CAB: Empathetic Dialogue Generation with Cognition, Affection and Behavior

Pan Gao[1] , Donghong Han[1,2]($\boxtimes$), Rui Zhou[3], Xuejiao Zhang[1],
and Zikun Wang[1]

[1] School of Computer Science and Engineering, Northeastern University,
Shenyang, China
handonghong@cse.neu.edu.cn
[2] Key Laboratory of Intelligent Computing in Medical Image of Ministry
of Education, Northeastern University, Shenyang, China
[3] Swinburne University of Technology, Hawthorn, Australia
rzhou@swin.edu.au

Abstract. Empathy is an important characteristic to be considered when building a more intelligent and humanized dialogue agent. However, existing methods did not fully comprehend empathy as a complex process involving three aspects: cognition, affection and behavior. In this paper, we propose CAB, a novel framework that takes a comprehensive perspective of cognition, affection and behavior to generate empathetic responses. For cognition, we build paths between critical keywords in the dialogue by leveraging external knowledge. This is because keywords in a dialogue are the core of sentences. Building the logic relationship between keywords, which is overlooked by the majority of existing works, can improve the understanding of keywords and contextual logic, thus enhance the cognitive ability. For affection, we capture the emotional dependencies with dual latent variables that contain both interlocutors' emotions. The reason is that considering both interlocutors' emotions simultaneously helps to learn the emotional dependencies. For behavior, we use appropriate dialogue acts to guide the dialogue generation to enhance the empathy expression. Extensive experiments demonstrate that our multi-perspective model outperforms the state-of-the-art models.

Keywords: Empathetic dialogue · Dialogue generation · Cognition affection and behavior

1 Introduction

Empathy is the ability to understand others' feelings, and respond appropriately to their situations . Previous studies have shown that empathetic dialogue models can improve user's satisfaction in several areas, such as customer service [14], healthcare community [26] and etc. Therefore, how to successfully implement empathy becomes one of the key issues to build an intelligent and considerate agent. In recent years, many studies have been conducted on the task of
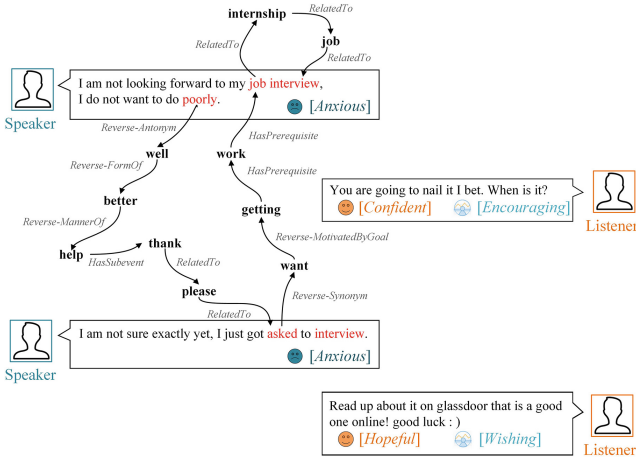
Fig. 1. A dialogue from the *EmpatheticDialogues* dataset. The cognitive ability is improved by retrieving entities (bold in black) and relationships (grey) from *ConceptNet* and building paths between critical keywords (red) to generate a high quality response under the influence of *anxious* and *confident* emotions and *wishing* dialogue act. (Color figure online)

empathetic dialogue generation, which are mainly divided into two categories: One is to enhance the understanding of a user's situation and emotion by leveraging knowledge from one or more external knowledge bases [11,15,21,25] or adding emotion causes as prior emotion knowledge [3,25]. This is to improve the cognitive ability. The issue of the existing work is that they overlook the importance of paths between users' critical keywords, which can actually reflect the contextual logic in the conversation. Although some studies [25] build paths between emotion concepts and cause concepts, they mainly focus on the causality aspect and ignore the fact that paths between any keywords can help. The second category is to design emotion strategies, such as mixture of experts [12], emotion mimicry [17] and multi-resolution emotions [10] to generate appropriate responses from the affection aspect. Unfortunately, these studies learn to respond properly mainly according to the speaker's emotion rather than both interlocutors' emotions. In this paper, we aim to improve the aforementioned weak aspects of the existing works to help advance the study of empathetic dialogue generation.

