

融合 Gate 过滤机制与深度 Bi-LSTM-CRF 的汉语语义角色标注

张苗苗 刘明童 张玉洁 徐金安 陈钰枫

北京交通大学计算机与信息技术学院 北京 100044

摘要 语义角色标注的传统方法采用基于句法特征的统计机器学习方法。由于依存句法可以表示词语之间的语义关系，故在语义角色标注中取得了较好的性能；但该方法存在特征抽取过程繁琐，难以捕捉句子中长距离依赖等问题。随着深度学习的兴起，研究者将基于双向长短时记忆（Bidirectional Long Short-Term Memory, Bi-LSTM）神经网络模型用于语义角色标注。该模型可以自动学习特征，并对词与词之间的远距离依赖关系进行有效建模。本文提出融合 Bi-LSTM-CRF 模型与依存句法特征的方法，并且引入 Gate 过滤机制对词向量表示进行调整，以达到利用句法特征提高语义角色标注精度的同时，规避特征工程的繁琐。CPB 上的实验结果表明，利用本文所提方法的汉语语义角色标注的 F1 值达到 79.53%，比前人的方法有了较为显著的提升。

关键词：汉语语义角色标注；Gate 过滤机制；Bi-LSTM-CRF；依存句法分析

中图分类号：TP391

开放科学（资源服务）标识码（OSID）



The Integration of Gated Filtering Mechanism and Deep Bi-LSTM-CRF for Chinese Semantic Role Labeling

ZHANG Miaomiao LIU Mingtong ZHANG Yujie XU Jinan CHEN Yufeng

The School of Computer and Information Technology, Beijing Jiaotong University, Beijing 100044, China

基金项目：北京交通大学人才基金（KKRC11001532）；国家自然科学基金（61370130，61473294）；北京市自然科学基金（4172047）。

作者简介：张苗苗（1992-），硕士，研究方向：自然语言处理、语义分析，Email：15120465@bjtu.edu.cn；刘明童（1993-），博士，研究方向：自然语言处理、复述；张玉洁（1961-），博士，教授，研究方向：自然语言处理、机器翻译、文本大数据处理；徐金安（1970-），博士，副教授，研究方向：自然语言处理、机器翻译；陈钰枫（1981-），博士，副教授，研究方向：自然语言处理、机器翻译。

Abstract The traditional statistical methods which based on the syntactic features algorithm were frequently used for the Chinese semantic role labeling. Since the dependency parsing provides semantic relations between words, better performances in semantic role labeling were achieved. However, hand-crafted feature extraction process was complicated in such methods and it is difficult to capture the long range dependences in a sentence. With the development of deep learning, researchers have applied the bidirectional long short-term memory (Bi-LSTM) model to semantic role labeling, which is capable of learning features automatically and capturing long-range dependence. This paper proposed a method of combining model (Bi-LSTM) with dependency structure and introduced a Gated filtering mechanism (GFM) to adjust the word representation. Experimental results on CPB showed that the proposed method achieved 79.53% of F1 in Chinese semantic role labeling and significantly outperformed the previous work.

Keywords: Chinese semantic role labeling; gated filtering mechanism; Bi-LSTM-CRF; dependency parsing

1 引言

在自然语言处理中，语义角色标注（Semantic role labeling, SRL）是语义分析的基础与核心技术，其主要目的是识别出句子中所有与谓词相关的语义角色，并且赋予它们相应的角色类型，如施事、受事、时间、地点、原因等。其中，核心的语义角色为 ARG0~ARG5，ARG0 表示动作的施事，ARG1 表示动作的受事，ARG2~ARG5 则根据谓语动词的不同具有不同的语义含义；其余的语义角色为附加语义角色，使用 ARGM-X 表示，如 ARGM-LOC 表示地点，ARGM-TMP 表示时间等。下面给出一个语义角色标注的例子，“[委员会]_{ARG0}[明天]_{ARGM-TMP}[将要]_{ARGM-ADV}[通过]_{Pred}[议案]_{ARG1}”，其中，“通过”为谓词，“委员会”、“明天”、“将要”、“议案”分别为其施事、发生的时间、修饰性副词以及受事。语义角色标注作为自然语言处理的底层技术，在信息抽取、问答系统、机器翻译等任务中

