entailment recognition performance and generalization ability.

[Key words] Chinese textual entailment; natural language inference; attention mechanism; Bi-directional Long Short-Term Memory (BiLSTM) network; HowNet; CiLin

DOI: 10.19678/j.issn.1000-3428.0056841

0 概述

文本蕴含识别也称为自然语言推理, 是一项基础而又重要的自然语言处理任务。该任务给定前提文本

(P)与假设文本(H), 然后从语义层面识别出P和H之间的单向推理关系, 包括蕴含、矛盾、中立3种关系^[1], 并且在文本蕴含识别任务上所取得的进展可进一步提高文本摘要、语义搜索、阅读理解和问答等任务的性能^[2]。

基金项目: 国家自然科学基金(61972417, 61872385); 中央高校基本科研业务费专项资金(18CX02134A, 19CX05003A-4, 18CX02137A)。

作者简介: 李世宝(1978—), 男, 副教授、硕士, 主研方向为无线通信、自然语言处理; 李 贺、赵庆帅、殷乐乐, 硕士研究生; 刘建航、黄庭培, 副教授、博士。

收稿日期: 2019-12-09 **修回日期:** 2020-01-17 **E-mail:** lishibao@upc.edu.cn

早期研究多数采用统计及机器学习方法进行文本蕴含识别^[3]。近年来,端到端训练的神经网络模型在阅读理解、情感分析和机器翻译等自然语言处理任务中表现出色,由此研究人员提出了一些组合多种神经网络和注意力机制的文本蕴含识别模型,并且经过大量数据训练,取得了相当高的训练精度。这些文本蕴含识别模型大致可以分为两类。一类侧重编码器设计,能够较好地对句子特征进行编码,并且整体网络结构简单,是早期的主流方向。文献[4]使用长短期记忆(Long Short-Term Memory, LSTM)网络分别学习前提和假设的句子表示并对其进行串联及分类。文献[5]进一步使用双向LSTM(Bi-directional LSTM, BiLSTM)网络对句子进行编码,并运用句子内自注意力机制更好地增强句子编码能力。文献[6]设计HBMP模型进行句子编码,相比使用单一的BiLSTM层效果更佳。另一类注重句间交互,引入句间注意力机制提取句子之间的语义信息,能够更好地识别句子之间的语义关系。文献[7]提出一个逐词注意力模型,对文本中隐含的前提和假设之间的关系进行编码。文献[8]对文献[7]做了进一步扩展,解决了学习注意力权重前提下单个向量表示的局限性问题。文献[9]提出双向多视角匹配机制,分析并对比了多种注意力机制的匹配策略。文献[10]提出一种增强序列推理模型(Enhanced Sequential Inference Model, ESIM),利用两层不同的BiLSTM网络和注意力机制对文本进行编码。

但是由于上述模型仅从训练数据中学习推理知识,受限于训练数据集规模,导致模型泛化能力较弱^[11],因此文献[12]利用WordNet作为外部知识,增强了模型在不同测试集上的泛化能力,为结合外部知识的相关研究提供了参考。文献[13]在模型中使用多个英文知识库,并引入知识图概念,提高了模型识别性能。然而,基于神经网络的文本蕴含识别在中文领域的研究较少^[14],且知网(HowNet)等中文外部语义知识库与WordNet知识库有较大区别,使得无法直接迁移并应用基于英文数据集的文本蕴含识别模型。因此,本文提出中文知识增强推理模型(Chinese Knowledge Enhanced Inference Model, CKEIM),通过分析中文语义知识库的特点,精确提取词级的语义知识特征以构建注意力权重矩阵,同时将其融入神经网络训练过程。

1 知识库处理

1.1 知识库结构分析

HowNet知识库^[15]的层次结构为词语-义项-义原,其包含2 000多个基础义原,将词语从语义上进行拆分,同一个词语的不同含义分为多个义项,每个义项又由多个义原组合而来。如图1所示,词语“苹

果”有多种含义,分别代表“电脑”“手机”“水果”以及“树木”。每一种含义表示一个义项,用多个义原加以注解。在HowNet中,使用“样式值”“携带”“特定品牌”和“电脑”等义原描述“苹果”表示电脑品牌时的义项,并以树状结构进行组织。本文忽略了义原之间的结构关系,并且对于每个词语,将其所有的义原归类为无序集。

