

DE-AA:基于词对距离嵌入和轴向注意力机制的实体关系联合抽取模型

张梦赢, 沈海龙

引用本文

张梦赢, 沈海龙. [DE-AA:基于词对距离嵌入和轴向注意力机制的实体关系联合抽取模型](#)[J]. 计算机科学, 2024, 51(12): 234-241.

ZHANG Mengying, SHEN Hailong. [Joint Extraction of Entities and Relations Based on Word-Pair Distance Embedding and Axial Attention Mechanism](#) [J]. Computer Science, 2024, 51(12): 234-241.

相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[融合语义和句法图神经网络的实体关系联合抽取](#)

Fusion of Semantic and Syntactic Graph Convolutional Networks for Joint Entity and Relation Extraction

计算机科学, 2023, 50(9): 295-302. <https://doi.org/10.11896/jsjcx.220700041>

DE-AA:基于词对距离嵌入和轴向注意力机制的实体关系联合抽取模型

张梦赢 沈海龙

东北大学理学院 沈阳 110819

(zhangmengying0316@163.com)

摘要 实体关系联合抽取为知识图谱的构建提供了关键的技术支持,而重叠关系问题一直都是联合抽取模型研究的重点。现有的方法大多采用多步骤的建模方法,虽然在解决重叠关系问题上取得了很好的效果,但产生了曝光偏差问题。为同时解决重叠关系和曝光偏差问题,提出了一种基于词对距离嵌入和轴向注意力机制的实体关系联合抽取方法(DE-AA)。首先,构建代表词对关系的表特征,加入词对距离特征信息优化其表示;其次,应用基于行注意力和列注意力的轴向注意力模型去增强表特征,在融合全局特征的同时能够降低计算复杂度;最后,将表特征映射到各关系空间中,生成特定关系下的词对关系表,并使用表格填充法为表中各项分配标签,以三重分类的方式进行三元组的抽取。在公开数据集 NYT 和 WebNLG 上评估了所提出的模型,实验结果表明其与其他基线模型相比取得了更好的性能,且在处理重叠关系或多重关系问题上优势显著。

关键词: 实体关系联合抽取; 轴向注意力机制; 词对距离嵌入; 表格填充法

中图分类号 TP391.1

Joint Extraction of Entities and Relations Based on Word-Pair Distance Embedding and Axial Attention Mechanism

ZHANG Mengying and SHEN Hailong

School of Science, Northeastern University, Shenyang 110819, China

Abstract The joint extraction of entities and relations provides key technical support for the construction of knowledge graphs, and the problem of overlapping relations has always been the focus of joint extraction model research. Many of the existing methods use multi-step modeling methods. Although they have achieved good results in solving the problem of overlapping relations, they have produced the problem of exposure bias. In order to solve the problem of overlapping relations and exposure bias at the same time, a joint entities and relations extraction method (DE-AA) based on word-pair distance embedding and axial attention mechanism is proposed. Firstly, the table features of the representative word-pair relation are constructed, and the word-pair distance feature information is added to optimize its representation. Secondly, the axial attention model based on row attention and column attention is applied to enhance the table features, which can reduce the computational complexity while fusing the global features. Finally, the table features are mapped to each relation space to generate the relation-specific word-pair relation table, and the table filling method is used to assign labels to each item in the table, and the triples are extracted by triple classification. The proposed model is evaluated on the public datasets NYT and WebNLG. Experimental results show that the proposed model achieves better performance than other baseline models, and has significant advantages in dealing with overlapping relations or multiple relations.

Keywords Joint extraction of entities and relations, Axial attention mechanism, Word-Pair distance embedding, Table filling method

1 引言

知识图谱以结构化的表现形式呈现出知识间的关联关系,具有较强的可解释性,近年来已成为人工智能领域研究的热点^[1]。实体关系抽取通过从自然语言文本中抽取结构化三元组信息(头实体,关系,尾实体),为知识图谱的构建提供了关键的技术支持,这些结构化的信息在智能问答、推荐系统等下游任务中也得到了广泛的应用^[2-4]。

早期的抽取模型大多是基于流水线式(Pipeline)的方法,主要是将整个三元组抽取过程分为命名实体识别(Named Entity Recognition)和关系分类(Relation Classification)两个相互独立的子任务,先抽取句子中包含的实体,再针对实体对进行关系分类任务^[5-6]。流水线式方法的主要问题是容易出现误差传播问题,并且两个子任务间缺乏有效的信息交互,同时也产生了大量的信息冗余^[7]。为了解决这些问题,学者们提出了实体和关系的联合抽取方法(Joint Extraction)^[8-9]。

尽管联合抽取方法在两个子任务上都表现出了更优的性能,但仍存在诸多挑战。首先,重叠关系问题一直都是联合抽取模型研究的重点^[10-11]。一个句子中某实体对间可能存在着多种关系,或某实体与其他多个实体间存在着关系,交错复杂的实体关系极大地影响着模型的抽取效果^[2]。其次,曝光偏差(Exposure Bias)问题普遍存在于联合抽取模型中^[12]。现有基于深度学习的联合抽取模型大多采用多模块-多步骤的建模方法,它们将对三元组的抽取分解为多个对其组成元素抽取的模块^[13-14],如先对关系建模,再依据不同的关系类型抽取实体对,或是先识别头实体,再根据头实体抽取对应的关系和尾实体。多步骤的建模方法对三元组的抽取不是一步完成的,这就导致了曝光偏差的产生。

