Proceedings of the 2024 27th International Conference on Computer Supported Cooperative Work in Design

Multi-Level Knowledge-Enhanced Prompting for Empathetic Dialogue Generation

Ziyin Gu[1,2,†], Qingmeng Zhu[1,2,†], Hao He[2,*], Zhipeng Yu[2]
, Tianxing Lan[1,2], Shuo Yuan[3],
[1]University of Chinese Academy of Sciences, Beijing, China
[2]Institute of Software, Chinese Academy of Sciences, Beijing, China
[3] Faculty of Information Engineering and Automation, Kunming University of Science and Technology, Yunnan, China
{ziyin2020,qingmeng,hehao21,yuzhipeng,tianxing2023}@iscas.ac.cn, yuanshuo@stu.kust.edu.cn

2024 27th International Conference on Computer Supported Cooperative Work in Design (CSCWD) | 979-8-3503-4918-4/24/$31.00 ©2024 IEEE | DOI: 10.1109/CSCWD61410.2024.10580095

***Abstract*—Empathetic dialogue systems can recognize users' emotions and provide appropriate responses, which are crucial for enhancing the user experience. However, existing empathetic dialogue systems often fall short in understanding some complex implicit emotions. To address this problem, we propose a multi-level knowledge-enhanced prompting approach to achieve more effective empathetic dialogue generation effect. We first acquire topic words and emotional keywords as low-level emotional knowledge. Next, we retrieve dialogue samples that are most similar in topic and emotional attributes, forming mid-level emotional knowledge. Subsequently, we guide a large language model (LLM) to generate high-level comprehensive emotional knowledge based on the information from the previous two levels and the dialogue context. Finally, based on the emotional knowledge, we further guide LLM to generate empathetic responses. The research results indicate that our multi-level knowledge-enhanced prompting approach outperforms other baselines.**

***Index Terms*—artificial intelligence; large language model; empathetic dialogue system; knowledge-enhanced prompting**

I. Introduction

In recent years, despite the rapid development of dialogue systems based on artificial intelligence (AI) technologies [1]–[3], [34]–[39], it is still noticeable that existing dialogue systems face challenges in accurately understanding and perceiving emotions [4]. Recent efforts have sought to address this issue by developing empathetic dialogue systems [5], enabling dialogue systems to understand, perceive, and appropriately respond to the context and emotions of others [6], [7].

Current empathetic dialogue systems commonly rely on extensive external knowledge for emotional perception and empathetic expression [8], [9]. This is analogous to how humans typically depend on experience and common-sense knowledge to identify and express emotions [10]. However, existing methods still struggle to comprehensively capture all potential emotions of the speaker and respond to all emotions because some latent emotions in dialogues are implicit and complex. Existing models have limitations in understanding certain implicit emotions, making it challenging to perceive all implicit emotions and learn the crucial points required for corresponding emotional responses from limited dialogue history [8].

Q. Zhu and Z. Gu contributed equally to this work. H. He is the corresponding author.

Therefore, in this paper, we focus on addressing the limitations of current empathetic dialogue systems in handling implicit emotions. To better understand implicit emotions, models need to possess emotion perception and expression capabilities at a level similar to that of humans. Considering that the human empathetic expression process typically follows the steps of 1) understanding the topics in the sentence, 2) identifying emotional keywords in the sentence, 3) determining the emotional aspects of the sentence based on common-sense, and 4) selecting appropriate emotions for empathetic responses, we propose a multi-level knowledge-enhanced prompting approach to simulate a similar process for LLM, thus generating more effective empathetic dialogues. Specifically, we first utilize a topic model to obtain topic words (similar to "step 1)" in the human empathetic expression process). We then employ the SHAP algorithm [35] to extract emotional keywords from the sentence (similar to "step 2)" in the human empathetic expression process). The combination of topic words and emotional keywords forms the low-level emotional knowledge. Next, we retrieve dialogue samples that have topic and emotional attributes most similar to the obtained knowledge and the dialogue context, creating mid-level emotional knowledge (analogous to the common-sense reasoning humans undergo for "step 3)"). Subsequently, we guide a large language model (LLM) to generate advanced comprehensive emotional knowledge based on the information from the first two levels and the dialogue context (similar to "step 3)" in the human empathetic expression process). Finally, considering that massive language models like GPT-3 [11] may only be accessible through APIs, making fine-tuning for this task impractical, we further prompt the LLM to generate responses based on the dialogue context and newly acquired exemplars (similar to "step 4)" in the human empathetic expression process).

