

# 基于 Highway-BiLSTM 网络的汉语谓语中心词识别研究

黄瑞章<sup>1,2</sup>, 靳文繁<sup>1</sup>, 陈艳平<sup>1,2</sup>, 秦永彬<sup>1,2</sup>, 郑庆华<sup>3</sup>

(1. 贵州大学计算机科学与技术学院, 贵州 贵阳 550025; 2. 贵州省公共大数据重点实验室, 贵州 贵阳 550025;  
3. 西安交通大学计算机科学与技术学院, 陕西 西安 710049)

**摘 要:** 针对汉语谓语中心词识别困难及唯一性的问题, 提出了一种基于 Highway-BiLSTM 网络的深度学习模型。首先, 通过多层 BiLSTM 网络叠加获取句子内部不同粒度抽象语义信息的直接依赖关系; 然后, 利用 Highway 网络缓解深层模型出现的梯度消失问题; 最后, 通过约束层对输出路径进行规划, 解决谓语中心词的唯一性问题。实验结果表明, 该方法有效提升了谓语中心词识别的性能。

**关键词:** 谓语中心词; 高速公路连接; 双向长短期记忆网络; 唯一性

**中图分类号:** TN391

**文献标识码:** A

**DOI:** 10.11959/j.issn.1000-436x.2021027

## Research on Chinese predicate head recognition based on Highway-BiLSTM network

HUANG Ruizhang<sup>1,2</sup>, JIN Wenfan<sup>1</sup>, CHEN Yanping<sup>1,2</sup>, QIN Yongbin<sup>1,2</sup>, ZHENG Qinghua<sup>3</sup>

1. College of Computer Science and Technology, Guizhou University, Guiyang 550025, China

2. Guizhou Provincial Key Laboratory of Public Big Data, Guiyang 550025, China

3. College of Computer Science and Technology, Xi'an Jiaotong University, Xi'an 710049, China

**Abstract:** Aiming at the problem of difficult recognition and uniqueness of Chinese predicate head, a Highway-BiLSTM model was proposed. Firstly, multi-layer BiLSTM networks were used to capture multi-granular semantic dependence in a sentence. Then, a Highway network was adopted to alleviate the problem of gradient disappearance. Finally, the output path was optimized by a constraint layer which was designed to guarantee the uniqueness of predicate head. The experimental results show that the proposed method effectively improves the performance of predicate head recognition.

**Keywords:** predicate head, highway connection, BiLSTM, uniqueness

## 1 引言

谓语中心词是句子的焦点, 是支配和关联句子其他语法成分的关键语法单元。在以谓语为中心的句法成分分析中, 需要根据谓语中心词来解析句子结构。识别谓语中心词是正确理解句子的前提。正确识别谓语中心词可以解析句子结构、获取句子的语义信息, 进一步支撑浅层句法分析的研究, 从而支撑机器翻译、信息检索、情感分析等自然语言处

理应用。

一个动词在句子中是否是谓语中心词既与它本身的语法属性有关, 也与它的上下文环境有关。动词本身的语法属性称为静态特征, 与识别谓语中心词有关的上下文环境称为动态特征。由于汉语句子结构松散, 传统观点认为汉语句子没有形式上的谓语中心词。

此外, 谓语中心词的识别还需要克服以下几个问题。1) 汉语是一种古老的象形文字, 缺少分词信

收稿日期: 2020-05-13; 修回日期: 2020-08-10

基金项目: 国家自然科学基金资助项目(No.U1836205, No.91746116); 贵州省科学技术基金重点资助项目(黔科合基础[2020]1Z055)

**Foundation Items:** The National Natural Science Foundation of China (No.U1836205, No.91746116), The Key Project of Science and Technology of Guizhou (No.[2020]1Z055)

息。比如,《现代汉语规范词典》没有收录“撞向”为动词,但收录了“通向”“流向”等词为动词。类似的情况有“下雨”被收录为词,而“下雪”没有。无法正确分词给谓语中心词的识别带来困难。2) 汉语句子结构松散。汉语句子通常包含几个动词,它们中的每一个都可以作为谓语中心词或状语短语来处理,很难识别句子中单词之间的依赖关系。3) 汉语单词中的兼义现象非常严重,存在很多名词、形容词动词化的用法,但没有形态特征来表示它们的动词用法,使区分它们之间的句法作用变得困难。4) 谓语中心词是句子的中心,识别谓语中心词需要对句子的高阶依赖关系进行建模。当前的序列模型难以捕获句子中的高阶依赖关系。在汉语谓语中心词识别方面,现有的序列标注模型还存在不足之处。例如长短时记忆(LSTM, long-short term memory)模型理论上能记忆长距离信息,但是在实际使用中,LSTM对长实体的识别性能较低。

本文主要的研究工作如下。

1) 针对汉语谓语中心词的特点,提出了一种基于深层双向长短时记忆(BiLSTM, bi-directional LSTM)的汉语谓语中心词识别模型。该模型利用4层BiLSTM结构获取句子的抽象语义特征和上下文语义依赖关系。与传统序列标注模型相比,深层BiLSTM模型能更好地获取句子内部不同粒度抽象语义信息,在汉语谓语中心词数据集上有更好的表现。

