

基于联合学习的语言粒度融合的重叠事件抽取方法

闫婧涛, 李旸, 王素格, 潘邦泽

引用本文

闫婧涛, 李旸, 王素格, 潘邦泽. 基于联合学习的语言粒度融合的重叠事件抽取方法[J]. 计算机科学, 2024, 51(7): 287-295.

YAN Jingtao, LI Yang, WANG Suge, PAN Bangze. [Overlap Event Extraction Method with Language Granularity Fusion Based on Joint Learning](#) [J]. Computer Science, 2024, 51(7): 287-295.

相似文献推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[针对系统调用的基于语义特征的多方面信息融合的主机异常检测框架](#)

Host Anomaly Detection Framework Based on Multifaceted Information Fusion of Semantic Features for System Calls

计算机科学, 2024, 51(7): 380-388. <https://doi.org/10.11896/jsjcx.230400023>

[基于深度确定性策略梯度与注意力Critic的多智能体协同清障算法](#)

Multi-agent Cooperative Algorithm for Obstacle Clearance Based on Deep Deterministic Policy Gradient and Attention Critic

计算机科学, 2024, 51(7): 319-326. <https://doi.org/10.11896/jsjcx.230600129>

[融合多图卷积与层级池化的文本分类模型](#)

Text Classification Method Based on Multi Graph Convolution and Hierarchical Pooling

计算机科学, 2024, 51(7): 303-309. <https://doi.org/10.11896/jsjcx.230400164>

[基于外部先验和自先验注意力的图像描述生成方法](#)

Image Captioning Generation Method Based on External Prior and Self-prior Attention

计算机科学, 2024, 51(7): 214-220. <https://doi.org/10.11896/jsjcx.230600167>

[一种基于YOLOX_s的雾天场景目标检测方法](#)

Foggy Weather Object Detection Method Based on YOLOX_s

计算机科学, 2024, 51(7): 206-213. <https://doi.org/10.11896/jsjcx.230400086>

基于联合学习的语言粒度融合的重叠事件抽取方法

闫婧涛¹ 李 旻² 王素格^{1,3} 潘邦泽¹

1 山西大学计算机与信息技术学院 太原 030006

2 山西财经大学金融学院 太原 030006

3 山西大学计算智能与中文信息处理教育部重点实验室 太原 030006

(yanjingtao0731@163.com)

摘要 事件抽取是一项重要的信息抽取任务,现有的事件抽取方法大多假设一个句子中仅出现一个事件,然而,在真实的场景下,重叠事件是难以避免的。文中提出了一种基于联合学习的语言粒度融合的重叠事件抽取方法。该方法设计了基于 token 数目逐层递增和逐层递减的策略,对不同语言粒度的片段进行表示,在此基础上,构建了渐进式语言粒度融合的句子表示。通过引入事件信息感知,建立了基于门控机制的语言粒度和事件信息融合的句子表示。最后,通过联合学习词间的片段关系和角色关系,实现对事件触发词、论元、事件类型和论元角色的判别。在 FewFC 和 DuEE1.0-1 数据集上进行了实验,所提 LGFEE 模型在事件类型判别任务上的 F1 值分别提高了 0.8% 和 0.6%,在触发词识别、论元识别、论元角色分类任务上也获得了较高的召回率和 F1 值,验证了其有效性。

关键词: 重叠事件抽取;语言粒度融合;联合学习;注意力机制;门控机制

中图分类号 TP391

Overlap Event Extraction Method with Language Granularity Fusion Based on Joint Learning

YAN Jingtao¹, LI Yang², WANG Suge^{1,3} and PAN Bangze¹

1 School of Computer and Information Technology, Shanxi University, Taiyuan 030006, China

2 School of Finance, Shanxi University of Finance and Economics, Taiyuan 030006, China

3 Key Laboratory Computational Intelligence and Chinese Information Processing of Ministry of Education, Shanxi University, Taiyuan 030006, China

Abstract Event extraction is a crucial task in information extraction. The existing event extraction methods generally assume that only one event occurs in a sentence. However, overlapping events are inevitable in real scenarios. Therefore, this paper designs an overlap event extraction method with language granularity fusion based on joint learning. In this method, a strategy of increasing and decreasing token number layer by layer is designed to represent fragments of different language granularity. On this basis, a sentence representation of progressive language granularity fusion is constructed. By introducing event information perception, the sentence representation of language granularity and event information fusion based on gating mechanism is established. Finally, through the joint study of the fragment relationship and role relationship between words, the identification of event triggering words, arguments, event types and argument roles is realized. The experiments conducted on the FewFC and DuEE1.0-1 datasets demonstrate that the LGFEE model proposed in this paper achieves an improvement of 0.8% and 0.6% in the F1 score for event type discrimination tasks, respectively. Furthermore, it also exhibits higher recall rates and F1 scores in trigger word recognition, argument recognition, and argument role classification tasks, which verifies the validity of LGFEE model.

Keywords Overlapping event extraction, Language granularity fusion, Joint learning, Attention mechanism, Gating mechanism

1 引言

事件抽取(Event Extraction, EE)是自然语言处理领域中一项既重要又具有挑战性的任务,旨在从非结构文本中抽取

结构化形式的事件信息,即抽取触发词和论元、判别事件类型和论元角色。事件抽取在医疗^[1]、金融^[2]、司法^[3]等领域有着广泛的应用,特别是金融领域,通过对新闻以及企业公告等文本进行事件抽取,可以抽取其中的事件、时间、地点、人物、

到稿日期:2023-07-17 返修日期:2023-12-09

基金项目:国家重点研发计划(2022QY0300-01);国家自然科学基金(62106130);山西省高等学校科技创新项目(2021L284)

This work was supported by the National Key Research and Development Program of China(2022QY0300-01), National Natural Science Foundation of China(62106130) and Scientific and Technological Innovation Programs of Higher Education Institutions in Shanxi(2021L284).

