



计算机工程
Computer Engineering
ISSN 1000-3428, CN 31-1289/TP

《计算机工程》网络首发论文

题目：融合词性语义扩展信息的事件检测方法
作者：严海宁，余正涛，黄于欣，宋燃，杨溪
DOI：10.19678/j.issn.1000-3428.0066880
网络首发日期：2023-04-26
引用格式：严海宁，余正涛，黄于欣，宋燃，杨溪. 融合词性语义扩展信息的事件检测方法[J/OL]. 计算机工程. <https://doi.org/10.19678/j.issn.1000-3428.0066880>



网络首发：在编辑部工作流程中，稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定，且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式（包括网络呈现版式）排版后的稿件，可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定；学术研究成果具有创新性、科学性和先进性，符合编辑部对刊文的录用要求，不存在学术不端行为及其他侵权行为；稿件内容应基本符合国家有关书刊编辑、出版的技术标准，正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性，录用定稿一经发布，不得修改论文题目、作者、机构名称和学术内容，只可基于编辑规范进行少量文字的修改。

出版确认：纸质期刊编辑部通过与《中国学术期刊（光盘版）》电子杂志社有限公司签约，在《中国学术期刊（网络版）》出版传播平台上创办与纸质期刊内容一致的网络版，以单篇或整期出版形式，在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊（网络版）》是国家新闻出版广电总局批准的网络连续型出版物（ISSN 2096-4188，CN 11-6037/Z），所以签约期刊的网络版上网络首发论文视为正式出版。



融合词性语义扩展信息的事件检测方法

严海宁^{1,2}, 余正涛^{1,2}, 黄于欣^{1,2}, 宋燃^{1,2}, 杨溪^{1,2}

(1. 昆明理工大学, 信息工程与自动化学院, 昆明 云南 650504;

2. 昆明理工大学, 云南省人工智能重点实验室, 昆明 云南 650504)

摘要: 事件检测是事件抽取中的关键步骤, 其依赖于触发词进行事件类型分类。目前主流的事件检测方法在稀疏标记数据上性能较差, 模型过度拟合密集标注的触发词, 而在稀疏标记的触发词或者未见过的触发词上失效。现有方法大多通过扩充更多训练实例来缓解这一问题。但扩充后的数据分布不均, 存在内置偏差, 仍然表现不佳。其原因在于稀疏标注数据的触发词难以识别, 导致其事件类型不能被正确分类。为此, 对词粒度扩展信息进行探索, 在不增加训练实例的条件下, 缩小候选触发词的范围, 并对候选触发词进行语义扩展, 挖掘候选触发词的上下文中蕴含的丰富语义, 以缓解稀疏标记数据的问题。首先, 通过词性筛选模块寻找候选触发词并对其进行语义扩展, 挖掘词粒度语义信息。其次, 融合句子粒度语义信息提升语义表征的鲁棒性, 最终通过 Softmax 分类器进行分类, 从而完成事件检测任务。在 ACE2005 和 KBP2015 数据集上的实验结果表明, 该模型在事件检测任务中的 F1 分数分别达到了 79.5% 和 67.5%, 有效提升了事件检测的性能。并且在稀疏标记数据实验中的 F1 分数达到了 78.5%, 明显改善了稀疏标记数据的问题。

关键词: 事件检测; 稀疏标记; 词性筛选; 语义扩展; 语义融合; 动态池化

开放科学(资源服务)标志码(OSID):



Fusing the Extended Semantic Information of Part-of-Speech for Event Detection

YAN Haining^{1,2}, YU Zhengtao^{1,2}, HUANG Yuxin^{1,2}, SONG Ran^{1,2}, YANG Xi^{1,2}

(1. Faculty of Information Engineering and Automation, Kunming University of Science and Technology,

Kunming, Yunnan 650504, China;

2. Yunnan Key Laboratory of Artificial Intelligence, Kunming University of Science and Technology,

Kunming, Yunnan 650504, China)

[Abstract] Event detection is one of the key steps in event extraction, which depends on the identified triggers for event type classification. Current mainstream event detection methods have poor performance on sparsely labeled data, which overfit the model with densely labeled triggers and fail on the sparsely labeled trigger or unseen triggers. Most previous methods mitigate this problem by adding more training examples. But the expanded data are distributed unevenly, have built-in biases, and still perform poorly. The reason is that triggers of sparsely labeled data are difficult to identify, resulting in their event types cannot be classified correctly. To this end, this paper explores word granularity expansion information to mitigate the impact of the problem of sparsely labeled data by reducing the range of candidate triggers, and mining the rich semantic information in the contexts without increasing the number of training instances. Firstly, a part-of-speech selection module is applied to find candidate triggers and extend their semantics, which digs out word granularity semantic information. And then, sentence granularity semantic information is incorporated to improve the robustness of semantic information. Lastly, event types classification is performed by Softmax function, which completes the event detection task. Experimental results on ACE2005 and KBP2015 datasets show that the model achieves F1 scores of 79.5% and 67.5% in the event detection task respectively, which improves the performance of event detection effectively. And the F1 score reaches 78.5% in the sparsely labeled data experiments, which alleviate the sparsely labeled data problem significantly.

[Key words] Event detection; sparsely labeled; part-of-speech selection; extended semantic information; semantic fusion; dynamic pooling

基金项目: 国家自然科学基金(U21B2027, 61972186, 62266028); 云南省重大科技专项计划(202202AD080003);

作者简介: 严海宁(1997—), 男, 硕士研究生, CCF 学生会员(会员号: G4204G) 主要研究方向: 自然语言处理、事件检测; 余正涛(通信作者), 教授, 博士, 博士生导师; 黄于欣, 副教授, 博士; 宋燃, 博士研究生; 杨溪, 硕士研究生。

E-mail: ztyu@hotmail.com

0 概述

事件检测 (Event Detection, ED) 是事件抽取的关键步骤之一, 其目标是在给定句子中识别事件触发词并将其分类为预定义的事件类型。其中触发词是最能清楚表达事件核心含义的词, 通常是一个名词或者动词^[1]。事件检测作为自然语言处理领域中的一项重要任务, 被广泛应用于知识图谱^[2]、自动内容抽取^[3]等领域。

事件检测任务依赖于识别出的触发词进行事件类型分类^[4-5]。现有的大数据中触发词的标注严重不平衡, 使模型过度拟合密集标注的触发词数据, 而稀疏标记的触发词数据往往得不到有效训练, 导致触发词为稀疏标记或未出现过的词时模型性能不佳。如下述例子 S1、S2 所示, 其中有下列划线的词语是触发词。

S1: EU will release 20 million euros in emergency humanitarian aid to Iraq.

