

# 改进 RoBERTa、多实例学习和双重注意力机制的关系抽取方法

王禹鸥<sup>1</sup>,苑迎春<sup>1,2\*</sup>,何振学<sup>1</sup>,王克俭<sup>1</sup>

(1.河北农业大学信息科学与技术学院,河北保定 071001; 2.河北省农业大数据重点实验室(河北农业大学),河北保定 071001)

**摘要:**针对远程监督关系抽取不能充分利用句子上下文高层信息、易带来噪声标注的问题,提出一种基于改进鲁棒优化的双向编码器表征预训练模型(robustly optimized bidirectional encoder representations from Transformers pretraining approach, RoBERTa)、多实例学习(multiple-instance learning, MI)和双重注意力(dual attention, DA)机制的关系抽取方法。在 RoBERTa 中引入全词动态掩码,获取文本上下文信息,获得词级别语义向量;将特征向量输入双向门控循环单元(bidirectional gated recurrent unit, BiGRU),挖掘文本深层次语义表征;引入多实例学习,通过学习实例级别特征缩小关系抽取类别范围;引入双重注意力机制,结合词语级注意力机制和句子级注意力机制的优势,充分捕捉句子中实体词语特征信息和对有效语句的关注度,增强句子表达能力。试验结果表明,在公开数据集纽约时报(New York Times, NYT)数据集和谷歌 IISc 远程监督(Google IISc distant supervision, GIDS)数据集中,关系抽取方法的  $F_1$  值分别为 88.63%、90.13%,均优于主流对比方法,能够有效降低远程监督噪声影响,实现关系抽取,为构建知识图谱提供理论基础。

**关键词:**远程监督;关系抽取;改进 RoBERTa;多实例学习;双重注意力机制

**中图分类号:**TP391

**文献标志码:**A

**引用格式:**王禹鸥,苑迎春,何振学,等.改进 RoBERTa、多实例学习和双重注意力机制的关系抽取方法[J].山东大学学报(工学版),2025,55(2):78-87.

WANG Yuou, YUAN Yingchun, HE Zhenxue, et al. A relation extraction method based on improved RoBERTa, multiple-instance learning and dual attention mechanism[J]. Journal of Shandong University (Engineering Science), 2025, 55(2):78-87.

## A relation extraction method based on improved RoBERTa, multiple-instance learning and dual attention mechanism

WANG Yuou<sup>1</sup>, YUAN Yingchun<sup>1,2\*</sup>, HE Zhenxue<sup>1</sup>, WANG Kejian<sup>1</sup>

(1. College of Information Science and Technology, Hebei Agricultural University, Baoding 071001, Hebei, China; 2. Hebei Province Key Laboratory of Agricultural Big Data, Hebei Agricultural University, Baoding 071001, Hebei, China)

**Abstract:** Aiming at the problem that distant supervision relation extraction could not make full use of the high-level information of sentence context and was easy to bring noise annotations, a relation extraction method based on improved robustly optimized bidirectional encoder representations from Transformers pretraining approach (RoBERTa), multiple-instance learning (MI) and dual attention (DA) mechanism was proposed. The full-word dynamic mask was introduced on the RoBERTa to obtain the text context information and the word-level semantic vector. The feature vectors were input into bidirectional gated recurrent unit (BiGRU) to mine the deep semantic representation of the text. Multiple-instance learning was introduced to narrow the range of relation extraction categories by learning instance-level features. Dual attention mechanism was introduced, which combined the advantages of word-level attention mechanism and sentence-level attention mechanism to fully capture the feature information of

收稿日期:2024-03-07

基金项目:国家自然科学基金资助项目(62102130)

第一作者简介:王禹鸥(2000—),女,河北廊坊人,硕士研究生,主要研究方向为自然语言处理。E-mail:20222060106@pgs.hebau.edu.cn

\* 通信作者简介:苑迎春(1970—),女,河北保定人,教授,博士生导师,博士,主要研究方向为智能信息处理与大数据研究。

E-mail:nd\_hd\_yc@163.com

entity words in the sentence, improved the model's attention of effective sentences, and enhanced the expression ability of sentences. The experimental results showed that the  $F_1$  value of the method reached 88.63% and 90.13% on the public dataset New York Times (NYT) and Google IISc distant supervision (GIDS), which were better than the mainstream comparison methods. It could effectively reduce the noise influence of distant supervision, realize the relation extraction, and lay a theoretical foundation for the construction of knowledge graph.

**Keywords:** distant supervision; relation extraction; improved RoBERTa; multiple-instance learning; dual attention mechanism

## 0 引言

近年来,互联网数据呈现爆炸式增长,如何从海量的非结构化数据中提取有用信息成为关键问题。传统搜索引擎工具虽然能够通过索引和检索技术帮助用户找到所需信息,但仍存在信息过载、搜索引擎无法准确理解用户意图导致结果不准确等问题<sup>[1]</sup>。信息抽取能够从大量非结构化文本中抽取结构化信息,实现对文本的深层次理解,为用户提供语义特征更丰富、更能体现用户需求的信息,提高查询效率。关系抽取是信息抽取的关键步骤<sup>[2]</sup>,从非结构化文本中抽取不同实体之间的关系,对知识图谱构建、信息检索、智能问答等具有重要作用<sup>[3]</sup>。

关系抽取方法通常可分为有监督关系抽取、无监督关系抽取和半监督关系抽取3种<sup>[4]</sup>。其中,基于深度学习的有监督关系抽取方法被广泛应用<sup>[5]</sup>,能自动学习文本特征,减少对预处理模块的依赖,提升关系抽取效果。但该方法需要大量人工标注数据进行训练,限制了大规模实体关系抽取。为减少有监督模型对标记训练语料库的依赖,文献[6]提出远程监督关系抽取方法,无需人工标注,可节省大量时间和成本,能够利用外部语料库学习文本特征,提高关系抽取精度。但远程监督假设2个实体若存在某种关系,则所有包含这2个实体的句子都必须表达为这种关系<sup>[7]</sup>。由于该假设未利用句子上下文语义特征,容易引起噪声标注问题,因此,解决噪声标注问题成为远程监督方法的研究重点。

