一种融合情感和策略信息的共情对话生成方法

朱振方[①] 李嘉欣[②] 徐富永*[②] 刘培玉[②] 张广渊[①]

[①](山东交通学院信息科学与电气工程学院 济南 250357)

[②](山东师范大学信息科学与工程学院 济南 250358)

摘 要：共情对话旨在为情感焦虑的对话系统聊天用户提供心理健康支持，因此，赋予对话系统共情能力是一个值得关注的问题。现有方法往往只能识别用户的情感状态，并不能根据聊天用户不同的情感状态生成有效的、具有同理心的回复，更不能缓解用户的不良情感。因此，在构建情感支持对话系统的研究中，如何动态地捕捉用户的细粒度情感特征并根据情感特征提供相应的心理支持，需要进一步地探索。该文提出一个情感和策略信息融合的共情对话生成方法，该方法首先使用情感分类网络动态感知用户的情感状态；然后利用支持策略准确地建模策略匹配网络，并根据对话上下文引入对话生成网络进行回复生成；最后，通过比较所提方法和当前较为先进的方法在相应数据集上的实验结果，验证所提方法的有效性以及情感支持的重要性。

关键词：共情对话；情感支持；对话生成；支持策略

中图分类号：TN911.7; TP183 **文献标识码**：A **文章编号**：1009-5896(2024)08-3382-08

DOI: 10.11999/JEIT231417

Empathetic Dialogue Generation via Sentiment and Support Strategy

ZHU Zhenfang[①] LI Jiaxin[②] XU Fuyong[②] LIU Peiyu[②] ZHANG Guangyuan[①]

[①](*School of Information Science and Electrical Engineering, Shandong Jiaotong University, Jinan* 250357, *China*)

[②](*School of Information Science and Engineering, Shandong Normal University, Jinan* 250358, *China*)

Abstract: Empathic dialogue aims to provide mental health support for anxious users, thus chatbots with empathic capabilities is a noteworthy issue. The existing methods can only identify users' sentiment states, but can not effectively generate empathetic responses according to users' different sentiment states and let alone effectively relieve users' bad emotions. Therefore, in the research of building sentiment support chatbots, how to dynamically capture users' fine-grained sentiment features and provide corresponding psychological support according to sentiment features needs to be further explored. This paper proposes an empathic dialogue generation method based on the fusion of emotion and strategy. Firstly, the sentiment classification network is used to dynamically perceive the user's sentiment state. Then the support strategy is used to accurately model the strategy matching network which is introduced according to the context of the conversation to generate the conversation. Finally, by comparing the experimental results of the proposed method and the current advanced methods on the corresponding datasets, the effectiveness of the proposed method and the importance of sentiment support are verified.

Key words: Empathic dialogue; Sentiment support; Dialogue generation; Support strategy

1 引言

共情对话系统需要挖掘对话上下文信息中用户的显式和隐式特征，建模用户对话情感，赋予对话系统共情能力，与人类建立更有效的沟通。此外，当面对用户的不良心理状态时，理解用户的情境并提供有效疏导则可以用于心理治疗等具有更广泛场景[1]。本文将探讨在用户对话过程中如何关注情感信息，并通过构建智能体来挖掘情感信息，以实现与用户共情的目标。

构建共情对话系统的关键在于如何将情感融入到回复中，较早的研究是Ghosh等人[2]应用情感标签和强度进行预测，能够根据所确定的情感策略产生反应。之后，Zhou等人[3]以一种自然而连贯的方式考虑情感，提出了一种情感聊天机(Emotional Conversation Machine, ECM)，基于不同的情绪标

收稿日期：2023-12-25；改回日期：2024-04-23；网络出版：2024-07-25

*通信作者： 徐富永 xfysec@163.com

基金项目：国家社科基金 (19BYY076)

Foundation Item: The National Social Science Foundation (19BYY076)

签自动生成回复。受ECM的启发，Wei等人[4]在ECM中加入了一个情感选择器，可以自动生成具有个性化特征的回复。目前的共情对话生成方法多遵循多任务学习的范式，通过联合训练一个情绪预测模块与对话生成模块，以实现具有共情约束的对话生成，但是这些方法在如何动态地捕捉用户的情感特征和如何缓解用户的负面情感两个问题上还存在极大的探索空间。

为了更好地解决上述两个问题，本文提出了一种融合情感和策略信息的共情对话生成模型ESFM(Emotional Strategy Fusion Model)，并在ESConv数据集[5]上进行了较为全面的实验。在基于自动指标和人工判断两方面的评测中，证明了所提模型产生的回复更具有相关性和同理心。本文主要贡献如下：