S2: EU will disburse 20 million euros in emergency humanitarian aid to Iraq.

句子 S1 中, 密集标注触发词数据 “release” 被模型正确识别后, 句子被分类为正确的事件类型 “Transaction||Transfer|Money”。而将句子 S2 中密集标注触发词 “release” 替换为稀疏标记的同义触发词 “disburse” 后, 不改变句子原本表达的含义, 模型却不能识别出正确的事件类型。

针对以上问题, 现有的研究认为生成更多的训练实例是一种解决方案。一些方法通过引导扩展出更多的训练实例^[6-8], 另一些方法使用更多数据进行远程监督^[9-11]。但是这些方法要么生成同质的语料库, 要么受制于低覆盖率的知识库, 导致数据本身分布不均, 训练的模型会存在偏置^[12], 且在稀疏标记数据上性能仍然较差。本文认为其原因可能在于稀疏标注数据的触发词难以识别, 导致其事件类型不能被正确分类。为此, 本文对词粒度扩展信息进行探索, 在不增加训练实例的条件下, 缩小候选触发词的范围, 并对候选触发词进行语义扩展, 从而提升触发词识别能力。同时, 融合不同粒度的语义信息, 增强语义表征的鲁棒性, 以缓解稀疏标记数据的问题。

本文提出一种融合词性语义扩展信息的事件检测方法 (Fusing the Extended Semantic Information of

Part-of-Speech for Event Detection, FESPOS-ED)。

通过词性筛选模块寻找特定词性的词, 以确定候选触发词的位置。无论是稀疏标记还是未见过的触发词, 都会以极大概率包含在特定词性中, 因此能更好地识别候选触发词且不受触发词样本数量的限制。并对候选触发词位置进行语义扩展, 挖掘候选触发词上下文中丰富的词粒度语义信息。最终, 融合句子粒度语义信息, 增强语义表征的鲁棒性, 进一步提升事件检测准确性。本文的主要贡献如下:

(1) 本文提出一种融合词性语义扩展信息的事件检测方法, 结合词粒度语义扩展信息及句子粒度语义信息, 提升语义表征的鲁棒性, 从而缓解稀疏标记数据的问题;

(2) 利用词性语义扩展方法, 在缩小候选触发词范围的同时扩展候选触发词在当前语境下的语义信息, 能更好地识别候选触发词且不受触发词样本数量的限制;

(3) 本文方法在 ACE2005 和 KBP2015 数据集上的实验结果相较于基线模型, 在 F1 值上分别提升了 1.2% 和 5.8%, 在稀疏标记数据实验中相较于基线模型 F1 值提升了 3.4%。

1 相关工作

在早期的研究中, 事件检测任务被视作基于触发词的分类问题, 其重点在于收集全局统计特征作为知识来源或决策基础^[13-14]。随着深度学习的发展, 许多研究利用深度神经网络来学习输入序列的上下文表征信息。在将文本上下文信息嵌入到低维空间后, 再利用这些特征识别触发词和事件类型。根据上下文信息不同, 可以分为结构化上下文信息和非结构化上下文信息。引入结构化上下文信息的方法通常引入事件参数信息^[15-16], 利用结构化文本描述事件信息。引入非结构化上下文信息的方法中通常使用卷积神经网络 (CNN)^[17-18]、图卷积神经网络 (GCN)^[19-21]、预训练语言模型^[10,22-23]等方法捕捉事件信息。

在针对标注不平衡数据问题的研究中, 通常有两类平衡思路: 一类是从数据方面进行平衡, 另一类是从分类损失上进行平衡。在数据方面的平衡策略, 研究者们利用半监督或弱监督的方法自动扩充训练集。文献[6]依靠复杂的预定义规则从并行的新闻流中引导产生更多训练实例。文献[7]采用 WordNet 和基于规则的方法来生成没有事件类型标签的开放域数据。文献[10]利用远程监督

