

基于自注意力机制模拟实体信息的实体关系抽取^①



何松泽, 王 婷, 梁佳莹, 陈永雄, 戴青江

(成都信息工程大学 计算机学院, 成都 610225)

通信作者: 王 婷, E-mail: wangting@cuit.edu.cn

摘 要: 在信息抽取领域, 从非结构化文本中抽取实体关系是一项基础且重要的任务, 且面临实体重叠和模型误差累积等挑战. 本文以关系为导向, 提出一种改进的实体关系联合抽取方法. 该方法将实体关系抽取任务分为关系抽取与实体抽取两个子任务. 在关系抽取任务上采用自注意力机制关注词与词之间的重要程度从而模拟实体信息, 并使用平均池化来表征整个句子信息; 在实体抽取任务上结合关系信息使用条件随机场识别该关系下的实体对. 本模型不仅能够利用存在关系必定存在实体对的思想解决实体对重叠问题, 还能够在训练过程中利用数据集中已知的关系使实体抽取模块不依赖于关系抽取模块的结果来训练, 从而在训练阶段避免误差累积. 最后, 在 WebNLG 和 NYT 公开数据集上验证了该模型的有效性.

关键词: 信息抽取; 深度学习; 注意力机制; 自然语言处理; 人工智能

引用格式: 何松泽, 王婷, 梁佳莹, 陈永雄, 戴青江. 基于自注意力机制模拟实体信息的实体关系抽取. 计算机系统应用, 2023, 32(2): 364-370. <http://www.c-s-a.org.cn/1003-3254/8963.html>

Entity Relation Extraction Simulation of Entity Information Based on Self-attention Mechanism

HE Song-Ze, WANG Ting, LIANG Jia-Ying, CHEN Yong-Xiong, DAI Qing-Jiang

(School of Computer Science, Chengdu University of Information Technology, Chengdu 610225, China)

Abstract: In the field of information extraction, it is a basic and important task to extract entity relations from unstructured texts, and challenges such as entity overlap and model error accumulation often appear. This study is relation-oriented, and it proposes an improved joint extraction method for entity relations. The method divides the entity relation extraction task into two subtasks: relation extraction and entity extraction. For the relation extraction subtask, a self-attention mechanism is adopted to evaluate the degree of association between words, so as to simulate entity information and represent the whole sentence information by the average pooling. For the entity extraction subtask, according to relation information, the conditional random field is used to identify the entity pairs under the relation. This method can not only solve the problem of entity overlap by using the idea that relation and entity pairs coexist but also perform training by using the known relation in the dataset to make the entity extraction module independent from the results of the relation extraction module during the training, so as to avoid error accumulation. Finally, the effectiveness of the model is verified on the public datasets of WebNLG and NYT.

Key words: information extraction; deep learning; attention mechanism; natural language processing (NLP); artificial intelligence

① 基金项目: 四川省科技厅重点研发项目 (2021YFG0031, 2022YFG0375); 四川省科技服务业示范项目 (2021GFW130); 2022 年度大学生创业训练计划 (202210621196, 202210621073k)

