

# 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<sup>1</sup>, 李旸<sup>2</sup>, 王素格<sup>1,3</sup>, 廖健<sup>1</sup>, 普瑞丽<sup>1</sup>, 潘邦泽<sup>1</sup>, 李德玉<sup>1,3</sup>

1.山西大学 计算机与信息技术学院, 山西 太原 030006

2.山西财经大学 金融学院, 山西 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, 山西 太原 030006

wsg@sxu.edu.cn

## 摘要

事件抽取是一项重要的信息抽取任务,旨在从非结构化文本中抽取结构化的事件信息。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,未考虑重叠事件,然而,在真实的场景下,重叠事件是不可避免的。因此,该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,该方法利用预训练语言模型BERT对句子进行初始表示,再以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,旨在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。进一步,使用门控机制,获得融合不同粒度特征和目标事件感知的句子表示。最后,在预测阶段,通过计算句子中词间距离感知得分,并行的预测词间的片段和角色关系,再利用关系标签解码,基于联合学习获得了事件触发词、论元及事件类型和论元角色。在公开的事件抽取数据集FewFC上进行实验,结果表明该文提出的方法对重叠事件的抽取是有效的。

**关键词:** 重叠事件抽取; 语言粒度融合; 联合学习

## Overlapping Event Extraction Method of Language Granularity Fusion Based on Joint Learning

Jingtao Yan<sup>1</sup>, Yang Li<sup>2</sup>, Suge Wang<sup>1,3</sup>, Jian Liao<sup>1</sup>, Ruili Pu<sup>1</sup>, Bangze Pan<sup>1</sup>, Deyu Li<sup>1,3</sup>

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

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

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

wsg@sxu.edu.cn

## Abstract

Event extraction is an important information extraction task, which aims to extract structured event information from unstructured text. Most of the existing event extraction methods assume that only one event occurs in a sentence, without considering overlapping events. However, overlapping events are inevitable in real scenarios. Therefore, this paper proposes an overlapping event extraction method of language granularity fusion based on joint learning. This method uses pre-trained language model BERT to perform initial representation of sentences, and then constructs fragment representations containing different token numbers in the way of increasing and decreasing token numbers layer by layer, respectively, to transfer discrete information contained in short texts to long texts. Then the abstract semantic information contained in the long text is transferred to the short text, and the sentence representation based on the language granularity fusion is established. Further, a gating mechanism is used to obtain a sentence representation that integrates multi-granularity features and

target event perception. Finally, in the prediction stage, by calculating the inter-word distance perception score in the sentence, the fragment and role relationship between the words are predicted in parallel, and then the relation label is decoded to obtain the event trigger word, argument, event type and argument role based on joint learning. Experiments on a public event extraction dataset FewFC show that the proposed method is effective for extracting overlapping events.

**Keywords:** Overlapping event extraction , Language granularity fusion , Joint learning

## 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。它在医疗 (Huang et al., 2020)、金融 (Yang et al., 2018)、司法 (Halterman et al., 2021) 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、金额等一系列核心要素, 可为决策者提供重要的参考信息, 使他们及时根据实际情况做出科学决策。

事件抽取包括普通事件抽取和重叠事件抽取。(Sheng et al., 2021) 将重叠事件抽取归纳为三种方式: (1) 一个词在多个事件中作为触发词; (2) 一个词作为论元, 在一个事件中扮演不同的角色; (3) 一个词作为论元, 在多个事件中扮演不同的角色。方式 (1) 称为触发词重叠事件抽取问题, 方式 (2) 和 (3) 称为论元重叠事件抽取问题。如图 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。

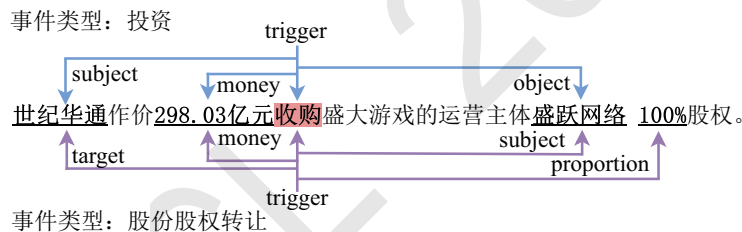


图 1. 重叠事件抽取实例

传统的事件抽取方法 (Nguyen et al., 2016; Liu et al., 2018) 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, 研究者开展对重叠事件抽取的相关研究, 主要采用基于 Pipeline 范式 (Yang et al., 2019; Sheng et al., 2021) 和基于 Joint 范式的事件抽取方法 (Cao et al., 2022)。基于 Pipeline 范式的事件抽取方法, 是在不同的阶段依次抽取触发词和论元, 忽略了触发词和论元之间的依赖关系, 容易导致错误传播。而现有的基于 Joint 范式的事件抽取方法 (Cao et al., 2022), 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的有效表示。

