两阶段问答范式的生物医学事件触发词检测

行 帅[1],熊玉洁[1],苏前敏[1],黄继汉[2]
1.上海工程技术大学 电子电气工程学院,上海 201620
2.上海中医药大学 药物临床研究中心,上海 201203

摘 要:现有的生物医学事件触发词检测存在以下缺陷:保留了与触发词无关的冗余信息;忽略了实体与事件之间的潜在关联性;传统方法容易受到数据稀缺性的影响。针对上述问题,提出了一种两阶段问答范式的生物医学事件触发词检测方法。在事件类型识别阶段,采用基于句法距离的注意力捕获更有意义的上下文特征,排除无关信息的干扰;为了有效利用实体中的潜在特征,采用全局统计的单词-实体-事件共现特征,指导事件类型感知注意力挖掘词与事件之间的强关联性。在触发词定位阶段,根据识别出的事件类型,制定问题回答该事件对应的触发词索引,从而利用丰富的问答数据库实现数据增强。在MLEE语料库上的结果表明,两阶段问答范式、句法距离和事件类型感知注意力都有效地提升了模型性能,所提出的模型取得了81.39%的*F*1分数,并在多个事件类型上的详细结果均优于其他基线模型。

关键词:生物医学事件;触发词检测;句法距离;单词-实体-事件共现特征;两阶段问答范式
文献标志码:A **中图分类号:**TP391.1 doi:10.3778/j.issn.1002-8331.2301-0152

Biomedical Event Trigger Detection Based on Two-Stage Question Answering Paradigm

XING Shuai[1], XIONG Yujie[1], SU Qianmin[1], HUANG Jihan[2]
1.School of Electronic and Electrical Engineering, Shanghai University of Engineering Science, Shanghai 201620, China
2.Center for Drug Clinical Research, Shanghai University of Traditional Chinese Medicine, Shanghai 201203, China

Abstract: The existing biomedical event trigger detection methods have the following defects: Redundant information unrelated to triggers are retained; potential correlations between entities and events are ignored; traditional methods are vulnerable to data scarcity. A biomedical event trigger detection based on two-stage question answering paradigm is proposed to address the above problems. In the event type identification phase, in order to exclude the interference of irrelevant information, the attention based on syntactic distance is allowed to capture more meaningful contextual features. In order to effectively utilize the potential features in the entities, the word-entity-event co-occurrence feature based on global statistics is used to guide event type aware attention to explore the strong relationship between words and events. In the trigger localization phase, the trigger index of the event in the sentence is answered according to the identified event type questions, thus leveraging the rich question answering database to achieve data enhancement. The results on the MLEE corpus show that the two-stage question answering paradigm, syntactic distance attention, and event type aware attention effectively improve the performance of the model, and the proposed model achieves 81.39% *F*1-score, outperforming other baseline models in terms of detailed results for multiple event types.

Key words: biomedical events; trigger detection; syntactic distance; word-entity-event co-occurrence feature; two-stage question answering paradigm

伴随着生物医学领域中自然语言文本的不断扩大,生物医学事件在改善生物医学研究方面发挥着重要作用,其中生物医学事件抽取旨在从非结构化的生物医学文本中高效地表征结构化事件信息。生物医学事件提取工具的开发使得更多的下游任务成为可能,例如医学搜索[1]、医学知识图谱构建[2]和特定领域的文

基金项目:上海市科技创新行动计划技术标准项目(21DZ2203100);国家自然科学基金(62006150)。
作者简介:行帅,男,硕士研究生,CCF学生会员,研究方向为自然语言处理;熊玉洁,通信作者,博士,副教授,主要研究方向为特征识别,E-mail:xiong@sues.edu.cn;苏前敏,博士,副教授,主要研究方向为生物医学信息处理;黄继汉,博士,讲师,主要研究方向为临床试验设计。
收稿日期:2023-01-16 **修回日期:**2023-05-09 **文章编号:**1002-8331(2024)10-0121-11

本挖掘[3]等。

生物医学事件一般是由事件类型、触发词、论元和论元角色组成。触发词通常是一个动词或短语,其标志着某个特定事件的发生,论元通常是生物医学实体或其他触发词(嵌套事件),论元角色是论元与其参与事件的关系。例如,图1的句子中包含两个事件:(1)事件类型Gene Expression,触发词expression,Theme角色论元vascular endothelial growth factor;(2)事件类型Positive Regulation,触发词induced,Theme角色论元expression,Cause角色论元Insulin。生物医生事件抽取可以分为触发词检测和论元检测两个步骤。文献[4-5]显示,超过60%的事件抽取错误归因于不正确的事件触发词检测,因此一个有效的触发词检测方法对于事件抽取至关重要。

大部分的生物医学事件触发词检测方法都存在某些缺陷。首先,依赖树中的句法结构对于生物医学事件触发词检测至关重要[6]。最近,对依赖树建模的方法主要使用图卷积网络(graph convolutional networks,GCN)学习相邻节点之间的句法依赖关系[7-9]。然而,大多数基于GCN的方法被设计为堆叠结构以捕获高阶上下文信息,导致与触发词无关的冗余信息被保留。Ahmad等[10]验证了依赖树中的句法距离可以有效地表征词间句法联系,因此本文考虑使用句法距离学习词间的句法依赖结构,通过注意力机制代替GCN捕获上下文表示,并使每个词只关注给定句法距离内的相关词,超过该距离的词则被视为噪声,从而保证每个词只与其存在高度依赖关系的词执行上下文交互,避免无关信息的干扰。

其次,Li等[11]指出实体信息能为生物医学事件触发词检测提供重要线索,但是现有的触发词检测方法忽略了实体信息与事件之间的潜在关联,实体信息要么被完全忽略,要么被不同的上下文所混淆。为了能够利用实体特征指导模型学习,一个天然的想法就是在变化的上下文中寻求一个保持稳定的决策条件。因此本文通过全局统计的方式从外部环境中提取一组词-实体-事件共现特征作为稳定决策,为词与事件之间的事件类型感知注意力提供明确的关联方向,从而利用实体与事件之间的潜在关联提高模型性能。

最后,传统的生物医学事件触发词检测方法是基于单阶段的,这种方法的泛化能力受到了限制[12],并且无法缓解数据稀缺带来的影响[13]。为了解决上述问题,本文引入了一种新颖的两阶段问答范式,分别是事件类型识别和触发词定位。它不仅可以利用问答的最新进展提升模型性能,并且拆分后的两个子任务降低了检测复杂度。此外,一个句子能与多个事件类型问题组合生成多个新数据,因此实现了在不需要引入额外医学文本的情况下,利用丰富的问答数据库实现数据增强的效果。

本文提出了一种两阶段问答范式的生物医学事件触发词检测方法。它将触发词检测任务分为事件类型识别阶段和触发词定位阶段。在识别阶段,首先根据事件类型问题标识、原始文本和事件类型名称定制一个问题查询。然后引入依赖树中的依存关系构造句法距离矩阵,融合句法距离后的注意力机制有能力关注句子中更有意义的上下文特征,避免无关信息干扰。为了利用实体与事件之间的潜在关联,从训练集中统计的单词-实体-事件共现特征指导事件类型感知注意力机制挖掘词与事件类型之间的强关联性。最后,根据丰富的语义特征回答句子中包含的事件类型。在定位阶段,首先为句子中识别出的事件类型分别定制简单的问题模板,然后回答每个事件类型在句子中对应的触发词索引。本文的贡献如下:

(1)提出了一种新的两阶段问答范式,在利用问答的最新进展提升检测性能的同时也利用丰富的问答数据库实现数据增强。

(2)提出了一种基于全局统计的单词-实体-事件共现特征指导模型学习,减弱了词与事件类型之间的异常关联。

(3)利用依赖树中的句法距离集成词之间更有意义的上下文特征,避免了无关信息的干扰。

1　相关工作

由于生物医学领域涵盖了更复杂的专业背景知识,生物医学领域的事件抽取比一般领域更具有挑战性。现有的生物医学事件触发词检测方法可以分为基于规则的方法、机器学习方法和神经网络方法。基于规则的方法主要侧重于从预定义的词典中发现一系列规则。然而,规则的定义需要大量的领域知识作为支撑,且消耗时间。机器学习方法将触发器检测任务视为词级别或句子级别的分类问题。TEES(Turku event extraction system)[14]是一种综合依赖性特征的生物医学事件提取系统。Pyysalo等[15]将手工设计的上下文依赖特征

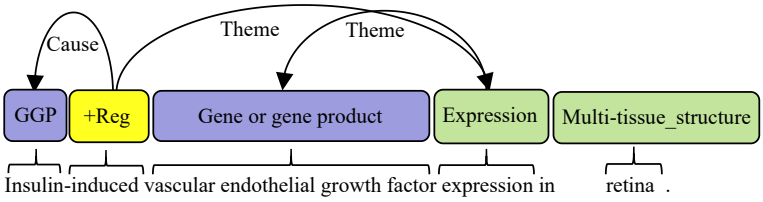

图1　包含两个事件的生物医学文本

Fig.1　Biomedical text containing two events

输入到支持向量机(support vector machine,SVM)中以完成生物医学事件提取。Zhou等[5]将外部语料库中学习的专业知识作为特征嵌入,以训练一种多核分类器。He等[16]使用SVM将复杂的特征结合起来,从而实现两阶段的生物医学事件触发词检测。然而,这些机器学习方法仍然需要大量的人力。

目前,基于神经网络的触发词检测已成为主流。Rahul等[17]使用循环神经网络提取句子中的复杂特征。Wang等[18]使用预训练模型FastText建立词级别的嵌入,并且结合双向长短期记忆网络(bi-directional long short-term memory,BiLSTM)和条件随机场(conditional random field,CRF)提高任务的性能。Wang等[19]将字符、词性、实体类型以及词嵌入作为特征嵌入,通过BiLSTM和CRF实现事件触发词检测。由于堆叠的学习方式容易丢失一些有效低级特征,Shen等[20]提出了一种基于卷积公路神经网络和极限学习机的检测框架。Wei等[21]为了克服触发词检测中单词表示的模糊性和隐藏层的特征提取不足问题,提出了一种基于上下文语境化的多层残差架构。He等[22]将触发词检测任务分解为两个简单的任务,针对每个阶段使用合适的分类器,缓解了数据不平衡。

上述方法虽然提高了触发词检测的性能,但忽视了依赖树中的句法结构。相比于主题词嵌入,基于依赖的词嵌入对触发词检测任务的贡献更大。He等[23]使用基于依赖的词嵌入和句子向量以丰富嵌入特征,并使用BiLSTM和注意力机制捕获句子中更重要的信息。Fei等[7]使用递归神经网络自动提取依存树中的句法特征,使用CRF解码整个句子的标签。Li等[8]将外部知识库的信息融入到基于树结构的递归神经网络中,从而获取更精准的特征表达。Li等[24]为了获得触发词标签之间的关联以及词嵌入和上下文特征之间的动态权重,引入了一种门控机制动态调整权重分布。此外,Zhao等[9]使用GCN捕获依赖树中的局部上下文,通过堆叠的超图聚合神经网络模块实现局部和全局上下文之间的细粒度交互。Huang等[25]基于生物医学知识库将知识划为分层知识图表示。

BERT(bidirectional encoder representations from transformers)[26]作为一种预训练模型,与传统的网络结构相比,它能够更好地捕获长期依赖关系。McCann等[27]展示了如何在适当的上下文中建模问答任务。基于问答的方式也逐渐成为主流。Wang等[28]将事件抽取任务建模为迭代问答任务,并且每个子任务使用相同的模型,从而达到联合学习的效果。

2　本文模型

如图2所示,本文模型主要由两个阶段组成,即事件类型识别和触发词定位。在识别阶段,首先,事件类型问题标识、原始文本和事件原型列表被连接起来作为问题的查询向量表示。然后,基于句法距离的注意力机制挖掘更有意义的上下文词特征表示,基于单词-实体-事件共现特征的事件类型感知注意力机制捕获词与事件之间的强关联性。最后,将事件类型感知上下文特征

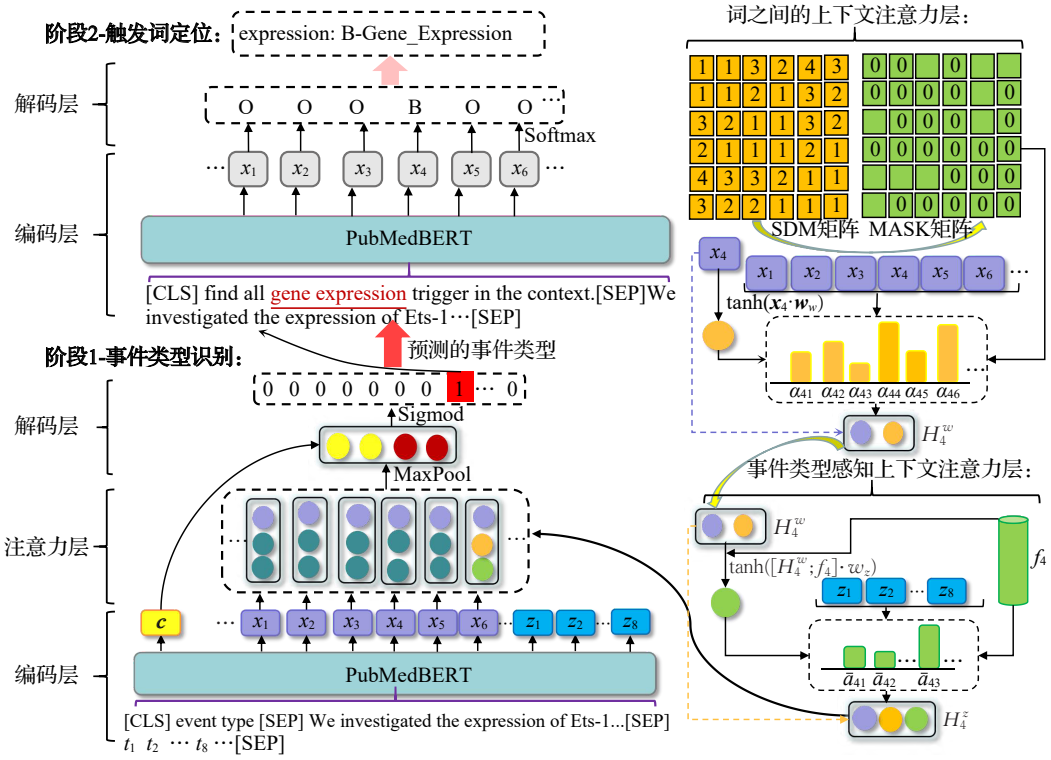

图2　本文提出的触发词检测架构

Fig.2　Proposed trigger detection architecture

与句子聚合特征联合起来回答句子中的候选事件类型。在定位阶段,首先,将每个候选事件类型的问题作为对句子的查询,使用与识别阶段相同的预训练编码器以获得每个词的深度上下文表示。然后,将深度上下文特征输入到全连接层,定位每个事件类型对应的触发词索引。

2.1　事件类型识别

2.1.1　上下文编码

给定由 L 个句子组成的生物医学文档,每个句子由 $W=\{w_1,w_2,\cdots,w_n\}$ 组成,其中 n 表示句子中单词的数量,“event type”表示事件类型识别的问题标识,事件原型列表 $T=\{t_1,t_2,\cdots,t_k,\text{null}\}$ 作为事件类型识别的辅助决策,其中 T 由所有事件类型名称和一个非事件类型名称“null”组成,k 是给定的事件类型的数量。本文将事件类型识别的问题标识、原始文本以及事件原型列表连接成顺序序列,如下所示:

[CLS] event type [SEP] W [SEP] T [SEP]

本文用预训练语言模型 PubMedBERT[29]编码上述序列,以获得词向量列表 $X=\{x_i,x_2,\cdots,x_n\}$ 以及事件原型向量列表 $Z=\{z_1,z_2,\cdots,z_{k+1}\}$。WordPiece 分词器可能将一个词拆分成多个子词,因此拆分后的第一个子词被作为该词的上下文表示。

2.1.2　基于句法距离的上下文特征

一个词在不同的上下文中可能表达不同的含义并触发不同的事件,因此句子中的事件类型识别也依赖于特定的上下文发生。然而,通过单一的注意力机制或 GCN 融合的上下文特征包含了与触发词无关的冗余信息。为了聚焦与事件类型相关的重要特征,当两个词间的句法距离大于阈值 α 时,它们将不参与彼此之间的语义交互。本文使用句子的依赖关系计算单词之间的句法距离,如图 3 所示。句法距离矩阵 $SDM\in\mathrm{R}^{n\times n}$ 中 SDM_{ij} 表示单词 w_i 到单词 w_j 之间的句法距离。通过构造基于图结构的掩码矩阵 $MASK$ 实现强依赖的词间语义信息交互:

$$MASK_{ij}=\begin{cases}0, & SDM_{ij}\leqslant\alpha\\ -\infty, & \text{otherwise}\end{cases} \tag{1}$$

从图 3 的依赖关系解析可知,induced 与 expression 之间的顺序距离为 4,与 Insulin 之间的顺序距离为 2。如果仅考虑顺序距离,所有顺序距离小于 4 的单词都将纳入计算范围,以至于融入不相关的噪声。本文方法引入句法距离,induced 到 expression 和 Insulin 之间分别存在一条依赖边,即 induced 与 expression 和 Insulin 之间的句法距离都为 1,因此能够减少无关上下文信息的交互,帮助识别复杂的语义结构。本文将掩码矩阵融合到注意力机制中来关注更有意义的上下文特征:

$$H_i^w=\sum_{j=1}^{n}e_{ij}\cdot x_j \tag{2}$$

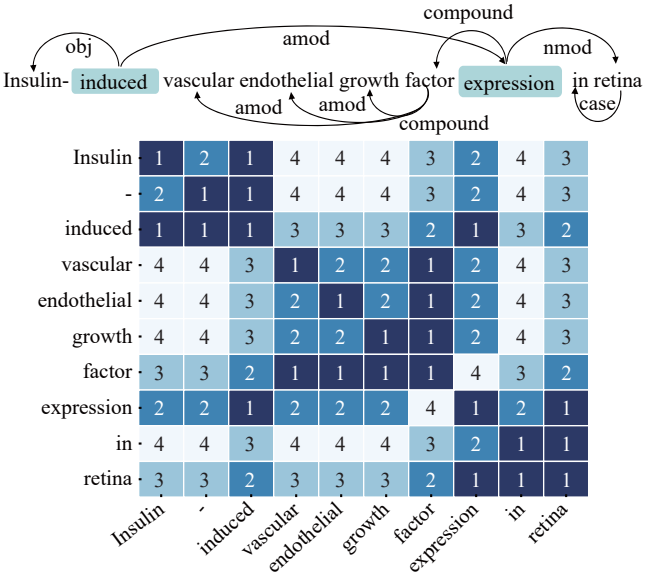

图 3　生物医学文本示例的句法距离矩阵

Fig.3　Syntactic distance matrix of biomedical text example

$$e_{ij}=\mathrm{softmax}(score(x_i,x_j)+MASK_{ij}) \tag{3}$$

$$score(x_i,x_j)=\frac{1}{\sqrt{d_j}}\tanh(x_i\cdot W_w)\cdot x_j \tag{4}$$

其中,d_j 是 x_j 的嵌入维度,$W_w\in\mathrm{R}^{d_j\times d_j}$ 是参数矩阵,$score(x_i,x_j)$ 表示词向量 x_i 与 x_j 的注意力分数,$score()$ 是打分函数,e_{ij} 表示 x_i 和 x_j 之间归一化后的注意力权值,H_i^w 是 w_i 的上下文特征表示。

2.1.3　单词-实体-事件共现特征

为了挖掘实体与事件之间的潜在特征,避免变化的上下文对模型的预测产生干扰,本文提出基于全局统计的单词-实体-事件共现特征指导模型学习,该特征是在不依赖任何外部数据的情况下从整个训练集中统计出来的。当触发词与实体之间的句法距离小于阈值 δ 时,它们之间存在强关联性。如图 4 所示,在句子 S1 中,存在一个待抽取事件,且触发词 expression 与 Gene or Gene Product 类型实体 Est-1 之间的句法距离是 1。然而,在句子 S2 中,虽然包含单词 expression,但与其最近的 Gene or Gene Product 类型实体 LRP-1 的句法距离为 2,不发挥触发作用。在句子 S3 中,虽然包含单词 expression,但不存在任何 Gene or Gene Product 类型的实体,也没有发挥触发作用。因此,根据单词-实体-事件之间的句法依存关系,可以帮助事件类型的识别。

为了获取词 w_i 的单词-实体-事件共现特征向量 $f_i=(f_i^1,f_i^2,\cdots,f_i^k,1)$,首先标记每个词对应的实体标签,非实体用 O 标记,然后统计训练文档中每个事件类型 t_j 对应的实体标签共现列表 Et_j,即遍历所有句子。如果句子中单词的事件类型是 t_j,则捕捉与该单词的句法距离在 δ 内的词的实体标签列表并去重,如果该列表中的元素数量不大于 δ,则将这些元素加入到 Et_j 中,大于

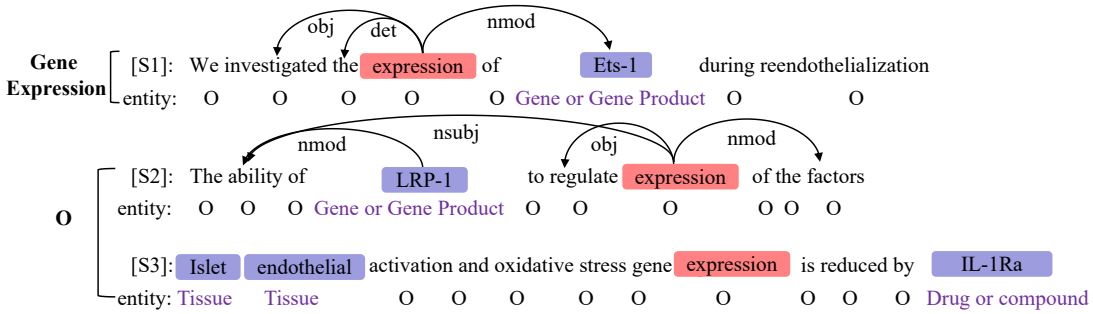

图4　包含Gene Expression事件的生物医学文本

Fig.4　Biomedical texts containing "Gene Expression" events

则舍弃。这种处理方式能专注每个事件类型中起到关键作用的实体类型,减弱噪声实体的影响。最后,如果与 w_i 的句法距离在 δ 内的词的实体标签列表 Ew_i 与 Et_j 存在交集,则 $f_i^j=1$,反之,则 $f_i^j=0$。

