• 计算机工程与应用 •

基于 BERT 多知识图融合嵌入的中文 NER 模型

张凤荔[1]，黄 鑫[1]，王瑞锦[1*]，周志远[1]，韩英军[2]

(1. 电子科技大学信息与软件工程学院 成都 610054；2. 四川中烟工业有限责任公司成都卷烟厂 成都 610066)

【摘要】针对目前特定领域知识图谱构建效率低、领域已有知识图谱利用率不足、传统模型提取领域语义专业性强实体困难的问题，提出了基于 BERT 多知识图融合嵌入的中文 NER 模型 (BERT-FKG)，实现了对多个知识图通过融合语义进行实体间属性共享，丰富了句子嵌入的知识。该模型在开放域和医疗领域的中文 NER 任务中，表现出了更好的性能。实验结果表明，多个领域知识图通过计算语义相似度进行相似实体的属性共享，能够使模型吸纳更多的领域知识，提高在 NER 任务中的准确率。

关 键 词 BERT； 中文命名实体识别； 医疗领域； 多知识图融合嵌入

中图分类号 TP391 文献标志码 A doi:10.12178/1001-0548.2021400

A Chinese NER Model Based on BERT with Multi Knowledge Graph Fusion and Embedding

ZHANG Fengli[1], HUANG Xin[1], WANG Ruijin[1*], ZHOU Zhiyuan[1], and HAN Yingjun[2]

(1. School of Information and Software Engineering, University of Electronic Science and Technology of China Chengdu 610054;

2. Chengdu Cigarette Factory of Sichuan China Tobacco Industry Co., Ltd Chengdu 610066)

Abstract Aiming at the problems of low efficiency in the construction of knowledge graph in specific fields, insufficient utilization of existing knowledge graph in the field, and difficulty in extracting domain semantic professional entities from traditional models, a Chinese named entity recognition (NER) model based on Bert (bidirectional encoder representations from transformers) multi knowledge graph fusion and embedding (BERT-FKG) is proposed in this paper. It realizes the attribute sharing among entities through semantic fusion for multiple knowledge graphs and enriches the knowledge of sentence embedding. The proposed model shows better performance in Chinese NER tasks in open domain and medical field. The experimental results show that multiple domain knowledge graphs share the attributes of similar entities by calculating semantic similarity, which can make the model absorb more domain knowledge and improve the accuracy in NER tasks.

Key words BERT; Chinese named entity recognition; medical field; multi knowledge graph fusion and embedding

由于知识图谱技术在多个领域表现出了强大的态势感知和关系表征的能力，各领域构建知识图谱的需求愈发强烈，而作为构建知识图谱的关键技术之一的命名实体识别 (named entities recognition，NER)，已经成为自然语言处理领域的热点研究方向。

传统的 NER 模型将 NER 视为序列标注任务，注重对文本数据中特定实体的提取，有 CNN-based[1]、RNN[2] 以及 BiLSTM-CRF[3] 等对序列文本处理效果不错的 NER 模型，但它们由于模型能力的限制，无法很好地完成特定领域内专业性强的实体抽取任务。

随着计算机性能的提高，研究人员提出了基于预训练的模型方法，即使用高性能计算机通过构筑千万级语料的预训练任务，对单词或字符向量进行预先训练，以提高模型的起点，如 wor2vec[4] 和 Glove[5]，但它们是静态的词袋模型，不能很好地从文本序列中捕获上下文背景信息。而文献 [6] 采用

收稿日期：2021－12－27；修回日期：2022－10－11

基金项目：国家自然科学基金 (61802033，61472064，61602096)；四川省区域创新合作项目 (2020YFQ0018)；四川省科技计划重点研发项目 (2021YFS0391，2020YFG0475，2020YFG0414)

作者简介：张凤荔 (1963－)，女，博士，教授，主要从事数据挖掘与网络信息安全方面的研究.

*通信作者：王瑞锦，E-mail：ruijinwang@uestc.edu.cn

BERT 作为预训练模型，在 NER 任务中取得了较好的效果。而谷歌公司面向所有 NLP 工作者开源了多个预先训练好的 BERT 模型，针对不同的任务特点，微调 BERT 的模型参数就可以达到很好的效果，这极大地促进了 NLP 领域的发展。

虽然目前在通用知识领域，各类模型在 NER 任务中已经取得了极高的成效，但对于一些专业知识要求高的特定领域，如医疗、金融、军事、民用航空等，由于传统模型不能很好地利用行业内已有的数据集作为先验知识，导致领域知识图谱的构建总需要从头开始训练模型，造成了大量的资源浪费。

针对特定领域知识图谱构建的问题，本文以医疗领域中文 NER 为主要场景，提出了一种基于 BERT 的外部知识图谱融合嵌入的 NER 模型 (BERT-FKG)，通过将领域内已有的知识图谱进行融合嵌入，结合 BERT 预训练模型的强大文本处理能力，实现对外部先验知识的充分学习，减少模型领域化的训练耗费，实现对领域文本中实体的准确识别。