通信作者:李旻(liyangprimrose@163.com)

金额等一系列核心要素,可为决策者提供重要的参考信息,根据实际情况及时做出科学决策。

事件抽取包括普通事件抽取和重叠事件抽取。Sheng等^[4]将重叠事件抽取归纳为:一个词在多个事件中作为触发词;一个词作为论元,在一个事件中扮演不同的角色或者在多个事件中扮演不同的角色。前者被称为触发词重叠事件抽取问题,后者被称为论元重叠事件抽取问题。图1所示的实例中,“收购”一词同时触发了“投资”和“股份股权转让”两个事件,属于触发词重叠事件抽取问题;“世纪华通”在“投资”事件中扮演“投资方”角色,在“股份股权转让”事件中扮演“受让转让公司”角色,属于论元重叠事件抽取问题。

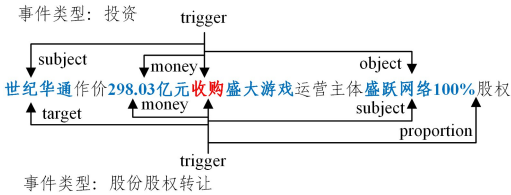


图1 重叠事件抽取实例

Fig. 1 Overlapping event extraction instances

传统的事件抽取方法^[5-6]将事件抽取视为序列标注任务,使用统一的解码器对句子进行一次标注,同时抽取触发词和论元,但有可能引起标记间的冲突,出现事件要素遗漏问题。因此,研究者开展了对重叠事件抽取方法的相关研究,主要采用基于 Pipeline 范式^[4]和基于 Joint 范式的事件抽取方法^[7]。基于 Pipeline 范式的事件抽取方法是在不同阶段依次抽取触发词和论元,容易忽略触发词和论元之间的依赖关系,导致错误传播。而现有的基于 Joint 范式的事件抽取方法^[7]未使用句子中不同语言粒度的片段表示,不能对句子进行深层次的有效表示。

语言粒度研究主要集中于字粒度、词粒度和句子粒度,然而,单一粒度的模式^[8]往往只会捕捉句子所包含的部分信息。因此,研究者开始探索不同语言粒度融合的方法,同时考虑多种粒度的信息,更全面地捕捉句子所包含的不同层次的语义和上下文信息。Xin等^[9]使用分词工具对文本进行切分来获取词粒度的信息,但分词可能会出现错误,影响后续的应用。Li等^[10]提出了非自回归翻译的多粒度优化方法,该方法应用随机屏蔽策略对每个源序列进行多次采样,关注每个源序列的不同部分,收集不同粒度翻译片段上模型的行为,以生成高质量的翻译,但该方法可能会破坏上下文的连续结构。

针对上述问题,本文提出了一种基于联合学习的语言粒度融合的重叠事件抽取方法(An Overlap Event Extraction Method with Language Granularity Fusion Based on Joint Learning, LGFEE),并建立了5个模块:基于BERT^[11]的句子表示模块、渐进式语言粒度融合模块、基于门控机制和注意力机制的信息融合模块、片段/角色标签预测模块、关系标签解码。

本文的贡献主要包括以下3个方面:

(1)提出了一种基于联合学习的语言粒度融合的重叠事件抽取方法(LGFEE),用于重叠事件抽取。

(2)提出了渐进式语言粒度融合方法,构建基于语言粒度

融合的文本表示,分别以 token 数目逐层递增和逐层递减的策略构建不同粒度的片段表示,融合了短文本包含的离散信息和长文本包含的抽象化语义信息。

(3)在公开的金融事件抽取数据集 FewFC 和 DuEE1.0-1 上进行实验,验证了所提模型的有效性。

2 相关工作

事件抽取包括触发词识别、论元识别、事件类型分类和论元角色分类4个子任务。依据4个子任务的实现过程,将事件抽取方法分为基于 Pipeline 范式的方法和基于 Joint 范式的方法^[12]。

2.1 基于 Pipeline 范式的事件抽取方法

基于 Pipeline 范式的事件抽取方法是将事件抽取任务转化为多阶段分类问题。首先识别触发词,再根据触发词判断事件类型;根据识别的触发词及事件类型的预测结果对事件论元进行抽取,并对论元角色进行分类。Chen等^[13]提出了DMCNN模型,该模型使用两个动态多池卷积神经网络,将词语级语义特征和句子级语义特征相结合,对触发词进行识别和分类。进一步,若句子中存在触发词,则再进行事件论元识别和分类。Yang等^[14]提出了PLMEE模型,该模型针对特定的事件角色分别进行抽取论元,以实现重叠事件论元的抽取。Sheng等^[4]提出了一种基于级联解码的重叠事件抽取框架CasEE,该框架依次执行事件类型检测、触发词抽取和论元抽取,在此基础上,将前一阶段的预测结果作为条件抽取重叠的目标,但该方法的后一阶段依赖于前一阶段,容易出现错误传播的问题。Xu等^[15]提出了MatEE模型,该模型首先通过分类得到触发词的开始和结束位置,然后建立事件矩阵,同时抽取触发词的多个事件,并抽取论元、建立论元间的关系,实现了重叠事件触发词的抽取。

2.2 基于 Joint 范式的事件抽取方法

基于 Joint 范式的事件抽取方法同时完成触发词和论元识别、事件类型和论元角色分类4个任务,在任务执行过程中,通过共享参数减少错误信息的传播,进一步利用全局特征提升模型的性能。Liu等^[5]提出了JMEE模型,该模型将句子从序列模式转换成句法依存图模式,并在句法依存图上使用GCN进行信息聚合,并采用联合方式,抽取事件触发词和论元。Nguyen等^[6]提出了JRNN模型,该模型使用两个循环神经网络从两个方向学习句子的表示,并将其应用于联合预测触发词和论元角色;为了捕获触发词和论元间的依赖关系,该模型引入记忆向量和记忆矩阵,存储标记过程中的预测信息。Li等^[16]提出了一种基于强化学习和对话指导的事件抽取方法,旨在从对话文本中逐步自动识别关联的事件和它们的论元。Cao等^[7]提出了OneEE模型,该模型将事件抽取转换为词与词间的关系识别任务,同时完成触发词和论元的抽取。

3 基于联合学习的语言粒度融合的重叠事件抽取 (LGFEE)

为了将不同语言粒度的文本信息和事件信息进行融合,本文构建了基于联合学习的语言粒度融合的重叠事件抽取

方法。首先,使用预训练语言模型 BERT^[11] 作为编码器,获得输入句子中词的嵌入表示;其次,提出了渐进式语言粒度融合模块,以 token 数目逐层递增和逐层递减的方式,分别构建包含不同 token 数目的片段表示,建立融合不同语言粒度的

句子表示;在此基础上,利用门控机制和注意力机制,将事件信息和不同粒度特征的文本表示融合,通过联合学习预测词间的片段关系和角色关系以及关系标签解码,获得事件触发词、论元、事件类型和论元角色。LGFE 模型架构如图 2 所示。

