

# Personality-affected Emotion Generation in Dialog Systems

ZHIYUAN WEN, The Hong Kong Polytechnic University, China

JIANNONG CAO, The Hong Kong Polytechnic University, China

JIAXING SHEN\*, Lingnan University, China

RUOSONG YANG, The Hong Kong Polytechnic University, China

SHUAIQI LIU, The Hong Kong Polytechnic University, China

MAOSONG SUN, Tsinghua University, China

Generating appropriate emotions for responses is essential for dialog systems to provide human-like interaction in various application scenarios. Most previous dialog systems tried to achieve this goal by learning empathetic manners from anonymous conversational data. However, emotional responses generated by those methods may be inconsistent, which will decrease user engagement and service quality. Psychological findings suggest that the emotional expressions of humans are rooted in personality traits. Therefore, we propose a new task, Personality-affected Emotion Generation, to generate emotion based on the personality given to the dialog system and further investigate a solution through the personality-affected mood transition. Specifically, we first construct a daily dialog dataset, Personality EmotionLines Dataset (**PELD**), with emotion and personality annotations. Subsequently, we analyze the challenges in this task, *i.e.*, (1) heterogeneously integrating personality and emotional factors and (2) extracting multi-granularity emotional information in the dialog context. Finally, we propose to model the personality as the transition weight by simulating the mood transition process in the dialog system and solve the challenges above. We conduct extensive experiments on PELD for evaluation. Results suggest that by adopting our method, the emotion generation performance is improved by 13% in macro-F1 and 5% in weighted-F1 from the BERT-base model.

CCS Concepts: • **Computing methodologies** → **Discourse, dialogue and pragmatics**.

Additional Key Words and Phrases: dialog systems, emotion, personality

## ACM Reference Format:

Zhiyuan Wen, Jiannong Cao, Jiaxing Shen, Ruosong Yang, Shuaiqi Liu, and Maosong Sun. 2018. Personality-affected Emotion Generation in Dialog Systems. *J. ACM* 37, 4, Article 111 (August 2018), 27 pages. <https://doi.org/XXXXXXX.XXXXXXX>

## 1 INTRODUCTION

Emotional intelligence can be considered a mental ability to reason validly with emotional information and the action of emotions to enhance thought [44]. Empowering the dialog system with emotional intelligence can enhance plenty of applications like self-disclosure in AI mental therapy,

\*Corresponding author

Authors' addresses: Zhiyuan Wen, [cszwen@comp.polyu.edu.hk](mailto:cszwen@comp.polyu.edu.hk), The Hong Kong Polytechnic University, Hong Kong, China; Jiannong Cao, [csjcao@comp.polyu.edu.hk](mailto:csjcao@comp.polyu.edu.hk), The Hong Kong Polytechnic University, Hong Kong, China; Jiaxing Shen, [jiaxingshen@LN.edu.hk](mailto:jiaxingshen@LN.edu.hk), Lingnan University, Hong Kong, China; Ruosong Yang, [csryang@comp.polyu.edu.hk](mailto:csryang@comp.polyu.edu.hk), The Hong Kong Polytechnic University, Hong Kong, China; Shuaiqi Liu, [cssqliu@comp.polyu.edu.hk](mailto:cssqliu@comp.polyu.edu.hk), The Hong Kong Polytechnic University, Hong Kong, China; Maosong Sun, [sms@tsinghua.edu.cn](mailto:sms@tsinghua.edu.cn), Tsinghua University, Beijing, China.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

© 2018 Association for Computing Machinery.

0004-5411/2018/8-ART111 \$15.00

<https://doi.org/XXXXXXX.XXXXXXX>

smart customer services, and virtual human in the Metaverse. To achieve this goal, it is necessary to enable the machine to understand users' emotions, generate appropriate emotions, and express them coherently in conversations.

Despite existing works are relatively mature in recognizing the emotions in dialog context [15, 40, 60] and rendering specific emotions in responses [10, 80], the performance of generating appropriate emotions is still limited. Following the problem settings in empathetic responding [54], most existing studies [35, 76, 78, 79] infer appropriate emotions for the response without modeling the personality of speakers. Consequently, emotions in responses might be inconsistent with the preceding context. As a result, users may feel they are still talking to rigid machines and reduce their engagement; an example is shown in Figure 1.

To fill in the research gap, we propose a new task: Personality-affected Emotion Generation, which is to empower the dialog system with personality traits and generate emotion for the response so that the emotions will always be consistent with the given personality trait. Affective instability is a core feature of personality disorder [64], which inspires us that a stable personality trait for the dialog system helps solve the inconsistency in emotional responses. Besides, it is shown that the personality, *i.e.*, the big-five personality model [11] can be represented as temperament in the Valence-Arousal-Dominance (VAD) space for emotions [45, 46].<sup>1</sup> These findings suggest that different personalities make different impacts on emotional expressions. Moreover, similar to the one-to-many nature [77] of the dialogue, multiple emotions can be appropriate for response in a similar conversation context, but only one can be selected for the response each time in the dialog system. Accordingly, personality provides an additional reasoning condition to narrow down the searching space so that simplifies the emotion generation.

However, the proposed task entails the following two challenges when applied to reality. The first one is heterogeneously integrating personality and emotional factors. Personality plays an important role in inferring appropriate emotions as discussed above. According to the trait theory, personality is usually represented as dimensions or spectrums with extremes at both ends. However, the aspects described by the dimensions are defined in psychological analysis, which is different from the ways we describe the emotional factors in conversations (*e.g.*, discrete emotion labels, or dimensional affective vectors). So, how to integrate the personality with the emotional factors to facilitate our task is difficult. The second is the extracting multi-granularity emotional information (*i.e.*, emotional factors) from the dialog context. Affective information from conversational data facilitates the emotion generation process. Nevertheless, it can be captured in multiple aspects from conversational data, *e.g.*, utterance-level semantic content, manually annotated emotion labels, and token-level emotional embeddings. Combining affective information in different granularities and different aspects is challenging.

Inspired by personality studies and mood state analysis in human-computer interaction (HCI) [17, 21, 45], we propose to simulate the mood transition process affected by the given personality traits for emotion generation to solve the above challenges. In detail, we first project the mood states and emotions of the dialog system as regions and discrete points in the VAD space. Then, the transition among the mood states is modeled as the shifting among the different regions, where the emotion factors within the dialog context are the shifting variation, and the personality is modeled as the weight of shifting. In this way, the influence of personality is described in the VAD emotional space where the emotional factors (*e.g.* mood states and emotions) are also represented. The mood transition process is modeled based on the psychology and HCI domain knowledge,

<sup>1</sup>It is Pleasure-Arousal-Dominance (PAD) in the original paper, PAD and VAD share the same meaning in the context of text understanding, we will use VAD for consistency henceforth.

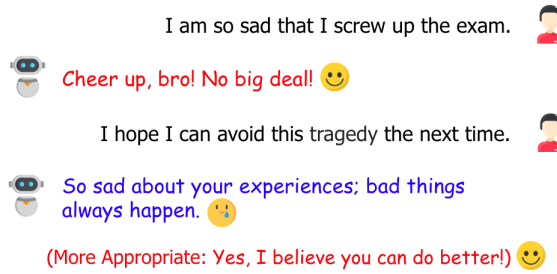


Fig. 1. A dialog snippet from a real conversation with a chatbot. The chatbot first optimistically encourages the user with a *Joy* emotion. But after the encouragement, the chatbot pessimistically shows empathy to the user in a *Sadness* emotion even the user expressed a positive attitude. A more appropriate emotional response here should still be in *Joy* so that consistent with the previous response.

while the parameters are learned from daily conversation data during training. To extract the multi-granularity affective information from the dialog context, we design an attention layer to align the semantic representations from BERT [12] as well as the emotion annotations at the utterance level, and the token-level VAD embeddings in different aspects. Finally, we observed that different people may express different emotions even in similar mood states or dialog contexts according to their personalities. So, we generate the response emotion (*i.e.*, a specific point corresponding to discrete emotion labels in the VAD space) within the predicted mood state region also conditioned on the given personality trait.

To facilitate related researches, we construct the **Personality EmotionLines Dataset (PELD)**, which includes 6,510 dialogue triples from 711 pieces of daily conversations with emotion labels and annotated personality traits. The emotion labels and personality annotations are collected and re-organized from other researchers [22, 51, 75] analyzing the script of a famous TV series *Friends*<sup>2</sup>. We further explore the mood transition patterns in different characters on PELD and observe that the personality traits influence more in the mood transitions related to negative emotions. Besides, we analyze the distributions of mood state transitions in PELD and the correlations between personality and mood transitions to facilitate future research.

We conduct extensive experiments on the PELD dataset for evaluation. The results verify that by integrating the personality and the mood transition regression modeling, our method achieves significantly better F-scores than the BERT-base model in Emotion Generation, especially for minority emotions. Besides, we also conduct an extensive ablation study to empirically validate the effectiveness of personality and mood transitions in the Emotion Generation task<sup>3</sup>.

Our key contributions are summarized as follows:

- To the best of our knowledge, we are the first to raise the effect of personality on generating appropriate emotion in dialog systems. We propose the task of personality-affected Emotion Generation, identify the challenging issues, and propose a solution by simulating the mood transition process in the dialog system inspired by psychological findings.
- We construct a dialog script dataset PELD with emotion and personality annotations from several existing corpora. Besides, we analyze the distributions of mood state transitions in

<sup>2</sup><https://en.wikipedia.org/wiki/Friends>

<sup>3</sup>Our code and dataset are released at: [github.com/preke/PELD](https://github.com/preke/PELD).

PELD and the correlations between personality and mood transitions to facilitate future research.

- We conduct extensive experiments on PELD to evaluate the effectiveness of our method. The results verify that integrating the personality and the mood transition regression significantly improves the F-scores than base models in Emotion Generation, especially in minority emotions.

In the following content, we first review the related studies in Section 2. Section 3 describes the background and preliminary knowledge of our work. Section 4 presents the problem definition of personality-affected emotion generation and the methodology we propose based on mood transition prediction. Section 6 introduces the experimental setup. Section 7 describes the evaluation results and analysis. Finally, we conclude and mention future work in Section 8.

## 2 RELATED WORKS

To enhance the understanding of the context behind our study and illustrate the distinctiveness of our work compared with prior research, we conducted a comprehensive review of the current literature on emotional dialog systems and personality influence on emotion expression.

### 2.1 Emotional Dialog Systems

The idea of an emotional dialog system was first proposed in Colby's 1975 publication [9], where a rule-based chatbot for simulating emotions was introduced. Present research on emotional dialog systems predominantly concentrates on three areas: (1) emotion recognition in conversations, (2) emotional response generation, and (3) empathetic response generation.

**2.1.1 Emotion Recognition in Conversations.** The ERC task aims to identify the emotion of each utterance from several pre-defined emotions given the transcript of a conversation along with speaker information on each constituent utterance [52]. Relevant research is becoming increasingly popular recently, largely because of its potential to extract opinions from the enormous amounts of conversational data accessible on diverse online platforms such as Facebook, YouTube, Reddit, Twitter, and other similar sites[52]. Conversational memory network (CMN), proposed by Hazarika et al. [19] for emotion recognition in conversation, utilizes distinct memories for each speaker for speaker-specific context modeling. Later, Hazarika et al. [18] improved upon this approach with an interactive conversational memory network (ICON), which interconnects these memories to model self and inter-speaker emotional influence. However, neither methods actually exploit the speaker's information of the target utterance for classification. This makes the model blind to speaker-specific nuances. DialogueRNN [40] aims to solve this issue by considering the speaker's information of the target utterance and, further, modeling self and inter-speaker emotional influence with a hierarchical multi-stage RNN with the attention mechanism. Dialogue Graph Convolutional Network (DialogueGCN) [15], a graph neural network-based approach is proposed by leveraging self and inter-speaker dependency of the interlocutors to model conversational context for emotion recognition. Through the graph network, DialogueGCN addresses context propagation issues present in the current RNN-based methods.

