Rational Sensibility: LLM Enhanced Empathetic Response Generation Guided by Self-presentation Theory

Linzhuang Sun[2,3], **Nan Xu**[1*], **Jingxuan Wei**[2,3], **Bihui Yu**[2,3], **Liping Bu**[2,3], **Yin Luo**[1]
[1]Beijing Wenge Technology Co., Ltd
[2]Shenyang Institute of Computing, Chinese Academy of Science, shenyang, China
[3]University of Chinese Academy, Beijing, China
sunlinzhuang21@mails.ucas.ac.cn

Abstract

Having the ability to empathize is crucial for accurately representing human behavior during conversations. Despite numerous research aim to improve the cognitive capability of models by incorporating external knowledge, there has been limited attention on the sensible and rational expression of the conversation itself, which are crucial components of the cognitive empathy. Guided by self-presentation theory in sociology, we have designed an innovative categorical approach that segregates historical dialogues into sensible and rational sentences and subsequently elucidate the context through the designed attention mechanism. However, The rational information within the conversation is restricted and the external knowledge used in previous methods have limitations of semantic contradiction and narrow vision field. Considering the impressive performance of LLM in the domain of intelligent agent. We employ LLaMA2-70b as a rational brain to analyze the profound logical information maintained in conversations, which assists the model assessing the balance of sensibility and rationality to produce quality empathetic responses. Experimental evaluations demonstrate that our method outperforms other comparable methods on both automatic and human evaluations.

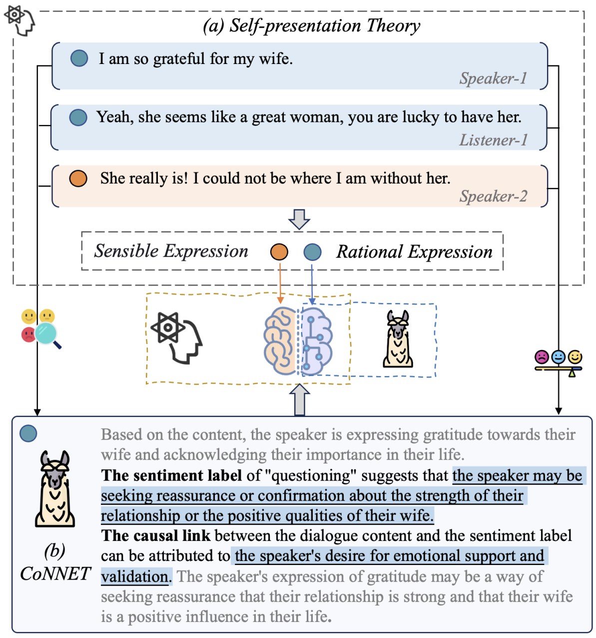

Figure 1: An example of empathetic response from EMPATHETICDIALOGUES dataset. (a) According to the sociological theory of self-presentation. Categorizing conversation sentences into sensible and rational groups to enhance the expression of the dialog's semantic message. (b) the Chain of Emotion-aware Empathetic prompting based on LLaMA2-70b.

1 Introduction

Empathetic response generation, the capacity to perceive the emotions of individuals and react accordingly, is integral in the pursuit of intelligent agents(Hofmann et al., 2010). As the core of individual emotions, sensibility and rationality play a significant role in empathy ability. Lack of sensibility makes it challenging to emotionally relate to users, And conversely, lack of rationality may result in emotional empathy and the symptom of unmitigated communion(Fritz and Helgeson, 1998). However, Rational Sensibility, known as cognitive empathy, allows for a better comprehension of the user while decreasing own negative emotional encounters(Smith, 2006). Although many research endeavors enhance the cognitive ability of models through the incorporation of external knowledge(Ghosal et al., 2020; Zhou et al., 2021; Sabour et al., 2022), limited attention has been directed towards the sensible and rational expression inherent within the conversational itself.

In sociology, the Self-presentation theory (Baumeister and Hutton, 1987; Jensen Schau and Gilly, 2003) emphasizes the utilization of emotionally charged or expressive language to accentuate the individual's personality traits and sensible con-

* Corresponding author, nan.xu@wenge.com

dition, thus effectively depict their personal state and projecting a desired character. Reversely, a distinct set of rational sentences primarily serves the purpose of transmitting information, expressing opinions, and engaging in rational thought processes. These sentences aim to maintain objectivity by focusing on delivering factual content and employing logical reasoning. As shown in Figure 1, we utilize the pre-trained emotion-cause model to identify sentences that are sensibly charged and combine them with objective-neutral sentences using designed attention mechanism.

However, the rational thinking of the dialog itself provides limited content and cannot provide the deep information hidden by the dialog, such as intention and purpose. Considering the method employed in previous research, such as COMET(Bosselut et al., 2019) and NRC_VAD(Mohammad, 2018), possess two limitations. Firstly, **semantic contradiction**. To guarantee a varied range of generated data, the incorporation of an external knowledge model may give rise to conflicting outcomes among multiple generated results. Secondly, **narrow field of vision**. Most research endeavors are limited to focusing on a single utterance to expanding external knowledge for a specific role.

