面向微博文本的融合字词信息的轻量级命名实体识别

陈淳 **李明扬** **孔芳***
苏州大学计算机科学与技术学院
江苏省苏州市干将东路333号苏州大学158信箱
{20195227037, 20175227067}@stu.suda.edu.cn, kongfang@suda.edu.cn

摘要

中文社交媒体命名实体识别由于其领域特殊性，一直广受关注。非正式且无结构的微博文本存在以下两个问题：一是词语边界模糊；二是语料规模有限。针对问题一，本文将同维度的字词进行融合，获得丰富的文本序列表征；针对问题二，提出了基于Star-Transformer框架的命名实体识别模型，借助星型拓扑结构更好地捕获动态特征；同时利用高速网络优化Star-Transformer中的信息桥接，提升模型的鲁棒性。本文提出的轻量级命名实体识别模型取得了目前Weibo语料上最好的效果。

关键词： 命名实体识别；中文社交媒体；星型-Transformer；高速网络

Lightweight Named Entity Recognition for Weibo Based on Word and Character

Chun Chen **Mingyang Li** **Fang Kong**
School of Computer Science and Technology, Soochow University
{20195227037, 20175227067}@stu.suda.edu.cn, kongfang@suda.edu.cn

Abstract

Chinese social media named entity recognition has been widely concerned due to its domain specificity. Informal and unstructured Weibo text has two issues to be addressed. First is the ambiguous word boundary; Second is the limited scale of corpus. To deal with the first problem, this paper places character and word embedding on the same dimension to obtain rich sequence representation. Aiming at the second problem, a named entity recognition model based on Star-Transformer framework is proposed to capture dynamic feature preferably, with the help of star topology structure. Besides, Highway Networks is also used to optimize the information connection in Star-Transformer, improving the robustness of the model. The lightweight named entity recognition model proposed in this paper achieves the best performance on Weibo corpus.

Keywords: Named Entity Recognition , Chinese Social Meida , Star-Transformer , Highway Networks

1 引言

*通讯作者：kongfang@suda.edu.cn

第十九届中国计算语言学大会论文集，第402页-第413页，海南，中国，2020年10月30日至11月1日。

命名实体识别(Named Entity Recognition, NER)旨在识别出非结构化文本序列中具有特殊含义的实体，并为这些实体分配相应的类别，比如人名、地名、组织机构名等等。由于命名实体识别在对话生成 (Reddy et al., 2019)、关系抽取 (Zelenko et al., 2003)、共指消解 (Clark and Manning, 2016)等任务中起着基础支撑作用，因此命名实体识别在自然语言处理(Natural Language Processing, NLP)领域得到了广泛的研究。

社交媒体领域的中文命名实体识别一直是亟待发展的热点任务之一，由于其领域特殊性，社交媒体的中文命名实体识别主要有3个难点：1) 相较于英文，汉语没有显式的词语边界，专有词汇也没有拼写变化等提示信息；2) 社交媒体领域多是不规范的短文本，新词和错词频繁出现，网络用语以及表情等噪声干扰较多；3)社交媒体领域的语料相较于规范的新闻类语料规模较小。

作为一个典型的序列标注问题，命名实体识别的神经网络模型通常包含三个组件：单词嵌入层、上下文编码器层以及解码器层，现有的命名实体识别模型一般通过不断优化这三个组件来寻求突破。循环神经网络（Recurrent Neural Network, RNN） (Zaremba et al., 2014)这一类编码器，尤其是双向长短期记忆网络(Bi-directional Long Short-Term Memory, BiLSTM) (Hochreiter and Schmidhuber, 1997)以序列作为输入，在上下文信息方面有着强大的学习能力，但是不能捕捉到较长距离的上下文依赖关系。Vaswani et al. (2017)提出的Transformer模型采用完全连接的自注意力结构来对远程上下文进行建模，能够较好地弥补RNN模型的缺点，而且Transformer具有更好的并行计算能力。但Yan et al. (2019)的实验证实Transformer不能很好地适用于NER任务中，原因是Transformer内部结构复杂并且采用全连接的注意力机制，这也导致性能的提升需要依赖大量的训练数据。而对于社交媒体领域的命名实体识别而言，难点之一就是语料规模过小，因此传统的Transformer无法取得预期的性能。因此，如何将Transformer较好地融入到NER任务中成为“当务之急”。

已有的中文社交媒体命名实体识别研究都是使用字粒度的Weibo语料 (Peng and Dredze, 2015)，本文同样的在Weibo语料上进行实验。此外，考虑到词粒度的语料包含更丰富且情境化的序列信息，我们根据Weibo语料中每个字的位置特征对语料进行了重新标注，整理出按词粒度划分的新Weibo语料。进一步的，本文将字粒度和词粒度置于同等维度作为编码层的输入，获得了较好的文本序列表示，有效缓解了中文边界模糊的问题。

本文的主要贡献如下：1)本文重新标注并整理出词粒度的中文Weibo语料，已开源供大家使用[0]；2)本文首次将轻量级Star-Transformer模型应用到中文社交媒体NER任务中，并且利用Highway Networks机制使Star-Transformer更高效地适配NER任务，取得了可观的性能。实验结果表明，本文提出的基于字词粒度的Star-Transformer with Highway Networks(STHN)模型可以大幅提升Transformer在社交媒体命名实体识别上的性能，并且在Weibo语料上取得目前最好的性能。

