文章编号：1003-0077(2021)06-0074-11

BSLRel：基于二元序列标注的级联关系三元组抽取模型

张龙辉[1]，尹淑娟[1]，任飞亮[1]，苏剑林[2]，明瑞成[1]，白宇佳[1]

(1. 东北大学 计算机科学与工程学院，辽宁 沈阳 110819；
2. 深圳追一科技有限公司，广东 深圳 518057)

摘 要：关系三元组抽取是构建大规模知识图谱的基础，近年来受到学术界和工业界的广泛关注。为了提高模型对重叠关系三元组和多槽值关系三元组的抽取能力，该文提出了一个基于神经网络的端到端的关系三元组抽取模型 BSLRel。其主要特点是将关系三元组抽取任务转化为级联的二元序列标注任务，并使用多信息融合结构 Conditional Layer Normalization 进行信息融合。实验结果显示，BSLRel 模型对重叠关系三元组和多槽值关系三元组具有较强的抽取能力。基于 BSLRel 模型，该团队参加了“2020 语言与智能技术竞赛”中的关系三元组抽取任务，并取得了第五名的成绩。

关键词：BSLRel 模型；重叠关系三元组抽取；多槽值关系三元组抽取

中图分类号：TP391 **文献标识码**：A

BSLRel: A Binary Sequence Labeling Based Cascading Relation Triple Extraction Model

ZHANG Longhui[1], YIN Shujuan[1], REN Feiliang[1], SU Jianlin[2], MING Ruicheng[1], BAI Yujia[1]

(1. School of Computer Science and Engineering, Northeastern University, Shenyang, Liaoning 110819, China;
2. Shenzhen Zhuiyi Technology Co. Ltd, Shenzhen, Guangdong 518057, China)

Abstract: Extracting relational triples is a basic task for large-scale knowledge graph construction. In order to improve the ability of extracting overlapped relation triples and multi-slot relation triples, this paper proposes BSLRel, an end-to-end relation triple extraction model based on neural network. Specifically, BSLRel model converts the relation triplet extraction task into a cascade binary sequence labeling task, which consists of a new multiple information fusion structure “Conditional Layer Normalization” to integrate information. With BSLRel, we participate in the “Relation Extraction” task organized by “the 2020 Language and Intelligence Challenge” and achieve Top 5 among all competitive models.

Keywords: BSLRel; overlapped relation triple; multi-slot relation triple

0 引言

将非结构化或半结构化的自然语言文本转化成结构化内容是信息抽取的目的。关系三元组抽取作为信息抽取的子任务，其目的是从自然语言文本(常常以句子为输入单位)中抽取两个实体及实体之间的关系。该任务是构建大规模知识图谱的关键，被广泛运用在信息检索、问答系统等相关任务中[1]。

在关系三元组抽取任务中，关系事实大多以形如“(subject, predicate, object)”或“(head, relation, tail)”的三元组形式表示。其中，subject(或 head)称为头实体，object(或 tail)称为尾实体，predicate(或 relation)表示头、尾实体之间的语义关系[2]。比如三元组“(孔乙己，作者，鲁迅)”可表示“孔乙己”和“鲁迅”之间存在“作者”关系这一事实。

当前关系三元组抽取研究方法主要分为两类：管道(pipeline)抽取方法和联合抽取方法。早期的

收稿日期：2020-08-03 **定稿日期**：2020-11-02

基金项目：国家重点研发计划项目(2018YFC0830701)；国家自然科学基金(61572120)；中央高校基本科研业务专项资金(N181602013)

方法主要是管道抽取方法，即先对输入句子中的实体进行识别；之后，再为识别出的实体分配合适的关系类型。这类管道抽取方法存在以下两点不足。

（1）**实体冗余**：由于先对抽取的所有实体进行两两配对，然后再进行关系分类，大量错误的候选实体对所带来的冗余信息，会提升错误率、增加计算复杂度。

（2）**任务间的交互缺失**：忽略了实体识别和关系抽取这两个任务之间的内在联系和依赖关系，无法充分利用输入信息。

因此，研究者提出了联合抽取的方法，即同时从输入句子中抽取实体及实体间的对应关系。这类方法可以充分利用实体识别以及关系抽取之间的交互信息，因而可以有效缓解管道方法的不足。然而，在关系三元组抽取任务中，同一句子中的多个三元组共享相同的头实体、关系或尾实体的情况是大量存在的。如三元组“（孔乙己，作者，鲁迅）”和三元组“（朝花夕拾，作者，鲁迅）”就共享关系和尾实体。研究者将此类具有某些信息共享的三元组称为重叠三元组。

目前基于联合抽取的方法并不能有效地处理重叠三元组抽取问题。因为重叠三元组的重叠情况多种多样，头尾实体或关系，任意一部分或两部分之间都会有重叠，甚至一个实体可能同时是一个三元组的头实体和另一个三元组的尾实体。而在联合抽取模型中常常使用序列标注的思路进行实体识别，即对每个单词赋予一个唯一的标注，比如 B/I/O 标注，用来区分该单词是否是某一实体的开始、结尾或者不是实体。因为无法将一个单词既标注为一个头实体同时又标注为一个尾实体，所以此架构很难处理重叠三元组的抽取问题。虽然管道抽取模型会遍历所有抽取出来的实体对来解决重叠三元组的抽取问题，但会引入大量错误的实体对而导致抽取性能大幅下降。

此外，“2020 语言与智能技术竞赛（2020 Language and Intelligence Challenge）”基于 DuIE 2.0 数据集[3]，举办了难度更大的关系三元组抽取任务。第一，该竞赛任务将关系三元组的 object 部分进行了复杂化扩展：首先，将某些关系对应的 object 由一个实体组成（本文将 object 中只有一个实体的三元组称为“单槽值三元组”）拓展到可以由多个实体组成，其中每个实体对应一种槽位（本文将 object 中有多个实体的三元组称为“多槽值三元组”）。然后，为每种关系类型都设置了 schema，以达到对三元组中的头、尾实体类别进行约束的目的。表 1 显示了该竞赛任务中不同类型关系三元组的样例，其中“@value”类型的槽位一定会在三元组中出现，而其他类型的槽位信息如果在句子中没有体现，则可以不出现在三元组中。

