智能系统学报
CAAI TRANSACTIONS ON INTELLIGENT SYSTEMS

双关系预测与特征融合的实体关系抽取模型

沈健, 夏鸿斌, 刘渊

引用本文:
沈健,夏鸿斌,刘渊. 双关系预测与特征融合的实体关系抽取模型[J]. 智能系统学报, 2024, 19(2): 462–471.
SHEN Jian, XIA Hongbin, LIU Yuan. Entity relation extraction model with dual relation prediction and feature fusion[J]. *CAAI Transactions on Intelligent Systems*, 2024, 19(2): 462–471.

在线阅读 View online: https://dx.doi.org/10.11992/tis.202204047

DOI: 10.11992/tis.202204047
网络出版地址: https://link.cnki.net/urlid/23.1538.tp.20231110.1125.004

双关系预测与特征融合的实体关系抽取模型

沈健[1], 夏鸿斌[1,2], 刘渊[1,2]
(1. 江南大学 人工智能与计算机学院, 江苏 无锡 214122; 2. 江苏省媒体设计与软件技术重点实验室, 江苏 无锡 214122)

摘 要: 现有分阶段解码的实体关系抽取模型仍存在着阶段间特征融合不充分的问题, 会增大曝光偏差对抽取性能的影响。为此, 提出一种双关系预测和特征融合的实体关系抽取模型(entity relation extraction model with dual relation prediction and feature fusion, DRPFF), 该模型使用预训练的基于 Transformer 的双向编码表示模型(bidirectional encoder representation from transformers, BERT)对文本进行编码, 并设计两阶段的双关系预测结构来减少抽取过程中错误三元组的生成。在阶段间通过门控线性单元(gated linear unit, GLU)和条件层规范化(conditional layer normalization, CLN)组合的结构来更好地融合实体之间的特征。在 NYT 和 WebNLG 这 2 个公开数据集上的试验结果表明, 该模型相较于基线方法取得了更好的效果。

关键词: 实体关系抽取; 关系三元组; 预训练模型; 双关系预测; 指针网络; 特征融合; 门控线性单元; 条件层规范化
中图分类号: TP391 **文献标志码:** A **文章编号:** 1673−4785(2024)02−0462−10

中文引用格式: 沈健, 夏鸿斌, 刘渊. 双关系预测与特征融合的实体关系抽取模型 [J]. 智能系统学报, 2024, 19(2): 462–471.
英文引用格式: SHEN Jian, XIA Hongbin, LIU Yuan. Entity relation extraction model with dual relation prediction and feature fusion[J]. CAAI transactions on intelligent systems, 2024, 19(2): 462–471.

Entity relation extraction model with dual relation prediction and feature fusion

SHEN Jian[1], XIA Hongbin[1,2], LIU Yuan[1,2]
(1. School of Artificial Intelligence and Computer, Jiangnan University, Wuxi 214122, China; 2. Jiangsu Key Laboratory of Media Design and Software Technology, Wuxi 214122, China)

Abstract: The staged decoding entity relation extraction model still has an insufficient feature fusion problem between stages, which increases the impact of exposure bias on the extraction performance. Herein, we propose a new entity relation extraction model with dual relation prediction and feature fusion (DRPFF). DRPFF uses a pretrained model of bidirectional encoder representation from transformers to encode texts, and a two-stage dual relation prediction structure is developed to reduce the false triples' generation. Between stages, a structure combining gated linear units and conditional layer normalization is utilized to fuse features better between entities. Experimental findings on two public datasets, NYT and WebNLG, demonstrate that the presented method has better results than the baseline methods.

Keywords: entity relation extraction; relational triple; BERT pretrained model; dual relation prediction; pointer network; feature fusion; gated linear unit; conditional layer normalization

实体关系抽取作为信息抽取的一个子任务, 其目的是从非结构化文本中抽取出结构化的关系三元组[1]。关系三元组通常以(S, P, O)的形式表示, 其中包含了主体(subject)、客体(object)以及这 2 个实体间的语义关系(predicate)。实体关系抽取是大规模构建知识图谱的基础, 被广泛运用于知识问答、信息检索和医学知识发现[2-3]等领域中。

当前实体关系抽取方法按照模型结构可分为

收稿日期: 2022−04−29. **网络出版日期:** 2023−11−13.
基金项目: 国家自然科学基金项目(61972182).
通信作者: 夏鸿斌. E-mail: hbxia@jiangnan.edu.cn.

流水线(pipline)抽取和联合(joint)抽取两种。流水线抽取方法将三元组抽取分成实体抽取和关系抽取 2 个子任务, 一般采取先抽取出实体, 配对后抽取出关系的流程, 在关系抽取时常采用基于特征、基于核函数和基于深度学习 3 种方法。Kambhatla[4] 通过最大熵模型整合从文本中提取出的各种特征, 以此预测关系; Zhao 等[5] 通过不同核函数的组合, 证明了核函数有非常好的组合特征; Soares 等[6] 使用了基于转换器模型(Transformer)[7] 的双向编码表示模型(bidirectional encoder representation from Transformer, BERT)[8] 编码后的实体向量, 通过匹配空缺(matching the blanks, MTB)任务训练模型来进行关系抽取。流水线方法的 2 个子任务间欠缺交互, 在实际抽取过程中会出现实体冗余, 误差传播等问题。联合抽取方法针对流水线抽取方法的缺点, 通过共享参数、联合解码等形式, 将 2 个子任务对应的模型整合成一个模型进行训练, 在一定程度上缓解了误差传播问题。罗欣等[9] 提出的模型, 在关系抽取器中融合了依存句法树, 标签学习器则使用关系抽取器的参数进行预训练, 能够有效提升模型的抽取性能, 但在处理复杂文本时效果较差。Zheng 等[10] 在一种基于新型标记方案的实体关系抽取模型(joint extraction of entities and relations based on a novel tagging scheme, NovelTagging)中将关系类别与实体类别结合起来, 能够联合解码出三元组, 但该模型采用序列标注来进行实体解码, 限定了某个实体只能被标注为主体或客体, 无法很好地处理重叠三元组问题。苗琳等[11] 的基于双向长短期记忆神经网络(bi-directional long short-term memory neural network, BiLSTM)、基于语义依存图的图注意力网络(adjacency matrix of senmantic dependency graph-graph attention network, SDA-GAT)和双向图卷积神经网络(bi-directional graph neural network, BiGCN)的抽取模型(Bi-LSTM+SDA-GAT+BiGCN, BSGB)在预测得到实体间的关系后, 通过 BiGCN 进一步整合关系信息, 一定程度上缓解了重叠三元组问题。

