Applied Intelligence (2022) 52:4744–4757
https://doi.org/10.1007/s10489-021-02660-4

Diverse dialogue generation by fusing mutual persona-aware and self-transferrer

Fuyong Xu[1] · Guangtao Xu[1] · Yuanying Wang[1] · Ru Wang[1] · Qi Ding[1] · Peiyu Liu[1] · Zhenfang Zhu[1]

Accepted: 4 July 2021 / Published online: 28 July 2021

Abstract
In recent years, personalized dialogue generation has attracted researchers' attention due to its wide range of applications. Although there has been much excellent research, current chatbots are known to generate uninteresting responses without personality. In a general way, persona information is adopted to mitigate these issues. Yet, how to integrate persona information reasonably based on historical dialogue in responses is still a severe challenge. Thus, we present a solution that consolidates the persona information memory process and pretrained method to generate personalized and fluency responses. Moreover, we adopt a reinforcement learning algorithm to joint each part of the dialogue model. Last but not least, we present a self-learning framework to explore the hypothetical space, and make the responses more personalized and fascinating. Extensive experiments on the large-scale dialogue public dataset ConvAI2 verify the effectiveness of our method.

Keywords Dialogue generation · Persona information · Self-learning · Reinforcement learning

1 Introduction

In the field of dialogue systems, [1] pointed out that chitchat models must not only have the ability to generate diverse responses but also establish an emotional state connection with interlocutors during the conversation. Each user has his or her own characteristics and habits. It is also very important to fully mine the user's persona information. Seq2Seq models [2–5] can fully integrate the context information of the dialogue and solve the problem that RNN s yield output with fixed data dimensions. This model can effectively improve the diversity of responses in the dialogue system. Although the Seq2Seq model has now been widely adopted in dialogue systems, there is still a long way to go before dialogue systems can understand natural human language and pass the Turing test [6]. There are still quite a few problems with adopting the Seq2Seq model to build dialogue language models; for example, the generated responses have a lower degree of personalization [7] and correlation with the dialogue history. The Seq2Seq neural network model applied for dialogue generation tends to generate safe and shared responses (e.g., "I don't know") [8]. [3] pointed out that the reason for the above problems is that the persona information associated with speakers was not combined into the dialogue generation process.

In some cases, the generated response does not need to reflect persona information, but it needs to combine the persona information properly on the basis of fully combining the historical dialog information. To improve the personality performance of dialogue models, we exploit the latent implicit personalized interactive information by a multihop attention mechanism for dialogue context and persona information. In this paper, we designed a persona information memory selection network (PMSN), which is trained by predefined persona information. When predicting the responses, we employ the PMSN to generate the most relevant persona information and input it into dialogue generation blocks to assist in the generation process. Figure 1 illustrates our method, the persona information W^A of speaker A is composed of L pieces of profile information $\{W_1^A, W_2^A \ldots W_L^A\}$. When we communicate with others, firstly we think about what kind of characters they are, and what kind of personality traits they have. These features will be generated by PMSN which is based on pre-defined persona information. The dialogue history $h_{n-1}^A = (x_1^A, x_1^B, x_2^A, x_2^B \ldots x_{n-1}^A, x_{n-1}^B)$ and the most correlative persona information will be adopted in the dialogue generation process. When we chat with others, not

✉ Peiyu Liu
liupy@sdnu.edu.cn

✉ Zhenfang Zhu
zhuzf@sdjtu.edu.cn

[1] Shandong Normal University, Shandong, 250300, China

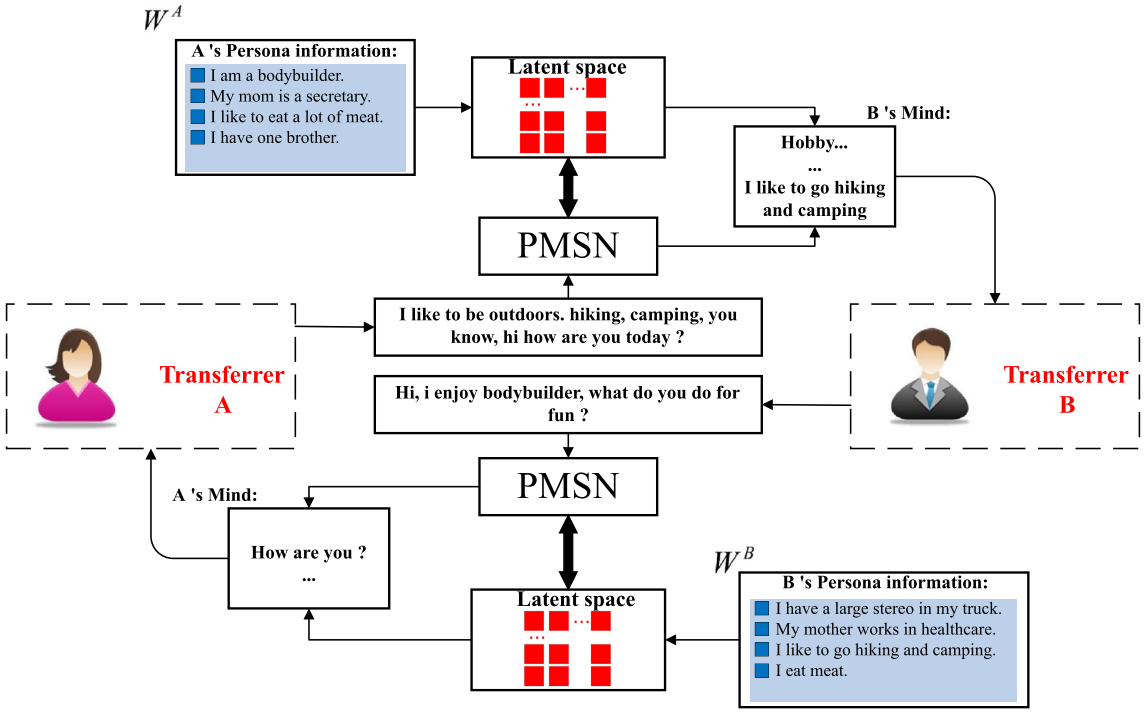


Fig. 1 Method Overview

all of the responses need to incorporate persona information, and appropriate integration of persona information is more suitable. Therefore, to improve the response diversity, we input the dialogue context into the PMSN to predict the most correlative persona information and then employ these selected persona features to assist in the dialogue generation process. The PMSN employs the multilayer perceptron (MLP) method to mine the historical persona information and adopts $W^* = MLP(W, x)$ to choose the most correlative persona information. The dialogue generation network Transferrer is a sequence prediction model based on dialogue history and the selected persona information to generate more personalized and diverse responses. Our basic dialogue model adopts the GPT-2 pretrained model because it involves a larger number of parameters and is trained by more data than the GPT model, which achieves an excellent performance in dialogue generation. Transferrer adopts a conditional probability $p(x_n^A \mid W^A, W^*, h_{n-1}^A)$ to predict the target sequence, where x_n^A denotes the target token. W^A means the persona information about speaker A, W^* denotes the most relativity persona information, and h_{n-1}^A is the $n-1$ rounds of historical dialogue.