表 1　2020 语言与智能技术竞赛三元组关系抽取任务样例（选自竞赛提供的数据集 DuIE 2.0[3]）

文本	《正大综艺》的主持人王雪纯既是 87 版《红楼梦》中晴雯的配音者，也是电影《外出》中的孙艺珍的配音者，还为韩剧《汉江怪物》做过配音。
“配音”关系的 Schema	subject_type：娱乐人物；predicate：配音；object_type：{@value：人物，inWork：影视作品}
“主持人”关系的 Schema	subject_type：电视综艺；predicate：主持人；object_type：{@value：人物}
单槽值三元组	subject：正大综艺；predicate：主持人；object：{@value：王雪纯}
多槽值三元组 1	subject：王雪纯；predicate：配音；object：{@value：晴雯，inWork：红楼梦}
多槽值三元组 2	subject：王雪纯；predicate：配音；object：{@value：孙艺珍，inWork：外出}

从本质上讲，“2020 语言与智能技术竞赛”中的单槽值三元组为没有重叠问题的普通三元组，多槽值三元组为一类头实体和关系重叠的重叠三元组。然而，由于 schema 约束的存在，在该竞赛任务中的多槽值三元组抽取和传统的重叠三元组抽取并不完全相同。显然，同重叠三元组一样，多槽值三元组广泛存在于真实文本中。对这些特殊类型的三元组进行有效抽取是提升关系三元组抽取任务性能的关键。

为有效地解决关系三元组抽取任务中的重叠三元组抽取问题和多槽值三元组抽取问题，本文提出了 BSLRel 模型（binary sequence labeling based cascading relation triple extraction model），一个基于神经网络的端到端的关系三元组抽取模型。该模型的主要特点是将关系三元组抽取任务转化为二元序列标注任务，并通过级联结构解决重叠三元组和

多槽值三元组的抽取问题。此外,本文还发现当前基于神经网络结构的模型在进行信息融合时并不能体现待融合信息之间的方向性。而方向对于准确理解输入信息的语义具有重要意义。为此,本文提出了一种名为 conditional layer normalization (CLN)的多信息融合结构,并将其应用在 BSLRel 模型中,实验证明,此方法可以取得比其他信息融合方法更好的实验结果。

基于 BSLRel 模型,我们参加了"2020 语言与智能技术竞赛"中的关系三元组抽取任务,最终 F_1 值为 0.808 3、取得了总竞赛排名第五的成绩。

综上,本论文工作主要贡献为以下两点:

(1) 提出 BSLRel 模型,将关系三元组抽取任务转化为了二元序列标注任务,并采用级联结构为重叠三元组和多槽值三元组的抽取问题提出了一种解决方案。

(2) 提出可以体现信息之间方向性的多信息融合结构 conditional layer normalization(CLN),并通过实验证明了此结构的有效性。

1 相关工作

当前关系三元组抽取方法按抽取过程可以分为管道抽取方法和联合抽取方法。

1.1 管道抽取方法

管道抽取方法是指通过流水线方式进行关系三元组抽取的一类方法。该类方法的主要特点是将关系三元组抽取任务转化为命名实体识别与关系分类这两个相对独立的子任务,并以流水线方式依次完成这两个子任务得到最终的关系三元组结果。即针对给定输入文本,该类方法首先识别文本中的命名实体;之后枚举各种候选实体对,并为每一个候选对进行关系预测;最后,根据预测结果得到输入文本最终的关系三元组结果。

早期的命名实体识别主要是采用基于规则的方法。规则通常由领域专家和语言学者耗费大量时间和精力来制定,且只适用于简单场景,领域的迁移性较差。后来产生了基于机器学习的命名实体识别方法,主要有隐马尔科夫模型、条件随机场模型等,这些模型都将命名实体识别任务作为一种序列标注任务,即将每个字或者词都标记为一个标签类别。这类方法比基于规则的方法有了很大改进,大幅减少了人工成本。但这些方法都需要研究者们人工提取有效的语法特征。基于神经网络的方法可大大降低模型对人工语法特征的依赖,因而一经提出便引起了研究者们的极大注意,有力地促进了命名实体识别任务的研究。Dong 等人[4]提出基于 BiLSTM 利用 CRF 来进行中文命名实体识别,捕获时间序列的特性,提高识别准确率。Ling 等人[5]在 BiLSTM 中引入注意力机制以关注长文本的局部特征。Yan 等人[6]通过改进的 Transformer 模型进一步提升了命名实体识别的性能。

对于关系分类研究,目前的方法按照使用的主要技术可以分为三类:基于特征向量的方法、基于核函数的方法和基于神经网络的方法。基于特征向量的方法通过从包含特定实体对的句子中提取语义特征,构造特征向量,然后使用支持向量机[7]、最大熵[8]等模型进行关系分类。基于核函数的方法充分利用句子的特定组织形式[9],通过设计核函数计算句子之间的相似度,并根据相似度进行分类。基于神经网络的关系分类方法主要使用各种深度神经网络方法进行关系分类。Zhou 等人[10]提出 Att-BiLSTM 模型,使用双向 LSTM 网络来获取句子中每个单词的隐状态输出,使用注意力机制来抽取单词级别特征,通过特定的向量来计算每个单词对句子表示贡献的权重,以得到句子的最终向量表示用以分类。Wu 等人[11]在输入文本中加入位置表示信息,利用预训练模型 BERT 取了得优异性能。

从关系三元组抽取过程来看,管道抽取方法会遍历所有候选实体对,因而理论上可以很好地处理重叠三元组的识别问题。但由于这类方法高度依赖命名实体识别的结果,容易造成误差传递。另外,由于整个过程将命名实体识别和关系分类当作两个相对独立的任务,没有充分利用二者之间的相关性,也进一步降低了此类方法在关系三元组抽取任务上的性能。

1.2 联合抽取方法

为了有效解决管道抽取方法中的误差传递问题,研究者们提出了联合抽取的方法。

早期的联合模型采用了基于人工特征的结构化学习[12-14],很大程度上依赖于人工制作的各类特征。神经网络的发展缓解了人工制作特征的问题,但其中一些方法仍然依赖于 NLP 工具(如 POS 标记器、依赖解析器)[15]。2016 年 Miwa 和 Bansal 等人[16]提出了一种基于递归神经网络的模型,将双向树结构的 LSTM-RNNs 叠加在双向顺序的 LSTM-

RNNs 上，获得了单词序列和依赖树子结构信息，此模型在实体识别和关系提取任务上共享参数，加强了两个任务之间的相关性。该方法虽然将两个任务整合到了同一个模型当中，但依然是两个分离的过程。

