

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^{1,2}, 金方焱²

(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: 隐式篇章关系识别研究因其难度大、普遍性高等特点是一项极具挑战性的任务。本文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。该模型同时兼顾了论元本身特征和论元间交互特征的作用, 并且分别进行了优化。论元编码部分整合了双向长短期记忆网络(Bidirectional Long Short-Term Memory, BiLSTM)和循环注意力卷积神经网络(Recurrent Attention Convolution Neural Network, RACNN), 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, 并且利用神经张量网络(Neural Tensor Network, NTN)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(Penn Discourse Treebank, PDTB)数据集上的实验结果表明, 该模型F1值均优于其他对比模型。

关键词: 隐式篇章关系识别; 双向长短期记忆网络; 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

中图分类号: TP391

文献标识码: A

文章编号: 0372-2112(XXXX)XX-0001-15

电子学报 URL: <http://www.ejournal.org.cn>

DOI: 10.12263/DZXB. xxxxxxxx

Implicit Discourse Relation Recognition Integrating Feature Coding and Phrase Interaction Perception

WANG Xiu-li^{1,2}, JIN Fang-yan²

(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 F1 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

1 引言

社会活动是篇章事件的基础, 篇章事件是社会活动的表现形式。在自然语言处理(Natural Language Processing, NLP)研究中, 自然语言数据显得尤为重要, 其中篇章作为语言信息一种不可或缺的表现方式, 受到

诸多研究者的高度关注。篇章文本信息量大, 价值高, 尤其是其中的隐式篇章关系包含着巨大的潜在利用价值。对篇章进行隐式篇章关系分析, 挖掘隐式篇章关系中隐含的重要信息, 进一步了解篇章间更深层的演化逻辑, 对自动问答^[1]、自动摘要^[2]、自动翻译^[3]等自然语言处理下游应用有着重要意义, 例如: 自动问答需要通

过整合篇章信息,针对相关提问提供符合逻辑的答复;自动摘要可以分析篇章关系以抽取文本中价值相对较高的语句作为选择;篇章连接词可以帮助提高自动翻译的性能等。

篇章一般指一个完整的语句或片段,其中包含许多内部构成元素如目标对象、事件内容等,篇章内部的逻辑语义即篇章关系,一般来说,篇章关系直接体现篇章中两个论元(Arg1和Arg2)间的关联关系。宾州篇章树库(Penn Discourse Treebank, PDTB)^[4]是目前应用最广泛的篇章语料库,其中 FullRawText 表示篇章, Arg1_RawText 和 Arg2_RawText 分别表示两个论元, Connective_RawText 表示篇章关系连接词。根据篇章间是否存在篇章关系连接词将篇章关系划分为显式篇章关系和隐式篇章关系。显式篇章关系表示论元间存在连接词,而隐式篇章关系则不存在关系连接词,相比于显式篇章关系其识别难度更高,是浅层篇章分析(Shallow Discourse Parsing)中最困难的部分之一,也是自然语言处理中一项极具挑战性的任务,旨在推断出篇章关系中缺少连接词的论元对的逻辑语义关系。因此,本文专注于隐式篇章关系的识别。

文本是一种无结构的数据,为了对篇章关系进行识别,需要将篇章关系文本数据转换为计算机能理解的结构化数据。而特征建模工具也从传统统计分析方法逐渐迭代到当前的深度学习方法。

在传统的统计分析研究中,学者利用特征工程从语料库中筛选出更有价值的特征信息输入到模型中进行训练以识别隐式篇章关系。特征工程往往没有固定的思维,根据研究领域的不同,其特征抽取方式也不尽相同。信息特征抽取工程需要丰富的领域背景知识以及大量的人工操作,进而才能从数据中挖掘有价值的特征信息,转换为计算机可理解的形式输入到模型中进行训练。这是传统统计分析研究的缺陷。另外,在模型训练之前需要先对数据进行预处理操作。其中,寻找合适的特征至关重要,这直接关系到最终模型的识别性能。特征类型可以分为离散型和连续型,对于离散型特征,需要将这些特征进行数值化;对于数值型数据,需要利用归一化对数据进行处理,以避免单位和尺度差异带来的偏向影响。随着特征数量不断增加,其维度也在不断变高,数据变得过于稀疏。通过引入特征筛选和降维可以避免特征维度上升所带来的维度灾难。不过,这样也会导致一些重要特征的丢失,进而影响隐式篇章关系识别性能。

NLP的最终目标是让机器更智能,以具备正确理解人类自然语言的能力,更有效地进行智能服务。在实现这个目标的过程中涉及到机器学习、大数据、深度学习等。首先,机器模型的学习训练需要大量有价值的数

据;其次,自然语言与语法信息密不可分,NLP利用句法分析、语义分析等相关语法分析技术能够帮助机器更好地学习到其中的价值信息。目前,NLP技术已经应用到了机器客服、智能助手、风险预警等多个领域。但目前对于篇章层次的研究相对较少,特别是其中涉及篇章间隐式逻辑关系的研究,在理论和实际应用研究^[5-8]中仍处于初级阶段。在隐式篇章关系识别研究方向,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)、注意力机制等模型来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。一般来说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信息,基于短语的注意力机制网络^[9]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($G(x) \rightarrow Mask$)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(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,捕获高维度的交互特征信息,进而提高隐式篇章关系识别效果。

2 相关研究

随着深度学习的快速发展,其在自然语言处理领域应用越来越广泛,基于深度学习的隐式篇章关系识别研究取得了不错的进展,其性能远超前于传统统计分析方法。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,借助显式篇章信息为隐式篇章信息带来的有用信息.

