

### 基于知识增强的命名实体识别方法研究

高翔, 唐积强, 朱俊武, 梁明轩, 李阳

#### 引用本文

高翔, 唐积强, 朱俊武, 梁明轩, 李阳 [基于知识增强的命名实体识别方法研究](#) [J]. 计算机科学, 2023, 50(6A): 220700153-6.

GAO Xiang, TANG Jiqiang, ZHU Junwu, LIANG Mingxuan, LI Yang. [Study on Named Entity Recognition Method Based on Knowledge Graph Enhancement](#) [J]. Computer Science, 2023, 50(6A): 220700153-6.

---

#### 相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

##### Similar articles recommended (Please use Firefox or IE to view the article)

#### [基于多特征融合的GRU-LSTM大学生就业动态预测](#)

College Students Employment Dynamic Prediction of Multi-feature Fusion Based on GRU-LSTM  
计算机科学, 2023, 50(6A): 220500056-6. <https://doi.org/10.11896/jsjcx.220500056>

#### [基于深度学习的超高频标签识别系统](#)

Tag Identification for UHF RFID Systems Based on Deep Learning

计算机科学, 2023, 50(6A): 220200151-6. <https://doi.org/10.11896/jsjcx.220200151>

#### [基于动态时空神经网络的城市交通流量预测方法](#)

City Traffic Flow Prediction Method Based on Dynamic Spatio-Temporal Neural Network

计算机科学, 2023, 50(6A): 220600266-7. <https://doi.org/10.11896/jsjcx.220600266>

#### [面向交通流量预测的时空Graph-CoordAttention网络](#)

Spatial-Temporal Graph-CoordAttention Network for Traffic Forecasting

计算机科学, 2023, 50(6A): 220200042-7. <https://doi.org/10.11896/jsjcx.220200042>

#### [基于多模态特征融合的时间序列异常检测](#)

Anomaly Detection of Time-series Based on Multi-modal Feature Fusion

计算机科学, 2023, 50(6A): 220700094-7. <https://doi.org/10.11896/jsjcx.220700094>

# 基于知识增强的命名实体识别方法研究

高翔<sup>1,2</sup> 唐积强<sup>3</sup> 朱俊武<sup>1</sup> 梁明轩<sup>1,2</sup> 李阳<sup>1,2</sup>

1 扬州大学信息工程学院 江苏扬州 225000

2 中国科学院计算技术研究所 北京 100190

3 国家计算机网络应急技术处理协调中心 北京 100029

(stevegao66@163.com)

**摘要** 命名实体识别作为自然语言处理中一项十分基础的任务,其目的是从一段用自然语言描述的文本中识别出相应的实体及类型。知识图谱作为以三元组形式存在的外部知识,已经在很多自然语言处理任务中得以应用并取得了良好效果。文中提出了一种基于知识图谱信息增强的注意力对齐命名实体识别方法,首先通过嵌入层和注意力机制嵌入知识图谱信息,获取知识图谱三元组信息的表示;其次通过 BERT-BiLSTM 获取句子的上下文表示;然后通过一种注意力对齐模块分配三元组权重融合知识图谱信息与句子信息的表示;最后通过 softmax 控制融合后的表示向量的预测输出,进而获取实体的标签。该方法有效避免了因知识图谱的融合而改变原句子的语义信息,同时也使得句子中的词向量具有丰富的外部知识。所提方法在中文通用数据集 MSRA 和医疗领域专用数据集 Medicine 上的 F1 值分别达到了 95.73% 和 93.80%,相比基线模型提升了 1.21% 和 1.3%。

**关键词** 命名实体识别;知识图谱增强;注意力机制;深度学习

中图法分类号 TP391

## Study on Named Entity Recognition Method Based on Knowledge Graph Enhancement

GAO Xiang<sup>1,2</sup>, TANG Jiqiang<sup>3</sup>, ZHU Junwu<sup>1</sup>, LIANG Mingxuan<sup>1,2</sup> and LI Yang<sup>1,2</sup>

1 College of Information Engineering, Yangzhou University, Yangzhou, Jiangsu 225000, China

2 Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China

3 National Computer Network Emergency Response Technical Team/Coordination Center of China, Beijing 100029, China

**Abstract** Named entity recognition is a very basic task in natural language processing, and its purpose is to identify the corresponding entities and types from a text described in natural language. As external knowledge in the form of triples, knowledge graphs have been applied in many natural language processing tasks and achieved good results. This paper proposes an attention-aligned named entity recognition method based on knowledge graph information enhancement. Firstly, the knowledge graph information is embedded through the embedding layer and attention mechanism to obtain the representation of the knowledge graph triple information. Secondly, the sentence is obtained through BERT-BiLSTM. Then, an attention alignment module is used to assign triple weights to fuse the representation of knowledge graph information and sentence information. Finally, the prediction output of the fused representation vector is controlled by softmax, and the label of the entity is obtained. This method effectively avoids changing the semantic information of the original sentence due to the fusion of knowledge graphs, and also enables the word vectors in the sentence to have rich external knowledge. The proposed method achieves F1 values of 95.73% and 93.80% on the Chinese general data set MSRA and the medical domain specific data set Medicine, respectively, achieving advanced performance.

