MDR: Memory Distillation and Reproduction for Personalized Dialogue Generation

Anonymous ACL submission

Abstract

Personalized dialogue generation requires chatbots to generate dialogue content that meets users' personality preferences and aligns with historical interactions. The long conversations pose difficulties for personalized and coherent responses, which becomes more challenging when most current systems generate responses by directly encoding features of various personalized information. To make better use of the correlation between encoded features and actual responses, in this paper the Memory Distillation and Reproduction (MDR) framework is proposed. For sentence feature encoding, we utilize the student encoder to align with and fit the response features encoded by the teacher encoder through knowledge distillation, enhancing the understanding of personality and complex contexts. For response generation, the decoding process is tailored to accommodate the contribution degree of response tokens. Therefore, MDR integrates users' historical dialogue and personalized knowledge to construct up-to-date user profiles. Extensive experiments are conducted on ConvAI2 and Baidu PersonaChat datasets, compared with 8 SOTA competitors through automatic evaluation. The results validate the superiority of MDR in terms of Coherence, Diversity and Consistency. Notably, MDR achieves BLEU-1 19.40 and Coh-Con.Score 37.14 on ConvAI2, and ROUGE-L 27.32 and S-Dist-2 92.08 on Baidu PersonaChat.

1 Introduction

Personalized Dialogue Generation (PDG) is critical for both theoretical research and various applications in the field of natural language processing (Wu et al., 2021; Ma et al., 2021). The main objective of PDG is to produce responses that are not only consistent with the characteristics of personas but also closely coherent with the queries from users (Zhong et al., 2022; Tang et al., 2023). Given user conversations, both dialogue context and user personas can be modeled to generate expected responses, thus the chat system will answer questions according to the specific roles set (see Figure 1).

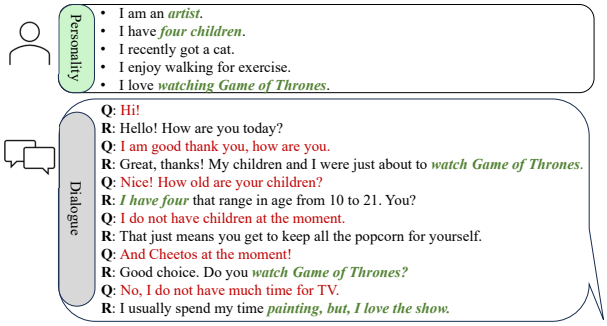

Figure 1: An example of PDG process. At the beginning, the dialogue agent is endowed with specific personality to answer questions (R). Then, the user conduct dialogue with the agent by iteratively querying (Q, marked in red). The content in the dialogue which is consistent with the agent's personality is highlighted in green.

Early works mainly focused on explicit persona modeling (Zhang et al., 2018; Song et al., 2019; Zheng et al., 2020; Song et al., 2021). They relied on pairing dialogue with user-descriptive sentences. However, sufficient data are difficult to collect in practice and the private information is limited, restricting the effectiveness to capture and express more subtle personalization. Subsequent works mostly developed automatic personas extracting (Li et al., 2016; Mazaré et al., 2018; Wu et al., 2019; Tang et al., 2023; Han et al., 2023) to help improve content consistency. Nevertheless, the information in users' historical conversations lacks full utilization. To make better use of historical information, recent researches treat the user dialogue history as an implicit profile (Ma et al., 2021; Wu et al., 2021; Zhong et al., 2022; Liu et al., 2023; Tang et al., 2024; Wang et al., 2024). To achieve this, some methods generated personalized responses by retrieving a subset of relevant historical dia-

logues and integrating them into a decoder (Ma et al., 2021; Wu et al., 2021; Zhong et al., 2022; Liu et al., 2023). However, they either used the most recent dialogues or based on the similarity of the current context, leading some potential weaknesses. Therefore, in practice important personal information may be lost, and unexpected behavior and unmotivated retrieval may exist. To take both persona information and historical conversations into account, (Tang et al., 2024; Wang et al., 2024) attempted to implicitly introduce them to the response generation process.

Unlike existing works, we facilitate PDG in two novel perspectives. First, except for extracting information form query, history, person and response, we carefully consider their pairwise correlation and tailor the encoder input. Second, beyond modeling the implicit but fixed user profiles, we further update the personality with historical dialogue in a knowledge distillation manner and leverage the updated user profiles to reproduce better responses.

In this paper, to effectively combine the explicit personality with the implicit personality within historical information, we propose the Memory Distillation and Reproduction (MDR) framework. Our MDR adopts a Pre-trained Language Model (GPT-2) as the backbone, and performs sentence-level and token-level alignment. To extract the features of dialogue memory, a learnable teacher encoder is used. Based on the similarity between encoded query and history features, a subset of relevant history features are obtained for student encoder. With a shared Memory Net structure (Madotto et al., 2018), the personality features and response features can be effectively updated, to enhance the understanding of long-term memory and complex sentences. To reproduce the characteristics of memory, the gradient of the decoding process is modified according to the contribution degree of response tokens, capturing personalized dialogue patterns and controlling the consistency and coherence for PDG. Experimental results on two PDG benchmark datasets demonstrate that our proposed MDR can outperform the SOTA competitors in terms of main evaluation metrics.

