doi:10.6043/j.issn.0438-0479.201803047

结合树形概率和双向长短期记忆的渐步性句法分析方法

谌志群*,鞠 婷,王 冰
(杭州电子科技大学认知与智能计算研究所,浙江 杭州 310018)

摘要:为有效解决数据的稀疏性问题,并考虑句法预测的内在层次性,提出了一个基于双向长短时记忆(bidirectional long short term memory,BLSTM)神经网络模型的渐步性句法分析模型.该模型将树形概率计算方法应用到对句法标签分类的研究中,利用句法结构和标签之间的层次关系,提出一种从句法结构到句法标签的渐步性句法分析方法,再使用句法分析树来生成句法标签的特征表示,并输入到 BLSTM 神经网络模型里进行句法标签的分类.在清华大学语义依存语料库上进行实验的结果表明,与链式概率计算方法以及其他依存句法分析器比较,依存准确率提升了 0~1 个百分点,表明新方法是可行、有效的.

关键词:树形概率计算方法;双向长短时记忆;渐步性;依存句法分析;句法标签分类
中图分类号:TP 391 **文献标志码**:A **文章编号**:0438-0479(2019)02-0243-06

句法分析作为自然语言处理最关键的一个环节,起到承接词法分析和语义分析的作用,是信息提取、机器翻译等应用不可或缺的一部分.目前在句法分析研究中,除了应用概率方法对上下文无关文法进行句法结构分析外,研究人员还提出了依存句法理论.依存句法通过分析语言单位内成分之间的依存关系揭示其句法结构[1],依存关系可通过依存树表示.Hays[2]和 Gaifman[3]便是最早使用依存句法理论的研究者.Yamada[4]提出使用确定性的自上而下的支持向量机方法分析字词依赖,将树库数据转换为依赖树,达到了较好的依存准确度.Lai 等[5]使用基于跨度的思想将 CYK(Cocke、Younger 和 Kasami)算法应用于中文的统计依存句法分析,并采取其他措施避免重复工作.然而这些基于统计的模型存在着严重的数据稀疏问题.近年来,神经网络和深度学习模型在很多领域都取得了显著的成果,并逐渐用于句法分析的研究中.Collobert[6]提出一个基于深度递归卷积图转换模型(GTN)的句法分析算法,假设将分析树逐层分解,那么网络可根据树的上一层对句子进行分析.Chen 等[7]提出了一个贪心的基于神经网络的句法分析模型,将传统的稀疏特征替换为稠密特征,避免大量特征工程.Durrett 等[8]提出一种将条件随机场和神经网络相结合的模型进行句法结构分析,在宾州树库上取得与传统模型可比的效果.Ma 等[9]提出将卷积神经网络(CNN)和长短时记忆(LSTM)两种方法相结合进行句法分析,分别提取字级和字符级向量特征,取得较佳的效果.王衡军等[10]提出利用 LSTM 网络和分片池化的 CNN 分别提取词向量特征和全局特征,并输入到前馈网络中训练,有效提升了依存准确率.但是,这些模型大多是对句法结构和句法标记进行统一建模,而忽略句法分析本身的内在层次性,这样会使输出层增大,计算代价变大,导致分析速度很慢.

针对数据稀疏以及忽略句法分析层次性的问题,本研究提出一个结合树形概率计算方法和双向长短时记忆(BLSTM)神经网络模型的从句法结构到句法标签的渐步性句法分析方法,考虑到句法标记预测对句法结构的依赖,将对句法结构和句法标签的预测分步进行,根据树形概率计算方法选取高质量特征生成输入向量,以有效解决数据的稀疏性问题,提高句法标签的分类准确率,从而提高依存准确度.