为此, 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模型 (LGFEE)。首先, 采用 BERT (Devlin et al., 2018) 作为编码器获得句子表示; 其次, 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; 然后, 基于门控机制和注意力机制进行信息融合, 获得融合不同粒度特征和目标事件感知的句

©2023 中国计算语言学大会

根据《Creative Commons Attribution 4.0 International License》许可出版

基金项目: 国家自然科学基金项目(62106130, 62076158, 62072294, 62272286); 山西省基础研究计划(20210302124084); 山西省高等学校科技创新项目(2021L284); CCF-智谱AI大模型基金(CCF-Zhipu202310)

子表示；最后，计算句子中词间的距离感知得分，并行的预测词间的片段关系和角色关系，再利用关系标签解码，获得事件的触发词、论元及事件类型和论元角色。

本文的贡献主要包括以下三个方面：

(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模型 (LGFEE)，用于重叠事件抽取。

(2) 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方式构建不同粒度的片段表示，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。

(3) 在公开的金融事件抽取数据集FewFC (Zhou et al., 2021)进行实验，实验结果验证了模型的有效性。

## 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四个子任务。依据四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 (Li et al., 2022a)。

### 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论元进行抽取，并对论元角色进行分类。(Chen et al., 2015)提出了DMCNN模型，该模型使用两个动态多池卷积神经网络分别对触发词和论元进行分类，将词语级语义特征和句子级语义特征结合，对触发词进行识别和分类，进一步，若句子中存在触发词再进论元的识别和分类。(Yang et al., 2019)提出了PLMEE模型，该模型针对特定的角色分别抽取论元，可以对重叠论元进行抽取。(Sheng et al., 2021)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，再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，容易出现错误传播问题。

### 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，该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四个任务。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性能。(Nguyen et al., 2016)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，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，引入记忆向量和记忆矩阵，存储标记过程中的预测信息。(Liu et al., 2018)提出JMEE模型，该模型将句子从序列的模式转换成句法依存图的模式，并在句法依存图上使用GCN进行信息聚合，再采用联合方式，抽取事件触发词和论元。(Li et al., 2022b)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 et al., 2022)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抽取。

## 3 基于联合学习的语言粒度融合的重叠事件抽取模型 (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模型。首先，使用预训练语言模型BERT作为编码器，获得输入句子的词的嵌入表示；其次，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，建立基于语言粒度融合的句子表示；然后，基于门控机制和注意力机制进行信息融合，将事件信息和融合了不同粒度特征的文本表示融合；最后，通过联合学习预测词间的片段关系和角色关系，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。模型架构如图 2所示。

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 $H^0 = (h_1^0, h_2^0, \dots, h_N^0) \in R^{N \times d}$ 。

### 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，本文基于语言粒度融合构建句子表示。具体来说，分别以token数目逐层递增 (Incremental-Phase, I-Phase) 和逐层递减 (Decremental-Phase, D-

Phase) 的方式, 依次构建不同粒度的片段表示。这两个阶段均包含 $L$ 层, 将I-Phase中第 $l$ 层的片段表示记作 $H^l = (h_1^l, h_2^l, \dots, h_T^l)$ , 将D-Phase中第 $l$ 层的片段表示记作 $H'^l = (h_1'^l, h_2'^l, \dots, h_T'^l)$ , 其中,  $T$ 是片段的数量,  $l$ 表示每个片段的token数, 每一层的结构如图 3所示。

