

文章编号: 1003-0077 (2011) 00-0000-00

基于分层输出神经网络的汉语语义角色标注*

王臻, 常宝宝, 穗志方

(北京大学 计算语言学教育部重点实验室, 北京 100871;

北京大学 计算语言学研究所, 北京 100871)

摘要: 语义角色标注是自然语言处理中的一项重要任务。当下针对中文语义角色标注的主流做法是通过基于特征的统计机器学习实现的。然而, 统计机器学习的方法需要引入经验性的人工特征, 这在一定程度上增加了工作量。深度学习在自然语言处理领域的应用使得特征的自动学习成为可能。文章尝试了一种适用于语义角色标注的深层神经网络架构, 该模型能自然地推广到其他标注任务。实验表明, 深度学习算法能够有效地用于语义角色标注任务, 但是我们仍然发现, 模型对语义层面知识的学习是相当有限的, 基于深度学习的方法还不能取代基于人工特征的统计机器学习算法。

关键词: 三语义角色标注; 深度学习; 特征向量

中图分类号: TP391

文献标识码: A

Chinese Semantic Role Labeling based on Neural Network with Optimized Output Layer

WANG Zhen, CHANG Baobao, SUI Zhifang

(Key Laboratory of Computational Linguistics (Peking University), Ministry of Education, Peking University, Beijing 100871, China;

Institute of Computational Linguistics, Peking University, Beijing 100871, China)

Abstract: Semantic role labeling is an important task of Chinese natural language processing. Using feature-based statistical machine learning to perform semantic role labeling is in popular nowadays. However, conventional machine learning methods heavily rely on manually designed features, which need intensive efforts. The development of deep learning algorithm in the field of natural language processing makes it possible to learn features automatically. This paper investigates semantic role labeling based on deep neural nets, which can be extended to other related work. Experimental results show that our algorithm is an effective method for Chinese semantic role labeling. However, because of the limitation to capture the knowledge of semantic sense, its performance still cannot reach conventional machine learning methods with manually designed features.

Key words: semantic role labeling; deep learning; feature vectors

1 引言

语义分析就是根据句子的句法结构和句中每个实词的词义, 推导出能够反映句子意义的某种形式化表示。对句子进行正确的语义分析, 一直是从事自然语言处理的人们追求的目标。随着自然语言处理基础技术, 如: 中文分词、词性标注、句法分析、机器学习等的逐步成熟, 以及语义分析在问答系统、信息抽取、机器翻译等领域的广泛应用, 这一方向越来越受到人们的重视。

目前深层语义分析很难做到, 人们更关注于浅层语义分析。浅层语义分析是深层语义分

* 收稿日期:

定稿日期:

基金项目: 国家重点基础研究发展计划 (2014CB340504); 国家自然科学基金 (61273318) (61375074)

作者简介: 王臻 (1990—), 男, 硕士研究生, 主要研究领域为自然语言处理。Email: wangzhen151@126.com; 常宝宝 (1971—), 男, 副教授, 主要研究领域为自然语言处理。Email: chbb@pku.edu.cn; 穗志方 (1970—), 女, 教授, 主要研究领域为自然语言处理。Email: szf@pku.edu.cn。

析的一种简化,它只标注与句子中谓词有关的成份的语义角色,如施事、受事、时间和地点等。

语义角色标注是浅层语义分析的一种实现方式。该方法并不对整个句子进行详细的语义分析,而只是根据特定的谓词去标注句子中的一些语义角色(论元),这些语义角色本身被赋予一定的语义含义。

目前大多数的语义角色标注任务基于的语料资源是提供完整句法树的宾州树库(Xue, 2002),以及完成人工语义角色标注的 Proposition Bank(Xue, 2003)。两者都有英文版和中文版,本文的工作是在上述的中文语料环境下完成的。

前人的工作大多需要产生完整的句法分析树作为预处理,例如(Sun and Jurafsky, 2004; Xue, 2008),再通过在树中抽取出很多人工设定的特征,结合传统的统计机器学习算法对语义角色进行分类标注。也有使用浅层句法分析的,例如(Sun and Sui, 2009),但是实验效果同样依赖于很多人工特征的使用。