2 Related Work

Dialogue Generation. Applying neural models has become the mainstream for dialogue generation (Gao et al., 2018; Ni et al., 2023; Mo et al., 2024). DialoGPT (Zhang et al., 2019) employs the GPT-2 (Radford et al., 2019) decoder. As DialoGPT is pre-trained on Reddit conversations, it can effectively capture the contextual information in dialogues, thereby generating interesting and human-like responses. However, DialoGPT does not allow for explicit style control over the generated responses. DPDP (He et al., 2024) introduces thinking-intuitive and analytical theory through two complementary planning systems. The first is an instinctive policy model for familiar contexts, while the second is a deliberative Monte Carlo Tree Search (MCTS) mechanism for complex and novel scenarios. COOPER (Cheng et al., 2024) is a new dialogue framework that coordinates multiple specialized agents, where each agent is respectively committed to a specific aspect of the dialogue goal so as to approach complex objectives. Generally, the above models focus on generating coherent responses and pay little attention to personality information.

Personalized Dialogue Generation. To possess personalized characteristics, there are three kinds of representative models for PDG. (1) By using well-defined persona attributes, models can effectively utilize different attributes and realize knowledge-enhanced dialogue generation (Gupta et al., 2020; Tang et al., 2021; Song et al., 2021; Han et al., 2023). However, the persona information is relatively insufficient. At the same time, as persona information is often set fixed, it cannot be associated with the dialogue in a timely manner. (2) By implicitly modeling personality traits from historical dialogue queries, models can incorporate personality in PDG without additional persona information (Wu et al., 2019; Ma et al., 2021; Zhong et al., 2022; Liu et al., 2023). However, in practice it is rather difficult to implicitly obtain personality from the dialogue history without reference objects. (3) Some models combine the implicit personality in historical dialogues with the defined persona attributes (Wang et al., 2024; Tang et al., 2024). Note that our MDR falls into this category.

3 Methodology

Knowledge Distillation in NLP. Knowledge distillation (Hinton et al., 2015) is a technique for model compression and knowledge transfer. Conventional teacher models have been trained on large-scale data, where student models are made to fit the output of teacher models (Sanh et al., 2019). (Wang et al., 2022) adopt an actor-critic-based approach to

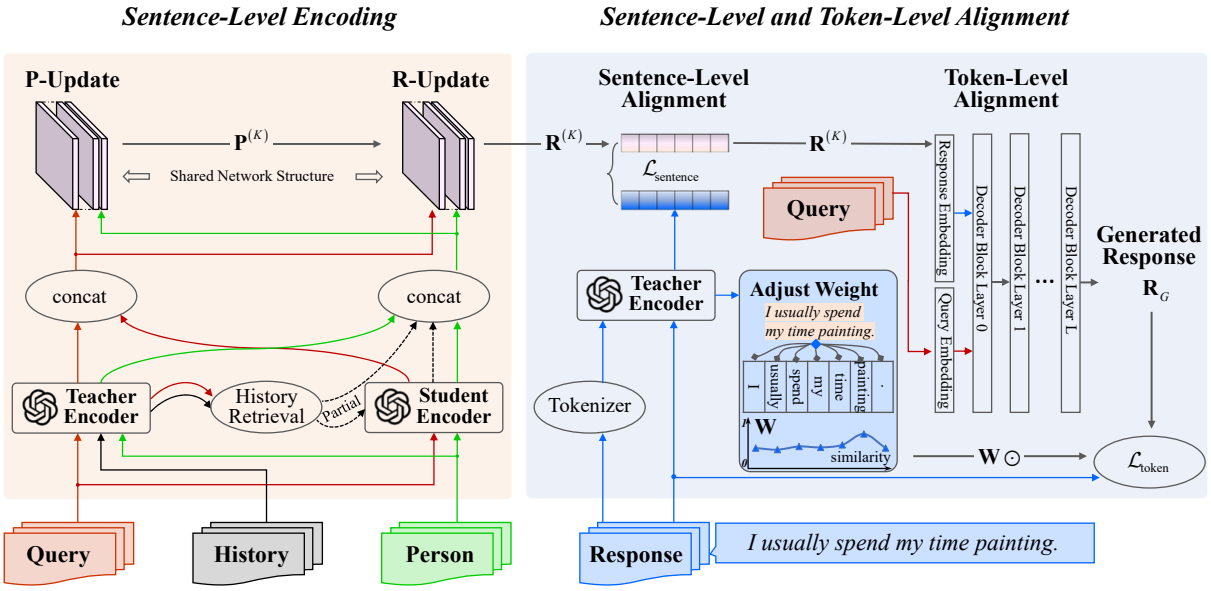

Figure 2: The overall framework of MDR. In sentence-level encoding, student and teacher encoders are employed based on GPT-2, to encode queries, historical dialogues and personality, and update the personality and response features through Memory Net. Then, the updated response is aligned with current response at both sentence-level and token-level.

select the appropriate knowledge to be transferred. TAPIR (Yue et al., 2024) distills the instruction-answering abilities of black-box Pretrained Language Models (PLMs) through multi-task curriculum planning. TAPIR demonstrates the potential of applying PLMs for knowledge distillation. It mainly focuses on aligning the output probability distributions or intermediate features of the teacher PLMs for the same input that the student PLMs learns. Our MDR tailors and enriches the input both for teacher PLMs and student PLMs.

The overall framework of MDR is shown in Figure 2, the proposed MDR mainly comprises a sentence encoding module and an alignment module.

3.1 Problem Definition and Preliminary

Given a T-round dialogue between two users. Without loss of generality, $\mathbf{Q}$, $\mathbf{P}$ and $\mathbf{R}$ denotes query-related, person-related and response-related features, respectively. The dialogue history $\mathbf{H}$ can be represented as $\{(\mathbf{Q}, \mathbf{R})\}_{i=0}^{T-1}$. The goal of PDG is to to generate personalized response $\mathbf{R}(T)$ with respect to query $\mathbf{Q}(T)$, by leveraging $\mathbf{H}$ and $\mathbf{P}$.

