文章编号:1008-1542(2023)01-0020-09

联合实体边界检测的命名实体识别方法

李晓腾[1],勾智楠[2],高　凯[1]

(1.河北科技大学信息科学与工程学院,河北石家庄　050018;2.河北经贸大学信息技术学院,河北石家庄　050061)

摘　要:针对传统命名实体识别方法无法有效利用实体边界信息的问题,提出一种联合实体边界检测的命名实体识别方法,即将实体边界检测作为辅助任务,增强模型对实体边界的判断能力,进而提升模型对实体的识别效果。首先,利用 Bert 预训练语言模型对原始文本进行特征嵌入获取词向量,并引入自注意力机制增强词对上下文信息的利用;其次,在命名实体识别任务的基础上,添加实体边界检测辅助任务,增强模型对实体边界的识别能力;再次,对比联合实体边界检测的命名实体识别方法与基线方法的有效性,并对测试结果进行消融实验;最后,进行样例分析,分析损失权重 β 对实体边界检测的影响。实验结果表明,在英文社交媒体数据集 Twitter-2015 上,联合实体边界检测的命名实体识别方法相较于基线模型取得了更高的精准率、召回率和 F1 值,其中 F1 值达到了 73.57%;并且,边界检测辅助任务提升了基线方法的检测效果。所提方法能有效利用实体边界信息,从而获得更好的实体识别效果,促进了人机交互系统的发展,对自然语言处理下游任务有重要意义。

关键词:自然语言处理;命名实体识别;实体边界检测;辅助任务;深度学习

中图分类号:TP391.1　　**文献标识码**:A　　**DOI**:10.7535/hbkd.2023yx01003

Named entity recognition method based on joint entity boundary detection

LI Xiaoteng[1], GOU Zhinan[2], GAO Kai[1]

(1. School of Information Science and Engineering, Hebei University of Science and Technology, Shijiazhuang, Hebei 050018, China; 2. School of Information Technology, Hebei University of Economics and Business, Shijiazhuang, Hebei 050061, China)

Abstract: To solve the problem that traditional named entity recognition methods cannot effectively utilize entity boundary information, a named entity recognition method based on joint entity boundary detection was proposed. The method took entity boundary detection as an auxiliary task, so that the model can enhance the ability of entity boundary recognition, and then

收稿日期:2022-02-21;修回日期:2022-12-25;责任编辑:王淑霞
基金项目:河北省自然科学基金面上项目(F2022208006);河北省高等学校科学技术研究项目(QN2020198)
第一作者简介:李晓腾(1994—),男,河北石家庄人,硕士研究生,主要从事自然语言处理方面的研究。
通信作者:高　凯教授。E-mail:gaokai@hebust.edu.cn

李晓腾,勾智楠,高凯.联合实体边界检测的命名实体识别方法[J].河北科技大学学报,2023,44(1):20-28.
LI Xiaoteng, GOU Zhinan, GAO Kai. Named entity recognition method based on joint entity boundary detection[J]. Journal of Hebei University of Science and Technology, 2023, 44(1): 20-28.

improve the effect of entity recognition. Firstly, the Bert pretraining language model was used to embed the features of the original text to obtain word vectors, and the self-attention mechanism was introduced to enrich the context features of words. Secondly, on the basis of named entity recognition task, an auxiliary entity boundary detection task was added to enhance the recognition ability of the model to the entity boundaries. Thirdly, the effectiveness of the named entity recognition method and the baseline method was compared, and the test results were from ablation experiments. Finally, the influence of loss weight β on entity boundary detection was analyzed by examples. The experimental results show that on the English social media dataset Twitter-2015, the named entity recognition method combined with entity boundary detection achieves higher accuracy, recall rate and F1 value than the baseline model, of which the F1 value can reach 73.57%. In addition, the boundary detection auxiliary task has a certain improvement effect on the baseline method. The proposed method can effectively utilize entity boundary information to obtain better entity recognition effect, and promote the development of human-computer interaction system, which is of great significance for downstream tasks of natural language processing.

Keywords: natural language processing; named entity recognition; entity boundary detection; auxiliary task; deep learning

命名实体识别(named entity recognition, NER)是指抽取文本序列中的“人名”、“地名”、“机构名”等实体，是一项重要的自然语言处理任务。命名实体识别任务广泛应用于其他自然语言处理任务，如信息抽取、信息检索、问答系统以及知识图谱构建等[1]。随着社交网络的快速发展，人们在社交网络上通过文字来表达自己的观点，浩如烟海的社交网络数据亟待处理，命名实体识别是结构化处理社交媒体数据中的关键技术，命名实体识别任务可以提取出社交网络数据中的“人名”、“地名”、“机构名”等实体，对社交媒体数据的归纳分类具有重要意义。

