

结合 Roberta 和 Bi-FLASH-SRU 的中文事件因果关系抽取^①



陈泉林, 贾 珺, 樊 硕

(军事科学院 战争研究院, 北京 100091)
通信作者: 贾 珺, E-mail: musicbirdy@sina.com

摘要: 针对现有基于填表的事件关系抽取方法填表数目过多、表格特征获取不充分的问题, 本文提出了结合 Roberta 和 Bi-FLASH-SRU 的中文事件因果关系抽取方法 TF-ChineseERE. 该方法通过制定填表策略, 利用文本中已标记关系, 将其转化为带有标签的表格; 借助 Roberta 预训练模型和本文提出的双向内置闪电注意力简单循环单元 (bidirectional built-in flash attention simple recurrent unit, Bi-FLASH-SRU) 获取主客体事件特征, 再利用表格特征循环学习模块挖掘全局特征, 最后进行表格解码获得事件因果关系三元组. 实验采用金融领域两个公开数据集进行验证, 结果表明, 本文提出的方法 F_1 值分别达到 59.2% 和 62.5%, 且 Bi-FLASH-SRU 模型训练速度更快, 填表数目更少, 证明了该方法的有效性.

关键词: 事件因果关系抽取; Roberta; Bi-FLASH-SRU; 闪电注意力; 填表式抽取

引用格式: 陈泉林, 贾珺, 樊硕. 结合 Roberta 和 Bi-FLASH-SRU 的中文事件因果关系抽取. 计算机系统应用, 2024, 33(6):259-267. <http://www.c-s-a.org.cn/1003-3254/9539.html>

Combining Roberta and Bi-FLASH-SRU for Chinese Event Causality Extraction

CHEN Quan-Lin, JIA Jun, FAN Shuo

(Institute of War, Academy of Military Sciences, Beijing 100091, China)

Abstract: Aiming at the problems of too many forms and insufficient acquisition of form features in the existing form-based event relationship extraction methods, this study proposes TF-ChineseERE, a Chinese event causality extraction method that combines Roberta and Bi-FLASH-SRU. The method transforms the text into labeled forms by formulating a form-filling strategy that takes advantage of the labeled relationships in the text. The study proposes the Roberta pre-training model and the bidirectional built-in flash attention simple recurrent unit (Bi-FLASH-SRU) to obtain the subject-object event features. It then uses the table feature recurrent learning module to mine the global features and finally performs table decoding to obtain event causality triples. The experiments are validated with two public datasets in the financial domain. The results show that the F_1 values of the proposed method reach 59.2% and 62.5%, respectively, with a faster training speed of the Bi-FLASH-SRU model and less number of filled forms, which proves the effectiveness of the method.

Key words: event causality extraction; Roberta; Bi-FLASH-SRU; flash-attention; table filling extraction

随着科技的快速发展, 人类步入了信息化时代, 同步产生的海量文本数据, 形成了天然的“信息富矿”。如

何从复杂繁琐的非结构数据中抽取有价值的结构化信息, 成为研究热点, 由此信息抽取技术应运而生. 事件

① 基金项目: 国家自然科学基金 (6203366); 全军军事类研究生资助课题 (JY2022C251)

收稿时间: 2024-01-05; 修改时间: 2024-02-04; 采用时间: 2024-02-23; csa 在线出版时间: 2024-04-30

CNKI 网络首发时间: 2024-05-07

关系抽取作为信息抽取的重要的一环,其目的是从非结构化文本中抽取三元组形式的知识,并表示为(事件,关系,事件)的形式.事件关系包含时序关系、因果关系、共指关系等,其中因果关系可以实现对文本的深层次的理解.因此,事件因果关系抽取可以在事理图谱、信息检索、情报分析进行广泛的应用.

早期的事件关系抽取主要采用基于特征工程的方法,尽管可以达到很好的效果,但所依赖人工对特征的构建,费时费力^[1-4].随着深度学习模型的推广,涌现出以序列标注、生成式模型、阅读理解和文本切片为基础的事件关系抽取方法,取得了不错的效果,但或多或少存在误差传递和暴露偏差^[5]、对上下文语义获取不充分,消耗更多的计算资源^[6]、难以处理边界模糊文本^[7]以及关系预测时噪声难以消除^[8]等问题.此后有学者提出了基于填表的方法,解决事件关系抽取中的关系重叠问题,消除了误差传递和暴露偏差,但仍存在标注复杂度高、解码效率较慢等问题^[9].

针对基于填表的事件关系抽取方法存在的问题,本文提出了一种结合 Roberta 和 Bi-FLASH-SRU 的中文事件因果关系抽取方法 TF-ChineseERE,贡献点在于:

1) 针对 CNN、RNN 和 LSTM 等网络对上下文语义获取不充分的问题,提出了内嵌闪电注意力简单循环单元,其可以更好地关注词到词之间的依存关系,有效提升了语义理解能力.同时降低了内嵌注意力层的时间、空间复杂度,有效提升了训练效率.

2) 相比于其他填表式抽取方法,本文所设计的填表策略需要填充数目为 $(n^2 - n)R$,有效减少了冗余信息.

3) 本文提出的方法在事件因果关系抽取公开数据集上的实验证明,该方法在 F_1 值上表现较优,可以完成事件因果关系抽取任务.