Psychological research shows that empathy is a complex mental process involving three aspects of interlocutors: cognition, affection and behavior [13]. Specifically, cognitive empathy refers to the ability to understand and interpret a user's situation [2]; affective empathy is an emotional reaction based on differentiating the emotions of oneself and others [13]; behavioral empathy means verbal or non-verbal forms of communication used in the empathetic dialogue [6]. Among the existing works, some only consider the aspects of congition and affection [21,28]; others mainly consider the aspect of behavior [1,27]. None of the

existing works had comprehensively considered all the three aspects (cognition, affection, behavior), which we believe are all important. In the following, we elaborate in detail with the example in Fig. 1. The dialogue in Fig. 1 shows that (1) **Cognition**: The speaker is *anxious* about attending a job interview. In the first turn, there exists a path between $<job, interview>$ with internship as a bridge to enhance the understanding of the keywords and the context. In the next turn, the paths between $<poorly, asked>$ and $<asked, job>$ are built to alleviate the problem that it is difficult to capture the contextual logic based on limited context. Thus, it can be seen that the paths , which establish the relationships between utterances, are critical to improve the cognitive ability. (2) **Affection**: In interpersonal conversations, responses are usually influenced by both interlocutors' emotions [5]. As shown in Fig. 1, in the second turn, instead of both sides falling into anxiety, the listener is able to perceive the speaker's emotion and accept the emotion difference between them, thus generating a response with more positive emotion (*hopeful*). Therefore, how to learn the emotional dependencies between the context and target response based on both interlocutors' emotions is critical for responding properly. (3) **Behavior**: Appropriate dialogue acts are used as communicative form to enhance empathy expression. For example, the listener inspires the speaker by *encouraging* and makes the speaker relaxed by *wishing*. Different from [27], we consider that all the responses (rather than some of them) are generated by the guiding of dialog act. In this way, we can guide dialogue generation better.

To this end, we propose a novel empathetic dialogue generation model including aspects of **C**ognition, **A**ffection and **B**ehavior (**CAB**) to achieve a comprehensive empathetic dialogue task. Specifically, since keywords are important to understand the contextual logic, our model builds paths between critical keywords through multi-hop commonsense reasoning to enhance the cognitive ability. Conditional Variational Auto Encoder (CVAE) model with dual latent variables is built based on both interlocutors' emotions, and then the dual latent variables are injected into the decoder together with the dialogue act features to produce empathetic responses from the perspective of affection and behavior. Our contributions are summarized as follows:

- To the best of our knowledge, we are the first to propose a novel framework for empathetic dialogue generation based on psychological theory from three perspectives: cognition, affection and behavior.
- We propose a context-based multi-hop reasoning method, in which paths are established between critical keywords to acquire implicit knowledge and learn contextual logic.
- We present a novel CVAE model, which introduces dual latent variables to learn the emotional dependencies between the context and target responses. After that, we incorporate the dialogue act features into the decoder to guide the generation.
- Experiments demonstrate that CAB generates more relevant and empathetic responses compared with the state-of-the-art methods.[1]

---

[1] Code and data are available at https://github.com/geri-emp/CAB.

2   Related Work

Recently, there has been numerous works in the task of empathetic dialogue generation proposed by Rashkin et al. [20]. Lin et al. [12] assign different decoders for various emotions, and fuse the output of each decoder with users' emotion weights. Majumder et al. [17] adopt emotion stochastic sampling and emotion mimicry to respond to positive or negative emotions for generating empathetic responses. Li et al. [10] construct an interactive adversarial learning network considering multi-resolution emotions and user feedback. Liu et al. [16] incorporate anticipated emotions into response generation via reinforcement learning. Gao et al. [3] adopt emotion cause to better understand the user's emotion. However, all of the above methods only consider the user's emotion and ignore the influence between both interlocutors' emotions in the dialogue.