近年来，随着深度学习的不断发展，许多深度学习方法被应用到命名实体识别任务中。YANG 等[2]结合双向长短时记忆网络(bi-directional long-short term memory，Bi-LSTM)和条件随机场[3](conditional random field，CRF)来处理命名实体识别问题，目前 Bi-LSTM+CRF 依然是一种常见的命名实体识别处理方法。REI[4]通过添加语言模型的辅助任务，学习文本中深层的语义、语法信息，帮助模型获得更强大的文本特征表示。LIN 等[5]提出利用迁移学习缓解 NER 任务中数据不足的问题，利用源域的大量有标注数据学习知识，然后利用迁移学习方法，将知识迁移到目标域，缓解目标域数据不足的问题。YANG 等[6]利用远监督方法产生的数据在新领域进行命名实体识别。ZHOU 等[7]提出利用对抗学习处理 NER 任务，在原始数据中添加扰动生成对抗样本，判别器判断样本的正负性，使得模型可以更好地处理文本中的噪声，提升了模型的鲁棒性。2018 年谷歌提出的 Bert 模型[8]，在 11 项 NLP 任务中获得了最优结果。随着 Bert 获得的巨大成功，涌现了许多对 Bert 改进的方法，如 BERT-WWM [9]，SpanBERT [10]，UNILM[11]，ViLBERT[12]和 K-BERT[13]等。多任务学习在 NER 任务中同样具有广泛应用。多任务学习是指将多个相关的任务联合在一起训练，通过共享任务之间的特征信息，获得一个更好的效果[14]。多任务学习中常见的参数共享方式有 2 种，一种是硬共享[15]，另一种是软共享。LI 等[16]添加情感分类任务作为立场检测的辅助任务，并引入靶向注意力机制提升立场检测效果。多任务学习在命名实体识别任务中也得到了广泛应用。LIN 等[17]提出了一种跨语言的多任务学习方式，缓解特定 NER 领域语料不足的问题。GREENBERG 等[18]针对生物医学领域数据不足的问题，提出了使用多类数据集来训练网络模型，增强模型的泛化能力。ZHAO 等[19]通过联合实体规范化任务，在 2 个任务之间增加反馈链路，提升了 NER 任务和实体规范化任务的效果。

命名实体识别领域虽然已有大量优秀的研究成果，但已有方法忽略了对实体边界信息的利用。实体边界信息是实体识别中的一项重要信息，对实体的正确识别有重要意义。为了充分利用实体边界信息，本文提出一种联合实体边界检测的命名实体识别方法(joint entity boundary detection named entity recognition，JEBD-NER)。在命名实体识别模型的基础上，通过引入实体边界检测任务，帮助模型学习到实体边界信息。此外，相似的实体有相似的上下文，因此为了增强词对上下文信息的利用，引入自注意力机制来丰富词的上下文信息，进一步提升模型对实体的识别能力。

1　联合实体边界检测的命名实体识别模型(JEBD-NER)

1.1　任务定义

命名实体识别任务需要在一段文本序列 S 中判断出其中的实体，并对这些实体分类。同其他研究者一

致，本文将该任务视为序列标注任务，模型需要判断出 S 中的实体并对其分类，且判断出实体的边界信息。本文定义文本序列为 $S=(s_1,s_2,\cdots,s_n)$，其中 n 为文本序列长度。$Y=(y_1,y_2,\cdots,y_n)$，为文本序列对应的标签。$Z=(z_1,z_2,\cdots,z_n)$，为实体边界检测任务的标签。其中标签 Y 和 Z 遵循 BIO2 标注原则。B 代表实体的首个字符，I 表示实体的中间或者结尾字符，O 表示非实体字符。

1.2 模型结构

JEBD-NER 方法的模型结构如图 1 所示，整体可分为 3 部分：Bert 编码层、Self-Att 层、多任务学习层。首先，Bert 编码层将原始文本输入转换成词向量 $\boldsymbol{X}$ 供 Self-Att 层使用；其次，Self-Att 层通过自注意力机制增加词对上下文信息的利用，并将融合上下文信息的文本特征向量 $\boldsymbol{A}$ 传入多任务学习层；最后，多任务学习层联合了命名实体识别任务与实体边界检测任务，并利用文本特征向量 $\boldsymbol{A}$ 分别输出实体和实体边界预测结果。