1　相关工作

自 Google 于 2018 年推出 BERT[6] 以来，研究人员不断对其进行优化，但都侧重于预训练任务和编码器优化。而近年来越来越多的研究证明，在 NER 任务中，尤其是对中文，在输入的句子中嵌入词汇相关的信息能够有效地提升模型的准确率，如 Lattice-LSTM[7]、FLAT[8]、Lex-BERT[9] 等。而同时包含了词汇信息和语义背景信息的知识图谱理论上可以取得更好的效果，因此如何利用更多的领域知识图谱信息提高 BERT 模型的 NER 能力是本文的研究方向。

1.1　知识嵌入

随着预训练语言模型的发展，越来越多的人意识到，赋予训练句子一些先验知识能够有效提升模型性能，而知识图作为语义知识的集合，包含丰富的实体信息。目前已有对模型嵌入知识图的尝试，ERNIE[10] 是融合实体信息的先驱，证明了知识嵌入能够加强语义的表征，且 ERNIE2.0[11] 又对嵌入框架进行了优化。而 LUKE[12] 也将句子中的实体信息通过位置编码嵌入到模型输入，使模型获得实体注意力，取得了不错的效果，但它们忽略了实体之间的关系。COMET[13] 使用百科知识中的三元组作为语料库来训练 GPT[14] 进行常识学习，但这样的训练效率不高，模型并不能很好地吸纳外部知识。K-BERT[15] 通过软位置编码和可见矩阵将知识图嵌入进句子中，并且屏蔽了 Transformer 可能产生的错误注意力，但它只能嵌入单个知识图；KEPLER[16] 通过句子实体与源文本融合加强实体的表示，最终计算文本融合和模型的总体损失来取得更好的知识嵌入效果。上述文献证明了外部知识的嵌入对提升基于 Transformer 模型的 NER 性能有帮助，但它们都只能嵌入有限的知识，而无法充分利用领域内更多的知识图谱。

1.2　知识融合

在早期的知识融合的工作中，字符串相似性是主要的融合依据。RDF-AI[17] 设计了一套包括数据预处理、候选实体匹配、相似实体融合、实体互连以及文本后处理模块的相似实体比较框架，并通过字词特征进行相似性比较。LIMES[18] 对于实体相似度以三角形不等式计算值来评估，将高度相似实体对抽取出来进一步计算实际相似度，并返回具有最高实际字符串相似度的实体对。HolisticEM[19] 基于实体的重叠属性和相邻实体构建了一个潜在实体对的图，然后通过图中的局部和全局属性计算实体对的实际相似度。TranE[20] 将实体和关系归一化后，在同一个向量空间获得实体向量表示作为实体相似性度量。文献 [21] 基于嵌入学习和实体相似性传播的方法，实现了对跨实体的实体间相似度比较。

随着 BERT 模型的诞生和发展，预训练模型使得实体词向量包括了丰富的字符和图信息以及领域知识等。本文通过 BERT 词向量计算嵌入知识中实体对的余弦相似度进行实体对齐，证明该方法能简单有效地完成知识融合工作，并且避免嵌入更多的模型，降低了模型整体的复杂度。

2　问题定义

定义句子$s=[w_0,w_1,w_2,\cdots,w_n]$作为一个句子标记序列，使用 BIO(begin, inside, out) 作为标记方案，$C=\{c_1,c_2,\cdots,c_n\}$为标记中预定义的实体类别，n 是句子的长度。在本文中，英语标记是在单词层面上使用的，而汉语标记是在字符层面上使用的。每个标记w_i都包含在词汇$\mathbb{V}$中，$w_i\in\mathbb{V}$对于知识图谱，定义为KG，是三元组$\varepsilon=(w_i,r_j,w_k)$的集合，其中$w_i$和$w_j$是实体的名称，$r_j\in\mathbb{V}$是它们之间的关系。所有的三元组都是 KG 的组成。

如图 1 所示，本文的命名实体识别任务首先以

待提取文本、标签以及外部知识图谱为输入，在训练阶段根据文本与标签，在外部知识图谱作为先验知识的帮助下训练模型的权重，在预测阶段计算文本对应的预测标签并与真实标签进行比较从而优化模型，最终输出文本中存在的目标实体。用符号简述为：

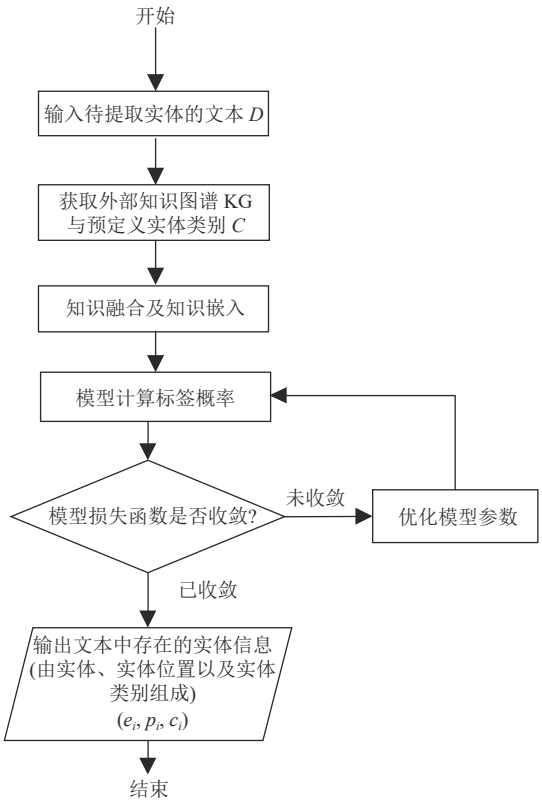

图 1 任务流程图

1) 输入目标领域的自然语言文本$S=\{s_1,s_2,\cdots,s_n\}$，$s_i=\{w_{i0},w_{i1},\cdots,w_{in}\}$；

