



计算机科学

COMPUTER SCIENCE

对话场景下的情感引导问题生成模型

胥备, 许鹏

引用本文

胥备, 许鹏. 对话场景下的情感引导问题生成模型[J]. 计算机科学, 2024, 51(11): 265-272.

XU Bei, XU Peng. Emotion Elicited Question Generation Model in Dialogue Scenarios[J]. Computer Science, 2024, 51(11): 265-272.

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>

[基于视觉语义与提示学习的多模态情感分析模型](#)

Multimodal Sentiment Analysis Model Based on Visual Semantics and Prompt Learning

计算机科学, 2024, 51(9): 250-257. <https://doi.org/10.11896/jsjcx.230600047>

[基于改进高斯混合变分自编码器的半监督情感音乐生成](#)

Semi-supervised Emotional Music Generation Method Based on Improved Gaussian Mixture Variational Autoencoders

计算机科学, 2024, 51(8): 281-296. <https://doi.org/10.11896/jsjcx.230500124>

[基于注意力机制和提示学习联合训练的上下位关系识别研究](#)

Study on Hypernymy Recognition Based on Combined Training of Attention Mechanism and Prompt Learning

计算机科学, 2024, 51(6A): 230700226-5. <https://doi.org/10.11896/jsjcx.230700226>

[基于多任务联合训练的长文本多实体情感分析](#)

Long Text Multi-entity Sentiment Analysis Based on Multi-task Joint Training

计算机科学, 2024, 51(6): 309-316. <https://doi.org/10.11896/jsjcx.230400001>

对话场景下的情感引导问题生成模型

胥备^{1,2} 许鹏¹

1 南京邮电大学计算机学院、软件学院、网络空间安全学院 南京 210023

2 江苏大数据安全与智能处理重点实验室 南京 210023

摘要 人机对话系统已在多种智能服务场景中得到广泛应用。当前的人机对话系统可以感知对话者的情感,并根据上下文给出具备特定情感的响应。但是,具备特定情感的响应难以确保能够有效地引导人们产生特定的情感,例如,一个具备“高兴”情感的响应并不能保证人们产生高兴的情感。在一些场景中,人机对话系统需要引导用户达到某种特定的情感状态,以利于对话的持续开展或提升交互效率,如对话心理陪护或在线智能教学。当前的人机对话系统仅针对“积极/消极”等粗粒度情感引导进行了探索,难以应对细粒度情感引导任务。同时,针对对话的心理研究指出,“问题”会显著影响对话方情感的走向。基于上述背景,提出了一种对话场景下的情感引导问题生成模型。该模型基于GPT预训练模型,将需要引导对话方产生的情感作为情感知识引入模型的响应生成过程之中,同时引入了上下文情感感知机制和常识知识融合机制,并采用多任务学习的方法增强了模型的情感感知能力和对话响应生成能力。鉴于这是首次提出面向细粒度情感引导的问题生成任务,因此构建了情感引导数据集用于训练和实验,并且提出了基于提示学习的自动评价方法。最终,自动评价和人工评价的结果表明,所提模型能有效地生成问题,以引导对话方产生特定的情感。

关键词:情感引导;问题生成;情感对话;提示学习;多任务学习

中图分类号 TP181

Emotion Elicited Question Generation Model in Dialogue Scenarios

XU Bei^{1,2} and XU Peng¹

1 School of Computer Science, Nanjing University of Posts and Telecommunications, Nanjing 210023, China

2 Jiangsu Key Laboratory of Big Data Security & Intelligent Processing, Nanjing 210023, China

Abstract Human-machine dialog systems have been widely used in intelligent services. Existing human-machine dialog systems can perceive the interlocutor's emotional state and give a response with an appropriate emotion based on context. However, it is difficult to ensure that a response with a specific emotion can elicit the same emotion from people. For example, a response with a “joy” emotion does not guarantee that people will experience a “joy” emotion. In some scenarios, human-machine dialogue systems need to guide users to reach a specific emotional state to facilitate the continuous development of a conversation or improve interaction efficiency, such as dialogue psychological escort or online intelligent teaching. Current human-computer dialogue systems focus on coarse-grained emotion eliciting, such as “positive/negative”, and therefore are difficult to handle fine-grained emotion eliciting. On the other side, research on dialogue psychology indicates that “questions” in a conversation can significantly affect the emotions of interlocutors. Based on the above background, a question-generation model for emotional elicitation in dialogue scenarios is proposed. This model is based on the GPT pre-trained model and incorporates the knowledge of the emotion to be elicited into the response generation. The model also introduces a contextual emotional perception mechanism and a common sense knowledge fusion mechanism and uses multi-task learning to enhance the emotion perception ability and conversation response generation ability. Given that it is the first time to propose a question generation task for fine-grained emotion eliciting, an emotional eliciting dataset has been constructed for training and experiments. An automatic evaluation method based on prompt learning has been designed. Finally, automatic evaluation and human evaluation demonstrate that the proposed model can generate questions that can effectively elicit target emotions.

Keywords Emotion eliciting, Question generation, Emotional dialogue, Prompt learning, Multi-task learning

到稿日期:2023-10-07 返修日期:2024-03-10

基金项目:江苏省高校自然科学基金面上项目(21KJB520017)

This work was supported by the Natural Science Foundation of the Jiangsu Higher Education Institutions of China(21KJB520017).