**Keywords** Named entity recognition, Knowledge graph enhancement, Attention mechanism, Deep learning

## 1 引言

如今,伴随着人工智能领域的高速发展,抽取非结构化文本中包含的信息已经成为自然语言处理(Natural Language

Processing, NLP)领域重要的基础任务之一。其中,命名实体识别(Named Entity Recognition, NER)任务作为一个基础性的研究课题,近年来研究方法层出不穷。此任务由 Rau<sup>[1]</sup>于 1991 年首次提出,随后在信息抽取等领域得到了广泛应用,

基金项目:国家 242 信息安全计划项目(2021A008);北京市科技新星计划交叉学科合作课题(Z191100001119014);国家重点研发计划重点专项(2017YFC1700300, 2017YFB1002300);国家自然科学基金(61702234)

This work was supported by the National 242 Information Security Program(2021A008), Beijing NOVA Program(Cross-discipline, Z191100001119014), National Key Research and Development Program of China (2017YFC1700300, 2017YFB1002300) and National Natural Science Foundation of China(61702234).

通信作者:唐积强(tjq@cert.org.cn)

并且作为一个序列标注子任务被应用到测评任务中。目前,该任务普遍被建模为一个序列标注模型,旨在通过序列标注的方法抽取并识别出文本中特定类型的实体,如人名、地名、组织等,以便为一些下游任务奠定基础。

早期的命名实体识别模型主要使用基于统计学和机器学习的方法,其中,条件随机场(Conditional Random Field, CRF)和隐马尔可夫模型(Hidden Markov Model, HMM)是统计学模型在命名实体识别任务中最常用的两种方法并且取得了良好的效果。具体而言,Xu等<sup>[2]</sup>在条件随机场的基础上建立了中文分词、命名实体识别和词性标注3个系统。这3个系统系统采用了大量语言特征。对于分割任务,根据每个字符的置信度调整 BIO 标签。基于统计学的方法具有全局最优等优点,但同时也存在着训练时间长、局部标签依赖等不可忽视的缺点。

近年来,基于神经网络而衍生出的深度学习被广泛应用于命名实体识别任务中。这种方法不再像之前的方法那样需要大量的手工特征工程以及领域专家提供的领域知识。其中,BERT<sup>[3]</sup>模型作为一种基于多层 Transformers 结构的预训练语言模型,已经在自然语言处理的多个领域取得了 SO-TA 的效果,该模型在词向量的特征表示上融合了诸多其他模型的优点并摒弃了相关的缺点。越来越多基于该模型衍生出的各种预训练模型被应用于 NER 任务,如 RoBERTa<sup>[4-6]</sup>, Merge-BERT<sup>[7]</sup>等。但目前的大多数方法只是简单地将预训练语言模型和 LSTM 与预测层相结合,这些方法存在着不能充分学习各个词的语义信息,忽略了外部知识对句子语义的正面作用等一系列问题。

针对如上所述的问题,本文提出了一种基于知识增强的注意力对齐命名实体识别方法(Knowledge Augmented Attention Alignment NER, KANER)。该方法首先通过嵌入层和注意力机制嵌入知识图谱信息,获取知识图谱三元组信息的表示;然后将其与 BERT-BiLSTM 获取的句子上下文表示进行融合,使得句子具有更多的外部知识;最后通过 softmax 层预测最终的标签分数。为了验证模型的有效性,本文在通用的中文 NER 数据集 MSRA 和医疗领域 NER 数据集上进行了实验,均取得了卓越的效果。

总的来说,本文的主要贡献如下:

(1)通过嵌入层与自注意力机制嵌入知识图谱信息,有效避免了由于外部知识的引入而改变句子的语义;

(2)运用一种注意力对齐方式分配知识图谱三元组的权重,从而可以有效地融合知识图谱信息和句子信息,使句子中的词向量具有更加丰富的外部知识。

本文首先给出相关工作;第2节给出本文的方法;第3节详述实验结果并加以分析;最后总结全文。

## 2 相关工作

### 2.1 命名实体识别

由于各个语种之间文字构成的差异,命名实体识别任务也存在着区别。对于中文而言,由于中文数据集中各个汉字之间的排列非常紧密,没有空格,这使得中文领域命名实体识别的难度大大增加;而对于英文而言则恰恰相反,由于其单词构造的特殊性,每个单词之间存在空格,其难度相较于中文而言较为简单<sup>[8]</sup>。除此之外,对于其他的小语种领域,虽然用来进行训练的数据集可能很少,但也有大批学者进行了相应的