2.1.4　事件类型感知上下文特征

本文将事件原型列表作为辅助决策,通过事件类型感知注意力机制捕获词与事件类型之间的语义关联性。为了防止关联语义受到上下文变化的影响,单词-实体-事件共现特征向量通过直接和间接的方式指引注意力的关联方向。对于词 w_i,直接方式:将 H_i^w 与 f_i 级联起来作为注意力机制的目标特征输入;间接方式:给定一个非事件类型"null"对应的one-hot编码 β,其中 $\beta=(0,0,\cdots,1)\in\mathrm{R}^{1\times|T|}$,如果 $f_i^{m_i}=0$(m_i 是词向量 x_i 与事件原型列表向量 Z 之间的初始注意力分数向量 $\boldsymbol{\Gamma}_i$ 中最大值对应的下标),则将 β 作为 x_i 与 Z 之间注意力分数向量 $\overline{\boldsymbol{\Gamma}}_i$,反之,将 $\boldsymbol{\Gamma}_i$ 与 f_i 点积后的结果作为 $\overline{\boldsymbol{\Gamma}}_i$。这种方式使每个词只关注符合单词-实体-事件共现的事件类型,词 w_i 的事件类型感知上下文特征表示 H_i^z 为:

$$H_i^z=\mathrm{softmax}(\overline{\boldsymbol{\Gamma}}_i)\cdot Z \tag{5}$$

$$\overline{\boldsymbol{\Gamma}}_i=\begin{cases}\boldsymbol{\Gamma}_i\cdot f_i, f_i^{m_i}=1\\ \beta, f_i^{m_i}=0\end{cases} \tag{6}$$

$$m_i=\arg\max(\boldsymbol{\Gamma}_i) \tag{7}$$

$$\boldsymbol{\Gamma}_i=(\rho(x_i,z_1,f_i),\rho(x_i,z_2,f_i),\cdots,\rho(x_i,z_{k+1},f_i)) \tag{8}$$

$$\rho(x_i,z_j,f_i)=\frac{1}{\sqrt{d_j}}\tanh([H_i^w;f_i]\cdot W_z)\cdot z_j \tag{9}$$

其中,d_j 是 z_j 的嵌入维度,$W_z\in\mathrm{R}^{(d+|T|)\times d}$ 是可学习的参数矩阵,$\rho(x_i,z_j,f_i)$ 表示词向量 x_i 与事件类型向量 z_j 的注意力分数,$\rho(\cdot)$ 是打分函数,$\mathrm{softmax}(\overline{\boldsymbol{\Gamma}}_i)$ 是 w_i 的事件类型感知注意力权值向量。

事件类型感知注意力权值向量的目的是给予相关的事件类型更多关注。为了实现这一目的,本文将句子中每个词对应的事件类型构造为向量表示形式,使用注释过的事件类型向量作为真实的注意力分数向量,即该词真实对应的事件类型获得全部的注意力,然后将其作为监督训练事件类型感知注意力机制。

2.1.5　事件类型识别解码

由于事件类型识别是句子级别的,需要将捕捉到的词级别的事件类型感知上下文特征表示 $H^z=\{H_1^z,H_2^z,\cdots,H_n^z\}$ 送入最大池化层,以获得句子级别的事件类型感知上下文特征表示 Ts:

$$Ts=MaxPool(H^z) \tag{10}$$

其中,$MaxPool(\cdot)$ 是最大池化函数,它将词的事件类型感知上下文特征映射到由句子表示的事件类型感知语义特征。

在PubMedBERT编码的序列中,第一个输入[CLS]作为句子聚合后的向量表示 c。本文将句子级别的事件类型感知上下文特征 Ts 与句子聚合向量 c 级联起来输入到一个具有非线性激活函数的全连接层中,以回答句子中包含的事件类型:

$$E_{\mathrm{tp}}=f(W_{\mathrm{e}}[c;Ts]+b_{\mathrm{e}}) \tag{11}$$

其中,$f(\cdot)$ 是sigmoid激活函数,$W_{\mathrm{e}}\in\mathrm{R}^{2d\times|T|}$ 是参数矩阵,b_{e} 是偏置向量,$E_{\mathrm{tp}}\in\mathrm{R}^{1\times|T|}$ 是句子对应所有事件类型的类别概率。

2.2　触发词定位

2.2.1　上下文编码

给定句子中的词 $W=\{w_1,w_2,\cdots,w_n\}$ 和第2.1节预测的事件类型列表 $T=\{t_1,t_2,\cdots,t_z\}$,为了能定位句子中每个事件类型对应的触发词,本文针对每个事件类型构造了一个简单的问题模板:"find all [event type] in the context."。与事件检测相同,本文将问题模板与句子连接成序列,如下所示:

[CLS] find all t_i in the context. [SEP] W [SEP]

相关研究[30]表示,多任务能有效地提升模型性能,因此本文使用与第2.1节相同的PubMedBERT作为编码器对整个序列进行编码,以获得句子中每个词的向量表示 $X=\{x_1,x_2,\cdots,x_n\}$。

2.2.2　触发词定位解码

由于一个生物医学触发词可能由多个词组成,为了能够准确定位触发词,本文将答案转为"BIO"的形成,即对于每个问题,解码器只回答当前词是"B""I""O"中

的一种即可,这有效地降低了预测的复杂度。本文将句子中每个词的向量表示 x_i 输入到一个具有softmax激活函数的全连接网络,以定位该词的概率分布:

$$p(\boldsymbol{y}_i|\boldsymbol{x}_i)=\text{softmax}(\boldsymbol{W}_{\text{tr}}\boldsymbol{x}_i+\boldsymbol{b}_{\text{tr}}) \tag{12}$$

其中,$\boldsymbol{W}_{\text{tr}}\in\mathrm{R}^{d\times 3}$ 是参数矩阵,$\boldsymbol{b}_{\text{tr}}$ 是偏置向量,$p(\boldsymbol{y}_i|\boldsymbol{x}_i)$ 表示单词 $\boldsymbol{w}_i$ 对应标记类型 $\boldsymbol{y}_i$ 的类别概率。

2.3 损失函数

假设生物医学文档训练集由 L 个句子组成,n 是句子中的单词数量。本文的损失函数由三部分组成:(1)事件类型识别的损失函数;(2)触发词定位的损失函数;(3)监督事件类型感知注意力机制的损失函数。首先,在事件类型识别中,用 $J(\theta)$ 衡量目标与输出之间的二元交叉熵。然后,触发词定位的损失函数是目标与输出之间的多分类交叉熵 $Y(\theta)$。最后,本文将平方误差函数 $D(\theta)$ 作为监督事件类型感知注意力机制的损失。模型整体的损失函数 $Loss$ 构造如下所示:

$$J(\theta)=\sum_{j=1}^{L}\sum_{i\in\triangledown}\text{lb}\,p_{\text{tp}}(E_{\text{tp}i}^{*}|E_{\text{tp}i}) \tag{13}$$

$$Y(\theta)=\sum_{j=1}^{L}\sum_{i=1}^{n}\text{lb}\,p(\boldsymbol{y}_i|\boldsymbol{x}_i) \tag{14}$$

$$D(\theta)=\sum_{j=1}^{L}\sum_{i=1}^{n}(\bar{\boldsymbol{\alpha}}^{*}_{i}-\bar{\boldsymbol{\alpha}}_i)^2 \tag{15}$$

$$Loss=Y(\theta)+J(\theta)+D(\theta) \tag{16}$$

其中,$\triangledown$ 是事件类型集合,$E_{\text{tp}i}^{*}$ 是第 i 个事件类型对应的真实值,$E_{\text{tp}i}$ 是第 i 个事件类型对应的类别概率,$\bar{\boldsymbol{\alpha}}_i$ 是预测的词 $\boldsymbol{w}_i$ 的事件类型感知注意力权值向量,$\bar{\boldsymbol{\alpha}}_i^{*}$ 是注释的词 $\boldsymbol{w}_i$ 的事件类型感知注意力权值向量。

