

隐变量与情绪原因增强的情感支持对话模型

陶然, 汪红松, 郭志德

(华南师范大学人工智能学院, 广东 佛山 528225)

摘要: 情感支持对话系统在客服对话和自动心理治疗等应用领域具有关键价值。然而, 现有的情感支持对话系统通常忽视了2个重要问题: ①它们一般依赖于简单的情绪标签来描述用户状态, 较少捕捉用户的情绪原因; ②它们通常专注于优化策略选择, 可能忽略了响应的相关性和多样性问题。因此, 提出一个带有隐变量的情绪原因增强的模型。首先, 引入情绪知识补充情绪相关信息, 以建立用户的细粒度情绪状态, 为推理情绪原因提供线索; 其次, 利用连续隐变量描述输入数据的潜在特征, 并使用记忆模式让解码器直接关注这些信息, 从而提高响应的相关性和多样性。实验结果表明, 模型在自动评估和人工评估中均超越了现有基线。在策略预测准确度指标ACC、响应相关性指标B-2和多样性指标D-2上都取得了明显提升, 进一步验证了模型的有效性。

关键词: 情感支持对话; 隐变量; 情绪原因; 常识知识; 记忆模式

DOI: 10.11907/rj.242038

扫描二维码阅读全文:

中图分类号: TP391.1

文献标识码: A

文章编号: 1672-7800(2026)003-0078-08



An Emotion Cause Enhanced Model with Latent Variable for Emotional Support Conversation

TAO Ran, WANG Hongsong, GUO Zhide

(School of Artificial Intelligence, South China Normal University, Foshan 528225, China)

Abstract: Emotional Support Conversation (ESC) systems are crucial in many applications, such as customer service interactions and automated psychological therapy. However, existing ESC systems typically overlook two important issues: (a) they generally rely on simple emotion labels to describe user states, capturing the reasons behind users' emotions less frequently; (b) they often focus on optimizing strategy selection, potentially neglecting the relevance and diversity of responses. Therefore, this paper proposes an Emotion Cause enhanced Model With Latent Variable (ECLV). First, the paper introduces commonsense knowledge to supplement emotion-related information, establishing a fine-grained emotional state of the user to provide clues for inferring the emotion causes. Subsequently, continuous latent variables are used to describe the latent features of the input data, and a memory scheme is employed to allow the decoder to directly focus on this information, thereby improving response relevance and diversity. The model surpasses existing baselines in both automatic and human evaluations, with significant improvements in strategy prediction accuracy (ACC), response relevance (B-2), and diversity (D-2) metrics, validating the model's effectiveness.

Key Words: emotional support conversation; latent variable; emotion cause; commonsense knowledge; memory schema

0 引言

情感支持对话 (Emotional Support Conversation, ESC) 系统的目标不仅是安慰用户并表达共情, 还需要通过对话来缓解用户的情绪困扰, 帮助用户识别问题和解决困难。在现代社会中, 许多用户面临情绪困扰问题, 但家人和朋

友的帮助作用有限, 而专业心理咨询又往往费用较高。因此, 利用聊天机器人自动缓解用户情绪问题具有很好的发展前景, 对 ESC 系统的需求也在不断增加。为了改进情感支持对话, 本文提出了一种带有隐变量的情绪原因增强机制的情感支持对话模型 (Emotion Cause enhanced Model with Latent Variable, ECLV)。该模型成功提高了 ESC 系统的策略预测准确度, 并增强了对话响应的相关性和多样

收稿日期: 2025-01-16

作者简介: 陶然 (2000-), 男, 华南师范大学人工智能学院硕士研究生, 研究方向为自然语言处理; 汪红松 (1975-), 男, 博士, 华南师范大学人工智能学院讲师、硕士生导师, 研究方向为自然语言处理; 郭志德 (2000-), 男, 华南师范大学人工智能学院硕士研究生, 研究方向为自然语言处理。本文通讯作者: 陶然。

性。图1是一个情感支持对话示例,展示了模型的效果。其中,蓝色文字是使用的支持策略,黄色文字是情绪原因及其解决办法。

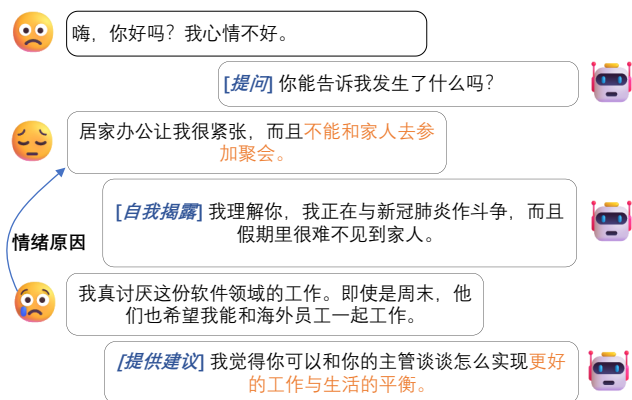


Fig. 1 Example of emotional support conversation

图1 情感支持对话示例

本文的主要贡献如下:①提出了ECLV模型,该模型能够识别情绪原因并利用连续隐变量生成ESC任务的响应;②为了更准确地识别情绪原因,使用常识知识作为因果线索,同时采用记忆模式技术使解码器充分关注隐变量中的信息;③通过与其他先进模型的严格比较,实验结果展示了本文模型的优越性。

1 相关工作

1.1 情感支持对话

近年来,情感^[1-2]和共情^[3-4]对话生成领域受到了广泛关注。然而,这些领域的研究仍不足以提供完整的情感支持。情感支持对话提出了更为先进和复杂的功能要求,不仅需要表达情绪及同情心,还需要通过适当的支持策略来缓解用户的负面情绪,同时深入了解问题并提供帮助。近年来,研究者们对ESC任务进行了多项研究,例如,Liu等^[5]提出了一个具有规范对话流程和支持策略的ESC框架。同时,还引入ESConv数据集,一个大规模的众包数据集,包含了求助者和支持者之间的多轮对话,并标注各种注释以便理解情感支持。此外,在ESConv数据集上训练了大型预训练对话模型,如Blenderbot^[6]和DialogPT^[7],使其能够学习情感支持过程和策略,并生成情感支持回答。