In conclusion, we propose a novel multi-level knowledge-enhanced prompting approach, comprising a first-stage for multi-level knowledge-enhanced prompting generation and a second-stage for empathetic dialogue generation. By simulating the intricate thought process of humans through this multi-level knowledge-enhanced prompting approach, we achieve more efficient and comprehensive emotional perception and empathetic expression effect, addressing the limitations of existing empathetic dialogue systems in handling implicit emotions. Through comprehensive automatic and human evaluations, our experimental results demonstrate a significant advantage of our approach in empathetic dialogue generation task compared to relevant baseline methods.

II. Related Work

Early studies have demonstrated that knowledge contributes to ensuring the knowledgeable and engaging responses in dialogue systems [16], [19]. In order to more effectively recognize the speaker's emotions and generate empathetic responses, numerous previous approaches have been proposed. Current empathetic dialogue systems commonly rely on external knowledge to enhance empathetic expression [12]. Sabour et al. leverage common-sense knowledge to gather more information about dialogue context and use this additional information to further enhance empathetic expression in generated responses [12]. Li et al. propose EmpDG, utilizing coarse-grained dialogue-level and fine-grained token-level emotion information to generate more empathetic responses [13]. Zhu et al. design an additional layer based on the Transformer architecture specifically for topic detection to generate more empathetic responses with the topic knowledge [14]. Similarly, Some approaches focus on knowledge related to emotional reasoning, such as knowledge of emotion causes [20], [21], the intent of empathy [22], and emotion-related knowledge in emotion understanding [23].

With the development of large language models, some studies of large language models for empathetic dialogue have been explored. Lee et al. explore the performance of GPT-3 in generating empathetic responses based on prompt-based context learning [24]. Qian et al. infer users' emotional states through a two-stage process while introducing the common-sense knowledge base ATOMIC as external knowledge to enhance emotional awareness [25]. Fu et al. combine context learning with common-sense knowledge (desires, reactions, and intentions) to enhance ChatGPT's causal reasoning ability, thereby improving emotional understanding [26].

However, existing methods still struggle to comprehensively capture all implicit emotions. Therefore, this paper proposes a multi-level knowledge-enhanced prompting approach, aiming to deeply stimulate the reasoning and emotional expression capabilities of large language models for more efficient empathetic dialogue generation.

III. Method

In this section, we introduce the proposed multi-level knowledge-enhanced prompting approach for generating empathetic dialogue generation. Our goal is to design an effective prompting method to enhance the empathetic expression capability of large language models. The proposed multi-stage dialogue prompting framework is illustrated in Figure 1. It consists of a first-stage process for multi-level knowledge-enhanced prompting generation and a second-stage process for dialogue generation, both utilizing the same large language model (LLM). In the first-stage process, we use topic model and the SHAP algorithm to acquire two types of low-level emotional knowledge (see Part B). Next, we retrieve dialogues most similar to the dialogue context in terms of topics and emotional attributes, forming mid-level emotional knowledge (see Part C). Subsequently, by designing a suitable prompt, we guide a large language model (LLM) to generate advanced high-level emotional knowledge based on the information from the first two levels and the dialogue context (see Part D), and this multi-level emotional knowledge will be used as a kind of knowledge-enhanced prompting for the second-stage. Then, the second-stage process produces emotionally rich responses by considering the dialogue context and exemplars retrieved based on the three levels of emotional knowledge (see Part E).

A. Task Definition

Formally, the dialogue context consists of alternating utterances between the speaker and the listener, defined as $C = \{U_1, ..., U_{n-1}\}$, where U_i represents the i^{th}utterance, and n denotes the number of utterances in a dialogue. Our objective is to assume the role of the listener and generate an empathetic, coherent, and informative response, denoted as Y, corresponding to U_n.

