

doi:10.3969/j.issn.1673-9833.2020.03.003

# 融合深度神经网络与统计学习的 印地语词性标注方法研究

王连喜<sup>1,2</sup>, 钟 准<sup>2</sup>, 丁曾强<sup>2</sup>, 邓致妍<sup>1,3</sup>, 李 霞<sup>1,2</sup>

(1. 广东外语外贸大学 广州市非通用语种智能处理重点实验室, 广东 广州 510006;

2. 广东外语外贸大学 信息科学与技术学院, 广东 广州 510006;

3. 广东外语外贸大学 日语语言文化学院, 广东 广州 510006)

**摘要:** 针对统计模型受限于标注语料规模且不能捕获标注序列的上下文信息问题, 提出一种融合深度学习和统计学习的印地语词性标注模型。该模型具有3层逻辑结构, 首先在词表示层采用神经网络框架训练出印地语单词的形态特征, 并利用 word2vec 方法对语料训练生成具有语义信息的低维度稠密实数词向量, 然后在序列表示层将形态特征和词向量作为神经网络模型的输入并进行训练, 得到输入序列的信息特征, 最后在 CRF 推理层利用神经网络模型的输出状态和当前的转移概率矩阵作为 CRF 模型的参数, 最终得到最优的标签序列。对提出的方法与其他方法进行了对比实验, 结果表明融合深度学习和统计模型的方法较其他几种统计模型的性能有显著的提升。

**关键词:** 印地语; 词性标注; 深度神经网络; 统计学习; 条件随机场

**中图分类号:** TP391.1

**文献标志码:** A

**文章编号:** 1673-9833(2020)03-0017-06

**引文格式:** 王连喜, 钟 准, 丁曾强, 等. 融合深度神经网络与统计学习的印地语词性标注方法研究 [J]. 湖南工业大学学报, 2020, 34(3): 17-22.

## Research on Hindi Part-of-Speech Tagging Based on Deep Neural Network and Statistical Learning

WANG Lianxi<sup>1, 2</sup>, ZHONG Zhun<sup>2</sup>, DING Zengqiang<sup>2</sup>, DENG Zhiyan<sup>1, 3</sup>, LI Xia<sup>1, 2</sup>

(1. Guangzhou Key Laboratory of Multilingual Intelligent Processing, Guangdong University of Foreign Studies,

Guangzhou 510006, China; 2. School of Information Science and Technology, Guangdong University of

Foreign Studies, Guangzhou 510006, China; 3. Faculty of Japanese Language and Culture,

Guangdong University of Foreign Studies, Guangzhou 510006, China)

**Abstract:** In view of the limitation exhibited by the statistical model in its size of the annotated corpus with its inability to capture the context information of the tagging sequence, a Hindi part-of-speech tagging model with deep learning and statistical learning combined has thus been proposed. The model has a three-layer logical structure.

**收稿日期:** 2020-04-17

**基金项目:** 国家社会科学基金资助项目 (17CTQ045), 广州市科技计划基金资助项目 (202006020302), 广东省软科学基金资助项目 (2019A101002108), 广东省普通高校“人工智能”重点领域专项基金资助项目 (2019KZDZX1016), 广东外语外贸大学特色创新基金资助项目 (师生共研类) (19SS01)

**作者简介:** 王连喜 (1985-), 男, 湖南常宁人, 广东外语外贸大学副研究员, 博士, 硕士生导师, 主要从事自然语言处理与数据挖掘方面的教学与研究, E-mail: wanglianxi@gdufs.edu.cn

**通信作者:** 邓致妍 (1994-), 女, 广东清远人, 广东外语外贸大学教师, 主要从事印度舆情和印度语言文化方面的教学与研究, E-mail: 201610140@oamail.gdufs.edu.cn

First, the deep neural network framework is used to extract the morphological features of Hindi words in the word representation layer, together with using the word2vec method to generate with the low-dimensional dense real number word vectors with semantic information to be generated by adopting the word2vec method. Next, in the sequence presentation layer, the information features of the input sequence can be obtained with morphological features and word vectors used as the input of the deep neural network model tested by training. Finally, in the CRF inference layer, with the output state of the deep neural network model and the current transition probability matrix as the parameters of the CRF model, the optimal label sequence can be finally obtained. The results show that, compared with other statistical models, the performance of the proposed method has been significantly improved.

