

# 融合特征编码和短语交互感知的隐式篇章关系识别

王秀丽<sup>1,2</sup>, 金方焱<sup>1</sup>

(1. 中央财经大学信息学院, 北京 102206; 2. 国家金融安全教育部工程研究中心, 北京 102206)

**摘要:** 隐式篇章关系识别难度大、普遍性高。从论元编码和论元交互角度入手, 提出了一种融合特征编码和短语交互感知的隐式篇章关系识别模型。该模型兼顾了论元本身特征和论元间交互特征的作用, 并分别进行了优化。论元编码部分整合了双向长短时记忆网络和循环注意力卷积神经网络, 能够更全面地捕获论元全局和局部特征; 论元交互部分从短语层级考虑论元间的语义关系建模, 构建了短语级交互注意力机制, 并利用神经张量网络深入挖掘其中的关系模式, 更能体现出论元间潜在的更深层次的关联关系。在宾州篇章树库数据集上的实验结果表明, 该模型  $F_1$  值均优于其他模型。

**关键词:** 隐式篇章关系识别; 双向长短时记忆网络; 循环注意力卷积神经网络; 短语级交互注意力; 神经张量网络

**基金项目:** 教育部哲学社会科学研究重大课题攻关项目(No.22JZD011); 中央财经大学新兴交叉学科建设项目  
**中图分类号:** TP391 **文献标识码:** A **文章编号:** 0372-2112(2024)04-1377-12  
**电子学报 URL:** <http://www.ejournal.org.cn> **DOI:** 10.12263/DZXB.20220246

## Implicit Discourse Relation Recognition Integrating Feature Coding and Phrase Interaction Perception

WANG Xiu-li<sup>1,2</sup>, JIN Fang-yan<sup>1</sup>

(1. College of Information, Central University of Finance and Economics, Beijing 102206, China;

2. Engineering Research Center of State Financial Security, Ministry of Education, Beijing 102206, China)

**Abstract:** Implicit discourse relation recognition is a challenging task because of its difficulty and universality. From the perspective of argument coding and argument interaction, an implicit discourse relation recognition model integrating feature coding and phrase interaction perception is proposed. The model considers both the characteristics of argument itself and the interaction characteristics between arguments, and optimizes separately. The part of argument coding incorporates bidirectional long short-term memory (BiLSTM) and recurrent attention convolution neural network (RACNN), which can capture global and local features of arguments in a more comprehensive way; in the part of argument interaction, the semantic relationship between arguments is modeled from phrase level, and a mechanism of phrase-level interactive attention is constructed. Also, neural tensor network (NTN) is used to dig into the relational pattern, which can better reflect the potential deeper relational relationship between arguments. Experimental results on penn discourse treebank (PDTB) dataset show that the  $F_1$  values of this model are superior to other comparison models.

**Key words:** implicit discourse relation recognition; bidirectional long short-term memory; recurrent attention convolution neural network; phrase-level interactive attention; neural tensor network

**Foundation Item(s):** Key Projects of Philosophy and Social Sciences Research, Ministry of Education of the People's Republic of China (No. 22JZD011); Emerging Interdisciplinary Project of CUFU

## 1 引言

篇章间隐式逻辑关系的研究<sup>[1-3]</sup>仍处于初级阶段, 部分学者利用卷积神经网络(Convolutional Neural Network, CNN)、循环神经网络(Recurrent Neural Network,

RNN)、注意力机制等模型对论元本身进行特征建模, 但大部分模型结构单一, 论元语义理解不够深入, 无法挖掘高质量特征, 并且忽略了论元间语义交互特征。另一部分学者考虑到论元交互特征的重要性, 构建多层

复杂网络、交互注意力网络等模型来挖掘论元间交互语义特征,但缺乏对论元真实语义表达形式的深入理解,仅仅从论元词级、片段级等维度对论元进行交互建模,导致语义交互特征在关系识别上表现力不足.基于短语的注意力机制网络<sup>[4]</sup>在文本分类、自然语言推理等任务上取得了当时同类模型最好的效果,为了进一步提高隐式篇章关系识别效果,本文同时对论元本身特征编码和论元间交互特征挖掘进行了改进,提出一种融合特征编码和短语交互感知的隐式篇章关系识别模型.本文的主要贡献如下:

(1)从特征挖掘的角度入手,设计了一种融合双向长短时记忆网络(Bidirectional Long Short-Term Memory, BiLSTM)和基于迭代反馈机制( $G(x) \rightarrow \text{Mask}$ )的循环注意力卷积神经网络(Recurrent Attention Convolution Neural Network, RACNN)的编码模型,该模型利用BiLSTM和改进的RACNN分别捕获全局特征和深层次的局部特征,采用并行方式进行特征融合,避免论元特征的编码过程中的部分语义信息丢失;

(2)利用语法分析树设计了一种全新的短语级交互注意力模型,该模型融入了论元依存句法信息,从短语层级更好地挖掘论元间有效的语义特征;

(3)利用神经张量网络(Neural Tensor Network, NTN)捕获论元间线性和非线性交互特征,挖掘论元之间深层次的交互模式,捕获高维度交互信息,提高隐式篇章关系识别效果.

## 2 相关研究

目前,基于深度学习的隐式篇章关系识别研究主要分为三个方向:论元编码、论元交互以及外部信息监督.基于论元编码的方法注重于论元语义的学习;基于论元交互的方法在论元语义学习的基础上对论元之间的语义关系进行显式建模;基于外部信息监督的方法引入了显式篇章信息.

### 2.1 基于论元编码的方法

基于论元编码的方法关注论元本身语义特征的挖掘,可以利用多种神经网络模型分别从各个角度进行论元特征编码建模.文献[5]提出一种浅层卷积神经网络模型,利用平均池化、最大池化和最小池化对论元对进行联合局部特征提取,但该局部特征仅停留在表面层级,缺乏对局部重要特征的进一步筛选.除了局部特征,部分学者还关注论元全局特征.文献[6]提出一种基于前馈网络以及LSTM的模型,通过LSTM对论元对进行编码,捕获论元词的历史特征信息,但忽略了论元未来信息的影响;为了同时考虑论元前后语义信息,文献[7,8]利用BiLSTM对论元级上下文进行建模以捕获全局特征.为了同时考虑论元全局特征和局部特征的

优势,文献[9]融合内部和外部注意力模型,以充分关注模式整合了论元内部和外部知识信息;文献[10]在特征编码阶段利用LSTM-CNNs结构挖掘论元充分语义信息;文献[11]提出了一种基于动态块最大池化模型,使用CNN保留局部信息,提取多粒度特征,同时使用BiLSTM保留全局上下文信息获取论元上下文语义表示.此类方法兼顾了论元全局和局部语义信息对篇章关系的影响,但论元局部特征挖掘不够深入,存在部分局部噪声特征干扰,局部特征卷积过程也会导致全局特征部分语义丢失.

