Emotion Recognition in Conversation via Dynamic Personality

Yan Wang[1], **Bo Wang**[1,2*], **Yachao Zhao**[1], **Dongming Zhao**[3], **Xiaojia Jin**[3], **Jijun Zhang**[3], **Ruifang He**[1], **Yuexian Hou**[1]
[1] College of Intelligence and Computing, Tianjin University, Tianjin, China
[2] Institute of Applied Psychology, Tianjin University, Tianjin, China
[3] China Mobile Communication Group Tianjin Co.,Ltd
{wy_31, bo_wang, zhaoyachao, rfhe, yxhou}@tju.edu.cn
waitman_840602@163.com
{jinxiaojia, zhangjijun}@tj.chinamobile.com

Abstract

Emotion recognition in conversation (ERC) is a field that aims to classify the emotion of each utterance within conversational contexts. This presents significant challenges, particularly in handling emotional ambiguity across various speakers and contextual factors. Existing ERC approaches have primarily focused on modeling conversational contexts while incorporating only superficial speaker attributes such as names, memories, and interactions. Recent works introduce personality as an essential deep speaker factor for emotion recognition, but relies on static personality, overlooking dynamic variability during conversations. Advances in personality psychology conceptualize personality as dynamic, proposing that personality states can change across situations. In this paper, we introduce ERC-DP, a novel model considering the dynamic personality of speakers during conversations. ERC-DP accounts for past utterances from the same speaker as situation impacting dynamic personality. It combines personality modeling with prompt design and fine-grained classification modules. Through a series of comprehensive experiments, ERC-DP demonstrates superior performance on three benchmark conversational datasets.

Keywords: Emotion Recognition, Dynamic Personality, Big Five Personality

1. Introduction

Emotion is a intricate construct that encapsulates an individual's psychological state, interweaving their thoughts, feelings, and behaviors. It frequently finds expression in natural language, serving as a reflection of one's emotional condition. With the advent of instant messaging and social media, the text data in conversational format has witnessed a significant surge. As a result, Emotion Recognition in Conversation (ERC) plays an increasingly important role in the realm of Natural Language Processing (NLP). It aims to deduce the emotion of the speaker engaged in a conversation with one or multiple interlocutors (Kim and Vossen, 2021).

Unlike vanilla emotion recognition of plain text, ERC relies on conversational contexts, which is what early research focused more on. In recent studies, more approaches have expanded the scope of analysis, taking into account speaker factors such as the speaker's names, conversational dependencies, and memory traces to discern the emotion within each speaker's utterance. For instance, DialogueGCN (Ghosal et al., 2019) leverages self and inter-speaker dependency, and CoMPM (Lee and Lee, 2022) extracts the speaker's memory to represent the speaker's knowledge. Additionally, recent researches note that personality is a deep factor in recognizing an individual's emotion, asserting their significance over shallow individual factors (Wen et al., 2021; Chen et al., 2022).

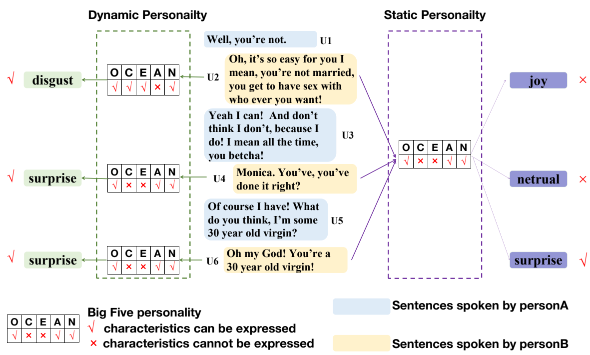

Figure 1: The comparison between the dynamic personality and static personality. Dynamic personality assigns a corresponding personality state to each utterance. Static personality utilize the same personality profile to represent a given speaker across all their utterances. Compared with dynamic personality, static personality will lead to inaccurate emotion prediction.

However, there are two limitations for personality-based method. Firstly, personality modeling has not been incorporated into ERC tasks. Secondly, more importantly, prevailing methods in other fields of emotion recognition overwhelmingly rely on static personality(Wen et al., 2021). This hypothesis largely disregards the role of momentary expres-

*Corresponding Author: bo_wang@tju.edu.cn

LREC-COLING 2024, pages 5711–5722

sions of personality. Meanwhile, there is a substantial body of research evidence supports the existence of short-term within-person variations in personality (Fleeson, 2007; Fleeson and Gallagher, 2009). Recent theoretical advances in personality psychology posit that personality can best be conceptualized as a dynamic system (Sosnowska et al., 2020). And dynamic personality theories are best suited to predict behavior, affect, and situational experiences (Hecht et al., 2023).

Dynamic personality theories incorporate stability and variability of personality, stemming from the person-situation-debate (Fleeson and Noftle, 2009). This theoretical framework emerged in response to the observations that static personality traits are inadequate predictors of momentary behavior. Instead, situational factors have demonstrated superior efficacy in explaining behavioral manifestations at a specific moment. Such momentary expressions of personality in particular situations are termed as **personality states** (Baumert et al., 2017). Personality states delves into an individual's present state, contrasting with the broader, general aspect of their dispositions (Schutte et al., 2003). Empirical findings reveal that personality states swiftly, extensively, and significantly fluctuate within the typical individual, while also showing associations with the characteristics of given situations (Fleeson, 2007). It is evident that situations wield a significant influence on dynamic personality, given their role in modulating the intra-individual variations in cognitive, affective, and behavioral responses (Beckmann and Wood, 2017).