Wei 等[12] 提出了一种基于级联二元标注结构的实体关系抽取模型(a novel cascade binary tagging framework for relational triple extraction, CasRel), 该模型将解码过程分为 2 个阶段, 第 1 阶段抽取得到主体, 第 2 阶段通过相加的方式将主体特征表示与文本特征融合, 再使用指针网络(一种级联二元标注结构)进行关系和客体的联合抽取。指针网络的易堆叠特性使得实体能够在不同标注层被预测出, 有效应对了重叠三元组问题, 但该模型的分阶段解码引入了曝光偏差, 即在训练时, 2 个阶段的输入使用的都是真实样本, 而在推理过程中, 第 2 阶段会使用第 1 阶段的预测值作为输入, 训练和推理的搜索空间不一致, 最终导致了模型的性能下降。同时, 特征融合过程中简单的相加操作导致了特征信息的丢失, 进一步扩大了曝光偏差的影响。Wang 等[13] 提出了基于片段对的单阶段实体关系联合抽取模型(single-stage joint extraction of entities and relations through token pair, TPLinker), 通过握手标注方式(handshaking tagging scheme)单次解码出实体和关系, 从结构上避免了曝光偏差, 但其相对复杂的解码结构增加了计算开销。

针对上述问题, 本研究提出了一种双关系预测和特征融合的实体关系抽取模型(entity relation extraction model with dual relation prediction and feature fusion, DRPFF)。在使用 BERT[8] 对模型进行编码后, 采用解码过程较为简单的指针网络搭建两阶段的双关系预测结构, 第 1 阶段通过指针网络直接对所有关系下的主、客体进行建模, 单次解码出主、客体及其对应关系, 并按照关系种类对主、客体进行配对, 获得候选三元组集; 第 2 阶段, 使用融合后的实体对特征来对实体对间的关系再次预测, 用于进一步排除候选集中的错误三元组。针对分阶段解码引入的曝光偏差, 模型采用了门控线性单元(gated linear unit, GLU)[14] 和条件层规范化(conditional layer normalization, CLN)[15] 组合的结构, 在强调了实体对间的方向性特征的同时, 强化了模型的拟合能力, 能够有效减少曝光偏差对抽取性能的影响。

1　相关工作

重叠三元组问题可以分为单实体重叠(single entity overlap, SEO)和实体对重叠(entity pair overlap, EPO), Normal 对应文本中只含有普通三元组, 如图 1 所示。SEO 例句中 “Jakarta”和 “Jusuf Kalla”2 个客体共享一个主体“Indonesia”, 即发生了单一实体的重叠。EPO 例句中实体对(“Bakso”, “Indonesia”)之间存在“region”和“country”2 种关系, 即同一实体对间存在多种关系。

当前实体关系抽取方法的实体解码方式可以分为序列标注、片段排列和指针网络 3 种。序列标注[16-17] 的常见形式为先通过编码器获得文本的上下文信息, 再使用条件随机场(conditional random field, CRF)[18] 或是 softmax 进行解码。其主要问题是单一标注层无法处理重叠三元组问题,

例如上文提及的 NovelTagging 模型。Zeng 等[19]在其模型(learning the extraction order of multiple relational facts in a sentence with reinforcement learning, OrderCopyRE)中采用了复制(Copy)[20]机制来应对 NovelTagging 中无法处理的重叠三元组问题，并利用强化学习(reinforcement learning, RL)来提取关系三元组抽取顺序，有效提高了抽取性能。片段(span)由 2 个指针构成，实体片段即由实体在文本序列中对应的首部和尾部 2 个指针构成。基于片段排列的方法是先枚举文本序列中所有可能为实体的片段，再计算这些片段是实体的概率。该方式能够较好地应对重叠三元组问题，但枚举片段过程中会出现大量冗余，导致计算量的增加。Eberts 等[21]在提出的基于预训练变换器和片段的实体关系抽取模型(span-based joint entity and relation extraction with transformer pre-training, SpERT)中，采用了基于片段排列的方法，该模型通过最大池化(max-pooling)来融合实体片段的嵌入信息和长度信息，并使用一个实体片段过滤器来过滤掉冗余的实体片段。指针网络则是使用多个标签序列来对同一个文本标注。Li 等[22]结合机器阅读理解(machine reading comprehension, MRC)框架进行命名实体识别，通过 2 个二分类器来预测实体片段，这种标注结构即是指针网络。Yu 等[23]提出了一种基于片段的先提取后标注的方式(extract-then-label method with span-based scheme, ETL-span)，通过指针网络来预测实体片段，能够有效处理重叠三元组问题。王泽儒等[24]的级联标注模型中(novel pointer tagging cascade strategy, NPTCS)模型在 CasRel 的主体抽取部分加入了实体类型的预测，在抽取性能上取得了一定的提升。

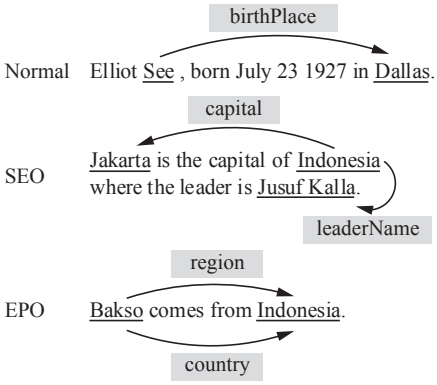

图 1　重叠三元组问题

Fig. 1　Overlapping triple problem

2　DRPFF 模型

2.1　模型框架

DRPFF 分为编码器和解码器两个部分。在编码器部分，使用预训练 BERT 模型对输入文本进行分词和编码。在解码器部分，考虑到序列标注难以处理重叠三元组问题，以及片段排列方法的计算开销，本研究选择了指针网络作为实体解码的方式，并以此构建了一种两阶段的双关系预测结构，来处理重叠三元组、实体对冗余、关系重合等问题。GLU 和 CLN 组合的特征融合结构用来强化阶段间的特征融合，减少曝光偏差的影响。

如图 2 所示，DRPFF 模型分为 2 个部分：1)BERT 编码层，用于获得文本的高维编码向量；2)双关系预测解码层，分为实体抽取模块和关系抽取模块 2 个部分。实体抽取模块，双关系预测结构的第 1 阶段，采用指针网络同时解码出实体对候选集R和关系候选集P_1；关系抽取模块，双关系预测结构的第 2 阶段，融合主体特征E_S和客体特征E_O得到实体对特征$E_{S,O}$，对实体对间的关系进行再次预测，整合得到关系候选集P_2。

