

文章编号: 2095-2163(2020)03-0021-06

中图分类号: TP391

文献标志码: A

# 基于 ERNIE 的命名实体识别

张 晓<sup>1</sup>, 李业刚<sup>1</sup>, 王 栋<sup>1</sup>, 史树敏<sup>2</sup>

(1 山东理工大学 计算机科学与技术学院, 山东 淄博 255049; 2 北京理工大学 计算机学院, 北京 100081)

**摘要:** 针对深度学习方法处理命名实体识别任务时,经典的词向量表示只能将其映射为单一向量,无法在上下文中表示出字的歧义性这一问题,提出 ERNIE-BiGRU-CRF 模型,该模型通过知识增强语义表示 ERNIE 预训练模型增强字的语义感知表示。引入多元数据知识生成语义向量,然后将字向量输入到 GRU 层提取特征,最后通过 CRF 层得到标签序列。实验结果表明,该模型在人民日报语料库中  $F_1$  值达到了 94.46%。

**关键词:** 命名实体识别; 知识增强语义表示; 门控循环单元网络; 条件随机场

## Named Entity Recognition based on ERNIE

ZHANG Xiao<sup>1</sup>, LI Yegang<sup>1</sup>, WANG Dong<sup>1</sup>, SHI Shumin<sup>2</sup>

(1 School of Computer Science and Technology, Shandong University of Technology, Zibo Shandong 255049, China;

2 School of Computer Science and Technology, Beijing Institute of Technology, Beijing 100081, China)

**[Abstract]** For the problem of processing the named entity recognition in the deep learning method, traditional word embedding method map words or chars into a single vector, which can't represent the ambiguity of the word in the context. The ERNIE-BiGRU-CRF model is proposed. The model expresses the semantic perceptual representation of the enhanced words of the ERNIE pre-training model through Enhanced Representation from kNnowledge IntEgration. The multivariate data knowledge is introduced to generate the semantic vector, the word embedding is input to the GRU layer to extract features, and the label sequence is obtained through the CRF layer. The experimental results show that the  $F_1$  value of the model in the People's Daily corpus reaches 94.46%.

**[Key words]** Named Entity Recognition; Enhanced Representation from kNnowledge IntEgration; Gated Recurrent Unit; Conditional Random Field

## 0 引言

命名实体识别<sup>[1]</sup> (Named Entity Recognition, NER) 是自然语言处理<sup>[2]</sup> (Natural Language Processing, NLP) 中的一项基础性任务,目的是从非结构化数据文本中提取出人名、地名、组织机构名和数字表达等特定类型的实体。NER 在问答系统<sup>[3]</sup>、句法分析<sup>[4]</sup>、机器翻译<sup>[5]</sup>等高级 NLP 任务中有广泛的应用价值。英文命名实体识别和中文命名实体识别相比,起步相对较早,且英文单词间有空格等明显的标识符,实体识别过程中不涉及分词等问题,只需要考虑词本身的特征,识别起来较中文来说难度相对较小。而中文实体结构复杂、形式多样、边界模糊,且中文字在不同的语境下还会存在一词多义、一字多性的情况。同时与分词等其它 NLP 任务相互影响。这些都增加了中文命名实体识别的难度,使得中文命名实体识别更具有研究和应用价值。

针对中文命名实体识别任务,本文在传统深度

学习命名实体识别中引入了融合知识增强语义表示 ERNIE 预训练语言模型,提出了一种 ERNIE-BiGRU-CRF 神经网络模型。ERNIE 通过对训练数据中的词法结构、语法结构、语义信息进行统一建模,极大地增强了通用语义表示能力。实验结果表明,ERNIE-BiGRU-CRF 模型在人民日报语料库上  $F_1$  值达到了 94.46%。与之前最好的 Feng(2018) 模型相比,人名实体、地名实体  $F_1$  值分别提高了 3.26 个百分点和 2.96 个百分点。