收稿时间: 2022-06-30; 修改时间: 2022-07-29; 采用时间: 2022-09-07; csa 在线出版时间: 2022-11-14

CNKI 网络首发时间: 2022-11-15

实体关系抽取是指从非结构化文本中抽取形式如头实体、关系和尾实体的三元组的信息。它是知识图谱及一些下游工作的基础。目前该任务主要存在着3种类型的实体重叠问题,分别是:单个实体重叠(single entity overlap, SEO)、实体对重叠(entity pair overlap, EPO)以及无实体重叠(normal)。单个实体重叠是指一个实体与其他多个实体存在关系;实体对重叠是指一个实体对存在多种关系;无实体重叠是指一个实体与另一个实体之间有且仅有一个关系,并且另一个实体也是唯一的。目前来说,三元组抽取方法大体分为管道(pipeline)和联合(joint)两种方式。管道方法主要采用先抽取实体,再抽取关系的方式来进行实体关系抽取。早期,该类方法主要采用卷积神经网络(convolutional neural network, CNN)^[1]进行实体关系抽取,其将无监督的词向量以及位置向量作为输入特征来进行实体关系抽取并取得了一定效果,但存在卷积核大小固定,抽取到的特征单一的缺点。Shen等人^[2]将注意力机制方法引入到CNN模型中进行关系抽取,充分利用了位置等信息从而提升了模型效果。在这之后,基于循环神经网络的方法也逐渐被应用到这个任务上,并能够很好地解决文本长依赖问题。Qin等人^[3]使用了引入注意力机制的双向循环门控单元,能够捕获更有价值的字符级信息。Xu等人^[4]提出了基于最短依存路径的长短时记忆网络对文本进行关系分类,能够收集该最短路径上的异构信息。采用管道方式纵然能够取得一定的实体关系抽取效果,但结构上不可避免地存在误差累积的问题。采用联合抽取的方式将实体识别与关系抽取任务进行联合建模,同时完成实体和关系的抽取,可以在一定程度上避免这种误差累积问题。在早期尝试中,Miwa等人^[5]提出了一种树形结构来提取实体及实体关系。其主要利用了词序信息以及依存树结构信息,将长短时记忆模型转换为树形结构来进行建模。但该方式并不是完全同步,实体与关系的抽取还是存在先后性。而Zheng等人^[6]将该问题转换为一个序列标注问题,利用长短时记忆网络就可以建模,大大简化了问题的复杂性。但是该模型缺点也十分明显,以该方式进行实体关系抽取难以解决实体重叠问题。Bekoulis等人^[7]在此基础上引入了多头注意力机制对该问题进行建模,从而部分解决实体关系重叠问题。Wei等人^[8]提出了CasRel方法,该方法通过共享参数将三元组抽取任务视为一个多任务学习。其将该任务建模成3个级别问题,从而更有效的解决实体重叠问题,但存在曝光偏差问题,即训练阶段同时训练,而在预测时则按照一定的

顺序进行依次抽取。TPLinker^[9]则是基于握手标注策略的端到端序列标注模型,巧妙地将联合抽取转化为token pair的链接问题,避免了曝光偏差问题。

上述所采取模型方法中,主要存在着模型结构复杂、需特殊的标注方法、需精心设计抽取次序等抽取设计问题,以及实体重叠和误差累积问题。在本文中,我们对这些问题进行实验探究,提出以关系为导向的一种无复杂设计结构、无特殊标注方法且简单有效的实体关系联合抽取方法,该方法针对上述问题主要贡献如下。

(1) 所提出的方法在解决实体对重叠问题上效果较好,能够在关系数较多样本量较少的情况下,保持较好的表现。

(2) 本文提出在没有获得实体前,通过采用自注意力机制关注字符(词)信息生成模拟实体信息,进而抽取实体关系的方法,该模型有效的提升三元组抽取效果。

(3) 在保证三元组抽取效果的前提下,本文极大地简化了抽取模型,无复杂设计部分与较为特殊的训练方式,易于理解。

1 相关工作

对于两种抽取方法(管道式与联合)又可以分别细分为以关系为导向(先关系后实体)和以实体信息为导向(先实体后关系)。这里主要介绍采用以关系为导向的处理流程来进行实体关系联合抽取,例如:CopyRE^[10]与HRL^[11]均是在预测阶段先抽取关系再抽取实体方法。不同的是,CopyRE是以一种复制机制来抽取三元组信息方法,但其只能复制实体的最后一个单词。而HRL则应用了一种分层强化学习的方式来抽取三元组的信息。RSAN^[12]则提出了一种基于关系注意力机制的方法来进行实体关系三元组的抽取。该方法利用注意力机制为每个关系构建特定的句子表征,并据此抽取特定关系下的实体信息。然而,上述以关系为导向的实体关系联合抽取法所采取的方式都较为复杂,忽略了一些简单却有效的方法和理论研究。例如,Yu等人^[13]使用双仿射注意力机制能够让实体首尾边界进行交互,从而达到提升实体抽取效果;Wu等人^[14]通过提取实体的字符向量与句子语义信息从而获得较为全面的信息来提升关系抽取效果;Peng等人^[15]对实体信息与上下文信息对关系抽取影响进行了深入研究,发现实体信息对于关系抽取的效果有重要的影响。

本文对上述方法进行简化和改进,将有效的方法与研究保留(如:RSAN模型中利用平均池化方法提取

整个句子信息,引入实体信息思想等),构建了有效且简单的实体关系三元组的联合抽取模型,该模型不仅保留了上述方法的优点,还有效地提取存在实体重叠问题的文本中的三元组.