**2.1.2 Emotional Response Generation.** In dialog systems, emotional response generation is mainly concerns with incorporating emotional information into responses [37]. Research related to emotional response generation has gained popularity recently, following Microsoft's introduction of Xiaoice, a chit-chat robot capable of recognizing users' emotional needs, back in 2014 [81]. Emotional Chatting Machine [80] was proposed to exploit the deep learning approach in building a large-scale emotionally aware conversational bot. Emotional embeddings are also trained based

on context and then integrated into response generation [58]. Some researchers [10] control the emotional response generation with both categorical emotion representations and continuous word representations in VAD space [47]. Moreover, affectively diverse beam search [2] is proposed for decoding in response generation. Besides, reinforcement learning is also adopted to encourage response generation models to render specified emotions. Li et al. [29] combined reinforcement learning with emotional editing constraints to generate meaningful and customizable emotional replies. Sun et al. [62] also use an emotion tag to partially reward the model for expressing specified emotion.

**2.1.3 Empathetic Dialogue Generation.** Despite the existing advancement in emotion response generation, it is impractical always to specify response emotions for dialog systems in real application scenarios. Therefore, empathetic dialogue generation aims to understand the emotion of users and respond to them appropriately. Rashkin et al. [54] propose a benchmark for empathetic dialogue generation and a dataset (EMPATHETICDIALOGUES) of 25k conversations grounded in emotional situations. Followingly, the Mixture of Empathetic Listeners (MoEL) [35] was proposed as a novel end-to-end approach for modeling empathy in dialogue systems. Their model first captures the user's emotions and outputs an emotion distribution. Based on this, MoEL will softly combine the output states of the appropriate Listener(s), which are each optimized to react to certain emotions and generate an empathetic response. It is stated that empathetic responses often mimic the emotion of the user to a varying degree, depending on its positivity or negativity and content [39]. So, they use emotion stochastic sampling and emotion mimicry to generate responses that are appropriate and empathetic for positive or negative statements. EmpTransfo [76] utilizes OpenAI-GPT for language generation. The authors also show that the history of emotions and other metadata can improve the quality of generated conversations by the dialog system. To understand expressed empathy in text-based, asynchronous conversations, Sharma et al. [59] develop a multi-task RoBERTa-based [36] bi-encoder model for identifying empathy and extracting rationales underlying its predictions. Zheng et al. [78] state that empathy expression is multi-dimensional and is influenced by various factors. Therefore they propose CoMAE modeling communication mechanism, dialog act, and emotion together for empathetic response generation. To simulate the emotional interaction among humans, Wei et al. [69] design an emotion selector to learn the proper emotion for responses from massive dialogue pairs. But the emotional expression is subjective; for the same post, different users may have different emotions in their responses. So, the pattern learned only from online dialogues ignores the user information and turns out to be impractical.

## 2.2 Personality Effects on Emotions

Besides existing studies in emotional dialog systems, some researchers in related disciplines also consider projecting the effect of the personality of emotions in designing conversation agents. Research about personality effects on emotions is mainly distributed in cognitive science and Human-Computer Interaction (HCI). Due to the limited dialog corpus with personality annotations, related works in text-based conversation are few. However, some researchers also investigate how factual persona information influences emotion expressions in conversation.

**2.2.1 Personality Effects on Emotions in Cognitive Science.** Personality affects the way of communication in various manners including both linguistic style [38] and acoustic traits [50]. As it feels more natural to interact with a machine that has its own personality, implanting personality into dialogue agents would possibly increase the social attachment [26]. While emotion is a complex psychological experience of an individual's state of mind as interacting with people or environmental influences [17]. The **Pleasure-Arousal-Dominance (PAD)** [46] or **Valence-Arousal-Dominance**

(VAD) emotion temperament model shows three nearly orthogonal dimensions providing a comprehensive description of emotional states. Based on this, several psychologists studied the relationship between human emotional factors and personality factors. However, most of them are rule-based models [23] and probabilistic models [1]. Mehrabian et al. [45] utilized the five factors of personality [11] to represent the VAD temperament model through linear regression analysis.

**2.2.2 Personality Effects on Emotions in HCI.** To integrate the analysis above into Artificial Intelligence, some researchers in HCI borrow the ideas above and design facial emotional expressions for humanoid robots. An emotion generation model [21] was proposed that represents a robot's internal state. This model can assess the robot's individuality through mood transitions. The mood transition process of robots was also introduced in this work. Later on, a method of the smooth transition of emotional states for a robotic face was developed by Han et al. [16]. The proposed system can generate emotional behaviors of a robot in a more natural manner based on the 2-D emotional model. Subsequently, they continued to employ five factors of personality in a 2D (pleasure-arousal) scaling model to represent a robotic non-verbal emotional model [17]. Based on previous works, Masuyama et al. [42] studied the three stages (core affects, emotion, and mood state) robotic emotional model with the OCEAN model as the personality factors based on the 2D (Pleasant and Arousal) scaling model. Then, a personality-affected robotic emotional model and the emotion-affected associative memory model for the robot [43] are introduced for robots expressing emotions. The robots in this work provide non-verbal emotional interaction with users where the pre-defined personalities of robots affect their propensity for simulated mood transitions.

**2.2.3 Effects of Persona Information on Emotions.** While in the research of text-based dialog systems, early work in [3] utilizes models of emotions and personality encoded as Bayesian networks to generate empathetic behaviors or speech responses to users in conversation. Though the VAD space is adopted to model emotions in some research [2, 10, 71], integrating the psychological personality into the dialog system is rare due to the data shortage for personality analysis in conversation. Consequently, the psychological personality influence on emotion in dialogues is still an open problem. It is worth mentioning that, Zhong et al. propose a new task for persona-based empathetic conversations and present the first empirical study on the impact of persona on empathetic responding [79]. The authors use the description of persona as the additional input to generate empathetic responses based on the statement that persona has been shown to be highly correlated to personality, which in turn influences empathy. Although the impact of persona on generating appropriate emotions is validated in the results, as a more straightforward factor to affect emotion generation, the influence of personality in emotional dialog systems has not been studied before.

### 3 PRELIMINARIES

In this section, we introduce some concepts from psychology and affective computing for a better explanation of our problem and method.

#### 3.1 Personality, Mood-state, and Emotion

Emotion (short-term), mood-state (mid-term), and personality (long-term) thus represent three levels of the affective module that interact with each other [24]. Specifically, emotion is the term used in psychology to describe short-term variations in internal mental state [3]. Beyond discrete emotions, which are typically short-term, mood states are a powerful way to model emotional shifts and explain affective influences over longer periods of time [24]. Personality characterizes the long-term patterns of thought, emotion, and behavior associated with an individual [3, 73].

Table 1. Numeric vectors of Basic Emotions in the VAD Space [56].

| Basic Emotions | (Valence, Arousal, Dominance) |
|----------------|-------------------------------|
| Anger          | (-0.51, 0.59, 0.25)           |
| Disgust        | (-0.60, 0.35, 0.11)           |
| Fear           | (-0.62, 0.82, -0.43)          |
| Joy            | (0.81, 0.51, 0.46)            |
| Neutral        | (0.00, 0.00, 0.00)            |
| Sadness        | (-0.63, -0.27, -0.33)         |
| Surprise       | (0.40, 0.67, -0.13)           |

### 3.2 Basic Emotions in the VAD Space

There are two fundamental approaches to model emotions in psychological and affective computing literature: categorical and dimensional models. Categorical emotion models assume that all humans have a given discrete set of basic emotions (*i.e.*, *Anger*, *Disgust*, *Fear*, *Joy*, *Sadness*, and *Surprise* [13]). These models are easily adopted for classification tasks in emotion analysis problems [15, 67]. However, different researchers propose various standards of categorization [55], making it difficult to unify emotion analysis research. Dimensional approaches, on the other hand, often refer to Russell and Mehrabian's Valence-Arousal-Dominance (VAD) model [46]. According to this model, emotional states can be described relative to three fundamental emotional dimensions: Valence (the degree of pleasure or displeasure of an emotion), Arousal (level of mental activity, ranging from low engagement to ecstasy), and Dominance (extent of control felt in a given situation). Accordingly, emotions are characterized by three dimensions, each of which spans an interval of real-valued numbers indicating the strength and orientation of each dimension [5].

To better utilize the strength in different approaches, discrete emotions are projected into the dimensional space as shown in Table 1<sup>4</sup>. Consequently, a wide range of existing emotion analysis models can be embedded into the VAD space [4, 30, 31]. Besides, emotions in the VAD space can have multiplex comparisons. For example, all *Anger*, *Disgust* and *Fear* are negative emotions according to the valence dimension, but the intensity of *Anger* is higher than *Disgust* and *Fear* referring to the arousal dimension.

### 3.3 Mood States in the VAD Space

Mood states provide the historical context of recent emotions [41], and they are not usually characterized by their direction at a person or event. While someone might show an emotion *Anger* toward a specific object (*e.g.*, a colleague), as the specific emotion dissipates, they may feel a generally

<sup>4</sup>*Fear* and *Joy* correspond to *Terrified* and *Happy* in the original reference.

Table 2. The OCEAN personality traits and description [11].

| Factor            | Description  |
|-------------------|--|
| Openness          | Openminded, imaginative, and sensitive.            |
| Conscientiousness | Scrupulous, well-organized.                        |
| Extraversion      | The tendency to experience positive emotions.      |
| Agreeableness     | Trusting, sympathetic, and cooperative.            |
| Neuroticism       | The tendency to experience psychological distress. |

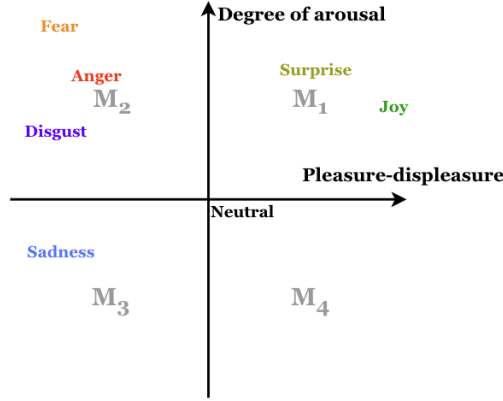


Fig. 2. Mood domains and emotions in the VAD space. We only illustrate the plane of Valence (Pleasure-displeasure) and the Arousal (Degree of arousal) axis here because the mood domains are split only by these two dimensions [21].  $M_1$ ,  $M_2$ ,  $M_3$ , and  $M_4$  are different quadrants representing different mood states.

bad mood [65]. Mehrabian et al. [46] give specific names to the resulting different octants in VAD space and describes the diagonally opposite octants as Exuberant/Bored, Dependent/Disdainful, Relaxed/Anxious, Docile/Hostile. Thus mood-states are not points but octants of the VAD space. More practically, moods can also be represented in four categories by dividing the VAD spaces into four domains through the axis of pleasure-displeasure and degree of arousal, as shown in Figure 2. Emotions from  $M_1$  (e.g., *Joy*) are with positive polarity and high intensity. While *Sadness* from  $M_3$  is with negative polarity and low intensity, for instance.