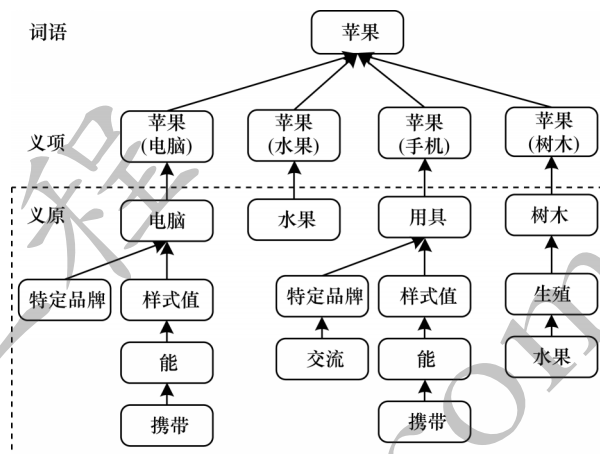


图1 HowNet结构示意图

Fig.1 Schematic diagram of HowNet structure

同义词词林^[16]将中文同义词以及同类词进行划分和归类,分为5层层次结构,其中1层~4层仅包含分类编码信息,没有具体词汇。本文使用改进版同义词词林^[17]将原子词群中较抽象的概念提取到1层~4层的分类节点中,使1层~4层的分类节点包含代表整个类别的具体概念,如提取“生物”“人”“植物”“动物”等抽象概念并将其放置于大、中、小3类高层节点中。

为方便表述,本文使用的符号定义如下:HowNet中所有的义原、义项和词语集合分别表示为 \mathcal{E} 、 \mathcal{S} 和 \mathcal{W} ;语料库包含 $K=|\mathcal{E}|$ 个义原、 $M=|\mathcal{S}|$ 个义项以及 $N=|\mathcal{W}|$ 个词语; \mathcal{S}^w 表示词语 $w \in \mathcal{W}$ 所包含的义项集合; $\mathcal{E}^w \subseteq \mathcal{E} = \{e_k\}_{k=1}^K$ 表示词语 w 所包含义原的无序集合; \mathcal{E}^s 表示义项 $s \in \mathcal{S}$ 所包含义原的无序集合。

1.2 知识库特征提取

通过HowNet知识库结构可知,在使用HowNet知识库计算两个词语的相似度时,如“苹果”和“梨”的相似度,由于“梨”只有“水果”1种义项,而“苹果”有“电脑”“水果”“手机”“树木”4种义项,这样就会出现4个差距相当大的相似度,如果取相似度的最大值或者平均值,则会造成误差,而在相互为多义词的词语之间计算相似度会产生更大的误差。本文设计一个义项选择器结构,如图2所示,在模型训练时根据词语 p_s 的上下文信息 p_1 、 p_o 等,计算出其与各个义原匹配的概率,进而选出当前词语 p_s 所代表的真正义项,然后进行词语间的相似度计算。

