

基于句法结构的神经网络复述识别模型

刘明童¹ 张玉洁^{1,†} 徐金安¹ 陈钰枫¹

1. 北京交通大学计算机与信息技术学院, 北京 100044

† 通讯作者, E-mail: yjzhang@bjtu.edu.cn

摘要 复述识别涉及到重要的自然语言语义理解方法。已有基于神经网络的复述识别方法将句子看作序列化结构, 并采用序列化结构神经网络进行语义组合计算。这些方法没有考虑句法结构对语义计算的影响。对此, 提出基于句法结构的神经网络复述识别模型, 设计基于树结构的神经网络模型进行语义组合计算, 使得语义表示从词语级扩展到短语级。进一步, 提出基于短语级语义表示的句法树对齐机制, 利用跨句子注意力机制提取特征。最后, 设计自注意力机制捕获全局信息增强语义表示, 有效建模基于句法结构的上下文信息。在公开英语复述识别数据集 Quora 上进行了评测实验, 实验结果显示, 复述识别性能得到改进, 达到 89.3% 的精度, 表明提出的基于句法结构的语义组合计算, 以及基于短语级语义表示的跨句子注意力机制和自注意力机制在改进复述识别性能方面的有效性。

关键词 复述识别; 句法结构; 树结构神经网络; 注意力机制

中图分类号 TN914

A Neural Paraphrase Identification Model Based on Syntactic Structure

LIU Mingtong¹, ZHANG Yujie^{1,†}, XU Jinan¹, CHEN Yufeng¹

1. School of Computer and Information Technology, Beijing Jiaotong University, Beijing 100044

†Corresponding Author, E-mail: yjzhang@bjtu.edu.cn

Abstract Paraphrase identification involves natural language semantic understanding. Most previous methods regarded sentences as sequential structures, and used sequential neural network for semantic composition. These methods do not consider the influence of syntactic structure on semantic computation. In this paper, the authors proposed a neural paraphrase identification model based on syntactic structure, and designed a tree-based neural network model for semantic composition, which extends the semantic representation from word level to phrase level. Furthermore, this paper proposed a syntactic tree alignment mechanism based on phrase-level semantic representation, and extracted features by using cross-sentence attention mechanism. Finally, this paper used a self-attention mechanism to enhance semantic representation, which can effectively model context information based on syntactic structure. Experimental results on Quora paraphrase dataset show that the performance of paraphrase identification has been improved to 89.3% accuracy. Deep analysis further confirms that the proposed approach can improve model accuracy.

Key words paraphrase identification; syntactic structure; tree-structured neural network; attention mechanism

自然语言理解是自然语言处理的终极目标, 其判定标准包括复述、翻译、问答和文摘, 复述处理的复杂性和重要性可见一斑^[1]。复述识别是复述处理的关键技术之一, 广泛应用于机器翻译^[2,3]、自动问答^[4,5]和文本摘要^[5,6]等自然语言处理任务中, 如机器翻译评测中识别同义表达, 提高评测系统质量; 自动问答中识别相同意思的问句, 提高问答系统的性能; 文本摘要中检测句子相似性, 从而更好进行句子压缩和生成。近年来, 基于神经网络的复述研究日益成为关注热点, 成为自然语言处理的重要方向之一。本文主要针对基于神经网络的复述识别方法开展研究。

早期的复述识别方法主要是基于特征工程的方法^[7-10], 如利用机器翻译评价指标 BLEU 值计算句子相

国家自然科学基金(61876198,61370130,61473294), 中央高校基本科研业务费专项资金(2018YJS025,2015JBM033), 科学技术部国际科技合作计划(K11F100010) 资助

收稿日期: 2010-12-03; 修回日期: 2010-04-12; 网络出版时间:

网络出版地址:

似度^[9]，利用语义角色标注信息作为特征^[10]，该类方法依赖于句子间组成部分是否相同，其主要问题是难以捕获表面形式差异大而语义相同的复述现象。后来，基于神经网络的模型在语义计算和特征提取上显示出了优势^[11-17]，研究者利用神经网络将句子表示为低维连续的语义向量，然后将语义向量作为特征用于复述识别，该类方法避免了复杂的特征工程，并且可以提取句子全局特征。基于神经网络的模型表明了句子间的语义相关性越大，它们在语义空间上的距离也越接近，其主要研究句子的语义表示方法，即从词语到句子的语义组合方法，同时利用神经注意力机制提取跨句子特征。

