基于词序列的中文实体识别方法

刘　鑫，姚俊萍，李晓军，郭　毅，李　雪
（火箭军工程大学，陕西 西安 710025）

摘　要：针对现有词汇增强中文命名实体识别方法存在的引入干扰词汇、无法处理词汇信息冲突以及词汇信息利用不充分的问题，提出了一种基于词序列的中文命名实体识别方法。通过构建词序列，剔除干扰词汇，优化了词序列潜在词；利用位置编码技术融合潜在词位置信息，基于词序列注意力机制，解决词汇冲突问题。实验表明：该模型在 Resume 数据集、MSRA 数据集、OntoNotes 数据集、Weibo 数据集分别取得 96.38、95.77、82.23、71.25 的 F1 值，结果优于其他主流方法。

关键词：中文命名实体识别；词汇增强；BERT；注意力机制；位置编码
中图分类号：TP391　　**文献标志码：**A　　**文章编号：**2097-2741(2024)06-044-08

Chinese Named Entity Recognition Based on Lexicon Sequence

LIU Xin, YAO Junping, LI Xiaojun, GUO Yi, LI Xue
(Rocket Force University of Engineering, Xi'an 710025, Shaanxi)

ABSTRACT: In order to address the problems of interference words, inability to handle lexicon information conflicts, and insufficient utilization of lexicon information in existing lexicon enhanced Chinese named entity recognition (NER) methods, an NER based on lexicon sequence was proposed. Firstly, interference words were removed by constructing a lexicon sequence, which can optimize the potential words in the lexicon sequence. Secondly, the position information of potential words can be fused by utilizing positional encoding. Finally, lexicon conflicts were solved by processing positional encoding infused protential location information with the attention mechanism of lexicon sequence. Experimental results showed that F1 scores achieved by the proposed model on dataset Resume, MSRA, OntoNotes and Weibo were 96.38, 95.77, 82.23 and 71.25 respectively, which outperformed other mainstream methods.

KEYWORDS: Chinese named entity recognition; lexicon enhancement; BERT; attention mechanism; positional encoding

命名实体识别（Named Entity Recognition，NER）是自然语言处理中的一项重要任务，最早在 1987 年的信息理解会议（Message Understanding Conference，MUC）上作为实体关系分类的子任务被提出。NER 旨在从自然语言文本中识别出具有特定意义的实体，并准确识别其类型，主要包括人名、组织名、地名等。NER 的主要目标是确定实体的边界和类型，从非结构化文本中提取

收稿日期：2024-07-08
基金项目：火箭军工程大学青年项目（2022QN-S008）；军队科研资助项目
第一作者：刘鑫（1991—），女，讲师，主要从事自然语言处理研究。E-mail：liuxinalice@163.com

引用格式：刘鑫，姚俊萍，李晓军，等. 基于词序的中文实体识别方法[J]. 火箭军工程大学学报，2025，38（6）：44-51.

有价值的信息，应用于许多自然语言处理（Natural Language Processing，NLP）下游任务中，如信息检索、知识图谱、问答系统、舆情分析、生物医疗、推荐系统等。

由于中文文本不具有明显的词汇分割特性，与英文命名实体识别任务相比，中文命名实体识别（Chinese Named Entity Recognition，CNER）方法[1]通常是先使用现有的中文分词系统（Chinese Word Segmentation，CWS）对文本进行分词，然后利用序列标注模型对分词结果进行标注。然而，中文构词灵活、语义多样化，现有CWS系统不可避免地产生实体错误分割，导致实体边界检测和实体类别打标错误。为了避免实体分割错误引入误差干扰，大多数CNER模型都是基于字。字级别命名实体识别方法取消了分词步骤，直接按字切分语句，有效解决了分词错误传播问题。研究表明，针对中文命名实体识别任务，基于字的方法要显著优于基于词的方法，但字级别命名实体识别依旧面临词汇及词汇边界信息稀疏的挑战。为解决这一问题，研究者开始探索在基于字的NER模型中引入词汇信息，即词汇增强，以提升中文实体识别的性能。