In ERC tasks, predicting the emotion of the current utterance constitutes a form of momentary behavior. Therefore, when utilizing personality as a speaker-specific factor, it is logical to utilize its momentary expression, specifically, personality states. Given the significant role of situations in explaining personality states, and as Lewis (Lewis, 1999) argued, the most effective way to study personality change is to scrutinize behavior in context. In ERC tasks, it is conceptually intuitive to consider a speaker's past utterances of current utterance as the situation within the ongoing conversation. This approach enables the computation of the personality states of the current utterance attributed to the same speaker, thus providing a more granular and contextually rich analysis of the speaker's emotion.

To this end, we introduce ERC-DP, a novel framework for emotion recognition in conversation guided by dynamic personality. We choose the Big Five model of personality because it has been proved to have cross-cultural applicability and has been widely used. Our model comprises three components: a personality recognition module, a prompt design module, and a fine-grained classification module. (1) In the personality recognition module, a Big Five personality classifier is trained using the Essays dataset (Pennebaker and King, 1999). We obtain a dynamic personality state by considering the past utterances of the same speaker in the preceding conversation. (2)In the prompt design module, we embed the personality state as a prompt into context and encode it using a pretrained language model. (3) Finally, inspired by BERT-ERC (Qin et al., 2023), we divide the input features into three parts according to special tokens: past features, current features, and future features and perform classification to obtain emotional labels. It is worth noting that the personality prompt is the part of the current feature.

To verify the effectiveness of our method, we conduct evaluations of ERC-DP using three benchmark datasets for ERC: IEMOCAP (Busso et al., 2008), MELD (Poria et al., 2019), and EmoryNLP (Zahiri and Choi, 2017). The experimental results demonstrate that our ERC-DP outperforms currently advanced methodologies in the domain of emotion recognition in conversation.

Our contributions are summarized as follows:

(1) By amalgamating dynamic personality theories with ERC, we deviate from the conventional use of static personality, choosing instead dynamic personality states. This involves incorporating past utterances as situations, which represent personality states as momentary responses in particular situations to predict emotion.

(2) We introduce a novel concept of a personality prompt, which amalgamates personality states using prompt with contextual information. This allows for the adaptation of our proposed framework across diverse scenarios.

(3) Our model demonstrates superior performance on three benchmark datasets, setting new benchmarks in terms of results. The experimental verification not only substantiates the effectiveness of our approach, but also highlights the significant contribution of our dynamic personality strategy.

2. Related Work

Traditional strategies for Emotion Recognition in Conversation can be classified into two groups based on the usage of speaker information as additional features for classification: Content Modeling methods and Speaker Modeling methods.

Content Modeling methods primarily analyze the content of the conversation. HiGRU (Jiao et al., 2019) employs a lower-level GRU to model the word-level inputs and an upper-level GRU to capture the contexts of utterance-level embeddings. DAG-ERC (Shen et al., 2021b) formulates a directed acyclic graph to capture the information exchange between the broader conversational background and the immediate contextual elements.

Speaker Modeling methods not only consider the context but also introduce speaker information as additional features. The majority of contemporary approaches adhere to this methodology. Existing works typically incorporate speaker names into the context. EmoBERTa (Kim and Vossen, 2021) incorporates the context of the query utterance and speakers' names into PLM during finetuning, allowing it to delve into contextual information. EmotionFlow (Song et al., 2022) encodes utterances by concatenating the context and names with an auxiliary question to learn speaker-specific features. And some works focus on the correlation between speakers. DialogueGCN (Ghosal et al., 2019) leverages self and inter-speaker dependency of the speakers to model conversational context for emotion recognition. COSMIC (Ghosal et al., 2020) learns interactions between speakers participating in a conversation and leads towards a better understanding of the emotional dynamics and other aspects of the conversation. SGED (Bao et al., 2022) designs a novel speaker modeling scheme that explores intra- and inter-speaker dependencies, effectively exploiting speaker information. Although some works do not utilize direct speaker information, speaker-level information can also be reflected in the model, such as DialogueCRN (Hu et al., 2021) and DialogueTRM (Mao et al., 2021), they utilize speaker-level information indirectly through speaker-level clues or context.

3. Methodology

3.1. Problem Statement

Given a conversation, denoted as C, consisting of M utterances and S speakers. Emotion Recognition in Conversation aims to predict the emotion of each utterance across a predetermined emotion set Y. The conversation can be represented as a sequence, $C = [(u_1, s_1), ..., (u_M, s_M)]$, where each utterance u_i, comprises several tokens. Each utterance u_i is spoken by a unique speaker $s_i \in S$, which is annotated with an emotion label $y_i^e \in Y$.

3.2. Overall Architecture

Our model is designed around three key modules: the personality recognition module, the prompt design module, and the fine-grained classification module, as illustrated in Figure 2. The personality recognition module captures the dynamic personality of the current query utterance, taking into account the past utterances of the same speaker. The output personality state is a 5-D vector, representing the Big Five personality. In the prompt design module, we convert personality state into a prompt text and integrate it with context. This boosts model flexibility while reducing compute costs compared to handling personality as a separate feature. In the fine-grained classification module, we divide the input features according to special tokens into past features, current features, and future features to obtain position-aware results. Since the personality state refers to the current utterance, we only embed the prompt text into the current features. The final classification result is derived from the combination of these three features.

3.3. Personality Recognition Module

Due to the lack of personality annotations in most emotion recognition datasets, our personality recognition model, denoted as F_p, is constructed using the personality recognition dataset Essays (Pennebaker and King, 1999). This dataset employs the Big Five Model to annotate each sample, yielding in a 5-D personality score where each dimension has values of either 0 or 1.