研究<sup>[9-10]</sup>。早期传统的命名实体识别任务使用最广泛的是 CRF 模型。Li 等<sup>[11]</sup>最早在命名实体识别任务中加入了 CRF,他们通过大量词汇测试将具有特征归纳的 CRF 应用于印地语命名实体识别任务并取得了不错的效果。Konkol 等<sup>[12]</sup>基于 CRF 构造了一个捷克语命名实体识别系统,该系统被构造为输出平面或分层命名实体,从而可以使用捷克语的所有已知系统进行评估。此外,隐马尔可夫模型(HMM)也是早期运用于 NER 任务的一种方法。Li 等<sup>[13]</sup>提出了一种条件隐马尔可夫模型(Conditional Hidden Markov Model, CHMM),在此前常见的马尔可夫模型上,该模型考虑到通过预训练的语言模型具有较强的上下文表示能力这一特征,以此来增强模型的能力。CHMM 是用于多源标签去噪的 HMM 变体。它将真实实体标签建模为隐藏变量,并从观察到的噪声标签中推断出它们。同时还使用替代训练方法进一步完善 CHMM。然而,这些方法需要大量的人工来提取相关的特征,并且需要大量的专家知识标注样本,在模型的训练速度上存在着明显的缺陷。

近年来,伴随着神经网络和预训练语言模型的高速发展,基于该方法的模型被越来越多的学者运用到 NER 任务上。该方法不再像之前的方法那样需要大量的手工特征工程以及领域专家提供的领域知识,已经成为了 NER 任务的主流趋势。Collobert 等<sup>[14]</sup>将卷积神经网络(CNN)对隐藏向量进行特征提取,并且构造了一种句子级别的似然函数来估算模型的效果,该方法通过对特征的提取在 NER 任务上取得了不错的效果。Peng 等<sup>[15]</sup>提出了一种将 LSTM 和 CRF 相结合的模型联合训练命名实体识别和分词任务,该工作的  $F$  值比之前发布的相关结果提高了近 5%。Liu 等<sup>[16]</sup>通过在混合半马尔可夫 CRF 架构上添加一个简单的模块,进而大大提升了分段神经命名实体识别的性能。Lample 等<sup>[17]</sup>使用一种 BiLSTM-CRF 的结构和另一种基于转换的结构,依赖于从未注释的语料库中学习到的无监督单词表示,在 4 种语言的命名实体识别任务中展现出优异的效果。此外,基于卷积神经网络(CNN)<sup>[18]</sup>和混合神经网络(HNN)<sup>[19]</sup>的方法在命名实体识别任务上的应用也非常广泛,并且都展示出了较好的效果。但基于简单神经网络模型的命名实体识别方法存在着单词语义信息不足、训练速度慢的缺点。自 Google 的 Devlin 等<sup>[3]</sup>提出 BERT 模型之后,将预训练的语言模型与神经网络相结合成为了研究热点。该方法可以大大减少模型的训练参数,只需要在特定任务上进行微调即可得到卓越的效果。

### 2.2 知识图谱融合

考虑到基于预训练语言模型和神经网络相结合的命名实体识别方法虽然可以在大规模开放领域语料上进行预训练并获得单词的语言表示,然后在特定的下游任务中微调,从而吸收特定领域的知识,但是由于预训练语言模型和下游微调之间存在着差异,该方法在知识驱动的任务上表现不佳,于是学者们开始尝试将知识图谱融入模型中,从而改善模型在特定领域的表现。

在预训练语言模型出现之前,一些研究已经尝试将知识图谱与单词的词向量相结合,Wang 等<sup>[20]</sup>基于 Word2vec<sup>[21]</sup>的思想提出了一种将实体和对应的关系联合嵌入到同一个连续的向量空间中的新颖的方法。Han 等<sup>[22]</sup>应用卷积神经网络和知识图谱完成任务,以此联合学习了文本和知识的双重表示。伴随着预训练语言模型的出现,展开了对将知识图谱

与之结合的研究,但总体来说该方面的工作很少。Zhang等<sup>[23]</sup>提出的名为ERNIE的方法是该方向研究的第一个成果。该方法利用大规模文本语料库和知识图谱来训练增强语言表示模型,使得其可以同时充分利用词汇、句法和知识信息,但缺点是忽略了实体之间的关系。Bosselut等<sup>[24]</sup>提出了一种名为COMET的方法,该方法使用知识图谱中的三元组作为原始的语料库训练GPT<sup>[25]</sup>,从而进行知识的学习,但存在着训练时间过长、效率低下的缺点。