针对现有联合抽取模型存在的问题和不足,本文提出了一个基于词对距离嵌入和轴向注意力机制的联合抽取模型。模型最终会为每个关系生成一个词对关系表,通过表格填充法标注出头尾实体开始和结束位置对应的词对,以三重分类的方式对三元组进行联合抽取,该模型能够同时处理重叠关系和曝光偏差问题。首先,基于上下文表示和词对的相对距离特征信息生成一个二维的词对关系表;其次,为使词对充分交互,融合表中全局特征信息,应用基于行注意力和列注意力的轴向注意力模型对表特征进行建模,进一步将表特征映射到各关系空间中,从而为每个关系生成一个词对关系表;最后,为表中各项填充合适的标签,其标签指示了该特定关系的词对是否是头尾实体边界位置的词对,基于表格填充的结果进行搜索查询能够提取到所有的三元组。

本文的主要工作和贡献如下:

1)提出了融合词对距离嵌入的词对关系表建模方法。通过分析一个句子中三元组头尾实体的位置信息,认为组成头实体和尾实体的边界词对其相对距离是敏感的,因此加入词对距离特征信息进一步优化词对关系表的特征表示。

2)引入轴向注意力机制对词对表特征进行增强。轴向注意力模型产生于图像处理领域,本文提出将其应用到实体关系抽取任务中。模型主要采用双向的(行和列)轴向注意力模型来捕获表中全局信息,同时降低了计算复杂度。

3)分别在两类公开数据集 NYT 和 WebNLG 上进行了实验,本文模型的结果均优于其他基线模型。同时,针对重叠关系和多重关系的复杂场景进行了实验,结果也进一步证实了本文方法的有效性。

2 相关工作

2.1 实体关系抽取方法

实体关系抽取模型按照其建模方法的不同主要分为流水线式模型和联合抽取模型^[2]。流水线式模型将三元组抽取分为命名实体识别和关系分类两个子任务^[15-16]。首先,关系分类任务是基于命名实体识别任务而进行的,上一阶段任务产生的错误会积累传播给下一阶段的任务而造成误差传播问题^[17];其次,一个句子中往往只有少数实体间存在语义关系,而枚举所有可能的实体对进行关系分类造成了大量的信息冗余;另外,现有大量的实验已经证明关系信息对命名实体识别任务是有利的,而流水线式模型不能关联两个子任务的信息^[18]。

为避免流水线式模型的弊端,研究者们相继提出了各种联合抽取的模型,即以端到端的方式同时提取实体和关系。2016年,Miwa和Bansal^[19]提出一种基于序列和树状LSTM的联合抽取模型,该模型采用多步骤的抽取方法,通过参数共享的形式实现了模块间的交互。Zheng等^[20]将联合抽取任务转化为标记问题,采用特定的标签集标记出句子中三元组的头尾实体。Wang和Lu^[21]将联合抽取视为一种表格填充任务,在表中对角线和上下三角位置填充不同的标签,同时标注出句子中的实体及实体间的关系,以联合解码的形式来抽取三元组。这些模型虽实现了实体和关系的联合抽取并取得了不错的效果,但它们都不能完全解决重叠关系问题。重叠关系类型可分为以下3类:正常(Normal)、单实体重叠(SEO)和实体对重叠(EPO)^[6]。如图1所示,(a)中例句可抽取出两个三元组,且两个三元组之间没有包含相同的实体,这种属于正常情形;如果一个句子中某实体和其他多个实体间存在着关系,则属于单实体重叠的情况,如(b)中实体 Stephen Chow 和 China 之间存在 National 的关系,而又和 Hong Kong 之间存在 Born in 的关系;一个句子中某实体对之间存在着多种关系则属于实体对重叠情况,如(c)中实体 China 和 Beijing 之间同时存在关系 Capital City 和 Contains。

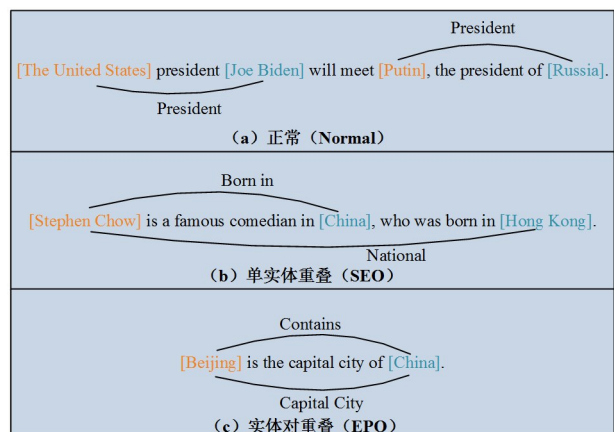


图1 重叠关系类型示例

Fig.1 Examples of overlapping relations types

重叠关系问题影响着模型抽取的性能,最近的一些联合抽取方法一直致力于解决这个问题。Dai等^[22]基于位置注意力机制,依据不同的查询位置生成不同的上下文表示,通过序列标记的方法解决重叠关系的抽取。Wei等^[10]将关系建模为头实体到尾实体的映射函数,并提出了一个新的标记框架。Yuan等^[11]基于关系感知注意力机制创建特定关系的句子表示,并识别对应的头尾实体。Zeng等^[23]提出一种基于复制机制的 seq2seq 模型,并采用两种不同的解码策略来联合抽取三元组。Eberts和Ulges^[24]采用跨度分类的方法来抽取实体,并依据实体的跨度表征、实体间的上下文特征和实体的长度嵌入进行关系分类任务。Fu等^[8]将文本序列建模为关系图,并应用图卷积网络来联合学习实体和关系,在重叠关系预测上有了很大的改进。Zheng等^[25]将整个抽取过程分解为关系筛选、实体抽取和头尾实体对齐3个子任务。这些模型在处理重叠关系问题上优势明显,但大多采用多步骤的建模方法,如实体域映射到关系域、关系域映射到实体域,

