

# Facilitating Multi-turn Emotional Support Conversation with Positive Emotion Elicitation: A Reinforcement Learning Approach

Jinfeng Zhou<sup>1,2\*</sup> Zhuang Chen<sup>1†</sup> Bo Wang<sup>2‡</sup> Minlie Huang<sup>1</sup>

<sup>1</sup>The CoAI Group, DCST, Institute for Artificial Intelligence, State Key Lab of Intelligent Technology and Systems,

<sup>1</sup>Beijing National Research Center for Information Science and Technology, Tsinghua University, Beijing 100084, China

<sup>2</sup>College of Intelligence and Computing, Tianjin University, Tianjin, China

jfzhou.mail@gmail.com, zhchen-nlp@mail.tsinghua.edu.cn, bo\_wang@tju.edu.cn, aihuang@tsinghua.edu.cn

## Abstract

Emotional support conversation (ESC) aims to provide emotional support (ES) to improve one's mental state. Existing works stay at fitting grounded responses and responding strategies (e.g., *question*), which ignore the effect on ES and lack explicit goals to guide emotional positive transition. To this end, we introduce a new paradigm to formalize multi-turn ESC as a process of positive emotion elicitation. Addressing this task requires finely adjusting the elicitation intensity in ES as the conversation progresses while maintaining conversational goals like coherence. In this paper, we propose SUPPORTER, a mixture-of-expert-based reinforcement learning model, and well design ES and dialogue coherence rewards to guide policy's learning for responding. Experiments verify the superiority of SUPPORTER in achieving positive emotion elicitation during responding while maintaining conversational goals including coherence.

## 1 Introduction

Emotional support (ES) aims to reassure a person to recover from emotional distress and improve one's mental state (Burleson, 2003). It is a manifestation of emotional intelligence in social interactions (Heaney and Israel, 2008; Atoum and Al-Shoboul, 2018). Endowing ES into social dialogue systems for building helpful and trustful agents is an emerging trend (Huang et al., 2020; Rains et al., 2020).

To achieve this goal, a typical practice is modeling empathy, which aims to perceive and understand the situation and feelings of others (Keskin, 2014). Yet, the empathetic conversation (Rashkin et al., 2019) is inherently deficient in providing ES as (1) Lack of consideration of multi-turn conversation. Just making empathetic responses in each single dialogue turn leads to ignoring the user's feedback and mental state changes in multi-turn

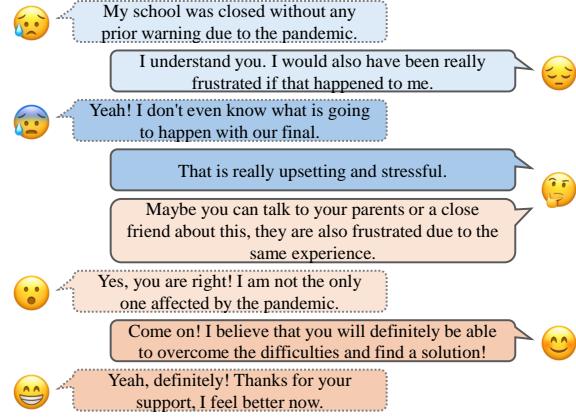


Figure 1: A simplified multi-turn ESC example between the user (*left*) and agent (*right*). The agent progressively adjusts the intensity of *empathy* and *elicitation* to achieve the goal of improving the user's mental state.

interaction. (2) Lack of awareness of emotional elicitation. Only emanating emotional resonance fails to help users jump out of negative mental states. Although Liu et al. (2021) design emotional support conversation (ESC) task promising to remedy these deficiencies, existing works (Tu et al., 2022; Cheng et al., 2022; Peng et al., 2022) stay at fitting grounded responses and responding strategies (e.g., *question*) while ignoring the effects of such efforts on ES. They do not fully model the essential working mechanism of ESC and lack explicit goals to guide a user's emotion to a positive transition in the multi-turn process. Thus, they are still insufficient to lay out an entire ESC process and cannot effectively improve one's mental state.

To this end, we introduce multi-turn ESC with positive emotion elicitation, a new paradigm aims to progressively empathize and elicit users to reach a better mental state through multi-turn conversation. Addressing this task is challenging (an example is in Figure 1): **First**, in a realistic multi-turn ESC, the user's emotions often transit towards positive (e.g., the user's emotion starts with negative and ends with positive, i.e., “*My school was closed*”

\* Work done during internship at the CoAI Group.

† Equal contribution.

‡ Corresponding author.

→ “*I feel better now*”) with fluctuation (e.g., the user’s negative emotions in the first two turns gradually deepen, i.e., “*My school was closed*” → “*I don’t even know*”), which requires the agent to equip with the mechanism dealing with complex situations to respond satisfactorily (Shibata et al., 2014; Yoshino and Kawahara, 2015). **Second**, for ES, the ES response requires a delicate balance between empathy and elicitation. Only empathizing without eliciting falls into a negative emotional cycle, while the opposite setting brings a sense of distance in communication. They need to be progressively and purposefully adjusted in ongoing interactions, e.g., the agent expresses empathy of varying emotional polarity (*negative* → *negative* → *positive*) and carefully increase the intensity of elicitation (*only empathy* → *weak elicitation* → *strong elicitation*). **Third**, for language expression, the ES response purposefully elicits positive emotions but should not undermine general conversational goals like coherence. Making an eliciting response that is out of the dialogue context, e.g., replacing “*I understand you. I would ... happened to me.*” with “*Come on! I believe ... find a solution!*”, may cause users to resent and block useful feedback.

In this paper, we propose **SUPPORTER**<sup>1</sup> to facilitate multi-turn emotional **SUPPORT** conversation with positive emotion **Elicitation** using a mixture-of-expert(MoE) based Reinforcement learning(RL). MoE designs heuristic experts associated with specific tasks to learn diverse semantics by characterizing dialogue context, where: (1) To cope with the user’s emotional fluctuation in the ongoing conversation, experts are devised as positive and negative experts as a whole; (2) To inspire ES of responding, the emotion experts of MoE are designed to predict the user’s emotional states that are possibly transited to; (3) To inspire the expression of responding, the keyword experts of MoE are designed to predict the keywords that maintain the dialogue coherence. With experts as candidates, our RL agent learns conversational semantic encoding policy and purposefully selects experts with expert selection policy for response generation. To achieve the goal of positive emotion elicitation during responding while maintaining conversational goals like coherence, we optimize policy by carefully constructing the rewards: (1) ES rewards consider the conversation progress to dynamically adjust the elicitation

intensity of positive emotion; (2) Dialogue coherence rewards involve keyword-level and sentence-level guides to finely maintain coherence.

Our contributions are summarized as follows:

(1) We introduce a new paradigm by carefully dissecting the challenges of formalizing multi-turn ESC as a process of positive emotion elicitation.

(2) We propose SUPPORTER, an MoE-based RL model with carefully constructed ES and dialogue coherence rewards, elicits positive emotion during responding while maintaining dialogue coherence.

(3) Extensive experiments show the superiority of SUPPORTER with automatic, interactive human, and novel ES and dialogue coherence evaluations.

## 2 Related Work

**Empathetic Conversation** To construct a warm dialogue system, a milestone is to endow it with empathy (Rashkin et al., 2019). Considering affective empathy (Lin et al., 2019; Majumder et al., 2020; Li et al., 2020, 2022), i.e., perceiving the user’s emotion, and cognitive empathy (Zheng et al., 2021; Sabour et al., 2022; Zhou et al., 2022), i.e., understanding the user’s situation, puts the psychological theory of empathy into practice. Limited by focusing on a single-turn empathy and lack of emotional induction, it is difficult to achieve the higher goal of improving the user’s mental state due to failure to help one jump out of the negative situation.

**Emotional Support Conversation** To remedy above deficiencies, Liu et al. (2021) design ESC for providing ES in interactions. Our work is related to existing works on ESC but differs in task definition as we focus on enhancing the elicitation effect of positive emotion of responses instead of responding strategy prediction (e.g., *question*) and grounded response generation. Although fusing knowledge (Tu et al., 2022; Peng et al., 2022) and planning strategy (Cheng et al., 2022) are beneficial for word-overlap metrics (e.g., *Bleu*), we argue whether the gains serve to ES is opaque and less convincing due to lacking corresponding evaluation mechanisms.

