



结合单词-字符引导注意力网络的中文旅游文本命名实体识别

西尔艾力·色提^{1,2}, 艾山·吾买尔^{1,2}, 王路路^{1,2}, 吐尔根·依布拉克^{1,2},
马喆康^{2,3}, 买合木提·买买提^{1,2}

(1.新疆大学 信息科学与工程学院, 乌鲁木齐 830046; 2.新疆大学 新疆多语种信息技术重点实验室, 乌鲁木齐 830046;
3.新疆大学 软件学院, 乌鲁木齐 830046)

摘 要:传统基于词向量表示的命名实体识别方法通常忽略了字符语义信息、字符间的位置信息,以及字符和单词间的关联关系。提出一种基于单词-字符引导注意力网络(WCGAN)的中文旅游命名实体识别方法,利用单词引导注意力网络获取单词间的序列信息和关键单词信息,采用字符引导注意力网络捕获字符语义信息和字符间的位置信息,增强单词和字符间的关联性与互补性,从而实现中文旅游文本中命名实体的识别。实验结果表明,WCGAN方法在 ResumeNER 和 TourismNER 基准数据集上的 F 值分别为 93.491% 和 92.860%,相比 Bi-LSTM+CRF、Char-Dense 等方法识别效果更好。

关键词:命名实体识别;字符引导注意力网络;单词引导注意力网络;字符语义;信息互补;位置信息

开放科学(资源服务)标志码(OSID):



中文引用格式:西尔艾力·色提,艾山·吾买尔,王路路,等.结合单词-字符引导注意力网络的中文旅游文本命名实体识别[J].计算机工程,2021,47(2):39-45.

英文引用格式:Xieraili Seti, Aishan Wumaier, WANG Lulu, et al. Named entity recognition for Chinese tourism texts combining word-character guided attention network[J]. Computer Engineering, 2021, 47(2): 39-45.

Named Entity Recognition for Chinese Tourism Texts Combining Word-Character Guided Attention Network

Xieraili Seti^{1,2}, Aishan Wumaier^{1,2}, WANG Lulu^{1,2}, Tuergeren Yibulayin^{1,2}, MA Zhekang^{2,3}, Maihemuti Maimaiti^{1,2}

(1.College of Information Science and Engineering, Xinjiang University, Urumqi 830046, China;
2.Xinjiang Key Laboratory of Multi-language Information Technology, Xinjiang University, Urumqi 830046, China;
3.College of Software, Xinjiang University, Urumqi 830046, China)

[Abstract] The traditional Named Entity Recognition (NER) methods based on word vector usually neglect the character semantics of Chinese characters, the position information between characters, and the dependence between characters and words. To address the problem, this paper proposes a NER method based on Word-Character Guided Attention Network (WCGAN) for Chinese tourism texts. The method uses the Word-Guided Attention Network (WGAN) to obtain the sequence information between words and further capture the significant word information. The Character Guided Attention Network (CGAN) is used to obtain the information about character semantics and position between characters, and thus enhance the relevance and complementarity between words and characters to realize the recognition of named entities in Chinese tourism texts. Experimental results on the two benchmark datasets of ResumeNER and TourismNER show that the F values of the WCGAN method are 93.491% and 92.860% respectively, and the proposed method has better recognition performance than Bi-LSTM+CRF, Char-Dense and other methods.

[Key words] Named Entity Recognition (NER); Character Guided Attention Network (CGAN); Word Guided Attention Network (WGAN); character semantics; information complementary; location information

DOI: 10.19678/j.issn.1000-3428.0056522

基金项目:国家自然科学基金(61262060, 61662077);国家重点研发计划(2017YFB1002103);新疆维吾尔自治区重点实验室开放课题(2018D04019)。

作者简介:西尔艾力·色提(1994—),男,硕士研究生,主研方向为自然语言处理、命名实体识别;艾山·吾买尔,副教授;王路路,博士研究生;吐尔根·依布拉克,教授、博士生导师;马喆康,硕士研究生;买合木提·买买提(通信作者),实验师、博士。

收稿日期:2019-11-07 **修回日期:**2020-01-16 **E-mail:**572984640@qq.com

0 概述

命名实体识别(Named Entity Recognition, NER)作为自然语言处理的基本任务之一,受到国内外研究人员的广泛关注,并且随着深度学习技术的不断发展,其取得了重要的研究成果。例如, COLLOBERT^[1]、LAMPLE^[2]等人将命名实体识别转换为序列标注问题,并利用深度学习技术捕获命名实体的词和字符信息,实现文本中命名实体的精准识别。通过将深度学习技术与字符信息相结合对文本序列标注数据进行建模,能够有效提高命名实体识别性能,然而现有中文命名实体识别方法^[3-4]多数使用单词向量特征对命名实体进行表示,忽略了汉字字符特征的表征能力。针对该问题,国内外研究人员也进行了一系列关于字符特征表示的研究。例如, LU^[5]、DONG^[6]等人利用字符信息和单词信息提高了命名实体的识别精度,但其没有考虑单词与字符以及字符与字符之间的位置关系。本文提出一种单词-字符引导注意力网络(Word-Character Guided Attention Network, WCGAN)的旅游文本命名实体识别方法。将不同层次的引导注意力网络进行集成学习以获取命名实体的单词和字符信息,同时捕获单词与字符、字符与字符之间的位置信息。