2 相关研究

RNN一类的神经网络模型由于其顺序特征而被应用在NLP任务中，而其中应用最为广泛的BiLSTM模型已经成为主流编码器。Huang et al. (2015)等首先引入BiLSTM和CRF模型来解决序列标注问题，从那时起，BiLSTM模型被广泛应用于NER领域，Chiu and Nichols (2016)、Dong et al. (2016)、Lample et al. (2016)以及Ma and Hovy (2016)的研究都是基于此模型。

在字词编码方面，已有研究均是以字粒度为主，词粒度为辅，没有将二者放在同等维度上考虑。具有代表性的相关研究有：Zhang and Yang (2018)引入Lattice结构将所有与词典匹配的潜在单词信息整合到字符序列中，获得较好的向量表示；Gong et al. (2020)以字粒度作为输入，将词语边界信息融入到BiLSTM和CRF中，以此弥补字粒度输入信息不足的缺点；Gui et al. (2019)利用CNN对不同窗口大小的潜在单词进行编码；Peng et al. (2019)优化了Lattice结构的潜在词向量表示，获得了较快的运算速度以及更好的命名实体识别性能。

尽管BiLSTM模型在NER领域获得了不小的成就，但是它必须逐一计算token的表示，这极大地阻碍了GPU的并行性利用，而且BiLSTM无法捕捉到较长距离的上下文依赖关系。2017年以来，Transformer (Vaswani et al., 2017)逐渐在NLP各个任务中占据主导地位，例如机器翻译 (Vaswani et al., 2017)、语言建模 (Radford et al., 2018)以及预训练模型 (Devlin et al.,

[0]词粒度划分的weiboNER语料：https://github.com/cchen-nlp/weiboNER

2019)等等。然而Transformer在NER任务中效果不佳，Yan et al. (2019)提出了TENER模型，引入了方向感知、距离感知和无比例关注，同时定制了命名实体识别专属的Transformer编码层，使得Transformer在NER任务上获得较好性能。Li et al. (2020)的FLAT模型将Lattice结构转换为平面结构，并且设计合适的Transformer位置编码，进一步提升了NER性能。

本文结合字与词粒度各自的优势，将二者放在同等维度作为下层编码器的输入，同时引入Star-Transformer代替传统模型中的BiLSTM模型，加入Highway Networks机制进行结点信息的自我桥接，通过Star-Transformer特有的注意力连接和门控机制的动态调整服务于社交媒体领域的命名实体识别。

3 基于Star-Transformer 的命名实体识别框架

将命名实体识别看作是序列标注问题之后，实体采用BMES规则标注，实体的开头标注为B(Beginning)，实体内部单元标注为M(Median)，实体的结尾标注为E(End)，其他的词标注为O(Other)。

图 1给出了本文提出的基于字词粒度的Star-Transformer with Highway Networks模型完整框架，从图中可以看出STHN模型可以分为两个部分：第一部分是字词粒度的嵌入式表示，模型以同等维度的字向量和词向量作为输入，其中字向量需要经过Self-Attention做初步处理；第二部分是STHN模型。下面逐个介绍STHN模型中各个组成部分。