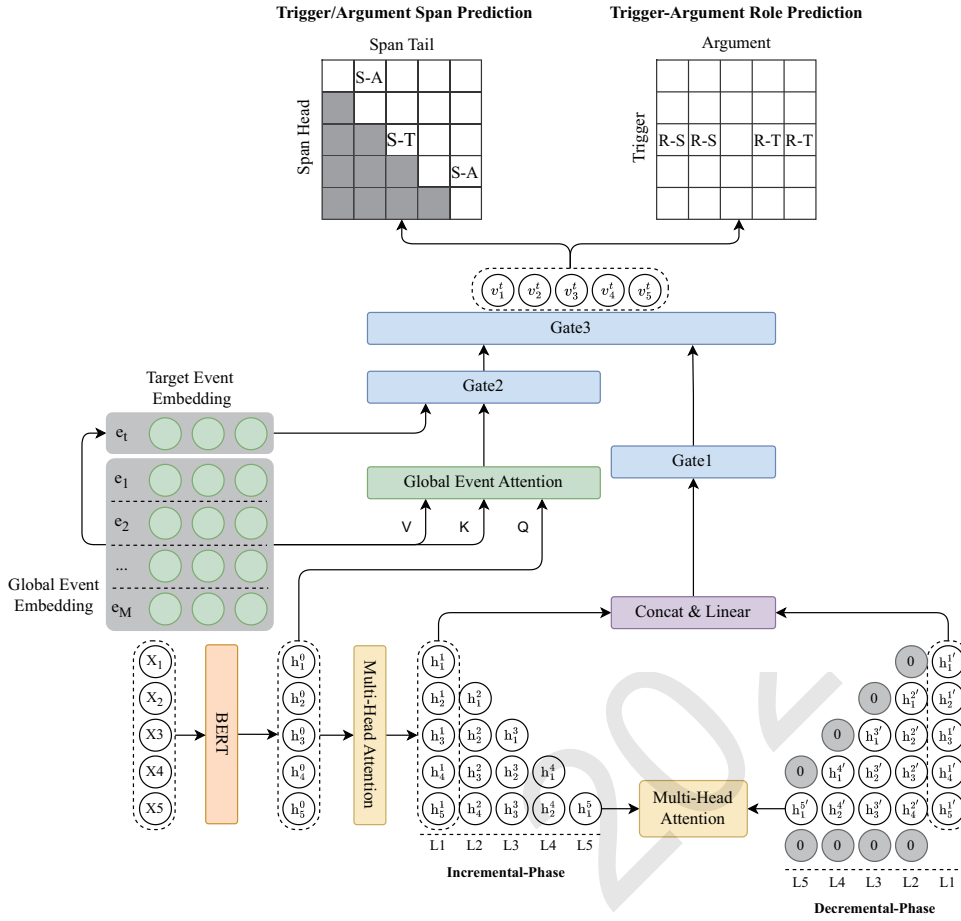


图 2. 基于联合学习的语言粒度融合的重叠事件抽取模型 (LGFEE)

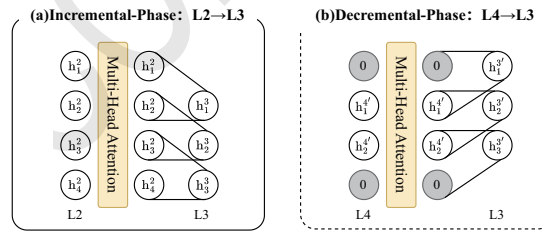


图 3. Incremental-Phase和Decremental-Phase中构建L3层的详细结构

在I-Phase和D-Phase的每一层中, 首先利用Transformer (Vaswani et al., 2017)的多头注意力机制增强片段表示, 句子在第 $l$ 层经过多头注意力的表示 $H^{l-multihead}$ 如下:

$$H^{l-multihead} = \text{MultiHead}(Q, K, V) = \text{Concat}(head_1, \dots, head_h)W^O \quad (1)$$

$$head_i = \text{Attention}(QW_i^Q, KW_i^K, VW_i^V) = \text{softmax}\left(\frac{QW_i^Q(KW_i^K)^T}{\sqrt{d_k}}\right)VW_i^V \quad (2)$$

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

**I-Phase:** 由于句子中的短文本包含的信息相对较少，无法直接表达上下文的语义，需要通过模型的分层组合和传递，逐步形成更全面、准确的长文本语义信息。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，L1层的片段表示对应于1-token的片段，对于更高的层，为了避免破坏连续的上下文，利用内核大小为2的卷积神经网络，以token数目递增的方式从第 $l-1$ 层迭代地生成连续的 $l$ -token片段表示。I-Phase中第 $l$ 层中的片段表示为 $H^l$ ，其计算方式如下：

$$H^l = \begin{cases} f(H^0) & , l = 1 \\ \text{Conv}(f(H^{l-1})) & , 1 < l \leq L \end{cases} \quad (3)$$