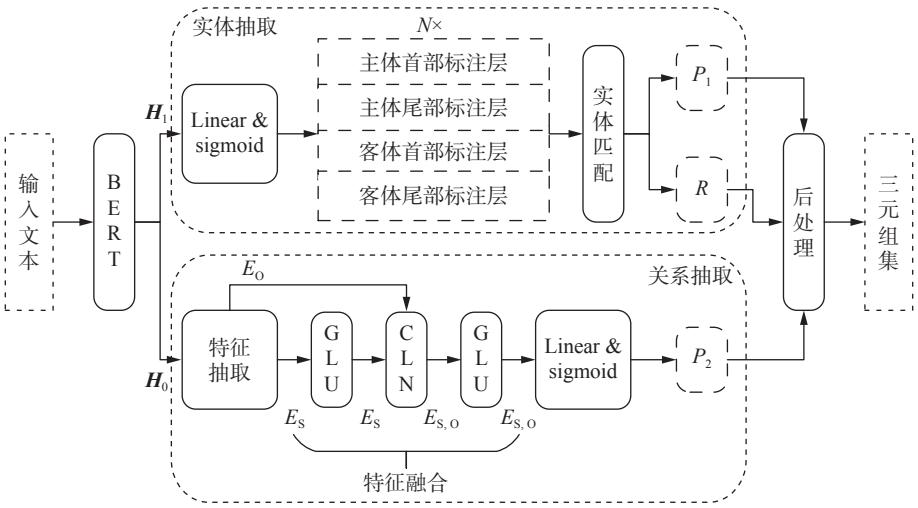

图 2　DRPFF 模型框架

Fig. 2　Framework of DRPFF

2.2　BERT 编码层

BERT[8] 模型由多个双向 Transformer[7] 组件堆叠而成, 因其采用掩码语言模型(masked language model, MLM)和下一句预测(next sentence prediction, NSP)任务来进行训练, 所以能够生成深度的双向语言表征。

对于给定输入文本T, 通过 BERT 的分词工具将其划分为长度为$n+2$的词序列$T^*=\{t_{\mathrm{CLS}},t_1,t_2,\cdots,t_{n-1},t_n,t_{\mathrm{SEP}}\}$, CLS 标识句子的开头, SEP 标识句子的末尾。BERT 的每个双向 Transformer 组件在输出前都会进行层规范化(layer normalization, LN)操作。将T^*按照词表转换后输入到 BERT 中进行编码, 取 BERT 最后一个组件未经层规范化的输出$\boldsymbol{H}_0$和经过层规范化后的输出$\boldsymbol{H}_1$, 作为共享编码供后续操作使用, 具体形式如下

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

$$\boldsymbol{H}_1=\ln(\boldsymbol{H}_0) \tag{2}$$

式(1)为 LN 的计算公式, 其中$\mathrm{E}[\boldsymbol{x}]$为求$\boldsymbol{x}$的均值, $\mathrm{Var}[\boldsymbol{x}]$为求$\boldsymbol{x}$的方差, *为向量间乘积, ε为保持分母不为 0 的一个极小常量; γ和β分别为缩放变量和平移变量, 这 2 个可学习的变量能够保留原有数据特征以及加速训练。LN 在训练过程中能够缓解梯度爆炸、梯度消失等问题。式(2)为$\boldsymbol{H}_0$到$\boldsymbol{H}_1$的计算过程。

2.3　实体抽取模块

由图 2 可知, 本模块使用指针网络来分别对所有关系种类下的主体和客体进行建模, 以此分别预测主体和客体的首部和尾部。给定关系种类数为N, 模型基于输入编码$\boldsymbol{H}_1$, 通过线性层和激活函数生成$(n+2)\times4\times N$个激活概率, 代表着每个分词在不同关系类别下, 都要对主体首部、主体尾部、客体首部和客体尾部这 4 个类别进行一次二分类预测, 具体计算公式如下

$$\boldsymbol{p}_{sh}^{i,r}=\sigma(\boldsymbol{W}_{sh}^{r}\boldsymbol{H}_1^{i}+\boldsymbol{b}_{sh}^{r}) \tag{3}$$

$$\boldsymbol{p}_{st}^{i,r}=\sigma(\boldsymbol{W}_{st}^{r}\boldsymbol{H}_1^{i}+\boldsymbol{b}_{st}^{r}) \tag{4}$$

$$\boldsymbol{p}_{oh}^{i,r}=\sigma(\boldsymbol{W}_{oh}^{r}\boldsymbol{H}_1^{i}+\boldsymbol{b}_{oh}^{r}) \tag{5}$$

$$\boldsymbol{p}_{ot}^{i,r}=\sigma(\boldsymbol{W}_{ot}^{r}\boldsymbol{H}_1^{i}+\boldsymbol{b}_{ot}^{r}) \tag{6}$$

式中: i为分词序号; r为关系序号; $\boldsymbol{p}_{sh}^{(\cdot)}$、$\boldsymbol{p}_{st}^{(\cdot)}$、$\boldsymbol{p}_{oh}^{(\cdot)}$、$\boldsymbol{p}_{ot}^{(\cdot)}$分别代表主体的首部和尾部以及客体的首部和尾部的预测概率; σ代表 sigmoid 激活函数; $\boldsymbol{W}_{(\cdot)}^{r}$和$\boldsymbol{b}_{(\cdot)}^{r}$为关系$r$对应线性层的可训练参数。该结构允许同一个实体同时被标注为主体和客体, 也允许同一实体在不同关系对应的标注层中被同时标注出, 能够有效处理重叠三元组问题。同时, 实体抽取过程中抽取的关系类别是实体配对时的重要筛选依据。

在得到主、客体首部和尾部的预测概率后, 给定一个二分类阈值θ_e。以主体首部的预测为例, 将概率矩阵中大于θ_e的元素置为 1, 反之置为 0。置为 1 的元素, 将其对应文本的分词片段标注为主体首部。同理可以抽取到主体的尾部以及客体的首部和尾部。实体首部和尾部的匹配按照就近原则匹配得到实体片段。抽取到主体和客体后, 在对应关系下将其进行两两匹配, 整理得到候选三元组集。为方便后续操作, 将候选三元组集划分为实体对候选集R和关系候选集P_1, 这 2 个集合按照索引序号相对应。

实体抽取模块的实体配对方式, 只能防止不同关系间的主体和客体进行配对, 无法处理关系重合问题, 需要关系抽取模块来进一步地筛选。

2.4　关系抽取模块

将图 3 的例句 1 输入到实体抽取模块中, 会在关系“contains”对应的标注层中抽取得到主体“Virginia”和“Michigan”, 客体“Norfolk”和“Detroit”, 按照两两匹配的规则可以得到 4 个候选实体对, 生成了 2 个错误实体对, 即由关系重合情况导致的错误三元组生成。此外, 实体抽取模块在实际抽取过程中产生的标注错误也会导致错误三元组的产生。因此, 需要关系抽取模块来进一步筛选掉候选三元组集中的错误三元组。

