

块的模型 OneRel, 该模型包括特定关系角标注策略和基于分数的分类器两部分. 该模型先使用标注策略进行标注, 并利用分类器进行打分, 最后得到实体关系三元组, 该模型仍然存在关系冗余和泛化能力不足的问题.

3 RGPNRE 模型

3.1 模型的结构及训练流程介绍

RGPNRE 模型不仅能够提取出头实体和尾实体, 还可以同时预测它们之间的关系, 从而一次性抽取三元组. 模型主要包含 4 部分: RoBERTa 编码器、实体的全局指针^[2]、关系预测模块^[7]及特定关系的全局指针^[2], 整体结构如图 1 所示.

模型使用 RoBERTa 预训练模型及全局指针来进行实体关系抽取. 首先, 将待抽取的句子 S , 如 S 为“久坐伤害身体, 散步改善身体”, 输入到 RoBERTa 编码器

中进行正常训练, 得到词嵌入表示后, 然后在其中加入干扰 r_{adv} , 进行对抗训练 (详见第 3.3 节对抗训练), 得到句子的语义表示 H . 接着, 利用实体的全局指针来得到头实体开头到结尾位置的分数 $P_{(s_head, s_tail)}$ 和尾实体开头到结尾位置的分数 $P_{(o_head, o_tail)}$. 随后对句子 S 中可能存在的关系进行预测, 得到一组潜在的关系集合 $R_{potential}$. 接下来, 遍历关系集合 $R_{potential}$ 中的所有关系, 将集合中的每个关系 r 放入到特定关系的全局指针, 得到针对每个关系 r , 对头实体开始位置到尾实体开始位置和头实体结束位置到尾实体结束位置的打分 $P_{(s_head, o_head|r)}$ 和 $P_{(s_tail, o_tail|r)}$. 最后, 预测出所有可能存在的组合, 如 (久坐, 伤害, 身体)、(久坐, 改善, 身体)、(散步, 伤害, 身体) 等, 并输出其中满足 $P_{(s_head, s_tail)}$ 、 $P_{(o_head, o_tail)}$ 、 $P_{(s_head, o_head|r)}$ 和 $P_{(s_tail, o_tail|r)}$ 均大于 0 的部分. 然后取它们的交集作为最后的输出, 即 (久坐, 伤害, 身体) 和 (散步, 改善, 身体).