### 2.2 基于论元交互的方法

基于论元交互的方法重点对论元间语义交互进行特征建模.大部分学者利用交互式注意力机制进行论元间交互语义挖掘.文献[12]利用交互式注意力机制实现词级别和论元级别信息交互,强化了论元之间的交互性.文献[13]利用堆叠式注意力机制,将论元的自注意力特征与论元之间交互式注意力进行融合,同时考虑了论元本身特征和论元间交互特征.文献[14]通过静态交互注意力机制学习论元间的非对称关系,利用NTN获取论元深层交互模式.文献[15]利用交互式记忆注意力机制分别模拟人类双向阅读和重复阅读的特性,充分挖掘论元交互语义信息.该类方法仅从词级角度进行论元交互.文献[16]引用不同层次的表征学习,利用深层文本表征和句子分隔符提高论元表征,通过双边多角度匹配计算论元之间的关系.文献[17]通过挖掘论元对潜在的几何结构信息和语义信息同时挖掘论元本身编码和论元间特定关系模式,融合两种信息相互强化获取更好的语义.文献[18]整合词级、子词级以及句级论元信息,利用CNN进行论元交互建模.上述方法从词级、句级、几何等角度理解论元间的语义信息,但忽略了论元短语级语义表示,文献[4]表明短语能更好地表达论元的重要特征,更有效地体现论元间的语义关系.

### 2.3 基于外部信息监督的方法

基于外部信息监督的方法通常利用外部显式数据辅助隐式篇章关系识别.文献[19]利用远程监督自动标注显式篇章数据,对论元进行显式编码以识别隐式篇章关系;文献[20]从篇章之外的外部知识库中挖掘更广泛的语境,通过构建论元之间相互关系矩阵和隐式知识矩阵帮助隐式篇章关系理解;文献[21]通过在隐式篇章论元对间插入显式标记并且通过预训练模型进行评分,然后利用显式篇章关系识别器进行关系识别,选择评分最高的显式标记用作最终的隐式关系识别.该类方法局限性在于需要大量外部显式信息监督,且无法有效挖掘隐式性关系本身的语义规则.

综上所述,基于论元编码的方法注重语义表示的

学习,主要从论元本身的局部信息、所处段落的上下文信息以及整体的语法结构上进行特征挖掘,但缺乏论元间显式交互语义建模;基于论元交互的方法在论元编码的基础上,主要从词级、句级或者几何角度进行论元交互语义关系挖掘,而短语级研究较少;基于外部信息监督的方法需要大量的显式外部信息作为辅助,且无法挖掘隐式关系本身的语义特征.因此,本文整合论元编码和论元交互各自的优势,提出一种融合特征编码和短语交互感知的方法.从特征挖掘的角度出发设计一种全新的论元编码模型,在兼顾论元全局特征与局部特征的前提下进一步提高论元本身特征的质量;从短语层次角度出发设计一种全新的短语交互注意力机制以挖掘论元间真实有效的语义交互信息,以进一步提高隐式篇章关系识别效果.

### 3 模型

如图 1 所示,本文提出的融合特征编码和短语交互感知的隐式篇章关系识别模型包括 5 个环节:

(1)输入层:将论元  $Arg_1$  和  $Arg_2$  映射为词向量矩阵作为模型的输入.

(2)论元编码层:将(1)中论元的向量表示通过由 BiLSTM 和 RACNN 融合的编码模型进行编码解析语义,输出论元融合特征.

句法分析层:利用语法分析树解析论元  $Arg_1$  和  $Arg_2$  原始文本,获取其语法结构,选择第三层作为论元短语级表示,转换为 0-1 矩阵形式.

(3)论元交互层:利用(2)中分析层输出的论元短语分割矩阵将(2)中编码层输出的论元隐藏层表示转换为短语级特征表示,然后搭建短语级交互注意力层挖掘论元间的关联关系,输出  $Arg_1$ - $Arg_2$  与  $Arg_2$ - $Arg_1$  的注意力语义特征.

(4)NTN:将(3)中的注意力语义特征输入到 NTN 中,探索论元交互关系中的特有模式,进一步捕捉更深层次的交互关系,输出论元对不同关系的分数向量.

(5)输出层:将(4)中的输出输入到全连接层中,经过 softmax 完成关系识别任务.

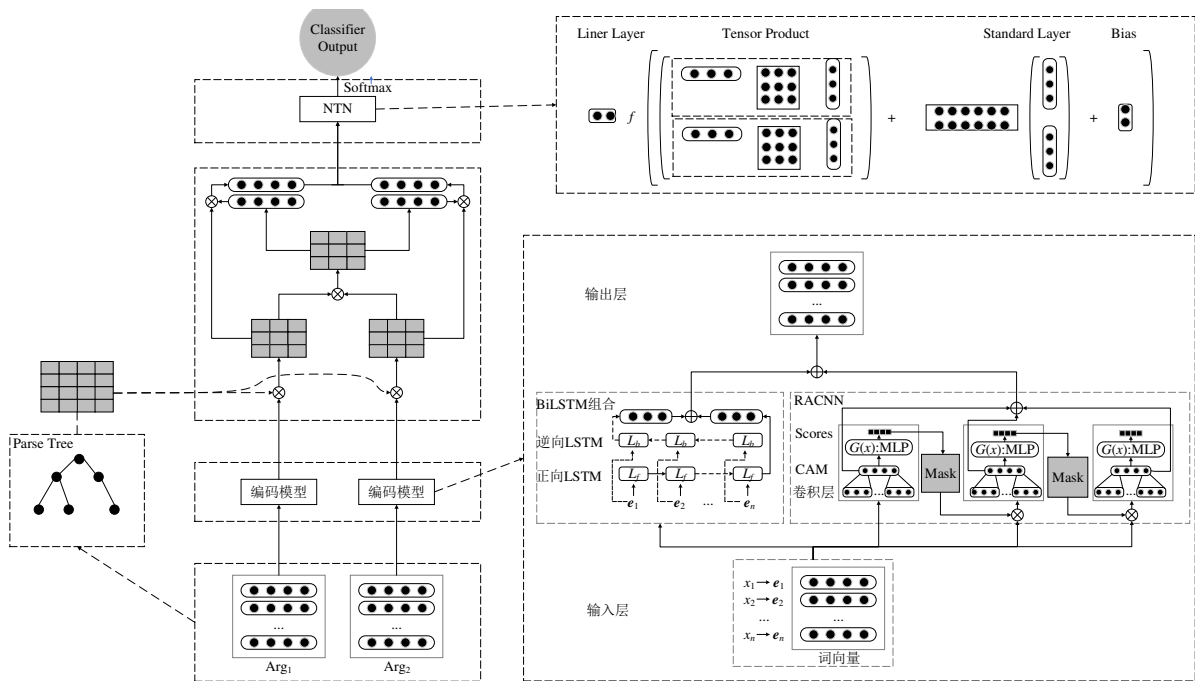


图 1 融合特征编码和短语交互感知的隐式篇章关系识别模型

#### 3.1 基于 BiLSTM-RACNN 的论元编码

##### 3.1.1 BiLSTM

LSTM 只关心文本中当前词的历史信息而忽略了未来信息. BiLSTM 考虑了当前词的前后文语义信息,可以获取到整个文本序列的信息充分挖掘文本全局特征,如图 2 所示.

给定一个包含  $n$  个词的论元  $Arg = \{x_1, x_2, \dots, x_n\}$ , 通过训练将  $Arg$  向量化,映射为  $E_m = (e_1, e_2, \dots, e_n)$ , 其中

$e_i = (w_1, w_2, \dots, w_m), w \in [-1, 1], m$  为指定的词向量维度. 将  $E_m$  输入到 BiLSTM 中,分别输出正向结果  $fw$  和反向结果  $bw$ ,连接  $fw$  和  $bw$  作为 BiLSTM 输出:

$$output_i = (fw \oplus bw) \quad (1)$$

$$H = (output_1, output_2, \dots, output_n) \quad (2)$$

其中,  $output_i$  表示词向量  $e_i$  的输出,  $H$  表示  $E_m$  的输出.