## 1 相关工作

一些研究将深度学习模型与远程监督关系抽取相结合。文献[8-11]分别利用分段卷积神经网络、图卷积神经网络、双向长短期记忆(bidirectional long short-term memory, BiLSTM)网络及双向门控循环单元(bidirectional gated recurrent

unit, BiGRU)捕获文本特征信息,减少标签错误标注问题。为提高远程监督关系抽取准确度,预训练语言模型、多实例学习(multiple-instance learning, MI)、注意力机制等技术也用于深度学习网络中。文献[12-13]将双向编码器表征(bidirectional encoder representations from Transformers, BERT)法应用于临床医学、生物医学领域关系抽取任务中,增强上下文信息语义表示。鲁棒优化的双向编码器表征预训练模型(robustly optimized bidirectional encoder representations from Transformers pretraining approach, RoBERTa)改进BERT静态掩码策略,采用字符级动态掩码技术,使模型全方位学习输入层的文本向量,获得比BERT模型更强的表征学习能力。文献[14-17]利用RoBERTa学习不同的语言表征,为给定输入生成具有代表性的词嵌入,提高远程监督关系抽取的有效性;文献[18]在RoBERTa的基础上增加模型预训练数据集,同时增加训练步数,提高地质领域远程监督关系抽取效果。多实例学习方法能够利用多个实例从不同上下文中学习关系特征,减少噪声标注的影响。文献[18-19]利用多实例学习方法学习同一实体对可能存在的不同关系,将最大预测概率关系类型作为包的标签,减少噪声标签的影响。注意力机制通过调配信息权重解决远程监督中的数据噪声问题。文献[20-21]采用句子级注意力机制的远程监督关系抽取方法动态赋予权重,降低噪声数据的影响;文献[22]利用词级注意力机制为表达语义关系更重要的词分配权重,提高关系抽取准确率;文献[23]提出一种分层注意力机制,利用句子级和包级注意力机制增强特征信息较差的目标包特征,提高关系抽取性能。

上述文献主要利用某一种方法对深度学习网络模型进行改进,综合利用预训练模型、注意力机制等方法的关系抽取也得到一些学者研究:文献[24]使用RoBERTa作为编码器,结合基于注意力机制的原型网络预测关系类型,提高桥梁检测和健康领域的关系抽取精度;文献[25]利用RoBERTa生成文本向量,结合交叉传递注意力机制解决实体

类型缺失问题,在3个公开数据集上均优于对比模型;文献[26]利用加权求和获取 RoBERTa 编码的输出,采用多头注意力机制捕获事实描述和元素之间的关系,提高法律文件数据集的关系抽取效果;文献[27]采用 RoBERTa 获取语料库的文本向量,提出实体注意力网络选择重要词,在公开数据集的准确率优于对比方法。上述方法均采用 RoBERTa 和注意力机制结合的方法提高关系抽取精度,但 RoBERTa 以字符为单位进行掩码,导致词的语义特征表达不完整,模型不能充分捕捉词语的上下文信息。仅基于词语级注意力机制的方法没有考虑不同句子对关系抽取的影响,仅基于句子级注意力机制的关系抽取方法未考虑不同词语实体在句子中的重要程度,更多情况下,不同的实体词语和句子都会影响关系抽取效果。

为提高远程监督关系抽取效果,本研究提出一种基于改进的 RoBERTa、多实例学习和双重注意力

(dual attention, DA) 机制(improved RoBERTa, MI and DA, IR\_MI\_DA) 关系抽取方法。该模型改进 RoBERTa,将以字符为粒度的词嵌入改进为采用全词掩码技术词嵌入,以获取实体词级别的特征表示,更好保留词的完整语义信息;引入多实例学习,利用实体对的多个实例,从不同上下文中学习关系特征,缩小关系类别的抽取范围;提出结合词语级注意力和句子级注意力的双重注意力机制,充分捕捉句子中实体特征信息,提高模型对有效语句的关注度,提升关系抽取精度。通过试验验证模型的有效性。

## 2 关系抽取方法 IR\_MI\_DA

本研究提出的关系抽取方法 IR\_MI\_DA 分为5部分,分别为改进的 RoBERTa 层、BiGRU 层、多实例学习层、双重注意力机制层和解码层,模型总体框架如图1所示。