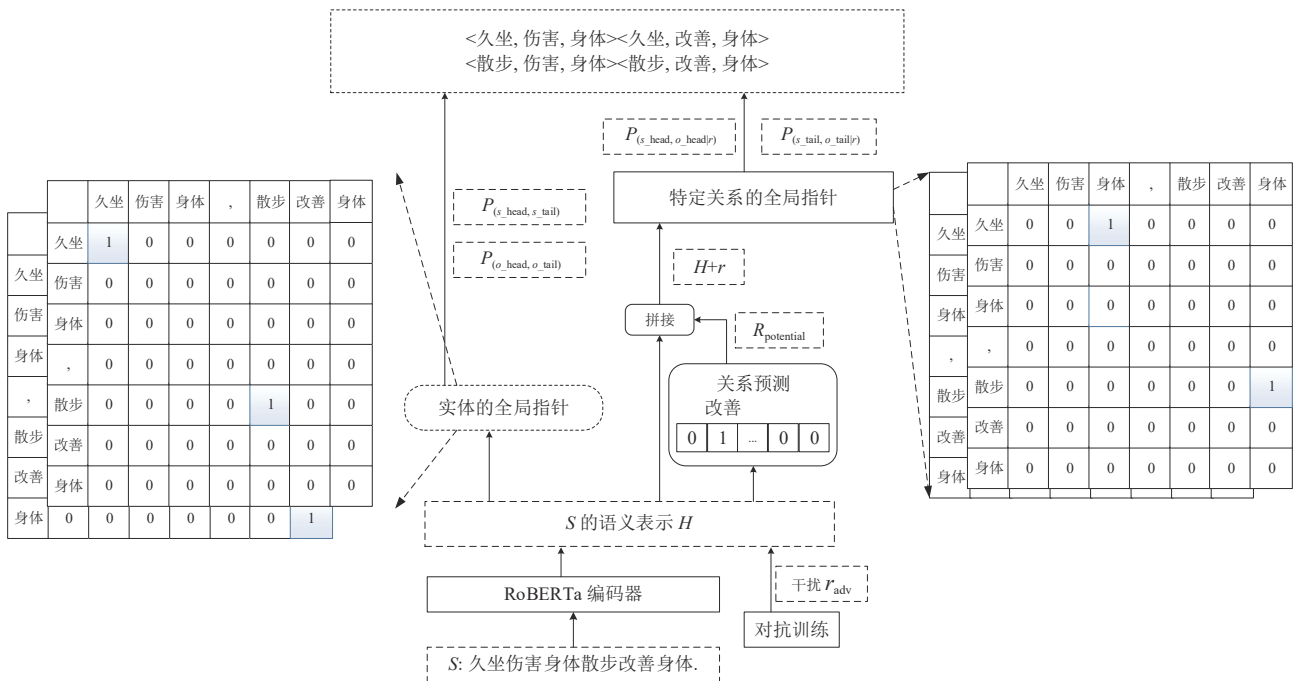


图 1 RGPNRE 模型体系结构图

具体的训练策略是在训练时, 让标注的五元组 ($S_{head}, S_{tail}, r, O_{head}, O_{tail}$) 的 $P_{(s_head, s_tail)}$ 、 $P_{(o_head, o_tail)}$ 、 $P_{(s_head, o_head|r)}$ 和 $P_{(s_tail, o_tail|r)}$ 均大于 0, 其余的则均小于 0. 预测时枚举所有可能的五元组, 输出 $P_{(s_head, s_tail)}$ 、 $P_{(o_head, o_tail)}$ 、 $P_{(s_head, o_head|r)}$ 和 $P_{(s_tail, o_tail|r)}$ 均大于 0 的部分, 接着取它们的交集作为最终的输出.

3.2 RoBERTa 编码器

Liu 等人^[1]提出了 RoBERTa 模型, 该模型能够得到词嵌入, 它是以 BERT 模型^[11]为基础的优化版本, 经过更长时间的训练及使用更多和更长的训练序列和数据而得到. 该模型采用双向 Transformer^[12], 可以获取上下文特征, 拥有更强大的表达信息的能力. 相较于 BERT

以字为单位进行切分, RoBERTa 模型以词为单位进行切分, 使其在中文语料上获得了更好的效果。

RoBERTa 模型经过改进, 可以得到不同种类的预训练模型, 如 RoBERTa-mid、RoBERTa-zh-base 和 RoBERTa-zh-Large 等。而本文因为是针对中文进行实体关系抽取, 所以选择了 RoBERTa-zh-Large 作为预训练模型。该模型对输入的句子进行编码, 将 Transformer 编码结构定义为 $Trans(x)$ 。它的输入舍弃了下一段预测任务, 其他的方面和 BERT 模型相同, 包含词嵌入、片段嵌入和位置嵌入。接着, 输入向量 E 经过多层 Transformer 网络, 得到文本向量表示 H_N 。最终, H_N 被赋予后续步骤的输入向量, 即 $H=H_N$ 。 H_N 的计算如式 (1) 和式 (2) 所示:

$$H_1 = Trans(E) \quad (1)$$

$$H_N = Trans(H_{N-1}) \quad (2)$$

3.3 对抗训练

2013 年, Bruna 等人^[13]首次提出了对抗样本的概念, 即在原始输入数据中添加微小的干扰, 以使模型犯错误, 导致可能输出错误的结果。为了探究对抗样本产生的原因, Goodfellow 等人^[14]基于神经网络线性特点提出了线性假设解释对抗样本的存在。在 2015 年, 他首次提出了对抗训练的概念, 即在原始输入 x 上添加一个干扰 r_{adv} 来生成对抗样本, 并使用这些对抗样本进行训练。以前的研究理论部分比较依赖直觉, 但 Madry 等人^[15]在 2018 年总结了以前的工作并提出了一种高鲁棒性方案, 即最小化外部风险、最大化内部损失。这种方法明显增强了模型的防御能力。