1 相关工作

早期的关系抽取主要采取模式匹配的方法^[1-4],但需要人工制定规则,费时费力,迁移性差.随着深度学习的发展,使得关系抽取任务取得了突破性的进展.以 CasRel^[5]、BiRTE^[10]、PRGC^[11]以及文献[12]为代表的基于标注的方法很好地解决了三元组重叠问题,模型的计算效率也比传统的关系抽取方法有了很大提升,但是误差传递和暴露偏差问题没有解决,仍采用的是流水线式抽取,一定程度上影响了三元组抽取的精度.

基于生成的方法不同于以往的抽取范式,从 2018 年以 Seq2Seq 的模型框架形式被提出^[13-16],但因为相比于抽取式模型没有明显的提升而没有形成热点,直到近年来 GPT、T5 等大型生成式预训练模型的推广,使得基于生成的方法具有更高的性能.生成式模型迁移能力更强,适配于不同场景、不同任务,对于零样本以及少样本低资源场景下性能更为突出^[6,7,17,18].但这种方法往往需要消耗更多的计算资源,在某些需要低成本的垂直领域,费效比较低.同时某些边界模糊的上下文,生成式模型也难以处理,输出错误的三元组信息^[7].

以 SpanRE^[19]、SpERT^[20]、PURE^[8]以及文献[21]为代表的基于片段的方法,将关系抽取分为片段列举、向量表示、标签分类预测以及关系抽取 4 个步骤.其中 PURE 不同于其他模型,采用流水线式抽取思路,用不同的编码器捕捉实体和关系抽取两个子任务的信息,创新地在输入层引入实体信息,同时利用跨句的文本依赖信息,有效地提升了关系抽取的性能,超越了很多联合抽取模型.但在预测关系时,需要逐个判断片段对,大幅度降低了计算效率.文献[22]以基于片段的方法为基础,将关系抽取看作多轮问答,进而从文本中识别答案来转化成关系抽取的三元组形式.但仍没解决流水线式抽取所带来的问题.

基于填表的方法^[23,24]通常为数据集中每一个关系创建一个表格,表中每一个单元表示一个单词对的关系,通过构建模型训练,准确的填充表格,再根据填表策略对表格进行解码,输出对应的三元组.与基于标注的方法相比,其中 TPLinker^[25]很好地解决了暴露偏差和误差累积的问题,但填表的复杂度较高,存在较多的冗余信息;GRTE^[9]在 TPLinker 的基础上,减少了填表的数目,很好地挖掘全局的特征,不足的是模型解码部分参数大,推理效率下降.本文将中文文本的特性与基于填表的方法相融合,在填表数目和模型体量上进行优化,同时避免了其他关系抽取方法中存在的误差传递、暴露偏差等问题.

2 方法论

2.1 填表策略

给定一个中文句子 $S = [t_1, t_2, \dots, t_n]$, 为每一个关系 r 制作一个表格 $table_r$, 其中表格 $table_r$ 大小为 $n \times n$, 关系 r 包含于关系集合 R . 考虑中文在编码时多单词的特性,为每一个表格中的单元设计的不同的标签.标签

集合 L 定义如下:

$$L = \{ \text{"None"}, \text{"Head"}, \text{"Tail"} \}$$

其中, 标签“None”表示两个单词间 (t_i, t_j) 没有关联关系, 标签“Head”表示两个单词间 (t_i, t_j) 均为两个事件的头单词, 标签“Tail”表示两个单词间 (t_i, t_j) 均为两个事件的尾单词.

