



基于双层 CNN-BiGRU-CRF 的事件因果关系抽取

郑巧夺, 吴贞东, 邹俊颖

(四川师范大学 计算机科学学院, 成都 610101)

摘 要: 针对现有事件因果关系抽取方法关系边界识别能力弱和文本语义表征不足的问题, 提出一种基于双层 CNN-BiGRU-CRF 深度学习模型的事件因果关系抽取方法。将因果关系抽取任务转换为两次序列标注任务分别由两层 CNN-BiGRU-CRF 模型完成, 上层模型用于识别事件因果关系语义角色词, 其标注结果作为特征输入下层模型划分因果关系边界。在每层模型中, 采用突发事件样本数据对 BERT 模型进行微调, 形成文本表示模型以获取语义特征向量矩阵, 利用卷积神经网络和双向门控循环单元分别提取局部和全局深层特征, 并将上述特征在每个时间序列步进行线性加权融合以增强语义表征能力, 最终基于残差思想将高分度特征输入 CRF 模型解码完成序列标注任务。在中文突发事件语料集上的实验结果表明, 与 BiLSTM-Att-规则特征、GAN-BiGRU-CRF 等因果关系抽取方法相比, 该方法的事件因果关系抽取效果更好, F 值达到 91.81%, 能有效实现事件因果关系的准确抽取。

关键词: 因果关系抽取; 深度学习; 卷积神经网络; 特征融合; 突发事件

开放科学(资源服务)标志码(OSID):



中文引用格式: 郑巧夺, 吴贞东, 邹俊颖. 基于双层 CNN-BiGRU-CRF 的事件因果关系抽取[J]. 计算机工程, 2021, 47(5): 58-64, 72.

英文引用格式: ZHENG Qiaoduo, WU Zhendong, ZOU Junying. Event causality extraction based on two-layer CNN-BiGRU-CRF[J]. Computer Engineering, 2021, 47(5): 58-64, 72.

Event Causality Extraction Based on Two-Layer CNN-BiGRU-CRF

ZHENG Qiaoduo, WU Zhendong, ZOU Junying

(College of Computer Science, Sichuan Normal University, Chengdu 610101, China)

[Abstract] The existing event causality extraction methods have poor performance in relationship boundary recognition and are limited by insufficient semantic representation of texts. To solve the problem, this paper proposes a method of event causality extraction based on two-layer CNN-BiGRU-CRF deep learning model. The task of causality extraction is transformed into two sequential annotation tasks, which are respectively completed by two-layer CNN-BiGRU-CRF model. The upper model is used to identify the semantic role words of event causality, and the annotation results are input as features into the lower model to divide the causality boundary. In both models, the emergency sample data is used to fine-tune the BERT model to form a text representation model to obtain the semantic feature vector matrix. Convolutional Neural Network(CNN) and Bi-directional Gated Recurrent Unit(BiGRU) are used to extract local and global deep-layer features respectively, and the above features are linearly weighted and fused in each time series step to enhance the semantic representation. Finally, based on the idea of residuals, high resolution features are input into the CRF model for decoding and to complete the task of sequence annotation. Experimental results on Chinese emergency corpus show that compared with the causality extraction methods such as BiLSTM-Att-rule feature and GAN-BiGRU-CRF, the proposed method has better causality extraction performance with F value reaching 91.81%. It can effectively achieve the accurate extraction of event causality.

[Key words] causality extraction; deep learning; Convolutional Neural Network(CNN); feature integration; emergency event

DOI: 10.19678/j.issn.1000-3428.0057361

0 概述

事件作为文本类信息的重要表现形式, 近年来在

自然语言处理(Natural Language Processing, NLP)领域应用较广泛, 各类基于事件的理论模型^[1]和应用技术

基金项目: 国家自然科学基金青年科学基金项目(11905153)。

作者简介: 郑巧夺(1993—), 男, 硕士研究生, 主研方向为自然语言处理; 吴贞东, 副教授、硕士; 邹俊颖, 讲师、博士。

收稿日期: 2020-02-10 修回日期: 2020-04-07 E-mail: qiaoduo@sina.com

相继出现^[2-3]。一个文本通常由多个事件组成, 这些事件由同一个主题串联, 事件之间存在因果关系、时序关系等多种语义关系。其中, 事件因果关系是较为常见和重要的一类关系, 其体现出理解知识后进行推理的过程, 反映了事件先后相继、由因及果的联系^[4]。准确、高效地理解文本事件中的因果关系, 对于掌握文本中事件的演变过程、预判事件后期发展以及文本处理的上层应用具有重要的现实意义。

在事件因果关系识别方面, 国内外已提出众多基于规则的方法和基于机器学习的方法^[5-6], 但这两类方法存在特征工程代价大、事件因果关系识别覆盖面低等问题。文献[7]提出一种层叠机器学习模型, 可分步解决事件因果关系识别的复杂问题, 但其输入与输出缺乏深层语义特征, 不能结合上下文推理事件隐式因果关系。文献[8]通过使用深度学习技术, 有效解决了事件因果关系抽取过程中深层语义信息推理时隐式因果关系缺失的问题。文献[9]将双向长短时记忆(Bidirectional Long Short-Term Memory, BiLSTM)神经网络模型用于事件因果关系的识别, 充分结合语义特征推理出上下文因果关系。由此可见, 深度学习在事件因果关系识别领域已取得一定成果, 但还存在关系边界识别模糊和文本表征能力不足等问题。

