

基于反馈图卷积神经网络的事件检测方法

刘浏^{1,2}, 丁鲲^{1†}, 刘珊珊¹, 刘茗¹

(1. 国防科技大学第六十三研究所, 江苏 南京 210007;
2. 宿迁学院 信息工程学院, 江苏 宿迁 223800)

摘要:事件检测是自然语言处理领域的重要任务之一,其结果可以有效支撑信息抽取、文本分类和事件推理等下游任务。预训练语言模型BERT在事件检测任务上取得了显著的成绩,然而该类方法无法有效获取长距离和结构化的文本信息。为了缓解该问题,本文提出基于反馈网络的图卷积神经网络模型进行文本结构信息捕获,同时这种新方法能够有效解决图卷积神经网络带来的语义信息衰减性问题。本文首先使用BERT预训练模型获取文本的语义特征,然后使用融入反馈网络的图卷积神经网络提取文本的句法结构特征,最终使用多分类器实现对事件触发词的识别和分类。公开数据集ACE 2005上的实验结果表明,本文提出的事件检测方法在事件触发词识别和分类任务上的 F_1 值分别达到了74.46%和79.49%,较现有工作平均提高了4.13%和4.79%。

关键词:事件检测;图神经网络;BERT;句法分析;自然语言处理
中图分类号:TP391 **文献标志码:**A

Event Detection Method Based on Feedback Graph Convolutional Networks

LIU Liu^{1,2}, DING Kun^{1†}, LIU Shanshan¹, LIU Ming¹

(1. The Sixty-third Research Institute, National University of Defense Technology, Nanjing 210007, China;
2. School of Information Engineering, Suqian University, Suqian 223800, China)

Abstract: Event detection is one of the most important tasks in the field of natural language processing (NLP). Its result is the key information supporting downstream tasks, such as information extraction, text classification and event reasoning. BERT model has achieved remarkable achievements in event detection. However, it cannot effectively obtain long-distance and structured text information. To alleviate this problem, feedback-based GCNs network is proposed to capture text structure information in this paper, and it can solve the problem of semantic information attenuation caused by GCNs. This paper first uses BERT model to obtain semantic features of the text, then adopts GCNs integrated into the feedback network to extract the syntactic structure features of the text, and finally employs multiple classifiers to identify and classify event trigger words. The experimental results on the open

* 收稿日期:2022-06-27

基金项目:国家自然科学基金资助项目(71901215), National Natural Science Foundation of China(71901215);中国博士后科学基金资助项目(2021MD703983), China Postdoctoral Science Foundation (2021MD703983);江苏省“333工程”培养资金资助项目(BRA2020418), Jiangsu Province “333 Projects” (BRA2020418);国防科技大学科研计划项目(ZK20-46);江苏省高等学校自然科学研究面上项目(20KJB413003), Natural Science Foundation of the Jiangsu Higher Education Institutions of China (20KJB413003);宿迁市科技计划项目(K202128), Science and Technology Program of Suqian (K202128)

作者简介:刘浏(1988—),男,江苏宿迁人,国防科技大学博士后,博士

† 通信联系人, E-mail: dingkun18@nudt.edu.cn

dataset ACE 2005 show that the F_1 value of the event detection method proposed in the task of event trigger word recognition and classification has reached 74.46% and 79.49%, respectively, which gains an average increase of 4.13% and 4.79% compared with the existing work.

Key words: event detection; graph neural network; BERT; syntactic parsing; natural language processing

在大数据时代,信息以爆炸式速度增长,而它们大多以非结构化文本形式存在,如微博、新闻以及军事情报等.如何有效利用文本中蕴含的有价值信息成为新的挑战.事件检测作为自然语言处理技术一个重要子任务,其目标就是要从这些不规则的文本中提取出结构化信息用以支持知识图谱、信息检索、推荐系统等下游任务.

事件检测的主要目标是识别文本中的事件触发词并对事件进行分类.触发词主要表现为动词,动词通常是一个句子的核心词汇,能够直接体现事件的类型和状态.ACE会议(自动内容抽取测评会议)则将触发词描述为触发事件发生的词^[1].例如下面的句子:

S: Many refugees have *left* Ukraine.

(许多难民乌克兰.)

