

# 基于 BERT – BLSTM – CRF 的政务领域 命名实体识别方法

杨春明<sup>1 3</sup> 魏成志<sup>1</sup> 张 晖<sup>2</sup> 赵旭剑<sup>1</sup> 李 波<sup>1</sup>

(1. 西南科技大学计算机科学与技术学院 四川绵阳 621010; 2. 西南科技大学理学院 四川绵阳 621010;  
3. 四川省大数据与智能系统工程技术研究中心 四川绵阳 621010)

**摘要:** 政务领域的命名实体通常是一些政务事项名,这类实体与开放域实体比较,具有长度较长、实体并列、别称等特点,目前还未见公开可用的训练数据集。构建了具有 25 176 个句子的政务领域命名实体识别数据集,并提出一种基于 BERT – BLSTM – CRF 的神经网络识别模型,该模型在不依赖人工特征选择的情况下,使用 BERT 中文预训练模型,然后采用 BLSTM – CRF 识别实体。实验结果表明,该模型识别效果优于 CRF, BLSTM – CRF, CNN – BLSTM – CRF,  $F_1$  值达到 92.23%。

**关键词:** 政务事务 命名实体识别 BLSTM CRF BERT

**中图分类号:** TP391 **文献标志码:** A **文章编号:** 1671 – 8755(2020)03 – 0086 – 06

## A Method of Named Entity Recognition in Government Affairs Based on BERT – BLSTM – CRF

YANG Chunming<sup>1 3</sup>, WEI Chenzhi<sup>1</sup>, ZHANG Hui<sup>2</sup>, ZHAO Xujian<sup>1</sup>, LI Bo<sup>1</sup>

(1. School of Computer Science and Technology, Southwest University of Science and Technology, Mianyang 621010, Sichuan, China; 2. School of Science, Southwest University of Science and Technology, Mianyang 621010, Sichuan, China; 3. Sichuan Big Data and Intelligent System Engineering Technology Research Center, Mianyang 621010, Sichuan, China)

**Abstract:** The named entities in the government affairs are some service items, and they have the characteristics of long length, entity juxtaposition, abbreviations, nicknames, etc. At present, there is no publicly available training data set. In this paper, a government domain named entity recognition data set with 25 176 sentences was constructed, and a neural network method based on BERT – BLSTM – CRF was proposed. In this model, BERT Chinese pre-training model was used without relying on the selection of artificial features, and then BLSTM – CRF was used for named entity recognition. The experimental results show that the recognition accuracy is better than that of CRF, BLSTM – CRF, CNN – BLSTM – CRF, and the  $F_1$  value reaches 92.23%.

**Keywords:** Government affairs; Named entity recognition; BLSTM; CRF; BERT

在基于知识图谱的问答系统中,需要将问句中的命名实体进行识别后与知识库中的事实进行映射,通过计算相似度或推理来找到问句的正确答案。在政务服务事项的自动问答系统中,问句及知识库

中的命名实体通常是一些事项名(如建设施工企业资质、辐射安全许可、苗木检疫审批、生育服务证等),这类命名实体长度相对较长,常为多个意义词的并列,在用户的问句中有时表现为缩写或别称等,

收稿日期: 2020 – 02 – 26

基金项目: 教育部人文社科基金(17YJCZH260); 赛尔网络下一代创新项目(NGH20170901, NGH20180403)

第一作者简介: 杨春明(1980—),男,副教授,研究方向为自然语言处理、知识工程、大数据技术, E-mail: yangchunming@swust.edu.cn

如“生育服务证”也称“准生证”。

实体中的这些特征主要是由词语组合、同义词、多义词等导致的,这需要有丰富的训练语料来构建含有丰富句法和语义信息的词语表示,才能提高识别精度。由于该领域目前还没有已标注的专业语料库,致使已有的基于规则和词典的方法、基于统计机器学习的方法效果不佳。针对以上问题,本文以政务领域命名实体识别为目标,构建了政务领域命名实体识别数据集,并将通用领域大规模语料上学习到的文本特征迁移到政务领域,提出了结合中文预训练模型的命名实体识别方法,对比了已有的方法,在构建的语料上取得了较好的效果。