2.1 基于论元编码的方法

基于论元编码的方法关注于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.Zhang等^[10]最早将神经网络应用于隐式篇章关系识别,提出一种浅层卷积神经网络模型,该模型利用平均池化、最大池化和最小池化对论元对进行联合局部特征提取,单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.除了局部特征以外,部分学者还关注论元全局特征在篇章关系识别中的影响.Rutherford等^[11]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,凡子威等^[12]和Dai等^[13]利用BiLSTM对论元级上下文进行建模以捕获整合上下文信息的全局特征,进而提高篇章识别效果.为了同时考虑论元全局特征和局部特征的优势,Zhang等^[14]融合内部注意力模型和外部注意力模型,以充分关注模式整合了论元内部信息和外部知识信息;Popa等^[15]在特征编码阶段利用LSTM-CNNs结构挖掘论元充分语义信息;Guo等^[16]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,进而影响篇章关系识别效果.

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研究.徐扬等^[17]利用交互式注意力机制实现词级别信息交互和论元级别信息交互,进一步强化了论元之间的交互性;阮慧彬等^[18]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;Guo等^[19]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;徐昇等^[20]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,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

互研究,另外,部分学者还从其他角度进行了相关研究.Liu等^[21]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;He等^[22]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.Bai等^[23]整合词级、子词级以及句级论元信息,利用卷积神经网络进行论元交互建模,较大地提高了关系识别的性能.上述方法从词级、句级、几何等角度去理解论元间的语义信息,但忽略了论元短语级语义表示的考虑,Wu等^[9]研究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信息.

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.唐裕婷等^[24]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识别;Guo等^[25]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;Kurfali等^[26]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,并且选择评分最高的显式标记用作最终的隐式关系识别以提高模型识别性能.该类方法局限在于需要大量外部显式信息监督,且无法有效挖掘隐式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文编码的基础上,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,但目前短语层级的交互语义关系挖掘研究较少,一般来说短语更能体现论元间的真实语义关系;而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,本文将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 Arg1 和 Arg2 映射为词向量矩阵作为模型的输入。
- (2.1)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。
- (2.2)句法分析层:利用语法分析树解析论元 Arg1 和 Arg2 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(2.2)中输出的论元短语分

割矩阵将(2.1)中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 Arg1-Arg2 与 Agr2-Arg1 的注意力语义特征。

(4)NTN:将(3)中的注意力语义特征输入到神经张量网络中,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。

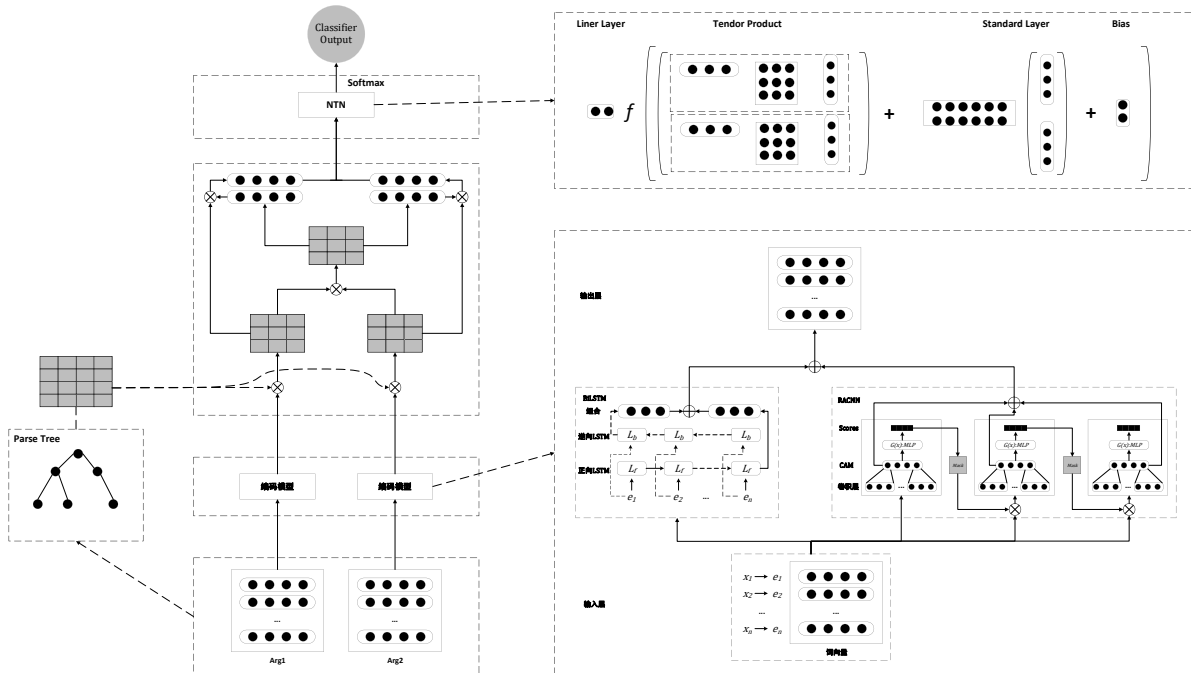


图1 融合特征编码和短语交互感知的隐式篇章关系识别模型

3.1 基于 BiLSTM-RACNN 的论元编码

3.1.1 BiLSTM

LSTM 是由 Hochreiter 等^[27]首次提出,用于解决梯度爆炸的问题.但 LSTM 只关心文本中当前词的历史信息而忽略了未来信息.因此,本文采用 BiLSTM 替代 LSTM,不仅解决了梯度问题,还考虑了当前词的前后文语义信息. BiLSTM 是一个正向 LSTM 和一个反向 LSTM 的组合,可以获取到整个文本序列的信息,在隐式因果关系抽取任务中可以充分挖掘文本全局特征,进而能更加准确地识别事件的关系类型.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 f_w 和反向结果 b_w ,连接 f_w 和 b_w 作为 BiLSTM 输出:

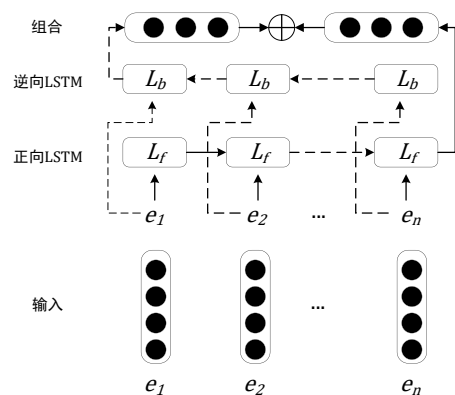


图2 BiLSTM

$$output_i = [f_w \oplus b_w] \quad (1)$$

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

其中: $output_i$ 表示词向量 e_i 的 BiLSTM 输出, H 表示 E_m 的 BiLSTM 输出。

3.1.2 RACNN 变体

RACNN 是由 Fu 等^[28]提出的一种卷积神经网络结构,应用于图像识别. 本文提出一种 RACNN 变体,如图 3 所示,将其应用于文本局部特征提取. 相比于用于文本局部特征提取的传统 CNN,该变体采用了迭代反馈机制 ($G(x) \rightarrow Mask$),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 $Em = \{e_1, e_2, \dots, e_n\}$;其次,用 $r \times m$ 的过滤器来对 Em 进行卷积,过滤器的数量为 u ,提取出 Em 局部特征:

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

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 Em 中从 i 到 $i+r-1$

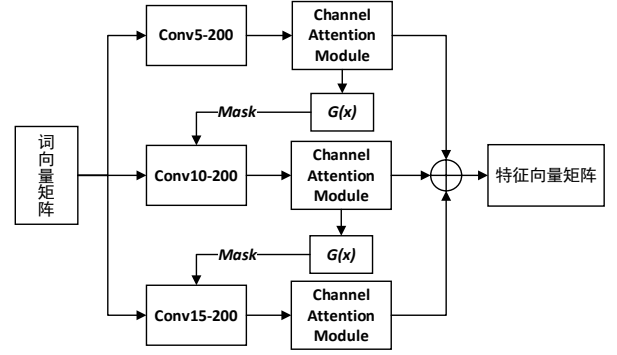


图3 RACNN 变体

$r-1$ 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 :

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

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

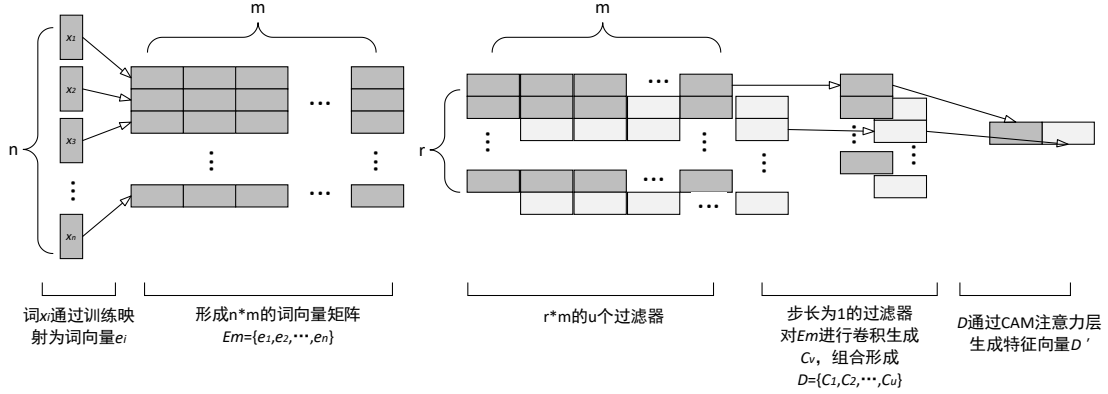


图4 RACNN 单次卷积

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 :

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

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

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

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

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

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 score_i ,最后生成一个矩阵 $Mask$,用于识别分数低的词:

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

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

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 $Mask$ 反馈到下一次卷积,卷积层的输入更新为 Em' :

$$Em' = Mask \bullet Em \quad (14)$$

如上述步骤迭代,直到三次卷积操作完成,将三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

在目前的研究成果中,基于论元交互的隐式篇章

关系是被模型拥有较好的性能,这些模型主要从词级、片段级、论元级或者段落级进行相互语义关系建模,而忽略了短语级的相互语义关系.短语是指由句法、语义和语用三个层面上组合起来的语言单位,它是大于词而又不成句的语法单位.一般来说短语能更好地表达论元的重要特征,进而合理体现论元间的语义关系信息,基于短语的注意力机制网络^[9]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.

3.2.1 词级交互注意力机制

Herzog等^[29]认为世界不是连续而被感知,反而是离散化的.他们提出了一种新的框架解释了人类视觉处理的时间分辨率有多快,各个物体的特征都被量化编码,最后同步地被意识以离散的方式进行感知.在这个感知的过程中,人在识别图像过程中包括收集信息和理解信息两个阶段.其实,在日常生活中,人们在阅读过程中也是需要反复阅读两个论句,然后才能发现其中的潜在关系.阅读两个论句的顺序不同,阅读者所

理解的信息也会有所差异,进而可能做出不同的判定.因此,Guo等^[19]认为可以从正反两个阅读顺序去挖掘两个论句之间的相互语义关系,即交互注意力.从词级角度出发,根据 Arg1 和 Arg2 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 Arg1 第*i*个词和 Arg2 中第*j*个词的语义关系分数,如下所示.

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

其中, $S \in R^{n_1 \times n_2}$, n_1 表示 Arg1 的词级长度, n_2 表示 Arg2 的词级长度.

3.2.2 句法分析树

句法分析树(Parse Tree)是指确定句子结构以及各层次句法成分之间关系的树状结构,它可以确定一个句子中哪些是短语,词的属性等.在本文中由 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".