Given the training set $\{x_i, y_i^p\}_{i=1}^N$, x_i is a sentence in the dataset and $y_i^p \in R^5$ is the personality score of x_i. We build a prompt-based sentence encoder upon BERT (Devlin et al., 2019) to obtain personality feature representations. We construct a prompt $Prompt_p =$ *The personality is* $[MASK]$. to prepend to the input in order to obtain better results.The full input of the encoder is $Prompt_p \oplus x_i$, where $\oplus$ is the concatenation operation. We first feed input into PLM and get the classification vector $H_{\text{cls}} \in R^d$ as a representation of dynamic personality, and d is the dimension of a token embedding. We employ a fully connected layer followed by the ReLU activation function for projection. Subsequently, we incorporate a Dropout layer (Srivastava et al., 2014) to mitigate overfitting, followed by an MLP for classification. In other words, the prediction $\hat{y}_i^p$ can be computed as follows:

$$\hat{y}_i^p = MLP(Dropout(RELU(FC(H_{\text{cls}})))) \quad (1)$$

The model is optimized using the Binary Cross-Entropy loss function, which is commonly used for binary classification tasks. y represents the true target value, while p corresponds to the predicted probability, indicative of the likelihood that the sample belongs to the positive class. The loss is calculated as follows:

$$Loss_p = -(y \cdot \log p) + (1 - y) \cdot \log(1 - p)) \quad (2)$$

Then, we employ the trained personality recognition model F_p, derived above, to predict the personality state of each utterance in the emotion dataset. Given the significance of the situation in the dynamic personality theory, we represent the situation using the concatenation of the current query utterance u_i and the past utterances $C = \{u_k | 0 <= k < i, s_k = s_i\}$ from same speaker s_i. The concatenated utterances

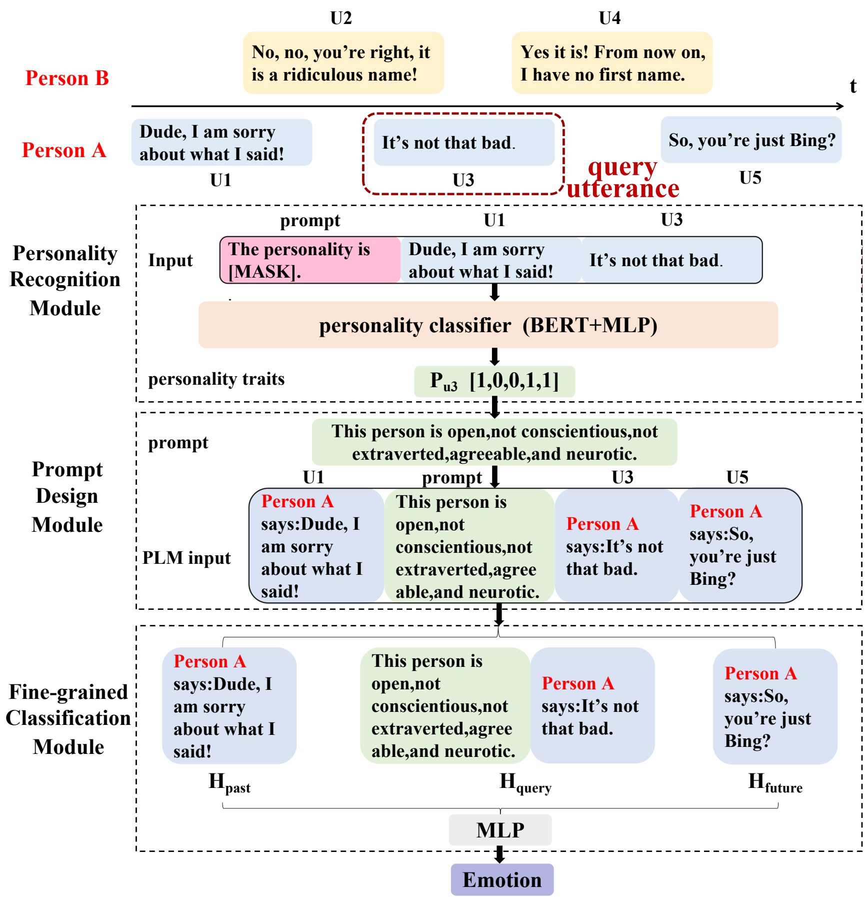

Figure 2: The architecture of our ERC-DP: In the personality recognition module, we obtain personality state. Then, in the prompt design module, we flexibly integrate dynamic personality prompt with content. In the fine-grained classification module, we finally classify by combining past, current, and future features.

is $X_i = [c_1, ..., c_n, u_i], c_1, ..., c_n \in C$, n is the number of past utterances by the current query utterance of same speaker s_i. Therefore, the input of each utterance is $Prompt_p \oplus X_i$, and the personality state assigned to the current utterance u_i is $P_{\mathrm{u_i}} = F_p(Prompt_p \oplus X_i)$, $P_{\mathrm{u_i}} \in R^5$, which represents the 5-dimensional dynamic personality.

3.4. Prompt Design Module

In the prompt design module, we first convert the personality states $P_{\mathrm{u_i}}$ obtained based on the personality recognition module into a $Prompt_{\mathrm{u_i}}$. The Big Five dimensions of openness, conscientiousness, extraversion, agreeableness, and neuroticism are translated into corresponding adjectives, which are open, conscientious, extraverted, agreeable, and neurotic. If the corresponding value of the factor is 1, then we use the corresponding adjective. If it is 0, we prepend `not` before the corresponding adjective. We use openness as an example:

$$P_{\mathrm{u_i-open}} = \begin{cases} 0 & \textit{not open} \\ 1 & \textit{open} \end{cases} \tag{3}$$

Therefore, given an OCEAN personality state of $[1, 0, 0, 1, 1]$, the prompt $Prompt_{\mathrm{u_i}}$ of this utterance u_i is "This person is open, not conscientious, not extraverted, agreeable, and neurotic."

To further concentrate on the information from the same speaker, we only consider the context from the same speaker when incorporating context information. When constructing the input, we incorporate as much context as possible, the construction of the input is detailed in Algorithm 1.Through this algorithm, we can obtain input for classifying the current sentence u_i, which includes dynamic personality and context.

