

Citation: Lili Yin, Xiaoqiang Du, and Hengwen Gu. Empathetic response generation based on topic shifting. *Journal of Harbin Institute of Technology (New Series)*. DOI: 10.11916/j.issn.1005-9113.2024067

Empathetic Response Generation Based on Topic Shifting

Lili Yin¹, Xiaoqiang Du¹, and Hengwen Gu^{2*}

(1. College of Computer Science and Technology, Harbin University of Science and Technology, Harbin 150080, China;
2. No. 703 Research Institute of China Shipbuilding Industry Corporation, Harbin 150001, China)

Abstract: The current dialogue system can be sensitive to the emotions in the user's words, to generate an empathetic response and help calm the user's emotions. But in some cases, eliciting empathetic responses may not adequately mitigate the adverse effects that the current conversation topic is having on users. The dialogue system will continue the conversation with the user under this uncomfortable topic, which will lead to a worse chat situation or even an impasse. To solve this problem, a dialogue system that can change the topic autonomously according to the user's emotions is proposed in this paper. Specifically, the dialogue system first collects the emotional semantic information of the users and then detects it according to the emotion classification module. Once the detection results show that the user is in a bad mood, the topic change module selects a new topic from the context to shift to and generates a response. This not only helps to calm the user's mood but also move the conversation to a new area, steering the user away from the uncomfortable topic. The experimental results show that the proposed method incurs less cost in terms of content quality, but improves the emotional perception ability. Additionally, it endows the dialogue system with the ability to change the topic, and improves the user's dialogue experience.

Keywords: Topic Shifting; Empathetic Response; Emotion Classification

CLC number: TP18

Document code: A

Article ID: 1005-9113(2025)00-0000-09

0 Introduction

Incorporating human emotions into conversational AI systems is a challenge of long-term significance. The ability to recognize and express emotions is a key component of human intelligence and the basis for establishing natural human-computer interaction^[1]. Lately, there has been a growing interest in empathetic communication within discussions, particularly within natural language processing and question-answering systems^[2]. Incorporating emotional intelligence into conversational AI systems will enhance their ability to engage in more human-like interactions^[3]. It can identify and track the speaker's emotional state during a conversation and generate emotionally reasonable and natural responses^[4]. This technology is applicable across various domains, including commercial client support, jurisprudence, healthcare, and additional spheres of

dialogue. The development of conversational artificial intelligence with emotional cognition and expression capabilities is a long-term challenge and goal for AI research to reproduce human intelligence^[5].

In this work, the unit of analysis is the dialogue. A dialogue typically contains contextual cues^[6], which trigger the emotion of the current utterance, such as reasons or circumstances. The methods proposed by Lin^[7] and Majumder^[8] gauged the speaker's emotions by deciphering the total meaning and feelings conveyed within the discourse. These methods are prone to the wrong response, which is attributed to ignoring the nuances of human emotion in the conversation. Recent work on empathetic response has been devoted to sensing the emotional vocabulary in the speaker's context through deep learning models^[9-10]. However, the current technology of empathetic dialogue still faces some challenges. Current methods fall short in grasping the intricate emotional nuances within the scenario and the

Received 2024-08-28.

Sponsored by the Natural Science Foundation of Heilongjiang Provincial (Grant No. LH2023F033), and the Science and Technology Innovation Talent Project of Harbin (Grant No. 2022CXRC006).

* Corresponding author. Hengwen Gu, Senior Engineer at No. 703 Research Institute of China Shipbuilding Industry Corporation. Email: guhengwen234@126.com.

underlying semantic meanings in individual experiences, and the same utterance may express different emotional connotations depending on what different people think^[11]. To further improve emotion recognition and understanding in conversation, models need to better analyze contextual cues and capture the semantics of personal experiences. In addition, since the current dialogue system will only perform sentiment analysis based on the user's input to generate responses, in the event of a bad topic, it will only continue to generate responses under the current topic, allowing the dialogue context to then deteriorate. Developing a dialogue system that can strategically decide to shift the topic in line with the present context is a challenging area of study.

Most of the current emotion-aware dialogue systems focus on the overall semantics of the contextual language and the emotional words in the communicator's language as static vectors to perceive subtle emotions^[12-13], however, ignoring the impact of the current chat topic on the user's emotions. To tackle the aforementioned challenges, we propose a novel framework designed to facilitate topic shifts in conversational systems, enabling them to transition seamlessly into chat contexts. The framework performs emotion classification on user input and guides the model to conduct more accurate emotion classification by introducing a dynamic relational graph convolutional network, thereby facilitating better topic shifts and generating more empathetic responses. The emotion classification module will classify the emotions captured by the dynamic correlation graph convolutional network, and the dialogue system will work normally when positive or neutral emotions are detected. When extremely negative emotions are detected, the conversation system will choose to shift the topic, and the topic shift module will select other topics in the context to generate a response. Although shifting the topic may diminish the coherence of the conversation, it can eliminate the unfavorable subject matter, enhance the scalability of dialogue, and provide comfort to the user's feelings by offering a different viewpoint.