1 相关工作

传统NER方法通常采用机器学习技术或者人工标注特征^[7]进行命名实体识别。文献[8]提出一种基于SVM语料库检索模式和重复MWEs的NER识别方法,实验结果表明该方法优于基于规则的命名实体识别方法。文献[9]建立了条件随机场(Conditional Random Field, CRF)命名实体识别系统,在捷克语、西班牙语等语料库上的测试结果表明其具有较好的识别效果。然而,人工参数设计特征对上述方法的性能有较大影响,因此识别稳定性相对较差。

随着深度学习技术的快速发展,其在特征学习中表现出强大的自学习能力并在命名实体识别中得到广泛应用^[10-11]。文献[12]使用字符卷积神经网络(CharCNN)捕获了文本中的字符特征,并用字符特征表示命名实体,实现命名实体的准确识别。文献[13]使用字符序列标注方法对中文文本进行标注,提高了中文命名实体的识别精度。文献[14]将Max margin神经网络用于中文社交网络文本并识别出其中的命名实体。文献[15]提出一种半监督的命名实体识别模型,该模型在中文社交网络进行跨领域识别并取得较好的识别效果。文献[16]将双向LSTM与条件随机场相结合对文本序列进行标注,实现命名实体的准确识别。文献[17]提出一种任务感知神经语言模型,对文本序列中的命名实体进行权重赋值,提高了命名实体的识别精度。文献[18]

提出一种用于文本序列标记的半监督多任务学习方法,实现文本中命名实体的精确识别。上述方法虽然提高了命名实体的识别精度,但多数是基于单一特征对文本中的命名实体进行表示,忽略了大量的细节信息,而本文提出的WCGAN方法,从单词和字符两个方面对文本序列进行建模实现信息互补,同时强调单词与字符之间的关联性,进一步使用字符信息突出单词的关键性。

2 基于WCGAN的命名实体识别

汉字字符是单词的基本组成单位,其不仅具有独立的语义信息,而且可以通过汉字字符推断出单词的词义,即单词和汉字信息对中文命名实体的识别具有重要作用。如图1所示,本文将多个汉字字符按照一定的位置信息组合形成单词和命名实体。“山海”由“山”和“海”两个汉字字符组成,且具有一定的位置关系,而“山海”和“长城”两个单词经过前后位置的组合形成关于地址的命名实体。因此,字符及字符的位置信息有助于进一步突出单词信息,也可以提高字符与单词之间的关联性。在图1中, Char-Tag 和 Word-Tag 表示文本中命名实体的标签, B表示命名实体的初始单元, M表示命名实体的中间单元, E表示命名实体的结尾单元, O表示非命名实体, ComName表示旅游地址。

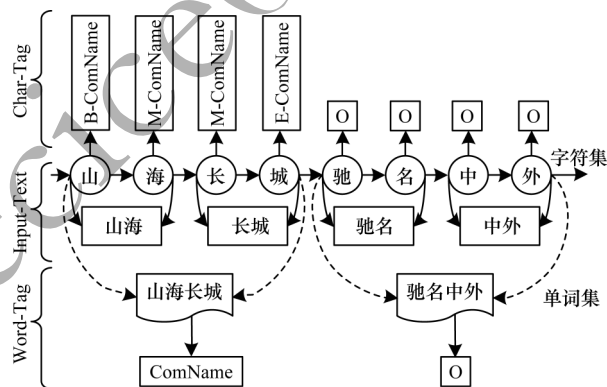


图1 单词和字符的标签信息

Fig.1 Label informations of words and characters

2.1 单词引导注意力网络模型

单词引导注意力网络(Word Guided Attention Network, WGAN)模型先借助词嵌入技术^[19]将文本中的每个单词(命名实体和非实体)映射到一个密集的低维向量空间中,再通过引导注意力的双向长短期记忆(Bi-directional Long Short-Term Memory, Bi-LSTM)网络捕获文本中命名实体的单词信息。Bi-LSTM网络^[6,20]能获取文本的上下文信息以及文本中命名实体的单词信息,使得单词与文本之间建立失联关系。WGAN模型的具体步骤如下:

1)在Bi-LSTM网络中, T 时刻隐藏层 H_t 的计算

公式如式(1)所示:

$$H_T = o_T \times \tanh C_T \quad (1)$$

其中, o_T 表示 T 时刻输出门, C_T 表示 T 时刻记忆单元的激活状态。 o_T 和 C_T 的计算公式如式(2)所示:

$$\begin{aligned} o_T &= \sigma(W_o[h_{T-1}, x_T] + b_o) \\ C_T &= f_T \times C_{T-1} + i_T \times \widehat{C}_T \end{aligned} \quad (2)$$