In the field of intelligent agent, LLM has achieved impressive performance as the vital component of the system for information processing and decision making(Wang et al., 2023b; Shen et al., 2023; Hong et al., 2023). Inspired by that, we employed the LLM as a rational brain to provide logical reasoning for empathetic replies, so as to help the model balance sensible and rational cognition, and further make empathetic replies that are more consistant with the listener's role. Particularly, we have devised and implemented an innovative approach called the **C**hain **o**f Emotio**n**-awar**e** Empatheti**c** promp**t**ing(CoNECT) within the task of generating empathetic responses, as shown in Figure 1. CoNECT employs an affective indicator to facilitate the evaluation and correlation of contextual connections, relying on the extensive knowledge within LLM. This approach enables empathetic reasoning and boasts notable advantages compared to prior external sources of information: 1) **sensible reasoning**: The capability to examine how emotional labels are expressed and used in historical context, and analyze the underlying psychological processes happening. 2) **Global vision**: A comprehensive perspective in terms of perceiving the emotions and intentions of different roles within dialogues on a global scale.

To this end, we present **LLM** enhanced emp**a**thetic response generation **m**odel guided **b**y self-presentation theory, **Lamb**. The model consists of three modules: 1) **sensible knowledge enhancement**: Given the historical conversation, we utilize the emotional label annotated by a sentiment classifier SEEK (Wang et al., 2022) as a reference point and conduct reasoning using the LLaMA2-70b large-scale language model to grasp the sentiment understanding conveyed in the surrounding context. In addition, we continue to employ the COMET model to enhance conventional common sense knowledge; 2) **sense knowledge enhancement**: Considering the problems of bias and hallucination in LLM, we took measures to reduce the negative effects of CoNECT data on the accuracy of the original conversation history. Drawing upon the principles of self-presentation theory from the field of sociology, we employed the EMOTION-CAUSE approach to distinguish sensible and rational sentence expressions within the given context. Subsequently, we accomplished the collaborative modeling of sentences attributed with sensible and rational components by utilizing a designed attention mechanism; and 3) **sense-sensibility aware decoder**: We employ a cross-attention mechanism that allows the decoder to perceive CoNECT data, COMET commonsense knowledge, and weighted historical conversation data. Moreover, we integrate the knowledge selection mechanism suggested by Cai et al. (2023) to further enhance semantic coherence.

Our main contributions are listed as follows:

- Guided by sociological theories of self-presented theory, we focus on the sensible and rational expression of the conversation itself to enhance cognitive empathy.
- We introduce the chain of emotion-aware empathetic data into the empathetic response generation task, which provides sensible reasoning and role-aware sensible reasoning in the global field of view.
- We propose a novel model, Lamb. Lamb uses external knowledge and weighted context to understand speaker's emotional state and generate appropriate empathetic responses.
- Experiments demonstrate that Lamb generates

more empathetic responses compared with the state-of-the-art methods.

2 Related Work

The objective of the empathetic response generation task is to equip the model with the capability to deliver suitable emotional values to users while generating empathetic responses. Numerous researchers have presented a range of strategies to tackle this task.

Rashkin et al. (2019) open source the EMPATHETICDIALOGUES dataset for the empathetic response generation task by scientifically training the annotators and standardizing the evaluation of the model's effectiveness. Sabour et al. (2022) introduces commonsense knowledge inference to this task for the first time by enriching the commonsense information in the historical conversations with the pre-trained COEMT model. Li et al. (2022) on the other hand, also considers the introduction of external knowledge, but unlike the standard structure of Sabour et al. (2022), it uses the structure of graph neural networks for task modeling.

In terms of the fine-grained aspect of encoding, Wang et al. (2022) considers that the sentiment labels of each dialog round in the data can help the model capture user sentiment more accurately, and proposes a multi-granularity sentiment prediction loss function that jointly supports model learning. In addition, Kim et al. (2022) further extracts the key information of word dimensions to realize more accurate empathetic responses.

In addition, Chen and Liang (2022) uses the emotional cause recognition algorithm guided by psychological theory to successfully uncover key utterances in the conversation process. Meanwhile, Zhao et al. (2023) not only perceives the emotions of others, but also evaluates its own emotional state. Qian et al. (2023a), on the other hand, innovatively breaks down the empathetic response task into a two-step task, first confirming the semantic information to be responded to, and then adding the emotional coloring in the second step.

3 Method

We provide a general overview of Lamb in Figure 2. Lamb consists of three models: 1) sensible knowledge enhancement, 2) rational representation enhancement, 3) sensibility-rationality balance decoder.