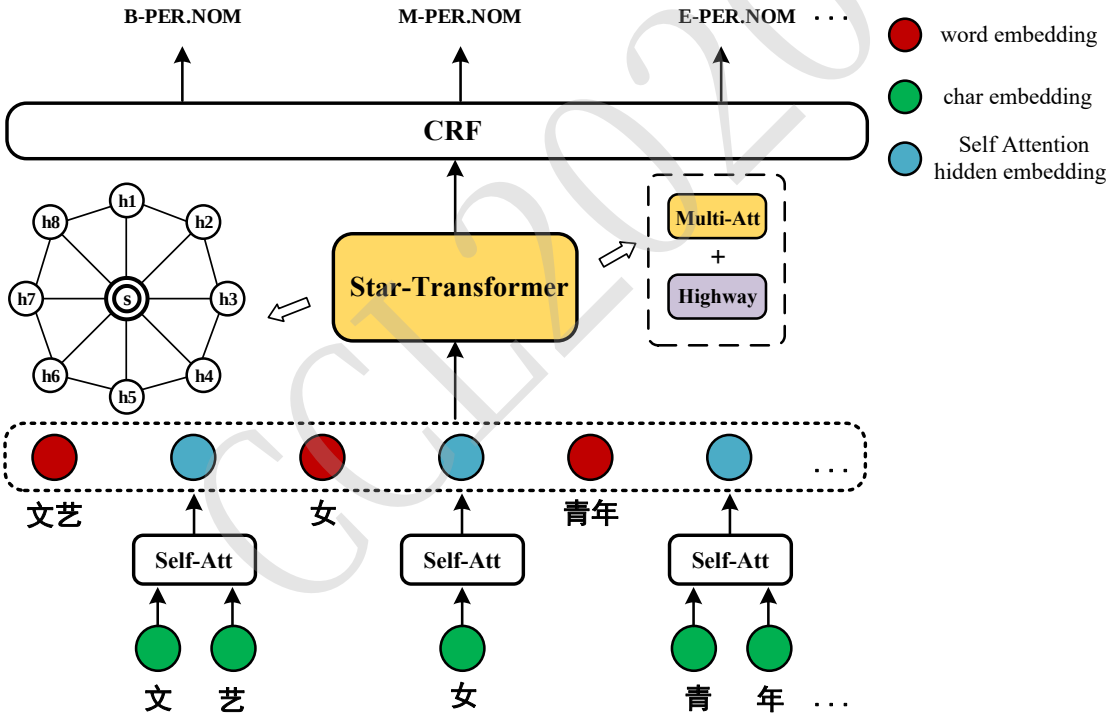

图 1. STHN模型图

3.1 字、词粒度的模型输入

在编码阶段，原始数据通过查找字或词向量表转化为字或词向量序列。对于文本中的字与词的向量表示，我们使用2018年预训练的词向量 (Li et al., 2018)，该词向量使用Word2vec 中的Skip-Gram模型训练，维度为300。该词向量包括百度百科、中文维基百科、人民日报、微博、知乎等多领域的字词特征，具体的模型训练设置如表 1所示。

字词向量表查找的过程是让原始文本中每一个字符或者单词在表上查找相对应的字词向量，如果某个字符或单词在表中不存在，则被赋予一个随机值。

考虑到现有研究都是以字粒度作为编码层的输入，并且He and Wang (2008)、Liu et al. (2010)和Li et al. (2014)的工作验证了基于字粒度要优于词粒度，但是基于字粒度的嵌入式表示

Window Size	Dynamic Window	Sub-sampling
5	Yes	1e-5
Low-Frequency Word	**Iteration**	**Negative Sampling**
10	5	5

表 1. SGNS模型训练参数设置

存在识别结果的标签不连续的情况，而基于词粒度的嵌入式表示具有显式的词汇边界，可以有效缓解社交媒体语料中中的边界模糊问题。本文将字粒度与词粒度放在同等维度上作为输入，其中字粒度需要先经过Self-Attention做初步特征提取，这部分与Yan et al. (2019)是相同的。

本文模型中最终编码层的输入是词粒度的嵌入式表示与经过特征提取的字粒度嵌入式表示的结合，如公式(1)∼(2)所示:

$$char^{'} = SelfAtt(char) \tag{1}$$

$$h_i = [word_i; char_i^{'}] \tag{2}$$

3.2 Star-Transformer模型

传统Transformer的注意力连接为全连接结构，如图 2(a)所示，而命名实体识别任务旨在识别出特定含义的实体，并且社交媒体语料中实体密度较稀疏，并不需要时刻关注句子序列中所有的结点，即传统Transformer的全连接结构在命名实体识别任务里存在信息冗余的现象，这些多余的信息不仅会降低运算速度，甚至会对命名实体识别任务起到反作用。因此传统Transformer对于社交媒体的实体识别任务来说并不合适。为了降低模型的复杂性，Guo et al. (2019)提出用星型拓扑结构代替全连通结构来简化架构，如图 2(b)所示。其中每两个相邻结点通过一个共享中继结点进行连接，因此，模型复杂性从二次降低到线性，同时保留捕获局部成分和长期依赖关系的能力。本节将详细介绍Star-Transformer的相关内容。

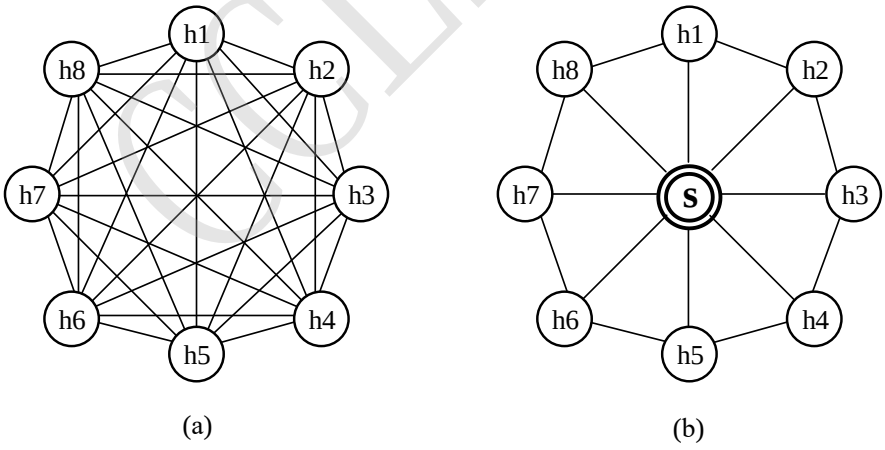

图 2. 传统Transformer(a)与Star-Transformer(b)结点连接方式图

3.2.1 Multi-Head Attention

Transformer (Vaswani et al., 2017)首先使用h个注意力头对一个输入序列分别进行单独的自我注意，然后对每个注意力头进行连接和线性变换操作，称为多头注意力机制（Multi–Head Attention）。一般来说，多头注意力机制可以用查询（query）到一系列键（key）值（value）对的映射来描述。

首先介绍放缩点积注意力Scaled Dot-product Attention，其本质上是使用了点积进行相似度计算。给定一个向量序列X，我们可以使用一个查询向量Q软选择相关信息，如公