**Positive Emotion Elicitation Conversation** To free users from emotional distress and advance the conversation towards an optimistic state, positive emotion elicitation is an intuitive solution (Mishara et al., 2007; Jiang et al., 2021). Previous works (Hasegawa et al., 2013; Lubis et al., 2018, 2019a,b) posit the emotional elicitation process as an ideal single-turn dialogue with linear emotional changes

---

<sup>1</sup>The project repository is available at <https://github.com/jfzhouyoo/Supporter>

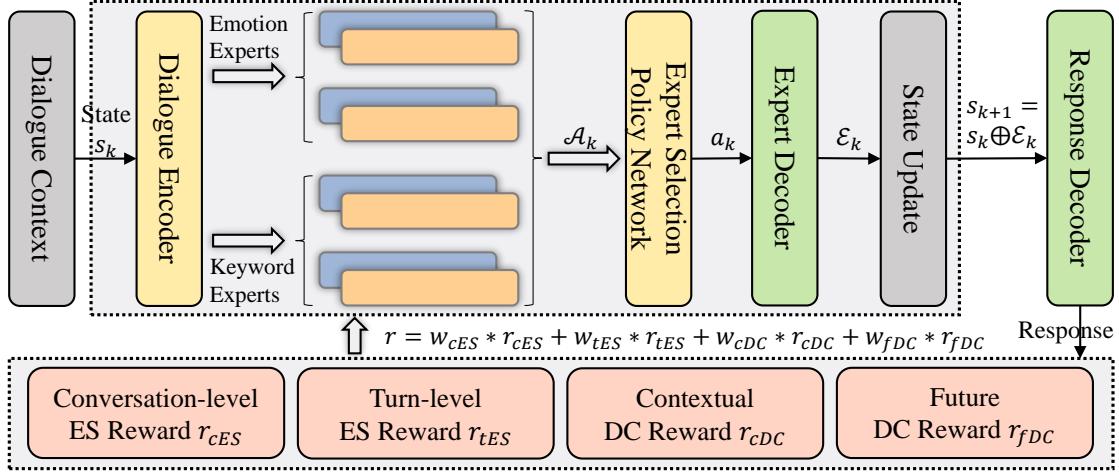


Figure 2: The architecture of the proposed SUPPORTER model. *DC* is an abbreviation for *Dialogue Coherence*.

(Wang et al., 2022). However, realistic scenarios often involve multi-turn interactions with complex emotional fluctuations. To weaken the previous strong hypothesis, we extend positive emotion elicitation to ESC by well defining challenges, and take it as a real-world application of the solution.

### 3 Preliminaries

At the  $t$ -th turn of dialogue, given dialogue context  $C_t = \{x_1, y_1, \dots, x_{t-1}, y_{t-1}, x_t\}$ , our goal is to generate the response  $y_t$  which serves to improve the user’s mental state. To equip this ability, the response generation process should achieve specific goals related to ES and language expression.

**ES for Positive Emotion Elicitation** Providing effective elicitation during multi-turn ESC suffers from two issues: First, the elicitation intensity of positive emotion needs to be adjusted progressively as the conversation progresses. Maintaining weak elicitation (e.g., “*I understand you*”) or strong elicitation (e.g., “*Come on*”) may fail to shake one’s mental state. Second, the elicitation effect of positive emotion needs to be indirectly verified by the feedback from the user’s next turn utterance. It means the elicitation intensity should consider the future fluctuation of the user’s emotional states. In this work, we construct conversation-level and turn-level ES rewards to guide the model’s learning of elicitation policy and conduct corresponding automatic and interactive human evaluations for measuring the ES performance of responding.

**Language Expression for Dialogue Coherence** The purpose of generative processes to enhance elicitation induces two attendant issues: First, with-

out proper controls may lead to greedily pursuing the goals of elicitation while discarding the contextual coherence, e.g., “*Come on!*” with strong elicitation as a response in the context of the user continuing to express negative emotions. Second, whether the response meets the user’s expectations needs feedback from the user’s future utterance. It means maintaining coherence with future dialogue is also crucial. In this work, we construct contextual and future dialogue coherence rewards to guide the model’s learning of bi-coherent expressions and perform the automatic and interactive human evaluation of conversational goals including coherence.

### 4 Methodology

In Figure 2, our SUPPORTER takes dialogue context as input to construct state sequence, which is encoded by a dialogue encoder as the conversational semantic encoding policy. The mixture-of-expert associated with emotion and keyword prediction tasks characterize state semantics to yield action candidates of the expert selection policy, which are purposefully selected for inducing state update. We use the updated state to generate response and further optimize the policy by measuring how well the response reaches the goal of ES and dialogue coherence with the well-designed parallel rewards.

#### 4.1 Multi-task Mixture-of-Expert

As a key component of SUPPORTER, we first introduce the structure of multi-task mixture-of-expert.

**Dialogue Encoder** Following Liu et al. (2021), the dialogue encoder is implemented with BlenderBot (Roller et al., 2021). Given an input sequence  $X$ , we concatenate all input tokens and prepend

with a  $[CLS]$  token, e.g., for the dialogue context, getting  $[CLS] \oplus x_1 \oplus y_1 \dots \oplus x_{t-1}$ . The sequence is fed into the dialogue encoder to obtain the hidden state  $\mathbf{H}_X$ . We denote the sequence representation derived from  $[CLS]$  as  $\mathbf{h}_X$ .

**Emotion Experts** To track possible transitions of user’s emotional states, emotion experts are associated with contextual and future user emotion predictions. We extract  $M$  fine-grained emotional reactions for each utterance in the corpus, which are inferred from COMET (Bosselut et al., 2019) using the “*xReact*” relation. Since emotional reactions are often emotional words (e.g., *happy*, *sad*), we use VAD (Mohammad, 2018) to identify the emotional polarity of each word according to its valence as a positive or negative emotional category. The high-frequency categories are finally retained as supervised labels for the emotion prediction task.

We divide contextual emotion experts into positive and negative emotion experts, which are two MLP transforming  $\mathbf{H}_X$  into  $\mathbf{H}_{X, pos}$  and  $\mathbf{H}_{X, neg}$ :

$$\begin{aligned}\mathbf{H}_{X, pos} &= \text{MLP}_{pos}(\mathbf{H}_X), \\ \mathbf{H}_{X, neg} &= \text{MLP}_{neg}(\mathbf{H}_X).\end{aligned}\quad (1)$$

We project the  $[CLS]$  representations  $\mathbf{h}_{X, pos}$  and  $\mathbf{h}_{X, neg}$  of positive and negative experts to predict positive and negative emotion, respectively:

$$\begin{aligned}P_{pos} &= \text{softmax}(\mathbf{W}_{pos} \mathbf{h}_{X, pos}), \\ P_{neg} &= \text{softmax}(\mathbf{W}_{neg} \mathbf{h}_{X, neg}),\end{aligned}\quad (2)$$

which is supervised by the positive and negative emotions collected in the  $e_{pos}^*$  and  $e_{neg}^*$  sets of the user’s last utterance in the dialogue context using cross-entropy loss:

$$\begin{aligned}L_{pos}^{ctx-emo} &= -\frac{1}{|e_{pos}^*|} \sum_{i=1}^{|e_{pos}^*|} \log P_{pos}(e_i^*), \\ L_{neg}^{ctx-emo} &= -\frac{1}{|e_{neg}^*|} \sum_{i=1}^{|e_{neg}^*|} \log P_{neg}(e_i^*).\end{aligned}\quad (3)$$

Note that an utterance may be inferred to the emotions with different polarities due to cognitive differences (Westbrook et al., 2011; Zhou et al., 2022). For future emotion experts, we adopt the above method to get  $L_{pos}^{ftr-emo}$  and  $L_{neg}^{ftr-emo}$  losses and train them to predict the positive and negative emotions of the user’s future utterance (i.e., next turn utterance). In this way, emotion experts can learn various emotion-level features by  $L_{emo}$  loss:  $L_{emo} = L_{pos}^{ctx-emo} + L_{neg}^{ctx-emo} + L_{pos}^{ftr-emo} + L_{neg}^{ftr-emo}$ .

**Keyword Experts** To meet the need for dialogue coherence, keyword experts are associated with keyword predictions that act on maintaining coherence with contextual and future utterances. Here, a bidirectional emotion keyword graph  $\mathcal{G}$  is constructed, which is also used in coherence rewards designing (a construction example is in Appendix A). We extract the salient keywords of each utterance in the corpus as vertices using a rule-based approach (Tang et al., 2019), and employ VAD to identify the emotional polarity of each keyword. The pointwise mutual information (PMI) (Church and Hanks, 1989) is adopted to construct bidirectional edges by characterizing the association between keyword pairs, where the *forward* edge depicts the keyword pairs extracted from the context and response, and the *backward* edge depicts the ones are from the future utterance and response. We further construct *positive* edges to describe the keywords with positive tail vertices, and *negative* edges are negative ones. Finally, each head vertex selects the tail vertices with the top PMI scores for building connections. The vertices of  $\mathcal{G}$  serve as supervised labels for the keyword prediction task.