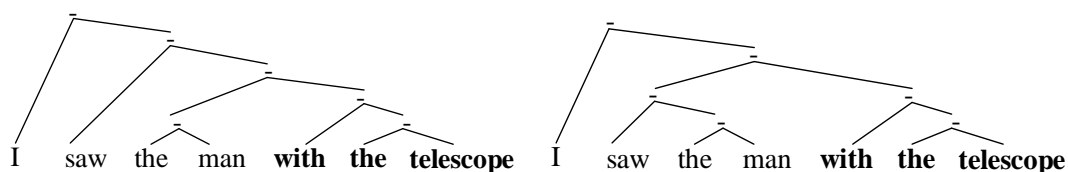


图 1 相同词序列的不同句法结构

Fig.1 Sentences with same word orders but different syntactic structures

已有模型在复述语义组合计算和跨句子特征提取上仍存在以下问题。一方面，在计算句子语义表示时，没有考虑句法结构信息。其将句子看作序列化结构，进行语义组合计算，该方法难以捕获细微语义差别，如“Show me the fights from **New York to Florida.**”和“Show me the fights from **Florida to New York.**”。同时，序列化的方法无法体现词语在句法结构上的语义关系。而自然语言句子是按照一定的句法规则生成，即使构成词语和词序完全相同的句子，如果句法结构不同，其语义也会有所不同。如图 1 所示，两个句子在序列上完全一致，但有不同的句法结构，在左边句法结构中，“with the telescope”修饰“the man”这个主语，语义解释为“该男士有一个望远镜”；在右边的句法结构中，“with the telescope”修饰“saw”这个动作，语义解释为“我用望远镜看一个男士”。由此可见，句法结构在理解句子语义上有重要作用。另一方面，在学习跨句子的特征时，序列化结构无法构建短语的表示，其主要关注句子间词语级的匹配信息^[11,13]，而无法有效计算短语级的匹配信息。针对以上问题，本文提出基于句法结构的复述识别模型，设计树结构神经网络模型进行语义组合计算，构建短语语义表示，并基于短语表示设计跨句子注意力和自注意力机制学习特征。

本文剩余部分组织如下：第一节介绍相关研究；第二节针对基于句法结构的复述识别模型，描述基于句法树的神经网络语义组合计算方法，以及基于短语级语义表示的注意力机制；第三节介绍神经网络复述识别模型的实现和训练细节；第四节介绍评测实验和结果分析；最后，对本文研究进行总结。

1 相关研究

复述识别是指判定给定的两个句子 P 和 Q 在语义上是否一致的问题。在基于神经网络的复述识别方面，文献^[12,16,17]考虑句子中的 n -gram 特征，利用卷积神经网络建模，并采用注意力机制提取跨句子特征，相比传统基于特征工程的方法取得了更好的性能。基于循环神经网络 BiLSTM 的模型也广泛应用于复述识别，文献^[11,13]提出基于词语级语义匹配的方法，其主要关注句子间词语级的语义匹配信息，采用注意力机制学习词对齐信息，从而提取跨句子特征。文献^[13]扩展了先前的注意力机制，设计多个注意力层用于提取跨句子特征，并采用 BiLSTM 对注意力信息进行融合，增加了网络深度，进一步改进了复述识别的性能。文献^[14]在神经网络模型中融合语义角色信息，以利用更多的特征用于辅助复述识别。这些方法实现了高效的特征表示和提取方法，复述识别精度得到提升。其主要问题是，在学习句子表示时，将句子看作序列化结构，语义组合没有考虑句法结构信息，从而无法捕获具有语义修饰关系的词语进行语义组合计算。而句子是按照句法结构递归生成的，句法结构知识可以更好地解释句子的语义表示。另一方面，已有模型在提取跨句子特征时，只关注了词语级的匹配信息作为特征，而忽视了短语级的语义匹配，而短语级的匹配信息在识别句子语义关系有重要的作用。针对这些问题，本文研究基于句法结构的复述识别模型，设计基于句法结构的神经网络语义组合计算模型，同时研究基于短语级语义表示的跨句子特征提取方法，采用神经注意力机制学习语义表示，以有效利用句子内部和和句子间的特征，改进复述识别的精度。