## 1 相关工作

命名实体识别方法主要有基于规则和词典的方法<sup>[1-2]</sup>、基于统计的方法<sup>[3]</sup>以及规则和统计相结合的方法<sup>[4-6]</sup>、基于深度学习的方法<sup>[7-12]</sup>。

基于规则的方法可在特定的语料上取得较高的识别效果,但是识别效果依赖大量人工规则的制定,受领域限制较大,当领域差别较大时,制定的规则通常无法移植。

基于统计的方法是一种序列化标注的方法,该方法中每个词可以有若干类候选标签,这些标签也对应着在实体中的位置信息,通过分类模型CRF, HMM, SVM等对词进行实体标注,最后通过标注确定实体。在特定领域中,通常采用基于规则和统计相结合的方法,如文献[6]采用CRF结合规则的方法,在医学病历的实体识别上准确率和召回率分别达到了91.03%和87.26%。该类方法的精度主要依赖于领域和已标注语料的规模。

基于深度学习的方法是为了解决命名实体识别的数据稀疏和语义问题,常见的有LSTM与CRF相结合的模型<sup>[7]</sup>、卷积神经网络(CNN)<sup>[9]</sup>等。这类方法主要通过深度学习获取词语的向量表示,由于在学习过程考虑了上下文信息,使得词向量包含了语义信息,解决了传统离散表示带来的数据稀疏问题。如Qin等<sup>[11]</sup>提出一种结合BLSTM网络和CRF的深度学习模型,用于提取电子病历中实体的方法,在i2b2/VA公开数据集中 $F_1$ 值最高达到了85.37%。Santos等<sup>[13]</sup>人在词性标注(POS)的基础上,利用CNN提出了一种CharWNN架构来对字符进行处

理,取得了较好的效果。Long等<sup>[14]</sup>结合CNN, BLSTM, CRF在SIGHAN 2006 Bakeoff-3语料库上提高了命名实体识别的性能。Chiu等<sup>[15]</sup>提出混合双向LSTM和CNN的神经网络模型,利用CNN训练得到字符级向量,同时结合词向量,再将二者进行组合作为BLSTM深层神经网络模型的输入。在CoNLL-2003和OntoNotes5.0语料上取得很好的效果, $F$ 值为91.62%和86.28%。基于深度学习的方法通过字符级特征与词向量相结合的方法,减少了特征工程的工作量,同时结合了上下文的语义信息,提高命名实体的性能,但该方法同样需要大量的标注语料。

政务领域命名实体主要指对政务领域服务办事事项,在政务领域自动问答、事务检索、主题发现等任务中通常需要进行命名实体识别,用于进行相似度的计算或推理。办事项中的部分实体具有词长度较长、意义并列(如建设施工企业资质、辐射安全许可、苗木检疫审批、生育服务证等)等情况,使用时通常表现为缩写或别称(如“生育服务证”也称“准生证”),使得该领域命名实体识别比通用的命名实体识别更为复杂。目前还未见专门针对政务领域中命名实体识别研究及相关的数据集。

政务实体的词语较长、意义并列是由通用领域中一般词语的组合导致的,同一事项的缩写和别称通常表现为多种形式,根据上下文,还会存在同义词和多义词的情况,这需要对词的表示具有丰富的句法和语义信息。针对这一问题,可使用预训练语言模型的方法,将从通用领域学习到的包含有上下文语义信息的词向量表示迁移到政务领域。因此,本文提出一种基于BERT(Bidirectional Encoder Representations from Transformers)预训练的神经网络模型,使用BERT中文预训练政务领域词向量作为BLSTM网络的输入,最后通过CRF层获取全局最优标记序列来确定实体。本文构建了政务领域4913个办事项、25176个句子的命名实体识别语料,在该语料上对比了传统的基于规则和统计相结合的方法、基于BLSTM-CRF的方法、基于CNN-BLSTM-CRF的方法以及基于BERT的方法与本文方法在未使用任何人工特征的情况下模型在训练语料上的效果。