Contextual keyword experts are transformed similarly to emotion experts, and their  $[CLS]$  representations  $\mathbf{h}_{X, pos}^{ctx-kws}$  and  $\mathbf{h}_{X, neg}^{ctx-kws}$  can be obtained from positive and negative keyword experts  $\mathbf{H}_{X, pos}^{ctx-kws}$  and  $\mathbf{H}_{X, neg}^{ctx-kws}$ , respectively. We infer the one-hop neighbors of contextual keywords from the “*forward-positive*” and “*forward-negative*” relations respectively in  $\mathcal{G}$  to enhance the perception of the target keywords in the golden response. Specifically, we use attention (Bahdanau et al., 2015) to obtain fused embeddings  $e_{pos}^{ctx-kws}$  and  $e_{neg}^{ctx-kws}$ :

$$\begin{aligned}e_{pos}^{ctx-kws} &= \text{Attention}(\mathbf{h}_{X, pos}^{ctx-kws}, \mathbf{E}_{pos}^{ctx-kws}), \\ e_{neg}^{ctx-kws} &= \text{Attention}(\mathbf{h}_{X, neg}^{ctx-kws}, \mathbf{E}_{neg}^{ctx-kws}),\end{aligned}\quad (4)$$

where  $\mathbf{E}_{pos}^{ctx-kws}$  and  $\mathbf{E}_{neg}^{ctx-kws}$  are positive and negative neighbor embedding matrices that share parameters with the dialogue encoder. We then concatenate  $e_{pos}^{ctx-kws}$  and  $e_{neg}^{ctx-kws}$  with  $\mathbf{H}_{X, pos}^{ctx-kws}$  and  $\mathbf{H}_{X, neg}^{ctx-kws}$  respectively at the token level, and use an MLP layer to fuse them to obtain keyword-enhanced experts  $\mathbf{H}_{X, pos-kws}^{ctx-kws}$  and  $\mathbf{H}_{X, neg-kws}^{ctx-kws}$ :

$$\begin{aligned}\mathbf{H}_{X, pos-kws}^{ctx-kws}[i] &= \text{MLP}(\mathbf{H}_{X, pos}^{ctx-kws}[i] \oplus e_{pos}^{ctx-kws}) \\ \mathbf{H}_{X, neg-kws}^{ctx-kws}[i] &= \text{MLP}(\mathbf{H}_{X, neg}^{ctx-kws}[i] \oplus e_{neg}^{ctx-kws})\end{aligned}\quad (5)$$

Further, we take the positive and negative key-

words in the golden response as supervision to optimize the  $L_{pos}^{ctx-kws}$  and  $L_{neg}^{ctx-kws}$  losses adopting cross-entropy (this process can refer to above emotion prediction task). Similarly, multi-hop reasoning on  $\mathcal{G}$ , i.e., “forward → forward → backward-positive” and “forward → forward → backward-negative” (clarified in Appendix A), is performed to obtain keywords coherent with the future utterance. Taking the positive and negative keywords in future utterance as the prediction target, the keyword-enhanced future keyword experts can be optimized by  $L_{pos}^{ftr-kws}$  and  $L_{neg}^{ftr-kws}$  losses. In this way, keyword experts can learn various expression-level features by  $L_{kws}$  loss:  $L_{kws} = L_{pos}^{ctx-kws} + L_{neg}^{ctx-kws} + L_{pos}^{ftr-kws} + L_{neg}^{ftr-kws}$ .

**Multi-task Training** To make the experts retain the primitive semantics without hindering their respective diversity, we give them a minor constraint. Specifically, we average the representations of emotion and keyword experts to get  $\mathbf{h}_{X,exp}$ , and make it close to sequence representation  $\mathbf{h}_X$  by optimizing the MSE loss with a minor hyperparameter  $\alpha$ :

$$L_{mse} = \frac{\alpha}{d_h} \sum_{i=1}^{d_h} (\mathbf{h}_X[i] - \mathbf{h}_{X,exp}[i])^2, \quad (6)$$

where  $d_h$  is the dimension of  $\mathbf{h}_X$ . Then, we jointly train the multi-task MoE by optimizing  $L_{exp}$  loss:

$$L_{exp} = L_{emo} + L_{kws} + L_{mse}. \quad (7)$$

## 4.2 MoE-based Reinforcement Learning

We use the standard reinforcement learning framework (Sutton and Barto, 2018) as the backbone.

**State** We concatenate the dialogue context and the extracted keywords as the initial state  $s_1 \in \mathcal{S}$ , i.e.,  $s_1 = \{C, C_{kws}\}$  (we omit the subscript  $t$  of dialogue context  $C_t$  for simplicity). At each step, the prompt token sequence  $\mathcal{E}$  generated by the policy determined expert (i.e., action) triggers an update of the state. We record the observed state  $s_k \in \mathcal{S}$  at  $k$ -th step, i.e.,  $s_k = \{C, \mathcal{E}_1, \dots, \mathcal{E}_{k-1}\}$ , which is encoded by the dialogue encoder to get  $\mathbf{H}_{S,k}$  and  $\mathbf{h}_{S,k}$ . We concatenate sequence representations of historical states to obtain current state embedding  $s_k = \mathbf{h}_{S,1} \oplus \dots \oplus \mathbf{h}_{S,k}$ . If  $k$  is smaller than the set maximum iteration steps  $K$ , we pad  $s_k$  with zeros for fixing dimension. Note that when  $k > 1$ , we discard the keywords  $C_{kws}$  because: (1) It has already acted on the first iteration; (2) The input sequence length is limited due to the constraint of the pre-trained model (i.e., BlenderBot).

**Action** The action space  $\mathcal{A}_k$  at  $k$ -th step is defined as the multi-task associated experts transformed by state  $s_k$ . At state  $s_k$ , our agent learns to choose an expert in  $\mathcal{A}_k$  as expert action  $a_k$ . We utilize a BlenderBot-based dialogue decoder to generate expert prompt  $\mathcal{E}_k$  of  $a_k$ .

**Policy** Besides the above dialogue encoder as the semantic encoding policy network, we design an expert selection policy network using REINFORCE with baseline (Sutton and Barto, 2018) that includes an actor network and a value network. Actor learns an expert finding policy  $\pi_\varphi(a_k, s_k, \mathcal{A}_k)$  which selects the appropriate expert action  $a_k$  based on the current state  $s_k$  and action space  $\mathcal{A}_k$  by emitting the probability distribution of actions in  $\mathcal{A}_k$ . The value network measures the value  $Q_\delta(s_k)$  of state  $s_k$  as the baseline in REINFORCE. Their network structures are defined as:

$$\begin{aligned} \mathbf{o}_k &= \eta((\eta(s_k \mathbf{W}_1) \mathbf{W}_2)), \\ \pi_\varphi(a_k, s_k, \mathcal{A}_k) &= \phi(\mathbf{A}_k \odot \mathbf{o}_k \mathbf{W}_\varphi), \\ Q_\delta(s_k) &= \mathbf{o}_k \mathbf{W}_\delta, \end{aligned} \quad (8)$$

where  $\eta(\cdot)$  is an ELU activation function with a dropout layer,  $\odot$  is the hadamard product,  $\phi(\cdot)$  is the softmax function.  $\mathbf{A}_k$  is a binarized vector for pruning the action space, and we set it as a full-one vector due to the small number of experts.

**Rewards** To guide policy learning, we reward the decision made at each step by measuring how well the response generated from updated state  $s_{k+1}$  provides ES and maintains dialogue coherence.