To speed up the convergence of our model and improve the performance of the model, we integrate the reinforcement learning method based on a Markov decision process (MDP) into the learning process, apply fine-tuning and optimize the parameters of our model. Two Transferrers are initialized by allowing them to chat with each other. In addition, we design a reward mechanism that suits interlocutors' dialogue preferences to generate interesting, personalized and smooth dialogue responses. With the exploration of agents, the PMSN will be completed with persona information. A set of successful dialogs allows the two interlocutors to enhance their understanding of each other through the content of the dialogue and the characteristics of the interlocutors. In summary, our contribution is threefold:

- We propose a solution for personalized dialogue generation. Our method can generate abundant personalized and smooth responses from predefined persona information, and distinguish the most relevant profile features from noise data effectively.
- We design an optimized method named Self-Transferrer framework to optimize our personalized dialogue model. The Self-Transferrer optimize method joint each part of our model to generate more personalized and smooth responses.
- Extensive experiments on ConvAI2 conducted to validate the superiority of our model can generate a more personalized response than other popular baselines in terms of different metrics.

The remainder of our paper is organized as follows. In Section 2, we describe the related work of dialogue systems and personalized systems. Sections 3 and 4 introduce the PMSN and Transferrer architecture we employed in our personalized dialogue generation process separately. Besides, we present the Self-Transferrer framework to joint each part of the dialogue model to improve the

personality and fluency of responses in Section 5. Finally, Section 6 shows the experimental details, results, and further analysis to verify the effectiveness of our method. Section 7 concludes our work and suggests ways to improve dialogue quality in the future.

2 Related work

Due to the availability of various related large-scale datasets [9, 10], the task of personalized dialogue generation has made considerable progress [36]. Table 1 illustrates a round of conversations in the ConvAI2 dataset, and the persona information is expressed in a few descriptive sentences. Dialog systems based on generation methods employ an encoder-decoder architecture, which encodes the dialogue context and adopts a Seq2Seq model to predict the target responses. Generative-based dialogue systems still have certain problems, such as nonsmooth and repeated responses [11, 35, 37]. The responses generated by these dialogue systems are relatively fixed and lack flexibility, making them very difficult to apply in chatbots. According to research in cognitive science, effective communication creates similar activation maps in the brains of the two interlocutors [12], suggesting that understanding interlocutors' persona information and emotional states is an essential process for generating high-quality conversations.

In the field of personalized dialogue generation, the traditional method of constructing personalized dialogue systems focuses on psychological characteristics, such as the "Big Five" personality traits [13]. Modeling the psychological features and collecting the corresponding dialogue context is very difficult. These limitations have hindered the development of personalized dialogue systems. Recent studies have tried to build personalized dialogue generation models in a data-driven manner. [8] first incorporated persona information, and the persona information was converted into a dense vector for the subsequent dialogue generation task, which could effectively reduce the number of general responses and increase the variety of responses. Nevertheless, all of the above methods rely too much on labeled data, the training cost is high, and the features of the training data are sparse, so [11] proposed a new reinforcement learning process based on a Markov decision process, which effectively reduced the number of regular responses and effectively increased the variety of responses. [14] input persona information and historical dialogue into a neural network model and aimed to generate more meaningful responses.

Traditional datasets about personalized dialogue generation do little encourage the dialogue model to engage in understanding natural language and maintaining persona information. Inspired by this mind, [9] proposed a dialogue dataset based on persona information, and they further proposed two generative models: persona-Seq2Seq and a generative profile memory network, which were employed to incorporate persona information into responses. In the work of [15], the researchers pointed out that dialogue agents fail to engage users, especially when trained on an end-to-end dialogue system. For the above reasons, they introduced a new dataset providing more persona informa-

Table 1 A sample of the dataset, every sample includes persona information of each speaker and dialogue context

Persona A	Persona B
I like to ski	I am an artist
My wife does not like me anymore	I have four children
I have went to Mexico 4 times this year	I recently got a cat
I hate Mexican food	I enjoy walking for exercise
I like to eat cheetos	I love watching Game of Thrones
[Persona A:] Hi	
[Persona B:] Hello ! How are you today ?	
[Persona A:] I am good thank you, how are you.	
[Persona B:] Great, thanks ! My children and I were just about to watch Game of Thrones.	
[Persona A:] Nice ! How old are your children?	
[Persona B:] I have four that range in age from 10 to 21. You?	
[Persona A:] I do not have children at the moment.	
[Persona B:] That just means you get to keep all the popcorn for yourself.	
[Persona A:] And Cheetos at the moment!	
[Persona B:] Good choice. Do you watch Game of Thrones?	
[Persona A:] No, I do not have much time for TV.	
[Persona B:] I usually spend my time painting: but, I love the show.	

tion and dialogue context based on persona information. With the development of a pretrained model, [16] proposed a new approach named TransferTransfo that used the Transformer model to improve the fluency of responses. To fuse the target persona information into the decoding process effectively and balance its contribution, [17] proposed an attention routing structure, which can make more effective use of personalized sparse data in the process of model training. There are still many meaningful challenges in building a personalized dialogue system, for example, how to generate an informative response with multiple relevant personalities without losing fluency and coherence. To address this issue, [18] presented a model that incorporates recurrent personality interactions among response decoding steps to fuse appropriate persona information. Dialogue generation processes need to incorporate the interlocutors' persona information in most cases, whereas in some specific cases, it is not necessary to incorporate their persona information [19]. To incorporate more coherent personality information into the dialogue generation process, [20] predefined several profile key-value pairs, including name, gender, age, location, etc., and distinctly expressed a profile value in the response. Although the above methods have achieved positive results, most models focus too much on deliberately imitating human responses and generate responses that are excessively related to persona information. To address these issues, our work generates responses with the most relevant persona information without losing fluency and coherence. [21] proposed P2 BOT, which incorporates mutual persona perceptions to improve the quality of personalized dialogue generation, and there are also more open-domain dialogue systems with reinforcement learning methods [22–25]. Moreover, integrating the task of dialogue generation into task-oriented dialogue systems, which provides users with a more interesting experience while completing tasks, has also attracted the attention of researchers [26, 27].

In summary, researchers have performed many meaningful studies in the field of personalized dialogue generation. Nevertheless, many challenges still exist, such as (1) how to enhance the diversity expression of the dialogue generation process by mining the personalized expression of historical dialogue, (2) how to model the dialogue process with the development of a pretrained model, and (3) how to combine reinforcement learning methods to make the training process of data-driven dialogue systems more effective. The reason why cause these issues is mostly due to the personality characteristics of interlocutors are limited and there has no good way to explore the persona features from limited persona information and dialogue history context. To address these challenges, we design our model from the following three aspects. First, we exploit deep implicit personalized interaction information by the multihop attention mechanism, it can explore the relation between given profile and dialogue history. Second, we apply the advantages of the pretrained language model to enhance semantic representation based on the limited persona information. Third, we design a self-learning process based on a reinforcement learning method, which can help improve the learning efficiency of the dialogue system.