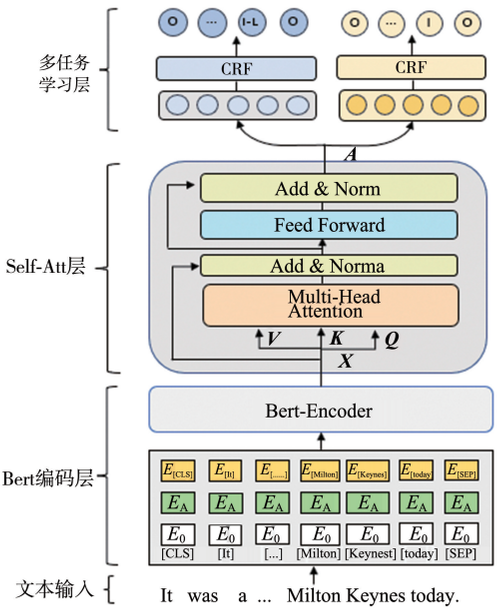

图 1 实体边界检测命名实体模型结构

Fig. 1 JEBD-NER model architecture

1.2.1 基于 Bert 预训练语言模型的特征嵌入层

如图 1 中 Bert 编码层所示，为了增强对原始文本的嵌入能力，Bert 采用 token 嵌入、segment 嵌入、position 嵌入联合表示的方法来增强字符级、词级、句级的特征信息表示。Bert-Encoder 则是使用 Transformer 的编码器。定义 $\boldsymbol{S}'=(\boldsymbol{s}_0,\boldsymbol{s}_1,\cdots,\boldsymbol{s}_{n+1})$，为 Bert 编码器的输入，其中 $\boldsymbol{s}_0$ 和 $\boldsymbol{s}_{n+1}$ 分别代表文本序列的开始字符[CLS]和结束字符[SEP]。$\boldsymbol{s}_i$ 由 token 嵌入、segment 嵌入、position 嵌入构成。$\boldsymbol{X}=(\boldsymbol{x}_0,\boldsymbol{x}_1,\cdots,\boldsymbol{x}_{n+1})$，作为 Bert 编码器的输出，即词的特征向量，$\boldsymbol{x}_i\in\mathbb{R}^d$ 是 $\boldsymbol{s}_i$ 的特征向量，d 是特征向量维度。

1.2.2 基于 Self-Attention 的上下文语义交互层

对于文本序列而言，如何有效利用上下文信息是识别实体的关键。因为对于相似的上下文而言，其中的实体类型是相似的。例如，“我的家乡在河北石家庄”，其中“河北石家庄”是地点实体。“我的家乡在济南”，其中“济南”是地点实体。由上述 2 个例子可知，在上下文相似的情况下，实体类型是相似的。如何有效利用上下文信息是判断实体的关键，因此本文引入 Self-Attention 机制来增强词对上下文信息的利用。

如图 1 中 Self-Att 层所示，为了有效利用上下文信息，引入 Self-Attention 机制[20]。Self-Attention 机制是一种自注意力方法，其中注意力模块计算公式如式(1)所示：

$$\text{Attention}(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V})=\text{softmax}\left(\frac{\boldsymbol{Q}\boldsymbol{K}^{\mathrm{T}}}{\sqrt{d_k}}\right)\boldsymbol{V}\ , \tag{1}$$

式中：$\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V}$ 分别指注意力机制中的查询向量、键值向量、权值向量；d_k 为输入向量的维度。在使用自注意力机制时，通常会使用多个注意力网络并行计算，每个注意力称为一个注意力头。第 i 个注意力头计算公式如式(2)所示：

$$\text{Head}_i=\text{Attention}(\boldsymbol{Q}\boldsymbol{W}_i^q,\boldsymbol{K}\boldsymbol{W}_i^k,\boldsymbol{V}\boldsymbol{W}_i^v)。 \tag{2}$$

式中：$\boldsymbol{W}_i^q,\boldsymbol{W}_i^k,\boldsymbol{W}_i^v$ 为第 i 个注意力头的线性映射变换权重。最终的 h 个注意力头拼接结果为[Head_1，Head_2，…，Head_h]。

在本文模型中，文本序列 $\boldsymbol{X}=(\boldsymbol{x}_0,\boldsymbol{x}_1,\cdots,\boldsymbol{x}_{n+1})$作为多头自注意力的输入，最终经过多头自注意力机制得到文本序列特征 $\boldsymbol{A}=(\boldsymbol{A}_0,\boldsymbol{A}_1,\cdots,\boldsymbol{A}_{n+1})$。

1.2.3 联合实体边界检测的多任务学习层

为了更好地利用实体边界信息，本文提出联合实体边界检测的命名实体识别方法。实体边界信息指的是实体词组在文本序列中的位置信息，即文本序列中实体词组开始到结束的位置信息。命名实体识别任务需要同时识别出实体词组的边界信息和实体类别信息。因此，提升模型对实体词组的边界识别能力可以在一定程度上促进命名实体的识别效果。受多任务学习策略启发，在命名实体识别任务基础上，引入实体边界

检测辅助任务。实体边界检测任务是与命名实体识别任务高度相关的任务，可以帮助模型有效学习实体边界信息。在本方法中采用硬共享的方式来共享参数信息，硬共享是目前应用最广泛的共享机制，它把多个任务的数据表示嵌入到同一个特征语义空间之中，多个任务之间共享模型底层参数，从而使得底层参数学习到多个任务的知识，提升实验效果。

如图 1 中多任务学习层所示，该层有 2 个分支：命名实体识别分支和实体边界检测。命名实体识别分支在图左侧，作为主任务，其根据输入的文本序列特征 $\boldsymbol{A}$ 预测出实体结果；实体边界检测分支在图右侧，作为辅助任务，其根据输入的文本序列特征 $\boldsymbol{A}$ 预测实体边界结果。命名实体识别任务的标签定义为 $\boldsymbol{Y}=(y_1,y_2,\cdots,y_n)$，边界检测任务的标签定义为 $\boldsymbol{Z}=(z_1,z_2,\cdots,z_n)$，在训练过程中，根据预测结果与标签之间的损失来优化文本序列特征 $\boldsymbol{A}$。

