Automatically Select Emotion for Response via Personality-affected Emotion Transition

Zhiyuan Wen, Jiannong Cao, Ruosong Yang, Shuaiqi Liu, Jiaxing Shen
Department of Computing,
The Hong Kong Polytechnic University
Kowloon, Hong Kong, China
{cszwen, csjcao, csryang, cssliu, jiaxshen}@comp.polyu.edu.hk

Abstract

To provide consistent emotional interaction with users, dialog systems should be capable to automatically select appropriate emotions for responses like humans. However, most existing works focus on rendering specified emotions in responses or empathetically respond to the emotion of users, yet the individual difference in emotion expression is overlooked. This may lead to inconsistent emotional expressions and disinterest users. To tackle this issue, we propose to equip the dialog system with personality and enable it to automatically select emotions in responses by simulating the emotion transition of humans in conversation. In detail, the emotion of the dialog system is transitioned from its preceding emotion in context. The transition is triggered by the preceding dialog context and affected by the specified personality trait. To achieve this, we first model the emotion transition in the dialog system as the variation between the preceding emotion and the response emotion in the **V**alence-**A**rousal-**D**ominance (VAD) emotion space. Then, we design neural networks to encode the preceding dialog context and the specified personality traits to compose the variation. Finally, the emotion for response is selected from the sum of the preceding emotion and the variation. We construct a dialog dataset with emotion and personality labels and conduct emotion prediction tasks for evaluation. Experimental results validate the effectiveness of the personality-affected emotion transition.[1]

1 Introduction

Emotional intelligence can be considered a mental ability to reason validly with emotional information, and the action of emotions to enhance thought (Mayer, 2004). Hence, to create dialog systems with emotional intelligence during communication, it is necessary to enable the machine to understand the emotion of users, select appropriate response emotions and express in conversation.

Existing works either focus on rendering specified emotions in responses (Zhou et al., 2018; Colombo et al., 2019), or understanding the emotion of users and respond empathetically (Zandie and Mahoor, 2020; Zhong et al., 2020; Lin et al., 2019); but how to automatically select the emotion for response is seldom discussed. Wei et al. (2019) proposes to learn appropriate emotional responses from massive anonymous online dialogues. However, trained on conversations from different speakers, the dialog system ignores the individual difference of expressing emotions. This may lead to inconsistent emotional interactions and disinterest users as they may feel they are still talking to rigid machines.

In a dialog system, automatically selecting the emotion for response is to decide an emotion to be expressed facilitating the emotional response generation. Emotion selection can be modeled as the emotion transition (Thornton and Tamir, 2017), which refers to how the preceding emotion changes to the next, of the dialog system reacting to the dialog context. To achieve it like humans, it requires long-term patterns of thought, and behavior associated with an individual (Ball, 2000). Mehrabian (1996a) shows that the personality, *e.g.*, the big-five personality model (Costa and McCrae, 1992) also can be represented as temperament in the **V**alence-**A**rousal-**D**ominance (VAD) space for emotions (Mehrabian, 1996b).[2] The finding suggests that different personalities make different impacts on emotional expressions. Inspired by these works, we propose a personality-affected emotion

---

[2]It is Pleasure-Arousability-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.

transition model to endow personality to the dialog system, enabling it to select emotions that react to the dialog context affected by its given personality.

In our method, we model the emotion transition of the dialog system as the variation in the VAD space from its preceding emotion to the next emotion in the response to users. We first obtain the preceding emotion of the dialog system from the dialog context and project it into the VAD space as an emotion vector. Simultaneously, we endow a personality trait, a 5-dimension vector representing the strength of each dimension in the big-five personality traits, to the dialog system. Then, we design neural networks to encode the dialog context and the personality traits into the VAD space to compose the variation of emotion. Finally, the emotion for response is selected based on the sum of the preceding emotion and the variation.

To facilitate related researches, we construct the **P**ersonality **E**motion**L**ines **D**ataset (PELD), which includes 6,510 dialogue triples of daily conversations with emotion labels and annotated personality traits. The emotion labels and personality annotations are adopted from other researches (Poria et al., 2018; Zahiri and Choi, 2017; Jiang et al., 2019) analyzing the script of a famous TV series *Friends*[3]. We conduct emotion prediction tasks on the PELD dataset to evaluate the effectiveness of our method. The results suggest that the personality-affected emotion transition does contribute to better accuracy in emotion selection. To summary, our contributions are as follows:

- We raise the problem of automatically select the emotion for response in conversation and propose a new perspective to solve it through personality-affected emotion transition.