有着广泛应用，语义角色标注方法的深入研究以及自动标注精度的提高对自然语言处理的整体发展有着重要意义。

传统的基于特征的统计机器学习方法，通常将语义角色标注任务转换为有监督的分类问题，主要分为基于短语结构句法分析以及基于依存句法分析的两种语义角色标注方法。由于依存句法分析能够表达词语之间的依赖关系，并且强调动词在句子中的重要作用，故能为语义角色标注提供更有效的信息。但是，传统的基于特征的方法通常需要抽取很多人工设定的特征，因此存在特征稀疏、模型复杂和容易过拟合等问题。随着神经网络的发展，其自动学习特征规避传统特征工程的优势，近年来在自然语言处理中得到有效应用。2011 年，Collobert 等^[1]提出一种基于卷积神经网络（Convolutional Neural Networks, CNN）的模型，在对英文进行词性标注、语块分析、实体识别和语义角色标注任务上取得了较好的效果。

本文采用在两层 Bi-LSTM-CRF 神经网络模型基础上, 引入依存句法特征的方法实现汉语语义角色标注。该方法 (1) 首先将词、词性以及依存句法等特征映射为实数向量, 并将该向量作为神经网络的输入; (2) 为了减小 n-gram 对词向量表示的影响, 引入 Gate 过滤机制对特征向量进行调整; (3) 然后采用两层 Bi-LSTM 获取待标注词在句子中的上下文信息; (4) 为了防止出现标注偏置问题, 引入标签转移概率矩阵, 并且使用 CRF 融合全局标签信息得到最优标注序列。

本文组织如下: 第 2 节介绍语义角色标注的相关研究; 第 3 节介绍引入依存句法特征的两层 Bi-LSTM-CRF 语义角色标注方法; 第 4 节介绍网络模型的训练方式; 第 5 节通过实验验证所提方法的性能, 并与已有方法比较; 最后对本文工作进行总结, 并提出今后研究方向。

2 相关研究

在语义角色标注研究的初始阶段, 大多数的语义角色标注系统都是利用短语结构句法分析。随着依存句法分析技术的不断进步, 基于依存句法分析的语义角色标注方法也逐渐发展起来。依存句法分析所给出的依存树与传统的短语结构句法树相比, 在结构上更加扁平化, 使得语义角色与谓词在句法树上的距离相对缩短, 从而谓词-论元信息呈现更加直接的关系。此外, 基于依存句法分析的语义角色标注方法不仅可以利用短语之间的依存关系, 而且只需要关注与谓词有依存关系的短语, 学习和预测这些短语的语义角色。

Hacioglu^[2] 最早实现了利用依存句法分析进行语义角色标注; 而在 CoNLL 2008 和 CoNLL 2009 两届依存句法分析和语义角色标注联合评测任务中, 更是强调利用依存句法结构进行语义角色标注。在 CoNLL 2009 评测中, 汉语语义角色标注的最好成绩 F1 值达到 78.60%^[3]。为了降低依存句法分析错误对语义角色标注性能的影响, Jin^[4] 等提出只利用高质量的句法分析结果进行汉语语义角色标注的方法。国内学者汪红林^[5]、王步康^[6]、王鑫^[7] 等也对基于依存句法分析的汉语语义角色标注方法展开研究。此外, Yang^[8] 等对句中包含多个谓词的现象进行了研究, 充分考虑了多谓词之间的相互关系和共享论元的全局信息, 提出了基于判别式重排序的多谓词语义角色标注方法, 显著提升了共享论元分类的效果。

上述方法主要基于支持向量机、条件随机场以及最大熵模型等机器学习算法, 针对基本特征及特征组合对语义角色标注进行研究。这些方法直接使用词、词性等符号信息, 需要人工指定特征, 难以进一步抽取更有效的特征, 并且难以利用特征之间的语义相关性。此外, 由于词特征数量较多, 使得模型的特征矩阵维度很高, 容易导致模型的过拟合。

自 2006 年深度学习兴起以来, 很多学者开始使用深度神经网络方法解决自然语言处理的问题。在语义角色标注方面, Collobert 等^[1] 使用 CNN 模型在英语上的工作表明, 减少了大量的特征抽取工作, 并且与传统的基于特征的机器学习方法相比, 实验结果接近英语语义角色标注的最好水平。Zhou 等^[9] 使用深度 Bi-LSTM 模型对英语语义角色标注进行了研究。Roth 等^[10]