(1)从对话历史中动态捕捉对话系统用户细粒度的情感状态，获取用户当前正确的心理状态。

(2)使用了具有策略支持的数据集ESConv，在模型中通过匹配策略，使生成的回复能够缓解对话系统用户的焦虑并提供情感支持。

(3)在公开数据集上进行了实验，并与其他基线方法进行比较，结果显示，本文所提出的方法在困惑度指标上提升了2%左右。

2　问题来源

共情对话系统旨在捕捉用户的情感，生成针对用户情感表达的回复，与用户进行有效沟通，从而创建更具共情能力的对话系统[6]。Zhou等人[3]首次提出了共情对话系统模型，该模型能够根据预定义的情感准确地在回复中表达出指定的情感。此后，为了增强对话系统与用户的共情能力，共情反馈任务引起了广泛的研究兴趣，Rashkin等人[7]认为通过了解用户的情感可以生成更准确的共情反应；为此，Lin等人[8]设计了一个专门的解码器来响应理解到的情感；Majumder等人[9]采用情感模仿[10]的概念，让回复更有同理心。随着预训练模型在各个领域的广泛应用，大型预训练模型也在共情对话中发挥了重要作用，Radford和Zandie等人[11,12]使用GPT[13]来生成共情回复，取得了更好的效果。

上述工作在自动和评价指标都证明了共情对话在理解用户情感方面发挥着积极的作用，能更好地获得用户的信任。然而，通过对回复数据进行大量的分析，模型在情感分析方面还存在欠缺，对情感追踪的粒度不足使得模型准确定位用户的实际情感信息，从而影响了后续生成任务。

除此之外，共情回复的任务致力于表达情感，如快乐或悲伤，但面对不好的情感时，无法通过适当的支持技能来减少用户的痛苦，这降低了共情对话系统的实用性。为了解决这个问题，一些传统的对话系统利用人为设计的规则和策略来提供情感支持反应[14]。基于这些数据集所开展的研究，让对话系统具有了更强的情感支持能力。因此，本文在产生支持性反应时，使用含有行为策略引导的对话数据集，根据现有的带有支持策略的对话数据集，让不同的用户情感和策略进行有效匹配，使模型能根据用户的情感需要进行情感支持。

3　方法

一个对话可以被定义为一系列的话语序列$d=\{d_1,d_2,\cdots,d_n\}$，其中n为话语的总数。每个话语d_j包含一个单词序列$\{w_1^j,w_2^j,\cdots,w_i^j\}$，其中$i$为话语$d_j$中的最大单词数。假设$r$是目标响应，$\boldsymbol{S}$是用户在本次对话的情景状态，比如用户当前的情景状态为在一次考试中失败，则用户与系统的对话都会围绕这个主题展开。因此，本文将共情对话生成定义为语言序列预测的任务，共情对话的目标是基于当前的对话情景$\boldsymbol{S}$和对话的话语序列d生成富有合适情感的回复，达到对话系统与用户共情的目的，任务的公式化定义为估计概率$P(r|d,\boldsymbol{S})$。

本文提出的ESFM模型如图1所示，在该模型中，首先建立了一个情感分类网络，将对话情景传递到多层注意力模块中，以捕获多粒度的情感信息，并最终通过向量隐藏状态进行情感分类；然后，为了使生成的回复更加自然并更具情感支持，我们设计了策略匹配网络，通过3层模块来探究策略与回复之间的条件关系，并得到低维向量表示；最后，本文建立了对话生成网络，通过多头注意力机制的关系建模，使预测更接近相关目标回复。下文将详细阐述这3个子网络。

3.1　情感分类网络

为了在对话生成的过程中更好地动态捕捉用户的情感特征，了解用户的心理状态，在对话开始之前将情景信息$\boldsymbol{S}$输入到嵌入层进行学习建模，这样可以将输入转换为更好的表示。单词的输入嵌入是词嵌入$\boldsymbol{E}_w$和位置嵌入$\boldsymbol{E}_p$的和

$$\begin{aligned}\boldsymbol{E}&=\boldsymbol{E}_w+\boldsymbol{E}_p\\ \boldsymbol{E}_p&=\{e_1^p,e_2^p,\cdots,e_m^p\}\\ \boldsymbol{E}_w&=\{e_1^w,e_2^w,\cdots,e_m^w\}\end{aligned}\tag{1}$$