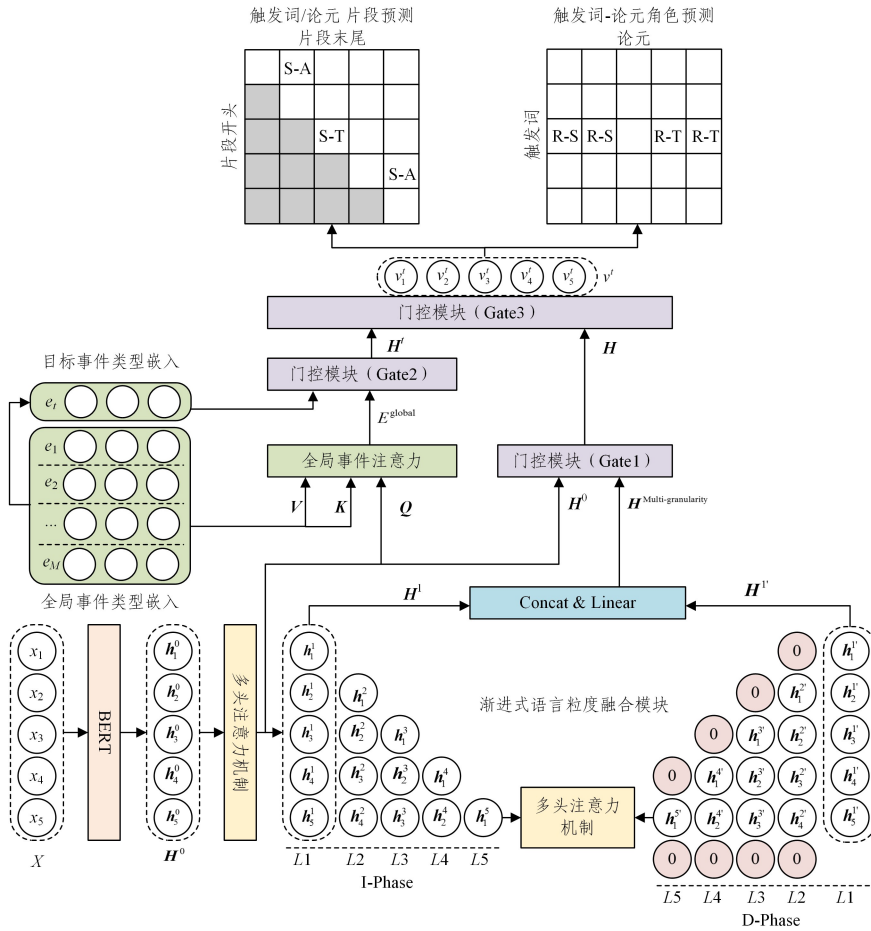


图 2 基于联合学习的语言粒度融合的重叠事件抽取方法(LGFEE)

Fig. 2 Overlap event extraction method with language granularity fusion based on joint learning(LGFEE)

句子的初始表示:给定一个具有 N 个词的输入句子 $X = (x_1, x_2, \dots, x_N)$,经过 BERT^[11] 编码后,再使用最大池化生成每个词的嵌入表示 $H^0 = (h_1^0, h_2^0, \dots, h_N^0) \in \mathbb{R}^{N \times d}$ 。

3.1 基于语言粒度融合的文本表示

为了更好地捕获句子的语义信息,本文提出了渐进式语言粒度融合模块,分别以 token 数目逐层递增(Incremental-Phase, I-Phase)和逐层递减(Decremental-Phase, D-Phase)两种方式,依次构建不同粒度的片段表示。由于卷积操作可以有效地提取文本中的局部特征,但也存在无法充分捕捉句子中较长范围的依赖关系和上下文信息的问题,因此引入多头注意力机制,使模型同时关注输入序列的不同位置并学习它们之间的关系。这有助于模型学习长距离依赖关系和上下文信息,故在 I-Phase 和 D-Phase 的每一层中,都利用卷积操作和多头注意力机制对上下文信息进行建模。

I-Phase 和 D-Phase 两个阶段均包含 L 层,其中,第 l 层的片段均为包含 l 个连续 token 的片段,片段表示分别为 $H^l = (h_1^l, h_2^l, \dots, h_T^l)$ 和 $H^{l'} = (h_1^{l'}, h_2^{l'}, \dots, h_T^{l'})$,其中, T 是片段的数量。图 3 给出了 I-Phase 和 D-Phase 中构建的第三层的详细结构。

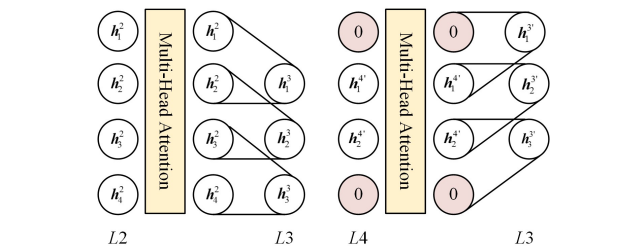


图 3 I-Phase 和 D-Phase 中构建 L3 层的详细结构
Fig. 3 Detailed structure of L3 layer is built in I-Phase and D-Phase

I-Phase:由于字粒度的文本相对较细,无法直接捕捉上下文的语义信息,需要采取分层组合和传递的方式,生成长文本的表示,逐步获取更全面、准确的上下文语义信息。因此,它的目标是将输入的文本逐层组合成不同长度的片段,并在不同层次上对这些片段进行特征提取和表示学习,将短文本包含的离散信息传递给长文本。具体来说,第一层的片段表示对应于 1-token 的片段,表示如下:

$$H^1 = f(H^0) \tag{1}$$

其中, $f(\cdot)$ 是多头注意力机制的简写,本文利用 Transformer^[17] 的多头注意力机制来增强片段表示,计算式如下:

$$\begin{aligned}
\mathbf{H}^l \text{multithread} &= \text{MultiHead}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) \\
&= \text{Concat}(\text{head}_1, \dots, \text{head}_i) \mathbf{W}^O \quad (2) \\
\text{head}_i &= \text{Attention}(\mathbf{Q}\mathbf{W}_i^Q, \mathbf{K}\mathbf{W}_i^K, \mathbf{V}\mathbf{W}_i^V) \\
&= \text{softmax}\left(\frac{\mathbf{Q}\mathbf{W}_i^Q (\mathbf{K}\mathbf{W}_i^K)^T}{\sqrt{d_k}}\right) \mathbf{V}\mathbf{W}_i^V \quad (3)
\end{aligned}$$