- We construct a dialog script dataset with emotion and personality labels and analyze the patterns of emotion transitions in our dataset to facilitate related researches.

- We evaluate the effectiveness of our proposed method on emotion prediction tasks and analyze the effects of personality and emotion transition respectively.

2   Related Works

Our research is related to the emotional dialog systems, and the personality influence on emotion ex-

pression in psychology and Human-Computer Interaction (HCI). So, we review existing works in the two aspects as follows.

2.1   Emotional Dialog Systems

The concept of the emotional dialog system first occurred in (Colby, 1975), where a rule-based emotion simulation chatbot was proposed. Microsoft introduced the Xiaoice (Zhou et al., 2020), an empathetic social chatbot that is able to recognize users' emotional needs, in 2014. Related researches become popular recently since Zhou et al. (2018) proposed the Emotional Chatting Machine to exploit the deep learning approach in building a large-scale emotionally aware conversational bot. Most existing works focus on incorporating specified emotion factors into neural response generation. Shantala et al. (2018) trains emotional embeddings based on context and then integrated them into response generation. Colombo et al. (2019) controls the emotional response generation with both categorical emotion representations and continuous word representations in VAD space (Mohammad, 2018). Moreover, Asghar et al. (2018) proposes an affectively diverse beam search for decoding. Besides, reinforcement learning is also adopted to encourage response generation models to render specified emotions. Li et al. (2019) combines reinforcement learning with emotional editing constraints to generate meaningful and customizable emotional replies. (Sun et al., 2018) also uses an emotion tag to partially rewarding the model to express specified emotion.

However, it is impractical to always specify response emotions for dialog systems in real application scenarios. To simulate the emotional interaction among humans, Wei et al. (2019) designs 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 to be impractical.

2.2   Personality Effects on Emotions

Emotion is a complex psychological experience of an individual's state of mind as interacting with people or environmental influences (Han et al., 2012). The **P**leasure-**A**rousal-**D**ominance (PAD) (Mehrabian, 1996b) or **V**alence-**A**rousal-**D**ominance (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 (Johns and Silverman, 2001) and probabilistic models (André et al., 1999). Mehrabian (1996a) utilized the five factors of personality (Costa and McCrae, 1992) to represent the VAD temperament model through linear regression analysis. This finding is widely used to design robots having non-verbal emotional interaction with users (Han et al., 2012; Masuyama et al., 2018), where the pre-defined personalities of robots affect their propensity of simulated emotion transitions.

To integrate the analysis above into Artificial Intelligence, some researchers in HCI borrow the idea and design facial emotional expressions for humanoid robots. Ball (2000) utilizes models of emotions and personality encoded as Bayesian networks to generate empathetic behaviors or speech responses to users in conversation. Han et al. (2012) employed five factors of personality to a 2D (pleasure-arousal) scaling model to represent a robotic emotional model. Masuyama et al. (2018) introduces an emotion-affected associative memory model for robots expressing emotions. While in NLP, though the VAD space is adopted to model emotions in some researches (Mohammad, 2018; Colombo et al., 2019; Asghar et al., 2018), the personality influence on emotion in dialogues is still an open problem.

3 Methodology

3.1 Problem Definition

We research on enabling the dialog system to automatically select emotions for response through the personality-affected emotion transition.

Formally, a dyadic emotional conversation between the user and the dialog system contains the dialog context $C = \{U_1, U_2, ..., U_{n-1}\}$ including all the preceding $n - 1$ utterances from both the user and the dialog system, the preceding emotion E_i expressed in $U_i \in C$ which is the last utterance from the dialog system, and the response emotion E_r for the dialog system to facilitate generating the next emotional response U_n to the user. We specify a personality trait P_n to the dialog system and enable it to select response emotion E_r through the personality-affected emotion transition model

F_{ET}:

$$E_r = F_{ET}(E_i|P_n, C) \tag{1}$$

where E_r is transitioned from E_i. The transition is triggered by the preceding dialog context C and affected by the specified personality trait P_n. In the following content, we will introduce how we model this process in detail.

3.2 Preliminaries

3.2.1 Emotions in the VAD space

Assuming in the problem above, emotions in all emotional utterances can be categorized into the six basic emotions: *Anger*, *Disgust*, *Fear*, *Joy*, *Sadness*, and *Surprise* (Ekman and Davidson, 1994). We project the basic emotions into the Valence-Arousal-Dominance (VAD) space as Table 1 refer to the analysis result in (Russell and Mehrabian, 1977)[4]. The VAD space indicates emotion intensity in three different dimensions, where the valence measures the positivity/negativity, arousal the excitement/calmness, and dominance the powerfulness/weakness. As for the utterances with no explicit emotion, we use the *Neutral* with (0.00, 0.00, 0.00) as the VAD vector.