2 基于句法结构的神经网络复述识别模型

我们提出基于句法结构的神经网络复述识别模型，采用基于树结构神经网络的语义组合计算方法、以及基于注意力机制的特征提取方法，主要框图如图 2 所示。模型输入两个句子 $P=\{p_1,\dots,p_i,\dots,p_m\}$ 和 $Q=\{q_1,\dots,q_j,\dots,q_n\}$ ，其中 p_i 和 q_j 分别表示句子 P 和 Q 中的第 i 和 j 个词语， m 和 n 分别表示句子 P 和 Q 的长度。最后，模型采用一个分类器预测句子 P 和 Q 的语义关系。下面主要从基于句法结构的语义组合计算方法、跨句子的短语级注意力机制和自注意力机制三个模块详细介绍我们的模型。

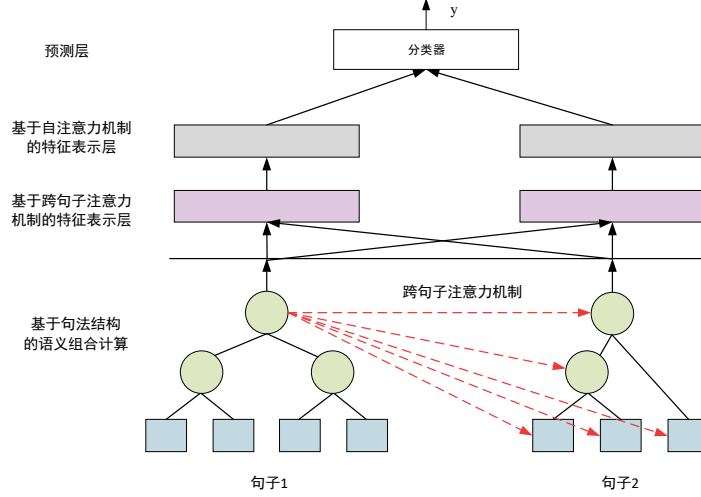


图 2 本文提出的基于句法结构的复述识别模型整体框架

Fig. 2 The proposed paraphrase identification framework based on syntactic structure

2.1 基于句法结构的语义组合计算方法

我们采用短语树作为句法结构指导语义组合计算，短语句法树刻画了句子内部词语之间的层次关系，即先由词语构成短语，然后由短语组合成整个句子。在短语句法树中，每一个叶子节点表示词语，每一个非叶节点表示构成的短语，根节点集成了所有孩子节点的信息，通常被看作整个句子的表示。其中，非叶节点仅包含两个孩子节点，左孩子节点 l 和右孩子节点 r ，孩子节点可以是词语，也可以是已构成的短语。

我们设计基于 TreeLSTM^[18]的神经网络结构作为语义组合计算模型，通过门控制机制（gate mechanism）选择输入信息，在合成短语表示时选择性的关注诸如动词、名词等重要节点，从而更好的编码短语的表示。在 TreeLSTM 神经网络中，每一个 TreeLSTM 神经元包含两个向量，记忆单元向量 c 和隐藏状态向量 h 。我们定义建立短语的左孩子节点的记忆单元向量和隐藏层向量为 (c^l, h^l) ，定义建立短语的右孩子节点的记忆单元向量和隐藏层向量为 (c^r, h^r) 。对于新合成的短语，定义 c^{ph} 和 h^{ph} 为合成短语的记忆单元向量和隐藏层向量， c^{ph} 和 h^{ph} 的计算如公式（1）-（3）所示。

$$\begin{bmatrix} i \\ f^l \\ f^r \\ o \\ \tilde{c} \end{bmatrix} = \begin{bmatrix} \sigma \\ \sigma \\ \sigma \\ \sigma \\ \tanh \end{bmatrix} \left(W^{tree} \cdot \begin{bmatrix} h^l \\ h^r \end{bmatrix} + b^{tree} \right) \quad (1)$$

$$c^{ph} = f^l \odot c^l + f^r \odot c^r + i \odot \tilde{c} \quad (2)$$

$$h^{ph} = o \odot \tanh(c^{ph}) \quad (3)$$

其中 σ 表示 sigmoid 函数， \odot 表示两个向量的逐点乘积操作。 i 、 f^l 、 f^r 、 o 和 \tilde{c} 分别表示输入门、左孩子节点对应的遗忘门、右孩子节点对应的遗忘门、输出门和候选记忆单元。在我们的结构中，短语表示是按照自底向上的规则组合生成的，从而递增的得到整个句子的表示，以及句法树中所有节点的表示。在此，句子 P 和 Q 被表示为 $h_p = \{h_{p_1}, \dots, h_{p_i}, \dots, h_{p_{2m-1}}\}$ 和 $h_q = \{h_{q_1}, \dots, h_{q_j}, \dots, h_{q_{2n-1}}\}$ ，其中， h_{p_i} 和 h_{q_j} 分别是 P 和 Q 中的第 i 和 j 个节点的语义向量。在长度为 m 的句子所对应的短语句法树中，有 m 个叶子节点和 $m-1$ 个非叶节点，最后共得到 $2m-1$ 个节点。接下来，我们将利用 h_p 和 h_q 学习跨句子特征。