将依存句法路径作为特征加入到 LSTM 模型中, 在英语方面取得了不错的效果。在汉语语义角色标注方面, 王臻等^[11]使用分层输出的神经网络模型, 将角色识别和角色分类分开进行, 最终 F1 值达到 64.19%; Wang 等^[12]使用 Bi-LSTM 模型, 在不引入任何其他资源的情况下, 取得了当时最好的结果; Sha 等^[13]引入依存信息用于论元关系的识别, 在 CPB 语料上 F1 值取得了 77.69%; Guo 等^[14]重点关注句法路径信息并使用 Bi-LSTM 对其进行建模, 从而提高了系统的性能。王瑞波等^[15]使用多特征融合的神经网络结构来构建汉语框架语义角色识别模型。

近几年, 性能较好的语义角色标注系统大多基于 Bi-LSTM 神经网络模型。受到 LSTM^[16]中的门控单元以及 Tu 等^[17]在机器翻译模型中使用的上下文门 (Context gates) 的启发, 本文提出在 Bi-LSTM 层之前, 加入 Gate 过滤机制

对词向量表示进行调整; 并提出将依存句法特征融合到 Bi-LSTM 模型中的方法。CPB 上的实验结果表明, 我们的方法在使用自动依存句法分析时, 汉语语义角色标注的 F1 值达到 79.53%, 比前人方法有了较为显著的提升。

3 汉语语义角色标注模型

本文采用图 1 所给出的神经网络结构来构建汉语语义角色标注框架。自底向上, (1) 首先, 将词、词性、谓词、是否为谓词、词到谓词的距离和依存关系六种特征的实数向量拉直拼接, 作为当前词的特征向量表示; (2) 然后 Gate 过滤机制对输入信息进行调整; (3) 之后两层 Bi-LSTM 获取每个词长距离的上下文特征; (4) 最后 CRF 层考虑单词标签之间的制约关系, 加入标签转移概率矩阵, 给出全局最优标注序列。

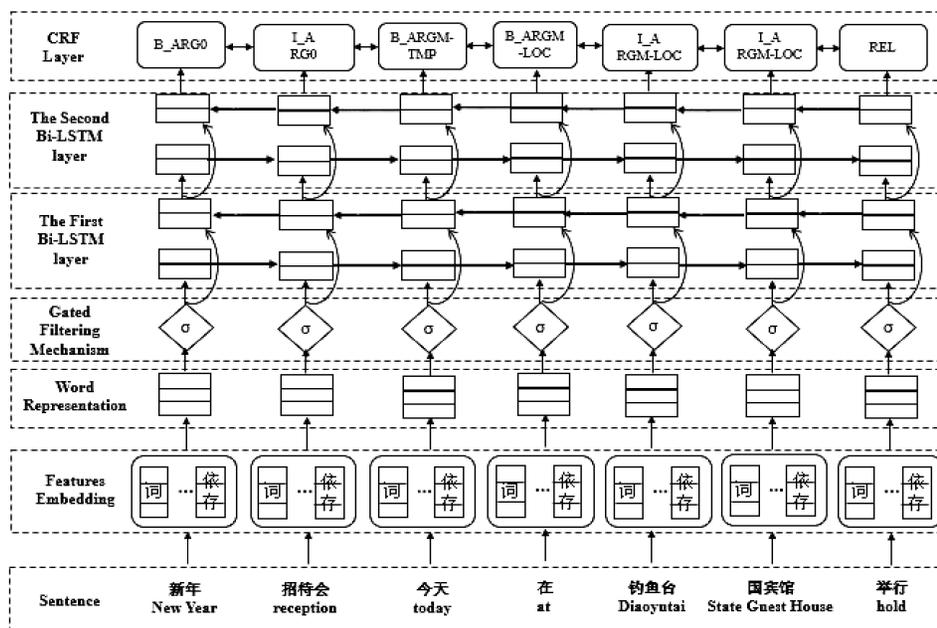


图1 融合Gate过滤机制与两层Bi-LSTM-CRF模型的汉语语义角色标注框架

3.1 依存关系特征

依存句法分析是通过分析句子中各个成分之间的修饰关系来表示其句法结构，如主谓关系和动宾关系等。依存语法中句子的核心成分是中心谓语动词，句子中的其他词语均受中心词的支配，而中心词本身不会受到句子中其他成分的支配。基于依存句法分析所构建的依存树，能够表达词语之间的依赖关系，并且强调动词在句子中的重要作用，这和语义角色标注任务在一定程度上是相通的。因此，本文在语义角色标注过程中，引入依存关系特征。

