

融合语义和句法图神经网络的实体关系联合抽取

衡红军, 苗菁

引用本文

衡红军, 苗菁. 融合语义和句法图神经网络的实体关系联合抽取[J]. 计算机科学, 2023, 50(9): 295-302.

HENG Hongjun, MIAO Jing. [Fusion of Semantic and Syntactic Graph Convolutional Networks for Joint Entity and Relation Extraction](#) [J]. Computer Science, 2023, 50(9): 295-302.

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)

[上下文信息融合与噪声自适应的异常检测方法](#)

Anomaly Detection Method Based on Context Information Fusion and Noise Adaptation

计算机科学, 2023, 50(7): 237-245. <https://doi.org/10.11896/jsjcx.220700078>

[基于多事件语义增强的情感分析](#)

Sentiment Analysis Based on Multi-event Semantic Enhancement

计算机科学, 2023, 50(5): 238-247. <https://doi.org/10.11896/jsjcx.220400256>

[基于社交网络图节点度的神经网络个性化传播算法研究](#)

Study on Degree of Node Based Personalized Propagation of Neural Predictions for Social Networks

计算机科学, 2023, 50(4): 16-21. <https://doi.org/10.11896/jsjcx.220300274>

[细粒度语义知识图谱增强的中文OOV词嵌入学习](#)

Fine-grained Semantic Knowledge Graph Enhanced Chinese OOV Word Embedding Learning

计算机科学, 2023, 50(3): 72-82. <https://doi.org/10.11896/jsjcx.220700249>

[SS-GCN:情感增强和句法增强的方面级情感分析模型](#)

SS-GCN: Aspect-based Sentiment Analysis Model with Affective Enhancement and Syntactic Enhancement

计算机科学, 2023, 50(3): 3-11. <https://doi.org/10.11896/jsjcx.220700238>

融合语义和句法图神经网络的实体关系联合抽取

衡红军 苗菁

中国民航大学计算机科学与技术学院 天津 300300

(henghjcauc@163.com)

摘要 实体关系抽取任务是信息抽取的核心任务,它对于有效地从爆炸性增长的数据中提取出关键性的信息有着不可替代的作用,也是构建大规模知识图谱的基础任务,因此研究实体关系抽取对各种自然语言处理任务具有重要意义。尽管现有的基于深度学习方法的实体关系抽取已经有了很成熟的理论和较好的性能,但依然还存在着误差累积、实体冗余、交互缺失、三元组重叠等问题。语义信息和句法信息对自然语言处理任务都具有重要作用,为了充分利用这些信息以解决上述提到的问题,提出了一种融合语义和句法图神经网络的二元标记实体关系联合抽取模型 FSSRel(Fusion of Semantic and Syntactic Graph Convolutional Networks Binary Tagging Framework for Relation triple extraction)。该模型分为三个阶段进行:第一阶段,对三元组主体的开始结束位置进行预测标记;第二阶段,分别通过语义图神经网络和句法图神经网络提取语义特征和句法特征,并将其融合进编码向量;第三阶段,对语句的每种关系的客体位置进行预测标记,完成最终三元组的提取。实验结果表明,在 NYT 数据集和 WebNLG 数据集上,该模型的 F1 值较基线模型分别提升了 2.5% 和 1.6%,并且在拥有重叠三元组和多三元组等问题的复杂数据上也有良好的表现。

关键词: 实体关系联合抽取;语义信息;句法依存分析;图卷积神经网络

中图法分类号 TP391

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

HENG Hongjun and MIAO Jing

College of Computer Science and Technology, Civil Aviation University of China, Tianjin 300300, China

Abstract Entity and relation extraction task is the core task of information extraction. It plays an irreplaceable role in effectively extracting key information from explosive growth data, and is also the basic task of building a large-scale knowledge graph. Therefore, the research on entity relationship extraction task is of great significance for various natural language processing (NLP) tasks. Although the existing entity and relation extraction based on deep learning method has a very mature theory and good performance, there are still some problems, such as error accumulation, entity redundancy, lack of interaction, entity and relation overlap. Semantic information and syntactic information play an important role in NLP tasks. In order to make full use of them to solve the above problems, a fusion of semantic and syntactic graph convolutional networks binary tagging framework for relation triple extraction(FSSRel) is proposed. The model is divided into three stages. In the first stage, the start and end positions of the triple body are predicted. In the second stage, semantic features and syntactic features are extracted by semantic graph neural network and syntactic graph neural network respectively, and fused into the coding vector. In the third stage, the object position of each relation of the statement is predicted and marked to complete the extraction of the final triple. Experimental results show that the F1 value of the model increases by 2.5% and 1.6% respectively compared with the baseline model on the NYT dataset and the WebNLG dataset, and it also performs well on complex data with overlapping triples and multiple triples.

Keywords Joint entity and relation extraction, Semantic information, Syntactic dependency analysis, Graph convolution neural network

到稿日期:2022-07-05 返修日期:2022-12-06

基金项目:国家自然科学基金(U1333109)

This work was supported by the National Natural Science Foundation of China(U1333109).

通信作者:苗菁(602890196@qq.com)

1 引言

实体关系抽取任务的目的是从非结构化的文本中抽取实体并对实体之间的关系进行预测分类,使得最终能够提取出(主体,关系,客体)形式的三元组,从而能够自动获取到文本中的基础信息,为构建知识图谱提供基本的结构化数据,避免使用昂贵的人工标注语料。实体关系抽取任务在现代化社会中为处理纷杂繁多的数据提供了不可或缺的重要助力,这种必要性也使得该任务成为了文本摘要、语义网标注、搜索引擎、智能问答系统等自然语言处理任务的基础工作。实体关系抽取任务的工作流程如下:针对语句“Elliot was born in Dallas”,可以抽取主体“Elliot”和客体“Dallas”,并预测它们的关系是“birthplace”,最终提取三元组(Elliot, birthplace, Dallas)作为输出结果。

现在常用的基于深度学习的实体关系抽取方法主要分为两种:流水线(Pipeline)方法和联合抽取(Joint)方法。流水线方法的主要流程是将命名实体识别(Named Entity Recognition)和关系抽取(Relation Extraction)两个子任务分开进行,即先抽取实体,再对实体之间的关系进行预测。这种方法虽然能够自由组合两个子任务的模型,更加灵活,但同时也使得这两个任务之间缺少了必要的联系,由此产生了以下缺陷:1)误差累积,关系抽取阶段的训练过程无法与实体识别阶段的训练进行参数交互,导致实体识别时的误差会直接传播到关系抽取模块中,并且无法通过训练纠正,这种误差的不断累积,会直接影响关系抽取的准确率;2)实体冗余,实体识别任务没有经过关系的约束就抽取大量的实体,但其中的很多实体并不包含关系抽取阶段所要抽取的关系类型,并不符合当前任务的抽取目标,这部分多余的实体会给整个抽取任务带来大量冗余,从而增大抽取难度,提高错误率,增大计算量;3)交互缺失,流水线方法将本来紧密关联的两个任务完全分离,无法获取它们之间的内在联系和依赖关系,使得它们之间的信息无法交互^[1]。