本文的触发词检测由两个阶段共同构建,即事件类型识别阶段和触发词定位阶段,$J(\theta)$ 和 $D(\theta)$ 分别由识别阶段的事件类型感知注意力层和识别解码层产生,$Y(\theta)$ 是由定位阶段的定位解码层生成。模型的训练并不是对每个阶段进行独立的学习,而是综合两个阶段的三个损失进行联合优化和训练,最终的优化目标是使模型整体的 $Loss$ 尽可能最小。在训练时,触发词定位阶段采用正确的事件类型构造问题模板定位句子中的触发词,计算触发词定位的损失;在验证和测试时,触发词定位阶段采用事件类型识别阶段的预测结果定位触发词。

3 实验

3.1 数据集

为了评估本文模型的性能,在生物医学事件抽取语料库MLEE[15]中进行了触发词检测的实验。MLEE语料库已经广泛应用在生物医学事件抽取任务中,它从PubMed摘要中生成了262个生物医学文档、2 608个句子和6 677个事件。相关的事件类型分为解剖、分子、一般和计划四大类,一共包含了19个预定义的事件类型和14个预定义的实体。为了在划分数据集时避免人为因素的干扰,本文使用MLEE语料库提供的原始拆分标准作为训练集、验证集和测试集。MLEE数据集的划分情况总结如表1所示。在本文实验中,训练集用于训练模型以及确定模型权重参数,验证集用于调整模型的超参数和对模型的能力进行初步评估,并从验证集的所有epoch中选择最优性能的模型作为最终的训练结果。此外,本文所有的对比实验结果都在测试集中进行。MLEE的事件类型静态分布如表2所示。

表1 MLEE数据集的统计信息

Table 1 Statistical information for MLEE dataset

标注	训练集	验证集	测试集	全部
文档	131	44	87	262
句子	1 271	457	880	2 608
事件	3 296	1 175	2 206	6 677

表2 MLEE数据集的事件类型统计

Table 2 Event type statistics for MLEE dataset

事件大类	事件类型	训练集	测试集
Anatomical	Cell Proliferation	82	43
	Development	202	98
	Blood Vessel Development	540	305
	Death	57	36
	Breakdown	44	23
	Remodeling	22	10
	Growth	107	56
Molecular	Synthesis	13	4
	Gene Expression	210	132
	Transcription	16	7
	Catabolism	20	4
	Phosphorylation	26	3
	Dephosphorylation	2	1
General	Localization	282	133
	Binding	102	56
	Regulation	362	178
	Positive Regulation	654	312
	Negative Regulation	450	233
Planned	Planned Process	407	175

3.2 参数设置与评价指标

本文模型使用Pytorch深度学习框架和Python3编程语言。Linux服务器配置:CPU为Intel® Xeon® W-2223 CPU@3.60 GHz,GPU为Quadro RTX 48 GB,内存为128 GB。模型是基于PubMedBERT实现的,其中batch_ size设置为32,epoch设置为30,学习率设置为2E−5,词间的句法距离阈值 α 和实体间的句法距离阈值 δ 都设置为2,Stanford解析器[31]生成MLEE中每个句子的依赖树结构,Adam被作为随机梯度下降算法,最大的句子长度由每个batch的最大长度动态决定。为了评估模型的性能,本文沿用之前研究使用的触发词检测匹配原则:如果触发词被正确检测,则触发词的偏移量完全

匹配标注的触发词,触发词对应的事件类型完全匹配标注的事件类型。本文使用NLP任务中常用的评价指标,即精确率(precision,P)、召回率(recall,R)和$F1$值评估模型性能。精确率是针对模型预测结果而言的,其表示的是在所有预测为触发词的样本中实际为触发词的比例。召回率是针对实际样本而言的,其表示的是在所有实际为触发词的样本中被模型预测为触发词的比例。然而,仅考虑精确率和召回率其中的一项无法反映出模型性能的好坏,因此选用$F1$值作为模型性能的总体评价指标,其表示精准率和召回率的调和平均数。相关计算公式如下:

$$P=\frac{TP}{TP+FP} \tag{17}$$

$$R=\frac{TP}{TP+FN} \tag{18}$$

$$F1=2\times\frac{P\times R}{P+R} \tag{19}$$

其中,TP、FP、FN分别表示模型预测为正类的正样本数量,预测为负类的负样本数量,以及预测为负类的正样本数量。

3.3　实验结果

为了证明本文模型的有效性,将它与以下最先进的方法进行比较:(1)Pyysalo等[15]提出了一种人为设计特征与SVM分类器相结合的模型;(2)Zhou等[5]将外部语料库中的生物医学领域的知识作为特征嵌入,来训练一种多核分类器;(3)Wang等[18]通过BiLSTM-CRF模型将触发词检测视为序列标注任务;(4)He等[16]为了缓解类别不平衡问题,将触发词检测划分为识别阶段和分类阶段,并且为每个阶段匹配更合适的特征嵌入;(5)Shen等[20]将端到端的卷积神经网络和极限学习机结合,从多个角度学习高维语义特征;(6)Li等[24]融合了基于依赖的词嵌入、门控机制和注意力层,并充分利用上下文标签之间的关联;(7)Fei等[7]提出了一种基于递归神经网络的模型以学习依赖树中的句法信息,并使用CRF对句子的预测进行联合编码;(8)Diao等[32]结合细粒度双向长短期存储器和SVM,通过充分捕捉各层次的语义特征来检测触发词;(9)He等[22]引入基于注意力机制的BiLSTM提取语义特征,并使用在线算法对语义进行分类;(10)Wei等[21]提出了一种基于多层残差的BiLSTM架构,通过反复提取特征以避免梯度的爆炸或消失;(11)He等[33]结合句子嵌入和多层次关注的方式解决事件类型歧义的问题。

不同模型在MLEE数据集上的结果如表3所示。为了保证实验结果的公平性和可信性,所有模型的预测结果判定标准都严格遵守第3.2节中描述的触发词检测匹配原则。从表3中可以观察到:

(1)文献[15]和文献[5]使用SVM对生物医学触发词进行分类,并且在传统的机器学习方法中取得了最优异的性能,但是需要人工去构造大量的特征作为模型的输入。与这些方法相比,深度学习的方法在触发词识别方面具有更好的效果,例如文献[20]的$F1$分数比文献[5]增加了2.25个百分点,文献[24]的$F1$分数比文献[15]高4.91个百分点,本文模型的$F1$分数分别比文献[15]和文献[5]增加了5.72个百分点和3.07个百分点。造成这一差异的主要原因是深度学习的方法可以更好地学习句子中潜在的语义特征。

表3　不同模型的触发词检测结果

Table 3　Trigger detection results of different models

单位:%

模型	P	R	$F1$
文献[15]	70.65	81.46	75.67
文献[5]	75.35	81.60	78.32
文献[18]	77.89	78.28	78.08
文献[16]	80.35	79.16	79.75
文献[20]	80.06	81.25	80.57
文献[24]	80.74	80.42	80.58
文献[7]	81.12	79.15	80.28
文献[32]	80.03	81.54	80.66
文献[22]	80.88	79.65	80.26
文献[21]	79.89	**81.61**	80.74
文献[33]	82.01	78.02	79.96
本文	**82.03**	80.75	**81.39**

(2)文献[24]、文献[7]和本文模型是融合了依赖解析的触发词检测模型。与其他没有基于句法信息的模型相比,它们普遍拥有较高的$F1$分数。文献[24]的$F1$分数比文献[18]高了2.50个百分点,文献[7]的$F1$分数比文献[16]高了0.53个百分点。文献[7]与文献[18]相比,只是将基于序列的循环神经网络替换成基于树结构的递归神经网络就提高了识别性能。就$F1$分数而言,本文模型分别比文献[24]和文献[7]高0.81个百分点和1.11个百分点。这些结果印证了依赖信息能够为触发词检测提供关键信息。

(3)文献[22]、文献[33]和本文模型都引入了不同方式的注意力机制。与没有注意力机制的模型相比,文献[22]比文献[18]的$F1$值高了2.18个百分点,相比于文献[16],文献[33]的$F1$分数增加了0.21个百分点。此外,本文模型的$F1$分数分别比文献[22]和文献[33]增加1.13个百分点和1.43个百分点。这说明了基于注意力的模式给触发词检测带来实质性的帮助,并且本文提出的注意力能够捕获更有意义的上下文信息。