2 模型概述

依据所划分的两个任务,本文所提出的方法分为两个模块:关系抽取模块与实体抽取模块,如图1所示.

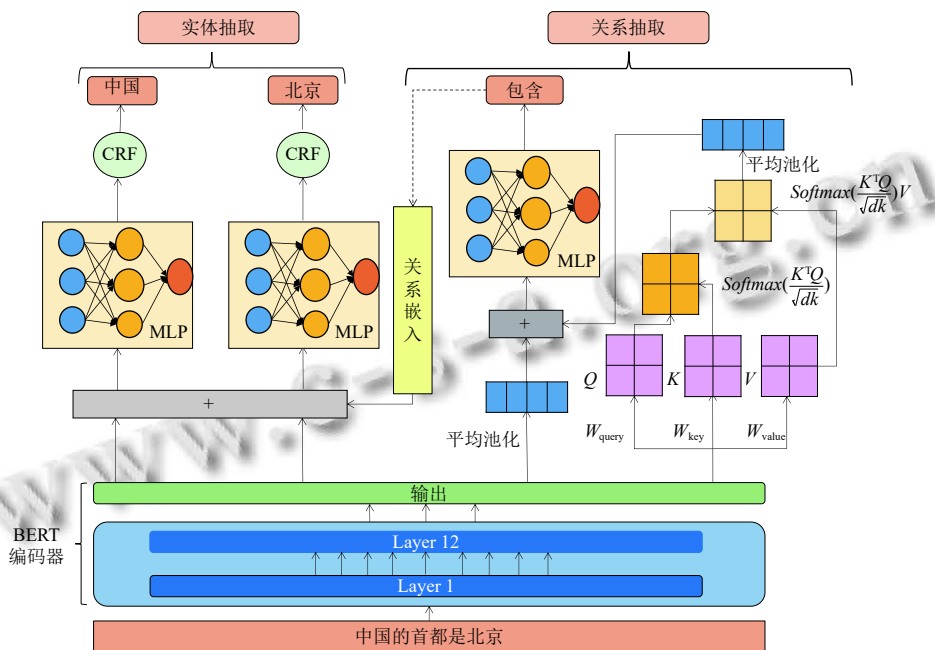


图1 模型整体结构

从模型整体上来说,如图1所示,对于一个给定的句子,首先将该句子输入进编码器获取到由每个词所对应的词向量,这里使用BERT预训练模型作为共享编码器得到编码输出 $Y_{enc}(s) = \{h_1, h_2, \dots, h_n | h_i \in R^{d \times 1}\}$,其中 d 为向量维度, n 为句子的单词个数.例如,输入“中国的首都是北京”,就会将其转换为行数为8列数为768的词向量矩阵.接着将由这些词向量所组成的矩阵输入到关系抽取模块中,在关系抽取模块中,我们认为每个词都可能蕴含实体信息,由此如果需要知道在某个关系下的实体信息,就是要知道词与词之间的重要性,因为一个关系存在则必定能对应一组实体对,那么在该关系下这组实体对之间关联程度,应当大于与其他实体的关联程度.为了获得这种关联度,也可称作为模拟实体信息,这里我们采用自注意力机制与平均池化的方法来模拟实体信息.具体来说,将词向量所组成的矩阵分别乘以3个不同的可训练的参数矩阵,由此获得如图所示的查询矩阵(Q)、关键字矩阵(K)、值矩阵(V)3个矩阵,再利用 Q 矩阵与 K 矩阵计算出注意力权重矩阵.该矩阵可以表征在某个关系下词与词