##### 3.1.2 RACNN 变体

RACNN<sup>[22]</sup> 是一种卷积神经网络结构. 本文提出一

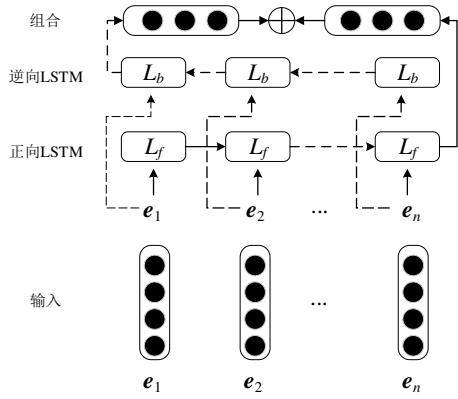


图2 BiLSTM

种 RACNN 变体,如图 3 所示. 该变体采用了迭代反馈机制 ( $G(x) \rightarrow \text{Mask}$ ), 连续对文本词向量进行掩膜 (Mask) 处理, 模拟人注意力特性, 能够更好地捕获文本中关键的局部特征.

RACNN 单次卷积层如图 4 所示, 首先将包含  $n$  个词的输入文本中的词  $x_i$  通过训练映射为词向量  $e_i$ , 组合形成词向量矩阵  $E_m = (e_1, e_2, \dots, e_n)$ ; 其次, 用  $r \times m$  的过滤器来对  $E_m$  进行卷积, 过滤器的数量为  $u$ , 提取出  $E_m$

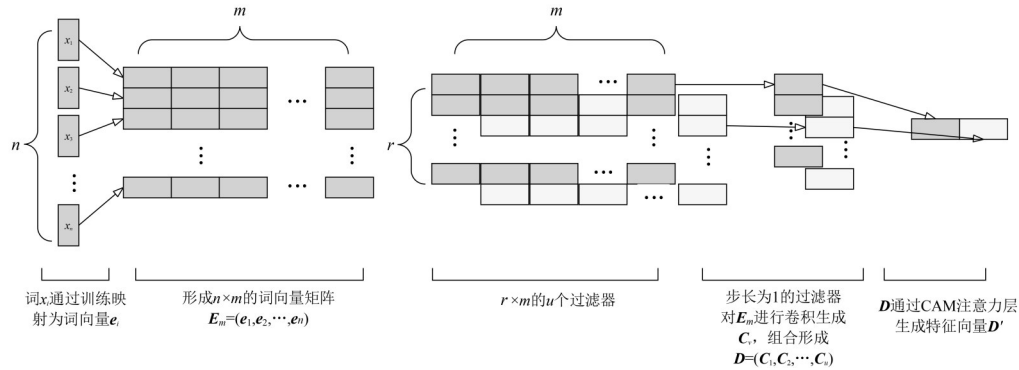


图4 RACNN 单次卷积

如图 3 所示, 将  $D$  输入到 RACNN 的注意力层, 采用 CAM 提取新的局部特征  $D'$  来代替  $D$ :

$$D_{\text{Max}} = \text{Maxpool}(D, \text{axis} = 1) \quad (6)$$

$$D_{\text{Avg}} = \text{Avgpool}(D, \text{axis} = 1) \quad (7)$$

$$D'_{\text{Max}} = w_2 \cdot (w_1 \cdot D_{\text{Max}} + b_1) + b_2 \quad (8)$$

$$D'_{\text{Avg}} = w_2 \cdot (w_1 \cdot D_{\text{Avg}} + b_1) + b_2 \quad (9)$$

$$D' = \text{sigmoid}(D'_{\text{Max}} + D'_{\text{Avg}}) \quad (10)$$

其中,  $\text{axis} = 1$  表示按维度 1 对  $D$  池化;  $w_1, b_1, w_2, b_2$  为共享 MLP 的参数, 最终  $D'$  表示为一个长度为  $u$  的局部特征向量:

$$D' = (d_1, d_2, \dots, d_u) \quad (11)$$

RACNN 中迭代反馈函数  $G(x)$  可以通过全连接层实现, 先计算文本中词  $x_i$  对应的分数  $\text{score}_i$ , 最后生成一

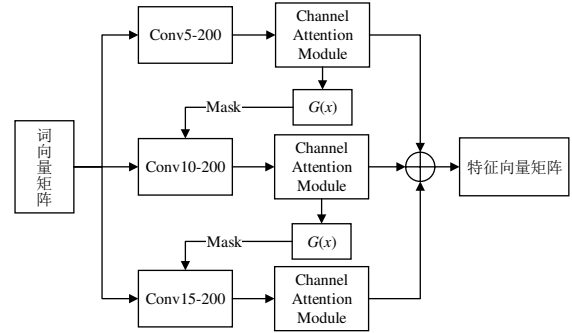


图3 RACNN 变体

局部特征:

$$c_{v(i)} = \text{relu}(W_v \cdot e_{(i:i+r-1)} + b_v) \quad (3)$$

其中,  $v \in (1, u)$  表示过滤器的序号;  $W_v$  表示第  $v$  个  $r \times m$  的过滤器;  $b_v$  表示偏置项;  $e_{(i:i+r-1)}$  表示  $E_m$  中从  $i$  到  $i+r-1$  行向量;  $c_{v(i)}$  表示第  $v$  个过滤器提取出的局部特征. 过滤器按步长为 1 进行滑动, 最后得到局部特征集合  $C_v$ , 所有的局部特征集合组合成局部特征矩阵  $D$ :

$$C_v = (c_{v(1)}, c_{v(2)}, \dots, c_{v(n-r+1)}) \quad (4)$$

$$D = (C_1, C_2, \dots, C_u) \quad (5)$$

个矩阵  $F_{\text{Mask}}$ , 用于识别分数低的词:

$$\text{Scores} = \{\text{score}_1, \text{score}_2, \dots, \text{score}_n\} \quad (12)$$

$$F_{\text{Mask}} = \text{Matrix}(\text{IODS}(\text{Scores})) \quad (13)$$

其中, IODS 表示获取 Scores 中分数较低值索引的函数; Matrix 表示将索引列表映射为矩阵.

将  $F_{\text{Mask}}$  反馈到下一次卷积, 卷积层的输入更新为  $E_m'$ :

$$E_m' = F_{\text{Mask}} \cdot E_m \quad (14)$$

如上述步骤迭代, 直到 3 次卷积操作完成, 将 3 次卷积操作输出的特征向量  $D'$  组合, 形成最终的局部特征向量:

$$E = (D'_1, D'_2, D'_3) \quad (15)$$

其中,  $D'_i$  表示第  $i$  次卷积操作输出的特征向量  $D'$ .

### 3.2 短语级交互注意力模型

#### 3.2.1 词级交互注意力机制

从词级角度出发,根据  $Arg_1$  和  $Arg_2$  两个论元的隐藏层表示计算两个论句之间词对的语义关系矩阵,每个矩阵元素表示  $Arg_1$  中第  $i$  个词和  $Arg_2$  中第  $j$  个词的语义关系分数.

$$S(i,j)=(h_i^1)^T \cdot h_j^2 \quad (16)$$

其中,  $S \in \mathbf{R}^{n_1 \times n_2}$ ,  $n_1, n_2$  分别表示  $Arg_1$  和  $Arg_2$  的词级长度.

#### 3.2.2 句法分析树

句法分析树(parse tree)可以确定一个句子中哪些是短语,词的属性等.由 Stanford Parser 来实现.如图 5 所示:

Sentence = “a little girl wearing a green dress with blue dots carries a box by its handle”.