针对流水线方法存在的以上种种问题,学者们逐渐开始研究加强两个子任务之间的联系。由此,联合抽取方法被提出,并在近几年得到不断发展。实体关系联合抽取的方法能够进一步利用实体识别和关系抽取两个子任务之间的潜在信息,相比流水线方法往往能够取得更好的效果。

而实体关系联合抽取的方法同样也面临着许多挑战,比如重叠三元组的问题。现实中,文本普遍拥有不止一个三元组,这些三元组之间互相重叠,造成抽取任务困难。若模型处理重叠三元组问题的能力有限,那么该模型就会无法适应许多数据集,导致严重的局限性和抽取误差。根据三元组重叠问题的类型可以将数据分为3类:普通类(Normal)、实体对重叠类(Entity Pair Overlap, EPO)和单一实体重叠类(Single Entity Overlap, SEO)。如图1所示,Normal类表示训练语句不包含重叠情况;EPO类表示同一实体对拥有多种关系;SEO类表示语句中实体与其他多个实体都有关系。现有的研究工作中,处理重叠三元组问题的能力是评价一个模型的重要指标。

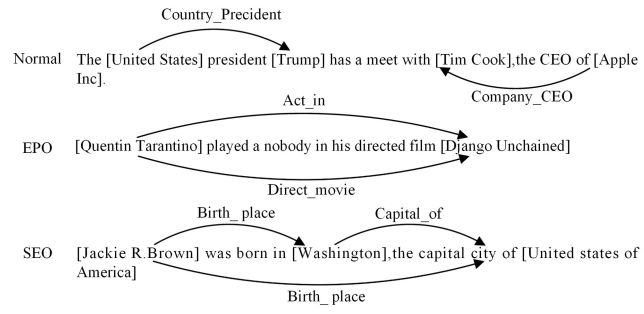


图1 重叠三元组类型实例

Fig. 1 Triple overlap type example

要有效地缓解以上问题,除了模型本身要具有处理这些问题的能力外,还要进一步挖掘文本中的潜在信息,增强模型的理解能力,使其能够在不同的语境中准确分辨三元组问题。对于自然语言处理任务,如何同时兼顾到语义信息和句法信息,对其充分地挖掘并使用是其中不可或缺的关键问题。而在文本中,不同的单词所包含的语义信息和句法信息的重要程度也有所不同。比如,名词具有更重要的语义信息,而介词则往往包含更多的句法信息^[2]。并且,即使相同词性的不同单词,其所拥有的语义特征和句法特征通常也截然不同。因此,准确合理地提取并利用语义信息和句法信息是提升实体关系抽取任务准确性的关键一环,即如何针对每个单词,考虑其包含的语义和句法特征,对这两种特征实现合理的融合,从而最终实现语义和句法的互补,丰富文本的表示信息。

图结构数据是由节点和边组成的数据类型,它能够灵活地表示各个节点之间的关系。因此对于自然语言任务中的语料,使用图神经网络结构,能够不局限于前后位置的依赖关系,更精确获取各个位置单词之间的特征信息^[3]。另一方面,注意力机制的引入使得图神经网络能够区分不同边的重要程度,从而增强表达能力,更加重视影响大的节点,忽略影响小的节点。因此,当通过语义分析和句法分析获取到单词之间的联系时,就可以利用注意力机制图卷积神经网络充分挖掘出潜在的语义信息和句法信息,进而增强模型的表达能力,使模型能够适应更复杂的语境。

基于对以上问题的分析,本文旨在研究通过图神经网络获取到语义特征和句法特征,具体如下:利用自注意力机制获取单词间的语义权重,并融合先验的关系信息特征,构建语义注意力机制图卷积神经网络得到语义特征;利用句法依存关系和依存类型信息获取单词间的句法权重,构建句法注意力机制图卷积神经网络得到句法特征;将两种特征向量进行融合,并融入编码向量中,从而挖掘到更多的文本表示信息。

结合上文的研究,同时受 Wei 等^[4]的基础模型启发,本文提出一种将语义的注意力机制图卷积神经网络(Attentive Graph Convolutional Networks)^[5]和句法的注意力机制图神经网络的输出特征进行融合的二元标记实体关系联合抽取模型 FSSRel。本文的贡献主要有以下4个方面:1)构建语义图神经网络获取语义特征,从而充分利用语义信息;2)构建句法图神经网络获取句法特征,从而充分利用句法信息;3)对两种特征信息进行融合,实现信息互补,并将其融合进入特征编码当中;4)在 NYT 数据集和 WebNLG 数据集

上的实验结果表明, FSSRel 模型的 F1 值较基线模型有明显提升。

2 相关工作

实体关系联合抽取主要可以分为参数共享的联合抽取和联合解码的联合抽取两种方法。基于参数共享的联合抽取方法主要是在实体识别和关系抽取两个子任务的训练过程中共享训练参数,从而使原本分离的两个任务产生联系。基于联合解码的联合抽取方法是通过提出新标注的序列模型,对三元组整体进行建模,从而使模型能够同时抽取实体和关系,进一步加强内部联系。

2016年, Miwa等^[6]首次将神经网络的方法用于对实体和关系进行联合抽取的任务中。然而在他们提出的模型中,两个子任务仅共享了 BiLSTM 编码层的参数,因此严格意义上并不能称为联合抽取。Eberts等^[7]在输出端通过片段排列来对实体进行分类,然后再进行关系分类,虽然缓解了片段排列的高复杂度问题,但关系分类仍有实体冗余问题。Zeng等^[8]设计了一种以 Copy 机制为基础的序列到序列学习的模型,编码器将语句转换为固定长度的语义向量,然后解码器读入此向量并直接生成三元组。但该模型只能抽取单一单词的实体,对多单词实体无能为力。Bekoulis等^[9]将关系抽取任务当作一个多头选择问题,从而能够判断每个实体与文中其他实体之间存在的关系。此外,其不再将关系抽取任务当作一个每种关系互斥的多分类任务,而是看作每个关系独立的多个二分类任务,从而判断每一对实体是否可能存在多个关系,一定程度上缓解了重叠关系问题,并且还加入了对抗网络的收益。Yu等^[10]为了增强实体边界位置之间的关联,提出了一个分层的边界标记器,从而可以将起始点和结束点的位置标记出来。Niu等^[11]使用两阶段的图卷积神经网络,分别通过迭代多个特征编码器和利用节点注意力机制来捕获单词的多粒度表示和区域特征以及注意力相关矩阵,从而获取实体、关系和三元组之间的相互作用关系。Yan等^[12]提出了一种分割滤波器网络,通过设立一个共享分区以达到正确地建模