2) 利用 Highway 连接缓解深层模型的梯度消失的问题。随着深度网络层数的不断增加,输入信息在通过网络层到达网络的末端时,可能出现梯度消失的情况。本文通过 Highway 网络的引入有效地缓解了训练深层模型时梯度消失的情况。

3) 汉语谓语中心词的唯一性问题。单个句子中通常只有一个谓语中心词,但可以有多个动词。为了解决这个问题,本文在模型的输出中加入约束层,通过约束函数对输出路径进行约束,确保谓语中心词的输出唯一性。

## 2 相关工作

谓语中心词在句中起到组织句法或者语义信息的中心作用,如主语、时间、原因和形式等。识别谓语中心词是理解句子的关键。然而,在汉语自然语言处理领域,关于谓语中心词识别的研究工作却很少。现有工作主要采用基于规则的方法、基于统计学习的方法和规则与统计相结合的方法。

在基于规则的方法中,Luo等<sup>[1]</sup>从各种词性的词作为谓语时的语法特点出发,讨论谓语的识别策略,通过规则的方法来判别和确定句子的中心谓语及其相应边界。Li等<sup>[2]</sup>提出了一种利用句子的主语和谓语之间的句法关系来识别谓语中心词的方法。该方法除了利用谓语中心词候选项的静态语法特征和动态语法特征外,还加入了对规则间相互影响的考虑,较之前的方法从更高的句法层次上进行了分析。但是特征的应用过程相对复杂,计算量较大,对于一些特殊的句型可能产生错误的结果。Sui等<sup>[3-4]</sup>出了一种折中的汉语句子分析方法——骨架依存分析法,利用句子级对齐的双语语料库中英汉语谓语中心词的对应来寻找汉语句子的谓语中心词。但是其仅对例句集中的汉语单句识别了谓语中心词,没有从大规模已标注的谓语中心词的汉语例句中实现谓语中心词自动抽取。

在基于统计学习的方法中,陈小荷等<sup>[5]</sup>采用统计的方法对50万字的语料库识别了核心谓语。Wang等<sup>[6]</sup>组合谓语动词的多个特征,并使用最大熵分类器对谓语中心词进行自动识别。谌志群<sup>[7]</sup>提出了一种基于统计学原理的汉语句子谓语自动识别概率模型,通过对语料库中句子的谓语所处上下文环境的细致分析,选择影响谓语出现的语境特征,在此基础上通过构建统计模型来计算谓语出现的概率,识别汉语句子的谓语。

在规则与统计学习相结合的方法中,Gong等<sup>[8]</sup>将整个谓语识别的过程分为语片捆绑、谓语粗筛选和谓语精筛选3个阶段。首先,在识别之前加入语片捆绑的预处理工作,有效排除了一些准谓语;然后,用规则的方法进行粗筛选,降低了精筛选的复杂度;最后,利用特征学习的方法进行谓语精筛选,有效解决了规则的不完备和特征重要度排序的问题。但是,该方法还存在一些谓语误识的情况,对于复杂结构的汉语句子不能完全正确识别。另外,Han等<sup>[9]</sup>提出一种融合词法与句法特征、结合C4.5机器学习和规则进行谓语识别的方法。该方法表明句法特征能有效提升谓语识别效果。李琳等<sup>[10]</sup>利用大规模的藏语语料库训练得到藏语词向量,其结果表明词向量特征可显著提高藏语谓语动词短语的识别效果。目前,谓语中心词识别的研究大部分还是使用传统方法,难以建模高阶依赖信息。

在汉语信息抽取领域,与谓语中心词识别相关的任务还有命名实体识别和语义角色标注。其中,

命名实体识别通常采用序列标注模型进行识别，如隐马尔可夫模型 (HMM, hidden Markov model)<sup>[11]</sup>、条件随机场 (CRF, conditional random field)<sup>[12]</sup>和 LSTM<sup>[13]</sup>。近年来，基于深度学习模型的实体识别得到了广泛研究。比如，Li 等<sup>[14]</sup>采用双向 LSTM-CRF 结构，在生物医学实体识别上取得了良好的效果。与命名实体识别相比，谓语中心词识别更强调谓语中心词作为句子中心的语法功能，在识别上需要依赖句子的整体结构和语义特征，在输出路径中需要保证标注实体的单一性。

语义角色标注 (SRL, semantic role labeling)<sup>[15]</sup>是浅层语义分析中的一种主要实现方式。该方法主要是对给定句子中存在的每个谓语进行分析，并标注其相应的语义成分。传统的 SRL 方法采用基于句法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 2 种语义角色标注方法。比如，Koomen 等<sup>[16]</sup>和 Tackstrom 等<sup>[17]</sup>采用线性规划或动态规划的方式获得句子的全局约束。随着深度学习的兴起，研究者将基于 BiLSTM 模型用于语义角色标注任务。比如，Zhou 等<sup>[18]</sup>使用深度 BiLSTM 模型对英文语义角色标注进行了研究。Guo 等<sup>[19]</sup>重点关注句法路径信息并使用 BiLSTM 对其进行建模，从而提高了 SRL 系统的性能。王瑞波等<sup>[20]</sup>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。Strubell 等<sup>[21]</sup>提出了一种基于语言信息的自我注意神经网络模型，它将多头自我注意与多任务学习相结合，包括依赖分析、词性标注、谓语检测和 SRL。