3.2 词的特征向量表示

为了减少特征选择的大量工作，本文只考虑当前词、当前词的词性、谓词、当前词到谓词的距离以及依存关系特征。另外，针对句中可能存在多个相同谓词的情况，设置是否为谓词 (0/1) 这一特征，即：若当前词为谓词，则为 1，否则为 0。对当前词、词性这两个特征设置特征窗口 [-1,1]，表示抽取当前词和左右各一个词及其词性作为特征。将上述特征以实数向量表示，拉直拼接后生成词的特征向量表示。

3.3 Gate 过滤机制

本文对当前词及词性特征设置了特征窗口，即利用固定上下文窗口内的词语表示当前词，相当于利用了 n-gram 结构。但是固定窗口内的词不一定每次都出现在一起，有的可能只出现少数次。因此，加入一个 Gate 过滤层可以对固定窗口内的信息进行调整，使得固定窗口内的词嵌入以一定概率出现，不再依赖于固定窗口内左右词的词向量的共同作用。

Gate 过滤机制由 Sigmoid 神经网络层和逐点乘积操作组成，如公式 (1) 定义所示， \odot 表示逐点乘积操作， x_t 是第 t 个词的原始特征向量表示，经过 Gate 过滤机制 g_t 的选择，得到输出 z_t 。

$$z_t = g_t \odot x_t \quad (1)$$

g_t 的定义如下：

$$g_t = \sigma(W_g x_t + b_g) \quad (2)$$

其中， σ 取 sigmoid 函数， W_g 表示当前层的权值矩阵， b_g 表示偏置向量。Sigmoid 层的输出在 0 到 1 之间，定义了信息通过的程度。

3.4 两层 Bi-LSTM 模型

长短时记忆神经网络 (Long Short-Term Memory, LSTM) 是一种特殊的循环网络结构，可以有效地解决 RNN 中存在的梯度消失和梯度爆炸问题。LSTM 专门设计了记忆单元 (memory cell) 用于保存历史信息。通过输入门 (input gate)、遗忘门 (forget gate) 以及输出门 (output gate) 实现对历史信息的选择性利用与更新，并且能够有效捕捉长距离依赖特征。而 Bi-LSTM (Bidirectional LSTM) 进一步结合历史信息与将来信息，分别从左右两个方向合并特征，以提升对文本全局语义的捕捉能力。

设 h 为 LSTM 单元的输出， C 为 LSTM 记忆单元的值， z 为输入数据。LSTM 单元内相关状态的更新以及输出如下：

$$\tilde{C}_t = \tanh(W_c z_t + U_c h_{t-1} + b_c) \quad (3)$$

$$i_t = \sigma(W_i z_t + U_i h_{t-1} + b_i) \quad (4)$$

$$f_t = \sigma(W_f z_t + U_f h_{t-1} + b_f) \quad (5)$$

$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t \quad (6)$$

$$o_t = \sigma(W_o z_t + U_o h_{t-1} + b_o) \quad (7)$$

$$h_t = o_t \odot \tanh(C_t) \quad (8)$$

其中, i_t 、 f_t 、 o_t 、 \tilde{C}_t 、 C_t 分别表示输入门、遗忘门、输出门、当前时刻记忆单元状态候选值以及状态值; W 、 U 分别表示输入数据和上一时刻 LSTM 单元输出的权值矩阵, b 表示偏置向量; σ 表示逻辑回归中的 sigmoid 函数。

在序列标注问题上, 将来信息与历史信息同样有助于序列的预测。因此, 本文采用 Bi-LSTM, 其基本思想是通过构造两个隐藏层分别从正反两个方向捕获上下文信息。Bi-LSTM 首先计算正向隐藏层序列 \vec{h}_t , 然后计算反向层序列 \overleftarrow{h}_t , 最后将正向序列 \vec{h}_t 与反向序列 \overleftarrow{h}_t 拼接, 得到输出 r_t :

$$r_t = [\vec{h}_t; \overleftarrow{h}_t] \quad (9)$$

为了获取更深层次的语义信息, 并且考虑到训练时间与标注效果之间的平衡, 本文使用两层 Bi-LSTM 框架, 将第一层 Bi-LSTM 的输出作为第二层 Bi-LSTM 神经网络层相应节点的输入。

