

基于类型语义提示的事件检测方法

丁远远^{1,3}, 张顺香^{1,2,3}, 文 华^{1,3}, 焦熠璇^{1,3}, 张基旭^{1,3}, 曹宇轩^{1,3}

(1. 安徽理工大学计算机科学与工程学院, 淮南 232001; 2. 淮南师范学院计算机学院, 淮南 232038; 3. 合肥综合性国家科学中心人工智能研究院, 合肥 230026)

摘 要: 针对现有研究将事件检测过程分解为触发词识别和分类两个阶段性任务, 从而引发误差传递的问题, 本文提出一种基于类型语义提示的事件检测方法。通过将事件类型作为提示信息来引导模型从事件文本中抽取与事件类型对应的触发词, 从而并行执行触发词的识别和分类, 缓解任务间误差传递的问题。首先利用跨注意力机制处理事件文本表征和事件类型提示模板, 获得融合事件文本信息的提示表征; 然后计算提示表征与事件文本表征间的余弦相似度, 得到与事件类型对应的触发词在事件文本中位置的概率分布; 最后基于位置的概率分布确定触发词的位置, 从而同时实现触发词的识别与分类。在 ACE2005 和 MACCROBAT-EE 数据集上的实验结果表明, 本文方法在事件检测任务中的 F_1 值均有提升。

关键词: 事件检测; 提示学习; 跨注意力机制; 余弦相似度; 概率分布

中图分类号: TP391.1

文献标志码: A

Event Detection Method Based on Type-Semantic Prompts

DING Yuanyuan^{1,3}, ZHANG Shunxiang^{1,2,3}, WEN Hua^{1,3}, JIAO Yixuan^{1,3}, ZHANG Jixu^{1,3}, CAO Yuxuan^{1,3}

(1. School of Computer Science and Engineering, Anhui University of Science and Technology, Huainan 232001, China; 2. School of Computer, Huainan Normal University, Huainan 232038, China; 3. Institute of Artificial Intelligence, Hefei Comprehensive National Science Center, Hefei 230026, China)

Abstract: Addressing the issue of error propagation in existing research that decomposes the event detection process into two staged tasks of trigger recognition and classification, this paper proposes an event detection method based on type-semantic prompts. This method uses event types as prompt information to guide the model in extracting triggers corresponding to the event types from event text. It enables the parallel execution of trigger recognition and classification tasks, thereby mitigating the issue of error propagation between tasks. Firstly, the cross-attention mechanism is utilized to process the representation of the event text and the prompt template consisting of event types, obtaining a fused prompt representation that integrates the event text information. Then, the cosine similarity between the prompt representation and the event context representation is computed to obtain the probability distribution of the trigger positions corresponding to the event types in the event text. Finally, the position

基金项目: 国家自然科学基金(62076006); 认知智能全国重点实验室开放课题(COGOS-2023HE02); 安徽高校协同创新项目(GXXT-2021-008)。

收稿日期: 2024-06-30; **修订日期:** 2024-10-16

of the trigger corresponding to the event type is determined based on the probability distribution of positions, thus achieving parallel execution of trigger recognition and classification tasks. Experimental results on the ACE2005 and MACCROBAT-EE datasets demonstrate an improvement in the F_1 score of the proposed method in event detection tasks.

Key words: event detection; prompt learning; cross-attention mechanism; cosine similarity; probability distribution

引言

事件抽取旨在从非结构化或半结构化文本中抽取结构化的信息,可以为网络舆情分析、知识图谱构建和智能问答等任务提供有力支持。事件抽取包括事件检测(Event detection, ED)和事件论元抽取两个子任务。其中,事件检测的目标是识别出事件文本中的触发词并确定触发词所属的事件类型,即触发词识别和触发词分类^[1]。图1为一段ACE2005数据集中的事件文本^[2],用红框标注的文本“sentence”为事件触发词。事件检测的任务之一是触发词识别,即确定“sentence”在文本中的起始和终止位置,图中第一个“sentence”的识别结果为“起始:21 终止:22”,第二个“sentence”的识别结果为“起始:35 终止:36”。第二个任务是触发词分类,即为触发词分配相应的事件类型,图中“sentence”的事件类型为“Justice-sentence”。

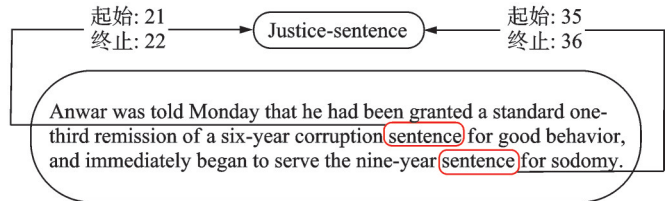


图1 事件检测示例

Fig.1 Example of event detection

当前主流的事件检测模型^[3-4]通常采用两阶段的任务执行策略,即首先识别出给定事件文本中的触发词,然后将被识别出的触发词归类到相应的事件类型中。然而,由于任务之间的串联性质,使得触发词识别阶段的错误会逐渐传递并影响触发词分类任务,从而导致误差累积现象。

针对这一问题,本文提出了一种基于类型语义提示的事件检测方法(Type-semantic prompts, TSP)。TSP将提示学习方法引入到事件检测过程,即使用事件类型作为事件检测过程中的提示信息,使得模型能够根据事件类型提供的语义信息直接从事件文本中提取对应触发词的跨度,从而实现并行执行触发词识别和分类任务的目的,缓解任务之间误差逐渐传递的现象,提高事件检测的准确率。本文基于BART(Bidirectional and auto-regressive transformer)预训练语言模型,首先获取事件上下文的语义表征,然后使用跨注意力机制处理事件上下文表征和事件类型提示模板,获得包含事件上下文语义信息的提示表征。然后计算事件上下文表征与提示表征之间的余弦相似度,得到某一事件类型对应的触发词在事件文本中起始和终止位置的概率分布。最后,通过阈值调优和启发式规则对概率分布进行解码,确定事件类型下触发词起始和终止位置,实现触发词识别及分类任务的并行执行,提高事件检测的准确率。本文的主要贡献包括:

(1) 与主流的事件检测方法不同,本文采用单阶段性处理方法来执行事件检测任务。该方法通过直接解码出事件文本中与事件类型相对应的触发词跨度信息,实现了触发词识别和分类任务的并行处理,缓解了任务之间误差逐渐传递的现象,从而提高事件检测的准确率。

(2) 提出了一种基于事件类型语义提示的事件检测方法。该方法采用了提示学习的策略,将事件类型作为事件检测过程中的提示信息,并使用BART模型对提示信息与文本进行编解码,提升对于事件的语义分析和关系建模,从而改善事件检测结果的准确性和模型的泛化能力。

1 相关工作

1.1 事件检测

近年来,事件检测受到了国内外学者的广泛关注。在早期的研究中,事件检测方法主要围绕特征工程展开。例如,文献[3]利用多级门限注意力机制来提取句子级和文档级事件之间的特征。文献[5]在事件抽取的过程中将预训练字符向量和词性标注向量作为特征输入模型,并利用长短期记忆(Long short-term memory, LSTM)网络获得事件文本的语义增强特征。然而,由于基于特征工程的方法需要依赖人为选择和设计特征,导致此类方法需要消耗大量的成本且相关模型的通用性较差。

随着深度学习的发展,人们开始将基于神经网络的方法应用于事件检测任务中。例如,Qiu等^[6]在文本特征提取模块中,将社交媒体不同类型的数据作为关系图的节点,从而更好地挖掘事件文本中的语义信息并提高事件检测的准确率。文献[7]使用卷积神经网络(Convolutional neural network, CNN)和双向门控循环单元(Bidirectional gated recurrent unit, BiGRU)来提取局部特征,并运用注意力机制提取全局特征。文献[8]在模型中以图的形式融合了实体类型及事件要素信息,实现了多语义信息的融合。相较于基于特征工程的方法,深度学习模型具有更加强大的自动特征学习能力。然而,深度学习模型通常需要大量的标注数据进行训练,以获得良好的性能。如果训练数据有限,可能会导致模型的泛化能力下降。

1.2 预训练语言模型和提示学习

BART是一种基于Transformer架构的预训练语言模型,其采用了双向和自回归的训练方式,这种训练方式使得BART学习到更全面和多样的语言表示。基于BART模型的编解码特点,Lewis等^[9]提出将BART模型应用于文本生成和理解任务当中。此后,越来越多的研究者将BART模型应用于文本生成任务。例如,Lu等^[10]将BART模型应用于问答任务中,为事件抽取任务设计了动态的问题模板。Ma等^[11]将事件论元抽取定义为文本生成任务,利用BART的双向和自回归的训练方式来生成与论元角色对应的事件论元。

预训练语言模型的兴起,使得提示学习在自然语言处理领域引起了广泛的关注。例如,Nguyen等^[12]在事件论元抽取任务中通过构造图来捕获文本中实体之间的关系,再根据图中的信息设计软提示;Dai等^[13]构建了前向迭代提示加后向迭代提示模型,为事件论元抽取提供更多的提示信息;Wang等^[14]在事件检测过程中设计了多步骤提示模块,并在模块当中引入了知识增强本体。相较于基于深度学习的事件检测方法,基于提示学习加预训练语言模型的方法具有更好的上下文感知及泛化能力。Liu等^[15]以预训练语言模型为基础,并将提示信息置于事件文本之前,根据事件类型的不同动态生成不同的前缀提示;针对实体之间的长距离依赖关系,Liu等^[16]提出了采用BART加链路预测的方式来联系各实体。

尽管现有的事件检测方法已取得一定进展,但仍有不足。现有方法通常采用两阶段式的任务处理方式,即先进行事件触发词识别,再对识别出的触发词进行分类,导致出现误差逐渐传递的现象。此外,现有模型过度关注上下文信息的获取,忽视了外部知识的辅助作用,导致模型泛化性有待提升。本文在事件检测任务中,将事件类型作为模型的提示信息,提示模型直接从事件文本中抽取出与事件类型对应的触发词,从而实现触发词识别与分类任务的并行处理,在缓解事件检测模型误差累积现象的同时提升模型的泛化性能。

