Learning to Know Myself: A Coarse-to-Fine Persona-Aware Training Framework for Personalized Dialogue Generation

Yunpeng Li[1,2], Yue Hu[1,2*], Yajing Sun[1,2], Luxi Xing[1,2], Ping Guo[1,2], Yuqiang Xie[1,2], Wei Peng[1,2]

[1]Institute of Information Engineering, Chinese Academy of Sciences, Beijing, China
[2]School of Cyber Security, University of Chinese Academy of Sciences, Beijing, China
{liyunpeng, huyue, sunyajing, guoping, xieyuqiang, pengwei}@iie.ac.cn, xingluxi.xlx@gmail.com

Abstract

A critical challenge for open-domain dialogue agents is to generate persona-relevant and consistent responses. Due to the nature of persona sparsity in conversation scenarios, previous persona-based dialogue agents trained with Maximum Likelihood Estimation tend to overlook the given personas and generate responses irrelevant or inconsistent with personas. To address this problem, we propose a two-stage coarse-to-fine persona-aware training framework to improve the persona consistency of a dialogue agent progressively. Specifically, our framework first trains the dialogue agent to answer the constructed persona-aware questions, making it highly sensitive to the personas to generate persona-relevant responses. Then the dialogue agent is further trained with a contrastive learning paradigm by explicitly perceiving the difference between the consistent and the generated inconsistent responses, forcing it to pay more attention to the key persona information to generate consistent responses. By applying our proposed training framework to several representative baseline models, experimental results show significant boosts on both automatic and human evaluation metrics, especially the consistency of generated responses.

Introduction

Despite the recent success of open-domain dialogue generation by training from large volumes of human context-response pairs data, achieving human-machine conversation is still in its infancy (Li et al. 2016). The performance of the dialogue agent is limited, which makes communicating with them a rather unsatisfying experience for human beings due to the lack of a consistent personality.

To alleviate the inconsistency issue, (Zhang et al. 2018a) introduces the Persona-Chat dataset to endow dialogue agents with predefined personas, which is comprised of a set of sentences describing personal information (as shown in the top of Figure 1). Various approaches have been explored to incorporate explicit personas into dialogue generation, such as using latent variables (Song et al. 2019; Wu et al. 2020) and large pre-training models (Wolf et al. 2019; Liu et al. 2020; Song et al. 2021). Typically, these persona-based dialogue models are trained with Maximum Likelihood Estimation (MLE) to mimic the human context-response pairs in the training corpus. Favoring to generate high-frequency tokens, they still face the persona inconsistency issue due to the nature of persona sparsity in conversation scenarios.

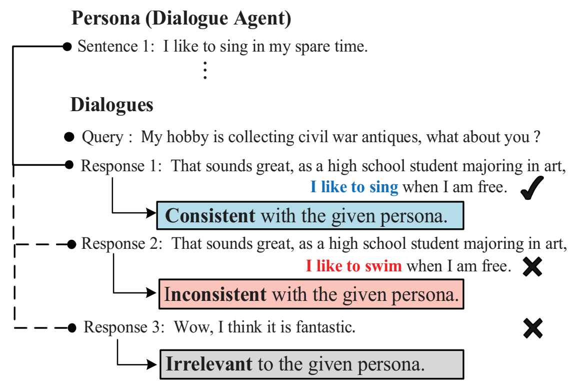

Figure 1: Examples of persona-irrelevant and persona-inconsistent responses.

As a matter of fact, human responses are not always persona-relevant. Even if a certain response talks about the persona, it would also contain much persona-irrelevant information. Therefore, the persona sparsity problem also exists in the crowd-sourced dataset such as Persona-Chat (Zhang et al. 2018a) as all the responses are collected from real conversations. As illustrated in Figure 1, in response 1, only four words "*I like to sing*" are relevant to the persona sentence "*I like to sing in my spare time.*". The problem of persona sparsity hampers the consistency of existing persona-based dialogue models mainly in two aspects:

On the one hand, considering the coarse-grained persona consistency, dialogue agents overlook the persona sentences and generate persona-irrelevant responses. Illustrated in Figure 1, the response 3 fails to perceive the persona sentence 1 and leads to a persona-irrelevant reaction. As the training data only contains a limited number of responses relevant to speakers' personas, the models cannot attend to the persona sentences and may even regard them as noises. On the other hand, considering the fine-grained persona consistency, dialogue agents lack extra attention to the key persona information and generate persona-relevant but inconsistent responses. As shown in response 2 in Figure 1, the wrong generated word "swim" turns the good consistent response into an inconsistent one. The reason is that training with MLE,

*Corresponding Author

the objective function provides equal supervision signal to all the words in the response, weakening the influence of persona-relevant words. Intuitively, a high-quality persona-based dialogue agent should have capabilities: 1) Generate a persona-relevant response given the dialogue context, which is the prerequisite for persona consistency. 2) Pay attention to key persona information to ensure a consistent persona-relevant response. However, existing methods on persona consistency (Shum et al. 2020; Song et al. 2020b,a; Li et al. 2020) only focus on the penalty of inconsistent responses, leaving out the consideration of both coarse-grained consistency and fine-grained consistency, which should be improved in a progressive process.

Different from the existing dialogue agents, we humans can easily maintain consistency during the conversation because of self-consciousness. According to the *Self-Consciousness Theory*[1] (Rödl 2007), a self-conscious human should have the capacity to avoid misidentification, which means he can not only pick himself out but also avoid taking another person to be him. Motivated by the above analysis and this theory, in this work, we propose a coarse-to-fine persona-aware training framework to improve the persona consistency of a dialogue agent progressively. It consists of two training stages: **Learning to Know Myself** and **Learning to Avoid Misidentification**. The central idea of our model is to improve response consistency from learning to know myself coarsely to learning to avoid misidentification finely. Specifically, in the coarse-grained training stage, a set of new persona-dense query-response pairs is first constructed and used to train the dialogue agent, making it highly sensitive to the personas to generate persona-relevant responses. Then, in the fine-grained training stage, the dialogue agent is further trained with a contrastive learning paradigm, which explicitly perceives the difference between consistent and constructed inconsistent responses. Here, the constructed responses are obtained by revising the persona-related words in the golden consistent responses. By this way, the fine-grained training method forces the dialogue agent to pay more attention to the key persona information to generate consistent responses.