之间的关联程度,获得注意力权重矩阵后与 V 矩阵相乘得到包含这种关联程度的向量矩阵,也可称作包含模拟实体信息的向量矩阵.接着对包含模拟实体信息的向量矩阵进行平均池化来获得代表整体的模拟实体信息向量.最后将由编码器输出的词向量矩阵平均池化后所获得的能够代表整体句子信息向量与模拟实体信息向量拼接,拼接后的混合向量送入到神经网络中(即图1中的MLP)来识别出句子中“包含”这一关系.在获得“包含”这一关系后,通过向量随机初始化的方式(即关系嵌入)就能找到该词所对应的关系向量.再将得到的关系向量与词向量结合输入到实体抽取模块中,在该模块中采用了条件随机场(CRF)这一方法.具体来说,关系向量与词向量矩阵结合得到向量矩阵通过两组神经网络进行变换,将得到的结果再输入进条件随机场进行实体的抽取,得到了图中的实体对(中国,北京),最终该实体对与关系进行组合就抽取到句子中所包含的三元组.在训练阶段可以利用数据集中已知的关系使实体抽取模块不依赖于关系抽取模块的结果来训练,使其在训练阶段降低误差累积.而在预测

阶段则需要先识别出关系,依据关系找出实体.

2.1 关系抽取模块

关系抽取模块主要是完成句子中关系的识别,一个句子中会存在多个关系,因此该模块主要完成多分类任务,又由存在一个关系必定存在一对实体对的思想,所以在预测阶段,关系抽取模块对于后面的实体抽取模块起到了奠定基础的作用.而以关系为导向的实体关系抽取模型,少有在关系抽取上引用实体信息,这是由于关系抽取阶段实体信息未知.在本文中我们提出模拟实体的思想来进行引入实体信息,也即是每个词都可能蕴含实体信息,将每个词约视作实体,由此如果需要知道在某个关系下的实体信息,就是要知道词与词之间的重要性,因为一个关系存在则必定能对应一组实体对,那么在该关系下这组实体对之间关联程度,应当大于与其他实体的关联程度,简化思想即是求词与词之间的关联程度.如果两个字符的关联度大则说明这两个字符对关系具有一定的重要性,可以根据输入的编码单词向量计算出的模拟实体信息预估.由此,在获得单词向量的表示之后,可以使用自注意力机制关注字符与字符之间的重要程度从而模拟所需要的实体信息,并对已经获取到这种重要性的词向量矩阵进行平均池化来表示整个句子的实体信息.具体公式如下:

$$Entity_v = Avgpool\left(\text{Softmax}\left(\frac{QK^T}{\sqrt{dk}}\right)V\right) \in R^{d \times 1} \quad (1)$$

$$Q = W_{\text{query}} Y_{\text{enc}}(s) \quad (2)$$

$$V = W_{\text{value}} Y_{\text{enc}}(s) \quad (3)$$

$$K = W_{\text{key}} Y_{\text{enc}}(s) \quad (4)$$

$$\begin{cases} h^{\text{avg}} = Avgpool(Y_{\text{enc}}(s)) \in R^{d \times 1} \\ p_{\text{rel}} = \text{Sigmoid}(W_r[h^{\text{avg}}; Entity_v]) \in R^{n_r \times 1} \end{cases} \quad (5)$$

其中, $W_{\text{query}} \in R^{d_k \times d}$, $W_{\text{key}} \in R^{d_k \times d}$, $W_{\text{value}} \in R^{d \times d}$ 和 $W_r \in R^{n_r \times 2d}$ 皆是可训练参数矩阵, dk 是指线性变换后的维度. h^{avg} 是指代表整体句子信息向量, p_{rel} 是指正确答案为某关系的概率大小, $Entity_v$ 是指代表整体的模拟实体信息向量. 式 (1) 表示使用自注意力机制获取表征字符与字符之间的信息矩阵,并通过平均池化得到整体的字符信息从而模拟实体信息. 式 (5) 表示将式 (1) 生成的整体字符信息作为一个特征与能够表征整体句子信息的向量进行拼接并通过神经网络进行分类从而获取

句子关系信息.

2.2 实体抽取模块

从关系抽取模块中获取到句子关系向量后,将其与句子中的每个单词向量进行拼接,继而使用条件随机场识别出实体.实体的标注采用普通常用的标注方法即 BIO 标注法.计算标签 $y = \{y_1, y_2, \dots, y_n\}$ 的概率公式如下:

$$\begin{cases} V_{\text{mix}} = \text{Concat}(Y_{\text{enc}}(s), u_j) \in R^{2d \times n} \\ O = W_o V_{\text{mix}} + b_o \\ P(y|s) = \frac{\exp\left(\sum_i (O_{i,y_i} + T_{y_{i-1},y_i})\right)}{\sum_{y' \in Y_s} \exp\left(\sum_i (O_{i,y'_i} + T_{y_{i-1}',y'_i})\right)} \end{cases} \quad (6)$$