通信作者:胥备(xubei@njupt.edu.cn)

1 引言

对话系统由图灵于1950年在“图灵测试”中首次提出^[1]。当前,智能对话系统已在心理陪护、在线教育 and 电子商务等领域广泛应用。情感是对话中蕴含的重要语义元素。人工智能的目的之一是使机器能够感知人类的情感并与人进行情感交互^[2]。以往的研究工作表明,通过主动感知并引导对话方的情感状态,能够极大地提升用户满意度,改善交互体验,减少对话的中断率^[3]。例如,带有情感引导的心理治疗聊天机器人问诊能够增加用户信任,进一步促进聊天的深度,提升治疗效果^[4]。

对话中的情感引导属于情感计算范畴,需要融合自然语言理解、情感识别和文本生成这3方面的研究。近年来,随着预训练语言模型的快速发展以及高质量人工标注数据集的涌现,这3方面的研究均取得巨大进展,出现了一批情感对话生成工作。Ghosh等^[5]首次将情感因素引入神经生成模型中。他们定义了一个能量项,通过上下文、情感强度和情感因素来生成响应。Zhou等^[6]首次在对话生成的工作中引入情感因素,提出了ECM(Emotional Chatting Machine)模型,使用序列到序列模型架构,通过情感类别向量建模抽象的情感特征表示,并引入内部情感记忆模块和外部情感词典分别捕获解码生成过程中的情感状态以及动态地选择合适的情感词,从而生成带有指定情感类别的响应。部分工作专注于共情对话生成,即不需要指定模型生成响应的情感,而是让模型根据上下文生成具备适合当下场景的情感的响应。例如,Sabour等^[7]通过引入常识知识库ATOMIC^[8]来深层次分析对话上下文,以生成带有合适情感并且知识丰富的响应;Zhou等^[9]提出的CASE系统可通过构建常识认知图和情感概念图,将用户的认知和情感在粗粒度和细粒度层面进行对齐,生成更具共情和信息丰富的回复。但上述研究聚焦于生成包含特定情感的响应,较难有效地进行特定情感的引导。

当前的情感引导研究聚焦于粗粒度的情感方向,如积极和消极。Lubis等^[10]构建了一个反映积极情感诱导策略的对话语料,在无需任何对话策略的情况下利用神经网络进行积极情感的激发;Wang等^[11]改进负对数损失,提出积极情感引导损失来鼓励模型在保证对话平稳的条件下往积极情感方向进行引导;Jiang等^[12]引入情感激发因子平衡反应生成过程中的情感,也利用该因子平衡积极和消极情感对生成响应的影响,进而达到更精确的情感控制;Gong等^[13]通过CVAE引入潜在变量来捕获对话中的变化,以提高对积极情绪的引导效果。针对多轮情感支持对话,Zhou等^[14]提出了一种基于混合专家和强化学习的模型SUPPORTER,可通过选择合适的专家和调整情感表达强度有效地引导用户从负面情绪转变为积极情绪。然而,上述研究专注于激发粗粒度的正向积极情感而没有关注如高兴、沮丧、难过、愤怒等细粒度情感,从而导致情感过于同质化,较难支撑相对复杂细致的情感引导场景。为了产生更好的引导作用,部分场景中需要更细粒度的情感引导。

另一方面,现有对话系统较少关注不同句子类型在对话中的辅助作用。心理学研究指出,“问题”是一种能够有效

引导情感的对话方式。例如在对话中,通过提出合适的“问题”,提问者不但被认为体现了倾听、理解等关系,还会增加人际间的好感^[15]。McEvoy等^[16]指出,在对话过程中通过“问题”展现出好奇心是建立有意义交流和平等理解关系的基础,有利于情感引导;Svikhnushina等^[17]同样关注到了“问题”在对话情感中的引导作用,构建数据集分析了问题意图和行为,以支撑“问题”对共情对话的情感影响。

本文提出了一种可以在开放域对话中进行细粒度情感引导的问题生成模型(Question Generation Model for Emotion Eliciting, QGM-EE),并标注了一个用于细粒度情感引导任务的数据集。该模型生成的响应以“问题”的形式反馈给对话人。该模型通过不同的情感表征融合和多任务学习使生成的问题具有悲伤、快乐、愤怒、恐惧4类情感引导功能,并且结合常识知识使问题具有更多的有效信息。通过基于提示学习的自动评估和人工评估,验证了模型所生成的问题能够在一定程度上有效地引导情感。

2 相关工作

2.1 对话系统

对话系统包括基于规则、检索和生成3类。

Weizenbaum^[18]在1966年提出基于规则的聊天机器人ELIZA,首先解析出关键词,然后使用带有关键词的模版进行填充生成响应。Colby等^[19]在ELIZA的基础上引入愤怒和恐惧的情感来应对不同的对话情景,并且是第一个通过简单“图灵测试”的对话系统。此后,越来越多的基于规则的聊天机器人被提出^[20]。

基于检索的聊天机器人通过信息匹配或检索技术从高质量语料库中选择最合适的语句作为响应,具有较强的系统可控性^[21]。现有基于检索的聊天机器人大多包括预检索和重排序两步。

基于生成的方法大多采取序列到序列架构,通过编码器给定的上下文得到对话语义信息,然后传给解码器生成响应。在基于生成的方法中,越来越多的研究者使用注意力机制控制对话系统生成融入某种特定因素的响应,如融入常识知识^[22-23]、个性化特征和情感因素等^[6,24]。

在对话中操控聊天机器人响应中蕴含的情感,可以使聊天机器人更加智能。Zhou等^[6]首次将情感因素引入大规模生成式的对话系统中。Li等^[25]在编码过程中利用情感关键词,并在解码过程中利用主题关键词使生成的带有情感的响应符合上下文。Huang等^[26]研究了3种引入情感因素来控制响应生成的方式。Ma等^[27]提出一个情感控制单元,并通过情感策略选择机制来解决生成响应过程中的情感偏移问题,即生成响应的情感与上下文情感不一致。上述研究主要关注生成包含特定情感的响应。

2.2 问题生成

问题生成旨在根据上下文和答案生成一段可以用该答案来回答的问题。早期的问题生成是问答系统的辅助任务之一,用来为问答系统生成额外的监督数据。此后问题生成被应用于辅助教学、意图阐明等应用中。传统问题生成基于一长段比较长的上下文,通过控制回答问题的答案跳数生成复杂