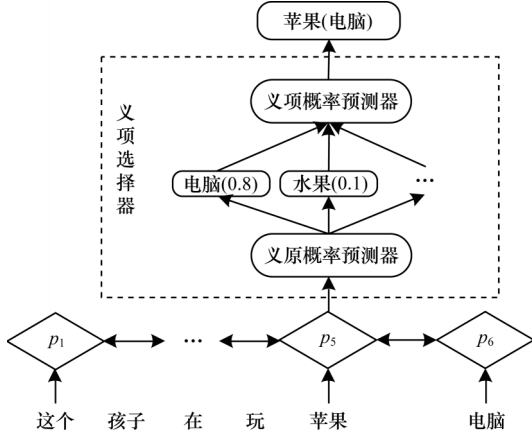


图2 义项选择器结构

Fig.2 Structure of sense selector

对于义原概率预测器,假设给定上下文 $w_1, w_2, \dots, w_i, \dots, w_n$, 经过 BiLSTM 网络编码后,得到词语 w_i 的上下文向量 $\mathbf{g}_i \in \mathbf{R}^{H_0}$, H_0 为向量 \mathbf{g}_i 的维度。将 \mathbf{g}_i 输入到预测器中,可生成词语 w_i 包含的每个义原 e_k 与 \mathbf{g}_i 的关联度。由于义原是最小的语义单位,任何两个不同的义原之间不存在语义重叠,因此词语 w_i 包含的义原 e_k 事件为独立,当前词语 w_i 在上下文语境中编码后得到的上下文向量 \mathbf{g}_i 与义原 e_k 的关联度可表示为 $\varphi'_k(e_k | \mathbf{g}_i)$, 如式(1)所示:

$$\varphi'_k(e_k | \mathbf{g}_i) = \sigma((\mathbf{g}_i^T \mathbf{v}'_k + b'_k), e_k \in \mathcal{E}^{w_i}) \quad (1)$$

其中: $\mathbf{v}'_k \in \mathbf{R}^{H_0}$ 表示当前词语 w_i 所包含义原 e_k 的 200 维嵌入向量,由 SAT 模型^[18]训练的义原向量进行初始化; $b'_k \in \mathbf{R}$ 表示训练参数,且随着网络训练进行更新。由于该预测器的输出为概率值,表示当前词语 w_i 包含的各义原 e_k 与上下文向量 \mathbf{g}_i 关联的概率值,并且网络层数为单层,因此将 sigmoid 作为激活函数 $\sigma(*)$ 。

对于义项概率预测器,由于一个义项会包含一个或多个义原,因此根据在义原概率预测器中得到的 e_k 与 \mathbf{g}_i 的关联概率 φ'_k ,可进一步得到词语 w_i 中各义项 s 与 \mathbf{g}_i 匹配的概率 $P(s | \mathbf{g}_i)$, 如式(2)所示:

$$P(s | \mathbf{g}_i) = \frac{1}{|\mathcal{E}^s|} \sum_{e_k \in \mathcal{E}^s} \varphi'_k(e_k | \mathbf{g}_i), s \in \mathcal{S}^{w_i} \quad (2)$$

由以上设计的义项选择器可计算出词语 w_i 在当前语境下概率最大的义项 s_{\max} 。在确定义项后,根据混合层次结构的 HowNet 词语相似度计算方法^[19],精确计算出词语 w_i 和 w_j 在具体语境中的相似度 Z_{ij} 。

由于提取的同义词词林特征包含词语相似度特征以及词语上下位特征,因此将基于信息内容的词林词语相似度计算方法^[17]作为相似度特征的提取方法,使提取到的词语 w_i 和 w_j 之间的相似度表示为 C_{ij} 。同义词词林共 5 层结构,设层次结构中两个词语之间的距离

为 n ,可提取的词语间上下位信息具体包括:

1) 上位词特征 S_{ij} : 如果一个词语 w_j 是另一个词语 w_i 的上位词,该特征取值为 $1-n/4$, 如 [苹果, 水果]=0.75, [苹果, 植物]=0.5。

2) 下位词特征 X_{ij} : 与上位词特征正好相反, 如 [水果, 苹果]=0.75。

最终将外部语义知识库中提取到的 4 个特征组合成特征向量 $\mathbf{R}_{ij} = [Z_{ij}, C_{ij}, S_{ij}, X_{ij}]$, 并将其作为外部语义知识融入神经网络中。

2 CKEIM 模型

在 CKEIM 模型中,设前提语句 $\mathbf{p} = [p_1, p_2, \dots, p_{l_p}]$, 假设语句 $\mathbf{h} = [h_1, h_2, \dots, h_{l_h}]$, 首先对语句进行编码,使用预训练的词向量对词语进行初始化,然后输入 BiLSTM 网络进行编码,获得上下文相关的隐层向量 $\bar{\mathbf{p}}_i$ 和 $\bar{\mathbf{h}}_j$, 如式(3)、式(4)所示:

$$\bar{\mathbf{p}}_i = \text{BiLSTM}(E(\mathbf{p}), i), \forall i \in \{1, 2, \dots, l_p\} \quad (3)$$

$$\bar{\mathbf{h}}_j = \text{BiLSTM}(E(\mathbf{h}), j), \forall j \in \{1, 2, \dots, l_h\} \quad (4)$$