其中， $f(\cdot)$ 表示公式(1)的简写。此外，使用ReLU和Norm层获得最终的片段表示。

**D-Phase:** 由于在同一上下文中，长文本的特征可以提供与短文本特征互补的上下文信息。因此，D-Phase的目标是将长文本逐层分解成不同长度的片段，将长文本包含的抽象化语义信息传递给短文本。具体来说，以I-Phase中第 $L$ 层的输出作为D-Phase第 $L$ 层的输入， $H^{L'}$ 是 $H^L$ 经过多头注意力层增强得到的片段表示。第二阶段采用token数目递减的方式，适当的填充零向量，叠加内核大小为2的卷积神经网络，从第 $l-1$ 层迭代地构建连续的 $l$ -token片段表示。如图3(b)所示，例如，D-Phase中L3层的片段表示为 $(h_1^{3'}, h_2^{3'}, h_3^{3'})$ ，它是基于L4层的片段表示构建的，即 $(0, h_1^{4'}, h_2^{4'}, 0)$ 。类似的，D-Phase中第 $l$ 层的片段表示 $H^{l'}$ 可以由公式(4)获得：

$$H^{l'} = \begin{cases} f(H^l) & , l = L \\ \text{Conv}'(f(H^{l+1'})) & , 1 \leq l < L \end{cases} \quad (4)$$

最后，将I-Phase与D-Phase中L1层的句子表示拼接，作为融合不同粒度特征的句子表示 $H^{\text{Multi-granularity}}$ ：

$$H^{\text{Multi-granularity}} = W_c[H^1; H^{1'}] + b_c \quad (5)$$

其中， $W_c$ 和 $b_c$ 是可学习的参数矩阵。

### 3.2 基于门控机制和注意力机制的信息融合

由于本文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。

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

$$H = \text{Gate1}(H^0, H^{\text{Multi-granularity}}) = g_1 \odot H^0 + (1 - g_1) \odot H^{\text{Multi-granularity}} \quad (6)$$

其中， $W_{g_1}$ 和 $b_{g_1}$ 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。

参考(Cao et al., 2022)，为了减少计算资源，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 $E = (e_1, e_2, \dots, e_M) \in R^{M \times d}$ ，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 $E^{\text{global}}$ ：

$$Q, K, V = H^0 W^Q, E W^K, E W^V \quad (7)$$

$$E^{\text{global}} = \text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_h}}\right)V \quad (8)$$

其中， $W^Q$ 、 $W^K$ 、 $W^V$ 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 $E^{global}$ 的基础上，融合目标事件类型嵌入 $e_t$ ，生成目标事件感知的句子表示 $H^t$ ：

$$H^t = \text{Gate2}(E^{global}, e_t) = g_2 \odot E^{global} + (1 - g_2) \odot e_t \quad (9)$$

$$g_2 = \text{Sigmod}(W_{g_2}[E^{global}; e_t] + b_{g_2}) \quad (10)$$

其中， $e_t \in E$ 表示目标事件类型嵌入， $W_{g_2}$ 和 $b_{g_2}$ 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，将公式 (9)和公式 (6)分别得到的句子表示 $H^t$ 和 $H$ 进一步融合，得到融合目标事件 $e_t$ 的句子最终表示 $V^t = (v_1^t, v_2^t, \dots, v_N^t) \in R^{N \times d}$ ，其中， $v_i^t$ 是句子中词的嵌入表示。

$$V^t = \text{Gate3}(H^t, H) = g_3 \odot H^t + (1 - g_3) \odot H \quad (11)$$

$$g_3 = \text{Sigmod}(W_{g_3}[H^t; H] + b_{g_3}) \quad (12)$$

其中， $W_{g_3}$ 和 $b_{g_3}$ 是可学习的参数矩阵。

### 3.3 片段/角色标签预测

本文参照 (Cao et al., 2022)的工作，将事件抽取任务转化为目标事件的词间关系检测任务，该任务旨在判别句子中的每个词对  $(x_i, x_j)$  的片段关系 $S$ 和角色关系 $R$ ，其中，片段关系 $S$ 表示 $x_i$ 和 $x_j$ 分别是触发词片段 ( $S - T$ ) 或论元片段 ( $S - A$ ) 的起始和结束词，角色关系 $R$ 表示包含 $x_j$ 的论元在包含 $x_i$ 的触发词所触发的事件中扮演某个论元角色，记作 $R - *$ ，其中 $*$ 表示论元角色。为了更好的联合预测词对之间的片段关系 $s \in S$ 和角色关系 $r \in R$ ，将词对表示和距离信息融合，对每个词对  $(x_i, x_j)$  的嵌入表示  $(v_i^t, v_j^t)$ ，分别计算片段分数 $c_{ij}^s$ 和角色分数 $c_{ij}^r$ ，计算方式如下：