通过序列标注的方法来进行关系三元组抽取是另一种受研究者们关注较多的联合抽取方法。2017 年 Zheng 等[17]根据关系种类对实体类型标签进行了相应的扩展，并将关系三元组抽取问题转化为完全的基于序列标注的命名实体识别问题。以句子“中国的首都是北京”为例，“北京”在传统的实体识别任务中对应的两个标注分别是“B、I”，但在该模型中，对应的标注为“B—首都—尾实体、I—首都—尾实体”。2019 年，Yuan 等[18]首先生成每个关系下特定的句子表示，然后对每个句子表示进行基于序列标注的命名实体识别，抽取出该关系下的头尾实体。

上述两种方法都不依赖于任何 NLP 工具，可以有效解决管道抽取方法中的误差传递问题，但都无法应对同种关系下存在多个三元组的情况。

Liu 等[19]提出了另外一种基于序列标注的关系三元组抽取方法。首先采用基于序列标注的实体识别模块识别出所有的实体对，再通过一个三维的关系分类模块识别出任意实体对之间具有的关系。这种方法与管道抽取方法类似，都可以涵盖任何一种三元组重叠情况，而且可以通过共享两个模块之间的编码层使模型整体性能大大提升。但缺点是其使用的三维关系分类模块过于庞大，使得模型收敛困难，需要较大的内存空间。

从当前国内外的最新研究成果来看，基于联合抽取的方法，尤其是基于序列标注的联合抽取方法，已成为关系三元组抽取的主流方法，并在实际应用中取得了当前的最佳实验结果。

2　模型介绍

本文将提出的关系三元组抽取模型称为 BSLRel(binary sequence labeling based cascading relation triple extraction model)，一个基于神经网络的端到端的关系三元组抽取模型。其主要思路是将关系三元组抽取任务转化为二元序列标注任务，并通过级联结构解决重叠三元组和多槽值三元组的抽取问题。在模型中，首先在输入句子中进行三元组的头实体识别，之后再对每个头实体识别其对应的关系以及尾实体。BSLRel 模型将关系和 object 中的“@value”槽位进行共同抽取，避免了过长的级联结构，因而，BSLRel 模型不仅可以解决重叠三元组和多槽值三元组的抽取问题，还可以通过缩短级联长度的方式，缓解错误累积问题。并且，由于模型在整个过程中将三元组抽取问题分解为两个相对独立的子问题，所以从整体模型层面来讲，并不需要对重叠三元组抽取以及多槽值三元组抽取进行特殊的处理，其流程与普通的单槽值三元组抽取完全一致。这大大降低了模型的复杂度，使得本文提出的模型具有良好的关系三元组抽取性能以及很好的系统鲁棒性。

本文提出的 BSLRel 模型的整体结构如图 1 所示，可以看出，该模型从功能上可以分为句子表示层、单槽值抽取层和多槽值抽取层三个模块。接下来将对这些模块分别进行详细介绍。

2.1　句子表示层

在句子表示层，本文使用预训练的 BERT 模型为输入句子中的每个 token 生成一个向量表示。得到一个针对输入句子所有 token 的向量序列，本文把这个向量序列记为 H_1。如图 1 所示，H_1 将在随后的多个处理模块中被共享使用。

2.2　单槽值抽取层

单槽值抽取层的目的是抽取形式为“(subject，predicate，@value)”的单槽值三元组，抽取流程为先抽取所有的 subject 实体，再抽取每个 subject 对应的 predicate 和 object 结构体的@value 槽位。

(1) **subject 的抽取**。如图 1 所示，这一模块首先对 H_1 采用 Layer Normalization(LN)方法进行归一化，并把归一化的结果记为 H_2，这一过程如式(1)、式(2)所示。

$$\ln(x)=\frac{x-E[x]}{\sqrt{\mathrm{Var}[x]+\varepsilon}}*\gamma+\beta \tag{1}$$

$$H_2=\ln(H_1) \tag{2}$$

其中，x 为待进行归一化的数据，$E[x]$ 和 $\mathrm{Var}[x]$ 分别为 x 的均值和方差，ε 代表一个大于 0 的极小常数，目的是为了防止公式中分母部分趋近于 0 时所导致的计算错误，因此，$\sqrt{\mathrm{Var}[x]+\varepsilon}$ 即为 x 的标准偏差，γ 和 β 是可训练的参数向量，形状与 x 相同。$*$ 代表元素级别的乘法。

