

DRE-3DC: 基于三维表征建模的篇章级关系抽取模型

王宇¹, 王震¹, 温立强¹, 李伟平¹, 赵文²

(1. 北京大学软件与微电子学院, 北京 100871; 2. 北京大学软件工程国家工程研究中心, 北京 100871)

摘要: 篇章级关系抽取任务旨在从非结构化文档的多个句子中提取事实, 是构建领域知识库和知识问答应用的关键环节, 相较于句子级关系抽取, 该任务既要求模型能够基于文档结构特征捕获实体间的复杂交互, 还要应对严重的关系类别长尾分布问题. 现有基于表格的关系抽取模型主要对文档进行“实体/实体”二维建模, 采用多层卷积网络或局部注意力机制提取实体间的交互特征, 由于未显式对关系语义进行解耦建模, 使得模型无法避免类别重叠影响和捕获关系的方向性特征, 导致缺乏实体交互的充分语义信息. 针对上述挑战, 本文提出了一种基于三维表征建模的篇章级关系抽取模型 DRE-3DC (Document-Level Relation Extraction with Three-Dimensional Representation Combination Modeling), 对二维表格建模方式进行扩展, 形成“实体/实体/关系”三维表征建模, 采用基于形变卷积的三重注意力机制有效区分和聚合不同语义空间下的实体间及实体与关系的交互表征, 自适应地增强模型对文档结构特征的聚合. 同时, 采用多任务学习方法增强模型对文档整体关系类别组合的感知来缓解篇章级关系抽取任务中的关系类别长尾分布问题. 在 DocRED 和 Revisit-DocRED 两个篇章级关系抽取数据集上进行的实验结果表明, DRE-3DC 模型性能良好, 并通过消融实验、对比分析和实例分析, 验证了本文所提方法的有效性.

关键词: 篇章级关系抽取; 三维表征; 三重注意力; 形变卷积网络; 多任务学习

基金项目: 国家重点研发计划 (No.2020YFC0833300)

中图分类号: TP391.1

文献标识码: A

文章编号: 0372-2112(2024)09-2950-11

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

DOI: 10.12263/DZXB.20221187

DRE-3DC: Document-Level Relation Extraction with Three-Dimensional Representation Combination Modeling

WANG Yu¹, WANG Zhen¹, WEN Li-qiang¹, LI Wei-ping¹, ZHAO Wen²

(1. School of Software and Microelectronics, Peking University, Beijing 100871, China;

2. National Engineering Research Center for Software Engineering, Peking University, Beijing 100871, China)

Abstract: The task of document-level relation extraction aims to extract facts from multiple sentences of unstructured documents, which is a key step in the construction of domain knowledge graph and knowledge answering application. The task requires that the model not only capture the complex interactions between entities based on the structural features of documents, but also deal with the serious long-tail category distribution problem. Existing table-based relation extraction models try to solve this issue, but they mainly model documents in two-dimensional “entity/entity” space, and use multi-layer convolutional network or restricted self-attention mechanism to extract the interaction features between entities, which cannot avoid the influence of category overlap and capture the directional features of relationships, resulting in the lack of decoupled semantic information of interaction. For the above challenges, this paper proposes a new document-level relation extraction model, named DRE-3DC (Document-Level Relation Extraction with Three-Dimensional Representation Combination Modeling), in which the “entity/entity” modeling extend to the form of three-dimensional “entity/entities/relationship” modeling method. Based on the deformable convolution in triple attention mechanism, the model effectively distinguishes and integrates the interaction features under different semantic space and adaptively captures the document structural features. At the same time, we propose a multi-task learning method to enhance the perception of relation category combination of documents to alleviate the long-tail distribution problem. The experimental results reveal better score on DocRED

and Revisit-DocRED dataset respectively. The effectiveness of the proposed method was verified by ablation experiment, comparative analysis and example analysis.

Key words: document-level relation extraction; three-dimensional representation; triplet attention; deformable convolution; multi-task learning

Foundation Item(s): National Key R&D Program of China (No.2020YFC0833300)

1 引言

篇章级关系抽取^[1]任务的目标是在给定含有 N 个句子的文档 $D = \{s_i\}_{i=1}^N$, p 个实体的实体集合 $\varepsilon = \{e_i\}_{i=1}^p$ 以及关系类别集合 R 的基础上,从文档中抽取所有符合约束条件的实体关系三元组 $\{(e_i, r, e_j) | e_i, e_j \in \varepsilon, r \in R\}$, 其中 $s_i = \{w_j\}_{j=1}^M$ 表示文档中由 M 个单词 w 组成的第 i 个句子, $e_i = \{m_j\}_{j=1}^Q$ 表示由 Q 个实体提及 m 组成的第 i 个实体. 由于实体对间可能含有多种语义关系,通常将篇章级关系抽取任务视作多分类问题,相较于句子级关系抽取^[2],篇章级关系抽取需要模型具有更强的建模和推理能力来有效捕捉篇章文本中存在的句内和句间的实体复杂交互,即相关实体在不同语义空间下的关联,来确定目标实体对所属的关系类别,如图 1 所示,图中多个部分表示同一实体的不同提及,具有关系的相关实体可以来自不同句子,通过桥实体连接之间的语义. 例如,为了确定 Ninoy Aquino International Airport 与 Philippines 之间的关系,需要借助桥实体 Manila 进行推理,其中 Manila 与 Philippines 之间的关系 1 对确定 Ninoy Aquino International Airport 与 Philippines 之间的关系是有益的,而关系 2 对推理所起的作用则是有害的.

<p>AirAsia Zest <节选自维基百科: 亚洲激爽航空公司> [1] Zest Airways, Inc., operated as AirAsia Zest(formerly Asian Spirit and Zest Air), was a low-cost airline based in Pasay City, Metro Manila in the Philippines. [2] It operated scheduled domestic and international tourist services, mainly feeder services linking Manila and Cebu with 24 domestic destinations in support of the truck route operations of other airlines. [3] In 2013, the airline became an affiliate of Philippines AirAsia operating their brand separately. [4] Its main base was Ninoy Aquino International Airport, Manila. [5] The airline was founded as Asian Spirit, the first airline in the Philippines to be run as a cooperative. [6] On August 16, 2013, the Civil Aviation Authority of the Philippines (CAAP), the regulating body of the Government of the Republic of the Philippines for civil aviation, suspended Zest Air flights until further notice because of safety issues. [7] Less than a year after AirAsia and Zest Air's strategic alliance, the airline was rebranded as AirAsia Zest. [8] The airline was merged into AirAsia Philippines in January 2016.</p>
<p>头实体: Manila/Metro Manila 尾实体: Philippines 关系⁽¹⁾: located in the administrative territorial entity (Wikidata ID:131) 关系⁽²⁾: country (Wikidata ID:17) 证据相关句子: [1]</p>
<p>头实体: Philippines 尾实体: Manila/Metro Manila 关系⁽³⁾: contains administrative territorial Entity (Wikidata ID:150) 证据相关句子: [1]</p>
<p>头实体: Ninoy Aquino International Airport 尾实体: Manila/Metro Manila 关系⁽⁴⁾: located in the administrative territorial entity (Wikidata ID:131) 证据相关句子: [4]</p>
<p>头实体: Ninoy Aquino International Airport 尾实体: Philippines 关系⁽⁵⁾: located in the administrative territorial entity (Wikidata ID:131) 证据相关句子: [1][4]</p>