## 2 BERT - BLSTM - CRF 政务领域命名实体识别模型

模型首先通过 BERT 预训练语言模型得到含有语义信息的字向量表示,再将其作为 BLSTM 层的输入,进一步进行语义编码,最后通过 CRF 层解码出一个全局最优的标记序列。模型整体结构如图 1 所示。

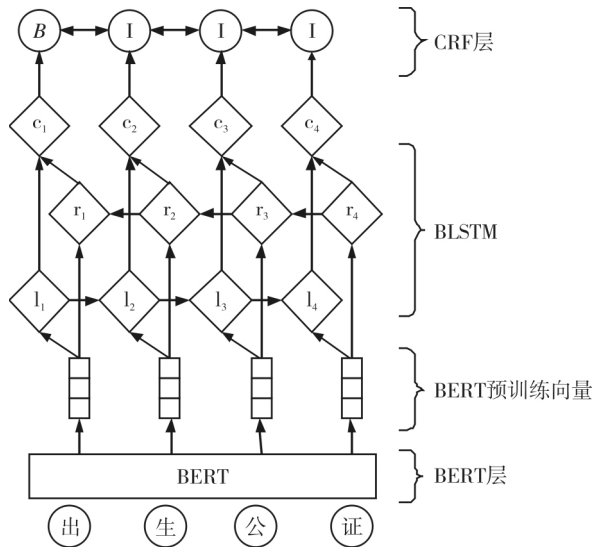


图1 基于 BERT - BLSTM - CRF 政务领域命名实体识别模型

Fig. 1 Named entity recognition model in government affairs based on BERT - BLSTM - CRF

BERT<sup>[16]</sup>是一个预训练语言模型,它在大规模语料上采用 12 层或者 24 层的双向 Transformer 编码结构,基于注意力机制来对文本建模。由于采用双向编码,使得训练出的结果反映了词与句子中其他词语的语义关系,能够表征出词语的多义性,可获得更好的词分布表示。大规模语料的预训练语言模型可以为某些数据规模较少的自然语言处理任务提供模型参数,提升训练效果。使用时只需对预训练的网络参数进行微调以适合当前的任务,即可将大规模语料中学习的特征迁移到具体任务中。

政务领域中的实体通常由通用领域中一般词语组合而来,其别称和缩写在日常文本中经常使用,因此本文使用 BERT 将通用领域学习到的文本特征迁移到政务领域,将政务领域文本数据结合通用领域数据训练得到对应的词向量,将其作为 BLSTM 层的输入。

由于政务领域中的命名实体构成复杂,其实体的子序列也可能是命名实体,词语前后的关联性强,

在进行网络训练时需要充分考虑文本序列的上下文信息。循环神经网络(RNN)采用定向循环的方式来处理序列化数据,通过输入层、隐藏层和输出层的顺序进行信息传递,使得序列中的节点能够“记忆”前文信息。但 RNN 在序列过长时存在长期依赖性问题,即会出现梯度消失,不能学习到长期依赖的特征。长短时记忆网络(Long-Short-Term Memory, LSTM)<sup>[17]</sup>改进了 RNN,在隐层通过单元状态和门控来解决长期依赖问题,其隐藏层的结构如图 2 所示。

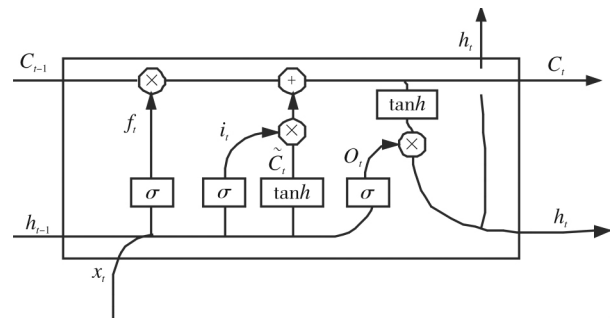


图2 LSTM 隐含层的结构图

Fig. 2 Structure diagram of LSTM hidden layer

LSTM 的网络结构可以表示为:

$$f_t = \sigma(W_f \cdot [h_{t-1} \ x_t] + b_f) \quad (1)$$

$$i_t = \sigma(W_i \cdot [h_{t-1} \ x_t] + b_i) \quad (2)$$

$$\tilde{C}_t = \tanh(W_c \cdot [h_{t-1} \ x_t] + b_c) \quad (3)$$

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t \quad (4)$$

$$o_t = \sigma(W_o \cdot [h_{t-1} \ x_t] + b_o) \quad (5)$$

$$h_t = o_t \cdot \tanh(C_t) \quad (6)$$

其中  $i, f, o$  分别表示输入门、遗忘门和输出门,  $W_i, W_f, W_o$  是对应的权重矩阵,  $[h_{t-1} \ x_t]$  表示连接这两个向量  $b_i, b_f, b_o$  是遗忘门的偏置项,  $\sigma$  是 sigmoid 函数。

从网络结构可知,单向的 LSTM 重点获取了序列中的历史信息,针对命名实体而言,进行序列标注时一个字的标签与该字所在句子的上下文都有关系。BLSTM(Bi-directional Long Short-Term Memory)是一个双向循环神经网络,由前向 LSTM 和后向 LSTM 组成,从两个方向来处理输入序列,每个时刻的前向和后向 LSTM 隐含层都连接到同一个输出层,最终将前向 LSTM 输出的字向量  $\vec{h}_t$  及后向的字向量  $\overleftarrow{h}_t$  拼接成  $h_t = [\vec{h}_t \ \overleftarrow{h}_t]$ 。

针对政务实体结构前后关联较强的问题,本文采用 BLSTM 来获取实体在句子上下文中的信息,将其结果输出作为实体标签分类的输入。

BLSTM 只获取了含有上下文信息的字向量,并没有标签之间的关系,因此不能简单根据字向量来决策标签。本文采用 CRF 考虑标签之间的相邻关系来获取全局最优的标签序列。

对于一个句子序列  $X = [x_1, x_2, \dots, x_n]$ , 在  $t$  时刻输入  $x_t$ , BLSTM 层的输出为预测  $x_t$  标签的概率分布  $c_t$ 。因此,对于输入序列  $X$ , BLSTM 层的输出为  $p_{n \times k} = [c_1, c_2, \dots, c_n]$ , 其中  $p$  为  $n \times k$  大小的概率矩阵,  $n$  是单词个数,  $k$  是标签的个数,  $p_{i \times j}$  表示第  $i$  个单词的第  $j$  个标签的概率。对于模型得到的标签序列  $y = [y_1, y_2, \dots, y_n]$  最优标签序列的计算公式如式(7):

$$s(X, y) = \sum_{i=1}^n A_{y_i y_{i+1}} + \sum_{i=1}^n P_{i y_i} \quad (7)$$

其中  $A$  为转移矩阵,  $A_{ij}$  表示由标签  $i$  到标签  $j$  的转移概率,  $P_{i y_i}$  表示第  $i$  个位置 softmax 输出为  $y_i$  的概率,  $A$  是一个大小为  $(n+2) \times (n+2)$  的矩阵, 因为额外增加了起始位置和结束位置。对于政务领域, 标签出现的顺序即是前后标签存在的依赖关系, 对每个训练样本  $X$  求出所有可能的标签序列  $y$  的得分  $s(X, y)$ , 对其归一化:

$$p(y | X) = \frac{e^{s(X, y)}}{\sum_{\tilde{y} \in Y_X} e^{s(X, \tilde{y})}} \quad (8)$$

其中,  $\tilde{y}$  表示真实的标记值,  $Y_X$  是对于输入序列  $X$  所有可能出现的预测序列观测值的集合。该标记序列的似然函数为:

$$\log(p(y | X)) = s(X, y) - \log\left(\sum_{\tilde{y} \in Y_X} e^{s(X, \tilde{y})}\right) \quad (9)$$