## 1 相关工作

在命名实体识别中常用的机器学习算法有条件随机场 (Conditional Random Fields, CRF) 模型、支持向量机 (Support Vector Machine, SVM) 模型等。这类机器学习方法通常依赖复杂的特定人工特征,需要人工制作特征模板,导致识别方法泛化能力弱,识别方法不通用。

随着深度学习的兴起,端到端的神经网络模型

**基金项目:** 国家自然科学基金面上项目(61671064)。

**作者简介:** 张 晓(1995-),男,硕士研究生,主要研究方向:自然语言处理、命名实体识别、迁移学习;李业刚(1975-),男,博士,副教授,硕士生导师,主要研究方向:自然语言处理、机器翻译。

**通讯作者:** 李业刚 Email:lyg8256@qq.com

收稿日期: 2019-11-18

占据了命名实体识别方法的主流。Hammerton<sup>[6]</sup>首先通过 LSTM 网络结构进行命名实体识别的研究。Huang 等人<sup>[7]</sup>通过 LSTM 网络结构和 CRF 相结合,弥补了 LSTM 未能考虑样本输出之间关系的缺陷,该模型在 NER 等序列标记任务中取得了很好的效果。目前许多 NER 框架都是基于 LSTM-CRF 模型的。通过在 LSTM-CRF 模型上融入各种特征,可以进一步提升命名实体识别效果。Ma 等人<sup>[8]</sup>采用了 CNN 提取字符级别的特征,通过字词结合的方式处理命名实体识别任务获得了不错的效果。李丽双等人<sup>[9]</sup>在生物领域内利用 BiLSTM-CNN-CRF 模型取得了当时最高的  $F_1$  值。李明扬等人<sup>[10]</sup>通过使用 Transformer 结构中的多头注意力机制结合 BiLSTM-CRF 模型在微博文本领域的命名实体识别中取得了很好的效果。Strubell 等人<sup>[11]</sup>通过利用空洞卷积网络进行命名实体识别,有效提取序列信息的同时缩短了训练的时间。

将深度学习算法引用到命名实体识别任务中的一个关键步骤是用词向量表达文本中词语<sup>[12]</sup>。其中,word2vec<sup>[13]</sup>的提出吸引了大量科研人员的关注,word2vec 可以通过训练神经网络模型将词语表示为固定长度的低维向量,得到分布表示特征。训练好的向量中蕴涵着一些潜在的语义信息。将预训练词向量作为字词特征输入到神经网络模型中,提高了命名实体识别效果。Lample 等人提出的 BiLSTM-CRF 模型,使用了大量无标记的语料在无监督学习中预训练生成 word2vec 词向量,模型中没有使用附加词典便取得了很好的效果。但 word2vec 无法表示字的歧义性。例如在句子“他每天都重复地做着那些重活”中“重”字表达了不同的意思,类似于 word2vec 等工作聚焦于上下文无关的词向量建模,不能分别区分其含义,针对这个问题,科研人员提出了各种预训练语言模型来进行词表示,Peters 等人<sup>[14]</sup>提出了一种基于 BiLSTM 网络结构的词向量表征模型,该模型通过上下文的语境变动而改变,动态地增强词的表示,表征词汇的语法语义层面特征;Devlin 等人<sup>[15]</sup>提出的 BERT 模型,通过预测屏蔽的词,利用双向 Transformer 网络来获取预训练语言模型;2019 年,Sun 等人<sup>[16]</sup>针对 BERT 在处理中文时难以学习出更大语义单元的完整表示,提出了基于知识增强的 ERNIE 模型直接对先验语义知识单元进行建模,增强了模型语义表示能力。

鉴于 ERNIE 预训练语言模型的强大表义能力,针对中文命名实体识别任务,本文引入了 ERNIE 预

训练语言模型,提出了一种融合知识增强语义表示的 ERNIE-BiGRU-CRF 深度神经网络模型。结合 GRU 网络对文本进行深层特征提取,结果表明,ERNIE-BiGRU-CRF 模型在人民日报语料库上  $F_1$  值达到了 94.46%。