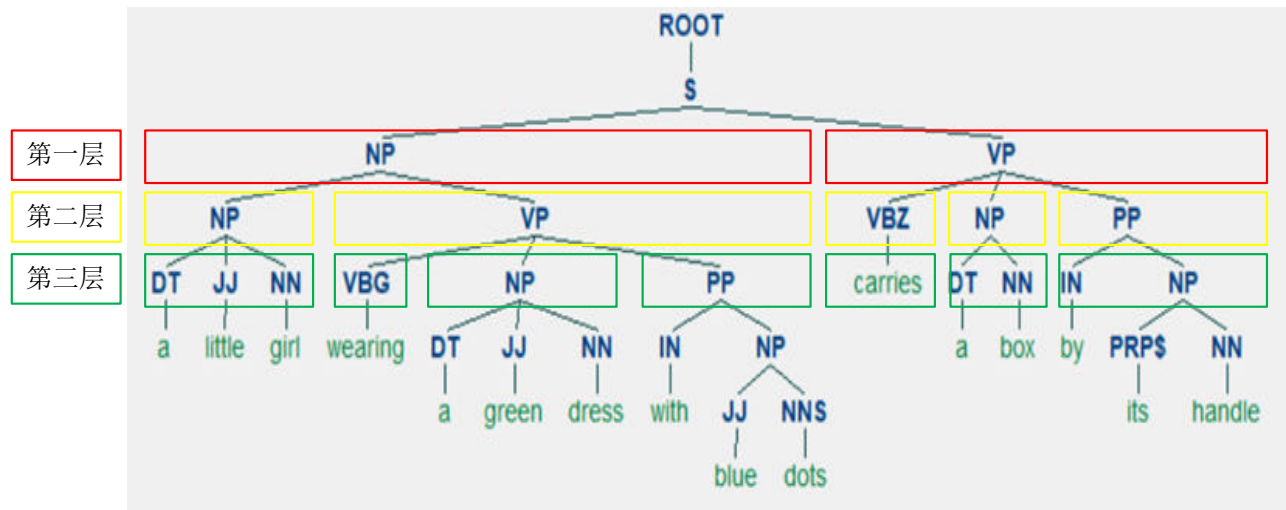


图5 句法分析树

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语句短语级表示,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

短语级交互注意力机制仅仅从论元单词角度进行

语义关联分析,而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.因此,本文设计了一种短语级交互注意力模型,能够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,使得能够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 Arg1 和 Arg2 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 Q1 和 Q2,然后通过矩阵乘法变换将 Arg1 和 Arg2 的隐藏层表示转换为短语级表示 P1 和 P2,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 Arg1-Arg2 和 Arg2-Arg1 的注意力语义分数,最后与 P1 和 P2 相结合分别形

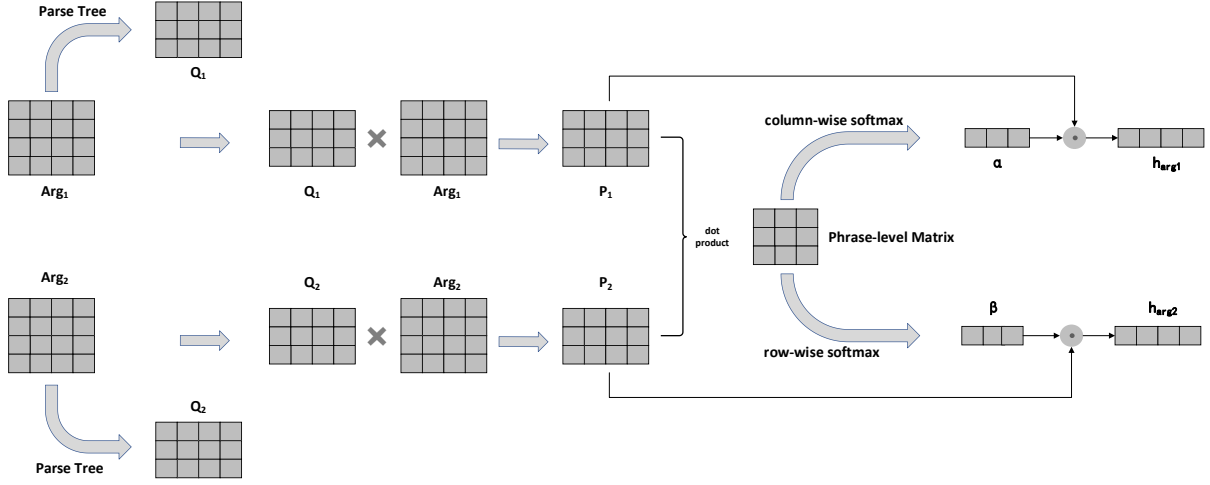


图6 短语级交互式注意力模型

成 Arg1 和 Arg2 对应的交互语义表示。

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.

捕获论元间交互模型的方法有双线性模型和单层网络等.双线性模型解决了弱论元向量相互作用的问题,通过双线性结构简单有效地整合了论元之间的交互.但是,双线性模型只能构建论元间线性交互模式而缺乏对非线性交互特征的挖掘.具体实现如下所示.

$$S_{bilinear}(h_{arg1}, R, h_{arg2}) = h_{arg1}^T W_R h_{arg2} \quad (17)$$

其中, $W_R \in R^{d \times d}$ 是该关系 R 分数的唯一系数.

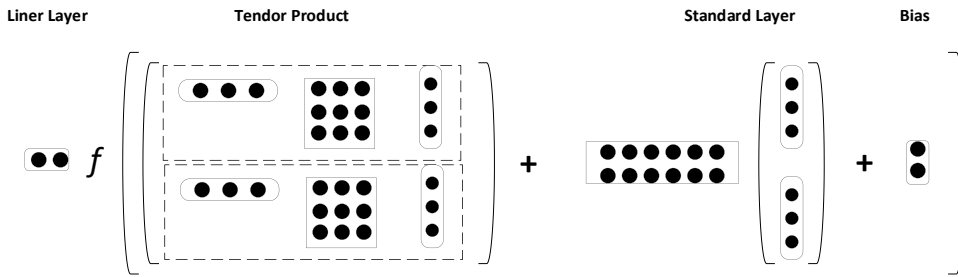


图7 神经张量网络模型

$$S(h_{arg1}, R, h_{arg2}) = u_R^T f(h_{arg1}^T W_R^{[1:k]} h_{arg2} + V_R \begin{bmatrix} h_{arg1} \\ h_{arg2} \end{bmatrix} + b_R) \quad (19)$$

其中, f 是标准非线性函数 \tanh , $W_R^{[1:k]} \in R^{d \times d \times k}$ 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. $h_{arg1}^T W_R^{[1:k]} h_{arg2}$ 指的是双线性张量项,生成一个向量 $h \in R^{k \times 1}$,其中 h 的每个元素

单层网络是一种通过标准单一网络的非线性模式来隐式地连接论元,将两个论元对应参数合并到一起,解决了两个论元参数不相互作用的问题,缺点是只提供了论元之间的弱交互关系,不能挖掘丰富的特征.具体实现如下所示.

$$S_{single}(h_{arg1}, R, h_{arg2}) = u_R^T f(V_R \begin{bmatrix} h_{arg1} \\ h_{arg2} \end{bmatrix} + b_R) \quad (18)$$

其中, f 是标准非线性函数 \tanh , $u_R \in R^{k \times 1}$, $V_R \in R^{k \times d}$, $b_R \in R^{k \times 1}$.