**Keywords:** Hindi; part-of-speech tagging; deep neural network; statistical learning; CRF

## 0 引言

词性标注(part-of-speech tagging), 又称词类标注, 是指以上下文关系为前提, 给句子中的每个词标注出合理词性的过程<sup>[1]</sup>。词性标注是自然语言处理任务的重要预处理工作之一, 也是信息抽取、词法分析、语法分析、语义分析等研究的基础。词性标注的准确率将直接影响文本检索、文本分类、机器翻译等研究的效果。词性标注本身是一个难度比较大的问题, 因为很多词语通常有多种词性, 但是在考虑词语所处上下文语境的情况下, 这些词语的词性是唯一的, 可以明确附以名词、动词、形容词等具体词性标签。

印地语是印欧语系-印度语族下的一种语言, 是南亚次大陆上使用人口最多的语言, 也是印度的官方语言之一。以使用人口数量和使用国家数量来计算, 印地语是世界排名第8的语言, 但由于印度曾被殖民的缘故, 印地语的使用地位比英语要低得多, 并未成为众多使用人口的第一语言。因此, 印地语与其他低资源语言一样, 存在标注语料稀缺、标注难度及代价较高、规范性较差等问题, 导致大规模的标注资源不易获取。

许多研究表明<sup>[1-2]</sup>, 目前对于通用语种的词性标注方法已较为成熟, 但由于资源稀缺问题和语言的特殊性, 无法直接应用于印地语词性标注上。印地语词性标注目前面临如下挑战: 一是印地语中有大量借词(其来源包括梵语、英语、乌尔都语、阿拉伯语、波斯语、土耳其语、葡萄牙语及普什图语), 导致语料中有大量外来词; 二是印地语承袭了古代梵文多变性特点, 词语的格和词性比较复杂, 变化灵活。

## 1 相关研究概述

针对印地语词性标注问题, 近年来科研工作者们已提出了多种相关的词性标注方法, 如基于规则的方

法、基于HMM(hidden markov model)的词性标注方法等<sup>[3-7]</sup>。如H. Agarwal等<sup>[8]</sup>利用条件随机场对印地语词性进行标注训练, 并对语料库中每个单词的词根和可能的词性进行了标记。A. Dalal等<sup>[9]</sup>以最大熵马尔可夫模型同时使用多个特征来捕获与序列标记任务相关的词语词性, 并提出了一个基于语言标签和组块分析统计的印地语词性标注方法。隐马尔可夫模型、最大熵模型和条件随机场等统计模型的依赖关系学习能力强, 但是需要定义大量特征模板, 否则容易陷入局部最优解。N. Garg等<sup>[10]</sup>提出了基于规则的印地语词性标注方法, 并对包含了30个印地语标准词性标签的语料进行了实验, 结果显示, 其精度达到了87.55%。S. Asopa等<sup>[11]</sup>利用规则与组块相结合的方法进行词性标注, 在有限语料的情况下提高了标注效果。D. Modi等<sup>[12]</sup>结合29个词性标签和多个词法特征规则, 提出了基于规则的印地语词性标注方法, 该方法也取得了一定的效果。基于规则的方法针对特定领域的准确率往往较高, 但是规则制定较难且不能穷尽, 所以其泛化效果通常会受到限制。

自神经网络模型流行以来, 一些学者也在尝试利用深度学习来学习长距离信息之间的上下文信息, 并以此解决词性标注、命名实体识别等序列标注问题。如J. Yousif等<sup>[13]</sup>提出了基于多层感知器神经网络的印地语词性标注, 并利用反向传播学习算法对标注结果进行纠错。R. Narayan等<sup>[14]</sup>采用神经网络对印地语词性标注进行了研究, 实验准确率高达91.30%, 效果明显比其他印地语词性标注方法好。此外, N. Mishra等<sup>[15]</sup>提出了一种混合的印地语词性标注方法, 该方法首先借助WordNet字典标记印地语单词, 然后采用基于规则的方法为未标记的单词分配标签标记, 最后利用HMM模型来消除歧义。D. Modi等<sup>[16]</sup>融合基于统计和基于规则的方法, 在一个小规模标注语料上取得了88.15%的平均准确率。