两个子任务双向交互的目的。Zheng等^[13]设计了一种联合抽取的标记方案,使模型能够从语句中直接提取三元组,但这个模型也有设计缺陷,它无法处理具有重叠三元组问题的数据。Wang等^[14]提出了一种新的握手标记策略,分别标记实体头到实体尾、主体头到客体头、主体尾到客体尾,将联合抽取任务分解为序列标记子任务,以解决训练和预测阶段暴露偏差的问题。Sui等^[15]使用 transformer^[16]作为解码器,将多个相同的 transformer 堆叠,使用多头自注意力机制模拟各三元组之间的关系,通过多头互注意力融合句子信息并给句子做注意力表示。

以上工作尽管已经取得了不错的成果,推动了实体关系联合抽取领域的进步和发展,但均未将重点放在如何充分挖掘语义信息和句法信息上,也没有关注如何针对不同情境将两种信息进行合理融合。在自然语言处理任务中,若无法获取到准确有效的表示信息来增强模型的理解能力,那么模型就难以分辨繁杂的实体位置和关系类别,难以处理复杂文本中的冗余重叠信息,导致无法突破瓶颈以进一步提升性能。对于这些问题的处理,也是本文的研究重点和 FSSRel 模型的优势所在。

3 模型方法

FSSRel 模型的整体预测过程主要可以分为 3 个阶段,其总体结构如图 2 所示。首先,将输入语句进行 BERT 编码,融合上下文信息,对三元组主体的位置进行标记预测,即主体标记阶段;然后,将每种关系视为节点,与文本的单词节点连接构成图卷积神经网络输出融合关系信息的单词向量,进而将自注意力驱动的图卷积神经网络(A-GCN)输出作为语义特征,同时将句法依存关系的注意力机制图卷积神经网络(A-GCN)输出作为句法特征,对这两种特征进行信息融合并取主体部分,作为融合特征向量,该阶段为融合语义特征和句法特征阶段;最后,对输入语句序列的每种关系的客体位置进行标注,即关系预测和客体标记阶段。下文将按顺序对这 3 个阶段进行介绍。

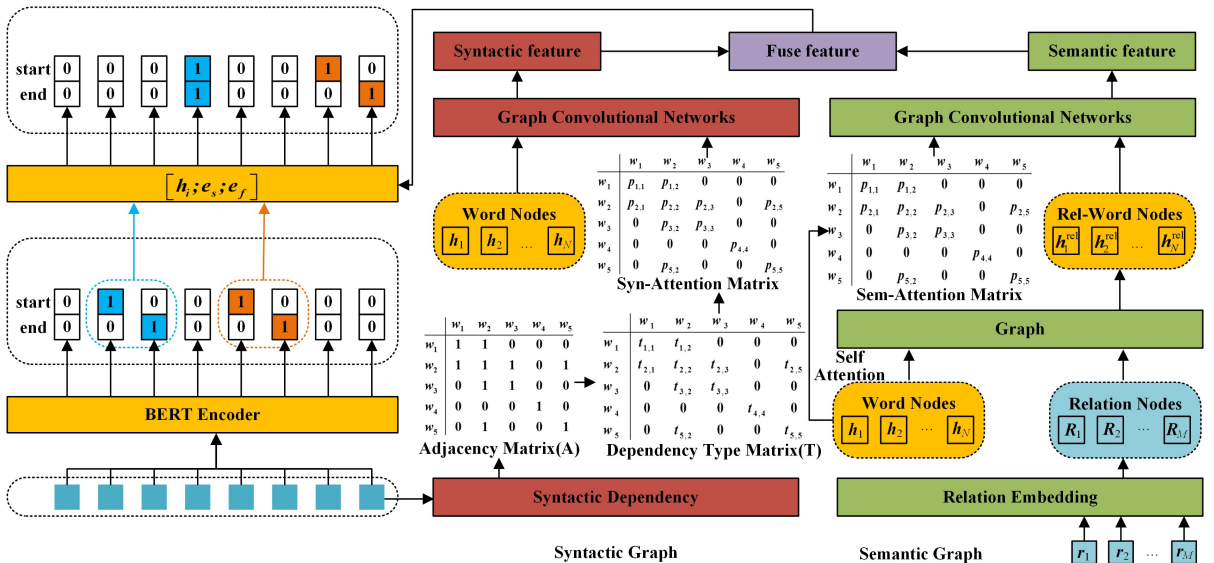


图 2 主模型结构

Fig. 2 Main model structure

3.1 主体标记

3.1.1 BERT 词编码

BERT^[17] 是于 2018 年提出的一种预训练模型,相较于一般的编码方式,其能够生成深层的双向语言特征,获取到上下文信息,在自然语言处理任务中有着优异的表现。对于给定的训练语句 $\mathbf{X} = \{w_1, w_2, \dots, w_N\}$, 用 BERT 预训练模型对该语句进行编码,得到向量表示,如式(1)所示:

$$[h_1, h_2, \dots, h_N] = B[w_1, w_2, \dots, w_N] \quad (1)$$

其中, w_i 是 BERT 中用 WordPiece 将每个单词分割标记得到的细粒度子单词, N 是语句长度, B 是 BERT 预训练模型。将 w_i 进行 BERT 编码后得到模型的输入向量 h_i 。

3.1.2 主体标记

主体标记阶段的目标是对 BERT 编码后的输入语句序列直接进行解码,从而对主体的起始和结束位置分别进行二元标记,即为训练语句中每个位置分配一个二元标记(1/0), 1 表示是主体的起始或结束位置, 0 则表示不是,以此识别出语句中所有可能的实体。具体如式(2)、式(3)所示:

$$p_i^{\text{start}_s} = \text{Sigmoid}(W_{\text{start}_s} h_i + b_{\text{start}_s}) \quad (2)$$

$$p_i^{\text{end}_s} = \text{Sigmoid}(W_{\text{end}_s} h_i + b_{\text{end}_s}) \quad (3)$$