### 3 方法

本节将具体介绍KANER的具体结构及做法。首先进行

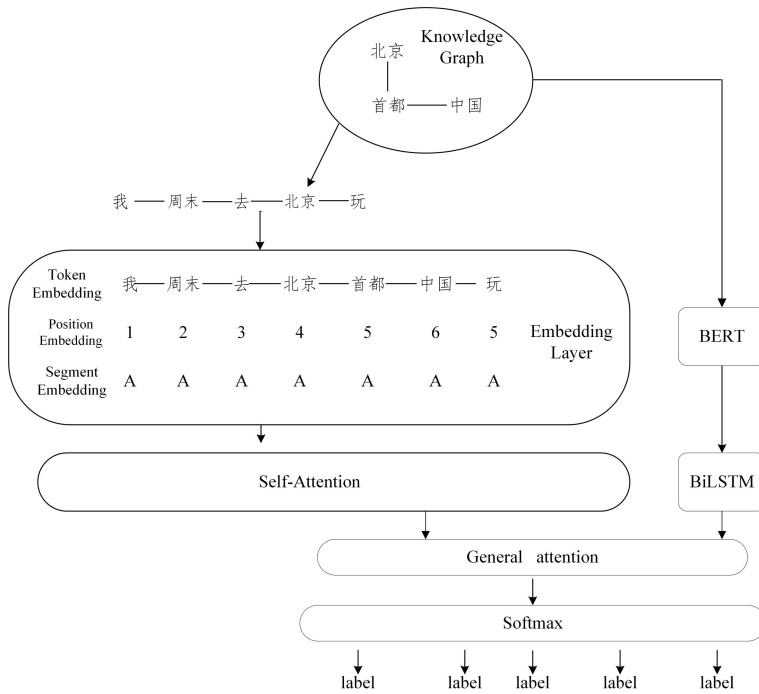


图1 基于知识图谱增强的命名实体识别模型流程图

Fig. 1 Flowchart of named entity recognition model based on knowledge graph enhancement

#### 3.2.1 知识图谱与文本嵌入层

如图1所示,首先将知识图谱中与句子相对应的三元组插入到原句中从而生成融合知识图谱的新句子。对于一个给定句子 $s=[x_0, x_1, x_2, \dots, x_n]$ ,查找其中的每一个词 $x_i, i \in (0, n)$ 是否在知识图谱中有对应的三元组,若有,则在相应位置插入三元组。假设该词在知识图谱中三元组的表示形式为 $K=[(x_i, r_{i_0}, x_{i_0}), \dots, (x_i, r_{i_k}, x_{i_k})]$ ,原句则变为: $s=[x_0, x_1, \dots, x_i(r_{i_0}, x_{i_0}), \dots, (r_{i_k}, x_{i_k}), \dots, x_n]$ 的形式,经过此步骤形成了一个融入知识图谱三元组的新句子。

其次将融合知识图谱信息的新句子输入嵌入层。由于增加了知识图谱信息,因此直接使用BERT位置嵌入会造成语义丢失进而影响模型的效果,受Liu等<sup>[26]</sup>的方法启发,我们设计一个新的嵌入层。如图1所示,该嵌入由TokenEmbedding, PositionEmbedding和SegmentEmbedding3个部分组成。其中,Tokenembedding表示将输入的单词表示为词向量,由于中文的特殊构造,在处理时可以分割为词向量或者单个字向量。在本文中,选用北京大学开源分词工具pkuseg对文本进行分词从而得到词向量,在开始时加入的CLS1标志可以用于后期的分类任务;Positionembedding用于标记每个单词在句子中的位置,由于知识图谱的加入,按照BERT

任务描述,然后详细介绍模型结构及知识图谱融合过程。

#### 3.1 任务描述

给定一个输入句子 $x=\{\omega_1, \omega_2, \dots, \omega_n\}$ ,其中 $\omega_i$ 表示经过分词之后的每个单词,模型输出每个单词 $\omega_i$ 对应的实体类型及在每个实体中的位置,例如,输出集合 $y \in \{B-LOC, I-LOC, B-PER, I-PER, O\}$ 。由于本文模型加入了知识图谱,输入是句子 $x$ 和单词对应的知识图谱中的三元组,预训练语言模型Bert被用于生成和训练词向量。

#### 3.2 模型结构

本文模型主要由知识图谱与文本嵌入层、自注意力层、知识图谱融合层组成,具体的模型图如图1所示。

类型的编码设计了新的位置编码,对于知识图谱中新加入的三元组中的词,由于其加入位置紧接在被加入词的后方,因此将其位置编码按照原来句子中该词之后的位置对三元组进行编码,而未加入知识图谱的原句子中相邻词位置编码也被编码为相同的数字,后期可以通过训练模型学习得到单词的位置信息;Segmentembedding是段落标记,用于区分做分类任务时两个不同的输入句子,对于在相同句子中的单词给予相同的段落标记 $E_i$ 。

在嵌入层之后紧接着一个双向长短期记忆网络BiLSTM层,使用LSTM可以动态捕获隐藏向量的数据信息,同时可以记忆保存数据信息。在本文中,使用Graves等<sup>[27]</sup>提出的改进LSTM机制,这种改进的机制可以选择性地对记忆单元的信息进行过滤和保存,从而解决了原有机制存在的问题。