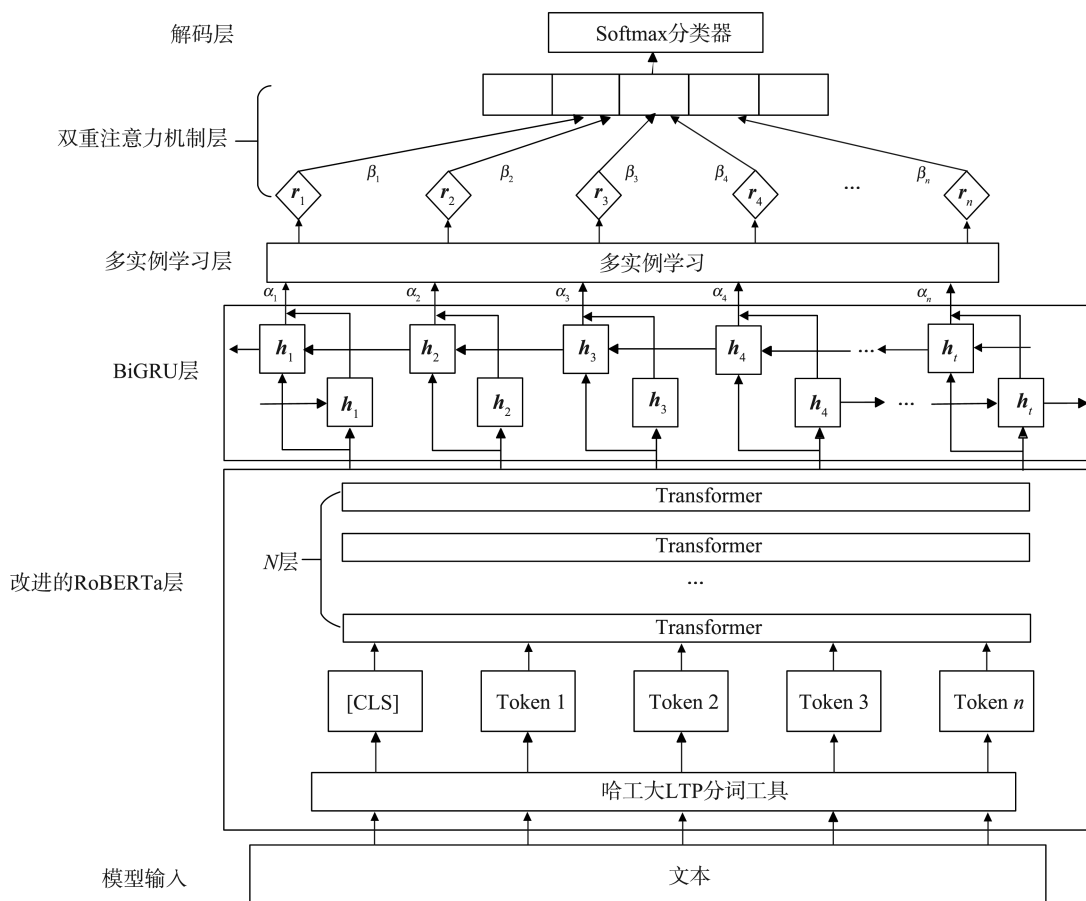


图1 关系抽取方法结构图

Fig.1 Relation extraction method structure diagram

### 2.1 改进的 RoBERTa 层

RoBERTa 以字符为单位进行动态掩码操作,原有基于 Wordpiece 的分词方式将一个完整的词切分为若干个字符,在生成训练样本时,这些字符被随

机掩码,使词的语义特征表达不充分,不能利用模型捕捉实体词语的语义信息。

本研究在 RoBERTa 的基础上,通过添加分词操作进行改进,即将以字符为单位的掩码操作改为

全词掩码操作,在全词掩码中,如果一个完整词的部分字符被掩码,则同属于这个词的其他部分也被掩码,即全词掩码。使用哈工大语言技术平台(language technology platform, LTP)分词工具将语料进行分词,随机遮蔽一部分词,即对组成同一个词的字符全部掩码。该方法以词为单位的掩码操作代替以字符为单位的掩码操作,改变样本生成过程,更好地保留词的完整语义信息,捕捉文本深层次语义特征,得到较为全面的词向量和位置向量。

假设文本语料库为  $X = \{x_1, x_2, \dots, x_n\}$ , 其中  $x_i$  为第  $i$  条语料,  $i = 1, 2, \dots, n$ 。令  $Y_{ij}$  为  $x_i$  的第  $j$  个词, 则  $x_i = \{Y_{i1}, Y_{i2}, \dots, Y_{im}\}$ ,  $j = 1, 2, \dots, m$ , 将  $Y_{im}$  表示为字特征  $C_{im} = \{c_1, c_2, \dots, c_a\}$ , 其中  $c_a$  为第  $m$  个词的第  $a$  个字特征。模型的输入向量

$$\tilde{Y}_{im} = Y_{im} + S_{im} + P_{im}, \quad (1)$$

式中:  $S_{im}$  为句子特征,  $S_{im} = \{s_1, s_2, \dots, s_a\}$ , 其中  $s_a$  为第  $i$  条语料的第  $a$  个句子特征;  $P_{im}$  为位置特征,  $P_{im} = \{p_1, p_2, \dots, p_a\}$ , 其中  $p_a$  为第  $i$  条语料的第  $a$  个位置特征。输入计算方式流程图如图 2 所示, 其中  $\tilde{C}_{im}$  为得到字特征、句子特征和位置特征后的文本向量。

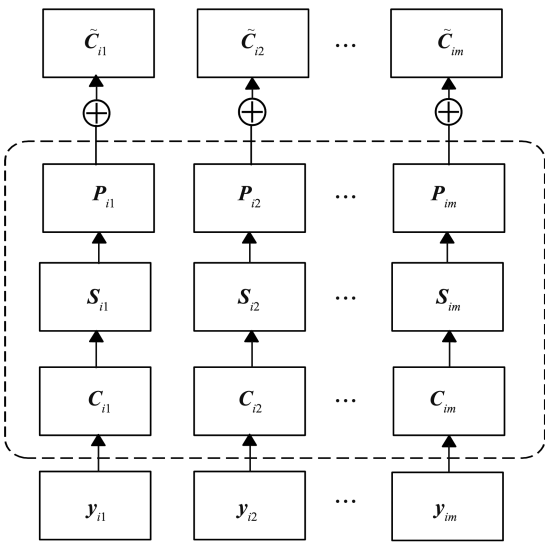


图2 输入计算方式

Fig.2 Input calculation method

RoBERTa 的  $N$  个 Transformer 将语义信息的特征表示  $\tilde{E}_{im}$  转换为融合上下文语义信息的特征表示  $\tilde{E}_{im}^n$ , 即

$$\tilde{E}_{im}^n = \text{Transformers}(\tilde{E}_{im}), \quad (2)$$

式中:  $\text{Transformers}(\cdot)$  为 Transformer 处理过程;  $n$  为 Transformer 数量,  $n = 1, 2, \dots, N$ 。改进的 RoBERTa 处理过程如图 3 所示。通过分词掩码操

作,改进的 RoBERTa 可以学习词级别的语义信息。

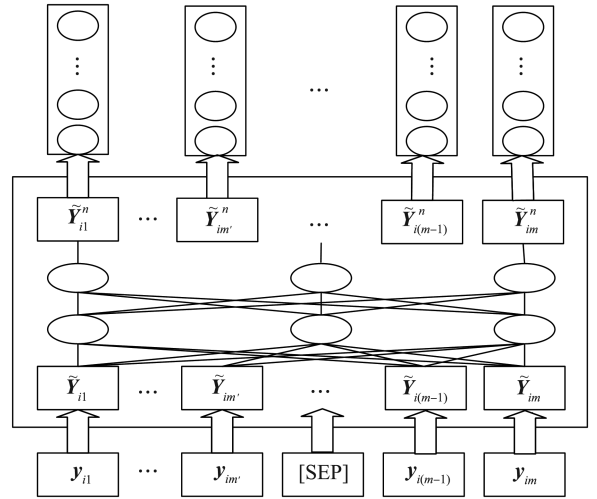


图3 改进的 RoBERTa 处理过程

Fig.3 Improved RoBERTa processing

### 2.2 BiGRU 层

为充分获取句子上下文的深层语义信息,本研究采用 BiGRU 模型对输入向量进行特征提取。BiGRU 模型包含重置门、更新门和 2 个隐状态。重置门和更新门的输出均通过 Sigmoid 激活函数。重置门的作用是判断前一时刻所需遗忘的隐状态信息,允许模型选择性遗忘一些历史信息;更新门的作用是决定保留多少前一时刻的信息,用于对当前时刻的隐状态进行补充。