式(3)∼(4)所示：

$$Attention(Q,K,V) = softmax(\frac{QK^T}{\sqrt{d_k}})) \cdot V \tag{3}$$

$$K = XW^K, V = XW^V \tag{4}$$

其中，W^K，W^V是对应向量的学习参数。然后我们可以将多头注意力机制定义成公式(5)∼(6)：

$$MultiAtt = (z_1 \oplus z_2 \oplus \cdots \oplus z_h) \cdot W^o \tag{5}$$

$$z_i = Attention(QW_i^Q, KW_i^K, VW_i^V) \tag{6}$$

其中，$\oplus$表示向量连接操作，W^o, W_i^Q, W_i^K, W_i^V是对应向量的学习参数。

3.2.2 Star-Transformer Encoder

Star-Transformer (Guo et al., 2019)的星型拓扑结构如图 2(b)所示，由一个中继结点s和n个卫星结点组成。第i个卫星结点h_i的状态表示文本序列中第i个token的特征。中继结点s充当虚拟中心，在所有卫星节点之间收集和散布信息。

Star-Transformer提出了基于time step的循环更新方式：每个token由输入向量初始化，中继结点初始为所有token的平均值，每个token依次通过多头注意力机制更新。在更新卫星结点的时候，每个卫星结点h_i的状态根据其相邻的结点更新，包括上一轮的上一个结点的隐态h_{i-1}^{t-1}；上一轮该结点的隐态h_i^{t-1}；上一轮下一个结点的隐态h_{i+1}^{t-1}；本结点的向量表示e^i；上一轮的中继结点状态s^{t-1}，具体过程如公式(7)∼(9)所示：

$$C_i^t = [h_{i-1}^{t-1}; h_i^{t-1}; h_{i+1}^{t-1}; e^i; s^{t-1}] \tag{7}$$

$$h_i^t = MultiAtt(h_i^{t-1}, C_i^t, C_i^t) \tag{8}$$

其中，C表示第i个卫星结点的上下文信息，在更新完信息后，使用层归一化操作处理卫星节点信息：

$$h_i^t = LayerNorm(ReLU(h_i^t)) \tag{9}$$

在更新中继结点s时，中继结点s将汇总所有卫星结点h_i的信息以及之前的状态，如公式(10)∼(11)所示：

$$s^t = MultiAtt(s^{t-1}, [s^{t-1}; H^t], [s^{t-1}; H^t]) \tag{10}$$

$$s^t = LayerNorm(ReLU(s^t)) \tag{11}$$

通过交替更新卫星结点h_i和中继结点s的信息，Star-Transformer在减少了注意力连接的前提下，依旧可以捕获句子序列中的局部特征和长期依赖关系，能够较好地融入到命名实体识别任务中。

3.3 Highway Networks

高速网络（Highway Networks）(Srivastava et al., 2015)是一种能够在信息传递之间进行平滑切换的神经网络，它能够有效解决网络深度加深，梯度信息回流受阻，造成网络训练困难的问题。Dauphin et al. (2017)、Gehring et al. (2017)以及Wu et al. (2019)验证了LSTM类型的门控单元在序列学习任务中是有效的。而Chai et al. (2020)证明了高速网络类型的门控机制有助于增强Transformer组件。

考虑到Star-Transformer用星型拓扑结构代替全连通结构，减少了相对较多的计算，我们对Star-Transformer中的每个卫星结点h_i利用高速网络进行信息的自我桥接，使得每一层Star-Transformer都能够充分利用上一层的卫星节点信息，这样的自我桥接可以看作是特征的动态调整，图 3给出了我们在Star-Transformer内部加入的高速网络结构。

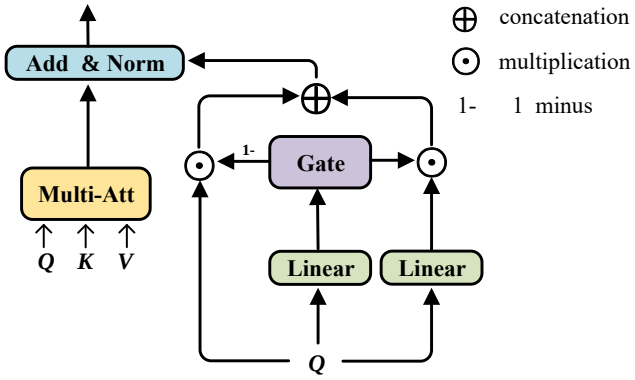

图 3. 高速网络结构图

我们在Star-Transformer计算完多头注意力之后，进入层归一化之前，加入一个新的输入分支$HW(h_i)$，这个输入分支就是上文所说的卫星结点h_i的自我桥接，如公式(12)∼(14)所示：

$$gated = \sigma(w_1 h_i + b_1) \tag{12}$$

$$f(h_i) = w_2 h_i + b_2 \tag{13}$$

$$HW(h_i) = [(1 - gated) \cdot h_i + gated \cdot f(h_i)] \tag{14}$$

其中，w_1, w_2表示门控机制的权重参数，b_1, b_2表示门控机制的偏差参数，σ为激活函数。

接着我们将使用高速网络增强的表征来丰富原有的多头注意力结果，如公式(15)所示：

$$H_i = LayerNorm(HW(h_i) + MultiAtt(h_i, C_i, C_i)) \tag{15}$$

最后，经过高速网络进行自我桥接之后的卫星节点信息和多头注意力计算结果相加并经过层归一化，得到新的卫星节点H_i。

4 实验设置和结果分析

本文使用命名实体识别社交媒体领域的Weibo数据集，通过不同的设置对前文所述的模型进行实验，并对实验结果进行讨论与分析，最终采用准确率$Precision$、召回率$Recall$和综合指标$Micro - F_1$值 (DBL, 1992)对标注结果进行评价。