Several studies have incorporated external knowledge into empathetic dialogue generation. Li et al. [11] employ multi-type knowledge to explore implicit information and construct an emotional context graph to improve emotional perception. Liu et al. [15] prepend the retrieved knowledge triples to the gold responses in order to get proper responses. However, these approaches retrieve knowledge triples without fully considering the contextual meaning of the words. Although Wang et al. [25] adopt *ConceptNet* to explore the emotional causality by commonsense reasoning between the emotion clause and the cause clause, the logical relationships between other utterances may be ignored. Sabour et al. [21] use *ATOMIC* for commonsense reasoning to better understand the user's situation and feeling, but reasoning on a whole dialogue history may neglect the important role of keywords in the context. To overcome the above proposed shortcomings, we propose a context-based multi-hop commonsense reasoning method to enrich contextual information and reason about the logical relationships between utterances.

3   Method

3.1   Task Formulation and Overview

In empathetic dialogue generation, each dialogue consists of a dialogue history $C = [S_1, L_1, S_2, L_2, \ldots, S_{N-1}, L_{N-1}, S_N]$ of 2N-1 utterances and a gold empathetic response $L_N = [w_N^1, w_N^2, \ldots, w_N^n]$ of n words, where S_i and L_i denote the i-th utterance of speaker and listener respectively. Our goal is to generate an empathetic response $R = [r_1, r_2, \ldots, r_m]$ based on the dialogue history C, the speaker's emotion e_s, the listener's emotion e_l, and the listener's dialogue act a_l.

We provide a overview of CAB in Fig. 2, which consists of five components: (a) **Emotional Context Representation**. The predicted emotions, e_s and e_l, are fed into context C by emotional context encoder to obtain the emotional context representation $\hat{\boldsymbol{H}}_S$ and $\hat{\boldsymbol{H}}_L$; (b) **Affection**. Then prior network and posterior network capture dual latent variables $\boldsymbol{z}_s$ and $\boldsymbol{z}_l$, based on $\hat{\boldsymbol{H}}_S$ and $\hat{\boldsymbol{H}}_L$ in the test and training phase; (c) **Cognition**. To build paths P, we leverage *ConceptNet* to acquire external knowledge and incorporate it into C to obtain a
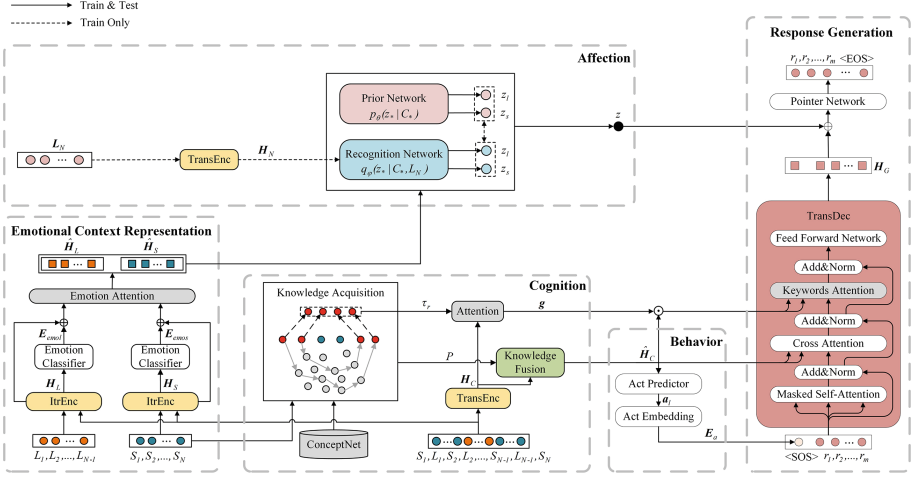
Fig. 2. The overall architecture of CAB.

knowledge-enhanced context representation $\hat{\boldsymbol{H}}_C$; **(d) Behavior**. The dialogue act features $\boldsymbol{E}_a$ are distilled based on a predictor and the embedding layer; **(e) Response Generation**. The three-stage decoder generates an empathetic response R based on the aspects of affection, cognition and behavior.

We evaluate the model on *EmpatheticDialogues* [20], which is a publicly available benchmark dataset for empathetic dialogue generation. However, dialogues in this dataset do not contain labels of emotion and dialogue act for each listener's utterance, and we annotate emotion and dialogue act by Emoberta [7] and EmoBERT [27], respectively, to support the studies in this paper.

