·人工智能与模式识别·　　文章编号：1000-3428(2022)03-0107-08　　文献标志码：A　　中图分类号：TP391

融合词典与对抗迁移的越南语事件实体识别

薛振宇[1,2]，线岩团[1,2]，余正涛[1,2]，高盛祥[1,2]，普浏清[1,2]

（1.昆明理工大学 信息工程与自动化学院，昆明 650500；2.昆明理工大学 云南省人工智能重点实验室，昆明 650500）

摘　要：针对越南语事件标注语料稀缺且标注语料中未登陆词过多导致实体识别精度降低的问题，提出一种融合词典与对抗迁移的实体识别模型。将越南语作为目标语言，英语和汉语作为源语言，通过源语言的实体标注信息和双语词典提升目标语言的实体识别效果。采用词级别对抗迁移实现源语言与目标语言的语义空间共享，融合双语词典进行多粒度特征嵌入以丰富目标语言词的语义表征，再使用句子级别对抗迁移提取与语言无关的序列特征，最终通过条件随机场推理模块标注实体识别结果。在越南语新闻数据集上的实验结果表明，在源语言为英语和汉语的情况下，该模型相比主流的单语实体识别模型和迁移学习模型的实体识别性能有明显提升，并且在加入目标语义标注数据后，相比单语实体识别模型的F1值分别增加了19.61和18.73个百分点。

关键词：实体识别；对抗迁移；双语词典；多粒度特征；序列特征

开放科学（资源服务）标志码（OSID）：

中文引用格式：薛振宇，线岩团，余正涛，等.融合词典与对抗迁移的越南语事件实体识别[J].计算机工程，2022，48(3)：107-114，145.

英文引用格式：XUE Z Y，XIAN Y T，YU Z T，et al. Vietnamese event entity recognition combining dictionary and adversarial transfer[J].Computer Engineering，2022，48(3)：107-114，145.

Vietnamese Event Entity Recognition Combining Dictionary and Adversarial Transfer

XUE Zhenyu[1,2]，XIAN Yantuan[1,2]，YU Zhengtao[1,2]，GAO Shengxiang[1,2]，PU Liuqing[1,2]

（1.Faculty of Information Engineering and Automation，Kunming University of Science and Technology，Kunming 650500，China；
2.Yunnan Key Laboratory of Artificial Intelligence，Kunming University of Science and Technology，Kunming 650500，China）

【Abstract】The problem of the scarcity of Vietnamese event annotated corpus，comprising several unregistered words，reduces the accuracy of entity recognition.This study proposes an entity recognition model that combines a dictionary and adversarial transfer.It uses Vietnamese as the target language，and English and Chinese as the source languages.Furthermore，the entity tagging information of the source language and bilingual dictionary are used to improve the entity recognition of the target language.The semantic space is shared between the source and target languages by word-level adversarial transfer.Moreover，multi-granular features are embedded into bilingual dictionary to enrich the semantic representation of target language words，and sentence-level adversarial transfer is used to extract language-independent sequence features.Finally，the entity recognition result is marked by a Conditional Random Field（CRF）inference module.The experimental results on the Vietnamese news dataset demonstrate that the proposed model has improved entity recognition compared to the mainstream monolingual entity recognition model and transfer learning model when the source languages are English and Chinese.After adding the target semantic annotation data，the F1-score of the monolingual entity recognition transfer learning model when the source languages are English and Chinese increased by 19.61% and 18.73%，respectively.

【Key words】entity recognition；adversarial transfer；bilingual dictionary；multi-granular feature；sequence feature

DOI：10. 19678/j. issn. 1000-3428. 0060466

0　概述

越南语事件实体识别主要包括对越南语新闻文本中人名、地名、组织机构名、特定政治概念名等实体类型标签的自动识别，是越南语新闻信息检索、自动问答、机器翻译等任务的重要基础。目前，多数事件实体识别系统采用基于双向长短时记忆（Bidirectional Long Short-Term Memory，BiLSTM）网络和条件随机场

基金项目：国家自然科学基金（61972186，61762056，61472168）；云南省重大科技专项计划（202002AD080001）；云南省高新技术产业专项（201606）。

作者简介：薛振宇（1996—），男，硕士研究生，主研方向为自然语言处理、跨语言信息检索；线岩团，副教授、硕士；余正涛，教授、博士；高盛祥，副教授、博士；普浏清，硕士研究生。

收稿日期：2021-01-04　　**修回日期：**2021-03-01　　**E-mail：**gaoshengxiang.yn@foxmail.com

(Conditional Random Field,CRF)的组合模型BiLSTM-CRF[1]进行实体识别。该模型在高资源语言事件实体识别任务上具有较好的性能,在高资源语言分别为英语和汉语的情况下,使用单语事件标注语料进行训练,所得F1值为91.23和90.78,并且在越南语公共数据集VLSP2016[2]上也取得了87.33的F1值。但是,该模型在越南语事件实体识别任务上的性能较差,主要原因为相较于公共数据集VLSP2016,越南语新闻事件数据集中加入了政治概念名这一特定的事件实体类别,扩大了实体标签的搜索空间,增加了模型对于实体标签的预测难度。同时,由于越南语新闻语料较少且人工标注越南语事件实体困难,因此导致越南语事件标注语料稀缺且标注语料中未登录词过多。然而,缺少用于训练的标注语料会使得模型训练不充分,引起模型过拟合,最终降低越南语事件实体识别的F1值。

目前,一些研究人员利用基于迁移学习思想的多任务学习、词级对抗实现双语词嵌入表示、双语词典实现双语词嵌入表示、两层对抗迁移等模型来提升越南语事件实体识别效果。多任务学习模型[3-4]是所有任务共享一个编码层,通过共享编码层进行知识迁移,但是由于不同语言的序列结构不同,当同时编码两种不同资源的语言信息时,编码器不能保证提取到与语言无关的序列信息从而对高资源语言的标注信息进行较好的迁移。词级对抗实现双语词嵌入表示模型[5-8]仅对两种语言的预训练词向量进行对抗训练以将两种语言映射到同一语义空间中,忽略了两种语言的序列特征信息,无法充分地利用源语言的序列特征辅助目标语言进行实体识别。双语词典实现双语词嵌入表示模型[9-11]使用大规模双语词典对齐源语言与目标语言的词向量空间,从而将源语言标注信息迁移至目标语言空间上,但人工构造大规模双语词典相对困难且该模型未考虑双语翻译的一词多义问题。两层对抗迁移模型[12]基于BiLSTM-CRF网络,使用词级对抗迁移将两种语言融入同一语义空间,利用了句子级对抗迁移提取与语言无关的序列特征,但是目标语言词语义表征单一且提取与语言无关的序列特征效果较差。

为更好地将源语言序列信息迁移到目标语言语义空间中,进而利用源语言序列特征辅助目标语言进行实体识别。本文针对上述多任务学习模型和词级对抗实现双语词嵌入表示模型中存在无法提取与语言无关的序列特征问题,以及两层对抗迁移模型中存在与语言无关的序列特征提取效果较差的问题,使用融合多头注意力的句子级对抗迁移方式,句子级鉴别器用来区分目标语言语义空间中句子的真实来源,即判断句子是否来源于源语言句子或目标语言句子,使用多头注意力特征共享编码器混淆句子级鉴别器,从而提取到与语言无关的序列特征,实现将两种语言的序列信息映射到同一目标语言语义空间中。

