arXiv:2303.01795v1 [cs.CL] 3 Mar 2023

PAGE：对话情绪原因蕴含的位置感知图模型

Xiaojie Gu[1], *Renze Lou*[2],*Lin Sun*[1*]、*Shangxin Li*[1]

[1]杭州城市大学计算机科学系，杭州，中国[2] 宾夕法尼亚州立大学计算机科学与工程系，大学城，美国

摘要

对话**因**果**情**绪**蕴**含（C_2E_2）是一项旨在识别对话中与目标情绪相对应原因的任务。对话中话语的顺序会影响因果推断。然而，目前大多数位置编码策略都忽略了话语和说话人之间的顺序关系。为解决这一问题，我们设计了一种新颖的位置感知图，用于编码整个对话，从而充分建模话语之间的因果关系。全面的实验表明，我们的方法在两个具有挑战性的测试集上始终达到最先进的性能，证明了模型的有效性。我们的源代码可在 Git Hub[1] 上获取。

索引词— 情绪原因蕴含、图神经网络、位置编码

1. 引言

对于传递特定情绪的目标话语，C_2E_2 旨在从对话历史中识别出导致目标情绪的因果语句。这项新颖任务对于设计当前的对话智能体至关重要，例如生成富有同理心的回应 [1] 和提供情绪咨询 [2]。此外，它还为提升情感模型的可解释性提供了一种潜在途径 [3]。

Lee 等人 [4] 首次提出了情绪原因抽取（ECE），指出了这项任务的重要性，并将其置于句子级分类范式中进行研究。早期研究采用基于规则的策略 [5]和传统机器学习方法 [6, 7]来处理这一问题。作为 ECE 的进一步发展，C_2E_2 考虑了更具挑战性的对话场景。迄今为止，只有少数研究 [3, 8]关注了 C_2E_2。Porial 等人 [3]仅将一个目标话语与其他话语配对，这会丢失上下文信息，并破坏话语之间的因果关系。Li 等人 [8] 利用常识知识来辅助识别原因。根据我们的观察，上下文信息对话语理解有显著影响；如果不考虑话语之间的位置，就难以判断它们之间是否存在相应的因果关系 [9, 10]。对此，一种简单的方法是将绝对位置嵌入直接与话语表示拼接 [11, 12]。然而，这种方案常常会导致聚合无信息量的上下文，从而引入推断噪声（即因果无关上下文）。为此，Xia 等人 [13] 采用多头注意力机制对位置信息进行加权。Ding 等人 [9] 将窗口大小设定为仅考虑目标话语的邻接关系。然而，这些位置编码策略针对的是文档，而非对话。因此，它们忽视了至关重要的说话人间依赖关系，而这种依赖关系对于理解对话不可或缺 [14]。

为解决上述问题，我们提出了 PAGE（**P**位置感知**A**图模型，用于**G**情绪原因**E**蕴含），并设计了一种新颖的相对位置编码方案，以区分不同说话者的话语，从而进行更好的推理。直观而言，相对位置在基于对话的因果推断中发挥着至关重要的作用。例如，来自不同说话者的两个与因果关系相关的话语分别是“嘿，你明天想去看电影吗？”和“听起来是个好主意。”。如果顺序颠倒，我们可能就无法意识到“好主意”指的是“去看电影”。因此，我们根据不同说话者之间话语的相对距离构建位置关系。此外，情绪与原因之间存在明确的拓扑关系 [15], ，因此我们利用图神经网络对整个对话上下文进行编码。我们在最新的基准数据集上评估了所提出的方法。实验结果表明，与其他强基线方法相比，我们的方法取得了有竞争力的表现。总而言之，本文的主要贡献如下：

- 我们提出了一种位置编码策略，可增强 C_2E_2 中的情绪原因蕴含能力和对话上下文理解能力。我们设计了一种新颖的位置感知图，以更好地聚合整个对话。
- 我们进行了全面的实验，以验证所提方法的有效性，并进行了深入分析。我们的模型在基准数据集上取得了最先进的性能，尤其是在 IE测试集上的 F1分数绝对提升了 8.1%。