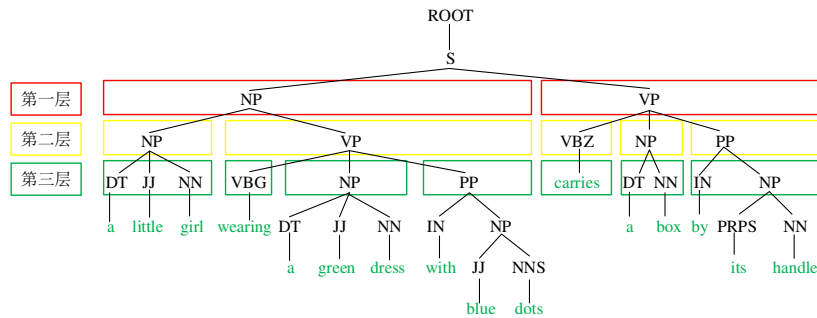


图5 句法分析树

第一层:“(a little girl wearing a green dress with blue dots) (carries a box by its handle)”

第二层:“(a little girl) (wearing a green dress with blue dots) (carries) (a box) (by its handle)”

第三层:“(a little girl) (wearing) (a green dress) (with blue dots) (carries) (a box) (by its handle)”

其中,对于小于或等于3个词的短语不再向下分解.第三层相比于前两层来说,短语划分更加明显,能够更好地表示句子短语级信息.因此,本模型选择第三层作为短语级表示,将论元隐藏层词级表示转换为短语级表示,加入到模型中进行训练.

#### 3.2.3 短语级交互注意力机制

短语一般比单词更具有语义表现力,更能体现论

元的重要语义信息.本文设计了短语级交互注意力模型,从短语级角度挖掘论元间交互特征,呈现表现力更强的语义交互注意力信息,在隐式篇章关系识别中更具有表现价值,如图6所示.

首先,利用句法分析树对  $Arg_1$  和  $Arg_2$  文本进行句法分析,选择第三层分别构建短语级表示矩阵,即  $Q_1$  和  $Q_2$ ,然后通过矩阵乘法变换将  $Arg_1$  和  $Arg_2$  的隐藏层表示转换为短语级表示  $P_1$  和  $P_2$ ,然后通过点乘生成短语级交互信息矩阵.分别通过按行和按列对短语级交互信息矩阵进行 softmax,分别得出  $Arg_1$ - $Arg_2$  和  $Arg_2$ - $Arg_1$  的注意力语义分数,最后与  $P_1$  和  $P_2$  相结合分别形成  $Arg_1$  和  $Arg_2$  对应的交互语义表示.

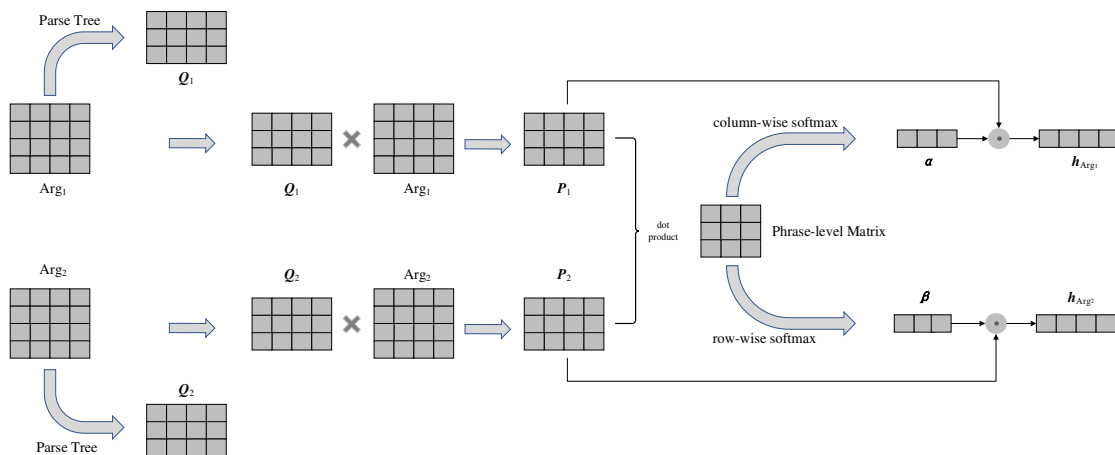


图6 短语级交互式注意力模型

### 3.3 神经张量网络

篇章关系不同的核心在于其内部特征的差异性,而这种差异性可以通过论元间的交互模型来体现.如何高效捕获论元之间的交互模型对论元之间的关系识别至关重要.

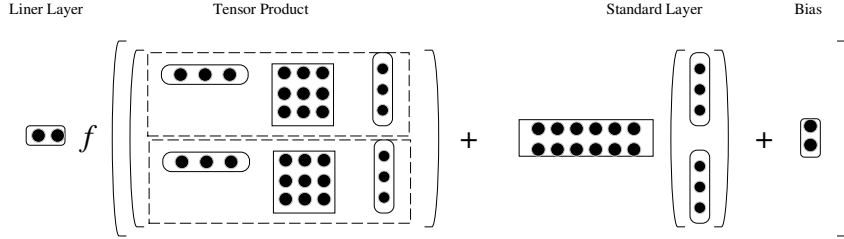


图7 神经张量网络模型

$$S(\mathbf{h}_{\text{arg}_1}, R, \mathbf{h}_{\text{arg}_2}) = \mathbf{u}_R^T f(\mathbf{h}_{\text{arg}_1}^T \mathbf{W}_R^{[1:k]} \mathbf{h}_{\text{arg}_2} + \mathbf{V}_R \begin{bmatrix} \mathbf{h}_{\text{arg}_1} \\ \mathbf{h}_{\text{arg}_2} \end{bmatrix} + \mathbf{b}_R) \quad (17)$$

其中,  $f$  是标准非线性函数  $\tanh$ ;  $\mathbf{W}_R^{[1:k]} \in \mathbf{R}^{d \times d \times k}$  是一个由  $k$  个  $d \times d$  的矩阵构成的张量, 每个矩阵称为张量片, 所有的张量片可以同时捕获论元间交互特征, 进而抽取论元间深层次的交互语义特征;  $\mathbf{h}_{\text{arg}_1}^T \mathbf{W}_R^{[1:k]} \mathbf{h}_{\text{arg}_2}$  指的是双线性张量项, 生成一个向量  $\mathbf{h} \in \mathbf{R}^{k \times 1}$ , 其中,  $\mathbf{h}$  的每个元素  $h_i = \mathbf{h}_{\text{arg}_1}^T \mathbf{W}_R^{[i]} \mathbf{h}_{\text{arg}_2}$  ( $i=1, 2, 3, \dots, k$ ),  $\mathbf{W}_R^{[i]}$  是随机参数, 随着  $i$  的不断增大, 参数也在不断增多, 模型也变得更加复杂. 另外, 剩余参数是标准神经网络的组成部分:  $\mathbf{u}_R \in \mathbf{R}^{k \times 1}$ ,  $\mathbf{V}_R \in \mathbf{R}^{k \times d}$ ,  $\mathbf{b}_R \in \mathbf{R}^{k \times 1}$ .

### 3.4 隐式篇章关系识别

在各个篇章关系上, 采用随机抽样的方法来选择每次输入的训练集. 对于每个类别的分类器, 采用 AdaDelta 优化器来对参数更新, 利用学习率动态迭代更新的方式来帮助模型达到更好的拟合效果. 另外, 通过交叉熵 (cross-entropy) 损失函数度量每个样本的预测误差. 给定一组论元对 ( $\text{Arg}_1$  和  $\text{Arg}_2$ ) 及其类别标签  $y$ , 则其损失函数定义如下:

$$L(\hat{y}, y) = - \sum_j y_j \log(\text{Pr}(\hat{y}_j)) \quad (18)$$

其中,  $\text{Pr}(\hat{y}_j)$  表示第  $j$  个标签的预测概率,  $C$  是篇章关系类别的数量.

