

基于Electra预训练模型并融合依存关系的中文事件检测模型

尹宝生, 孔维一

引用本文

尹宝生, 孔维一. 基于Electra预训练模型并融合依存关系的中文事件检测模型[J]. 计算机科学, 2024, 51(6A): 230600158-6.

YIN Baosheng, KONG Weiyi. [Electra Based Chinese Event Detection Model with Dependency Syntax Tree](#) [J]. Computer Science, 2024, 51(6A): 230600158-6.

相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[融合标签知识的中文医学命名实体识别](#)

Chinese Medical Named Entity Recognition with Label Knowledge

计算机科学, 2024, 51(6A): 230500203-7. <https://doi.org/10.11896/jsjcx.230500203>

[基于多尺度注意力的遥感影像建筑物提取研究](#)

Study on Building Extraction from Remote Sensing Image Based on Multi-scale Attention

计算机科学, 2024, 51(5): 134-142. <https://doi.org/10.11896/jsjcx.230200134>

[基于动态负采样的图卷积协同过滤推荐模型](#)

Dynamic Negative Sampling for Graph Convolution Network Based Collaborative Filtering Recommendation Model

计算机科学, 2023, 50(11A): 230200149-7. <https://doi.org/10.11896/jsjcx.230200149>

[基于BERT和多特征门控机制的口语理解联合方法](#)

Joint Method for Spoken Language Understanding Based on BERT and Multiple Feature Gate Mechanism

计算机科学, 2023, 50(11A): 230300002-6. <https://doi.org/10.11896/jsjcx.230300002>

[基于核心句的端到端事件共指消解](#)

End-to-End Event Coreference Resolution Based on Core Sentence

计算机科学, 2023, 50(11): 185-191. <https://doi.org/10.11896/jsjcx.221000078>

基于 Electra 预训练模型并融合依存关系的中文事件检测模型

尹宝生 孔维一

沈阳航空航天大学人机智能研究中心 沈阳 110136

(54951941@qq.com)

摘要 事件检测是信息提取领域的一个重要研究方向。现存的事件检测模型受到语言模型训练目标的限制,只能被动地获取词与词之间的依赖关系,使得模型在训练的过程中过多地关注与训练目标不相关的成分,从而导致检测结果错误。以往的研究表明,充分理解上下文信息对于基于深度学习的事件检测技术至关重要。因此,在 Electra 预训练模型的基础上,引入 KVMN 网络来捕捉单词之间的依赖关系,以增强单词的语义特征,并采用了一种门控机制来加权这些特征。然后,为了解决中文事件检测中模型识别错误决策的问题,在输入中加入负样本,对不同样本加入不同程度的噪声,使模型学习更好的嵌入表示,有效提高了模型对未知样本的泛化能力。最后,在公共数据集 LEVEN 上的实验结果表明,该方法优于现有方法,取得了 93.43% 的 F1 值。

关键词 事件检测;依存关系;键值记忆网络;门控机制;负采样

中图分类号 TP391

Electra Based Chinese Event Detection Model with Dependency Syntax Tree

YIN Baosheng and KONG Weiyi

Human-Machine Intelligence Research Center, Shenyang Aerospace University, Shenyang 110136, China

Abstract Event detection is an important research direction in the field of information extraction. The existing event detection models are limited by the training targets of language models, and the dependency relationship between words can only be acquired passively, so the models pay more attention to the unrelated components during training, resulting in the wrong detection results. Previous studies show that fully understanding contextual information is crucial for deep learning-based event detection techniques. In this paper, we introduce the KVMN network to capture the dependencies between words and enhance the semantic features of words, and a gating mechanism is adapted to weight these features. Then, in order to solve the problem of the model's identification of wrong decisions, negative samples are added to the input, and different levels of noise are added for different samples, so that the model could learn a better embedding representation, effectively improving the model's ability to generalise unknown samples. Finally, experimental results on the public dataset LEVEN show that this method is superior to the existing methods and achieves a F1 score of 93.43%.