图 1 篇章级关系抽取任务实例

目前,篇章级关系抽取模型主要分为两大类:基于序列的模型^[3-9]和基于结构的模型^[10-16]. 其中,基于结构的模型又可细分为:(1)基于图结构的模型^[10-12]. 这类方法首先采用启发式规则^[10-12]或篇章解析工具^[13]构建篇章级图结构,然后利用图神经网络的消息传递框架建模和丰富实体对间关联.(2)基于表格结构的模型. 这类方法将图像语义分割^[14]的思想引入到篇章级关系抽取任务中,首先按照实体在篇章中出现的先后顺序排列成二维实体表格,然后利用多层卷积神经网络^[13]或利用注意力机制^[15,16]来建模实体间的关联.

然而,当前基于表格的建模工作在捕捉实体间交互表征时,忽略了由于二维建模方式导致的关系重叠和关系语义方向性问题. 同一实体对间同时可能存在多种关系,且不同关系对模型推理所起作用不同,此外同一实体在不同关系中的语义作用也有所不同,这些都带来有效捕捉实体交互特征的困难. 如图 2 所示,针对实体 E3 和 E4,当 E3 作为头实体、E4 作为尾实体时,二者间存在 131(类别编号)和 17 两种不同的关系类别;在 E3 与 E4 有关的 131 和 17 关系类别中, E3 作为头实体,而在 150 关系类别中, E3 则作为尾实体.

Entity	E0	E1	E2	E3	E4
E0			159		17
E1			131		17
E2				131	17
E3			150		17 131
E4				150	

图 2 二维表格建模方式存在的关系重叠和方向性问题

此外,篇章级关系抽取数据集相较于句子级抽取数据集存在更加严重的关系类别长尾分布问题,如图 3 所示,在 DocRED 数据集 96 种关系类别中最频繁的关系占总三元组的 55.12%, 60 种关系的训练三元组出现频次小于 200. 现有工作往往采用逐个对实体对预测其关系类别的方式,长尾分布问题必然导致模型对文档整体关系表达的感知不平衡,使得模型表现不佳^[17].

针对上述问题,本文提出:

(1)在原二维实体对表格^[14]的基础上,扩展得到

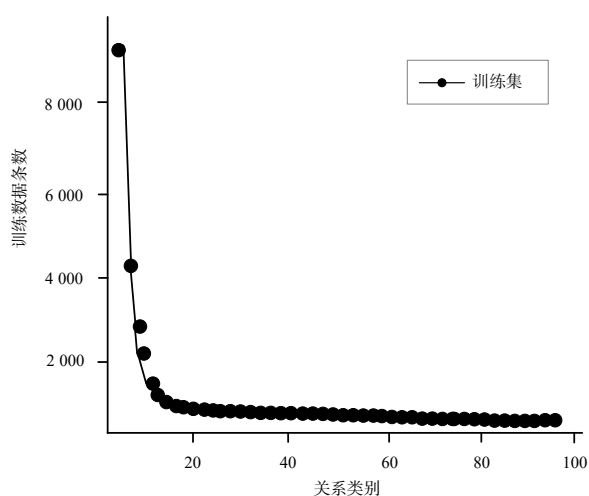


图3 DocRED数据集的长尾分布

“实体/实体/关系”三维表格；

(2)通过基于形变卷积的三重注意力模块区分和聚合不同语义空间下的交互表征,自适应地捕获篇章结构信息；

(3)通过多任务学习方式解决长尾分布问题,增强模型对篇章整体关系表达的感知能力。

在 DocRED^[18]和 Revisit-DocRED^[19]两个篇章级关系抽取数据集上进行了实验,实验结果表明 DRE-3DC 模型性能良好,并且通过消融、实例以及误差分析验证了本文所提法的有效性。

通常关系抽取任务中的关系泛指任何存在语义关联的实体对间的语义类别,本文提出将二维表格建模方式扩展为“实体/实体/关系”三维表征的建模方式,则将这种语义类别进一步细化,包括了重叠特性和方向特性,更能有效表示在特定模式约束下导致的实体对间语义关联。而且,这种三维表征建模方式具有一定程度的泛化性,对于未考虑关系类别重叠特性和方向特性的相关工作都有一定的借鉴和参考价值。

2 相关工作

2.1 基于序列的模型

基于序列的模型在篇章级关系抽取任务上的优点是序列更接近文档上下文表示形式,模型多采用 Transformer 架构,可以隐式建模长距离依赖性,并结合注意力机制进行推理。文档信息以序列形式一次性输入,可以利用前馈网络并行计算提高模型计算速度。注意力机制在序列模型中应用普遍,用来辨别在整个文档推理表示中的重要句子信息,捕捉关键上下文语义信息。基于序列的模型可以同时多个句子进行阅读、推理和聚合,这样就充分利用了多粒度的推理信息,即实体级、句子级和文档级,具备多句联合推理能力。但受

限于模型的序列架构,在直观表示多跳推理的实体关系结构方面会有所欠缺。

2.2 基于结构的模型

基于图结构的方法能够清楚表示句内实体关系推理和句间实体关系推理;图结构方法可以解决关系重叠和实体重叠问题(共指),能够利用共指信息为推理提供先验知识;通过建立跨句子实体间的联系,从而解决 RNN-based 编码器模型捕获长依赖信息时的不足。但是,图结构节点和边的设计要求比较高,比较复杂精巧;由于文本本身语言序列结构和图网络的异构性,图结构方法将上下文推理阶段和结构推理阶段隔离开来,这意味着上下文表示不能受益于模型结构指导。U-net^[20]是一种神经网络结构,主要用于图像语义分割。U-net 模型的基本结构由两条路径组成,一条路径是收缩路径,也被称为编码器或分析路径,它类似于常规的卷积网络,提供分类信息;第二种是扩展路径,也称为解码器或合成路径。轴向注意力机制可用于图像分类和密集预测,近来也被用于改进三元组两跳关系的推理,使用轴向注意模块^[15,21]作为特征提取器,能够关注两跳逻辑路径内的元素,并捕获三元组之间的相互依赖关系。语义分割和目标检测任务中,上下文信息在理解问题中至关重要,十字交叉注意力机制^[16,22]能够以一种非常高效的方式获得全图像的上下文信息。具体地说,对于每个像素,一种新颖的交叉注意模块在其交叉路径上获取所有像素的上下文信息,通过进一步的循环操作,最终可以捕获整个图像每个像素间的依赖性。

