doi: 10.6043/j.issn.0438-0479.201701010

**基于混合树结构神经网络的**

**隐式篇章关系识别**

郑江龙，陈锦秀\*

(厦门大学信息科学与技术学院，福建 厦门 361005)

**摘要**: 隐式篇章关系识别的主要挑战是如何表示两个文本单元的语义信息。将句子中的单词转换为具有语义信息的单词向量并根据句子语法结构组合单词向量获取句子语义信息。而句子的语义信息往往由语法树中的信息焦点(谓词部分)所决定，所以关注这些信息焦点可以提升篇章关系识别的效果。为了增强信息焦点的作用，引入树状长短时记忆 (tree-structured long short-term memory, Tree-LSTM) 网络，使用其遗忘门(forget gate)的特性区别对待不同孩子节点的信息。最后利用神经张量网 (neural tensor network, NTN)来计算两个句子语义向量之间的关系。基于PDTB2.0（Penn Discourse Treebank）语料数据进行实验，实验结果表明混合树结构神经网络比传统模型在F-score上提高了3.0%。

**关键词**： 隐式篇章关系识别; 信息焦点；树状长短时记忆网络；神经张量网

**中图分类号：**TP 391 **文献标志码：** A

篇章关系识别是自动识别两个文本单元之间结构和语义的逻辑关系，文本单元可以是子句、句子或段落。该任务在自然语言处理的很多领域有着十分重要的应用，如机器理解[1]、问答系统[2]、文本摘要[3]等。根据文本单元之间是否存在连接词，篇章关系识别任务分为显示和隐式篇章关系识别。显示篇章关系中存在连接词，所以比较容易识别, 而隐式篇章关系缺乏明确的连接词，识别难度大。隐式篇章关系识别任务是利用相关技术识别不存在连接词的两个文本单元的篇章关系。

最初为了克服缺乏人工标注数据的问题,一些学者如March[4]、Saito[5]和Blair-Goldensohn[6]等从生语料库中获取显示篇章关系单元，并去除其连接词来合成隐式篇章关系数据，进而用这些数据研究隐式篇章关系。显示篇章关系的连接词本身蕴含着对关系类别的判定属性，而隐式篇章关系的识别则更多依赖于上下文的语义知识，所以通过人工合成数据训练的模型与通过自然数据训练的模型仍然存在很大的差异。

[[1]](#footnote-1) 随着标注语料库的出现，许多基于特征提取的有监督学习方法和半监督学习方法相继被提出并应用于隐式篇章关系识别。典型的语料库有修饰结构理论篇章树库(RST discourse treebank, RST-DT)[7]、宾州篇章树库(penn discourse treebank, PDTB)[8]和图库 (GraphBank)[9]。这些研究学者主要集中于使用传统的特征提取方法研究隐式篇章关系识别，前人通过研究和探索积累了丰富的特征[10-20]。2009年Pitler等[10]收集了一些语言特征, 如词极特征等;Lin等[11]引入了带有语义特性的依赖式规则和产生式规则;Brain等[18]对过去常用的单词对特征进行改造，从Gigaword语料库中对单词对进行统计得到一个更加密集的聚类单词对特征。国内也有一些研究者基于特征提取方法运用监督学习方法识别隐式篇章关系 [21-25]。2014年刘初等[23]提出基于自训练的半监督策略来识别隐式篇章关系，并对常用的9种篇章关系特征进行组合分析。前人的工作已经证实了现有特征具有很好的效果，然而特征提取方法需要投入大量的人力成本，同时特征提取方法难以获得句子焦点信息。本研究从特征集合中抽取部分特征构建baseline系统，同时从中选出部分特征提升混合树模型识别效果。