3.2 Sentence-Level Encoding

We inherit the merit of knowledge distillation and adopt GPT-2 to construct student and teacher encoders. Note that in our framework, the focus of knowledge transfer has shifted from direct imitation of input-output to imitation of feature generation across different inputs. That is, given current $\mathbf{Q}$, our student PLM leverages $\mathbf{Q}$, $\mathbf{P}$, and $\mathbf{H}$ in the past T-round dialogue to generate current $\mathbf{R}$ feature that matches the ground-truth $\mathbf{R}$ feature generated by our teacher PLM.

Teacher Encoder. To conduct a deep semantic understanding of the responses and extract feature vectors that can represent the essential knowledge of the responses, we only use $\mathbf{R}$ to train encoder and decoder by GPT-2. Thus, the response features encoded by the encoder can be used by the decoder to restore the original response:

$$\mathbf{R}_{\text{TE}} = \text{Encoder}_{\text{teacher}}(\mathbf{R}), \tag{1}$$

$$\hat{\mathbf{R}} = \text{Decoder}_{\text{teacher}}(\mathbf{Q}, \mathbf{R}_{\text{TE}}), \tag{2}$$

$$\mathcal{L}_{\text{teacher}} = -\log p(\hat{\mathbf{R}}|\mathbf{Q}, \mathbf{R}). \tag{3}$$

Before training the whole network of MDR, the teacher encoder is preliminarily trained and its parameters are frozen throughout the subsequent training process.

Student Encoder. Given $\mathbf{Q}$ (the specific requirements of the query), $\mathbf{P}$ (the nuances of the personality), and $\mathbf{H}$ (the semantic content of the dialogue history) as input, the student encoder leverages GPT-2 to encode the semantic features associated with responses.

Since the previous trained teacher encoder can summarize the semantics of the input itself, we first

use it to encode $\mathbf{Q}$, $\mathbf{P}$, and $\mathbf{H}$ as follows:

$$\mathbf{Q}_{\text{TE}} = \text{Encoder}_{\text{teacher}}(\mathbf{Q}), \quad (4)$$

$$\mathbf{P}_{\text{TE}} = \text{Encoder}_{\text{teacher}}(\mathbf{P}), \quad (5)$$

$$\mathbf{H}_{\text{TE}} = \text{Encoder}_{\text{teacher}}(\mathbf{H}). \quad (6)$$

To make better information filtering on dialogue history, given $\mathbf{Q}_{\text{TE}}(T)$ and $\mathbf{H}_{\text{TE}}$ from teacher encoder, we utilize the cosine similarity between $\mathbf{Q}_{\text{TE}}(T)$ and $\mathbf{H}_{\text{TE}}$ to retrieve the most similar content $\mathbf{H}(\text{index})$. Further, we also refer to the historical content that is nearest to the current query in T-round dialogue, i.e., $(\mathbf{Q}(T-1), \mathbf{R}(T-1))$. Then, we combine these two parts and obtain a partial historical content $\mathbf{H}_{\text{P}}$:

$$\mathbf{H}_{\text{P}} = \text{Concat}(\mathbf{H}(\text{index}), \mathbf{Q}(T-1), \mathbf{R}(T-1)),$$
$$\mathbf{H}(\text{index}) = \arg\max_{\text{index}}(\text{cos\_sim}(\mathbf{Q}_{\text{TE}}(T), \mathbf{H}_{\text{TE}})). \quad (7)$$

Subsequently, the corresponding output of student encoder can be obtained:

$$\mathbf{Q}_{\text{SE}} = \text{Encoder}_{\text{student}}(\mathbf{Q}), \quad (8)$$

$$\mathbf{P}_{\text{SE}} = \text{Encoder}_{\text{student}}(\mathbf{P}), \quad (9)$$

$$\mathbf{H}_{\text{PSE}} = \text{Encoder}_{\text{student}}(\mathbf{H}_{\text{P}}). \quad (10)$$

Concatenating the encoding results from student and teacher encoders, we obtain the embedded query feature $\mathbf{Q}_{\text{E}}$ and embedded personality feature $\mathbf{P}_{\text{E}}$ as follows:

$$\mathbf{Q}_{\text{E}} = \text{Concat}(\mathbf{Q}_{\text{TE}}, \mathbf{Q}_{\text{SE}}), \quad (11)$$

$$\mathbf{P}_{\text{E}} = \text{Concat}(\mathbf{P}_{\text{TE}}, \mathbf{P}_{\text{SE}}, \mathbf{H}_{\text{PTE}}, \mathbf{H}_{\text{PSE}}), \quad (12)$$

where $\mathbf{H}_{\text{PTE}}$ denotes the features in $\mathbf{H}_{\text{TE}}$ correspond to $\mathbf{H}_{\text{P}}$. Therefore, the encoding process is not just about simple transformation but rather about capturing the complex correlations between the input elements and the responses.