2) 获取领域外部知识图谱$\mathrm{KG}=\{K_1,K_2,\cdots,K_n\}$，$K_i=\{\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n\}$

3) 获取数据集中预定义的实体类别$C=\{c_1,c_2,\cdots,c_n\}$；

4) 输出实体信息、实体位置和所属类别对的集合$\{(e_1,p_1,c_1),(e_2,p_2,c_2),\cdots,(e_n,p_n,c_n)\}$。其中，$e_i$是实体信息，$p_i=(d_i,b_i)$是出现在文档中的领域实体的位置信息，$b_i$是实体$e_i$在句子$d_i$中的起止位置信息，$c_i$表示所属的实体类别。

3 模型设计

3.1 模型架构

模型的整个结构如图 2 所示，主要包括知识融合、知识嵌入、混合编码以及最后的 BERT 层训练 4 个部分。知识融合主要是将外部知识库中领域的多个外部知识图谱，通过预训练好的 BERT 模型计算知识相似度进行融合，从而进一步领域化先验知识，支撑对输入句子更丰富的知识嵌入；知识嵌入部分主要负责生成句子树，构造出下一步 BERT 模型所需的输入；混合编码是根据 BERT 模型的输入结构，构造出相应的输入序列，核心工作是生成遮蔽矩阵；BERT 层则是采用 Google 预训练好的 BERT-Chinese 模型并且通过训练语料进行下游 NER 任务微调参数，通过最小化预测标签的交叉熵，实现对模型参数的调优，最终在开发集和测试集上评估模型的实体抽取性能。

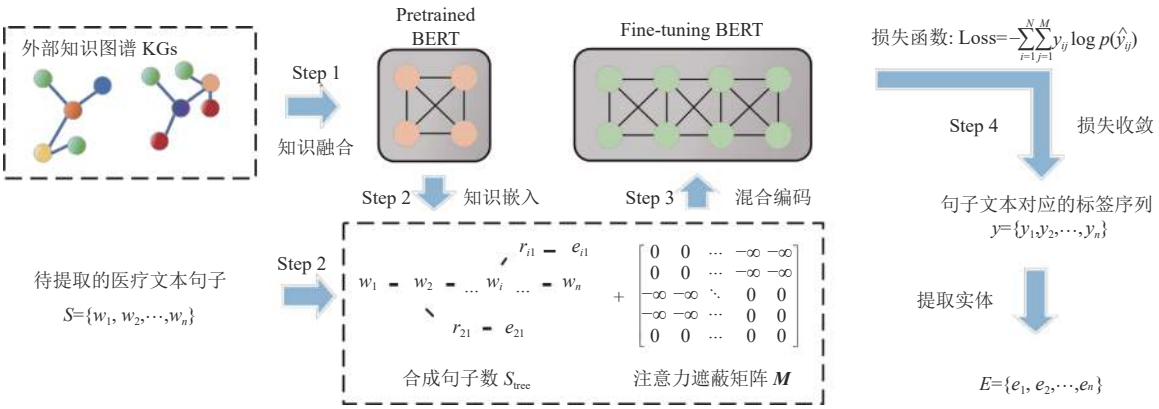

图 2 模型架构图

3.2 外部知识图融合

外部知识图融合，即将多个知识图谱通过融合手段转化为单个知识图，使相似实体间能够形成联系，从而在知识嵌入时可以让句子获得来自相似实体的更多知识。知识图融合流程如图 3 所示。

首先将所有的领域知识图汇集到一个字典中，假设字典中有着大量的语义上非常相似的实体，如“帕金森综合征、帕金森病、震颤麻痹，老年痴呆、

阿尔茨海默病等”。本文通过将图的信息按三元组分段输入到模型中的 BERT 层训练获得词向量，再将对应词向量保存到字典中，然后转化为聚类任务，通过计算实体之间的语义相似度作为判定实体间相似度的依据，最后根据设定的阈值将超出阈值的实体归为同一实体，使两个实体的属性共享，达到知识图谱融合的目的。

由于要计算实体间的语义相似度，而 BERT 模型生成的句子嵌入向量中的每个词语都包含了丰富的上下文和训练语料中的背景语义，并且已经有研究人员在研究文本相似和同义词匹配的任务中通过 BERT 模型生成的句嵌入向量取得了不错的效果。同时，为了保证本文模型的尽量简化，也采用 BERT 模型获取词向量。