#### 3.2.2 自注意力层

在BERT基础模型中,是通过堆叠多层Transformers模型来进行编码,其中的自注意力机制用来更好地提取向量的特征。而本文模型融合了知识图谱的信息,三元组编码后的向量在进行自注意力计算时会影响打分函数的判断,使得原本不相关的词之间得分不为零,进而影响模型的效率。因此,受Liu等<sup>[26]</sup>的启发,我们重新优化了自注意力机制,具体如下:

$$Q^{i+1} = h^i W_q, K^{i+1} = h^i W_k, V^{i+1} = h^i W_v \quad (1)$$

$$F^{i+1} = \text{softmax}\left(\frac{K^{i+1T} Q^{i+1} - S}{\sqrt{d_k}}\right) \quad (2)$$

$$h^{i+1} = V^{i+1} F^{i+1} \quad (3)$$

其中,式(1)中的 $W_q, W_k, W_v$ 是可以训练的参数矩阵, $h^i$ 是第*i*层的隐藏向量,由此可以得到该注意力机制的3个向量 *Query*、*Key* 和 *Value*。然后在式(2)中使用第*i*+1层的每个 *Query* 向量 $q_i$ 分别和各个 *Key* 向量 $k_i$ 点乘得到系数矩阵后减去 *S* 来计算权重。这里的 *S* 是一个 mask 矩阵,该矩阵是对称的,在加入知识图谱后,为了防止知识图谱带来的负面影响,即可能由于三元组的加入改变原句子的语义,从而导致注意力权重分布不均匀,所以在后面加上之前遮蔽矩阵的数值。对于输入句子[我,周末,去,北京,首都,中国,玩],mask 矩阵的维度为 $7 \times 7$ ,每行和每列都对应于该句中的每个词。由于[首都,中国]两个词是知识图谱中加入的,因此只有“北京”在自注意力时对这两个词可见。在矩阵中, $S_{1,5}$ 和 $S_{1,6}$ 位置被赋值为0,其他单词对于这两个词均不可见,相应位置被复制为 $+\infty$ ,如式(4)所示:

$$S_{ij} = \begin{cases} 0, & x_i \text{ 与 } x_j \text{ 关联} \\ +\infty, & x_i \text{ 与 } x_j \text{ 不关联} \end{cases} \quad (4)$$

$\sqrt{d_k}$ 代表隐藏向量的维度。最后进行 softmax 归一化,此步可以使得在做第一次注意力计算时原本不相关的词注意力得分为0,其他位置不受影响正常进行运算。随后几次运算时,由于每个隐藏向量之间的彼此运算,可以促使整个句子中的每个词向量都可以间接融合知识图谱的相关信息。最后使用式(3)将注意力机制运算过后得到的权重与 *Value* 进行运算得到最终的隐藏向量表示。

### 3.2.3 知识图谱融合层

为了知识图谱能够更好地融入模型,本模型在此层运用了一种注意力对齐方式来计算知识图谱中三元组的权重。本文将前几步输出的隐藏向量与将原句子经过 BERT-LSTM<sup>[27]</sup>模型后输出的隐藏向量按照该注意力对齐方式进行运算,使得知识图谱与原句子融合,具体的做法为:

$$a_i(s) = \frac{\exp(\text{score}_s)}{\sum_j \exp(\text{score}_j)} \quad (5)$$

$$\text{score}_s = h_s^T W_a h_t \quad (6)$$

$$c_t = \sum a_i h_s \quad (7)$$

$$h_t' = \text{softmax}(c, h_t) \quad (8)$$

其中, $h_s$ 和 $h_t$ 分别表示加入知识图谱后的句子得到的隐藏向量和原句子经过 BERT-LSTM 嵌入后得到的隐藏向量。通过 score 打分函数也就是对齐方式来计算带有知识图谱的隐藏向量对每个原句子隐藏向量的权重,经过 softmax 归一化后与原隐藏向量相乘得到最终的隐藏表示。经过此步可以使得句子具有丰富的语义信息,更好地融合外部知识。最后通过预测层得出最终的预测标签。

## 4 实验

### 4.1 数据集及参数设置

本文采用了中文命名实体识别公开数据集 MSRA\_NER<sup>[28]</sup>,该数据集由微软发布,包含了人名(PER)、地名(LOC)和组织名(ORG)3种类型的实体。此外,还采用了一个医疗领域的

专业命名实体识别数据集 Medicine\_NER,该数据集是由 CCKS 2017<sup>[29]</sup>发布的临床命名实体识别(CNER)任务,目标是从电子病历中提取与医学相关的实体名称。将每个数据集划分成训练集(Train)、验证集(Dev)和测试集(Test),具体信息如表1所列。

表1 实验数据集

Table 1 Experimental datasets

| dataset  | Train  | Dev   | Test  |
|----------|--------|-------|-------|
| MSRA     | 20 865 | 2 319 | 4 637 |
| Medicine | 6 920  | 940   | 756   |

本文数据集采用的标注方式为 BIO 标注,其中 B 表示一个实体的开头,I 表示实体的中间部分,O 表示非实体。

本文采用的深度模型框架为 pytorch,BERT 生成的词向量维度为 786,训练和测试的 batch\_size 为 10,学习衰退率(decay\_rate)为 0.98,权重衰退(weight\_decay)为 $1 \times 10^{-6}$ ,模型训练的学习率(Learning\_rate)为 $2 \times 10^{-5}$ 。