收稿日期:2018-03-22 **录用日期**:2018-08-06
* **通信作者**:chenzq@hdu.edu.cn
引文格式:谌志群,鞠婷,王冰.结合树形概率和双向长短期记忆的渐步性句法分析方法[J].厦门大学学报(自然科学版),2019,58(2):243-248.
Citation:CHEN Z Q,JU T,WANG B. A step-by-step syntactic analysis method based on tree-like probability and bidirectional long short term memory[J]. J Xiamen Univ Nat Sci,2019,58(2):243-248.(in Chinese)

1 基于树形概率的句法标签分类方法

在句法分析里,通常需要根据特征对组块/叶节点进行分类,贴上相应的句法标签,而特征通常是指句法分析树上的节点,比如该节点的词语本身、词性、位置信息等.当使用这些特征计算某个节点的标签时,一般方法是计算出该节点多个有可能的句法标签对应的概率值,从中找到概率值最大的句法标签,即为该节点的句法标签.通常在计算时并不会使用到所有的特征,节点 w 的句法标签值 $sc(w)$ 的计算过程为

$$sc(w) \approx \underset{1\leqslant i\leqslant k}{\operatorname{argmax}} \hat{P}(l_i \mid f_1, f_2, \cdots, f_j), \quad 0 < j < n, \tag{1}$$

其中,$\hat{P}$ 为从语料库数据集里统计出的经验观测概率,l_i 为句法标签,f 为特征项.式(1)中的计算过程就是如图 1 所示条件概率的计算过程,如果某个特征 f_i 对句法标签 l_i 的影响很大,那么它的条件概率 $\hat{P}(l_i \mid f_1, f_2, \cdots, f_i)$ 就会使整体的概率值向变大的趋势发展,反之亦是.

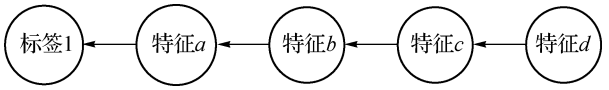

图 1 链式概率计算方法示意图

Fig. 1 A schematic diagram of chained probability calculation method

链式概率计算方法并没有充分地捕捉到某些特征的信息量.如果两个特征是互相独立的并且携带有大量对于计算整体概率有用的信息,但由于是链式的概率计算方法,排在前面的概率值会严重影响后面概率值的作用.将链式概率计算方法转换为树形的概率计算方法,亦即使两个稀疏的独立特征分开,共同依赖于前驱节点,放到树里来说即共同依赖于父节点,其概率计算方法如图 2 所示.

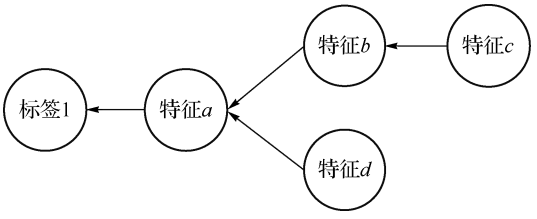

图 2 树形概率计算方法

Fig. 2 Tree-like probability calculation method

树形概率计算方法与链式概率计算方法的不同点在于链式概率计算方法里,每个特征项 f_i 的概率值都要与它前面的特征项 f_{i-1} 的概率值相乘,但是在树形的计算方式上,特征项 f_i 只与那些与它在同一条路径上的特征项的概率值相乘,并且用根节点的概率值作为分母.树形概率计算方法并不会将所有的分子分母都约去.在概率树上的叶节点都会出现在分子上,而每一个分叉点上的节点都至少会在分母上出现一次.如图 2 所示,有一个整体概率值和 4 个特征概率值,特征 b 和特征 d 是相互独立的,共同依赖于特征 a;特征 c 依赖于特征 b.则图 3 的概率计算如式(2)所示.

$$P(l \mid a,b,c,d) \approx \frac{\hat{P}(l \mid a,b,c)\,\hat{P}(l \mid a,d)}{\hat{P}(l \mid a)}, \tag{2}$$

由此可见,在树形的概率计算方法里,某个稀疏项的概率值不会决定整个算式的最终计算结果.

在句法分析里,如果先预测出句法结构,则可以使用上述概率计算原理来作为特征选取的标准,如图 3 所示句子"世界第八大奇迹出现"的句法结构,链式概率计算方法无法将依存于"奇迹"的两个独立的节点"世界"与"大"区分开,而树形概率计算方法中,节点"出现"依存于虚拟根节点"ROOT",节点"奇迹"依存于前驱节点"出现",节点"世界"和节点"大"相互独立,共同依存于前驱节点"奇迹",节点"第"和节点"八"依存于节点"大".在选取节点"世界"的特征时,可以选取与之有依存关系的节点"奇迹"、节点"出现",而不再选取只是与之在位置上相连的节点"第"或者节点"八".