We evaluated the validity of the proposed approach on the empathetic dialogue benchmark with the EmpatheticDialogues datasets and compared it to current state-of-the-art empathetic dialogue models. The results show that the proposed method can effectively change the topic, get rid of the bad topic,

accurately understand the conversation, and produce grammatically fluent empathetic feedback.

The contribution of this study can be summarized as follows: A new framework is proposed for the dialogue system, which realizes the topic shift by classifying the emotion of the dialogue; There are few researches on topic shifting in conversation system, but the feasibility of this research is proved by experiments.

1 Related Work

1.1 Empathetic Dialogue Generation

The objective of creating empathetic dialogues necessitates that systems acquire the ability to convey emotions in a manner that is empathetically fitting. Rashkin et al.^[14] introduced affective representations generated by pre-trained affective classifiers to learn and express specific affective types in dialogue. Lin et al.^[7] determined the nuanced emotional patterns in replies by leveraging a consortium of specialists, thereby enhancing the capacity for empathic reactions within a dialogue model. Majumder et al.^[8] simulated the speaker's emotions to a certain extent to produce an empathetic response. These various methods concentrate on the general ambiance of the dialogue, neglecting the nuanced feelings. Detecting nuanced sentiments, Gao et al.^[12] and Kim et al.^[15] developed a sentiment analysis tool aimed at identifying emotionally charged terms and discerning delicate emotional states. Wang et al.^[16] used a fine-grained encoding strategy to make the model more sensitive to the flow of emotions in a conversation. Yang et al.^[17] employed a contextually adaptive convolutional neural network to direct a system in discerning the interplay of emotional content and semantic meaning. In addition, some methods guide dialogue models to empathic response abilities by introducing common sense knowledge. Sabour et al.^[10] used common sense to infer subtle emotions in conversations. Li et al.^[18] incorporated general knowledge and sentiment-related lexical information to enable models to clearly comprehend and articulate emotions. Jiang et al.^[19] extracted knowledge with emotional representation from common sense knowledge, guided the interactive reasoning between conversation history and common sense knowledge and imitated human reasoning based on common sense to generate empathic responses. Zhou et al.^[20] introduced common-sense cognitive

maps and emotion concept maps to learn subtle emotions in a dialogue.

1.2 Topic Shifting

A majority of current advisory conversational systems primarily reply to users' statements in an effort to more accurately discern their indicated preferences or inquiries, subsequently offering tailored suggestions. This responsive suggestion dialogue system faces practical constraints, given that individuals often lack defined inclinations toward novel topics or unfamiliar objects. Wang et al.^[21] proposed a goal-driven recommendation dialogue system in which, given a specified target topic (e.g., movie, music, food), the dialogue system will actively shift the topic of the conversation toward its set topic. In Ref. [22], Wang et al. proposed a coherent dialogue planning method that employed a stochastic mechanism to capture the sequential evolution of the conversational trajectory, generate more coherent statements, and realize the transfer of the target topic with a higher success rate. Wang et al.^[23] proposed a new goal-constrained bi-directional planning method that plans appropriate dialogue paths through foresight and retrospection, thus enabling topic transfer.

In addition, Liu et al.^[24] proposed a spoken conversation system that combines the two modes of small talk and task orientation, enabling switching from one mode to the other. Bang et al.^[25] put forward an example-based dialogue system. By retrieving samples in the database, matching samples are selected for reply. If there is no suitable sample, topics that users are interested in are selected and inserted into the template to generate reply. Sakai et al.^[26] proposed a Twitter-based system for topic selection within dialogue platforms, designed to spark discussions on topics that pique users' curiosity while being new to them.

With the progress of dialogue, the topic shift at the discourse level will naturally occur in the context of continuous multi-round dialogue. Nevertheless, existing retrieval-focused systems rely solely on local thematic vocabulary for representing the context of discourse, yet they are unable to seize critical overarching subject-related indicators on the discourse level. Xu et al.^[27] proposed a novel topic perception solution for modeling multiple rounds of dialogue, which identifies and isolates sections of topic-aware text without supervision, enabling the resulting model

to detect prominent topic shifts necessary for discourse management and to efficiently monitor the interaction.