Keywords Event detection, Dependency, Key-value memory network, Gating mechanism, Negative sampling

1 引言

事件检测是事件抽取的一个重要子任务,它直接影响事件元素提取的效果^[1],其目标是从文本中自动提取事件触发词,然后将其对应的事件类型进行分类。经过事件检测识别出来的事件可以找出发生的事件及其之间的因果关系,用于辅助构建事理图谱、信息检索以及生成事件脉络等^[2-3]。

当前事件检测面临以下几个方面的挑战:(1)目前的事件检测模型只关注句子级的上下文,而事件触发词经常分散在不同的句子中,模型只能关注句子层面的语义,忽略了上下文信息,导致模型检测准确性和收敛性均欠佳^[4];(2)一段文本可能会出现多个相同或不同类型的事件,即多事件问题;(3)随着深度神经网络的规模和复杂性的增长,保证模型不会学习到训练集中的虚假样例正成为一个需要解决的问题^[5]。Wang

等^[6]提出了模型的辨别问题即混淆了负样本和正样本。结果显示,假正例和假负例的误差分别为 48.99% 和 34.41%。

针对以上难点,本文的工作主要体现在两个方面:(1)构建一种利用 Electra 作为语义特征提取器的事件检测模型,针对辨别问题,在输入中添加负样例,选择其中蕴含更多语义知识的负样例扩充训练样本,增强模型对关键信息的关注,提高模型对未知样本的泛化能力,增强模型的鲁棒性;(2)利用键值记忆网络捕获单词之间的依存关系,利用依存句法信息为模型提供句法知识,作为上下文信息的补充,将其与编码后的语义信息采用门控机制进行相加,融入了依存关系后让模型更好地利用不同的词语之间的语义相关性,提高事件检测的性能。

2 相关工作

事件检测任务的发展过程从基于深度学习的方法,发展

基金项目:辽宁省教育厅项目(LJKMZ20220536)

This work was supported by the Project of Liaoning Provincial Department of Education(LJKMZ20220536).

通信作者:孔维一(1120444186@qq.com)

到现在采用 NLP 第四范式的方法进行事件检测。

2.1 深度学习的方法

随着深度学习的发展,越来越多的预训练模型被用于事件检测任务中。Gu 等^[7]采用多任务学习的方法,引入词语歧义任务来提升事件检测模型的联合学习能力,两个任务可以相互学习,进而增强事件检测的能力。Liu 等^[8]从数据端进行一个多语言的联合学习,之后对中英文进行联合事件检测,通过把一个语言翻译成另一语言,较好地利用了不同语言之间的互补关系,数据得以加倍,解决了数据稀疏问题,但增加了模型的训练成本。

Liu 等^[9]将图像信息与文本级别上的词和短语进行对齐,从而丰富文本模态的语义,帮助模型对词进行消歧,虽然加入了图片信息之后,能够识别出一些分类上的错误,但若 not 结合图片信息,模型难以识别出某类事件。且不同领域和任务需要不同的图片信息,重复定义工作量大。Luo 等^[10]利用知识增强的方法,针对数据集中各类事件分布不均衡的缺点,从区分度信息和泛化度信息的角度进行了一个对抗的训练,使得模型不会过于依赖触发词本身进行分类,而是必须依赖周边环境进行分类。Tong 等^[11]通过引入开放域知识,实现了对于长尾的极少出现的触发词的性能提升。

2.2 预训练和提示学习的方法

NLP 第四范式的到来,使得越来越多的学者将该预训练和提示学习应用到事件检测任务中。Lu 等^[12]针对目前大多数事件检测方法严重依赖训练实例,几乎忽略了事件类型的相关性,往往会遭受数据短缺的困扰,并且无法辨识新的事件类型等问题,将事件检测描述为一个事件本体的生成过程,将事件实例与事件本体中预定义的事件类型连接起来,利用事件类型之间的关联来丰富事件本体,并进一步诱发更多的事件-事件关联。Tong 等^[13]将事件抽取定义为问题抽取范式,利用事件类型和论元角色之间的语义,提高事件抽取的性能。Nguyen 等^[14]将事件抽取任务作为生成式任务,采用事件类型作为动态前缀,来解决生成方法存在的两个重大挑战:使用次优提示和静态事件类型信息。Chen 等^[15]将提示学习应用到事件抽取任务中,模板包含了实体类型的描述信息、事件关键词信息,以及事件模板信息。DMBERT 提取序列的特征,并使用动态池化层来获得每个候选对象特定于触发器的表示^[16],相比 DMCNN 在 F1 值上提升了 2%,但该模型并没有考虑到词间依赖信息。本文引用 DMBERT 模型以及融合特征的方法,利用键值记忆网络捕获单词之间的依存关系,并将依存句法信息作为上下文信息的补充,将其与编码后的语义信息采用门控机制进行相加。

3 融合依存关系的法律事件检测模型

本文在现存事件检测的研究基础上提出了基于 Electra 预训练模型并融合依存关系的中文法律事件检测模型 EWD-ED(Event With Dependency Event Detection)。