3 Persona information memory selection network (PMSN)

To better integrate persona information into the dialogue generation process, the persona information will be input to the PMSN for memorization before the dialogue starts. Memorizing persona information is a process of paying attention to the persona information. To reduce the error caused by the memorization process, the multihop attention method is adopted to calculate the attention of the character information. The calculation process is as follows:

First, inspired by the attention mechanism, our model calculates the attention score between d_t(Query) and each w_i(Persona) by $e_{ti} = d_t^T w_i$, and then employs the softmax function to normalize the attention score, as shown in formula 1.

$$a_{ti} = softmax\left(d_t^T w_i\right) \tag{1}$$

Second, we adopt an attention weight to measure the matching degree of the current dialogue context and persona information, and the attention weight a_{ti} and its corresponding w_i weighted sum are employed as the attention output of the $t-th$ dialogue sequence. The calculation formula is shown in formula 2. Each person's information has an output vector C_i (by another embedding matrix C).

$$Attention\left(h_t, \mathrm{W}, \mathrm{C}\right) = \sum_i a_{ti} C_i \tag{2}$$

The process of calculating the attention is essentially a weighted sum function. If it employs only one-hop attention, there will exist a certain error. Some features that are not highly relevant to the current context are discarded, and there is an error in this selection process. The calculated attention matrix cannot indicate the degree of association between the target sentence and the current context well.

Finally, inspired by [19, 38, 39], we adopt a multihop attention structure, where the attention output of the i hops is shown in formula 3.

$$m^i = m^{i-1} + Attention^{i-1} \tag{3}$$

Where m_0=d_t. Finally m_3 is employed as our memorized persona. In our experiments, $i = 3$ achieves a better performance than $i = 1$ or 2, whereas there is no crucial increase when i = 4, 5 or 6.

When selecting the persona representation related to the current case W^*, a linear transformation is applied to the output of the multihop attention to obtain the persona information, and the formula is as follows:

$$W^* = softmax\left(W_p\left[m^3\right]\right) = MLP\left(\left[m^3\right]\right)$$

Where W_p is the weighted matrix when selecting the most relevant persona information, and the selected persona information is adopted in the process of dialogue generation.

3.1 Training

It is necessary to label the dialogue context before training the PMSN, which adopts TFIDF to compute the relevance between the dialogue context and each person's persona information. The inverse document frequency is computed by formula 4:

$$idf_i = \frac{1}{(1 + \log(1 + tf_i))} \tag{4}$$

Where tf_i is the index of Glove vocabulary, W^* is the predicted persona information that is most relevant to the current context, p_i is the labeled persona,and the loss function is as follows:

$$\mathcal{L}_{pmsn} = \frac{1}{N}\sum_{i=1}^{N}\left|W_i^* - p_i\right| \tag{5}$$

4 Transferrer

According to the research of [9, 11, 21], we treat the task of dialogue generation as a sequence prediction process. Our method adopts the l GPT2 language model [28] to initialize our dialogue model. Compared with GPT [29] models, GPT2 increases the scale of training data and enriches the content of the models. They are all based on the Transformer model. When training the GPT2 model, the effectiveness of unsupervised learning is also verified.

The complete dialogue generation process is shown in Fig. 2. The Transferrer adopts the decoder structure in the Transformer model [30], processes dialogue-related context information and generates responses. The dialogue-related context information includes persona information W^A, historical dialogue information h_n^A, and persona information w^{A^*} with the highest degree of correlation with the current context. We employ the MLE to predict the next token of the response sequence, and the loss function is given in formula 6.

$$\mathcal{L}_{mle} = \sum_t \log p_\theta\left(x_{n,t}^A \mid W^A, w^{A^*}, h_n^A, x_{n,<t}^A\right) \tag{6}$$

Where θ is the parameter of Transferrer, $x_{n,t}^A$ means the t-th token in x_n^A, and $x_{n<t}^A$ indicates the token before t-th token. Formula 6 applies to both A and B, and we mention A for the sake of brevity (B is the same as below).

During the prediction process, beam search is applied to store the top-ranked response candidates $\{\hat{x}_n^{A^*}\}$, and the Transferrer further chooses the candidate that maximizes the length-normalized score as the prediction as follows:

$$\hat{x}_n^{A^*} = argmax\frac{\log p_\theta(\hat{x}_n^{A^*} \mid W^A, w^{A^*}, h_n^A)}{|\hat{x}_n^{A^*}|} \tag{7}$$

To improve the generalizability of the model and find a more powerful and robust feature representation that will benefit the process of dialogue generation, inspired

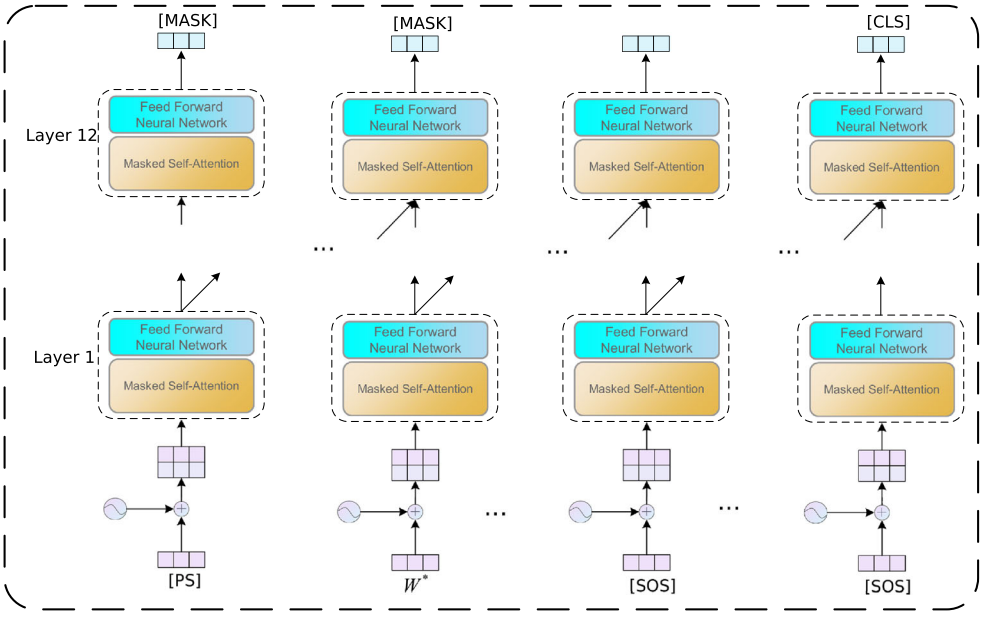


Fig. 2 Transferrer architecture, every block is decoder block in transformer

by [21], we set an auxiliary task to optimize the dialogue generation model: next utterance classification. In addition to generating more appropriate responses, a [CLS] token is added at the end of the generated sequence, and a classifier is added to the last layer of the model to determine whether the responses generated by the system are appropriate responses. The method of classification randomly selects interference item data and trains the classifier to distinguish between normal replies and interference items, and formula 8 is extended as follows:

$$x_n^{\mathcal{A}^*} = \arg\max_{\hat{x}_n^A}(\alpha \cdot \frac{\log p_\theta(\hat{x}_n^{\mathcal{A}} \mid w^{\mathcal{A}}, h_n^{\mathcal{A}})}{|\hat{x}_n^{\mathcal{A}}|} + (1-\alpha) \cdot \log p_\theta(y_n = 1 \mid w^{\mathcal{A}}, h_n^{\mathcal{A}}, \hat{x}_n^{\mathcal{A}})) \quad (8)$$

Where θ is the shared parameter of the dialogue generation task and the auxiliary task, $y_n = 1$ indicates that x_n^A is predicted as the next personalized utterance, and α is a hyper-parameter.

