第 38 卷　第 2 期
2023 年 4 月

天津科技大学学报
Journal of Tianjin University of Science & Technology

Vol. 38　No. 2
Apr. 2023

DOI:10.13364/j.issn.1672-6510.20220221

基于 LNBC 模型的中文命名实体识别

马永军，王　野
（天津科技大学人工智能学院，天津 300457）

摘　要：针对中文命名实体识别中融合词典信息准确率提升不足的问题，使用在模型内部融合词典信息的策略，并结合预训练语言模型 NEZHA 增强文本的嵌入表示，提出一种基于 LNBC（LE-NEZHA-BiLSTM-CRF）模型的中文命名实体识别方法. 首先通过词典树匹配所有潜在的词，然后采用面向中文理解的神经语境表征模型（NEZHA）进行融合嵌入表示，将训练得到的字词融合向量输入双向长短期记忆（BiLSTM）网络进行特征提取，获取长距离的语义信息，最后通过条件随机场（CRF）层降低错误标签输出的概率. 实验结果表明，该方法在 MSRA 数据集和 Resume 数据集中的 F_1 值分别为 95.71%和 96.11%，较其他对比模型均有提高.

关键词：命名实体识别；词典信息；双向长短期记忆网络；条件随机场
中图分类号：TP391　　**文献标志码**：A　　**文章编号**：1672-6510（2023）02-0050-06

Chinese Named Entity Recognition Based on LNBC Model

MA Yongjun，WANG Ye
（College of Artificial Intelligence，Tianjin University of Science & Technology，Tianjin 300457，China）

Abstract：Aiming at the problem of insufficient accuracy improvement of fusion dictionary information in Chinese named entity recognition，in this article，we propose a Chinese named entity recognition method based on LE-NEZHA-BilSTM-CRF（LNBC）model by using the strategy of fusion dictionary information inside the model and combining with the pre-trained language model NEZHA to enhance the embedded representation of text. In our proposed model，firstly，the dictionary tree is used to match all potential words，and then the neural context representation model for Chinese comprehension（NEZHA）is used for fusion and embedding representation. The trained word-word fusion vector is input into the bidirectional Long short-term Memory Network（BiLSTM）for feature extraction，and the long-distance semantic information is obtained. Finally，conditional random field（CRF） layer is used to reduce the probability of mislabeled output. The experimental results showed that the F_1 value of the proposed method in MSRA dataset and Resume dataset was 95.71% and 96.11%，respectively，which is higher than that of other comparison models.

Key words：named entity recognition；dictionary information；bidirectional long short-term memory network；conditional random field

中文命名实体识别（NER）是自然语言处理的一个重要研究方向，具体指在大量文本中识别出具有特定意义的实体，比如文本中的人名、地名、机构名、日期、时间等. 命名实体识别[1]是人工智能的关键技术，它为以后的知识图谱构建、信息检索、机器翻译、智能问答等奠定了坚实的基础，实体抽取的结果将直接关系到后续自然语言处理的质量和效率[2]. 命名实体识别方法主要包括以下 3 类：基于规则和词典的方法、基于统计机器学习的方法、基于深度学习的方法. 与传统的统计机器学习方法相比，基于深度学习的方法具有更少的人工依赖和更好的泛化能力，目前，深度学习已成为命名实体识别的主要研究方向，并在命名实体识别任务中得到了广泛的应用.

在开始中文命名实体识别任务之前，一般要对中

收稿日期：2022–09–27；**修回日期**：2022–12–02
基金项目：天津市教委社会科学重大项目（2017JWZD19）
作者简介：马永军（1970—），男，吉林长春人，教授；**通信作者**：王　野，硕士研究生，wangye@mail.tust.edu.cn

文文本进行分词，将其转换为词级别的序列标注. 然而，分词的错误往往会带来错误的实体边界，这给中文命名实体识别带来了很多困难和挑战. 字符级别表示不需要对中文进行分词，可以避免分词错误带来的影响，相较于词级别表示，字符级别表示会获得更细腻的特征，不会出现未登录词的问题[3]. Yan 等[4]提出用 Transformer（变换器）模型建模字符级别的信息，依次通过编码层和解码层，最终获得了较好的识别效果.

