高技术通讯 2023 年 第33 卷 第7 期:669-682

doi:10.3772/j.issn.1002-0470.2023.07.001

基于对比学习增强句子语义的事件检测方法①

梁 冬②*** 张 程* 史 骁③* 谭文婷*** 吕存驰*** 赵晓芳*****

(*中国科学院计算技术研究所 北京 100190)

(**中国科学院大学 北京 100049)

(***中科苏州智能计算技术研究院 苏州 215028)

摘 要 事件检测旨在识别文本中提到的事件及其类型。基于触发词的事件检测方法需要额外的人工成本标注事件触发词。本文从无触发词的文本语义提取出发,提出了一种基于对比学习增强句子语义的事件检测方法。该方法首先在事件检测数据集上通过自监督学习对预训练的语言模型基于转换器的双向编码表示器(BERT)调优,提高语言模型的领域适应性。然后利用掩码(mask)操作和丢弃(dropout)操作构建自监督对比样例,增加监督对比样例,实现了自监督对比和监督对比2种句子语义增强的方法。此外在训练过程中自动调整对比损失和事件分类的交叉熵损失的权重,以降低人工调参的成本,同时提高模型收敛速度。在自动内容抽取(ACE)2005 中英文语料上的实验结果表明,本文方法比先前无触发词事件检测方法取得更好的结果,与利用预训练 BERT 模型微调的事件检测方法相比也具有优势。

关键词 事件检测;自监督对比学习;监督对比学习;语义增强;自动调整权重

0 引 言

事件检测的目的是识别出文本中特定类型的事件[1-2],例如在句子“Liana Owen drove 10 hours from Pennsylvania to attend the rally in Manhattan with her parents.”中,需要识别出“Demonstrate”和“Transport”两种事件类型。基于触发词的识别和分类是当前常用的检测方法,这些方法将事件检测任务看成词的分类问题[3-8],通过对句子中每个词的分类,预测该词是否为一个事件触发词以及其触发的事件类型。比如上述句子要先识别出触发词“rally”和“drove”,然后检测出其分别触发了“Demonstrate”和“Transport”事件。

然而,基于触发词的事件检测方法存在2个方面的问题:(1)构造训练数据比较耗时,标注人员不仅需要标注事件类型,还需要标注事件触发词。如上述句子需要标注{rally:Demonstrate, drove:Transport}。根据自动内容抽取(automatic context extraction, ACE)事件评估项目[9],触发词最能表述一个事件的发生。但是从已知的句子中挑选最能代表事件发生的词是需要专业的综合判断,往往会花费较多的精力,特别是对于一个很长的句子,尤其是文档级别的事件检测任务。(2)过度依赖触发词的事件检测,会丢失很多语义信息。先前的工作[10-11]表明文本所属的事件类型不仅依赖于触发词特征,还取决于文本的上下文语义特征。同类事件的发生可能是不同的触发词引起的,而相同的触发词根据其语义环境会触发不一致的事件。过度依赖触发词特征,容易导致事件类型的模糊性,难以正确识别已知事件类型未见过的触发词,从而使事件检测准确率

① 国家重点研发计划(2021YFF0703800)资助项目。

② 女,1989 年生,博士生;研究方向:自然语言处理,大数据分析,分布式训练;E-mail:liangdong@ict.ac.cn。

③ 通信作者,E-mail:shixiao@ict.ac.cn。

(收稿日期:2022-08-10)

下降。

针对上述问题,研究人员设计了无触发词的事件检测方法[12-14]。具有注意力机制的类型感知偏差神经网络(type-aware bias neural network with attention mechanisms, TBNNAM)[12]提出触发词的识别不是事件检测任务的必需步骤,并研究了无触发词的事件检测,利用基于事件类型的注意力机制来弥补触发词缺失造成的重要线索信息的丢失。Doc2EDAG[13]设计了一个无触发词的文档级别事件抽取方法,通过对整篇文档内容的分类来识别事件类型,以缓解事件标注的困难性。文献[14]实现了在无触发词条件下定位文档中的事件表述中心句并判断事件类型的方法。这些方法简化了事件检测的流程,能够降低数据标注的压力,但是由于监督学习需要大规模地标注数据才能学习到强健的语义表示,而目前可用的事件检测数据资源比较匮乏,这些模型在提取事件信息的语义特征方面还有待提高。最近兴起的对比学习是一种强大的表示提取方法,受到了越来越多的关注,在自然语言处理(natural language processing,NLP)领域也得到广泛应用[15]。对比学习为文本语义表示提供了一种新思路,在监督数据有限的情况下,可以有效地提取文本的语义信息。

本文提出了一种基于对比学习增强句子语义的事件检测方法。该方法在事件检测任务中忽略触发词识别这一中间步骤,降低了事件标注成本,简化了事件检测任务的流程,同时引入对比学习增强句子语义表示,提高事件检测结果的可靠性。本文的主要贡献总结为以下 4 点。

(1)利用掩码(mask)操作和丢弃(dropout)操作构建自监督对比样例,增加监督对比样例,实现了自监督对比和监督对比 2 种句子语义增强的方法,并进一步分析了不同对比学习方法的效果及收敛速度。

(2)在事件检测数据集上通过自监督学习对预训练的语言模型基于转换器的双向编码表示器(bidirectional encoder representations from transformers, BERT)调优,增强事件文本字符级别语义表示和提取,提高了模型在事件检测任务上的领域适应性。

(3)在训练过程中自动调整对比损失和事件检测的交叉熵损失的权重,能够更快地获得较优的结果。

(4)在 ACE 项目[9]的中英文数据集上进行了详尽的实验分析,证实了本文方法的有效性。

本文的组织结构如下。第 1 节介绍事件检测,预训练语言模型,对比学习相关研究工作;第 2 节阐述基于对比学习增强句子语义的事件检测方法;第 3 节通过实验对比验证上述方法的有效性;第 4 节对全文进行总结。

1 相关工作

1.1 事件检测

事件检测是事件抽取的一个重要子任务,目的是识别出文本中提到的事件类型。这一问题受到了研究人员的广泛关注,现有的方法大多是基于触发词的识别和分类实现的,通常遵循监督学习范式,可以分为基于特征工程的方法和基于表示学习的方法。

基于特征工程的事件检测方法是通过人工设计的精巧特征将事件分类线索转化为特征向量,输入到随机森林(random forest, RF)或者支持向量机(support vector machine, SVM)之类的模型中进行事件识别。例如,文献[16]利用词汇、句法、额外知识等特征。文献[17]结合了相关文档的全局特征和局部决策特征。文献[18,19]提出了跨事件或跨实体的特征。文献[4]采用一个联合模型来捕获触发词和参数的组合特征。

随着深度学习技术的发展,近几年来的事件检测方法倾向于基于神经网络的表示学习方法。例如,文献[1,20]利用卷积神经网络(convolutional neural networks,CNNs)学习事件特征信息。文献[2,5,21,22]利用循环神经网络(recurrent neural networks, RNNs)学习词在事件中上下文语义信息。文献[23]结合 CNNs 和 RNNs 提取不同层面的事件特征。文献[24]利用注意力机制编码参数信息。文献[6,25]利用图神经网络探索了句法信息特征。文献[26,27,28]学习了文档级别的事件线索。

然而,上述的这些方法均需要额外注释触发词,增加了获取事件检测训练数据的成本,限制了数据驱动的深层网络应用。为了缓解事件标注的困难,受先前工作[12-14]的启发,本文开展了无触发词的事件检测研究工作。

1.2　预训练语言模型

近年来,预训练语言模型的发展将 NLP 领域的研究带入到一个新阶段,从海量语料中学习的通用语言表示,能够显著提升下游任务。ELMo[29]、GPT[30]、BERT[31]、GPT2[32]等模型的相继提出,不断刷新了各大 NLP 任务的排行榜,并表明采用更多的训练数据可以不断地提高模型性能。

预训练语言模型通过自监督学习,充分利用大量无监督的文本数据,编码语言知识,能够根据上下文动态捕捉单词的语义。本文使用大规模的预训练语言模型 BERT[31]作为文本编码器,并在事件检测数据集上通过对随机掩码字符预测的调优训练,提高文本字符级别的事件语义表示,从而进一步提升模型在事件检测任务上的领域适应性。