 深度学习在图像和语音领域得到广泛应用并取得显著效果，Rumelhart等 [26]首次提出了词向量，利用大量无标注数据训练得到具有语义信息的单词向量，这也是统计自然语言领域的重要创举。为了处理复杂的任务，需要训练能够捕获语法、语义甚至语用等信息的句子向量。关于这方面的模型可以分为三类：词袋模型[27]、序列模型[28]、树结构模型[29-30]。其中：1）词袋模型是将句子中所有连续的单词对应的单词向量相加求得平均值作为句子向量表示，这种方法忽略了句子中单词的次序和语法结构等;2）序列模型经常选用递归神经网络(recursive neural network , RNN) 来处理任意长度的单词序列，这种模型可以捕获到单词的序列语义；3）树结构模型根据句子的语法结构来构建树结构的神经网络捕获句子的语义、语法、单词序列等信息。例如Socher等[29]采用了广义的RNN神经网络构建一棵语法树，学习短语的特征向量。Tai等[30]将LSTM推广到树形神经网络，即提出Tree-LSTM。他们的实验表明Tree-LSTM在情感分析任务中效果比RNN提升很多。Tree-LSTM设置了灵活的遗忘门非常适合增强语法结构中信息焦点的作用。最近研究学者Ji等[31]利用Socher等[29]提出的RNN模型结合句子语法解析树构建神经网络，将单词转换为单词向量作为输入，自底向上计算最终获取两个文本单元的句子向量，最后将两文本单元的句子向量作为特征使用。即使Ji等[31]增强了实体向量的关注，也只比基准实验提高了0.6%，因此实体不是影响篇章关系的主要因素。每个句子中都有一些特定的信息焦点对整个句子的语义起着重要作用。在语法中，谓词一般是句子的信息焦点，它往往决定了整个句子的主题含义。但是Ji等[31]使用的RNN模型同等对待语法树中所有孩子节点信息，因此不能有效获取信息焦点和句子真实语义，进而影响篇章识别效果。同时RNN在处理长序列时性能较差，会导致反向梯度信息丢失。因此本研究提出基于混合树结构神经网络的隐式篇章关系识别模型，该模型组合了从大量无标注语料训练得到的单词向量、用于学习句子向量的Tree-LSTM以及用于计算句子向量之间关系的神经张量网(neural tensor network, NTN)。该混合树结构神经网络模型可以通过模仿人类大脑方式学习语法知识表示，同时能够自动识别隐式篇章关系。基于PDTB语料库进行试验，实验结果表明该模型比RNN模型识别效果更好，比传统模型的F-score提升了3%，同时也证实了句子中信息焦点对隐式篇章识别是非常重要的。

#  1模型的提出

## 1.1 混合树结构模型

每个句子中都有一些特定的信息焦点对整个句子的语义起着重要作用。在语法中，谓词一般是句子的信息焦点，它往往决定了整个句子的主体含义。谓词在句子中有特定的表现形式，如连系动词+表语构成复合谓词出现在句子中。例1中“my favorite pet”和“a cat in the pan”就是表语性的名词短语，与前面的系动词在句子中共同充当谓语，影响整个句子意思；或者通过系动词+形容词构成形容词谓语，例2中“kind”是一个形容词表语，与系动词“was”一起构成形容词谓语。

例1 Arg1: Ada is my favorite pet.

 Arg2: It is a cat in the pan.

例2 Arg1: Susan gave Jack the coat.

 Arg2: She was kind.

成分树是通过将句子不断按主语-谓语的方式分裂，得到一棵解析树。因此，使用成分树结构构建神经网络来计算两个文本单元的句子向量时，能够更加方便增强谓语在整个句子中的作用。进一步地说成分树可以提高动词表语、名词表语或形容词表语对句子向量的影响程度。如图1中虚线框住的信息是两个文本单元中需要关注的部分，这部分对句子语义起着重要作用。因此，若构建一个知识向量模型能够区别对待句子中的信息焦点和非信息焦点，则能够准确地获取句子的语法语义信息进而提高篇章关系的识别效果。为了解决上述问题，本研究引入Tree-LSTM神经网络来获取每个篇章单元的语法语义含义。Tree-LSTM模型利用树状结构通过组合输入单词向量和孩子单元获取具有语法语义特性的句子向量。



图 1 解析树中的信息焦点

Fig 1 An example of information focus in the parse tree.

另外，本研究还采用了NTN模型计算两个文本单元句子向量之间的关系。与标准的NTN模型不同，本研究对该模型做了一些修改，使得NTN不再单纯依赖于两个文本单元的句子向量，而是可以融入特征向量提升篇章识别的效果。

图2是基于混合树结构神经网络的隐式篇章关系识别模型，图中Arg1和Arg2分别表示隐式篇章关系候选实例中的两个文本单元。第一层是将Arg1 和Arg2中的单词转换为低维单词向量，一般单词向量可以随机初始化或者利用大量无标准语料库通过神经网络学习得到。因为这个不是本论文的研究重点，所以使用Lai[33]训练好的word2Vec作为初始单词向量。该单词向量作为Tree-LSTM模型的输入，计算句子向量。后续章节将详细介绍该神经网络如何能够学习良好的句子向量和进行篇章关系识别。



图2 基于深层语义神经网络的隐式篇章关系识别模型

Fig 2 The framework of our proposed hybrid tree structured neural network for implicit discourse relation recognition