LN 方法是神经网络架构常用的数据归一化方法之一，目的是把数据分布映射到一个确定的区间，加速收敛，缓解梯度爆炸或消失的问题，读者可以参

考 BERT 原始论文[20]获得关于 BERT 和 LN 的更多内容，在此不再赘述。

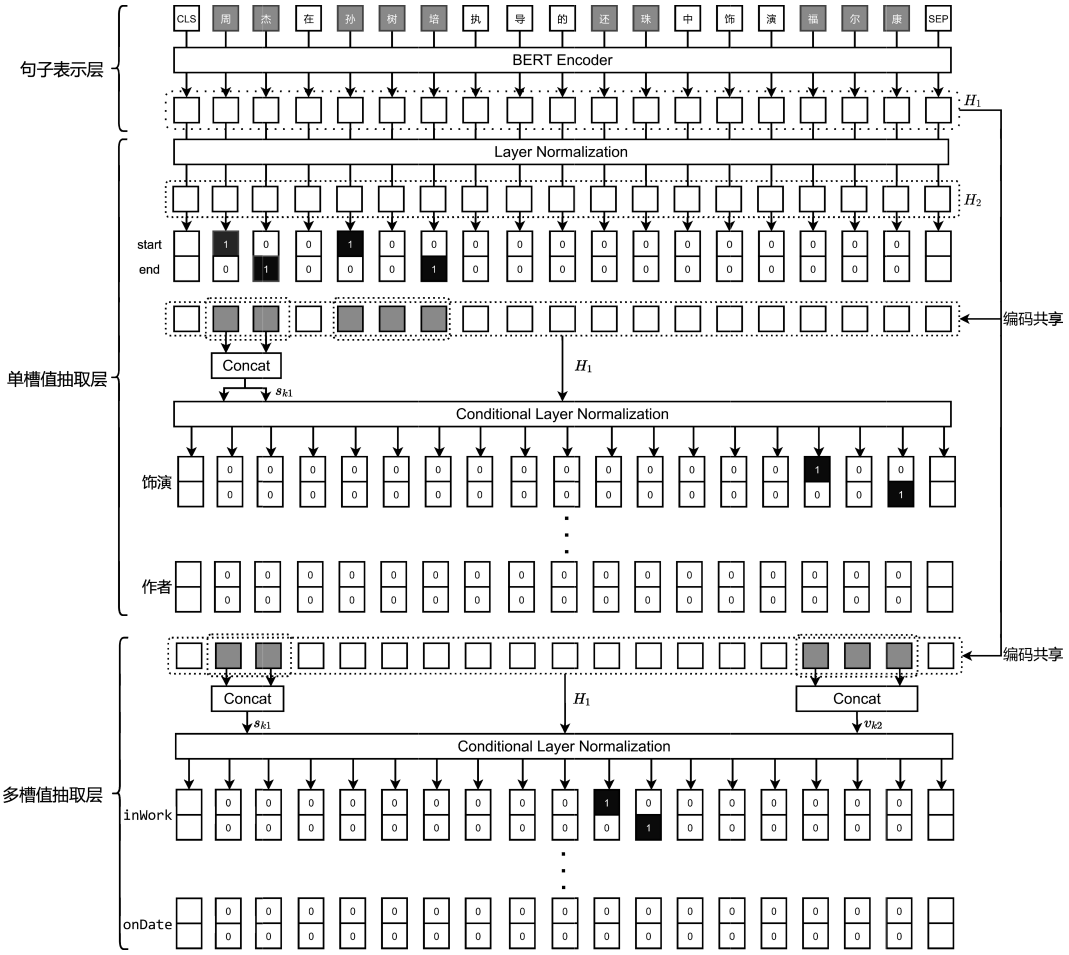

图 1 BSLRel 模型的整体架构

之后，模型将基于H_2为输入句子的每个 token 生成两个概率，分别代表对应的 token 为某个 subject 实体的开头或者结尾的可能性。这里本文将输入句子中第 i 个 token 对应的两个概率分别记作$p_{\text{start}}^{s,i}$和$p_{\text{end}}^{s,i}$。当对应的概率值大于预定义的实体边界识别阈值 θ 时，将对应的 token 标注为 1，否则标注为 0。具体的计算过程如式(3)、式(4)所示。

$$p_{\text{start}}^{s,i}=\sigma(\boldsymbol{W}_{\text{start}}^{s}\boldsymbol{H}_2^i+b_{\text{start}}^{s}) \tag{3}$$

$$p_{\text{end}}^{s,i}=\sigma(\boldsymbol{W}_{\text{end}}^{s}\boldsymbol{H}_2^i+b_{\text{end}}^{s}) \tag{4}$$

其中，$\boldsymbol{H}_2^i$ 为句子编码$\boldsymbol{H}_2$ 中的第 i 个 token 对应的向量，$W_{(\cdot)}^{s}$为可训练权重，$b_{(\cdot)}^{s}$为可训练偏置，σ 为 sigmoid 激活函数。本文使用 sigmoid 而非 softmax 作为激活函数是为了保证在概率序列$p_{\text{start}}^{s,i}$和$p_{\text{end}}^{s,i}$中，每个元素都有可能大于预定义的实体边界识别阈值 θ，即保证了模型可以同时识别出句子中的多个 subject 实体。

在本模块中，BSLRel 模型根据二分类的交叉熵计算关于 subject 的损失 Loss_s，其中二分类的交叉熵计算如式(5)所示，p 为预测出的概率值，q 为正确的二元标注。Loss_s 的计算过程如式(6)所示。其中，L 为句子长度，$q_{(\cdot)}^{s,i}$为句子的第 i 个 token 对应的正确标注。

$$\text{ce}(p,q)=-[q\log p+(1-q)\log(1-p)] \tag{5}$$

$$\text{Loss\_s}=-\frac{1}{2L}\sum_{i=0}^{L}\sum_{pos\in\{\text{start,end}\}}\text{ce}(p_{\text{pos}}^{s,i},q_{\text{pos}}^{s,i}) \tag{6}$$

(2) **predicate 和@value 的抽取**。本模块中，模型首先将句子编码H_1和句子中第 k_1 个 subject 的实体表示s_{k1}通过 CLN 进行信息融合，产生新的句子编码H_1^s。然后H_1^s 经过线性层和 sigmoid 激活函数，生成 $2M$ 个二元标注序列，其中，M 为关系的种类数。本文将序列中的元素记作$p_{\text{start}}^{pv,i,j}$和$p_{\text{end}}^{pv,i,j}$，分别表示在句子中第 i 个 token 是第 j 种关系的条件下，该 subject 对应的@value 槽值的开头和结尾的概率。

在本模块中，一个关键操作就是将前面处理层得到的信息进行有效融合。传统的多信息融合方

法，如多信息相加、相乘等，常常忽略信息之间的方向性，但在关系三元组中，这种方向性却是一个重要信息。以表 1 为例，在 schema 约束下，“（正大综艺，主持人，王雪纯）”是一组正确的三元组，而“（王雪纯，主持人，正大综艺）”是一组错误的三元组，是如果将“正大综艺”“主持人”“王雪纯”这三个信息按照传统的信息融合策略，直接相乘或相加，则会得到相同的融合结果。换句话说，正确的三元组和错误的三元组在分别进行信息融合之后，得到两个相同的结果，这显然是不合理的。因此，传统的信息融合方式并不能体现信息之间的方向性，使用一个能体现关系三元组各个部分之间方向性的多信息融合结构具有重要意义。

受图像处理中流行的多信息融合结构：conditional batch normalization（CBN）[21] 启发，本文提出了一种新的信息融合方法：conditional layer normalization（CLN）。CLN 和 CBN 的做法类似，都是将归一化结构中对应的偏置和权重变成关于待融合条件的函数。CLN 的具体计算如式（7）所示。

$$\mathrm{cln}(y,c_\beta,c_\gamma)=\frac{y-E[y]}{\sqrt{Var[y]+\varepsilon}} * W_1 c_\gamma + W_2 c_\beta \tag{7}$$

其中，y 为输入 CLN 结构的特征信息，c_β 和 c_γ 分别为输入的两个待融合的条件信息。

从式（7）可以看出，CLN 通过两个不同的可训练权重 W_1 和 W_2，将条件信息 c_γ 和 c_β 映射到不同的空间，以此体现记录条件的方向信息。另外，本文的实验部分表明，两个条件的传入位置交换之后，模型效果会大打折扣，这也验证了 CLN 结构对两个条件的方向信息的敏感性。

如图 1 所示，本模块利用 CLN 结构对 s_{k1} 和 H_1^s 进行信息融合，并产生新的句子编码 H_1^s。整个计算如式（8）所示。

$$\boldsymbol{H}_1^s=\mathrm{cln}(H_1,s_{k1},s_{k1}) \tag{8}$$

其中，s_{k1} 代表句子中第 k_1 个 subject 的实体编码。模型将该 subject 实体的开始和结束位置在句子编码 H_1^s 中的对应向量拼接起来作为整个实体编码。此时只有 1 个待融合条件 s_{k1}，因此，本模块将 CLN 结构的两个输入条件都设置为 s_{k1}。

