文章编号：1003-0077（2017）00-0000-00

融入自注意力机制的社交媒体命名实体识别

李明扬，孔芳*
（苏州大学 计算机科学与技术学院，江苏 苏州 215006）

摘要：针对中文社交媒体命名实体识别的结果远不如传统领域的识别效果，以及近年来中文社交媒体命名实体识别研究逐渐倾向于使用外部知识与联合训练，而忽视了进一步提取文本中的特征，该文提出了一种基于深度学习、结合双向长短时记忆和自注意力机制的命名实体识方法。在 Weibo NER 公开语料上的对比实验表明了我们所提出方案的有效性，实验表明在不使用外部资源和联合训练的情况下，F_1值为 58.76%。

关键词：命名实体识别；中文社交媒体；自注意力机制

中图分类号：TP391 **文献标识码：**A

Combining Self-attention Mechanism for Named Entity Recognition in Social Media

LI Ming Yang , KONG Fang*
(School of Computer Science and Technology, Soochow University, Suzhou, Jiangsu 215006, China)

Abstract: For the problems of the result of named entity recognition for Chinese social media is much less than the traditional field, and Chinese social media NER in recent years gradually tend to use external knowledge and jointly training, thus ignoring the further extracting features from the text. The article puts forward a method of named entity recognition based on deep neural networks that combines a bi-directional long short-term memory with self-attention mechanism. Comparative experiments on the Weibo NER released corpus show the effectiveness of our proposed approach, and show that without using external knowledge and transfer learning, our method achieved 58.76% in F_1-score.

Key words: Named Entity Recognition; Chinese Social Media; Self-attention Mechanism

0 引言

命名实体识别（Named Entity Recognition，NER）是指识别出非结构化文本中出现的包括人名、地名、组织机构名等实体的指称。作为信息抽取的核心技术之一，命名实体识别在依赖信息抽取技术的诸如知识库自动构建、问答系统中有广泛的应用场景和应用价值。国外对于英文命名实体识别的研究开始的比较早，并且由于英文命名实体识别只需考虑词本身的特征而不涉及分词问题，识别的难度相对较低。相反，由于中文内在的特殊性，如单元词汇边界模糊，实体结构复杂、表现形式多样、缺乏显式的单词边界和其他提示命名实体的线索等，增加了中文命名实体识别的难度。

目前，为了减少对语言学知识的依赖和避免繁琐的特征工程，命名实体识别逐渐由使用传统的统计学习方法转移到应用深度学习的方法，借助搭建多层神经网络结构来学习文本中潜在的相关信息。

收稿日期：2017-03-16；**定稿日期**：2017-04-26
*通信作者：kongfang@suda.edu.cn
基金项目：国家自然科学基金（61472264）；人工智能应急项目（61751206）；国家重点研发计划子课题（2017YFB1002101）

近年来，针对社交媒体的中文命名实体研究成为热点。社交媒体的中文命名实体识别主要有三个难点：①相对于英文，中文缺乏显式的词汇边界和固有的定冠词，专有词汇也没有拼写变化。②社交媒体上发布的往往是不规范的短文本，新词、错词的出现更为频繁，网络用语、表情等噪音更多。③语料规模更小（例如本文所使用的 Weibo NER 语料的训练集是 MSRA 语料训练集的 1/30）。因此如何在规模较小、混杂很多噪声的语料上尽可能地获取更多有效特征来提升中文社交媒体命名实体识别的性能具有很重要的研究价值。

我们证明，在双向长短时记忆网络-条件随机随机场（LSTM-CRF）序列标注模型基础上，加上多头自注意力机制，在多个不同子空间捕获上下文相关信息，从而理解句子结构，能够提升不规范文本的实体识别标注性能。后续章节中统一使用 LSTM-Self_Att-CRF 代表本文提出的方法。在不依赖外部资源和联合学习的实验配置下，LSTM-Self_Att-CRF 方法取得的F_1值为 58.39%，通过对最后结果的分析，我们发现我们的模型识别出了更多训练集中未出现的实体。

本文后续内容安排如下：第一节介绍中文社交媒体相关的研究，第二节详细介绍基于 LSTM-Self_Att-CRF 的中文社交媒体命名实体识别模型以及引入的 Self-Attention 机制，第三节给出实验过程及实验结果的详细分析，第四节是结论。