基于上述研究,在改进对上下文的理解和选择适当的支持策略上,Xu等^[8]挖掘了先验知识在范例和策略序列方面的潜力,并利用隐变量来建模一对多的策略关系。Peng等^[9]利用用户反馈鼓励或惩罚策略,旨在改善策略选择。Deng等^[10]提出了一个知识增强的混合主动性框架(knowledge-enhanced mixed-initiative framework, KEMI)用于ESC,从大规模的心理知识图谱中检索实际案例知识。尽管这些研究在识别用户情绪和优化策略选择方面取得了进步,但忽略了2个重要问题:①依赖对话级别的简单

情绪标签来表示用户状态。考虑到人类情绪的复杂性,必须考虑情绪的潜在原因,以深入地理解用户状态并提供有针对性的情感支持。②尽管在策略选择方面取得了改进,但很少关注响应的相关性和多样性。在ESC任务中,不仅策略选择很重要,响应的相关性和多样性也同样重要,忽视上述的2个问题会损害模型的情感支持能力。

为了解决上述问题,本文提出一种ECLV模型以改进情感支持对话。高质量的ESC需要准确识别用户产生消极情绪的原因,并采用适当的支持策略来解决相应问题。此外,处于消极情绪状态的用户对对话内容可能非常敏感。因此,ESC响应必须与上下文高度相关且具有良好的多样性。为此,ECLV通过识别用户的情绪原因来捕捉其细粒度情绪状态,通过引入隐变量优化了整个响应生成过程,同时提升了响应的相关性和多样性。具体而言,模型首先利用常识知识生成模型COMET为给定的话语生成各种与情绪相关的知识,并将其作为因果线索,弥补了用户的情绪语句和其他语句之间的推理差距,从而得到准确的情绪原因^[11]。随后,引入一个知识引导的条件变分自编码器来预测隐变量的概率分布,利用连续隐变量描述输入数据的潜在特征,并使用记忆模式让解码器直接关注这些信息,从而提高响应的相关性和多样性^[12]。

1.2 因果情感蕴含

近年来,情感推理任务(如识别对话中情感背后的原因)已经得到了一定的研究关注。Poria等^[13]提出了在对话中识别情绪原因的任务,并定义了2个子任务:因果跨度提取和因果情感蕴含。其中,因果情感蕴含旨在识别对话中导致目标话语的非中性情感的因果话语。然而,先前的研究在深入理解对话上下文和精准推断情绪成因方面存在局限性,因此研究者开始引入外部知识来辅助情绪成因推理。例如,Zhao等^[14]提出了知识桥接的因果交互网络,利用常识知识捕捉上下文中的深层次语义关系,获取因果线索以准确地推理情绪原因。COMET是一个基于transformer的用于自动构建常识性知识库的模型,本文利用COMET引入常识知识,让其作为因果线索弥补用户情感子句和上下文中其他语句之间的推理差距。

1.3 对话系统中的隐变量

近期对话系统工作中,隐变量被广泛用于提高响应的可控性和多样性,并解决对话系统中的一对多问题。例如,Bao等^[15]将离散隐变量引入预训练对话模型,提出了PLATO。其通过引入一个K方式($K=20$)的分类隐变量,在多个下游响应生成任务中表现出显著的增益。除了离散隐变量外,连续隐变量也常用于增强大规模预训练对话模型。例如,Chen等^[16]提出了一种用于对话生成的预训练隐变量编码器-解码器模型,在响应的相关性和多样性方面均优于先前的最先进方法。Tu等^[17]和Xu等^[8]都在ESC中利用隐变量来预测策略的概率分布,获得混合策略表示,以建模上下文和策略的一对多映射关系。本文引入

连续隐变量来描述输入数据的丰富特征,并使用记忆模式使解码器能够直接关注隐变量中输入数据的信息。这种创新方法旨在提高 ESC 中支持性响应的相关性和多样性。

2 模型

本文设计的模型 ECLV 包括 4 个主要组件:①情绪原因提取器。使用 COMET 生成与用户情绪相关的常识知

识,将其作为推理线索,与上下文结合得到准确的情绪原因。②多源编码器。使用基于 Transformer 的多层编码器来编码来自多个源的输入。③隐变量模块。将编码的输入传递给先验网络和识别网络,以分别获得隐变量的先验分布和后验分布。④记忆模式解码器。将隐变量和编码器输出一起输入到解码器,并将隐变量投影到额外的记忆向量中,让解码器直接关注隐变量中的信息,最后依次预测响应中的词。具体如图 2 所示。

