

基于位置交互感知网络的多任务情绪原因对抽取方法

付明睿, 李卫疆

引用本文

付明睿, 李卫疆. 基于位置交互感知网络的多任务情绪原因对抽取方法[J]. 计算机科学, 2024, 51(11A): 231000086-9.

FU Mingrui, LI Weijiang. [Multi-task Emotion-Cause Pair Extraction Method Based on Position-aware Interaction Network](#) [J]. Computer Science, 2024, 51(11A): 231000086-9.

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)

[基于多任务学习的复杂城市遥感图像道路提取](#)

Road Extraction from Complex Urban Remote Sensing Images Based on Multi-task Learning
计算机科学, 2024, 51(11A): 240300095-8. <https://doi.org/10.11896/jsjcx.240300095>

[基于多视角的图像文本情感分析](#)

Sentiment Analysis of Image-Text Based on Multiple Perspectives

计算机科学, 2024, 51(11A): 231200163-8. <https://doi.org/10.11896/jsjcx.231200163>

[基于BERT模型和图注意力网络的方面级情感分析](#)

Aspect-based Sentiment Analysis Based on BERT Model and Graph Attention Network

计算机科学, 2024, 51(11A): 240400018-7. <https://doi.org/10.11896/jsjcx.240400018>

[对话场景下的情感引导问题生成模型](#)

Emotion Elicited Question Generation Model in Dialogue Scenarios

计算机科学, 2024, 51(11): 265-272. <https://doi.org/10.11896/jsjcx.231000002>

[基于AU的多任务学生情绪识别方法研究](#)

Study on Multi-task Student Emotion Recognition Methods Based on Facial Action Units

计算机科学, 2024, 51(10): 105-111. <https://doi.org/10.11896/jsjcx.240300059>

基于位置交互感知网络的多任务情绪原因对抽取方法

付明睿 李卫疆

昆明理工大学信息工程与自动化学院 昆明 650500

昆明理工大学云南省人工智能重点实验室 昆明 650500

(a2813176983@126.com)

摘要 情绪原因对抽取任务旨在同时抽取情感子句和原因子句。已有的方法把情绪原因对抽取看作情绪抽取、原因抽取和情绪原因对抽取3个独立的任务,不能有效捕捉到任务之间的联系。此外,现有的两阶段模型存在误差传播问题,并且情感子句和原因子句间相对位置分布不平衡。文中提出了一个新的基于BERT、情感词典和位置感知交互模块的情绪原因对抽取模型MK-BERT。该模型首先用情感词典增强的BERT进行文本编码;其次,为了解决标签位置不平衡问题,根据情感子句和原因子句间的相对距离设计位置感知交互模块,以捕捉位置信息并构建情绪原因对的特征;最后,通过情绪预测模块和原因预测模块间交互编码,充分挖掘多个任务间的共享信息。在中文情绪原因对抽取数据集上进行实验,结果表明,所提模型可以有效地抽取情绪原因对,并且在位置不平衡样本上取得良好性能。

关键词:情感分析;情绪原因对抽取;多任务学习;情感词典;位置感知

中图分类号 TP391

Multi-task Emotion-Cause Pair Extraction Method Based on Position-aware Interaction Network

FU Mingrui and LI Weijiang

Faculty of Information Engineering and Automation, Kunming University of Science and Technology, Kunming 650500, China

Yunnan Key Laboratory of Artificial Intelligence, Kunming University of Science and Technology, Kunming 650500, China

Abstract The task of emotion-cause pair extraction is to extract emotion clauses and reason clauses simultaneously. Previous methods regard emotion-cause pair extraction as three independent tasks of emotion extraction, cause extraction, and emotion-cause pair extraction, which cannot effectively capture the connection between tasks. In addition, the existing two-stage models suffer from error propagation problems, and the relative position distribution between emotion clauses and reason clauses is unbalanced. This paper proposes a new emotional reason pair extraction model MK-BERT based on BERT, sentiment lexicon and position-aware interaction module. The model first uses the BERT enhanced by the sentiment lexicon for document encoding. In order to solve the problem of label position imbalance, a position-aware interaction module is designed according to the relative distance between the emotion clause and the reason clause to capture the position information and construct the characteristics of the emotion-cause pair. Then, through interactive encoding between the emotion prediction module and the reason prediction module, the shared information among multiple tasks is fully mined. Experimental results on the Chinese emotion-reason pair extraction dataset show that the proposed model can effectively extract emotion-reason pairs and achieve good performance on positionally imbalanced samples.

Keywords Sentiment analysis, Emotion-Cause pair extraction, Multi-task learning, Sentiment lexicon, Position aware

1 引言

情绪原因分析(ECA)作为情感分析(SA)的一项基本任务,旨在分析文本中隐含的情绪原因,在舆情分析、在线评论等方面具有重大作用。但是由于人工标记的昂贵成本,使得这项任务难以应用到现实中。情绪原因分析包括情绪抽取(EE)、情绪原因抽取(ECE)和情绪原因对抽取(ECPE)3个子任务。作为情绪原因分析的一项新任务,情绪原因对抽取既需要对文档全面理解,又需要对具体方面进行细致分析。如何在没有情感注释的情况下挖掘文本潜在的情绪和原因,

成为了ECPE任务的关键。

Lee等^[1]提出情绪原因抽取任务要求对情绪进行人工注释,然后抽取出相应的原因,这不仅制约着它在现实任务场景下的使用,也忽略了情感与原因之间的内在联系。Xia等^[2]提出了情感原因对抽取这一新的任务,以解决情感原因抽取任务所产生的问题。以往对ECPE任务研究的方法大致可以划分为两阶段模型和端到端模型。Xia等^[2]所提出的两阶段模型给出了子句抽取模型和对抽取模型,首先通过该模型分别抽取情感子句与原因子句,然后该模型通过笛卡尔积获取情感子句与原因子句的全部情绪原因配对,再利用逻辑回归

基金项目:国家自然科学基金(62066022)

This work was supported by the National Natural Science Foundation of China(62066022).