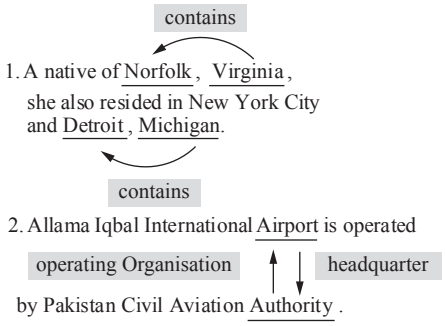

图 3　关系重合情况和实体对间的方向性问题

Fig. 3　The same relation situation, and the orientation problem between entity pair

本模块首先根据实体对集R中的主、客体片段的信息, 从$\boldsymbol{H}_0$中抽取得到主、客体的首部和尾部的特征向量, 再融合这些特征向量得到实体对的特征向量, 最后进行关系的再预测。该过程中需要处理 2 个问题, 一是在推理阶段, 本模块使用实体抽取模块预测因得到R而导致的曝光偏差, 二是主体和客体间的方向性问题。图 3 例句 2 中的实体“Airport”和“Authority”, 在主、客体顺序不同的情况下分别对应了“headquater”和“operating Organisation”2 个关系, 若和 CasRel 一样通过

相加来融合特征，会导致实体对（“Airport”, “Authority”）和（“Authority”, “Airport”）融合的特征相同，无法区分。DRPFF 使用了 GLU 和 CLN 来处理这些问题。

2.4.1 门控线性单元（GLU）

本研究使用的 GLU 源于门控卷积（gated CNN, GCNN）[14]，其计算方式如下

$$\boldsymbol{Y} = \text{conv1d}_1(\boldsymbol{X}) \otimes \sigma(\text{conv1d}_2(\boldsymbol{X})) \tag{7}$$

式中：conv1d_1和conv1d_2是 2 个结构相同的一维卷积；⊗代表元素级的乘法。输入$\boldsymbol{X}$通过 2 个卷积得到 2 个输出，其中的一个进行激活操作，再与另一个未激活的输出做元素级的乘法，最后得到输出$\boldsymbol{Y}$。GLU 中带激活函数的卷积限制了另一个卷积的信息输出，而不带激活函数的卷积则保证了该部分梯度不易消失。

本模块根据实际试验结果对 GLU 的原公式做出了如下改动

$$\text{glu}(\boldsymbol{A}, \boldsymbol{B}) = (\boldsymbol{W}_1\boldsymbol{A} + \boldsymbol{b}_1) \otimes (\boldsymbol{W}_2\boldsymbol{B} + \boldsymbol{b}_2) \tag{8}$$

式中：$\boldsymbol{W}_1$、$\boldsymbol{b}_1$和$\boldsymbol{W}_2$、$\boldsymbol{b}_2$分别为 2 个结构相同的线性层的可训练参数；$\boldsymbol{A}$和$\boldsymbol{B}$代表 2 个形状相同的输入向量。相较于式（7），本研究使用线性层替换了一维卷积，同时拓宽了输入的范围，$\boldsymbol{A}$、$\boldsymbol{B}$为同一个向量时，式（8）与式（7）相同，本质为通过一个非线性变换来强化输入的特征，保证其梯度不易消失，以提高模型的拟合能力；$\boldsymbol{A}$、$\boldsymbol{B}$作为不同向量输入时，式（8）强调了$\boldsymbol{A}$、$\boldsymbol{B}$间的方向性，$\boldsymbol{A}$、$\boldsymbol{B}$的输入顺序不同会导致不同的融合结果。

2.4.2 条件层规范化（CLN）

张龙辉等[15]根据条件批规范化（conditional batch normalization, CBN）[25]的启发，在对序列进行 LN 操作的同时，将额外条件$\boldsymbol{c}_1$和$\boldsymbol{c}_2$融入到 LN 的过程中。CLN 的计算公式为

$$\text{cln}(\boldsymbol{y}, \boldsymbol{c}_1, \boldsymbol{c}_2) = \frac{\boldsymbol{y} - \text{E}[\boldsymbol{y}]}{\sqrt{\text{Var}[\boldsymbol{y}] + \varepsilon}} * \boldsymbol{W}_\gamma \boldsymbol{c}_1 + \boldsymbol{W}_\beta \boldsymbol{c}_2 \tag{9}$$

式中：$\boldsymbol{W}_\gamma$、$\boldsymbol{W}_\beta$为可训练参数；$\boldsymbol{c}_1$和$\boldsymbol{c}_2$为待融合条件。对比式（1），CLN 将条件$\boldsymbol{c}_1$和$\boldsymbol{c}_2$通过线性变换映射到 2 个不同的空间，作为缩放变量γ和平移变量β加入到$\boldsymbol{y}$的规范化过程中。CLN 中变量的不同输入顺序，同样会导致不同的融合结果，能够体现变量间的方向性特征。

2.4.3 特征融合和关系预测

根据消融试验的结果，本研究选择了图 2 中的特征融合结构，以R中的实体对a、b为例，a为主体，b为客体。根据实体片段信息，从$\boldsymbol{H}_0$中抽取出a和b对应的首、尾部向量$\boldsymbol{h}^a_{\text{sh}}$、$\boldsymbol{h}^a_{\text{st}}$、$\boldsymbol{h}^b_{\text{oh}}$、$\boldsymbol{h}^b_{\text{ot}}$，进行拼接得到$a$和$b$的特征向量$[\boldsymbol{h}^a_{\text{sh}}; \boldsymbol{h}^a_{\text{st}}]$和$[\boldsymbol{h}^b_{\text{oh}}; \boldsymbol{h}^b_{\text{ot}}]$，分别以$\boldsymbol{E}_a$和$\boldsymbol{E}_b$表示。首先$\boldsymbol{E}_a$通过 GLU 和残差进行特征的强化，计算过程为

$$\boldsymbol{E}_a = \text{glu}(\boldsymbol{E}_a, \boldsymbol{E}_a) + \boldsymbol{E}_a \tag{10}$$

然后使用 CLN 融合$\boldsymbol{E}_a$和$\boldsymbol{E}_b$，得到实体对特征向量$\boldsymbol{E}_{a,b}$：

$$\boldsymbol{E}_{a,b} = \text{cln}(\boldsymbol{E}_a, \boldsymbol{E}_b, \boldsymbol{E}_b) \tag{11}$$

接着，$\boldsymbol{E}_{a,b}$再次通过 GLU 和残差进行特征的强化：

$$\boldsymbol{E}_{a,b} = \text{glu}(\boldsymbol{E}_{a,b}, \boldsymbol{E}_{a,b}) + \boldsymbol{E}_{a,b} \tag{12}$$