*通讯作者。[1]

https://github.com/XiaojieGu/PAGE

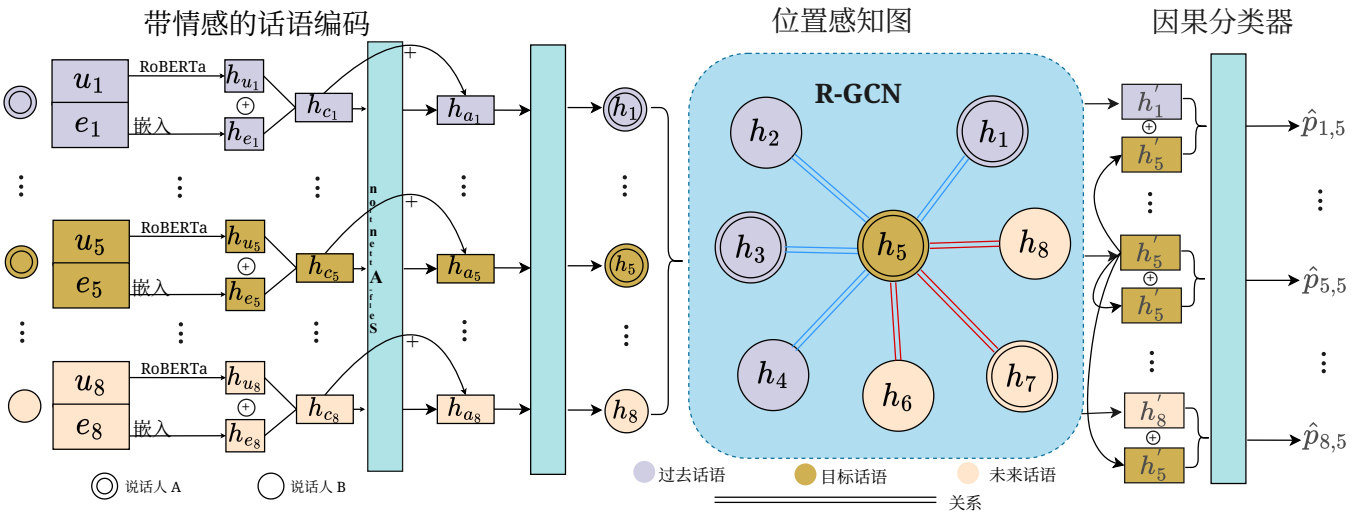

图1。PAGE 的整体架构。

2. 方法

如图1所示，我们的框架由三个部分组成，分别是带情感的话语编码（§ 2.1）、位置感知图（PaG）（§ 2.2）和因果分类器（§ 2.3）。给定一个对话 $C = \{u_1, u_2, \ldots, u_k\}$，其中存在说话者 $S_t \in \{A, B\}$的目标话语 $u_t \in C$。每个 u_t 都标注有一种特定的非中性情绪（例如，“快乐”、“愤怒”、“悲伤”）。我们的目标是识别 u_t的对话历史中的因果语句。也就是说，判断一个 u_i （$i \leq t$）是否是 u_t 的原因。

2.1. 带情感的语句编码

对于由 m个词语组成的话语 $u_n = \{w_1, w_2, \ldots, w_m\}$，我们在其开头和结尾分别添加两个特殊标记 $[CLS]$ 和 $[SEP]$。然后，我们使用 RoBERTa[16] 进行句子级编码，并将最后一层的隐藏状态作为词级表示：

$$h_w = \text{RoBERTa}\left([CLS], w_1, w_2, \ldots, w_m, [SEP]\right), \quad (1)$$

其中，$h_w \in \mathbb{R}^{m \times d_R}$ 和 $d_R = 768$ 表示 RoBERTa 中词级表示的维度。然后，我们通过对 h_w进行线性投影 $W_u \in \mathbb{R}^{d_R \times d_u}$，得到话语表示 $h_u \in \mathbb{R}^{d_u}$。

对话中的情感信息可以在说话者之间传递 [17]，，这有助于蕴含判断 [18]；因此，我们将情感嵌入与话语表示拼接，即 $h_c = h_e \oplus h_u$。为了从多个方面捕捉语句特征，我们采用多头自注意力机制 [19]，，这有助于句子级情感分析 [20]。Q、K 和 V 向量的值与 h_c相同：

$$head_N = \text{softmax}\left(\frac{QK^T}{\sqrt{d_u}}\right)V, \quad (2)$$

其中，$head_N \in \mathbb{R}^{\frac{d_u}{N}}$，N 是头的数量。我们将各个头拼接起来，得到注意力输出 h_a，并通过 $x = h_a + h_c$将其与 h_c 相加。然后，接入一个带有单个隐藏层和残差连接的多层感知机（MLP）：