其中, W_o 表示输出门的权重矩阵, b_o 表示输出门的偏置值, f_T 表示 T 时刻忘记门的输出, i_T 表示 T 时刻记忆门的值, \widehat{C}_T 表示 T 时刻临时记忆单元的状态。相关计算公式如式(3)所示:

$$\begin{aligned} f_T &= \sigma(W_f[h_{T-1}, x_T] + b_f) \\ i_T &= \sigma(W_i[h_{T-1}, x_T] + b_i) \\ \widehat{C}_T &= \tanh(W_c[h_{T-1}, x_T] + b_c) \end{aligned} \quad (3)$$

2) 通过上述公式可得到第 j 步文本中命名实体的隐藏层特征集合 H_j^w 。

3) 将注意力网络(Attention)^[21-22] 嵌入隐藏层的最后一层, 进一步捕获命名实体的上下文语义, 同时突出关键词的显著性并分配注意力权重。利用单层感知机将 H_j^w 表示为 v_j^w , 计算公式如式(4)所示:

$$v_j^w = \tanh(W_w H_j^w + b_w) \quad (4)$$

其中, W_w 、 b_w 分别表示单层感知机的权重矩阵和偏置矩阵。通过注意力机制计算单词注意力系数 α_j^w , 如式(5)所示:

$$\alpha_j^w = \frac{\exp(v_j^w A^w)}{\sum_j \exp((v_j^w)^T A^w)} \quad (5)$$

其中, A^w 表示随机初始化单词的上下文向量。对单词的注意力系数进行加权求和, 得到命名实体中词的注意力表示形式 u^w , 如式(6)所示:

$$u^w = \sum_j \alpha_j^w H_j^w \quad (6)$$

4) 根据 WGAN 捕获文本中显著的命名实体单词, 最终得到命名实体的单词表示信息。

2.2 字符引导注意力网络模型

在中文文本中字符是单词的基本组成单位, 当文本的语义相差较小时, 字符信息显得尤为重要, 且字符之间的排列也会影响命名实体的准确提取。本文设计一种基于字符引导注意力网络(Character Guided Attention Network, CGAN)的字符信息提取模型。CGAN 模型的具体步骤如下:

1) 利用 CNN^[23] 和双向独立循环神经网络(Bi-IndRNN)^[24] 提取中文文本中命名实体的字符信息。

2) 采用位置注意力网络(Pos-Attention)进一步捕获字符信息, 同时获取命名实体中字符之间的位置信息, 在字符与字符之间建立位置关联, 从而提高命名实体的识别效果。

3) 替换 CNN 中卷积层预定义的 1-of- m 并对其进行编码^[23], 将预定义输入的字符长度设置为 225,

对超过预定义长度的字符进行裁剪, 空缺的字符用 0 进行填充, 计算公式如式(7)所示:

$$C_{xy}^l = \sum_{a=0}^{m-1} \sum_{b=0}^{m-1} W_{ab} y_{(x+a)(y+b)}^l \quad (7)$$

其中, 输入向量为 x_i , C_{xy}^l 表示第 l 个 CNN 层的输出特征, m 表示 CNN 层中卷积层的滤波器大小。

4) 基于池化层进一步捕获命名实体的字符信息。池化层的输出如式(8)所示:

$$P_{xy}^l = \text{Globalaverage } C_{xy}^l \quad (8)$$

其中, P_{xy}^l 表示池化层的输出, Globalaverage 表示全局平均池化, 本文将池化层和卷积层的步长设置为 1。

本文利用 CNN 提取命名实体中字符的空间信息, 为进一步获取命名实体中字符的时序信息, 将池化层输出的特征向量输入双向独立循环神经网络。双向独立循环神经网络层 T 时刻的输出如式(9)所示:

$$H_T = \sigma(WP_{xy}^l + U \odot H_{T-1} + b) \quad (9)$$

其中, σ 表示激活函数, U 表示输入层到隐藏层的权重矩阵。

将双向独立循环神经网络的输出作为位置注意力机制的输入^[25], 进一步捕获字符在上下文中的位置信息并分配相应的权重值, 以提高字符信息的表征能力和命名实体的识别精度, 具体计算公式如式(10)所示:

$$u^c = \text{SoftMax}(W_{s2} \tanh(W_{s1} H_T)) \quad (10)$$

其中, u^c 表示命名实体中字符的表示形式, W_{s2} 、 W_{s1} 表示权重矩阵。

通过字符引导注意力网络模型, 本文捕获了文本中命名实体的字符信息, 同时获取了字符之间的位置信息, 并在字符之间建立了关联关系。在使用 CNN 提取字符的空间信息时, 本文设计的 CharCNN 结构与文献[26]的 CharCNN 结构有所差异, 将原结构的最大池化层和全连接层变换为全局平均池化层和双向独立循环神经网络层。