1.3　对比学习

对比学习的主要思想是最小化给定样本(称为锚点)和正样例之间的距离,并最大化锚点和负样例之间的距离,从而更好地提取数据表示特征。如何构造正负样例是对比学习的关键问题。自监督对比学习没有标签信息,通过数据增强构造相似的样本,原数据和增强后的数据为正样例,和其他样本增强后的数据为负样例。图 1 所示为自监督对比学习示意图,虚线圆圈为同一颜色(或同一形状)的实线圆圈增强后的相似样本。监督对比学习利用类别标签关系构造正负样例,同一类别的数据为正样例,不同类别的数据为负样例。图 2 所示为监督对比学习示意图,不同颜色(或不同形状)属于不同的类别,本文利用数据增强操作提高同类别的样本数量,增加监督对比的正样例。

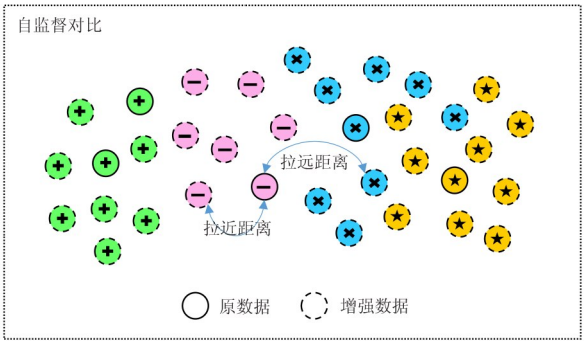


图 1　自监督对比学习示意图

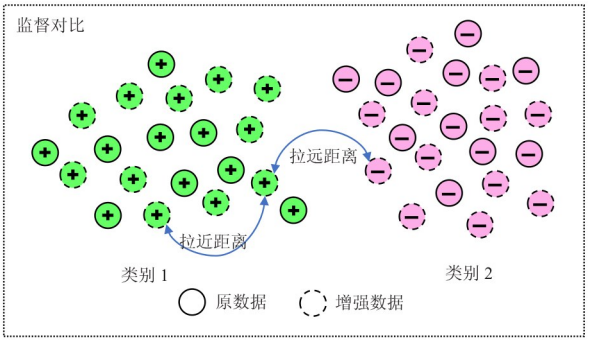


图 2　监督对比学习示意图

最近的研究工作[15,33-34]将对比学习应用在 NLP 领域,通过对比学习训练文本的编码器。文献[15]通过 dropout[35]操作设计了一个简单的对比学习框架,在文本语义相似度任务上取得先进的性能。文献[33]设计了一个监督对比损失函数,并对预训练语言模型微调。文献[34]通过 2 种数据增强方式构建自监督对比样例:对原始文本随机删除或掩码操作,利用 dropout 操作对特征层向量增强。

在事件检测任务中,对比学习的应用也有一定的研究。文献[36,37]采用三元组损失函数[38]将标签分类下的实例间的关系距离作为一个有效的监督信号,使得类别内部的样本距离小于不同类别样本的距离,但是没有探索无标签信息的句子语义距离关系。文献[39]利用文本中同一事件的参数比其他词更相近,事件相关的语义结构图比事件无关的语义结构距离更远的思想,构造自监督对比信号在大量的无监督数据上指导训练事件检测与抽取任务的预训练语言模型,没有关注对比学习在下游任务增强句子语义的能力。

本文在无触发词的事件检测研究中,为了增强句子的语义表示,实现了自监督对比和监督对比 2 种方式的表示学习方法,并实验分析了效果的差异性。

2　基于对比学习增强句子语义的事件检测方法

本文设计并实现了基于句子语义提取的事件检测框架。首先通过自监督学习对预训练的语言模型 BERT 调优,提高模型在事件检测任务上的领域适应性。然后在此基础上提出了 2 种增强句子语义的事件检测方法,分别是自监督对比学习和监督对比学习,通过对比信号提高句子语义的表达能力,从而提升事件检测方法的效果。

2.1　基于句子语义提取的事件检测基础架构

根据最近事件检测的研究进展,本文采用 BERT 编码器学习输入文本的向量表示。图 3 所示为事件检测基础模型架构,该模型包括输入层、BERT 编码器、Dropout 层、线性分类器和输出层。对于给定的句子 S,首先在句子 S 的开始和结束位置添加特殊字符,构建扩展的序列“[CLS] S [SEP]”,利用 BERT 编码器对序列进行编码,输出序列第一个字符的最后一层的隐含层向量作为整个句子的语义表示;然后将句子的语义表示向量进行 dropout 操作,提高模型的鲁棒性,避免模型的过拟合;最后将 dropout 操作后句子表示输入到一个线性分类器中,进行事件类型的识别。线性分类器输出一个数组,长度为事件类型总个数,每个索引位置分别对应一类事件,数组元素值为 1 表示属于该索引位置对应的事件。图中表示句子“Liana Owen drove 10 hours from Pennsylvania to attend the rally in Manhattan with her parents.”预测为“Demonstrate”和“Transport” 2 种事件类型。

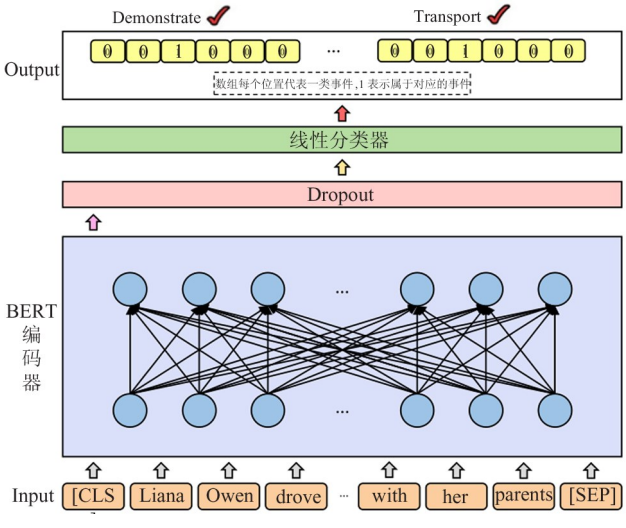


图 3　事件检测基础模型架构

事件检测任务采用二进制交叉熵损失函数。首先将线性分类器的输出经过 sigmoid 激活函数处理,归一化到(0,1)的范围,获取文本所属每一类事件的概率 $P(y_j \mid x_i)$,如式(1)所示,其中 $f_\theta(x_i)$ 表示输入的文本序列 x_i 经过 BERT 编码器、Dropout 层、线性分类器输出的结果,θ 为模型参数。然后在每个类别上计算负对数概率作为事件检测的损失函数 L_{ed},如式(2)所示,N 为每次迭代训练批数据量大小,K 为事件类型个数。

$$P(y_j \mid x_i) = \frac{1}{1 + \exp(-f_\theta(x_i))} \tag{1}$$

$$L_{\text{ed}} = -\frac{1}{N}\sum_{j=1}^{K}\sum_{i=1}^{N}[y_j \log P(y_j \mid x_i) + (1 - y_j)\log(1 - P(y_j \mid x_i))] \tag{2}$$

基于触发词的事件检测方法需要对每个字符的隐含层向量进行分类,本文在事件检测时忽略触发词,只需要对整个句子语义编码进行分类即可,相对更简单易于理解。为了满足句子可能描述多个事件的需求,本文通过判断句子预测为某类事件概率 $P(y_j \mid x_i)$ 是否大于阈值 T 来确定句子所属事件类型,若概率超过阈值则属于该事件类型,从而支持多标签事件检测。

2.2　面向事件检测的预训练语言模型调优

面向事件检测任务,本文采用自监督学习方法对预训练语言模型 BERT 调优,提高模型的领域适应性。具体方式为:使用事件检测数据集,通过自监督预测随机掩码字符的方式,对预训练语言模型做调优训练。相比于直接将预训练模型应用在下游任务,本文的调优训练能够增强模型对事件文本字符级别语义的提取,有助于下游任务对比学习的句子语义表示。