2.3 卷积神经网络和注意力机制

自 2012 年 AlexNet^[23]问世以来,卷积神经网络席卷了计算机视觉领域,后续改进的卷积机制包括膨胀卷积^[24]、形变卷积^[25,26]等来增大感受野和适应任务形变需求,广泛应用在图像识别、目标定位与检测、语义分割等方面。在篇章级关系抽取方面利用卷积神经网络,优势体现在既能充分利用字符级和词级的语义特征,又可以用到语序、上下文及文档整体的语义特征。近年来,随着神经网络的发展,注意力机制得以充分应用在实体关系抽取任务上^[27]。在面向表格类型数据方面,将注意力机制转移到图像最重要区域和忽略无关部分的方法称为视觉注意力机制^[28],视觉系统使用这种机制来帮助高效有效地分析和理解复杂场景^[29]。面对跨句实体间关系联合推理问题,注意力机制可以有效捕捉关键上下文语义信息,排除不相关上下文的噪声干扰,以获取更多有价值的判断关系类别所需要的细节特征。

2.4 多任务学习

多任务学习^[30](multi-task learning)是一种联合多

个任务同时学习来增强模型表示和泛化能力的一种手段,目前大都通过参数共享来实现多任务学习. 已有的工作提出了很多参数共享策略,其中使用的较多的有硬共享、软共享和分层共. 硬共享^[31-33]是目前应用最为广泛的共享机制,它把多个任务的数据表示嵌入到同一个语义空间中,再为每个任务使用一任务特定层提取任务特定表示. 硬共享实现起来比较简单,适合处理有较强相关性的任务,但遇到弱相关任务时常常表现较差. 软共享^[34-36]为每个任务都学习一个网络,但每个任务的网络都可以访问其他任务对应网络中的信息,例如表示、梯度等. 软共享机制非常灵活,不需要对任务相关性做任何假设,但是由于为每个任

务分配一个网络,常常需要增加很多参数. 分层共享^[37]是在网络的低层做较简单的任务,在高层做较困难的. 分层共享比硬共享要更灵活,同时所需的参数又比软共享少,但为多个任务设计高效的分层结构则依赖专家经验.

3 DRE-3DC 篇章级关系抽取模型

DRE-3DC 模型主要包括“实体/实体/关系”三维语义聚合模块、基于形变卷积的三重注意力机制的语义交互模块以及增强全局关系感知的多任务学习关系分类模块,这三个模块在任务中起到了编码器,语义交互和解码器的作用,如图4所示.

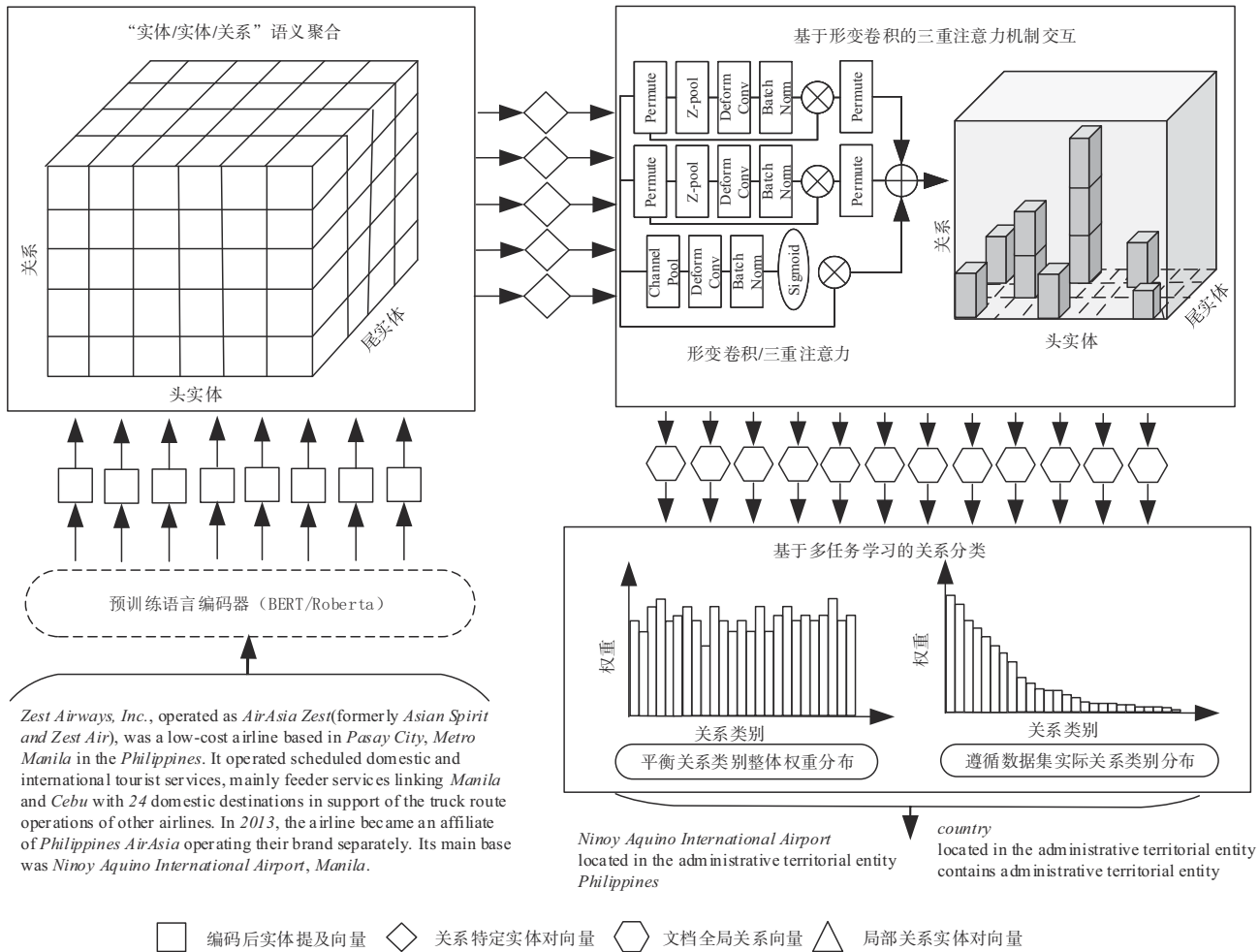


图4 DRE-3DC 模型概览

具体来说,模型接收文本输入,通过预训练语言模型编码模块提取原始语义特征,然后通过基于三维“实体/实体/关系”的表格形式进行头-尾实体、头实体-关系、尾实体-关系三种语义空间建模,之后通过基于形变卷积和三重注意力机制的语义交互模块增强实体对间对语义特征交互,然后通过与原始三维特征进行加和