3.1 Task Formulation

Given a dialogue history $U = [S_1, L_1, S_2, L_2, ..., L_{N-1}, S_N]$ of $2N$-1 utterances, the goal of the empathetic response generation is to predict the overall sentiment label e and generate empathetic response $Y = [y_1, y_2, ..., y_{m_y}]$. In dialogue history U, $S_i = [s^i_1, s^i_2, ..., s^i_{m_{s_i}}]$ and $L_i = [l^i_1, l^i_2, ..., l^i_{m_{l_i}}]$ represent the $i-$th utterance of speaker and listener,composed by m_{s_i} and m_{l_i} tokens respectively.

3.2 Sense Representation Enhancement

Drawing on the theory of self-presentation within the realm of sociology, we segregate the historical context into two groups. The first group encompasses sentences that have been discovered to underscore individual characteristics by sensible expression, through the utilization of a pretrained emotion-cause model RECCON(Poria et al., 2021). On the other hand, the second group encompasses sentences that primarily emphasize the rational reasoning. When selecting sentences that emphasize specific characteristics, our purpose is to align the sensible experiences of both the speaker and listener through global sentiment markers, which enables the listener to better understand the speaker's emotion and respond more empathy. And since global sentiment labels are not provided in the training data, We utilize the Welivita and Pu (2020) as a reference and apply the SEEK(Wang et al., 2022) model to assign comprehensive sentiment labels for each conversation.

According to Figure 2, we set U as input of the SEEK model, and obtain predicted global sentiment label e_{ano} accordingly. Next, we use e_{ano} as the target sentiment of the model RECCON, and look for cues that evoke the sentiment in U. The final output is the sensible sentences D that implies the expression of the sentiment, which is a subset of U.

$$e_{ano} = SEEK(U) \tag{1}$$

$$D = RECCON(e_{ano}, U) \tag{2}$$

We feed U and D into the Bart encoder to obtain the context representation $R_U \in \Re^{l \times d}$ and the cause representation $R_D \in \Re^{s \times d}$:

$$R_U = Bart_{enc}(U) \tag{3}$$

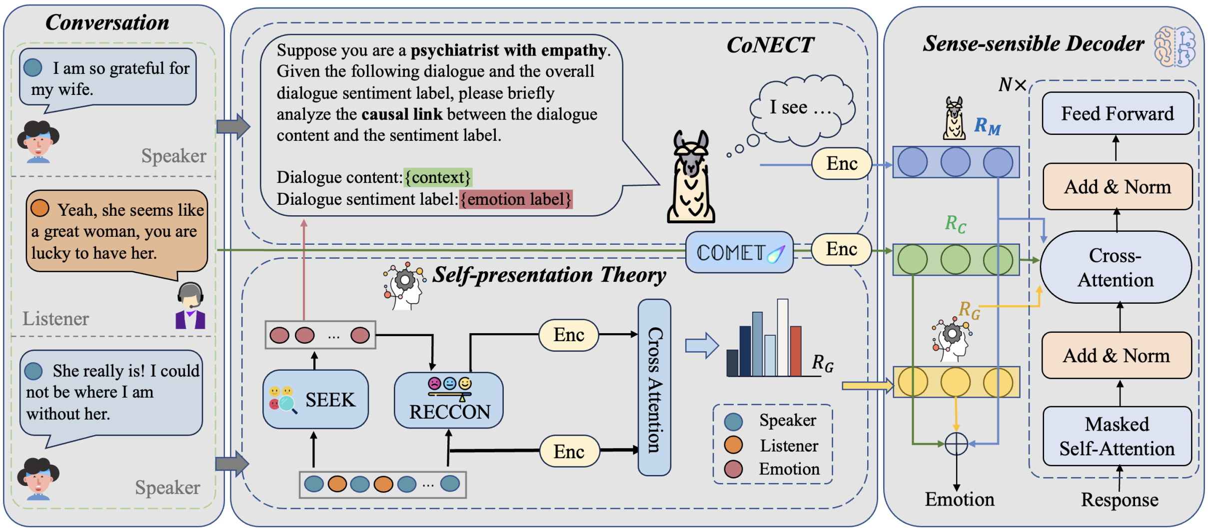

Figure 2: The model architecture of Lamb, which consists of three parts: (a) Sensible Representation Enhancement, (b) Rational Representation Enhancement by CoNECT and COMET, (c) Sensibility-rationality Balance Decoder.

$$R_D = Bart_{enc}(D) \tag{4}$$

Next, to assist the model in discerning between these two categories of sentences in historical dialogues, we incorporate the method of joint modeling through the employment of the attention mechanism to assign distinct weights to them, and output finally representation $R_G \in \Re^{l \times d}$.

$$R_G = softmax(\frac{(W_q R_U) \cdot (W_k R_D)^T}{\sqrt{2d}}) \cdot (W_v R_D) \tag{5}$$

, where $W_q \in \Re^{d \times d}$, $W_k \in \Re^{d \times d}$ and $W_v \in \Re^{d \times d}$ are randomly initialised tensor matrix.

3.3 Sensible Representation Enhancement

In this module, we dedicated to enhancing external knowledge by CoNECT data and COMET's common sense knowledge extension.