其中e_i^{w}, e_i^{p}分别为语言序列中第i个位置的词嵌入表示和位置嵌入表示，m为语言序列的长度。

为了从对话情景信息中捕捉对话情感的长距离依赖关系，采用多头注意力机制对对话情景进行建模，得到对话情感状态特征h_s

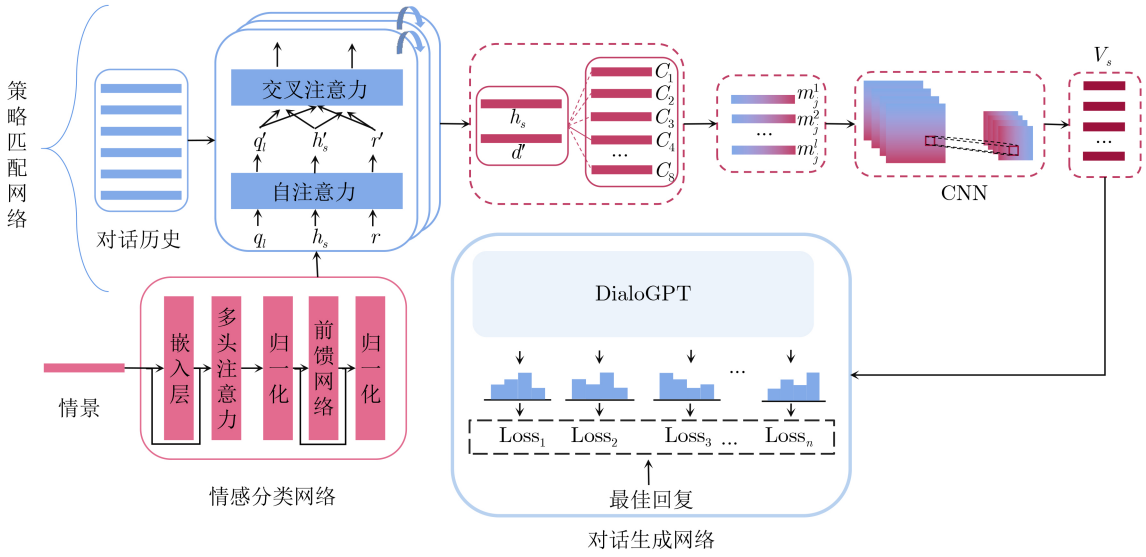

图 1 ESFM模型示意图

$$h_s = \mathrm{FFN}(\mathrm{MultiHead}(\boldsymbol{S},\boldsymbol{S},\boldsymbol{S}))$$
$$\mathrm{FFN}(x) = \max(0, x\boldsymbol{W}_1 + b_1)\boldsymbol{W}_2 + b_2$$
$$\mathrm{MultiHead}(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V}) = \mathrm{Concat}(\mathrm{head}_1,\cdots,\mathrm{head}_h)\boldsymbol{W}$$
$$\mathrm{head}_i = \mathrm{Softmax}\left(\frac{\boldsymbol{Q}\boldsymbol{W}_i^{\boldsymbol{Q}}\cdot(\boldsymbol{K}\boldsymbol{W}_i^{\boldsymbol{K}})^{\mathrm{T}}}{|d_K|}\right)\cdot\boldsymbol{V}\boldsymbol{W}_i^{\boldsymbol{V}} \quad (2)$$

为了更好地理解用户情感，本文在对话数据集中提取了主要的11种情感类别。情感分类情况如左半部分所示。并利用预训练模型Generative Pre-trained Transformer(GPT)[13]作为情感分类基础模型。在预训练阶段，GPT模型通过大规模的无监督学习从文本数据中学习语言的统计特征和上下文关系。在下游任务中，预训练好的GPT模型通过微调或迁移学习，适应特定的任务需求。

3.2 策略匹配网络

为使模型依据用户的情感状态和对话历史，使用正确的策略连贯地输出回复，隐藏在对话过程中的会话风格对回复的生成非常有帮助。因此，受到Qian等人[15]的启发，本文设计了一个策略匹配网络，旨在从用户的情感中建模用户会话风格，并提供情感支持。使用回复策略的数据集ESConv生成支持性回复。该数据集使用8种不同的策略[5]，并提供了每个策略的详细定义，具体定义如表1右半部分所示。

形式上，给定一个对话历史$d^{'}=\{d_0,d_1,\cdots,d_{i+1}\}$，$i\in(0,n-1)$，用户的对话情感状态$h_s$和响应候选$r$，其中$r$中包含了匹配策略$C_n$。策略匹配网络的目的是得到一个风格匹配特征向量$\boldsymbol{g}\{h_s,r,d^{'}\}$，用来度量响应候选$r$的风格一致性。策略匹配网络通过一个3层模块实现：

(1)表示层：提取n层情感状态和候选响应的多粒度语义表示，其目的是获得多粒度上下文表示和交叉注意表示。具体来说，以第j个语句为例，首先通过Word2Vec来初始化d_j的嵌入表示得到E_J，并采用多层自注意力的方式来获得d_j的高维语义表示q_j^l

$$q_j^l = \mathrm{FNN}\left(\mathrm{MultiHead}\left(\boldsymbol{E}_J,\boldsymbol{E}_J,\boldsymbol{E}_J\right)\right) \quad (3)$$