在句子S中,事件检测的主要任务是识别动词“left”为触发词并将其分类为“Transport”.

深度学习在欧氏空间数据特征的挖掘上取得了成功,但是在现实应用场景中数据大多来自非欧氏空间,例如社交网络中,用户之间非规则化的复杂关系.为此,有学者提出了以节点和边形成的图结构学习模型,即图神经网络(Graph Neural Network, GNN)^[2]. GNN模型能够学习非欧氏空间中数据之间的依赖关系,同时通过聚合邻居节点的信息进而丰富其表征能力.因此在自然语言处理领域,通常使用GNN模型获取具有词间依赖关系的结构化信息.同时, GNN随着隐藏层深度的增加,其节点可感知更远的邻居节点信息,因此GNN能够获取长距离的节点信息.

近年来,预训练语言模型BERT^[3]在文本分类、机器翻译以及自动问答等领域任务上展示了卓越性能,成为实现事件检测任务的重要技术手段.虽然BERT相较于ELMo^[4]、GPT^[5]等其他预训练模型具有更强的语义理解能力和语义特征提取能力,但是BERT无法有效捕获长距离的结构化文本特征.为此,本文提出了一种改进图神经网络模型,即反馈图卷积神经网络模型(FB-GCNs).本文的主要贡献在

于:首先使用GCNs与预训练语言模型BERT相结合,从而提高模型捕获长距离和结构化信息的能力;其次,设计反馈网络模块用以缓解由GCNs模型带来的语义特征衰减问题;然后,通过在公开数据集上开展对比实验和消融实验,验证了本文模型的可行性和优越性.同时,消融实验也进一步验证了随着隐藏层的增加,GCNs模型对语义信息的传递呈线性衰减趋势.

1 相关工作

与传统的统计学模型不同,深度学习模型作为具有非线性表示能力的深层次网络,能够自动挖掘数据特征,因此许多学者将其应用于事件检测任务.Chen等^[6]和Nguyen等^[7]最早将深度卷积神经网络(CNN)应用于事件检测任务,并取得了不错的效果.Nguyen等^[8]和Jagannatha等^[9]使用循环神经网络(RNN)进行事件检测,其中前者利用BiRNN进行文本表征,然后采用联合学习模型对事件触发词和论元进行识别.后者则利用BiRNN获取词嵌入,并利用条件随机场实现对电子病历中包含的医疗事件进行检测.Hong等^[10]提出了一种名为自约束学习方法,其利用生成对抗网络生成虚假特征从而解决事件检测过程中可能由伪关系引起的错误.Liu等^[11]为减少触发词标注带来的时间成本,提出了一种无须触发词标注的事件检测方法.该方法使用LSTM学习词的上下文环境特征,同时联合注意力机制获取的触发词局部信息,最终通过多任务分类器实现对事件类型的判断.然而,基于深度学习的事件检测技术依然面临着数据标注的问题,因为深度学习模型的训练严重依赖大规模的有标签实验数据的支撑.同时,它难以处理像图数据这样的非欧氏空间问题.因此,许多自然语言处理的专家开始越来越关注图神经网络在事件检测任务上的应用.Nguyen和Grishman^[12]认为之前的神经网络主要研究的是文本序列特征,对文本的语义结构特征研究相对较少,因此提出基于图

神经网络的事件检测方法.Cui等^[13]认为经典的图神经网络忽略了依赖关系标签的信息,因此设计了一种强化边信息的图卷积神经网络用于事件检测.该方法利用边代表的节点依赖关系更新节点信息,同时又使用节点信息更新边表征.

BERT模型凭借着强大的语义表征能力成为较成功的预训练语言模型之一.目前许多事件检测任务都是基于该模型开展研究的,并取得了显著的成果.Yang等^[14]将文本中的词、位置以及段嵌入输入到BERT模型获取具有语义信息的编码,然后使用多分类器作用于编码进行触发词抽取.Wadden等^[15]利用BERT构建一个DyGIE++框架,进行跨文本建模,并捕获句内和跨句子上下文信息.Liu等^[16]和Du等^[17]为了缓解模型对数据的依赖以及错误传播的问题,将事件抽取任务构造为基于BERT的机器阅读理解模型.两者不同的是,前者使用非监督方法进行问题

生成,而后者使用人工方式设计问题.但是这类方法往往对长距离的信息挖掘存在不足.因此本文利用预训练模型获取文本的语义特征,并输入图神经网络,从而实现语义特征和长距离的结构特征相结合.

