

基于多通道压缩双线性池化的情感-原因 句子对提取模型

黄晋¹ 许实¹ 蔡而聪¹ 吴志杰¹ 郭美美¹ 朱佳^{2,†}

1. 华南师范大学计算机学院, 广州 510631; 2. 浙江师范大学智能教育技术与应用重点实验室, 金华 321004;

† 通信作者, E-mail: jiazhu@zjnu.edu.cn

摘要 提出一个基于多通道压缩双线性池化的模型, 对文档中的候选情感-原因句子对进行排序。该模型利用图注意力网络提取包含位置信息的情感特定化表示和原因特定化表示, 通过局部关系学习模块, 进一步学习情感与原因句子之间的局部关系表示, 再使用多通道压缩双线性池化来融合学习情感-原因候选句子对表示。最后, 对候选句子对进行排序。实验结果表明, 与最新模型相比, 所提模型在多方面表现更优。

关键词 情感分析; 情感-原因句子对提取; 图注意力网络; 局部关系提取; 多通道压缩双线性池化

An Emotion-Cause Pair Extraction Model Based on Multichannel Compact Bilinear Pooling

HUANG Jin¹, XU Shi¹, CAI Ercong¹, WU Zhijie¹, GUO Meimei¹, ZHU Jia^{2,†}

1. School of Computer Science, South China Normal University, Guangzhou 510631; 2. The Key Laboratory of Intelligent Education Technology and Application of Zhejiang Normal University, Jinhua 321004; † Corresponding author, E-mail: jiazhu@zjnu.edu.cn

Abstract The authors propose a model based on multichannel compact bilinear pooling to rank pair candidates in a document. The proposed model firstly extracts the emotion-specific and cause-specific representation containing position information via graph attention network, then further learns the local relation representation between emotion clause and cause clause through the local relation-aware module. Finally, these representations are fused via multichannel compact bilinear pooling to learn the emotion-cause pairs representation for effective ranking. Experimental results show that the proposed approach achieves the best performance among all compared approaches on the task.

Key words sentiment analysis; emotion-cause pair extraction; graph attention network; local relation-aware module; multichannel compact bilinear pooling

随着社交媒体的发展, 情感分析领域越来越受到研究人员的关注。研究文本中隐含的情感及其成因, 对商业决策、舆情管控和信息预测等应用场景具有重大的应用价值。对此, Lee等^[1]首先提出情感-原因提取任务: 在文档中, 给定一个已标注的情感句子, 提取出导致情感发生的原因句子。针对这一任务, 有人采用基于语法规则的方法^[2-3], 也有人采用基于机器学习的方法^[4-6]。近年来, 随着深度学习的发展, 也有不少人采用深度记忆网络^[7]、互

注意力神经网络^[8]和多层次神经网络^[9]等基于深度神经网络的方法。

上述方法都需要将标注好情感标签的句子作为模型输入。Xia等^[10]认为这种先标注句子的情感, 再根据标注好的句子去检测对应原因的方法限制了其实际应用。为了使情感-原因提取任务更具有泛用性, 他们提出情感-原因句子对提取任务: 给定一个文档, 提取出所有的情感句子以及对应的原因句子。他们提出一种两步走的方法, 用于提取情感-

原因句子对,但存在的误差传播问题。针对这一问题, Ding 等^[11]提出一种情感-原因句子对二维矩阵方法, Wei 等^[12]提出一种情感-原因句子对排序方法, Fan 等^[13]则从序列标注的角度出发,提出一个基于转移系统的解决方案。

我们发现,上述模型存在句子间缺乏高效信息交互和融合的问题。为了解决这一问题,本文进一步考虑句子间的相对位置信息和局部关系,结合多通道信息融合的思想,提出一个基于多通道压缩双线性池化的情感-原因句子对提取模型。

1 任务定义

给定一个包含若干句子的文档 $d=[c_1, c_2, \dots, c_{|d|}]$ ($|d|$ 是文档的句子数),任务目标是提取出情感-原因句子对集合:

$$p = \{\dots, (c^{\text{emo}}, c^{\text{cau}}), \dots\},$$

其中, $c^{\text{emo}} \in d$, 是一个情感句子; $c^{\text{cau}} \in d$, 是其对应的原因句子。值得注意的是,一个文档中可能包含若干不同的情感-原因句子对,且一个情感句子可能有多个对应的原因句子。

2 模型与方法

如图 1 所示,本文模型的整体框架主要由 3 个部分组成:第一部分为带有位置信息的情感、原因特定化表示学习,分别学习包含位置信息的情感、原因句子向量表示;第二部分建模每一个候选句子对中情感句子与原因句子之间的局部关系;第三部分使用多通道压缩双线性池化方法来融合情感特定化表示、原因特定化表示和局部关系表示,并进一步将候选句子对排序,得出最后结果。