3.5. Fine-grained Classification Module

We feed input into SimCSE (Gao et al., 2021) to obtain the last hidden vector $H_{\mathrm{total}} \in R^{l*d}$, l is the number of tokens and d is the dimension of a token embedding. We divide tokens l into three parts according to the position of $<$ *s* $>$ and $<$ */s* $>$, representing past features, query features and future features. Given $[f_1, ...f_l]$ denote the features of input, the start and end positions of

Dataset	Conversations			Utterances		
	Train	Val	Test	Train	Val	Test
IEMOCAP	120		31	5810		1623
MELD	1038	114	280	9989	1109	2610
EmoryNLP	713	99	85	9934	1344	1328

Table 1: Statistics for the three datasets.

the query feature are a and b respectively. We derive past features H_{past}, query features H_{query}, and future features H_{future} by taking the mean of the corresponding token embeddings as follows:

$$H_{\text{past}} = \text{mean}(f_1, \ldots, f_{a-1}) \quad (4)$$

$$H_{\text{query}} = \text{mean}(f_a, \ldots, f_b) \quad (5)$$

$$H_{\text{future}} = \text{mean}(f_{b+1}, \ldots, f_l) \quad (6)$$

We concatenate past features H_{past}, query features H_{query}, and future features H_{future} to obtain the position-aware context representations.

$$H_{\text{result}} = [H_{\text{past}}, H_{\text{query}}, H_{\text{future}}] \quad (7)$$

H_{result} is passed to a following classifier to predict emotion $\hat{y}_i^e$ using Focal Loss (Lin et al., 2017).

$$\hat{y}_i^e = MLP(Dropout(Tanh(FC(H_{\text{result}})))) \quad (8)$$

4. Experiments

4.1. Experimental Setup

In the personality recognition module, we fine-tune the BERT model for a batch size of 16 and apply the Adam optimizer (Kingma and Ba, 2014) with a learning rate of 5e-4. In the prompt design module, we employ SimCSE, with the first eight encoder layers frozen as the PLM, and apply Adam optimizer with a learning rate of 9e-6. We select the model that demonstrates the best performance on the validation set. All experiments are conducted on a computing infrastructure equipped with four P100 GPUs with 16GB memory.[1]

4.2. Datasets

We train the personality recognition module on the Essays dataset (Pennebaker and King, 1999), which contains 2468 anonymous essays tagged with the author's personality traits. These stream-of-consciousness essays were penned by volunteers in a controlled environment, who were subsequently asked to label their own Big Five personality traits. The personality labels only employ 0 and 1 to represent whether the sentence can express the corresponding characteristics.

[1]Code: `https://github.com/sevenn31/ERC-DP`.

Algorithm 1 Building an input of current query utterance

Require: $[(u_1, s_1), ..., (u_M, s_M)]$:the conversation
u_i:the current query utterance
max_length: the max length
lst_p:the list of past utterances
L_p:the length of lst_p
lst_f:the list of future utterances
L_f:the length of future lst_f
Initialize empty lists lst_p and lst_f,$num = 0$
$input =$ < s > $+ Prompt_{u_i} + s_i +$ *says:* $+ u_i +$ < /s >
for j from $i-1$ to 0 **do**
  if $s_i = s_j$ **then**
    $X_{\text{past}} = s_j +$ *says:* $+ u_j$
    $\text{Append} X_{\text{past}}$ to lst_p
  end if
end for
for k from $i+1$ to M **do**
  if $s_i = s_k$ **then**
    $X_{\text{future}} = s_k +$ *says:* $+ u_k$
    $\text{Append} X_{\text{future}}$ to lst_f
  end if
end for
while $num < L_p$ or $num < L_f$ **do**
  if $num < L_p$ **then**
    if $\text{len}(lst_p[num] + input) < max\_length$ **then**
      $input = lst_p[num] + input$
    else
      break
    end if
  end if
  if $num < L_f$ **then**
    if $\text{len}(input + lst_f[num]) < max\_length$ **then**
      $input = input + lst_f[num]$
    else
      break
    end if
  end if
  $num = num + 1$
end while
return $input$

For ERC tasks, we benchmark ERC-DP on three conversational emotion recognition datasets: MELD (Poria et al., 2019), EmoryNLP (Zahiri and Choi, 2017) and IEMOCAP (Busso et al., 2008). The statistics of datasets are presented in Table 1.

MELD is a dataset based on the Friends TV show, comprising more than 1400 conversations and 13000 utterances labeled with one of seven emotions. The emotion classes include anger, disgust, sadness, joy, surprise, fear, and neutral.

EmoryNLP is another dataset also based on the Friends TV show. This dataset encompasses 97 episodes, 897 scenes, and 12606 utterances, each annotated with one of seven emotions: neutral, joyful, peaceful, powerful, scared, mad, and sad.

IEMOCAP is a dataset of two-person conversations among ten speakers. The dataset encompasses data collected from five separate sessions, featuring interactions between five pairs of speakers. Each spoken utterance encapsulates one of six distinct emotional states: neutral, happiness, sadness, anger, frustration, and excited.

Evaluation Metrics: Given the class imbalance present in all three benchmarks, we adopt the weighted-F1 score as the evaluation metric for all experiments in this paper.

4.3. Baselines

We classify the model into two types based on the usage of speaker information as additional features for classification: Content Modeling methods and Speaker Modeling methods. So we select representative models among these two types as baselines, especially Speaker Modeling methods, which closely align with our approach.

Content Modeling methods: **HiGRU** (Jiao et al., 2019) utilizes HiGRU-f and HiGRU-sf to obtain word level inputs and the long-range contextual information. **DAG-ERC** (Shen et al., 2021b) utilizes a directed acyclic neural network to model the information flow between long-distance conversation background and nearby context.