## 2 ERNIE-BiGRU-CRF 模型

ERNIE-BiGRU-CRF 模型的整体架构如图 1 所示。该模型首先通过知识增强语义表示的 ERNIE 预训练语言模型获得输入的语义表示,获取的字向量输入到双向 GRU 层提取句子级别特征,最后经过 CRF 层进行序列标注,获取全局最优的标签序列。

与之前的主流命名实体识别模型相比,ERNIE-BiGRU-CRF 模型最主要的区别是加入了知识增强的语义表示 ERNIE 预训练语言模型,ERNIE 模型通过对词、实体等语义单元的掩码学习得到完整概念的语义表示。表征了字的歧义性,增强了模型语义表示能力。

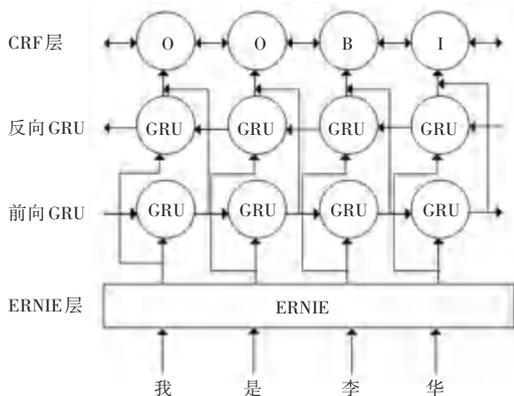


图 1 基于 ERNIE-BiGRU-CRF 的命名实体识别框架

Fig. 1 Named entity recognition framework based on ERNIE-BiGRU-CRF

### 2.1 ERNIE 预训练语言模型

近年来,通过神经网络训练无监督文本的预训练语言模型有效地提高了命名实体识别、情感分析和问答系统等各类 NLP 任务。早期的工作集中于上下文无关词向量的建模,之后的 Glove<sup>[17]</sup>、Elmo 等模型构建了语句级的语义表示,2018 年 Google 的 Devlin 等人<sup>[15]</sup>提出的 BERT 模型利用 Transformer 编码器的 self-attention 双向建模,取得了更好的性能。但以上的几种模型关注的焦点在于处理原始语言信号上,很少利用语义知识建模,导致模型难以获取更大语义单元的完整语义表示,在中文上表现更为突出。针对这个问题,Sun 等人<sup>[16]</sup>提出了基于知识增强的 ERNIE 模型。ERNIE 模型通过对大规模的语料中的词、实体及实体关系进行建

模,学习真实世界的语义关系。

BERT 模型采用双向 Transformer 作为编码器,并提出了掩码语言模型 (Masked Language Model, MLM) 用来捕捉词级别的表示。掩码语言模型用来训练一个深度双向文本表示向量,其主要思想是通过随机的方式来任意屏蔽序列中的 15% 单词作为训练样本,利用编码器通过序列中其它未被屏蔽的单词预测这些被屏蔽单词的原始词汇。BERT 中的 Transformer 基于注意力机制进行建模,Transformer 编码单元如图 2 所示。

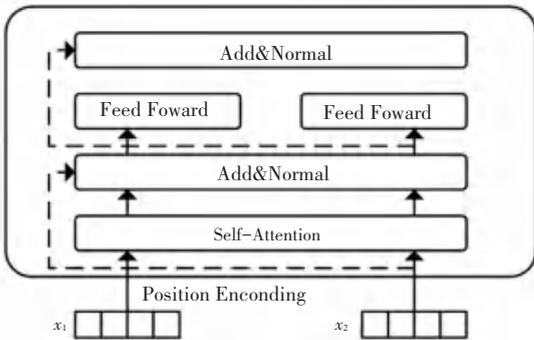


图 2 Transformer 编码单元

Fig. 2 Transformer coding unit

Transformer 最主要的模块是自注意力部分,其相应数学公式可写为:

$$Attention(Q, K, V) = softmax(\frac{QK^T}{\sqrt{d_k}})V. \quad (1)$$

其中, \$Q, K, V\$ 表示输入字向量矩阵, \$d\_k\$ 为输入字向量的维度,通过计算输入序列中每个词与这个序列中其它词的相互关系,反映序列中不同词之间的关联性及其重要程度,然后调整每个词的权重获得每个词的新的表征。