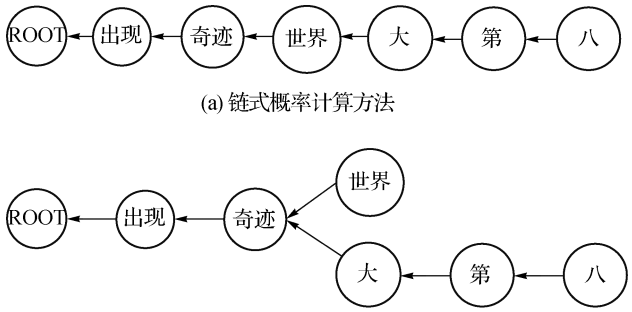

图 3 概率计算方法示例

Fig. 3 Example of probability calculation method

根据上述思路,在进行句法标签处理时选取特定的节点作为特征进行分类,具体到某一个节点来说,包括该节点的词语本身、父节点的词语本身、父节点的句法标签、祖父节点词语本身、祖父节点的词性、子节点的词语本身、子节点的词性等特征.

2　BLSTM 模型实现及训练

本研究首先直接使用语料库数据集里已有的依存结构,从而将句法结构和句法标签的分析分开进行,充分利用句法体现其句法分析的渐步性,再结合树形概率计算方法选取词语、词性以及距离作为特征生成分布式向量,输入到 BLSTM 模型里进行句法标签的分类工作,具体流程如图 4 所示.

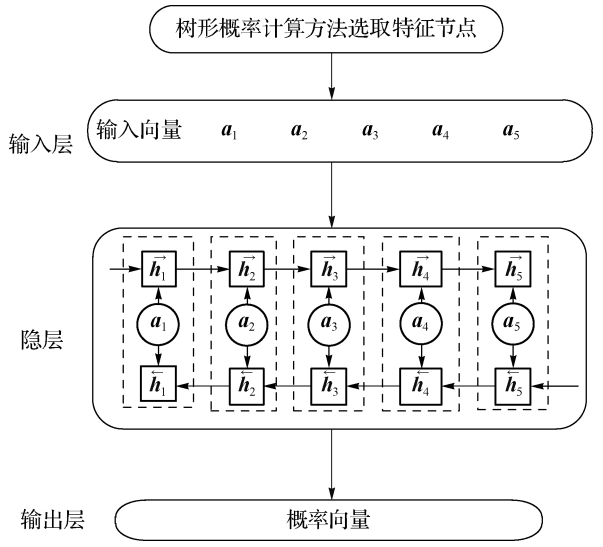

图 4　神经网络结构

Fig. 4　Neural network structure

2.1　输入层

在神经网络模型里,词语向量、词性向量和距离向量被映射为分布式向量作为模型的输入.对于选取的特征节点的词语组块 $\boldsymbol{x}_i \in \mathbf{R}^{3d_a}$,由词语的分布式向量 $\boldsymbol{e}_{wi} \in \mathbf{R}^{d_a}$、词性的分布式向量 $\boldsymbol{e}_{pi} \in \mathbf{R}^{d_a}$ 以及距离向量 $\boldsymbol{l}$ 组合作为该组块的分布式向量表示,其中,d_a 为分布式向量的维度,计算过程如式(3)所示.

$$\boldsymbol{a}_i = g_1(\boldsymbol{W}_a[\boldsymbol{e}_{wi};\boldsymbol{e}_{pi};\boldsymbol{l}] + \boldsymbol{b}_a), \tag{3}$$

其中:$1 \leqslant i \leqslant n$,$n$ 为句子中选取的特征节点数;$\boldsymbol{a}_i \in \mathbf{R}^{d_a}$;$\boldsymbol{W}_a \in \mathbf{R}^{d_a \times 3d_a}$ 是权重矩阵;$\boldsymbol{b}_a$ 是偏执向量,维度为 d_a;激活函数 g_1 采用 sigmoid 函数.

2.2　隐层和输出层

如图 4 所示,模型将输入向量输入到隐层中.标准的 LSTM 递归神经网络一般从一个方向计算隐层向量序列 $\boldsymbol{h}$,而 B LSTM 模型则从两个不同方向计算独立隐层,最后将输出合并起来输入到输出层里.其中前向隐层向量为 $\overrightarrow{\boldsymbol{h}}$,后向隐层向量为 $\overleftarrow{\boldsymbol{h}}$,隐层在第 i 时刻的输出向量为 $\boldsymbol{v}_i$,计算方式如式(4).

$$\boldsymbol{v}_i = \overrightarrow{\boldsymbol{h}}_i + \overleftarrow{\boldsymbol{h}}_i, \tag{4}$$

其中,$\overrightarrow{\boldsymbol{h}}_i \in \mathbf{R}^{d_h}$,$\overleftarrow{\boldsymbol{h}}_i \in \mathbf{R}^{d_h}$.