2.3 WCGAN 方法

本文提出的 WCGAN 方法主要由 WGAN 和 CGAN 模型联合组成, 其能同时捕获文本命名实体的单词和字符信息, 并使两者之间形成信息互补, 同时使用字符信息增强命名实体的显著性, 进一步强调单词中字符之间的位置信息。本文首先使用 WGAN 方法学习文本的单词特征, 获得文本中命名实体的局部信息; 其次采用注意力机制引导 Bi-LSTM 对嵌入网络文本进行编码, 并对关键信息分配更高的权重比, 以降低冗余信息且使其更多关注显著性单词特征; 然后通过 CGAN 模型捕获文本中命名实体的字符信息, 使其更好地挖掘出不同命名实体之间的细微差别, 进一步捕获命名实体中字符之间的位置信息, 同时强调了字符与单词之间的关联性, 并与 WGAN 模型实现信息互补, 以增强命名实体之间的辨识度; 最后利用 CRF 对文本中的命名实体进行识别。WCGAN 网络结构如图 2 所示。

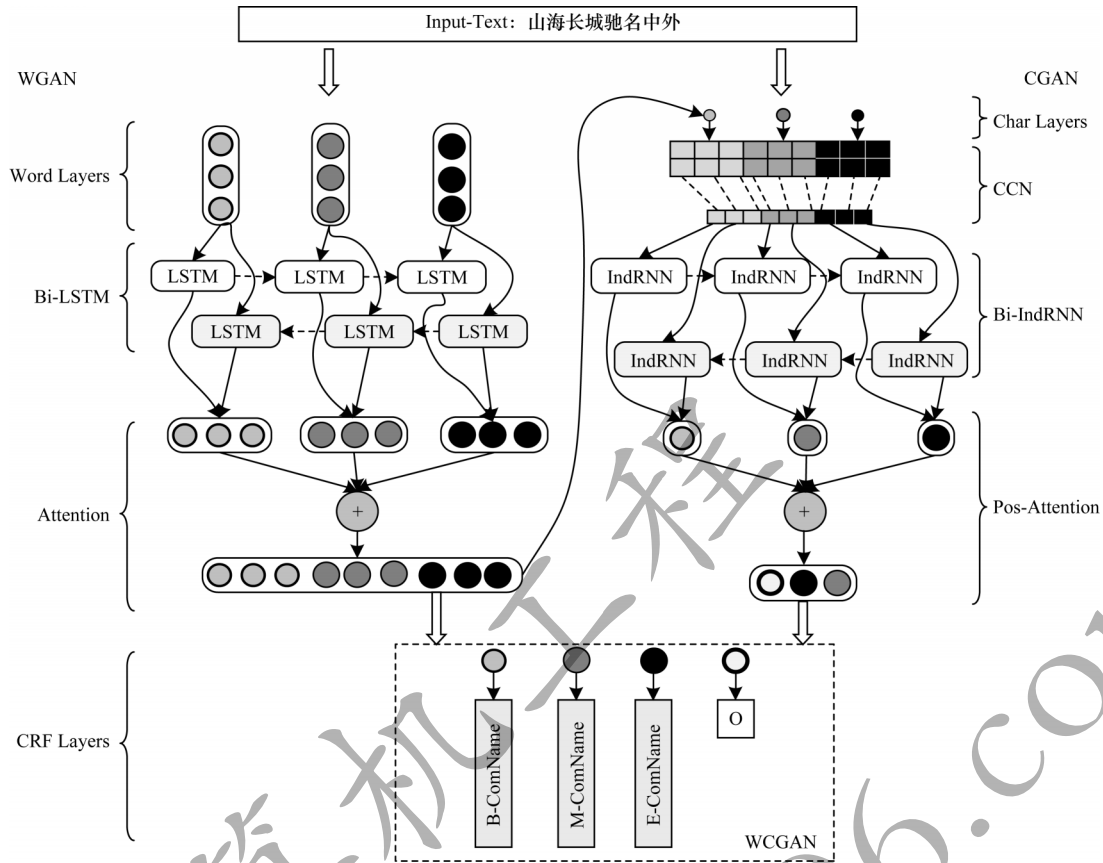


图2 WCGAN网络结构

Fig.2 Network structure of WCGAN

WCGAN方法的具体步骤如下:

1)通过WGAN得到命名实体的词表示形式 u^w ,利用CGAN得到命名实体中的字符表示形式 u^c 。

2)将词和字符信息进行融合,获得字符增强型的命名实体特征 u ,计算公式如式(11)所示:

$$u = u^w \cdot u^c \quad (11)$$

3)利用CRF获取命名实体的标签预测值^[27],计算公式如式(12)所示:

$$\text{Score}(\mathbf{x}, \mathbf{y}) = \sum_{j=1}^n \mathbf{u}_j \sum_{j=1}^{n+1} \mathbf{s}_{y_{j-1}, y}$$

$$P(\mathbf{x}, \mathbf{y}) = \exp(\text{Score}(\mathbf{x}, \mathbf{y})) \sum_{y'} \exp(\text{Score}(\mathbf{x}, \mathbf{y}')) \quad (12)$$