$$h_n = \sigma\left(MLP(x)\right) + x, \quad (3)$$

其中，多层感知机的输出维度为 $\mathbb{R}^{d_u}$，而 $\sigma(\cdot)$ 是一个 Sigmoid 函数。

2.2. 位置感知图

为了缓解长对话中话语之间的长期依赖关系，我们利用图神经网络进行位置感知编码。我们设计了一个有向图，记为 $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{R})$。对话中的话语 $u_n \in \mathcal{V}$ 表示一个节点，其初始表示为 h_n，而 $r_{o,t} \in \mathcal{R}$ 是边 $(u_o, r_{o,t}, u_t)$ 的类型 $\in \mathcal{E}$，其中 u_t 表示目标话语，u_o ($1 \leq o \leq n$) 表示其他话语。

相对位置在话语之间因果信息的传递中发挥着重要作用；因此，我们根据相对位置关系来表示节点之间边的类型。此外，同一说话者或不同说话者的话语之间的序列关系，有助于理解话语中的信息并增强蕴涵能力。为此，我们构建一种基于相对距离的$r_{o,t} D_{o,t}$ 位置关系，用以表示

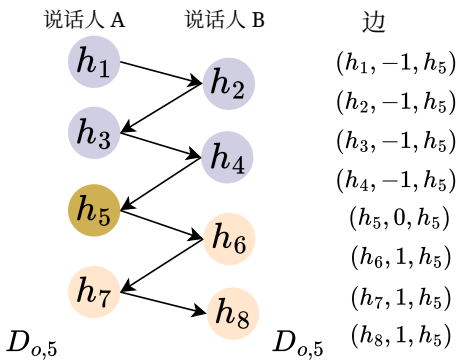

图 2。相对位置编码示例如图所示。目标话语为 u_5，窗口大小为 $w = 1$。紫色表示过去话语，橙色表示目标话语，奶油色表示未来话语。

目标话语 u_t 与相邻话语 u_o之间的距离：

$$D_{o,t} = \begin{cases} \frac{o-t}{2} & S_o = S_t \\ -1 & S_o \neq S_t \; and \; t = o \pm 1. \\ \frac{o-t-1}{2} & others \end{cases} \tag{4}$$

如图 2 所示，在说话者相同或不同的情况下，目标话语 u_5 与周围话语 u_3 和 u_4 的相对距离相同，即 $D_{3,5} = D_{4,5}$，因此它们被视为具有相同的位置关系，即 $r_{3,5} = r_{4,5}$。对于话语 u_j，其中 $j > t$，目标话语之后的话语相比目标话语之前的部分并不占主导地位 [21], ，但它们仍可作为对话理解的补充上下文信息。我们不能完全忽略它们，因此将目标话语之后的部分视为具有相同的相对位置关系。本文将其设为 1。此外，先前研究表明，大多数原因都位于情绪 [10], 附近；如果两个话语相距较远，它们之间的因果关系就会减弱。因此，与目标话语附近的话语相比，我们无需给予它们同等的关注。为缓解这一问题，我们设置了一个窗口大小，使得 u_t 与 u_j 之间的相对位置关系相同。简言之，我们得到从 u_o 到 u_t 的相对位置关系 $r_{o,t}$：

$$r_{o,t} = \begin{cases} -w & D_{o,t} < -w \\ D_{o,t} & D_{o,t} \geq -w \; and \; o \leq t. \\ 1 & o > t \end{cases} \tag{5}$$

我们使用 R-GCN [22], ，它能够整合节点之间的不同关系，以获得有利于位置感知变换的最终话语表示：

$$h_t^{'} = \sigma \left(\sum_{r \in \mathcal{R}} \sum_{o \in \mathcal{N}_t^r} \frac{1}{c_{t,r}} W_r h_o + W_0 h_t \right), \tag{6}$$

测试集	对话	Utt.	Avg.	正样本对	负样本对
DD	225	2,405	10	1,894	26,814
IE	16	665	41	1,080	11,305

表1。RECCON测试集的统计数据，其中“DD”和“IE”分别代表RECCON-DD和RECCON-IE测试集。平均值表示每段对话的话语平均数量。

其中，$\mathcal{N}_t^r$表示关系 r下节点 t 的邻居节点集合，$c_{t,r}$ 是正则化常数，$W_r \in \mathbb{R}^{d_u \times d_u}$ 是用于通过关系 r变换邻域节点 o的可训练参数。