(1) Conversation-level ES Reward: aims to dynamically adjust the elicitation intensity of positive emotion as the conversation progresses defined as:

$$\begin{aligned} PED_{cES} &= f_{ES}(y) - f_{ES}(c_t), \\ r_{cES} &= \sum_{t=1}^T \cos\left(\frac{\pi}{2} \cdot \frac{t}{MT}\right) \cdot PED_{cES}. \end{aligned} \quad (9)$$

Here,  $f_{ES}(\cdot)$  measures the positive emotion level of an utterance using the emotion classification model developed by Hartmann (2022). The model is trained on six datasets containing diverse text types and achieves 66% accuracy for emotion classification. Positive emotion scores are collected as positive level. We encourage the positive emotion distance  $PED_{cES}$  of the generated response  $y$  and the contextual user’s post  $c_t$ : (a) is non-negative, i.e., expressing empathy (equal to 0) or elicitation (greater than 0) is the underlying requirement; (b)

synchronously increases with the dialogue turn  $t$ , i.e., the early stage of the conversation is dominated by empathy, and the latter is elicitation.  $MT$  is the maximum turn of conversation,  $T$  is current turn.

(2) Turn-level ES Reward: aims to capture the feedback of user's next turn emotion defined as:

$$PED_{tES} = |f_{ES}(y) - f_{ES}(c_f)|, \\ r_{tES} = \cos\left(\frac{\pi}{2} \cdot \frac{T}{MT}\right) \cdot \cos\left(\frac{\pi}{2} \cdot PED_{tES}\right). \quad (10)$$

Here,  $PED_{tES}$  measures the relative positive emotion distance between the generated response  $y$  and the user's future (i.e., next turn) utterance  $c_f$ . We encourage  $PED_{tES}$  to get smaller with the approaching of current turn  $T$  to  $MT$ , i.e., supervising smooth elicitation in the latter stage and improving tolerance to emotional fluctuations.

(3) Contextual Dialogue Coherence Reward: aims to constrain generated response  $y$  to maintain coherence with context  $C$  by measuring their coherence at keyword-level and sentence-level. First, we reconstruct a dataset (Liu et al., 2021) containing coherent and incoherent context-response pairs, where the response of the incoherent pairs is an utterance randomly sampled from the dataset. Next, a BERT-based (Devlin et al., 2019) text classification model  $f_{cDC}$  is trained by feeding sentence-keyword pairs and achieves 85% accuracy. We take the coherence probability as the coherence score, the reward is defined as:

$$r_{cDC} = f_{cDC}(C \oplus C_{kws}, y \oplus y_{kws}) \cdot e^{\frac{N_{c,kws}}{|y_{kws}|} - 1}, \quad (11)$$

where  $y_{kws}$  is the keyword set of  $y$  and  $N_{c,kws}$  is the number of keywords in  $y_{kws}$  that are the *forward* neighbors of contextual keywords in  $\mathcal{G}$ .

(4) Future Dialogue Coherence Reward: aims to introduce the consideration of coherence with the user's future utterance  $c_f$ . Similarly, we reconstruct a dataset (Liu et al., 2021) containing coherent and incoherent future utterance-response pairs and train another text classification model  $f_{fDC}$  which achieves 77% accuracy. The reward is defined as:

$$r_{fDC} = f_{fDC}(c_f \oplus c_{f_{kws}}, y \oplus y_{kws}) \cdot e^{\frac{N_{f,kws}}{|y_{kws}|} - 1}, \quad (12)$$

where  $N_{f,kws}$  is the number of keywords in  $y_{kws}$  that have a *backward* relation with keywords  $c_{f_{kws}}$  of  $c_f$  in  $\mathcal{G}$ .

(5) Total reward. The total reward is  $r = w_{cES} * r_{cES} + w_{tES} * r_{tES} + w_{cDC} * r_{cDC} + w_{fDC} * r_{fDC}$ .

Corpus Info.	#Dialogues	1,053
	#Utterances	31,410
	Avg. length of dialogues	29.8
	Avg. length of utterances	17.8
Graph $\mathcal{G}$ Info.	#Split Ratio	8:1:1
	#Keywords	2,433
	Avg. forward neighbors	21.24
	Avg. backward neighbors	21.17
	Avg. positive neighbors	33.94
	Avg. negative neighbors	8.46

Table 1: Statistics of our dataset after preprocessing.

### 4.3 Optimization

We set  $K$ -step iterations, and the goal of agent learning is to maximize the expected cumulative reward:  $J_\theta = \mathbb{E}_\pi \left[ \sum_{k=1}^K \gamma^k r_{k+1} \right]$ , where  $\theta$  is the learned parameter and  $\gamma$  is the discount coefficient. The agent is optimized by  $L_{agent}$  loss and its policy gradient is defined as:

$$\nabla_\theta J_\theta = \mathbb{E}_\pi [\nabla_\theta \log \pi_\varphi(a_k, s_k, \mathcal{A}_k)(G - Q_\delta(s_k))], \quad (13)$$

where  $G$  is the discounted cumulative reward from the initial state to the terminal state. Finally, we take the hidden state  $\mathbf{H}_{S,K+1}$  of the state  $s_{K+1}$  to generate the response, where the decoder is optimized by  $L_{gen}$  loss:

$$L_{gen} = - \sum_{m=1}^M \log P(y_m | \mathbf{H}_{S,K+1}, y_{<m}). \quad (14)$$

**Warm Start** We use the pretrained small version of BenderBot for initializing our model. The initial state is used as input to fine-tune the model for warm start by optimizing  $L_{warm} = L_{exp} + L_{gen}$ .

**Joint Training** Our model is finally jointly trained by optimizing  $L_{joint}$  loss:

$$L_{joint} = L_{agent} + L_{gen} + \frac{1}{K+1} \sum_{k=1}^{K+1} L_{exp,k} \quad (15)$$

## 5 Experiments

### 5.1 Experimental Setup

**Dataset** Our experiments are conducted on the widely used ESConv (Liu et al., 2021), a multi-turn conversation dataset for ES. In a conversation, the user confides personal negative situation, and the supporter provides comfort and support to improve the user's mental state. The statistics of ESConv and graph  $\mathcal{G}$  after preprocessing are in Table 1.

Models	PPL↓	B-1↑	B-2↑	B-3↑	D-1↑	D-2↑	D-3↑	cES↑	tES↑	cDC↑	fDC↑	Len
MoEL	112.34	18.14	6.77	3.22	2.43	17.03	38.08	0.658	0.390	0.391	0.384	20.36
MIME	68.49	15.89	6.58	3.27	2.02	10.51	22.60	0.598	0.370	0.450	0.412	19.44
BlenderBot-Joint	<b>14.78</b>	17.97	7.17	3.31	4.56	24.65	49.71	0.611	0.398	0.710	0.459	17.69
MISC	16.16	-	7.31	-	4.41	19.71	-	-	-	-	-	-
GLHG	15.67	19.66	7.57	3.74	3.50	21.61	-	-	-	-	-	-
Bart-Joint	16.05	<b>19.99</b>	<b>7.92</b>	<b>3.93</b>	4.24	21.98	43.33	0.635	0.402	<b>0.723</b>	<b>0.475</b>	18.85
<b>SUPPORTER</b>	15.37	19.50	7.49	3.58	<b>4.93</b>	<b>27.73</b>	<b>53.78</b>	<b>0.743</b>	<b>0.409</b>	0.681	0.472	18.37
w/o EmoExperts	15.35	18.32	7.12	3.38	4.79	27.20	53.01	0.711	0.392	0.679	0.460	18.14
w/o KwsExperts	15.54	17.76	6.74	3.19	4.69	26.16	50.92	0.728	0.394	0.636	0.443	17.72
w/o Multi-Task	15.49	16.79	6.54	3.18	4.78	27.17	53.45	0.651	0.399	0.651	0.450	16.48
w/o ESRewards	15.46	18.49	7.10	3.36	4.69	26.92	52.49	0.664	0.391	0.660	0.457	18.41
w/o DCRewards	15.43	17.28	6.80	3.25	4.80	27.45	53.04	0.707	0.401	0.652	0.448	17.12
w/o ExpertPolicy	15.54	18.30	7.23	3.54	4.75	27.23	52.85	0.683	0.395	0.657	0.454	18.54
Warm-Start Only	15.03	17.42	6.74	3.21	4.67	26.24	51.82	0.629	0.402	0.644	0.444	17.35
w/o Warm-Start	15.01	17.98	6.86	3.18	4.55	26.06	51.62	0.673	0.403	0.638	0.453	18.26

Table 2: Automatic evaluation results. “Len” indicates the average length of the generated responses.