本文提出一种基于双层CNN-BiGRU-CRF模型的事件因果关系抽取方法。将事件因果关系抽取分解为两次序列标注任务并由双层CNN-BiGRU-CRF模型完成, 结合基于Transformer的双向编码器表示(Bidirectional Encoder Representation from Transformer, BERT)模型与卷积神经网络(Convolutional Neural Network, CNN)增强事件文本特征的表达能力, 在模型中引入残差思想^[10-11], 将文本信息局部特征与全局特征在每个时间序列步中进行线性加权融合, 以提升模型对重要语义特征的提取能力。

1 相关理论

1.1 事件

在自动内容提取(Automatic Content Extraction, ACE)中, 事件被定义为包含参与者的具体发生的事情, 用来描述状态的改变。目前, 对事件的研究主要分为事件内部组成结构研究和事件外部关联研究。事件内部组成结构研究是指对时间、地点和参与者等事件论元进行研究, 事件外部关联研究是指对事件因果关系和时序关系等多种语义关系进行研究。

1.2 事件因果关系

事件因果关系由原因事件和结果事件两部分构成。根据文本上下文是否出现因果连接词, 分为带标记因果关系和无标记因果关系^[7]。根据文本上下文是否同时出现原因和结果, 分为显式因果关系(原因和结果同时出现)和隐式因果关系(只出现原因或结果)。显式因果关系又分为一因一果、一因多果、多因一果和多因

多果4类。对于隐式因果关系, 需根据文本上下文和深层语义推断事件隐藏的原因或结果。

1.3 BERT 模型

BERT^[12]模型的训练过程分为两个阶段。第一阶段采用双层双向Transformer模型并基于MLM和NSP两种策略进行预训练, 以达到双向和深度动态地利用文本上下文信息的目的。MLM策略指给定1个句子, 随机隐藏句子中部分词, 根据剩余词预测被隐藏的部分, 使模型能结合上下文语义信息推测句子中任何词。NSP策略指给定文本中任意2个句子, 判断其顺序关系并构建二分类模型, 使得BERT模型能学习句子级别关系。第二阶段将BERT模型进行微调应用于下游任务, 以适应领域任务的需求。

BERT模型由多层Transformer编码模块堆叠而成。Transformer编码单元是BERT模型的核心, 其结构如图1所示。其中, 输入文本的向量矩阵由字符表征、句子表征和位置表征构成。

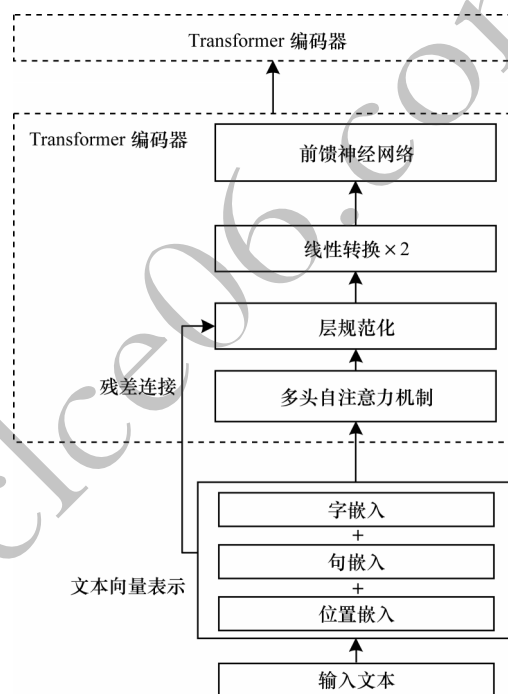


图1 Transformer编码单元结构

Fig.1 Structure of Transformer encoder unit

将向量矩阵输入多头自注意力(Multi-head Self-attention)模块, 利用目标字特征向量 Q 、目标字上下文特征向量 K 和原始向量 V ^[13]计算并增强语义表达。其中, 自注意力机制的计算思想为: Q 与 K 先进行相似性计算获取注意力权重, 再基于注意力权重向量重构原始向量 V , 以区分 V 中目标词上下文对其不同的重要程度。相关计算公式如下:

$$\text{attention_output} = \text{Attention}(Q, K, V) \quad (1)$$

$$\text{Attention}(Q, K, V) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (2)$$

其中, d_k 为向量维度。将 Q 与 K 作用生成的注意力向

量通过 Softmax 激活函数形成注意力得分, 然后对 V 中每个向量与其对应的注意力得分的点乘结果进行加权求和, 得到输入目标字的自注意力加权后的原始向量。

为扩展模型处理不同场景的能力, 文献[13]提出先对 Q 、 K 和 V 进行 h 次不同的投影, 映射维度均为 d_k , 再拼接不同自注意力模块输出结果, 经过线性映射输出形成多头自注意力, 相关计算公式如下:

$$\text{MultiHead}(Q, K, V) = \text{Concat}(\text{head}_1, \text{head}_2, \dots, \text{head}_h)W^Q \quad (3)$$

$$\text{head}_i = \text{Attention}(QW_i^Q, KW_i^K, VW_i^V) \quad (4)$$

多头自注意力机制模块输出后, 进行残差连接与标准化处理。残差连接使模型具备深度能力, 标准化处理是使神经网络节点进行 0 均值 1 方差映射^[14], 相关计算公式如下:

$$\text{sub_layer_output} = \text{LayerNorm}(x + (\text{SubLayer}(x))) \quad (5)$$