1 相关研究

中文社交媒体命名实体识别的相关工作专注于面向规模较小的标注语料进行有效的监督学习，现有的中文社交媒体的命名实体识别方法大都是在传统的命名实体识别方法上：①引入外部资源（字典、知识库、维基百科等）；②将相关任务（分词、词性标注）进行联合训练，解决包含噪声的文本短语切分性能下降对命名实体识别的影响；③使用迁移学习将训练完成的传统领域的命名实体识别模型放入到社交媒体领域中。详细的相关研究如下。

Peng and Dredez[1]在 2015 年首先发布了一个中文社交媒体语料库：Weibo NER corpora，用于命名实体实体的相关研究，随后他们提出了将基于“字+位置”的 embedding 与 NER 任务联合训练的模型。Peng and Dredez[2]将外部资源和联合训练相结合，在 2016 年提出了将中文分词表征作为特征的方法从而来提高命名实体识别的性能。

He and Sun[3]在 2016 年提出把句子级别的得分（F-score）放入损失函数中，实验结果表明在不使用外部资源的情况下取得了较高的准确率（Precision），但是召回率（Recall）较低。之后，He and Sun[4]在 2017 年又提出了使用跨领域数据集的半监督联合模型，实验表明提升了识别性能，但是未能超过 Peng and Dredez 在 2016 年提出的模型。

本文从文本自身出发，在不使用外部资源、不进行多任务联合训练的情况下，在使用传统的 Bi-LSTM+CRF 模型中加入 Self-Attention 机制进一步捕获文本的特征，从而提高社交媒体的中文命名实体识别的性能。

2 命名实体识别框架

与大多数实体识别方法相同，本文也将实体识别任务转化为一个序列标注问题，实体开头的单元标注为B，实体内的单元标注为I，其他的标注为O。为了避免汉语分词工具在不规范文本上的性能下降对实体识别任务的错误传播，我们效仿 He and Wang[5]等人的工作，以字的粒度进行命名实体识别。图 1 给出了 LSTM-Self_Att-CRF 模型，该模型包含三个部分：①字粒度的表示层②基于 BiLSTM 的上下文的序列编码层 ③融入自注意力机制的 CRF 解码层。

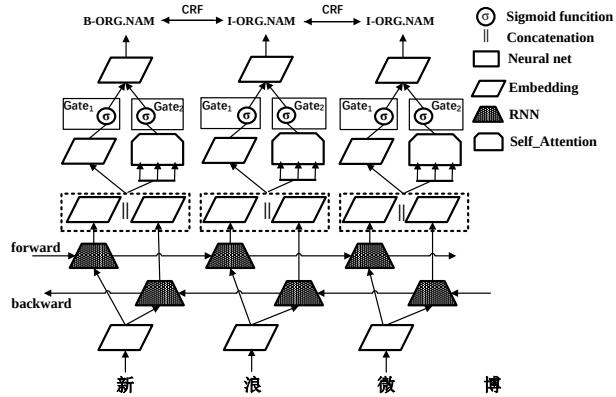

图 1 LSTM-Self_Att-CRF 模型

2.1 字的分布式表示

在编码阶段，原始数据通过查找字向量表转化为字向量序列。其中，本文所使用的字向量表包括 3103 个常用汉字和一些特殊字符（如数字、标点等）以及它们的分布式向量表示。该字向量表使用了 word2vec[6]工具，将无标签中文 Gigaword 数据集训练成相应的数值向量表。

使用 word2vec 训练字向量的过程中，本文选用是 Skip-gram 模型，字的频数（min_count）设置为 100，字向量训练时上下文扫描窗口（window）设置为 5，负采样（negative）的值设为 10，每个字的向量维度（size）设置为 200。

字向量表查找的过程是让原始文本中每一个字符在表上查找相对应的字向量，如果某个字在表中不存在，则被赋予一个随机值。

2.2 LSTM

长短时记忆（long short-term memory，LSTM）很好地解决了传统循环神经网络（RNN）在训练过程中存在的梯度消失和梯度爆炸的问题，同时可以更好地对长距离依赖关系进行建模[7]。图 2 是一个典型的 LSTM 单元结构。