### 3 模型构建

本文的 Highway-BiLSTM 网络结构如图 1 所示。自底向上描述如下。1)使用预训练的维基百科字向量将输入的文本序列映射为向量，作为当前词的特征向量表示；2)经过 4 层 BiLSTM<sup>[22]</sup>获取句子内部不同粒度抽象语义信息的直接依赖关系；3)为了防止训练深层 BiLSTM 模型时出现梯度消失的问题，层与层之间使用 Highway 网络连接；4)通过一个 Softmax 层进行归一化处理；5)通过约束层保证谓语中心词的唯一性，得到最优标注序列。句子中包含的谓语中心词用标签  $\{B, I\}$  表示，其中谓语中心词的开始用标签  $B$  标记，其余部分用标签  $I$  标记。在模型的输出部分，使用  $P$  表示预测的句子，模型

预测出的谓语中心词使用标签  $B-I$  标记，句子中的其他成分使用标签  $O$  标记。

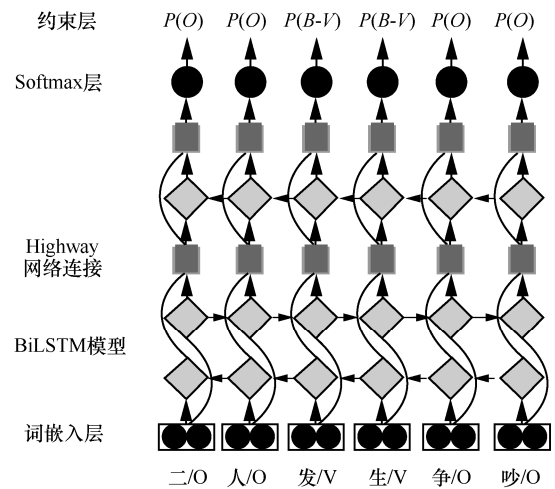


图 1 Highway-BiLSTM 网络结构

### 3.1 深度双向长短时记忆模型

一方面，传统的循环神经网络 (RNN, recurrent neural network) 在处理长序列数据时容易出现梯度消失或梯度爆炸问题。另一方面，对于基本的 RNN 来说，它能够处理一定的短期依赖，但是无法处理长期依赖问题。而 LSTM 通过引入门结构解决了 RNN 的长期依赖问题。由于汉语谓语中心词缺少形态特征、句子结构松散且形式多样、单个句子可能存在多个动词等情况，传统的序列标注模型无法很好地建模句子的高阶依赖特征，获取句子的全局信息。针对上述问题，本文提出利用多层堆叠的 BiLSTM 构建谓语中心词识别模型。LSTM 通过 3 个不同的门来调节单元状态中的信息流，即输入门、遗忘门、输出门控制着信息流的更新与利用。本文设输入的信息流为  $x_{l,t}$ ，表示第  $l$  层和  $t$  时刻到 LSTM 的输入；细胞状态为  $c_{l,t}$ ；LSTM 在  $t$  时刻的输出信息为  $h_{l,t}$ ；激活函数为  $\sigma$ ；权重矩阵为  $W$ ；偏置向量为  $b$ 。

首先，决定从细胞状态中丢弃什么信息。这个决策是通过一个称为“遗忘门”的层来完成的。该门会读取  $h_{l,t}$  和  $x_{l,t}$ ，使用 sigmoid 函数输出一个 0~1 的数值，输出状态  $c_{l,t}$  中每个细胞的数值，1 表示完全保留，0 表示完全舍弃。

$$i_{l,t} = \sigma(W_i^l [h_{l,t} + \delta_l, x_{l,t}] + b_i^l) \quad (1)$$

$$O_{l,t} = \sigma(W_o^l [h_{l,t} + \delta_l, x_{l,t}] + b_o^l) \quad (2)$$

然后，确定什么样的新信息被存放在细胞状态中。信息包含两部分：一部分是 sigmoid 函数，称

为“输入门”，决定更新什么值；另一部分是  $\tanh$  函数，用于创建一个新的候选值向量  $\bar{C}_{l,t}$ ，该向量会被加入状态中。这样就能用这 2 个信息产生对状态的更新。

$$f_{l,t} = \sigma(W_f^l[h_{l,t} + \delta_l, x_{l,t}] + b_f^l + 1) \quad (3)$$

$$\bar{C}_{l,t} = \tanh(W_c^l[h_{l,t} + \delta_l, x_{l,t}] + b_c^l) \quad (4)$$

更新旧细胞状态过程如下。 $\bar{C}_{l,t-1}$  更新为  $\bar{C}_{l,t}$ 。把旧状态与  $f_{l,t}$  相乘，丢弃确定需要丢弃的信息。然后加上  $i_{l,t} \circ \bar{C}_{l,t}$ ，得到新的候选值，根据每个状态的更新程度进行变化，其中  $\circ$  表示点乘。