2 本文方法

本文提出了一种基于类型语义提示的事件检测方法TSP,其整体结构如图2所示。

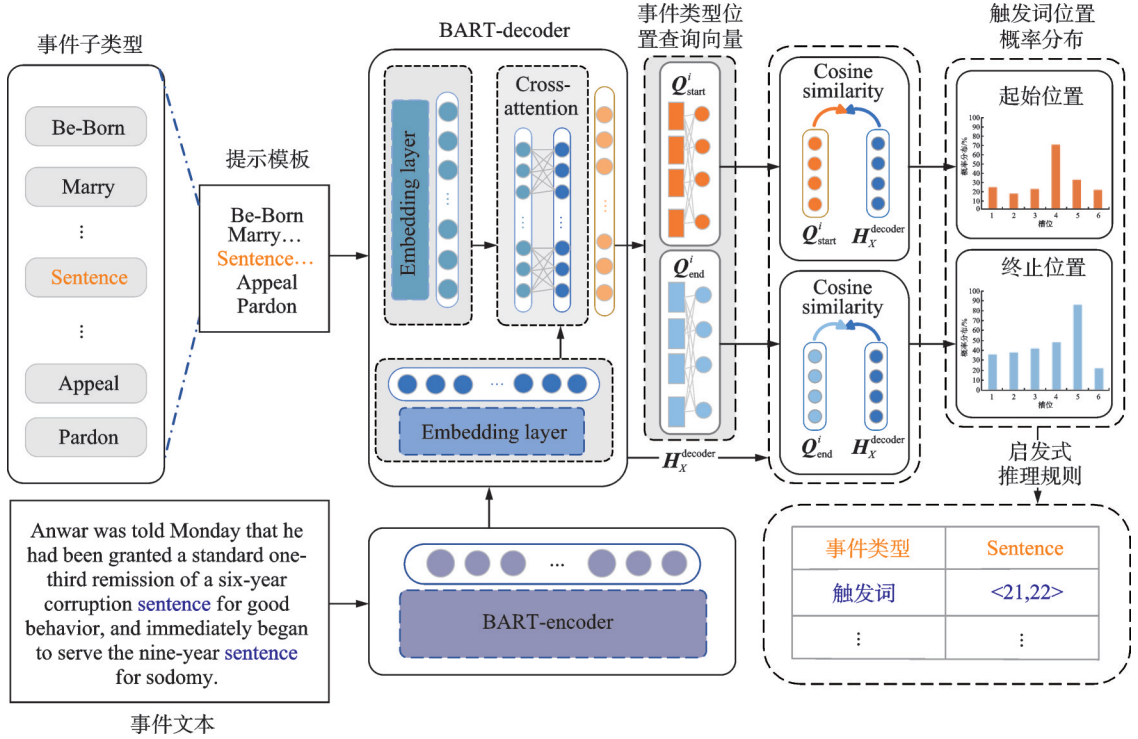


图2 TSP结构

Fig.2 Structure of TSP

2.1 任务定义

TSP方法采用单阶段性处理策略来执行事件检测任务,即使用事件类型作为提示文本,提示模型直接从事件文本中抽取出与该事件类型对应的触发词位置信息,从而同时完成触发词识别及触发词分类任务。

类型语义即事件类型 e 所对应的语义信息。假设数据集中的事件类型集合为 E ,则有 $e \in E$ 。本文使用一个 n 维向量表示每个事件类型,对于事件类型 $e \in E$,可以表示为一个 n 维向量 $E(e) \in \mathbb{R}^n$ 。

2.2 提示生成

本文基于事件类型文本设计了两种类型的提示模板,分别是类型独立性模板和类型相关性模板,并进一步在ACE2005数据集上对这两个模板的性能进行了验证。如表1所示,ACE2005数据集当中共标注了8种事件类型、33种事件子类型。

(1) 类型独立性模板

从任务角度考虑,本文模型的最终任务是确定事件文本中每个事件类型对应的触发词及其位置。为了实现这一目标,采用了较为直接的方法,即从每个事件子类型入手,设计了每个事件子类型相互独立的类型独立性模板 $\tilde{P}$ 。该模板的生成过程为

$$\tilde{P} = \tilde{f}_{\text{prompt}}(e) = \{(\text{Be-Born}), (\text{Marry}), (\text{Divorce}), (\text{Injure}), (\text{Die}), \dots\} \quad (1)$$

式中 $\tilde{f}_{\text{prompt}}$ 表示类型独立性模板的创建函数,“Be-Born”“Marry”“Divorce”“Injure”及“Die”等为数据集当中给出的事件子类型文本。

(2) 类型相关性模板

根据表1的数据,可以观察到每个事件类型下都存在多个与之对应的事件子类型。因此,可以通过将

表 1 ACE2005 数据集中的事件类型及事件子类型
Table 1 Event types and event subtypes in ACE2005 dataset

事件类型	事件子类型
Life	Be-Born, Marry, Divorce, Injure, Die
Movement	Transport
Transaction	Transfer-Ownership, Transfer-Money
Business	Start-Org, Merge-Org, Declare-Bankruptcy, End-Org
Conflict	Attack, Demonstrate
Contact	Meet, Phone-Write
Personnel	Start-Position, End-Position, Nominate, Elect
Justice	Arrest-Jail, Release-Parole, Trial-Hearing, Charge-Indict, Sue, Convict, Sentence, Fine, Execute, Extradite, Acquit, Appeal, Pardon

所有事件子类型拼接的方式来隐式建模事件类型之间的相关性。本文将 33 种事件子类型作为提示文本的基本单元,并采用直接拼接的方式将这些事件子类型组合成一个提示文本 P , 拼接过程为

$$P = f_{\text{prompt}}(e) = \{\text{Be-Born Marry Divorce Injure Die} \dots\} \quad (2)$$

式中: f_{prompt} 表示类型相关性模板的创建函数; “Be-Born” “Marry” “Divorce” “Injure” 及 “Die” 等为数据集当中给出的事件子类型。

通过直接拼接的方式设计提示模板, 可以将每个事件子类型的信息融合在一起, 从而更好地捕获各个事件子类型之间的隐式交互, 形成一个更全面、更综合的提示文本。在该模板中, 每个事件子类型都以其独立的形式存在, 这意味着模型在解读提示模板时也能够将每个子类型的信息视为独立的特征, 从而体现事件子类型之间的独立性。

2.3 编解码

由于 BART 模型中的跨注意力机制能使解码后的表征包含更多的语义信息, 这一特点为事件检测任务提供更多的语义支持, 因此 TSP 使用 BART 作为预训练语言模型。

在 TSP 方法中, 分别将事件文本 X 和提示模板 P 输入到预训练语言模型 L 的编码器 (Encoder) 和解码器 (Decoder) 中进行处理, 其中 $X = \{x_1, x_2, \dots, x_n\} \in D$ 且 x_i 为 X 当中的第 i 个单词。其中集合 D 为 ACE2005 数据集。

在编码阶段, 编码器对事件文本 X 中的每个单词 x_i 及其位置进行编码, 将其转化为对应的向量表示 c_i , 得到事件文本表征 H_X^{encoder} , 这一过程可以表示为

$$H_X^{\text{encoder}} = L_{\text{encoder}}(X) = \{c_1, c_2, \dots, c_n\} \quad (3)$$