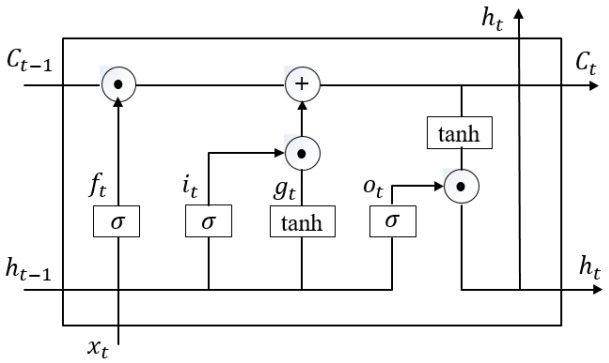

图 2 LSTM 单元结构

LSTM 的隐藏层由特殊构建的记忆单元（Cell）构成。每个 Cell 由以下四个部分组成：①循环连接的 Cell；②用于控制输入信号流量的输入控制门i；③用于控制流向下一个单元的信号强度的输出门o；④用于控制遗忘之前 Cell 状态的遗忘门f。每个时刻t各个单元的计算如式（1）~（6）所示。

$$i_t = \sigma(W_i \cdot [h_{t-1}\,;\, x_t] + b_i)\#(1)$$

$$f_t = \sigma\big(W_f \cdot [h_{t-1}\,;\, x_t] + b_f\big)\#(2)$$

$$o_t = \sigma(W_o \cdot [h_{t-1}\,;\, x_t] + b_o)\#(3)$$

$$g_t = tanh(W_c \cdot [h_{t-1}\,;\, x_t] + b_c)\#(4)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot g_t\#(5)$$

$$h_t = o_t \odot \tanh(c_t)\#(6)$$

其中，$\odot$表示元素级乘法计算，σ表示 sigmod 函数，W_i、W_f、W_o、b_i、b_f、b_o分别表示输入门、遗忘门、输出门的权值矩阵和偏置项。

2.3 Self-Attention

在字向量经过 Bi-LSTM 提取特征后，我们需要使用 Self-Attention 机制来进一步获取文本的特征。本文使用 Multi-head Attention（该 attention 的详细内容会在下一个章节介绍）从多角度、多层次地获取文本自身的相关特征，通过输入三组相同的 Bi-LSTM 的隐藏层输出h_i，便得到经 attention 处理后的信息向量y_i。在解码之前，将h_i和y_i按照一定的权重进行相加，从而得到结合上下文特征与自身特征的信息向量z_i。计算如式（7）~（8）所示。

$$h_i = \overrightarrow{h_i} || \overleftarrow{h_i}\#(7)$$

$$y_i = Multi\_Head(h_i, h_i, h_i)\#(8)$$

$$z_i = Gate_1 * h_i + Gate_2 * y_i\#(9)$$

考虑到句子中每个位置的标注对于上下文的依赖程度不同，我们引入门控机制来学习句子中每个位置所占权重，该门控机制由$sigmoid$单元组成。其中，$Gate_1$、$Gate_2$分别表示h_i、y_i所占的权重向量。为了得到的信息向量能进行正则化处理，我们设置$Gate_1 + Gate_2 = 1$。

2.4 Self-Attention 介绍

注意力机制是一种用来分配有限地信息处理能力的选择机制，其特点为选择性地关注某些重要的信息，相应的忽略同一时刻接收到的其他信息[8]。相对应于文本处理的表现为将较高的权重分配给重要的文字，而将较小的权重赋给其他的文字。attention 函数$Attention(Q, K, V)$的本质可以被描述为一个查询（query）到一系列（键 key--值 value）对的映射。自注意力（Self-Attention）机制就是在序列内部做 attention，寻找序列内部的联系，即：$Attention(X, X, X)$，X就是输入序列。

注意力机制广泛的运用于神经机器翻译上，Google 机器翻译团队在 2017 年发表的《Attention is all you need》[9]论文中提出了 Transform 结构，并且该结构在机器翻译领域取得了最佳的效果。考虑到中文微博 NER 语料规模较小，且语料中存在大量的不规范文本，需要从多角度、多层级的视角提取更多的文本自身的特征，所以使用 Transform 结构中的多头注意力机制应该有助于微博文本的命名实体识别。