将句子第  $t$  个词语的词向量  $x_t$  作为 GRU 的输入,获取第  $t$  个词的门控循环单元(gated recurrent unit, GRU)各个门单元状态的特征值,计算式为

$$r_t = \sigma(W_r x_t + W_r h_{t-1} + b_r), \quad (3)$$

$$z_t = \sigma(W_z x_t + W_z h_{t-1} + b_z), \quad (4)$$

$$\tilde{h}_t = \tanh(W x_t + (r_t \odot h_{t-1}) W + b_{\tilde{h}}), \quad (5)$$

$$h_t = z_t \odot \tilde{h}_t + (1 - z_t) \odot h_{t-1}, \quad (6)$$

式中,  $r_t$  和  $z_t$  分别为重置门与更新门的输出向量,  $\tilde{h}_t$  为候选隐状态输出,  $h_t$  为当前时刻隐状态输出,  $h_{t-1}$  为前一时刻的隐状态输出,  $\sigma$  为 BiGRU 模型运算过程,  $W_r$ 、 $W_z$  分别为重置门、更新门的权重矩阵,  $b_r$ 、 $b_z$  分别为重置门、更新门的偏置参数,  $W$  为权重矩阵,  $b_{\tilde{h}}$  为候选隐状态的偏置参数,  $\odot$  代表矩阵相乘。

BiGRU 由正向 GRU 和反向 GRU 组成,可以学习到前一时刻和后一时刻与当前状态之间的序列关系。通过 BiGRU 可以获取第  $t$  个词语的正向输出状态  $\tilde{h}_t$  和反向输出状态  $\tilde{h}_t$ , 2 个方向的输出状态进行拼接, 即  $h_t = \tilde{h}_t + \tilde{h}_t$ , 构成 BiGRU 的输出状态, 由此获取句子中词语实体的高维语义表示, 挖掘文本潜在的代表关系。

### 2.3 多实例学习层

远程监督中产生的大量噪声标注会对关系抽取效果造成影响,误差累积和传播直接影响关系抽取精度。为缓解上述问题,本研究在 BiGRU 模型后引入多实例学习层。在多实例学习中,将文本中包含同一实体对的所有句子(实例)看成一个包。若同一实体对在不同实例中表达不同的关系,则多实例学习能够利用多个实例进行学习,从不同上下文中学习关系特征,缩小关系抽取类别范围,减少噪声标注的影响。

远程监督核心思想如下。假设训练样本分为  $T$  个包,即  $\{\mathbf{M}_1, \mathbf{M}_2, \dots, \mathbf{M}_T\}$ 。第  $i$  个包中包含  $q$  个实例,即  $\mathbf{M}_i = \{\mathbf{m}_i^1, \mathbf{m}_i^2, \dots, \mathbf{m}_i^q\}$ ,每个输入的包之间相互独立。以包中的一个实例作为输入,给定参数  $\theta$ ,由神经网络输出向量,由 Softmax 分类器计算置信度并转换为预测概率,计算式为

$$p(r|\mathbf{m}_i^j, \theta) = \frac{e^{o_r}}{\sum_{r=1}^n e^{o_r}}, \quad (7)$$

$$J(\theta) = \sum_{i=1}^T \lg p(y_i|\mathbf{m}_i^j, \theta), \quad (8)$$

$$j^* = \arg \max_j p(y_i|\mathbf{m}_i^j, \theta), \quad (9)$$

式中:  $p(r|\mathbf{m}_i^j, \theta)$  为给定输入实例  $\mathbf{m}_i^j$  和参数  $\theta$  的条件下,预测关系  $r$  的概率;  $o_r$  为预测关系  $r$  的相关得分;  $J(\theta)$  为所有包中实例对数概率之和;  $y_i$  为实例  $\mathbf{m}_i^j$  的真实标签;  $j^*$  为实例的关系预测概率。

## 2.4 双重注意力机制层

### 2.4.1 词语级注意力机制

对于关系抽取任务,不同词语在句子中的重要程度不同,其对应的实体对也不同。因此,本研究通过计算实体对之间的相关性,建立词注意力机制,捕获实体上下文语义特征。

假设输入词语级注意力机制的向量为  $\mathbf{Z}$ ,  $\mathbf{Z} = [\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_d]$ ,其中  $\mathbf{z}_d$  为向量  $\mathbf{Z}$  中的第  $d$  个词语,计算向量  $\mathbf{Z}$  中每个实体对的注意力权重,句子向量

$$\mathbf{x} = \sum_{i=1}^d \alpha_i \mathbf{z}_i, \quad (10)$$

式中:  $d$  为句子长度;  $\alpha_i$  为上下文实体对的权重,  $\alpha_i \in [0, 1]$ ,  $\sum_i \alpha_i = 1$ 。

利用评分函数计算词语与实体对的相关程度

$$\mu_i = \tanh(\mathbf{W}_{\text{word-att}}[\mathbf{z}_i \mathbf{h}_{\text{ch}} \mathbf{h}_{\text{et}}] + \mathbf{b}_{\text{word-att}}), \quad (11)$$

式中,  $\mathbf{W}_{\text{word-att}}$  为词语相关性权重矩阵,  $\mathbf{b}_{\text{word-att}}$  为偏置值,  $\mathbf{h}_{\text{ch}}$  为实体对的头实体,  $\mathbf{h}_{\text{et}}$  为实体对的尾实体。通过 Softmax 函数学习对  $\mu_1, \mu_2, \dots, \mu_i$  进行归一

化操作,得到权重

$$\alpha_i = \frac{\exp(\mu_i)}{\sum_{k=1}^i \exp(\mu_k)}. \quad (12)$$

利用最大池化策略,选择句子最显著特征,得到句子表示

$$\mathbf{S} = \text{maxpool}(\mathbf{x}), \quad (13)$$

式中 maxpool 为最大池化操作。

### 2.4.2 句子级注意力机制

词语级注意力机制已经学习到实体深层次的特征信息,句子级注意力机制在词语级注意力机制的基础上充分捕捉对有效语句的关注度,挖掘相同实体对的不同语义,增强句子表达能力,有助于判断句子对应预测关系的准确性。本研究中句子级注意力机制策略包含计算相同实体对句子与预测关系的相似度,赋予相似度高的句子更多的权重,构建句子级别的向量矩阵,降低噪声句子对关系预测的影响。