其中, head_i 为第 i 个头经过注意力得到的句子表示; $\mathbf{Q}, \mathbf{K}, \mathbf{V}$ 均为句子表示; $\mathbf{W}_i^Q, \mathbf{W}_i^K, \mathbf{W}_i^V, \mathbf{W}^O$ 表示可学习的参数矩阵; h 表示注意力头数; $d_k = d/h$ 。

对于更高的层, 为了避免破坏连续的上下文, 以 token 数目逐层递增的方式从第 $l-1$ 层迭代地生成第 l 层连续的 l -token 片段表示。利用卷积核为 2 的滤波器与第 $l-1$ 层经多头注意力机制增强的片段表示 $\mathbf{h}_{i,i+1}^{(l-1)\text{-multithread}}$ 进行卷积运算, 得到第 l 层中第 i 个片段的表示 \mathbf{h}_i^l , 其计算式如下:

$$\mathbf{h}_i^l = \text{ReLU}(\mathbf{W}_a \mathbf{h}_{i,i+1}^{(l-1)\text{-multithread}} + \mathbf{b}_a) \quad (4)$$

其中, \mathbf{W}_a 和 \mathbf{b}_a 是可学习的参数矩阵, $\mathbf{h}_{i,i+1}^{(l-1)\text{-multithread}}$ 表示句子中从 i 到 $i+1$ 的片段表示。采取步长为 1 的卷积操作, 在长度为 T 的序列上计算结束后, 可以得到 $T-1$ 个输出, 将其拼接可得到 I-Phase 中第 l 层的句子表示 \mathbf{H}^l :

$$\mathbf{H}^l = [\mathbf{h}_1^l; \mathbf{h}_2^l; \dots; \mathbf{h}_{T-1}^l] \quad (5)$$

其中, $[\cdot]$ 表示拼接操作。

D-Phase: 由于在同一上下文中, 长片段的特征可以提供与短片段特征互补的上下文信息。因此, D-Phase 旨在将较长片段中包含的抽象化语义信息传递给下一层的较短片段, 该信息即可作为 I-Phase 中对应片段的补充信息, D-Phase 中每一层的片段数均与 I-Phase 中对应层的片段数相同。具体来说, 以 I-Phase 中第 L 层的输出作为 D-Phase 第 L 层的输入, D-Phase 第 L 层的输入 $\mathbf{H}^{L'}$ 计算式如下:

$$\mathbf{H}^{L'} = f(\mathbf{H}^L) \quad (6)$$

其中, $f(\cdot)$ 类似于式(1)的简写。

与 I-Phase 类似, D-Phase 的第二阶段采用 token 数目逐层递减的方式, 从第 $l+1$ 层迭代地构建第 l 层连续的 l -token 片段表示。在 $l+1$ 层经多头注意力机制增强的文本表示 $\mathbf{H}^{(l+1)'\text{-multithread}}$ 两端填充零向量, 填充零向量后的新表示 $\mathbf{H}^{(l+1)'\text{-new}}$ 如式(7)所示:

$$\mathbf{H}^{(l+1)'\text{-new}} = [0; \mathbf{h}^{(l+1)'\text{-multithread}}; 0] \quad (7)$$

在 l 层, 利用卷积核为 2 的滤波器与 $\mathbf{h}_{i,i+1}^{(l+1)'\text{-new}}$ 进行卷积运算得到 l 层的第 i 个片段的表示 \mathbf{h}_i^l , 其计算式如下:

$$\mathbf{h}_i^l = \text{ReLU}(\mathbf{W}_b \mathbf{h}_{i,i+1}^{(l+1)'\text{-new}} + \mathbf{b}_b) \quad (8)$$

其中, $\mathbf{W}(1)_b$ 和 \mathbf{b}_b 是可学习的参数矩阵。

通过零向量填充后, 序列的长度变为 $T+2$, 采取步长为 1 的卷积操作, 可以得到 $T+1$ 个输出, 将其拼接得到 D-Phase 中第 l 层的句子表示 $\mathbf{H}^{l'}$ 如下:

$$\mathbf{H}^{l'} = [\mathbf{h}_1^l; \mathbf{h}_2^l; \dots; \mathbf{h}_{T+1}^l] \quad (9)$$

其中, $[\cdot]$ 表示拼接操作。

在图 3(b)中, D-Phase 中第 3 层的片段表示基于第 4 层的片段表示 $(0, \mathbf{h}_1^4, \mathbf{h}_2^4, 0)$ 构建, 记作 $(\mathbf{h}_1^3, \mathbf{h}_2^3, \mathbf{h}_3^3)$ 。

最后, 将 I-Phase 与 D-Phase 中第一层的句子表示拼接, 作为融合不同粒度特征的句子表示 $\mathbf{H}^{\text{Multi-granularity}}$:

$$\mathbf{H}^{\text{Multi-granularity}} = \mathbf{W}_c [\mathbf{H}^1; \mathbf{H}^{L'}] + \mathbf{b}_c \quad (10)$$

其中, \mathbf{W}_c 和 \mathbf{b}_c 是可学习的参数矩阵, $[\cdot]$ 表示拼接操作。

3.2 基于门控机制和注意力机制的信息融合

由于本研究的目标是预测目标事件类型的词间关系, 因此, 本节基于注意力机制和门控机制融合事件信息和文本信息, 为目标事件生成相应的句子表示。其中, 注意力机制用于建模事件间的交互以及事件与文本间的交互; 门控机制用于融合两类特征并过滤掉不必要的信息。

融合不同粒度特征的句子表示: 采用门控模块 Gate1, 在经 BERT^[11] 编码的句子表示 \mathbf{H}^0 的基础上, 融合不同粒度特征句子表示 $\mathbf{H}^{\text{Multi-granularity}}$, 得到融合不同粒度特征句子的新表示 \mathbf{H} :

$$\begin{aligned}
\mathbf{H} &= \text{Gate1}(\mathbf{H}^0, \mathbf{H}^{\text{Multi-granularity}}) \\
&= \mathbf{g}_1 \odot \mathbf{H}^0 + (1 - \mathbf{g}_1) \odot \mathbf{H}^{\text{Multi-granularity}} \quad (11)
\end{aligned}$$

$$\mathbf{g}_1 = \text{Sigmoid}(\mathbf{W}_{g_1} [\mathbf{H}^0; \mathbf{H}^{\text{Multi-granularity}}] + \mathbf{b}_{g_1}) \quad (12)$$