或头实体域映射到关系和尾实体域,这些多步骤的抽取顺序和方法都很难避免曝光偏差问题。为同时解决曝光偏差和重叠关系问题,Wang等^[12]采用词对链接的方法对齐每个关系下实体对的边界词来实现联合抽取。Ren等^[26]基于两类全局特征为每个关系构建一个表特征,采用表格填充的方法进行联合抽取。综上所述,为了解决重叠关系问题,大部分方法在模型设计上都存在曝光偏差问题。因此,本文提出了一个新颖的表格填充联合抽取模型,同时解决了以上两个问题。

2.2 轴向注意力模型

轴向注意力模型最初被提出应用于图像处理领域^[27],其将输入视为二维像素矩阵,并在特定的轴向上使用自注意力机制来捕捉全局特征信息。相比传统的自注意力模型^[28],其计算复杂度从 $O(nm \times nm)$ 降为 $O(nm + nm)$,这种优势在 n 和 m 较大时非常明显。采用轴向注意力既降低了计算复杂度,也实现了全局感受野。

在实体关系抽取任务中对轴向注意力模型的研究很

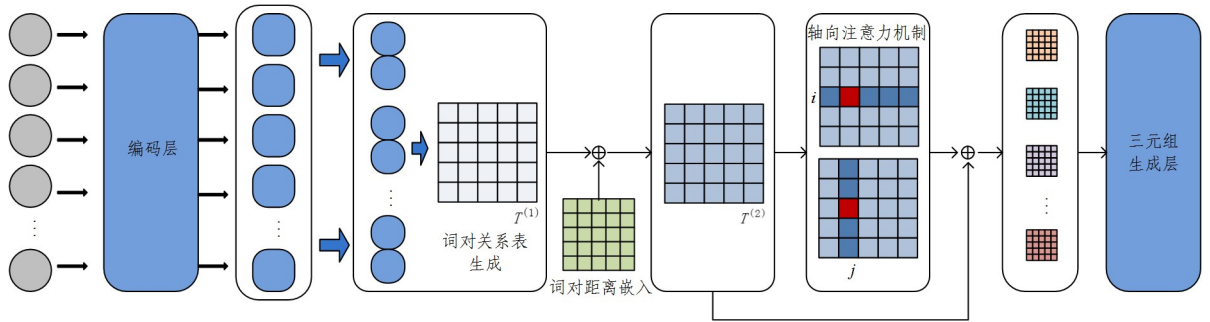


图2 模型整体框架

Fig. 2 Overall framework of the proposed model

3.1 任务定义

给定一个长度为 L 的句子 $S = \{x_i\}_{i=1}^L$,已知一个预定义关系集 $R = \{r_i\}_{i=1}^{|R|}$,关系集中关系的数量为 $|R|$ 。实体关系联合抽取任务的目的是从句子 S 中抽取出所有可能的三元组 $T = \{(h_i, r_i, t_i)\}_{i=1}^N$,这里 h_i 和 t_i 分别代表头实体和尾实体, r_i 表示它们之间的关系。

3.2 表格填充策略

本文模型最终会为每个关系生成一个表特征,因此表格填充的标签既要保证对实体对的有效识别,同时也要能够解决重叠关系的问题。在研究了过去的标记方案后^[21,26,29],最终选择以下标签集。以句子“John and Edward Thomas are from New York City, USA”为例,句子中包含的三元组有:(Edward Thomas, live in, New York City), (Edward Thomas, live in, New York), (Edward Thomas, live in, USA), (John, live in, New York City), (John, live in, New York)和(John, live in, USA)。该句子的表格填充结果如图3所示。

本文所选的表格填充标签分为以下5种组合类型:

1)MMB和MME:针对头实体和尾实体均是多词组成的情形。两个标签分别用于标记头实体及尾实体的开始词对和结束词对,后续的组合标签与之含义相同。例如:对于句中包含的三元组(Edward Thomas, live in, New York City)和(Edward Thomas, live in, New York),可以看出两个三元组

少见。为有效实现表特征增强策略,本文提出将基于文本序列生成的词对关系表视为二维输入,并应用基于列注意力和行注意力的双轴向注意力模型来建模,使其获得更好的特征表示。

3 模型介绍

本文提出的DE-AA模型由4部分组成:编码层、融合词对距离嵌入的表特征生成模块、基于轴向注意力机制的表特征增强模块和三元组生成层。在编码层,基于BERT模型生成上下文特征表示;在融合词对距离嵌入的表特征生成模块,通过特征拼接的方式构建代表词对关系的表特征,在此基础上融入词对距离信息来优化表特征表示;在基于轴向注意力机制的表特征增强模块,应用轴向注意力进一步增强表特征表示,并将其映射到各关系空间中;在三元组生成层,在对应关系下为表中各项填充最大概率标签,从而标注出头尾实体开始和结束位置对应的词对。DE-AA模型的框架如图2所示。

的头实体和尾实体均是多词组成的实体,因此在关系为“live in”的填充表格中,词对(Edward, New)会被标记为MMB,而词对(Thomas, City)和(Thomas, York)均会被标记为MME。

2)MSB和MSE:针对头实体是多词组成的而尾实体是单个词组成的情形。例如:对于句中出现的三元组(Edward Thomas, live in, USA),词对(Edward, USA)会被标记为MSB,而词对(Thomas, USA)会被标记为MSE。

3)SMB和SME:针对头实体是单个词组成的而尾实体是多词组成的情形。例如:对于句中出现的三元组(John, live in, New York City)和(John, live in, New York)而言,词对(John, New)的位置会被标记为SMB,而词对(John, City)和(John, York)的位置均会被标记为SME。

4)S-S:针对头实体和尾实体均是由单个词组成的情形。例如:对于句子中出现的三元组(John, live in, USA),词对(John, USA)会被标记为S-S。

5)-:除上述的情况外,表中所有其他的单元格将被标记为“-”。

对于表格填充结果,通过正向和反向搜索来提取三元组。正向搜索路径包括:MMB→MME, MSB→MSE, SMB→SME和S-S;反向搜索路径包括:MME→MMB, MSE→MSB, SME→SMB。对于例句中的SEO情形,如三元组(John, live in, New York City)和(John, live in, New York),采用正向