预测过程中, 最大化输出整体概率的一组序列用(10)表示:

$$y^* = \operatorname{argmax}_{\tilde{y} \in Y_X} s(X, \tilde{y}) \quad (10)$$

### 3 实验及结果分析

#### 3.1 实验环境

本文所有实验的环境如表1所示。

表1 实验环境

Table 1 Experiment conditions

环境	配置
操作系统	Ubuntu 16.04
CPU	I7-8700K
GPU	GTX1080Ti(11GB)
内存	64 GB
Python	3.5
Tensorflow	1.12
CRF	CRF++ 0.58

#### 3.2 评价标准

本文采用的评价标准包括准确率  $P$  (Precision)、召回率  $R$  (Recall) 和  $F_1$  ( $F_1$ -measure) 3 个指标, 计算公式如式(11)、式(12)、式(13):

$$P = \frac{\text{正确标注的实体个数}}{\text{标注为实体的总个数}} \times 100\% \quad (11)$$

$$R = \frac{\text{正确标注的实体个数}}{\text{语料中实体的总个数}} \times 100\% \quad (12)$$

$$F_1 = \frac{2 \times (P \times R)}{P + R} \times 100\% \quad (13)$$

#### 3.3 数据集构建与标注策略

由于针对政务领域的命名实体识别没有公开的数据集, 为了验证模型的有效性, 从某省政务服务网上抓取了政务领域办事事项、常见问题共计 1 687 条, 然后进行数据清洗、事项拆分、问句组合、BIO 标注, 最后构建了具有 25 176 个句子的政务命名实体识别的数据集。

数据清洗主要是去除特殊符号, 例如: “《” “(” “)” “》”。由于获取的政务事项数据中存在一个事项包含多个子项情况, 如“学历、学位公证”这一事项, 包含了“学历公证”“学位公证”两个办事事项, 因此, 人工拆分了 1 687 个具体办事事项, 最终确定 4 913 个命名实体。然后, 再对清洗后的政务事项中的常见问题进行分析, 从中提取了约 100 种问题提问方式, 最后与具体办事事项进行组合, 采用 BIO 标注方式标注语料, 构建了针对政务命名实体识别的数据集。其中 B 为命名实体的开始部分, I 为命名实体的中间部分, O 为非命名实体成分。构建过程及标注后的语料格式如图3所示。

#### 3.4 实验及结果分析

为了验证本文模型的有效性, 对比了基于 CRF、BLSTM-CRF、CNN-BLSTM-CRF、BERT 的命名实体识别方法。基于 CRF 的命名实体识别采用 70% 数据用作训练集, 30% 用作测试集的划分方式, 数据规模如表2所示。基于 BLSTM-CRF、CNN-BLSTM-CRF、BERT 以及 BERT-BLSTM-CRF 的命名实体识别实验, 采用 80% 用作训练集, 10% 用作验证集, 10% 用作测试集的划分方式。数据规模如表3所示。