近年来，研究者们提出了多种词汇增强方法，根据融合词汇信息的方式不同，可将词汇增强的中文实体识别方法分为4类，分别是基于动态架构、基于自适应嵌入、基于图神经网络和基于预训练模型的词汇增强方法。基于动态架构的方法，是将词汇信息注入到特殊的模型结构中，包括Lattice LSTM（Lattice Long Short-Term Memory）[2]、LR-CNN（Lexicon Rethinking Convolutional Neural Networks）[3]、FLAT（Flat-Lattice Transformer）[4]。Lattice LSTM首次引入词汇来处理中文命名实体识别任务，通过使用格子结构将潜在词整合到基于字的LSTM网络模型中。LR-CNN利用动态CNN结构捕获词汇信息，并采用Rethinking机制更新潜在词的权重分配比，有效解决了Lattice LSTM模型无法有效处理词汇信息冲突的问题。FLAT将Lattice结构从一个有向无环图展平为Flat-Lattice Transformer结构，并对Transformer模型中的传统位置编码方式进行改进，有效保留了位置信息，同时Transformer模型中的attention机制不但可以建立远程依赖关系，而且允许每个字与其自匹配词进行交互，解决了Lattice LSTM模型无法充分捕获中文文本语义信息和不能并行处理问题。基于自适应嵌入的方法，是将词汇信息融合到模型的嵌入层，包括WCLSTM（Word Character LSTM）[5]、Simple-Lexicon[6]。WC-LSTM通过给每个字添加潜在词信息，解决了Lattice LSTM模型训练时无法并行化的问题。Simple-Lexicon减少了复杂的网络层次，使得模型易于实现和训练。基于图神经网络的方法，利用图结构融合词汇信息，包括CGN（Collaborative Graph Network）[7]、LGN（Lexicon-based Graph Neural Network）[8]。CGN率先将图注意力神经网络（Graph Attention Network，GAT）[9]和自动构建语义图引入CNER任务，利用图结构获取多角度全方位词汇信息，不仅提升了CNER任务的准确率，还大大降低了时间成本。LGN网络结构将CNER任务转化为图节点分类问题，利用词汇信息丰富局部语义信息，并添加全局节点捕获全局语义信息和长距离依赖，通过不断地递归聚合实现节点与连接边的信息更新，提升了模型性能。基于预训练模型的方法，是通过BERT（Bidirectional Encoder Representations from Transformers）预训练模型增强词汇，包括Lex-BERT（Lexicon-Enhanced BERT）[10]、LEBERT（Lexicon Enhanced BERT）[11]。Lex-BERT通过整合外部词典信息来增强实体识别能力。LEBERT在BERT的基础上进行改进，引入Lexicon Adapter层，构建字符-词汇对（Char-Words Pair）序列，将潜在词直接融合到BERT的中间层中，从而提高对实体边界和类型的识别性能。

上述模型中，LEBERT模型在多项中文NER任务中取得了最先进的结果[12-15]。然而，LEBERT模型仍存在一些不足。首先，LEBERT模型为了发现和引入更全面的潜在词，针对每个字的自匹配词汇不可避免存在干扰词汇。以“南京市长江大桥”为例，其在LEBERT模型中的字词匹配示意如图1所示。可以看出，结合词汇边界和上下文信息，应该对“江大”进行过滤。其次，LEBERT模型无法处理词汇信息冲突问题，虽然采用注意力机制能够给不同词汇分配不同权重系数，但没有考虑词汇之间可能存在冲突情况。“南京市长”和“长江大桥”，两个词汇的语义信息不可能同时需要引入。最后，LEBERT模型没有区分字与词汇之间的位置关系，导致词汇边界信息利用不充分。

针对这些问题，本文提出一种基于词序列的

中文实体识别方法 SLEBERT（Sequential Lexicon Enhanced BERT）。通过构建词序列，剔除干扰词汇，并使用位置编码技术，有效利用词汇边界信息和语义信息。最后，通过利用词序列注意力机制方法，有效解决词汇信息冲突问题。在4个公开的中文NER数据集上进行多次实验，验证了该方法的有效性。

1 基于词序列的中文实体识别方法

1.1 模型框架

首先，构建词序列，基于词序列剔除干扰词汇，优化字词匹配的潜在词。然后，利用位置编码技术，融入潜在词的位置信息。最后，基于词序列注意力机制，解决词汇冲突问题。图2为模型的整体框架。