2 事件检测模型

图1为事件检测模型框架,主要由语义特征提取、结构特征提取、反馈网络以及事件检测等模块组成.语义提取模块使用预训练语言模型挖掘句子级的上下文语义信息.结构特征提取模块是使用由文本句法分析结果转化的邻接矩阵构造的图神经网络,提取文本的结构化信息.反馈网络包含多个反馈层,每个反馈层由归一化函数、激活函数组成.最后,全连接层使用多任务分类器实现对文本中的事件触发词的识别和分类.

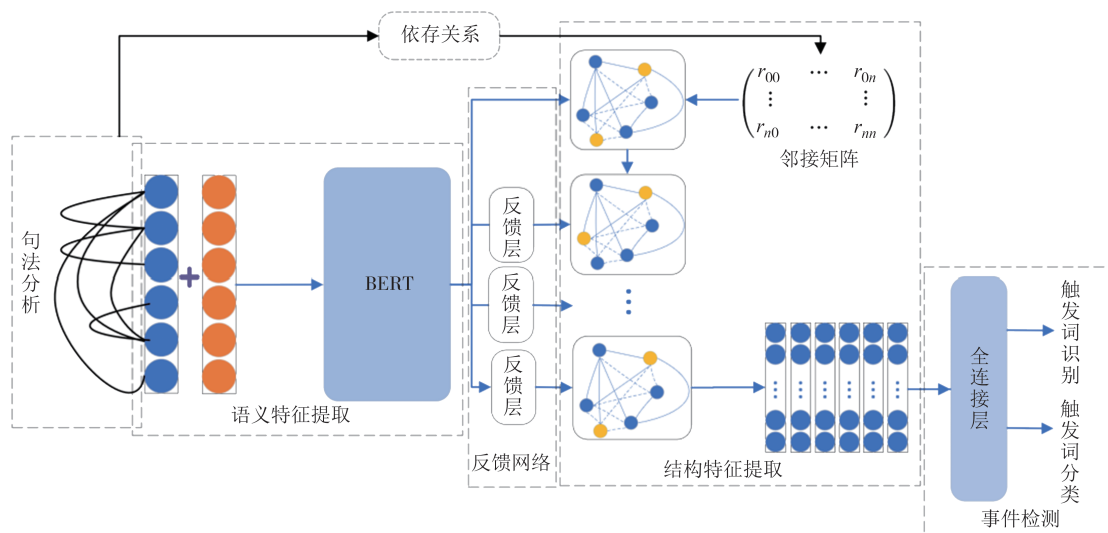


图1 FB-GCNs模型框架图

Fig.1 Illustration of FB-GCNs architecture

2.1 语义特征

语义特征提取模块主要包括词编码和BERT模型两个部分.本文的词编码由两个部分组成,即Token embeddings和Position embeddings.前者通过查询WordPiece embeddings方法^[17]形成的编码词典获取不同词的数值表示,其中该词典包含英文常用的30 000个单词的编码结果.后者通过三角函数表示不同词在句子中相对位置信息.需要说明的是,由于本文重点研究句子级的事件检测任务,因此没有加入段嵌入(Segment embeddings).

BERT预训练模型是将Token embeddings和

Position embeddings作为输入,使用叠加的Transformer encoder作为语义分析的核心模块.本文使用的BERT模型是采用文献[18]中的BERT_Base结构,其包含Transformer Block的层数为12,隐藏层的维度为768,Self-Attention头数为12.

2.2 结构特征

为了获取文本的结构化特征,本文使用图神经网络捕获不同词汇之间的依赖关系.首先,使用句法分析工具抽取原始文本中词间依赖关系,然后将其转化为邻接矩阵,最后使用邻接矩阵对图神经网络中节点关系进行初始化.