(4)文献[20]、文献[24]、文献[32]和本文模型都引入了实体类型特征。与不引入实体类型的模型相比,文献[20]比文献[7]的$F1$分数高了0.29个百分点,文献[32]比文献[22]增加了0.40%。另外,本文模型的$F1$分数分别比文献[20]、文献[24]和文献[32]高0.82个百分点、0.81个百分点和0.73个百分点。实验结果说明了基于全局统计的实体类型特征能够有效地辅助神经网络做出决策。

表4　不同模型在19种事件类型上的详细结果

Table 4　Detailed results of different models on 19 event types

单位:%

事件类型	文献[16]			文献[15]			文献[5]			本文		
	P	*R*	*F*1	*P*	*R*	*F*1	*P*	*R*	*F*1	*P*	*R*	*F*1
Cell Proliferation	78.4	67.4	72.5	63.8	69.8	66.7	78.4	67.4	72.5	**94.4**	**79.1**	**86.1**
Development	72.9	80.4	76.5	68.1	**83.5**	75.0	69.3	81.4	74.9	**77.1**	82.7	**79.8**
Blood Vessel Development	98.2	94.7	96.4	85.7	96.3	96.0	**98.7**	**97.3**	**98.0**	97.3	95.7	96.5
Death	68.9	86.1	76.5	56.9	**94.3**	71.0	**72.1**	88.6	**79.5**	70.0	77.8	73.7
Breakdown	85.7	54.5	66.7	80.0	34.8	48.5	80.0	34.8	48.5	**100.0**	**65.2**	**78.9**
Remodeling	83.3	50.0	62.5	85.7	60.0	70.6	85.7	60.0	70.6	**87.5**	**70.0**	**77.8**
Growth	79.7	83.9	81.7	69.1	83.9	75.8	77.1	83.9	80.3	**96.3**	**92.9**	**94.6**
Synthesis	66.7	50.0	57.1	33.3	50.0	40.0	40.0	50.0	44.4	**80.0**	**100.0**	**88.9**
Gene Expression	**87.1**	92.0	89.5	83.8	**93.9**	88.6	84.7	92.4	88.4	86.6	**93.9**	**90.1**
Transcription	**100.0**	16.7	28.6	25.0	14.3	18.2	0	0	0	50.0	**28.6**	**36.4**
Catabolism	25.0	**33.3**	28.6	0	0	0	16.7	**33.3**	22.2	**50.0**	**33.3**	**40.0**
Phosphorylation	50.0	**100.0**	66.7	50.0	**100.0**	66.7	**75.0**	**100.0**	**85.7**	**75.0**	**100.0**	**85.7**
Dephosphorylation	0	0	0	0	0	0	**100.0**	**100.0**	**100.0**	0	0	0
Localization	81.9	78.2	80.0	79.9	83.5	81.6	80.9	**85.7**	**83.2**	**82.5**	78.2	80.3
Binding	**92.7**	76.9	78.4	84.0	76.4	**80.0**	81.1	**78.2**	79.6	87.5	75.0	77.8
Regulation	61.9	**61.6**	61.8	46.5	60.4	52.5	56.5	53.1	54.7	**63.4**	57.3	**60.2**
Positive Regulation	78.0	83.8	80.8	67.9	**86.7**	76.1	71.6	86.4	78.3	**81.7**	84.8	**83.2**
Negative Regulation	**80.3**	75.3	77.3	74.4	77.0	75.7	77.1	78.8	78.0	78.9	**87.4**	**83.0**
Planned Process	**75.9**	71.1	**78.4**	53.9	75.0	62.7	56.6	**75.6**	64.7	68.7	57.7	62.7

(5)与表3中的这些模型相比,本文提出的模型实现了81.39%的*F*1分数,显著优于其他最先进模型。本文模型的精度仍然是一个有竞争力的结果,并且三个评价指标都大于80%也说明本文模型的稳定性。这主要归因于基于句法距离的上下文注意力机制避免了噪声的干扰,基于单词-实体-事件共现特征的事件类型感知注意力机制聚合了词与事件之间的关联信息。此外,两阶段问答范式在降低了任务复杂度的同时也缓解了数据稀缺问题。

3.4　在事件类型上的详细比较

为了分析本文提出的模型在每个事件类型上的检测性能,表4揭示了文献[16]、文献[15]、文献[5]和本文模型在所有19类事件类型上的详细实验结果。

从表4中可以看出:(1)在*F*1分数方面,本文模型在19类事件类型中的13类都具有最优的性能。在"Synthesis""Catabolism"和"Transcription"事件类型中,本文模型的*F*1分数分别比三个基线的平均分数高了41.7个百分点、23.1个百分点和20.8个百分点。(2)在召回率方面,本文模型在10类事件类型中优于其他基线模型,相比于三个基线模型的均值,在"Cell Proliferation""Breakdown"和"Growth"事件类型中,本文模型分别多识别出了10.9个百分点、23.8个百分点和9.0个百分点的正实例。(3)在精确率方面,本文模型在11类事件类型上表现出色。总之,在19类事件类型上的详细结果证明了本文模型的有效性。

为了进一步做误差分析,本文模型在每个事件类型上的预测结果如图5中的混淆矩阵所示。其中混淆矩阵中的颜色密度与预测的实例数匹配,即实例数越大颜色越深,混淆矩阵中的横坐标和纵坐标是由19类事件类型的缩写和一个表示非事件类型的"Other"组成,例如坐标"Bvd"是"Blood Vessel Development"的缩写。如图5所示,除了"Cell""Pos""Reg"和"Trans"类别外,剩余类别中均未出现其他事件类型的预测实例,这也印证了模型区分不同事件类型的能力。此外,影响模型性能的主要因素是"Plan""Neg"和"Reg"等类别中的一部分实例被预测为非事件类型,这主要是因为原始文档中存在一些触发词有时被标记为特定事件类型,有时不被标记,但是它们却共享高度相似的上下文语境和词-实体-事件共现特征,这种模糊性导致模型预测错误。

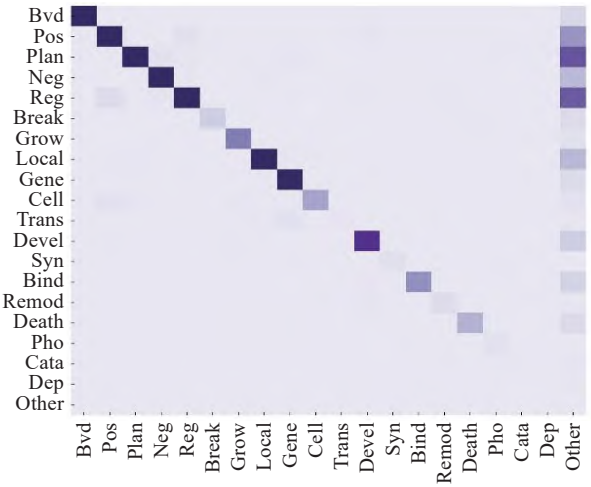

图5　预测值与真值的混淆矩阵

Fig.5　Confusion matrix of predicted value and true value

3.5　不同组件对模型的影响

为了进一步验证不同组件对模型的影响,本文设计了不同的子网络以评估各个组件的有效性。如表5所示,单阶段是直接将触发词检测视为序列标注任务的子网络,两阶段是指两阶段问答范式子网络,两阶段+AT1是加入了上下文注意力的子网络,两阶段+MASK-AT1是加入了句法距离的上下文注意力的子网络,两阶段+MASK-AT1+AT2是加入事件类型感知注意力的子网络,两阶段+MASK-AT1+Entity-AT2是本文提出的完整的模型。