Our main contributions are as follows: (1) We propose a novel two-stage coarse-to-fine persona-aware training framework to improve the persona consistency of dialogue agents progressively. (2) To the best of our knowledge, this is the first work to empower the persona consistency simultaneously considering both aspects of coarse-grained and fine-grained. (3) By applying our proposed training framework to several representative baseline models, experimental results show significant boosts on evaluation metrics, especially the consistency of generated responses.

Related Work

Persona-based Dialogue Open-domain chit-chat dialogue generation has attracted growing interest in recent years, where personalization is one of the key points. For making the conversation more engaging, it requires the personalized dialogue agent generate consistent responses with its dialogue history and the given personas(Li et al. 2016). To achieve this goal, (Zhang et al. 2018a) introduces the Persona-Chat dataset, with explicit persona texts to each dialogue. Following this line, (Song et al. 2019) integrates persona texts into the *Persona-CVAE* model for generating diverse responses. (Wolf et al. 2019) builds a dialogue agent *TransferTransfo*, leveraging the advantage of large-scale pretraining model. However, these models still face inconsistency and are hard to evaluate.

To solve this problem, (Welleck et al. 2019) propose the Dialogue Natural Language Inference (DNLI) dataset, which can be used to evaluate the consistency performance of dialogue agents. Then plenty of works start to focus on improving the persona consistency, which can be broadly categorized into reward-based methods and rewriting-based methods. Reward-based methods (Song et al. 2020b; Liu et al. 2020) introduce additional consistency reward information to perform reinforcement learning, while rewriting-based methods (Shum et al. 2020; Song et al. 2020a) perform rewriting operation on the generated sketches to improve the consistency of dialogue. Nevertheless, these models only focus on the penalty of the persona-inconsistent responses without considering the coarse-grained persona consistency introduced in Sec. 1 that dialogue agents overlook the personas to generate persona-irrelevant responses.

Contrastive Learning for Generation Recently, contrastive learning has been a popular method for representation learning. Perceiving the difference between positive and negative input samples, the learned representation can well enhance the generation results of strongly grounded tasks such as machine translation (Wei et al. 2021). However, the semantic inconsistency between dialogue context and response makes it hard to generate better only by enhancing input representation. Motivated by Noise-Contrastive Estimation(NCE) (Gutmann and Hyvärinen 2012), our work belongs to another category of methods that directly performs contrastive learning in the output side, while existing works (Dai and Lin 2017; Cai et al. 2020) concentrate on increasing the diversity of generated texts. Different from these methods, our work aims to improve the fine-grained persona consistency of the dialogue agent. The negative samples are generated by manipulating golden consistent responses instead of being sampled randomly, which effectively forces the dialogue agent to pay attention to the key persona information.

Methodology

Overview

In this work, we propose a coarse-to-fine persona-aware training framework to improve the persona consistency of dialogue agents progressively. Formally, the persona-based dialogue generation task is defined as follows: given a set of personalized speaker information $P = \{P_1, P_2, \ldots, P_T\}$, where T is the number of persona sentences, and a user query Q, the persona-based dialogue agent $\mathcal{M}_\theta$ is trained to generate a response $Y = \{y_1, y_2, \ldots, y_n\}$ consistent with the persona, where n is the number of words of the response.

As shown in Figure 2, our framework consists of two training stages, where:(1) **Learning to Know Myself** is

[1] https://plato.stanford.edu/entries/self-consciousness

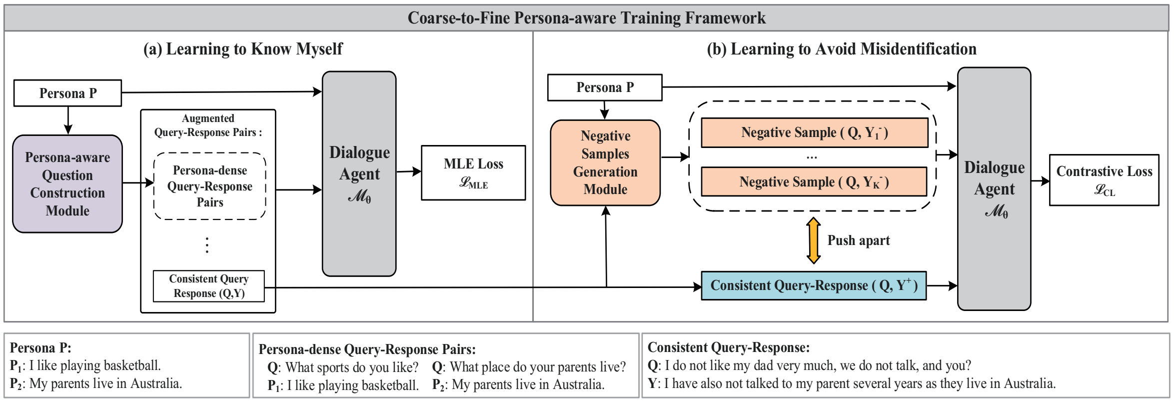

Figure 2: Illustration of our proposed Persona-aware Training Framework adapted to persona-based dialogue agents. (a) is the Coarse-grained training stage, while (b) is the Fine-grained stage.

the coarse-grained stage that trains the dialogue agent to answer the generated persona-aware questions, making it highly sensitive to the personas to generate persona-relevant responses. (2) **Learning to Avoid Misidentification** is the fine-grained stage which further trains the dialogue agent by explicitly perceiving the difference between consistent and constructed inconsistent responses, forcing it to pay attention to the key persona information to generate persona-consistent responses.