总体上看, 印地语自然语言处理技术还不够成熟, 国内外开展印地语词性标注方法研究的工作相对较少, 并且目前基于统计的词性标注方法对于缺乏大规模人工标注语料的印地语词性标注而言, 还不能取得明显的效果。基于此, 本课题组考虑将统计学习模型和深度神经网络模型进行结合, 通过深度学习模型对待标注序列的上下文信息进行捕获, 然后采用条件随机场 (conditional random field, CRF) 对整个待标注序列的局部特征进行信息线性加权, 从而更好地识别序列信息前后的依赖关系。为能更好地捕获上下文信息, 并使得输出序列符合一些基本依赖关系的约束, 本文研究一种融合深度神经网络和统计学习的印地语词性标注模型。

## 2 印地语词性标注

词性标注是自然语言处理领域的一种典型序列标注任务, 其本质就是对线性序列中每个元素根据上下文内容进行分类的问题。词性标注的过程, 就是给定一个一维线性输入序列  $X = \{x_1, x_2, x_3, \dots, x_n\}$ , 通过利用规则、统计模型或深度学习模型, 对该序列中的每个元素  $x_i$  打上给定标签集中的某个标签  $y_i$ , 从而得到输入序列的相应标签序列  $Y = \{y_1, y_2, y_3, \dots, y_n\}$ 。例如, 给出一个印地语句子: स्वास्थय आप अपने स्वास्थय को लेकर अक्सर परेशान रहेंगे |, 通过词性标注方法进行标注后, 可以得到如下的相应标注结果: स्वास्थय/NN आप/PRP अपने/PRF स्वास्थय/NN को/PSP लेकर/VAUX अक्सर/RB परेशान/JJ रहेंगे/VM | PTT。

在处理序列标注任务时, 目前学术界较为流行的做法, 是将统计模型与深度学习模型相结合<sup>[17-18]</sup>, 这样, 一方面可以学习待标注序列中的依赖约束关

系, 另一方面也可以考虑到长距离的上下文关系, 从而很好地结合两种模型的优点。

### 2.1 LSTM 和 BiLSTM

虽然循环神经网络 (recurrent neural network, RNN) 常被用于解决序列标注问题, 但是该模型存在“长距离依赖”问题, 且容易在网络训练过程中出现梯度消失和梯度爆炸问题。为了解决长序列信息依赖问题, S. Hochreiter 等<sup>[19]</sup> 提出了长短期记忆网络 (long short-term memory, LSTM)。LSTM 通过输入门、遗忘门和输出门 3 种类型的门结构来控制不同时刻的状态和输出, 通过门控控制单元的状态信息控制之前信息和当前信息的记忆和遗忘程度, 所以该模型可以选择性地保存序列信息。但是单向的 LSTM 模型 (如图 1) 只能获取序列的上文信息, 无法获取序列的下文信息, 且往往会更偏向于最近的输入表示信息。

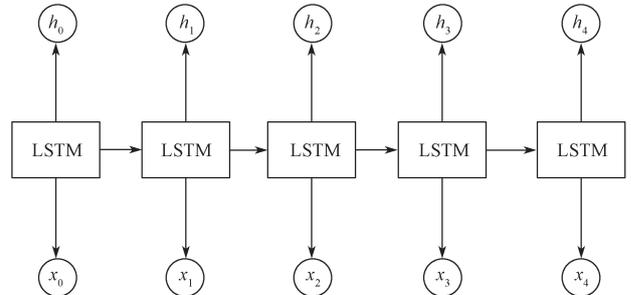


图 1 LSTM 模型的结构

Fig. 1 Structure of LSTM model

为解决 LSTM 无法同时从正向和反向获取序列信息的问题, C. Dyer 等<sup>[20]</sup> 提出了双向长短期记忆模型 BiLSTM, 其网络结构如图 2 所示。BiLSTM 模型可以从正向和反向两个方向同时对序列进行建模, 不仅可以保存上文信息, 还可以考虑到下文信息。