4.1 实验数据集

本文采用了Peng and Dredze (2015)公开的Weibo NER语料，该语料是按照字粒度划分的，我们根据语料中的位置特征，整理出了对应的按照词粒度划分的Weibo NER语料，且将标注方式从BIO标注转换成了BMESO标注。

我们在整理语料过程中还引入了词性标注（part-of-speech tagging）特征，希望能够通过对语料中不同词性的区别来优化命名实体识别的结果。语料中采用Stanford Parser的词性标注器进行标注，使用的模型是chinese-distsim.tagger (de Marneffe et al., 2014)。我们对比了将一整个句子进行标注的方式以及对单个词标注的方式，最终采用更加准确的融合句法信息的标注方式。为了模型对比的公平性，本文的实验部分没有使用词性标注等外部信息。

更新后的Weibo NER语料的字粒度结构不变，包含训练集、开发集和测试集共1890句。表 2详细地给出了该语料原本的字粒度结构以及我们整理之后词粒度结构，从中我们可以清晰地看到Weibo语料规模较小，带标记字符的数量表明了待识别的实体数目也相对较少。

Weibo NER语料标注的实体类型包括PER、ORG、LOC和GPE，且每个类型分别有特定实体（named entity, NE）和指代实体（nominal mention, NM），表 3给出了Weibo NER语料中各个类别的分布情况。特定实体即为传统领域中需要识别出来的实体，比如人名的特定实体

Type	Train	Dev	Test
Sentence	1350	270	270
Character	73378	14509	14842
Word	45678	9026	9143
Char with label	4951	971	1078
Label percent	6.71%	6.69%	7.26%

表 2. Weibo NER数据集结构

有詹天佑、钱钟书以及舒淇等，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阿姨、妹纸以及皇上等。社交媒体类语料中常常会有特定实体和指代实体混合出现的情况，这是其与规范的新闻类语料差别最大的地方，这种特殊的结构无疑增加了社交媒体领域命名实体识别的难度。

Type	Train	Dev	Test	All
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

表 3. Weibo NER数据集Label分布

4.2 实验参数设置

本文实验采用Pytorch 0.4.1框架，并用NVIDIA的1080GPU进行加速。使用的预训练词向量参数在表 1中已经给出，模型的查询表使用预训练得到的向量进行初始化，其他参数均采用均匀分布的随机函数初始化。

表 4给出了模型的参数值，我们使用Adam（Adaptive moment estimation）来优化所有可训练的参数；为了保证字词二者的同一性，使用的字词嵌入式表示维度都是300；神经网络的隐藏层维度均设为300；多头注意力机制的头数head为5(维度300可被head整除)；Star-Transformer层数为6层；整个模型的学习率learning rate设置为0.0005，学习率减少步长lr_decay设置为0.05，所有神经网络的dropout设置为0.5，L2正则化参数设置为1e-8。

Parameter	Value	Parameter	Value
char emb size	300	learning rate	0.0005
word emb size	300	lr_decay	0.05
hidden dim	300	dropout	0.5
Multi head	5	batch size	10
star layer	6	regularization	1e-8

表 4. 超参数设置

4.3 实验结果及分析

表 5给出了本文的模型在社交媒体Weibo语料上的实验结果对比，其中STAR和STHN分别表示本文提出的基于Star-Transformer的模型以及利用高速网络优化的Star-Transformer的模型。

	Level	Models	NE(%)	NM(%)	Overall(%)
Peng and Dredze (2015)	char	CRF	51.96	61.05	56.05
Peng and Dredze (2016)	char	LSTM	55.28	62.97	58.99
He and Sun (2017a)	char	LSTM	50.60	59.32	54.82
He and Sun (2017b)	char	LSTM	54.50	62.17	58.23
Zhang and Yang (2018)	char+Lattice	LSTM	53.04	62.25	58.79
Gui et al. (2019)	char+Lattice	CNN	**57.14**	**66.67**	59.92
Peng et al. (2019)	char+Lattice	LSTM	56.99	61.41	**61.24**
Yan et al. (2019)	char	Transformer	–	–	58.39
Li et al. (2020)	char+Lattice	Transformer	–	–	**63.42**
Our work	char	LSTM	53.16	60.70	55.76
	char	Transformer	46.90	53.45	48.96
	char	STAR	51.28	62.02	55.08
	char	STHN	52.32	64.53	56.63
	char+word	LSTM	58.82	69.32	64.88
	char+word	Transformer	53.02	64.11	59.79
	char+word	STAR	57.87	**72.04**	66.58
	char+word	STHN	**61.58**	69.45	**68.15**

表 5. 中文社交媒体命名实体识别实验结果对比(F_1)

STHN模型模型在特定实体NE的识别上性能达到了61.58%，比已有最好的模型结果高出了约4.44%；对于Weibo语料中特有的指代实体NM的识别，Star-Transformer模型获得了72.04%的F_1值；整体上本文提出的STHN模型取得了目前最好的综合性能68.15%，比之前最好的FLAT模型高出4.73%。

