

第 5 章 基于分离-融合范式的跨事件信息构建与利用

以 ACE2005 数据集为代表的句子级别数据集被主流的事件抽取研究广泛采用,但随着真实世界文本内容的复杂性不断增加,事件要素信息越来越多地跨越出现在一个文档中的多个句子中。因此,研究同时适用于句子级别和文档级别事件要素抽取的通用模型成为新的研究热点和挑战。与上一章节不同,该研究遵循章节2.2所述的关于事件要素抽取的第二种定义,即不使用实体提及的标注信息,而在预测要素类型的同时识别出其文段范围。

现有通用文本级别事件要素抽取模型存在各自的局限性,而基于多词元链接的建模方法能够有效避免这些模型的内在局限,并展现出良好的性能潜力。因此,本章研究利用多词元链接方法进行通用文本级别的事件要素抽取。然而该方法中的单一事件建模方式缺乏对跨事件依赖关系的利用,多事件建模方式需要更长的要素类型序列编码和额外的链接操作,增加了预测推理的难度。为此,本章提出了一种基于分离-融合范式的跨事件信息构建与利用方法,并相应提出了一种新的多词元链接模型,避免了当前方法的各自局限,且实现了从不同级别文本中高效利用跨事件依赖信息。实验结果证明了本章提出的模型在所有句子级和文档级数据集上都优于目前性能最优的事件要素抽取模型。

5.1 引言

最近,基于抽取式^[95,129-131]和生成式^[123,125,154]的提示学习方法在同时适用于句子级别和文本级别的事件要素抽取任务上取得了显著的性能提升。然而,前者在抽取具有相同要素类型的多个事件要素时,受到提示模版中预先设定的重复要素槽位数量的限制,后者在抽取长距离事件要素的性能表现不佳。此外,这两类方法的性能表现均较为依赖于提示模版的设计质量。

因此,不同于基于提示学习的方法,最近的一些研究^[74,155,156]将输入文本和所有涉及的要素类型串接成自然语言序列,然后联合编码,并进一步实施多词元¹链接操作,从而实现事件要素抽取或通用信息抽取的建模。其中,所有事件要素文段和要素类型的并行链接操作能够直接高效地抽取具有相同要素类型的多个事件要素,并提升

¹ “词元”对应于英文表达中的“token”,这里特指经过预训练语言模型分词器分词后得到的语义单元。

长文本中分散的事件要素间的交互。此外，该类方法不需要精心设计的提示模版。根据每次同时抽取的事件数量，可以进一步将该类方法划分为两种建模方式：单事件抽取和多事件抽取。

第一种方式的研究^[74,156]将待抽取文本和只在目标事件中出现的要素类型串接后作为输入，然后实施两种链接操作去建模目标事件。如图5.1所示，其中被红色标记的为事件触发词，被黄色标记的为事件要素类型，不同颜色的箭头表示不同种类的链接操作，事件要素“James McDade”在 *Die* 事件和 *Attack* 事件中扮演了不同的要素类型，其抽取建模过程是分离的。虽然这种单事件抽取的方式便于进行简单的链接预测，但其忽略了建模不同事件间的显著关联依赖^[122,130,131]。

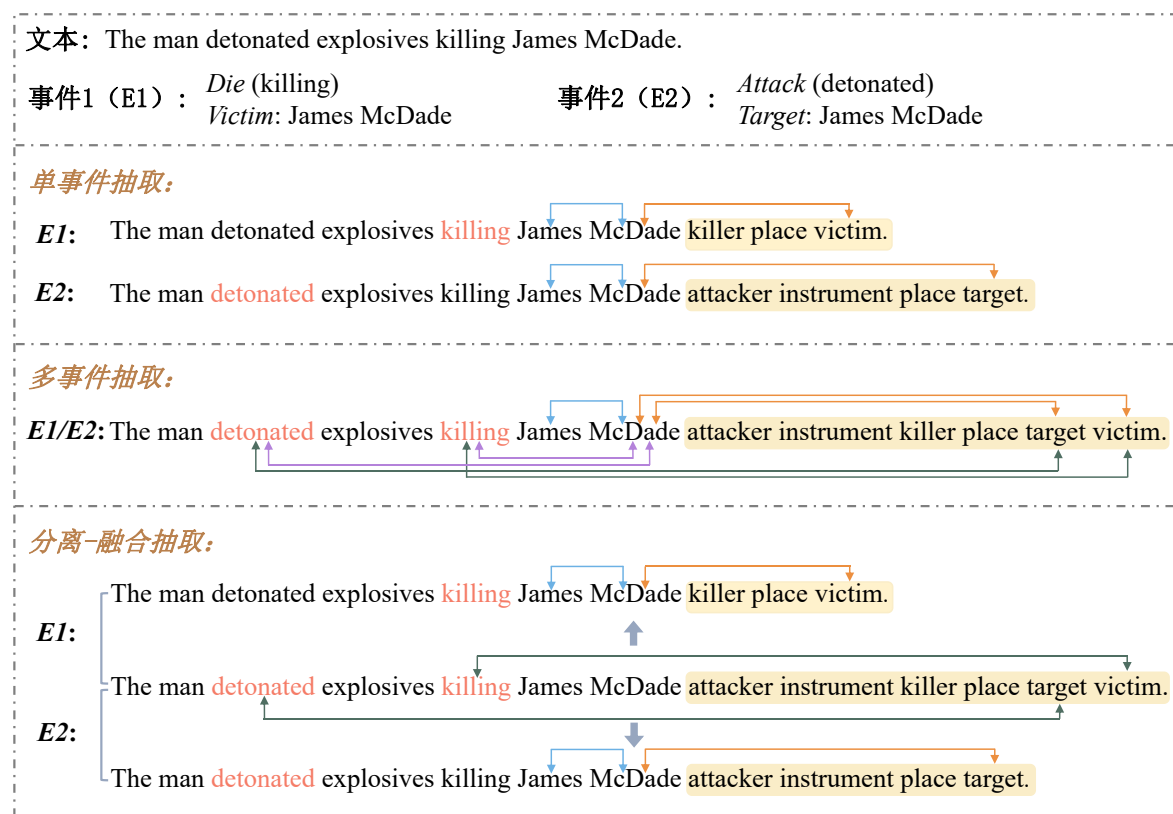


图 5.1 基于不同建模方式的多词元链接事件要素抽取示意图

因此，第二种方式的研究^[155]一次性抽取多个事件的事件要素。然而，两个新的问题随之而来。首先，与输入文本串接的不再是只在单个事件中出现的要素类型，而是在所有事件中出现的要素类型，这使得通过链接预测选择出正确的要素类型变得更加困难。其次，与单事件抽取方式相比，其需要两种额外的链接操作来确定抽取的事件要素文段和对应的要素类型归属的事件触发词，导致预测阶段积累更多的错误。图5.1展示了同时从 *Die* 事件和 *Attack* 事件中抽取事件要素“James McDade”时，其

额外增加的触发词-要素文段和触发词-要素类型链接操作。

为了利用上述两种抽取方式的优点，本章设计了一种基于分离-融合范式的跨事件信息构建与利用方法，其先分离跨事件信息获取和事件要素抽取的过程，然后将获取的跨事件信息融合到最终的事件要素抽取中。通过这种分离-融合范式，最终的事件要素抽取可以同时保留单事件抽取方式的简单链接预测和多事件抽取方式的跨事件依赖关系建模能力。图5.1说明了这一范式，其中中间部分从事件要素抽取过程中分离出来，通过触发词-要素类型链接获取跨事件信息，向上和向下的箭头表示跨事件信息的融合。