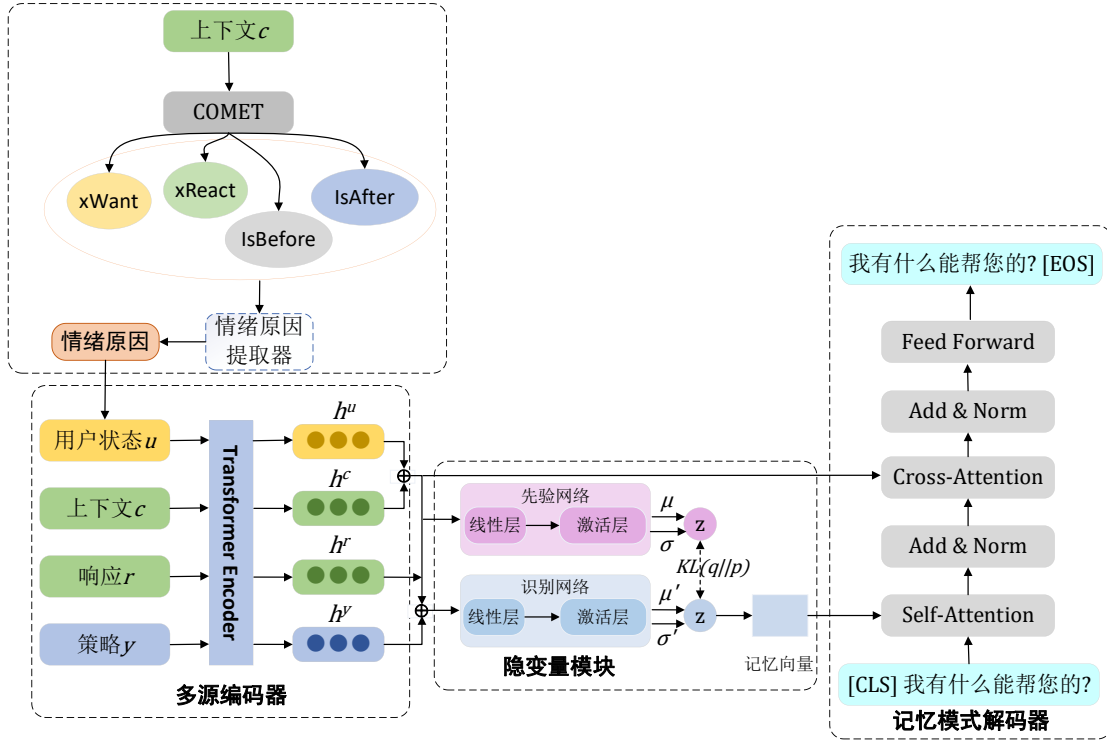


Fig. 2 Overall architecture of the ECLV model

图 2 ECLV 模型的总体架构

2.1 任务定义

假设第 t 轮对话中来自用户(求助者)和系统(支持者)的话语分别为 (a_t, b_t) , 那么给定对话上下文 $c = a_1, b_1, a_2, b_2, \dots, a_t$, 情感支持响应语句就是 b_t 。其中, 每个话语由几个词组成 $a_i = w_1^i, w_2^i, \dots, w_M^i$, 每个系统话语 b_i 都使用一个支持策略 y_i 。本文将最后使用的支持策略表示为 y 。用户境况 s 则大致概括了用户遇到的问题 and 情绪状态, 并简单地描述用户当前状况。本文的任务流程如下: 首先, 从对话上下文 c 中识别出 a_t 中负面情绪的原因 ec , 与境况 s 一起作为用户状态 u ; 然后, 利用上下文 c 和用户状态 u 得到隐变量 z ; 最后, 通过隐变量生成合适的情感支持响应语句 b_t 。

2.2 情绪原因提取器

识别用户负面情绪的原因对于深入理解其情绪状态并制定有效解决方案至关重要, 这是提供情感支持的核心环节。然而, 现有的方法往往难以把握上下文的复杂性, 在推理情绪原因方面准确度不高。为了解决这个挑战, 本文使用 ATOMIC-2020^[18] 作为常识知识库, 其包含了广泛

的日常推理知识, 为情绪原因提供了有价值的因果线索。后续将在 ATOMIC-2020 上训练 COMET 来增强这一点, 使生成的常识知识能够与对话上下文紧密对齐。

本文使用的常识知识有 $xWant$ 、 $xReact$ 、 $IsBefore$ 和 $IsAfter$ 4 种。其中, $xWant$ 表示事件主体 x 在事件发生后可能想做什么; $xReact$ 表示事件主体 x 在事件发生后的情绪感受; $IsBefore$ 和 $IsAfter$ 则表示事件发生之前或之后的事件。这些知识能够帮助模型理解用户的目的、感受以及事件的时间顺序, 有助于获取逻辑清晰且内容丰富的原因信息。随后, 上下文及生成的常识知识被输入到一个现成的情绪原因检测器^[14]中, 该检测器在情绪原因的基准数据集 RECCON-DD^[13]上进行了训练, 以识别用户话语中的情绪原因 ec 。

2.3 多源编码器

本文模型使用基于 Transformer^[19] 的多层编码器来编码来自多个源的输入。首先, 基于 c 的定义, 进一步对其进行表示, 如公式(1)所示:

$$c = (\text{CLS}, a_1, \text{SEP}, b_1, \text{SEP}, \dots, a_t) \quad (1)$$

其中,CLS是起始标记,SEP是两个话语之间的分隔标记, i 代表当前的对话轮次。

随后,将用户状态 u (包括用户境况 s 和情绪原因 ec)与上下文 c 进行连接,并输入到编码器中获得输入的隐藏表示,如公式(2)所示:

$$H_c = \text{Enc}(s, ec, \text{CLS}, a_1, \text{SEP}, b_1, \text{SEP}, \dots, a_i) \quad (2)$$

其中,Enc是Transformer编码器, $H_c \in \mathbb{R}^{l \times d_h}$ 是上下文的隐藏状态, l 表示上下文中的标记数量, d_h 表示隐藏状态的大小。为了简化整段上下文的表示,使用序列的第一个标记(即CLS标记)的隐藏状态作为上下文表示,即 $h^c = H_c^0$ 。

类似地,将其他输入也馈送到编码器中,并获得它们的隐藏表示,如公式(3)-(4)所示:

$$H_r = \text{Enc}(\text{CLS}, b_i, \text{SEP}) \quad (3)$$

$$H_y = \text{Enc}(\text{CLS}, y, \text{SEP}) \quad (4)$$

其中, H_r 和 H_y 分别是回答语句和支持策略的隐藏状态,也使用CLS标记的隐藏状态作为最终表示。

此外,与传统的Transformer编码器相比,本文模型加入了更丰富的位置嵌入,以提供位置信息,同时,捕捉更深层次的上下文语义关系。参考Chen等^[16]的研究发现,绝对位置嵌入的效果优于相对位置嵌入。为此,本文在模型中新增加了2种绝对位置嵌入方式:轮次绝对嵌入和角色绝对嵌入。最终,每个标记的输入嵌入由其标记嵌入与对应的轮次、角色嵌入相加得到。