自监督预测使用的掩码语言模型损失函数[33]如式(3)所示,其中,$P(x_m)$ 为句子中一个掩码字符的预测概率,M 为一个批量数据中掩码字符的总个数。

$$L_{\text{mlm}} = -\frac{1}{M}\sum_{m=1}^{M}\log(P(x_m)) \tag{3}$$

算法 1 为预训练语言模型调优算法的流程。首先(第 2 和 3 行)对批数据中每条样本进行随机掩码操作,构建掩码字符预测的训练数据;然后(第 4 ~7 行)前向计算损失函数 L_{mlm},并利用损失函数的梯度更新 BERT 编码器的参数。本文随机掩码字符操作参照文献[33]的操作,随机选取 10% 的字符,如果某个位置的字符被选择,则以 80% 的概率用“[MASK]”字符替换,10% 的概率替换成随机字符,10% 的概率保持原字符不变。

算法 1　预训练语言模型调优算法

输入:自监督训练集 $\{x^{(n)}\}_{n=1}^{N}$,迭代次数 S

输出:更新后的模型参数 θ

1　for $s = 1 \cdots S$ do

2　　从 $\{x^{(n)}\}_{n=1}^{N}$ 中选择一批数据 $B = \{x^{(k)}\}_{k=1}^{K}$

3　　对 B 中每条样本进行随机掩码操作,构建模型调优的训练数据 $T_B = \{x_{\text{mask}}^{(k)}, y_{\text{mask}}^{(k)}\}_{k=1}^{K_{\text{mask}}}$

4　　利用 T_B 前向计算预测掩码字符的损失 L_{mlm}

5　　计算损失函数的梯度 $\nabla_\theta L_{\text{mlm}}$

6　　利用梯度 $\nabla_\theta L_{\text{mlm}}$ 计算模型新参数 θ'

7　　$\theta \leftarrow \theta'$

8　end for

2.3　基于自监督对比的事件检测算法

在自监督对比学习中,由于没有标签信息,本文采用 2 种数据增强方式构建自监督对比学习的正样例。如图 4 所示,在 BERT 编码器的输入端,通过句子的 mask 操作增强样例;在 BERT 编码器的输出端,通过句子特征层向量的 dropout 操作增强样例。本文对原文本的 mask 操作方法和上述预训练语言模型调优阶段的掩码方法相同。句子特征层向量的 dropout 操作如下:对 BERT 编码器输出的文本表示向量,以 0.3 的概率丢弃。原文本和自身数据增强后的文本属于同源样本,用来构造对比学习的正样例,和其他文本增强的数据属于不同源样本,构造为负样例。

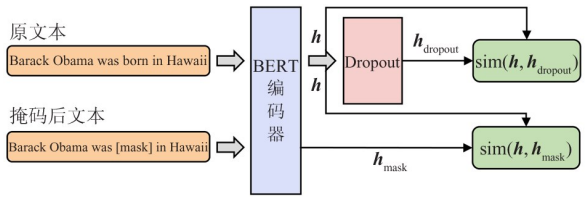


图 4　自监督对比学习构建正样例的 2 种方式

基于上述构造的正负样例,通过自监督方式对相近的语义进行理解和辨析[15,34],拉近同源样本在语义空间的距离,拉远不同源样本在语义空间的距离,从而提高样本语义表示能力。

自监督对比损失如式(4)所示,其中,N 为每个迭代训练中批量数据的句子总个数;τ 为温度系数,控制类间距离的超参数;$\text{sim}(\boldsymbol{h}_i, \bar{\boldsymbol{h}}_i)$ 表示输入 2 个向量 $\boldsymbol{h}_i$ 和 $\bar{\boldsymbol{h}}_i$ 的余弦相似度,$\boldsymbol{h}_i$ 为原文本经 BERT 编码器输出的向量。在使用 mask 操作构建正样例时,$\bar{\boldsymbol{h}}_i$ 为原文本数据增强后经 BERT 编码器输出的向量;在使用 dropout 操作构建正样例时,$\bar{\boldsymbol{h}}_i$ 为 BERT 编码器输出的原文本向量数据增强后的表示。

$$L_{\text{uns\_cl}} = -\frac{1}{N}\sum_{i=1}^{N}\log\frac{\exp(\text{sim}(\boldsymbol{h}_i, \bar{\boldsymbol{h}}_i)/\tau)}{\sum_{j=1}^{N}\mathbb{I}_{i\neq j}\exp(\text{sim}(\boldsymbol{h}_i, \bar{\boldsymbol{h}}_j)/\tau)} \tag{4}$$

通常情况下,基于自监督对比的事件检测的联合损失函数如式(5)所示,λ_1 为权重超参数。

$$L_{\text{ed\_uns}} = (1-\lambda_1)L_{\text{ed}} + \lambda_1 L_{\text{uns\_cl}} \tag{5}$$

但是需要大量的工作分析不同损失的量级、验证测试等才能设置合理的权重 λ_1。

为了降低手动设置联合学习损失函数权重的实验成本,文献[40]利用多任务内在的不确定性自动学习不同损失的权重。受此启发,本文在训练过程中自动调整损失函数的权重,减少了固定权重参数调优的过程,同时在事件检测的精度及模型收敛速度上都有所提升。

在动态调整损失函数权重的方式下,基于自监督对比的事件检测的联合损失函数如式(6)所示,其中 σ_1、σ_2 和 β_1 在训练过程中动态变化,σ_1 和 σ_2 决定损失函数的权重,β_1 为联合损失函数的正则项。

$$L_{\text{ed\_uns}} = \frac{1}{2\sigma_1}L_{\text{ed}} + \frac{1}{2\sigma_2}L_{\text{uns\_cl}} + \beta_1 \tag{6}$$

算法 2 为基于自监督对比的事件检测算法的流程。首先(第 3 ~7 行)对批数据中每条样本进行数据增强操作构建其正样例对,生成自监督对比学习的训练数据;然后(第 8 ~ 13 行)对自监督对比任务和事件检测任务联合学习。

算法 2 基于自监督对比的事件检测算法

输入：训练集 $\{x^{(n)}, y^{(n)}\}_{n=1}^{N}$，迭代次数 S

输出：更新后的模型参数 θ

```
for s = 1…S do
  从 {x^(n), y^(n)}_{n=1}^N 中选择一批数据 B = {x^(k), y^(k)}_{k=1}^K
  if 选择 mask 操作增强样例：
    对 B 中每条样本进行随机掩码操作，构建该样本的正样例对
  else 选择 dropout 操作增强样例：
    对 B 中每条样本在特征向量层 dropout 操作，构建该样本的正样例对
  生成自监督对比学习的一批训练数据 B_uns = {(x_i, x_j)^(k), y^(k)}_{k=1}^{K_uns}
  利用 B_uns 前向计算自监督对比损失 L_uns_cl
  利用 B 前向计算事件检测的交叉熵损失 L_ed
  计算多任务的联合损失 L_ed_uns
  计算损失函数的梯度 ∇_θ L_ed_uns
  利用梯度 ∇_θ L_ed_uns 计算模型新参数 θ′
  θ ← θ′
end for
```

2.4 基于监督对比的事件检测方法

在监督对比学习中，可以通过事件类型标签信息构造监督对比学习的正负样例。属于同一个事件类型的文本的语义比较相似，其向量的距离也更近，相反属于不同事件类型的文本表示的向量距离更远。因此，属于同一类别的文本为正样例，属于不同类别的文本为负样例。由于事件检测数据的稀疏，事件类型的繁多，每个批量数据中正样例的数量比较少。为了增加监督对比学习的正样例，本文同样采用原始文本 mask 操作和文本特征层 dropout 操作来增强样例，并赋予增强的样例和原文本相同的事件类别标签。

基于上述构造的正负样例，通过监督对比区分不同事件类型的文本语义空间[33]，拉近同类型文本在语义空间的距离，拉远不同类型文本在语义空间的距离，从而提高样本在事件语义空间表示能力。监督对比损失如式(7)所示。

$$L_{s\_cl} = -\frac{1}{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\mathbb{I}_{y_i=y_j}\log\frac{\exp(\mathrm{sim}(\boldsymbol{h}_i,\boldsymbol{h}_j)/\tau)}{\sum_{n=1}^{N}\mathbb{I}_{i\neq n}\exp(\mathrm{sim}(\boldsymbol{h}_i,\boldsymbol{h}_n)/\tau)} \tag{7}$$