其中, \mathbf{W}_{g_1} 和 \mathbf{b}_{g_1} 是可学习的参数矩阵, \odot 和 $[\cdot]$ 分别表示元素相乘和拼接操作。

为了减少计算资源, 本文参考 Cao 等^[7] 的工作, 使用正、负采样策略对事件类型进行采样。给定一组采样的事件类型, 随机初始化其嵌入表示 $\mathbf{E} = (\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_M) \in \mathbb{R}^{M \times d}$, 其中, M 为采样事件类型的个数, 包含一个正类型和 $M-1$ 个负类型。利用注意力机制对事件间、事件与文本间的交互进行建模, 获得句子的全局事件嵌入表示 $\mathbf{E}^{\text{global}}$:

$$\mathbf{Q}, \mathbf{K}, \mathbf{V} = \mathbf{H}^0 \mathbf{W}^Q, \mathbf{E} \mathbf{W}^K, \mathbf{E} \mathbf{W}^V \quad (13)$$

$$\mathbf{E}^{\text{global}} = \text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d_h}}\right) \mathbf{V} \quad (14)$$

其中, $\mathbf{W}^Q, \mathbf{W}^K, \mathbf{W}^V$ 是可学习的参数矩阵。

目标事件感知的句子表示: 采用门控模块 Gate2, 在全局事件嵌入 $\mathbf{E}^{\text{global}}$ 的基础上, 融合目标事件类型嵌入 \mathbf{e}_t , 生成目标事件感知的句子表示 \mathbf{H}' :

$$\mathbf{H}' = \text{Gate2}(\mathbf{E}^{\text{global}}, \mathbf{e}_t) = \mathbf{g}_2 \odot \mathbf{E}^{\text{global}} + (1 - \mathbf{g}_2) \odot \mathbf{e}_t \quad (15)$$

$$\mathbf{g}_2 = \text{Sigmoid}(\mathbf{W}_{g_2} [\mathbf{E}^{\text{global}}; \mathbf{e}_t] + \mathbf{b}_{g_2}) \quad (16)$$

其中, $\mathbf{e}_t \in \mathbf{E}$ 表示目标事件类型嵌入, \mathbf{W}_{g_2} 和 \mathbf{b}_{g_2} 是可学习的参数矩阵。

基于不同粒度特征和目标事件感知的句子表示: 采用门控模块 Gate3, 将式(10)和式(6)分别得到的句子表示 \mathbf{H} 和 \mathbf{H}' 进一步融合, 得到融合目标事件 \mathbf{e}_t 的句子最终表示 $\mathbf{V}^t = (\mathbf{v}_1^t, \mathbf{v}_2^t, \dots, \mathbf{v}_N^t) \in \mathbb{R}^{N \times d}$, 其中, \mathbf{v}_i^t 是句子中词的嵌入表示。

$$\mathbf{V}^t = \text{Gate3}(\mathbf{H}, \mathbf{H}') = \mathbf{g}_3 \odot \mathbf{H} + (1 - \mathbf{g}_3) \odot \mathbf{H}' \quad (17)$$

$$\mathbf{g}_3 = \text{Sigmoid}(\mathbf{W}_{g_3} [\mathbf{H}; \mathbf{H}'] + \mathbf{b}_{g_3}) \quad (18)$$

其中, \mathbf{W}_{g_3} 和 \mathbf{b}_{g_3} 是可学习的参数矩阵。

引入多个门控模块, 模型可以更精确地控制信息的流动和影响。首先利用 Gate1 动态地选择句子的初始表示和融合不同语言粒度的文本表示; 然后利用 Gate2 过滤目标事件信息和全局事件信息; 最后通过门控模块 Gate3 融合事件表示和文本表示, 可以使每个门控机制根据不同的任务或上下文条件动态地选择输入的信息, 提高模型在不同情况下的泛化能力, 使模型更好地适应不同的数据分布。

进行分类。BERT-CRF 在 BERT 基础上添加条件随机场 (CRF) 捕获标签依赖关系, Du 等^[21] 将其用于文档级事件抽取。BERT-CRF-joint 是将 BIO 标注方案扩展为类型和角色的联合标注 B/I/O-type-role。以上方法会产生标签冲突, 不利于重叠事件抽取任务。

(2) 重叠事件抽取方法: PLMEE^[14] 针对特定的角色分别抽取论元, 可以抽取出重叠论元。CasEE^[4] 依次执行事件类型检测、触发词抽取和论元抽取, 其根据前一步的预测结果抽取了重叠事件目标。OneEE^[7] 将事件抽取表示为词与词间的关系识别, 在一个阶段内同时识别触发词或论元的词间关系, 并行地进行网格标注, 抽取重叠目标。MatEE^[15] 使用事件矩阵同时抽取共享触发词的多个事件, 可以抽取出重叠触发词。

4.3 参数设置

本文模型利用 BERT^[11] 模型对文本进行编码, 隐层向量维度为 768, 优化器采用 AdamW。BERT 的学习率为 2×10^{-5} , 其他模块的学习率为 1×10^{-3} 。多头注意力机制中并行头的数量 h 设置为 8, 层数 L 设置为 12, 采样的事件类型的

个数 M 设置为 6 (其中正、负样本的个数分别为 1 和 5)。批量大小设置为 8, 训练 30 轮。

5 实验结果及分析

5.1 比较实验结果分析

为了验证本文提出的 LGFEE 模型的有效性, 利用 4.2 节介绍的比较方法, 与本文方法进行对比, 设计了下面 3 类实验: (1) 将本文方法与重叠事件抽取方法分别在 FewFC 数据集和 DuEE1.0-1 数据集上进行对比实验; (2) 将本文方法与普通事件抽取方法在 FewFC 数据集上进行对比实验; (3) 将本文方法与 4.2 节介绍的所有方法在 FewFC-overlap 和 FewFC-normal 数据集上进行对比实验。