2.4 隐变量模块

以往的情感支持对话系统在进行对话生成时,直接将编码器的输出馈送到解码器中,用于生成回答。这样的做法虽然简单高效,但是可能难以充分利用丰富的输入信息,特别是复杂的上下文中蕴含的信息。在变分自编码器(Variational Autoencoder, VAE)和条件变分自编码器(Conditional Variational Autoencoder, CVAE)中,隐变量用来描述输入数据的潜在特征或抽象表示,引入隐变量能够表示丰富的输入信息。

引入隐变量模块使模型能够进行分层生成,具体实现如下:①隐变量采样阶段:将编码器输出的输入表示输入先验网络 $p_\theta(z|u, c)$,采样得到隐变量 z ;②回答生成阶段:将隐变量 z 与编码器输出共同输入解码器,通过 $p(r|c, u, z)$ 生成回答 r 。

根据文献[16]的研究,在引入隐变量的生成模型中,隐变量 $z \sim p_\theta(z|u, c)$ 负责捕捉输出语句的高层语义特征,自回归解码过程则负责生成低层句法结构和词汇细节。这种分层生成机制通过显式建模不同层级的语言特征,使模型能够更有效地学习对话上下文的深层次语义,并提高生成回答的语义相关性和表达多样性。

对于隐变量的建模,使用原始的条件变分自编码器可能不适合需要多种知识的ESC,因此,本文使用知识引导的条件变分自编码器来建模隐变量,使其能够学习输入中

的知识^[12]。该框架包含2个核心组件:先验网络用于生成隐变量的先验分布;识别网络则负责推断其后验分布。先验网络和识别网络使用相同的结构,包括一个线性层和一个激活层,将多源编码器中获得的各种编码表示作为输入,转换为先验或后验分布的均值和方差。其中,线性层用于将输入的维度扩展为原本的两倍,因为最后需要将其分割为均值和方差两部分。同时,激活层用于为输入增加非线性变换,以便网络能够更好地学习和表达复杂的输入信息。

本文将多源编码器输出的用户状态 u 和上下文 c 拼接后输入先验网络。假设隐变量 z 在各维度上都遵循高斯分布,那么先验网络的计算过程如公式(5)所示:

$$p_\theta(z|u, c) \sim N(\mu, \sigma^2 \mathbf{I}) \quad (5)$$

其中, θ 是先验网络的参数, $N(\mu, \sigma^2 \mathbf{I})$ 是一个均值为 μ 和方差为 σ 的高斯分布。

为逼近隐变量的真实后验分布,识别网络需额外引入响应 r 和支持策略 y 作为输入。识别网络的计算过程如公式(6)所示:

$$q_\phi(z|r, y, u, c) \sim N(\mu', \sigma'^2 \mathbf{I}) \quad (6)$$

其中, ϕ 是识别网络的参数, $N(\mu', \sigma'^2 \mathbf{I})$ 是一个均值为 μ' 和方差为 σ' 的高斯分布。

先验网络和识别网络的参数随着模型的训练不断更新,在这个过程中先验网络和识别网络可以更好地将输入映射为分布参数,从而得到更准确的隐变量。

通过上述步骤优化先验网络和识别网络后,模型就可以从网络输出的分布中采样隐变量 z 。由于直接采样操作不可导,本文采用重参数化技术实现梯度传播:首先从标准正态分布 $N(0, 1)$ 中采样一个参数 ϵ ,再通过先验(识别)网络输出的均值 μ 和方差 σ 计算隐变量 $z = \mu + \sigma \times \epsilon$ 。这种方法使采样过程与可训练参数 μ 和 σ 建立可导关系,模型可以通过反向梯度传播进行优化。

在模型训练时,解码器使用识别网络采样的隐变量 z 来生成响应 r ,同时使用多层感知机(Multilayer Perceptrons, MLP)预测 $y' = \text{MLP}_y(z, c, u)$,在生成回答语句 r 的基础上额外生成策略 y ,可以使隐变量 z 捕获到有关策略 y 的信息。而在推理阶段,则使用先验网络采样的隐变量 z 生成响应 r 。

2.5 记忆模式解码器

与编码器类似,本文使用基于Transformer^[19]的多层解码器进行响应生成。由于隐变量从多个输入中学习了大量的信息,这些信息需要在解码器中得到充分利用,因此,本文研究使用了Li等^[20]提出的记忆模式技术,帮助解码器从隐变量中提取出丰富的输入信息。为了使隐变量能够在解码阶段得到更有效的利用,提出了记忆(Memory)方法,其优势在于允许解码器在每一层直接关注隐变量的潜在信息。具体而言,隐变量 $z \in \mathbb{R}^d$ 被映射到一个额外的记忆向量 $W_m z$,作为解码器关注的一个额外键值对。记忆

向量 $W_M z$ 的计算公式如公式(7)所示:

$$z_M = \begin{bmatrix} z_{key} \\ z_{value} \end{bmatrix} = W_M z \quad (7)$$

其中, $W_M \in \mathbb{R}^{d_k \times d_v}$ 是权重矩阵。隐变量经过一个线性层和形状变化,被分割为2个块,分别代表键和值。得到的键和值被称为记忆向量,在解码器的所有层中被共享和传播。随后,在解码器的每个自注意力层中,在键 k 和值 v 前面加上记忆向量,如公式(8)所示:

$$H^{k+1} = \text{Multihead}(H^k, [z_{key}, k], [z_{value}, v]) \quad (8)$$

其中, Multihead 表示多头注意力, H^k 表示解码器第 k 层的隐藏状态。记忆向量相当于在解码过程的每一层中添加了一个虚拟标记,它与原本的键和值合并,参与了自注意力的计算,这对解码器充分利用隐变量中的信息至关重要。具体而言,Transformer 解码器中每个自注意力层的键值对都会进行注意力计算,记忆向量与键值对合并后也会参与每一层的计算过程,使得隐变量能够被解码器直接关注。通过自注意力层得到的编码表示,会在后续的交叉注意力层中与编码器输出进行交叉注意力计算,计算过程中也能充分利用记忆向量中的信息。交叉注意力层的输出结果会在随后的逐步解码中被使用,相当于记忆向量间接对解码器的每一步生成进行指导。