2.2.1 句法分析

本文使用NLP工具spaCy对句子进行句法分析,并得到词之间的依存关系,如图2所示.通常依存关系是二元非对称的,其中以动词或动词短语为中心词(Root).句子与中心词之间存在依存路径,图中的圆弧表示词之间的存在依存关系,例如(left, Ukraine)的关系为dobj,表示后者是前者的直接宾语.

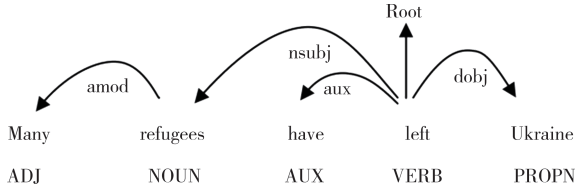


图2 句法分析示例图

Fig.2 An example of syntactic dependency parsing

句子中的词间依赖关系具有多样性,示例中便存在5种关系.关系的非对称性又决定了依存路径的有向性.为了能够将这种依存路径表示为可计算的邻接矩阵 $A_{n \times n}$,其元素表示为 $A_{i,j} = \begin{cases} 1, (v_i, v_j) \in E \\ 0, (v_i, v_j) \notin E \end{cases}$.本文将存在依存路径的两个词关系值表示为1,不存在依存路径的值表示为0.依存路径分为直接和间接路径两种,图中(left, Ukraine)为直接路径,由(left, refugees)→(refugees, Many)推导的(left, Many)的路径为间接路径.

由于词间的依存关系是相互,因此在邻接矩阵的转化过程中将路径依存关系转化为对称的无向图.同时为每一个词添加环,使得自身原始信息得以保留.

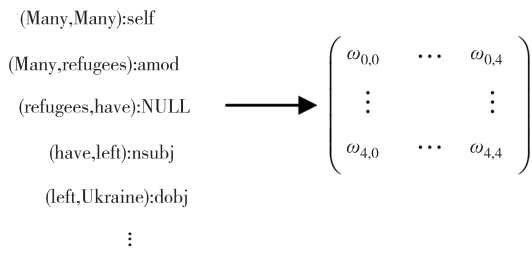


图3 基于句法依赖关系的邻接矩阵

Fig.3 Adjacency matrix based on syntactic dependency

2.2.2 图卷积神经网络

为便于描述,本文采用文献[12]的描述形式将图表示为 $G = (V, E)$,其中 V 表示节点(词向量)的集合, E 表示边(词间关系)的集合.定义一个矩阵 $X \in \mathbb{R}^{m \times n}$ 表示句子词嵌入, n 表示词的数量, m 表示词向量 $x_v \in V$ 的维度.

如图4所示,原始的GCNs模型中的节点向量更

新是由邻居节点聚合而成,而本文则通过上一节中的邻接矩阵确定用来聚合的邻居节点.图神经网络的层次代表需要更新的次数,其中对第 $k+1$ 层节点进行更新的公式如下所示:

$$h_v^{(k+1)} = \text{ReLU} \left(\sum_{u \in N(v)} W_{L(u,v)}^{(k)} h_u^{(k)} + b_{L(u,v)}^{(k)} \right) \quad (1)$$

式中: $N(v) \in V$ 表示节点 v 的邻居集合, $L(u,v)$ 表示节点 v 与邻居节点之间存在的依存路径,若存在则为1,否则为0. v 表示隐藏层数, $W_{L(u,v)}^{(k)} \in \mathbb{R}^{n \times n}$ 是可学习参数, $b_{L(u,v)}^{(k)}$ 为偏置参数,ReLU为激活函数, $h_v^{(1)} = x_v$.