### 3.4 Personalities in the VAD Space

The big five personality traits (OCEAN, shown in Table 2) are widely used for psychological analysis. Higher scores for openness are correlated with philosophical and free thought, as well as an interest in the arts, music, and cinema [14, 57]. Those who score low here may be more practical, realistic, or close-minded [11]. Individuals with high conscientiousness tend to be well-organized, which may be expressed through discussions of work or school-related responsibilities [14, 74]. Those who score lower on this dimension may appear impulsive or unreliable. Those with high extraversion are likely to talk about friends and are willing to have interpersonal interaction. However, low extraversion people may be more independent and may focus more on solo activities [11, 48]. Agreeableness is associated with being friendly and good-natured, while those who score low may be selfish or rude. Swearing is highly correlated with low agreeableness [57, 74]. High neuroticism is strongly linked to anxiety and depression, while low neuroticism is linked to emotional stability. This dimension may be expressed through feelings such as fear, sadness, or frustration [11, 14].

Positioning a personality within a VAD space could have been a rather difficult task since there is no mathematically-correct way to make the conversion. Luckily, this transformation can be based on empirical data. [45] proposed a temperament model shown in Equation 1 derived through linear regression to show the VAD scales of personality traits, where  $O, C, E, A, N$  are the strength of the big-five personality traits.



$$\begin{aligned}
 P_V &= 0.21E + 0.59A + 0.19N \\
 P_A &= 0.15O + 0.30A - 0.57N \\
 P_D &= 0.25O + 0.17C + 0.60E - 0.32A
 \end{aligned} \tag{1}$$

It is seen that VAD scores can be estimated with moderate accuracy when only big-five scores are available, and investigators also wish to analyze their data in relation to the VAD temperament model.

## 4 PERSONALITY-AFFECTED EMOTION GENERATION

### 4.1 Problem Statement

We define Personality-affected Emotion Generation as enabling the dialog system to generate appropriate emotions for response facilitated by the given personality trait.

Formally, we study a dyadic emotional conversation between the user and the dialog system, which contains the dialog context  $C = \{(U_1, E_1), (U_2, E_2), \dots, (U_{n-1}, E_{n-1})\}$  including all the preceding  $n - 1$  utterances, where  $E_i$  is the emotion label for each utterance  $U_i$ . The personality trait  $P$  of the dialog system is also given. Our objective is to let the dialog system generate an appropriate emotion  $E_n$  for the upcoming response  $U_n$  to the user. We evaluate the appropriateness of the generated  $E_n$  by comparing it with the ground truth emotion  $E_n'$  in the dialog corpus.  $E_n'$  is generated by the speaker annotated with the same personality  $P$  in the given context.

Based on the problem definition and the preliminaries above, we design the Personality-affected Mood Transition model as illustrated in Figure 3. Our model includes two modules: Mood Transition Regression and Emotion Generation. We will introduce both modules in the following subsections.

### 4.2 Mood Transition Regression

**4.2.1 Personality Effect on Mood Transition.** Mood dynamically changes with the accumulation of past emotions. Changing the mood transition of robots can produce various characterizations in them as well as individuality that enables the robots to demonstrate a wide range of behaviors [21]. Moreover, the transition of mood depends not only on the user's emotions but also on the mood and personality of the dialog system[17]. Therefore, we first model the personality of the dialog system as the weighting parameters of the mood state transition.

In our model, the mood states are represented by  $M_1, M_2, M_3, M_4$ , and *Neutral* referring to Figure 2, where the *Neutral* especially represent the coordinate origin of the VAD space<sup>5</sup>. We use mood vectors in Table 3 to represent the mood states in our model. Therefore, the mood transition is the shifting among different octants in the VAD space [66].

Besides, the personality of the dialog system is specified as a 5-dimensional vector  $P = [O, C, E, A, N]$  in our method representing the strength in Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism, respectively. The temperament of personality in the VAD space (shown in Equation 1) is widely used as the weighting parameter for mood transition of robots [4, 17, 43], which is also adopted by us as the preliminary personality traits of the dialog system.

In detail, we first obtain  $P_V, P_A, P_D$  referring to Equation 1 as the preliminary affecting parameters. However, these numeric coefficients are summarized from the analysis of questionnaire results from 72 participants [45], which are not suitable for directly adopted in the model design. Our intent is to learn how personality affects the mood transition process in conversation from the dialog corpus. Therefore, we choose  $P_V, P_A, P_D$  as prior parameters and learn suitable coefficients  $\mathcal{P}_V, \mathcal{P}_A, \mathcal{P}_D$  from conversational data through an adaptation dense layer  $D_A$ .

<sup>5</sup>We generally regards  $M_1, M_2, M_3, M_4$ , and *Neutral* are different regions (octants) in the VAD space

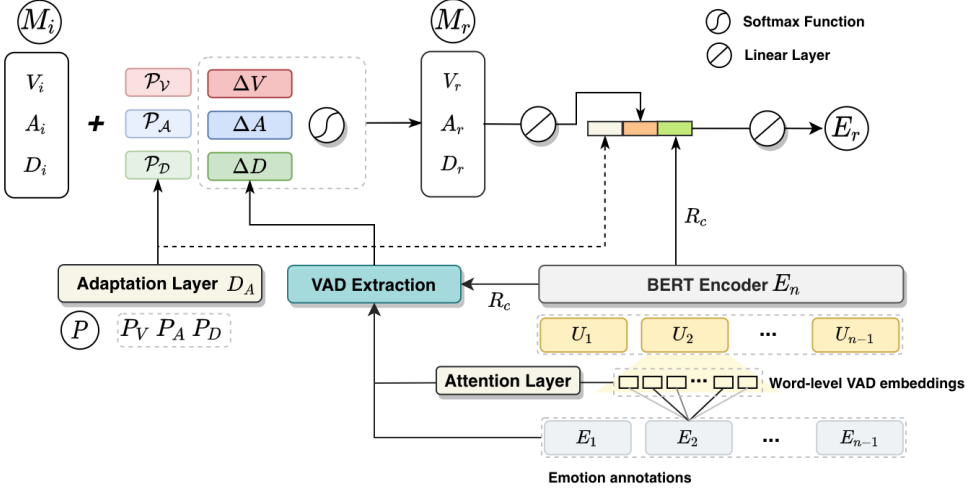


Fig. 3. The Model Illustration. The upper-left part is the mood transition regression, where the initial mood state  $M_i$  changes to the  $M_r$  in the upcoming response. The personality trait  $P$  is transformed to the transition weights  $\mathcal{P}_V, \mathcal{P}_A, \mathcal{P}_D$ , and the extracted affective information from the dialog context is changed to the transition variables  $\Delta V, \Delta A, \Delta D$ . Then, the upper-right part is the emotion generation module, where the response  $E_r$  is generated through integrating  $M_r, P$ , and the dialog context  $R_c$ . The lower part illustrates how we extract affective information by the Attention Layer aligning the emotion annotations, word-level VAD embeddings, and the semantic representations from the input dialog context.

Table 3. Mood VAD vectors representing different mood states

| Mood States | (Valence, Arousal, Dominance) |
|-------------|-------------------------------|
| $M_1$       | (1.0, 1.0, 0.0)               |
| $M_2$       | (-1.0, 1.0, 0.0)              |
| $M_3$       | (-1.0, -1.0, 0.0)             |
| $M_4$       | (1.0, -1.0, 0.0)              |
| Neutral     | (0.00, 0.00, 0.00)            |

**4.2.2 Affective Information Extraction.** In human-robot interaction, users' emotional expressions can be treated as trigger inputs to drive the robotic mood transition [17]. Inspired by this idea, we extract the affective information from the utterances in the dialog context to facilitate the mood transition process in our model.

Extracting the affective information from utterances has been studied by previous researchers. Buechel et al.[6] annotate a corpus of over 10,000 English sentences employing the dimensional VAD model of emotion. Park et al. [49] predict VAD vectors with the supervision of categorical emotion annotations by minimizing the EMD (Earth Mover's Distance) loss between the predicted VAD score distribution and the categorical emotion distributions. However, both methods overlook the affective information in the semantic content. So, we design an affective extraction module to integrate the semantic content, token-level VAD embeddings, and emotion annotations for each utterance together.

Specifically, at the utterance level, we first fine-tune the pre-trained BERT<sup>6</sup> [12] encoder  $E_n$ , a famous pre-trained language model whose performance is widely validated in many natural language understanding tasks, on daily conversation corpus by concatenating all the utterances in  $C$  with the [SEP] to form the input and obtain the semantic representation  $R_c$  from the dialog content.

$$R_c = E_n(U_1[SEP]U_2, \dots, U_{n-1}[SEP]) \quad (2)$$

Then, at the token-level, for each utterance  $U_i$  in  $C$ , we design an attention layer using the VAD embeddings  $[V_i, A_i, D_i]$  of the emotion annotation  $E_i$  to attend the VAD embeddings of all the tokens  $[t_1, t_2, \dots, t_m]$  in  $U_i$ . The VAD embeddings for the emotion annotations and each word in the utterance are obtained by the manual VAD annotations for English words [47]. The attention weights  $W_j$  are calculated by the softmax function over the inner product between the VAD embeddings of  $E_i$  and  $t_j$ :

$$\begin{aligned} W_j &= \frac{e^{E_i^T t_j}}{\sum_{j=1}^m e^{E_i^T t_j}} \\ A_i &= \sum_{j=1}^m W_j t_j \end{aligned} \quad (3)$$

Our principle here is that not all words are affected by emotion. For example, function words are mostly only affected by syntax while topical/factual words are dominated by topics or facts [37, 72]. Therefore, a higher attention weight means the current word is more important in expressing emotion. The attention output  $A_i$  combines the token-level (word VAD embeddings) and the utterance-level (utterance emotion annotations) affective information for each sentence.

Finally, we sum the average among all the  $E_i$  for each utterance, and the average among all the output  $A_i$  from the attention layer together as the affective information extracted from the dialog context. Besides, the semantic representation  $R_c$  is also concatenated with the affective information and is fed into a linear layer to obtain the mood variation  $\Delta V, \Delta A, \Delta D$ .

$$\begin{aligned} E &= \sum_{i=1}^{n-1} \frac{E_i}{n-1} \\ A &= \sum_{i=1}^{n-1} \frac{A_i}{n-1} \\ \Delta V, \Delta A, \Delta D &= \text{Linear}((w_E E + w_A A) \oplus R_c) \end{aligned} \quad (4)$$

where  $w_E$  and  $w_A$  are hyperparameter weights to balance the two aspects.

**4.2.3 Mood State Regression.** We utilize the regression methods to predict the response mood  $M_r$ , which aims to describe the distance among different mood states in the VAD space. Firstly, we obtain the numeric values of the  $M_r$  by Table 3. The predicted mood state  $M_{r'} : V_r', A_r', D_r'$  is calculated by adding the preceding mood state  $M_i : V_i, A_i, D_i$  and the variation of the mood weighted by the personality coefficients  $\mathcal{P}_V, \mathcal{P}_A, \mathcal{P}_D$ . What is more, we add a softmax function to the mood variation  $\Delta V, \Delta A, \Delta D$  to constrain their numeric scale into  $[0 - 1]$  for the transition regression.

<sup>6</sup>Here we adopt the pre-trained BERT-base model from Huggingface (<https://huggingface.co/>).

$$\begin{aligned}
V'_r &= V_i + \mathcal{P}_V * \text{Softmax}(\Delta V) \\
A'_r &= A_i + \mathcal{P}_A * \text{Softmax}(\Delta A) \\
D'_r &= D_i + \mathcal{P}_D * \text{Softmax}(\Delta D)
\end{aligned} \tag{5}$$

### 4.3 Emotion Generation

After predicting the mood state of the upcoming response, we generate the appropriate emotion that the dialog system responds to the user. The dialog context acts as a set of parameters that influence a person to speak an utterance while expressing a particular emotion [51]. Besides, emotion expressions are also affected by the personality trait.