其中，l为多层自注意力的层数。$\boldsymbol{E}_J$为上下文语句的嵌入表示Word2Vec(d_j)。通过第l层的自注意力模块，得到了不同粒度下单词向量的第$l+1$层表示。此外，本文认为回应风格也是基于用户当前的心理状态。例如，对于用户高兴或者悲伤的回复应该是不同的。鉴于此，让对话的上下文表征q_j^l关注对应的对话情感状态h_s，以获得多层交叉注意力表示

$$\boldsymbol{u}_j^l = \mathrm{MultiHead}\left(q_j^l,h_s,h_s\right) \quad (4)$$

表 1 情感分类和情感支持策略

情感分类	情感支持策略
焦虑 内疚 抑郁 嫉妒 厌恶 痛苦 恐惧 愤怒 羞耻 悲伤 紧张	(1)提问：指通过疑问的方式询问用户的感受或是遇到的困难。 (2)重述或转述：对用户的状态进行简单的重述，可以帮助他们更清楚地看到自己的情况。 (3)回复感受：清晰地表达和描述用户的感受。 (4)自我披露：与用户分享相似的经历来表达同理心。 (5)确认和保证：肯定用户的优点和能力，并提供安慰和鼓励。 (6)提供建议：提供关于如何改变的建议告诉他们该怎么做。 (7)信息：为用户提供有用的信息，例如数据、意见、资源等。 (8)其他：使用其他类别的支持策略对用户提供帮助。

(2)匹配层：在每个语义层次上进行匹配，其目的是获得一个样式匹配矩阵M_s，该矩阵在多个粒度下测量候选响应与每个对话历史上下文的样式匹配度。具体来说，给定对话情感状态h_s和第j个对话及历史d_j，计算上下文表示和交叉注意力表示的匹配矩阵，公式为

$$m_j^l = \frac{q_j^l \cdot \boldsymbol{u}_j^{l^{\mathrm{T}}}}{\sqrt{d}}$$
$$\boldsymbol{M}_s = \{m_j^0, m_j^1, \cdots, m_j^l\} \tag{5}$$

(3)提取层：动态融合对话情感状态与所有对话之间的匹配信号，并对$\boldsymbol{M}_s$进行特征的维度变换，其目的是通过卷积神经网络(Convolutional Neural Networks, CNN)从匹配矩阵中提取匹配特征，然后将提取的特征线性映射到相应的维度

$$\boldsymbol{V}_s = \mathrm{Conv2D}(\boldsymbol{M}_s, k) \tag{6}$$

其中，k 是卷积的大小。通过上述计算，得到了策略、对话历史和情感特征之间的匹配信号。为了进一步获得匹配信号，使用自注意力机制得到$\boldsymbol{g}\{h_s, r, d^{'}\}$

$$\boldsymbol{W}_s = \mathrm{softmax}(\mathrm{MLP}(\tanh(\mathrm{MLP}(\boldsymbol{V}_s))))$$
$$\boldsymbol{g}\{h_s, r, d^{'}\} = \sum_{\dim=0} \boldsymbol{W}_s \cdot \boldsymbol{V}_s \tag{7}$$

其中，$\boldsymbol{W}_s$ 表示注意力权重。

通过上述策略匹配模块，针对捕获的情感进行有针对性的对话生成，从一定程度上可以实现针对心理状态变化的动态建模。

3.3 对话生成网络

本文将对话生成的任务看成序列预测的任务，采用预训练的语言模型 DialoGPT[16]初始化模型，与GPT-2[12]预训练模型一样，DialoGPT被表述为一个自回归语言模型，并使用多层转换器作为模型体系结构。然而，与GPT-2不同的是，DialoGPT解决了GPT-2在对话生成任务中所遇到的挑战，在保证回复的自然性和多样性的同时，还能有效地保留上下文的信息。DialoGPT将多回合对话会话建模为一个长文本，并将生成任务建模成序列预测任务。

在对话生成网络的每一层中，进行多头自注意力操作：

$$r_j^l = \mathrm{FNN}(\mathrm{MultiHead}(\boldsymbol{V}_s \cdot \boldsymbol{W}_j^Q, \boldsymbol{V}_s \cdot \boldsymbol{W}_j^K, \boldsymbol{V}_s \cdot \boldsymbol{W}_j^V)) \tag{8}$$

$\boldsymbol{W}_j^Q$, $\boldsymbol{W}_j^K$, $\boldsymbol{W}_j^V$ 分别是第j 层注意力机制的查询、键、值变换矩阵。每一层的r^l 融合了来自策略情感匹配网络中的匹配信息，对最后一层的隐藏层状态进行线性变换，并将特征映射到词表，获得最后生成的回复y' 。

4 实验

本节将详细介绍实验中采用的数据集、基线模型、评估指标和实验结果。

4.1 数据集

本文使用了情感支持对话数据集ESConv[5]，相对于之前的情感对话数据集，ESConv数据集能在多次的对话互动回合使用更直接的情感支持策略。

4.2 评估指标

本文采用了自动评价和人工评估指标来测试模型的性能。

4.2.1 自动评估指标

(1)困惑度perplexity(PPL)[17]：目前大多数语言生成模型都是通过最小化负对数似然值(negative log-likelihood, NLL)来训练的，这一指标被广泛用作公共标准。具体来说，给定生成模型$P_G(Y|X)$，给定输入$X = \{X_1, X_2, \cdots, X_n\}$和预测序列$Y = \{Y_1, Y_2, \cdots, Y_n\}$。困惑度的计算公式为

$$\mathrm{PPL} = \exp\left(-\frac{\sum_{i-1,t-1}^{n;|Y_i|} \log P_G(y_{i,t}|X_i, Y_{i<t})}{\sum_{i-1}^{n} |Y_i|}\right) \tag{9}$$