事件检测通常分为两个子任务:触发词识别和触发词分类。本文提出的事件检测模型如图 1 所示,主要由以下 4 个部分构成。

句子编码层:采用预训练语言模型 Electra 作为语义特征提取器,对包含正样本和负样本的输入进行编码,得到句子的

语义特征向量表示。

依存语义图交互层:对句子进行依存句法分析^[17],获得根据句子构建的语义图的结构特征,将其作为键值记忆网络的输入。

门控融合层:首先对融入了依存关系的句子的表示进行池化操作,获得统一长度的句子表示,之后将句子编码层句子的嵌入表示与包含依存关系的句子嵌入表示进行相加,作为事件检测任务的特征输入。

事件类型分类层:判断句子中的每个词是否为触发词,以及其对应的事件类型。

3.1 句子编码层

本文采用哈工大讯飞联合实验室发布的中文的、基于大规模法律文本训练出的 ELECTRA 预训练模型。Yao 等^[23]的研究指出,一个好的模型应该能够通过逻辑回归来区分数据和噪声。在 Li 等^[19]提出的在微调前给预训练模型参数增加噪音,提高预训练语言模型在下游任务的效果的基础上,扩充训练样本,为正样本与负样本采用不同程度的噪音,且将使用词之间关系作为可见矩阵限制负样例注入对输入文本的影响。其中负样例的定义如式(1)所示,CEL 为一致性增强损失, w_p 是每个正确的可分类的触发词, w_i 求和符号内的 W_i 为每个不可分类(none)的触发词, E 表示期望,即从伯努利分布 $P(W)$ 中随机抽样 k 个 w_i , 表示整体的不可分类触发词。Devlin 等^[20]提出可以近似地最大化 softmax 的对数概率,使模型学习到高质量的向量表示。

$$CEL(W_p) + \sum_{i=1}^k E_{w_i \sim P(W)} [CEL(w_i)] \quad (1)$$

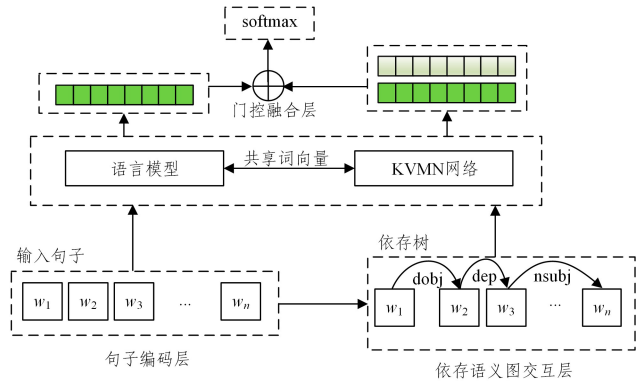


图 1 融合依存关系的事件检测模型图

Fig. 1 Diagram of event detection model with dependencies

对于给定的一篇文章 D , 可以用句子序列 $D = \{S_1, S_2, \dots, S_k\}$ 表示, 负样本也包含在集合 D 中。为了保持句子之间的连续性, 输入文档的句子开头和结尾用语言模型的特殊符号 $[CLS]$ 标记, 两个句子之间使用特殊符号 $[SEP]$ 进行分割, $D = \{[CLS], S_1, S_2, [SEP], \dots, S_k\}$, 其中 k 表示文章中句子的个数。每个句子经过 Electra 编码后的最后一层隐层状态作为句子的词向量表示, 可以用序列 $\{c_{i,1}, c_{i,2}, \dots, c_{i,N_i}\}$ 表示, 其中 c_{i,N_i} 为第 i 个句子中第 N_i 个词对应的向量, 维度为 d 。编码过程如图 2 所示, 针对模型难以覆盖句子较长或篇章级文本, 本文对编码后的词向量采用动态池化操作。对于给定的一个句子 S , c_i 表示文章中的第 i 个词, 用如下公式对 $\{c_1, c_2, \dots, c_i\}$ 进行动态池化操作, 得到新的句子特征向量表示 L_d 。本文进行的动态多池化如式(2)所示:

$$L_d = \max\{c_1, c_2, \dots, c_i\} \quad (2)$$

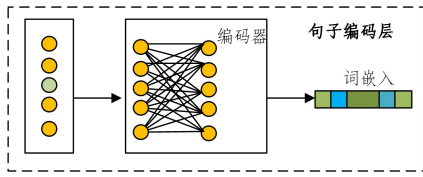


图2 句子编码图

Fig. 2 Sentence encoding diagram

3.2 依存语义图交互层

StanfordCoreNLP^[21]是斯坦福提供的一系列自然语言处理工具,提供了分词、词性标注、命名实体识别、依存句法分析、语义角色标注等丰富、高效、精准的自然语言处理技术。