Speaker Modeling methods: **DialogueRNN** (Majumder et al., 2019) models the speaker identity, historical conversation, and emotions of past utterances using RNNs. **COSMIC** (Ghosal et al., 2020) presents a new framework that incorporates various elements of commonsense and builds upon them to learn interactions between interlocutors participating in a conversation. **DialogXL** (Shen et al., 2021a) relies on enhanced memory to store longer historical context and captures intra- and inter- speaker dependencies with self-attention. **DialogueCRN** (Hu et al., 2021) designs multi-turn reasoning modules to extract and integrate speaker clues and situation clues, executing an intuitive retrieval process and a deliberate reasoning process. **SKAIG** (Li et al., 2021) enhances targeted utterances with information inferred from the past context and intentions implied by the future context between speakers. **EmoBERTa** (Kim and Vossen, 2021) models contextual information by simply prepending speaker names to utterances and inserting separation tokens between the utterances in a conversation. **CoMPM** (Lee and Lee, 2022) uses a special token to distinguish the speaker in CoM and tracks speaker's pre-trained memory in PM. **SACL** (Hu et al., 2023) employs situation-aware and speaker-aware features to form the context representation, applies contrast-aware adversarial training to generate worst-case samples and uses joint class-spread contrastive learning to extract structured representations. **BERT-ERC** (Qin et al., 2023) explores contextual information with speaker names and adapts the PLM to the task. **EmotionIC** (Yingjian et al., 2023) captures global contextual dependencies with identity information, extracts speaker- and temporal-aware local contextual information, and employs SkipCRF to simulate emotional propagations.

5. Results and Analysis

5.1. Main Results

We compare ERC-DP with state-of-the-art text-based ERC methods, and the results are shown in Table 2. Upon reviewing existing methodologies, it can be seen that the Speaker Modeling methods are better than the Content Modeling methods, whether directly modeling speaker traits or implicitly capturing speaker information. Our method ERC-DP, leverages dynamic personality as speaker information and also exhibits remarkable performance across all three datasets.

For the MELD dataset, ERC-DP surpasses the current state-of-the-art method, BERT-ERC, by a margin of 0.23% in terms of weighted F1 score. For the EmoryNLP dataset, ERC-DP surpasses the current state-of-the-art method, EmotionIC, by a margin of 0.99% in terms of the weighted F1 score. Moreover, on the IEMOCAP dataset, ERC-DP demonstrates superior performance over most of the baseline models, with an improvement of 69.64%. We acknowledge that our results on the IEMOCAP dataset are not the best, so we delve deeper into the reasons behind this in section 5.5, providing a detailed analysis of the results.

5.2. Ablation Study

We also conducted an ablation study on three datasets to validate the effectiveness of our proposed theory of dynamic personality for the ERC task, and the results are shown in Table 3.

The first row in the table displays the results derived in the absence of a personality recognition module, relying solely on contextual information. The second row displays the performance when utilizing static personality. The term "static personality" refers to the personality recognition achieved through the aggregation of all utterances from a speaker within a conversation. Consequently, in a given conversation, all utterances from the same speaker are characterized with the same static personality. The last row displays our proposed dynamic personality, which identifies personality

Type	Method	MELD	EmoryNLP	IEMOCAP
Content Modeling methods	HiGRU (Jiao et al., 2019)	56.81	34.48	58.54
	DAG-ERC (Shen et al., 2021b)	63.65	39.02	68.03
Speaker Modeling methods	DialogRNN+RoBERTa (Majumder et al., 2019)	63.20	37.75	63.92
	COSMIC (Ghosal et al., 2020)	65.21	38.11	65.28
	DialogueCRN (Hu et al., 2021)	63.42	38.91	66.33
	SKAIG (Lee and Lee, 2022)	65.18	38.88	66.98
	DialogXL (Shen et al., 2021a)	62.41	34.73	66.20
	EmoBERTa (Kim and Vossen, 2021)	66.51	-	68.57
	COMPM (Lee and Lee, 2022)	66.52	38.93	69.46
	EmotionIC (Yingjian et al., 2023)	66.40	40.01	69.61
	SACL (Hu et al., 2023)	66.45	39.65	69.22
	BERT-ERC (Qin et al., 2023)	67.11	39.84	**71.70**
	ERC-DP(our model)	**67.34**	**40.10**	69.64

Table 2: Comparison with the state-of-the-art methods on three datasets (%).

Method	MELD	EmoryNLP	IEMOCAP
No Personality	66.70	39.10	68.14
Static Personality	67.08	39.47	69.02
Dynamic Personality	**67.34**	**40.10**	**69.64**

Table 3: Ablation study on three datasets (%).

	O	C	E	A	N
MELD	70.20	84.60	82.00	78.20	65.00
EmoryNLP	75.40	81.20	82.40	80.40	63.80
IEMOCAP	63.40	75.60	71.80	79.60	67.20

Table 4: The accuracy between personality model and manual assessments (%).

states by combining the current query utterance with past utterances from the same speaker in the conversation. In this way, even the same speaker will show different personalities due to different situations (Fleeson, 2001), which the theory of dynamic personality also shows. Based on the data presented in Table 3, it is clear that the incorporation of personality information, whether static or dynamic, significantly enhances performance. Notably, dynamic personality method yields superior outcomes compared to static personality method. Overall, the results demonstrate the effectiveness of our model of dynamic personality.

5.3. The Study of Personality Recognition Module

To evaluate the generalization performance of this personality model, we randomly selected one hundred utterances from the test sets of the three datasets and manually annotated them for the Big Five personality traits. Notably, the assessment was not performed on individual conversational utterances but rather on the concatenation of past utterances and the current query utterance from the same speaker. Table 4 compares our model ERC-DP predictions to human judgments across the three datasets.