其中$|Y_i|$ 表示Y_i 中的单词数量，$y_{i,t}$ 是Y_i的第i个单词，$Y_{i<t}$ 是Y_i的前$t-1$ 个单词。在计算过程中，困惑度反映了输出与目标句之间的拟合程度，困惑度越低，意味着语言模型的性能越好。

(2)多样性Dist[18]：多样性用于度量在开放域对话模型中生成的短语的唯一性，假设该模型在测试阶段生成N 个句子$y^* = \{Y_1^*, Y_2^*, \cdots, Y_N^*\}$，计算$y^*$中所有长度为$k$的短语，由此生成短语集合计算如下：

$$G_K = \bigcup_{i=1}^{N} \mathrm{count}(k - \mathrm{grams}(Y_i^*)) \tag{10}$$

其中count$(\cdot)$为计数函数，k-grams$(\cdot)$表示的不同粒度的词的数量。

用Dist-k(Distinct-k)衡量生成回复的多样性:

$$\mathrm{Dist} - k(y^*) = \frac{|G_k|}{\sum_{s \in G_k} \sum_{i=1}^{N} C(Y_i^*, s)} \tag{11}$$

其中，$C(Y_i^*, s)$ 代表s 出现的次数。为了更准确地评估模型，我们同时采用了Dist-1和Dist-2两个指标分别衡量不同词粒度(1-gram, 2-gram)的句子多样性质量。

(3)双语评估替换BLEU[19]：通常用来度量一组对话回复中句子集合与人工标注回复中的句子集合

的相似程度，BLEU根据n-gram可以划分为多个评价指标，常见的有BLEU-1, BLEU-2, BLEU-3, BLEU-4等4种，数字表示连续单词的个数。BLEU-1衡量的是单词级别的准确性，高阶BLEU可以衡量句子的流畅性。为了更精确地衡量模型的性能，本文采用BLEU-2和BLEU-4指标同时验证。

(4)Rouge[20]：通过将自回归生成的文本与目标回复进行比较计算，得出相应的分值，以衡量二者之间的“相似度”，评估生成响应的词汇和语义方面的性能。

(5)准确度Accuracy(ACC)：通过测量正例和负例中预测正确数量占总数量的比例来测试支持策略的准确性。在评估模型时，准确度越高，表明该模型选择响应策略的能力越好。

4.2.2 人工评估指标

为了减少实验结果的偶然性，需要人工评论来检查这些反应是否合格。为此，招募了4名在对话生成方面表现出色的研究人员作为注释者。注释者将所提模型生成的对话反应与每个基线进行比较，从3个方面来判断反应：(1)反应的流畅性：生成的反应是否符合清晰的条件反射和正确的语法标准。(2)回答相关性：回答内容是否与对话主题或问题相关。(3)同理心反应：同理心反应意味着在输入查询问题时，模型生成的答案应在一定的程度上理解用户的心理状态，并进行情感疏导。

对于反应流畅性和反应相关性两个标准，我们使用“1～5”5个分值，分值越高，代表回复的质量越高。其中1表示糟糕，5表示完美。对于同理心反应的评估标准，本文采用“–1，0，1”3个取值空间，其中–1表示矛盾，0表示中立，1表示具有同理心。详细的自动评估实验结果如表2所示，人工评估实验结果如表3所示。

4.3 基线模型

本节将提出的对话生成模型与以下方法进行了比较。

(1)Transformer[21]作为基本的Seq2Seq模型，接受用户的对话历史信息生成对应的回复。

(2)MT Transformer[7]是一种多任务的Transformer，它将情感预测视为一种额外的学习任务，通过多任务学习情感识别。

(3)MoEL[8]结合多个解码器的输出状态，以增强对不同情感的反应共情，并生成含有情感信息的回复。

(4)MIME[9]考虑基于极性的情感集群和情感模拟来产生共情反应。

(5)DialoGPT[16]是ESConv数据集上的基本模型，使用现有的预训练模型对对话历史进行建模，生成带有情感支持的策略回复。

(6)DCKS[22]结合共情知识选择的自适应模块，通过控制反应和处境之间的一致性来提高回复的质量。

4.4 实验细节

在本实验环境中，我们使用NVIDIA RTX3090 GPU，内存大小为24GB。为了获得更好的实验结果，并将单词嵌入的维度设置为768，学习率设置为5e-5。为了公平比较，所有被比较的模型均使用12层注意力机制，并使用AdamW优化器。