5 Self-transferrer & fine-tuning

Although the supervised dialogue generation model can imitate a speaker's personalized responses well based on the training data, it cannot fully allow the machine to fully understand natural language. Therefore, we try to match the two speakers randomly and encourage the Transferrer to learn a policy that could yield the maximum reward. A dialogue is carried out, and we encourage the Transferrer to learn a strategy that can obtain the greatest reward through reinforcement learning. We further optimize the model by fine-tuning, which employs reinforcement learning to maximize the reward function. We apply self-play to simulate the interaction between two Transferrers. Transferrers adopt the historical dialogue context and the persona information of the interlocutors for complete exploration. The exploration is shown in Fig. 3, and the details are explained below.

According to the work of [21], we divide the two conversational individuals into users and agents. The process of self-play is the process by which the agent optimizes parameters θ. User $\mathcal{A}$ starts the conversation, and $\mathcal{B}$ will reply as the agent. Inspired by the work of [11, 21], we introduce some necessary formulations for modeling our problem with reinforcement learning. A policy defines the behavior of the learnable agent at a specific time and computes the conditional probability of a certain action taken in a certain state, and the formula is expressed as $p_\theta = (a_n^B \mid s_n^B)$. The policy is responsible for mapping a state to an action. The reward defines the temporary income of the agent. After the agent takes a certain action, the environment sends a reward to the agent at each time step. The goal of the value function is to judge which action is better from a long-term perspective, indicating the long-term expectations after taking a certain action. A state contains the overall persona information of the interlocutors, the persona information most relevant to the current context, and the dialog history. Here, we define the state as a triple $s = (W, h, W^*)$ such that the state of agent $\mathcal{B}$ in the n-th round is expressed as $s_n^{\mathcal{B}} = (W^{\mathcal{B}}, h_n^{\mathcal{B}}, W^*)$. An action is taken by the agent according to a policy. In our

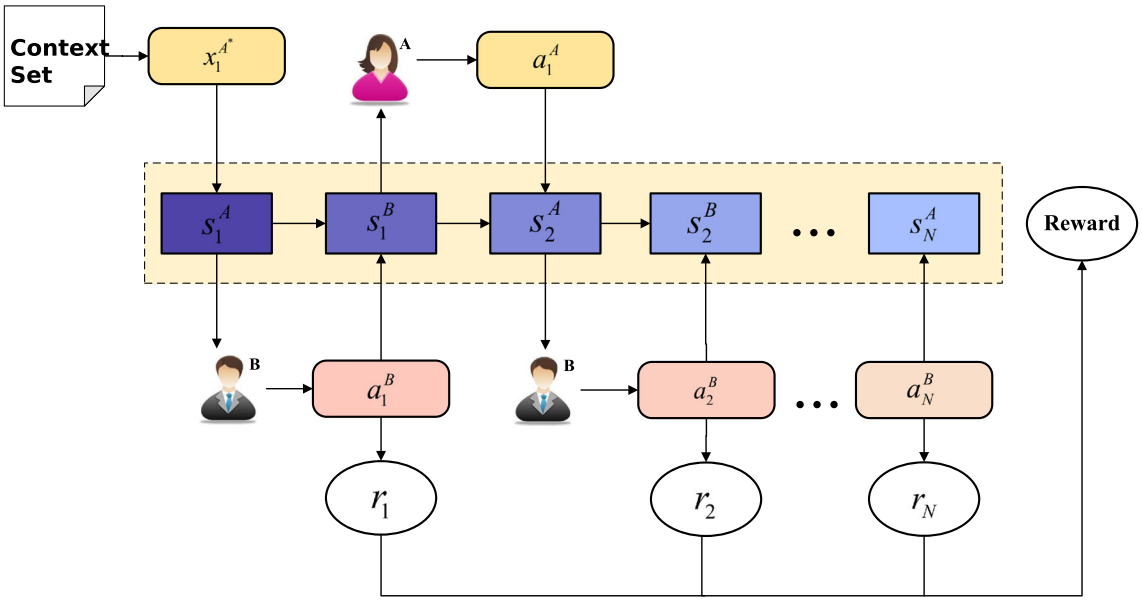


Fig. 3 The exploring process. Agent A select context from context database to start conversation as it is difficult to generate high quality sequence without dialogue history

personalized dialogue generation task, we regard the action as the response of agent $\mathcal{B}$ to the question of user $\mathcal{A}$, and the action is defined as $a_n^{\mathcal{B}}$. The agent learns from the dialogue history and chooses the best answer from $a_{t+1}^1 \ldots \cdot a_{t+1}^K$ for each time step $t = 1, ..., T$. After the agent takes an action, the agent receives a reward from the environment as the hidden state h_{t+1} for the next time step. Then, the agent sets a new action set $a_{t+1}^1 \ldots \cdot a_{t+1}^K$ and chooses the proper answer according to the state and policy. For user $\mathcal{A}$, the state is updated when receiving $\mathcal{B}$'s response, and the Transferrer method is employed to generate a response. We adopt a gradient policy [29] of the Transferrer, and it can output the policy function directly.

We define the sequence of exploratory processes as $\tau = \{s_1, a_1, s_2, a_2 \ldots s_T, a_T\}$, where $s_1 = (W^{\mathcal{A}}, h_0^{\mathcal{A}}, W^*)$, $s_2 = (W^{\mathcal{B}}, h_1^{\mathcal{B}}, W^*)$, and user $\mathcal{A}$ and agent $\mathcal{B}$ alternately interact with the environment to update their states. According to the Markov decision process, the probability of the occurrence of a certain sequence τ_i is shown in formula 9.

$$p_\theta(\tau_i) = p(s_1) \prod_{t=1}^{T} p_\theta(a_t \mid s_t)\, p(s_{t+1} \mid s_t, a_t) \tag{9}$$

This formula has an expected value of reward for each episode sequence and estimates the expected value of the return for each episode sequence by the action-value function, as shown in formula 10.

$$\bar{R}_\theta = E_{\tau \sim p_\theta(\tau)}[R(\tau)] \tag{10}$$

Where $R(\tau)$ is the true reward during exploration and the target is the maximum expected value of the reward. We employ the gradient policy to optimize the parameters in the next section.

5.1 Policy gradient

To obtain the maximum expected value of the reward, the likelihood ratio trick is adopted to update the parameter θ through a gradient ascent method, where the gradient of the expected value is shown in formula 11.

$$\nabla \bar{R}_\theta = E_{\tau \sim p_\theta(\tau)}[R(\tau) \nabla \log p_\theta(\tau)] \tag{11}$$

where θ is the parameter and the update method of θ is shown in formula 12.

$$\theta \leftarrow \theta + \eta \nabla \bar{R}_\theta \tag{12}$$

As mentioned above, the action space is infinite. The reinforcement algorithm is adopted to approximate formula 11 by sampling an action from the policy distribution. Furthermore, by subtracting a baseline, [31], applied the mean reward of a mini-batch as the reward to reduce the variance. The agent samples the tokens one by one through multinomial sampling over the output distribution. Multinomial sampling provides more diversity than beam search sampling.

5.2 Reawrd shaping

According to the work of [21], high-quality personalized dialogue generation models should focus on modeling human language and mutual persona perception between interlocutors. In the process of reinforcement learning, when the reward in the environment is too sparse, it may be slow to find a solution to the problem by simply relying on the agent to explore and learn, so experience can be integrated into the reward design process to be more effective in solving the problem and speeding up convergence. We design two reward processing schemes as follows.