为增强模型线性表达能力, 对每个字的增强语义向量进行两次线性变换, 输入前馈神经网络 FFN, 完成 Transformer 编码。若将上一步输出表示为 Z , 则 FFN 表达式为:

$$\text{FFN}(Z) = \max(0, ZW_1 + b_1)W_2 + b_2 \quad (6)$$

由于 Transformer 模型摒弃了传统循环神经网络 (Recurrent Neural Network, RNN) 的结构及其对时间序列抽象化的能力, 因此在特征表示中融入相对位置信息来保持其对时序信息的处理能力, 具体相对位置信息如下:

$$\text{PE}_{(\text{pos}, 2i)} = \sin(\text{pos}/10000^{2i/d_{\text{model}}}) \quad (7)$$

$$\text{PE}_{(\text{pos}, 2i+1)} = \cos(\text{pos}/10000^{2i/d_{\text{model}}}) \quad (8)$$

其中, i 为单词维度, pos 为单词位置, d_{model} 为 embedding 维度, \sin 函数和 \cos 函数均表示不同频率的计算。

1.4 双向门控循环单元

门控循环单元 (Gated Recurrent Unit, GRU) 能有效克服循环神经网络中的梯度问题与长期依赖问题^[15], 其在长短时记忆 (Long Short-Term Memory, LSTM) 神经网络的基础上合并细胞状态与隐藏层状态, 并将遗忘门与输入门结合为更新门, 输出门更名为重置门。GRU 和 LSTM 均通过门结构保留重要特征, 从而实现信息的长距离传播。由于 GRU 结构更简单, 其相较 LSTM 少一个门且参数量也更少, 因此 GRU 的整体训练速度快于 LSTM^[16-17]。GRU 单元结构如图 2 所示。

在图 2 中, z_t 和 r_t 分别表示控制神经元信息读写的更新门和重置门。更新门用于控制前一时刻的状态被带入当前状态的程度, 重置门用于控制前一时刻状态信息被写入当前候选集 \tilde{h}_t 的程度, 其包含确定当前记忆内容与确定当前时刻隐藏层保留信息^[18]两部分, 具体计算公式如下:

$$z_t = \sigma(w_z \times [h_{t-1}, x_t]) \quad (9)$$

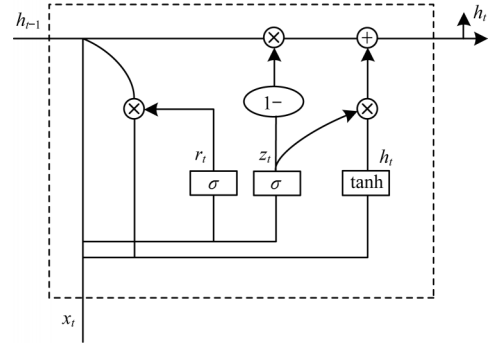


图2 门控循环单元结构

Fig.2 Structure of gated recurrent unit

$$r_t = \sigma(w_r \times [h_{t-1}, x_t]) \quad (10)$$

$$\tilde{h}_t = \tanh(w_h \times [r_t \times h_{t-1}, x_t]) \quad (11)$$

$$h_t = (1 - z_t) \times h_{t-1} + z_t \times \tilde{h}_t \quad (12)$$

$$y_t = \sigma(w_o \times h_t) \quad (13)$$

其中, h_t 为当前 GRU 单元的激活值, y_t 为 t 时刻的输出, σ 和 \tanh 为激活函数, w_z 、 w_r 、 w_h 和 w_o 为对应的权值参数, \tilde{h}_t 为 GRU 单元在 t 时刻的激活值, 其同时受重置门 r_t 、 $t-1$ 时刻 GRU 单元激活值 h_{t-1} 以及当前时刻输入 x_t 的控制。

由于 GRU 网络不能有效利用输入文本的上下文信息, 因此增加反向输入的逆序 GRU 对文本的逆向语义进行学习, 然后将正反向 GRU 合并构建双向门控循环单元 (Bidirectional Gated Recurrent Unit, BiGRU) 以充分利用上下文信息挖掘因果关系, 并采用正向与反向传递连接的方法进行合并, 计算公式如下:

$$H = \sum_{i=1}^t \text{concat}(H_f, H_b) \quad (14)$$

其中, H_f 、 H_b 分别为正向与反向 GRU 在时刻 t 的输出向量, H 为正向与反向 GRU 隐层全部时间序列连接后的最终输出。

1.5 卷积神经网络

卷积神经网络^[19]主要基于动物视觉的局部感受野原理而构建。本文利用 CNN 特性提取输入句子中每个字词的局部特征。若给定序列长度为 t 的句子 $S = \{w_1, w_2, \dots, w_t\}$, 其经过 BERT 模型生成对应的字向量 $V = \{v_1, v_2, \dots, v_t\}$, 每个字向量 v_i 对应维度 $d=768$, 且任意句子均能构成向量矩阵 $R^{n \times d}$ 并作为 CNN 的输入, 在因果关系抽取任务中, $K = \{k_1, k_2, \dots, k_n\}$ 为卷积核集合, n 为卷积核数量, 则对输入的字向量矩阵利用卷积核 k_i 进行卷积操作后, 目标字向量可形成局部特征向量 f_i , 其表达式为:

$$f_i = f(k_i v_{t:t+j-1} + b) \quad (15)$$

其中, v 为输入的字向量矩阵, j 为卷积核 k_i 的窗口大小, b 为偏置值, f 为 ReLU 激活函数。

通过式 (15) 使整个卷积核集合 K 作用于窗口中心字向量, 并对该字向量 v_i 形成不同的局部特征 f , n 个卷积操作表示如下:

$$\mathbf{v}_i = \mathbf{F} = \{\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_n\} \quad (16)$$

其中, \mathbf{F} 为目标字 \mathbf{v}_i 经过 n 个卷积操作后形成的上下文特征集。

由于 \mathbf{F} 是多特征高维度向量, 因此采用池化操作降低其维度。因果语义角色词作为显著特征, 可采用最大池化操作保留特征, 其表达式为:

$$\mathbf{m} = \max(\mathbf{F}) \quad (17)$$

将每个特征向量 \mathbf{f}_i 经最大池化操作后保留的特征 \mathbf{m} 进行全连接以固定其维度输出, 最终形成中心字向量 \mathbf{v}_i 的局部上下文表征, 其表达式为:

$$\mathbf{v}_i = \mathbf{M} = \{\mathbf{m}_1, \mathbf{m}_2, \dots, \mathbf{m}_n\} \quad (18)$$

在中心字向量卷积过程中, 根据窗口大小, 用“0”补齐每个句子边界以避免出现窗口中心字左边或者右边无字的情况, 导致卷积到空值。重复上述中心字向量卷积过程, 同时, 对输入长度为 t 的句子向量矩阵, 利用卷积集合 K 扫描整个句子序列形成整个句子的局部特征集, 其表达式如下:

$$\mathbf{S} = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_t\} = \{\mathbf{M}_1, \mathbf{M}_2, \dots, \mathbf{M}_t\} \quad (19)$$

1.6 CRF 模型

GRU 仅考虑长远的上下文信息, 忽略了标签之间的依赖关系。条件随机场 (Conditional Random Field, CRF) 模型能在考虑标签之间约束关系的情况下, 获得全局最优标签序列^[20]。

CRF 解码标签序列描述的条件概率模型为 $P = (Y|X)$, 其中, Y 为输出变量 (状态序列), X 为输入变量 (观测序列)。利用训练数据通过极大似然估计可得到条件概率模型 $P(Y|X)$, 输入 x 后计算得到条件概率 $P(y|x)$ 最大的输出序列 y 。

在训练 CRF 模型时, 使用极大似然函数估计形成条件概率模型, 其对数似然函数为:

$$L = \sum_{i=1}^n \ln(P(y_i | x_i)) + \frac{\lambda}{2} \|\theta\|^2 \quad (20)$$

$$P(y|x) = \frac{1}{Z(x)} \exp \sum_{k=1}^K \mathbf{w}_k f_k(y, x) \quad (21)$$

$$Z(x) = \sum_y \exp \sum_{k=1}^K \mathbf{w}_k f_k(y, x) \quad (22)$$

其中, f_k 为关于 x 和 y 的特征函数, \mathbf{w}_k 为训练权重, $Z(x)$ 为归一化因子。通过极大似然函数训练获取合适的参数使对数似然 P 最大, 解码函数用于搜索条件概率最高时对应的标签序列 Y , 其表达式为:

$$Y = \arg \max p(x/y, w) \quad (23)$$

2 双层 CNN-BiGRU-CRF 模型

2.1 模型结构

本文提出的双层 CNN-BiGRU-CRF 模型结构如图 3 所示, 主要包括数据预处理层、上层 CNN-BiGRU-CRF 模

型、下层 CNN-BiGRU-CRF 模型以及输出。因果关系抽取任务被分解为语义角色词识别任务与边界特征词识别任务, 分别由上下两层 CNN-BiGRU-CRF 模型完成, 上层模型的输出序列作为新特征被传递给下层模型。

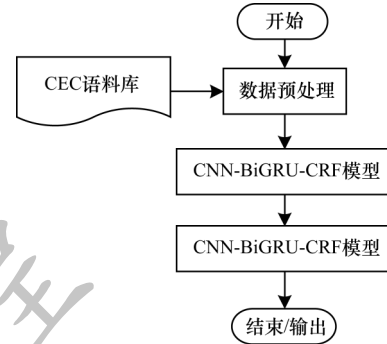


图3 本文模型结构

Fig.3 Structure of the proposed model

上层 CNN-BiGRU-CRF 模型用于完成分解任务中的因果关系语义角色词识别任务。采用 CEN-t 标注方案, 其中 C 表示原因事件, E 表示结果事件, N 表示无关事件, t 表示普通文本。基于因果关系的对应方式不同, 将上层 CNN-BiGRU-CRF 模型的标注结果作为特征输入下层 CNN-BiGRU-CRF 模型, 进一步划分因果关系边界。采用 BIO 标注方案, B 表示因果关系的开始词, I 表示因果关系的中间词或结尾词, O 表示无关联词。输入句子的预测输出序列格式为: 上层模型标注结果/下层模型标注结果。其中, 标注结果含义如表 1 所示。