神经张量网络(NTN)通过整合了上述方法的观点和优势,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.

$h_i = h_{arg1}^T W_R^{[i]} h_{arg2}$ ($i=1, 2, 3, \dots, k$), $W_R^{[i]}$ 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: $u_R \in R^{k \times 1}$, $V_R \in R^{k \times d}$, $b_R \in R^{k \times 1}$.

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,本文采用 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,本文

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(Arg1和Arg2)及其类别标签 y , 则其损失函数定义如下:

$$L(\hat{y}, y) = -\sum_j^C y_j \log(\Pr(\hat{y}_j)) \quad (20)$$

其中, $\Pr(\hat{y}_j)$ 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. 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, 大多数学者利用 l_1 来进行求解, 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 θ_0 , 其余参数记为 θ_1 , 具体最小化训练目标如下:

$$J(\theta) = \frac{1}{n} \sum_{r=1}^n L(\hat{y}^r, y^r) + \lambda_0 \|\theta_0\| + \lambda_1 \|\theta_1\|_2^2 \quad (21)$$

其中, λ_0 表示 l_1 正则项参数, λ_1 表示 l_2 正则项参数.

4 实验与分析

4.1 数据集

PDTB是语言联盟提供的, 当前数据量最大、应用最广的篇章关系语料库, 并已经成为了篇章关系识别的基准库. PDTB的原始数据来源于宾州华尔街日报, 于2006年和2008年相继发布了PDTB1.0和PDTB2.0, 第二个版本在第一个版本的基础上做了部分调整. 在目前的隐式篇章关系识别研究领域中, 众多学者均选择PDTB2.0进行相关研究. 考虑到该语料库的广泛性以及为了更好地和其他研究成果进行对比, 本文选择该语料库进行一系列相关研究, 通过对比实验验证本文模型在该语料库上的表现性能.

PDTB包含40600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Arg1和Arg2之间潜在的逻辑关系.

(e. g. 1) Arg1: There's no question that some of those workers and managers contracted asbestos-related diseases.

Arg2: But you have to recognize that these events took place 35 years ago.

(wsj_0003)

(e. g. 2) Arg1: Before boarding the buses again.

Arg2: [Implicit=But] This time, it was for dinner and dancing -- a block away.

(wsj_0010)

PDTB关系类型一共有5种, 除了Explicit(显式)和Implicit(隐式)外, 还存在另外三种关系类型: 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相关研究, 为了能更好地和前人进行比较, 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	18459
Implicit	16224
AltLex	624
EntRel	5210
NoRel	254
Total	40600

表2 PDTB第一层级关系数据分布

"CLASS"	Explicit	Implicit	AltLex	Total
"TEMP."	3612	950	88	4650
"CON."	3581	4185	276	8042
"COM."	5516	2832	46	8394
"EXP."	6424	8861	221	15506
Total	19133	16828	634	36592

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中的统计数据.

4.2 评价指标

为了更好地展现预测效果, 由混淆矩阵衍生出几个评测指标: Precision(精确率)、Recall(召回率)、F1-Score(F1值)和Accuracy(准确率), 计算公式如下所示.

1) Precision, 表示所有被识别为正样本中原本为正样本的比例. 精确率越高, 被分类正确的正样本比率越高.

$$Precision = TP / (TP + FP) \quad (22)$$

2) Recall, 表示原本为正样本的样本中被识别为正样本的比例. 召回率越高, 原本是正样本被识别为正样本的比率越高.

$$Recall = TP / (TP + FN) \quad (23)$$

3) F1-Score, 为了能够兼顾精确率和召回率进而设置的一个综合指标, 普适性较强.

$$F1 = \frac{2 * Precision * Recall}{Precision + Recall} \quad (24)$$

4) Accuracy, 表示预测正确的样本占总样本的比例, 即正样本被预测为正样本和负样本被预测为负样本所有样本占总样本的比例.

$$Accuracy = (TP + TN) / (P + N) \quad (25)$$

通常 F1 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, 因此本文也采用 Precision、Recall 和 F1 三个指标进行对比实验. 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, 计算方式有三种: Micro-Average (微平均)、Macro-Average (宏平均) 以及 Weighted (加权). Micro-Average 指先将多个混淆矩阵所对应的位置求平均, 然后计算 Micro-Precision、Micro-Recall 和 Micro-F1; Macro-Average 指先对每个混淆矩阵计算相应的 Precision、Recall 和 F1, 然后将这些对应指标求平均; Weighted 指对 Macro-Average 中每个混淆矩阵对应的指标给一个相应的权重, 这个权重一般是由混淆矩阵对应类的数据量占总体数据量的比例决定的, 适合各类别分布不均衡的数据集. 根据本文语料库的数据样本分布情况, 本文模型评价指标选择 Weighted 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	16G
CPU	Intel Core i7-6700	开发环境	Tensorflow 1.4.1
GPU	GeForce GTX 2080Ti		

4.4 参数选择

参照先前的研究成果, 为了能够更好地进行对比, 本模型配置了多个参数设置, 如表 4 所示.

其中, 隐藏层是指 LSTM 中的 hiddenlayer; dropout_ratio 是指特征过滤力度, 一般设置为 0.3, 能有效防止过拟合; K_slices 表示 NTN 中的切片数量; Masksize 表示 RACNN 中掩膜大小, 帮助提高局部卷积质量.