通信作者:李卫疆(hrbrichard@126.com)

模型筛选出候选子句对,抽取适当的情感原因配对。在端到端模型上,该方法将情感子句和原因子句并行编码到特征空间,然后区分笛卡尔积生成的所有候选子句对是否是有效情感原因子句对。端到端模型由于能缓解两阶段模型来自第一步的误差而导致第二步性能下降的误差传递问题,且能将所有子任务的先验知识统一编码,因此比两阶段模型效果更好。例如,表1列出了一个情绪原因对抽取实例。

表1 情绪原因对抽取实例

Table 1 Extraction instance of emotion and cause pair

C1:	在今年2月成立了这家摄影工作室
C2:	随着顾客的订单越来越多
C3:	虽然经常熬夜修片
C4:	但王高兴却乐在其中
C5:	可这层快乐始终蒙着一层担忧
C6:	父母还不知道她辞职的事

Candidate Pairs
{(c4,c2),(c5,c6)}

表1中有6个子句。c4子句包含“快乐”情绪,它的原因子句为c2;c5子句包含“担忧”情绪,它对应的原因子句为c6。情绪原因对抽取的最终结果为:{(c4,c2),(c5,c6)}。

在ECPE任务中,关键的难点在于如何处理情绪与原因的内在联系。先前的模型通过给情感子句和原因子句间的距离设置一个阈值来筛选出一定距离大小的情绪原因对。如图1所示,虽然有将近95.8%的情绪原因对的情感子句和原因子句距离小于或等于2,但是仅考虑相对位置信息而忽略语义依赖性会加大情绪和原因的相对距离分布不平衡所造成的问题。一方面,尽管已有模型指出情感子句与原因子句之间存在关联,但是未充分利用两者间的关联信息,亦未考虑子任务情绪抽取和情绪原因抽取是否有助于主任务情绪原因对抽取。另一方面,在语言表达中,有效情感原因对中情感与原因子句在情感上具有一致性。情感与原因之间还存在着错综复杂的多对多映射,情感的原因可能不止一个,相同的原因也可能受到多重情感的刺激。

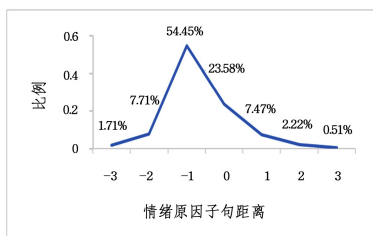


图1 情绪原因相对位置

Fig. 1 Relative distance of emotion and cause

情绪抽取、情绪原因抽取和情绪原因对抽取这3项任务的交互是复杂的。本文在关注情绪和原因关系的同时,发现感情也是有其特殊的指向的,并且影响其原因。因此在抽取子句时,既要把原因和情感联系起来,又要加以区别。在现有的大部分多任务学习的框架中,先前的共享机制未必能够有效利用任务之间丰富的隐含信息。

如图2所示,为了解决以上问题,充分利用3个任务间的交互性以及情绪和原因抽取的一致性,本文提出了一种新的

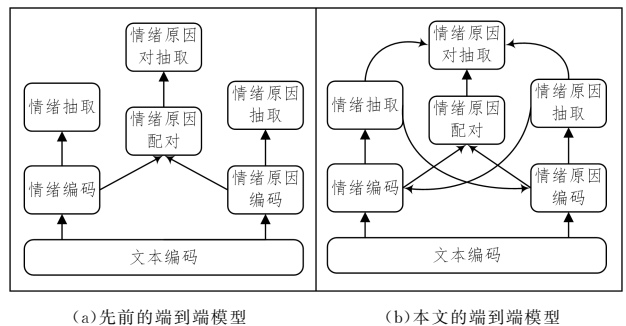
多任务情绪原因对抽取模型。其分为4个模块,使用各种简单有效的机制来捕获交互性,提出包含位置感知交互编码模块、情绪预测模块、原因预测模块和情绪原因对预测模块的多任务端到端模型。具体而言,首先将情感词典知识引入位置感知交互编码模块,以获得外部知识增强的句子编码,再通过两个PAIM模块捕捉相对位置,分别对情绪和原因编码;其次,考虑到情绪和原因子句间的一致性有助于寻找上下文因果关系,情绪预测模块和原因预测模块对情绪和原因分别进行预测,进行任务间的交互;最后,通过交叉注意力^[3]机制将情绪过滤器与原因过滤器所得到的子句进行信息交换,进一步提炼精细表示。本文在ECPE的基准数据集上评估方法的有效性。实验结果表明,该方法优于现有方法。总之,本文的贡献如下:

1)提出了一种新的包含情绪抽取、情绪原因抽取和情绪原因对抽取3个任务相交互的网络,用于捕获情感子句和原因子句的信息,从而更好地利用3个任务相互间的关系。

2)引入情感词典来增强子句的表示,使句子在编码时包含丰富的情感信息和上下文语义信息。

3)引入位置感知交互模块(PAIM)来捕捉相对位置,来缓解样本相对位置的不平衡问题,并结合在情绪抽取和情绪原因抽取模块中得到的预测,学习以数据相对位置为主的情感原因对表示,从而更好地捕捉子句对之间的关联。

4)实验结果表明,本文方法在标准的ECPE数据集上达到最先进的性能,并在情绪和原因的相对距离分布不平衡样本上更加稳健。



(a) 先前的端到端模型

(b) 本文的端到端模型

图2 端到端模型

Fig. 2 End-to-End model

2 相关工作

2.1 情绪原因抽取

情绪原因对抽取有基于统计学习方法、机器学习方法和深度学习方法。早期,Gui等^[4]利用SVM分类器构建数据集并构造开放数据集。Li等^[5]提出,具体情感的语境有助于寻找对应理由的有价值的线索,共同注意力模块的设计是为了利用执行ECE任务时情感的语境。Yu等^[6]注意到子句间的关系对任务的积极作用,从而提出分层框架。该框架考虑到了情感描述与从句间语义信息间的相互作用和从句间的关联。除文档内容外,Ding等^[7]发现情感描述及子句间标签信息及相对位置信息对情感原因的抽取具有重要意义。Xia等^[8]用Transformer作子句编码器建模子句和进一步抽取出的情感原因间的关系。这些方法都是基于词袋模型的。Hu等^[9]利用图卷积网络对子句语义与结构信息进行编码,并且在情感原因的抽取上获得卓越表现。

2.2 情绪原因对抽取

情绪原因对抽取任务旨在抽取触发情绪表达的潜在原因,其主要包括两阶段模型和端到端模型。

两阶段模型在第一阶段把原因抽取和情绪抽取看作两个独立的任务,分别完成两个子任务,在第二阶段通过笛卡尔积将情感子句和原因子句配对,然后通过过滤器筛选出具有关系的情感原因对。先前的工作^[2]首次提出了情绪原因对抽取任务,并提出了两阶段模型。具体来说,该方法首先分别抽取情感子句和原因子句,然后训练分类器过滤错误的候选对。由于它存在错误传播,因此这种两阶段模型的系统对于 ECPE 来说不是最优的。此外,子句抽取模型和对抽取模型是独立的,具有不同的特征空间和输出标签空间,导致不能充分利用相关任务的互利性。