表1 标注结果含义

Table 1 Meaning of annotation results

预测序列	一次标注含义	二次标注含义
t/O	非事件词	非因果关系边界词
C/B	原因事件词	因果关系边界开始词
C/I	原因事件词	因果关系边界中间词或结尾词
E/I	结果事件词	因果关系边界中间词或结尾词
N/O	其他事件词	非因果关系边界词

由两次分解任务的序列标注结果可识别因果关系对, 具体为: 1) 结合模型预测输出的序列和表 1 中一次标注结果含义, 可确定原因事件词、结果事件词等因果语义角色词; 2) 结合模型预测输出的序列和表 1 中二次标注结果含义, 可确定因果关系边界词, 即因果关系对的对应关系。

2.2 CNN-BiGRU-CRF 模型

CNN-BiGRU-CRF 模型结构如图 4 所示, 主要包括 embedding 层、CNN-BiGRU 层和 CRF 层。由图 4 可以看到, 本文列举的输入句子及其经上下层模型标注后的输出结果。

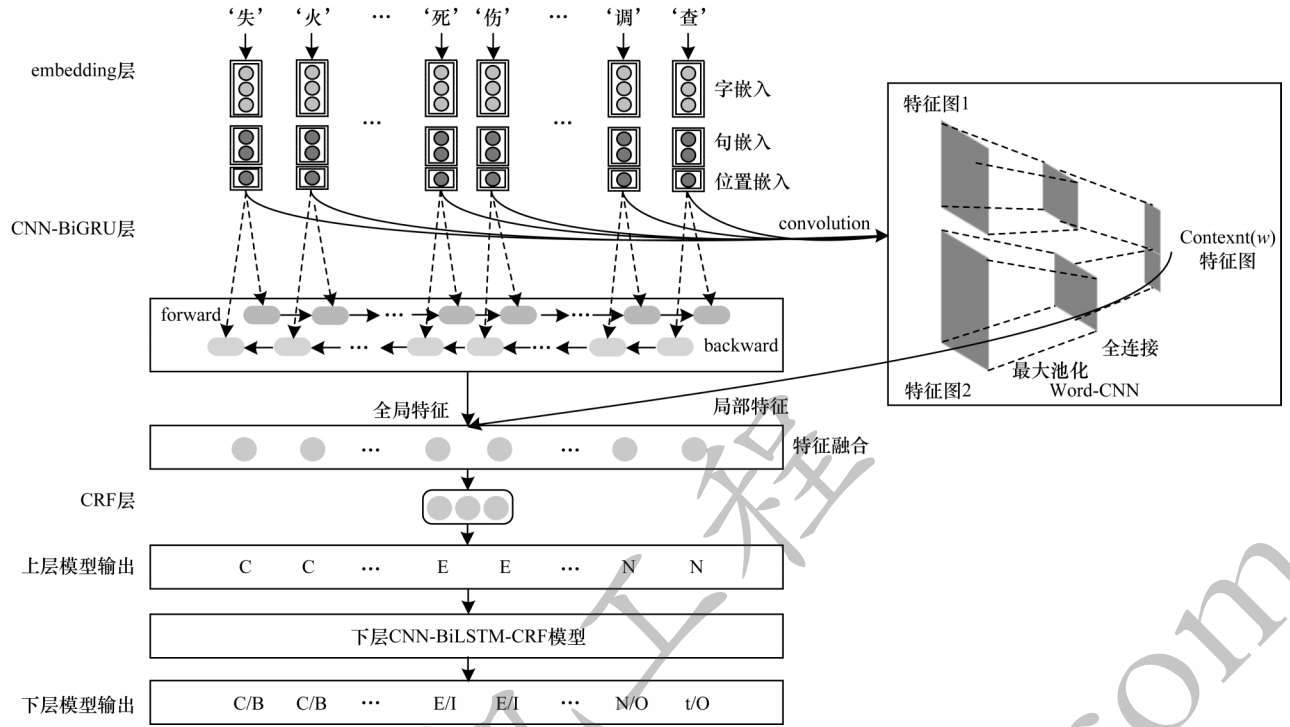


图4 CNN-BiGRU-CRF模型结构

Fig.4 Structure of CNN-BiGRU-CRF model

2.2.1 embedding层

以突发事件因果关系抽取为例,使用突发事件样本数据对BERT模型进行微调以生成突发事件域的文本表示模型,再利用该模型对输入本层的样本句子生成特征向量矩阵,并加上首尾标签与位置索引。若将输入本层的样本句子中每个字用 w_i 表示,则句子 $S=\{w_1, w_2, \dots, w_t\}$ 由BERT模型生成的向量表示为矩阵 V ,其表达式如下:

$$V = \begin{bmatrix} \text{SEP}_1 & \text{SEP}_2 & \dots & \text{SEP}_{768} \\ v_{11} & v_{12} & \dots & v_{1 \times 768} \\ v_{21} & v_{22} & \dots & v_{2 \times 768} \\ \vdots & \vdots & \dots & \vdots \\ v_{t1} & v_{t2} & \dots & v_{t \times 768} \\ \text{CLS}_1 & \text{CLS}_2 & \dots & \text{CLS}_{768} \\ \text{index}_1 & \text{index}_2 & \dots & \text{index}_{768} \end{bmatrix} \quad (24)$$