ERNIE 模型全称为 Enhanced Representation through Knowledge Integration, ERNIE 采用了和 BERT 模型相同的多层双向 Transformer 作为编码器提取特征的基本单元。放弃了 BERT 模型的单一基于字掩码策略,通过采用字掩码、短语掩码和实体掩码三种级别的掩码策略,隐式地学习实体属性、实体关系等知识信息,模型得以学习完整概念的语义表示。

基于 BERT 和 ERNIE 的不同 masking 策略如图 3 所示。对于 BERT 模型,其中掩码掉的只是一些字,训练时学到的更多的是字与字之间的关系,例如图 3 中所示的“尔”与“哈”、“滨”之间的局部关系,在 ERNIE 模型中,则能掩码一些连续的 tokens。除了能学习到 BERT 之间的局部关系,还能学到“哈尔

滨”与“省会”间的知识信息。

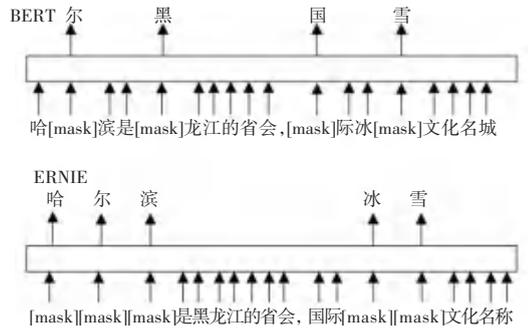


图 3 基于 BERT 和 ERNIE 的不同 masking 策略

Fig. 3 The different masking strategy between BERT and ERNIE

在训练过程中,ERNIE 模型利用词法分析工具对数据进行字、词、实体等不同粒度的切分,根据切分边界对连续的 token 进行随机掩码操作,让模型在语义、知识等不同层次上学到相应的关系。在训练语料上,模型数据预处理也加入了基于资讯类、论坛对话类、百科类等构造具有上下文关系的句子对数据,进一步地提升了模型语义表示能力。

### 2.2 门控循环单元网络

循环神经网络 (Recurrent Neural Network, RNN) 是一类特殊的神经网络,可以动态地捕捉序列信息,其节点具有一定记忆能力,由于网络中当前时刻的输出也与上一时刻的输出值有关,对上下文相关的序列数据文本有很强的处理能力,因此对处理命名实体识别这类序列标注问题有较好的效果,但实际训练过程中 RNN 容易产生梯度问题,使得 RNN 难以处理序列过长的的问题。

为解决 RNN 在处理序列过长时出现梯度上的问题,一系列的 RNN 变体相继被提出,目前主流的变体包括长短期记忆网络<sup>[18]</sup> (Long Short-Term Memory, LSTM) 和门控循环单元<sup>[19-20]</sup> (Gated Recurrent Unit, GRU) 等, LSTM 通过引入输入门、遗忘门和输出门三个门控机制和记忆单元控制信息传递,有效地解决了 RNN 产生的梯度爆炸或梯度消失问题。Cho 等人<sup>[21]</sup>在 LSTM 的基础上,对门控单元进行进一步简化,将 LSTM 中遗忘门和输入门合并为更新门,混合细胞状态和隐藏状态。简化后的 GRU 结构更为简单,减少了网络中的参数,在取得了与 LSTM 模型相当的效果的同时缩短了模型训练时间。GRU 的形式化表示见式(2)~式(5):