Overall, in this research field, some dialogue systems are able to capture the subject words in discourse during multiple rounds of dialogue, so that the topic of dialogue can shift with the passive movement of users. Some task-led conversation systems do not perceive the topic in the user's speech, but instead turn the topic to the system's set topic. Another part of the dialogue system will actively change the topic, but the logic of changing the topic is that there is no matching reply in the dialogue system database. Unlike these approaches, our dialogue system can actively shift topics based on the user's emotions, and the topics chosen are those that appear in the user's context and are not set.

2 Method

We provide an overview of Topic Change Emotion Semantic (TCES) dialogue system in Fig.1, which consists of six components: a) Context encoder: used to understand context semantics, and generate context word embeddings and context semantic representations; b) Emotion semantic integrator: combining emotion embeddings with contextual semantics embeddings to generate emotion-semantic representations; c) Syntactic dependency module: by introducing a dependency tree, the correlation between words is fed back into emotion-semantic representations to generate correlation representations; d) Emotion classification module: emotion prediction based on context semantic representation and correlation representation; e) Topic change module: judgment of whether the topic is shifted or not based on predicted sentiments, and generation of replies based on concepts in the conversation combined with templates in the event of a topic shift; f) Dialogue generation: generates empathetic responses based on context semantic representations and correlation representations. See below for specific methods.

2.1 Task Formulation

Considering the conversational framework $X = [S_1, S_2, \dots, S_N]$ between two speakers, the model is tasked with precisely discerning the emotional and semantic nuances within the conversational framework and produce an empathetic response $Y = [y_1, y_2, \dots, y_j, y_N]$. Here, $S_i = [w_1^i, w_2^i, \dots, w_m^i]$ represents the i -th

utterance containing m words. Y is a response to N words.

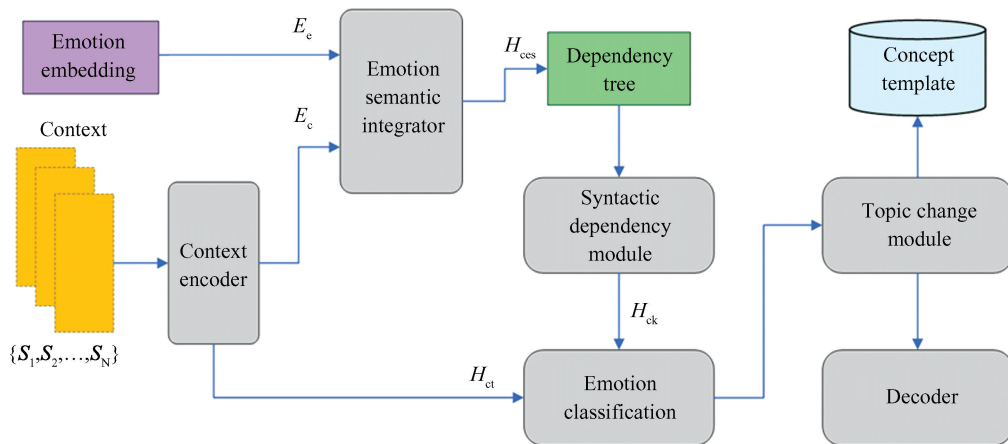


Fig.1 Overview of the proposed model

2.2 Context Encoder

As in previous approaches^[10,18], we concatenate segments in the context of a conversation and use [CLS] as a prefix for the whole sequence to form contextual input $C = [\text{CLS}] \oplus S_1 \oplus S_2 \oplus \dots \oplus S_N$. Here, \oplus represents the connector. This study translates context C into vector representations of words E_c to incorporate it into the model. Subsequently, word embeddings, position embeddings, and situation embeddings results in the creation of composite semantic embeddings E_{CM} . These situation embeddings aid in identifying the roles of speakers or respondents and are initialized stochastically. In order to grasp the meaning behind the dialogue, the study feeds semantic embeddings E_{CM} into a context encoder E_{Enect} to derive representations H_{ct} of the dialogue's context;

$$H_{ct} = E_{\text{Enect}}(E_{CM}) \quad (1)$$

where $H_{ct} \in \mathbf{R}^{L \times d}$, L is the length of the context sequence and d denotes the size of the hidden layer of the encoder.

2.3 Emotion Semantic Integrator

Linguistic research^[28] indicates that the context significantly shapes the meaning of words, and interpreting words in isolation can result in misinterpretations. Take, for instance, the term 'well', devoid of context, it typically conveys a favorable emotion. However, within the phrase 'well, we should go home' its function as an exclamation amplifies the statement's intensity, conveying no particular emotion. Such employment of language does not convey an affirmative connotation. Consequently, there is a need for a contextual adaptation of both emotional tone and meaning.