3.3.1 COMET

Following previous works, we present the pre-trained model COMET to generate common sense knowledge, which is constructed based on two general knowledge graphs, ATOMIC(Sap et al., 2019) and ConceptNet(Speer et al., 2017), which cover a huge amount of social general knowledge. Given the last utterance S_N of the historical context U, we alternately select xIntent, xEffect, xWant, xReact and xNeed cognitive attributes for inference, obtaining C_r commonsense knowledge where $r \in \{xIntent, xEffect, xWant, xReact, xNeed\}$:

$$C_r = COMET(S_N) \tag{6}$$

Meanwhile, in order to take into account the semantic diversity of the generated content, we embed the commonsense knowledge in the decoder, so as to provide more effective complementary information to the model.

3.3.2 CoNECT

The CoT technology of large language models has shown significant application in many domains. In the task of empathetic response generation, the CoT data can compensate the deficiencies existed in previous common sense knowledge, such as contextual semantic conflict and narrow field of view. By bootstrapping the massive knowledge of large models, our designed CoNECT can provide contextually semantically coherent, grammatically structured and complementary information.

As shown in Figure 2, the prompt we used can be divided into three parts:

Character. We have assigned LLaMA2 to cosplay the character of a compassionate psychologist who is expected to possess theoretical knowledge and behavioral paradigms in professional domain.

Causal Chain. In order to improve contextual coherence and achieve a more precise comprehension of dialogues, we extended the scope of external knowledge sources by incorporating the complete text into the prompt.

Global Sentiment Label. In conjunction with the Causal Chain, it offers guidance to logical framework of the analytical model.

In practice, we use the prompt C_P as the input of LLaMA2-70b, and the output of the LLM is the CoNECT data C_M.

$$C_P = Prompt_{template}(U, e_{ano}) \tag{7}$$

$$C_M = LLaMA2 - 70b(C_P) \tag{8}$$

Afterwards, we pass C_M through the encoder in order to acquire unified feature representation $R_M \in \Re^{c \times d}$, where c is the length of CoNECT data.

$$R_M = Bart_{enc}(C_M) \tag{9}$$

3.4 Emotion Classification

For each COMET relation sequence, we append the $[CLS]$ token to the first position, then fed into the encoder getting feature matrix $R_C \in \Re^{l_{intent} \times d}$.

$$R_C = Bart_{enc-r}([CLS] \oplus C_r) \tag{10}$$

After that, we utilize average pooling to get the knowledge vector $p_k \in \Re^d$:

$$p_k = Average - pooling(R_C) \tag{11}$$

To optimize knowledge acquisition, we strive for the fusion of historical conversations, empathetic thought chains, and common sense knowledge.

$$R_F = Bart_{enc}(R_G \oplus R_M \oplus p_k), \tag{12}$$

$$P_{e_{pre}} = softmax(W_\theta(Average(R_F))) \tag{13}$$

We subsequently pass R_F through a linear layer $W_\theta \in \Re^{d \times q}$, followed by a softmax calculation to produce the emotion category distribution $P_{e_{pre}} \in \Re^q$:

During training, we optimize these weights by minimizing the Cross-Entropy loss between the emotion category and the ground truth label e_{tar}:

$$\pounds_{emo} = -log(P_{e_{pre}}(e_{tar})) \tag{14}$$

3.5 Sense-sensibility Aware Decoder

In terms of response generation, the sensible rich historical dialogue information and the CoT data of LLMs provide the model with sensible and rational messages respectively. In addition, to control the semantic collision between different COMET knowledge, we adopt the refiner method introduced by Cai et al. (2023). Particularly, given COMET representation, the refiner uses the competition and broadcasting mechanism to obtain a unified feature vector δ_m, which is the input of the next decoder module after concatenation with h_k. h_k is the residual from the previous block.

The target response $Y = [y_1, y_2, ..., y_M]$ with length M, which is generated token by token by the Bart decoder based above representation C_R and E_F. In training time, we adopt the standard negative log-likelihood(NLL) loss on the target response Y:

$$\pounds_{nll} = -\sum_{M}^{t=1} \log(y|(C_R, E_F), y_{<t}) \tag{15}$$

3.6 Training Objectives

In the training phase, all the parameters of our model are optimized with $\pounds_{emo}$ and $\pounds_{nll}$:

$$\pounds = \pounds_{nll} + \pounds_{emo} \tag{16}$$

4 Experiments

4.1 Datasets

We evaluate the model effects on the EMPATHETICDIALOGUES dataset(Rashkin et al., 2019), a dataset containing twenty-five thousand sets of multi-round dialogues, each containing two roles: speaker and listener, and global sentiment labels and multi-round dialogues alternating between the two roles. There are a total of 32 uniformly distributed sentiment labels, with an average of 4.31 dialogue rounds per dataset and an average length of 15.2 words per dialogue.

To ensure the fairness of the comparison experiments, we used the same dataset division as the previous research method, dividing the dataset into training, validation and test sets in an 8:1:1 ratio.