According to our observation, different people may express different emotions in similar mood states or dialog contexts according to their personalities. For example, under a stressful context, individuals who exhibit high levels of Neuroticism will tend to be irritable and express anxiety or anger, while a person who ranks lower on the neurotic level will peacefully exhibit emotions with lower arousability.

Therefore, we concatenate the predicted mood state  $M'_r$ , the personality trait  $P$ , and the semantic representation  $R_c$  of the dialog context as the input. Here, to project the mood state and the personality trait into the same feature space as  $R_c$ , we adopt two linear layers  $Linear_m$  and  $Linear_p$  respectively on  $M'_r$  and  $P$ . Then, the  $E_r$  is generated from  $M'_r$  through a linear classification layer  $Linear_{cls}$  conditioned on both  $R_c$  and  $P$ , as described in Formula 6.

$$E_r = Linear_{cls}(Linear_m(M'_r) \oplus Linear_p(P_n) \oplus R_c) \tag{6}$$

### 4.4 Joint Training

As the mood transition prediction part is the preceding module of the emotion generation part, we jointly optimize the mood transition prediction part and the emotion generation part during training so that the performance of the latter part could be enhanced.

First, we calculate the mean square error (MSE) as the mood loss  $\mathcal{L}_{mood}$  between the regression results in Formula 5 and the mood state vectors  $[V_r, A_r, D_r]$  of actual  $M_r$  in Table 3.

$$\mathcal{L}_{mood} = \sqrt{(V_r - V'_r)^2} + \sqrt{(A_r - A'_r)^2} + \sqrt{(D_r - D'_r)^2} \tag{7}$$

For the emotion generation, based on our observation of dialog corpus that discrete emotions are unequally distributed in daily conversations (e.g., *Neutral* is the majority among all the emotions, while *Fear*, *Disgust* are the minority), we use the Focal Loss[33] between the generated emotion and the ground truth emotion to address class imbalance during training, as shown in 8.

Focal Loss applies a modulating term to the cross-entropy loss in order to focus learning on hard misclassified examples. It is a dynamically scaled cross-entropy loss, where the scaling factor decays to zero as confidence in the correct class increases. Intuitively, this scaling factor can automatically down-weight the contribution of easy examples during training and rapidly focus the model on hard examples. Formally, the Focal Loss adds a factor  $(1 - p_t)^\gamma$  to the standard cross-entropy criterion, where the  $p_t$  is the probability of generating the correct emotion. Setting  $\gamma > 0$  reduces the relative loss for well-classified examples ( $p_t > 0.5$ ), putting more focus on hard, misclassified examples.

$$\mathcal{L}_{emo} = FL(p_t) = -\alpha(1 - p_t)^\gamma \log(p_t) \tag{8}$$

The  $\alpha$  is set to be 0.5 and  $\gamma$  to be 2 according to the empirical results in the original paper [33].

Finally, we minimize the mood loss and emotion loss simultaneously with adaptive weights  $w_1, w_2$  learned during training.

$$\mathcal{L} = w_1 \mathcal{L}_{mood} + w_2 \mathcal{L}_{emo} \tag{9}$$

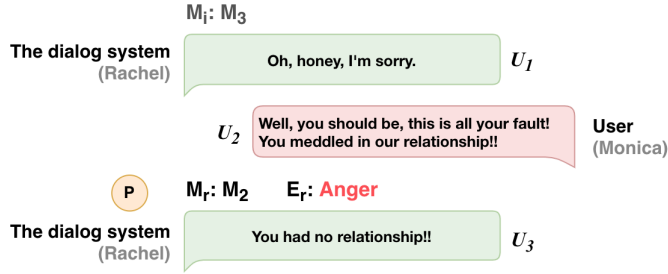


Fig. 4. A triple example in PELD. The dyadic conversation between Rachel and Monica (two main roles in *Friends*,  $P$  is the personality of Rachel. In this example, the dialog system is set as Rachel and generate the emotion *Anger* in the response  $U_3$  to the user set as Monica.

## 5 THE PELD DATASET

### 5.1 Dataset Construction & Statistics

To facilitate related research, we construct the **Personality EmotionLines Dataset (PELD)**, an affective dialog dataset with personality traits for speakers and emotion annotations for utterances. In PELD, each sample is represented as a dyadic dialog triple  $\{(U_1, U_2, U_3), M_i, M_r, E_r, P\}$ , as shown in Figure 4.  $M_i$  and  $M_r$  are mood states expressed in  $U_1$  and  $U_3$ ,  $P$  is the personality trait, and  $E_r$  is the emotion label to be generated.

According to the book *Television Dialogue: The sitcom Friends vs natural conversation* [53], television conversations and natural conversations are basically the same in terms of linguistic features. This classic script is widely analyzed in many dialog researches [22, 27, 28]. Therefore, we construct the triples in PELD from the dialog scripts in *Friends* based on existing studies.

Specifically, the utterances and their emotion labels are mainly adopted from the dialogues in the MELD [51] and the EmoryNLP dataset [75], two famous datasets analyzing emotional expressions in *Friends*. The VAD vector of each mood state refers to Table 3. To keep consistency, each dialog triple in PELD is constructed within the same dialogue in the original datasets. The personality traits in our dataset are adopted from the personality annotations in 711 different dialogues [22]. We only keep the personality traits of the six main roles in *Friends* for confidence, as these annotations are most frequent. Referring to the annotations, a role may exhibit different aspects of its personality in different conversation scenarios. Thus, for each of the main roles, we average their annotated personality traits in all the dialogues by  $P = \frac{1}{K} \sum_{i=1}^K P_i$  for simplification, where  $K$  is the number of annotations. The averaged results are shown in Table 4.

We also calculate the standard deviation of the personality traits on each dimension. It suggests that Extroversion and Conscientiousness share the largest distinction among the five traits in the six main roles. These two traits, especially Extroversion, play important roles in mood state variation and emotion generation according to the previous discussion. Therefore, it also reflects the availability of PELD in the aspect of personality.

We split the PELD into **Train**, **Valid**, and **Test** set with portion around 8:1:1. There are 6,527 triples in PELD. The total number of unique utterances in PELD (10,468) is less than the sum of the original MELD (13,708) and the EmoryNLP (9,489). The reason is there are lots of overlap between these two datasets. Besides, not all dialogues include the six main roles and are suitable for constructing triples. The overall statistics of the dataset are shown in Table 5<sup>7</sup>. The average length

<sup>7</sup>According to the annotations in PELD, there is no utterance in  $M_4$ .

Table 4. Personality vectors of *Friends* main roles in PELD.

| Roles    | Personality Traits (O,C,E,A,N)      |
|----------|-------------------------------------|
| Chandler | [0.648, 0.375, 0.386, 0.58, 0.477]  |
| Joey     | [0.574, 0.614, 0.297, 0.545, 0.455] |
| Monica   | [0.713, 0.457, 0.457, 0.66, 0.511]  |
| Phoebe   | [0.6, 0.48, 0.31, 0.46, 0.56]       |
| Rachel   | [0.635, 0.354, 0.521, 0.552, 0.469] |
| Ross     | [0.722, 0.489, 0.6, 0.533, 0.356]   |
| Std      | [0.059, 0.093, 0.120, 0.065, 0.068] |

of utterances is 9.32 in the whole PELD, which also conforms to the length of short sentences in daily conversation.

Similar to existing emotional conversation datasets [7, 32], PELD also suffers the emotion imbalance issue. Utterances labeled as *Neutral* are the majority (44.6%), while *Fear* and *Disgust* only take a small portion (6.9% and 1.9%). Though it reflects the actual emotion distribution in daily conversation, it also brings challenges to machine learning models to identify and generate emotions. We tried several automatic methods for data augmentation like synonym substitution, back-translation, or the Easy Data Augmentation (EDA) proposed in [68]. But most of the synthetic samples are either odd or the same as the original samples. The reason might be there are limited options for short sentences as utterances in conversation to replace synonyms, add or delete words.

The mood state distribution in PELD is the sum of emotions within mood states. All the triples are equally distributed among the six main roles, ranging from 977 (Phoebe) to 1159 (Rachel). So, there is no explicit dominant personality in our dataset, making the model more robust when trained on it.

## 5.2 Mood Transitions in PELD

After constructing PELD, we further explore the dataset in the aspect of mood transitions. As the triples in PELD are constructed for analyzing the transitions between  $M_i$  in  $U_1$  and the  $M_r$  in  $U_3$ , we show both the mood states and emotion distributions in the  $U_1$  and  $U_3$  in Table 6, respectively.

We can see that for both mood states and emotions, the distributions in  $U_1$  and  $U_3$  are similar, which means the transition of mood states and emotions are equitable in PELD triples, *i.e.*, no explicit mood state or emotion transition dominates during the dialogues. It also conforms to the nature of daily conversations. Moreover, the proportions of all mood states and emotions are also similar to the overall statistics of PELD, which suggests that the mood states and emotions in PELD are also average distributed in the triples.

Since mood transitions are affected by the personality traits as discussed above, we exhibit the mood transition patterns for different roles with different personality traits in Figure 5. In general, among the six transition matrixes, most of the first columns are in deeper colors, which indicates most transitions occur from other mood states to *Neutral* as it is the majority in PELD. Besides, blocks with deeper color are also more likely to occur around or in the diagonals of the transition matrixes; it suggests the preceding mood states tend to transition to the same or similar ones.

As for individual differences, nearly 60% of  $M_3$  in Joey and Phoebe transmit to *Neutral* and  $M_1$ , which shows they can better handle the negative mood states. This also correlates to the lower Extraversion compared to others in Table 4. Specifically, over 80% of  $M_1$  in Joey remains to be

Table 5. Basic Statistics in PELD.

| Basic Statistics       | Train | Valid | Test | Total |
|------------------------|-------|-------|------|-------|
| #Triple                | 5286  | 588   | 653  | 6527  |
| #Unique Uttr.          | 9273  | 1529  | 1679 | 10468 |
| Avg. Uttr. Length      | 9.26  | 9.33  | 8.95 | 9.32  |
| #Emotions              | Train | Valid | Test | Total |
| Anger                  | 1857  | 238   | 247  | 2342  |
| Disgust                | 316   | 30    | 30   | 376   |
| Fear                   | 1100  | 118   | 132  | 1350  |
| Joy                    | 2883  | 321   | 345  | 3549  |
| Neutral                | 7066  | 782   | 880  | 8728  |
| Sadness                | 1086  | 120   | 141  | 1347  |
| Surprise               | 1550  | 155   | 184  | 1889  |
| #Mood States           | Train | Valid | Test | Total |
| Neutral                | 7066  | 782   | 880  | 8728  |
| $M_1$                  | 4433  | 476   | 529  | 5438  |
| $M_2$                  | 3273  | 386   | 409  | 4068  |
| $M_3$                  | 1086  | 120   | 141  | 1347  |
| $M_4$                  | -     | -     | -    | -     |
| #Triples of Main Roles | Train | Valid | Test | Total |
| Chandler               | 864   | 107   | 117  | 1088  |
| Joey                   | 929   | 96    | 100  | 1125  |
| Monica                 | 847   | 95    | 111  | 1053  |
| Phoebe                 | 789   | 90    | 98   | 977   |
| Rachel                 | 934   | 97    | 128  | 1159  |
| Ross                   | 923   | 103   | 99   | 1123  |

*Neutral* and  $M_1$ , which especially shows Joey is an optimistic person. As for the ratio of unpleasure mood states  $M_2$  and  $M_3$  change to pleasure mood state  $M_1$ , Phoebe achieves the highest.