使用 tanh 函数作为激活函数 g_2 来进行隐层的计算,其中 $\boldsymbol{h}$ 计算方式如式(5)所示.

$$\boldsymbol{h}_i^d = g_2(\boldsymbol{W}_h^d \boldsymbol{a}_i + \boldsymbol{U}\boldsymbol{h}_{i-1}^d + \boldsymbol{b}_h^d), \tag{5}$$

其中:$\boldsymbol{W}_h^d \in \mathbf{R}^{d_h \times d_a}$ 是第 d 个索引对应的 $\boldsymbol{a}_i$ 的权重矩阵,$d \in \{0,1\}$ 代表在隐层的两个不同的方向;$\boldsymbol{U}$ 对应 $i-1$ 时刻隐层输出 $\boldsymbol{h}^d$ 的权重矩阵;$\boldsymbol{b}_h^d \in \mathbf{R}^{d_h}$ 是第 d 个索引对应的偏执向量.

输出层增加到最后一层隐层的顶部来评分依赖弧,评分公式如式(6)所示.

$$\boldsymbol{SC}(\boldsymbol{x}_i, \boldsymbol{y}_i) = \boldsymbol{W}_o^d \boldsymbol{v}_i + \boldsymbol{b}_o^d, \tag{6}$$

其中:$\boldsymbol{W}_o^d \in \mathbf{R}^{L \times d_h}$ 为依赖弧的权重矩阵;$\boldsymbol{b}_o^d \in \mathbf{R}^L$ 为偏执向量;$\boldsymbol{SC}(\boldsymbol{x}_i, \boldsymbol{y}_i) \in \mathbf{R}^L$ 为输出向量,代表句法标签的评分,输出向量的每一维度的值为对应状态的概率值;L 为句法标签的个数.

2.3　训练似然性

使用最大边距准则(max-margin criterion)来训练模型.给出一个训练距离 $(\boldsymbol{x}_i, \boldsymbol{y}_i)$,用 $Y(\boldsymbol{x}_i)$ 来表示可能的状态序列的集合,用 $\boldsymbol{y}_i$ 来表示词语组块 $\boldsymbol{x}_i$ 正确的状态序列.最大边距训练方法的目的是找到参数集合 Θ 使正确的状态序列 $\boldsymbol{y}_i$ 与不正确的状态序列 $\hat{\boldsymbol{y}} \in Y(\boldsymbol{x}_i)$ 评分之差至少为边距损失函数 $\nabla(\boldsymbol{y}_i, \hat{\boldsymbol{y}})$,公式为

$$S(\boldsymbol{x}_i, \boldsymbol{y}_i; \Theta) \geqslant S(\boldsymbol{x}_i, \hat{\boldsymbol{y}}) + \nabla(\boldsymbol{y}_i, \hat{\boldsymbol{y}}), \tag{7}$$

其中,$S(\boldsymbol{x}_i, \boldsymbol{y}_i; \Theta) = \sum_{k=1}^{L} SC_k(\boldsymbol{x}_i, \boldsymbol{y}_i; \Theta)$, (8)

$$\nabla(\boldsymbol{y}_i, \hat{\boldsymbol{y}}) = \sum_{j}^{n} k\{b(\boldsymbol{y}_i, x_i^j) \neq b(\hat{\boldsymbol{y}}, x_i^j)\}, \tag{9}$$

Θ 为模型参数 $\{\boldsymbol{W}_a, \boldsymbol{b}_a, \boldsymbol{b}_h^d, \boldsymbol{U}, \boldsymbol{W}_h^d, \boldsymbol{W}_o^d, \boldsymbol{b}_o^d\}$,$b(\boldsymbol{y}_i, x_j^i)$ 为组块 $\boldsymbol{x}_i$ 里的第 j 个词语在状态序列 $\boldsymbol{y}_i$ 里对应的标签,k 为损失参数,损失函数的值与该词语在状态序列 y_i 里的数量和该词语在状态序列 $\hat{\boldsymbol{y}}$ 里的数量有关.