2.1 句子向量编码

对于一个给定的文档 $d=[c_1, c_2, \dots, c_{|d|}]$, 包含 $|d|$ 个句子。对于每个句子 $c_i=(w_{i,1}, w_{i,2}, \dots, w_{i,|c_i|})$, 包含 $|c_i|$ 个单词(其中 i 表示该句子在文档中的位置)。我们以 BERT 为句子编码器去获取句子 c_i 的向量表示 r_i , 因此文档 d 的句子编码向量为 $[r_1, r_2, \dots, r_{|d|}]$ 。

2.2 带有位置信息的情感、原因特定化表示学习

本文模型采用两路图注意力网络(graph attention network, GAT)^[14]对句子进行学习,并分别产生包含位置信息的情感特定化表示以及原因特定化表示,即该句子作为情感或原因时的向量表示。在情感与

原因特定化表示之间进行两两配对,产生情感-原因候选句子对。

2.2.1 复数值句子位置嵌入

相对位置信息对情感-原因句子对提取任务具有重要作用^[15]。现有研究多将位置信息表示为一个独立的向量,没有在训练中将位置信息融入情感、原因特定化表示之中。Wang 等^[16]提出一种复数值嵌入的方法,对绝对位置信息和相对位置信息进行建模。受此工作启发,本文改进复数值嵌入方法,并运用在句子的表示学习中,使得情感、原因句子特定化表示具有更强的位置相关性。假设句子 c_i 的向量表示为 $r_i=[x_{i,1}, x_{i,2}, \dots, x_{i,D}]$, 其中 D 是向量 r_i 的维度,复数值位置嵌入定义如下:

$$\begin{aligned} r_i^{\text{pos}} &= r_i e^{i(\omega \text{pos} + \theta)} \\ &= [x_{i,1} e^{i(\omega_1 \text{pos} + \theta_1)}, x_{i,2} e^{i(\omega_2 \text{pos} + \theta_2)}, \dots, x_{i,D} e^{i(\omega_D \text{pos} + \theta_D)}], \end{aligned} \quad (1)$$

其中, $\text{pos}(1 \leq \text{pos} \leq |d|)$ 是句子 c_i 在文档 d 中的位置, ω 和 θ 是可学习的参数, ω 决定句子对位置 pos 的敏感程度。根据欧拉公式

$$e^{ix} = \cos x + i \sin x, \quad (2)$$

复数值句子位置嵌入可以进行如下转化:

$$r_{i,\text{real}}^{\text{pos}} = r_i \cos(\omega \text{pos} + \theta), \quad (3)$$

$$r_{i,\text{imag}}^{\text{pos}} = r_i \sin(\omega \text{pos} + \theta), \quad (4)$$

其中, $r_{i,\text{real}}^{\text{pos}}$ 和 $r_{i,\text{imag}}^{\text{pos}}$ 分别是复数值嵌入的实数和虚数部分。为了能用于神经网络的输入,我们将这两部分拼接起来,作为最终的复数位置嵌入向量 r_i^{pos} :

$$r_i^{\text{pos}} = [r_{i,\text{real}}^{\text{pos}}; r_{i,\text{imag}}^{\text{pos}}]. \quad (5)$$

2.2.2 基于图注意力网络的情感/原因特定化表示学习

为了加强句子之间的信息交互,本文模型采用两路 GAT 对句子信息 r_i' 进行全局更新,分别获取融合位置信息的情感特定化表示 r_i^{emo} 和原因特定化表示 r_i^{cau} 。

以情感特定化表示学习为例说明处理流程。将 r_i 与 r_i^{pos} 拼接成 r_i' , 作为第 i 个句子的向量表示:

$$r_i' = [r_i; r_i^{\text{pos}}]. \quad (6)$$

本文模型将文档中每个句子 c_i 拼接所得的 r_i' 当成一个节点,节点之间(包括自身)两两连接,构成全连

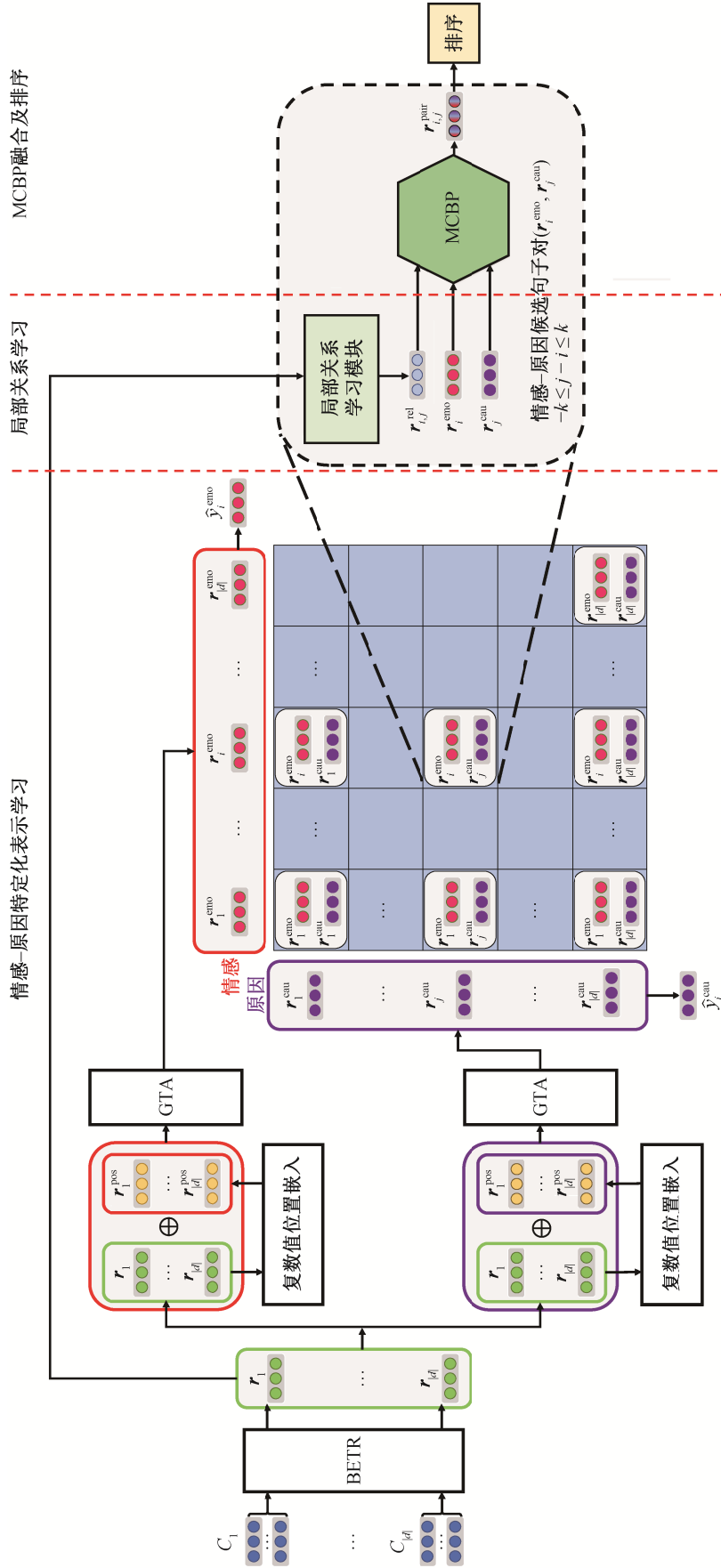


图 1 本文模型整体框架
Fig. 1 Overview of the proposed model

接图, 并利用 GAT 来深度聚合节点与邻接节点之间的信息。GAT 的实现公式如下:

$$e_{ij} = \mathbf{w}^T \tanh([\mathbf{W}\mathbf{r}'_i; \mathbf{W}\mathbf{r}'_j]), \quad (7)$$

$$\alpha_{ij} = \frac{\exp(\text{LeakyReLU}(e_{ij}))}{\sum_{k \in N_i} \exp(\text{LeakyReLU}(e_{ik}))}, \quad (8)$$

$$\mathbf{r}_i^{\text{emo}} = \text{ReLU}(\sum_{j \in N_i} \alpha_{ij} \mathbf{W}\mathbf{r}'_j), \quad (9)$$

其中, \mathbf{w}^T 和 \mathbf{W} 是可学习参数, e_{ij} 是注意力互相关系数, \tanh , LeakyReLU 和 ReLU 是激活函数, N_i 是节点 i 的邻接节点, α_{ij} 是节点 i 与 j 之间的注意力权重。

通过 GAT 的信息交互, 得到句子 c_i 的情感特定化表示 $\mathbf{r}_i^{\text{emo}}$ 。同理, 通过另一路 GAT 的信息交互, 得到原因特定化表示 $\mathbf{r}_i^{\text{cau}}$ 。然后, 将 $\mathbf{r}_i^{\text{emo}}$ 和 $\mathbf{r}_i^{\text{cau}}$ 分别输入两个 sigmoid 函数中, 获取情感预输出 \hat{y}_i^{emo} 和原因预输出 \hat{y}_i^{cau} :

$$\hat{y}_i^{\text{emo}} = \text{sigmoid}(\mathbf{W}^{\text{emo}} \mathbf{r}_i^{\text{emo}} + b^{\text{emo}}), \quad (10)$$

$$\hat{y}_i^{\text{cau}} = \text{sigmoid}(\mathbf{W}^{\text{cau}} \mathbf{r}_i^{\text{cau}} + b^{\text{cau}}). \quad (11)$$

其中, \mathbf{W}^{emo} , b^{emo} , \mathbf{W}^{cau} 和 b^{cau} 都为可学习参数, sigmoid 为激活函数。