表5　不同组件对模型的影响

Table 5　Impact of different components on model

单位:%

模型	*P*	*R*	*F*1
两阶段+MASK-AT1+Entity-AT2	**82.03**	80.75	**81.39**
两阶段+MASK-AT1+AT2	79.42	**81.53**	80.46
两阶段+MASK-AT1	79.74	80.53	80.13
两阶段+AT1	79.72	79.37	79.54
两阶段	79.62	79.31	79.47
单阶段	79.14	78.44	78.79

(1)由表5可以观察到,两阶段比单阶段拥有更好的性能,*F*1分数高了0.68个百分点,印证了本文提出的两阶段问答范式的有效性。它不仅利用PubMedBERT提升了模型的性能,并且拆分的两个子任务在降低问题复杂度的同时也利用丰富的问答数据库实现数据增强。

(2)为了验证基于句法距离的上下文注意力给模型带来的优势,由表5可知,两阶段+AT1的*F*1分数比两阶段高0.07个百分点,两阶段+MASK-AT1比两阶段+AT1的*F*1分数高0.59个百分点。这说明依赖树中的句法距离为变化的上下文提供了关键信息,导致携带的重要语义信息能够帮助模型更多地关注词的基本论点。

(3)为了证明基于单词-实体-事件共现特征的事件类型感知注意力机制的有效性,由表5可知,两阶段+MASK-AT1+AT2比两阶段+MASK-AT1拥有更出色的性能。这是因为事件类型感知上下文注意力机制加强了词与事件类型之间的交互。两阶段+MASK-AT1+Entity-AT2比两阶+MASK-AT1+AT2的*F*1值高了0.93个百分点。优异的结果主要归功于事件与实体之间的强关联性被单词-实体-事件共现特征表示出来,并且在不受变化上下文干预的情况下,纠正事件类型感知注意力权重。

3.6　超参数敏感性分析

本文根据在MLEE语料库上的生物医学事件触发词检测结果验证超参数对本文模型的影响。本文描述了两个最重要的超参数的敏感性分析,包括词间的句法距离阈值α和实体间的句法距离阈值δ。图6展示了两个超参数在取不同值时的三个评价指标分数,每个子图中的结果都是在固定δ下获得的。

词间的句法距离阈值α:从图6可以发现,在*F*1值和精确率方面,随着α的值从1变为2,在变化δ的情况下模型性能都有所提升。然而,伴随着α从2逐渐增大,所提出的模型性能开始有所下降。在召回率方面,从每个子图中发现,伴随着α的逐渐增大,召回率在一

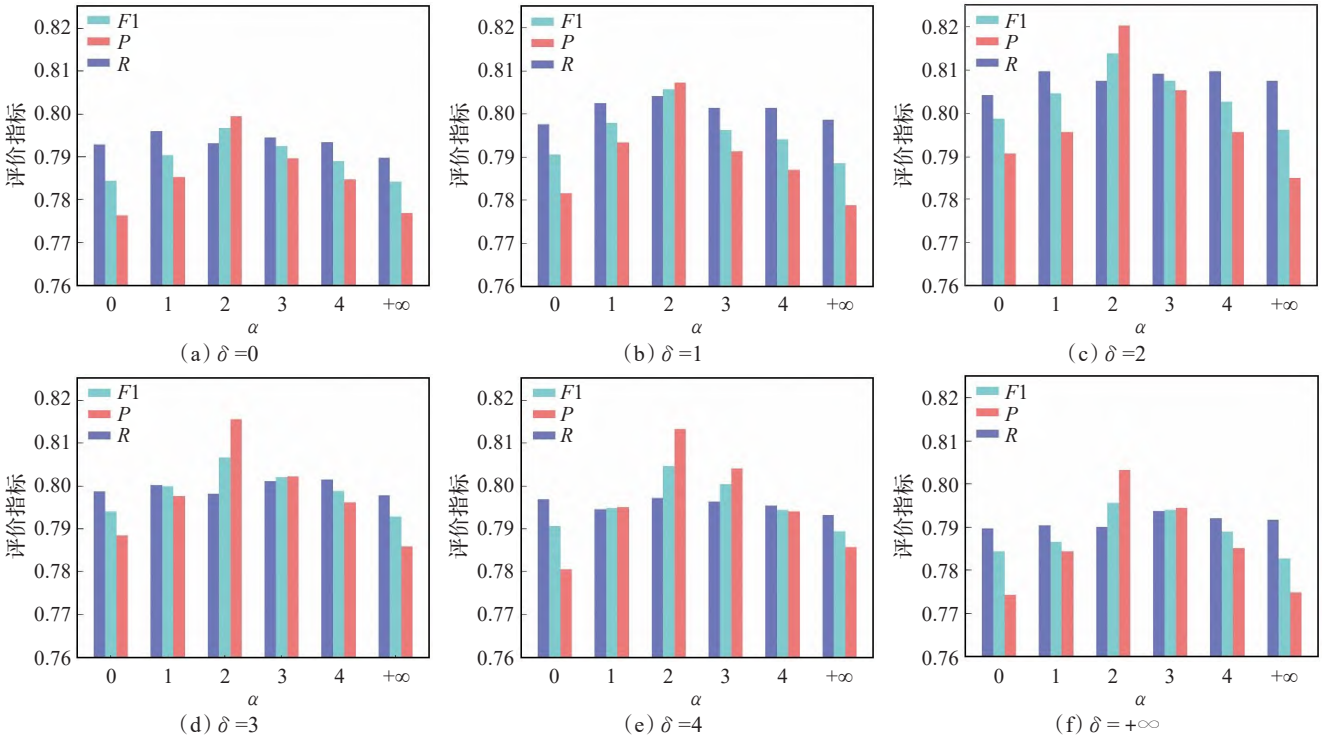

图6　超参数敏感性分析

Fig.6　Hyperparametric sensitivity analysis

定的区间呈现小幅度波动。产生这一现象的原因主要是当 α 过小时,依存关系中相隔较远且重要的词被遗忘,当 α 过大时,更多的噪声被融合到上下文特征中,从而导致更多的单词被误判为触发词,降低了模型性能。此外,当 α 值为0时的模型性能明显低于 α 值为1时的模型性能。这是因为当 α 为0时,每个词的上下文注意力权值仅关注该单词本身,即基于句法距离的词间注意力机制失去了原有的作用。同时,当 α 为 $+\infty$ 时,句法距离并没有发挥任何作用,即基于句法距离的词间注意力机制退化为普通的注意力机制。因此,其模型性能低于 α 值为4时的模型性能。

实体间的句法距离阈值 δ :从图6(b)到图6(e)的结果中可以发现,模型的 $F1$ 值、精确率以及召回率先呈现上升趋势,在图6(c)中的整体性能达到最优后,则呈现下降趋势。导致这一趋势的主要原因是伴随着 δ 的增加,更多的实体类型被误认为与事件类型之间具有强关联性,从而忽略了真正对事件类型起绝对作用的实体类型,导致句子中出现误判的事件类型,影响后续的触发词定位。此外,当 δ 值为0时,无法找出与事件类型共现的实体特征,即起关键作用的实体类型被忽视。当 δ 值为 $+\infty$ 时,句子中所有的实体类型都被误认为是与事件类型相关的共现特征,淡化了起关键作用的实体类型,导致事件类型识别的模糊性。因此,如图6(a)和图6(f)所示, δ 值为0和 $+\infty$ 时的模型性能是最差的。

3.7　两阶段问答范式的可行性

为了证明所提出的两阶段问答范式的可行性,本文在测试阶段不使用预测出的事件类型,而是引入每个句子的真实事件类型生成问题模板。模型结果如图7所示,它的三个评价指标 P、R 和 $F1$ 分数分别是89.17%、90.85%和90.00%。这个结果说明了一个高性能的事件类型识别方法对于触发词的检测至关重要,也印证了两阶段问答范式的有效性与可行性。