(1) 重叠事件抽取方法在 FewFC 和 DuEE1.0-1 数据集上的实验结果

将本文方法与重叠事件抽取方法分别在 FewFC 和 DuEE1.0-1 数据集上进行对比实验, 实验结果如表 2 所列。

表 2 不同重叠事件抽取方法在 FewFC 和 DuEE1.0-1 数据集上的对比实验结果

		FewFC				DuEE1.0-1				
		PLMEE	CasEE	OneEE	MatEE	LGFEE	PLMEE	CasEE	OneEE	LGFEE
TI	P	83.7	89.4	88.7	—	88.5	85.8	90.3	87.7	87.8
	R	85.8	87.7	88.7	—	89.4	89.8	88.5	90.8	92.4
	F1	84.7	88.6	88.7	—	89.0	87.8	89.4	89.2	90.0
AI	P	74.3	72.8	75.4	—	75.7	76.0	72.5	72.4	73.8
	R	67.3	73.1	77.0	—	77.4	71.8	69.8	74.0	74.6
	F1	70.6	72.9	76.2	—	76.5	73.8	71.1	73.2	74.2
TC	P	75.6	77.9	79.1	80.4	79.3	85.6	89.9	87.6	87.5
	R	74.5	78.5	80.3	78.9	81.7	89.6	88.4	90.8	92.3
	F1	75.1	78.2	79.7	79.6	80.5	87.6	89.1	89.2	89.8
AC	P	72.5	71.3	74.0	74.6	73.6	75.5	71.6	69.5	70.9
	R	65.5	71.5	72.9	70.5	73.4	71.4	69.0	72.9	73.4
	F1	68.8	71.4	73.4	72.5	73.5	73.4	70.3	71.2	72.2

从表 2 的结果可以看出: 与具有先进性和代表性的重叠事件抽取方法 CasEE 和 OneEE 相比, 本文提出的 LGFEE 模型在 FewFC 和 DuEE1.0-1 数据集上的召回率和 F1 值均较高。

相较于 CasEE, 在事件类型分类 (TC) 任务上, LGFEE 模型在 FewFC 数据集上的召回率和 F1 值分别提高了 3.2% 和 2.3%, 在 DuEE1.0-1 数据集上的召回率和 F1 值分别提高了 3.9% 和 0.7%; 在论元角色分类 (AC) 任务上, LGFEE 模型在 FewFC 数据集上的召回率和 F1 值分别提高了 1.9% 和 2.1%, 在 DuEE1.0-1 数据集上的召回率和 F1 值分别提高了 4.4% 和 1.9%。原因主要在于, 基于联合学习的方法在事件抽取过程中共享参数, 减少了错误信息的传播, 可以更好地实现重叠事件抽取。

相较于 OneEE, 在论元识别 (AI) 任务上, LGFEE 模型在 FewFC 数据集上的召回率和 F1 值分别提高了 0.4% 和 0.3%, 在 DuEE1.0-1 数据集上的召回率和 F1 值分别提高了 0.6% 和 1.0%; 在事件类型分类 (TC) 任务上, LGFEE 模型在 FewFC 数据集上的召回率和 F1 值分别提高了 1.4% 和 0.8%, 在 DuEE1.0-1 数据集上的召回率和 F1 值分别提高了 1.5% 和 0.6%。原因主要在于, 基于不同语言粒度融合构建

文本表示, I-Phase 采取分层组合和传递的方式, 生成文本的表示, 有助于获取更全面的上下文语义信息; D-Phase 将长文本包含的抽象化语义信息传递给短文本, 作为 I-Phase 中对应片段的补充信息, 获得了更全面、准确的文本表示, 从而提升了整体性能。

相较于 MatEE, 本文提出的 LGFEE 模型在 FewFC 数据集上的召回率和 F1 值均较高。在事件类型分类 (TC) 任务上, LGFEE 模型在 FewFC 数据集上的召回率和 F1 值分别提高了 2.8% 和 0.9%; 在论元角色分类 (AC) 任务上, LGFEE 模型在 FewFC 数据集上的召回率和 F1 值分别提高了 2.9% 和 1.0%, 原因主要在于, 一方面, 基于联合学习的方法在事件抽取过程中减少了错误信息的传播; 另一方面, LGFEE 模型可以同时提取事件中的重叠触发词和重叠论元, 而 MatEE 仅能抽取重叠触发词。

在事件类型分类 (TC) 任务中, LGFEE 模型在 DuEE1.0-1 数据集上的性能优于 FewFC。主要原因是, 在事件类型判别任务中, 模型更关注捕捉不同事件类型之间的特征, 在 FewFC 数据集中, 10 个事件类型均属于金融领域, 具有相似的特征和语义; 而在 DuEE1.0-1 数据集中, 事件类型分别属于竞赛行为、产品行为、人生、司法行为 4 类, 不同类别间事件类型

差异较大,模型可能更容易区分。而在论元识别(AI)和论元角色分类(AC)中, LGFEE 模型在 FewFC 数据集上的性能优于 DuEE1.0-1,主要原因是 DuEE1.0-1 数据集中包含 31 种事件类型,62 种论元角色,模型的计算复杂度较高,更容易过拟合,导致性能受限。

(2)普通事件抽取方法在 FewFC 数据集上的实验结果

将本文方法与普通事件抽取方法在 FewFC 数据集上进行对比实验,实验结果如表 3 所列。

表 3 普通事件抽取方法在 FewFC 上的对比实验结果

Table 3 Experimental results of general event extraction methods on FewFC

		BERT-softmax	BERT-CRF	BERT-CRF-joint	LGFEE
TI	P	89.8	90.8	89.5	88.5
	R	79.0	80.8	79.8	89.4
	F1	84.0	85.5	84.4	89.0
AI	P	74.6	75.1	76.1	75.7
	R	62.8	64.3	63.5	77.4
	F1	68.2	69.3	69.2	76.5
TC	P	80.2	81.7	80.7	79.3
	R	61.8	63.6	63.0	81.7
	F1	69.8	71.5	70.8	80.5
AC	P	72.5	72.9	74.2	73.6
	R	60.2	61.8	61.2	73.4
	F1	65.8	66.9	67.1	73.5

从表 3 的结果可以看出:与普通事件抽取方法相比,本文的 LGFEE 模型在 FewFC 数据集上的召回率和 F1 值均较高,说明该模型对于解决重叠问题是有效的。

(3)在 FewFC-overlap 和 FewFC-normal 数据集上的实验结果

为了进一步验证本文所提模型的性能,本文分别在重叠事件数据集 FewFC-overlap 和非重叠事件数据集 FewFC-normal 上进行实验,实验结果分别如表 4 和表 5 所列。