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)

Table 1: Emotions in the VAD Space.

3.2.2 Personalities in the VAD space

Meanwhile, the big-five personality traits (OCEAN, shown in Table 2) are widely used for psychological analysis. Mehrabian (1996a) proposed a temperament model shown in Equation 2 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.

---

[4]*fear* and *Joy* correspond to *Terrified* and *Happy* in the reference table.
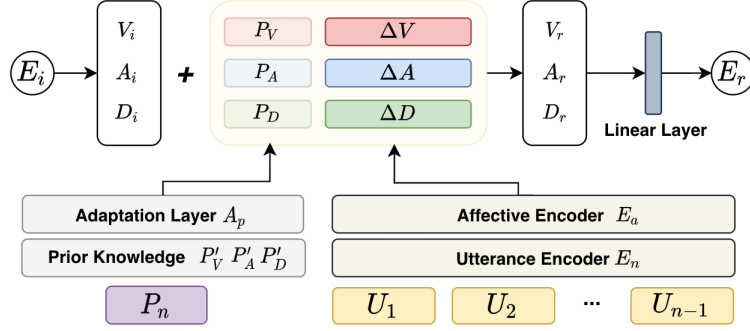
Figure 1: The Model Illustration

$$P_V = 0.21E + 0.59A + 0.19N$$
$$P_A = 0.15O + 0.30A - 0.57N \quad (2)$$
$$P_D = 0.25O + 0.17C + 0.60E - 0.32A$$

3.3 Personality-affected Emotion Transition

Based on the problem definition and the preliminaries above, we design the Personality-affected Emotion Transition model as illustrated in Figure 1. Our model mainly include three modules: the personality effect on emotions in the left lower part, the context encoding in the right lower part, and the emotion transition in the top half in Figure 1. We will introduce these three modules in detail as follow.

3.3.1 Personality Effect on Emotions

In our model, the personality of the dialog system is specified as a 5-dimensional vector $P_n = [O, C, E, A, N]$ representing the strength in Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism, respectively.

The temperament of personality in the VAD space (shown in Equation 2) is widely used as weighting parameters for emotion transition of robots in HCI works (Han et al., 2012; Masuyama et al., 2018). However, the numeric coefficients in Equation 2 are summarized from analysis of questionnaire results from 72 participants (Mehrabian, 1996a), which are not suitable to directly adopted

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.

Table 2: The OCEAN personality traits and description (Costa and McCrae, 1992)

as hyper-parameters in the model design. Hence, we choose to adopt the analysis results in Equation 2 as prior knowledge and learn suitable coefficients for personality by neural networks. First, we still calculate P'_V, P'_A, P'_D from the personality P_n by Equation 2; then we use P'_V, P'_A, P'_D as initialized input for an adaptation layer A_p to learn the weighting parameters P_V, P_A, P_D that suitable for the training data.

3.3.2 Context Encoding

The dialog context acts as a set of parameters that may influence a person to speak an utterance while expressing a certain emotion (Poria et al., 2018). In the VAD space, the emotion transition is regarded as the variation from one point (the preceding emotion) to another point (the next emotion). Thus, we generate the emotion transition variations $\Delta V, \Delta A, \Delta D$ from the semantic representations of the preceding dialog context C.

$$R_c = E_n(U_1) \oplus E_n(U_2)... \oplus E_n(U_{n-1})$$
$$\Delta V, \Delta A, \Delta D = E_a(R_c) \quad (3)$$

We fine-tune the pre-trained RoBERTa[5] (Liu et al., 2019) encoder, a famous pre-trained language model whose performance is widely validated in many natural language understanding tasks, to first extract the semantic representations $E_n(U_1), ..., E_n(U_{n-1})$ of all $n-1$ utterances in C. Then, we concatenate the semantic representations of utterances to obtain the overall context semantics R_c. Finally, $\Delta V, \Delta A, \Delta D$ are calculated by feeding R_c into an affective encoder E_a, which extract the affective information from R_c in the aspect of V, A, D, respectively.