1 模型结构

为有效利用源语言的已标注信息提升目标语言的实体识别效果,本文提出融合词典与对抗迁移的越南语事件实体识别模型。使用词级对抗迁移方法将源语言预训练词向量线性映射到目标语言语义空间中,词级鉴别器用来区分目标语言语义空间中词的真实来源,即判断词是否来源于线性映射前的源语言词或真实的目标语言词,线性映射层与词级鉴别器相互对抗混淆以使得线性映射层不断优化,从而实现将两种语言的词级信息映射到同一目标语言语义空间中。

对于目标语言句子而言,本文模型针对两层对抗迁移模型中存在的目标语言词语义表征单一的问题,融入目标语言字符级特征,并且引入小规模双语词典中词义互为补充的源语言互译词的词级特征,使目标语言词得到更丰富的语义表征。不同语言对的同一个词往往有不同的解释,例如:越南语词“thợ rèn”的中文解释是“铁匠”,该词在越南语中通常不是作为一个人名出现的,但是根据英越词典,该越南语词的一个英文解释是“smith”,而该英文解释在英语中常常是作为人名出现的;越南语词“Phật sơn”的中文解释是“佛画”,该词在越南语中通常不是作为一个地名出现的,但是根据汉越词典,该越南语词的一种汉语解释是“佛山”,而该汉语解释在汉语中常常是作为具体地名出现的。由于双语词典中不一定包含所有的目标语言词及其对应的源语言词构成的互译词对,并且会存在少量的目标语言未登录词,因此为解决这一问题,本文将未登录的目标语言词通过线性映射转化到源语言空间上,利用转化后的目标语言词表征作为源语言词的语义表征融入模型中。对于源语言句子而言,将源语言词通过使用词级对抗迁移方法优化后的线性映射层转化到目标语言空间后,融入源语言字符级特征,由于转化后的源语言词可能失去源语言词本身的语义信息,因此将转化前源语言词的语义表征融入模型中以补充该词缺失的语义信息。

本文提出的融合词典与对抗迁移的越南语事件实体识别模型结构如图1所示,该模型主要由词级对抗迁移模块、融合双语词典的多粒度特征嵌入模块、句子级对抗迁移模块、CRF推理模块等4个部分组成。首先,在词级对抗迁移训练过程中令线性映射层与词级鉴别器相互对抗混淆以使得线性映射层不断优化。然后,提取并融合目标语言句子中的目标语言词级特征、目标语言字符级特征与通过双语词典找到的对应源语言词级特征,以及源语言句子中的源语言词级特征、源语言字符级特征与该句子通过优化后的线性映射层后的源语言词级特征。最后,在句子级对抗迁移训练过程中,将多头注意力特征共享编码器与句子级鉴别器相互对抗混淆,不断优化共享编码器,从而使得多头注意力特征共享编码器提取到与语言无关的序列特征信息。将与语言无关的序列特征输入多头注意力上下文编码器中提取全局信息,衡量每个词在整个句子中的重要性程度,进而通过CRF对整个句子的输出进行联合建模。

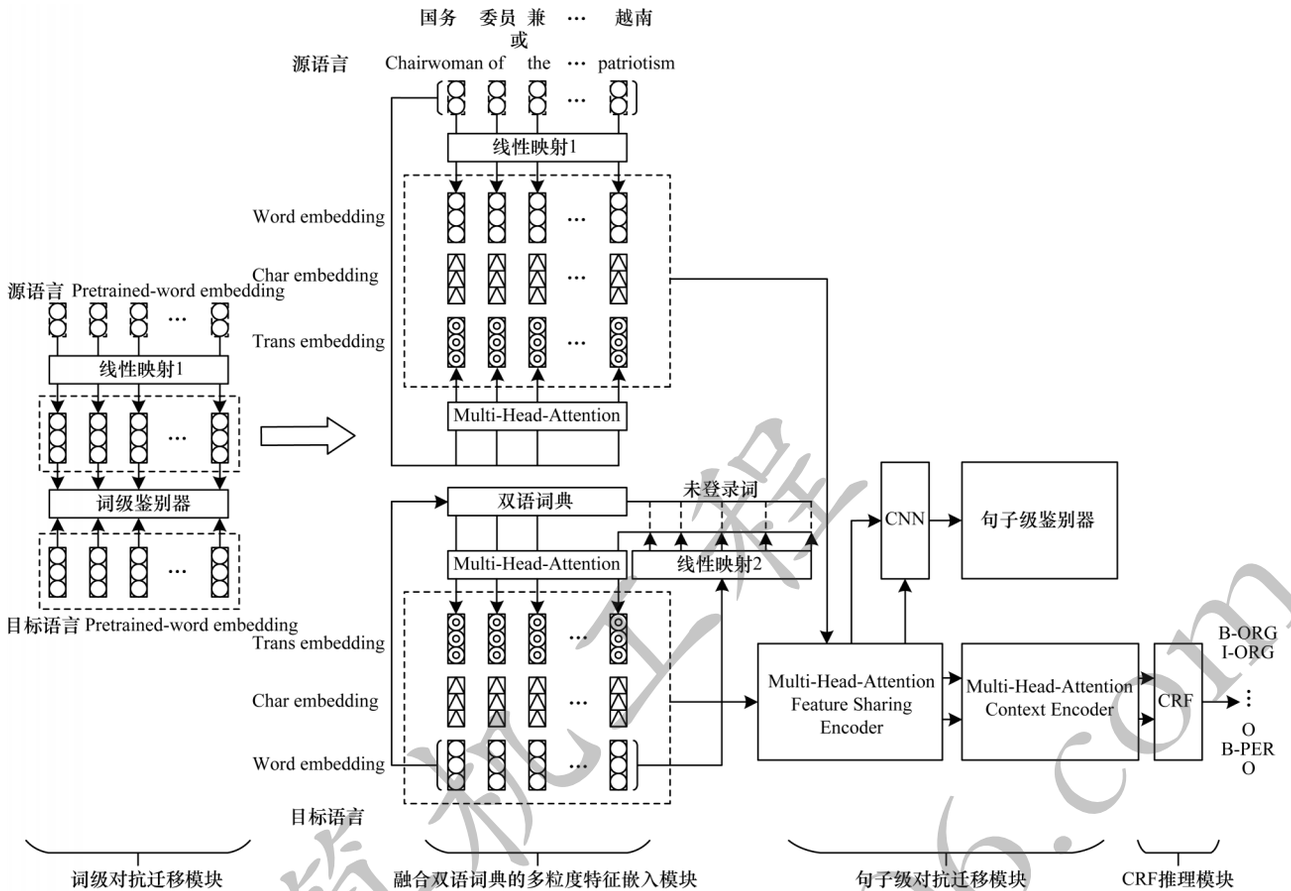

图1　融合词典与对抗迁移的越南语事件实体识别模型结构

Fig.1　Structure of Vietnamese event entity recognition model combining dictionary and adversarial transfer