假设句子集  $\mathbf{S} = \{\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_k\}$ ,  $\mathbf{s}_k$  为句子集中第  $k$  个句子。这  $k$  个句子中包含相同的实体对。为了表示  $k$  个包含相同实体对的句子与预测关系的相关程度,将  $\mathbf{S}$  表示为  $k$  个句子的加权和,即

$$\mathbf{S} = \sum_{i=1}^k \eta_i \mathbf{s}_i, \quad (14)$$

式中,  $\eta_i$  为句子集中每个句子与预测关系的匹配度,  $\eta_i \in [0, 1]$ ,  $\sum_i \eta_i = 1$ 。通过计算句子向量与实体对对应关系类型的内积,利用 Softmax 函数学习归一化操作,得到句子级注意力权重

$$\beta_i = \text{Softmax}(\mathbf{s}_i \mathbf{A} \mathbf{v}_r), \quad (15)$$

式中,  $\mathbf{A}$  为随机初始化的加权对角矩阵,  $\mathbf{v}_r$  为预测关系  $r$  的向量表示。

## 2.5 解码层

使用 Softmax 分类器从句子  $\mathbf{S}$  的离散集合类  $\mathbf{Y}$  中预测关系标签  $\hat{y}$ ,将文本  $\mathbf{t}_{\text{xt}}$  进行非线性变换。预测关系标签  $\hat{y}$  的计算式为

$$\hat{y} = \arg \max_y \hat{p}[y|\mathbf{S}]. \quad (16)$$

式中:  $\hat{p}[y|\mathbf{S}]$  为在句子为  $\mathbf{S}$  的条件下,预测标签  $y$  的概率,  $\hat{p}[y|\mathbf{S}] = \tanh(\mathbf{W}^{(s)} \cdot \mathbf{t}_{\text{xt}} + \mathbf{b}^{(s)})$ ,其中  $\mathbf{W}^{(s)}$  为权重矩阵,  $\mathbf{b}^{(s)}$  为偏置向量。

利用损失函数计算真实关系标签  $\hat{y}$  的负对数似然值

$$L(\boldsymbol{\theta}) = -\frac{1}{v} \sum_{i=1}^m t_i \lg(q_i) + \lambda \|\boldsymbol{\theta}\|_{\text{F}}^2, \quad (17)$$

式中:  $t_i$  为一次性学习的真实值,  $t_i \in \mathbf{R}^m$ ;  $q_i$  为使用

Softmax 函数估计每个关系类别的概率,  $q_i \in \mathbf{R}^m$ ;  $\nu$  为关系类别的数量;  $\lambda$  为 L2 正则化参数;  $\theta$  为参数矩阵; 下标 F 代表 Frobenius 矩阵范数。

### 3 试验与分析

#### 3.1 数据集与评估标准

本试验选用目前使用广泛的远程监督数据集纽约时报 (New York Times, NYT) 数据集和谷歌 IISc 远程监督 (google IISc distant supervision, GIDS) 数据集。NYT 数据集经过知识库 Freebase 中关系和关系实例的对齐处理。知识库 Freebase 由形式化语言表示的实体和实体关系所构成的大量三元组组成。NYT 数据集作为远程监督关系抽取任务常用数据集, 其训练集中共存在 53 类实体关系, 包含 570 088 个实例和 293 162 个实体对, 测试集包含 172 448 个实例和 96 678 个实体对。GIDS 数据集通过扩展 Google 关系提取语料库开发, 数据集的每个实体对都有实例, 其训练集中存在 5 类关系, 包含 11 297 个实例和 6 498 个实体对, 测试集中包含 5 663 个实例和 3 274 个实体对。NYT 数据集涵盖领域广泛, 经过严格筛选和清除, 研究人员可以利用这些数据进行大规模关系抽取任务, 且 NYT 数据集不仅提供实体对之间的关系标签, 还提供关系类型标签, 能够进行多关系抽取任务, 更好地理解实体之间的关系。因此, NYT 数据集为远程监督关系抽取的典型数据集。

本研究选用 3 类评价指标进行评价, 分别为: 精确率  $P$ 、召回率  $R$ 、 $F_1$  值。3 类评价指标的计算式分别为

$$P = \frac{T_p}{T_p + F_p} \times 100\%, \quad (18)$$

$$R = \frac{T_p}{T_p + F_n} \times 100\%, \quad (19)$$

$$F_1 = \frac{2PR}{P+R} \times 100\%, \quad (20)$$

式中,  $T_p$  为模型正确预测为正例的样本数量,  $F_p$  为模型错误地将负例预测为正例的样本数量,  $F_n$  为模型错误地将正例预测为负例的样本数量。

#### 3.2 试验环境

本试验采用的硬件环境为 Intel Core i7-10700 2.10 GHz 处理器、16 GB 内存、NVIDIA GeForce RTX 3060 显卡, 软件环境为 Windows 11 操作系统、Pycharm 开发环境、Python 3.8 编程语言。经过测试, 模型批量大小设置为 32、学习率设置为 0.01 时,

学习效果最佳。

#### 3.3 远程监督关系抽取方法对比试验结果分析

为验证本研究提出的方法对远程监督关系抽取的有效性, 选用多类别逻辑回归模型 (multiclass logistic regression model, MLRO)<sup>[6]</sup>、分段卷积神经网络 (piece-wise convolutional neural network, PCNN)<sup>[28]</sup>、基于双向门控循环单元的词注意力模型 (bi-directional gated recurrent unit based word attention model, BGWA)<sup>[29]</sup>、基于实体特征增强和注意力原型网络模型 (entity feature enhancement and attention-based network model, EFEAPN)<sup>[24]</sup>、图中神经网络中基于知识图谱的关系提取模型 (relation extraction using knowledge graph context in a graph neural network, RECON)<sup>[30]</sup>、基于远程监督生成对抗网络和知识图谱图注意力网络模型 (distant supervised generative adversarial network and knowledge graph graph attention network model, DSGAN-KGGAT)<sup>[31]</sup>、基于 RoBERTa 实体嵌入和图注意力网络增强句子表示的关系抽取模型 (RoBERTa entity embedding and graph attention networks enhanced sentence representation for relation extraction, REEGAT)<sup>[32]</sup> 进行比较, 试验结果如表 1、2 所示。

表 1 NYT 数据集上模型试验结果对比  
Table 1 Comparison of model experimental results in NYT dataset 单位: %

模型	$P$	$R$	$F_1$ 值
MLRO	25.82	26.38	25.82
PCNN	57.61	62.15	59.80
BGWA	72.59	69.93	71.24
EFEAPN	74.85	72.39	73.60
RECON	85.49	87.36	86.41
DSGAN-KGGAT	87.15	87.20	87.17
REEGAT	87.72	87.69	87.70
IR_MI_DA	<b>88.63</b>	<b>88.59</b>	<b>88.61</b>

表 2 GIDS 数据集上模型试验结果对比  
Table 2 Comparison of model experimental results in GIDS dataset 单位: %

模型	$P$	$R$	$F_1$ 值
MLRO	30.25	29.41	29.82
PCNN	61.57	68.23	64.73
BGWA	75.61	71.05	73.26
EFEAPN	79.45	77.30	78.36
RECON	85.29	86.72	85.97
DSGAN+KGGAT	86.33	87.94	87.13
REEGAT	88.61	89.50	89.05
IR_MI_DA	<b>89.37</b>	<b>90.91</b>	<b>90.13</b>