训练目标的目的是使正确篇章关系具有更大概率. 由于 NTN 的高维特性导致挖掘的关系特征具有很强的稀疏性, 一般可以利用  $l_0$  和  $l_1$  范数来实现稀疏约束, 但是对  $l_0$  的求解相对困难, 利用  $l_1$  进行求解, 并且, 利用  $l_2$  正则化避免过拟合问题. 将模型中 NTN 张量片参数记为  $\theta_0$ , 其余参数记为  $\theta_1$ , 具体最小化训练目标如下:

NTN 整合了双线性模型和单层网络的优势, 可以捕获论元间更高维的线性和非线性交互特征. 通过构建神经张量网络, 挖掘论元之间深层次的交互模式, 捕获高维度的交互特征, 进而提高隐式篇章关系识别效果. 具体实现过程如图 7 所示.

$$J(\theta) = \frac{1}{n} \sum_{r=1}^n L(\hat{y}^r, y^r) + \lambda_0 \|\theta_0\|_1 + \lambda_1 \|\theta_1\|_2^2 \quad (19)$$

其中,  $\lambda_0$  表示  $l_1$  正则项参数,  $\lambda_1$  表示  $l_2$  正则项参数.

## 4 实验与分析

### 4.1 数据集

宾州篇章树库 (Penn Discourse Treebank, PDTB)<sup>[23]</sup> 是目前应用最广泛的篇章语料库. 本文选择 PDTB2.0 进行相关研究.

PDTB 包含 40 600 个样本数据, 分为 25 个章节. 其中对篇章关系连接词和论元进行了明显标注, 显式篇章关系连接词本身是原始数据中的一部分, 而隐式篇章关系连接词本来并不存在于原始数据中, 为了能够展现篇章潜在的隐式关系, 标注人员手动增加了符合隐式篇章关系的连接词, 即隐式篇章关系连接词. 如下例所示, e.g.1 表示显式篇章关系数据, 其中 “But” 是标注好的关系连接词. e.g.2 表示隐式篇章关系数据, 其中 “But” 是根据语义信息插入的连接词, 暗示篇章论元  $\text{Arg}_1$  和  $\text{Arg}_2$  之间潜在的逻辑关系.

(e.g.1)  $\text{Arg}_1$ : There’s no question that some of those workers and managers contracted asbestos-related diseases.

$\text{Arg}_2$ : But you have to recognize that these events took place 35 years ago. (wsj\_0003)

(e.g.2)  $\text{Arg}_1$ : Before boarding the buses again.

$\text{Arg}_2$ : [Implicit=But] This time, it was for dinner and dancing — a block away. (wsj\_0010)

PDTB 关系类型共有 5 种, 除了 Explicit (显式) 和 Implicit (隐式) 外, 还存在另外 3 种关系类型: AltLex (可选关系)、NoRel (没有关系) 和 EntRel (实体关系). 另外, 数据集对显式、隐式和可选关系扩展了一个三层关系结构. 其中, 第一层级关系包括 Contingency (偶然关

系)、Expansion(扩展关系)、Temporal(时序关系)以及Comparison(比较关系),剩下两层分别包括16和23种关系. 鉴于大部分学者都只针对第一层关系开展研究,为了更好地进行比较,本文研究也集中在第一层关系. PDTB整体数据分布和第一层级关系数据分布分别如表1和表2所示.

表1 PDTB整体数据分布

PDTB Relations	No. of tokens
Explicit	18 459
Implicit	16 224
AltLex	624
EntRel	5 210
NoRel	254
Total	40 600

表2 PDTB第一层级关系数据分布

"CLASS"	Explicit	Implicit	AltLex	Total
"TEMP."	3 612	950	88	4 650
"CON."	3 581	4 185	276	8 042
"COM."	5 516	2 832	46	8 394
"EXP."	6 424	8 861	221	15 506
Total	19 133	16 828	634	36 592

值得注意的是,表1中各个关系类别的数量和表2中统计数据存在一定差距,这是第一层关系下可能同时存在多种二级关系导致的,因此表2中的统计数据会略大于表1.

#### 4.2 评价指标

通常  $F_1$  比 Accuracy 能更好地评价模型性能,也是本研究领域经典的评价指标,因此本文采用3个指标即 Precision、Recall 和  $F_1$  进行对比实验. 在多分类任务中,为了能够更好地表现模型识别效果,最终的评价指标是由每一类单独的评价指标计算得来的,计算方式有3种:微平均、宏平均和加权. 根据语料库的数据样本分布,本文评价指标选择加权计算方式.

此外,考虑到选择一个方法训练的最优模型在未知数据上表现不好,本文在实验中对每种方法都训练3次,选取评估指标的平均值作为最终的指标,这样既避免了单一模型的随机性,又提高了模型的鲁棒性.

#### 4.3 实验环境

本文实验环境如表3所示.

表3 实验环境

名称	配置	名称	配置
操作系统	Windows10	内存	16 GB
CPU	Intel Core i7-6700	开发环境	Tensorflow 1.4.1
GPU	GeForce GTX 2080Ti		

#### 4.4 参数选择

基本参数设置如表4所示.

表4 基本参数

参数	值	参数	值
词向量维度	50	隐藏层大小	50
Epoch	50	损失函数	交叉熵
优化函数	Adadelata	Batch_size	10
学习率	0.3	L1	1E-5
dropout_ratio	0.3	L2	1E-5
K_slices	3	滑动窗口	5,10,15
Masksize	15	滑动窗口数量	200

其中,隐藏层是指LSTM中的hiddenlayer; dropout\_ratio是指特征过滤力度;K\_slices表示NTN中的切片数量;Masksize表示RACNN中掩膜大小.

将输入论元长度设定为50,如果原论元长度大于50,则进行裁剪,如果原论元长度小于50,则进行补0填充. 对于词向量词典中未出现的词,由[-1,1]均匀分布随机初始化生成向量表示. RACNN模块中滑动窗口过小会导致卷积过程侧重于对细节的捕捉,但是忽略了局部特征的整体把握,而滑动窗口过大时,卷积过程则侧重于局部特征整体性而忽略了细节特征,实验验证选择5,10,15能达到最好的效果;RACNN模块中Masksize过小时,上一个卷积的反馈对下一个卷积的特征聚焦效果不明显,而Masksize过大时,上一个卷积的反馈会导致大部分特征丢失,实验验证Masksize设置为15能达到更好的效果. 另外用Adadelata优化函数来优化损失,采用的是学习率递减的方式,为了能够在Epoch为50的前提下模型成功收敛并且效果显著,因此对学习率的最高点进行了调整对比,如表5所示. 当学习率最高点设置为0.3时收敛效果最好.

表5 学习率

学习率	1	0.5	0.3	0.1
$F_1/\%$	48.07	49.40	49.99	47.22

另外,NTN模块中不同的切片能够获取不同的论元交互模式特征,切片越多,模型计算越复杂,难度越大,为此,对切片数量K\_slices进行调整对比,如表6所示. 当切片数量选择3时效果最优.