其中: $\mathbf{s}_{y_{j-1}, y}$ 表示CRF转移矩阵, $P(\mathbf{x}, \mathbf{y})$ 表示标签预测概率; \mathbf{u}_j 表示第 j 个命名实体的特征信息,且 $\mathbf{u} = \{u_1, u_2, \dots, u_n\}$ 。

可见,本文提出的WCGAN方法利用字符信息能有效增强命名实体的辨识度。

3 实验结果与分析

为验证WCGAN方法的有效性,在两组中文实验数据集上进行测试验证并与WGAN和CGAN模型进行比较,同时使用精确率、召回率和F值作为评

价指标确保结果的正确性和一致性。

3.1 实验数据

本文使用的数据集为:1)ResumeNER公开数据集^[4],该数据集包括国家、教育、人名、组织和职业5类命名实体,其中,命名实体数量为16 565,训练集数据量为9 380,验证集数据量为2 185,测试集数据量为5 000;2)TourismNER自定义数据集,该数据集是笔者从各旅游网站收集的旅游景点、人物、地址3类命名实体,其中,命名实体数量为13 840,训练集数据量为5 536,验证集数据量为4 152,测试集数据量为4 152。

本文对WCGAN方法中相关初始化参数进行设置。WGAN词嵌入维度为300维。CGAN采用1-of- m 编码策略,扩展单元数量为512,当其少于512个汉字字符时使用0进行填充,当其多于512时只需映射512个字符,该参数设置与CharCNN^[26]类似。优化函数为Adam,迭代次数为200,Dropout为0.25,学习率为0.02,当迭代20次后将学习率设置为0.001。

3.2 实验方法

为进一步验证WCGAN方法的优越性,将其与目前主流的中文命名实体识别方法进行比较:

1)CRF^[28]:该方法先对文本中的命名实体进行

标注,再使用CRF对各类命名实体的标签进行预测。

2)Bi-LSTM+CRF^[20]:该方法将获得的外部单词嵌入CRF扩展的Bi-LSTM中进行编码,提高了模型运行效率,并增加了命名实体识别的准确度。相关超参数的设置如下:学习率为0.005,优化函数为SGD,丢码率为0.5,隐藏层维度为100,外部词嵌入维度为100。

3)Bi-LSTM+CNN+CRF^[12]:该方法利用卷积神经网络提取命名实体的字符信息,并将不同字符之间的信息进行组合,输入Bi-LSTM中对每个单词的上下文进行建模,然后使用CRF模型进行解码,其在无需特征工程的情况下提高了命名实体的识别效果。相关超参数的设置如下:词嵌入维度为100,优化函数为SGD,学习率为0.015,批处理大小为10。

4)Char-Dense^[27]:该方法利用密集连接网络捕获命名实体的字符信息,能有效减少训练时间,并提高字符位置信息利用率及模型鲁棒性和有效性。该方法使用Adam作为优化函数。

5)CAN-NER^[4]:该方法使用卷积神经网络捕获文本中命名实体的字符信息,并利用自注意力机制和GUR模型来刻画相邻字符之间的关联性和句子的上下文语义信息。

6)LSTM+CRF^[6]:该方法利用不同通道来捕获中文社交网络中的命名实体,采用随机化嵌入词向量的方式将命名实体嵌入到网络层中,并将不同的通道参数进行共享,使得该方法能更好地捕获文本中命名实体的上下文信息。

3.3 实验结果

3.3.1 WGAN中不同识别模型对识别效果的影响

为验证WGAN方法中WGAN和CGAN模型对命名实体识别的影响,以ResumeNER和TourismNER数据集为基准数据集进行实验测试,结果如表1所示。

表1 不同识别模型的实验结果对比

Table 1 Comparison of experimental results of different recognition models %

模型与 方法	ResumeNER数据集			TourismNER数据集		
	精确率	召回率	F值	精确率	召回率	F值
WGAN	92.020	90.754	91.372	91.561	90.293	90.921
CGAN	92.880	91.625	92.253	92.145	90.859	91.480
WCGAN	94.127	92.863	93.491	93.491	92.238	92.860

由表1可以看出:1)WCGAN方法的F值在两种数据集上均表现最好,分别为93.491%和92.860%,相比WGAN模型分别提高了2.119和1.939个百分点,其主要原因为WCGAN方法使用字符信息增加了关键词的关注度,并进一步强调单词中字符与字符之间的关联性和位置信息,从而实现了信息互补,相比CGAN模型分别提高了1.238和1.380个百

分点,其主要原因为CGAN模型丢失了大量的单词信息;2)CGAN模型相比WGAN模型F值更高,其主要原因为WGAN模型在捕获单词信息时,忽略了命名实体单词中字符与字符之间的关联性以及命名实体之间的细微差别,从而导致WGAN模型的识别效果劣于CGAN模型。