Based on the personality assessment results in Table 4, it is evident that the human evaluations on the MELD dataset and the EmoryNLP dataset closely align with the model outcomes, particularly in terms of conscientiousness, extroversion, and agreeableness. However, the results in the IEMOCAP dataset generally exhibit lower alignment. This discrepancy could be due to the increased length of the utterances used as situation, which potentially intensifies the ambiguity of dynamic personality assessment. Additionally, the IEMOCAP dataset differs from the other two as it is derived from actual conversations rather than scripted TV shows, resulting in more mundane utterances. Across all three datasets, neuroticism tends to be consistently low, which might be attributed to its inherent difficulty in assessing accurately.

5.4. Case Study

In an effort to delve deeper into the significance of dynamic personality, we performed an analysis by extracting specific conversations, as illustrated in Figure 3. The current query utterance, denoted as $U2$, is spoken by Rachel, which is "No. Sorry". When predicting static personality, we use all of Rachel's utterances in this conversation as input, essentially concatenating $U2$ and $U6$. The resulting static Big Five personality can express openness, conscientiousness, and agreeableness, predicting a neutral emotion. On the other hand, when considering dynamic personality, we only use past utterances with the same speaker before the

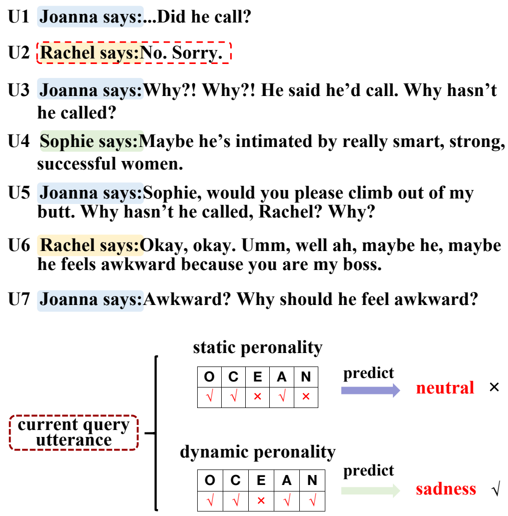

Figure 3: Using a conversation in the MELD dataset as the case study, we find that for the current query utterance, the personality state obtained by not considering subsequent utterances is better than the static personality using the global context.

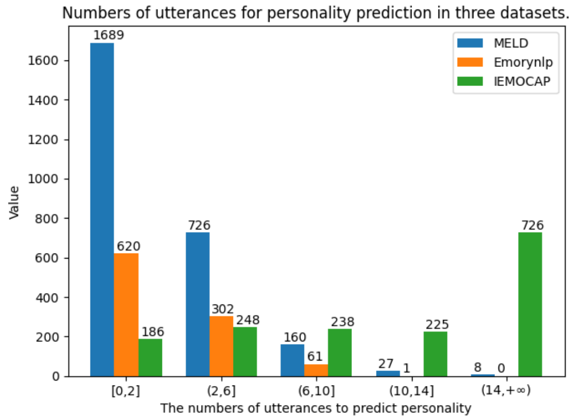

Figure 4: Utterance numbers for personality prediction of current utterance in test set of the three datasets. The X-axis represents range of utterance numbers when predicting dynamic personality for the current utterance, while the Y-axis represents the number of data points in the test set.

query utterance. For the current query utterance $U2$, using only $U2$ as input in this conversation, the resulting dynamic Big Five personality can express openness, conscientiousness, agreeableness, and neuroticism. And the predicted emotion is sadness, which is the same as the true label.

The key difference is the additional neuroticism captured in the dynamic personality, which is absent in the static personality. This is because dynamic personality does not consider $U6$, that is, "Okay, okay. Umm, well ah, maybe he, maybe he feels awkward because you are my boss". In other words, for the momentary behavior of predicting the query utterance $U2$, if the subsequent utterance $U6$ is considered, the personality that is meant to represent the current speaker's information becomes skewed, ultimately leading to erroneous emotion prediction results. In essence, this succinct example illustrates that a dynamic personality approach is more effective than a static personality one.

5.5. The Discussion of IEMOCAP

Although our model achieved a result of 69.64% on the IEMOCAP dataset, surpassing the majority of existing models, it did not attain optimal performance. We attribute this to potential inaccuracies in the dynamic personality predictions for the current query utterance, which could stem from two primary factors: the length of conversations and the nature of the conversations in IEMOCAP dataset.

Firstly, the conversations in IEMOCAP are longer and only have two speakers. We use the past utterances of the query utterance as the situation to predict dynamic personality, which means that as a conversation progresses, the input text for prediction becomes increasingly substantial. The figure 4 shows the number of utterances of the same speaker as a situation when predicting dynamic personality, t is the number of utterances. It is evident that the IEMOCAP dataset contains a greater number of longer utterances used as contexts for personality prediction. When obtaining dynamic personality from these extensive contexts, the risk of ambiguity escalates potentially impact our model's performance.

Secondly, we find that the nature of conversations in the IEMOCAP dataset, being derived from real-world conversations, presents an added layer of complexity compared to the MELD and EmoryNLP datasets, which are based on scripted conversations from movies and TV shows. Real-life conversations are subject to a broader range of influences, including background knowledge and environmental contexts, making them more diverse and intricate. Consequently, accurately capturing and representing the nuanced internal changes in these conversations using the Big Five personality traits becomes a challenging work.

6. Conclusion

In this paper, we propose ERC-DP, a novel methodology that integrates dynamic personality theory into Emotion Recognition in Conversation task. Our model comprises three modules: a personality recognition module, a prompt design module, and a fine-grained classification module, all of which work in harmony to extract personality traits and contextual features. Extensive experiments conducted on three widely recognized benchmarks and our approach performs well. In addition, through designed ablation experiments, we validate the superior effectiveness of dynamic personality over static personality. And our analysis of the IEMOCAP dataset reveals limitations in our model's performance with long conversations.