搜索策略,当检测到词对(John,New)的标签 SMB 时,根据搜索路径首先会查询到词对(John, York)的标签 SME,此时搜索进程将结束,这就导致了对三元组(John, live in, New York City)的遗漏。添加反向搜索路径能够保证对三元组查询搜索的完整性。

本文的标记策略仅标记构成头尾实体开始和结束位置的词对(边界词对),因此生成的标记矩阵具有稀疏性,这极大地简化了解码的过程。同时,该方法可以很好地处理重叠关系问题。对于 EPO 情形,能通过不同的关系解码器进行抽取。对于 SEO 情形,如果两个三元组中的关系不同,可通过不同的关系解码器进行解码;如果关系相同,但该种情况只是单实体重叠,另一实体间存在不同的边界词,因此依然可以处理。

	...	New	York	City	,	USA
John		SMB	SME	SME		S-S
and						
Edward		MMB				MSB
Thomas			MME	MME		MSE
...						

图3 表格填充示例

Fig. 3 Examples of table filling

3.3 编码层

为有效增强全局上下文感知,获得更好的句子语义特征表示,本文采用基于 Transformer 架构的预训练语言 BERT 模型对输入句子进行编码^[30]。具体而言,对于输入句子 $S = \{x_1, x_2, \dots, x_L\}$,应用模型得到上下文特征表示 $H = \{h_1, h_2, \dots, h_L\}$,具体计算式如式(1)所示:

$$H = Bert(S) \quad (1)$$

3.4 融合词对距离嵌入的表特征生成模块

为获得两个词之间的关系表示,采用特征拼接的形式。对于词对 (h_i, h_j) ,首先将 h_i 和 h_j 按照某一维度进行特征拼接,然后经过线性层输出词对表示 h_{ij} ,计算式如式(2)所示。对于一个句子序列,每两个词生成一个关系表示,在整体上共同构成了一个二维的表特征,即图 2 中的 $T^{(1)}$ 。

$$h_{ij} = \mathbf{W}_1 [h_i; h_j] + \mathbf{b}_1 \quad (2)$$

其中, \mathbf{W}_1 和 \mathbf{b}_1 为可训练的参数矩阵和偏置向量, $h_{ij} \in R^{d_h}$, d_h 表示隐藏层维度大小。

对于一个正确的三元组而言,通常认为头实体和尾实体不是同一实体,且在大多数句法结构中,两实体之间应包含着关系指示词汇,同时实体的跨度均是大于等于 1 的。因此,当词与词的相对距离非常小时,两词分别构成头实体及尾实体的开始边界词的概率几乎为零。基于以上考虑,认为词与词

之间的关系对其相对距离是敏感的,因此加入距离特征信息进一步优化词对关系表的特征表示。具体的计算式如式(3)所示:

$$l_{ij} = \mathbf{W}_2 \tanh(drop([h_{ij}; d_{ij}])) + \mathbf{b}_2 \quad (3)$$

其中, d_{ij} 表示第 i 位置和第 j 位置词对的相对距离嵌入, \mathbf{W}_2 和 \mathbf{b}_2 是可训练的参数矩阵和偏置向量。将融合词对相对距离嵌入的表特征记为 $T^{(2)}$ 。

3.5 基于轴向注意力机制的表特征增强模块

为增强表特征表示,捕捉表中词与词之间一些潜在的依赖关系,并有效地融合表中的全局信息,本文提出使用轴向注意力模型进行建模的一种新颖应用方法^[27]。通过图 3 中表格填充的例子可以看出,针对头实体和尾实体至少一个是单个词组成的情况,开始位置和结束位置的词对标签位于同一行、同一列或同一位置,且整个表格的标记是具有稀疏性的,采用轴向注意力模型进行建模既能保证模型的效果也能降低计算复杂度。给定表 $T^{(2)}$ 中某词对的关系表示 l_{ij} ,使用基于行注意力和列注意力的轴向注意力模型进行计算,如式(4)所示:

$$\begin{aligned} Att_{row} &= MultiHead(l_{ij}, row_i, row_i) \\ Att_{col} &= MultiHead(l_{ij}, col_j, col_j) \\ a_{ij} &= Att_{row} + Att_{col} \end{aligned} \quad (4)$$

其中, $MultiHead$ 表示多头自注意力机制, row_i 和 col_j 分别代表第 i 行和第 j 列。

最终,对于表中 (i, j) 位置的词对,将词对关系表示 l_{ij} 、轴向注意力机制的输出结果 a_{ij} 和对称方向上轴向注意力机制的输出结果 a_{ji} 进行拼接得到最终的表示,并将其输入到线性层生成特定关系下的表特征,如式(5)所示:

$$t_r(i, j) = \mathbf{W}_r ([a_{ij}; l_{ij}; a_{ji}]) + \mathbf{b}_r \quad (5)$$

其中, $r \in R$ 表示某个特定的关系; \mathbf{W}_r 和 \mathbf{b}_r 分别是可训练的权重矩阵和偏置向量; $t_r(i, j) \in R^{1 \times C}$, C 表示标签集中标签的数量,这里 $C=8$ 。

3.6 三元组生成层

基于关系特定的表特征,为表中的每项填充合适的标签。给定 $t_r(i, j)$,其表示在关系 r 下表中的 (i, j) 位置的词对在各个标签下的得分,应用 softmax 和 argmax 函数进行预测,计算式如式(6)所示,其中 $l \in C$ 。

$$\begin{aligned} P_r(y_{i,j}) &= \text{softmax}(t_r(i, j)) \\ Table_r(i, j) &= \text{argmax}_l (P_r(y_{i,j}) = l) \end{aligned} \quad (6)$$