在给定训练集的数量大小 m 的情况下,模型的正则化目标函数是损失函数 $J(\Theta)$ 与 L_2 范式之和:

$$J(\Theta) = \frac{1}{m}\sum_{i=1}^{m} l_i(\Theta) + \frac{\lambda}{2}\|\Theta\|_2^2,$$

其中,λ 为正则化参数,$l_i(\Theta)$ 如式(9)所示.

$$l_i(\Theta) = \max_{\boldsymbol{y} \in Y(\boldsymbol{x}_i)} (S(\boldsymbol{x}_i, \hat{\boldsymbol{y}}; \Theta) + \nabla(\boldsymbol{y}_i, \hat{\boldsymbol{y}})) - S(\boldsymbol{x}_i, \boldsymbol{y}_i; \Theta), \tag{10}$$

可见通过减小该目标函数的值,正确状态序列的概率值增加,不正确状态序列的概率值则会减小.

由于折页损失,目标函数是不可微的,因此使用次梯度方法.采用对角变量进行参数的优化,在时刻 i 第 j 个参数 $\Theta_{i,j}$ 的更新步骤如式(11)所示.

$$\Theta_{i,j}=\Theta_{i-1,j}-\frac{\alpha}{\sqrt{\sum_{\tau=1}^{i}g_{\tau,j}^{2}}}g_{i,j}, \tag{11}$$

其中:α 为初始学习率,将其设定为 0.2;$g_{\tau,j}$ 是参数 Θ_j 在时刻 τ 的次梯度. 为了降低过拟合的可能性,使用丢包(dropout)来泛化模型,设定隐层上的 dropout 值为 0.2.

3 实 验

3.1 语料库数据集设置

在实验的数据准备阶段,为了将句法结构和句法标签的分析分开进行,使用清华大学的语义依存语料库数据集,保存其依存结构后将前 80%的句子作为训练集,10%的句子作为验证集,10%的句子作为测试集. 使用 BLSTM 模型在训练集上进行训练,在验证集上验证模型的效果并进行参数调整,在测试集上进行测试. 为了尽可能准确地验证提出的特征选择方法,实验直接使用了树库的依存结构,在实验过程中不再独立进行依存结构的分类.

模型超参数的设定如表 1 所示,并使用表 1 中的参数进行向量的预训练,其他所有参数均在区间[−0.05,0.05]里均匀采样获得. 树库的数据格式示例如表 2 所示.

表 1 模型超参数设定

Tab. 1 Super parameter setting of the model

超参数	设定值
词语分布式向量维度	100
词性分布式向量维度	100
隐层向量维度	100
BLSTM 隐层层数	2
正则化参数 λ	0.2

表 2 语料库数据集示例

Tab. 2 Corpus dataset example

序号	词语	粗粒度词性	细粒度词性	中心词个数	依存关系
1	世界	名词	名词	5	限定
2	第	数词	数词	4	限定
3	八	数词	数词	2	连接依存
4	大	形容词	形容词	5	限定
5	奇迹	名词	名词	6	存现体
6	出现	动词	动词	0	核心成分

3.2 评测指标设计

在清华大学语义依存语料库数据集里,句子数大约有 20 000,词的个数大约有 160 000,共统计出 69 个依存标签. 实验对这些标签进行了分类,在句法分析树里一个节点可以有多个句法标签,但不允许有两个或者两个以上的标签来自于同一类. 按照测试集里不同句法标签的分布比例,如果有某个句法标签的分布比例大于等于总句法标签数量的 10%,则将其单独成为一类,否则,将所有的低频句法标签组为一类. 分布比例的计算式定义为:分布比例=测试集里该句法标签出现的数量/测试集里所有的句法标签数量×100%. 在所有的句法标签里,只有“核心成分”和“限定”各自单独成为一类,因此,将语料库数据集里的句法标签分为了 3 类.

对每个句法标签分别计算其分类准确率 P'、召回率 R 和综合指标 $F1$. 具体计算式为:

$$P'=A/B\times 100\%,$$

$$R=A/C\times 100\%,$$

$$F1=2\times P\times R/(P+R)\times 100\%,$$

其中,A 为某类正确的句法标签总数,B 为自动标注为该类句法标签的词语总数,C 为标准答案中该类句法标签的词语总数.