2.3. 因果分类器

我们将目标话语 u_t 和另一话语 u_o的最终表示进行拼接，然后使用一个包含单个隐藏层和 Sigmoid 函数的多层感知机，得到逻辑值 $\hat{}p_{o,t}$ 并使用交叉熵损失函数进行二分类：

$$\hat{p}_{o,t} = \sigma \left(\text{MLP} \left(h_o^{'} \oplus h_t^{'} \right) \right). \tag{7}$$

3. 实验

3.1. 实验设置

数据集。 我们使用了近期提出的 C_2E_2, 基准数据集 RECCON [3], ，其中对话中的每个话语都附有人工标注的情绪标签。该数据集取样自两个广泛用于对话情绪识别任务的数据集 [23, 24]。为全面评估模型，RECCON 包含两个测试集：RECCON-DD 和 RECCON-IE。值得注意的是，RECCON-DD 的数据来源与训练集相同，而 RECCON-IE 则不同。换言之，RECCON-IE 代表一项具有挑战性的跨领域评估。具体而言，如表1所示，IE 中每段对话的平均长度都长于 DD。此外，IE 对话中的情绪转变比 DD 更频繁，因此要求模型具备更复杂的情绪理解能力[3]。遵循 Poria 等人 [3], 的做法，我们省略了那些罕见出现的未来原因，并采用 RECCON 的“Fold-1”作为负样本。

实现细节。 训练时，我们选用交叉熵损失函数，并将学习率设为 3e-5，批量大小设为 4。话语表示的维度 d_u 和多层感知机的隐藏层大小均设为 300，注意力头数和窗口大小分别设为 6 和 3。对于超参数$c_{t,r}$，我们将其固定为 2。

模型	DD			IE		
	负类 F1	正类 F1	宏平均 F1	负类 F1	正类 F1	宏平均 F1
基线[3]	88.74	64.28	76.51	95.67	28.02	61.85
ECPE-MLL[25]	94.68	48.48	71.59	93.55	20.23	57.65
ECPE-2D[9]	94.96	55.50	75.23	**97.39**	28.67	63.03
RankCP[26]	**97.30**	33.00	65.15	92.24	15.12	54.75
KEC♣ [8]	95.74(±0.05)	66.76(±0.33)	81.25(±0.17)	86.08(±0.46)	19.72(±1.71)	52.9(±0.8)
PAGE	95.80(±0.06)	**68.80**(±0.11)	**82.30**(±0.05)	96.41(±0.25)	**45.96**(±0.82)	**71.19**(±0.52)
-不使用 PaG	93.36(±0.46)	52.94(±0.97)	73.15(±0.31)	84.53(±2.0)	21.62(±0.32)	53.07(±0.89)

表 2。RECCON 上的结果。我们使用宏平均 F1 作为整体指标，而正类 F1 和负类 F1 分别表示正负样本对上的 F1 分数。 ♣：由于 KEC [8] 仅报告了 DD 上的结果，我们在 IE 测试集上运行 KEC 算法并报告结果。

3.2. 结果

我们将所提出的方法与强基线进行比较。如表 2 所示，前四个基线的分数由 Poria 等人 [3], 报告，这些分数是多次重复实验中表现最佳的一次结果。为便于比较，我们遵循 Li 等人 [8] （即 KEC）的做法，报告五次随机运行所得的平均 F1分数及相应方差。

总体而言，除了正负样本数量类别不平衡导致的负类 F1 外，如表 1 所示，我们在两个测试集的两项主要指标（即正类 F1 和宏平均 F1）上均取得了最先进的性能。与其他方法相比，我们的相对位置编码策略能够充分考虑说话者间依赖关系，从而有效增强对话语的理解。如表 1 所示，IE 集中的对话比 DD 集中的对话更长，因此 IE 集中的情绪原因转移更多。因此，与 DD 集相比，检测 IE 集中的原因更具挑战性。得益于 PaG 结构在聚合长期上下文信息方面的优势，我们的模型在 IE 上仍然取得了出色的结果，与其他基线方法相比表现差距显著。由于现实世界应用中的对话通常较为冗长， [27], 这一令人鼓舞的表现证明了我们方法的鲁棒性和实用性。

为进一步证明我们的位置编码策略的有效性，我们进行了额外的消融研究。如表 2 最后一行所示，移除 PaG 后，性能显著下降，尤其是在 IE测试集中的长篇对话中。这一结果表明，位
置对对话中的情绪原因推断有益，尤其有助于处理较长的对话，展现了我们模型出色的泛化能力和潜在的实用性。