近些年来,随着深度学习的兴起,人们越来越关注对特征表示的自动学习,即把特征的学习任务交由机器经过计算得到,而不是由领域专家事先人工确定,这在很大程度上减少了对任务的人为干预,也大大减轻了人们的工作量。(Collobert, 2008)的工作正是在这种环境下完成的,他使用深度神经网络的框架完成语义角色标注,摆脱了传统上对人工设定特征的依赖,但是原文工作是在英文环境下完成的,同样的方法是否适合于中文环境迄今为止还没有过完整论述。

本文尝试使用神经网络实现中文语义角色标注。章节安排具体如下:第2节介绍了语义角色标注的相关工作;第3节介绍了基于深度学习的语义角色标注网络框架;第4节介绍了深度学习网络的训练方式;第5节为实验结果及分析;最后对本文工作进行总结,并指出将来工作的方向。

2 相关工作

中文语义角色标注起步较晚,早期的工作侧重于使用在英文语义角色标注中取得成效的方法,比如(Sun and Jurafsky, 2004)使用 Collins parser 得到完整的句法树后移植了大部分在英文工作中用到的特征,继而使用 SVM 算法,但受制于当时的语料规模,其实验结果与后续工作并没有可比性。

在大规模语料 CPB (Chinese Proposition Bank) 建成之后,(Xue, 2005)、(Xue, 2008)在工作中加入了一些有助于性能提升的特征,并对中文语义角色标注做了系统性的研究。

(Ding and Chang, 2008)侧重于语义角色分类,采用层次结构,将这一过程分为三个子任务。(Chen, 2008)使用直推式 SVM 算法运用到中文语义角色标注工作中。上述工作中无一例外地使用了语句的完整句法结构信息,这些信息对语义角色边界的识别以及分类都是很有用的,所以语义角色标注的效果受制于自动句法分析器的性能。(Sun and Sui, 2009)引入浅层句法分析处理语义角色标注,避免了使用自动句法分析器所带来的性能损失,在语义组块一级进行语义角色识别和分类标注。(Sun, 2010)对这一方法进行了改进,加入了几种新的路径特征以及扩大了语义组块,使得总的 F1 值达到了 76.46%,这也是迄今为止获知的在中文语义角色标注上最好的性能结果。

上述所有方法的特点是使用了很多人工设定的特征,这有赖于对特征工程的研究,是比较费时费力的。在英文中,(Collobert, 2008)的工作打破了这一传统,文章使用深度学习框架自动学习特征的向量表示,使得与任务相关的信息更容易被捕获,避免了过多地人为介入。实验结果也是到达了 74%之多,比较接近英文语义角色标注的最好水平。

本文受上述英文工作的启发,尝试使用深度学习方法进行中文语义角色标注,通过自动特征学习避免了繁复的人工特征设定,在中文环境下获得了不错的实验效果。

3 基于分层输出的神经网络框架

中文语义角色标注可以视为对输入句子中的每个词指定一个标签。传统机器学习方法的性能很大程度上受制于所选的特征，正是由于这个原因，对该问题的研究很多都偏向于特征工程，而这些工作需要大量的人工介入，费时费力。

为了使得语义角色标注脱离特征选择的繁复工作，本文使用了神经网络的方法。该神经网络以每个待标注的句子作为输入，在多层隐层中获取输入的抽象表示，最后通过输出层得到可能性最高的标签序列，完成整个标注工作。网络的框架如图 1 所示，Lookup Table 层提取出每个词的特征向量，固定窗口内的特征向量进行拼接作为下方神经网络的输入。对于最后的结果，我们可以使用 Viterbi 算法进行解码得到最后的标注序列。