采用端到端模型来解决 ECPE 的任务已经有很多工作。Wei 等^[10]强调子句间关系的重要性,对观点进行排名,并且选取最高排名的几对作为情感原因配对,提出了一种新的端到端方法,在 2D 表示方案中表示情绪原因,并利用 2D-Transformer 对不同情绪原因对的相互作用进行建模。另外, Ding 等^[11-12]认为,只有当情绪被识别后才能更好地去理解原因的产生。为了解决这一问题,他们提出了两种不同的方法分别基于子句进行分析。Yuan 等^[13]设计了一个新型标记方案,提出了基于 BiLSTM 序列标记模型,用于抽取情感原因对。Tang 等^[14]认为,目前的研究尚未发现情绪检测与 ECPE 之间的关系,为此,其提出了一种面向 ED 与 ECPE 工作的多任务学习框架。Wu 等^[15]提出了多任务学习神经网络,共同进行情绪预测、原因预测以及情绪原因关系预测任务等,探讨了这几项任务间的相互关系。Song 等^[16]将 ECPE 视为链接预测问题,设计了一个端到端多任务学习模型,用于预测情绪原因对。Yu 等^[17]提出了一种相互支持的多任务模型,在模型中加入两项子任务,建构情绪抽取与原因抽取交互作用。Chen 等^[18]采用多任务学习的方法共同提取情绪、原因和情绪-原因对,清晰地建模了不同任务之间的相互作用。该框架采用了循环同步机制,使不同任务间的相关性能被恰当使用,从而迭代地提高它们的预测。Chen 等^[19]构造了一个具有 3 种基于距离的依赖关系的情绪原因候选对的图,并使用 GCN 来建模局部邻域候选对之间的依赖关系。Wu 等^[20]提出了一个新标记框架 PTF,用于端到端地求解 ECPE 任务。PTF 把子句进行关系分类并将情感与原因相融合,使整个 ECPE 成功整合成全局优化的统一标注任务。Huang 等^[21]提出了一种基于跨度关联预测的深度神经网络模型,用于情绪原因对提取。其使用跨度表示方法,从语法习惯的角度出发结合跨度的概念来处理这项任务,同时利用跨度关联匹配方法强化句子深层含义,并在文档中匹配句子以精确得到候选情绪原因对。

综上所述,先前的端到端方法取得了良好的效果,但没能很好地缓解样本位置不平衡问题。本文设计了位置感知模块 (PAIM),用于捕捉情绪和原因之间的位置相关性,以解决相对位置问题。与以往隐式捕获部分相互性的方法不同,本文模型显式地捕捉了上述情感和原因子句之间的所有交互信息。

3 MK-BERT 模型

给定一个包含 n 个子句的文档 $D = \{c_1, c_2, \dots, c_n\}$ 。子句

表示为 $c_i = \{\omega_1, \omega_2, \dots, \omega_m\}$,其中 m 为子句的长度。通过模型抽取出情感子句 c_e 和原因子句 c_c 构成的情感原因对 $\langle c_e, c_c \rangle$ 。然后,对具有因果关系的情感原因对预测其概率分布。最后,若 c_e 和 c_c 构成具有因果关系的情感原因对,则赋予标签“1”,否则赋予标签“0”。模型整体框架如图 3 所示。

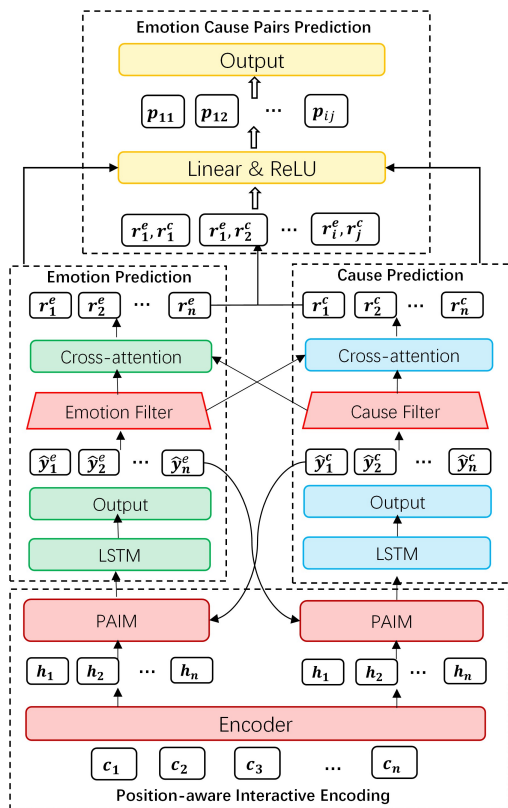


图3 MK-BERT 模型

Fig. 3 MK-BERT model

3.1 词典增强的相对位置感知交互编码模块

本文将整个文档输入到经过预训练的 BERT^[22] 中,将 jieba 分词工具分词得到的词汇在情感词典中检索,若某个词汇在情感词典中,便将该词加到相应子句的末尾。为了表示单个子句,本文在每个子句的开头插入 [CLS] 标记,每个子句结尾插入 [SEP] 符号。本文还使用区间分隔符嵌入来区分文档中的多个子句。对于子句 c_i ,根据 i 是奇数还是偶数来分别嵌入 E_A 或 E_B 标记。例如,对于文本 $[c_1, c_2, c_3, c_4, c_5]$,本文分别嵌入 $[E_A, E_B, E_A, E_B, E_A]$ 。最终子句 c_i 以 $c_i = ([CLS], \omega_1, \omega_2, \dots, \omega_m, [SEP])$ 的形式输入到 BERT 中。输出的 h_i 用于表示子句 c_i ,整个文档的表示记为 $H = \{h_1, h_2, \dots, h_n\}$ 。

3.2 位置感知交互模块

如图 4 所示,为了捕捉情感与原因之间的位置相互性来解决上下文建模阶段的相对位置问题,本文设计了感知交互模块 PAIM,该模块包含前馈网络、自注意力和层归一化的 Transformer 编码器架构。此外,本文将自注意力与可训练的相对位置项结合起来作为其核心组件,对相对位置进行建模。

文档中的子句并不独立存在,一般来说,掌握上下文线索可以帮助本文更好地理解当前子句。因此,本文应用 Transformer 的编码器模块,通过将其他子句的信息合并到当前子句中生成新的子句表示,这使本文能够从文档的角度理解当前子句。Transformer 的标准编码器包括一个由 N 个相同层组成的堆栈,其中每个层有两个子层。第一子层是多头

自注意力层,第二子层是全连接的前馈网络。

融合相对位置的多头自注意力层。本文采用多头注意力,对于每个句子 c_i ,首先将其句子表示 $h_i \in R^{d_h}$ 送到3个不同的完全连接层中,以生成 query, key, value, 表示为 q_i, k_i 和 v_i ,具体表达式如下:

$$q_i = W_Q h_i \quad (1)$$

$$k_i = W_K h_i \quad (2)$$

$$v_i = W_V h_i \quad (3)$$

$$s_{ij} = \frac{q_i k_j}{\sqrt{d}} \quad (4)$$

其中, $W_Q \in R^{d_q \times d_h}, W_K \in R^{d_k \times d_h}, W_V \in R^{d_v \times d_h}$ 为可训练参数, d_k, d_q, d_v 分别为 key, query, value 向量的维度。注意力分数 s_{ij} 与标准自注意力机制一样。