其中, $u_j \in R^{d \times 1}$ 是指关系向量, $W_o \in R^{n_l \times 2d}$ 和 $b_o \in R^{n_l \times n}$ 都是可训练参数矩阵, $O \in R^{n_l \times n_l}$ 是发射矩阵, $T \in R^{n_l \times n_l}$ 是转移矩阵, Y_s 是指序列中可能的标签, n_l 是指标签个数,然后再使用维特比算法解码.最后,将关系与实体对进行组合,获取三元组.

2.3 模型学习

在训练阶段,利用数据集的已知信息对模型进行训练,两个子任务不相互依赖,即在训练阶段,实体抽取模块不需要使用由关系模块中所获得的关系来进行训练,而是直接使用数据中真实的关系,由此,训练阶段避免了由关系不正确而导致的误差累积问题;在预测阶段,则是需要先通过关系抽取模块获得关系,再利用获得的关系进行实体的抽取,最后将获得实体与关系进行组合,将所抽取的头实体、尾实体与关系组合起来获得(头实体,关系,尾实体)这样的三元组.各部分以及总的损失函数如式 (7)–式 (9) 所示:

$$L_{\text{rel}} = -\frac{1}{n_r} \sum_{i=1}^{n_r} (x'_i \log P_{\text{rel}} + (1 - x'_i) \log(1 - P_{\text{rel}})) \quad (7)$$

$$L_{\text{ent}} = -\log \left(\frac{\exp\left(\sum_i (O_{i,y_i} + T_{y_{i-1},y_i})\right)}{\sum_{y' \in Y_s} \exp\left(\sum_i (O_{i,y'_i} + T_{y_{i-1}',y'_i})\right)} \right) \quad (8)$$

$$L = \alpha L_{\text{rel}} + \beta L_{\text{ent}} \quad (9)$$

我们尝试最小化 L 的损失函数,其中 x' 表示在关系抽取模块中正确的标签, n_r 是指关系个数, L 为总损失函数.若仔细调整每个子损失的权重,模型性能可能会更好,但为了简单起见,我们分配了相等的权重 ($\alpha = \beta = 1$).

3 实验分析

3.1 数据集及参数设置

数据集采用 NYT^[16] 和 WebNLG^[17] 数据集. NYT 数据集与 WebNLG 数据集分别可分为两种类型: 一类为实体边界标注完全的类型, 另一类型为仅标注实体最后一个单词. 本文使用的类型为实体边界标注完全类型, 即在 NYT 数据集上与 WebNLG 数据集上的实体边界均标注完全. NYT 数据集来源于纽约时报 New

York Times (NYT) 所标注的语料, 有 24 个有效关系. WebNLG 为自然语言生成任务而构建的数据集, 拥有 239 个有效关系. 具体信息如表 1 所示. 对于不同的数据集, 实验设定参数并不相同, 主要的实验参数设定与实验环境配值分别如表 2、表 3 所示.

表 1 数据集信息

数据集	训练集 (句子数)	测试集 (句子数)	验证集 (句子数)	关系数
NYT	56 195	5 000	5 000	24
WebNLG	5 019	703	500	239

表 2 实验参数

数据集	批处理大小	最大训练次数	神经元丢弃率	预训练模型微调学习率	下游任务学习率	句子最大长度	梯度的最大裁剪范数	权重衰减	逐步预热策略	逐步预热比率	优化器
NYT	64	100	0.7	0.000 01	0.000 1	100	2	0.01	余弦	0.1	BertAdam
WebNLG	32	100	0.3	0.000 1	0.001	100	2	0.01	余弦	0.1	BertAdam

表 3 实验环境

操作系统	PyTorch版本	CPU型号	GPU型号	Cuda版本
Ubuntu 18.04	1.8.1	Inter core i9	NVIDIA RTX A4000	11.1

我们采用标准查准率, 召回率以及 $F1$ 值作为我们的评判方法. 评判标准为所抽取的三元组中实体对与关系如果能够相互匹配, 则视为正确.