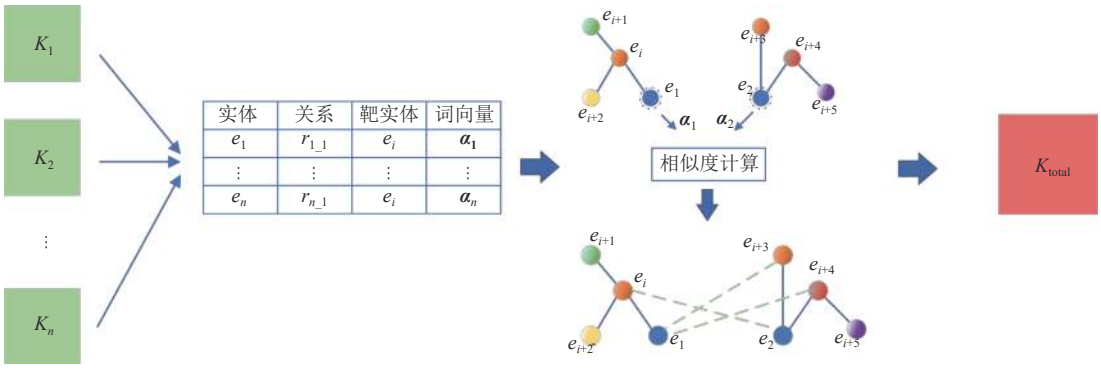

图 3　知识图谱融合流程图

BERT 词向量的生成过程如图 4 所示，首先对模型中的 BERT 模块进行词向量训练，虽然 Google 预训练好的 BERT-Chinese 模型已经有了基于维基百科知识的词向量信息，但对于特定领域的专业词汇词向量仍需要通过领域文本来进行训练获得。当文本分词后转化为 BERT 嵌入输入到 BERT 模型中进行词向量计算。受文献 [6] 启发，取最后 4 层的隐藏层拼接求和来获得输入句子的句向量，再按字分为单独的字向量，最后根据词的位置和跨度信息组合成相应的词向量，如图 5 所示。

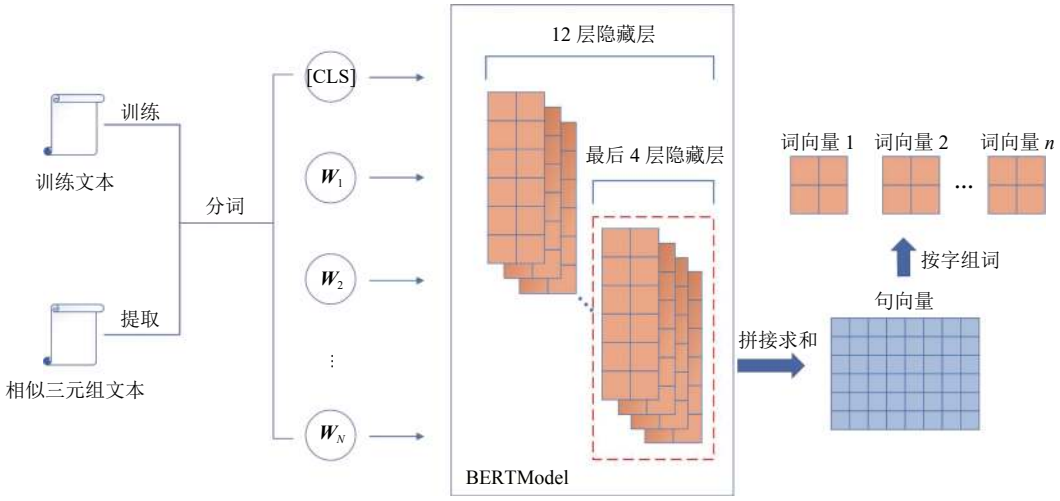

图 4　词向量生成过程

对每个词向量进行归一化处理，将其置于同一个向量空间中进行比较。由于每个词的词向量都包含着各自的语义背景，因此词向量间的近似程度可以用两个词的夹角θ来表示，通过计算夹角余弦的值来获取词语之间的相似程度，最终根据设定的阈值对满足条件的相似实体进行属性共享，实现相似知识的融合。实体余弦相似度的计算公式为：

$$\cos(\theta)=\frac{\boldsymbol{W}_j\boldsymbol{W}_k}{|\boldsymbol{W}_j||\boldsymbol{W}_k|}=\frac{\sum_{i=1}^{n}(x_i y_i)}{\sqrt{\sum_{i=1}^{n}(x_i)^2}\sqrt{\sum_{i=1}^{n}(y_i)^2}}$$

$$\boldsymbol{W}_j=(x_1,x_2,\cdots,x_n),\boldsymbol{W}_k=(y_1,y_2,\cdots,y_n) \tag{1}$$

式中，θ是向量$\boldsymbol{W}_j$、$\boldsymbol{W}_k$的夹角；$\boldsymbol{W}_j$、$\boldsymbol{W}_k$分别为两个实体的词向量。