字符级别表示会获得更细腻的特征，词级别表示能够获得更充分的语义信息，由此出现了很多字词融合表示的研究，这些研究兼顾了字符级别表示和词级别表示的优点. Zhang 等[5]使用字词融合表示的方法，提出晶格–长短期记忆（Lattice-LSTM）网络模型，通过使用晶格结构的长短期记忆（LSTM）网络将字词信息整合在一起，将不同的词信息与字符信息进行融合来提升实体的识别效果. Li 等[6]提出平面晶格变换器（FLAT）模型，引入了全新的位置编码方法来表示原有的晶格结构，无损地融合字词信息. 以上研究通过更改模型结构来巧妙地融合字词信息，但是也带来了推理速度变慢的问题.

在字词融合中引入实体词典，可以提高实体识别的准确率. 胡新棒等[7]通过实体词典融入先验知识，提高字词融合的表征能力，以减少未登录词对模型的影响，提高模型的迁移能力，与上述通过更改结构的模型相比，提高了模型的推理速度，减少了运算资源. 吴雅娟等[8]根据特定领域中实体种类繁多和边界模糊等特点，结合实体词典来提高实体识别效果，将词典中的词级信息和字符级信息在同等维度下进行拼接后输入模型的编码层，丰富了向量的表示. 胡婕等[9]提出将知识库信息添加到词典中，该知识库从百科知识图谱下载，其内部含有海量的数据并含有大量的实体信息，在输入编码器之前，将字嵌入向量、文本嵌入向量和位置向量相加作为输入的嵌入表示，最后通过解码器获得最优的序列标签. Liu 等[10]提出利用 Lexicon Adapter 的结构将词典信息融合到模型内部，通过内部融合可以学习更多的语义信息，获得更好的识别效果.

自 Lample 等[11]提出双向长短期记忆网络和条件随机场（CRF）结合的方法以来，该方法在命名实体识别领域得到广泛的应用，谢腾等[12]在此基础上提出结合预训练模型 BERT（变换器双向编码表示模型）的方法，利用动态词向量提高上下文信息的表征能力，解决一词多义的问题. 李军怀等[13]提出结合预训练模型轻量级变换器双向编码表示模型（ALBERT）的方法，在解决一词多义问题的同时加快了运算速度. 但对中文命名实体识别来说，大多数融合词典所使用的预训练模型都是基于英文语料进行训练，适合中文的模型较少，中文命名实体识别效果有待提高.

为了进一步提高中文命名实体识别的准确率，本文提出 LNBC（LE-NEZHA-BiLSTM-CRF）中文命名实体识别模型，在 Lexicon Adapter 的基础上，首先使用 LE-NEZHA（Lexicon-NEZHA）将实体词典的词向量通过词典适配器的方式嵌入面向中文理解的神经语境表征模型（NEZHA）内部进行融合，具体是指在各层 Transformer 之间融合词典信息，然后通过双向长短期记忆网络（BiLSTM）提取长距离的语义信息，避免局部上下文语义的特征缺失，最后通过条件随机场（CRF）合法性约束错误标签输出的概率，得到最终的识别结果.

1　LNBC 模型

本文提出的基于 LNBC 的中文命名实体识别模型具体分为 LE-NEZHA 层、BiLSTM 层和 CRF 层 3 个部分，该模型的结构如图 1 所示.