其中, $(E(\mathbf{p}), i)$ 表示前提语句 \mathbf{p} 的第 i 个词语初始化后的词向量, $(E(\mathbf{h}), j)$ 表示假设语句 \mathbf{h} 的第 j 个词语初始化后的词向量。

根据 HowNet 知识库中提取的词语相似度特征 Z_{ij} , 结合 BiLSTM 网络输出的隐层向量 $\bar{\mathbf{p}}_i$ 和 $\bar{\mathbf{h}}_j$ 的内积 $U_{ij} = (\bar{\mathbf{p}}_i)^T \bar{\mathbf{h}}_j$ 可计算出知识增强的协同注意力权重 M_{ij} , 如式(5)所示:

$$M_{ij} = U_{ij} + \lambda Z_{ij} \quad (5)$$

其中, λ 为超参数。将得到的注意力权重归一化后可得到注意力权重矩阵 μ_{ij} 和 η_{ij} , 进而得到前提语句中的每个词语在假设语句中相关语义信息的向量表示 \mathbf{p}_i^m , 以及假设语句中每个词语在前提语句中相关语义信息的向量表示 \mathbf{h}_j^m , 如式(6)、式(7)所示:

$$\mathbf{p}_i^m = \sum_{j=1}^{l_h} \mu_{ij} \bar{\mathbf{h}}_j, \mu_{ij} = \frac{\exp(M_{ij})}{\sum_{k=1}^{l_h} \exp(M_{ik})} \quad (6)$$

$$\mathbf{h}_j^m = \sum_{i=1}^{l_p} \eta_{ij} \bar{\mathbf{p}}_i, \eta_{ij} = \frac{\exp(M_{ij})}{\sum_{k=1}^{l_p} \exp(M_{kj})} \quad (7)$$

将上下文向量 $\bar{\mathbf{p}}_i(\bar{\mathbf{h}}_j)$ 与经过外部知识融合的向量 $\mathbf{p}_i^m(\mathbf{h}_j^m)$ 拼接后进行比较并进一步拼接两者之差和两者逐位乘,再把从外部语义知识中获取的特征向量 \mathbf{R}_{ij} 加权后拼接,输入非线性映射函数 F 中,得到加权后的关系向量 \mathbf{p}_i^c 和 \mathbf{h}_j^c , 如式(8)、式(9)所示:

$$\mathbf{p}_i^c = F \left(\left[\bar{\mathbf{p}}_i; \mathbf{p}_i^m; \bar{\mathbf{p}}_i - \mathbf{p}_i^m; \bar{\mathbf{p}}_i \odot \mathbf{p}_i^m; \sum_{j=1}^{l_h} \mu_{ij} \mathbf{R}_{ij} \right] \right) \quad (8)$$

$$h_j^c = F \left(\left[\bar{h}_j; h_j^m; \bar{h}_j - h_j^m; \bar{h}_j \odot h_j^m; \sum_{i=1}^{l_p} \eta_{ij} R_{ji} \right] \right) \quad (9)$$

其中, F 是前馈神经网络, 其以 ReLU 为激活函数。

经过拼接并降维后得到的关系向量包含了大量词语间的推理信息, 需要进行组合并提取出语句级的推理信息, 因此再用一层 BiLSTM 网络进一步收集并组合这些信息, 提取前提和假设文本之间语句级的推理关系, 得到包含语句级推理关系的向量 p_i^v

和 h_j^v , 如式(10)、式(11)所示:

$$p_i^v = \text{BiLSTM}(p_i^c, i) \quad (10)$$

$$h_j^v = \text{BiLSTM}(h_j^c, j) \quad (11)$$

首先将输出的隐层向量 p_i^v 和 h_j^v 输入池化层, 分别经过均值池化、最大池化和加权池化得到固定长度的向量。然后将其拼接后输入到多层感知机 (Multilayer Perceptron, MLP) 中进行分类。最后以最小化交叉熵损失为目标, 对整个模型进行端到端训练。CKEIM 模型结构如图3所示。