Moreover, to highlight the individual differences in mood transitions among the six main roles in detail, we also show the standard deviations (Std) of each row in the transition matrixes of the six main roles, as shown in Figure 6. The red bar chart shows the Std of the infinite norms of rows in the mood transition matrix, which indicates the diversity of the most probable mood states from

Table 6. Mood States and Emotions distributions in PELD Triples

| Tri.Moods | Neutral | $M_1$ |          | $M_2$ |      |         | $M_3$   |
|-----------|---------|-------|----------|-------|------|---------|---------|
| $M_i$     | 2922    | 1842  |          | 1325  |      |         | 438     |
| $M_r$     | 2778    | 1763  |          | 1492  |      |         | 494     |
| Tri.Emos  | Neutral | Joy   | Surprise | Anger | Fear | Disgust | Sadness |
| $E_i$     | 2922    | 1245  | 597      | 752   | 458  | 115     | 438     |
| $E_r$     | 2778    | 1129  | 634      | 859   | 489  | 144     | 494     |

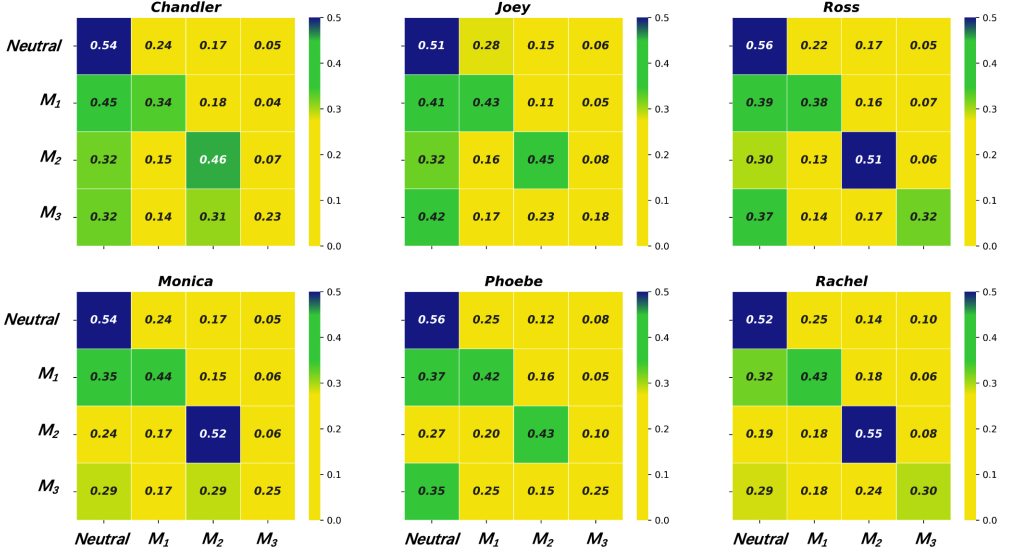


Fig. 5. Mood transition matrixes of the six main roles in PELD. Each row in a matrix shows the ratios of the current mood state  $M_i$  is transferred to the next mood state  $M_r$ .

the same mood state in transitions of different roles. While the blue bar chart shows the Std of the L2-norms, which generally describes the difference in how different roles transfer from one mood state to others. Detailed calculations of the standard deviations are shown below:

$$\begin{aligned}
 Std(||\mathcal{M}^i||_2) &= \sqrt{\frac{1}{m} \sum_{j=1}^m \left( \sum_{k=1}^n (|\mathcal{R}_{jk}|^2) - \mu_2^i \right)^2} \\
 Std(||\mathcal{M}^i||_{inf}) &= \sqrt{\frac{1}{m} \sum_{j=1}^m (\max ||\mathcal{R}_j|| - \mu_{inf}^i)^2}
 \end{aligned} \tag{10}$$

where  $||\mathcal{M}^i||_2$  is the L2-norm of the  $i$ -th mood state among the six roles, while  $||\mathcal{M}^i||_{inf}$  is the corresponding infinit-norm.  $m$  is the number of the roles,  $n$  is the number of columns in each mood transition matrix.  $\mathcal{R}$  is the  $i$ -th row in the mood transition matrix, indicating the result mood states of the  $i$ -th mood state's transition.  $\mu_2^i$  and  $\mu_{inf}^i$  are the mean of  $||\mathcal{M}^i||_2$  and  $||\mathcal{M}^i||_{inf}$ , respectively.

Both charts show similar patterns of mood transitions. Unpleasure moods ( $M_2$  and  $M_3$ ) vary the most in different roles, while people are more common when process *Neutral* and pleasure mood  $M_1$  in conversation. Besides, Unpleasure mood states ( $M_2$  and  $M_3$ ) are relatively higher than pleasure mood states and *Neutral* on average, which means the individual difference is larger. So, we can infer that personality traits influence more in mood transitions from negative emotions.

### 5.3 Personality-aware Mood State Transitions

Although we have done an extensive analysis of the mood transitions of different characters, these conclusions may have certain limitations because their personality traits are specially designed by the scriptwriters for the TV series.



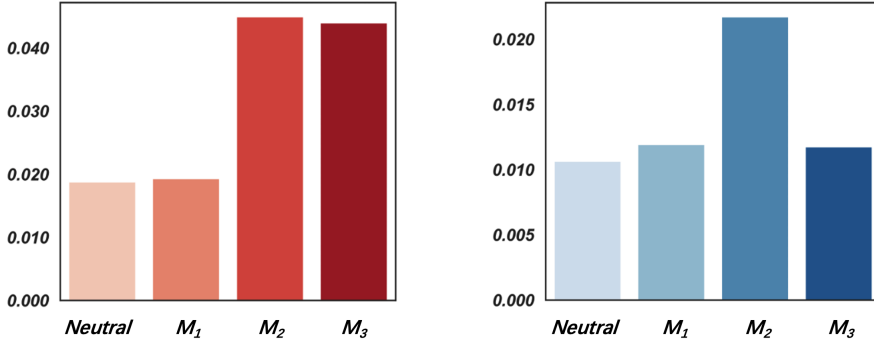


Fig. 6. The standard deviations of the infinity norm (red) and the L2-norm (blue) of each row in mood transition matrixes of the six main roles in PELD. In both figures, a higher bar indicates a greater disparity among the roles on the mood transition pattern in the corresponding row.

In order to explore more of the influence of personality on mood transitions, we investigate the personality-aware mood state transitions. Specifically, we calculate the Spearman correlations between mood state transitions and different personality traits in PELD, as shown in Figure 7.

Figure 7 shows the mood state variation from  $U_1$  to  $U_3$  in the triples in PELD, where we illustrate the variations in separate V-A dimensions<sup>8</sup>. We can see that Conscientiousness (C) has a higher positive correlation with mood state variation in the V-dimension. It suggests that the mood states of speakers with higher Conscientiousness tend to be more positive during conversations. In other words, if these speakers initially experience *Fear*, *Anger*, *Sadness*, or *Disgust*, they are more likely to transition toward *Surprise* or *Joy* as the conversation progresses. On the contrary, speakers with higher Openness (O) have a higher negative mood state variation in the V-dimension, which means these speakers are more likely to be first positive and then relatively negative in the conversations. In terms of emotion, these speakers are more likely to be first in *Surprise* or *Joy*, then become *Fear*, *Anger*, *Sadness*, or *Disgust*.

It is worth noting that the current conclusions we draw from analyzing mood state transition patterns based on the Big Five in PELD still have certain limitations. Personality expression, especially in conversational contexts, is affected by various cultural backgrounds, social factors, and even different conversation situations. Therefore, any analysis results from a single conversation dataset will have specific limitations.

However, our intent is not to derive a universal conclusion about the influence of personality on mood transitions in conversations from a social science perspective. The methods we present, such as calculating mood transition matrices, comparing the norms of these matrices, and examining the correlation between changes in different affective dimensions (V, A) and the intensity of different personality traits, can be used to analyze various conversation datasets with Big Five personality and affective annotations.

## 6 EXPERIMENT

### 6.1 Evaluation Task

To validate the effectiveness of our proposed method, we conducted the Emotion Generation task on PELD. Emotion Generation requires the model to generate the appropriate response emotion in

<sup>8</sup>the mood states are only in V-A spaces as explained in Section 3

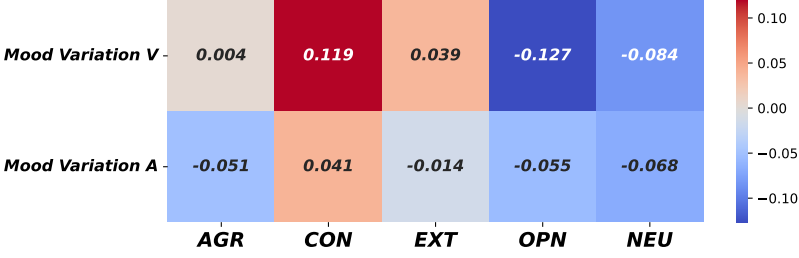


Fig. 7. Spearman correlations among mood state variations (from  $U_1$  to  $U_3$  in PELD's triples) and personality traits. The mood variations are in Valence and Arousal dimensions.

the upcoming utterance based on the preceding dialog context in a dyadic conversation scenario. The emotions here are discrete labels of basic emotions (*i.e.*, *Anger*, *Disgust*, *Fear*, *Joy*, *Neutral*, *Sadness*, and *Surprise*).

We evaluate the performance by F-scores of single emotions. Besides, the overall performance is also measured from two aspects: the macro averaged F-score (**m-avg**), and the weighted averaged F-score (**w-avg**) of all seven emotions. A higher m-avg indicates the model performs relatively better in predicting each category, while a higher w-avg means the model predicts mood states or emotions with larger proportions in the dataset better. Our evaluation is under the assumption that emotions expressed in the upcoming utterance are the emotions generated by the speakers.

## 6.2 Ablation Study Setting

Although there are existing studies working on emotion recognition in conversation [34, 61], there is no existing model to solve the Emotion Generation task issued in our work. Therefore, we conduct an ablation study to evaluate the effectiveness of different modules of our model design and the ways we utilize the personality and model the mood transition. The ablation study compares the performances of the following baseline models:

**BERT-base:** BERT-base [12] is a famous pre-trained language model with 12 transformer layers, over 110 million parameters, and 12 attention heads. It is capable of capturing bidirectional context for performing a wide range of downstream NLU tasks. Bert-base is pre-trained with a large corpus of text consisting of around 3.3 billion words and 2,500 million sentences. This vast corpus allowed BERT to learn a wide range of linguistic patterns and structures, making it a highly effective pre-trained model for natural language processing tasks. We use the pre-trained BERT-base model, corresponding to the  $E_n$  in our model, to encode the preceding dialog context to obtain the semantic representation as input, then directly predict the emotions for response through a classification head.

**BERT-Mood:** based on the vanilla BERT-base model, we add a mood state classification as the auxiliary task to enhance Emotion Generation. In BERT-Mood, we first use the BERT-base model to encode the preceding dialog context to obtain the semantic representation. Then, the semantic representation is fed into two classification heads: Emotion Generation and mood state classification. The sum of the losses from these two classification tasks will be back-propagated to the BERT-base model and update its parameters. This baseline model is to verify if the mood state information helps the Emotion Generation task.