最后，模型使用$\boldsymbol{E}_{a,b}$对实体对间的关系再次预测。由于 EPO 问题的存在，单一实体对信息输入后，模型需要根据该输入对关系表上的每个关系种类进行一个二分类预测。实体对a、b对关系r的预测概率计算公式为

$$\boldsymbol{p}^{a,b}_r = \sigma(\boldsymbol{W}_r \boldsymbol{E}_{a,b} + \boldsymbol{b}_r) \tag{13}$$

式中$\boldsymbol{W}_r$和$\boldsymbol{b}_r$为关系种类r下的可训练参数。在得到预测概率后，给定二分类阈值θ_r，对概率矩阵进行和实体抽取时相同的二值化操作。值为 1 的元素的下标，在关系表中对应的关系即构成实体对a、b的关系集。R中实体对都进行关系预测后，所有获得的关系集组合构成关系集P_2，P_2和P_1相交后即可得到修正的关系集，与R匹配后即可输出最终结果。

2.5 损失计算

DRPFF 在解码过程中采用的均为二分类结构，根据二元交叉熵L_{bce}公式、实体抽取模块损失L_{entity}、关系抽取模块损失L_{relation}和模型整体损失L_{total}的计算公式如下

$$L_{\text{bce}}(p, q) = -[q \log p + (1 - q) \log(1 - p)] \tag{14}$$

$$L_{\text{sh}} = -\frac{1}{N(n+2)} \sum_{i}^{n+2} \sum_{r}^{N} L_{\text{bce}}(\boldsymbol{p}^{i,r}_{\text{sh}}, \boldsymbol{q}^{i,r}_{\text{sh}}) \tag{15}$$

$$L_{\text{entity}} = L_{\text{sh}} + L_{\text{st}} + L_{\text{oh}} + L_{\text{ot}} \tag{16}$$

$$L_{\text{relation}} = -\frac{1}{N \times L} \sum_{a,b}^{R} \sum_{r}^{N} L_{\text{bce}}(\boldsymbol{p}^{a,b}_r, \boldsymbol{q}^{a,b}_r) \tag{17}$$

$$L_{\text{total}} = L_{\text{entity}} + L_{\text{relation}} \tag{18}$$

式（14）中p为预测值，q为真实值。式（15）中，L_{sh}代表主体首部的损失，$\boldsymbol{p}^{i,r}_{\text{sh}}$为第$i$个分词片段在关系$r$下为主体首部的概率，$\boldsymbol{q}^{i,r}_{\text{sh}}$为对应真实标签值。主体尾部以及客体的首部和尾部的损失，分别对应L_{st}、L_{oh}、L_{ot}，其计算过程与L_{sh}相同。式（16）为 4 个部分的损失之和。式（17）中L为集合R的大小，$\boldsymbol{p}^{a,b}_r$为实体对a、b对关系r的预测概率，$\boldsymbol{q}^{a,b}_r$为a、b在关系r下的真实标签。式（18）为 2 个模块损失的总和。

3　试验及分析

3.1　数据集和评价指标

本研究试验采用的是 2 个通用数据集: NYT[26] 和 WebNLG[27]。这 2 个数据集中包含了各种含重叠三元组的文本, 其具体构成如表 1 所示。NYT 数据集的关系分类较少, 且实体对间的关系类别多为地域包含和人际关系, 含关系重合情况的文本较多。WebNLG 数据集的关系分类更加详细, 种类更多, 含关系重合情况的文本较少。在测试集中, NYT 数据集中含关系重合情况的文本有 143 条, 占 2.9%; WebNLG 数据集有 7 条, 占 1.0%。

表 1　各数据集统计信息

Table 1　Statistical data information for each dataset

数据集	样本数			关系数
	训练集	验证集	测试集	
NYT	56 195	4 999	5 000	24
WebNLG	5 019	500	703	171

表 2 和表 3 为 2 个数据集的测试集根据重叠三元组问题和单文本包含三元组个数进行划分后的构成。Normal 代表文本中不存在重叠三元组, SEO、EPO 为上文提到的 2 种重叠三元组情况, Q(与关系种类数 Q 不同)代表单个文本中所包含的关系三元组数。

表 2　测试集中不同重叠三元组问题样本的分布信息

Table 2　Distribution information of samples containing different overlapping triplet problems in test set

数据集	Normal	SEO	EPO
NYT	3 266	1 297	978
WebNLG	246	457	26

表 3　测试集中包含不同三元组数量样本的分布信息

Table 3　Distribution information of samples containing different numbers of triples in test set

数据集	Q=1	Q=2	Q=3	Q=4	$Q\geqslant 5$
NYT	3 244	1 045	312	291	108
WebNLG	266	171	131	90	45

模型抽取性能采用实体关系抽取任务中常用的 F_1(F1-score)、精确率(Precision)、召回率(Recall)作为评价指标, 其中 F_1 为精确率和召回率的调和平均值, 是模型性能的综合评价指标。

在模型计算复杂度方面采用乘加累积操作数(multiply-accumulate operations, MACs)作为量化标准, 同时增加了解码器参数量、模型单轮平均训练时间、单条测试文本平均推理时间这 3 个评价指标, 来对模型的计算性能进行综合对比。

3.2　试验环境和模型参数

本试验环境如下: CPU 为 Intel(R) Core(TM) i9-10900K, GPU 为 GeForce RTX 2060, 内存为 DDR4 16 GB, 开发环境为 Windows 10 64 位系统, PyTorch1.7.1。

本试验使用了 bert-base-cased 预训练模型, 输入文本最大长度为 100, 预训练模型输入最大序列长度为 512, batch_size 设置为 4, 用于抽取实体首部和尾部的二分类阈值 θ_e, 以及用于抽取实体对间关系的二分类阈值 θ_r, θ_e和θ_r的值均设为 0.6, 采用 Adam[28] 优化器, 学习率设置为 0.000 01, 对于 NYT 数据集, epoch 设置为 100, 对于 WebNLG 数据集, epoch 设置为 200。

3.3　对比试验结果和分析

为了综合评估 DRPFF, 试验中将其与几种较为先进的基线模型进行了比较。

基于关系图的实体关系抽取模型(modeling text as relational graphs for joint entity and relation extraction, GraphRel): Fu 等[29] 利用图卷积网络(graph convolutional network, GCN)来进行关系三元组抽取, 包含 1p 和 2p 2 个阶段, 1p 阶段初步预测出实体对间的关系, 2p 阶段使用加权 GCN 为 1p 预测的关系建立完整加权图, 进一步提升预测效果。