上述工作主要是针对计算机视觉领域展开的。直到 2017 年, Kurakin 等人^[16]在原先提出的 FGSM 算法的基础上进行了改进, 得到了针对自然语言处理领域的 FGM 算法, 并根据具体的梯度进行处理, 从而快速生成对抗样本。相比于 FGM 的简洁明了, Madry 等人提出了 PGD 算法, 该算法进行多次迭代, 以求得最优的扰动。Zhu 等人^[17]基于 PGD 算法, 提出了类似的算法 FreeLB, 这个算法通过扩大训练数据并执行多次迭代操作来提高鲁棒性和泛化能力。最近 Jiang 等人^[18]提出了 SMART, 它采用对抗正则化实现对抗学习, 同时提出一个新的目标函数, 并提出了优化算法来解决它。

本文使用的是 FGM 算法。该算法通过引入一定的干扰, 生成了模型很容易预测错误的样本, 通过对这些

样本进行大量训练, 提高了模型的泛化能力。它的主要流程如下。

1) 首先正常地计算前向传播损失, 然后反向传播推导梯度。

2) 得到嵌入层的梯度, 根据公式计算出 r_{adv} , 然后将它加到原始嵌入样本中, 以生成对抗样本。

3) 通过使用对抗本来计算新的损失, 并反向传播计算对抗样本梯度并加到步骤 1 的梯度上。

4) 恢复原始嵌入状态。

5) 使用步骤 3 中获得的梯度来更新模型参数。

计算干扰的相关公式如式 (3) 和式 (4) 所示:

$$g = \nabla_x L(\theta, x, y) \quad (3)$$

$$r_{adv} = \epsilon \cdot \frac{g}{\|g\|_2} \quad (4)$$

其中, g 是嵌入层的梯度, r_{adv} 是干扰。

3.4 实体的全局指针

模型使用实体的全局指针方法来抽取实体关系三元组。该方法将实体的开始、结尾和实体类型视作一个整体进行判断。因为传统的指针网络在命名实体识别任务中通常由两个模块分别识别实体的开始和结尾。这种传统方式的问题在于, 训练和预测的过程中使用了不同的信息, 即预测时只能根据上下文信息预测开始和结束位置, 没有全局的视角。

实体的全局指针将命名实体识别任务分成了提取和分类两个子任务。在前者中, 提取文本片段作为实体, 在后者中, 识别实体的类型。从这个角度来看, 提取可以被视为只有一种实体类型的命名实体识别任务。因此本文使用注意力机制和相对位置编码 RoPE^[19]设计了评分矩阵, 引入 RoPE 的目的是提高模型对实体长度和跨度的敏感度。这样做的目的是使模型能够更好地区分真正的实体。

具体步骤如下 (详见式 (5)–式 (10))。

给定一个包含 n 个词的句子 $X = [x_1, x_2, \dots, x_n]$, 需要通过 RoBERTa() 对其进行处理。将原始输入转化为新矩阵 $H \in R^{n \times d}$ 的编码表示, 其中 d 为表示维度。具体而言, H 由每个单词的编码 h_i 表示。 w_q 、 b_q 、 w_k 和 b_k 分别是查询向量和键值向量对应的权重和偏差, $q_i \in R^{1 \times d}$ 是第 i 个词的查询向量, $k_i \in R^{1 \times d}$ 是第 i 个词的键值向量, $[q_i; k_i; q_j; k_j]$ 是片段表示, $\omega_\alpha \in R^{4v}$ 是可以训练的权重, $s_{i,j}^{\text{extract}}$ 和 $s_{i,j}^{\text{classify}}$ 分别表示对从 i 到 j 的实体抽取进

行打分和对从 i 到 j 的实体是否为 α 类型进行打分, 两个得分相加可以得到对 α 类型的实体的总分数 $s_{i,j}^\alpha$.

$$h_1, h_2, \dots, h_n = \text{RoBERTa}(x_1, x_2, \dots, x_n) \quad (5)$$

$$q_i = w_q h_i + b_q \quad (6)$$

$$k_i = w_k h_i + b_k \quad (7)$$

$$s_{i,j}^{\text{extract}} = q_i^T + k_j \quad (8)$$

$$s_{i,j}^{\text{classify}} = \omega_\alpha^T [q_i; k_i; q_j; k_j] \quad (9)$$

$$s_{i,j}^\alpha \equiv s_{i,j}^{\text{extract}} + s_{i,j}^{\text{classify}} \quad (10)$$