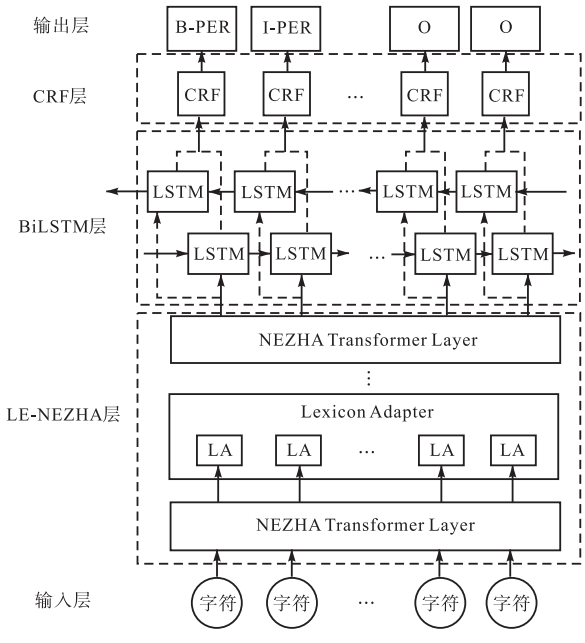

图 1　LE-NEZHA-BILSTM-CRF 模型的结构

Fig. 1　Structure of LE-NEZHA-BLSTM-CRF model

1.1　LE-NEZHA 层

NEZHA[14]模型是基于 Transformer 架构的预训练模型，它在 BERT 的基础上增加了相对位置编码，使模型能够更好地学习信息之间的传递；NEZHA 使

用更适合中文的全词掩码，在中文上的表现更加出色，并且使用混合精度训练和批量训练的分层自适应时刻优化器（LAMB）提升训练效率. LE-NEZHA 与 NEZHA 相比，首先从模型的输入来说，中文句子不再是以字符为最小单位，而是要通过词典树的结构被转换成字符和单词序列对表示，将字符和实体词典特征作为输入部分. 其次，在 Transformer 层之间附加了一个词典适配器，可以将词典知识有效地融入 NEZHA 内部.

1.1.1　词典树

中文句子通常被表示为一串字符序列，仅包含字符级别的特征. 为了利用词典信息，将字符序列扩展为字词对序列，将每个字符与分配给它的词进行配对，并将汉语句子转换成字词对序列；然后基于实体词典新建词典树，遍历整个中文句子中的所有字符序列，并与词典树进行匹配，得到所有潜在的词.

1.1.2　词典适配器

将词典适配器添加到 Transformer 层中，使词典信息在模型内部融合得更加充分，将字符和单词对输入到词典适配器，依次通过注意力层和归一化层，最后将加权后的词汇信息融入字符信息中. 词典适配器的结构如图 2 所示.

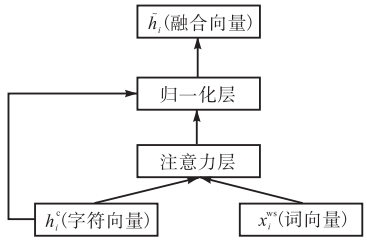

图 2　词典适配器的结构

Fig. 2　Structure of the lexicon adapter

图中词典适配器的输入表示为 $(\boldsymbol{h}_i^{\mathrm{c}}, x_i^{\mathrm{ws}})$，其中 $\boldsymbol{h}_i^{\mathrm{c}}$ 为字符向量，是 NEZHA 中某个 Transformer 层的输出，$x_i^{\mathrm{ws}} = \{x_{i1}^{\mathrm{w}}, x_{i2}^{\mathrm{w}}, ..., x_{im}^{\mathrm{w}}\}$ 是一组词嵌入的集合，为了能够将字符向量和词向量有效融合，对词向量信息进行非线性变换.

$$\boldsymbol{v}_{ij}^{\mathrm{w}} = \boldsymbol{W}_2[\tanh(\boldsymbol{W}_1 x_{ij}^{\mathrm{w}} + b_1)] + b_2 \tag{1}$$

式中：$\boldsymbol{v}_{ij}^{\mathrm{w}}$ 为通过非线性变化的词向量；b_1 和 b_2 都代表非线性变换缩放的偏置；$\boldsymbol{W}_1$ 是 $d_{\mathrm{c}} \times d_{\mathrm{w}}$ 的矩阵，d_{c} 为 BERT 中隐藏层的大小，d_{w} 为整个词典词嵌入的维度大小；$\boldsymbol{W}_2$ 是 $d_{\mathrm{c}} \times d_{\mathrm{c}}$ 的矩阵.

词向量通过注意力层，其中引入了字词对的注意力机制，分配给第 i 个字符的所有 $\boldsymbol{v}_{ij}^{\mathrm{w}}$ 的词汇信息表示为 $V_i = (\boldsymbol{v}_{i1}^{\mathrm{w}}, \cdots, \boldsymbol{v}_{im}^{\mathrm{w}})$，其中 m 是分配给单词的总数量，每个词的相似度（a_i）计算公式为

$$a_i = \mathrm{softmax}(\boldsymbol{h}_i^{\mathrm{c}} \boldsymbol{W}_{\mathrm{attention}} V_i^{\mathrm{T}}) \tag{2}$$

$$z_i^{\mathrm{w}} = \sum_{j=1}^{m} a_{ij} \boldsymbol{v}_{ij}^{\mathrm{w}} \tag{3}$$

其中：$\boldsymbol{W}_{\mathrm{attention}}$ 是双线性注意力的权重矩阵，z_i^{w} 是所有单词的加权总和.