### 4.2 知识图谱

本文采用的知识图谱为 CN-Dbpedia<sup>[29]</sup>,该知识图谱是复旦大学知识工作实验室开发的大型开放领域百科知识库,涵盖数千万实体和数亿关系。精炼的 CN-DBpedia 总共包含 517 万个三元组。为了验证模型在特定领域的的能力,采用医疗领域知识图谱 medical\_KG 在医疗数据集 Medicine-NER 上进行实验。

### 4.3 评价指标

本文采用精确率(precision)、召回率(recall)和 F1 值作为模型的评价指标。具体公式如下:

$$\text{Precision} = \frac{\text{correct}}{\text{pred\_entities\_num}}$$

$$\text{Recall} = \frac{\text{correct}}{\text{gold\_entities\_num}}$$

$$F1 = 2 \frac{\text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}}$$

其中,correct 表示预测正确的标签数量,pred\_entities\_num 表示预测的标签总数量,gold\_entities\_nums 表示样本中的标签总数量。

### 4.4 总体实验结果及分析

为了验证模型的有效性,在 MSRA 数据集上选取了几个对比模型,具体如下:

(1)BiLSTM-CRF<sup>[30]</sup>:该模型是最基础的神经网络命名实体识别模型,由双向长短期记忆网络加条件随机场组成。

(2)BERT-BiLSTM-CRF<sup>[31]</sup>:将 BERT 预训练模型与双向长短期记忆神经网络和条件随机场相结合,可以减少模型的训练参数。

(3)BERT-BiLSTM-self-attention-CRF<sup>[32]</sup>:在 CRF 层预测前加上一层自注意力机制,以此捕捉隐藏向量之间的相关程度。

(4)Lattice-LSTM-CRF 模型<sup>[33]</sup>:该模型将字词信息融合并且考虑了字的潜在信息,可有效避免分词错误的传递,在中文命名实体识别任务上达到了良好的效果。

(5)BERT-BiGRU-CRF 模型<sup>[34]</sup>:该模型构建字向量用于表征语义特征,根据上下文动态生成语义向量。

表2列出了本文模型和对比模型在 MSRA 数据集上的实验数据对比。

表2 本文模型和对比模型在 MSRA 数据集上的实验结果

| 模型名称                           | P     | R     | F1    |
|--------------------------------|-------|-------|-------|
| BiLSTM-CRF                     | 89.88 | 87.93 | 88.89 |
| BERT-BiLSTM-self-attention-CRF | 94.30 | 94.74 | 94.52 |
| Lattice-LSTM-CRF               | 93.57 | 92.79 | 93.18 |
| BERT-BiLSTM-CRF                | 94.48 | 94.48 | 94.48 |
| BERT-BiGRU-CRF                 | 94.19 | 94.16 | 94.18 |
| KANER                          | 95.58 | 95.89 | 95.73 |

由表2可以看出,本文的KANER在MSRA数据集上实现了最佳的效果,在测试集上的P值、R值和F1值分别达到了95.58%、95.89%和95.73%。由于在KANER中加入了知识图谱,引入了外部知识,将其与原句子的隐藏向量相融合,可以使得隐藏向量具有丰富的语义信息,从而提高模型的准确率,这也证明了知识图谱的加入可以有效地提高命名实体识别任务效率,给模型带来正面影响。

在所有对比模型中,没有预训练语言模型的BiLSTM-CRF模型的效果最差,3个值均未超过90%,由此可见预训练的语言模型,例如BERT对命名实体识别任务的准确性起到了较为明显的提升作用。

通过对Lattice-LSTM-CRF和BERT-BiLSTM-CRF两个实验模型的对比发现,后者比前者在3个值上均提高了1%左右,这说明BERT对于句子特征的抽取能力要强于Lattice,通过抽取提取的特征比字词融合的特征更准确。

通过对BERT-BiLSTM-CRF和BERT-BiGRU-CRF两个模型的对比可以看出,前者在3个值上略高于后者,说明LSTM在特征的提取上略好于GRU模型。

而通过BERT-BiLSTM-CRF和BERT-BiLSTM-self-attention-CRF两个模型的对比可以看出,在LSTM层后加一层自注意力机制对于模型效果的提升并不明显,这可能是由于LSTM网络与自注意力机制相互影响的结果。

为了验证引入知识图谱后对于专业领域内命名实体识别任务的有效性,采用Medicine-NER数据集进行实验,对比不引入知识图谱的BERT-BiLSTM-CRF模型的实验结果,如表3所列。

表3 对比模型在Medicine数据集上的实验结果

| 模型名称            | P     | R     | F1    |
|-----------------|-------|-------|-------|
| BERT-BiLSTM-CRF | 91.90 | 93.11 | 92.50 |
| KANER           | 93.45 | 94.16 | 93.80 |

通过对比可以看出,本文模型在该特定领域数据集上的3个值分别达到了93.45%、94.16%、93.80%,均好于未融入知识图谱的命名实体识别模型。这证明了外部知识对特定领域命名实体识别模型的有效性。