In terms of meaning, we apply a weighted calibration of semantics to the context word embeddings E_c in order to derive adaptive semantic vectors E_{cs} that $E_{cs} \in \mathbf{R}^{L \times c_s}$. c_s is the hidden size of the adaptive semantic vectors E_{cs} . In terms of sentiment, this paper converts words denoting sentiment into emotion embeddings E_e and combines them with context word embeddings to obtain fused sentiment vectors E_{ce} , $E_{ce} \in \mathbf{R}^{L \times c_e}$, where c_e denotes the number of sentiment categories.

Subsequently, the research aims to merge emotion embeddings E_{ce} with adaptive semantic vectors E_{cs} into the encoder, aiming to acquire representations that capture emotional semantics. By doing so, the model is able to account for emotional content and semantic meaning simultaneously during training, leading to a more comprehensive grasp of contextual word meanings.

$$H_{ces} = E_{\text{Eneces}}(E_{cs} \oplus E_{ce}) \quad (2)$$

where E_{Eneces} is the encoder and H_{ces} is the representation of the emotion-semantic. $H_{ces} \in \mathbf{R}^{L \times (c_s + c_e)}$.

2.4 Syntactic Dependency Module

To improve the accuracy of emotion classification and the quality of responses generated, dependency trees are introduced to reflect the grammatical dependencies between related words. In this paper, the representation of the emotion-semantic, parts of speech, and dependency types are connected to form a guidance vector V_k . Where $V_k \in \mathbf{R}^{L \times L \times (c_s + c_e + 2d_t)}$, d_t is the size of the part of speech embeddings or dependency embeddings. Then, we calculate the correlation probability between vertex m and vertex n to get the correlation representation H_{ck}

between node m and node n .

$$\mathbf{p}_{m,n} = \frac{w_{m,n} \cdot \exp(\mathbf{V}_k[m] \cdot \mathbf{V}_k[n])}{\sum_{n=1}^L w_{m,n} \cdot \exp(\mathbf{V}_k[m] \cdot \mathbf{V}_k[n])} \quad (3)$$

$$\mathbf{H}_{ck} = \text{ReLU}\left(\sum_{j=1}^L \mathbf{p}_{m,n} (\mathbf{W}_v(\mathbf{E}_{cs} \oplus \mathbf{E}_{ce})[n] + \mathbf{b}_v)\right) \quad (4)$$

Among them, $\mathbf{p}_{m,n} \in \mathbf{R}^{L \times L \times 1}$, $\mathbf{H}_{ck} \in \mathbf{R}^{L \times (c_s + c_e)}$. $w_{m,n}$ is 1 when node m and node n are directly related in the dependency tree and $w_{m,n}$ is 0 otherwise. \mathbf{W}_v and \mathbf{b}_v are trainable parameters, and ReLU is the ReLU activation function.

2.5 Emotion Classification

To cope with the problem of poor accuracy in emotion classification, recent research^[10] has proposed an approach that identifies and analyzes sentiment words during the communication process and then transforms them into static metrics, intending to achieve precise capture of subtle sentiments, thereby improving the overall accuracy of emotion classification. The study employs a pair of networks, sharing an identical structure yet distinct parameters, to integrate semantic context and correlated features for the purpose of classifying emotions. Below, we will use the processing of context semantic representations as a concrete example.

We use a hidden representation of the [CLS] tag in the context representation to classify emotions in the case of providing emotional labels e^* for each conversation.

$$\mathbf{h}_{ct} = \mathbf{H}_{ct}[0] \quad (5)$$

$$\mathbf{P}_{ct} = \text{Softmax}(\mathbf{W}_e \mathbf{h}_{ct}) \quad (6)$$

In this paper, \mathbf{h}_{ct} first passes through a linear layer and then performs Softmax operation to generate an emotion category distribution $\mathbf{P}_{ct} \in \mathbf{R}^{c_e}$. Where $\mathbf{h}_{ct} \in \mathbf{R}^d$, $\mathbf{W}_e \in \mathbf{R}^{d \times c_e}$ is the weight vector of the linear layer. Similarly, we obtain a relational representation of the emotion category distribution $\mathbf{P}_{ck} \in \mathbf{R}^{c_e}$. These two probability distributions are added together to get the final emotion \mathbf{P}_{emo} .