图1中为某个关系的表格标注示例, 在事件因果关系中, 较为普遍的情况是前因后果, 所以对于表格矩阵

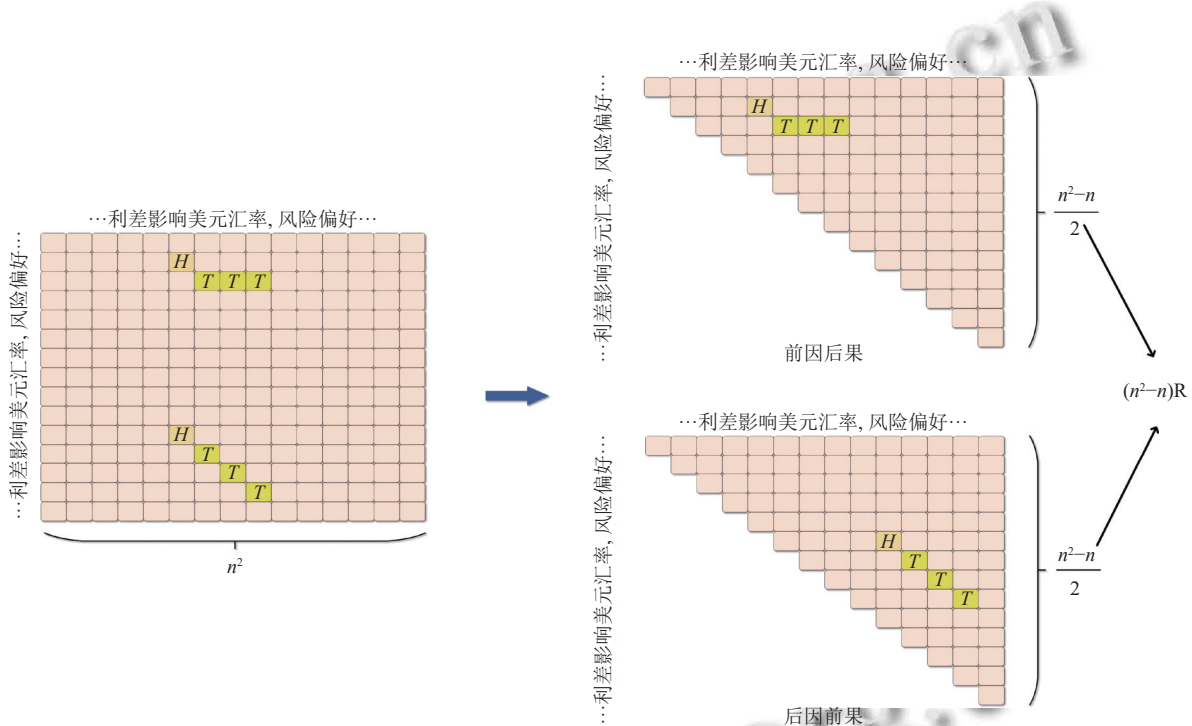


图1 表格标注示例及解码策略

本文填表数目为 $(n^2 - n)R$, 低于 TPLinker 的 $\frac{n^2 + n}{2}$ $(2R + 1)$ 个和 GRTE 的 $n^2 R$ 个.

2.2 模型描述

模型的架构如图2所示, 主要包括数据预处理、编码模块、Bi-FLASH-SRU层、学习表格特征的循环模块以及解码算法.

2.2.1 数据预处理

数据集文本中可能包括特殊字符, 通过使用分词器对文本进行字符级的分词, 去除冗余信息后, 得到序列化文本, 输入编码模块.

2.2.2 编码模块

得到序列文本后, 使用预训练模型对其进行编码, 输出由字符向量、位置向量和分句向量组成的动态词

来说, 下三角区域的标签均为“None”, 导致矩阵稀疏, 浪费了存储空间的同时, 还导致填表慢. 但又不能对下三角区域直接丢弃, 因为可能会出现后因前果的情况发生. 所以模型的优势在于, 为后因前果进行填表映射, 同时考虑多对因果所带来的标签覆盖问题, 为每一个关系 r 映射一个反关系 r' , 同样地制作一个表格 $table_{r'}$, 将 $table_r$ 下三角区域的表格单元对应的填充到 $table_{r'}$ 的上三角区域, 并丢弃下三角区域以及对角线区域.

向量, 作为全局特征 e .

2.2.3 Bi-FLASH-SRU层

为了进一步获得主体事件特征 e_1 和客体事件特征 e_2 , 将编码模块输出的全局特征 e 作为 Bi-FLASH-SRU 的输入, 进行二次上下文语义捕捉. 使用两个独立的 Bi-FLASH-SRU 来获取主客体事件特征, 如式(1).

$$\begin{cases} e_1 = \text{Bi-FLASH-SRU}(e) \\ e_2 = \text{Bi-FLASH-SRU}(e) \end{cases} \quad (1)$$

Bi-FLASH-SRU 是由内置注意力简单循环单元(SRU++)^[26]改进而来. SRU++内置注意力层使用的缩放点积注意力, 当输入序列越长时, 计算储存矩阵会导致时间和空间复杂度几何式增长. 针对此问题, 引入闪电注意力机制(flash-attention)^[27], 提出了内嵌闪电注意

力简单循环单元, 模型结构如图3所示.

图3中, 768、512 分别表示输入向量维度和注意力维度. 闪电注意力机制主要在重新构建注意力计算, 将输入信息以块的形式计算, 在输入块上完成多次运

行, 从而实现 Softmax 的缩减; 以及储存向前传递的 Softmax 归一化因子, 以便在后向传递中重新快速计算芯片上的注意力, 这比从 HBM 中读取中间注意力矩阵的传统方法速度更快, 闪电注意力机制结构如图4所示.