3.3.2 WGAN中不同命名实体对识别效果的影响

为验证WGAN方法的准确性,对ResumeNER和TourismNER数据集中不同命名实体的识别效果进行实验测试。不同命名实体的混淆矩阵如图3所示。从图3中可以较为直观地看出不同命名实体在ResumeNER和TourismNER数据集中的分类效果,其中O在图3(a)和图3(b)中的F值均为最高,其主要原因为中文文本中非实体占比较大。

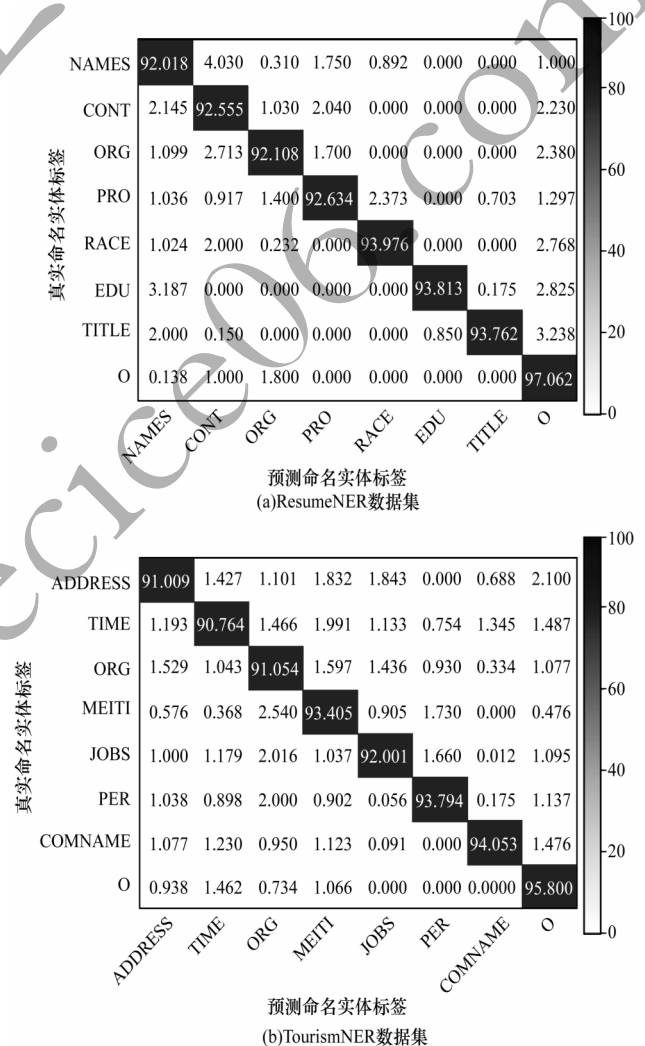


图3 不同命名实体的分类结果

Fig.3 Classification results of different named entities

3.3.3 WGAN在小样本数据集上的识别效果

为验证WGAN方法对小样本数据集中命名实体的识别效果,以ResumeNER基准数据集为基础,

分别使用1%、10%、20%、30%和40%的数据作为训练数据并迭代20次,剩余数据中测试和验证数据的数量均等,实验结果如图4所示。

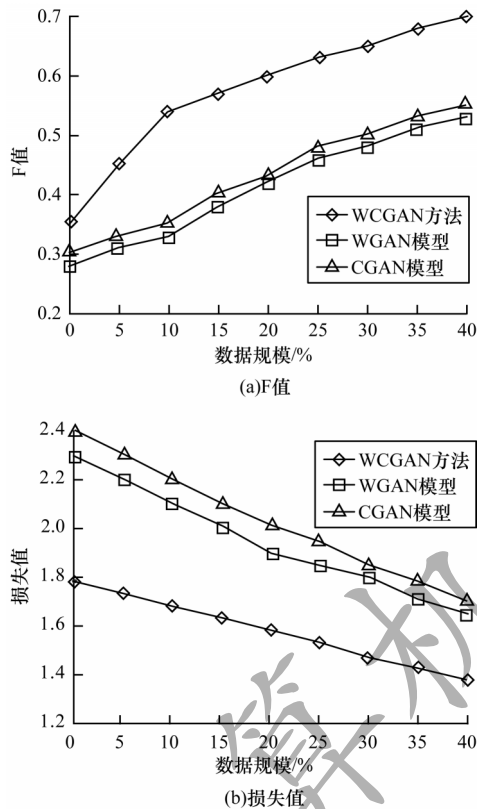


图4 在小样本数据集上的识别结果

Fig.4 Recognition results on small sample datasets

由图4可以看出:1)随着训练数据集的增加,F值逐渐升高,WCGAN方法的F值相比CGAN和WGAN模型更高,其主要原因为WCGAN方法对字符和单词建立关联性,且使用不同模型学习相应的特征信息形成信息互补;2)随着数据集规模的增加,3种方法的损失值也有所下降,且WCGAN方法的损失值始终处于最低状态,这表明在相同超参数的情况下,WCGAN方法能更好地学习命名实体的特征。