4.2 Baselines

We compare our method with the following state-of-art baselines: 1) Transformer(Vaswani et al., 2017): An original Transformer, which is trained to optimize the negative log-likelihood loss. 2) Multi-TRS(Rashkin et al., 2019): A variation of the Transformer optimized for emotion, where the joint generation loss and emotion classification loss are jointly optimized. 3) MoEL(Lin et al., 2019): The model designs a corresponding decoder for each emotion to generate a response, and all the

Table 1: Results of automatic evaluation. Optimal outcomes are highlighted in bold, and suboptimal outcomes are indicated by underlining within each assessment indicator.

Models	PPL ↓	B-1 ↑	B-2 ↑	B-3 ↑	B-4 ↑	R-1 ↑	R-2 ↑	Dist-1 ↑	Dist-2 ↑	Acc ↑
Transformer	37.62	18.07	8.34	4.57	2.86	17.22	4.21	0.36	1.35	-
Multi-TRS	37.50	18.78	8.55	4.70	2.95	16.85	4.21	0.35	1.27	33.95
MoEL	36.60	18.07	8.30	4.37	2.65	18.24	4.81	0.59	2.64	31.74
MIME	37.24	18.60	8.39	4.54	2.81	17.08	4.05	0.47	1.66	30.96
EmpDG	37.43	19.96	9.11	4.74	2.80	18.02	4.43	0.46	1.99	31.65
CEM	36.33	16.12	7.29	4.06	2.03	15.77	4.50	0.62	2.39	36.84
SEEK	36.78	10.77	4.40	2.02	1.08	12.74	2.94	0.68	2.81	42.74
CASE	35.20	15.59	7.22	3.80	2.24	17.33	4.67	0.65	3.37	38.99
E-CORE	33.03	-	-	-	-	-	-	0.72	3.49	42.59
KEMP	36.39	16.72	7.17	3.77	2.33	16.11	3.31	0.66	3.07	36.57
CAB	34.36	19.23	8.55	4.36	2.57	17.50	4.13	1.13	4.23	40.52
DCKS	**18.58**	18.75	9.12	5.38	3.57	19.14	5.45	1.57	6.02	48.69
Lamb	18.86	22.00	10.49	**6.07**	**3.97**	19.55	5.47	1.80	7.73	**53.44**
w/o CoNECT	20.30	21.69	10.38	5.95	3.83	**19.81**	**5.49**	1.98	8.79	48.35
w/o Self-pres	20.26	**22.28**	**10.69**	6.05	3.85	19.50	5.33	**2.29**	**10.17**	51.01

Table 2: Results of human evaluation.

Models	Coh.	Emp.	Inf.	Cont.
EmpDG	3.22	3.10	2.99	3.07
CEM	3.41	3.49	3.12	3.18
SEEK	3.40	3.62	3.19	3.33
KEMP	3.56	3.66	3.35	3.52
DCKS	4.10	3.94	3.76	3.87
Lamb	**4.34**	**4.17**	**4.76**	**4.46**

results are synthesised to produce the final result. 4) MIME(Ghosal et al., 2020): Depending on the positive or negative polarity of the emotion and the contextual content, the model is able to simulate the user's emotions and enable the generation of empathetic responses by introducing randomness. 5) EmpDG(Li et al., 2020): A model that includes an empathetic information generator and a sentiment discriminator. The function of the generator is to ensure that the model has the ability to generate diverse content, while the function of the discriminator is to ensure that the information produced by the generator matches the empathetic sentiment in the context. 6) CEM(Sabour et al., 2022): The first model introduces the COMET pre-trained model to acquire common sense knowledge in empathetic response generation task, and uses both intent and cognitive types to categorise the knowledge. 7) SEEK(Wang et al., 2022): A model focuses on sentence-level sentiment information and utilizes attention mechanisms to provide the model with multi-granularity sentiment label information. 8) CASE(Zhou et al., 2023): A model employ external message, COEMT and ConceptNet, to enhance the ability of cognitive and emotion. 9) E-CORE(Fu et al., 2023): A method fouces on exploring intrinsic sentiment by emotion correlation learning, utilization, and supervising. 10) KEMP(Li et al., 2022): A model that uses ConceptNet and VRC-NED as external knowledge sources and performs contextual modelling through a graph neural network structure. 11) CAB(Gao et al., 2023): A model split the empathy response generation into three parts: cognition, affection and behavior. 12) DCKS(Cai et al., 2023): A model uses adaptive module for commonsense knowledge selection to ensure consistency between the model responses and the history context.

Table 3: Results of Human A/B test.

Models	Win	Loss	Tie
Lamb vs EmpDG	48.8%	16.3%	34.9%
Lamb vs CEM	40.2%	19.0%	40.8%
Lamb vs SEEK	41.7%	19.5%	38.8%
Lamb vs KEMP	45.3%	18.5%	36.2%
Lamb vs DCKS	37.4%	20.3%	42.3%

4.3 Implementation Details

Table 4: Compared with LLM.