$$c_{ij}^s = \text{Score}(W_{s_1}v_i^t, W_{s_2}v_j^t) \quad (13)$$

$$c_{ij}^r = \text{Score}(W_{r_1}v_i^t, W_{r_2}v_j^t) \quad (14)$$

其中，Score为距离感知评分函数， $W_{s_1}$ 、 $W_{s_2}$ 、 $W_{r_1}$ 、 $W_{r_2}$ 是可学习的参数矩阵。

最后，联合训练片段分数 $c_{ij}^s$ 和角色分数 $c_{ij}^r$ ，训练目标是 minimized 事件类型集合 $E$ 中的所有事件的Circle loss (Sun et al., 2020)，得到总损失 $\mathcal{L}$ 如下：

$$\begin{aligned} \mathcal{L} = & \sum_{t \in E} \left\{ \sum_{s \in S} [\log(1 + \sum_{(i,j) \in \Omega^s} e^{-c_{ij}^s}) + \log(1 + \sum_{(i,j) \notin \Omega^s} e^{c_{ij}^s})] \right. \\ & \left. + \sum_{r \in R} [\log(1 + \sum_{(i,j) \in \Omega^r} e^{-c_{ij}^r}) + \log(1 + \sum_{(i,j) \notin \Omega^r} e^{c_{ij}^r})] \right\} \end{aligned} \quad (15)$$

其中， $\Omega^s$ 、 $\Omega^r$ 分别表示片段关系 $s$ 和角色关系 $r$ 的词对集合，关注分数大于0的词对。

### 3.4 关系标签解码

通过第 3.3节的片段/角色标签预测，一方面，可以利用目标事件类型的触发词片段 ( $S - T$ ) 或论元片段 ( $S - A$ ) 的始末索引，解码得到触发词或论元的片段；另一方面，可以利用获得的角色关系 ( $R - *$ ) 进一步匹配触发词和论元。在图 4所示的实例中，“收购”为触发词片段，“世纪华通”和“盛跃网络”为论元片段；“世纪华通”和“盛跃网络”表示触发词“收购”所触发的“投资”事件中的论元，对应论元角色分别为“投资方 ( $R - S$ )”和“被投资方 ( $R - O$ )”。

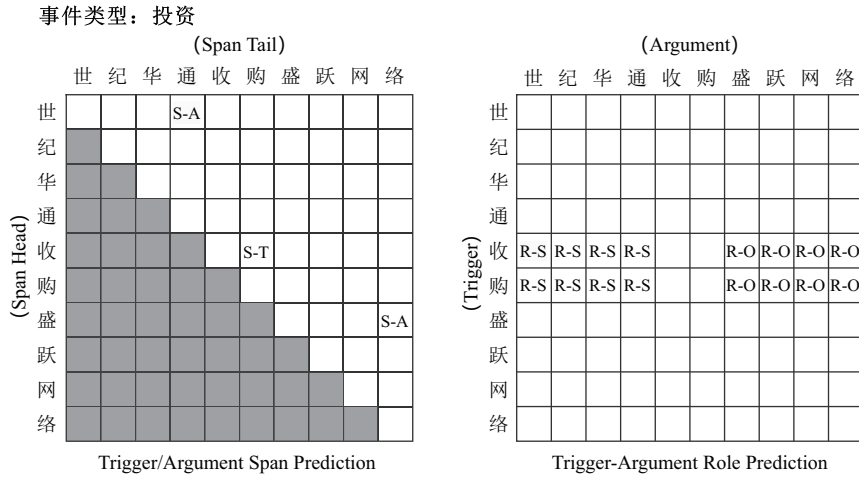


图 4. 关系标签预测结果

## 4 实验

### 4.1 数据集与评价指标

**数据集：** 本文选择中文金融事件抽取数据集FewFC (Zhou et al., 2021)，该数据集中标注了10个事件类型和18种论元角色，具有触发词重叠和论元重叠问题的句子分别占13.5%和21.7%。数据集中训练集、验证集和测试集的比例为8: 1: 1，统计情况如表 1所示。

