

基于事件语境的文本情感原因对特征提取

叶苑莉, 张 灵[†], 陈云华

(广东工业大学 计算机学院, 广东 广州 510006)

摘要: 现有的情感原因对提取任务(ECPE)大多采用将情感从句逐一与原因从句匹配的方法, 或专注于候选对的排序方法, 忽略了影响情感因果关系成立的从句的事件语境, 导致模型在理解情感因果关系时产生偏差, 并且无法捕捉长距离的因果关系. 为此, 本文提出了基于注意力机制和情感从句卷积核的分层模型, 将原始文档的事件语境特征嵌入到情感原因对特征提取器中, 以创建一个集成和增强的特征. 首先, 将情感分析得到的情感从句类别特征作为卷积核. 然后, 利用文档的事件语境特征提取情感原因对. 本文方法在中文数据集的F1分数上有1.38%~6.08%的提升, 在英文数据集的F1分数上有2.35%~7.27%的提升, 说明情感分析和因果事件语境对于情感原因对提取的有效性.

关键词: 情感原因对; 情感分析; 注意力机制; 事件语境

引用格式: 叶苑莉, 张灵, 陈云华. 基于事件语境的文本情感原因对特征提取. 控制理论与应用, 2022, 39(7): 1315 – 1323

DOI: 10.7641/CTA.2021.10408

Feature extraction of emotion-cause pairs in text based on event context

YE Yuan-li, ZHANG Ling[†], CHEN Yun-hua

(College of Computer, Guangdong University of Technology, Guangzhou Guangdong 510006, China)

Abstract: For emotion-cause pair extraction (ECPE) task, most of the existing works only match the emotion clause and the corresponding cause clause one by one, or focus on sorting candidate pairs, ignoring the event context among clauses that would impact significantly on the establishment of emotion causality. This leads to the deviation of the model in learning emotion causality. They also fail to capture long-span causality hidden among plenty of clauses. To address this issue, we propose a hierarchical model based on attention mechanism and the convolution kernel of emotion clause. In our method, we derive event context features from the original documents and embed them into emotion-cause pair feature extractors to create an integrated and enhanced feature. Firstly, the category features of emotion clause obtained from sentiment analysis are used as convolution kernel. Then, the event context features of documents are presented to extract emotion-cause pairs. The experimental results show that the F1 score on the Chinese benchmark emotion cause corpus are improved from 1.38% to 6.08% compared with the state-of-the-art approaches. Meanwhile, the F1 score on the English benchmark emotion cause corpus are improved from 2.35% to 7.27% compared with the state-of-the-art approaches. They verify the effectiveness of sentiment analysis and causal event context in ECPE.

Key words: emotion-cause pair; sentiment analysis; attention mechanism; event context

Citation: YE Yuanli, ZHANG Ling, CHEN Yunhua. Feature extraction of emotion-cause pairs in text based on event context. *Control Theory & Applications*, 2022, 39(7): 1315 – 1323

1 引言

情感原因提取任务(emotion cause extraction, ECE)^[1-2]是自然语言处理中具有重要价值的任务之一. 传统的情感分析(sentiment analysis)只关注情感极性的分类, 没有研究产生该情感的原因, 而任何情感的产生都是有原因的, 提取情感原因可以增加情感分

类模型的可解释性.

构成情感原因的的事件一般由多个单词构成, 因此词汇特征^[1]不适合ECE任务, 于是Gui等人^[2-3]提出了以从句为单位的ECE任务. Xia等人^[4]指出情感与原因之间存在相互联系, 不应只识别原因从句, 于是提出情感原因对提取任务(emotion-cause pair extra-

收稿日期: 2021-05-16; 录用日期: 2021-10-25.

[†]张灵. E-mail: 1252875930@qq.com; Tel.: +86 13535483864.

本文责任编辑: 赵冬斌.

广东省交通运输厅科技项目(科技-2016-02-030), 智慧交通跨域关联大数据挖掘与指导决策关键技术研究与应用基金项目(20151BAB207043), 广东省自然科学基金项目(2021A1515012233)资助.

Supported by the Scientific and Technological Project of Transportation Department of Guangdong Province (Sci-Tec-2016-02-030), the Research and Application of Key Technologies of Intelligent Transportation cross Domain Association Big Data Mining and Decision Guidance (20151BAB 207043) and the National Natural Science Foundation of Guangdong Province (2021A1515012233).

ction, ECPE), 将情感从句和原因从句匹配成情感原因对(emotion-cause pair, EC-Pair).

本文考虑在ECPE任务中EC-Pair在给定文档中的事件关系, 即构成情感因果关系的从句的事件语境, 提出了分层模型ECPE-EKC (ECPE-emotion kernel convolutional neural network). 特别地, 采用基于情感从句卷积核(emotion kernel, EK)构造的卷积神经网络(convolutional neural network, CNN)^[5]来提取EC-Pair的因果事件语境. 与基线模型^[4,6-7]相比, ECPE的F1分数^[8](衡量二分类模型精确度的一种指标)在中文数据集上有1.38%~6.08%的提升, 在英文数据集上有2.35%~7.27%的提升, 说明ECPE-EKC能有效地利用情感类别特征和因果事件语境来提高EC-Pair特征的提取质量.

2 相关工作

Lee等人^[1]最先提出ECE任务, 并采用规则的方法完成该任务. 随后, 许多基于规则^[9-14]和机器学习^[15-19]的方法被用于挖掘情感和原因之间的关系. 然而, 这些方法大多依赖于人工设计的特征, 难以涵盖所有复杂的语义关系(如, 嵌套关系、事件关系).

Gui等人^[2-3]采用多项式核完成从句级别ECE, 但该方法忽略了从句之间的关系, 一种解决方法是采用神经网络编码从句之间的语义关系^[20-24]; 另一种方案是提取情感词与上下文的关系^[25-26]. 由于ECE样本少, 仅从模型内部得到的情感特征有限, 于是引入了外部情感信息^[27-28]、全局标签信息^[29]等辅助特征.

Xia等人^[4]的ECPE-2Step模型有两步独立的反向传播网络, 第1步分别提取情感从句和原因从句, 第2步根据第1步的结果(包括正确和错误的结果)提取EC-Pair, 因此第1步反向传播的错误会延续到第2步, 于是诞生了许多端到端(同一个反向传播网络)的注意力框架. Tang等人^[6]的多层次注意力机制不仅能编码同一个从句的内部关系, 还能编码两个从句之间的关系; Ding等人^[30]的2D-Transformer模型可以过滤不相关的EC-Pair; Wei等人^[31]提出堆叠式图注意力和基于核的相对位置来对候选从句进行排序; Diao等人^[32]通过注意力机制将情感词与其同义词融合, 得到强化的从句特征表示. 除了注意力机制, 其他神经网络方法也被用于编码情感从句和原因从句之间的关系^[7,33-36].