$$z_j^c = \sigma(W_{zx}x_j^c + W_{zh}h_{j-1}^c + b^c), \quad (2)$$

$$r_j^c = \sigma(W_{rx}x_j^c + W_{rh}h_{j-1}^c + b^c), \quad (3)$$

$$\tilde{h}_j^c = \tanh(W_{cx}x_j^c + W_{ch}(r_j^c \cdot h_{j-1}^c + b^c)), \quad (4)$$

$$h_j^c = (1 - z_j^c) \cdot h_{j-1}^c + z_j^c \cdot \tilde{h}_j^c. \quad (5)$$

其中,  $z_j^c$  表示一个更新门, 控制下一时刻的信息流入;  $r_j^c$  表示重置门, 控制信息的保留与丢失;  $z_j^c$  和  $r_j^c$  决定隐藏状态的输出;  $\tilde{h}_j^c$  表示当前时刻的候选状态;  $x_j^c$  表示当前时刻的输入;  $h_{j-1}^c$  表示上一时刻的隐藏层状态;  $\sigma$  表示 sigmoid 激活函数;  $W$  和  $b^c$  分别代表每个单元模型参数中的权重矩阵和偏置向量; “ $\cdot$ ”表示 Hadamard 乘积。

### 2.3 条件随机场

条件随机场 (Canonical Correlation Analysis, CRF) 是一种概率化无向图模型。CRF 可以通过一组随机变量来获取另一组随机变量的条件概率分布, 在处理词性标注和序列划分等问题上有很好的效果, CRF 主要通过全局优化的思想, 通过考虑标签之间的依赖关系获取一个全局最优标签序列, 避免了如出现标签 B-PER 后面接入了 I-LOC 等不合理情况。因此, 在双向 GRU 的输出后面接入 CRF 层, 用来计算标签序列。

定义一个大小为  $n \times k$  的矩阵  $P$  是句子  $X = \{W_1, W_2, \dots, W_n\}$  输入到双向 GRU 网络层中的输出, 其中  $n$  表示单词个数,  $k$  表示标签种类中不同标签的个数,  $P_{i,j}$  表示一个句子中第  $i$  个单词输入的第  $j$  个标签的概率。对于给定的一串文本序列  $x = \{x_1, x_2, x_3, \dots, x_n\}$  和其对应预测标记序列  $y = \{y_1, y_2, y_3, \dots, y_n\}$ , 定义得分表示为:

$$s(x, y) = \sum_{i=1}^n (W_{y_{i-1}, y_i} + P_{i, y_i}), \quad (6)$$

$$P_i = W_s h^{(t)} + b_s, \quad (7)$$

其中,  $W$ ,  $W_{i,j}$  分别代表转换矩阵和标签转移分数;  $P_{i,y_i}$  表示一个字符的第  $y_i$  个标签的得分情况;  $h^{(t)}$  为双向 GRU 层  $t$  时刻输入数据  $x^{(t)}$  的隐藏状态。

在 CRF 训练中, 采取最大似然估计函数标记序列, 即:

$$L = \sum_{i=1}^n \log(P(y_i | x_i)) + \frac{\lambda}{2} \|\theta\|^2, \quad (8)$$

模型训练完成后, 采用维特比算法求解出概率值最大的标签作为预测标签。

$$P(y | x) = \frac{e^{s(x, y)}}{\sum_{y \in Y_x} e^{s(x, y)}}. \quad (9)$$

## 3 实验

### 3.1 实验数据集与评价指标

本文实验数据集采用的是 1998 年 1 月份人民日报语料库, 由北京大学计算语言学研究所和富士

通公司共同制作。实验过程中将其随机划分为训练集、测试集和验证集三部分。训练集、测试集、验证集分别占总数据集的 75%、15% 和 10%。该数据集作为中文命名实体识别中的经典数据集广泛地应用在科学研究中。

本文采用了 BIO 方式标注数据集, 人民日报语料库中包含人名、地名、机构名三种实体类型。其中, B 开头的代表该实体类型标签的首字, I 开头代表该类型标签的非首字, O 表示不属于实体部分。在预测实体边界的时候, 实体类别共 7 类, 见表 1。