OrderCopyRE: Zeng 等[20] 提出的强化学习和复制机制相结合的模型。

ETL-span、CasRel: Yu 等[23] 和 Wei 等[12] 提出的两种利用指针网络进行联合抽取的模型。

TPLinker: Wang 等[13] 提出了握手标注方式, 能够单次解码出三元组, 避免了曝光偏差问题。

3.3.1　整体性能对比

由表 4 可知, 对比 TPLinker, DRPFF 虽然在 NYT 数据集的召回率上低了 0.9%, 但在精确率和 F_1 上分别提升了 1.4% 和 0.2%, 整体性能更有优势; 同时, DRPFF 在 WebNLG 数据集的 3 个指标上分别提升了 1.9%、0.8%、1.4%, 能够有效应对 WebNLG 这种体量小且关系总数大的数据集。综合来看, DRPFF 的抽取性能更好且泛化能力更强。

3.3.2　细节性能对比

为了验证 DRPFF 在各种情况下的抽取性能, 在细节试验中按照 2 种方式对测试集进行划分后, 对各模型的 F_1 进行了对比。表 5 为样本所含关系三元组数 Q 进行划分后进行试验得到的结果, 表 6 为按照重叠三元组问题划分后进行试验得到的结果。

文中将 Normal 类型和 Q=1 的文本称为简单文本, 其他情况的文本则称为复杂文本。综合表 5 和表 6 中的数据来看, DRPFF 在所有区间和分类上的 F_1 值均达到了 90% 以上, 在处理简单文本和复杂文本上, 相比 CasRel 和 TPLinker 都会更有优势。

表 4　各模型整体性能对比

Table 4　Overall performance comparison of each model　%

模型	NYT			WebNLG		
	精确率	召回率	F_1	精确率	召回率	F_1
GraphRel	63.9	60.0	61.9	44.7	41.1	42.9
OrderCopyRE	77.9	67.2	72.1	63.3	59.9	61.6
ETL-span	84.9	72.3	78.1	84.0	91.5	87.6
CasRel	89.7	89.5	89.6	93.4	90.1	91.8
TPLinker	91.3	**92.5**	91.9	91.8	92.0	91.9
DRPFF(本文)	**92.7**	91.6	**92.1**	**93.7**	**92.8**	**93.3**

表 5　细节性能对比 1

Table 5　Detail performance comparison 1　%

模型	NYT					WebNLG				
	Q=1	Q=2	Q=3	Q=4	$Q\geqslant5$	Q=1	Q=2	Q=3	Q=4	$Q\geqslant5$
GraphRel	71.0	61.5	57.4	55.1	41.1	66.0	48.3	37.0	32.1	32.1
OrderCopyRE	71.7	72.6	72.5	77.9	45.9	63.4	62.2	64.4	57.2	55.7
ETL-span	88.5	82.1	74.7	75.6	76.9	82.1	86.5	91.4	89.5	91.1
CasRel	88.2	90.3	91.9	94.2	83.7	89.3	90.8	94.2	92.4	90.9
TPLinker	90.0	**92.8**	93.1	**96.1**	90.0	88.0	90.1	94.6	93.3	91.6
DRPFF(本文)	**90.4**	**92.8**	**93.5**	95.8	**90.1**	**90.1**	**91.9**	**95.3**	**94.8**	**93.0**

表 6　细节性能对比 2

Table 6　Detail performance comparison 2　%

模型	NYT			WebNLG		
	Normal	SEO	EPO	Normal	SEO	EPO
GraphRel	69.6	51.2	58.2	65.8	38.3	40.6
OrderCopyRE	71.2	69.4	72.8	65.4	60.1	67.4
ETL-span	88.5	87.6	60.3	87.3	91.5	80.5
CasRel	87.3	91.4	92.0	89.4	92.2	94.7
TPLinker	90.1	93.4	94.0	87.9	92.5	95.3
DRPFF(本文)	**90.4**	**93.5**	**94.3**	**90.4**	**93.8**	**95.9**

3.3.3　模型计算性能对比

本试验对 CasRel、TPLinker、DRP(即 DRPFF 不使用特征融合结构)和 DRPFF 在 2 个数据集上的计算性能进行了对比。由于上述 4 个模型都使用 BERT 进行编码，试验中只比较模型解码器的 MACs 和参数量；训练时间为各模型在批量设为 4 的情况下，单轮训练所需的平均时间；推理时间为测试集上每条测试文本所需的平均推理时间。

表 7 为各模型计算性能的对比结果。TPLinker 虽然有着不错的抽取性能，但由于单阶段的复杂解码结构，在实际计算性能上相比 CasRel 会有很大的下降，训练时间和推理时间是 CasRel 的数倍。DRP 和 DRPFF 相比 CasRel，在 MACs、训练时间和推理时间上增长的相对较少。DRP 证明了双关系预测结构相比 CasRel 的解码结构更有优势，DRPFF 则证明了特征融合结构的有效性。综合全表的数据来看，DRPFF 在 2 个数据集的单轮训练时间上对比 CasRel 分别增加了 3.13 min 和 0.31 min，测试文本的平均推理时间分别增加了 0.6 ms 和 1.8 ms，F_1 分别提升了 2.5% 和 1.5%，以更少的计算复杂度和计算时间增量，在抽取性能上达到了比 TPLinker 更好的提升。

3.4　消融试验

为了分析 DRPFF 各组件的性能以及 CLN 和 GLU 不同应用方式对模型性能的影响，消融试验中测试了如下模型在 2 个数据集上的性能，其中 GLU_C 代表由一维卷积构成的门控线性单元，其 2 个输入相同；GLU_L 代表由线性层构成的门控线性单元，GLU_L1 代表 2 个输入相同的 GLU_L，GLU_L2 则是 2 个输入不同 GLU_L，试验中的门控线性单元均带有残差结构；CLN1 代表 2 个 CLN 的待融合条件相同的，CLN2 则代表 2 个待融合条件不同的 CLN。

DRP：只使用双关系预测结构抽取三元组，在特征融合时使用向量拼接的方式。

表 7　各模型计算性能对比

Table 7　Comparison of computing performance of each model

数据集	模型	MACs	解码器参数	训练时间/min	推理时间/ms	F_1/%
NYT	CasRel	**19.66**	**38450**	**22.47**	**12.9**	89.6
	TPLinker	6523.55	6017426	72.53	35.5	91.9
	DRP	37.79	110712	24.65	13.3	90.7
	DRPFF(本文)	41.31	3634296	25.60	13.5	**92.1**
WebNLG	CasRel	**135.27**	**264536**	**1.54**	**13.1**	91.8
	TPLinker	9944.30	6695684	21.78	90.1	91.9
	DRP	269.22	788823	1.71	14.6	92.7
	DRPFF(本文)	272.63	4199511	1.85	14.9	**93.3**