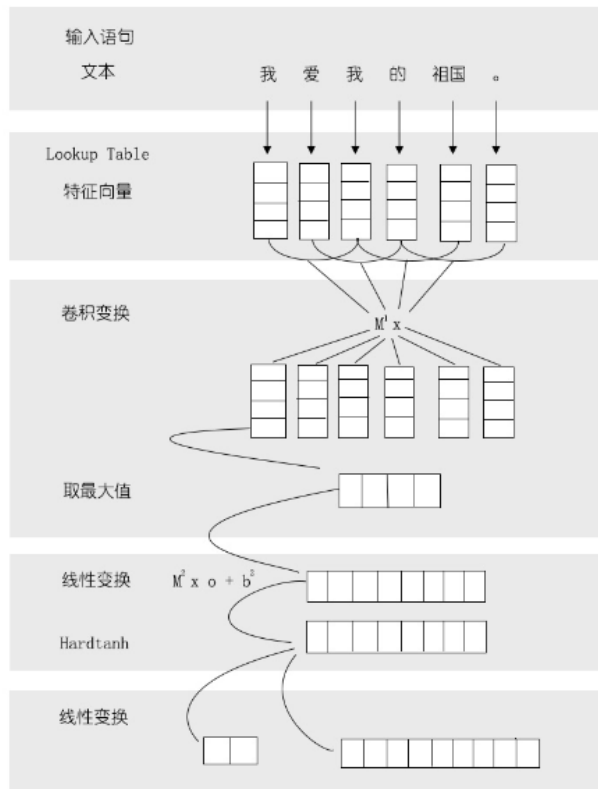


图 1 网络框架

本次工作对模型的主要改进集中于对网络输出层的改进。传统的深度学习方法使用的都是单层输出层进行分类标记，(Collobert, 2008) 一文正是这样做的。然而语义角色标注不同于一般的标注问题，因为它有牵扯到非语义角色和语义角色，语义角色细分又可以分为多种不同的语义角色标记。通过上面的分析，我们可以构造两个层次对语义角色标注问题加以描述：角色识别和角色分类，由此引入了两个网络输出层，一层用来进行角色识别，一层用来表示角色分类，两层结合以给出输入语句的标注序列，这也是本文工作的重心，我们在 3.2 节中会对这一部分重点进行解释说明。

3.1 中文语义角色标注

语义角色标注是浅层语义分析的一种实现方式。该方法并不对整个句子进行详细的语义分析，而只是根据特定的谓词去标注句子中的一些语义角色（论元），这些语义角色本身被赋予一定的语义含义。

目前大多数的语义角色标注任务基于的语料资源是提供完整句法树的宾州树库 (Xue, 2002)，以及完成人工语义角色标注的 Proposition Bank (Xue, 2003)。在 Proposition Bank 中，核心的语义角色有 Arg0~5 共六种，前缀 ArgM 表示非核心论元角色，后面跟一些附加标记表示具体的语义类别，如 ArgM_LOC 表示地点，ArgM_TMP 表示时间等。

图 2 对一个标注实例具体进行说明：民众/已经/充分/掌握/事情/真相。其中，“掌握”

为谓词，用 Rel 表示；“民众”为施事，用 Arg0 表示；“事情真相”为受事，用 Arg1 表示；“已经”和“充分”分别表示发生的时间和程度，用 ArgM_TMP 和 ArgM_MNR 表示。下面的语法树分析了句子的结构，并标记出了关于谓词各个语义角色。

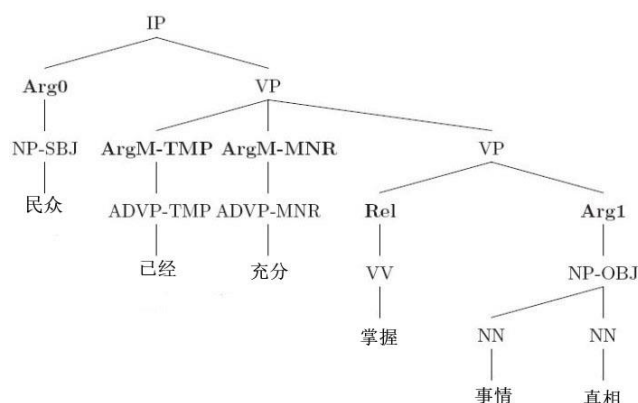


图2 语义角色标注示例

所以对于在 Chinese Proposition Bank 标注体系下进行的语义角色标注任务，我们的目标是给定一个输入语句和相关谓词，使得系统能够自动地为每个词语标注语义角色（如上例中的 Arg0、Arg1、ArgM_TMP、ArgM_MNR），并力求达到好的标注效果。