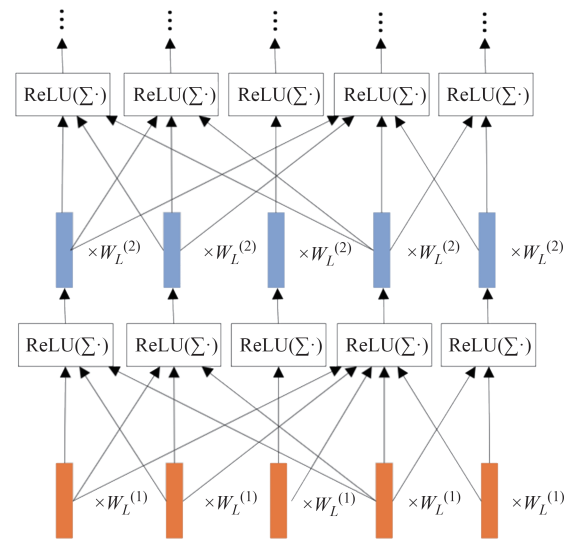


图4 图卷积神经网络结构图

Fig.4 Illustration of GCNs architecture

2.3 反馈网络

受残差网络^[19]的启发,本文设计了一种反馈网络用以缓解GCNs模型的退化问题以及破坏 x_v 的语义信息.如图5所示,与上一节中在第 $(k+1)$ 隐藏层的节点 x_v 的更新公式不同,经过反馈网络后的公式为:

$$\tilde{h}_v^{(k+1)} = \alpha^{(k+1)} \times \text{LN}(h_v^{(k+1)}) + \beta^{(k+1)} \times \text{LN}(x_v) \quad (2)$$

式中: $\text{LN}(x) = \frac{x - E[x]}{\sqrt{\text{Var}[x] + \epsilon}}$, α 和 $\beta \in [0, 1]$ 为超

参数.

反馈网络通过在GCNs模型每一层中引入由BERT模型获取的语义信息,从而减少由GCNs模型节点信息聚合带来的语义信息衰减和破坏.在语义特征与结构化特征融合过程中采用了归一化方法避免两者向量化计算时出现数值空间差异较大的问题,并能够加快模型收敛.同时使用权重系数对语义特征和结构特征进行调和,使其达到最优的状态.

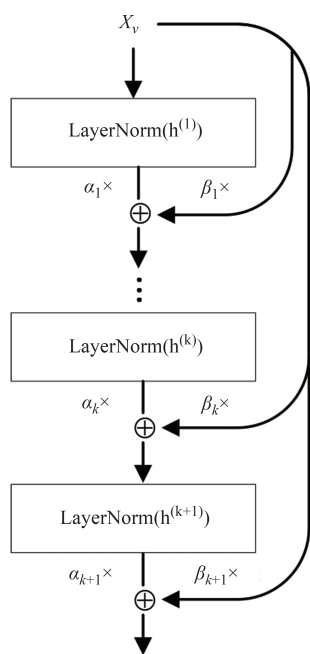


图5 反馈网络

Fig.5 Feedback network

2.4 事件检测

由于事件的类型是多样的,因此本文使用多分类器对样本中的词进行预测.GCNs模型中的节点是对样本句中词的量化表示,因而使用如下公式:

$$\hat{y} = \text{softmax}(\tilde{h}W + b) \in \mathbb{R}^T \times n \quad (3)$$

式中: $\tilde{h}$ 为GNN中的节点表示向量,即包含语义和结构信息的词向量, W 为可训练参数, b 为偏置向量.

本文将交叉熵作为损失函数,其公式如下:

$$L = \left(- \sum_{i=0}^{N-1} \log \left(\frac{e^{y_i}}{\sum_j e^{\hat{y}_i^j}} \right) \right) / N \quad (4)$$

式中: N 表示训练batch中的样本数量, y_i 表示样本的真实标签概率, $\hat{y}_i^j$ 表示第*i*个样本的被判断为类别*j*的概率.

3 实验

3.1 实验数据与评估指标

本文实验采用的是ACE 2005数据集,该数据集包含来自不同领域的新闻、微博、广播等文本^[20].ACE 2005数据集包含8种事件类型,33种子类型.为了实验的公平性,使用文献[21]中的方法共抽取599篇文本中的15 715个句子作为实验数据,其中14 180个句子作为训练集样本,863个句子作为验证集样

本,672个句子作为测试集样本.

本文使用精准率 P (precision)、召回率 R (recall)、 F_1 三个指标对实验结果进行评估.由于事件检测由触发词识别和分类组成,因此分别使用两组相同的指标进行评估.

3.2 实验环境与模型训练参数

本文实验采用的编程语言是Python 3.8,使用的深度学习框架为PyTorch1.8.0+CUDA 11.1,预训练语言模型使用的是BERT-base-uncased版本的BERT模型,图计算框架采用的是DGL 0.8.1+CUDA 11.0,句法分析工具采用的是spaCy v3.0(加载的模型为en_core_web_lg-3.2.0).FB-GCNs模型训练的具体参数配置见表1.