数据集	重叠	普通	句子	事件
训练集	1560	5625	7185	10277
验证集	205	694	899	1281
测试集	210	688	898	1332
合计	1975	7007	8982	12890

表 1. 数据集统计结果

**评价指标：** 本文按照 (Du and Cardie, 2020; Sheng et al., 2021; Cao et al., 2022)的严格匹配评价标准，对事件抽取中的四个子任务：触发词识别 (TI)、论元识别 (AI)、事件类型分类 (TC)、论元角色分类 (AC)，采用准确率 (P)、召回率 (R) 和F1值 (F1) 进行评估。

### 4.2 基线方法

我们将本文提出的模型与最先进的基线方法进行了综合比较，基线可以分为以下两组：

(1) 普通事件抽取方法：这些方法将事件抽取任务转换为序列标注任务，为每个词分配一个标签。BERT-softmax (Devlin et al., 2018)，该方法采用BERT学习文本表示，并使用隐层表示对触发词和论元进行分类。BERT-CRF，该方法在BERT上添加了条件随机场 (CRF) 捕获标签依赖关系，(Du and Cardie, 2020)将其用于文档级事件抽取。BERT-CRF-joint，该方法将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，无法解决事件抽取的重叠问题。

(2) 重叠事件抽取方法：PLMEE (Yang et al., 2019)，该方法针对特定的角色分别抽取论元，可以抽取出重叠论元。CasEE (Sheng et al., 2021)，该方法依次执行事件类型检测、触发词抽取和论元抽取，根据前一步的预测结果抽取了重叠事件目标。OneEE (Cao et al., 2022)，该方法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，并行的进行网格标注抽取重叠目标。

### 4.3 参数设置

本文模型利用Chinese-BERT模型对句子进行编码，隐层向量维度为768，优化器采用AdamW。BERT的学习率为 $2e-5$ ，其他模块的学习率为 $1e-3$ 。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，在FewFC数据集上训练模型30次。

## 5 实验结果及分析

### 5.1 比较实验及结果分析

为了验证本文提出的LGFEE模型的有效性，进行了下面三个实验：（1）在FewFC数据集上进行事件抽取的整体性能实验；（2）在FewFC数据集的重叠事件上进行实验；（3）在FewFC数据集的普通事件上进行实验。利用第4.2节介绍的比较方法，与本文方法进行对比。

#### 5.1.1 事件抽取的整体性能实验

为了验证本文提出的LGFEE模型在事件抽取中的整体性能，在FewFC数据集上进行了实验，其中触发词识别（TI）、论元识别（AI）、事件类型分类（TC）、论元角色分类（AC）的对比实验结果如表2所示。

		BERT-softmax	BERT-CRF	BERT-CRF-joint	PLMEE	MQAEE	CasEE	OneEE	LGFEE
TI	P	89.8	<b>90.8</b>	89.5	83.7	89.1	89.4	88.7	88.5
	R	79.0	80.8	79.8	85.8	85.5	87.7	88.7	<b>89.4</b>
	F1	84.0	85.5	84.4	84.7	87.4	88.6	88.7	<b>89.0</b>
AI	P	74.6	75.1	<b>76.1</b>	74.3	70.3	72.8	75.4	75.7
	R	62.8	64.3	63.5	67.3	68.3	73.1	77.0	<b>77.4</b>
	F1	68.2	69.3	69.2	70.6	69.3	72.9	76.2	<b>76.5</b>
TC	P	80.2	<b>81.7</b>	80.7	75.6	79.7	77.9	79.1	79.3
	R	61.8	63.6	63.0	74.5	76.1	78.5	80.3	<b>81.7</b>
	F1	69.8	71.5	70.8	75.1	77.8	78.2	79.7	<b>80.5</b>
AC	P	72.5	72.9	<b>74.2</b>	72.5	68.2	71.3	74.0	73.6
	R	60.2	61.8	61.2	65.5	66.5	71.5	72.9	<b>73.4</b>
	F1	65.8	66.9	67.1	68.8	67.3	71.4	73.4	<b>73.5</b>