To capture the up-to-date personality, we design the P-Update module based on Memory Net. As illustrated in Figure 3, this module is utilized to update the personality information that varies with the historical conversations. In this module, there are a total of K HOPs. For $\text{HOP}^{(0)}$, $\mathbf{P}^{(0)}$ is initialized according to a normal distribution. Taking $\mathbf{P}^{(0)}$ as a query for $\mathbf{Q}_{\text{E}}$, then we can use the obtained results as a new query for $\mathbf{P}_{\text{E}}$ and get a personality variant, $\delta\mathbf{P}$. Iteratively, the updated personality feature is $\mathbf{P}^{(k+1)} = \mathbf{P}^{(k)} + \delta\mathbf{P}$, for $k \in [0, K-1]$. In a similar way, we design the R-Update module and finally obtain the response embedding $\mathbf{R}^{(K)}$.

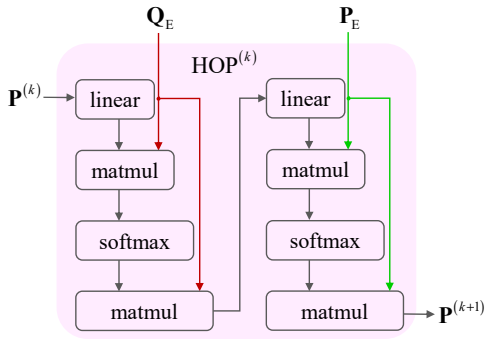

Figure 3: An illustration of the P-Update for $\text{HOP}_{(k)}$. Matrix multiplication is abbreviated to matmul. Note that R-Update and P-Update adopt shared network structure.

3.3 Sentence-Level and Token-Level Alignment

Sentence-Level Alignment. We align $\mathbf{R}^{(K)}$ with $\mathbf{R}_{\text{TE}}$ through the Mean Squared Error (MSE) loss so as to ensure the capacity of feature representation in the knowledge distillation process:

$$\mathcal{L}_{\text{sentence}} = \text{MSE}(\mathbf{R}^{(K)}, \mathbf{R}_{\text{TE}}). \quad (13)$$

Token-Level Alignment. We perform token-level alignment to fine-tune the GPT-2 decoder. As shown in Figure 2, $\mathbf{R}^{(K)}$ and $\mathbf{Q}$ are fed into the network and the generated response $\mathbf{R}_{\text{G}}$ is obtained. Instead of directly calculating the cross-entropy loss between $\mathbf{R}_{\text{G}}$ and $\mathbf{R}$, we further obtain a weight distribution $\mathbf{W}$ for each vocabulary item based on the similarity. In specific, we apply tokenizer and encode each token of $\mathbf{R}$ using the teacher encoder. Then the acquired token-level response features are used to calculate the cosine similarity with sentence-level response feature $\mathbf{R}_{\text{TE}}$ and obtain $\mathbf{W}$.

To make $\mathbf{W}$ normally distributed and facilitate convergence, we further adjust the obtained $\mathbf{W}$. The mean μ is set as the maximum value within $\mathbf{W}$, and the standard deviation σ can be adjusted freely. The objective of adjust weight is to control the requirements for personalization and coherence in generation. Using the adjust weight in a dot product way, the token-level loss is:

$$\mathcal{L}_{\text{token}} = -\frac{1}{N}\sum_{i=1}^{N} w_i \log p(\mathbf{R}_{\text{Tok}}(i)| \mathbf{R}_{\text{Tok}}(<i), \mathbf{Q}, \mathbf{R}^{(K)}), \quad (14)$$

where N represents the number of tokens in the current response and $\mathbf{R}_{\text{Tok}}(i)$ represents a token-level response with the ith token.

3.4 Training Process

The training process is divided into three stages. As mentioned before, firstly the teacher encoder is trained with $\mathcal{L}_{\text{teacher}}$ using response data. Next, sentence-level alignment is carried out through training with $\mathcal{L}_{\text{sentence}}$, performing knowledge distillation. Finally, token-level alignment is carried out through training with $\mathcal{L}_{\text{token}}$, to make the generated response match the truth response.

4 Experiments

4.1 Datasets

We conduct experiments on two widely used benchmark datasets to evaluate the performance of PDG. ConvAI2 (Dinan et al., 2020) is an English dialogue dataset that the conversations therein encompasses abundant personal information. Baidu PersonaChat (Zhang et al., 2018) is a Chinese dataset of personalized dialogue collected and open-sourced by Baidu Co., Ltd. Data preprocessing is conducted following (Tang et al., 2023), and train/valid/test splitting strategy is shown in Table 1.

Dataset	Train	Valid	Test
ConvAI2	43410	4213	2138
Baidu PersonaChat	376016	19923	4456

Table 1: The statistics of datasets.

4.2 Competitors

Non-Personalized Models. As a foundation model of our MDR, the pre-trained GPT-2 (Radford et al., 2019) has exhibited remarkable proficiency in diverse text generation tasks including dialogue related applications.

Models Based on Persona. BoB (Song et al., 2021) leverages the Bert model to perform personalized dialogue generation, and integrates the consistent generation task into the dialogue generation task. CLV (Tang et al., 2023) aggregates the dense persona descriptions into sparse categories, and these categories are then combined with the history query to generate personalized responses. PersonaPKT (Han et al., 2023) directly acquires implicit persona-specific features by representing each persona as a continuous vector from a limited number of dialogue samples generated by the same persona.

Models Based on Dialogue History. DHAP (Ma et al., 2021) utilizes historical memory to store and construct dynamic query-aware user profiles grounded in dialogue history for PDG. MSP (Zhong et al., 2022) enhances PDG by retrieving similar dialogues from similar users via a user refiner and a topic refiner.

Models Based on Persona and Dialogue History. MIDI-Tuning (Wang et al., 2024) models both agents and users through two adapters established on LLMs. The adapters alternate in using their respective utterances and are adjusted through a round-level memory cache mechanism. MORPHEUS (Tang et al., 2024) generates a persona codebook to build a posterior distribution of the role-related information, which concisely represents the roles within the latent space.