In our future research, we will delve into attention mechanisms and other approaches to address the limitations of our model when applied to long conversations. Additionally, we will take into account various conversational factors like that impact the situation to yield more precise dynamic personality results, ultimately improving the overall accuracy of emotion prediction.

Limitations

This work has three limitations: (1) Excessively long utterances can undermine the precision of our dynamic personality prediction, consequently impacting the outcomes of emotion forecasting. (2) Numerous factors influence the situation, and we must also account for contextual elements, information about other speakers, and more. (3) While the Big Five personality traits serve as a valuable initial framework, they may not offer a comprehensive portrayal of the speaker.

Ethical Considerations

All models in this paper are trained on the public corpus. The used datasets do not contain personal information or unethical language. We recruit evaluators to evaluate the results of personality experiments, and require them to be able to identify the Big Five personalities contained in the sentences. Compensation is $5 per hour. We also ensure the anonymization of the human evaluation.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (62376188, 62272340, 62276187, 62376192, 62166022), Yunnan Fundamental Research Projects (Grant No. 202101AT070077).

References

Yinan Bao, Qianwen Ma, Lingwei Wei, Wei Zhou, and Songlin Hu. 2022. Speaker-guided encoder-decoder framework for emotion recognition in conversation. In *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI-22*, pages 4051–4057. International Joint Conferences on Artificial Intelligence Organization. Main Track.

Murray R Barrick and Michael K Mount. 1991. The big five personality dimensions and job performance: a meta-analysis. *Personnel psychology*, 44(1):1–26.

Anna Baumert, Manfred Schmitt, Marco Perugini, Wendy Johnson, Gabriela Blum, Peter Borkenau, Giulio Costantini, Jaap JA Denissen, William Fleeson, Ben Grafton, et al. 2017. Integrating personality structure, personality process, and personality development. *European Journal of Personality*, 31(5):503–528.

Nadin Beckmann and Robert E Wood. 2017. Dynamic personality science. integrating between-person stability and within-person change.

Carlos Busso, Murtaza Bulut, Chi-Chun Lee, Abe Kazemzadeh, Emily Mower, Samuel Kim, Jeannette N Chang, Sungbok Lee, and Shrikanth S Narayanan. 2008. Iemocap: Interactive emotional dyadic motion capture database. *Language resources and evaluation*, 42:335–359.

Yirong Chen, Weiquan Fan, Xiaofen Xing, Jianxin Pang, Minlie Huang, Wenjing Han, Qianfeng Tie, and Xiangmin Xu. 2022. Cped: A large-scale chinese personalized and emotional dialogue dataset for conversational ai. *arXiv preprint arXiv:2205.14727*.

Vishal Chudasama, Purbayan Kar, Ashish Gudmalwar, Nirmesh Shah, Pankaj Wasnik, and Naoyuki Onoe. 2022. M2fnet: Multi-modal fusion network for emotion recognition in conversation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4652–4661.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

William Fleeson. 2001. Toward a structure-and process-integrated view of personality: Traits as density distributions of states. *Journal of personality and social psychology*, 80(6):1011.

William Fleeson. 2007. Situation-based contingencies underlying trait-content manifestation in behavior. *Journal of personality*, 75(4):825–862.

William Fleeson and Patrick Gallagher. 2009. The implications of big five standing for the distribution of trait manifestation in behavior: fifteen experience-sampling studies and a meta-analysis. *Journal of personality and social psychology*, 97(6):1097.

William Fleeson and Erik E Noftle. 2009. In favor of the synthetic resolution to the person–situation debate. *Journal of Research in Personality*, 43(2):150–154.

Tianyu Gao, Xingcheng Yao, and Danqi Chen. 2021. SimCSE: Simple contrastive learning of sentence embeddings. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 6894–6910, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

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, Online. Association for Computational Linguistics.

Deepanway Ghosal, Navonil Majumder, Soujanya Poria, Niyati Chhaya, and Alexander Gelbukh. 2019. DialogueGCN: A graph convolutional neural network for emotion recognition in conversation. 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 154–164, Hong Kong, China. Association for Computational Linguistics.

Martin Hecht, Kai T Horstmann, Manuel Arnold, Ryne A Sherman, and Manuel C Voelkle. 2023. Modeling dynamic personality theories in a continuous-time framework: An illustration. *Journal of Personality*, 91(3):718–735.

Kai T Horstmann, John F Rauthmann, Ryne A Sherman, and Matthias Ziegler. 2021. Distinguishing simple and residual consistencies in functionally equivalent and non-equivalent situations: Evidence from experimental and observational longitudinal data. *European Journal of Personality*, 35(6):833–860.

Dou Hu, Yinan Bao, Lingwei Wei, Wei Zhou, and Songlin Hu. 2023. Supervised adversarial contrastive learning for emotion recognition in conversations. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 10835–10852, Toronto, Canada. Association for Computational Linguistics.

Dou Hu, Lingwei Wei, and Xiaoyong Huai. 2021. DialogueCRN: Contextual reasoning networks for emotion recognition in conversations. 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)*, pages 7042–7052, Online. Association for Computational Linguistics.

Wenxiang Jiao, Haiqin Yang, Irwin King, and Michael R. Lyu. 2019. HiGRU: Hierarchical gated recurrent units for utterance-level emotion recognition. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 397–406, Minneapolis, Minnesota. Association for Computational Linguistics.

Taewoon Kim and Piek Vossen. 2021. Emoberta: Speaker-aware emotion recognition in conversation with roberta. *CoRR*, abs/2108.12009.

Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Joosung Lee. 2022. The emotion is not one-hot encoding: Learning with grayscale label for emotion recognition in conversation. In *Interspeech*.