其中, t 为序列长度,768为字向量维度,SEP为句首标签,CLS为句尾标签,index为位置索引, v 为BERT模型生成的字向量。输入每个样本句子通过式(24)可转换为对应的特征向量矩阵。

2.2.2 CNN-BiGRU层

将上层生成的序列向量矩阵同时输入CNN与BiGRU可分别提取句子深层局部特征与长距离特征。为丰富重点语义特征的表达信息,借鉴残差思想将提取的局部特征和全局特征在每一个时间序列步进行融合。假设句子序列长度为 t ,经CNN提取的特征 $S\text{-CNN}=\{M_1, M_2, \dots, M_t\}$,经BiGRU形成输出 $S\text{-BiGRU}=\{H_1, H_2, \dots, H_t\}$,则在每一个时间序列步融合的特征向量表达式如下:

$$C = \text{TimeDistributed}(S\text{-CNN}, S\text{-BiGRU}) \quad (25)$$

其中, C 为融合后的特征向量,TimeDistributed为按照时间序列进行特征整合的残差融合函数。按照时间序列 t ,采用线性加权法将句子局部特征 $S\text{-CNN}$ 的 M_i 依次整合到句子全局特征 $S\text{-BiGRU}$ 的 H_i 中,计算公式如下:

$$c_i = \sum_{j=1}^i w \times \text{concat}(M_j, H_j) \quad (26)$$

对于火灾、地震、食物中毒等原因事件角色词,以及死亡、受伤等结果事件角色词和部分显著的边界特征词,可在句子按照时间序列进行特征整合的过程中加大其特征权重 w 。

由上述可知,利用残差思想来丰富句子特征以及采用线性加权法突出句子中重点语义特征,可得到因果关系高区分度特征,使整个模型更容易捕获重点语义单元,从而优化模型抽取效果。

2.2.3 CRF层

将上层输出的每个字所对应全部标签的发射概率矩阵和初始化的转移概率矩阵作为原始CRF模型的训练参数,采用最大似然函数训练CRF模型可确定标签序列。

3 实验与结果分析

3.1 实验语料集

本文实验选取上海大学实验室自制的开源中文突发事件语料库(Chinese Emergency Corpus, CEC)作为语料集。CEC语料集基于国务院颁布的分类体系所得到,共包括地震、火灾、食物中毒、交通事故以及恐怖袭击5个类别。其中,从互联网上收集的CEC语料共332篇,均采用XML语言作为标注格式,其对突发事件、突发事件要素以及突发事件对应关系的标注较全面。

通过去HTML标签、格式处理、事件句转换、分词、分字以及数据集自动标注等数据预处理操作将CEC语料集转换为模型训练的样本数据,并对文本数据进行增强处理,得到的CEC语料集中有11 908个事件,1 093对因果关系,共构成4 446条因果关系样本数据,按照7:1:2的数量比例分为训练集、验证集与测试集。CEC语料集具体数量信息与本文模型超参数设置分别如表2和表3所示。

表2 CEC语料集具体数量信息

Table 2 Specific quantity information of CEC corpus

类别	地震	火灾	交通事故	恐怖袭击	食物中毒	总数
文本	124	150	170	98	122	664
事件	2 004	2 432	3 604	1 646	2 222	11 908
原因事件	204	128	438	176	184	1 130
结果事件	442	166	614	254	176	1 652

表3 本文模型超参数设置

Table 3 Hyperparameter setting of the proposed model

参数	参数设置
学习率	0.001
迭代次数	80
批大小	32
Dropout	0.5
GRU层数	2
GRU隐藏单元数	128
BERT embedding维度	768
序列长度	200
CNN内核尺寸	5
CNN过滤器个数	32
CNN激活函数	ReLU

3.2 评价指标

为验证本文方法在事件因果关系上的抽取效果并便于分析,以整体标注准确率、召回率和F值作为抽取效果的评价指标。在评估事件因果关系抽取正确性的过程中,若事件因果关系成对出现,则需同时考虑一对事件因果关系中原因事件标注与结果事件标注,不能仅以标注结果为准,因此在判断一对事件因果关系抽取的正确性时,要求原因事件标注和结果事件标注同时正确(如果是一对多个结果,则要求多个结果全部正确),否则视为抽取错误。

3.3 结果分析

3.3.1 不同模型的对比

为验证本文模型的性能,分别将本文模型与其他模型从单层和双层、是否引入CNN两方面进行对比,并对不同文本表示模型进行比较分析。在CEC语料集上的事件因果关系抽取中,进行5次实验并