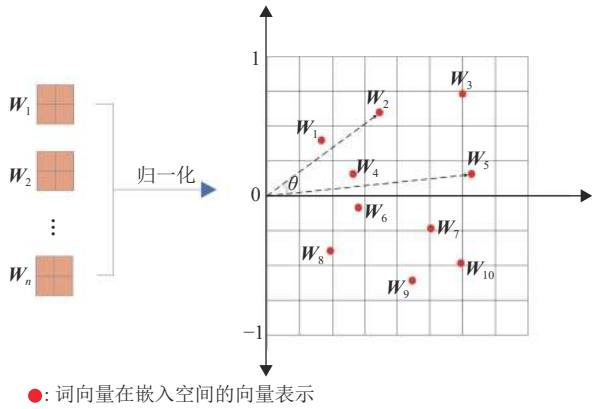

图 5 基于词向量的实体语义相似度比较

3.3 知识嵌入

在外部知识如图融合后，就需要对句子进行知识嵌入。具体地说，给定一个文本，经过分词器分词后获得句子$s=[w_0,w_1,w_2,\cdots,w_n]$和一个融合知识图 K，经过知识层输出一个句子树$t=\{w_0,w_1,\cdots,w_i[(r_{i0},w_{i0}),\cdots,(r_{ik},w_{ik})],\cdots,w_n\}$。这个过程可以分为知识查询和知识链接两个步骤。在知识查询中，句子s中的所有实体名称被依次选中，从融合知识图 K 中检索它们相对应的三元组关键词。获取到关键词后可以构建出对应的候选链接三元组集$E=[(w_i,r_{i0},w_{i0}),\cdots,(w_k,r_{ik},w_{ik})]$。知识链接过程需要通过知识融合生成的词典进行关键词匹配，对匹配到的关键词进行实体链接。知识链接过程将E中的三元组链接到它们相应的位置，将外部三元组知识注入到句子s中，并生成一个句子树t。句子中每个有知识注入的实体作为根节点且被嵌入的三元组实体不能再作为根节点扩展其他的分支，而其他词作为叶子结点，注入知识的实体间关系作为它们的边。具体结构可写作：

$$\begin{aligned}s_{\text{tree}}=\{&w_1,w_2[(r_{21},e_{21}),(r_{22},e_{22})],\\&\cdots,w_i[(r_{i1},e_{i2}),\cdots,(r_{im},e_{im})],\cdots,w_n\}\end{aligned}$$

如：2 型糖尿病，旧称非胰岛素依赖型糖尿病，是一种慢性代谢疾病，多在 35～40 岁之后发病。常见症状有烦渴、频尿、不明原因的体重减轻，常通过注射胰岛素治疗。转化为句子树之后变为：{2 型糖尿病 [(类型,疾病),(简称,糖尿病)(治疗手段,注射胰岛素)]，……。常,见,症,状,有,烦渴 [(类型,症状)]、频尿 [(类型,症状)]、不,明,原,因,的,体重减轻 [(类型,症状)]，常通过注射胰岛素 [(类型,治疗手段),(主治,糖尿病)],治,疗。}这样就使得输入的句子获得了除本身语义以外的知识，丰富了输入的信息，使得 BERT 模型能够学习到更多的知识。

3.4 混合编码

混合编码的功能是将知识嵌入部分生成的句子树转换成模型可计算的嵌入表示，为了使该嵌入表示可以输入到 BERT 模型中，本文采用与 K-BERT[15] 相同的句子树嵌入方法，可公式化为：

$$\text{Model\_Emb}=\text{TreeToEmb}(s_{\text{tree}}) \tag{2}$$

其主要将句子树转化为 BERT 可计算的输入结构，分 3 个部分：token 嵌入、位置嵌入和段落嵌入，由于嵌入层的输入是句子树，而不是 BERT 传统的标记序列。因此，为了将句子树转化为序列，并且保留句子的结构信息，位置嵌入采用软位置编码，并以此构建注意力遮蔽矩阵来防止 BERT 模型因嵌入的知识而产生错误的注意力。

如图 6 所示，句子树转 BERT 嵌入编码时，在 token 嵌入部分，为了能输入 BERT 模型，句子按顺序生成输入序列，在分支结点部分分支上的知识按分支顺序依次排列在结点内容之后，但句子树无法保证原本的语义结构，因此在位置嵌入部分还需要引入软位置嵌入。软位置嵌入如图 6 的句子树编号所示，每一段分支接续结点的编号继续编号，这样可以保证每一段分支的语义信息不丢失，但这样的位置编码直接输入到 BERT 模型中会造成模型内部的混乱，Transformer 层的多头注意力机制会出现剧烈的抖动计算，而注意力遮蔽矩阵的引入则很好地解决了这个问题。

如图 6 的遮蔽矩阵所示，注意力遮蔽矩阵根据句子树的位置编码生成，使得每一段 token 的嵌入只受到来自句子树中同一分支的上下文影响，而屏蔽了不同分支之间的知识。遮蔽矩阵的生成公式为：

$$M_{ij}=\begin{cases}0 & \text{当}w_i\text{和}w_j\text{处在同一分支上时}\\ -\infty & \text{当}w_i\text{和}w_j\text{处在不同分支上时}\end{cases} \tag{3}$$