B. Low-level Emotional Knowledge Acquisition

We adopt an advanced approach utilizing a Neural Topic Model (NTM) to extract the topic words from the dialogue context. Specifically, we introduce a Variational Autoencoder (VAE) [27] as a component of the NTM, comprising an encoder and a decoder. Initially, for each input dialogue, we represent it as a word bag containing all the words present in the conversation. Subsequently, we employ the VAE encoder to map the word bag into an intermediate representation denoted as μ and $\log\sigma$. This step is analogous to data compression, where μ signifies the mean of the encoding, and $\log\sigma$ represents the standard deviation of the encoding. Assuming each dialogue content has k topics, with each topic being a probability distribution denoted as θ, these topic distributions θ are obtained by sampling and decoding from μ and $\log\sigma$:

$$\mu = f_\mu(x_{bow}), \log\sigma = f_\sigma(x_{bow}), z \sim N(\mu, \sigma^2) \qquad (1)$$

Each topic k has a probability distribution over words in the vocabulary, forming a distribution θ of topic words, with the selection of topic words sampled from the vocabulary probability distribution of each topic:

$$\theta = softmax(z), p_w = softmax(W_\phi \theta) \qquad (2)$$

In summary, we capture the deep-layer information of topics in a probabilistic manner, treating the obtained topic words as a form of low-level emotional knowledge. This serves as the first-level knowledge foundation, enabling large language models to comprehend the emotional nuances in dialogue context more profoundly.

On the other hand, we use SHAP to extract another type of low-level emotional knowledge. In principle, SHAP explains the model's output by calculating the "Shapley values" for

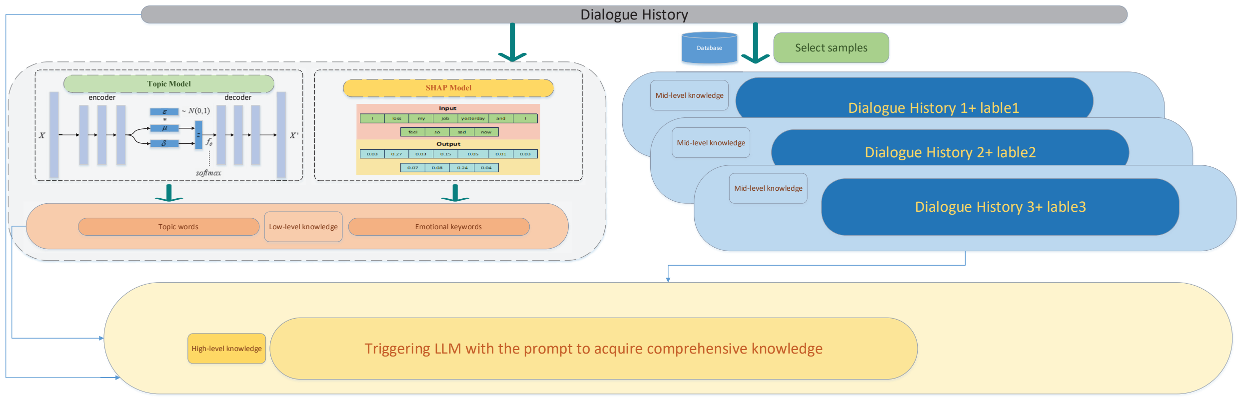

Fig. 1. The first-stage for multi-level knowledge-enhanced prompting generation.

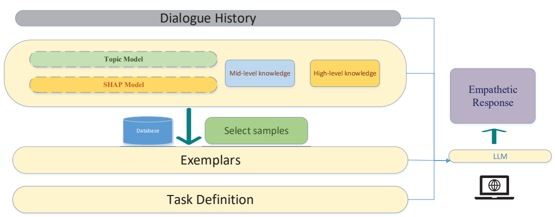

Fig. 2. The second-stage for empathetic dialogue generation

each feature, representing the importance of each feature to the model's prediction. These Shapley values indicate the significance of features for predictions. Therefore, SHAP can assist in identifying the impact of different words on emotional predictions, which are referred to as SHAP weight scores.