遵循设计的范式，本章提出了一种新颖的多词元链接事件要素抽取模型去获取分离的跨事件信息并进一步两阶段地融合到最终的事件要素预测 (**Separate acquisition of cross-event information and Two-fold Fusion, Sep2F**)。为了分离跨事件信息获取和事件要素抽取过程，本章设计了两个多词元链接模块。具体来说，针对多事件中的每个事件，引入一个链接模块来连接事件触发词和其中出现的要素类型。因此，不同事件触发词的特征表示并行聚合了在各自事件中出现的要素类型，提供了关键的跨事件信息。同时，引入另一个链接模块抽取目标事件的事件要素。其利用两种链接操作获取事件要素的文段范围和对应的要素类型。在此基础上，本章提出了一个两阶段融合模块将获取的跨事件信息融合到目标事件的事件要素抽取中。具体地，其首先动态融合来自上述两个链接模块的文本特征表示。然后，利用融合的文本特征表示获取跨模块词元链接分数。该链接分数被进一步融合到最终的预测中。这两阶段的融合过程相互影响，提供了显著的性能贡献。本章的主要贡献可以概括如下：

- 提出了一种新的基于分离-融合范式的跨事件信息构建与利用方法，其能够同时建模跨事件依赖信息和保留单事件抽取方式的优点。
- 基于分离-融合范式，提出了一种新的多词元链接事件要素抽取模型，其避免了当前通用事件要素抽取方法的各自局限，且实现了从不同级别文本中通用高效地利用跨事件信息。具体地，设计了两个链接模块，以分别获取跨事件信息和抽取目标事件的事件要素。此外，引入了一个两阶段融合模块，以确保在要素抽取过程中充分利用到获取的跨事件信息。
- 在四个不同文本级别的数据集上进行了充分的实验，实验结果表明本章所提的模型均显著超越目前性能最优的事件要素抽取基线方法。

5.2 方法设计

本章使用 (X, T, C) 表示每个数据实例，其中 X 是输入文本， T 是目标事件， C 表示在文本 X 中除目标事件 T 以外的其他事件。具体地， T 进一步表示为 (e, t, \mathcal{R}^e) ，其中 e 是事件类型， t 是事件触发词， \mathcal{R}^e 为只在事件 e 中出现的要素类型集合。类似地， C 可表示为 $\{(\tilde{e}_i, \tilde{t}_i, \mathcal{R}^{\tilde{e}_i}) \mid i \leq |C|\}$ ，其为目标事件 T 提供了跨事件信息。基于此，本章所提模型旨在抽取目标事件 T 的要素集合 \mathcal{A} ，其集合 \mathcal{A} 中的每个事件要素 $a^{(r)}$ 表示要素类型为 $r \in \mathcal{R}^e$ 的 X 中文本片段。

如图5.2所示，本章提出的模型由三个模块组成：多事件链接构造、目标事件链接构造和两阶段融合，其遵循了本章设计的分离-融合范式。具体地，前两个模块将跨事件信息的获取和目标事件的要素抽取过程分离，最后一个模块实现了获取的跨事件信息的融合。接下来，本章将具体介绍各个模块。

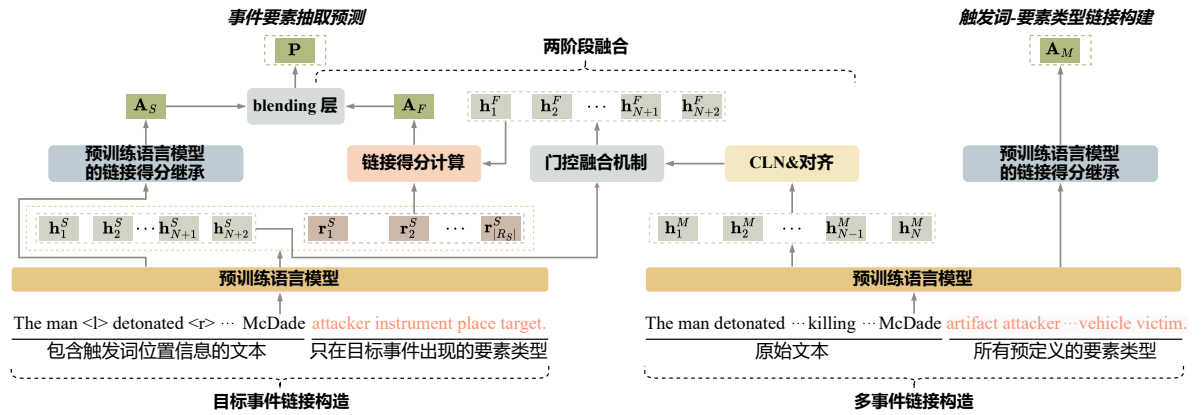


图 5.2 Sep2F 模型架构图

5.2.1 多事件链接构造

本模块通过聚合多事件的要素类型信息来获取跨事件信息，其中多事件包含目标事件 T 和所有其他事件 C 。为了实现这种信息聚合，本模块并行地建立不同的事件触发词和它们对应的事件要素类型之间的链接。首先，联合编码待抽取文本和给定数据集中预定义的所有要素类型。然后，根据事件中的触发词-要素类型链接引入标签矩阵和得分矩阵。最后，给出相应的训练损失。

编码 首先，本章利用对应的事件要素名称描述每个事件要素类型，即单个自然语言描述的词。对于名称中包含多个词的少数要素类型，本模型在训练过程中学习额外

的特殊词元进行表示。然后，拼接给定数据集中预定义的所有要素类型的名称描述，记作 R_M 。在此之后，使用预训练语言模型（PLM）对序列 R_M 和输入文本 X 进行联合编码如下：

$$\mathbf{E}^M = (\mathbf{h}_1^M, \dots, \mathbf{h}_N^M, \mathbf{r}_1^M, \dots, \mathbf{r}_{|R_M|}^M) = \text{PLM}(X \oplus R_M) \quad (5.1)$$

其中 \mathbf{h}_n^M ($1 \leq n \leq N$) 和 \mathbf{r}_n^M ($1 \leq n \leq |R_M|$) 分别为输入文本 X 和事件要素序列 R_M 中的词元向量。

标签矩阵 为了学习到多事件各自的要素类型信息，本模块设计了标签矩阵 $\mathbf{L}_M \in \mathbb{B}^{(N+|R_M|) \times (N+|R_M|)}$ 。对于 T 或 C 中的每个事件，假设其触发词的开始和结束词元向量分别是 \mathbf{h}_i^M 和 \mathbf{h}_j^M 。进一步，对于该事件中出现的每个要素类型，假设其对应的词元向量是 \mathbf{r}_k^M 。然后，实施链接操作来连接触发词和该要素类型。具体来说，分别构造链接词元对 $(i, k + N)$ 和 $(k + N, j)$ ，并相应地设置 $L_M[i][k + N]$ 和 $L_M[k + N][j]$ 为“True”。对于没有链接的词元对，本模块将 \mathbf{L}_M 中的相应值标记为“False”。图5.3展示了 \mathbf{L}_M 的一个示例，其中被红色标记的为事件触发词词元，被黄色标记的为所有预定义的事件要素类型词元，“-”表示“False”。根据图5.3可知，“detonated”触发的事件中存在要素类型 *Target*，而“killing”触发的事件中存在要素类型 *Victim*。