3.3 窗口大小的影响

窗口大小可能会对模型性能产生潜在影响。因此，我们通过实验考察不同窗口大小下的性能波动。请注意

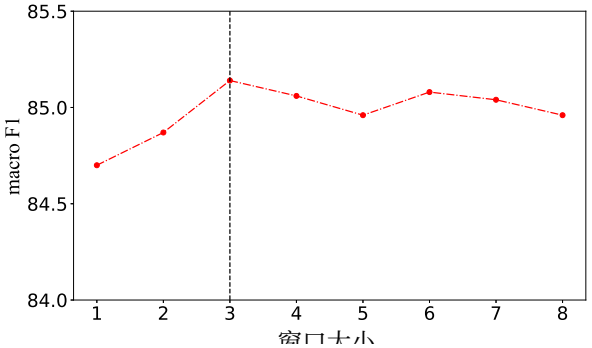

图 3。不同窗口大小下 DD 验证集上的性能。

RECCON 的验证集与 DD 测试集来自相同来源，而 IE 没有对应的验证集。图 3 展示了实验结果。从理论上讲，较大的窗口大小会使图中包含更多位置类别，但也会增加图网络的复杂度。窗口过小时，只有目标话语的邻接关系包含有效的位置信号。如图 3 所示，将窗口大小从 1 增加到 3 时，性能有所提升，这是因为多样的位置类别有助于上下文理解。然而，较大的窗口大小（> 3）会带来负面影响，这是由于远距离话语（即因果无关上下文）产生了推断噪声。因此，我们在 PAGE 中选择窗口大小 $= 3$ 。

4. 结论

本文提出一种用于 C_2E_2任务的位置感知图，有助于情绪原因蕴含。具体而言，我们的框架采用一种融入说话者间依存关系的有效位置编码策略，从而增强复杂情绪原因推理能力。我们的 PAGE 模型在两个具有挑战性的测试集上均取得了最先进的性能；与以往模型相比，宏平均 F1 分别提升了 1.1% 和 8.1%。