3.5 标签转移概率矩阵的引入

在语义角色序列标注任务中, 相邻词的标签间实际存在着很强的依赖关系。例如, 在 IOB 序列表示法中, 标签 B_ARG0 之后只能是 I_ARG0、O 或者 B_X, 而其余的标签都是不合语法的; 而标签 I_ARG0 之前只能是 B_ARG0 或 I_ARG0。因此, 单独考虑每个词的标记得分将会不合适。为充分利用序列标签之间的依赖关系, 我们参照 Collobert 等^[1] 工作设计实现引入标签转移概率矩阵的方法, 在实验中使用转移概率矩阵 A 排除不可能的标签序列。

对于一个输入句子序列 $x=(x_1, x_2, \dots, x_n)$, 其标签序列为 $y=(y_1, y_2, \dots, y_n)$, P 为双向 LSTM 神经网络的计算得分矩阵, P_{ij} 表示句中第 i 个词标为第 j 个角色标签的得分, θ 是模型中需要学习的参数。令 A 表示标签之间的转移概率矩阵, 则元素 A_{ij} 表示标签 i 在下一时刻转移到标签 j 的概率, 对于不可能发生转移的元素赋值为 -10000, 其余转移则在模型训练中获得。另外, 用 $A_{y_0 y_1}$ 表示序列的第一个标签为 y_1 的概率。于是, 该标签序列的得分定义为:

$$s(x, y, \theta) = \sum_{i=1}^n (A_{y_{i-1} y_i} + P_{i, y_i}) \quad (10)$$

在图 1 的整体框架中, 在神经网络计算出输入序列中各个词的标签概率之后, 加上一层 CRF^[18], 对整个标签序列进行全局归一化处理, 找到概率最大的最优序列。本文使用 Viterbi 算法来推断最优标签序列。

4 模型训练

本文使用 Taskar 等^[19] 提出的最大间隔准则训练语义角色标注模型, 以提升似然估计的准确度。设输入句子为 x , 该句子正确的标签序列为 y , 用 $Y(x)$ 表示 x 所有可能标签序列的集合; 定义 $\tilde{y} \in Y(x)$ 表示 $Y(x)$ 中预测得分最高的一个标签序列, \tilde{y} 的计算公式为:

$$\tilde{y} = \arg \max_{y \in Y(x)} s(x, y, \theta) \quad (11)$$

结构损失函数 $\Delta(y, \tilde{y})$ 定义如下:

$$\Delta(y, \tilde{y}) = \sum_t \eta 1\{y_t \neq \tilde{y}_t\} \quad (12)$$

其中, n 为句子长度, η 为比例调节参数, $1\{y_t \neq \tilde{y}_t\}$ 表示当 $y_t \neq \tilde{y}_t$ 时为 1, 否则为 0。该损失函数表示对于输入句子 x , 标签预测错误数

的线性相关值。设训练集为 T ，定义正则化的目标函数 $J(\theta)$ 如下：

$$J(\theta) = \frac{1}{|T|} \sum_{(x,y) \in T} l(\theta) + \frac{\lambda}{2} \|\theta\|_2^2 \quad (13)$$

其中， $\frac{\lambda}{2} \|\theta\|_2^2$ 为 $L2$ 范数正则化， λ 为正则化系数。 $l(\theta)$ 的定义如下：

$$l(\theta) = \max(0, s(x, \tilde{y}, \theta) + \Delta(y, \tilde{y}) - s(x, y, \theta)) \quad (14)$$

最后，本文采用随机梯度下降法 (Stochastic Gradient Descent, SGD) 优化目标函数。

5 实验设计与结果分析

5.1 实验数据

本文采用 Chinese PropBank (CPB) 语义角色标注集，实验数据设置如下：训练集 648 个文件 (chtb_081.fid - chtb_899.fid)；开发集 40 个文件 (chtb_041.fid - chtb_080.fid)；测试集 72 个文件 (chtb_001.fid - chtb_040.fid, chtb_900.fid - chtb_931.fid)。

在语义角色标注性能评估中，采用评测常用的 P (准确率)、R (召回率) 和 F 值 (综合指标 F 值) 等评测指标。

5.2 依存关系的获取

本文采用两种方法获取依存关系特征：(1) 正确依存关系：使用 Penn2Malt 对 CTB 的短语结构句法树进行转换，得到高质量的依存句法分析树，从而获得正确的依存关系；(2) 自动依存分析结果：使用 the Stanford Parser^① 对 CTB 中的句子进行依存句法分析，从而获得自动分析的依存关系。