命名实体识别分支　在该分支中，模型将文本序列特征输入 CRF 层，输出对实体的预测标签序列。将文本序列特征 $\boldsymbol{A}$ 经过线性层(LN)控制维度，得到新序列特征 $\boldsymbol{W}$，具体公式如式(3)所示：

$$\boldsymbol{W}=\mathrm{LN}(\boldsymbol{A})。\tag{3}$$

考虑到文本序列标签之间的依赖关系，本文采用 CRF 结构学习标签之间的依赖关系。给定特征 $\boldsymbol{W}$，得到预测序列标签 y' 的概率如式(4)－式(6)所示：

$$P(y'\mid\boldsymbol{W})=\frac{\exp(\mathrm{score}(w,y'))}{\sum\limits_{y^*\in Y^*}\exp(\mathrm{score}(w,y^*))},\tag{4}$$

$$\mathrm{score}(w,y')=\sum_{i=0}^{n}T_{y'_i,y'_{i+1}}+\sum_{i=1}^{n}E_{w_i,y'_i},\tag{5}$$

$$E_{w_i,y'_i}=\boldsymbol{W}^{y'_i}\cdot w_i。\tag{6}$$

式中：Y^* 是所有可能标签序列集合，每种可能的序列得分 $\mathrm{score}(w,y')$ 由 $T_{y_i',y'_{i+1}}$ 和 E_{w_i,y'_i} 共 2 部分构成。其中：$T_{y_i',y'_{i+1}}$ 是标签 y'_i 到 y'_{i+1} 之间的转移得分；E_{w_i,y'_i} 是第 i 个词预测为 y'_i 的发射得分；$\boldsymbol{W}^{y'_i}$ 是预测为 y'_i 时的权重参数。

实体边界检测分支　在该分支中，模型将文本序列特征输入 CRF 层，输出实体边界的预测标签序列。首先，将文本序列特征 $\boldsymbol{A}$ 经过线性层(LN)控制维度，得到新序列特征 $\boldsymbol{W}'$。随后，经过 CRF 层学习标签之间的依赖关系。给定特征 $\boldsymbol{W}'$，得到序列标签 z' 的概率如式(7)－式(9)所示：

$$P(z'\mid\boldsymbol{W}')=\frac{\exp(\mathrm{score}(w',z'))}{\sum\limits_{z^*\in Z^*}\exp(\mathrm{score}(w',z^*))},\tag{7}$$

$$\mathrm{score}(w',z')=\sum_{i=0}^{n}T_{z'_i,z'_{i+1}}+\sum_{i=1}^{n}E_{w'_i,z'_i},\tag{8}$$

$$E_{w'_i,z'_i}=\boldsymbol{W}^{z'_i}\cdot w'_i。\tag{9}$$

式中：Z^* 是所有可能标签序列合集，每种可能的序列得分 $\mathrm{score}(w',z')$ 由 $T_{z'_i,z'_{i+1}}$ 和 $E_{w'_i,z'_i}$ 共 2 部分构成。其中：$T_{z'_i,z'_{i+1}}$ 是标签 z'_i 到 z'_{i+1} 之间的转移得分；$E_{w'_i,z'_i}$ 是第 i 个词预测为 z'_i 的发射得分；$\boldsymbol{W}^{z'_i}$ 是预测为 z'_i 时的权重参数。

1.3　模型训练

在模型训练过程中，采用命名实体识别任务损失结合实体边界检测任务损失的方式共同来优化网络参数，其损失函数如式(10)－式(12)所示：

$$\mathrm{loss}_{\mathrm{NER}}=-\frac{1}{N}\sum_{j=1}^{N}\log P(y'_j\mid\boldsymbol{W}_j),\tag{10}$$

$$\mathrm{loss}_{\mathrm{EBD}}=-\frac{1}{N}\sum_{j=1}^{N}\log P(z'_j\mid\boldsymbol{W}'_j),\tag{11}$$

$$\mathrm{loss}=\mathrm{loss}_{\mathrm{NER}}+\beta\cdot\mathrm{loss}_{\mathrm{EDB}}。\tag{12}$$

式中：$\mathrm{loss}_{\mathrm{NER}}$ 是命名实体识别任务损失；$\mathrm{loss}_{\mathrm{EDB}}$ 是实体边界检测任务损失；β 是实体边界检测任务的损失权重系数。

2 实验设计

2.1 数据集和评价指标

为了验证联合实体边界检测的命名实体识别方法的有效性，本文在国际公开数据集 Twitter-2015[21]上进行实验验证。Twitter-2015 是命名实体识别任务中经典的公开数据集，本文选取了 Twitter-2015 数据集中的文本数据来验证模型的有效性。Twitter-2015 来源于 Twitter，包含了“Person”、“Location”、“Organization”、“Misc”共 4 类实体。其数据划分具体情况如表 1 所示，4 类实体在训练集、验证集和测试集中分布情况如表 2 所示。