$\boldsymbol{H}_1^s$ 中第 i 个 token 对应的向量 $\boldsymbol{H}_1^{s,i}$ 经过线性层和激活层，生成该 token 在第 j 种关系下，是@value 槽值的开头或者结尾的概率 $p_{\mathrm{start}}^{pv,i,j}$ 和 $p_{\mathrm{end}}^{pv,i,j}$，计算如式（9）、式（10）所示。

$$p_{\mathrm{start}}^{pv,i,j}=\sigma(\boldsymbol{W}_{\mathrm{start}}^{pv,j}\ \boldsymbol{H}_1^{s,i}+b_{\mathrm{start}}^{pv,j}) \tag{9}$$

$$p_{\mathrm{end}}^{pv,i,j}=\sigma(\boldsymbol{W}_{\mathrm{end}}^{pv,j}\ \boldsymbol{H}_1^{s,i}+b_{\mathrm{end}}^{pv,j}) \tag{10}$$

其中，$W_{(\cdot)}^{pv,j}$ 和 $b_{(\cdot)}^{pv,j}$ 为线性层中的可训练参数。

在这一层中，模型根据二分类交叉熵计算关于 predicate 和@value 的损失 Loss_pv，具体过程如式（11）所示，其中，$q_{(\cdot)}^{pv,i,j}$ 在第 j 种关系下，第 i 个 token 对应的正确标注。

$$\mathrm{Loss\_pv}=-\frac{1}{2LM}\sum_{i=0}^{L}\sum_{j=0}^{M}\sum_{\mathrm{pos}\in\{\mathrm{start,end}\}} ce(p_{\mathrm{pos}}^{pv,i,j},q_{\mathrm{pos}}^{pv,i,j}) \tag{11}$$

2.3　多槽值抽取层

多槽值抽取层的作用是依据关系的 schema 约束，对在单槽值抽取层中预测的单槽值三元组的其他槽位进行补充抽取。

在本层中，首先将句子编码 H_1 作为句子的特征信息，句中第 k_1 个 subject 的编码 s_{k1} 及其对应的第 k_2 个 object 的@value 槽位编码 v_{k2} 作为两个不同条件，输入 CLN 结构中进行信息融合，产生新的句子编码 H_1^{sv}，然后 H_1^{sv} 经过线性映射和 sigmoid 层，生成 $2(S-1)$ 个二元序列标注，其中，S 为槽位类型的种类数（因为“@value”槽位在单槽值抽取层中已经被抽取出来，所以只需再针对其余的 S−1 个槽位进行抽取即可）。本文将标注序列中的元素记作 $p_{\mathrm{start}}^{o,i,t}$ 和 $p_{\mathrm{end}}^{o,i,t}$，分别代表在第 t 种槽位类型下，句中第 i 个 token 是该三元组的槽位实体的开头或结尾的概率。

利用 CLN 结构生成句子编码 H_1^{sv} 的过程，如式（12）所示。

$$H_1^{sv}=\mathrm{cln}(\mathrm{H}_1,s_{k1},v_{k2}) \tag{12}$$

对比式（8）和式（12）可知，与单槽值抽取层不同，多槽值抽取层中的 CLN 结构的两个条件输入 c_γ 和 c_β 是不同的，式（12）中，subject 和@value 的编码分别作为 c_β 条件和 c_γ 条件传入 CLN 结构。

之后，基于信息融合后的句子编码 H_1^{sv}，模型将生成句子中第 i 个 token 是该三元组的槽位实体的开头或结尾的概率 $p_{\mathrm{start}}^{o,i,t}$ 和 $p_{\mathrm{end}}^{o,i,t}$，如式（13）、式（14）所示。

$$p_{\mathrm{start}}^{o,i,t}=\sigma(\boldsymbol{W}_{\mathrm{start}}^{o,t}\ \boldsymbol{H}_1^{sv}+b_{\mathrm{start}}^{o,t}) \tag{13}$$

$$p_{\mathrm{end}}^{o,i,t}=\sigma(\boldsymbol{W}_{\mathrm{end}}^{o,t}\ \boldsymbol{H}_1^{sv}+b_{\mathrm{end}}^{o,t}) \tag{14}$$

其中，$\boldsymbol{W}_{(\cdot)}^{o,t}$ 和 $b_{(\cdot)}^{o,t}$ 为可训练参数。

在这一层中，模型将根据二分类交叉熵计算槽位损失 Loss_o，具体过程如式（15）所示。

$$\mathrm{Loss\_o}=-\frac{1}{2L(S-1)}\sum_{i=0}^{L}\sum_{t=0}^{\mathrm{S-1}}\sum_{\mathrm{pos}\in\mathrm{start,end}} ce(p_{\mathrm{pos}}^{o,i,t},q_{\mathrm{pos}}^{o,i,t}) \tag{15}$$

其中，$q_{\mathrm{pos}}^{o,i,t}$ 为概率值 $p_{\mathrm{pos}}^{o,i,t}$ 对应的正确二元标注。

2.4 训练过程

BSLRel 模型将关系三元组的抽取任务转化为级联的二元序列标注任务，因此本文将所有子任务的平均损失作为模型的整体损失 Loss，计算过程如式(16)所示。

$$\text{Loss}=(\text{Loss\_s}+\text{Loss\_pv}+\text{Loss\_o})/3 \quad (16)$$

本文采用 Adam 算法进行梯度下降训练。在训练阶段，BSLRel 模型采用 Teacher Forcing 策略进行训练：①在 predicate 和“@value”槽位的抽取结构中(即式(8)中)，传入训练集中真实的 subject 实体s_{k1}进行条件融合；②在多槽值抽取层(即式(12)中)，传入训练集中真实的 subject 实体s_{k1}和“@value”槽位v_{k2}进行条件融合。而在预测阶段，BSLRel 模型只使用预测到的信息进行更深层次的运算。

2.5 其他数据处理操作

2.5.1 实体识别

在 BSLRel 模型中，实体信息以两个分别代表实体的开始位置和结束位置的二元标注序列存储，因此从这两个二元标注序列中解码出所有实体是不可缺少的步骤。

当句子中包含多个实体时，模型首先根据预测出的二元标注序列，选取出所有可能的实体的开始位置和结束位置。然后从前到后遍历所有开始位置，选取距离每一个开始位置最近且在其后面的结束位置进行配对，如果没有符合条件的结束位置，则不进行配对。