$$C_{l,t} = i_{l,t} \circ \bar{C}_{l,t} + f_{l,t} \circ C_{l,t-1} \quad (5)$$

最后，运行一个 sigmoid 函数来确定输出细胞状态的哪个部分，通过  $\tanh$  处理细胞状态，并与 sigmoid 门的输出相乘，仅输出确定输出的那部分。

$$h_{l,t} = O_{l,t} \circ \tanh(C_{l,t}) \quad (6)$$

其中， $\delta_l$  为 1 或 -1，表示 LSTM 在第  $l$  层的方向性。在实验中根据 Zhou 等<sup>[23]</sup>的方法以交织模式堆叠 LSTM，设每层的特定输入  $x_{l,t}$  和方向性  $\delta_l$  分别为

$$x_{l,t} = \begin{cases} [W_{\text{emb}}(w_t), W_{\text{mask}}(t = v)], l = 1 \\ h_{l-1,t}, l > 1 \end{cases} \quad (7)$$

$$\delta_l = \begin{cases} 1, l \text{ 为偶数} \\ -1, \text{其他} \end{cases} \quad (8)$$

### 3.2 Highway 连接

本文在实验中使用了 4 层的 BiLSTM（即 8 层 LSTM）堆叠模型取得了较好的效果，其中一个关键要素是使用封闭的“Highway 连接”<sup>[23-24]</sup>。相比于 RNN，LSTM 能够在一定程度上缓解梯度消失的问题。理论和经验表明，神经网络的深度是其成功的关键。然而，随着网络层数的不断增加，训练变得更加困难，在网络传输中会出现一个问题：当输入的信息通过许多层，到达网络的末端（或起点）时，信息可能会“消失”。这里的“消失”实际是由于链式法则下多个小于 1 的数值相乘导致的。针对这个问题，本文使用 Highway 连接各层 BiLSTM，使特征信息能够在多个层面上传输。Highway 网络受到 LSTM 网络的启发，同样使用自适应门控单元来调节信息流。典型的神经网络是一个仿射变换加一个非线性函数，即  $y = H(x, W_H)$ 。在深层模型的层间连接中，为每一个层的输出添加转换门和进位门，形成 Highway 网络。其定义为

$$y = H(x, W_H) \circ T(x, W_T) + x \circ C(x, W_C) \quad (9)$$

Highway 网络的本质是通过 shortcut 机制实现深层网络的训练。如图 2 所示，shortcut 机制选择合适的层进行转换，通过该机制可以使信息在许多层之间流动而不会衰减，即使在深层模型中，使用 Highway 连接也可以通过简单的梯度下降直接进行训练。本文实验通过转换门  $r_l$  来控制层与层之间线性和非线性变换的权重。

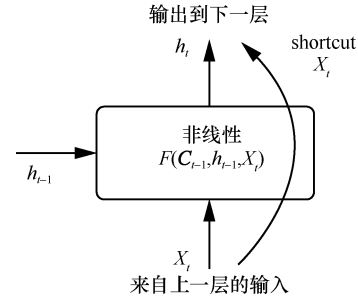


图2 Highway 连接

最后，输出  $h_{l,t}$  改为

$$r_{l,t} = \sigma(W_r^l[h_{l,t-1}, x_t] + b_r^l) \quad (10)$$

$$h'_{l,t} = O_{l,t} \circ \tanh(C_{l,t}) \quad (11)$$

$$h_{l,t} = r_{l,t} \circ h'_{l,t} + (1 - r_{l,t}) \circ W_h^l x_{l,t} \quad (12)$$

### 3.3 Dropout 机制

在训练样本一定的情况下，模型的复杂性越高，就会有越多的参数，训练出的模型越容易产生过拟合的现象。为了避免过拟合，在本文的模型中使用 Gal 等<sup>[25]</sup>所描述的 Dropout 机制，通过 Dropout 提高了模型的泛化能力。定义为

$$\tilde{h}_{l,t} = r_{l,t} \circ h'_{l,t} + (1 - r_{l,t}) \circ W_h^l x_{l,t} \quad (13)$$

$$h_{l,t} = z_l \circ \tilde{h}_{l,t} \quad (14)$$

其中， $z_l$  在  $l$  层上跨时间步共享，以免沿序列放大噪声。

### 3.4 谓语中心词的唯一性

通常，识别任务是给定一个句子作为输入预测一个序列  $y$ ，每个  $y_i \in y$  都属于一个离散的标签集合  $T$ 。句子中包含的谓语中心词用标签  $\{B, I\}$  表示，其中谓语中心词的开始用标签  $B$  标记，其余部分用标签  $I$  标记。为了避免在单个句子中谓语中心词数量大于一个的现象，本文在模型全连接层之后使用 Softmax 层进行归一化处理，并通过约束层对谓语中心词的输出路径进行约束。对于每一个句子，预

测它的谓语中心词结构，在所有可能的解空间  $Y$  中找到得分最高的标签序列。其约束函数定义为

$$\hat{y} = \arg \max f(y) \quad (15)$$

其中， $y \in Y$ ，得分函数  $f(y)$  的输入条件为

$\sum_{i=1}^n \log p(y_i | w)$ 。为了加入额外的信息，比如，结构一致性、语法输入等，在实验中利用惩罚项来增加得分函数，即

$$f(y) = \sum_{i=1}^n \log p(y_i | w) - \sum_{c \in C} c(w, y_{i,t}) \quad (16)$$