上述方法在筛选EC-Pair时, 大多将情感从句逐一匹配原因从句, 即只专注于比较两个从句之间的关系, 没有考虑使EC-Pair因果关系成立的条件. Chen等人^[37]发现EC-Pair不一定在所有场景下都存在因果关系. 说明基于排序的筛选方法^[31]没有考虑因果关系成立的条件, 而注意力框架^[6,30-31]大多只专注于两个从句之间的关系. 但是文献^[37]需要额外制作负样本和人工标注, 不适合真实应用场景, 而本文的EK可以

自动提取从句之间的因果事件语境.

对于将词向量作为卷积核的方法, Li等人^[38]将因果触发词(cause, trigger, lead to)的词向量作为单词卷积核, 以提取原因特征. 虽然单词卷积核比普通卷积核更能提取原因特征, 但是他们的词向量池中因果框架的数量有限, 难以应对千变万化的因果关系, 而本文的EK是根据文档动态产生的, 因此更能适应不同的情感因果情景.

针对现有方法的不足, 本文提出了ECPE-EKC, 以此对ECPE存在的问题进行改进:

1) 加入情感类别的检测.

ECPE任务没有检测具体的情感类别, 模型只知道情感从句而不知道具体的情感, 相当于情感从句是黑盒子, 而模型还需要根据这个黑盒子识别原因从句, 因此原因的提取质量不高, 因为模型无法理解情感原因与具体情感的关系;

2) 增加事件的因果语境考虑.

目前工作^[4,6,31]在筛选EC-Pair时大多只关注情感从句和原因从句之间的关系, 忽略了使EC-Pair因果关系成立的从句的事件语境, 导致模型在理解因果关系时产生偏差(见第3.5小节). 本文的研究工作, 正是为了解决这个问题.

本文搭建分层框架的步骤为: 首先, 前两层使用双向长短期记忆网络(bi-directional long short-term memory, BiLSTM)^[39]分别提取从句级别和文档级别的语义特征, 得到用于情感分析的从句级别情感特征. 接着, 第3层, 通过注意力机制编码情感从句的类别特征, 作为该文档的EK. 然后, 第4层, 使用EK对从句进行卷积操作, 以提取原因事件特征. 最后, 通过融合分层的特征来提取EC-Pair.

3 ECPE-EKC模型

3.1 ECPE问题描述

设有若干篇文档, 第 $i(1 \leq i \leq M)$ 篇文档为 d_i , d_i 的第 $j(1 \leq j \leq N)$ 个从句为 $c_{i,j}$, $c_{i,j}$ 共有 P 个单词. 对于情感分析, 从句类别数为7, 包括6类情感标签^[40]和1类非情感标签^[6]; 对于ECPE, 需要同时正确提取情感从句和对应的原因从句.

如图1所示, 每篇文档以从句为单位进行文本预处理和词嵌入操作, 得到 $c_{i,j}$ 的词向量集 Z_j .

3.2 词编码器

在本文中, 词编码器由一个BiLSTM模块和一个注意力机制模块构成. 具体地, 记 $c_{i,j}$ 的BiLSTM输出为 $\mathbf{h}_{i,j}^o \in \mathbb{R}^{P \times 2H}$, $\mathbf{h}_{i,j}^i \in \mathbb{R}^{2 \times 1 \times H}$, 其中 $\mathbf{h}_{i,j}^o$ 包含 $c_{i,j}$ 中的每个单词的隐层状态, 表示该从句的单词特征, $\mathbf{h}_{i,j}^i$ 为BiLSTM最后一个状态的输出, 表示该从句的内部语义特征, H 为隐藏层特征维度. 经过注意力后得到 $c_{i,j}$ 的注意力权重向量 $\mathbf{a}_{i,j} \in \mathbb{R}^{1 \times P}$, $\mathbf{a}_{i,j}$ 表示每个

单词对该从句语义表达的重要程度. 如式(1)–(3)所示:

$$\mathbf{h}_{i,j}^o, \mathbf{h}_{i,j}^l = \text{BiLSTM}_{i,j}^{\text{TL}}(\vec{Z}_j, \overleftarrow{Z}_j), \quad (1)$$

$$\mathbf{a}_{i,t} = \frac{\exp(g(\mathbf{h}_{i,t}^o \circ \mathbf{h}_{i,t}^l))}{\sum_{j=1}^P \exp(g(\mathbf{h}_{i,j}^o \circ \mathbf{h}_{i,j}^l))}, t \in [1, P], \quad (2)$$

$$\mathbf{o}_{i,j}^{\text{TL}} = (\mathbf{a}_{i,j} \cdot \mathbf{h}_{i,j}^o) \circ \mathbf{h}_{i,j}^l, \quad (3)$$

式(1)中“ \rightarrow ”, “ \leftarrow ”分别为BiLSTM的两个计算方向, 式(2)中函数 g 为两层感知机(multi-layer perceptron, MLP), $g(x) = \text{MLP}(x) = \sigma_2(\sigma_1(x))$, σ_1 和 σ_2 分别为线性转换函数, “ \circ ”表示沿最后一个维度合并两个矩阵, 式(3)中“ \cdot ”为矩阵乘法, $\mathbf{o}_{i,j}^{\text{TL}} \in \mathbb{R}^{1 \times 4H}$ 为 $c_{i,j}$ 的词编码器输出. 特别地, 在计算式(2)时, 先将 $\mathbf{h}_{i,j}^l$ 的维度扩展为 $P \times 2H$, 当应用函数 g 后, 最后一维的变化依次为: $4H, 2H, 1$, 表示将单词特征压成一个实数, 然后使用softmax分配注意力权重, 得到 $c_{i,j}$ 中单个单词的注意力权重 $\mathbf{a}_{i,t} \in \mathbb{R}^1$, 合并单词注意力权重得到 $\mathbf{a}_{i,j}$.

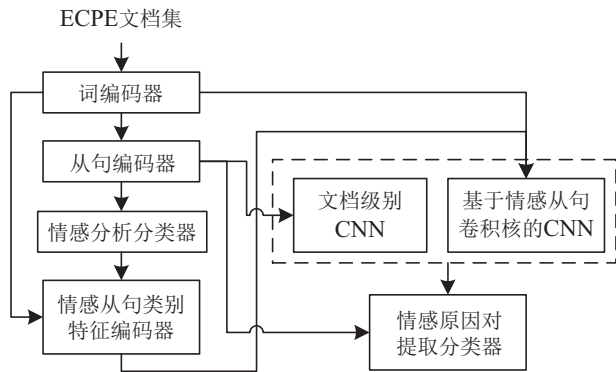


图 1 总体框架图

Fig. 1 Overall framework

3.3 从句编码器

该模块由一个BiLSTM构成. 记 d_i 的BiLSTM输出为 $\tilde{\mathbf{h}}_i^o \in \mathbb{R}^{N \times 2H}$, $\tilde{\mathbf{h}}_i^l \in \mathbb{R}^{2 \times 1 \times H}$, 其中 $\tilde{\mathbf{h}}_i^o$ 包含 d_i 中的每个从句的隐层状态, 表示从句之间的语义特征, $\tilde{\mathbf{h}}_i^l$ 表示BiLSTM最后一个状态的输出. 如式(4)–(5)所示:

$$\tilde{\mathbf{h}}_i^o, \tilde{\mathbf{h}}_i^l = \text{BiLSTM}_i^{\text{CL}}(\vec{\sigma}_i^{\text{TL}}, \overleftarrow{\sigma}_i^{\text{TL}}), \quad (4)$$

$$\mathbf{o}_i^{\text{CL}} = g(\tilde{\mathbf{h}}_i^o). \quad (5)$$

词编码器以从句为单位输入到模型中, 得到的 $\mathbf{o}_{i,j}^{\text{TL}}$ 为该从句的语义特征, 而从句编码器则以文档为单位计算, 因此, 式(4)中, 沿第1个维度合并同一个文档所有从句的词编码器输出 $\mathbf{o}_{i,j}^{\text{TL}} \in \mathbb{R}^{1 \times 4H}$, 得到该文档的从句编码器输入 $\mathbf{o}_i^{\text{TL}} \in \mathbb{R}^{N \times 4H}$. 式(5)中, $\mathbf{o}_i^{\text{CL}} \in \mathbb{R}^{N \times H}$ 为 d_i 的从句编码器输出, 表示该文档的语义特征.

3.4 情感从句类别特征编码器

将 \mathbf{o}_i^{CL} 通过线性层, 从而得到情感分析预测矩阵

$\mathbf{P}_i^e \in \mathbb{R}^{N \times \text{eclass}}$, 其中eclass为从句的情感类别数. 采用注意力机制得到含有情感类别特征的注意力权重矩阵 $\mathbf{A}_{i,j}^e \in \mathbb{R}^{N \times P}$, 取出其中的情感从句注意力向量 \mathbf{A}_i^e , 相关过程如式(6)–(7)所示:

$$\mathbf{P}_i^e = \tanh(g(\mathbf{o}_i^{\text{CL}})), \quad (6)$$

$$\mathbf{A}_{i,t}^e = \frac{\exp(\text{argmax}(\mathbf{P}_{i,t}^e) * \mathbf{a}_{i,t})}{\sum_{j=1}^N \exp(\text{argmax}(\mathbf{P}_{i,j}^e) * \mathbf{a}_{i,j})}, \quad (7)$$

$t \in [1, N]$,

式(7)中“ $*$ ”为矩阵对应元素相乘, “argmax”表示取出最大特征的索引.

3.5 基于情感从句卷积核的CNN

原因事件具有顺承关系, 因此寻找原因的时候不能脱离其他从句的事件语境. 图2(a)和图2(b)原因从句的事件均为“回忆”, 但情感却不一样, 如果只看情感从句和原因从句, 那么这个因果关系是缺少事件成分的, 导致模型无法理解为什么原因相同却得出不同的情感. 因此理解情感因果关系的关键是理解背后发生的事件.

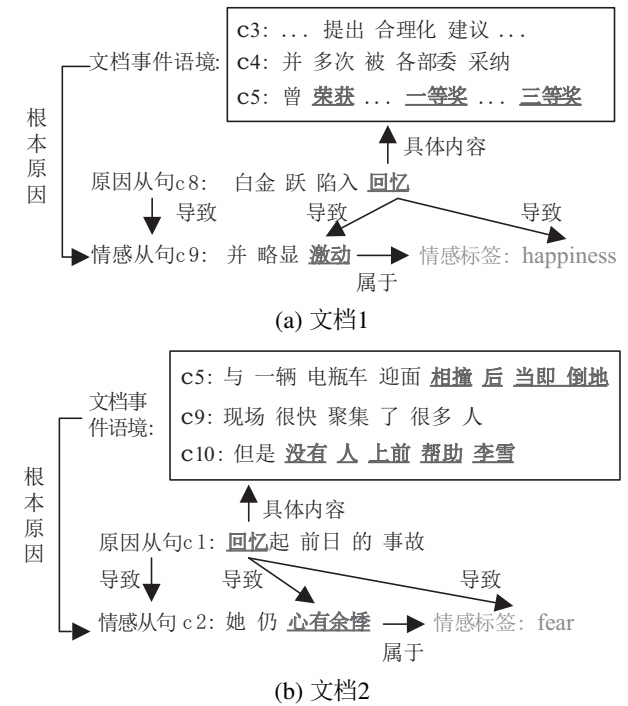


图 2 事件语境与EC-Pair的关系图

Fig. 2 The relationship between event context and EC-Pair

情感从句卷积核的CNN(emotion kernel CNN, EKCNN)的目的在于提取情感对应的原因, 因此卷积核需要具备提取情感因果关系的能力, 而情感原因是与情感特征关联的, 因此将 \mathbf{A}_i^e 作为EK. 在一次卷积中, EK依次扫描一篇文档的所有从句, 模型根据EK提供的情感信息有目的地重新整理文档的事件关系, 从而学习使情感因果关系成立的事件语境. ECPE数

据集中情感对应的原因只有一个,所以采用最大池化提取最高的特征,作为原因事件特征.文献[6,23]采用的将其余从句逐一与情感从句比较从而配对原因从句的方法没有考虑因果事件语境,文献[29]采用的将从句重排序的方法则打乱了从句原有的事件顺承关系,这会干扰模型对因果关系的理解.

在ECPE数据集中,一个文档通常存在多个事件,但不是每个事件都与情感原因有关,而且情感从句和原因从句是有距离的,如果不过滤文档中无关的事件,就难以提取有效的原因信息,于是采用跳跃连接和事件阈值来改善原因特征的提取过程.

设 d_i 中情感从句个数为 n^e ($1 \leq n^e \leq 3$),则模型会得到 n^e 个情感从句的注意力向量,即 $\mathbf{A}_i^e \in \mathbb{R}^{n^e \times P}$.