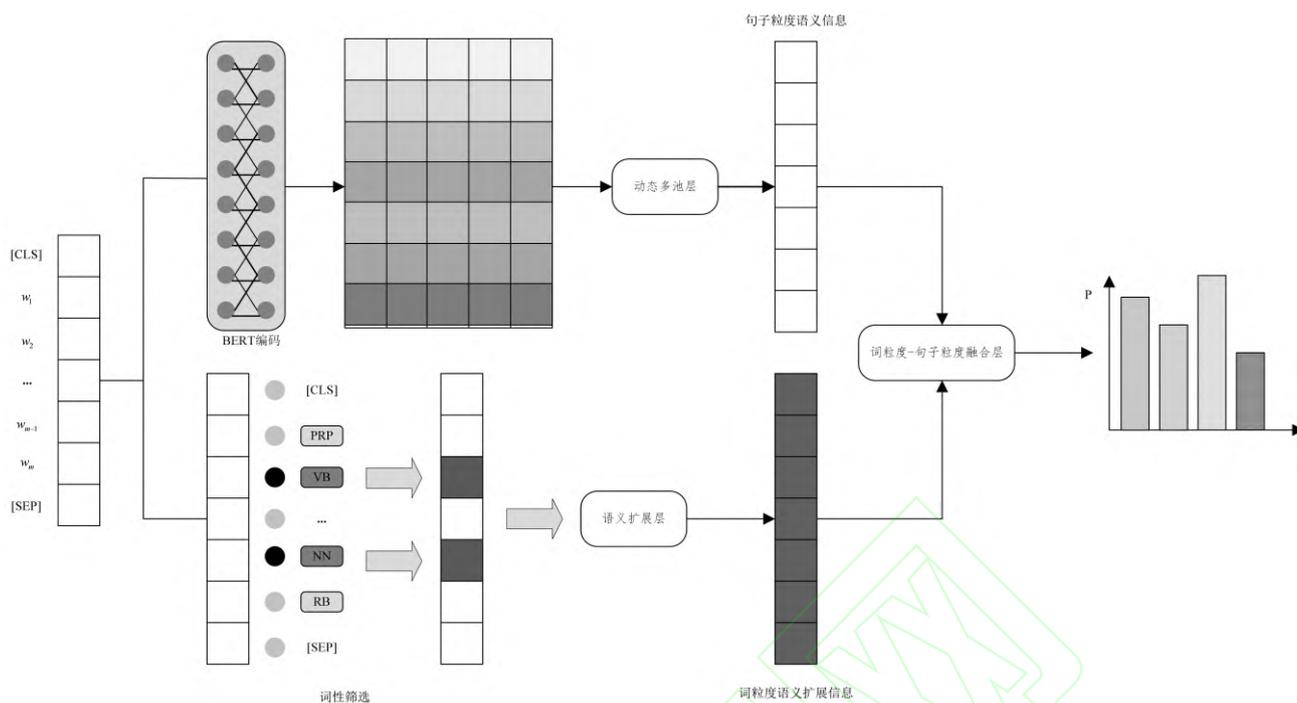


图 1 融合词性语义扩展信息的事件检测框架图

Fig.1 The structure of Fusing the Extended Semantic Information of Part-of-Speech for Event Detection

方法，基于触发器的潜在实例发现策略和对抗性训练方法来协同获得更加多样化和准确的训练数据。文献[11]利用远程监督从知识库中现有的结构化事件知识中生成大规模数据。在分类损失方面的平衡策略，文献[21]提出了一个带有解耦分类重新平衡机制的语法增强型 GCN 框架，根据样本数量模拟负幂律分布重新调整分类器权重。文献[24]提出的损失通过鼓励少数类别拥有更大的边际来扩展现有的软边际损失。文献[25]利用贝叶斯不确定性估计计算样本和类别的不确定性来度量最大边距，以缓解类别不平衡造成的影响。

2 模型

本文针对标记数据不平衡，导致稀疏标记数据不能得到有效训练的问题，提出一种融合词性语义扩展信息的事件检测方法，其结构如图 1 所示。具体来说，分为三个模块：1) 句子粒度语义编码模块；2) 词粒度语义扩展码模块；3) 词粒度-句子粒度语义融合模块。

2.1 句子粒度语义编码

首先，利用 Word Price Model 标记化获得输入文本序列中 token 和位置嵌入的总和，这一做法可以确保为任意字符序列生成确定性的切分，具体过程如下：

$$S_w = (w_1, w_2, \dots, w_{m-1}, w_m) \quad (1)$$

$$S_t = ([CLS], t_1, t_2, \dots, t_{n-1}, t_n)$$

其中， S_w 表示输入文本序列， m 代表文本序列中包含的单词数量； S_t 表示输入序列中 token 序列， n 代表序列中包含的 token 数量。

随后，将文本序列 S_w 输入到 BERT 中，进一步编码得到隐藏状态，具体表示如下：

$$h_t = \text{BERT}(S_t; \theta_{S_t}^{\text{BERT}}) \quad (2)$$

其中， $h_t = (h_1, h_2, \dots, h_{n-1}, h_n)$ 表示文本序列中每个 token 对应的隐状态， $\theta_{S_t}^{\text{BERT}}$ 表示 BERT 中的可学习参数。

接下来，分别提取每个 token 对应隐藏状态左右两端的上下文特征信息，并利用动态多池保留每个 token 左右两部分的极大值。与传统的最大池化相比，动态多池化可以在不丢失最大池化值的情况下保留更多有价值的信息^[9]，具体表示如下：

$$E_1 = \max([h_1, h_2, \dots, h_j])$$

$$E_2 = \max([h_{j+1}, h_{j+2}, \dots, h_n]) \quad (3)$$

$$E_s = [E_1; E_2]$$

其中， $[\]$ 表示拼接操作； $\max(\cdot)$ 表示最大池化操作； E_1 表示第 j 个 token 左边部分上下文信息特征； E_2 表示第 j 个 token 右边部分上下文信息特征； E_s

表示句子粒度语义特征。

2.2 词粒度语义扩展信息

词性筛选

本文利用词性标注工具 Stanford CoreNLP 对输入序列进行筛选, 得到满足词性的词语位置, 具体表示如下:

$$\text{Loc}_i = \begin{cases} 1, \text{Pos}_{w_i} \in \text{Set}_p \\ 0, \text{Pos}_{w_i} \notin \text{Set}_p \end{cases} \quad (4)$$

其中, Set_p 表示词性集, 具体包括动词词性和名词词性; Pos_{w_i} 表示第 i 个单词的词性; Loc_i 为 0 表示该位置的词语不满足词性; Loc_i 为 1 表示该位置的词语满足词性, 即候选触发词位置。

语义扩展

在得到候选触发词位置后, 为了扩展候选触发词在当前语境下的语义信息, 受到完形填空任务^[26]的启发, 使用掩码预训练语言模型对每个候选触发词位置进行遮盖。

首先, 在输入文本序列中加入占位符, 每次只对一个满足词性的候选触发词位置进行遮盖, 具体过程表示如下:

$$S_w = ([\text{CLS}], w_1, \dots, w_{m-1}, w_m, [\text{SEP}]) \quad (5)$$

$$S_{\text{mask}} = ([\text{CLS}], w_1, \dots, [\text{MASK}], w_m, [\text{SEP}])$$