取每次实验的平均值作为最终值,不同模型的实验结果如表4所示。其中,前缀S-与D-分别代表单层模型与双层模型,模型1为基础模型。由表4可知,双层模型5的评价指标高于单层模型2,其原因是双层模型对事件因果关系抽取任务进行分步处理,其第二层结构专用于学习边界特征,提升整个模型识别因果关系边界能力并简化了单个深度学习模型任务复杂性。引入残差结构(CNN-BiGRU输出线性加权整合后结构)的模型5的各评价指标均高于未引入残差结构的模型3,模型5在增强特征表达能力后F值提高5.02个百分点,特别是召回率提升7.76个百分点,说明残差结构的应用有效丰富了因果关系抽取任务中重点语义特征,提高模型的特征提取性能。微调BERT模型后所得模型5的F值较模型4提高2.95个百分点,验证了BERT模型的有效性。

表4 不同模型的评价指标结果

Table 4 Evaluation index results of different models %

序号	模型	准确率	召回率	F值
1	S-BiGRU-CRF-BERT	83.58	81.73	82.64
2	S-CNN-BiGRU-CRF-BERT	84.52	88.15	86.30
3	D-BiGRU-CRF-BERT	87.83	85.77	86.79
4	D-CNN-BiGRU-CRF-w2v	88.09	89.65	88.86
5	D-CNN-BiGRU-CRF-BERT	90.16	93.53	91.81

本文针对CNN重要参数调节对模型5的影响进行实验,取5次实验的平均值作为最终值,实验结果如图5所示。可以看出,在其他参数相同的情况下,CNN的卷积核窗口大小为5时模型5的F值最高,卷积核数量为32时模型5的F值趋于稳定。由上述结果可知:CNN卷积核窗口设置过大不利于提取句子局部特征,而卷积核窗口设置过小又无法充分提取句子局部特征,所提取特征缺少上下文语义信息;若卷积核数量越多,则CNN提取的特征平面信息越丰富,当卷积核数量为128时计算量较大,此时模型5的F值仅较卷积核数量为32时高出0.1个百分点。

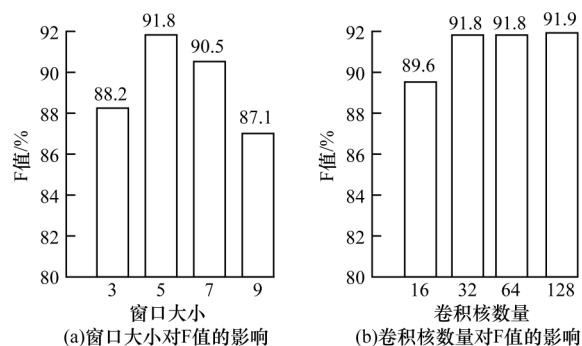


图5 CNN参数设置对模型5的F值影响

Fig.5 Influence of CNN parameter setting on F value of the model 5

3.3.2 不同方法的对比

将本文方法分别与基于双向BiLSTM网络的维吾尔事件因果关系抽取方法^[9](BiLSTM-Att-规则特征方法)、基于注意力机制的BiLSTM网络模型突发事件演化关系抽取方法^[10](BiLSTM-Att-演化模型方法)以及融合对抗学习的因果关系抽取方法^[21](GAN-BiGRU-CRF方法)等先进的因果关系抽取方法进行对比分析,实验结果如表5所示。可以看出,本文方法的召回率较其他方法更高,这是因为本文更注重丰富与突出重点因果关系语义特征。本文方法的F值高于其他3种方法,其原因为:1)本文采用的双层模型比其他方法采用的单层模型更适用于因果关系抽取任务,能充分挖掘出语义特征细节信息;2)本文方法采用的残差特征加权融合方式突出了因果关系的语义角色特征与关系边界特征,使模型更易训练和学习;3)相较其他方法的文本表示模型,本文引入的BERT模型具有更强大的文本表示能力。

表5 不同方法的评价指标结果

Table 5 Evaluation index results of different methods %

方法	准确率	召回率	F值
BiLSTM-Att-规则特征方法	89.19	83.19	86.09
BiLSTM-Att-演化模型方法	86.30	87.90	87.10
GAN-BiGRU-CRF方法	93.17	89.25	91.17
本文方法	90.16	93.53	91.81

4 结束语

本文提出一种基于残差思想的双层模型因果关系抽取方法。对复杂的事件因果关系抽取问题进行分解,分别采用两个深度学习模型分步处理以增强关系边界的识别。在模型内部结构中,利用BERT结构提取丰富的语义特征,再基于残差思想结合CNN与BiGRU模型将特征进行线性加权融合增强语义表征能力。在中文突发事件语料集上的实验结果表明,该方法的事件因果关系抽取效果较现有的事件因果关系抽取方法更好,能有效解决关系抽取覆盖面低与语义表征不足的问题。后续将进一步扩充实验语料集,从多领域特征融合着手提升模型的抽取效果。