其中，T 为每次迭代训练批量数据中属于同一类样本对的个数，$\mathbb{I}_{y_i=y_j}$ 表示第 i 个样本和第 j 个样本属于同一类事件。

基于监督对比的事件检测的联合损失函数如式(8)所示，λ_2 为权重超参数。

$$L_{ed\_s} = (1-\lambda_2)L_{ed} + \lambda_2 L_{s\_cl} \tag{8}$$

同样地，动态地调整事件检测的二进制交叉熵损失和监督对比损失的权重，基于监督对比的事件检测的联合损失如式(9)所示，权重参数 σ_3、σ_4 和正则项 β_2 在训练过程中自行调整。

$$L_{ed\_s} = \frac{1}{2\sigma_3}L_{ed} + \frac{1}{2\sigma_4}L_{s\_cl} + \beta_2 \tag{9}$$

算法 3 所示为基于监督对比学习的事件检测算法的流程。首先(第 3 ~ 7 行)对批数据中每条样本进行数据增强操作增加同类事件的样本数；然后(第 8 行)利用事件标签信息生成监督对比学习的训练数据；最后(第 9 ~ 14 行)对监督对比任务和事件检测任务联合学习。算法 3 和算法 2 的主要区别是数据增强的目的不同，算法 2 中是为了构建样本的正样例对，而算法 3 是为了增加同类事件的样本数。

算法 3 基于监督对比的事件检测算法

输入：训练集 $\{x^{(n)}, y^{(n)}\}_{n=1}^{N}$，迭代次数 S

输出：更新后的模型参数 θ

```
for s = 1…S do
  从 {x^(n), y^(n)}_{n=1}^N 中选择一批数据 B = {x^(k), y^(k)}_{k=1}^K
  if 选择 mask 操作增强样例：
    对 B 中每条样本进行随机掩码操作，增加同类型样本数量
  else 选择 dropout 操作增强样例：
    对 B 中每条样本在特征向量层 dropout 操作，在表示层增加同类型样本数量
  利用事件类型标签信息构建对比学习的正负样例
  生成监督对比学习的一批训练数据 B_s = {(x_i, x_j)^(k), y^(k)}_{k=1}^{K_s}
  利用 B_s 前向计算监督对比损失 L_s_cl
  利用 B 前向计算事件检测的交叉熵损失 L_ed
  计算多任务的联合损失 L_ed_s
  计算损失函数的梯度 ∇_θ L_ed_s
  利用梯度 ∇_θ L_ed_s 计算模型新参数 θ′
  θ ← θ′
end for
```

3　实验与分析

本节将介绍实验数据集、评价指标、实验设置、实验结果等内容。

3.1　实验数据集

本文在 ACE 2005 中英文数据集上进行了实验评估。中英文语料均包含 8 大类、33 个子类的事件样本,本文分析了细粒度划分的子类事件检测的效果。根据之前的工作[4,12,20,24],对英文数据集划分方式如下:先从不同类型文档中随机选择 30 篇文章作为验证集,然后随机选择 40 篇作为测试集,剩余的 529 篇文档作为训练集。同样方式对中文数据集划分:先从不同类型文档中随机选择 64 篇文章作为验证集,再随机选择 64 篇作为测试集,剩余的 521 篇文档作为训练集。表 1 所示为实验数据集包含文档、句子、事件个数的情况。

利用 Stanford CoreNLP 将每篇文档切分成句子,并根据 ACE 2005 原始语料的注释为每条句子分配标签。如果一个句子不包括任何事件类型,则分配一个“Negative”的标签。如果一个句子中包含多个不同类型的事件,则保留每个类型的标签;如果句子中包含多个同一类型的事件,则针对同一事件类型,只保留一个标签。在 ACE 语料中,这种情况占比不超过 3%[12]。表 2 所示为 ACE 中英文语料中无触发词注释的几个样例。

表 1　ACE 2005 中英文语料的文档、句子、事件统计结果

数据集	划分	文档数	句子数	事件数
ACE 2005 中文语料	Train	64	5483	2333
	Dev	64	665	299
	Test	521	672	303
ACE 2005 英文语料	Train	529	14 669	3903
	Dev	30	873	433
	Test	40	711	393

表 2　ACE 2005 中英文语料的无触发词注释的样例

数据集	样例句子	样例标签
ACE 2005 中文语料	目前,已有约 20 人在冲突中丧生	{Conflict, Die}
	两女站在门前欲将男子拉住,但男子却分别把她们咬伤及推跌,逃之夭夭。	{Injure}
	他好奇了,问道:那么其他呢	{Negative}
ACE 2005 英文语料	Liana Owen drove 10 hours from Pennsylvania to attend the rally in Manhattan with her parents	{Demonstrate, Transport}
	The boss fired his secretary today	{End-Position}
	But I want to ask you a question	{Negative}

3.2　评价指标

根据先前的工作[4,12,20,24],本文采用准确率 P、召回率 R 和 F_1 度量评估所提出事件检测算法的效果。

准确率 P 指在预测的事件中(不包括“Negative”类别)正确预测事件的概率,如式(10)所示。

$$P = \frac{\text{预测事件中预测标签等于真实标签的总个数}}{\text{测试集中所预测出事件的总个数}} \tag{10}$$

召回率 R 指在所有真实的事件中(不包括“Negative”类别)正确预测出的概率,如式(11)所示。

$$R = \frac{\text{预测事件中预测标签等于真实标签的总个数}}{\text{测试集中真实事件总个数}} \tag{11}$$

F_1 度量计算如式(12)所示。

$$F_1 = \frac{2 \times P \times R}{P + R} \tag{12}$$

3.3　实验设置

本文采用 BERT 模型的基础配置:隐含层的大小为 768、层数为 12。训练时采用 Adam 优化器,模型的超参数设置为:序列最大长度为 128、批量数据的大小为 32、学习速率为3×10^{-5},对比学习的温度系数 τ 在{0.1, 0.2, 0.3, 0.4, 0.5}集合中取值。

在 ACE 2005 英文语料的实验中,对比了同样无触发词的事件检测方法 TBNNAM\Bias[12]、TBNNAM[12],以及传统需要标注触发词的基于 BERT 编码器的事件检测方法 DYGIE ++[26]、BERT-CRF。

在 ACE 2005 中文语料的实验中,对比了传统需要标注触发词的基于 BERT 编码器的事件检测方法 MCEE[41]、JMCEE[41]、BERT-CRF[8]。

TBNNAM:基于注意力机制的类型感知偏差神经网络,利用 LSTM 编码字符的上下文,通过类型感知的注意力弥补触发词信息的缺失,并设计了偏差损失函数增强正样本的影响。TBNNAM\Bias 在计算损失函数时不使用偏差项。

DYGIE ++ :一个基于 BERT 编码器的多任务学习框架,可以捕获句子内和跨句子的上下文语义。DYGIE ++ 、BERTFinetune 是在事件检测任务上微调预训练 BERT 模型。

BERT-CRF:传统的序列标注的事件检测方法,文本字符输入 BERT 编码器输出的序列向量,通过 CRF 层对序列字符所属事件触发词识别和分类。

MCEE、JMCEE:基于 BERT 编码器的事件提取方法。MCEE 是管道式模型,先进行触发词的识别和分类,然后再进行事件元素的识别。JMCEE 是联合模型,同时预测文本的事件触发词和事件元素。本文对比了 MCEE、JMCEE 事件检测(触发词的识别和分类)的结果。

3.4 实验结果

3.4.1 事件检测结果的量化分析

表 3 和表 4 分别展示了不同方法在 ACE 2005 中英文数据集上的实验结果。第 1 组是对比的基准实验,第 2 组是本文提出的方法,包括只使用交叉熵损失 L_{ed} 的基础方法,使用交叉熵损失和对比损失联合训练增强句子语义的方法,以及预训练语言模型调优的基础上利用对比学习增强句子语义的方法。