Implementation Details. Our framework is implemented using the GPT-2 architecture throughout all components: the teacher encoder-decoder (12-layer Transformer), student encoder, and generation-phase decoder all maintain 768-dimensional embeddings and hidden states. Training is performed on a single NVIDIA A40 GPU with fixed hyperparameters (maximum learning rate of 1e-4 with linear warmup), while exhibiting varying phase durations: 10 minutes per epoch for teacher encoder training, 2 hours per epoch for student encoder training, and 15 minutes per epoch for decoder fine-tuning, with a maximum of 5 epochs per phase. To reduce inference latency, we cache the distilled response features after completing student encoder training. All reported results represent a single execution instance with fixed random seeds (seed=2022). The pre-trained models used in these experiments include gpt2[1], gpt2-chinese-cluecorpussmall[2], bert-base-uncased[3], llama-2-7b[4].

4.3 Evaluation Metrics

Coherence. Widely used BLEU-1 (Papineni et al., 2002) and ROUGE-L (Lin and Och, 2004) measuring the similarity between the generated responses and the ground truth responses are mainly considered for evaluating the coherence of dialogues. Moreover, Coh-Con.Score (Coh-

[1] https://huggingface.co/gpt2

[2] https://huggingface.co/uer/gpt2-chinese-cluecorpussmall2

[3] https://huggingface.co/google-bert/bert-base-uncased

[4] https://huggingface.co/yahma/llama-7b-hf

	BLEU-1	ROUGE-L	Con.S	C-Dist-1	C-Dist-2	S-Dist-1	S-Dist-2	Coh-Con.S
GPT-2	6.77	10.96	56.71	7.35	28.13	68.22	88.81	13.29
BoB	7.85	12.46	62.47	7.24	26.41	63.85	85.02	15.97
DHAP	7.21	9.90	64.27	9.24	**30.98**	69.86	90.23	16.04
MSP	8.19	11.67	65.81	**10.49**	29.96	65.79	89.43	15.4
PersonaPKT	8.70	11.08	60.58	6.30	26.72	-	-	24.87
CLV	11.85	15.10	71.72	5.63	26.91	71.24	92.89	23.01
MORPHEUS	12.67	16.18	**73.19**	5.89	28.74	-	-	31.57
MIDI-tuning	15.63	15.93	55.66	6.96	25.68	61.08	82.83	32.31
MDR(ours)	**19.40**	**19.34**	55.86	5.62	20.25	**72.89**	**93.68**	**37.14**

Table 2: Automatic evaluation on ConvAI2. The best result is in bold face. "-" represents the result is not applicable.

	BLEU-1	ROUGE-L	Con.S	C-Dist-1	C-Dist-2	S-Dist-1	S-Dist-2	Coh-Con.S
GPT-2	10.53	11.29	49.37	5.64	24.98	51.93	84.06	12.14
BoB	14.26	13.30	58.13	5.36	27.45	52.91	82.93	16.33
DHAP	12.96	12.54	55.21	**6.23**	25.37	57.09	85.44	12.30
MSP	15.84	14.06	61.52	5.37	**28.41**	54.06	86.24	14.37
PersonaPKT	13.82	15.57	53.95	2.98	21.83	-	-	19.86
CLV	**24.77**	22.33	60.74	2.42	22.96	60.27	88.15	18.15
MORPHEUS	19.70	24.64	62.45	3.07	23.05	-	-	29.93
MIDI-tuning	21.41	26.50	90.04	3.01	22.58	51.02	79.06	76.66
MDR(ours)	22.41	**27.32**	**93.50**	2.47	18.75	**70.75**	**92.08**	**86.12**

Table 3: Automatic evaluation on Baidu PersonaChat. The best result is in bold face. "-" represents the result is not applicable.

	BLEU-1	ROUGE-L	Con.S	S-Dist-1
(1)	13.87	12.40	30.24	**81.43**
(2)	**21.89**	**19.48**	29.34	56.09
(3)	14.05	12.85	31.10	81.36
(4)	18.79	17.66	**38.25**	66.21

Table 4: Automatic evaluation on ConvAI2 using LLaMA as the backbone, comparing four configurations: (1) LLaMA baseline, (2) LLaMA with sentence-level alignment, (3) LLaMA with both sentence and token level alignment, and (4) MDR. The best result is in bold face.

Con.S) (Tang et al., 2023) is used for evaluation.

Diversity. We adopt Distinct-n (n=1, 2) (Li et al., 2015) to evaluate the diversity of the generated response. Specifically, C-Dist-1/2 and S-Dist-1/2 are adopted, where the former is used to evaluate individual dialogue responses, while the latter for multiple responses (in this paper, five responses are generated).

Consistency. We use Con.Score (Con.S) (Tang et al., 2023) to measure the consistency between the generated responses and persona information.

Note that as PersonaPKT and MORPHEUS only generate a single candidate response, the corresponding S-Dist is not applicable. All the evaluation metrics are positive oriented. For MDR, the standard deviation σ is set to 0.5.