## 1.2 句子向量的语义获取

采用Tree-LSTM神经网络作为句子语义学习模型，以单词向量作为输入并自底向上学习得到句子向量。在句子语法解析树中非叶子结点的值是以其孩子节点和单词向量作为输入计算得到。相比于RNN神经网络，Tree-LSTM能够解决两方面问题：首先引入了记忆单元能够解决长序列梯度信息反向传播的丢失问题；另外加入了遗忘门能够选择性地组合孩子节点信息，因此可以加强信息焦点在整个句子中的作用。

如图3所示，一个Tree-LSTM单元*j*含有一个记忆单元***c****j*、隐含状态***h****j*、输入门***i****j*、输出门***o****j*和每个孩子*k*对应的遗忘门***f****jk*。其中，***x****j*是单词*j*对应的单词向量，作为Tree-LSTM的输入。调节每个孩子节点*k*遗忘门***f****jk***,**的值，在篇章关系识别中有利于关注信息焦点。如图1中，在成分树中左边的孩子节点是名词，右边孩子节点是动词短语对整个句子语义起着关键作用，所以可以加大右孩子节点在句子的影响程度。在模型中减小***f****j*1并加大*fj*2。



图3 Tree-LSTMs单元结构

Fig 3: A Binary Tree-LSTM unit: updating the memory cell c1 and hidden state h1 with two children

这里利用成分树来构建二元Tree-LSTM网络，得到一个获取句子语义的模型。对于任意一个二元Tree-LSTM的节点*j，****c****jk****、h****jk*分别表示其第*k*个孩子节点的记忆单元和隐含状态，*k*取值为1或2，从而可得二元Tree-LSTM网络的转换公式如下:

$i\_{j}=σ(W^{(i)}x\_{j}+\sum\_{k=1}^{2}U\_{k}^{\left(i\right)}h\_{jk}+b^{(i)})$

$f\_{jk}=σ(W^{(f)}x\_{j}+\sum\_{l=1}^{2}U\_{kl}^{\left(f\right)}h\_{jl}+b^{(f)})$

$o\_{j}=σ(W^{(o)}x\_{j}+\sum\_{k=1}^{2}U\_{k}^{\left(o\right)}h\_{jk}+b^{(o)})$

$u\_{j}=tanh(W^{(u)}x\_{j}+\sum\_{k=1}^{2}U\_{k}^{\left(u\right)}h\_{jk}+b^{(u)})$

$c\_{j}=i\_{j} \bigodot\_{}^{}u\_{j}+\sum\_{k=1}^{2}f\_{jk}⨀ c\_{jk}$

$h\_{j}=o\_{j}\bigodot\_{}^{}tanh⁡(c\_{j})$

其中，上式中的函数$σ$是sigmoid函数；$b^{(i)}$**、**$b^{(f)}$**、**$b^{(o)}$**、**$b^{(u)}$表示偏置项；$U\_{k}^{\left(i\right)}$**、**$U\_{kl}^{\left(f\right)}$**、**$U\_{k}^{\left(o\right)}$**、**$U\_{k}^{\left(u\right)}$表示孩子节点隐含值的权重；$W^{(i)}$**、**$W^{(f)}、W^{(o)}、W^{(u)}$分别表示输入向量在不同结构中的权重。$i\_{j}$**、**$o\_{j}$**、**$c\_{j}$分别代表Tree-LSTM中第*j*个节点计算得到的输入门、输出门、记忆单元的值。$f\_{jk}$表示第*j*个节点的第*k*个孩子节点的遗忘门值。模型中每个单元节点最多只有两个孩子节点，当计算模型中的任何一个单元时，其两个孩子节点分别采用不同的参数矩阵，以实现对其孩子节点的更加细粒度的调节，例如，左孩子节点对应的是名词，右孩子节点对应动词，那么通过训练可以增大右孩子节点的权重使模型更加关注动词信息。从模型空间复杂度分析，其参数数量规模仅为*O*(4|*h*|2)，|*h*|表示结点隐含状态的向量长度，因而该模型体系具有一定的可行性。最后我们将通过二元Tree-LSTM模型逐步训练得到的两个句子向量传递给NTN模型作为输入进一步识别两个文本单元的篇章关系。

##  隐式篇章关系的识别

 NTN模型是Socher[32]提出并成功地应用在实体关系的识别中, 受该工作的启发，将两个文本单元的句子向量(***s***1*,* ***s***2)作为输入，使用NTN判断两个文本单元之间是否存在某种隐式篇章关系*r*及对应的概率大小。如图4所示，直观地显示了NTN的结构，NTN中引入了一个二元线性张量积将2个句子单元用矩阵相乘的方式紧密关联在一起，代替了传统神经网络直接将两个输入向量首尾相连的方式。根据下面的NTN模型公式，可以计算两文本之间的篇章关系属于关系的概率值。



图4 神经张量网络NTN的结构视图

Fig. 4 Visualization of Neural Tensor Network

 for discourse relation Classification.