可控的问题^[28-29]。但传统对话领域的问题生成并没有得到太多关注。大部分研究仅将问题生成作为对话生成过程中的子任务。Wang等^[30]首次提出在对话中生成问题的重要性,并设计了软解码器和硬解码器在疑问词、话题词和普通词3种类别上产生概率分布,以生成更有意义的问题。Gao等^[31]关注到问答对话中问题的相关性,提出了一个端到端的神经网络模型建模代对齐和会话流,以生成高度对话性的问题。

3 数据集

由于当前没有可以直接应用于细粒度情感引导问题生成任务的数据集,我们利用现有对话数据集,构建了一个用于此任务的数据集 Emotion eliciting data (EE-Data)。

3.1 数据来源

本文使用了以下高质量多轮对话数据集。

1) DailyDialog^[32]。该数据集内容来源于英语学习网站中用于练习英语日常聊天的高质量对话。相比从推特、微博等社交媒体爬取的 Post-Reply Pair 数据, DailyDialog 涵盖的生活主题更加丰富,语法更加严谨。

2) EmpatheticDialog^[33]。该数据集内容为共情对话,其中每段对话设置了一个场景和两个角色 (Speaker 和 Listener)。两个角色根据给定的情感标签开展基于该场景的对话,即 Speaker 先带着指定情感描述其在当前情景下的处境,然后 Listener 给予共情响应。

3) Cornell Movie-Dialogs Corpus^[34]。该数据集为康奈尔大学从电影剧本中搜集整理的电影对话语料数据集,具有角色多和噪声小的优点。

数据集 EE-Data 的构建过程如下:

从 DailyDialog, EmpatheticDialog, Cornell Movie-Dialogs Corpus 这 3 个数据集的每段对话中检索“问句”作为 Question。每个 Question 的前一句作为上下文 Context, 后一句作为对 Question 的响应 Response, 得到一个三元组形式数据 $\langle Context, Question, Response \rangle$ 。为平衡数据集长度分布,我们删除了 Context 多于 50 个字、Question 多于 30 个字或者少于 3 个字的数据。

3.2 情感选择和数据标注

Ekman^[35]提出了人类最基础的 6 类情感,即愤怒、厌恶、恐惧、快乐、悲伤、惊喜。由于数据集分布不均,以及我们认为厌恶一定程度上接近愤怒,惊喜一定程度上接近快乐,因此剔除了在数据集中出现频率不高的厌恶和惊喜。最终选择悲伤、快乐、愤怒、恐惧代表人类的基础情感来进行情感引导。我们使用 Bert-Base^[36]和 RoBERTa-Base^[37]作为数据标注模型,并加载了 Huggingface 社区中准确率最高的模型权重进行微调。为了客观和准确,从数据集中每类情感随机挑出 1200 条数据进行人工标注,其中 1000 条作为微调数据,200 条作为测试数据。最终的结果如表 1 所列。

表 1 微调后的情感分类器效果

Model	Recall	Precision	Accuracy	F1-macro
Bert	0.895	0.897	0.895	0.895
RoBERTa	0.836	0.837	0.836	0.836

使用上述微调完的模型对构建的数据集中的 Context 和 Response 进行情感自动标注。为保证数据质量,仅在 Bert-Base^[36]和 Roberta-Base^[37]针对同一条数据给出相同情感分类结果时,才保留该条数据。最终从标注完的数据中为每个情感随机抽取 10000 条作为训练集,150 条作为验证集,150 条作为测试集,得到 40000 条训练数据、600 条测试和验证数据。

4 研究方法

4.1 任务定义

以往的情感对话生成为两阶段任务。任务给出了用户的上下文和情感,然后生成一个带有给定情感的响应。将面向情感引导的问题生成定义为一个三阶段任务,其数据格式可以表示为 $\langle Context, Question, Response \rangle$ 的三元组。Context 和 Response 来自说话者 A, Question 来自说话者 B, Response 所表达的情感可以被视为说话者 A 在听到说话者 B 说的 Question 后产生的情感。因此任务可被定义为:给定一段上下文 Context 以及 Response 的情感标签 Emotion-Response, 模型生成能够引导出 Emotion-Response 情感的 Question。

4.2 模型

基于自回归结构的 GPT 模型已经被证明在对话场景中具有突出的生成能力^[38-39],因此,本文模型 (QGM-EE) 设计了一个标准的 GPT2^[40] 自回归模块和一个情感预测模块,如图 1 所示。

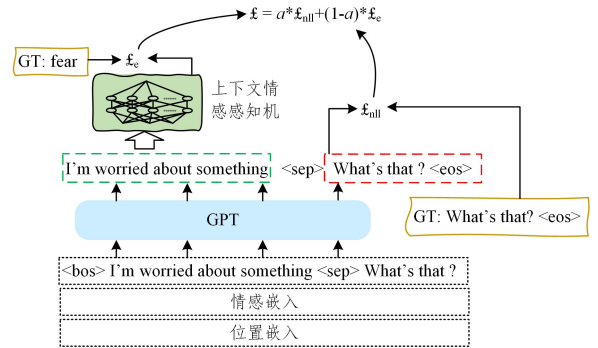


图 1 QGM-EE 模型结构图

Fig. 1 Architecture of QGM-EE model

本文模型以“上下文-响应”对,即 (X, Y) , 作为输入。 X 和 Y 是分别表示上下文和响应的序列,其中 $X = x_1, x_2, \dots, x_n$, $Y = y_1, y_2, \dots, y_m$ 。自回归语言模型将 X 和 Y 连接成一个句子作为输入。模型逐字生成 $\hat{Y} = \hat{y}_1, \hat{y}_2, \dots, \hat{y}_m$ 。除 (X, Y) 之外,模型还将 Response 中的情感 e 作为输入,并将 e 映射到一个低维向量,与词嵌入进行合并,那么任务的最终目标就是最小化 $\mathcal{L} = -\log P(Y|X, e)$ 。同时,我们为 X 添加了情感预测的辅助任务,即在模型生成 Y 的过程中,预测上下文 X 的情感类别,增强模型的情感感知能力,以帮助生成高质量的 Y 。