然后, 利用掩码预训练语言模型对带有占位符 [MASK] 的文本序列 S_{mask} 进行预测, 具体表示如下:

$$H_{\text{mask}} = \text{BERT}_{\text{mask}}(S_{\text{mask}}; \theta_{S_{\text{mask}}}^{\text{BERT}_{\text{mask}}}) \quad (6)$$

$$P_{\text{mask}} = \text{softmax}(H_{\text{mask}})$$

其中, H_{mask} 表示带有占位符的文本序列 S_{mask} 的隐藏状态; P_{mask} 表示预测单词的概率分布。这里模型得到的预测单词表征是固定的, 不会跟随模型一起进行训练。如果预测单词表征是动态的, 那么每一次训练得到的词粒度语义特征都会出现变化, 当遇到一些未被训练的输入序列仍会使用之前产生的表征, 这样可能会导致错误的识别。

随后, 取 Top-k 预测单词的表征作为词粒度语义扩展特征, 具体表示如下:

$$\mathbf{K} = \text{Largestk}(p_{\text{mask}}, k) \quad (7)$$

$$E_w = [\mathbf{K}_1; \mathbf{K}_2; \dots; \mathbf{K}_k]$$

其中, $\text{Largestk}(\cdot)$ 返回候选词中最大的 k 个元素, k 是超参数; $\mathbf{K} = (\mathbf{K}_1, \mathbf{K}_2, \dots, \mathbf{K}_k)$ 表示 Top-k 预测

单词的表征; E_w 表示词粒度语义扩展特征。

2.3 词粒度-句子粒度语义融合

在得到句子粒度语义特征 E_s 和词粒度语义特征 E_w 后, 将其进行融合, 得到最终鲁棒的语义表征, 具体过程如下:

$$E = \text{MLP}([E_s; E_w]; \theta_E^{\text{MLP}}) \quad (8)$$

其中, $[\cdot]$ 表示拼接操作; $\text{MLP}(\cdot)$ 表示多层感知机; θ_E^{MLP} 代表多层感知机中的可学习参数; E 表示最终鲁棒的语义表征。

利用最终的融合语义特征经过 Softmax 操作, 得到最终的事件类型概率分布, 并根据分布概率判断输入序列的事件类型, 具体计算过程如下:

$$P = \text{softmax}(E)$$

$$\mathcal{L} = -\frac{1}{n} \sum_{i=1}^n q_i \cdot \log(P_i) \quad (9)$$

其中, P 表示预测事件类型概率分布; \mathcal{L} 表示损失函数; q_i 是真实标签的 one-hot 编码。

3 实验及结果分析

3.1 数据集

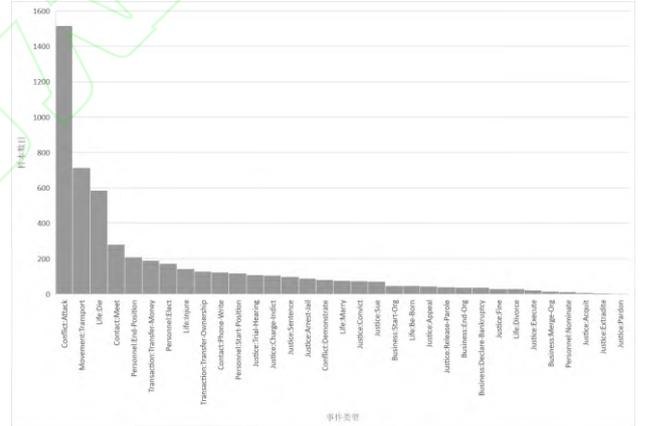


图 2 ACE2005 中事件类型样本大小分布
Fig.2 The Sample size distribution of event detection in ACE2005 dataset

本文实验中采用 ACE2005^[11]和 KBP2015^[27]语料库作为数据集来评估所提出的模型。ACE2005 数据集是事件检测任务广泛使用的一个基准数据集, 其中包括来自不同领域的文档集合, 例如文本新闻、广播对话和博客等。该数据集中包含 599 个文档并定义了 34 种事件类型, 同时提供对应的事件触发词、事件参数和事件类型注释。为避免数据预处理对模型性能产生较大影响, 本文采用与以前工作^[17,22,28]相同的方式分割数据, 训练集、验证集和测试集的文件数量分别为 529、30 和 40。

KBP2015 数据集是来自 2015 年文本分析会议 (TAC) 的事件 Nugget 检测评估的评估数据。数据集包含 360 个文档, 定义了 39 种事件类型。本文采用与官方^[27]相同的数据分割方式, 训练集和测试集分别包括 158 个文档和 202 个文档, 并将训练集的一个子集作为验证集, 约占训练集的 20%。

以 ACE2005 数据集为例, 该数据集中数据标注情况存在不平衡现象, 统计情况具体如图 2 所示。位于头部的“Conflict-Attack”事件类型的触发词数量远远高于位于尾部的事件类型总和。其中尤为显著的是数量排名最后五类事件类型中, 触发词数量仅有个位数。

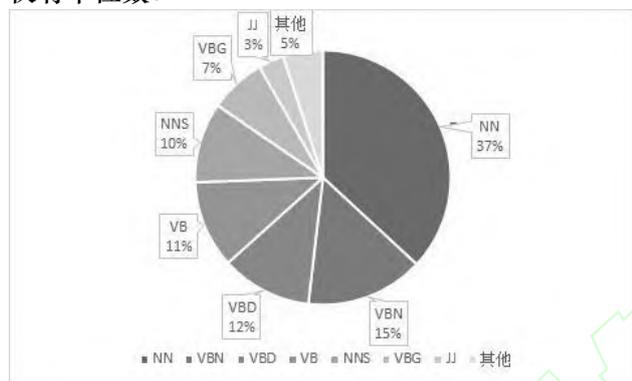


图 3 ACE2005 数据集触发词词性统计

Fig.3 The statistics of ACE2005dataset trigger part-of-speech

触发词是句子中最能清楚表达事件核心含义的词, 通常是一个名词或动词^[1]。以 ACE2005 数据集为例, 对数据集中的触发词词性进行统计, 具体结果如图 3 所示。名词词性包括: 单数名词形式 (NN)、复数名词形式 (NNS) 等; 动词词性包括: 过去分词 (VBN)、动词过去式 (VBD)、动词基本形式 (VB)、动名词和现在分词 (VBG) 等。由统计结果可以看出, 触发词中名词和动词大约占总数的 92%。因此, 对动词或名词词性的特定位置进行语义扩展, 可以缩小候选触发词范围, 并且不受数据集中样本数量的影响。