2　融合词典与对抗迁移的事件实体识别

2.1　词级对抗迁移模块

为更好地利用源语言的标注信息,这一模块利用词级对抗迁移的方式将源语言与目标语言语义上对齐的词映射在同一语义空间中。该训练过程是无监督的训练过程,在参考ZHANG等[13-14]利用无监督的方式学习双语词向量的工作基础上,本文使用词级对抗训练的方式来自动对齐源语言与目标语言的词表示。在得到预训练好的目标语言词向量 $\boldsymbol{V}_{\mathrm{t}}=\{\boldsymbol{v}_1^{\mathrm{t}},\boldsymbol{v}_2^{\mathrm{t}},\cdots,\boldsymbol{v}_N^{\mathrm{t}}\}\in\mathbb{R}^{N\times d_{\mathrm{t}}}$($\boldsymbol{v}_i^{\mathrm{t}}$是目标语言词$w_i^{\mathrm{t}}$的向量表示,$N$是词向量所含词的数目,$d_{\mathrm{t}}$是目标语言词向量的维度大小)与预训练好的源语言词向量 $\boldsymbol{V}_{\mathrm{s}}=\{\boldsymbol{v}_1^{\mathrm{s}},\boldsymbol{v}_2^{\mathrm{s}},\cdots,\boldsymbol{v}_M^{\mathrm{s}}\}\in\mathbb{R}^{M\times d_{\mathrm{s}}}$($\boldsymbol{v}_j^{\mathrm{s}}$是源语言词$w_j^{\mathrm{s}}$的向量表示,$M$是词向量所含词的数目,$d_{\mathrm{s}}$是源语言词向量的维度大小)的情况下,使用一个线性映射函数f将源语言映射到目标语言空间中:

$$\tilde{\boldsymbol{V}}_{\mathrm{s}}=f(\boldsymbol{V}_{\mathrm{s}})=\boldsymbol{V}_{\mathrm{s}}\boldsymbol{U} \tag{1}$$

其中:$\boldsymbol{U}\in\mathbb{R}^{d_{\mathrm{s}}\times d_{\mathrm{t}}}$是转换矩阵;$\tilde{\boldsymbol{V}}_{\mathrm{s}}\in\mathbb{R}^{M\times d_{\mathrm{t}}}$是映射后的源语言词向量。为对$\tilde{\boldsymbol{V}}_{\mathrm{s}}$进行归一化操作,使用奇异值分解的方法[15]限定转换矩阵$\boldsymbol{U}\in\mathbb{R}^{d_{\mathrm{s}}\times d_{\mathrm{t}}}$为正交矩阵:

$$\boldsymbol{U}=\boldsymbol{A}\boldsymbol{B}^{\mathrm{T}},\boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{B}^{\mathrm{T}}=\mathrm{SVD}(\tilde{\boldsymbol{V}}_{\mathrm{s}}\boldsymbol{V}_{\mathrm{s}}^{\mathrm{T}}) \tag{2}$$

为自动优化映射函数f,使用一个多层感知器D作为词级鉴别器。将映射后的源语言词向量与目标语言词向量输入到鉴别器中,每一个词向量对应的输出是一个单纯的数值。通过最小化如式(3)所示的交叉熵损失函数来训练词级鉴别器:

$$\begin{aligned}L_{\mathrm{dis}}^{w}=&-\frac{1}{I_{\mathrm{t;\,s}}}\cdot\sum_{i=0}^{I_{\mathrm{t;\,s}}}(y_i\cdot\log_a(D(w_i^*))+\\&(1-y_i)\cdot\log_a(1-D(w_i^*)))\\y_i=&\,\delta_i(1-2\varepsilon)+\varepsilon\end{aligned} \tag{3}$$

其中:$D(w_i^*)$表示词w_i^*来自目标语言的概率,当w_i^*来自目标语言时,$\delta_i=1$,否则$\delta_i=0$;$I_{\mathrm{t;\,s}}$表示目标语言词与源语言词的数目之和;ε表示附加在词标签上的平滑值。

映射函数f与词级鉴别器在训练过程中互相对抗混淆对方,通过最小化如式(4)所示的交叉熵损失函数来训练映射函数f,使得映射函数f的参数趋于最优:

$$\begin{aligned}L_f^{w}=&-\frac{1}{I_{\mathrm{t;\,s}}}\cdot\sum_{i=0}^{I_{\mathrm{t;\,s}}}((1-y_i)\cdot\log_a(D(w_i^*))+\\&y_i\cdot\log_a(1-D(w_i^*)))\\y_i=&\,\delta_i(1-2\varepsilon)+\varepsilon\end{aligned} \tag{4}$$

借鉴GOODFELLOW等[16]在深度对抗神经网络训练过程中的优化策略,使用随机梯度下降法在训练过程中对线性映射函数和词级鉴别器进行优化,不断减小损失函数L_{dis}^{w}与L_f^{w}的值。参考

CONNEAU等[8]在词级对抗训练后，基于映射后的源语言空间和目标语言空间，找到k个出现频率最高的源语言词和分别与之距离相近（语义相近）的目标语言词来构建双语词典，利用双语词典进一步优化转换矩阵$\boldsymbol{U}$。但是，考虑到该方法找到的语义上能够完全相同的源语言与目标语言词对的数量较少，因此在词级对抗后，使用预先构建好的外部双语词典，词典内有语义相同的m个源语言与目标语言词对。同时，在得到上述k个源语言与目标语言词对的基础上，去除该k个词对中源语言词在词典中有对应的词对，将剩下的词对与词典中的m个词对相结合后重构一个新的双语词典，从而利用新的双语词典并以有监督的方式进一步优化转换矩阵$\boldsymbol{U}$。

2.2 融合双语词典的多粒度特征嵌入模块

在利用词级对抗迁移的方式对齐源语言与目标语言的词表示后，输入源语言与目标语言的句子表示，将源语言的句子表示通过训练好的线性映射层映射到目标语言语义空间中。此时，认为映射后的源语言句子和目标语言句子来自同一种语言，可以利用源语言的标注信息来对目标语言的句子进行标注，但是对事件实体进行标注不仅需要词级特征，而且需要字符级特征和句子内上下文特征，因此分别针对源语言与目标语言的特点提取词级特征和字符级特征。不同语言对的同一个词往往有不同的解释，为使目标语言和映射后的源语言获得更丰富的语义表示，分别利用双语词典引入目标语言词翻译后的词嵌入和直接引入映射前的源语言词嵌入的方式使得目标语言得到更多源语言的语义信息。

输入目标语言句子表示$x^{\mathrm{t}}=\{w_1^{\mathrm{t}},w_2^{\mathrm{t}},\cdots,w_Q^{\mathrm{t}}\}$与源语言句子表示$x^{\mathrm{s}}=\{w_1^{\mathrm{s}},w_2^{\mathrm{s}},\cdots,w_Y^{\mathrm{s}}\}$后，首先使用$\boldsymbol{V}_{\mathrm{t}}$和$\tilde{\boldsymbol{V}}_{\mathrm{s}}$将句子中的每一个词表示$w_i^{\mathrm{t}}$和$w_j^{\mathrm{s}}$初始化为词向量。将每一个目标语言词$w_i^{\mathrm{t}}$与源语言词$w_j^{\mathrm{s}}$分别随机初始化为字符向量$\boldsymbol{w}_i^{\mathrm{tc}}=\{\boldsymbol{c}_1^{\mathrm{t}},\boldsymbol{c}_2^{\mathrm{t}},\cdots,\boldsymbol{c}_E^{\mathrm{t}}\}$和$\boldsymbol{w}_j^{\mathrm{sc}}=\{\boldsymbol{c}_1^{\mathrm{s}},\boldsymbol{c}_2^{\mathrm{s}},\cdots,\boldsymbol{c}_R^{\mathrm{s}}\}$，然后使用CNN来提取字符向量的特征。