表 4 不同方法在 FewFC-overlap 数据集上的实验结果

Table 4 Experimental results of different methods on FewFC-overlap

	TI	AI	TC	AC
BERT-softmax	76.5	56.1	49.0	53.5
BERT-CRF	77.9	61.0	52.4	58.4
BERT-CRF-joint	77.8	58.8	52.0	56.8
PLMEE	80.7	63.2	66.6	61.4
CasEE	89.0	71.5	74.9	70.3
OneEE	87.5	75.2	77.2	71.9
LGFEE	88.1	75.1	8.0	72.6

表 5 不同方法在 FewFC-normal 数据集上的实验结果

Table 5 Experimental results of different methods on FewFC-normal

	TI	TC	AI	AC
BERT-softmax	86.9	79.9	76.2	74.1
BERT-CRF	88.4	80.8	74.9	72.8
BERT-CRF-joint	86.9	79.9	76.1	74.0
PLMEE	86.4	79.7	75.7	74.0
CasEE	88.4	80.2	74.0	72.3
OneEE	89.1	80.6	76.6	73.9
LGFEE	89.1	81.7	77.5	74.2

从表 4 可以看出,对重叠事件,本文方法在事件类型分类

(TC)和论元角色分类(AC)任务中的性能优于基线方法。相较于普通事件抽取方法,本文采用词间关系识别的方法进行事件抽取,可以有效避免标签冲突,提升了重叠事件的抽取性能,而 BERT-CRF-joint 仅能为每个词分配一个标签,不能有效地解决事件的重叠问题。相较于已有的重叠事件抽取方法,本文方法基于语言粒度融合,学习到了语义更加丰富的句子表示,同时,采用联合学习的方式对重叠触发词和重叠论元进行有效的联合抽取,而 PLMEE 仅能对重叠论元进行抽取。CasEE 采用基于 Pipeline 的事件抽取方法,由于后一阶段依赖前一阶段,因此可能导致错误传播问题。

如表 5 所列,本文提出的 LGFEE 模型在普通事件数据集上整体仍然可以取得令人满意的结果。

5.2 消融实验

为了验证本文提出的 LGFEE 模型各个模块的性能,本节在 FewFC 数据集上进行了消融实验。

-MHA:在 LGFEE 的基础上,仅去除 I-Phase 和 D-Phase 的多头注意力机制。

-D-Phase:在 LGFEE 的基础上,仅去除逐层递减阶段 D-Phase。

-Gate1:在 LGFEE 的基础上,用加法操作代替门控机制 Gate1。

模型在 FewFC 数据集上的消融实验结果如表 6 所列。

表 6 消融实验结果

Table 6 Results of ablation experiment

		-MHA	-D-Phase	-Gate1	LGFEE
TI	P	87.0	87.6	88.3	88.5
	R	89.3	87.5	89.0	89.4
	F1	88.2	87.5	88.7	89.0
TC	P	76.1	75.7	78.7	79.3
	R	81.0	79.1	80.6	81.7
	F1	78.5	77.4	79.6	80.5
AI	P	74.4	72.6	74.3	75.7
	R	76.9	77.8	76.7	77.4
	F1	75.6	75.1	75.5	76.5
AC	P	72.5	71.2	72.8	73.6
	R	73.1	72.8	73.1	73.4
	F1	72.8	72.0	73.0	73.5

由表 6 的实验结果可以看出:

(1)去除任一模块,事件抽取 4 个子任务的 F1 值均会下降,说明引入任一模块均有利于事件抽取任务。

(2)-MHA 方法在 4 个子任务上的 F1 值相比本文方法均有下降,说明使用多头注意力可以增强文本间的信息交互。

(3)相较于其他模块,-D-Phase 方法的 F1 值下降最为明显。D-Phase 将长文本包含的抽象化语义信息传递给短文本,说明同时考虑多种语言粒度的信息,可以更全面地捕捉句子所包含的语义和上下文信息。

(4)-Gate1 方法相较于本文方法在 4 个子任务上的 F1 值均有下降,说明设计的门控模块 Gate1 对信息的过滤是有效的,可以保留更多有利于事件抽取的上下文信息。

5.3 层数对事件抽取的影响

为了验证 I-Phase 和 D-Phase 中不同层数 L 对性能的影响,本节针对不同层数在 FewFC 数据集上进行了实验。使用 F1 值对上述 4 个任务的性能进行评估,实验结果如表 7

所列。

表7 不同层数下的实验结果

Table 7 Experimental results with different number of layers

层数	TI	AI	TC	AC
16	88.5	75.6	79.6	72.7
14	88.5	76.0	79.5	72.5
12	89.0	76.5	80.5	73.5
10	88.7	75.6	80.3	73.0
8	88.5	75.7	79.6	73.1
6	87.8	76.0	78.9	73.2
4	88.2	76.0	79.3	73.1
2	87.8	76.6	79.4	73.1
1	87.6	75.6	78.9	73.0

由表7的实验结果可以看出,层数并不是越多越好,当

表8 案例分析样例

Table 8 Example of case analysis

文本	厉健律师提示,根据司法解释,暂定,在2017年4月12日至2018年5月10日买入上峰水泥股票,且在2018年5月11日后继续持有或卖出该股票的受损投资者可以索赔
黄金标签	事件1:【类型】投资;【触发词】买入;【论元】obj:上峰水泥 事件2:【类型】股份股权转让;【触发词】卖出;【论元】target-company:上峰水泥;collateral:股票 事件3:【类型】起诉;【触发词】索赔;【论元】obj-org:上峰水泥
CasEE	事件1:【类型】起诉;【触发词】索赔;【论元】obj-org:上峰水泥
OneEE	事件1:【类型】投资;【触发词】买入;【论元】obj:上峰水泥
LGFE	事件1:【类型】投资;【触发词】买入;【论元】obj:上峰水泥 事件2:【类型】起诉;【触发词】索赔;【论元】obj-org:上峰水泥

为了更好地发掘研究中存在的潜在问题,本文对实验结果进行了错误分析。通过此项工作,希望能够更加全面、客观

地评估本文实验结果的科学性和可靠性,为未来的相关研究提供有益的参考依据。表9列出了一个错误分析的样例。

表9 错误分析样例

Table 9 Example of error analysis

文本	黄金标签	LGFE
罗一鸣于2019年9月4日向公司董事会提交《茂名石化实华股份有限公司详式权益变动报告书》,告知公司因其成为神州永丰和东方永兴的控股股东,而实现对公司的间接收购,为公司实际控制人	事件:【类型】收购;【触发词】收购;【论元】sub-per:罗一鸣;date:2019年9月4日;obj-org:茂名石化实华股份有限公司	事件:【类型】收购;【触发词】收购;【论元】date:2019年9月4日;obj-org:东方永兴