4.3 输入表示

假设一段对话三元组中的 Context 和 Question 被表示为 (C, Q) , 其中 $C = \{c_1, c_2, c_3, \dots, c_m\}$, $Q = \{q_1, q_2, q_3, \dots, q_n\}$ 。我们使用字节对编码算法 (Byte Pair Encoding) 对 C 和 Q 进行

标记,并将经过词标记的 C 和 Q 串联起来,然后在开头添加一个特殊的 $[bos]$ 标记作为开始信号。在 C 和 Q 之间插入一个 $[sep]$ 标记来区分不同的说话者,并在最后添加一个 $[eos]$ 标记。因此输入被表示为 $I = [bos, c_1, c_2, \dots, c_m, sep, q_1, q_2, \dots, q_n, eos]$ 。此外,我们为输入加入了可学习的位置编码。受 Huang 等^[22] 的启发,本文设计了一个与词嵌入相同的向量维度的矩阵,该矩阵可以在训练期间学习以表示每种情感。通过在解码时将情感嵌入注入 GPT,本文模型生成的问题可以激发出对话对方相应的情感。最终,定义了词嵌入编码 $E \in \mathbb{R}^{|V| \times d}$ 、可学习的情感词嵌入 $Emo \in \mathbb{R}^{|K| \times d}$,以及位置嵌入 $P \in \mathbb{R}^{1024 \times d}$ 。其中 $|V|$ 表示词表的大小, $|K|$ 表示情感的类别数, d 表示向量嵌入的维度。最终,模型的输入被标示为词嵌入、情感嵌入和位置嵌入之和,如式(1)所示:

$$Emb(I) = E([C, Q]) + P([C, Q]) + Emo(e) \quad (1)$$

4.4 上下文情感感知机

GPT 模型预训练时没有包含情感感知任务。考虑到情感信息较难学习,如果直接将情感嵌入结合到输入中,则很难将一个普通的生成模型转换为一个情感感知生成模型。在情感对话生成场景中,说话者的上下文情感信息对结果的准确性非常重要。因此,对话人当前的情感状态是决定生成的问题能否成功引导特定情感的关键因素。例如,当一个人处于开心和生气两种不同的情感状态时,使他们达到开心状态的响应很可能是不同的。同时,研究表明,同领域的多任务学习策略可以有效地提高其他任务的表现^[41-42]。

基于以上分析,本文提出上下文情感感知机来增强模型对句子情感的感知能力。我们使用三元组中 $Context$ 的最后一个词的隐藏层向量作为其语义表征。在经过一个全连接层后映射到情感类别数,模型输出 $Context$ 的情感概率,如式(2)~式(5)所示:

$$H^l = Trans(H^{l-1}) \quad (2)$$

$$H^0 = Emb(I) \quad (3)$$

$$H_c = Truncate(H^l) \quad (4)$$

$$P_e = Softmax(W_p H_c) \quad (5)$$

其中, $H^l \in \mathbb{R}^{len \times d}$ 表示第 l 层 Transformer 块的输出, $W_p \in \mathbb{R}^{d \times |s|}$ 表示全连接层的可学习映射参数, P_e 表示上下文 $Context$ 的概率分布。使用交叉熵损失来最小化预测概率和 $Context$ 的真实值之间的分布,如式(6)所示:

$$\mathcal{L}_e = -\log P_e(e^*) \quad (6)$$

4.5 训练目标

最终,模型逐词生成问题: $Q = q_1, q_2, \dots, q_n$, 如式(7)所示:

$$P(q_t | C; q_{<t}, e) = Softmax(W_v H_{q_{<t}}^l) \quad (7)$$

其中, $H_{q_{<t}}^l$ 表示在生成第 t 步时的隐藏层表示, $W_v \in \mathbb{R}^{d \times |V|}$ 是一个将隐藏层映射到词表大小的输出层可学习矩阵参数。在训练任务中,有两个学习目标:1)优化上下文情感识别;2)训练一个自回归语言模型,最小化负对数损失,如式(8)所示。

$$\mathcal{L}_{\text{nil}} = -\sum_{t=1}^T \log P(y_t | C; q_{<t}, e) \quad (8)$$

式(9)给出了最终结合两个任务后的损失函数。

$$\mathcal{L} = \alpha * \mathcal{L}_{\text{nil}} + (1 - \alpha) * \mathcal{L}_e \quad (9)$$

其中超参数 α 用于平衡两个训练任务,设置为 0.75。

5 实验

5.1 实验设置

实验使用 Pytorch 和 Huggingface 社区的开源权重初始化训练模型。本文选择了基础版的 GPT2 作为 QGM-EE 的主体结构,包含 127M 的参数,由 12 层的 Transformer Block 组成,有 50257 大小的词表,模型的中间隐藏层和上下文情感感知机的维度均为 768 维。训练时,设置学习率为 1×10^{-5} ,批大小为 16,轮次设置为 20,在一块 12 GB 显存 NVIDIA RTX 3060 GPU 上进行训练和测试。测试时,采用 Multinomial Sampling 的生成方式,最大生成长度限制为 100 个 token。

5.2 对比模型

由于缺少针对细粒度情感引导工作,因此本文构建了若干对比模型来验证 QGM-EE 的有效性。为保证 QGM-EE 和对比模型不受基础模型性能差异的影响,对比的模型均基于预训练的 GPT2。