Models	BLEU-2	BLEU-4
LLaMA2-7B	5.94	1.38
ChatGPT(+ 0-short)	6.19	1.86
ChatGPT(+ 1-short)	6.79	2.12
ChatGPT(+ 5-short)	7.85	2.65
GPT-3(+ 0-short)	6.88	2.22
GPT-3(+ 1-short)	6.71	2.16
GPT-3(+ 5-short)	8.51	2.83
GPT-3.5(+ 0-short)	8.51	2.80
GPT-3.5(+ 1-short)	5.62	1.99
GPT-3.5(+ 5-short)	9.37	3.26
Lamb	**10.49**	**3.97**

We implement all the models using PyTorch and use the encoder and decoder from base version of BART in our work as the same as Cai et al. (2023). The emotion cot data are derived from LLaMA2-70b model with Nvidia-A100 GPU and we train Lamb models using Adam optimizer with initial learning rate 0.00005 on Nvidia-4090 GPU. In practice, we set batch size to 16 and we use the same 8:1:1 train/valid/test dataset split as provided by Cai et al. (2023).

4.4 Evaluation Metrics

4.4.1 Automatic Evaluation

Following previous work, we choose Perplexity(PPL), corpus-level BLEU(B-n), sentence-level ROUGE(R-n) and Distinct-n(Dist-n) as our main automatic metrics. 1) Perplexity is used to assess the fluency and comprehensibility of the text generated by the model, and a lower PPL value means that the text is more coherent and natural. 2) BLEU and ROUGE describe the similarity of the text, the higher the score, the more similar the text generated by the model is to the ground-truth text.3) Dist-n is used to assess the diversity of the text, with higher metric values indicating that the model is able to produce more diverse and richer representation. 4) Acc is used to measure the accuracy of the model for emotion classification.

4.4.2 Human Evaluation

To further assess the quality of the response generated by Lamb, we use a manual evaluation method and keep same with previous research to ensure fairness. Particularly, we randomly select 100 sets of conversations from the data and give them to three graduate students for evaluation. The rating scale included the following three dimensions and each of which was on a scale of 1 to 5. 1) Coherence(Coh.): Measure how relevant the text generated by the model is to inputs. 2) Empathy(Emp.): Used to measure the extent to which the model's responses understand and express human emotions and opinions. 3) Informativeness(Inf.): Measuring the amount of information contained in the responses. 4) Continuity(Con.): The intensity of the speaker's desire to continue the conversation.

4.5 Automatic Evaluation Results

Table 1 shows the overall evaluation results on automatic metrics. Our model outperforms the baselines in every metric. The remarkable improvement in the DIST-2 metric suggests that our model responses are more diverse, mitigating the appearance of hollow platitudes. And the highest scores on the BLEU-n, ROUGE-N metrics reflect better consistency between model responses and historical conversations, while also maintaining diversity. In terms of emotion classification, our model improves by three points, demonstrating the ability of our method to capture the sentiment of the speaker.

4.6 Ablation Studies

In order to test the validity of the affective thought chain and self-presentation theory, we conducted several ablation experiments, as shown at the end of Table 1. The following are several variant models:

1) **w/o CoNECT**: the cot data derived from LLM is removed.

2) **w/o Self-pres**: the self-presentation module used for emotion enhancement is removed.

Removing the self-presentation theory resulted in a 19.6% decrease in the model's emotion classification accuracy, along with a corresponding decrease in BLEU-N and Dist-N scores, suggesting that the direct introduction of sentiment thought chain data results in the model neglecting to analyse the history of the conversation itself.

4.7 Human Evaluation Results

Table 2 reports the results of human evaluation.

Benefiting from the CoNECT, the model's ability to empathize is enhanced, and the volume of information in the response is also increased. Furthermore, reinforced by self-presentation theory, the model is able to focus on the sensibility clues in history content, which improves the contextual

Table 5: An example of EMPATHETICDIALOG and the responses generated from xxx and other SOTA models. Blue text and orange text denotes the rational and emotional expression separately.

Context	Speaker: one of the times i remember feeling the most blissed out in life was right after the birth of my first child. Listener: that is a very blessed day. it is something you will never forget. Speaker: obviously there are the demands of a new child – but that feeling of finally meeting someone you waited so long for, and the love surrounding the whole situation. truly something to remember.
Emotion **CoNECT**	Grateful Based on the content of the dialogue, it is clear that the **speaker** is expressing feelings of gratitude and contentment. The use of phrases such as "most blissed out in life" and "something you will never forget" suggest a strong **positive emotion**. Additionally, the speaker mentions the birth of their **first child**, which is a significant life event that can evoke powerful emotions. The sentiment label "grateful" accurately captures the speaker's emotional state. Overall, the causal link between the dialogue content and the sentiment label is strong and straightforward.
KEMP **SEEK** **DCKS**	I am sure you will do great ! That is a great accomplishment ! That is great to hear. i am happy for you.
Lamb	Yes, it is a great feeling to finally meet someone you waited so long for.
Gold	Yes, I could not agree more. it is remarkable how your feeling suddenly change.

coherence of the generated response and thus motivates the Speaker's willingness to continue the conversation. Additionally, the pairwise response comparison results are shown in Table 3. The results also confirm that the responses generated by Lamb are more preferred by human judges.