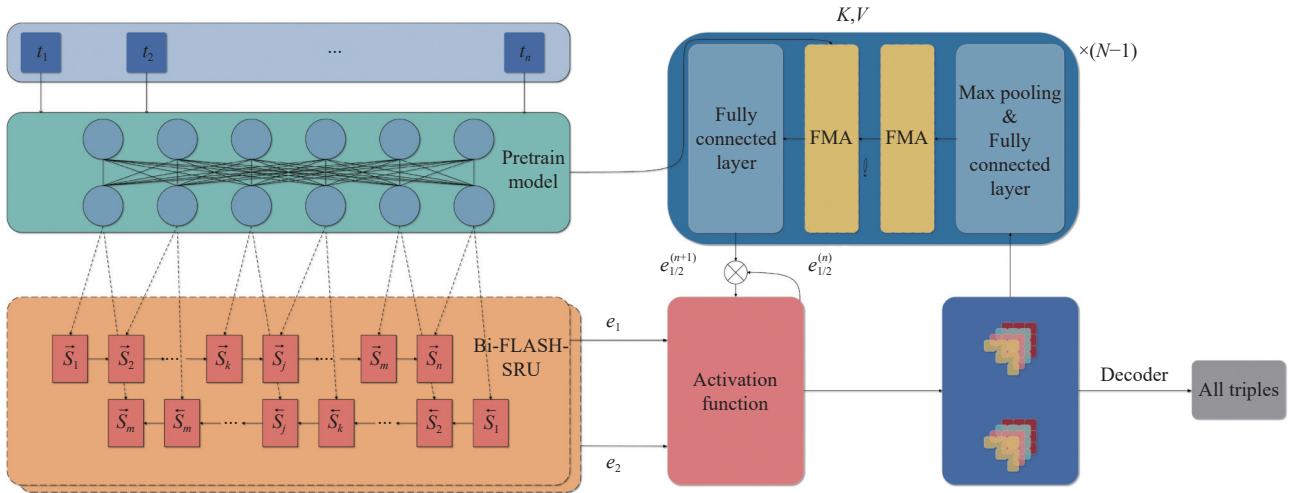


图2 模型结构图

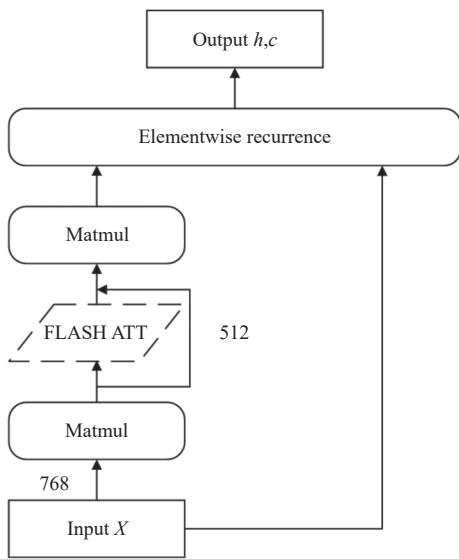


图3 FLASH-SRU 模型结构

2.2.4 学习表格特征的循环模块

主要分为以下3步.

第1步: 为了获得表格特征, 将 Bi-FLASH-SRU 层所获得的主体事件特征 e_1 和客体事件特征 e_2 作为输入. 由于要进行若干次迭代的计算, 令第 t 轮迭代中, 主体事件特征和客体事件特征分别为 $e_1^{(t)}$ 和 $e_2^{(t)}$, 关系 re 的表格特征为 $tf_{re}^{(t)}$, 每一个单词对 (t_i, t_j) 所代表的表格特征

如式 (2).

$$tf_{re}^{(t)}(i, j) = W_{re} \text{ReLU}(e_{1,j}^{(t)} \circ e_{2,j}^{(t)}) + b_{re} \quad (2)$$

其中, \circ 代表哈达玛积运算, ReLU 是激活函数, $e_{1,j}^{(t)}$ 和 $e_{2,j}^{(t)}$ 分别为在第 t 轮迭代单词 t_i 和 t_j 的特征, W_{re} 和 b_{re} 分别为训练权重和训练偏置.

第2步: 将第1步的局部关联特征进行拓展, 深入挖掘表格的全局特征, 形成新的表特征. 将所有的事件关系进行拼接, 形成一个大的表格特征 $tf^{(t)}$, 然后用最大池化和全连接得到与主客体事件相关的表格特征 $tf_1^{(t)}$ 和 $tf_2^{(t)}$, 如式 (3).

$$\begin{cases} tf_1^{(t)} = W_1 \maxpool_1(tf^{(t)}) + b_1 \\ tf_2^{(t)} = W_2 \maxpool_2(tf^{(t)}) + b_2 \end{cases} \quad (3)$$

其中, $W_{1/2}$ 和 $b_{1/2}$ 分别表示训练权重和训练偏置.

第3步: 用基于 Transformer 模型来挖掘关系以及单词对之间的全局关联, 同时考虑传统多头注意力机制计算过慢, 结合闪电注意力机制, 引入 FMA (flash-multi-attention). 首先挖掘事件关系之间的关联, 在主客体事件相关的表格特征 $tf_{1/2}^{(t)}$ 上进行 FMA 计算; 再挖掘单词对的全局关联, 在全局特征 e 和 FMA 计算过后的主客体事件相关的表格特征 $tf_{1/2}^{(t)}$ 上进行 FMA 计算; 最

后利用全连接模块和激活函数生成新的主客体事件特征, 如式 (4).

$$\begin{cases} \hat{t}_{1/2}^{(t)} = \text{FMA}(t_{1/2}^{(t)}) \\ \hat{e}_{1/2}^{(t+1)} = \text{FMA}(t_{1/2}^{(t)}, e, e) \\ e_{1/2}^{(t+1)} = \text{ReLU}(\hat{e}_{1/2}^{(t+1)}W + b) \end{cases} \quad (4)$$

为了缓解梯度消失, 使用残差连接生成最终的主客体事件特征, 如式 (5).

$$e_{1/2}^{(t+1)} = \text{LayerNorm}(e_{1/2}^{(t)} + e_{1/2}^{(t+1)}) \quad (5)$$

最后所得到的主客体事件特征作为输入进入下一循环, 继续学习表格特征.