3.2 分层输出结构

针对序列标注问题构造神经网络，传统上的做法是将网络的输出层表示为单层结构 (Collobert, 2008)，将所有的标记一起进行考虑。对于语义角色标注问题而言，即将各类语义角色与非语义角色合并考虑，最后的输出层有语义角色个数+1 个节点，分别表示每一类标注的得分，这样的结构如图 3 所示。

这种方法对于处理某些简单问题是比较合适的，将输出表示为单层网络也意味着所有的标记之间是平等的关系。然而，对于中文语义角色标注问题来说，将所有标记同等考虑是不恰当的，因为非语义角色和各类语义角色是两个层级的概念，如果顺序考察这一任务，我们先得确定一个词语是语义角色，然后再对它是什么语义角色进行判断，而单层神经网络是不能捕捉到这样的两层语义关系。

基于上面的原因，对于本次工作塑造的神经网络的输出层，我们采用图 4 的结构表示神经网络输出。针对最后一层隐藏层，通过两个不同的线性变换得到两个不同的输出层，其中一个输出层有两个节点，表示这个词是语义角色或是非语义角色的得分，另一个输出层有语义角色个数个节点，表示这个词是每种语义角色的得分。这两个部分是独立存在的，也就是说两者的线性变换之间没有严格的关联，不过在训练中需要将两层的输出合并进行考虑，这在第 4 部分会有详细的说明。

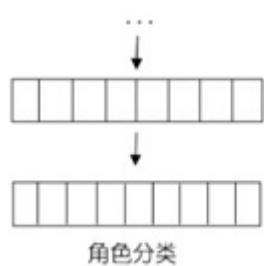


图3 传统输出层

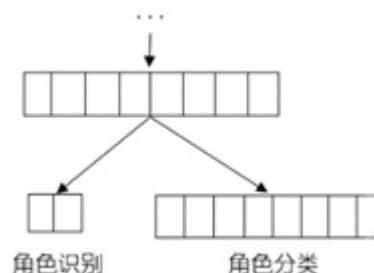


图4 改进输出层

本文的做法很好地模拟了中文语义角色标注实施过程中的两个步骤，将角色识别和角色分类分开进行考虑。我们也在实验结果中发现，这种分层输出的方法比传统处理序列标注任务的单层输出效果要好，这也证明了我们设计神经网络的合理性。

3.3 提取中文特征向量

本文工作中使用的特征是词、词到标注词的距离、词到特定谓词的距离，针对这三个特征分别维护了 Lookup Table，从中可以得到每个特征的向量表示。假设特征的集合是 \mathbf{D} ，则特征的向量表示存储在一个 $\mathbf{M} \in \mathbb{R}^{d \times |\mathbf{D}|}$ 的矩阵之中， d 是特征向量的维数， $|\mathbf{D}|$ 是特征集合的大小。

以词特征为例，具体阐述这个过程。假设给定了一个中文句子 $c_{[1:n]}$ ，每个词设为 c_i ， $1 \leq i \leq n$ 。对于每个 $c_i \in \mathbf{D}$ ，有它在 \mathbf{D} 中的索引 k_i ，用 I_{k_i} 表示只在第 k_i 位为 1，其余位皆为 0 的 $|\mathbf{D}|$ 维向量，则关于 c_i 可以得到它的特征向量 $W_{c_i} \in \mathbb{R}^d$ ：

$$W_{c_i} = M I_{k_i} \quad (1)$$

值得说明的是，如果一个词没有出现在 \mathbf{D} 的集合内，可以将其标记为 **UNKNOWN**，将 **UNKNOWN** 加入到集合 \mathbf{D} 中作为所有未出现的词特征，这也意味着它们将共享同一特征向量。

3.4 窗口向量拼接

工作中引入窗口的概念，对每个词特定窗口内所有词的特征向量进行拼接作为该词的最终的特征向量输入下层神经网络。如此，每个词在网络中不再是孤立的存在，语义上的相关性在某种程度上可以通过这个方式捕捉到。