1)GPT-INSERT。该模型^[26] 在指定情感的对话生成任务基础上,采用了在编码器端输入句子的前面和后面拼接情感词和在解码过程中添加情感词的情感表达方式。我们实现了该模型,并将情感词作为一个特殊的情感词 e 直接插入模型的输入中。输入的格式为 $[bos, c_1, \dots, c_m, sep, e, sep, q_1, \dots, q_n, eos]$ 。

2)GPT-INSERT-EP。为了探究 $Context$ 的情感感知对模型的影响,设计了基于 GPT-INSERT 的 GPT-INSERT-EP,添加了用来感知 $Context$ 情感的多任务学习架构以增强模型学习能力。EP 表示情感感知(Emotional perception)。

3)GPT-Embedding(GPT-EMB)。为探究情感向量的融合形式,设计了 GPT-EMB。该模型在解码的每一步,将情感词嵌入添加到输入中,以便模型的生成。GPT-EMB 和 QGM-EE 的主要区别为 GPT-EMB 没有多任务输出头模块。

5.3 自动评价

本文采用 Rouge 作为句子自动评估标准。相比普通的情感响应生成,情感引导生成属于三阶段的任务,较难对生成结果进行直接评价。如图 2 所示,GPT3.5 无法直接给出正确答案,因此本文采用 Prompt 策略进行情感预测准确性评估^[43]。

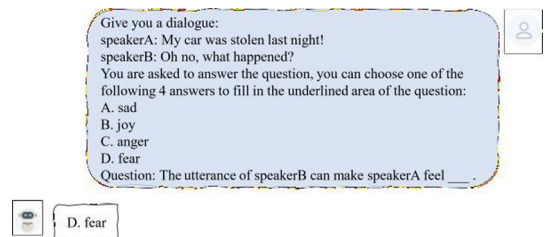


图 2 基于 GPT3.5 的情感引导评估

Fig. 2 Emotional elicited evaluation based on ChatGPT3.5

使用提示模板 “speaker A: [x]. speaker B: [y]. speaker A: It makes me so [mask].” 将 $Context$ 和模型生成的 $Question$ 结合成一句带有 Mask 的通顺句子,Mask 位置用于模型

预测 *Response* 的情感。本文利用语言模型的生成能力来推断生成的问题是否成功激发目标情感。如图 3 中的例子所示,将被模板包裹的句子 *X* 送入模型,模型随后给出概率最高的情感标签。其中 *C* 表示 *Context*, *Q* 表示 *Question*。

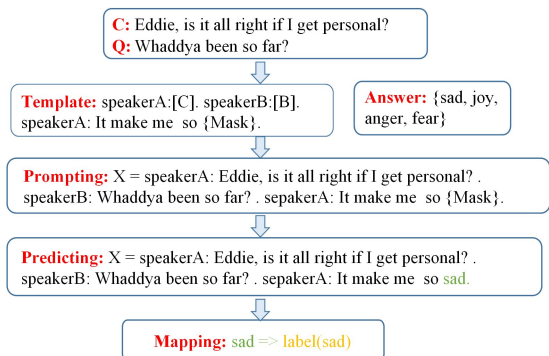


图 3 基于 Prompt 的自动评估

Fig. 3 Automated evaluation based on Prompt

表 4 列出了各模型的 Rouge 值和整体准确率。表 2—表 4 的统计数据表明 QGM-EE 在准确率和召回率上优于其他对比模型。相较于使用情感词拼接方式的 GPT-INSERT-EP, 本文模型在精确率和召回率上各有 3.4% 的提升; 相较于不使用上下文感知多任务学习的 GPT-EMB, 本文模型在精确率和召回率上分别有 0.6% 和 0.9% 的提升。这表明需要引导的情感信息的融入方式是影响最终模型效果的重要因素。此外, 感知上下文的多任务学习机制也能为模型带来一定的性能提升。结果表明, QGM-EE 将情感词向量与句子词向量相结合的情感融入方式以及上下文感知的多任务学习机制有助于生成具有情感引导作用的句子。

表 2 各情感类别上 4 个模型的精确率

Table 2 Precision of 4 models in each emotion category

Emotion	GPT-INSERT	GPT-INSERT-EP	GPT-EMB	QGM-EE
sad	0.514	0.567	0.621	0.594
joy	0.526	0.514	0.477	0.574
angry	0.333	0.328	0.368	0.357
fear	0.441	0.476	0.533	0.496
avg	0.454	0.471	0.499	0.505

表 3 各情感类别上 4 个模型的召回率

Table 3 Recall of 4 models in each emotion category

Emotion	GPT-INSERT	GPT-INSERT-EP	GPT-EMB	QGM-EE	Avg
sad	0.247	0.253	0.240	0.273	0.253
joy	0.533	0.507	0.613	0.567	0.555
angry	0.553	0.567	0.593	0.613	0.582
fear	0.373	0.400	0.380	0.413	0.392
avg	0.427	0.432	0.457	0.466	0.446

表 4 自动评估结果

Table 4 Results of automatic evaluation

模型	Rouge	准确率
GPT-INSERT	0.1526	0.4267
GPT-INSERT-EP	0.1280	0.4316
GPT-EMB	0.1456	0.4566
QGM-EE	0.1357	0.4667

根据 Rouge 结果, 具有上下文情感感知头和情感向量融合的模型比直接拼接情感词的模型性能略差。因此, 虽然上下文

情感感知多任务学习和情感词融合能改善模型的情感感知能力, 但同时也略微降低了模型的内容生成能力。此外, 根据召回率, 对于 sad, angry 和 fear 这 3 类情感, 带有上下文情感感知的 QGM-EE 的性能优于没有感知上下文情感的模型, 并且与拼接情感词相比, 情感词向量融合具有更好的效果。