StanfordCoreNLP 的依存句法分析工具有两种准确率,一种是仅考虑箭头是否指向了正确的词,而不考虑箭头的标签是否正确,这种准确率称为 UAS(Unlabeled Accuracy Score)。另一种是既考虑箭头是否正确,又考虑箭头的标签是否正确,这种准确率称为 LAS(Labeled Accuracy Score),StanfordCoreNLP 的依存句法分析工具的 UAS 可达到92.3, LAS 可达到 89.7。图 3 为使用 StanfordCoreNLP 进行依存句法分析的例子。

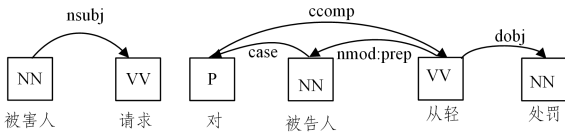


图3 StanfordCoreNLP 语义图示例

Fig. 3 Example of StanfordCoreNLP semantic graph

首先将每个句子输入 StanfordCoreNLP 工具进行依存句法分析处理,得到有关系的词与词的三元组形式的嵌套列表,三元组中包含词 1、词 2、词 1 与词 2 的关系,分别以字符串的形式表示,之后将字符串 1 和字符串 2 送入语言模型中得到其对应的词嵌入表示,将字符串 3 也就是每个单词之间的依赖关系送入语言模型中进行建模。之后统计句子中有关系的词在句子中的位置,以及对应的关系的类别,生成图 4 中的两个矩阵(w_p 和 w_q),分别记录句子中有关系的词的位置,以及对应的关系类型,其中矩阵 w_p 中 0 表示两个词之间有关系,1 表示没有关系。矩阵 w_q 由关系字典中的值构成,之后将两个矩阵与预训练模型的输出一起作为键值记忆网络的输入。

KVMN 网络是由 Wu 等^[22]提出的,本文采用门控机制进行依存句法信息的融合,使用依存关系增强文本的表示,其核心思想为将字与字之间的关联矩阵作为键集,字与字之间的依赖关系作为值集,均作为键值记忆网络的输入。KVMN 网络的输入有 3 个,第一个是矩阵 w_p ,第二个是矩阵 w_q ,第三个是原始句子的词嵌入表示。最后经过 KVMN 网络得到两个输出,一个用来表示获得上下文以及依存句法关系的句子的嵌入表示,另一个就是原始句子的词嵌入表示。其计算公式为:

$$o_i = \sum_{j=1}^q p_{i,j} * e_{i,j}^v \quad (3)$$

$$p_{i,j} = \frac{\exp(h_i * e_{i,j}^k)}{\sum_{j=1}^q \exp(h_i * e_{i,j}^k)} \quad (4)$$

其中, $e_{i,j}^v$ 为值集对应的词嵌入表示, $p_{i,j}$ 为每个句子隐层输出

与键集对应的词嵌入表示进行融合归一化的结果。将 KVMN 网络的两个输出进行加权求和,得到携带其单词依赖信息且包含依存关系的语义特征 L_d 。

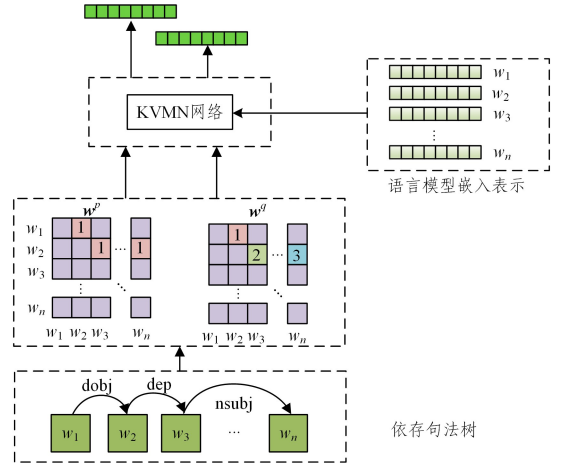


图4 依存语义交互图

Fig. 4 Dependency semantic interaction diagram

3.3 门控融合层

获得原始的编码器输出 L_r 和经过依存关系增强后的语义特征 L_d 后,将它们组合起来以获得集成的信息即句子最终的嵌入表示 u_i ,作为之后的事件类型分类层的输入。我们对 4 种融合策略进行了实验,其中门控相加效果最佳。