5.2.1 RS.1

In the task of personalized dialogue generation, the responses generated by the system must conform to human language characteristics and dialogue rules so that the generated responses will be meaningful. According to the work of [21], such rules can be measured by pretrained models. Therefore, we employ reward shaping based on the pretrained model (GPT2). The reward for the actions taken by learner $\mathcal{B}$ in sequence τ is

$$R_1\left(a_n^{\mathcal{B}}\right) = \frac{1}{\left|a_n^{\mathcal{B}}\right|} \sum_t \log p_\theta\left(a_{n,t}^{\mathcal{B}} \mid a_{n,<t}^{\mathcal{B}}\right) \tag{13}$$

5.2.2 RS.2

The language is evaluated separately, and the coherence of the dialogue context is not fully considered. Therefore, a reasonable dialogue generation model should fully integrate the historical dialogue information to generate more meaningful responses, and we employed the auxiliary task in the previous article to design the reward, as shown in formula 14.

$$R_2\left(a_n^{\mathcal{B}}\right) = \log p_\theta\left(y_n = 1 \mid a_n^{\mathcal{B}}, s_n^{\mathcal{B}}\right) \tag{14}$$

It is safe to assume that human responses are always more natural and personalized than dialogue agents. $y_n = 1$ is the signal indicating that the generated responses $a_n^{\mathcal{B}}$ are predicted to be the next personalized utterance. In summary, the final reward is as follows:

$$r = \beta_1 R_1 + \beta_2 R_2 \tag{15}$$

Where β_1 and β_2 are hyper-parameters.

Algorithm 1 Self-transferrer framework with policy gradient.

Initial Agent $\mathcal{A}$ and $\mathcal{B}$
for each sample in datasets **do**
  Agent $\mathcal{A}$ select dialogue context $x_1^{A^*}$ from context database
  $s_1^{\mathcal{B}} = (W^{\mathcal{B}}, h_1^{\mathcal{B}}, W^*)$
  for each $i \in [1, T]$ **do**
    $a_i^{\mathcal{B}}, s_i^{\mathcal{B}} = Transferrer(s_i^{\mathcal{A}})$
    $a_i^{\mathcal{A}}, s_i^{\mathcal{A}} = Transferrer(s_i^{\mathcal{B}})$
    $r = \beta_1 R_1 + \beta_2 R_2$
  end for
  $\tau = \{s_1, a_1, s_2, a_2 \ldots s_T, a_T\}$
  $\nabla \bar{R}_\theta = E_{\tau \sim p_\theta(\tau)}[R(\tau) \nabla \log p_\theta(\tau)]$
  $\theta \leftarrow \theta + \eta \nabla \bar{R}_\theta$
end for

6 Experiment

6.1 ConvAI2 dataset and preparation

Our experiment is based on the large-scale ConvAI2 dataset, which incorporates interlocutor persona information, and a new test set is added to the PERSONA-CHAT dataset proposed by [9]. The dialogues are randomly paired and given persona information selected from the persona information pool. The training set contains more than 10,000 sets of multiround dialogues, including approximately 160,000 sentences. Each set of multiround dialogues contains at least five sentences describing the speaker's profile information.

6.2 Baselines

The baselines for comparison are divided into three categories: dialogue generation models without persona information, dialogue generation models based on persona information, and dialogue generation models based on a pretrained model.

STSA [4]: In the architecture of the Seq2Seq model with an attention mechanism, the encoder is responsible for encoding the dialogue context and calculating the semantic vector c_t of each time step, and the decoder is responsible for linearly transforming the generated semantic vector and generating a response. This method does not consider that persona information plays a significant role in the dialogue generation process.

Per-STSA [9]: On the basis of integrating the attention mechanism, persona information is integrated into the process of dialogue generation and effectively improves the diversity of the response; i.e., $x = \forall p \in P \| x$, where $\|$ denotes concatenation.

Dia-CVAE [32]: This dialogue model without profile information adopts a hidden variable to obtain potential features in the dialogue generation process and aims at increasing the diversity of the generated responses.

Per-CVAE [19]: This method employs a memory-augmented architecture to exploit persona information from context and incorporates a conditional variational autoencoder model together to generate diverse and sustainable conversations.

TransferTransfo [16]: This model combines transfer learning and the Transformer model and fine-tunes the pretrained model by optimizing the multitask objective function.

Transformer MemNet [10]: This dialogue generation process employs two trained models, namely, knowledge selection and dialogue prediction.

KIC [33]: This method is combined with knowledge-aware pointer networks and a recurrent knowledge interaction hybrid generator.

P2 BOT [21]: This dialogue system incorporates mutual persona perception and reinforcement learning methods to improve the personality of the generated response.

6.3 Experimental settings

In our experiments, the RNN is a two-layer GRU with a 768-dimensional hidden state, and the dimension of word embedding is set to 768. The vocabulary size is limited to 50,256. The mini-batch size is 16, and the Adam optimizer is adopted with an initial learning rate of 0.001. All the parameters are initialized by sampling from a uniform distribution. For the PMSN, the hidden size is 768, the number of epochs is 50, and the maximum length of the persona sequence is 15. For Transferrer, the size of the hidden state is 512, the batch size is 256, the beam size is 3, the maximum length is 256, the position embedding size is 512, 100 epochs are employed, and the learning rate is set to 6.25e-5. From the reinforcement learning process, the learning rate is 0.5, $\beta_1 = 0.4$, $\beta_2 = 0.6$, and $\alpha = 0.1$.

6.4 Automatic evaluation

Evaluation is an important task when building an open-domain dialogue system [34]. Automatically evaluating an open-domain dialogue generation model is still a challenging task. Inspired by [9], we employ official automatic metrics to evaluate our model:

Table 2 Experiment about diversity

Method	N=1			N=5			N=10		
	Dtinct-1	Dtinct-2	P.Cover	Dtinct-1	Dtinct-2	P.Cover	Dtinct-1	Dtinct-2	P.Cover
STSA	.0126	.0464	.0026	.0031	.0142	.0057	.0018	.0089	.0071
Per-STSA	.0157	.0745	.0091	.0036	.0213	.0217	.0021	.0139	.0193
Dia-CVAE	.0366	.2080	.0021	.0090	.0875	.0029	.0050	.0663	.0048
Per-CVAE	.0384	.0738	.0167	.0120	.1240	.0410	.0075	.0779	.0395
Transfertransfo	.0410	.0524	.0124	.0315	.0320	.0119	.0230	.0251	.0120
Transformer MemNet	.0393	.0471	.0120	.0301	.0317	.0121	.0219	.0225	.0103
KIC	.0407	.0476	.0122	.0326	.0351	.0121	.0247	.0251	.0120
P2 BOT	.0381	.3953	.0116	.0320	.0361	.0124	.0290	.0312	.0117
Our(PMSN+GPT2+RL)	**.0453**	**.0750**	.0109	**.0423**	**.0801**	**.0480**	**.0367**	**.0720**	**.0402**

N means the number of generated responses in each turn

Bold entries denote the cases in which our method is better than compared baselines

PPL, F1: PPL(Perplexity), The basic idea is to assign a higher probability value to the sentences in the test set and adopt the perplexity to measure the fluency of the dialogue and the intelligibility of the generated response. The lower the perplexity is, means the generated responses more fit the sentence of the test set, and the smoother and easier it is to understand the generated response. The definition of PPL is shown as below:

$$\mathrm{PPL}(S) = P\left(W_1, W_2, W_3 \ldots W_N\right)^{-\frac{1}{N}} \tag{16}$$

The F1 score reflects the precision and recall. We adopt the following metrics to prove the effectiveness of our method.