2.6 训练目标

ECLV 模型通过最大化变分下界 $\mathcal{L}_{ELBO}$ 进行训练,包括3个部分:先验和后验分布的KL散度项,响应 r 的重构损失,以及支持策略 y 的重构损失。具体如公式(9)所示:

$$\begin{aligned} \mathcal{L}_{ELBO} &= \mathcal{L}_{kl} + \mathcal{L}_r + \mathcal{L}_y \\ &= -KL(q_{\phi}(z|c, u, r, y) \| p_{\theta}(z|c, u)) \\ &\quad + E_{q_{\phi}(z|c, u, r, y)}[\log p(r|z, c, u, y)] \\ &\quad + E_{q_{\phi}(z|c, u, r, y)}[\log p(y|z, c, u)] \end{aligned} \quad (9)$$

在优化过程中,KL散度项的逐渐减小与最大化变分下界的目标相矛盾。因此,需要使KL散度项保持负值以确保变分下界的最大化。

文献[21]已表明,当使用隐变量模型(例如CVAE)生成对话时,会出现隐变量消失的问题,这使得简单的优化策略难以在隐变量 z 中编码有用的信息。为了缓解这个问题,本文采用了2种策略:①KL退火^[21]:在训练过程中,KL散度项的权重从0逐渐增加到1,防止模型在训练后期忽略隐变量的信息,只依赖于自回归的解码生成;②词袋损失^[12]:要求解码器预测响应 r 中的词袋 r_{bow} ,鼓励隐变量以非自回归的方式预测响应中的词,从而使隐变量能够捕获响应的全局信息。

本文假设,即使存在与响应 r 相关的支持策略 y, r_{bow} 的预测仍然仅依赖于 z, c 和 u 。因此,对于带有词袋损失的ECLV模型,改进的变分下界 $\mathcal{L}_{ELBO}$ 如公式(10)所示:

$$\begin{aligned} \mathcal{L}_{ELBO} &= \mathcal{L}_{kl} + \mathcal{L}_r + \mathcal{L}_y + \mathcal{L}_{r_{bow}} \\ &= -KL(q_{\phi}(z|c, u, r, y) \| p_{\theta}(z|c, u)) \end{aligned}$$

$$\begin{aligned} &+ E_{q_{\phi}(z|c, u, r, y)}[\log p(r|z, c, u, y)] \\ &+ E_{q_{\phi}(z|c, u, r, y)}[\log p(y|z, c, u)] \\ &+ E_{q_{\phi}(z|c, u, r, y)}[\log p(r_{bow}|z, c, u)] \end{aligned} \quad (10)$$

与传统CVAE模型使用的变分下界相比,本文用于模型优化的变分下界包含了额外的组件:支持策略的重构损失和词袋损失。通过将策略 y 的重构损失整合到当前损失函数中,ECLV可以更有效地将与策略相关的信息编码到隐变量 z 中,而不是仅限于表面的上下文 c 和用户状态 u 。基于此,模型能够更好地利用那些能提供有效情感支持的支持策略。同时,将词袋损失包含在损失函数中,有助于减少对编码器输出和解码器自回归的依赖。相反地,模型充分利用了隐变量 z 中封装的更多信息。这一方法有效缓解了隐变量消失的问题,并有利于提高生成的响应质量。

3 实验

3.1 数据集和评估指标

(1)数据集。本文使用ESConv数据集^[5]评估ECLV模型的性能。其包含多种高质量的注释,如用户的情绪类型、问题类型和用户境况等。对话中使用的支持策略一共有8种(含提问、情感反映和提供建议等),支持者使用的每一种策略都会在对话中标注出来,并且8种策略的使用频率几乎是均匀分布的。由于ESConv数据集中的每个对话相对较长,平均包含29.8个话语,因此需要将每个对话切分成包含8个话语的片段,其中最后一个话语是支持者的回应。随后,将数据集按6:2:2的比例随机划分为训练集、验证集和测试集,与官方的ESConv数据集划分方式一致。ESConv包括模型训练所需的基本标签,如情感支持策略、情绪类型和用户境况。ESConv数据集中1053个样例的整体统计结果,如表1所示。

Table 1 Statistical information of the ESConv dataset

表1 ESConv数据集的统计信息

Category	Total	Supporter	Seeker
# Dialogues	1053	-	-
Avg. Minutes per Chat	22.6	-	-
# Workers	854	425	532
# Utterances	31410	14855	16555
Avg. length of dialogues	29.8	14.1	15.7
Avg. length of utterances	17.8	20.2	15.7

(2)自动评估指标。①本文使用策略预测的准确性(ACC)作为一个评估指标。②使用BLEU-2(B-2)、BLEU-3(B-3)、BLEU-4(B-4)^[22]和Rouge-L(R-L)^[23]来评估生成的响应和参考响应之间的文本相似性。其中,B-n指标评估生成响应和参考响应之间的n-gram重合度,R-L则基于最长公共子序列来评估相似度。③对于响应的多样性,使用Distinct-1(D-1)和Distinct-2(D-2)来衡量,计算响应中独特n-gram的数量^[24]。

(3)人工评估指标。按照Liu等^[5]提出的方法,本文聘

请了3名具有语言学或心理学背景的工作人员,并从测试集中随机选择50个对话样本,分别使用本文的模型ECLV和基线模型来生成响应。为确保评估客观性,3名专业人员需要根据以下5项指标进行双盲测评:①流畅度(Flu.):哪个响应更流畅?②识别度(Ide.):哪个响应在识别用户问题方面更有帮助?③安慰度(Com.):哪个模型的响应更能安慰用户?④建议度(Sug.):哪个响应提供了更有帮助的建议?⑤总体(Ove.):总的来说,哪个模型的响应提供了更有效的情感支持?