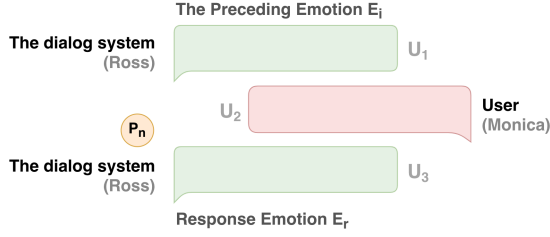[5]Here we adopt the pre-trained RoBERTa-base model.

Figure 2: A triple example in PELD. The dyadic conversation between Ross and Monica (two main roles in *Friends*, P_n is the personality of Ross. The dialog system is set as Ross and talk with the user set as Monica in this example.

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]

Table 3: Personalities of *Friends* main roles in PELD.

3.3.3 Emotion Transition

After we obtain the weighting parameters P_V, P_A, P_D and the emotion transition variation $\Delta V, \Delta A, \Delta D$, the emotion for response is generated by the sum of the VAD vectors of the preceding emotion and the weighted variation, as shown in Equation 4.

$$
\begin{aligned}
V_r &= V_i + P_V \cdot \Delta V \\
A_r &= A_i + P_A \cdot \Delta A \\
D_r &= D_i + P_D \cdot \Delta D \\
E_r &= F_c(V_r, A_r, D_r)
\end{aligned}
\tag{4}
$$

where the V_i, A_i, D_i are the VAD vectors of E_i, and the V_r, A_r, D_r are the emotion transition results in the VAD space. To alleviate the errors of using the numeric value in calculated VAD vectors, we add a linear layer F_c to transform V_r, A_r, D_r into a probability distribution on the discrete emotion categories. The output E_r is the emotion with the largest probability.

4 The PELD Dataset

4.1 Dataset Construction & Statistics

To facilitate related researches, we construct the **P**ersonality **E**motion**L**ines **D**ataset (**PELD**), an emotional dialog dataset with personality traits for speakers. As labeling online conversation on social media with speakers' personalities is time-consuming and may cause privacy issues, we turn to research on the dialogue script of a famous TV series *Friends*. This classic script is widely analyzed in many dialog researches (Li et al., 2016; Li and Choi, 2020; Jiang et al., 2019).

In PELD, each sample is represented as a dialog triple ($C = \{U_1, U_2, U_3\}, \{E_i, E_r\}, P_n$), shown in Figure 2) as a dyadic conversation. E_i and E_r

are emotions expressed in U_1 and U_3, respectively. The utterances and their emotion labels are mainly adopted from the dialogues in the MELD (Poria et al., 2018) and the EmoryNLP dataset (Zahiri and Choi, 2017), two famous datasets analyzing emotional expressions in *Friends*. 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 (Jiang et al., 2019). Refer to the annotations, a role may exhibit different aspects of its personality in different dialogues. We only keep the personality traits of the six main roles in *Friends* for confidence as these annotations are most frequent. For each of the main roles, we average their annotated personality traits in all the dialogues by $P_n = \frac{1}{K} \sum_{i=1}^{K} P_i$ for simplification, where K is the number of annotations. The averaged results are shown in Table 3.

We split the PELD into **Train**, **Valid**, and **Test** set with portion around 8:1:1. The total number of utterances in PELD (10,648) is less than the sum of the original MELD (13,708) and the EmoryNLP (9,489), as not all dialogues are suitable to construct triples including main roles. The overall statistics of the dataset is shown in Table 4.

Similar to existing emotional conversation datasets (Li et al., 2017; Busso et al., 2008), PELD also suffers the emotion imbalance issue. Utterances labeled as *Neutral* are the majority, while *Fear* and *Disgust* only take a small portion. Though it reflects the real 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 EDA proposed in (Wei and Zou, 2019). But most of the synthetic samples are either odd or the same as the original samples. The

Basic Statistics	Train	Valid	Test	Total
#Triple	5273	586	651	6510
#Unique Uttr.	9306	1518	1675	10468
Avg. Uttr. Length	9.26	9.33	8.95	9.32
#Emotion	**Train**	**Valid**	**Test**	**Total**
Anger	1863	236	241	2340
Disgust	312	32	32	376
Fear	1101	114	131	1346
Joy	2863	326	344	3533
Neutral	7055	756	890	8701
Sadness	1088	121	136	1345
Surprise	1537	173	179	1889
#Sentiment	**Train**	**Valid**	**Test**	**Total**
Positive	4400	499	523	5422
Neutral	7055	756	890	8701
Negative	4364	503	540	5407
#Triple of Main Roles	**Train**	**Valid**	**Test**	**Total**
Chandler	880	97	108	1085
Joey	912	109	102	1123
Monica	850	94	107	1051
Phoebe	782	87	103	972
Rachel	921	112	123	1156
Ross	928	87	108	1123

Table 4: Basic Statistics in PELD.

reason might be there are limited options for short sentences as utterances in conversation to replace synonyms, add or delete words.