2.2.3 情感-原因候选句子对生成

在得到文档的情感特定化表示 \mathbf{r}^{emo} 和原因特定化表示 \mathbf{r}^{cau} 后, 模型进一步生成情感-原因候选句子对。两个句子之间的相对距离是情感-原因句子对的关键, 如果太大, 它们之间的因果关系就较小, 则组成一个情感-原因句子对的可能性就很小。因此, 本模型设定一个窗口范围 k , 只有相对距离在 k 之内的句子对才可以成为候选情感-原因句子对。于是, 可以从文档 d 中构建一个候选情感-原因句子对集合:

$$\mathbf{P} = \{(\mathbf{r}_i^{\text{emo}}, \mathbf{r}_j^{\text{cau}}) | -k \leq j - i \leq k\}, \quad (12)$$

其中, $\mathbf{r}_i^{\text{emo}}$ 是句子 c_i 的情感特定化表示, $\mathbf{r}_j^{\text{cau}}$ 是句子 c_j 的原因特定化表示, $j - i$ 表示句子 c_i 与 c_j 的相对位置。

2.3 局部关系表示学习

对于每一个候选句子对 $(\mathbf{r}_i^{\text{emo}}, \mathbf{r}_j^{\text{cau}})$, 现有研究大多只关注于句子 c_i 和 c_j 本身, 对两个句子之间的局部信息有所忽略, 没有进一步进行提取学习, 导致当它们之间的相对距离较大时, c_i 与 c_j 的联系较弱, 无法提取其相互关系。为了克服这一缺点, 本文把句子 c_i 与 c_j 之间的句子也考虑进来, 提出一种局部

关系表示学习方法, 用于增强 c_i 与 c_j 之间的联系。

如图 2 所示, 局部关系学习模块通过包含注意力机制的 BiLSTM 模型来学习两个句子之间的局部信息。针对每一候选句子对(情感句子 c_i 与原因句子 c_j , 假设 $i \leq j$), 为了尽可能保留句子原有的信息, 提取从 c_i 到 c_j 之间的所有句子表示 $\mathbf{r} = [\mathbf{r}_i, \mathbf{r}_{i+1}, \dots, \mathbf{r}_j]$ 作为 BiLSTM 模型的输入, 从而学习到对应的句子隐藏表示 \mathbf{H} 。然后, 通过注意力机制对 \mathbf{H} 进行融合, 得到情感-原因之间的局部关系表示 $\mathbf{r}_{i,j}^{\text{rel}}$:

$$\mathbf{H} = [\mathbf{h}_i, \mathbf{h}_{i+1}, \dots, \mathbf{h}_j] = \text{BiLSTM}([\mathbf{r}_i, \mathbf{r}_{i+1}, \dots, \mathbf{r}_j]), \quad (13)$$

$$\mathbf{e} = \tanh(\mathbf{H}\mathbf{W} + b), \quad (14)$$

$$\alpha_i = \text{softmax}(e_i) = \frac{\exp(e_i)}{\sum_{t=i}^j \exp(e_t)}, \quad (15)$$

$$\mathbf{r}_{i,j}^{\text{rel}} = \sum_{t=i}^j \alpha_t \mathbf{h}_t. \quad (16)$$

其中, $[\mathbf{h}_i, \mathbf{h}_{i+1}, \dots, \mathbf{h}_j]$ 是句子 c_i 到 c_j 对应的隐藏表示, \mathbf{W} 和 b 是可学习参数, \tanh 是激活函数, \mathbf{e} 是注意力系数, α_i 是注意力权重, $\mathbf{r}_{i,j}^{\text{rel}}$ 是句子 c_i 与 c_j 之间的局部关系表示。

2.4 多通道压缩双线性池化(MCBP)

在求候选句子对的表示时, 现有方法只简单地将句子的情感特定化表示和原因特定化表示拼接起来, 不足以学习情感句子与原因句子之间的联系。双线性池化(bilinear pooling)^[17]是特征融合的一种方法, 通过求取两个向量外积的方式来建模特征的高阶统计信息, 从而捕获特征之间的关系。这种方法在计算机视觉细粒度分类任务中取得不错的效果。然而, 两个向量的外积维度很高, 存在学习的参数数量过大问题。在此基础上, Fukui 等^[18]提出多模态压缩双线性池化(multimodal compact bilinear pooling, MCB)方法, 有效地降低了维度, 并将其运用于跨模态看图答题领域, 将图像和文本两个模态的特征向量有效地融合, 取得很好的效果。