其中, $p_i^{\text{start}_s}$ 和 $p_i^{\text{end}_s}$ 分别表示语句序列中第 i 个词是主体起始位置和结束位置的概率。模型设置一个阈值,若大于这个阈值则预测标签置为 1,小于该阈值则预测标签置为 0。

主体标记模块优化以下似然函数来识别语句 \mathbf{X} 中的主体 s ,如式(4)所示:

$$p_{\theta_i}(s|\mathbf{X}) = \prod_{r \in \{\text{start}_s, \text{end}_s\}} \prod_{i=1}^N (p_i^r)^{I\{y_i^r=1\}} (1-p_i^r)^{I\{y_i^r=0\}} \quad (4)$$

其中, N 是输入语句的长度; $y_i^{\text{start}_s}$ 和 $y_i^{\text{end}_s}$ 分别表示输入语句序列中第 i 个词是主体起始位置和结束位置的二元标记值; $I\{z\}$ 表示判断函数,当 z 表示的事件为真, $I\{z\}$ 的值为 1,否则为 0。

3.1.3 实体抽取

对输入训练语句序列的全部单词标记完成后,需要提取出被标记的主体,供在下文的两个阶段中使用。本文通过“就近原则”对主体 s 进行实体抽取,即对每个实体起始位置 $y_i^{\text{start}_s}$ 被标记为 1 的单词,不考虑该单词位置之前的单词,而是直接向后寻找语句序列中离该词最近的实体结束位置 $y_i^{\text{end}_s}$ 被标记为 1 的单词,这两个单词之间的跨度部分就是要识别的主体 s 。由于本文模型实际上是对语句所有位置进行标记预测,并求出损失值来更新梯度,因此在经过多轮的训练拟合后,实体的起始和结束位置被正确识别出来的准确率明显提升,从而能够完整、准确地抽取出语句中的全部实体。这种标记方式在多实体语句中能发挥更大的作用,在最终抽取客体时利用的也是这个原则。如图 3 所示,在这句话中根据第一个起始位置标签为 1 的单词“Jack”向后寻找离它最近的结束位置标签为 1 的单词“Lee”,可以得到实体“Jack Lee”。同理,也可以得到实体“New York”。

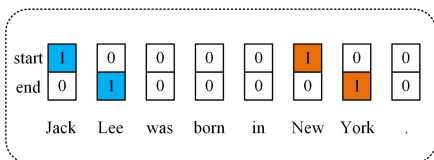


图 3 实体抽取过程

Fig. 3 Entity extraction process

3.2 融合语义特征和句法特征

不同于上一阶段中只是从文本中抽取主体,对关系的预测抽取过程显然更加复杂,要考虑的信息也更多,因此要从文本中获取到更多有用的语义特征和句法特征,需要降低被冗余实体和关系影响的误差,提高准确率。

利用自注意力机制能够捕获单词节点之间的语义信息,提取出重要的内部语义特征,加强模型的语义理解能力。通过自注意力获取单词节点之间权重的具体公式如式(5)所示:

$$p_{i,j}^{\text{sem}} = \text{Softmax}(W_q h_i \cdot W_k h_j) \quad (5)$$

从而能够得到一个语义注意力权重矩阵(Sem-Attention Matrix)。随着多层图神经网络的迭代,语义注意力权重 $p_{i,j}^{\text{sem}}$ 也在不断更新。

为了加强关系信息对抽取过程的约束,模型将关系类型作为先验知识引入,加强关系信息的约束,使模型能够更清晰地理解不同情境下的语义信息,减小上一阶段提取的冗余实体的影响,拉近单词与关系之间的联系,从而能够更有针对性地进行实体关系抽取,提升对关系预测的准确程度。

将单词和所有的预设关系视为两种类型的节点,将每种关系类型进行 one-hot 编码后映射为高维向量,如式(6)所示:

$$[R_1, R_2, \dots, R_M] = W_r E([r_1, r_2, \dots, r_M]) + b_r \quad (6)$$

其中, M 是预设关系的数量, r_i 表示 one-hot 编码后的关系类型, R_i 表示映射后的关系高维向量,将其表示为关系节点向量。

将单词和关系两种类型节点之间进行全连接,如图 4 所示,连接线的颜色深浅代表关系的重要程度。

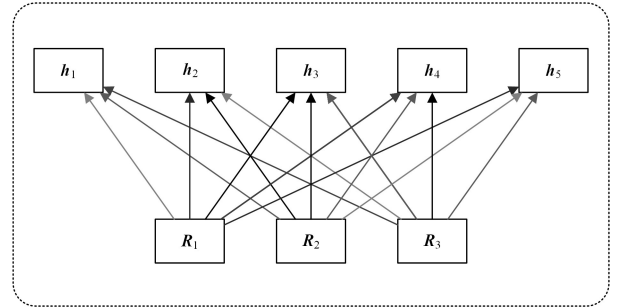


图 4 单词和关系节点连接图结构

Fig. 4 Structure of word and relation node connection

在第 l 层图卷积网络中,则针对第 $l-1$ 层神经网络输出单词节点 h_i , 对所有的关系节点计算注意力权重,更新节点表示,具体公式如式(7)、式(8)所示:

$$p_{i,j}^{\text{rel}} = \frac{\exp(W_h h_i \cdot W_R R_j)}{\exp(\sum_{j=1}^M W_h h_i \cdot W_R R_j)} \quad (7)$$

$$h_i^{\text{rel}} = h_i + \sum_{j=1}^M p_{i,j}^{\text{rel}} W_{\text{rel}} R_j \quad (8)$$

其中, W_h 和 W_R 将 h_i 和 R_j 映射为相同维度的向量; $p_{i,j}^{\text{rel}}$ 是 h_i 和 R_j 两节点之间的注意力权重; h_i^{rel} 是更新后的单词节点表示,该节点向量融合了针对不同关系类型的重要程度信息。通过融合了关系信息的新节点向量 h_i^{rel} 和语义注意力 $p_{i,j}^{\text{sem}}$ 就能够获取融合语义信息的输出向量:

$$h_i^{\text{sem}} = \text{Sigmoid}(\sum_{j=1}^N p_{i,j}^{\text{sem}} (W_{\text{sem}} \cdot h_j^{\text{rel}} + b^{\text{sem}})) \quad (9)$$