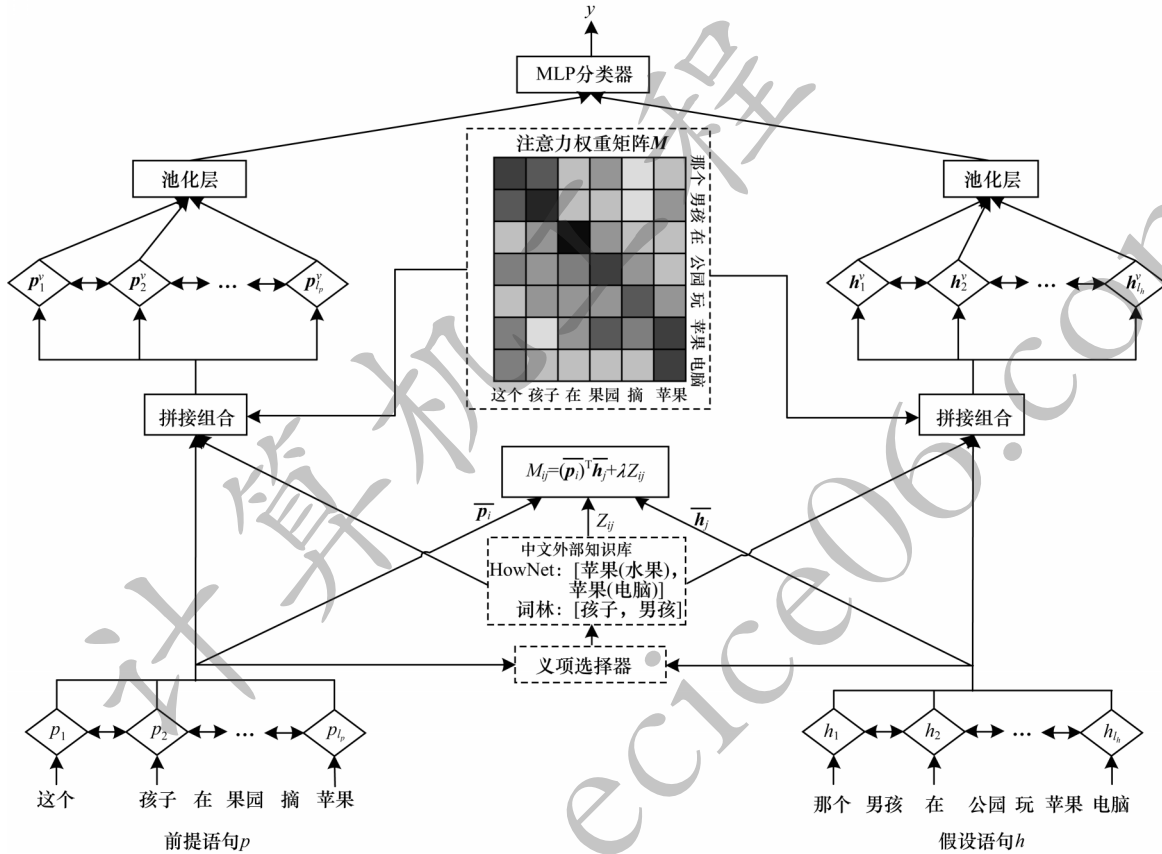


图3 CKEIM 模型结构

Fig.3 Structure of CKEIM model

3 实验结果与分析

3.1 实验数据集

本文在 RITE 和 CNLI 两个中文文本蕴含数据集上进行相关实验。

1) RITE: 该数据集收集了 2012 年和 2014 年日本国立情报学研究所组织的 NTCIR 项目发布的 2 600 个文本对, 其中, 训练集中有 2 000 个文本对, 测试集中有 600 个文本对。数据集涵盖多音字、同音字等多种语言现象, 包括人文、历史、政治等多种题材, 并且以识别准确率作为评价标准。

2) CNLI: 该数据集由 SNLI 数据集通过机器翻译和人工筛选得到, 包括 145 000 个文本对, 其中, 训练集有 125 000 个文本对, 验证集和测试集各有 10 000 个文