3.5 关系预测模块

关系预测模块的主要任务是得到句子可能存在的关系集合 $R_{\text{potential}}$. 该模块的输入是经过 RoBERTa 和对抗训练处理后得到的语义向量 H , 输出是潜在的关系集合 $R_{\text{potential}}$. 公式如下:

$$h_{\text{avg}} = \text{Avgpool}(h) \in R^{d \times 1} \quad (11)$$

$$P_i = \sigma(W_i h_{\text{avg}} + b_i) \quad (12)$$

其中, Avgpool 代表平均池化操作, W_i 代表权重, b_i 代表偏差, 而 σ 代表 Sigmoid 函数. 本文将关系预测视为多标签的二分类任务. 如果某个关系对应的概率超过阈值 α (设为 0.5), 则为其分配标签 1; 否则分配标签 0. 因此, 在接下来的特定关系的全局指针络中, 只需要使用标签为 1 的关系, 而不需要预测所有关系.

3.6 特定关系的全局指针

特定关系的全局指针在前面步骤获得的关系集合 $R_{\text{potential}}$ 进行处理, 遍历关系集合 $R_{\text{potential}}$ 里面的每种关系, 针对每种关系 R , 进行头实体和尾实体开始位置以及头实体和尾实体结束位置的打分预测. 选择这两个位置是为了确保头实体和尾实体之间没有嵌套实体. 由于这个模型需要针对不同种类的关系预测实体位置, 因此在每个关系的语义表示中都加入了相应的关系向量. 具体公式如下:

$$h_r = H + R_i \quad (13)$$

其中, H 是通过 RoBERTa 和对抗训练后得到的句子语义表示, $R_i \in R^{n \times d}$ 是对于某种特定关系 r 的矩阵, 将 R_i 加入句子的语义表示 H 中, 得到新的特定关系 r 的句子语义表示 H_r . 接下来的评分方法与全局关联评分方

法相同, 只不过在这里我们预测的是头实体和尾实体的开始位置和结束位置.

3.7 损失函数

模型采用稀疏版的多标签交叉熵作为损失函数. 通常情况下, 多标签分类被视为多个二元分类问题来处理. 然而, 该方法可能会导致类别分布不均衡, 从而影响实验结果. 因此在训练实体关系联合抽取模型时, 采用交叉熵来解决多标签分类任务中的类别不均衡现象. 另外, 如果使用多向量标记正负类别, 每一步训练都需要创建一个巨大的矩阵, 并将其传入 GPU, 这将导致昂贵的创建和传输成本. 因此, 本文选择实现稀疏版的多标签交叉熵, 即只传递正类对应的下标, 因为正类数目较少, 所以可以大大减少矩阵的尺寸. 稀疏版的损失函数实现如下:

$$L = \log \left(1 + \sum_{i \in P} e^{-S_i} \right) + \log \left(1 + \sum_{i \in N} e^{S_i} \right) \quad (14)$$

$$\log \left(1 + \sum_{i \in N} e^{S_i} \right) = \log \left(1 + \sum_{i \in A} e^{S_i} \right) + \log \left(1 - \frac{\sum_{i \in P} e^{S_i}}{\sum_{i \in A} e^{S_i}} \right) \quad (15)$$

其中, P 、 N 和 A 分别是正类、负类和全部集合, S_i 是对 i 进行打分. 为了书写方便简洁, 令 $\log \left(1 + \sum_{i \in A} e^{S_i} \right) = a$, $\log \left(1 - \frac{\sum_{i \in P} e^{S_i}}{\sum_{i \in A} e^{S_i}} \right) = b$, 可以得到损失函数实现方式的简洁写法:

$$L = a + b + \log(1 - e^{b-a}) \quad (16)$$

4 实验与分析

4.1 实验数据集

实验使用基于 schema 的中文医学信息抽取数据集 CMelE (Chinese medical information extraction)^[20], 该数据集包括两个训练语料: 一个是儿科训练语料, 涵盖了 518 种儿科疾病; 另一个是百种常见疾病训练语料, 包含 109 种常见疾病. 该数据集总共约有 7.5 万个三元组数据, 其中大约有 2.8 万个关于疾病的句子, 以及 53 种定义好的关系类型. 虽然该数据集已经相对完备, 但仍需进行预处理才能直接使用. 经过预处理后得到最终使用的整体数据集, 包括 14339 条用于训练的句子以及 1421 条用于测试的句子. 为了进一步验证本