整体语义注意力机制图卷积神经网络结构如图5所示。

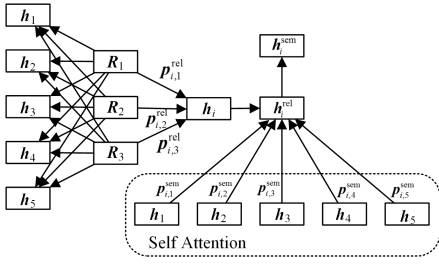


图5 语义图神经网络结构

Fig. 5 Semantic GCN structure

3.2.1 通过句法注意力机制图卷积神经网络获取句法信息

仅使用语义信息可能会忽视许多单词之间潜在的结构信息,因此需要通过句法依存分析技术,对文本中单词之间的句法关系进行解析,获得它们之间的依存结构信息。同时,不同的连接词和不同的句法依存关系对于实体关系抽取任务的重要程度也有所不同,因此只是简单的运用生成的句法依存树而不对连接的重要程度作区分,会降低模型的表达能力。为了更充分地利用句法特征,本文通过注意力机制来区分不同连接之间的重要程度,即使用句法依存关系驱动的注意力机制图卷积神经网络(A-GCN)获取更加准确的句法信息,结构图如图6所示。

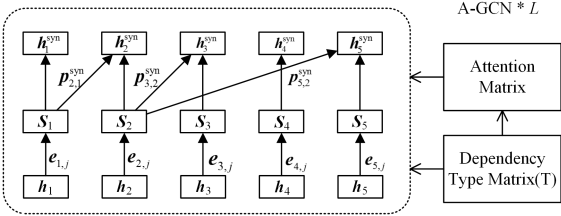


图6 句法图神经网络结构

Fig. 6 Syntactic GCN structure

首先对数据使用 Standard CoreNLP Toolkits(SCT)^[18] 进行处理可以获取句法依存树,处理示例结果如图7所示。

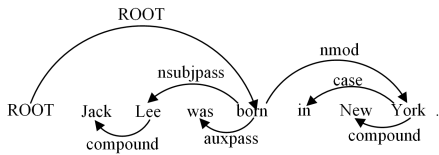


图7 句法依存树

Fig. 7 Syntactic dependency tree

针对长度为 N 的输入语句 $\mathbf{X} = \{w_1, w_2, \dots, w_N\}$, 根据它的句法依存树可以构建一个邻接矩阵(Adjacency Matrix) $\mathbf{A} = (a_{i,j})_{N \times N}$, 如果在依存树中两个词 w_i 和 w_j 之间具有句法依存关系,则将 $a_{i,j}$ 值置为 1, 否则为 0。

不同的句法依存关系所包含的信息也有所不同。为了充分利用这些信息,在 \mathbf{A} 矩阵的基础上构建了一个邻接关系矩阵(Dependency Type Matrix) $\mathbf{T} = (t_{i,j})_{N \times N}$, $t_{i,j}$ 表示经过 one-hot 编码后词之间的依存关系,当两个词之间没有依存关系时, $t_{i,j}$ 为 0。然后将每种依存关系 $t_{i,j}$ 编码成为高维向量得到新的向量表示 $e_{i,j}$ 。在第 l 层图卷积网络中,将第 $l-1$ 层图神经网络输出 h_i 和 $e_{i,j}$ 拼接得到中间向量 s_i , 如式(10)所示:

$$s_i = [h_i; e_{i,j}] \quad (10)$$

从而能够计算第 l 层图神经网络中 x_i 和 x_j 的得分权重 $p_{i,j}^{\text{syn}}$, 如式(11)所示:

$$p_{i,j}^{\text{syn}} = \frac{a_{i,j} \cdot \exp(s_i \cdot s_j)}{\sum_{j=1}^N a_{i,j} \cdot \exp(s_i \cdot s_j)} \quad (11)$$

最终能够得到一个句法注意力权重矩阵(Syn-Attention Matrix)。 $p_{i,j}^{\text{syn}}$ 代表句法依存图中不同连接的重要程度权重,它不只是将连接简单地置为(0/1)二元标记,因此包含了更丰富的句法信息。句法图神经网络中矩阵的变化如图8所示。

	w_1	w_2	w_3	w_4	w_5		w_1	w_2	w_3	w_4	w_5		w_1	w_2	w_3	w_4	w_5
w_1	1	1	0	0	0	w_1	$t_{1,1}$	$t_{1,2}$	0	0	0	w_1	$p_{1,1}$	$p_{1,2}$	0	0	0
w_2	1	1	1	0	1	w_2	$t_{2,1}$	$t_{2,2}$	$t_{2,3}$	0	$t_{2,5}$	w_2	$p_{2,1}$	$p_{2,2}$	$p_{2,3}$	0	$p_{2,5}$
w_3	0	1	1	0	0	w_3	0	$t_{3,2}$	$t_{3,3}$	0	0	w_3	0	$p_{3,2}$	$p_{3,3}$	0	0
w_4	0	0	0	1	0	w_4	0	0	0	$t_{4,4}$	0	w_4	0	0	0	$p_{4,4}$	0
w_5	0	1	0	0	1	w_5	0	$t_{5,2}$	0	0	$t_{5,5}$	w_5	0	$p_{5,2}$	0	0	$p_{5,5}$

Adjacency Matrix(A) Dependency Type Matrix(T) Syn-Attention Matrix

图8 矩阵变化

Fig. 8 Matrix change

由于图神经网络的输出 h_i 不断更新,因此注意力权重矩阵也在不断地更新。根据权重,可以得到融合了句法信息的新输出向量,如式(12)所示:

$$h_i^{\text{sem}} = \text{Sigmoid}\left(\sum_{j=1}^N p_{i,j}^{\text{syn}} (\mathbf{W}_{\text{syn}} \cdot \tilde{h}_j + b')\right) \quad (12)$$

其中, \tilde{h}_j 是 h_j 融合 $e_{i,j}$ 来加强依存关系信息获得的输出,如式(13)所示:

$$\tilde{h}_j = h_j + \mathbf{W}_T e_{i,j} \quad (13)$$

其中, \mathbf{W}_T 将 $e_{i,j}$ 映射为与 h_j 相同的维度。

3.2.2 融合语义特征和句法特征

在文本进行 BERT 编码时,可以从句首得到 [CLS] 向量 e_{cls} , 其常常被运用在句子级自然语言处理任务中,包含语句的整体信息。将语义特征和句法特征中的每个单词向量分别与 e_{cls} 计算线性相似度,进而进行归一化得到权重,该权重能够体现该单词在整句话中语义信息和句法信息的相对重要程度,从而可以对单词的语义特征和句法特征进行融合。若模型中图卷积神经网络有 L 层,则具体公式如式(14)、式(15)所示:

$$\alpha_i^{\text{sem}}, \alpha_i^{\text{syn}} = \text{Softmax}(\{e_{\text{cls}} \mathbf{W} h_i^{L-\text{sem}}, e_{\text{cls}} \mathbf{W} h_i^{L-\text{syn}}\}) \quad (14)$$

$$h_i^{\text{fuse}} = \alpha_i^{\text{sem}} h_i^{L-\text{sem}} + \alpha_i^{\text{syn}} h_i^{L-\text{syn}} \quad (15)$$