Another way to alleviate the imbalance issue is to expand the granularity of emotion to sentiment. As mentioned in **3.2**, in the VAD space, the Valence dimension of emotions measures the positivity and negativity, we can categorize the emotions into sentiments according to the values of Valance; i.e., positive emotions: *Joy* and *Surprise*; negative emotions: *Anger*, *Disgust*, *Fear*, and *Sadness*. Thus, the distribution of sentiments in PELD is also shown in Table 4. Besides, dialog triples of six main roles (each triple corresponds to a main role with the personality trait) are averagely distributed in all train, valid, and test sets in PELD.

4.2 Emotion Transitions in PELD

After constructed PELD, we further explore the dataset in the aspect of emotion transitions. As the triples in PELD are constructed for analyzing the emotion transitions between E_i in U_1 and the E_r in U_3. Table 5 shows the emotion and sentiment distributions in the U_1 and U_3, respectively. Be-

sides, we also count the sentiments of emotions in U_1 and U_3 denoted as S_1 and S_3. We can see that for both emotion and sentiment, the distributions in U_1 and U_3 are similar, which means the transition of emotions and sentiments are equitable in PELD triples. Besides, the proportions of all emotions and sentiments are also similar to the overall statistics of PELD, which suggests that the emotions and sentiments in PELD are also average distributed in the triples.

Since emotion transitions are affected by the personality traits as discussed above, we exhibit the emotion transition patterns for different roles with different personality traits in Figure 3. Although the emotion transitions are also correlated to the dialog context, we can still find patterns through these transition matrixes[6].

In general, among the six transition matrixes, all the first columns are in deeper colors, which indicates most transitions occur from other emotions to *Neutral* as it is the majority emotion in PELD. Besides, blocks with deeper color also more likely to occur around or in the diagonals of the transition matrixes; it suggests the preceding emotions tend to transition to the same or similar emotions. As for individual roles, 0.59 of the *Anger* from Rachel remains the same in dialog triples, while for other roles, most *Anger* emotions are transferred to *Neutral* and *Anger*. Besides, most *Surprise* from Ross transfers to the *Neutral*, *Joy*, and *Surprise*, but most *Surprise*s of the other five roles tend to transfer to only *Surprise* and *Neutral*.

Moreover, to highlight the individual differences of emotion transitions among the six main roles in detail, we also show the standard deviations (Std) of each row in the emotion transition matrixes of the six main roles, as shown in Figure 4. The red bar chart shows the Std of the infinite norms of rows in the emotion transition matrix, which indicates the diversity of the most probable emotions from the same emotion in emotion transfers 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 emotion to other emotions.

Both charts show similar patterns of emotion transitions. *Anger*, *Surprise*, and *Disgust* vary the most in different roles, while people are more common when process *Neutral* and *Joy* emotions in