首先介绍放缩点积注意力（Scaled Dot-Product

Attention）机制，其本质上是使用点积进行相似度计算的注意力机制。图 3 给出了 Scaled Dot-Product Attention 的计算方式，如式（10）所示：

$$Attention(Q,K,V) = softmax\left(\frac{QK^T}{\sqrt{d_K}}\right)V, \#(10)$$

其中，Q,K,V均为向量形式，且$Q \in \mathbb{R}^{n\times d_K}, K \in \mathbb{R}^{m\times d_K}, V \in \mathbb{R}^{m\times d_V}$，$d_K$表示 Q,K 的第二维度。该 attention 层本质上是将$n \times d_K$的序列 Q 编码成了一个新的$n \times d_V$的新序列，$\sqrt{d_K}$起到了调节作用，控制 Q,K 的内积不会太大，确保该 attention 为软分布注意力（soft attention[10]）。

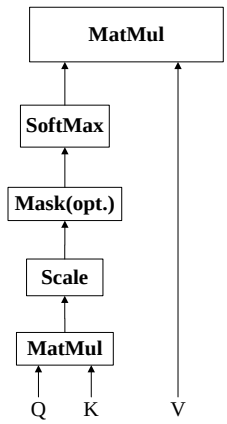

图 3 放缩点积注意力机制

在此基础上，考虑到一个 attention 机制无法从多角度、多层面的捕获到重要的特征，所以需要使用多头注意力（Multi-Head Attention）机制。

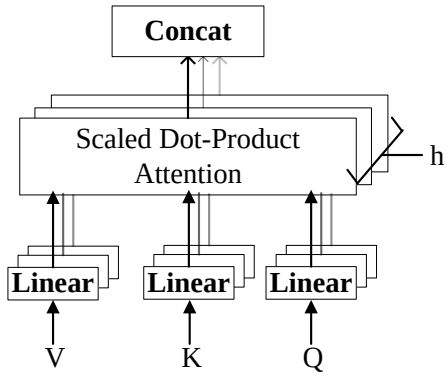

图 4 多头注意力机制

多头注意力机制在参数不共享的前提下将Q,K,V通过参数矩阵映射后再做 Scaled Dot-Product Attention，并将这个过程重复做h次，最后将结果进行拼接，从而获得较全面的特征信息。

图 4 给出了 Multi-Head Attention 的计算方式，该 attention 的计算如式（11）～（12）所示。

$$head_i = Attention\left(QW_i^Q, KW_i^k, VW_i^v\right)\#(11)$$

$$MultiHead(Q,K,V) = Concat(head_1, \dots, head_h)\#(12)$$

其中，$W_i^Q \in \mathbb{R}^{d_K\times \tilde{d}_K}, W_i^k \in \mathbb{R}^{d_K\times \tilde{d}_K}, W_i^v \in \mathbb{R}^{d_v\times \tilde{d}_v}$，$Concat$表示将每次的结果进行拼接。

2.5 条件随机场模型

条件随机场（Conditional Random Filed，CRF）模型是给定一组输入随机变量条件下另一组输出随机变量的条件概率分布模型，其特点是假设输出随机变量构成马尔可夫随机场[11]。本文使用链式条件随机场对 Bi-LSTM 和 Self-Attention 生成的信息向量进行解码，该模型广泛应用于序列标注问题中，计算给定输入序列下标记序列的分布[12]。对应观察序列x，对应的标注序列y的链式条件随机场模型的概率计算公式如式（13）～（14）所示。

$$p(y|x) = \frac{1}{Z(x)}\exp\left(\sum_{i,k}\lambda_k f_k(y_{i-1}, y_i, x) + \sum_{i,k} u_k g_k(y_i, x)\right)(13)$$

$$Z(y|x) = \sum_{y} exp\left(\sum_{i,k}\lambda_k f_k(y_{i-1}, y_i, x) + \sum_{i,k} u_k g_k(y_i, x)\right)\#(14)$$

其中，$Z(x)$是归一化因子，$f_k(y_{i-1}, y_i, x)$表示观察序列x中位置i和$i-1$的输出节点的特征，$g_k(y_i, x)$表示位置i的输入节点和输出节点的特征，λ和u分别表示这两个特征函数的权重。

最终的解码阶段通过 CRF 中的标准维特比算法[13]预测全局最优标注序列，计算如式（15）～（16）所示。

$$s_{char} = f(W_{out}h_i, b_{out})\#(15)$$

$$s(X, Y', \theta) = \sum_{i=1}^{n}\left(A_{y'_{i-1},\ y'_i} + s_{char}(i)\right)\#(16)$$