We compute the SHAP weight scores for different words, represented mathematically as:

$$sc_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|N|!(|N| - |S| - 1)!}{|N|!} [f(S \cup \{i\}) - f(S)] \tag{3}$$

where sc_i is the contribution score of the i-th word in the dialogue context, N is the set of words, S is the set of words excepted i-th word, and f is a sentiment classifier.

We consider words with sc above a threshold as emotional keywords. We extract several emotional keywords from the context of the dialogue as a second category of low-level emotional knowledge. We believe that emotional keywords can reflect key points of emotional interaction in the dialogue context. Similar to topic words, they will also establish a first-level emotional knowledge base for a deeper understanding of emotional content in dialogue by large language models.

C. Mid-level Emotional Knowledge Acquisition

Next, intuitively speaking, utilizing knowledge from similar themes or dialogue contexts can help the LLM generate contextually relevant and factually accurate knowledge sentences. Therefore, we assume that selecting appropriate dialogue samples as reference knowledge is crucial for generating high-quality empathetic responses. We consider similar samples as mid-level emotional knowledge. We propose a query-based sample selection method designed to search for similar samples from empathetic dialogue generation dataset based on an input query (q). To ensure that the selected samples are relevant to the query, we use a pre-trained sentence encoder (SE) [15]to obtain representations of the query and each data sample (d_i) in empathetic dialogue generation dataset. We then calculate the dot product of their representations to measure the similarity between the query and each sample:

$$Sim(q, d_i) = SE(t+h)^{\mathrm{T}} \cdot SE(t_i + h_i) \tag{4}$$

Finally, we select the top n samples with the highest similarity scores to q. Due to the relatively small size of the empathetic dialogue dataset, this selection process can be efficiently performed.

D. High-level Emotional Knowledge Acquisition

Next, based on low-level emotional knowledge and mid-level emotional knowledge, we generate high-level emotional knowledge using a large language model.

We ask LLM to generate high-level emotional knowledge with the following template:

> Given the dialogue context [X], we know that the emotional keywords are [SC words], the corresponding theme information is [Topic words]], and the most relevant similar samples are [D1, D2, D3]. The corresponding emotional labels are [L1, L2, L3]]. Based on the above information and common-sense knowledge, summarize the key points of emotional interactions to the given dialogue context.

E. The Second-Stage Process for Dialogue Generation

First, based on all multi-level knowledge (q), we proceed to retrieve semantically more relevant samples as reference knowledge exemplars. We continue to employ a query-based sample selection method, searching for similar samples from the dataset. Finally, we choose the top 2 samples with the highest similarity scores to q as reference exemplars. Next, we generate empathetic dialogues based on exemplars, task descriptions, dialogue context, and the multi-level knowledge foundation.The second-stage process produces emotionally rich responses with the following prompt:

> Given the task definition [EG] and the dialogue context [X], based on the all above information and common-sense knowledge, give your response.

IV. Experiments

A. Dataset

We conduct experiments on the EMPATHETICDIALOGUES (ED) dataset [28]. The ED dataset, tailored for the multi-turn empathetic dialogue generation task, comprises 24,850 dialogic rounds featuring 32 distinct emotion types. Each dialogue unfolds through successive utterances from a speaker and a listener. The speaker talks about their situation and feelings, while the listener endeavors to discern and respond empathetically to the speaker's emotional expressions.

B. Evaluation metrics

We follow previous related studies, conducting both automatic and human evaluations.

Automatic evaluation metrics. We employ various automatic metrics to evaluate response generation performance. Emotion understanding accuracy is measured using emotion prediction accuracy (Acc). Additionally, we utilize Distinct-n (Dist-1/2) [29], BERTScore (PBERT, RBERT, FBERT) [30], and BLEU-n (B-2/4) [31] for a comprehensive evaluation. Distinct-n gauges the diversity of generated language by measuring the proportion of distinct n-grams. Higher Distinct-n values indicate greater linguistic diversity. BERTscore evaluates similarity between generated and reference sentences using cosine similarity, providing detailed precision, recall, and F1 scores (PBERT, RBERT, FBERT). This metric excels in capturing nuanced language details. BLEU-n (B-2/4) calculates precision, emphasizing overlapping n-grams, effective in assessing adequacy and fluency. Higher BLEU-n scores, including B-2 and B-4, signify improved similarity and relevance, highlighting the model's linguistic accuracy.