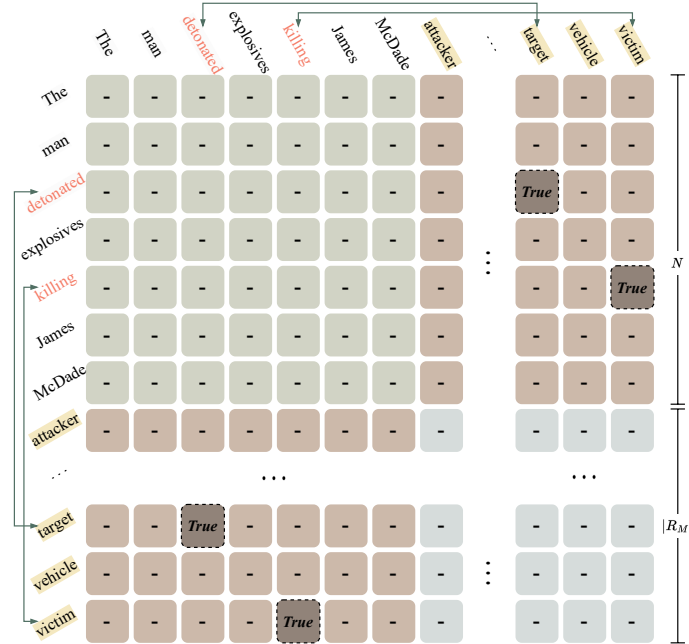


图 5.3 标签矩阵 \mathbf{L}_M 示例图

得分矩阵 受 Tang 等人^[157] 研究工作的启发，本模块继承基于 Transformer 架构的

预训练语言模型的多头自注意力结果作为不同词元对之间的链接分数。具体来说，根据选定的预训练语言模型的最后一层，获取其未经过 softmax 归一化的多头自注意力权重，并进行均值计算：

$$\mathbf{A}_M = \frac{1}{P} \sum_p \frac{\mathbf{Q}_p \mathbf{K}_p^\top}{\sqrt{d_h}} \quad (5.2)$$

其中 P 为多头的数目， d_h 为多头自注意机制中查询向量和键向量的维度， \mathbf{Q}_p 和 \mathbf{K}_p 分别为查询矩阵和键矩阵， $\mathbf{A}_M \in \mathbb{R}^{(N+|R_M|) \times (N+|R_M|)}$ 表示多事件的触发词-要素类型链接得分。

训练损失 本模块使用训练损失 \mathcal{L}_{TR} 去指导模型学习不同事件触发词及其相应出现的要素类型之间的链接，如下所示：

$$\begin{aligned} \mathcal{L}_{\text{TR}} = - \frac{1}{(N + |R_M|)^2} \sum_i \sum_j (L_{i,j}^M \log \sigma(A_M[i][j]) + \\ (1 - L_{i,j}^M) \log (1 - \sigma(A_M[i][j]))) \end{aligned} \quad (5.3)$$

其中 $\sigma(\cdot)$ 为 sigmoid 激活函数。当 $L_M[i][j]$ 为 “True” 时，设置 $L_{i,j}^M$ 为 1，否则设置 $L_{i,j}^M$ 为 0。

5.2.2 目标事件链接构造

为了抽取给定目标事件 (e, t, \mathcal{R}^e) 的事件要素，本模块对输入文本和只在事件 e 中出现的要素类型进行联合编码，然后定义对应的多词元链接标签矩阵和得分矩阵，并给出相应的训练损失计算。

编码 与多事件链接构造模块不同，本模块拼接所有 \mathcal{R}^e 集合中的事件要素类型的名称描述，记作 R_S 。然后，遵循 Ma 等人^[95]的研究工作在文本 X 中插入两个特殊词元 $\langle \ell \rangle$ 和 $\langle r \rangle$ ，以便标记触发词 t 的位置：

$$X_S = (x_1, \dots, \langle \ell \rangle, t, \langle r \rangle, \dots, x_{|X|}) \quad (5.4)$$

在此之后，将包含触发词位置信息的文本 X_S 和要素类型序列 R_S 进行串接，并利用

另一个预训练语言模型对其进行编码：

$$\mathbf{E}^S = (\mathbf{h}_1^S, \dots, \mathbf{h}_{N+2}^S, \mathbf{r}_1^S, \dots, \mathbf{r}_{|R_S|}^S) = \text{PLM}(X_S \oplus R_S) \quad (5.5)$$

其中 \mathbf{h}_n^S ($1 \leq n \leq N+2$) 和 \mathbf{r}_n^S ($1 \leq n \leq |R_S|$) 分别为 X_S 和 R_S 中的词元向量。

标签矩阵 为了标记给定目标事件中的事件要素文段范围和要素类型，本模块定义了标签矩阵 $\mathbf{L}_S \in \mathbb{B}^{(N+|R_S|+2) \times (N+|R_S|+2)}$ 。对于目标事件中的每个事件要素，假设其文段的开始和结束词元向量分别是 \mathbf{h}_i^S 和 \mathbf{h}_j^S ，其要素类型对应的词元向量是 \mathbf{r}_k^S 。基于此，首先构造链接词元对 (i, j) 和 (j, i) 并设置 $L_S[i][j]$ 和 $L_S[j][i]$ 为 “True”，以用来标记事件要素的文段。同时，通过要素文段-要素类型链接操作来标记事件要素对应的要素类型信息。相应地，构造两个链接对 $(i, k+N+2)$ 和 $(k+N+2, j)$ ，并设置 $L_S[i][k+N+2]$ 和 $L_S[k+N+2][j]$ 为 “True”。对于没有链接的词元对，将 \mathbf{L}_S 中的相应值设置为 “False”。图5.4展示了 \mathbf{L}_S 的一个示例，其中被红色标记的为事件触发词词元，被黄色标记的为只在该事件触发词对应的事件类型中出现的事件要素类型词元，“-”表示 “False”。根据图5.4可知，在 “detonated” 触发的事件中，存在事件要素 “James McDade”，其对应的要素类型为 *Target*。

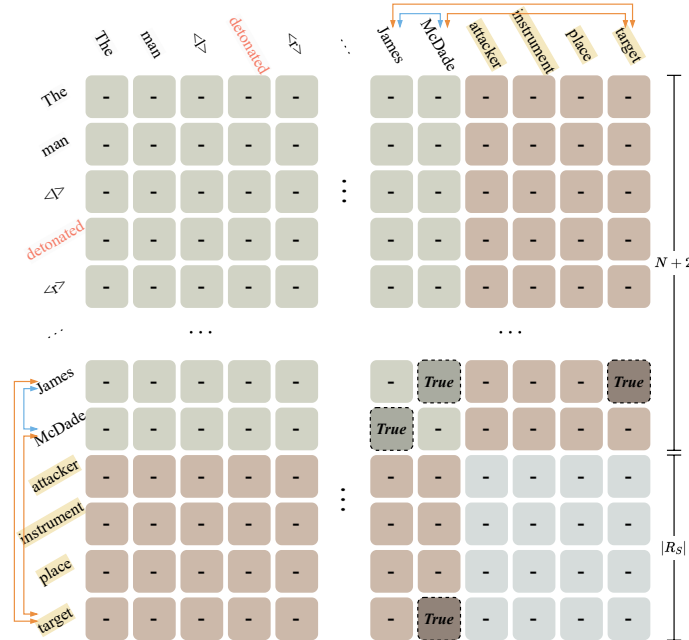


图 5.4 标签矩阵 \mathbf{L}_S 示例图

得分矩阵 与多事件链接构造模块中的公式5.2相同，本模块同样根据所使用的预训练语言模型的最后一层，利用其未通过 softmax 归一化的平均多头自注意权重作为

多词元间的链接分数，表示为 $\mathbf{A}_S \in \mathbb{R}^{(N+|R_S|+2) \times (N+|R_S|+2)}$ 。

训练损失 为了利用到获取的跨事件信息，首先采用两阶段融合模块（其将在下一模块中作详细介绍）来得到最终的词元链接预测矩阵：

$$\mathbf{P} = \text{TFF}(\mathbf{E}^M, \mathbf{E}^S, \mathbf{A}_S) \quad (5.6)$$

其中 TFF 指两阶段融合模块。然后，可以得到训练损失 \mathcal{L}_{EAE} 如下：

$$\mathcal{L}_{\text{EAE}} = - \frac{1}{(N + |R_S| + 2)^2} \sum_i \sum_j (L_{i,j}^S \log P[i][j] + (1 - L_{i,j}^S) \log (1 - P[i][j])) \quad (5.7)$$