4.8 Comparing with LLM

Furthermore, we compared Lamb with strong LLMs. As shown in Table 4, Our model outperforms compared methods in BLUE-2 and BLUE-4. For fairness, we maintain the setting of GPT-series LLM are consistent of Qian et al. (2023b), and ensure that the configuration of LLaMA2-7B aligns with Wang et al. (2023a).

4.9 Case Study

The responses generated by Lamb and other models are shown in Table 5. The response generated by KEMP is not relevant to the history of the dialog. Although SEEK and DCKS responses expressed congratulations to the speaker, they did not establish a connection to concrete things in the conversation and therefore contained relatively few semantic information.

In our method, CoNECT analyzes the speaker's state of mind from a global perspective, and speculates that "the happiest thing in life" and "something you will never forget" are the key words in the conversation, which add color to the joy brought by the birth of the first child. The keywords emphasize the joy of the birth of the first child. By further integrating the self-presentation theory, Lamb recognized the speaker's joyfulness while closely focusing on the key messages in the conversation, which effectively stimulated the speaker's willingness to continue the communication.

5 Conclusions

In this paper, we improve empathetic response generation by introducing emotional inference data and utilizing self-presentation theory to enhance the semantic representation of history context. The standardized automatic and human evaluation demonstrate the high quality of generated response from Lamb, and further ablation experiments demonstrate the effectiveness of our modules.

References

Roy F. Baumeister and Debra G. Hutton. 1987. Self-Presentation Theory: Self-Construction and Audience Pleasing, pages 71–87. Springer New York.

Antoine Bosselut, Hannah Rashkin, Maarten Sap, Chaitanya Malaviya, Asli Celikyilmaz, and Yejin Choi. 2019. COMET: Commonsense transformers for automatic knowledge graph construction. In in Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, pages 4762–4779.

Hua Cai, Xuli Shen, Qing Xu, Weilin Shen, Xiaomei Wang, Weifeng Ge, Xiaoqing Zheng, and Xiangyang Xue. 2023. Improving empathetic dialogue generation by dynamically infusing commonsense knowledge. In Findings of the Association for Computational Linguistics: ACL 2023, pages 7858–7873.

Yangbin Chen and Chunfeng Liang. 2022. Wish I can feel what you feel: A neural approach for empathetic response generation. In Findings of the Association for Computational Linguistics: EMNLP 2022, pages 922–933.

Heidi L Fritz and Vicki S Helgeson. 1998. Distinctions of unmitigated communion from communion: self-neglect and overinvolvement with others. Journal of Personality and Social Psychology, 75:121.

Fengyi Fu, Lei Zhang, Quan Wang, and Zhendong Mao. 2023. E-core: Emotion correlation enhanced empathetic dialogue generation. In in Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing, pages 10568–10586.

Pan Gao, Donghong Han, Rui Zhou, Xuejiao Zhang, and Zikun Wang. 2023. Cab: Empathetic dialogue generation with cognition, affection and behavior. In Database Systems for Advanced Applications: 28th International Conference, DASFAA 2023, page 597–606.

Deepanway Ghosal, Navonil Majumder, Alexander Gelbukh, Rada Mihalcea, and Soujanya Poria. 2020. COSMIC: COmmonSense knowledge for eMotion identification in conversations. In Findings of the Association for Computational Linguistics: EMNLP 2020, pages 2470–2481.

Stefan G Hofmann, Alice T Sawyer, and Angela Fang. 2010. The empirical status of the “new wave” of cognitive behavioral therapy. Psychiatric Clinics, 33(3):701–710.

Sirui Hong, Xiawu Zheng, Jonathan Chen, Yuheng Cheng, Jinlin Wang, Ceyao Zhang, Zili Wang, Steven Ka Shing Yau, Zijuan Lin, Liyang Zhou, et al. 2023. Metagpt: Meta programming for multi-agent collaborative framework. arXiv preprint arXiv:2308.00352.

Hope Jensen Schau and Mary C Gilly. 2003. We are what we post? self-presentation in personal web space. Journal of consumer research, 30:385–404.

Wongyu Kim, Youbin Ahn, Donghyun Kim, and Kyong-Ho Lee. 2022. Emp-RFT: Empathetic response generation via recognizing feature transitions between utterances. In in Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pages 4118–4128.

Qintong Li, Hongshen Chen, Zhaochun Ren, Pengjie Ren, Zhaopeng Tu, and Zhumin Chen. 2020. EmpDG: Multi-resolution interactive empathetic dialogue generation. In in Proceedings of the 28th International Conference on Computational Linguistics, pages 4454–4466.