Human evaluation metrics. Human evaluation assumes significance because automatic metrics may not always align with human judgment. Sixty dialogues were randomly selected from the test dataset. We choose competitive models as baselines. Taking into account the responses generated by these models, we recruite three evaluators to score the generated responses based on four aspects: empathy, informativeness, coherence, and fluency. All aspects were rated on a Likert scale (1: not at all, 3: somewhat, 5: very much). These four aspects are: 1) Empathy: whether the response demonstrates an understanding of the user's feelings and experiences and is appropriately expressed; 2) Informativeness: whether the response contains more valuable information; 3) Coherence: whether the response is logically coherent and contextually relevant; 4) Fluency: whether the response is easy to read.

C. Baselines and Ablation Study

Our baseline methods include MIME [32], MoEL [33], EmpDG [17], KEMP [8], Cem [6], We also test with GPT-3 davinci and another version of GPT-3.5 (text-davinci-003). As for ChatGPT, we use the ”gpt-3.5-turbo”, provided through the OpenAI API, which serves as the foundational model for ChatGPT. The temperature setting was configured to 0. Additionally, we investigate the effects of different components of our method through the following ablation studies. We combine various components of multi-level knowledge with a prompt consisting of a task definition for the empathetic dialogue generation task, enabling the LLM to generate empathetic responses based on the dialogue context.

D. Automatic metrics results

Table I presents the results of our method on automatic evaluation metrics. We observe a significant improvement of our proposed approach over Moel across all automatic metrics. This demonstrates that, compared to Moel, which relies solely on dialogue history to infer emotional states and generate responses, our method, leveraging multi-level knowledge, achieves more effective emotional responses beyond generating fluent sentences based on semantic mapping alone. Additionally, when compared to MIME, our method exhibits higher diversity and more accurate emotional understanding, indicating that our multi-level knowledge significantly outperforms MIME's polarity-based emotional clustering, emotional imitation, and random emotional mixing methods in the domain of expressing emotional information.

Furthermore, our method outperforms EmpDG overall, suggesting that multi-level emotional knowledge aids in understanding more complex implicit emotions more effectively than the coarse-grained dialogue-level information used by EmpDG to identify implicit emotions. Moreover, compared to KEMP, our method shows improvement, indicating that the multi-level emotional knowledge obtained through our prompting approach is more effective than the simple emotional lexicon knowledge contained in KEMP.

The comparison of our method with other approaches based on large language models suggests that, in harnessing the capabilities of large language models, multi-level knowledge contributes to inspiring comprehensive emotional understanding and expression by LLMs. Overall, our proposed method surpasses powerful baselines, underscoring the necessity of extracting and integrating multi-level emotional knowledge and validating the effectiveness of our approach in generating empathetic dialogues. In the ablation experiments section, we first test the effectiveness of incorporating topic knowledge. The results indicate an improvement in both Bleu and BERTScore metrics, suggesting that by introducing topic knowledge, LLM successfully focuses on specific information in the dialogue context, thereby generating more explicit responses. Further analysis of the experimental results incorporating emotional keywords suggests that, compared to topic knowledge, emotional keywords can further enhance the empathetic understanding and expression capabilities of LLMs. This may be

TABLE I
RESULTS OF AUTOMATIC EVALUATION.

	Models	Acc	Dist-1	Dist-2	B-2	B-4	PBERT	RBERT	FBERT
Baselines	Moel	30.75	0.47	2.16	6.95	1.99	85.57	85.98	85.76
	MIME	31.21	0.45	1.83	6.78	1.94	85.29	86.05	85.66
	EmpDG	30.64	0.47	1.98	7.17	2.03	85.59	86.20	85.88
	KEMP	36.57	0.66	3.26	5.99	1.82	85.33	85.97	85.64
	CEM	37.81	0.64	2.86	5.64	1.70	85.77	86.04	85.89
	GPT-3	-	2.07	9.08	6.88	2.22	85.62	86.10	85.84
	GPT-3.5	-	2.37	11.84	8.51	2.80	**87.37**	87.27	87.31
	ChatGPT	-	2.72	17.12	6.19	1.86	86.79	87.91	87.33
Ablation Experiments	Low-level-topic	-	2.73	17.40	6.54	2.30	87.20	87.83	87.39
	Low-level-SHAP	-	2.74	17.42	6.92	2.50	87.23	87.95	87.58
	Mid-level	-	2.83	17.51	6.77	2.14	87.07	87.86	87.49
	High-level	-	**2.98**	**18.38**	**7.64**	**2.57**	**87.29**	**88.04**	**87.63**