Diff-k-grams: Based on the idea of [19], Distinct-K is adopted to measure the diversity of the generated response. The idea of this metric is to calculate the number of different k-grams in the generated responses.

Persona Coverage(P.Cover): Inspired by [19], we adopt this metric to measure the coverage of persona information in the generated responses. Suppose that there are M pieces of predefined interlocutor persona information $\{p_1, p_2 \ldots p_M, \}$, the generated responses are $\left\{\hat{y}_1, \hat{y}_2 \ldots \hat{y}_N\right\}$, and the definition of P.Cover is as follows:

$$\mathcal{C}_{\mathrm{per}} = \frac{\sum_i^N \max_{j\in[1,M]} \mathcal{S}\left(\hat{y}_i, p_j\right)}{N} \mathcal{S}\left(\hat{y}_i, p_j\right) = \frac{\sum_{w_k \in W}\left(f_k\right)}{|W|} \tag{17}$$

where W is the shared word set of $\hat{y}_i$and p_j, and$|W|$means the length of the shared word list. N means the number of generated responses in each turn. The method to compute the f_k is $f_k = \frac{1}{(1+\log(1+tf_i))}$.

The final experimental score is obtained by averaging the model on the test set, and the results are shown in Table 2. When N=1, the diversity of our model is higher than those of the baselines, and the P.Cover of the generated response

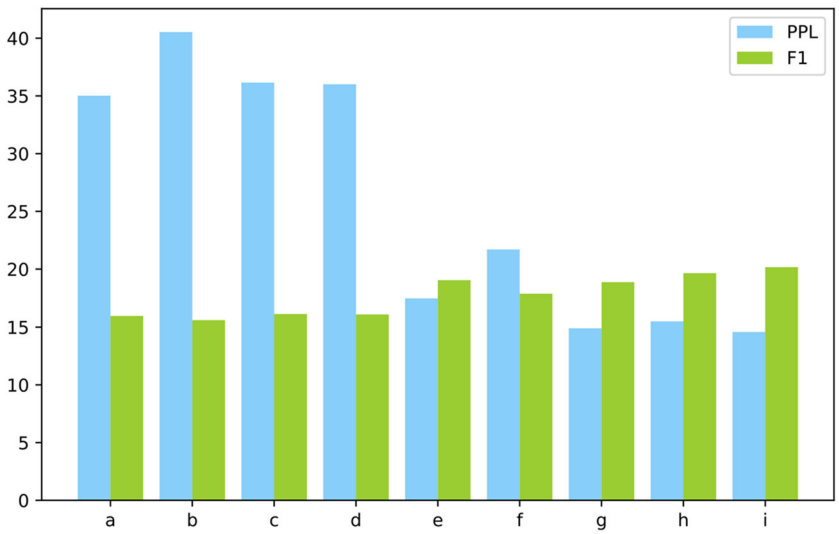


Fig. 4 Experiment about the dialogue quality. **a** STSA, **b** Per-STSA, **c** Dia-CVAE, **d** Per-CVAE, **e** TransferTransfo, **f** Transformer MemNet, **g** KIC, **h** P2 BOT, **i** Our(PMSN+GPT2+RL)

Table 3 Ablation tests

Variant	Dtinct-1	Dtinct-2	P.Cover
Transferso-Pre-Transformer	.0297	.0331	.0125
↪ +PMSN	.0314	.0452	.0259
↪ +RS.1	.0354	.0569	.0412
↪ +RS.2	.0423	.0801	.0480

is slightly lower than that of Per-CVAE, TransferTransfo and P2 BOT. When N=5 or 10, since we employ the GPT2 pretrained model and the reinforcement learning method in the dialogue generation process, the performance of our model is better than the other baselines in PPL and F1 as shown in Fig. 4. Our method outperforms almost all the baselines on personalized metrics due to the Self-Transferrer framework focus on generating more diversity and fluency dialogues. As the dialogue continues, the amount of information involved in the dialogue increases, and the P.Cover shows a downward trend due to the limited amount of persona information of the interlocutors that usually contains 5 pieces of persona information for each speaker.

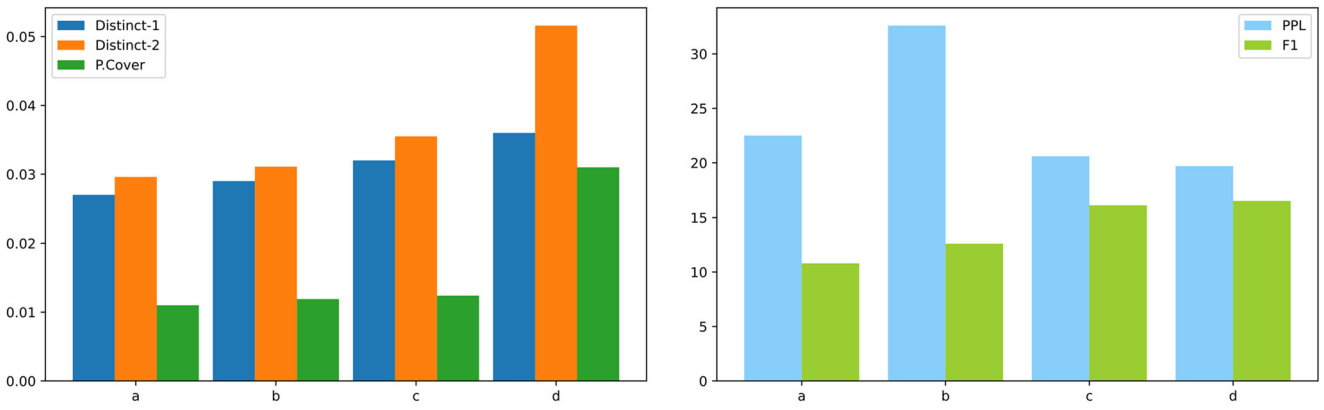
6.5 Additional analysis of the pretrained model

After employing the PMSN, the diversity of the generated responses shows a certain improvement. The purpose of adopting RS.1 is to make the predicted responses conform to human dialogue characteristics and dialogue rules. Table 3 shows that the predicted response complexity has been improved to a certain extent, the response complexity has been reduced, and the language can be explained clearly. The purpose of adopting RS.2 is to allow the reply to fully integrate the historical information of the dialogue so that it contains richer content. As shown in Table 3, after the introduction of RS.2, the persona information contained in the reply becomes richer. We compare our method with a previous method that deletes the pretrained model, and the result is shown in Fig. 5. Due to the RS.2 and the PMSN, the personalized metrics achieve significant improvement. After we remove the pre-trained model, our method is also slightly better than the other methods in PPL and F1. These improvements benefit from the multihop attention mechanism and the exploration of the RL algorithm. It also can prove the effectiveness of pre-trained model GPT-2 from Fig. 5.

6.6 Human evaluation

The automatic evaluation metrics show that our method can effectively combine persona information to generate a variety of interesting responses. We employ human evaluation to better evaluate the diversity of the responses generated by our method. The following results are based on N =5.