最后将加权后的词向量和字符向量融合表示为

$$\tilde{h}_i = \boldsymbol{h}_i^{\mathrm{c}} + z_i^{\mathrm{w}} \tag{4}$$

LE-NEZHA 利用词典树和词典适配器等结构充分融合了字符信息和词典信息，将输出的融合向量作为 BiLSTM 模型的输入.

1.2　BiLSTM 层

LSTM 模型[15]经常被用于处理时序信息，可以有效解决循环神经网络（recurrent neural network，RNN）会产生梯度消失或爆炸的问题，通过门控机制能更好地提取长距离的语义信息，BiLSTM 由正向 LSTM 和反向 LSTM 组合而成，对输入进行前向计算和后向计算，得到不同的结果，通过向量拼接得到最终的隐藏层表示，最终会获得两个方向的上下文特征，能够学习符合上下文语境的语义信息，LSTM 的结构如图 3 所示，图中：“×”代表乘操作，“+”表示加操作，“σ”和“tanh”分别表示 sigmoid 和 tanh 激活函数. 使用双向长短期记忆模型可以更好地捕捉较长距离的依赖关系，可以更好地捕捉到双向的语义依赖，避免发生局部特征语义缺失的问题.

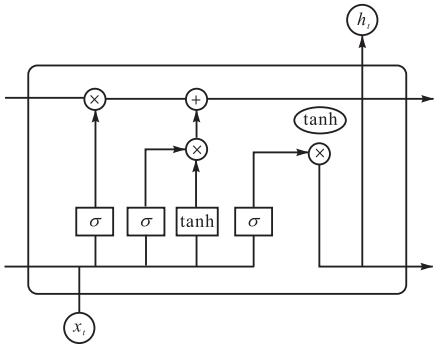

图 3　LSTM 的结构

Fig. 3　Structure of LSTM

LSTM 的计算过程可以概括为，对细胞状态中的信息进行选择性的遗忘和记忆，在后续时刻使有用的信息得以传递，而无用的信息被丢弃，并在每个时间步都会输出隐层状态. 其中遗忘、记忆与输出由通过上个时刻的隐层状态和当前输入计算出来的遗忘门、

记忆门与输出门来控制，各状态的计算公式为

$$f_t=\sigma(\boldsymbol{W}_f\cdot[h_{t-1},x_t]+\boldsymbol{b}_f) \tag{5}$$

$$i_t=\sigma(\boldsymbol{W}_i\cdot[h_{t-1},x_t]+\boldsymbol{b}_i) \tag{6}$$

$$\tilde{C}_t=\tanh(\boldsymbol{W}_c\cdot[h_{t-1},x_t]+\boldsymbol{b}_c) \tag{7}$$

$$C_t=f_t\odot C_{t-1}+i_t\odot\tilde{C}_t \tag{8}$$

$$o_t=\sigma(\boldsymbol{W}_o[h_{t-1},x_t]+\boldsymbol{b}_o) \tag{9}$$

$$h_t=o_t\odot\tanh(C_t) \tag{10}$$

其中：f_t 为遗忘门，i_t 为记忆门，$\tilde{C}_t$ 为临时的细胞状态，C_t 为当前时刻细胞状态，o_t 为输出门，h_t 为隐层状态，h_{t-1} 为前一个时刻的隐层状态，$\boldsymbol{W}$ 为权重矩阵，x_t 为当前时刻的输入词，$\boldsymbol{b}$ 为偏置向量.

1.3　CRF 层

条件随机场简称 CRF，是一种用于预测序列的判别式模型. 目前，线性链条件随机场[16]是使用最广泛的模型，该模型是在给定一组输入序列的条件下，求另一组输出序列的条件概率分布，与仅考虑局部最优的 softmax 函数相比，CRF 通过训练可以获得全局条件下的最优标注序列.