本对, 分类结果包括蕴含、矛盾和中立 3 种关系, 每个文本对具有唯一的标签, 并且以识别准确率作为评价标准。

3.2 实验参数设置

CKEIM 模型的实验参数设置如下: 单词嵌入与 BiLSTM 网络隐藏层的维度均为 200 维, 使用腾讯 AI 实验室^[20]发布的 200 维中文词向量初始化词嵌入, 词表中不存在的词语会随机初始化; 将 Adam^[21]作为随机梯度优化策略, 初始化学率为 0.000 4, 批处理 (Mirri-batch) 大小为 32; 利用哈尔滨工业大学的 LTP 语言云^[22]作为中文分词工具。

为便于对比, ESIM 等基线模型的实验参数、实验环境、实验设置与 CKEIM 模型相同。首先从集合 {0.2,

0.5, 1, 1.5, 2, 5, 10, 15, 20} 中依次选取计算注意力权重的超参数 λ 并在训练集上进行训练。然后根据测试集的实验结果选择效果最优的 λ 值,再以此 λ 值为基础,通过调整 $\pm(0.1\sim 0.5)$ 的步长改变 λ 值,直到找到实验效果最优的 λ 值。最后得到CNLI数据集的最佳 λ 值为1, RITE数据集的最佳 λ 值为20。

3.3 结果分析

表1为本文CKEIM模型与BiLSTM+广义池化层模型、BiLSTM+句内注意力模型^[5]、BiLSTM+最大池化层模型^[6]、HBMP模型^[6]和ESIM模型^[10]在大规模CNLI数据集上的识别准确率对比。由于在神经网络中结合外部知识的文本蕴含识别模型都是基于英文知识库,无法在中文数据集上进行比较,因此本文选择其基础模型ESIM作为主要基线模型,同时与其他经典模型进行对比。表2为CKEIM模型与BiLSTM+广义池化层模型、BiLSTM+句内注意力模型、HBMP模型和ESIM模型在小规模RITE数据集上的识别准确率对比。

表1 6种模型在CNLI数据集上的识别准确率对比

Table 1 Comparison of recognition accuracy of six models on CNLI dataset %

模型	识别准确率
BiLSTM+广义池化层模型	77.4
BiLSTM+句内注意力模型	77.5
BiLSTM+最大池化层模型	78.5
HBMP模型	79.6
ESIM模型	80.5
CKEIM模型	81.4

表2 5种模型在RITE数据集上的识别准确率对比

Table 2 Comparison of recognition accuracy of five models on RITE dataset %

模型	识别准确率
BiLSTM+广义池化层模型	65.3
BiLSTM+句内注意力模型	66.8
HBMP模型	64.6
ESIM模型	69.0
CKEIM模型	72.3

由表1可以看出,CKEIM模型在大规模CNLI数据集上达到81.4%的识别准确率,比ESIM模型提高了0.9个百分点,也明显优于其他经典模型。由表2可以看出,CKEIM模型在小规模RITE数据集上优势更加明显,超出ESIM模型3.3个百分点。可见,CKEIM模型在两种数据集上都要优于ESIM基线模型及其他经典模型,且在小规模RITE数据集上优势更加明显,因此其泛化能力更强,从而说明本文CKEIM模型通过将中文外部语义知识融入神经网络,提高了模型识别准确率和泛化能力。

为比较不同规模训练集下外部语义知识对识别效果的影响,更进一步验证模型识别准确率和泛化能力,将CNLI数据集进行拆分并在整个训练集上随机采样得到原训练集4%、15%、50%和100%数据规模的训练集,然后分别在这些训练集上对本文CKEIM模型与BiLSTM+广义池化层模型、BiLSTM+句内注意力模型、HBMP模型和ESIM模型进行训练,得到如表3所示的实验结果,并据此得到如图4所示的柱状图,其中:对于4%和15%数据规模的训练集,设置最佳 λ 值为5;对于50%和100%数据规模的训练集,设置最佳 λ 值为1。

表3 5种模型在不同训练集规模下的识别准确率对比

Table 3 Comparison of recognition accuracy of five models at different training set scales %