In the human evaluation process, we randomly selected 10 workers to test our dialogue model, these workers with high-level language skills and know nothing about our methods. We randomly sample 200 profile-question-response pairs from the test set, and the repeated pairs were filtered out. These workers chat with different chatbots follow by given persona information, and score the quality of generated responses and employ the average score as the last result. In our human evaluation, 1 means only fluent in terms of grammar and vocabulary, 2 means the responses



(a) (b)

Fig. 5 Our method compares with previdous methods without pre-trained model. **a** TransferTransfo, **b** Transformer-MemNet, **c** P2 BOT, **d** Our(PMSN+RL)

Table 4 Result of human evaluation

Model	1(%)	2(%)	3(%)	4(%)
TransferTransfo	**43.6**	41.2	11.7	3.5
P2 BOT	22.6	28.4	**41.3**	7.7
Our	24.9	20.1	31.7	**23.3**

generated by the system are related to the given persona information, 3 means the reply contains comprehensive and diverse information about the interlocutors, 4 means the responses are consistent with the given persona information. The results are shown in Table 4. From the table, our model can generate more personalized and consistent responses. Finally, we provide a dialogue example, as shown in Table 5, to prove more directly the superiority of our method compared with other different methods in the personalized dialogue generation process.

7 Conclusion and future work

The BOT-PMSN method we proposed is based on the persona information of the speaker, adopts the Self-Transferrer framework and persona information to assist in dialogue generation, and then introduces a reward signal in the dialogue process. The signal enhances the persona perception between humans and machines and realizes the task of personalized dialogue generation. The dialogue generation model that we trained can effectively understand natural language. Experiments on the large-scale dialogue public dataset ConvAI2 verify the effectiveness of our method. In future work, we will consider sustainability in the dialogue generation process and mine the rich expressions between persona and dialogue context. In the task of dialogue generation, meaningful dialogue also needs to fully consider the transfer of the speaker's emotional states. Therefore, emotional analysis will be added to subsequent work to increase the diversity of the generated responses.

References

1. Shum H-Y, He X, Li D (2018) From eliza to xiaoice: challenges and opportunities with social chatbots. Front Inf Technol Electron Eng 19(1):10–26
2. Sutskever I, Vinyals O, Le QV (2014) Sequence to sequence learning with neural networks. In: Ghahramani Z, Welling M, Cortes C, Lawrence ND, Weinberger KQ (eds) Advances in Neural Information Processing Systems 27: Annual Conference on Neural Information Processing Systems 2014, Montreal, pp 3104–3112
3. Vinyals O, Le QV (2015) A neural conversational model. CoRR arXiv:1506.05869
4. Shang L, Lu Z, Li H (2015) Neural responding machine for short-text conversation. In: Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers). Association for Computational Linguistics, pp 1577–1586
5. Serban IV, Sordoni A, Bengio Y, Courville AC, Pineau J (2016) Building end-to-end dialogue systems using generative hierarchical neural network models. In: Schuurmans D, Wellman MP (eds) Proceedings of the thirtieth AAAI conference on artificial intelligence. AAAI Press, Phoenix, pp 3776–3784
6. Turing AM (1957) Computing machines and intelligence. Mind 59:236–241
7. Jiang S, de Rijke M (2018) Why are sequence-to-sequence models so dull? understanding the low-diversity problem of chatbots. In: Chuklin A, Dalton J, Kiseleva J, Borisov A, Burtsev MS (eds) Proceedings of the 2nd International Workshop on Search-Oriented Conversational AI, SCAI@EMNLP 2018. Association for Computational Linguistics, Brussels, pp 81–86

Table 5 Sampled dialogues from different models

Persona	i love film
	i have a dog named pedro
	i am five feet tall
	i like to eat muffins
Context	what's your hobby, do you like play sports?
STSA	I don't know
Per-STSA	I love file
Per-CVAE	I love file and i do not like play sports
TransferTransfo	I do not have time for playing sports, do you like to listen to music
P2 BOT	I do not like play sports, and i like watching movies
Our	My hobby is watching movies at home, and i do not like play sports outside