其中，给定输入  $w$  和长度  $t$  的前缀  $y_{i,t}$ ，每个函数  $c$  都应用非负惩罚。

## 4 实验

### 4.1 数据集

本文的实验使用汉语谓语中心词数据集。该数据来源于“中国裁判文书网”中的 762 篇法院刑事判决书。与李婷等<sup>[26]</sup>的标注规范相同，在标注规范中，把谓语中心词分为以下几种模式。

#### 模式 1 单个谓语中心词

由于汉语单词之间没有分隔符，在词的划分上存在歧义，因此本文的单个谓语中心词，以词典的收录为准。例如，“取得”在字典中被收录为词，“取出”却没有。那么只有“取得”属于模式 1，标注为单个谓语中心词。

#### 模式 2 复合结构的谓语中心词

汉语句子结构中经常使用重复的表达式来构成复合结构的词，如“跑一跑”“洗洗手”等。

#### 模式 3 同义并列的谓语中心词

同义动词通常同时使用来作为谓语中心词，如“驱车/行驶”“开发/建设”“抓捕/归案”等。此外，连续的动词表达相反的语义，但属于偏正关系的，如“进进出出”也被标记为一个谓语中心词。

#### 模式 4 带修饰或带补语的谓语中心词

当句子中的动词带有时态标记、补语或修饰符时，将谓语中心词标注在括号中，如“王某取出一把尖刀”，标记为“王某[(取)出]一把尖刀”。

#### 模式 5 其他特殊表达的谓语中心词

当句子中存在名词做动词、形容词做动词，以及谚语和成语或典故等时，如“张某[心生不满]”，成语“心生不满”如果切分，则会引起歧义，所以单独标为谓语中心词。该模式还可以用于处理主语—谓语从句，其形式为名词（代词）+动词（形容词）。

例如，“我[开心]”等。

通过上述 5 种标注模式，本文共标注 7 022 条句子，标记 7 022 个谓语中心词，其中，模式 1 有 4 959 个，模式 2 有 24 个，模式 3 有 272 个，模式 4 有 1 651 个，模式 5 有 116 个。

### 4.2 实验参数设置

本文的模型由 8 个 LSTM 层（4 个正向 LSTM 层和 4 个反向 LSTM 层）和一个用于预测输出分布的 Softmax 层组成，层与层之间使用 Highway 连接。根据文献[27]，本文模型中所有权重矩阵都用随机标准正交矩阵初始化。模型参数设置如表 1 所示。

表 1 参数设置

参数	设定值
词向量维度/维	100
上下文标记维度/维	128
LSTM 隐含层维度/维	100
Batchsize	20
LSTM 层数/层	8
学习率	0.001
Dropoutrate	0.2

在实验训练阶段使用了预训练的汉语维基百科词向量嵌入初始化，每个字经过预训练后的维度都是 100 维，并且在训练期间进行更新。维基百科未覆盖的字将替换为随机初始化的<UNK>嵌入。实验设置每次输入模型的最大句长为 128，即上下文标记维度。LSTM 隐含层维度设置为 100，表示用于记忆和存储过去状态的节点个数。Batchsize 大小为 20，其含义是每个批次将 20 个句子输入到模型中进行训练，完成一次参数计算和更新。在训练模型时，如果学习率设置过小，模型的收敛将变得十分缓慢；如果学习率设置过大，模型可能在最小值附近振荡，甚至无法收敛。经过实验对比，学习率设置为 0.001 时，模型可以快速收敛。为了防止模型过拟合，本文使用了 Dropout 机制，设置 Dropoutrate 取值区间为[0, 1]。经过实验对比，当 Dropoutrate=0.2 时，实验结果最优。在训练阶段使用的优化器是 AdaDelta<sup>[28]</sup>，其中  $\rho = 0.95$ ， $\epsilon = 1 \times 10^{-6}$ 。所有的模型都经过 100 次训练，并根据验证集结果停止。实验结果如表 2 所示。

从表 2 的实验结果可以看出，Highway+ BiLSTM+ Softmax 模型取得了最好的效果。第一组实验使用的是传统的序列标注模型 CRF。CRF 通过

特征模板扫描整个句子，它更多考虑的是整个句子局部特征的线性加权组合。CRF 计算的是一种联合概率，优化的是整个序列，而不是将每个时刻的最优结果拼接起来。第二组实验加入了 BiLSTM，BiLSTM 的优势是可以同时捕捉正反 2 个方向的长距离信息，建模上下文的依赖关系。通过实验对比可以看到，BiLSTM 在 CRF 的基础上有了较大的提升。但是 BiLSTM 只能学习到某个特定维度的特征。第三组实验是李婷等<sup>[26]</sup>最近的工作，在第二组实验的基础上加入了 Attention 机制，Attention 机制可以获取全局与局部的联系，不会像 RNN 模型那样对长期依赖的捕捉受到序列长度的影响。在谓词中心词约束部分，文献[26]使用卷积神经网络对序列标注结果进行二分类，得到最终的识别结果。第三组实验使用了多层 BiLSTM 叠加获取句子的全局信息，其中层与层之间使用残差模块连接<sup>[29]</sup>，性能较第二组实验有近 3% 的提升。第四组实验中在使用多层 BiLSTM 模型堆叠的同时，利用 Highway 连接来缓解梯度消失的问题。通过实验结果可以看出，Highway 连接比残差模块有 1% 的性能提升。与文献[26]所做的工作相比，本文的模型是端到端的，不需要分步实验就能得到最终的序列标注结果。