具体地，如果每个词通过第一步得到的特征向量是 W_{c_i} ，那么经过窗口拼接的特征向量被转化为：

$$z_{c_i}^1 = (W_{c_{i-d/2}} \dots W_{c_i} \dots W_{c_{i+d_{win}/2}})^T \quad (2)$$

其中 d_{win} 表示窗口的大小。

3.5 卷积变换

因为每个句子的长度都是不确定的，所以在网络中引入一个卷积变换层对输入特征向量进行归一化变换。具体地，设变换矩阵为 M^1 ，卷积变换之后的输出为：

$$z_t^2 = \max_t (M^1 Z^1)_{i,t} \quad (3)$$

其中， $1 \leq t \leq n$ ， n 表示待标注语句的长度； $1 \leq i \leq h_1$ ， h_1 表示第一层隐层的节点个数。

3.6 深层网络结构

后续的神经网络计算，使用如下公式进行线性变换：

$$z^l = M^{l-1} z^{l-1} + b^{l-1} \quad (4)$$

其中 M^{l-1} 表示 $l-1$ 层到 l 层的权值矩阵， b^{l-1} 表示 $l-1$ 层到 l 层的阈值向量。

工作中隐层使用的激活函数是 **hardtanh**，其定义如下所示：

$$f(x) = \begin{cases} -1, & x \leq -1 \\ x, & -1 \leq x \leq 1 \\ 1, & x \geq 1 \end{cases} \quad (5)$$

这是受到前人工作的启发而设定的。

3. 7 标注策略

虽然我们的工作目标是给每个词进行 ArgO~5 或是 ArgM-X 的标记,但是单纯使用这样的标记,我们并不能得到语义角色的边界信息,因为每个语义角色的开头和结尾都应该是有据可循的,单纯使用语义角色的标注显然会忽略掉这一点。

所以,针对上面情况的分析,我们在实验中采用 IOBES 的标注策略,I-ARGX 表示一个语义角色的内层部分,B-ARGX 表示语义角色的开始部分,E-ARGX 表示语义角色的结尾,S-ARGX 表示单个词形成的语义角色,0 表示不属于语义角色的部分。而在预测阶段,我们会将该标注策略转换为仅表示语义角色的标注结果。

4 基于词标签的网络训练方式

本次工作采用了基于词标签的训练方式对深度学习网络进行训练,具体地,我们是通过最大化似然函数进行参数求解的,随机梯度下降能够很好地运用在问题的求解中。用 θ 表示网络中的所有参数, x 表示一个特定的待标记的词, y 表示对应的正确标记,实际上,我们试图最大化的是下面这个式子:

$$\sum_{(x,y) \in T} \log P(y|x, \theta) \quad (6)$$

其中, T 是所有的训练数据。

在这种训练方法中,我们单独考虑句子中每个词的标签。对于每个词 x ,网络的第一个输出层输出 $f_1(x, i_1, \theta)$,其中 i_1 为 0 或 1,分别表示 x 是语义角色和非语义角色的得分,运用 softmax 操作进行归一化,我们可以得到当前词 x 是语义角色或是非语义角色的概率:

$$p_1(i_1|x, \theta) = \frac{e^{f_1(x, i_1, \theta)}}{\sum_j e^{f_1(x, j, \theta)}} \quad (7)$$

网络的第二个输出层输出 $f_2(x, i_2, \theta)$,其中 i_2 为对应每一种语义角色类型,分别表示 x 是每种语义角色的得分,运用 softmax 操作进行归一化,我们可以得到当前词 x 是每种语义角色的概率:

$$p_2(i_2|x, \theta) = \frac{e^{f_2(x, i_2, \theta)}}{\sum_j e^{f_2(x, j, \theta)}} \quad (8)$$

我们通过简单的概率公式可以计算得到 x 属于每种类别的概率:

$$p(i|x, \theta) = \begin{cases} p_1(1|x, \theta), & i \text{ 表示非语义角色} \\ p_1(0|x, \theta) \times p_2(i|x, \theta), & i \text{ 表示某种语义角色} \end{cases} \quad (9)$$