为了模型训练输入的统一标准, 需要将输入论元长度设定为 50, 如果原本论元长度大于 50, 则进行截

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	Adadelta	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

剪, 如果原本论元长度小于 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 模块中滑动窗口过大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, 本模型采用 Adadelta 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	48.07
0.5	49.40
0.3	49.99
0.1	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	F1(%)
1	47.16
2	48.08
3	49.99
4	49.22

4.5 对比方法

为了更好地对比本文模型性能, 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:该模型是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:该模型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:该模型利用句法分析树将基于词级论元交互模型改进为短语级论元交互模块,进而测试短语级论元语义交互相比于词级论元交互的优势。

(4)BiLSTM+RACNN+IA+NTN:该模型在词级论元交互模型的基础上增加了神经张量网络,用来测试神经张量网络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:Guo等^[16]提出一种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。该模型从特征提取的角度入手,验证了模型的有效性,可用于对比实验。

(2)Syntax+self-attention:凡子威等^[12]提出一种融合双仿射深度神经网络和BiLSTM的隐式篇章关系分类模型,该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。该模型考虑了文本依存句法信息特征,能提升分类任务的效果,可用于对比实验。

(3)TLAN:徐昇等^[20]提出了一个三层注意力网络模型用于隐式篇章关系识别,该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,最终生成论元的最终表示。

(4)Bai's: Bai等^[23]提出一种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: Liu等^[21]提出一种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,该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: Guo等^[19]提出一种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以提高隐式篇章关系识别效果。

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如表7和表8所示。

表7 PDTB四分类实验结果(%)

方法	Precision	Recall	F1
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 ^[16]	47.59	51.70	47.79
Syntax+self-attention ^[12]	46.67	54.31	47.42
TLAN ^[20]	52.40	52.75	47.32
TIASL ^[19]	49.54	53.39	48.95
本文模型	49.27	53.46	49.99

由表7可知,加入词级论元交互注意力机制的方法(BiLSTM+RACNN+IA)比BiLSTM+RACNN方法的F1值高0.7%;加入短语级论元交互注意力机制的方法(BiLSTM+RACNN+ParseMatrix+IA)比BiLSTM+RACNN+IA方法的F1值高0.67%;加入神经张量网络的方法(BiLSTM+RACNN+IA+NTN)比BiLSTM+RACNN+IA方法的F1值高0.03%。在于其他研究成果的对比中,本文模型

相比Guo等提出的DC-BCNN模型F1值提高2.2%;相比于凡子威等提出的Syntax+self-attention模型F1值提高了2.57%;相比于徐昇等提出的TLAN模型F1值提高了2.67%;相比于Guo等提出的TIASL模型F1值提高了1.04%。

横向比较中,BiLSTM+RACNN+IA方法比BiLSTM+RACNN方法增加了词级论元交互模块,不仅仅考虑了

论元本身的语义特征,还考虑了论元之间的相互语义关系,在隐式篇章关系识别过程中具有一定的促进作用;BiLSTM+RACNN+ParseMatrix+IA方法与BiLSTM+RACNN+IA方法相比,改进在于将词级论元交互模块转换为短语级论元交互,一般来说短语相比于词语对论元语义的表现力更强,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;BiLSTM+RACNN+IA+NTN方法在BiLSTM+RACNN+IA方法的基础上增加了神经张量网络模块,神经张量网络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,进而提高了隐式篇章关系识别效果.

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,进而能提高隐式篇章关系识别性能;与Syntax+self-attention模型相比,Syntax+self-attention模型的优势在于输入中加入了依存句法信息,但是这也仅仅是在论元本身特征抽

取上进行提升,缺乏论元之间交互特征的捕获,导致于无法获取论元间深层次的交互关系模式,进而影响隐式篇章关系识别效果;与TLAN模型相比,TLAN模型利用自注意力机制捕获论元本身的特征,获取论元中内部各个词之间的关联关系,并且通过交互注意力机制及带有记忆的注意力机制对论元进行处理,不过在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,而本文在解决了这些问题的同时,也进一步考虑了交互特征向量更深层次的关系模式挖掘;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.

表8 PDTB二分类实验结果(%)

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 ^[16]	32.00	48.66	67.80	25.08
Syntax+self-attention ^[12]	32.30	47.40	68.99	23.21
TLAN ^[20]	32.52	48.98	67.73	22.22
TIASL ^[19]	32.46	50.53	68.87	29.19
本文模型	32.55	51.85	69.07	30.18

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(F1值),可以得出如下结论:

(1)在四种关系分类中,所有模型在比较关系和时序关系上的表现效果相对比较差.其可能的原因是这两种关系在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,进而识别效果不佳.从四种关系样本分布也可以看出,随着训练样本的增加,模型识别的准确率也在上升,这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,从横向比较角度来看,后两种消融模型(BiLSTM+RACNN+ParseMatrix+IA和BiLSTM+RACNN+IA+NTN)均比前两种消融模型(BiLSTM+RACNN和BiLSTM+RACNN+IA)表现性能更好,分别验证了短语级交互注意力机制和神经张量网络在隐式篇章识别过程中的重要作用.从纵向比较角度来看,TIASL模型在四种模型中表现最佳,F1值分别

达到了50.53%和68.87%,该模型重点关注论元交互部分,利用词级交互和神经张量网络深层次挖掘论元交互语义特征.DC-BCNN和Syntax+self-attention模型都是针对于论元编码部分的研究,缺乏对论元交互进一步挖掘;TLAN模型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,由于本身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一种短语级交互注意力机制,旨在能够更好地挖掘论元短语间的真实语义信息,并且利用神经张量网络高维度挖掘交互关系模式,分别在两种关系实验上F1值达到了51.85%和69.07%,较四种消融模型和四种对比模型的性能更好.

4.7 模型效率分析

模型性能优劣主要取决于实验效果好坏和时间效率高低,为了验证本文模型的实际性能,本节从模型训练过程中准确率和损失值变化来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.