当 $n^e = 1$ 时, $\mathbf{A}_i^e \in \mathbb{R}^{1 \times P}$,将文档 d_i 中所有从句的 $\mathbf{h}_{i,j}^o \in \mathbb{R}^{P \times 2H}$ 进行合并,得到 $\mathbf{h}_i^o \in \mathbb{R}^{N \times P \times 2H}$,再通过线性层,得到卷积输入 $\mathbf{h}_i^{\text{in}} \in \mathbb{R}^{1 \times NP \times 2H}$,卷积通道为1.接着将 \mathbf{A}_i^e 通过线性层,并进行维度扩展,得到 N 个(与从句数相同)通道为1的情感从句卷积核 $\mathbf{h}_i^{\text{kernel}} \in \mathbb{R}^{P \times 2H}$.相关过程如式(8)所示:

$$\mathbf{h}_i^{\text{map}} = \text{FC}_2(\text{FC}_1(\text{Maxpool2D}(\text{ReLU}(f^{P \times 2H}(\mathbf{h}_i^{\text{in}}, \mathbf{h}_i^{\text{kernel}}))))), \quad (8)$$

式(8)中 $f^{P \times 2H} \in \mathbb{R}^{N \times ((NP+N)-P+1) \times 1}$ 为卷积操作, $\text{Maxpool2D} \in \mathbb{R}^{N \times 1 \times 1}$ 是沿最后两个维度的最大池化操作,池化窗口大小为 $((NP+N)-P+1) \times 1$,表示保留一个最重要的特征, FC_1 和 FC_2 分别表示两个全连接层, $\mathbf{h}_i^{\text{map}} \in \mathbb{R}^{N \times 1 \times 2H}$ 为 d_i 的卷积输出.

当 $1 < n^e \leq 3$ 时, $\mathbf{A}_i^e \in \mathbb{R}^{2 \times P}$ 或 $\mathbf{A}_i^e \in \mathbb{R}^{3 \times P}$,将每个情感从句单独作为卷积核进行卷积操作,此时得到 $\mathbf{h}_{i,n^e}^{\text{map}} \in \mathbb{R}^{N \times 1 \times 2H}$,将 $\mathbf{h}_{i,n^e}^{\text{map}}$ 沿最后一个维度进行拼接,得到 $\mathbf{h}_i^{\text{map}} \in \mathbb{R}^{N \times 1 \times 2n^e H}$,再通过最大池化层和全连接层,得到 d_i 的卷积输出 $\mathbf{h}_i^{\text{map}} \in \mathbb{R}^{N \times 1 \times 2H}$.

在批训练的应用下,将所有文档的 $\mathbf{h}_i^{\text{map}}$ 沿批次方向进行拼接,再进行维度扩展,得到EKCNN的输出 $\mathbf{h}^{\text{CNN1}} \in \mathbb{R}^{B \times N \times N \times 2H}$, B 表示batchsize.

为了能适应不同的事件距离,本文方法将时序卷积网络^[41](temporal convolutional network, TCN)应用到EKCNN中,从而实现跳跃连接,以提取多尺度原因事件信息. TCN由因果卷积和空洞卷积构成,设空洞卷积的跳跃因子为 $\varepsilon \in 1, 2, 4, \dots, 2^{L-1}$, L 为TCN的层数(L 的选择见第4.4小节),当 $\varepsilon = 1$ 时退化为普通卷积.式(8)使用TCN后,卷积操作如图3所示.

在每层TCN后,过滤低于事件阈值 τ 的特征,表示这些特征与情感原因事件无关,为了适应不同的事件语境, τ 取 \mathbf{A}_i^e 的均值,如式(9)–(10)所示:

$$\tau = \text{Mean}(\mathbf{A}_i^e), \quad (9)$$

$$\mathbf{h}_i^{\text{map}} = I[|\mathbf{h}_{i,j}^{\text{map}}| > \tau], \quad (10)$$

式(9)中Mean为均值计算,式(10)中 I 为指示函数,表

示只保留大于 τ 的特征.

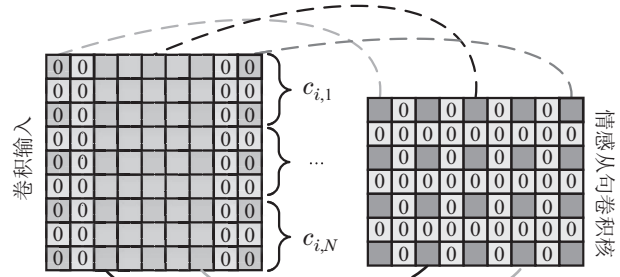


图3 使用TCN后的EKCNN结构图

Fig. 3 EKCNN structure after using TCN

3.6 文档级别CNN

EKCNN主要捕捉情感因果关系的事件语境特征,而文档级别CNN(document CNN, DCNN)则用来对从句进行特征强化,得到文档的融合语义特征,从而进一步帮助模型理解因果关系.

将 $\mathbf{o}_i^{\text{CL}} \in \mathbb{R}^{N \times H}$ 进行维度扩展,得到DCNN的卷积输入 $\tilde{\mathbf{h}}_i^{\text{in}} \in \mathbb{R}^{1 \times N \times 2H}$,卷积通道为1,如式(11)所示:

$$\tilde{\mathbf{h}}_i^{\text{map}} = \text{FC}(\text{Maxpool2D}(\text{ReLU}(f^{3 \times 2H}(\tilde{\mathbf{h}}_i^{\text{in}}, \tilde{\mathbf{h}}_i^{\text{kernel}}))))), \quad (11)$$

式中: $f^{3 \times 2H} \in \mathbb{R}^{1 \times (N+2 \times 3-3+1) \times 1}$, $\tilde{\mathbf{h}}_i^{\text{kernel}} \in \mathbb{R}^{3 \times 2H}$,池化操作的窗口大小为 $(N+2 \times 3-3+1) \times 1$, $\text{Maxpool2D} \in \mathbb{R}^{1 \times 1 \times 1}$,进行维度扩展后再接入全连接层FC,得到 d_i 的输出 $\tilde{\mathbf{h}}_i^{\text{map}} \in \mathbb{R}^{N \times N \times 2H}$.在批训练的应用下,DCNN的输出为 $\mathbf{h}^{\text{CNN2}} \in \mathbb{R}^{B \times N \times N \times 2H}$.

3.7 训练目标函数

将 \mathbf{h}^{CNN1} , \mathbf{h}^{CNN2} 沿最后一个维度进行合并,得到 $\mathbf{o}^{\text{CNN}} \in \mathbb{R}^{B \times N \times N \times 4H}$.然后将 \mathbf{o}^{CNN} 与该批次所有从句的 \mathbf{o}_i^{CL} 进行残差连接,以缓解深度网络训练时秩崩溃的问题,得到 $\mathbf{o}^{\text{net}} \in \mathbb{R}^{B \times N \times N \times 6H}$,如式(12)所示,接着通过MLP层的函数 g ,得到ECPE的预测矩阵 $P^{\text{ec}} \in \mathbb{R}^{B \times N \times N}$.

$$\mathbf{o}^{\text{net}} = \sum_{i=1}^B (\alpha_1(\mathbf{o}_i^{\text{CL}}) \circ \alpha_2(\mathbf{o}_i^{\text{CL}})) \circ \mathbf{o}^{\text{CNN}}, \quad (12)$$

式中 α_1 和 α_2 分别为线性转换函数.情感分析为多分类任务,ECPE为二分类任务,均采用交叉熵损失函数,模型总体损失函数为两个任务交叉熵损失函数之和.