式中，i 和 j 是w_i和w_j在句子的硬位置编码结果；M_{ij}是遮蔽矩阵的元素。

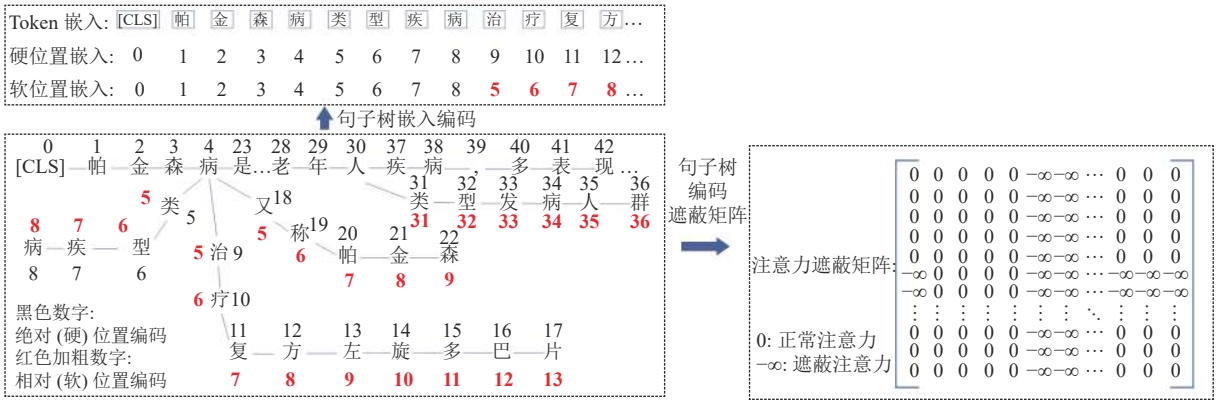

图 6　将句子树转换成嵌入表示和注意力遮蔽矩阵的过程

3.5　BERT 层

在对包含外部知识的句子进行嵌入编码并生成注意力遮蔽矩阵后，将其输入到进行 NER 的 BERT 模型中。而如 3.4 节所述，若不对 BERT 的 Transformer 层的注意力加以控制将会造成模型训练的混乱，因此参考 KBERT[15] 对 BERT-Chinese 模型的 Transformer 的注意力计算模块，嵌入注意力遮蔽矩阵，具体公式为：

$$\text{Attention}(\boldsymbol{Q}',\boldsymbol{K}',\boldsymbol{V}') = \text{Softmax}\left(\frac{\boldsymbol{Q}'\boldsymbol{K}'^{\mathrm{T}} + \boldsymbol{M}}{\sqrt{d_k}}\right)\boldsymbol{V}' \tag{4}$$

式中，$\boldsymbol{Q}',\boldsymbol{K}',\boldsymbol{V}'$是自注意力计算中的 Query、Key、Value 参数；$\boldsymbol{M}$ 是注意力遮蔽矩阵；$\sqrt{d_k}$是比例因子，用以抵消计算中的点积大幅增长的影响。添加注意力遮蔽矩阵后，如式 (3) 所示，如果两个 token 属于句子树的同一分支，则$M_{ij}=0$，即注意力计算的结果不受影响。而当两个 token 不属于同一分支，对应注意力遮蔽矩阵中的值为负无穷，这使得通过函数获得的注意值接近 0，这样就避免了不同分支间产生的错误的注意力。

模型总体的损失函数为：

$$L = -\sum_{i=1}^{N}\sum_{j=1}^{N} y_{ij}\log p(\hat{y}_{ij}) \tag{5}$$

式中，N 表示句子的长度；y_{ij}是实际识别标签的概率分布；$\hat{y}_{ij}$是模型输出的识别标签的概率分布。

4　实　验

4.1　实验设定

本节采用 Msra-NER 和 Medical-NER 两个数据集来评估模型在开放域和医疗领域 NER 的效果。MSRA-NER 是微软发布的 NER 数据集。这个任务是识别文本中的实体名称，包括人名、地名、机构名等。而 Medical-NER 是 CCKS 201 716 中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疗相关的实体名称。在外部知识库方面，选择 CN-Dbpedia、中文词语语义图谱 HowNet、疾病百科图谱 Medical。

CN-DBpedia 是复旦大学知识工作实验室开发的大规模开放域百科 KG，覆盖数千万实体、数亿关系。本文所使用的是精炼版的 CN-DBpedia，即去除了实体名称长度小于 2 或包含特殊字符的三元组，总共包含 517 万个三元组。

HowNet 是一个大规模的汉语词汇和概念的语言知识库，其中每个汉语词汇都用语义类型注释。与 K-BERT 一样，把 {word, contain, sememes} 作为知识三元组，为句子中的中文分词补充语义知识。本文使用的是精炼的 HowNet，共包含 52 576 个三元组。

Medical-NER 是由文献 [15] 提供的疾病知识图谱，包括了疾病名称、症状以及医疗部位等信息，共 13 864 条三元组。

本文的基线模型选择了基于单知识图嵌入的 K-BERT 和中文预训练模型 BERT-Chinese，并且在模型设置上采用 transformer 层 L=12、多头注意力为 12、嵌入向量的隐藏维数 H=768。