对于目标语言句子而言：如果目标语言句子中的一个词w_i^{t}通过双语词典能找到对应的源语言翻译词组w_i^T，w_i^T可以是由一个翻译词或多个翻译词构成。为更好地结合源语言翻译词的语义信息，需要编码所有的翻译词。使用$\boldsymbol{V}_{\mathrm{s}}$将每一个源语言翻译词初始化为词向量，将一个源语言翻译词组中包含的所有词向量的集合$\{\boldsymbol{t}_1,\boldsymbol{t}_2,\cdots,\boldsymbol{t}_i,\cdots,\boldsymbol{t}_l\}$看作$\boldsymbol{T}\in\mathbb{R}^{d\times l}$，其中，$l$表示源语言翻译词组中词的数目。考虑到源语言翻译词组中每个词的词义与原始对应的目标语言词的词义有不同的差异，为了尽可能强化与目标语言词的词义相接近的源语言翻译词的权重，在参考SUKHBAATAR等[17]在问答任务中引入基于注意力机制的工作基础上，将一个目标语言词向量$\boldsymbol{w}_i^{\mathrm{t}}\in\mathbb{R}^d$与其翻译词组向量$\boldsymbol{T}\in\mathbb{R}^{d\times l}$作为输入，通过式(5)进行计算：

$$\boldsymbol{p}=\sum_{j=1}^{l}\alpha_j\boldsymbol{t}_j \tag{5}$$

其中：$\boldsymbol{p}\in\mathbb{R}^d$；$l$为翻译词组中词的数目；$\alpha_j\in[0,1]$表示$\boldsymbol{t}_j$的权重且$\sum_j\alpha_j=1$。

使用一个线性层计算每一个$\boldsymbol{t}_j$与其对应的目标语言词向量$\boldsymbol{w}_i^{\mathrm{t}}$的语义相似程度，语义越相似，得分函数的值越大。得分函数计算如下：

$$g_j=\tanh(\boldsymbol{W}_{\mathrm{att}}\boldsymbol{w}_i^{\mathrm{t}}+\boldsymbol{U}_{\mathrm{att}}\boldsymbol{t}_j+\boldsymbol{b}_{\mathrm{att}}) \tag{6}$$

其中：$\boldsymbol{W}_{\mathrm{att}},\boldsymbol{U}_{\mathrm{att}}\in\mathbb{R}^d$；$\boldsymbol{b}_{\mathrm{att}}\in\mathbb{R}^{1\times 1}$。

将得到的$g_1,g_2,\cdots,g_l$输入softmax函数得出翻译词组中每个词的重要性分布$\alpha_1,\alpha_2,\cdots,\alpha_l$：

$$\alpha_j=\frac{\exp(g_j)}{\sum_{z=1}^{l}\exp(g_z)} \tag{7}$$

如果目标语言句子中的一个词w_i^{t}通过双语词典未能找到对应的源语言翻译词组w_i^T，则利用一个线性映射函数将目标语言句子中的词w_i^{t}转换到源语言语义空间上，将转换后得到的$\boldsymbol{p}_i$视作w_i^{t}对应的源语言翻译词向量：

$$\boldsymbol{p}_i=\boldsymbol{M}\boldsymbol{w}_i^{\mathrm{t}} \tag{8}$$

其中：$\boldsymbol{M}$表示线性映射矩阵。最小化如式(9)所示的损失函数以优化$\boldsymbol{M}$：

$$\mathrm{Loss}(\boldsymbol{M})=\sum_{i=1}^{f}\left\|\boldsymbol{p}_i-\boldsymbol{M}\boldsymbol{w}_i^{\mathrm{t}}\right\|_2 \tag{9}$$

在得到优化的$\boldsymbol{M}$后，对于每一个不在双语词典中的目标语言词o_i，可以使用式(10)计算出对应的源语言翻译词向量：

$$\boldsymbol{p}_i^{\mathrm{o}}=\boldsymbol{M}\boldsymbol{o}_i \tag{10}$$

为能从不同的表示子空间中学习句子内部源语言翻译词之间的依赖关系，捕获句子的内部结构，模型使用多头注意力机制对得到的$\boldsymbol{P}_i=\{\boldsymbol{p}_1,\boldsymbol{p}_2,\cdots,\boldsymbol{p}_i,\cdots,\boldsymbol{p}_Q\}$进行建模，将得到的多头注意力结果作为该句子最终的源语言翻译词嵌入$\boldsymbol{W}_i^{\mathrm{v}}=\{\boldsymbol{w}_1^{\mathrm{v}},\boldsymbol{w}_2^{\mathrm{v}},\cdots,\boldsymbol{w}_i^{\mathrm{v}},\cdots,\boldsymbol{w}_Q^{\mathrm{v}}\}$。

对于源语言句子而言：使用$\boldsymbol{V}_{\mathrm{s}}$将句子$x^{\mathrm{s}}=\{w_1^{\mathrm{s}},w_2^{\mathrm{s}},\cdots,w_Y^{\mathrm{s}}\}$中的每一个词表示$w_i^{\mathrm{s}}$初始化为词向量$\boldsymbol{w}_i^{\mathrm{ss}}$，使用多头注意力机制对初始化后的词向量集合$\boldsymbol{W}_i^{\mathrm{ss}}=\{\boldsymbol{w}_1^{\mathrm{ss}},\boldsymbol{w}_2^{\mathrm{ss}},\cdots,\boldsymbol{w}_i^{\mathrm{ss}},\cdots,\boldsymbol{w}_Y^{\mathrm{ss}}\}$进行建模，与上述从不同的表示子空间中学习句子内部源语言翻译词之间的依赖关系所使用的多头注意力机制建模过程一致，将得到的多头注意的结果作为该句子的映射前源语言词嵌入$\boldsymbol{W}_i^{\mathrm{vs}}=\{\boldsymbol{w}_1^{\mathrm{vs}},\boldsymbol{w}_2^{\mathrm{vs}},\cdots,\boldsymbol{w}_i^{\mathrm{vs}},\cdots,\boldsymbol{w}_Y^{\mathrm{vs}}\}$。

在得到目标语言字符嵌入、目标语言词嵌入和相应的源语言翻译词嵌入以及源语言字符嵌入、源语言词嵌入和相应的映射前源语言词嵌入后，借鉴多粒度嵌入算法[18]，分别针对源语言与目标语言的词嵌入和字符嵌入进行联合学习。但若只进行简单的词与字符向量的拼接会导致不准确的编码，则本文使用注意力机制自适应地依赖权重强化或