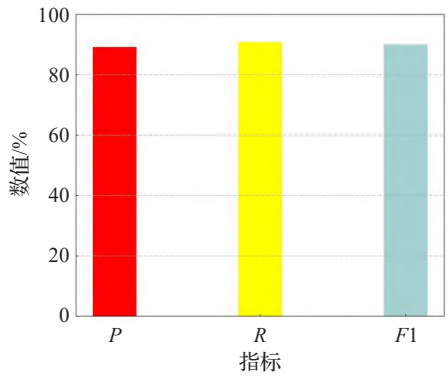

图7　真实事件类型的结果图

Fig.7　Results graph for real event types

4　结束语

本文提出了一种两阶段问答范式的生物医学事件触发词检测方法。它将触发词检测任务分为事件类型识别阶段和触发词定位阶段,引入依赖树中的依存关系构造句法距离矩阵,融合句法距离后的注意力机制有能力只关注词之间更重要的上下文特征;基于全局统计的单词-实体-事件共现特征指导注意力机制挖掘词与事件类型之间的强关联性。同时,两阶段的方式在利用问答的最新进展提升检测性能的同时也有效缓解了数据稀缺问题。在MLEE语料库上的各种实验证明了本文模型的有效性与可行性。未来,针对训练语料库的不足,本文考虑将迁移学习应用在触发词检测中。针对触发词定位错误,本文将探索更加有效的定位方法。

参考文献:

[1] ANANIADOU S, THOMPSON P, NAWAZ R, et al. Event-based text mining for biology and functional genomics[J]. Briefings in Functional Genomics, 2015, 14(3): 213-230.

[2] ROTMENSCH M, HALPERN Y, TLIMAT A, et al. Learning a health knowledge graph from electronic medical records[J]. Scientific Reports, 2017, 7(1): 1-11.

[3] SPANGHER A, PENG N, MAY J, et al. Enabling low-resource transfer learning across COVID-19 corpora by combining event-extraction and co-training[C]//Proceedings of the 1st Workshop on NLP for COVID-19 at ACL 2020, 2020.

[4] WANG J, ZHANG J, AN Y, et al. Biomedical event trigger detection by dependency-based word embedding[J]. BMC Medical Genomics, 2016, 9(2): 123-133.

[5] ZHOU D, ZHONG D, HE Y. Event trigger identification for biomedical events extraction using domain knowledge[J]. Bioinformatics, 2014, 30(11): 1587-1594.

[6] CUI S, YU B, LIU T, et al. Edge-enhanced graph convolution networks for event detection with syntactic relation[J]. arXiv:2002.10757, 2020.

[7] FEI H, REN Y, JI D. A tree-based neural network model for biomedical event trigger detection[J]. Information Sciences, 2020, 512: 175-185.

[8] LI D, HUANG L, JI H, et al. Biomedical event extraction based on knowledge-driven tree-LSTM[C]//Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1, Long and Short Papers), 2019: 1421-1430.

[9] ZHAO W, ZHANG J, YANG J, et al. A novel joint biomedical event extraction framework via two-level modeling of documents[J]. Information Sciences, 2021, 550: 27-40.

[10] AHMAD W U, PENG N, CHANG K W. GATE: graph attention transformer encoder for cross-lingual relation and event extraction[C]//Proceedings of the 35th AAAI Conference on Artificial Intelligence, 2021: 12462-12470.

[11] LI L, LIU Y. Exploiting argument information to improve

biomedical event trigger identification via recurrent neural networks and supervised attention mechanisms[C]//Proceedings of the 2017 IEEE International Conference on Bioinformatics and Biomedicine, 2017: 565-568.

[12] LI F, PENG W, CHEN Y, et al. Event extraction as multi-turn question answering[C]//Findings of the Association for Computational Linguistics: EMNLP 2020, 2020: 829-838.

[13] RONG W, NIE Y, SHEN Y, et al. Biological event trigger identification with noise contrastive estimation[J]. IEEE/ACM Transactions on Computational Biology and Bioinformatics, 2017, 15(5): 1549-1559.

[14] TEES15. BJÖRNE J, SALAKOSKI T. TEES 2.2: biomedical event extraction for diverse corpora[J]. BMC Bioinformatics, 2015, 16(16): 1-20.

[15] PYYSALO S, OHTA T, MIWA M, et al. Event extraction across multiple levels of biological organization[J]. Bioinformatics, 2012, 28(18): 575-581.

[16] HE X, LI L, LIU Y, et al. A two-stage biomedical event trigger detection method integrating feature selection and word embeddings[J]. IEEE/ACM Transactions on Computational Biology and Bioinformatics, 2017, 15(4): 1325-1332.

[17] RAHUL P V S S, SAHU S K, ANAND A. Biomedical event trigger identification using bidirectional recurrent neural network based models[J]. arXiv:1705.09516, 2017.

[18] WANG Y, WANG J, LIN H, et al. Biomedical event trigger detection based on bidirectional LSTM and CRF[C]//Proceedings of the 2017 IEEE International Conference on Bioinformatics and Biomedicine, 2017: 445-450.

[19] WANG Y, WANG J, LIN H, et al. Bidirectional long short-term memory with CRF for detecting biomedical event trigger in FastText semantic space[J]. BMC Bioinformatics, 2018, 19(20): 59-66.

[20] SHEN C, LIN H, FAN X, et al. Biomedical event trigger detection with convolutional highway neural network and extreme learning machine[J]. Applied Soft Computing, 2019, 84: 105661.

[21] WEI H, ZHOU A, ZHANG Y, et al. Biomedical event trigger extraction based on multi-layer residual BiLSTM and contextualized word representations[J]. International Journal of Machine Learning and Cybernetics, 2022, 13(3): 721-733.

[22] HE X, REN Y, TAI P, et al. A two-stage biomedical event trigger detection method based on hybrid neural network and sentence embeddings[J]. IEEE Access, 2021, 9: 81926-81935.

[23] HE X, LI L, WAN J, et al. Biomedical event trigger detection based on BiLSTM integrating attention mechanism and sentence vector[C]//Proceedings of the 2018 IEEE International Conference on Bioinformatics and Biomedicine, 2018: 651-654.

[24] LI L, HUANG M, LIU Y, et al. Contextual label sensitive gated network for biomedical event trigger extraction[J]. Journal of Biomedical Informatics, 2019, 95: 103221.

[25] HUANG K H, YANG M, PENG N. Biomedical event extraction with hierarchical knowledge graphs[J]. arXiv: 2009.09335, 2020.

[26] DEVLIN J, CHANG M W, LEE K, et al. BERT: pre-training of deep bidirectional transformers for language understanding [J]. arXiv:1810.04805, 2018.

[27] MCCANN B, KESKAR N S, XIONG C, et al. The natural language decathlon: multitask learning as question answering [J]. arXiv:1806.08730, 2018.

[28] WANG X D, WEBER L, LESER U. Biomedical event extraction as multi-turn question answering[C]//Proceedings of the 11th International Workshop on Health Text Mining and Information Analysis, 2020: 88-96.

[29] GU Y, TINN R, CHENG H, et al. Domain-specific language model pretraining for biomedical natural language processing [J]. ACM Transactions on Computing for Healthcare, 2021, 3(1): 1-23.

[30] WADDEN D, WENNBERG U, LUAN Y, et al. Entity, relation, and event extraction with contextualized span representations[J]. arXiv:1909.03546, 2019.

[31] KLEIN D, MANNING C D. Accurate unlexicalized parsing [C]//Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics, 2003: 423-430.

[32] DIAO Y, LIN H, YANG L, et al. FBSN: a hybrid fine-grained neural network for biomedical event trigger identification[J]. Neurocomputing, 2020, 381: 105-112.

[33] HE X, TAI P, LU H, et al. A biomedical event extraction method based on fine-grained and attention mechanism[J]. BMC Bioinformatics, 2022, 23(1): 1-17.