(1)相加

$$u_i = L_r + L_d \quad (5)$$

(2)拼接

$$u_i = L_r \oplus L_d \quad (6)$$

(3)门控相加

门模块用于将不同上下文中的信息融合到事件检测模型中。不同的上下文信息可以帮助模型更好地理解上下文语义。

$$k = \sigma(w_1 * h_i^r + w_2 * h_i^d) \quad (7)$$

$$\sigma(x) = 1 / (1 + \exp(-x)) \quad (8)$$

其中, w_1 和 w_2 是可训练的矩阵,之后使用如下公式来控制应该从这两种表示中合并多少信息,+表示逐元素相加。

$$u_i = (k * h_i^r) + (1 - k) * h_i^d \quad (9)$$

(4)门控拼接

$$u_i = (k * h_i^r) \oplus (1 - k) * h_i^d \quad (10)$$

其中, \oplus 表示向量拼接。

3.4 事件类型分类层

对包含依存关系的句子序列的嵌入表示,我们将每个词视为一个事件触发词候选对象,然后将事件触发词候选对象通过一个全连接层得到 108 个类别的分数,如式(11)所示:

$$H_r = W_r * h_i + b_r \quad (11)$$

其中, W_r 为参数矩阵, b_r 为偏置, h_i 为第 i 个句子的隐层向量表示。

计算得到分数之后,通过 Softmax 层对结果进行归一化,得到词对应的事件类型概率。词 W_k 属于事件类别 t 的概率如式(12)所示:

$$P(t | W_k, \theta) = \frac{e^{H_r}}{\sum_{j=1}^T e^{H_r}} \quad (12)$$

其中, θ 为模型参数, H_t 为词 W_k 属于类别 t 的分数, T 为所有的事件类型数量。

3.5 模型训练与优化

本文使用交叉熵作为损失函数, 以更好地优化模型参数, 加快模型的学习速度, 模型的损失计算如式(13)所示:

$$L_{\text{cd}} = -\frac{1}{N} \sum_{j=1}^N \sum_{m=1}^M y_{j,m} \log y_{j,m}^* + (1 - y_{j,m}) \log (1 - y_{j,m}^*) \quad (13)$$

其中, $y_{j,m}^*$ 为第 j 个样本预测为第 m 类事件的预测值, $y_{j,m}$ 为第 j 个样本为第 m 类事件的真实值, N 为样本总数, M 为预定义事件类别总数。

本文采用 Adam 优化算法^[23]对参数进行更新, 并在输入层之后使用 Dropout 机制随机失活一定比例的神经元, 以解决过拟合问题^[24]。

4 实验

4.1 数据集

本文采用 Yao 等^[18]公开的大规模的中国法律事件检测数据集 LEVEN 进行实验。该数据集是最大的法律事件检测数据集, 它包括 8116 份法律文件和 108 个事件类型中的 150 977 个人类注释的事件提及。不仅是与指控有关的事件, LEVEN 还涵盖了一般事件, 这些事件对法律案件的理解至关重要。LEVEN 将法律事件分为 108 个事件类型, 且数据集仅对这 108 个父数据类型进行分类, 其中包括面向指控的事件和一般事件。LEVEN 数据集中事件类型及其分布比率如表 1 所列。

表 1 LEVEN 事件类型及其分布比率

Table 1 LEVEN event types and their distribution

父事件类型	事件类型数	事件提及数	百分比/%	子事件类型
一般行为	40	68616	45.40	售卖、购买、雇佣
禁止性行为	40	43021	28.50	赌博、偷盗、强奸
司法相关	13	20709	19.70	逮捕、投降
危险结果	7	6832	4.50	死亡、受伤
事故	4	2742	1.80	交通事故、火灾
自然灾害	4	57	0.03	洪涝、山崩

4.2 模型评价方法

本文将事件检测任务视为一个基于词语的多分类任务, 具体操作为给定一个句子, 将句子中的每个词语分为 109 类中的 1 类。与以往的工作相同, 本文采用准确率 (Precision, P)、召回率 (Recall, R) 和 F1 值作为评价指标^[25], 其定义如下所示, 其中 TP (True Positive) 为真正例, FP (False Positive) 为假正例, FN (False Negative) 为假反例。

$$P = \frac{TP}{TP + FP} \quad (14)$$

$$R = \frac{TP}{TP + FN} \quad (15)$$

$$F1 = \frac{2PR}{P + R} \quad (16)$$

4.3 对比实验

本文选择了表 2 所列的模型作为基线参与比较。BiLSTM+CRF^[18]和 BERT+CRF^[18]是序列标注模型。为了验证多事件关联建模的有效性, 两个模型都采用条件随机场 (CRF) 作为输出层。DMBERT^[21]是一个基于 BERT 的模型。