社交媒体类Weibo语料没有规范的文本内容，这使得词与词的边界更加模糊，比任何领域都迫切地需要词粒度信息的输入。就传统的LSTM模型而言，我们基于字与词粒度的实验已经有了不小的突破，综合性能为64.88%，比同样以LSTM作为主模型的Lexicon结构 (Peng et al., 2019)高了约3.64%。类似的，在融入了词粒度信息后，相同的模型在NE、NM以及整体上都能获得明显的提升。

表 6给出了本文实验的详细结果，在两种粒度上，STAR模型的三个指标都是明显高于Transformer模型的，这进一步验证了Star-Transformer在命名实体识别任务上的有效性。除此以外，基于字词粒度的STAR模型在召回率R值上作用显著，比相同条件下的LSTM模型高了约6.17%，而引入高速网络的STHN模型又比STAR模型高出了约1.24%。

Level	Models	P(%)	R(%)	F1(%)
Char	LSTM	60.86	51.45	55.76
	Transformer	57.70	42.51	48.96
	STAR	58.95	51.69	55.08
	STHN	60.00	**53.62**	**56.63**
Char + Word	LSTM	**75.66**	56.79	64.88
	Transformer	65.40	55.06	59.79
	STAR	70.64	62.96	66.58
	STHN	72.63	**64.20**	**68.15**

表 6. 详细实验结果对比

但是Star-Transformer的P值明显低于LSTM模型，虽然Star-Transformer已经是轻量级的Transformer，但本质上还是一个多连接计算注意力的模型，正是这样的机制使得Star-Transformer充分理解了句子的结构，识别出了更多的实体、提升了实体类型判别的准确度，但

同时存在过度识别的现象，从而导致P值的降低。

从解决上述问题的角度考虑，我们利用高速网络对Star-Transformer的每个卫星节点进行信息的自我桥接，实验结果显示STHN模型相较于STAR模型在P值上提升了近2%，拉近了与LSTM模型的距离。与此同时，STHN模型进一步提升了R值和F_1值。由此可见，高速网络的门控机制可以有效缓解Transformer的过度识别问题，同时带来命名实体识别性能的提升。

4.4 NE、NM结果对比分析

表 7给出了三个模型分别在特定实体(NE)和指代实体(NM)上面的实验结果。对于特定实体NE来说，整体趋势和表 6实验结果一致，LSTM仅在P值上占优势；Star-Transformer的引入带来了R值的大幅度提升，约为4.27%；我们最终的STHN模型进一步优化了Star-Transformer，在R和F_1值上都达到了最高值，虽然P值没有超越LSTM，但是已经尽可能将差距最小化。

表 7NM相关的数据体现了Star-Transformer对指代实体NM的识别性能，STAR模型在P、R和F_1三个指标上都比STHN模型的结果高。此外，STAR模型的R值比LSTM模型高出了约6.18%，可见相较于特定实体，Star-Transformer更适合用在指代实体NM的识别上。这也进一步验证了我们将Star-Transformer应用到社交媒体领域命名实体识别的有效性。

Models	NE			NM		
	P(%)	R(%)	F1(%)	P(%)	R(%)	F1(%)
LSTM	**71.92**	49.76	58.82	**77.22**	62.89	69.32
STAR	62.30	54.03	57.87	75.28	**69.07**	**72.04**
STHN	69.23	**55.45**	**61.58**	70.37	68.56	69.45

表 7. NE和NM详细结果分析

4.5 STHN效用分析

我们进一步统计并分析了Weibo语料上的实验结果，发现结合了Highway Networks的Star-Transformer模型能够识别出更多的Single类实体，表 8展示了相关数据。STHN模型比LSTM多识别出了67个Single实体，在数据量较小的weibo语料中占了约20%。这样超强的学习能力在带来性能提升的同时，也存在着过度识别的问题——将一些本不是实体的词识别为实体，从而导致P值的降低。

	总数	LSTM	STAR	STHN
Single	327	258	299	**325**
Else	175	98	**142**	89

表 8. 识别结果分析

表 9列举了几个在实验结果中出现的典型案例，很多在LSTM模型中被预测错误的实体，STHN模型能够准确地将其预测出来。Star-Transformer的连接机制能够捕获丰富的序列上下文信息，使得模型更好地理解句子结构，从而能够正确识别出更多的实体，这也是使R值大幅提升的关键。

Sentence	1.平洲玉器街... 2.中国女足打好基础再说吧...			
Word	平洲	玉器	街	中国女足
LSTM	B-LOC.NAM	M-LOC.NAM	E-LOC.NAM	S-PER.NAM
STHN	**S-LOC.NAM**	**B-LOC.NAM**	**E-LOC.NAM**	**S-ORG.NAM**

表 9. 识别案例分析

5 结论

本文根据公开的Weibo字粒度语料划分出了Weibo词粒度语料，并且提出了字词融合的方法，将字粒度与词粒度放在同等维度上加以考虑作为下层神经网络的输入，获得词语边界明确的句子表征。此外，我们分析了传统Transformer在命名实体识别任务上的劣势，并首次将Star-Transformer应用到社交媒体领域的命名实体识别任务中。由于Star-Transformer独特的星型拓扑结构，以及Highway Networks的动态特征调整，我们的STHN模型能够较好地理解句子序列的上下文信息，正确识别出更多的实体，在社交媒体领域的Weibo语料上取得了目前最好的效果。

社交媒体领域的Weibo语料规模较小，未来可以考虑将STHN模型应用到更多领域中，比如新闻领域，利用规模较大的语料深入研究Transformer在命名实体识别任务中的应用。

参考文献

Yekun Chai, Jin Shuo, and Xinwen Hou. 2020. Highway transformer: Self-gating enhanced self-attentive networks. *CoRR*, abs/2004.08178.