5. 参考文献

[1] Jamin Shin、Peng Xu、Andrea Madotto 和 Pascale Fung，《通过预判用户的情感倾向生成共情式回应》，载于*ICASSP*。IEEE，2020年，第 7989–7993 页。[2] Siyang Liu、Chujie Zheng、Orianna Demasi、Sahand Sabour、Yu Li、Zhou Yu、Yong Jiang 和 Minlie Huang，《迈向情感支持对话系统》，*arXiv* 预印本 *arXiv:2106.01144*，2021年。[3] Soujanya Poria、Navonil Majumder、Devamanyu Hazarika、Deepanway Ghosal、Rishabh Bhardwaj、Samson Yu Bai Jian、Pengfei Hong、Romila Ghosh、Abhinaba Roy、Niyati Chhaya 等，《识别对话中的情绪原因》，认知计算，第 13 卷，第 5 期，第 1317–1332 页，2021年。[4] Sophia Yat Mei Lee、Ying Chen 和 Chu-Ren Huang，《用于情绪原因检测的文本驱动规则系统》，载于 *NAACL HLT 2010*，2010年，第 45–53 页。[5] Ying Chen、Sophia Yat Mei Lee、Shoushan Li 和 Chu-Ren Huang，《运用语言结构检测情绪原因》，载于*COLING*，2010年，第 179–187 页。[6] Diman Ghazi、Diana Inkpen 和 Stan Szpakowicz，《检测情绪表达句中的情绪刺激因素》，载于 *CICLing*。Springer，2015年，第 152–165 页。[7]Irene Russo、Tommaso Caselli、Francesco Rubino、Ester Boldrini、Patricio Martínez-Barco 等，《Emocause：一种易于适配的情绪原因语境处理方法》，载于*ACL*，2011年。[8] Jiangnan Li、Fandong Meng、Zheng Lin、Rui Liu、Peng Fu、Yanan Cao、Weiping Wang 和 Jie Zhou，《中性话语也是原因：利用社会常识知识增强对话中的因果情绪蕴含》，*arXiv* 预印本 *arXiv:2205.00759*，2022年。[9] Zixiang Ding、Rui Xia 和 Jianfei Yu，《ECPE-2D：基于联合二维表示、交互和预测的情绪原因对抽取》，载于*ACL*，2020年，第 3161–3170 页。[10] Hanqi Yan、Lin Gui、Gabriele Pergola 和 Yulan He，《位置偏差缓解：一种用于情绪原因抽取的知识感知图模型》，载于*ACL*，2021年，第 3364–3375 页。[11] Zixiang Ding、Huihui He、Mengran Zhang 和 Rui Xia，《从独立预测到重排序预测：将相对位置和全局标签信息融入情绪原因识别》，载于*AAAI*，2019年，第 6343–6350 页。[12] Xiangju Li、Shi Feng、Daling Wang 和 Yifei Zhang，《基于多重注意力神经网络的上下文感知情绪原因分析》，《基于知识的系统》，第 174 卷，第 205–218 页，2019年。[13] Rui Xia、Mengran Zhang 和 Zixiang Ding，《RTHN：用于情绪原因抽取的 RNN-Transformer 层次网络》，*IJCAI*，2019年。[14] Deepanway Ghosal、Navonil Majumder、Soujanya Poria、Niyati Chhaya 和 Alexander Gelbukh，《DialogueGCN：用于对话情绪识别的图卷积神经网络》，载于*EMNLP-IJCNLP*，2019年，第 154–164 页。[15] Agnes Moors，《自动建构性评价作为情绪的候选原因》，《情绪评论》，第 2 卷，156，2010年 2，139号，第 – 页。[16] Yinhan Liu、Myle Ott、Naman Goyal、Jingfei Du、Mandar Joshi、Danqi Chen、Omer Levy、Mike Lewis、Luke Zettle- moyer 和 Veselin Stoyanov，“RoBERTa：一种经过稳健优化的 BERT 预训练方法”，*arXiv* 预印本 *arXiv:1907 11692*，2019年。.[17] Xiaohui Song、Liangjun Zang、Rong Zhang、Songlin Hu 和 Longtao Huang，“Emotionflow：捕捉对话层面的情绪变化”，载于 *ICASSP*。IEEE，2022年，第 8542–8546 页。[18] Guimin Hu、Guangming Lu 和 Yi Zhao，“基于文档级上下文的双向层次注意力网络，用于情绪原因抽取”，载于*EMNLP*，2021年，第 558–568 页。[19] Ashish Vaswani、Noam Shazeer、Niki Parmar、Jakob Uszkoreit、Llion Jones、Aidan N Gomez、Lukasz Kaiser 和 Illia Polosukhin，“Attention is all you need”，载于神经信息处理系统进展，第 30 卷，2017年。[20] Sheng-Chieh Lin、Wen-Yuh Su、Po-Chuan Chien、Ming-Feng Tsai 和 Chuan-Ju Wang，“用于情感分析的自注意力情感句子嵌入”，载于*ICASSP*。IEEE，2020年，第 1678–1682 页。[21]Jiangnan Li、Zheng Lin、Peng Fu 和 Weiping Wang，“ 过去、现在与未来：通过心理学知识的结构建模进行对话情绪识别”，载于*EMNLP*，2021年，第 1204–1214 页。[22] Michael Schlichtkrull、Thomas N Kipf、Peter Bloem、Rianne van den Berg、Ivan Titov 和 Max Welling，“利用图卷积网络对关系数据建模”，载于欧洲语义网会议。Springer，2018年，第 593–607 页。[23]Carlos Busso、Murtaza Bulut、Chi-Chun Lee、Abe Kazemzadeh、Emily Mower、Samuel Kim、Jeannette N Chang、Sungbok Lee 和 Shrikanth S Narayanan，“IEMOCAP：交互式情绪双人动作捕捉数据库”，语言资源与评估，第 42 卷，359，2008年4，335第。– 页。[24] Yanran Li、Hui Su、Xiaoyu Shen、Wenjie Li、Ziqiang Cao 和 Shuzi Niu，“DailyDialog：人工标注的多轮对话数据集”，载于*IJCNLP*（第 *1* 卷：长篇论文），2017年，第 986–995 页。[25] Zixiang Ding、Rui Xia 和 Jianfei Yu，“基于滑动窗口多标签学习的端到端情绪原因对抽取”，载于*EMNLP*，2020年，第 3574–3583 页。[26] Penghui Wei、Jiahao Zhao 和 Wenji Mao，“面向端到端情绪原因对抽取的有效分句建模”，载于*ACL*，2020年，第 3171–3181 页。[27] Wenxiang Jiao、Michael Lyu 和 Irwin King，“通过注意力门控层次记忆网络实现实时情绪识别”，载于*AAAI*，2020年，第 8002–8009 页。