模型	识别准确率			
	4%数据规模的训练集	15%数据规模的训练集	50%数据规模的训练集	100%数据规模的训练集
BiLSTM+广义池化层模型	55.9	63.1	71.0	77.4
BiLSTM+句内注意力模型	55.2	68.0	72.5	77.5
HBMP模型	59.0	65.2	70.7	79.6
ESIM模型	59.5	68.1	76.6	80.5
CKEIM模型	63.0	71.8	78.1	81.4

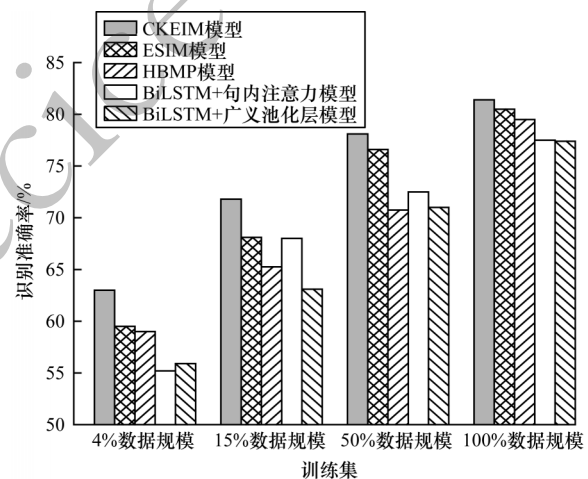


图4 不同训练集规模下的模型识别准确率对比

Fig.4 Comparison of model recognition accuracy at different training set scales

由表3和图4的数据可以看出,当训练数据非常有限时,即只有4%数据规模的训练集时,ESIM模型的识别准确率为59.5%,而CKEIM模型为63.0%,超出其3.5个百分点。在15%、50%和100%数据规模的训练集下识别准确率始终都高于ESIM模型,提升比例分别为3.7%、1.5%和0.9%,整体呈递减趋势。该结果进一步说明融合外部语义知识的CKEIM模型

具有更好的识别性能和泛化能力,且训练数据量越小,增强效果越明显。

4 结束语

本文提出一种基于外部语义知识的CKEIM模型,从HowNet和同义词词林知识库中提取外部知识,建立注意力权重矩阵并组成特征向量融入神经网络训练过程中,增强模型识别性能和泛化能力。实验结果表明,CKEIM模型在不同规模的训练集下的识别准确率均优于对比模型,并且训练数据量越小,其识别准确率和泛化能力的增强效果越明显。后续工作可将中文知识图谱融入神经网络训练过程中,进一步提高CKEIM模型的识别准确率及鲁棒性。