上述方法都是用于两个向量表示之间的融合, 而本文模型中有 3 个向量表示需要融合。为了有效地融合句子 c_i 的情感特定化表示 $\mathbf{r}_i^{\text{emo}}$ 、句子 c_j 的原因特定化表示 $\mathbf{r}_j^{\text{cau}}$ 以及局部关系表示 $\mathbf{r}_{i,j}^{\text{rel}}$ 这 3 个表示的信息, 我们将 Fukui 等^[18]的方法改进到本模型中, 提出一种多通道压缩双线性池化(multichannel compact bilinear pooling, MCBP)方法来求取候选句子对表示 $\mathbf{r}_{i,j}^{\text{pair}}$, 如图 3 所示。

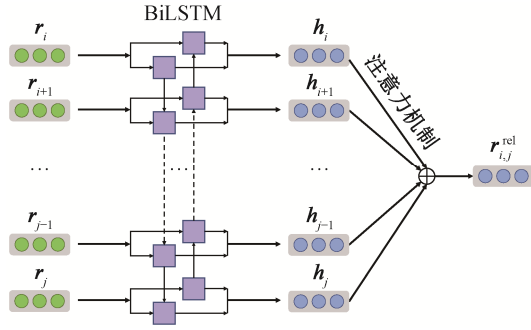
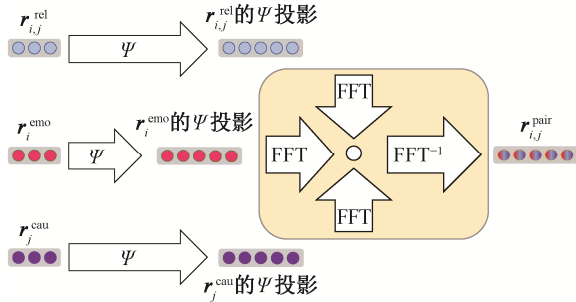


图 2 局部关系学习模块

Fig. 2 Local relation-aware module



Ψ 表示计数简述函数(Count Sketch), FFT和 FFT^{-1} 分别代表傅里叶变换和傅里叶逆变换

图 3 多通道压缩双线性池化

Fig. 3 Multichannel compact bilinear pooling

由于3个通道向量之间的外积维度过高,我们利用计数简述函数 Ψ ^[18-19],将其从高维空间投影到较低维空间,从而减少参数数量。3个向量的外积的 Ψ 投影等价于每个向量先进行 Ψ 投影,再进行卷积操作:

$$\Psi(r_i^{\text{emo}} \otimes r_j^{\text{cau}} \otimes r_{i,j}^{\text{rel}}) = \Psi(r_i^{\text{emo}}) * \Psi(r_j^{\text{cau}}) * \Psi(r_{i,j}^{\text{rel}}), \quad (17)$$

其中, \otimes 代表外积操作,*代表卷积操作。根据卷积原理,时域内的卷积等价于频域内的乘积,式(17)可以改写为

$$r_{i,j}^{\text{pair}} = \text{FFT}^{-1}(\text{FFT}(r_i^{\text{emo}}) \odot \text{FFT}(r_j^{\text{cau}}) \odot \text{FFT}(r_{i,j}^{\text{rel}})), \quad (18)$$

其中,FFT和 FFT^{-1} 分别代表傅里叶变换和傅里叶逆变换, \odot 表示按位相乘。

2.5 排序与预测

通过MCBP模块,我们得出每个候选情感-原因句子对的表示 $r_{i,j}^{\text{pair}}$ 。然后,将 $r_{i,j}^{\text{pair}}$ 输入sigmoid函数中,获取情感-原因句子对输出 $\hat{y}_{i,j}^{\text{pair}}$:

$$\hat{y}_{i,j}^{\text{pair}} = \text{sigmoid}(W^{\text{pair}} r_{i,j}^{\text{pair}} + b^{\text{pair}}). \quad (19)$$

在测试阶段,为了提取所有潜在的情感-原因句子对,本文模型采用一种基于情感词典的抽取方法^[12]。对于分数排名前 N 的降序列表 $\{p_1, p_2, \dots, p_N\}$,首先将拥有最高分数的 p_1 提取出来作为预测结果。若在其他候选句子对中的情感句子也包含情感词典中的词,则也将其提取出来作为预测结果,从而得到多个情感-原因句子对预测结果。

本文模型采用下式中的排序损失函数检测候选句子对的排序分数:

$$L^{\text{pair}} = -\sum_{p_{i,j} \in P} -(y_{i,j}^{\text{pair}} \log \hat{y}_{i,j}^{\text{pair}} + (1 - y_{i,j}^{\text{pair}}) \log (1 - \hat{y}_{i,j}^{\text{pair}})), \quad (20)$$

$y_{i,j}^{\text{pair}}$ 是情感-原因句子对 $p_{i,j}^{\text{pair}}$ 的真实值,即当 $p_{i,j}^{\text{pair}}$ 为一个情感-原因句子对时, $y_{i,j}^{\text{pair}}=1$,否则 $y_{i,j}^{\text{pair}}=0$ 。