其中当 $L_S[i][j]$ 为 “True” 时，设置 $L_{i,j}^S$ 为 1，否则设置 $L_{i,j}^S$ 为 0。

5.2.3 两阶段融合

本模块利用聚合了跨事件要素类型信息的文本表示来增强目标事件的要素抽取。具体地，首先动态融合来自上述两个链接模块的文本表示，即第一阶段融合。然后，使用融合的文本表示来获得目标事件的另一个词元链接得分矩阵，其不同于 \mathbf{A}_S 。这两个词元链接得分矩阵进一步被融合为最终的链接预测，即第二阶段融合。

第一阶段融合 由于 \mathbf{E}^M 中的文本向量涉及多事件信息，本模块首先利用条件层归一化（Conditional Layer Normalization, CLN）^[158,159] 来获取与目标事件（触发词）相关的上下文向量。对于目标事件触发词 t ，本模块简单地将其输入到一个参数冻结的预训练语言模型中，并使用得到的第一个词元向量来表示该触发词，记作 \mathbf{t} 。然后，对于 \mathbf{E}^M 中的每个词元向量 \mathbf{h}_n^M ($1 \leq n \leq N$)，可得到相应的向量 $\hat{\mathbf{h}}_n^M$ ($1 \leq n \leq N$) 如下：

$$\boldsymbol{\alpha}_t = \mathbf{t} \mathbf{W}_\alpha + \mathbf{b}_\alpha \quad (5.8)$$

$$\boldsymbol{\beta}_t = \mathbf{t} \mathbf{W}_\beta + \mathbf{b}_\beta \quad (5.9)$$

$$\hat{\mathbf{h}}_n^M = \text{CLN}(\mathbf{h}_n^M, \boldsymbol{\alpha}_t, \boldsymbol{\beta}_t) = \boldsymbol{\alpha}_t \odot \left(\frac{\mathbf{h}_n^M - \mu}{\sigma} \right) + \boldsymbol{\beta}_t \quad (5.10)$$

其中 $\mathbf{W}_\alpha \in \mathbb{R}^{d_1 \times d_1}$ ， $\mathbf{W}_\beta \in \mathbb{R}^{d_1 \times d_1}$ ， $\mathbf{b}_\alpha \in \mathbb{R}^{d_1}$ 和 $\mathbf{b}_\beta \in \mathbb{R}^{d_1}$ 为可训练的参数， μ 和 σ 分别为根据 \mathbf{h}_n^M 中的向量元素计算得到的平均值和标准偏差， d_1 为向量 \mathbf{t} 的维度。然后，

在 $(\hat{\mathbf{h}}_1^M, \dots, \hat{\mathbf{h}}_N^M)$ 中插入零向量 $\mathbf{0}$ ，以便与文本向量 \mathbf{E}^S 进行对齐：

$$(\bar{\mathbf{h}}_1^M, \dots, \bar{\mathbf{h}}_{N+2}^M) = (\hat{\mathbf{h}}_1^M, \dots, \mathbf{0}, \hat{\mathbf{h}}_i^M, \dots, \hat{\mathbf{h}}_j^M, \mathbf{0}, \dots, \hat{\mathbf{h}}_N^M) \quad (5.11)$$

其中 i 和 j 对应于触发词 t 的词元向量在 $(\mathbf{h}_1^S, \dots, \mathbf{h}_{N+2}^S)$ 中的开始和结束位置。最后，动态融合不同的文本向量。给定两个词元向量 \mathbf{h}_n^S 和 $\bar{\mathbf{h}}_n^M$ ($1 \leq n \leq N+2$)，本模块使用门控融合机制以获取词元向量 \mathbf{h}_n^F ：

$$\mathbf{g}_n = \sigma([\mathbf{h}_n^S; \bar{\mathbf{h}}_n^M] \mathbf{W}_G) \quad (5.12)$$

$$\mathbf{h}_n^F = \mathbf{g}_n \odot \mathbf{h}_n^S + (1 - \mathbf{g}_n) \odot \bar{\mathbf{h}}_n^M \quad (5.13)$$

其中 $\sigma(\cdot)$ 为 sigmoid 函数， $\mathbf{W}_G \in \mathbb{R}^{2d_1 \times d_1}$ 为可训练学习的矩阵， \odot 表示元素依次相乘操作。

第二阶段融合 首先，串接融合的文本向量 $(\mathbf{h}_1^F, \dots, \mathbf{h}_{N+2}^F)$ 和要素类型对应的词元向量 $(\mathbf{r}_1^S, \dots, \mathbf{r}_{|R_S|}^S)$ ，表示为 \mathbf{F} 。然后，本模块计算另一个词元链接得分矩阵 $\mathbf{A}_F \in \mathbb{R}^{(N+|R_S|+2) \times (N+|R_S|+2)}$ 如下：

$$\mathbf{F}_Q = \mathbf{F} \mathbf{W}_Q, \mathbf{F}_K = \mathbf{F} \mathbf{W}_K \quad (5.14)$$

$$\mathbf{A}_F = \mathbf{F}_Q \mathbf{F}_K^\top \quad (5.15)$$

其中 $\mathbf{W}_Q \in \mathbb{R}^{d_1 \times d_2}$ 和 $\mathbf{W}_K \in \mathbb{R}^{d_1 \times d_2}$ 为可训练学习的矩阵。最后，根据 \mathbf{A}_F 和章节 5.2.2 中得到的得分矩阵 \mathbf{A}_S ，使用 blending 层^[160] 融合如下：

$$\mathbf{P} = \sigma(\mathbf{A}_S + \mathbf{A}_F - \tau) \quad (5.16)$$

其中 τ 为可训练的参数。

5.2.4 训练和推理

训练 本章提出模型的整体训练损失可根据如下得到：

$$\mathcal{L} = \alpha \mathcal{L}_{\text{EAE}} + (1 - \alpha) \mathcal{L}_{\text{TR}} \quad (5.17)$$

Algorithm 1 目标事件的要素推理**Input:** 预测的链接对集合 $\mathcal{C}((s_k, e_k)), k \in (1, \dots, |\mathcal{C}|)$.**Output:** 事件要素集合 \mathcal{A} .

```

1:  $cand\_span = set()$ .
2:  $start2r\_dict = \{\}, end2r\_dict = \{\}$ .
3: for  $(s_k, e_k) \in \mathcal{C}$  do
4:   if  $s_k \leq N + 2$  and  $e_k \leq N + 2$  then
5:     将元组  $(s_k, e_k)$  添加到集合  $cand\_span$ 
6:   else if  $s_k \leq N + 2$  and  $e_k > N + 2$  then
7:     将键值对  $(s_k, e_k - N - 2)$  添加到字典  $start2r\_dict$ 
8:   else if  $s_k > N + 2$  and  $e_k \leq N + 2$  then
9:     将键值对  $(e_k, s_k - N - 2)$  添加到字典  $end2r\_dict$ 
10:  end if
11: end for
12: for  $(s, e) \in cand\_span$  do
13:   for  $r \in (\text{set}(start2r\_dict[s]) \cap \text{set}(end2r\_dict[e]))$  do
14:     将元组  $(s, e, r)$  添加到集合  $\mathcal{A}$ 
15:   end for
16:   for  $r \in (\text{set}(start2r\_dict[e]) \cap \text{set}(end2r\_dict[s]))$  do
17:     将元组  $(e, s, r)$  添加到集合  $\mathcal{A}$ 
18:   end for
19: end for