Learning to Know Myself

Due to the nature of persona sparsity in conversation scenarios, dialogue agents trained with MLE often overlook the given persona sentences and generate persona-irrelevant responses. To solve this problem, as shown in Figure 2(a), we first proposed a novel **Persona-aware Question Construction Module** to generate persona-dense query-response pairs. Then the dialogue agent is trained to answer the persona-aware questions. The two essential parts are how to construct the persona-aware questions and leverage them for the training of agents.

Persona-aware Question Construction Module This module aims at constructing a set of persona-dense query-response pairs for coarse-grained persona-aware training. With the help of DNLI dataset (Welleck et al. 2019), for each description sentence P_i in a persona P, we associate it with a human-labeled relation triple (e_1, r, e_2) which contains the main fact conveyed by a persona sentence. Then based on this relation triple, a question Q_{P_i} about P_i is constructed by our pre-defined templates. Typically, the relation r in triple (e_1, r, e_2) is the form of $r_1\_r_2$, where r_1 is a verb and r_2 is a noun. Thus, the template we use is "What r_2 do e_1 r_1 ?" For example, as shown in Figure 2, the persona P_1 "`I like playing basketball.`" is associated with the relation triple "`[I, like_sports, basketball]`", while "`My parents live in Australia.`" is associated with the relation triple "`[My parents, live_place, Australia]`". The constructed questions are "*What sports do you like?*" and "*What place do your parents live?*". For generalization, for persona sentences not in DNLI, we can use a relation extraction model (a RoBERTa-large(Liu et al. 2019) model)[2] trained on DNLI to achieve this goal, whose input is the persona sentence P_i and output is the relation r. Then, the entity e_1 will be extracted by keyword matching to form the query.

Finally, for each persona sentence P_i in a persona P, we construct the persona-aware question Q_{P_i}. The Q_{P_i} will be used as a dialogue query while the persona sentence P_i used as its corresponding response to construct the persona-dense query-response pairs $\{(Q_{P_i}, P_i)\}_{i=1}^{T}$.

Coarse-grained Persona-aware Training In the coarse-grained persona-aware training stage, the constructed persona-dense query-response pairs $\{(Q_{P_i}, P_i)\}_{i=1}^{T}$ will be used together with original query-response pairs in the training corpus to augment it. And the augmented pairs will be utilized to train the dialogue agent $\mathcal{M}_\theta$. The loss function of the target response $Y = \{y_1, y_2, \ldots, y_n\}$ is:

$$\mathcal{L}_{\text{MLE}}(Y|Q,P;\theta)] = -\frac{1}{n}\sum_{j=1}^{n} logp(y_j|y_{<j}, Q, P;\theta) \quad (1)$$

where $\mathcal{M}_\theta$ can be parameterized by any dialogue agent. During coarse-grained persona-aware training, as the new query-response pairs are persona-dense, the dialogue agent is forced to be sensitive to the personas, which can be considered as the prerequisite for improving the fine-grained persona consistency.

Learning to Avoid Misidentification

It's observed that the dialogue agent trained with MLE treats all tokens in the responses as equal importance, which may erroneously lead to the generation of inconsistent words in the persona-relevant responses. Inspired by (Rödl 2007), humans need to have the ability to perceive the fine-grained persona information and avoid misidentification correctly. In fine-grained persona-aware training, we propose to construct context-coherent but inconsistent negative samples. The dialogue agent is further trained to improve fine-grained persona consistency with a contrastive learning paradigm,

[2]The fine-tuned RoBERTa-large model achieves 82.3% micro-F1 on our filtered 41 relations which are pre-defined in DNLI.

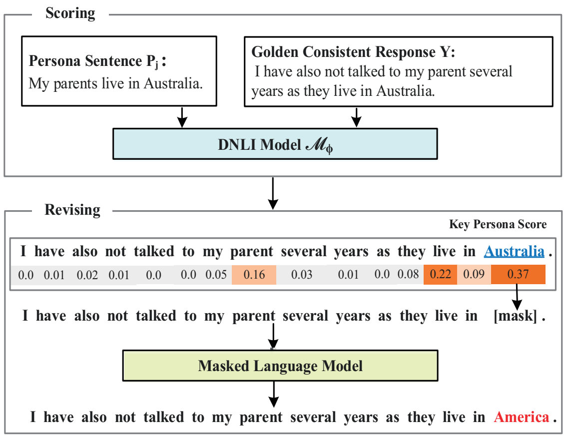

Figure 3: Illustration of our Negative Samples Generation Module. In this example, we select the word "Australia" with the top-1 highest key persona score and revise it to "America" to get a negative inconsistent response.

which explicitly perceives the difference between the positive and the constructed negative samples, forcing it to pay more attention to the key persona information.

Negative Samples Generation Module Ideal negative samples should be context-coherent, but persona-inconsistent. Motivated by (Park et al. 2021), we first detect the key persona word(s) in a golden consistent response with a consistency score, then revise the selected word(s) to construct negative samples, as illustrated in Figure 3.