**Baselines** (1) *MoEL* (Lin et al., 2019): An empathetic conversation model that uses multiple decoders to capture possible user emotions for generating. (2) *MIME* (Majumder et al., 2020): An empathetic conversation model that mimics user’s emotions during responding. (3) *BlenderBot-Joint* (Liu et al., 2021): An ESC model that prepends a predicted strategy token on the backbone of BlenderBot. (4) *MISC* (Tu et al., 2022): An ESC model that fuses commonsense. (5) *GLHG* (Peng et al., 2022): A commonsense-based ESC model that designs a global-to-local graph. (6) We design *Bart-Joint* by replacing the backbone of BlenderBot-Joint with Bart (Lewis et al., 2020). It achieves comparable performance to *MultiESC* (Cheng et al., 2022) as its replacement since MultiESC’s code is unavailable.

**Implementation Details** We implement all models with Pytorch, and all pretrained models (i.e., BlenderBot, Bart) use small versions. We set the number of steps  $K = 2$  and reward weights  $w_{cES} = w_{cDC} = 0.1, w_{tES} = w_{fDC} = 1.0$  (selected using a grid-search approach with two values {0.1, 1.0} for each hyperparameter). We extract  $M = 10$  emotional reactions for each utterance. The maximum number of conversation turn  $MT$  is set to 10. The discount factor  $\gamma$  is 0.99, the hyperparameter  $\alpha$  is 1e-5, and the batch size is 16. We use Adam optimizer (Kingma and Ba, 2015) with an initial learning rate of 2e-5 and a linear warmup of 120 steps for training on a GPU-V100 machine. The warm start stage is trained for 5 epochs, and the joint training stage is set to 3 epochs. The decoding settings are consistent with Liu et al. (2021). For a fair comparison, all baselines with available codes are reproduced under the same setting.

## 5.2 Automatic Evaluation

We adopt Perplexity (PPL), Bleu (B-n) and Distinct (D-n) to evaluate the general generation quality and diversity of the models. To measure how well the generated responses achieve goals, we define (1) ES scores containing conversation-level (*cES*) and turn-level (*tES*), i.e.,  $r_{cES}$  and  $r_{tES}$ , measure the elicitation intensity of positive emotion involving conversation progress and the perceived intensity to the user’s next turn emotion; (2) Dialogue coherence scores containing contextual (*cDC*) and future (*fDC*), i.e.,  $r_{cDC}$  and  $r_{fDC}$ , measure the coherence with the context and the user’s future utterance.

**Overall Performance** In Table 2, compared with all baselines, our SUPPORTER achieves the most diverse expressions and highest ES (12.9% outperforms the second best MoEL on *cES*) while maintaining competitive dialogue quality (*PPL*, *Bleu*) and coherence (*cDC*, *fDC*). Supportive responses generated by MoEL are often accompanied by low diversity and low coherence due to the retelling of generic responses (e.g., “*I am glad I could help you*” with high positive emotion) that are found from its outputs. Bart-based models benefit from robust sequence modeling (Lewis et al., 2020) with inherent advantages in coherence and Bleu but perform poorly in ES and diversity. The contextual coherence (*cDC*) of our SUPPORTER is inferior to BlenderBot-Joint, which is acceptable as ES for positive emotion elicitation needs to sacrifice a little coherence to jump out of negative topics.

**Ablation Study** In Table 2: First, we remove the emotion experts (w/o EmoExperts), keyword experts (w/o KwsExperts), and the multi-task as-

SUPPORTER vs.	BlenderBot-Joint			Bart-Joint			w/o EmoExperts			w/o ExpertPolicy		
	Win	Lose	Tie									
Fluency	<b>67.5<sup>‡</sup></b>	23.7	8.8	<b>66.5<sup>‡</sup></b>	26.5	7.0	<b>44.5<sup>†</sup></b>	40.0	15.5	<b>42.9<sup>†</sup></b>	37.5	19.6
Informativeness	<b>55.2<sup>‡</sup></b>	40.7	4.1	<b>56.7<sup>‡</sup></b>	38.8	4.5	<b>48.6<sup>‡</sup></b>	36.8	14.6	<b>38.5</b>	35.9	25.6
Coherence	<b>53.8<sup>‡</sup></b>	31.8	14.4	<b>45.4</b>	43.8	10.8	<b>53.7<sup>‡</sup></b>	35.7	10.6	<b>55.1<sup>‡</sup></b>	32.4	12.5
Supportiveness	<b>59.2<sup>‡</sup></b>	34.1	6.7	<b>51.4<sup>‡</sup></b>	37.6	11.0	<b>54.5<sup>‡</sup></b>	33.4	12.1	<b>51.4<sup>‡</sup></b>	34.3	14.3
Overall	<b>56.5<sup>‡</sup></b>	30.4	13.1	<b>48.6<sup>‡</sup></b>	37.1	14.3	<b>50.0<sup>‡</sup></b>	34.3	15.7	<b>49.6<sup>‡</sup></b>	32.1	18.3

Table 3: Results of interactive human evaluation (%).  $\dagger/\ddagger$  denote  $p$ -value  $< 0.1/0.05$  (statistical significance test).

sociated with the experts (w/o Multi-Task), respectively. Emotion experts mainly act on ES, including *cES* and *tES*. Keyword experts contribute significantly to dialogue coherence, including *cDC* and *fDC*. Multi-task training endows experts with specific abilities and thus has an impressive impact on overall performance. **Second**, we remove the ES rewards (w/o ESREwards) and dialogue coherence rewards (w/o DCRewards), respectively. The former improves positive support, and the latter maintains grounded expression. Therefore, besides achieving their own goals, they also benefit dialogue diversity and quality, respectively. Moreover, we replace the expert selection policy network with random sampling (w/o ExpertPolicy). Random experts lead to uncertainty in decision-making and thus damage overall performance, especially on ES and coherence. **Third**, we test using only warm start and without joint training (Warm-Start Only) as well as without warm start and only joint training (w/o Warm-Start). The former reaches comparable or even worse results than the baselines, and the latter greedily achieves the goal of maximizing the rewards resulting in low dialogue quality.

### 5.3 Interactive Human Evaluation

We recruited three crowdsourcing workers and exposed them to 100 negative situations randomly sampled from the test set. They were asked to engage in multi-turn conversation with the models to simulate the process of seeking ES and to choose the better one (Win) from a model pair by considering five aspects, respectively: (1) Fluency: which bot’s response is more fluent and understandable? (2) Informativeness: which bot’s response is more diverse and specific, and contains more information? (3) Coherence: which bot’s response is more coherent with context in a multi-turn conversation? (4) Supportiveness: which bot provides more effective ES, i.e., is more likely to elicit users to change their emotions from negative to positive? (5) Overall: generally, which bot is more preferred?

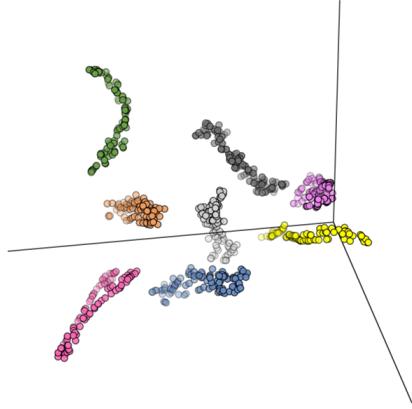


Figure 3: Latent space visualization of experts. Separate clusters show MoE has diverse and specific semantics.

As in Table 3, from the comparison with baselines, we found that a single incoherent response (*cDC* in Table 2) has less impact on the coherence of the overall multi-turn conversation. Comparisons with variants of SUPPORTER demonstrate that key components of our model, i.e., emotion experts and expert selection policy, lead to significant advantages in the overall performance.

### 5.4 Qualitative Analysis

**Specificity of Experts** To analyze the quality of the experts, we show the specificity of the experts learned by SUPPORTER. As shown in Figure 3, we visualize the latent space of experts using t-SNE on 200 conversation samples. The latent space distributions of multi-task-associated experts are clearly separated and clustered in specific regions. Some overlap is also intuitive due to the similarity between experts with the same polarity, e.g., contextual and future positive emotion experts. This verifies our MoE has diverse and specific semantics and the superiority of multi-task learning.

**Adjustability of Elicitation** To further explore the adjustability of elicitation intensity of positive emotion in multi-turn conversation, we analyze the trend of positive emotion distance with the dialogue

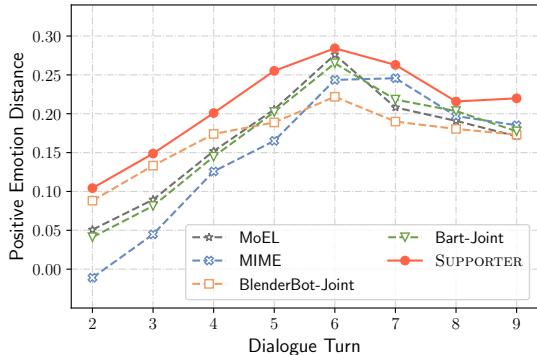


Figure 4: SUPPORTER progressively enhances the elicitation intensity and exhibits robust adjustment ability in the later stage of the conversation.