TABLE II
RESULTS OF HUMAN EVALUATION

	Models	Empathy	Informativity	Coherence	Fluency
Baselines	Moel	3.08	3.5	2.75	3.17
	MIME	3.58	3.17	2.82	3.42
	EmpDG	3.25	3.08	3.00	3.08
	KEMP	3.17	3.83	3.00	3.25
	CEM	3.42	3.67	3.17	3.08
	ChatGPT	3.50	3.59	3.42	3.59
Ablation Experiments	Low-level-topic	3.67	3.50	3.42	3.17
	Low-level-SHAP	4.00	**3.75**	3.67	3.42
	Mid-level	3.67	3.25	3.59	**3.67**
	High-level	**4.17**	**3.75**	**4.00**	3.58

attributed to the fact that emotional keywords knowledge is more applicable to empathetic understanding and expression.

Next, we investigate the experimental results of incorporating mid-level emotional knowledge. Surprisingly, the results demonstrate that standalone mid-level knowledge does not bring about significant performance improvements as expected. The corresponding reason may lie in the quality issues of the dataset, as empathetic dialogue scenarios are more diverse, and retrieving similar dialogues does not guarantee obtaining directly effective information; instead, it may introduce some noise, thereby weakening certain performance aspects. However, it is noteworthy that it contributes to improving diversity, indicating that standalone mid-level knowledge introduces new information.

Finally, by introducing high-level knowledge, we further validated our hypothesis. The results show that, our method successfully integrates standalone mid-level knowledge and other effective knowledge, comprehensively and effectively enhancing the overall quality of responses, achieving the best results in all ablation experiments.

E. Human evaluation results

Table II reports the results of human evaluation, and we find that responses generated based on high-level knowledge exhibit remarkable performance in terms of empathy compared to all baselines. This suggests that by acquiring multi-level knowledge, the model can approach a level of understanding and expression of implicit emotions closer to that of humans. This may be attributed to the design of our multi-level knowledge-enhanced prompting method, which resembles human thinking processes. In human evaluations, we also observe that responses generated based on high-level knowledge received superior scores in fluency, demonstrating that the incorporation of knowledge and high-level knowledge organization contributes to capturing the key points of empathetic dialogues and expressing them coherently.

V. CONCLUSION

We propose a novel multi-level knowledge-enhanced prompting approach for empathetic dialogue generation task. By designing multi-level prompts for emotional understanding and empathetic expression phrases, our approach enables LLMs to delve deeper into the emotions within the dialogue context. Consequently, it addresses the shortcomings of existing empathetic dialogue systems. Through extensive experiments, we have demonstrated the significant advantages of our method in empathetic dialogue generation task compared to other baselines.

ACKNOWLEDGMENT

This research is supported by Science and Technology Innovation 2030- Major Project “New Generation Artificial Intelligence (2030)”(2022ZD0116407). It is also supported by National Natural Science Foundation of China (62101552).

REFERENCES

[1] J. Gao, M. Galley, and L. Li, “Neural approaches to conversational AI,” in Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics: Tutorial Abstracts. Melbourne, Australia: Association for Computational Linguistics, Jul. 2018, pp. 2–7.

[2] Y. Zhang, S. Sun, M. Galley, Y.-C. Chen, C. Brockett, X. Gao, J. Gao, J. Liu, and B. Dolan, “DIALOGPT : Large-scale generative pre-training for conversational response generation,” in Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations. Online: Association for Computational Linguistics, Jul. 2020, pp. 270–278.