表 1 Twitter-2015 划分信息

Tab. 1 Divided information of Twitter-2015

数据集名称	训练集	验证集	测试集
Twitter-2015	4 000	1 000	3 257

表 2 Twitter-2015 实体统计信息

Tab. 2 Entitiesstatistics information of Twitter-2015

实体类型	Twitter-2015		
	训练集	验证集	测试集
Person	2 217	552	1 816
Location	2 091	522	1 697
Organization	928	247	839
Misc	940	225	726

采用精准率(Precision，P)、召回率(Recall，R)和 F1 值来评估命名实体识别模型的有效性。

2.2 对比基线模型与参数设置

为了验证 JEBD-NER 模型的有效性，本文对比了经典的命名实体识别基线模型。

BiLSTM-CRF[2] 命名实体识别任务中经典的基线模型，使用 Bi-LSTM 提取字特征，并利用 CRF 层学习序列之间转移关系，提升模型对实体识别效果。

CNN-BiLSTM-CRF[22] 使用 CNN 学习字符级特征，将字符级特征与词嵌入拼接后作为 Bi-LSTM 的输入，后接 CRF 获得最佳的预测标签序列。

HBiLSTM-CRF[23] 使用堆叠 LSTM 层抽取字符级特征，将字符级特征和词嵌入拼接作为 Bi-LSTM 的输入，后接 CRF 层获得最佳的预测标签序列。

Bert-CRF 使用 Bert 对原始文本序列进行词嵌入，得到文本序列的词向量；利用 CRF 学习文本序列之间的转移概率对最后结果预测输出。

Bert-Bi-LSTM-CRF[24] 使用 Bert 对原始文本序列进行词嵌入，得到文本序列的词向量；添加 Bi-LSTM 网络来学习上下文信息，丰富词向量表征信息；利用 CRF 层学习文本序列的转移概率，对最后结果预测输出。

Bert-Self-Att-CRF 使用 Bert 对原始文本序列进行词嵌入，得到文本序列的词向量；使用 Self-Attention 机制学习上下文信息；利用 CRF 层学习文本序列的转移概率，并对最后结果预测输出。

本文代码均使用 Pytorch 框架实现，所使用的显卡为 NVIDIA GeForce GTX TITAN X，显存大小为 12 211 MB。实验中所使用的预训练语言模型为 Bert-base-cased，具体参数信息如表 3 所示。

表 3 参数设置信息

Tab. 3 Parameters setting

参数名称	β	Epoch	Batch-Size	Weight-Decay	Learning-Rate
参数值	0.9	30	32	1e-3	5e-5

2.3 消融实验设计

为进一步说明实体边界检测辅助任务对模型的增益作用，设计实验来验证实体边界检测辅助任务对模型的提升效果。首先，选取 3 组基线模型，分别是 Bert-CRF，Bert-BiLSTM-CRF，Bert-Self-Att-CRF；其次，在基线模型上添加边界检测辅助任务(使用“+EBD”标识)；最后，将添加边界检测辅助任务的基线模型与原始基线模型结果进行对比，观察边界检测辅助任务对最终实验结果的提升效果。

2.4 样例分析

为了直观地展示联合实体边界检测的命名实体识别方法的效果，选取 3 组样例来说明其有效性，选取 Bert-Self-Att-CRF(表中记作 Bert-Self-Att)与 JEBD-NER 进行对比分析。

2.5 损失权重 β 分析

受文献[25]启发，设置实验探索实体边界检测任务的损失权重对 JEBD-NER 方法的影响。模型中其他参数固定，实体边界检测任务损失权重是唯一的变量，其变化范围为(0.1，1.0)，按 0.1 依次递增。为了细粒

度展示实体边界检测任务损失权重对实验结果的影响，将 4 类实体识别的 F1 值使用 4 种不同颜色柱状图展示，并将 4 类实体综合 F1 值（图中记作 Aug-F1）使用蓝色折线图展示。

3　实验结果分析

3.1　对比基线模型结果分析

本文提出的模型与上述基线模型在 Twitter-2015 数据集上进行实验对比，结果如表 4 所示，其中，“各类实体 F1 结果”是指模型在 4 类实体上各自的 F1 值结果；“4 类实体综合结果”是指模型在数据集上的整体实验结果，包括 3 部分，分别是精准率（P）、召回率（R）和 F1 值。本文以“4 类实体综合结果”中 F1 值为首要评价指标。

表 4　Twitter-2015 实验结果

Tab. 4　Experimental results on the Twitter-2015

方法名称	各类实体 F1 值/%				4 类实体综合结果/%		
	PER	LOC	ORG	MISC	P	R	F1
BiLSTM-CRF	76.77	72.56	41.33	26.80	68.14	61.09	64.42
CNN-BiLSTM-CRF	80.86	75.39	47.77	32.61	66.24	68.09	67.15
HBiLSTM-CRF	82.34	76.83	51.59	32.52	70.32	68.05	69.17
Bert-CRF	84.87	80.69	59.49	37.37	70.95	73.82	72.36
Bert-BiLSTM-CRF	84.77	80.58	60.41	39.30	71.06	74.73	72.84
Bert-Self-Att-CRF	84.28	79.92	61.15	38.14	70.66	74.52	72.54
JEBD-NER	**84.89**	**81.33**	**62.59**	**40.84**	**72.10**	**75.09**	**73.57**