3.3.4 命名实体识别方法性能对比

为验证WCGAN方法的命名实体识别性能优势,将其与目前主流的中文命名实体识别方法进行对比,实验结果如表2所示。由表2可以看出:1)与CRF方法相比,LSTM+CRF、Bi-LSTM+CRF和Bi-LSTM+CNN+CRF方法具有更好的识别效果,其主要原因为这些方法使用组合模型更好地捕获了文本中命名实体的深层抽象信息,并在一定程度上解决了浅层机器学习方法使用人工参与设置特征带来的误差问题;2)与Char-Dense方法相比,CAN-NER方法在两个基准数据集上的F值分别提高了2.640和2.953个百分点,结果表明注意力机制能更好地捕获文本命名实体的序列信息,并对文中的关键信息进行突出,验证了注意力机制

在命名实体识别方面的有效性;3)WCGAN方法在两种基准数据集均取得了较好的识别效果,其主要原因为WCGAN方法能捕获命名实体的单词信息、字符信息以及字符与字符之间的位置信息,并且强调了字符与单词之间的关联性。

表2 7种命名实体识别方法的性能对比

Table 2 Performance comparison of seven NER methods

方法	ResumeNER 数据集			TourismNER 数据集		
	精确率	召回率	F 值	精确率	召回率	F 值
CRF	79.326	78.061	78.688	78.530	77.277	77.898
LSTM+CRF	83.963	82.709	83.331	81.608	80.249	80.923
Bi-LSTM+CRF	85.011	83.754	84.378	82.893	81.635	82.259
Bi-LSTM+CNN+CRF	87.384	86.125	86.750	84.712	83.464	83.733
Char-Dense	88.049	86.881	87.461	85.503	84.241	84.867
CAN-NER	90.739	89.472	90.101	88.398	87.131	87.760
WCGAN	94.127	92.863	93.491	93.491	92.238	92.860

3.3.5 可视化结果

为验证WCGAN方法能更好地捕获命名实体的字符信息以及增强单词的表征能力,本文以TourismNER基准数据集为实验数据,同时展示了多种命名实体识别方法的可视化结果,颜色越亮表示字符越重要。由图5可以看出,WCGAN方法能有效捕获命名实体的关键字符信息,在旅游命名实体“山海长城驰名中外”中“山”、“海”、“长”和“城”4个字符就被突出显示。由此可见,WCGAN方法在捕获细节特征方面效果更好。

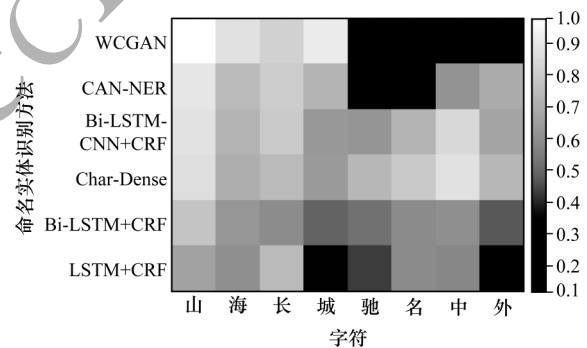


图5 字符信息的可视化结果

Fig.5 Visual results of character information

4 结束语

本文提出基于单词-字符引导注意力网络的中文旅游命名实体识别方法。通过单词引导注意力网络和字符引导注意力网络模型捕获命名实体的单词特征和字符信息,同时在单词和字符之间建立关联关系,并利用字符信息增强单词特征的表征能力,使得字符信息和单词特征间形成信息互补。在ResumeNER和

TourismNER 基准数据集上的实验结果表明,WCGAN方法能有效提高中文命名实体的识别效果。下一步将在无外部嵌入特征的情况下,研究基于图卷积神经网络的中文旅游文本命名实体识别方法。