2.2 基于短语级语义表示的跨句子注意力机制

我们基于短语级语义表示，设计跨句子注意力机制学习两个句子间的匹配信息。跨句子注意力机制从另一个句子中提取相关信息，给定句子 P 和 Q 中每一个节点的表示，我们首先计算每一对节点 i 和 j 的语义相关性得分，采用双仿射注意力机制^[19]，其计算方法如公式（4）所示。

$$A_{ij} = h_{p_i} W^a h_{q_j} + \langle U^p, h_{p_i} \rangle + \langle U^q, h_{q_j} \rangle \quad (4)$$

其中， W^a 、 U^p 和 U^q 表示学习参数， $\langle x, y \rangle$ 表示两个向量之间的内积操作， A_{ij} 表示 P 中第 i 个节点和 Q 中第 j 个节点的语义相关度得分。

接下来，我们对相关度得分做归一化处理，对每一个节点，计算其在另一个句子中的语义对齐信息，计算公式如（5）和（6）所示。

$$h_{p_i} = \sum_{j=1}^{2n-1} \frac{\exp(A_{ij})}{\sum_{k=1}^{2n-1} \exp(A_{ik})} h_{q_j} \quad (5)$$

$$h_{q_j} = \sum_{i=1}^{2m-1} \frac{\exp(A_{ij})}{\sum_{k=1}^{2m-1} \exp(A_{kj})} h_{p_i} \quad (6)$$

其中， h_{p_i} 表示 P 中节点 i 经过注意力在 Q 中得到的语义对齐信息， h_{q_j} 表示 Q 中节点 j 经过注意力在 P 中得到的语义对齐信息。直观上，若节点 i 和 j 在语义上更为相关，其对应的语义表示也更为接近，从而得到较大的注意力权重。最终， h_{p_i} 可以概率化地选择 h_{q_j} 的信息作为对齐的语义向量。

为了增强跨句子的特征，我们进一步设计了信息融合层，计算公式如（7）-（10）所示

$$\bar{h}_{p_i} = [h_{p_i}, h_{p_i}, |h_{p_i} - h_{p_i}|, h_{p_i} \odot h_{p_i}] \quad (7)$$

$$h_{p_i}^c = \text{ReLU}(W^c \bar{h}_{p_i} + b^c) \quad (8)$$

$$\bar{h}_{q_j} = [h_{q_j}, h_{q_j}, |h_{q_j} - h_{q_j}|, h_{q_j} \odot h_{q_j}] \quad (9)$$

$$h_{q_j}^c = \text{ReLU}(W^c \bar{h}_{q_j} + b^c) \quad (10)$$

其中， W^c 和 b^c 表示学习参数， $h_{p_i}^c$ 和 $h_{q_j}^c$ 分别表示融合后节点 i 和节点 j 的表示。在此，我们首先计算逐点的减法和乘积操作，用于捕获语义相似性，然后利用非线性变换 ReLU 函数进一步将各个表示结果进行融合，并降低语义向量表示的维度。由此，句子 P 和 Q 对应句法树中的节点被重新表示为 $h_p^c = \{h_{p_1}^c, \dots, h_{p_i}^c, \dots, h_{p_{2m-1}}^c\}$ 和 $h_q^c = \{h_{q_1}^c, \dots, h_{q_j}^c, \dots, h_{q_{2n-1}}^c\}$ 。

2.3 自注意力机制

针对每一个句子，我们采用自注意力机制（self-attention mechanism）进一步捕获基于句法结构的上下文信息，即使两个节点在句法树上路径较远，仍然可以捕获相关的信息，从而增强基于跨句子注意力机制的语义表示。对句子 P ，我们采用公式（4）中的注意力机制方法计算句法树中任意两个节点的语义相关性得分 S_{ij} ，然后计算基于自注意力机制的语义向量，计算公式如（11）所示。

$$h_{p_i}^c = \sum_{j=1}^{2m-1} \frac{\exp(S_{ij})}{\sum_{k=1}^{2m-1} \exp(S_{ik})} h_{p_j}^c \quad (11)$$