图 4 给出了情感引导概率分布的占比。其中纵轴表示想要引导对话方达到的目标情感; 横轴表示生成的问题实际引导出的情感; 数字表示每种情况的比例, 每行相加为 1。从图中对角线看, 引导的目标情感为 joy 和 angry 时, 大部分结果能被成功引导, 其成功占比大于 sad 和 fear。同时, 考虑到 sad, angry 和 fear 都是较为负面的情感, QGM-EE 很少在需要引导正面情感(负面情感)时, 错误地引导出负面情感(正面情感)。

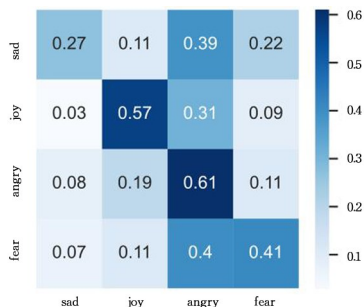


图 4 QGM-EE 的情感引导概率分布

Fig. 4 Emotion elicited probability distribution of QGM-EE

表 5 列出了从 *Context* 情感到目标情感的引导成功率。其中, 第一列为 *Context* 情感, 第一行为期望引导的目标情感。可以观察到, “从 sad 到 sad” 和 “从 sad 到 fear” 的引导成功率接近。在试图维持 *Context* 的情感, 即使三元组中 *Response* 的情感与 *Context* 的情感相同时, sad, joy 和 angry 的成功率较高, 而 fear 的成功率较低。这表明通过提出一个问题保持悲伤、高兴和生气相对容易, 保持恐惧相对较难。一个可能的原因是, fear 通常伴随着孤独, 当交流发生时则能缓解孤独感。同时, fear 的产生与环境密切相关, 通过问题形式的语句则难以维持产生 fear 的环境和氛围。表 5 同时表明, 与从 fear 中引导 angry 或 sad 相比, 从 fear 中引导 joy 相对容易。这可能来自于心理补偿效应^[44]。

表 5 从 *Context* 情感引导为目标情感的成功率

Table 5 Success rates eliciting from *Context* emotion to target emotion

目标情感	sad	joy	angry	fear
sad	0.633	0.375	0.688	0.000
joy	0.119	0.777	0.476	0.396
angry	0.157	0.206	0.852	0.260
fear	0.333	0.643	0.118	0.116

整体来看, 受复杂心理激发机制和情感主观感受影响, 且细粒度情感划分增加了引导难度, 引导成功率仅接近 50%。但考虑到 QGM-EE 很少在需要引导正向情感或负向情感时引导出相反情感, 该模型的整体性能优于传统的“正向/负向”情感激发模型。

5.4 人工评估

本文进行了人工评估以进一步验证 QGM-EE 的有效性。

我们随机选择每种引导情感模型生成的 400 条响应,并进行手动评分。征集了 5 名在校大学生进行评估,要求评估人员感受生成的响应是否可能导致相应的情感。如果可能,则标记为 1,否则标记为 0。为保证所评估情感互不干扰,每次评估后,评估人员需要休息 3~5min。

将“成功引导率”定义为标记 1 的响应的百分比,同时要求评估人员检查响应是否适合上下文的适当性,如果适合则标记为 1,否则标记为 0。

表 6 表明所提模型有 60%左右的平均引导成功率。另一方面,sad 情感的适当性较高但是成功引导率较低。因此,sad 情感相比其他 3 类情感较难被成功引导。fear 情感的适当性较低,但成功引导率较高,因此 fear 情感相对容易引导。

表 6 人工评估结果

Table 6 Results of human evaluation

	sad	joy	angry	fear	avg
成功率	0.525	0.625	0.600	0.625	0.593
适当性	0.850	0.875	0.675	0.675	0.775

5.5 常识知识融合

在上述 QGM-EE 生成的问题中,我们发现部分生成的问题长度偏短,含有的实体和知识不够丰富。例如,QGM-EE 会生成类似于“what?”或者“what’s happened?”这样的含信息量较少的反问回复。因此用常识知识对 *Context* 的话语进行常识推断,以生成具有高质量的带有常识知识的问题。本文使用 ATOMIC^[8] 作为常识知识库,该知识库包含一个基于日常事件进行知识推理的集合。对于一个事件,ATOMIC 能够推断出当事人在指定关系下的某种合理反应。ATOMIC 包含多类关系,如“反应”“想要怎么样”。例如,一个人如果说“我本来有机会挣很多钱,但手机却在关键时刻坏了”,通过“反应”的关系可以推断出他此刻应该感到沮丧和后悔,通过“想要怎么样”的关系,可以推断出他想买一个新的更耐用的手机。

我们采用预训练 BART 在 ATOMIC-2020 数据集上微调得到的 COMET-BART 模型。具体地,使用了表示“想要怎么样”关系(xWant)、“当前的事件对其产生了什么影响”关系(xEffect)、“对事件可能的反应”关系(xReact),以及“需要什么以应对将要发生的事”关系(xNeed)。将 *Context* 作为事件,并将其与关系拼接,如“my car was broken, <xWant>”,使用基于该知识库预训练的 COMET-BART 可以生成相应关系下的常识知识,如图 5 所示。

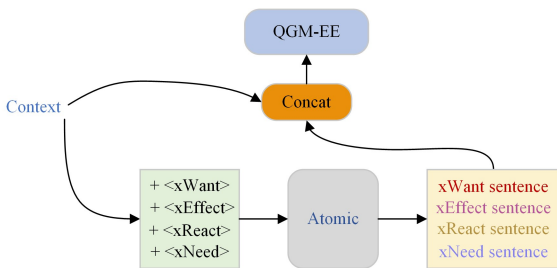


图 5 融合常识知识的 QGM-EE 结构图

Fig. 5 Architecture of QGM-EE combining common sense knowledge

情感以及信息量更加丰富。

表 7 加入常识推断后的例子

Table 7 Examples after adding common sense inference

Context	When my pet died,I felt like I lost my family member, my best friend
Ground truth	I'm sorry to hear that. What happens?
Question-CK	Oh no, what's happened?
Question+CK	I'm sorry to hear that. How did your family die?