 $g\left(s\_{1},r,s\_{2}\right)=U\_{R}ρ(s\_{1}^{T}W\_{r}^{\left[1:K\right]}s\_{2}+V\_{r}\left[\genfrac{}{}{0pt}{}{s\_{1}}{s\_{2}}\right]+b\_{r})$ (1)

 上式中的$ρ$是指非线性函数tanh，$W\_{r}^{[1:K]}\in R^{d×d×k}$表示张量。而公式(1)中的$s\_{1 }^{T}W\_{r}^{[1:k]}s\_{2}$是一个二元线性张量积其结果是一个向量$v\in R^{k}$, 向量***v***中每个值都是通过一个切片计算得到，$v\_{i}=s\_{1 }^{T}W\_{r}^{[1:k]}s\_{2}$, *i=* 1*,…,k*。对于每一个篇章关系*r*，NTN网络包含的参数有：$U\_{r}\in R^{k}$、$W\_{r}^{[1:k]}$、$V\_{r}\in R^{k×2d}$和$b\_{r}\in R^{k}$。

 $g\left(s\_{1},r,s\_{2}\right)=U\_{r}ρ(s\_{1 }^{T}W\_{r}^{[1:k]}s\_{2}+V\_{r}\left[\genfrac{}{}{0pt}{}{s\_{1}}{s\_{2}}\right]+M\_{r}F+b\_{r})$ (2)

 公式（2）是NTN函数的改进，使得NTN模型可以引入特征向量。$M\_{r}$表示特征的参数，$F$是特征向量。因此利用该改进的NTN，可以将近年来研究的特征加入到模型中，进一步提高识别效果。根据相关工作的研究成果，模型中加入了有效的特征：单词对特征、依赖树规则、产生式规则和布朗聚类.

 模型的最顶层采用softmax 分类器来识别两个文本单元的篇章关系。如果在PDTB四个隐式篇章关系中属于关系(***s***1, *r*,***s***2)的概率最大，则文本单元之间的关系定义为*r*，否则不属于*r*。当使用两种隐式篇章关系做分类时，多元分类器会变为的二元分类器。

## 1.4 训练目标

训练目标是使得正确篇章关系具有更大概率。模型的所有参数设为$θ$，所有参数需要通过学习得到。这些参数一部分来自NTN模型:{***U****r*，***W****r*，***V****r***，*b****r*}. 另一部分参数来自Tree-LSTM:{***b***(*o*)，***W***(*o*)，$U\_{l}^{(o)}$，***b***(*i*) ，***W***(*i*)，$U\_{l}^{(i)}$，***b***(*f*) ，***W***(*f*) ，$U\_{l}^{(f)}$}*。*$g\_{r}^{(i)}=g(s\_{i}^{\left(i\right)},r,s\_{2}^{(i)})$表示第*i*个文本单元对或样本通过混合树结构神经网络模型计算得到在关系*r*上的分值。通过最小化下面的目标函数来训练模型：

$$J\left(θ\right)=-\frac{1}{N}[\sum\_{i=1}^{N}\sum\_{r}^{L}1\left\{y^{\left(i\right)}=r\right\}log\frac{e^{g\_{r}^{\left(i\right)}}}{\sum\_{r}^{R}e\_{r}^{\left(i\right)}}]+λ\left‖θ\right‖\_{2}^{2}$$

式中：*N*表示训练数据的数量；***L***表示隐式篇章关系的集合，本研究主要基于PDTB第一层关系，因此***L***={Expansion、Temporal、Contingency、Comparison}；*r*表示关系集***L***中的第*r*个隐式篇章关系，如：Expansion；*y(i)*表示样本*i*在语料库中对应的关系；$λ$是正规项的参数以避免模型出现过度拟合。