为了更好地学习情感特定化表示和原因特定化表示,本文模型引入一个辅助损失函数来监督情感句子预输出 \hat{y}_i^{emo} 和原因句子预输出 \hat{y}_i^{cau} :

$$L^{\text{aux}} = -\sum_{i=1}^{|d|} y_i^{\text{emo}} \log \hat{y}_i^{\text{emo}} - \sum_{i=1}^{|d|} y_i^{\text{cau}} \log \hat{y}_i^{\text{cau}}, \quad (21)$$

其中, y_i^{emo} 和 y_i^{cau} 分别是文档中句子 c_i 的情感和原因标签值。即当 c_i 为情感句子时, $y_i^{\text{emo}}=1$,否则 $y_i^{\text{emo}}=0$ 。 y_i^{cau} 同理。

本文模型使用 L^{pair} 和 L^{aux} 以及L2正则化的加权和作为最终损失函数:

$$L = \lambda_1 L^{\text{pair}} + \lambda_2 L^{\text{aux}} + \lambda_3 \|\theta\|^2, \quad (22)$$

其中, L 是最终损失函数, $\lambda_1, \lambda_2, \lambda_3 \in (0, 1)$ 是权重, θ 代表模型中的所有参数, $\|\theta\|^2$ 是L2正则化。

3 实验与结果分析

3.1 数据集与评价指标

本文使用Xia等^[10]提供的基准数据集进行实验,验证模型的有效性。随机选择90%的数据作为训练数据,剩下的10%作为测试数据,重复进行20次实验,取平均值作为报告结果。使用准确率 P ,召回率 R 和F1值作为评价指标。此外,分别评估模型在情感句子提取和原因句子提取这两个子任务中的表现。

3.2 实验设置

本文模型词向量的维度为200,复数值位置向量的维度为200,产生候选句子对的窗口大小 k 设置为8,情感特定化表示 r_i^{emo} 、原因特定化表示 r_j^{cau} 以

及局部关系表示 $r_{i,j}^{rel}$ 的维度都设置为 500, 经 MCBP 整合后的候选句子对表示的维度设置为 3000。在训练阶段, 模型采用 Adam 优化器^[20]更新参数, 批处理大小和学习率分别设为 4 和 0.001。对于 Loss 函数, 参数 λ_1, λ_2 和 λ_3 都设置为 1, L2 正则化设置为 1×10^{-5} 。与 Wei 等^[12]一样, 本文也将 ANTUSD^[21]作为情感词典。

3.3 对比方法

本文对比的基本方法如下: 1) ECPE-2steps^[10]是一个两步走流水线框架, 有 Indep, Inter-CE 和 Inter-EC 共 3 种设置; 2) ECPE-2D^[11]是一个联合框架, 是 ECPE-2steps 的升级版, 它以端对端的形式整合了 2D 情感-原因句子对的表示、交互及预测, 有 Inter-EC-, Inter-EC+WC 和 Inter-EC+CR 共 3 种设置; 3) Transition-based^[13]用一个基于转移系统的模型, 将情感-原因句子对提取转化为一个类似句法解析的有向图结构, 句子编码器分别使用 LSTM 和 BERT; 4) RANKCP^[12]是一个端对端的方法, 它从排序的角度出发, 建模句子之间的内在联系, 句子编码器分别使用多层次 RNN 和 BERT, 它是情感-原因句子对提取任务中最先进的基准模型。

3.4 实验结果

表 1 展示在情感-原因句子对提取以及两个子任务(情感句子提取和原因句子提取)中结果的比较。值得注意的是, ECPE-2D, Transition-based 和

RANKCP 在使用 BERT 作为句子编码器来获得句子表示时, 效果有所提升, 表 1 中主要与其 BERT 版本进行比较。与所有基准模型对比, 本文模型在情感-原因句子对提取任务中的 F1 值、准确率和召回率都是最高的, 充分证明了本文模型的有效性。本文模型的 F1 值、准确率和召回率分别比 RANKCP (表现最优的基准模型)提升 1.88%, 2.87%和 0.85%。对比 RANKCP 模型, 本文模型的提升主要在于准确率。可能的原因是, 本文方法关注到情感句子与原因句子之间的局部关系信息, 并能有效地将情感特定化表示、原因特定化表示以及关系表示融合在一起, 最终提升了模型的准确率。

此外, 在情感句子提取和原因句子提取两个子任务中, 本文模型的表现也最佳。在情感句子抽取子任务中, 本文模型的准确率略低于 RANKCP 模型, 召回率略低于 ECPE-2D 模型, 但并不影响 F1 值的表现, 本文模型的 F1 值仍为最高值 90.76%。在原因句子抽取子任务中, 本文模型的 F1 值、准确率和召回率都最高。由于情感-原因句子对抽取任务和两个子任务具有相关性, 也从另一方面证明了本文方法的有效性。