操作得到用于多任务学习关系分类的特征输出,输入到关系分类模块中. 其中使用求和来连接交互模块和解码特征具有两个优势:首先,求和不会增大特征图的大小和通道的数量,能够减少下一层中可训练参数的数量;其次,具有自注意力机制求和结构的跳接可被视为远程残差连接,能够提高模型的分类精度. 在 DRE-

3DC 中,分类解码器得到的高层特征后通过多任务学习方式输出最终关系分类结果.

3.1 语义聚合模块

对于长度为 l 的文档 $D=[x_t]_{t=1}^l$, 其中 x_t 是在位置 t 处的字符, 使用特殊标记来表示实体提及, 即在实体提及的开始和结束位置用特殊标记“*”标记. 通过预训练语言模型来获得该处的上下文语义向量 H . 如果文档长度超过预训练语言模型的最大长度, 文档将被编码为多个重叠块并将重叠块的上下文化嵌入计算平均值.

$$H = \text{PrLM}(x_1, \dots, x_t) = [h_1, \dots, h_t] \in \mathbf{R}^{l \times d} \quad (1)$$

其中, d 为预训练语言模型的词向量维度大小. 本文使用 bert-base-cased 和 roberta-large 作为预训练模型编码句子语义向量. 先将输入文本的原始句子序列按字节对进行编码和分段标记. 文本经过编码分段后得到 token_ids 和 segment_ids, 其中 token_ids 和 segment_ids 分别代表对应语言模型词表中的字符序号和分段序号. 此后, 通过查询模型的词向量表得到对应位置的语义向量作为模块的输入.

进一步将实体提及开始处特殊标记“*”的嵌入向量作为实体提及向量 h_m . 对于有 m 个实体提及的实体 e 而言, 采用 logsumexp 方式聚合该实体对应的多个实体提及得到实体向量 h_e :

$$h_e = \log \sum_{j=1}^{N_{e_i}} \exp(h_{m_j}) \in \mathbf{R}^d \quad (2)$$

已有研究表明^[38], 实体周围出现的上下文信息对确定目标实体对的关系类别至关重要. 因而, 使用上下文池化得到目标实体的上下文注意力权重,

$$A_{e_i} = \sum_{j=1}^{N_{e_i}} (a_{m_j}) \quad (3)$$

其中, $a_{m_j} \in \mathbf{R}^{N \times l}$, 然后使用哈达马积得到目标实体对的注意力权重分布, 并以此作为查询向量, 再到原始上下文向量表中查询得到目标实体的上下文向量 $c^{(s,o)} \in \mathbf{R}^d$.

$$q^{(s,o)} = \sum_{i=1}^N (A_{e_i}^i \circ A_{e_o}^i) \quad (4)$$

$$c^{(s,o)} = H^T q^{(s,o)} \quad (5)$$

在此基础上, 通过加和操作融合实体表征和上下文表征得到最终的实体向量表示, 其中 W_s 和 W_c 分别是实体表征和上下文表征对应的变换矩阵.

$$z_s = \tanh(W_s h_{e_s} + W_c c^{(s,o)}) \quad (6)$$

接着采用分组双线性映射得到原始的实体交互表征, 即先把原实体表征向量表示平均分为 k 个组, 然后通过相应的变换矩阵 G 得到原始交互的实体对表征 $g^{(s,o)}$. 使用分组双线性映射的优势在于可以减少参数量, 降低模型过拟合的风险并加快模型的训练速度.

$$g_i^{(s,o)} = \sum_{j=1}^k (z_s^i \cdot W_{g_i}^j z_o^j) + b_i \quad (7)$$

$$g^{(s,o)} = [g_1^{(s,o)}, g_2^{(s,o)}, \dots, g_d^{(s,o)}] \quad (8)$$

其中, $G = W_{g_i}^j \in \mathbf{R}^{\frac{d}{k} \times \frac{d}{k}}$, $i \in (1, d)$, $j \in (1, k)$. 对于每种关系 $r \in R$ 通过一个变换矩阵 W_r 转换为新的关系特定实体表示:

$$T_i = \phi(GW_{r_i} + b_{r_i}) \quad (9)$$

其中, R 为预定义的关系类型集合, W_r 和 b_r 为可训练参数, 其中 $W_{r=1}^{|R|} \in \mathbf{R}^{d \times d}$, $b_{r=1}^{|R|} \in \mathbf{R}^{1 \times d}$, ϕ 为激活函数.

3.2 语义交互模块

在得到“实体/实体/关系”三维表格之后, 为了捕获关系具有的方向性特征, 使用三重注意力机制^[39]分别捕获目标实体在头实体和尾实体、头实体和关系、尾实体和关系三种语义空间下的特征. 首先通过 Z 池化进行维度缩减和特征提取:

$$Z\text{-pool}(x) = [\text{Maxpool}(x), \text{Avgpool}(x)] \quad (10)$$

其中 x 分别是“实体/实体/关系”三维表格在头实体和尾实体、头实体和关系、尾实体和关系三种语义空间下的切面. 在得到不同空间下的实体表征之后, 通过平均操作得到含有不同语义空间特征的实体表征:

$$y = \frac{1}{3} (\overline{x_1 w_1} + \overline{x_2 w_2} + \overline{x_3 w_3}) = \frac{1}{3} (y_1 + y_2 + y_3) \quad (11)$$

由于表格中的实体排序是按实体在文档中出现的先后次序排列, 存在一定的不规则性和隐含了一定的文档结构特征, 而普通卷积不能很好的自适应的捕获这一文档结构特征. 因而在此基础上, 为了自适应地聚合实体交互信息, 在捕获文档结构特征的同时避免无关实体的影响, 使用形变卷积^[26]替换普通卷积. 形变卷积主要由两部分构成, 分别是形变卷积和形变池化:

$$y(p) = \sum_{k=1}^k w_k \cdot x(p + p_k + \Delta p_k) \cdot \Delta m_k \quad (12)$$

$$y_k = \sum_{j=1}^{n_k} x(p_{k_j} + \Delta p_k) \cdot \Delta \frac{m_k}{n_k} \quad (13)$$

其中, Δp_k 和 Δm_k 分别是位置 k 处可学习的偏置和缩放量.