Scoring In a golden consistent response, the key persona word(s) contribute most to the persona consistency. The scoring process aims to select the words that will lead the persona-consistent response into a persona-irrelevant one if removing them. Specifically, we fine-tune a RoBERTa-large (Liu et al. 2019) model $\mathcal{M}_\phi$ on the DNLI dataset (Welleck et al. 2019), which can estimate the sentence level entailment category of a response for the given persona sentence. Given a consistent response $Y = \{y_1, y_2, \dots, y_n\}$ and its corresponding persona sentence P_j, for each word y_i in response Y, we calculate the *Entailment* score between the persona sentence P_j and (1) the golden response: Y (2) the golden response removing the word y_i: $Y_{/i}$. The difference between two scores is defined as the key persona score of y_i:

$$C_k(y_i) = p(E|[P_j; Y]; \phi) - p(E|[P_j; Y_{/i}]; \phi) \quad (2)$$

Revising After scoring, we select top-k words with highest key persona score as the key persona words. Each key persona word y_i will be replaced with a "$[mask]$" token and regenerated by a Masked Language Model(MLM):

$$y_i^{'} = MLM(Y_{/i}) \quad (3)$$

where the MLM is a pre-trained language model without fine-tuning. If the MLM predicts the original word, the second most likely word is used instead. Thus the generated negative sample is $Y^- = \{y_1^{'}, y_2^{'}, \dots, y_n^{'}\}$. Varying k from 1 to K, we can generate K negative response $\{Y_i^-\}_{i=1}^K$.

	ConvAI2	Persona-Chat	DuLeMon
Language	English	English	Chinese
Training Pairs	131,438	65,719	187,997
Validation Pairs	7,801	7,801	6,110
Testing Pairs	-	7,512	6,118
Avg. # persona P_i	4.49	4.50	3.99

Table 1: Data statistics for ConvAI2, Persona-Chat and DuLeMon datasets. The hidden testing set of ConvAI2 is not publicly available.

Then $\{(Q, Y_i^-)\}_{i=1}^K$ will be used as the negative pairs while (Q, Y) is the positive pair for fine-grained training.

Fine-grained Persona-aware Training Intuitively, the dialogue agent is expected to assign higher conditional probabilities for the positive samples and lower conditional probabilities for those negative samples. Given N_{pos} positive pairs and N_{neg} negative pairs, we perform contrastive learning in the output side, inspired by (Gutmann and Hyvärinen 2012; Dai and Lin 2017). As the dialogue agent $\mathcal{M}_\theta$ is trained in our coarse-grained persona-aware training stage, we do not utilize the target model during training. The contrastive loss function can be defined as:

$$\begin{aligned}\mathcal{L}_{\text{CL}}(Y|Q,P;\theta) = &-\frac{1}{N_{pos}}\sum_{i=1}^{N_{pos}} log[D(Q, Y_i^+);\theta] \\ &-\frac{1}{N_{neg}}\sum_{j=1}^{N_{neg}} log[1 - D(Q, Y_j^-);\theta]\end{aligned} \quad (4)$$

$$D(Q, Y;\theta) = \sigma(log(p(Y|Q,P;\theta))/|l_Y|) \quad (5)$$

where the N_{pos} is 1, N_{neg} is K, $|l_Y|$ is the length of Y and $\sigma(\cdot)$ is the sigmoid activation function. The loss function $\mathcal{L}_{CL}[P_\theta(Y|Q,P)]$ reaches its lower bound when the agent can clearly perceive the difference between positive and negative samples, which indicates that the dialogue agent can pay more attention to the key persona information.

The contrastive loss function only focuses on improving the fine-grained persona consistency. To both ensure the fluency and the consistency of the generated responses, the whole training objective function we use in the fine-grained training stage is the combination of the generation MLE loss and the contrastive loss, which can be defined as:

$$\begin{aligned}\mathcal{L}(Y|Q,P;\theta) = \mathcal{L}_{\text{MLE}}(Y|Q,P;\theta) + \\ \gamma\mathcal{L}_{\text{CL}}(Y|Q,P;\theta)\end{aligned} \quad (6)$$

where $\gamma = 1$ when Y is consistent with the persona P, and $\gamma = 0$ when Y is persona-irrelevant.

Experiments

Experimental Setup

Datasets We conduct persona-based dialogue generation experiments on three publicly available datasets: **ConvAI2** (Dinan et al. 2019), **Persona-Chat** (Zhang et al. 2018a) and **DuLeMon** (Xu et al. 2022). Persona-Chat is obtained from crowd workers on Amazon Mechanical Turk, while ConvAI2 can be considered as an extended version of Persona-Chat. DuLeMon is a crowd-sourced Chinese dataset, where the persona is mainly from the translation and rewriting of the persona in Persona-Chat. We summarize the key statistics of these three datasets in Table 1.

Automatic Metrics To compare different models, we employ BLEU (Papineni et al. 2002) and perplexity (PPL) (Zhang et al. 2018a) to measure the quality and fluency of responses. Lower PPL and higher BLEU mean better quality of the generated responses. We also use Distinct-n (Li et al. 2016) (Dist, n=1,2) which is the ratio of unique n-grams among the corpus, and Entropy-n[3] (Zhang et al. 2018b) (Ent, n=1,2) that is the entropy obtained via the n-gram distribution in a sentence to examine the model's ability to generate diverse responses. Besides, we adopt average Consistency Score(**C**) (Madotto et al. 2019) to evaluated the consistency between a response and persona sentences. This metric follows the default setting and use the output of an NLI model trained on the DNLI dataset (Welleck et al. 2019). For Persona-Chat and ConvAI2 datasets, the NLI model is a RoBERTa-large (Liu et al. 2019), while for DuLeMon, it is a cross-lingual XLM-RoBERTa-large (Conneau et al. 2020).

Human Evaluations We recruit five professional annotators for human evaluation. These annotators are proficient in language tasks but know nothing about the models. We sample 100 $\{$persona, query, response$\}$ tuples for each model's evaluation. Human annotators are asked to evaluate responses from three aspects: (1) Fluency. (2) Coherence. (3) Persona Consistency. The scores for the Fluency and Coherence metrices have three scales, in which 1/2/3 indicates unacceptable/moderate/satisfactory, respectively. The Persona Consistency metric also has three scales, where 1 means the response is persona-consistent, 0 means persona-irrelevant, and -1 means contradicted.