# 2实 验

## 2.1 数据设置

为了验证基于深层语义神经网络的隐式篇章识别关系的模型的有效性，混合树结构神经网络模型使用PDTB语料库作为训练和测试数据。与大部分研究学者一样主要研究PDTB中的第一层关系，包括: Expansion、Temporal、Contingency 和Comparison。另外在PDTB有2.2%的隐式篇章关系实例被标注为两种以上的关系，将这些实例在PDTB语料库中的第一关系作为其隐式篇章关系。例如文本实例被认为具有Temporal和Contingency关系，则定义该实例的篇章关系为Temporal。所有隐式篇章关系数据的02-21章节作为训练数据，00-01章节作为验证数据，22-23章节的数据作为测试用例。表1显示了训练数据、测试数据和验证数据在每种隐式篇章关系中的分布情况。与Pitler等[10]和Ji等[31]的相同，实验将EntRel关系的数据当做Expansion关系数据。表1的Exp+Ent表示Expansion和EntRel数据分布情况。

表 1 训练数据和测试数据设置

Tab. 1 The number of train data, Dev data and

test data for each relation type in PDTB corpus.

|  |  |
| --- | --- |
|  | 数据量 |
| Exp+Ent | Exp. | Cont. | Comp. | Temp. |
| 训练集 | 10 806 | 6 878 | 3 183 | 1 883 | 686 |
| 验证集 | 1 145 | 656 | 279 | 193 | 55 |
| 测试集 | 980 | 562 | 269 | 145 | 70 |

模型的训练算法使用自适应梯度下降算法(AdaGrad) [34]，并设置学习速率为0.05,最小批次为100，正规化项的参数值设为10-4。通过验证集验证后，将NTN模型的切片设置为3，单词向量的维度设置为50，Ji等[31]的模型的单词向量在这个维度下也是效果最好的。为了避免训练数据中正样本与负样本数据的不平衡性问题，随机从正样本集合中选择正样本继续添加到正样本集合中使得训练过程正样本和负样本具有相同比例。

##  与基准系统的比较

 之前大多数相关研究是基于特征工程和浅层统计分析，为了评估模型有效性，首先构建了两个传统学习模型作为基准系统。从两个文本单元中提取有效特征，这些特征在前人的工作中已经被证明是有效的。这些特征包括: First-Last-First3、Inquirer Tags、Production Rules、 Dependency Rules、 Polarity、Verbs、Modality、NER 和 Unigram。利用最大熵模型进行有监督学习并利用协同训练算法训练半监督学习模型。表2中的第1行、第2行展示该基准系统的结果。混合树结构神经网络模型与基准系统的比较结果可以看出混合树神经网络模型比基准提升了很多。

Tree-LSTM模型通过成分树构建网络模型有助于学习两文本单元的句子向量。RNN是采用Socher等[29]提出的迭代神经网络。为了验证Tree-LSTM模型能更好地捕获两文本之间的语义信息，将Ji等[31]使用到的RNN模型作为基准系统和混合树神经网络模型进行比较。表2中的第3行和第4行显示了RNN模型和Tree-LSTM的结果，可以看出在其他层设置都一样的情况下，使用Tree-LSTM作为句子向量捕获的模型效果比RNN模型更好。该结果也进一步验证了利用Tree-LSTM关注句子中的信息焦点能够找到两个篇章关系中的语义链，这也是Tree-LSTM效果好于RNN的原因。

表 2 与基准系统的比较

Table 2 Comparison results with various baseline systems.

|  |  |
| --- | --- |
| **方法** | **F-score/%** |
| **Exp+Ent** | **Exp.** | **Cont.** | **Comp.** | **Temp.** |
| Surface features + Maximum Entropy  | 66.67 |  | 49.80 | 33.60 | 24.32 |
| Surface features + Co-training | 66.80 |  | 50.07 | 33.99 | 23.37 |
| RNN + NTN | 77.87 | 59.50 | 44.98 | 29.53 | 23.26 |
| TreeLSTM + NTN | 79.22 | 66.09 | 45.10 | 28.28 | 30.03 |
| RNN + NTN + Surface features + Brown Cluster | 78.73 | 66.62 | 50.93 | 31.21 | 29.14 |
| TreeLSTM + NTN + Surface features + Brown Cluster | **80.19** | **69.81** | **53.47** | **35.03** | **30.21** |

为了进一步比较，在混合树结构神经网络模型中引入了与Ji等[31]相同的特征, 不仅包含了Lin等[11]提出的一些特征，还包含了Rutherford[20]等提出的Brown cluster特征。模型使用了与Ji等[31]一样的特征：500个单词对特征、600个布朗聚类特征、100个依赖树规则特征和100个产生式规则特征。如表2的第五行与第六行所示，引入了特征以后混合树结构的神经网络模型在Expansion、Contingency、Comparison和Temporal四个篇章关系中相比RNN模型基准系统F值提高了3%左右。表2中的Exp+Ent表示将EntRel和Expansion实例共同作为Exp实例，Exp表示单独使用Expansion实例作为正例。