然后将 H_X^{encoder} 作为解码器输入, 并获得了语义增强的事件上下文表征 H_X^{decoder} , 该过程为

$$H_X^{\text{decoder}} = L_{\text{decoder}}(H_X^{\text{encoder}}; H_X^{\text{encoder}}) \quad (4)$$

在解码阶段, 解码器接收事件文本表征 H_X^{encoder} 作为一个输入, 同时接收提示模板 P 作为第二个输入。在解码过程中, 对事件上下文表征和提示模板进行跨注意力交互计算, 得到特定事件文本感知的提示模板表征 H_P^{decoder} 。该过程实现对提示模板中不同事件类型和事件文本信息的融合, 最终生成特定事件文本感知的提示模板表征, 从而提高模型对特定事件文本的感知能力和表征学习效果。此过程为

$$H_P^{\text{decoder}} = L_{\text{decoder}}(H_X^{\text{encoder}}; P) = \{p_1, p_2, \dots, p_n\} \quad (5)$$

式中 p_i 为经过解码后的提示模板中第 i 个单词的表征。

由于提示模板中的一个事件子类型可能会被分词器分割成多个单词, 因此需要通过平均池化操作

获得该事件子类型对应的表征。该平均池化过程表示为

$$E_i = \text{MeanPooling}(H_p^{\text{decoder}}[E_{\text{start}}^i; E_{\text{end}}^i + 1]) \quad (6)$$

式中: E_i 为提示模板中第 i 个事件子类型的表征; E_{start}^i 和 E_{end}^i 分别表示第 i 个事件子类型经过分词后得到的起始位置和终止位置。

2.4 基于跨度抽取的训练目标

本文将事件检测任务定义为跨度抽取任务,旨在抽取出与提示模板中事件子类型对应的触发词起始和终止位置,因此分别设定了触发词起始位置和终止位置的训练目标。

针对某一事件子类型的表征 E_i , 本文分别为 E_i 的起始位置和终止位置设定参数矩阵,目的是获得该事件子类型对应的起始位置和终止位置的查询向量 Q_{start}^i 、 Q_{end}^i , 计算过程为

$$Q_{\text{start}}^i = E_i * W_{\text{start}} \quad (7)$$

$$Q_{\text{end}}^i = E_i * W_{\text{end}} \quad (8)$$

式中: W_{start} 和 W_{end} 为可学习的参数矩阵; “*” 表示点积操作。

分别计算查询向量 Q_{start}^i 、 Q_{end}^i 和事件上下文表征 H_X^{decoder} 的余弦相似度,以获得事件子类型对应的事件文本中的触发词位置概率分布 P_{start}^i 和 P_{end}^i , 计算过程为

$$P_{\text{start}}^i = \text{CosineSimilarity}(Q_{\text{start}}^i, H_X^{\text{decoder}}) \quad (9)$$

$$P_{\text{end}}^i = \text{CosineSimilarity}(Q_{\text{end}}^i, H_X^{\text{decoder}}) \quad (10)$$

由于本文方法能够同时抽取出某一事件子类型对应的多个触发词,且数据集中仅有约 24.5% 样本存在多种事件子类型的情况,其余每条样本下出现的事件子类型数量大多只有一种,即存在严重的类别不平衡现象。因此,本文方法采用了基于 softmax 和交叉熵的多标签分类损失函数,避免了多次超参数调优。

触发词起始和终止位置的损失函数分别为 $\text{Loss}_{\text{start}}^i$ 和 $\text{Loss}_{\text{end}}^i$, 参照文献[17]进行推导可得

$$\text{Loss}_{\text{start}}^i = \ln \left(1 + \sum_{j \in \Omega_{\text{neg}}} e^{P_{\text{start}}^j} \right) + \ln \left(1 + \sum_{k \in \Omega_{\text{pos}}} e^{-P_{\text{start}}^k} \right) \quad (11)$$

$$\text{Loss}_{\text{end}}^i = \ln \left(1 + \sum_{j \in \Omega_{\text{neg}}} e^{P_{\text{end}}^j} \right) + \ln \left(1 + \sum_{k \in \Omega_{\text{pos}}} e^{-P_{\text{end}}^k} \right) \quad (12)$$

式中: j 为属于该事件子类型下的除真实触发词起始位置外的所有位置; k 为属于该事件子类型下的真实触发词起始位置; Ω_{pos} 和 Ω_{neg} 分别表示样本的正、负类别集合。

最后将所有类型对应的损失函数进行求和,得到最终的损失函数 $\text{Loss}_{\text{total}}$, 计算过程为

$$\text{Loss}_{\text{total}} = \sum_{i=0}^{32} \text{Loss}_{\text{start}}^i + \sum_{i=0}^{32} \text{Loss}_{\text{end}}^i \quad (13)$$

式中: i 表示从 0 到 32, 涵盖了所有事件子类型。

2.5 触发词跨度推理规则

对于 2.4 节中得到的每种事件子类型对应触发词的起始位置和终止位置概率分布 P_{start}^i 和 P_{end}^i , 本文采用启发式的规则进行推理。

对于每个事件子类型对应的触发词,首先通过 sigmoid 函数将其起始位置概率分布和终止位置概率分布 P_{start}^i 和 P_{end}^i 分别转化为预测值 $\text{Start}_{\text{preds}}$ 和 $\text{End}_{\text{preds}}$, 默认阈值设置为 0.5。计算过程为

$$\text{Start}_{\text{preds}} = \text{sigmoid}(P_{\text{start}}^i) > 0.5 \quad (14)$$

$$\text{End}_{\text{preds}} = \text{sigmoid}(P_{\text{end}}^i) > 0.5 \quad (15)$$

然后,将起始预测值 $\text{Start}_{\text{preds}}$ 为 1 的位置保存在列表中,并从该位置出发遍历 K 个跨度内的终止预测值 $\text{End}_{\text{preds}}$ 为 1 的位置,并结合起始预测值保留进列表内,其中 K 的选择是根据数据集中所有触发词