对每个训练样本 X，求出所有可能的序列标注 y 的得分 $S(X,y)$，然后对所有得分进行归一化.

$$S(X,y)=\sum_{i=0}^{n}A_{y_i,y_{i+1}}+\sum_{i=1}^{n}P_{i,y_i} \tag{11}$$

$$p(y|X)=\frac{\mathrm{e}^{s(X,y)}}{\sum_{\tilde{y}\in Y_X}\mathrm{e}^{s(X,\tilde{y})}} \tag{12}$$

其中：P_{i,y_i} 为 softmax 输出的第 i 个位置是 y_i 的概率，y_i 为从 y_i 到 y_{i+1} 的转移概率.

由于每个最大概率值对应的标签并不都是 softmax 输出的，因此当一个预测标签得分很高时，还需要考虑转移概率相加是否最大，避免输出错误标签. 其中分子上的 y 是正确的标注序列，取对数可以求得 y 在所有序列中所占的最大概率.

$$\log(p(y|X))=S(X,y)-\log\left(\sum_{\tilde{y}\in Y_X}\mathrm{e}^{s(X,\tilde{y})}\right) \tag{13}$$

优化的目标就是最大化式(13)(即真实标签对应的最大概率值)，训练完毕后进行预测时，根据训练好的参数求出所有可能的 y 序列对应的 S 得分，得到最佳的预测结果 y^*.

$$y^*=\underset{\tilde{y}\in Y_X}{\operatorname{argmax}}\,S(X,\tilde{y}) \tag{14}$$

2　实　验

本实验在公开新闻数据集 MSRA 和简历数据集 Resume 上进行，MSRA 数据集比 Resume 数据集规模更大、实体数目更多；融合词典采用腾讯词向量词典[17].

2.1　MSRA 数据集

微软亚洲研究院公开的 MSRA[18]数据集，是中文命名实体识别最常用的中文数据库之一. 该数据集中标注有人名(PER)、地名(LOC)、组织机构名(ORG)3 类实体. 采用 BIO 标注格式，对每个文本中的实体，“B”代表实体的开始字符，“I”代表实体的中间字符或结尾字符，“O”代表非实体. 实验中，训练集包含 41 728 句语料，验证集包含 4 636 句语料，测试集包含 4 365 句语料.

2.2　Resume 数据集

新浪财经网公开的中文 Resume 数据集，是根据简历数据筛选和标注的公开数据集，该数据集包含 1 000 余份简历摘要，其中标注有 8 类实体，分别为人名(NAME)、国籍(CONT)、籍贯(LOC)、种族(RACE)、专业(PRO)、学位(EDU)、机构(ORG)、职称(TITLE). 采用 BMEOS 标注格式，其中“B”代表实体的开始字符，“M”代表实体的中间字符，“E”代表实体的结束字符，“O”代表非实体，“S”代表单一字符的实体.

2.3　评价标准

本实验采用准确率(precision，P)、召回率(recall，R)和 F_1 值(F_1-score，F_1)作为评价标准，准确率是指预测为正确的样本数中实际为正确样本的概率，召回率是指实际为正确的样本中被识别为正确样本的概率，F_1 值则是结合了两者，一般被作为对模型整体评估的指标.

$$P=N_{\mathrm{TP}}/(N_{\mathrm{TP}}+N_{\mathrm{FP}})\times100\%$$

$$R=\frac{N_{\mathrm{TP}}}{N_{\mathrm{TP}}+N_{\mathrm{FN}}}\times100\%$$

$$F_1=\frac{2PR}{P+R}\times100\%$$

其中：N_{TP} 表示识别出的正确的实体数，N_{FP} 表示识别出的错误标记的实体数，N_{FN} 表示未识别出来的实体数.

2.4　实验过程

2.4.1　实验环境

实验硬件配置以及实验环境：GPU 为 GeForce RTX 2080 Ti，内存为 32 GB，操作系统为Ubunt 18.04 版本，Python 版本为 3.6，PyTorch 版本为 1.17.0.

2.4.2　参数设置

在本实验中，训练模型使用 NEZHA 的预训练权

重(base 版本),含有 12 个 Transformer 层、768 维隐藏层和 12 个头的注意力机制,最大文本序列长度选择 128,BiLSTM 隐藏层为 128,其他具体参数设置见表 1.