---

[6]Here, we analyze the personality-affected emotion transition based on roles rather than the numeric traits in Table 2 to avoid numeric observation errors.
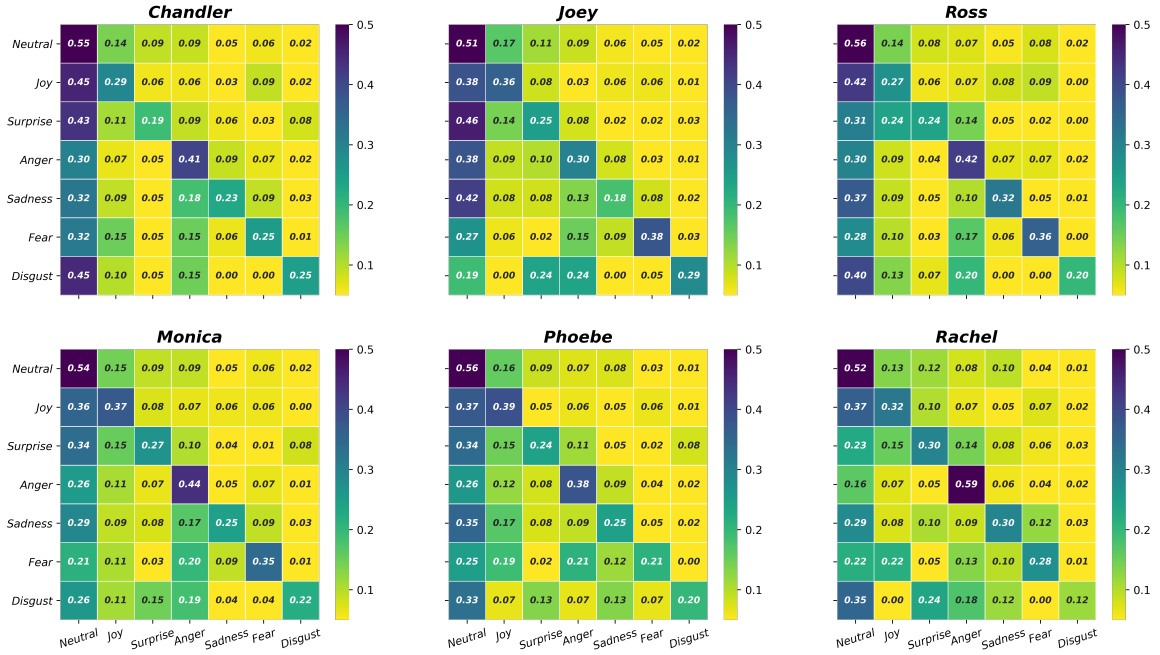
Figure 3: Emotion transition matrixes of the six main roles in PELD. Each row in a matrix shows the ratios of the current emotion E_i is transferred to the next emotion E_r.

conversation. Besides, negative emotions (*Anger*, *Sadness*, *Fear*, and *Disgust*) are relatively higher than positive emotions and *Neutral* on average. So, we can infer that the personality traits influence more in the emotion transfers from negative emotions.

5 Experiment

5.1 Evaluation Tasks

To validate the effectiveness of our proposed emotion generation model, we set two evaluation tasks: Emotion Prediction and Sentiment Prediction on PELD. Emotion Prediction requires the model to predict the emotion in the upcoming utterance based on the preceding dialog context in a dyadic conversation scenario; while Sentiment Prediction has the same setting except to predict the sentiment in the upcoming utterance.

For both tasks, we evaluate the prediction per-

Tri.Emos	Neutral	Joy	Surprise	Anger	Sadness	Fear	Disgust
E_i	2910	1242	597	751	438	457	115
E_r	2771	1123	634	858	493	487	144

Tri.Sentis	Neutral	Positive		Negative	
S_i	2910	1839		1761	
S_r	2771	1757		1982	

Table 5: Emotions in PELD Triples

formance by F-scores of single emotion or sentiment. Besides, the overall performance is also measured from two aspects with the macro averaged (**m-avg**) and the weighted averaged (**w-avg**) F-scores. A higher m-avg indicates the model performs relatively better predicting all categories, while a higher w-avg indicates the model predicts emotions or sentiments with larger proportions in the dataset better.

5.2 Ablation Study Setting

Although plenty methods (Majumder et al., 2019; Ghosal et al., 2020, 2019) has been proposed to analyze emotions in dialogues of *Friends*, most of their targets are to recognize the emotions of utterances in conversation. Compared with emotion recognition, the problem setting of selecting emotion is different and it is more difficult to select the appropriate emotion in response without knowing the response content. So, instead of comparing with other emotion recognition models, we turn to conduct ablation studies to evaluate the effectiveness of different parts of our model design. The ablation study compares the performances of the following models:

RoBERTa: RoBERTa (Liu et al., 2019) is a famous pre-trained language model designed for natural language understanding. Its performance

Methods	Anger	Disgust	Fear	Joy	Neutral	Sadness	Surprise	m-avg	w-avg
RoBERTa	0.218	0.000	0.107	0.214	0.453	0.122	0.126	0.177	0.287
RoBERTa-P	0.178	0.000	0.047	**0.265**	0.517	0.110	0.053	0.167	0.352
PET-VAD	0.190	**0.081**	0.115	0.188	0.474	0.000	**0.179**	0.175	0.309
PET-CLS	**0.320**	0.070	**0.140**	0.198	**0.528**	**0.155**	0.098	**0.203**	**0.424**

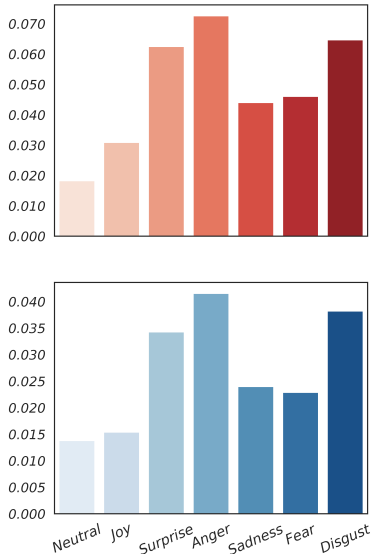Table 6: Results for Emotion Prediction.