4 实验

4.1 ECPE数据集

实验采用两个数据集,文献[4]的中文数据集来源于新浪城市新闻,共有2105篇文档;文献[7]的英文数据集来源于NTCIR-13ECE,共有2843篇文档.两个数据集均标有EC-Pair标签、从句的情感标签以及情感关键词.

4.2 实验设置

使用数据集提供的word2vec进行词编码.词向量维度为200,相对位置维度为50,中文从句数为75,中

文单词数为40, 英文从句数为40, 英文单词数为30, H 为128, B 为16, 初始学习率为0.001, dropout为0.3, L2正则化为 $1e-4$, 优化器为Adam, epoch为300. 本文采用文献[4]的数据集划分方法进行十折交叉验证, 取测试集上的平均结果作为每个评价指标的最终结果.

本文对比的基线模型为: Indep, Inter-CE, Inter-EC^[4], LAE^[6], E2E-PextE, E2E-PextC^[7].

4.3 模型评价指标

模型评价指标采用文献[4]定义在ECPE上的精确度(precision)、召回率(recall)、F1分数. 模型的最终结果主要对比F1分数. 记情感分析和ECPE同时预测正确的比例为EE.

4.4 TCN层数 L 的实验结果

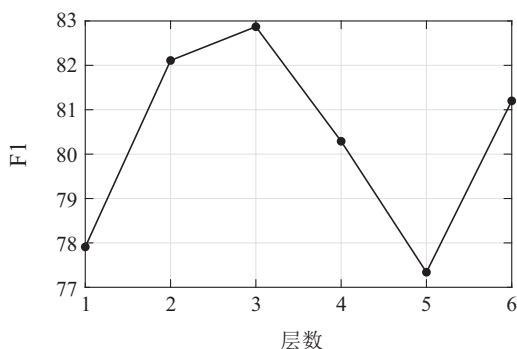
TCN比普通卷积拥有更广阔的感受野, 可以连接距离更远的事件的特征, 但是TCN中的空洞卷积由于零填充的原因可能会遗失部分特征, 所以需要多层的空洞卷积来弥补这一缺陷.

TCN层数的实验结果如图4所示. 虽然当 $L = 5$ 时, ECPE的F1分数最高, 但是此时情感分析的F1分数和EE的F1分数最低, 所以 L 不能取5. 当 $L = 4$ 时, 尽管ECPE的F1分数下降了0.17%, 但此时其他两项都处于较高值, 所以后续实验均取 $L = 4$.

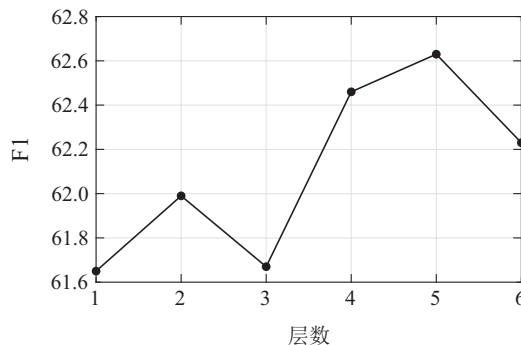
4.5 实验结果

在中文数据集上, 实验结果如表1所示, 本文方法ECPE的F1分数有1.38%~6.08%的提升, 说明情感分析有助于提高模型对情感因果关系的理解. 情感类别信息使得情感从句不再是黑盒子, 从而使模型提取出有效的原因信息. LAE模型在提取EC-Pair时没有利用情感类别信息, 情感从句依然是黑盒子, 并且该模型忽略了因果关系成立的事件语境, 而本文将情感从句作为卷积核, 这种动态的卷积核能帮助模型学到丰富的因果模式, 并且跳跃连接和事件阈值有效地去除不相关的语境信息, 从而提高ECPE结果.

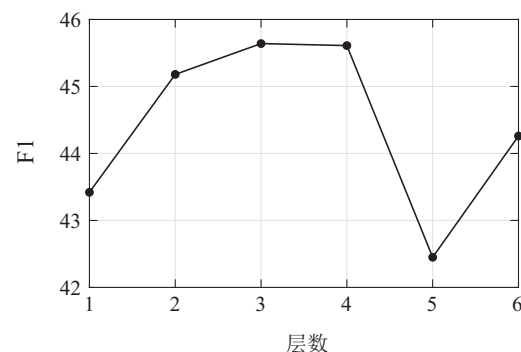
在英文数据集上, 如表2所示, 由于英文语法体系与中文有很大的差异, 因此整体的ECPE F1分数偏低. 实验结果显示, 本文方法有2.35%~7.27%的提升, 说明事件语境的研究有利于跨语言的因果提取, 模型可以提取不同语言的原因事件.



(a) Emotion F1分数



(b) ECPE F1分数



(c) EE F1分数

图4 不同TCN层数 L 的实验结果

Fig. 4 The result of different TCN layer L

4.6 ECE和ECPE任务对比

文献[4, 7]在ECPE任务中都引入了ECE任务(包括情感提取和原因提取两个子任务), 作为模型的辅助任务. 尽管文献[7]把ECPE和ECE放在同一个反向传播网络中, 但是情感提取和原因提取这两个过程是无法进行信息交流的, 需要在EC-Pair筛选阶段才能将这两个子任务结合, 而本文的情感分析不仅能帮助模型理解具体情感的差异, 还能提高原因提取的质量, 因此实验结果均高于文献[4]和文献[7](见表1-2).

4.7 因果事件语境的有效性

为了进一步验证本文模型对EC-Pair的提取效果, 将表1中ECPE F1分数最高的基线模型E2E-PextC的两个辅助任务去除, 同时将本文ECPE-EKC模型的情感分析去除, 从而对比在没有其他任务的情况下, EC-Pair的提取效果. 实验结果如表3所示, 本文方法仍有0.79%的提高, 说明在提取EC-Pair时考虑从句之间的事件语境关系, 而不仅仅是情感从句和原因从句之间的关系, 可以提高EC-Pair的提取质量.

4.8 基于BERT模型的对比实验

分别向 $c_{i,j}$ 的首尾两端添加[CLS]和[SEP]^[42]. 编号为单数和偶数的从句分别标记为0和1^[43]. 为了检测模型捕捉长距离因果关系的能力, 分别在情感从句和原因从句之间添加若干个由PAD标识符(PAD为补零标记, 添加后不会引入额外的因果关系)构成的从句, 记为ECPE-PAD.