```

其中 α ($0 < \alpha < 1$) 为权重参数。

推理 基于词元链接预测矩阵 \mathbf{P} ，对于其中的每一个链接对 (i, j) ，如果对应预测值 $P[i][j]$ ($1 \leq i, j \leq N + |R_S| + 2$) 超过设定的阈值参数 δ ，则将该链接对添加到链接对集合 \mathcal{C} 中。然后，使用 \mathcal{C} 作为输入，运行如算法1所示的推理算法，得到目标事件的要素集合 \mathcal{A} 。对于 \mathcal{A} 中的每个元素 $(start, end, role)$ ， $start$ 和 end 分别表示要素的开始和结束位置， $role$ 为对应的要素类型在 R_S 中的索引。

5.3 实验评估

5.3.1 实验设置

1. 数据集 为了评估所提模型，本章将在以下数据集上进行实验：

- ACE2005^[138] - 该数据集详细介绍见章节3.3.1。本章实验使用其事件标注数据进行句子级别事件要素抽取的评估。在数据预处理上，与章节4.4.1不同，本章遵循 PAIE^[95] 和 TabEAE^[131] 等通用事件要素抽取模型的设置，使用 Du 等人^[72] 的预处理脚本。该脚本遵循 DyGIE++^[161] 模型的设置。

- RAMS^[93] - 该文档级别事件要素抽取数据集来源于新闻文章。与原始数据集将同一上下文中的多个事件视为不同的实例不同，本章实验遵循 TabEAE 的预处理过程，聚合每个实例的同一上下文中出现的其他事件的标注信息。
- WikiEvents^[94] - 该文档级别数据集来源英文维基百科。本章实验遵循 PAIE 的预处理程序，采用以每个目标事件触发词为中心的预定义窗口去截断过长的上下文内容，避免超出预训练语言模型的输入长度限制。
- MLEE^[162] - 该文档级别事件要素抽取数据集基于生物医学领域的英文出版物摘要标注得到。在预处理方面，遵循 TabEAE 参照 Trieu 等人^[163]的预处理设置。此外，由于 MLEE 数据集中不存在验证集，本章实验遵循 TabEAE 使用训练集调整模型的超参数。

对于所有文档级别数据集，遵循 He 等人^[131]设置预定义的窗口长度为 250，将每个文档拆分为多个上下文片段。本章实验所用数据集的具体统计数据如表 5.1 所示。

表 5.1 数据集详细统计信息

数据集	ACE2005	RAMS	WikiEvents	MLEE
# 事件				
训练集	4,202	7,329	3,241	4,442
验证集	450	924	345	-
测试集	403	871	365	2,200
# 事件要素				
训练集	4,859	17,026	4,552	5,786
验证集	605	2,188	428	-
测试集	576	2,023	566	2,764
# 事件类型	33	139	50	23
# 要素类型	22	65	59	8

2. 评价指标 遵循之前通用文本级别事件要素抽取研究工作^[95,131]，本章采用两个指标来衡量模型性能。(1) 要素识别 F1 得分 (Argument Identification F1 score, Arg-I): 当其事件要素文段的边界与事件中任意一个标注的要素相匹配，则认为该事件要素被正确识别。基于该识别准则，本章进一步计算对应的 F1 值评估要素识别性能，其 F1 值计算参照章节 3.3.1。(2) 要素分类 F1 得分 (Argument Classification F1 score, Arg-C): 当其事件要素文段的边界和要素类型都与事件中任意一个标注的要素相匹配，则认为该事件要素被正确分类。基于该分类准则，本章进一步计算对应的 F1 值评估要素

分类性能，其 F1 值计算参照章节3.3.1。在实验中，使用三个不同的随机种子训练本章提出的 Sep2F 模型，并通过平均值计算得到相应的性能结果。

3. 实验配置 为了便于与最近的研究工作进行公平比较，本章实验在模型中使用 RoBERTa-base 和 RoBERTa-large^[105] 作为预训练语言模型，其分别基于 Huggingface 上发布的“FacebookAI/roberta-base”文件²和“FacebookAI/roberta-large”文件³进行参数初始化。基于此两种模型，均采用配备了线性学习率调度器的 Adam 优化器和比率为 0.1 的 warmup 策略。 δ 为等式5.7中二分类预测值 $P[i][j]$ 的推断阈值，因此实验中遵循常规的二分类设置，在所有数据集上均设置 δ 为 0.5。在 RAMS 数据集上设置训练轮次为 50，在其他数据集上设置训练轮次为 100。在 ACE2005 数据集上设置批数据大小为 16，在其他数据集上设置批数据大小为 8。对于 ACE2005， d_2 设置为 32，而在其他数据集上 d_2 设置为 64。此外，基于每个数据集的验证集，使用网格搜索调整各自模型的学习率和训练权重 α 。具体地，在区间 $\{0.5, 0.6, 0.7, 0.8, 0.9\}$ 内调整权重 α ，在区间 $\{3, 4, 5\} \times 10^{-5}$ 和 $\{1, 2, 3\} \times 10^{-5}$ 内分别搜索基于 RoBERTa-base 和 RoBERTa-large 的模型的最佳学习率。表5.2列出了在不同数据集上最终选择的训练权重和学习率，其中表格上半部分为基于 RoBERTa-base 的模型参数详情，表格下半部分为基于 RoBERTa-large 的模型参数详情。对于基于 RoBERTa-base 的模型，不同数据集上的训练均在相同实验服务器上完成，具体配置为：CPU 型号 Intel(R) Xeon(R) Gold 5320，主频 2.20GHz；内存 256G；GPU 计算卡型号 NVIDIA GeForce RTX 3090，显存 24G。对于基于 RoBERTa-large 的模型，其具体配置为：CPU 型号 Intel(R) Xeon(R) Silver 4314，主频 2.40GHz；内存 128G；GPU 计算卡型号 NVIDIA A100 PCIE，显存 40G。本章实验均基于 Pytorch^[139] 框架完成。

表 5.2 超参数设置

超参数	ACE2005	RAMS	WikiEvents	MLEE
学习率	3e-5	5e-5	4e-5	4e-5
权重 α	0.7	0.7	0.8	0.6
学习率	2e-5	3e-5	2e-5	1e-5
权重 α	0.8	0.5	0.7	0.8

4. 基线模型 本章选择以下模型作为与所提的 Sep2F 模型进行实验比较的基线，其均在句子级别和文档级别数据集上评估了事件要素抽取任务的性能：

²<https://huggingface.co/FacebookAI/roberta-base>

³<https://huggingface.co/FacebookAI/roberta-large>

- EEQA^[72] - 该模型将事件要素抽取任务转化为机器阅读理解问题进行解决。
- BART-Gen^[94] - 该模型将事件要素抽取建模为基于提示模版的条件生成任务。
- PAIE^[95] - 该模型提出了抽取式的提示学习方法，其利用要素类型对应的槽位表示查询对应的要素文段。
- APE(Single)^[154] - 该模型通过将提示模版中的要素类型映射至对应的实体类型，实现了跨事件共享知识的学习。
- PGAD^[164] - 该模型引入了扩散模型，以获取多样化的上下文感知提示表示。
- EDGE^[130] - 该模型同时设计了事件内和跨事件的图，以丰富要素类型语义信息。
- TabEAE^[131] - 该模型将多事件要素抽取转换为表格生成，以捕捉事件间的关联。

最近，基于情境学习 (In-Context Learning, ICL)^[90] 的 ChatGPT 在各种自然语言处理任务中表现出色。但是，目前缺乏对不同事件要素抽取数据集的全面评估，尤其是文档级别数据集。因此，本章按照 Han 等人^[89] 的设置，构建 5-shot ICL 提示模版作为每个测试样本的输入，并使用 OpenAI 接口⁴获取事件要素抽取的结果。具体地，本章使用两个版本的 ChatGPT 评估事件要素抽取的性能：*gpt-3.5-turbo* 和 *gpt-4*，其分别由 *gpt-3.5* 和 *gpt-4* 模型构建。