## 4.5 消融实验结果及分析

### 4.5.1 mask矩阵对模块的影响

由于mask矩阵可以控制各个隐藏向量之间的关联性,防止由于知识图谱的加入使得原本不相关的向量之间自注意力得分过高而影响打分机制的判断,因此本节对比了不加入遮蔽矩阵与原模型的效果,结果如表4所列。由表4可以看出,在自注意力时若不加入mask矩阵S,模型在两个数据集

上的表现均有所下降。这是因为mask矩阵可以控制每个单词彼此之间的关联性,使得在做第一次自注意力计算时不相关的词之间的注意力分数为零,随后在做第二次注意力计算时可以使这些单词彼此之间间接地融入知识图谱的相关知识。通过加入该矩阵可以使注意力机制有更好的打分判断。

表4 mask矩阵对模型的影响

Table 4 Effect of mask matrix on model

| model           | MSRA  |       |       | Medicine |       |       |
|-----------------|-------|-------|-------|----------|-------|-------|
|                 | P     | R     | F1    | P        | R     | F1    |
| KANER           | 95.58 | 95.89 | 95.73 | 93.45    | 94.16 | 93.80 |
| KANER without S | 94.23 | 94.56 | 94.39 | 92.86    | 93.32 | 93.01 |

### 4.5.2 融合方式对模型的影响

在将知识图谱嵌入之后,本文通过一种注意力对齐方式(gerenal)将其与基础的命名实体识别模型融合,在本节中将比较3.2.4节中的融合方式和直接将两个隐藏向量进行拼接(concat)在MSRA数据集上的效果,结果如图2所示。

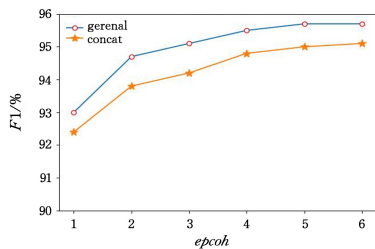


图2 两种结合方式在MSRA数据集上的F1值

Fig. 2 F1 values of two binding methods on MSRA dataset

可以看出,通过注意力对齐的方式的效果要好于直接拼接,这是因为该对齐方式可以有效地选择加入知识图谱后句子隐藏向量的权重,以此更好地将其与原句子的隐藏向量进行融合,这种方式可以更有效地利用知识图谱。

**结束语** 本文提出了一种将知识图谱嵌入与基础命名实体识别模型相融合的方法。该模型将知识图谱利用修改后的自注意力机制进行嵌入,通过遮蔽矩阵来判断三元组的影响范围,避免由于三元组的加入而改变了语义信息,然后将嵌入后的隐藏向量与BERT模型输出后的原句子的隐藏向量运用注意力对齐方式进行融合,经过预测层得到最后的标签。实验证明,模型的正确率和F1值在两个数据集上相比基线模型均有所提高,证明了知识图谱对命名实体识别任务的正面影响和这种融合方式的有效性。下一步将考虑如何在BERT输出后的下层模型上进行改进以更好地提升模型效果。

## 参考文献

- [1] RAU L F. Extracting company names from text [C]//Proceedings The Seventh IEEE Conference on Artificial Intelligence Application. 1991:29-32.
- [2] XU Z, QIAN X, ZHANG Y, et al. CRF-based hybrid model for word segmentation, NER and even POS tagging [C]//Proceedings of the Sixth SIGHAN Workshop on Chinese Language Processing. 2008.
- [3] DEVLIN J, CHANG M W, LEE K, et al. Bert: Pre-training of deep bidirectional transformers for language understanding [C]//Proceedings of NAACL-HLT. 2019:4171-4186.
- [4] TONG C, PENG H, DAI Q, et al. Improving Natural Language Understanding by Reverse Mapping Bytepair Encoding [C]//