例如，在图 1 中的样例中，以“周”为开始位置，在“周”之后，且离“周”最近的结束位置为“杰”，因此可以匹配到实体“周杰”。

2.5.2 数据预处理

为了提高模型在不规范文本上的泛化能力，同时也为了丰富可训练样本，在模型训练之前，本文通过 3 种数据预处理操作来破坏句子的规范性：①利用结巴分词工具随机删除或增加句中的若干词语；②将多个短句合并成一个长句，或者将长句拆分成多个短句；③将句中的关系三元组随机替换为其他相同关系的头尾实体。实验证明，这种破坏句子规范性的预处理操作可以有效提升模型性能。

2.5.3 模型融合和答案后处理

模型的融合是提升模型性能的有效方法。在本次比赛中，我们把句子表示层使用的预训练模型分别设置为 NEZHA-large，RoBERTa-large 和 BERT，再使用投票法得出三元组的最终预测结果。

此外，为了提升答案的召回率，我们将 train 数据集和 dev 数据集包含的所有三元组标注构建成一个三元组数据集，对模型生成的答案按照远程监督的思想进行补充。并利用规则对一些错误的答案进行纠正或删除，以提升精确率。

3 实验及讨论

3.1 数据集介绍

本文采用 2020 语言与智能技术竞赛关系三元组抽取任务提供的 DuIE 2.0 数据集，对提出的方法进行性能验证。该数据集是当前规模最大的中文关系三元组抽取数据集，其中包含超过 21 万的中文句子及 48 个已定义好的 schema 约束。该数据集来源于微博、贴吧、百度知道等真实场景，因此包含了部分并不规范的口语化文本，对模型的泛化能力具有很高的要求。并且 DuIE 2.0 不仅包含大量的重叠三元组，还对关系三元组中的 object 实体进行了多槽值扩展(可参考前文表 1 中提供的样例)。

3.2 实验设置

在本文方法中，一些基本的超参数设置如下。输入的最大句子长度设置为 256，训练 epoch 设置为 20，batch 设置为 20，判断实体边界的识别阈值 θ 设置为 0.5。训练过程中采用 Adam 算法进行优化，初始学习率设置为 2e-5，损失函数采用二分类的交叉熵。

3.3 实验结果

在本节中，首先对 BSLRel 模型的整体性能进行评价，然后评价模型对重叠三元组以及多槽值三元组的抽取能力，之后对 2.5 节中的各种数据处理操作和模型融合操作进行消融实验，评价各个部分对模型的贡献，最后对本文提出的 CLN 方法的有效性进行验证。

需要注意的是，在本次比赛中，官方排名以及得分都是根据 DuIE 2.0 数据集中的 test 集合，而该集合目前只发布了句子集合，并没有发布正确的三元组标注结果，只能将预测的结果提交到比赛网站进行评测，但是，目前评测系统已经关闭。因此，本文

的大部分实验均在 DuIE 2.0 的 dev 集合上进行。

本文将选用 2020 语言与智能技术竞赛官方提供的一个可对多槽值三元组进行抽取的系统 InfoExtractor 2.0 作为基线系统。InfoExtractor 2.0 设计了一种结构化的标记策略来直接调优预训练语言模型 ERNIE，通过这种策略可以一次性提取多个重叠的实体对。

3.3.1　模型的整体表现

本节分别对比了 InfoExtractor 2.0、未经 2.4 节介绍的数据处理操作的 BSLRel 模型和经过 2.4 节介绍的数据处理操作的 BSLRel 模型这三个系统在 dev 集合和 test 集合上的性能，结果如表 2 所示。从中可以看出，与 InfoExtractor 2.0 相比，BSLRel 模型在不经任何数据处理和模型融合时，在 dev 集合上取得了更好的结果。具体的，BSLRel 模型在 precision 和 recall 这两个指标上相对于基线系统而言分别提升了 2.57%、2.76%。

在最终的线上结果中，本文模型经过数据处理和模型融合之后，取得了 precision 值为 84.23%、recall值为 77.7%、最终 F_1 值为 80.83%的结果，在所有参赛队伍中成绩总排名列第 5（以模型在 test 集合的 F_1 分数为最终排名依据）。

关于表 2 需要说明两点：①线上评测系统已经关闭。因此，BSLRel 单模型在 test 集合上的分数没有测试到。②将预测的 test 集合的三元组标注情况提交到线上进行评测时，评测系统会使用一些方式保证评价的公平性，比如，构建同名词库提高标注对错判断的容错率、对于特殊字符进行特殊处理、人工校验等等。因此表 2 所示的模型在 dev 集合和 test 集合之间的分数不具有可比较性。

表 2　BSLRel 和 InfoExtractor 2.0 整体对比

数据集	InfoExtractor 2.0			BSLRel			BSLRel＋数据处理＋模型融合		
	Prec./%	Rec./%	F_1/%	Prec./%	Rec./%	F_1/%	Prec./%	Rec./%	F_1/%
dev	68.03	72.50	70.20	70.60	75.26	72.87	74.30	78.29	76.24
tyest	69.23	73.20	71.15	—	—	—	84.23	77.70	80.83

3.3.2　模型在三元组类型上的消融实验

为了确定模型是否能对重叠三元组和多槽值三元组进行有效抽取，本文从 DuIE 2.0 的 dev 集合中抽取出四类句子的集合，**集合 A**：既不包含重叠三元组，也不包含多槽值三元组。**集合 B**：仅包含重叠三元组。**集合 C**：仅包含多槽值三元组。**集合 D**：两种类型的三元组都包含。我们对这四类句子集合使用 BSLRel 模型和 InfoExtractor 2.0 进行测试，结果如表 3 所示。

表 3　BSLRel 和 InfoExtractor 2.0 在不同句子集合上的评分

类别	InfoExtractor 2.0			BSLRel		
	Prec./%	Rec./%	F_1/%	Prec./%	Rec./%	F_1/%
集合 A	71.51	74.23	72.84	71.14	74.83	72.93
集合 B	69.23	73.20	71.15	71.10	75.06	73.02
集合 C	67.82	70.90	69.32	70.50	74.79	72.58
集合 D	67.14	70.48	68.76	71.91	73.01	72.45

可以看出，InfoExtractor 2.0 在四个集合上的表现波动明显，F_1 最大相差 4.08%。而 BSLRel 模型的分数则差距不大，F_1 最大分差为 0.57%。因此，相比于 InfoExtractor 2.0，BSARel 模型对于句子中包含的三元组类型更具有鲁棒性。