本文采用一个可训练偏差项对相对位置进行建模,然后将其与标准的自注意力得分进行汇总,以得到最终的自注意力表示。

通过 PAD(Position-aware Distance)公式确定输入的情绪与原因之间的距离, $|p-q|$ 表示子句 i 和 j 中情绪词与原因词的绝对值距离, m 表示情绪词或原因词的长度。

$$PAD_{p,q} = |p-q| - \left\lfloor \frac{m}{2} \right\rfloor \quad (5)$$

$$e_{ij} = s_{ij} + f(PAD_{p,q}) \quad (6)$$

$$f(p-q) = \begin{cases} g(PAD_{p,q}), & \text{if } |i-j| \leq W \\ b, & \text{otherwise} \end{cases} \quad (7)$$

其中, $f(p-q)$ 是限制窗口 W 内相对距离的偏差函数,进一步

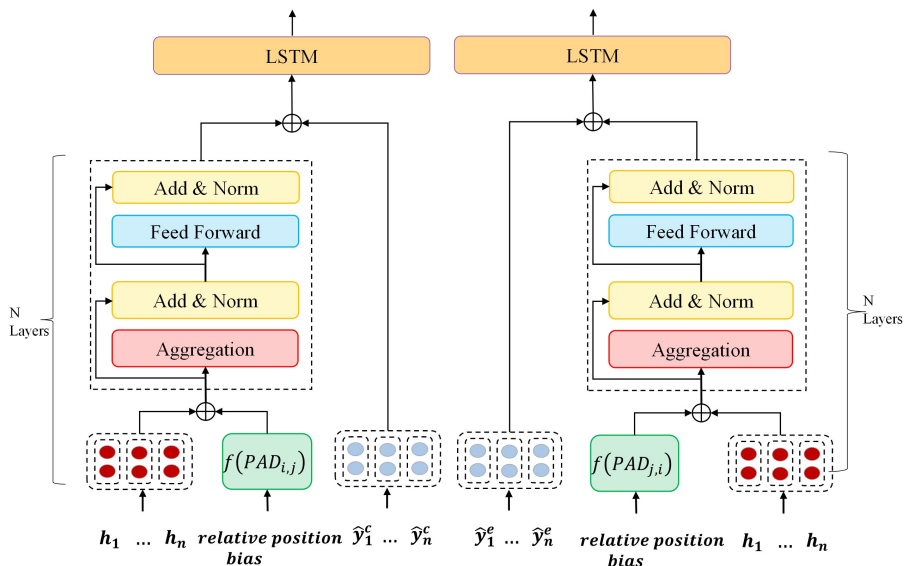


图4 位置感知交互模块(PAIM)

Fig. 4 Position-aware interaction module

3.3 情绪子句和原因子句预测模块

由于情绪与相应原因之间具有明显的交互性,因此本文引入预测模块来交换所选择的子句信息。两个编码器将自适应地交换子句信息,以完善编码器对它们的具体表示。在交换和细化之后,它们更容易消除噪声特征以获得更细化的特定表示。为此,本文计算子句 c_i 属于情感或原因的概率 \hat{y}_i^e 和 \hat{y}_i^c ,计算式如下:

$$\hat{y}_i^e = \sigma(W^e r_i^e + b^e) \quad (14)$$

$$\hat{y}_i^c = \sigma(W^c r_i^c + b^c) \quad (15)$$

将其映射到表示对应的位置偏置。

在获得调节注意分数 e_{ij} 后,经过 softmax 操作,再将值向量进行聚合,然后输入到编码块中:

$$z_i = \sum_j \text{Softmax}(e_{ij}) v_j \quad (8)$$

注意力子层之后是全连接的前馈网络层:

$$e_i = W_2 (\text{ReLU}(W_1 z_i + b_1)) + b_2 \quad (9)$$

其中, $W_1 \in R^{d_f \times d_v}, W_2 \in R^{d_v \times d_f}, b_1 \in R^{d_f}$ 和 $b_2 \in R^{d_f}$ 是可训练参数。

为了帮助 Transformer 训练,本文在每个子层的输出处添加一个残差连接,然后进行层标准化。

$$r_i = \text{LayerNorm}(x_i + \text{sublayer}(x_i)) \quad (10)$$

其中, x_i 是 layer 的输入,而 $\text{sublayer}(x_i)$ 是子层的输出。

最后,Transformer 的编码器输出一组句子嵌入,表示为 $rR = [r_1, r_2, \dots, r_i]$ 。

直观地说,不同的编码器可以捕获不同类型的特征。因此,本文使用两个子句级位置感知模块 PAIM^e 和 PAIM^c 分别获得每个子句 c_i 的情绪表示 R^e 和原因表示 R^c 。

$$R^e = \text{PAIM}^e(H) \quad (11)$$

$$R^c = \text{PAIM}^c(H) \quad (12)$$

情绪子句编码的输入包括以下两个部分:来自上下文编码模块的子句表示 $\{r_i\}_{i=1}^n$,循环得到的来自原因预测任务的原因标签嵌入 $\{\hat{y}_i^c\}_{i=1}^n$,将两个部分拼接起来,输入到 PAIM 模块来编码特定情绪的上下文信息。原因子句编码类似。

$$r_i^e = [r_i; \hat{y}_i^c] \quad (13)$$

其中, W^e, W^c, b^e, b^c 为可学习参数, σ 为 sigmoid 函数。

为了得到更精细的子句间交互信息,首先将子句按照 \hat{y}_i^e 和 \hat{y}_i^c 进行排序,分别得到情感和原因对应的两个排名,并分别选取两个排名中 top_N 的子句进行交换。

$$r_i^e = top_N(r^e, \hat{y}_i^e) \quad (16)$$

$$r_i^c = top_N(r^c, \hat{y}_i^c) \quad (17)$$

然后采用共享的 cross-attention 模块,将 cross-attention 机制、前馈网络和层归一化封装起来,得到表示:

$$r^e = \text{Cross-attention}(r^e, r_i^c) \quad (18)$$

$$r_j^c = \text{Cross-attention}(r_i^e, r_i^c) \quad (19)$$

3.4 情绪原因对预测模块

如图 5 所示,对于候选对的选择,本文设置了超参数 W

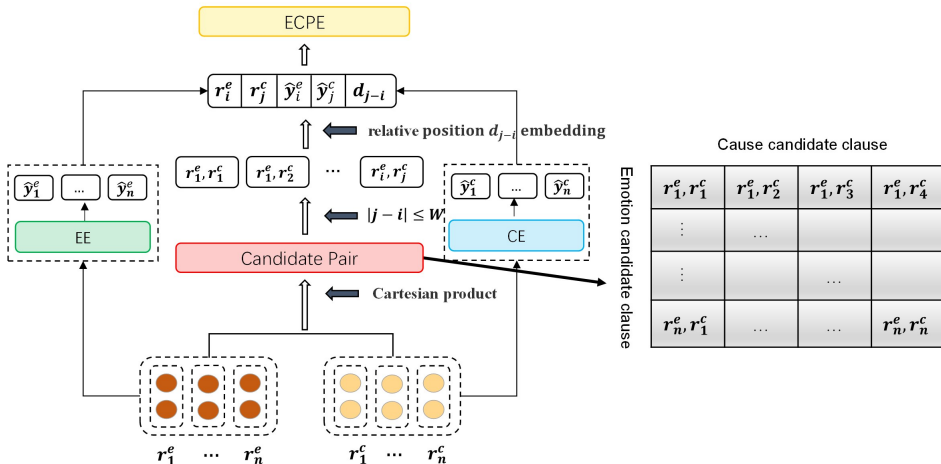


图 5 情绪原因对预测模块

Fig. 5 ECPE module

考虑到情绪抽取和原因抽取任务的一致性,对于一个候选子句对 $(c_i^e, c_j^c) \in P$, 本文将情绪和原因的表示连接, 以获得对表示 z_{ij} 。