From Sect. 3.2 to Sect. 3.7, we introduce CAB briefly due to space limit. More model and experiment details are in the full version [4].

3.2 Emotional Context Encoder

Input Representation. We divide the dialogue history into two segments $C_S = [S_1, S_2, \ldots, S_N]$ and $C_L = [L_1, L_2, \ldots, L_{N-1}]$. Following the previous work [12], we first gain the embedding of speaker context, listener context, global context and gold response respectively. Then the embedding of speaker context and listener context are fed into the Transformer-based inter-encoder (**ItrEnc**) to obtain $\boldsymbol{H}_S$ and $\boldsymbol{H}_L$, and the Transformer encoder (**TransEnc**) encodes the embedding of global context and gold response into $\boldsymbol{H}_C$ and $\boldsymbol{H}_N$.

Emotion Classification. To understand the emotions of the speaker and the listener, we project the hidden representations of the first token from $\boldsymbol{H}_S$ and $\boldsymbol{H}_L$ into the emotion category distribution P_s and P_l to predict their emotions. Then we send the emotions to a trainable emotion embedding layer to obtain the emotion states embedding matrix $\boldsymbol{E}_{emos}$ and $\boldsymbol{E}_{emol}$.

Emotion Self-attention. To make the latent variables in Sect. 3.3 incorporate both interlocutors' emotions, $\boldsymbol{H}_S$ and $\boldsymbol{H}_L$ are concatenated with $\boldsymbol{E}_{emos}$ and $\boldsymbol{E}_{emol}$ and then fed into a self-attention layer followed by a linear layer to obtain the emotional context representation $\hat{\boldsymbol{H}}_S$ and $\hat{\boldsymbol{H}}_L$.

3.3   Prior Network and Recognition Network (Affection)

We introduce dual latent variables $\boldsymbol{z}_* \in \{\boldsymbol{z}_s, \boldsymbol{z}_l\}$ in CVAE, mapping the input sequences $C_* \in \{C_S, C_L\}$ into the output sequence L_N via $\boldsymbol{z}_*$. Taking speaker as an example, we illustrate how to realize the prior network and the recognition network. The **prior network** $p_\theta(\boldsymbol{z}_s | C_S)$ is parameterized by 3-layer MLPs to compute the mean μ_s' and variance $\sigma_s'^2$ of $\boldsymbol{z}_s$. The network structure of the **recognition network** $q_\varphi(\boldsymbol{z}_s | C_S, L_N)$ is the same as that of the prior network, except that the input also includes $\boldsymbol{H}_N$. In order to learn the emotional dependencies based on both interlocutors' emotions, we fuse $\boldsymbol{z}_s$ and $\boldsymbol{z}_l$ due to the emotional similarity coefficient β between $\boldsymbol{E}_{emos}$ and $\boldsymbol{E}_{emol}$ to obtain $\boldsymbol{z} = \beta \cdot \boldsymbol{z}_s + (1-\beta) \cdot \boldsymbol{z}_l$.

3.4   Knowledge Acquisition and Fusion (Congnition)

Knowledge Acquisition. We first obtain the keyword set τ_{all} of size $\boldsymbol{cw}$ from C_S based on the TextRank algorithm [18]. Then we build paths as follows:

   a. Take one keyword in τ_{all} as the head entity $h_i \in \tau_{all}$, then feed the embedding of h_i and speaker context into ItrEnc to extract the semantic features of h_i. The Top-K knowledge triples in *ConceptNet* associated with h_i are retrieved based on a score and removed relation set [11].

   b. To ensure that the triples are logically related to other keywords τ_{other}, we first obtain the semantic features of $h_j \in \tau_{other}$ like step a. After ranking the triples by relevance between tail entity and h_j, we select Top-k triples. If the tail entity is same as h_j, which indicates there exists a one-hop path between h_i and h_j, we add them to the final keywords set τ_r (e.g. red circles in Fig. 2). If not, the tail entity is added to τ_{all} to continue finding the paths by repeating step a and b. Finally, we retain some paths P (e.g. the paths connected by grey arrows in Fig. 2) for futher fusion. The attention weight vector $\boldsymbol{g}$ is calculated to measure importance of each word in C with τ_r by the attention mechanism.