Base Models Because the coarse-to-fine persona-aware training framework is model-agnostic, we apply it to four representative baseline models to validate the effectiveness:

- **Per-Seq2Seq**: A RNN-based Seq2Seq dialogue generation model with context mechanism (Zhang et al. 2018a).
- **Transformer**: A widely-used dialogue generation model based on self-attention mechanism (Vaswani et al. 2017).
- **GPT2**: A pre-trained generation model (Radford et al. 2019) finetuned on personalized datasets.
- **DialoGPT**: A pre-trained generation model (Zhang et al. 2020) specified on dialog domain finetuned on personalized datasets.

The input of the baseline models is the concatenation of the persona sentences and the input query. We compare the original results of these base models with the results after training with our framework.

Compared Methods We also compare the experimental results with other state-of-the-art methods:

- **TransferTransfo** (Wolf et al. 2019)[4] achieves the SOTA performance on automatic metrics in ConvAI2 competition (Dinan et al. 2019).
- **LIC** (Golovanov et al. 2020)[5] achieves the best human evaluations in ConvAI2 competition.
- ρ^2 **bot** (Liu et al. 2020) models the mutual persona perception by reinforcement learning.
- **GPT-D**[3] (Cao et al. 2022) improves the personalized dialogue generation by data augmentation.

Among these compared methods, ρ^2 **bot** and **GPT2-D**[3] are the state-of-the-art (SOTA) methods of ConvAI2 and Persona-Chat, respectively.

Implement Details In our experiments, all the experiments are conducted on NVIDIA GTX 1080Ti GPUs. The implementation of these base models and our training framework are based on ParlAI (Miller et al. 2017). In the negative samples construction module, the hyperparameter K is set as 3. During the responses generation, the beam size is set as 3 with length penalty 1.0. Other detailed configurations of each method are given in Appendix[6].

Results

Main Results Table 2 reports the automatic results comparison on the validation set of ConvAI2 and the testing set of Persona-Chat, while Table 3 reports the results on the testing set of DuLeMon. In specifically, **K** denotes the coarse-grained training stage "Learning to Know Myself" and **A** denotes the fine-grained stage "Learning to Avoid Misidentification". From the results, we can observe that: (1) Our training framework can improve all the base models' performance in response fluency and distinctiveness. (2) The remarkable improvement on consistency score **C** demonstrates the effectiveness of our training framework for improving the persona consistency of dialogue agents. The simplest Per-Seq2Seq achieves competitive persona consistency results with the pre-trained model GPT2 using our training framework. A detailed analysis for the reason is in Section 4.3. Moreover, the human evaluation results are reported in Table 4. The agreement rate from raters is 94.5%, 88.7%, 100% @3, indicating the validity of scores. The human evaluation results are consistent with the automatic evaluation results, which demonstrates the superiority of our training framework. We also conducted Wilcoxon signed-rank tests between our method and the baselines and the results show the improvements are significant with $p < 0.05$.

Few-shot Setting Considering the DuLeMon dataset is relatively larger, we also conduct experiments in a few-shot setting, which are reported in Table 3. We train the dialogue agents using 1/4, 1/8 and 1/16 training data and test them on the original testing set. As the Table shows, our framework can stably improve the performance of base models all the evaluation metrics in the few-shot setting. Surprisingly, even using only 1/8 training data, the persona consistency of the dialogue agents can have competitive results with the agents training with full training data.

[3]https://github.com/golsun/NLP-tools

[4]http://github.com/huggingface/transfer-learning-conv-ai

[5]https://github.com/atselousov/transformer_chatbot

[6]The detailed appendix will be available online.

	Model	PPL	BLEU-1	BLEU-2	BLEU-3	BLEU-4	Dist-1	Dist-2	Ent-1	Ent-2	C
(a)	Per-Seq2Seq	31.50	19.65	9.48	5.61	3.31	1.27	4.67	3.95	5.31	25.47
	Transformer	32.32	18.61	9.24	5.26	3.12	1.00	4.17	4.02	5.43	12.28
	TransferTransfo	17.51	19.13	9.57	5.49	3.61	1.98	8.76	4.32	6.25	46.47
	LIC	-	-	-	-	-	**3.72**	**11.84**	5.32	6.89	39.55
	GPT2	15.18	17.86	9.26	4.94	3.21	2.27	9.12	5.17	6.65	50.99
	ρ^2 bot*	15.12	20.28	9.83	5.91	3.77	2.15	8.67	4.25	6.37	43.85
	DialoGPT	13.52	17.72	8.97	5.01	3.13	2.43	9.70	5.20	6.67	55.19
	Per-Seq2Seq+K+A	27.52	**20.54**	**10.46**	**6.38**	**3.92**	2.01	8.20	4.32	5.96	46.88
	Transformer+K+A	30.63	19.33	9.69	5.79	3.51	1.36	5.76	4.21	5.61	39.84
	GPT2+K+A	14.41	18.95	9.20	5.43	3.36	2.35	10.26	5.22	6.77	54.56
	DialoGPT+K+A	**13.19**	19.70	9.85	5.70	3.52	2.37	11.14	**5.39**	**7.10**	**63.90**
(b)	Per-Seq2Seq	34.66	20.13	10.02	5.67	3.51	0.96	4.74	4.06	5.66	22.96
	Transformer	37.21	18.12	8.47	4.95	2.87	0.56	2.02	3.76	4.93	11.24
	GPT2	15.97	18.74	9.17	5.41	3.40	2.01	8.26	4.49	6.19	47.86
	GPT2-D^{3*}	15.69	-	-	-	-	2.27	9.80	4.21	6.43	55.70
	DialoGPT	13.79	17.81	9.36	5.59	3.56	2.38	9.12	4.78	6.05	49.05
	Per-Seq2Seq+K+A	29.87	**21.70**	**11.41**	**6.51**	**4.03**	1.67	6.53	4.27	5.81	41.72
	Transformer+K+A	33.75	19.82	9.85	5.37	3.18	1.33	5.64	4.15	5.83	28.71
	GPT2+K+A	14.87	19.91	10.35	6.13	3.86	2.35	9.73	**5.04**	**6.54**	**62.22**
	DialoGPT+K+A	**13.44**	20.59	10.67	6.28	3.95	**3.39**	**10.85**	4.95	6.40	61.97

Table 2: Summary of results on two English datasets: (a) ConvAI2 validation set, (b) Persona-Chat testing set. The best results in each column are in bold, while the second are underlined. * denotes the previous SOTA models.