Qintong Li, Piji Li, Zhaochun Ren, Pengjie Ren, and Zhumin Chen. 2022. Knowledge bridging for empathetic dialogue generation. In in Proceedings of the AAAI Conference on Artificial Intelligence, volume 36, pages 10993–11001.

Zhaojiang Lin, Andrea Madotto, Jamin Shin, Peng Xu, and Pascale Fung. 2019. MoEL: Mixture of empathetic listeners. In 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), pages 121–132.

Saif Mohammad. 2018. Obtaining reliable human ratings of valence, arousal, and dominance for 20,000 English words. In in Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, volume 1, pages 174–184.

Soujanya Poria, Navonil Majumder, Devamanyu Hazarika, Deepanway Ghosal, Rishabh Bhardwaj, Samson Yu Bai Jian, Pengfei Hong, Romila Ghosh, Abhinaba Roy, Niyati Chhaya, et al. 2021. Recognizing emotion cause in conversations. Cognitive Computation, 13:1317–1332.

Yushan Qian, Bo Wang, Shangzhao Ma, Wu Bin, Shuo Zhang, Dongming Zhao, Kun Huang, and Yuexian Hou. 2023a. Think twice: A human-like two-stage conversational agent for emotional response generation. In in Proceedings of the 2023 International Conference on Autonomous Agents and Multiagent Systems, page 727–736.

Yushan Qian, Weinan Zhang, and Ting Liu. 2023b. Harnessing the power of large language models for empathetic response generation: Empirical investigations and improvements. In Findings of the Association for Computational Linguistics: EMNLP 2023, pages 6516–6528.

Hannah Rashkin, Eric Michael Smith, Margaret Li, and Y-Lan Boureau. 2019. Towards empathetic open-domain conversation models: A new benchmark and dataset. In in Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, pages 5370–5381.

Sahand Sabour, Chujie Zheng, and Minlie Huang. 2022. Cem: Commonsense-aware empathetic response generation. In in Proceedings of the AAAI Conference on Artificial Intelligence, volume 36, pages 11229–11237.

Maarten Sap, Ronan Le Bras, Emily Allaway, Chandra Bhagavatula, Nicholas Lourie, Hannah Rashkin, Brendan Roof, Noah A Smith, and Yejin Choi. 2019. Atomic: An atlas of machine commonsense for if-then reasoning. In in Proceedings of the AAAI conference on artificial intelligence, volume 33, pages 3027–3035.

Yongliang Shen, Kaitao Song, Xu Tan, Dongsheng Li, Weiming Lu, and Yueting Zhuang. 2023. Hugginggpt: Solving ai tasks with chatgpt and its friends in huggingface. arXiv preprint arXiv:2303.17580.

Adam Smith. 2006. Cognitive empathy and emotional empathy in human behavior and evolution. The Psychological Record, 56:3–21.

Robyn Speer, Joshua Chin, and Catherine Havasi. 2017. Conceptnet 5.5: An open multilingual graph of general knowledge. In in Proceedings of the AAAI conference on artificial intelligence, volume 31, page 4444–4451.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Ł ukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In in Proceedings of the 31st Conference on Neural Information Processing System, volume 30.

Lanrui Wang, Jiangnan Li, Zheng Lin, Fandong Meng, Chenxu Yang, Weiping Wang, and Jie Zhou. 2022. Empathetic dialogue generation via sensitive emotion recognition and sensible knowledge selection. In Findings of the Association for Computational Linguistics: EMNLP 2022, pages 4634–4645.

Lanrui Wang, Jiangnan Li, Chenxu Yang, Zheng Lin, and Weiping Wang. 2023a. Enhancing empathetic and emotion support dialogue generation with prophetic commonsense inference. arXiv preprint arXiv:2311.15316.

Lei Wang, Chen Ma, Xueyang Feng, Zeyu Zhang, Hao Yang, Jingsen Zhang, Zhiyuan Chen, Jiakai Tang, Xu Chen, Yankai Lin, et al. 2023b. A survey on large language model based autonomous agents. arXiv preprint arXiv:2308.11432.

Anuradha Welivita and Pearl Pu. 2020. A taxonomy of empathetic response intents in human social conversations. In in Proceedings of the 28th International Conference on Computational Linguistics, pages 4886–4899.

Weixiang Zhao, Yanyan Zhao, Xin Lu, and Bing Qin. 2023. Don't lose yourself! empathetic response generation via explicit self-other awareness. In Findings of the Association for Computational Linguistics: ACL 2023, pages 13331–13344.

Jinfeng Zhou, Chujie Zheng, Bo Wang, Zheng Zhang, and Minlie Huang. 2023. CASE: Aligning coarse-to-fine cognition and affection for empathetic response generation. In Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics, pages 8223–8237.

Pei Zhou, Pegah Jandaghi, Bill Yuchen Lin, Justin Cho, Jay Pujara, and Xiang Ren. 2021. Probing commonsense explanation in dialogue response generation. In Findings of the Association for Computational Linguistics: EMNLP 2021, pages 4132–4146.