Knowledge Fusion. We first convert the paths into sequences. Then the sequences are fed into the two-layer Bi-GRU to obtain the knowledge representation $\boldsymbol{H}_k$. Finally, following previous work [21], we concatenate $\boldsymbol{H}_k$ with context at token-level to learn the knowledge-enhanced context representation $\hat{\boldsymbol{H}}_C$.

3.5   Dialogue Act Predictor and Representation (Behavior)

To guide the communicative form of empathetic dialogue generation, our model uses the first token of $\hat{\boldsymbol{H}}_C$ to predict dialogue act $\boldsymbol{a}_l$. Then, $\boldsymbol{a}_l$ is fed into the embedding layer to learn the dialogue act embedding representation $\boldsymbol{E}_a$.

3.6   Response Generation

Finally, the aforementioned information $\boldsymbol{E}_a$, $\boldsymbol{g}$, $\boldsymbol{z}$ and $\hat{\boldsymbol{H}}_C$ are applied at the Transformer-based decoder (**TransDec**) through the following three stages: **(1)** The embedding of the start-of-sequence token $\boldsymbol{E}_{SOS}$ and $\boldsymbol{E}_a$ are fed into a linear layer, then the high-level act features are adopted to guide the generation. **(2)** We design a multi-head keywords attention, which takes the output of the cross-attention layer as query, the dot-product over $\boldsymbol{g}$ and $\hat{\boldsymbol{H}}_C$ as key and value. Then TransDec outputs the hidden state $\boldsymbol{H}_G$. **(3)** To learn the emotional dependencies, we concatenate $\boldsymbol{z}$ and $\boldsymbol{H}_G$ at token-level and use pointer network [23] to output the probability distribution of each word in the vocabulary.

3.7   Training Objectives

We jointly optimiaze the emotion classification loss, dialogue act prediction loss, the loss of CVAE model and bag-of-word loss as:

$$\mathcal{L} = \gamma_1 \mathcal{L}_s + \gamma_2 \mathcal{L}_l + \gamma_3 \mathcal{L}_a + \gamma_4 \mathcal{L}(C_*, C_N; \theta, \varphi) + \gamma_5 \mathcal{L}_{bow} \tag{1}$$

where γ_1, γ_2, γ_3, γ_4 and γ_5 are hyper-parameters.

4   Experiments

4.1   Experimental Setup

Baselines. We compare our model with the state-of-the-art models as follows: (1) **Transformer** [22]: The vanilla Transformer with the pointer network trained by optimizing the generation loss. (2) **Multi-Trans** [20]: A variant of Transformer that includes emotion classification loss in addition to the generation loss to jointly optimize the model. (3) **MOEL** [12]: A model that includes several Transformer decoders, and the outputs are softly combined to generate responses. (4) **MIME** [17]: A model adopting emotion mimicry and emotion clusters to deal with positive or negative emotions. (5) **EmpDG** [10]: A generative adversarial network that considers multi-resolution emotion and introduces discriminators to supervise the training in semantics and emotion. (6) **KEMP** [11]: A model that uses two-type knowledge to help understand and express emotions. (7) **CEM** [21]: A method for generating empathetic responses by leveraging commonsense to improve the understanding of interlocutors' situations and feelings.

Implementation Details. We implement all models in PyTorch[2] with GeForce GTX 3090 GPU, and train models using Adam optimization [8] with a mini-batch size of 16. All common hyper-parameters are the same as the work in [12]. We adopt 300-dimensional pre-trained 840B GloVE vectors [19] to initialize the word embeddings, which are shared between the encoders and the decoder. The hidden size is 300 everywhere, and the size of latent variable is 200. We use the

---

[2] https://pytorch.org/.

Table 1. Results of the automatic evaluation, and w/o Cog/Aff/Beh indicate ablation experiments and the best results of all models are bold.