并且可以发现，BSLRel 模型在包含重叠三元组的集合 B 上的表现甚至要略好于没有包含重叠三元组的集合 A，这是因为 BSLRel 模型从设计之初就以更好的抽取重叠三元组为重要目标。综上，可以证明 BSLRel 模型可以对重叠三元组和多槽值三元组进行有效抽取。

同时可以发现，虽然 BSLRel 模型在 4 种句子集合上的表现差别不大，但是在涉及多槽值三元组抽取的 C 和 D 集合上的表现却略差于另外两个集合，我们认为这主要是因为多槽值的抽取结构处于整个级联结构的最后一层，因此容易受到误差传递的影响。

3.3.3　模型关于数据处理的消融实验

本节将对 BSLRel 模型中 CLN、数据增强、模型融合、答案后处理等四个模块对三元组抽取性能的贡献进行相应的评价。

我们将 BSLRel 模型中利用 CLN 结构进行信息融合的方法改为将条件向量与句子向量进行拼接实现融合的方法，并以后者作为 baseline 系统。之后，在此 baseline 系统之上依次进行 CLN 方法替

换、加入数据增强、加入模型融合、加入答案后处理等操作，最终的实验结果如表4所示。

表4 四种策略在dev集合上的表现

模型	Prec./%	Rec./%	F_1/%
baseline	69.23	74.21	71.63
＋CLN	70.62	75.26	72.87
＋数据增强	72.46	75.28	73.84
＋模型融合	73.51	77.45	75.43
＋答案后处理	**74.30**	**78.29**	**76.24**

从表4可知，CLN结构和模型融合为BSLRel模型带来的性能涨幅最大，分别为1.24%和1.56%。

3.3.4 模型关于CLN的消融实验

本节在模型中对比了以下四种信息融合结构：

- 基于拼接的信息融合方法，即将subject和@value的实体表示拼接在句子表示后面，并将该方法记作Mod_con。
- 基于相加的信息融合方法，即将subject实体表示、@value的实体表示和句子表示三者相加，并将该方法记作Mod_add。
- 基于向量输入的信息融合方法，即将subject和object的实体表示加和成为一个向量表示，然后作为单条件传入CLN结构，并将该方法记作CLN_single。
- 条件输入位置进行交换的CLN方法，即将subject实体和@value实体在CLN结构中的传入位置进行交换，并将此时的模型记作CLN_rev。

相应的实验结果如表5所示，从中可以得出以下结论。

表5 不同条件融合方式在dev集合的表现

方法	Prec./%	Rec./%	F_1/%
Mod_add	69.03	73.87	71.36
Mod_con	69.23	74.21	71.63
CLN_single	69.92	74.87	72.31
CLN_rev	70.38	75.03	72.63
BSLRel	**70.60**	**75.26**	**72.87**

（1）基于CLN结构的信息融合方法CLN_single、CLN_rev取得了明显优于传统的信息融合方法Mod_add和Mod_con的效果。这证明了本文提出的CLN方法具有更好的信息融合能力。

（2）Mod_add、CLN_single这两种忽略信息方向性的多信息融合方法，得分低于Mod_con、CLN_rev这两种可以体现信息方向性的方法。这证明了使用一个对信息方向性敏感的多信息融合结构的重要性。

（3）比较CLN_rev和BSLRel的结果可知，将subject实体和@value实体在CLN结构中的传入位置进行交换之后，模型的性能受到了一定的影响，也证明了CLN结构对于输入条件信息输入位置的敏感性。

3.4 BSLRel模型的进一步讨论

首先，对BSLRel模型目前存在的不足进行分析。

（1）为了分析BSLRel模型对于不同长度句子的三元组抽取能力，本文将DuIE 2.0的dev集合中的句子按照长度从短到长分为5类，分别对这5类进行预测，并统计分数，结果如图2所示，可以发现BSLRel模型在长文本上的抽取效果相较于短文本明显较差。我们认为主要原因在于，BSLRel模型虽然可以通过级联的二元序列标注结构对重叠三元组和多槽值三元组进行有效抽取，但同时因为这种基于二元序列标注的输出形式，导致输出矩阵过于稀疏，0/1标注的类别严重不平衡，并且句子序列越长，这种类别不平衡的问题就越严重，因此就有了上述结果。

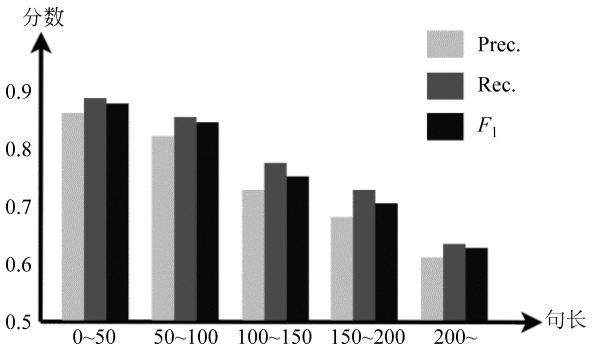

图2 BSLRel在不同句子长度上的表现

（2）对BSLRel模型在三元组中subject、predicate、@value槽位和其他其余槽位的抽取能力进行分析，结果如图3所示。从中可以发现，模型对subject的抽取能力相比于其他部分而言是最差的。我们认为，这是因为模型在对subject进行抽取时，相较于其他部分并没有接收到更多的有效信息，即模型在不明确三元组的其他信息，只对subject单独

进行抽取时，效果往往不会很好。所以，不难得到进一步的结论：虽然 BSLRel 模型通过尽量缩短级联结构的方式，来避免误差传递问题，但是由于模型对于 subject 的抽取部分处于整个级联结构的最底层，且效果最差，因此，BSLRel 模型仍然会存在一定的误差传递问题。

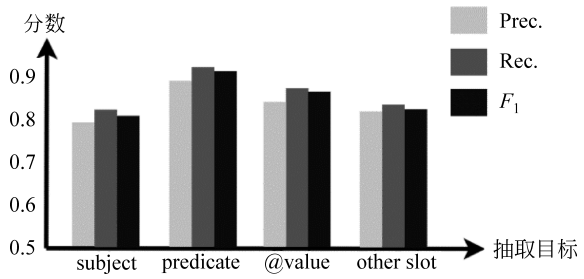

图 3　BSLRel 在不同抽取目标上的表现

然后，对本文中将重叠三元组和多槽值三元组进行区分的原因做进一步的解释。

直观来看，只需要将关系和 object 的槽位类型进行拼接，即可将多槽值三元组转化为多个传统的三元组。比如，表 1 中的单槽值三元组就可以转化成“(正大综艺，主持人_@value，王雪纯)”这种传统的三元组，多槽值三元组可以转化成两个共享头实体的传统三元组：“(王雪纯，配音_@value，晴雯)”和“(王雪纯，配音_inWork，红楼梦)”。