表 1 实体标签类型

Tab. 1 Entity label type

实体类别	首字符	非首字符
人名	B-PER	I-PER
地名	B-LOC	I-LOC
机构名	B-ORG	I-ORG

实验中性能评价指标采用准确率 (*Precision*,  $P$ )、召回率 (*Recall*,  $R$ ) 和  $F_1$  值 ( $F_1 - score$ ,  $F_1$ ) 三项, 具体见公式 (10):

$$P = \frac{T_p}{T_p + F_p} \times 100\%,$$

$$R = \frac{T_p}{T_p + F_n} \times 100\%,$$

$$F_1 = \frac{2PR}{P + R} \times 100\%. \quad (10)$$

其中,  $T_p$  表示识别正确的实体数量;  $F_p$  表示未正确识别的实体数量;  $F_n$  表示实体但未能识别的数量。  $F_1$  值综合  $P$  和  $R$  评估指标, 用于反映整体指标。

### 3.2 实验设置

实验中的硬件环境配置见表 2。

表 2 实验环境

Tab. 2 Experimental environment

操作系统	Ubuntu
CPU	Intel (R) Core (TM) i7-7700HQ@2.8 Hz
GPU	GTX1060ti
Python	3.6
Tensorflow-gpu	1.14.0

在模型训练过程中, 使用随机梯度法 (Stochastic gradient descent, SGD) 优化模型, 在模型中加入 *Dropout* 层减少模型过拟合问题, *Dropout* 设为 0.5, *batch-size* 为 64, 学习率为 0.001, 学习率的更新采

用 *step decay*, 衰减率为 0.05, L2 正则化参数为 0.000 1。模型训练过程中利用 GTX1060ti 加速处理。

### 3.3 实验结果与分析

为了验证本文提出的基于 ERNIE 的命名实体识别方法的有效性和泛化性, 对比了以下几种主流算法, 评估模型的性能。对此可做解析分述如下。

(1) CRF。CRF 是机器学习中处理命名实体识别任务效果最好且应用最为广泛的算法。

(2) BLSTM-CRF。该模型是序列标注任务中经典神经网络模型, 采用预训练好 word2vec 向量作为 BLSTM 网络的输入进行特征提取, 最后将其特征矩阵输入到 CRF 中完成序列标注。

(3) Feng (2018)<sup>[18]</sup>。该模型是 Feng 等人在 2018 年提出的一种命名实体识别模型, 在 BLSTM-CRF 模型的基础上加入了词语的前后缀信息和领域词典, 并将其转换为词的分布表示特征。

(4) ERNIE-BiGRU-CRF 模型。本文提出的 ERNIE-BiGRU-CRF 模型, 通过引入了 ERNIE 预训练语言模型, 增强模型语义表示能力, 通过学习海量文本中的潜在知识, 提高了命名实体识别效果。

在实验中, 为验证 ERNIE 对实体识别效果的影响, 对比了采用预训练好的词向量输入到 BiGRU-CRF 模型中进行训练的效果。

表 3~表 5 中分别展示了地名、人名、组织名三种实体准确率、召回率、 $F_1$  值的对比情况, 与现有的方法相比, 本文提出的 ERNIE-BiGRU-CRF 模型三种实体的  $F_1$  值分别达到了 93.73%、96.92% 和 91.56%, 总体  $F_1$  值达到了 94.46%。其中, 组织类实体识别的  $F_1$  值相对于其它两种相对较低, 分析是因为组织类实体容易出现地名嵌套、歧义实体等干扰信息, 导致模型预测出现误差。而人名这类实体结构相对简单, 一般由姓和名两部分组成, 更易获取中文命名实体识别的特征, 故  $F_1$  值高于机构名和地名两类实体。