Models	PPL	DIST-1	DIST-2	EmoSA	EmoLA	ActA
Transformer	34.11	0.49	1.91	-	-	-
Multi-Trans	36.42	0.43	1.85	28.91	-	-
MOEL	36.59	0.60	3.12	32.33	-	-
MIME	37.52	0.32	1.22	34.88	-	-
EmpDG)	37.37	0.45	1.89	32.45	-	-
KEMP	36.39	0.66	3.08	36.57	-	-
CEM	36.11	0.66	2.99	39.07	-	-
CAB	34.36	**1.13**	**4.23**	**40.52**	**72.23**	41.72
W/o Cog	**33.88**	0.94	3.33	39.42	71.82	**43.09**
W/o Aff	34.98	1.12	3.97	34.25	-	37.25
W/o Beh	34.79	1.06	3.83	40.05	72.20	-

KL annealing of 15,000 batches to achieve the best performance. During test, the batch size is 1 and the maximum greedy decoding steps is 50.

Automatic Evaluation Metrics. We choose the widely used PPL [24], Distinct-1, Distinct-2 [9] as our main automatic metrics. **PPL** is used to estimate the generation quality of a model in general. **Distinct-1** and **Distinct-2** are used to measure the diversity of responses. Since emotion accuracy of speaker/listener (**EmoSA**/**EmoLA**) reflects the understanding of both interlocutors' emotions and dialogue act accuracy (**ActA**) can determine whether the proper dialogue acts are chosen to produce responses, we also report these metrics.

4.2   Results and Analysis

Automatic Evaluation Results. The overall automatic evaluation results are shown in the Table 1. Our model CAB outperforms the baselines on all metrics significantly. The lower PPL score implies that CAB has a higher quality of generation generally, reflecting the importance of considering empathy from multi-perspective. The remarkable improvements in Distinct-1 and Distinct-2 suggest that the introduction of external knowledge can be beneficial in improving the understanding of dialogue history and thus generating a wider variety of response. The higher accuracy of emotion classification verifies the validity of modelling both interlocutors' emotions separately.

Ablation Study. As shown in the bottom part of Table 1, we also conduct ablation experiments to explore the effect of each component. From the results, we can observe that all metrics decrease except for PPL, especially Distinct-1 and Distinct-2, when commonsense knowledge acquisition and fusion are removed (**w/o Cog**), suggesting that the paths capture additional information to enhance

cognitive ability, thus improving the quality and diversity of responses. The increasing PPL score may be due to the introduction of knowledge, which may have an impact on the fluency of the generated responses. In addition, we find that only considering the speaker's emotion by removing the latent variable of listener (**w/o Aff**) yields lower emotion accuracy and higher PPL score, and thus it is difficult to generate appropriate responses without understanding both interlocutors' emotions exactly. All metrics decrease when we remove the classification of dialogue act and the dialogue act features fused at the decoder (**w/o Beh**), indicating the emphasis of the dialogue acts in improving empathy.

5   Conclusions

In this paper, we build paths by leveraging commonsense knowledge to enhance understanding of the user's situation, considering both interlocutors' emotions and guiding responses generation through dialogue act, namely by generating empathetic responses from three perspectives: cognition, affection and behavior. Extensive experiments based on benchmark metrics have shown that our method CAB outperforms the state-of-the-art methods, demonstrating the effectiveness of our method in improving empathy of the generated responses.

References

1. Chen, M.Y., Li, S., Yang, Y.: EmpHi: generating Empathetic Responses with Human-like Intents. In: NAACL-HLT 2022, pp. 1063–1074 (2022)
2. Elliott, R., Bohart, A.C., Watson, J.C., Murphy, D.: Therapist empathy and client outcome: An updated meta-analysis. Psychotherapy **55**(4), 399–410 (2018)
3. Gao, J., et al.: Improving Empathetic Response Generation by Recognizing Emotion Cause in Conversations. In: EMNLP 2021, pp. 807–819 (2021)
4. Gao, P., Han, D., Zhou, R., Zhang, X., Wang, Z.: CAB: empathetic dialogue generation with cognition, affection and behavior. arXiv preprint arXiv:2302.01935 (2023)
5. Ghosal, D., Majumder, N., Poria, S., Chhaya, N., Gelbukh, A.F.: DialogueGCN: a graph convolutional neural network for emotion recognition. In: EMNLP-IJCNLP, pp. 154–164 (2019)
6. Gladstein, G.A.: Empathy and counseling outcome: an empirical and conceptual review. Couns. Psychol. **6**(4), 70–79 (1977)
7. Kim, T., Vossen, P.: Emoberta: speaker-aware emotion recognition in conversation with roberta. (2021) arXiv preprint arXiv:2108.12009
8. Kingma, D.P., Ba, J.: Adam: a Method for Stochastic Optimization. In: 3rd International Conference on Learning Representations (2019)
9. Li, J., Galley, M., Brockett, C., Gao, J., Dolan, B.: A Diversity-Promoting Objective Function for Neural Conversation Models. In: NAACL HLT 2016, pp. 110–119 (2016)