表1 中文数据集十折交叉验证平均运行结果

Table 1 Average results of ten-fold cross validation for Chinese datasets

模型	Emotion/%			情感提取/%			原因提取/%			ECPE/%			EE/%
	P	R	F1	P	R	F1	P	R	F1	P	R	F1	
Indep	—	—	—	84.01	78.76	81.20	68.92	59.37	63.67	59.96	54.13	56.75	—
Inter-CE	—	—	—	82.34	79.81	81.02	68.82	58.85	63.35	58.73	54.54	56.42	—
Inter-EC	—	—	—	85.63	79.22	82.25	70.23	58.18	63.54	64.09	54.56	58.84	—
LAE	72.64	70.07	70.84	—	—	—	—	—	—	63.54	50.86	56.38	39.22
E2E-PextE	—	—	—	83.35	77.37	80.11	72.13	51.98	60.13	60.46	61.80	61.01	—
E2E-PextC	—	—	—	83.93	77.89	80.68	70.92	52.76	60.21	61.58	61.00	61.08	—
本文方法	82.51	79.28	80.29	—	—	—	—	—	—	66.14	59.17	62.46	45.61

* P表示精确度(precision), R表示召回率(recall), F1是基于P和R的调和平均, EE表示情感分析和ECPE同时预测正确的比例, 下同.

* Indep, Inter-CE, Inter-EC没有进行文献[4]的EC-Pair过滤步骤, 下同.

* 情感提取和原因提取为ECE的两个子任务.

* “—”表示该模型没有完成情感分析, 下同.

表2 英文数据集十折交叉验证平均运行结果

Table 2 Average results of ten-fold cross validation for English datasets

模型	Emotion/%			情感提取/%			原因提取/%			ECPE/%			EE/%
	P	R	F1	P	R	F1	P	R	F1	P	R	F1	
Indep	—	—	—	70.20	67.07	68.55	62.59	51.81	56.60	49.78	40.55	44.49	—
Inter-CE	—	—	—	69.87	68.14	68.95	61.07	52.48	56.37	48.88	41.90	45.04	—
Inter-EC	—	—	—	70.69	68.36	69.44	61.43	52.80	56.58	49.73	42.09	45.38	—
LAE	56.88	52.58	54.14	—	—	—	—	—	—	54.26	37.08	43.76	23.07
E2E-PextE	—	—	—	71.52	68.33	69.84	67.29	42.00	51.42	53.62	45.76	49.20	—
E2E-PextC	—	—	—	71.17	65.89	68.26	65.33	43.49	51.90	51.61	47.60	49.41	—
本文方法	59.71	54.60	57.04	—	—	—	—	—	—	60.25	45.36	51.76	27.85

表3 仅ECPE平均运行结果(中文数据集)

Table 3 Average results of ECPE only (Chinese datasets)

模型	ECPE/%		
	P	R	F1
E2E-PextC (去除ECE)	60.22	61.33	60.55
本文方法(去除情感分析)	65.11	57.99	61.34

受GPU硬件限制, 实验仅在中文数据集上进行, B 为4, 单词数为30, 去掉从句数大于45的文档, 并冻结BERT预训练模型的参数. BERT词向量维度为768, 模型的学习率为 $2e-5$, L2正则化为 $1e-5$, epoch为20, 采用bert-base-Chinese模型, 其余参数与第4.2小节相同. 对比的模型为ECPE-2D (Inter-EC+WC+BERT)^[30], 其中WC为该文献提出的窗口约束二维Transformer.

如表4所示, ECPE-2D (Inter-EC+WC+BERT)和本文方法(+BERT)在原始数据集的F1分数上效果接近, 但是当使用ECPE-PAD时, ECPE-2D效果大幅下降. WC假设原因从句和情感从句的位置是邻近的, 通

过计算窗口内从句对的Transformer特征来过滤EC-Pair, 而窗口外的从句对则忽略不计, 因此当EC-Pair距离较远的时候, 该方法难以捕捉长距离因果关系. 而本文方法采用TCN连接长距离事件, 再通过事件阈值来筛选EC-Pair, 从而捕捉长距离因果关系.

表4 基于BERT的平均运行结果(中文数据集)

Table 4 Average results based on BERT (Chinese datasets)

模型	ECPE/%		
	P	R	F1
ECPE-2D (Inter-EC+WC+BERT)	67.33	63.40	65.13
ECPE-2D (Inter-EC+WC+BERT, ECPE-PAD)	66.74	24.72	35.77
本文方法(+BERT)	69.51	62.38	65.68
本文方法(+BERT, ECPE-PAD)	69.05	53.95	60.57

* 第1行和第3行使用原始中文数据集, 第2行和第4行使用改造后的中文数据集ECPE-PAD.

表 5 正确样本
Table 5 Correct sample

从句	EC-Pair & Emotion 标签	Inter-EC	LAE	本文方法
c210,26 每次听见她这么说 c210,27我心里 难受	(27,26)&sadness	(27,26)&-	(22,25)(27,25)(27,26) &sadness	(27,26)&sadness
c965,4 孩子被扎虽然 愤怒	(4,4)&anger	(4,4)&-	null&>null	(4,4)&anger
c1349,6 都委婉拒绝 c1349,7父母十分 悲伤	(7,6)&sadness;	(7,6)&-;	null&sadness	(7,6)&sadness;
c1349,8 看到身体残疾的儿子整日待在家里 c1349,9不免深深地为他的未来 忧虑	(9,8)&sadness	null&-	null&>null	(9,8)&sadness

* “null”表示识别不出情感或EC-Pair, 下同。
* 从句编号加粗的为原因从句, 加下划线的为情感从句, 下同。

表 6 错误样本
Table 6 Error sample

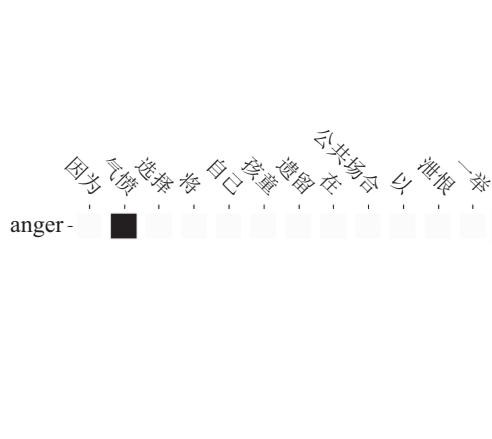
从句	EC-Pair & Emotion 标签	本文方法
c1089,4 但妻子每次都矢口否认 c1089,5这让李志辉十分 恼火 c1089,8两人经常因此事发生争吵 c1089,9为解心中 憋屈	(5,4)&anger	(9,8)&disgust
c1266,2 张鹏 兴奋 地告诉妈妈自己已经通过了面试 c1266,4今年25岁的张鹏腼腆地说道我特别 喜欢 开车	(2,2)&happiness	(4,4)&happiness (c2,c4同时得到happiness)
c1905,7 学得太差 不好意思	(7,7)&sadness	null&>null

4.9 案例分析

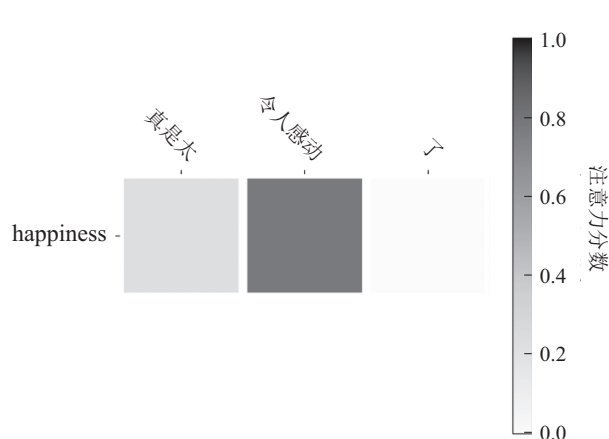
4.9.1 自注意力热力图

图5展示了自注意力对情感从句中情感特征的关注程度. 对于图5(a)和图5(b), 自注意力都能准确地集

中地落在情感关键词上, 说明自注意力能有效地提炼出情感从句的情感特征. 图5(c)和图5(d)展示了两种错误类型, 前者注意力分散, 没有充分地落在情感关键词上, 后者注意力尽管有关注情感关键词, 但总体分布比较均匀, 没有突出重点.