3.3 关系分类模块

在使用语义交互模块进行语义增强后, 为了确定一个实体对 (e_h, e_t) 的语义关系类型, 首先通过一个全连接神经网络将其映射到新的表示空间, 然后使用双线性映射得到最终该实体对所属的关系类别概率:

$$z_h = \tanh(W_h e_h + T_{h,t}) \quad (14)$$

$$z_t = \tanh(W_t e_t + T_{h,t}) \quad (15)$$

$$p(e_h, e_t) = \sigma(z_h W_r z_t + b_r) \quad (16)$$

其中, $W \in \mathbf{R}^{d \times d}$, $b_r \in \mathbf{R}^{1 \times d}$.

为了更有效地解决多标签和样本不平衡问题,采用自适应阈值损失作为训练目标. 具体来说,它引入一个额外的阈值关系类别 TH,并通过增加高于阈值关系的正关系对数 P_T 和减少低于阈值关系的负关系对数 N_T 来优化损失. 其中 logit 是 Sigmoid 函数之前最后一层的输出.

$$L_{\text{each}} = - \sum_{r \in P_r} \log \left(\frac{\exp(\text{logit}_r)}{\sum_{r' \in P_r \cup \{\text{TH}\}} \exp(\text{logit}_{r'})} \right) - \log \left(\frac{\exp(\text{logit}_{\text{TH}})}{\sum_{r' \in N_T \cup \{\text{TH}\}} \exp(\text{logit}_{r'})} \right) \quad (17)$$

为了缓解长尾分布问题,采用多任务学习方法,把篇章整体关系类别识别和单个实体对所属关系类别识别两个任务的数据表示嵌入到同一个语义空间中,再为每个任务使用任务特定层提取任务特定表示来增强模型对篇章全局关系表达的感知. 因此,在“实体/实体/关系”三维表格语义增强之后,采用全局平均池化得到文档整体的语义表示,作为全局关系分类模块对输入并采用交叉熵损失函数确定整体所包含的关系类别:

$$d_{\text{all}} = \text{GlobalAveragePooling}(T) \quad (18)$$

$$P(d_{\text{all}}) = \text{sigmoid}(d_{\text{all}} W_{\text{all}} + b_{\text{all}}) \quad (19)$$

其中, $W_{\text{all}} \in \mathbf{R}^{d_r \times d_r}$, $b_{\text{all}} \in \mathbf{R}^{1 \times d_r}$ 的相应损失函数为:

$$L_{\text{global}} = \text{CrossEntropy}(P_{d_{\text{all}}}, R_{d_{\text{all}}}) \quad (20)$$

选择在验证集上表现最佳的损失函数超参作为最终多任务学习中各损失函数的比例.

DRE-3DC 模型在训练阶段整体目标函数为:

$$L_{\text{train}} = L_{\text{each}} + L_{\text{global}} \quad (21)$$

4 实验与评估

4.1 实验环境

本文采用 DocRED^[18] 和 Revisit-DocRED^[19] 两个数据集进行实验评估. 其中 DocRED 是第一个大规模用于篇章级关系抽取的数据集,基于 Wikipedia 和 Wiki-data 构建. DocRED 包含 3 053 个实例用于训练、1 000 个实例用于验证和 1 000 个实例用于测试,共包含 97 种关系类型,每个文档平均包含 26 个实体.

Revisit-DocRED 是针对 DocRED 数据集构建时采用远程监督标注方式所导致漏标及长尾分布问题,通过专家手工标注 96 份验证集作为测试集来准确评估模型真实性能,包括 Revise 数据集和 Scratch 数据集. 其中 Revise 数据集用于评估模型在原始远程监督标注初始修正后的模型表现,数据集中含有大量的漏标实体对和关系类别长尾分布现象,而 Scratch 数据集是通过专家手工逐个实体对核对并补充原 Revise 数据集里漏标的实体对得到的,其数据标注的分布情况更符合真实

状态下的长尾分布现象,因此,模型在其上的精确率、召回率和 F1 值更加符合实际应用时模型的表现. 实验所用数据集的相关统计信息如表 1 所示.

表 1 实验所用数据集的相关统计信息

统计信息	DocRED	Revisit-DocRED
训练集	3 053	3 053
验证集	998	902
测试集	1 000	96
关系种类	97	97
平均实体数量	19.5	19.7
平均元组数量	12.6	34.5

实验采用小批量梯度下降方式训练模型,批量大小为 2;使用 AdamW^[40] 优化器,学习率设置为 10^{-5} ;在验证集上确定超参数,通过训练 30 轮,选取在验证集上效果表现最好的模型. 使用的预训练语言模型为 [BERT-base, cased] 和 [Roberta-large],分别包含 1.1 亿和 3.35 亿个参数,句子最大长度设置为 512. 实验硬件环境为 CPU 为 8 核 Intel Core i9-9900k,内存 128 GB, GPU 为 GeForce RTX 3090.

4.2 实验结果

由于本文着重解决表格模型抽取方法中的特征聚合混淆和局部化的问题,因此将重点对比所提出模型与其它基于表格结构模型的性能差异. 基于此,在 DocRED 和 Revisit-DocRED 上分别进行了实验来验证模型的有效性. 实验结果如表 2 和表 3 所示. 其中, B-b 和 Rb-l 分别代表使用 BERT-base 语言模型和 RoBERTa-large 语言模型的词向量进行参数初始化的模型.

在 DocRED 数据集的实验中,主要是通过对比所提出模型与其它模型的整体性能,来进一步验证和评估 DRE-3DC 模型的有效性. 从表 2 中可以观察到,相较于现有基于表格的抽取模型,本文所提出的模型在不使用远程监督数据集时,无论是基于 BERT-base 编码器还是 Roberta-large 编码器,模型都取得最高 IgnF1 及 F1 值.

在 Revisit-DocRED 上的实验是通过对比所提出模型与其它模型的性能来验证和评估模型对长尾分布问题的有效性. 实验结果如表 3 所示. 从表中可以观察到,相较于现有基于表格的抽取模型,本文所提出的模型在较少损失精确率的前提下,大幅提高了召回率,使得模型在 Revise 和 Scratch 两个数据集上都取得了当前最高的召回率和 F1 值,证明了模型对缓解长尾分布问题的有效性. 此外,还可以观察到本文所提出模型相较于 DocuNet 在 Revisit-DocRED 中的 Scratch 数据集上的 F1 提升效果大于在 DocRED 上的性能提升,表明本文所提出模型在 DocRED 数据集上的 F1 值性能提升主要来自数据集中对长尾关系类别部分,进一步说明了本文模型对于缓解长尾分布问题的有效性.