Models	D-1	B-2	$cES$	$tES$	$cDC$	$fDC$
$SUPPORTER_{K=1}$	4.40	7.55	0.801	0.382	0.668	0.466
$SUPPORTER_{K=2}$	4.93	7.49	0.743	0.409	0.681	0.472
$SUPPORTER_{K=3}$	5.22	6.71	0.699	0.405	0.657	0.459
$SUPPORTER_{K=4}$	5.05	6.10	0.673	0.413	0.594	0.431

Table 4: Parameter analysis for iteration steps  $K$ . SUPPORTER outperforms the best baselines in most settings.

turns, i.e.,  $PED = f_{ES}(y) - \frac{1}{T} \sum_{t=1}^T f_{ES}(c_t)$ . As shown in Figure 4, the PED score of all models tends to rise first and then fall. In the early stage of the conversation ( $turn < 6$ ), SUPPORTER keeps the same trend as the empathy model (i.e., MoEL, MIME) and gradually increases the intensity of elicitation. This is attributed to our encouragement that it should progressively transform the conversation from empathy-dominated to elicitation-dominated. In the later stage of the conversation ( $turn > 6$ ), SUPPORTER still maintains a higher level of elicitation than baselines and shows robust adjustment ability.

### 5.5 Parameter Analysis

We further analyze the impact of the number of iteration steps  $K$ . In Table 4, with the increase of steps, diversity and  $tES$  show an upward trend, while other metrics show a downward one. This happens possibly because the informativeness of the generated responses increases with selected experts, making it possible to lose focus and thus lead to poor dialogue quality. Furthermore, SUPPORTER outperforms the best baselines in most cases, confirming its effectiveness.

## 6 Conclusions

In this paper, we introduce a new paradigm to formalize multi-turn ESC as a process of positive emotion elicitation and propose an MoE-based rein-

forcement learning model SUPPORTER with well-designed ES and dialogue coherence rewards. Extensive experiments verify the superiority of our model in providing effective ES for positive emotion elicitation while maintaining conversational goals including coherence. Our work will facilitate future work to develop ESC with positive emotion elicitation for improving the users’ mental state.

## Limitations

We discuss three limitations of this work as follows.

The first one is the instability of reinforcement learning. Reward-driven policy learning is an essential advantage of this work because it is better equipped with the positive emotion-driven process of ESC than existing works and can model flexible ESC expression beyond the training data. However, this flexibility also suffers from instability, which calls for additional knowledge or strategies to refine the learning process.

The second one is the need for further reference to psychological theory. An advantage of our work is to learn posterior ESC patterns integrating the dialogue context and future feedback in the form of rewards. However, there is still other valuable prior knowledge to be referred from psychology studies, e.g., the CBT (cognitive-behavioral therapy) methods. This kind of prior knowledge can be used as additional knowledge to refine the learning process as mentioned in the first limitation.

The third one is that the reward design can be further optimized. The ideal case is to construct a high-quality dataset with human-feedback labels for training reward model (e.g., the constructed example of ChatGPT). At the same time, the larger parameter of the reward model, the more conducive it is to learn a robust policy and avoid it overfitting to the reward function. However, such optimizations need a trade-off with cost.

## Ethical Considerations

In this paper, the ESConv dataset used in our experiments is a publicly-available benchmark for emotional support conversation, which does not contain sensitive and personal information as well as unethical language. Our work builds on this dataset to study positive emotion elicitation to improve the user’s mental state. Therefore, we focus on constructing a dialogue system to provide emotional support from families and friends in the daily scenarios limited by this dataset rather than profes-

sional psychological counseling or psychological treatment. For risky non-daily scenarios such as self-harm or suicide-related conversations, we do not claim that the dialogue system we built has a treatment or improvement effect on them. Additionally, we also ensure the anonymity of our interactive human evaluation. We believe our work meets ACL’s Code of Ethics.

## Acknowledgements

This work was supported by the National Science Foundation for Distinguished Young Scholars (with No. 62125604). This work was also supported by the Guoqiang Institute of Tsinghua University, with Grant No. 2020GQG0005. This work was also supported by Tsinghua Precision Medicine Foundation.

This work was also supported by the National Natural Science Foundation of China (with No. 62272340, 61876128, 62276187).

## References

- Adnan Yousef Atoum and Rasha Ahmed Al-Shoboul. 2018. Emotional support and its relationship to emotional intelligence. *Advances in social sciences research journal*, 5(1).
- Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2015. Neural machine translation by jointly learning to align and translate. In *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*.
- Antoine Bosselut, Hannah Rashkin, Maarten Sap, Chaitanya Malaviya, Asli Celikyilmaz, and Yejin Choi. 2019. COMET: commonsense transformers for automatic knowledge graph construction. In *Proceedings of the 57th Conference of the Association for Computational Linguistics, ACL 2019, Florence, Italy, July 28- August 2, 2019, Volume 1: Long Papers*, pages 4762–4779. Association for Computational Linguistics.
- Brant R Burleson. 2003. Emotional support skills. In *Handbook of communication and social interaction skills*, pages 569–612. Routledge.
- Yi Cheng, Wenge Liu, Wenjie Li, Jiashuo Wang, Ruihui Zhao, Bang Liu, Xiaodan Liang, and Yefeng Zheng. 2022. Improving multi-turn emotional support dialogue generation with lookahead strategy planning. *CoRR*, abs/2210.04242.
- Kenneth Ward Church and Patrick Hanks. 1989. Word association norms, mutual information and lexicography. In *27th Annual Meeting of the Association for Computational Linguistics, 26-29 June 1989, University of British Columbia, Vancouver, BC, Canada, Proceedings*, pages 76–83. ACL.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019, Minneapolis, MN, USA, June 2-7, 2019, Volume 1 (Long and Short Papers)*, pages 4171–4186. Association for Computational Linguistics.
- Jochen Hartmann. 2022. Emotion english distilroberta-base. <https://huggingface.co/j-hartmann/emotion-english-distilroberta-base/>.
- Takayuki Hasegawa, Nobuhiro Kaji, Naoki Yoshinaga, and Masashi Toyoda. 2013. Predicting and eliciting addressee’s emotion in online dialogue. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics, ACL 2013, 4-9 August 2013, Sofia, Bulgaria, Volume 1: Long Papers*, pages 964–972. The Association for Computer Linguistics.
- Catherine A Heaney and Barbara A Israel. 2008. Social networks and social support. *Health behavior and health education: Theory, research, and practice*, 4:189–210.
- Minlie Huang, Xiaoyan Zhu, and Jianfeng Gao. 2020. Challenges in building intelligent open-domain dialog systems. *ACM Trans. Inf. Syst.*, 38(3):21:1–21:32.
- Hao Jiang, Yutao Zhu, Xinyu Zhang, Zhicheng Dou, Pan Du, Te Pi, and Yantao Jia. 2021. Emotion eliciting machine: Emotion eliciting conversation generation based on dual generator. *CoRR*, abs/2105.08251.
- Sevgi Coşkun Keskin. 2014. From what isn’t empathy to empathic learning process. *Procedia-Social and Behavioral Sciences*, 116:4932–4938.
- Diederik P. Kingma and Jimmy Ba. 2015. Adam: A method for stochastic optimization. In *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*.
- Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Veselin Stoyanov, and Luke Zettlemoyer. 2020. BART: denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, ACL 2020, Online, July 5-10, 2020*, pages 7871–7880. Association for Computational Linguistics.
- Qintong Li, Hongshen Chen, Zhaochun Ren, Pengjie Ren, Zhaopeng Tu, and Zhumin Chen. 2020. Empdg: Multi-resolution interactive empathetic dialogue generation. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 1–11. Association for Computational Linguistics.

*Conference on Computational Linguistics, COLING 2020, Barcelona, Spain (Online), December 8-13, 2020*, pages 4454–4466. International Committee on Computational Linguistics.

Qintong Li, Piji Li, Zhaochun Ren, Pengjie Ren, and Zhumin Chen. 2022. **Knowledge bridging for empathetic dialogue generation**. In *Thirty-Sixth AAAI Conference on Artificial Intelligence, AAAI 2022, Thirty-Fourth Conference on Innovative Applications of Artificial Intelligence, IAAI 2022, The Twelveth Symposium on Educational Advances in Artificial Intelligence, EAAI 2022 Virtual Event, February 22 - March 1, 2022*, pages 10993–11001. AAAI Press.