表 1　实验参数设置

Tab. 1　Experimental parameter setting

参数	详情
batch_size	32
epoch	16
dropout	0.4
learning rate	2×10^{-5}
词嵌入维度	200
融合词的最大个数	5

2.5　结果与分析

2.5.1　模型实验

模型在 MSRA 数据集上的识别结果见表 2.

表 2　模型在 MSRA 数据集上的识别结果

Tab. 2　Recognition results of the model on MSRA dataset

实体类型	P/%	R/%	F_1/%
全部实体	95.60	95.81	95.71
PER	96.98	97.52	97.25
LOC	97.09	95.83	96.46
ORG	91.07	93.91	92.47

由表 2 可知,模型对组织机构类型的实体识别的准确率相对较低,原因可能在于大量地名与组织机构名存在实体嵌套的情况,而且特定的组织机构简称比较多,也是影响识别效果的原因之一,对人名和地名的识别都有着较好的准确率.

模型在 Resume 数据集上的识别结果见表 3. 由表 3 可知,采用 LNBC 模型后,各类实体都取得了比较好的识别效果.

表 3　模型在 Resume 数据集上的识别结果

Tab. 3　Recognition results of the model on Resume dataset

实体类型	P/%	R/%	F_1/%
全部实体	95.70	96.53	96.11
NAME	99.10	98.72	98.90
CONT	94.29	96.03	95.15
LOC	97.67	99.26	98.46
RACE	93.33	93.33	93.33
PRO	94.12	99.74	96.84
EDU	96.33	99.06	97.67
ORG	93.77	95.03	94.40
TITLE	93.69	94.55	94.12

2.5.2　对比实验

为了验证模型的有效性,在 MSRA 数据集上分别和近几年相关研究的模型进行对比,具体对比结果见表 4.

表 4　MSRA 数据集对比实验结果

Tab. 4　MSRA dataset comparison experimental results

模型	P/%	R/%	F_1/%
Lattice + LSTM + CRF[5]	93.57	92.79	93.18
AKE[7]	94.72	93.76	94.24
KG + EntityBERT + CRF[9]	87.23	89.01	88.11
BERT + BiLSTM + CRF[12]	94.38	94.92	94.65
ALBERT + BGRU + CRF[13]	95.16	94.58	94.87
LNBC	95.60	95.81	95.71

由表 4 可知:为验证融入词典信息的有效性,设置了 BERT + BiLSTM + CRF、ALBERT + BGRU + CRF 和 LNBC 模型的对比实验,结果表明向模型融入词典信息提供先验知识可以有效地提高实体识别的效果. 为验证 LNBC 模型的有效性,分别与近几年主流融合词典信息的中文命名实体模型 Lattice + LSTM + CRF、AKE、KG + EntityBERT + CRF,在相同数据集上进行对比实验,由于 Lattice-LSTM-CRF 模型只是通过改造 LSTM 进行动态选择词汇和融合词典信息,效果不如在预训练模型中融合词典信息,本文方法相较于该模型 F_1 值提高了 2.53%. 相较于 AKE 模型和 KG + EntityBERT + CRF 模型,本文方法 F_1 值分别提高了 1.47%、7.60%,充分说明了本文模型融合的词典信息与字词向量拼接后送入模型相比,融合得更加充分,提供了更多的实体边界和词汇边界,准确率获得了大幅提高.

在 Resume 数据集上也进行相应的对比实验,对比结果见表 5. 实验结果表明,LNBC 模型相较于其他对比模型在自建数据集上的 F_1 值都有明显的提高,实体识别的准确率得到大幅增长,说明了本文模型的有效性.

表 5　Resume 数据集对比实验结果

Tab. 5　Resume dataset comparison experimental results

模型	P/%	R/%	F_1/%
BILSTM + CRF[19]	92.50	94.30	93.40
BERT + BILSTM + CRF	95.16	95.86	95.51
NEZHA + BILSTM + CRF	95.30	96.13	95.7
LNBC	95.70	96.53	96.11

3　结　语

本文提出一种基于 LNBC(LE-NEZHA-BiLSTM-CRF)模型的中文命名实体识别方法,解决了中文命名实体识别中融合词典信息准确率提升不足的问题. 使用预训练模型 NEZHA 作为文本的嵌入表示,并在模型内部融合词典信息后,将其输入到 BiLSTM

网络中进行特征提取，获取长距离的语义信息，最后通过 CRF 修正后输出标签结果，在两个不同的数据集上进行验证，F_1 值均高于对比模型，实验结果均证明了 LNBC 模型的有效性. 下一步的工作重点是继续优化模型的表现，进行食品营养领域中实体抽取的探索，为构建领域知识图谱打下基础.