##  与其他系统的比较

表3 与其他有监督和半监督学习模型的比较

Table 3: Comparison Results with the competitive systems.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **方法** | **Exp.** | **Cont.** | **Com.** | **Temp.** |
| **F-score** | **准确率** | **F-score** | **准确率** | **F-score** | **准确率** | **F-score** | **准确率** |
| Pitler等[10] | 76.42 | 63.62 | 47.13 | 67.30 | 21.96 | 56.59 | 16.76 | 63.49 |
| Zhou等[12] | 70.11 | 54.54 | 47.16 | 48.96 | 31.79 | 58.22 | 20.30 | 55.48 |
| Park等[17] | 79.22 | 69.14 | 49.82 | 72.09 | 31.32 | 74.66 | 26.57 | 79.32 |
| Biran等[18] | 75.87 | 62.84 | 46.94 | 68.09 | 25.40 | 63.36 | 20.23 | 68.35 |
| Ji等[31] | 80.02 | **69.80** | 52.78 | **76.95** | **35.93** | 70.27 | 27.63 | 87.11 |
| 本研究 | **80.19** | 66.94 | **53.47** | 68.55 | 35.03 | **78.01** | **30.21** | **90.73** |

如表3所示，该表列出了前人的研究结果：F-score和准确率。“Exp.”表示使用了EntRel和Expansion共同作为正例。这里只列出了使用相同的语料库PDTB、相同的训练、验证和测试集的参考文献的结果, 如Pitler[10]、Zhou[12]、Park[17]、Biran[18]、Ji[31]等。国内一些研究学者的模型比较方式和数据分布与本文不同，因此没有加入对比。在这些模型中, Ji[31]的模型在Expansion、Contingency、Comparison和Temporal四个分类中效果都是最好的。与Ji[31]的模型比较，我们的模型在以上四个关系上的F值都有明显的提升。

# 4讨 论

与传统基于统计分析和特征工程的模型不同，混合树结构神经网络模型尝试引入单词向量到隐式篇章识别关系中。相比于少量的人工标注数据，无标注数据大量存在。因此，可以使用自然语言模型从大规模无标注数据中学习得到单词向量。实验使用了Lai等[33]已经训练好的word2Vec作为初始向量。而Ji等[31]是直接采用PDTB语料库来训练word2Vec向量，这种向量本身能够提供更好的篇章关系信息。考虑到单词向量的影响，实验中使用一些不同的单词向量来初始化模型，这些模型在篇章关系识别上进行比较并选择效果较好的单词向量作为初始化向量。

此外，实验表明随着单词维度的增加，篇章识别效果没有得到明显的提升。其中的原因是随着单词维度的增加虽然单词向量的信息会增多，但是有些无用的信息也会随之被引入进来。因此在实验中将单词维度设置为50。

基于Tree-LSTM的模型比基于RNN模型的效果更显著，这是因为Tree-LSTM克服了RNN在长序列中反向传播会出现消失的问题。更重要的是基于成分树的Tree-LSTM模型能够帮助模型区别对待语法结构中的不同成分，增强信息焦点的影响程度，因此可以通过该模型获取到更加有效的语义信息。表2的结果证明了即使没有加入特征，混合树结构神经网路在隐式篇章关系识别上依然效果显著。在表3中与其他模型的比较中，该模型在四种关系中依然能取得一些提升。

 举一些例子说明为什么关注句子的关键信息有利于提升篇章关系识别的效果。

 例 3: Arg1: The estimated improvement in Colgate’s U.S. operations (*took some analysts by surprise*).

Arg2: Colgate’s household products business (*has been a weak performer*).

 例 4: Arg1: His organization urged trucking firms to (*halt all deliveries into the Bay area*) yesterday, except for emergency medical supplies.

Arg2: Some foodstuff shipments will probably *resume* Thursday.

Tree-LSTM模型能够正确识别出例3的隐式篇章关系是Contingency, 一部分原因是该模型能够关注句子中的VP部分“took some analysts by surprise”和“has been a weak performer”，增强该部分的影响程度。相比之下RNN同等对待语法结构中的所有孩子节点，因此不能降低“The estimated improvement”在句子的影响度，这样会导致错误的识别该篇章关系为Comparison. 例4中Tree-LSTM能够通过关注“*halt all deliveries into the Bay area*”和“resume”部分的信息正确识别出篇章关系为Comparison.