① <https://nlp.stanford.edu/software/lex-parser.shtml>

5.3 超参数设置

神经网络模型中超参数设置对模型性能会产生显著影响。参考前人工作，我们对超参数设定如下：

(1) 维度设置：设置词性以及谓词两个特征的向量维度为 150，词性、是否为谓词、当前词到谓词的距离、依存关系四个特征的向量维度分别设置为 50。

(2) 其他超参数设置：学习率 $a=0.01$ ，隐藏层节点个数 $H_{\text{dim}}=300$ 。关于 Gate 过滤机制的输出特征的向量表示，设置 Dropout 的丢弃率为 0.2；设置 Bi-LSTM 层的 Dropout 丢弃率为 0.5。最大间隔准则中的比例调节参数 $\eta=0.2$ ，正则化系数 $\lambda=0.0002$ 。

本文使用 Glorot 等^[20]提出的方法对词的特征向量表示进行初始化，然后在模型训练过程中学习更新。

5.4 实验结果分析

本文提出基于 Bi-LSTM 深度神经网络学习模型进行汉语语义角色标注的方法，我们首先对方法中的 Gate 过滤机制的作用进行评测实验，实验结果如表 1 所示。从表中结果看出，加入 Gate 过滤机制后，F 值从 77.74% 提高到 79.40%，提升了 1.66%，模型的性能得到显著提升，证明了 Gate 过滤机制的有效性。同时，显示了 Gate 过滤机制与 LSTM 中的三种门 (input, forget, output) 具有不同的作用。我们对此分析如下，Gate 过滤机制主要用于对词向量表示进行调整，而 LSTM 则是通过门使上下文信息选择性地通过，解决信息保存问题以

获取长距离的上下文信息。Gate 过滤机制与 LSTM 中的门控单元在功能上互补,从而使得语义角色标注的性能得到显著提升。

表1 加入Gate过滤机制前后的性能对比

模型	P (%)	R (%)	F (%)
未加Gate	79.58	75.99	77.74
加入Gate	82.44	76.57	79.40

句子:	...	其中	,	出口额	最	大	的	商品	是	服装	...
正确标签:	...	O	O	B_ARG0	B_ARGM-ADV	REL	O	B_ARG0	O	O	...
不加Gate:	...	O	O	B_ARGM-LOC	B_ARGM-ADV	REL	O	B_ARG0	O	O	...
加入Gate:	...	O	O	B_ARG0	B_ARGM-ADV	REL	O	B_ARG0	O	O	...

图2 加入Gate过滤机制前后的标注对比结果

在语义角色标注模型中,我们引入了依存关系特征。为了验证依存关系特征对模型性能的影响,我们分别使用正确依存关系和自动依存分析结果进行实验,评测结果如表2所示。当采用正确依存关系特征时,语义角色标注的F值达到79.67%;而采用自动依存分析结果的依存关系时,F值有所下降,但依然达到79.53%,比未利用依存关系的模型高出0.13%,证明本文利用依存关系特征方法的有效性。

表2 利用依存关系特征方法的性能比较

模型	P (%)	R (%)	F (%)
未加依存关系	82.44	76.57	79.40
正确依存关系	80.70	78.84	79.67 (+0.27)
自动依存分析结果	82.62	76.66	79.53 (+0.13)

此外,我们将本文所得到的语义角色标注模型的最优性能与前人工作进行了比较,对比结果如表3所示。从表中我们可以看出,采用

图2为加入Gate过滤机制前后实验中的一组标注结果,关注句子“... 其中, 出口额最大的商品是服装...”中的谓词“大”的相关论元“出口额”的标注结果。未加Gate过滤机制时,“出口额”受固定窗口中左边词“其中,”的影响,得到错误的标签“B_ARGM-LOC”;加入Gate过滤机制后,模型对词向量表示进行调整,“出口额”标注为正确标签“B_ARG0”。

Gate过滤机制与两层Bi-LSTM-CRF相结合的模型,并且加入依存关系特征后,模型的F值比前人最好的结果提高了1.84%,证明了本文方法的有效性。

表3 CPB标注集上语义角色标注方法的结果对比

模型	P (%)	R (%)	F (%)
Yang等 ^[5]	--	--	75.31
Wang等 ^[9]	--	--	77.09
Sha等 ^[10]	--	--	77.69
本文	82.62	76.66	79.53 (+1.84)