4.4 Main Results

Our proposed MDR is compared with 8 competitors on English and Chinese dialogue datasets, and the performance is shown in Table 2 and 3. Among 8 evaluation metrics, MDR reaches 4 and 6 best results on ConvAI2 and Baidu PersonaChat, respectively. The main results and analyses are briefly summarized in three aspects:

- **(1) Coherence.** In terms of coherence, the performance of our MDR is promising in BLEU-1, Rouge-L and Coh-Con.S. Specifically, compared to the SOTAs with the same GPT2 backbone, improvements of +6.73, +3.16 and +5.57 respectively in BLEU-1, Rouge-L and Coh-Con.S are achieved by MDR on ConvAI2, and improvements of +2.68 and +56.19 respectively in Rouge-L and Coh-Con.S are achieved by MDR on Baidu PersonaChat.

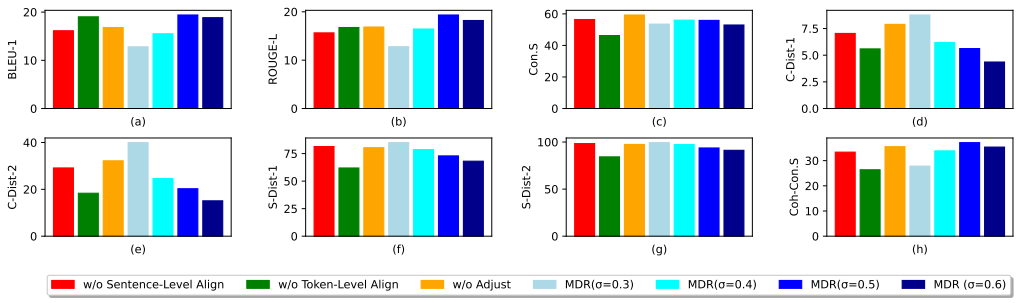

Figure 4: Ablation results on ConvAI2.

This is because, MDR employs knowledge distillation to enable the generated responses to be close to the real responses, enhancing the capacity of feature representation. Even compared with MIDI-tuning which uses a stronger Vicuna-7B as the backbone, MDR also shows advantages in most metrics.

- **(2) Consistency.** Particularly, it can be found that MDR achieves an overwhelming result in Con.S on Baidu PersonaChat, demonstrating MDR can well integrate persona information into response generation with the design of P-Update and R-Update. However, with respect to Con.S on ConvAI2, a declined Con.S should be noticed. This is because MDR takes the historical utterances of the same character as implicit personality traits and combines them with explicit personality data, thus the final personality information containing historical utterances. When training the Con.S classifier, the data with a label of 0 includes the user's response in previous turn, resulting in a lower Con.S result.

- **(3) Diversity.** The performance of MDR obviously varies in C-Dist and S-Dist. For S-Dist-1/2, MDR outperforms other models, which indicates that MDR can generate more diverse and flexible responses than other models when facing the same situations. However, C-Dist-1/2 results of MDR are lower than most models. As only one response for the query is considered when calculating C-Dist, the lower C-Dist results of MDR indicate that our model may make some sacrifices in improving consistency and coherence.

4.5 Ablation Study

To verify the effectiveness of the sentence-level alignment, token-level alignment and adjust weight, ablation experiment is conducted. Figure 4 illustrates the ablation results. Clearly, removing each component leads to a decline in performance. When we remove the token-level alignment and only adopt the response features obtained through sentence-level alignment to decode the responses, we can still obtain satisfactory generation results.

With respect to the impact of adjust weight, MDR variants with different σ are included for comparison. As mentioned in Section 3.3, although the standard deviation σ can be adjusted freely, we experimentally find that σ can influence the model characteristic in PDG. Within a specific range, as σ increases, the coherence of the generated content improves, yet the diversity diminishes. This can be attributed to the fact that when an appropriate σ is selected, it will perform reasonable weighting on the response tokens in $\mathcal{L}_{\text{token}}$, thereby emphasizing important information in the response to improve coherence.

4.6 Effects of MDR Adaptation in LLMs

We systematically investigate the adaptability of MDR to LLMs by constructing a minimalistic yet representative experimental framework. Our architecture employs: LLaMA-2-7B as the decoder, BERT-base as the teacher and student encoders, LoRA for parameter-efficient fine-tuning. As demonstrated in Table 4, comparing four configurations: (1) the base LLaMA model, (2) LLaMA with sentence-level alignment, (3) LLaMA with both sentence and token level alignment, and (4) MDR. The evaluation reveals distinct performance patterns across configurations: Configuration (2) achieves optimal performance on text similarity metrics (BLEU-1: 21.89; ROUGE-L: 19.48), demonstrating that sentence-level alignment effectively enhances response quality by better matching reference texts, though at the cost of reduced diversity; Configuration (3) shows improved diversity

Role Data	I **enjoy hiking**. I've never been abroad. I'm a bank teller. I've never been to college. My favorite phone is the iPhone.	I **like watching cooking** shows. I'm not good at swimming at all. I enjoy roller coasters and skydiving. I'm a vegetarian, and I like Hummers.	I help tend the fields. I've **three Arabian horses. I like to listen to rock music while working. I like to ride horses.**
Dialogue History	Q: Hi, what are you up to? A: I'm getting ready to go out. How about you? What are you doing? Q: I'm **traveling** with my girlfriend right now.	Q: I'm a carpenter, I started my own business. A: That's great! Q: How old are you? I feel you are so young. A: I just turned 17. Q: Oh, you're really young. **what do you do**?	Q: I spent my most of the time in study. A: What are you studying? Q: I am studying horses. A: That's interesting. Q: What is your favorite band? A: I **like rock**. Q: Me too.
Golden-Response	Do you like **traveling** everywhere? I do too	I **don't have a job** yet. I just **moved** here and haven't found one yet.	Do you **like horses**? I have three.
CLV	I often **go out** with my girlfriend too.	I'm out running.	I don't like jazz.
MORPHEUS	Do you like **traveling** everywhere? I do too.	I'm still a student.	Do you like **animals**?
MDR	I **love to travel** too. I have been traveling since a baby. How has your day been?	I am a **stay at home mom**.	I am a **huge gamer**.