[3] 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 Annual Meeting of the Association for Computational Linguistics, pp. 5370–5381.

[4] Y. Ma, K. L. Nguyen, F. Z. Xing, and E. Cambria, “A survey on empathetic dialogue systems,” Information Fusion, vol. 64, 2020, pp. 50–70.

[5] Mao Yan Chen, Siheng Li, and Yujiu Yang. 2022. "Emphi: Generating empathetic responses with human like intents," in Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL 2022, Seattle, WA, United States, July 10-15, 2022, pp. 1063–1074.

[6] S. Sabour, C. Zheng, and M. Huang, "CEM: Commonsense-aware empathetic response generation," In Proceedings of the AAAI Conference on Artificial Intelligence, 2022. pp. 423–425.

[7] S. Liu, C. Zheng, O. Demasi, S. Sabour, Y. Li, Z. Yu, Y. Jiang, and M. Huang, "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 (Volume 1: Long Papers). Online: Association for Computational Linguistics, Aug. 2021, pp. 3469–3483.

[8] Q. Li, P. Li, Z. Ren, P. Ren, and Z. Chen, "Knowledge bridging for empathetic dialogue generation," in Proceedings of the AAAI Conference on Artificial Intelligence, 2022, pp. 39–45.

[9] L. Zhu, G. Pergola, L. Gui, D. Zhou, and Y. He, "Topic-driven and knowledge-aware transformer for dialogue emotion detection," in Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers). Online: Association for Computational Linguistics, Aug. 2021, pp. 1571–1582.

[10] P. Zhong, D. Wang, and C. Miao, "Knowledge-enriched transformer for emotion detection in textual conversations," 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). Hong Kong, China: Association for Computational Linguistics, Nov. 2019, pp. 165–176.

[11] T. B. Brown, B. Mann, N. Ryder, M. Subbiah, J. Kaplan, P. Dhariwal,A. Neelakantan, P. Shyam, G. Sastry, and A. Askell, "Language Models are Few-Shot Learners," in The 34th Conference on Neural Information Processing Systems (NeurIPS), 2020.

[12] S. Sabour, C. Zheng, and M. Huang, "CEM: Commonsense-aware empathetic response generation," in Proceedings of the AAAI Conference on Artificial Intelligence, 2022.

[13] Q. Li, H. Chen, Z. Ren, P. Ren, Z. Tu, and Z. Chen, "EmpDG: Multiresolution interactive empathetic dialogue generation," in Proceedings of the 28th International Conference on Computational Linguistics. Barcelona, Spain (Online): International Committee on Computational Linguistics, Dec. 2020, pp. 4454–4466.

[14] L. Zhu, G. Pergola, L. Gui, D. Zhou, and Y. He, "Topic-driven and knowledge-aware transformer for dialogue emotion detection," in Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers). Online: Association for Computational Linguistics, Aug. 2021, pp. 1571–1582.

[15] Vladimir Karpukhin, Barlas Oguz, Sewon Min, Patrick Lewis, Ledell Wu, Sergey Edunov, Danqi Chen, and Wen-tau Yih. "Dense passage retrieval for open-domain question answering," in Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP). 2020, pp. 6769–6781.

[16] Byeongchang Kim, Jaewoo Ahn, and Gunhee Kim. 2019. Sequential latent knowledge selection for knowledge-grounded dialogue. In International Conference on Learning Representations.

[17] Q. Li, H. Chen, Z. Ren, P. Ren, Z. Tu, and Z. Chen, "EmpDG: Multiresolution interactive empathetic dialogue generation," in Proceedings of the 28th International Conference on Computational Linguistics. Barcelona, Spain (Online): International Committee on Computational Linguistics, Dec. 2020, pp. 4454–4466.

[18] Lundberg, Scott M., and Su-In Lee. "A unified approach to interpreting model predictions" Advances in neural information processing systems 30.2017.

[19] Mojtaba Komeili, Kurt Shuster, and Jason Weston. 2021. Internet-augmented dialogue generation. arXiv preprint arXiv:2107.07566.

[20] Gao, J., Liu, Y., Deng, H., Wang, W., Cao, Y., Du, J., Xu, R. "Improving empathetic response generation by recognizing emotion cause in conversations," in Findings of the association for computational linguistics: EMNLP 2021. pp. 807-819.