MLRO 模型是基于特征模型的远程监督关系抽取方法,在进行特征选择和自然语言处理工具处理数据时存在误差传播问题,捕捉深层次语义信息及泛化能力较弱,不适用于大规模数据的关系抽取。

PCNN、BGWA、EFEAPN、RECON、DSGAN-KGGAT、REEGAT 模型均为基于神经网络的远程监督关系抽取方法,神经网络可以自动提取数据特征,不受先验知识影响,提高关系抽取性能。PCNN 模型采用 Word2vec 方法进行向量编码,将句子分割成 3 部分进行卷积操作,更好地提取文本特征,相较于 MLRO 模型,在 NYT、GIDS 数据集上  $F_1$  值均提升 30 个百分点以上。BGWA 模型在 PCNN 模型的基础上添加 BiLSTM 模型和实体注意力机制,使用加权投票的方法进行特征组合,关系抽取效果进一步提升。相较于 BGWA 模型, EFEAPN 模型添加 RoBERTa,利用 BiLSTM 神经网络进行特征提取,利用基于注意力机制的原型神经网络进行关系预测,在 NYT、GIDS 数据集上  $F_1$  值均有提升,最多提升 5.10 个百分点。RECON、DSGAN-KGGAT、REEGAT 模型都以知识图谱为基础进行特征学习。在知识图谱中,实体属性能够帮助模型更好地学习实体特征,增强实体语义表示。相较于未融合知识图谱上下文信息的模型,融合后的模型远程监督关系抽取效果进一步提升。RECON 模型虽然学习实体的属性信息,但没有考虑实体与实体属性之间的关联,不能调整实体特征的重要性。DSGAN-KGGAT 模型利用知识图谱学习实体特征,采用图注意力机制为实体信息分配权重,使特征向量更加合理,与 RECON 模型相比,在 NYT、GIDS 数据集上  $F_1$  值均有提升,最多提升 1.16 个百分点。REEGAT 模型采用 RoBERTa 动态获取文本向量,融合图神经网络和多头自注意力机制对词向量和关系向量加权,进一步丰富向量含义,与 DSGAN-KGGAT 模型相比,在 NYT、GIDS 数据集上  $F_1$  值均有提升,最多提升 1.92 个百分点。图注意力机制通常考虑全局信息计算节点之间的注意力权重,每个节点表示都依赖整个图的信息,而不仅仅依赖其相邻节点。这种全局依赖在处理大规模图数据时可能变得困难。本研究采用的双重注意力机制不依赖整个文本的全局信息;词语级注意力机制在计算实体对之间的相关程度时,只需要考虑包含此实体对句子的上下文信息;句子级注意力机制赋予包含实体对句子与预测关系匹配度高的句子更多权重,

以此降低噪声数据的影响。此外,本研究模型利用改进 RoBERTa 学习文本的词向量特征,深度挖掘文本信息的潜在特征,增强模型的学习表示能力。相较于 REEGAT 模型,在 NYT、GIDS 数据集上,本研究模型的  $F_1$  值均有提升,最多提升 1.08 个百分点。

### 3.4 消融试验

为验证不同预训练模型、多实例学习、不同注意力机制对远程监督关系抽取效果的影响,本研究采取以下 3 组试验进行验证。

#### 3.4.1 不同预训练模型对远程监督关系抽取效果的影响

为验证 IR\_MI\_DA 有助于提升远程监督关系抽取效果,对比深度双向语言模型(embeddings from language models, ELMo)、BERT、RoBERTa 及改进的 RoBERTa。试验结果如表 3、4 所示。

表 3 NYT 数据集上不同预训练模型试验结果对比  
Table 3 Comparison of experimental results with different pre-trained models in NYT dataset 单位:%

模型	$P$	$R$	$F_1$ 值
ELMo_MI_DA	80.35	83.62	81.92
BERT_MI_DA	83.35	85.62	84.47
RoBERTa_MI_DA	86.25	87.39	86.82
IR_MI_DA	<b>88.63</b>	<b>88.59</b>	<b>88.61</b>

注:ELMo\_MI\_DA 为使用 ELMo、多实例学习和双重注意力机制的模型;BERT\_MI\_DA 为使用 BERT、多实例学习和双重注意力机制的模型;RoBERTa\_MI\_DA 为使用 RoBERTa、多实例学习和双重注意力机制的模型。

表 4 GIDS 数据集上不同预训练模型试验结果对比  
Table 4 Comparison of experimental results with different pre-trained models in GIDS dataset 单位:%

模型	$P$	$R$	$F_1$ 值
ELMo_MI_DA	81.63	82.57	82.10
BERT_MI_DA	85.94	86.68	86.31
RoBERTa_MI_DA	87.00	87.34	87.17
IR_MI_DA	<b>89.37</b>	<b>90.91</b>	<b>90.13</b>

试验结果表明,本研究模型在使用改进的 RoBERTa 后,关系抽取效果显著提升,在 NYT 数据集上, $P$ 、 $R$ 、 $F_1$  值分别为 88.63%、88.59%、88.61%,在 GIDS 数据集上, $P$ 、 $R$ 、 $F_1$  值分别为 89.37%、90.91%、90.13%。ELMo 使用向量拼接方式融合上下文特征,只在最后一层长短期记忆网络上做拼接,特征提取和融合效果较差;BERT 利用双向特性解决 ELMo 特征融合较差的问题,将文本转换为高维向量,相较于 ELMo,在 NYT、GIDS 数据集上  $F_1$  值均提升 2 个百分点以上;RoBERTa 改进 BERT 的静态掩码策略,采用动态掩码技术,使模型全方位学习输入层的文本向量,获得比 BERT 更强的表征学

习能力,相较于 BERT,在 NYT、GIDS 数据集上  $F_1$  值均有提升,最多提升 2.35 百分点。改进后的 RoBERTa 在 RoBERTa 的基础上添加全词掩码技术,得到当前词在上下文语境的动态语义,同时赋予词在句子中的位置信息,加强词的表征能力,相较于 RoBERTa,在 NYT、GIDS 数据集上  $F_1$  值均有提升,最多提升 2.96 百分点,显著提高模型在远程监督关系抽取的性能。

### 3.4.2 多实例学习对远程监督关系抽取的影响

为验证多实例学习对远程监督关系抽取的有效性,本研究在没有添加多实例学习 (IR\_DA) 和添加多实例学习 (IR\_MI\_DA) 的模型上进行对比试验,试验结果如表 5、6 所示。