Joosung Lee and Wooin Lee. 2022. CoMPM: Context modeling with speaker's pre-trained memory tracking for emotion recognition in conversation. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 5669–5679, Seattle, United States. Association for Computational Linguistics.

Michael Lewis. 1999. On the development of personality. *Handbook of personality: Theory and research*, 2:327–346.

Jiangnan Li, Zheng Lin, Peng Fu, and Weiping Wang. 2021. Past, present, and future: Conversational emotion recognition through structural modeling of psychological knowledge. In *Findings of the Association for Computational Linguistics: EMNLP 2021*, pages 1204–1214, Punta Cana, Dominican Republic. Association for Computational Linguistics.

Jingye Li, Donghong Ji, Fei Li, Meishan Zhang, and Yijiang Liu. 2020. Hitrans: A transformer-based context-and speaker-sensitive model for emotion detection in conversations. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 4190–4200.

Zaijing Li, Fengxiao Tang, Ming Zhao, and Yusen Zhu. 2022. EmoCaps: Emotion capsule based model for conversational emotion recognition. In *Findings of the Association for Computational Linguistics: ACL 2022*, pages 1610–1618, Dublin, Ireland. Association for Computational Linguistics.

Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. 2017. Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision*, pages 2980–2988.

Navonil Majumder, Soujanya Poria, Devamanyu Hazarika, Rada Mihalcea, Alexander Gelbukh, and Erik Cambria. 2019. Dialoguernn: An attentive rnn for emotion detection in conversations. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, pages 6818–6825.

Yuzhao Mao, Guang Liu, Xiaojie Wang, Weiguo Gao, and Xuan Li. 2021. DialogueTRM: Exploring multi-modal emotional dynamics in a conversation. In *Findings of the Association for Computational Linguistics: EMNLP 2021*, pages 2694–2704, Punta Cana, Dominican Republic. Association for Computational Linguistics.

Goreti Marreiros, Ricardo Santos, Carlos Ramos, José Neves, and José Bulas-Cruz. 2008. Abs4gd: a multi-agent system that simulates group decision processes considering emotional and argumentative aspects. In *AAAI spring symposium on emotion, personality and social behaviour*. AAAI.

Aire Mill, Anu Realo, and Jüri Allik. 2016. Retrospective ratings of emotions: The effects of age, daily tiredness, and personality. *Frontiers in psychology*, 6:2020.

Daniel J Ozer and Veronica Benet-Martinez. 2006. Personality and the prediction of consequential outcomes. *Annu. Rev. Psychol.*, 57:401–421.

James W Pennebaker and Laura A King. 1999. Linguistic styles: language use as an individual difference. *Journal of personality and social psychology*, 77(6):1296.

Soujanya Poria, Devamanyu Hazarika, Navonil Majumder, Gautam Naik, Erik Cambria, and Rada Mihalcea. 2019. MELD: A multimodal multi-party dataset for emotion recognition in conversations. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 527–536, Florence, Italy. Association for Computational Linguistics.

Xiangyu Qin, Zhiyu Wu, Tingting Zhang, Yanran Li, Jian Luan, Bin Wang, Li Wang, and Jinshi Cui. 2023. Bert-erc: Fine-tuning bert is enough for emotion recognition in conversation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 13492–13500.

Nicola S Schutte, John M Malouff, Elida Segrera, Amanda Wolf, and Larissa Rodgers. 2003. States reflecting the big five dimensions. *Personality and Individual Differences*, 34(4):591–603.

Weizhou Shen, Junqing Chen, Xiaojun Quan, and Zhixian Xie. 2021a. Dialogxl: All-in-one xlnet for multi-party conversation emotion recognition. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 13789–13797.

Weizhou Shen, Siyue Wu, Yunyi Yang, and Xiaojun Quan. 2021b. Directed acyclic graph network for conversational emotion recognition. 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)*, pages 1551–1560, Online. Association for Computational Linguistics.

Xiaohui Song, Liangjun Zang, Rong Zhang, Songlin Hu, and Longtao Huang. 2022. Emotionflow: Capture the dialogue level emotion transitions. In *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 8542–8546.

Joanna Sosnowska, Peter Kuppens, Filip De Fruyt, and Joeri Hofmans. 2020. New directions in the conceptualization and assessment of personality—a dynamic systems approach. *European Journal of Personality*, 34(6):988–998.

Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. 2014. Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, 15(1):1929–1958.

Zhiyuan Wen, Jiannong Cao, Ruosong Yang, Shuaiqi Liu, and Jiaxing Shen. 2021. Automatically select emotion for response via personality-affected emotion transition. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 5010–5020, Online. Association for Computational Linguistics.

Liu Yingjian, Li Jiang, Wang Xiaoping, and Zeng Zhigang. 2023. Emotionic: Emotional inertia and contagion-driven dependency modelling for emotion recognition in conversation. *arXiv preprint arXiv:2303.11117*.

Sayyed M Zahiri and Jinho D Choi. 2017. Emotion detection on tv show transcripts with sequence-based convolutional neural networks. *arXiv preprint arXiv:1708.04299*.

Dong Zhang, Liangqing Wu, Changlong Sun, Shoushan Li, Qiaoming Zhu, and Guodong Zhou. 2019. Modeling both context-and speaker-sensitive dependence for emotion detection in multi-speaker conversations. In *IJCAI*, pages 5415–5421.

Duzhen Zhang, Xiuyi Chen, Shuang Xu, and Bo Xu. 2020. Knowledge aware emotion recognition in textual conversations via multi-task incremental transformer. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 4429–4440.

Lixing Zhu, Gabriele Pergola, Lin Gui, Deyu Zhou, and Yulan He. 2021. 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)*, pages 1571–1582, Online. Association for Computational Linguistics.