1.2 构建词序列

针对中文实体识别任务，基于字的NER模型无法拥有词汇的边界信息和语义信息，影响判定实体的边界和类型。基于词的NER模型可以捕捉到词与词之间的语义关系，但存在分词错误，且在处理不规则词和新词时比较困难。将基于字的模型和基于词的模型结合起来，在基于字的模型中引入词汇信息，即词汇增强，可以使模型具有较好的鲁棒性和识别精度。虽然词汇增强的中文实体识别方法在性能方面取得了显著成果[16-20]，但仍存在一些挑战。如引入的词汇信息存在干扰词汇、无法解决词汇冲突问题以及词汇的语义信息和边界信息利用不充分等。因此，本文提出了构建词序列，基于构建的词序列完成词汇增强，不仅充分利用了词汇边界信息，同时也剔除了干扰词汇，提高了潜在词质量。而且，构建词序列过程中是通过词汇树发现和引入词汇，不需要依赖分词工具，从而避免了分词错误对实体识别的影响。构建词序列的具体过程可以用算法1描述。

算法1 构建词序列

输入：词汇树 $\boldsymbol{T}$，中文句子的字序列 $\boldsymbol{C}=(c_1, c_2, \cdots, c_n)$

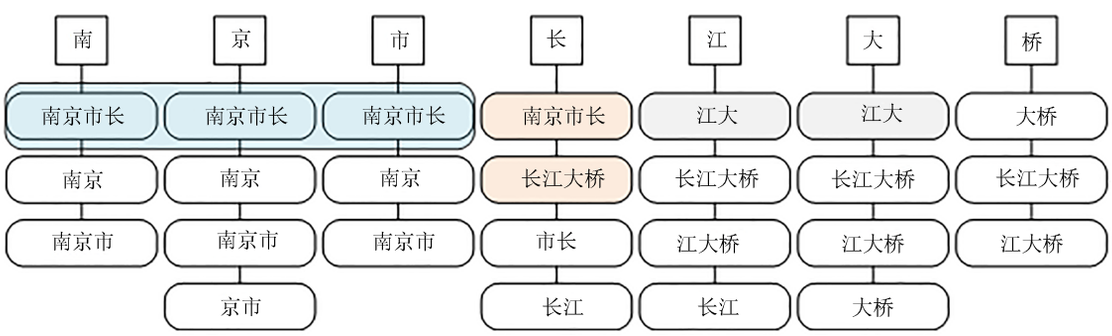

图1 LEBERT模型中的字词匹配示意

Fig. 1 Character-word pair in LEBERT

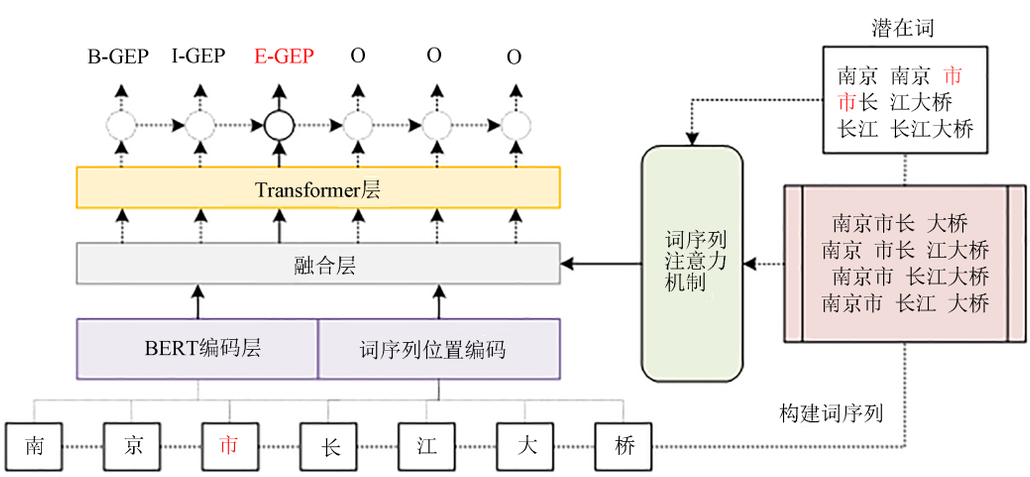

图2 基于词序列的中文实体识别方法整体框架

Fig. 2 Architecture of SLEBERT

输出：k组词序列$LS=(ls_1, ls_2, \cdots, ls_n)$

a）def DFS（index, current_sequence）：

i. if index >= k:

1. LS.append（current_sequence）

2. return

ii. if not LC[index]:

1.DFS（index+1, current_sequence）

iii. else:

1. for word in LC[index]:

1）current_sequence. append（word）

2）DFS（index + len（word）, current_sequence）

3）current_sequence.pop（） //移除word实现回溯

b）LC=[] //每个字的首字词汇序列

c）for character in C:

i. potential_word = buildBeginWithWordSet（character, T）

ii. LC.append（potential_word）

d）LS=[]

e）for index in range（k）: //每个字递归调用DFS

i. DFS（index, []）

f）print（LS）//输出构建的所有词序列

丰富的词汇信息可能带来词序列规模较大，且词序列之间存在冗余信息。以“南京市长江大桥”为例，其词序列构建及优化示例如图3所示。可以看出，词序列[南京，市长，江大桥]包含词序列[市长，江大桥]。此外，相比词序列[南京市，江大]，词序列[南京市，长江大桥]语义更准确。在保证不丢失上下文信息的情况下，通过计算每个序列的覆盖率，完成词序列的语义优化。覆盖率计算公式为：

$$cov_i = \frac{\sum_{j=0}^{m} \text{len}(lc_{ij})}{\text{len}(S)}$$

式中：S表示输入的整个句子；lc_{ij}表示句子中第i组序列中以第j个字为首的词汇；m表示每组序列中的词汇总数；cov_i表示第i组序列的覆盖率。词序列优化的详细过程可以用算法2描述。

算法2 优化词序列

输入：k组词序列$LS=(ls_1, ls_2, \cdots, ls_n)$

输出：top组词序列$LS=(ls_1, ls_2, \cdots, ls_{top})$

a）cov = {}

b）for index, sequence in enumerate（LS）：

i. cov = sum（len（word）for word in sequence） //计算每组序列的覆盖率

ii. cov/= n

iii. cov[index] = cov

c）cov = dict（sorted（covs.items（）, key=lambda item:item[1] , reverse=True）） //按覆盖率降序排序

d）array = []

e）for rank,index in enumerate（cov.keys（））：

i. array.append（LS[index]）

ii. if len（array）>top: //取前TOP组词序

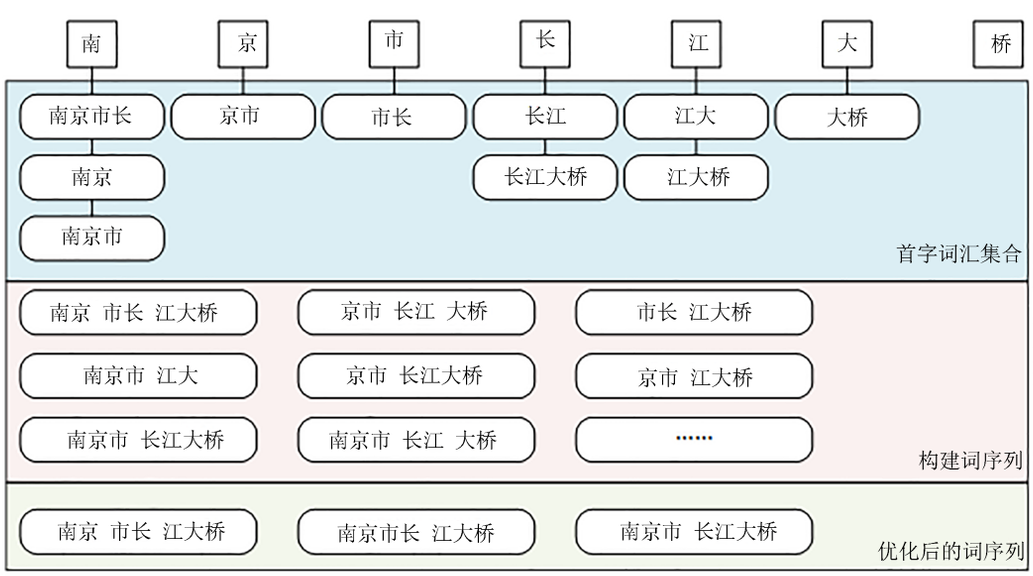

图3 词序列构建及优化示例

Fig. 3 Example of lexicon sequence construction and optimization

列 1. break

f）***LS***= array

g）print（***LS***）#输出优化后的词序列

1.3 融入位置编码

考虑到字词匹配的潜在词中，同一词汇对不同位置的字的边界贡献各不相同，使用位置编码技术，为每个潜在词添加一个额外的向量，用来标注字在句子中的相对位置信息，以便模型能够理解潜在词对不同字的边界信息。具体计算过程如下：

$$X_{ij}=e^{w}(LS_{ij})$$
$$p_i=\cos(i/1\,000^{2i/d_e})$$
$$q_i=\sin(i/1\,000^{2i/d_e})$$
$$\boldsymbol{P}_i=p_i+q_i$$
$$x_{ij}^{'}=x_{ij}+\boldsymbol{P}_i$$