3.2 对比模型

(1)Blenderbot-Joint^[5]。在ESConv数据集上进行微调的Blenderbot-small模型。

(2)GLHG^[25]。使用全局到局部的层次图网络来捕获多源信息,并将这些信息的层次关系建模为进行ESC的一种方式。

(3)FADO^[9]。利用用户反馈来鼓励或惩罚策略以更好地选择策略,并关注策略流动到上下文中。

(4)KEMI^[10]。使用新颖的知识增强的混合主动性框架用于ESC,从心理健康知识图中检索外部知识,以更好地选择支持策略。

3.3 参数设置

本文模型使用多层Transformer作为编码器和解码器^[19]。其中,编码器的嵌入层大小为1024,前馈网络层大小为4096,编码器层数为12,注意力头数为16,解码器的各项参数与编码器一致。此外,模型使用Adam优化器进行训练,其中, $\beta_1 = 0.9$, $\beta_2 = 0.999$, $\varepsilon = 1e - 8$ 。学习率设置为 $3e - 5$,并从 $1e - 7$ 开始进行190步的线性预热,训练时的批次大小为16。为了稳定优化,总KL退火步骤为1000,并且KL退火速率 α 为 1×10^{-3} 。最终,模型进行了4个epoch的训练,并根据验证集上的损失选择了最佳模型。对于解码算法,本文借鉴了Liu等^[5]提出的方法,采用Top- k 和Top- p 采样方法,并设置 $k = 30$, $p = 0.9$,温度 $\tau = 0.6$,与基线模型保持了较高的一致性,以便更好地与其他基线模型进行比较。

3.4 实验结果

3.4.1 自动评估

本文模型与其他基线模型在ESConv数据集上的自动评估结果,如表2所示。其中,Blenderbot-joint模型的评估结果参考了Peng等^[9]的实验结果,而其他基线模型的结果均来自原始论文,并保持数据集的训练、验证和测试分割不变。

由表2可知,ECLV模型在所有指标上均超越基线模型,以下将分别从策略选择和响应生成2个方面展开分析:

(1)策略选择方面,ECLV模型在ACC上取得了最好的表现,对比FADO模型提高了0.55%。与其他基线模型相比,ECLV加入情绪原因,建模细粒度的用户情感状态,有

Table 2 Automatic evaluation results

表2 自动评估结果

Model	ACC/%	B-2	B-3	B-4	D-1	D-2	R-L
Blenderbot-joint	28.37	5.52	2.16	1.29	2.71	19.38	15.51
GLHG	-	7.57	3.74	2.13	3.50	21.61	16.37
FADO	32.90	8.00	4.00	2.32	3.84	21.84	17.53
KEMI	-	8.31	-	2.51	-	-	17.05
ECLV(Ours)	33.45	8.91	4.97	3.17	3.94	22.82	21.38

助于选择正确的支持策略。在情感支持对话任务中,模型需要引导用户情绪从消极向积极转变,这就要求对用户情绪状态进行细粒度建模^[26]。然而,传统的情绪原因分析往往过于笼统,难以为支持策略的选择提供有效指导。为此,本文提出的情绪原因模块创新性地融合了与用户情绪和事件信息相关的常识知识,这种设计能生成内容准确、逻辑严谨的情绪原因分析,进而帮助模型选择更合适的支持策略。实验结果表明,该模块使ACC指标获得提升,但提升幅度有限。这主要源于ESC任务中上下文与支持策略之间存在一对多映射关系——同一对话情境下可能存在多个等效的有效策略。当模型选择的策略与人工参考响应不同(但同样有效)时,会导致自动评估指标未能充分反映模型的实际能力。后续人工评估证实,ECLV模型确实能够选择合适的情感支持策略。

(2)响应生成方面,ECLV模型在代表响应相关性和多样性的指标上都取得了最好的结果。实验数据显示,对比KEMI模型,本文模型在B-2指标上提升了0.6%。对比FADO模型,则在R-L和D-2指标上分别提升3.85%和0.98%,其增幅均显著高于ACC指标。该结果有效验证了隐变量模块对提升响应相关性和多样性的帮助。ESC任务涉及多种输入,包括对话上下文和用户状态等。这些信息直接输入到解码器中,可能会产生大量噪声,导致解码器难以充分学习输入信息。本文通过引入连续隐变量来描述丰富的输入数据的潜在特征,并采用记忆模式使解码器能够直接关注这些隐变量的输入信息,从而有效建立编码器和解码器之间的信息传递机制。

综上所述,ECLV模型在策略预测和响应生成2个方面均展现了最优性能,这得益于情绪原因模块与隐变量模块的协同作用。隐变量模块既保持了情绪原因模块对策略预测精度的提升,同时进一步增强了响应的相关性和多样性。

3.4.2 人工评估

本文选择GLHG和FADO模型进行进一步的人工评估,结果如表3所示。人工评估的结果与自动评估的结果一致,模型ECLV在所有指标上都优于GLHG和FADO。与GLHG相比,ECLV模型在流畅度(Flu.)指标上有明显的优势,这表明隐变量帮助模型理解对话上下文的深层语义,生成流畅自然的文本。此外,与FADO模型相比,识别度(Ide.)取得了较好的改进,这表明情绪原因为模型识别用户问题提供了线索,模型获得情绪原因后,更容易识别出

用户遇到的情绪问题。

Table 3 Human evaluation results

表 3 人工评估结果

Comparisons	Aspect	Win	Lose	Tie
ECLV vs. GLHG	Flu.	48.7	24.5	26.8
	Ide.	41.8	30.3	27.9
	Com.	44.0	12.5	43.5
	Sug.	46.7	27.2	26.1
	Ove.	42.6	21.0	36.4
	Flu.	29.0	17.8	53.2
ECLV vs. FADO	Ide.	43.5	29.5	27.0
	Com.	41.3	32.5	26.2
	Sug.	31.5	25.8	52.7
	Ove.	41.0	27.6	31.4

3.5 消融实验

为了验证 ECLV 中每个组件的有效性,本文进行了消融实验,结果如表 4 所示。其中,ECLV w/o ec 表示移除情绪原因模块的变体,ECLV w/o z 表示移除隐变量模块的变体,ECLV w/o ms 表示从解码器中移除记忆模式技术的变体。