文方法在重叠关系中的抽取能力,根据 CasRel 里面重叠三元组的定义,将整体数据集中的句子进一步划分为正常情况 (Normal)、单实体重叠 (single entity overlap, SEO) 和实体对重叠 (entity pair overlap, EPO) 这 3 种类型。这里的正常情况是指不存在实体重叠。实体重叠类型信息的详细情况如表 1 所示。

表 1 数据集详情

| 类型 | 数据集 | 测试集 |
|--------|------|-----|
| Normal | 5506 | 549 |
| EPO | 167 | 13 |
| SEO | 8806 | 869 |

4.2 对比实验

将 RGPNRE 模型和 CasRel、OneRel、GPLinker 等几个关系抽取基准模型进行了比较。

1) CasRel^[7]: 一种经过改进和优化的实体关系联合抽取模型,首次定义了实体重叠问题,并给出了解决方法,在传统方法基础上实现了显著提升。

2) OneRel^[10]: 目前最先进的模型。该模型提出了以往未曾提出的特定标注策略,对文本进行标记,来帮助抽取实体关系三元组。

3) GPLinker^[2]: CasRel 模型的作者之一在自然语言处理领域不断探索,积累了足够的素材,并且结合近几年提出的 SOTA 模型的设计,构建的新的实体关系抽取模型。本文提出的模型也是以该模型为基础进行改进的。

在实验中,设置批量大小为 2,即每次处理两个句子进行训练,设置训练轮数为 50,即训练 50 轮。学习率为 $2E-5$,其他参数基本设为默认值。为了保证实验的公平,所有模型的参数设置都是一样的。

本文使用准确率 P (Precision)、召回率 R (Recall) 和 $F1$ 值作为评估指标。在抽取的三元组和真实的三元组完全匹配,即头实体、关系和尾实体必须全部正确时,才认为抽取结果正确。

实验结果如表 2 所示,可以看到 RGPNRE 模型在整体数据集上,准确率相比基准模型提升了至少 4.57%,但是召回率比 OneRel 模型低 2.47%,不过 RGPNRE 模型的 $F1$ 值是最高的,提升了至少 2.01%。说明 RGPNRE 模型的整体效果最好,但是召回率不如 OneRel 模型,还有进一步提升的空间。

主要的原因是由于 CasRel 模型先抽取头实体,再抽取尾实体和关系,这样的分步骤操作会导致误差传

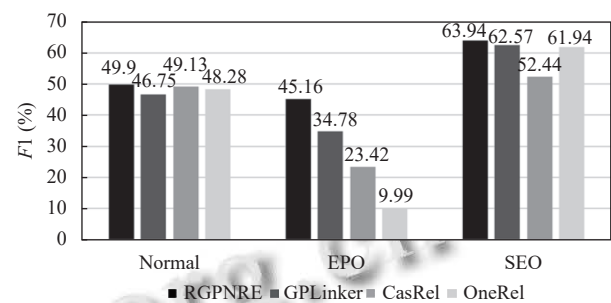
播,从而影响了其 $F1$ 值的表现。OneRel 模型虽然没有采用分步抽取的方法,但它没有考虑到全局的关系,因此 $F1$ 值也会受到影响。GPLinker 模型虽然没有受到以上因素的影响,但其泛化能力较弱。本文提出的 RGPNRE 模型利用 RoBERTa 模型作为编码器增强了对语句的编码能力,并加入对抗训练的方法提升了模型的泛化能力,因而取得了较好的性能。

表 2 CMeIE 数据集上对比实验结果 (%)

| 模型 | P | R | $F1$ |
|----------|--------------|--------------|--------------|
| CasRel | 55.35 | 46.66 | 50.59 |
| OneRel | 59.42 | 57.96 | 58.68 |
| GPLinker | 63.83 | 55.30 | 59.26 |
| RGPNRE | 68.40 | 55.49 | 61.27 |

4.3 实体重叠模式对比实验

为了进一步验证 RGPNRE 模型处理不同的实体重叠类型的能力,将模型分别应用于 Normal、SEO 和 EPO 这 3 个不同实体重叠类型数据集,进行进一步实验,实验结果如图 2 所示。