Zhaojiang Lin, Andrea Madotto, Jamin Shin, Peng Xu, and Pascale Fung. 2019. **Moel: Mixture of empathetic listeners**. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, EMNLP-IJCNLP 2019, Hong Kong, China, November 3-7, 2019*, pages 121–132. Association for Computational Linguistics.

Siyang Liu, Chujie Zheng, Orianna Demasi, Sahand Sabour, Yu Li, Zhou Yu, Yong Jiang, and Minlie Huang. 2021. **Towards emotional support dialog systems**. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing, ACL/IJCNLP 2021, (Volume 1: Long Papers), Virtual Event, August 1-6, 2021*, pages 3469–3483. Association for Computational Linguistics.

Nurul Lubis, Sakriani Sakti, Koichiro Yoshino, and Satoshi Nakamura. 2018. **Eliciting positive emotion through affect-sensitive dialogue response generation: A neural network approach**. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, (AAAI-18), the 30th innovative Applications of Artificial Intelligence (IAAI-18), and the 8th AAAI Symposium on Educational Advances in Artificial Intelligence (EAAI-18), New Orleans, Louisiana, USA, February 2-7, 2018*, pages 5293–5300. AAAI Press.

Nurul Lubis, Sakriani Sakti, Koichiro Yoshino, and Satoshi Nakamura. 2019a. Dialogue model and response generation for emotion improvement elicitation. In *Proc. 33rd Conf. Neural Inf. Process. Syst.(NIPS)*, pages 1–11.

Nurul Lubis, Sakriani Sakti, Koichiro Yoshino, and Satoshi Nakamura. 2019b. **Positive emotion elicitation in chat-based dialogue systems**. *IEEE ACM Trans. Audio Speech Lang. Process.*, 27(4):866–877.

Navonil Majumder, Pengfei Hong, Shanshan Peng, Jiankun Lu, Deepanway Ghosal, Alexander F. Gelbukh, Rada Mihalcea, and Soujanya Poria. 2020. **MIME: mimicking emotions for empathetic response generation**. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, EMNLP 2020, Online, November 16-20, 2020*,

pages 8968–8979. Association for Computational Linguistics.

Brian L Mishara, François Chagnon, Marc Daigle, Bogdan Balan, Sylvaine Raymond, Isabelle Marcoux, Cécile Bardon, Julie K Campbell, and Alan Berman. 2007. Which helper behaviors and intervention styles are related to better short-term outcomes in telephone crisis intervention? results from a silent monitoring study of calls to the us 1-800-suicide network. *Suicide and Life-Threatening Behavior*, 37(3):308–321.

Saif M. Mohammad. 2018. **Obtaining reliable human ratings of valence, arousal, and dominance for 20, 000 english words**. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, ACL 2018, Melbourne, Australia, July 15-20, 2018, Volume 1: Long Papers*, pages 174–184. Association for Computational Linguistics.

Wei Peng, Yue Hu, Luxi Xing, Yuqiang Xie, Yajing Sun, and Yunpeng Li. 2022. **Control globally, understand locally: A global-to-local hierarchical graph network for emotional support conversation**. In *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI 2022, Vienna, Austria, 23-29 July 2022*, pages 4324–4330. ijcai.org.

Stephen A Rains, Corey A Pavlich, Bethany Lutovsky, Eric Tsetsi, and Anjali Ashtaputre. 2020. Support seeker expectations, support message quality, and supportive interaction processes and outcomes: The case of the comforting computer program revisited. *Journal of Social and Personal Relationships*, 37(2):647–666.

Hannah Rashkin, Eric Michael Smith, Margaret Li, and Y-Lan Boureau. 2019. **Towards empathetic open-domain conversation models: A new benchmark and dataset**. In *Proceedings of the 57th Conference of the Association for Computational Linguistics, ACL 2019, Florence, Italy, July 28- August 2, 2019, Volume 1: Long Papers*, pages 5370–5381. Association for Computational Linguistics.

Stephen Roller, Emily Dinan, Naman Goyal, Da Ju, Mary Williamson, Yinhan Liu, Jing Xu, Myle Ott, Eric Michael Smith, Y-Lan Boureau, and Jason Weston. 2021. **Recipes for building an open-domain chatbot**. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume, EACL 2021, Online, April 19 - 23, 2021*, pages 300–325. Association for Computational Linguistics.

Sahand Sabour, Chujie Zheng, and Minlie Huang. 2022. **CEM: commonsense-aware empathetic response generation**. In *Thirty-Sixth AAAI Conference on Artificial Intelligence, AAAI 2022, Thirty-Fourth Conference on Innovative Applications of Artificial Intelligence, IAAI 2022, The Twelveth Symposium on Educational Advances in Artificial Intelligence, EAAI 2022 Virtual Event, February 22 - March 1, 2022*, pages 11229–11237. AAAI Press.

Tomohide Shibata, Yusuke Egashira, and Sadao Kurohashi. 2014. Chat-like conversational system based on selection of reply generating module with reinforcement learning. In *Situated Dialog in Speech-Based Human-Computer Interaction, 5th International Workshop on Spoken Dialogue Systems, IWSDS 2014, Napa, CA, USA, January 18-20, 2014, Signals and Communication Technology*, pages 63–69. Springer.

Richard S Sutton and Andrew G Barto. 2018. *Reinforcement learning: An introduction*. MIT press.

Jianheng Tang, Tiancheng Zhao, Chenyan Xiong, Xiaodan Liang, Eric P. Xing, and Zhiting Hu. 2019. Target-guided open-domain conversation. In *Proceedings of the 57th Conference of the Association for Computational Linguistics, ACL 2019, Florence, Italy, July 28- August 2, 2019, Volume 1: Long Papers*, pages 5624–5634. Association for Computational Linguistics.

Quan Tu, Yanran Li, Jianwei Cui, Bin Wang, Ji-Rong Wen, and Rui Yan. 2022. MISC: A mixed strategy-aware model integrating COMET for emotional support conversation. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, ACL 2022, Dublin, Ireland, May 22-27, 2022, pages 308–319. Association for Computational Linguistics.

Shihang Wang, Xinchao Xu, Wenquan Wu, Zheng-Yu Niu, Hua Wu, and Haifeng Wang. 2022. Towards multi-turn empathetic dialogs with positive emotion elicitation. *CoRR*, abs/2204.10509.

David Westbrook, Helen Kennerley, and Joan Kirk. 2011. *An introduction to cognitive behaviour therapy: Skills and applications*. Sage.

Koichiro Yoshino and Tatsuya Kawahara. 2015. Conversational system for information navigation based on POMDP with user focus tracking. *Comput. Speech Lang.*, 34(1):275–291.

Chujie Zheng, Yong Liu, Wei Chen, Yongcai Leng, and Minlie Huang. 2021. Comae: A multi-factor hierarchical framework for empathetic response generation. In *Findings of the Association for Computational Linguistics: ACL/IJCNLP 2021, Online Event, August 1-6, 2021*, volume ACL/IJCNLP 2021 of *Findings of ACL*, pages 813–824. Association for Computational Linguistics.

Jinfeng Zhou, Chujie Zheng, Bo Wang, Zheng Zhang, and Minlie Huang. 2022. CASE: aligning coarse-to-fine cognition and affection for empathetic response generation. *CoRR*, abs/2208.08845.

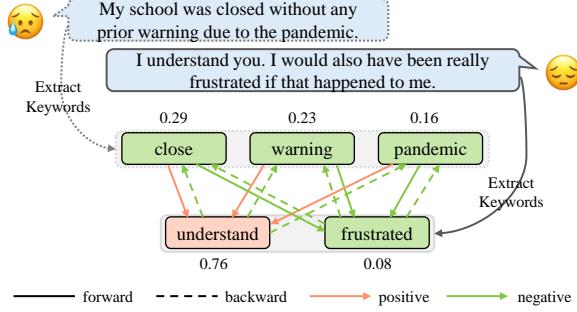


Figure 5: A construction example of the bidirectional emotion keyword graph  $\mathcal{G}$ . The valence (i.e., the number in the figure) is used to identify **positive** and **negative** keywords.

## A Bidirectional Emotion Keyword Graph

A construction example of the bidirectional emotion keyword graph  $\mathcal{G}$  is in Figure 5.