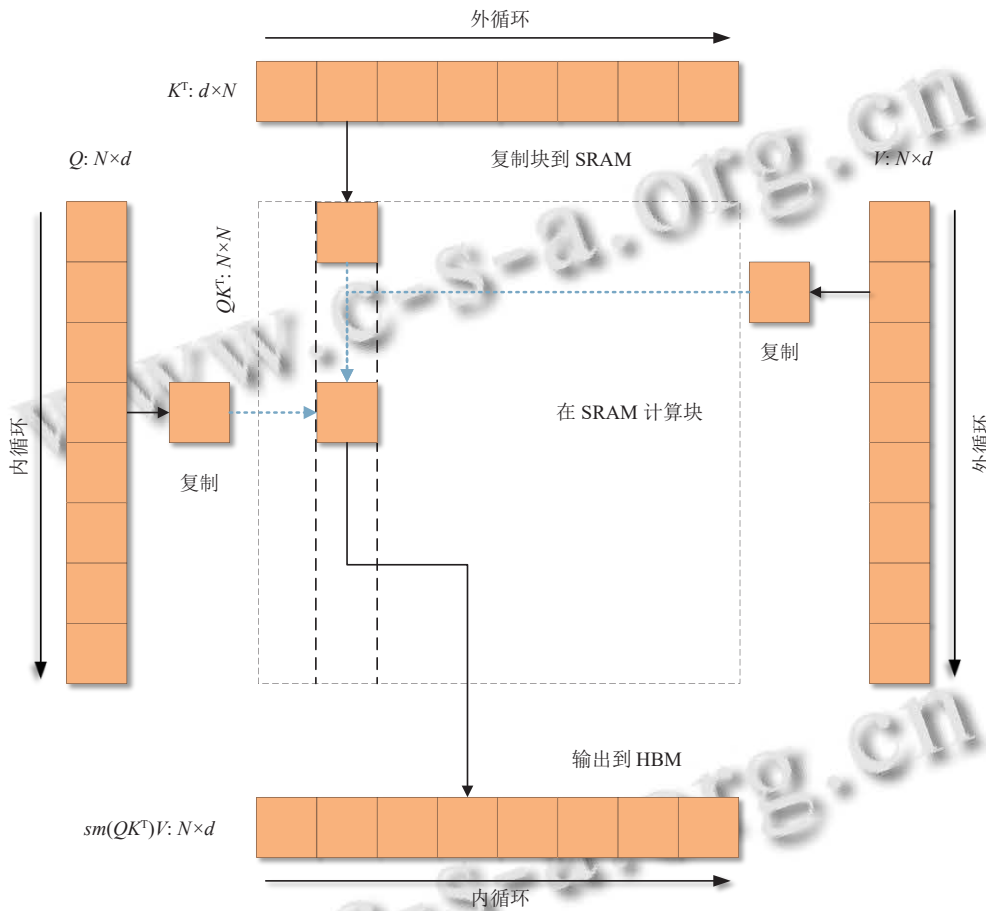


图4 闪电注意力机制图

2.2.5 解码算法

结束 T 轮循环后所获得的表格特征 $tf^{(T)}$ 作为输入, 首先进行如式 (6) 的计算操作.

$$\begin{cases} \hat{table}_r(i, j) = \text{Softmax}(tf_r^{(T)}(i, j)) \\ table_r(i, j) = \arg \max_{l \in L} (\hat{table}_r(i, j)[l]) \end{cases} \quad (6)$$

其中, $\hat{table}_r(i, j) \in \mathbb{R}^{|L|}$ 和 $table_r(i, j)$ 代表关系表格中单词对 (t_i, t_j) 的标签结果.

之后使用表格解码算法进行三元组提取, 并将关系 r 映射的反关系 r' 进行解码, 形成三元组集合, 表格解码算法如算法 1 所示.

3 实验分析

3.1 数据集与评价标准

为了验证模型的有效性, 采用同花顺面向金融领域的事件因果关系抽取 (THS) 和 CCKS2021 金融领域事件因果关系抽取 (CCKS2021) 公开数据集进行验证. 同花顺面向金融领域的事件因果关系抽取数据集被划分为训练集 2 000 条、验证集 500 条、测试集 500 条. CCKS2021 金融领域事件因果关系抽取数据集被划分为训练集 5 000 条、验证集 1 000 条、测试集 1 000 条.

实验使用精确率 P 、召回率 R 和 F_1 值对模型进行评价. F_1 值的计算公式如下:

$$F_1 = \frac{2 \times P \times R}{P + R} \quad (7)$$

算法 1. 表格解码算法

输入: 关系集合 R , 句子 $S = \{t_1, t_2, \dots, t_n\}$, 所有的 $table_{r_1} \in \mathbb{R}^{n \times n}$ 对应每个关系 $r_1 \in R$, 所有的 $table_{r_2} \in \mathbb{R}^{n \times n}$ 对应每个反关系 $r_2 \in R$, r_1 与 r_2 互相对应。
输出: 输出所有的预测三元组集合 RT 。

```

 $H_{1/2} \leftarrow \emptyset, R_{1/2} \leftarrow \emptyset, RT \leftarrow \emptyset$  /*定义临时三元组集合  $H_{1/2}$  和  $R_{1/2}$ , 并初始化所有集合*/
for 每一个  $r_{1/2} \in R$  do
    定义临时集合  $WP_{r_{1/2}}^H, WP_{r_{1/2}}^T$  /* $H, T$  代表  $table_{r_{1/2}}$  中单词对组成的标签结尾字符*/
for 每一个  $(t_i, t_j) \in WP_{r_{1/2}}^H$  do // 正向搜索
    1) 搜索满足以下条件的单词对  $(t_k, t_m): i \leq k, j \leq m$ , 同时  $table_{r_{1/2}}(t_i, t_j)$  和  $table_{r_{1/2}}(t_k, t_m)$  相匹配且  $(t_i, t_j)$  和  $(t_k, t_m)$  是最接近的
    2) 添加满足条件的三元组  $(t_{i \dots k}, r, t_{j \dots m})$  到  $H_1$ , 添加满足条件的三元组  $(t_{j \dots m}, r, t_{i \dots k})$  到  $H_2$ 
结束 for 循环
for 每一个  $(t_k, t_m) \in WP_{r_{1/2}}^T$  do // 逆向搜索
    1) 搜索满足以下条件的单词对, 搜索策略参考正向搜索
    2) 添加满足条件的三元组  $(t_{i \dots k}, r, t_{j \dots m})$  到  $T_1$ , 添加满足条件的三元组  $(t_{j \dots m}, r, t_{i \dots k})$  到  $T_2$ 
结束 for 循环
结束 for 循环
 $RT \leftarrow H_{1/2} \cup T_{1/2}$ 
返回  $RT$  // 输出结果
    
```