3.5 情感-原因特定化表示学习的有效性验证

为了检验模型情感特定化表示和原因特定化表示学习的有效性, 我们比较 BiLSTM 与 GAT 两种神经网络及不同位置嵌入方法组合的效果, 如表 2 所

表 1 本文模型与基准模型比较结果(%)
Table 1 Comparison of proposed method with baseline models (%)

方法	情感-原因句子对提取			情感句子提取			原因句子提取			
	F1	P	R	F1	P	R	F1	P	R	
ECPE-2Steps ^[10]	Indep	58.18	68.32	50.82	82.10	83.75	80.71	62.05	69.02	56.73
	Inter-CE	59.01	69.02	51.35	83.00	84.94	81.22	61.51	68.09	56.34
	Inter-EC	61.28	67.21	57.05	82.30	83.64	81.07	65.07	70.41	60.83
ECPE-2D(BERT) ^[11]	Inter-EC	67.47	70.73	64.86	88.88	86.22	91.82	70.96	73.46	68.79
	Inter-EC+WC	68.89	72.92	65.44	89.10	86.27	92.21	71.23	73.36	69.34
	Inter-EC+CR	68.37	69.35	67.85	88.78	85.48	92.44	70.87	72.72	69.27
Transition-based ^[13]	LSTM	64.34	65.15	63.54	82.56	80.80	84.39	66.36	67.42	65.34
	BERT	67.99	73.74	63.07	84.74	87.16	82.44	69.74	75.62	64.71
RANKCP ^[12]	RNN	66.10	66.98	65.46	85.48	87.03	84.06	68.24	69.27	67.43
	BERT	73.60	71.19	76.30	90.57	91.23	89.99	76.15	74.61	77.88
本文模型		75.48	74.06	77.15	90.76	90.67	90.91	77.81	77.14	78.68

说明: 粗体数字为同一评价指标的最高值, 下同。

示,其中“绝对位置嵌入”代表对不同的句子位置 $pos(1 \leq pos \leq |d|)$ 进行随机初始化,从而得到其位置嵌入。

在不使用神经网络进行情感特定化表示和原因特定化表示学习时,模型的效果很差,说明句子间信息交互的重要性。GAT的效果比BiLSTM好,说明将文档建模成全连接图并进行信息交互的方法是有效的,图注意力机制使句子之间的相关性更强。

对GAT而言,不使用位置嵌入时模型的效果最差,表明加入位置嵌入的必要性。使用绝对值位置嵌入与不使用位置嵌入的效果相差不大,这是由于绝对值位置嵌入实际上只是建模了句子的绝对位置关系,并不能很好地建模句子之间的相对位置关系,因此并没有带来明显的效果提升。本文采用的复数值位置嵌入方法能有效地将句子的位置信息融入句子向量表示中,使得情感特定化表示和原因特定化表示更具位置相关性。

3.6 局部关系表示模块有效性验证

为了检验局部关系表示模块的有效性,本文根据真实标签中情感句子与原因句子的相对距离,将测试集划分成相对距离 ≤ 1 和相对距离 > 1 两个子集,并在这两个子集上对比去掉局部关系模块和完整模型的结果(表3)。可以看到,对于相对距离 ≤ 1 的子集,去掉关系感知模块之后,模型效果只有小幅下

表2 不同神经网络和位置嵌入方式的效果(%)

Table 2 Effect of different neural networks and position embeddings (%)

神经网络	位置嵌入方式	F1	P	R
不使用	不使用	48.41	46.28	50.86
	不使用	67.43	64.73	70.58
BiLSTM	绝对位置嵌入	69.26	67.43	71.56
	复数值位置嵌入	71.38	70.06	73.12
	不使用	72.87	71.03	74.86
GAT	绝对位置嵌入	73.25	71.41	75.23
	复数值位置嵌入	75.48	74.06	77.15

表3 局部关系表示模块的效果比较(%)

Table 3 Effect of local relation-aware module (%)

相对距离	方法	F1	P	R
≤ 1	去掉局部关系	80.98	77.95	84.39
	完整模型	82.18	85.05	84.52
> 1	去掉局部关系	33.04	31.11	35.43
	完整模型	44.20	46.60	42.83

表4 不同融合方式的效果(%)

Table 4 Effect of different fusion methods (%)

融合方式	F1	P	R
按位相加	70.35	67.76	73.32
拼接	73.44	71.71	75.40
MCBP	75.48	74.06	77.15

降;对于相对距离 > 1 的子集,因为情感句子与原因句子的相对距离变大,增大了识别难度,所以在去掉关系感知模块之后,识别效果大大下降(F1, P和R分别下降11.16%, 15.49%和7.4%),说明对于相对距离较大的情况,本文局部关系学习模块充分地考虑到两者之间其他句子的信息,从而提高了模型整体的效果。