3.2 参数设置

本文提出模型的实现基于 pytorch 框架, 利用两张 NVIDIA 3080Ti-12GB GPU 进行训练。实验中采用 Adam 作为优化器。实验参数设置如表 1 所示。

如 2 节所述, 本文模型分为三个模块 1) 句子粒度语义编码模块; 2) 词粒度语义扩展模块; 3) 词粒度-句子粒度语义融合模块。其中, 只有第二个模块中的参数不参与更新, 不具有梯度, 其余模块参数均具有梯度。词粒度语义扩展模块参数不具有

梯度的原因是: 词粒度语义扩展信息旨在扩展符合当前输入序列上下文语义信息的词粒度信息表征, 而不能受到其他输入序列表征的干扰, 否则将会与当前语境产生偏差, 可能会导致错误的识别。

3.3 评价指标

本文实验选取事件检测人物研究中常用的 3 个评价指标: P、R 和 F1 分数来评估结果。

1) 准确率 (Precision) 表示正确预测的事件在总预测的事件中的比例。

2) 召回率 (Recall) 表示正确预测的事件在所有事件样本中的比例。

3) F1 分数根据准确率和召回率计算得来, 具体为 $F1 = 2RP / (R + P)$ 。

表 1 实验参数设置

Table 1 Experimental parameter setting

参数	值
BERT embedding size	1024
Batch size	16
Epoch	100
Dropout	0.3
Top-K	30
Learning rate	8e-6
BERT layer	4
MLP dropout	0.1
MLP learning rate	0.001

3.4 基线模型

为了充分验证本文提出模型的性能, 选取了近年来主流的事件检测模型与 FESPOS-ED 用上述三种评价指标拉进行公平地比较。这些基线方法如下:

1) **DMCNN^[17]模型**: 在事件检测任务中首次引入了动态多池化层作为特征提取器, 该特征提取器更加关注事件触发词与事件要素信息从而保留更重要的信息。

2) **JRNN^[29]模型**: 为了避免管道模型中的误差传播问题采用双向 RNN 来学习更加丰富的句子表示, 同时也考虑到了事件触发词与事件要素信息之间的联系。

3) **dbRNN^[30]模型**: 在使用 RNN 方法的同时应用了树结构与序列结构来提高事件检测模型的性能, 丰富了每个 token 的信息表示。

4) **GCN-ED^[31]模型**: 首次采用基于依赖树的卷积神经网络模型来改进事件检测, 并且提出一种新颖的基于实体提及的卷积向量聚合方法。

5) **JMEE^[19]模型**: 引入依存句法树方法并使用

基于注意力机制的卷积神经网络对图信息进行建模,从而解决事件触发词歧义的问题并提升事件检测的效果。

6) **EE-GCN^[32]模型**:通过融合句法结构和类型依赖标签并且以上下文相关的方式更新关系表示来改进卷积神经网络模型。

7) **Gated-GCN^[20]模型**:采用 BERT 进行编码,并利用门控机制根据候选触发词的信息过滤图卷积神经网络模型中的噪声信息。

8) **Adv-DMBERT^[10]模型**:采用一种对抗训练机制,不仅可以从候选集中提取实例信息,还可以提高事件检测模型在嘈杂环境中的性能。

9) **EKD^[12]模型**:引入开放域触发知识,为未见过/稀疏标记的触发词提供额外的语义支持,并改进触发识别。

10) **SEGCN-DCR^[24]模型**:采用一种带有解耦分类重新平衡机制的语法增强型卷积神经网络方法,来提升在数据分布不均衡的情况下事件检测模型的性能。

3.5 对比实验

为了充分验证本文提出模型的性能,利用上述基线在相同的实验条件下进行充分的对比。在 ACE2005 数据集的实验结果,如表 2 所示。

表 2 ACE2005 数据集实验结果
Table 2 Experimental results on ACE2005dataset

Model	触发词识别			触发词分类		
	P	R	F1	P	R	F1
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
dbRNN▼	74.1	69.8	71.9	74.1	69.8	71.9
GCN-ED△	-	-	-	77.9	68.6	73.1
JMEE△	80.2	72.1	75.9	76.3	71.3	73.7
EE-GCN△	-	-	-	76.7	78.6	77.6
GatedGCN△	78.8	76.3	77.6	78.8	76.3	77.6
Adv-DMBERT★※	-	-	-	77.6	71.8	74.6
EKD★	-	-	-	79.1	78.0	78.6
SEGCN-DCR△	-	-	-	77.1	79.6	78.3
FESPOS-ED(Ours)	79.5	83.9	81.6	77.3	81.8	79.5

★表示扩充训练实例的方法;▼表示基于循环神经网络的方法;△表示基于图神经网络的方法;※表示引入动态多池层的方法。

在 ACE2005 数据集上的实验结果表明,本文提出模型在触发词识别任务中的性能有明显提升,且在事件检测任务上的性能要优于其他对比模型,召回率和 F1 分数都有所提升,尤其是对比先进的

SEGCN-DCR^[24]模型,召回率和 F1 分数分别提升了 2.2 个百分点和 1.2 百分点。这充分说明了 FESPOS-ED 在事件检测任务上的有效性和先进性。在 ACE2005 数据集上,EKD^[12]模型的准确率最高,是因为其引入了丰富的开放域触发词知识并利用师生模型减少注释中的内在偏差,但其召回率较低。从实验结果可以看出:

1) 与扩充训练实例的方法 Adv-DMBERT^[10]和 EKD^[12]相比,本文提出模型在 F1 分数上均有所提升,分别提高了 4.9%和 0.9%。这可能是因为他们生成的语料库都是同质,仍然存在内置偏差,并且在一定程度上受限于知识库的低覆盖。而本文提出方法不扩充句子级训练实例,而在词粒度上进行语义扩展,避免引入噪声干扰模型训练,且可以挖掘候选触发词上下文中的丰富语义。

2) 与基于循环神经网络的方法 JRNN^[29]、dbRNN^[30]相比,文本模型性能更加优异,在 F1 分数上分别提升了 10.2%、7.6%。这可能是因为循环神经网络无法解决句子长距离依赖问题,而本文模型同时利用词粒度和句子粒度语义信息,有效地避免此问题。

3) 与基于图神经网络的方法 GCN-ED^[31]、JMEE^[19]、EE-GCN^[32]、GatedGCN^[20]、SEGCN-DCR^[21]相比,本文提出模型性能均有所提升,在 F1 分数上分别提高了 6.4%、5.8%、1.9%、1.9%、1.2%。这可能是因为 GCN 不能为每个邻居分配不同的权重,在卷积时对所有邻居赋予同等的重要性,不能根据节点重要性分配不同的权重;并且 GCN 将特征编码为高阶向量,会引入过多额外参数。而本文提出方法,着重对符合触发词词性的候选触发词进行语义扩展,关注特定位置符合语境的上下文信息预测,且没有过多参数。

4) 与引入动态池化层的方法 DMCNN^[17]和 Adv-DMBERT^[10]相比,本文提出方法在 F1 分数上分别提升了 10.4%和 4.9%。这可能是因为引入动态池化层的方法仅关注句子粒度语义信息,没有重视与触发词相关的词粒度语义扩展信息。

综上所述,FESPOS-ED 仅在词粒度上进行语义扩展,利用词性筛选缩小候选触发词范围并扩展特定位置的语义信息,不仅考虑候选触发词信息,而且还充分考虑了候选触发词与当前语境的关联;同时,融合词粒度语义扩展信息和句子粒度语义信息增强了分类能力,因此性能优于对比基线模型。

表 3 KBP2015 数据集实验结果
Table 3 Experimental results on KBP2015 dataset

Model	触发词识别			触发词分类		
	P	R	F1	P	R	F1
DMCNN	77.4	48.2	59.8	71.3	45.8	55.8
JRNN	66.7	53.5	59.4	63.1	54.0	58.2
TAC Top	82.0	52.0	63.7	75.2	47.4	58.4
GCN-ED	-	-	-	70.3	50.6	58.8
Adv-DMBERT	-	-	-	71.0	54.6	61.7
FESPOS-ED(Ours)	83.5	65.3	73.3	77.8	59.6	67.5

为了进一步验证所提方法的有效性,在 KBP2015 数据集上进行了实验,实验结果如表 3 所示。

在 KBP2015 数据集上的实验结果表明,本文提出模型在触发词识别与触发词分类任务上的性能都有不同程度的提升。其中在触发词识别任务上,准确率、召回率和 F1 分数分别达到了 83.5%、65.3% 和 73.3%,在触发词分类任务上准确率、召回率和 F1 分数分别达到了 77.8%、59.6% 和 67.5%。TAC Top^[27]是在 TAC KBP2015 事件检测中获得排名第一的结果,与之相比,本文提出模型也表现出优异的性能。

3.6 稀疏标记数据实验

当数据标记不平衡时,整体性能无法衡量模型在稀疏标记数据上的表现。本实验中,根据训练数据的分布拆分测试集并对 ACE2005 数据集中数量小于 10 条数据的类别进行了测试,并与基线模型在相同实验条件下进行比较。实验结果如表 4 所示。

表 4 稀疏标记数据实验结果
Table 4 Experimental results on ACE2005 dataset

Model	稀疏标记数据		
	P	R	F1
EE-GCN	61.5	57.1	59.2
GatedGCN	62.3	60.6	61.4
SEGCN-DCR	61.9	64.7	63.3
Adv-DMBERT	71.9	78.6	75.1
FESPOS-ED(Ours)	76.7	80.5	78.5

由表 4 可知,随着实验数据的减少,所有模型的性能都会显著下降。但是在稀疏标记数据上,本文提出方法在各项指标上都有所改善,与扩充训练实例的方法 Adv-DMBERT^[10]和在分类损失上进行平衡的方法 SEGCN-DCR^[21]相比,在三个指标上都有不同程度的提升,在 F1 分数上分别提高了 3.4% 和 15.2%。可以观察到在召回率上的改善最为显著,这可以解释为词性筛选机制可以缩小候选触发词范

围,在这些候选触发词中极大概率包含真正的触发词,从而使更多稀疏标记的触发词被模型识别。并且扩充候选触发词语义信息,挖掘触发词的上下文中蕴含的丰富语义信息进行训练,进而提升事件类型分类性能。实验结果表明,该方法能够缓解稀疏标记数据得不到有效训练而造成的影响。

3.7 消融实验

为了进一步探究模型中各个子网络的具体作用及其对于最终结果的影响,进行了消融实验。实验中使用 ACE2005 数据集,具体结果如表 5 所示。

1) 对词性筛选模块的消融

(-)词性筛选模块:①消除词性筛选模块,对输入序列的每个词进行语义扩展;②消除词性筛选模块,对输入序列进行随机语义扩展。

从实验结果可以看出消除词性筛选模块后,模型性能受到损伤,原因是无论是对整个输入序列进行语义扩展还是对输入序列进行随机语义扩展,都有可能引入与触发词无关的噪声,从而损伤模型性能。

表 5 ACE2005 数据集消融实验
Table 5 Ablation study on ACE2005 dataset

FESPOS-ED		P	R	F1
		77.3	81.8	79.5
(-)词性筛选模块	所有词	76.1	77.1	76.6
	随机词	75.5	76.7	76.1
(-)词粒度语义扩展信息		73.6	78.1	75.8
(-)句子粒度语义特征		60.7	30.8	40.9
(-)融合模块		76.0	81.6	78.7