表1 实验参数配置

Tab.1 Parameter configurations of our experiment

参数	数值
Epoch	10
Batch size	8
Learning rate	4e-5
Dropout	0.2
Warmup proportion	0.1
GCN layers	5
GCN input	768
Hidden size	768

3.3 实验结果与分析

本文的实验主要包括对比实验和消融实验,其中对比实验用于将FB-GCNs模型与已有的主流事件检测模型进行性能比较,而消融实验则为了验证模型中反馈网络以及GCNs模型结构对性能的影响程度.

3.3.1 对比试验

为了有效评估FB-GCNs模型的性能,本文选择将其与6种不同类型的事件检测模型进行比较,这6种模型分别是:

DMCNN^[6]:是一种动态多池化卷积神经网络,用以捕获包含多事件的句子中有效信息.

TBNNAM^[11]:为了避免对触发进行标注,使用注意力机制计算聚集事件类型的句子表示,从而实现了一种无须触发词的事件检测任务.

BERT_QA^[17]:主要思想是通过为事件抽取不同任务设计相应的问答对,从而将多分类问题转化为

机器阅读理解.

GCN-ED^[12]:首次将GCNs模型应用于事件检测任务,它除了利用句法依赖关系表征词的上下文信息,还提出了一种整合实体提及的聚合方法.

JMEE^[22]:将注意力机制应用于GCNs模型的信息聚合,从而实现事件触发词和论元的联合抽取.

Joint3EE^[23]:为了防止错误传递,让命名实体识别、触发词检测以及论元分类三个任务共享BiRNN模型对句子的表示能力,进而实现三者任务的联合学习.

表2是本文模型以及各基准模型在ACE 2005数据集上的测试结果.该实验结果表明,本文模型相较于其他模型在触发词分类上的精确率 P 、召回率 R 以及 F_1 值分别平均提高了1.33%、4.24%及4.13%,在触发词的识别上也分别提高了4.68%、4.59%及4.79%.但是在触发词分类的精确率上相较于TBNNAM、GCN-ED以及JMEE等模型有所下降,这是由于TBNNAM模型采用了无触发词标注方法,避免了有触发词标注错误给模型训练带来的影响.但是由于其没有能力捕获文本的结构化信息,因此在综合指标 F_1 上明显弱于本文模型.JMEE模型相较于本文模型在GCNs模型的聚合上做出了改进,因此在结构化信息捕获方面具有一定优势.GCN-ED模型也是图神经网络模型,由于其将实体信息与触发词关联起来,从而促进了对事件分类精确率的提高.但是由于它们相较于本文模型在语义信息捕获能力上的不足,因此在综合指标 F_1 值上显示出一定的劣势.

表2 ACE 2005数据集上事件检测结果

Tab.2 Event detection results on ACE 2005

模型	触发词分类			触发词识别		
	P	R	F_1	P	R	F_1
DMCNN	75.60	63.60	69.10	80.40	67.70	73.50
TBNNAM	76.20	64.50	69.90	—	—	—
BERT_QA	71.12	73.70	72.39	74.29	77.42	75.82
GCN-ED	77.90	68.80	73.10	—	—	—
JMEE	76.30	64.50	69.90	80.20	72.10	75.90
Joint3EE	68.00	71.80	69.80	70.50	74.50	72.50
FB-GCN	75.51	73.42	74.46	80.61	78.41	79.49

3.2.2 消融实验

为了验证GNN模型在语义特征上的衰减性和破坏性,这里采用不同隐藏层数量的GCNs模型在事件触发词分类和识别上的 F_1 值作为验证指标.从图6

和图7可以看出,随着GCNs模型隐藏层深度的不断增加,其对事件触发词分类和识别的性能呈现线性下降趋势.这表明随着模型深度的增加,GNNs模型对语义特征具有显著的衰减和破坏作用.在不多于两个隐藏层的GCNs模型中,其性能还是保持较高的水平,这是由于首层接收的是带有语义特征的文本向量,而浅层模型在获得结构化特征的同时依然保留着部分语义信息.