对句子 Q ，我们也采用了相同的自注意力机制。直观上， $h_{p_i}^c$ 和 $h_{q_j}^c$ 可以捕获对应句法树中每一个节点的信息，增强节点基于全局上下文的语义表示。我们也采用了公式（7）-（10）的融合层，进一步融合基于自注意力机制的语义表示，节点 i 和节点 j 分别被表示为 $h_{p_i}^s$ 和 $h_{q_j}^s$ 。

最后，我们融合基于跨句子注意力机制和自注意力机制的信息来表示句法树中的每一个节点，有效捕获两个句子基于句法结构的语义信息作为复述识别的特征，计算公式如（12）和（13）所示。

$$h_{p_i} = h_{p_i}^c + h_{p_i}^s \quad (12)$$

$$h_{q_j} = h_{q_j}^c + h_{q_j}^s \quad (13)$$

2.4 叶子节点表示方法

我们设计基于 BiLSTM^[20] 的叶子节点表示方法，为了获取更加丰富的上下文信息表示单词特征，我们

结合 LSTM 每一时刻的隐藏层向量 h_t 和记忆单元向量 c_t 表示每一个叶子节点的词语。

$$\vec{h}_t, \vec{c}_t = \overrightarrow{LSTM}(x_t, (\vec{h}_{t-1}, \vec{c}_{t-1})) \quad (14)$$

$$\vec{h}_t, \vec{c}_t = \overleftarrow{LSTM}(x_t, (\vec{h}_{t+1}, \vec{c}_{t+1})) \quad (15)$$

对当前时刻 t ，我们整合 \vec{c}_t 和 \bar{c}_t 获得记忆单元表示 c_t ，整合 \vec{h}_t 和 \bar{h}_t 获得隐藏层表示 h_t 。由于序列化的 LSTM 网络在计算当前时刻的信息时传递了历史信息，因此，同一个单词在不同的上下文中，会得到不同的词向量，从而可以学习到基于上下文信息的词语特征表示，参与基于句法结构的语义组合计算。

2.5 预测层

参考前人的工作^[11,13]，我们采用了平均池化和最大池化操作分别整合 P 和 Q 对应句法树中节点的信息，从而得到一个长度固定的语义向量，计算公式如 (16) - (18) 所示。

$$V_{p_{mean}} = \frac{1}{2m-1} \sum_{i=1}^{2m-1} h_{p_i}; V = \max_{i=1}^{2m-1} h_{p_i} \quad (16)$$

$$V_{q_{mean}} = \frac{1}{2n-1} \sum_{j=1}^{2n-1} h_{q_j}; V = \max_{j=1}^{2n-1} h_{q_j} \quad (17)$$

$$V = [V_{p_{mean}}; V_{p_{max}}; V_{q_{mean}}; V_{q_{max}}] \quad (18)$$

最后，我们将语义向量 V 输入到多层感知机分类器，预测两个句子互为复述的概率 $\Pr(\bullet | P, Q)$ 。

$$\Pr(\bullet | P, Q) = \text{soft max}(W^2 \text{ReLU}(W^1 V + b^1) + b^2) \quad (19)$$

3 模型训练和实施细则

我们利用上面提出的基于句法结构的语义计算方法，搭建神经网络复述识别模型，模型的训练目标和实施细则描述如下。

首先，我们利用端到端的方法训练整个模型。目标函数采用交叉熵损失函数，添加 l_2 正则化项防止过拟合，通过超参数 λ 调整目标函数权重：

$$\ell(\theta) = -\frac{1}{N} \sum_{i=1}^N \log \Pr(y^{(i)} | P^{(i)}, Q^{(i)}) + \frac{1}{2} \lambda \|\theta\|_2^2 \quad (20)$$

其中 θ 表示模型中所有的参数， N 表示训练集大小， $(P^{(i)}, Q^{(i)})$ 表示句对， $y^{(i)}$ 是相应句对的标签， $y^{(i)}=1$ 表示句对间的语义关系是复述， $y^{(i)}=0$ 表示句对间的语义关系是非复述。