(a) 正确样本(情感标签: anger)



(b) 正确样本(情感标签: happiness)

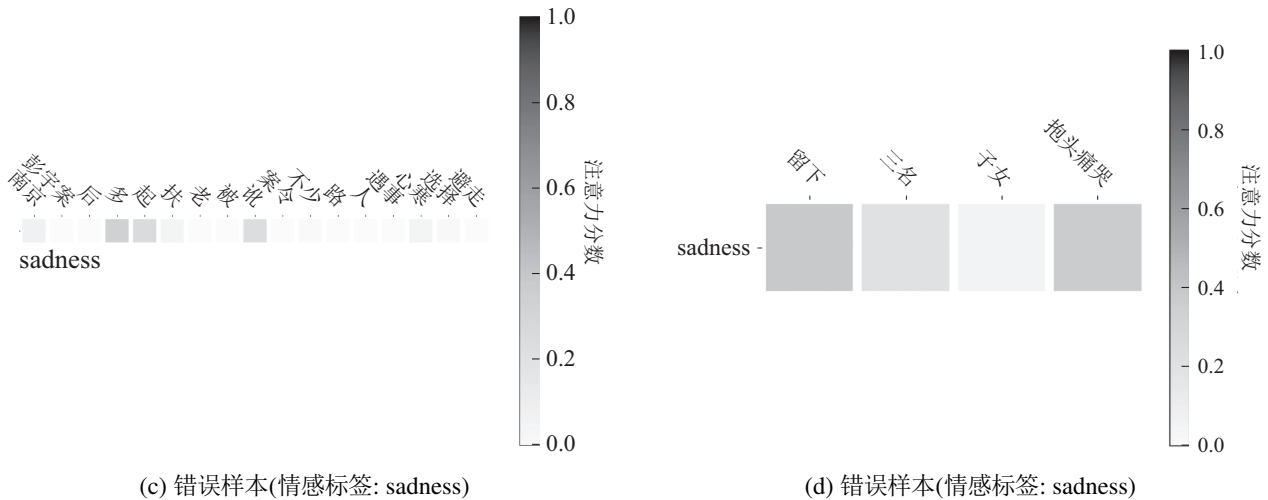


图5 情感从句自注意力热力图

Fig. 5 Self-attention heatmap of emotion clause

4.9.2 样本分析

以中文数据集为例,表5列举了本文方法识别正确的样本.在第210号文档中,LAE错误提取出了标签以外的两对EC-Pair;在第965号文档中,LAE识别不出任何的标签;在第1349号文档中,Inter-EC和LAE都无法完整地找出EC-Pair.

以中文数据集为例,表6列举了3种错误类型.在第1089号文档中,本文方法识别出另一种没有标注的情感“disgust”,并认为 $c_{1089,8}$ 是该情感的原因从句,经过分析, $c_{1089,8}$ 和 $c_{1089,9}$ 情感因果关系确实成立,说明ECPE数据集中数据标注有遗漏. $c_{1905,7}$ 的情感为“sadness”,但“不好意思”和“伤心”是完全不同的情感,说明ECPE数据集中数据标注存在歧义.在第1266文档中,只有 $c_{1266,2}$ 进行了情感标注,而本文却同时测出了 $c_{1266,2}$ 和 $c_{1266,4}$ 的“happiness”情感.

5 总结

本文提出了基于注意力机制的情感从句类别特征编码器和基于情感从句卷积核的CNN的分层模型ECPE-EKC,将情感类别特征融入ECPE,使情感从句不再是黑盒子,从而使模型更准确地学到情感从句的情感信息;将因果事件语境融入原因提取过程,使EC-Pair的提取不再是脱离从句上下文的独立匹配过程,从而使模型更准确地理解情感因果关系.

参考文献:

- [1] LEE S, CHEN Y, HUANG C. A text-driven rule-based system for emotion cause detection. *Human Language Technologies: Conference of the North American Chapter of the Association for Computational Linguistics*. Los Angeles: ACL, 2010: 45 – 53.
- [2] GUI L, WU D, XU R, et al. Event-driven emotion cause extraction with corpus construction. *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*. Austin: ACL, 2016: 1639 – 1649.
- [3] GUI L, XU R, LU Q, et al. Emotion cause extraction, a challenging task with corpus construction. *Chinese National Conference on Social Media Processing*. Nanchang, Jiangxi: CCIS, 2016: 98 – 109.
- [4] XIA R, DING Z. Emotion-cause pair extraction: a new task to emotion analysis in texts. *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. Florence: ACL, 2019: 1003 – 1012.
- [5] LECUN Y, BOTTOU L, BENGIO Y, et al. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 1998, 86(11): 2278 – 2324.
- [6] TANG H, JI D, ZHOU Q. Joint multi-level attentional model for emotion detection and emotion-cause pair extraction. *Neurocomputing*, 2020, 409(7): 329 – 340.
- [7] SINGH A, HINGANE S, WANI S, et al. An end-to-end network for emotion-cause pair extraction. *Proceedings of the 11th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*. Online: ACL, 2021: 84 – 91.
- [8] VAN RIJSBERGEN, C.J. *Information Retrieval. 2nd Edition*. London: Butterworth, 1979: 1 – 208.
- [9] LI W, XU H. Text-based emotion classification using emotion cause extraction. *Expert Systems with Applications*, 2014, 41(4): 1742 – 1749.
- [10] GAO K, XU H, WANG J. Emotion cause detection for chinese micro-blogs based on ECOCC model. *Pacific-Asia Conference on Knowledge Discovery and Data Mining*. Ho Chi Minh City: Springer, 2015, 9078: 3 – 14.
- [11] GAO K, XU H, WANG J. A rule-based approach to emotion cause detection for Chinese micro-blogs. *Expert Systems with Applications*, 2015, 42(9): 4517 – 4528.
- [12] MEI Y, LEE S. *A Linguistic Approach to Emotion Detection and Classification*. HongKong: The Hong Kong Polytechnic University, 2010.
- [13] CHEN Y, LEE S, LI S, et al. Emotion cause detection with linguistic constructions. *Proceedings of the 23rd International Conference on Computational Linguistics*. Beijing: Coling 2010 Organizing Committee, 2010: 179 – 187.
- [14] LEE S, CHEN Y, HUANG C, et al. Detecting emotion causes with a linguistic rule-based approach. *Computational Intelligence*, 2013, 29(3): 390 – 416.
- [15] GUI L, YUAN L, XU R, et al. Emotion cause detection with linguistic construction in chinese weibo text. *CCF International Con-*