**One-hop Reasoning on Graph  $\mathcal{G}$**  For the contextual keyword “*close*”, its one-hop neighbor reasoned by the “*forward-positive*” relation is “*understand*”, and the one reasoned by the “*forward-negative*” relation is “*frustrated*”. Further, the one-hop neighbors reasoned by the “*forward*” relation are the union of the one-hop neighbors of the above two relations, i.e., “*understand*” and “*frustrated*”. For the keyword “*frustrated*” of the response, it cannot reason the one-hop neighbor using the “*backward-positive*” relation. Therefore, its one-hop neighbors reasoned by the “*backward*” relation are the same as the one-hop neighbors reasoned by the “*backward-negative*” relation, i.e., “*close*”, “*warning*”, and “*pandemic*”.

**Multi-hop Reasoning on Graph  $\mathcal{G}$**  Taking the “*forward*  $\rightarrow$  *forward*  $\rightarrow$  *backward-positive*” multi-hop reasoning as an example, using the “*forward*” relationship for the contextual keywords to perform one-hop reasoning can obtain the set of neighbors that contain the keywords of the response, which we regard as the extended keyword set of the response determined by the context. Using the keywords in this set as a starting point to perform the second-hop reasoning by the “*forward*” relation can result in the expanded keyword set of the user’s future utterance (i.e., the user’s next turn utterance) determined by the response. Further, similarly, the third-hop reasoning is performed using the “*backward-positive*” relation to determine the extended positive keywords set of the response coherent to the future utterance.

## B Case Study

In Table 5, two cases in multi-turn conversation generated by five models are selected. We found that the advantage of SUPPORTER is that it is more likely to elicit users’ emotions to transit positively by adaptively adjusting the intensity of empathy and elicitation. For example, low turns have a tone of strong empathy (e.g., “*that is so awful*”, “*you are struggling financially and that you have not been able to afford your kids presents*”). As the conversation deepens, the elicitation intensity of positive emotion progressively increases (e.g., “*do you know anyone who has gone through similar situations?*”, “*it is okay to be proud of yourself for being able to make them*”), and finally the expression with strong elicitation will help users jump out of negative situations (e.g., “*I am sure you will be successful in the new job market*”, “*you can really enjoy seeing them happy*”). During these multi-turn conversations, SUPPORTER effectively provides ES with diverse expressions while maintaining the dialogue coherence, being consistent with the observations from automatic and interactive human evaluations.

Situation#1	I have been laid out for not performing. But the actual situation is the customer was satisfied my service. But the expectation of the management was not to do that.
Context (1 <sup>st</sup> -2 <sup>nd</sup> turn)	<p>User: Hi!</p> <p>Bot: Hi can I help you today? I am here to listen. Do you have anything happen recently that has you feeling badly?</p> <p>User: I am laid off.</p>
MoEL	I am sorry to hear that. What is going on with your thoughts?
MIME	I am sorry to hear that. I am sorry to hear that. I am so sorry to hear that.
BlenderBot-Joint	I am sorry to hear that. Is there something you would like to talk about?
Bart-Joint	I have been laid off before. I was laid off too.
<b>SUPPORTER</b>	<b>Oh no, that is so awful. I am so sorry to hear that. I know that feeling well, especially when you have this situation.</b>
Ground-Truth	Oh I am really sorry to hear that, did you have the same job for a long time? That sounds very difficult to deal with.
Context (3 <sup>rd</sup> -5 <sup>th</sup> turn)	<p>User: I was attending a customer. He was having a grievance which I sorted out. But the management did not like that. For the past 10 years I have been with this job.</p> <p>/*.....*/</p> <p>Bot: That is really unfair and hard to deal with are you close to any family?</p> <p>User: I am not in any close to any family related to job.</p>
MoEL	I think that it is great you have a great job!
MIME	It sounds like you are trying to get your job that you are not alone.
BlenderBot-Joint	I can tell you are really feeling badly about this.
Bart-Joint	I know that is hard to deal with. I am glad that you are reaching out to me to help out!
<b>SUPPORTER</b>	<b>That is a really difficult position to be in, do you know anyone who has gone through similar situations?</b>
Ground-Truth	Do you have any close friends to talk to about any new job prospects?
Context (6 <sup>th</sup> -9 <sup>th</sup> turn)	<p>User: I have few friends, I have been talking to them. They also tell I was not any wrong. For the experience and qualification, I will be getting a better job.</p> <p>/*.....*/</p> <p>Bot: Oh, wow then you do have a lot of contacts and some support or at least understanding. Music is a great way to relax and that is very positive in your life and current situation.</p> <p>User: I am so confident I will be having a better job within the next few months.</p>
MoEL	I have found that online too. I do not know if I have heard of my work. I found out online that I could talk to you. I would suggest looking for a new job.
MIME	I am glad to hear that! I bet you could try to get some breathing techniques.
BlenderBot-Joint	It is very difficult to think about it right now but I would definitely encourage you to take some deep breaths and take some time to reflect on your progress.
Bart-Joint	I am so happy that you are feeling better and have found some new ways to relax.
<b>SUPPORTER</b>	<b>I am glad that I could be of some help to you. It sounds like you are a very strong and intelligent person. I am sure you will be successful in the new job market.</b>
Ground-Truth	That is a great outlook. Thank you for sharing today.
Situation#2	I don't have enough money to buy my kids Christmas gifts.
Context (1 <sup>st</sup> -3 <sup>rd</sup> turn)	<p>User: Hi!</p> <p>Bot: How are things going?</p> <p>User: Not good. I am depressed.</p> <p>Bot: Is there anything in particular that has been bothering you recently?</p> <p>User: I do not have enough money to buy my 5 kids X-mas gifts.</p>
MoEL	I am sorry to hear that. I am sorry you are having this right now.
MIME	I understand how you are feeling. I know how hard it must be.
BlenderBot-Joint	So you are feeling depressed because of covid and not enough money to buy gifts?
Bart-Joint	I am sorry to hear that. I have had to give up my job to help my kids during the holidays.
<b>SUPPORTER</b>	<b>It seems to me that you are struggling financially and that you have not been able to afford your kids presents.</b>
Ground-Truth	I am really sorry to hear that, but I am sure your love for your kids can still shine through.
Context (4 <sup>th</sup> -5 <sup>th</sup> turn)	<p>User: I do love my kids, but they are expecting some toys and stuff from Santa.</p> <p>Bot: Would it be possible to look at some more home made stuff for this year that will still be fun and meaningful?</p> <p>User: I can not craft. If I made them something it would be so bad.</p>
MoEL	I am so sorry to hear that. I am sorry you are having to go through this.
MIME	I understand how you feel. I have been there too. I know how hard it is to the next year. I am sure there are some options out there.
BlenderBot-Joint	Have you considered doing crafts with your kids?
Bart-Joint	Have you considered getting some more craft supplies? Maybe something that you can use to make a list of things you would like to get your kids for Christmas?
<b>SUPPORTER</b>	<b>I know how hard it can be to make things. I know it is a terrible feeling to feel bad about yourself, but it is okay to be proud of yourself for being able to make them.</b>
Ground-Truth	I know it can be really hard to feel like you can not make anything worthwhile, but I am sure there's something you could come up with that could still put a smile on your kids' faces.
Context (6 <sup>th</sup> -7 <sup>th</sup> turn)	<p>User: Do you have any craft ideas?</p> <p>Bot: Maybe you could look at involving your kids in the process. The feeling of doing something together as a family can make it way more fun. It could be a shared experience your family can look back on fondly later on.</p> <p>User: That is a good idea. I would like to have it done before Christmas. How about some wooden toys or a cardboard box fort?</p>
MoEL	I am so glad to hear that. I hope that you can do something fun!
MIME	I think you can do it! I have been through this, and you are not alone?
BlenderBot-Joint	That sounds like a great idea! Maybe even something for them to look forward to after Christmas. I know I love the idea of making toys and toys!
Bart-Joint	Yes, I think you should try to get it done before Christmas, but maybe even after that. You can see how hard it is to find a good time to do something.
<b>SUPPORTER</b>	<b>I think that would be a great idea! That way you can feel like your kids are being cared for and you can really enjoy seeing them happy.</b>
Ground-Truth	I have seen some people make small ornaments filled with candies and treats and other small goodies. So even if all of the gifts feel small, they can feel like they are getting a lot of them. And you could even get them to have fun decorating and painting the ornaments!

Table 5: Cases generated from baselines and SUPPORTER. /\*.....\*/ indicates that some turns of dialogue are omitted.