其中，$s_{char}(i)$表示输入字符x_i经过双向 LSTM 网络得到的标注的概率分布，W_{out}和b_{out}为全连接层的映射矩阵及偏置向量，f为$softmax$函数。A表示标注状态的转移矩阵，例如，$A_{y'_{i-1},\ y'_i}$表示从标注状态y'_{i-1}到y'_i的转移概率。$s(X, Y', \theta)$表示所有候选标注路径的得分值，最终取其中的最大值作为最后的输出标签序列。

3 实验设置与结果分析

前文介绍了基于深度神经网络的命名实体识别框架，并引入了 Self-Attention 机制进一步获取文本自身的特征、理解句子结构，从而丰富文本的表征。本节将使用 Peng and Dredez 公开发布的 Weibo NER 数据集，通过不同的设置对模型进行实验，并对实验结果进行讨论分析。

3.1 实验数据集

本文采用了 Peng and Dredez[1]公开的已经划分好的 Weibo NER 语料，其中包含训练集、开发集和测试集共 1890 句。表 1 详细地给出了该语料的结构，可以看出该语料整体规模不大，从带标记比例值可以看出待识别的实体数目也存在相对较少的问题。

表 1 Weibo NER 数据集结构

	训练集	开发集	测试集
句子数	1350	270	270
字符数	73378	14509	14842
带标记字符数	4951	971	1078
标记比例（%）	6.71	6.69	7.26

该微博语料待识别的实体类型包括 GPE、LOC、ORG 和 PER，且每个类型分别都有特定实体（Named Entity，NE）和指代实体（Nominal Mention，NM），具体含义如表 2 所示。

表 2 Weibo NER 语料标签含义

特定实体		指代实体	
标签	含义	标签	含义
GPE.NAM	地理政治特定实体	GPE.NOM	地理政治指代实体
LOC.NAM	地名特定实体	LOC.NOM	地名指代实体
ORG.NAM	组织名特定实体	ORG.NOM	组织名指代实体
PER.NAM	人名特定实体	PER.NOM	人名指代实体

该微博语料中各个标签在训练集、开发集和测试集的分布如表 3 所示。

特定实体即为传统领域中的要进行识别的实体，例如，人名的特定实体有：马化腾、李开复等。而指代实体是将名词性的指代词作为实体，例如，人名的指代实体有：妈妈、女朋友等。特定实体与指代实体混合出现在语料中，这又增加了实体识别的难度。

表 3 Weibo NER 语料分布

	训练集	开发集	测试集	总计
GPE.NAM	205	26	47	278
GPE.NOM	8	1	2	11
LOC.NAM	56	6	19	81
LOC.NOM	51	6	9	66
ORG.NAM	183	47	39	269
ORG.NOM	42	5	17	64
PER.NAM	574	90	111	775
PER.NOM	766	208	170	1144

3.2 实验设置

实验中采用了 Pytorch 0.3.0 框架，并用 NVIDIA 的 1080GPU 进行了加速。具体的模型参数配置如下。

（1）生成预训练字向量的参数已经在本文 2.3 节阐述过了，这里就不赘述了。

（2）训练 LSTM-Self_Att-CRF 模型：模型的查询表初始化为上述（1）中预训练得到的字向量。其他参数均采用均匀分布的随机函数初始化成较小的实数。模型中双向 LSTM 的输入是（100 × max_sen_len ×200）的张量，其中第一维度表示 batch size，第二维表示在 batch size 个句子中最大的句子长度，是一个变量，第三维表示隐藏层的维度。LSTM 网络的输出部分将生成（100 × max_sen_len ×400）的张量，其中 400 是前向和后向两个 LSTM 的 Cell 拼接而成的向量大小。其中，优化函数选择随机梯度下降（Stochastic Gradient Decent，SGD）算法[14]，学习率lr设置为 0.015，学习率减少步长lr_decay设置为 0.05，LSTM 的层数$lstm_layer$设置为 1，$dropout$设置为 0.5，Self-Attention 中的d_K设置为 4。

3.3 实验结果及分析

实验采用准确率P、召回率R和的F_1值对识别结果进行评价[15]。其中，F_1值能够综合评价模型的性能。三种评价指标的计算如式（17）～（19）所示。

$$P = \frac{\text{正确识别的实体数}}{\text{识别的实体数}} \#(17)$$

$$R = \frac{\text{正确识别的实体数}}{\text{样本的实体数}} \#(18)$$

$$F_1 = \frac{2 \times \text{准确率} \times \text{召回率}}{\text{准确率} + \text{召回率}} \#(19)$$