3.2 实验环境与参数

本文实验环境为操作系统 Ubuntu 18.04, GPU NVIDIA Tesla T4, RAM 32 GB, 硬盘 100 GB, 使用 Python 3.8.16, PyTorch 1.9 深度学习环境。

经过多轮实验对比后, 选取模型参数如表 1 所示. 并使用 Adam 优化器, 自适应调整学习率, 同时根据不同的数据集设定不同的学习表格特征模块的循环次数, THS 数据集设定为 3 轮次, CCKS2021 数据集设定为 2 轮次。

表 1 超参数配置

超参数名称	数值
Batch_size	4
Learning_rate	2E-5
Max_seq_len	200
Bi-FLASH-SRU_hidden_size	512
Bi-FLASH-SRU_proj_size	768
Bi-FLASH-SRU_num_layers	2
Num_epoch	20
Loss	交叉熵函数

3.3 实验设计

为了验证本文提出模型的优越性, 我们对比了很多取得 SOTA 的关系抽取模型, 有 copyRE^[12]、CasRel^[5]、

TPLinker^[25]、SPN^[17]、GRTE^[9]以及文献^[28], 这些基线模型的实验结果, 是在开源代码提供或根据文章复现的条件下产生的. 同时选取了不同的词向量模型, 如 ELMO、GloVe 进行了对比实验. 最后在同等参数条件的情况下, 选取不同神经单元, 对文中提出的 Bi-FLASH-SRU 进行实验验证。

3.4 实验结果分析

为了确保实验结果的可信性, 降低随机误差, 设定随机种子, 取 5 次实验的平均结果作为最终实验结果, 实验结果如表 2 所示, 其中 TF-ChineseERE_(GloVe) 表示模型在编码模块使用 GloVe 词向量, TF-ChineseERE_{Bi-SRU++} 表示模型在 Bi-FLASH-SRU 层使用 Bi-SRU++ 进行替换。

表 2 实验结果 (%)

模型	THS			CCKS2021		
	P	R	F ₁	P	R	F ₁
copyRE	43.8	49.3	46.4	54.8	53.4	54.1
CasRel	54.3	57.7	55.9	59.2	58.0	58.6
TPLinker	56.3	58.4	57.3	60.1	59.3	59.7
文献 ^[28]	56.5	58.6	57.5	60.2	59.8	60.0
SPN	57.9	58.0	57.9	60.9	59.9	60.4
GRTE	58.1	59.0	58.5	62.5	60.2	61.3
TF-ChineseERE _(GloVe)	55.3	59.7	57.4	60.3	61.0	60.6
TF-ChineseERE _(ELMO)	57.1	58.6	57.8	61.5	60.3	60.9
TF-ChineseERE _(Roberta)	58.5	59.9	59.2	62.3	62.7	62.5
TF-ChineseERE _{Bi-LSTM}	57.1	59.2	58.1	61.2	61.5	61.3
TF-ChineseERE _{Bi-GRU}	57.2	58.9	58.0	60.9	61.2	61.0
TF-ChineseERE _{Bi-SRU++}	58.1	58.8	58.4	61.5	61.7	61.6

3.4.1 基线模型实验结果分析

由表 2 可知, 本文所提模型是十分高效的. 与其他取得 SOTA 的关系抽取模型相比, 在两个公开数据集上均取得了最高的 F_1 值. 相比 GRTE 和 TPLinker 同类表格式抽取模型, 两个公开数据集在 F_1 值上平均提高了 1% 和 1.7%, 同时在填表策略上, 本文模型填表数目是最少的. 综上所述, 本文模型在关系抽取上取得好的表现是很有意义的。

3.4.2 词向量实验结果分析

为了验证不同词向量对于本文模型的影响, 故采用 GloVe 和 ELMO 进行对比. 由表 2 第 6-8 行可知, 两组数据集在使用 Roberta 作为编码模块的词向量模型, 均取得了最高的 F_1 值. 尽管 Roberta 和 ELMO 都能产生动态词向量, 根据任务进行微调, 融合领域知识来解决一词多义的问题, 但 ELMO 是使用 Bi-LSTM 来

获得单词的上下文语义,而基于 Transformer 的 Roberta,其特征提取能力更强,所以效果更优.相对于静态词向量的 GloVe,仅是基于全局的词频统计的特征工具,效果肯定劣于动态词向量.但与表 2 第 1-5 行中一些取得 SOTA 的模型相比,还是取得了不错的效果.