**BERT-MT:** BERT-MT (Mood Transition) integrates the mood transition module (as described in Section 5.2) into BERT-base. In BERT-MT, the transitioned mood state is concatenated with the dialog context to help generate  $E_r$ , where we use both the regression loss of the mood state generation and the classification loss of Emotion Generation for supervision. In this baseline model, the personality is ablated to its effectiveness. To eliminate the influence of the model scale, we keep the number of parameters the same as our model and randomly initialize the personality-related parameters.

**BERT-P:** In BERT-P, we concatenate the preceding context representation from BERT-base and the specified personality trait vector together. Then, the concatenated representation is fed into the classification head for Emotion Generation. This baseline model is to evaluate whether merely using the personality trait vector can enhance Emotion Generation.

### 6.3 Implementation Details

To facilitate the reproduction of our model, we add the implementation details as follows. The dialog flows in PELD are fed into the models in batches with a size of 16. In encoding the preceding dialog context, we pad all the utterances with [PAD] to the MAX\_LEN of 128. We adopt the pre-trained BERT-base model from huggingface<sup>9</sup>.

We set the warm-up as 0.05 of the total training times. Besides, we use the Adam [25] as the optimization algorithm in training. The learning rate for all the models is set as  $1e-5$ . All the models for testing are selected referring to the best performance on the Validation sets in 50 epochs of training.

To ensure the reliability of results, we run each experiment 10 times with different random seeds and record the average performances and the standard deviations, where the random seeds are used to split the datasets and initialize the model parameters. Lastly, we released our code and data at github<sup>10</sup>.

## 7 RESULTS AND ANALYSIS

In this section, we report and analyze the experimental results of Emotion Generation in our ablation study by answering several research questions:

**RQ1:** Does our model outperform baseline methods in Emotion Generation?

**RQ2:** What is the correlation between the mood transition process and Emotion Generation?

**RQ3:** Are certain personalities easier to predict mood transition and Emotion Generation?

**RQ4:** Are there any cases to show how our model generates the emotions for the response?

### 7.1 Does our model improve baseline methods with the mood transition and personality in Emotion Generation?

We report the performance of Emotion Generation on the test set of the PELD in Table 7. To show the significance of the outperformance of our method, we conduct the Welch's t-test [70] between our model and all the baseline models. The Welch's t-test, also known as Welch's unequal variances t-test, is a statistical test used to compare the means of two independent groups when the assumption of equal variances is violated. It is a modification of the traditional Student's t-test that allows for unequal variances between the groups being compared. We employed Welch's t-test to assess the statistical significance of our model's superior performance over other baseline models.

<sup>9</sup>huggingface.com

<sup>10</sup><https://github.com/preke/PELD>

Table 7. Result comparison among baseline methods and our model in Emotion Generation. The first lines in baseline results are the F-scores of Emotion Generation (averaged over ten runs), while the second lines are the P-values  $p$  by calculating Welch's t-test between each baseline and our method. The green results indicate that our model outperforms the corresponding baselines with  $p \leq 0.05$ , where we consider the gain in performance to be statistically significant.

| Methods   | Anger        | Disgust      | Fear         | Joy          | Neutral      | Sadness      | Surprise     | m-avg        | w-avg        |
|-----------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
| BERT      | 0.318        | 0.012        | 0.226        | 0.278        | 0.513        | 0.212        | 0.109        | 0.242        | 0.375        |
|           | 0.05         | 0.02         | 0.29         | 0.05         | 0.03         | 0.03         | 0.03         | 0.03         | 0.03         |
| BERT-Mood | 0.252        | 0.113        | 0.227        | 0.248        | 0.468        | 0.288        | 0.107        | 0.242        | 0.344        |
|           | 0.01         | 0.00         | 0.36         | 0.03         | 0.00         | 0.05         | 0.00         | 0.01         | 0.00         |
| BERT-P    | 0.267        | 0.096        | 0.159        | 0.320        | 0.494        | <b>0.299</b> | 0.119        | 0.254        | 0.349        |
|           | 0.05         | 0.04         | 0.05         | 0.01         | 0.05         | 0.03         | 0.01         | 0.05         | 0.03         |
| BERT-MT   | 0.271        | 0.099        | 0.173        | <b>0.334</b> | 0.507        | 0.239        | <b>0.127</b> | 0.247        | 0.368        |
|           | 0.05         | 0.02         | 0.40         | 0.03         | 0.04         | 0.03         | 0.02         | 0.39         | 0.04         |
| Our Model | <b>0.323</b> | <b>0.167</b> | <b>0.229</b> | 0.291        | <b>0.545</b> | 0.254        | 0.114        | <b>0.269</b> | <b>0.392</b> |

Since these results were derived from different models, we cannot guarantee that their variances are equal or similar. Therefore, we opted for Welch's t-test instead of the standard t-test.

As we can see in Table 7, the comprehensive performance of all models is moderately low, which indicates the task's difficulty. The reason is that Emotion Generation is a seven-classes generation task without knowing the response content. Besides, it also suffers from the imbalance issue, as shown in the data distribution in Table 6. When we look at the average F-scores, w-avgs of all models are higher than the m-avgs. It also verifies that generating the majority emotion (*i.e.*, *Neutral*) is easier than other emotions. Our model statistical-significantly outperforms all other baseline models in w-avg, *Anger*, *Disgust*, and *Neutral*. Besides, it also surpasses multiple baseline models in m-avg, *Surprise*, *Joy*, and *Sadness*. The results verify that our model design efficiently utilizes the mood transition process and the personality to improve Emotion Generation.

The lowest performance occurs when we only use the BERT-base model for the Emotion Generation task. The m-avg is **0.242** and the F-score of *Disgust* is only **0.012**. However, when we integrate the mood state transition and the personality trait into the BERT model, the performances are immediately improved, especially in generating *Disgust*. It suggests that our method improves the base model by raising the performance of generating minority emotions.

We first compare the performances of BERT, BERT-Mood, and BERT-MT to illustrate how different utilizations of mood states influence Emotion Generation. By comparing BERT and BERT-Mood, we can see that integrating the mood transition prediction task decreases the F-scores of the majority of emotions like *Anger*, *Joy* and *Neutral* but increase the rest of minorities. Consequently, although the m-avg remains the same, the w-avg goes down. So, predicting the mood state helps generate minority emotions.

Then we look at BERT-MT compared with BERT-Mood, F-scores of *Fear*, *Sadness*, and *Fear* decrease, while the F-scores of all other emotions grow up. Besides, both the m-avg and the w-avg improve than BERT-Mood. So, integrating the mood transition process in Emotion Generation fixed the performance issue of the majority of emotions in BERT-Mood.

Then, we focus on the personality. Comparing BERT-P and BERT, we can see that: although integrating the personality trait can slightly improve the m-avg, some majority emotions, *e.g.*, *Neutral*, *Anger* are lower so that the w-avg also decreases. However, when utilizing the personality

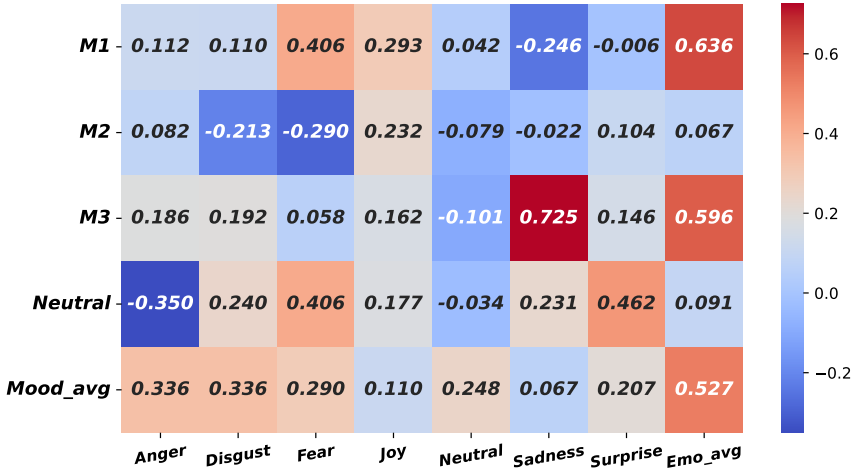


Fig. 8. Spearman correlations among F-scores of mood transition prediction and Emotion Generation.

as the mood transition weight in our model, we can see the performance is much better. So, in the Emotion Generation task, directly concatenating the personality trait causes unstable improvement, but modeled as the mood transition weight, the personality helps achieve robust and significant enhancement.

## 7.2 What is the correlation between the mood transition process and Emotion Generation?

To elucidate the correlation between the mood transition process and Emotion Generation, we present the impact of distinct mood transition processes on the generation of diverse emotions. Specifically, we conduct experiments of Emotion Generation with 10 random seeds and calculate the Spearman correlation coefficient between the results of mood transition (the F-scores of accurately predicting each mood in the response) and the F-scores of different emotions. The correlation is illustrated in Figure 8. In particular, the last row in Table 8 indicates the Spearman correlation coefficient between the average F-scores of all the mood transition predictions and the F scores of each emotion generation. Similarly, the last column in Table 8 is the Spearman correlation coefficient between the average F-scores of all emotion generations and the F scores of each mood transition prediction.

In general, Emotion Generation is partially conditioned on the results of mood transition; accurate prediction of mood transition means that the information provided to Emotion Generation is correct. The last row of Figure 8 shows that the overall predictions for mood transitions are positively correlated with the generation of all emotions. Besides, the last column shows that the transition predictions of all single mood states are also positively correlated with Emotion Generation.

The last column shows that the F-score of predicting *M1* has the highest correlation with the average result of Emotion Generation. We infer the reason is that emotions in *M1* (*Joy* and *Surprise*) take a large proportion of all emotions in PELD. Accurately predicting *M1* helps to correctly generate these emotions. Moreover, the high correlation between *M3* and *Emo\_avg* shows that *M3* only corresponds to the minority emotion *Sadness*, and correctly predicting *M3* will help eliminate interferences in Emotion Generation.

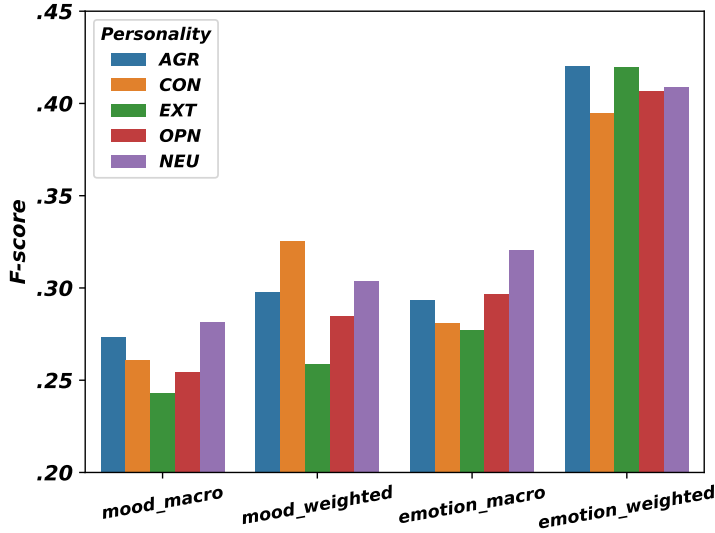


Fig. 9. This bar plot illustrates the F-scores of predicting Mood Transition (mood\_macro and mood\_weighted) and Emotion Generation (emo\_macro and emo\_weighted) for speakers with different personalities using our method.