弱化每个粒度，更有效地学习输入的特征并丰富单词嵌入。

对于目标语言句子表示 $x^{\mathrm{t}}=\{w_1^{\mathrm{t}},w_2^{\mathrm{t}},\cdots,w_Q^{\mathrm{t}}\}$ 而言，预测其中一个词 w_i^{t} 是依据：1）该词本身，表示为 $\boldsymbol{w}_i^{\mathrm{t}}\in\mathbb{R}^d$；2）通过CNN提取到该词所包含的字符特征，表示为 $\boldsymbol{w}_i^{\mathrm{tc}}=\{\boldsymbol{c}_1^{\mathrm{t}},\boldsymbol{c}_2^{\mathrm{t}},\cdots,\boldsymbol{c}_E^{\mathrm{t}}\}\in\mathbb{R}^d$；3）该词对应的源语言翻译词级特征，表示为 $\boldsymbol{w}_i^{\mathrm{v}}\in\mathbb{R}^d$。引入注意力机制以提取对句子语义有重要作用的词，然后依据其加权重要程度在词粒度与字符粒度之间产生一个知识聚合的单一向量 $\boldsymbol{s}_i^{\mathrm{t}}$，具体计算如下：

$$[\boldsymbol{u}_i^{\mathrm{t}},\boldsymbol{u}_i^{\mathrm{tc}},\boldsymbol{u}_i^{\mathrm{v}}]=\sigma[(\boldsymbol{W}_m\cdot\boldsymbol{w}_i^{\mathrm{t}}+\boldsymbol{b}_m),(\boldsymbol{W}_m\cdot\boldsymbol{w}_i^{\mathrm{tc}}+\boldsymbol{b}_m),(\boldsymbol{W}_m\cdot\boldsymbol{w}_i^{\mathrm{v}}+\boldsymbol{b}_m)]$$

$$\alpha_i^m=\frac{\exp(\boldsymbol{u}_i^m)}{\sum_{m^*\in\{\mathrm{t,tc,v}\}}\exp(\boldsymbol{u}_i^{m^*})},\forall m\in\{\mathrm{t,tc,v}\}$$

$$\boldsymbol{s}_i^{\mathrm{t}}=\alpha_i^{\mathrm{t}}\cdot\boldsymbol{w}_i^{\mathrm{t}}+\alpha_i^{\mathrm{tc}}\cdot\boldsymbol{w}_i^{\mathrm{tc}}+\alpha_i^{\mathrm{v}}\cdot\boldsymbol{w}_i^{\mathrm{v}}\tag{11}$$

其中：$[\boldsymbol{u}_i^{\mathrm{t}},\boldsymbol{u}_i^{\mathrm{tc}},\boldsymbol{u}_i^{\mathrm{v}}]$ 为注意力向量；$\boldsymbol{W}_m$ 为平均权重；α_i^m 为注意力权重值；$\boldsymbol{b}_m$ 为偏置项；$\boldsymbol{s}_i^{\mathrm{t}}$ 为融合多粒度特征后的向量；$\boldsymbol{w}_i^{\mathrm{t}}$、$\boldsymbol{w}_i^{\mathrm{tc}}$ 和 $\boldsymbol{w}_i^{\mathrm{v}}$ 维度相同。

对于源语言句子表示 $x^{\mathrm{s}}=\{w_1^{\mathrm{s}},w_2^{\mathrm{s}},\cdots,w_Y^{\mathrm{s}}\}$ 而言，预测其中一个词 w_i^{s} 是依据：1）该词本身，表示为 $\boldsymbol{w}_i^{\mathrm{s}}\in\mathbb{R}^d$；2）通过CNN提取到该词所包含的字符特征，表示为 $\boldsymbol{w}_i^{\mathrm{sc}}=\{\boldsymbol{c}_1^{\mathrm{s}},\boldsymbol{c}_2^{\mathrm{s}},\cdots,\boldsymbol{c}_R^{\mathrm{s}}\}\in\mathbb{R}^d$；3）该词对应的映射前源语言词级特征，表示为 $\boldsymbol{w}_i^{\mathrm{vs}}\in\mathbb{R}^d$。同样依据加权重要程度产生一个知识聚合的单一向量 $\boldsymbol{s}_i^{\mathrm{s}}$，具体计算如下：

$$[\boldsymbol{u}_i^{\mathrm{s}},\boldsymbol{u}_i^{\mathrm{sc}},\boldsymbol{u}_i^{\mathrm{vs}}]=\sigma[(\boldsymbol{W}_n\cdot\boldsymbol{w}_i^{\mathrm{s}}+\boldsymbol{b}_n),(\boldsymbol{W}_n\cdot\boldsymbol{w}_i^{\mathrm{sc}}+\boldsymbol{b}_n),(\boldsymbol{W}_n\cdot\boldsymbol{w}_i^{\mathrm{vs}}+\boldsymbol{b}_n)]$$

$$\alpha_i^n=\frac{\exp(\boldsymbol{u}_i^n)}{\sum_{n^*\in\{\mathrm{s,sc,vs}\}}\exp(\boldsymbol{u}_i^{n^*})},\forall n\in\{\mathrm{s,sc,vs}\}$$

$$\boldsymbol{s}_i^{\mathrm{s}}=\alpha_i^{\mathrm{s}}\cdot\boldsymbol{w}_i^{\mathrm{s}}+\alpha_i^{\mathrm{sc}}\cdot\boldsymbol{w}_i^{\mathrm{sc}}+\alpha_i^{\mathrm{vs}}\cdot\boldsymbol{w}_i^{\mathrm{vs}}\tag{12}$$

其中：$[\boldsymbol{u}_i^{\mathrm{s}},\boldsymbol{u}_i^{\mathrm{sc}},\boldsymbol{u}_i^{\mathrm{vs}}]$ 为注意力向量；$\boldsymbol{W}_n$ 为平均权重；α_i^n 为注意力权重值；$\boldsymbol{b}_n$ 为偏置项；$\boldsymbol{s}_i^{\mathrm{s}}$ 为融合多粒度特征后的向量；$\boldsymbol{w}_i^{\mathrm{s}}$、$\boldsymbol{w}_i^{\mathrm{sc}}$ 和 $\boldsymbol{w}_i^{\mathrm{vs}}$ 维度相同。

2.3　句子级对抗迁移模块

在得到融合多粒度特征后的目标语言句子嵌入 $\boldsymbol{S}_i^{\mathrm{t}}=\{\boldsymbol{s}_1^{\mathrm{t}},\boldsymbol{s}_2^{\mathrm{t}},\cdots,\boldsymbol{s}_i^{\mathrm{t}},\cdots,\boldsymbol{s}_Q^{\mathrm{t}}\}$ 与通过线性映射后的源语言句子嵌入 $\boldsymbol{S}_i^{\mathrm{s}}=\{\boldsymbol{s}_1^{\mathrm{s}},\boldsymbol{s}_2^{\mathrm{s}},\cdots,\boldsymbol{s}_i^{\mathrm{s}},\cdots,\boldsymbol{s}_Y^{\mathrm{s}}\}$ 后，本文使用多头注意力作为特征共享编码器来对这两种句子分别提取句子级特征。但由于不同的语言有不同的词序和句子结构，共享编码器不能保证提取到的特征是与语言无关的序列特征，而且两种语言的标注资源不平衡，编码器更倾向于提取标注资源较多的语言（源语言）的特征，而该特征并不一定有助于目标语言的实体标注识别，因此本文使用句子级对抗迁移的方式使得特征共享编码器可以提取到更多与语言无关的序列特征。