表 2 在 DocRED 数据集上的实验结果

数据集类型	模型	验证集		测试集	
		Ign $F1$	$F1$	Ign $F1$	$F1$
不使用远程 监督数据集	NC-DRE-B-b ^[9]	60.84	62.75	60.59	62.73
	DocuNet-B-b ^[14]	59.86	61.83	59.93	61.86
	KD-DocRE-B-b ^[15]	60.08	62.03	60.04	62.08
	Dense-CCNet-B-b ^[16]	60.72	62.74	60.46	62.55
	MRN-B-b ^[41]	59.74	61.61	59.52	61.74
	Ours-B-b	61.19	63.11	61.23	63.14
	DocuNet-Rb-l ^[14]	62.23	64.12	62.39	64.55
	KD-DocRE-Rb-l ^[15]	62.16	64.19	62.57	64.28
	Ours-Rb-l	62.51	64.59	62.88	65.17
使用远程 监督数据集	DocuNet-NA-Rb-l ^[14]	63.26	65.21	63.29	65.44
	ATLOP-NA-Rb-l ^[7]	63.41	65.33	63.54	65.47
	KD-DocRE-NA-Rb-l ^[15]	63.38	65.64	63.63	65.71
	SSAN-NA-Rb-l ^[8]	63.76	65.69	63.78	65.92
	Ours-NA-Rb-l	64.24	66.34	63.93	66.19

表 3 在 Revisit-DocRED 数据集上的实验结果

模型	Revise			Scratch		
	P	R	$F1$	P	R	$F1$
BiLSTM ^[18]	50.2	46.7	48.4	66.6	22.8	33.9
GAIN ^[12]	60.0	56.8	58.3	81.1	28.1	41.8
ATLOP ^[7]	66.3	59.1	62.5	90.3	29.5	44.5
SSAN ^[8]	63.1	61.3	62.2	84.5	30.1	44.5
DocuNet ^[14]	66.9	59.9	63.2	89.1	29.3	44.1
KD-DocRE ^[15]	65.4	62.9	64.1	88.4	29.4	44.2
Ours	63.2	67.4	65.2	85.5	31.6	46.2

4.3 实验分析

4.3.1 消融研究

为了分析模型各部分所起的作用,本文分别在 DocRED 和 Revisit-DocRED 上进行了消融实验. 实验结果如表 4 和表 5 所示. 其中, Ours 代表本文所提出的完整模型, -Triplet Attn 代表去掉三重注意力后的模型, -Deform Conv 表示用常规卷积替换形变卷积进行特征融合, -Global Relation 表示去掉用于全局关系感知的多任务学习模块. 从实验结果中可以看到完整模型在 2 个数据集上都取得了最高的 $F1$ 和 Ign $F1$ 值.

本文所提出模型由于添加了三重注意力,在 2 个数据集上的 $F1$ 值分别提升 0.46% 和 0.9%; 因为采用了形变卷积, $F1$ 值分别提升了 0.76% 和 0.5%; 由于设计了多任务学习模块, $F1$ 值分别提升了 0.51% 和 0.6%. 这表明三重注意力的引入可以进一步提高模型的解耦语义聚合能力,形变卷积的引入可以更有效地实现自适应结构特征融合,多任务学习模块进一步提升了模型对长尾类别实体对的召回. 综上,模型各模块的有效性得到了充分证明,说明本文所提出的方法可以有效地减

少不同语义空间下噪声对目标实体对特征融合的影响,更准确地分类各实体对所属关系类别.

表 4 在 DocRED 数据集上的消融实验结果

模型变体	验证集			
	P	R	Ign $F1$	$F1$
Ours	66.62	62.57	62.51	64.59
-Triplet Attn	64.52	63.74	61.24	64.13
-Deform Conv	64.21	63.46	61.05	63.83
-Global Relation	65.24	62.31	62.02	64.08

表 5 在 Revisit-DocRED 数据集上的消融实验结果

模型变体	Scratch 数据集		
	P	R	$F1$
Ours	85.5	31.6	46.2
-Triplet Attn	84.3	30.2	45.3
-Deform Conv	83.7	31.4	45.7
-Global Relation	87.6	30.8	45.6

4.3.2 超参鲁棒性研究

为了探究卷积核大小对聚合交互特征的影响,在使用同样的超参数以及训练相同轮数的前提下,调整卷积核大小 K 对该影响因素进行了分析.

从图 5 中可以看出,随着卷积核大小 K 的变化, $F1$ 值呈现先升后降的变化趋势. 表明按照实体在文档中出现顺序排列的表格结构具有一定的文档结构特征,一开始随着卷积核感受野的扩大,模型可以有效的聚合相关实体交互特征,但感受野过大也会导致噪声对模型性能产生负面影响.

4.3.3 实例分析

为了进一步观察和分析本文所提出模型的实际效果,选取具有代表性的篇章级关系抽取模型 DocuNet 和

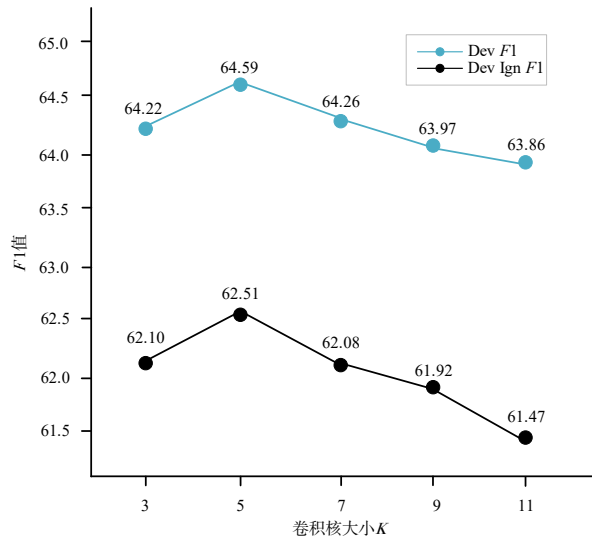


图 5 卷积核大小 K 对性能的影响

KD-DocRE 作为比较对象,且二者同样是利用表格结构进行语义分割的关系抽取模型.实例对比实验挑选了 3 个典型样例,如表 6 所示.从表 7 中可以观察到本文模型与基线模型在抽取三元组方面的结果差异.

通过实验可以看到,针对样例 1,DocuNet 模型由于使用多层卷积对特征进行聚合,在样例中出现了特征混淆,导致主语 Dancer 的特征融合误差引起三元组关系类别错误,同时也无法抽取到所有三元组.而本文模型由于融合了关系表达向量,可以正确地抽取出所有头实体,并基于此抽取出所有正确的三元组.