4.5 实验结果

本次实验含有7种类型的自动指标和3种类型的人工评估指标，它们从不同的方面反映了模型的性能。从表2可以看出，Transformer模型的评价性能最差，因为它没有任何特定的优化目标来学习共情的能力，而且没有在对话的上下文历史中学习到关键信息。MT Transformer、MoEL和MIME的性能也表现不佳，尽管它们3个能捕获上下文中的关键信息且能进行情感预测，但他们只能获取对话水平的静态情感标签和粗粒度情感信号，这也是现有研究工作存在的不足之处。更重要的是，这3种模型并没有使用具有策略支持的情感数据集，这也意味着他们不具备在回复中安慰用户的能力。通过引入

表 2 不同模型在ESConv数据集中的自动评价性能

	ACC	PPL	D-1	D-2	B-2	B-4	R-L
Transformer	–	89.61	1.29	6.91	6.53	1.37	15.17
MT Transformer	–	89.52	1.28	7.12	6.58	1.47	14.75
MoEL	21.72	133.1	2.33	15.26	5.93	1.22	14.65
MIME	20.26	47.51	2.11	10.94	5.23	1.17	14.74
DialoGPT	28.57	20.4	4.12	17.72	5.78	1.74	16.39
DCKS	30.74	21.83	4.26	18.20	6.58	2.03	15.77
ours	30.4	19.82	4.30	18.23	6.45	2.04	16.28

表 3 不同模型在ESConv数据集中的人工评价性能

模型	流畅性	相关性	同理心
Transformer	0.62	0.31	0.29
MT Transformer	0.78	0.34	0.82
MoEL	0.36	0.80	0.33
MIME	1.13	0.27	0.35
DialoGPT	1.84	0.62	1.04
DCKS	1.80	0.67	1.14
ours	1.79	0.71	1.15

支持策略，DialoGPT相比于前面的基线模型有着更好的表现力，这也说明回复中策略支持的重要性。如表2不同模型在ESConv数据集中的自动评价性能所示，可以看到本文模型比DialoGPT模型更有效。DCKS在BLEU-2和情绪分类指标上有更好的效果，但在困惑度等其他指标上本文模型的评估结果更优。在模型中加入策略匹配网络，使上下文对话和情感与策略相匹配。比较结果表明，考虑情感支持对话的策略复杂性是有益的。

表3为本次实验的人工评估结果，结果显示，在表达同理心方面，所提模型得到了最高的分数，这说明所提模型更有效地融合利用了支持策略数据集，能帮助学习策略选择的特征，生成一致的表达。

对于对话质量，评价结果意味着本文所提模型能更好地拟合数据集，准确地推断和模拟未知数据。此外，模型在Dist、BLEU等度量上明显比其他方法更好，这表明反应的多样性和对话的相关程度更高。ESFM模型在PPL中不突出的主要原因是模型挖掘了更细粒度的情感信号，这通常导致比DialoGPT产生复杂的词汇反应。DCKS由于引入了外部常识知识，减弱了回复的情感匹配能力和句子流畅性。

4.6 实验分析

4.6.1 消融实验

上述实验结果表明，本文模型在许多方面都优于基线。为了更准确地验证模型中不同成分的贡献，本文进行了消融研究，分别删除了情感分类网络和策略匹配网络，实验结果如表4所示。

可以观察到，当去除情感分类网络时，性能明显下降，这表明情感的细粒度提取对生成回复质量是有益的。如果情感的分类只涵盖褒贬大类，缺少情感强度量化信息，那么在具体的对话回复中，会出现回复单一，无法与用户产生共鸣。此外，当策略情感匹配网络被移除时，模型将失去对对话中的情感策略进行建模的能力，从而导致建模对话仅基于语义上下文。语言表现的下降证明了情感策略的建模在对话过程中起着重要的作用。不同的支持策略代表了不同的对话语气和方法，从实验中可以看出，策略匹配的越正确，回复的共情能力和情绪缓解方面越好。因此，本文模型可以更好地模拟长而复杂的对话，并缓解用户的焦虑情感。

4.6.2 样例分析

由于在基线比较中，只有DialoGPT运用了含有支持策略的数据集，并且比其他比较模型有明显的性能提升。本文使用DialoGPT与ESFM进行实际的响应生成比较，比较结果如表5所示。根据两方面进行评估：(1)策略选择的准确性：如图2所示，根据与参考生成答案的对比，我们的模型更能正确选择支持策略，这表明模型中的策略匹配模块的有效性和重要性。(2)回复的质量：两个模型生成的回复都是流畅的，但ESFM更贴合给定情景和对话的主题，对情感进行细粒度挖掘能更好地提升回复的相关性。同时，情感的动态挖掘也同样对策略的选择起着关键作用。

4.6.3 策略选择准确度

除了样例分析之外，本文还试图对策略选择准确度进行深度探究，分析模型与DialoGPT在不同候选答案数量上策略选择的准确度，从而对模型进行全面评估，实验结果如图2所示。