的最大跨度来设计的。通过这样的推理方式,避免了重复枚举起始位置和终止位置,从而提高了模型推理的效率。

3 实验与结果分析

3.1 实验数据集及评价指标

本文选用了在事件抽取领域应用广泛的 ACE2005 数据集^[2]和医学领域的 MACCROBAT-EE 数据集^[18]进行实验。ACE2005 数据集由语言数据联盟发布,涵盖了 2003~2005 年期间新闻、博客和广播等多个领域的事件文本,包括实体、关系和事件注释。该数据集包含 599 篇不同领域的文档,涵盖 33 种事件类型。MACCROBAT-EE 是一个临床事件抽取数据集,包含来自 PubMed 公司的 200 个临床病例文档,共定义 13 种数据类型。ACE2005 和 MACCROBAT-EE 数据集的具体划分比例如表 2 所示。

本文选取了与 Du 等^[19]工作中相同的评价指标:如果事件触发词的偏移量与标准触发词的偏移量相匹配,则该触发词被正确识别;如果事件类型也与标准触发词的事件类型匹配,则该触发词被正确分类。触发词识别和分类的结果使用精确率 p 、召回率 R 和综合评估指标 F_1 进行评估。

表 2 数据集划分比例

Table 2 Dataset partition ratio

数据集	训练集	验证集	测试集
ACE2005	529	30	40
MACCROBAT-EE	160	20	20

3.2 模型参数设置及实验环境

本文选择了 BART 作为预训练语言模型,并使用了两种不同规模的模型。其中,BART-base 模型采用 6 层的 Encoder 和 Decoder 结构,而 BART-large 模型则采用 12 层的 Encoder 和 Decoder 结构。在实验中,设置了批次大小为 16 的训练批次,并使用学习率为 $2e-5$ 的优化器进行训练。此外,还设置了权重衰减为 0.01,并使用 AdamW 优化器进行参数更新。整个训练过程共进行了 1 000 个步骤。本文的实验环境如表 3 所示。

表 3 实验环境

Table 3 Experimental environment

类型	详细配置
操作系统	Windows 11(64 位)
CPU	intel(R)Xeon(R)Gald 5218
GPU	RTX 4090(24 GB)
Pytorch 版本	2.1.1+CU121
CUDA 版本	12.0

3.3 对比试验

为了评估本文模型的性能,选取了在事件检测领域较先进的模型作为对比模型进行实验:

- (1) DMCNN^[20]:提出了动态多池化卷积神经网络,使用动态多池层来保存关键的信息。
- (2) JRNN^[21]:使用了双向的循环神经网络(Recurrent neural network, RNN),提取更多的语义信息。
- (3) JMEE^[22]:使用自注意力机制构建图神经网络,提升事件检测的准确率。
- (4) OneIE^[23]:设计了联合神经网络模型,结合文本的全局信息,通过构造全局最优图来进行信息抽取。
- (5) EEQA^[19]:首次将事件抽取问题表述为问答/阅读理解问题,以端到端的方式进行事件抽取。
- (6) DEGREE(ED)^[24]:提出了一种基于条件生成的少样本端到端的事件提取模型,通过融合事件文本和提示文本当中的信息进行事件抽取。
- (7) IFPT^[25]:一种融合实体信息、实体类型信息作为提示的模型。
- (8) PromptLoc^[26]:提出了触发词定位公式,并且通过对比学习的方式来区分上下文中真实的触发词。

TSP 方法与对比模型均在 ACE2005 数据集上进行实验,TSP 使用的提示模板为类型相关性模板。为

了体现实验的公正性,本文分别基于BART-base和BART-large预训练语言模型(Pre-trained language model, PLM)设计了TSP模型。实验结果如表4所示。

表 4 ACE2005数据集上的对比实验结果

Table 4 Comparative experimental results on ACE2005 dataset

%

模型	PLM	触发词识别			触发词分类		
		p	R	F_1	p	R	F_1
DMCNN	—	80.4	67.7	73.5	75.6	63.6	69.1
JRNN	—	68.5	75.7	71.9	66.0	73.0	69.3
JMEE	—	80.2	72.1	75.9	76.3	71.3	73.7
OneIE	BERT	—	—	78.2	74.9	74.5	74.7
EEQA	BERT	74.3	77.4	75.8	71.1	73.7	72.4
DEGREE(ED)	BART-base	—	—	—	—	—	73.3
IFPT	BART-base	—	—	75.1	—	—	71.9
PromptLoc	BERT	—	—	—	73.6	74.2	73.9
TSP(Ours)	BART-base	80.6	81.4	81.0	75.2	75.9	75.6
TSP(Ours)	BART-large	83.0	81.1	82.0	78.7	76.9	77.8

由表4中的实验数据得出以下结论:

(1) TSP方法在触发词识别和分类任务上的准确率均有显著提升,表明该方法在事件检测任务中的有效性。此外,与基于BART-base的实验结果相比,TSP模型在基于BART-large的实验结果中表现更为出色。

(2) 与基于深度学习的DMCNN、JRNN和JMEE模型相比,TSP方法在事件检测任务中表现出最优的性能。这是因为相比于DMCNN、JRNN和JMEE模型,TSP方法在获取事件上下文语义信息的同时,还利用事件类型的语义信息作为模型的提示信息。因此,该方法的整体性能优于其他3种方法。

(3) 与对比模型中性能最优的OneIE模型相比,TSP方法的整体性能有显著提升。提升的主要原因是OneIE模型采用两阶段处理策略执行事件检测任务,导致触发词识别阶段的误差传递并影响触发词分类阶段。而TSP方法的单阶段处理策略避免了这种误差累积现象,从而提高了事件检测的准确性。