其次，关于输入词向量表示，本文中，我们采用三种类型的向量表示当前输入单词 x_i ：预训练的词向量 v_i^{pre} 、可学习的词向量 v_i^{learn} 和词性标签向量 v_i^{pos} 。其中，和已有工作一致^[11-13]，预训练的词向量是 Glove 840B^[21]。最后，我们采用一个非线性变换获得三个部分整合的词向量表示 x_i ：

$$x_i = \text{ReLU}(W^e [v_i^{pre}; v_i^{learn}; v_i^{pos}] + b^e) \quad (21)$$

最后，关于模型实施具体细节描述如下。我们设置预训练的词向量为 300 维，可学习的词向量为 30 维，词性标签向量为 50 维，所有的 ReLU 层为 300 维。设置 TreeLSTM 结构的隐藏层为 300 维。在实验中，预训练的词向量保持固定不变。我们采用 Adam^[22] 优化算法，设置初始学习率大小为 $2e-4$ ， β_1 为 0.9， β_2 为 0.999。在每一轮迭代中，学习率以 0.95 的频率衰减。设置 l_2 正则化强度为 $1e-6$ ，训练 batch 的大小为 64。为了防止过拟合，我们使用了 dropout^[23] 正则化技术，设置词向量输入层的 drop 率为 0.2，ReLU 层输出层的 drop 率为 0.07。我们对预训练的词向量和最终整合的词向量使用了 batch normalization^[24] 正则技术。其中，dropout 层作用在 batch normalization 层之后。

4 评测实验与结果分析

为了验证本文所提方法的有效性，我们在英语复述识别数据 Quora 上进行了评测实验。

4.1 实验数据

本文使用英语复述识别数据集 Quora 作为实验数据，我们采用了标准的数据划分^[9]规则。原始数据集

没有句法标注信息，因此我们采用伯克利句法分析器¹预处理数据，获取句子的短语句法结构树。我们也采用了斯坦福句法分析器获取句法结构，但不同的句法分析器得到了相近的复述识别精度，可能的原因是这些句法分析器本身有相近的句法分析精度。表 1 给出了 Quora 英语复述数据的统计信息。

表 1 Quora 英语复述数据集统计信息

Table 1 Statistical information of Quora English paraphrase data

数据集	训练集	开发集	测试集	词语数	平均句长	
					P	Q
Quora	384K	10K	10K	107K	12	12

4.2 评测实验结果

和先前工作一致^[11-13]，我们使用准确度评价模型性能，利用开发集选择性能最好的模型参数用于测试阶段。我们和已有公开发表的基于神经网络的复述识别系统进行了性能比较，比较结果如表 2 所示。

表 2 和已有基于神经网络的复述识别系统的性能比较结果

Table 2 Experiment results on paraphrase identification compared with previous models

系统	测试集
Siamese-CNN ^[11]	79.60
Multi-Perspective-CNN ^[11]	81.38
Siamese-LSTM ^[11]	82.58
Multi-Perspective-LSTM ^[11]	83.21
L.D.C ^[11]	85.55
BiMPM ^[11]	88.17
DIIN ^[12]	89.06
AF-DMN ^[13]	88.72
本文提出的模型	89.30

从表2的结果可以看出，相比于文献^[11]提出的基于序列化结构BiLSTM语义组合计算和基于神经注意力的词对齐复述识别系统BiMPM，我们的模型在准确度取得了1.13点的提高。相比于文献^[12]提出的基于CNN模型的语义表示和融合方法DIIN，我们的模型提高了0.24个点。相比于文献^[13]提出的基于序列化结构的多层神经注意力机制的复述识别系统AF-DMN，我们的模型提高了0.58个点。和已有基于序列化结构的复述识别模型的性能比较表明，本文提出的基于句法结构的神经网络复述识别模型，采用树结构的神经网络模型进行语义组合计算，学习短语级的语义表示和对齐，对于改进复述识别系统的有效性。

4.3 模型各个部分有效性分析

我们分析了模型中各个部分对最终复述识别精度的影响，实验结果展示在表3中。

表3 模型各个部分有效性分析结果

Table 3 Effect analysis of each component

系统	开发集	测试集	参数大小
(1) 无注意力机制	86.2	86.1	4.1M
(2) +跨句子注意力机制	89.2	88.8	4.6M
(3) +自注意力机制	89.5	89.3	5.1M

从表3的实验结果来看，基于短语级语义表示的注意力机制通过学习两个句子间短语级的语义匹配信息，有效改进了复述识别的性能。模型（1）没有使用注意力机制，未考虑句子间的短语对齐信息，仅使用句法树根节点的语义表示作为整个句子的表示，复述识别精度达到 86.1%。模型（2）加入了基于短语