E-single: 只使用 DRP 的实体抽取模块。

SO2R: 删除 DRP 实体抽取模块中的关系抽取。

DRP-GLU_C、DRP-GLU_L1、DRP-GLU_L2、DRP- CLN: 在 DRP 的基础上, 依次使用 GLU_C、GLU_L1、GLU_L2 和 CLN1 来进行特征融合。

DRP-CLN-3F: 在 DRP 的基础上, 特征融合时加入平均池化后的$\boldsymbol{H}_0$, 主、客体的特征向量作为两个待融合条件参与到池化后向量的 CLN2 过程中。

DRP-GG: 在 DRP 的特征融合过程中, 主体特征先通过 GLU_L1 强化特征表达, 再通过 GLU_L2 融合主、客体特征。

DRP-GC: 在 DRP 的特征融合过程中, 主体特征先通过 GLU_L1 强化特征表达, 再通过 CLN1 融合主、客体特征。

DRP-CG: 在 DRP 的特征融合过程中, 先使用客体特征加入到主体的 CLN1 过程中, 得到实体对特征后通过 GLU_L1 强化特征表达。

表 8 为消融试验中各模型在 2 个数据集上的整体性能。整个消融试验可以分为 3 组: 第 1 组为双关系预测结构有效性的试验, 包括 E-single、SO2R 和 DRP。第 2 组为单独使用 CLN 或 GLU 进行特征融合的试验, 包括 DRP-GLU_C、DRP-GLU_L1、DRP-GLU_L2、DRP-CLN、DRP-CLN-3F、DRP-GG。第 3 组为将 CLN 和 GLU 组合进行特征融合的试验, 包括 DRP-GC、DRP-CG、DRPFF。

表 8　消融试验结果

Table 8　Results of ablation study　　%

模型	NYT			WebNLG		
	精确率	召回率	F_1	精确率	召回率	F_1
E-single	88.1	**91.8**	90.0	93.9	91.9	92.9
SO2R	89.0	90.9	89.9	90.8	**93.2**	92.0
DRP	91.0	90.5	90.7	93.6	91.9	92.7
DRP-GLU_C	91.5	91.4	91.4	93.6	92.1	92.9
DRP-GLU_L1	91.3	91.5	91.4	93.2	92.9	93.0
DRP-GLU_L2	92.3	91.3	91.8	93.6	91.9	92.8
DRP-CLN	91.7	91.5	91.6	93.8	92.2	93.0
DRP-CLN-3F	89.8	90.7	90.2	93.2	91.1	92.1
DRP-GG	91.9	91.2	91.5	93.9	92.2	93.0
DRP-GC	91.9	91.7	91.8	**94.0**	93.1	**93.5**
DRP-CG	92.2	**91.8**	92.0	**94.0**	92.6	93.3
DRPFF(本文)	**92.7**	91.6	**92.1**	93.7	92.8	93.3

第 1 组试验中, E-single 在 2 个数据集上的差异化表现, 表明了其并不适合处理含关系重合情况的文本。SO2R 因实体配对时没有关系的限制, 实际生成的实体对数量大, 答案覆盖面广, 因此召回率高; 但 SO2R 的关系分类器的过滤性能有限, 导致了很多冗余实体无法过滤掉。DRP 则结合了 E-single 和 SO2R 的结构特点, 通过两次关系预测, 使其在处理关系重叠问题和冗余实体的问题上更加均衡。

第 2 组的试验结果证明了 CLN 和 GLU 都能对特征融合产生正面的影响。DRP-CLN 和 DRP-CLN-3F 的结果证明了 CLN 的 2 个额外条件为同一个时才对模型有正向的增幅效果。

第 3 组试验中, 3 个模型在 2 个数据集上的抽取效果均比单一使用 GLU 或 CLN 的模型更好。在 NYT 数据集上, DRPFF 是该数据集上表现最

好的模型。在 WebNLG 数据集上，DRP-GC 取得了最高的精确率和 F_1，是该数据集上表现最好的模型。DRPFF 却呈现轻微过拟合态势，3 个指标对比 DRP-GC 均出现了 0.2%~0.3% 的下滑。

综合 2 个数据集上各模型的整体表现，考虑到当前大数据计算的普遍性后，文中最后选择了 DRPFF 的模型结构。DRPFF 在含样本量更大的 NYT 数据集上取得了最好的效果，虽然在 WebNLG 数据集上效果会稍弱于 DRP-GC，但仍能够体现 GLU 和 CLN 组合结构的优势。

4 结束语

本研究提出了一种双关系预测和特征融合的实体关系抽取模型（DRPFF）。该模型通过两阶段的双关系预测结构，成功减少了抽取过程中错误三元组的产生，同时明确了主体和客体的特征信息，为特征融合做了准备。GLU 和 CLN 组合的特征融合结构，能够有效提高阶段间的特征融合程度，减少曝光偏差带来的影响。在 2 个数据集上的试验结果表明，相较于当前先进的基线模型，DRPFF 的性能表现占优，且泛化能力更强。未来的工作中，将改进实体解码结构以及实体对匹配机制，完善模型在应对简单文本时性能的不足。

参考文献：

[1] 张勇, 高大林, 巩敦卫, 等. 用于关系抽取的注意力图长短时记忆神经网络 [J]. 智能系统学报, 2021, 16(3): 518–527.
ZHANG Yong, GAO Dalin, GONG Dunwei, et al. Attention graph long short-term memory neural network for relation extraction[J]. CAAI transactions on intelligent systems, 2021, 16(3): 518–527.

[2] 杨志豪, 洪莉, 林鸿飞, 等. 基于支持向量机的生物医学文献蛋白质关系抽取 [J]. 智能系统学报, 2008, 3(4): 361–369.
YANG Zhihao, HONG Li, LIN Hongfei, et al. Extraction of information on protein-protein interaction from biomedical literatures using an SVM[J]. CAAI transactions on intelligent systems, 2008, 3(4): 361–369.

[3] 范智渊, 何璇, 梁品, 等. 中文医学文献的实体关系提取研究及在糖尿病医学文献中的应用 [J]. 生物医学工程学杂志, 2021, 38(3): 563–573.
FAN Zhiyuan, HE Xuan, LIANG Pin, et al. Research on entity relationship extraction of Chinese medical literature and application in diabetes medical literature[J]. Journal of biomedical engineering, 2021, 38(3): 563–573.

[4] KAMBHATLA N. Combining lexical, syntactic, and semantic features with maximum entropy models for extracting relations[C]//Proceedings of the ACL 2004 on Interactive Poster and Demonstration Sessions. Morristown: Association for Computational Linguistics, 2004: 178−181.

[5] ZHAO Shubin, GRISHMAN R. Extracting relations with integrated information using kernel methods[C]//Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics-ACL '05. Morristown: Association for Computational Linguistics, 2005: 419−426.