将目标语言句子嵌入 $\boldsymbol{S}_i^{\mathrm{t}}=\{\boldsymbol{s}_1^{\mathrm{t}},\boldsymbol{s}_2^{\mathrm{t}},\cdots,\boldsymbol{s}_i^{\mathrm{t}},\cdots,\boldsymbol{s}_Q^{\mathrm{t}}\}$ 与源语言句子嵌入 $\boldsymbol{S}_i^{\mathrm{s}}=\{\boldsymbol{s}_1^{\mathrm{s}},\boldsymbol{s}_2^{\mathrm{s}},\cdots,\boldsymbol{s}_i^{\mathrm{s}},\cdots,\boldsymbol{s}_Y^{\mathrm{s}}\}$ 分别输入多头注意力特征共享编码器中，得到目标语言句子嵌入的多头注意力结果 $\boldsymbol{H}_{\mathrm{t}}=\{\boldsymbol{h}_1^{\mathrm{t}},\boldsymbol{h}_2^{\mathrm{t}},\cdots,\boldsymbol{h}_Q^{\mathrm{t}}\}$ 与源语言句子嵌入的多头注意力结果 $\boldsymbol{H}_{\mathrm{s}}=\{\boldsymbol{h}_1^{\mathrm{s}},\boldsymbol{h}_2^{\mathrm{s}},\cdots,\boldsymbol{h}_Y^{\mathrm{s}}\}$。

基于得到的目标语言序列特征 $\boldsymbol{H}_{\mathrm{t}}=\{\boldsymbol{h}_1^{\mathrm{t}},\boldsymbol{h}_2^{\mathrm{t}},\cdots,\boldsymbol{h}_Q^{\mathrm{t}}\}$ 与源语言的序列特征 $\boldsymbol{H}_{\mathrm{s}}=\{\boldsymbol{h}_1^{\mathrm{s}},\boldsymbol{h}_2^{\mathrm{s}},\cdots,\boldsymbol{h}_Y^{\mathrm{s}}\}$，使用句子级鉴别器预测输入模型的一个句子是否来自目标语言或源语言。对于一个句子表示 x^*，首先使用特征共享编码器提取序列特征 $\boldsymbol{H}=\{\boldsymbol{h}_1^*,\boldsymbol{h}_2^*,\cdots,\boldsymbol{h}_n^*\}$，然后将特征输入带有最大池化的CNN中得到 x^* 的整体向量表示，最后将向量表示输入多层感知器 $\tilde{D}$ 中以预测 x^* 来自目标语言的可能性。通过最小化如式（13）所示的交叉熵损失函数来训练句子级鉴别器：

$$L_{\mathrm{dis}}^x=-\frac{1}{\tilde{I}_{\mathrm{t;\,s}}}\cdot\sum_{i=0}^{\tilde{I}_{\mathrm{t;\,s}}}(\tilde{y}_i\cdot\log_a(\tilde{D}(x_i^*))+(1-\tilde{y}_i)\cdot\log_a(1-\tilde{D}(x_i^*)))$$

$$\tilde{y}_i=\tilde{\delta}_i(1-2\eta)+\eta\tag{13}$$

其中：当 x_i^* 来自目标语言时，$\tilde{\delta}_i=1$，否则 $\tilde{\delta}_i=0$；$\tilde{I}_{\mathrm{t;\,s}}$ 表示目标语言句子与源语言句子数目之和；η 表示附加在句子标签上的平滑值。

特征共享编码器与句子级鉴别器在训练过程中互相对抗混淆对方，试图使 $\tilde{D}$ 分辨不出 x_i^* 具体来自何种语言以使特征共享编码器的参数得到优化。同时，转换句子标签，最小化如式（14）所示的交叉熵损失函数以优化特征共享编码器中的参数：

$$L_{\mathrm{c}}^x=-\frac{1}{\tilde{I}_{\mathrm{t;\,s}}}\cdot\sum_{i=0}^{\tilde{I}_{\mathrm{t;\,s}}}((1-\tilde{y}_i)\cdot\log_a(\tilde{D}(x_i^*))+\tilde{y}_i\cdot\log_a(1-\tilde{D}(x_i^*)))$$

$$\tilde{y}_i=\tilde{\delta}_i(1-2\eta)+\eta\tag{14}$$

2.4　CRF推理模块

在特征共享编码器提取到与语言无关的序列特征后，可以利用所有目标语言与源语言已标注的训练数据训练一个仅针对目标语言的实体识别器。将得到的特征送入基于多头注意力的上下文编码器中来重新捕获每个词的上下文语义依赖关系，然后使用CRF作为最后的输出层[19-21]，给每个事件实体打上预测的标签。

首先在得到共享编码器提取到的序列特征 $\boldsymbol{H}=\{\boldsymbol{h}_1^*,\boldsymbol{h}_2^*,\cdots,\boldsymbol{h}_n^*\}$ 后，将 $\boldsymbol{H}=\{\boldsymbol{h}_1^*,\boldsymbol{h}_2^*,\cdots,\boldsymbol{h}_n^*\}$ 输入多头注意力上下文编码器中进行注意力计算，计算过程与基于多头注意力特征共享编码器中的计算过程相似，结果得到上下文特征序列 $\tilde{\boldsymbol{H}}=\{\tilde{\boldsymbol{h}}_1,\tilde{\boldsymbol{h}}_2,\cdots,\tilde{\boldsymbol{h}}_n\}$。然后使用线性层 ℓ 将每一个 $\tilde{\boldsymbol{h}}_i$ 转换成一个分数向量 $\boldsymbol{y}_i$，$\boldsymbol{y}_i$ 中每一个维度代表一个标签的预测得分。最后将分数向量序列 $\boldsymbol{Y}=\{\boldsymbol{y}_1,\boldsymbol{y}_2,\cdots,\boldsymbol{y}_n\}$ 送入CRF层。标签序列 $Z=\{z_1,z_2,\cdots,z_n\}$ 的得分计算如下：

$$\mathrm{Score}(x,\boldsymbol{Y},Z)=\sum_{i=1}^{n}(\boldsymbol{R}_{z_{i-1},z_i}+\boldsymbol{Y}_{i,z_i})\tag{15}$$

其中：$\boldsymbol{R}$表示转换矩阵；$\boldsymbol{R}_{p,q}$表示从标签p到标签q的转换得分；$\boldsymbol{Y}_{i,z}$表示将第i个单词打上标签z的得分。

对于已标注的标签序列Z，通过式(16)计算得到CRF的损失函数：

$$L_{\mathrm{crf}}=\log_a\sum_{Z'\in\tilde{Z}}\mathrm{e}^{\mathrm{Score}(x,\boldsymbol{Y},Z')}-\mathrm{Score}(x,\boldsymbol{Y},Z) \tag{16}$$

其中：$\tilde{Z}$包含所有可能的标签路径。

通过最小化损失函数$L'=L_{\mathrm{c}}^{x}+L_{\mathrm{crf}}$对特征共享编码器、上下文编码器和CRF进行联合优化，使用随机梯度下降法最小化L_{dis}^{x}和L'。

3 实验结果与分析

3.1 实验数据与参数设置

本文提出一种融合词典与对抗迁移的越南语事件实体识别模型，在属于低资源语言范畴内的越南语上进行模型性能评估。越南语数据集采用人工构造的越南语新闻数据集，数据集中包含预定义的人名、地名、组织机构名、特定政治概念名等实体类型。针对作为目标语言的越南语，分别选用属于高资源语言范畴内的英语和汉语作为源语言。实验中用到的目标语言和源语言新闻数据集篇章(Paragraph)数与句子(Sentence)数的详细统计信息如表1所示，其中“—”表示实验中未设置英语新闻和汉语新闻的验证集与测试集。

表1 数据集篇章数与句子数统计

Table 1 Statistics of the number of paragraphs and sentences in the dataset

数据集	类型	训练集	验证集	测试集
越南语新闻	Paragraph	1 300	165	172
	Sentence	17 923	2 100	2 433
英语新闻	Paragraph	4 521	—	—
	Sentence	39 887	—	—
汉语新闻	Paragraph	4 832	—	—
	Sentence	40 026	—	—