之后再运用上面提到的最大似然方法进行训练就可以得到整个网络模型的各个参数。

5 实验

5. 1 实验设置

我们采用中文 Proposition Bank 的数据集,同前人的工作一样,使用 Proposition Bank

中的 02-21 分块作为训练数据，第 24 分块作为开发集数据，第 23 块用于测试集。本次实验用到的语义角色有 ArgO^{~5}，以及其余的十三种修饰角色 ArgM-X，如 ArgM-LOC 修饰地点，ArgM-TMP 修饰时间等。

5. 2 超参数设置

对于深层网络中超参数的选择，我们使用经验知识和前人工作进行判断，(Yoshua Bengio, 2012) 对深层网络常用的参数设置给出了经验上的指导，本文的工作受上文和之前工作的启发，对网络架构中的超参数进行了人为设定。

工作中训练网络的学习率设为 0.001，窗口拼接时的窗口大小设为 5，词表大小设置为 10000 (使用最频繁出现的前 10000 个词)，特征向量的长度分别为 50、20、20，两层隐藏层的节点数目分别是 200 和 100。以上就是所有的参数设定。

5. 3 词向量初始化

对于待标记词到给定谓词距离以及当前词到待标记词距离这两个特征，我们都是使用随机初始化的方式对它们的特征向量进行初始化，因为对于它们而言我们没有很多先验知识去得到有意义的特征向量。对于词特征，我们当然也可以通过随机初始化的方式对它进行赋值，这样造成的代价是网络的训练时间会更长。由于目前大规模中文语料的存在，对词向量的初始化我们有了更好的选择。

word2vec 是 google 的一个基于上下文共现信息学习词向量的一个实用的工具，借助它在 **giga** 语料上学习得到的词向量，我们可以对网络中用到的词的特征表示进行有意义的初始化，因为实验表明该工具生成的词向量在一定程度上是能够把握词语之间的语义信息的。使用这种词的特征表示也使得网络的训练时间大幅度降低。

5. 4 实验结果

上文提到，我们在标记时使用的是 IOBES 的标注策略，该标注策略实际上给标注结果添加了一种可能性限制。比如说，对于 B-ARGO 之后只能出现 I-ARGO 或是 E-ARGO，其余的标注结果都是非法的，对于 I-ARGO，其后仅有 I-ARGO 和 E-ARGO 合法，我们可以对每种情况进行列举，得出关于标记序列的一些限制，使用这些限制，我们才能得到合法的标记序列。

具体地，我们在预测阶段可以得到每个词关于所有标记的得分 (在 IOBES 标注体系之下)，我们再根据上述提到的限制生成一个限制矩阵 $A_{m \times m}$ ， m 表示在 IOBES 标注体系下的标记个数，矩阵元素 a_{ij} 表示从标记 i 到标记 j 的可能性，我们这里对其赋值 0 表示可能的转移，赋值 $-\infty$ 表示不可能的转移。基于上述得分以及转移矩阵，我们可以使用 Viterbi 算法进行解码，得到最大值的序列就是我们在预测中给定的标记序列，上述的赋值保证了该标记序列的合法性。之后，再对 IOBES 标记序列进行转化，变为仅使用语义角色的序列，使用此序列与标准答案进行比较，得到我们关心的准确率、召回率和 F 值。

表 1 给出了本文工作得到的实验结果，其中的 baseline 是仅使用一层输出层进行标记的结果，我们可以看到本文工作对输出层的改进使得实验结果有了明显的提升。

	Precision	Recall	F score
baseline	0.6125	0.6238	0.6181
NNsr1	0.6446	0.6392	0.6419