- ference on Natural Language Processing and Chinese Computing*. Berlin: Springer, 2014, 496: 457 – 464.
- [16] RUSSO I, CASELLI T, RUBINO F, et al. EMO Cause: an easy-adaptable approach to extract emotion cause contexts. *Proceedings of the 2nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis*. Portland: ACL, 2011: 153 – 160.
- [17] NEVIAROUSKAYA A, AONO M. Extracting causes of emotions from text. *Proceedings of the 6th International Joint Conference on Natural Language Processing*. Nagoya: AFNLP, 2013: 932 – 936.
- [18] SONG S, MENG Y. Detecting concept-level emotion cause in microblogging. *WWW'15 Companion: Proceedings of the 24th International Conference on World Wide Web*. New York: ACM, 2015: 119 – 120.
- [19] GHAZI D, INKPEN D, SZPAKOWICZ S. Detecting emotion stimuli in emotion-bearing sentences. *International Conference on Intelligent Text Processing and Computational Linguistics*. Cairo: Springer, 2015, 9042: 152 – 165.
- [20] YU X, RONG W, ZHANG Z, et al. Multiple level hierarchical network based clause selection for emotion cause extraction. *IEEE Access*, 2019, 7: 9071 – 9079.
- [21] XIA R, ZHANG M, DING Z. RTHN: A rnn-transformer hierarchical network for emotion cause extraction. *Proceedings of the 28th International Joint Conference on Artificial Intelligence*. Macao: IJCAI, 2019: 5285 – 5291.
- [22] XIAO X, WEI P, MAO W, et al. Context-aware multi-view attention networks for emotion cause extraction. *IEEE International Conference on Intelligence and Security Informatics*. Shenzhen: IEEE, 2019: 128 – 133.
- [23] LI X, FENG S, WANG D, et al. Context-aware emotion cause analysis with multi-attention-based neural network. *Knowledge-based Systems*, 2019, 174: 205 – 218.
- [24] LIANG L, JI X, REN F. Attention-based bi-lstm-crf network for emotion cause extraction in texts. *IEEE International Conference on Mechatronics and Automation*. Beijing: IEEE, 2020: 1670 – 1675.
- [25] LI X, SONG K, FENG S, et al. A co-attention neural network model for emotion cause analysis with emotional context awareness. *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*. Brussels: ACL, 2018: 4752 – 4757.
- [26] DIAO Y, LIN H, YANG L, et al. Multi-granularity bidirectional attention stream machine comprehension method for emotion cause extraction. *Neural Computing and Applications*, 2019, 32(12): 8401 – 8413.
- [27] HU J, SHI S, HUANG H. Combining external sentiment knowledge for emotion cause detection. *CCF International Conference on Natural Language Processing and Chinese Computing*. Dunhuang: Springer, 2019, 11838: 711 – 722.
- [28] FAN C, YAN H, DU J, et al. A knowledge regularized hierarchical approach for emotion cause analysis. *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing*. Hong Kong: ACL, 2019: 5614 – 5624.
- [29] DING Z, HE H, ZHANG M, et al. From independent prediction to reordered prediction: Integrating relative position and global label information to emotion cause identification. *Proceedings of the 33rd Association for the Advancement of Artificial Intelligence Conference on Artificial Intelligence*. Hawaii: AAAI, 2019, 33(1): 6343 – 6350.
- [30] DING Z, XIA R, YU J. ECPE-2D: emotion-cause pair extraction based on joint two-dimensional representation, interaction and prediction. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. Online: ACL, 2020: 3161 – 3170.
- [31] WEI P, ZHAO J, MAO W. Effective inter-clause modeling for end-to-end emotioncause pair extraction. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. Online: ACL, 2020: 3171 – 3181.
- [32] DIAO Y, LIN H, YANG L, et al. Emotion cause detection with enhanced-representation attention convolutional-context network. *Soft Computing*, 2021, 25(2): 1297 – 1307.
- [33] DING Z, XIA R, YU J. End-to-end emotion-cause pair extraction based on sliding window multi-label learning. *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing*. Online: EMNLP, 2020: 3574 – 3583.
- [34] JIA X, CHEN X, WAN Q, et al. A novel interactive recurrent attention network for emotion-cause pair extraction. *The 3rd International Conference on Algorithms, Computing and Artificial Intelligence*. Sanya: ACM, 2020: 1 – 9.
- [35] WU S, CHEN F, HUANG Y, et al. A multi-task learning neural network for emotioncause pair extraction. *The 24th European Conference on Artificial Intelligence*. Santiago: IOS Press, 2020, 325: 2212 – 2219.
- [36] CHEN X, LI Q, WANG J. A unified sequence labeling model for emotion cause pair extraction. *Proceedings of the 28th International Conference on Computational Linguistics*. Barcelona: ACL, 2020: 208 – 218.
- [37] CHEN X, LI Q, WANG J. Conditional causal relationships between emotions and causes in texts. *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing*. Online: ACL, 2020: 3111 – 3121.
- [38] LI P, MAO K. Knowledge-oriented convolutional neural network for causal relation extraction from natural language texts. *Expert Systems with Applications*, 2019, 115: 512 – 523.
- [39] SCHUSTER M, PALIWAL ALIWAL K. Bidirectional recurrent neural networks. *IEEE Transactions on Signal Processing*, 1997, 45(11): 2673 – 2681.
- [40] EKMAN P. Expression and the nature of emotion. *Approaches to Emotion*, 1984: 319 – 344.
- [41] BAI S, KOLTER J Z, KOLTUN V. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. *arXiv Preprint*. ArXiv: 1803.01271, 2018.
- [42] BERT: pre-training of deep bidirectional transformers for language understanding. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Minneapolis: ACL, 2019, 1: 4171 – 4186.
- [43] LIU Y, LAPATA M. Text summarization with pretrained encoders. *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing*. Hong Kong: ACL, 2019: 3730 – 3740.

作者简介:

叶苑莉 硕士研究生, 目前研究方向为自然语言处理, E-mail: 872265931@qq.com;

张 灵 教授, 目前研究方向为人工智能及机器学习算法, E-mail: 1252875930@qq.com;

陈云华 副教授, 目前研究方向为深度学习、神经形态类脑计算、计算机视觉, E-mail: yhchen@gdut.edu.cn.