表6 切片数量

切片数量	1	2	3	4
$F_1/\%$	47.16	48.08	49.99	49.22

#### 4.5 对比方法

本文设置了横向和纵向两种对比方式. 横向对比指的是本文模型与其次级模型进行比较,即消融实验,可以验证模型中模块优化的有效性;纵向比较指的是

本文模型与当前领域最前沿的研究成果进行比较,主要从论元编码、论元交互的角度进一步验证模型的核心竞争力,并且对于不适合进行对比实验的研究成果进行解释说明。

#### 4.5.1 横向对比

(1)BiLSTM+RACNN:基于论元编码的隐式篇章关系识别模型,作为基准模型,可以作为其他模型的对比模型。

(2)BiLSTM+RACNN+IA:在基于论元编码模型的基础上增加了词级论元交互模块,用来测试论元交互对隐式篇章关系识别的影响。

(3)BiLSTM+RACNN+ParseMatrix+IA:利用句法分析树将基于词级论元交互模型改进为短语级,进而测试短语级论元交互相比于词级论元交互的优势。

(4)BiLSTM+RACNN+IA+NTN:在词级论元交互模型的基础上增加了NTN,用来测试NTN在促进隐式篇章关系识别效果中的表现力。

#### 4.5.2 纵向对比

(1)DC-BCNN<sup>[11]</sup>:基于动态块最大池化的BiLSTM-CNN模型,将通过BiLSTM获取的文本语义信息输入到CNN框架中提取特征,最终用于关系分类任务。

(2)Syntax+self-attention<sup>[7]</sup>:融合双仿射深度神经网络和BiLSTM的隐式篇章关系分类模型,运用双仿射神经网络识别句法信息,再通过BiLSTM对句法信息和原文本进行特征提取,最终用于关系分类任务。

(3)TLAN<sup>[15]</sup>:三层注意力网络模型,利用交互式注意力机制和记忆注意力机制分别模拟人类双向阅读和重复阅读的特性,生成论元的最终表示。

(4)Bai's<sup>[18]</sup>:增强不同粒度文本表示的模型,通过引入基于上下文的词向量加强输入,并通过多层卷积神经网络建模不同长度的文本片段。不过该模型通过获取论元各种片段(字符级、子词级、词级)作为模型的输入表示,与本文模型输入为论元原始片段存在差异性,因此可比性略低。

(5)BMGF-RoBERTa<sup>[16]</sup>:深度融合上下文表示、双边多角度匹配以及全局信息融合的模型,引用不同层次的表征学习,深层文本表征和句子分隔符可以提高论元表征,双边多角度匹配可以计算论元之间的关系而不用考虑繁重的特征工程,全局信息融合模型结合多头注意力和门控机制获取论元的关键信息。不过该模型输入借助了Bert模型,需要利用“CLS”、“EOS”等标示,无法与本文模型输入统一。

(6)TIASL<sup>[14]</sup>:结合静态交互注意力机制和NTN的隐式篇章关系识别模型,通过交互注意力机制学习论元之间的非对称关系,利用具有稀疏约束的NTN获取论元深层交互模式,挖掘论元隐式篇章关系模式。

## 4.6 结果分析

本文模型在PDTB上进行了多个对比实验,其中四分类实验结果和二分类实验结果分别如表7和表8所示。

表7 PDTB四分类实验结果

单位:%

方法	Precision	Recall	$F_1$
BiLSTM+RACNN	46.25	52.60	47.84
BiLSTM+RACNN+IA	49.35	52.35	48.54
BiLSTM+RACNN+ParseMatrix+IA	50.08	52.91	49.21
BiLSTM+RACNN+IA+NTN	48.09	53.52	48.57
DC-BCNN <sup>[11]</sup>	47.59	51.70	47.79
Syntax+self-attention <sup>[7]</sup>	46.67	54.31	47.42
TLAN <sup>[15]</sup>	52.40	52.75	47.32
TIASL <sup>[14]</sup>	49.54	53.39	48.95
本文模型	49.27	53.46	49.99

由表7可知,BiLSTM+RACNN+IA比BiLSTM+RACNN增加了词级论元交互模块,不仅考虑了论元本身的语义特征,还考虑了论元之间的相互语义关系, $F_1$ 值高0.7%;BiLSTM+RACNN+ParseMatrix+IA与BiLSTM+RACNN+IA相比,改进在于将词级论元交互模块转换为短语级论元交互,在论元交互方面也更能够捕获到论元间隐含的语义关联关系, $F_1$ 值高0.67%;BiLSTM+RACNN+IA+NTN在BiLSTM+RACNN+IA的基础上增加了NTN,结合了线性和非线性特征挖掘的优势,能够在论元交互的基础上进一步挖掘论元间更深层次的关系模式, $F_1$ 值高0.03%。

纵向比较中,与DC-BCNN模型相比,本文模型在论元编码环节通过并接方式进行特征交融避免了特征丢失问题,而且还考虑了论元间的交互性, $F_1$ 值高2.2%;Syntax+self-attention模型缺乏论元之间交互特征的捕获,本文模型 $F_1$ 值高2.57%;TLAN模型在整个过程中缺乏对非重要特征的去除,另外在论元交互中没有考虑到论元中短语在语义表达上的作用,并且利用池化操作获取注意力向量可能会导致语义不充分问题,而本文模型在解决以上问题的同时,考虑了交互特征向量更深层次的关系模式挖掘, $F_1$ 值高2.67%;与TIASL模型相比,本文模型在论元特征挖掘时不仅利用BiLSTM获取全局特征,也利用RACNN获取局部特征,丰富了论元特征表示。另外,利用句法分析树将词级交互注意力机制改进为短语级交互注意力机制,能够在进一步理解论元语义的同时也能更易于挖掘论元间隐含的关系特征, $F_1$ 值高1.04%。

由表8可以看出在二分类实验中各个对比模型在四个关系上的识别效果( $F_1$ 值),可以得出如下结论:

(1)在四种关系分类中,所有模型在对比关系和时序关系上的表现效果相对较差。原因是这两种关系在

表 8 PDTB 二分类实验结果

单位:%

方法	对比关系	偶然关系	扩展关系	时序关系
BiLSTM+RACNN	29.24	48.17	67.69	28.74
BiLSTM+RACNN+IA	32.00	48.62	68.43	29.44
BiLSTM+RACNN+ParseMatrix+IA	32.46	50.78	68.77	29.65
BiLSTM+RACNN+IA+NTN	32.36	50.61	68.99	29.66
DC-BCNN <sup>[11]</sup>	32.00	48.66	67.80	25.08
Syntax+self-attention <sup>[7]</sup>	32.30	47.40	68.99	23.21
TLAN <sup>[15]</sup>	32.52	48.98	67.73	22.22
TIASL <sup>[14]</sup>	32.46	50.53	68.87	29.19
本文模型	32.55	51.85	69.07	30.18

训练集中所占的比例小,分别占约 15% 和 6%。训练样本少导致模型对该类关系样本学习的特征不充分。从四种关系样本分布可以看出,随着训练样本的增加,模型识别的准确率也在上升,表明模型训练语料库的大小对隐式篇章关系识别至关重要。

(2)偶然关系和扩展关系方面,横向比较来看,后两种模型(BiLSTM+RACNN+ParseMatrix+IA 和 BiLSTM+RACNN+IA+NTN)均比前两种模型(BiLSTM+RACNN 和 BiLSTM+RACNN+IA)性能更好,分别验证了短语级交互注意力机制和 NTN 在隐式篇章关系识别中的重要作用。纵向比较来看,TIASL 的  $F_1$  值分别达到了 50.53% 和 68.87%,其重点关注论元交互部分,利用词级交互和 NTN 深层次挖掘论元交互语义特征。DC-BCNN 和 Syntax+self-attention 都是针对论元编码部分研究,缺乏对论元交互进一步挖掘;TLAN 很好地发挥了