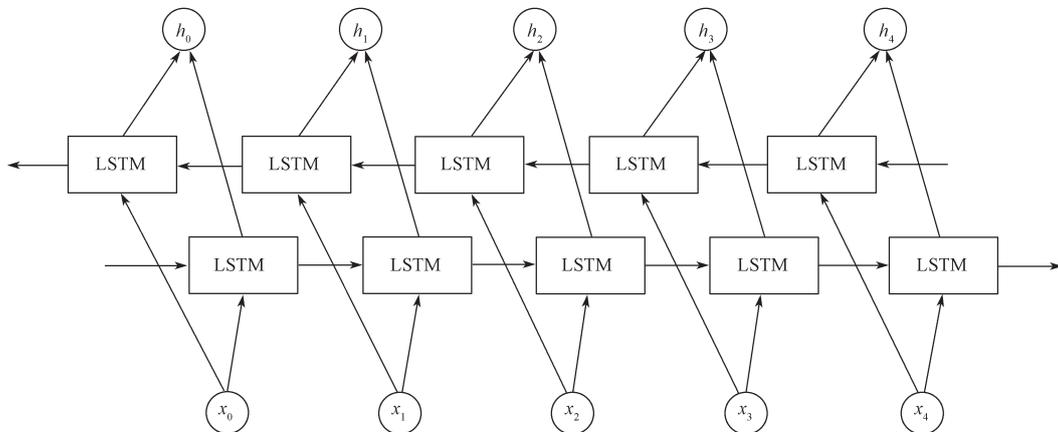


图 2 BiLSTM 模型的结构

Fig. 2 Structure of BiLSTM model

### 2.2 CRF

条件随机场<sup>[21]</sup> (CRF) 是一种经典的基于无向

图的条件概率模型, 也常被用于序列标注问题。CRF 模型既具有判别式模型的优点, 又具有生成式模型考

虑到的上下文标注间的转移概率,以序列化形式进行全局参数优化和解码的特点,其结合了隐马尔可夫模型的优点和最大熵模型的优点,解决了其他判别式模型难以避免的标注偏置问题,是传统统计序列标注的强力模型。

假设随机变量序列  $X=\{x_1, x_2, x_3, \dots, x_n\}$  和  $Y=\{y_1, y_2, y_3, \dots, y_n\}$  均为线性链表示的随机变量序列,在给定  $X$  的情况下,  $Y$  的条件概率分布  $P(Y|X)$  构成条件随机场,即满足马尔可夫性:

$$P(Y_i|X, Y_1, Y_2, \dots, Y_n) = P(Y_i|X, Y_{i-1}, Y_{i+1}).$$

在词性标注问题中,  $X$  表示观测序列,  $Y$  表示相应的标注序列或状态序列。

对于线性链 CRF, 给定一个输入序列(观测序列)  $X$ , 通过一组特征函数集合来对序列  $X$  每个元素的可能标签序列组合  $Y$  的得分进行计算, 计算公式为

$$score(Y|X) = \sum_{j=1}^m \sum_{i=1}^n \lambda_j f_j(y_{i-1}, y_i, X, i). \quad (1)$$

式(1)中:  $x_i$  为序列  $X$  中的第  $i$  个元素;

$y_i$  为第  $i$  个元素的标签;

$y_{i-1}$  为第  $i-1$  个元素的标签;

$f_j$  为特征函数;

$\lambda_j$  为特征权重。

在得到每个可能的标签序列的分数之后,可以通过幂运算和 softmax 对序列组合得分进行指数和归一化处理,从而获得标注序列的概率值。归一化方法如式(2)所示。

$$P(Y|X) = \frac{\exp[score(Y|X)]}{\sum_{Y'} \exp[score(Y'|X)]} = \frac{1}{Z(x)} \exp[score(Y|X)]. \quad (2)$$

### 2.3 基于深度学习和 CRF 的词性标注方法

由前面的描述可知, 神经网络模型和 CRF 在序列标注任务中各有优缺点: 神经网络模型能够捕获待标注序列的上下文信息, 但不能有效识别信息之间的依赖关系; CRF 不考虑长远的上下文信息, 但更多考虑整个序列局部特征的线性加权信息, 对具有依赖关系的信息识别效果较好。

为了能更好地捕获待标注序列的上下文信息和输出序列的依赖关系约束, 本文提出基于深度学习和 CRF 的印地语词性标注方法, 首先通过神经网络模型(如 CNN、LSTM、BiLSTM 等)训练得到输入序列的信息特征, 然后利用 CRF 对上一步的输出信息添加一些约束来保证词性标签预测的合理性。