Table 4 Ablation study results

表 4 消融评估结果

Model	ACC/%	B-2	B-3	B-4	D-1	D-2	R-L
ECLV w/o ec	30.50	8.23	4.20	2.60	3.48	19.31	20.71
ECLV w/o z	31.12	8.56	4.50	2.76	3.25	17.43	20.91
ECLV w/o ms	30.62	8.25	4.21	2.66	3.39	18.38	20.54
ECLV	33.45	8.91	4.97	3.17	3.94	22.82	21.38

(1)ECLV w/o z。在移除隐变量模块后,ECLV w/o z 在 D-1 和 D-2 指标上显著下降,这表明隐变量模块对增强响应多样性具有重要作用。与响应相关性相关的其他指标也不同程度地下降,进一步证明了隐变量模块能够有效提升响应的相关性。

(2)ECLV w/o ec。在移除情绪原因模块后,对比原始模型 ECLV,ECLV w/o ec 在策略 ACC 指标上下降了 2.95%,这表明识别情绪原因对于选择正确的支持策略至关重要。如果不能准确理解用户负面情绪产生的原因,就很难抓住情绪问题的关键,选择正确的支持策略。以原始模型 ECLV 为基准,对比 ECLV w/o z 和 ECLV w/o ec 在 BLEU 和 R-L 指标上的下降幅度,可以发现 ECLV w/o ec 在这些指标上有更大的下降幅度,这表明情绪原因也会影响响应相关性。这是由于情绪原因可以帮助模型理解上下文的深层次语义和用户的情绪状态,生成与上下文紧密相关的响应。

(3)ECLV w/o ms。在移除记忆模式技术后,ECLV w/o ms 在响应相关性和多样性指标上都显著下降,这表明记忆模式允许解码器充分利用隐变量中的信息,可以有效提高响应的相关性和多样性。

3.6 词袋损失分析

为了探索词袋损失对 ECLV 性能的影响,本文手动设置了不同的词袋损失权重 α 。如表 5 所示,ECLV 的性能在添加词袋损失($\alpha = 1.0$)后达到峰值。然而,当词袋损失被移除($\alpha = 0$)时,ECLV 的性能显著下降。这表明使用词袋损失对模型的性能至关重要。词袋损失的加入鼓励模型

以非自回归的方式生成,模型无法像自回归生成那样利用之前生成的信息,所以会更多地学习到隐变量中包含的信息。

Table 5 Analysis results of bag-of-words loss

表 5 词袋损失分析结果

Model	ACC/%	B-2	B-4	D-1	R-L
ECLV ($\alpha = 0$)	30.42	8.53	2.90	3.60	20.51
ECLV ($\alpha = 0.2$)	31.91	8.59	2.97	3.78	20.94
ECLV ($\alpha = 0.5$)	32.37	8.55	2.98	3.71	20.86
ECLV ($\alpha = 0.8$)	32.62	8.78	3.06	3.58	21.18
ECLV ($\alpha = 1.0$)	33.13	8.91	3.17	3.94	21.38

3.7 策略重构损失分析

为了研究策略重构损失对 ECLV 性能的影响,本文手动设置了策略损失 β 的不同权重。若 $\beta = 0$,则意味着模型仅将支持策略作为识别网络的输入条件,而在训练过程中不进行策略重构。如表 6 所示,当添加策略重构损失($\beta = 1.0$)后,ECLV 的性能达到峰值。这表明仅使用策略作为输入条件,难以使隐变量捕获与策略相关的信息。相比之下,添加策略重构损失后,模型能够学习与支持策略高度相关的隐变量。由于支持策略在很大程度上决定了情感支持回答的有效性,因此隐变量中所学习的支持策略信息对模型性能至关重要。

Table 6 Analysis results of strategy reconstruction loss

表 6 策略重构损失分析结果

Model	ACC/%	B-2	B-4	D-1	R-L
ECLV ($\beta = 0$)	30.74	8.53	2.76	3.55	20.99
ECLV ($\beta = 0.2$)	30.75	8.79	2.86	3.67	21.20
ECLV ($\beta = 0.5$)	31.28	8.61	2.91	3.71	20.68
ECLV ($\beta = 0.8$)	32.45	8.78	3.16	3.87	21.17
ECLV ($\beta = 1.0$)	33.13	8.91	3.17	3.94	21.38

3.8 案例研究

表 7 中展示了一个情感支持对话的实例,便于比较 ECLV 模型生成的回应与基线模型的回应。从对话内容中可以看出,用户正在尝试在全职工作的同时获取许可证。GLHG 与基准响应相似,采用了“自我披露”策略,对求助者表示同情。相比之下,由于 ECLV 模型融入了情绪原因和隐变量,其对求助者的情绪状态有了更深入的理解,提供了更具技巧的情感支持。ECLV 模型采用了“肯定和安慰”策略,不仅指出了求助者的沮丧情绪,还鼓励并保证他或她一定会摆脱这种情况。

4 结语

ESC 任务不仅需要表达共情和安慰用户,还需要选择正确的支持策略来提供相关性和多样性强的响应。因此,本文提出了 ECLV 模型来改进情感支持对话。具体而言,ESC 任务首先需要识别和解决用户的情绪问题。模型通过引入情绪原因模块来深刻理解用户状态,从而找到解决问题的关键。其次,ESC 任务不能使用无关响应刺激用户,而是需要保证响应的高度相关性和多样性。模型通过

Table 7 Emotional support conversation example

表7 情感支持对话实例

Dialogue history	Situation	Seeker	I want to get a proper qualification know work but I can't seem to find...
		Seeker	Have there, I have a work issue that I'm hoping you can help me with.
		Supporter	[Question] Sure, how can I help you?
		Seeker	I want to get a qualification but I'm working full-time at the moment... ..
Response		Seeker	Have you ever had a situation like this?
		MISC	[Self-disclosure] I have been through it and I can tell you that it was very painful.
			[Affirmation-and-Reassurance] I have not. It sounds like you are frustrated about the situation, and that you want to work out a way to get through it. I'm sure you'll get there.
		ECLV	[Self-disclosure] I have not but I know people who are in your situation.
Ground-Truth			who are in your situation.