本文以 Bi-LSTM + CRF 为基本模型，第二个模型是在基本模型的基础上加入位置特征（position-Embedding），第三个模型，即 LSTM-Self_Att-CRF 模型是在第二个模型基础上再加入 Self-Attention。另外，本文的模型与已有模型的对比如表 4 所示。

表 4 模型对比

模型	使用外部知识	联合训练
Peng and Dredez(2015)	否	是
Peng and Dredez(2016)	是	是
He and Sun(2017a)	否	否
He and Sun(2017b)	是	是
Bi-LSTM + CRF	否	否
Bi-LSTM + pos + CRF	否	否
LSTM-Self_Att-CRF	否	否

表 5 给出了各个模型的实验结果。从表中可以看出，我们提出的 LSTM-Self_Att-CRF 模型在综合性能F_1值上达到了 58.76%，均高于其他现有模型，除了比使用了外部知识、将分词联合训练的 Peng and Dredez(2016)的模型低了 0.23%。所以实验结果也证明了：引入多头注意力（Multi-Head Attention）机制能在多个不同子空间捕获上下文相关信息，从而理解句子结构，能够提升不规范文本的实体识别标注性能。

表 5 中文社交媒体命名实体识别实验结果

模型	NE	NM	Overall
Peng and Dredez(2015)	51.96	61.05	56.05
Peng and Dredez(2016)*	**55.28**	**62.97**	**58.99**
He and Sun(2017a)	50.60	59.32	54.82
He and Sun(2017b)*	54.50	62.17	58.23
Bi-LSTM + CRF	44.44	55.34	49.86
Bi-LSTM + pos + CRF	52.57	57.38	54.95
LSTM-Self_Att-CRF	**56.10**	**61.62**	**58.76**

表 6、表 7 列出了本文提出的模型和目前性能最高的 Peng and Dredez(2016)模型的详细对比结果，两张表分别从特定实体（NE）与指代实体（NM）的P、R和F_1值进行比较。从两张表可以看出，我们的模型在特定实体的识别性能上略高于 Peng and Dredez(2016)模型，但是在指代实体识别的性能上比他的模型低了约 1.35%。

表 6 NE 识别详细实验对比结果

	特定实体		
模型	P	R	F_1
Peng and Dredez(2016)	66.67	47.22	55.28
LSTM-Self_Att-CRF	63.90	50.00	56.10

表 7 NM 识别详细实验对比结果

	指代实体		
模型	P	R	F_1
Peng and Dredez(2016)	74.48	54.55	62.97
LSTM-Self_Att-CRF	69.18	55.55	61.62

具体的，我们的模型因为引入了 Self-Attention（详细的 Self-Attention 的效用见 3.4 节），模型在文本自身上充分捕获有用信息，从而在召回率R值上的表现都优于他的模型，尤其在特定实体识别上，我们的R值高了他的模型约 2.78%。但是，因为 Peng and Dredez(2016)模型使用了外部资源和联合训练，所以他的模型在准确率P值的表现上均优于我们的模型，尤其在指代实体识别上的P值高出了我们约 5.3%。

3.4 Self-Attention 的效用分析

关于 Self-Attention 的效用，本文将未加入 Self-Attention 的第二组实验和加入 Self-Attention 的第三组实验进行详细地比较，并分析加入 Self-Attention 的效用。实验结果如表 8、9 所示。

表 8 NE 识别详细实验对比结果

	特定实体		
模型	P	R	F_1
Bi-LSTM + pos + CRF	63.39	44.90	52.57
LSTM-Self_Att-CRF	63.90	50.00	56.10

表 9 NM 识别详细实验对比结果

	指代实体		
模型	P	R	F_1
Bi-LSTM + pos + CRF	63.97	52.02	57.38
LSTM-Self_Att-CRF	69.18	55.55	61.62

实验结果显示，在特定实体识别任务中，加入 Self-Attention 后，模型学习到了更多的上下文本特征，从而召回率R值得到了明显的提升。在指代实体识别任务中，加入 Self-Attention 后，准确率p值与召回率R值均得到了明显的提升，原因是指代实体相较于特定实体更加缺乏显式的实体边界，而 Self-Attention 能在句子中直接连接两个任意的元素，因此，远距离的元素可以通过更短的路径相互作用，这使得文本信息在网络中广泛地传播，从而提升识别性能。