假设  $X=\{x_1, x_2, x_3, \dots, x_n\}$  表示一个印地语句子, 其中  $x_i$  为句子  $X$  的第  $i$  个词,  $Y=\{y_1, y_2, y_3, \dots, y_n\}$  为句子  $X$  的词性标注序列。基于深度学习和 CRF 的印地语词性标注模型主要包括词表示层、序列表示层和 CRF 推理层。

1) 词表示层。根据输入数据  $X$  进行数据预处理, 并利用 word2vec 方法对大规模语料进行训练生成具有语义信息的低维度稠密实数词向量。

2) 序列表示层。将词向量传入到神经网络模型中进行训练, 得到输入序列的信息特征。

3) CRF 推理层。利用神经网络模型的输出状态和当前的转移概率矩阵作为 CRF 模型的参数, 最终获得标签序列  $Y$  的概率。

## 3 实验及分析

### 3.1 实验数据

词性标注方法的效果通常受如下因素影响: 一是受语料规模和语料所属领域范围的影响, 语料规模越大, 领域范围越集中, 词性标注效果越好; 二是受词性标记集合大小的影响, 词性标记集合的粒度越小, 标注效果随之降低; 三是受训练语料规模的影响, 标注的训练语料规模越大, 学习算法的学习效果越好, 越能得到好的结果。由于目前印地语缺乏公开的词性标注语料, 且相同模型对于不同词性标记集合的标注效果也会出现较大差异, 因此结合印地语语法特点及后续印地语自然语言处理任务要求, 项目组邀请印地语专家设计了表 1 所示印地语词性标注体系。

表 1 印地语词性标注体系  
Table 1 Part-of-speech tagging system of Hindi

序号	词性类别	符号	序号	词性类别	符号
1	一般名词	NN	12	后置词	PSP
2	时间名词	TN	13	量词	OTG
3	地点名词	LN	14	集合数词	CN
4	人名	NNP	15	基数词	QTC
5	反身代词	PRF	16	序数词	QTO
6	指示代词	PRP	17	连接词	CC_CCD
7	主动词	VM	18	从属词	CC_CCS
8	助动词	VAUX	19	疑问词	QW
9	系动词	LV	20	程度副词	INTF
10	形容词	JJ	21	标点符号	PTT
11	副词	RB			

由表 1 可知, 该体系包含 21 种词性类别构成的词性标记集合, 且具有相同词性的细粒度标记类别, 如名词包含了一般名词、时间名词、地点名词、人名 4 种, 动词包含了主动词、助动词、系动词 3 种。

同时, 邀请了 7 名印地语专业人员组成词汇标注

小组,对来自印地语的新闻文本进行人工标注。在人工标注时,每个句子均是由多个印地语专业人员达成标注一致性的结果,最终形成 5 950 条标注句子集,合计 114 127 个单词。

### 3.2 实验方案

本实验中,将词性标注集的 90% 作为训练集(共 5 355 个句子,108 756 个单词),10% 作为测试集(共 595 个句子,5 371 个单词)。训练集不仅用于训练词性标注模型,还用于训练词向量矩阵和字符向量矩阵。

为了评估词性标注方法的性能,使用准确率 *Accuracy* 作为本实验的效果评估指标,计算公式为

$$Accuracy = \frac{\text{正确标注的单词数量}}{\text{总的单词数量}}$$

### 3.3 模型参数设置与评估方法

模型实现使用 Python 语言及 TensorFlow 框架。词性标注模型经过多次实验调整后,选择最佳结果设置了如表 2 所示的超参数。在实验的正则化方法中,Dropout 通常会随机删除一些神经元,以防止模型出现过度拟合的情况。

表 2 模型超参数

Table 2 Hyperparameters of the model

组成部件	参数	数值
CNN	窗口大小	3
	过滤器数目	30
词向量	维度	400
BiLSTM	单元数	200
Dropout	CNN	0.5
	BiLSTM	0.5
优化器	损失函数	Adam
	学习率	0.01