3.7 损失函数

模型损失函数定义如下:

$$\begin{aligned} L_{\text{Triple}} &= -\frac{1}{L \times L \times |R|} \sum_{i=1}^L \sum_{j=1}^L \sum_{r=1}^{|R|} \log(y_r(i, j)) \\ &= Table_r(i, j) \end{aligned} \quad (7)$$

其中, $Table_r(i, j)$ 表示关系 r 下表中的 (i, j) 位置词对标签填充的预测结果, $y_r(i, j)$ 表示该位置的真实标签。

4 实验及结果分析

4.1 数据集及评估指标

在两个广泛使用的公开数据集 NYT^[31] 和 WebNLG^[32] 上评估本文模型。其中, NYT 数据集由纽约时报语料库通过

远程监督的方式生成, WebNLG 数据集产生于自然语言生成任务。根据标注标准的不同, 两个数据集分别有两个版本: 1) 标注实体的最后一个单词; 2) 标注整个实体的跨度。将第一个版本的数据集记为 NYT* 和 WebNLG*, 将第二

个版本的数据集记为 NYT 和 WebNLG。此外, 为评估模型处理复杂场景(重叠关系或多重关系)的能力, 本文进一步按重叠关系类型和句子中包含三元组的数量对数据集进行了拆分。4 个数据集的详细统计结果如表 1 所列。

表 1 数据集的统计信息

Table 1 Statistical information of datasets

Datasets	Sentences			Overlapping relations			Number of triples					Triples	Relations
	Train	Valid	Test	Normal	SEO	EPO	N=1	N=2	N=3	N=4	N \geq 5		
NYT*	56 195	4 999	5 000	3 266	1 297	978	3 244	1 045	312	291	108	8 110	24
WebNLG*	5 019	500	703	245	457	26	266	171	131	90	45	1 591	171
NYT	56 195	5 000	5 000	3 071	1 273	1 168	3 089	1 137	300	317	157	8 616	24
WebNLG	5 019	500	703	239	448	6	256	175	138	93	41	1 607	216

表 2 不同数据集上的实验结果对比

Table 2 Comparison of experimental results on different datasets

Models	NYT*			WebNLG*			NYT			WebNLG		
	Prec.	Rec.	F1	Prec.	Rec.	F1	Prec.	Rec.	F1	Prec.	Rec.	F1
CasRel	89.7	89.5	89.6	93.4	90.1	91.8	—	—	—	—	—	—
RSAN	—	—	—	—	—	—	85.7	83.6	84.6	80.5	83.8	82.1
ETL-Span	84.9	72.3	78.1	84.0	91.5	87.6	85.5	71.1	78.0	84.3	82.0	83.1
GraphRel _{1p}	62.9	57.3	60.0	42.3	39.2	40.7	—	—	—	—	—	—
GraphRel _{2p}	63.9	60.0	61.9	44.7	41.1	42.9	—	—	—	—	—	—
CopyRE _{Multi}	61.0	56.6	58.7	37.7	36.4	37.1	—	—	—	—	—	—
EmRel	91.7	92.5	92.1	92.7	93.0	92.9	92.6	92.7	92.6	90.2	87.4	88.7
TPLinker	91.3	92.5	91.9	91.8	92.0	91.9	91.4	92.6	92.0	88.9	84.5	86.7
Ours	92.9	91.8	92.3	94.2	93.7	93.9	91.7	93.0	92.4	90.7	89.2	90.0

本文选取准确率(Prec.)、召回率(Rec.)和 F1 值(F1)作为模型的评价指标。对数据集 NYT* 和 WebNLG* 采用部分匹配策略, 即只有当关系、头实体和尾实体的最后一个词都正确时, 才认为提取的三元组是正确的。对数据集 NYT 和 WebNLG 采用精确匹配策略, 即只有当两个实体和关系的整个跨度都精确匹配时, 预测的三元组才被认为是正确的。

4.2 实验环境

所有实验在 PyTorch2.0.0, Python3.8, Cuda11.8 的环境下进行。训练过程所使用的 GPU 为 RTX 4090 (24 GB)。对于预训练语言模型, 采用 BERT-Base-Cased 的版本, 保持默认的参数设置。在 NYT 和 WebNLG 两类数据集上设置批次大小分别为 12 和 6, 学习率分别为 1×10^{-5} 和 5×10^{-5} 。所有参数都通过 Adam 算法进行优化。式(3)中的 dropout 概率为 0.1, 多头注意力机制头数为 8, 最大序列长度设置为 100。

4.3 基线模型

为有效验证本文模型的性能, 选取以下基线模型进行比较:

1) CasRel^[10]: 在解决重叠关系问题上提出了一个新的级联二进制标记框架, 将关系建模为头实体映射到尾实体的函数。

2) RSAN^[11]: 应用基于关系的注意力机制创建特定关系的上下文表示, 并基于此进行序列标记提取对应的头实体和尾实体。

3) ETL-Span^[33]: 将联合抽取分为两个关联的子任务, 先识别头实体, 再为每个头实体确定对应的尾实体和关系。

4) GraphRel^[8]: 将文本序列建模为关系图, 提出了一种基于图卷积网络的端到端联合抽取模型。

5) CopyRE^[23]: 提出一种基于 seq2seq 的联合抽取模型, 采用复制机制从句子中提取三元组, 有效地解决了

重叠关系的问题。

6) EmRel^[34]: 同时对文本序列和关系进行建模, 通过注意力机制来交互融合头实体、尾实体和关系之间的信息进而捕获三重关联的表示。

7) TPLinker^[12]: 将联合抽取任务建模为词对链接问题, 并引入一种新的标记方案对每个关系类型下实体对的边界词进行对齐。

4.4 实验结果

本文模型与其他基线模型在所有数据集上的实验结果如表 2 所列, 其中每列加粗标记表示最优结果。可以看出, 本文模型(DE-AA)在各数据集上的效果都有所提升。其中, 在数据集 WebNLG, WebNLG* 和 NYT* 上的 F1 值取得了最好结果, 比最佳的基线模型分别提升了 1.3%, 1.0% 和 0.2%, 但在 NYT 数据集上, 其 F1 值要低于 EmRel 模型。整体而言, 模型 DE-AA 在 WebNLG 类数据集上的表现更优, 原因在于: 首先, WebNLG 和 WebNLG* 数据集中分别包含了 216 和 171 种关系, 这给三元组抽取带来了很大的难度, 而本文模型采用一种词对关系表的建模方法, 该方法最终会为每个关系生成一个表特征, 一定程度上有效地捕获了三元组中头实体、关系和尾实体的三重依赖关系, 提高了模型抽取的准确性; 其次, 本文模型从三重分类的角度来实现三元组的联合抽取, 也避免了曝光偏差问题的产生。实验结果充分说明了本文方法在三元组抽取任务是有效的。

表 3 消融实验结果

Table 3 Ablation experimental results

Models	Prec.	Rec.	F1
DE-AA	94.2	93.7	93.9
-词对距离嵌入	93.6	93.3	93.4
-轴向注意力机制	93.3	93.3	93.4

4.5 消融实验

本文进行了消融实验来研究模型中词对距离嵌入和轴向注意力机制两组件的有效性。表3列出了模型DE-AA在WebNLG^{*}数据集上的消融实验结果,该数据集包含较多的关系类型,可以代表更多现实世界中的场景。依据消融实验结果可以得出结论:词对距离嵌入和轴向注意力机制对三元组抽取任务的性能有显著影响。具体而言,在构建词对关系表示时不考虑词与词间的相对距离信息,只依赖词本身的特征信息进行简单的拼接操作生成词对关系表示,模型性能将有所下降,说明词对距离信息的嵌入对于实体关系抽取任务是有利的,它优化了词对关系表特征,提高了实体对识别的准确性。当缺少轴向注意力模块时,模型的准确率下降更加明显。这是因为缺少轴向注意力模块会导致无法实现词对关系表中全局信息的融合,也无法捕获头尾实体组成词之间的长距离依赖关系,进而导致三元组抽取准确率的降低。

4.6 复杂场景分析

为了验证本文模型在处理重叠关系和多重关系(句子中包含多个三元组)方面的能力,在NYT^{*}和WebNLG^{*}数据集上做了更详细的实验。根据重叠关系类型的不同对模型进行了分类测试,与基线模型的对比结果如表4所列。可以看出,DE-AA几乎在所有测试项上都优于其他基线模型。在处理Normal类型的句子时,DE-AA在数据

集WebNLG^{*}上的效果要低于CasRel,但在处理SEO和EPO类型的句子时F1值较CasRel分别提升了2.4%和1.2%。DE-AA在场景SEO或EPO上的得分要高于Normal的情形。实验结果充分说明了DE-AA在处理重叠关系问题上的显著优势。

表4 从不同重叠类型的句子中抽取三元组的F1值

Table 4 F1-score of extracting triples from sentences with different overlapping types

Models	NYT [*]			WebNLG [*]		
	Normal	SEO	EPO	Normal	SEO	EPO
CasRel	87.3	91.4	92.0	89.4	92.2	94.7
GraphRel _{1p}	67.4	49.9	56.5	62.7	36.4	39.5
GraphRel _{2p}	69.6	51.2	58.2	65.8	38.3	40.6
CopyRE _{Mul}	66.0	48.6	55.0	—	—	—
TPLinker	90.1	93.4	94.0	87.9	92.5	95.3
Ours	90.4	94.0	94.2	89.1	94.6	95.9

根据句子所含三元组数量的不同,对模型进行了分类测试。与基线模型的对比结果如表5所列,其中 N 表示句子中包含1,2,3,4或5个及以上数量的三元组。可以看出,整体上DE-AA的效果要优于基线模型,且当处理三元组数量大于等于5的句子时,其在WebNLG^{*}和NYT^{*}数据集上的效果较最优基线模型分别提升了3.1%和1.8%。实验结果也证实了DE-AA具有处理多重关系的能力。

表5 抽取句子中不同数量三元组的F1值

Table 5 F1-score of extracting different numbers of triples from sentences

Models	NYT [*]					WebNLG [*]				
	$N=1$	$N=2$	$N=3$	$N=4$	$N \geq 5$	$N=1$	$N=2$	$N=3$	$N=4$	$N \geq 5$
CasRel	88.2	90.3	91.9	94.2	83.7	89.3	90.8	94.2	92.4	90.9
GraphRel _{1p}	69.1	59.5	54.4	53.9	37.5	63.8	46.3	34.7	30.8	29.4
GraphRel _{2p}	71.0	61.5	57.4	55.1	41.1	66.0	48.3	37.0	32.1	32.1
CopyRE _{Mul}	67.1	58.6	52.0	53.6	30.0	59.2	42.5	31.7	24.2	30.0
TPLinker	90.0	92.8	93.1	96.1	90.0	88.0	90.1	94.6	93.3	91.6
Ours	90.5	93.1	93.7	95.7	91.8	89.0	94.1	95.7	94.5	94.7