表 2 实验结果

模型	准确率	召回率	F 值
CRF	75.26%	58.12%	65.59%
BiLSTM+CRF	76.05%	71.45%	73.68%
BiLSTM+Attention+CRF	76.60%	74.48%	76.58%
Residual + BiLSTM + Softmax	80.135%	78.517%	79.318%
Highway + BiLSTM + Softmax	<b>80.757%</b>	<b>80.095%</b>	<b>80.424%</b>

### 4.3 实验结果分析

#### 4.3.1 句子长度对预测结果的影响

本文在实验中比较了不同的句子长度对实验结果的影响，如表 3 所示。随着句子长度的增加，准确率、召回率、F 值均呈下降趋势。这也充分说明了模型在长距离语义依赖中的表现略差，长距离语义建模成为限制模型性能的一大因素。

表 3 不同句长下的实验结果

句长/字	句子数/条	准确率	召回率	F 值
1~6	1119	78.632%	84.404%	81.416%
7~14	3395	78.482%	75.573%	77.000%
15~25	2748	74.396%	71.628%	72.986%
≥26	2923	69.654%	73.704%	71.622%

#### 4.3.2 词向量分析

本文使用汉语维基百科语料为预训练词向量，得到约 38 万个字的字向量，可以覆盖 99% 的训练集和验证集。在查找表中不能找到的词被映射为 <UNK> 字符，采用  $[-0.05, 0.05]$  的均匀分布随机化。与随机初始化相比，使用大规模语料预训练词向量后 F 值从 79.436% 提升至 80.424%，如表 4 所示。

表 4 不同词向量的实验结果

实验方法	准确率	召回率	F 值
预训练词向量	80.757%	80.095%	80.424%
随机初始化	79.790%	79.085%	79.436%

如图 3 所示，采用不同的词向量后 F 值随着迭代次数的增加而上升，在每一轮迭代中采用预训练的词向量的模型效果都优于随机初始化的模型。预训练的词向量能够更好地表达每个字的特征表示。

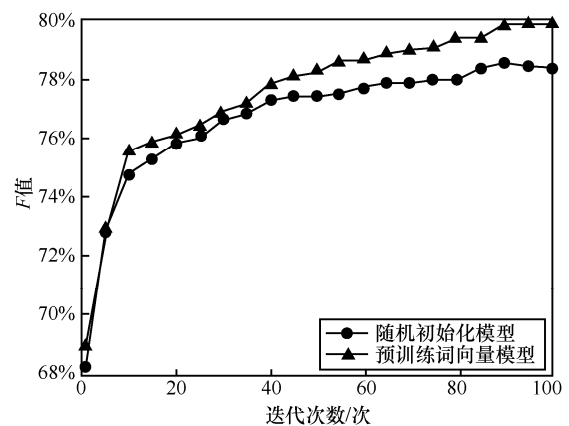


图 3 不同词向量对 F 值的影响

#### 4.3.3 层间连接分析

本节实验选取 Wiki-100 initialized 词向量的模型进行实验，对比了无连接、残差连接与 Highway 连接的表现，如表 5 所示。

表 5 有无 Highway 连接的实验结果

实验方法	准确率	召回率	F 值
无连接	77.819%	78.580%	78.198%
残差连接	80.135%	78.517%	79.318%
Highway 连接	80.757%	80.095%	80.424%

Highway 连接与残差连接具有一个相同的特性，即过 shortcut 机制来实现深度网络的训练。从表 5 可以看出，Highway 连接有更好的表现，其原因是 Highway 连接比残差连接对跳转连接有更多的

控制, Highway 连接启发于 LSTM 的门控机制, 它的转换门和进位门门控结构允许信息在跳转层和使用之间有一个学习的平衡。通过实验结果可以看出, Highway 连接比残差连接有 1% 的性能提升。

通过 Highway 连接可以缓解训练深层模型时梯度消失的情况, 同时说明 Highway 连接在本文模型中是有效的。为了进一步说明 Highway 连接对深层模型的重要性, 本文在实验中改变了层与层之间的连接方式, 使用残差连接代替 Highway 连接。如图 4 所示, 当迭代次数不断增加时, 使用残差连接的模型准确率会先上升然后达到饱和, 迭代次数继续增加准确率会下降, 而使用 Highway 连接的模型准确率随着迭代次数的增加而不断上升, 最终在 100 次迭代后达到饱和。