图 2 RGPNRE 不同重叠类型的 $F1$ 指标对比

由图 2 的实验结果可以发现,在 Normal、SEO 和 EPO 这 3 种实体重叠类型下,RGPNRE 模型相较于 3 种基准模型, $F1$ 值都是最高的。

在 Normal 数据集、SEO 数据集、EPO 数据集上,相比于排在第 2 位的基准模型, $F1$ 值分别提升了 3.15%、10.38% 和 1.37%。由于 EPO 重叠类型的数据量比较小,因此模型之间的差距较大,不一定具有说服力。但是从 SEO 重叠类型的结果来看,RGPNRE 模型能够更好地处理单实体重叠类型的三元组,印证了模型使用的实体全局指针能够很好地处理实体重叠问题。

4.4 不同三元组数量对比实验

本文在 CMeIE 数据集上验证了不同基准模型在从含有不同数量三元组的句子中提取三元组的能力。根据三元组的数量不同,将 CMeIE 数据集中的句子分

为5类:包含1、2、3、4或大于5个三元组.实验结果如表3所示.

表3 CMeIE数据集上 RGP NRE 不同三元组数量的 F1 指标对比

| 模型 | N=1 | N=2 | N=3 | N=4 | N≥5 |
|----------|---------------|---------------|---------------|---------------|---------------|
| RGP NRE | 0.4955 | 0.4865 | 0.4989 | 0.5276 | 0.6789 |
| CasRel | 0.5185 | 0.4339 | 0.4532 | 0.4402 | 0.5305 |
| OneRel | 0.5056 | 0.4888 | 0.5268 | 0.5051 | 0.6495 |
| GPLinker | 0.5022 | 0.4802 | 0.4960 | 0.5018 | 0.6649 |

由表3的实验结果来看,本文提出的模型在处理小于等于3个三元组的句子时能力一般,但在处理大于等于4个三元组的句子时能力有明显提升,这表明本文模型在处理复杂句子场景中有很高的优势.

4.5 消融实验

为了验证本文提出的 RoBERTa 编码器、对抗训练和全局指针模块的有效性,进行了消融实验,表4中的4行分别是本文提出的模型 RGP NRE、本文模型去掉对抗训练后的模型 RG-Adv,本文模型去掉 RoBERTa-zh-Large 编码器而使用 BERT-Base-Chinese 编码器的模型 RG-BERT 以及本文模型去掉全局指针而像 CasRel 模型一样使用两个模块抽取实体的模型 RG-GP 的实验结果.

表4 CMeIE数据集上 RGP NRE 的消融实验指标对比 (%)

| 模型 | P | R | F1 |
|---------|--------------|--------------|--------------|
| RGP NRE | 68.40 | 55.49 | 61.27 |
| RG-Adv | 61.18 | 58.68 | 59.90 |
| RG-BERT | 67.22 | 53.65 | 59.67 |
| RG-GP | 60.25 | 51.16 | 55.33 |

对比表4中第1行和第2行,加入了对抗训练后,准确率和 F1 值有一定提升,但是召回率略有下降.因为进行对抗训练时,产生了大量易出错的样本,而经过这些样本的训练,提升了模型的识别未知样本的准确率,但是同时降低了召回率.对比表4中第1行和第3行,发现不使用 BERT 编码器而改用 RoBERT 编码器后,准确率、召回率和 F1 值都有一定的提升.因为 RoBERT 编码器本就是在 BERT 编码器的基础上进行改进的,并且其以词为单位进行编码能够更好地适应中文数据集.对比第1行和第4行,发现使用全局指针抽取后,大大提升了模型的准确率和召回率和 F1 值.因为全局指针抽取将实体的开始、结尾和类型作为一个整体进行抽取,避免了分头尾抽取可能导致训练和预测过程中使用不同信息,从而产生误差.

5 结束语

为了解决实体关系三元组抽取任务中存在的信息表达能力较差、实体重叠和泛化能力较差等问题,本文提出了一种加入对抗训练的基于 RoBERTa 的全局指针网络实体关系联合抽取模型 RGP NRE.该模型可以更好地理解输入的信息,能够更好地处理实体重叠问题,并拥有更强的泛化能力.实验结果表明,该模型在中文医学文本实体关系抽取数据集 CMeIE 上表现良好,它能够准确地从句子中提取出多个关系三元组,并有效解决了对信息表达能力较差、实体重叠和泛化能力较差的问题.但是仍然存在曝光误差、召回率提升较少、无法很好地处理三元组较少句子等问题,后续会探索新方法来解决这些问题.