# 结 论

混合树结构神经网络模型用于识别隐式篇章关系，该模型是由单词向量、用于获取句子向量的Tree-LSTM以及进行关系识别的NTN组成。该模型可以捕获两个文本单元的语义、语法和词汇信息，这些有助于提升隐式篇章关系识别的效果。实验结果显示了该模型比传统基于特征工程的方法在F-score上提高了3.0%。在未来将进一步验证基于依赖树的Tree-LSTM是否比基于成分树的Tree-LSTM更好。

# 参考文献:

1. NARASIMHAN K, BARZILAY R. Machine comprehension with discourse relations[C]// Association for Computational Linguistics. Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing. Beijing: ACL, 2015:1253-1262.
2. VERBERNE S, BOVES L, OOSTDIJK N, et al. Evaluating discourse-based answer extraction for why-question answering[C]//SIGIR. International ACM SIGIR Conference on Research and Development in Information Retrieval. Amsterdam: SIGIR, 2007:735-736.
3. LOUIS A, JOSHI A, NENKOVA A. Discourse indicators for content selection in summarization[C]// Association for Computational Linguistics. Meeting of the Special Interest Group on Discourse and Dialogue. Uppsala: ACL, 2010:147-156.
4. MARCU D, ECHIHABI A. An unsupervised approach to recognizing discourse relations[C]// Association for Computational Linguistics. Proceedings of the 40th Annual Meeting on Association for Computational Linguistics. Philadelphia: ACL, 2002: 368-375.
5. SAITO M, YAMAMOTO K, SEKINE S. Using phrasal patterns to identify discourse relations[C]// Association for Computational Linguistics. Proceedings of the Human Language Technology Conference of the NAACL. Sydney: ACL, 2006: 133-136.
6. BLAIR-GOLDENSOHN S, MCKEOWN K, RAMBOW O. Building and refining rhetorical-semantic relation models[C]//HLT-NAACL. The Conference of the North American Chapter of the Association for Computational Linguistics: Human Language. Rochester: HLT-NAACL, 2007: 428-435.
7. CARLSON L, MARCU D, OKUROWSKI M E. Building a discourse-tagged corpus in the framework of rhetorical structure theory[C]// Proceedings of the Second SIGdial Workshop on Discourse and Dialogue. Aalborg：Sigdial，2001:1-10.
8. PRASAD R, MILTSAKAKI E, DINESH N, et al. The penn discourse treebank 2.0 annotation manual[J]. Proceedings of Lrec, 2007, 24(1) :2961-2968.
9. WOLF F, GIBSON E. Representing discourse coherence: a corpus-based study[J]. Computational Linguistics, 2005, 31(2): 249-287.
10. PITLER E, LOUIS A, NENKOVA A. Automatic sense prediction for implicit discourse relations in text[C]// Association for Computational Linguistics. Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP. Suntec: ACL, 2009: 683-691.
11. LIN Z, KAN M Y, NG H T. Recognizing implicit discourse relations in the Penn Discourse Treebank[C]// Association for Computational Linguistics. Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing. Suntec: ACL, 2009: 343-351.
12. ZHOU Z M, LAN M, NIU Z Y, et al. The effects of discourse connectives prediction on implicit discourse relation recognition[C]// Association for Computational Linguistics. Proceedings of the 11th Annual Meeting of the Special Interest Group on Discourse and Dialogue. Uppsala: ACL, 2010: 139-146.
13. WANG W T, SU J, TAN C L. Kernel based discourse relation recognition with temporal ordering information[C]//Association for Computational Linguistics. Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics. Uppsala: ACL, 2010: 710-719.
14. HERNAULT H, BOLLEGALA D, ISHIZUKA M. A semi-supervised approach to improve classification of infrequent discourse relations using feature vector extension[C]// Association for Computational Linguistics. Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing. Uppsala: ACL, 2010: 399-409.
15. HERNAULT H, BOLLEGALA D, ISHIZUKA M. Semi-supervised discourse relation classification with structural learning[C]//Springer Berlin Heidelberg. International Conference on Intelligent Text Processing and Computational Linguistics. Berlin Heidelberg: Springer Berlin Heidelberg, 2011: 340-352.
16. LAN M, XU Y, NIU Z Y. Leveraging synthetic discourse data via multi-task learning for implicit discourse relation recognition[C]// Association for Computational Linguistics. Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics. Sofia: ACL, 2013: 476-485.
17. PARK J, CARDIE C. Improving implicit discourse relation recognition through feature set optimization[C]// Association for Computational Linguistics. Proceedings of the 13th Annual Meeting of the Special Interest Group on Discourse and Dialogue. Jeju Island, Korea: ACL, 2012: 108-112.