During training, this study optimizes these weights by minimizing the cross entropy (CE) loss between the emotion category distribution P and the ground reality label.

$$\mathcal{L}_{ct} = -\log(\mathbf{P}_{ct}(e^*)) \quad (7)$$

$$\mathcal{L}_{ck} = -\log(\mathbf{P}_{ck}(e^*)) \quad (8)$$

2.6 Topic Change Module

Most of the current research on topic shifting

aims to solve the problem that the database of the conversation system can not find a matching response or the user is tired of the current topic, and does not pay attention to the emotional aspects of the user. To solve the topic that users are in a bad mood when chatting, a topic change module is designed according to emotion classification. In this paper, the various categories of emotions are first labeled with scores, with positive strong emotions set to 1 and negative strong emotions set to -1 . All other emotions are set to 0, defined as neutral emotions. Then, the three emotions predicted by the part of the emotion classification module were weighted and combined to obtain the emotion score. Finally, the topic change is judged according to the emotion score. If the emotion score is less than 0, the topic change is carried out. Formally, the topic change module extracts the flow of concepts from the context of the conversation. $F = \{f_1, f_2, \dots, f_N\}$ represents the concept observed in the dialogue, where f_i corresponds to the concept set of the i -th discourse, collecting all the entities in S_i , i.e. $f_i = \{C_{i1}, C_{i2}, \dots, C_{im}\}$. After that, the extracted entities and the topic entity in the current user dialogue are converted into word vectors, and the cosine similarity between each entity in the concept set and the topic entity in the current dialogue is calculated respectively, and then multiplied by the weight of each concept set. Finally, the entity with the best result is selected and inserted into the created template to generate a reply.

$$\alpha_i = 1 - \frac{i}{1 + 2 + \dots + N} \quad (9)$$

Among them, the weights of the concept set α_i are shown above. Here we set the weight of the more distant sentences higher because entities that are further away from the topic of the current chat are better able to stay away from the topic of the moment that user feel uncomfortable. If the emotion score is great than 0, then no topic transfer is performed and the dialogue system generates the response through the decoder.

2.7 Response Generation

Finally, the decoder generates responses based on contextual semantic representations and correlation representations, which incorporate emotion semantics, parts of speech, and dependencies.

$P(y_t | y < t, C) = \text{Dec}(\mathbf{E}_{y < t}, \mathbf{H}_{ct} \oplus \mathbf{H}_{ck})$ (10) where $\mathbf{E}_{y < t}$ denotes the generated labeled embedding. We use the standard negative log-likelihood loss for

the target response y :

$$\mathcal{L}_{\text{gen}} = - \sum_{t=1}^T \log P(y_t | C, y < t) \quad (11)$$

All parameters of the proposed model are trained and optimized based on the weighting of the above three losses.

$$\mathcal{L} = r_1 \mathcal{L}_{\text{gen}} + r_2 \mathcal{L}_{\text{ct}} + r_3 \mathcal{L}_{\text{ck}} \quad (12)$$

where r_1 , r_2 , and r_3 are the hyperparameters which are used to control for the effects of the three losses. To highlight the effect of grammatical correlation on the dialogue system, we set $r_1 = 1$, $r_2 = 1$, and $r_3 = 1.5$.

3 Experiments

In this section, we present the details of the datasets, the comparison methods, and the implementation of our model. By typing a sentence, the dialogue system generates a response. Here, we first automatically evaluate the accuracy of the dialogue system in judging the emotion expressed in the user's sentences, and the fluency and diversity of the dialogue system in generating responses. We then manually evaluated the performance of the dialogue system's responses in four areas.

3.1 Datasets

In terms of data sets, there is no data set that focuses on evaluating topic change, but only one that evaluates empathetic response generation. Therefore, we conducted experiments on the Empathetic Dialogues (ED) dataset^[14]. ED is a large-scale multi-round conversation dataset that contains 25 k empathic conversations between speakers and listeners, providing 32 uniformly distributed emotion labels. During dialogues, individuals share their private encounters, allowing others to deduce their circumstances and feelings, leading to a compassionate response. Following the method of Rashkin et al.^[14], we partitioned the training set/validation set/test set in the ratio of 8 : 1 : 1.

3.2 Baselines

We chose the following baseline model to compare with the TCES model presented above:

1) MoEL^[7]: An extension of the transformer model that pairs a decoder with each predicted emotion, and then generates a response with a meta-decoder combining the output of each decoder.

2) MIMÉ^[8]: An empathetic dialogue system considers polar emotion categorization and emotion

replication, reflecting user sentiments in its replies.

3) KEMP^[18]: An empathetic dialogue system grounded in knowledge incorporates ConceptNet to augment the understanding of implicit emotions, enhancing the representation with supplementary insights.

4) CEM^[10]: An empathetic conversational framework that considers the emotional and cognitive dimensions of empathy, enriched with general knowledge.

5) SEEK^[16]: An empathetic dialogue model that uses fine-grained coding strategies to capture emotional dynamics in conversation, designs a new framework to simulate the interaction of knowledge and emotion.

6) CASE^[20]: A model for empathetic conversation through common-sense cognitive graphs and emotion concept graphs, integrates the user's cognitive processes and emotional states across macro and micro levels of detail.