参考文献：

[1] 陈曙东，欧阳小叶. 命名实体识别技术综述[J]. 无线电通信技术，2020，46(3)：251–260.

[2] 张栋，陈文亮. 基于上下文相关字向量的中文命名实体识别[J]. 计算机科学，2021，48(3)：233–238.

[3] 郑洪浩，宋旭晖，于洪涛，等. 基于深度学习的中文命名实体识别综述[J]. 信息工程大学学报，2021，22(5)：590–596.

[4] YAN H，DENG B C，LI X N，et al. TENER：adapting transformer encoder for named entity recognition[EB/OL]. (2019–12–10)[2022–09–20]. https://arxiv.org/pdf/1911.04474.pdf.

[5] ZHANG Y，YANG J. Chinese NER using lattice LSTM[EB/OL]. (2018–07–05)[2022–09–20]. https://arxiv.org/pdf/1805.02023.pdf.

[6] LI X N，YAN H，QIU X P，et al. FLAT：Chinese NER using flat-lattice transformer[EB/OL]. (2020–05–23)[2022–09–20]. https://arxiv.org/pdf/2004.11795.pdf.

[7] 胡新棒，于溆乔，李邵梅，等. 基于知识增强的中文命名实体识别[J]. 计算机工程，2021，47(11)：84–92.

[8] 吴雅娟，牛甲奎，解红涛，等. 基于词典与字向量融合的井控领域命名实体识别[J]. 海南大学学报(自然科学版)，2022，40(2)：125–133.

[9] 胡婕，胡燕，刘梦赤，等. 基于知识库实体增强 BERT 模型的中文命名实体识别[J]. 计算机应用，2022，42(9)：2680–2685.

[10] LIU W，FU X Y，ZHANG Y，et al. Lexicon enhanced Chinese sequence labeling using bert adapter[EB/OL]. (2021–12–26)[2022–09–20]. https://arxiv.org/pdf/2105.07148.pdf.

[11] LAMPLE G，BALLESTEROS M，SUBRAMANIAN S，et al. Neural architectures for named entity recognition[EB/OL]. (2016–04–06)[2022–09–20]. https://arxiv.org/abs/1603.01360.

[12] 谢腾，杨俊安，刘辉. 基于 BERT-BiLSTM-CRF 模型的中文实体识别[J]. 计算机系统应用，2020，29(7)：48–55.

[13] 李军怀，陈苗苗，王怀军，等. 基于 ALBERT-BGRU-CRF 的中文命名实体识别方法[J]. 计算机工程，2022，48(6)：89–94.

[14] WEI J Q，REN X Z，LI X G，et al. NEZHA：neural contextualized representation for Chinese language understanding[EB/OL]. (2021–11–19)[2022–09–20]. https://arxiv.org/pdf/1909.00204.pdf.

[15] HOCHREITER S，SCHMIDHUBER J. Long short-term memory[J]. Neural computation，1997，9(8)：1735–1780.

[16] 曹依依，周应华，申发海，等. 基于 CNN-CRF 的中文电子病历命名实体识别研究[J]. 重庆邮电大学学报(自然科学版)，2019，31(6)：869–875.

[17] 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，Volume 2(Short Papers). New Orleans：Association for Computational Linguistics，2018：175–180.

[18] LEVOW G A. The third international Chinese language processing bakeoff：word segmentation and named entity recognition[C]//Proceedings of the Fifth SIGHAN Workshop on Chinese Language Processing. 2006：108–117. Sydney：Association for Computational Linguistics，2006：108–117.

[19] 司逸晨，管有庆. 基于 Transformer 编码器的中文命名实体识别模型[J]. 计算机工程，2022，48(7)：66–72.

责任编辑：周建军