注意力机制在论元编码和论元交互中的作用,由于模型复杂度的限制未能达到很出众的效果。本文模型在论元编码的基础上,进一步挖掘论元交互特征,利用句法分析树提出了短语级交互注意力机制,并且利用 NTN 高维度挖掘交互关系模式,在两种关系实验上  $F_1$  值分别达到了 51.85% 和 69.07%,较其他模型的性能更好。

#### 4.7 效率分析

本节从模型训练过程中准确率和损失值变化对模型进行效率对比分析。以四分类实验为基础,对 DC-BCNN、Syntax+self-attention、TLAN、TIASL 以及本文模型的训练集损失值和验证集准确率进行了比较,数据曲线由可视化工具 TensorBoard 生成,如图 8 和图 9 所示。其中,Step(步)表示训练步数,指的是每个 Epoch(迭代)中一个 batch\_size(批量大小)数据的训练过程。

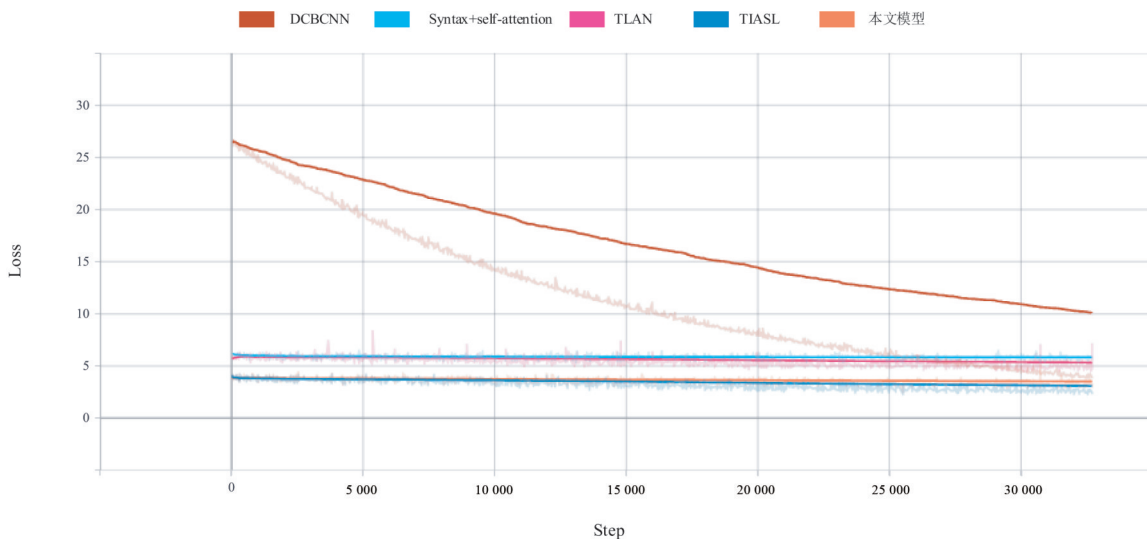


图 8 训练步数-损失值(训练集)

由图 8 可知,在训练过程中,除了 DC-BCNN 外,其他模型训练集损失值均快速地收敛,并且趋于平稳状态,原因可能是 DC-BCNN 中动态块最大池化过程中特

征动态学习相比于其他模型特征静态学习较为缓慢;另外可以看出,训练集损失值变化层级可以分为三个等级:等级一(DC-BCNN)、等级二(Syntax+self-attention

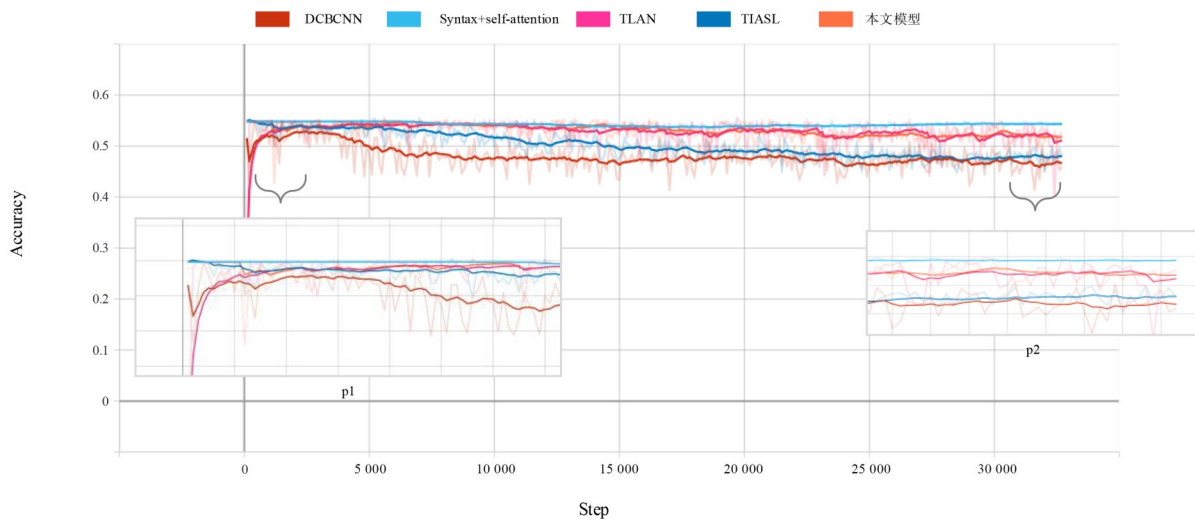


图9 训练步数-准确率(验证集)

和TLAN)以及等级三(本文模型和TIASL),其中,等级三损失值最小,本文模型相比TIASL的训练集损失值收敛速度和变化无较大差异。由图9可知,在训练初始阶段(图中p1所示),Syntax+self-attention的验证集准确率上升速度最快,并且保持稳定,其次是本文模型和TIASL,最后是DC-BCNN和TLAN,其中,TIASL和DC-BCNN在训练中期阶段出现准确率下降趋势,这可能是模型在验证过程中出现了部分过拟合现象;从图9中p2可以看出,最后阶段本文模型的验证集准确率稳定在前列。

基于多种神经网络,本文模型兼顾了论元编码和论元交互的优势,引入句法分析树构建了短语级交互注意力机制,进一步提高了论元间语义交互信息准确性,同时还利用NTN更深层次挖掘论元间高维度特征模式,模型参数不断增加,结构也变得更加复杂,因此在上述训练集损失值和验证集准确率分析中未达到所有模型中的最高效率,但仍然处于前列,实现了以最小的时间代价换取最多的性能。

## 5 结束语

基于上述多种神经网络模型,提出一种融合特征编码和短语交互感知的隐式篇章关系识别模型。整合了论元编码、论元交互和关系模式挖掘方面的优势,利用BiLSTM-RACNN、短语级交互注意力机制和NTN分别对论元编码和论元交互两个模块进行了改进优化,既能保证论元本身特征的全面性和高质量,还更深层地挖掘了论元间短语级交互语义关系和高维度的关系特征模式。在PDTB数据集上的实验表明,本文模型在四分类和二分类任务中均优于其他模型。