3.3 Implementation Details

This paper implements the model using PyTorch^[29] and initializes word embeddings using pre-trained GloVe vectors^[30] of dimension 300. The model is optimized using the Adam optimizer^[31], $\beta_1 = 0.9$, $\beta_2 = 0.98$. In the training phase, the learning rate is initialized to 0.0001, and we varied this value during training according to Vaswani et al.^[32]. The model in this paper was trained on an NVIDIA Geforce RTX 4090 GPU using a batch size of 16 and an early stopping strategy. The model converged after 18400 training rounds.

3.4 Automatic Evaluation

In the model generation evaluation, we used emotion accuracy as an assessment criterion for the consistency between the underlying true emotion labels and the predicted emotion labels. For assessing the quality of model generation, we choose the widely used Perplexity (PPL) and Distinct-1/2 (Dist-1/2). PPL signifies the degree of certainty the model has regarding its potential reply options. When the confidence higher, the PPL lower. Dist-1/2 denotes the proportion of distinct one-word words/double-word words in all the generated results, indicating the diversity of the generated results.

3.5 Automatic Result And Analysis

Table 1 shows the results of the automatic evaluation of the model in this paper and other comparative models. From Table 1, it can be seen

that TCES outperforms the previous best by 1.11% in emotion classification. It demonstrates that concentrating on the link between emotional content and semantics enhances the model’s ability to classify emotion more accurately. Table 2 shows the results of ablation experiments for the model in this paper. By analyzing Tables 1 and 2, we can see that the template response strategy used in Module *C* resulted in an increase of 1.34 for the Dist-1 indicator and 3.22 for the Dist-2 indicator. After removing this module, the Dist-1/2 values of the model are 0.69 and 3.17 respectively, which are at normal level. In terms of PPL, due to the introduction of the *B* module, the confusion score of TCES is 37.27, which is lower than the other models. However, this module can effectively improve the accuracy of emotion classification. In Table 2, the w/o *A* denotes no emotion semantic integrator, w/o *B* denotes no syntactic dependency module, and w/o *C* denotes no topic change module.

Table 1 Results of automatic evaluation

Models	Acc(%)	PPL	Dist-1	Dist-2
MoEL	30.62	36.93	0.44	2.10
MIME	31.36	37.09	0.47	1.90
KEMP	39.31	36.89	0.55	2.29
CEM	39.11	36.11	0.66	2.99
SEEK	41.85	37.09	0.73	3.23
CASE	40.20	35.37	0.74	4.01
TCES	42.96	37.27	2.03	6.39

Table 2 Ablation studies

Models	Acc(%)	PPL	Dist-1	Dist-2
TCES	42.96	37.27	2.03	6.39
w/o <i>A</i>	40.32	36.89	1.95	5.94
w/o <i>B</i>	40.16	36.10	1.90	5.81
w/o <i>C</i>	42.96	37.27	0.69	3.17

In order to verify the validity of adaptive semantic vector and fused sentiment vector, module *A* is removed. The findings indicated a marked decrease in the precision of emotional detection and the variety of language used. This shows the important role of adaptive semantic vectors in accurate understanding and information expression in dialogue and the integration of sentiment vectors in capturing emotion. To verify the validity of syntax dependencies, the *B* module is removed. The findings indicated a marked

decrease in the precision of emotions and the range of language used. This implies that grammatical dependencies greatly affect the recognition of emotions and meaning, which are crucial for conveying informational reactions. After removing the *C* module, we found that the diversity of the dialogue system decreased significantly. This shows that the *C* module can also significantly improve the quality of reply generation. In addition, we found that the removal of modules *A* and *B* can improve fluency, but reduce the diversity of response generation, indicating that the sentences generated by the ablation model are relatively simple.

3.6 Human Evaluation

In order to start from the user’s point of view, this study designed an experiment of human evaluation and convened 15 evaluators to test these systems. The 15 evaluators were evenly distributed between the ages of 20 to 30, 30 to 40 and 40 to 50. Evaluators used each dialogue system for 10 min and then rated each dialogue system on a scale of 1 to 5, with the final score being averaged. We asked the evaluators to talk mainly about recent bad things and bad moods, but the format of the conversation was free. We selected the three strongest baseline models to compare TCES’ performance on the following questions: a) Are you satisfied with the chat content? b) Are you satisfied with the topic? c) Does the system detect your mood? d) Has your mood improved? The experimental results are shown in Table 3.