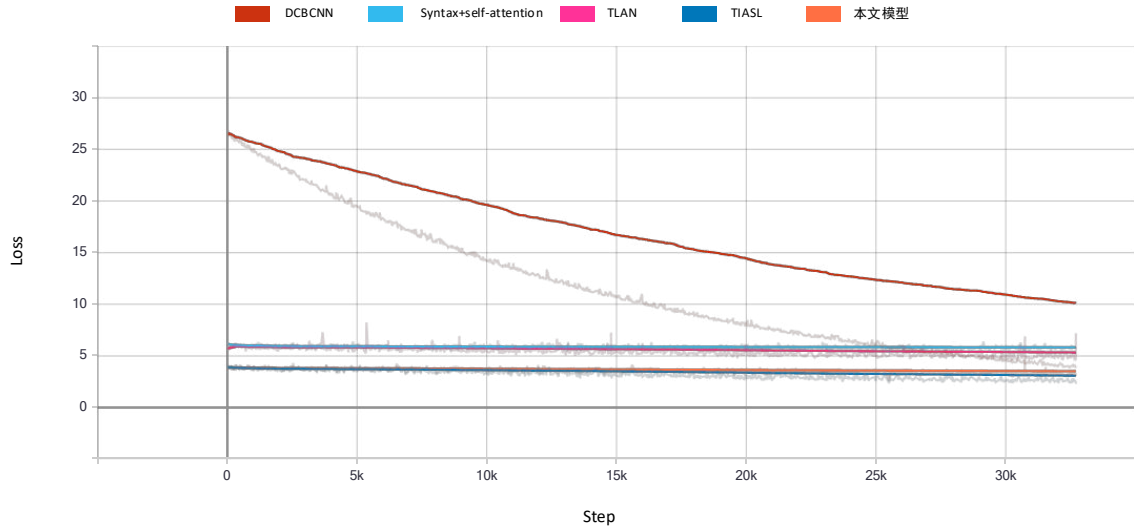


图 8 训练步数-损失值(训练集)

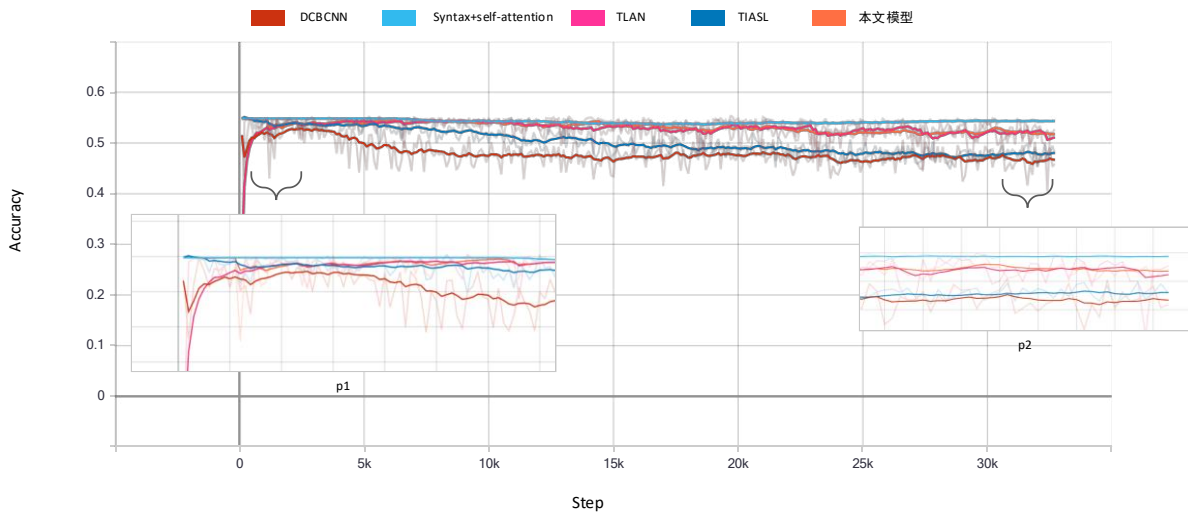


图 9 训练步数-准确率(验证集)

由图 8 可知, 在训练过程中, 除了 DC-BCNN 模型外, 其他模型训练集损失值均快速地收敛, 并且趋于平稳状态, 该状况出现原因可能是 DC-BCNN 模型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 和 TLAN) 以及等级三 (本文模型和 TIASL), 其中, 等级三损失值最小, 本文模型相比于 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 模型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 模型, 最后是 DC-

BCNN 和 TLAN 模型, 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 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.

基于多种神经网络的基础上, 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, 同时还利用神经张量网络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里的最高效率, 但仍然处于前列位置, 实现了以最小的时间代价换取最多的实验

性能.

5 结束语

本文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.该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和短语级交互注意力机制+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.在基于PDTB数据集的实验表明,本文模型在四分类和二分类任务中均优于其他对比模型,充分说明了本文提出的模型在隐式篇章关系识别任务上的有效性.

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,论元编码部分主要基于传统的RNN、CNN模型,本文也只是在此基础上进一步改进,但缺乏利用其他最新的变种模型进行探索,比如Bi-GRU、Tree-LSTM等.另外,在论元交互部分,本文设计了一种短语级交互注意力机制,相比于前人研究中使用的传统交互注意力机制效果更好,但考虑到论元中短语重要程度各不相同,根据短语重要性在交互过程中加入加权影响,是一个值得更加深入优化的研究方向,旨在进一步帮助提高隐式篇章关系识别性能.

本文主要是从论元本身、论元交互的特征表示角度入手,研究不同语义特征对隐式篇章关系识别的影响.模型虽然相比于对比模型有一定的提升,但还未达到理想状态,因此未来有如下几个可能的研究方向.

(1)考虑将论元外部特征加入到模型输入中,研究表明,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.

(2)由于本文采用传统的BiLSTM获取论元全局特征,可以尝试Bi-GRU、Tree-LSTM等多种不同的模型进一步对比特征抽取效果.