同时,本文也存在一些不足之处,与大部分相关研究成果一样,论元编码主要基于RNN、CNN模型,只在

此基础上进一步改进,缺乏利用Bi-GRU、Tree-LSTM等最新变种模型进行探索。论元交互部分,本文设计了一种短语级交互注意力机制,相比于前人研究使用的传统交互注意力机制效果更好,但论元中短语重要程度各不相同,可根据短语重要性在交互过程中加入加权影响;考虑将论元外部特征加入到模型输入中,文本外部特征依然存在很多有价值的信息,比如依存句法分析中的单词标签信息等;尝试Bi-GRU、Tree-LSTM等多种不同的模型进一步对比特征抽取效果;在论元交互部分,词对、短语对等层级也是值得研究的方向;提高模型的普适性,增加中文数据集进行对比实验;篇章间上下文信息同样会对模型产生一定的影响。

## 参考文献

- [1] ZHOU P, SHI W, TIAN J, et al. Attention-based bidirectional long short-term memory networks for relation classification[C]//Proceedings of the Annual Meeting of the ACL. Stroudsburg: ACL, 2016: 207-212.
- [2] 李志欣, 孙亚茹, 唐素勤, 等. 双路注意力引导图卷积网络的关系抽取[J]. 电子学报, 2021, 49(2): 315-323.  
LI Z X, SUN Y R, TANG S Q, et al. Dual attention guided graph convolutional networks for relation extraction[J]. Acta Electronica Sinica, 2021, 49(2): 315-323. (in Chinese)
- [3] ZHANG Y, MENG F, LI P, et al. Context tracking network: Graph-based context modeling for implicit discourse relation recognition[C]//Proceedings of the Conference of the North American Chapter of the ACL: Human Language Technologies. Stroudsburg: ACL, 2021: 1592-1599.
- [4] WU W, WANG H, LIU T, et al. Phrase-level self-attention networks for universal sentence encoding[C]//Proceedings of the Conference on Empirical Methods in Natural Lan-

- guage Processing. Stroudsburg: ACL, 2018: 3729-3738.
- [5] ZHANG B, SU J, XIONG D, et al. Shallow convolutional neural network for implicit discourse relation recognition [C]//Proceedings of the Conference on Empirical Methods in Natural Language Processing. Stroudsburg: ACL, 2015: 2230-2235.
- [6] RUTHERFORD A, DEMBERG V, XUE N. A systematic study of neural discourse models for implicit discourse relation[C]//Proceedings of the Conference of the European Chapter of the ACL. Stroudsburg: ACL, 2017: 281-291.
- [7] 凡子威, 张民, 李正华. 基于 BiLSTM 并结合自注意力机制和句法信息的隐式篇章关系分类[J]. 计算机科学, 2019, 46(5): 214-220.
- FAN Z W, ZHANG M, LI Z H. BiLSTM-based implicit discourse relation classification combining self-attention mechanism and syntactic information[J]. Computer Science, 2019, 46(5): 214-220. (in Chinese)
- [8] DAI Z, HUANG R. Improving implicit discourse relation classification by modeling inter-dependencies of discourse units in a paragraph[C]//Proceedings of the Conference of the North American Chapter of the ACL: Human Language Technologies. Stroudsburg: ACL, 2018: 141-151.
- [9] ZHANG B, XIONG D, SU J, et al. Learning better discourse representation for implicit discourse relation recognition via attention networks[J]. Neurocomputing, 2018, 275: 1241-1249.
- [10] POPA D N, PEREZ J, HENDERSON J, et al. Implicit discourse relation classification with syntax-aware contextualized word representations[C]//Proceedings of the Florida Artificial Intelligence Research Society. New York: AAAI, 2019: 203-208.
- [11] GUO F, HE R, DANG J. Implicit discourse relation recognition via a BiLSTM-CNN architecture with dynamic chunk-based max pooling[J]. IEEE Access, 2019, 7: 169281-169292.
- [12] 徐扬, 周文瑄, 阮慧彬, 等. 基于层次化表示的隐式篇章关系识别[J]. 南京大学学报: 自然科学版, 2019, 55(6): 1000-1009.
- XU Y, ZHOU W X, RUAN H B, et al. Hierarchical representation for implicit discourse relation recognition[J]. Journal of Nanjing University (Natural Sciences), 2019, 55(6): 1000-1009. (in Chinese)
- [13] 阮慧彬, 徐扬, 孙雨, 等. 基于堆叠式注意力机制的隐式篇章关系识别[J]. 山西大学学报: 自然科学版, 2020, 43(3): 508-516.
- RUAN H B, XU Y, SUN Y, et al. Stacked-attention based implicit discourse relation recognition[J]. Journal of Shanxi University (Natural Sciences), 2020, 43(3): 508-516. (in Chinese)
- [14] GUO F, HE R, JIN D, et al. Implicit discourse relation recognition using neural tensor network with interactive attention and sparse learning[C]//Proceedings of the Conference on Computational Linguistics. Stroudsburg: ACL, 2018: 547-558.
- [15] 徐昇, 王体爽, 李培峰, 等. 运用多层注意力神经网络识别中文隐式篇章关系[J]. 中文信息学报, 2019, 33(8): 12-19, 35.
- XU S, WANG T S, LI P F, et al. Multi-layer attention network based Chinese implicit discourse relation recognition[J]. Journal of Chinese Information Processing, 2019, 33(8): 12-19, 35. (in Chinese)
- [16] LIU X, OU J, SONG Y, et al. On the importance of word and sentence representation learning in implicit discourse relation classification[C]//Proceedings of the Conference on Artificial Intelligence. Amsterdam: Elsevier, 2020: 3830-3836.
- [17] HE R, WANG J, GUO F, et al. Transs-driven joint learning architecture for implicit discourse relation recognition [C]//Proceedings of the Annual Meeting of the ACL. Stroudsburg: ACL, 2020: 139-148.
- [18] BAI H, ZHAO H. Deep enhanced representation for implicit discourse relation recognition[C]//Proceedings of the Conference on Computational Linguistics. Stroudsburg: ACL, 2018: 571-583.
- [19] 唐裕婷, 李艳斌, 刘露, 等. 面向细粒度隐式篇章关系识别的远距离监督特征学习算法[J]. 北京大学学报: 自然科学版, 2019, 55(1): 91-97.
- TANG Y T, LI Y B, LIU L, et al. Feature learning by distant supervision for fine-grained implicit discourse relation identification[J]. Acta Scientiarum Naturalium Universitatis Pekinensis, 2019, 55(1): 91-97. (in Chinese)
- [20] GUO F, HE R, DANG J, et al. Working memory-driven neural networks with a novel knowledge enhancement paradigm for implicit discourse relation recognition[C]//Proceedings of the Conference on Artificial Intelligence. New York: AAAI, 2020: 7822-7829.
- [21] KURFALI M, ÖSTLING R. Let's be explicit about that: Distant supervision for implicit discourse relation classification via connective prediction[C]//Proceedings of the Workshop on Understanding Implicit and Underspecified Language. Stroudsburg: ACL, 2021: 1-10.
- [22] FU J L, ZHENG H L, MEI T. Look closer to see better:

Recurrent attention convolutional neural network for fine-grained image recognition[C]//Proceedings of the Conference on CVPR. Piscataway: IEEE, 2017: 4476-4484.

- [23] PRASAD R, DINESH N, LEE A, et al. The penn discourse treebank 2.0[C]//Proceedings of the Conference on Language Resources and Evaluation. Marrakech: LREC, 2008: 2961-2968.

#### 作者简介



**王秀丽** 男,中央财经大学信息学院/国家金融安全教育部工程研究中心教授,博士生导师.主要研究方向为金融科技与安全,大数据与人工智能.

E-mail: wangcufe@163.com



**金方焱** 男,中央财经大学信息学院硕士研究生.主要研究方向为自然语言处理.