¹ <https://github.com/slavpetrov/berkeleyparser>

级语义表示的跨句子特征，利用注意力机制学习两个句子基于句法树的语义对齐关系，在准确度上提高了2.7个百分点，这表明引入跨句子的短语级特征对复述识别十分重要。模型（3）进一步加入了自注意力机制，可以捕获更多句法上下文信息增强语义表示，准确度达到89.3%的精度，进一步改进了复述识别的性能。相比于未使用句法结构的复述识别模型，文献^[13]提出的基于序列化结构的多个跨句子和自注意力层，我们基于句法结构的模型仅使用了一个跨句子注意力层和一个自注意力层，含有更少的参数，但取得了更高的精度。最终，实验结果表明，本文提出基于句法结构进行语义组合计算的模型，以及基于短语表示的跨句子注意力机制和自注意力机制，可以有效学习语义特征，提高复述识别的精度。

4.4 实例分析

为了进一步验证模型对短语的表示和对齐能力，我们进行了可视化分析，给定互为复述的句对{*P*:“a molecule of water is two hydrogen atoms and one oxygen atom.”和 *Q*:“2 hydrogen, 1 oxygen atoms combine to make up a molecule of water.”}。图3给出了对应的句法树，图4给出了两个句子间的句法树对齐结果。从图3和4中可以看到，尽管两个句子句法结构差异较大，模型仍有效捕获了该句对间的短语对齐信息，如：*P*中短语节点21“two hydrogen atoms and one oxygen atom”和*Q*中短语节点11“2 hydrogen, 1 oxygen atoms”在语义表示上对齐概率较大。可视化的分析结果证明了，本文所提出的基于句法结构的语义组合计算方法和基于短语级语义表示的对齐方法，可以有效学习跨句子语义特征，从而改进复述识别性能。

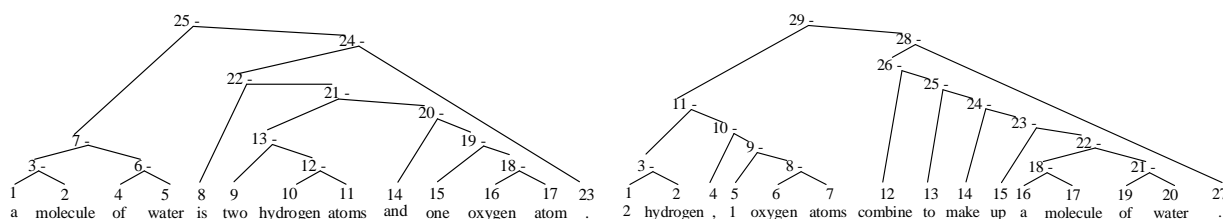


图3 复述句对 *P* 和 *Q* 的短语句法树

Fig. 3 The syntactic trees of paraphrase sentences *P* and *Q*

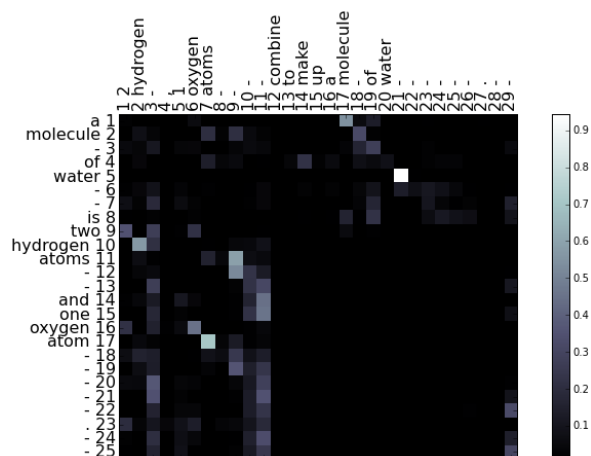


图4 互为复述句对的短语级对齐结果

Fig.4 The phrase-level attention results for paraphrase sentences *P* and *Q*.

5 结语

本文提出了一个基于句法结构的神经网络复述识别模型，探索了句法结构信息在神经网络语义组合计算方面的有效性，扩展了基于序列化结构的词语级语义表示到基于句结构的短语级语义表示。同时，设计基于短语级语义表示的注意力机制建模跨句子语义信息，引入自注意力机制建模基于句法的上下文信息，有效结合了句子内部和句子间的特征用于复述识别。实验结果显示本文所提模型明显提高了复述识别的精度。针对未来的研究工作，我们需要探索模型深度在改进短语级语义匹配方面的有效性。同时，需要探索在基于句法结构的模型中融入预训练语言模型的技术，如 BERT^[25]，以进一步改进复述识别性能。