[21] Qian, Yushan, et al. "Empathetic Response Generation via Emotion Cause Transition Graph." ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2023, pp. 1-5

[22] Welivita, Anuradha, and Pearl Pu. "A Taxonomy of Empathetic Response Intents in Human Social Conversations." in Proceedings of the 28th International Conference on Computational Linguistics. 2020, pp. 4886-4899.

[23] Wang, Lanrui, et al. "Empathetic Dialogue Generation via Sensitive Emotion Recognition and Sensible Knowledge Selection." Findings of the Association for Computational Linguistics: EMNLP 2022. 2022, pp. 4634-4645

[24] Pan, Rui, et al. "Improving Empathetic Dialogue Generation with Semantics Decoupling," in Proceedings of the 11th International Joint Conference on Knowledge Graphs. 2022, pp. 55-63.

[25] Qian, Yushan, Wei-Nan Zhang, and Ting Liu. "Harnessing the Power of Large Language Models for Empathetic Response Generation: Empirical Investigations and Improvements." arXiv preprint arXiv:2310.05140 (2023).

[26] Fu, Yahui, et al. "Reasoning before Responding: Integrating Commonsense-based Causality Explanation for Empathetic Response Generation." arXiv preprint arXiv:2308.00085 (2023).

[27] F. Nan, R. Ding, R. Nallapati, and B. Xiang, "Topic Modeling with Wasserstein Autoencoders," in Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Florence, Italy:Association for Computational Linguistics, 2019, pp. 6345–6381.

[28] H. Rashkin, E. M. Smith, M. Li, and Y.L. Boureau, "Towards empathetic open-domain conversation models: A new benchmark and dataset," in Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Florence, Italy: Association for Computational Linguistics, Jul. 2019, pp. 5370–5381.

[29] Li, Jiwei, et al. "A Diversity-Promoting Objective Function for Neural Conversation Models," in Proceedings of NAACL-HLT. 2016, pp.110-119.

[30] Zhang, Tianyi, et al. "BERTScore: Evaluating Text Generation with BERT." International Conference on Learning Representations. 2019.

[31] Kishore Papineni, Salim Roukos, Todd Ward, and WeiJing Zhu. 2002. "Bleu: a method for automatic evaluation of machine translation," in Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics. 2002, pp. 311–318.

[32] R. Mihalcea, and S. Poria, "MIME: MIMicking emotions for empathetic response generation," in Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP). Online: Association for Computational Linguistics, Nov. 2020, pp. 8968–8979.

[33] Z. Lin, A. Madotto, J. Shin, P. Xu, and P. Fung, "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). Hong Kong, China: Association for Computational Linguistics, Nov. 2019, pp.121–132.

[34] W. Qiang, J. Li, B. Su, J. Fu, H. Xiong, JR. Wen. "Meta attention-generation network for cross-granularity few-shot learning," International Journal of Computer Vision, 2023, 131(5): 1211-1233.

[35] W. Qiang, J. Li, B. Su, C. Zheng, H. Xiong. "Robust local preserving and global aligning network for adversarial domain adaptation," IEEE Transactions on Knowledge and Data Engineering, 2021, 35(3): 3014-3029.

[36] W. Qiang, J. Li, C. Zheng, B. Su, H. Xiong. "Interventional contrastive learning with meta semantic regularizer," International Conference on Machine Learning. PMLR, 2022: 18018-18030.

[37] W. Qiang, J. Li, C. Zheng, B. Su. "Auxiliary task guided mean and covariance alignment network for adversarial domain adaptation," Knowledge-Based Systems, 2021, 223: 107066.

[38] J. Li, W. Qiang, C. Zheng, B. Su, H. Xiong. "Metaug: Contrastive learning via meta feature augmentation," International Conference on Machine Learning. PMLR, 2022: 12964-12978.

[39] W, Qiang, H, Zhang, J, Zhang, L, Jing. "TSVM-M3: twin support vector machine based on multi-order moment matching for large-scale multi-class classification," Applied Soft Computing, 2022, 128: 109506.