Table 5: A case study on ConvAI2.

after fine-tuning, but suffers from degraded coherence, indicating that direct fine-tuning may compromise some aligned response features; Configuration (4) with adjusted weight loss shows superior consistency preservation, indicating this mechanism enables LLaMA to better utilize fitted features while optimally balancing coherence and diversity, ultimately achieving the highest overall response consistency.

4.7 Case Study

Table 5 shows some cases on ConvAI2. In the left case, it can be seen that the result "love to travel" generated by MDR caters to the question, is in line with the personality trait "enjoy hiking", aligns with the semantics of the real response, and also includes a rhetorical question "How has your day been?". In the middle case, the generated response "stay at home mom" is a coherent answer to the question "what do you do", corresponds to the personality trait "like watching cooking shows", and is also consistent with the real responses "don't have a job" and "moved". In the right case, since the query is a meaningless sentence "Me too", the response "huge gamer" generated by MDR corresponds to the personality of being fond of various things and loving to play. The sentences in the golden response also express the personality of being fond of playing.

5 Conclusion

In the paper, we propose memory distillation and reproduction (MDR) framework for PDG. To better leverage the memory within dialogue history, we propose to align the response features of student PLM and teacher PLM through knowledge distillation. The personality traits and response contents are updated through Memory Net, accurately grasping the personality in the current dialogue turn. To enhance the coherence of generated response, we propose to reproduce the response in a token perspective with adjust weight. Extensive experiments on both English and Chinese dialogue datasets verify the effectiveness of MDR, demonstrating its great potential in PDG. In the future, the introduction of more powerful foundation LLMs in our framework will be investigated and the trade-off between efficiency and personalized performance will be considered.

Limitations

MDR significantly advances personalized dialogue generation through its knowledge distillation framework that extracts response features from user profiles and dialogue history, combined with an adaptive loss weighting mechanism to effectively utilize these refined features for response generation. However, we identify three key limitations for future investigation: (1) multimodal dialogue scenarios are not currently supported and require future exploration; and (2) while we validated MDR's compatibility with LLaMA, computational constraints limited our model selection, larger foundation models could further optimize feature alignment and generation quality. These limitations represent natural extensions rather than fundamental shortcomings of the proposed framework.

Ethics Statement

This work develops personalized dialogue generation technology under strict ethical guidelines. All training data is sourced from carefully vetted public datasets, with guarantees that no personally identifiable information is collected or stored. While our system enhances user experience through personalized interactions, we fully acknowledge its potential risks, including: (1) the possible generation of deceptive/harmful "pseudo-personalized" content, and (2) the amplification of social biases present in training data. To mitigate these concerns, we implement multiple safeguards: (1) employing distillation methods to align learned features with authentic responses as closely as possible; (2) designing adaptive weighting mechanisms that prioritize ground-truth tokens; and (3) releasing resources with strictly enforced usage guidelines. We emphasize the necessity for ongoing ethical oversight as the technology advances, particularly regarding potential misuse in sensitive domains.

References

Yi Cheng, Wenge Liu, Jian Wang, Chak Tou Leong, Yi Ouyang, Wenjie Li, Xian Wu, and Yefeng Zheng. 2024. COOPER: Coordinating specialized agents towards a complex dialogue goal. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 17853–17861.

Emily Dinan, Varvara Logacheva, Valentin Malykh, Alexander Miller, Kurt Shuster, Jack Urbanek, Douwe Kiela, Arthur Szlam, Iulian Serban, Ryan Lowe, and 1 others. 2020. The second conversational intelligence challenge (ConvAI2). In *The NeurIPS'18 Competition: From Machine Learning to Intelligent Conversations*, pages 187–208.

Jianfeng Gao, Michel Galley, and Lihong Li. 2018. Neural approaches to conversational AI. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*, pages 1371–1374.

Prakhar Gupta, Jeffrey P Bigham, Yulia Tsvetkov, and Amy Pavel. 2020. Controlling dialogue generation with semantic exemplars. *arXiv preprint arXiv:2008.09075*.

Xu Han, Bin Guo, Yoon Jung, Benjamin Yao, Yu Zhang, Xiaohu Liu, and Chenlei Guo. 2023. PersonaPKT: Building personalized dialogue agents via parameter-efficient knowledge transfer. In *Proceedings of The Fourth Workshop on Simple and Efficient Natural Language Processing (SustaiNLP)*, pages 264–273.

Tao He, Lizi Liao, Yixin Cao, Yuanxing Liu, Ming Liu, Zerui Chen, and Bing Qin. 2024. Planning like human: A dual-process framework for dialogue planning. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 4768–4791.

Geoffrey Hinton, Oriol Vinyals, and Jeffrey Dean. 2015. Distilling the knowledge in a neural network. In *NeurIPS Deep Learning and Representation Learning Workshop*.

Jiwei Li, Michel Galley, Chris Brockett, Jianfeng Gao, and Bill Dolan. 2015. A diversity-promoting objective function for neural conversation models. *arXiv preprint arXiv:1510.03055*.

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*.

Chin-Yew Lin and Franz Josef Och. 2004. Automatic evaluation of machine translation quality using longest common subsequence and skip-bigram statistics. In *Proceedings of the 42nd annual meeting of the association for computational linguistics*, pages 605–612.