表 8 的统计数据表明,加入了常识知识后(QGM-EE+Commonsense Knowledge,简称为 QGM-EE+CK),模型生成的问题在内容的信息丰富程度上有一定提升,但整体情感引导准确率略有下降。因此,我们在引入常识知识后对模型的不同情感引导能力进行了分别实验,结果如表 9 所列。其中,第一行为引导前的 *Context* 情感;第一列为期望的引导后情感。每个单元格记录了 QGM-EE+CK 的准确率减去 QGM-EE 的准确率。

表 8 生成的问题句子平均长度以及情感引导准确率

Table 8 Average length of the generated questions and the accuracy of emotion eliciting

	INSERT	INSERT-EP	EMB	QGM-EE	QGM-EE+CK
句子平均长度	6.030	6.003	6.223	6.125	7.450
准确率	0.4267	0.4316	0.4566	0.4667	0.4580

表 9 表明,引入常识知识后,从 sad 引导 sad 和从 joy 引导 joy 时,成功率会有一定上升,主要原因为其 *Context* 情感和目標情感一致。但是因为常识推断总是会默认与 *Context* 情感保持一致,对于引导非 *Context* 情感的其他情感时,可能会有反向结果,导致情感引导准确率略有下降。因此,在进行与 *Context* 情感不一致的引导任务时,增加常识知识可能会影响部分情感引导的效果。

表 9 添加推理知识前后引导成功率比较

Table 9 Comparison of elicited success rates before and after adding inference knowledge

	sad	joy	anger	fear
sad	0.034	-0.250	-0.313	0.111
joy	-0.024	0.074	-0.032	-0.021
anger	-0.020	-0.144	-0.056	0.000
fear	-0.185	-0.429	-0.059	0.000

结束语 本文提出了一个能够生成带有细粒度情感引导作用问题的模型,以实现悲伤、快乐、愤怒和恐惧 4 类情感的引导功能。所提模型使用上下文情感感知的多任务学习策略来提高情感引导的效率,并使用基于提示的策略进行自动评估。作为首个细粒度情感引导工作,我们构建了一个三元组<*Context*,*Question*,*Response*>形式的情感引导对话数据集。研究表明,QGM-EE 由于在解码过程中将情感词向量嵌入与词嵌入融合,以及使用和任务相关的多任务学习来提高情感感知能力,能够生成具有良好情感引导能力的问题。

同时,我们发现不同的情感在被引导时具有不同的特征。例如,从 fear 中引导 joy 比从 fear 中引导 angry 或 sad 更容易,该特征符合心理补偿效应;通过问题较难使对话方保持 fear 情感,说明对话中的 fear 情感具备快速消退的特点。

如表 7 中的示例所示,加入知识推理后,模型生成的答案

本文进一步使用常识知识来辅助模型生成。实验表明,加入常识知识推理能有效增加生成问题的信息量,以及可以提高在进行 joy 和 sad 的上下文情感一致的情感引导时的准确率。但在进行与上下文情感不一致的引导任务时,增加常识知识可能会轻微影响情感引导的效果。

参 考 文 献

- [1] TURNING A M. Computing machinery and intelligence [M]. Springer Netherlands,2009;23-65.
- [2] PICARD R W. Affective Computing [J]. User Modeling and User-Adapted Interaction,2002,12(1):85-89.
- [3] CHRISTENSEN H, GRIFFITHS K M, KORTEN A E, et al. A comparison of changes in anxiety and depression symptoms of spontaneous users and trial participants of a cognitive behavior therapy website[J]. Journal of Medical Internet Research,2004,6(4):e46-e46.
- [4] PRENDINGER H, MORI J, ISHIZUKA M. Using human physiology to evaluate subtle expressivity of a virtual quizmaster in a mathematical game[J]. International journal of human-computer studies,2005,62(2):231-245.
- [5] GHOSH S, CHOLLET M, LAKSANA E, et al. Affect-LM: A Neural Language Model for Customizable Affective Text Generation[C]//Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). 2017:634-642.
- [6] ZHOU H, HUANG M L, ZHANG T, et al. Emotional chatting machine: emotional conversation generation with internal and external memory[C]//Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence and Thirtieth Innovative Applications of Artificial Intelligence Conference and Eighth AAAI Symposium on Educational Advances in Artificial Intelligence. 2018:730-738.
- [7] SABOUR S., ZHENG C J, HUANG M L. Cem: Commonsense-aware empathetic response generation[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2022, 36(10): 11229-11237.
- [8] HWANG J D, BHAGAVATULA C, LEBRAS R, et al. Comet-atomic 2020: on symbolic and neural commonsense knowledge graphs[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2021, 35(7):6384-6392.
- [9] ZHOU J, ZHENG C, WANG B, et al. CASE: Aligning Coarse-to-Fine Cognition and Affection for Empathetic Response Generation[C]//Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). 2023:8223-8237.
- [10] LUBIS N, SAKTI S, YOSHINO K, et al. Eliciting positive emotion through affect-sensitive dialogue response generation: a neural network approach[C]//Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence and Thirtieth Innovative Applications of Artificial Intelligence Conference and Eighth AAAI Symposium on Educational Advances in Artificial Intelligence. 2018:5293-5300.
- [11] WANG S, XU X, WU W, et al. Towards Multi-Turn Empathetic Dialogs with Positive Emotion Elicitation [J]. arXiv: 2204.10509,2022.
- [12] JIANG H, ZHU Y, ZHANG X, et al. Emotion Eliciting Machine: Emotion Eliciting Conversation Generation based on Dual Generator[J]. arXiv:2105.08251,2021.
- [13] GONG Z, MIN Q, ZHANG Y. Eliciting Rich Positive Emotions in Dialogue Generation[C]//Proceedings of the First Workshop on Social Influence in Conversations (SICon 2023). 2023:1-8.
- [14] ZHOU J F, CHEN Z, WANG B, et al. Facilitating Multi-turn Emotional Support Conversation with Positive Emotion Elicitation: A Reinforcement Learning Approach[C]//Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). 2023:1714-1729.
- [15] HUANG K, YEOMANS M, BROOKS A W, et al. It doesn't hurt to ask: Question-asking increases liking[J]. Journal of Personality and Social Psychology,2017,113(3):430-452.
- [16] MCEVOY P, PLANT R. Dementia care: using empathic curiosity to establish the common ground that is necessary for meaningful communication [J]. Journal of Psychiatric and Mental Health Nursing,2014,21(6):477-482.
- [17] SVIKHNUSHINA E, VOINEA I, WELIVITA A, et al. A taxonomy of empathetic questions in social dialogs[C]//Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). 2022:2952-2973.
- [18] WEIZENBAUM J. ELIZA—a computer program for the study of natural language communication between man and machine [J]. Communications of the ACM,1966,9(1):36-45.
- [19] COLBY K M, WEBER S, HILF F D. Artificial paranoia[J]. Artificial Intelligence,1971,2(1):1-25.
- [20] THORAT S A, JADHAV V. A review on implementation issues of rule-based chatbot systems[C]//Proceedings of the International Conference on Innovative Computing & Communications (ICICC). 2020.
- [21] YAN R, LI J, YU Z. Deep learning for dialogue systems: Chit-chat and beyond[J]. Foundations and Trends © in Information Retrieval,2022,15(5):417-589.
- [22] LIU S, CHEN H, REN Z, et al. Knowledge diffusion for neural dialogue generation[C]//Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). 2018:1489-1498.
- [23] ZHOU H, YOUNG T, HUANG M L, et al. Commonsense knowledge aware conversation generation with graph attention [C]//IJCAI. 2018:4623-4629.
- [24] SONG H, ZHANG W N, HU J, et al. Generating persona consistent dialogues by exploiting natural language inference[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2020:8878-8885.
- [25] LI J Y, SUN X. A Syntactically Constrained Bidirectional-Asynchronous Approach for Emotional Conversation Generation [C]//Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. 2018:678-683.
- [26] HUANG C Y, ZAIANE O R, TRABELSI A, et al. Automatic