参考文献

- [1] 赵世奇,刘挺,李生. 复述技术研究[J]. 软件学报,2009,08:2124-2137.
- [2] Callisonburch C , Koehn P , Osborne M . Improved statistical machine translation using paraphrases[C]// Proc Human Language Technology Conference-north American Chapter of the Association for Computational Linguistics. 2006.
- [3] Liang Zhou, Chin-Yew Lin, and Eduard Hovy. 2006. Re-evaluating Machine Translation Results with Paraphrase Support. In Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing (EMNLP '06). Association for Computational Linguistics, Stroudsburg, PA, USA, 77–84.
- [4] S. Zhao, M. Zhou, and T. Liu. Learning Question Paraphrases for QA from Encarta Logs[C]. Proceedings of IJCAI. 2007:17961800
- [5] Regina Barzilay and Kathleen R McKeown. 2001. Extracting paraphrases from a parallel corpus. In Proceedings of the 39th annual meeting on Association for Computational Linguistics. Association for Computational Linguistics, 50–57.
- [6] Inderjeet Mani. 1999. Advances in automatic text summarization. MIT press.
- [7] Rada Mihalcea, Courtney Corley, Carlo Strapparava, et al. 2006. Corpus-based and knowledge-based measures of text semantic similarity. In AAAI, Vol. 6. 775–780.
- [8] Zornitsa Kozareva and Andrés Montoyo. 2006. Paraphrase identification on the basis of supervised machine learning techniques. In Advances in natural language processing. Springer, 524–533.
- [9] Finch A, Hwang Y S, Sumita E. Using machine translation evaluation techniques to determine sentence-level semantic equivalence[C]//Proceedings of the Third International Workshop on Paraphrasing (IWP2005). 2005..
- [10] 吴晓锋, 宗成庆. 基于语义角色标注的新闻领域复述句识别方法[J]. 中文信息学报, 2010, 24(5): 3-10.
- [11] Zhiguo Wang, Wael Hamza, and Radu Florian. Bilateral multi-perspective matching for natural language sentences. corr abs/1702.03814, 2017.
- [12] Yichen Gong, Heng Luo, and Jian Zhang. Natural language inference over interaction space. arXiv preprint arXiv:1709.04348, 2017.
- [13] Chaoqun Duan, Lei Cui, Xinchu Chen, FuruWei, Conghui Zhu, and Tiejun Zhao. Attention-fused deep matching network for natural language inference. In IJCAI, pages 4033–4040, 2018.
- [14] 李天时, 李琦, 王文辉, 等. 基于外部记忆单元和语义角色知识的文本复述判别模型[J]. 中文信息学报, 2017, 31(6): 33-40.
- [15] 刘明童, 张玉洁, 徐金安, 等. 开放域上基于深度语义计算的复述模板获取方法[J]. 中文信息学报, 2018, 32(2): 94-101.
- [16] Fan M, Lin W, Feng Y, et al. A Globalization-Semantic Matching Neural Network for Paraphrase Identification[C]//Proceedings of the 27th ACM International Conference on Information and Knowledge Management. ACM, 2018: 2067-2075.
- [17] Yin W, Schütze H, Xiang B, et al. Abcnn: Attention-based convolutional neural network for modeling sentence pairs[J]. Transactions of the Association for Computational Linguistics, 2016, 4: 259-272.
- [18] Tai K S, Socher R, Manning C D. Improved semantic representations from tree-structured long short-term memory networks[J]. arXiv preprint arXiv:1503.00075, 2015.
- [19] Timothy Dozat and Christopher D. Manning. Deep biaffine attention for neural dependency parsing[J]. arXiv preprint arXiv:1611.01734, 2016.
- [20] Schmidhuber J, Hochreiter S. Long short-term memory[J]. Neural Comput, 1997, 9(8): 1735-1780.
- [21] Jeffrey Pennington, Richard Socher, and Christopher Manning. 2014. Glove: Global vectors for word representation. In Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP), pages 1532–1543.
- [22] Kingma D P, Ba J. Adam: A Method for Stochastic Optimization[J]. arXiv preprint arXiv:1412.6980, 2014.
- [23] Srivastava N, Hinton G, Krizhevsky A, et al. Dropout: a simple way to prevent neural networks from overfitting[J]. Journal of Machine Learning Research, 2014, 15(1):1929-1958.
- [24] Ioffe S, Szegedy C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift[J]. 2015:448-456.
- [25] Devlin J, Chang M W, Lee K, et al. Bert: Pre-training of deep bidirectional transformers for language understanding[J]. arXiv preprint arXiv:1810.04805, 2018.