此外，UnifiedEAE^[165] 和 APE^[154] 等研究工作同时利用了多个数据集提升目标数据集的性能，以受益于额外的数据资源。因此，本章未将其作为主实验结果的比较基线，而是另作单独的分析。

5.3.2 性能结果

表5.3展示了在四个数据集上的事件要素抽取实验结果，其中粗体表示在各自数据集上的最优性能，下划线表示在各自数据集上的次优性能结果，预训练模型所在列中的“b”和“l”分别表示对应的为 base 版本和 large 版本。从表5.3中可以观察到，本章提出的基于 RoBERTa-large 的模型在所有数据集上都取得了当前最优的性能。具体地，与之前性能最优的 TabEAE 模型相比，其在 ACE2005、RAMS、WikiEvent 和 MLEE 数据集上，分别在 Arg-I/Arg-C F1 指标上取得了 1.6%/2.0%、1.4%/1.0%、2.6%/2.8%

⁴<https://platform.openai.com/>

和 2.4%/2.5% 的性能提升。而相比于使用了 base 版本预训练模型的最优基线，本章提出的基于 RoBERTa-base 的模型在 ACE2005、RAMS 和 WikiEvent 数据集上分别在 Arg-I/Arg-C F1 指标上取得了 0.9%/2.9%、1.4%/2.4% 和 3.7%/4.1% 的性能提升。此外，本章所提的 base 模型在 WikiEvents 和 MLEE 数据集上超越了所有使用 large 版本预训练模型的基线，且在 ACE2005 和 RAMS 数据集上相比 large 基线模型也展现了具有竞争力的性能。因此，可以得知本章提出的 Sep2F 模型在处理不同文本级别的事件要素抽取任务时均表现出色。

表 5.3 四种数据集上的事件要素抽取性能结果

模型	预训练模型	ACE2005		RAMS		WikiEvents		MLEE	
		Arg-I	Arg-C	Arg-I	Arg-C	Arg-I	Arg-C	Arg-I	Arg-C
ChatGPT	GPT-3.5	35.6	30.0	25.1	19.3	13.6	11.6	15.7	11.8
ChatGPT	GPT-4	37.5	33.2	23.8	20.9	16.7	15.3	16.9	14.9
BART-Gen	BART-b	59.6	55.0	50.9	44.9	47.5	41.7	-	-
PAIE	BART-b	73.6	69.8	54.7	49.5	68.9	63.4	-	-
APE(Single)	BART-b	74.1	70.1	54.8	49.6	66.2	62.1	-	-
PGAD	BART-b	74.1	70.5	53.7	45.9	69.6	64.1	-	-
EDGE	BART-b	75.3	70.6	55.2	49.7	68.2	62.8	-	-
Sep2F	RoBERTa-b	76.2	73.5	56.6	52.1	<u>73.3</u>	<u>68.2</u>	<u>76.8</u>	<u>75.6</u>
EEQA	RoBERTa-l	72.1	70.4	51.9	47.5	60.4	57.2	70.3	68.7
BART-Gen	BART-l	69.9	66.7	51.2	47.1	66.8	62.4	71.0	69.8
PAIE	BART-l	75.7	72.7	56.8	52.2	70.5	65.3	72.1	70.8
PAIE	RoBERTa-l	76.1	73.0	57.1	52.3	70.9	65.5	72.5	71.4
APE(Single)	BART-l	75.3	72.9	56.3	51.7	70.6	65.8	-	-
TabEAE	RoBERTa-l	<u>77.2</u>	<u>75.0</u>	<u>57.3</u>	<u>52.7</u>	71.4	66.5	75.1	74.2
Sep2F	RoBERTa-l	78.8	77.0	58.7	53.7	74.0	69.3	77.5	76.7

另外，可以观察到不同版本的 ChatGPT 在事件要素抽取任务上的性能都明显落后于当前有监督模型。尤其在处理更具挑战性的文档级别事件要素抽取任务时，性能差距通常会进一步扩大。因此，提升 ChatGPT 等大语言模型在事件要素抽取任务上的性能表现仍有待深入研究。

5.3.3 不同多词元链接模型性能比较

如章节 5.1 中所述，基于多词元链接操作的现有事件要素抽取模型主要存在两种方式：单事件抽取和多事件抽取。然而，这些方法将事件要素抽取任务作为端到端的通用信息抽取或事件抽取中的一个子任务进行建模。因此，其实验设置和训练数据集与单独处理事件要素抽取任务存在差异。为了更加公平地比较不同的多词元链接模型

的性能，本章根据这些多词元链接模型的建模方式设计以下变体模型：(1) **SingleE** 在目标事件链接构造模块中移除两阶段融合，以抽取单个事件的要素信息，对应于单事件抽取方式。(2) **SingleE-L** 在模型输入中串接给定数据集的所有预定义要素类型名称与待抽取文本，其他设置与 **SingleE** 相同。(3) **MultiE** 同时抽取文本中多个事件的事件要素，对应多事件抽取方式。具体地，将事件要素抽取转化为多个 < 事件触发词, 要素类型, 要素文段 > 三元组抽取问题，并参照 Tang 等人在 UniRel^[157] 模型中基于多词元组成的实体进行关系三元组抽取的代码实现⁵。由于本章实验在事件要素抽取中使用正确标注的事件触发词作为输入，MultiE 将在推理过程中根据标注信息纠正事件触发词的文段范围预测错误。此外，与 MultiE 输入文本串接的为只在待抽取事件中出现的要素类型名称。(4) **MultiE-L** 将所有预定义的要素类型名称与待抽取文本串接，其他设置与 MultiE 相同。

表 5.4 四种数据集上不同多词元链接模型在 Arg-C F1 (%) 指标上的结果对比

模型	ACE2005 [1.35]	RAMS [1.25]	WikiEvents [1.78]	MLEE [3.32]
SingleE	75.4	53.0	66.5	74.4
SingleE-L	72.9	49.1	55.8	73.0
MultiE	69.9	49.3	47.9	57.2
MultiE-L	68.1	40.5	26.5	8.4
Sep2F	77.0	53.7	69.3	76.7

表5.4为上述的多词元链接模型在四个数据集上关于 Arg-C F1 (%) 指标的实验结果，其均使用 RoBERTa-large 进行编码，且 “[.]” 表示每个对应数据集中单个数据实例存在的平均事件数目。首先，从表5.4中可以总结出本章提出的模型在四个数据集上的性能均超越了所有变体模型。接下来，给出详细比较如下：

- **SingleE vs Sep2F**: 与本章提出的 Sep2F 模型相比，未利用跨事件信息的 SingleE 的 Arg-C F1 指标在 ACE2005、RAMS、WikiEvents 和 MLEE 数据集上分别下降了 1.6%、0.7%、2.8% 和 2.3%。
- **MultiE vs Sep2F**: MultiE 在事件要素抽取任务上未表现出有竞争力的性能，尤其是在 WikiEvents 和 MLEE 数据集上，MultiE 和本章所提的 Sep2F 的性能存在显著差距。