参考文献

- [1] GUO Maosheng, ZHANG Yu, LIU Ting. Research advances and prospect of recognizing textual entailment and knowledge acquisition [J]. Chinese Journal of Computers, 2017, 40(4): 119-140. (in Chinese)
郭茂盛,张宇,刘挺. 文本蕴含关系识别与知识获取研究进展及展望[J]. 计算机学报, 2017, 40(4): 119-140.
- [2] MACCARTNEY B, MANNING C D. Natural language inference[M]. Stanford, USA: Stanford University, 2009.
- [3] ZHANG Zhichang, YAO Dongren, LIU Xia, et al. Textual entailment recognition fused with syntactic structure transformation and lexical semantic features [J]. Computer Engineering, 2015, 41(9): 199-204. (in Chinese)
张志昌,姚东任,刘霞,等. 融合句法结构变换与词汇语义特征的文本蕴含识别[J]. 计算机工程, 2015, 41(9): 199-204.
- [4] BOWMAN S R, ANGELI G, POTTS C, et al. A large annotated corpus for learning natural language inference [EB/OL]. [2019-11-15]. <https://arxiv.org/abs/1508.05326>.
- [5] CHEN Qian, ZHU Xiaodan, LING Zhenhua, et al. Recurrent neural network-based sentence encoder with gated attention for natural language inference [EB/OL]. [2019-11-15]. <https://arxiv.org/abs/1708.01353>.
- [6] TALMAN A, YLI-JYRÄ A, TIEDEMANN J. Natural language inference with hierarchical BiLSTM max pooling architecture [EB/OL]. [2019-11-15]. <https://arxiv.org/abs/1808.08762v1>.
- [7] ROCKTÄSCHEL T, GREFFENSTETTE E, HERMANN K, et al. Reasoning about entailment with neural attention [EB/OL]. [2019-11-15]. <https://arxiv.org/abs/1509.06664>.
- [8] WANG Shuohang, JIANG Jing. Learning natural language inference with LSTM [EB/OL]. [2019-11-15]. <https://arxiv.org/abs/1512.08849>.
- [9] WANG Z, HAMZA W, FLORIAN R. Bilateral multi-perspective matching for natural language sentences [EB/OL]. [2019-11-15]. <https://arxiv.org/abs/1702.038>.
- [10] CHEN Qian, ZHU Xiaodan, LING Zhenhua, et al. Enhanced LSTM for natural language inference [C]//Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics. Philadelphia, USA: Association for Computational Linguistics, 2017: 1657-1668.
- [11] GLOCKNER M, SHWARTZ V, GOLDBERG Y. Breaking NLI systems with sentences that require simple lexical inferences [C]//Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics. Philadelphia, USA: Association for Computational Linguistics, 2018: 650-655.
- [12] CHEN Qian, ZHU Xiaodan, LING Zhenhua, et al. Neural natural language inference models enhanced with external knowledge [C]//Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics. Philadelphia, USA: Association for Computational Linguistics, 2018: 2406-2417.
- [13] WANG X, KAPANIPATHI P, MUSA R, et al. Improving natural language inference using external knowledge in the science questions domain [C]//Proceedings of AAAI Conference on Artificial Intelligence. Palo Alto, USA: AAAI Press, 2019: 7208-7215.
- [14] TAN Yongmei, LIU Shuwen, LÜ Xueqiang. CNN and BiLSTM based Chinese textual entailment recognition [J]. Journal of Chinese Information Processing, 2018, 32(7): 11-19. (in Chinese)
谭咏梅,刘姝雯,吕学强. 基于CNN与双向LSTM的中文文本蕴含识别方法[J]. 中文信息学报, 2018, 32(7): 11-19.
- [15] DONG Zhendong, DONG Qiang. HowNet—a hybrid language and knowledge resource [C]//Proceedings of International Conference on Natural Language Processing and Knowledge Engineering. Washington D. C.: IEEE Press, 2003: 820-824.
- [16] MEI Jiaju. Synonym CiLin [M]. Shanghai: Shanghai Lexicographic Publishing House, 1983. (in Chinese)
梅家驹. 同义词词林[M]. 上海: 上海辞书出版社, 1983.
- [17] PENG Qi, ZHU Xinhua, CHEN Yishan, et al. IC-based approach for calculating word semantic similarity in CiLin [J]. Application Research of Computers, 2018, 35(2): 400-404. (in Chinese)
彭琦,朱新华,陈意山,等. 基于信息内容的词林词语相似度计算[J]. 计算机应用研究, 2018, 35(2): 400-404.
- [18] NIU Yibin, XIE Ruobing, LIU Zhiyuan, et al. Improved word representation learning with sememes [C]//Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics. Philadelphia, USA: Association for Computational Linguistics, 2017: 2049-2058.
- [19] LIU Jiangming, XU Jinan, ZHANG Yujie. An approach of hybrid hierarchical structure for word similarity computing by HowNet [C]//Proceedings of the 6th International Joint Conference on Natural Language Processing. Washington D. C., USA: IEEE Press, 2013: 927-931.
- [20] SONG Y, SHI S M, LI J, et al. Directional skip-gram: explicitly distinguishing left and right context for word embeddings [C]//Proceedings of 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Philadelphia, USA: Association for Computational Linguistics, 2018: 175-180.
- [21] KINGMA D P, BA J. Adam: a method for stochastic optimization [EB/OL]. [2019-11-15]. <https://arxiv.org/abs/1412.6980>.
- [22] CHE Wanxiang, LI Zhenghua, LIU Ting. LTP: a Chinese language technology platform [C]//Proceedings of the 23rd International Conference on Computational Linguistics: Demonstrations. Philadelphia, USA: Association for Computational Linguistics, 2010: 13-16.