##### 3.4.1 基于 CRF 命名实体识别实验结果

本文选取了以字、词为单位, 结合词性、边界特征来构建训练集和测试集, 使用 CRF++ 0.58 进行

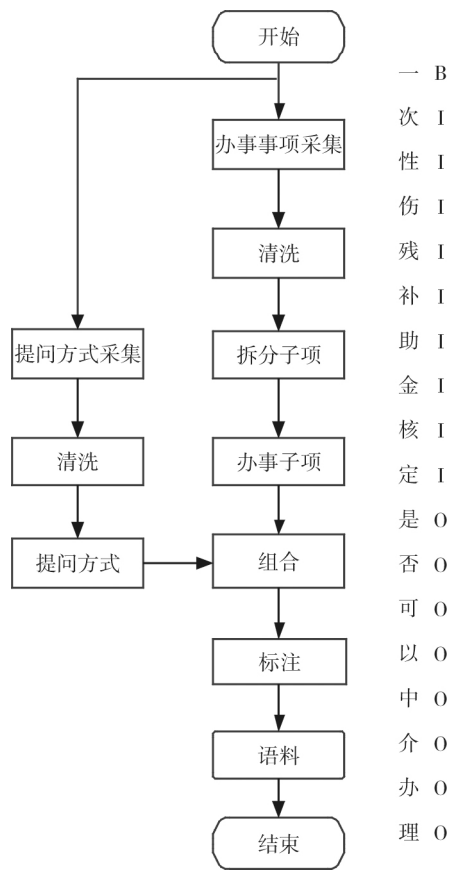


图 3 数据集构建流程图及标注格式

Fig. 3 Dataset construction flow chart and annotation format

表 2 基于 CRF 的命名实体识别实验数据规模

Table 2 CRF – based named entity recognition experiment datasize

项目	长度	句子数	句子平均长度
训练集	324 512	17 579	18.5
测试集	138 609	7 595	18.3

表 3 基于深度学习的命名实体识别实验数据规模

Table 3 Data size of Deep Learning

项目	长度	句子数	句子平均长度
训练集	368 806	20 043	18.4
验证集	47 017	2 577	18.2
测试集	47 322	2 556	18.5

训练和测试,为防止过度拟合,代价参数为 1.5,其他使用默认参数。字、词特征主要选择当前字、词的前、后第一、第二个字、词,同时选择了对应位置词的词性。实验结果如表 4 所示。

表 4 基于 CRF 的政务领域命名实体识别实验结果

Table 4 Experimental results of named entity recognition based on CRF in government affairs

项 目	$P/\%$	$R/\%$	$F_1/\%$
CRF + 字级	89.15	78.17	83.30
CRF + 字级 + 词性	88.79	79.48	83.88
CRF + 字级 + 边界	88.88	78.76	83.51
CRF + 字级 + 词性 + 边界	89.10	80.00	84.31
CRF + 词级 + 词性	92.25	80.45	85.95

从以上实验结果来看,以词级为单位并结合词性特征来进行政务领域命名实体识别的效果最好,准确率(  $P$  )和召回率(  $R$  )都有提升,  $F$  值为 85.95%。这反映了政务实体通常是由词语组合而成,且具有组合的基础词性一致的特点,其准确率有了近 3 个百分点的提升。而字级特征加词性和边界特征的组合的效果差异不大,这表明词对政务实体更敏感,但所有这些特征的选择依赖人工选取以及领域知识。

### 3.4.2 基于深度学习的命名实体识别实验结果

为了对比深度学习方法在政务实体识别中的优势,本文在 CRF 的基础上增加了不同的深度学习方法,并进行了对比。具体比较了 BLSTM – CRF, CNN – BLSTM – CRF, BERT 以及 BERT – BLSTM – CRF,其中预训练语言模型中采用的是 BERT – Base。根据实验硬件条件的限制以及多次训练后的对比,在 BERT 和 BLSTM 中的参数设置如下:最大序列长度:128;一次训练使用的样本数:32;学习率:  $2 \times 10^{-5}$ ;训练周期:3;Dropout:0.5;隐层维度:200。

CNN 的字符嵌入大小为 30,过滤器大小为 30,过滤器长度为 3,dropout 为 0.5,学习率为 0.01,训练周期为 20。实验结果如表 5 所示。

从表 5 可以看到,增加了 BLSTM 的效果比基于 CRF 的命名实体识别在选择合适的特征模板( CRF + 词 + 词性)下精度有所降低,这是因为政务领域中命名实体中词组合比较灵活,与神经网络模型自动学习实体特征相比,采用特征模板从窗口提取的特征往往更有效。在增加 CNN 抽取单词的局部特征信息,将字符级表示向量与词嵌入向量连接作为 BLSTM 网络的输入后,识别精度有所提升,  $F_1$  值达到 91.54%,这表明词语的局部信息嵌入向量后,其语义信息更加丰富。

表5 基于深度学习的命名实体识别实验结果

Table 5 Experimental results of named entity recognition based on deep learning

模型	$P/\%$	$R/\%$	$F_1/\%$
BLSTM-CRF	87.58	83.67	85.58
CNN-BLSTM-CRF	90.38	92.73	91.54
BERT	86.85	92.00	89.35
BERT-BLSTM-CRF	92.72	91.73	92.23