Top-k是指包含k个候选回复，本次实验将k值设为1，2，3，4等4种选择。如图2所示，本文模型的Top-k策略预测精度始终超过比较模型DialoGPT，且在候选回复为3时，策略的预测精度接近60%，表明我们的模型在策略匹配模型建模是有意义的。策略预测的精确程度决定着生成响应的准确性。

5 讨论和分析

ChatGPT作为一种先进的大型语言模型为对话系统带来了前所未有的语义理解和响应生成能力，显著提升了与人类用户的交互体验。不少研究

表 4 消融实验，分别去掉情感分类网络和策略匹配网络

Model	D-1	B-2	R-L
ours	4.30	6.45	16.28
W/o 分类	3.75	6.28	16.10
W/o 策略	4.10	6.19	16.19

表 5 ESFM和DialoGPT产生响应的对比

情景:	I'm fearful of where I'll be living in the future.(我害怕我将来会住在哪里)
用户:	I'm looking for someone new to talk to. I don't really feel like I can express my actual feelings in my current living arrangement. (我在一个新的倾诉对象。在我目前的生活安排中，我真的觉得自己无法表达自己的真实感受)
参考:	I understand. I have had to stay by myself for most of the year due to the pandemic.That has been hard. (我理解。由于疫情，我不得不一个人待了一年的大部分时间。这很难) [自我披露]
DialogGPT:	Have a great night and tare care of yourself.(度过一个美好的夜晚好好照顾自己) [回复感受]
ESFM:	I know there are tough times and we all need to be heard.(我知道现在很艰难，我们都需要被倾听) [自我披露]

者对其在各种传统自然语言处理任务中的表现进行研究分析，然而它对情感的理解能力尚未探讨。下文通过实验统计将本文模型和ChatGPT进行对比分析。

本文随机准备了60个问题并分别在ESFM模型和ChatGPT上进行实验。为了得到充分的验证。这些问题中消极情绪、积极情绪和中性情绪的分布比为1:1:1，并随机邀请3位相关领域研究人员对两个模型的回复效果进行打分。回复结果指标分为相关性和同理心效果。实验结果如表6所示。

根据表6可知，ChatGPT在积极情感中表现较好，能准确地回复用户的情绪和问题。然而在消极情感中，ChatGPT在同理心方面表现欠缺。这说明了本文模型在对消极情感用户的响应方面有更好的疏导作用。

6 结论

本文提出了一种新的情感策略融合模型。该模型可以在对话中动态地捕捉到用户的细粒度情感，通过捕捉用户的情感倾向，并基于其情感依赖进行策略的匹配建模，这是大多数现有工作所忽略的。本文在具有情感支持的数据集上进行了大量实验，且模型在大部分指标上优于其他先进的模型，取得了具有竞争力的性能，证明了该模型的有效性。此外，个性化信息的融入可以让共情对话系统更加个性化、拟人化，这也将是我们未来研究方向的重点工作。

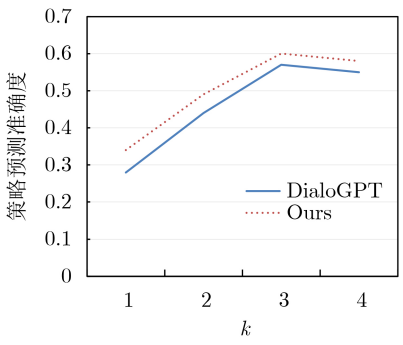

图 2 Top-k策略预测准确度

表 6 ESFM模型和ChatGPT人工评价性能

模型	情感特性	相关性	同理心
ESFM	积极	0.91	0.34
	消极	0.86	0.90
	中性	0.82	0.80
ChatGPT	积极	0.95	0.36
	消极	0.85	0.85
	中性	0.85	0.81

参 考 文 献

[1] 黄宏程, 苏美丹, 寇兰, 等. 基于主从博弈的多方人机交互对话心理模型[J]. 电子与信息学报, 2023, 45(5): 1758–1765. doi: 10.11999/JEIT220441.
HUANG Hongcheng, SU Meidan, KOU Lan, *et al*. Multi-party human-computer interaction dialogue psychology model based on stackelberg game[J]. *Journal of Electronics & Information Technology*, 2023, 45(5): 1758–1765. doi: 10.11999/JEIT220441.

[2] 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, Vancouver, Canada, 2017: 634–642. doi: 10.18653/v1/P17-1059.

[3] ZHOU Hao, HUANG Minlie, ZHANG Tianyang, *et al*. Emotional chatting machine: Emotional conversation generation with internal and external memory[C]. Proceedings of the 32nd AAAI Conference on Artificial Intelligence, New Orleans, USA, 2018: 730–739. doi: 10.1609/AAAI.V32I1.11325.

[4] WEI Wei, LIU Jiayi, MAO Xianling, *et al*. Emotion-aware chat machine: Automatic emotional response generation for human-like emotional interaction[C]. Proceedings of the 28th ACM International Conference on Information and Knowledge Management. Beijing, China, 2019: 1401–1410. doi: 10.1145/3357384.3357937.