18. BIRAN O, MCKEOWN K. Aggregated word pair features for implicit discourse relation disambiguation[C]//Association for Computational Linguistics. Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics. Sofia: ACL, 2013: 69-73.
19. FISHER R, SIMMONS R G. Spectral Semi-Supervised Discourse Relation Classification[C]// Association for Computational Linguistics. Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing. Beijing: ACL. 2015: 89-93.
20. RUTHERFORD A, XUE N. Discovering Implicit Discourse Relations Through Brown Cluster Pair Representation and Coreference Patterns[C]// Association for Computational Linguistics. Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistic. Gothenburg: ACL. 2014, 645: 2014.
21. 徐凡, 朱巧明, 周国栋. 基于树核的隐式篇章关系识别[J]. 软件学报, 2013, 24(5): 1022-1035.
22. 洪宇,严为绒,车婷婷,梁颖红,姚建民,朱巧明,周国栋.平行推理机制：一种隐式篇章关系检测方法[J]. 软件学报, 2014, 25(11): 2528-2555.
23. 刘初,陈锦秀,基于组合特征的自训练隐式篇章关系的识别技术[J]. 厦门大学学报(自然科学版), 2014, 53(2): 0438-0479.
24. 李生, 孔芳, 周国栋.基于PDTB体系的隐式篇章关系识别[J]. 中文信息学报, 2016, 30(4): 81-89.
25. 朱珊珊,洪宇,丁思远,严为绒, 姚建民,朱巧明. 基于训练样本集扩展的隐式篇章关系分类[J]. 中文信息学报. 2016, 30(5): 111 – 120.
26. RUMELHART D E, HINTON G E, WILLIAMS R J. Learning representations by back-propagating errors[J]. Cognitive modeling, 1986, 323(6088):533-536.
27. FOLTZ P W, KINTSCH W, LANDAUER T K. The measurement of textual coherence with latent semantic analysis[J]. Discourse processes, 1998, 25(2/3): 285-307.
28. MIKOLOV T. Statistical language models based on neural networks[D]. PhD thesis: Brno University of Technology, 2012.
29. SOCHER R, LIN C C, MANNING C, et al. Parsing natural scenes and natural language with recursive neural networks[C]//ICML. Proceedings of the 28th international conference on machine learning (ICML-11). Bellevue, Washington: ICML. 2011: 129-136.
30. TAI K S, SOCHER R, MANNING C D. Improved Semantic Representations From Tree-Structured Long Short-Term Memory Networks[J]. Computer Science, 2015, 5(1):36.
31. JI Y, EISENSTEIN J. One vector is not enough: Entity-augmented distributional semantics for discourse relations[J]. Transactions of the Association for Computational Linguistics, 2015, 3(1): 329–344.
32. SOCHER R, CHEN D, MANNING C D, et al. Reasoning with neural tensor networks for knowledge base completion[C]//International Conference on Intelligent Control and Information Processing. Beijing: ICICIP, 2013: 926-934.
33. LAI S, LIU K, HE S, ET AL. How to generate a good word embedding[J]. IEEE Intelligent Systems, 2016, 31(6): 5-14.
34. DUCHI J, HAZAN E, SINGER Y. Adaptive subgradient methods for online learning and stochastic optimization[J]. Journal of Machine Learning Research, 2011, 12(7): 2121-2159.

**A Hybrid Tree Structured Neural Network for Implicit Discourse Relation Recognition**

ZHENG Jianglong , CHEN Jinxiu \*

(School of Information Science and Engineering, Xiamen University, Xiamen 361005, China)

**Abstract:** The most critural challenge of implicit discourse relation recognition lies in how to represent the semantic information of each discourse argument. However, the semantic value of the sentence is mainly decided by its specific information focus in linguistics. Therefore, the discourse relation may mostly depend on links between information focuses. Intuitively, we cannot give equal treatment to every phrase branches during composition up the syntactic parse tree. To resolve the problem, we introduce the Tree-Structured Long Short-Term Memory(Tree-LSTM) network to selectively incorporate information from each child to compute the distributed semantic representation of two arguments. Consequently, it can emphasize those informative predicative branches that indicate the “focus” of a sentence. Then the Neural Tensor Network(NTN) is used to predict the semantic correlation between these two discourse arguments across multiple dimensions. Experimental results on PDTB corpus show that our model has achieved some improvement on the task of discourse relation recognition.

**Key words:** implicit discourse relation recognition；specific information；tree-structured long short-term memory(Tree-LSTM)；neural tensor network(NTN)

1. [↑](#footnote-ref-1)