在单独使用BERT模型时,效果并不理想,这是因为只迁移了通用领域的词语信息,而加入BLSTM后,获取了词语前后的历史信息,融入了词语的上下文信息,可以明确多义词在上下文中的语义, $F_1$ 达到92.23%。比基于CRF提高了3.4%,比BLSTM-CRF提高了3.77%。可见,通过大规模语料的预训练,可以有效提高政务领域命名实体的识别精度。

#### 4 结论

政务命名实体识别可以有效提高政务服务自动问答系统的效果,其具有一定规模的训练数据集是提高识别率的前提。本文针对省级政务服务事项,构建了具有25 176条句子的政务领域命名实体识别的数据集。通过使用预训练语言模型,提出了结合BERT中文预训练模型的BERT-BLSTM-CRF的深度学习模型,对比了基于CRF结合人工定义特征的方法、基于BLSTM-CRF模型的方法、基于CNN-BLSTM-CRF模型的方法等,在构建的政务领域数据集上,取得了较好的效果, $F$ 值达到了92.23%。由于该领域的实体具有动态性,在问答任务中还需要考虑实体间的关系,因此,在今后的工作中会增加不同行政区划的部门、机构信息来丰富训练集,同时深入研究政务领域实体链接方面的工作。

#### 参考文献

- [1] 闫丹辉,毕玉德. 基于规则的越南语命名实体识别研究[J]. 中文信息学报, 2014, 28(5): 198-205.
- [2] 皇甫晶,王凌云. 基于规则的纪传体古代汉语文献姓名识别[J]. 图书情报工作, 2013, 57(3): 120-124.
- [3] 乐娟,赵玺. 基于HMM的京剧机构命名实体识别算法[J]. 计算机工程, 2013, 39(6): 266-271.
- [4] 何炎祥,罗楚威,胡彬尧. 基于CRF和规则相结合的地理命名实体识别方法[J]. 计算机应用与软件, 2015, 32(1): 179-185.
- [5] 张金龙,王石,钱存发. 基于CRF和规则的中文医疗机构名称识别[J]. 计算机应用与软件, 2014(3): 159-162.
- [6] 栗伟,赵大哲,李博,等. CRF与规则相结合的医学病历实体识别[J]. 计算机应用研究, 2015, 32(4): 1082-1086.
- [7] 李明浩,刘忠,姚远哲. 基于LSTM-CRF的中医医案症状术语识别[J]. 计算机应用, 2018, 38(S2): 47-51.
- [8] 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): 261-268.
- [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): 116-122.
- [10] ZENG D, SUN C, LIN L, et al. LSTM-CRF for drug-named entity recognition[J]. Entropy, 2017, 19(6): 283.
- [11] QIN Y, ZENG Y. Research of clinical named entity recognition based on Bi-LSTM-CRF[J]. Journal of Shanghai Jiaotong University (Science), 2018, 23(3): 392-397.
- [12] MA X, HOVY E H. End-to-end sequence labeling via Bi-directional LSTM-CNNs-CRF[C]// Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, Berlin, 2016: 1064-1074.
- [13] SANTOS, C D, ZADROZNY, B. Learning character-level representations for part-of-speech tagging[C]// Proceedings of the 31st International Conference on International Conference on Machine Learning. JMLR.org, 2014, 32(II): 1818-1826.
- [14] LONG S, YUAN R, YI L, et al. A method of Chinese named entity recognition based on CNN-BILSTM-CRF model[C]// International Conference of Pioneering Computer Scientists, Engineers and Educators. Springer, Singapore, 2018: 161-175.
- [15] CHIU J P, NICHOLS E. Named entity recognition with bidirectional LSTM-CNNs[J]. Transactions of the Association for Computational Linguistics, 2016, 4(1): 357-370.
- [16] DEVLIN J, CHANG M W, LEE K, et al. Bert: Pre-training of deep bidirectional transformers for language understanding[C]// Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2019, Minneapolis, MN, USA, June 2-7, 2019, Volume 1 (Long and Short Papers). Association for Computational Linguistics, 2019: 4171-4186.
- [17] ZAREMBA W, SUTSKEVER I, VINYALS O. Recurrent neural network regularization[J]. CoRR, 2014, abs/1409.2329.