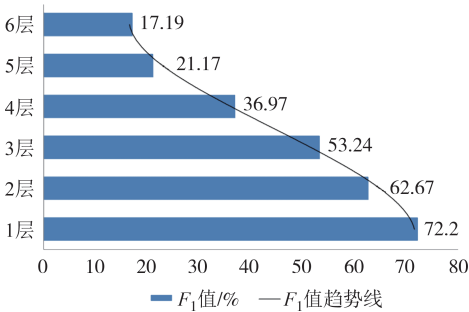


图6 GNNs模型深度对触发词分类 F_1 值影响

Fig.6 Influence of GNNs model depth on F_1 value of trigger classification

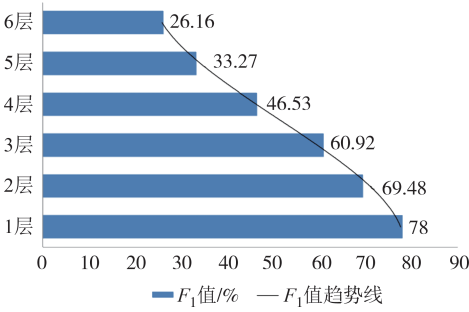


图7 GNN模型深度对触发词识别 F_1 值影响

Fig.7 Influence of GNN model depth on F_1 value of trigger identification

为了验证在事件检测任务中,文本的语义信息与结构信息都具有重要作用,同时也为了进一步验证本文所提反馈网络对GCNs模型具有一定的提升作用,即能够防止语义信息的衰减和破坏,本文对FB-GCNs模型进行消融实验.该实验主要将缺少捕获结构信息能力的BERT+FineTune模型、失去反馈网络的BERT+GCNs模型以及本文的完整模型FB-GCNs三者进行比较.由于在对比实验中已经与多种GNNs模型进行比较,因此消融实验中没有重复与GNNs模型的比较.从图8可以看出,本文提出的FB-GCNs模型相较于其他两个模型具有明显的优势.由此可以表明,合理地利用语义特征和结构体特征对于事件检测具有重要的意义.

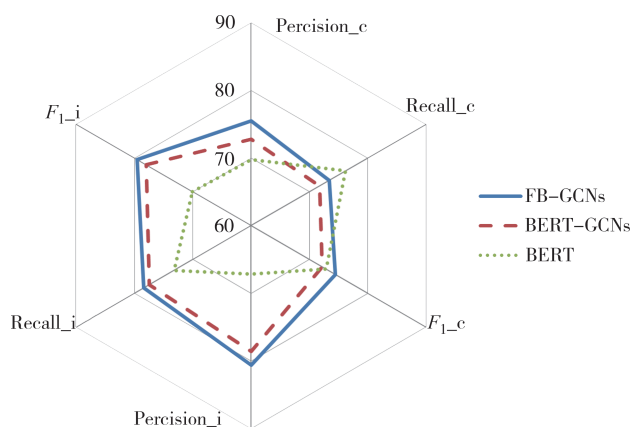


图 8 消融实验结果对比

Fig.8 Comparison of ablation results

4 结论

事件检测是重要的自然语言处理子任务之一, 本文为了能够有效利用文本的语义和结构特征从而提升事件检测的性能, 提出了基于反馈图卷积神经网络的事件检测模型. 该模型通过设计一种反馈网络从而缓解了由于图神经网络节点表示的多层次聚合带来的语义信息衰减和破坏问题. 最终在公开数据集上的实验结果表明, 本文模型能够有效地对触发词进行识别与分类. 但是本文并未对不同的图神经网络模型在事件检测任务上的性能差异进行深入研究和验证, 因此未来将对图神经网络模型展开深入研究.