式中：LS_{ij}表示第i个字基于第j组词序列获取的潜在词；e^{w}表示在预训练向量表中查找的词向量；d_e表示词向量大小；i表示句子中的第i个字，也表示字的位置；$\boldsymbol{P}_i$表示每个位置编码向量，是正弦和余弦函数的组合。为了对齐词向量和BERT模型中Transformer隐藏层输出，对词向量使用了非线性变换，得到每个词汇的隐藏向量$\boldsymbol{v}_{ij}^{h}$：

$$\boldsymbol{v}_{ij}^{h}=\boldsymbol{W}_2(\tanh(\boldsymbol{W}_1x_{ij}^{'}+b_1))+b_2$$

式中：$\boldsymbol{W}_1$是一个维度$d_h\times d_e$的矩阵；$\boldsymbol{W}_2$是一个$d_h\times d_h$维度矩阵；d_h表示BERT隐藏层大小。

1.4 引入词序列注意力机制

每个字与多个词汇配对，每个词汇对应一组词序列。考虑到具体语境中不同词序列的贡献各不相同。例如，“南京市长江大桥”，词序列“南京市，长江大桥”贡献应该最高，因为最接近真实分割。此外，考虑到词汇之间可能存在的冲突问题，例如“南京市长”和“长江大桥”的引入存在冲突，意味着如果当中有一个词汇接近真实语义，即词汇增强的注意力权重较高，则另外一个词汇增强的注意力权重应该很低。为了能够在词汇集合中挑选出最相关的词汇，引入词序列注意力机制，如图4所示。

具体来说，使用$\boldsymbol{A}=(a_1,a_2,\cdots,a_j)$表示权重向量，$a_j$表示第$j$组词序列的注意力权重。首先通过加权平均求得每组词序列向量，然后采用自注意力机制计算每组序列的权重值。具体计算过程如下：

$$s_j=\frac{1}{n}\sum_{i=1}^{n}x_{ij}^{'}$$
$$Q=s_jW_Q$$
$$K=s_jW_K$$
$$a_j=\mathrm{softmax}\left(\frac{QK^{\mathrm{T}}}{\sqrt{d_e}}\right)$$

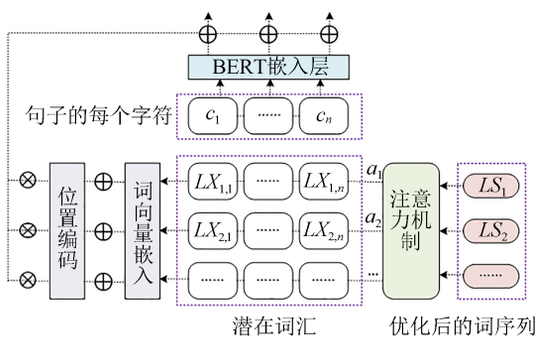

图 4 基于词序列实现词汇增强

Fig. 4 Lexicon enhancement based on lexicon sequence

最后，通过加权词汇隐藏向量得到$\boldsymbol{z}_i^{h}$，与字向量隐含向量$\boldsymbol{h}_i^{c}$关联，实现词汇信息注入。使用$H^{l}=(h_1^{l},h_2^{l},\cdots,h_n^{l})$表示字级别的BERT模型中第$l$层Transformer的输出。$\tilde{\boldsymbol{h}}_i^{l}$表示融入词汇信息后的$l$层的向量输出。

$$\boldsymbol{z}_i^{h}=\sum_{j=1}^{TOP}a_j\boldsymbol{v}_{ij}^{h}$$
$$\tilde{\boldsymbol{h}}_i^{l}=\tilde{\boldsymbol{h}}_i^{l}+\boldsymbol{z}_i^{h}$$