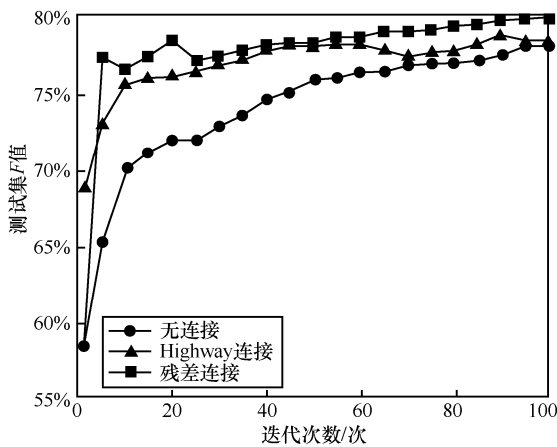


图 4 不同层连接方式对 F 值的影响

#### 4.3.4 比较层数对模型性能的影响

本节选取了 Wiki-100 词向量的模型进行实验, 探究模型层数对模型性能的影响。在使用 4 层的 BiLSTM 叠加后实验效果达到最佳。如表 6 所示, 4 层模型比 2 层模型提升了近 1.3% 的 F 值。从实验结果可以看出, 深层神经网络的特征提取能力更强, 在序列标注任务上优于浅层神经网络。

表 6 不同层数在验证集上的实验结果

层数/层	准确率	召回率	F 值
4	80.757%	80.095%	80.424%
2	80.743%	77.508%	79.092%

如图 5 所示, 不同层数的模型训练时的损失值下降速率也不同。从图 5 可以看出, 设置模型层数为 4 层时收敛速度达到最快。

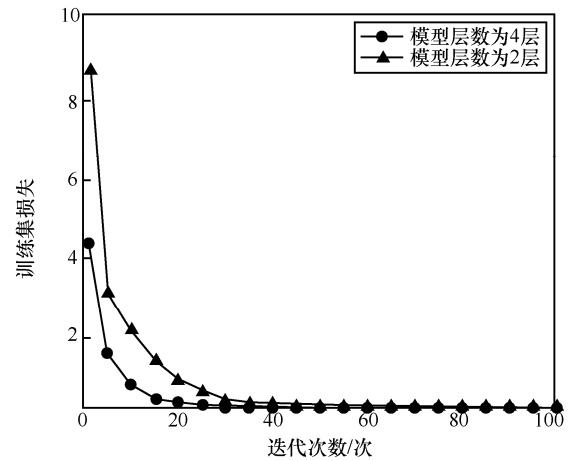


图 5 不同层数对训练集损失的影响

## 5 结束语

本文针对汉语谓语中心词进行识别研究, 使用深层神经网络模型对句子进行建模, 在汉语谓语中心词数据集上进行实验, F 值达到 80.424%, 并且通过输出路径的约束解决了中心词的唯一性问题。在未来的研究中, 需要利用更深层的神经网络模型来获取句子的结构信息, 以进一步提升识别性能。此外, 下一阶段工作将研究模型与输出路径间的高阶依赖问题, 通过在神经网络模型中加入全局约束条件以及构建知识库规则, 使模型在训练过程中能够自动学习句子的结构特点, 以支撑与谓语中心词相关的研究工作。

### 参考文献:

- [1] LUO Z S, ZHENG B X. An approach to the automatic analysis and frequency statistics of Chinese sentence patterns[J]. Journal of Chinese Information Processing, 1994, 8(2): 1-9.
- [2] LI G C, MENG J. A method of identifying the predicate head based on the correspondence between the subject and the predicate[J]. Journal of Chinese Information Processing, 2005, 19(1): 1-7.
- [3] SUI Z F, YU S W. The research on recognizing the predicate head of a Chinese simple sentence in EBMT[J]. Journal of Chinese Information Processing, 1998, 12(4): 39-46.
- [4] SUI Z F, YU S W. The acquisition and application of the knowledge for recognizing the predicate head of a Chinese simple sentence[J]. Acta Scientiarum Naturalium Universitatis Pekinensis, 1998, 34(2-3): 221-229.
- [5] 陈小荷, 石定栩. 汉语句子的主题-主语标注[C]//全国计算语言学联合学术会议. 北京: 中国计算机学会, 1997: 102-108. CHEN X H, SHI D X. Topic-subject labeling of Chinese sentences[C]//National Computer Linguistics Joint Academic Conference. Beijing: China Computer Federation, 1997: 102-108.
- [6] WANG H L, ZHOU G D. Feature engineering for predicate identification and classification in semantic analysis[J]. Computer Engineering