依据这个思路，本次比赛的一些参赛队伍将“多槽值三元组”和“重叠三元组”这两个概念进行统一，使用“先抽取多个头实体重叠的传统三元组，再将这些重叠三元组还原成所需要的多槽值三元组”的模型来进行多槽值三元组的抽取。但是这种模型的处理方式将会带来以下问题：

(1) 将关系和槽位类型进行拼接的做法，忽略了@value 槽位的必要性，会导致模型进行错误抽取。比如在表 1 所示的样例中，上述系统将抽取出错误的三元组“(王雪纯，配音_inWork，汉江怪物)”。因为“@value”类型的槽位是 object 结构体不可缺少的，并且文本中没有出现该三元组对应的“@value”类型的槽位，即并没有提及“王雪纯在《汉江怪物》中为哪个角色配音”，因此这个看似正确的三元组不应该被抽取出来。

(2) 将关系和槽位类型进行拼接的做法，将会使“@value”槽位无法和其余槽位进行正确匹配。例如，对于表 1 中的句子，上述系统将会抽取出“(王雪纯，配音_@value，晴雯)”、“(王雪纯，配音_inWork，红楼梦)”、“(王雪纯，配音_@value，孙艺珍)”和“(王雪纯，配音_inWork，外出)”这四个三元组，很明显，上述系统并不能确认两个“@value”槽位的三元组和另外两个“inWork”槽位之间的对应关系。因此，即使上述模型可以正确地抽取出这四个传统三元组，但仍将它们无法还原成表 1 所示的多槽值三元组。

基于上述两个问题的考虑，本文将多槽值三元组和重叠三元组这两个概念进行了区分。

4　结语

本文提出了 BSLRel，一种级联的关系三元组抽取模型，将关系三元组抽取任务转化为了二元序列标注任务，并针对关系三元组的实体对之间的方向性，提出了一种有效的多信息融合方法 CLN(conditional layer normalization)。

实验结果显示，BSLRel 模型可以对重叠关系三元组和多槽值关系三元组进行有效抽取。基于本文所提出的模型，我们最终在“2020 语言与智能技术竞赛”举办的关系三元组抽取任务中取得了较好的比赛成绩。

在实验中，本文也发现基于二元序列标注的 BSLRel 模型存在着类别不平衡问题和误差传递问题，这将是我们进一步改进模型的方向。

参考文献

[1] 庄传志，靳小龙，朱伟建，等.基于深度学习的关系抽取研究综述[J]. 中文信息学报，2019，33(12)：1-18.

[2] 白龙，靳小龙，席鹏弼，等. 基于远程监督的关系抽取研究综述[J]. 中文信息学报，2019，33(10)：10-17.

[3] Li S，He W，Shi Y，et al. DuIE：A large-scale Chinese dataset for information extraction[C]//Proceedings of CCF International Conference on Natural Language Processing and Chinese Computing，2019：791-800.

[4] Dong C H，Zhang J J，Zong C Q. Character-based LSTM-CRF with radical-level features for Chinese named entity recognition[C]//Proceedings of the 24th International Conference on Computer Processing of Oriental Languages，2016：239-250.

[5] Ling L，Zhihao Y，Pei Y，et al. An attention-based BiLSTM-CRF approach to document-level chemical named entity recognition[J]. Bioinformatics，2018，34(8)：1381-1388.

[6] Yan H，Deng B，Li X，et al. TENER：Adapting transformer encoder for named entity recognition[J]. arXiv preprint arXiv：1911.04474，2019.

[7] Zhao S，Grishman R. Extracting relations with inte-

grated information using kernel methods[C]//Proceedings of the 43th Annual Meeting on Association for Computational Linguistics, 2005: 419-426.

[8] Kambhatla Nanda. Combining lexical, syntactic, and semantic features with maximum entropy models for extracting relations[C]//Proceedings of the ACL 2004 or Interactive Paster and demanstration sessions, 2004: 22-25.

[9] Zelenko D, Aone C, Richardella A. Kernel methods for relation extraction[J]. Journal of Machine Learning Research, 2003, 3(3): 1083-1106.

[10] Zhou P, Shi W, Tian J, et al. Attention-based bidirectional long short-term memory networks for relation classification[C]//Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, 2016: 120-129.

[11] Wu S, He Y. Enriching pre-trained language model with entity information for relation classification [C]//Proceedings of the 29th Conference on Information and Knowledge Management, 2019: 2361-2364.

[12] Yu X, Lam W. Jointly identifying entities and extracting relations in encyclopedia text via A Graphical Model Approach[C]//Proceedings of the 23rd International Conference on Computational Linguistics: Poster, 2010: 1399-1407.

[13] 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, 2014: 402-412.

[14] Makoto Miwa, Yutaka Sasaki. Modeling joint entity and relation extraction with table representation[C]//Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, 2014: 1858-1869.

[15] 宋睿,陈鑫,洪宇,等. 基于卷积循环神经网络的关系抽取[J]. 中文信息学报, 2019, 33(10): 64-72.

[16] 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, 2016: 1105-1116.

[17] Zheng S, Wang F, Bao H, et al. Joint extraction of entities and relations based on a novel tagging scheme [C]//Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, 2017: 1227-1236.

[18] Yuan Y, Zhou X, Pan S, et al. A relation-specific attention network for joint entity and relation extraction [C]//Proceedings of the 29th International Joint Conferences on Artificial Intelligence, 2020: 4054-4060.

[19] Liu J, Chen S, Wang B, et al. Attention as relation: Learning supervised multi-head self-attention for relation extraction[C]//Proceedings of the 29th International Joint Conference on Artificial Intelligence and Seventeenth Pacific Rim International Conference on Artificial Intelligence, 2020: 3787-3793.

[20] Devlin J, Chang M W, Lee K, et al. BERT: Pre-training of deep bidirectional transformers for language understanding [C]//Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2019: 4171-4186.

[21] De Vries H, Strub F, Mary Jérémie, et al. Modulating early visual processing by language [J]. Neural Information Processing Systems,2017,11(19): 6594-6604.

张龙辉(1997—),硕士研究生,主要研究领域为自然语言处理。
E-mail: 2650603623@qq.com

任飞亮(1976—),通信作者,博士,副教授,主要研究领域为自然语言处理、知识图谱构建、智能问答。
E-mail: renfeiliang@cse.neu.edu.cn

尹淑娟(1998—),硕士研究生,主要研究领域为自然语言处理。
E-mail: 3045816640@qq.com