模型最后的解码模块采用CRF层来捕获连续标签之间的依赖关系，生成对应的标签序列。给定一个句子的字序列

$$C=(c_1,c_2,\cdots,c_n)$$

字序列对应的真实标签序列Y为

$$Y=(y_1,y_2,\cdots,y_n)$$

在经过所有Transformer层之后得到最后一层的隐藏输出$H^{l}=(h_1^{l},h_2^{l},\cdots,h_n^{l})$，则标签的概率计算为

$$\boldsymbol{O}=W_0H_L+b_0$$
$$p(y|c)=\frac{\exp\left(\sum_i(\boldsymbol{O}_{i,y_i}+\boldsymbol{T}_{y_{i-1},y_i})\right)}{\sum_{\tilde{y}}\exp\left(\sum_i(\boldsymbol{O}_{i,\tilde{y}_i}+\boldsymbol{T}_{\tilde{y}_{i-1},\tilde{y}_i})\right)}$$

式中：W_0和b_0是矩阵参数；$\boldsymbol{T}_{\tilde{y}_{i-1},\tilde{y}_i}$、$\boldsymbol{T}_{y_{i-1},y_i}$、$\boldsymbol{O}_{i,y_i}$、$\boldsymbol{O}_{i,\tilde{y}_i}$分别为$\boldsymbol{T}$、$\boldsymbol{O}$的子项；$\boldsymbol{T}$为分数矩阵$\boldsymbol{O}$的转置；$\tilde{y}$表示所有可能的标签序列。

2 实验与评估

基于4个公开数据集，对5个主流算法进行对

比实验，验证本文方法的有效性，并通过消融实验检验SLEBERT模型增强词汇的能力。

2.1 数据集

4个公开的中文命名NER数据集分别是OntoNotes、MSRA、Weibo和Resume，各数据集信息如表1所示。OntoNotes数据集是一个大规模的多语言语料库，包含人名、地名、组织机构、日期、货币、时间、数量等18种不同的实体类型。MSRA数据集来自新闻领域，由中国微软亚洲研究院发布，共包括人名、组织、组织机构等3种实体类型。Weibo数据集来自社交媒体领域，数据集规模较大，具有较高的噪声和语言变异性，分为人名、地名、组织机构、地缘政治等4种实体类型。Resume数据集来自新浪财经的中文简历数据，分为人名、地名、组织机构、国家地区、职业、种族、教育等8种不同类型的实体。

表1 数据集统计信息

Tab. 1 Statistics of dataset

数据集	数据类型	训练集/万个	验证集/万个	测试集/万个
OntoNotes	Sentence	1.57	0.43	0.43
	Char	29.19	20.05	20.81
MSRA	Sentence	4.64	0.44	0.44
	Char	216.99	17.26	17.26
Weibo	Sentence	0.14	0.027	0.027
	Char	7.38	1.45	1.48
Resume	Sentence	0.38	0.046	0.048
	Char	12.41	1.39	1.51

2.2 实验设置

本文选用召回率R、精确率P、以及F1（$F1=\frac{2RP}{R+P}$）作为评测指标，用以衡量模型的好坏。实验中，字向量输入使用Song[21]等训练的200维预训练词嵌入，词汇树是预训练词嵌入的词汇表。在BERT模型的第1层和第2层Transformer之间应用了基于词序列的词汇增强，并在训练过程中对BERT进行微调。采用了Adam 算法优化参数，初始学习率设置为1×10^{-5}，epoch设置为20，Weibo和Resume数据集batch size设置为20，OntoNote和MSRA数据集batch size设置为10。并使用dropout算法防止模型过拟合。

2.3 实验结果与分析

本文采用Lattice LSTM、LR-CNN、CGN、NFLAT、LEBERT词汇增强模型作为对比算法进行对比实验，以上模型分别使用了LSTM、CNN、图神经网络、Transformer、BERT等方法实现，覆盖了几种常用的特征提取技术。

2.3.1 对比实验

表2展示了本文算法SLEBERT与对比算法的实验结果。结果表明，SLEBERT的性能指标明显优于其他算法。SLEBERT在OntoNotes数据集、MSRA数据集、Weibo数据集和Resume数据集上F1值分别提升0.2%、0.07%、0.5%和0.3%。

2.3.2 消融实验

为了验证文中算法SLEBERT各个部分的贡献，在上述4个数据集上进行消融实验，实验结果如表3所示，可以看出，在OntoNotes数据集、Weibo数据集和Resume数据集上，没有使用位置编码和注意力机制，对比LEBERT模型，F1分别提高了0.18%、0.31%和0.16%。加入位置编码模块后，在MSRA数据集和Weibo数据集上，F1分

表2 本文算法与对比算法在各数据集上的对比结果

Tab. 2 Comparison results of the proposed algorithm and comparative algorithms on different datasets %