⁵<https://github.com/wtangdev/UniRel>

- **SingleE vs SingleE-L**: 在不同数据集上 SingleE-L 的 Arg-C F1 指标均低于 SingleE, 其验证了在文本和要素类型名称间进行链接预测时, 串接更多的要素类型会降低预测的性能。此外, 图5.5进一步展示了这两个模型在不同数据集上要素类型串接长度和 Arg-C F1 指标的关联性, 其中“增加的要素类型长度”表示 SingleE-L 输入中拼接的要素类型序列相比于 SingleE 输入中拼接的最长要素类型序列增加了的长度, “Arg-C F1 下降值”表示 SingleE 和 SingleE-L 的性能差距。根据图5.5可以观察到, 由于 SingleE-L 与 SingleE 在 RAMS 和 WikiEvents 数据集上的要素类型串联比其他数据集更长, 导致其在这两个数据集上的 Arg-C F1 指标差异也会相应地更加显著。
- **MultiE vs MultiE-L**: 与 SingleE 和 SingleE-L 的比较结果相比, MultiE 和 MultiE-L 的性能差距进一步增加, 其增加程度和不同数据集上单个数据实例的平均事件数目密切相关。例如, MLEE 数据集上单个数据实例的平均事件数超过三个, 导致 MultiE-L 性能的大幅下降。
- **SingleE-L vs MultiE-L**: 尽管 MultiE-L 模型同时抽取多个事件的事件要素, 以便考虑不同事件间的关联性。但是, 当 MultiE-L 和 SingleE-L 保持相同的要素类型串接时, 其 Arg-C F1 指标在所有数据集上均低于 SingleE-L 模型。此结果表明, MultiE-L 无法高效地在单个数据实例中处理更多的词元链接预测。因此, 其跨事件信息带来的性能提升未能弥补更多累积错误导致的性能下降。此外, WikiEvents 和 MLEE 数据集上单个实例的平均事件数量相比其他数据集较多, 导致在这两个数据集上 SingleE-L 和 MultiE-L 的性能差异更大。

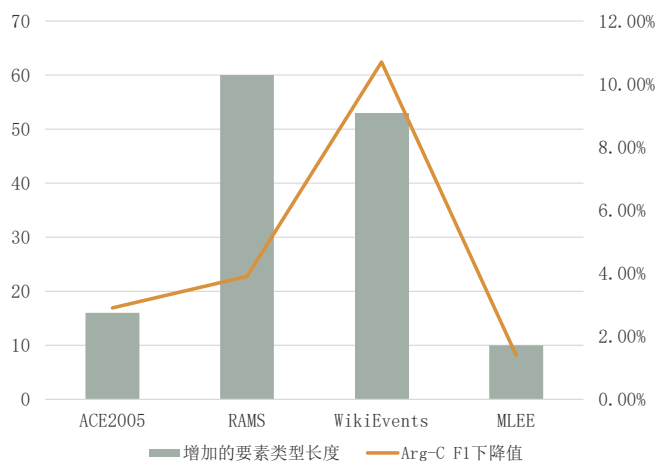


图 5.5 要素类型串接长度与 Arg-C F1 指标的关联性

5.3.4 单事件/多事件实例性能分析

遵循 He 等人^[131] 的实验设置, 本节根据目标抽取事件所在上下文中存在的其他事件的数量 (即 $|C|$), 将每个数据集的测试实例分为两组。当 $|C| = 0$ 时, 实例中只存在目标抽取事件。如果 $|C| > 0$, 则实例中存在多个事件。然后, 基于分组得到各自的统计和性能信息如表 5.5 所示, 其中 $[\cdot]$ 中的值表示对应分组中测试实例的数量, Sep2F-b 和 Sep2F-l 分别表示基于 RoBERTa-base 和 RoBERTa-large 的 Sep2F 模型。值得注意的是, 表中其他的基线模型均基于 RoBERTa-large 进行构建。从表 5.5 中可以观察到: (1) 在两个分组中, 基于 RoBERTa-large 的 Sep2F 模型在 Arg-C F1 指标上均超越了之前性能最优的基线。这种性能的提升可以归因于本章所提的分离-融合范式, 其同时保留了单事件抽取方式和多事件抽取方式的优点。(2) 基于 RoBERTa-base 的 Sep2F 模型在 WikiEvents 和 MLEE 数据集包含多个事件的实例分组上表现出了令人印象深刻的性能结果, 其超越了两个最具性能竞争力的基于 RoBERTa-large 的 SingleE 和 TabEAE。这验证了本章提出的模型在处理具有多个事件的实例时的良好表现。(3) 尽管 SingleE 和 PAIE 模型均未利用到跨事件信息, 但与 PAIE 相比, SingleE 在所有数据集上均表现出更好的性能结果, 其验证了基于多词元链接的模型解决事件要素抽取任务的潜力和优势。

表 5.5 四种数据集上单事件/多事件实例对应的 Arg-C F1 (%) 指标结果

模型	ACE2005		RAMS		WikiEvents		MLEE	
	$ C = 0$ [185]	$ C > 0$ [218]	$ C = 0$ [587]	$ C > 0$ [284]	$ C = 0$ [114]	$ C > 0$ [251]	$ C = 0$ [175]	$ C > 0$ [2025]
SingleE	74.8	76.0	52.9	53.3	68.4	65.5	83.1	73.8
MultiE	70.8	69.1	50.0	47.5	45.5	49.1	53.7	57.5
PAIE	71.0	73.9	52.7	52.1	65.3	65.4	78.9	70.1
TabEAE	73.4	76.1	52.9	52.5	67.3	66.2	81.1	73.6
Sep2F-b	72.8	74.0	52.1	52.0	66.8	69.0	85.5	74.9
Sep2F-l	75.4	78.3	53.2	54.6	68.8	69.5	84.3	76.1

5.3.5 消融研究

本章通过消融研究分析基于 RoBERTa-large 的 Sep2F 模型中不同组件的有效性, 包括第一阶段融合 (First-fold Fusion)、第二阶段融合 (Second-fold Fusion)、Blending 层 (Blending Layer)、文本中存在的其他事件信息 (Surrounding Events C) 和训练损

失 \mathcal{L}_{TR} (Training Loss \mathcal{L}_{TR})。根据表5.6所示的在 Arg-C F1 (%) 指标上的消融结果, 所有组件均能提升本章提出的 Sep2F 模型的性能。具体地, 可以观察到: (1) 如果移除第一阶段融合或第二阶段融合, Sep2F 模型在各个数据集上的性能均将下降。这表明两个阶段的融合均对性能提升贡献显著。(2) 移除两个阶段中任意一阶段的融合, Sep2F 模型在各个数据集上, 尤其是 WikiEvents 和 MLEE 数据集, 取得与之前性能表现最优的事件要素抽取基线模型相当的 Arg-C F1 指标。这展现了本章提出的模型即使利用简单的融合方法, 仍然能在事件要素抽取任务上表现良好, 进一步验证了分离-融合范式的优势。(3) 当移除 Blending 层并直接对两个词元链接得分矩阵进行加和, Sep2F 模型的 Arg-C F1 指标在所有数据集上均下降了 1.0% 以上, 这表明 Blending 层可以较好地集成不同的词元链接预测分数。(4) 如果移除给定目标事件上下文中的其他事件 C , 并在多事件链接构建模块中保留目标事件触发词和在该事件中出现的要素类型之间的链接, 则可以发现 Sep2F 模型在所有数据集上的性能均出现下降, 从而明确地证明了跨事件依赖信息对于模型的性能增益。此外, 可以发现 Sep2F 模型在 ACE2005 和 RAMS 数据集上性能略有下降, 而在 WikiEvents 和 MLEE 数据集上则性能下降的更加显著。该观察结果可以归因于表5.4中展示的各个数据集中单个实例的平均事件数目的差异。与 ACE2005 和 RAMS 数据集相比, 本章提出的模型可以利用到 WikiEvents 和 MLEE 数据集上更多事件的跨事件信息。因此, 与其他两个数据集相比, 移除目标事件周围的其他事件信息会导致 Sep2F 模型在 WikiEvents 和 MLEE 数据集上更显著的性能下降。(5) 如果移除多事件链接构造模块中的训练损失 \mathcal{L}_{TR} (即设置公式5.17中的训练权重参数 α 为 1), 则意味着 Sep2F 模型无法通过调整 \mathcal{L}_{TR} 而利用到上下文中其他事件的要素类型信息。从消融结果可以看出, 由此丢失的跨事件信息使得模型的 Arg-C F1 指标在 ACE2005、RAMS、WikiEvents 和 MLEE 数据集上分别下降了 1.2%、1.8%、2.3% 和 1.9%。