4.7 三元组不同元素的抽取结果

表6列出了CasRel和本文模型DE-AA在数据集NYT^{*}和WebNLG^{*}上对三元组不同元素的提取性能。 (h, r, t) 表示一个三元组,其中 h 表示头实体, t 表示尾实体, r 表示 h 与 t 之间的关系。当预测三元组 (h, r, t) 中的头实体和尾实体都正确时,认为元素 (h, t) 的提取是正确的。类似地,当预测三元组 (h, r, t) 中的关系是正确的,认为元素 r 的提取是正确的。从表中可以看出,对于 (h, t) 的提取,DE-AA的所有测试得分都高于CasRel,F1值在数据集NYT^{*}和WebNLG^{*}上分别提升了2.9%和2.0%。在关系 r 抽取任务中,DE-AA在数据集NYT^{*}上的F1值略低于CasRel,但在WebNLG^{*}数据集上的效果显著优于CasRel,F1值提升了2%。整体结果表明,本文模型能够识别更多的实体和关系,在三元组抽取任务上获得了更好的性能。此外,对比 (h, t) 抽取和 r 抽取的得分可以发现,对 (h, t) 抽取的效果要普遍低于 r ,这也反映了命名实体识别的难度要高于关系分类任务。

表6 三元组不同元素的抽取结果

Table 6 Results of extracting different elements of triples

Models	Components	NYT [*]			WebNLG [*]		
		<i>Prec.</i>	<i>Rec.</i>	<i>F1</i>	<i>Prec.</i>	<i>Rec.</i>	<i>F1</i>
CasRel	(h, t)	89.2	90.1	89.7	95.3	91.7	93.5
	r	96.0	93.8	94.9	96.6	91.5	94.0
	(h, r, t)	89.7	89.5	89.6	93.4	90.1	91.8
Ours	(h, t)	92.9	92.4	92.6	95.8	95.3	95.5
	r	96.7	95.5	94.4	96.7	95.3	96.0
	(h, r, t)	92.9	91.8	92.3	94.2	93.7	93.9

4.8 模型效率

表7中列出了模型TPLinker、-DE-AA(无词对距离嵌入和轴向注意力模块的DE-AA模型)和DE-AA在数据集WebNLG^{*}上训练时间、推理时间和F1值的比较结果。模型在训练和测试时的批次大小分别设置为6和1,其他实验环境与4.2节保持一致。表中的Training Time(s)指训练一个epoch所需的时间,Inference Time(s)指预测整个测试集所需的时间。可以看出,模型DE-AA的训练和推理时间都少于TPLinker,且F1值要比TPLinker高2%,这充分说明了本文

模型的效率和学习能力。另外,去除词对距离嵌入和轴向注意力模块的-DE-AA模型虽然训练时间和推理时间短很多,但抽取效果与DE-AA相比也大大降低。说明词对距离嵌入和轴向注意力模块的存在能够很大程度上提升模型的抽取性能。

表7 模型效率的比较

Table 7 Comparison of model efficiencies

Models	Training Time	Inference Time	F1
TPLinker	486	24.4	91.9
-DE-AA	60	9.6	92.8
DE-AA	114	19.4	93.9

4.9 案例研究

图4给出了模型DE-AA和-DE-AA在WebNLG*数据

Texts	Ground Truth	-DE-AA	DE-AA
The [Apollo 11] program was organized by [NASA] and included [Buzz Aldrin] as one of its crew members.	(11, operator, NASA) (Aldrin, was a crew member of, 11)	(11, operator, NASA) (11, backup pilot, Aldrin)	(11, operator, NASA) (Aldrin, was a crew member of, 11)
[Antwerp International Airport] serves the city of [Antwerp] in [Belgium] where the leader is [Phillipe of Belgium]	(Antwerp, country, Belgium) (Airport, cityServed, Antwerp) (Belgium, leaderName, Belgium)	(Antwerp, country, Belgium) (Airport, cityServed, Antwerp)	(Antwerp, country, Belgium) (Airport, cityServed, Antwerp) (Belgium, leaderName, Belgium)

图4 DE-AA和-DE-AA模型的案例研究

Fig. 4 Case studies of DE-AA and -DE-AA

结束语 为解决实体关系抽取任务中存在的重叠关系和曝光偏差问题,本文提出了一种基于词对距离嵌入和轴向注意力机制的联合抽取模型(DE-AA)。该模型主要通过构建词对的关系表示,并将其映射到特定的关系空间中,以三重分类的方式对三元组进行抽取。模型在生成词对关系表示的基础上加入了词对的距离特征信息,并应用双向(行和列)的轴向注意力模型去进一步增强其特征表示,充分考虑了词对的关联特征,保证了全局信息的融合,同时也降低了计算复杂度。在两类公开数据集上对所提出的模型进行了评估,结果证实了模型的有效性,详细的实验也进一步说明了该模型具有处理包含重叠关系或多重关系句子的能力。未来,将针对联合抽取模型进行更深入的研究,提升模型的效果和计算效率,并将其应用到更多的实际领域。

参考文献

[1] HOGAN A, BLOMQUIST E, COCHEZ M, et al. Knowledge graphs[J]. ACM Computing Surveys, 2021, 54(4): 1-37.

[2] NASAR Z, JAFFRY S W, MALIK M K. Named entity recognition and relation extraction: State-of-the-art[J]. ACM Computing Surveys, 2021, 54(1): 1-39.

[3] HAO Y, ZHANG Y, LIU K, et al. An end-to-end model for question answering over knowledge base with cross-attention combining global knowledge[C]// Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). 2017: 221-231.

[4] WANG X, HE X, CAO Y, et al. Kgat: Knowledge graph attention network for recommendation[C]// Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. 2019: 950-958.

集上的两个抽取示例。由于该数据集模型的预测结果仅标注出实体的最后一个词,为方便观察,在原文本中标注出了实体的真实跨度,使用中括号并加粗进行表示,图中实体或关系预测错误的情形用红色标记,其余预测正确的实体用加粗表示,预测正确的关系用其他不同颜色标记。可以看出,第一句话和第二句话中均包含了多个三元组,且都存在SEO(单实体重叠)的重叠关系情景,属于复杂语句。从结果可以看出,DE-AA模型都能准确提取关系,而-DE-AA在第一句话中未能正确预测出Apollo 11和Buzz Aldrin之间的crew member关系,在第二句话中没有预测出Belgium和Phillipe of Belgium之间存在的leaderName的关系。缺少词对距离嵌入和轴向注意力模块的-DE-AA模型很难完全识别出所有正确的三元组,这也进一步证明了本文模型的有效性和合理性。

[5] LI J, SUN A, HAN J, et al. A survey on deep learning for named entity recognition[J]. IEEE Transactions on Knowledge and Data Engineering, 2020, 34(1): 50-70.