- dialogue generation with expressed emotions[C]// Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics; Human Language Technologies, Volume 2(Short Papers). 2018;49-54.
- [27] MA Z Q, YANG R, DU B, et al. A control unit for emotional conversation generation [J]. IEEE Access, 2020, 8: 43168-43176.
- [28] GAO Y, BING L, CHEN W, et al. Difficulty controllable generation of reading comprehension questions[J]. arXiv:1807.03586, 2018.
- [29] FEI Z C, ZHANG Q, GUI T, et al. CQG: A simple and effective controlled generation framework for multi-hop question generation[C]// Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). 2022;6896-6906.
- [30] WANG Y S, LIU C, HUANG M L, et al. Learning to Ask Questions in Open-domain Conversational Systems with Typed Decoders[C]// Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). 2018;2193-2203.
- [31] GAO Y F, LI P, KING I, et al. Interconnected Question Generation with Coreference Alignment and Conversation Flow Modeling[C]// Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. 2019;4853-4862.
- [32] LI Y R, SU H, SHEN X, et al. DailyDialog: A Manually Labeled Multi-turn Dialogue Dataset [C] // Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 1: Long Papers). 2017;986-995.
- [33] RASHKIN H, SMITH E M, LI M, et al. Towards Empathetic Open-domain Conversation Models: A New Benchmark and Dataset[C]// Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. 2019;5370-5381.
- [34] DANESCU-NICULESCU-MIZIL C, LEE L. Chameleons in Imagined Conversations: A New Approach to Understanding Coordination of Linguistic Style in Dialogs[C]// Proceedings of the 2nd Workshop on Cognitive Modeling and Computational Linguistics. 2011;76-87.
- [35] EKMAN P. Emotion in the human face: guidelines for research and an integration of findings[M]. Pergamon Press, 1972.
- [36] DEVLIN J, CHANG M W, LEE K, et al. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding [C]// Proceedings of NAACL-HLT. 2019;4171-4186.
- [37] LIU Y, OTT M, GOYAL N, et al. Roberta: A robustly optimized bert pretraining approach[J]. arXiv:1907.11692, 2019.
- [38] ZHANG Y Z, SUN S, GALLEY M, et al. DIALOGPT: Large-Scale Generative Pre-training for Conversational. Response Generation[C]// Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations. 2020;270-278.
- [39] WANG Y D, KE P, ZHENG Y, et al. A large-scale chinese short-text conversation dataset[C]// Natural Language Processing and Chinese Computing; 9th CCF International Conference. Springer International Publishing, 2020;91-103.
- [40] RADFORD A, WU J, CHILD R, et al. Language models are unsupervised multitask learners[J]. OpenAI blog, 2019, 1(8):9.
- [41] LIU X D, HE P, CHEN W, et al. Multi-Task Deep Neural Networks for Natural Language Understanding[C]// Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. 2019;4487-4496.
- [42] XU W R, GU X, CHEN G. Generating emotional controllable response based on multi-task and dual attention framework[J]. IEEE Access, 2019, 7;93734-93741.
- [43] LIU P F, YUAN W Z, FU J L, et al. Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing [J]. ACM Computing Surveys, 2023, 55(9): 1-35.
- [44] BÄCKMAN L, DIXON R A. Psychological compensation: a theoretical framework[J]. Psychological Bulletin, 1992, 112(2): 259-283.



XU Bei, born in 1986, Ph.D, associate professor, is a member of CCF (No. P1014M). His main research interests include affective computing and natural language processing.

(责任编辑:何杨)