Jason P. C. Chiu and Eric Nichols. 2016. Named entity recognition with bidirectional lstm-cnns. *Trans. Assoc. Comput. Linguistics*, 4:357–370.

Kevin Clark and Christopher D. Manning. 2016. Improving coreference resolution by learning entity-level distributed representations. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, ACL 2016, August 7-12, 2016, Berlin, Germany, Volume 1: Long Papers*. The Association for Computer Linguistics.

Yann N. Dauphin, Angela Fan, Michael Auli, and David Grangier. 2017. Language modeling with gated convolutional networks. In Doina Precup and Yee Whye Teh, editors, *Proceedings of the 34th International Conference on Machine Learning, ICML 2017, Sydney, NSW, Australia, 6-11 August 2017*, volume 70 of *Proceedings of Machine Learning Research*, pages 933–941. PMLR.

1992. *Proceedings of the 4th Conference on Message Understanding, MUC 1992, McLean, Virginia, USA, June 16-18, 1992*. ACL.

Marie-Catherine de Marneffe, Timothy Dozat, Natalia Silveira, Katri Haverinen, Filip Ginter, Joakim Nivre, and Christopher D. Manning. 2014. Universal stanford dependencies: A cross-linguistic typology. In Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Hrafn Loftsson, Bente Maegaard, Joseph Mariani, Asunción Moreno, Jan Odijk, and Stelios Piperidis, editors, *Proceedings of the Ninth International Conference on Language Resources and Evaluation, LREC 2014, Reykjavik, Iceland, May 26-31, 2014*, pages 4585–4592. European Language Resources Association (ELRA).

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: pre-training of deep bidirectional transformers for language understanding. In Jill Burstein, Christy Doran, and Thamar Solorio, editors, *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019, Minneapolis, MN, USA, June 2-7, 2019, Volume 1 (Long and Short Papers)*, pages 4171–4186. Association for Computational Linguistics.

Chuanhai Dong, Jiajun Zhang, Chengqing Zong, Masanori Hattori, and Hui Di. 2016. Character-based LSTM-CRF with radical-level features for chinese named entity recognition. In Chin-Yew Lin, Nianwen Xue, Dongyan Zhao, Xuanjing Huang, and Yansong Feng, editors, *Natural Language Understanding and Intelligent Applications - 5th CCF Conference on Natural Language Processing and Chinese Computing, NLPCC 2016, and 24th International Conference on Computer Processing of Oriental Languages, ICCPOL 2016, Kunming, China, December 2-6, 2016, Proceedings*, volume 10102 of *Lecture Notes in Computer Science*, pages 239–250. Springer.

Jonas Gehring, Michael Auli, David Grangier, Denis Yarats, and Yann N. Dauphin. 2017. Convolutional sequence to sequence learning. In Doina Precup and Yee Whye Teh, editors, *Proceedings of the 34th International Conference on Machine Learning, ICML 2017, Sydney, NSW, Australia, 6-11 August 2017*, volume 70 of *Proceedings of Machine Learning Research*, pages 1243–1252. PMLR.

Zhaoheng Gong, Ping Chen, and Jiang Zhou. 2020. Integrating boundary assembling into a DNN framework for named entity recognition in chinese social media text. *CoRR*, abs/2002.11910.

Tao Gui, Ruotian Ma, Qi Zhang, Lujun Zhao, Yu-Gang Jiang, and Xuanjing Huang. 2019. Cnn-based chinese NER with lexicon rethinking. In Sarit Kraus, editor, *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI 2019, Macao, China, August 10-16, 2019*, pages 4982–4988. ijcai.org.

Qipeng Guo, Xipeng Qiu, Pengfei Liu, Yunfan Shao, Xiangyang Xue, and Zheng Zhang. 2019. Star-transformer. In Jill Burstein, Christy Doran, and Thamar Solorio, editors, *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019, Minneapolis, MN, USA, June 2-7, 2019, Volume 1 (Long and Short Papers)*, pages 1315–1325. Association for Computational Linguistics.

Hangfeng He and Xu Sun. 2017a. F-score driven max margin neural network for named entity recognition in chinese social media. In Mirella Lapata, Phil Blunsom, and Alexander Koller, editors, *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics, EACL 2017, Valencia, Spain, April 3-7, 2017, Volume 2: Short Papers*, pages 713–718. Association for Computational Linguistics.

Hangfeng He and Xu Sun. 2017b. A unified model for cross-domain and semi-supervised named entity recognition in chinese social media. In Satinder P. Singh and Shaul Markovitch, editors, *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, February 4-9, 2017, San Francisco, California, USA*, pages 3216–3222. AAAI Press.