4.2　实验指标

通过召回率、准确率、$F1$ 评分对命名实体标注进行评价。

召回率 (R) 表示正确抽取的实体在实际实体中的比例：

$$R = \frac{\text{TP}}{\text{TP}+\text{FN}}$$

式中，TP 为 true positive，即实体标签预测结果与实际实体一致的样本；FN 为 false negative，即未

预测出的实体标签的样本。

准确率 (P) 表示正确抽取的实体在所有抽取实体中的比例：

$$P = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}}$$

式中，FP 为 false positive，即错误预测的实体标签样本。

$F1$ 分数 ($F1$) 是一种综合衡量命名实体识别结果召回率和准确性的指标：

$$F1 = \frac{2RP}{R+P}$$

4.3 实验结果

分别在开放域数据集 Msra 和特定医疗领域数据集上通过与基线模型 K-BERT 和 BERT 在两个数据集上 NER 任务的性能比较，评估了本文知识融合模块对开放域中文 NER 以及中文医疗 NER 的改进效果，最后通过比较模型间的收敛速度评估知识融合加速模型训练的效果。上述两个数据集都分为 3 个部分：训练、开发和测试。使用训练部分来微调模型，然后在开发和测试部分评估其性能。实验结果如表 1 和表 2所示。

表 1 模型在 MSRA 数据集的 NER 实验结果

模型	知识图谱	Msra-Ner					
		开发集			测试集		
		P	R	$F1$	P	R	$F1$
BERT-Chinese	(With out knowledge)	0.938	0.950	0.945	0.936	0.943	0.936
K-BERT	HowNet	0.958	0.954	0.958	0.951	0.956	0.945
	Cn-Dbpedia	0.961	0.960	0.963	0.953	0.956	0.957
(本文)BERT-FKG	Cn-Dbpedia + HowNet(with knowledge fusion)	**0.971**	**0.965**	**0.968**	**0.958**	**0.961**	**0.963**

表 2 模型在中文医疗数据集的 NER 实验结果

模型	知识图谱	Medical-Ner					
		开发集			测试集		
		P	R	$F1$	P	R	$F1$
BERT-Chinese	(With out knowledge)	0.919	0.931	0.925	0.919	0.931	0.925
K-BERT	Cn-Dbpedia	0.937	0.941	0.939	0.939	0.938	0.938
	Medical	0.939	0.942	0.941	0.940	0.944	0.942
(本文)BERT-FKG	Cn-Dbpedia + Medical(with knowledge fusion)	**0.945**	**0.947**	**0.945**	**0.950**	**0.947**	**0.950**

从表 1 的实验结果可以看到知识图谱的嵌入能够有效提升 BERT 模型的中文 NER 性能，而本文模型在融合了 HowNet 和 Cn-Dbpedia 的知识后，在 MSRA 数据集上的 NER 性能相较于 K-BERT 取得了一定的提升，证明了两个知识图谱之间的相似实体的属性共享能够为句子提供更丰富的语义嵌入，从而使得 BERT 模型中的注意力能够学习到更多的语义知识，提高对文本的理解能力。

在特定医疗领域数据集的实验结果如表 2 所示，从结果上可以看出知识融合在特定领域的中文 NER 取得了更明显的性能提升，融合了 Cn-Dbpedia 和 Medical 两个知识图谱中的相似实体的多知识嵌入相比于单个知识图谱的嵌入为句子中的医疗实体提供了更多的医疗语义知识，既能嵌入疾病的类型、别名信息还能嵌入疾病的治疗方式、易感人群等背景信息，在进行医疗命名实体识别时，能够使 BERT 模型更有效地识别医疗实体边界。

如图 7 所示，在医疗领域做中文 NER 时，多知识图的融合可以在一定程度上加速模型的收敛，并且在相同的训练步数下取得更高的精度，证明了多知识图的嵌入在进行知识抽取任务时能够为模型提供先验知识，使得模型能够更快地完成对已有知识的吸纳，进而减少重复提取已有知识的耗费。

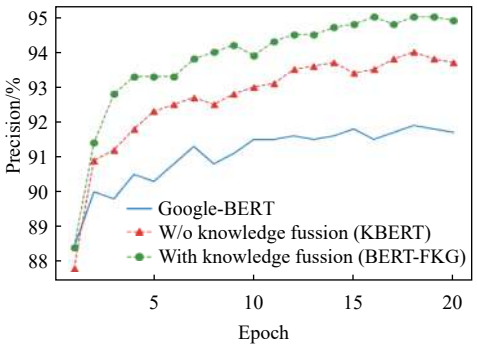

图 7 模型性能对比

5 结　论

本文提出了一种知识融合方法来对 K-BERT 的嵌入单个知识图进行改进，实现多个领域知识图的知识融合和已有领域相似实体的属性共享。BERT-FKG 将多个知识图通过知识融合对关联知识进行聚合，提高知识嵌入的效率。在医疗领域场景下的实验结果表明，在医疗领域进行 NER 任务时，知识的融合嵌入相较嵌入单个的知识图任务的准确率更高，证明在进行特定领域的 NER 任务时，利用相关领域内已有的知识图谱进行融合嵌入能提升模型识别性能，同时在开放域数据集的实验也证明了该方法的有效性。

针对目前的不足，未来可以在以下两个方面进行进一步研究：1) 知识融合方法的改进，可以考虑基于图网络的方法直接对知识图谱进行图级融合，以期待更好的知识融合效果；2) 将模型拓展到知识图谱构建应用中，通过知识融合利用模型构建过程抽取出来的三元组更新知识库，理论上可以加速领域知识图谱构建的领域化并且提升图谱质量。

参　考　文　献

[1] STRUBELL E, VERGA P, BELANGER D, et al. Fast and accurate entity recognition with iterated dilated convolutions[EB/OL]. [2021-10-11]. https://arxiv.org/pdf/1702.02098.pdf.