- and Application, 2010, 46(9): 134-137.
- [7] 谌志群. 汉语句子谓语识别的自动识别方法研究[J]. 计算机工程与应用, 2007, 43(17): 176-178.  
CHEN Z Q. Study on recognizing predicate of Chinese sentences[J]. Engineering and Applications, 2007, 43(17): 176-178.
- [8] GONG X J, LUO Z S, LUO W H. Recognizing the predicate head of Chinese sentences[J]. Journal of Chinese Information Processing, 2003, 17(2): 7-13.
- [9] HAN L, LUO S L, PAN L M. High accuracy Chinese predicate recognition method combining lexical and syntactic feature [J]. Journal of Zhejiang University (Engineering Edition), 2014, 48(12): 2107-2114.
- [10] 李琳, 赵维纳, 泽旺宽卓. 基于词向量特征的藏语谓语动词短语识别模型[J]. 电子技术与软件工程, 2019(4): 242-243.  
LI L, ZHAO W N, ZEWANG K Z. Tibetan predicate verb phrase recognition model based on word vector features[J]. Electronic Technology & Software Engineering, 2019(4): 242-243.
- [11] RAVINER L, JUANG B. An introduction to hidden Markov models[J]. IEEE ASSP Magazine, 1986, 3(1): 4-16.
- [12] LAFFERTY J, MCCALLUM A, PEREIRA F C N. Conditional random fields: probabilistic models for segmenting and labeling sequence data[J]. Proceedings of the 18th International Conference on Machine Learning. New York: ACM Press, 2001: 282-289.
- [13] HAMMERTON J. Named entity recognition with long-short term memory[C]//Proceedings of the 17th Conference on Natural Language Learning at HLT-NAACL. New York: ACM Press, 2003: 172-175.
- [14] LI F, ZHANG M, TIAN B, et al. Recognizing irregular entities in biomedical text via deep neural networks[J]. Pattern Recognition Letters, 2018, 105: 105-113.
- [15] CARRERAS X, MARQUEZ L. Introduction to the CoNLL-2005 shared task: semantic role labeling[C]//Proceedings of the 19th Conference on Computational Natural Language Learning. Boston: Association for Computational Linguistics, 2005: 152-164.
- [16] KOOMEN P, PUNYAKANOK V, ROTH D, et al. Generalized inference with multiple semantic role labeling systems[C]//Proceedings of the 9th Conference on Computational Natural Language Learning. Boston: Association for Computational Linguistics, 2005: 181-184.
- [17] TACKSTROM O, GANCHEV K, DAS D. Efficient inference and structured learning for semantic role labeling[J]. Transactions of the Association for Computational Linguistics, 2015, 3: 29-41.
- [18] 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 and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers). Boston: Association for Computational Linguistics, 2015: 1127-1137.
- [19] GUO J, CHE W, WANG H, et al. A unified architecture for semantic role labeling and relation classification[C]//Proceedings of the 26th International Conference on Computational Linguistics: Technical Papers. [S.n.:s.l.], 2016: 1264-1274.
- [20] 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]. 汉语信息学报, 2017, 31(1): 147-154.  
WANG R B, LI J H, LI G C, et al. Chinese frame semantic role recognition based on Dropout regularization[J]. Journal of Chinese Information Processing, 2017, 31(1): 147-154.
- [21] STRUBELL E, VERGA P, ANDOR D, et al. Linguistically-informed self-attention for semantic role labeling[C]// Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. Boston: Association for Computational Linguistics, 2018: 5027-5038.
- [22] HE L, LEE K, LEWIS M, et al. Deep semantic role labeling: what works and what's next[C]//Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Boston: Association for Computational Linguistics, 2017: 473-483.
- [23] PUNDAK G, SAINATH T N. Highway-LSTM and recurrent highway networks for speech recognition[C]//Interspeech, [S.n.:s.l.], 2017.
- [24] ZHANG Y, CHEN G, YU D, et al. Highway long short-term memory RNNs for distant speech recognition[C]//2016 IEEE International Conference on Acoustics, Speech and Signal Processing. Piscataway: IEEE Press, 2016: 5755-5759.
- [25] GAL Y, GHAHRAMAN Z. A theoretically grounded application of dropout in recurrent neural networks[J]. Advances in Neural Information Processing Systems, 2016, 29: 1019-1027.
- [26] 李婷, 秦永彬, 黄瑞章, 等. 基于神经网络的汉语谓语动词识别研究[J]. 数据采集与处理, 2020, 35(3): 582-590.  
LI T, QIN Y B, HUANG R Z, et al. Research on Chinese predicate verb recognition based on neural network[J]. Journal of Data Acquisition & Processing, 2020(3): 582-590.
- [27] SAXE A M, MCCLELLAND J L, GANGULI S. Exact solutions to the nonlinear dynamics of learning in deep linear neural networks[J]. arXiv Preprint, arXiv:1312.6120, 2013.
- [28] ZEILER M D. ADADELTA: an adaptive learning rate method[J]. arXiv Preprint, arXiv:1212.5701, 2012.
- [29] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]// IEEE Conference on Computer Vision & Pattern Recognition. Los Alamitos: IEEE Computer Society, 2016: 770-778.

## [作者简介]



黄瑞章 (1979-), 女, 天津人, 博士, 贵州大学副教授, 主要研究方向为数据融合分析、文本挖掘、网络挖掘和知识发现。

靳文繁 (1996-), 男, 贵州毕节人, 贵州大学硕士生, 主要研究方向为自然语言处理、信息抽取。

陈艳平 (1980-), 男, 贵州长顺人, 博士, 贵州大学副教授, 主要研究方向为人工智能、自然语言处理。

秦永彬 (1980-), 男, 山东招远人, 博士, 贵州大学教授, 主要研究方向为大数据治理与应用、多源数据融合。

郑庆华 (1969-), 男, 浙江嵊州人, 博士, 西安交通大学教授, 主要研究方向为大数据知识工程、网络舆情监测。