由表 4 可知，联合实体边界检测命名实体识别方法相较于基线方法实验结果最优。JEBD-NER 方法在各类实体 F1 值中均表现最优，在 4 类实体综合结果中，3 个评价指标均达到最优。相较于最优基线模型 Bert-BiLSTM-CRF，精准率提升了 1.04%，召回率提升了 0.36%，F1 值提升了 0.73%，这表明本文方法是有效的，对实体识别能力有提升；相较于 Bert-CRF 模型，精准率提升了 1.15%，召回率提升了 1.27%，F1 值提升了 1.21%。这表明增加 Self-Attention 机制以及实体边界检测辅助任务后，模型增强了对上下文信息和实体边界信息的利用，实体识别能力有较大的提升；相较于 Bert-Self-Att-CRF 模型，精准率提升了 1.44%，召回率提升了 0.57%，F1 值提升了 1.03%。这表明增加实体边界检测辅助任务后，模型对实体边界信息能够有效利用，提升了对实体的识别效果。综上分析可知，本文提出的联合实体边界检测的命名实体识别方法是有意义的，模型中的 Self-Attention 机制增强了单词对上下文信息的利用；实体边界检测辅助任务，提升了模型对实体边界的识别能力。

3.2　消融实验结果分析

实验结果如表 5 所示。由表 5 可知，基线模型 Bert-CRF 在添加边界检测辅助任务后，精准率提升了 0.73%，召回率提升了 0.41%，F1 值提升了 0.57%；基线模型 Bert-BiLSTM-CRF 在添加边界检测辅助任务后，精准率提升了 0.42%，召回率提升了 0.7%，F1 值提升了 0.61%。基线模型 Bert-Self-Att-CRF 添加边界检测辅助任务，即本文方法 JEBD-NER，相较于 Bert-Self-Att-CRF 精准率提升了 1.44%，召回率提升了 0.57%，F1 值提升了 1.03%。综上可知，在添加边界检测辅助任务后，3 个基线模型实验效果均有不同程度的提升，进一步说明了本文提出的边界检测辅助任务对于命名实体识别模型是有效的，并且对不同模型的实验效果均有提升效果。

表 5　Twitter-2015 消融实验结果

Tab. 5　Ablation results on the Twitter-2015

方法名称	各类实体 F1 值/%				4 类实体综合结果/%		
	PER	LOC	ORG	MISC	P	R	F1
Bert-CRF	84.87	80.69	59.49	37.37	70.95	73.82	72.36
Bert-BiLSTM-CRF	84.77	80.58	60.41	39.30	71.06	74.73	72.84
Bert-Self-Att-CRF	84.28	79.92	61.15	38.14	70.66	74.52	72.54
Bert-CRF＋EDB	84.63	80.56	60.35	39.30	71.68	74.23	72.93
Bert-BiLSTM-CRF＋EDB	84.33	81.24	61.03	38.13	71.48	74.52	72.97
JEBD-NER	**84.89**	**81.33**	**62.59**	**40.84**	**72.10**	**75.09**	**73.57**

3.3 样例结果分析

如表 6 所示，每张样例表中分别有样例文本、真实标签、本文方法的预测结果以及 Bert-Self-Att 方法的预测结果。为直观对比，对预测结果添加底纹，绿色表示预测正确，红色表示预测错误。

表 6 样例预测结果对比表

Tab. 6 Comparison table of Sample prediction results

序 号	样例文本	真实标签	JEBD-NER	Bert-Self-Att
样例 1	Governor McCrory presents NC flag to Ruger CEO Mike Fifer at jobs announcement in Mayodan.	O B-PER O B-LOC O O B-ORG O B-PER I-PER O O O O B-LOC	O I-PER O B-LOC O O B-ORG O B-PER I-PER O O O O B-LOC	B-PER I-PER O B-LOC O O B-ORG O B-PER I-PER O O O O B-LOC
样例 2	Blackhawks Rozsival turns ＃Stars Travis Moen sideways in the second period.	B-ORG B-PER O B-ORG B-PER I-PER O O O O O O O	B-ORG B-PER O B-ORG B-PER I-PER O O O O O O O	B-ORG B-PER O O B-PER I-PER O O O O O O O
样例 3	MH17 report not released as it proves Russia not responsible for crash	B-MISC O O O O O O B-LOC O O O O	B-MISC O O O O O O B-LOC O O O O	O O O O O O O B-LOC O O O O