表 5 NYT 数据集上多实例学习试验结果对比  
Table 5 Comparison of experimental results of multiple instance learning in NYT dataset 单位:%

模型	$P$	$R$	$F_1$ 值
IR_DA	88.27	88.15	88.21
IR_MI_DA	<b>88.63</b>	<b>88.59</b>	<b>88.61</b>

表 6 GIDS 数据集上多实例学习试验结果对比  
Table 6 Comparison of experimental results of multiple instance learning in GIDS dataset 单位:%

模型	$P$	$R$	$F_1$ 值
IR_DA	88.78	89.61	89.19
IR_MI_DA	<b>89.37</b>	<b>90.91</b>	<b>90.13</b>

从表 5、6 中可以看出:添加多实例学习后,在 NYT 数据集上, $P$ 、 $R$ 、 $F_1$  值分别提升 0.36、0.44、0.40 百分点;在 GIDS 数据集上, $P$ 、 $R$ 、 $F_1$  值分别提升 0.59、1.30、0.94 百分点。结果表明,多实例学习能够利用实体对的多个实例,从不同上下文中学习关系特征,丰富关系特征表示,缩小关系类别范围,减少噪声数据的影响。使用多实例学习相较未使用多实例学习,关系抽取效果有一定提升。

### 3.4.3 不同注意力机制对远程监督关系抽取效果的影响

为验证提出的双重注意力机制对远程监督关系抽取的有效性,本研究在没有添加注意力机制 (IR\_MI)、添加单层句子级注意力机制 (IR\_MI\_A) 和添加词语、句子双重注意力机制 (IR\_MI\_DA) 上进行对比试验,试验结果如表 7、8 所示。

表 7 NYT 数据集上不同注意力机制试验结果对比  
Table 7 Comparison of experimental results of attention mechanism ablation in NYT dataset 单位:%

模型	$P$	$R$	$F_1$ 值
IR_MI	82.95	83.72	83.33
IR_MI_A	84.87	85.07	84.97
IR_MI_DA	<b>88.63</b>	<b>88.59</b>	<b>88.61</b>

表 8 GIDS 数据集上不同注意力机制试验结果对比  
Table 8 Comparison of experimental results of attention mechanism ablation in GIDS dataset 单位:%

模型	$P$	$R$	$F_1$ 值
IR_MI	83.90	85.37	84.63
IR_MI_A	86.41	88.04	87.21
IR_MI_DA	<b>89.37</b>	<b>90.91</b>	<b>90.13</b>

由表 7、8 可知,本研究模型在结合词语级和句子级双重注意力机制后,在 NYT、GIDS 数据集上  $F_1$  值均达到最高,分别为 88.61%、90.13%。建立句子级注意力机制能够提高模型对有效语句的关注度,为正确标注的文本赋予更高的权重,尽可能衰减错误标注样本的信息,减少错误标注带来的噪声问题,在 NYT、GIDS 数据集上  $F_1$  值比未使用注意力机制的模型均有提升,最多提升 2.58 百分点。在双重注意力机制中,添加词语级注意力机制能使模型在学习文本的隐藏特征时综合考虑长距离的句法依赖,捕获句子中关键词的语义信息,增强句子的语义表征能力,因此,结合句子级注意力机制能够提高句子预测关系的准确性。相较于不使用注意力机制、单独使用句子级注意力机制,在 NYT、GIDS 数据集上  $F_1$  值均提升 2 个百分点以上,说明了双重注意力机制对远程监督关系抽取的有效性。

## 4 结论

针对远程监督关系抽取带来的噪声标注问题,本研究提出远程监督关系抽取方法 IR\_MI\_DA。该模型利用改进的 RoBERTa 和 BiGRU 对文本向量进行深层次特征提取;多实例学习通过学习实例级别的特征,缩小关系类别范围;双重注意力机制中,词语级注意力机制能够捕获实体上下文特征信息,提高对关系分类起决定作用的词权重,句子级注意力机制能够学习实体共现句的上下文特征,为每个句子学习注意力权重。结果表明,在 NYT、GIDS 数据集上,本研究提出的模型  $P$ 、 $R$ 、 $F_1$  值均达到最高,能够有效降低远程监督的噪声影响,更好地实现关系抽取,为构建领域知识图谱提供参考。

### 参考文献:

[1] 袁泉,陈昌平,陈泽,等.基于 BERT 的两次注意力机制远程监督关系抽取[J].计算机应用,2024,44(4):1080-1085.  
YUAN Quan, CHEN Changping, CHEN Ze, et al. Twice attention mechanism distantly supervised relation extraction based on BERT [J]. Journal of Computer