表 5.6 四种数据集上关于 Arg-C F1 (%) 指标的消融研究结果

模型	ACE2005	RAMS	WikiEvents	MLEE
Sep2F	77.0	53.7	69.3	76.7
- First-fold Fusion	74.8	51.2	66.3	75.6
- Second-fold Fusion	72.5	50.2	67.8	74.2
- Blending Layer	75.9	52.2	66.6	75.9
- Surrounding Events C	76.5	52.6	66.2	74.3
- Training Loss \mathcal{L}_{TR}	75.8	51.9	67.0	74.8

5.3.6 与资源增强模型的性能比较

本章将 Sep2F 模型与聚焦于利用跨数据集知识的 UnifiedEAE 和 APE 模型作比较，其中 APE 还额外人工构建了学习重叠知识的提示模版。图5.6展示了性能比较结果，其中上半部分的图展示的为基于 base 版本预训练语言模型的比较，而下半部分的图展示的为基于 large 版本预训练语言模型的结果。在使用 base 版本的预训练语言模型时，本章提出的 Sep2F 模型在所有三个数据集上均优于 UnifiedEAE 和 APE 模型。此外，当使用 large 版本预训练模型时，Sep2F 模型在 ACE2005 和 WikiEvents 数据集上的性能均优于 APE，并在 RAMS 数据集上也取得与 APE 相当的 Arg-C F1 指标。因此，可以得知尽管参与比较的基线受益于额外的训练资源和人为设计的经验知识，本章提出的模型在使用不同版本预训练语言模型时仍能在绝大部分数据集上实现最佳性能，进一步证明了其在事件要素抽取任务上的优越性。

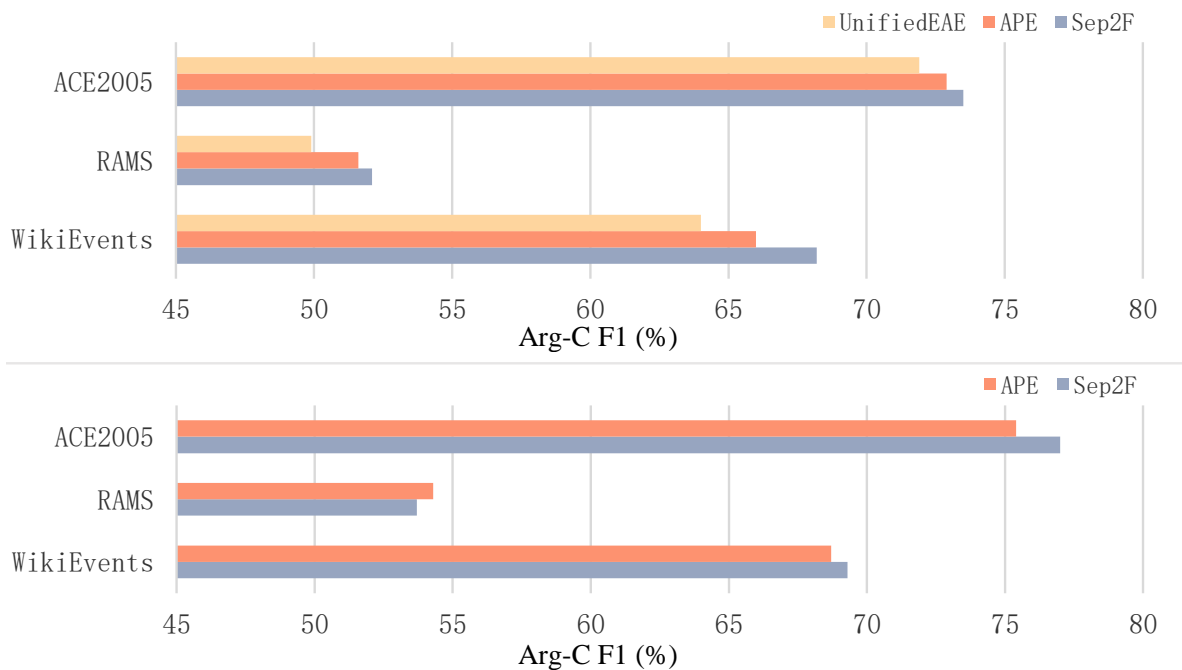


图 5.6 三种数据集上 Sep2F 与资源增强模型的性能结果对比

5.3.7 案例研究

本章节基于 WikiEvents 数据集中的一个具体实例进行定性的案例分析。图5.7展示了本章提出的 Sep2F 模型和 SingleE 模型在该实例上的事件要素抽取结果，其不同事件的触发词被不同颜色标记，且这两个模型均基于 RoBERTa-large 进行构建和训练。

可以观察到, SingleE 模型未能预测出由“booming”触发的 *Attack* 事件中的两个事件要素, 但 Sep2F 模型给出了正确的预测结果。本章推测其原因为 Sep2F 模型可以利用到由“killed”触发的事件中的要素类型信息。同样, 由于跨事件要素类型 *Destroyer* 提供了重要的依赖信息提示, Sep2F 模型避免了事件触发词“death”的干扰, 将“He”的要素类型识别为 *Killer* 而不是 *Victim*, 但 SingleE 模型给出了错误的抽取结果。

<p>The diagram illustrates semantic relations between entities and events in the text. Arrows indicate relations: 'Victim' from 'person' to 'killed'; 'Target' from 'killed' to 'bombing'; 'Killer' from 'Tsarnaev' to 'bombing'; 'Attacker' from 'Tsarnaev' to 'bombing'; 'Destroyer' from 'destruction' to 'death'; and 'Killer' from 'He' to 'death'.</p>		
事件类型: <i>Attack</i>	事件触发词: booming	
标注的Attacker: <i>Tsarnaev</i>	SingleE: <i>N/A</i> ×	Sep2F: <i>Tsarnaev</i>
标注的Target: <i>person</i>	SingleE: <i>N/A</i> ×	Sep2F: <i>person</i>
事件类型: <i>Die</i>	事件触发词: death	
标注的Killer: <i>He</i>	SingleE: <i>N/A</i> ×	Sep2F: <i>He</i>
标注的Victim: <i>N/A</i>	SingleE: <i>He</i> ×	Sep2F: <i>N/A</i>

图 5.7 WikiEvents 数据集上的案例分析

5.3.8 训练权重超参数分析

本章研究公式 5.17 中训练权重 α 对于 Sep2F 模型性能的影响。表 5.7 展示了不同 α 取值下 Sep2F 模型在各个数据集上的 Arg-C F1 性能指标, 可以观察到本章提出的 Sep2F 模型在不同数据集上均性能稳定, 证明了其良好的鲁棒性。

表 5.7 四种数据集上不同 α 取值对应的 Arg-C F1 (%) 指标结果

训练权重 α	ACE2005	RAMS	WikiEvents	MLEE
0.5	76.1	53.7	68.6	76.6
0.6	76.8	53.0	68.7	76.4
0.7	76.7	53.1	69.3	76.6
0.8	77.0	52.7	69.3	76.7
0.9	76.4	52.8	67.9	76.4

5.4 本章小结

本章提出了一种新颖的多词元链接事件要素抽取模型。具体地，提出了两个链接模块来分离跨事件信息的获取和目标事件的要素抽取过程。此外，提出了一种新的两阶段融合模块，以确保获取的跨事件信息能够有效地增益目标事件的要素抽取。因此，本章所提模型利用了跨事件依赖线索并保留了单事件抽取方式的优点。本章在四个不同级别的基准数据集上进行了充分实验，结果表明提出的模型在所有数据集上均取得了事件要素抽取的最优性能。