从表 3 可以看出,在 ACE 2005 中文语料上,本文只使用 L_{ed} 的基础方法,优于传统的 MCEE(BERT-Pipeline)方法, F_1 值有 2.3% 的提升。在增加了对比损失后,和只使用 L_{ed} 的基础方法相比, F_1 值有 1.8% ~3.2% 的提升,甚至最高的 F_1 值超过了利用事件检测和事件元素提取多任务学习的方式增益事件识别效果的 JMCEE(BERT-Joint)。和传统 BERT-CRF 的事件检测方法相比,本文提出的只使用 L_{ed} 的基础方法和增加了对比表示的方法没有优势,主要原因是 BERT-CRF 将触发词识别和分类转化为序列标注任务,避免了触发词识别错误后传影响整体事件检测效果的问题。在预训练语言模型调优的基础上对比语义增强方法,和 JMCEE(BERT-Joint)相比, F_1 值提升 0.1% ~3.5%,和只使用 L_{ed} 的基础方法相比, F_1 值提升 3.1% ~6.5%,并在监督对比学习方法中达到和传统方法 BERT-CRF 相匹配的效果,甚至在利用 mask 操作增强样例的情况下取得更优的 F_1 值。分析其原因为:在无触发词的事件检测方法中,事件类型的识别主要依赖文本句子的语义表示;对比学习使不同语义的句子在向量空间具有更清晰的辨识度;预训练语言模型的调优,增强了句子字符级别事件信息的表示,提高了模型在下游任务的领域适应性。

表 3 在 ACE 2005 中文语料上不同方法的实验对比

	方法	P/%	R/%	F_1/%
基准	MCEE(BERT-Pipeline)	72.6	68.2	70.3
	JMCEE(BERT-Joint)	76.4	71.7	74.0
	BERT-CRF	77.4	74.2	75.8
本文方法	L_{ed}	72.1	71.6	71.9
	L_{ed_uns} (dropout)	72.1	74.3	73.2
	L_{ed_s} (dropout)	69.5	77.6	73.3
	L_{ed_uns} (mask)	71.7	75.3	73.4
	L_{ed_s} (mask)	73.5	74.9	**74.2**
	$L_{mlm} + L_{ed\_uns}$ (dropout)	76.9	72.6	74.7
	$L_{mlm} + L_{ed\_s}$ (dropout)	78.4	72.9	75.6
	$L_{mlm} + L_{ed\_uns}$ (mask)	69.9	78.9	74.1
	$L_{mlm} + L_{ed\_s}$ (mask)	74.5	78.9	**76.6**

从表 4 可以看出,在 ACE 2005 英文语料上,本文只使用 L_{ed} 的基础方法,优于同样无触发词的事件检测方法 TBNNAM\Bias, F_1 值提升 1.6%,但低于使用偏差损失函数增强正样本的 TBNNAM 方法。这说明事件检测任务的正负样本(包含事件信息的样本为正样本,根据表 1 占比不到 50%)的偏差对事件识别的效果影响很大,而本文的重点在于文本语义的表示学习,没有针对这一问题做优化。在增加了对比损失后,和只使用 L_{ed} 的基础方法相比, F_1 值有 2.0% ~4.0% 的提升,并达到了和 TBNNAM 相匹配的效果,最高的 F_1 值(基于 mask 操作增强样例的监督对比学习)提升 1.6%。和 DYGIE ++ 、BERT Finetune 相比,本文只使用 L_{ed} 的基础方法略差,主

要因为本文采用的方法缺失触发词信息,在增加了对比损失后,F_1 值提升 0.0% ~1.9%,表明增强句子语义的方法可以提高触发词缺失而下降的效果。在预训练语言模型调优的基础上使用对比语义增强的方法,和只使用 L_{ed} 的基础方法相比,F_1 值有 4.2% ~ 5.9% 的提升,并达到和传统方法 BERT-CRF 相匹配的效果,甚至在利用 mask 操作增强样例的监督对比学习方法中取得更优的 F_1 值。

表 4　在 ACE 2005 英文语料上不同方法的实验对比

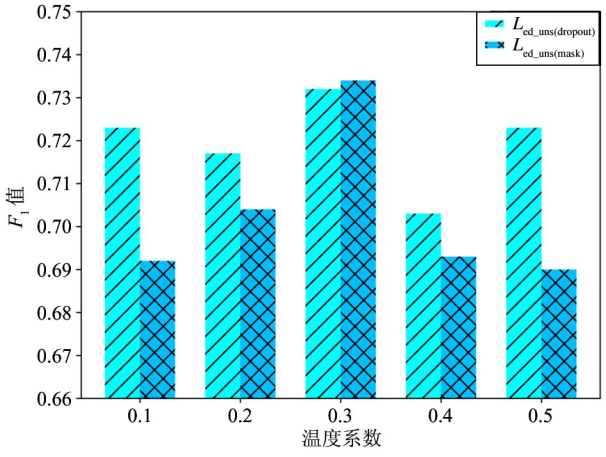
	方法	P/%	R/%	F_1/%
基准	TBNNAM\Bias	76.6	59.8	67.2
	TBNNAM	76.2	64.5	69.9
	DYGIE ++,BERT Finetune	-	-	69.7
	BERT-CRF	70.4	73.3	71.8
本文方法	L_{ed}	67.8	68.7	68.3
	L_{ed_uns} (dropout)	69.2	70.2	69.7
	L_{ed_s} (dropout)	70.7	70.5	70.6
	L_{ed_uns} (mask)	68.2	72.0	70.1
	L_{ed_s} (mask)	67.4	75.1	**71.0**
	$L_{mlm} + L_{ed\_uns}$ (dropout)	72.6	70.0	71.2
	$L_{mlm} + L_{ed\_s}$ (dropout)	69.8	74.1	71.9
	$L_{mlm} + L_{ed\_uns}$ (mask)	73.2	70.2	71.7
	$L_{mlm} + L_{ed\_s}$ (mask)	71.6	73.0	**72.3**

从表 3 和表 4 还发现:同自监督对比学习相比,利用标签信息的监督对比学习的事件检测效果更好。一方面是因为监督对比将同一类别的样本作为正样例,相比于将同源样本作为正样例,更能在语义空间区分不同事件类型的语义距离。另一方面是因为监督对比采用数据增强的方式增加了每次迭代训练中的正样例数量。此外,不管是监督对比还是自监督对比,利用 mask 操作的数据增强具有更好的检测效果,特别是基于 mask 操作增强样例的监督对比学习在中英文语料上都取得了相对较优的 F_1 值。

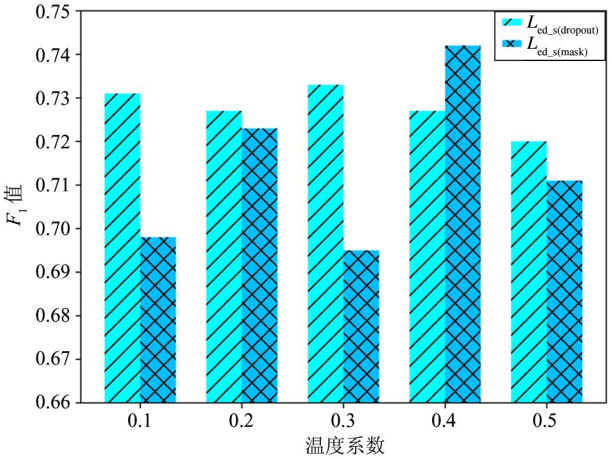
3.4.2　对比学习温度系数的影响分析

温度系数 τ 是对比损失中一个重要的参数,控制着模型对负样本的区分度,直接影响模型的效果。本文实验分析了在{0.1, 0.2, 0.3, 0.4, 0.5}中取不同的值对事件检测效果的影响。由于更小的值如 0.07 在利用 dropout 操作增强样例的自监督对比学习中收敛速度较慢,不在本文对比分析的范围。