[5] LIU Siyang, ZHENG Chujie, 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. doi: 10.18653/V1/2021.ACL-LONG.269.

[6] 车万翔, 窦志成, 冯岩松, 等. 大模型时代的自然语言处理: 挑战、机遇与发展[J]. 中国科学: 信息科学, 2023, 53(9): 1645–1687. doi: 10.1360/SSI-2023-0113.
CHE Wanxiang, DOU Zhicheng, FENG Yansong, *et al*. Towards a comprehensive understanding of the impact of large language models on natural language processing: Challenges, opportunities and future directions[J]. *SCIENTIA SINICA Informationis*, 2023, 53(9): 1645–1687. doi: 10.1360/SSI-2023-0113.

[7] 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, Florence, Italy, 2019: 5370–5381. doi: 10.18653/V1/P19-1534.

[8] LIN Zhaojiang, MADOTTO A, SHIN J, *et al*. MoEL: Mixture of empathetic listeners[C]. Proceedings of the 2019 Conference on Empirical Methods in Natural Language

Processing and the 9th International Joint Conference on Natural Language Processing, Hong Kong, China, 2019: 121–132. doi: 10.18653/V1/D19-1012.

[9] MAJUMDER N, HONG Pengfei, PENG Shanshan, *et al*. MIME: MIMicking emotions for empathetic response generation[C]. Proceedings of 2020 Conference on Empirical Methods in Natural Language Processing, 2020: 8968–8979. doi: 10.18653/V1/2020.EMNLP-MAIN.721.

[10] HESS U and FISCHER A. Emotional mimicry: Why and when we mimic emotions[J]. *Social and Personality Psychology Compass*, 2014, 8(2): 45–57. doi: 10.1111/spc3.12083.

[11] RADFORD A, WU J, CHILD R, *et al*. Language models are unsupervised multitask learners[J]. *OpenAI Blog*, 2019, 1(8): 9.

[12] ZANDIE R and MAHOOR M H. Emptransfo: A multi-head transformer architecture for creating empathetic dialog systems[C]. Proceedings of the Thirty-Third International Florida Artificial Intelligence Research Society Conference, North Miami Beach, USA, 2020: 276–281.

[13] RADFORD A, NARASIMHAN K, SALIMANS T, *et al*. Improving language understanding by generative pre-training[J]. 2018.

[14] YU Dian and YU Zhou. MIDAS: A dialog act annotation scheme for open domain HumanMachine spoken conversations[C]. Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics, 2021: 1103–1120. doi: 10.18653/V1/2021.EACL-MAIN.94.

[15] QIAN Hongjin, DOU Zhicheng, ZHU Yutao, *et al*. Learning implicit user profile for personalized retrieval-based chatbot[C]. Proceedings of the 30th ACM International Conference on Information & Knowledge Management, Queensland, Australia, 2021: 1467–1477. doi: 10.1145/3459637.3482269.

[16] ZHANG Yizhe, SUN Siqi, 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. doi: 10.18653/V1/2020.ACL-DEMOS.30.

[17] ZHANG Saizheng, DINAN E, URBANEK J A, *et al*. Personalizing dialogue agents: I have a dog, do you have pets too?[C]. Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, Melbourne, Australia, 2018: 2204–2213. doi: 10.18653/V1/P18-1205.

[18] LI Jiwei, GALLEY M, BROCKETT C, *et al*. A diversity-promoting objective function for neural conversation models[C]. Proceedings of Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, San Diego, USA, 2016: 110–119. doi: 10.18653/V1/N16-1014.

[19] PAPINENI K, ROUKOS S, WARD T, *et al*. BLEU: A method for automatic evaluation of machine translation[C]. Proceedings of the 40th Annual Meeting on Association for Computational Linguistics, Philadelphia, USA, 2022: 311–318. doi: 10.3115/1073083.1073135.

[20] LIN C Y. ROUGE: A package for automatic evaluation of summaries[C]. Proceedings of Text Summarization Branches Out, Barcelona, Spain, 2004: 74–81.

[21] 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, Long Beach, USA, 2017: 6000–6010.

[22] CAI Hua, SHEN Xuli, XU Qing, *et al*. Improving empathetic dialogue generation by dynamically infusing commonsense knowledge[C]. Proceedings of Findings of the Association for Computational Linguistics, Toronto, Canada, 2023: 7858–7873. doi: 10.18653/V1/2023.FINDINGS-ACL.498.

朱振方：男，教授，博士生导师，研究方向为信息检索、自然语言处理.

李嘉欣：女，硕士生，研究方向为自然语言处理.

徐富永：男，博士生，研究方向为数据挖掘、自然语言处理.

刘培玉：男，教授，博士生导师，研究方向为信息检索、自然语言处理.

张广渊：男，教授，研究方向为数字图像处理、模式识别.

责任编辑：陈　倩