	Full Data				1/4 Data				1/8 Data				1/16 Data			
Model	PPL	B-4	D-2	C	PPL	B-4	D-2	C	PPL	B-4	D-2	C	PPL	B-4	D-2	C
Per-Seq2Seq	18.85	6.86	12.01	48.86	33.10	4.83	7.00	41.39	44.01	3.55	4.32	32.46	62.33	2.62	1.44	16.35
Transformer	26.31	3.71	5.56	41.68	34.61	2.69	2.02	36.16	46.55	2.41	1.19	23.14	71.30	1.41	0.47	12.62
GPT2	**5.56**	5.78	12.20	64.12	6.92	4.96	10.57	60.03	8.02	3.47	10.02	57.32	8.75	2.43	9.02	61.98
Per-Seq2Seq+K+A	18.41	**7.29**	**13.31**	53.20	29.85	5.42	10.35	49.04	38.26	**4.34**	7.05	45.57	48.55	3.29	4.62	34.80
Transformer+K+A	25.76	4.02	6.23	45.87	31.23	3.67	3.70	37.24	42.55	2.82	2.04	29.46	62.97	2.31	0.65	23.27
GPT2+K+A	5.58	6.94	13.09	**69.57**	**6.71**	**5.77**	**12.77**	**66.23**	**7.67**	4.30	**12.48**	**62.11**	**8.05**	**3.85**	**11.04**	**65.84**

Table 3: Summary of results on the testing set of DuLeMon dataset when using different size of training data. The best results in each column are in bold, while the second best results are underlined.

	Model	Fluency	Coherence	Consistency
(a)	GPT2	2.378	2.143	0.508
	LIC	2.358	2.150	0.416
	ρ^2 bot	2.417	**2.233**	0.487
	GPT2+K+A(ours)	**2.482**	2.156	**0.598**
(b)	Transformer	2.231	2.013	0.470
	GPT2	2.467	2.108	0.573
	GPT2+K+A(ours)	**2.503**	**2.172**	**0.632**

Table 4: The results of human evaluation with 100 samples randomly selected from (a) testing set of PersonaChat (b) testing set of DuLeMon.

More Analysis

We further analyze the contributions made by our framework by answering the following questions: (1) **Q1**: How do the two training stages affect the persona consistency of dialogue agents? (2) **Q2**: Is the proposed negative samples construction strategy better than others?

Analysis of Two Training Stages We first study how differently the two training stages **K** and **A** influence the persona consistency of the base models Per-Seq2Seq and GPT2 in Table 5. Here, the **Entail.** and **Contra.** are the proportion of *entailment* and *contradiction* of all the generated responses calculated by the trained NLI model. We conduct ablation studies on two versions of Persona-Chat, **Original Persona** and **Revised Persona**. The latter paraphrases the persona sentences to test whether a model really understands the meaning of a persona. From the results, we can see that (1) Our training framework can improve the persona consistency on both original persona setting and revised persona setting, which means the improvement is from enhancing the dialogue agent's ability for persona understanding. (2) With the coarse-grained training stage **K**, the entailment score rises significantly while the contradiction score slightly goes up. It demonstrates the effectiveness of **K** to make the dialogue agent highly sensitive to the personas to generate persona-relevant responses. (3) The contradiction score drops when we only use **A**, or both use **K+A** to

	Original Persona						Revised Persona					
Model	**PPL**	**BLEU-4**	**Dist-2**	**Entail.↑**	**Contra.↓**	**C**	**PPL**	**BLEU-4**	**Dist-2**	**Entail.↑**	**Contra.↓**	**C**
Per-Seq2Seq	34.66	3.51	4.74	40.88	17.92	22.96	35.77	3.12	4.34	24.00	28.63	-4.63
+K	**29.36**	3.98	5.87	58.64	22.95	35.69	**32.17**	**3.66**	**6.23**	**42.15**	27.52	14.68
+A	37.35	3.26	6.03	42.91	**11.52**	31.39	37.35	3.26	5.76	25.33	**13.91**	11.42
+K +A	29.87	**4.03**	**6.53**	**59.99**	18.27	**41.72**	32.56	3.52	6.18	36.05	16.87	**19.25**
GPT2	15.97	3.40	8.26	71.38	23.52	47.86	18.55	3.32	8.21	30.48	24.44	6.04
+K	**14.56**	**3.89**	9.28	78.21	25.57	52.70	**15.28**	3.65	8.98	**48.21**	25.23	22.98
+A	17.86	3.19	9.67	72.28	**16.53**	55.71	18.86	3.08	9.13	37.57	**15.57**	21.99
+K +A	14.87	3.86	**9.73**	**80.59**	18.37	**62.22**	15.75	**3.77**	**9.65**	47.59	16.36	**31.23**

Table 5: Ablation study on the testing set of Persona-Chat dataset with Original Persona and Revised Persona.

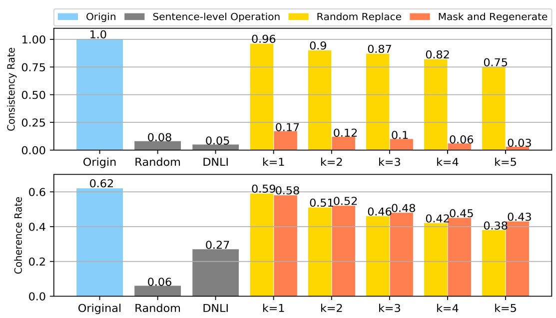

Figure 4: The average consistency rate and coherence rate of different negative samples construct strategies.

train the dialogue agent, proving that the fine-grained training stage **A** can force the dialogue agents pay attention to the key persona information to generate consistent responses.