表 2 对比实验结果

Table 2 Model comparison results

Model	P	R	F1
BERT ^[18]	84.19	84.31	84.25
BiLSTM+CRF ^[18]	84.74	83.33	84.03
BERT+CRF ^[18]	83.82	84.56	84.19
DMBERT ^[25]	84.77	86.22	85.48
EEQA ^[26]	85.09	83.59	84.33
LifelongED ^[27]	81.82	47.37	60.00
UIE ^[28]	56.48	26.15	35.75
EWDED	93.75	93.37	93.43

与 BERT 不同, DMBERT 使用动态多池化层来聚合特征。但 BERT 模型参数过多, 在训练少量数据时容易出现过拟合等问题。2020EEQA^[22]引入了一种新的基于问题和答案的事件提取范式, 以端到端的方式提取事件参数, 这种方法解决了联合提取模型中仍然存在的错误传播问题。2021LifelongED^[23]定义并提出了一个模型, 通过在旧事件类型和新事件类型之间传递环知识来缓解严重的类型不平衡。2022UIE^[25]是一个用于信息抽取的统一文本到结构生成框架, 它对不同的信息抽取任务进行统一建模, 但它不适用于事件提取的子任务。

4.4 采用不同预训练模型的结果

为了研究不同预训练模型与键值记忆网络的拟合度, 本文使用了 5 种预训练模型。表 3 列出了不同预训练模型在测试集上的实验结果。从结果可以看出, ELECTRA 模型效果最好, 其次是 bert-base-chinese 模型。首先, 对 ELECTRA 模型进行了微调, 使模型在不丢失从大规模通用数据中学习到语义信息的情况下, 对法律文本的适应性更强。其次, 它使用生成器模式进行训练, 相当于对输入进行过滤, 使判别器的任务更加困难, 从而学习到更好的词嵌入表示。鉴别器进行二分类而不是多分类, 降低了模型的复杂度, 并且在相同的超参数设置下, 其训练时间是 bert-base-chines 的一半。最后, 虽然通过结合 transformer-XL 模型结构解决了长文本问题^[24], 但是 XLNET 模型和 Lawformer 模型都采用掩码机制, 而 ELECTRA 同时利用自身和上下文进行预测, 其结果更有效。

表 3 不同预训练模型结果

Table 3 Comparison of different pretrained models results

模型	P	R	F1
bert-base-chinese	92.04	92.07	91.64
ELECTRA ^[30]	93.75	93.37	93.43
XLNET ^[31]	76.16	87.94	84.21
Lawformer ^[32]	88.25	88.94	87.16

4.5 采用不同阶数依存关系的结果

本文分析了依存句法知识对不同顺序的作用, 实验结果如表 4 所列。实验结果表明, 使用一阶依存句法知识的模型取得了最好的效果。二阶依存句法知识扩展了模型的视野, 却引入了过多不相关的上下文信息^[26], 从而影响了模型的性能。

表 4 不同阶数依存关系结果

Table 4 Results of dependence of different orders

依存关系阶数	P	R	F1
一阶	93.75	93.37	93.43
二阶	88.9	89.29	88.78

4.6 采用不同融合方法的结果

本研究将依存关系作为输入的上下文表示,并在语义层面进行增强,采用4种语义增强方法进行对比,其结果如表5所列。从表中可以看出,直接相加的方法改进了模型的语义感知,同时保留了句子的原始语义。除特别说明外,本文其余方法均采用门控相加法。

表5 不同融合方法结果

Table 5 Results of different fusion methods

融合方法	P	R	F1
相加	90.66	91.4	91.29
拼接	93.6	93.32	93.36
门控相加	93.75	93.37	93.43
门控拼接	92.66	92.63	92.62

4.7 消融实验

为了探究所有方法对模型的影响程度,本文设计了如表6所列的消融实验。首先是在基础模型上添加了一阶的依存关系,其效果优于不包含上下文信息的模型,之后将未加噪声的负样本与正样本送入模型进行编码,并采用依存关系进行语义增强,效果虽然略差于单一模型,但是经过噪声处理后,模型效果更佳,可见其可以提升模型对不同样本的辨识能力,