(3)在论元交互部分,本文只考虑了词级和短语级交互的比较,因此,词对、短语对等层级也是值得研究的方向.

(4)为了提高模型的普适性,增加中文数据集进行对比实验研究.

(5)本文仅考虑到篇章内部的语义关系信息,篇章间上下文信息同样会对模型产生一定的影响.

总之,隐式篇章关系识别在自然语言处理领域具有很高的研究价值,值得人们不断地、更深入地探索与研究.

参考文献

- [1] YOSHIDA Y, SUZUKI J, HIRAO T, et al. Dependency-based discourse parser for single-document summarization [C]//Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). Doha, Qatar: ACL, 2014: 1834-1839.
- [2] SUN M, CHAI J Y. Discourse processing for context question answering based on linguistic knowledge[J]. Knowledge-Based Systems, 2007, 20(6): 511-526.
- [3] LI J J, CARPUAT M, NENKOVA A. Assessing the discourse factors that influence the quality of machine translation[C]//Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers). Baltimore, MD: ACL, 2014: 283-288.
- [4] PRASAD R, DINESH N, LEE A, et al. The penn discourse treebank 2.0[C]//Proceedings of the International Conference on Language Resources and Evaluation. Marrakech, Morocco: LREC, 2008: 2961-2968.
- [5] ZHOU P, SHI W, TIAN J, et al. Attention-based bidirectional long short-term memory networks for relation classification[C]// Proceedings of the 54th Annual Meeting of the ACL. Stroudsburg, PA: ACL, 2016:207-212.
- [6] 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 Zhi-xin, SUN Ya-ru, TANG Su-qin, et al. Dual attention guided graph convolutional networks for relation extraction[J]. Acta Electronica Sinica, 2021, 49(2): 315-323. (in Chinese)
- [7] 冯建周, 宋沙沙, 王元卓, 等. 基于改进注意力机制的实体关系抽取方法[J]. 电子学报, 2019, 47(8): 1692-1700.
FENG Jian-zhou, SONG Sha-sha, WANG Yuan-zhuo, et al. Entity relation extraction based on improved attention mechanism[J]. Acta Electronica Sinica, 2019, 47(8): 1692-1700.(in Chinese)
- [8] ZHANG Y, MENG F, LI P, et al. Context tracking network: graph-based context modeling for implicit discourse relation recognition[C]//Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Online: ACL, 2021: 1592-1599.
- [9] WU W, WANG H, LIU T, et al. Phrase-level self-attention networks for universal sentence encoding[C]//Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. Brussels, Belgium: ACL, 2018: 3729-3738.
- [10] ZHANG B, SU J, XIONG D, et al. Shallow convolutional

- neural network for implicit discourse relation recognition [C]//Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing. Lisbon, Portugal: ACL, 2015: 2230-2235.
- [11] RUTHERFORD A, DEMBERG V, XUE N. A systematic study of neural discourse models for implicit discourse relation[C]//Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers. Valencia, Spain: ACL, 2017: 281-291.
- [12] 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 Zi-wei, ZHANG Min, LI Zheng-hua. BiLSTM-based implicit discourse relation classification combining self-attention mechanism and syntactic information[J]. Computer Science, 2019, 46(5): 214-220.(in Chinese)
- [13] DAI Z, HUANG R. Improving implicit discourse relation classification by modeling inter-dependencies of discourse units in a paragraph[C]//Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics : Human Language Technologies. New Orleans, Louisiana: ACL, 2018: 141-151.
- [14] 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.
- [15] Popa D N, Perez J, Henderson J, et al. Implicit discourse relation classification with syntax-aware contextualized word representations[C]//Proceedings of the Thirty-Second International Florida Artificial Intelligence Research Society Conference. Sarasota, Florida: AAAI Press, 2019: 203-208.
- [16] 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.
- [17] 徐扬, 周文, 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.
YANG Xu, ZHOU Wen-xuan, RUAN Hui-bin, et al. Hierarchical representation for implicit discourse relation recognition[J]. Journal of Nanjing University(Natural Sciences), 2019, 55(6): 1000-1009.(in Chinese)
- [18] 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 Hui-bin, XU Yang, SUN Yu, et al. Stacked-attention based implicit discourse relation recognition[J]. Journal of Shanxi University(Natural Sciences), 2020, 43(3): 508-516.(in Chinese)
- [19] 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 27th International Conference on Computational Linguistics. Santa Fe, New Mexico: ACL, 2018: 547-558.
- [20] 徐, 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 Sheng, WANG Ti-shuang, LI Pei-feng, 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)
- [21] 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 29th International Joint Conference on Artificial Intelligence. Amsterdam: Elsevier, 2020:3830-3836.
- [22] HE R, WANG J, GUO F, et al. Transs-driven joint learning architecture for implicit discourse relation recognition [C]//Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. Online: ACL, 2020: 139-148.
- [23] BAI H, ZHAO H. Deep enhanced representation for implicit discourse relation recognition[C]//Proceedings of the 27th International Conference on Computational Linguistics. Santa Fe, New Mexico: ACL, 2018: 571-583.
- [24] 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 Yu-ting, LI Yan-bin, LIU Lu, 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)
- [25] 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 AAAI Conference on Artificial Intelligence. New York: AAAI Press, 2020: 7822-7829.
- [26] KURFALI M, ÖSTLING R. Let's be explicit about that:

Distant supervision for implicit discourse relation classification via connective prediction[C]//Proceedings of the 1st Workshop on Understanding Implicit and Underspecified Language (UnImplicit 2021). Bangkok, Thailand (online): ACL, 2021: 1-10.

- [27] HOCHREITER S, SCHMIDHUBER J. Long short-term memory[J]. Neural Computation, 1997, 9(8):1735-1780.
- [28] 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 CVPR. Piscataway, NJ: IEEE, 2017:4438-4446.
- [29] HERZOG M H, KAMMER T, SCHARNOWSKI F. Time slices: what is the duration of a percept?[J]. PLoS biology, 2016, 14(4): 1-12.

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.



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.