表 2. 不同方法在FewFC数据集中的实验结果(%)

从表2的结果可以看出：

(1) 本文提出的方法在召回率（R）和F1值上均优于其他对比方法。

(2) 与普通事件抽取方法相比，本文的LGFEE模型在F1值上取得了较好的性能，说明该模型对于解决重叠问题是有效的。相较于BERT-CRF-joint，本文的LGFEE模型在事件类型分类（TC）任务上的F1值提高了9.7%，在论元角色分类（AC）任务上的F1值提高了6.4%。

(3) 与具有先进性和代表性的重叠事件抽取方法CasEE和OneEE相比，本文的LGFEE模型在召回率和F1值上均取得了良好的性能。具体而言，相较于CasEE，本文的LGFEE模型在事件类型分类（TC）任务上的召回率和F1值分别提高了3.2%和2.3%，在论元角色分类（AC）任务上分别提高了1.9%和2.1%；相较于OneEE，本文的LGFEE模型在事件类型分类（TC）任务上的召回率和F1值也分别提高了1.4%和0.8%。主要原因是，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，可以更好的实现重叠事件抽取，同时，基于语言粒度融合构建文本表示，有助于学习更加丰富的语义信息，提升了整体性能。

#### 5.1.2 在重叠事件数据上的事件抽取性能

为了进一步验证本文所提模型LGFEE在重叠事件上的性能，在FewFC数据集中的重叠事件数据上进行了实验，对比实验结果如表3所示。

	BERT-softmax	BERT-CRF	BERT-CRF-joint	PLMEE	MQAEE	CasEE	OneEE	LGFE
TI	76.5	77.9	77.8	80.7	83.6	<b>89.0</b>	87.5	88.1
AI	56.1	61.0	58.8	63.2	62.1	71.5	<b>75.2</b>	75.1
TC	49.0	52.4	52.0	66.6	70.4	74.9	77.2	<b>78.0</b>
AC	53.5	58.4	56.8	61.4	60.1	70.3	71.9	<b>72.6</b>

表 3. 不同方法在FewFC中重叠事件数据上的实验结果(%)

对重叠事件来说, 该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, 无法解决事件的重叠问题。相较于已有的重叠事件抽取方法, 本文基于语言粒度融合, 学习到语义更丰富的句子表示, 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地抽取, 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, 后一阶段依赖于前一阶段, 有可能导致错误传播问题。

### 5.1.3 在普通事件数据上的事件抽取性能实验

如表 4所示, 本文提出的LGFE模型在普通事件数据上的性能仍可取得令人满意的结果。

	BERT-softmax	BERT-CRF	BERT-CRF-joint	PLMEE	MQAEE	CasEE	OneEE	LGFE
TI	86.9	88.4	86.9	86.4	89.0	88.4	<b>89.1</b>	<b>89.1</b>
AI	76.2	74.9	76.1	75.7	74.2	74.0	76.6	<b>77.5</b>
TC	79.9	80.8	79.9	79.7	<b>82.0</b>	80.2	80.6	81.7
AC	74.1	72.8	74.0	74.0	72.3	72.3	73.9	<b>74.2</b>

表 4. 不同方法在FewFC中普通事件数据集上的实验结果(%)

## 5.2 消融实验

为了验证本文提出的LGFE模型各个模块的性能, 本节进行了消融实验。

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

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

-Gate1: 在LGFE的基础上, 用加法操作代替门控机制Gate1。

模型消融在FewFC数据集的实验结果如表 5所示。

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

表 5. 模型消融实验结果(%)

由表 5的实验结果可以看出:

(1) 去除任一模块, 事件抽取四个子任务的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, 它在四个子任务上的F1值相较本文的方法均有下降, 说明使用多头注意力可以增强文本间的信息交互。

(3) -D-Phase方法, 相较于其他模块, 它的F1值下降最为明显, 说明D-Phase将长文本包含的抽象化语义信息传递给短文本, 提供了有利的上下文信息。

(4) -Gate1方法, 相较本文的方法在四个子任务上的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 $L$ 对模型性能的影响, 本节针对不同层数进行了实验, 使用F1指标对上述四个任务的性能进行评估, 实验结果如表 6所示。

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

表 6. 不同层数的实验结果比较(%)