通过进一步对比未加 Self-Attention 和添加 Self-Attention 的实验结果，我们统计了以下三种识别情况：①前者错误识别，后者正确识别，且边界正确；②前者错误识别，后者正确识别，但边界错误； ③前者正确识别，但存在边界错误，后者正确识别，且边界正确。统计结果如表 10 所示。

表 10 三种识别情况分布

	个数	占比（%）
情况 1	25	78.12
情况 2	4	12.50
情况 3	3	9.37

由表 10 可知，在引入 Self-Attention 后，主要解决的问题就是在不规范文本中未能识别出的实体，在经过 Self-Attention 从多个不同子空间捕获上下文相关信息，理解了句子结构后，从而正确的识别了实体。例如下面的例子。

例：“你这个不孝子哈哈哈”是测试集中的例句。其中，“不孝子”是人名指代实体，在未引入 Self-Attention 前未能识别。引入 Self-Attention 后，经过从不同角度多层次的捕获上下文信息，前文中“你这个”提升了“不孝子”在文本中的权重，从而该实体被正确的识别出来。

4 结论

本文提出了一种中文社交媒体命名实体识别的模型：LSTM-Self_Att-CRF 模型。在不使用外部资源、不进行多任务联合训练的情况下，充分捕获文本自身的特征，实验结果表明本文提出的 Bi-LSTM+CRF 模型很有效的使用 Self-Attention 进行文本自身的特征提取，取得了与 Peng and Dredez(2016) 模型相当的性能，最终的F_1值为 58.76%。

未来的工作我们希望将迁移学习应用到社交媒体的中文命名实体识别任务中，利用传统领域规模较大的数据集以及传统领域已经取得的较高的F_1值，从而进一步提高任务的性能。

参考文献

[1] Peng N, Dredze M. Named entity recognition for chinese social media with jointly trained embeddings[C]//Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing. 2015: 548-554.

[2] Peng N, Dredze M. Improving named entity recognition for chinese social media with word segmentation representation learning[J]. arXiv preprint arXiv:1603.00786, 2016.

[3] He H, Sun X. F-score driven max margin neural network for named entity recognition in chinese social media[J]. arXiv preprint arXiv:1611.04234, 2016.

[4] He H, Sun X. A Unified Model for Cross-Domain and Semi-Supervised Named Entity Recognition in Chinese Social Media[C]//AAAI. 2017: 3216-3222.

[5] He J, Wang H. Chinese named entity recognition and word segmentation based on character[C]//Proceedings of the Sixth SIGHAN Workshop on Chinese Language Processing. 2008.

[6] Goldberg Y, Levy O. word2vec explained: Deriving mikolov et al.'s negative-sampling word-embedding method[J]. arXiv preprint arXiv:1402.3722, 2014.

[7] Hochreiter S, Schmidhuber J. LSTM can solve hard long time lag problems[C]//Advances in neural information processing systems. 1997: 473-479.

[8] 冯辉. 视觉注意力机制及其应用研究[D]. 华北电力大学(北京), 2011.

[9] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[C]//Advances in Neural Information Processing Systems. 2017: 6000-6010.

[10] Yao L, Torabi A, Cho K, et al. Video description generation incorporating spatio-temporal features and a soft-attention mechanism[J]. arXiv preprint arXiv:1502.08029, 2015.

[11] Huang Z, Xu W, Yu K. Bidirectional LSTM-CRF models for sequence tagging[J]. arXiv preprint arXiv:1508.01991, 2015.

[12] 洪铭材，张阔，唐杰,等. 基于条件随机场(CRFs)的中文词性标注方法[J]. 计算机科学，2006, 33(10):148-151.

[13] Forney G D. The viterbi algorithm[J]. Proceedings of the IEEE, 1973, 61(3): 268-278.

[14] Bottou L. Stochastic gradient descent tricks[M]//Neural networks: Tricks of the trade. Springer, Berlin, Heidelberg, 2012: 421-436.

[15] 陈治纲，何丕廉，孙越恒，等. 基于向量空间模型的文本分类系统的研究与实现[J]. 中文信息学报，2005，19(1)：37-42.

李明扬（1995—），硕士研究生，主要研究领域为自然语言处理，命名实体识别，实体链接。
E-mail：20175227067@stu.suda.edu.cn

孔芳（1977—），博士，教授，主要研究领域为机器学习，自然语言处理，篇章分析。
E-mail：kongfang@suda.edu.cn