(4) 与基于提示学习的EEQA、DEGREE(ED)、IFPT和PromptLoc四种模型相比,TSP方法的性能均有提升。这是由于TSP模型通过交叉注意力机制实现事件文本和提示文本之间的信息交互,使得TSP模型能够挖掘更多的语义信息,从而为事件检测任务提供更加全面的信息,进而实现更好的效果。

此外,基于BART-base和BART-large的TSP模型处理每条样本的平均时间分别为0.014 s和0.052 s,这表明TSP模型在处理样本时间上具备优异的性能,能够在短时间内高效地完成大量样本的处理任务。

3.4 消融实验

为了研究TSP方法各个组成部分的作用,本节在ACE2005数据集上进行消融实验,包括以下几个实验设置:

(1) w/o prompt表示删除类型相关性提示模板,该过程首先使用BART编码器编码事件文本,再对事件类型进行独热编码,利用两个线性层将事件文本表征映射为事件类型起始位置和结尾位置的概率分布,最后计算损失函数并进行推理。

- (2) w/o prompt(With scattered templates)表示将 TSP 中的类型相关性提示模板替换为类型独立性模板。
- (3) w/o cross-attention(With relevance template)表示删除了提示模板和事件上下文的跨注意力交互,即将提示表征与文本表征进行余弦相似度计算。此时提示模板为类型相关性模板。
- (4) w/o cross-attention(With scattered templates)表示删除了提示模板和事件上下文的跨注意力交互,即将提示表征与文本表征进行余弦相似度计算。此时提示模板为类型独立性模板。
- (5) w/o multi-label loss 表示将多标签损失函数替换为二元交叉熵损失函数。
- (6) TSP 表示本文所提的完整模型,使用的提示模板为类型相关性模板。
- 对于以上 5 个实验设置,分别在相同的实验环境下进行实验,实验结果如表 5 所示。

表 5 ACE2005数据集消融实验结果

Table 5 Results of ablation experiments on ACE2005 dataset

%

模型	PLM	触发词识别			触发词分类		
		p	R	F_1	p	R	F_1
w/o prompt	BART-base	77.3	80.9	79.0	72.0	75.4	73.7
	BART-large	80.3	83.1	81.7	75.5	78.2	76.8
w/o prompt(With scattered templates)	BART-base	78.9	76.9	77.9	74.6	72.7	73.6
	BART-large	76.1	74.9	75.5	72.8	71.7	72.2
w/o cross-attention(With relevance template)	BART-base	82.4	74.4	78.2	78.6	71.0	74.6
	BART-large	79.8	80.1	80.0	76.0	76.4	76.2
w/o cross-attention(With scattered templates)	BART-base	76.1	80.6	78.3	71.4	75.7	73.5
	BART-large	76.1	79.7	77.8	72.7	76.2	74.4
w/o multi-label loss	BART-base	79.8	79.4	79.6	74.1	74.1	74.6
	BART-large	80.5	83.9	82.1	75.5	78.7	77.0
TSP(Ours)	BART-base	80.6	81.4	81.0	75.2	75.9	75.6
	BART-large	83.0	81.1	82.0	78.7	76.9	77.8

由表 5 中的实验结果得出以下结论:

- (1) 类型相关性提示模板对结果的影响较为显著。w/o prompt 和 w/o prompt(With scattered templates)模型的准确率均有明显下降,体现了类型相关性模板的积极作用。该模板在保持每个事件类型独立性的同时,实现事件类型间的语义交互,这种交互有助于模型更准确地理解和识别触发词,进而提高事件检测的准确率。
- (2) 交叉注意力层对结果也存在显著影响。w/o cross-attention(With relevance template)模型的 F_1 值有所下降,这是由于缺少事件文本表征与提示模板表征间的跨注意力交互,导致模型的信息交互能力下降。此外,w/o cross-attention(With scattered templates)模型的准确率下降幅度更为明显,这是因为将事件子类型分散后,各个事件子类型未进行信息交互,导致获得的语义信息减少,从而降低了模型的性能。
- (3) 多标签损失函数对结果也存在的影响。基于 BART-base 的 w/o multi-label loss 模型的 F_1 值降低,这是由于本文将事件检测定义为多标签分类任务,使用多标签损失函数能捕捉标签间的相关性和样本的多样性,提高了模型性能。而基于 BART-large 的模型在触发词识别任务上的 F_1 值提升 0.1%,这是因为 BART-large 的模型规模和参数量较大,具备更强的泛化能力和拟合能力。

3.5 模型泛化性验证实验

为了验证 TSP 模型的泛化能力,本文在医学领域的 MACCROBAT-EE 数据集上进行了对比及消融实验,实验中所使用的提示模板为 MACCROBAT-EE 数据集中定义的 13 种数据类型的拼接文本。

3.5.1 对比试验

实验引入 Lu 等^[27]提出的序列到结构的生成模型 Text2Event,以端到端的方式进行事件抽取;Ma 等^[18]提出的基于生成模型的 Vanilla DICE 方法,通过提及增强技术进行提及边界识别。对比实验结果如表 6 所示。

表 6 MACCROBAT-EE 数据集上的对比实验结果
Table 6 Comparative experimental results on MACCROBAT-EE dataset %

模型	PLM	触发词识别			触发词分类		
		p	R	F_1	p	R	F_1
Text2Event	T5-large	—	—	—	66.6	60.5	63.4
DEGREE(ED)	BART-large	71.9	66.3	69.0	67.5	62.5	65.0
Vanilla DICE	T5-large	65.0	74.0	69.2	60.5	70.2	65.0
TSP(Ours)	BART-large	67.7	77.9	72.4	63.8	73.5	68.3

由表 6 中的数据可知:TSP 方法在 MACCROBAT-EE 数据集上的实验结果明显优于对比模型,呈现出较高的事件检测准确率。相较于在对比模型中表现最优的 Vanilla DICE 模型,TSP 方法在触发词识别任务上的 F_1 值提升了 3.2%,在触发词分类任务上的 F_1 值提升了 3.3%,表明 TSP 方法具有出色的泛化能力。