2) 对语义信息模块的消融

(-)词粒度语义扩展信息:消除词粒度语义扩展信息模块,仅使用句子粒度语义信息进行事件检测;(-)句子粒度语义信息:消除句子粒度语义信息,仅使用词粒度语义扩展信息进行事件检测。

实验结果表明,模型在 F1 分数上有不同程度地下降,分别下降了 3.7% 和 38.6%,这表明在本文提出模型中,句子粒度语义信息和词粒度语义扩展信息都是不可或缺的。词粒度语义扩展信息能够提供候选触发词的语义扩展信息,句子粒度语义信息能够提供输入序列的上下文信息,两种粒度的语义信息融合生成鲁棒的语义表征对事件检测更加有效。

3) 对融合模块的消融

(-)融合模块:消除融合模块,仅将词粒度语义扩展信息和句子粒度语义信息进行拼接。

由实验结果可以看出拼接操作缺少句子粒度语

义信息和词粒度语义信息的交互过程，不能有效融合两种语义信息。

3.8 实例分析

为了更好地验证本文提出的模型，本节从ACE2005数据集中挑选了若干条数据。(1)组数据中，为了验证本文提出模型的鲁棒性，将两条数据中的触发词换为同义的未见过的触发词进行测试。

(2)组数据都来源于ACE2005数据集中的稀疏标记数据，为了验证本文提出模型的有效性，对两条稀疏标记数据进行测试。

由实验结果可以看出，在(1)组数据中，(a)句的原始触发词是“Bankrupt”，将其替换为数据集中未见过的同义词“Insolvent”；(b)句的原始触发词是“fallen”，将其替换为未见过的同义词“deceased”。由于触发词在极大概率上是动词或名词，本文提出模型利用词性筛选缩小候选触发词范

围，在未见过的触发词识别及分类上有天然的优势。无论是什么样的触发词都有极大概率包括在特定词性内，因此在未见过的样本测试中模型依旧能够正确识别，有效地说明了该模型具有良好的泛化性。

在(2)组数据中，(c)句对应的触发词是“fined”，事件类型为“Justice:Fine”，该类型在数据集中仅有5条样本；(d)句对应的触发词是“amnesty”，事件类型为“Justice:Pardon”，该类型在数据集中仅有1条样本，本文提出模型都能够正确识别触发词和事件类型。实验结果表明该模型在稀疏标记数据中，不仅能够利用特定词性的词粒度语义扩展信息正确识别触发词，并且还能有效利用输入序列的词粒度-句子粒度语义信息正确识别事件类型，充分说明了FESPOS-ED的有效性和先进性。

表 6 ACE2005 数据集实例分析
Table 6 Case study on ACE2005 dataset

	文本	真实事件类型	预测事件类型
(1)未见过触发词实例	(a) <u>Bankrupt</u> energy giant Enron is considering legal action against its own bankers. <u>Insolvent</u> energy giant Enron is considering legal action against its own bankers	Business Declare-Bankruptcy	Business Declare-Bankruptcy
	(b) The commander in chief told the family of one <u>fallen</u> Marine in a private moment. The commander in chief told the family of one <u>deceased</u> Marine in a private moment.	Life Die	Life Die
(2)稀疏标记触发词实例	(c) Iraq <u>fined</u> millions of dollars for falsifying documents to buy uranium from the West African country of Niger.	Justice Fine	Justice Fine
	(d) More than ninety prisoners have been set free so far under a government <u>amnesty</u> .	Justice Pardon	Justice Pardon

4 结束语

针对数据标注不平衡导致稀疏标记数据得不到有效训练的问题，本文提出融合词性语义扩展信息的事件检测方法(FESPOS-ED)通过词性语义扩展，缩小候选触发词范围的同时挖掘候选触发词的上下文丰富语义，并融合句子粒度语义信息进行识别和分类，避免训练受到样本数量的限制，缓解稀疏标记数据带来的影响。实验表明，本文提出模型在ACE2005和KBP2015数据集上都具有良好性能，在稀疏标记数据上也取得了具有竞争性的效果。在未来的工作中，我们将进一步在无触发词的事件检测方面进行探索和研究。

参考文献

[1] Linguistic Data Consortium. ACE (Automatic Content Extraction) English Annotation Guidelines for Events

Version 5.4.3[EB/OL]. [2022-12-07]. <https://www ldc.upenn.edu/>.

[2] Xiang W, Wang B. A survey of event extraction from text[J]. IEEE Access, 2019, 7: 173111-173137.

[3] Ahn D. The stages of event extraction[C] //Proceedings of the Workshop on Annotating and Reasoning about Time and Events. Stroudsburg, PA : Association for Computational Linguistics, 2006: 1-8.

[4] 李中秋,洪宇,王捷,周国栋.基于实体画像增强网络的事件检测方法[J].中文信息学报,2022,36(8):81-91.LI Zhongqiu, HONG Yu, WANG Jie, ZHOU Guodong. Entity Profile Enhancement Network for Event Detection[J]. Journal of Chinese Information Processing,2022,36(8):81-91.

[5] 陈斌,周勇,刘兵.基于卷积双向长短期记忆网络的事件触发词抽取[J].计算机工程,2019,45(1): 153-158. CHEN Bin, ZHOU Yong, LIU Bing. Event Trigger Word Extraction Based on Convolutional Bidirectional Long