将所有句法标签放在一起评价时,依存分析的结果采用依存弧准确率(UAS)和依存关系准确率(LAS)作为评价指标.

$$\mathrm{UAS}=X/Y\times 100\%,$$

$$\mathrm{LAS}=Z/Y\times 100\%,$$

其中,X 为核心节点正确的词数,Y 为总词数,Z 为核心节点正确、并且对应依存关系类型也正确的词数.

4 结果分析

本节将通过实验结果验证本文中所给出模型的句法分析效果,为解决数据稀疏性对结果分析的影响,特将树形概率计算方法和链式概率计算方法两种特征提取方法进行对比实验,结果如表 3 所示. 可以看出树形概率计算方法的实验效果明显好于链式概率计算方法,观察“其他依存句法标签”类,树形概率计算方法的优势更为明显,平均高出了 4～5 个百分点,相较于链式概率计算方法能够更好地处理稀疏句法标签的分类问题. “核心成分依存句法标签”的实验效果最好,达到了 90%的准确率,说明基于结构化的词语、词性以及丰富的上下文信息能够极好地反映句子中的关联信息.

表 3　句法标签实验结果

Tab. 3　Experimental results of syntactic label　　%

方　法	核心成分			限定句法			其他		
	P'	R	F_1	P'	R	F_1	P'	R	F_1
链式概率计算方法	85.37	83.77	85.39	79.61	81.24	81.31	82.37	82.90	83.49
树形概率计算方法	**90.47**	**88.83**	**90.65**	**83.41**	**85.59**	**85.98**	**86.19**	**85.56**	**88.87**

在对所有标签数据进行实验中,baseline1 选择采用经典的基于图的依存分析器 MSTParser,baseline2 系统采用经典的基于转移的依存分析器 MaltParser. 此外,本文中还与其他准确率较高的中文依存分析器进行了对比,分别包括:由 CNN 和 LSTM 两种方法相结合的句法分析器[9]、基于图的层次长短期记忆神经网络(hierarchy long-short memory term model, HLSTM)句法分析模型[11]、基于 BLSTM 的句法分析模型[12]. 另外,针对句法分析层次性问题,将本研究的模型与引入层次成分分析的句法分析模型[13]再进行对比. 由于模型的实验直接使用语料库数据集里的依存结构,而前 3 种句法分析模型则是统一对依存结构和句法标签进行建模预测,为了实验的客观性,在分析实验结果时,去除前 3 种句法分析模型中错误的结果,同时也去除在模型实验结果里对应的句子,其依存分析结果如表 6 所示.

表 4　依存分析实验结果

Tab. 4　Dependency parsing experimental results　　%

模　型	准确率	
	UAS	LAS
baseline1	82.11	81.35
baseline2	82.72	81.27
CNN+LSTM[9]	88.83	85.30
HLSTM[11]	87.77	86.42
BLSTM[12]	87.76	86.34
层次成分分析法[13]	84.17	79.62
链式概率计算方法+BLSTM	82.33	79.73
树形概率计算方法+BLSTM	87.83	84.52

从表 4 可以看出,本研究提出模型的依存准确率均高于两个 baseline 系统约 5 个百分点;树形概率结合 BLSTM 最终的依存准确率远远高于链式概率计算方法,因此,无论是对不同成分进行实验对比还是在整体上对比,前者效果均好于后者;文献[13]中所提的引入层次成分分析的句法分析方法,在考虑内部结构的基础上提出解决长距离依赖问题,从最后实验结果对比得知,本文中所提出的方法略高于该方法,可看出本文在长距离依赖问题上没有很突出;对比文献[12]仅使用 BLSTM 模型的结果,本文中提出的模型依存准确率略高,但优势并不明显,可能存在数据集的差距;对比文献[11]和文献[12]基于神经网络模型存在的忽略句法分析本身的内在层次性的问题,观察发现依存准确率略高,可看出所提出模型具有一定的有效性;最后对比文献[9]结合 CNN 和 LSTM 的模型准确率高于本文中所提出模型,达到了最佳,而本文模型与之差距较大,还有待进一步提高. 另外影响依存句法分析准确率的还有语料库数据集上存在一些错误的标注的原因,一些句法标签通过人工进行识别很容易判断出标注是错误的,比如句子"六位数的薪水可以雇佣很多精英",训练集里给"雇佣"的标注是"关系主体",而正确的标注应该是"核心成分".