Analysis of Negative Samples Construction To answer the question **Q2**, we use 1000 responses to compare different negative samples construction strategies: (1) Randomly select other responses as negative samples. (2) Use the human-labeled negative samples from DNLI (Welleck et al. 2019) dataset. (3) Randomly replace words. (4) Mask and regenerate the key persona words, which is our proposed method. The first two strategies are sentence-level while the last two are token-level. Ideal negative samples should be context-coherent but persona-inconsistent. Thus, to evaluate the quality of negative samples, we use two NLI models to calculate the average coherence rate and consistency rate respectively, which are:

$$Rate(\text{Coh./Con.}) = \frac{1}{N}\sum_{i=1}^{N} \text{NLI}(R_i, C_i) \quad (7)$$

Here $\text{NLI}(R_i, C_i) = 1$ when C_i entails response P_i, otherwise 0. The NLI model used for coherence are trained on InferConvAI2 (Dziri et al. 2019), where C_i is the previous utterance of R_i. For consistency, it is trained on DNLI (Welleck et al. 2019), where C_i is the persona sentence. The results are shown in Figure 4, where k denotes the num of changed words. The result suggests that: (1) Our method can well detect the key persona words, as the consistency rate of the generated samples drops significantly compared to the original response and random words replace. (2) In comparison to sentence-level strategies, our method can better keep the coherence of the generated samples. (3) When k improves, the generated responses tend to be general which means context-incoherent and persona-irrelevant. Thus, in our work, to guarantee the quality of the generated samples, we construct 3 negative samples for each response.

Dialogue	
Persona	I drive a ford pickup truck. I am very conservative. I have three guns and love hunting. My family lives down the street from me. I go to church every Sunday.
Context	How are you? Being an old man, I am slowing down these days.
Response	
Seq2Seq	I am good, I am well, and you?
Seq2Seq+K+A	I am good, just got back from **my ford truck**.
GPT2	I am good, **staying at home** this weekend.
GPT2+K+A	I am good. I just got back from a long day of **hunting**, driving my **ford** car.

Table 6: Sampled responses from different models.

Case Study We present some of the generated examples from different models in Table 6, which well demonstrates that our framework can improve the persona consistency of dialogue agents to generate persona-relevant and persona-consistent responses.

Conclusion

In this paper, we propose a two-stage coarse-to-fine persona-aware training framework to improve the persona consistency of a dialogue agent progressively. Specifically, our framework first trains the dialogue agent to answer the constructed persona-aware questions. Then the dialogue agent is further trained with a contrastive learning paradigm by explicitly perceiving the difference between the consistent and the generated inconsistent responses, forcing it to pay more attention to the key persona information to generate consistent responses. By applying our proposed training framework to several representative base models, experimental results show significant boosts on evaluation metrics, especially the consistency of generated responses.

Acknowledgements

We thank all anonymous reviewers for their constructive and insightful comments and we have made some modifications. This work is supported by the National Natural Science Foundation of China (No.U21B2009).

References

Cai, H.; Chen, H.; Song, Y.; Ding, Z.; Bao, Y.; Yan, W.; and Zhao, X. 2020. Group-wise Contrastive Learning for Neural Dialogue Generation. In *Findings of EMNLP 2020*, 793–802. Online.

Cao, Y.; Bi, W.; Fang, M.; Shi, S.; and Tao, D. 2022. A Model-agnostic Data Manipulation Method for Persona-based Dialogue Generation. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 7984–8002. Dublin, Ireland: Association for Computational Linguistics.

Clancey, W. J. 2021. The Engineering of Qualitative Models. Forthcoming.

Conneau, A.; Khandelwal, K.; Goyal, N.; Chaudhary, V.; Wenzek, G.; Guzmán, F.; Grave, E.; Ott, M.; Zettlemoyer, L.; and Stoyanov, V. 2020. Unsupervised Cross-lingual Representation Learning at Scale. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 8440–8451. Online: Association for Computational Linguistics.

Dai, B.; and Lin, D. 2017. Contrastive Learning for Image Captioning. In Guyon, I.; von Luxburg, U.; Bengio, S.; Wallach, H. M.; Fergus, R.; Vishwanathan, S. V. N.; and Garnett, R., eds., *NIPS 2017*, 898–907.

Dinan, E.; Logacheva, V.; Malykh, V.; Miller, A.; Shuster, K.; Urbanek, J.; Kiela, D.; Szlam, A.; Serban, I.; Lowe, R.; Prabhumoye, S.; Black, A. W.; Rudnicky, A.; Williams, J.; Pineau, J.; Burtsev, M.; and Weston, J. 2019. The Second Conversational Intelligence Challenge (ConvAI2). arXiv:1902.00098.

Dziri, N.; Kamalloo, E.; Mathewson, K.; and Zaiane, O. 2019. Evaluating Coherence in Dialogue Systems using Entailment. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, 3806–3812. Minneapolis, Minnesota: Association for Computational Linguistics.

Golovanov, S.; Tselousov, A.; Kurbanov, R.; and Nikolenko, S. I. 2020. Lost in Conversation: A Conversational Agent Based on the Transformer and Transfer Learning. In Escalera, S.; and Herbrich, R., eds., *The NeurIPS '18 Competition*, 295–315. Cham: Springer International Publishing. ISBN 978-3-030-29135-8.

Gutmann, M.; and Hyvärinen, A. 2012. Noise-Contrastive Estimation of Unnormalized Statistical Models, with Applications to Natural Image Statistics. *J. Mach. Learn. Res.*, 13: 307–361.