实验中使用准确率(P)、召回率(R)和F1值(F)作为评价指标[20]，指标计算公式如式(17)~式(19)所示：

$$P=\frac{T_{\mathrm{P}}}{T_{\mathrm{P}}+F_{\mathrm{P}}}\times 100\% \tag{17}$$

$$R=\frac{T_{\mathrm{P}}}{T_{\mathrm{P}}+F_{\mathrm{N}}}\times 100\% \tag{18}$$

$$F=\frac{2\times P\times R}{P+R}\times 100\% \tag{19}$$

其中：F_{N}代表模型未能识别出的实体个数；F_{P}代表模型识别出的非实体个数；T_{P}代表模型正确识别出的实体个数。

对越南语、英语和汉语新闻语料均使用FastText[22]工具分别训练其各自的单语词嵌入，实验超参数设置如表2所示。

表2 超参数设置

Table 2 Setting of hyperparameters

参数名称	参数取值
单语词嵌入维度	300
字符级CNN滤波器个数	25
字符级CNN滤波器宽度	2~3
多头注意力的头数h	8
Dropout率	0.5
词级鉴别器的平滑值ε	0.1
词级对抗训练轮次	5
句子级鉴别器的平滑值η	0.3
句子级对抗训练轮次	60
CRF实体识别器的训练轮次	60
训练批次大小	20
初始化学习率	0.01
优化器	SGD

3.2 对比实验

3.2.1 对比实验设置

为验证本文模型的有效性，将其与单语实体识别模型和主流基线模型进行比较：

1)单语实体识别模型[1]。仅使用目标语言标注语料进行训练，利用目前比较流行的BiLSTM-CRF神经网络进行越南语事件实体识别。

2)多任务学习模型[4]。使用多任务学习的方式实现目标语言的实体标注，通过使用权重共享的上下文编码器将源语言的标注信息迁移到目标语言上，从而提升越南语的实体标注准确率。

3)词级对抗实现双语词嵌入表示模型[8]。仅使用词级对抗迁移的方式将源语言映射到目标语言空间，然后利用所有的源语言和越南语的标注信息对越南语文本进行实体识别。在将源语言映射到目标语言空间后：直接使用两种语言的所有标注信息训练实体识别器对越南语进行标注，记为词级对抗实现双语词嵌入表示模型1；先使用越南语的标注信息训练实体识别器，再使用源语言的标注信息进行调优，记为词级对抗实现双语词嵌入表示模型2。

4)双语词典实现双语词嵌入表示模型[11]。使用预先构造好的双语词典对齐源语言与目标语言的词向量空间，通过最近邻搜索算法找到与源语言词距离最近的目标语言词作为该源语言词的翻译词。利用翻译词和其源语言词对应的标签训练融合自注意力机制的BiLSTM-CRF网络对越南语文本进行实体识别。

5)两层对抗迁移模型[12]。利用BiLSTM-CRF网络，首先使用词级对抗迁移的方式将源语言映射到目标语言空间上，然后使用句子级对抗迁移的方式使得共享编码器提取与语言无关的序列特征，最后融合上下文语义信息训练实体识别器对越南语进行标注。

3.2.2　无目标语言标注数据情况下的跨语言迁移

比较在无目标语言(越南语)标注数据的情况下,本文模型与对比模型在性能上的差异。在进行句子级对抗迁移训练时,移除输入的目标语言(越南语)句子的标签信息,在只有源语言标注数据的情况下对句子鉴别器进行优化,训练出针对目标语言的实体识别器。以英语和汉语作为源语言对越南语进行实体识别,实验结果如表3所示。

表3　无目标语言标注数据情况下的实体识别性能

Table 3　Entity recognition performance without target language annotation data　%

模型	英语			汉语		
	P	*R*	*F*	*P*	*R*	*F*
多任务学习模型	24.55	23.67	24.10	11.93	9.48	10.56
词级对抗实现双语词嵌入表示模型1	25.44	21.33	23.20	16.89	14.73	15.74
双语词典实现双语词嵌入表示模型	27.48	25.99	26.71	23.68	22.34	22.99
两层对抗迁移模型	37.96	38.34	38.15	25.77	24.12	24.92
本文模型	45.82	44.71	45.26	44.90	42.81	43.83

从表3的对比结果可以看出,本文模型在源语言为英语或汉语的情况下的实体识别性能均优于对比模型。与仅包含权重共享的上下文编码器的多任务学习模型相比,本文模型不仅加入了语言共享的上下文编码器,而且使用多级对抗训练的方式促使两种语言的词进行语义对齐,基于双语词典融入多粒度特征信息,使用特征共享编码器提取与语言无关的序列特征。因此,在源语言为英语和汉语的情况下,本文模型的F1值增加了21.16和33.27个百分点,提升效果显著。与词级对抗实现双语词嵌入表示模型和两层对抗迁移模型相比,本文模型不仅使用词级对抗和句子级对抗迁移,更重要的是加入了基于双语词典及注意力的多粒度特征嵌入。因此,本文模型在准确率、召回率和F1值上均有一定程度的提升。与双语词典实现双语词嵌入表示模型相比,本文模型在其基础上加入了多级对抗迁移,提升了两种语言词的语义对齐效果,从而使得最终的实体识别性能有所提升。

3.2.3　有目标语言标注数据情况下的跨语言迁移

在有目标语言(越南语)标注数据的情况下,比较本文模型与对比模型的性能差异。以英语和汉语作为源语言对越南语进行实体识别,实验结果如表4所示。

表4　有目标语言标注数据情况下的实体识别性能

Table 4　Entity recognition performance with target language annotation data　%

模型	英语			汉语		
	P	*R*	*F*	*P*	*R*	*F*
单语实体识别模型	71.22	69.67	70.44	71.22	69.67	70.44
多任务学习模型	71.38	69.93	70.65	70.12	68.97	69.54
词级对抗实现双语词嵌入表示模型1	71.06	69.59	70.32	71.20	69.62	70.40
词级对抗实现双语词嵌入表示模型2	71.26	70.01	70.63	71.18	69.45	70.30
双语词典实现双语词嵌入表示模型	77.34	78.91	78.12	78.55	79.17	78.86
两层对抗迁移模型	81.89	79.41	80.63	79.46	79.90	79.68
本文模型	90.45	89.66	90.05	89.31	89.03	89.17

从表4的对比结果可以看出,词级对抗实现双语词嵌入表示模型和单语实体识别模型在利用目标语言标注数据进行训练的基础上,直接加入源语言标注数据可能会降低模型性能。这也说明了在用于训练的目标语言标注数据不足时,模型会对噪声更加敏感,在加入源语言标注数据的同时也引入了噪声影响模型性能。当源语言与目标语言属于同一语系时,多任务学习模型的识别结果优于单语实体识别结果;反之,结果则相反。