$$z_{ij} = \{r_i^e, r_j^c; \hat{y}_i^e; \hat{y}_j^c; d_{j-i}\} \quad (20)$$

$$r_{ij} = \text{ReLU}(\mathbf{W}_p^T z_{ij} + \mathbf{b}_p) \quad (21)$$

其中, d_{j-i} 是子句 c_i^e 和 c_j^c 的相对位置嵌入, \mathbf{W}_p 和 \mathbf{b}_p 是可训练的参数。

然后,将对表示 r_{ij} 馈送到 sigmoid 分类器中, 以获得候选对表示 (c_i^e, c_j^c) 是情绪原因对的概率 \hat{y}_{ij} :

$$\hat{y}_{ij} = \sigma(\mathbf{W}_p r_{ij} + \mathbf{b}_p) \quad (22)$$

本文分别计算情绪和原因的二进制交叉熵损失, 情绪原因对抽取任务的损失函数使用预测的分布 \hat{y}_{ij} 和标签 y_{ij} 之间的二进制交叉熵损失得到:

$$\mathcal{L}_{\text{emo}} = - \sum_k (y_k^e \log(\hat{y}_k^e) + (1 - y_k^e) \log(1 - \hat{y}_k^e)) \quad (23)$$

$$\mathcal{L}_{\text{cau}} = - \sum_k (y_k^c \log(\hat{y}_k^c) + (1 - y_k^c) \log(1 - \hat{y}_k^c)) \quad (24)$$

$$\mathcal{L}_{\text{pair}} = - \sum_{(c_i^e, c_j^c) \in \mathcal{P}} (y_{ij} \log(\hat{y}_{ij}) + (1 - y_{ij}) \log(1 - \hat{y}_{ij})) \quad (25)$$

对于给定的训练数据集, 本文联合训练 3 个任务。训练目标是 minimized 损失函数, 计算式如下:

$$\mathcal{L} = \mathcal{L}_{\text{pair}} + \lambda_e \mathcal{L}_{\text{emo}} + \lambda_c \mathcal{L}_{\text{cau}} + \lambda \|\theta\|^2 \quad (26)$$

其中, λ_e 和 λ_c 是权重系数, λ 是 L_2 正则化损失 $\|\theta\|^2$ 的系数。

4 实验与分析

4.1 实验数据和评价指标

本文在 ECPE 的基准数据集上评估模型。该数据集由 Xia^[2] 基于中文情绪原因语料库构建。其中文档数量为 1945, 具体数据集的统计数据如表 2 所列。本文选择了 90% 的数据进行训练, 10% 的数据用于测试。为了获得可信的结果, 本文将每个实验重复 10 次, 并选取平均值作为结果。选取 P, R, F1 值是评估指标, 其中 F1 平衡预测结果的准确性和召回率, 是主要指标。

来控制生成候选对的窗口 ($|j - i| \leq W$)。如果第 c_i 条和第 c_j 条子句之间的绝对距离小于或等于窗口 W , 本文将作为 ECPE 的训练样本 \mathcal{P} 。

表 2 中文数据集数量统计

Table 2 Number of Chinese datasets statistics

Total number of documents	1945
The number of documents containing 1 pair	1746
The number of documents containing 2 pairs	177
The number of documents containing 3 or more pairs	22
Average number of clauses per document	14.77
Average number of pairs per document	1.11
Maximum number of clauses per document	73

4.2 实验设置

为了评估模型的有效性, 本文使用 200 维单词嵌入来初始化单词向量, 在新浪新闻构成的 ECPE 数据集^[2] 上进行实验, 并选取 HowNet 和 ANTUSD 作为情感词典。此外, 本文将相对位置嵌入设置为 50, 超参数 W 设为 4, dropout 率设为 0.1 并应用于单词嵌入层。本文模型是基于 AdamW 优化器进行训练的, 其中批大小设置为 4, 学习率设置为 2×10^{-5} , 模型的目标函数中, L_2 正则化系数 λ 设为 $\times 10^{-5}$, 损失权重 λ_e 和 λ_c 都设为 1。本文在工作中采用 BERT-Chinese 作为子句编码器, 子句表示的维数被设置为 200。在 Transformer 中, 编码器模块的层数设置为 2, 隐藏状态的维度设置为 200, 相对距离设为 4。在实验中, 本文将数据分为 10 部分, 并采用 10 倍交叉验证。本文使用精度 P , 召回率 R 和 $F1$ 得分作为评估指标。

$$P = \frac{\sum \text{correct_pairs}}{\sum \text{predict_pairs}} \quad (27)$$

$$R = \frac{\sum \text{correct_pairs}}{\sum \text{annotated_pairs}} \quad (28)$$

$$F1 = \frac{2 * P * R}{P + R} \quad (29)$$

4.3 对比方法

现有的模型分为两阶段和端到端的方法。下文将本文的方法与以下几种方法进行比较。

1) 两阶段方法

Indep: 含有 Inter-CE 和 Inter-EC 两个变体, 采用两阶段框架。首先用 BiLSTM 对子句进行编码, 再分别用 BiLSTM 对情绪和原因子句进行建模来抽取情感子句和原因子句,

然后进行情绪原因配对和过滤。

Inter-CE:利用原因抽取的预测来改进情绪抽取。

Inter-EC:利用情绪抽取的预测来改进原因抽取。

MAM-SD:是一种相互支持的多任务模型,可以模拟情绪抽取和原因抽取之间的交互作用。

2)端到端方法

MTNECP:是一个统一的多任务学习框架。它在任务间共享有用的特征,并利用位置感知的情感信息进行原因抽取。

TDGC^[23]:是一个基于转换的模型。它通过构建带有标记边的有向图来识别情绪及相应的原因。

ECPE-2D:包含 Inter-EC+WC 和 Inter-EC+CR。基于多任务学习,ECPE-2D 使用多层变换器来进一步捕获子句之间的连接。

Inter-EC+WC:基于 Inter-EC 的端到端模型。使用窗口约束的 2D-transformer 来模拟不同情绪原因对之间的相互作用。

Inter-EC+CR:基于 Inter-EC 的另一个端到端模型 ECPE-2D。二路交互 2D-transformer 用于模拟不同情绪原因对之间的相互作用。