Li, J.; Galley, M.; Brockett, C.; Spithourakis, G.; Gao, J.; and Dolan, B. 2016. A Persona-Based Neural Conversation Model. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 994–1003. Berlin, Germany: Association for Computational Linguistics.

Li, M.; Roller, S.; Kulikov, I.; Welleck, S.; Boureau, Y.-L.; Cho, K.; and Weston, J. 2020. Don't Say That! Making Inconsistent Dialogue Unlikely with Unlikelihood Training. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 4715–4728. Online: Association for Computational Linguistics.

Liu, Q.; Chen, Y.; Chen, B.; Lou, J.-G.; Chen, Z.; Zhou, B.; and Zhang, D. 2020. You Impress Me: Dialogue Generation via Mutual Persona Perception. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 1417–1427. Online: Association for Computational Linguistics.

Liu, Y.; Ott, M.; Goyal, N.; Du, J.; Joshi, M.; Chen, D.; Levy, O.; Lewis, M.; Zettlemoyer, L.; and Stoyanov, V. 2019. RoBERTa: A Robustly Optimized BERT Pretraining Approach. *CoRR*, abs/1907.11692.

Madotto, A.; Lin, Z.; Wu, C.-S.; and Fung, P. 2019. Personalizing Dialogue Agents via Meta-Learning. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 5454–5459. Florence, Italy: Association for Computational Linguistics.

Miller, A. H.; Feng, W.; Fisch, A.; Lu, J.; Batra, D.; Bordes, A.; Parikh, D.; and Weston, J. 2017. ParlAI: A Dialog Research Software Platform. *arXiv preprint arXiv:1705.06476*.

Papineni, K.; Roukos, S.; Ward, T.; and Zhu, W.-J. 2002. Bleu: a Method for Automatic Evaluation of Machine Translation. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, 311–318. Philadelphia, Pennsylvania, USA: Association for Computational Linguistics.

Park, C.; Jang, E.; Yang, W.; and Park, J. 2021. Generating Negative Samples by Manipulating Golden Responses for Unsupervised Learning of a Response Evaluation Model. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 1525–1534. Online: Association for Computational Linguistics.

Radford, A.; Wu, J.; Child, R.; Luan, D.; Amodei, D.; Sutskever, I.; et al. 2019. Language models are unsupervised multitask learners. *OpenAI blog*, 1(8): 9.

Rödl, S. 2007. *Self-Consciousness*. Harvard University Press.

Shum, M.; Zheng, S.; Kryscinski, W.; Xiong, C.; and Socher, R. 2020. Sketch-Fill-A-R: A Persona-Grounded Chit-Chat Generation Framework. In *Proceedings of the 2nd Workshop on Natural Language Processing for Conversational AI*, 118–131. Online.

Song, H.; Wang, Y.; Zhang, K.; Zhang, W.-N.; and Liu, T. 2021. BoB: BERT Over BERT for Training Persona-based Dialogue Models from Limited Personalized Data. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, 167–177. Online.

Song, H.; Wang, Y.; Zhang, W.-N.; Liu, X.; and Liu, T. 2020a. 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*, 5821–5831. Online: Association for Computational Linguistics.

Song, H.; Zhang, W.; Cui, Y.; Wang, D.; and 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, Macao, China, August 10-16, 2019*, 5190–5196. ijcai.org.

Song, H.; Zhang, W.-N.; Hu, J.; and Liu, T. 2020b. Generating Persona Consistent Dialogues by Exploiting Natural Language Inference. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(05): 8878–8885.

Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, L. u.; and Polosukhin, I. 2017. Attention is All you Need. In Guyon, I.; Luxburg, U. V.; Bengio, S.; Wallach, H.; Fergus, R.; Vishwanathan, S.; and Garnett, R., eds., *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc.

Wei, X.; Weng, R.; Hu, Y.; Xing, L.; Yu, H.; and Luo, W. 2021. On Learning Universal Representations Across Languages. In *9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021*. OpenReview.net.

Welleck, S.; Weston, J.; Szlam, A.; and Cho, K. 2019. Dialogue Natural Language Inference. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 3731–3741. Florence, Italy: Association for Computational Linguistics.

Wolf, T.; Sanh, V.; Chaumond, J.; and Delangue, C. 2019. TransferTransfo: A Transfer Learning Approach for Neural Network Based Conversational Agents. *CoRR*, abs/1901.08149.

Wu, B.; Li, M.; Wang, Z.; Chen, Y.; Wong, D. F.; Feng, Q.; Huang, J.; and Wang, B. 2020. Guiding Variational Response Generator to Exploit Persona. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 53–65. Online: Association for Computational Linguistics.

Xu, X.; Gou, Z.; Wu, W.; Niu, Z.-Y.; Wu, H.; Wang, H.; and Wang, S. 2022. Long Time No See! Open-Domain Conversation with Long-Term Persona Memory. *arXiv preprint arXiv:2203.05797*.

Zhang, S.; Dinan, E.; Urbanek, J.; Szlam, A.; Kiela, D.; and Weston, J. 2018a. 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)*, 2204–2213. Melbourne, Australia: Association for Computational Linguistics.

Zhang, Y.; Galley, M.; Gao, J.; Gan, Z.; Li, X.; Brockett, C.; and Dolan, B. 2018b. Generating Informative and Diverse Conversational Responses via Adversarial Information Maximization. In *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, NIPS'18, 1815–1825. Red Hook, NY, USA: Curran Associates Inc.

Zhang, Y.; Sun, S.; Galley, M.; Chen, Y.-C.; Brockett, C.; Gao, X.; Gao, J.; Liu, J.; and Dolan, B. 2020. DIALOGPT : Large-Scale Generative Pre-training for Conversational Response Generation. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, 270–278. Online: Association for Computational Linguistics.