8. Li J, Galley M, Brockett C, Gao J, Dolan B (2016) A diversity-promoting objective function for neural conversation models. In: Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Association for Computational Linguistics, pp 110–119
9. Zhang S, Dinan E, Urbanek J, Szlam A, Kiela D, Weston J (2018) Personalizing dialogue agents: I have a dog, do you have pets too? In: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Association for Computational Linguistics, pp 2204–2213
10. Dinan E, Roller S, Shuster K, Fan A, Auli M, Weston J (2019) Wizard of wikipedia: Knowledge-powered conversational agents. In: 7th International Conference on Learning Representations, ICLR 2019. OpenReview.net, New Orleans. https://openreview.net/forum?id=r1l73iRqKm
11. Li J, Monroe W, Ritter A, Jurafsky D, Galley M, Gao J (November 2016) Deep reinforcement learning for dialogue generation. In: Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, Austin, pp 1192–1202
12. Hasson U, Ghazanfar AA, Galantucci B, Garrod S, Keysers C (2012) Brain-to-brain coupling: a mechanism for creating and sharing a social world. Trends Cogn Ences 16(2):114–121
13. Mairesse F, Walker M (2007) PERSONAGE: Personality generation for dialogue. In: Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics. Association for Computational Linguistics, Prague, pp 496–503
14. Kottur S, Wang X, Carvalho V (2017) Exploring personalized neural conversational models. In: Sierra C (ed) Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI 2017. ijcai.org, Melbourne, pp 3728–3734
15. Mazaré P-E, Humeau S, Raison M, Bordes A (2018) Training millions of personalized dialogue agents. In: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, Brussels, pp 2775–2779
16. Wolf T, Sanh V, Chaumond J, Delangue C (2019) Transfertransfo: A transfer learning approach for neural network based conversational agents. CoRR arXiv:1901.08149
17. Zheng Y, Zhang R, Huang M, Mao X (2020) A Pre-Training Based Personalized Dialogue Generation Model with Persona-Sparse Data. Proceedings of the AAAI Conference on Artificial Intelligence 34:9693–9700. https://doi.org/10.1609/aaai.v34i05.6518
18. Lin X, Jian W, He J, Wang T, Chu W (2020) Generating informative conversational response using recurrent knowledge-interaction and knowledge-copy. In: Jurafsky D, Chai J, Schluter N, Tetreault JR (eds) Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, ACL 2020, Online. Association for Computational Linguistics, pp 41–52. https://doi.org/10.18653/v1/2020.acl-main.6
19. Song H, Zhang W, Cui Y, Wang D, Liu T (2019) Exploiting persona information for diverse generation of conversational responses. In: Kraus S (ed) Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI 2019. ijcai.org, Macao, pp 5190–5196
20. Zheng Y, Zhang R, Huang M, Mao X (2020) A pre-training based personalized dialogue generation model with persona-sparse data. In: The Thirty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2020, The Thirty-Second Innovative Applications of Artificial Intelligence Conference, IAAI 2020, The Tenth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2020. AAAI Press, New York, pp 9693–9700
21. Liu Q, Chen Y, Chen B, Lou J-G, Chen Z, Zhou B, Zhang D (2020) You impress me: Dialogue generation via mutual persona perception. In: Jurafsky D, Chai J, Schluter N, Tetreault JR (eds) Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, ACL 2020, Online. Association for Computational Linguistics, pp 1417–1427
22. Saha T, Chopra S, Saha S, Bhattacharyya P (2020) Reinforcement learning based personalized neural dialogue generation. In: Yang H, Pasupa K, Leung AC-S, Kwok JT, Chan JH, King I (eds) Neural information processing - 27th international conference, ICONIP 2020, Proceedings, part IV, Communications in Computer and Information Science, vol 1332. Springer, Bangkok, pp 709–716
23. den Hengst F, Hoogendoorn M, van Harmelen F, Bosman J (2019) Reinforcement learning for personalized dialogue management. In: Barnaghi PM, Gottlob G, Manolopoulos Y, Tzouramanis T, Vakali A (eds) 2019 IEEE/WIC/ACM International Conference on Web Intelligence, WI 2019. ACM, Thessaloniki, pp 59–67. https://doi.org/10.1145/3350546.3352501
24. He W, Sun Y, Yang M, Ji F, Li C, Xu R (2021) Multi-goal multi-agent learning for task-oriented dialogue with bidirectional teacher-student learning. Knowl Based Syst 213:106667. https://doi.org/10.1016/j.knosys.2020.106667
25. Saha T, Gupta D, Saha S, Bhattacharyya P (2020) Towards integrated dialogue policy learning for multiple domains and intents using hierarchical deep reinforcement learning. Expert Syst Appl 162:113650. https://doi.org/10.1016/j.eswa.2020.113650
26. Luo L, Huang W, Zeng Q, Nie Z, Sun X (2019) Learning personalized end-to-end goal-oriented dialog. In: The Thirty-Third AAAI Conference on Artificial Intelligence, AAAI 2019, The Thirty-First Innovative Applications of Artificial Intelligence Conference, IAAI 2019, The Ninth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2019. AAAI Press, Honolulu, pp 6794–6801
27. Zhang B, Xu X, Li X, Ye Y, Chen X, Wang Z (2020) A memory network based end-to-end personalized task-oriented dialogue generation. Knowl Based Syst 207:106398. https://doi.org/10.1016/j.knosys.2020.106398
28. Radford A, Wu J, Child R, Luan D, Amodei D, Sutskever I (2019) Language models are unsupervised multitask learners
29. Radford A, Narasimhan K, Salimans T, Sutskever I (2018) Improving language understanding by generative pre-training
30. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser L, Polosukhin I (2017) Attention is all you need. In: Guyon I, von Luxburg U, Bengio S, Wallach HM, Fergus R, Vishwanathan SVN, Garnett R (eds) Advances in neural information processing systems 30: Annual conference on neural information processing systems 2017, Long Beach, pp 5998–6008
31. Weaver L, Tao N (2001) The optimal reward baseline for gradient-based reinforcement learning. In: Breese JS, Koller D (eds) UAI '01: Proceedings of the 17th conference in uncertainty in artificial intelligence, university of washington. Morgan Kaufmann, Seattle, pp 538–545
32. Zhao T, Zhao R, Eskénazi M (2017) Learning discourse-level diversity for neural dialog models using conditional variational autoencoders. In: Barzilay R, Kan M-Y (eds) Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, ACL 2017, Volume 1: Long Papers. Association for Computational Linguistics, Vancouver, pp 654–664
33. Lin X, Jian W, He J, Wang T, Chu W (2020) Generating informative conversational response using recurrent knowledge-interaction and knowledge-copy. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. Association for Computational Linguistics, Online, pp 41–52. https://www.aclweb.org/anthology/2020.acl-main.6

34. Deriu J, Rodrigo A, Otegi A, Echegoyen G, Rosset S, Agirre E, Cieliebak M (2021) Survey on evaluation methods for dialogue systems. Artif Intell Rev 54(1):755–810. https://doi.org/10.1007/s10462-020-09866-x
35. Liu C-W, Lowe R, Serban I, Noseworthy M, Charlin L, Pineau J (2016) How NOT to evaluate your dialogue system: An empirical study of unsupervised evaluation metrics for dialogue response generation. In: Su J, Carreras X, Duh K (eds) Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, EMNLP 2016. The Association for Computational Linguistics, Austin, pp 2122–2132
36. Wu B, Li M, Wang Z, Chen Y, Wong DF, Feng Q, Huang J, Wang B (2020) Guiding variational response generator to exploit persona. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. Association for Computational Linguistics, pp 53–65
37. Song H, Wang Y, Zhang W-N, Liu X, Liu T (2020) Generate, delete and rewrite: A three-stage framework for improving persona consistency of dialogue generation. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. Association for Computational Linguistics, pp 5821–5831
38. Wang R, Li L, Wang P, Tao X, Liu P (2020) Feature-aware unsupervised learning with joint variational attention and automatic clustering. In: 25th International Conference on Pattern Recognition, ICPR 2020. IEEE, Virtual Event / Milan, pp 923–930
39. Wang R, Li L, Tao X, Dong X, Wang P, Liu P (2021) Trio-based collaborative multi-view graph clustering with multiple constraints. Inf Process Manag 58(3):102466. https://doi.org/10.1016/j.ipm.2020.102466

Fuyong Xu was born in 1997. He is pursuing the master's degree in the College of Computer Science, Shandong Normal University, China. His research focuses on open-domain dialogue generation, reinforcement learning, and sentiment analysis.

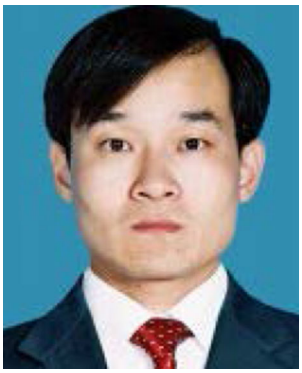
Guangtao Xu is a master student in the School of Information Science and Engineering at Shandong Normal University. His research field is sentiment analysis.

Yuanying Wang was born in 1997. At present, she is studying computer technology in Shandong Normal University of China. Her research focus on dialogue generation and dialogue system.

Ru Wang received the BE degree in computer science and technology from Qufu Normal University and the ME degree in computer application technology from Shandong Normal University in 2015 and 2018. he is currently working toward the PhD degree at the Shandong Normal University, China. He has published several peer-reviewed papers such as the ICPR, IPM, PRL etc. His current research interests include machine learning and data mining. He serves as an reviewer for several journals and conferences like the JMLC, MIPR, PRCV etc.

Qi Ding was born in 1997. She is currently pursuing the master's degree in the College of Computer Science and Engineering, Shandong Normal University, China. Her research interests include recommender system, data mining and robustness research.

Peiyu Liu received the Master degree from East China normal university in 1986. He is currently a Second-level professor, doctoral supervisor, with the School of information science and engineering, Shandong Normal University, China. He is the national outstanding science and technology worker, middle-aged and young expert with outstanding contribution in Shandong province, famous teacher of Shandong province. His research interests include network information security, information retrieval, natural language processing, and artificial intelligence.



Zhenfang Zhu received the Ph.D. degree from Shandong Normal University in 2012, and he was a postdoctoral fellow at Shandong University between 2012 and 2015. He is currently an Associate Professor and a Master's Supervisor with the School of information science and electrical engineering, Shandong Jiao Tong University, China. His research interests include network information security, natural language processing, and applied linguistics.