学习到更优秀的特征表示。

表6 消融实验结果

Table 6 Results of ablation experiments

模型	P	R	F1
DMBERT	84.77	86.22	85.48
DMBERT+依存关系	90.79	91.02	91.22
DMBERT+负样本	90.62	90.77	90.93
DMBERT+依存关系+负样本	92.04	92.07	91.64
ELECTRA+动态多池化	84.54	84.76	85.65
ELECTRA+动态多池化+依存关系	93.62	93.3	93.35
ELECTRA+动态多池化+负样本	93.58	93.44	93.36
ELECTRA+动态多池化+负样本+依存关系	93.49	93.12	93.19
EWDED	93.75	93.37	93.43

4.8 案例分析

表7列出了模型未使用依存关系加强对句子结构的理解时的触发词分类结果,及添加后的结果和门模块对词之间依存关系的关注程度。从结果中可以看出仅使用预训练模型(PLM)的情况下,模型会对触发词进行错误分类。依据门模块(Gate)对依存关系(dep)的关注程度的结果可以看出,给预训练模型提供了句子完整的依存关系后可以让模型更好地关注到与词与词之间的关系,获得词语更好的向量表示。

表7 案例分析结果

Table 7 Results of case study

句子	语言模型	Gold	添加依存关系	门控系数
2014年,公安机关接受 报案 后立案侦查,于5月16日电话通知吴某到公安机关 接受 调查,吴某随后到大新镇派出所接受调查,如 实供述 犯罪事实。	投案 受损 提供	报警/报案 同意/接受 供述	报警/报案 同意/接受 供述	0.6736 0.6179 0.5459

如表7所列,针对输入句子中“报案”触发词,模型对依存句法产生的关注程度对0.6736,说明门控机制可以更关注句子的语义信息,产生了正确的分类结果。这进一步表明,即使是在具有丰富先验知识预训练模型做编码器的情况下,仍可以通过预训练语言模型+依存关系框架获得进一步的性能提升。

4.9 实验参数设置

本文中的BERT,BiLSTM-CRF,BERT+CRF,DMBERT的实验结果均引用Yao等^[18]论文中的实验参数,其余模型的参数设置如表8所列。

表8 模型参数

Table 8 Model parameters

Model	Batch Size	Learning Rate	Hidden Size
EEQA	8	1×10^{-3}	768
LifelongED	4	1×10^{-3}	768
UIE ^[29]	16	2×10^{-5}	768
EWDED	8	1×10^{-3}	768

不同预训练模型的实验参数设置如表9所列。

表9 不同预训练模型参数

Table 9 Parameters of different pretrained models

Model	Batch Size	Learning Rate	Hidden Size
bert-base-chinese	64	1×10^{-3}	768
ELECTRA	8	1×10^{-3}	768
XLNET	16	1×10^{-3}	768
Lawformer	8	$1e \times 10^{-3}$	768

模型,该模型由4部分组成:句子编码层、依存语义图交互层、门控融合层、事件类型分类层。针对文本中事件触发词经常分散在不同的句子中,本文采用依存关系作为上下文表示,并采用门控机制对其进行加权,以解决神经网络模型语义缺失等问题。针对模型的辨别问题,本文添加负样本,且采用针对不同样本添加不同噪声的方法,来增强模型获得输入的高质量向量表示能力以及模型的泛化能力和纠错能力。实验结果表明通过构建依存图兼顾词性与语义的事件检测方法可以解决语义缺失等问题,提升了事件检测的效果。中文事件检测任务中,除了以上问题,还有数量较多的未知事件触发词和事件类型稀疏问题。未知事件触发词指在训练集中未出现的触发词,事件类型稀疏指某类事件的触发词在训练集中出现次数较少。这些问题阻碍了事件检测任务性能的提升,因此本文的下一步研究方向是探索如何通过基于篇章信息和字符信息融合,来解决事件触发词和事件类型稀疏问题。

参考文献

- [1] DAI J H, PENG R Y, XU L, et al. A Review of Information Extraction Based on Deep Neural Networks [J]. Journal of Southwest Normal University(Natural Science Edition), 2022, 47(4): 1-11.
- [2] LIU P, WEI H Z, LU X L, et al. Constructing a Mine Disaster Event Detection Model Based on a New Convolutional Neural Network [J]. Journal of Chinese Information Science, 2020, 34(10): 59-68.
- [3] LI R, ZHAO W, YANG C, et al. Treasures outside contexts: Im-