数据集	OntoNotes			MSRA			Weibo			Resume		
模型	P	R	F1	P	R	F1	P	R	F1	P	R	F1
Lattice-LSTM	76.35	71.56	73.88	93.57	92.79	93.18	53.04	62.25	58.79	94.81	94.11	94.46
LR-CNN	76.40	72.60	74.45	94.50	92.93	93.71	57.14	66.67	59.92	95.37	94.84	95.11
CGN	76.27	72.74	74.46	94.01	92.93	93.63	56.45	68.32	65.18	94.27	94.59	94.43
NFLAT	75.17	79.37	77.21	94.92	94.19	94.55	59.10	63.76	61.94	95.63	95.52	95.58
LEBERT	—	—	82.08	—	—	95.70	—	—	70.75	—	—	96.08
SLEBERT	**81.94**	**82.53**	**82.23**	**95.88**	**95.67**	**95.77**	**73.83**	**68.85**	**71.25**	**96.32**	**96.44**	**96.38**

别提高了0.07%、0.17%。加入词序列注意力机制模块后，在Weibo数据集和Resume数据集上F1分别提高了0.23%、0.18%。

从消融实验结果可以看出：1）使用基于词序列的词汇增强与LEBERT相比，能够剔除干扰词汇，更充分利用词汇语义信息和边界信息；2）使用位置编码，能够考虑同一词汇对不同位置字的影响程度，进而提高模型性能；3）词序列注意力机制解决了词汇冲突问题，能够更有针对性地突出相关词汇，提升模型准确性。

2.3.3 不同词序列TOP值的性能对比

为了验证SLEBERT模型在不同词序列TOP值下的有效性和鲁棒性，分别对4个公开数据集按照不同阈值进行验证，实验结果如表4所示。可以看出，随着词序列TOP值的增加，句子中每个字匹配的潜在词数量也在不断增加，SLEBERT模型可以更好地增强词汇信息。考虑到LEBERT模型中每个字匹配的潜在词最多3个，因此SLEBERT模型的F1值在TOP=3以后基本稳定。

表3 对比LEBERT实验和消融实验F1结果

Tab. 3 Comparison between the F1 results of LEBERT experiment and ablation experiment %

模型	OntoNotes	MSRA	Weibo	Resume
LEBERT	82.08	95.70	70.75	96.08
-位置编码 -注意力机制	82.23	95.70	70.97	96.21
-注意力机制	82.23	95.77	71.09	96.21
SLEBERT	82.23	95.77	71.25	96.38

表4 不同词序列TOP值的对比实验

Tab. 4 Comparative experiment results of TOP values of different lexicon sequences %

TOP值	OntoNotes	MSRA	Weibo	Resume
1	82.07	95.48	69.81	96.21
3	82.23	95.77	71.25	96.38
5	82.23	95.77	71.25	96.38
10	82.23	95.77	71.25	96.38

3 结论

针对现有词汇增强中文命名实体识别方法存在的问题，本文提出了一种基于词序列的中文实体识别方法。通过构建词序列，结合位置编码技术和词序列注意力机制的方法，有效利用词汇边界信息和语义信息，解决了词汇信息冲突问题。在4个公开的中文NER数据集上进行试验，结果表明：SLEBERT的性能指标明显优于其他主流算法，证明了SLEBERT方法对中文实体识别任务的有效性。但SLEBERT的泛化能力还需在不同领域的中文NER数据集上进行对比和验证，未来工作将进一步评估SLEBERT在不同领域数据集上的泛化能力、计算成本以及效率。

参考文献：

[1] ZHU P，CHENG D W，YANG F Z，et al. Improving Chinese named entity recognition by large-scale syntactic dependency graph[J]. IEEE / ACM Transactions on Audio，Speech，and Language Processing，2022，30：979-991.

[2] ZHANG Y，YANG J. Chinese NER using lattice LSTM[C]// Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics. Stroudsburg，PA，USA：ACL，2018：1554-1564.

[3] GUI T，MA R T，ZHANG Q，et al. CNN-based Chinese NER with lexicon rethinking[C]// Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence. California，USA：IJCAI，2019：4982-4988.

[4] LI X N，YAN H，QIU X P，et al. FLAT：Chinese NER using flat-lattice transformer[C]// Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. Stroudsburg，PA，USA：ACL，2020：6836-6842.

[5] LIU W，XU T G，XU Q H，et al. An encoding strategy based word-character LSTM for Chinese NER[C]// Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics：Human Language Technologies. Stroudsburg，PA，USA：ACL，2019：2379-2389.