Figure 4: The standard deviations of the infinity norm (red) and the L2-norm (blue) of each row in emotion transition matrixes of the six main roles in PELD.

is widely validated in many downstream tasks. We here use pre-trained RoBERTa, 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 emotion for response through a classification head.

RoBERTa-P: We concatenate the personality vector of the speakers with the dialog context representation by RoBERTa as the feature, then predict the response emotion. This method is to evaluate whether personality influences the expression of emotions.

PET-VAD: As emotions can be represented by both discrete category labels or vectors in the VAD spaces. PET-VAD is set to compare the different usages of emotion VAD vectors in our model. During training, PET-VAD regressions the VAD vectors of target emotions by minimizing the Mean

Squared Error (MSE) between generated vectors and the VAD vectors of ground truth emotions. The prediction output of PET-VAD is the closest neighbor emotions of generated VAD vectors measured by MSE.

PET-CLS: This is our method Personality-affected Emotion Transition with a classifier after obtaining the VAD vector of generated emotion. PET-CLS predicts emotions in the upcoming utterances as described in Section 3.

For RoBERTa, RoBERTa-P, and PET-CLS directly outputting discrete emotions, we adopt the Focal loss (Lin et al., 2017) to relieve the imbalanced emotion prediction.

6 Results and Analysis

In this section, we report and analyze the experimental results on the Test set of PELD in our ablation study. All results are chosen by the best performance on the Valid set within 50 epochs training.

6.1 Results for Emotion Prediction

The results on the Emotion Prediction task are reported in Table 6. First of all, as a seven-classes prediction task also suffered from the imbalance issue, the overall performance is moderately low, which also indicates the difficulty of the task. As for the averaged F-scores, PET-CLS improves both the w-avg and m-avg by a large margin from all other methods, which verifies our personality-affected emotion transition method.

In detail, all models perform better on emotions with larger portions (*Neutral* and *Joy*), as they are more probable to occur in the response emotion. Moreover, PET-VAD and PET-CLS achieve moderately higher F-score on the minority emotions (*Anger*, *Sadness*, *Disgust*, *Fear*, and *Surprise*), which shows that the emotion transition process is more important generating these minority emotions. It also verifies the finding in Section 4.2.

Methods	Negative	Neutral	Positive	m-avg	w-avg
RoBERTa	0.415	0.430	0.323	0.389	0.390
RoBERTa-P	0.401	**0.505**	0.176	0.361	0.430
PET-CLS	**0.492**	0.474	**0.327**	**0.431**	**0.445**

Table 7: Results for Sentiment Prediction.

On the other hand, although PET-VAD is based on the designed personality-affected emotion transition, most single emotion F-scores of PET-VAD are lower than RoBERTa or RoBERTa-P. We discuss the possible reasons as follows. One reason might be that the imbalance emotion issue cannot be alleviated in directly regression the emotion VAD vectors. Another reason might be that the value of emotion VAD vectors in Table 1 are estimated rather than precisely calculated, and the distance among different emotions in the theoretical VAD space is not similar to those in the emotion distribution in daily conversation.

6.2 Results for Sentiment Prediction

As predicting the emotions for the upcoming responses is difficult due to the multiple imbalanced categories, we also report the results on the Sentiment Prediction task in Table 7. Besides, different from the analysis above, which categorizes emotions by their portions in PELD, sentiment is another aspect of emotion analysis. As the sentiments are not directly described in the VAD spaces, we only report the results for RoBERTa, RoBERTa-P, and the PET-CLS. Besides, we only change the output size of PET-CLS from 7 (for emotions) to (3 for sentiments) and preserve the emotion transition process in this task.