5　结　论

本研究的主要工作是针对数据稀疏以及神经网络模型忽略句法结构和标签之间层次关系,结合树形概率计算方法提取句法结构中的上下文信息,并将对句法结构的分析和句法标签的分析分开处理,提出一种从句法结构到句法标签的渐步性句法分析方法. 同时考虑到数据的稀疏性问题,树形概率计算方法能够使区分能力强的特征在整体概率值的计算中更好地发挥作用,而最终的实验可证明该方法在不同的句法标签类别上取得了较高的准确率. 然而未考虑到句法分析需要依赖长距离的信息,在进一步的研究中可从这个方面改善模型,更好地提高句法分析的准确性.

参考文献:

[1]　刘海涛. 依存语法和机器翻译[J]. 语言文字应用,1997,23(3):89-93.

[2]　HAYS D G. Dependency theory: a formalism and some

observations[J]. Language,1964,40(4):511-525.

[3] GAIFMAN H. Dependency systems and phrase-structure systems[J]. Information and Control, 1965, 8(3): 304-337.

[4] YAMADA H. Statistical dependency analysis with support vector machines[C]// Proceedings of the 8th International Workshop on Parsing Technologies. Nancy: International Workshop on Parsing Technologies,2003:195-206.

[5] LAI T B Y, HUANG C, ZHOU M, et al. Span-based statistical dependency parsing of chinese[C]// Proceedings of the 6th Natural Language Processsing Pacific Rim Syposium (NLPRS2001). Tokyo: National Center of Sciences,2001:677-684.

[6] COLLOBERT R. Deep learning for efficient discriminative parsing[C]// International Conference on Artificical Intelligence and Statistics. Lauderdate: AISTATS, 2011: 224-232.

[7] CHEN D, MANNING C D. A fast and accurate dependency parser using neural networks[C]// Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing. Doha: Association for Computational Linguistics, 2014:740-750.

[8] DURRETT G, KLEIN D. Neural CRF Parsing[EB/OL]. [2018-03-01]. http://arxiv.org/abs/1507.03641.

[9] MA X, HOVY E. Neural probabilistic model for non-projective MST parsing[C]// Proceeding of the 8th International Point Conference on Natural Language Processing. Taipei: IJCNLP, 2017:59-69.

[10] 王衡军,司念文,宋玉龙,单义栋. 结合全局向量特征的神经网络依存句法分析模型[J]. 通信学报,2018,39(2):53-64.

[11] WANG W, CHANG B. Improved graph-based dependency parsing via hierarchical LSTM networks[C]// China National Conference on Chinese Computational Linguistics. Yantai: Springer International Publishing, 2016:25-32.

[12] ZHANG X, CHENG J, LAPATA M. Dependency parsing as head selection[EB/OL]. [2018-03-01]. http://arxiv.org/pdf/1606.01280.pdf.

[13] 张丹,周俏丽,张桂平. 引入层次成分分析的依存句法分析[J]. 沈阳航空航天大学学报,2017,34(1):76-82.

A step-by-step syntactic analysis method based on tree-like probability and bidirectional long short term memory

CHEN Zhiqun*, JU Ting, WANG Bing

(Institute of Cognitive and Intelligent Computing, Hangzhou Dianzi University, Hangzhou 310018, China)

Abstract: In order to effectively solve the problem of data sparseness and inherent level of syntactic prediction, an incremental stepwise dependency parsing model based on bidirectional long short term memory (BLSTM) is proposed. This paper applying the tree-like probability calculation method to the study of syntactic tag classification, using the hierarchical relationship between syntactic structure and tag, proposes a step-by-step syntactic analysis method from syntactic structure to syntax tag, using syntactic analysis tree to generate the characteristics of the syntactic tag which are input into the BLSTM model to classify syntactic tags. Compared with other syntactic analysis methods and chained probability calculation method on the Semantic Dependency Corpus dataset of Tsinghua University, the dependency accuracy rate is improved by 0-1 percent. It shows that the new method is feasible and effective.

Keywords: tree-like probability calculation method; bidirectional long short term memory; step-by-step; dependency parsing; syntactic label classification