[6] MA R T，PENG M L，ZHANG Q，et al. Simplify the usage of lexicon in Chinese NER[C]// Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. Stroudsburg，PA，USA：ACL，2020：5951-5960.

[7] SUI D B，CHEN Y B，LIU K，et al. Leverage lexical knowledge for Chinese named entity recognition via collaborative graph network[C]// Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th Inter-

national Joint Conference on Natural Language Processing. Stroudsburg, PA, USA: ACL, 2019: 3830-3840.

[8] GUI T, ZOU Y C, ZHANG Q, et al. A lexicon-based graph neural network for Chinese NER [C]// Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing. Stroudsburg, PA, USA: ACL, 2019: 1040-1050.

[9] VELIČKOVIĆ P, CUCURULL G, CASANOVA A, et al. Graph attention networks[EB / OL]. (2018-02-04)[2024-05-29]. http://arxiv.org/abs/1710.10903.

[10] ZHU W, CHEUNG D. Lex-BERT: Enhancing BERT based NER with lexicons[C]// Proceedings of the International Conference on Learning Representations (ICLR), 2021.

[11] LIU W, FU X Y, ZHANG Y, et al. Lexicon enhanced Chinese sequence labelling using BERT adapter[J]. Journal of Artificial Intelligence Research, 2021, 50(1): 123-134.

[12] 胡滨，耿天玉，邓赓，等. 基于知识蒸馏的高效生物医学命名实体识别模型[J]. 清华大学学报(自然科学版)，2021，61(9)：936-942.
HU B, GENG T Y, DENG G, et al. Faster biomedical named entity recognition based on knowledge distillation[J]. Journal of Tsinghua University (Science and Technology), 2021, 61(9): 936-942.

[13] 琚生根，李天宁，孙界平. 基于关联记忆网络的中文细粒度命名实体识别[J]. 软件学报，2021，32(8)：2545-2556.
JU S G, LI T N, SUN J P, Chinses fine-grained name entity recognition based on memory networks[J]. Journal of Software, 2021, 32(8): 2545-2556.

[14] 王庆人，王银子，仲红，等. 面向中文的字词组合序列实体识别方法[J]. 清华大学学报(自然科学版)，2023，63(9)：1326-1338.
WANG Q R, WANG Y Z, ZHONG H, et al. Chinese-oriented entity recognition method of character vocabulary combination sequence[J]. Journal of Tsinghua University (Science and Technology), 2023, 63(9): 1326-1338.

[15] DONG C H, ZHANG J J, ZONG C Q, et al. Character-based LSTM-CRF with radical-level features for Chinese named entity recognition[C]// 5th CCF Conference on Natural Language Processing and Chinese Computing. Beijing, China: CCF, 2016: 239-250.

[16] WU S, SONG X N, FENG Z H, et al. NFLAT: Non-flat-lattice transformer for Chinese named entity recognition [EB / OL]. (2022-11-14)[2024-09-12]. https://arxiv.org/abs/2205.05832.

[17] DEVLIN J, CHANG M W, LEE K, et al. BERT: pre-training of deep bidirectional transformers for language understanding[C]// Proceedings of Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Stroudsburg, PA, USA: ACL, 2019: 4171-4186.

[18] 李妮，关焕梅，杨飘，等. 基于BERT-IDCNN-CRF的中文命名实体识别方法[J]. 山东大学学报(理学版)，2020，55(1)：102-109.
LI N, GUAN H M, YANG P, et al. BERT-IDCNN-CRF for named entity recognition in Chinese [J]. Journal of Shandong University (Natural Science), 2020, 55(1): 102-109.

[19] XIE T, YANG J N, LIU H. Chinese entity recognition based on BERT-BILSTM-CRF model [J/OL]. Journal of Artificial Intelligence Research, 2021, 10(2): 150-165[2024-03-12]. https://doi.org/10.1016/j.jar.2021.02.001.

[20] LONG K F, ZHAO H, SHAO Z Z, et al. Deep neural network with embedding fusion for Chinese named entity recognition[J]. ACM Transactions on Asian and Low-Resource Language Information Processing, 2023, 22(3): 1-16.

[21] SONG Y, SHI S M, LI J, et al. Directional skip-gram: Explicitly distinguishing left and right context for word embeddings[C]// Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Stroudsburg, PA, USA: ACL, 2018: 175 - 180.