3.5.2 消融试验

本节在 MACCROBAT-EE 数据集上进行消融实验,消融模型的配置与 3.4 节相同。实验结果如表 7 所示。

表 7 MACCROBAT-EE 数据集上的消融实验结果
Table 7 Results of ablation experiments on MACCROBAT-EE dataset %

模型	PLM	触发词识别			触发词分类		
		p	R	F_1	p	R	F_1
w/o prompt(With scattered templates)	BART-large	62.2	81.3	70.5	58.2	76.0	65.9
w/o cross-attention(With relevance template)	BART-large	65.0	77.7	70.8	61.1	73.0	66.5
w/o multi-label loss	BART-large	64.0	80.0	71.1	60.2	75.2	67.2
TSP(Ours)	BART-large	67.7	77.9	72.4	63.8	73.5	68.3

由表 7 中的数据可知:在针对 TSP 方法进行的消融实验中,当移除事件相关性提示模板、交叉注意力机制以及多标签损失函数这 3 个核心组件后,模型的性能均出现了不同程度的下降。这一实验结果有力地验证了 TSP 方法中各个模块对于提升事件检测任务准确率的有效性。

3.6 少样本下泛化性验证实验

为了探究 TSP 在少样本情况下的泛化能力,本文继续在 ACE2005 数据集上进行了少样本实验,实验使用 BART-base 模型,并分别采用 1%、2%、5%、10%、20%、50%、80% 和 100% 不同比例的训练数据,实验结果如图 3 所示。

由图3的实验结果可以得出以下结论:随着训练数据比例的增加,TSP模型在触发词识别和分类任务上的性能逐渐提升。在最小的训练数据比例(1%)下,模型的性能较低;当训练数据比例增至10%时,模型的性能实现了显著的提升;当训练数据进一步增至20%时,模型性能同样得到了可观的提升。这一系列的结果不仅说明了TSP模型在少样本学习中的稳健性,也进一步证实了其在少样本情况下的良好泛化能力。

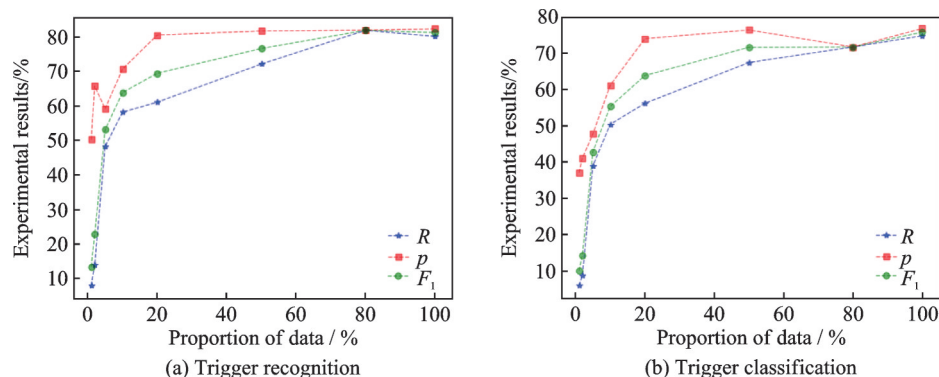


图3 ACE2005数据集上不同样本比例的实验结果

Fig.3 Experimental results with different sample proportions on ACE2005 dataset

4 结束语

针对当前主流事件检测方法存在任务间误差逐渐传递的问题,本文提出了一种基于类型语义提示的事件检测方法TSP。该方法在事件检测过程中引入交叉注意力机制,实现了事件文本与事件类型信息的深度融合,促使模型综合考虑了事件上下文的语义信息和事件类型模板的提示信息,进而增强模型对事件类别判断的准确性与可靠性。此外,TSP能够并行完成触发词识别和分类任务,有效缓解了任务间误差逐渐传递的问题,提高了事件检测的准确率。在ACE2005和MACCROBAT-EE数据集上的实验结果表明,本文所提出的TSP方法在事件触发词识别和触发词分类任务上的准确率均得到了显著的提升,并且在少样本情况下模型也表现出了优异的性能,展现出了良好的泛化能力。本文的局限在于其研究的是特定领域的事件检测工作,未来的研究将深入探讨开放领域的事件检测方法,以拓展其适用范围至更广泛的事件类型和应用场景,从而进一步增强模型的泛化能力。

参考文献:

- [1] 张紫月,王羽,徐建.基于图卷积网络融合依存信息的事件检测方法[J].计算机应用研究,2023,40(10):2967-2971,3006.
ZHANG Ziyue, WANG Yu, XU Jian. Event detection based on dependency information and graph convolutional network[J]. Application Research of Computers, 2023, 40(10): 2967-2971, 3006.
- [2] DODDINGTON G R, MITCHELL A, PRZYBOCKI M A, et al. The automatic content extraction (ACE) program-tasks, data, and evaluation[C]//Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC'04). Lisbon, Portugal: European Language Resources Association (ELRA), 2004, 2(1): 837-840.
- [3] 潘丽敏,李筱雅,罗森林,等.融合多级语义特征的双通道GAN事件检测方法[J].北京理工大学学报,2021,41(3): 295-302.
PAN Limin, LI Xiaoya, LUO Senlin, et al. Double-channel GAN with multi-level semantic correlation for event detection[J]. Transactions of Beijing Institute of Technology, 2021, 41(3): 295-302.
- [4] 王红,吴浩正.基于句法和全文信息增强的中文事件检测方法[J].数据采集与处理,2022,37(5):1059-1069.
WANG Hong, WU Haozheng. Chinese event detection with syntax and full text information enhancement[J]. Journal of Data