Table 3 Results of human evaluation

Models	Questions			
	a	b	c	d
CEM	2.52	3.35	2.82	3.23
SEEK	2.36	3.65	3.16	3.68
CASE	2.82	3.78	3.36	3.47
TCES	2.71	3.89	3.27	3.95

3.7 Human Result And Analysis

The user believes that the proposed model performs better on problems b and d and worse on problems a and c. Question a assesses the fluency and variety of responses generated by each dialogue system, and question c assesses the ability of each dialogue system to perceive emotion. Because CASE is a dialogue system built on common sense cognitive map and emotion concept map, and the system

synchronizes users' cognitive and emotional states across both broad and detailed scales. So the system performs best on these two issues. Question b assesses which of the two methods is preferred by users, and question d assesses which of the two methods is more soothing to users. According to the results of manual evaluation, the system can properly detect the negative emotions caused by the dialogue topic and change the topic to improve the user's impression of the dialogue topic.

4 Conclusions and Future Work

In this study, we propose a topic change empathetic dialogue system based on emotion classification, which performs topic shifting by recognizing the user's conversational mood to escape from some topics that put the user in a bad mood. The systematic assessment outcomes demonstrate the efficacy of our methodology for generating empathetic responses while adeptly navigating conversational topics. However, our work is not yet complete, and in the future, we would like to find a way for the dialogue system to generate replies about a topic through a topic in context.

References

- [1] Mao Y, Cai F, Guo Y, et al. Incorporating emotion for response generation in multi-turn dialogues. *Appl Intell*, 2022, 52(7): 7218–7229. DOI: 10.1007/s10489-021-02819-z.
- [2] Wang Y H, Hsu J H, Wu C H, et al. Transformer-based empathetic response generation using dialogue situation and advanced-level definition of empathy. 2021 12th International Symposium on Chinese Spoken Language Processing (ISCSLP). Hong Kong, 2021:1–5. DOI:10.1109/ISCSLP49672.2021.9362067.
- [3] Huang M, Zhu X, Gao J. Challenges in building intelligent open-domain dialog systems. *ACM Transaction on Information Systems*, 2020, 38(3): 1–32. DOI: 10.1145/3383123.
- [4] Lv X, Yang Y, Qin D, et al. Artificial intelligence service recovery: The role of empathic response in hospitality customers' continuous usage intention. *Computers in Human Behavior*, 2022, 126: 106993. DOI: 10.1016/j.chb.2021.106993.
- [5] Ni J, Young T, Pandelea V, et al. Recent advances in deep learning based dialogue systems: a systematic survey. *Artificial Intelligence Review*, 2023, 56(4): 3055–3155. DOI: 10.1007/s10462-022-10248-8.
- [6] Ghosal D, Majumder N, Poria S, et al. DialogueGCN: A graph convolutional neural network for emotion recognition in conversation. *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, 2019: 154–164. DOI:10.18653/v1/D19-1015.
- [7] Lin Z, Madotto A, Shin J, et al. MoEL: Mixture of empathetic listeners. *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*. 2019:121–132. DOI:10.18653/v1/D19-1012.
- [8] Majumder N, Hong P, Peng S, et al. MIME: MIMicking emotions for empathetic response generation. *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 2020:8968–8979. DOI: 10.18653/v1/2020.emnlp-main.721.
- [9] Li Q, Chen H, Ren Z, et al. EmpDG: Multi-resolution interactive empathetic dialogue generation. *Proceedings of the 28th International Conference on Computational Linguistics*. Barcelona, 2020:4454–4466. DOI:10.18653/v1/2020.coling-main.394.
- [10] Sabour S, Zheng C, Huang M. CEM: Commonsense-aware empathetic response generation. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2022, 36(10):11229–11237. DOI:10.1609/aaai.v36i10.21373.
- [11] Chen M Y, Li S, Yang Y. EmpHi: Generating empathetic responses with human-like intents. *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics; Human Language Technologies*. Seattle, 2022:1063–1074. DOI: 10.18653/v1/2022.naacl-main.78.
- [12] Gao J, Liu Y, Deng H, et al. Improving empathetic response generation by recognizing emotion cause in conversations. *Findings of the Association for Computational Linguistics; EMNLP 2021*. Punta Cana, 2021: 807–819. DOI: 10.18653/v1/2021.findings-emnlp.70.
- [13] Kim W, Ahn Y, Kim D, et al. Emp-RFT: Empathetic response generation via recognizing feature transitions between utterances. *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics; Human Language Technologies*. Seattle, 2022:4118–4128. DOI:10.18653/v1/2022.naacl-main.303.
- [14] Rashkin H, Smith E M, Li M, et al. Towards empathetic open-domain conversation models: A new benchmark and dataset. *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. Florence, 2019:5370–5381. DOI: 10.18653/v1/P19-1534.
- [15] Kim H, Kim B, Kim G. Perspective-taking and pragmatics for generating empathetic responses focused on emotion causes. *Proceedings of the 2021 Conference on*