加入源语言标注数据会引入噪声的主要原因在于源语言与目标语言的语言表达和序列结构不相同。双语词典实现双语词嵌入表示模型利用预先构造好的双语词典对齐源语言与目标语言的词向量空间,找到源语言词的翻译词,从而实现源语言到目标语言的转换,减弱数据噪声。两层对抗迁移模型使用共享编码器提取到与语言无关的序列特征,从而达到减弱源语言标注数据噪声的问题。从实验结果可以看出:这两种模型的F1值相较单语实体识别模型均有大幅提升;本文模型不仅利用双语词典融入了多粒度特征信息,而且使用基于多头注意力的特征共享编码器提取与语言无关的序列特征,分别在源语言为英语和汉语的情况下,相较单语实体识别模型的F1值增加了19.61和18.73个百分点,提升效果明显。以上实验结果证明了本文模型能利用源语言标注数据提升目标语言事件实体识别性能。

4 结束语

本文提出一种融合词典与对抗迁移的越南语事件实体识别模型，利用词级对抗迁移训练将源语言和目标语言映射到同一语义空间中，通过双语词典及注意力进行多粒度特征嵌入使得目标语言和映射后的源语言获得更丰富的语义表示，高度关注对实体识别有用的信息。同时，考虑到不同语言有不同的语言表达和序列结构，因此利用句子级对抗迁移训练以使得基于多头注意力的特征共享编码器可以提取到与语言无关的序列特征。实验结果表明，本文模型在属于低资源语言范畴内的越南语新闻数据集上相较于当前主流的单语实体识别模型和迁移学习模型效果均有显著提升。但是本文模型相比汉语、英语等高资源语言的单语实体识别模型在F1值上相对较低，下一步将考虑在其中加入篇章级对抗迁移训练以融入源语言篇章级信息，同时构建针对越南语事件实体识别任务的无监督预训练跨语言模型，进一步提升实体识别性能。

参考文献

[1] LAMPLE G, BALLESTEROS M, SUBRAMANIAN S, et al. Neural architectures for named entity recognition[C]// Proceedings of 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Stroudsburg, USA: Association for Computational Linguistics, 2016: 260-270.

[2] DOAN X D, DANG T T, NGUYEN L M. Effectiveness of character language model for Vietnamese named entity recognition[C]//Proceedings of the 32nd Pacific Asia Conference on Language, Information and Computation. Stroudsburg, USA: Association for Computational Linguistics, 2018: 157-163.

[3] YANG Z, SALAKHUTDINOV R, COHEN W. Multi-task cross-lingual sequence tagging from scratch[EB/OL]. [2020-12-28]. https://arxiv. org/pdf/1603. 06270. pdf.

[4] LIN Y, YANG S Q, STOYANOV V, et al. A multi-lingual multi-task architecture for low-resource sequence labeling[C]// Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics. Stroudsburg, USA: Association for Computational Linguistics, 2018: 799-809.

[5] FANG M, COHN T. Learning when to trust distant supervision: an application to low-resource POS tagging using cross-lingual projection[C]//Proceedings of the 20th SIGNLL Conference on Computational Natural Language Learning. Stroudsburg, USA: Association for Computational Linguistics, 2016: 178-186.

[6] WANG D, PENG N, DUH K. A multi-task learning approach to adapting bilingual word embeddings for cross-lingual named entity recognition[C]//Proceedings of the 8th International Joint Conference on Natural Language Processing. Stroudsburg, USA: Association for Computational Linguistics, 2017: 383-388.

[7] SHI G, FENG C, HUANG L F, et al. Genre separation network with adversarial training for cross-genre relation extraction[C]//Proceedings of 2018 Conference on Empirical Methods in Natural Language Processing. Stroudsburg, USA: Association for Computational Linguistics, 2018: 1018-1023.

[8] CONNEAU A, LAMPLE G, RANZATO M A, et al. Word translation without parallel data[EB/OL]. [2020-12-28]. http://arxiv. org/abs/1710. 04087.

[9] FANG M, COHN T. Model transfer for tagging low-resource languages using a bilingual dictionary[C]// Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics. Stroudsburg, USA: Association for Computational Linguistics, 2017: 587-593.

[10] ZIRIKLY A. Cross-lingual transfer of named entity recognizers without parallel corpora[C]//Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing. Stroudsburg, USA: Association for Computational Linguistics, 2015: 390-396.

[11] XIE J T, YANG Z L, NEUBIG G, et al. Neural cross-lingual named entity recognition with minimal resources[C]// Proceedings of 2018 Conference on Empirical Methods in Natural Language Processing. Stroudsburg, USA: Association for Computational Linguistics, 2018: 369-379.

[12] HUANG L F, JI H, MAY J. Cross-lingual multi-level adversarial transfer to enhance low-resource name tagging[C]// Proceedings of 2019 Conference of the North. Stroudsburg, USA: Association for Computational Linguistics, 2019: 3823-3833.

[13] ZHANG M, LIU Y, LUAN H B, et al. Adversarial training for unsupervised bilingual lexicon induction[C]//Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics. Stroudsburg, USA: Association for Computational Linguistics, 2017: 1959-1970.

[14] CAO P F, CHEN Y B, LIU K, et al. Adversarial transfer learning for Chinese named entity recognition with self-attention mechanism[C]//Proceedings of 2018 Conference on Empirical Methods in Natural Language Processing. Stroudsburg, USA: Association for Computational Linguistics, 2018: 182-192.

[15] XING C, WANG D, LIU C, et al. Normalized word embedding and orthogonal transform for bilingual word translation[C]//Proceedings of 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Stroudsburg, USA: Association for Computational Linguistics, 2015: 1006-1011.

[16] GOODFELLOW I, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial nets[C]//Proceedings of the 27th Annual Conference on Neural Information Processing Systems. New York, USA: ACM Press, 2014: 2672-2680.

[17] SUKHBAATAR S, WESTON J, FERGUS R. End-to-end memory networks[C]//Proceedings of the 28th International Conference on Neural Information Processing Systems. New York, USA: ACM Press, 2015: 2440-2448.

[18] YIN R C, WANG Q, LI P, et al. Multi-granularity Chinese word embedding[C]//Proceedings of 2016 Conference on Empirical Methods in Natural Language Processing. Stroudsburg, USA: Association for Computational Linguistics, 2016: 981-986.

（下转第145页）

（上接第 114 页）

[19] 张应成，杨洋，蒋瑞，等. 基于BiLSTM-CRF的商情实体识别模型[J]. 计算机工程，2019，45(5)：308-314.
ZHANG Y C，YANG Y，JIANG R，et al. Commercial intelligence entity recognition model based on BiLSTM-CRF [J]. Computer Engineering，2019，45(5)：308-314. (in Chinese)

[20] 何阳宇，晏雷，易绵竹，等. 融合CRF与规则的老挝语军事领域命名实体识别方法[J]. 计算机工程，2020，46(8)：297-304.
HE Y Y，YAN L，YI M Z，et al. Named entitiy recognition method for Laotian in military field combining CRF and rules[J]. Computer Engineering，2020，46(8)：297-304. (in Chinese)

[21] 买买提阿依甫，吾守尔·斯拉木，帕丽旦·木合塔尔，等. 基于BiLSTM-CNN-CRF模型的维吾尔文命名实体识别[J]. 计算机工程，2018，44(8)：230-236.
Maimaitiayifu，Silamu Wushouer，Muhetaer Palidan，et al. Uyghur named entity recognition based on BiLSTM-CNN-CRF model[J]. Computer Engineering，2018，44(8)：230-236. (in Chinese)

[22] JOULIN A，GRAVE E，BOJANOWSKI P，et al. Bag of tricks for efficient text classification[C]//Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics. Stroudsburg，USA：Association for Computational Linguistics，2017：427-431.

编辑　陆燕菲