- Short Term Memory Network[J]. Computer Engineering, 2019, 45(1): 153-158.
- [6] Ferguson J, Lockard C, Weld D S, et al. Semi-Supervised Event Extraction with Paraphrase Clusters[C]//Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers). Stroudsburg, PA: Association for Computational Linguistics, 2018: 359-364.
- [7] Araki J, Mitamura T. Open-domain event detection using distant supervision[C] //Proceedings of the 27th International Conference on Computational Linguistics. Stroudsburg, PA : Association for Computational Linguistics, 2018: 878-891.
- [8] Cao Y, Hu Z, Chua T S, et al. Low-Resource Name Tagging Learned with Weakly Labeled Data[C]//Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing. Stroudsburg, PA: Association for Computational Linguistics, 2019: 261-270.
- [9] Chen Y, Liu S, Zhang X, et al. Automatically labeled data generation for large scale event extraction[C]// Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Stroudsburg, PA : Association for Computational Linguistics, 2017: 409-419.
- [10] Wang X, Han X, Liu Z, et al. Adversarial training for weakly supervised event detection[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). Stroudsburg, PA: Association for Computational Linguistics,2019: 998-1008.
- [11] Zeng Y, Feng Y, Ma R, et al. Scale up event extraction learning via automatic training data generation[C]//Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence and Thirtieth Innovative Applications of Artificial Intelligence Conference and Eighth AAAI Symposium on Educational Advances in Artificial Intelligence. Palo Alto, CA: AAAI Press, 2018: 6045-6052.
- [12] Tong M, Xu B, Wang S, et al. Improving Event Detection via Open-domain Trigger Knowledge[C]//Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. Stroudsburg, PA: Association for Computational Linguistics, 2020: 5887-5897.
- [13] Ji H, Grishman R. Refining event extraction through cross-document inference[C]/ Proceedings of the 46rd Annual Meeting of the Association for Computational Linguistics, Hlt. Stroudsburg, PA : Association for Computational Linguistics, 2008: 254-262.
- [14] Qin Y, Zhang Y, Zhang M, et al. Feature-Rich Segment-Based News Event Detection on Twitter[C]// International Joint Conference on Natural Language Processing. Stroudsburg, PA : Association for Computational Linguistics, 2013: 302-310.
- [15] Liu S, Chen Y, He S, et al. Leveraging framenet to improve automatic event detection[C]//Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Stroudsburg, PA: Association for Computational Linguistics, 2016: 2134-2143.
- [16] Liu S, Chen Y, Liu K, et al. Exploiting argument information to improve event detection via supervised attention mechanisms[C]// Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Stroudsburg, PA: Association for Computational Linguistics, 2017: 1789-1798.
- [17] 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 (Volume 1: Long Papers). Stroudsburg, PA: Association for Computational Linguistics, 2015: 167-176.
- [18] Nguyen T H, Grishman R. Modeling skip-grams for event detection with convolutional neural networks[C]//Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing. Stroudsburg, PA : Association for Computational Linguistics, 2016: 886-891.
- [19] Liu X, Luo Z, Huang H Y. Jointly Multiple Events Extraction via Attention-based Graph Information Aggregation[C]// Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. Stroudsburg, PA : Association for Computational Linguistics, 2018: 1247-1256.

- [20] Nguyen T N, Nguyen T H. Event Detection: Gate Diversity and Syntactic Importance Scores for Graph Convolution Neural Networks[C]//Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing. Stroudsburg, PA: Association for Computational Linguistics, 2020: 5405-5411.
- [21] Hu B, Liu Y, Chen N, et al. SEG-CN-DCR: A syntax-enhanced event detection framework with decoupled classification rebalance[J]. *Neurocomputing*, 2022, 481: 55-66.
- [22] Liu J, Chen Y, Liu K, et al. How does context matter? On the robustness of event detection with context-selective mask generalization [C]//Findings of the Association for Computational Linguistics: EMNLP 2020. Stroudsburg, PA: Association for Computational Linguistics, 2020: 2523-2532.
- [23] Li R, Zhao W, Yang C, et al. Treasures Outside Contexts: Improving Event Detection via Global Statistics[C]//Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing. Stroudsburg, PA: Association for Computational Linguistics, 2021: 2625-2635.
- [24] Cao K, Wei C, Gaidon A, et al. Learning imbalanced datasets with label-distribution-aware margin loss[C]//Proceedings of the 33rd International Conference on Neural Information Processing Systems. Cham: Springer International Publishing, 2019: 1567-1578.
- [25] Khan S H, Hayat M, Bennamoun M, et al. Cos-t-sensitive learning of deep feature representations from imbalanced data[J]. *IEEE transactions on neural networks and learning systems*, 2017, 29(8): 3573-3587.
- [26] Petroni F, Rocktäschel T, Riedel S, et al. Language Models as Knowledge Bases?[C]//Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing. Stroudsburg, PA : Association for Computational Linguistics, 2019: 2463-2473.
- [27] Mitamura T, Liu Z.Z, Hovy E, Overview of TAC-KBP 2015 Event Nugget Track, in: TAC, 2015.
- [28] 苗佳, 段跃兴, 张月琴, 张泽华. 基于 CNN-BiGRU 模型的事件触发词抽取方法[J]. *计算机工程*, 2021, 47(9): 69-74,83.
MIAO Jia, DUAN Yueqing, ZHANG Yueqin, ZHANG Zehua. Method for Extracting Event Trigger Words Based on the CNN-BiGRU Model[J]. *Computer Engineering*, 2021, 47(9): 69-74,83.
- [29] Nguyen, T.H., Cho, K., Grishman, R., 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, pp. Stroudsburg, PA: Association for Computational Linguistics, 2016:300 – 309.
- [30] Sha L, Qian F, Chang B, et al. Jointly extracting event triggers and arguments by dependency-bridge RNN and tensor-based argument interaction[C]//Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence and Thirtieth Innovative Applications of Artificial Intelligence Conference and Eighth AAAI Symposium on Educational Advances in Artificial Intelligence. Palo Alto, CA: AAAI Press, 2018: 5916-5923.
- [31] Nguyen T H, Grishman R. Graph convolutional networks with argument-aware pooling for event detection[C]//Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence and Thirtieth Innovative Applications of Artificial Intelligence Conference and Eighth AAAI Symposium on Educational Advances in Artificial Intelligence. Palo Alto, CA: AAAI Press, 2018: 5900-5907.
- [32] Cui S, Yu B, Liu T, et al. Edge-Enhanced Graph Convolution Networks for Event Detection with Syntactic Relation[C]// Findings of the Association for Computational Linguistics: EMNLP 2020. Stroudsburg, PA : Association for Computational Linguistics, 2020: 2329-2339.