Jingzhou He and Houfeng Wang. 2008. Chinese named entity recognition and word segmentation based on character. In *Third International Joint Conference on Natural Language Processing, IJCNLP 2008, Hyderabad, India, January 7-12, 2008*, pages 128–132. The Association for Computer Linguistics.

Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural Computation*, 9(8):1735–1780.

Zhiheng Huang, Wei Xu, and Kai Yu. 2015. Bidirectional LSTM-CRF models for sequence tagging. *CoRR*, abs/1508.01991.

Guillaume Lample, Miguel Ballesteros, Sandeep Subramanian, Kazuya Kawakami, and Chris Dyer. 2016. Neural architectures for named entity recognition. *CoRR*, abs/1603.01360.

Haibo Li, Masato Hagiwara, Qi Li, and Heng Ji. 2014. Comparison of the impact of word segmentation on name tagging for chinese and japanese. In Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Hrafn Loftsson, Bente Maegaard, Joseph Mariani, Asunción Moreno, Jan Odijk, and Stelios Piperidis, editors, *Proceedings of the Ninth International Conference on Language Resources and Evaluation, LREC 2014, Reykjavik, Iceland, May 26-31, 2014*, pages 2532–2536. European Language Resources Association (ELRA).

Shen Li, Zhe Zhao, Renfen Hu, Wensi Li, Tao Liu, and Xiaoyong Du. 2018. Analogical reasoning on chinese morphological and semantic relations. In Iryna Gurevych and Yusuke Miyao, editors, *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, ACL 2018, Melbourne, Australia, July 15-20, 2018, Volume 2: Short Papers*, pages 138–143. Association for Computational Linguistics.

Xiaonan Li, Hang Yan, Xipeng Qiu, and Xuanjing Huang. 2020. FLAT: chinese NER using flat-lattice transformer. *CoRR*, abs/2004.11795.

Zhangxun Liu, Conghui Zhu, and Tiejun Zhao. 2010. Chinese named entity recognition with a sequence labeling approach: Based on characters, or based on words? In De-Shuang Huang, Xiang Zhang, Carlos A. Reyes García, and Lei Zhang, editors, *Advanced Intelligent Computing Theories and Applications. With Aspects of Artificial Intelligence, 6th International Conference on Intelligent Computing, ICIC 2010, Changsha, China, August 18-21, 2010. Proceedings*, volume 6216 of *Lecture Notes in Computer Science*, pages 634–640. Springer.

Xuezhe Ma and Eduard H. Hovy. 2016. End-to-end sequence labeling via bi-directional lstm-cnns-crf. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, ACL 2016, August 7-12, 2016, Berlin, Germany, Volume 1: Long Papers*. The Association for Computer Linguistics.

Nanyun Peng and Mark Dredze. 2015. Named entity recognition for chinese social media with jointly trained embeddings. In Lluís Màrquez, Chris Callison-Burch, Jian Su, Daniele Pighin, and Yuval Marton, editors, *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, EMNLP 2015, Lisbon, Portugal, September 17-21, 2015*, pages 548–554. The Association for Computational Linguistics.

Nanyun Peng and Mark Dredze. 2016. Improving named entity recognition for chinese social media with word segmentation representation learning. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, ACL 2016, August 7-12, 2016, Berlin, Germany, Volume 2: Short Papers*. The Association for Computer Linguistics.

Minlong Peng, Ruotian Ma, Qi Zhang, and Xuanjing Huang. 2019. Simplify the usage of lexicon in chinese NER. *CoRR*, abs/1908.05969.

Alec Radford, Karthik Narasimhan, Tim Salimans, and Ilya Sutskever. 2018. Improving language understanding by generative pre-training.

Revanth Reddy, Danish Contractor, Dinesh Raghu, and Sachindra Joshi. 2019. Multi-level memory for task oriented dialogs. In Jill Burstein, Christy Doran, and Thamar Solorio, editors, *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019, Minneapolis, MN, USA, June 2-7, 2019, Volume 1 (Long and Short Papers)*, pages 3744–3754. Association for Computational Linguistics.

Rupesh Kumar Srivastava, Klaus Greff, and Jürgen Schmidhuber. 2015. Highway networks. *CoRR*, abs/1505.00387.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *CoRR*, abs/1706.03762.

Felix Wu, Angela Fan, Alexei Baevski, Yann N. Dauphin, and Michael Auli. 2019. Pay less attention with lightweight and dynamic convolutions. In *7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019*. OpenReview.net.

Hang Yan, Bocao Deng, Xiaonan Li, and Xipeng Qiu. 2019. TENER: adapting transformer encoder for named entity recognition. *CoRR*, abs/1911.04474.

Wojciech Zaremba, Ilya Sutskever, and Oriol Vinyals. 2014. Recurrent neural network regularization. *CoRR*, abs/1409.2329.

Dmitry Zelenko, Chinatsu Aone, and Anthony Richardella. 2003. Kernel methods for relation extraction. *Journal of machine learning research*, 3(Feb):1083–1106.

Yue Zhang and Jie Yang. 2018. Chinese NER using lattice LSTM. In Iryna Gurevych and Yusuke Miyao, editors, *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, ACL 2018, Melbourne, Australia, July 15-20, 2018, Volume 1: Long Papers*, pages 1554–1564. Association for Computational Linguistics.