其中, $h_i^{L-\text{sem}}$ 和 $h_i^{L-\text{syn}}$ 表示经过 L 层语义图神经网络和句法图神经网络后的输出向量; α_i^{sem} 和 α_i^{syn} 分别是语义特征和句法特征的重要程度权重,根据这两个权重可以将语义特征和句法特征进行融合从而得到新的输出向量 h_i^{fuse} 。

3.2.3 融合所有特征

为了使下文中抽取主体 s 引导的三元组中客体和关系的过程更加精确,要结合主体 s 的信息。

首先将原始输入向量的主体 s 最大池化,获取到主体特征向量 e_s , 如式(16)所示:

$$e_s = \text{MaxPooling}(\{h_i | w_i \in s\}) \quad (16)$$

同时结合主体 s 的位置信息,将融合特征向量的主体跨度内所有单词最大池化,得到该主体的语义和句法的融合特征 e_f , 如式(17)所示:

$$e_f = \text{MaxPooling}(\{h_i^{\text{fuse}} | w_i \in s\}) \quad (17)$$

将这些特征全部融合进原始编码向量中,如式(18)所示:

$$\mathbf{h}_i^* = \text{Tanh}(\mathbf{W}[\mathbf{h}_i; \mathbf{e}_i; \mathbf{e}_r] + \mathbf{b}) \quad (18)$$

其中, $;$ 指矩阵拼接。

3.3 关系预测和客体标记

在上一模块中获得了新的输入后,要对语句序列进行客体的标记。

对客体的预测过程与对主体的预测过程大致相同,不同的是,针对预定义的关系集合 \mathbf{R} 中的每种关系 r ,要结合前一模块的预测主体 s ,组成 (s, r) 主体关系对,根据不同的关系对客体 o 进行标记。语句中有多个主体,则预测过程会根据主体迭代多次。具体如式(19)、式(20)所示:

$$\mathbf{p}_i^{\text{start}_o} = \text{sigmoid}(\mathbf{W}_{\text{start}_o} \mathbf{h}_i^* + \mathbf{b}_{\text{start}_o}) \quad (19)$$

$$\mathbf{p}_i^{\text{end}_o} = \text{sigmoid}(\mathbf{W}_{\text{end}_o} \mathbf{h}_i^* + \mathbf{b}_{\text{end}_o}) \quad (20)$$

其中, $\mathbf{p}_i^{\text{start}_o}$ 和 $\mathbf{p}_i^{\text{end}_o}$ 分别表示语句序列中第 i 个词是三元组中客体的起始位置和结束位置的概率。

对于关系 r 的客体标记模块,优化以下似然函数来识别给定的输入语句 \mathbf{X} 和主体 s 的客体 o 的跨度,如式(21)所示:

$$\mathbf{p}_{\theta_o}(\mathbf{o} | \mathbf{s}, \mathbf{r}, \mathbf{X}) = \prod_{i \in (\text{start}_o, \text{end}_o)} \prod_{i=1}^N (\mathbf{p}_i^{\text{start}_o})^{I(y_i^{\text{start}_o}=1)} (1 - \mathbf{p}_i^{\text{end}_o})^{I(y_i^{\text{end}_o}=0)} \quad (21)$$

其中, N 是输入语句的长度, $y_i^{\text{start}_o}$ 和 $y_i^{\text{end}_o}$ 分别是输入语句序列中第 i 个词是客体起始位置和结束位置的二元标记。

根据文献[4]中的工作,本文设计的目标函数具体如式(22)所示:

$$\begin{aligned} & \prod_{j=1}^{|\mathbf{D}|} \left[\prod_{(s, r, o) \in \mathbf{T}_j} \mathbf{p}((s, r, o) | \mathbf{X}_j) \right] \\ &= \prod_{j=1}^{|\mathbf{D}|} \left[\prod_{s \in \mathbf{T}_j} \mathbf{p}_{\theta_s}(s | \mathbf{X}_j) \prod_{(r, o) \in \mathbf{T}_j | s} \mathbf{p}_{\theta_o}((r, o) | s, \mathbf{X}_j) \right] \\ &= \prod_{j=1}^{|\mathbf{D}|} \left[\prod_{s \in \mathbf{T}_j} \mathbf{p}_{\theta_s}(s | \mathbf{X}_j) \prod_{r \in \mathbf{T}_j | s} \mathbf{p}_{\theta_o}(\mathbf{o} | s, r, \mathbf{X}_j) \prod_{r \in \mathbf{R} \setminus \mathbf{T}_j | s} \mathbf{p}_{\theta_o}(\mathbf{o}_0 | s, r, \mathbf{X}_j) \right] \end{aligned} \quad (22)$$

其中, $|\mathbf{D}|$ 表示整个训练集; \mathbf{T}_j 是训练语句 \mathbf{X}_j 中的三元组集合, $\mathbf{T}_j | s$ 表示 \mathbf{T}_j 中主体为 s 的三元组集合; $(r, o) \in \mathbf{T}_j | s$ 表示该集合中的一个 (r, o) 对; $\mathbf{R} \setminus \mathbf{T}_j | s$ 表示除 \mathbf{T}_j 中主体为 s 的三元组的关系外,其余所有关系的集合,这些关系在语句 \mathbf{X}_j 中没有相应的客体,用 \mathbf{o}_0 表示。 \mathbf{p}_{θ_s} 和 \mathbf{p}_{θ_o} 由式(4)和式(21)计算得到。

4 数据和实验

4.1 实验数据

本文使用 NYT^[19]和 WebNLG^[20]两个数据集对模型的有效性进行评估。NYT 数据集是通过远程监督方法获得的纽约时报的标注语料,拥有 118 万条数据,有 24 种预定义的关系。WebNLG 数据集是为自然语言生成任务(NLG)而提出的,有 246 种预定义的关系。

Zeng 等^[8]对这两个数据集进行了过滤处理,下文对比实验中使用的也是该数据集。处理后的数据集中 NYT 训练集中有 56 195 条语句,验证集中有 5 000 条语句,测试集中有 5 000 条语句;WebNLG 训练集中有 5 019 条语句,验证集中有 500 条语句,测试集中有 703 条语句。