- Acquisition and Processing, 2022, 37(5): 1059-1069.
- [5] 刘伟, 马亚威, 彭艳, 等. 基于图注意力和表指针网络的中文事件抽取方法[J]. 模式识别与人工智能, 2023, 36(5): 459-470.
LIU Wei, MA Yawei, PENG Yan, 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.
 - [6] QIU Z, MA C, WU J, et al. An efficient automatic meta-path selection for social event detection via hyperbolic space[C]// Proceedings of the ACM on Web Conference 2024. [S.l.]: ACM, 2024: 2519-2529.
 - [7] 王侃, 王孟洋, 刘鑫, 等. 融合自注意力机制与CNN-BiGRU的事件检测[J]. 西安电子科技大学学报, 2022, 49(5): 181-188.
WANG Kan, WANG Mengyang, LIU Xin, et al. Event detection by combining self-attention and CNN-BiGRU[J]. Journal of Xidian University, 2022, 49(5): 181-188.
 - [8] 魏建香, 陆谦, 韩普, 等. 基于多语义信息融合的事件检测模型[J]. 数据分析与知识发现, 2023, 7(12): 64-74.
WEI Jianxiang, LU Qian, HAN Pu, et al. Event detection model based on semantic information fusion[J]. Data Analysis and Knowledge Discovery, 2023, 7(12): 64-74.
 - [9] LEWIS M, LIU Y, GOYAL N, et al. BART: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension[C]//Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. [S.l.]: Association for Computational Linguistics, 2020: 7871-7880.
 - [10] LU D, RAN S, TETREAULT J, et al. Event extraction as question generation and answering[C]//Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics. Toronto, Canada: Association for Computational Linguistics, 2023: 1666-1688.
 - [11] MA Y, WANG Z, CAO Y, et al. Prompt for extraction? PAIE: Prompting argument interaction for event argument extraction [C]//Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics. Dublin, Ireland: Association for Computational Linguistics, 2022: 6759-6774.
 - [12] NGUYEN C, MÃN H, NGUYEN T. Contextualized soft prompts for extraction of event arguments[C]//Proceedings of Findings of the Association for Computational Linguistics: ACL 2023. Toronto, Canada: Association for Computational Linguistics, 2023: 4352-4361.
 - [13] DAI L, WANG B, XIANG W, et al. Bi-directional iterative prompt-tuning for event argument extraction[C]//Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing. Abu Dhabi, United Arab Emirates: Association for Computational Linguistics, 2022: 6251-6263.
 - [14] WANG S, ZHENG J, CAI F, et al. MsPrompt: Multi-step prompt learning for debiasing few-shot event detection[J]. Information Processing & Management, 2023, 60(6): 103509.
 - [15] LIU X, HUANG H, SHI G, et al. Dynamic prefix-tuning for generative template-based event extraction[C]//Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics. Dublin, Ireland: Association for Computational Linguistics, 2022: 5216-5228.
 - [16] LIU J, LIANG C, XU J, et al. Document-level event argument extraction with a chain reasoning paradigm[C]//Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics. Toronto, Canada: Association for Computational Linguistics, 2023: 9570-9583.
 - [17] SU J, ZHU M, MURTADHA A, et al. ZLPR: A novel loss for multi-label classification[EB/OL]. (2022-08-05). <https://doi.org/10.48550/arXiv.2208.02955>.
 - [18] MA M D, TAYLOR A K, WANG W, et al. DICE: data-efficient clinical event extraction with generative models[C]// Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics. Toronto, Canada: Association for Computational Linguistics, 2023: 15898-15917.
 - [19] DU X, CARDIE C. Event extraction by answering (almost) natural questions[C]//Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP). [S.l.]: Association for Computational Linguistics, 2022: 671-683.
 - [20] 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. Beijing, China: Association for Computational Linguistics, 2015: 167-176.

- [21] 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. San Diego, California, USA: Association for Computational Linguistics, 2016: 300-309.
- [22] 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. Brussels, Belgium: Association for Computational Linguistics, 2018: 1247-1256.
- [23] LIN Y, JI H, HUANG F, et al. A joint neural model for information extraction with global features[C]//Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. [S.l.]: Association for Computational Linguistics, 2020: 7999-8009.
- [24] HSU I, HUANG K H, BOSCHEE E, et al. DEGREE: A data-efficient generation-based event extraction model[C]//Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Seattle, United States: Association for Computational Linguistics, 2018: 1890-1908.
- [25] 苏杭, 胡亚豪, 潘志松. 利用提示调优融合多种信息的低资源事件抽取方法[J]. 计算机应用研究, 2024, 41(2): 381-387, 400.
SU Hang, HU Yahao, PAN Zhisong. Low-resource event extraction method using multi-information fusion with prompt tuning [J]. Application Research of Computers, 2024, 41(2): 381-387, 400.
- [26] LIU J, SUI D, LIU K, et al. Learning with partial annotations for event detection[C]//Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics. Toronto, Canada: Association for Computational Linguistics, 2023: 508-523.
- [27] LU Y, LIN H, XU J, et al. Text2Event: Controllable sequence-to-structure generation for end-to-end event extraction[C]//Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing. [S.l.]: Association for Computational Linguistics, 2021: 2795-2806.

作者简介:



丁远远(2000-),女,硕士研究生,研究方向:Web挖掘、事件检测。



张顺香(1970-),通信作者,男,教授,博士生导师,研究方向:Web挖掘、语义搜索和复杂网络,E-mail:sx-zhang@aust.edu.cn。



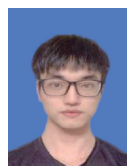
文华(2002-),男,硕士研究生,研究方向:Web挖掘、情感分析。



焦熠璇(2000-),女,硕士研究生,研究方向:自然语言处理、文本分类。



张基旭(1999-),男,硕士研究生,研究方向:目标检测、图像分割和6D姿态估计。



曹宇轩(2000-),男,硕士研究生,研究方向:目标检测、多模态情感分析。

(编辑:刘彦东)