结束语 本文提出了一个基于 Electra 的新的事件检测

- proving event detection via global statistics[C]//Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing. 2021:2625-2635.
- [4] LIU X, HUANG H, SHI G, et al. Dynamic prefix-tuning for generative template-based event extraction [J]. arXiv: 2205.06166, 2022.
- [5] MIKOLOV T, SUTSKEVER I, CHEN K, et al. Distributed representations of words and phrases and their compositionality [J]. arXiv:1310.4546, 2013.
- [6] BARBANO C A, DUFUMIER B, TARTAGLIONE E, et al. Unbiased Supervised Contrastive Learning [J]. arXiv: 2211.05568, 2022.
- [7] GU S, CHU Y, ZHANG W, et al. Research on System Log Anomaly Detection Combining Two-way Slice GRU and GA-Attention Mechanism[C]//2021 4th International Conference on Artificial Intelligence and Big Data (ICAIBD). IEEE, 2021: 577-583.
- [8] LIU W, NGUYEN T H. Similar but not the same: Word sense disambiguation improves event detection via neural representation matching[C]//Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. 2018: 4822-4828.
- [9] LIU J, CHEN Y, LIU K, et al. Event detection via gated multilingual attention mechanism [C] // Proceedings of the AAAI Conference on Artificial Intelligence. 2018.
- [10] LUO R Y, CHEN J F, YIN X X. A review of the application of machine learning in automatic detection of earthquake events [J]. Advances in Geophysics. 2021, 36(3): 923-932.
- [11] TONG M, WANG S, CAO Y, et al. Image enhanced event detection in news articles[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2020:9040-9047.
- [12] LU Y, LIN H, HAN X, et al. Distilling discrimination and generalization knowledge for event detection via delta-representation learning[C]//Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. 2019:4366-4376.
- [13] TONG M, XU B, WANG S, et al. Improving event detection via open-domain event trigger knowledge [C] // Association for Computational Linguistics. 2020.
- [14] 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 (Volume 2: Short Papers). 2015:365-371.
- [15] 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). 2015:167-176.
- [16] 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. 2016:300-309.
- [17] ZHU P P, WANG Z Q, LI S S, et al. Chinese Event Detection based on Text Information and Bi-GRU [J]. Computer Science, 2020, 47(12): 233-238.
- [18] YAO F, XIAO C, WANG X, et al. LEVEN: A Large-Scale Chinese Legal Event Detection Dataset [J]. arXiv: 2203.08556, 2022.
- [19] LI X, LI F, PAN L, et al. DuEE: a large-scale dataset for Chinese event extraction in real-world scenarios[C]//CCF International Conference on Natural Language Processing and Chinese Computing. Cham:Springer, 2020: 534-545.
- [20] DEVLIN J, CHANG M W, LEE K, et al. Bert: Pre-training of deep bidirectional transformers for language understanding[J]. arXiv:1810.04805, 2018.
- [21] HSU I H, HUANG K H, BOSCHÉE 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. 2022:1890-1908.
- [22] WU C, WU F, QI T, et al. NoisyTune: A Little Noise Can Help You Finetune Pretrained Language Models Better [J]. arXiv: 2202.12024, 2022.
- [23] TESNIÈRE L. Elements of structural syntax[M]. John Benjamins Publishing Company, 2015.
- [24] CHEN Y, YANG H, LIU K, et al. Collective event detection via a hierarchical and bias tagging networks with gated multi-level attention mechanisms[C]// Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. 2018: 1267-1276.
- [25] TIAN Y, CHEN G, SONG Y. Enhancing aspect-level sentiment analysis with word dependencies[C]// Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume. 2021:3726-3739.
- [26] WANG X, WANG Z, HAN X, et al. MAVEN: A massive general domain event detection dataset[J]. arXiv:2004.13590, 2020.
- [27] DU X, CARDIE C. Event extraction by answering (almost) natural questions[J]. arXiv:2004.13625, 2020.
- [28] YU P, JI H, NATARAJAN P. Lifelong event detection with knowledge transfer[C]//Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing. 2021:5278-5290.
- [29] LU Y, LIU Q, DAI D, et al. Unified Structure Generation for Universal Information Extraction[J]. arXiv:2203.12277, 2022.
- [30] CLARK K, LUONG M T, LE Q V, et al. Electra: Pre-training text encoders as discriminators rather than generators[J]. arXiv:2003.10555, 2020.
- [31] YANG Z, DAI Z, YANG Y, et al. XLNet: Generalized Autoregressive Pretraining for Language Understanding [J]. arXiv: 1906.08237.
- [32] XIAO C, HU X, LIU Z, et al. Lawformer: A Pre-trained Language Model for Chinese Legal Long Documents [J]. arXiv: 2003.10555, 2020.



YIN Baosheng, born in 1975, professor. His main research interests include deep learning and natural language processing.



KONG Weiyi, born in 1999, postgraduate. Her main research interests include event detection and so on.