表 3 地名识别实验结果

Tab. 3 Experimental results of local name recognition %

模型	指标		
	精确率	召回率	$F_1$
CRF	91.92	88.95	90.06
LSTM-CRF	93.31	89.16	91.18
Feng(2018)	93.74	87.98	90.77
GRU-CRF	93.05	89.78	91.39
The proposed	93.45	94.01	93.73

表 4 人名识别实验结果

Tab. 4 Experimental results of personal name recognition %

模型	指标		
	精确率	召回率	$F_1$
CRF	83.82	77.56	80.33
LSTM-CRF	88.17	85.64	86.88
Feng(2018)	98.23	89.49	93.66
GRU-CRF	89.42	86.18	87.77
The proposed	96.87	96.98	96.92

表 5 组织名识别实验结果

Tab. 5 Experimental results of organization name recognition %

模型	指标		
	精确率	召回率	$F_1$
CRF	80.64	82.43	81.52
LSTM-CRF	83.57	86.78	85.14
Feng(2018)	97.52	89.34	93.25
GRU-CRF	84.91	86.15	85.52
The proposed	91.91	91.21	91.56

ERNIE-BiGRU-CRF 模型在  $P$ 、 $R$ 、 $F_1$  值三个方面均优于与 CRF 模型。原因是利用 GRU 网络在训练的过程中能够捕捉长距离信息, 能够更加有效地提取特征信息。同时 ERNIE-BiGRU-CRF 模型避免了手动构造特征模板等不足, 解决了传统识别方法过于依赖人工特征的问题, 实现了端到端的命名实体识别。

与 BiGRU-CRF 模型、BLSTM-CRF 模型相比, 地名、人名、组织名三类实体的  $F_1$  值均所提高, 其中提高幅度最大的是人名类实体, 分别提高了 10.04% 和 9.15%。说明基于知识增强的 ERNIE 模型能够通过建模海量数据中的实体概念等先验知识, 增强字的语义感知表示, 更好地表示字的语义信息。例如在句子“罗阳汝信韦典徐达欧阳春汤洪高安启元”中, 正确的实体划分应为“罗阳|汝信|韦典|徐达|欧阳春|汤洪高|安启元”, 但是在 BiGRU-CRF 模型中, 未能识别出安启元这个名字, 研究分析是因为安字作为姓氏出现的场景较少, 利用 word2vec 训练的词向量可能更会偏向于“平安”的意思。而在 ERNIE-BiGRU-CRF 模型中, 由于采用了知识增强语义表示 ERNIE 预训练模型, ERNIE 模型通过对词、实体等语义单元的掩码学习到完整概念的语义表示。表征了字的多义性, 根据不同的语义生成不同的语义向量, 正确识别出人名。

对比 Feng 等人(2018)的模型, ERNIE-BiGRU-

CRF模型在地名实体、人名实体  $F_1$  值均有所提高,说明ERNIE的特征抽取能力较强。与Feng(2018)模型相比,还少去了添加字典等构造外部特征的过程,从而使模型具有更好的鲁棒性与更强的泛化能力。

本文提出的ERNIE-BiGRU-CRF模型更具有竞争力,明显优于本文对比的其它方法,在人民日报语料库实验对比中取得了最佳的结果。证明了该模型在中文命名实体识别中的有效性。

#### 4 结束语

针对传统词向量不能表征字的歧义性问题和模型难以获取更大语义单元的完整语义表示等问题,本文提出了ERNIE-BiGRU-CRF模型。ERNIE预训练模型采用了多层双向Transformer作为编码器提取特征的基本单元,通过采用字掩码、短语掩码和实体掩码三种级别的掩码策略动态生成字的上下文语义表示,比传统的词向量更能增强模型语义表示能力,最终提高了命名实体识别效果。

该模型存在的问题是在一些缺乏大规模标记数据的特定领域中,由于上下文信息不足,且存在缩写、歧义实体等现象时,会出现提取错误的情况,下一步的研究方向可以考虑深度学习结合迁移学习方法来解决。