参考文献

- Liu YH, Ott M, Goyal N, *et al.* RoBERTa: A robustly optimized BERT pretraining approach. Proceedings of the 2020 International Conference on Learning Representations. Addis Ababa: OpenReview.net, 2020.
- Su JL, Murtadha A, Pan SF, *et al.* Global pointer: Novel efficient span-based approach for named entity recognition. arXiv:2208.03054, 2022.
- Mintz M, Bills S, Snow R, *et al.* Distant supervision for relation extraction without labeled data. 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. Suntec: Association for Computational Linguistics, 2009. 1003–1011.
- Li Q, Ji H. Incremental joint extraction of entity mentions and relations. Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Baltimore: Association for Computational Linguistics, 2014. 402–412.
- Miwa M, Sasaki Y. Modeling joint entity and relation extraction with table representation. Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing. Doha: Association for Computational Linguistics, 2014. 1858–1869.
- Zheng SC, Wang F, Bao HY, *et al.* Joint extraction of entities and relations based on a novel tagging scheme. Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Vancouver: Association for Computational Linguistics, 2017. 1227–1236.

- 7 Wei ZP, Su JL, Wang Y, *et al.* A novel cascade binary tagging framework for relational triple extraction. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. Association for Computational Linguistics, 2020. 1476–1488.
- 8 Wang YC, Yu BW, Zhang YY, *et al.* TPLinker: Single-stage joint extraction of entities and relations through token pair linking. Proceedings of the 28th International Conference on Computational Linguistics. Barcelona: Association for Computational Linguistics, 2020. 1572–1582.
- 9 Zheng HY, Wen R, Chen X, *et al.* PRGC: Potential relation and global correspondence based joint relational triple extraction. 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). Association for Computational Linguistics, 2021. 6225–6235.
- 10 Shang YM, Huang HY, Mao XL. OneRel: Joint entity and relation extraction with one module in one step. Proceedings of the 36th AAAI Conference on Artificial Intelligence. AAAI Press, 2022. 11285–11293.
- 11 Devlin J, Chang MW, Lee K, *et al.* BERT: Pre-training of deep bidirectional Transformers for language understanding. Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers). Minneapolis: Association for Computational Linguistics, 2019. 4171–4186.
- 12 Vaswani A, Shazeer N, Parmar N, *et al.* Attention is all you need. Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach: Curran Associates Inc., 2017. 6000–6010.
- 13 Bruna J, Szegedy C, Sutskever I, *et al.* Intriguing properties of neural networks. Proceedings of the 2nd International Conference on Learning Representations. Banff: OpenReview.net, 2014.
- 14 Goodfellow IJ, Shlens J, Szegedy C. Explaining and harnessing adversarial examples. Proceedings of the 3rd International Conference on Learning Representations. San Diego: ICLR, 2015.
- 15 Madry A, Makelov A, Schmidt L, *et al.* Towards deep learning models resistant to adversarial attacks. Proceedings of the 6th International Conference on Learning Representations. Vancouver: OpenReview.net, 2018.
- 16 Kurakin A, Goodfellow IJ, Bengio S. Adversarial machine learning at scale. Proceedings of the 5th International Conference on Learning Representations. Toulon: OpenReview.net, 2017.
- 17 Zhu C, Cheng Y, Gan Z, *et al.* FreeLB: Enhanced adversarial training for natural language understanding. Proceedings of the 8th International Conference on Learning Representations. Addis Ababa: OpenReview.net, 2020.
- 18 Jiang HM, He CP, Chen WZ, *et al.* SMART: Robust and efficient fine-tuning for pre-trained natural language models through principled regularized optimization. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. Association for Computational Linguistics, 2020. 2177–2190.
- 19 Su JL, Lu Y, Pan SF, *et al.* RoFormer: Enhanced Transformer with rotary position embedding. arXiv:2104.09864, 2021.
- 20 Guan TF, Zan HY, Zhou XB, *et al.* CMelE: Construction and evaluation of Chinese medical information extraction dataset. Proceedings of the 9th CCF International Conference on Natural Language Processing and Chinese Computing. Zhengzhou: Springer, 2020. 270–282.

(校对责编: 孙君艳)