两个数据集将语句划分为普通类(Normal)、实体对重叠类(EPO)和单一实体重叠类(SEO)3类数据后的分布情况如表1所列。

表1 数据集统计

Table 1 Dataset statistics

Databases	NYT		WebNLG	
	Train	Test	Train	Test
Normal	37 013	3 266	1 596	246
EPO	9 782	978	227	26
SEO	14 735	1 297	3 406	457
ALL	56 195	5 000	5 019	703

4.2 参数设置

图卷积网络的层数设置为 2 层,在训练期间使用 Adam 优化器,其余参数设置如表 2 所列。

表2 参数设置

Table 2 Parameter setting

Parameter	NYT	WebNLG
Batch size	2	4
Learning rate	0.000 005	0.000 009
Max epoch	100	120
Threshold	0.5	0.7

4.3 实验分析

4.3.1 整体结果分析

本文选择了包括 NovelTagging 模型^[13]、CopyRE 模型^[8]、GraphRel 模型^[21]、SPointer 模型^[22]、CopyR_{RL} 模型^[23]、Relation-Aware 模型^[24]、CasRel 模型^[4]在内的多个模型进行对比实验。对比模型简介如下:

- 1) NovelTagging 模型设计了一种能够从语句中直接提取三元组的方法,但它无法处理存在重叠问题的语句。
- 2) CopyRE 模型提出了一种以 Copy 机制为基础的序列到序列的学习模型,能够处理一定的重叠问题。
- 3) GraphRel 模型针对联合抽取设计了一种图神经网络模型,并针对三元组重叠问题进行了研究。
- 4) SPointer 模型提出了一种利用双指针模块的端到端的联合抽取模型。
- 5) CopyR_{RL} 模型引入强化学习,通过多关系任务学习和 copy 机制来解决关系重叠问题,提取所有关系事实。
- 6) Relation-Aware 模型基于改进的图卷积神经网络,先识别实体跨度,再利用关系感知的注意力机制获取实体间关系。
- 7) CasRel 模型设计了一种级联二元标记模型,将关系视为主体映射到客体的函数上,在三元组重叠问题上效果显著。

FSSRel 模型在与其它对比模型在 NYT 数据集和 WebNLG 数据集上的性能对比结果如表 3 所列,其中对比模型的实验结果均来自原始论文实验结果。

表3 结果对比

Table 3 Comparison of results

(单位:%)

Method	NYT			WebNLG		
	P	R	F1	P	R	F1
NovelTagging	62.4	31.7	42.0	52.5	19.3	28.3
CopyR _{OneDecoder}	59.4	53.1	56.0	32.2	28.9	30.5
CopyR _{MultiDecoder}	61.0	56.6	58.7	37.7	36.4	37.1
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
SPointer	72.8	69.0	70.9	38.7	37.5	38.1
CopyR _{RL}	77.9	67.2	72.1	63.3	59.9	61.6
Relation-Aware	83.2	64.7	72.8	66.4	62.7	64.5
CasRel	89.7	89.5	89.6	93.4	90.1	91.8
Ours	92.3	92.0	92.1	94.5	92.3	93.4

从整体可以看出,FSSRel 模型在经过多层语义注意力机制的图卷积神经网络和句法注意力机制的图卷积神经网络后,挖掘到了比其他模型更深层次的语义特征和句法特征,并在不同单词上区分这两种特征的重要程度再进行融合,使得各个单词之间的联系更加紧密且丰富,模型表达能力得到增强。在 NYT 数据集和 WebNLG 数据集上,FSSRel 模型对比基线模型 CasRel 的 F1 值分别提升了 2.5% 和 1.6%,模型提取实体关系的能力提升明显。

4.3.2 消融实验

为研究 FSSRel 模型获取到的语义信息和句法信息对实验结果的影响和有效性,在 WebNLG 数据集上进行了消融实验,结果如表 4 所列。其中 -sem 表示模型去掉语义注意力机制的图卷积神经网络获取语义特征后的实验结果;-syn 表示模型去掉句法注意力机制的图卷积神经网络获取句法特征后的实验结果;-rel 表示语义图神经网络中不添加关系信息的实验结果;-fuse 表示不对语义特征和句法特征进行融合,将其仅进行简单拼接的实验结果。可以看出,不同模块的缺失均会对最终结果产生一定的影响,证明了各个模块的有效性。

表 4 消融实验

Table 4 Ablation experimental results (单位:%)

Method	P	R	F1
Ours	94.5	92.3	93.4
-sem	93.9	91.9	92.8
-syn	94.1	91.8	92.9
-rel	93.8	92.3	93.0
-fuse	93.8	92.4	93.1

4.3.3 不同三元组重叠类型语句结果分析

将测试集的数据分为普通类(Normal)、实体对重叠类(EPO)和单一实体重叠类(SEO)3类数据后,分别对性能效果进行测试,结果如图 9—图 11 所示。

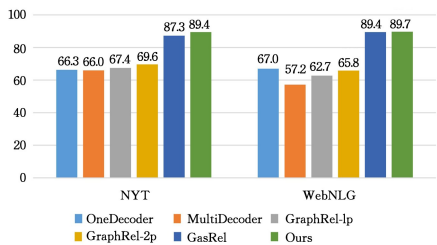


图 9 Normal 类数据 F1(%) 值

Fig. 9 F1 value of Normal data

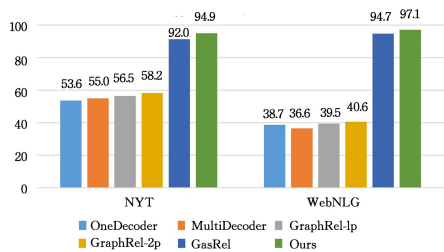


图 10 EPO 类数据 F1(%) 值

Fig. 10 F1 value of EPO data

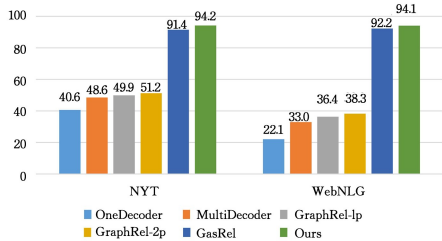


图 11 SEO 类数据 F1(%) 值

Fig. 11 F1 value of SEO data

从实验的对比结果可以看出,FSSRel 模型在这 3 种不同类型的数据上均能发挥出更强的抽取能力,并且在 EPO 和 SEO 这种包含复杂重叠三元组问题的数据上效果提升更明显,证明了 FSSRel 模型在获取潜在的语义信息和句法信息后,增强了对这些问题的处理能力和辨别能力,拥有了更强的普适性,更能够适应现实的复杂语境。