表 1 实验结果

通过表 1 我们可以发现，改进的框架模型不管是在准确率还是召回率上，相对于单层输

出的模型都有比较明显的提升，而对于 F 值，改进后的框架在原有的基础上提升了 0.02 之多，这也证明了我们分离角色识别和角色分类的改进是有效的。

虽然如此，我们不得不承认该实验结果离目前基于传统机器学习方法的中文语义角色标注的最好水平还有一段距离，这也是需要我们去总结和反思的地方。

5. 5 词向量

本次实验的另一个发现是，经过深层学习网络的训练，我们得到了更好的词向量表示，以中国为例，表 2 列出了分别在训练前后与中国一词最近的十个词语。

训练前	训练后
作为	新加坡
亚洲	日本
国家	印度尼西亚
乃至	菲律宾
界	印度
全球	马来西亚
来自	朝鲜
世界	陆
本土	泰国
东亚	印尼

表 2 词语相关度

通过表 2 我们可以发现，经过针对语义角色标注的深层神经网络的训练，我们得到的词向量能更好地反映出词语的语义关联。表 2 中训练前得到的结果还比较杂乱无章，虽然大多表示的都是地点，但是都无法与中国保持在一个层级之上；而在训练之后，这一现象有了很大的改观，得到的词语基本都是国家名称，这是一个很大的改进，说明了我们的网络在得到词向量方面具备很强的能力，这虽然不是本次工作的重心，但是对以后的工作方向也是很有启发的。

6 总结

本文尝试使用深层神经网络模型进行中文语义角色标注的探索，首次将角色识别和角色分类两个步骤以网络层的形式引入深度学习模型。在测试语料上的实验结果表明，该方法在一定程度上可以对语义角色进行标注，引入的两层输出模型性能上也超越了单层输出的模型。该方法有效地规避了人工设定特征的步骤，一定程度上减少了工作量，同时也对词向量产生了不错的补充与改进。然而，我们不能忽略的是，本文对中文语义角色标注任务的处理效果并不理想，与主流方法的差距还是比较明显的。

下一步，我们将参考前人工作试图在网络中引入人工设计的语义特征信息，并适当扩充网络的规模和复杂性，尝试从更深层次的语义层面入手，挖掘出更多的语义信息，帮助进行中文语义角色标注任务，从而提升整个工作的性能。

参考文献

- [1] Nianwen Xue. Building a Large-Scale Annotated Chinese Corpus[C]//Proceedings of the 19th international conference on Computational linguistics. 2002: 1-8.
- [2] Nianwen Xue and Martha Palmer. Annotating the propositions in the Penn Chinese Treebank[C]//Proceedings of the second SIGHAN workshop on Chinese language processing. 2003: 47-54.

- [3] Honglin Sun and Daniel Jurafsky. Shallow semantic parsing of Chinese[C]//In Proceedings of NAACL-HLT. 2004.
- [4] Nianwen Xue. Labeling Chinese predicates with semantic roles[J]. Computational Linguistics, 2008, 34(2): 225-255.
- [5] Weiwei Sun and Zhifang Sui. Chinese function tag labeling[C]//In Proceedings of the 23rd Pacific Asia Conference on Language, Information and Computation. 2009.
- [6] Weiwei Sun, Zhifang Sui, Meng Wang and Xin Wang. Chinese semantic role labeling with shallow parsing[C]//In Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing. 2009: 1475-1483.
- [7] Collobert Ronan and Weston Jason. A unified architecture for natural language processing: Deep neural networks with multitask learning[C]//Proceedings of the 25th international conference on machine learning. 2008: 160-167.
- [8] Nianwen Xue and Martha Palmer. Automatic semantic role labeling for Chinese verbs[C]//In Proceedings of the 19th International Joint Conference on Artificial Intelligence. 2005.
- [9] Weiwei Ding and Baobao Chang. Improving Chinese semantic role classification with hierarchical feature selection strategy[C]//In Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing. 2008.
- [10] Weiwei Ding and Baobao Chang. Fast semantic role labeling for Chinese based on semantic chunking[C]//Proceedings for of the 22nd International Conference on Computer Processing of Oriental Languages. 2009:79-90.
- [11] Yaodong Chen, Ting Wang, Huowang Chen, and Xishan Xu. Semantic role labeling of Chinese using transductive svm and semantic heuristics[C]//In Proceedings of the Third International Joint Conference on Natural Language Processing. 2008.
- [12] Weiwei Sun. Improving Chinese semantic role labeling with rich syntactic features[C]//Proceedings of the ACL 2010 Conference Short Papers. 2010: 168-172.
- [13] Weiwei Sun. Semantics-driven shallow parsing for Chinese semantic role labeling[C]//Association for Computational Linguistics (ACL). 2010.
- [14] Bengio, Y. Practical recommendations for gradient-based training of deep architectures. In NN: Tricks of the Trade[M]. 2012: 437-478.