3.2 实验效果及分析

本文主要选取 CopyRE 等以关系为导向的模型与 TPLinker 等目前在 WebNLG 数据集与 NYT 数据集效果较好的模型作为比较参照对象, 具体结果如表 4 所示.

表 4 模型比较结果 (%)

模型	NYT			WebNLG		
	查准率	召回率	$F1$ 值	查准率	召回率	$F1$ 值
CopyRE	61.0	56.6	58.7	37.7	36.4	37.1
HRL	78.1	77.1	77.6	—	—	—
ETL-span ^[18]	85.5	71.7	78.0	84.3	82.0	83.1
RSAN	85.7	83.6	84.6	80.5	83.8	82.1
TPLinker	91.4	92.6	92.0	88.9	84.5	86.7
CasDE ^[19]	89.9	91.4	90.6	88.0	88.9	88.4
PMEI ^[20]	88.4	88.9	88.7	80.8	82.8	81.8
本文	89.2	91.5	90.3	88.9	88.6	88.8

通过对比效果可以发现, 在上述模型中, 本模型效果在以关系为导向的模型 (CopyRE、HRL、RSAN 与 CasDE) 中表现效果与 CasDE 表现效果相近. 具体来说, 本模型在 NYT 数据集上, $F1$ 值相较于 CasDE 效果低了 0.3%, 但在 WebNLG 数据集上 $F1$ 值则比 CasDE 高出 0.4%. 究其具体原因, CasDE 模型与本模型结构虽然相似, 但却有着根本不同, CasDE 模型主要采取一

种级联的方式, 即是在抽取完关系的基础上, 再结合关系去抽取头实体, 最后结合抽取到的头实体与关系再去抽取尾实体. 这样做的好处在于增加了实体抽取部分特征的多样性, 但坏处也同样明显, 如若中间头实体抽取错误很有可能导致尾实体抽取错误. 而本文模型在实体抽取模块并不采用这种级联结构, 是将实体成对抽取, 这样有效地避免了这种错误. 但是这样做在关系数较少的情况下, 容易过拟合, 因而本模型在 NYT 数据集下效果略低于 CasDE. 对比其他并不以关系为导向的模型 (ETL-span、TPLinker、PMEI), 本文模型效果在 WebNLG 数据集上表现较好而在 NYT 数据集上表现欠佳. 具体对比 TPLinker 模型, 本文模型在 NYT 数据集上 $F1$ 值低了 1.7%, 而在 WebNLG 数据集 $F1$ 值上高了 2.1%. 主要原因可能是由于 TPLinker 是采用了基于握手标注策略的端到端序列标注模型. 具体来说, 该模型将实体头部与实体尾部链接, 头实体头部与尾实体头部链接, 头实体尾部与尾实体尾部链接, 链接所采用的方式为构建一个矩阵, 在矩阵中进行标记. 接下来只需要训练模型来预测每个关系下的矩阵, 这样就可以实体关系同时训练, 同时预测, 不存在训练与预测顺序不一致问题 (即曝光偏差问题), 从而使其在某些数据集上有更好的表现. 而对于本文所提出的模型, 仍然存在一定的曝光偏差问题, 因此在 NYT 数据集上效果差于 TPLinker. 但在 WebNLG 数据集上, 本文提出的方法在召回率上高于 TPLinker, 这说明虽然本模型存在一定曝光偏差问题, 但总体模型结构仍有可取之处. 总之, TPLinker 精心设计了一组特殊的标注方式,

将该任务转化为序列标注任务从而到达实体关系可以同时抽取的目的,而对于本文所提出方法没有采用特殊标注方式,而是按照传统的标注方式对实体进行标注,训练阶段,实体抽取模块不需要使用由关系模块中所获得的关系来进行训练,而是直接使用数据中真实的关系,由此,训练阶段避免了由关系不正确而导致的误差累积问题,预测阶段再共同互相协助预测方式进

行预测.最后,综合对比 NYT 和 WebNLG 的识别结果,可以看出在关系数较多的情况下本文模型效果较好.这是由于关系数较多不容易导致过拟合的问题,因此该方法对于解决关系数较多的文本中的问题效果较好.

我们更进一步研究了其解决 SEO、EPO 以及 normal 问题的能力,以及存在一个或多个关系时模型抽取实体关系三元组的能力.具体结果如表 5 所示.