3.4.3 神经元实验结果分析

不同循环神经元对比实验分析:为了验证不同循环神经元对模型的影响,选取 Bi-LSTM、Bi-GRU、Bi-SRU++进行对比实验,如表 2 第 10-12 行所示.由表 2 所知,本文所提模型的 F_1 值在两个数据集上均取得最高,说明 Bi-FLASH-SRU 能全面提取上下文的语义,内置的闪电注意力很好地感知了词间的依赖关系,有助于提升事件关系抽取的性能.在模型中使用 Bi-SRU++获取特征尽管在 F_1 值上与使用 Bi-FLASH-SRU 的较为相近,但还是体现了内置闪电注意力的优越性. Bi-LSTM 仅能获取上下文语义, Bi-GRU 不能完全解决梯度消失的问题,且不可并行计算,所以在 F_1 值上相差较大.

Bi-FLASH-SRU 的计算效率分析:为了验证 Bi-FLASH-SRU 的计算效率,在与其他循环神经元对比时,记录了 5 次训练模型的时间,并绘制了 F_1 值和最小平均训练时间 (Min_Time) 与平均训练时间 (Time) 的散点图,如图 5 所示.

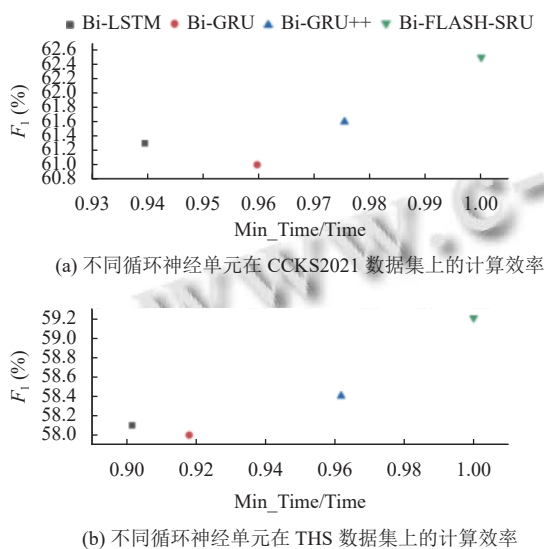


图 5 不同循环神经元的计算效率

平均训练时间小的神经元横坐标值更大,最大值为 1,纵坐标 F_1 值越高代表神经元性能越好,因此处于散点图右上方的神经元评价越高.

由图 5 所知, Bi-FLASH-SRU 显著高于其他神经单元.与 Bi-SRU++相比,说明了闪电注意力有效提升了训练速度.尽管 Bi-GRU 由于简化了门结构,裁剪了参数量,但和 Bi-LSTM 同样存在不可并行化处理以及没有内置注意力,所以训练速度和抽取性能都劣于 Bi-FLASH-SRU.

Bi-FLASH-SRU 层数影响及闪电注意力机制的有效性实验分析:为了验证内置闪电注意力机制的有效性以及 Bi-FLASH-SRU 不同层数对模型效果的影响,基于 THS 数据集进行了不同层数、不同条件下的闪电注意力机制的对比实验,如图 6 所示.其中 num_head 表示多头注意力的数量, attn_n_layers 表示每 N 层内置闪电注意力.

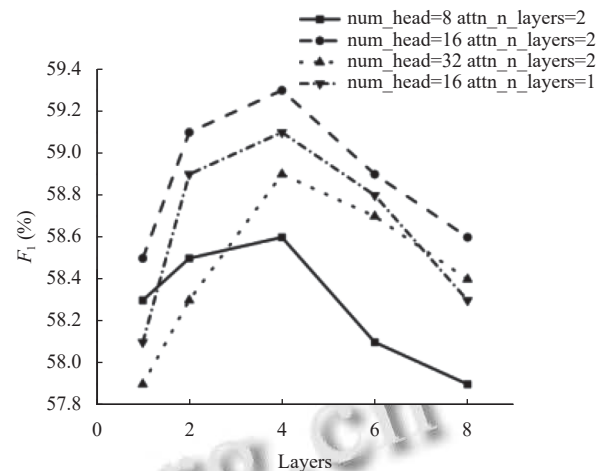


图 6 不同层数、不同条件下的内置闪电注意力的有效性

由图 6 可知,当 Bi-FLASH-SRU 达到 4 层时,4 种条件下均出现 F_1 值的顶峰,此后在增加层数, F_1 值出现退化现象,说明模型层数不宜过多,需要根据任务实际情况而设定层数.当 num_head=16、attn_n_layers=2 时,模型表现最好,证明了注意力机制并不是越多越好,多头数量也应该设定恰当,这样才能提高模型的迁移与泛化能力.

4 结论与展望

本文提出了一个新的基于表格关系抽取方法和模型,并应用于事件关系抽取任务,在两个公开数据集上进行实验并取得了良好的抽取效果.相比于先前的表格抽取式方法,采用关系映射和裁剪表格的方法,使填表数目进一步缩小,比 TPLinker 减少了 $\frac{n^2+n}{2}+$

$2nR$, 比 GRTE 减少了 nR ; Bi-FLASH-SRU 在完整获取上下文语义和捕捉词间关系, 相比于基线系统, 在两个公开数据集上, 将 F_1 值分别提升了 0.7% 和 1.2%, 同时计算效率也提升了约 3%。

需要注意的是, 内置闪电注意力机制和模型层数的使用需要根据任务进行设定, 才能达到最优效果。本文仅在事件因果关系抽取任务上进行了实验, 在接下来工作中, 将尝试在其他关系的事件关系抽取任务中进行应用, 并考虑与 GPT 等生成式预训练模型, 取得更好的抽取效果。