3.7 MCBP有效性验证

通过对比向量表示之间不同的融合方式来验证本文MCBP模块的有效性。对于3个输入(情感特定化表示 r_i^{emo} 、原因特定化表示 r_j^{cau} 和局部关系表示 $r_{i,j}^{cau}$),采取按位相加、拼接和MCBP这3种融合方式来学习候选句子对的最终向量表示,实验结果列于表4。可以看出,按位加的方法不能很好地获取最终表示,所以效果最差。拼接的方法直接将向量拼接起来,不会引入太多噪声信息,所以效果好一点。本文提出的MCBP融合方法将3个特征表示有效地融合起来,3个评价指标(F1, P和R)都明显提高,得到更有效的情感-原因候选句子对表示。

4 结语

本文提出一个基于多通道压缩双线性池化的模型(MCBP),对文档中的候选句子对进行排序,并通过消融实验验证每一个模块的有效性。本文的主要贡献如下: 1) 所提模型利用GAT提取包含位置信息的情感特定化表示和原因特定化表示; 2) 构建局部关系学习模块,对情感句子与原因句子之间的局部关系进行学习,增强远距离句子之间的联系; 3) 使用MCBP,有效地融合多路特征向量表示并进行排序。实验结果表明,在所有参与对比的方法中,本文方法表现最优。

参考文献

[1] Lee S Y M, Chen Y, Huang C R. A text-driven rule-based system for emotion cause detection // Proceedings of the NAACL HLT 2010 Workshop on Compu-

- tational Approaches to Analysis and Generation of Emotion in Text. Los Angeles, 2010: 45–53
- [2] Russo I, Caselli T, Rubino F, et al. Emocause: an easy-adaptable approach to emotion cause contexts // Association for Computational Linguistics. Portland, OR, 2011: 153–160
- [3] Gui L, Yuan L, Xu R, et al. Emotion cause detection with linguistic construction in chinese weibo text // CCF International Conference on Natural Language Processing and Chinese Computing. Berlin: Springer, 2014: 457–464
- [4] Ghazi D, Inkpen D, Szpakowicz S. Detecting emotion stimuli in emotion-bearing sentences // International Conference on Intelligent Text Processing and Computational Linguistics. Cham: Springer, 2015: 152–165
- [5] Song S, Meng Y. Detecting concept-level emotion cause in microblogging // Proceedings of the 24th International Conference on World Wide Web. Firenze, 2015: 119–120
- [6] Cheng X, Chen Y, Cheng B, et al. An emotion cause corpus for chinese microblogs with multiple-user structures. ACM Transactions on Asian and Low-Resource Language Information Processing, 2017, 17 (1): 1–19
- [7] Gui L, Hu J, He Y, et al. A question answering approach for emotion cause extraction // Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing. Copenhagen, 2017: 1593–1602
- [8] 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, 2018: 4752–4757
- [9] Xia R, Zhang M, Ding Z. RTHN: a RNN-transformer hierarchical network for emotion cause extraction // Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence. Macao, 2019: 5285–5291
- [10] 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, 2019: 1003–1012
- [11] 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. Seattle, 2020: 3161–3170
- [12] Wei P, Zhao J, Mao W. Effective inter-clause modeling for end-to-end emotion-cause pair extraction // Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. Seattle, 2020: 3171–3181
- [13] Fan C, Yuan C, Du J, et al. Transition-based directed graph construction for emotion-cause pair extraction // Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. Seattle, 2020: 3707–3717
- [14] Veličković P, Cucurull G, Casanova A, et al. Graph attention networks // Proceedings of the 7th International Conference on Learning Representations. Vancouver, 2018: 1–12
- [15] 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, 2016: 1639–1649
- [16] Wang B, Zhao D, Lioma C, et al. Encoding word order in complex embeddings [C/OL] // Proceedings of the 8th International Conference on Learning Representations. (2019–09–26) [2021–04–26]. <https://openreview.net/forum?id=Hke-WTVtwr>
- [17] Lin T Y, RoyChowdhury A, Maji S. Bilinear CNN models for fine-grained visual recognition // Proceedings of the IEEE International Conference on Computer Vision. Santiago, 2015: 1449–1457
- [18] Fukui A, Park D H, Yang D, et al. Multimodal compact bilinear pooling for visual question answering and visual grounding // Conference on Empirical Methods in Natural Language Processing. Austin, 2016: 457–468
- [19] Charikar M, Chen K, Farach-Colton M. Finding frequent items in data streams // International Colloquium on Automata, Languages, and Programming. Berlin: Springer, 2002: 693–703
- [20] Kingma D P, Ba J. Adam: a method for stochastic optimization // Proceedings of the 3th International Conference on Learning Representations. (2014–12–22) [2020–05–20]. <https://arxiv.org/abs/1412.6980>
- [21] Wang S M, Ku L W. ANTUSD: a large Chinese sentiment dictionary // Proceedings of the Tenth International Conference on Language Resources and Evaluation. Portorož, 2016: 2697–2702