### 3.4 实验结果及分析

为了对比不同方法在印地语词性标注任务中的效果,本文采用相同的标注语料在 TnT<sup>[22]</sup>、HMM<sup>[5]</sup>、CRF、BiLSTM、BiLSTM+LAN<sup>[23]</sup> 和 CNN+LSTM+CRF 等模型上进行实验。其中,TnT、HMM、CRF 为基于统计学习的方法,BiLSTM 和 BiLSTM+LAN 为基于深度学习的方法,但是 BiLSTM+LAN 是一种融合标签信息的(label attention network)深度学习的方法,主要是将标签进行词嵌入表示,然后将其传入 BiLSTM 进行训练,这样就能够捕捉到更长的标签依赖关系。后面几种为深度神经网络与统计学习模型相结合的方法。

由图 3 所示的多个模型的词性标注结果可以看出,CRF 是基于统计学习模型中效果最好的,与 HMM 方法相比,其性能约提高了 22%。BiLSTM 模型的效果较 CRF 和 BiLSTM+LAN 方法的差,而

CNN+LSTM+CRF 能够得到更好的效果。

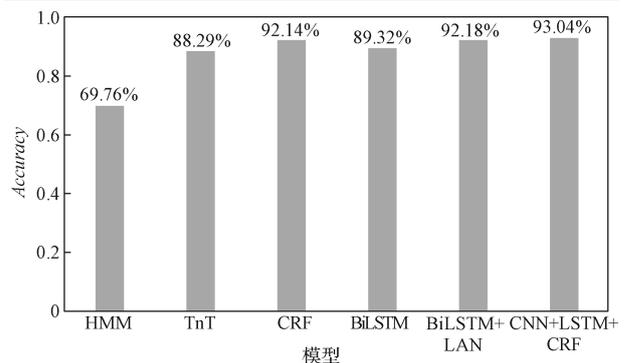


图 3 不同模型的词性标注结果

Fig. 3 Tagging result comparison of different models

整体上看,在深度学习模型基础上加上 CRF,其效果并未比深度学习模型有很大的提升,一个可能的原因是神经网络编码器已有很强序列信息编码能力,在此基础上加上 CRF 并未引入更多有效信息。

虽然提出的方法能够取得较好的效果,但相较于其他语种的词性标注结果尚有一定的提升空间。其原因在于,当前实验的标注语料规模还不够大,没有充分发挥出深度神经网络在大规模语料中的学习优势。

## 4 结语

本文提出了一种融合深度神经网络和统计学习的印地语词性标注方法。首先,通过词嵌入方法对印地语单词进行词向量表示;然后,将词向量作为深度神经网络模型的输入进行训练,进而获取单词的上下文信息;最后,通过 CRF 模型解码深度神经网络模型的输出,获取最优标注序列。实验结果表明,提出的方法较传统统计方法能得到更好的效果。

在未来研究工作中,将尝试端到端的训练模型,并进一步从词法分析角度挖掘印地语本身的语言特征,进而辅助深度学习模型的学习效果,从而提高印地语词性标注的准确性。

### 参考文献:

- [1] 金国哲,崔荣一.一种新的朝鲜语词性标注方法[J].中文信息学报,2018,32(10):53-58,68.  
JIN Guozhe, CUI Rongyi. A Novel Korean POS Tagging Method[J]. Journal of Chinese Information Processing, 2018, 32(10): 53-58, 68.
- [2] 帕力旦·吐尔逊,房鼎益.融合形态特征的最大熵维吾尔语词性标注[J].西北大学学报(自然科学版),2015,45(5):721-726.  
PALIDAN Tuerxun, FANG Dingyi. Fusion of Morphological Features for Uyghur Part-of-Speech