由表 6 可知，样例 1 中，对于“Governor”单词，本文方法成功预测为 O，而 Bert-Self-Att 方法误将“Governor”识别为人名实体，错误地拓宽了实体边界，表明在实体边界检测辅助任务的作用下，本文方法对实体边界有更好的判断能力。样例 2 中，对于“＃Stars”单词，本文方法成功预测为 B-ORG，正确识别出该词为组织实体，而 Bert-Self-Att 方法错误地将该词识别为 O，识别为非实体单词，表明在增加实体边界检测任务后，本文方法对实体的识别效果也有所增强。样例 3 中，对“MH17”单词，本文方法成功预测为 B-MISC，正确识别出该词为其他类实体，而 Bert-Self-Att 方法错误地将该词识别为 O，表明本文方法能更好识别出特殊类实体，说明在实体边界检测辅助任务的帮助下，模型对实体的识别能力也进一步得到提升。综合以上 3 组样例分析可知，在添加实体边界检测辅助任务后，方法不仅对实体边界的识别能力有所提升，而且对实体识别效果也同步变好。因此可以验证实体边界检测辅助任务对命名实体识别任务是有增益作用的。

3.4 损失权重 β 影响结果分析

损失权重 β 影响结果如图 2 所示。由图 2 可知，当实体边界检测损失权重 β 为 0.9 时，可获得最佳综合 Aug-F1 值为 73.57%，4 类实体各自的 F1 值也相对最优。当实体边界检测损失权重 β 过大或过小都无法得到最优实验结果。分析可知，当实体边界检测损失权重 β 过大时，实体边界检测任务将影响主任务命名实体识别的学习过程，导致实体识别效果变差；当实体边界检测损失权重 β 过小时，将无法起到应有的效果，对主任务命名实体识别效果提升没有作用。因此，选择合适的实体边界检测损失权重 β 也是实验中的重要环节。

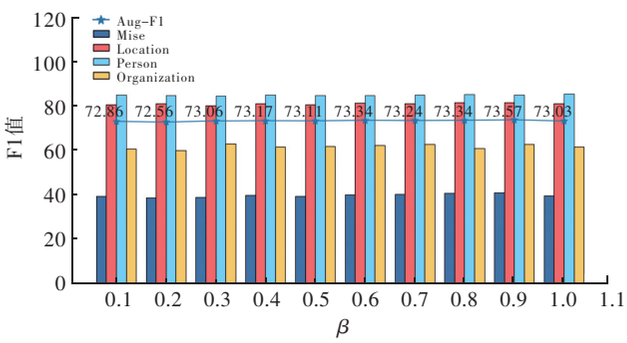

图 2 损失权重 β 对实验结果的影响

Fig. 2 Effect of β on experimental results

4 结 语

本文提出一种联合实体边界检测的命名实体识别方法，解决传统命名实体识别方法无法有效利用实体边界信息的问题。首先，使用 Bert 对原始文本进行词嵌入，获得词向量；其次，引入 Self-Attention 机制增强对上下文信息的利用能力，并引入实体边界检测辅助任务来提升模型对实体边界的判断能力，进而帮助模型增强实体识别效果；再次，对比了 JEBD-NER 模型与经典的命名实体识别基线模型的有效性，并对测试结果进行消融实验；最后，进行样例分析，分析了损失权重 β 对实体边界检测的影响。通过在 Twitter-2015 数据集上的实验证明了联合实体边界检测的命名实体识别方法是有效的。通过样例分析可知，所提方法不仅可以提升实体边界的识别能力，实体的识别效果也同步变好。同时，选择合适的损失权重 β 对于实体边界检测

也很重要。

所提方法虽然在当前数据集上的实体识别能力有一定提升，但其对“Misc”类实体无法很好识别，因为“Misc”类实体包含多种类型的实体。现有方法的实体识别能力仍有较大提升空间，下一步将探索利用迁移学习来提升模型对“Misc”类实体的识别能力，采用数据增强方法缓解数据受限问题。

致　谢

在此感谢清华大学智能技术与系统国家重点实验室徐华老师对本项工作给予的建设性意见及帮助。

参考文献/References：

[1] LI Jing，SUN Aixin，HAN Jianglei，et al. A survey on deep learning for named entity recognition[J]. IEEE Transactions on Knowledge and Data Engineering，2022，34(1)：50-70.

[2] YANG Xuemin，GAO Zhihong，LI Yongmin，et al. Bidirectional LSTM-CRF for biomedical named entity recognition[C]//2018 14th International Conference on Natural Computation，Fuzzy Systems and Knowledge Discovery (ICNC-FSKD). Huangshan：IEEE，2018：239-242.

[3] LAFFERTY J，MCCALLUM A，PEREIRA F C N. Conditional random fields：Probabilistic models for segmenting and labeling sequence data[C]//Proceedings of the Eighteenth International Conference on Machine Learning. Williamstown：Morgan Kaufmann Publishers Inc.，2001：282-289.

[4] REI M. Semi-supervised multitask learning for sequence labeling[C]//Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1：Long Papers). Vancouver：Association for Computational Linguistics，2017：2121-2130.

[5] LIN B Y，LU W. Neural adaptation layers for cross-domain named entity recognition[C]// Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. Brussels：Association for Computational Linguistics，2018：2012-2022.

[6] YANG Yaosheng，CHEN Wenliang，LI Zhenghua，et al. Distantly supervised NER with partial annotation learning and reinforcement learning[C]//Proceedings of the 27th International Conference on Computational Linguistics. Santa Fe：Association for Computational Linguistics，2018：2159-2169.