由表 6的实验结果可以看出, 当层数 $L$ 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 (Sheng et al., 2021)、OneEE (Cao et al., 2022)模型与本文模型进行比较, 表 7展示了一个实例分析。

句子	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】ob:上峰水泥。 事件2: 【类型】起诉; 【触发词】索赔; 【论元】obj-org:上峰水泥。

表 7. 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”这三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“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。

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 8展示了一个错误预测的实例。

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。

表 8. 错误分析实例

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。LGFEE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在于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。

## 6 结论

本文针对重叠事件抽取问题，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，该方法首先利用预训练语言模型BERT对句子进行初始表示。其次，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。然后，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。最后，在预测阶段，融合文本的位置信息，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。本文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到事件抽取模型中，以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文本时计算时间较长的问题，并寻求更有效的解决方案。

## 参考文献

- Hu Cao, Jingye Li, Fangfang Su, Fei Li, Hao Fei, Shengqiong Wu, Bobo Li, Liang Zhao, and Donghong Ji. 2022. Oneee: a one-stage framework for fast overlapping and nested event extraction. In *Proceedings of the 29th International Conference on Computational Linguistics*, pages 1953–1964.
- Yubo Chen, Liheng Xu, Kang Liu, Daojian Zeng, and Jun Zhao. 2015. Event extraction via dynamic multi-pooling convolutional neural networks. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 167–176.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.
- Xinya Du and Claire Cardie. 2020. Document-level event role filler extraction using multi-granularity contextualized encoding. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 8010–8020.
- Andrew Halterman, Katherine Keith, Sheikh Sarwar, and Brendan O’Connor. 2021. Corpus-level evaluation for event qa: the indiapoliceevents corpus covering the 2002 gujarat violence. *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*.
- Kung-Hsiang Huang, Mu Yang, and Nanyun Peng. 2020. Biomedical event extraction with hierarchical knowledge graphs. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 1277–1285.
- Qian Li, Jianxin Li, Jiawei Sheng, Shiyao Cui, Jia Wu, Yiming Hei, Hao Peng, Shu Guo, Lihong Wang, Amin Beheshti, et al. 2022a. A survey on deep learning event extraction: approaches and applications. *IEEE Transactions on Neural Networks and Learning Systems*.
- Qian Li, Hao Peng, Jianxin Li, Jia Wu, Yuanxing Ning, Lihong Wang, Philip S. Yu, and Zheng Wang. 2022b. Reinforcement learning-based dialogue guided event extraction to exploit argument relations. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 30:520–533.
- Xiao Liu, Zhunchen Luo, and He-Yan Huang. 2018. Jointly multiple events extraction via attention-based graph information aggregation. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 1247–1256.
- Thien Huu Nguyen, Kyunghyun Cho, and Ralph Grishman. 2016. Joint event extraction via recurrent neural networks. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 300–309.
- Jiawei Sheng, Shu Guo, Bowen Yu, Qian Li, Yiming Hei, Lihong Wang, Tingwen Liu, and Hongbo Xu. 2021. Casee: a joint learning framework with cascade decoding for overlapping event extraction. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 164–174.

- Yifan Sun, Changmao Cheng, Yuhan Zhang, Chi Zhang, Liang Zheng, Zhongdao Wang, and Yichen Wei. 2020. Circle loss: A unified perspective of pair similarity optimization. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 6398–6407.
- Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Advances in Neural Information Processing Systems*, 30.
- Hang Yang, Yubo Chen, Kang Liu, Yang Xiao, and Jun Zhao. 2018. Dcfee: a document-level chinese financial event extraction system based on automatically labeled training data. In *Proceedings of ACL 2018, System Demonstrations*, pages 50–55.
- Sen Yang, Dawei Feng, Linbo Qiao, Zhigang Kan, and Dongsheng Li. 2019. Exploring pre-trained language models for event extraction and generation. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 5284–5294.
- Zhiwei Yang, Jing Ma, Hechang Chen, Yunke Zhang, and Yi Chang. 2021. Hitrans: a hierarchical transformer network for nested named entity recognition. In *Findings of the Association for Computational Linguistics: EMNLP 2021*, pages 124–132.
- Yang Zhou, Yubo Chen, Jun Zhao, Yin Wu, Jiexin Xu, and Jinlong Li. 2021. What the role is vs. what plays the role: semi-supervised event argument extraction via dual question answering. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 14638–14646.