10. Li, Q., Chen, H., Ren, Z., Ren, P., Tu, Z., Chen, Z.: EmpDG: multi-resolution Interactive Empathetic Dialogue Generation. In: ICCL-20, pp. 4454–4466 (2020)
11. Li, Q., Li, P., Ren, Z., Ren, P., Chen, Z.: Knowledge Bridging for Empathetic Dialogue Generation. In: AAAI-22, pp. 10993–11001 (2022)
12. Lin, Z., Madotto, A., Shin, J., Xu, P., Fung, P.: MoEL: Mixture of Empathetic Listeners. In: EMNLP-IJCNLP, pp. 121–132 (2021)
13. Liu, C., Wang, Y., Yu, G., Wang, Y.: A review of relevant theories of empathy and exploration of new dynamic models. Adv. Psychol. Sci. (5), 9 (2009)
14. Liu, S., Zheng, C., Demasi, O., Sabour, S., Huang, M.: Towards Emotional Support Dialog Systems. In: ACL/IJCNLP(1), pp. 3469–3483 (2021)
15. Liu, Y., Maier, W., Minker, W., Ultes, S.: Empathetic dialogue generation with pre-trained roberta-gpt2 and external knowledge. (2021) arXiv preprint arXiv:2109.03004
16. Liu, Y., Du, J., Li, X., Xu, R.: Generating Empathetic Responses by Injecting Anticipated Emotion. In: ICASSP, pp. 7403–7407 (2021)
17. Majumder, N., Hong, P., Peng, S., Lu, J., Poria, S.: MIME: MIMicking Emotions for Empathetic Response Generation. In: EMNLP(1), pp. 8968–8979 (2020)
18. Mihalcea, R., Tarau, P.: TextRank: bringing Order into Text. In: EMNLP, pp. 404–411 (2004)
19. Pennington, J., Socher, R., Manning, C.D.: Glove: global Vectors for Word Representation. In: EMNLP, pp. 1532–1543 (2014)
20. Rashkin, H., Smith, E.M., Li, M., Boureau, Y.: Towards Empathetic Open-domain Conversation Models: a New Benchmark and Dataset. In: ACL 2019, pp. 5370–5381 (2019)
21. Sabour, S., Zheng, C., Huang, M.: CEM: commonsense-Aware Empathetic Response Generation. In: AAAI-22, pp. 11229–11237 (2022)
22. Vaswani, A., et al.: Attention is All you Need. In: NIPS-17, pp. 5998–6008 (2017)
23. Vinyals, O., Fortunato, M., Jaitly, N.: Pointer Networks. In: NEURIPS 2015, pp. 2692–2700 (2015)
24. Vinyals, O., Le, Q.V.: A neural conversational model. (2015) arXiv preprint-arXiv:1506.05869
25. Wang, J., Li, W., Lin, P., Mu, F.: Empathetic response generation through graph-based multi-hop reasoning on emotional causality. Knowl. Based Syst. **233**, 107547 (2021)
26. Wang, L., et al.: Cass: Towards building a social-support chatbot for online health community. Proc. ACM Hum. Comput. Interact. **5**(CSCW1), 1–31 (2021)
27. Welivita, A., Pu, P.: A Taxonomy of Empathetic Response Intents in Human Social Conversations. In: COLING 2020, pp. 4886–4899 (2020)
28. Zheng, C., Liu, Y., Chen, W., Leng, Y., Huang, M.: CoMAE: a Multi-factor Hierarchical Framework for Empathetic Response Generation. In: Findings of the Association for Computational Linguistics: ACL/IJCNLP, pp. 813–824 (2021)