参考文献

- [1] Linguistic Data Consortium. ACE (Automatic Content Extraction) English Annotation Guidelines for Events Version 5.4. 3 [EB/OL]. (2005-07-01)[2020-06-18]. <https://www ldc.upenn.edu/>.
- [2] SCARSELLI F, GORI M, TSOI A C, et al. The graph neural network model [J]. IEEE Transactions on Neural Networks, 2009, 20(1): 61-80.
- [3] DEVLIN J, CHANG M W, LEE K, et al. BERT: pre-training of deep bidirectional transformers for language understanding [EB/OL]. 2018; arXiv: 1810.04805. <https://arxiv.org/abs/1810.04805>.
- [4] PETERS M, NEUMANN M, IYYER M, et al. Deep contextualized word representations [C]//Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers). New Orleans, Louisiana. Stroudsburg, PA, USA: Association for Computational Linguistics, 2018: 2227-2237.
- [5] RADFORD A, NARASIMHAN K, SALIMANS T, et al. Improving language understanding by generative pre-training [EB/OL] [2022-06-25]. https://cdn.openai.com/research-covers/language-unsupervised/language_understanding_paper.pdf.
- [6] 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 (Volume 1: Long Papers). Beijing, China. Stroudsburg, PA, USA: Association for Computational Linguistics, 2015: 167-176.
- [7] 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. Austin, Texas. Stroudsburg, PA, USA: Association for Computational Linguistics, 2016: 886-891.
- [8] 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. Stroudsburg, PA, USA: Association for Computational Linguistics, 2016: 300-309.
- [9] JAGANNATHA A N, YU H. Bidirectional RNN for medical event detection in electronic health records [C]//Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. San Diego, California. Stroudsburg, PA, USA: Association for Computational Linguistics, 2016: 473.
- [10] HONG Y, ZHOU W X, ZHANG J L, et al. Self-regulation: employing a generative adversarial network to improve event detection [C]//Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Melbourne, Australia. Stroudsburg, PA, USA: Association for Computational Linguistics, 2018: 515-526.
- [11] LIU S, LI Y, ZHANG F, et al. Event detection without triggers [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). PA, USA: Association for Computational Linguistics, 2019: 735-744.
- [12] NGUYEN T, GRISHMAN R. Graph convolutional networks with argument-aware pooling for event detection [J]. Proceedings of

- the AAAI Conference on Artificial Intelligence, 2018, 32(1): 5900–5907.
- [13] CUI S Y, YU B W, LIU T W, et al. Edge-enhanced graph convolution networks for event detection with syntactic relation [EB/OL]. 2020; arXiv: 2002.10757. <https://arxiv.org/abs/2002.10757>.
- [14] YANG S, FENG D W, QIAO L B, et al. Exploring pre-trained language models for event extraction and generation [C]// Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Florence, Italy. Stroudsburg, PA, USA: Association for Computational Linguistics, 2019: 5284–5294.
- [15] WADDEN D, WENNERBERG U, LUAN Y, et al. Entity, relation, and event extraction with contextualized span representations [EB/OL]. 2019; arXiv: 1909.03546. <https://arxiv.org/abs/1909.03546>.
- [16] LIU J A, CHEN Y B, LIU K, et al. Event extraction as machine reading comprehension [C]// Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP). Online. Stroudsburg, PA, USA: Association for Computational Linguistics, 2020: 1641–1651.
- [17] DU X Y, CARDIE C. Event extraction by answering (almost) natural questions [EB/OL]. 2020; arXiv: 2004.13625. <https://arxiv.org/abs/2004.13625>.
- [18] WU Y H, SCHUSTER M, CHEN Z F, et al. Google's neural machine translation system: bridging the gap between human and machine translation [EB/OL]. 2016; arXiv: 1609.08144. <https://arxiv.org/abs/1609.08144>.
- [19] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition [C]// 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, NV, USA: IEEE, 2016: 770–778.
- [20] DODDINGTON G, MITCHELL A, PRZYBOCKI M A, et al. 2004. The automatic content extraction (ACE) program—tasks, data, and evaluation [J]. Proc Lrec, 2004, 2(1): 1–4.
- [21] ZHANG T T, JI H, SIL A. Joint entity and event extraction with generative adversarial imitation learning [J]. Data Intelligence, 2019, 1(2): 99–120.
- [22] LIU X, LUO Z C, HUANG H Y. Jointly multiple events extraction via attention-based graph information aggregation [EB/OL]. 2018; arXiv: 1809.09078. <https://arxiv.org/abs/1809.09078>.
- [23] NGUYEN T M, NGUYEN T H. One for all: neural joint modeling of entities and events [J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2019, 33(1): 6851–6858.