引入隐变量模块,提升了响应与上下文的相关性和多样性。与以前的研究不同,ECLV首先引入情绪相关的知识作为推理线索,建立对用户情绪状态更细致的理解,从而准确推理情绪原因。随后,利用连续隐变量来描述输入数据的潜在特征,并应用记忆模式将隐变量学习的信息融入到解码器中。自动评估和人工评估的结果表明,ECLV超越了现有的最先进模型。进一步通过实验分析证明,ECLV能够利用情绪原因和隐变量选择正确的支持策略,并生成更相关和多样的响应。但本文未能验证模型在中文场景下的效果,因为目前情感支持对话的中文数据集资源仍然较为匮乏,这使得模型的训练成为一大挑战。在未来的工作中,将进一步探索如何建立针对中文的情感支持对话数据集。

参考文献:

- [1] XUE H, LIANG Y, MU B, et al. E-chat: emotion-sensitive spoken dialogue system with large language models [DB/OL]. <https://export.arxiv.org/abs/2401.00475v3>.
- [2] WEN Z, CAO J, SHEN J, et al. Personality-affected emotion generation in dialog systems [J]. *ACM Transactions on Information Systems*, 2024, 42(5): 1-27.
- [3] YANG Z, REN Z, YUFENG W, et al. Exploiting emotion-semantic correlations for empathetic response generation [DB/OL]. <https://arxiv.org/abs/2402.17437>.
- [4] WANG Y, CHEN C, YANG Z, et al. CTSM: combining trait and state emotions for empathetic response model [C]//Proceedings of the 2024 Joint International Conference on Computational Linguistics, 2024: 4214-4225.
- [5] LIU S, ZHENG C, DEMASI O, et al. Towards emotional support dialog systems [C]//Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing, 2021: 3469-3483.
- [6] ROLLER S, DINAN E, GOYAL N, et al. Recipes for building an open-domain chatbot [DB/OL]. <https://arxiv.org/abs/2004.13637>.
- [7] ZHANG Y, SUN S, GALLEY M, et al. DialoGPT: large-scale generative pre-training for conversational response generation [DB/OL]. <https://arxiv.org/abs/1911.00536>.
- [8] XU X, MENG X, WANG Y. PoKE: prior knowledge enhanced emotional support conversation with latent variable [DB/OL]. <https://arxiv.org/abs/2210.12640>.
- [9] PENG W, QIN Z, HU Y, et al. FADO: feedback-aware double controlling network for emotional support conversation [J]. *Knowledge-Based Systems*, 2023, 264: 110340.
- [10] DENG Y, ZHANG W, YUAN Y, et al. Knowledge-enhanced mixed-initiative dialogue system for emotional support conversations [C]//Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics, 2023: 4079-4095.
- [11] BOSSELUT A, RASHKIN H, SAP M, et al. COMET: commonsense transformers for automatic knowledge graph construction [C]//Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, 2019: 4762-4779.
- [12] ZHAO T, ZHAO R, ESKENAZI M. Learning discourse-level diversity for neural dialog models using conditional variational autoencoders [C]//Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, 2017: 654-664.
- [13] PORIA S, MAJUMDER N, HAZARIKA D, et al. Recognizing emotion cause in conversations [J]. *Cognitive Computation*, 2021, 13: 1317-1332.
- [14] ZHAO W, ZHAO Y, LI Z, et al. Knowledge-bridged causal interaction network for causal emotion entailment [C]//Proceedings of the AAAI Conference on Artificial Intelligence, 2023: 14020-14028.
- [15] BAO S, HE H, WANG F, et al. PLATO: pre-trained dialogue generation model with discrete latent variable [C]//Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, 2020: 85-96.
- [16] CHEN W, GONG Y, WANG S, et al. DialogVED: a pre-trained latent variable encoder-decoder model for dialog response generation [C]//Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics, 2022: 4852-4864.
- [17] TU Q, LI Y, CUI J, et al. MISC: a mixed strategy-aware model integrating comet for emotional support conversation [C]//Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics, 2022: 308-319.
- [18] HWANG J D, BHAGAVATULA C, LE B R, et al. COMET-ATOMIC 2020: on symbolic and neural commonsense knowledge graphs [C]//Proceedings of the AAAI Conference on Artificial Intelligence, 2021: 6384-6392.
- [19] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [C]//Proceedings of the 31st International Conference on Neural Information Processing Systems, 2017: 5998-6008.
- [20] LI C, GAO X, LI Y, et al. Optimus: organizing sentences via pre-trained modeling of a latent space [C]//Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, 2020: 4678-4699.
- [21] BOWMAN S R, VILNIS L, VINYALS O, et al. Generating sentences from a continuous space [C]//20th SIGNLL Conference on Computational Natural Language Learning, 2016: 10-21.
- [22] PAPINENI K, ROUKOS S, WARD T, et al. Bleu: a method for automatic evaluation of machine translation [C]//Proceedings of the 40th annual meeting of the Association for Computational Linguistics, 2002: 311-318.
- [23] LIN C Y. Rouge: a package for automatic evaluation of summaries [C]//Workshop on Text Summarization Branches Out, 2004: 74-81.
- [24] LI J, GALLEY M, BROCKETT C, et al. A diversity-promoting objective function for neural conversation models [C]//Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics, 2016: 110-119.
- [25] PENG W, HU Y, XING L, et al. Control globally, understand locally: a global-to-local hierarchical graph network for emotional support conversation [C]//International Joint Conference on Artificial Intelligence, 2022: 4324-4330.
- [26] XU B, XU P. Emotion elicited question generation model in dialogue scenarios [J]. *Computer Science*, 2024, 51(11): 265-272.

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.

(责任编辑:陈维捷)