We further look at the correlations among specific mood states and emotions. The accurate prediction of  $M3$  is highly correlated with the generation of *Sadness*, and it is the highest correlation value among all. This is also due to  $M3$  only corresponding to one emotion *Sadness*, as mentioned before. However, we also noticed that the results of some mood transitions are negatively correlated with Emotion Generation, such as *Neutral* mood states and *Anger*, or  $M2$  and *Fear*. It indicates that when these mood transitions are accurately predicted, some Emotion Generation will be decreased. We infer the reason behind this might be that these two different losses  $\mathcal{L}_{mood}$  and  $\mathcal{L}_{emo}$  in our model inevitably conflict with each other. However, because the introduction of mood transition has improved the Emotion Generation as a whole, as mentioned above, we retain the current design.

### 7.3 Do certain personalities easier to predict mood transition and Emotion Generation

We analyze the results of our method (averaged from 10 random seeds) on the test set based on different personalities. The results are shown in Figure 9.

Specifically, for the triples in the test set, we manually labeled the speakers with binary big-five personality categories according to the original annotations in FriendsPersona. It means if the triples occur in FriendsPersona, we directly adopt the annotations; otherwise, we manually label the personality categories of the speakers.

The results in Figure 9 show that the mood transition results of different personalities (both mood\_macro and mood\_weighted) vary a lot. The mood states of speakers with NEU are easier to predict, while speakers with EXT are more difficult to be captured, relatively. We infer the reason might be: NEU speakers are more sensitive in conversation. Their mood states will be easier to be influenced by the dialog content. Therefore, the dialog content as the mood transition variables in our method provides more valuable information. While EXT speakers tend to be outgoing and talkative in most dialog contexts, their subtle mood states transitions are more difficult to capture in our method.

As for the Emotion Generation, *emotion\_macro* of different personalities also vary, but *emotion\_weighted* of different personalities are more similar to each other. The reason is that correctly predicting the majority emotion *Neutral* is counted more in calculating *emotion\_weighted* compared with *emotion\_macro*. However, *Neutral* is expressed the most among all speakers regardless of their personalities. Besides, the patterns of speakers of EXT and NEU remain similar in *emotion\_macro* due to their characteristics.

#### 7.4 Is there any cases to show how our model generate the emotions for the response?

We conduct a case study and analyze the result samples of our method on the test set to show how our method works in real conversation scenarios. As shown in Table 8, the Utterance 1 and the Ground Truth Response are from the same speaker with a known personality.

First, we show two representative samples where our model correctly generates the emotion for the response. The first sample occurs when the two speakers are talking about the fetal movement, both speakers are in the *Joy* emotion in the context, so our model generates the *Joy* for the response and is verified in the ground truth. In the second sample, Rachel is sad in utterance 1; after being comforted and helped by Michael in utterance 2, she becomes happy in the response. This sample verifies that our model is able to generate the *Joy* emotion with the personality of Rachel. With relatively higher Extroversion, Rachel's mood state is easier to be influenced by others and express emotion with higher Arousal.

Moreover, we also show an example where our model makes mistakes. In the third case, our model wrongly generates the *Angry* because the two speakers argue angrily in the context. However, the content of Rachel's response clearly demonstrates her willingness to attend the lecture. This case also illustrates the limitation of our method: without knowing and modeling enough background knowledge of the speakers, it is difficult to generate appropriate emotion in conversations.

## 8 CONCLUSION AND FUTURE WORK

In this work, we raise a new task of personality-affected emotion generation and propose a new perspective to solve it through personality-affected mood transition. Besides, we construct a dialog script dataset PELD with emotion and personality labels to facilitate related research. We conduct extensive experiments on PELD to evaluate the effectiveness of our method. The results verify that integrating the personality and the mood transition regression significantly improves the performance in emotion generation, especially in minority emotions.

In future research, we intend to focus on two issues: (1) the personality effects on emotions in the multi-modality scenario, and (2) personality effects on response generation. Facial expressions, voices, gestures, and environment information are also vital in emotional interaction, but they are not captured in purely text-based dialog systems. Besides, as seen from statistics in PELD, the most common emotion in the dialog scripts is still *Neutral*. One possible reason is that other subtle affective information is not captured in the text. Therefore, our future works will continue to investigate the personality effects on emotions in the multi-modality scenario. Besides, the influence of personality on language usage is also studied in existing works [8, 63]. To construct intelligent dialog systems with personality, it is also important to investigate how the given personality influences the semantic content in response generation [20].

## 9 ACKNOWLEDGEMENT

This work is conducted at the Research Institute for Artificial Intelligence of Things (RIAIoT) at PolyU and supported by the GRF-RGC General Research Fund 2021/22 (No.15204921), and financial supported by Lingnan University (LU) (DB23A4) and Lam Woo Research Fund at LU (871236).

Table 8. Emotion generation illustration on dialog samples in PELD.

| Utterance 1   | Utterance 2   | Generated Emotion | Ground Truth Response  |
|---|---|-------------------|--|
| <b>Joey:</b> Oh, yeah. he's got that great baby smell. get a whiff of his head. ( <b>Joy</b> )  | <b>Caroline:</b> I think my uterus just skipped a beat! ( <b>Joy</b> )  | <b>Joy</b>        | <b>Joey:</b> What'd I tell you? What'd i tell you? ( <b>Joy</b> )  |
| <b>Rachel:</b> Oh, look at me, look at me. Oh, I'm on a date with a really great guy, all I can think about is Ross and his cat and his... Julie. I just want to get over him. God, I just, why can't I do that? ( <b>Sadness</b> ) | <b>Machael:</b> Oy. Look, I've been through a divorce, trust me, you're gonna be fine. You just can't see it now because you haven't had any closure. ( <b>Sadness</b> )  | <b>Joy</b>        | <b>Rachel:</b> Yeah! Closure. That's what it is, that's what I need. God, you're brilliant! ( <b>Joy</b> ) |
| <b>Rachel:</b> Is that funny? Am I supposed to be laughing? ( <b>Anger</b> )  | <b>Ross:</b> I don't know, you thought 'See you Saturday' was funny. Look honey, Mark is in fashion okay, I like having a friend that I can share this stuff with. You guys would never want to go to a lecture with me. ( <b>Anger</b> ) | <b>Anger</b>      | <b>Rachel:</b> Pa-haa!! I would love to go with you. ( <b>Joy</b> )  |

## REFERENCES

- [1] Elisabeth André, Martin Klesen, Patrick Gebhard, Steve Allen, and Thomas Rist. 1999. Integrating models of personality and emotions into lifelike characters. In *International Workshop on Affective Interactions*. Springer, 150–165.
- [2] Nabiha Asghar, Pascal Poupart, Jesse Hoey, Xin Jiang, and Lili Mou. 2018. Affective neural response generation. In *European Conference on Information Retrieval*. Springer, 154–166.
- [3] G Ball and J Breese. 2000. Emotion and personality ina conversational character. *Embodied Conversational Agents*. MIT Press, Cambridge (2000).
- [4] C Bauerhenne, A Gammoudi, M Moussaoui, J Reichelt, and J Wang. 2020. Emotional states and personality profiles in Conversational AI. (2020).
- [5] Sven Buechel and Udo Hahn. 2016. Emotion analysis as a regression problem—dimensional models and their implications on emotion representation and metrical evaluation. In *ECAI 2016*. IOS Press, 1114–1122.
- [6] Sven Buechel and Udo Hahn. 2017. Emobank: Studying the impact of annotation perspective and representation format on dimensional emotion analysis. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers*. 578–585.
- [7] 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, 4 (2008), 335–359.
- [8] Kaitlin E Cannava, Andrew C High, Susanne M Jones, and Graham D Bodie. 2018. The stuff that verbal person-centered support is made of: Identifying linguistic markers of more and less supportive conversations. *Journal of Language and Social Psychology* 37, 6 (2018), 656–679.
- [9] Kenneth Mark Colby. 1975. *Artificial paranoia: a computer simulation of paranoid process*. Pergamon Press.
- [10] Pierre Colombo, Wojciech Witon, Ashutosh Modi, James Kennedy, and Mubbasir Kapadia. 2019. Affect-driven dialog generation. *arXiv preprint arXiv:1904.02793* (2019).



- [11] Paul T Costa and Robert R McCrae. 1992. Normal personality assessment in clinical practice: The NEO Personality Inventory. *Psychological assessment* 4, 1 (1992), 5.
- [12] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805* (2018).
- [13] Paul Ed Ekman and Richard J Davidson. 1994. *The nature of emotion: Fundamental questions*. Oxford University Press.
- [14] Howard S Friedman and Margaret L Kern. 2014. Personality, well-being, and health. *Annual review of psychology* 65 (2014), 719–742.
- [15] Deepanway Ghosal, Navonil Majumder, Soujanya Poria, Niyati Chhaya, and Alexander Gelbukh. 2019. Dialoguecn: A graph convolutional neural network for emotion recognition in conversation. *arXiv preprint arXiv:1908.11540* (2019).
- [16] Meng-Ju Han, Chia-How Lin, and Kai-Tai Song. 2010. A design for smooth transition of robotic emotional states. In *2010 IEEE Workshop on Advanced Robotics and its Social Impacts*. IEEE, 13–18.
- [17] Meng-Ju Han, Chia-How Lin, and Kai-Tai Song. 2012. Robotic emotional expression generation based on mood transition and personality model. *IEEE transactions on cybernetics* 43, 4 (2012), 1290–1303.
- [18] Devamanyu Hazarika, Soujanya Poria, Rada Mihalcea, Erik Cambria, and Roger Zimmermann. 2018. Icon: Interactive conversational memory network for multimodal emotion detection. In *Proceedings of the 2018 conference on empirical methods in natural language processing*. 2594–2604.
- [19] Devamanyu Hazarika, Soujanya Poria, Amir Zadeh, Erik Cambria, Louis-Philippe Morency, and Roger Zimmermann. 2018. Conversational memory network for emotion recognition in dyadic dialogue videos. In *Proceedings of the conference. Association for Computational Linguistics. North American Chapter. Meeting*, Vol. 2018. NIH Public Access, 2122.
- [20] Minlie Huang, Xiaoyan Zhu, and Jianfeng Gao. 2020. Challenges in building intelligent open-domain dialog systems. *ACM Transactions on Information Systems (TOIS)* 38, 3 (2020), 1–32.
- [21] Chika Itoh, Shohei Kato, and Hidenori Itoh. 2009. Mood-transition-based emotion generation model for the robot’s personality. In *2009 IEEE International Conference on Systems, Man and Cybernetics*. IEEE, 2878–2883.
- [22] Hang Jiang, Xianzhe Zhang, and Jinho D Choi. 2019. Automatic Text-based Personality Recognition on Monologues and Multiparty Dialogues Using Attentive Networks and Contextual Embeddings. *arXiv preprint arXiv:1911.09304* (2019).
- [23] Michael Johns and Barry G Silverman. 2001. How emotions and personality effect the utility of alternative decisions: a terrorist target selection case study. *Center for Human Modeling and Simulation* (2001), 10.
- [24] Henrik Kessler, Alexander Festini, Harald C Traue, Suzanne Filipic, Michael Weber, and Holger Hoffmann. 2008. SIMPLEX—simulation of personal emotion experience. *Affective Computing* (2008), 255–270.
- [25] Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).
- [26] Kwan Min Lee, Wei Peng, Seung-A Jin, and Chang Yan. 2006. Can robots manifest personality?: An empirical test of personality recognition, social responses, and social presence in human–robot interaction. *Journal of communication* 56, 4 (2006), 754–772.
- [27] Changmao Li and Jinho D Choi. 2020. Transformers to Learn Hierarchical Contexts in Multiparty Dialogue for Span-based Question Answering. *arXiv preprint arXiv:2004.03561* (2020).
- [28] Jiwei Li, Michel Galley, Chris Brockett, Georgios P Spithourakis, Jianfeng Gao, and Bill Dolan. 2016. A persona-based neural conversation model. *arXiv preprint arXiv:1603.06155* (2016).
- [29] Jia Li, Xiao Sun, Xing Wei, Changliang Li, and Jianhua Tao. 2019. Reinforcement Learning Based Emotional Editing Constraint Conversation Generation. *arXiv preprint arXiv:1904.08061* (2019).
- [30] Minglei Li, Qin Lu, Yunfei Long, and Lin Gui. 2017. Affective State Prediction of Contextualized Concepts. In *IJCAI 2017 Workshop on Artificial Intelligence in Affective Computing*. PMLR, 45–57.
- [31] Minglei Li, Qin Lu, Yunfei Long, and Lin Gui. 2017. Inferring affective meanings of words from word embedding. *IEEE Transactions on Affective Computing* 8, 4 (2017), 443–456.
- [32] Yanran Li, Hui Su, Xiaoyu Shen, Wenjie Li, Ziqiang Cao, and Shuzi Niu. 2017. Dailydialog: A manually labelled multi-turn dialogue dataset. *arXiv preprint arXiv:1710.03957* (2017).
- [33] 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*. 2980–2988.
- [34] Zijie Lin, Bin Liang, Yunfei Long, Yixue Dang, Min Yang, Min Zhang, and Ruifeng Xu. 2022. Modeling Intra-and Inter-Modal Relations: Hierarchical Graph Contrastive Learning for Multimodal Sentiment Analysis. In *Proceedings of the 29th International Conference on Computational Linguistics*. 7124–7135.
- [35] Zhaojiang Lin, Andrea Madotto, Jamin Shin, Peng Xu, and Pascale Fung. 2019. MoEL: Mixture of Empathetic Listeners. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*. Association for Computational Linguistics, Hong Kong, China, 121–132. <https://doi.org/10.18653/v1/D19-1012>