在样例 2 中,由于单词 University 涉及到多个三元组,且 city 与 state 这 2 个关系较难区分,KD-DocRE 模型由于限制了特征聚合的方式,仅考虑了头尾实体的少量交互信息,因此只能抽取到部分三元组.而本文模型引入的三维表征建模方式进一步地解耦了不同语义空

间信息,使得模型可以准确地分类所有实体对.

从样例 3 的结果中可以发现本文所提出的模型虽然弥补了部分 KD-DocRE 模型中头尾实体特征聚合受限问题,但会发生错误识别二者所属关系类别的情况,因此得到了 1 个错误的三元组.经过分析,导致该错误一方面是由于句子表达信息的缺失,另一方面是没有有效构建实体对方向约束.未来的研究可以考虑构建实体对方向约束,进一步提升模型的鲁棒性和泛化能力.

表 6 实例分析例句

例句	文本内容	包含的三元组
例句 1	English is the language in Great Britain and United States. A Loyal Character Dancer was published by Soho Press in the United States.	(Language,spokenIn,Britain), (Dancer,publisher, Press),(States,language,language),(Press,country, States)
例句 2	Cornell University in Ithaca, New York is the publisher of Administrative Science Quarterly. The University is affiliated with the Association of American Universities.	(University,city,York), (University,state,York), (University,affiliation,Universities), (Quarterly,publisher,University)
例句 3	Elliot See was born in Dallas, which is a country in Texas. He attended the University of Texas at Austin, which is affiliated to the University of Texas system. The University of Texas at Austin will be part of the Big 12 Conference competition.	(See, birthplace, Dallas), (See,almaMater,Austin), (Austin,compete in,Conference), (Dallas,partsType, Texas)

表 7 实例分析结果

例句	方法	输出
例句 1	DocuNet	(Language, spokenIn, Britain), (Dancer, publisher, Press), (Dancer, country, States),
	本文方法	(Language, spokenIn, Britain), (Dancer, publisher, Press), (States, language, language), (Press, country,States)
例句 2	KD-DocRE	(University, city, York), (University, affiliation, Universities), (Quarterly, publisher, University)
	本文方法	(University, city, York), (University, state, York), (University, affiliation, Universities), (Quarterly, publisher, University)
例句 3	KD-DocRE	(See, birthplace, Dallas), (See, almaMater, Austin), (Austin, compete in, Conference)
	本文方法	(See, birthplace, Dallas), (See, almaMater, Austin), (Austin, compete in, Conference), (Dallas, part, Texas)

5 结论

本文提出了一种新的篇章级关系抽取模型 DRE-3DC,对实体对二维表格建模进行扩展,设计了“实体/实体/关系”三维表征建模方式,采用三重注意力机制区分实体在关系头尾处的语义,使得相同关系在头尾实体区别下的语义得到充分表示,同时采用形变卷积网络应对表格建模方法的几何形变问题,有效增强了实体关系所涉及的文档结构信息聚合能力,并且采用多任务学习增强模型对文档整体关系类别组合的感知能力,缓解篇章级关系抽取任务的关系类别长尾分布问题.实验结果表明,DRE-3DC模型性能达到了良好的效果,并通过消融实验、对比分析和实例分析,验证了本文所提方法的有效性.

参考文献

- [1] 冯钧,魏大保,苏栋,等.文档级实体关系抽取方法研究综述[J].计算机学报,2022,49(10):224-242.
FENG J, WEI D B, SU D, et al. Survey of document-level entity relation extraction methods[J]. Computer Science, 2022, 49(10): 224-242. (in Chinese)
- [2] 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 J Z, SONG S S, WANG Y Z, et al. Entity relation extraction based on improved attention mechanism[J]. Acta Electronica Sinica, 2019, 47(8): 1692-1700. (in Chinese)
- [3] SAHU S, ANAND A, ORUGANTY K, et al. Relation extraction from clinical texts using domain invariant convolutional neural network[C]//Proceedings of the 15th Workshop on Biomedical Natural Language Processing. Stroudsburg: Association for Computational Linguistics, 2016: 206-215.
- [4] JIA R, WONG C, POON H. Document-level N-ary relation extraction with multiscale representation learning[C]//Proceedings of the 2019 Conference of the North. Stroudsburg: Association for Computational Linguistics, 2019: 3693-3704.
- [5] XU Y, YANG Z H, SONG Y W, et al. Star-BiLSTM-LAN for document-level mutation-disease relation extraction from biomedical literature[C]//2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). Piscataway: IEEE, 2020: 357-362.
- [6] WANG H, FOCKE C, SYLVESTER R, et al. Fine-tune bert for DocRED with two-step process[EB/OL]. (2019-09-26) [2022-08-20]. <https://arxiv.org/abs/1909.11898>.
- [7] ZHOU W X, HUANG K, MA T Y, et al. Document-level relation extraction with adaptive thresholding and localized context pooling[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2021, 35(16): 14612-14620.
- [8] XU B F, WANG Q, LYU Y J, et al. Entity structure within and throughout: Modeling mention dependencies for document-level relation extraction[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2021, 35(16): 14149-14157.
- [9] ZHANG L, CHENG Y D. NC-DRE: Leveraging non-entity clue information for document-level relation extraction [EB/OL]. (2022-04-01) [2022-08-10]. <http://arxiv.org/abs/2204.00255>.
- [10] CHRISTOPOULOU F, MIWA M, ANANIADOU S. Connecting the dots: Document-level neural relation extraction with edge-oriented graphs[C]//Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). Stroudsburg: Association for Computational Linguistics, 2019: 4924-4935.
- [11] 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)
- [12] ZENG S, XU R X, CHANG B B, et al. Double graph based reasoning for document-level relation extraction [C]//Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP). Stroudsburg: Association for Computational Linguistics, 2020: 1630-1640.
- [13] WANG H L, QIN K, LU G M, et al. Document-level relation extraction using evidence reasoning on RST-GRAPH [J]. Knowledge-Based Systems, 2021, 228: 107274.
- [14] ZHANG N Y, CHEN X, XIE X, et al. Document-level relation extraction as semantic segmentation[EB/OL]. (2021-08-22) [2022-08-20]. <https://arxiv.org/abs/2106.03618>.
- [15] TAN Q Y, HE R D, BING L D, et al. Document-level relation extraction with adaptive focal loss and knowledge distillation[C]//Findings of the Association for Computational Linguistics: ACL 2022. Stroudsburg: Association for Computational Linguistics, 2022: 1672-1681.
- [16] ZHANG L, CHENG Y D. A densely connected criss-cross attention network for document-level relation extraction[EB/OL]. (2022-03-26) [2022-08-10]. <http://arxiv.org/abs/2203.05111>.

org/abs/2203.13953.