表 5 解决实体重叠或存在多个关系时的具体效果

数据集	指标	EPO	SEO	Normal	关系数=1	关系数=2	关系数=3	关系数=4	关系数≥5
NYT	查准率	0.928	0.927	0.847	0.844	0.922	0.943	0.946	0.881
	召回率	0.924	0.899	0.931	0.933	0.917	0.897	0.941	0.828
	F1值	0.926	0.913	0.887	0.886	0.919	0.920	0.943	0.859
WebNLG	查准率	0.960	0.905	0.830	0.808	0.909	0.909	0.921	0.871
	召回率	0.960	0.885	0.900	0.887	0.909	0.896	0.876	0.847
	F1值	0.960	0.895	0.863	0.845	0.909	0.903	0.898	0.858

由表 5 可知,本文所提出的模型在解决 SEO、EPO 问题上均好于处理 normal 问题.在 NYT 和 WebNLG 数据集上的 F1 值均在 90% 左右,说明该模型能够有效地解决实体重叠问题.并且当句子中存在 2 到 4 个关系时, F1 值也能保持在 90% 左右,这说明当句子中存在 2 到 4 个关系时,该方法效果较好,能够有效地抽取出实体关系.为进一步地研究本文方法解决实体重叠的能力,我们主要在关系数较多的 WebNLG 数据集上与以关系为导向的方法 (CasDE、RSAN 模型) 进行比较,结果如图 2 所示.

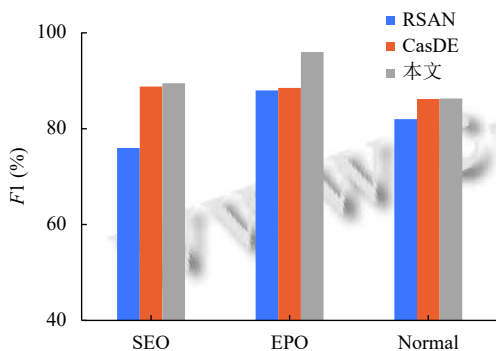


图 2 模型对比

如图 2 所示,在解决实体重叠问题,该方式相比于 RSAN 模型与 CasDE 模型都有提升,特别是在解决 EPO 问题效果最为显著,说明该方法在解决实体对重叠问题上效果较好,这是因为模型是以关系为导向,在给定关系的情况下,模型会尽可能地去识别实体对,并不会出现由于级联的方式带来的误差,从而在解决实

体对重叠问题上效果较好.综合来看,本文所提出的方法相对于一些以关系为导向的实体抽取方法,在处理各个实体重叠问题上并没有落后于大部分方法.

3.3 消融实验

为探究模型的各部分是否都起到了应有的作用,我们在 WebNLG 数据集上进行消融实验探究,其结果如表 6 所示.

表 6 消融实验结果

指标	自注意力机制	条件随机场	自注意力机制与条件随机场	原模型
查准率	0.879	0.451	0.471	0.889
召回率	0.878	0.861	0.854	0.886
F1值	0.879	0.592	0.607	0.888

由上述消融实验可以看出,各部分均起到了效果,其中条件随机场对于约束实体边界起到了较为重要的效果,自注意力机制的运用也对模型起到了积极提高的作用,这说明各部分均无冗余,均对模型的最终效果有着贡献.

4 结束语

本文主要借由实体信息对于关系抽取的效果有重要的影响的思想,采用以关系为导向的策略,来指导模型的构建,使得该方法更加有效的解决实体重叠问题.且从训练方式来看,两个模块在训练阶段中独立训练,缓解误差累积问题.提出引入模拟实体信息思想,模型采用自注意力机制关注字符信息以此来模拟实体信息,更加有效的提升整体效果,并且采用较为常用的方法

来进行三元组抽取,无复杂设计部分与较为特殊的训练方式,也易于理解与使用.且由消融实验可以看出自注意力机制对于模拟实体信息是有效的.