E2EECPE:是 ECPE 任务的典型多任务学习方法。它将情感原因对抽取作为一个链接预测任务,并使用辅助情感抽取和原因抽取来促进对抽取。

RankCP:是一个基于排名的模型。它利用子句之间的相关性和基于核的相对位置嵌入进行有效的排序。

ECPE-MLL:基于滑动窗口多标签学习,并集成以情绪和原因为枢轴的两个任务抽取,以在预测阶段获得最终结果。

UTOS^[24]:通过设计统一的标记方案,将 ECPE 任务重新定义为统一的序列标记问题,并提出了一个统一的面向目标的序列到序列模型来解决统一的序列标签问题。

MTST-ECPE^[25]:提出了一种多任务序列标记框架,该框架可以通过将情绪的距离编码到一个新的标记方案中,同时抽取具有相关原因的情绪。

4.4 评估和比较

不同方法在中文数据集上的实验结果如表 3 所列,

可以看出:

1)总体上看,端到端模型要优于两阶段模型。端到端模型能充分考虑到多任务间的交互性,从而减少误差传播。相比其他模型,本文 MK-BERT 模型在 ECPE 上的 F1 分数取得最好表现。此外,P 和 R 评分优于大部分模型,证明了其有效性。本文模型在主任务情绪原因对抽取任务上均优于其他方法,在两项子任务中得分也名列前茅。此外,RankCP 和 ECPE-2D 在 EE 上与 ECPE 上的性能较差。同时,在原因抽取上表现良好的方法一般在情绪原因对抽取上表现较好,各方法对原因抽取的性能显著低于情绪抽取。本文猜测原因在于原因抽取对情绪原因对抽取性能瓶颈的制约,因此改善原因抽取比改善情绪抽取对提高情绪原因对抽取性能更为有利。

2)在多个任务中,MK-BERT 均取得了不错的成绩。本文方法在 F1 得分上相比现有 EE,CE 和 ECPE 任务上的最佳模型,分别增加了 0.99%,1.91%和 2.39%。其原因在于 MK-BERT 内部 3 种任务相互作用后,EE 牺牲掉部分的性能帮助 ECPE 与 CE 获得更加明显的改进。结合情感词汇增强子句的表达之后,在子任务上的表现均得到了明显提升。结果表明,借助情感词典的词汇扩充子句表示输入预训练 BERT 能有效加强子句级语境特征表示能力。与以往最先进的方法相比,由于使用情感词汇,MK-BERT 对 ECPE 任务中的 F1 得分有明显提升。由于 MK-BERT 能有效利用 3 个子任务间的相互性,当总体任务性能得到改善时,MK-BERT 能够取得更加优秀的表现。

3)在 F1 性能方面,基于 BERT 的方法 RANKCP (BERT)和 ECPE-MLL(BERT)是目前较好的两个模型。本文模型在 R 值上超过了它们很多,因为本文在执行 ECPE 任务之前通过加入相对位置来筛选出主要的情感原因对,这既有利于模型专注于寻找具有情绪原因关系的对,又有利于缓解样本位置不平衡的问题。此外,由于端到端模型位置感知交互模块输出的子句表示,每个任务生成特定于任务的表示,本文模型在各个子任务上也有很好的表现。

表 3 不同方法在中文数据集上的实验结果

Table 3 Experimental results of different methods on Chinese dataset

Models	Sub Tasks						Main Task			
	Emotion Extraction			Cause Extraction			Pairs Extraction			
	P	R	F1	P	R	F1	P	R	F1	
Two-step	Indep	83.75	80.71	80.71	69.02	56.73	62.05	68.32	50.82	58.18
	Inter-CE	84.94	81.22	81.22	68.09	56.34	61.51	69.02	51.35	59.01
	Inter-EC	83.64	81.07	82.30	70.41	60.83	65.07	67.21	57.05	61.28
	MAM-SD	85.54	81.41	83.39	72.02	63.75	67.51	69.63	57.99	63.20
End-to-end	MTNECP	86.62	83.93	85.20	74.00	63.78	68.44	68.28	58.94	63.21
	TDGC	80.80	84.39	82.56	67.42	65.34	66.36	65.15	63.54	64.34
	Inter-EC+WC	85.11	82.37	83.65	71.33	62.85	66.72	71.18	59.84	64.94
	Inter-EC+CR	85.12	82.20	83.58	72.72	62.98	67.38	69.60	61.18	64.96
	E2EECPE	85.95	79.15	82.38	70.62	60.30	65.03	64.78	61.05	62.80
	RankCP	87.03	84.06	85.48	69.27	67.43	68.24	66.98	65.46	66.10
	TDGC(BERT)	87.16	82.44	84.74	75.62	64.71	69.74	73.74	63.07	67.99
	Inter-EC+WC(BERT)	86.27	92.21	89.10	73.36	69.34	71.23	72.92	65.44	68.89
	Inter-EC+CR(BERT)	85.48	92.44	88.78	72.72	69.27	70.87	69.35	67.85	68.37
	RankCP(BERT)	91.23	89.99	90.57	74.61	77.88	76.15	71.19	76.30	73.60
ECPE-MLL(BERT)	86.08	91.91	88.86	73.82	79.12	76.30	77.00	72.35	74.52	
UTOS(BERT)	88.15	83.21	85.56	76.71	73.20	74.71	73.89	70.62	72.03	
MTST-ECPE(BERT)	85.83	80.94	83.21	77.64	72.36	74.77	75.78	70.51	72.91	
MK-BERT	92.91	90.25	91.56	79.06	77.36	78.20	77.48	76.35	76.91	

4.5 消融实验

为了评估各个组件对模型性能的影响,本文依次删除各个组件以证明其有效性。

MK-BERT w/o lex:在模型的基础上去除情感词典。

MK-BERT w/o inter:在模型的基础上去除情绪和原因预测模块基于过滤器和交叉注意力的交互信息。

MK-BERT w/o PAIM:在模型的基础上去除位置感知交互模块 PAIM。

最终消融实验的结果如表 4 所列。

表 4 消融实验结果

Table 4 Result of ablation experiment

Models	P	R	F1
MK-BERT	77.48	76.35	76.91
MK-BERT w/o lex	73.05	75.10	74.06
MK-BERT w/o inter	71.27	75.70	73.42
MK-BERT w/o PAIM	72.56	75.45	73.97

从表 4 可以看出,去掉任意模块,性能都有所下降。

1)相比原模型, MK-BERT w/o inter 的 F1 指标下降 3.49%,说明了通过过滤器和交叉注意力相交互的情绪和原因预测模块的有效性,与其他单层交互层模型相比有很大的提升。两个子任务情绪预测和原因预测对主任务情绪原因对预测有积极影响。情绪预测的结果有助于原因预测,两者之间存在相互指认的关系。