图 5 所示为在 ACE 2005 中文语料上基于对比学习事件检测的 F_1 值随温度系数 τ 变化情况,其中图 5(a)和图(b)分别为温度系数取不同值时自监督对比学习和监督对比学习的事件检测方法的效果。同样,本文也测试了在 ACE 2005 英文语料上基于对比学习事件检测的 F_1 值随温度系数 τ 变化情况,如图 6 所示。从图中可以看到,温度系数的选择受数据增强方式、对比学习方式、数据集特点等多种因素的影响。在中文语料上,利用 dropout 操作增强样例的 2 种对比学习方法的温度系数倾向于取值 0.3,而利用 mask 操作增强样例的 2 种对比学习方法的最优温度系数分别为 0.3、0.4。在英文语料上,利用 dropout 操作增强样例的 2 种对比学习方法

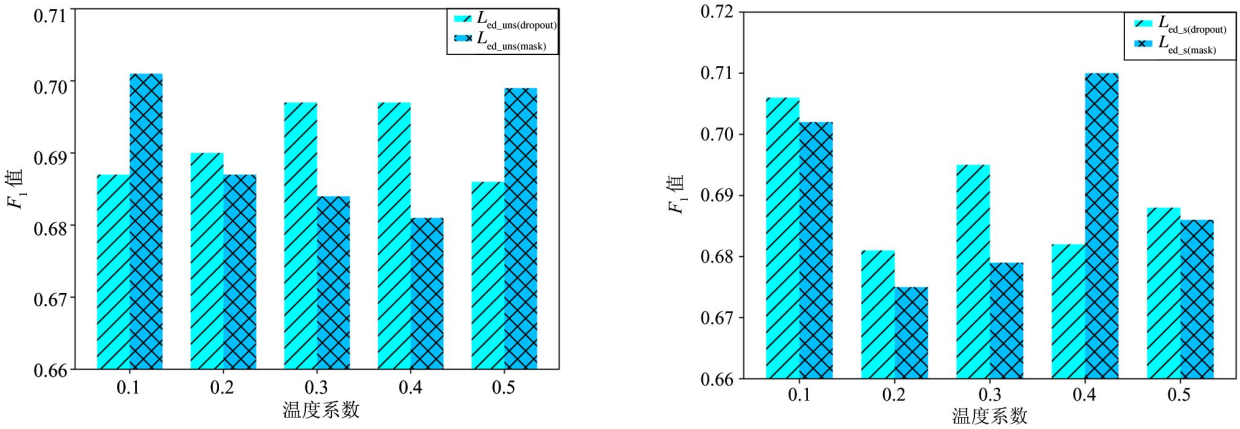


(a)基于自监督对比学习的事件检测方法



(b)基于监督对比学习的事件检测方法

图 5　在 ACE 2005 中文语料上事件检测 F_1 随温度系数变化情况

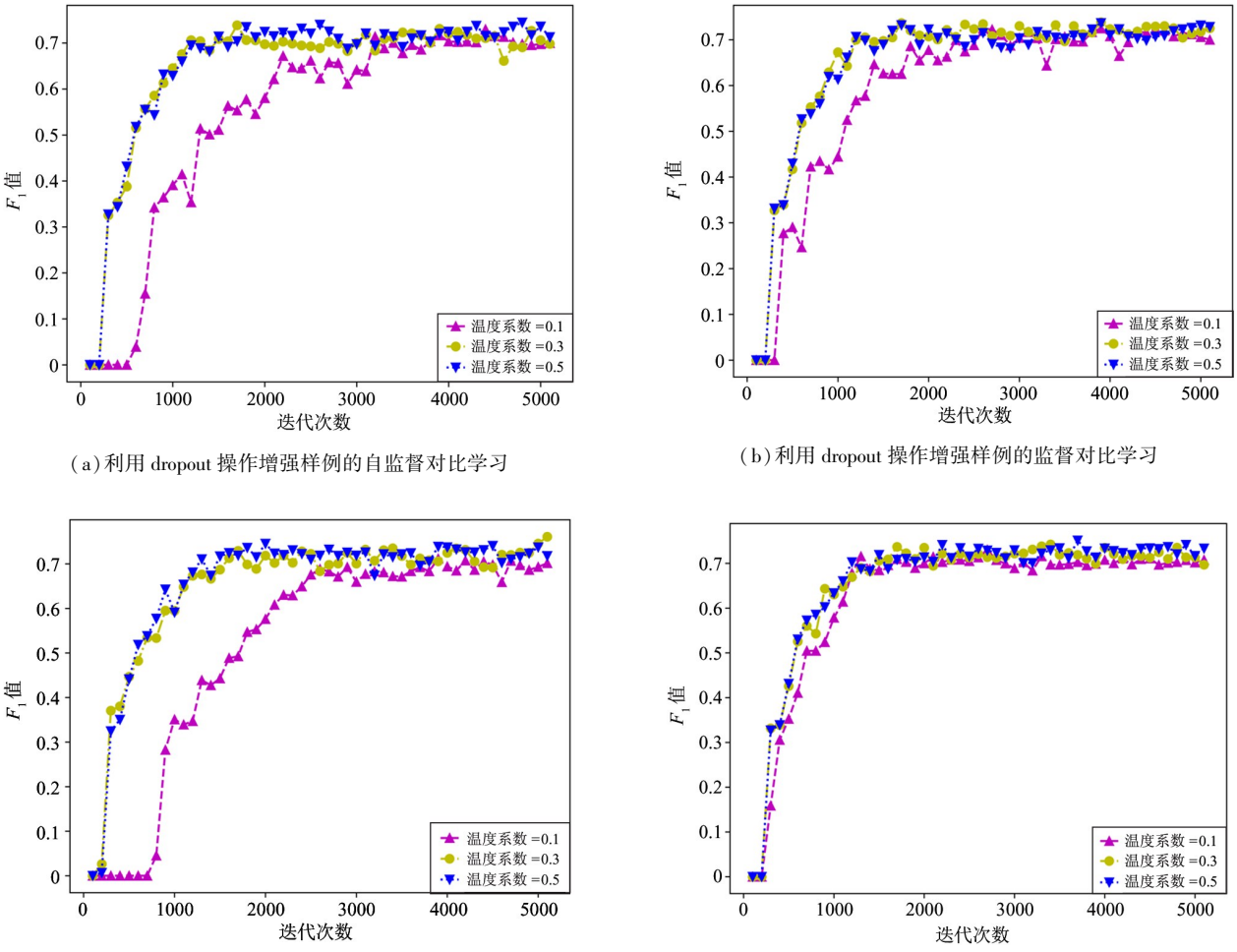


(a)基于自监督对比学习的事件检测方法　　(b)基于监督对比学习的事件检测方法

图 6　在 ACE 2005 英文语料上事件检测 F_1 随温度系数变化情况

的温度系数倾向于取值 0.4(或 0.3)、0.1,而利用 mask 操作增强样例的 2 种对比学习方法的最优温度系数分别为 0.1、0.4。

本文进一步分析了不同取值的温度系数对模型收敛速度的影响。图 7 所示为在 ACE 2005 中文语料上温度系数对不同对比学习方法收敛速度的影



(a)利用 dropout 操作增强样例的自监督对比学习　　(b)利用 dropout 操作增强样例的监督对比学习

(c)利用 mask 操作增强样例的自监督对比学习　　(d)利用 mask 操作增强样例的监督对比学习

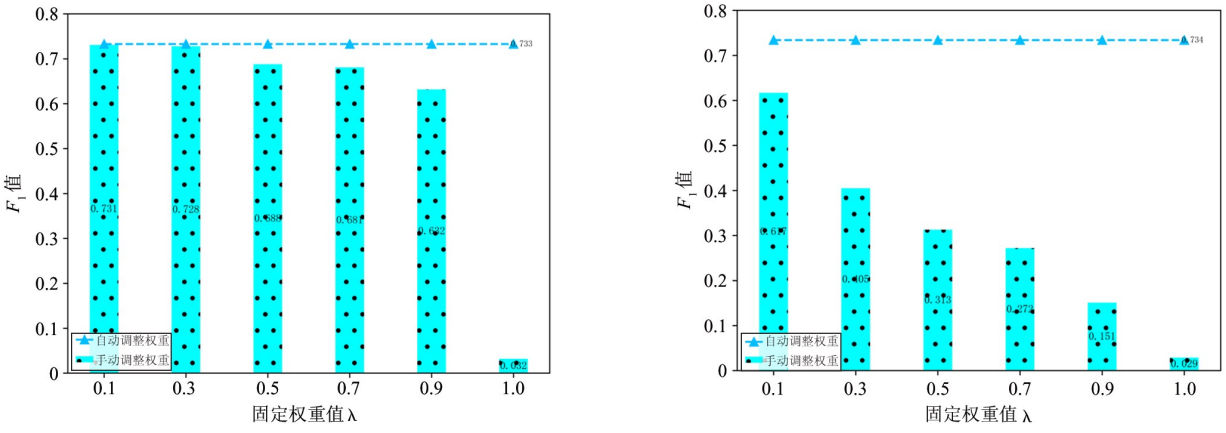
图 7　在 ACE 2005 中文语料上温度系数对不同对比学习方法收敛速度的影响