- Tagging Based on Maximum Entropy Model[J]. Journal of Northwest University(Natural Science Edition), 2015, 45(5): 721-726.
- [3] BRANTS T. TnT: A Statistical Part-of-Speech Tagger[C]//Proceedings of the Sixth Conference on Applied Natural Language Processing. Morristown: Association for Computational Linguistics, 2000: 224-231.
- [4] MISHRA N, MISHRA A. Part of Speech Tagging for Hindi Corpus[C]//2011 International Conference on Communication Systems and Network Technologies. Jammu: IEEE, 2011: 554-558.
- [5] JOSHI N, DARBARI H, MATHUR I. HMM Based POS Tagger for Hindi[C]//In Proceedings of 2nd International Conference on Artificial Intelligence and Soft Computing. Bangalore: Springer India, 2013: 341-349.
- [6] KUMAR D, JOSAN G S. Part of Speech Taggers for Morphologically Rich Indian Languages: A Survey[J]. International Journal of Computer Applications, 2010, 6(5): 32-41.
- [7] SINGH J, GARCHA L S, SINGH S. A Survey on Parts of Speech Tagging for Indian Languages[J]. International Journal of Advanced Research in Computer Science and Software Engineering, 2017, 7(4): 209-213.
- [8] AGARWAL H, AMNI A. Part of Speech Tagging and Chunking with Conditional Random Fields[C]// International Workshop on Replication in Empirical Softwar, Department of Computer Science. Hyderabad: Springer India, 2006: 1-4.
- [9] DALAL A, NAGARAJ K, SAWANT U, et al. Hindi Part-of-Speech Tagging and Chunking: A Maximum Entropy Approach[C/OL]//NLP AI Machine Learning Workshop: Part of Speech Tagging and Chunking for Indian Languages. Hyderabad: Springer India, 2006. [2020-03-24]. <http://pdfs.semanticscholar.org/7047/5e66b0d1cd5966d84027d88e99e95fd8dfa4.pdf>.
- [10] GARG N, GOYAL V, PREET S. Rule Based Hindi Part of Speech Tagger[C]//Proceedings of COLING 2012: Demonstration Papers. Mumbai: Springer India, 2012: 163-174.
- [11] ASOPA S, ASOPA P, MATHUR I, et al. Rule Based Chunker for Hindi[C]//2016 2nd International Conference on Contemporary Computing and Informatics. Noida: IEEE, 2016: 442-445.
- [12] MODI D, NAIN N. Part-of-Speech Tagging of Hindi Corpus Using Rule-Based Method[C]//Proceedings of the International Conference on Recent Cognizance in Wireless Communication & Image Processing. New Delhi: Springer India, 2016: 241-247.
- [13] YOUSIF J, SAINI D K. Hindi Part-of-Speech Tagger Based Neural Networks[J]. Journal of Computing, 2011, 3(2): 59-66.
- [14] NARAYAN R, CHAKRAVERTY S, SINGH V P. Neural Network Based Parts of Speech Tagger for Hindi[J]. IFAC Proceedings Volumes, 2014, 47(1): 519-524.
- [15] MISHRA N, JAIN S. POS Tagging of Hindi Language Using Hybrid Approach[C]//International Conference on Next Generation Computing Technologies. Singapore: Springer, 2017: 723-735.
- [16] MODI D, NAIN N, NEHRA M. Part-of-Speech Tagging for Hindi Corpus in Poor Resource[J]. Journal of Multimedia Information System, 2018, 5(3): 147-154.
- [17] MA X Z, HOVY E. End-to-End Sequence Labeling Via Bi-Directional LSTM-CNNs-CRF[EB/OL]. 2016: arXiv:1603.01354. [2020-01-20]. <https://arxiv.org/abs/1603.01354>.
- [18] HUANG Z, XU W, YU K. Bidirectional LSTM-CRF Models for Sequence Tagging[J/OL]. Computer Science, 2015. [2020-03-20]. <https://arxiv.org/pdf/1508.01991v1.pdf>.
- [19] HOCHREITER S, SCHMIDHUBER J. Long Short-Term Memory[J]. Neural Computation, 1997, 9(8): 1735-1780.
- [20] DYER C, BALLESTEROS M, LING W, et al. Transition-Based Dependency Parsing with Stack Long Short-Term Memory[C]//In Proceedings of ACL-2015 (Volume 1: Long Papers). Beijing: Association for Computational Linguistics, 2015: 334-343
- [21] MIKOLOV T, SUTSKEVER I, CHEN K, et al. Distributed Representations of Words and Phrases and Their Compositionality[J]. Advances in Neural Information Processing Systems, 2013, 26: 3111-3119.
- [22] GADDE S, YELETI M. Improving Statistical POS Tagging Using Linguistic Features for Hindi and Telugu[C/OL]// International Conference on Natural Language Processing, ICON-2008. [2020-03-20]. <https://researchweb.iiit.ac.in/~mehervijay.yeleti/papers/icon08pos.pdf>.
- [23] CUI L Y, ZHANG Y. Hierarchically-Refined Label Attention Network for Sequence Labeling[C]// Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing(EMNLP-IJCNLP). Hong Kong: Association for Computational Linguistics, 2019: 4106-4119.

(责任编辑: 廖友媛)