- Empirical Methods in Natural Language Processing. Online and Punta Cana, 2021; 2227 – 2240. DOI; 10.18653/v1/2021.emnlp-main.170.
- [16] Wang L, Li J, Lin Z, et al. Empathetic dialogue generation via sensitive emotion recognition and sensible knowledge selection. Findings of the Association for Computational Linguistics: EMNLP. Abu Dhabi, 2022; 4634–4645. DOI: 10.18653/v1/2022.findings-emnlp.340.
- [17] Yang Z, Ren Z, Wang Y, et al.. Exploiting emotion-semantic correlations for empathetic response generation. Findings of the Association for Computational Linguistics: EMNLP, 2023, 2023; 4826–4837. DOI: 10.18653/v1/2023.findings-emnlp.320.
- [18] Li Q, Li P, Ren Z, et al. Knowledge bridging for empathetic dialogue generation. Proceedings of the AAAI Conference on Artificial Intelligence, 2022, 36 (10); 10993–11001. DOI: 10.1609/aaai.v36i10.21347.
- [19] Jiang L, Wu D, Li Y, et al. Emp-USIR: A unidirectional synchronous interactive reasoning model for empathetic dialogue. 2023 International Joint Conference on Neural Networks (IJCNN). Gold Coast, 2023; 1–9. DOI: 10.1109/IJCNN54540.2023.10191512.
- [20] Zhou J, Zheng C, Wang B, et al. CASE: Aligning coarse-to-fine cognition and affection for empathetic response generation. Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Toronto, 2023; 8223–8237. DOI: 10.18653/v1/2023.acl-long.457.
- [21] Wang J, Lin D, Li W. Follow me: Conversation planning for target-driven recommendation dialogue systems. arXiv; 2208.03516 [cs.CL] 2022. DOI: 10.48550/arXiv.2208.03516.
- [22] Wang J, Lin D, Li W. Dialogue planning via brownian bridge stochastic process for goal-directed proactive dialogue. Findings of the Association for Computational Linguistics; ACL. Toronto, 2023; 370 – 387. DOI: 10.18653/v1/2023.findings-acl.25.
- [23] Wang J, Lin D, Li W. Target-constrained bidirectional planning for generation of target-oriented proactive dialogue. ACM Transactions on Information Systems, 2024, 42(5): 124;1–124;27. DOI: 10.1145/3652598.
- [24] Liu Y, Ultes S, Minker W, et al. Unified conversational models with system-initiated transitions between chit-chat and task – oriented dialogues. Proceedings of the 5th International Conference on Conversational User Interfaces (CUI '23). Association for Computing Machinery, New York, 2023; Article 33; 1–9. DOI: 10.1145/3571884.3597125.
- [25] Bang J, Han S, Lee K, et al.. Open-domain personalized dialog system using user – interested topics in system responses. 2015 IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU). Scottsdale, 2015; 771–776. DOI: 10.1109/ASRU.2015.7404866.
- [26] Sakai Y, Matsumoto M. Twitter-based selection of topics that users are interested in but are not familiar with. 2021 IEEE/SICE International Symposium on System Integration (SII). Iwaki, 2021; 769–774. DOI: 10.1109/IEEECONF49454.2021.9382765.
- [27] Xu Y, Zhao H, Zhang Z. Topic – aware multi-turn dialogue modeling. Proceedings of the AAAI Conference on Artificial Intelligence. 2021, 35(16); Article no. 16. DOI: 10.1609/aaai.v35i16.17668.
- [28] Johns B T, Dye M, Jones M N. The influence of contextual diversity on word learning. Psychon Bull Rev, 2016, 23(4): 1214 – 1220. DOI: 10.3758/s13423-015-0980-7.
- [29] Paszke A, Gross S, Massa F, et al. PyTorch: An imperative style, high-performance deep learning library. Advances in Neural Information Processing Systems 32 (NeurIPS 2019). 2019; arXiv: 1912.01703. DOI: 10.48550/arXiv.1912.01703.
- [30] Pennington J, Socher R, Manning C. GloVe: Global vectors for word representation. Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). Doha, 2014; 1532–1543. DOI: 10.3115/v1/D14-1162.
- [31] Kingma D P, Ba J. Adam: A method for stochastic optimization. Proceedings of the the 3rd International Conference for Learning Representations. San Diego, 2015; arXiv: 1412.6980. DOI: 10.48550/arXiv.1412.6980.
- [32] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need. NIPS '17; Proceedings of the 31st International Conference on Neural Information Processing Systems. New York; ACM, 2023; 6000 – 6010. DOI: 10.48550/arXiv.1706.03762.