参考文献

- [1] HAN Songqiao, HAO Xiaoling, HUANG Hailiang. An event-extraction approach for business analysis from online Chinese news[J]. Electronic Commerce Research and Applications, 2018, 28(3): 244-260.
- [2] ZHANG Yijie, LI Peifeng, ZHU Qiaoming. Joint learning for identifying temporal and causal relations between events[J]. Computer Engineering, 2020, 46(7): 65-71. (in Chinese) 张义杰, 李培峰, 朱巧明. 面向事件时序和因果关系识别的联合学习方法[J]. 计算机工程, 2020, 46(7): 65-71.
- [3] XIANG Wei, WANG Bang. Survey of Chinese event extraction research[J]. Computer Technology and Development, 2020, 30(2): 1-6. (in Chinese) 项威, 王邦. 中文事件抽取研究综述[J]. 计算机技术与发展, 2020, 30(2): 1-6.
- [4] ZHOU Xingfa. Causal relation extraction of Uyghur events [D]. Xinjiang: Xinjiang University, 2018. (in Chinese) 周兴发. 维吾尔语事件间因果关系抽取[D]. 新疆: 新疆大学, 2018.
- [5] KIM H B, JOUNG J, KIM K. Semi-automatic extraction of technological causality from patents [J]. Computers and Industrial Engineering, 2018, 115(1): 532-542.
- [6] ZHAO Sendon, LIU Tin, ZHAO Sicheng, et al. Event causality extraction based on connectives analysis [J]. Neurocomputing, 2016, 173(1): 1943-1950.
- [7] ZHONG Jun, YU Long, TIAN Shengwei, et al. Causal relation extraction of Uyghur emergency events based on cascaded model [J]. Acta Automatica Sinica, 2014, 40(4): 771-779. (in Chinese) 钟军, 禹龙, 田生伟, 等. 基于双层模型的维吾尔语突发事件因果关系抽取[J]. 自动化学报, 2014, 40(4): 771-779.
- [8] LI Peifeng, HUANG Yilong, ZHU Qiaoming. Global optimization to recognize causal relations between events[J]. Journal of Tsinghua University (Science and Technology), 2017, 57(10): 1042-1047. (in Chinese) 李培峰, 黄一龙, 朱巧明. 使用全局优化方法识别中文事件因果关系[J]. 清华大学学报(自然科学版), 2017, 57(10): 1042-1047.
- [9] TIAN Shengwei, ZHOU Xingfa, YU Long, et al. Causal relation extraction of Uyghur events based on bidirectional long short-term memory model [J]. Journal of Electronics and Information Technology, 2018, 40(1): 200-208. (in Chinese) 田生伟, 周兴发, 禹龙, 等. 基于双向LSTM的维吾尔语事件因果关系抽取[J]. 电子与信息学报, 2018, 40(1): 200-208.
- [10] PENG Yuqing, SONG Chubai, YAN Qian, et al. Research on Chinese text classification based on hybrid model of VDCNN and LSTM [J]. Computer Engineering, 2018, 44(11): 190-196. (in Chinese) 彭玉青, 宋初柏, 闫倩, 等. 基于VDCNN与LSTM混合模型的中文文本分类研究[J]. 计算机工程, 2018, 44(11): 190-196.
- [11] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, et al. Deep residual learning for image recognition [C]//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2016: 25-32.
- [12] DEVLIN J, CHANG M W, LEE K, et al. BERT: pre-training of deep bidirectional transformers for language understanding [EB/OL]. [2019-12-25]. <https://arxiv.org/abs/1810.04805>.
- [13] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [EB/OL]. [2019-12-25]. https://www.researchgate.net/publication/317558625_Attention_Is_All_You_Need.

(下转第72页)

(上接第 64 页)

- [14] JIMMY L B, JAMIE R K, GEOFFREY E. H, et al. Layer Normalization[EB/OL]. [2019-12-25]. <https://arxiv.org/abs/1607.06450>.
- [15] CHO K, VAN M B, GULCEHRE C, et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation[EB/OL]. [2019-12-25]. <https://www.oalib.com/paper/4082023#.YD8uWPmuVB4>.
- [16] YANG Piao, DONG Wenyong. Chinese named entity recognition method based on BERT embedding [J]. Computer Engineering, 2020, 46(4): 40-45, 52. (in Chinese) 杨飘, 董文永. 基于 BERT 嵌入的中文命名实体识别方法[J]. 计算机工程, 2020, 46(4): 40-45, 52.
- [17] RANA R. Gated recurrent unit for emotion classification from noisy speech [EB/OL]. [2019-12-25]. https://www.researchgate.net/publication/311842569_Gated_Recurrent_Unit_GRU_for_Emotion_Classification_from_Noisy_Speech.
- [18] LI Lishuang, WAN Jia, ZHENG Jieqiong, et al. Biomedical event extraction based on GRU integrating attention mechanism[J]. BMC Bioinformatics, 2018, 19(9): 177-184.
- [19] RUSSAKOVSKY O, DENG J, SU H, et al. Imagenet large scale visual recognition challenge[J]. International Journal of Computer Vision, 2015, 115(3): 211-252.
- [20] LAMPLE G, BALLESTEROS M, SUBRAMANIAN S, et al. Neural architectures for named entity recognition[EB/OL]. [2019-12-25]. https://www.researchgate.net/publication/305334469_Neural_Architectures_for_Named_Entity_Recognition.
- [21] FENG Chong, KANG Liqi, SHI Ge, et al. Causality extraction with GAN[J]. Acta Automatica Sinica, 2018, 44(5): 811-818. (in Chinese) 冯冲, 康丽琪, 石戈, 等. 融合对抗学习的因果关系抽取[J]. 自动化学报, 2018, 44(5): 811-818.

编辑 宋 圆