- Applications, 2024, 44(4): 1080-1085.
- [2] GRISHMAN R. Information extraction: techniques and challenges [C]//Information Extraction A Multidisciplinary Approach to an Emerging Information Technology. Berlin, Germany: Springer, 1997: 10-27.
- [3] 刘峤, 李杨, 段宏, 等. 知识图谱构建技术综述[J]. 计算机研究与发展, 2016, 53(3): 582-600.  
LIU Qiao, LI Yang, DUAN Hong, et al. Knowledge graph construction techniques [J]. Journal of Computer Research and Development, 2016, 53(3): 582-600.
- [4] 王传栋, 徐娇, 张永. 实体关系抽取综述[J]. 计算机工程与应用, 2020, 56(12): 25-36.  
WANG Chuandong, XU Jiao, ZHANG Yong. Survey of entity relation extraction [J]. Computer Engineering and Applications, 2020, 56(12): 25-36.
- [5] 李枫林, 柯佳. 基于深度学习框架的实体关系抽取研究进展[J]. 情报科学, 2018, 36(3): 169-176.  
LI Fenglin, KE Jia. Research progress of entity relation extraction base on deep learning framework[J]. Information Science, 2018, 36(3): 169-176.
- [6] MINTZ M, BILLS S, SNOW R, et al. Distant supervision for relation extraction without labeled data [C]//Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP. Singapore:ACL, 2009: 1003-1011.
- [7] BONAN M, RALPH G, LI W. Distant supervision for relation extraction with an incomplete knowledge base [C]//Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Atlanta, USA: ACL, 2013: 777-782.
- [8] 郑志蕴, 徐亚媚, 李伦, 等. 融合位置特征注意力与关系增强机制的远程监督关系抽取[J]. 小型微型计算机系统, 2023, 44(12): 2678-2684.  
ZHENG Zhiyun, XU Yamei, LI Lun, et al. Distantly supervised relation extraction with position feature attention and relation enhancement[J]. Journal of Chinese Computer Systems, 2023, 44(12): 2678-2684.
- [9] 张欢, 李卫疆. 基于类型注意力和 GCN 的远程监督关系抽取 [J]. 计算机工程与科学, 2024, 46(2): 316-324.  
ZHANG Huan, LI Weijiang. Distant supervision relation extraction based on type attention and GCN[J]. Computer Engineering & Science, 2024, 46(2): 316-324.
- [10] 崔仕林, 闫蓉. 基于 SoftLexicon 和注意力机制的中文因果关系抽取 [J]. 中文信息学报, 2023, 37(4): 81-89.  
CUI Shilin, YAN Rong. Chinese causality extraction based on SoftLexicon and attention mechanism[J]. Journal of Chinese Information Processing, 2023, 37(4): 81-89.
- [11] 李浩, 刘永坚, 解庆, 等. 基于多层次注意力机制的远程监督关系抽取模型 [J]. 计算机科学, 2019, 46(10): 252-257.  
LI Hao, LIU Yongjian, XIE Qing, et al. Distant supervision relation extraction model based on multi-level attention mechanism [J]. Computer Science, 2019, 46(10): 252-257.
- [12] WEI Q, JI Z C, SI Y Q, et al. Relation extraction from clinical narratives using pre-trained language models [C]//American Medical Informatics Association Annual Symposium. Washington, USA: AMIA, 2019: 1236-1245.
- [13] SU P, VIJAY-SHANKER K. Investigation of improving the pre-training and fine-tuning of BERT model for biomedical relation extraction [J]. BMC Bioinformatics, 2022, 23(1): 120.
- [14] FENG P, ZHANG X, ZHAO J, et al. Relation extraction based on prompt information and feature reuse [J]. Data Intelligence, 2023, 5(3): 824-840.
- [15] FAN C Y. The entity relationship extraction method using improved RoBERTa and multi-task learning [J]. Computers, Materials & Continua, 2023, 77(2): 1719-1738.
- [16] YE Q, CAI T T, JI X, et al. Subsequence and distant supervision based active learning for relation extraction of Chinese medical texts [J]. BMC Medical Informatics and Decision Making, 2023, 23(1): 34.
- [17] 张鲁, 段友祥, 刘娟, 等. 基于 RoBERTa 和加权图卷积网络的中文地质实体关系抽取 [J]. 计算机科学, 2024, 51(8): 297-303.  
ZHANG Lu, DUAN Youxiang, LIU Juan, et al. Chinese geological entity relation extraction based on RoBERTa and weighted graph convolutional networks [J]. Computer Science, 2024, 51(8): 297-303.
- [18] ZHOU G D, SU J, ZHANG J, et al. Exploring various knowledge in relation extraction [C]//Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics. Ann Arbor, USA: ACL, 2005: 427-434.
- [19] JIANG X, WANG Q, LI P, et al. Relation extraction with multi-instance multi-label convolutional neural networks [C]//Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics. Osaka, Japan: COLING, 2016: 1471-1480.
- [20] 刘哲. 基于句子级注意力机制的远程监督实体关系抽取 [D]. 南昌: 江西财经大学, 2021.

- LIU Zhe. Distant supervised entity relationship extraction based on sentence-level attention mechanism[D]. Nanchang: Jiangxi University of Finance and Economics, 2021.
- [21] YUAN Y, LIU L, TANG S, et al. Cross-relation cross-bag attention for distantly-supervised relation extraction [C]//Proceedings of the Thirty-Third AAAI Conference on Artificial Intelligence and Thirty-First Innovative Applications of Artificial Intelligence Conference and Ninth AAAI Symposium on Educational Advances in Artificial Intelligence. Honolulu, USA: AAAI, 2019: 419-426.
- [22] 王红, 李晗, 李浩飞. 民航突发事件领域本体关系抽取方法的研究[J]. 计算机科学与探索, 2020, 14(2): 285-293.
- WANG Hong, LI Han, LI Haofei. Research of relation extraction method of civil aviation emergency domain ontology[J]. Journal of Frontiers of Computer Science and Technology, 2020, 14(2): 285-293.
- [23] ZHANG J, CAO M L. Distant supervision for relation extraction with hierarchical attention-based networks[J]. Expert Systems with Applications, 2023, 220: 119727.
- [24] LI R, XIAO Q, YANG J X, et al. Few-shot relation extraction via the entity feature enhancement and attention-based prototypical network[J]. International Journal of Intelligent Systems, 2023, 1: 1186977.
- [25] ZHAI Z W, FAN R L, HUANG J, et al. A novel joint extraction model based on cross-attention mechanism and global pointer using context shield window[J]. Computer Speech & Language, 2024, 87: 101643.
- [26] HUANG S J, CHEN Y, ZHOU E J, et al. A RoBERTa based model for identifying non-substantive factual elements of the case [C]// 2021 2nd International Conference on Intelligent Computing and Human-Computer Interaction (ICHCI). Shenyang, China: IEEE, 2021: 65-71.
- [27] ZHU Z D, SU J D, HONG X B. Improving relation extraction using semantic role and multi-task learning [C]//Proceedings of the Knowledge Graph and Semantic Computing: Knowledge Graph and Cognitive Intelligence. Singapore: Springer, 2021: 93-105.
- [28] ZENG D, LIU K, CHEN Y, et al. Distant supervision for relation extraction via piecewise convolutional neural networks[C]//Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing. Lisbon, Portugal: ACL, 2015: 1753-1762.
- [29] JAT S, KHANDELWAL S, TALUKDAR P. Improving distantly supervised relation extraction using word and entity based attention [EB/OL]. (2018-04-19) [2024-03-05]. <https://arxiv.org/abs/1804.06987>
- [30] BASTOS A, NADGERI A, SINGH K, et al. RECON: relation extraction using knowledge graph context in a graph neural network [C]//Proceedings of the Web Conference 2021. Ljubljana, Slovenia: ACM, 2021: 1673-1685.
- [31] 赵晋斌, 王琦, 马黎雨, 等. 基于知识图谱的远程监督关系抽取降噪方法[J]. 火力与指挥控制, 2023, 48(10): 160-169.
- ZHAO Jinbin, WANG Qi, MA Liyu, et al. A noise reduction method for distant supervision relation extraction based on knowledge graph[J]. Fire Control & Command Control, 2023, 48(10): 160-169.
- [32] CAI F Z, HU Q, ZHOU R J, et al. REEGAT: RoBERTa entity embedding and graph attention networks enhanced sentence representation for relation extraction [J]. Electronics, 2023, 12(11): 2429.

(编辑:孙亚彤)