[6] SOARES L B, FITZGERALD N, LING J, et al. Matching the blanks: distributional similarity for relation learning[C]//Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Stroudsburg: Association for Computational Linguistics, 2019: 2895−2905.

[7] 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. New York: ACM, 2017: 6000−6010.

[8] CHUNG Y A, ZHU Chenguang, ZENG M. SPLAT: speech-language joint pre-training for spoken language understanding[C]//Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Stroudsburg: Association for Computational Linguistics, 2021: 4171−4186.

[9] 罗欣, 陈艳阳, 耿昊天, 等. 基于深度强化学习的文本实体关系抽取方法 [J]. 电子科技大学学报, 2022, 51(1): 91–99.
LUO Xin, CHEN Yanyang, GENG Haotian, et al. Entity relationship extraction from text data based on deep reinforcement learning[J]. Journal of University of Electronic Science and Technology of China, 2022, 51(1): 91–99.

[10] ZHENG Suncong, WANG Feng, BAO Hongyun, 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. Stroudsburg: Association for Computational Linguistics, 2017: 1227–1236.

[11] 苗琳, 张英俊, 谢斌红, 等. 基于图神经网络的联合实体关系抽取 [J]. 计算机应用研究, 2022, 39(2): 424–431.
MIAO Lin, ZHANG Yingjun, XIE Binhong, et al. Joint entity relation extraction based on graph neural network[J]. Application research of computers, 2022, 39(2): 424–431.

[12] WEI Zhepei, SU Jianlin, WANG Yue, et al. A novel cascade binary tagging framework for relational triple extraction[C]//Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. Stroudsburg: Association for Computational Linguistics, 2020: 1476−1488.

[13] WANG Yucheng, YU Bowen, ZHANG Yueyang, et al. TPLinker: single-stage joint extraction of entities and relations through token pair linking[C]//Proceedings of the 28th International Conference on Computational Linguistics. Stroudsburg: International Committee on Computational Linguistics, 2020: 1572–1582.

[14] DAUPHIN Y N, FAN A, AULI M, et al. Language modeling with gated convolutional networks[C]// Proceedings of the 34th International Conference on Machine

Learning. Sydney: ICML, 2017: 933−941.
[15] 张龙辉, 尹淑娟, 任飞亮, 等. BSLRel: 基于二元序列标注的级联关系三元组抽取模型 [J]. 中文信息学报, 2021, 35(6): 74–84.
ZHANG Longhui, YIN Shujuan, REN Feiliang, et al. BSLRel: a binary sequence labeling based cascading relation triple extraction model[J]. Journal of Chinese information processing, 2021, 35(6): 74–84.
[16] ZHANG Yue, YANG Jie. Chinese NER using lattice LSTM[C]//Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics. Stroudsburg: Association for Computational Linguistics, 2018: 1554−1564.
[17] LI Xiaonan, YAN Hang, QIU Xipeng, et al. FLAT: Chinese NER using flat-lattice transformer[C]//Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. Stroudsburg: Association for Computational Linguistics, 2020: 6836−6842.
[18] LAFFERTY J D, MCCALLUM A, PEREIRA F C N. Conditional random fields: probabilistic models for segmenting and labeling sequence data[C]//Proceedings of the Eighteenth International Conference on Machine Learning. New York: ACM, 2001: 282−289.
[19] ZENG Xiangrong, HE Shizhu, ZENG Daojian, 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. Stroudsburg: Association for Computational Linguistics, 2019: 367–377.
[20] ZENG Xiangrong, ZENG Daojian, HE Shizhu, 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. Stroudsburg: Association for Computational Linguistics, 2018: 506–514.
[21] EBERTS M, ULGES A. Span-based joint entity and relation extraction with transformer pre-training[C]//Proceedings of the 24th European Conference on Artificial Intelligence. Santiago de Compostela: ECAI, 2020: 2006−2013.
[22] LI Xiaoya, FENG Jingrong, MENG Yuxian, et al. A unified MRC framework for named entity recognition[C]//Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. Stroudsburg: Association for Computational Linguistics, 2020: 5849−5859.
[23] YU Bowen, ZHANG Zhenyu, SU Jianlin, et al. Joint extraction of entities and relations based on a novel decomposition strategy[C] // Proceedings of the 24th European Conference on Artificial Intelligence. Santiago de Compostela: ECAI, 2020: 2282−2289.
[24] 王泽儒, 柳先辉. 基于指针级联标注的中文实体关系联合抽取模型 [J]. 武汉大学学报(理学版), 2022, 68(3): 304–310.
WANG Zeru, LIU Xianhui. Joint model of Chinese entity-relation extraction based on a pointer cascade tagging strategy[J]. Journal of Wuhan University (natural science edition), 2022, 68(3): 304–310.
[25] DE VRIES H, STRUB F, MARY J, et al. Modulating early visual processing by language[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. New York: ACM, 2017: 6597−6607.
[26] RIEDEL S, YAO Limin, MCCALLUM A. Modeling relations and their mentions without labeled text[C]//Proceedings of the 2010th European Conference on Machine Learning and Knowledge Discovery in Databases-Volume Part III. New York: ACM, 2010: 148−163.
[27] GARDENT C, SHIMORINA A, NARAYAN S, et al. Creating training corpora for NLG micro-planners[C]//Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics. Stroudsburg: Association for Computational Linguistics, 2017: 179–188.
[28] KINGMA D P, BA J L. Adam: a method for stochastic optimization[C]//Proceedings of the 3rd International Conference on Learning Representations. San Diego: ICLR, 2015.
[29] FU T J, LI P H, MA Weiyun. 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. Stroudsburg: Association for Computational Linguistics, 2019: 1409–1418.

作者简介:

沈健, 硕士研究生, 主要研究方向为自然语言处理、实体关系抽取。E-mail: 1452112297@qq.com。

夏鸿斌, 教授, 博士, 江苏省特色化软件人才培养专委会委员, 江南大学人工智能与计算机学院副院长。主要研究方向为大数据分析与处理、个性化推荐系统、自然语言处理。主持工信部、江苏省科研项目 2 项, 参加国家科技支撑项目 1 项、江苏省自然科学基金重点项目 1 项。获江苏省教育成果二等奖、教育部科技成果奖、中国商业联合会科技进步特等奖。发表学术论文 20 余篇。E-mail: hbxia@jiangnan.edu.cn。

刘渊, 教授, 博士生导师, 中国网络空间安全协会会员, 江南大学人工智能与计算机学院院长, 主要研究方向为网络安全、社交网络。作为项目负责人完成省部级科研项目多项。获中国商业联合会科技奖特等奖、无锡市有突出贡献的中青年专家称号。发表学术论文 40 余篇。E-mail: lyuan1800@sina.com。