参考文献

- 1 Zhou GD, Su J, Zhang J, *et al.* Exploring various knowledge in relation extraction. Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics. Ann Arbor: ACL, 2005. 427–434.
- 2 Culotta A, McCallum A, Betz J. Integrating probabilistic extraction models and data mining to discover relations and patterns in text. Proceedings of the 2006 Human Language Technology Conference of the NAACL, Main Conference. New York: NAACL, 2006. 296–303.
- 3 Zelenko D, Aone C, Richardella A. Kernel methods for relation extraction. Journal of Machine Learning Research, 2003, 3: 1083–1106.
- 4 Bunescu R, Mooney R. A shortest path dependency kernel for relation extraction. Proceedings of the 2005 Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing. Vancouver: ACL, 2005. 724–731.
- 5 Wei Z, Su J, Wang Y, *et al.* A novel hierarchical binary tagging framework for joint extraction of entities and relations. arXiv:1909.03227, 2019.
- 6 Lu YJ, Liu Q, Dai D, *et al.* Unified structure generation for universal information extraction. Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics. Dublin: ACL, 2022. 5755–5772.
- 7 Ye HB, Zhang NY, Deng SM, *et al.* Contrastive triple extraction with generative Transformer. Proceedings of the 2021 AAAI Conference on Artificial Intelligence. AAAI, 2021. 14257–14265.
- 8 Zhong ZX, Chen DQ. A frustratingly easy approach for entity and relation extraction. Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. ACL, 2021. 50–61.
- 9 Ren FL, Zhang LH, Yin SJ, *et al.* A novel global feature-oriented relational triple extraction model based on table filling. Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing. Punta Cana: ACL, 2021. 2646–2656.
- 10 Ren FL, Zhang LH, Zhao XF, *et al.* A simple but effective bidirectional framework for relational triple extraction. Proceedings of the 15th ACM International Conference on Web Search and Data Mining. Tempe: WSDM, 2022. 824–832.
- 11 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. ACL, 2021. 6225–6235.
- 12 方义秋, 刘飞, 葛君伟. 基于 Mogrifier LSTM 的序列标注关系抽取方法. 计算机工程, 2022, 48(10): 81–87, 94.
- 13 Zeng XR, Zeng DJ, He SZ, *et al.* Extracting relational facts by an end-to-end neural model with copy mechanism. Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics. Melbourne: ACL, 2018. 506–514.
- 14 Zeng XR, He SZ, Zeng DJ, *et al.* Learning the extraction order of multiple relational facts in a sentence with reinforcement learning. Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing. Hong Kong: ACL, 2019. 367–377.
- 15 Zeng DJ, Zhang HR, Liu QY. CopyMTL: Copy mechanism for joint extraction of entities and relations with multi-task learning. Proceedings of the 2020 AAAI Conference on Artificial Intelligence. New York: AAAI, 2020. 9507–9514.
- 16 Nayak T, Ng HT. Effective modeling of encoder-decoder architecture for joint entity and relation extraction. Proceedings of the 2020 AAAI Conference on Artificial Intelligence. New York: AAAI, 2020. 8528–8535.
- 17 Sui DB, Zeng XR, Chen YB, *et al.* Joint entity and relation extraction with set prediction networks. IEEE Transactions on Neural Networks and Learning Systems, 2023. [doi: 10.1109/TNNLS.2023.3264735]
- 18 Cabot PLH, Navigli R. REBEL: Relation extraction by end-to-end language generation. Proceedings of the 2021 Findings of the Association for Computational Linguistics. Punta Cana: ACL, 2021. 2370–2381.
- 19 Dixit K, Al-Onaizan Y. Span-level model for relation

- extraction. Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Florence: ACL, 2019. 5308–5314.
- 20 Eberts M, Ulges A. Span-based joint entity and relation extraction with transformer pre-training. Proceedings of the 24th European Conference on Artificial Intelligence. Santiago De Compostela: ECAI, 2020. 2006–2013.
- 21 郑肇谦, 韩东辰, 赵辉. 单步片段标注的实体关系联合抽取模型. 计算机工程与应用, 2023, 59(9): 130–139. [doi: 10.3778/j.issn.1002-8331.2112-0418]
- 22 Li XY, Yin F, Sun ZJ, *et al.* Entity-relation extraction as multi-turn question answering. Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Florence: ACL, 2019. 1340–1350.
- 23 Wang YJ, Sun CZ, Wu YB, *et al.* UniRE: A unified label space for entity relation extraction. Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing. ACL, 2021. 220–231.
- 24 Shang YM, Huang HY, Mao XL. OneRel: Joint entity and relation extraction with one module in one step. Proceedings of the 2022 AAAI Conference on Artificial Intelligence. AAAI, 2022. 11285–11293.
- 25 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: ACL, 2020. 1572–1582.
- 26 Lei T. When attention meets fast recurrence: Training language models with reduced compute. Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing. Punta Cana: ACL, 2021. 7633–7648.
- 27 Dao T, Fu DY, Ermon S, *et al.* FlashAttention: Fast and memory-efficient exact attention with IO-awareness. Proceedings of the 36th Advances in Neural Information Processing Systems. New Orleans: NeurIPS, 2022.
- 28 程春雷, 邹静, 叶青, 等. 融入关系形式化概念的端到端三元组抽取. 计算机工程与应用, 2023, 59(9): 182–189. [doi: 10.3778/j.issn.1002-8331.2201-0418]
- (校对责编: 孙君艳)