响。从图中可以看到,不管是 dropout 操作还是 mask 操作的数据增强方式,相对于监督对比学习,自监督对比学习的收敛速度更容易受温度系数的影响,比如温度系数取值 0.1 时和较大的取值(如 0.3、0.5)相比,自监督对比学习的收敛速度明显慢很多。值得注意的是,和利用 dropout 操作增强样例的对比学习方法相比,利用 mask 操作增强样例的对比学习更容易收敛。以温度系数取值 0.1 为例说明如下:在自监督对比学习中,利用 dropout 操作增强样例的方式大概需要 3000 多次迭代才能收敛,而利用 mask 操作增强样例的方式大概需要迭代 2500 次;在监督对比学习中,利用 dropout 操作增强样例的方式大概需要迭代 2700 次才能收敛,而利用 mask 操作增强样例的方式仅需要迭代 1000 多次。

3.4.3　自动调整权重的效果分析

本文在训练过程中自动调整对比损失函数和事件检测的交叉熵损失函数的权重,和手动设置权重的方式相比,自动调整权重能够快速地达到较优的效果。

图 8 所示为手动调整权重和自动调整权重的效果(F_1 值)对比,柱状形为使用式(5)或(6)手动调整损失函数权重的效果,虚线表示使用式(8)或(9)自动调整损失函数权重的效果。手动调整损失函数权重时,λ 取值为{0.1, 0.3, 0.5, 0.7, 0.9, 1.0}。测试方法是在 ACE 2005 中文语料上利用 dropout 操作增强样例的监督对比学习和利用 mask 操作增强样例的自监督对比学习 2 种方法,实验显示的结果是训练 30 个 epoch 在测试集上获得的 F_1 值。从图中可以看到手动调整权重时 λ 取值 0.1 最优,这对于有经验的专业人员,可能会快速地找到 λ 的最优值为 0.1,但对于经验不足的调参人员可能需要多次的实验才能找到最优的权重分配,而动态调整权重的方式不需要如此繁琐的调参过程就能获得较优的 F_1 值。



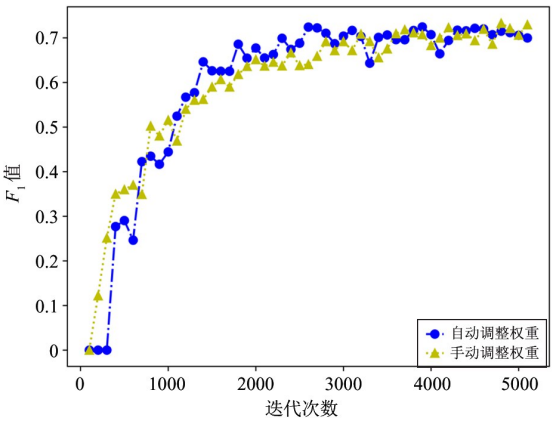
(a)利用 dropout 操作增强样例的监督对比学习　　(b)利用 mask 操作增强样例的自监督对比学习

图 8　手动调整权重和自动调整权重的效果(F_1 值)对比

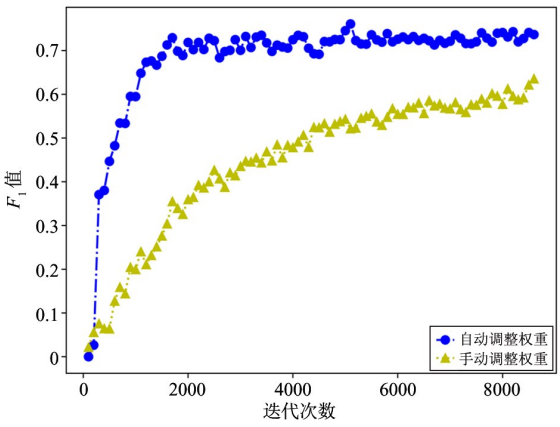
图 9 对比了手动调整权重和自动调整权重的收敛速度。手动调整权重的 λ 取值0.1,曲线显示的是在验证集上 F_1 值的变化情况。从图中可以看到,利用 dropout 操作增强样例的监督对比学习,自动调整权重和手动调整权重收敛速度相差不大,经过 3000 多次的迭代训练均能收敛;利用 mask 操作增强样例的自监督对比学习,自动调整权重的方式收敛速度更快,大概需要 2000 多次迭代训练,而手动调整权重的方式经过 8000 多次迭代训练还不能达到收敛状态。

4　结　论

面对事件检测数据稀疏、标注昂贵的问题,本文探索了无触发词事件检测的语义提取方法,介绍了基于对比学习增强句子语义的事件检测方法。该方法在事件检测数据集上通过自监督学习对预训练的语言模型 BERT 调优,并利用 mask 操作和 dropout 操作构建自监督对比样例,增加监督对比样例,实现了自监督对比和监督对比 2 种句子语义增强的方法,同时在训练过程中自动调整对比损失和事件分



(a)利用 dropout 操作增强样例的监督对比学习



(b)利用 mask 操作增强样例的自监督对比学习

图 9　手动调整权重和自动调整权重的收敛速度对比

类的交叉熵损失的权重。在 ACE 2005 中英文语料上的实验结果表明,本文提出的方法和只使用交叉熵损失的方法相比,F_1 值在中文语料上有 3.1% ~ 6.5% 的提升,在英文语料上有 4.2% ~5.9% 的提升;相较于基准方法,F_1 值也具有明显的优势。在模型收敛速度方面,自监督对比学习比监督对比学习更容易受温度系数的影响,相同的对比学习方法中利用 mask 操作增强样例的方式更容易收敛,特别是对于取值较低的温度系数。此外,本文采用的动态调整损失函数权重的方法,能够降低人工调参成本,同时更快地达到较优的结果。未来工作中,本研究将进一步探讨无触发词事件检测相关技术以及事件元素的提取。

参考文献

[1] NGUYEN T H, GRISHMAN R. Event detection and domain adaptation with 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: Association for Computational Linguistics, 2015: 365-371.

[2] GHAEINI R, FERN X L, HUANG L, et al. Event nugget detection with forward-backward recurrent neural networks[C]// Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics. Berlin: Association for Computational Linguistics, 2016: 369-373.

[3] FENG X C, HUANG L F, TANG D Y, et al. A language-independent neural network for event detection[C]// Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics. Berlin: Association for Computational Linguistics, 2016:66-71.

[4] LI Q, JI H, HUANG L. Joint event extraction via structured prediction with global features[C]// Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics. Sofia: Association for Computational Linguistics, 2013:73-82.

[5] 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: Association for Computational Linguistics, 2016:300-309.

[6] NGUYEN T H, GRISHMAN R. Graph convolution networks with argument-aware pooling for event detection [C]// Proceedings of the 32th AAAI Conference on Artificial Intelligence and 30th Innovative Applications of Artificial Intelligence Conference. New Orleans: AAAI Press, 2018:5900-5907.

[7] JI Y Z, LIN Y F, GAO J W, et al. Exploiting the entity type sequence to benefit event detection[C]// Proceedings of the 23rd Conference on Computational Natural Language Learning. Hong Kong: Association for Computational Linguistics, 2019:613-623.

[8] 田梓函,李欣. 基于 BERT-CRF 模型的中文事件检测方法研究[J]. 计算机工程与应用,2021,57(11):135-139.

[9] GEORGE D, ALEXIS M, MARK P, et al. The automatic content extraction (ACE) program-task, data, and evaluation[C]// Proceedings of the 2004 International Conference on Language Resources and Evaluation. Lisbon: Association for Computational Linguistics, 2004:837-840.