- [17] DU Y K, MA T F, WU L F, et al. Improving long tailed document-level relation extraction via easy relation augmentation and contrastive learning[EB/OL]. (2022-05-21) [2022-08-10]. <https://arxiv.org/abs/2205.10511>.
- [18] YAO Y, YE D M, LI P, et al. DocRED: A large-scale document-level relation extraction dataset[C]//Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Stroudsburg: Association for Computational Linguistics, 2019: 764-777.
- [19] HUANG Q Z, HAO S B, YE Y, et al. Does recommend-revise produce reliable annotations? An analysis on missing instances in DocRED[C]//Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Stroudsburg: Association for Computational Linguistics, 2022: 6241-6252.
- [20] RONNEBERGER O, FISCHER P, BROX T. U-Net: Convolutional networks for biomedical image segmentation [M]//Lecture Notes in Computer Science. Cham: Springer International Publishing, 2015: 234-241.
- [21] WANG H Y, ZHU Y K, GREEN B, et al. Axial-deeplab: Stand-alone axial-attention for panoptic segmentation[C]//Computer Vision-ECCV 2020. New York: ACM, 2020: 108-126.
- [22] HUANG Z L, WANG X G, HUANG L C, et al. CCNet: criss-cross attention for semantic segmentation[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway: IEEE, 2019: 603-612.
- [23] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks[C]//Proceedings of the 26th Annual Conference on Neural Information Processing Systems. Cambridge: MIT Press, 2012: 1106-1114.
- [24] YU F, KOLTUN V. Multi-scale context aggregation by dilated convolutions[EB/OL]. (2015-11-23) [2022-08-10]. <http://arxiv.org/abs/1511.07122>.
- [25] DAI J F, QI H Z, XIONG Y W, et al. Deformable convolutional networks[C]//2017 IEEE International Conference on Computer Vision (ICCV). Piscataway: IEEE, 2017: 764-773.
- [26] ZHU X Z, HU H, LIN S, et al. Deformable ConvNets V2: More deformable, better results[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2019: 9308-9316.
- [27] GUO Z J, ZHANG Y, LU W. Attention guided graph convolutional networks for relation extraction[C]//Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Stroudsburg: Association for Computational Linguistics, 2019: 241-251.
- [28] GUO M H, XU T X, LIU J J, et al. Attention mechanisms in computer vision: A survey[J]. Computational Visual Media, 2022, 8(3): 331-368.
- [29] 吴绿, 张馨月, 唐茉, 等. Focus+Context 语义表征的场景图像分割[J]. 电子学报, 2021, 49(3): 596-604.
- WU L, ZHANG X Y, TANG M, et al. Focus+Context semantic representation in scene segmentation[J]. Acta Electronica Sinica, 2021, 49(3): 596-604. (in Chinese)
- [30] ZHANG Z H, YU W H, YU M X, et al. A survey of multi-task learning in natural language processing: regarding task relatedness and training methods[EB/OL]. (2022-04-07) [2022-08-10]. <https://arxiv.org/abs/2204.03508>.
- [31] COLLOBERT R, WESTON J. A unified architecture for natural language processing: Deep neural networks with multitask learning[C]//Proceedings of the 25th international conference on Machine learning. New York: ACM, 2008: 160-167.
- [32] SUBRAMANIAN S, TRISCHLER A, BENGIO Y, et al. Learning general purpose distributed sentence representations via large scale multi-task learning[EB/OL]. (2018-03-30) [2022-08-10]. <http://arxiv.org/abs/1804.00079>.
- [33] LIU X D, HE P C, CHEN W Z, et al. Multi-task deep neural networks for natural language understanding[C]//Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Stroudsburg: Association for Computational Linguistics, 2019: 4487-4496.
- [34] MISRA I, SHRIVASTAVA A, GUPTA A, et al. Cross-stitch networks for multi-task learning[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2016: 3994-4003.
- [35] RUDER S, BINGEL J, AUGENSTEIN I, et al. Latent multi-task architecture learning[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2019, 33(1): 4822-4829.
- [36] LIU P F, QIU X P, HUANG X J. Recurrent neural network for text classification with multi-task learning[C]//Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence. New York: ACM, 2016: 2873-2879.
- [37] HASHIMOTO K, XIONG C M, TSURUOKA Y, et al. A joint many-task model: Growing a neural network for multiple NLP tasks[C]//Proceedings of the 2017 Conference on Empirical Methods in Natural Language Process-

ing. Stroudsburg: Association for Computational Linguistics, 2017: 1923-1933.

- [38] PENG H, GAO T Y, HAN X, et al. Learning from context or names? An empirical study on neural relation extraction[C]//Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP). Stroudsburg: Association for Computational Linguistics, 2020: 3661-3672.
- [39] MISRA D, NALAMADA T, ARASANIPALAI A U, et al. Rotate to attend: Convolutional triplet attention module [C]//2021 IEEE Winter Conference on Applications of Computer Vision (WACV). Piscataway: IEEE, 2021: 3139-3148.
- [40] KINGMA D P, BA J. Adam: A method for stochastic optimization[EB/OL]. (2014-12-22) [2022-08-10]. <https://arxiv.org/abs/1412.6980>.
- [41] LI J Y, XU K, LI F, et al. MRN: A locally and globally mention-based reasoning network for document-level relation extraction[C]//Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021. Stroudsburg: Association for Computational Linguistics, 2021: 1359-1370.



李伟平 男,1973年3月,辽宁凌源人.北京大学教授、博士生导师,主要研究方向为大数据分析 and 信息抽取.

E-mail: wpli@ss.pku.edu.cn



赵文 男,1967年出生,辽宁大连人.现为北京大学软件工程国家工程研究中心研究员、博士生导师,主要研究领域为知识图谱、软件工程和软件安全.

E-mail: zhaowen@pku.edu.cn

作者简介



王宇 男,1978年出生,辽宁沈阳人.北京大学软件与微电子学院2020级博士研究生,主要研究方向为知识图谱构建和自然语言处理.

E-mail: wangyu_cn@stu.pku.edu.cn



王震 男,1994年出生,山东济南人.北京大学软件与微电子学院博士研究生,主要研究方向为知识图谱构建和自然语言处理.

E-mail: wangzhen@stu.pku.edu.cn



温立强 男,1991年出生,山西吕梁人.北京大学软件与微电子学院博士研究生,主要研究方向为知识图谱问答和智能检索.

E-mail: wenlq@pku.edu.cn