- Conference on Computational Natural Language Learning.
- [5] RAFFEL C, SHAZEER N, ROBERTS A, et al. Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer[J]. *The Journal of Machine Learning Research*, 2020, 21(1): 5485-5551.
- [6] LIU Y H, OTT M, GOYAL N, et al. Roberta: A robustly optimized bert pretraining approach[J]. arXiv:1907.11692, 2019.
- [7] FISHER J, VLACHOS A. Merge and label: A novel neural network architecture for nested NER[C]//Proceedings of the 57th Conference of the Association for Computational Linguistics (ACL 2019). Florence, Italy, Volume 1: Long Papers, 2019: 5840-5850.
- [8] OUYANG X, CHEN S, ZHAO H, et al. A multi-Cross Matching Network for Chinese Named Entity Linking in Short Text [J]. *Journal of Physics: Conference Series*. IOP Publishing, 2019, 1325(1): 012069.
- [9] XIA C, ZHANG C, YANG T, et al. Multi-grained named entity recognition[C]//57th Annual Meeting of the Association for Computational Linguistics(ACL 2019). Association for Computational Linguistics(ACL), 2020: 1430-1440.
- [10] NI J, FLORIAN R. Improving Multilingual Named Entity Recognition with Wikipedia Entity Type Mapping [C]// Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, 2016: 1275-1284.
- [11] LI W, MCCALLUM A. Rapid development of Hindi named entity recognition using conditional random fields and feature induction[J]. *ACM Transactions on Asian Language Information Processing (TALIP)*, 2003, 2(3): 290-294.
- [12] KONKOL M, KONOPÍK M. CRF-based Czech named entity recognizer and consolidation of Czech NER research[C]//International Conference on Text, Speech and Dialogue. Berlin: Springer, 2013: 153-160.
- [13] LI Y, SHETTY P, LIU L, et al. BERTifying the Hidden Markov Model for Multi-Source Weakly Supervised Named Entity Recognition[C]//Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics, 2021.
- [14] COLLOBERT R, WESTON J, BOTTOU L. Natural Language Processing(almost)from scratch[J]. *Journal of Machine Learning Research*, 2011, 12(1): 2493-2537.
- [15] PENG N, DREDZE M. Improving named entity recognition for Chinese social media with word segmentation representation learning[C]//Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, 2016: 149-155.
- [16] LIU T, YAO J G, LINC Y. Towards improving neural named entity recognition with gazetteers[C]//Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, 2019: 5301-5307.
- [17] LAMPLE G, BALLESTEROS M, SUBRAMANIAN S, et al. Neural Architectures for Named Entity Recognition[C]//HLT-NAACL, 2016.
- [18] DONG X, QIAN L, GUAN Y, et al. A multiclass classification method based on deep learning for named entity recognition in electronic medical records[C]//2016 New York Scientific Data Summit(NYSDS). IEEE, 2016: 1-10.
- [19] SHAO Y, HARDMEIER C, NIVRE J. Multilingual named entity recognition using hybrid neural networks[C]//The Sixth Swedish Language Technology Conference(SLTC). 2016.
- [20] WANG Z, ZHANG J, FENG J, et al. Knowledge graph and text jointly embedding[C]//Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing(EMNLP). 2014: 1591-1601.
- [21] MIKOLOV T, CHEN K, CORRADO G, et al. Efficient Estimation of Word Representations in Vector Space[J]. arXiv: 1301.3781, 2013.
- [22] HAN X, LIU Z, SUN M. Joint representation learning of text and knowledge for knowledge graph completion [J]. arXiv: 1611.04125, 2016.
- [23] ZHANG Z, HAN X, LIU Z, et al. ERNIE: Enhanced language representation with informative entities[J]. arXiv: 1905.07129, 2019.
- [24] BOSSELUT A, RASHKIN H, SAP M, et al. COMET: Commonsense Transformers for Automatic Knowledge Graph Construction[J]. 2019.
- [25] JARS I, PANAGET F. Improving Spoken Language Understanding with information retrieval and active learning methods [C]//IEEE International Conference on Acoustics, IEEE, 2008.
- [26] LIU W, ZHOU P, ZHAO Z, et al. K-BERT: Enabling Language Representation with Knowledge Graph[C]//Proceedings of the AAAI Conference on Artificial Intelligence, 2020: 2901-2908.
- [27] GRAVES A, MOHAMED A R, HINTON G. Speech recognition with deep recurrent neural networks[C]//International Conference on Acoustics, Speech, and Signal Processing. IEEE, 2013: 6645-6649.
- [28] WANG M, YANG L, HUA X S. MSRA-MM: Bridging research and industrial societies for multimedia information retrieval[J]. *Microsoft Research Asia*, 2009, 5(6).
- [29] EGAN M E, GLCKNER-PAGEL J, AMBROSE C A, et al. Calcium-pump inhibitors induce functional surface expression of Delta F508-CFTR protein in cystic fibrosis epithelial cells[J]. *Nature Medicine*, 2002, 8(5): 485.
- [30] XU B, XU Y, LIANG J, et al. Cn-dbpedia: A never-ending chinese knowledge extraction system [C]//International Conference Industrial, Engineering and other Applications Applied Intelligent Systems, 2017: 428-438.
- [31] WU H, CHEN P, LI W, et al. Construction of Power Communication Network Knowledge Graph with BERT-BiLSTM-CRF Model Based Entity Recognition[C]//2021 IEEE 6th International Conference on Computer and Communication Systems(ICCCS). IEEE, 2021.
- [32] MAO M Y, WU C, ZHONG Y X, et al. BERT named entity recognition model with self-attention mechanism[J]. *CAAI Transactions on Intelligent Systems*, 2020, 15(4): 772-779.
- [33] ZHANG Y, YANG J. Chinese NER using lattice LSTM[EB/OL]. [2019-01-02]. <https://arxiv.org/pdf/1805.02023.pdf>.
- [34] YANG P, DONG W Y. Chinese named entity recognition method based on BERT embedding [J]. *Computer Engineering*, 2020, 46(4): 40-45, 52.



**GAO Xiang**, born in 1996, postgraduate. His main research interests include natural language processing and named entity recognition.



**TANG Jiqiang**, born in 1981, master. His main research interests include natural language processing and network security.