参考文献

- [1] COLLOBERT R, WESTON J, BOTTOU L, et al. Natural language processing (almost) from scratch [EB/OL]. [2019-10-17]. <https://blog.csdn.net/tuqinag/article/details/43889795>.
- [2] LAMPLE G, BALLESTEROS M, SUBRAMANIAN S, et al. Neural architectures for named entity recognition [EB/OL]. [2019-10-17]. <https://arxiv.org/abs/1603.01360>.
- [3] YANG Y S, ZHANG M S, CHEN W L, et al. Adversarial learning for Chinese NER from crowd annotations [EB/OL]. [2019-10-17]. <https://arxiv.org/abs/1801.05147>.
- [4] ZHU Y Y, WANG G X, KARLSSON B. CAN-NER: convolutional attention network for Chinese named entity recognition [EB/OL]. [2019-10-17]. <https://arxiv.org/abs/1904.02141>.
- [5] LU Yanan, ZHANG Yue, JI Donghong. Multi-prototype Chinese character embedding [C]//Proceedings of the 10th International Conference on Language Resources and Evaluation. Philadelphia, USA: ACL Press, 2016: 855-859.
- [6] DONG Chuanhai, ZHANG Jiajun, ZONG Chengqing, et al. Character based LSTM-CRF with radical-level features for Chinese named entity recognition [C]//Proceedings of International Conference on Computer Processing of Oriental Languages. Berlin, Germany: Springer, 2016: 239-250.
- [7] NADEAU D, SEKINE S. A survey of named entity recognition and classification [J]. *Linguisticae Investigationes*, 2007, 30(1): 3-26.
- [8] SINGH T D, BANDYOPADHYAY S. Web based Manipuri corpus for multiword NER and Reduplicated MWEs identification using SVM [C]//Proceedings of International Conference on Computational Linguistics. Beijing: Chinese Information Society of China, 2013: 35-42.
- [9] KONKOL M, KONOPÍK M. CRF-based Czech named entity recognizer and consolidation of Czech NER research [C]//Proceedings of International Conference on Text, Speech and Dialogue. Berlin, Germany: Springer, 2013: 153-160.
- [10] SANTOS C D, GUIMARAES V, NITEROI R J, et al. Boosting named entity recognition with neural character embeddings [EB/OL]. [2019-10-17]. <https://arxiv.org/abs/1505.05008>.
- [11] ZHANG Yingcheng, YANG Yang, JIANG Rui, et al. Commercial intelligence entity recognition model based on BiLSTM-CRF [J]. *Computer Engineering*, 2019, 45(5): 308-314. (in Chinese)
张应成, 杨洋, 蒋瑞, 等. 基于 BiLSTM-CRF 的商情实体识别模型 [J]. *计算机工程*, 2019, 45(5): 308-314.
- [12] MA X Z, HOVY E. End-to-end sequence labeling via bi-directional LSTM-CNN-CRF [EB/OL]. [2019-10-17]. <https://arxiv.org/abs/1603.01354>.
- [13] CHEN W L, ZHANG Y J, ISAHARA H. Chinese named entity recognition with conditional random fields [C]//Proceedings of the 5th SIGHAN Workshop on Chinese Language Processing. Washington D. C., USA: IEEE Press, 2006: 118-121.
- [14] HE Hangfeng, SUN Xu. F-score driven max margin neural network for named entity recognition in Chinese social media [EB/OL]. [2019-10-17]. <https://arxiv.org/abs/1611.04234>.
- [15] HE Hangfeng, SUN Xu. A unified model for cross-domain and semi-supervised named entity recognition in Chinese social media [C]//Proceedings of AAAI' 17. Palo Alto, USA: AAAI Press, 2017: 3216-3222.
- [16] HUANG Zhiheng, XU Wei, YU Kai. Bidirectional LSTM-CRF models for sequence tagging [EB/OL]. [2019-10-17]. <https://arxiv.org/abs/1508.01991>.
- [17] LIU L Y, SHANG J B, XU F, et al. Empower sequence labeling with task-aware neural language model [EB/OL]. [2019-10-17]. <https://arxiv.org/abs/1709.04109>.
- [18] REI M. Semi-supervised multitask learning for sequence labeling [EB/OL]. [2019-10-17]. <https://arxiv.org/abs/1704.07156>.
- [19] GE L, MOH T S. Improving text classification with word embedding [C]//Proceedings of IEEE International Conference on Big Data. Washington D. C., USA: IEEE Press, 2018: 1-7.
- [20] ANH L T, ARKHIPOV M Y, BURTSEV M S. Application of a hybrid Bi-LSTM-CRF model to the task of Russian named entity recognition [C]//Proceedings of Conference on Artificial Intelligence and Natural Language. Berlin, Germany: Springer, 2017: 91-103.
- [21] JIANG X, HAVAELI M, CHARTRAND G, et al. On the importance of attention in meta-learning for few-shot text classification [EB/OL]. [2019-10-17]. <https://arxiv.org/abs/1806.00852>.
- [22] BROCKNER J, HJELLE L, PLANT R W. Self-focused attention, self-esteem, and the experience of state depression [J]. *Journal of Personality*, 1985, 53(3): 425-434.
- [23] ZHANG X, ZHAO J B, LECUN Y. Character-level convolutional networks for text classification [EB/OL]. [2019-10-17]. <https://arxiv.org/abs/1509.01626>.
- [24] LI S, LI W Q, COOK C, et al. Independently Recurrent Neural Network (IndRNN): building a longer and deeper RNN [EB/OL]. [2019-10-17]. <https://arxiv.org/abs/1803.04831>.
- [25] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [C]//Proceedings of Proceedings of the 31st International Conference on Neural Information Processing Systems. New York, USA: ACM Press, 2017: 6000-6010.
- [26] KIM Y. Convolutional neural networks for sentence classification [C]//Proceedings of EMNLP'14. Washington D. C., USA: IEEE Press, 2014: 1746-1751.
- [27] LEE C, KIM Y B, LEE D, et al. Character-level feature extraction with densely connected networks [EB/OL]. [2019-10-17]. <https://arxiv.org/abs/1806.09089>.
- [28] KRÁL P. Features for named entity recognition in Czech language [C]//Proceedings of International Conference on Knowledge Discovery and Information Retrieval. Berlin, Germany: Springer, 2011: 1-8.