In general, we can see that the prediction F-scores of sentiments are higher than emotion predictions. Besides, the prediction of negative emotions is much easier than predicting positive emotions in all three methods. It may because although the numbers of sentiments are similar, the categories of negative emotions (*Anger*, *Sadness*, *Fear*, and *Disgust*) are more than positive emotions (*Joy* and *Surprise*). Equipped with our model design, PET-CLS outperforms both RoBERTa and RoBERTa-P excepted for the neutral sentiment. It suggests that the personality-affected emotion transitions also facilitate sentiment prediction. However, only concatenating the personality vectors with context representation, RoBERTa-P improves the F-scores

of Neutral but decreases the Positive and Negative. Hence, direct concatenation limits the effect of personality information in sentiment prediction.

7 Conclusion and Future Work

In this work, we raise the problem of automatically selecting the emotion for response considering the individual differences in conversation and propose a new perspective to solve it through personality-affected emotion transition. Besides, we construct a dialog script dataset PELD with emotion and personality labels to facilitate related researches. We also validate our personality-affected emotion transition model in emotion prediction experiments.

Facial expressions, voices, gestures, and environment information are also vital in emotional interaction, but they are not captured in the 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.

7.1 Acknowledgement

References

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*, pages 150–165. Springer.

Nabiha Asghar, Pascal Poupart, Jesse Hoey, Xin Jiang, and Lili Mou. 2018. Affective neural response generation. In *European Conference on Information Retrieval*, pages 154–166. Springer.

Gene Ball. 2000. Emotion and personality in a conversational agent. *Embodied conversational agents*, pages 189–219.

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):335–359.

Kenneth Mark Colby. 1975. *Artificial paranoia: a computer simulation of paranoid process*. Pergamon Press.

Pierre Colombo, Wojciech Witon, Ashutosh Modi, James Kennedy, and Mubbasir Kapadia. 2019. Affect-driven dialog generation. *arXiv preprint arXiv:1904.02793*.

Paul T Costa and Robert R McCrae. 1992. Normal personality assessment in clinical practice: The neo personality inventory. *Psychological assessment*, 4(1):5.

Paul Ed Ekman and Richard J Davidson. 1994. *The nature of emotion: Fundamental questions.* Oxford University Press.

Deepanway Ghosal, Navonil Majumder, Alexander Gelbukh, Rada Mihalcea, and Soujanya Poria. 2020. Cosmic: Commonsense knowledge for emotion identification in conversations. *arXiv preprint arXiv:2010.02795*.

Deepanway Ghosal, Navonil Majumder, Soujanya Poria, Niyati Chhaya, and Alexander Gelbukh. 2019. Dialoguegcn: A graph convolutional neural network for emotion recognition in conversation. *arXiv preprint arXiv:1908.11540*.

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):1290–1303.

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*.

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*, page 10.

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*.

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*.

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*.

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*.

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

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)*, pages 121–132, Hong Kong, China. Association for Computational Linguistics.

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*.

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

Naoki Masuyama, Chu Kiong Loo, and Manjeevan Seera. 2018. Personality affected robotic emotional model with associative memory for human-robot interaction. *Neurocomputing*, 272:213–225.

John D Mayer. 2004. (2004) what is emotional intelligence?

Albert Mehrabian. 1996a. Analysis of the big-five personality factors in terms of the pad temperament model. *Australian journal of Psychology*, 48(2):86–92.

Albert Mehrabian. 1996b. Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in temperament. *Current Psychology*, 14(4):261–292.

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)*, pages 174–184.

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*.

James A Russell and Albert Mehrabian. 1977. Evidence for a three-factor theory of emotions. *Journal of research in Personality*, 11(3):273–294.

Roman Shantala, Gennadiy Kyselov, and Anna Kyselova. 2018. Neural dialogue system with emotion embeddings. In *2018 IEEE First International Conference on System Analysis & Intelligent Computing (SAIC)*, pages 1–4. IEEE.

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)*, pages 1–6. IEEE.

Mark A Thornton and Diana I Tamir. 2017. Mental models accurately predict emotion transitions. *Proceedings of the National Academy of Sciences*, 114(23):5982–5987.

Jason Wei and Kai Zou. 2019. Eda: Easy data augmentation techniques for boosting performance on text classification tasks. *arXiv preprint arXiv:1901.11196*.

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*, pages 1401–1410.

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*.

Rohola Zandie and Mohammad H Mahoor. 2020. Emptransfo: A multi-head transformer architecture for creating empathetic dialog systems. *arXiv preprint arXiv:2003.02958*.

Peixiang Zhong, Yan Zhu, Yong Liu, Chen Zhang, Hao Wang, Zaiqing Nie, and Chunyan Miao. 2020. Endowing empathetic conversational models with personas. *arXiv preprint arXiv:2004.12316*.

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*.

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):53–93.