[7] ZHOU J T，ZHANG H，JIN D，et al. Dual adversarial neural transfer for low-resource named entity recognition[C]//Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Florence：Association for Computational Linguistics，2019：3461-3471.

[8] DEVLIN J，CHANG M W，LEE K，et al. BERT：Pre-training of deep bidirectional transformers for language understanding[C]//Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics：Human Language Technologies，Volume 1 (Long and Short Papers). Minneapolis：Association for Computational Linguistics，2019：4171-4186.

[9] CUI Yiming，CHE Wanxiang，LIU Ting，et al. Pre-training with whole word masking for Chinese BERT[J]. IEEE/ACM Transactions on Audio，Speech，and Language Processing，2021，29：3504-3514.

[10] JOSHI M，CHEN Danqi，LIU Yinhan，et al. Spanbert：Improving pre-training by representing and predicting spans[J]. Transactions of the Association for Computational Linguistics，2020，8：64-77.

[11] DONG Li，YANG Nan，WANG Wenhui，et al. Unified language model pre-training for natural language understanding and generation [C]//Advances in Neural Information Processing Systems 32. Vancouver：Curran Associates，Inc.，2019：13042-13054.

[12] LU J，BATRA D，PARIKH D，et al. ViLBERT：Pretraining task-agnostic visiolinguistic representations for vision-and-language tasks [C]//Advances in Neural Information Processing Systems 32. Vancouver：Curran Associates，Inc.，2019：13-23.

[13] LIU Weijie，ZHOU Peng，ZHAO Zhe，et al. K-BERT：Enabling language representation with knowledge graph[C]//The Thirty-Fourth AAAI Conference on Artificial Intelligence. New York：AAAI Press，2020：2901-2908.

[14] ZHANG Yu，YANG Qiang. An overview of multi-task learning[J]. National Science Review，2018，5(1)：30-43.

[15] CHEN Z，BADRINARAYANAN V，LEE C Y，et al. Gradnorm：Gradient normalization for adaptive loss balancing in deep multitask networks[C]//Proceedings of the 35th International Conference on Machine Learning. Stockholmsmässan：PMLR，2018：794-803.

[16] LI Y J，CARAGEA C. Multi-task stance detection with sentiment and stance lexicons[C]//Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). Hong Kong：Association for Computational Linguistics，2019：6299-6305.

[17] 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 (Volume 1：Long Papers). Melbourne：Association for Computational Linguistics，2018：799-809.

[18] GREENBERG N，BANSAL T，VERGA P，et al. Marginal likelihood training of bilstm-crf for biomedical named entity recognition from disjoint label sets[C]//Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. Brussels：Association for Computational Linguistics，2018：2824-2829.

[19] ZHAO Sendong，LIU Ting，ZHAO Sicheng，et al. A neural multi-task learning framework to jointly model medical named entity recognition and normalization[C]//The Thirty-Third AAAI Conference on Artificial Intelligence. Honolulu：AAAI Press，2019：817-824.

[20] VASWANI A,SHAZEER N,PARMAR N,et al. Attention is all you need[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach, California:Curran Associates Inc. ,2017:6000-6010.

[21] ZHANG Qi,FU Jinlan,LIU Xiaoyu,et al. Adaptive co-attention network for named entity recognition in tweets[C]//Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence and Thirtieth Innovative Applications of Artificial Intelligence Conference and Eighth AAAI Symposium on Educational Advances in Artificial Intelligence. New Orleans:AAAI Press,2018:5674-5681.

[22] MA X Z,HOVY E. End-to-end sequence labeling via bi-directional LSTM-CNNs-CRF[C]// Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1:Long Papers). Berlin:Association for Computational Linguistics, 2016:1064-1074.

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

[24] 顾亦然,霍建霖,杨海根,等. 基于 BERT 的电机领域中文命名实体识别方法[J]. 计算机工程,2021,47(8):78-83.
GU Yiran,HUO Jianlin,YANG Haigen,et al. BERT-based Chinese named entity recognition method in motor field[J]. Computer Engineering,2021,47(8):78-83.

[25] AAKERBERG A,JOHANSEN A S,NASROLLAHI K,et al. Single-loss multi-task learning for improving semantic segmentation using super-resolution[C]//Computer Analysis of Images and Patterns. Cham:Springer,2021:403-411.

向本期载文的审稿专家致谢

本期《河北科技大学学报》共发表论文 11 篇,这些论文的发表是与有关专家的认真审读、细查资料、推敲分析、中肯评价分不开的。对此,本刊编辑部特向这些专家表示敬意,对他们的辛勤劳动表示感谢。

本期载文的审稿专家名单如下(按姓名的汉语拼音字母顺序排列):

白志民　毕　赛　丁万昱　高慧媛　古双喜　郝　喆
何　辉　蒋　凡　李东武　李铁军　廉保旺　马履中
孙晓东　王　顶　王建秀　王晓远　汪超群　吴亚光
肖福仁　杨金显　湛永钟　赵东升　张付祥　张立新
张　宁　张文君　郑建勇

(本刊编辑部)