- [36] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692* (2019).
- [37] Yukun Ma, Khanh Linh Nguyen, Frank Z Xing, and Erik Cambria. 2020. A survey on empathetic dialogue systems. *Information Fusion* 64 (2020), 50–70.
- [38] François Mairesse and Marilyn A Walker. 2011. Controlling user perceptions of linguistic style: Trainable generation of personality traits. *Computational Linguistics* 37, 3 (2011), 455–488.
- [39] Navonil Majumder, Pengfei Hong, Shanshan Peng, Jiankun Lu, Deepanway Ghosal, Alexander Gelbukh, Rada Mihalcea, and Soujanya Poria. 2020. MIME: MIMicking emotions for empathetic response generation. *arXiv preprint arXiv:2010.01454* (2020).
- [40] Navonil Majumder, Soujanya Poria, Devamanyu Hazarika, Rada Mihalcea, Alexander Gelbukh, and Erik Cambria. 2019. Dialoguerrn: An attentive rnn for emotion detection in conversations. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 33. 6818–6825.
- [41] Robert P Marinier and John E Laird. 2007. Computational modeling of mood and feeling from emotion. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, Vol. 29.
- [42] Naoki Masuyama and Chu Kiong Loo. 2015. Robotic emotional model with personality factors based on Pleasant-Arousal scaling model. In *2015 24th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 19–24.
- [43] Naoki Masuyama, Chu Kiong Loo, and Manjeevan Seera. 2018. Personality affected robotic emotional model with associative memory for human-robot interaction. *Neurocomputing* 272 (2018), 213–225.
- [44] John D Mayer. 2004. (2004) What is Emotional Intelligence? (2004).
- [45] Albert Mehrabian. 1996. Analysis of the big-five personality factors in terms of the PAD temperament model. *Australian journal of Psychology* 48, 2 (1996), 86–92.
- [46] Albert Mehrabian. 1996. Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in temperament. *Current Psychology* 14, 4 (1996), 261–292.
- [47] Saif Mohammad. 2018. Obtaining reliable human ratings of valence, arousal, and dominance for 20,000 english words. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 174–184.
- [48] Gregory Park, H Andrew Schwartz, Johannes C Eichstaedt, Margaret L Kern, Michal Kosinski, David J Stillwell, Lyle H Ungar, and Martin EP Seligman. 2015. Automatic personality assessment through social media language. *Journal of personality and social psychology* 108, 6 (2015), 934.
- [49] Sungjoon Park, Jiseon Kim, Seonghyeon Ye, Jaeyeol Jeon, Hee Young Park, and Alice Oh. 2019. Dimensional Emotion Detection from Categorical Emotion. *arXiv preprint arXiv:1911.02499* (2019).
- [50] Tim Polzehl. 2016. *PERSONALITY IN SPEECH*. Springer.
- [51] Soujanya Poria, Devamanyu Hazarika, Navonil Majumder, Gautam Naik, Erik Cambria, and Rada Mihalcea. 2018. Meld: A multimodal multi-party dataset for emotion recognition in conversations. *arXiv preprint arXiv:1810.02508* (2018).
- [52] Soujanya Poria, Navonil Majumder, Rada Mihalcea, and Eduard Hovy. 2019. Emotion recognition in conversation: Research challenges, datasets, and recent advances. *IEEE Access* 7 (2019), 100943–100953.
- [53] Paulo Quaglio. 2009. Television dialogue: The sitcom Friends vs. natural conversation. Amsterdam, NL.
- [54] Hannah Rashkin, Eric Michael Smith, Margaret Li, and Y-Lan Boureau. 2018. Towards empathetic open-domain conversation models: A new benchmark and dataset. *arXiv preprint arXiv:1811.00207* (2018).
- [55] James A Russell. 1991. Culture and the categorization of emotions. *Psychological bulletin* 110, 3 (1991), 426.
- [56] James A Russell and Albert Mehrabian. 1977. Evidence for a three-factor theory of emotions. *Journal of research in Personality* 11, 3 (1977), 273–294.
- [57] H Andrew Schwartz, Johannes C Eichstaedt, Margaret L Kern, Lukasz Dziurzynski, Stephanie M Ramones, Megha Agrawal, Achal Shah, Michal Kosinski, David Stillwell, Martin EP Seligman, et al. 2013. Personality, gender, and age in the language of social media: The open-vocabulary approach. *PloS one* 8, 9 (2013), e73791.
- [58] Roman Shantala, Gennadiv Kyselov, and Anna Kyselova. 2018. Neural dialogue system with emotion embeddings. In *2018 IEEE First International Conference on System Analysis & Intelligent Computing (SAIC)*. IEEE, 1–4.
- [59] Ashish Sharma, Adam S Miner, David C Atkins, and Tim Althoff. 2020. A computational approach to understanding empathy expressed in text-based mental health support. *arXiv preprint arXiv:2009.08441* (2020).
- [60] JiaXing Shen, Mingyu Derek Ma, Rong Xiang, Qin Lu, Elvira Perez Vallejos, Ge Xu, Chu-Ren Huang, and Yunfei Long. 2020. Dual memory network model for sentiment analysis of review text. *Knowledge-Based Systems* 188 (2020), 105004.
- [61] Xiaohui Song, Longtao Huang, Hui Xue, and Songlin Hu. 2022. Supervised prototypical contrastive learning for emotion recognition in conversation. *arXiv preprint arXiv:2210.08713* (2022).
- [62] Xiao Sun, Xinmiao Chen, Zhengmeng Pei, and Fuji Ren. 2018. Emotional human machine conversation generation based on SeqGAN. In *2018 First Asian Conference on Affective Computing and Intelligent Interaction (ACII Asia)*. IEEE,

1–6.

- [63] Yla R Tausczik and James W Pennebaker. 2010. The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of language and social psychology* 29, 1 (2010), 24–54.
- [64] Timothy J Trull, Marika B Solhan, Sarah L Tragesser, Seungmin Jahng, Phillip K Wood, Thomas M Piasecki, and David Watson. 2008. Affective instability: measuring a core feature of borderline personality disorder with ecological momentary assessment. *Journal of abnormal psychology* 117, 3 (2008), 647.
- [65] A Wagner and E Briscoe. 2017. Psychological modeling of humans by assistive robots. In *Human Modelling for Bio-Inspired Robotics*. Elsevier, 273–296.
- [66] Steffen Walter, Jonghwa Kim, David Hrabal, Stephen Clive Crawcour, Henrik Kessler, and Harald C Traue. 2013. Transsituational individual-specific biopsychological classification of emotions. *IEEE Transactions on Systems, Man, and Cybernetics: Systems* 43, 4 (2013), 988–995.
- [67] Yan Wang, Jiayu Zhang, Jun Ma, Shaojun Wang, and Jing Xiao. 2020. Contextualized emotion recognition in conversation as sequence tagging. In *Proceedings of the 21th Annual Meeting of the Special Interest Group on Discourse and Dialogue*. 186–195.
- [68] Jason Wei and Kai Zou. 2019. Eda: Easy data augmentation techniques for boosting performance on text classification tasks. *arXiv preprint arXiv:1901.11196* (2019).
- [69] Wei Wei, Jiayi Liu, Xianling Mao, Guibing Guo, Feida Zhu, Pan Zhou, and Yuchong Hu. 2019. Emotion-aware Chat Machine: Automatic Emotional Response Generation for Human-like Emotional Interaction. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*. 1401–1410.
- [70] Bernard L Welch. 1947. The generalization of STUDENT’S problem when several different population variances are involved. *Biometrika* 34, 1-2 (1947), 28–35.
- [71] 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*. 5010–5020.
- [72] Zhiyuan Wen, Jiannong Cao, Ruosong Yang, and Senzhang Wang. 2020. Decode with template: Content preserving sentiment transfer. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*. 4671–4679.
- [73] Zhiyuan Wen, Jiannong Cao, Yu Yang, Haoli Wang, Ruosong Yang, and Shuaiqi Liu. 2023. DesPrompt: Personality-descriptive prompt tuning for few-shot personality recognition. *Information Processing & Management* 60, 5 (2023), 103422.
- [74] Tal Yarkoni. 2010. Personality in 100,000 words: A large-scale analysis of personality and word use among bloggers. *Journal of research in personality* 44, 3 (2010), 363–373.
- [75] 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* (2017).
- [76] Rohola Zandie and Mohammad H Mahoor. 2020. Emptransfo: A multi-head transformer architecture for creating empathetic dialog systems. *arXiv preprint arXiv:2003.02958* (2020).
- [77] Tiancheng Zhao, Ran Zhao, and Maxine Eskenazi. 2017. Learning discourse-level diversity for neural dialog models using conditional variational autoencoders. *arXiv preprint arXiv:1703.10960* (2017).
- [78] Chujie Zheng, Yong Liu, Wei Chen, Yongcai Leng, and Minlie Huang. 2021. Comae: a multi-factor hierarchical framework for empathetic response generation. *arXiv preprint arXiv:2105.08316* (2021).
- [79] Peixiang Zhong, Chen Zhang, Hao Wang, Yong Liu, and Chunyan Miao. 2020. Towards persona-based empathetic conversational models. *arXiv preprint arXiv:2004.12316* (2020).
- [80] Hao Zhou, Minlie Huang, Tianyang Zhang, Xiaoyan Zhu, and Bing Liu. 2018. Emotional chatting machine: Emotional conversation generation with internal and external memory. In *Thirty-Second AAAI Conference on Artificial Intelligence*.
- [81] Li Zhou, Jianfeng Gao, Di Li, and Heung-Yeung Shum. 2020. The design and implementation of xiaoice, an empathetic social chatbot. *Computational Linguistics* 46, 1 (2020), 53–93.