4.3.4 不同三元组个数语句结果分析

为了进一步验证 FSSRel 模型应对复杂数据的处理能力,本文设置 N 为测试语句中存在三元组的个数,将 NYT 和 WebNLG 两个数据集的测试集分别分为 $N=1, N=2, N=3, N=4, N \geq 5$ 这 5 类数据进行测试对比,具体结果如表 5 所列。从表中的结果可以看出,在包含不同三元组个数的数据中,包含多个三元组的数据 F1 值普遍比仅包含单一三元组的数据增长幅度更大,并且在 $N \geq 5$ 的数据上,两数据集的增长分别达到了 7.9% 和 3.3%,是各类数据中增长最多的,由此可以进一步证明模型的泛化能力。

表 5 不同数量三元组数据的 F1 值

Table 5 F1 value of different numbers of triplet data

(单位:%)

Method	NYT					WebNLG				
	$N=1$	$N=2$	$N=3$	$N=4$	$N \geq 5$	$N=1$	$N=2$	$N=3$	$N=4$	$N \geq 5$
CopyR _{OneDecoder}	66.6	52.6	49.7	48.7	20.3	65.2	33.0	22.2	14.2	13.2
CopyR _{MultiDecoder}	67.1	58.6	52.0	53.6	30.0	59.2	42.5	31.7	24.2	30.0
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
CasRel	88.2	90.3	91.9	94.2	83.7	89.3	90.8	94.2	92.4	90.9
Ours	89.3	95.3	93.9	96.5	91.6	89.9	91.5	95.6	95.2	94.2

结束语 本文通过语义注意力机制图卷积神经网络和句法注意力机制图卷积神经网络,分别获取到语义特征和句法特征,并对其进行融合,从而增强了模型的表达能力,提高了预测的准确性,并在 NYT 和 WebNLG 两个数据集上得到了验证。同时更多的实验也证明了 FSSRel 模型能够很好地处理存在重叠三元组和多三元组问题的语料。但本文模型依然存在一些不足的地方,比如,本文使用两个图卷积神经网络导致模型的时间复杂度和空间复杂度较大,将来会继续研究优化模型的工作流程以减小复杂度。对两种特征寻找更合理的融合方法同样具有研究空间和研究价值,未来也将围绕这些方面继续进行实验。

参考文献

[1] LI D M,ZHANG Y,LI D Y,et al. Review of entity relation extraction methods [J]. Journal of Computer Research and Deve-

- lopment, 2020, 57(7):1424-1448.
- [2] SHEN Y, MA X, TANG Y, et al. A Trigger-Sense Memory Flow Framework for Joint Entity and Relation Extraction[C]// Proceedings of the Web Conference 2021. 2021:1704-1715.
- [3] WU L, CHEN Y, SHEN K, et al. Graph neural networks for natural language processing: A survey[J]. arXiv: 2106. 06090, 2021.
- [4] WEI Z, SU J, WANG Y, et al. A novel cascade binary tagging framework for relational triple extraction [J]. arXiv: 1909. 03227, 2019.
- [5] TIAN Y, CHEN G, SONG Y, et al. Dependency-driven Relation Extraction with Attentive Graph Convolutional Networks[C]// Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers). 2021:4458-4471.
- [6] MIWA M, BANSAL M. End-to-End Relation Extraction using LSTMs on Sequences and Tree Structures[C]// Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). 2016.
- [7] EBERTS M, ULGES A. Span-based joint entity and relation extraction with transformer pre-training[J]. arXiv: 1909. 07755, 2019.
- [8] 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.
- [9] BEKOULIS G, DELEU J, DEMEESTER T, et al. Joint entity recognition and relation extraction as a multi-head selection problem[J]. Expert Systems with Applications, 2018, 114: 34-45.
- [10] 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.
- [11] NIU W, CHEN Q, ZHANG W, et al. GCN2-NAA: Two-stage Graph Convolutional Networks with Node-Aware Attention for Joint Entity and Relation Extraction[C]// ICMLC 2021: 2021 13th International Conference on Machine Learning and Computing. 2021.
- [12] YAN Z, ZHANG C, FU J, et al. A Partition Filter Network for Joint Entity and Relation Extraction[C]// Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing. 2021:185-197.
- [13] ZHENG S, F WANG, BAO H, et al. Joint Extraction of Entities and Relations Based on a Novel Tagging Scheme[C]// Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). 2017.
- [14] WANG J, LU W. Two are better than one: Joint entity and relation extraction with table-sequence encoders[J]. arXiv: 2010. 03851, 2020.
- [15] SUI D, CHEN Y, LIU K, et al. Joint Entity and Relation Extraction with Set Prediction Networks[J]. arXiv: 2011. 01675, 2020.
- [16] VASWANI A, SHAZEER N, PARMAR N, et al. Attention Is All You Need[J]. arXiv: 1706. 03762, 2017.
- [17] DEVLIN J, CHANG M W, LEE K, et al. Bert: Pre-training of deep bidirectional transformers for language understanding[J]. arXiv: 1810. 04805, 2018.
- [18] MANNING C, SURDEANU M, BAUER J, et al. The stanford core NLP natural language processing toolkit[C]// Proceedings of the 52th Annual Meeting of Association for Computational Linguistics: System Demonstrations. Stroudsburg: ACL Press, 2014:55-60.
- [19] RIEDEL S, YAO L, MCCALLUM A. Modeling relations and their mentions without labeled text[C]// Joint European Conference on Machine Learning and Knowledge Discovery in Databases. Berlin: Springer, 2010:148-163.
- [20] 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.
- [21] 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.
- [22] BAI C, PAN L, LUO S, et al. Joint extraction of entities and relations by a novel end-to-end model with a double-pointer module[J]. Neurocomputing, 2020, 377:325-333.
- [23] ZENG X, HE S, ZENG D, et al. Learning the extraction order of multiple relational facts in a sentence with reinforcement learning[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). 2019:367-377.
- [24] HONG Y, LIU Y, YANG S, et al. Improving graph convolutional networks based on relation-aware attention for end-to-end relation extraction[J]. IEEE Access, 2020, 8:51315-51323.



HENG Hongjun, born in 1968, Ph.D., associate professor, is a member of China Computer Federation. His main research interests include intelligent information processing, natural language and knowledge graph.



MIAO Jing, born in 1998, postgraduate. His main research interests include natural language processing and information extraction.