[10] LIU J, CHEN Y B, LIU K, et al. How does context matter? on the robustness of event detection with context-selective mask generalization[C]// Findings of the Associa-

tion for Computational Linguistics: EMNLP. Virtual: Association for Computational Linguistics, 2020:2523-2532.
[11] LIU J, CHEN Y F, XU J. Saliency as evidence: event detection with trigger saliency attribution[C]//Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics. Dublin: Association for Computational Linguistics, 2022:4573-4585.
[12] LIU S L, LI Y, ZHANG F, et al. Event detection without triggers[C]//Proceedings of NAACL-HLT 2019. Minneapolis: Association for Computational Linguistics, 2019: 735-744.
[13] ZHENG S, CAO W, XU W, et al. Doc2EDAG: an end-to-end document-level framework for Chinese financial event extraction[C]//Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing. Hong Kong: Association for Computational Linguistics, 2019:337-346.
[14] 王雷,李瑞轩,李玉华,等. 文档级无触发词事件抽取联合模型[J]. 计算机科学与探索,2021,15(12):2327-2334.
[15] GAO T Y, YAO X C, CHEN D Q. SimCSE: simple contrastive learning of sentence embeddings[C]//Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing. Punta Cana: Association for Computational Linguistics, 2021: 6894-6910.
[16] AHN D. The stages of event extraction[C]//Proceedings of the Workshop on Annotating and Reasoning about Time and Events. Sydney: Association for Computational Linguistics, 2006:1-8.
[17] JI H, GRISHMAN R. Refining event extraction through cross-document inference[C]// Proceedings of ACL-08: HLT. Columbus: Association for Computational Linguistics, 2008:254-262.
[18] LIAO S S, GRISHMAN R. Using document level cross-event inference to improve event extraction[C]//Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics. Uppsala: Association for Computational Linguistics, 2010:789-797.
[19] HONG Y, ZHANG J F, MA B, et al. Using cross-entity inference to improve event extraction[C]//Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies. Portland: Association for Computational Linguistics, 2011: 1127-1136.
[20] CHEN Y B, XU L H, 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: Association for Computational Linguistics, 2015: 167-176.
[21] 王凯,洪宇,邱盈盈,等. 融合上下文依赖和句子语义的事件线索检测研究[J]. 计算机科学与探索,2018,12(3):423-431.
[22] 马晨曦,陈兴蜀,王文贤,等. 基于递归神经网络的中文事件检测[J]. 信息网络安全,2018(5):75-81.
[23] 苗佳,段跃兴,张月琴,等. 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.
[24] LIU S L, CHEN Y B, 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. Vancouver: Association for Computational Linguistics, 2017:1789-1798.
[25] LAI V D, NGUYEN T N, NGUYEN T H. Event detection: gate diversity and syntactic importance scores for graph convolution neural networks[C]//Proceedings of the Conference on Empirical Methods in Natural Language Processing. Virtual: EMNLP, 2020:5405-5411.
[26] WADDEN D, WENNBERG U, LUAN Y, et al. Entity, relation, and event extraction with contextualized span representations[C]//Proceedings of the Conference on Empirical Methods in Natural Language Processing and the International Joint Conference on Natural Language Processing. Hong Kong: Association for Computational Linguistics, 2019:5784-5789.
[27] DU X Y, CARDIE C. Document-level event role filler extraction using multi-granularity contextualized encoding[C]//Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. Virtual: Association for Computational Linguistics, 2020:8010-8020.
[28] 朱培培,王中卿,李寿山,等. 基于篇章信息和Bi-GRU的中文事件检测[J]. 计算机科学,2020,47(12):233-238.
[29] PETERS M E, NEUMANN M, IYYER M, et al. Deep contextualized word representations[C]//Proceedings of the Conference of the North American Chapter of Association for Computational Linguistics: Human Language Technologies. New Orleans: Association for Computational Linguistics, 2018:2227-2237.
[30] RADFORD A, NARASIMHAN K, SALIMANS T, et al. Improving language understanding by generative pre-training[EB/OL]. [2022-02-16]. https://www.cs.ubc.ca/~amuham01/LING530/papers/radford2018improving.pdf.
[31] DEVLIN J, CHANG M W, LEE K, et al. BERT: pre-training of deep bidirectional transformers for language understanding[C]//Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Minneapolis: Association for Computational Linguistics,

2019:4171-4186.
[32] RADFORD A, WU J, CHILD R, et al. Language models are unsupervised multitask learners[J]. OpenAI Blog, 2019,1(8):9-32.
[33] GUNEL B, DU J, CONNEAU A, et al. Supervised contrastive learning for pre-trained language model fine-tuning [EB/OL]. (2020-11-03)[2022-02-16]. http://arxiv.org/pdf/2011.01403.pdf.
[34] LIU F Y, VULIĆ I, KORHONEN A, et al. Fast, effective and self-supervised: transforming masked language models into universal lexical and sentence encoders[EB/OL]. (2021-04-16)[2022-02-16]. http://arxiv.org/pdf/2104.08027.pdf.
[35] SRIVASTAVA N, HINTON G, KRIZHEVSKY A, et al. Dropout: a simple way to prevent neural networks from overfitting[J]. The Journal of Machine Learning Research, 2014,15(1):1929-1958.
[36] ZHENG J M, CAI F, CHEN W Y, et al. Taxonomy-aware learning for few-shot event detection[C]// Proceedings of the Web Conference 2021. Ljubljana: Association foe Computing Machinery, 2021:3546-3557.
[37] CAO Y W, PENG H, WU J, et al. Knowledge-preserving incremental social event detection via heterogeneous GNNs[C]// Proceedings of the Web Conference 2021. Ljubljana: Association foe Computing Machinery, 2021: 3383-3395.
[38] SCHROFF F, KALENICHENKO D, PHILBIN J. FaceNet: a unified embedding for face recognition and clustering[C]// Proceeding of the Computer Vision and Pattern Recognition Conference. Boston: IEEE, 2015:815-823.
[39] WANG Z Q, WANG X Z, HAN X, et al. CLEVE:contrastive pre-training for 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. Virtual: Association for Computational Linguistics, 2021:6283-6297.
[40] KENDALL A, GAL Y, CIPOLLA R. Multi-task learning using uncertainty to weigh losses for scene geometry and semantics[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018:7482-7491.
[41] XU N, XIE H H, ZHAO D Y. Anovel joint framework for multiple Chinese event extraction[C]// Proceeding of the 19th China National Conference on Computational Linguistics. Haikou: Technical Committee on Computational Linguistics, Chinese Information Processing Society of China, 2020:950-961.

Event detection based on enhanced sentence semantics via contrastive learning

LIANG Dong[*,**], ZHANG Cheng[*], SHI Xiao[*], TAN Wenting[*,**], LV Cunchi[*,**], ZHAO Xiaofang[*,**,***]
([*] Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190)
([**] University of Chinese Academy of Sciences, Beijing 100049)
([***] Institute of Intelligent Computing Technology, Suzhou, CAS, Suzhou 215028)

Abstract

Event detection aims to find the events and their types mentioned in the text. Previous work based on triggers requires additional labor to label event triggers. Considering that event detection mainly relies on the text semantics without triggers, an event detection method based on enhanced sentence semantics via contrastive learning is proposed. First, the pre-trained language model biderectional encoder representations from transformers (BERT) is fine-tuned with self-supervised learning for better domain adaptability. Then, a novel event detection method based on sentence semantic enhancement through contrastive learning is presented, using mask operation or dropout operation to construct the self-supervised contrastive samples and increase the supervised contrastive samples. Furthermore, the weight assignment between the cross-entropy loss and the contrastive loss is automatically adjusted during training to reduce the cost of manual parameter tuning and improve the convergence speed. Experimental results on automatic context extraction (ACE) 2005 Chinese and English corpus show that the proposed method is superior to the previous event detection methods without triggers, and outperforms the methods with triggers extracted by the pre-trained BERT model fine-tuning.

Key words: event detection, self-supervised contrastive learning, supervised contrastive learning, semantic enhancement, automatic weighting