Shuai Liu, Hyundong Cho, Marjorie Freedman, Xuezhe Ma, and Jonathan May. 2023. RECAP: Retrieval-enhanced context-aware prefix encoder for personalized dialogue response generation. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 8404–8419.

Zhengyi Ma, Zhicheng Dou, Yutao Zhu, Hanxun Zhong, and Ji-Rong Wen. 2021. One chatbot per person: Creating personalized chatbots based on implicit user profiles. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 555–564.

Andrea Madotto, Chien-Sheng Wu, and Pascale Fung. 2018. Mem2Seq: Effectively incorporating knowledge bases into end-to-end task-oriented dialog systems. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1468–1478.

Pierre-Emmanuel Mazaré, Samuel Humeau, Martin Raison, and Antoine Bordes. 2018. Training millions of personalized dialogue agents. *arXiv preprint arXiv:1809.01984*.

Fengran Mo, Kelong Mao, Ziliang Zhao, Hongjin Qian, Haonan Chen, Yiruo Cheng, Xiaoxi Li, Yutao Zhu, Zhicheng Dou, and Jian-Yun Nie. 2024. A survey of conversational search. *arXiv preprint arXiv:2410.15576*.

Jinjie Ni, Tom Young, Vlad Pandelea, Fuzhao Xue, and Erik Cambria. 2023. Recent advances in deep learning based dialogue systems: A systematic survey. *Artificial Intelligence Review*, 56(4):3055–3155.

Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. BLEU: a method for automatic evaluation of machine translation. In *Proceedings of the 40th annual meeting of the Association for Computational Linguistics*, pages 311–318.

Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, Ilya Sutskever, and 1 others. 2019. Language models are unsupervised multitask learners. *OpenAI Blog*.

Victor Sanh, L Debut, J Chaumond, and T Wolf. 2019. DistilBERT, a distilled version of BERT: Smaller, faster, cheaper and lighter. *arXiv preprint arXiv:1910.01108*.

Haoyu Song, Yan Wang, Kaiyan Zhang, Wei-Nan Zhang, and Ting Liu. 2021. BoB: BERT over BERT for training persona-based dialogue models from limited personalized data. *arXiv preprint arXiv:2106.06169*.

Haoyu Song, Wei-Nan Zhang, Yiming Cui, Dong Wang, and Ting Liu. 2019. Exploiting persona information for diverse generation of conversational responses. *arXiv preprint arXiv:1905.12188*.

Fengyi Tang, Lifan Zeng, Fei Wang, and Jiayu Zhou. 2021. Persona authentication through generative dialogue. *arXiv preprint arXiv:2110.12949*.

Yihong Tang, Bo Wang, Miao Fang, Dongming Zhao, Kun Huang, Ruifang He, and Yuexian Hou. 2023. Enhancing personalized dialogue generation with contrastive latent variables: Combining sparse and dense persona. *arXiv preprint arXiv:2305.11482*.

Yihong Tang, Bo Wang, Dongming Zhao, Jinxiaojia Jinxiaojia, Zhangjijun Zhangjijun, Ruifang He, and Yuexian Hou. 2024. MORPHEUS: Modeling role from personalized dialogue history by exploring and utilizing latent space. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 7664–7676.

Chenglong Wang, Yi Lu, Yongyu Mu, Yimin Hu, Tong Xiao, and Jingbo Zhu. 2022. Improved knowledge distillation for pre-trained language models via knowledge selection. In *Findings of the Association for Computational Linguistics: EMNLP 2022*, pages 6232–6244.

Jian Wang, Chak Tou Leong, Jiashuo Wang, Dongding Lin, Wenjie Li, and Xiaoyong Wei. 2024. Instruct once, chat consistently in multiple rounds: An efficient tuning framework for dialogue. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 3993–4010.

Chien-Sheng Wu, Andrea Madotto, Zhaojiang Lin, Peng Xu, and Pascale Fung. 2019. Getting to know you: User attribute extraction from dialogues. *arXiv preprint arXiv:1908.04621*.

Yuwei Wu, Xuezhe Ma, and Diyi Yang. 2021. Personalized response generation via generative split memory network. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1956–1970.

Yuanhao Yue, Chengyu Wang, Jun Huang, and Peng Wang. 2024. Distilling instruction-following abilities of large language models with task-aware curriculum planning. In *Findings of the Association for Computational Linguistics: EMNLP 2024*, pages 6030–6054.

Saizheng Zhang, Emily Dinan, Jack Urbanek, Arthur Szlam, Douwe Kiela, and Jason Weston. 2018. Personalizing dialogue agents: I have a dog, do you have pets too? *arXiv preprint arXiv:1801.07243*.

Yizhe Zhang, Siqi Sun, Michel Galley, Yen-Chun Chen, Chris Brockett, Xiang Gao, Jianfeng Gao, Jingjing Liu, and Bill Dolan. 2019. Dialogpt: Large-scale generative pre-training for conversational response generation. *arXiv preprint arXiv:1911.00536*.

Yinhe Zheng, Rongsheng Zhang, Minlie Huang, and Xiaoxi Mao. 2020. A pre-training based personalized dialogue generation model with persona-sparse data. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 9693–9700.

Hanxun Zhong, Zhicheng Dou, Yutao Zhu, Hongjin Qian, and Ji-Rong Wen. 2022. Less is more: Learning to refine dialogue history for personalized dialogue generation. *arXiv preprint arXiv:2204.08128*.