#### 参考文献

- [1] MARRERO M, URBANO J, SÁNCHEZ-CUADRADO S, et al. Named entity recognition: Fallacies, challenges and opportunities [J]. *Computer Standards & Interfaces*, 2013, 35(5): 482.
- [2] SEKINE S, NOBATA C. Definition, dictionaries and tagger for extended named entity hierarchy [C]//LREC. Lisbon: ELRA, 2004: 1977.
- [3] 于根,李晓戈,刘睿,等. 基于信息抽取技术的问答系统[J]. *计算机工程与设计*, 2017, 38(4): 1051.
- [4] 杨锦锋,于秋滨,关毅,等. 电子病历命名实体识别和实体关系抽取研究综述[J]. *自动化学报*, 2014, 40(8): 1537.
- [5] 张磊,杨雅婷,米成刚,等. 维吾尔语数词类命名实体的识别与翻译[J]. *计算机应用与软件*, 2015, 32(8): 64.
- [6] HAMMERTON J. Named entity recognition with long short-term memory [C]//Proceedings of the Seventh Conference on Natural Language Learning at HLT-NAACL 2003. Stroudsburg, PA, USA: ACL, 2003, 4: 1.
- [7] HUANG Z, XU W L, YU K, et al. Bidirectional LSTM-CRF models for sequence tagging [J]. *arXiv preprint arXiv: 1508.01991*, 2015.
- [8] MA Xuezhe, HOVY E. End-to-end sequence labeling via bi-directional LSTM-CNNs-CRF [J]. *arXiv preprint arXiv: 1603.01354*, 2016.
- [9] 李丽双,郭元凯. 基于CNN-BLSTM-CRF模型的生物医学命名实体识别[J]. *中文信息学报*, 2018, 32(1): 116.
- [10] 李明扬,孔芳. 融入自注意力机制的社交媒体命名实体识别[J]. *清华大学学报(自然科学版)*, 2019, 59(6): 461.
- [11] STRUBELL E, VERGA P, BELANGER D, et al. Fast and accurate entity recognition with iterated dilated convolutions [J]. *arXiv preprint arXiv: 1702.02098*, 2017.
- [12] BENGIO Y, SCHWENK H, SENÉCCAL J S, et al. A neural probabilistic language model [J]. *Journal of Machine Learning Research*, 2003, 3(6): 1137.
- [13] MIKOLOV T, CHEN K, CORRADO G S, et al. Efficient estimation of word representations in vector space [C]//Proceedings of International Conference on Learning Representations. Scottsdale, Arizona: ICLR, 2013: 1.
- [14] PETERS M E, NEUMANN M, IYYER M, et al. Deep contextualized word representations [C]//North American Chapter of the Association for Computational Linguistics. Louisiana, USA: dblp, 2018: 2227.
- [15] DEVLIN J, CHANG M, LEE K, et al. BERT: Pre-training of deep bidirectional transformers for language understanding [C]//North American Chapter of the Association for Computational Linguistics. Minneapolis: dblp, 2019: 4171.
- [16] SUN Y, WANG S, LI Y, et al. ERNIE: Enhanced representation through knowledge integration [J]. *arXiv preprint arXiv: 1904.09223*, 2019.
- [17] 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). Doha, Qatar: ACL, 2014: 1532.
- [18] 冯艳红,于红,孙庚,等. 基于BLSTM的命名实体识别方法[J]. *计算机科学*, 2018, 45(2): 261.
- [19] 石春丹,秦岭. 基于BGRU-CRF的中文命名实体识别方法[J]. *计算机科学*, 2019, 46(9): 237.
- [20] 李佰蔚. 基于GRU-CRF的中文命名实体识别方法研究[D]. 哈尔滨: 哈尔滨理工大学, 2019.
- [21] CHO K, Van MERRIENBOER B, GULCEHRE C, et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation [C]//Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing. Doha, Qatar: Association for Computational Linguistics, 2014: 1724.