6 结语

针对传统的基于特征的统计机器学习方法的局限性,本文给出了一种融合Gate过滤机制与深度Bi-LSTM-CRF的汉语语义角色标注模型,并且提出了引入依存关系特征的方法。我们在CPB上的实验结果表明:Gate过滤机制的

加入使 F 值提高了 1.66%，显著提升了语义角色标注的性能；依存关系特征对汉语语义角色标注的性能也有一定程度的提高；汉语语义角色标注模型最终所得到的 F 值达到 79.53%，超过已有研究，达到最好水平。今后，将扩大实验规模，进一步验证方法的有效性，并探索利用句法结构的新方法。

参考文献

- [1] Collobert R, Weston J, Karlen M, et al. Natural Language Processing (Almost) from Scratch[J]. Journal of Machine Learning Research, 2011, 12(1):2493-2537.
- [2] Hacioglu K. Semantic Role Labeling using Dependency Trees[C]. International Conference on Computational Linguistics. Association for Computational Linguistics, 2004:1273.
- [3] Hafdel L, Nugues P. Multilingual Semantic Role Labeling[C]. Thirteenth Conference on Computational Natural Language Learning: Shared Task. Association for Computational Linguistics, 2009:43-48.
- [4] Jin G, Kawahara D, Kurohashi S. Chinese Semantic Role Labeling using High-quality Syntactic Knowledge[C]. Eighth Sighan Workshop on Chinese Language Processing. 2015:120-127.
- [5] 汪红林, 王红玲, 周国栋. 基于依存关系的语义角色标注[J]. 计算机工程, 2009, 35(15):82-84.
- [6] 王步康, 王红玲, 袁晓虹, 等. 基于依存句法分析的中文语义角色标注[J]. 中文信息学报, 2010, 24(1):25-29.
- [7] 王鑫, 穗志方. 基于依存树距离识别论元的语义角色标注系统[J]. 中文信息学报, 2012, 26(2):40-45.
- [8] Yang H, Zong C. Multi-Predicate Semantic Role Labeling[C]. Conference on Empirical Methods in Natural Language Processing. 2014:363-373.
- [9] Zhou J, Xu W. End-to-end learning of Semantic Role Labeling using Recurrent Neural Networks[C]. Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics. 2015:1127-1137.
- [10] Roth M, Lapata M. Neural Semantic Role Labeling with Dependency Path Embeddings[C]. Meeting of the Association for Computational Linguistics. 2016:1192-1202.
- [11] 王臻, 常宝宝, 穗志方. 基于分层输出神经网络的汉语语义角色标注[J]. 中文信息学报, 2014, 28(6): 56-61.
- [12] Wang Z, Jiang T, Chang B, et al. Chinese Semantic Role Labeling with Bidirectional Recurrent Neural Networks[C]. Conference on Empirical Methods in Natural Language Processing. 2015:1626-1631.
- [13] Sha L, Li S, Chang B, et al. Capturing Argument Relationship for Chinese Semantic Role Labeling[C]. Conference on Empirical Methods in Natural Language Processing. 2016:2011-2016.
- [14] Jiang Guo, Wanxiang Che, Haifeng Wang, Ting Liu and Jun Xu. A Unified Architecture for Semantic Role Labeling and Relation Classification[C]. Proceedings of COLING 2016, 2016: 1264-1274.
- [15] 王瑞波, 李济洪, 李国臣, 等. 基于Dropout正则化的汉语框架语义角色识别[J]. 中文信息学报, 2017, 31(1):147-154.
- [16] Hochreiter S, Schmidhuber J. Long Short-Term Memory[J]. Neural Computation, 2013, 9(8):1735-1780.
- [17] Zhaopeng Tu, Yang Liu, Zhengdong Lu, Xiaohua Liu, Hang Li. Context Gates for Neural Machine Translation[C]. Proceedings of 2017 Association for Computational Linguistics, 2017:87-99.
- [18] Lafferty J D, McCallum A, Pereira F C N. Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data[C]. Eighteenth International Conference on Machine Learning. Morgan Kaufmann Publishers Inc. 2001:282-289.
- [19] Taskar B, Chatalbashev V, Koller D, et al. Learning Structured Prediction Models: A Large Margin Approach[C]. International Conference on Machine Learning. ACM, 2005:896-903.
- [20] Glorot X, Bengio Y. Understanding the Difficulty of Training Deep Feedforward Neural Networks[J]. Journal of Machine Learning Research, 2010, 9:249-256.