[6] KAMBAR M E Z N, ESMAEILZADEH A, HEIDARI M. A survey on deep learning techniques for joint named entities and relation extraction[C]// 2022 IEEE World AI IoT Congress (AI-IoT). IEEE, 2022: 218-224.

[7] ZHANG X W, WANG X, CHEN Z R, et al. Survey of Supervised Joint Entity Relation Extraction Methods[J]. Journal of Frontiers of Computer Science & Technology, 2022, 16(4): 713-733.

[8] FU T J, LI P H, MA W Y. Graphrel: Modeling text as relational graphs for joint entity and relation extraction[C]// Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. 2019: 1409-1418.

[9] YAN Z, ZHANG C, FU J, et al. A partition filter network for joint entity and relation extraction[J]. arXiv: 2108.12202, 2021.

[10] WEI J, SU J, WANG Y, et al. A novel cascade binary tagging framework for relational triple extraction [J]. arXiv: 1909.03227, 2019.

[11] YUAN Y, ZHOU X, PAN S, et al. A relation-specific attention network for joint entity and relation extraction[C]// International Joint Conference on Artificial Intelligence. 2021.

[12] WANG Y, YU B, ZHANG Y, et al. TPLinker: Single-stage joint extraction of entities and relations through token pair linking [J]. arXiv: 2010.13415, 2020.

[13] REN F, ZHANG L, ZHAO X, et al. A simple but effective bidirectional framework for relational triple extraction[C]// Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining. 2022: 824-832.

- [14] REN F,ZHANG L,YIN S,et al. A Conditional Cascade Model for Relational Triple Extraction[C]//Proceedings of the 30th ACM International Conference on Information & Knowledge Management. 2021;3393-3397.
- [15] GOYAL A,GUPTA V,KUMAR M. Recent named entity recognition and classification techniques:a systematic review[J]. Computer Science Review,2018,29:21-43.
- [16] LIU X,CHEN H,XIA W. Overview of named entity recognition [J]. Journal of Contemporary Educational Research, 2022, 6(5):65-68.
- [17] LI Q,JI H. Incremental joint extraction of entity mentions and relations[C]//Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics(Volume 1:Long Papers). 2014:402-412.
- [18] ZHAO K,XU H,CHENG Y,et al. Representation iterative fusion based on heterogeneous graph neural network for joint entity and relation extraction[J]. Knowledge-Based Systems,2021, 219:106888.
- [19] MIWA M,BANSAL M. End-to-end relation extraction using lstms on sequences and tree structures[J]. arXiv:1601.00770, 2016.
- [20] ZHENG S,WANG F,BAO H,et al. Joint extraction of entities and relations based on a novel tagging scheme[J]. arXiv:1706.05075,2017.
- [21] WANG J,LU W. Two are better than one:Joint entity and relation extraction with table-sequence encoders[J]. arXiv:2010.03851,2020.
- [22] DAI D,XIAO X,LYU Y,et al. Joint extraction of entities and overlapping relations using position-attentive sequence labeling [C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2019:6300-6308.
- [23] ZENG X,ZENG D,HE S,et al. Extracting relational facts by an end-to-end neural model with copy mechanism[C]//Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics(Volume 1:Long Papers). 2018:506-514.
- [24] EBERTS M,ULGES A. Span-based joint entity and relation extraction with transformer pre-training[J]. arXiv:1909.07755, 2019.
- [25] ZHENG H,WEN R,CHEN X,et al. PRGC:Potential relation and global correspondence based joint relational triple extraction [J]. arXiv:2106.09895,2021.
- [26] REN F,ZHANG L,YIN S,et al. A novel global feature-oriented relational triple extraction model based on table filling[J]. arXiv:2109.06705,2021.
- [27] HO J,KALCHBRENNER N,WEISSENBORN D,et al. Axial attention in multidimensional transformers [J]. arXiv:1912.12180,2019.
- [28] VASWANI A,SHAZEER N,PARMAR N,et al. Attention is all you need[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. 2017:6000-6010.
- [29] SHANG Y M,HUANG H,MAO X. Onerel:Joint entity and relation extraction with one module in one step[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2022:11285-11293.
- [30] DEVLIN J,CHANG M W,LEE K,et al. Bert:Pre-training of deep bidirectional transformers for language understanding[J]. arXiv:1810.04805,2018.
- [31] RIEDEL S,YAO L,MCCALLUM A. Modeling relations and their mentions without labeled text[C]//Machine Learning and Knowledge Discovery in Databases; European Conference, ECML PKDD 2010, Barcelona, Spain, September 20 – 24, 2010, Proceedings,Part III 21. Springer Berlin Heidelberg,2010:148-163.
- [32] GARDENT C,SHIMORINA A,NARAYAN S,et al. Creating training corpora for nlg micro-planning[C]//55th annual meeting of the Association for Computational Linguistics(ACL). 2017.
- [33] YU B,ZHANG Z,SHU X,et al. Joint extraction of entities and relations based on a novel decomposition strategy [J]. arXiv:1909.04273,2019.
- [34] XU B,WANG Q,LYU Y,et al. EmRel:Joint Representation of Entities and Embedded Relations for Multi-triple Extraction [C]//Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. 2022:659-665.



ZHANG Mengying, born in 2000, post-graduate. Her main research interests include knowledge graph and information extraction.



SHEN Hailong, born in 1971, Ph.D, associate professor. His main research interests include data analysis and intelligent computing.

(责任编辑:何杨)