[2] 孙悦清. 基于循环神经网络 RNN 的领域命名实体识别方法研究[D]. 武汉: 武汉理工大学, 2018.
SUN Y Q. Research on domain named entity recognition method based on RNN of recurrent neural network[D]. Wuhan: Wuhan University of Technology, 2018.

[3] LIU Z, YANG M, WANG X, et al. Entity recognition from clinical texts via recurrent neural network[J]. BMC Medical Informatics and Decision Making, 2017, 17(2): 53-61.

[4] MIKOLOV T, CHEN K, CORRADO G, et al. Efficient estimation of word representations in vector space[EB/OL]. [2021-10-19]. https://arxiv.org/pdf/1301.3781.pdf.

[5] PENNINGTON J, SOCHER R, MANNING C D. Glove: Global vectors for word representation[C]//Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). Stroudsburg: ACL, 2014: 1532-1543.

[6] DEVLIN J, CHANG M W, LEE K, et al. BERT: Pre-Training of deep bidirectional transformers for language understanding[EB/OL]. [2021-11-11]. https://arxiv.org/pdf/1810.04805.pdf.

[7] ZHANG Y, YANG J. Chinese NER using lattice LSTM[EB/OL]. [2021-11-15]. https://arxiv.org/pdf/1805.02023.pdf.

[8] LI X, YAN H, QIU X, et al. FLAT: Chinese NER using flat-lattice transformer[EB/OL]. [2021-11-20]. https://www.xueshufan.com/publication/3019125528.

[9] ZHU W, CHEUNG D. LEX-BERT: Enhancing bert based ner with lexicons[EB/OL]. [2021-11-26]. https://www.xueshufan.com/publication/3118353673.

[10] SUN Y, WANG S, LI Y, et al. ERNIE: Enhanced representation through knowledge integration[EB/OL]. [2021-11-30]. https://www.xueshufan.com/publication/2938830017.

[11] SUN Y, WANG S, LI Y, et al. ERNIE 2.0: A continual pre-training framework for language understanding [C]//Proceedings of the AAAI Conference on Artificial Intelligence. [S.l.]: AAAI, 2020, 34(5): 8968-8975.

[12] YAMADA I, ASAI A, SHINDO H, et al. LUKE: Deep contextualized entity representations with entity-aware self-attention[EB/OL]. [2021-12-11]. https://www.xueshufan.com/publication/3090325631.

[13] BOSSELUT A, RASHKIN H, SAP M, et al. COMET: Commonsense transformers for automatic knowledge graph construction[EB/OL]. [2021-12-18]. https://www.xueshufan.com/publication/2999524812.

[14] RADFORD A, NARASIMHAN K, SALIMANS T, et al. Improving language understanding by generative pre-training[EB/OL]. [2021-12-21]. https://s3-us-west-2.amazonaws.com/openai-assets/research-covers/language-unsupervised/language_understanding_paper.pdf.

[15] 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. [S.l.]: AAAI, 2020, 34(3): 2901-2908.

[16] WANG X, GAO T, ZHU Z, et al. KEPLER: A unified model for knowledge embedding and pre-trained language representation[J]. Transactions of the Association for Computational Linguistics, 2021, 9: 176-194.

[17] SCHARFFE F, LIU Y, ZHOU C. Rdf-AI: An architecture for rdf datasets matching, fusion and interlink[C]//IJCAI 2009 Workshop on Identity, Reference, and Knowledge Representation (IR-KR). Pasadena: [s.n.], 2009: 23.

[18] NGOMO A C N, AUER S. Limes-A time-efficient approach for large-scale link discovery on the web of data[EB/OL]. [2021-12-25]. https://www.ijcai.org/Proceedings/11/Papers/385.pdf.

[19] PERSHINA M, YAKOUT M, CHAKRABARTI K. Holistic entity matching across knowledge graphs[C]//2015 IEEE International Conference on Big Data. [S.l.]: IEEE, 2015: 1585-1590.

[20] BORDES A, USUNIER N, GARCIA-DURAN A, et al. Translating embeddings for modeling multi-relational data[J]. Advances in Neural Information Processing Systems, 2013, 26: 288-297.

[21] ZHANG Y, LIU L, FU S, et al. Entity alignment across knowledge graphs based on representative relations selection[C]//2018 5th International Conference on Systems and Informatics. [S.l.]: IEEE, 2018: 1056-1061.

编　辑　叶　芳