参考文献

- 1 Zeng DJ, Liu K, Lai SW, *et al.* Relation classification via convolutional deep neural network. Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics. Dublin: ACL, 2014. 2335–2344.
- 2 Shen YT, Huang XJ. Attention-based convolutional neural network for semantic relation extraction. Proceedings of the 26th International Conference on Computational Linguistics. Osaka: The COLING 2016 Organizing Committee, 2016. 2526–2536.
- 3 Qin PD, Xu WR, Guo J. Designing an adaptive attention mechanism for relation classification. Proceedings of 2017 International Joint Conference on Neural Networks (IJCNN). Anchorage: IEEE, 2017. 4356–4362.
- 4 Xu Y, Mon LL, Li G, *et al.* Classifying relations via long short term memory networks along shortest dependency paths. Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing. Lisbon: ACL, 2015. 1785–1794.
- 5 Miwa M, Bansal M. End-to-end relation extraction using LSTMs on sequences and tree structures. Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics. Berlin: ACL, 2016. 1105–1116.
- 6 Zheng SC, Wang F, Bao HY, *et al.* Joint extraction of entities and relations based on a novel tagging scheme. Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics. Vancouver: ACL, 2017. 1227–1236.
- 7 Bekoulis G, Deleu J, Demeester T, *et al.* Joint entity recognition and relation extraction as a multi-head selection problem. Expert Systems with Applications, 2018, 114: 34–45. [doi: [10.1016/j.eswa.2018.07.032](https://doi.org/10.1016/j.eswa.2018.07.032)]
- 8 Wei ZP, Su JL, Wang Y, *et al.* A novel cascade binary tagging framework for relational triple extraction. arXiv:1909.03227, 2019.
- 9 Wang YC, Yu BW, Zhang YY, *et al.* TPLinker: Single-stage joint extraction of entities and relations through token pair linking. Proceedings of the 28th International Conference on Computational Linguistics. Barcelona: International Committee on Computational Linguistics, 2020. 1572–1582.
- 10 Zeng XR, Zeng DJ, He SZ, *et al.* Extracting relational facts by an end-to-end neural model with copy mechanism. Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics. Melbourne: ACL, 2018. 506–514.
- 11 Takanobu R, Zhang TY, Liu JX, *et al.* A hierarchical framework for relation extraction with reinforcement learning. Proceedings of the 33rd AAAI Conference on Artificial Intelligence and 31st Innovative Applications of Artificial Intelligence Conference and 9th AAAI Symposium on Educational Advances in Artificial Intelligence. Honolulu: AAAI Press, 2019. 868.
- 12 Yuan Y, Zhou XF, Pan SR, *et al.* A relation-specific attention network for joint entity and relation extraction. Proceedings of the 29th International Joint Conference on Artificial Intelligence. Yokohama: IJCAI.org, 2021. 561.
- 13 Yu JT, Bohnet B, Poesio M. Named entity recognition as dependency parsing. arXiv:2005.07150, 2020.
- 14 Wu SC, He YF. Enriching pre-trained language model with entity information for relation classification. Proceedings of the 28th ACM International Conference on Information and Knowledge Management. Beijing: ACM, 2019. 2361–2364.
- 15 Peng H, Gao TY, Han X, *et al.* Learning from context or names? An empirical study on neural relation extraction. arXiv:2010.01923, 2020.
- 16 Riedel S, Yao LM, McCallum A. Modeling relations and their mentions without labeled text. Proceedings of the Joint European Conference on Machine Learning and Knowledge Discovery in Databases. Barcelona: Springer, 2010. 148–163.
- 17 Gardent C, Shimorina A, Narayan S, *et al.* Creating training corpora for NLG micro-planning. Proceedings of the 55th annual meeting of the Association for Computational Linguistics. Vancouver: ACL, 2017. 179–188.
- 18 Yu BW, Zhang ZY, Shu XB, *et al.* Joint extraction of entities and relations based on a novel decomposition strategy. Proceedings of the 24th European Conference on Artificial Intelligence. Santiago de Compostela: IOS Press, 2020. 2282–2289.
- 19 Ma LB, Ren HM, Zhang XL. Effective cascade dual-decoder model for joint entity and relation extraction. arXiv:2106.14163, 2021.
- 20 Sun K, Zhang RC, Mensah S, *et al.* Progressive multi-task learning with controlled information flow for joint entity and relation extraction. Proceedings of the 35th AAAI Conference on Artificial Intelligence. Palo Alto: AAAI Press, 2021. 13851–13859.

(校对责编:孙君艳)