2)MK-BERT w/o PAIM 的 F1 指标下降了 2.94%,这说明位置偏置项对任务有积极影响,通过结合了位置偏置项的 PAIM,每个子句在编码情绪特定表示时都聚焦于前面子句,最终的自注意力流与情感对应原因之间的相对位置一致,使得模型能够关注目标位置周围语义一致的子句,从而消除噪声。一方面,两个子任务 EE 和 CE 之间交换了高置信子句信息,有助于模型找到更好的优化方向。另一方面,完整模型在 ECPE 上的性能也优于使用单个 PAIM 进行子句表示的模型,说明用两个 PAIM 分别对情感子句和原因子句建模可以帮助模型获得更有意义的具体的情绪和具体的原因表示,进一步提高了情绪原因对抽取的性能。

3)MK-BERT w/o lex 的 F1 指标下降了 2.85%,这说明通过情感词典引入外部知识来增强子句表示有助于提升模型性能。

4.6 参数 W 的影响

本文通过选择情绪子句和原因子句间相对距离 $|j-i|$ 小于或等于窗口 W 的候选对 (c_i^e, c_j^r) 来筛选 ECPE 任务的训练集,而不是采用所有候选对作为训练集。

本文根据超参数 W 构造训练集,进一步探讨了超参数 W 的不同值对本文模型的影响,结果如图 6 所示。

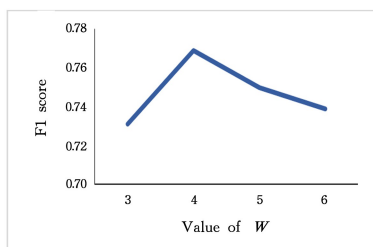


图 6 参数 W 的影响

Fig. 6 Effect of parameter W

当 W 设置为 4 时,本文模型实现了最佳性能。这证明本文窗口大小筛选情绪原因对的方法是有效的,因为它有效地缓解了标签位置不平衡的问题,并使本文模型能够专注于寻找具有情绪原因关系的候选对。

4.7 PAIM 层数的影响

PAIM 层数的不同会影响情绪原因对抽取时的性能,因此,本文对不同 PAIM 层数下的模型进行对比实验。为检验不同模块层数对模型性能所造成的影响,分别研究了不同层数情况下模型的表现。结果表明,当层数较多或较少时,分类效果下降。测试结果如表 5 所列。

表 5 PAIM 层数的影响

Table 5 Effect of the number of PAIM layers

Layers	Pairs Extraction		
	P	R	F1
1	73.82	75.26	74.53
2	77.48	76.35	76.91
3	73.20	75.09	74.12
4	73.37	74.81	74.08

可以看出,在层数为 2 时,模型取得了最佳的性能。此后随层数的增加参数也增加,性能越来越差。当 PAIM 模块的层数达到 2 时, MK-BERT 模型抽取情绪原因的性能最好。

4.8 位置不平衡样本的实验比较

如表 6 所列,为评估模型应对位置不平衡样本的能力,本文采取与 Bao 等^[26]同样的方法,选取情绪和原因子句相对距离大于 1 的不平衡样本 $test_{bias}$ 作为测试样本。通过评估可以看出, MK-BERT 在数据不平衡样本上取得了最佳结果。此外,去除 PAIM 的模型 MK-BERT w/o PAIM 比原模型下降 5.04%,这说明 PAIM 模块有助于缓解数据不平衡带来的问题。

表 6 在位置不平衡样本 $test_{bias}$ 的实验

Table 6 Experiments on $test_{bias}$

Models	$test_{bias}$	$test$
Inter-EC	34.28	67.83
RankCP	39.17	66.10
ECPE-2D	39.06	75.90
MK-BERT	45.46	76.20
MK-BERT w/o PAIM	40.42	67.81

4.9 实例分析

本文选择了 3 个例子来分析本文模型与在 ECPE 任务中性能较好的模型以及本文的消融模型之间的性能差异。表 7 列出了 3 个例子下不同模型的预测结果。

本文模型与较高性能的 RankCP 模型进行比较。对于第一个例子, RankCP 和本文模型都能很好地抽取出情绪原因对 (c_6, c_5) 。对于第二个例子,包含两个情绪原因对 (c_4, c_4) 和 (c_6, c_5) 。由于 RankCP 将预测的对进行排名,并将排名前三的对视为情绪原因对,无法抽取出正确的情绪原因对。相比之下,由于先验知识的引入,本文的 MK-BERT 成功地抽取了两个情绪原因对。在第三个例子中,虽然子句 c_6 包含了明确的情感词“忧心忡忡”,但这距离原因子句 c_{11} 的相对位置太大导致 RankCP 无法识别,而本文的 MK-BERT 由于位置增强的 PAIM 模块的引入和情绪原因对预测模块嵌入的位置信息,因此被抽取为预测结果。从上面的例子可以发现,与其他方法相比,本文方法实现了更好的性能。

与本文的消融模型 MK-BERT w/o PAIM 进行比较。在第一个例子和第二个例子中，本文模型和消融模型都正确地抽取了情绪原因对 (c_6, c_5) ，在第三个例子中，消融模型产生的情绪原因对更少，原因是消融模型忽略了位置不平衡样本

对 (c_6, c_{11}) 。这也解释了它的召回率较低，但精度高于本文模型的原因。相比之下，本文的完整模型成功地识别了所有 3 个例子。因此，本文 PAIM 模块缓解了当多个配对在一个文档中时的混淆问题。

表 7 不同模型间的案例研究
Table 7 Case studies of different models

Examples	Predicted results			Ground Truths
	RankCP	MK-BERT w/o PAIM	MK-BERT	
谈起最初的艰辛 (c_1) 王永涛觉得其实并不累 (c_2) 因为每天都生活在一种希望里面 (c_3) 总相信弟弟突然就能睁开眼睛 (c_4) 弟弟有希望 (c_5) 父母也就会更开心 (c_6)	(c_6, c_5)	(c_6, c_5)	(c_6, c_5)	(c_6, c_5)
从没想过出名希望平静生活记者发现 (c_1) 与前段时间相比 (c_2) 孙老师的精神状态好了很多 (c_3) 逐渐走出了失去女儿的悲伤中 (c_4) 面对这么多人的关心 (c_5) 他觉得很高兴 (c_6) 其实这段时间生活上多少还是有点影响的 (c_7)	(c_4, c_4)	$(c_4, c_4)(c_6, c_5)$	$(c_4, c_4)(c_6, c_5)$	$(c_4, c_4), (c_6, c_5)$
最有趣的是 (c_1) 刚当上外公的山姆和他的新女友罗冰·托马斯还生有一名 9 个月大的儿子 (c_2) 这意味着 (c_3) 提娅刚出生的女儿艾娃将必须称这名只比她 9 个月的男婴叫舅舅 (c_4) 尽管提娅的出生让山姆一家都高兴万分 (c_5) 但已经与山姆分手的前女友刚当上外婆的凯丽却有些忧心忡忡 (c_6) 如今 30 岁的凯丽是一名咖啡店员 (c_7) 王她说当年我无法应付成为少女妈妈后带来的压力 (c_8) 选择了逃跑 (c_9) 一直以来 (c_{10}) 我最坏的噩梦就是提娅会重复我的错误 (c_{11}) 年纪轻轻就怀孕生子 (c_{12}) 但万万没有想到噩梦还是变成了现实 (c_{13})	(c_5, c_5)	(c_5, c_5)	$(c_5, c_5)(c_6, c_{11})$	$(c_5, c_5), (c_6, c_{11})$