在这个实例中,罗一鸣是通过成为神州永丰和东方永兴的控股股东,实现对茂名石化实华股份有限公司的收购。LGFE模型误将“东方永兴”预测为被收购公司,造成这个错误的原因可能是文本自身语义信息复杂且存在因果关系,增加了模型预测的难度。

结束语 本文针对重叠事件抽取问题,提出了一种基于语言粒度融合联合学习的重叠事件抽取方法。该方法提出的渐进式语言粒度融合模块,以 token 数目逐层递增和逐层递减的方式,分别构建包含不同 token 数目的片段表示,将短文本包含的离散信息传递给长文本,再将长文本包含的抽象化语义信息传递给短文本,建立基于语言粒度融合的句子表示。此外,提出了基于门控机制和注意力机制进行信息融合,生成融合不同粒度特征和目标事件感知的句子表示。预测阶段,融合了文本的位置信息,通过联合学习的方式识别词间关系,再利用关系标签解码,获得事件触发词、论元及事件类型和论元角色。在公开数据集上的实验结果表明,该模型在事件抽取方面优于目前已有的同类方法,说明本文方法对重叠事件的抽取是有效的。我们将在未来的研究工作中探索两个方面的改进。其一,将因果关系识别嵌入事件抽取模型中,以

层数 L 为 12 时,本文提出的模型综合性能最好。

5.4 实例分析

为了进一步说明本文方法的有效性,本节选取了具有先进性和代表性的 CasEE^[4] 和 OneEE^[7] 模型与本文模型进行比较,表8列出了一个实例分析。在这个例子中,“上峰水泥”在“投资”“股份股权转让”和“起诉”这3个事件中都充当论元,CasEE模型仅识别出“起诉”事件,OneEE模型仅识别出“投资”事件,而本文提出的 LGFE 模型可以同时识别出“起诉”和“投资”两个事件。这说明构建基于语言粒度融合的文本表示融合短文本包含的离散信息和长文本包含的抽象化语义信息,可以有效地获得较为丰富的片段表示。

提高事件抽取的准确性和完整性。其二,将进一步探索模型在处理长篇文本时计算时间较长的问题,寻求更有效的解决方案。

参考文献

- [1] HUANG K H, YANG M, PENG N. Biomedical event extraction with hierarchical knowledge graphs[C]// Findings of the Association for Computational Linguistics (EMNLP 2020). 2020: 1277-1285.
- [2] YANG H, CHEN Y, LIU K, et al. DCFE: a document-level Chinese financial event extraction system based on automatically labeled training data[C]// Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (ACL 2018). 2018: 50-55.
- [3] HALTERMAN A, KEITH K A, SARWAR S M, et al. Corpus-level evaluation for event QA: the IndiaPoliceEvents corpus covering the 2002 Gujarat violence[C]// Findings of the Association for Computational Linguistics (ACL-IJCNLP 2021). 2021: 4240-4253.
- [4] SHENG J, GUO S, YU B, et al. CaSEE: a joint learning frame-

- work with cascade decoding for overlapping event extraction [C]//Findings of the Association for Computational Linguistics: ACL-IJCNLP. 2021;164-174.
- [5] LIU X, LUO Z, HUANG H. Jointly multiple events extraction via attention-based graph information aggregation [C] // Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. 2018;1247-1256.
- [6] NGUYEN T H, CHO K, GRISHMAN R. Joint event extraction via recurrent neural networks [C] // Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics; Human Language Technologies. 2016;300-309.
- [7] CAO H, LI J, SU F, et al. OneEE: a one-stage framework for fast overlapping and nested event extraction [C] // Proceedings of the 29th International Conference on Computational Linguistics. 2022;1953-1964.
- [8] LIU W, MA Y W, PENG Y, et al. Chinese event extraction method based on graph attention and table pointer network [J]. Pattern Recognition and Artificial Intelligence, 2023, 36(5): 459-470.
- [9] XIN M M, MA L, HU B F. Research on text classification fusing multi-granularity information [J]. Computer Engineering and Applications, 2023, 59(9): 104-111.
- [10] LI Y, CUI L, YIN Y, et al. Multi-granularity optimization for non-autoregressive translation [C] // Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing. 2022;5073-5084.
- [11] DEVLIN J, CHANG M W, LEE K, et al. Bert: pre-training of deep bidirectional transformers for language understanding [C] // Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. 2019;4171-4186.
- [12] LI Q, LI J, SHENG J, et al. A survey on deep learning event extraction: approaches and applications [J/OL]. <https://doi.org/10.1109/TNNLS.2022.3213168>.
- [13] CHEN Y, XU L, LIU K, et al. Event extraction via dynamic multi-pooling convolutional neural networks [C] // Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing. 2015;167-176.
- [14] YANG S, FENG D, QIAO L, et al. Exploring pre-trained language models for event extraction and generation [C] // Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. 2019;5284-5294.
- [15] XU J, XU W, SUN M, et al. Extracting trigger-sharing events via an event matrix [C] // Findings of the Association for Computational Linguistics (EMNLP 2022). 2022;1189-1201.
- [16] LI Q, PENG H, LI J, et al. Reinforcement learning-based dialogue guided event extraction to exploit argument relations [J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2022, 30: 520-533.
- [17] 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. 2017; 6000-6010.
- [18] SUN Y, CHENG C, ZHANG Y, et al. Circle loss: a unified perspective of pair similarity optimization [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020;6398-6407.
- [19] ZHOU Y, CHEN Y, ZHAO J, et al. What the role is vs. what plays the role: semi-supervised event argument extraction via dual question answering [C] // Proceedings of the AAAI Conference on Artificial Intelligence. 2021;14638-14646.
- [20] LI X, LI F, PAN L, et al. DuEE: a large-scale dataset for chinese event extraction in real-world scenarios [C] // Natural Language Processing and Chinese Computing: 9th CCF International Conference (NLPCC 2020). 2020;534-545.
- [21] DU X, CARDIE C. Document-level event role filler extraction using multi-granularity contextualized encoding [C] // Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. 2020;8010-8020.



YAN Jingtao, born in 1999, postgraduate. Her main research interest is natural language processing.



LI Yang, born in 1988, Ph.D, associate professor, is a member of CCF (No. 33621G). Her main research interests include text sentiment analysis and text mining.

(责任编辑:何杨)