结束语 本文提出了一种新的情绪原因对端到端模型 MK-BERT 联合执行情绪抽取、情绪原因抽取和情绪原因对抽取任务，并清晰地建模了不同工作间的相互作用。模型首先利用情感词汇融入 BERT 获取子句表示。然后，在位置感知交互模块通过带有相对位置信息的 PAIM 模块缓解样本位置不平衡问题，并用两个 PAIM 模块分别学习情感子句和原因子句的表示。接着，通过情绪预测和原因预测模块间交互编码，充分学习情绪和原因子句间的相互信息。最后，本文通过多任务交互缓解传统两阶段任务的误差传播问题，正确地抽取情感、原因和情感原因对。通过大量实验证明了本文模型的有效性。

通过对数据的分析发现，数据集中存在正负例样本不平衡问题。未来将计划用对比学习等特定的方法来解决标签不平衡问题，并计划开发更多的高级模型与学习到的位置信息相结合。

参考文献

[1] LEE S Y M, CHEN Y, HUANG C R. A text-driven rule-based system for emotion cause detection[C]//Proceedings of Computational Approaches to Analysis and Generation of Emotion in Text. 2010:45-53.

[2] XIA R, DING Z. Emotion-cause pair extraction: A new task to emotion analysis in texts[C]//Proceedings of the Association for Computational Linguistics. 2019:1003-1012.

[3] VASWANI A, SHAZEER N, PARMARN, et al. Attention is all you need[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. 2017:6000-6010.

[4] GUI L, HU J, HE Y, et al. A question answering approach for emotion cause extraction [C]//Proceedings of the Conference on Empirical Methods in Natural Language Processing. 2017:1593-1602.

[5] LI X, SONG K, FENG S, et al. A co-attention neural network model for emotion cause analysis with emotional context aware-

ness [C]//Proceedings of the Conference on Empirical Methods in Natural Language Processing. 2018:4752-4757.

[6] YU X Y, WENG E. Multiple Level Hierarchical Network-Based Clause Selection for Emotion Cause Extraction[J]. IEEE Access, 2019:9071-9079.

[7] DING Z, HE H, ZHANG M, et al. From Independent Prediction to Re-ordered Prediction: Integrating Relative Position and Global Label Information to Emotion Cause Identification[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2019, 33:6343-6350.

[8] XIA R, ZHANG M, DING Z, RTHN: A RNN-Transformer Hierarchical Network for Emotion Cause Extraction[C]//IJCAI. 2019:5285-5291.

[9] HU G, LU G, ZHAO Y. FSS-GCN: A graph convolutional networks with fusion of semantic and structure for emotion cause analysis[J]. Knowledge-Based Systems, 2020, 212(1):106584.

[10] WEI P, ZHAO J, MAO W. Effective inter-clause modeling for end-to-end emotion-cause pair extraction[C]//Proceedings of the Association for Computational Linguistics. 2020:3171-3181.

[11] DING Z, XIA R, YU J. ECPE-2D: Emotion-cause pair extraction based on joint two-dimensional representation, interaction and prediction [C]//Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. 2020:3161-3170.

[12] DING Z, XIA R, YU J. End-to-end emotion-cause pair extraction based on sliding window multi-label learning[C]//Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing. 2020:3574-3583.

[13] YUAN C, FAN C, BAO J, et al. Emotion-cause pair extraction as sequence labeling based on a novel tagging scheme[C]//Proceedings of the Conference on Empirical Methods in Natural Language Processing. 2020:3568-3573.

[14] TANG H, JI D, ZHOU Q. Joint multi-level attentional model for emotion detection and emotion-cause pair extraction[J]. Neurocomputing, 2020, 409:329-340.

[15] WU S, CHEN F, WU F, et al. A multi-task learning neural net-

- work for emotion-cause pair extraction[C]// Proceedings of the 24th European Conference on Artificial Intelligence. 2020:1-8.
- [16] SONG H, SONG D. An end-to-end multi-task learning to link framework for emotion-cause pair extraction[C]// Proceedings of the International Conference on Image, Video Processing and Artificial Intelligence. 2021:13-21.
- [17] YU J, LIU W, HE Y, et al. A Mutually Auxiliary Multitask Model with Self-Distillation for Emotion-Cause Pair Extraction [J]. IEEE Access, 2021, 99: 1-1.
- [18] CHEN F, SHI Z, YANG Z, et al. Recurrent synchronization network for emotion-cause pair extraction [J]. Knowledge-based Systems, 2022: 238.
- [19] CHEN Y, HOU W, LI S, et al. End-to-End Emotion-Cause Pair Extraction with Graph Convolutional Network [C]// International Conference on Computational Linguistics. 2020.
- [20] WU Z, DAI X, XIA R. Pairwise tagging framework for end-to-end emotion-cause pair extraction [J]. Frontiers of Computer Science, 2023, 17(2): 1-10.
- [21] HUANG W C, YANG Y X, PENG Z Y, et al. Deep Neural Networks Based on Span Association Prediction for Emotion-Cause Pair Extraction [J]. Sensors, 2022, 22(10).
- [22] 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. 2019: 4171-4186.
- [23] FAN C, YUAN C, DU J, et al. Transition-based directed graph construction for emotion-cause pair extraction [C]// Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. 2020: 3707-3717.
- [24] CHENG Z, JIANG Z, YIN Y, et al. A Unified Target-Oriented Sequence-to-Sequence Model for Emotion-Cause Pair Extraction [J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2021: 1-1.
- [25] FAN C, YUAN C, GUI L, et al. Multi-task Sequence Tagging for Emotion-Cause Pair Extraction via Tag Distribution Refinement [J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2021: 1-1.
- [26] BAO Y, MA Q, WEI L, et al. Multi-Granularity Semantic Aware Graph Model for Reducing Position Bias in Emotion-Cause Pair Extraction [J]. 2022.



FU Mingrui, born in 1997, MS candidate. His main research interests include sentiment analysis and machine learning.



LI Weijiang, born in 1969, Ph.D, professor. His main research interests include information retrieval and natural language processing.